FP7-ICT-2007-C
Objective ICT-2007.8.0
FET-Open 222107

NIW
Natural Interactive Walking

Publications Booklet

F. Fontana (UNIUD)
J. Cooperstock (McGill University)
A. Lécuyer (INRIA)
S. Serafin (AAU)
V. Hayward, Y. Visell (UPMC)

Sep. 30, 2011
v. 1.0

INTRODUCTION

This booklet has been produced in response to a recommendation coming at the end of the second review period. In that occasion, the project reviewers had lamented the practical difficulty in evaluating the consistency and merit of the literature generated by NIW.

The document contains all the publications that the project has acknowledged as by September 30, 2011. The consortium is confident in that this booklet will be able to disclose the actual value of the literature production that NIW has made possible to publish during its three years of life.

Federico Fontana, Udine, Sep. 30, 2011.

YEAR 1

1. L. Terziman, A. Lécuyer, S. Hillaire, J. Wiener, *Can Camera Motions Improve the Perception of Traveled Distance in Virtual Environments?*, IEEE International Conference on Virtual Reality, Lafayette, US, 2009.
2. G. Cirio, M. Marchal, T. Regia-Corte, A. Lécuyer, *The Magic Barrier Tape: a Novel Metaphor for Infinite Navigation in Virtual Worlds with a Restricted Walking Workspace*, ACM Symposium on Virtual Reality Software and Technology, Yokohama, Japan, 2009.
3. B. L. Giordano, S. McAdams, Y. Visell, J. R. Cooperstock, H.-Y. Yao, V. Hayward, *Non-visual identification of walking grounds*, Acoustics '08 123(5), p. 3412, Paris, France, 2008.
4. A. Law, B. Peck, Y. Visell, P. Kry, J. R. Cooperstock, *A Multi-modal Floor-space for Experiencing Material Deformation Underfoot in Virtual Reality*, IEEE International Workshop on Haptic Audio Visual Environments and Games (HAVE2008), Ottawa, ON, 2008.
5. S. Papetti, F. Fontana, M. Civolani, *A shoe-based interface for ecological ground augmentation*, Haptic and Audio Interaction Design Conference (HAID2009), Dresden, Germany, 2009.
6. C. Drioli, P. Polotti, D. Rocchesso, S. Delle Monache, K. Adiloğlu, R. Anniés, K. Obermayer, *Auditory representations as landmarks in the sound design space*, 6th Sound and Music Computing Conference (SMC2009), Porto, Portugal, 2009.
7. C. Drioli, D. Rocchesso, *Acoustic rendering of particle-based simulation of liquids in motion*, 12th Int. Conference on Digital Audio Effects (DAFx-09), Como, Italy, 2009.
8. S. Serafin, L. Turchet, R. Nordahl. *Extraction of ground reaction forces for real-time synthesis of walking sounds*, 4th Audio Mostly Conference, 2009.
9. Y. Visell, A. Law, J. Cooperstock, *Toward Iconic Vibrotactile Information Display via Floor Surfaces*. Proceedings of World Haptics, 2009.
10. Y. Visell, F. Fontana, B. Giordano, R. Nohrdahl, S. Serafin, R. Bresin, *Sound Design and Perception in Walking Interactions*. Int. J. Human Computer Interaction Studies, 67 (11), Elsevier, 2009.

YEAR 2

11. T. Regia-Corte, M. Marchal, A. Lécuyer, *Can You Stand on Virtual Grounds? A Study on Postural Affordances in Virtual Reality*, IEEE International Conference on Virtual Reality (IEEE VR), Waltham, US, 2010.
12. L. Terziman, M. Marchal, M. Emily, F. Multon, B. Arnaldi, A. Lécuyer, *Shake-Your-Head: Revisiting Walking-In-Place for Desktop Virtual Reality*, ACM Symposium on Virtual Reality Software and Technology (ACM VRST), Hong-Kong, China, 2010.
13. L. Turchet, M. Marchal, A. Lécuyer, S. Serafin, R. Nordhal, *Influence of auditory and visual feedback for perceiving walking over bumps*, ACM Symposium on Virtual Reality Software and Technology (ACM VRST), Hong-Kong, China, 2010.
14. M. Marchal, A. Lécuyer, G. Cirio, L. Bonnet, T. Regia-Corte, M. Emily, *Walking Up and Down in Immersive Virtual Worlds: Novel Interaction Techniques Based on Visual Feedback*, IEEE International Symposium on 3D User Interfaces (IEEE 3DUI), Waltham, US, 2010.
15. A. Law, J. Ip, B. Peck, Y. Visell, P. Kry, J. R. Cooperstock, *Multimodal floor for immersive environments*, ACM SIGGRAPH Emerging Technologies, New Orleans, LA, 2009.
16. R. Rajalingham and Y. Visell and J. R. Cooperstock, *Probabilistic Tracking of Pedestrian Movements via In-Floor Force Sensing*, Seventh Canadian Conference on Computer and Robot Vision (CRV), pp. 143-150, Ottawa, ON, 2010.
17. R. Bresin, A. deWitt, S. Papetti, M. Civolani, F. Fontana, *Expressive sonification of footstep sounds*, 3rd Interactive Sonification Workshop (ISon 2010), Stockholm, Sweden, Apr. 2010.
18. S. Papetti, F. Fontana, M. Civolani, A. Berrezag, V. Hayward, *Audio-tactile Display of Ground Properties Using Interactive Shoes*, Haptic and Audio Interaction Design Conference (HAID2010), Copenhagen, Denmark, Sep. 2010.

19. M. Civolani, F. Fontana, S. Papetti, *Efficient Acquisition of Force Data in Interactive Shoe Designs*, Haptic and Audio Interaction Design Conference (HAID2010), Copenhagen, Denmark, 2010.
20. K. Adiloğlu, C. Drioli, P. Polotti, D. Rocchesso, S. Delle Monache, *Physics-based and spike-guided tools for sound design*, 13th Int. Conference on Digital Audio Effects (DAFx-10), Graz, Austria, 2010.
21. S. Delle Monache, P. Polotti, D. Rocchesso, *A toolkit for explorations in sonic interaction design*, 5th Audio Mostly Conference, Piteå, Sweden, 2010.
22. R. Nordahl, A. Berrezag, S. Dimitrov, L. Turchet, V. Hayward, S. Serafin. *Preliminary experiment combining virtual reality haptic shoes and audio synthesis*. Haptics: Generating and Perceiving Tangible Sensations, 2010.
23. R. Nordahl, S. Serafin, L. Turchet. *Sound synthesis and evaluation of interactive footsteps for virtual reality applications*. Proceedings of the IEEE Virtual Reality Conference, 2010.
24. S. Serafin, L. Turchet, R. Nordahl. *Do you hear a bump or a hole? An experiment on temporal aspects in footsteps recognition*. Proceedings of the 13th Conference on Digital Audio Effects, 2010.
25. S. Serafin, L. Turchet, R. Nordahl, S. Dimitrov, A. Berrezag, V. Hayward. *Identification of virtual grounds using virtual reality haptic shoes and sound synthesis*. Proceedings of Eurohaptics symposium on Haptic and Audio-Visual Stimuli: Enhancing Experiences and Interaction, 2010.
26. L. Turchet, M. Marchal, A. Lécuyer, R. Nordahl, S. Serafin. *Influence of auditory and visual feedback for perceiving walking over bumps and holes in desktop VR*. Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology. ACM, 2010.
27. L. Turchet, R. Nordahl, A. Berrezag, S. Dimitrov, V. Hayward, S. Serafin. *Audio-haptic physically based simulation of walking on dfferent grounds*. Proceedings of IEEE International Workshop on Multimedia Signal Processing. IEEE Press, 2010.
28. L. Turchet, S. Serafin, S. Dimitrov, R. Nordahl. *Conflicting audio-haptic feedback in physically based simulation of walking sounds*. Haptic and Audio Interaction Design (HAID 2010), 2010.
29. L. Turchet, S. Serafin, S. Dimitrov, R. Nordahl. *Physically based sound synthesis and control of footsteps sounds*. In Proceedings of the 13th Conference on Digital Audio Effects, 2010.
30. L. Turchet, S. Serafin, R. Nordahl. *Examining the role of context in the recognition of walking sounds*. In Proceedings of Sound and Music Computing Conference (SMC2010), 2010.
31. Y. Visell, A. Law, S. Smith, J. Ip, J. Cooperstock, *Interaction Capture in Immersive Environments via an Intelligent Floor Surface*. Proceedings of IEEE Virtual Reality (VR'10), 2010.
32. Y. Visell, J. Cooperstock, *Design of a Vibrotactile Display via a Rigid Surface*. Proceedings of the IEEE Haptics Symposium, 2010.
33. Y. Visell, S. Smith, A. Law, R. Rajalingham, J. Cooperstock, *Contact Sensing and Interaction Techniques for a Distributed, Multimodal Floor Display*. Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI'10), 2010.
34. Y. Visell, A. Law, J. Cooperstock, *Touch is Everywhere: Floor Surfaces as Ambient Haptic Interfaces*. IEEE Transactions on Haptics, 2 (3), July-September, 2009.

YEAR 3

35. G. Cirio, M. Marchal, S. Hillaire, A. Lécuyer, *Six Degrees-of-Freedom Haptic Interaction with Fluids*, IEEE Transactions on Visualization and Computer Graphics, in press.
36. A. Lécuyer, M. Marchal, A. Hamelin, D. Wolinski, F. Fontana, M. Civolani, S. Papetti, S. Serafin, *"Shoes-Your-Style": Changing Sound of Footsteps to Create New Walking Experiences*, AIMI Workshop, CHItaly, Alghero, Italy, 2011.
37. G. Cirio, M. Marchal, A. LeGentil, A. Lécuyer, *"Tap, Squeeze and Stir" the Virtual World: Touching the Different States of Matter Through 6DoF Haptic Interaction*, (short paper) IEEE International Conference on Virtual Reality (IEEE VR), Singapore, 2011.

38. G. Cirio, M. Marchal, S. Hillaire, A. Lécuyer, *The Virtual Crepe Factory: 6DoF Haptic Interaction with Fluids*, ACM SIGGRAPH Emerging Technologies (ETech), Vancouver, Canada, 2011.
39. J. Pettré, O. Siret, M. Marchal, A. Lécuyer, *Joyman: an immersive and entertaining interface for virtual locomotion*, ACM SIGGRAPH ASIA Emerging Technologies (ETech), Hong-Kong, 2011.
40. L. Terziman, M. Marchal, F. Multon, B. Arnaldi, A. Lécuyer, *Comparing Virtual Trajectories Made in Slalom UsingWalking-In-Place and Joystick Techniques*, (short paper) Joint Virtual Reality Conference (Joint Eurographics Symposium on Virtual Environments-EuroVR Conference) (JVRC), 2011.
41. M. Marchal, J. Pettre, A. Lécuyer, *Joyman: a Human-Scale Joystick for Navigating in Virtual Worlds*, IEEE International Symposium on 3D User Interfaces (IEEE 3DUI), Singapore, 2011.
42. M. Otis, G. Millet, S. Beniak, J. R. Cooperstock, *Modeling of Lower Limbs for Vibrotactile Compensation*, Canadian Medical and Biological Engineering Conference, 2011.
43. G. Millet, M. Otis, G. Chaw, J. R. Cooperstock, *Initial Development of a Variable-Friction Floor Surface*, Canadian Medical and Biological Engineering Conference, 2011.
44. F. Fontana, M. Civolani, S. Papetti, V. dal Bello, B. Bank, *An exploration on the influence of vibrotactile cues during digital piano playing*, 8th Sound and Music Computing Conference (SMC2011), Padova, Italy, 2011.
45. S. Papetti, F. Fontana, M. Civolani, *Effects of ecological auditory and vibrotactile underfoot feedback on human gait: a preliminary investigation*, Haptic and Audio Interaction Design Conference (HAID2011), Kyoto, Japan, 2011.
46. F. Fontana, F. Morreale, T. Regia-Corte, A. Lécuyer, M. Marchal, *Auditory recognition of floor surfaces by temporal and spectral cues of walking*, 17th International Conference on Auditory Display (ICAD2011), Budapest, Hungary, 2011.
47. S. Papetti, M. Civolani, F. Fontana, *Rhythm'n'shoes: a wearable foot tapping interface with audio-tactile feedback*, Int. Conf. of New Interfaces for Musical Expression (NIME'11), Oslo, Norway, 2011.
48. S. Papetti, F. Avanzini, D. Rocchesso, *Numerical Methods for a Nonlinear Impact Model: A Comparative Study With Closed-Form Corrections*, IEEE Transactions on Audio, Speech, and Language Processing, 19(7), pp. 2146-2158, 2011.
49. R. Nordahl, S. Serafin, L. Turchet, *Sound synthesis and evaluation of interactive footsteps and environmental sounds rendering for virtual reality applications*. IEEE Transactions on Visualization and Computer Graphics, 17(9):1234-1244, 2011.
50. S. Serafin, L. Turchet, R. Nordahl. *Auditory feedback in a multimodal balancing task: walking on a virtual rope*. Proceedings of Sound and Music Computing Conference (SMC2011), Jul. 2011.
51. L. Turchet, S. Serafin. *An investigation on temporal aspects in the audio-haptic simulation of footsteps*. A. Vatakis et al. (Eds.): Time and Time Perception 2010, pages 101-115, 2011.
52. L. Turchet, S. Serafin, *A preliminary study on sound delivery methods for footstep sounds*. Proceedings of the 14th Conf. On Digital Audio Effects Conference, 2011.
53. Y. Visell, B. Giordano, G. Millet, J. Cooperstock, *Vibration Influences Haptic Perception of Surface Compliance During Walking*. PLoS ONE 6 (3), March, 2011.
54. Y. Visell, *Frictional stick-slip oscillation as a first passage problem*. Europhysics Letters (EPL) 95 (6), 2011.
55. Y. Visell, *Walking in virtual worlds: Physics, perception, and interface design*. Ph.D. Thesis, McGill University, Dept. of Electrical and Computer Engineering, 2011. (Introduction and Table of contents included.)
56. F. Fontana, Y. Visell, eds., *Walking with the senses - Perceptual techniques for walking in simulated environments*, Logos Verlag, Berlin, Germany, 2011, in press. (Not included.)

YEAR 1

Can Camera Motions Improve the Perception of Traveled Distance in Virtual Environments?

Léo Terziman*
Bunraku Team
INSA/INRIA
France

Anatole Lécuyer†
Bunraku Team
INRIA
France

Sébastien Hillaire‡
Bunraku Team
Orange Labs/INRIA/INSA
France

Jan M. Wiener§
Center for Cognitive Science
Freiburg University
Germany

ABSTRACT

This paper reports one experiment conducted to evaluate the influence of oscillating camera motions on the perception of traveled distances in virtual environments. In the experiment, participants viewed visual projections of translations along straight paths. They were then asked to reproduce the traveled distance during a navigation phase using keyboard keys. Each participant had to complete the task (1) with linear camera motion, and (2) with oscillating camera motion that simulates the visual flow generated by natural human walking. Taken together, our preliminary results suggest that oscillating camera motions allow a more accurate distance reproduction for short traveled distances.

Keywords: camera motion, walking, perception, distance, first-person-navigation

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Input devices and strategies, Interaction styles, User-centered design; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality

1 INTRODUCTION

Navigation is one of the main interaction tasks in Virtual Environments (VE). In the real world, different senses contribute to the sensation of walking: the vestibular sense, the proprioceptive sense and the efferent copies of the muscular commands [1]. However, most applications of Virtual Reality (VR) cannot afford expensive locomotion interfaces that would provide users with vestibular and proprioceptive cues. Even if passive physical motion appears to be a more effective cue, Harris et al. [2] showed that visual flow also influences the perception of self-motion.

Thus, visual simulations of human walking have been developed, in particular with the use of camera shifts calculated to reproduce the motions of the eyes of a human being during the walk, in response to the oscillations of the user's head produced at each step by natural walking [3][5]. Lécuyer et al. [5] proposed a compensated oscillating camera motion which automatically modifies the angle of the camera in order to keep a fixed point as the focus point of the camera. Using subjective questionnaires they found that oscillating motions were preferred by the participants over the "classical" linear motions. In a second survey, they found that compensated oscillating camera motion was also preferred by the participants over the regular oscillating camera motion.

Hillaire et al. [3] proposed to use an eye-tracking system in order to improve the use of compensated oscillating camera motions. They dynamically adapted the focus point of the camera to match the participants' gaze. By this means they were able to simulate a more realistic camera motion. Still using questionnaires they found that participants have globally preferred the compensated camera motion adapted with the eye-tracking system.

However, as for today, camera motions still lack formal evaluations of their intrinsic properties. For instance, what is the influence of using this type of camera motion on the perception of properties of VE, such as the perception of traveled distances, the sense of orientation, or the perception of scale?

In the remainder of this paper, we will focus on the effect of compensated oscillating camera motions on the perception of traveled distances in virtual environments.

2 EXPERIMENT

The goal of this experiment is to investigate if oscillating camera motions result in an increased accuracy of traveled distance perception in Virtual Environments (VE).

VE are often used in experiments that involve path integration or wayfinding. Indeed, VE enable to control more precisely the parameters of the experiment. For example, Meilinger et al. [6] used a photorealistic virtual environment representing a medieval city in order to study the influence of a set of different distracting tasks on the working memory task involved in memorizing a recently experienced route. Also with the help of VE, Riecke et al. [9] found that, in absence of vestibular and proprioceptive cues, visual flow information alone was sufficient to support path integration for simple navigations, composed of one or two segments. Using VE, some studies could even demonstrate an increase in path integration accuracy for more complex paths [11]. On the other hand, it appears that VR, lacking proprioceptive and vestibular sensations, can also result in systematic misperceptions of the general form of traveled paths [10].

In our experiment, participants were presented with visual projections of a straight path. For half of the trials, the camera moved linearly (as in a car simulation), and for the remaining trials the camera oscillates around three axes in addition to the linear movement in order to simulate visual flow generated by human walking. We used the same compensated oscillating camera motions as proposed in [5] and [3]. The participants' task was to estimate and reproduce the traveled distance.

2.1 Methods

2.1.1 Participants

Twelve participants aged between 22 and 59 (mean=30, standard deviation=12.6) performed this experiment. There were 8 men and 4 women. One participant was left-handed, all the other participants were right-handed. All participants were naive with respect to the experimental setup and to the purpose of the experiment.

2.1.2 Set-up

The experiment was carried out in a darkened room. Participants sat 3 meters in front of a half-cylindrical projection screen of 9.5 meters by 2.4 meters, but only the center of the screen was used (Figure 1).

*e-mail: leo.terziman@irisa.fr
†e-mail: anatole.lecuyer@irisa.fr
‡e-mail: sebastien.hillaire@irisa.fr
§e-mail: mail@jan-wiener.net

The resulting image was 3.25 meters wide and 2.4 meters high, and had a resolution of 1400×1050 pixels. The physical field of view was 55° horizontally and 45° vertically. We used monoscopic rendering, with a frame rate of 50 Hz. The projector used was a Barco Galaxy 7+.

Figure 1: Experimental setup

The experiment was performed within an empty 3D virtual environment. All possible landmarks were removed so to avoid that the sizes of virtual objects give information about depths. The ground was made of a green color (4db34d in hexadecimal RGB format), and the sky of a gray/blue color (e6e6f2). In order to provide participants with visual flow during navigation, a set of 100 000 gray points (4d804d) was displayed on the ground (see Figure 2). Each point had a radius of 5 cm, was randomly positioned in the environment, and had a random lifetime of 5 ± 3 seconds. Using these limited lifetime dots, we avoided all problems of pattern repetitions peculiar to the use of textures.

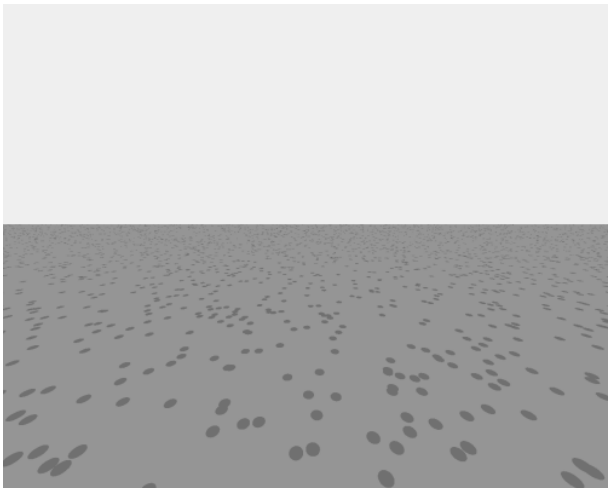


Figure 2: Visual display with random lifetime dots

2.1.3 Distracting task

In order to avoid time counting strategies when estimating traveled distance, participants had to complete a secondary task in parallel to the main task of estimating the traveled distance. During the passive navigation phase, they heard one to four random digits (from "1" to "9"), directly spoken into headphones every 2 ± 1 seconds. The task was to memorize all the digits.

2.1.4 Interaction

Directly after the completion of each passive navigation, participants had to enter the digits they heard with the keyboard and then press the "Enter" key in order to move on to the reproduction phase. During the reproduction phase, participants used the "Up" key (forward translation) and the "Down" key (backward translation) to move through the VE. As long as they pressed the key, the camera continued to move in the given direction at a constant velocity ($V_0 = 1.2\ m.s^{-1}$). The motion stopped when the key was released. Note that there was a short acceleration and deceleration phase when the keys were pressed or released in order to avoid any strong discontinuity in the motion's speed. When participants estimated that they reproduced the correct distance, they validated their choice with the "Enter" key and the next trial started.

2.1.5 Experimental plan

The participants were exposed to two sets of 28 trials: one set for each motion condition, e.g., linear camera motion vs. oscillating camera motion.

- The *linear camera motion* was a linear translation of the camera at a constant speed ($V_0 = 1.2\ m.s^{-1}$) as if the camera was on rails, or as if the user was driving a car.

- The *oscillating camera motion* was the same linear translation as in the linear camera motion ($V_0 = 1.2\ m.s^{-1}$), but additionally, the camera position was slightly moved on the vertical, lateral and forward axes depending on time. The camera motion used corresponds exactly to the model proposed by Lécuyer et al. [5] and Hillaire et al. [3]. The main parameters of the camera motion were set to: 80 cm step length, 6 cm oscillation amplitude on the vertical axis and 2.5 cm oscillation amplitude on the two other axes. The period on the lateral axis was twice the period on the two other axes. These values correspond approximately to the physical values observed for a person of 1.70 m height. The focus point was constantly set to the center of the screen. This point was used to compute a compensation of the camera as explained in [5] and [3].

Participants were divided into 2 groups. One group (n=6) was first presented with the Linear Camera Motion trials, while the other group (n=6) was first presented with the Oscillating Camera Motion trials.

In each condition, the participants were exposed to 7 successive blocks of 4 trials with different distances (5, 7, 10 and 13 meters). In each block, the presentation order of these trials was randomized. During a learning phase, prior to each condition (linear or oscillating camera motion), participants were provided with a set of 4 trials that did not enter in the final data set. The entire experiment lasted approximately 30 minutes.

2.2 Collected Data

The traveled distance and the distance reproduced by the participants (in meters) were automatically recorded at the end of each trial. The digits heard and entered for the distracting task were also recorded.

At the end of the experiment, participants were asked to fill out a subjective questionnaire in which they had to explain which strategies they used to complete the task. Using a seven-point Likert scale they were asked to evaluate both modes of camera motion according to five criteria: (1) *perception of distances*, (2) *immersion feeling*, (3) *impression of realism*, (4) *sensation of walking*, and finally (5) *simulator sickness*.

2.3 Results

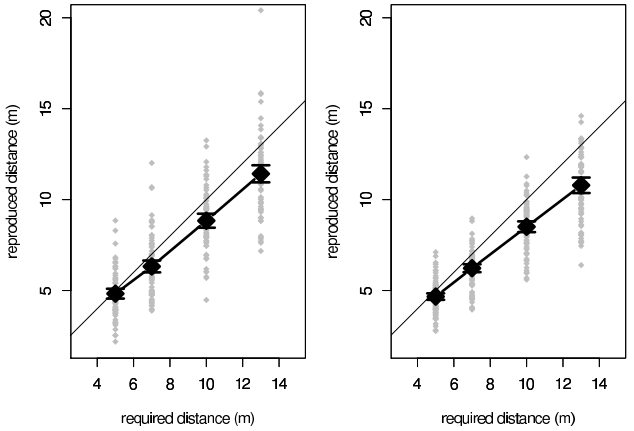
Two different error measures were analyzed: the signed distance error and the absolute distance error. The signed distance error corresponds to the reproduced distance minus the required distance. The absolute distance error corresponds to the absolute value of the reproduced distance minus the required distance.

2.3.1 Signed Distance Errors

The participants' performance in reproducing the navigated distance is characterized by an overall tendency to undershoot the distances (see Figure 3(a) and Figure 3(b)). Indeed, among the 12 participants, 8 participants undershot distances for all conditions and distances. Two participants overshot distances only for short distances and only for the linear camera motion. One participant overshot distances for all distances and only for the linear camera motion. Last, only one participant overshot all the distances for both conditions.

An ANOVA on signed distance errors (factors: navigated distance [5m, 7m, 10m, 13m] and camera motion [oscillating, linear]) revealed a significant effect of navigated distance ($F(1) = 19.0$, $p \ll .001$). This effect reflects that the participants' tendency to undershoot the required distance increased with the length of the traveled distance.

Moreover, as illustrated on Figure 3(a) vs. Figure 3(b), there appear to be more outliers and more overshootings with the linear camera motion than with the oscillating camera motion. This suggests that the use of oscillating camera motions can reduce variance and increase global accuracy. Indeed, for short distances (5m and 7m), the standard deviations of participants' signed responses were found significantly smaller in the oscillating camera motion condition (0.83) than in the linear camera motion condition (1.27) (t(11)=2.71, p=.02). For long distances (10m and 13m) the standard deviations of participants' signed responses did not differ significantly (linear motion condition: 1.60; oscillating condition: 1.38; t(11)=.98, p=.35).



(a) Linear camera motions (b) Oscillating camera motions

Figure 3: Distances reproduced by the participants vs. Required distances for both camera conditions

2.3.2 Absolute Distance Errors

An ANOVA (factors: navigated distance [5m, 7m, 10m, 13m] and camera motion [oscillating, linear]) revealed a significant effect of navigated distance ($F(3) = 16.58$, $p < .001$) as well as a significant interaction between navigated distance and camera motion ($F(3) = 3.02$, $p < .05$). To further characterize the nature of the interaction we pooled data over short distances (5m and 7m) and over long distances (10m and 13m) and analyzed the data again (see Figure 4). A T-test revealed a marginally significant difference for short distances between the oscillating and linear camera motion ($t(11) = 2.05$, $p = .065$). Together with the significant interaction described above, and the reduction in variability, this tendency suggests that distance reproduction performance for short distances is more accurate with oscillating camera motion than with linear camera motion. For long distances, however, both kinds of camera motion rendered similar results.

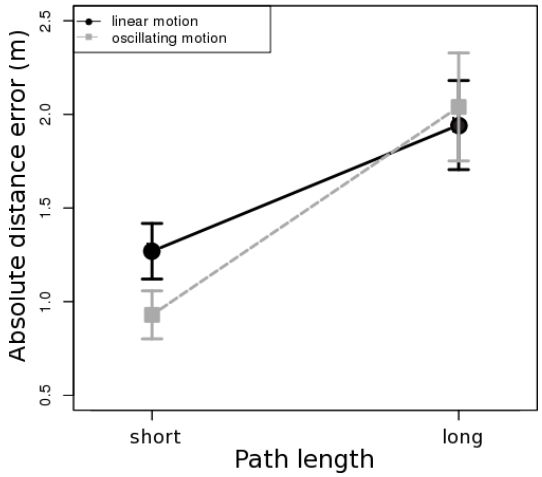


Figure 4: Absolute distance errors for short distances (5m, 7m) and long distances (10m, 13m)

2.3.3 Questionnaire

The results of the subjective questionnaire confirmed the results of Lécuyer et al. [5]. A Wilcoxon signed rank test showed significant differences for *distance perception* ($z = -2.17$, $p = .03$), *walking sensation* ($z = -2.68$, $p < .01$) and a marginally significant difference for *realism* ($z = -1.77$, $p = .08$). For these three criteria, on average participants seem to have preferred the oscillating camera motions (see Figure 5). However, the criteria of *immersion* and *simulator sickness* failed to be significant.

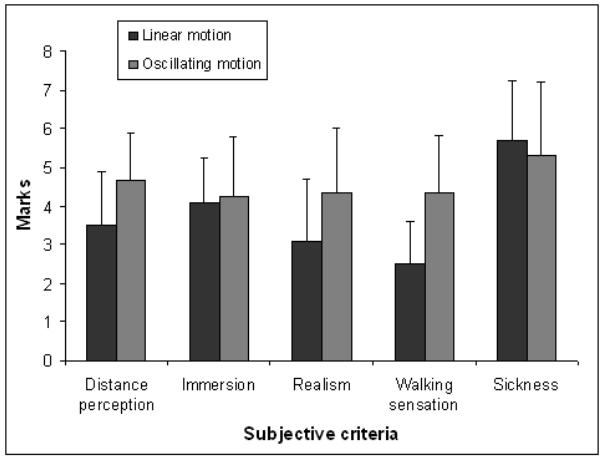


Figure 5: Subjective evaluation of oscillating camera motion vs. linear camera motion

The participants were also asked to explain the strategies they used in order to complete the main task. Eight reported to have

counted the number of steps during the oscillating motion, even if the secondary task made it complicated. In addition to this main strategy, five of them also tried to remember or "feel" the elapsed time for the linear motion or for both motions. Five participants reported to have tried to memorize the "rhythm" of occurrence of the digits of the secondary task, even if the digits appeared randomly. Three of the participants stated that they used a strategy involving the visual flow.

3 Discussion

Results from this study suggest that the oscillating camera motion increases the accuracy of the perception of traveled distance, at least for short distances. This is supported by the significant decrease in standard deviations of participants' responses and the marginally significant decrease of absolute distance error values, in the oscillating camera motion condition and for short distances.

Participants globally undershot traveled distances. Undershooting distances is a result that is often found and discussed in the literature [7] [4]. However, it is sometimes found that participants overshoot traveled distances, such as in [8]. The main differences between our experiment and experiments found in related work (e.g., Riecke et al. [8]) concern the generated visual flow and the field of view. For instance, the field of view used in our setup is smaller than the one used in Riecke et al. [8].

The difference in performance concerning the estimation of traveled distance between oscillating camera motion and linear camera motion is found to be marginally significant. This marginality might be due to the secondary task used. This secondary task is a classical distracting task in experimental studies on distance perception which avoids that participants count the elapsed time. But it also prevent them counting the number of steps with the oscillating camera motion. It is possible that without the secondary task, the accuracy of the reproduced distances could have been increased with the oscillating camera motions. Indeed, if participants would have been able to count precisely the number of steps, they would have been able to reproduce the distances very accurately by reproducing the same number of steps. Moreover, we found that the effect of camera condition was marginally significant only for short distances. One explanation could be that our secondary task was more difficult for the long paths. Indeed, the number of digits that participants had to remember increased with the length of the path. Thus, the complexity increased and it presumably introduced more variance for long paths.

It could also be interesting to conduct this experiment again with an eye-tracking system in order to adapt the oscillating camera motion to the point on screen the user is actually looking at, such as in [3]. This implementation has been subjectively preferred by participants in [3]. Thus, it might also improve the perception of traveled distance.

The results of our subjective questionnaires confirmed some of the results found in [5] and [3]. Our questionnaires showed that participants preferred the compensated oscillating camera motion over the linear camera motion. They found that oscillating camera motions increased both the distance perception and the walking sensation. Besides, the subjective questionnaires showed no significant differences between the oscillating camera motion and the linear camera motion for the immersion criterion. It is maybe a consequence of the limited field of view, or of the visual environment that was not really reallistic (random lifetime dots without other landmarks). The simulator sickness criterion also failed to be significant. One explanation could be the poor diversity of traveled paths. Indeed, traveled paths were only straight lines. Participants never had to turn or make complex trajectories. More complex paths could make them feel more dizzy. Another explanation is that the random lifetime dots seem to have disturbed the participants more than what we expected. The resulting visual flow could be tiring for the eyes and it might have had a high influence on the simulator sickness criterion, more than the camera motion condition.

4 Conclusion

This paper investigated the use of camera motion to improve the perception of traveled distances in VE. We found that the accuracy of the reproduced distances seems to be increased by compensated oscillating camera motions, at least for short traveled distances. Taken together, our results suggest a positive influence of the camera motions on the perception of distances and on the sensation of walking in VE.

Future work will investigate the influence of camera motions on the perception of distances for more complex paths, on the sense of orientation, and on the perception of other properties of virtual environents such as scale. We will also study whether dynamically adapting the focus point by measuring participants' gaze during navigation results in similar or even increased path integration performance.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme under FET-Open grant agreement n° 222107 NIW - Natural Interactive Walking, and the Volkswagen Foundation.

References

[1] L. R. Harris, M. Jenkin, and D. C. Zikovitz. Vestibular cues and virtual environments: Choosing the magnitude of the vestibular cue. In *Proceedings of the IEEE conference on Virtual Reality*, pages 229–236, 1999.

[2] L. R. Harris, M. Jenkin, D. C. Zikovitz, F. Redlick, P. M. Jaekl, U. Jasiobedzka, H. L. Jenkin, and R. S. Allison. Simulating self-motion I: Cues for the perception of motion. *Virtual Reality*, 6(2):75–85, 2002.

[3] S. Hillaire, A. Lécuyer, R. Cozot, and G. Casiez. Using an eye-tracking system to improve camera motions and depth-of-field blur effects in virtual environments. In *Proceedings of the IEEE conference on Virtual Reality*, pages 47–50, 2008.

[4] V. Interrante, B. Ries, and L. Anderson. Distance perception in immersive virtual environments, revisited. In *Proceedings of the IEEE conference on Virtual Reality*, pages 3–10, 2006.

[5] A. Lécuyer, J.-M. Burkhardt, J.-M. Henaff, and S. Donikian. Camera motions improve the sensation of walking in virtual environments. In *Proceedings of the IEEE conference on Virtual Reality*, pages 11–18, 2006.

[6] T. Meilinger, M. Knauff, and H. H. Bülthoff. Working memory in wayfinding - a dual task experiment in a virtual city. *Cognitive Science: A Multidisciplinary Journal*, 32(4):755–770, 2008.

[7] J. M. Plumert, J. K. Kearney, and J. F. Cremer. Distance perception in real and virtual environments. In *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pages 27–34, 2004.

[8] B. E. Riecke, J. Schulte-Pelkum, and H. H. Bülthoff. Perceiving simulated ego-motions in virtual reality - comparing large screen displays with hmds. *Proceedings of the SPIE*, 5666:344–355, 2005.

[9] B. E. Riecke, H. A. van Veen, and H. H. Bülthoff. Visual homing is possible without landmarks: A path integration study in virtual reality. *Presence: Teleoperators and Virtual Environments*, 11(5):443–473, 2002.

[10] B. E. Riecke and J. M. Wiener. Can people not tell left from right in VR? point-to-origin studies revealed qualitative errors in visual path integration. In *Proceedings of the IEEE conference on Virtual Reality*, pages 3–10, 2007.

[11] J. M. Wiener and H. A. Mallot. Path complexity does not impair visual path integration. *Spatial Cognition and Computation*, 6(4):333–346, 2006.

The Magic Barrier Tape: a Novel Metaphor for Infinite Navigation in Virtual Worlds with a Restricted Walking Workspace

Gabriel Cirio
INRIA Rennes*

Maud Marchal
INSA/INRIA Rennes*

Tony Regia-Corte
INRIA Rennes*

Anatole Lécuyer
INRIA Rennes*

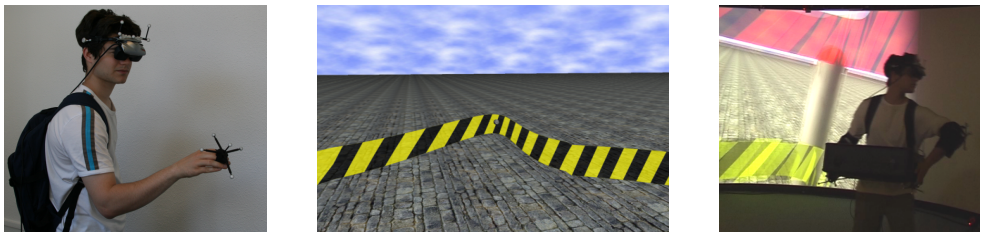


Figure 1: *The Magic Barrier Tape displays the boundaries of the real workspace as a virtual barrier tape, and uses a hybrid position/rate control to travel in the scene. The user (left) "pushes" on the Magic Barrier Tape (center) to move inside the scene when he reaches the workspace boundaries. Any tracked body part can be used to trigger the Magic Barrier Tape (right).*

Abstract

In most virtual reality simulations the virtual world is larger than the real walking workspace. The workspace is often bounded by the tracking area or the display devices. This paper describes a novel interaction metaphor called the Magic Barrier Tape, which allows a user to navigate in a potentially infinite virtual scene while confined to a restricted walking workspace. The technique relies on the barrier tape metaphor and its "do not cross" implicit message by surrounding the walking workspace with a virtual barrier tape in the scene. Therefore, the technique informs the user about the boundaries of his walking workspace, providing an environment safe from collisions and tracking problems. It uses a hybrid position/rate control mechanism to enable real walking inside the workspace and rate control navigation to move beyond the boundaries by "pushing" on the virtual barrier tape. It provides an easy, intuitive and safe way of navigating in a virtual scene, without break of immersion. Two experiments were conducted in order to evaluate the Magic Barrier Tape by comparing it to two state-of-the-art navigation techniques. Results showed that the Magic Barrier Tape was faster and more appreciated than the compared techniques, while being more natural and less tiring. Considering it can be used in many different virtual reality systems, it is an interaction metaphor suitable for many different applications, from the entertainment field to training simulations scenarios.

CR Categories: H.5.2 [Information Interfaces and Presentation]: User Interfaces—evaluation/methodology, input devices and strategies, interaction styles, user-centered design; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—artificial, augmented, and virtual realities;

Keywords: 3D interaction, Interaction metaphor, Navigation, Walking workspace, Hybrid position/rate control

*e-mail: {gcirio,mmarchal,tregiaco,alecuyer}@irisa.fr

Copyright © 2009 by the Association for Computing Machinery, Inc.
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions Dept, ACM Inc., fax +1 (212) 869-0481 or e-mail permissions@acm.org.
VRST 2009, Kyoto, Japan, November 18 – 20, 2009.
© 2009 ACM 978-1-60558-869-8/09/0011 $10.00

1 Introduction

The ability to get from one place to another is a fundamental requirement in both real and virtual environments. In many cases, navigation is not a goal, but rather a mean to reach a location in order to perform a task [Bowman et al. 2004]. In virtual reality, real walking is often the most appropriate navigation interface. It provides the most natural, intuitive and direct way of controlling one's position, matching vestibular and proprioceptive cues from the real movement with the visual feedback from the virtual movement, therefore producing an accurate multi-sensory perception of self-motion. It has been shown that real walking provides benefits over other locomotion techniques [Ruddle and Lessels 2009]. Besides, real walking is, after all, the locomotion interface that we use in our everyday life.

However, in most simulations the virtual world is larger than the real workspace. Navigation techniques become, paradoxically, both a strength and a weakness of current virtual reality systems. They allow individuals to easily and instantly travel long distances and to follow impossible virtual paths. However, they are unable to fully reproduce real life infinite walking capabilities since the user can quickly reach the boundaries of his real workspace. Hence, since real walking alone is not possible, the illusion of a virtual world is lost through the use of unnatural techniques usually coming from the 2D realm.

There are hardware and software-based approaches to overcome the problem of a restricted size workspace. Locomotion interfaces [Hollerbach 2002] such as treadmills often have major limitations that restrict their widespread use (huge size and weight, high cost, lack of accuracy). Navigation techniques based on input devices [Bowman et al. 2004] often fail at providing a simple, intuitive and immersive interaction.

In this paper, we propose a natural metaphor for locomotion in restricted size workspaces, using real walking in position control when inside the workspace, and an interaction technique in rate control at the limits of the workspace. The main idea is to use a well-known real world object, the barrier tape, and its well-known association to the "do not cross" message. The technique visually and clearly defines the workspace where the user can freely walk by surrounding it with virtual barrier tape. When the user reaches the virtual barrier tape, he can move farther in the scene by "pushing"

on the virtual barrier tape. Hence, the technique allows the navigation in an unlimited virtual space, allowing real walking when inside the workspace boundaries, providing an environment safe from collisions with the displays or tracking data loss, and this in a natural and efficient way, without break of immersion.

This paper starts with an overview of related work focusing on locomotion interfaces and, mainly, 3D navigation techniques. It is followed by a description of the design and implementation of the virtual barrier tape. Then, after describing extensions made to existing techniques to be able to compare our contribution with the current art, we report on the experiments conducted for the evaluation of the Magic Barrier Tape. The paper ends with a general discussion on the results of the experiments, some perspectives and a conclusion on this study.

2 Previous Work

There has been a large amount of previous work in the field of locomotion interfaces [Hollerbach 2002]. These devices allow users to be self-propulsed through a repetitive gait while staying in place by canceling the user's motion. Thus, they can provide kinesthetic feedback to the user. Some locomotion interfaces provide linear walking capabilities, such as treadmills and pedaling devices. Although the proprioceptive feedback matches real world movements, they are limited to a 1 degree of freedom (DOF) motion. Omnidirectional walking is possible with 3DOF treadmills and foot platforms [Hollerbach 2002], together with some foot-wearable devices [Iwata et al. 2007]. For the reasons mentioned in the introduction, namely their size and weight, their cost or their accuracy, they have not yet been widely adopted outside the laboratory.

Passive locomotion techniques were surveyed by Bowman et al. [Bowman et al. 2004]. These travel techniques allow the user to navigate inside a virtual world without moving from his real-world position. Among rate control techniques, the most common and widely used technique is the flying vehicle metaphor [Ware and Osborne 1990], often coupled to an input device such as a wand through which the user controls his speed and orientation inside the virtual world. Using position control, the eyeball-in-hand and the scene-in-hand [Bowman et al. 2004] techniques map the camera and the scene respectively to the user's hand. A clutching mechanism is required to navigate beyond one's arm length. The World-in-Miniature technique [Pausch et al. 1995] provides a hand-held miniature of the scene, through which the user can select the location he wants to navigate to, and then be taken to that location in the real size virtual environment. Although usually intuitive and efficient, these navigation techniques are often inadequate for real world simulation scenarios, since their metaphors do not match real world navigation.

Active locomotion techniques, oriented towards real walking, provide natural metaphors by adapting real walking to restricted size workspaces. Moreover, the vestibular and proprioceptive feedbacks produced by self-propulsion increase the realism of the techniques and therefore the degree of immersion of the simulation. The most basic example is the Walking-in-Place technique [Slater et al. 1995], where the user simulates the physical act of walking by stepping in place but without forward motion of the body, making a gesture that is interpreted as a virtual forward, backward or sidestep motion. However, although the technique has the advantage of not having to deal with workspace limits, it falls short regarding kinesthetic feedback. By scaling the user's speed along his intended direction of travel, the Seven League Boots [Interrante et al. 2007] technique implements a scaling technique for real walking in order to reach virtual places beyond the real world workspace. The technique does not solve the limited workspace problem, and precise navigation can become very difficult. The Step WIM [LaViola et al. 2001] takes a different approach on scaling by adapting the World-In-Miniature technique to real walking. The miniature world is drawn on the floor, and the user walks to the new destination on the miniature, instead of using his hands. Although the technique involves real walking, the metaphor might not be considered as natural. Resetting techniques such as the Freeze-backup [Williams et al. 2007] use a clutching approach by freezing the scene while the user recenters his position in the real world once he has reached the limits of the workspace. Other resetting techniques such as the 2:1-Turn [Williams et al. 2007] map a 360° virtual rotation to a 180° real rotation to keep walking on the same direction in the virtual world while taking the opposite direction in the real one. These resetting techniques are performed consciously by the user following a warning signal, which implies a break of immersion. Moreover, the resetting itself might feel unnatural: there is a sudden change in locomotion direction and orientation that does not correspond to the natural movement.

Redirected Walking [Razzaque et al. 2001] and Motion Compression [Nitzsche et al. 2004] techniques solve many of the aforementioned problems by forcing the user into walking in a curved path in the real world when walking in a straight line in the virtual world through the progressive rotation of the scene around him. In a sufficiently large workspace, and with a straight virtual path, the user can walk endlessly without reaching the limits of the real workspace. These techniques are natural and in some cases imperceptible. However, they require large workspaces, can be confusing when doing unpredictable or quick changes of direction, and may require distracting events. In practice, they are more suited for Head Mounted Displays and wide area tracking systems.

3 The Magic Barrier Tape

We propose a novel interaction metaphor, the Magic Barrier Tape, that brings a solution to immersive infinite walking in a restricted workspace through a natural and efficient metaphor.

Walking workspaces of virtual reality systems are often bounded by the tracking area, the display devices or by the walls of the immersive room. Hence, the Magic Barrier Tape has two fundamental objectives. The first one is to inform and display the limits of the workspace in a natural way, without break of immersion, in order to avoid the collision with physical objects outside the workspace boundaries or leaving the tracking area. The second one is to provide an integrated navigation technique to reach any location in the virtual scene, beyond the walking workspace.

To overcome the mismatch between the restricted size workspace and the potentially infinite size of the virtual scene, we followed the concept of hybrid position/rate control [Dominjon et al. 2005], used in a different context for object manipulation, where position control is used inside the available workspace for fine positioning, while rate control is used at the boundaries for coarse positioning. This concept can be found in common desktop applications and games where the mouse switches to rate control when it reaches the edge of the screen: in a file manager when doing multiple selection, or in top-view strategy games such as Starcraft when panning on the map. In our context, we applied the concept to navigation, with the available workspace being the walking workspace. The boundaries of the workspace are represented by a virtual barrier at mid body height textured with slanted black and yellow stripes, evoking the use of barrier tape and its implicit message: "do not cross".

The real workspace, delimited by the physical boundaries, is mapped to a virtual workspace inside the scene, delimited by the virtual barrier tape. Inside the workspace, we use position control: the user can freely walk, and objects inside the virtual workspace

can be reached and manipulated through real walking and real life movements. When reaching the boundaries of the workspace, we switch to rate control: the user can move farther in the scene by "pushing" on the virtual barrier tape, hence translating the virtual workspace in the scene. He can then perform a task inside the virtual workspace at the new location.

The Magic Barrier Tape concept is not subject to a specific technology. It can be implemented in many different virtual reality systems. Any object or body part can be used as an actuator for the virtual barrier tape, depending on the application, and the rate control law can be fitted to specific behavioral needs. In the remaining of this section, we detail the Magic Barrier Tape concept. We take as implementation example our own virtual reality environment, consisting of a Head Mounted Display (HMD) with a 1.5m radius cylindrical tracking space, and one of the user's hands as actuating object.

3.1 Display of the Workspace Limits

The boundaries of the workspace are displayed through 3 complementary visual cues: the main virtual barrier tape, the warning virtual barrier tape, and their grey shadow on the floor.

The main virtual barrier tape is presented as a band that matches the shape of the workspace boundaries, such as a square for a CAVE or a circle for a cylindrical tracking system. It is positioned at a safe distance ahead of them, high enough from the virtual floor so that the user does not need to look down to see the barrier tape, and low enough so that it does not occlude the user's forward vision. The boundaries of the workspace are therefore clearly and continuously visible. The tape is made slightly translucent so what would have been normally hidden by the tape is still discernible.

The warning virtual barrier tape appears when the user's body is close to the main tape, as a warning signal. This second tape has the same shape and origin than the main one, and has a red glow to capture the user's attention. For the same reason, it is positioned at the user's eyes height. The tape is fully transparent when the user is at a reasonably safe distance from the main tape, and becomes progressively opaque as the user gets closer, therefore making the warning signal also progressive, from dim to strong. The warning virtual barrier tape is complementary to the main tape, since it is triggered as a safety measure, and it gives an idea of when to stop walking and start "pushing".

The tapes shadow is drawn on the floor as if the barrier tapes were lit from above, in order to have a visual cue about the limits of the workspace when the user looks down. Hence, at least one of the 3 visual components is always visible at almost any viewing direction, which is particularly helpful with an HMD setup where there is usually a narrow field of view. Figure 2 shows the three components of the Magic Barrier Tape: the main barrier tape, the warning barrier tape (here visible) and the tapes' shadow.

In our virtual reality environment implementation, the main virtual barrier tape is 30 cm high and at 30 cm from the boundaries. It is shaped as a ring with a 1.2m radius and the center of the tracking area as origin. It is positioned at 1.3m from the virtual floor. The warning tape is activated when the user is at 30 cm from the main tape.

3.2 Navigation Through Rate Control

The Magic Barrier Tape allows the use of position control inside the workspace, and rate control at the boundaries. The user is switched from position control to rate control whenever his hand (or any other tracked body part) penetrates the boundaries represented

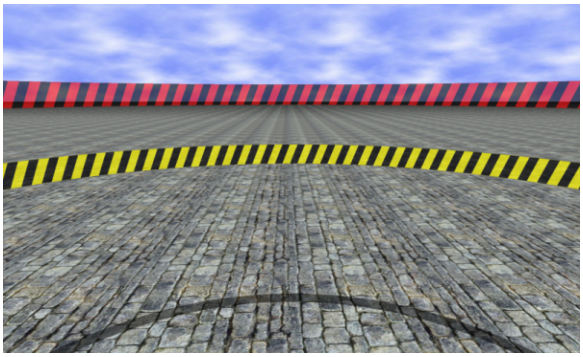


Figure 2: *The three Magic Barrier Tape visual cues to show the workspace boundaries: the main virtual barrier tape (middle), the warning tape (top) and the tapes shadow (bottom).*

by the virtual barrier tape. The speed of the resulting translation in the virtual scene is a function of the hand penetration distance. When the user's hand is pulled back inside the workspace, the user is switched back to position control.

The virtual barrier tapes (main and warning) are deformed when the user's body (preferentially, the hand) penetrates the boundaries. This elastic behavior allows the user to see how deep he is "pushing", and therefore to evaluate how fast he will move in the virtual scene. A visual feedback on the rate control is also important so the user can know where the neutral position is located [Dominjon et al. 2005].

The deformation follows the shape of a centered Gaussian curve D, of equation:

$$D(p) = p\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}$$

where p is the penetration length (in meters), and σ the standard deviation, which controls the "width" of the deformation. The virtual barrier tape is rotated so that the center of the Gaussian curve matches the penetration point P, the collision point between the hand and the virtual barrier tape. Therefore, as shown in Figure 3, the Gaussian deformation is centered around the penetration point, and its symmetry axis is given by the $\overrightarrow{OP}$ direction, where O is the center of the virtual barrier tape. Since the deformation follows the user's hand, the Gaussian curve has to be shifted to take into account the lateral deviation of the hand position H with respect to the Gaussian axis, as shown in Figure 3. The final result gives the impression of having an elastic region around the penetration point than can be deformed in any direction. This deformation direction, $\overrightarrow{PH}$, gives the travel direction of the virtual workspace (Figure 3).

The velocity V, a function of p, gives the speed of travel. It has the following equation:

$$V(p) = k * p^n$$

where k and n are constants. We use a polynomial function in order to have both slow speed when the user is close to the boundaries for small distances, and high speeds to move fast for distant targets. In our implementation, after preliminary testing, we used $\sigma = 0.15$, $k = 1.4$ and $n = 3$.

Our Magic Barrier Tape implementation provides both a safe walking environment and a natural and efficient navigation technique.

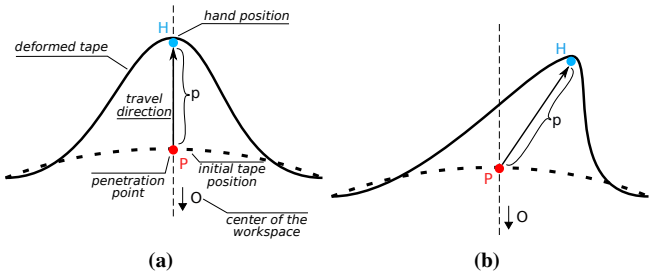


Figure 3: *The Gaussian deformation of the Magic Barrier Tape in top-view (a) and its shifted version (b) to follow the hand position.*

4 Extending Resetting Techniques for Omni-Directional Walking

In order to conduct the evaluation of the Magic Barrier Tape, we chose similar purpose techniques among existing navigation techniques. Among the surveyed active navigation techniques, based on real walking, only the resetting techniques developped by Williams et al. [Williams et al. 2007] provide collision free and infinite navigation capabilities. However, the resetting techniques were originally designed for straight paths and right angle turns, where in most virtual reality applications the user is allowed to freely explore his surrounding virtual environment, taking arbitrary paths and freely rotating around him. For fair comparison throughout the evaluation, since our Magic Barrier Tape technique enables such a navigation, we propose to add visual cues to these techniques in order to make them well suited for omni-directional navigation.

Extended Freeze-Backup Technique. In the original Freeze-backup technique, in order to reset his position the user has to walk backwards in a straight line, until he reaches the resetting position. Since he is not guided while walking backwards, paths can only be straight. Otherwise, he could reach the workspace boundaries prematurely and find himself "locked" in a very short path resetting loop.

In the extended Freeze-backup technique, backups now need to take the user to the center of the real workspace. Before the reset, the user can be at any position in the real workspace, and with any orientation. Hence, we propose to add visual cues to guide the user through his resetting motion, which is divided in two steps. First, the body needs to be oriented towards the resetting position. An horizontal segment is drawn on the screen representing the user's orientation with respect to the resetting position, like his shoulder line seen from above in the real workspace reference frame. The user has to change his orientation until the segment becomes parallel to his body. Then, as a second step, the user has to walk to the resetting position by following an arrow direction. The arrow becomes smaller as the user gets closer to the resetting position, indicating how far he is from his target. Through this mechanism, the user can reach the center of the real workspace from anywhere in the real workspace. Figure 4a shows the segment and the arrow drawn at the top of the screen.

Extended 2:1-Turn Technique. In the original 2:1-Turn technique, a 180° real rotation of the user is mapped to a 360° virtual turn, and the user stays on the same real path but on the opposite direction. Since real turns are always of 180°, walking paths need to be straight with eventually right angle turns to avoid the same "locking" problems mentioned above.

In the extended 2:1-Turn technique, real turns can no longer be of only 180°. The resetting angle is given by the non oriented angle between the viewing direction and the body position - resetting position vector. The virtual angle remains the same, 360°. For any orientation before resetting, two turning directions are possible: to the left and to the right. To each direction corresponds an angle, with usually one greater than the other. The direction with the largest angle is chosen, so that the rotation gain when mapping the turn to a 360° virtual turn is lower, and the illusion is therefore less perceivable. As show in Figure 4b, an arrow drawn at the top of the screen indicates the turning direction to the user.

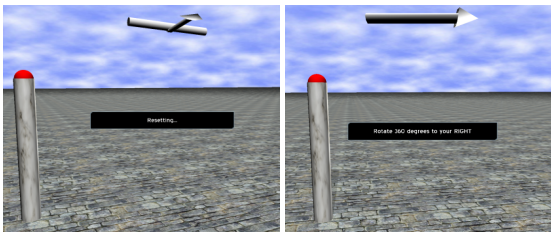


Figure 4: *The visual cues (top of the screen) from: (a) the extended Freeze-backup technique, giving the shoulder orientation and the walking direction, and (b) the extended 2:1-Turn technique, giving the rotation direction.*

5 Evaluation

In order to demonstrate its suitability for infinite navigation within a restricted workspace, we evaluated the Magic Barrier Tape by comparing it to two other existing navigation techniques that enable collision free infinite walking within a restricted workspace, namely the Freeze-Backup and the 2:1-Turn resetting techniques [Williams et al. 2007] with our extensions for omni-directional walking. We conducted two experiments, a pointing task and a path following task.

5.1 Experiment #1

In Experiment #1, our goal was to compare the 3 techniques over a pointing task where the user had to move from a central initial location to a new location, indicated by a target, as fast as possible. We a priori assumed that the Magic Barrier Tape will be faster, since rate control allows speeds greater than the average walking speed.

5.1.1 Description

Population. Twelve participants (1 female and 11 males) aged from 24 to 59 (mean = 30.3, sd = 5.7), took part in this experiment. Two of them were left-handed, and none of them had known perception disorders. They were all naïve to the purpose of the experiment.

Experimental Apparatus. The experiment was conducted in a closed room with dim light. We used the eMagin Z800 Head Mounted Display as display device, at 60 Hz and with stereoscopy enabled. The user was wearing an opaque fabric on top of the HMD to avoid seeing the surrounding real world. The user was carrying a backpack with the laptop computer running the application, and could therefore move freely (Figure 5). The user's head and hand were tracked by an ART ARTtrack2 infrared tracking system with

9 surrounding cameras for 360° tracking. The available tracking space was a cylinder with a 3m diameter and a 2.5m height.

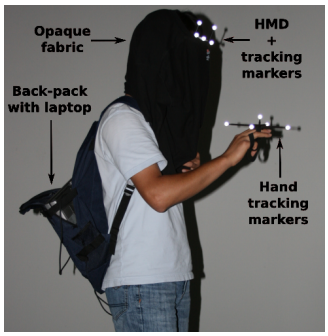


Figure 5: *A subject wearing the tracking equipment, the opaque fabric for occlusion and the backpack with the laptop computer.*

The scene consisted of a flat infinite floor with a rockwall texture, a cloudy blue sky, and a target made of a 1.4m high and 0.2m radius marble textured cylinder with a 0.2m radius red hemisphere on top, as illustrated in Figure 6.

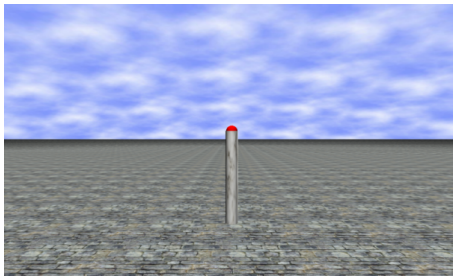


Figure 6: *Virtual environment used in Experiment #1 with a target.*

Procedure. Before the beginning of the experiment, the three techniques were explained to the subject. Before each trial, the subject had to go back to the initial position and orientation at the center of the workspace. Then, the experimenter launched the next trial. The participant was instructed to look for the target in the scene and to walk towards it until collision. For each technique, he had 3 trials for training. The experiment lasted about 30 minutes, including training trials.

Experimental Plan. Participants completed all the three technique conditions (Barrier Tape, Freeze-backup, and 2:1-Turn) and the order of the conditions was counterbalanced across participants. In each condition, the participants were exposed to 3 successive blocks of 6 trials (2 different distances $\times$ 3 different directions). The 3 possible directions are at 120° each, one of them being the user's initial direction, and the two possible distances are 2.2m and 5m. In each block, the presentation order of these trials was randomized. Participants completed a total of 54 trials (6 target positions $\times$ 3 technique conditions $\times$ 3 trials per condition). During a learning phase, prior to each technique condition, participants were exposed to 3 trials that did not enter in the final data set.

Collected Data. For each trial and each subject, we recorded the completion time (in seconds) and the amplitude of walking in the real world (in meters). The completion time is the time took by the subject to complete the trial. The amplitude of walking in the real world corresponds to the distance traveled when walking in the real *and* the virtual world. It is the movement in the real world that makes the user move forward in the virtual world (as opposed to a resetting movement where the user moves in the real world, but not in the virtual one).

5.1.2 Results

For the different comparison analyses, a correction for experiment-wise error was realized by using Bonferroni-adjusted alpha level ($p = 0.05$ divided by the number of tests). Thus, in order to compare the Barrier Tape technique to the two other techniques (Freeze-backup and 2:1-Turn) the alpha level was adjusted to $p = 0.025$.

Completion Time. Using the completion time data collected during the experiment, we conducted a statistical analysis. For each participant, statistics (mean M, standard deviation SD) were computed on the 18 trials in each condition. A one-way within subject design ANOVA (Techniques: Barrier Tape, Freeze-backup, 2:1-Turn) on the mean completion time (in seconds) revealed a significant main effect of the technique ($F(2, 22) = 183.22, p < 0.001$) (Figure 7). Follow up t tests revealed that completion time in the Barrier Tape technique ($M = 6.37$ sec, $SD = 1.30$ sec) was significantly shorter than in the Freeze-backup technique ($M = 21.49$ sec, $SD = 3.11$ sec, $t(11) = -19.15$, $p < 0.001$). Similarly, completion time in the Barrier Tape technique was significantly shorter than in the 2:1-Turn technique ($M = 14.54$ sec, $SD = 2.41$ sec, $t(11) = -14.61, p < 0.001$).

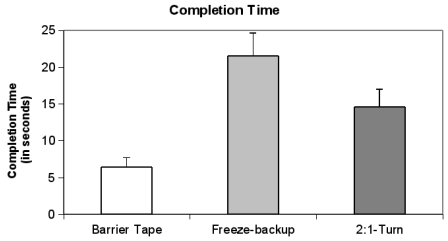


Figure 7: *Experiment #1: mean and standard deviation of the completion time (in seconds) for the three techniques (Barrier Tape, Freeze-backup, and 2:1-Turn).*

Amplitude of Walking in the Real World. An ANOVA on the mean amplitude of walking in the real world (in meters) revealed a significant main effect of the technique ($F(2, 22) = 434.75$, $p < 0.001$). Follow up t tests revealed that the amplitude of walking in the real world in the Barrier Tape technique ($M = 1.46$ m, $SD = 0.16$ m) was significantly shorter than in the Freeze-backup technique ($M = 4.42$ m, $SD = 0.30$ m, $t(11) = -30.13$, $p < 0.001$). Similarly, the amplitude of walking in the real world in the Barrier Tape technique was significantly shorter than in the 2:1-Turn technique ($M = 3.37$ m, $SD = 0.23$ m, $t(11) = -20.80$, $p < 0.001$).

5.2 Experiment #2

In the second experiment, our goal was to compare the 3 techniques over a path following task where the user had to follow a path delimited by two virtual walls, as fast as possible and as accurately as possible by trying to stay right between the two walls. We a priori assumed that the Magic Barrier Tape will be faster, as in Experiment #1, but less precise due to the controlability of rate control [Zhai 1995].

5.2.1 Description

Population. The population that participated in this experiment was the same as for Experiment #1.

Experimental Apparatus. We used the same experimental apparatus than in Experiment #1, except that we replaced the targets by two possible paths: a 2.5m radius circle or a 6m side length square (Figure 8). The walls at both sides of the path were in a semi-transparent blue material, 1m high, and at 1m from the path, creating a 2m wide corridor. Red arrows on the floor indicated the direction to follow, and a red 1m high and 0.2m radius cylinder indicated the start and finish position. The paths alternated throughout the experiment. Figure 9 shows the virtual scene with the circular path as seen from the user's point of view.

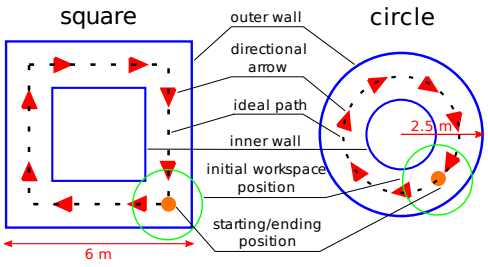


Figure 8: *Two paths used in Experiment #2, in top-view, with the initial cylindrical walking workspace position.*

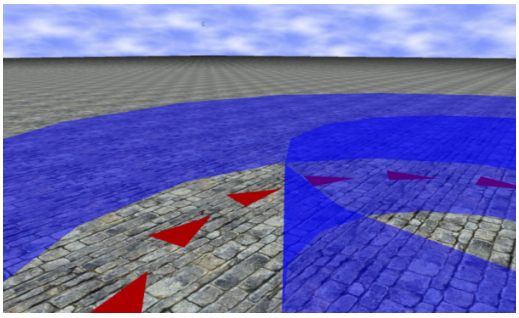


Figure 9: *The virtual environment used in the Experiment #2 with circular path, walls and directional arrow cues on the ground.*

Procedure. Before each trial, the subject had to go back to the initial position and orientation at the center of the workspace. Then, the experimenter launched the next trial. The participant was instructed to try to follow the path right between the two walls in the direction given by the arrows on the floor, until he reached the target cylinder. For each technique, he had 2 trials for training. The experiment lasted about 30 minutes, including training trials.

Experimental Plan. Participants completed all the three technique conditions (Barrier Tape, Freeze-backup, 2:1-Turn) and the order of the conditions was counterbalanced across participants. In each condition, the participants were exposed to 2 successive blocks of 2 trials with 2 different paths (square and circle). In each block, the presentation order of these trials was randomized. Participants completed a total of 12 trials (2 paths $\times$ 3 technique conditions $\times$ 2 trials per condition). During a learning phase, prior to each technique condition, participants were exposed to 2 trials that did not enter in the final data set.

Collected Data. Along with the same data as in Experiment #1, we also collected the path deviation. The path deviation (in m^2) is given by the area delimited by the subject's path in the virtual scene and the ideal path (exactly between the two walls).

5.2.2 Results

For the different comparison analysis, a correction for experiment-wise error was realized by using Bonferroni-adjusted alpha level ($p = 0.05$ divided by the number of tests). Thus, in order to compare the Barrier Tape technique to the two other techniques (Freeze-backup and 2:1-Turn) the alpha level was adjusted to $p = 0.025$.

Completion Time. An ANOVA on the mean completion time (in seconds) revealed a significant main effect of the technique ($F(2, 22) = 84.01$, $p < 0.001$) (Figure 10). Follow up t tests revealed that completion time in the Barrier Tape technique ($M = 31.62$ sec, $SD = 9.71$ sec) was significantly shorter than in the Freeze-backup technique ($M = 99.54$ sec, $SD = 21.63$ sec, $t(11) = -12.06$, $p < 0.001$). Similarly, completion time in the Barrier Tape technique was significantly shorter than in the 2:1-Turn technique ($M = 52.33$ sec, $SD = 6.59$ sec, $t(11) = -6.48$, $p < 0.001$).

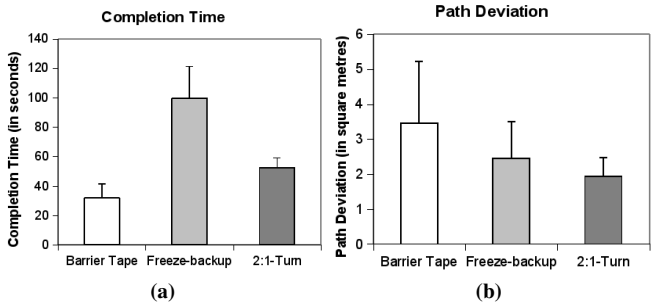


Figure 10: *Experiment #2: means and standard deviations of the completion time (a) and the path deviation (b) for the three techniques (Barrier Tape, Freeze-backup, and 2:1-Turn).*

Path Deviation. An ANOVA on the mean path deviation (in square meters) revealed a significant main effect of the technique $F(2, 22) = 4.77$, $p = 0.019$ (Figure 10). Follow up t tests revealed that the path deviation in the Barrier Tape technique ($M = 3.46$ m^2, $SD = 1.76$ m^2) was not significantly different from the path deviation in the Freeze-backup technique ($M = 2.45$ m^2, $SD = 1.04$ m^2, $t(11) = 1.72$, $p = 0.1143$.). By contrast, the analysis indicated that the path deviation in the 2:1-Turn technique ($M = 1.93$ m^2, $SD = 0.54$ m^2) was significantly lower than in the Barrier Tape technique, $t(11) = 2.81$, $p = 0.017$.

Amplitude of Walking in the Real World. An ANOVA on the mean amplitude of walking in the real world (in meters) revealed a significant main effect of the technique ($F(2, 22) = 379.81$, $p < 0.001$). Follow up t tests revealed that the amplitude of walking in the real world in the Barrier Tape technique ($M = 6.81$ m, $SD = 1.33$ m) was significantly shorter than in the Freeze-backup technique ($M = 19.03$ m, $SD = 1.25$ m, $t(11) = -32.63$, $p < 0.001$). Similarly, the amplitude of walking in the real world in the Barrier Tape technique was significantly shorter than in the 2:1-Turn technique ($M = 13.61$ m, $SD = 1.54$ m, $t(11) = -13.17$, $p < 0.001$).

5.3 Subjective Questionnaire

After both experiments, a preference questionnaire was proposed in which participants had to grade from 1 to 7 the 3 techniques according to 6 subjective criteria: *easiness of use*, *fatigue*, *navigation speed*, *navigation precision*, *general appreciation* and *naturalness*. Figure 11 shows the means and standard deviations of the 3 techniques for each of the subjective criteria.

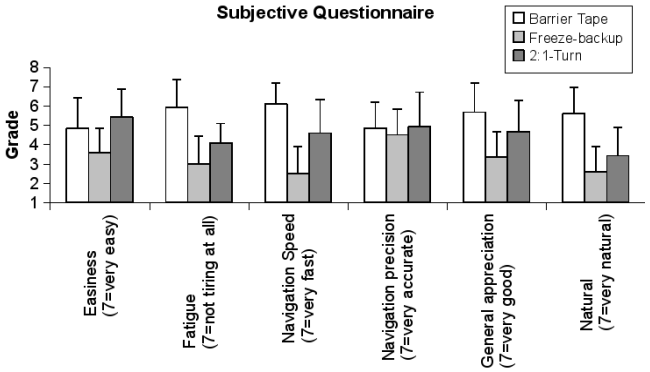


Figure 11: *Mean and standard deviation of subjective ratings about the different criteria for the three techniques.*

Wilcoxon signed rank tests with Bonferroni correction showed significant differences: for the *fatigue*, between the Barrier Tape and the Freeze-backup techniques ($z = 2.69$, $p = 0.007$) and between the Barrier Tape and the 2:1-Turn techniques ($z = 2.41$, $p = 0.016$); for the *naturalness*, between the Barrier Tape and the Freeze-backup techniques ($z = 2.77$, $p = 0.006$) and between the Barrier Tape and the 2:1-Turn techniques ($z = 2.53$, $p = 0.011$); for the *navigation speed*, only between the Barrier Tape and the Freeze-backup techniques ($z = 2.82$, $p = 0.005$); and for the *general appreciation*, only between the Barrier Tape and the Freeze-backup techniques ($z = 2.65$, $p = 0.008$).

6 General Discussion

Both experiments showed that the Magic Barrier Tape is faster compared to the other techniques. Indeed, results show that Experiment #1 trials were completed more than 3 times faster with the Magic Barrier Tape than with the Freeze-backup technique, and more than 2 times faster than with the 2:1-Turn technique. In Experiment #2, completion time using the Magic Barrier Tape was also roughly 3 and 2 times faster respectively. This result is consistent with the user's impression from the questionnaire regarding the navigation speed of the different techniques (Figure 11). It is mainly due to the fact that there is no time lost in the resetting of the position when using the Magic Barrier Tape, and that the control law allows navigation speeds greater than the average walking speed. Completion times could be further reduced by tuning the control law for greater speeds, although controlling the Magic Barrier Tape could become increasingly difficult, as testified by 3 users which complained about an acceleration behavior that was sometimes hard to control.

The experiments also showed that users walked less when using the Magic Barrier Tape than with the other 2 techniques, which was expected due to the use of rate control at the boundaries of the workspace. However, an interesting observation can be made when considering that trials were completed significantly faster with the Magic Barrier Tape. If we do a ratio between the amplitude of walking in the real world and completion time, in a per user basis, we obtain similar values for the Magic Barrier Tape ($M = 0.24$, $SD = 0.04$), the Freeze-backup ($M = 0.21$, $SD = 0.04$) and the 2:1-Turn ($M = 0.24$, $SD = 0.04$) techniques in Experiment #1, as well as in Experiment #2 with ($M = 0.22$, $SD = 0.036$), ($M = 0.20$, $SD = 0.037$) and ($M = 0.26$, $SD = 0.052$) respectively. Hence, the amount of "useful walking", contributing to moving forward in the virtual world, relative to time is as large with the Magic Barrier Tape as with the other techniques. If we consider that walking speeds are the same for the 3 techniques, users spend roughly the same percentage of the total time doing useful walking with the Magic Barrier Tape technique than with the other 2 techniques.

Experiment #2 showed that the Magic Barrier Tape was less precise when following a given path, with a higher path deviation when compared to the 2:1-Turn technique (roughly 2 times less precise). We cannot conclude on the comparison with the Freeze-backup technique, since results were not significantly different. Again, these results were expected. By nature and design, the use of the Magic Barrier Tape is meant for coarse positioning. The user gets close enough to the navigation target in order to have it inside his workspace, and can then reach it by fine positioning navigation through real walking. As explained by one of the subjects of the Experiment #2, when asked about the strategies he used: "I sent the barrier tape as far as possible without going into the walls in order to take advantage of the workspace". However, path deviation could be improved by allowing users to cutomize their control law, like when they choose the mouse speed in desktop computers. Moreover, Zhai [Zhai 1995] observed that, with sufficient training, rate control and position control can achieve similar performances. Hence, further user training on the Magic Barrier Tape rate control might improve its mean path deviation.

Overall, users graded the Magic Barrier Tape higher in all criteria of the questionnaire where comparisons were significantly different. We highlight that 6 subjects complained about having cybersickness when using the 2:1-Turn technique, while 2 said it made them loose balance. Many subjects found the Freeze-backup technique exhausting and frustrating. It is also important to note that 2 subjects had a very hard time using the Magic Barrier Tape. They used an inadequate strategy, and complained about the control law. They might have needed a longer training time, or more guidance on the strategy to adopt. They consistently graded it lower than the other techniques in every criteria of the questionnaire.

In a nutshell, the Magic Barrier Tape is faster than the Freeze-backup and the 2:1-Turn techniques, but is less precise when using it in rate control. The 2:1-Turn technique is the most precise, but seems to induce cybersickness to users, as well as stability issues. There is a general dissatisfaction with the Freeze-backup technique, mainly due to its physical exertion and slow speed, leading to a frustrating experience. People generally prefer the Magic Barrier Tape, and find it more natural and less tiring.

7 Perspectives

Through user feedback on the experiments and our own observations, we found some ways of potentially improving the Magic Barrier Tape.

The user could use any part of his body in order to "push" on the virtual barrier tape. Figure 1 (right) shows the virtual barrier tape being triggered by using the elbow when the user's hands are busy carrying an object. One could think about using the shoulders, the pelvis or the feet, since we often naturally use these body parts when we are unable to use our hands.

In their "Bubble" technique [Dominjon et al. 2005], a hybrid position/rate control haptic interaction technique for devices with re-

stricted workspace, Dominjon et al. successfully used haptic feedback to represent the workspace boundaries and their virtual elasticity. In the RubberEdge technique [Casiez et al. 2007], Casiez et al. used a passive haptic feedback through an elastic ring on top of a tracking surface such as a touchpad to allow the user to switch from position to rate control when reaching the elastic boundaries. Similarly, the Magic Barrier Tape could be augmented with haptic feedback when "pushing" on the tape. A possibility could be the use of passive haptics through tangible objects such as queue barriers with retractable belts as one could find in airports and queue-up places. The queue barriers could follow the workspace boundaries, and the virtual barrier tape would match the queue barriers position. Since retractable belts are elastic, the haptic feedback of the virtual barrier tape elastic deformation would be straightforward. Many users complained about the translation speed when using the Magic Barrier Tape, since the acceleration could be hard to control. A solution to this problem could be the use of a discrete approach. The control law, according to the hand penetration, would deliver one of three discrete velocities, corresponding to a human walking, jogging or running. Side-stepping human velocities could be used when moving in a direction orthogonal to the body orientation. The translation speed would therefore be more predictable, although capped by the running speed.

Last, although the barrier tape is made semi-transparent to reduce occlusion, visibility might be reduced in environments were the dominant color is close to the tape color. A way to enhance visibility in such cases would be to use different tape textures using the complementaries of the dominant colors of the surrounding environment, in order to emphasize the presence of the Magic Barrier Tape while increasing the visibility of the scene.

8 Conclusion

This paper introduces the Magic Barrier Tape, a new interaction metaphor for navigating in a potentially infinite virtual scene while confined to a restricted walking workspace. Using the barrier tape metaphor and its "do not cross" implicit message, the walking workspace is surrounded with virtual barrier tape in the virtual scene. The technique uses a hybrid position/rate control mechanism: real walking is used inside the workspace, while rate control navigation is used to move beyond the boundaries by "pushing" on the virtual barrier tape. Moreover, the technique naturally informs the user about the boundaries of his walking workspace, providing a walking environment safe from collisions and tracking problems.

We conducted two experiments in order to evaluate the Magic Barrier Tape by comparing it to other state-of-the-art navigation techniques previously extended for omni-directional navigation. In Experiment #1 participants had to walk to a target, while in Experiment #2 they had to navigate inside a scene following a path. Results showed that the Magic Barrier Tape was faster than the other techniques. Experiment #2 results confirmed that, by design, navigation through rate control with the Magic Barrier Tape is not meant for precise path following, but rather for coarse positioning between fine positioning tasks. Overall, the Magic Barrier Tape was more appreciated, while being more natural and less tiring.

Future work will focus on exploring the different perspectives highlighted in this paper, namely the use of haptics for a more compelling and immersive experience, and the use of discrete velocities to produce a more predictable and realistic motion in rate control.

Acknowledgements

The authors would like to thank A.R.T. GmbH and Haption for their help with the hardware setup. This work was supported by the European Community under FP7 FET-Open grant agreement n°222107 NIW - Natural Interactive Walking.

References

BOWMAN, D. A., KRUIJFF, E., LAVIOLA, J. J., AND POUPYREV, I. 2004. *3D User Interfaces: Theory and Practice*, 1 ed. Addison-Wesley Professional, Aug.

CASIEZ, G., VOGEL, D., PAN, Q., AND CHAILLOU, C. 2007. RubberEdge: reducing clutching by combining position and rate control with elastic feedback. In *Proceedings of the ACM symposium on User interface software and technology*, ACM, Newport, Rhode Island, USA, 129–138.

DOMINJON, L., LECUYER, A., BURKHARDT, J., ANDRADE-BARROSO, G., AND RICHIR, S. 2005. The "Bubble" technique: interacting with large virtual environments using haptic devices with limited workspace. In *Proceedings of World Haptics conference*, 639–640.

HOLLERBACH, J. 2002. Locomotion interfaces. In *Handbook of Virtual Environments*, K. M. Stanney, Ed. 1232.

INTERRANTE, V., RIES, B., AND ANDERSON, L. 2007. Seven league boots: A new metaphor for augmented locomotion through moderately large scale immersive virtual environments. In *Proceedings of the IEEE Symposium on 3D User Interfaces*, 167—170.

IWATA, H., YANO, H., AND TOMIYOSHI, M. 2007. String walker. In *ACM SIGGRAPH 2007 emerging technologies*, ACM, San Diego, California, 20.

LAVIOLA, J. J., FELIZ, D. A., KEEFE, D. F., AND ZELEZNIK, R. C. 2001. Hands-free multi-scale navigation in virtual environments. In *Proceedings of the ACM symposium on Interactive 3D graphics*, ACM, 9–15.

NITZSCHE, N., HANEBECK, U. D., AND SCHMIDT, G. 2004. Motion compression for telepresent walking in large target environments. *Presence: Teleoper. Virtual Environ. 13*, 1, 44–60.

PAUSCH, R., BURNETTE, T., BROCKWAY, D., AND WEIBLEN, M. E. 1995. Navigation and locomotion in virtual worlds via flight into hand-held miniatures. In *Proceedings of ACM SIGGRAPH*, ACM, 399–400.

RAZZAQUE, S., KOHN, Z., AND WHITTON, M. C. 2001. Redirected walking. In *Proceedings of Eurographics*.

RUDDLE, R. A., AND LESSELS, S. 2009. The benefits of using a walking interface to navigate virtual environments. *ACM Trans. Comput.-Hum. Interact. 16*, 1, 1–18.

SLATER, M., USOH, M., AND STEED, A. 1995. Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Trans. Comput.-Hum. Interact. 2*, 3, 201–219.

WARE, C., AND OSBORNE, S. 1990. Exploration and virtual camera control in virtual three dimensional environments. In *Proceedings of ACM SIGGRAPH*, 175–183.

WILLIAMS, B., NARASIMHAM, G., RUMP, B., MCNAMARA, T. P., CARR, T. H., RIESER, J., AND BODENHEIMER, B. 2007. Exploring large virtual environments with an HMD when physical space is limited. In *Proceedings of the ACM symposium on Applied perception in graphics and visualization*, ACM, Tubingen, Germany, 41–48.

ZHAI, S. 1995. *Human performance in six degree of freedom input control*. PhD thesis, University of Toronto.

ACOUSTICS2008/57
Non-visual identification of walking grounds

B. L. Giordano[a], S. Mcadams[a], Y. Visell[b], J. Cooperstock[b], H.-Y. Yao[c] and V. Hayward[c]
[a]Centre for Interdisciplinary Research in Music Media & Technology (CIRMMT) - Schulich School of Music - McGill Univ., 555 Sherbrooke Street West, QC H3A1E3 Montreal, Canada
[b]Centre for Intelligent Machines - CIRMMT - Dept. of Electrical and Computer Engineering - McGill Univ., 3480 University Street, QC H3A 2A7 Montreal, Canada
[c]Centre for Intelligent Machines - McGill Univ., 3480 University Street, QC H3A 2A7 Montreal, Canada
bruno.giordano@music.mcgill.ca

We investigated the role of haptic, proprioceptive and auditory information in the non-visual identification of walking grounds. We selected four solid materials (e.g., marble) and four aggregate materials (e.g., fine gravel). Five observers identified the materials in each of four experimental conditions: multisensory, haptic, proprioceptive, and auditory. In the auditory condition, they were presented with walking sounds they produced. In the other conditions, observers walked blindfolded on the materials. In the haptic and proprioceptive conditions auditory information was masked. In the proprioceptive condition haptic information was masked. No masking took place in the multisensory condition. In all conditions, solids and aggregates were seldom confused, and aggregates were better identified than solids. Chance identification performance was observed only for solids in the presence of simultaneous haptic and auditory masking, suggesting a secondary role of proprioceptive information. In the proprioceptive condition, identification of aggregates was better than chance, suggesting a significant role of proprioception. Identification of aggregates was at its best in the haptic condition, and not in the multisensory condition, suggesting either an inconsistency of information between auditory and non-auditory channels, or conflicting cognitive strategies for the exploitation of these sources of information in multisensory contexts.

Number of words in abstract: 195
Keywords: Walking, Multisensory integration, Sound source perception, Haptics, Proprioception
Technical area: Psychological and Physiological Acoustics (PP)
Session: PP10 - Auditory perception of sound source properties
PACS #1: 43.66.Lj Perceptual effects of sound (see also 43.71.-k)
PACS #2: 43.66.Jh Timbre, timbre in musical acoustics
PACS #3: 43.66.Ba Models and theories of auditory processes
Presentation: Oral presentation preferred (Invited paper)
Special facility: Audio equipment
Best student paper competition: no
Send notice to: Giordano Bruno (bruno.giordano@music.mcgill.ca)

HAVE 2008 - IEEE International Workshop on
Haptic Audio Visual Environments and their Applications
Ottawa, Canada, 18-19 October 2008

A Multi-modal Floor-space for Experiencing Material Deformation Underfoot in Virtual Reality

Alvin W. Law[1], Benjamin V. Peck[2], Yon Visell[1], Paul G. Kry[2], Jeremy R. Cooperstock[1]

[1]Centre for Interdisciplinary Research in Music Media and Technology,
Centre for Intelligent Machines, Dept. of Electrical and Computer Engineering,
E–mail: alvinlaw@cim.mcgill.ca, yon@cim.mcgill.ca, jer@cim.mcgill.ca.

[2]School of Computer Science,
E–mail: benjamin.peck@mail.mcgill.ca, kry@cs.mcgill.ca.

3480 University Street, McGill University,
Montreal, Canada

Abstract – *We present a floor-space design that provides the impression of walking on various terrains by rendering graphical, audio and haptic stimuli synchronously with low latency. Currently, interactive floors tend to focus on visual and auditory feedback but have neglected to explore the role of haptics. Our design recreates these three modalities in a direct manner where the sensors and reactive cues are located in the same area. The goal of this project is the creation of a realistic and dynamic area of ground that allows multiple, untethered users to engage in intuitive interaction via locomotion in a virtual or augmented reality environment.*

Keywords – *interactive floor, deformation models, virtual materials*

I. INTRODUCTION

Virtual reality (VR) stimulates our senses to convince us that the surrounding world is a real, yet mutable, environment. One area that has much room for improvement is that of realistic, interactive locomotion in VR spaces. Presented with a new environment, curiousity naturally encourages people to walk around and explore. Locomotion is a critical part of familiarizing oneself with a new place, not just because it allows us to see around corners, but also because information is gathered through the act of walking itself. The sound of crunching leaves underfoot, or the sensation of gravel shifting as it is walked upon, can inform us about where we are, and aid our decisions during navigation. Our feet are the only parts of our bodies in constant contact with the surrounding environment. This project aims to take advantage of the lifetime of experience we have acquired in this way.

The objective of our work is to generate perceptually realistic simulations of ground areas, by modeling their physically interactive properties. If this effort is successful, we might provide people with the sensation of walking on remote terrains, such as arctic snow or desert sand, as appropriate to specific applications, and to change these sensations dynamically. Our floor design (see Figure 1) builds on prior work in the field, and is especially adapted for multi-modal virtual reality. It is intended to look, sound, and feel realistic, drawing on all three modalities simultaneously, and with sufficiently low latency, to provide an immersive experience.

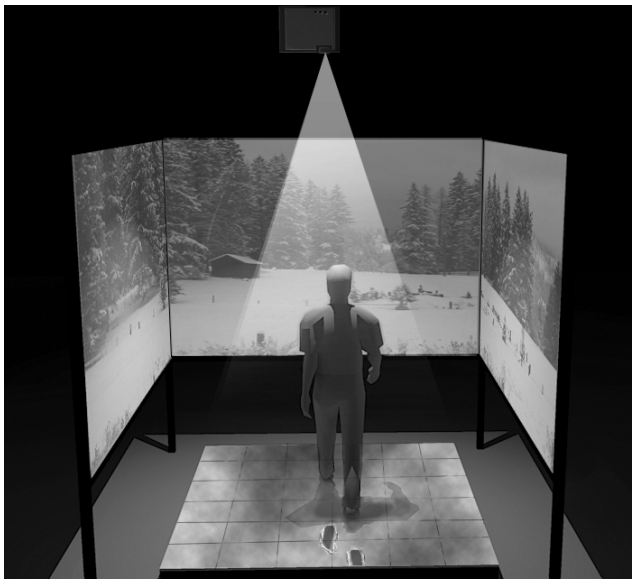


Fig. 1. The floor being constructed for our lab's existing CAVE-like environment adds a fourth dynamic, multi-modal surface.

II. RELATED WORK

As our contribution focuses on integrating and improving both a floor-space and a real-time material-deformation model, relevant work in these respective domains is presented.

A. Interactive Floors

A number of research groups have investigated the development of floors with integrated sensing capabilities, and in some cases, with integrated auditory or visual display capabilities. Many of these use force-sensing resistor (FSR) arrays, arranged in a matrix, to acquire foot-floor contact forces, as in

the system we present here. One example is the touch sensitive dance floor by Pinkston [1], which also provides auditory feedback. Other approaches to sensing that have been used in musically interactive floors include the piezoelectric wire network developed by Paradiso et al. [2]. Closer in spirit to the present contribution, Cook's PholieMat [3] uses a sensor mat as a controller for the real-time synthesis of footstep sounds generated by walking on different virtual materials.

Several current systems make use of visual feedback in response to users' footsteps. Commercial products[1] exist that track users with a stereo camera and display advertisements via top-down projection. Using projection and video capture from beneath a glass floor surface, the Wisdom Well [4] creates an educational play area that overcomes the common problem of shadows. The ADA project [5] involves large sensing tiles with integrated color illumination that respond to the patterns of people walking through the space in which it is installed. Although limited in resolution compared to systems utilizing video projection, the project resulted in an interactive floor much larger in scale than previous work.

A comparison of further examples is given in the book by Miranda and Wanderley [6].

B. Deformable Terrain Models

Previous work on simulating deformable terrain can generally be categorized into one of two methods: height-field-based or particle-based approaches. The former represent terrain as a matrix of vertical displacement values, while particle systems simulate the collection of granular elements that make up the terrain. Particle-based models have achieved highly realistic results, but remain impractical for real-time rendering of any significant terrain quantity [7][8]. By simplifying the problem to two dimensions, height fields have been successful in achieving real-time frame rates [9][10]. There are, however, corresponding limitations in simulating certain phenomena accurately, such as the interactions with concave objects. The remainder of this section will discuss literature relevant to our project objectives.

Sumner et al. [9] represent a deformable terrain with a regular grid of cells containing height values. Their simulation consists of three steps: collision detection by ray-tracing, displacement, and erosion. All material in the collision is displaced into the surrounding cells of the height field. The erosion step further spreads material to cells outside of the collision area wherever the slope is sufficiently steep. The type of material being simulated is determined by parameters such as compression coefficient and the terminal angle at which material will slide into a neighbouring cell. For example, sand does not compress but slides at shallow slopes, while snow compresses and only slides at steep slopes. This method yields effective results, but is intended for off-line animation. Our approach is influenced by this work but focuses on interactive real-time deformation.

Onoue et al. [11] introduced the height span map (HS-map), which overcomes the limitations of the standard height field by adding multiple pairs of heights, or *spans*, of material per cell. Ground deformation is calculated by algorithms based on those of Sumner et al. [9]. The resulting model is convincing, but the HS-map must be recalculated whenever the orientation of a rigid object changes; this is computationally expensive and leads to suboptimal display rates (7-14 Hz on a 3 GHz Pentium 4).

Bell et al. [7] use massive collections of particles to simulate discrete elements of ground material. This model does not take into account any cohesive forces, making it effective only for granular materials such as sand. Zhu et al. [8] modify a particle-based water model and incorporate viscosity by tracking the cohesive forces. A unified surface is calculated from the collection of particles, yielding an impressive simulation of completely fluid or semi-fluid bodies. However, neither of these models are currently implemented in real time.

Zeng et al. [10] optimize collision detection with a height field model by rendering the depth information of the rigid object to a texture for comparison against the values in the height field. The velocity with which a rigid object enters the ground is taken into account when applying a deformation. The momentum of material in each cell is tracked, and the slide velocity is calculated using Mohr-Coulomb equations. This creates a much more physically based simulation than previous attempts, as the ground material continues to move dynamically from the momentum transferred by colliding rigid objects. Our method, in contrast, computes a static solution for the terrain at each time step using a simpler approach that runs on the GPU, has higher frame rates, and can handle larger terrain resolutions.

C. Haptic Interaction with Granular Models

Benes et al. [12] present a visual and haptic model of a real-time deformable granular material, based on a height-field model of sand. Interaction with the material is afforded by two different manually operated haptic desktop devices (Force Dimension model Omega and Phantom). In this configuration, the physical device controls a virtual proxy object that mediates the deformation. By comparison, the system presented here is a direct interaction paradigm.

III. SYSTEM DESIGN

The example of snow offers a compelling description of our objectives. When people step in snow, they perceive the *crunch* beneath their feet, they hear the distinctive sound of its compression, and they see the deformation of material around their feet. Our floor is intended to reproduce these three sensations simultaneously and instantaneously.

Our work extends a prototype [13] that successfully displays haptic and audio feedback, using FSRs as inputs. We employ a Vicon motion capture system to track the position and orientation of the walker's feet. This is necessary to provide real-time deformable graphics of the terrain model with sufficient accuracy for convincing graphical feedback. Since

[1] www.displax.com, www.vertigo-systems.com

the imprint is highly dependent on the amount of force being applied, optical tracking alone is insufficient, and FSRs must also be used. Additional benefits arise from the sharing of data between optical and FSR sensing paradigms, as described in Section D.

The current design attempts to address limitations of previous architectures. In particular, we localize all feedback modalities to the floor itself, which we believe is of importance to provide a compelling sense of engagement, all the more so in the case of multiple simultaneous users. The element of haptic feedback is, of course, a major distinguishing aspect of our work. This is motivated in part by past research, which has shown that in the absence of visual or auditory cues, characteristic vibrations of ground materials are sufficient for material identification [14].

A. Haptic-Auditory Sensing and Actuation

To determine position and orientation of the foot, the motion capture system tracks markers attached to the shoe. But actual footsteps are determined far more accurately by the FSRs, which can easily distinguish between a foot hovering just above the floor from one that is in contact. The footground interactions *during* contact are also captured with the floor sensors. This pressure information during a footstep has an important effect on all three feedback modalities.

The floor design consists of thirty six square tiles, each measuring 30.5 cm per side. These are arranged in a 6×6 matrix in the center of our CAVE-like environment.

Our initial prototype used a D-Box motion controller for haptic actuation. Although this allowed for DC motion displacement, latency considerations required us to switch to vibrotactile actuators that vibrate, but do not physically displace the tiles. Discarding the motion controller reduced the sensing-to-actuation latency from 100 ms to less than 20 ms. The reader may wonder whether the familiar experience of walking on complex, compliant ground materials, such as snow, can be simulated by such a device. Fortunately, our past research has shown that in the absence of visual or auditory cues, characteristic vibrations of ground materials are sufficient for material identification [14], even without the benefit of proprioceptive information, for example, the sense of one's feet sinking into snow. As applied weight causes a granular material to compress, a weak reaction force is present compared to the applied force. When the limit of compression is reached, the reaction force equals the applied force, with certain characteristic contact vibrations. Although a vibrating actuator cannot generate the displacement effects during compression, it can reproduce the contact vibrations, which convey a significant amount of perceptual information related to the ground material.The results of that study also suggest that vibrotactile and auditory signals are more important as carriers of information about material identity than is proprioceptive information.

Each tile is attached to a Clark Synthesis Vibrotactile Transducer (Silver Model) with 100 W of vibrating power. These act as the voice coil of a speaker, but fixed to the diaphragm. Low frequencies are perceived as haptic stimuli while higher frequencies are perceived as audio. Although there is a small band of overlap between the two, in general, the modalities can be treated as independent. The tiles are suspended on a square ring of 0.635 cm polyurethane foam to isolate vibrations from the rest of the system, as illustrated in Figure 2. Rubber feet can be installed to isolate the vibrations from the floor. Given the limitations of resolution available from a 6×6 tile array, there will likely be situations in which the walker has both feet on the same tile. In this case, both feet will sense the same vibrations, although these will likely be experienced stronger by the active stepping foot due to the mechanics of pressure distribution between the feet during a natural walking motion.

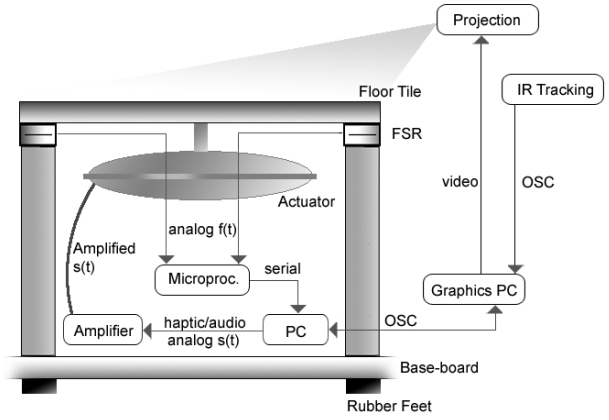


Fig. 2. Side view of a single tile. The components underneath the actuator actually span multiple tiles. Also shown are communication lines between different components. FSR voltage values are denoted by $f(t)$ and the synthesized audio-haptic signal by $s(t)$.

An FSR is recessed in the foam at each corner of the tile, as shown in Figure 3. Taking the relative force readings from the four FSRs allows the location of the center of pressure to be approximated. The sensors can be multiplexed into the analog inputs of an Arduino Diecimila board, supporting up to four tiles of four sensors each, with refresh rates above 250 Hz per sensor. The Arduino is connected by USB to a Mac Mini, which runs the simulation software responsible for haptic and acoustic feedback. USB-based multichannel sound cards are used to transmit dedicated signals to each actuator. Each set of four tile actuators is powered by a four-channel 400 W Class-D audio amplifier. Size concerns influenced the choice of components: we wanted these to fit beneath the actuators, making the height of the floor less than 20 cm, and mostly encapsulated.

B. Haptic-Auditory Synthesis

Haptic and auditory synthesis are accomplished in parallel, as tactile and audio signals are both generated by means of a real-time, lumped signal processing model of interactions with

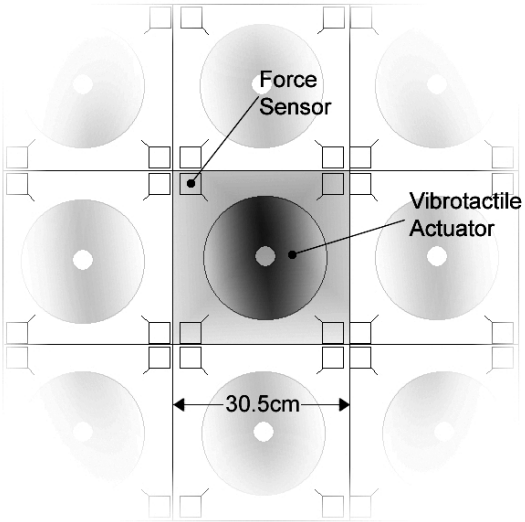


Fig. 3. Top view of the tile array. Each tile contains one actuator in the center and four force sensors located in the corners.

a granular medium, together with a control algorithm matched to the force-sensing floor interface. The model has been described in a prior publication [13], and is based on existing work in footstep sound synthesis by Fontana and Bresin [15]. A measurement scheme has also previously been developed to capture the reaction forces of real world granular ground materials via a custom instrumented shoe. We have used these recordings to identify parameter values of the synthesis model to best match the interactive behavior of these materials.

C. Graphics

Our graphical simulation of the deformable terrain focuses on the interaction between the foot and the ground material. The simulation is displayed on the floor by a ceiling-mounted Hitachi CPX5 projector.

We model the ground with a height field and take advantage of optimizations of Zeng et al. [10] to accelerate collision detection. Our model further optimizes rendering of the height field by using video RAM to store all terrain data and using the GPU to compute lighting normals dynamically. Additionally, we perform all deformations of the height field using only the GPU. Taking advantage of the GPU's stream processing architecture leads to faster simulations and larger terrains than those possible with previous methods. Our deformation algorithm is mapped to the GPU by combining OpenGL shading language programs and render to texture techniques, and is described after our discussion of the model and rendering method, below.

C.1 Model Description and Rendering

The virtual environment consists of a height field and a set of rigid objects, e.g., a model of a foot or shoe, each represented by a 3D triangle mesh. The height field consists of a planar grid of vertices and a floating point texture containing the vertical displacement values for each of the vertices (each pixel of the texture corresponds to one vertex in the plane). We use a texture of dimension 512×512 to simulate a 3.3 m^2 area, which translates to a density of approximately 8 pixels/cm^2. Both the vertex grid and the height map texture are stored in video RAM. This accelerates the rendering process and allows for dynamic computation of lighting normals on the GPU.

The height field uses two height maps. This is required because shader operations are not performed in-place but must be written to a separate texture from that used as input. The first map, or *active* map is used for on-screen rendering and as input for the deformation operation. Terrain deformation is produced in a different sequence of rendering passes (described in Section C.2) which write to the second *inactive* height map. After each simulation iteration, these height maps switch roles (see Figure 4), a process known as "ping-ponging." To render the height field, the ground plane vertices are sent through the graphics pipeline as triangle strips. A vertex shader displaces each vertex in the plane by the amount described in the active height map [16].

Every time the height field is rendered, the vertex shader computes new normals at each vertex by sampling the surrounding height map values. Dynamic computation of normals saves storage space and simplifies deformation operations. Note that the planar field of vertices that makes up the height field mesh remains completely static for the duration of the simulation; only the height maps need to be modified in order to deform the ground.

C.2 Terrain Deformation

The deformation algorithm is triggered once foot pressure is detected on the floor. The first step involves calculating the intersection volume between the foot and the ground. This is done by the collision shader. Following the approach of Zeng et al. [10], the depth buffer information of the foot is rendered to a texture. The depth texture of the foot corresponds to a small subsection of the height map. We only process data in this region, and as such, the computation time for the deformation algorithm is dependent only on the size of the collision area rather than the terrain as a whole. The collision shader compares the depth texture to the active height map. This is done by rendering a single quadrilateral where each pixel corresponds to a sample in the active height map. The output of this shader is rendered to a texture that we refer to as the collision map.

The next step is to calculate the deformation of the ground and produce a new height field. This is done by the deformation shader. Input for this shader consists of the collision map, the active height map, the velocity of the foot, and the compression rate of the material being simulated. This last value defines the percentage of colliding material that is removed from the simulation. The deformation shader initially decreases the height map value by the amount described in the collision map, eliminating the collision with the foot.

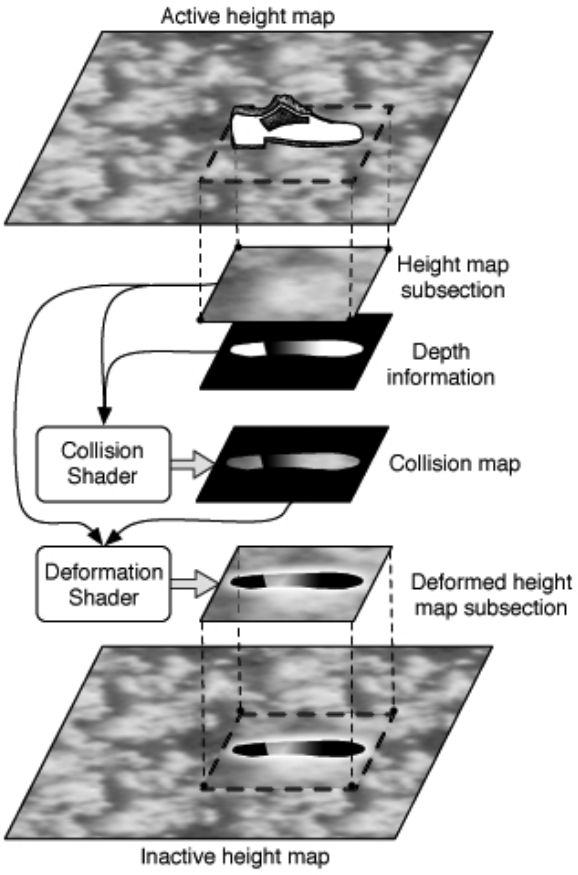


Fig. 4. One simulation step of our graphic model. At the end of each step, the active and inactive height maps are swapped.

To simulate the movement of material out from under the object, a 1D Gaussian blur is applied by the GPU to the collision map data in the direction of movement [17]. The blur width increases with the magnitude of tangential velocity. In addition, a radially symmetric Gaussian blur, independent of the tangential velocity, is also applied, and added to the value sampled from the height map. Depending on the tangential velocity of the object, this blurring either creates a mounding of material next to the foot in the direction of motion (see Figure 5A), or a uniform mounding around the area of collision (see Figure 5B). The resulting deformation map is rendered to the inactive height map in the area defined by the collision, and is displayed when the height maps are swapped. This process is visually effective and provides a reasonable compromise between physical accuracy and computational requirements.

D. Information Integration

Separate software is used to render the graphics and the vibrotactile signals. A Java application reads motion capture data to produce the height field simulation while Max/MSP reads FSR data to generates the acoustic and vibration outputs. These two programs exchange data via OpenSoundControl messages.

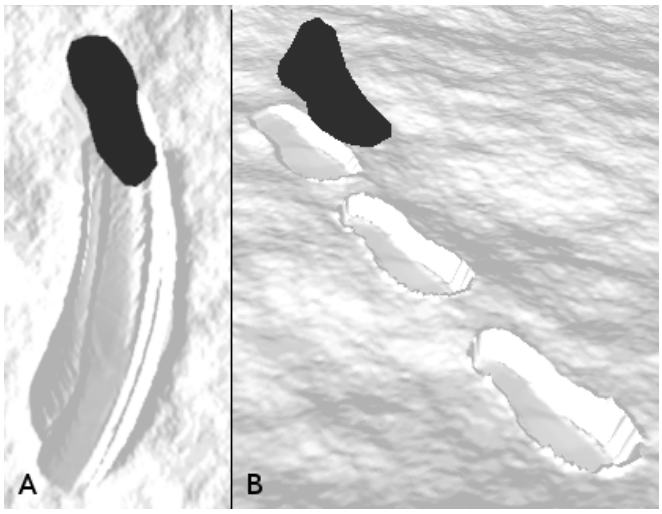


Fig. 5. Screenshots of: A) Deformation due to tangential motion through the terrain. B) Uniform mounding footprints left by the rigid model of a shoe.

While the FSRs provide initial sensor values that determine compression of the ground material, the synthesis model used for vibrotactile and acoustic actuation does not maintain long-term state. It may thus be beneficial to provide information to this synthesis algorithm from the graphical model, as the latter includes a detailed representation of the areas of ground that have already been compressed by footsteps. This would permit modification of auditory and haptic feedback to reflect the level of prior compression to the ground surface. For example, a footstep on compressed snow disperses its energy over a shorter vertical range and time period than on uncompressed snow. In return, the tile system can inform the graphical simulation of applied pressure by the foot to enrich the information provided by the motion capture system, as noted in Section A. This is critically important to achieve realistic effects, as, for example, a light tap of the foot should not produce the same impression as a hard stomp.

IV. ONGOING AND FUTURE WORK

A 2×2 tile prototype of the floor as seen in Figure 6 is presently being adapted and expanded to a 6×6 configuration for deployment in our CAVE-like environment. The effectiveness of the haptic and auditory simulation of snow was already validated through informal studies of the prototype without graphics. Almost all people when asked about what they were walking on comfirmed that it indeed felt like snow.

With regard to performance, we are able to render, on an nVidia GeForce 8800 GTS, a 3.3 m^2 area using a 512×512 height map at 260 Hz. This figure drops to 180 Hz when the

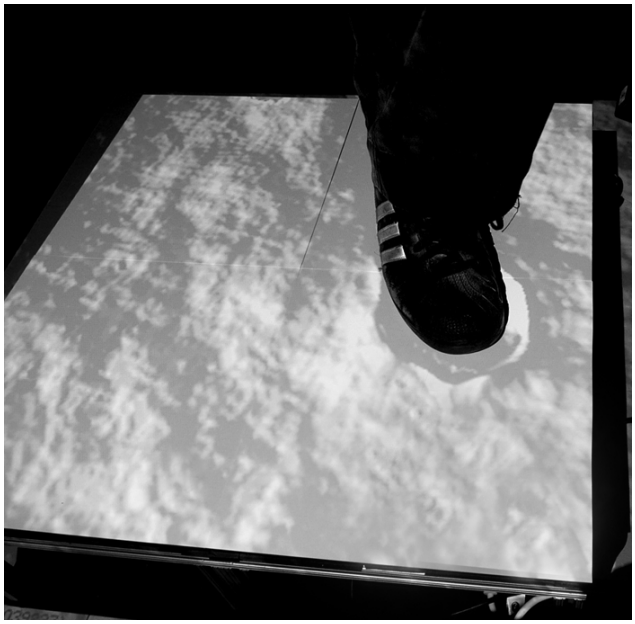


Fig. 6. A footstep causes snow to deform around the shoe (displayed on our prototype).

model is deformed by one foot, and 168-175 Hz with two feet. At 1024×1024 resolution, the height map runs at 90 Hz while rendering and 75 Hz under active deformation.

While the use of simple image filters to recreate deformation-related phenomena is very efficient, some tuning is required to achieve perceptually convincing results. Importantly, additional parameters need to be introduced to the deformation algorithm to support ground materials other than snow. This may also benefit from implementing a more physically based model [10].

We currently project an orthographic image of the terrain onto the floor. Shadows, which may occlude the graphics of interest, are inevitable with overhead projection, and back-projection is infeasible with the sensors and actuators in place. We are therefore considering the use of multiple overlapping projectors, possibly incorporating an active shadow removal process [18] to cope with this problem. The effects of shadows and relative contributions of each modality to users' "sense of immersion" can be assessed using methods such as presence questionnaires, as described by Slater et al. [19].

V. CONCLUSIONS

The system design presented in this paper introduces a novel approach to supporting natural walking activity on virtual terrains. Our prototype applies real-time synthesis to data obtained from both motion capture and force-sensing resistors to provide complementary audio, haptic, and visual stimuli that is highly engaging. While the integration of acoustic and haptic actuation, real-time granular deformation models, and detailed sensing of footsteps, is only now being completed, we are confident that each of these components is sufficiently developed and robust. The resulting fully controllable platform can be used for exploring various questions related to the interpretation of sensory information from the activity of locomotion.

A. Acknowledgments

Funding for this project has been provided by grants from the Natural Sciences and Engineering Research Council (NSERC), Les Fonds Quebecois de recherce sur la nature et les technologies (FQRNT), and the Canada Foundation for Innovation. The support of these agencies is gratefully acknowledged.

VI. REFERENCES

[1] R. Pinkston, "A Touch Sensitive Dance Floor/MIDI Conroller," *Proc. of the ICMC,* 1995.

[2] J. Paradiso, C. Abler, K. Hsiao, M. Reynolds, "The Magic Carpet: Physical Sensing for Immersive Environments," *Proc. of the ACM SIGCHI Conference* March 1997.

[3] P. Cook, "Modelling Bills Gait: Analysis and Parametric Synthesis of Walking Sounds," *Proc. of the Conf. on Virtual, Synthetic and Entertainment Audio,* July 2002.

[4] K. Gronbaek, O. Iversen, K. J. Kortbek, L. Aagaard, "iGameFloor - a Platform for Co-Located Collaborative Games," Proc. of the Intl. Conf. on Advances in Computer Entertainment June 2007.

[5] T. Delbruck, A. M. Whatley, R. Douglas, K. Eng, K. Hepp and P. F. M. J. V. Verschure, "A Tactile Luminous Floor for an Interactive Autonomous Space" *Robotics and Autonomous Systems 55: 433-443.* 2007.

[6] M. M. Wanderley, E. R. Miranda, "New Digital Musical Instruments: Control and Interaction Beyond the Keyboard," *A-R Editions, Middleton, WI,* 2006.

[7] N. Bell, Y. Yu, P. Mucha, "Particle-based Simulation of Granular Materials," *Proc. of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* 2005.

[8] Y. Zhu, R. Bridson, "Animating Sand as a Fluid," *ACM SIGGRAPH 2005 Papers* 2005.

[9] R. Sumner, J. O'Brien, J. Hodgins, "Animating Sand, Mud, and Snow," *The Proc. of Graphics Interface '98* 1998.

[10] Y. Zeng, C. Tan, M. Yang, C. Chiang, C. Chang, W. Tai, "A Momentum-Based Deformation System for Granular Material," *Computer Animation and Virtual Worlds* September 2007.

[11] K. Onoue, T. Nishita, "Virtual Sandbox," *Proc. of the 11th Pacific Conference on Computer Graphics and Applications* 2003.

[12] B. Benes, E. Dorjgotov, L. Arns, G. Bertoline, "Granular Material Interactive Manipulation: Touching Sand with Haptic Feedback," *Winter School of Computer Graphics Plzen, Czech republic,* 2006.

[13] Y. Visell, J. Cooperstock, B. Giordano, K. Franinovic, A. Law, S. McAdams, K. Jathal, F. Fontana, "A Vibrotactile Device for Display of Virtual Ground Materials in Walking," *Proc. of the Eurohaptics Conf.* June 2008.

[14] B. L. Giordano, S. McAdams, Y. Visell, J. Cooperstock, H. Yao, V. Hayward, "Non-Visual Identification of Walking Grounds," *Proc. of Acoustics* 2008.

[15] F. Fontana, R. Bresin, "Physics-based Sound Synthesis and Control: Crushing, Walking and Running by Crumpling Sounds," *Proc. of the Colloq. on Musical Infomatics* May 2003.

[16] E. d'Eon, "Deformers," *GPU Gems* 2004.

[17] I. Buck, T. Purcell, "A Toolkit for Computation on GPUs," *GPU Gems* 2004.

[18] S. Audet, J. Cooperstock, "Shadow Removal in Front Projection Environments Using Object Tracking," *Proc. of IEEE Workshop on Projector Camera Systems* 2007.

[19] M. Slater, M. Usoh, A. Steed, "Depth of Presence in Virtual Environments," *Presence: Teleoperators and Virtual Environments,* 3, 1994.

A shoe-based interface for ecological ground augmentation

Stefano Papetti, Federico Fontana, Marco Civolani
Dept. of Computer Science, University of Verona
Strada Le Grazie 15, 37134 Verona, Italy
name.surname@univr.it

ABSTRACT

The prototype of a wearable shoe-based interface is presented, which provides the user with auditory cues of ground. Data supplied by sensors embedded in the shoes drive a set of physics-based synthesis models running on a laptop, whose audio output is sent to a pair of shoe-mounted loudspeakers. By informing the synthesis models with ecological properties of grounds, neutral floors can be interactively augmented so as to react like they were made of a different material.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces—*Auditory (non-speech) feedback, Prototyping*

General Terms

Foot interfaces, Actuated shoes, Sonic interaction design, Ecological cues

INTRODUCTION

Among the interactive experiences that have been proposed in literature to convey information through non visual modalities, those which are mediated by the feet are still a negligible minority. Yet feet are, along with the hands, the most accurate tactile sensors we have in our external body. Augmented ground interactions based on auditory and tactile feedback represent an alternative to current, mostly visual [4], trends: by relieving the user from looking down to the floor, foot-based interfaces would afford a rich exchange of information taking place during everyday foot-based tasks, enhancing the modalities in which the environment can be experienced.

A closed-loop foot-ground interaction paradigm is proposed, which relies neither on vision nor on traditional locomotion tasks. The basic idea is to detect the foot action of users standing on a perceptually neutral floor (i.e., flat and homogeneous, such as a pavement made of concrete) and provide them with non-visual cues capable of augmenting the ground properties, so as to manipulate the ecological perception of the physical floor. Ecological cues of ground include its material (wood, marble, carpet, etc.) as well as its homogeneous or aggregate (e.g., a pathway covered with gravel or with snow) nature. A prototype application exists which already implements this paradigm, in the form of active tiles providing ecological auditory and haptic feedback in response to a force exerted over them by the foot [7]. On the other hand, our prototype aims at realizing a similar paradigm in a cost-effective wearable interface.

INTERACTION AND PHYSICAL DESIGN COMPONENTS

Our prototype (Fig. 1) is currently able to capture foot pressure thanks to a couple of FSR sensors inserted in each insole (one at the tip and one at the heel), but does not yet provide haptic feedback. Actuated shoes should avoid interferences and afford normal foot activity, however wearability requirements impose technological restrictions that to date we have solved only partially. Moreover, the forthcoming addition of haptic feedback poses further complications.

Synthetic feedback, in a multimodal augmented reality context, gives rise to a credible and unitary perception setting only if displayed no later than 10 ms after the real event [1]. In our case, the feedback is provided by a set of *physics-based models* capable of describing different foot-ground interaction events in terms of their physical dynamics. Holding this remark, we chose to minimize every possible source of temporal latency, therefore avoiding digital wireless communication (e.g., Bluetooth, Wi-Fi) from and to the shoes.

REAL-TIME FEEDBACK

In the last few years, physics-based models have progressively gained popularity in the field of non-visual human-computer interaction with everyday objects [3, 6]. To date, many models exist which are capable of simulating the dynamics of ecological events having a precise phenomenological auditory as well as haptic connotation, such as impact, friction, rolling. Once their parameters are properly tuned, by obeying to the physical laws governing the modeled event/process, they can simulate e.g. a displacement signal in response to an excitation: this signal can be interpreted as a source of air pressure perturbation, producing sound, as well as of floor vibrations for use in shoe-based haptic actuators.

The models chosen for our prototype come from the Sound Design Toolkit (SDT) [1], a software package which provides *patches* and *externals* for the DSP environments Max/MSP and Pure Data. We have found particularly suitable for experimentation the *continuous crumpling* model, which has already been exploited for synthesizing credible sounds of

[1] Freely available at http://closed.ircam.fr

Figure 1. The complete interface: the backpack hosts a laptop and an Arduino USB acquisition board which senses data from the shoes. The sensed data drive a set of real-time physics-based models running on the laptop and providing low-latency audio feedback through the shoe-mounted speakers and, in the short, through vibration actuators embedded into the shoes.

walking on aggregate grounds, such as snow and gravel [2, 7]. The crumpling model is based on a quite general impact model, on top of which a statistics of temporal impact events is superimposed. Homogeneous floors have been object of sonification experiences as well, aiming at adding resonances to neutral floors in order to change the perceived ground material. For instance, a floor made of concrete can be augmented so as to resonate like a wooden floor. In this case we have considered the aforementioned *impact* model, where a resonating object is excited by noise bursts triggered by the sensors. The rationale behind this idea, still prone to be validated, is that noise describes the tight sequence of microscopic impacts resulting from a foot-floor interaction distributed along space and time. Finally, a *friction* model is currently being fine-tuned envisioning its application to the rendering of foot sliding gestures.

CONCLUSION AND FUTURE WORK

The potential advantages of wearing actuated shoes are manifold. Scenarios where they could be profitably used include navigation aids to normally gifted as well as visually impaired people; auditory and haptic interventions for manual work in sensory deprived settings (e.g., astronautic/undersea labor, teleoperation); rehabilitation of persons affected by foot-level proprioceptive disorders. Finally, actuated shoes may contribute to improve the immersion in virtual environments, for simulation and gaming/entertainment purposes.

While our interface still needs to undergo a proper evaluation process, at this stage we can claim that some preliminary informal tests reported positive and sometimes enthusiastic user's feedback, while other users were instead more skeptical about the possibility of making profitable use of the interface.

At this point, the inclusion of the tactile modality is of primary importance. However, while physics-based synthesis provides an ideal source of force feedback for an audio-haptic interface, the engineering of active soles that must be thin, meanwhile powerful enough to generate credible stimuli with a moderate energy consumption, needs a major shift forward in the physical design of mobile haptic actuators.

Apart from some pioneering explorations, the psychophysics of foot-floor interactions is still unknown to a wide extent, especially from a quantitative point of view [5]. Should the proposed prototypes be of help in improving upon this lack of knowledge – for instance by providing opportunities for designing experiments with good level of control – then augmented shoes may ultimately find their place among the components that today are employed in multimodal augmented reality setups.

The research leading to these results has received funding from the European Community's Seventh Framework Programme under FET-Open grant agreement n°222107 NIW - Natural Interactive Walking.

REFERENCES

1. B. D. Adelstein, D. Begault, M. Anderson, and E. Wenzel. Sensitivity to haptic-audio asynchrony. In *ICMI'03: Proc. Int. Conf. on Multimodal Interfaces*, pages 73–76, New York, NY, USA, 2003. ACM.
2. R. Bresin, S. D. Monache, F. Fontana, S. Papetti, P. Polotti, and Y. Visell. Auditory feedback from continuous control of crumpling sound synthesis. In *CHI 2008 Workshop on Sonic Interaction Design*, pages 23–28, Florence, Italy, 2008.
3. P. R. Cook. Modeling Bill's gait: Analysis and parametric synthesis of walking sounds. In *Proc. Conf. on Virtual, Synthetic and Entertainment Audio*, Espoo, Finland, 2002.
4. T. Delbrück, A. M. Whatley, R. Douglas, K. Eng, K. Hepp, and P. Verschure. A tactile luminous floor for an interactive autonomous space. *Robotics and Autonomous Systems*, 55(6):433–443, 2007.
5. B. L. Giordano, S. McAdams, Y. Visell, J. Cooperstock, H.-Y. Yao, and V. Hayward. Non-visual identification of walking grounds. *J. of the Acoustical Society of America*, 123(5):3412, 2008.
6. D. Rocchesso, R. Bresin, and M. Fërnstrom. Sounding objects. *IEEE Multimedia*, 10(2), 2003.
7. Y. Visell, J. R. Cooperstock, B. L. Giordano, K. Franinovic, A. Law, S. McAdams, K. Jathal, and F. Fontana. A vibrotactile device for display of virtual ground materials in walking. In *Haptics: Perception, Devices and Scenarios*, number 5024 in Lecture Notes in Computer Science, pages 420–426. Springer, Berlin/Heidelberg, 2008.

AUDITORY REPRESENTATIONS AS LANDMARKS IN THE SOUND DESIGN SPACE

Carlo Drioli
Pietro Polotti
University of Verona
{carlo.drioli,pietro.polotti}
@univr.it

Davide Rocchesso,
Stefano Delle Monache
IUAV University of Venice
roc@iuav.it
stefano.dellemonache@gmail.com

Kamil Adiloğlu,
Robert, Anniés,
Klaus Obermayer
Berlin Institute of Technology
{kamil, robokopp, oby}
@cs.tu-berlin.de

ABSTRACT

A graphical tool for the timbre space exploration and interactive design of complex sounds by physical modeling synthesis is presented. It is built around an auditory representation of sounds based on spike functions and provides the designer with both a graphical and an auditory insight. The auditory representation of a number of reference sounds, located as landmarks in a 2D sound design space, provides the designer with an effective aid to direct his search along the paths that lie in the proximity of the most inspiring landmarks.

1 INTRODUCTION

Sound synthesis techniques are nowadays available which offer a high degree of naturalness and expressiveness. Sometimes, however, this comes at a price of a consistent complexity in both the control of real time performance, and the parametric tuning required in the sound design process. One such case is physical modeling, in which the process that produces the sound is represented by means of more or less simplified equations of the underlying physical laws. As a result, the parameters involved in the modeling should in principle be easy to understand because they have a real counterpart, and should be easy to control because our sensorial experience should mediate the action-reaction patterns to which they relate[1]. Nonetheless, the most accurate and expressive models available today are often described in terms of detailed physical relations and parameters that are not always accessible or understandable to the non specialists. Moreover, the nonlinear nature of the phenomena under observation may sometimes lead to numerical schemes that do not always reflect the behavior of the real systems in the whole parameters space. These considerations motivate the search for new tools to aid the synthesis parametric tuning, and are of particular interest in our opinion in the case of physical modeling audio synthesis.

In this paper we propose a graphical tool for the timbre space exploration and interactive design of complex sounds by physical modeling synthesis that is intended to assist the sound designer. It is aimed at exploiting the auditory representation of sounds based on spike functions and provides the designer with both a graphical and an auditory insight that may be used in place of, or combined with, the set of low-level physical parameters of the models. The graphical tool adopts a sonic landscape paradigm, in which the auditory representation of a number of reference sounds can be located as landmarks in a 2D sound design space, and provides an effective aid to direct the search along the paths that lie in the proximity of the most inspiring sonic landmarks.

The use of terminology and metaphors referring to the environment and landscapes has a rather old tradition in the field of sounds perception, especially when referred to the perception of ecological and everyday sounds. In the late 70's, the Canadian composer R. Murray Schafer introduced the term "soundscape", defined as the auditory equivalent to landscape[2], and Barry Truax published his Handbook for Acoustic Ecology[3]. The term soundscape perception is also used in a scientific context to characterize how inhabitants perceive, experience and appraise their sonic environment.

Our use of the terms "sonic landscape" and "sonic landmark" however refers to a particular organization of sounds in a 2D sonic space. The sound synthesis framework we rely on is based on a class of physical models for everyday sounds, which includes low level events (impacts and friction) and high level ones (bouncing, breaking, rolling, crumpling). This Physically-based Sound Design Tools (SDT from now on) is being developed and supported within a number of EU funded research projects on audio synthesis and sound design (Sounding Objetcs (SOB), Closing the Loop of Sound Evaluation and Design (CLOSED), Natural interactive walking (NIW))[4, 5]. An example of graphical interface for the parametric tuning of friction sounds, in-

SMC 2009, July 23-25, Porto, Portugal
Copyrights remain with the authors

cluded in the SDT and implemented in Max/Msp, is shown in Figure 1. The aim of the tool proposed here is to add to this class of tuning interfaces a perceptual representation output layer (spike functions) and a tool to perform parametric interpolations using reference auditory representations in the perceptual space. The representation of sounds based on the spike functions provides the designer with both a graphical and an auditory insight. The latter point is of extreme interest since it is based on a sort of "auditory representation of sound". The peculiarity of the spike representation is, in fact, to be lossy. This aspect turns to be an advantage, whereas the reconstructed sounds maintain the temporal articulation of the original ones but they are decolored from their original timbre characteristics. Any reconstruction of a sound has a "watery character", so that all the sounds become timbrically similar, even if distinguishable according to their temporal structure. In this way, the sound designer has at disposal a kind of "auditory-radiography" of the sounds that can be compared as sonic archetypes, helping the designer in her/his sound space exploration.

The paper is organized as follows: first, a description of the sound synthesis engine and of the spike-based auditory representation is given in Section 2; Section 3 describes the interpolation sound space and the graphical tool proposed to assist the sound designer; in Section 4, same sound design examples obtained with the tool are illustrated; Section 5 contains the conclusions and future issues.

2 DESIGN AND IMPLEMENTATION OF THE SPIKE-BASED PARAMETER INTERPOLATION TOOL

The interactive interpolation tool presented is organized as a client-server distributed application, in which the client side hosts the user graphical front-end based on the auditory spike sound representation, and the server side hosts the SDT audio synthesis engine. In the following we provide some details on the sound synthesis algorithm used to generate the reference and the new sounds, on the spike analysis framework used to graphically represent the sounds, and the properties of the interpolation in the 2D space.

2.1 The sound synthesis engine

The sound synthesis chosen to illustrate our design tool is a physical modeling implementation of friction sounds synthesis, included in the aforementioned SDT package.

The scope of the SDT is to provide a platform of sound synthesis tools that interaction designers can easily exploit in their sketching activities and that can be run on common real time software such as Max/MSP and Pd. The aim is also to provide the patches with a set of side tools to easily manage projects and that can be of help when working with

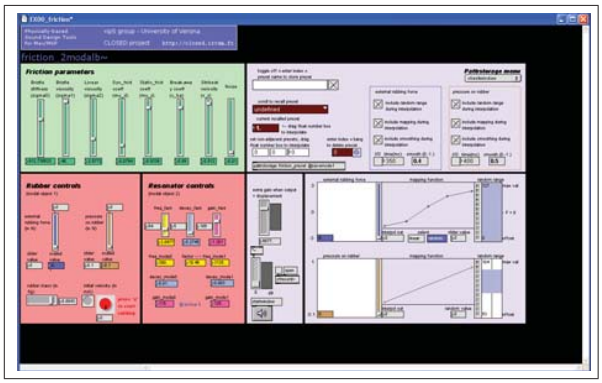


Figure 1. The friction MAX/Pd patch, contained in the SDT synthesis tools package. The user can tune up to 27 synthesis parameters representing the driving forces, the characteristics of the modal resonators, and the characteristics of the nonlinear interaction.

acquisition boards and sensors. Most of the models contained in the SDT, aimed at reproducing sounds from solid objects interaction, are structured as two resonating objects interacting by means of a contact model. The friction model specifically referred to here, relies on a description of the average behavior of a multitude of micro-contacts made by hypothetical bristles extending from each of two sliding surfaces. A modal decomposition is adopted for both interacting objects, leading to a first parametric subset including modes frequency, decay factors and gain. The remaining low-level parameters are related to the interaction mechanisms and to the interaction force specification. To gain an insight of the phenomenological role of the low-level physical parameters of the friction model, and of what a sound designer can be asked to deal with, a description is given in Table 1, and are visible in the SDT tuning interface of Figure 1 . Further details on the friction model can be found in [4], Chap. 8.

Sym.	Physical Description	Phenomenological Description
σ_0	bristle stiffness	affects the evolution of mode lock-in
σ_1	bristle dissipation	affects the sound bandwidth
σ_2	viscous friction	affects the speed of timbre evolution and pitch
σ_3	noise coefficient	affects the perceived surface roughness
μ_d	dynamic friction coeff.	high values reduce the sound bandwidth
μ_s	static friction coeff.	affects the smoothness of sound attack
v_s	Stribeck velocity	affects the smoothness of sound attack
f_N	normal force	high values give rougher and louder sounds

Table 1. A phenomenological guide to the friction model parameters.

2.2 Spike representation

In a classical signal representation approach, overlapping discrete blocks are used. However, in this method, in particular for non-stationary signals, a small shift of a block can cause a large change in the representation, depending on where an acoustic event falls within the block. A sparse, shiftable representation method based on atom-like filter functions can solve the problem. Hence, following [6], a sound signal $x(t)$ can be approximated as a linear combination of K filter functions, the so-called spikes. In this study, we employed the Gammatone functions $\gamma_{f_k,t_k}(t)$ with amplitudes a_k and residual $\epsilon_{K+1}(t)$:

$$x(t) \quad = \quad \sum_{k=1}^{K} a_k \gamma_{f_k,t_k}(t) + \epsilon_{K+1}(t). \qquad (1)$$

A Gammatone function is defined by its center frequency f_k and its filter order. In a Gammatone filter bank, the filter center frequencies are distributed across the frequency axis in proportion to their bandwidth. The shape of the magnitude characteristics of the fourth order Gammatone filter approximates the magnitude characteristics of the human auditory filter in a proper way [7]. Hence, these filter functions have a biological background.

Each spike $s_k = (t_k, f_k, a_k)$ is composed of the temporal offset t_k, the center frequency f_k of the corresponding Gammatone filter and the amplitude a_k.

For a pre-determined number of spikes K, the optimal coefficients a_k, f_k, t_k, in terms of a minimal residual signal $\epsilon(t)$ have to be found. We employ matching pursuit [8, 6] to iteratively determine the spikes s_k and to minimize the residual. The algorithm is initialized by setting the residual

$$\epsilon_1(t) := x(t). \qquad (2)$$

Then for $1 \leq k \leq K$, the optimal time offset t_k and center frequency f_k are determined iteratively so that the Gammatone filter $\gamma_{f_k,t_k}(t)$ maximally correlates with the signal $\epsilon_k(t)$:

$$(f_k, t_k) = \text{argmax}_{f,t^*} < \epsilon_k(t), \gamma_{m_{f,t^*}}(t) > . \qquad (3)$$

In order to perform the scalar product shown in Equation 3 between the filters and the signal, the filters should have certain window lengths in time domain. In his auditory toolbox, Slaney [9] used fixed window lengths for each filter in the filterbank. However, the energy levels decrease much slower for the low frequency filters than for the high frequency filters. Therefore, we adapted the window lengths considering the center frequencies. We calculated the positions within each filter, where the total energy in the time envelope falls to its thousandth. We used these position values in time as the window lengths to calculate the scalar products. Hence, the amplitude of the k^{th} spike is defined to be the scalar product between the corresponding filter, with its window length, and the residual signal, as follows:

$$a_k := < \epsilon_k(t), \gamma_{f_k,t_k}(t) > . \qquad (4)$$

Then we update the residual by

$$\epsilon_{k+1}(t) = \epsilon_k(t) - a_k \gamma_{f_k,t_k}(t) \qquad (5)$$

Finally, for $k = K$ we yield Equation 1.

By varying K within Equation 1, the SNR values of the spike code corresponding to the sparsity of the representation can be changed. Increasing K increases the SNR of the spike code. Small K (high sparsity) decreases the SNR.

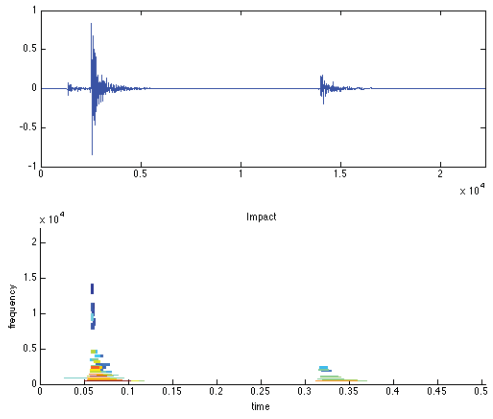


Figure 2. For an impact sound, we display the sound wave and the spike code. This picture clearly indicates how the spike code captures the skeleton of the sound.

Figure 2 shows the wave form and the spike code representation of an *impact* sound. In this example, a Gammatone filter bank of $M = 256$ filters is used for generating the spike code consisting of $K = 32$ spikes. Note that salient areas in the wave of this sound is coded with more spikes than other areas. This indicates that the spike representation captures the signal characteristics properly.

3 THE SONIC SPACE, SONIC LANDMARKS AND GUI

The sonic landscape is organized as a 2D space in which reference sounds (sonic landmarks) are located. The organization of the reference sounds may rely on perceptual criteria, derived on a statistical basis (e.g., clustering), or may be arbitrarily decided by the user.

The sonic landmarks in the sonic space form a 2D scatter points set. The sound designer may chose to generate a new

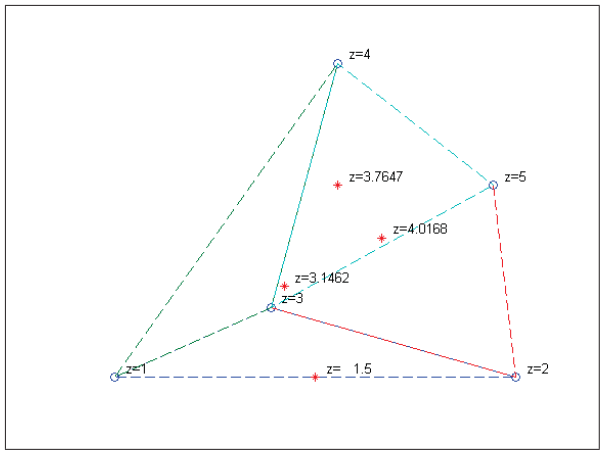


Figure 3. Scatter points interpolation: Delaunay triangulation (dashed lines) on a set of data points (circles), and four interpolated points (stars).

sound in the vicinity of a set of sonic landmarks inspired by their sonic properties. Depending on the coordinates of the new position indicated by the user, the new set of synthesis parameters is generated through an interpolation scheme accounting for the neighboring sonic landmarks. Consider a set of landmark points

$$\{(x_1, y_1, \mathbf{z}_1), (x_2, y_2, \mathbf{z}_2), \ldots, (x_n, y_n, \mathbf{z}_n)\}$$

where (x_i, y_i) is the position in the 2D space, and $\mathbf{z}_i$ is the value of the synthesis parameters vector. A convenient interpolation scheme for this class of problems is based on the following steps (for sake of simplicity, we consider a scalar parameter, z):

- Compute a grid of triangles connecting the scatter points together. One possibility, which is used here, is the Delaunay triangulation.
- For a new interpolated point, find the three vertices (x_{i1}, y_{i1}, z_{i1}), (x_{i2}, y_{i2}, z_{i2}) and (x_{i3}, y_{i3}, z_{i3}) defining the triangle in which the point is confined and compute the coefficients A, B, C, D for the equation of the plane defined by the triangle $Ax + By + Cz + D = 0$.
- Finally, compute the interpolated value through linear interpolation: $f(x, y) = -\frac{A}{C}x - \frac{B}{C}y - \frac{D}{C}$

Figure 3 gives an example of such interpolation for a small set of data points (circles) and a scalar z parameter. The interpolated data are depicted as stars.

A graphical user front-end (GUI) was implemented in Matlab and is shown in Figures 4 and 5. Its main frame represents the 2D sonic space with the sonic landmarks located in pre-defined spots (the figures show a configurations in which the landmarks were disposed as vertices of a rectangle). The location of the reference sounds can be arbitrarily decided and is specified in a configuration file loaded ad the beginning of each session. Once the landmarks sounds have been loaded and the auditory representation computed, the user can perform a number of actions, including:

- Creating a new sound by choosing with the mouse a position in the proximity of the sonic landmarks whit desirable characteristics
- Listening to landmarks and new sounds by clicking on the spike plot
- Deleting newly generated sounds which are not of interest for the user
- Saving the parametric setting of a new sound as presets in the XML-based format used by the SDT GUI
- Checking network connectivity with the Max/Msp or Pd server running the SDT sound synthesis engine

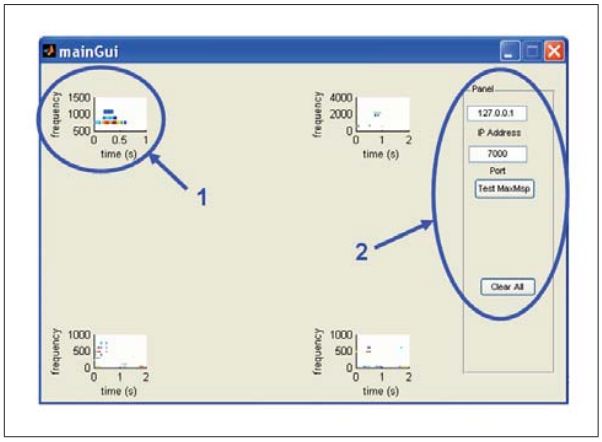


Figure 4. Matlab GUI. Spike representation of four sounds (sonic landmarks) representative of the entire sound space generated by means of the SDT friction model. By clicking in the graphical area around a spike representation it is possible to listen to the corresponding sound (see the area highlighted by circle 1). Circle 2 shows the control panel for the communication with the SDT model in Max/MSP

The spike-based GUI is connected through an OSC network layer to a Max/Msp or Pd server running the SDT sound synthesis engine. When the user confirms the position of a new sound to be generated, the GUI starts the communication with the SDT server, proceeding through the following steps: first, the new interpolated parameters are sent to the server, which updates the controller values; when done, the SDT synthesis engine is started and an auidio file is generated; finally, the GUI loads the audio file, generates the

new auditory representation, and plot the result in the 2D sonic space.

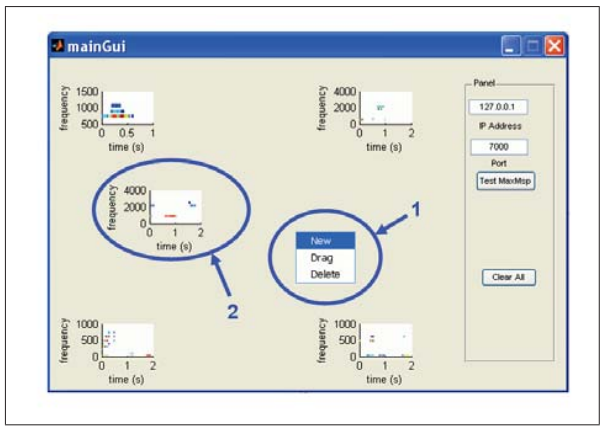


Figure 5. Creation of new sounds. By clicking in any part of the Matlab GUI it is possible to start the computation of a new sound, whose SDT synthesis parameters are calculated according to the graphical position with respect to the four sonic landmarks (see circle 1). The interpolation of the parameters of the four sonic landmarks for the creation of a new sound is linear. The new parameter set is stored as an SDT preset in xml format and acquired by the SDT model in Max/MSP. Circle 2 highlights the Spike representation of a new sound ready generated to be played in Max/MSP.

4 INTERPOLATION EXAMPLES

This section presents the result of a sound design experiment in which six reference sounds (the sonic landmarks) where organized in the 2D space as shown in Figure 6 (the spike representation of the reference sounds, and the Delaunay triangulation generated with this configuration, are shown). Six new sounds were generated by choosing six interpolation positions (depicted as red stars). The synthesis sounds resulting from the interpolated parameters, converted into the spike-based representation, are represented in Figure 7.

Perceptually, the results are in good agreement with the user's expectations, and the dynamical and spectral characteristics are recognized as actually deriving from the characteristics of the neighboring landmarks. The audio files corresponding to the sound landmarks and to the interpolated sounds are available for download at the link provided in the footnote [1] .

[1] http://mordente.sci.univr.it/∼carlodrioli/SoMuCo09/Experiments.htm

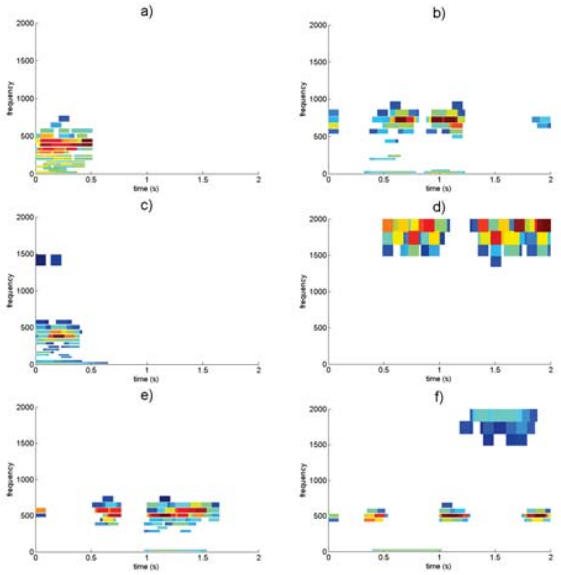


Figure 7. Interpolation example: new sounds (a), b) and c)) are generated by interpolation of points in the same triangle; the new sounds d), e) and f) are located in three different triangles

5 CONCLUSIONS

These results show that the visualization platform helps the sound designer to easily navigate through the soundscape and to find the desired sound fastly. In a following step, this visualization platform will be extended to perform a similar navigation within the soundscape towards a pre-defined (eg. recorded) sound with certain characteristics automatically. A probabilistic method will be defined to model the distribution of the input parameters of the sound model. Given the pre-defined sound, this method will automatically find the optimal input parameters for the sound model, which will produce the closest possible sound to the given sound, in terms of a spike distance. Hence, a distance measure will be defined to measure the distance between the spike code of the pre-defined sound and the spike code of the current sound produced by the sound model.

6 ACKNOWLEDGEMENTS

This work is part of the research carried out within two EU funded project: CLOSED (Closing the Loop of Sound Evaluation and Design) and NIW (Natural Interactive Walking).

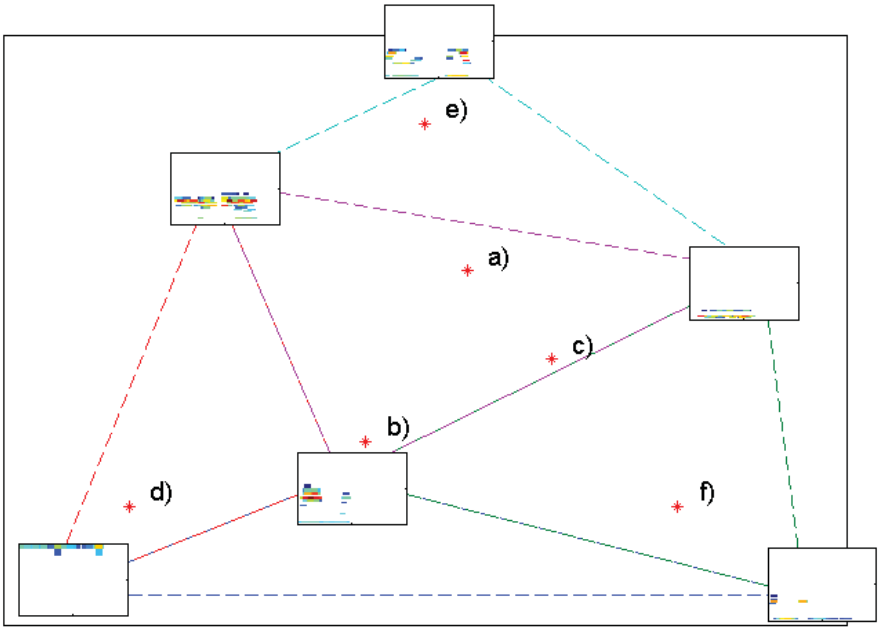


Figure 6. Interpolation example: the new sounds a), b) and c) are generated by interpolation of points in the same triangle; the new sounds d), e) and f) are located in three different triangles

7 REFERENCES

[1] Cumhur Erkut, Vesa Välimäki, Matti Karjalainen, and Henri Penttinen, *Physics-based Sound Synthesis*, vol. Sound to Sense, Sense to Sound. A State of the Art in Sound and Music Computing, chapter 8, pp. 303–343, Logos Verlag, Berlin, 2008.

[2] Murray Schafer, *The Soundscape: Our Sonic Environment and the Tuning of the World*, Rochester, Vermont: Destiny Books, 1994.

[3] Berry Truax, *Handbook for Acoustic Ecology*, ARC Publications, 1978.

[4] Davide Rocchesso and Federico Fontana Eds., *The Sounding Object*, Mondo Estremo, 2003.

[5] Stefano Delle Monache et Al., "Sound design tool: Algorithms for ecologically-founded sound synthesis," *submitted to Sound and Music Computing Conference 09*, 2009.

[6] E. Smith and M. S. Lewicki, "Efficient coding of time-relative structure using spikes," *Neural Computation*, vol. 17, pp. 19–45, 2006.

[7] R. D. Patterson and J. Holdsworth, "A functional model of neural activity patterns and auditory images," *Advances in Speech, Hearing and Language Processing*, vol. 3-B, pp. 547–563, 1992.

[8] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[9] M. Slaney, "A matlab toolbox for auditory modeling work," *Interval Research Corporation*, 1998.

ACOUSTIC RENDERING OF PARTICLE-BASED SIMULATION OF LIQUIDS IN MOTION

Carlo Drioli,
Dept. of Computer Science,
University of Verona
Verona, Italy
carlo.drioli@univr.it

Davide Rocchesso,
Dept. of Arts and Industrial Design,
IUAV University of Venice
Venice, Italy
roc@iuav.it

ABSTRACT

This paper presents an approach to the synthesis of acoustic emission due to liquids in motion. First, the models for the liquid motion description, based on a particle-based fluid dynamics representation, and for the acoustic emission are described, along with the criteria for the control of the audio algorithms through the parameters of the particles system. Then, the experimental results are discussed for a configuration representing the falling of a liquid volume into an underlying rigid container.

1. INTRODUCTION

In a wide range of applications, the exploitation of the physics underlying a natural process is an effective choice to understand, to control or to represent the process itself (examples can be found in the studies on phonation, predictive control of industrial plants, or animation through computer graphics). Physically-based process modeling has also proven to be an effective mean for real-time synthesis of sounds in many fields, including musical performance, virtual and augmented reality, HCI, interactive simulations of physical processes, computer graphics and animation[1].

In this paper we discuss the generation of the sound produced by a liquid in motion, in the particular case in which the simulation is based on physically-based numerical models of the underlying fluid dynamics. The topic is particularly interesting for applications based on computer graphics and animation, where realistically animated fluids can substantially improve the perceived quality of rendering and interaction. However, even though these models are highly effective for the computation of the parameters needed to graphically represent the motion of a fluid, (usually with a frame rate of 30 fps), their computational load at audio sampling frequencies becomes prohibitive. Thus, we decided to adopt a hybrid solution, in which the particles motion is used to control, by means of opportune mappings, the audio synthesis algorithms used for the elementary sounds, such as the ones due to the formation of bubbles under the liquid surface or the impacts emphasized by resonant cavities, responsible for the acoustic emission in liquids. In the design of the algorithms, a "cartoonification" approach was followed, with the intention of emphasizing the prototypical characteristics of the events, instead of targeting the acoustic realism typical of the audio sampling techniques[2]. However, all solutions were supported by an an accurate investigation on the physics underlying the events under study. This approach is somehow related to the PhISEM approach by P. Cook [3], in which "off-line" particle interaction simulations are performed to derive statistical collision distributions and to drive the synthesis of sounds produced by pools of interacting objects (as in maraca, sleigh bell, whistles), including water drops sounds.

The literature concerning noise emission due to liquids in motion or to liquid-liquid and liquid-solids interactions, addresses a number of phenomena ranging from the case of a single drop impact event (dripping), to more complex ones involving large liquid masses and complex evolutions in time. In general, large scale phenomena with complex time evolution are analyzed and modeled by looking at the whole phenomenon as the sum of a large number of elementary phenomena. One of the most important and acoustically relevant is the formation of air bubbles under the liquid surface, due to drops or solid objects impacting on a liquid surface [4, 5, 6]. In its simplest form, the acoustic emission is principally due to the initial impact and to the pulsation of the bubble. Depending on drop mass, impact angle, velocity, and other factors, the simple dripping event may turn into a variety of slightly different phenomena including liquid drop bouncing, spreading, and splashing (the formation of many secondary droplets at impact time), that have been accurately described and measured [7, 8]. Single-drop splashing is the simplest event which can be characterized by a structured time evolution in which the repetition of elementary events can be observed (principal drop impact followed by secondary droplet impacts). Other larger scale phenomena include sloshing, a term used in the literature to indicate the noise associated to a liquid in automotive fuel tanks which undergoes continuous shaking [9], and breaking waves [10].

This study will consider an experimental configuration representing the falling of a liquid volume into an underlying rigid container. This type of setup is typical of a class of benchmark tests known in the literature as "breaking dam", in which the dam walls holding the liquid suddenly break apart, provoking a flood wave to fill the basin.

2. METHOD

We focus here on a particle-based fluid simulation method known as smoothed particle hydrodynamics (SPH), which has interested many researchers and practitioners in the field of computer graphics and animations, due to the realistic nature of the animations of gas and liquids [11, 12]. In SPH, fluids are modeled with particles representing small fluid volumes. Each particle is described by a set of physical quantities (i.e. position, velocity, acceleration, pressure, and density) which are updated at each simulation step based on the interaction with neighbor particles. This interaction is ruled by the Navier-Stokes equations for the conservation of mass and momentum. Moreover, external forces such as the ones due to gravity or collisions with solids objects, can be accounted for in the simulation.

To address in a rigorous way the reconstruction of the acoustic field due to a system of particles in motion, one should address

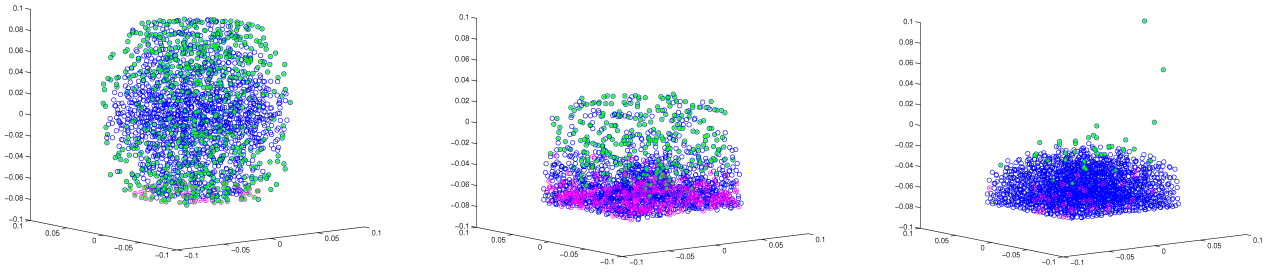


Figure 1: Three frames of the particles system representing the falling of a liquid into an empty container. a) impact of falling liquid with the empty container, b) container is filling up, and c) toward resting conditions. Different colors represent given properties or events of interest: particles with null velocity along the vertical axis are depicted in blue, falling particles are in green, and particles for which an abrupt change in the vertical velocity component (collisions) are in pink.

the calculation at audio rate of pressure wave radiation at liquid boundaries. However, this would be impractical due to computational complexity, because of the high temporal and spatial resolution requirements that this would imply. Our approach is to adopt commonly used computation rates for particles update (usually 10 to 30 frames per second) and to detect events for which acoustic emission is predictable. When such an event is detected, a corresponding acoustic event is triggered. A similar approach is found in the literature to treat the rendering of the vortex sound which accompanies air volumes in motion [13].

A preliminary step of this study concerned the selection of a minimal set of low level models representing basic events responsible for acoustic emission in liquids. In particular, the investigation focused on sounds originating from the formation of single resonating cavities (bubbles) under the liquid surface (such as in dripping, or pouring), and from surface impacts (such as those occurring between two liquid volumes or between a solid and a liquid surface) [8, 14].

In the following, we will first discuss the models used for the synthesis of the elementary acoustic events, then the analysis of the particles system and the reconstruction of the sound due to the liquid evolution will be illustrated.

2.1. Low level models for the synthesis of elementary acoustic events

The typical sound originating from a liquid drop falling into a liquid at rest is due to the formation, below the surface, of air bubbles with pulsating radius. The pulsation frequency is in the audible range, and rises with time. The law that describes the impulse response of such an event represents the radiated acoustic pressure as a damped sinusoid with rising frequency:

$$p(t) = a \sin(2\pi f(t)t) e^{-dt} \tag{1}$$

where $f(t)$ is the time-varying pulsation frequency, d is the damping factor, a is the amplitude, and t is time. The simulation of a single bubble relies on the following formulas, reported in [14]: initial frequency is $f_0 = 3/r$ (where r is the bubble radius), the damping factor is $d = 0.043 f_0 + 0.0014 f_0^{3/2}$, and the instantaneous frequency is $f(t) = f_0(1 + \sigma t)$ (where σ is the parameter which governs the slope of the frequency). The initial bubble radius is determined by a number of factors, i.e. the presence of a resting liquid level which permits the formation of the bubble under the surface, and the mass of the impacting solid or liquid. The frequency slope is related to the depth of bubble formation under the liquid surface.

At the time at which the falling drop or solid impacts the surface, a brief impulsive noise is also generated. This noise is characterized by resonances well visible in the spectrum due to the fact that the impact impulse excites cavities that are originated by air trapped between the impacting surfaces. Due to the non-rigid nature of the volumes involved in the impact, the resonances exhibit a time-varying nature and typically it is possible to observe, in the spectrograms of experimental recordings, non stationary "formant" patterns. Consequently, we decided to model this noise source by means of a subtractive synthesis in which one or two second order IIR filters are excited by a short impulse at the instant of impact. A similar model, based on a LPC analysis/synthesis scheme, was used in [15] to represent the sound of clapping hands, which has very similar perceptual characteristics to the impact noise under discussion. Figure 2 shows the spectrograms of a bubble event and of an impact event.

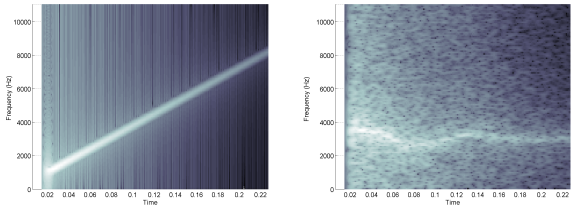


Figure 2: Spectrograms of the basic sound events employed: bubble (on the left) and impact (on the right)

2.2. Analysis and processing of the particles motion

At each frame, the particles analysis and sound generation process involves three main tasks:

1. the state of particles is updated through the SPH solver
2. the new particles configuration is analyzed, and events which are known to lead to sound emission are detected and classified
3. the detected events are mapped into low level audio event triggers with opportune synthesis parameters

A detailed description of the technique to solve the equations for the particles motion via the SPH method can be easily found in the literature (e.g., [12]). We will discuss here in detail the other two points.

The analysis of the instantaneous fluid configuration has a central role in the whole process. To this purpose, a set of relevant configurations and events, commonly observed in experimental situations involving liquids in motion, was identified. Among the configurations that were considered interesting, two are worth noting: a) the impact of a falling liquid volume with an empty container, and b) the impact of liquid volumes with the surface of a resting liquid in the container. In the first case, there is no possibility for the bubble formation phenomenon to occur, and only the impact sound due to the collision of the liquid volume with the solid surface may arise (Figure 1, panel a)). In the second case, together with the sound from liquid-liquid impact, sound from the formation of cavities under the liquid surface (bubbles) is also generated (Figure 1, panels b) and c)). The system evolution proceeds with particles falling until a terminal quiet state is reached, with some of the particles possibly bouncing a few times.

During the evolution, each particle is monitored and a list of neighbor particles is generated. This allows to obtain interesting information, such as the emergence of clusters of particles proceeding tight together (small volumes), isolated particles (drops), or the formation of noticeable configuration such as liquid-solid and liquid-air interfaces. In the experiment discussed here, during the particles analysis we rely on motion information (particles velocity, position, acceleration) and lists of neighbor particles. Based on these parameters, the following features are evaluated:

- the presence of resting liquid in the container. When the container is not empty, an estimation of the level h of liquid fallen up to the current frame is produced
- the presence of particles undergoing a collision (this is based on the acceleration)
- the identification of clusters of particles and of singleton particles in air
- formation of boundary configurations, i.e. horizontal liquid-air interfaces, at the bottom or top of a liquid volume

Once completed, the analysis of the particles configuration for the present frame is followed by the generation of control signals for the triggering of audio synthesis. The mapping between the states of the liquid description and the audio events was designed by searching for coherent relations between the parameters of particles motion and the acoustic parameters of the audio algorithms. The principal rules used to this aim are the following:

- the generation of sounds originated from the formation of air bubbles is dependent on the presence of resting liquid in the container, and is due only if $h > 0$
- for each collision detected, an impact sound is triggered, and a bubble sound is if $h > 0$. The trigger instants are modeled by a statistical uniform distribution within the audio frame corresponding to the analyzed video frame
- when clusters of collisions are detected, a bubble sound and an impact sound event are triggered, with resonance frequency inversely proportional to the cluster dimension
- for increasing values of h (increasing levels of liquid at rest in the container), an increment of the parameters controlling the bubble dimension and the bubble frequency slope (bubble collapse) is allowed.

In Figure 3, the spectrogram of the sound synthesis resulting from the adoption of the criteria described above is shown. It is possible to observe that when the first collisions occur (at around 0.3 sec), only audio impact events are generated, without any bubble sound. After a few frames (at around 0.4 sec), when the level of liquid in the container has reached a given threshold, bubble sound events, recognizable from the characteristic rising pitch, are also visible in the spectrogram. It can also be observed how the slope of the bubble pitch is lower at the beginning, for low values of h, and increases as the container fills up.

3. EXPERIMENTAL RESULTS AND DISCUSSION

Numerical simulations of the illustrated splashing event were conducted, and the analysis and audio synthesis steps were assessed. The simulation of the particles system was based on a software developed by Takashi Amada within a project for the real time animation of liquids[1]

The mean values of the statistical distributions of bubbles control parameters (bubble radius and frequency slope) and of impact resonances, were tuned empirically until the audiovisual results produced were perceptually in agreement with the average everyday experience of the experiment under study. Based on informal listening tests, we can say that the audio synthesis, even if isolated from the visual rendering, is well recognized and evokes a physical event involving the pouring of liquid into a bowl. The quality of the audio rendering is far from being as realistic as a sampled sound, however the perceptual information needed to convey the sensation of the source event is contained in the sound. With this respect, an important role is attributed in our opinion to the perception of the increasing level of liquid in the container, rendered by the parametric variations of the sound events, in particular of the bubble events. The audiovisual result, in which the synchronization between the particles motion and the timbre variations of the acoustic events is clearly perceived, is coherent with our perception of the phenomenon.[2]

4. CONCLUSIONS

In this paper we discussed the audiovisual rendering of a simple simulated experiment involving liquids in motion and based on a physically based approach. In particular, a numerical model of the fluid-dynamics based on a particle representation was used to represent the falling of a volume of liquid into an underlying empty container. The experimental results show that even by using simple models of the basilar acoustic events involved, the control based on physical models contribute to effectively render the phenomenon and to evoke the experience of complex events such as the pouring of a liquid into a container.

For the future, further developments are foreseen in the refinement of the acoustic models representing the basic sound events, with the aim of improving the degree of acoustic realism (e.g. by using data-driven sound models), and in the design of the control mappings, which are at present optimized on an empirical basis.

5. ACKNOWLEDGEMENTS

This work is part of the research carried out within two EU funded

[1] Available at http://www.ss.iij4u.or.jp/∼amada/fluid/

[2] Some audio files generated from the simulations can be downloaded here: http://mordente.sci.univr.it/∼carlodrioli/liquids_modeling

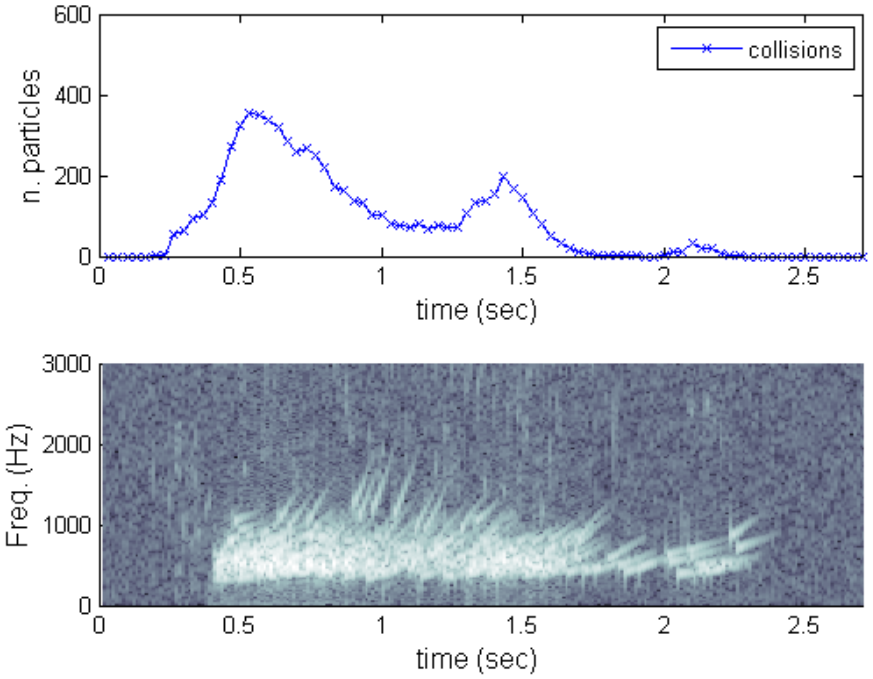


Figure 3: Result of the triggering of acoustic events based on the analysis of the particles configuration. Upper Figure shows the number of particles undergoing a collision. Lower Figure shows the spectrogram of the sound resulting from the combination of impact events and of bubble formation events.

project: CLOSED (Closing the Loop of Sound Evaluation and Design) and NIW (Natural Interactive Walking).

6. REFERENCES

[1] K. van den Doel, P. Kry, and K. Pai, "Foleyautomatic: Physically-based sound effects for interactive simulation and animation," *In Proc. ACM SIGGRAPH 01*, pp. 537–544, 2001.

[2] Davide Rocchesso, Roberto Bresin, and Mikael Fernström, "Sounding objects," *IEEE MultiMedia*, vol. 10, no. 2, pp. 42–52, 2003.

[3] P. R. Cook, "Physically informed sonic modeling (PHISM): Synthesis of percussive sounds," *Computer Music Journal*, vol. 21, no. 3, pp. 38–49, 1997.

[4] M. Minnaert, "On musical air bubbles and the sounds of running water," *Philos. Mag.*, vol. 16, pp. 235–248, 1933.

[5] Jeffrey A. Nystuen, Jr. Leo H. Ostwald, and Herman Medwin, "The hydroacoustics of a raindrop impact," *The Journal of the Acoustical Society of America*, vol. 92, no. 2, pp. 1017–1021, 1992.

[6] Hugh C. Pumphrey, L. A. Crum, and L. Bjorno, "Underwater sound produced by individual drop impacts and rainfall," *The Journal of the Acoustical Society of America*, vol. 85, no. 4, pp. 1518–1526, 1989.

[7] M. Rein, "Phenomena of liquid drop impact on solid and liquid surfaces," *Fluid Dynamics Research*, vol. 12, no. 2, pp. 61–93, 1993.

[8] G. J. Franz, "Splashes as sources of sound in liquids," *The Journal of the Acoustical Society of America*, vol. 31, no. 8, pp. 1080–1096, 1959.

[9] S. aus der Wiesche, "Computational slosh dynamics: theory and industrial application," *Computational Mechanics*, vol. 30, pp. 374–387, 2003.

[10] Steven L. Means and Richard M. Heitmeyer, "Low-frequency sound generation by an individual open-ocean breaking wave," *The Journal of the Acoustical Society of America*, vol. 110, no. 2, pp. 761–768, 2001.

[11] M. Muller, D. Charypar, and M. Gross, "Particle-based fluid simulation for interactive applications," *In Proc. ACM SIGGRAPH/Eurographics 03*, pp. 154–159, 2003.

[12] Douglas Enright, Stephen Marschner, and Ronald Fedkiw, "Animation and rendering of complex water surfaces," in *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 2002, pp. 736–744, ACM.

[13] Y.Dobashi, T.Yamamoto, and T.Nishita, "Real-time rendering of aerodynamic sound using sound textures based on computational fluid dynamics," *ACM Trans. on Graphics*, vol. 22, no. 3 (Proc. SIGGRAPH2003), pp. 732–740, 2003.

[14] K. van den Doel, "Physically-based models for liquid sounds," *ACM Transactions on Applied Perception*, vol. 2, no. 4, pp. 534–546, 2005.

[15] L. Peltola, C. Erkut, P. R. Cook, and V. Välimäki, "Synthesis of hand clapping sounds," *IEEE Trans. Audio, Speech and Language Proc.*, vol. 15, no. 3, pp. 1021–1029, 2007.

Extraction of ground reaction forces for real-time synthesis of walking sounds

Stefania Serafin, Luca Turchet and Rolf Nordahl
Medialogy, Aalborg University Copenhagen
Lautrupvang 15, DK-2700 Ballerup, sts,tur,rn@media.aau.dk

Abstract. A shoe-independent system to synthesize real-time footstep sounds on different materials has been developed. A footstep sound is considered as the result of an interaction between an exciter (the shoe) and a resonator (the floor). To achieve our goal, we propose two different solutions. The first solution is based on contact microphones attached on the exterior part of each shoe, which capture the sound of a footstep. The second approach consists on using microphones placed on the floor. In both situations, the captured sound is analysed and used to control a sound synthesis engine. We discuss advantages and disadvantages of the two approaches.

1 Introduction

Footsteps sounds represent important elements in movies and computer games. Usually such sounds are acquired from sound libraries or recorded by so-called Foley artists that put shoes in their hands and interact with different materials to simulate the act of walking. Recently, several algorithms have been proposed to simulate the sounds of walking algorithmically. One of the pioneers in this field is Perry Cook, who proposed a collection of physically informed stochastic models (PhiSM) simulating several everyday sonic events [5]. Among such algorithms the sounds of people walking on different surfaces were simulated [4]. A similar algorithm was also proposed in [7], where physically informed models simulate several stochastic surfaces.

Recently, in [6] a solution based on granular synthesis was proposed. The characteristic events of a footstep sounds were reproduced by simulating the so-called ground reaction force,i.e., the reaction force supplied by the ground at every step. The results presented in this paper are 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.

The solution proposed in this paper is inspired by the Sound of Touch installation [10]. In this project, a hand-held wand embedded with a microphone is used to capture the sound produced by the interaction with the wand and different physical textures. Such sounds are manipulated using digital convolution, to create interesting audio effects.

The role of sound in creating multimodal illusions in hand manipulations is well known. As an example, in the parchment skin illusion [8] the texture of the skin appears to change when subjects rub their hands together according to the variations of the auditory feedback, obtained by amplifying different frequency bands of the rubbing sounds.

Moreover, capturing in real-time and amplifying the sound of a hand rubbing a surface allows injured patients to substitute the lack of tactile feedback with auditory feedback [9].

The ultimate goal of this research is to recreate such illusions in foot-based interactions.

2 The setup

In this paper we are interested in augmenting traditional shoes with sensors in order to capture the act of walking and using it as input parameter for different sound synthesis algorithms. Shoes augmented with sensors have been widely developed both in the academic and commercial work. As an example, Paradiso and coworkers pioneered the development of shoes enhanced with sensors, able to capture 16 different parameters such as pressure, orientation, acceleration [11]. Such shoes were used for entertainment purpose as well as for rehabilitation studies [3]. The company Nike has also developed an accelerometer which can be attached to running shoes and connected to an iPod, in such a way that, when a person runs, the iPod tracks and reports different information.

The goal of our research is the development of a shoe-independent system able to produce synthesized

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

footsteps sounds in real time during the walk of a user.

We are particularly interested in developing a solution which requires a minimum amount of sensing technology.

The system proposed renders footsteps sounds on floors of different types of material giving to the user the impression of walking on a floor different from that the user is trampling on.

3 The setup

In this paper, we describe two different configurations of our sensing technology. Both configurations are based on the use of microphones. In the first configurations the microphones are attached to the shoes of the subjects, while on the second configuration they are placed on the floor. In Section 6 we discuss advantages and disadvantages of both configurations.

3.1 The first setup

In our preliminary experiments, we used a system composed by two small contact microphones, a sound card, a computer running Max/MSP[2] and a set of headphones.

We tried two different configurations: one with wired and one with wireless microphones. In both cases, the microphones were attached to the shoes of the test subjects. In the wired case, the two microphones (one for each shoe) were placed on the shoes by means of two adhesive gums (see Figure 1). Precisely, they were attached on the exterior part of each shoe, at about 4 cm from the sole and at the center between heel and toe. The microphones were connected to the sound card by means of wires, and the wires were attached to the trousers by means of velcro.

In detail, each microphone is a back-electret prepolarised condenser mic capsule (pressure receiver) with integrated impedance transducer. Precisely, we used the model KE4-211-1 by Sennheiser.[3]

The captured sounds were conveyed to the computer by means of a Fireface 800 sound card.[4]

In the second configuration, we used wireless contact microphones (Sennheiser SK 500 (evolution wireless G2)), that had better sensitivity and frequency response compared to the previous situation, and the obvious advantage that the users could move without any restriction. Each contact microphone was attached to the shoe in the same way as in the previous configuration. The wire connecting the microphone to the transceiver was attached to the clothes of the user.

[2]http://www.cycling74.com
[3]http://www.sennheiser.com/
[4]http://www.rme-audio.com/english/firewire/ff800.htm

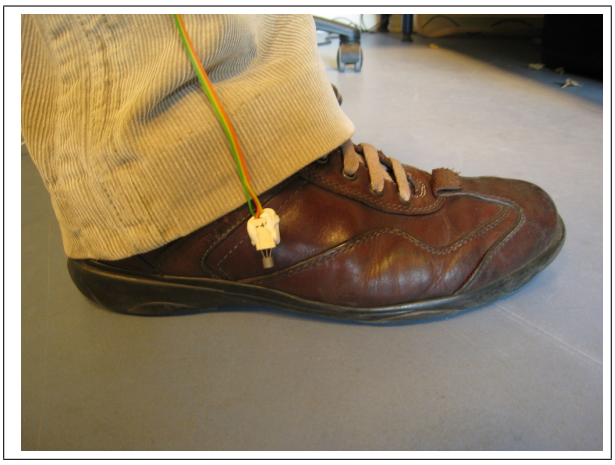


Figure 1: The contact microphone attached to the shoe.

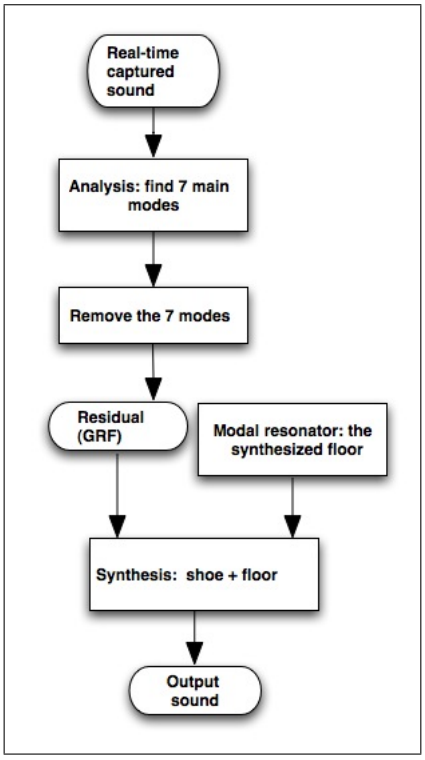


Figure 2: Block diagram of the analysis-synthesis algorithm used.

3.2 The second setup

In a second configuration we adopted a set of non-contact microphones placed on the floor. In our experiments we used the Shure BETA 91[5], a high performance condenser microphone with a tailored frequency response designed specifically for kick drums and other bass instruments. Its features made it a good candidate for our purpose of capturing the footsteps sounds. In our experiments we placed a couple of such microphones on the floor at 1.5 meters distance from each others. Such an approach still held the requisite of shoe independence, and made the requisite of wearability unnecessary.

4 Sound analysis

4.1 Extraction of the ground reaction force

A footstep sound is the result of multiple micro-impact sounds between the shoe and the floor. The set of such micro-events can be thought as an high level model of impact between an *exciter* (the shoe) and a *resonator* (the floor). In such a vision the sound captured by the microphones can be considered as a composition of both these two components.

To extract the contribution of the shoe to the sound we used two techniques. The first was based on modal analysis and resynthesis (see Figure 2), the second on linear predictive coding (LPC).

In order to achieve the final goal of producing the sensation of walking on floors made of different kinds of materials, we thought to remove the contribution of the resonator, keep the exciter and consider the latter as input for a new resonator that implements different kinds of floors. Subsequently the contribution of the shoe and of the new floor are summed in order to have a complete footstep sound. In the field of mechanics, such exciter is usually called ground reaction force (GRF), i.e., the reaction force supplied by the ground at every step. The aim of the phase of analysis consisted on finding some parameters that allowed us to extrapolate the exciter from the captured sound, i.e., finding the GRF from the acoustic waveform. Such an extrapolation consisted in removing, from the spectral representation of the sound, the main resonant frequencies, i.e., the modes. The first algorithm we used to remove such modes was by using a connection of notch filters. Such algorithm needed, as input parameters, the frequency and the bandwidth of the modes.

The algorithm has been implemented as a Max/MSP external. For computational efficiency, the external implements both the analysis and synthesis steps. The

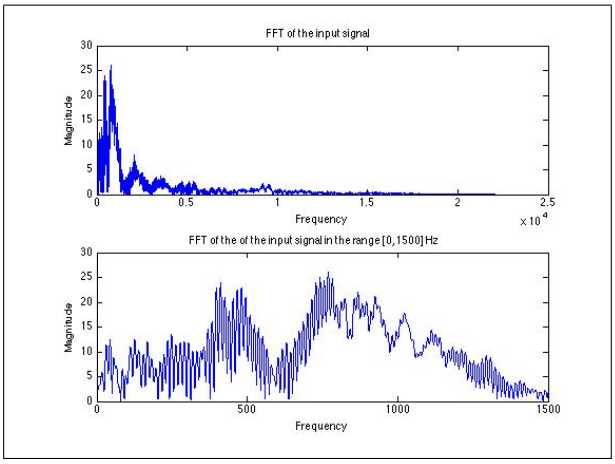


Figure 3: FFT of a sound of footstep on a concrete floor.

algorithm works as follows: the sound of a person walking is detected in real-time by the microphones described above. From this sound, the resonances corresponding to the impact of the shoe on the floor are removed, in order to extract the GRF. We assumed that the user utilizes our system walking on solid and homogeneous floors (as the concrete one for example). After some off-line analysis conducted on recorded sounds of footsteps on such types of floor, we concluded that, in average, the main modes (i.e., the main resonances present in the recordings) appear at low frequencies, and precisely they are included in the range of frequencies between 0 and 1500 Hz (see Figure 3). We established to consider seven main modes, two of which were chosen in the range (0,300) Hz, the other five in the range (300,1500) Hz. This choice is motivated by the fact that in average seven main resonances are detected in the recordings. Such resonances were detected in real-time, and removed by means of seven corresponding notch filters. We considered two distinct parts of the spectrum because we became aware of the great contribute of the modes at very low frequencies.

Figure 4 shows the time domain representation of an original single footstep sound (top) and or the same sound where the main resonances have been extracted. Such sound was used to test the algorithm offline, and obtained from the Hollywood Edge sound effects database.[6] The main two temporal components of the footstep sounds, i.e., the heel/toe event can be easily noticeable.

Figure 5 shows the spectrum of the residual sound after the removal of the seven main modes from the

[5] http://www.shure.com/

[6] www.hollywoodedge.com/

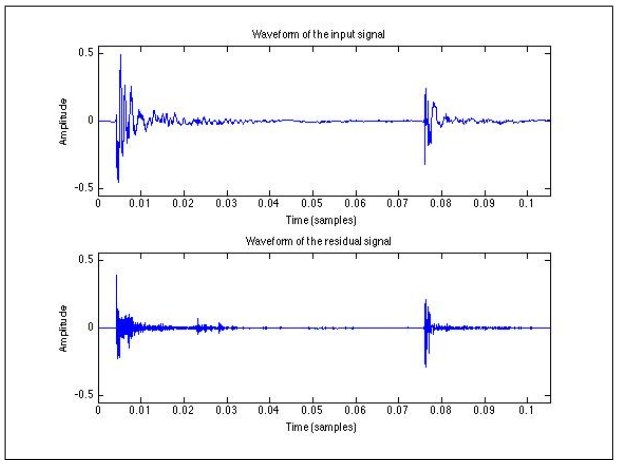


Figure 4: Time domain representation of a recorded footstep sound (top) and the corresponding residual (bottom). Notice how the two main temporal components of the footstep sound, i.e., heel and toe events, can be easily noticeable.

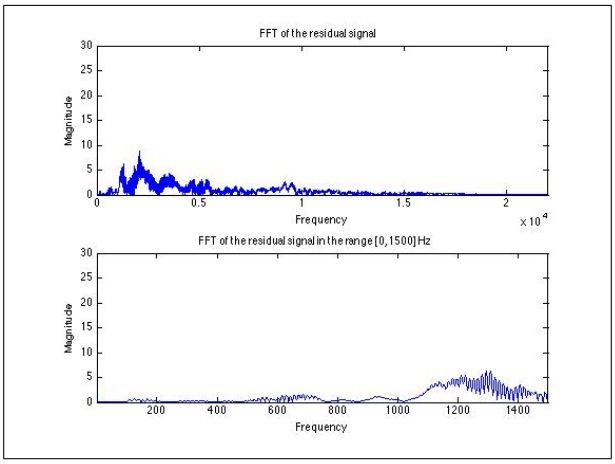


Figure 5: FFT of the residual sound analyzed in Figure 3.

sound of the footstep on the concrete floor analyzed in Figure 3.

The obtained residual sound was used as the exciter part of the real-time modal synthesizer described in the following section.

The second approach we followed to remove the main resonances of the sound and produce a residual, was the LPC. We implemented an external for Max/MSP able to produce in real-time the residual (i.e., the error signal) resulting from a LPC analysis conducted with order up to 31. Our idea was to give such a residual in input of a LPC resynthesis filter (of the same order of the analysis filter), modeling the resonance structure of different kinds of floor. We also implemented a LPC resynthesis filter in another external, whose coefficients were found thanks to an off line analysis conducted with MATLAB on recorded footstep sounds. For our analysis and resynthesis we chose an order equal to 10. Unfortunately we realized that the LPC analysis did not work as desired on the most part of the walking sounds we analyzed, since the estimated signal was very far from the original one, and consequently the signal error was very high. We noticed that the analysis performed better on some solid surfaces (in particular wood and concrete, see Figure 6). This was not a sufficient condition, since the LPC analysis on a metal floor produced bad results as 7 shows. The error signal was much higher on non homogeneous floors like gravel, dirt, leaves or snow.

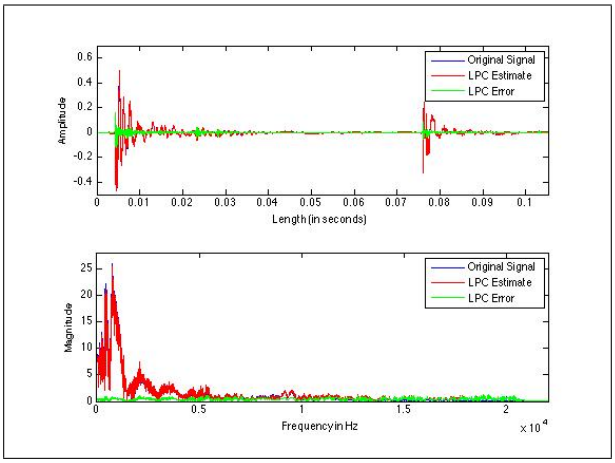


Figure 6: Original footstep sound on a concrete floor, its LPC estimate and the relative error (residual).

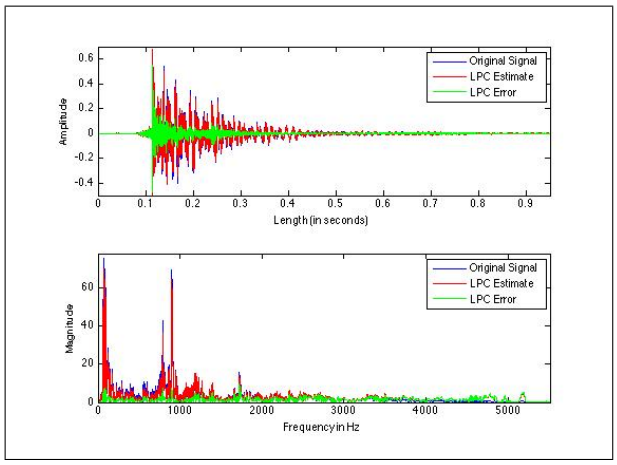


Figure 7: Original footstep sound on a metal floor, its LPC estimate and the relative error (residual).

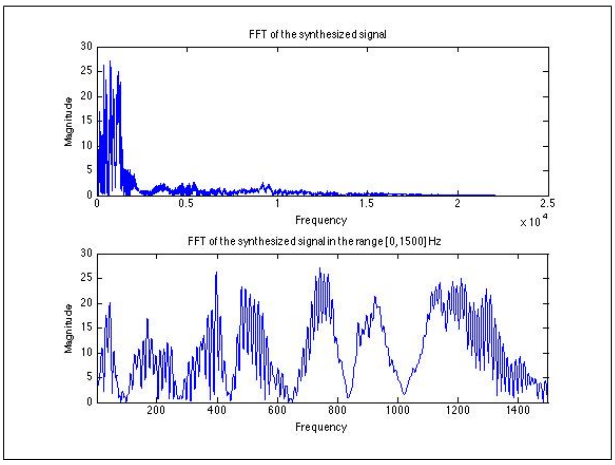


Figure 8: FFT of the synthesized signal.

5 Sound synthesis and manipulations

The first approach to synthesize the contribution of the floor (resonator) to the footsteps sounds was by using modal synthesis [1]. The modal synthesizer utilized the residual sound produced by the first analysis phase to excite the modes of the modal resonator.
A modal resonator with seven modes has been used. The parameters of the mode (amplitude, frequency and bandwidth) have been chosen after some off-line analysis on recorded footsteps sounds on various materials. Finally, the residual signal (contribution of the shoe) and the signal produced by the modal resonator (contribution of the new floor) have been summed in order to have a complete footstep sound.

Figure 8 shows the offline analysis on the sound of a footstep on a concrete floor synthesized with our system. From the figure it is easy to notice the seven modes added.

The quality of the synthesized sound turned out to be not ideal in the general case. Better results were found on the resynthesis of the same floor on which the residual extraction was performed.

The second approach consisted of providing as input the residual obtained from the LPC analysis, to a LPC resynthesis filter (of the same order of the analysis filter), modeling the resonance structure of different kinds of floor. After several trials we ascertained that such an approach did not allow us to produce the wanted result, except for the reconstruction of the same sound on which the LPC analysis was performed. Moreover the whole system was very sensitive to any background noise.

After having ascertained that the approaches above illustrated did not work, we concluded that the initial idea of extracting the sound of the shoe and use it as exciter for a resonator was not the right one. Such a conclusion was supported by the fact that it is difficult to extract the exact contribution of the shoe to the sound, and such a contribution is very different for each kind of shoe. In addition the shoes contribute a lot to the whole sound, especially in the case of homogeneous floors. For instance, the sound produced by high heels on a concrete floor is very different from the sound resulting by the walk with trainers on the same floor. We needed to extract from the sound a more general exciter that could be considered as GRF, and we found the solution in extracting the amplitude envelope from the footstep sound, as described in the following.

5.1 Extraction of ground reaction force

We decided to consider the amplitude envelope of the input sound as the overall GRF. An envelope extraction therefore corresponded to the calculation of the GRF. To perform envelope extraction we used a simple non-linear low-pass filter proposed in [12]:

$$e(n) = (1 - b(n))|x(n)| + b(n)e(n-1)$$

where

$$b = \begin{cases} b_{up} & \text{if } |x(n)| > e(n-1) \\ b_{down} & \text{otherwise} \end{cases}$$

In such an "envelope follower", the input signal is first rectified (absolute value) then the rectified signal is passed through a non-linear one-pole filter. If the rectified input is greater than the current output of the filter, a rapid "attack" tracking coefficient b_{up} is used.

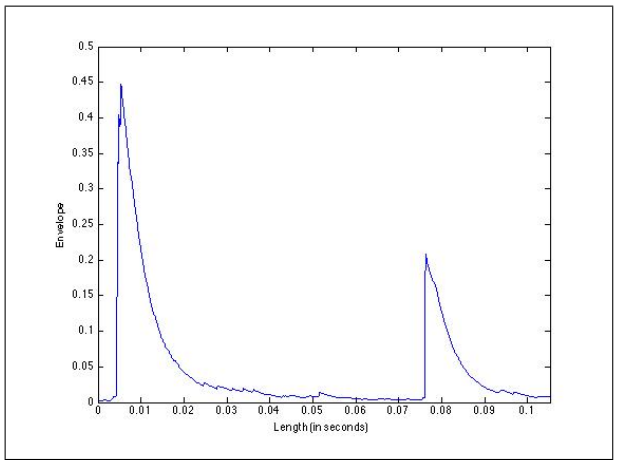


Figure 9: Envelope (GRF) extracted from the sound of Fig. 4.

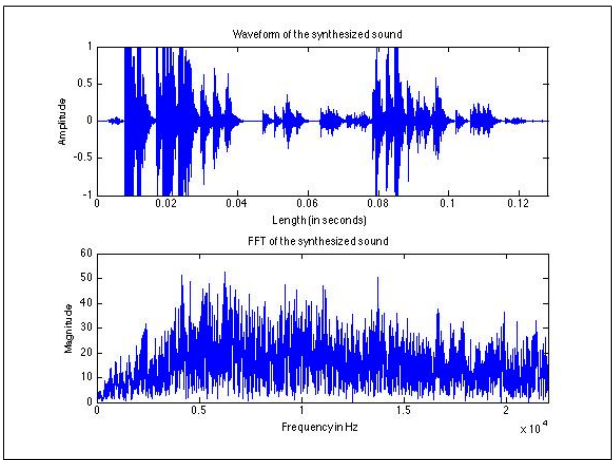


Figure 10: Synthesis of a footstep on little gravel using the GRF to control the PhiSM.

If the rectified input signal is less than the current filter output, a slower "release" coefficient b_{down} is used. The filter gain coefficient g is always set to $(1.0-b)$, to keep the total dc gain of the filter equal to 1.0. Typical values for a 22,050 Hz sample rate clapping/walking file are $b_{up} = 0.8$ and $b_{down} = 0.995$. We used such values for our purposes. Fig. 9 shows the envelope extracted form the footstep of Figure 4.

5.2 Sound synthesis algorithms

The GRF estimated with the technique described in the previous section was used to control two different sound synthesis algorithms, reproducing solid and aggregate surfaces respectively.

To synthesize solid surfaces, we used an algorithm first proposed in [2]. This algorithm physically simulates in real-time the contact between hard surfaces. In particular, we controlled one of the input parameters of the model, the impact force, by using the estimated GRF as described in the previous section. This allowed us to recreate realistic footsteps sounds. By varying the different parameters of the model it was possible to simulate the sounds of different surfaces, although a systematic mapping of physical to perceptual parameters is not in place yet.

To synthesize aggregate surfaces, we implemented the physically informed sonic models (PhiSM) [5]. The stochastic energy of the models is controlled by using the estimated GRF.

6 Testing the systems

Both approaches with the contact microphones attached to the shoes showed the same problems. First of all, during the act of walking the sound of the air against the microphones was also captured. Such an unwanted sound constituted a not negligible input error for our system. Moreover, other extra noises were clearly captured, as the noise of the trousers on the shoes, the rubbing of the two sleeves of trousers on each others, or the noise of the wires against the trousers. Finally, the system turned out to be also not so wearable or comfortable as we had imagined. Indeed, it was not so simple to attach the microphones in a right way to the shoes, as well as the wires to the trousers in order for the wires not to cause any extra sound. In addition, setting up the microphones on the users could take a long time.
Conversely, the use of the microphones on the ground did not show any particular problem, and the GRF was extracted in a clear way.

As concerns the quality of the synthesized sounds, good results have been obtained from the synthesis of big and little gravel (see Figure 10), as well as different kinds of wood (see Figure 11). The envelope extraction as GRF turned out to be the right choice for our purposes.

The systems have not been formally tested yet, but subjects which informally tried it commented on the naturaleness of the provided control.

7 Conclusions and future work

We introduced a real-time footsteps synthesizer controller by the user and which works independently from the shoe chosen. This is a feature that can distinguish our prototype from other apparati developed for the same purpose of producing, in real-time, synthesized

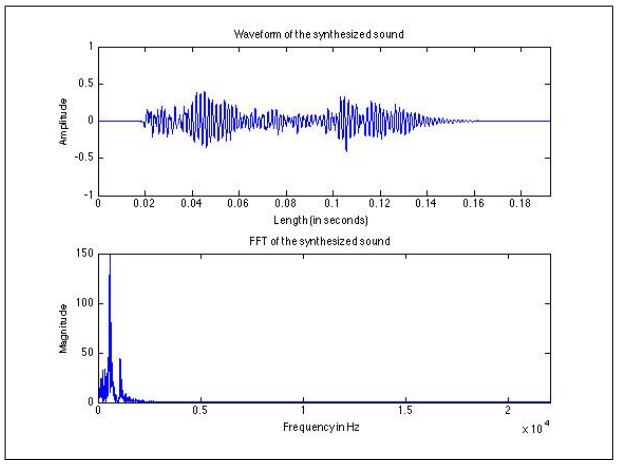


Figure 11: Synthesis of a footstep on wood floor using the GRF to control the algorithm first proposed in [2].

sounds of footsteps on different kinds of floors.

We described the different analysis and synthesis steps performed, hoping that the results of our experimentations can help researchers in related fields when choosing similar approaches.

One of the field of application of our research can be virtual reality. Indeed in future works we will test our system in a virtual environment to understand if it can improve the degree of immersion.

8 Acknowledgments

The authors gratefully acknowledge the support of the project NIW (FP7-FET-OPEN no. 222107).

References

[1] J.M. Adrien. *The missing link: Modal synthesis*. MIT Press Cambridge, MA, USA, 1991.

[2] F. Avanzini and D. Rocchesso. Modeling collision sounds: Non-linear contact force. In *Proc. COST-G6 Conf. Digital Audio Effects (DAFx-01)*, pages 61–66, 2001.

[3] AY Benbasat, SJ Morris, and JA Paradiso. A wireless modular sensor architecture and its application in on-shoe gait analysis. In *Sensors, 2003. Proceedings of IEEE*, volume 2, 2003.

[4] P. Cook. Modeling Bills 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.

[5] P.R. Cook. Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. *Computer Music Journal*, 21(3):38–49, 1997.

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

[7] 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.

[8] V. Jousmäki and R. Hari. Parchment-skin illusion: sound-biased touch. *Current Biology*, 8(6):190, 1998.

[9] M. Lanzetta, D. Perani, D. Anchisi, B. Rosen, M. Danna, P. Scifo, F. Fazio, and G. Lundborg. Early use of artificial sensibility in hand transplantation. *Scandinavian Journal of Plastic and Reconstructive Surgery and Hand Surgery*, 38(2):106–111, 2004.

[10] D. Merrill and H. Raffle. The sound of touch. 2007.

[11] J. Paradiso, K.Y. Hsiao, and E. Hu. Interactive music for instrumented dancing shoes. In *Proc. of the 1999 International Computer Music Conference*, pages 453–456, 1999.

[12] L. Peltola, C. Erkut, P.R. Cook, and V. Välimäki. Synthesis of hand clapping sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(3):1021–1029, 2007.

Toward Iconic Vibrotactile Information Display Using Floor Surfaces

Yon Visell* Alvin Law† Jeremy R. Cooperstock‡

McGill University, Department of Electrical and Computer Engineering,
Centre for Intelligent Machines and CIRMMT

ABSTRACT

This paper presents our preliminary research on haptic displays integrated in floor surfaces. We emphasize potential roles for the latter as vibrotactile communication channels that might be seamlessly integrated in everyday environments. We describe the interactive floor component that is the platform for our research, and an application we have developed to support the design of vibrotactile icon sets for display with the device. The results of a preliminary evaluation of this display method are presented, and future directions for this research are described.

1 INTRODUCTION

Tactile feedback is playing a widening role in computing applications in everyday environments, and holds special potential for improving interaction in the mobile computing domain, where attention is frequently at a premium, input devices are often small, and visual display opportunities are limited. However, the possibility of haptic interaction with artifacts or surfaces in our existing environments – such as architectural elements, transit points, or public furniture – has garnered less attention to date. The ubiquitous role that such features, including floor surfaces, play in the negotiation of our surroundings points to the possibility that they may be positioned to take on a more significant future role in haptic computing.

The research described here is motivated by the idea of exploring the latent potential in the use of floor surfaces for information display. Such surfaces are notable, from a haptic standpoint, as being among the few that are universally accepted as touchable, and with which we are most often in contact (albeit, frequently through our footwear).

1.1 Floor Surfaces as Situated Vibrotactile Displays

Recent years have seen growth in interest in haptic information displays for people in everyday environments. Most work in this area has addressed tactile displays for handheld or wearable information appliances. However, haptics has long found a role in the design of passive information displays in public spaces, notably on urban walking surfaces, where tactile ground surface indicators – passive haptic cues provided via patterns of bumps or other textures – are used to demarcate locations or paths of interest for visually impaired people or in situations in which low-lying features, such as stairs, may not be visible.

In a similar vein, many everyday ground surfaces could be profitably augmented with active displays of vibrotactile information to demarcate a location, event, condition, or process of interest. Such displays might find roles that are complementary to those that haptic displays for mobile devices have been developed for. Some simple end-user scenarios may be helpful to guide the discussion:

- A visually impaired pedestrian in an urban environment is traveling on foot and by public transportation. At a noisy urban crosswalk she receives timely information, in the form of a vibrotactile cue supplied near the curb, indicating the location and state of the crossing (figure 1.1). She reaches a building lobby, and locates the elevator. While ascending to her destination, she receives a vibrotactile cue from the elevator floor, indicating the floor number that has been reached, and instantly knows when to disembark.
- A dense crowd of pedestrians in a stadium is quickly approaching the turnstile exit. They cannot see the turnstile locations (the crowd is thick, and the lighting is poor), but vibrotactile indicators underfoot provide them with cues as to the direction of the nearest turnstile.

Figure 1: A possible end-user scenario: A crosswalk at which pedestrians receive vibrotactile cues (represented in green) indicating both the location of the crossing, and the state of the signal.

2 BACKGROUND

This project concerns vibrotactile information display to the foot via ground surfaces. We briefly review prior research in the areas of vibrotactile display, vibrotactile communication design, and haptics for the feet.

Diverse applications of vibrotactile displays have been proposed, in contexts ranging from aeronautics to web browsing. The engineering simplicity of many vibrotactile haptic devices relative to their force-reflecting counterparts has led to a shift in research emphasis toward matching displays to human capabilities and application contexts. The lightweight nature of these displays has made it possible to experiment with the presentation of such information to virtually any area of the skin, including the finger, hand, forearm, chest, abdomen, waist, back, forehead, and feet. As tactile sensory capabilities vary considerably by body location, an accounting

*e-mail: yon@cim.mcgill.ca
†e-mail:alvinlaw@cim.mcgill.ca
‡e-mail:jer@cim.mcgill.ca

of the location dependence of different aspects of tactile sensation is needed. Fortunately, many of the relevant issues have been reviewed in recent literature [11].

The design of vibrotactile communication systems for the transmission of information via touch has been studied by many researchers. In the 1950s, Geldard systematically investigated this topic, addressing problems such as display design relative to sensory capabilities and stimulation method, and stimulus set design [6]. He successfully trained several users in the understanding of a tactile language composed of 45 distinct symbols. Various authors have since approached the problem of developing design guidelines for the creation of vibrotactile symbol sets, based on combinations of perceptual, cognitive, and heuristic criteria, together with methods for their evaluation (see [15] for a review). Prior research in this area that has substantially informed our present work, as reviewed in section 4 below.

Although tactile sensation in the foot has been less studied than that of the hand, the foot has long been acknowledged as one of the most sensitive parts of the body to vibrotactile stimulation [28]. Moreover, its sensory physiology is broadly similar to that of the hand, including the same types of tactile mechanoreceptors as are present in the glabrous skin of the hand [22]. The main differences include typically much higher activation thresholds for the foot (by a factor of approximately eight [12]), and larger receptive fields (by a factor of approximately three [12]). Functional differences include the greater prehensile dexterity of the hand, the larger and more sustained forces that the feet are subjected to during locomotion, and the systematic differences in the types of activities that are performed with the hands (e.g., grasp, manipulation, and fine exploration) and the feet (e.g., stance, balance, and self-motion).

A wide range of interfaces for the feet have been previously engineered for human machine interaction (e.g., foot controlled transcribers, dental equipment, sewing machines) and human computer interaction (including foot controlled computer mice, sensing floors and shoes), but little work in these areas has aimed to profit from the integration of active haptic feedback in the interface.

A major area of recent research has concerned the engineering of locomotion interfaces for virtual environments (recently reviewed by Iwata [10] and Hollerbach [9]). However, research in this domain has predominantly focused on the challenging problems of stable, high-fidelity force-reflecting haptic interaction, with the aim of enabling the design of omnidirectional virtual walking experiences. The display of vibrotactile information underfoot for the purpose of increasing immersion for locomotion in virtual environments has only recently begun to be addressed [13]. Vibrotactile feedback via environmental surfaces has been integrated within simulation and entertainment systems (as, for example, in the vibration of vehicle simulator pedals or cockpits), but we are not aware of any systematic evaluations of the use of such displays as distinct communication channels.

Vibrotactile communication displays for presenting discrete, information-bearing stimuli to the soles of the feet, have received limited attention to date. Shoes for presenting informative stimuli to their wearers via integrated vibrotactile displays have been addressed by a few prior researchers [18, 25]. Ferber et al. investigated the design of tactile cues to aid tasks involving a force-feedback exercise machine, such as the maintenance of a target exercise rate [5]. A body of prior research has also addressed haptic feedback in the automobile cockpit (e.g., Enriquez et al. studied communicative feedback from car steering wheels [3]). Hayward and Rovan developed prototype floor tiles and in-shoe vibrotactile stimulators for providing additional feedback during computer music performance [17]. We are not aware of prior work on the design of vibrotactile information displays for pedestrians via the actuation of floor surfaces. Due, in part, to the modest technological requirements, it seems plausible that related devices, such as in-floor warning signals, have been the subject of past experimentation or invention, but we are not aware of any prior work of this nature.

3 Interactive Vibrotactile Floor Component

The floor component used in our work has been described in more detail in earlier publications [27, 26]. This device has been designed with the aim of enabling interactive vibrotactile information display to pedestrians standing or walking upon it. It is assumed that pedestrians are wearing their accustomed footwear. No special equipment need be worn in order to use the device.

In addition to actuating components, force sensing capabilities have been integrated to enable stimuli to be displayed in an interactive way, contingent upon the presence or movements of the individuals walking upon it. In separate work, we have used this interface to interactively synthesize the vibrotactile signatures that would normally be generated by walking on natural materials, such as snow or gravel [27, 13].

The device is designed to be simple to build, and to be adaptable to existing floor construction methods. The prototype, shown in figure 2, has been constructed from readily available, inexpensive materials. The tiles of the prototype shown are rigid polycarbonate of dimensions $30.5 \times 30.5 \times 1.25$ cm. Other prototypes have used plywood tiles. The tiles rest on a rigid substructure (constructed from aluminum extrusion in the model shown). An inertial motor type vibrotactile actuator (Clark Synthesis model TST-silver, single-unit retail price less than US$100) is rigidly attached to the center of the underside of each tile. A personal computer generates the vibrotactile signals in real time (described in section 4). These signals are output via the digital to analog converter of a computer audio interface, and sent to an amplifier that drives the actuators. To support display applications in which a stimulus is supplied contingent upon a user stepping upon the active area, force sensing has been integrated in the device. One force sensing resistor (Interlink model 403) encased in foam rubber 0.5 cm thick is positioned under each of the four corners of each tile. A microcontroller digitizes the force data (with a resolution of 10 bits and sampling rate of 250 Hz) and transmits it via a serial USB link to the computer.

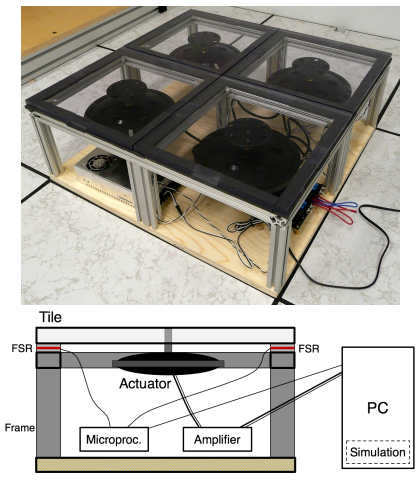


Figure 2: Top: 2×2 tile interactive floor prototype, showing the tile surface (polycarbonate sheet), vibrotactile actuator, force-sensing resistors, structural frame, and associated electronics. Bottom: Diagram of the same, including the PC running the software application.

4 Vibrotactile Icon Design Software: Parametrization and Synthesis

Vibrotactile icons can be defined as symbolic cues, in the form of temporally discrete vibrotactile stimuli, that are provided by a com-

putational artifact to inform a user about some object, place, event, or process of interest.

For a given display device, one question of research interest has been how to design such stimuli to most effectively make use of the tactile communication channel relative to the capabilities of the device and human perceptual apparatus. Prior researchers have formulated this goal through different inter-related problems, including that of maximizing information transfer (measured in bits per displayed stimulus) [19], that of creating the largest perceptually distinguishable stimulus set [14, 21], or of designing the largest set that can be separately identified at an acceptable average error rate within a task context of interest.

The design of such stimuli is frequently accomplished via a parametric representation of the family of vibrotactile signals from which the stimuli will be drawn, together with a method (heuristic or otherwise) for determining the parameters of those stimuli that are actually used. We adopt a musically-motivated design paradigm that has been introduced in prior literature on vibrotactile icons, where it has been evaluated positively. This approach views stimulus as composed of a short-time structure akin to a musical note (with a duration of perhaps 500 ms or less), and a longer time structure consisting of a motif or musical pattern composed of the note-like entities [7, 1, 24]. Parameters for describing a short-time signal of this type have often been borrowed from those that are employed within musical signal processing and sound synthesis. They can include variables such as fundamental frequency, harmonic content, waveform shape, amplitude temporal envelope, and duration. Other, psychophysically-inspired characteristics such as roughness (often specified as an amplitude modulation) or spectral centroid are also used. Longer time patterns are typically specified in terms of structural arrangements of such notes, characterized according to musical features such as rhythm, dynamics, and tempo [24, 1, 21]. Such structures may last one second or longer for each stimulus. The parameter ranges to be used can be limited by sensory and psychophysical considerations. Briefly, vibrotactile displays are often designed primarily to target the Pacinian corpuscles, which are cutaneous tactile receptors with a maximum sensitivity around 250 Hz and usable bandwidth of only a few hundred Hertz. Thus, the usable frequency bandwidth is much narrower than in audition, for which such parametric representations were first developed. Other constraints can be determined from perceptual effects such as temporal masking, dynamic range, and temporal event sensitivity, as reviewed in recent literature on vibrotactile display and interaction design [11, 7].

Spatial encoding is another interesting degree of freedom that is often used for design [20, 23]. The floor display discussed here is limited in this respect resolution, although it can be used to present distinct signals to each foot, at each location of a distributed floor area, or in response to each temporal footfall. Here, we ignore these degrees of freedom, and focus on the design of stimuli for an area of floor surface that is actuated in a spatially uniform way.

Additional considerations arise from the distinctive nature of the area of the skin that is used for display – in our case, the feet. Apart from the sensory differences mentioned in section 2, the design of such displays is complicated by the potential mobility of their users, and consequent variation in contact conditions between the user and the display. This has implications for application design in real world settings. Furthermore, in realistic settings, differences in vibrotactile transmission due to variations in users' footwear may be significant. Such issues are not addressed here, and in the preliminary study we have conducted (section 5), participants remained stationary on foot throughout, and all wore identical shoes.

4.1 Haptic Icon Designer

In order to explore the design of vibrotactile stimuli to be presented through a floor surface, we developed the Haptic Icon Designer application shown in figure 4.1. The application allows a designer to specify short-time, "note"-like stimulus properties, through parameters controlling fundamental frequency, duration, harmonic content, roughness (amplitude modulation depth), and amplitude temporal envelope (specified as a piecewise linear function). Harmonic structure is generated via a nonlinearity applied to a sinusoidal signal of the specified fundamental frequency. A sum of Chebyshev polynomials is used to design the nonlinearity, since in this representation, the amplitude of the nth harmonic frequency component of the signal (i.e., the nth multiple of the fundamental) is controlled by the corresponding polynomial coefficient [16]. Together with the real-time frequency spectrum display, this allows the designer to ensure that the stimulus lies within the target frequency band (centered on 250 Hz).

The interface allows to specify longer-time structures using a musical phrase metaphor, including rhythm, duration, note amplitude (accents), repetition, and duration. These phrases are designed within a "step sequencer" paradigm that is commonly used in digital music composition. The time domain of a single phrase is quantized into 24 steps. (The actual synthesized stimulus is continuous in nature.) Notes can begin at any step and possess durations given by integer numbers of steps (length of the blue bars shown in the interface). Only a single note is able to play at any time. The amplitude of each is specified at its onset. The duration of the entire phrase is specified in milliseconds, and the phrase may be repeated an arbitrary number of times. Notes within an icon's rhythmic pattern differ only in their amplitude and duration, so that all possess the same frequency characteristics. This is a significant constraint, but greatly limits the number of parameters that must be specified for each icon.

Once specified, the icon, or set of icons, may be saved in parametric form (in xml format) via the application. Icons may be interactively recalled as required for a given experiment or demonstration. The numerical parameters may be browsed (using a separate window) if desired.

The same application interface is used to perform real-time synthesis and to initiate playback of icons via the floor device. We have, in addition, made use of this capability to implement interactive demonstration scenarios in which, as a pedestrian steps onto the tile surface, he or she is automatically presented with a vibrotactile icon. In this case, an adaptive footstep onset detector is used to identify a step onto the surface of the tile. Such a mode of interaction is relevant to potential end user scenarios such as the vibrotactile crosswalk indicator described in section 1.1.

5 Preliminary Evaluation

The goal of this pilot experiment was to investigate the extent to which vibrotactile icons can be effective for information display via floor interfaces. Concretely, the aim was to examine the identifiability of a set of icons designed using the approach described in section 4. An initial set of twelve stimuli were specified using the Haptic Icon Designer. Each possessed a distinct rhythmic pattern of between two and five notes, repeated three or four times, with one accented note at the beginning. The duration of each icon was between 1.4 and 2.4 seconds. In addition, each icon was assigned a distinct roughness, fundamental frequency, and a frequency spectrum with most energy near 250 Hz. An informal pre-test was used to eliminate four of these stimuli that seemed too similar to the others, leaving eight icons to be used in the experiment.

The experimental design was similar to those used in prior research (e.g. [8]), although in this preliminary study, we only measured identification in a single condition. The experiment was guided by two working hypotheses: The first was that the musically-inspired method used to design the stimuli would allow participants to identify 8 different icons after a short period of training, mirroring the success that such design techniques have found in

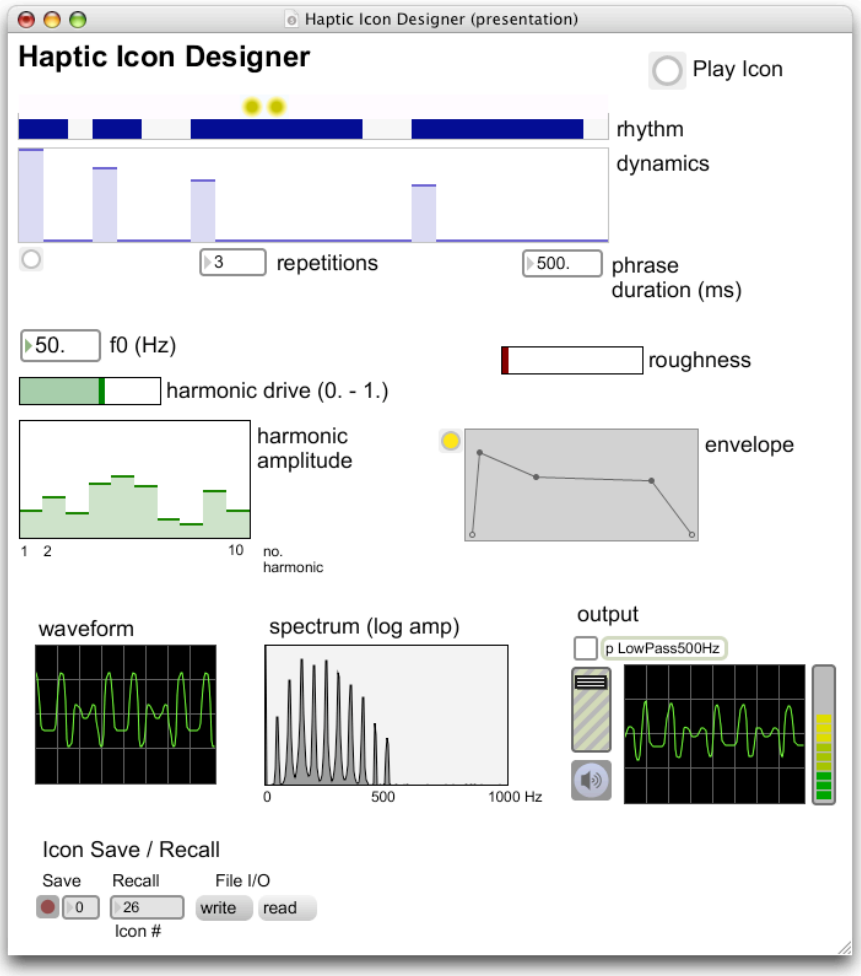


Figure 3: Screenshot of the haptic icon design software interface, during playback of an icon. The yellow dots above the note sequence (notes are represented by blue bars) denote the current time index along the musical phrase. User interface elements allow to control parameters defining the vibrotactile stimulus.

other areas of vibrotactile information display. The second was that greater confusion would be seen for shorter duration icons, due to information limitations, and for icons with more information near the onset of icon playback, perhaps due to the surprising nature of the stimulation method.

5.1 Methodology

Eight people between the ages of 20 and 39 years old took part in the experiment. Four of them were male and four were female. Five were university students. All participants were presented with the same task and stimulus sets. Each was given hard-soled men's dress shoes in his or her size to wear for the experiment. Apart from size differences, all the shoes were identical. The amplitude of vibration of the tile was adjusted for each participant, using a reference vibrotactile noise signal. At the beginning of the session, participants received instructions, together with an explanation and demonstration of the display function. At all subsequent stages (except during breaks), participants were required to wear headphones playing pink noise at a volume sufficient to mask the (generally low-level) sounds produced by the apparatus. The application used during the experiment ran on a personal computer, and was based on the vibrotactile icon design software described in section 4. In addition to the floor tile, the interface consisted of a graphical user interface with numbered buttons, one for each icon (figure 3).

The experiment was based on absolute identification, incorporating a set of 8 icons and a unique correct response for each, consisting of the numerical ID of the icon, ranging from 1 to 8. The same icons were used for all participants and all sessions of the experiment.

After an introduction to the device and interaction method, participants were given four minutes to interact with the icons used in the study, by standing on the floor tile, selecting a numerical ID, and receiving the stimulus corresponding to their selection.

The rest of the experiment was divided into six sessions. During each session, all stimuli were presented twice, in a different randomized order during each session. Thus, overall, each participant was asked to identify every vibrotactile icon a total of twelve times. Each session took an average of about four minutes to be completed. Participants were presented with stimuli sequentially during each session. At each presentation, they could press a button to play the stimulus up to four times before supplying a response. Feedback, in the form of the correct stimulus ID, was provided after each response was given. As in previous studies [8], the reason for providing feedback was to facilitate the assessment of recognition after learning and rate of learning throughout the experiment.

5.2 Results and Discussion

A log of the stimuli and responses was recorded by the application throughout the experiments. Participants were also interviewed following the experiment.

Although the icon set used in this preliminary experiment was small, the task was difficult, because it required participants to both distinguish the icons from each other, and to learn a symbolic association to the index of the icon, and to do so with very little training (a total of approximately 28 minutes, on average), after only 20 reinforced presentations of each stimulus.

The correct identification rate after six sessions of testing with enforced learning (feedback), averaged between all participants, was 55%, with a between-participant standard deviation of 25%. Chance performance would correspond to 12%. Improvements during the course of enforced learning varied considerably between individuals. One individual showed consistent improvement between sessions, attaining a 94% average correct identification rate in session 6, while others showed nearly no improvement. Average correct identification rates during each session are shown in table 1. The performance for each individual during session 6 is summarized in figure 5.2. Several participants reported feeling mental fatigue by the end of the study, which is one possible explanation for the increased average variation and reduced improvement during the last two sessions.

The results obtained are roughly comparable to published results on absolute identification of vibrotactile stimuli via manual interfaces after short periods of learning. For example, Enriquez et al. report average identification performance of 73% (relative to an expected chance performance rate of 33%) after an average of 20 minutes of training [4]. As noted below, a more direct comparison with performance using a manual vibrotactile display would be interesting.

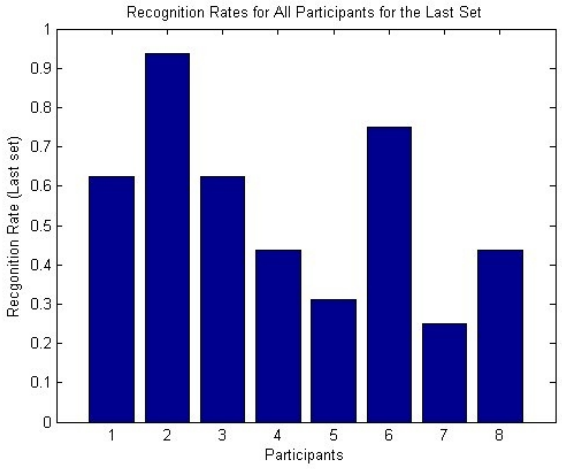


Figure 4: Identification rate (percent correct) for each participant, after 6 sessions of assessment with enforced learning.

The confusion matrix for the stimuli in the experiment is shown in figure 5.2, averaged between all sessions and participants. The least confused stimulus was identified at an average rate of 61%, while the most confused stimulus was identified at an average rate of only 25%. Comparing these confusions with the icon set did not reveal any easily discernable property of the least confused stimulus that might have caused it to be confused, nor was there any clear difference in identifiability of shorter and longer stimuli, or of stimuli with different numbers of notes. Due to the limited data available, it was not possible to infer more precise design guidelines, but we anticipate presenting further data in a final version of this paper.

Table 1: Average Performance During Each Session

Session	Average Correct	Standard Dev.
1	37%	15%
2	41%	13%
3	44%	16%
4	55%	13%
5	51%	24%
6	55%	23%

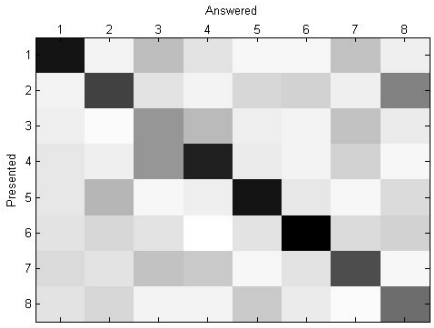


Figure 5: Confusion pattern for the 8 stimuli in the experiment, averaged between all sessions and participants.

Some participants felt that they were limited by the difficulty of the association part of the experiment – that is, the task of associating the stimuli to numbers from a list. Three of them suggested that the task would have been facilitated by a non-numerical semantic or mnemonic symbol, such as an animal name. Such an approach could hold potential for further improving the identification performance seen here (albeit in a way that does not speak to the design of the stimuli themselves). In adhering to a simple numerical labeling, we aimed to avoid introducing unknown effects linked to the relative fitness of the mnemonic(s) chosen. An earlier study by Enriquez et al. found no difference in learning and 2-week recall of arbitrarily assigned (abstract) vs. user-chosen (semantic) labels for 10 perceptually optimized vibrotactile icons [2]. The implication is that icon set selection may be more important than labeling, and that, given amenable conditions, an arbitrary association may be learned. Assuming this to be true does not, of course, rule out the possibility that the preliminary results presented here are degraded by the association task. However, it would be difficult to argue that these results are optimistic, which is arguably more important for a feasibility study such as ours.

Participants reported adopting diverse cognitive strategies for memorizing the signals, attending to frequency, perceived amplitude, pattern characteristics, or other features.

This study also aimed to assess the acceptability of the stimulation method to potential users. Participants were asked to rate the vibrotactile stimuli on a five point scale from 5 (comfortable) to 1 (uncomfortable). The average rating was 3.44 with a standard deviation of 1.1. No participants rated the stimuli as uncomfortable (although some reported growing tired of standing in place during each session.)

6 Conclusion

In this paper, we have described our continuing research on the display of vibrotactile information via floor interfaces. The low cost and robust nature of the components involved suggest that devices

based on this principle could become a meaningful addition to everyday contexts that might profit from the additional information channel. Other advantages include the ubiquitous nature of contact with the potential display surface, and the fact that users of such a system do not require any special equipment.

In order for this model of information display and interaction to achieve relevance and acceptance, design guidelines appropriate to it would need to be formulated. In turn, this would require a better understanding of the distinctive properties of the display relative to tactile perception via the feet.

6.1 Future work

As noted above, one can readily identify distinguishing features of the floor display that should be investigated further, not least effects that may be related to the unusually forceful coupling between the body and the actuated surface. This coupling allows, when desired, to transmit vibrations that propagate beyond the soles of the feet.

Two participants suggested the display might be improved if the note-level stimuli were less similar to oscillatory patterns, and more like transient knock or impact events. This is something we intend to explore through the development of a second, *Ecological* Haptic Icon Design application. In that approach, oscillatory note-level events are to be replaced by impact transients, which are synthesized in real time, and controlled through physical parameters such as hardness and contact shape. We intend to evaluate the relative fitness of such cues in a future study.

It would also be interesting to compare the floor display with a previously studied vibrotactile display (for example, one grasped in the hand) using stimuli that are as similar as possible between the two interfaces, and this is something we plan to address in future work.

Finally, the suitability of these cues to situations in which a person addressed by the display is either preoccupied with some form of workload, or is physically moving, remains to be determined.

REFERENCES

[1] L. M. Brown, S. A.Brewster, and H. C. Purchase. Tactile crescendos and sforzandos: Applying musical techniques to tactile icon design. In *Proc. of the ACM 2006 Conference on Human Factors in Computing Systems*, 2006.

[2] M. Enriquez and K. Maclean. The role of choice in longitudinal recall of meaningful tactile signals. In *Proc. of IEEE Haptics Symp.*, Reno, NV, 2008.

[3] M. Enriquez and K. E. MacLean. Impact of haptic warning signal reliability in a time-and-safety-critical task. In *Proc of the 12th Annual Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems*, 2004.

[4] M. Enriquez, K. E. MacLean, and C. Chita. Haptic phonemes: Basic building blocks of haptic communication. In *Proceedings of the 8th International Conference on Multimodal Interfaces*, 2006.

[5] A. R. Ferber, M. Peshkin, and J. E. Colgate. Using haptic communications with the leg to maintain exercise intensity. In *Proc. of Robot and Human interactive Communication*, 2007.

[6] F. A. Geldard. Adventures in tactile literacy. *Amer. Psychol.*, 12:115–124, 1957.

[7] E. Gunther, G. Davenport, and S. O'Modhrain. Cutaneous grooves: Composing for the sense of touch. In *Proc. of the Conf. on New Instruments for Musical Expression*, 2002.

[8] E. Hoggan and S. Brewster. Designing audio and tactile crossmodal icons for mobile devices. In *Proc. of ACM ICMI*, 2007.

[9] J. Hollerbach. Locomotion interfaces and rendering. In M. Lin and M. Otaduy, editors, *Haptic Rendering: Foundations, Algorithms and Applications*. A K Peters, Ltd, 2008.

[10] H. Iwata. Haptic interface. In A. Sears and J. A. Jacko, editors, *The Human-Computer Interaction Handbook*. Lawrence Erlbaum Assoc., New York, 2nd edition, 2008.

[11] L. A. Jones and N. B. Sarter. Tactile displays: Guidance for their design and application. *Human Factors*, 50(1):90–111, 2008.

[12] P. M. Kennedy and J. T. Inglis. Distribution and behaviour of glabrous cutaneous receptors in the human foot sole. *The Journal of Physiology*, 583(3), 2002.

[13] A. W. Law, B. V. Peck, Y. Visell, P. G. Kry, and J. R. Cooperstock. A multi-modal floor-space for displaying material deformation underfoot in virtual reality. In *Proc. of the IEEE Intl. Workshop on Haptic Audio Visual Environments and Their Applications*, 2008.

[14] J. Luk, J. Pasquero, S. Little, K. E. MacLean, V. Levesque, and V. Hayward. A role for haptics in mobile interaction: Initial design using a handheld tactile display prototype. In *Proc. of the ACM 2006 Conference on Human Factors in Computing Systems*, 2006.

[15] J. Pasquero. Survey on communication through touch. Technical Report TR-CIM-06.04, Center for Intelligent Machines, McGill University, 2006.

[16] C. Roads. *The Computer Music Tutorial*. MIT Press, 1996.

[17] J. Rovan and V. Hayward. Typology of tactile sounds and their synthesis in gesture-driven computer music performance. In M. Wanderley and M. Battier, editors, *Trends in Gestural Control of Music*. Editions IRCAM, 2000.

[18] A. Rovers and H. van Essen. FootIO – design and evaluation of a device to enable foot interaction over a computer network – design and evaluation of a device to enable foot interaction over a computer network. In *Proc. of Eurohaptics*, 2005.

[19] H. Z. Tan, N. I. Durlach, C. M. Reed, and W. M. Rabinowitz. Information transmission with a multifinger tactual display. *Perception and Psychophysics*, 61(6):993–1008, 1999.

[20] H. Z. Tan and A. Pentland. Tactual displays for wearable computing. *Personal Technologies*, 1:225–230, 1997.

[21] D. Ternes and K. E. MacLean. Designing large sets of haptic icons with rhythm. In *Proc. of Eurohaptics*, 2008.

[22] M. Trulsson. Microreceptive afferents in the human sural nerve. *Exp Brain Res*, 137:111–116, 2001.

[23] J. B. F. van Erp. Presenting directions with a vibrotactile torso display. *Ergonomics*, 48(3):302–313, 2005.

[24] J. B. F. van Erp and M. M. A. Spape. Distilling the underlying dimensions of tactile melodies. In *Proc. of Eurohaptics*, 2003.

[25] R. Velazquez. On-shoe tactile display. In *Proc. of IEEE Haptic Audio-Visual Environments*, 2008.

[26] Y. Visell, J. Cooperstock, and K. Franinovic. The ecotile: An architectural platform for audio-haptic simulation in walking. In *Proc. of the 4th Intl. Conf. on Enactive Interfaces*, 2007.

[27] Y. Visell, J. Cooperstock, B. L. Giordano, K. Franinovic, A. Law, S. McAdams, K. Jathal, and F. Fontana. A vibrotactile device for display of virtual ground materials in walking. In *Proc. of Eurohaptics 2008*, 2008.

[28] A. Wilska. On the vibrational sensitivity in different regions of the body surface. *Acta Physiologica Scandinavica*, 31:285–289, 1954.

Available online at www.sciencedirect.com
ScienceDirect
Int. J. Human-Computer Studies 67 (2009) 947–959
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Sound design and perception in walking interactions

Y. Visell[a,b,*], F. Fontana[c], B.L. Giordano[d], R. Nordahl[e], S. Serafin[e], R. Bresin[f]

[a]McGill University, Centre for Intelligent Machines and CIRMMT, Montreal, Canada
[b]Zurich University of the Arts, Zurich, Switzerland
[c]University of Verona, Dipartimento di Informatica, Verona, Italy
[d]McGill University, CIRMMT and Schulich School of Music, Montreal, Canada
[e]Medialogy, Aalborg University at Copenhagen, Ballerup, Denmark
[f]KTH Royal Institute of Technology, CSC School of Computer Science and Communication, Stockholm, Sweden

Received 20 September 2008; received in revised form 1 June 2009; accepted 21 July 2009
Available online 6 August 2009

Abstract

This paper reviews the state of the art in the display and perception of walking generated sounds and tactile vibrations, and their current and potential future uses in interactive systems. As non-visual information sources that are closely linked to human activities in diverse environments, such signals are capable of communicating about the spaces we traverse and activities we encounter in familiar and intuitive ways. However, in order for them to be effectively employed in human–computer interfaces, significant knowledge is required in areas including the perception of acoustic signatures of walking, and the design, engineering, and evaluation of interfaces that utilize them. Much of this expertise has accumulated in recent years, although many questions remain to be explored. We highlight past work and current research directions in this multidisciplinary area of investigation, and point to potential future trends.
© 2009 Elsevier Ltd. All rights reserved.

Keywords: Auditory display; Vibrotactile display; Interaction design; Walking interfaces

1. Introduction

Just as walking is fundamental to our negotiation of natural environments, it is of increasing relevance to interaction with computational systems. Contact interactions between our feet and the ground play important roles in generating information salient to locomotion control and planning in natural environments, and to the understanding of structures and events in them. Although much of this information is communicated as sound, the latter has been relatively neglected in past research related to walking in human–computer interaction. Consequently, a better understanding of the perception of walking sounds, and the way they may be rendered and displayed, is needed in order for new and existing human–computer interfaces to effectively make use of these channels. Such developments may hold potential to advance the state of the art in areas such as wearable computers, intelligent environments, and virtual reality. For example, in the ubiquitous computing domain, benefits could be foreseen for a range of new and emerging applications utilizing human locomotion and navigation as means for interaction with digital information (Gaye et al., 2006; Froehlich et al., 2009).

It is important to acknowledge that walking sounds have long played an important role in audiovisual media. In film, footsteps are acknowledged for their ability to signify unseen action, to lend a sense of movement to an otherwise static scene, and to modulate the perception of visible activities. In his seminal work on film sound, Chion writes 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 (Chion, 1994). The aim of this paper is to highlight the importance of interdisciplinary research surrounding

*Corresponding author at: McGill University, Centre for Intelligent Machines and CIRMMT, Montreal, Canada. Tel.: +1 514 967 1648; fax: +1 418 780 6764.
E-mail addresses: yon@cim.mcgill.ca (Y. Visell), federico.fontana@univr.it (F. Fontana), bruno.giordano@music.mcgill.ca (B.L. Giordano), rn@media.aau.dk (R. Nordahl), sts@imi.aau.dk (S. Serafin), roberto@kth.se (R. Bresin).

1071-5819/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2009.07.007

sound information in walking for the design of human-interactive systems. In retaining this focus, we address aspects of walking experiences that are seldom investigated in real or virtual contexts. Two potential future scenarios may lend concreteness to the discussion:

- A tourist using a smartphone is able to follow navigational cues that the device supplies by augmenting the sound of his footsteps as if he were walking along a cobblestone trail.
- A search and rescue worker is training in a virtual environment simulation of a dangerous rock canyon area. She receives realistic multimodal cues from the ground surface in the simulator, heightening her sense of immersion.

This article intends to point toward fundamental areas of knowledge needed to effectively realize such applications.

1.1. Foot–ground interactions and their signatures

It is almost a truism to say that self-motion is the most fundamental function of walking. Therefore, it is not surprising that the scientific literature has predominantly attended to questions linked to the biomechanics of human locomotion, and to the systems and processes underlying motor behavior on foot, including the integration of multisensory information subserving planning and control.

Walking is a periodic activity, and a single period is known as the gait cycle. Typical human walking rates are between 75 and 125 steps per minute, corresponding to a fundamental frequency of 1.25–2.08 Hz (Ekimov and Sabatier, 2008). It can be divided into two temporal phases—those of stance and swing. Stance can be characterized in terms of foot position and contact, decomposed into initial heel strike, followed by foot flat, heel off, knee flexion, and toe off (Li et al., 1991). The subsequent swing phase is composed of an initial swing, beginning at toe off. It proceeds to the mid-swing period, when the knee reaches maximum flexion, until the terminal swing, which begins when the tibia is vertical and ends when the reference foot touches the ground. Thus, the gait cycle is characterized by a mixture of postural attributes (e.g., the degree of flexion at the knee) and contact attributes (presence and degree of contact between the plantar area of the foot and the ground). One also distinguishes the several time scales involved, including those of the walking tempo or pace, the individual footstep, encompassing one stance period, and relatively discrete events such as heel strike and toe slap (Fig. 1).

The net force $\mathbf{F}$ exerted by the foot against the ground can be represented by a time varying spectrum $\mathbf{F}(\omega, t)$,

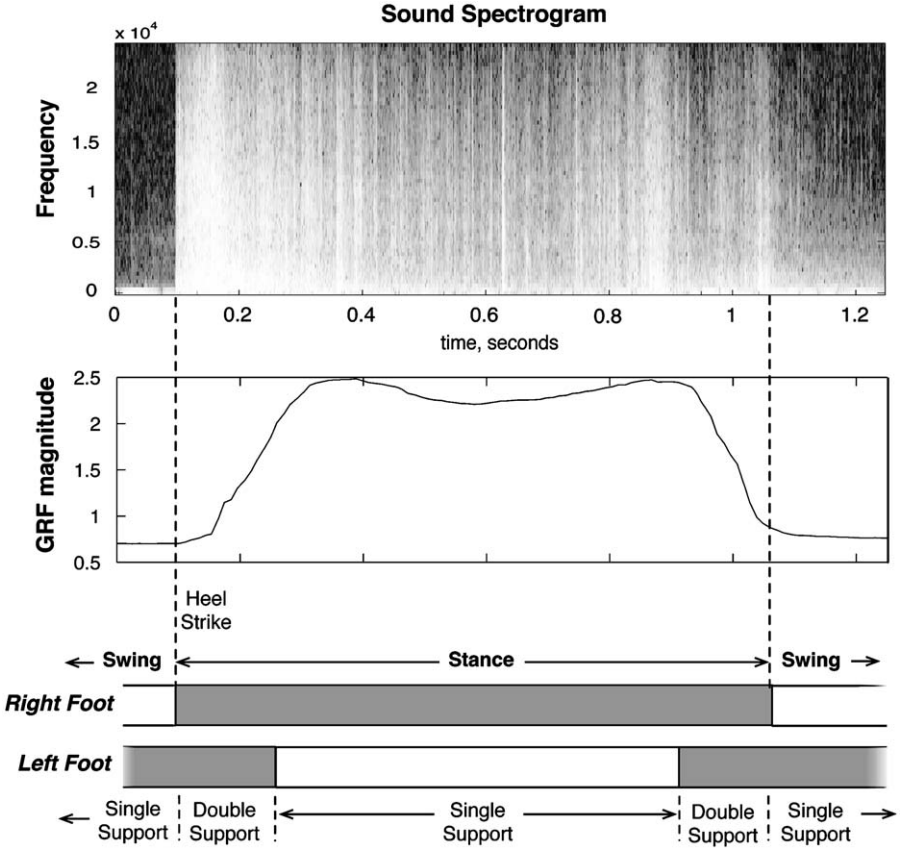


Fig. 1. (Left) Bottom: gait phase for left and right foot, showing more than 50% of the gait cycle. Middle: GRF for right foot (authors' recording). Top: spectrogram showing the acoustic signature resulting from a step onto gravel.

having components tangential and normal to the ground surface; ω denotes angular frequency and t is time. The term ground reaction force (GRF) is often used to refer to the low frequency information in **F**, below about 300 Hz. The GRF is essentially responsible for the center of mass movement of the individual. It is approximately independent of footwear type, but varies between individuals or walking styles (e.g., Galbraith and Barton, 1970). Higher-frequencies components of $\mathbf{F}(\omega, t)$ can be attributed to fast impacts between heel or toe and ground, sliding friction and contact variations between the shoe and ground (Ekimov and Sabatier, 2006). Unlike the GRF, these components can depend on footwear, on ground surface shape and material properties. They give rise to remote signatures in the form of airborne acoustic signals, seismic vibrations of the ground, and vibrations transmitted through the shoe to the foot, which have been studied in prior literature on acoustic (Li et al., 1991; Ekimov and Sabatier, 2006; Watters, 1965) and vibrational (Cress, 1978; Ekimov and Sabatier, 2008; Galbraith and Barton, 1970) signatures of human walking. These signals vary with the local material and spatial structure of the ground and with the temporal and spatial profile of interactions between the foot of the walker and the ground surface. Several phenomenological models for the contact interactions that produce them are reviewed in Section 3.3.

From a sensory standpoint, in addition to vision, the pedestrian receives sound information via the auditory channel, vibrational information via the tactile (touch) sensory receptors in the skin of the feet, and information about ground shape and compliance via the proprioceptive sense (the body's ability to sense the configuration of its limbs in space). Proprioception, vision, and the vestibular (balance) sense are integrated to inform the pedestrian about his motion in space.

1.2. Overview

As can be seen from the forgoing description, walking generates a great deal of multisensory information about the environment. Prior research has emphasized the influence of visual, haptic, vestibular, and proprioceptive information on control and planning of locomotion over predominantly flat surfaces (e.g., Wu and Chiang, 1996). In two respects, these studies provide a limited account of the complexity of walking in real world environments. Firstly, they have not addressed the range of ground surfaces and materials met outside the lab (e.g., to our knowledge, none has investigated locomotion on soil or gravel). Secondly, they ignore the information contained in sounds generated by walking on real world surfaces (e.g., acoustic information about the gender of a walker, Li et al., 1991). These limitations are addressed in human perception studies presented in Section 2. Notably, in VR contexts, when such layers of perceptual information are available, they are likely to contribute to a heightened sense of presence in the virtual environment, a subject addressed in Section 5.

The remainder of this paper describes developing research on the perception and design of non-visual signatures of walking. We focus on the simulation and perception of foot–ground contact interactions, conceived as carriers of information about ground surfaces and the walkers themselves. In the four sections that follow, we highlight research in these areas:

- The human perception of contact events, with an emphasis on walking sounds.
- Technologies for the interactive synthesis and display of virtual auditory and vibrotactile signatures of walking on natural materials.
- Efficient, physically based computational models for rendering such signals.
- The usability of such displays in human computer interaction and their impact on users' sense of presence in virtual environments.

Their diversity is suggestive of the interdisciplinary effort that is needed to inform future practice in the design of systems that make rich use of walking interactions.

2. Human perception

The information that reaches our senses is structured by the objects and events from which it originates. Probabilistic relationships between the properties of the objects and events in the environment on the one hand, and the structure of the sensory input on the other, are exploited by a perceiver to recover the properties of the surrounding environment. This function of perception is not limited to the visual system, but characterizes all of our senses. In the hearing domain, knowledge has recently accumulated on the perceptual ability to recover properties of the sound generating events in purely acoustical contexts (see Rosenblum, 2004; Lutfi, 2007, for recent reviews).

Locomotion usually produces audible sounds, comprising a number of qualitatively different acoustical events: isolated impulsive signals (e.g., from the impact of a hard heel onto marble); sliding sounds (e.g., a rubber sole sliding on parquet); crushing sounds (e.g., walking on snow); complex temporal patterns of overlapping impulsive signals (e.g., walking on gravel). Overall, the structure of such sounds is jointly determined by several properties of the source: the shape and material of the ground (e.g., brittle ice, gravel), the dynamical features of locomotion itself (e.g., speed, stability), the anthropometric and non-anthropometric properties of a walker (e.g., weight, legs length, but also and gender and emotion of a walker), and the properties of the foot surface in contact with the ground (e.g., size and hardness of the shoe sole). Walking thus conveys information about the properties of the sound source and, even in the absence of explicit training, listeners learn to recover properties of the walking event based on the features of the sound.

There are few published studies on the perceptual processing of walking sounds. Indeed, the major focus in the study of sound source perception has been on impact sounds, impulsive signals generated by a temporally limited interaction between two objects (e.g., mallet hitting a marimba bar). Nonetheless, this literature is relevant to understanding the hearing of walking sounds, for at least two reasons. Firstly, a walking sound is, more often than not, a sequence of isolated impact sounds, and similar strategies are likely applied to recover the properties of the sound source in both cases (e.g., interacting materials). Secondly, theoretical developments in the study of isolated impact sounds (e.g., hypotheses on the nature of inter-individual differences in source perception or on the factors determining the weighting of acoustical information) can, at least in principle, be extended to the perception of any natural sound-generating event.

In Section 2.1 we detail developments on the study of impact sounds. In Section 2.2 we present the literature on the perceptual processing of walking sound events.

2.1. Isolated impact sounds

The study of the perception of isolated impacts is the most developed area within the field of sound source perception (see Giordano et al., in press for a review of prior studies on impact sounds). Typically, research design in this field involves three stages (Li et al., 1991; Pastore et al., 2008). Firstly, the acoustical specification of the properties of the sound source is quantified (e.g., sound frequency is strongly dependent on the size of an object). At times, this analysis aims to quantify the perceptual performance of an ideal listener that perceives a source property through one or more sound features (Pastore et al., 2008; Giordano et al., in press). Secondly, perceptual data are modeled based on mechanical descriptors of the sound source (e.g., McAdams et al., 2004). This stage might consist in a quantification of the accuracy in the human perception of a target source property, or in the analysis of the statistical association between raw behavioral data and all of the manipulated source properties, independent of whether they are the target of perceptual judgment (e.g., material identification is strongly influenced by the size of an object Giordano and McAdams, 2006). Finally, behavioral data are modeled as a function of the sound features. This last modeling stage is of interest to the study of human processing of complex sounds, but also delivers to a sound designer important indications as to those properties of a sound necessary to deliver a perceptual effect.

In the literature on impact sounds, perception of the material of struck objects is linked with energy decay-related properties of the sound signals (e.g., velocity of the loudness decay Giordano and McAdams, 2006); perception of geometrical properties of struck objects is linked with the frequency of the spectral components (e.g., ratios of the frequency of specific spectral components, Lakatos et al., 1997); perception of the materials of impacting objects is linked with the spectral properties of the early portions of the sounds (e.g., impulsive signals with a high spectral center of gravity are perceived as generated with a hard hammer Giordano et al., in press).

Three recent developments in sound source perception aim at more than quantifying recognition performance and the mechanical and acoustical correlates of perception. Lutfi et al. (2005) investigated the extent to which real sounds can be accurately represented with simplified modal synthesis signals. Experiments compared real and synthetic signals in discrimination and source identification tasks, and investigated discrimination of signals synthesized with a variable number of resonant modes. Results indicate that simplified synthetic sounds, based on a small number of free parameters, yield similar perceptions as their real counterparts, and are frequently indistinguishable from them. Lutfi and Liu (2007) investigated the interindividual variability of the perceptual weight of acoustical information (e.g., the extent to which the frequency of the lowest spectral components affects the perceptual responses). They find that the across-tasks variation of perceptual weights (e.g., the extent to which the perceptual weight of the lowest frequency differs between across the identification of mallet hardness vs. material) is smaller than the across-listeners variation of perceptual weights. They take this result as evidence that participants adopt personalized styles in the weighting of acoustical information, independent of the particular task. They further show that similar performance levels can arise from widely different strategies in the weighting of acoustical information, and that interindividual differences in performance are strongly affected by internal noise factors rather than changes in weighting strategies. Finally, focusing on the estimation of the hardness of impacted objects, Giordano et al. (in press) investigated the influence of the accuracy and exploitability of acoustical information on its perceptual weighting. Studies of source perception reveal that listeners integrate information over both accurate and inaccurate acoustical features, and do not focus selectively onto the most accurate specifiers of a sound source property. It is thus hypothesized that the perceptual weight of an acoustical feature increases with its accuracy and decreases with its perceptual exploitability, as defined by feature-specific discrimination, memory and learning abilities. Both factors appear to interact in determining the weighting of acoustical information. In general, information is weighted in proportion to its accuracy, both in the presence and absence of feedback on response correctness. However, in the absence of feedback the most accurate information can become perceptually secondary, thus signaling limited exploitation abilities.

2.2. Acoustic and multimodal walking events

The study of the human perception of locomotion sounds has addressed several properties of the walking

sound source: the gender (Li et al., 1991; Giordano and Bresin, 2006) and posture of a walker (Pastore et al., 2008), the emotions of a walker (Giordano and Bresin, 2006), the hardness and size of the shoe sole (Giordano and Bresin, 2006), and the ground material (Giordano et al., 2008).

Li et al. (1991) investigated the perception of walkers' gender in untrained listeners. High identification performances were observed. Gender identification appeared related to shoe size, although this factor did not fully account for gender perception. From the acoustical standpoint, gender perception appeared related to the spectral properties of the footstep sounds: females (respectively males) were recognized by shallow (respectively sharp) spectra with a dominant high-frequency (low-frequency) component.

Giordano and Bresin (2006) asked untrained listeners to estimate several properties of the walking sound source: the gender and emotion of a walker (anger, happiness, sadness and fear), and the size and hardness of the shoe soles. The majority of participants recognized each of these attributes at higher-than-chance levels. Interestingly, recognition of gender, sole hardness and size, parameters strongly correlated with each other (female walkers wore smaller shoes with harder soles), was more accurate than the recognition of emotions. Consistent with the results of Li et al. (1991), estimation of gender, and of sole size and hardness, was based on spectral information (again, females were recognized in spectra with a predominant high-frequency component). Perception of emotions was instead strongly influenced by energetic and temporal features: the average pace and pace irregularity, and sound intensity.

Pastore et al. (2008) investigated the discrimination of upright and stooped walking postures in trained listeners. They analyzed the relationship between the mechanics, acoustics and perception of sound events, using the approach described in Section 2.1. The study of the source–acoustics relationship focuses on quantifying the posture discrimination performance afforded by a perceptual focus onto either isolated sound features or onto pairs of sound features (see Giordano et al., in press for a similar analysis conducted with impacted sound sources). They develop a hierarchical model of perceptual decision, based on pairs of sound descriptors. An ideal observer is assumed to be faced with the task of identifying which of two sounds is produced with an upright posture. This observer first considers the difference in the value of an acoustical descriptor between the two sound stimuli. If this difference exceeds a fixed threshold, a response is given. If not, the response is not guessed at random, but is based onto the computation of the difference between the two sound stimuli with respect to a second descriptor. Following this approach in the modeling of a simulated, ideal observer, recognition performance was maximized with pace as the first feature, and spectral amplitude of the heel impact in the 100–500 Hz range as the second.

Giordano et al. (2008) analyzed unimodal and multisensory non-visual identification of two classes of ground materials: solids (e.g., wood) and aggregates (gravels of different sizes). In the multisensory condition, participants walked blindfolded onto ground samples. In the vibrotactile condition, they were also presented with an acoustical masker over wireless headphones. In a proprioception condition, they were presented both the acoustical masker and a tactile masker, delivered through vibrotactile actuators mounted at the bottom of the shoe sole. In the auditory condition, participants did not walk on the materials, but were presented with their own footstep sounds. Overall, identification performance was at chance level only for solid materials in the proprioception condition: absent both auditory and vibrotactile information, solid materials could not be identified. The availability of all sources of non-visual information led to a small but consistent improvement in discrimination performance for solid materials. With aggregates, identification performance was best in the vibrotactile condition, and worst in the auditory condition. Discrimination in the multisensory condition was impaired, compared to that observed in the vibrotactile condition. Limited to the aggregate materials investigated, this result was interpreted as indicating the multisensory integration of incoherent information: auditory on the one hand, vibrotactile and proprioceptive on the other.

3. Augmented ground surfaces as walking interfaces

As noted in the preceding section, the identity of a natural ground surface that is walked upon may be communicated through several different non-visual sensory channels, including auditory, tactile, and proprioceptive channels. Furthermore, material identity may be preserved even when some modalities are absent. Consequently, one way to view the problem of designing an interface for non-visual walking signatures is as a tradeoff between the number of modalities addressed and fidelity at which they can be reproduced, versus the cost and effort required to do so.

One category of applications for the display of non-visual walking signatures in HCI aims to enable walking as a means of controlling self-motion in an immersive virtual environment (VE). In such a context, the convincing representation of ground surface properties is desirable, toward improving a user's sense of presence in the virtual environment (see Section 5). Another category can be identified with systems that utilize walking as a means of interaction with non-immersive systems. For example, such an interface may be designed to enable the use of walking sounds to provide navigational cues (as in the example in the Introduction), or to generate multimedia content (e.g., the PholieMat, described below).

Although somewhat orthogonal to the main content of this paper, we note that considerable research has been undertaken on robotic interfaces for walking in virtual

environments. This subject was recently reviewed by Iwata (2008) and Hollerbach (2008). The devices concerned consist of force-feedback interfaces that, when combined with a virtual environment (VE) simulator, provide the illusion that one is walking in a VE, when one is, in fact, staying in place. One type of configuration for such a device involves an omnidirectional treadmill interface, consisting of a belt that moves under the feet in such a way that the walker remains in place as one walks. Another consists of a pair of platforms attached to the feet and connected to a robotic mechanism capable of delivering forces to the feet. Although such devices are able to approximate the kinesthetic experience of walking (i.e., the sensory experience of bodily motion), it is important to note that they involve an intrinsic cue conflict between the inertial (vestibular) sensory capacities of the body and the visual and kinesthetic cues supplied by the device. Moreover, such devices do not attempt to represent the high frequency tactile or acoustic properties of a surface being walked upon.

The latter properties are the focus of the types of display described here. They consist of walking surfaces augmented with synthetic auditory and/or vibrotactile signals simulating these components of the response of real ground materials (Visell et al., 2007, 2008; Nordahl, 2006). Such devices (e.g., Fig. 5) attempt to compensate for the feedback channels they cannot display—specifically, the felt compliance and shape of the ground—by providing approximately realistic tactile and auditory feedback that is closely coordinated with the footsteps of their users. As indicated in examples presented in Section 5, in a virtual environment context, coordinated visual feedback via wall and/or floor surfaces can also be supplied. For the moment, we concentrate on the interactive auditory and tactile display elements.

A reasonable person may question whether the experience of walking on complex ground materials of varying material, such as marble, earth or snow, can possibly be simulated by a flat ground surface. However, the results of Giordano et al. related above (Giordano et al., 2008) indicate that for solid ground surfaces, vibrotactile and auditory signals are likely more important as carriers of information about material identity than proprioceptive information is. While proprioceptive information is very relevant for the identification of highly compliant materials, the same study suggests that the identity of such materials may be preserved to an acceptable level of accuracy without it. However, further research is needed on the effectiveness of such synthetic information channels at communicating ground properties.

3.1. Physical interaction design

The main components to be specified in the design of an augmented walking surface include the physical embodiment of the device, the sensors and actuators to be employed, and associated electronics for signal acquisition, amplification, and conditioning.

Two basic physical configurations of an augmented walking device can be envisaged (Fig. 2). The reader can undoubtedly envision a number of other possibilities combining these scenarios. The first type consists of a rigid surface instrumented with sensors and actuators. Users are allowed to walk on it wearing ordinary shoes. Such a surface might consist of a flat floor or an isolated surface, such as a stair step. The second type involves a shoe instrumented with sensors integrated in the sole or insole. Portable acoustic actuation can be supplied by a wearable 3D (binaural) auditory display or by wearable loudspeakers. Vibrotactile actuation can be accomplished with actuators integrated within a shoe sole or insole. To date, there has been limited research on such footwear (e.g., the stimulus masking shoes of Giordano et al., 2008), but the technologies involved lie within reach of the state of the art (Hayward and Maclean, 2007). Footwear type and material are relevant in both cases, because natural walking sounds depend on properties of both the shoe and ground (e.g., Li et al., 1991). However, such factors may be best considered in a case-based discussion, as the extent to which user footwear may be known or controlled likely depends upon the application scenario (e.g., virtual environment display vs. augmented reality display in a public space).

The most direct method of sensing involves the acquisition of foot–floor forces or contact regions. Other techniques involve the capture of secondary signatures of foot–ground contact, such as accelerations in a shoe sole or floor surface. A wide range of sensing technologies may be suitable. Examples that have been used for capturing foot–floor contacts include: force-sensing resistive materials (paper, rubber, or other substrates), composite structures of the same type (commercial force sensing resistors), strain gauge based load cells, piezoelectric elements, weaves or films, capacitive elements or films, MEMS accelerometers or gyrometers, and optical fiber composites.

As noted above, auditory display is readily accomplished using standard loudspeakers or head mounted auditory displays. Vibrotactile display, if less common, poses broadly similar requirements. It demands actuators with a frequency response overlapping most of the range from 20 to 1000 Hz (the approximate frequency band of greatest acuity of the human vibrotactile sense, Jones and Sarter, 2008). Moreover, a suitable mechanical design of the

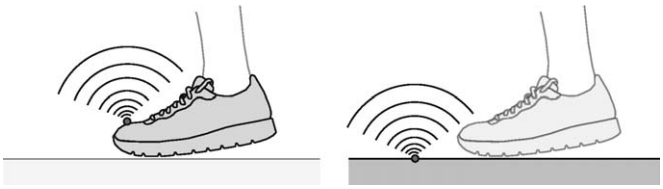


Fig. 2. A walking surface augmented using an instrumented shoe (left) or with an instrumented floor (right).

actuated surface and its structural isolation from the ground is needed to ensure good fidelity and power efficiency. A practical benefit of vibrotactile actuation is that the power requirements are much lower than for a kinesthetic display of comparable scale, in which large forces must be exerted at low frequencies. Among available actuator technologies, linear voice coil actuators, which consist of a magnetic inertial slug suspended on a membrane between a set of electromagnetic coils, are inexpensive, and can be made compact and powerful. Crucially, they permit independent control over stimulus amplitude and waveform. More detailed discussion of tactile actuator types can be found in recent literature (Hayward and Maclean, 2007; Visell et al., 2009).

The spatial distribution of the active display components is another salient factor. If a step is taken on any ground material, contact interactions occur at many sites along the shoe sole. This suggests a high spatial density of sensors and actuators may be required. However, limitations in spatial resolution of the display may be compensated if the interface is designed in such a way that different areas of the foot receive feedback in proportion to the force they are exerting against the tile. This is the case, for example, if the foot receives tactile feedback from a rigid floor surface in response to the force applied to that surface. The proportion may be interpreted as a measure of the responsibility of a given area of the foot for generating the feedback in question.

Commensurate with the coarse spatial resolution of the display, as noted below, for synthesis, a lumped interaction model is frequently adopted, in which the interaction is viewed as taking place through time without spatially distributed degrees of freedom. In such a case, all that may be required is a measurement of the net force applied by each foot at each instant in time. This can be accomplished with a network of sensors with a linear spatial resolution of approximately 30 cm, sufficient to distinguish the net force exerted by each foot.

3.2. Control design

The active components of the display consist of force sensors, actuators and drive electronics, and a computer running the control and sound and/or vibrotactile synthesis algorithm. The control mapping permits user actions captured through the device to determine the synthesis of sounds and/or vibrations.

A simplifying model regards the control mapping as an open loop (Fig. 3), to be calculated independently from the resulting output signals. Such an approximation is tantamount to the segregation of low-frequency input forces (generated by movements of the walker's lower appendages) from higher frequency acoustic and vibrotactile outputs (generated by material interaction with the ground). As described in the Introduction (see Fig. 1), such a separation is supported by prior literature characterizing the information content in comparable signals during walking on real materials (Ekimov and Sabatier, 2006, 2008).

3.3. Sound synthesis

Acoustic and vibrational signatures of locomotion are the result of more elementary physical interactions, including impacts, friction, or fracture events, between objects with certain material properties (hardness, density, etc.) and shapes. The decomposition of complex everyday sound phenomena in terms of more elementary ones has been an organizing idea in auditory display research during recent decades (Gaver, 1993). For present purposes, it is useful to draw a primary distinction between solid and aggregate ground surfaces, the latter being assumed to possess a granular structure, such as that of gravel.

A comprehensive phenomenology of footstep sounds accounting for diverse walking situations should consider various factors, including those described in Section 2.2. Ideally, a designer should have access to a sonic palette making it possible to manipulate all such parameters, including material properties, gestural, and emotional nuances of gait. While this is not yet possible, as reviewed below, there is much prior work on the synthesis of the sounds of contacting objects, including walking settings. Additionally, Section 4 reviews prior work on the control of walking sounds with emotional and gender-based parameters.

3.3.1. Solid surfaces

Sonic interactions between solid surfaces have been extensively investigated, and results are available which describe the relationship between physical and perceptual parameters of objects in contact (Klatzky et al., 2000; van den Doel et al., 2001). Such sounds are typically short in duration, with a sharp temporal onset and relatively rapid decay.

A common approach to synthesizing such sounds is based on a lumped source-filter model, in which an impulsive excitation $s(t)$, modeling the physics of contact, is passed through a linear filter $h(t)$, modeling the response

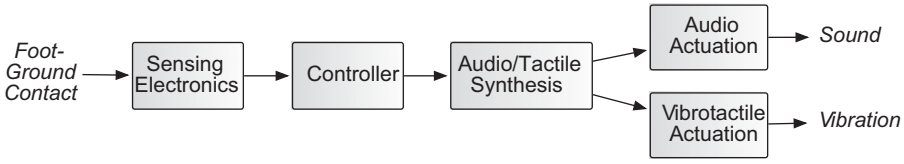


Fig. 3. Footstep interaction viewed as an open-loop process.

of the vibrating object as $y(t) = s(t) \star h(t)$, where $\star$ denotes convolution in time. Modal synthesis (Adrien, 1991) is one widely adopted implementation of this idea. It decomposes of the response model $h(t)$ in terms of the resonant frequencies f_i of the vibrating object (the modes). The response is modeled as a bank of filters with impulse response $h(t) = \sum_i a_i e^{-b_i t} \sin(2\pi f_i t)$, determined by a set of amplitudes a_i, decay rates b_i, and frequencies f_i.

Impacts between shoe and ground (for example, those occurring at heel strike and toe slap) provide the excitation source, while the resonator encompasses either or both of the floor surface itself and the shoe sole. The excitation corresponding to each impact $s(t)$ is assumed to possess a short temporal extent and an unbiased frequency response. In the simplest case, it can be taken to be a known, impulsive signal with total energy E. In a more refined approach, it may consist of a discrete-time model of the force between the two bodies, dependent on additional parameters governing the elasticity of the materials, their velocity of impact, and masses. The parameters governing such solid interactions can be used to specify the characteristics of each impact event, encoding the materials and other interaction parameters, for synthesis using existing models (Avanzini and Rocchesso, 2001).

3.3.2. Aggregate surfaces

The approach outlined above is not directly applicable to cases in which the ground surface does not consist of a solid body. Instead, footsteps onto aggregate ground materials, such as sand, snow, or ice fragments, belie a common temporal process originating with the transition toward a minimum-energy configuration of an ensemble of microscopic systems, by way of a sequence of transient events. The latter are characterized by energies and transition times that depend on the characteristics of the system and the amount of power it absorbs while changing configuration. They dynamically capture macroscopic information about the resulting composite system through time (Fontana and Bresin, 2003).

Physics provides a general formalization of such sounds in terms of: (i) the probabilistic distribution of the energies E of the short transients, which can be assumed to follow a power law $p(E) \propto E^{\gamma}$. The value of γ determines the type of noise produced by the process (for instance, in the case of crumpling paper it is $-1.6 < \gamma < -1.3$) (Sethna and Dahmen, 2001) and (ii) a model of the temporal density $N(t)$ of transients as a stationary Poisson process, under the assumption that the inter-transient event times τ are assumed to be independent (Papoulis, 1984): $P(\tau) = \lambda e^{\lambda \tau}$.

The parameters γ and λ together determine the macroscopic process dynamics. A simple view of this process is that each transient event consists of a microscopic solid impact with energy E. Thus, in addition, an individual transient can be assumed to possess a resonant response $h(t)$, which is specified in the same way as described above. The resulting parameters characterize each transient event independently of the evolution of the macroscopic system.

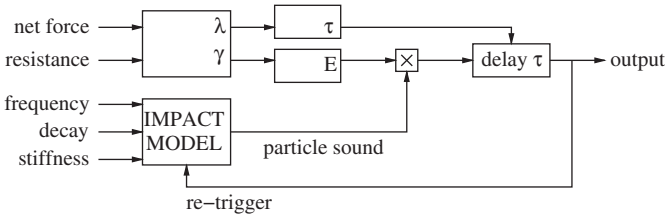


Fig. 4. Algorithm for the generation of sounds of contact with aggregate materials.

Taken together, intensity, arrival times, and impact parameters form a powerful set of independent parametric controls capable of rendering both the process dynamics, which is related to the temporal granularity of the interaction (and linked to the size of the foot, the walking speed, and the walker's weight), and the type of material the aggregate surface is made of (Fig. 4). Such controls enable the sound designer to choose foot–ground contact sounds from a particularly rich physically informed palette. Several models of this general type have been developed in order to mimic the sound of a footstep onto aggregate grounds (Cook, 2002; Fontana and Bresin, 2003; O'Modhrain and Essl, 2004). Section 3.5 reviews one case study in detail.

3.4. Augmented ground surfaces developed to date

The hardware technologies described above are well within the state of the art. As a result, a number of different augmented floor interfaces have been developed, with the largest application domains comprising artistic creation and entertainment. A comparative review of several floor interfaces that were developed for use in musical control was provided by Miranda and Wanderley (2006). Even more attention has been devoted to the development of distributed, sensing floor surfaces, without the explicit intent of generating sound, aided, in part, by the commercially availability of the necessary sensing technologies.[1]

A smaller number of devices have sought to re-create the experience of walking on virtual ground materials. Closest in spirit to the present contribution, Cook (2002) consists of a force-sensing floor mat used as a controller for the real-time synthesis of footstep sounds generated by walking on different ground surfaces. Nordahl (2006) investigated the integration within a VE of self-generated footstep sounds controlled by a set of instrumented sandals (reviewed in Section 5).

3.5. Example: Eco Tile

The Eco Tile, a floor component aimed at the interactive simulation of natural ground materials, is unique in its integration of force sensing in addition to acoustic and vibrotactile actuation (Visell et al., 2007, 2008). The

[1]Available from, e.g., TekScan, Inc.

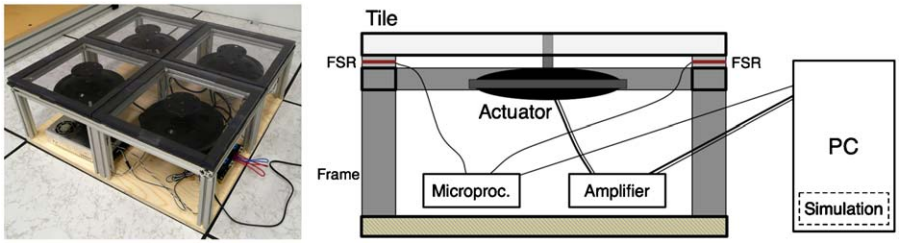


Fig. 5. Left: an image of the tile prototype, showing the tile surface (polycarbonate sheet), vibrotactile actuator, force-sensing resistors, structural frame, and associated electronics. Right: diagram of the same, including the PC running the floor material simulation.

prototype shown (Fig. 5) consists of a set of rigid $34 \times 34 \times 0.5$ cm polycarbonate tiles supported by a common structural frame. A linear voice coil actuator (Clark Synthesis model TST239), capable of driving the display over the amplitude and frequency ranges of interest, is rigidly attached to the underside of each tile. Auditory stimuli may be generated in two different ways. If the top of the tile is left untreated, it produces auditory feedback of usable quality as a byproduct of the vibration of the tile surface itself. Alternatively, a separate auditory display may be used. Force sensors are positioned beneath the four corners of each tile, and a single actuator is used to drive the tile surface.

This device has been used to provide the simulation of stepping onto an aggregate ground surface, whose response is synthesized in the manner described in Section 3.3, and controlled by driven by data captured from its sensors as we describe here. Consider a single tile. The vector of four force signals $\mathbf{f}(t)$ from its sensors are used to control the synthesis process. In this case, the distribution of impact events is modeled as a non-homogeneous Poisson random process with a rate parameter $\lambda(t)$ given by

$$\lambda(t) = Au(t)(1 + \tanh(Bu))/2, \tag{1}$$

$$u(t) = df_L(t)/dt, \quad f_L = \|\mathbf{f}_L\|. \tag{2}$$

Here A is a control gain parameter, $\mathbf{f}_L(t)$ are components of $\mathbf{f}(t)$ below about 300 Hz, and $(1 + \tanh(Bu))/2$ approximates a Heaviside step function when $B \gg 1$. This simple, force-derivative control scheme guarantees that a response is obtained primarily when the foot is coming down onto the tile, and the force exerted on the tile is increasing (Fig. 6). The total acoustic energy that can be generated by a single footstep is assumed to be a constant[2] value, $\mathscr{E}$. The amount E_i that is attributed to the ith impact event is determined by sampling from an exponential distribution $p(E) \propto E^{\gamma}$ with free parameter γ, ensuring that $\sum_k E_i = \mathscr{E}$ is satisfied.

Each virtual impact involves an inertial object striking a resonant object with the requisite energy. The force of impact $y(t)$ is determined by a simplified phenomenological equation known as the Hunt and Crossley (1975) model

$$y(t) = kx(t)^{\alpha} - \lambda x(t)^{\alpha}\dot{x}(t). \tag{3}$$

[2]For example, $\mathscr{E}$ may be considered to be a constant fraction of the potential energy difference of the body between mid-swing and stance.

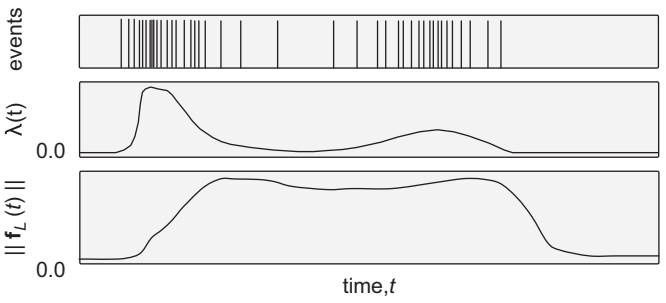


Fig. 6. Qualitative illustration of the control model of Eq. (2), relating the time derivative of the low frequency force signals $d\mathbf{f}_L/dt(t)$ (Bottom) to the event rate parameter $\lambda(t)$ (Middle) and a sampled event sequence (Top).

Here, $x(t)$ is the compression displacement and $\dot{x}(t)$ is the compression velocity. The impact force has parameters governing stiffness k, dissipation λ, and contact shape α. This force is coupled to a modal synthesis representation of the resonant object having the same structure as described above. An impact event is synthesized by initializing Eq. (3) with the velocity v_I of impact and integrating the composite system in time. See Rocchesso and Fontana (2003) for a more detailed discussion. Values for several of the synthesis and control parameters are obtained by measurement and analysis of measured responses of footsteps onto real granular materials (Visell et al., 2008).

In summary, as discussed at the beginning of this section, floor interfaces like the Eco Tile depend for their success on their ability to sustain two distinct illusions: First, that the foot is in contact with a compliant and/or composite material of definite properties that are distinct from those of the floor tile itself; Second, that the virtual physical interaction is distributed across the ground under the foot, rather than originating in a vibration of the ground surface that is (piecewise) constant across the latter.

4. Affective footstep sounds

In this section we present the main results of a recent study in which a model for the synthesis of natural footstep sounds was developed (DeWitt and Bresin, 2007), and preliminarily assessed. The starting point was the model of natural walking and running footstep sounds on aggregate materials that presented in Section 3.3. The pace of footsteps was controlled by tempo curves which were derived from studies in music performance, since strong

similarities between locomotion and music performance were found in prior research. A listening test for the validation of that model highlighted the way in which footstep sequences that were generated using expressive tempo curves, derived from music performance, were perceived as more natural by listeners compared to sequences having a constant pace. Using this study as starting point, we have developed a model of footstep sounds for simulating the presence of people walking in a virtual environment. The design choice was that the footstep sounds should communicate the gender, age, weight, and emotional intention of a virtual walker.

The sound synthesis model was tuned by ear to simulate different ground materials. Gravel, dirt, soft wood, snow, and grass-like settings were selected using the parameters λ and γ; in parallel, the impact parameters were set to reproduce rubber, glass, steel, and wood. The timing in footstep sequences was controlled by using a footstep tempo model developed after measurements of real walkers, who were asked to walk with emotional intentions (happiness, sadness, fear and anger), as well as with their natural (neutral) pace. In interactive listening tests, subjects could adjust pace and material to determine the gender of a virtual walker.

Results show that subjects associated both different pace and material to the two genders (Fig. 7). Female walkers were identified by faster pace (the time interval between to footsteps was about 0.5 s for females and 0.8 s for males), higher resonant frequency for impacts (glass and steel sounds for female; rubber and wood sounds for males) and for particle sounds (mainly gravel and snow sounds for females; dirt and soft wood sounds for males).

It was also tested how subjects would change the emotion of footstep sound sequences. Subjects could control the sound in a 2D activity-valence space in which pace characteristics (regularity and timing) were changed dynamically. Results are promising despite the existence of some confusion between angry and happy footstep sounds. This confusion could be overcome by improving the continuous control over the real-time change of the acoustical characteristics of the ground, thus allowing for a gradually changing perception of both the gender and emotional intention of a virtual walker.

5. VR applications and presence studies

Prior research has addressed issues related to the addition of auditory cues in virtual environments, and whether such cues may lead to a measurable enhancement of immersion in such environments. Most prior work in this area has focused on sound delivery methods (Storms and Zyda, 2000; Sanders and Scorgie, 2002), sound quantity and quality of auditory versus visual information (Chueng and Marsden, 2002) and 3D sound (Freeman and Lessiter, 2001; Vastfjall, 2003). Recent studies have investigated the role of auditory cues in enhancing self-motion and presence in VEs (Larsson et al., 2004; Kapralos et al., 2004; Väljamäe et al., 2005).

Self-generated sounds have been often used as enhancements to VEs and first-person 3D computer games—particularly in the form of footstep sounds accompanying self-motion or the presence of other virtual humans. A smaller number of examples, such as recent work of Law et al. (2008), have even aimed to provide multimodal cues linked to footstep events in such environments (Fig. 8). However, to our knowledge, the effect of such self-generated sounds on users' sense of presence had not been investigated prior to the authors' research in this area. The combination of physics-based rendering of walking sounds with contact-based sensing, as described in the preceding section, also appears to be novel.

5.1. Auditory feedback and motion

The algorithms described in Section 3.3 provided a basis for an evaluation carried out by the authors Nordahl (2006) on the role of interactive self-generated auditory feedback in virtual environments. The visual environment was reproduced using image based rendering techniques capturing part of the botanical garden in Prague. Physically modeled footstep sounds were controlled in real-time via a custom pair of sandals, enhanced with force sensors, which were worn by the user of the environment. The interest of this study was to understand to what extent the quality of auditory feedback would affect users' behavior in such a virtual reality system, and in particular, how and to what extent such interactive auditory feedback

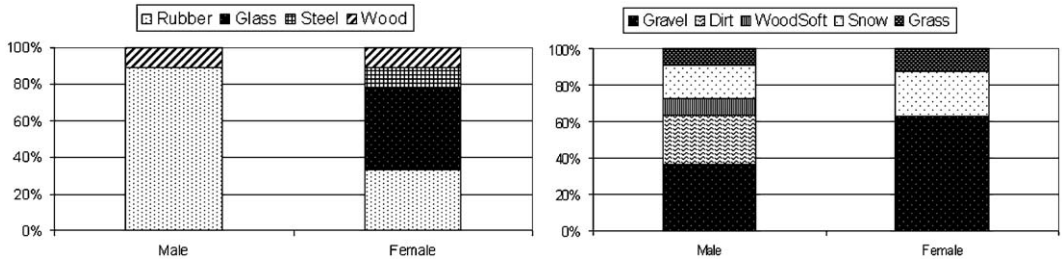


Fig. 7. Subjects' percentage choices of different ground materials in association to walker's gender. The left figure shows the choices for impact sound models. The right figure shows subjects' preferences for different tunings of the crumpling sound model.

Fig. 8. Depiction of the multimodal VE developed by Law et al. (2008), incorporating a CAVE-like visual VE with an auditory and haptic floor display based on the Eco Tile. Footstep events in an immersive snowy landscape are accompanied by visual, auditory, and haptic feedback.

might enhance the motion and presence of subjects in a VE. Prior work on environments simulated using image based rendering techniques has shown that subjects do not find the environments engaging, because of their lack of a dynamic temporal dimension (Turner et al., 2003). The authors were motivated by the belief that interactive auditory feedback can address such limitations.

This hypothesis was tested in an experiment with 126 subjects. Before entering the room, subjects were asked to wear a head mounted display and the instrumented sandals. Subjects were not informed about the purpose of the sensor-equipped footwear. Before beginning the experimental session, the subjects were told that they would enter a photo-realistic environment, where they could move around if they so wished. Furthermore, they were told that afterward they would be asked to fill out a questionnaire with several questions focused on what they remembered having experienced. No further guidance was given.

The experiment was performed as a between-subjects study including the following six conditions:

(1) Visual only. This condition had only uni-modal (visual) input.
(2) Visual with footstep sounds. In this condition, subjects had bi-modal perceptual input including auditory feedback with non-self-generated environmental sounds (audio-visual), comparable to earlier research (Nordahl, 2005).
(3) Visual with full sound. In this condition implies subjects were provided with environmental sounds, spatialized footstep sounds (using the VBAP algorithm) as well as rendering sounds from ego-motion (the subjects triggered sounds via their own footsteps).
(4) Visual with full sequenced sound. This condition was strongly related to condition 3. However, it was run in three stages: the condition started with bi-modal perceptual input (audio-visual) with static sound design. After 20 s, the rendering of the sounds from emotion was introduced. After 40 s the 3D sound started (in this case the sound of a mosquito, followed by other environmental sounds).
(5) Visual with sound + 3D sound. This condition introduced bi-modal (audio-visual) stimuli to the subjects in the form of static sound design and the inclusion of 3D sound (the VBAP algorithm using the sound of a mosquito as sound source). In this condition no rendering of ego-motion was conducted.
(6) Visual with music. In this condition the subjects were introduced to bi-modal stimuli (audio and visual) with the sound being a piece of music[3] described before. This condition was used as a control condition, to ascertain that it was not sound in general that may influence the in- or decreases in motion.

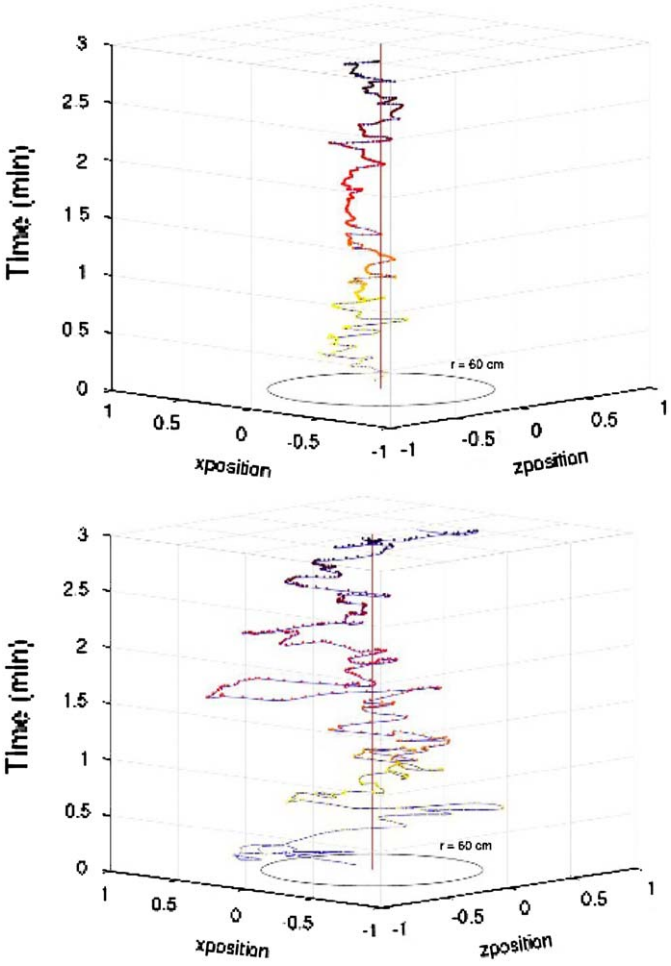


Fig. 9. Top: visualization over time of the motion of one subject over time with visual only condition (top) and full condition (bottom).

The results provided clear indications that footsteps sounds, when combined with environmental sounds,

[3]Mozart, Wolfgang Amadeus, Piano Quintet in E flat, K. 452, 1. Largo Allegro Moderato, Philips Digitals Classics, 446 236-2, 1987.

significantly enhance the motion of subjects in such a VE. The quantity of motion is clearly visible in Fig. 9, which shows subject position in the 2D plane, as acquired from a Polhemus magnetic tracker placed on the top of the head, respectively for one subject with *Visual only* stimuli (top) and with *Full condition* (bottom). The increase of movement exhibited by this subject in the *Full*condition is clearly noticeable. Results also indicated that footsteps sounds alone do not appear to cause a significant enhancement in subjects' motion. When comparing the results of the conditions *Visual only* versus *Visuals w. footsteps* (no significant difference) and the conditions *Full* versus *Sound+3D* (significant difference) there is an indication that the sound of footsteps benefits from the addition of environmental sounds. This result suggests that environmental sounds are implicitly necessary in a VE, and we assume that their inclusion is important to facilitate motion. Further detail is provided in the indicated references.

6. Conclusions

The interactive simulation of acoustic contact signatures generated by walking, which are highly salient to the experience of locomotion in diverse everyday environments, requires solving a number of design, engineering, and evaluation problems that make the realization of such interfaces a complex and multidisciplinary task. To effectively design the feedback channels involved, a solid base of knowledge on the perception of sound and vibrotactile information in walking events is needed, building on those studies discussed in this article. Conversely, we expect such knowledge to be further developed through experiments conducted in both real environments, and through virtual environments utilizing the current state of the art in acoustic and haptic display. The technologies, algorithms and methods for multimodal simulation and evaluation reviewed here are already capable of contributing to this process, but each can be further improved upon. For example, measurements relating the high frequency acoustic response of different ground surfaces to low frequency gait profiles (GRFs) would allow to refine the acoustic rendering techniques described in Section 3.3. Joint measurement of such attributes has only recently been broached in the literature (Visell et al., 2008). Control and rendering models can be refined to match the limitations of display devices and the perceptual capacities of their users, with the aim of compensating, as far as possible, missing sensory channels, such as proprioception. On the side of material attributes, a unification of rendering algorithms might be achieved by more carefully modeling of the physics of interaction with the materials involved, whether in a deterministic or stochastic setting. Such techniques have been successfully used in prior literature on everyday sound synthesis (Rath and Rocchesso, 2005). Open problems such as these can be expected to sustain the vitality of research in this emerging field for many years to come.

Acknowledgments

The authors gratefully acknowledge the support from the following sources: The European Commission for projects CLOSED (FP6-NEST-Path no. 029085), BrainTuning (FP6-NEST-Path no. 028570), and NIW (FP7-FET-Open no. 222107); The Quebec MDEIE for the project NIW; The ESF COST office for Actions no. 287 on Gesture Controlled Audio Systems (ConGAS) and no. IC0601 on Sonic Interaction Design (SID).

References

Adrien, J., 1991. The missing link: modal synthesis. In: Representations of Musical Signals. MIT Press, Cambridge, MA, USA, pp. 269–298.

Avanzini, F., Rocchesso, D., 2001. Modeling collision sounds: non-linear contact force. In: Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX-01), pp. 61–66.

Chion, M., 1994. Audio-Vision. Columbia University Press, New York, USA.

Chueng, P., Marsden, P., 2002. Designing auditory spaces to support sense of place: the role of expectation. In: Proceedings of the CSCW Workshop: The Role of Place in Shaping Virtual Community.

Cook, P.R., 2002. Modeling bill's gait: analysis and parametric synthesis of walking sounds. In: Proceedings of the Audio Engineering Society 22 Conference on Virtual, Synthetic and Entertainment Audio, AES, Espoo, Finland.

Cress, D.H., 1978. Terrain considerations and data base development for the design and testing of devices to detect intruder-induced ground motion. Technical Report M-78-1, U.S. Army Engineer Waterways Experiment Station, Vicksburg, Miss.

DeWitt, A., Bresin, R., 2007. Sound design for affective interaction. In: Affective Computing and Intelligent Interaction ACII 2007, pp. 523–533.

Ekimov, A., Sabatier, J., 2006. Vibration and sound signatures of human footsteps in buildings. Journal of the Acoustical Society of America 120 (2), 762–768.

Ekimov, A., Sabatier, J., 2008. A review of human signatures in urban environments using acoustic and seismic methods. In: Proceedings of IEEE Technologies for Homeland Security.

Fontana, F., Bresin, R., 2003. Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. In: Proceedings of the Colloquium on Musical Informatics, Florence, Italy, pp. 109–114.

Freeman, J., Lessiter, J., 2001. Hear, there and everywhere: the effects of multichannel audio on presence. In: Proceedings of ICAD 2001, pp. 231–234.

Froehlich, P., Simon, R., Baille, L., 2009. Mobile spatial interaction special issue. Personal and Ubiquitous Computing, May.

Galbraith, F., Barton, M., 1970. Ground loading from footsteps. Journal of the Acoustical Society of America 48 (5B).

Gaver, W.W., 1993. What in the world do we hear? An ecological approach to auditory source perception. Ecological Psychology 5 (1).

Gaye, L., Holmquist, L.E., Behrendt, F, Tanaka, A., 2006. Mobile music technology: Report on an emerging community. In: Proceedings of NIME'06.

Giordano, B.L., Bresin, R., 2006. Walking and playing: What's the origin of emotional expressiveness in music? In: Baroni, M., Addessi, A.R., Caterina, R., Costa, M. (Eds.), Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC9), Bologna, Italy.

Giordano, B.L., McAdams, S., 2006. Material identification of real impact sounds: effects of size variation in steel, glass, wood and plexiglass plates. Journal of the Acoustical Society of America 119 (2), 1171–1181.

Giordano, B.L., McAdams, S., Visell, Y., Cooperstock, J., Yao, H., Hayward, V., 2008. Non-visual identification of walking grounds. Journal of the Acoustical Society of America 123 (5), 3412.

Giordano, B.L., Rocchesso, D., McAdams, S., Integration of acoustical information in the perception of impacted sound sources: the role of information accuracy and exploitability. Journal of Experimental Psychology: Human Perception and Performance, in press.

Hayward, V., Maclean, K., 2007. Do it yourself haptics. Part I. IEEE Robotics and Automation Magazine 14 (4).

Hollerbach, J., 2008. Locomotion interfaces and rendering. In: Lin, M., Otaduy, M. (Eds.), Haptic Rendering: Foundations, Algorithms and Applications. AK Peters, Ltd.

Hunt, K.H., Crossley, F.R.E., 1975. Coefficient of restitution interpreted as damping in vibroimpact. ASME Journal of Applied Mechanics 42 (2), 440–445.

Iwata, H., 2008. Haptic interface. In: Sears, A., Jacko, J.A. (Eds.), The Human–Computer Interaction Handbook, second ed. Lawrence Erlbaum Associates, New York, 2008.

Jones, L.A., Sarter, N.B., 2008. Tactile displays: guidance for their design and application. Human Factors 50 (1), 90–111.

Kapralos, B., Zikovitz, D., Jenkin, M., Harris, L., 2004. Auditory cues in the perception of self-motion. In: Proceedings of the 116th AES convention.

Klatzky, R.L., Pai, D.K., Krotkov, E.P., 2000. Perception of material from contact sounds. Presence: Teleoperators and Virtual Environment 9 (4), 399–410.

Lakatos, S., McAdams, S., Caussé, R., 1997. The representation of auditory source characteristics: simple geometric form. Perception & Psychophysics 59 (8), 1180–1190.

Larsson, P., Västfjäll, D., Kleiner, M., 2004. Perception of self-motion and presence in auditory virtual environments. In: Proceedings of the Seventh Annual Workshop Presence, pp. 252–258.

Law, A.W., Peck, B.V., Visell, Y., Kry, P.G., Cooperstock, J.R., 2008. A multi-modal floor-space for displaying material deformation underfoot in virtual reality. In: Proceedings of the IEEE International Workshop on Haptic Audio Visual Environments and their Applications.

Li, X., Logan, R.J., Pastore, R.E., 1991. Perception of acoustic source characteristics: walking sounds. Journal of the Acoustical Society of America 90 (6), 3036–3049.

Lutfi, R.A., 2007. Human sound source identification. In: Yost, W.A., Fay, R.R., Popper, A.N. (Eds.), Auditory Perception of Sound Sources. Springer, New York, NY, pp. 13–42.

Lutfi, R.A., Liu, C.J., 2007. Individual differences in source identification from synthesized impact sounds. Journal of the Acoustical Society of America 122, 1017–1028.

Lutfi, R.A., Oh, E., Storm, E., Alexander, J.M., 2005. Classification and identification of recorded and synthesized impact sounds by practiced listeners, musicians, and nonmusicians. Journal of the Acoustical Society of America 118, 393.

McAdams, S., Chaigne, A., Roussarie, V., 2004. The psychomechanics of simulated sound sources: material properties of impacted bars. Journal of the Acoustical Society of America 115 (3), 1306–1320.

Miranda, E., Wanderley, M., 2006. New Digital Musical Instruments: Control And Interaction Beyond the Keyboard, AR Editions.

Nordahl, R., 2005. Auditory rendering of self-induced motion in virtual reality. M.Sc. Project Report, Department of Medialogy, Aalborg University Copenhagen.

Nordahl, R., 2006. Increasing the motion of users in photorealistic virtual environments by utilizing auditory rendering of the environment and ego-motion. In: Proceedings of Presence, pp. 57–62.

O'Modhrain, S., Essl, G., 2004. Pebblebox and crumblebag: tactile interfaces for granular synthesis. In: Proceedings of the NIME 2004, Hamamatsu, Japan, pp. 74–79.

Papoulis, A., 1984. Probability, Random Variables, and Stochastic Processes, second ed. McGraw-Hill, New York.

Pastore, R.E., Flint, J.D., Gaston, J.R., Solomon, M.J., 2008. Auditory event perception: the source-perception loop for posture in human gait. Perception & Psychophysics 70 (1), 13–29.

Rath, M., Rocchesso, D., 2005. Continuous sonic feedback from a rolling ball. IEEE Multimedia 12 (2), 60–69.

Rosenblum, L.D., 2004. Perceiving articulatory events: lessons for an ecological psychoacoustics. In: Neuhoff, J.G. (Ed.), Ecological Psychoacoustics. Elsevier Academic Press, San Diego, CA, pp. 219–248.

Rocchesso, D., Fontana, F., 2003. The Sounding Object. Edizioni di Mondo Estremo, Florence, Italy.

Sanders, R., Scorgie, M., 2002. The Effect of Sound Delivery Methods on a User's Sense of Presence in a Virtual Environment.

Sethna, J.P., Dahmen, K.A., 2001. Myers, crackling noise. Nature 410, 242–250.

Storms, R.L., Zyda, M.J., 2000. Interactions in perceived quality of auditory-visual displays. Presence: Teleoperators & Virtual Environments 9 (6), 557–580.

Turner, S., Turner, P., Carroll, F., O'Neill, S., Benyon, D., McCall, R., Smyth, M., 2003. Re-creating the Botanics: towards a sense of place in virtual environments. In: Proceedings of the Environmental Science Conference.

van den Doel, K., Kry, P., Pai, D., 2001. FoleyAutomatic: physically-based sound effects for interactive simulation and animation. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, pp. 537–544.

Väljamäe, A., Larsson, P., Västfjäll, D., Kleiner, M., 2005. Travelling without moving: Auditory scene cues for translational self-motion. In: Proceedings of ICAD'05.

Vastfjall, D., 2003. The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments. CyberPsychology & Behavior 6 (2), 181–188.

Visell, Y., Cooperstock, J., Franinovic, K., 2007. The ecotile: an architectural platform for audio-haptic simulation in walking. In: Proceedings of the 4th International Conference on Enactive Interfaces.

Visell, Y., Cooperstock, J., Giordano, B.L., Franinovic, K., Law, A., McAdams, S., Jathal, K., Fontana, F., 2008. A vibrotactile device for display of virtual ground materials in walking. In: Proceedings of Eurohaptics 2008.

Visell, Y., Law, A., Cooperstock, J., 2009. Touch is everywhere: floor surfaces as ambient haptic displays. IEEE Transactions on Haptics, 2 (3).

Watters, B.G., 1965. Impact noise characteristics of female hard-heeled foot traffic. Journal of the Acoustical Society of America 37, 619–630.

Wu, G., Chiang, J.H., 1996. The effects of surface compliance on foot pressure in stance. Gait & Posture 4 (2), 122–129.

YEAR 2

Can You Stand on Virtual Grounds? A Study on Postural Affordances in Virtual Reality

Tony Regia-Corte* INRIA Rennes, France
Maud Marchal * INSA/INRIA Rennes, France
Anatole Lécuyer * INRIA Rennes, France

ABSTRACT

The concept of affordance, introduced by the psychologist James Gibson, can be defined as the functional utility of an object, a surface or an event. The purpose of this article was to evaluate the perception of affordances in virtual environments (VE). In order to test this perception, we considered the affordances for standing on a virtual slanted surface. The participants were asked to judge whether a virtual slanted surface supported upright stance. The perception was investigated by manipulating the texture of the slanted surface (Wooden texture vs. Ice texture). Results showed an effect of the texture: the perceptual boundary (or critical angle) with the Ice texture was significantly lower than with the Wooden texture. These results reveal that perception of affordances for standing on a slanted surface in virtual reality is possible and comparable to previous studies conducted in real environments.

Index Terms: H.5.1 [nformation Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.1.2 [Information Systems]: User/Machine Systems—human factors; human information processing

1 INTRODUCTION

In daily life, we can notice that people are generally able to guide their activities in a suitable manner and without incident across the different interactions with their environment. During locomotion, for example, they succeed in both selecting optimal surfaces for supporting their weight and posture and in avoiding obstacles that could prevent locomotion. The concept of affordance, introduced by the American psychologist, James Gibson [3], defines very well this ability of the animal to guide behavior by perceiving the action possibilities offered by the surfaces of the environment. Thus, the affordance can be defined as the functional utility of an object, a surface or an event for an animal with given physical characteristics (height, weight) and some action capabilities (effectivities) defined according to the species and ontogenetic development. Similarly to the real world, virtual environments (VE) lead to a very large number of action possibilities. An important aspect regarding the interactions with VE is that these interactions involve both perception and action activities which are fundamental to affordances. Thus, the VE user has not only the possibility to perceive the VE and its different objects (via several perceptual modalities: vision, haptic, and audition) but he has also the possibility to act inside of the VE, by moving the body, walking, grasping and wielding virtual objects. These different aspects raise several important questions about the nature of VE: Are there affordances in the VE? Are we able to perceive affordances in the VE? Is there a difference between the perception of affordances in the VE and in the real world? And finally, what are the factors involved in the perception of these affordances? Therefore, the concept of affordance in the context of VE appears to be a wide and very interesting field of research for both VE researchers and psychologists. However, there is very little amount of studies in which the perception of affordances has been tested in VE. Consequently, the aim of the present study is to evaluate the perception of affordances when people are inside VE. In order to test the perception of affordances in such a condition, we have chosen to consider the perception of affordances for standing on a slanted surface. This perception is basic and fundamental in the interactions with our environment. In this paper, we begin by a review on the affordances in real and virtual worlds: the concept of affordance is explained, and the previous works on the perception of affordances in the context of Virtual Reality (VR) and postural activities are described. Regarding the experiment, participants were asked to judge whether a virtual slanted surface supported upright stance. This perception was investigated by manipulating the texture of the slanted surface (Wooden texture vs. Ice texture). Finally, results were analyzed in relation to previous works and different practical implications were suggested for several domains.

*e-mail: {tregiaco,mmarchal,alecuyer}@irisa.fr

2 AFFORDANCES IN REAL AND VIRTUAL WORLDS

2.1 The concept of Affordance

Gibson's work, mainly centered on the field of visual perception, is at the origin of the ecological approach to perception and action as opposed to the cognitive approach found in psychology. A fundamental tenet of the ecological approach is the claim that affordances are perceived directly [3]. In other words, the perception of affordances does not require mediation or internal processing by the perceiver. The direct perception of the affordance is possible because there is invariant information in the environment that uniquely specifies that affordance. A growing body of research has demonstrated that participants are capable of perceiving affordances to control their actions in various activities including stair climbing [8, 14], sitting on surfaces [8], walking through apertures [15], and walking up slopes [5]. Although these results allow a better understanding of the perception of affordances, there is still debate between researchers whether the affordance is an inherent property of the environment or an emergent property of the animal-environment system [13]. However, in both of these theoretical views, there is an agreement on the fact that the perception of affordances involves that the environmental properties (height, width, weight, distance, etc.) are not evaluated on an extrinsic scale with the units of measurement used in physics but are measured on an intrinsic scale according to certain relevant properties of the perceiver-actor, such as its own height, width and running speed [9]. Indeed, the aforementioned studies have demonstrated that perception of affordances is based on body-scaled information. In other words, actors perceive the properties of the environment in relation to themselves. In a study of the perception of stair climbing, Warren [14] asked observers to view stairs of different heights and judge which ones they could ascend in normal fashion. Warren found that observers judgments were consistent and accurate with respect to their actual stair-climbing capabilities; each persons maximum climbable riser height was a constant proportion (.88) of leg length. Studies of other actions identified similar invariant relationships between the critical action boundary and a relevant body part across actors of different sizes: sitting [8], and passing through apertures [15].

2.2 Affordances and Virtual Reality

Several researchers consider that the Gibson's ecological framework is a promising functional approach for defining the reality of experience in relation to the problem of designing virtual environments [2, 4]. For example, the perception of affordances could be a potential tool for sensorimotor assessment of physical presence, that is, the feeling of being physically located in a virtual place [7]. Therefore, Lepecq et al. [7] investigated the walk through a virtual aperture of variable widths. In the case of presence, the subjects body orientation, while walking, was hypothesized to be adapted to the width of the aperture and to their own shoulder width. The results of this study indicated that the locomotor postural patterns of subjects having to walk through a virtual aperture strongly resemble those of subjects who have to walk through a real aperture [15]. For most subjects, a behavioral transition from frontal walking to body rotation was observed as the width of the virtual aperture decreased. Finally, researchers have designed a conceptual model in order to evoke affordances in VE via sensory-stimuli substitution. Such a model can potentially guide VE designers in generating more ecologically valid designs [4].

2.3 Affordances for Standing on Surfaces

In the field of postural activities, different studies have shown that the stance can be an example of affordance; that is a given environment can afford stance for a given organism [3]. In a pioneering study, Fitzpatrick et al. [1] examined perception of affordance for supporting upright stance. The participants were asked to judge visually or haptically (i.e., by probing the surface with a hand-held rod while blindfolded) whether a slanted surface supports upright stance. In the experiments, participants stood at a distance of one meter from an inclined board, and either looked at the surface or explored the surface with the hand-held rod. Although participants were less confident and took longer to make haptic judgments in comparison to visual judgments, the perceptual boundary between supporting and not supporting did not differ for haptic and visual judgments (29.8 and 29.6 degrees respectively). The results also showed that the profiles of the responses time and confident judgments were similar for both perceptual systems: the exploration time increased and confidence decreased at the perceptual boundary. Moreover, this perceptual boundary was within a few degrees of the actual (behavioral) boundary on this behavior (approximately 30°). In a second experiment using ascending and descending methods of limits for the presentation of angles, the results also revealed that the perceptual boundaries occurred at steeper angles of inclination on descending trials than on ascending trials. This finding demonstrates a phenomenon known as enhanced contrast and suggests that perception of affordances in this task is a dynamical process [11]. In a more recent study using the same experimental paradigm, researchers have shown that the perception of affordance for supporting upright stance depended on height of center of mass [10]. In this study, participants performed the task while wearing a backpack apparatus to which masses were attached in different configurations. The developmental dimension was also examined in a study evaluating how children and adults perceived affordances for upright stance. The overall superiority of the adults relative to the children indicated clearly that there are developmental changes in the ability to perceive affordances [6]. Finally, we can notice that no studies have examined the effect of texture in this perception. This aspect will be explored in VE within the frame of the current paper.

3 Experiment

The aim of this experiment was to evaluate the perception of affordances in virtual reality. In order to investigate empirically this topic, we have chosen to focus our analysis on the perception of affordances for standing on a slanted surface. In this context, the perception was investigated by manipulating the texture of the slanted surface. To prevent an object from slipping down a slope, frictional force must be strong enough to overcome the pull of gravity. The amount of frictional force that is created depends on the coefficient of friction between the object and the surface of the slope. Thus, two contrasted textures (high-friction: Wooden vs. low-friction: Ice) were used for the slanted surface. The hypothesis of this experiment was that if the participant is able to perceive the affordance in the virtual reality, we should observe an effect of the texture: with a perceptual boundary (or critical angle) lower with the Ice texture than with the Wooden texture.

3.1 Participants

Twelve participants (2 females and 10 males) aged from 20 to 29 (mean = 24.9, SD = 2.8), took part in this experiment. All of them were right-handed, and none of them had known perception disorders. They were all naive to the purpose of the experiment.

3.2 Experimental Apparatus

The experiment was conducted in a closed room with dim light. We used the eMagin Z800 Head Mounted Display as display device, at 60 Hz and with stereoscopy enabled. The participant was upright in front of a table with the laptop computer running the application (Figure 1) and was wearing an opaque fabric on top of the HMD to avoid seeing the surrounding real world.

Figure 1: The experimental apparatus. Left: the participant wearing the HMD and the tracking equipment. Right: the participant equipped and wearing the opaque fabric in front of the table with the laptop computer.

The participants head was tracked by an ART ARTtrack2 infrared tracking system with 9 surrounding cameras for 360° tracking. The available tracking space was a cylinder with a 3 m diameter and a 2.5 m height.

3.3 The Virtual Environment

In the virtual environment, the participant was inside a room (width: 8.5 m × height: 4 m × length: 8.5 m) and stood upright 1 m from a slanted surface (width: 0.76 m × length: 1.56 m × thickness: 0.02 m). There were no contextual cues in the room. The floor of the room was displayed with a grey carpet, the walls and the ceiling with a brown paint. Two different textures were used for the slanted surface: a Wooden texture or an Ice texture (see Figure 2). The participant controlled the inclination of the slanted surface with the keyboard of the laptop computer. The participants virtual eye height (i.e., the position of the camera) corresponded to the actual participants eye height.

3.4 Procedure

The task in this experiment was to adjust the angle of inclination of the virtual slanted surface until the participant felt that it was just

barely possible for him (her) to stand on that surface with a normal upright posture. Normal upright posture was defined as standing with the feet flat (i.e., not on the toes) without bending at the hip or knees. Before the experiment, each participant was briefed about the task and was instructed to stay upright during experiment. The participant was allowed to move the head in order to explore the virtual environment. Once equipped with the HMD and the opaque fabric, the participant was conducted in front of the table with the laptop and the participants right hand was placed on the keyboard. The participant used their right hand fingers to press the computer keys. The participant could adjust the angle of the slanted surface with three keys: the up arrow to increase the inclination, the down arrow to decrease the inclination and the Enter key to validate the response. The resolution for one press on the up-down arrows was 0.25° and a continuous press on the keys was possible to adjust the inclination (5° / sec).

Figure 2: The participant controlled the inclination of the slanted surface with the keyboard of the laptop computer (Figure 1). The virtual environment was made up of a room with a slanted surface. Two different textures were used for the surface: a Wooden texture (left) and an Ice texture (right).

The method of adjustment was used for the measure of the critical inclination. For each trial, the angle of inclination of the surface was initially set at either the lowest angle of inclination (i.e., 0°) or the highest angle of inclination (i.e., 90°) and the participants adjusted the angle of inclination until they felt that the surface was set at the steepest angle that would support upright posture. Participants could view the surface for as long as they wished to determine whether they would be able to stand on the slanted surface. Once participants were satisfied with position of the surface, they pressed the Enter key to validate the response then a confirmation message appeared with a black screen and asked to press again the Enter key to confirm the response or to press the Space bar to return to the task. When the response was confirmed, the value of the inclination was recorded and the application displayed the next trial.

During the experiment, two different textures were used for the slanted surface: a Wooden texture and an Ice texture. No information was communicated to the participant about the texture of the slanted surface. Participants completed all the two texture conditions (Wooden and Ice) and the order of the conditions was counterbalanced across participants. In each condition, participants completed two ascending trials (in which the angle of inclination was initially set at 0°) and two descending trials (in which the angle of inclination was initially set at 90°). Ascending and descending trials alternated within a given condition, and the order of the sequence (i.e., whether an ascending or a descending trial was presented first in a given condition) was counterbalanced across participants. In this experiment, participants completed a total of 8 trials (2 texture conditions × 2 directions × 2 trials per condition). The duration of the experiment was approximatively 10 minutes.

3.5 Results

The mean angle of inclination chosen by the participants was considered as the perceptual boundary. A 2 (Texture: wooden vs. ice) × 2 (Direction: ascending vs. descending) repeated-measures ANOVA was conducted on these perceptual boundaries. The ANOVA revealed a significant effect of texture, $F(1,11) = 8.07$, $p = 0.016$ (see Figure 3), the perceptual boundary with the Ice texture ($M = 22.13°$, $SD = 8.52°$) was significantly lower than with the Wooden texture ($M = 27.60°$, $SD = 10.57°$).

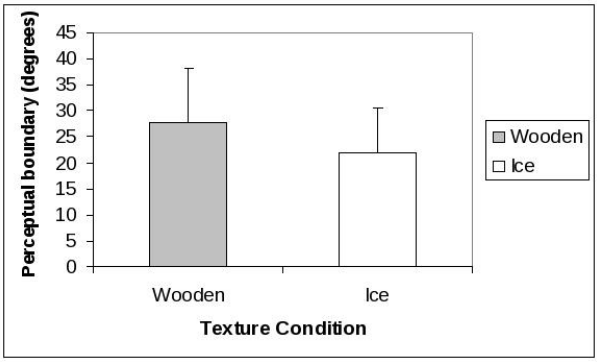


Figure 3: Perceptual boundary (or critical angle in degrees for standing on the slanted surface) as a function of the texture condition (Wooden and Ice).

The ANOVA also revealed a significant effect of direction, $F(1,11) = 6.83$, $p = 0.024$, the perceptual boundary occurred at a steeper angle of inclination when the surface was descending ($M = 26.09°$, $SD = 9.89°$), than when the surface was ascending ($M = 23.65°$, $SD = 8.34°$). The interaction between texture and direction was not significant ($F(1,11) = 1.38$, $p = 0.26$).

4 Discussion

This paper analyzed the perception of affordances for standing on a slanted surface in VR. During the experiment, participants were asked to judge whether a virtual slanted surface supported upright stance. Interestingly, participants showed a natural ability to perceive affordances in VR although they have no prior experience with the virtual slanted surface displayed. These results are interesting because they are consistent with the previous research conducted in real environments but also because they reveal several specificities. In our experiment, the texture of the slanted surface was manipulated. Results showed that the perceptual boundary with the Ice texture (22.13°) was significantly lower than with the Wooden texture (27.60°). This result revealed that the virtual information about friction was detected and used in VE. Thus, participants were able to differentiate visually a low-friction texture (Ice) from a high-friction texture (Wooden). Regarding the values of perceptual boundaries observed in our experiment, the comparison with previous works revealed that the perceptual boundary with the Wooden texture (27.60°) was close to those observed in real environments (29.25°, 31.40° and 31.13°) with a slight underestimation. A comparison of the values obtained both in real and virtual environments is provided in Table 1.

Furthermore, as in the previous works, our results also revealed that the perceptual boundaries occurred at steeper angles of inclination on descending trials than on ascending trials. This finding demonstrates a phenomenon known as enhanced contrast [11] and suggests that perception of affordances in this task is a dynamical process. This last result reinforces again the similarity observed between the perception in VE and real environments.

The slight underestimation observed between the perceptual boundaries in our experiment and previous works conducted in real environments is an interesting and paradoxal result. Indeed, we can imagine that people inside VE are aware to be in an unrealistic

		Perceptual boundaries (in degrees)		
		Mean	Ascending	Descending
Virtual Env.	Our study	27.60°	26.16°	29.04°
Real Env.	Fitzpatrick et al. [1]	29.25°	27.25°	30.45°
	Klevberg and Anderson [6]	31.40°	-	-
	Regia-Corte and Wagman [10]	31.13°	28.71°	33.56°

Table 1: Perceptual boundaries (or critical angle in degrees for standing on the slanted surface) for a Wooden texture as a function of the studies (in Virtual and Real environments). Mean, ascending and descending boundaries are distinguished.

world where their physical intregrity is not involved and where it is possible to risk dangerous behaviors. However, this slight underestimation indicated, on the contrary, that participants were more careful in VE. One possible explanation for this paradoxal result would be the presence of a time effect on the perception: at the beginning, the VE as a new environment involves a safety first effect with an underestimation of action possibilities. But during time and practice inside the VE, participants become more adapted and confident with a virtual perception reaching the real perception. And finally, participants adopt risky and dangerous behaviors leading to an overestimation of action possibilities. Thus, it would be interesting for the future research to consider the time factor in order to test this hypothesis.

It is important to notice that previous studies have shown that distances appear to be compressed in immersive virtual environments presented via head mounted display systems, relative to in the real world [12]. Thus, the slight underestimation observed in our study could indicate that the perception of affordances in VR would be also affected by the effect of compression. Research on this topic would be very interesting.

Regarding the practical implications of our study, the results suggest that people with motor impairments or balance disorders might improve their postural ability with specific activities in VR where they are confronted to different affordances. On the other hand, other practical implications would be in the context of urban projects, where the immersion in the 3D representations of buildings would allow to locate the uncomfortable affordances. The results observed in the present work call for additional investigations devoted to evaluate the perception of different affordances in VR (walking up slopes, stair climbing, gap crossing, and object reaching). It would be also interesting to conduct these investigations by considering different perceptual modalities (vision, haptic and audition).

5 Conclusion

The purpose of this work was to evaluate the perception of affordances in virtual reality taking as an example standing on a slanted surface. During the experiment, participants were asked to judge whether a virtual slanted surface supports upright stance. The task was to adjust the angle of inclination of the virtual slanted surface until the participant felt that it was just barely possible for him (her) to stand on that surface with a normal upright posture. The perception was investigated by manipulating the texture of the slanted surface. Thus, two contrasted textures (high-friction: Wooden vs. low-friction: Ice) were used for the slanted surface. Results showed an effect of the texture: the perceptual boundary (or critical angle) with the Ice texture was significantly lower than with the Wooden texture. Such results indicate that the perception of affordances for standing on a slanted surface in virtual reality is possible and comparable (with a slight underestimation) to previous studies conducted in real environments. Interestingly, it appears that virtual information about friction can be detected and used in VE. Finally, future research could investigate more precisely the role of texture and other factors on this perception in both virtual and real environments. Furthermore, future work could also evaluate the effects of order and transfer between real and virtual environments.

Acknowledgements

The authors would like to thank Mr. Laurent Bonnet and Mr. Gabriel Cirio for their help on the design of the virtual reality setup and experimental benchmark. This work was supported by the European community under FP7 FET- Open grant agreement n°222107 NIW Natural Interactive Walking.

References

[1] P. Fitzpatrick, C. Carello, R. C. Schmidt, and D. Corey. Haptic and visual perception of an affordance for upright posture. *Ecological Psychology*, 6(4):265–287, 1994.

[2] J. Flash and J. Holden. The reality of experience: Gibsons way. *Presence: Teleoperators and Virtual Environnements*, 7(1):90–95, 1998.

[3] J. Gibson. *The Ecological Approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc, original work published in 1979 edition, 1986.

[4] D. Gross, K. Stanney, and L. Cohn. Evoking affordances in virtual environments via sensori-stimuli substitution. *Presence: Teleoperators and Virtual Environnements*, 14(4):482–491, 2005.

[5] J. Kinsella-Shaw, B. Shaw, and M. Turvey. Perceiving walk-on-able slopes. *Ecological Psychology*, 4(4):223–239, 1992.

[6] G. Klevberg and D. Anderson. Visual and haptic perception of postural affordances in children and adults. *Human Movement Science*, 21(2):169–186, 2002.

[7] J.-C. Lepecq, L. Bringoux, J.-M. Pergandi, T. Coyle, and D. Mestre. Afforded actions as a behavioral assessment of physical presence in virtual environments. *Virtual Reality*, 13(3):141–151, 2009.

[8] L. Mark. Eyeheight-scaled information about affordances: A study of sitting and stair climbing. *Journal of Experimental Psychology: Human Perception and Performance*, 13(3):361–370, 1987.

[9] R. Oudejans, C. Michaels, F. Bakker, and M. Dolne. The relevance of action in perceiving affordances: Perception of catchableness of fly balls. *Journal of Experimental Psychology: Human Perception and Performance*, 22(4):879–891, 1996.

[10] T. Regia-Corte and J. Wagman. Perception of affordances for standing on an inclined surface depends on height of center of mass. *Experimental Brain Research*, 191(1):25–35, 2008.

[11] M. Richardson, K. Marsh, and R. Baron. Judging and actualizing intrapersonal and interpersonal affordances. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4):845–859, 2007.

[12] F. Steinicke, G. Bruder, K. Hinrichs, M. Lappe, B. Ries, and V. Interrante. Transitional environments enhance distance perception in immersive virtual reality systems. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, pages 19–26, 2009.

[13] T. Stoffregen. Affordances as properties of the animal-environment system. *Ecological Psychology*, 15(2):115–134, 2003.

[14] W. Warren. Perceiving affordances: Visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5):683–703, 1984.

[15] W. Warren and S. Whang. Visual guidance of walking through apertures: Body-scaled information for affordances. *Journal of Experimental Psychology: Human Perception and Performance*, 13(3):371–383, 1987.

Shake-Your-Head: Revisiting Walking-In-Place for Desktop Virtual Reality

Léo Terziman* Maud Marchal* Mathieu Emily† Franck Multon†
INSA / INRIA / DGA INSA / INRIA Université Rennes 2 / IRMAR M2S / Université Rennes2 / INRIA
Bruno Arnaldi* Anatole Lécuyer*
INSA / INRIA INRIA

Abstract

The Walking-In-Place interaction technique was introduced to navigate infinitely in 3D virtual worlds by walking in place in the real world. The technique has been initially developed for users standing in immersive setups and was built upon sophisticated visual displays and tracking equipments.

In this paper, we propose to revisit the whole pipeline of the Walking-In-Place technique to match a larger set of configurations and apply it notably to the context of desktop Virtual Reality. With our novel "Shake-Your-Head" technique, the user is left with the possibility to sit down, and to use small screens and standard input devices such as a basic webcam for tracking. The locomotion simulation can compute various motions such as turning, jumping and crawling, using as sole input the head movements of the user. We also introduce the use of additional visual feedback based on camera motions to enhance the walking sensations.

An experiment was conducted to compare our technique with classical input devices used for navigating in desktop VR. Interestingly, the results showed that our technique could even allow faster navigations when sitting, after a short learning. Our technique was also perceived as more fun and increasing presence, and was generally more appreciated for VR navigation.

CR Categories: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Input Devices and Strategies, Interaction Styles, User-centered Design; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality

Keywords: Walking, Walking-In-Place, Camera Motions, Head Tracking, First-Person-Navigation, Desktop Virtual Reality

1 Introduction

Navigation is one of the fundamental tasks needed for 3D interaction with Virtual Environments (VE) [Bowman et al. 2005]. The possibility to walk inside the VE is necessary in many applications of Virtual Reality (VR) such as for virtual visits (review of architectural and urban projects) or virtual training (technical procedures, military infantry).

In order to provide VR users with realistic sensations of walking while keeping them in their limited workspace, numerous types of VR interfaces have been proposed so far [Bowman et al. 2005]. Locomotion interfaces, such as unidirectional or omnidirectional treadmills [Stanney et al. 2002][Iwata and Yoshida 1999], are hardware solutions that enable the user to walk infinitely. But they are often costly, complex and cumbersome. Thus, many software interaction techniques have been developed based on classical 3D input devices such as joysticks or gamepads [Bowman et al. 2005]. But they induce very limited physical movements, and sometimes generate a sensation of "flying" in the virtual scene rather than walking [Lécuyer et al. 2006].

The Walking-In-Place (WIP) technique has been introduced by Slater et al. [Slater et al. 1995] to enable a real physical walking movement and an efficient navigation technique in 3D virtual environments. The user has to consciously walk in place while motions of his body are tracked and analyzed. The tracked Walking-In-Place motion is used as input for the locomotion simulation inside the VE. First implementations of WIP were all based on the processing of head positions using a neural network. More recent models track the positions of the heels or knees of the user to compute the resulting virtual locomotion [Feasel et al. 2008][Wendt et al. 2010]. However, all existing WIP techniques require the user to stand up, and they focus on immersive VR applications based on sophisticated tracking devices and head-mounted-displays or CAVE setups.

In this paper, we propose to revisit the whole pipeline of the WIP technique. Our intention is to extend it to match a larger set of configurations, by notably applying it to desktop setups. As a result, we propose to improve the WIP technique on three main aspects, corresponding to the three main innovations claimed in our paper:

- **Novel user interface capacities**. We introduce (1) the possibility for the user to stand or sit (using classical interaction paradigms of desktop VR with mouse), and (2) the possibility to use screens (with limited field of view) and low-cost tracking (webcam) by using the use of head movements as main input of WIP control laws.
- **Novel locomotion simulation**. We extend the range of possible motions with WIP to: walking, turning, jumping and crawling, using heuristics and control laws based exclusively on users' head movements.
- **Novel visual feedback of walking**. We introduce the possibility to combine our approach with more realistic visual feedback of walking, i.e. camera motions, that are known to improve the sensation of walking [Lécuyer et al. 2006].

2 Related Work

Navigating in a virtual environment can be achieved in numerous ways. The most straightforward solution probably consists in walking naturally in the real world, using for instance Head Mounted Display (HMD) together with a tracking system which measures head position for updating viewpoint (or virtual camera position and orientation) in the virtual environment. However, this solution can rarely be used due to the limited physical workspace in the real world, as well as limitations of the tracking devices.

To overcome this problem, numerous "locomotion interfaces" have been designed. The locomotion interfaces keep the user within

*e-mail: {lterzima|mmarchal|barnaldi|alecuyer}@irisa.fr
†e-mail: {Mathieu.Emily|Franck.Multon}@uhb.fr

	Input/Output Interfaces		Interaction Technique	
References	Output: Display	Input: Tracking	Control Law	Available Motions
[Slater et al. 1995]	HMD	Head	Neural network	FW, Turn
[Usoh et al. 1999]	HMD	Head	Neural network	FW, Turn
[Templeman et al. 1999]	HMD	Knees, waist, hands, head	Knee pattern	FW, BW, Diagonal, Sidestep, Turn
[Razzaque et al. 2002]	CAVE	Head	Neural network	FW, Turn
[Yan et al. 2004]	CAVE	Legs, waist, head	Leg speed relation	FW, Turn
[Feasel et al. 2008]	HMD	Shins	Filtered heels speed	FW (various speeds), Turn
[Wendt et al. 2010]	HMD	Shins	Biomechanical state machine	FW (various speeds), Turn

Table 1: *Overview of existing Walking-In-Place techniques.*

the limits of the physical workspace while allowing to navigate infinitely in the virtual world. A classical hardware approach consists in using a treadmill which can be either unidirectional (such as the Sarcos Treadport [Stanney et al. 2002]) or omni-directional (such as the Torus Treadmill [Iwata and Yoshida 1999]). Other locomotion interfaces can be based on feet-platforms [Iwata et al. 2001], walking inside a rotating sphere [Fernandes et al. 2003], or a set of moving tiles [Iwata et al. 2005]. However, the current locomotion interfaces are often too expensive, cumbersome, or complex for being integrated easily in both VR or desktop applications.

Thus, numerous interaction techniques have been developed to provide software-based navigation capabilities in VR. For instance, the World In Miniature (WIM) technique [Pausch et al. 1995] is based on a miniature graphical representation of the VE which can be used to select destinations in the VE. The virtual camera "flies" automatically to the new position selected in the WIM. The Magic Barrier Tape [Cirio et al. 2009] is another interaction metaphor which enables to navigate infinitely when immersed in a virtual scene, when located in a restricted walking workspace. The physical boundaries of the workspace are displayed in the virtual environment using a barrier tape that must not be crossed by the user but that can be moved in the VE with a joystick.

Besides, classical input devices such as mice, keyboards and joysticks can also be used within interaction metaphors to navigate from one virtual place to another [Bowman et al. 2005]. An efficient solution for navigating in 3D VE with a desktop configuration is the interaction paradigm proposed in most First-Person Shooter video games (FPS). It relies on mouse control (for orienting the viewpoint and the direction of motion), and keyboard keys (for forward, backward and lateral motions). However, such interaction techniques, relying exclusively on manipulated input devices, do not provide any proprioceptive or vestibular feedback of walking, and lead to a poor sensation of walking in the VE.

An efficient solution to enhance the sensation of walking in virtual worlds consists in playing with visual feedback and adding artificial camera motions [Lécuyer et al. 2006]. Such camera motions can simulate the visual flow corresponding to natural oscillations of the body and head during the walk. This approach also simulates the ocular reflexes present when walking [Hillaire et al. 2008]. Camera motions were shown to improve sensation of walking and perception of traveled distances in VE [Terziman et al. 2009].

Interestingly, novel commercial systems for optical tracking of body movements have appeared such as the EyeToy for Playstation 2 or, more recently, the Microsoft Kinect . Both systems are focused on interaction with multimedia content and video games but, for the moment, these systems have not been used for direct walking in 3D virtual environments such as when "walking in place".

Walking-In-Place (WIP) is an interaction paradigm designed for both an improved sensation of walking and an efficient navigation in virtual environments [Slater et al. 1995]. The WIP technique consists in consciously walking in place in the real world. Previous papers on WIP are presented in Table 1. The first implementation proposed for WIP was based on a HMD. Head tracking was used to predict and detect the steps of the user using a neural network [Slater et al. 1995]. When user's steps were detected, the viewpoint automatically moved forward in the VE. The advance direction was simply related to the head's direction. In the next implementations of WIP, the processing of head movements with a neural network was abandoned. The next systems could track the motions of the user's knees or heels. Then, using pattern recognition [Templeman et al. 1999], frequency analysis [Feasel et al. 2008] or biomechanically inspired state automaton [Wendt et al. 2010] the user's steps could be detected and translated into a variable advance speed depending on the steps frequency.

When considering Table 1, it seems that all the WIP techniques described in the literature so far are meant for highly immersive virtual environments and use either HMD or CAVE for visual display [Razzaque et al. 2002]. These setups use position/orientation tracking of body parts to detect footsteps and compute an advance speed [Feasel et al. 2008]. The few systems based on head movements are all using neural network as a processing scheme [Slater et al. 1995]. Turning in the virtual world is often achieved using the orientation of one of the tracked points (e.g., head, waist, chest) [Templeman et al. 1999].

Unfortunately, very few evaluations of WIP have been carried out up to now to compare it with other existing navigation techniques. Moreover, these rare studies have mostly used subjective questionnaires, excepted in [Whitton et al. 2005]. Considering the data available in the literature, it seems that WIP can increase presence compared to classical joystick-based interaction [Razzaque et al. 2002][Usoh et al. 1999]. WIP is also perceived as a more natural technique [Yan et al. 2004]. However, WIP might also result in an increased cybersickness [Usoh et al. 1999]. WIP seems also less efficient in term of usability than a real walk [Feasel et al. 2008].

3 Shake-Your-Head: Revisiting WIP for Desktop VR

We propose to revisit the whole pipeline of the Walking-In-Place technique to match a larger set of configurations and apply it to the context of desktop Virtual Reality. Our approach is schematized in Figure 1. The Figure highlights the main differences between our approach and the classical and existing WIP techniques.

With the Shake-Your-Head technique, the user can be standing or sitting (such as in traditional video games or desktop VR configurations). The user interacts with the system by means of head movements. These head movements can be captured using different tracking interfaces, but we insist on the use of low-cost optical tracking with standard webcams. The locomotion simulation proposes not only the computation of a virtual walking motion but also

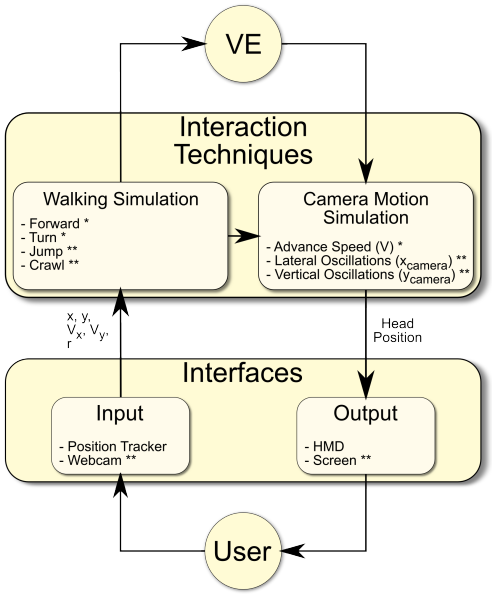


Figure 1: *Overview of our novel approach for Walking-In-Place (one star stresses improvement of existing component, two stars stress additional components).*

turning, jumping, and crawling possibilities. As a result, the user can perceive the locomotion in the virtual world by means of integrated virtual camera motions on the three axes of motion, to further enhance the sensation of walking.

In the following section, we will successively describe the different parts of our approach, namely: (1) the 3D user interface input/output, (2) the interaction techniques developed for the computation of the virtual locomotion, and (3) the visual feedback relying on camera motions.

3.1 Input/Output Interfaces

Our method proposes new features in terms of interfaces in order to extend the set of configurations where WIP can be applied, especially for Desktop VR. Thus, we propose to incorporate new devices for both input and output user interfaces in the VE.

3.1.1 Input: Tracking Based on Head Motions

The input interface of our method is only based on head motions. In our implementation, the input interface is reduced to a webcam, allowing the use of our method for Desktop configuration without any additional peripheral. However, our method can also be implemented with other classical VR tracking systems.

Use of Head Movements. The main concept of our method is to exploit the head oscillations as a transposition of the one observed during natural walking. While walking, the head of the user oscillates along the lateral, vertical and forward axes [Lécuyer et al. 2006]. The oscillations are strongly correlated to gait events and foot steps. Moreover, these oscillations also occur while walking in place and can be measured.

The head motions are classically retrieved in the existing WIP techniques thanks to the use of regular position trackers [Slater et al. 1995]. More generally, any tracking device can be used, as long as its accuracy is within the range of $1cm$. Moreover, the acquisition process of the required position does not have any influence, as long as the real time constraint is maintained. In our method, we propose the use of a video camera system to handle the tracking of the user head. Thus, our interaction technique can be deployed on a large scale at low cost for training purpose or video games for example.

Extracted Data. In our method, we propose to use 3 Degrees of Freedom (DOF) that can be easily accessed in the image frame provided by the webcam:

- The lateral position x (and the computed speed V_x);
- The vertical position y (and the computed speed V_y);
- The rotation of the head in the frontal plan r.

These three head motions are illustrated in Figure 2.

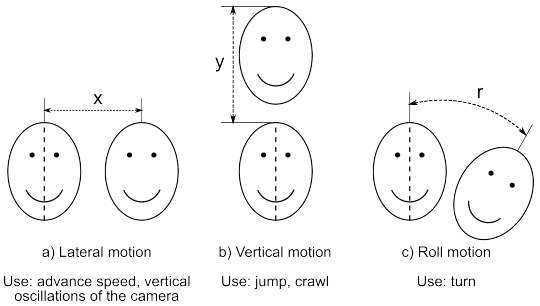


Figure 2: *Extracted head motions: (a) lateral motion, (b) vertical motion, (c) roll motion.*

Implementation. The use of the webcam to track the 3D position of the user head without using markers requires the implementation of real time constraints for the algorithms, i.e. more than 25 frames per seconds. Our implementation is based on the Camshift (Continuously Adaptive Mean Shift) algorithm [Bradski 1998] implemented in the OpenCV library. This algorithm is based on color tracking and is well-suited for real-time tracking of features of a given color, such as the face of the user.

While the user is standing in front of the webcam, our algorithm recognizes him as an ellipsoid. The position of the head (x, y) can be deduced from the center of the ellipse and the orientation angle of the head r is given by the angle with the vertical of the ellipse. However, x and y depend on the resolution of the used webcam. Thus, we compute a normalized position (x_n, y_n) on [-1; 1] on both axes. From the normalized positions of the head (x_n, y_n), we compute the instantaneous speeds V_x and V_y, and we use Kalman filters on all values to reduce the noise produced by the algorithm.

3.1.2 Output: Immersive and Desktop Visual Displays

Our method can be used with both immersive and regular screens as output interface with the VE. A requirement of desktop VR applications was to propose a technique that is usable with a limited Field of View (FOV). In regular WIP techniques, the provided FOV is always 360°, except in [Razzaque et al. 2002] where they used the redirected walking technique to simulate a 360° FOV in a 4-walls CAVE. In this paper, our technique is evaluated with both LCD laptop screens to fulfill the Desktop VR pre-requisite or video projected output. Our method can also be used with the classical output used with WIP, i.e. with HMDs or CAVEs.

3.2 Walking Simulation

3.2.1 Walking States

The main goal of our interaction technique is to translate the inputs of the user, i.e. head motions, into virtual motions in the VE. The user should be able to perform various motions while navigating in the VE. We implemented different locomotion states: walking, turning, jumping and crawling. To manage these different states, we added a state automaton to our algorithm (Figure 3). The state transitions are governed by the user head motions and the main inputs are the lateral velocity V_x and the vertical velocity V_y.

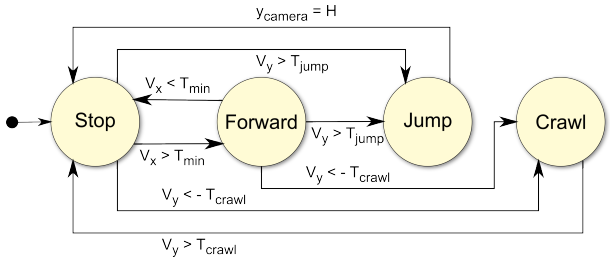


Figure 3: *Walking state automaton: the main inputs are the velocities V_x and V_y. V_x allows the transition from the Stop state to the Forward motion; V_y allows the transition to the Jump or Crawl states. The Turn motion is not represented as it can be activated in any state, depending on the head orientation.*

3.2.2 Forward State

The forward movements in the VE are governed by the lateral oscillations as main input. Our technique is designed to emphasis the idea of walking with a variating advance speed depending on the user head motions. The advance velocity V_a oscillates regularly, accordingly to the lateral head motions. One head oscillation period corresponds to one step. The footstep events are simulated by a null advance speed and correspond to a modification of the lateral velocity sign of the user's head. Thus, when the user's head reaches the maximal amplitude of the oscillations, the oscillating speed is null, as well as the advance speed, simulating a foot step.

For more realistic movements, we introduced two thresholds T_{min} and T_{max}. T_{min} allows to stop the forward movements when the lateral head motions are too small. T_{max} allows to avoid unrealistic high speed walks. The advance velocity V_a is computed in two steps in order to test these thresholds. The equations of V_a are:

$$\begin{aligned} V_{n_1} &= \frac{min(abs(V_x), T_{max})}{T_{max}} \\ V_{n_2} &= \begin{cases} 0 & if\ V_{n_1} < T_{min} \\ V_{n_1} & otherwise \end{cases} \\ V_a &= V_{n_2} * V_{max} \end{aligned} \quad (1)$$

Finally, the advance speed V of the camera inside the VE is adapted depending on the current locomotion state and is given by:

$$V = \begin{cases} V_a & if\ state = walk \\ 0.4 * V_a & if\ state = crawl \\ V_{max} & if\ state = jump \end{cases} \quad (2)$$

We chose normalized thresholds with the following values: $T_{min} = 0.05$ and $T_{max} = 0.5$. We also set V_{max} to 3.5 meters per second, corresponding to the maximal speed that can be achieved.

3.2.3 Jump and Crawl States

Comparing to existing WIP techniques, we chose to add 2 new states to the navigation possibilities in the VE: Jump and Crawl motions. The jump and crawl states are governed by the vertical oscillations of the user's head. If the vertical velocity exceeds normalized thresholds T_{jump} in upward direction and T_{crawl} in downward direction, the user can jump and crawl respectively in the VE. In practice, it means that the user will need to slightly jump or bend forward if he is seated, or jump or crouch down if he is standing. The user has to stand-up to stop crawling.

When a jump is detected, the vertical position of the camera is set to follow a classical parabolic trajectory defined by:

$$y_{camera} = \frac{1}{2} * g * t^2 + V * t \quad (3)$$

with g the gravity acceleration and t the time. The jumping state is left automatically while landing, i.e. when the camera reaches again its normal height H (known as the reference state when the algorithm starts). After preliminary testings, we set $T_{jump} = 0.3$ and $T_{crawl} = 0.4$. While the crawling state is activated, the vertical position of the virtual camera y_{camera} is lowered by 1 m.

3.2.4 Turn State

In parallel to Forward, Jump and Crawl states, the user has the possibility to turn inside the VE in order to modify his navigation direction. During a turn in a normal walk, the human body leans slightly in direction of the center of the turn to compensate the centrifugal force [Courtine and Schieppati 2003]. This phenomenon is often reproduced by video games players which can tend to lean in the direction of the turn even if it does not have any influence on their in-game trajectory. Thus, we choose to use this property to implement turns in our system as a control law based on the head orientation on the roll axis. To turn in the VE, the user has to lean his head in the left or right side respectively to turn left or right in the VE. The rotation speed V_r of the virtual camera is given by:

$$V_r = \begin{cases} V_{r_{max}} & if\ r > r_{max} \\ -V_{r_{max}} & if\ r < -r_{max} \\ 0 & otherwise \end{cases} \quad (4)$$

where r_{max} is the minimum angle of head inclination to start the rotation and $V_{r_{max}}$ is the maximal angular speed of the rotations. In our experiment, we set $r_{max} = 15°$ and $V_{r_{max}} = 45°/s$.

3.3 Visual Feedback Based on Camera Motions

To further emphasize the perception of walking in the VE, we extended the visual rendering of the WIP using camera motions driven by the user's head oscillations. There are existing models in the literature that make the virtual camera oscillating along the three axes ([Lécuyer et al. 2006] for example). However, the oscillations are totally independent from the user interactions.

We introduce a new model of camera motions adapted to the user's head motions. The camera oscillations along the different axes must follow the user in real time to maintain the coherency of the system. Thus we have implemented a novel visual feedback with camera motions along the vertical, lateral and advance axes.

3.3.1 Advance Oscillations

The advance speed V of the view point already oscillates. The camera motions are indeed intrinsically linked to the advance velocity

of the control law presented in paragraph 3.2.2. As a result, extra camera motion is not necessary along this axis and the advance camera velocity corresponds exactly to V.

3.3.2 Lateral Oscillations

In order to move in the VE, the user has to make his head oscillating from left to right. Thus, as the user moves in front of the screen, his view point of the scene is modified to follow the head oscillations.

The lateral oscillations of the camera are computed as a function of the user's position. If d is the distance of the user to the screen and α and β the opening angles of the webcam, the real world position of the user in front of the screen depends on the normalized coordinates x_n and y_n. The real world position of the user's head is given by the following coordinates:

$$\begin{cases} x_{real} & = x_n * d * tan(\alpha/2) \\ y_{real} & = y_n * d * tan(\beta/2) \end{cases} \quad (5)$$

Finally, the virtual camera is moved along the lateral axis by a distance x_{camera} equals to: $x_{camera} = A_x * x_{real}$. We set the scale factor $A_x = 1$ to match the user's head displacement and thus generate the illusion that the screen is a window through which the user can observe directly the VE. However, other values can be used to amplify the camera motions for example. The webcam used during the experiment was such as $\alpha = 60^\circ$ and $\beta = 45^\circ$.

3.3.3 Vertical Oscillations

The vertical oscillations of the camera can not be computed with the same algorithm as for lateral oscillations. In a desktop VR context the user can be seated and not be able to produce high vertical oscillations.

In our method, we propose to generate pseudo-sinusoidal vertical camera oscillations based on the current phase of the virtual gait cycle. Similarly to advance speed control law, the vertical amplitude y_{camera} of the camera oscillations is given by:

$$y_{camera} = V_{n_2} * y_{camera}^{max} \quad (6)$$

where y_{camera}^{max} is the amplitude of the vertical oscillations for velocities greater or equal to the T_{max} threshold. For smaller speeds, the amplitude of the oscillations is proportional to this maximum, thus increasing the perception of the variations in advance speeds. Using the same factor between the camera motions and the advance velocity forces the synchronization, resulting in a smooth final visual rendering. In our implementation, we set $y_{camera}^{max} = 15cm$.

3.4 Discussion

To summarize, our approach is composed of (1) an input interface based on the sole user's head movements, (2) a locomotion simulation in the VE composed of various possibilities such as jumping, crawling, turning, and (3) a visual feedback of walking relying on oscillating camera motions. The head motions are tracked along 3 Degrees of Freedom (DOF): lateral, vertical and roll axis (Figure 2). These different physical motions are transposed in virtual movements thanks to a locomotion automaton (Figure 3). We then added oscillating camera motions (Equations 2, 5 and 6) to the visual feedback to enhance the walking sensation. The different control laws were parametrized after preliminary testings. But of course some parameters can be modified in order to amplify/decrease some effects during the locomotion simulation. Besides, other movement possibilities could also be envisaged and added to our automaton such as running state or backward movement.

4 Evaluation

The evaluation of the proposed technique was performed using a comparison with classical techniques in Desktop VR. We chose keyboard and joystick peripherals for seated and standing positions respectively as they are often used in Desktop VR context. The experiments were conducted using 3D VE displayed on a screen and we investigated the effectiveness of our technique to travel complex paths composed of different gates placed in the VE.

In this paper, we choose to not compare our technique to existing WIP techniques, and instead used more common interfaces that followed our low cost requirement. We conducted the evaluation of the proposed technique in both immersive Standing Up (SU) position and Desktop Sitting Down (SD) position. The keyboard and joystick were chosen respectively as the control conditions in the SD position and the SU position. Both keyboard and joystick peripherals are referred as Control techniques (Ctrl) in the following paragraphs. Our technique is referred with the "WIP" suffix.

4.1 Method

4.1.1 Virtual Environment

The evaluation was performed within a 3D virtual environment without any contextual cues. The only landmarks were the gates that the user had to navigate through. A fog effect was added to mask the distant gates, allowing to perceive only the 2 or 3 closest gates. A texture on the ground provided useful visual flow information during the navigation. The scene was normally illuminated, and no shadows were drawn as illustrated in Figure 4.

Figure 4: *Virtual environment with different gates to represent the path that the user has to navigate through.*

4.1.2 Population

Twelve participants (10 males and 2 females) aged from 22 to 35 (mean=25.4, standard deviation=3.35) performed the experiment. Two participants were left-handed and none of them had known perception disorders. All participants were used to VEs but were naïve with respect to the proposed technique, as well as to the experimental setup and purpose of the experiment.

4.1.3 Experimental Conditions

The experiment was carried out in a room with controlled luminosity (using two projectors). There were two visual conditions corresponding to the two different configurations (SD and SU). The two configurations are illustrated in Figure 5. For the SU configuration, participants were at a distance of $1.5\ m$ in front of a $1.72\ m$ large and $1.24\ m$ height back-projected screen (physical field of view of 60° horizontally and 45° vertically). During the SD configuration, they sat in front of a 17 inches widescreen laptop screen (physical field of view of 30° horizontally and 18.5° vertically). In both cases, the resulting image had a resolution of 1600×1200 pixels. We used monoscopic rendering, with a frame rate of $60\ Hz$. The projector used was a DepthQ Stereoscopic.

(a) (b)

Figure 5: *Description of the experimental setup for (a) the Stand-Up (SU) configuration, and (b) the Sit-Down (SD) configuration.*

4.1.4 Experimental Apparatus

In our experiment, the goal was to compare our technique to classical interface devices (keyboard and joystick). During the experiment, the participants had to navigate in the VE through 3 different paths composed of 8 gates each. The 2 first paths, called Normal, were composed of 3 meters by 3 meters wide gates regularly disposed in order to form a slalom. The third path, called Steeple, was composed of 1.5 meters high tunnels and 80 cm high fences (4 of each). For this path, the users had to jump and crawl to pass the gates, while the 2 other paths only required a simple navigation.

The classical interface devices were implemented as follow:

- Keyboard: Forward and turn motions were triggered using the forward, left and right arrow keys. The jumping was triggered by the left Shift key and the crawling by the left Control key.
- Joystick: Forward and turn motions were triggered using the left joystick of a gamepad. Jumping and crawling were triggered with two of the right buttons.

To provide a fair comparison between WIP and Ctrl conditions, the oscillating camera motions were also implemented in the two Ctrl conditions. To do so, we used the best implementation described in [Lécuyer et al. 2006]. Moreover, the Ctrl condition advance speed was set to match the WIP condition average advance speed, which had been measured in a preliminary phase.

The participants had to press the "Enter" key at the beginning of each block of conditions to launch the experiment. At any time during the experiment they had the possibility to make a break by pressing the "Space" key. After each trial, the participants were automatically teleported to the beginning of the next trial. A black screen displayed 2 s notified the beginning of the new trial.

4.1.5 Experimental Plan

The participants were exposed to 4 blocks of 18 trials each: one block for each of the possible combinations of the experimental conditions. The combinations were the following: (1) Sitting Down, using the keyboard (SD-Ctrl); (2) Sitting Down, using the our technique (SD-WIP); (3) Standing Up, using the joystick (SU-Ctrl); (4) Standing Up, using the proposed technique (SU-WIP).

The participants were split in 4 groups equally composed of 3 people each. Two groups started with the SD configuration and the 2 other with the SU configuration. We counter-balanced the conditions between them, meaning that the group starting with Ctrl condition in the SD configuration starts with WIP condition in the SU configuration. For each of the 18 trials of a block, the participants had to navigate through the 3 paths. The 12 first trials were composed only of the normal paths (6 for each, in a random order), and the last 6 trials corresponded to the steeple path.

4.2 Collected Data

The WIP techniques are generally evaluated only with subjective questionnaires to evaluate presence or cybersickness. In our evaluation, we added new criteria based on the task performed by the participants. Thus, we measured for each participant the task completion time for each trial and the percentage of success for the different gates. A subjective questionnaire was also proposed.

4.3 Results

4.3.1 Task Completion Time

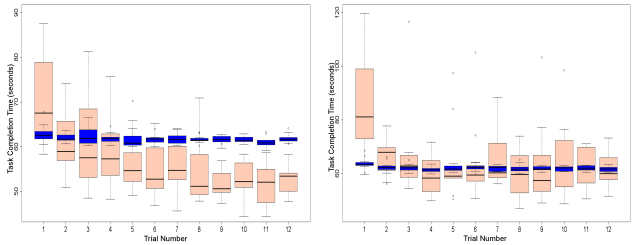
For each participant, the task completion time of each trial was measured for the different experimental conditions. An exploratory analysis was first performed. A Principle Component Analysis revealed the presence of one spurious individual who has been taken out from the analysis. A specific analysis was developed to study the learning effect of the two conditions (Joystick/Keyboard and WIP techniques). A linear model where all conditions are mixed was fitted to explain the relation between the task completion times and the trial number. It revealed that the slope of the linear regression was significantly lower than zero (p-value < 0.000001), reflecting a significant decrease in the task completion time as the number of trials increases. The same analysis, where the first trial was removed, showed that the slope was not significantly different from zero anymore. In the following paragraph, the first trial was removed from the analysis as it corresponds to a learning effect.

A two-way ANOVA was performed on the 2 different conditions (Joystick/Keyboard and WIP techniques) and the 2 different positions (Sit-Down and Stand-Up). A post-hoc analysis using Tukey's procedure was then performed. The ANOVA was achieved separately for the two different types of paths (normal/steeple).

Concerning the normal path, the two-way ANOVA accounting for the conditions and the positions revealed a significant dependency between the position and the task completion time ($F(1, 11) = 31.9981$, $p < 0.0001$) and between the condition and the task completion time ($F(1, 11) = 6.0449$, $p = 0.0143$). Interaction between condition and position was also considered as a significant factor to discriminate task completion time ($F(1, 11) = 27.1921$, $p < 0.0001$). Post-hoc analysis showed that the task completion time in the SD-WIP configuration ($M = 55.67s$) was significantly lower than in the SD-Ctrl configuration ($M = 61.72s$), adjusted p-value < 0.0001, in the SU-Ctrl configuration ($M = 62.06s$), adjusted p-value < 0.0001, and in the SU-WIP configuration ($M = 64.23s$), adjusted p-value < 0.0001. The other pairs of effects did not give any significant adjusted p-values.

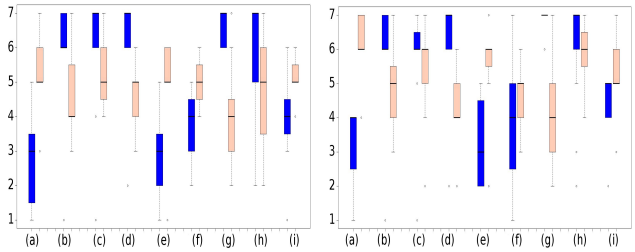
The results concerning the different conditions are represented in Figure 6 for Sit-Down and Stand-Up experiments respectively, for the normal path only. The results are ordered in function of the trials. The first trial was kept to illustrate the learning effect.

For the steeple path, the pre-analysis suggested the presence of a spurious individual and the existence of one learning trial. The two-way ANOVA accounting for the conditions and positions revealed a significant dependency between the position and the task completion time ($F(1, 11) = 8.5665$, $p < 0.005$), and the condition and the task completion time ($F(1, 11) = 11.8925$, $p < 0.001$. Interaction between condition and position was also considered as a significant factor to discriminate task completion time ($F(1, 11) = 8.7647$, $p < 0.005$). Post-hoc analysis showed than the task completion time in the SU-WIP configuration ($M = 86.07s$) was significantly higher than in the SD-Ctrl configuration ($M = 74.78s$), adjusted p-value < 0.0001, in the SU-Ctrl configuration ($M =$



(a) Sit-Down Experiment (b) Stand-Up Experiment

Figure 6: *Task completion time for the two different techniques on normal paths for (a) SD experiments and (b) SU experiments. The blue and red light colors correspond to the Ctrl and WIP conditions respectively. Each boxplot is delimited by the quartile (25% quantile and 75% quantile) of the distribution of the condition over the individuals. The median is also represented for each condition.*



(a) Sit-Down Experiment (b) Stand-Up Experiment

Figure 7: *Results for subjective rating for the two different techniques for (a) SD experiments and (b) SU experiments. The blue and red light colors correspond to the Ctrl and WIP conditions respectively. The subjective criteria are (a) Fun, (b) Easiness of Use, (c) Intuitive, (d) Accuracy, (e) Presence, (f) Walking realism, (g) Fatigue, (h) Cybersickness and (i) Global appreciation. Each boxplot is delimited by the quartile (25% quantile and 75% quantile) of the distribution of the condition over the individuals. The median is also represented for each condition.*

$74.72s$), adjusted p-values < 0.0001 and in the SD-WIP configuration ($M = 75.64s$), adjusted p-values < 0.0001. The other pairs of effects did not give any significant adjusted p-values.

4.3.2 Accuracy

For each participant and for each trial, the percentage of errors for the different paths was measured. The two types of path (Normal and Steeple) are separated. The resulting percentages are for the normal path: 0.61% of error for SD-WIP configuration, 0% of error for SD-Ctrl configuration, 1.39% of error for SU-WIP configuration, 0.09% of error for SU-Ctrl configuration.

The resulting percentages are for the steeple path: 20.49% of error for SD-WIP configuration, 7.81% of error for SD-Ctrl configuration, 27.28% of error for SU-WIP configuration, 13.19% of error for SU-Ctrl configuration. We found a significant effect between Ctrl and WIP techniques for the steeple path.

4.3.3 Subjective Questionnaire

After Sit-Down and Stand-Up configurations, a preference questionnaire was proposed in which participants had to grade from 1 (low appreciation) to 7 (high appreciation) the four different conditions (SD-Ctrl, SD-WIP, SU-Ctrl, SU-WIP) according to 9 subjective criteria: (a) Fun, (b) Easiness of Use, (c) Intuitive, (d) Accuracy, (e) Presence, (f) Walking realism, (g) Fatigue, (h) Cybersickness and (i) Global appreciation. Figure 7 shows the results concerning the grades (Likert-scale) obtained by the two different techniques for each of the subjective criteria, for the two experimental conditions (SD and SU). The grade 7 for Fatigue and Cybersickness respectively means that the technique does not induce any fatigue and does not imply any cybersickness feeling.

Concerning SD configuration, no significant effect was found for the following criteria: Intuitive ($p = 0.052$) and Cybersickness ($p = 0.12$). Concerning SU configuration, no significant effect was found for the following criteria: Intuitive ($p = 0.3$), Walking realism ($p = 0.19$) and Cybersickness ($p = 0.21$). We found a significant effect for all other criteria. In particular, our technique was better ranked for Fun, Presence and Global Appreciation, for both configurations.

5 Discussion

Our results suggest that our technique can allow efficient navigation even compared with standard and well-known input devices such as keyboards and gamepads. The participants could sometimes go even faster with WIP, without any strong loss in precision. The WIP-based interaction seems also fast to learn, after only a couple of trials. The technique is well appreciated and perceived as more immersive and more fun than classical configurations.

The quick learning of our technique could be explained by the fact that interfaces based on webcam are generally intuitive and simple to learn [Polaine 2005]. After the learning phase, WIP tends to become faster than the keyboard in sitting condition. One explanation could be that with our technique (but also with the joystick in the standing condition) we could observe that participants tended to turn without stopping their advance motion. On the contrary, with the keyboard condition, participants tended to walk and turn sequentially, which might have globally increased the task completion time. Another explanation could be that, in our implementation of WIP, the advance speed is influenced by the speed of lateral movements. The seated position allows the user to make faster oscillations than the standing position, and thus to accelerate the walking motion by making fast lateral oscillations. Interestingly, these faster motions did not impair the precision of users.

The longer task completion time observed for WIP in the steeple paths (involving jumping and crawling motions) could be due to unexpected behaviors which induced incorrect transitions in our locomotion automaton. Indeed, some participants acted as if they were ”anticipating” the jumps and bent down prior jumping. We could easily fix this problem in the future by using additional conditions in our automaton based on both speed and position.

Results from the questionnaire are very consistent with previous subjective evaluations of WIP [Usoh et al. 1999]. In our study, the WIP is more appreciated, and is perceived as more fun, and improving presence. As expected, more standard techniques (i.e. joystick and keyboard) are found easier to use, more precise, and less tiring (as they induce less physical movements). Interestingly, impression of cybersickness is not increased by WIP. This could be due to our desktop (and thus less immersive) configurations. Last, realism of walking in the VE was significantly improved only in the sitting condition. The perception of walking with WIP is actually quite complex, as participants wrote: ”we have the impression to be a video game character”, ”the motions are exaggerated”, or ”we really have the sensation of walking, and not running”. In the standing condition, some participants found that the physical motions were closer to ”skiing” or ”skating”, as they noticed that they did not lift

their feet from the ground but only oscillated their body. For these people we could further stress in the future that our implementation of WIP still works very well when lifting the feet and walking in place, as the oscillations of the head can be captured the same way in both situations (lifting the feet or not).

Taken together, our results suggest that our Shake-Your-Head technique could be used in a wide range of applications, when navigating in a 3D world, in sitting or standing configurations. It seems to be both a low-cost and an efficient paradigm that can match a lot of walking motions. It could thus be used for training in VR with more physical engagement (military infantry, vocational procedures), or more realistic virtual visits such as for project review in architecture or urban planning.

6 Conclusion

This paper proposed to revisit the whole pipeline of the Walking-In-Place technique to match a larger set of configurations and apply it notably to the context of Desktop Virtual Reality. We have introduced the Shake-Your-Head technique using as sole input the head movements of the user. It can be used in a desktop configuration with the possibility for the user to sit down and to navigate in the VE through small screens and standard input devices such as a basic webcam for tracking. We implemented various motions such as turning, jumping and crawling in the locomotion simulation. We also introduced the use of additional visual feedback based on camera motions to enhance the walking sensation.

We conducted an experiment in order to evaluate our technique compared to standard techniques such as keyboard and joystick. In this experiment, participants had to walk through a series a gates forming a slalom path. The evaluation was performed both in an immersive and desktop configurations. We notably found that WIP technique only requires a small learning time to allow faster navigation in seated position compared to the keyboard. Moreover, our technique was more appreciated and considered as more fun and inducing more presence than the other classical techniques.

Future work will focus on improving the underlying locomotion simulation. We could add new features such as running or backward movements in the VE. We could also improve the control laws with a more biomechanically-inspired description. Second, we also plan to conduct more evaluations, especially for further studying the perceptive aspects and measuring the influence of WIP on estimating spatial properties of VE.

References

BOWMAN, D. A., KRUIJFF, E., LAVIOLA, J. J., AND POUPYREV, I. 2005. *3D User Interfaces: Theory and Practice*. Addison-Wesley Boston (MA).

BRADSKI, G. R. 1998. Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal 2*, 2, 12–21.

CIRIO, G., MARCHAL, M., REGIA-CORTE, T., AND LÉCUYER, A. 2009. The magic barrier tape: a novel metaphor for infinite navigation in virtual worlds with a restricted walking workspace. In *Proceedings of the ACM Symposium on VRST*, 155–162.

COURTINE, G., AND SCHIEPPATI, M. 2003. Human walking along a curved path. i. body trajectory, segment orientation and the effect of vision. *Eu. J. of Neuroscience 18*, 1, 177–190.

FEASEL, J., WHITTON, M. C., AND WENDT, J. 2008. Llcm-wip: Low-latency, continuous-motion walking-in-place. In *IEEE Symposium on 3D User Interfaces*, 97 –104.

FERNANDES, K., RAJA, V., AND EYRE, J. 2003. Cybersphere: the fully immersive spherical projection system. *Communications of the ACM 46*, 9, 146.

HILLAIRE, S., LÉCUYER, A., COZOT, R., AND CASIEZ, G. 2008. Using an eye-tracking system to improve camera motions and depth-of-field blur effects in virtual environments. In *Proceedings of IEEE Virtual Reality*, 47–50.

IWATA, H., AND YOSHIDA, Y. 1999. Path reproduction tests using a torus treadmill. *Presence: Teleoperators and Virt. Env. 8*, 6, 587–597.

IWATA, H., YANO, H., AND NAKAIZUMI, F. 2001. Gait master: A versatile locomotion interface for uneven virtual terrain. In *Proceedings of IEEE Virtual Reality*, 131–137.

IWATA, H., YANO, H., FUKUSHIMA, H., AND NOMA, H. 2005. Circulafloor: A locomotion interface using circulation of movable tiles. In *Proceedings of IEEE Virtual Reality*, 223–230.

LÉCUYER, A., BURKHARDT, J.-M., HENAFF, J.-M., AND DONIKIAN, S. 2006. Camera motions improve the sensation of walking in virtual environments. In *Proceedings of IEEE Virtual Reality*, 11–18.

PAUSCH, R., BURNETTE, T., BROCKWAY, D., AND WEIBLEN, M. E. 1995. Navigation and locomotion in virtual worlds via flight into hand-held miniatures. In *Proceedings of the SIGGRAPH '95*, 399–400.

POLAINE, A. 2005. The flow principle in interactivity. In *Proceedings of the 2nd Au. conf. on Interactive Entertainment*, 158.

RAZZAQUE, S., SWAPP, D., SLATER, M., WHITTON, M. C., AND STEED, A. 2002. Redirected walking in place. In *Proceedings of the EGVE '02 workshop on Virtual environments*, 123–130.

SLATER, M., USOH, M., AND STEED, A. 1995. Taking steps: The influence of a walking technique on presence in virtual reality. *ACM Trans. on Computer-Human Interaction 2*, 3, 201–219.

STANNEY, K. M., ET AL. 2002. *Handbook of Virtual Environments: Design, Implementation, and Applications*. Lawrence Erlbaum Associates.

TEMPLEMAN, J. N., DENBROOK, P. S., AND SIBERT, L. E. 1999. Virtual locomotion: Walking in place through virtual environments. *Presence: Teleoperators and Virt. Env. 8*, 6, 598–617.

TERZIMAN, L., LÉCUYER, A., HILLAIRE, S., AND WIENER, J. M. 2009. Can camera motions improve the perception of traveled distance in virtual environments? In *Proceedings of IEEE Virtual Reality*, 131–134.

USOH, M., ARTHUR, K., WHITTON, M. C., BASTOS, R., STEED, A., SLATER, M., AND BROOKS, JR., F. P. 1999. Walking $>$ walking-in-place $>$ flying, in virtual environments. In *Proceedings of SIGGRAPH '99*, 359–364.

WENDT, J., WHITTON, M. C., AND BROOKS, F. 2010. Gud wip: Gait-understanding-driven walking-in-place. In *Proceedings of IEEE Virtual Reality*, 51 –58.

WHITTON, M. C., COHN, J., FEASEL, J., ZIMMONS, P., RAZZAQUE, S., POULTON, S., MCLEOD, B., AND JR, F. B. 2005. Comparing VE locomotion interfaces. In *Proceedings of IEEE Virtual Reality*, 123–130.

YAN, L., ALLISON, R., AND RUSHTON, S. 2004. New simple virtual walking method - walking on the spot. In *Proceedings of the IPT Symposium*.

Influence of auditory and visual feedback for perceiving walking over bumps and holes in desktop VR

Luca Turchet* AAU Copenhagen, Denmark

Maud Marchal† INSA/INRIA Rennes, France

Anatole Lécuyer‡ INRIA Rennes, France

Stefania Serafin§ AAU Copenhagen, Denmark

Rolf Nordahl¶ AAU Copenhagen, Denmark

Abstract

In this paper, we present an experiment whose goal is to investigate the role of sound and vision in the recognition of different surface profiles in a walking scenario. Fifteen subjects participated to two within-subjects experiments where they were asked to interact with a desktop system simulating bumps, holes and flat surfaces by means of audio, visual and audio-visual cues. Results of the first experiment show that participants are able to successfully identify the surface profiles provided through the proposed audio-visual techniques. Results of a second experiment in which conflictual audio-visual stimuli were presented, reveal that for some of the proposed visual effects the visual feedback is dominant on the auditory one, while for the others the role of dominance is inverted.

CR Categories: I.3.5 [Computer Graphics]: Physics Based Modeling—; I.6.5 [Simulation and Modeling]: Model Development— [H.5.5]: Information Interfaces and Presentation (e.g. HCI)—Sound and Music Computing.

Keywords: Bump and hole, pseudo-haptic, design, audio-visual.

1 Introduction

In most current virtual reality setups, users are restricted to walk on a flat surface. Only few locomotion interfaces are able to render uneven grounds, but they have the disadvantages of being costly and cumbersome [Hollerbach et al. 2003; Iwata et al. 2001]. Recently, research has shown that it is possible to simulate the act of walking on unflat surfaces only using visual cues [Marchal et al. 2010]. These results are a development of previous research on pseudo-haptic simulation [Lécuyer et al. 2004]. The main idea of the research described in [Lécuyer et al. 2004] was to investigate whether it was possible to simulate a bump or a hole by only using visual feedback.

In the sound synthesis community, all previous research on walking sounds has focused on flat surfaces [Cook 2002; Fontana and Bresin 2003; Farnell 2007; Miner and Caudell 2005]. Recently, a sound synthesis engine able to simulate footsteps sounds on aggregate and solid surfaces has been presented [Nordahl et al. 2010; Turchet et al. 2010]. Such sound engine is based on physical models synthesis and it works both offline and in realtime. In all the different surfaces simulated, the system energy parameter of the engine is controlled by a ground reaction force (GRF), i.e., the reaction force supplied by the ground at every step. During the offline working of the engine, such force is estimated from recordings on real footsteps sounds.

In [Serafin et al. 2010], researchers used such engine in order to run a preliminary experiment whose goal was to assess the role of temporal aspects in sonically simulating the act of walking on a bump or a hole. In particular, they investigated whether the timing between heel and toe and the timing between footsteps affected perception of walking on unflat surfaces. Results showed that it is possible to simulate a bump or a hole only by using temporal information.

In this paper, we are interested in understanding whether the addition of sounds to the visual feedback enhances the pseudo-haptic feedback for the simulation of unflat surfaces. Our hypothesis is that sound plays an important role in the recognition of the surface profile and that it can be dominant on the visual modality according to the visual stimulus presented.

*e-mail:tur@media.aau.dk

†e-mail:maud.marchal@irisa.fr

‡e-mail:anatole.lecuyer@irisa.fr

§e-mail:sts@media.aau.dk

¶e-mail:rn@media.aau.dk

2 Visual and auditory simulation

2.1 Visual feedback

The visual techniques used to simulate the act of walking over bumps and holes were the same proposed in [Marchal et al. 2010]: a straightforward modification of the camera's height (H), a modification of the camera's navigation velocity (V), a modification of the camera's orientation (O), and the combination of the three effects (HOV). Our simulations used a known mathematical profile, the Gaussian one, which was used for the simulations of both holes and bumps.

Since the velocity technique was used for the generation of the auditory feedback we briefly recall it here. The velocity effect is based on the variation of the camera velocity. Thus, the camera velocity is decreased when the user is going up and increased when the user is going down. We used a different algorithm for the ascending and descending cases. The algorithm computes the ratio R_{Velocity} applied between the real user velocity and the virtual camera velocity. The camera velocity is then modified following the equation:

$$\text{Velocity}^t = \text{Velocity}^{t-1} \,.\, R^t_{\text{Velocity}} \tag{1}$$

- Ascending case:

$$R^t_{\text{Velocity}} = exp(-R_{\text{AscendingV}} \,.\, \alpha^t) \tag{2}$$

where α is the tangent angle of the Gaussian curve and $R_{\text{AscendingV}}$ is a constant.

- Descending case:
This algorithm is designed to give a run up for a while after the bump or at the beginning of the hole. At time t, the ratio is updated regarding the difference between the user height in scene at times $t-1$ and t :

$$R^t_{\text{Velocity}} = R^{t-1}_{\text{Velocity}} + \Delta_{\text{Height}} \,.\, R_{\text{DescendingV}} \tag{3}$$

where the ratio $R_{\text{DescendingV}}$ is a constant. When the subject reaches the end of the descent, his speed is at a maximum.

If he is walking in a hole, then he starts to go up and his speed value will be given by the ascending algorithm. If the subject is on a bump, he will reach the plane ground after the bump. His speed ratio R_{Velocity} will start decreasing at 0.1 unit per second, until another bump/hole is reached or the ratio is back to normal.

2.2 Auditory feedback

As concerns the sounds, the goal was to add an auditory feedback consistent with the images provided at the screen, but since we were interested just in a proof of concept no interaction between the visual and sound engines was created.
The technique adopted to render bumps and holes at auditory level has been the placement of footsteps sounds at different temporal intervals, taking as reference the work of the visual engine set with the velocity effect. In order to create a coherent mapping between the space travelled by the visual rendering along the surface and the temporal distances between steps, we chose to divide the surface profile in equal parts covering the same amount of space, and to place in those points a step (see Figure 1). Subsequently the temporal intervals to reach each point located on the curve have been calculated according to the equations utilized for the velocity effect by the visual engine (see section 2.1). Precisely, such calculations have been performed using the same parameters defining the Gaussian curve (height and sigma), but modifying the values of the R factors regulating the lows for the ascending and descending case because of the necessity of synchronizing properly the two systems.

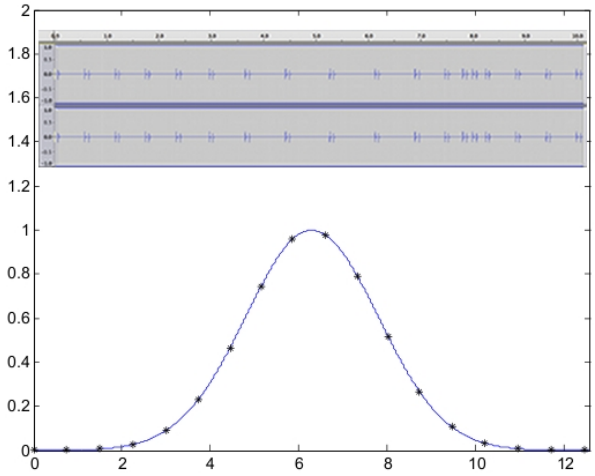


Figure 1: *Gaussian profile used for a bump simulation with the indication of the points in which the footstep sounds occur, and stereo waveform of the corresponding sound file.*

The input files for the sound synthesis engine (see section 1) resulting from such calculations, have been generated by means of MATLAB placing at different temporal patterns a single footstep sound (the same footstep was used to create the different stimuli presented to the subjects). The use of the same footstep sound 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 sound used was a recording of a real footstep on concrete. Such sound was chosen among those available in the Hollywood Edge sound effects library.[1]
For the purpose of this experiment, two types of surfaces, gravel and wood, were chosen. The reason for choosing two materials was to assess whether the surface type affected the quality of the results.

[1] www.hollywoodedge.com/

3 Experiment

We conducted two within-subjects experiments using three different kinds of stimuli: visual only, audio only, and audio visual.

1. Experiment 1: recognition of bumps, holes and flat surfaces by means of visual, audio and (coherent) audio-visual stimuli.
2. Experiment 2: recognition of bumps and holes in presence of incoherent audio-visual stimuli.

The images provided during the experiments were generated by the visual engine, while the sounds consisted of footsteps sounds generated by the offline use of the sound synthesis engine.
The goal of the first experiment was to investigate the ability of subjects to recognize the different walking surface profiles they were exposed to. One of our hypotheses was that the recognition would have improved using the bimodal stimuli rather than the single modalities alone.
The goal of the second experiment was to determine the role of dominance of the two modalities involved, in presence of incoherent stimuli. Our hypothesis was that the visual modality would have been dominant on the auditory modality for the H, O and HOV effects, while for the V effect the role of dominance would have been held by the auditory modality.

3.1 Participants

The experiments were performed by 15 participants, 12 men and 3 women, aged between 22 and 29 (mean=26.2,standard deviation=2.1662). All participants had normal or corrected vision and 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 40 minutes to complete the two experiments and to fill the questionnaire.

3.2 Task

During experiment 1 participants were given the list of three different surfaces (bump, hole, flat), presented as forced alternate choice, while during experiment 2 the possible choices were only two, i.e., bump or hole, in order to understand which modality was dominant. The task consisted of recognizing to which surface the walk corresponded after the presentation of the stimulus. Each stimulus was presented only one time before giving the answer, and when moving to the next stimulus participants could not change the answer to the previous stimuli.
After both experiments, a preference questionnaire was proposed in which participants had to grade from 1 (low appreciation) to 7 (high appreciation) three further effects (a bump provided at audio, visual and audio-visual level) according to 3 subjective criteria: efficiency of the simulation, realism of the simulation, and global appreciation of the effect. In the questionnaire we did not mentioned the words audio, visual or audio-visual, we simply called them condition 1, 2 and 3, in order to leave to the participants the task of creating the proper mapping.
At the end of the questionnaire, participants were also given the opportunity to leave an open comment on their experience interacting with the system.

3.3 Setup

All experiments were carried out in an acoustically isolated laboratory where the setup for the experiment was installed. It con-

sisted of a simple graphical user interface with which the participants were asked to interact. The interaction consisted of pressing keys on the keyboard in order to start each trial and to give the corresponding answer.
The virtual environment was a simple corridor with given dimensions (height=3.0m, length=12.6m, width=2.0m). There was a part in the center of the corridor where the height could be modified during the experiments: the user could experience either on a bump, a hole or a plane. To symbolize this variable part of the corridor, a transparent cube was represented on the ground with a height of 0.5m and a surface of 6.6mx2m, as illustrated in Figure 2. The variable height of the ground was not visible in order to exclude visual cues from the scene.

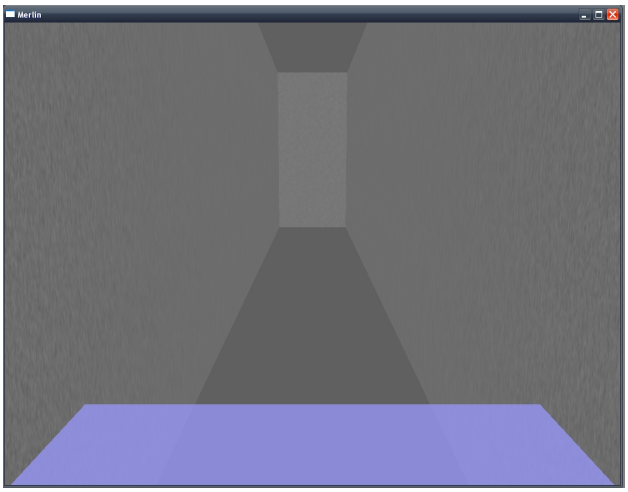


Figure 2: *A screenshot at the beginning of a trial.*

The computer utilized for the experiment had clock speed equal to 3.40 GHz, 2 GB of memory, it run the Windows XP operative system, and it was equipped with the SigmaTel Sound Card. Sounds were provided to the user by means of a set of headphones.[2]
During the experiments subjects were sitting on a chair, listening to the sounds through headphones and interacting with the interface.

3.4 Experimental plan

Participants were exposed to 105 trials in experiment 1, where 15 stimuli were repeated seven times. The stimuli consisted of bumps, holes and flat surfaces rendered by means of sounds (footsteps on two types of materials, gravel and wood), images (using the HOV technique), and the combinations of images and sounds.
During experiment 2 participants were exposed to 48 trials, where 8 audio-visual incoherent stimuli were repeated six times. In detail, only bumps ad holes were rendered: when a bump was presented at visual level a hole was provided at auditory level, and vice versa. The techniques used to render the bumps and holes at visual level were Height, Orientation, Velocity and HOV, while the sounds provided were the same presented in experiment 1 using wood as material.
The temporal durations of the trials were 10.5 seconds for bumps, 9 seconds for holes, and 9.5 seconds for the flat surfaces.

4 Results

For each participant, the percentage of correct answers was estimated for the different experimental conditions. Results are shown in tables 1, 2.
Concerning experiment 1, the first noticeable element emerging from results is the high percentages of correct answers for all the three modalities. The presentation of flat surfaces led to better performances rather than bumps and holes. The addition of audio stimuli did not produce higher percentages for the recognition of the surfaces rather than the visual only modality (which is already very high), and an in depth analysis performed by means of the Kruskal-Wallis test revealed that all the differences are not statistically significant.
Moreover no substantial differences have been found between stimuli involving footsteps sounds on wood and those involving footsteps sounds on gravel.

		Bump	Hole	Flat	%Correct answers
Visual	**Bump**	98	7	0	93.33
	Hole	7	98	0	93.33
	Flat	0	0	105	100
Audio 1	**Bump**	96	8	1	91.42
	Hole	8	97	0	92.38
	Flat	6	0	99	94.28
Audio 2	**Bump**	96	5	4	91.42
	Hole	4	98	3	93.33
	Flat	3	1	101	96.19
Audio-Visual 1	**Bump**	98	7	0	93.33
	Hole	8	97	0	92.38
	Flat	0	0	105	100
Audio-Visual 2	**Bump**	99	5	1	94.28
	Hole	8	97	0	92.38
	Flat	1	0	104	99.04

Table 1: *Confusion matrix illustrating the results for experiment 1. Legend: Visual: the HOV effect; Audio 1: sounds of footsteps on gravel; Audio 2: sounds of footsteps on wood; Audio-Visual 1: sounds of footsteps on gravel plus the HOV effect; Audio-Visual 2: sounds of footsteps on wood plus the HOV effect.*

As regards experiment 2, results show clearly that in presence of audio-visual conflicts audio is dominated by vision when H and O effects are presented. Conversely, vision is dominated by audio when V and HOV effects are presented. In particular the highest role of dominance has been found for audio stimuli respect visual stimuli provided by means of the Velocity effect.

	% Visual dominance	% Audio dominance
Bump H	85.55	14.45
Hole H	81.11	18.89
Bump O	72.22	27.78
Hole O	71.11	28.89
Bump V	8.88	91.12
Hole V	3.33	96.67
Bump HOV	30	70
Hole HOV	26.66	73.34

Table 2: *Table illustrating results for experiment 2. Legend: H: the Height effect; O: the Orientation effect; V: the Velocity effect; HOV: the combination of the three effects.*

Table 3 shows the results concerning the grades obtained by the three different modalities for each of the subjective criteria. Results clearly show better evaluations in the audio-visual modality (condition 3) rather than the audio (condition 1) and visual (condition 2) modalities alone. In particular an in depth analysis by means of t-test reveals that such differences are statistically significant for the efficiency criterium for condition 3 respect to condi-

[2]Sennheiser HD 600, http://www.sennheiser.com

tion 1 (p=0.01644), for the realism criterium for condition 3 respect condition 2 (p=0.02156), and for the appreciation criterium for condition 3 respect to condition 1 (p=0.02047) and 2 (p=0.006438). No significant differences in the criteria investigated have been found between condition 1 and 2.

	Audio		**Visual**		**Audio-Visual**	
	μ	σ	μ	σ	μ	σ
Efficiency	4.5333	1.7265	5.3333	1.3452	5.8667	0.9904
Realism	4.5333	1.5523	4.1333	1.5976	5.5333	1.5523
Appreciation	4.4667	1.7674	4.4667	1.3020	5.8667	1.3020

Table 3: *Table illustrating the answers of the questionnaire (average scores from a seven-point Likert scale, and relative standard deviation).*

5 General discussion

Results of experiment 1 show that the simulated surface profiles were recognized with high percentages in all the three modalities, and this is a confirm of the success of the techniques proposed in [Marchal et al. 2010] and [Serafin et al. 2010]. Although the addition of audio stimuli did not produce significant higher percentages rather than the visual only modality, the subjective questionnaire revealed, for the criteria investigated, a clear preference of the bimodal stimuli respect to the stimuli presented in the single modalities. In addition no significant differences were found between stimuli involving footsteps sounds on wood and those involving footsteps sounds on gravel and this is a further confirm of the results illustrated in [Serafin et al. 2010].
In case of conflictual audio-visual stimuli the relationships of dominance between modalities turned out to be very clear, according to the visual effect adopted. Our hypotheses were confirmed for all the effects with the exception of HOV for which, surprisingly, the auditory modality is dominant on the visual one. Presumably an explanation lies in the fact that the V effect was present in the HOV condition, and it is well known in the field of multimodal perception that the temporal resolution of the auditory system is better than the one of the visual system [Stein and Meredith 1993]. Our results confirm such behaviour of the perceptive system.
What emerges from an in depth analysis of results of experiment 2 is that participants on average were consistent in their choices: indeed in case of conflicts, they tended to choose always the same modality as guide (according to the visual effect presented).
Some of the subjects reported in their comments that in experiment 2 they noticed conflicts between what they heard and what they saw at screen. In particular one of them explicitly reported that in case of conflicts he was driven in his decision by the visual stimulus since not seeing the feet of the person walking in the animation, he imagined that the person would have been able to place the steps in an arbitrary way.

6 Conclusion and future work

In this paper, we described an experiment whose goal is to assess whether the combination of auditory and visual information affects the perception of walking on a hole or a bump using a desktop system. Results show that the surface profiles simulated by means of audio, visual and audio-visual stimuli were accurately recognized in all the three modalities, but the subjective questionnaire revealed a clear preference of the bimodal stimuli.
As concerns the auditory stimuli, they were created only varying temporal parameters of footsteps, such as the distance between steps. Other important aspects, such as amplitude and spectral information was not modified. These parameters are object of future investigations. 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 realism of the simulated experience. In future works we plan to integrate the visual system with the sound synthesis engine in order to have an audio-visual interactive simulation of the proposed surface profiles, both with a desktop system and an immersive VR configuration with HMD.

Acknowledgements

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. [3] This publication is supported by COST Action IC0601.

References

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

FARNELL, A. 2007. Marching onwards: procedural synthetic footsteps for video games and animation. *Proceedings of the Pure Data Convention*.

FONTANA, F., AND BRESIN, R. 2003. Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. *Proc. Colloquium on Musical Informatics*, 109–114.

HOLLERBACH, J., CHECCACCI, D., NOMA, H., YANAGIDA, Y., AND TETSUTANI, N. 2003. Simulating side slopes on locomotion interfaces using torso forces. In *Haptic Symposium*, Citeseer, 91–98.

IWATA, H., YANO, H., AND NAKAIZUMI, F. 2001. Gait master: A versatile locomotion interface for uneven virtual terrain. *vr*, 131.

LÉCUYER, A., BURKHARDT, J., AND ETIENNE, L. 2004. 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*, ACM, 246.

MARCHAL, M., LECUYER, A., CIRIO, G., BONNET, L., AND EMILY, M. 2010. Walking Up and Down in Immersive Virtual Worlds: Novel Interactive Techniques Based on Visual Feedback. In *Proceedings of IEEE Symposium on 3D User Interface*.

MINER, N., AND CAUDELL, T. 2005. Using wavelets to synthesize stochastic-based sounds for immersive virtual environments. *ACM Transactions on Applied Perception 2*, 4, 521–528.

NORDAHL, R., SERAFIN, S., AND TURCHET, L. 2010. Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In *Proceedings of IEEE Virtual Reality*.

SERAFIN, S., TURCHET, L., AND NORDAHL, R. 2010. Do you hear a bump or a hole? An experiment on temporal aspects in footsteps recognition. In *Proceedings of Digital Audio Effects Conference*.

STEIN, B. E., AND MEREDITH, M. A. 1993. *The Merging of the Senses*, 1st ed. MIT Press.

TURCHET, L., SERAFIN, S., DIMITROV, S., AND NORDAHL, R. 2010. Physically based sound synthesis and control of footsteps sounds. In *Proceedings of Digital Audio Effects Conference*.

[3] www.niwproject.eu

Walking Up and Down in Immersive Virtual Worlds: Novel Interactive Techniques Based on Visual Feedback

Maud Marchal* INSA/INRIA Rennes, France
Anatole Lécuyer * INRIA Rennes, France
Gabriel Cirio * INRIA Rennes, France
Laurent Bonnet * INRIA Rennes, France
Mathieu Emily † University Rennes 2, France

ABSTRACT

We introduce novel interactive techniques to simulate the sensation of walking up and down in immersive virtual worlds based on visual feedback. Our method consists in modifying the motion of the virtual subjective camera while the user is really walking in an immersive virtual environment. The modification of the virtual viewpoint is a function of the variations in the height of the virtual ground. Three effects are proposed: (1) a straightforward modification of the camera's height, (2) a modification of the camera's navigation velocity, (3) a modification of the camera's orientation. They were tested in an immersive virtual reality setup in which the user is really walking. A Desktop configuration where the user is seated and controls input devices was also tested and compared to the real walking configuration. Experimental results show that our visual techniques are very efficient for the simulation of two canonical shapes: bumps and holes located on the ground. Interestingly, a strong "orientation-height illusion" is found, as changes in pitch viewing orientation produce perception of height changes (although camera's height remains strictly the same in this case). Our visual effects could be applied in various virtual reality applications such as urban or architectural project reviews or training, as well as in videogames, in order to provide the sensation of walking on uneven grounds.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—evaluation/methodology, haptic I/O, input devices and strategies, interaction styles, user-centered design; H.5.1 [Information Interfaces and Presentation]: Multimedia Informations Systems—evaluation/methodology H.1.2 [Information Systems]: User/Machine Systems—human factors, human information processing

1 INTRODUCTION

Virtual Reality technologies immerse users inside a 3D synthetic world simulated in real-time by a computer. In such a virtual world, the user is given the possibility to manipulate virtual objects, and/or walk and explore virtual scenes.

Surprisingly, most current virtual reality (VR) setups restrict users to walk on flat workspaces. Whilst this might seem appropriate for walking inside virtual buildings or virtual streets, which are often flat, it becomes rapidly counter-immersive and unappropriate for any outdoor walking experience, such as when exploring a natural landscape. A main reason lies in the current difficulty to simulate, in the physical workspace, uneven grounds by means of mechanically actuated interfaces. As for today, few achievements have been reported on the design of devices that can render uneven grounds such as locomotion interfaces [2, 3, 5]. These interfaces remain costly, cumbersome and difficult to spread at the moment.

In videogames, the user is generally seated and interacts through input devices. Mouse and keyboards are often used to control avatar and walk in the 3D virtual world in "first-person view". In this case, a technique commonly employed when navigating on uneven grounds consists in constraining the motion of the virtual camera to follow the terrain. The camera stays always at the same height, according to ground level. This results in a continuous change in height of the view point, as if the user was "sliding" on the virtual ground.

In this paper, we study the use of such kind of visual techniques to simulate uneven terrains and provide the sensation of walking up and down in an immersive virtual environment (VE) while walking on flat real ground. The proposed techniques use only visual feedback and consist in modifying the motion of the camera as function of the relief of virtual grounds. Three techniques are proposed: (1) a modification of the camera's height, (2) a modification of the camera's advance speed, (3) a modification of the camera's orientation. These techniques are implemented and tested in two different configurations. The first one is an immersive virtual reality setup in which the user is really walking while visual feedback of an Head Mounted Display (HMD) is automatically modified by superimposing one or more of the aforementionned visual effects. The second configuration is a Desktop setup in which user is seated and controls the 3D walking with mouse/keyboard such as in videogames. We use these two setups to evaluate the influence of the different visual effects (and their combination) within various applications. The Desktop configuration can be considered as a control condition, in order to compare the use of visual techniques in an immersive situation (i.e., when the user is really walking) with a more classical desktop situation.

The remainder of the paper is organized as follows. First, the paper begins with a description of related work in the field of simulation of walking in virtual environments. Then, we describe the concept of our visual effects and how they were implemented for the simulation of two simple shapes: a bump and a hole. In the following parts, we describe the results of the experiment conducted to evaluate the efficiency of the techniques for simulating uneven terrains. The paper ends with a conclusion and a description of potential perspectives and applications.

2 RELATED WORK

As of today, the simulation of the physical sensation of walking on uneven grounds has mainly been proposed through locomotion interfaces. When using these locomotion devices, the user is self-propulsed through a repetitive gait, while his motion is compensated with an inverse motion produced by the device. Hence, the interface directly controls the position of the user in the virtual world. Most of the devices try to enable natural walking. In spite of the fact that there is a significant amount of locomotion devices specifically designed for VR systems and exploration of virtual worlds, most of them can only enable walking on flat surfaces, without ob-

*e-mail: {mmarchal,alecuyer,gcirio,lbonnet}@irisa.fr
†e-mail: mathieu.emily@uhb.fr

stacles. However, the action of walking over uneven terrain and cluttered environments is fundamental in our daily life (e.g. walking up and down the stairs), and critical on some occasions, such as when exploring outdoor environments. To date, only a few systems are capable of simulating human walking on non-flat ground.

The Sarcos Treadport [4], a treadmill with a mechanical tether attached to the back of the user, is an example of an attempt to provide a feeling of climbing slopes. Originally, the mechanical tether was used to compensate missing inertial forces and to simulate obstacles in the virtual path by applying forces on the user's torso. The concept was then extended so that the tether could also render the forces required to simulate a slope [2, 3]. A force on the opposite direction of motion was used when simulating walking uphill, with a magnitude equal to the horizontal component of the force in the real world case, and, analogously, a force was applied on the direction of motion when going downhill. Simulation of side slopes was also possible when applying lateral forces.

Leaving the kinesthetic simulation and entering the haptic realm, the ATLAS [9] treadmill, mounted on an actuated spherical joint, was able to provide slopes by allowing the pitching and rolling of the platform. With a different approach, the Ground Surface Simulator GSS [10] was able to simulate uneven terrain through a linear treadmill with a deformable belt. Six long platforms could locally raise the belt, allowing the display of small bumps up to 5° in slope. The Sarcos Biport and the GaitMaster [5], both made of foot motion platforms, could simulate uneven terrains but not inclined floors.

While these devices offer uneven terrain rendering to some extent, they all suffer from common limitations that restrict their widespread use, such as their huge size and weight, their cost or their lack of accuracy and degrees of freedom. Therefore, they have not yet been widely adopted outside the laboratory. Other smaller, less complex and more affordable locomotion devices exist that enable locomotion following the same motion compensation principle. Foot-wearable devices like the Powered-Shoes [6] and, more recently, the Gait Enhancing Mobile Shoe [1], compensate the user's motion without being attached to a bulky structure. However, they cannot render slopes or any kind of uneven terrain.

In order to simulate haptic sensations without haptic interfaces, other solutions have thus been proposed such as sensory substitution and pseudo-haptic feedback. Pseudo-haptic feedback was notably studied through the modification of the speed of a mouse cursor according to the "height" of a texture [7]. As the mouse cursor explored an image representing a top view of a texture, an acceleration (or deceleration) of the cursor indicated a negative (or positive) slope of the texture. Experimental evaluations showed that participants could successfully identify macroscopic textures such as bumps and holes, by simply using the variations of the motion of the cursor. Approaches using camera movements were also proposed in order to modify the locomotion sensations in virtual environments. For instance, Steinicke et al. [11] used redirected walking techniques in order to modify the trajectory orientation of the user in the real environement. Lécuyer et al. [8] used camera oscillations to improve the sensation of walking in VE. However, these approaches are not dedicated to walking sensations on uneven grounds.

In some "first-person view" videogames, the camera velocity is progressively scaled up or down whether the avatar is going up or down a hill, providing a slope information. This effect could be considered as a straight transposition of the aforementioned pseudo-haptic texture. However, to the authors' best knowledge there has been no study of the influence of such visual effects on the user's perception of heights and slopes in virtual environments, and these effects have never been implemented within an immersive VR setup when physically walking.

3 Novel Interactive Techniques Based on Visual Feedback

3.1 Concept of the Interactive Techniques

The objective of the interactive techniques is to reproduce the sensation of walking on an uneven ground without the use of any haptic or locomotion interface. The main idea consists in modifying the motion of the subjective camera while the user is walking in the VE. The concept is to control the camera position and orientation, depending on the uneven virtual terrain displayed either on the screen or on a HMD. The camera motion is adjusted as a function of the simulated height of the terrain on which the user is walking. The variations of the camera motion are used here to transpose the perception of climbing or descending a slope.

Three different types of modifications to the camera motion have been studied: height variation, orientation variation and velocity variation. The amplitude of the different effects is computed using the height information of the 3D virtual environment. Thus, the techniques can be used to simulate any uneven 3D terrain, assuming that we know its height map. The implemented algorithm computes an iterative solution (depending on the user motion) for the modification of the camera motion. When the user is moving in the VE, a theoretical displacement is measured and the amount of the camera motion is computed using this measurement. Then, the new position of the user is computed and transmitted to the camera position and/or orientation. The visual techniques described here recall the ones used in videogames. However, unlike most gaming situations, our intention is to use them when the user is actually walking, i.e. superimposed to the visual feedback of the virtual scene.

3.2 Implementation

The three different effects (Height, Orientation, Velocity) are displayed in Figure 1. The combination of the three effects was also implemented.

3.2.1 Height Variation

The height effect consists in modifying the subjective camera height with a translation along the vertical axis. This effect allows the user to move parallel to the ground surface during his navigation in the virtual world. This approach is already used in a lot of demos and videogames. The height variates following the equation:

$$\text{Height}^t = \text{Height}^{t-1} + \Delta_{\text{Height}} \,.\, R_{\text{Height}} \tag{1}$$

where Height^t represents the camera height value at time t where the image is updated. R_{Height} is the ratio applied to the difference of height Δ_{Height} between times $t-1$ and t. In our implementation, we introduced ratios in order to parametrize the amplitude of the different effects. In our experiment, we chose $R_{\text{Height}} = 0.5$.

3.2.2 Orientation Variation

The orientation effect consists in applying a variation in pitch angle to the subjective camera in order to look down when descending and up when ascending. This effect is supposed to mimick postural changes when walking on uneven grounds: the subject would have to compensate the orientation modification and adopt a new postural position. The camera angle at time t, Angle^t, is proportional to the tangent angle of the Gaussian curve α^t where the user is at time t:

$$\text{Angle}^t = \alpha^t \,.\, R_{\text{Orientation}} \tag{2}$$

where $R_{\text{Orientation}}$ is the ratio applied to the angle. In our experiment, we chose $R_{\text{Orientation}} = 0.5$.

3.2.3 Velocity Variation

The velocity effect is based on the variation of the camera velocity. In a real environment, a subject is generally going slower on ascending slopes, and faster on descending slopes. We try to transpose this effect in our experiment by modifying the camera motion when the user is walking in a VE. Thus, the camera velocity is decreased when the user is going up and increased when the user is going down. This effect could be considered as a straightforward transposition of the pseudo-haptic textures effect [7], adapted here to the simulation of walking on uneven reliefs at a first-person view. We used a different algorithm for the ascending and descending cases. The algorithms compute the ratio R_{Velocity} applied between the real user velocity and the virtual camera velocity. The camera velocity is then modified following the equation:

$$\text{Velocity}^t = \text{Velocity}^{t-1} \,.\, R^t_{\text{Velocity}} \tag{3}$$

- Ascending case:

$$R^t_{\text{Velocity}} = exp(-R_{\text{AscendingV}} \,.\, \alpha^t) \tag{4}$$

  where α is the tangent angle of the Gaussian curve and $R_{\text{AscendingV}} = 0.1$ in our experiments.

- Descending case:
  This algorithm is designed to give a run up for a while after the bump or at the beginning of the hole. At time t, the ratio is updated regarding the difference between the user height in scene at times $t-1$ and t :

$$R^t_{\text{Velocity}} = R^{t-1}_{\text{Velocity}} + \Delta_{\text{Height}} \,.\, R_{\text{DescendingV}} \tag{5}$$

  where the ratio $R_{\text{DescendingV}}$ is equal to 2.0 in our experiments. When the subject reaches the end of the descent, his speed is at a maximum. If he is walking in a hole, then he starts to go up and his speed value will be given by the ascending algorithm. If the subject is on a bump, he will reach the plane ground after the bump. His speed ratio R_{Velocity} will start decreasing at 0.1 unit per second, until another bump/hole is reached or the ratio is back to normal.

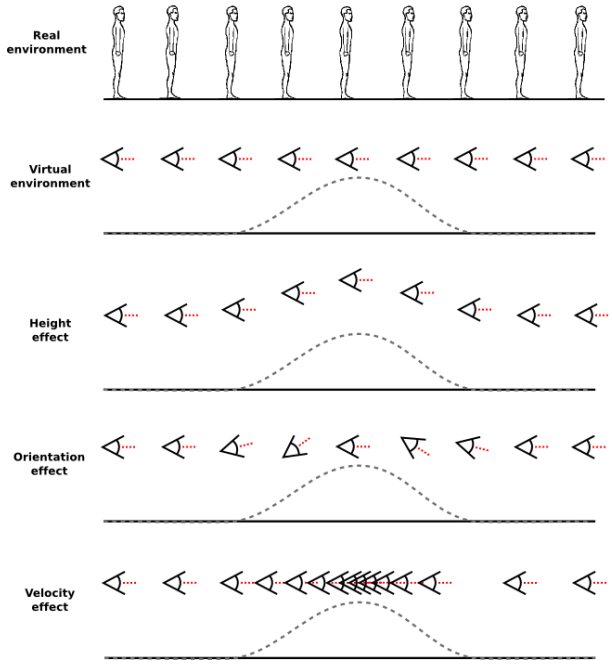


Figure 1: Principle of the three different effects: the user is walking on a flat environment while the VE is composed of a bump. The camera motion is modified in three different ways: height variation (the camera moves with a constant orientation and a constant distance to the virtual ground surface), orientation variation (the camera is oriented following the curvature of the slope), velocity variation (the camera velocity decreases as the user is going up a virtual bump and increases as the user is going down with a run up at the end of the bump).

3.3 Simulating Bumps and Holes

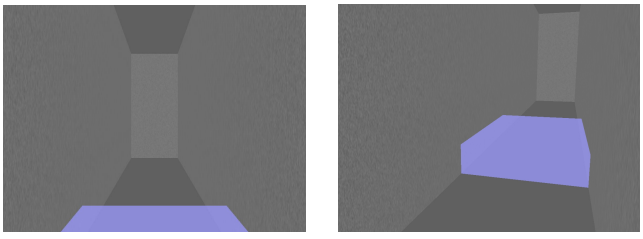
Our visual techniques were used to simulate two classical shapes: a bump and a hole. Our simulations used a known mathematical profile: a Gaussian profile, which defines the height maps of the shapes during the evaluations. It corresponds to a mathematical distribution of heights along a line perpendicular to the walking path. The same profile was used for the simulations of holes and bumps.

4 Evaluation: Efficiency of Visual Effects to Simulate Bumps and Holes

The investigation of the perception of 3D holes and bumps while walking in a VE was performed using an experimental protocol consisting of a comparison of the different effects. The experiments were conducted using 3D virtual environments displayed either on a HMD or on a screen.

4.1 Virtual Environment

The virtual environment is a simple corridor with given dimensions (height=3.0m, length=19.0m, width=2.0m). There is a part in the center of the corridor where the height can be modified during the experiments: the user can walk either on a bump, a hole or a plane. To symbolize this variable part of the corridor, a transparent cube is represented on the ground with a height of $0.5m$ and a surface of $3m \times 2m$, as illustrated in Figure 2. The variable height of the ground is not visible in order to exclude visual cues from the scene.



(a) (b)

Figure 2: (a) Description of the virtual environment composed of a corridor; (b) a transparent blue cube is placed in the center in order to represent to the participant the location of the height modifications on the ground surface.

4.2 Population and Visual Conditions

4.2.1 Group 1: Immersive VR Configuration with HMD

Twelve participants (4 females and 8 males) aged from 21 to 59 (mean=28.7, SD=11.0) were in Group 1 and exposed to a first visual condition. One of them was left-handed, and none of them had known perception disorders. They were all naïve to the purpose of the experiment.

For this group, the experiments were conducted in an immersive room large enough to walk straight forward almost 6 meters. We used the eMagin Z800 Head Mounted Display as display device, at 60 Hz and with stereoscopy enabled. The user was wearing an opaque fabric on top of the HMD to hide the surrounding real world. An unique wire was transmitting the data, allowing the user to move freely during the experiments, as illustrated in Figure 3. The user's head was tracked by an ART ARTtrack2 infrared tracking system with 9 surrounding cameras for tracking the entire path of the experiment (corresponding to the virtual corridor). The available tracking space dimensions were: height=2.5m, length=6m, width=3m.

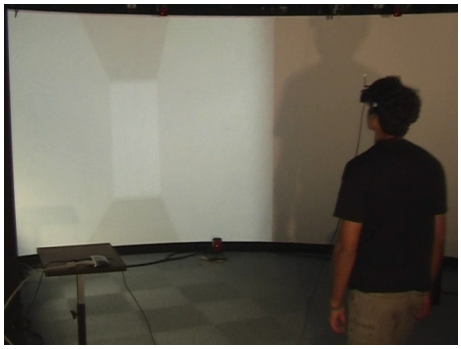


Figure 3: Configuration of the immersive room for the experiments using the HMD. The scene displayed on the HMD is also displayed on the screen in this picture as an illustration of what the user can see during the experiments.

4.2.2 Group 2: Desktop Configuration with Monitor Screen

Twelve participants (12 males) aged from 21 to 59 (mean=27.8, SD=6.1) were in Group 2 and exposed to a second visual condition. The twelve participants were all different from participants of Group 1. One of them was left-handed, and none of them had known perception disorders. They were all naïve to the purpose of the experiment and they were different from the experiments conducted with a HMD.

For this group, the experiments were conducted with a PC, by using a classical keyboard for the answers. There was no stereoscopic effect. This second group can be considered as a control population, to compare the use of visual techniques in an immersive configuration, i.e. when the user is really walking, with the desktop case. The different fields of view for the rendering process for Desktop and HMD configurations were identical.

4.3 Experimental Plan

In the experiment, our goal was to evaluate and compare the three different effects (Height, Orientation and Velocity) for the simulation of two canonical shapes: bumps and holes located on the surface ground of an immersive virtual environment. We also evaluated a fourth effect which is a combination of the three effects.

We used:

- three different *profiles*: Bump, Hole and Plane;
- two different types of walking locomotion: Forward and Backward *movements*;
- four visual *effects*: Height (H), Orientation (0), Velocity (V) and a combination of the three effects (HOV).

The experimental plan was made of the combinations [Profile x Movements] x 9 trials, for each effect (54 trials per effect). The subject alternates Forward and Backward movements, within a random sequence of [Bump, Hole, Plane] x [Forward, Backward] = 6 combinations. We performed between-subject design. The 4 series (one for each effect) are presented using a Latin square and a defined sequence [H-O-V-HOV], counterbalanced with 4 sub-groups. The 12 participants of each Group (Group 1 with HMD and 2 with PC) were thus equally divided into 4 sub-groups of 3 people each.

The motivation for testing backward movements relies on the fact that gait postures of human bodies are generally different when moving forward or backward on a slope. Thus, our hypothesis was that backward movements could potentially lead to different physical sensations for our visual effects.

4.4 Procedure

The experiment consists of 216 trials per participant (54 per effect). The subject has to go forward and then backward in the virtual corridor. At the end of each movement (either forward or backward), a black screen appears (either on the HMD or on the screen) and the participant can give his answer concerning the identified shape (hole, bump, or plane).

4.5 Results

For each participant, the percentage of correct answers was estimated for the different experimental conditions. An ANOVA on the 4 different effects was conducted on the percentage of correct answers. A post-hoc analysis using Tukey's procedure was then performed. ANOVA were achieved separately for the two experimental configurations (HMD and Desktop) and by differenciating Forward and Backward movements. The results concerning the different effects are represented in Figure 4 for the HMD and the Desktop configurations. Results concerning Forward and Backward movements are distinguished for each group, as they gave different values. The order in the sequence had no significative effect on the results.

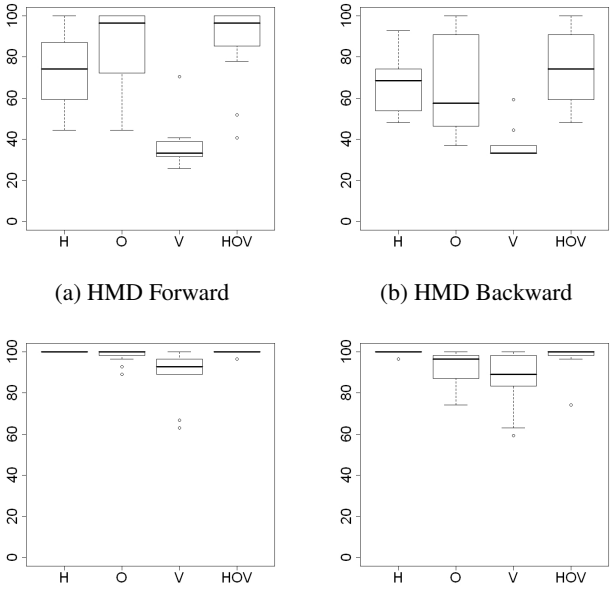
In the following paragraph, we present the results obtained for the four different configurations as a combination of HMD and Desktop groups, Forward and Backward movements.

The ANOVA accounting for the four different effects revealed a significant dependency between the effect and the probability of giving a correct answer for all the configurations.

For the Forward movement performed with the HMD configuration, the ANOVA performed between the four different effects revealed significant results for the effect ($F(3,11) = 19.447$, $p < 0.0001$). Post-hoc analysis showed that the percentage of correct responses in the Height condition ($M = 73\%$) was significantly higher than in the Velocity condition ($M = 37\%$), adjusted p-value < 0.001; the percentage of correct responses in the Orientation condition ($M = 85\%$) was significantly higher than in the Velocity condition, adjusted p-value < 0.001; and the percentage of correct responses in the HOV condition ($M = 87\%$) was significantly higher than in the Velocity condition, adjusted p-value < 0.001. The other pairs of effects did not give any significant adjusted p-values.

For the Backward movement performed with the HMD configuration, the ANOVA performed between the four different effects revealed significant results for the effect ($F(3,11) = 11.646$,

$p < 0.0001$). Post-hoc analysis revealed also significant differences: the percentage of correct responses in the Height condition ($M = 67\%$) was significantly higher than in the Velocity condition ($M = 37\%$), adjusted p-value $= 0.0004$; the percentage of correct responses in the Orientation condition ($M = 65\%$) was significantly higher than in the Velocity condition, adjusted p-value $= 0.0010$; and the percentage of correct responses in the HOV condition ($M = 75\%$) was significantly higher than in the Velocity condition, adjusted p-value < 0.0001. The other pairs of effects did not give any significant adjusted p-values.



(a) HMD Forward (b) HMD Backward

(c) Desktop Forward (d) Desktop Backward

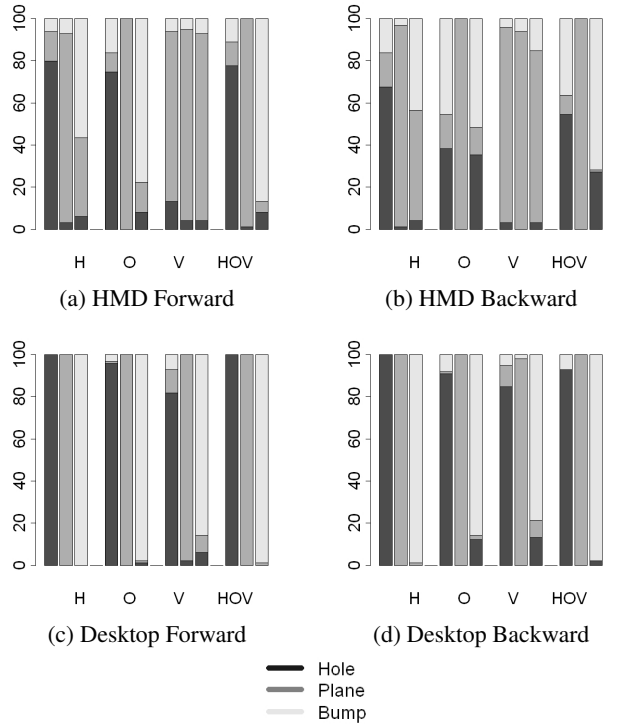
Figure 4: Results: Percentage of correct answers for HMD configuration ((a) and (b) boxplots) or Desktop configuration ((c) and (d) boxplots). (a) and (c) represent the results for Forward movements, (b) and (d) represent the results for Backward movements. The 4 different effects are represented on each picture: Height (H), Orientation (O), Velocity (V) and the combination of the three previous effects (HOV). Each boxplot is delimited by the quartile (25% quantile and 75% quantile) of the distribution of the effect over the individuals. The median is also represented for each effect.

For the Forward movement performed with the Desktop configuration, the ANOVA performed between the four different effects revealed significant results for the effect ($F(3,11) = 7.77$, $p = 0.0003$). Post-hoc analysis showed that the percentages of correct responses are significantly higher in Height condition ($M = 100\%$), HOV condition ($M = 99\%$), Orientation condition ($M = 98\%$) than in the Velocity condition ($M = 89\%$). Adjusted p-values are respectively 0.0008 (Height vs Velocity), 0.0012 (HOV vs Velocity) and 0.0072 (Orientation vs Velocity). The other pairs of effects did not give any significant adjusted p-values.

For the Backward movement performed with the Desktop configuration, the ANOVA revealed significant results for the effect ($F(3,11) = 3.8508$, $p = 0.016$). Post-hoc analysis revealed that the percentage of correct responses are significantly higher in the Height condition ($M = 99\%$) than in the Velocity condition ($M = 89\%$) with an adjusted p-value equals to 0.015. The other pairs of effects did not give any significant adjusted p-values.

At first glance, regarding the percentage of correct answers for the different effects, it seems that the sensation of bumps and holes was perceived among the participants. HMD and Desktop configurations give relatively different results. The Velocity effect with Desktop configuration gives namely higher percentages of correct responses compared to HMD configuration. Forward and Backward movements are distinguished for both configurations. Experiments conducted with a HMD and Backward movement globally obtained lower results compared to the experiments conducted with the same experimental configuration but with Forward movements. We can particularly notice the lower results for the two effects containing the Orientation effect (O and HOV) for Backward movements.

For HMD group, we can also notice the presence of two individuals (represented by individual dots on Figure 4.a and 4.b). These two individuals have obtained lower percentages of correct answers for the HOV effect and higher percentages for the Velocity effect, compared to the rest of the population, and could be considered as outliers.



(a) HMD Forward (b) HMD Backward

(c) Desktop Forward (d) Desktop Backward

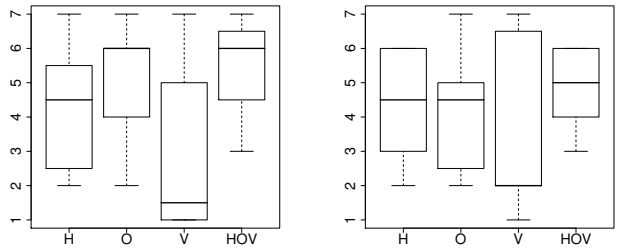
Figure 5: Results: Percentage of correct answers. Results are given for the 4 effects and the 3 different shapes (Hole, Plane and Bump in this order). The percentage of answers is decomposed for each shape, displaying additionally the incorrect shapes identified for each shape. The Forward and Backward movements are distinguished.

We conducted also an analysis concerning the percentage of correct answers for the different shapes identified (i.e. Hole, Plane and Bump). The results are reported in Figure 5 for HMD and Desktop configurations, and detailed for Forward and Backward movements. Experiments conducted with HMD contain a higher number of incorrect answers: the higher number of incorrect answers for each effect are Plane shape for Height effect, Bump/Hole shape for Orientation and HOV effects. Thus, the Orientation effect seems to have an influence on the shape perception. For Velocity effect with HMD configuration, almost all answers are incorrect: Plane shape solution is almost always chosen, meaning that holes and bumps are almost never detected. On the opposite side, shapes with Velocity effect on Desktop configuration are well recognized. Thus,

Velocity effect in an immersive situation leads to significantly different results, as observed also in Figure 4. Concerning backward movements with HMD configuration, we can notice that the percentage of incorrect answers is higher than for forward movements, for Height, Orientation and HOV effects.

4.6 Subjective Questionnaire

After both experiments, a preference questionnaire was proposed in which participants had to grade from 1 (low appreciation) to 7 (high appreciation) the four different effects (H, O, V, HOV) according to 4 subjective criteria: easiness of judgment, realism, cybersickness and global appreciation. Figures 6 and 7 show the results concerning the grades obtained by the four different effects for each of the subjective criteria, for HMD and Desktop configurations.



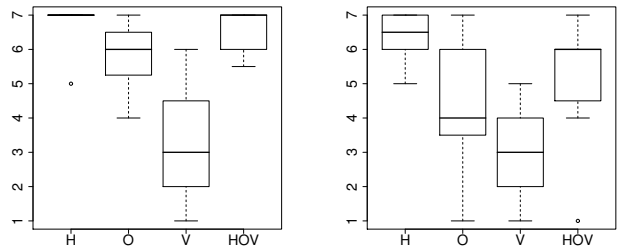
(a) Easiness of Judgment (b) Realism

(c) Cybersickness (d) Global appreciation

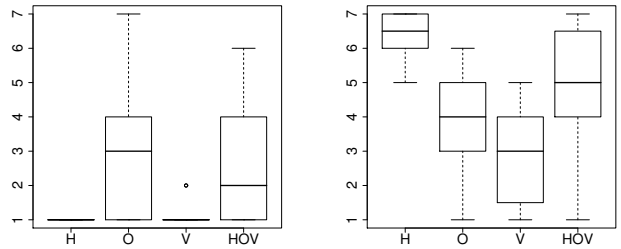
Figure 6: Results for subjective ratings for the four effects for HMD experiments: each boxplot is delimited by the quartile (25% quantile and 75% quantile) of the distribution of the effect over the individuals. The median is also represented for each effect. The 4 different effects are represented on each picture: Height (H), Orientation (O), Velocity (V) and the combination of the three effects (HOV).

As the number of modalities is high (7 grades), we assume data to be normally distributed and perform an ANOVA test to compare the four types of algorithm. Thus, an ANOVA on the 4 different effects was conducted on the grade of each criterion. ANOVA were performed separately for the two experimental configurations and were followed by a post-hoc analysis using Tukey's procedure. The ANOVA accounting for the four different effects revealed no significant dependency between the effect and the grading value for Realism ($F(3,11) = 1.30$, $p = 0.28$) and Cybersickness ($F(3,11) = 0.17$, $p = 0.91$) for HMD experiments.

Concerning global appreciation, the ANOVA performed between the four different effects revealed significant results for the Effect for both configurations ($F(3,11) = 13.27$, $p < 0.0001$ for Desktop configuration, $F(3,11) = 6.9$, $p < 0.0001$ for HMD configuration). The HOV effect obtains the best global appreciation for HMD experiments, followed by Orientation and Height effects. Post-hoc analysis shows that Velocity is significantly less apreciated than HOV (adjusted p-value 0.0003) and Orientation (adjusted p-value



(a) Easiness of Judgment (b) Realism



(c) Cybersickness (d) Global appreciation

Figure 7: Results for subjective ratings about the different criteria for the four effects for Desktop experiments: each boxplot is delimited by the quartile (25% quantile and 75% quantile) of the distribution of the effect over the individuals. The median is also represented for each effect. The 4 different effects are represented on each picture: Height (H), Orientation (O), Velocity (V) and the combination of the three effects (HOV).

0.013). The other pairs of effects did not give any significant adjusted p-values, which argues in favor of a difference between the Velocity effect and the three other effects.

Height effect obtains the best global appreciation for Desktop experiments, followed by HOV effect. Post-hoc analysis revealed that the Height condition is significantly more appreciated than Orientation (adjusted p-value 0.0008) and Velocity (adjusted p-value < 0.0001). The other pairs of effects did not give any significant adjusted p-values. Indeed, Height condition was less appreciated in HMD experiment. On the contrary, the other techniques were better accepted and fairly evaluated for HMD configuration.

Concerning Easiness of Judgment criterion, the ANOVA performed between the four different effects revealed significant results for the Effect for both configurations ($F(3,11) = 30.1$, $p < 0.0001$ for Desktop configuration, $F(3,11) = 4.53$, $p = 0.007$ for HMD configuration). Post-hoc analysis for HMD configuration gave similar results as obtained for global appreciation. Velocity effect is significantly less apreciated than HOV effect (adjusted p-value $= 0.0067$) and Orientation effect (adjusted p-value $= 0.036$). The other pairs of effects did not give any significant adjusted p-values, which argues in favor of a difference between the Velocity effect and the three other effects. However post-hoc analysis for Desktop configuration is slightly different than the one obtained for global appreciation. For Easiness of Judgment criterion, Velocity is significantly less apreciated than the three other conditions: Height effect (adjusted p-value < 0.0001), HOV effect (adjusted p-value < 0.0001) and Orientation effect (adjusted p-value < 0.0001).

Concerning Cybersickness and Realism criteria, the ANOVA performed between the four different effects revealed significant results for the Effect only for Desktop configurations ($F(3,11) = 6.56$, $p < 0.0001$ for Cybersickness, $F(3,11) = 14.6$, $p < 0.0001$

for Realism). Post-hoc analysis for Realism criterion revealed that Height effect is found more realistic than Orientation effect (adjusted p-value $= 0.0033$) and Velocity effect (adjusted p-value < 0.0001). Furthermore HOV effect is found more realistic than Velocity (adjusted p-value $= 0.0003$). The other pairs of effects did not give any significant adjusted p-values. Post-hoc analysis for Cybersickness criterion showed significant adjusted p-values between Orientation and Height effects (adjusted p-value $= 0.0025$) and between Orientation and Velocity effects (adjusted p-value=0.0058). These results argue in favor of the exaggerated perceptions of the Orientation for Desktop experiments due to parameter values (the different ratios explained in section 3), which seem to play a key role in the subjective appreciation of the participants.

Participants were also asked to evaluate the height of the shapes (Bump or Holes) of the experiments. Means and standard deviations of the participant answers are given in Table 1 for HMD configuration and in Table 2 for Desktop configuration. The real height of the Bump/Hole was 1.0 meter, with a ratio coefficient equal to 0.5 for Height, Orientation and Combination effects.

	H	O	V	HOV
Bump	0.32 (0.28)	0.79 (0.56)	0.06 (0.15)	0.68 (0.45)
Hole	-0.3 (0.32)	-0.77 (0.55)	-0.05 (0.12)	-0.59 (0.40)

Table 1: Means and standard deviations (in brackets) of the heights (in meters) given by the participants to holes and bumps for HMD configuration experiments. The four different effects (H, O, V, HOV) are distinguished.

	H	O	V	HOV
Bump	0.59 (0.52)	0.97 (0.54)	0.42 (0.46)	1.32 (1.12)
Hole	-0.58 (0.52)	-0.99 (0.53)	-0.42 (0.46)	-1.28 (1.10)

Table 2: Means and standard deviations (in brackets) of the heights (in meters) given by the participants to holes and bumps for Desktop configuration experiments. The four different effects (H, O, V, HOV) are distinguished.

The estimated values are globally lower for HMD configuration. Interestingly, participants gave the Orientation effect the highest height for HMD configuration. The Orientation effect is always evaluated with an over-estimation of the correct height, although there is no variation in the camera height. The height values are under-estimated for Height effect for HMD configuration but slightly over-estimated for Desktop configuration. For Desktop configuration, the highest height is given to the experiments conducted with the HOV effect. For HMD configuration, the Velocity effect conducts to a height value near to zero, but it is not the case for the Desktop configuration where the height change perception is better, as already observed in the results in Figures 4 and 5.

5 General Discussion

At first glance, results show that slope presence was identified for most effects in the immersive configuration. The slope appreciation greatly varies according to the experimental setup and, in a lesser way, according to the motion direction.

The Height and the Orientation effects yielded highly positive results in an immersive configuration. Users clearly felt a change in height, and could distinguish in most of the cases whether it was a bump or a hole. In the Forward case for the immersive setup, the Orientation effect shows better results than the Height effect. Although there was no change in height, users were able to perceive it more accurately than in trials where the height itself changed. The success and the accuracy of the Orientation effect was confirmed with the subjective questionnaire since users had no trouble in drawing the outline of the shapes they encountered during the experiments. When estimating the height of these shapes, results were not so far from real heights. The sum of the effects did not give better results than one effect taken alone for HMD configuration. However, the HOV effect was more appreciated in the subjective questionnaire for the immersive situation.

On the other hand, Height technique was less appreciated in HMD experiments. A possible explanation might be that users, particularly gamers and people familiar with navigation in VR, are used to see camera height variations when navigating in virtual uneven terrains in desktop environments. The Height effect is used in every desktop simulation involving slopes and landscapes. They have rarely or never been exposed to the other effects. Hence, they find the Height effect more natural and more appreciated. However, these same users have obviously spent less time in immersive simulations, and might not be used to the conditions of an immersive setup. Consequently, they are less trained for the Height effect under these conditions and did not perceive any "real inclines/declines" sensations. Hence, the other techniques were better accepted and fairly evaluated for the immersive configuration.

When immersed in a VE with an HMD setup, it appears that users do not well perceive change in height when subject to the Velocity effect. Hence, one could think that the direct transposition of the pseudo-haptic effect from the 2D to the 3D realm does not provide the expected visual cues in an immersive configuration. Thus, when using the non-immersive Desktop setup, the same Velocity effect, although not as efficient as the others, yields much better results than in the immersive setup, reaching up to an almost perfect score. A possible explanation for this behaviour might be related to the optical flow of the virtual scene. In the immersive setup, the walls of the corridor were situated at the sides of the user's field of view due to the use of an HMD, while in the Desktop setup the entire display was largely contained in the field of view. Hence, the optical flow visible on the walls had a greater effect in the Desktop setup. The Velocity effect might have better results with a different virtual scene and/or different HMD optics.

As expected, backward and forward movements led to different results. The different shapes were less identified for backward movements in an immersive configuration. A reason for this difference between the two directions of walking might be the tuning of the different parameters, especially when Orientation effect is used.

Parameters of the different effects seem to play a key role for letting the participant identify a bump or a hole. We chose to tune the parameters of our different effects based on the HMD configuration. This choice led us to non-optimized parameters for Desktop configuration, and could explain some differences in the results and subjective questionnaires (e.g. for the Orientation effect which parameter was probably too high). The choice of our parameter values was arbitrary but could be based on more sophisticated models or on the application objectives. Although a very simple and straightforward implementation of the orientation motion was enough to achieve a good performance with the Orientation effect, other models closer to real life motions and gaits might improve these results. The physically-based model of an avatar representing the user in the virtual world, coupled to the motion of the user in the real world, might provide changes in head orientation, and hence in camera orientation, that are closer to what the user might expect. The use of real data on head orientation of users walking up and down on slopes could also be an alternative solution to tune the camera parameters. In future work, it could also be interesting to vary the steepness of the bump/hole. To conclude, a higher degree of realism, something actually criticized by many users, as shown in the subjective questionnaire results, could improve the efficiency of the Orientation effect.

6 Conclusion and Perspectives

In this paper, we introduced novel interactive techniques to simulate the sensation of walking up and down in immersive virtual worlds based on visual feedback. Our method consists in modifying the motion of the virtual subjective camera as function of the variations in the height of the ground. This method has been widely used for deskop applications for videogames for example but never explored for providing real relief sensation when walking in an immersive virtual environment. In this paper, three effects were proposed: (1) a straightforward modification of the camera's height, (2) a modification of the camera's navigation velocity, (3) a modification of the camera's orientation. They were tested in an immersive virtual reality setup in which the user is really walking. A Desktop configuration where the user is seated and controls input devices was also used to compare the results.

The experiments were conducted to evaluate the influence of our visual techniques for the perception of simple and canonical shapes: virtual bumps and holes located on the ground. Experiments showed that changes in height and orientation of the camera are very efficient effects in an immersive configuration. The speed effect seems to be less perceived. Interestingly, in the immersive configuration, the consistent combination of all visual effects together led to the best results (although this result was not found significant) and was subjectively preferred by the participants. Experiments suggest also a strong perception of height changes caused by the orientation effect (although camera's height remains strictly the same in this case). This is confirmed by subjective questionnaire in which participants estimated a higher amplitude for bumps and holes simulated with orientation technique. This "orientation-height illusion" opens challenging questions in terms of human perception and challenges our interpretations.

Taken together, our results suggest that our visual techniques could be applied in an immersive virtual environment to simulate the sensation of walking on uneven surfaces. Our techniques could be used in various applications of virtual reality such as for urban and architectural reviewstraining, as well as in videogames in an immersive configuration.

Acknowledgements

This work was supported by the European community under FP7 FET-Open grant agreement n°222107 NIW Natural Interactive Walking.

References

[1] A. de Groot, R. Decker, and K. Reed. Gait enhancing mobile shoe (gems) for rehabilitation. In *Proceedings of World Haptics*, pages 190–195, 2009.

[2] J. Hollerbach, D. Checcacci, H. Noma, Y. Yanagida, and N. Tetsutani. Simulating side slopes on locomotion interfaces using torso forces. In *Proceedings of International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, page 91, 2003.

[3] J. Hollerbach, R. Mills, D. Tristano, R. Christensen, W. Thompson, and Y. Xu. Torso force feedback realistically simulates slope on treadmill-style locomotion interfaces. *International Journal of Robotics Research*, 20(12):939–952, 2001.

[4] J. Hollerbach, Y. Xu, R. Christensen, and S. Jacobsen. Design specifications for the second generation sarcos treadport locomotion interface. In *Proceedings of Haptics Symposium*, pages 1293–1298, 2000.

[5] H. Iwata, H. Yano, and F. Nakaizumi. Gait master: A versatile locomotion interface for uneven virtual terrain. In *Proceedings of IEEE Virtual Reality Conference*, pages 131–137, 2001.

[6] H. Iwata, H. Yano, and H. Tomioka. Powered shoes. In *Proceedings of SIGGRAPH 2006 Emerging technologies*, page 28, 2006.

[7] 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 SIGCHI Conference on Human factors in computing systems*, pages 239–246, 2004.

[8] A. Lécuyer, J.-M. Burkhardt, J.-M. Henaff, and S. Donikian. Camera motions improve sensation of walking in virtual environments. In *Proceedings of IEEE International Conference on Virtual Reality*, pages 11–18, 2006.

[9] H. Noma. Design for locomotion interface in a large-scale virtual environment. atlas: Atr locomotion interface for active self-motion. In *Proceedings of the ASME Dynamic Systems and control division*, pages 111–118, 1998.

[10] H. Noma, T. Sugihara, and T. Miyasato. Development of ground surface simulator for tel-e-merge system. In *Proceedings of IEEE Virtual Reality Conference*, page 217, 2000.

[11] F. Steinicke, G. Bruder, J. Jerald, H. Frenz, and M. Lapp. Estimation of detection thresholds for redirected walking techniques. *IEEE Transactions on Visualization and Computer Graphics*, 16(1):17–27, 2009.

Multimodal Floor for Immersive Environments

Alvin W. Law Jessica W. Ip Benjamin V. Peck Yon Visell Paul G. Kry Jeremy R. Cooperstock
McGill University, Montreal, Canada

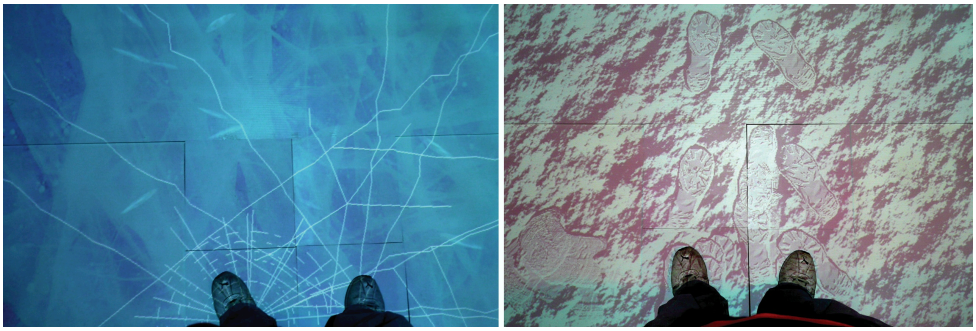


Figure 1: *Virtual ground surfaces exhibit their interactive, material-dependent characteristics when walked upon, through visual, auditory, and vibrotactile feedback. Shown are a frozen pond simulation (left) and a snowy field (right).*

We have developed an interactive system that allows untethered users to experience walking on virtual ground surfaces resembling natural materials. The demonstration consists of a multimodal floor interface for providing auditory, tactile and visual feedback to users' steps. It is intended for immersive virtual and augmented reality environments (VE) that provide the impression of walking over natural ground surfaces, such as snow and ice. To date, immersive environments with interactive floor surfaces have been largely focused on visual and auditory feedback linked to a VE simulation (e.g., [Gronbaek 2007]; see also the comparative review in [Miranda and Wanderley 2006]). However, while walking in natural environments, we receive continuous, multisensory information about the nature of the ground we walk on – the crush of dry leaves, the soft compression of grass. The static nature of floor surfaces in existing VEs typically bears little resemblance to a given natural ground material. This creates a perceptual conflict with the dynamic visual and/or auditory feedback that users are provided in the VE. This project illustrates a novel approach to reconciling such perceptual conflicts, based on multisensory feedback provided through a floor surface in response to users' steps.

The demonstration consists of a 6x6 foot, 36-tile floor surface. An array of force sensors within the floor (Interlink FSRs), is used to acquire the steps of users. Multimodal feedback is rendered, via physically-motivated models for the vibrotactile, auditory, and visual response of the material to the steps of a user [F. Fontana 2003; Law et al. 2008], and is displayed at the site of a footstep via an array of vibrotactile and audio actuators in the floor, as well as top-down video projection. Figure 1 and the accompanying video show a smaller 4 tile system along with a temporary projection surface as currently set up in our lab, pending completion of the larger system. The parts to construct the 36 floor tile array to be exhibited at SIGGRAPH have all arrived and are being assembled with planned completion in mid March. The final floor consists of a set of rigid wooden tiles with a durable, grey reflective paint. The prototype has been tested with hundreds of users during a McGill open house.

The two demonstration simulations consist of a frozen pond and field of snow (Figure 1). In the former, users can tip-toe over the virtual pond surface and observe fish swimming below. A hard step will cause the ice to fracture, accompanied by the sound of ice fracturing, the appearance of cracks, and synchronous vibrations. The snow setting allows one to leave footsteps onto virtual snow, with acoustic and vibrotactile response similar to the feeling of stepping onto real snow. The latter simulation requires motion capture (via a Vicon MCam2 system) of the posture of the feet in order to render the footsteps.

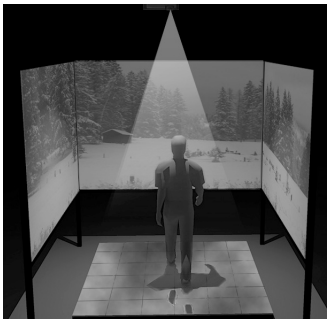


Figure 2: *The floor, in the configuration in which it is installed in the rear-projected CAVE-like environment of our lab.*

This system introduces a novel approach to affording natural walking activity in virtual environments, on virtual terrains. It may be used to enhance existing immersive VEs (Figure 2) or for augmented floors in other settings. Potential future applications can be envisioned in areas such as immersive VE training simulations, responsive floor-surfaces for entertainment parks, and interactive rehabilitation.

References

F. FONTANA, R. B. 2003. Physics-based sound synthesis and control. In *Proc. of the XIV Colloq. on Mus. Inf.*

GRONBAEK, K. 2007. Igamefloor: a platform for co-located collaborative games. In *ACE '07*.

LAW, A. W., PECK, B., VISELL, Y., KRY, P., AND COOPERSTOCK, J. 2008. A multi-modal floor-space for experiencing material deformation underfoot in virtual reality. In *Proc. IEEE HAVE'08*.

MIRANDA, E., AND WANDERLEY, M. 2006. *New Digital Musical Instruments: Control And Interaction Beyond the Keyboard.* AR Editions.

Copyright is held by the author / owner(s).
SIGGRAPH 2009, New Orleans, Louisiana, August 3–7, 2009.
ISBN 978-1-60558-726-4/09/0008

Probabilistic Tracking of Pedestrian Movements via In-Floor Force Sensing

Rishi Rajalingham, Yon Visell, Jeremy R. Cooperstock
Centre for Intelligent Machines and CIRMMT, McGill University
rishi@cim.mcgill.ca, yon@cim.mcgill.ca, jer@cim.mcgill.ca

Abstract

This article presents a probabilistic approach to the tracking and estimation of the lower body posture of users moving on foot over an instrumented floor surface. The latter consists of an array of low-cost force platforms providing intermittent foot-floor contact data with limited spatial resolution. We use this data to track body posture in 3D space using Bayesian filters with a switching state-space model. Potential applications of this work to person tracking and human-computer interaction are described.

1. Introduction

3D human posture tracking is a classic challenge in computer vision and pattern recognition. Computer vision techniques have been the most widely used for this purpose [7, 8], although several researchers have investigated human tracking via in-floor sensor arrays [10, 11, 1]. Common challenges of person tracking in these domains include the loss of 3D pose information through observation via the sensor array, and the underlying complex dynamics of human motion. Missing information plays an important role in both settings. Losses due to occlusion in video-based tracking are somewhat analogous to the loss of observations while feet are out of contact with the ground during tracking via a floor-based array. Overcoming such missing information has, in part, motivated the approach presented here.

Although motion capture and video can provide high-resolution 3D position information for human tracking, they are not always available and are prone to visual occlusion. Furthermore, state of the art methods for inferring human contact interactions from video provide inaccurate information about interaction forces between body and ground [3]. Such forces are highly characteristic of the individuals and activities generating them (as evidenced in the references given below).

Bayesian filtering provides a unifying view of diverse probabilistic tracking methods. It has been extensively applied to problems in object, person, or context tracking [4, 17]. The effectiveness of such methods stems from their ability to integrate information acquired over time in ways that respect the structure and dynamics of the object or individual being tracked. Nonparametric Bayesian filtering techniques, like the particle filter based model used here, make it possible to perform tracking without making unnecessary assumptions about the form taken by those distributions or the dynamics governing them, by maintaining many hypotheses in parallel.

The system presented here consists of a posture tracking system based on a distributed, sparse in-floor sensing array. Prior literature on the analysis of foot-floor contact forces has addressed applications to immersive interactive media [12], pedestrian identification [11, 16], gait and dance analysis [13], among others [21, 15, 9, 6]. Here, similar data is used for 3D kinematic tracking of users' lower bodies, a task that has received less attention in prior research on floor-sensing arrays. Murakita et al. utilized a Markov Chain Monte Carlo method to track pedestrians via an array of binary pressure sensors [10], albeit with much larger errors (typically 0.6 m) than what we achieve here (Sec. 4). Yin and Pai tracked whole-body movements via a high-resolution (and costly) floor sensing array, based on the similarity of force patterns to those recorded in a database of movements [21]. However, their system was limited to tracking relative to a predefined set of static poses.

Our intended applications involve both person tracking and interaction with floor-based touch displays via the feet [20]. In both cases, we are interested in tracking the 3D kinematics of the lower body of moving persons, with an emphasis on the locations of their feet. Although existing force sensing arrays can provide accurate information about foot-floor contact forces, the achievable resolution is often limited by cost constraints. In addition, mappings from patterns of foot-floor forces to body posture are complex and one-to-many. Effective use of prior knowledge about body structure, movement, and walking mechanics is required in order to track posture accurately.

2. System Configuration

The sensing floor (Fig. 1) consists of a 6×6 array of rigid tiles, 30 cm on each side. Each tile is instrumented with four resistive force sensors (Interlink model 402), which are located at the corners. Thus, the nominal linear sensing resolution of this array is 15 cm. In addition, each tile includes a wide bandwidth vibrotactile actuator [18], which although unused here, does modestly influence the sensor measurements, due to its weight [20]. Analog data from the force sensors are conditioned, amplified, and digitized via 32-channel, 16-bit data acquisition boards. Each sensor is sampled at a rate of up to 1 kHz transmitted over a low-latency Ethernet link. An array of six small-form-factor

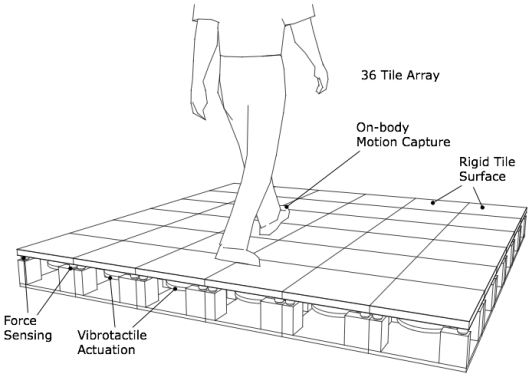


Figure 1. Illustration of the distributed floor interface, with components labelled.

computers is used for force data processing.

For applications that do not require kinematic tracking, we infer foot-floor contact loci using intrinsic contact sensing techniques [20], attaining an effective resolution of about 2 cm. However, such methods are incapable of tracking the location of a foot once it leaves the floor surface, and cannot resolve situations in which the feet overlap onto a single tile. Moreover, inference of lower body pose from in-floor force sensor data is challenging due to the loss of information inherent in this complex mapping. Such limitations have, in part, motivated the present work. In particular, we adopt the framework of Bayesian filtering in order to maintain continuity of lower body position estimates in dynamic settings, such as walking, where foot-ground contact is regularly interrupted.

Within this framework, our task remains challenging due to the high dimensional mapping from an articulated pedestrian pose to the observed force sensor values. In particular, this map is discontinuous for a sensing infrastructure consisting of independent, rigid tiles, as poses that are similar in nature, i.e. with footprints of similar 2D position and orientation, will lead to drastically different force observations when tile boundaries are crossed.

3. Tracking Problem and Algorithm

Consider a dynamic system with states x_t and observations z_t, both indexed by time. A Bayesian filter probabilistically estimate at time t the state x_t by sequentially updating a belief distribution $\mathrm{Bel}(x_t)$ over the state space, defined by $\mathrm{Bel}(x_t) = p(x_t|z_t, z_{t-1}, z_{t-2}, ...) = p(x_t|z_{1:t})$. Assuming the states comprise a Markov process, the Belief distribution at each subsequent time step can be obtained, using Bayes' Theorem, in terms of the belief state at the prior time step $t-1$, the assumed motion model $p(x_t|x_{t-1})$, and the likelihood $p(z_t|x_t)$ of the newly acquired observation z_t given the state x_t:

$$\mathrm{Bel}(x_t) \propto p(z_t|x_t) \int p(x_t|x_{t-1})\mathrm{Bel}(x_{t-1})\, dx_{t-1} \quad (1)$$

Bayesian filters can be distinguished, in part, by the form of the Belief distribution $\mathrm{Bel}(x_t)$ that is assumed. Here, we adopt a Monte-Carlo approach, in which $\mathrm{Bel}(x_t)$ is represented by a set of weighted samples, or particles, given by $S_t = \{(x_t^i, w_t^i), i = 1, 2, \ldots N_s\}$. Our method uses the Sampling-Importance-Resampling (SIR) algorithm [5], described in Alg. 1, and schematically illustrated in Fig 2.

For each new observation, SIR re-weighs the set of particles depending on their likelihood, evolves them in time using the assumed dynamic model, and updates the particle set S_t by resampling based on these weights.

Specifically, the likelihood function $L(x_t) = p(z_t|x_t)$ is computed in two steps as follows.

1. Using the observation model $\mathbf{H}$, defined in Section 3.1, that maps states into the observation space, generate expected observations $z_t^* = \mathbf{H}(x_t)$.

2. Using the similarity measure $S(z, z') = p(z|z')$, defined in Section 3.2, compute the likelihood as $L(x_t) \equiv S(z_t^*, z_t)$.

As seen in Eq.2, each particle is weighted according to its likelihood while also considering a prior distribution $p(x_t)$. As described in Section 3.2, the prior distribution $p(x_t)$ is useful for encoding constraints on states x_t, thus leading to an efficient search of the state space.

The attributed weight is then used to resample the particle set (see Eq. 3). This step distinguishes SIR from SIS (Sampling Importance Sampling), and is meant to eliminate degeneracy of particles; by concentrating on high weighted particles to create a new uniformly distributed particle set, the extreme case where most particles have negligible weight is avoided. The opposite extreme, consisting of a single strong hypothesis, can be avoided by the inclusion of a roughening process. Roughening consists of the addition of random, zero-mean, normally distributed noise to all particles after resampling (Eq. 5), to allow for a thorough search of the solution space. Additionally, particles

are propagated forward in time based on the assumed dynamic model (see Eq.4), which implicitly defines the motion probability $p(x_t|x_{t-1})$. Our dynamic model consists of two processes, accounting for the continuous and discrete aspects of the motion of interest. The dynamic model and roughening process are defined in detail in Section 3.3.

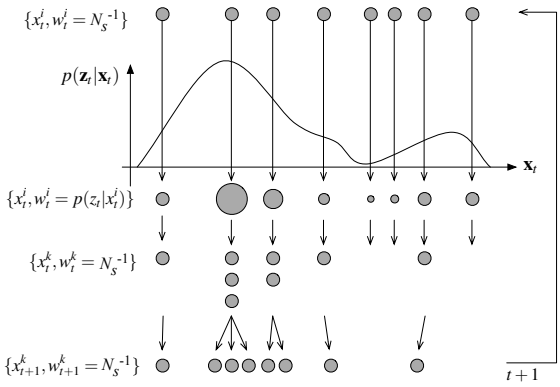


Figure 2. The SIR particle filter algorithm.

Algorithm 1 Sampling Importance Resampling

Initialize particles randomly: $S_0 \sim N(\mu_S, \sigma_S)$
while $t > 0$ **do**
Observe z_t.
for $i = 1$ to N_s **do**

$$\text{Likelihood:} \quad p(z_t|x_t^i) = S(z_t, H(x_t))$$
$$\text{Weight:} \quad w_t^i = p(x_t^i) \times p(z_t|x_t^i) \qquad (2)$$

end for
for $i = 1$ to N_s **do**

$$\text{Normalize:} \quad W = \textstyle\sum_i w_t^i, w_t^i \leftarrow W^{-1} w_t^i$$
$$\text{Resample:} \quad x_t^i \sim p(x_t^{i*}|w_t^i), w_{t+1}^i = N_s^{-1} \qquad (3)$$
$$\text{Draw:} \quad x_{t+1}^i \sim p(x_{t+1}|x_t^i) \qquad (4)$$
$$\text{Roughen:} \quad x_{t+1}^i \leftarrow x_{t+1}^i + \eta_{t+1}, \eta \sim N(0, \sigma_x) \qquad (5)$$

end for
$S_{t+1} = \{(x_{t+1}^i, w_{t+1}^i)\}_{i=1}^{N_s}$
end while

In our system for tracking via in-floor force measurements, the relevant variables consist of:

- Observations $\mathbf{z}_t$, consisting of a 12×12 array of force values, f_i.
- States, $\mathbf{x}_t$, describing kinematic lower-body poses, are 19-dimensional vectors: $\mathbf{x}_t$ = ($\phi_{l,t}$, $\dot{\phi}_{l,t}$, $\dot{\phi}_{l,t-1}$, $\phi_{r,t}$, $\dot{\phi}_{r,t}$, $\dot{\phi}_{r,t-1}$, β). They include planar midpoint coordinates $\mathbf{u}$ and orientations θ for each foot, where $\phi_l = (\mathbf{u}_l, \theta_l)$ and likewise for ϕ_r, along with first time derivatives. The state $\mathbf{x}_t$ also includes a binary-valued vector $\beta = (\beta_l, \beta_r)$, implemented as a quaternary variable, which indicates the foot-floor contact condition ($\beta_i = 1$ if there is contact) for the left and right feet.

Figure 3 illustrates this state description within a skeletal model. The algorithm definition is completed by specifying the observation model $\mathbf{z}_t = \mathbf{H}(\mathbf{x}_t)$, the likelihood model $p(\mathbf{z}_t|\mathbf{x}_t)$, and the motion model, $p(\mathbf{x}_t|\mathbf{x}_{t-1})$.

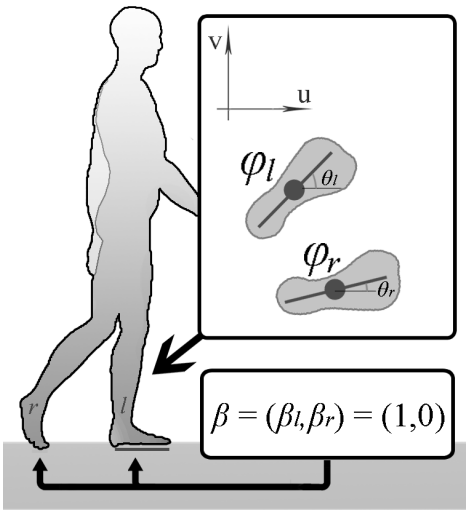


Figure 3. State description of lower body poses: $\mathbf{x}_t$ defines feet and foot-floor contact.

3.1 Observation model

We model expected observations $\mathbf{H}(\mathbf{x}_t)$ for a state $\mathbf{x}_t$ by simulating the mechanical forces associated with a pose. Ignoring shear forces, foot-tile contact results in a normal pressure distribution $p(\mathbf{u})$, where $\mathbf{u} = (u, v)$ are 2D coordinates on the floor. The pressure distribution on a tile is conveniently summarized by a contact centroid $p_c = (\mathbf{u}_c, F)$, where $F = \int d\mathbf{u}\, p(\mathbf{u})$ is the net normal force and $\mathbf{u}_c$ is the pressure centroid. A normal force with magnitude F, applied at $\mathbf{u}_c$, would give rise to the same force measurements f_i as $p(\mathbf{u})$ [20, 2] (Fig 4). Our observation model associates a pose $\mathbf{x}_t$ to a set of contact centroids, $p_{c,j}, j = 1, 2, \ldots, N_c$. One centroid is placed on each tile that a foot pose is determined to be in contact with ($\beta_i > 0$), and the total weight F of the user is partitioned among these centroids. The placement of a contact centroid is determined by the average position of foot to tile contact. Expected sensor readings f_i are obtained from the static equilibrium equations for each tile. Each pressure centroid p_c yields a contribution $f_{i,c} = d_i^{-1}(\sum_{j=0}^{4} d_j^{-1})^{-1}$, where $d_j = |\mathbf{u}_c - \mathbf{u}_j|$, and $\mathbf{u}_j$ is the sensor location. In this way, we predict observed force values $\mathbf{z}_t^* = \{f_{i,t}\}$ for a state $\mathbf{x}_t$.

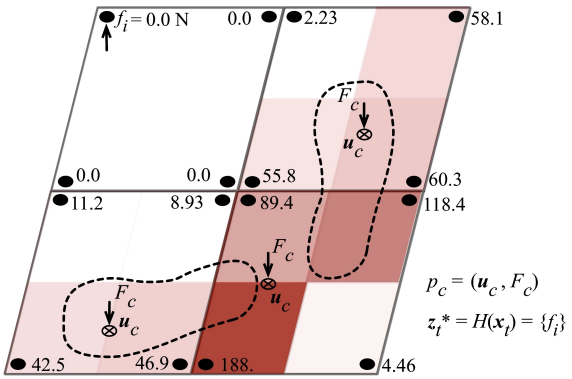


Figure 4. The observation model $H(\mathbf{x}_t)$ maps a pose $\mathbf{x}_t$ (illustrated here as two feet) to a set of contact centroids ($\mathbf{u}_c$). These are then converted to force sensor readings $\mathbf{z}_t = \{f_i\}$. Expected force observations are illustrated as grids, with each quadrant intensity proportional to corresponding force sensor value (given in Newtons).

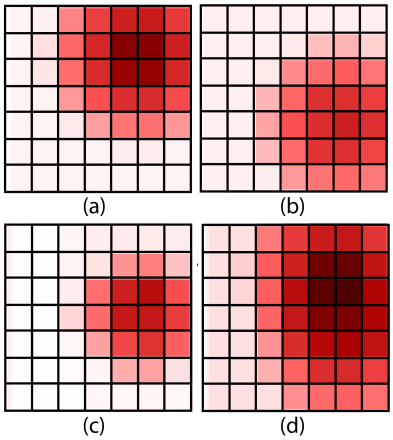


Figure 5. 2D pressure distributions for (a) $\mathbf{z}_t$, (b) $\mathbf{z}_t^*$, (c) $\mathbf{z}_t \cap \mathbf{z}_t^*$, (d) $\mathbf{z}_t \cup \mathbf{z}_t^*$. Pressure distributions are illustrated as grids, with each quadrant intensity proportional to corresponding force sensor value. The intersection ($\cap$) and union ($\cup$) of force observations are used in computing similarity $S(\mathbf{z}_t, \mathbf{z}_t^*)$.

3.2 Likelihood model

The likelihood function $L(\mathbf{x}_t) = p(\mathbf{z}_t|\mathbf{x}_t)$ describes the probability of observing force values $\mathbf{z}_t$ given the state $\mathbf{x}_t$. As described above, we use the observation model to generate expected observations, $\mathbf{z}_t^* = \mathbf{H}(\mathbf{x}_t)$. The likelihood function is then defined in terms of a normalized similarity measure $p(\mathbf{z}_t|\mathbf{x}_t) = S(\mathbf{z}_t^*, \mathbf{z}_t)$ between true force pattern observations $\mathbf{z}_t$ and expected force observations $\mathbf{z}_t^*$.

Similarity measure S The similarity measure $S(\mathbf{z}, \mathbf{z}') = p(\mathbf{z}|\mathbf{z}')$ models the probability of observing $\mathbf{z}$ if the true observation is $\mathbf{z}'$. Conventional similarity measures make use of metrics such as Euclidean or Mahalanobis distance functions. However, since $\mathbf{H}$ is a high-dimensional discontinuous map from states $\mathbf{x}_t$ to observations $\mathbf{z}_t$, these measures cannot properly gauge similarity between observation vectors $\mathbf{z}_t$, $\mathbf{z}_t^* = \mathbf{H}(\mathbf{x}_t)$. Moreover, the force observations are comparatively sparse, with most values being zero. As an alternative, we compute pair-wise similarity between such patterns, based on a measure of their area of overlap. Specifically, we employ a similarity measure that has proved useful in tracking via binary image masks [14], computing $S(\mathbf{z}^*, \mathbf{z})$ as the relative area of overlap between the true and expected 2D pressure distributions,

$$S(\mathbf{z}_t^*, \mathbf{z}_t) = \frac{\cap(\mathbf{z}_t^*, \mathbf{z}_t)}{\cup(\mathbf{z}_t^*, \mathbf{z}_t)} = \frac{1}{N_z} \sum_{i=1}^{N_z} \frac{\min(\mathbf{z}_t(i), \mathbf{z}_t^*(i))}{\max(\mathbf{z}_t(i), \mathbf{z}_t^*(i))}$$

Fig. 5 illustrates conceptual examples of observations $\mathbf{z}_t$, $\mathbf{z}_t^*$ as 2D pressure distributions, as well as their intersection $\cap(\mathbf{z}_t^*, \mathbf{z}_t)$ and union $\cup(\mathbf{z}_t^*, \mathbf{z}_t)$.

We note that the average overlap of these resulting pressure distributions is an effective metric for capturing similarities between force observations as a probability:

$$\begin{aligned} S(\mathbf{z}, \mathbf{z}) &= 1 \\ \mathbf{z}_1 \neq \mathbf{z}_2 &\Rightarrow 0 < S(\mathbf{z}_1, \mathbf{z}_2) < 1 \\ S(\mathbf{z}_1, \mathbf{z}_2) &= S(\mathbf{z}_2, \mathbf{z}_1) \end{aligned} \quad (6)$$

Postural constraints The likelihood model is modified to encode human postural constraints, via a prior distribution $p(\mathbf{x}_t)$. The latter is defined to consist of a set of independent postural priors for human walking, in the form of univariate Gaussian distributions $N(\mu_\varsigma, \sigma_\varsigma)$, $N(\mu_\Theta, \sigma_\Theta)$ over stance width $\varsigma = \|\mathbf{u}_l - \mathbf{u}_r\|$ and relative orientation $\Theta = |\theta_l - \theta_r|$ respectively. The postural prior is introduced as the product $p(\mathbf{x}_t) = N(\varsigma; \mu_\varsigma, \sigma_\varsigma) N(\Theta; \mu_\Theta, \sigma_\Theta)$. The prior distribution $p(\mathbf{x}_t)$ is applied when computing the particle weights from the likelihood L (see Eq. 2).

3.3 Dynamics Model

We model the movements of the lower body of a walker via the motion probability $p(\mathbf{x}_t|\mathbf{x}_{t-1})$. The state x_t consists of continuous configuration variables ϕ_i and discrete contact variables β_i. We therefore approximate foot motion in a hybrid (stochastic) framework, incorporating continuous, linear movements of the limb coupled to discrete state transitions reflecting changes of the foot-floor contact conditions.

Continuous, linear dynamics The state representing the left or right foot is denoted respectively by a vector $\phi_i = (\mathbf{u}_i, \theta_i)$ giving the midpoint and orientation of either the left ($i = l$) or right ($i = r$) foot at time t (Fig. 3. The dynamics used in our algorithm can be described by the following linear, discrete time system:

$$\phi_{i,t} = \phi_{i,t-1} + \beta_i \Gamma_i (\dot{\phi}_{t-1} dt + \eta_t^0) \quad (7)$$

$$\dot{\phi}_{i,t} = \beta_i [\alpha \dot{\phi}_{i,t-1} + (1-\alpha)(\dot{\phi}_{i,t-2} + \eta_{t-1}^1)] \quad (8)$$

The parameter β_i is the binary contact variable for foot i. Thus, ϕ_i is constant when there is contact ($\beta_i = 0$) and otherwise drifts, with position and velocity driven by additive Gaussian noise processes η_t^0 or η_t^1, where $\eta \sim N(0, \Sigma)$. For efficiency, we parametrize drift via a single noise process, defining $\eta \equiv \eta_t^0 = \eta_t^1 dt$ for all t. The noise covariance Σ is a 3×3 diagonal matrix with diagonal entries $\sigma_u, \sigma_v, \sigma_\theta$. To mimic walking, velocity drift in the direction that the foot is oriented is assumed to be larger. This non-isotropic drift is implemented through the factor Γ_i, a diagonal matrix with entries $(\cos\theta_i(\gamma + \sqrt{1-\gamma^2}), \sin\theta_i(\gamma - \sqrt{1-\gamma^2}), 1)$, where $0 < \gamma < 1$ is a dimensionless scalar defining the longitudinal bias. α is a dimensionless scalar defining the velocity noise mixing rate. It approximates the dynamics of a free foot during walking by means of a saturating linear drift velocity.

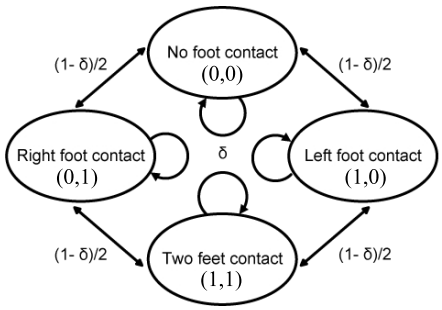


Figure 6. Stochastic state transition diagram approximating stepping motion.

Roughening As presented in (Alg. 1), noise $\eta \sim N(0, \sigma_x)$ is added to the continuous state components ϕ_i at each SIR step in order to avoid particle degeneracy.

Discrete state transition model The dynamic model includes discrete transitions from the foot-floor contact states $\beta = (\beta_l, \beta_r)$, where $\beta = 0$ or 1, via the stochastic process, shown in Figure 6. Despite its simplicity, this model is effective in approximating discrete stepping motion. δ is an empirically determined probability of no change in contact state β. All remaining transitions are symmetric, with transition probability $\frac{1-\delta}{2}$.

Foot	Error (m)	Windowed error (m)
Right (no contact)	0.1901	0.1245
Left (no contact)	0.1649	0.0336
Average (no contact)	**0.1775**	**0.0791**
Right (contact)	0.2025	0.1058
Left (contact)	0.1406	0.0306
Average (contact)	**0.1716**	**0.0682**
Right (all)	0.1982	0.1122
Left (all)	0.1490	0.0316
Average (all)	**0.1736**	**0.0719**

Table 1. RMS Position Error

4. Experiment and Results

The system described above was evaluated by measuring the absolute positions of the feet of pedestrians using data acquired synchronously via motion capture (Vicon Motion Systems). Reflective markers were attached to the walkers' shoes, providing an accurate estimate of 3D foot positions. Five recordings of walking sequences between 5.7 and 12.4 seconds in length were acquired via the apparatus described in Sec. 2. Synchronous motion capture and force data were recorded. Errors were computed based on maximum a posteriori (MAP) foot position estimates obtained from the tracking algorithm.

Figure 7 shows the state estimates and force observations at four stages of the walking sequence: the initial particle set has high variance, but is quickly narrowed down to a few hypotheses which are evolved based on our motion model. Figure 8 shows the resulting motion trajectories and errors for the two foot locations, in planar coordinates. Average RMS error values are reported in Table 1. The experimental parameters are given in Table 2. Position errors during foot-floor contact are found to be slightly less on average than when a foot is not in contact with the floor. Temporal alignment mismatches were found to have a large effect, so we also performed a windowed error calculation in which an acceptable time shift of 10 samples (at 20 Hz) was permitted. This greatly reduced RMS position errors (see Table 1), suggesting that system tracking performance may be most acceptable in situations in which temporal accuracy is not important. Tracking performance in more temporally demanding settings might be greatly improved if a better alignment can be achieved. Video documentation of these results is provided in the supplementary material, and available online here.

In addition to position estimates, this system provides continuous labels identifying the walker's right and left feet. During tracking of the sequences used for evaluation, left and right feet were continuously and coherently identified with 100% accuracy. The capability of this system to maintain and propagate these labels may be useful for appli-

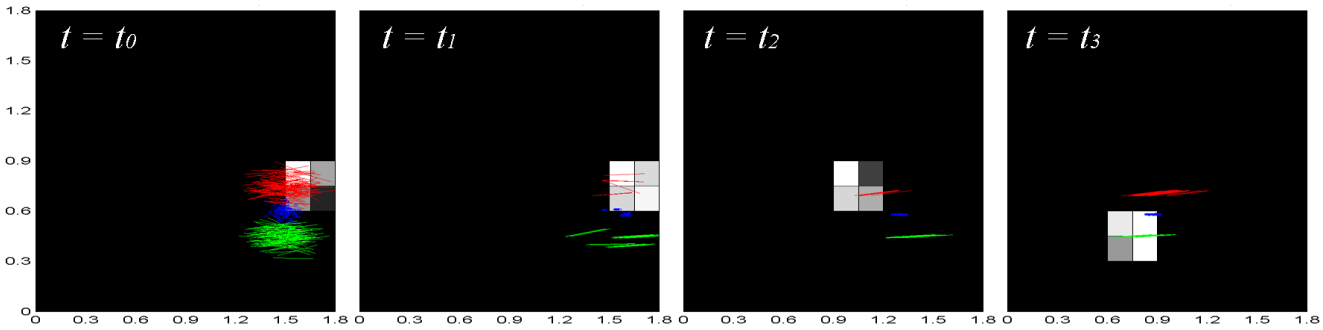


Figure 7. Force observations and pose estimates at 4 stages of the sample walking sequence. Observations are shown as 2D pressure distributions, with quadrant intensity proportional to force sensor readings. Poses are illustrated as green and red lines, corresponding to a top view of left and right feet estimates respectively, with blue circles corresponding to inferred center of mass.

SIR Algorithm	
Number of particles	$N_s = 200$
Initial Noise	$\sigma_S = 4.572$ cm
Roughening Position Noise	$\sigma_x = 0.003048$ cm/rad
Likelihood Model	
Postural prior stance size	$(\mu_\varsigma, \sigma_\varsigma) = (30.48, 30.48)$ cm
Postural prior stance angle	$(\mu_\Theta, \sigma_\Theta) = (0, \pi/2)$ rad
Dynamic Model	
Additive Position Noise	$\sigma_{u,v} = 0.3048$ cm
Additive Angle Noise	$\sigma_\theta = 0.003048$ rad
Longitudinal bias	$\gamma = 0.9$
Velocity mixing rate	$\alpha = 0.55$
Probability of no β change	$\delta = 0.55$

Table 2. Experimental parameters

cations including floor-based touch screen user interfaces, where it may be desirable to assign each foot a different functional operation, or to render a different response to left and right foot [20].

5. Conclusion

This paper presented Bayesian filtering techniques to track the lower body pose of a pedestrian from foot-floor interaction forces acquired via a coarse array of in-floor force sensors. The system achieves continuous and labeled tracking of the lower limbs of a walker via a coarse sensor array, by combining prior knowledge about the mechanical structure of the interface and a simple, but consistent, model of the dynamics of the feet during walking. In the experiment described above, our system never confused the left and right feet of the walker, and was able to track the locations of each with an average resolution on the order of 15 cm, and with improved resolution when feet are in contact with the surface. Potential applications of these techniques include tracking in smart environments in which motion capture is impractical (due to occlusion or other factors), and to interaction with distributed, floor-based touch surface interfaces for the feet [20, 19].

Despite the promising nature of the results presented above, further experiments are needed to evaluate the quality of the tracking. In addition, it is clear that the system itself can be improved in several respects. A higher-density sensor network would improve position estimates during contact, albeit at greater cost. A model for the non-contact portion of walking movement that is more sophisticated than the random drift model used in our system could significantly improve estimates in foot tracking when a foot is not in contact with the floor. The incorporation of additional prior knowledge about the kinematic constraints on lower limb positions during walking would also be expected to contribute improvements. In ongoing work, we are exploring the possibilities for optimally fusing information from in-floor force sensors with motion capture or other video sensors, in order to resolve data loss due to camera occlusion, or to provide contact forces and timing information that cannot be accurately estimated from video.

References

[1] M. Addlesee, A. H. Jones, F. Livesey, and F. S. Samaria. The orl active floor. *IEEE Personal Communications*, 4:35–41, 1997.

[2] A. Bicchi, J. Salisbury, and D. Brock. Contact sensing from force measurements. *The International Journal of Robotics Research*, 12(3):249, 1993.

[3] M. Brubaker, L. Sigal, and D. Fleet. Estimating Contact Dynamics. In *IEEE International Conference on Computer Vision (ICCV)*, 2009.

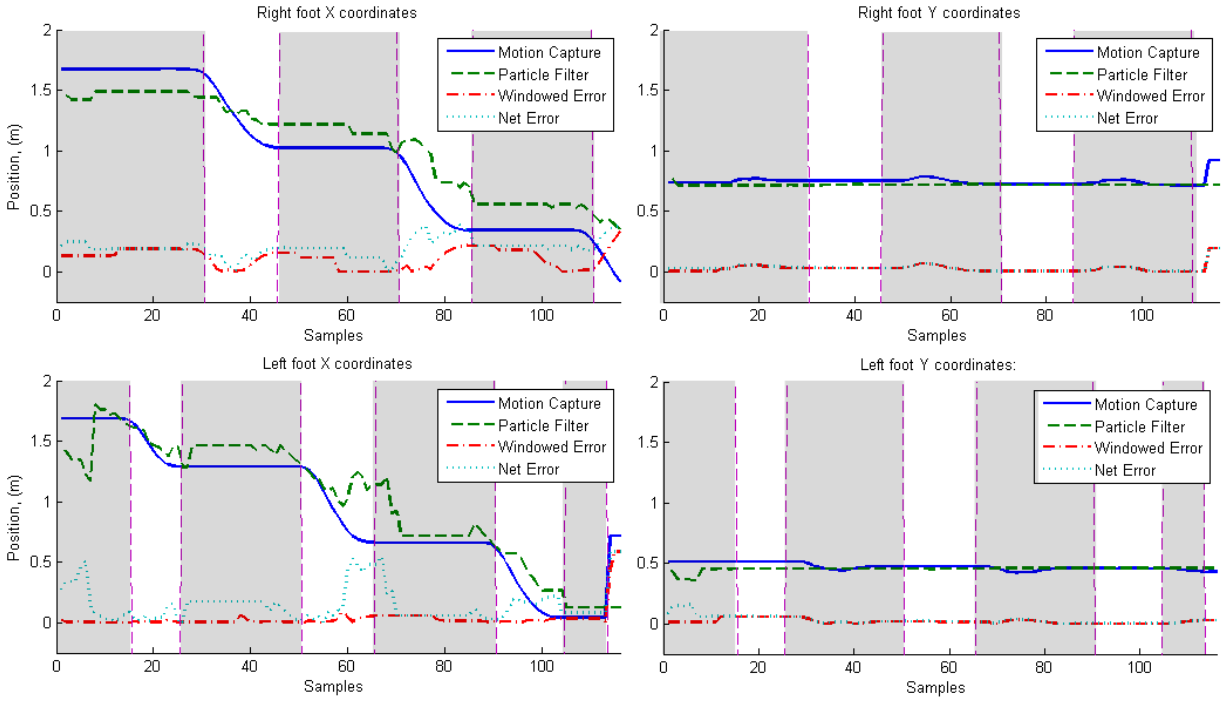


Figure 8. Sample results for a pedestrian crossing the apparatus, comparing measurements from motion capture data and estimates obtained from the particle filter. Results are shown for both feet ϕ_l, ϕ_r. The shaded domains illustrate contact states, with grey and white corresponding to $\beta_i = 1$, $\beta_i = 0$ respectively.

[4] D. Fox, J. Hightower, H. Kauz, L. Liao, and D. Patterson. Bayesian techniques for location estimation. In *Proceedings of The 2003 Workshop on Location-Aware Computing*, page 16, 2003.

[5] V. Fox, J. Hightower, L. Liao, D. Schulz, and G. Borriello. Bayesian filtering for location estimation. *Pervasive Computing, IEEE*, 2(3):24–33, July-Sept. 2003.

[6] L. Middleton, A. Buss, A. Bazin, and M. Nixon. A floor sensor system for gait recognition. In *Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, pages 171–176, 2005.

[7] T. Moeslund and E. Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81(3):231–268, 2001.

[8] T. Moeslund, A. Hilton, and V. Krueger. A survey of advances in vision-based human motion capture and analysis. *Computer vision and image understanding*, 104(2-3):90–126, 2006.

[9] H. Morishita, R. Fukui, and T. Sato. High resolution pressure sensor distributed floor for future human-robot symbiosis environments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2002), Lausanne, Switzerland*, pages 1246–1251, 2002.

[10] T. Murakita, T. Ikeda, and H. Ishiguro. Human tracking using floor sensors based on the Markov chain Monte Carlo method. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, 2004.

[11] R. J. Orr and G. D. Abowd. The smart floor: a mechanism for natural user identification and tracking. In *CHI '00: CHI '00 extended abstracts on Human factors in computing systems*, pages 275–276, New York, NY, USA, 2000. ACM.

[12] J. Paradiso, C. Abler, K.-y. Hsiao, and M. Reynolds. The magic carpet: physical sensing for immersive environments. In *CHI '97: CHI '97 extended abstracts on Human factors in computing systems*, pages 277–278, New York, NY, USA, 1997. ACM.

[13] R. F. Pinkston. A touch sensitive dance floor/midi controller. *The Journal of the Acoustical Society of America*, 96(5):3302–3302, 1994.

[14] I. Rius, X. Varona, F. X. Roca, and J. Gonzàlez. Posture constraints for bayesian human motion tracking. In *Articulated Motion and Deformable Objects*, pages 414–423. Springer Verlag, 2006.

[15] A. Schmidt, M. Strohbach, K. v. Laerhoven, A. Friday, and H.-W. Gellersen. Context acquisition based on load sensing. In *UbiComp '02: Proc of the 4th international conference on Ubiquitous Computing*, pages 333–350, London, UK, 2002. Springer-Verlag.

[16] J. Suutala, K. Fujinami, and J. Roening. Gaussian Process Person Identifier Based on Simple Floor Sensors. In *Proceedings of the 3rd European Conference on Smart Sensing and Context*, page 68. Springer-Verlag, 2008.

[17] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT Press, 2005.

[18] Y. Visell and J. Cooperstock. Design of a Vibrotactile Display via a Rigid Surface. In *Proc. of IEEE Haptics Symposium*, 2010.

[19] Y. Visell and J. Cooperstock. Visell, Y. and Law, A. and Smith, S. and Ip, J. and Cooperstock, J. In *Proc. of IEEE Haptics Symposium*, 2010.

[20] Y. Visell, A. Law, S. Smith, R. Rajalingham, and J. Cooperstock. Contact sensing and interaction techniques for a distributed multimodal floor display. In *Proc. of IEEE 3DUI*, 2010.

[21] K. Yin and D. K. Pai. Footsee: an interactive animation system. In *SCA '03: Proc of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 329–338, 2003.

EXPRESSIVE SONIFICATION OF FOOTSTEP SOUNDS

Roberto Bresin, Anna deWitt

KTH Dept. of Speech Music and Hearing
Stockholm, Sweden
roberto@kth.se, dewitt@kth.se

Stefano Papetti, Marco Civolani, Federico Fontana

University of Verona, Department of Computer Science
Verona, Italy
Name.Surname@univr.it

ABSTRACT

In this study we present the evaluation of a model for the interactive sonification of footsteps. The sonification is achieved by means of specially designed sensored-shoes which control the expressive parameters of novel sound synthesis models capable of reproducing continuous auditory feedback for walking. In a previous study, sounds corresponding to different grounds were associated to different emotions and gender. In this study, we used an interactive sonification actuated by the sensored-shoes for providing auditory feedback to walkers. In an experiment we asked subjects to walk (using the sensored-shoes) with four different emotional intentions (happy, sad, aggressive, tender) and for each emotion we manipulated the ground texture sound four times (wood panels, linoleum, muddy ground, and iced snow). Preliminary results show that walkers used a more active walking style (faster pace) when the sound of the walking surface was characterized by an higher spectral centroid (e.g. iced snow), and a less active style (slower pace) when the spectral centroid was low (e.g. muddy ground). Harder texture sounds lead to more aggressive walking patters while softer ones to more tender and sad walking styles.

1. INTRODUCTION

The strong relationship between body motion and sound production has been researched and documented in recent years. Most of the research in this field has been conducted looking at music performance (for an overview see [1]). Little work has been done in the area of everyday sounds produced by the human body and how they could influence the body action itself. In the work presented in this paper the focus is on the relationship between walking and footstep sounds. The main idea is to investigate if the sound produced by human feet while walking can influence the walk style.

In previous works about footstep sounds it has been shown how different sounds are preferred for walkers of different genders [2], and that when walking with different emotional intentions humans make variations of timing and sound level in the same way as found in expressive music performance [3]. It has also been found that there are timbres of musical instruments which are more suitable for certain emotions [4, 5, 6].

In this work we would like to go a step further, and investigate how the sound produced by the impact of shoes on the ground influences the walking style when walking with different emotional intentions. Our hypothesis is that, within the same emotion, walkers change their walking strategy depending on the impact sound of the shoe on the ground. We therefore designed an experiment that is presented in the following section.

The scientific context of this work is that of interactive sonification [7]: as reported in the following sections, data collected from sensors on the shoes have been used for direct real time control of sound models for the auditory display of different ground textures.

2. EXPERIMENT

An experiment was designed in which subjects were asked to walk with four emotional intentions (happy, aggressive, tender, sad) along a path which emitted footstep sounds corresponding to four different grounds. We choose the four emotions happiness, sadness, aggressiveness, and tenderness mainly for two reasons. First they have been investigated in several studies on emotional expression in music (for an overview see [8]). Second, they cover the four quadrants of the two-dimensional activity-valence space [4], and therefore give a quite comprehensive overview on the use of musical parameters such as tempo, that is one parameter that can be analyzed also in walking.

Figure 1: *Active shoes prototype.*

2.1. Subjects

For the experiment we gathered 16 subjects, 13 males and 3 females of the Italian nationality, with average age 27,875. 13 subjects had a musical background and played an instrument. All subjects listen to music. They were not compensated for their participation.

2.2. Equipment

A pair of prototype active shoes (Figure 1) was used for the experiment. The shoes capture the foot pressure that is used for informing different real-time synthesis models, whose output feeds the shoe-mounted haptic actuators and loudspeakers.

Our prototype relies on simple, inexpensive technology both at the sensing and at the actuation side.

For each shoe (see Figure 2), force sensing is operated by a couple of Interlink 400 force sensing resistors, one set in the front part of the insole, one on the rear part. The force data from both shoes are digitized by an Arduino Duemilanove acquisition board, that cuts off measurement noise in the force values depending on

Copyright http://creativecommons.org/licenses/by/3.0/

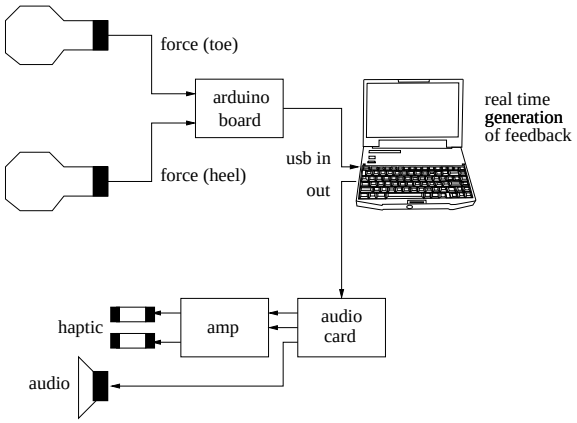


Figure 2: *Connections of physical components realizing one active shoe.*

a threshold, which can be set based on the user's weight and foot size [9].

The digital data are sent, via USB, to a MacBook that generates and conveys the resulting feedback to an RME FireFace 800 eight-channel audio card. From here, the signal goes to a small loudspeaker mounted over each shoe, and in parallel to a conventional stereo audio amplifier, which provides the needed power to form the final output for the haptic actuators, two for each shoe. In total, three signal channels are involved to feed each shoe.

On-shoe small loudspeakers allow for a precise recognition of the sound source position. On the other hand, they cannot emit low frequencies. The haptic actuators compensate for the lack of depth in the sound. By implementing a novel design of recoil-type actuators, they generate vibro-tactile components ranging 40 to 10 kHz, furthermore consume less than 5 W of power to produce thrust up to 5 N, resulting in very high accelerations. Due to their design and structure they can be immersed in the inside of a sole and are able to support the weight of a person [10].

Two subwoofers were placed at the two ends of a 6.8 m long walking path along which subjects were asked to walk, e.g. one subwoofer was positioned at the back of the subject and the other one at the front.

2.3. Stimuli

We used four sounds corresponding to grounds of different nature. Two sounds were directly generated by the foot impact on the ground (real-word textures: wood panels and linoleum floor), and the other two sounds were synthesized by an algorithm triggered by sensors on the shoes (augmented-reality textures: iced snow and muddy ground).

The real-world sound texture for wood panels was realized by putting a catwalk of wooden panels on the floor. The linoleum floor was theoriginal floor of the experiment room. The auditory feedback of iced snow and muddy ground were synthesized in real-time, as explained in the next section, and reproduced by the loudspeakers on the shoes and by the two subwoofers. A video demonstrating augmented-reality sound texture generation and the equipment used for the experiment is available on-line[1].

In the next section we briefly introduce the sound models used for synthesis of iced snow and muddy ground.

[1] `http://www.youtube.com/watch?v=S3P2lKVDMXk`

2.3.1. Sound synthesis

The software used for the synthesis of instantaneous and continuous feedback is a crucial aspect of the system. This software has been coded for the open source real time synthesis environment Pure Data[2]. Due to its flexibility and additional libraries, Pure Data allows to interface the Arduino board directly with the audio card, meanwhile taking in charge of all the needed processing. The patch used for the synthesis of mud sounds accounts for the tuning of parameters for virtual "soft" impacts, each triggered and controlled by the contact force received from the force sensors on the shoes.

All the Pure Data patches used in connection with the active shoes are part of the the Sound Design Toolkit (SDT [11]), a library of ecological synthesis models maintained at the University of Verona, which includes a palette of sound generation algorithms for the virtual display of different floor properties [12].

The experiment involved the synthesis of audio-tactile cues of

- **snow**, rendered by starting from a physically-based model called *crumpling*, that synthesizes in real time granular sequences of elementary hard impact events [13];
- **mud**, resulting from a specific tuning of soft impacts generated in real time.

These attributes are currently object of subjective evaluation, aimed at systematically understanding the sense of realism provided by the active shoes. However, for the purpose of this experiment only the macroscopic differences in the spectral and temporal features between such sounds were of interest, as any reference to the displayed material property was not exposed to the subjects during the walking task.

2.4. Procedure

The way the experiment was conducted was straightforward and the subjects were not introduced into the modality of the experiment. Subjects were asked to wear the active shoes, and to walk along a signed path 6.8 m long in a manner to express a specific emotion. Subjects were asked to walk with happiness, aggressiveness, sadness and tenderness. They did this for each of the four different grounds, wood panels, linoleum, iced snow, and muddy ground, for a total of 4x4 walks. The order of the combinations between emotions and grounds were randomized for each subject.

In order to maximize the possible amount of footsteps, subjects were asked to start their walk using their left foot, and data were logged starting from the first step with the right foot.

Sound examples produced by one subject in all sixteen conditions are available on-line[3].

At the end of the experimental session, subjects were asked about their personal data, and a question about how they experienced the test.

3. RESULTS

The active shoes allow for the logging of the values of several data-driven features. For the purpose of the experiment presented in this paper, five features were logged: heel-to-toe Inter-Onset-Interval (IOI) for both left and right foot, heel-to-heel IOI between right and left foot, toe-to-heel IOI for the right foot, and total footstep number.

First we conducted an analysis of variance. A two-way ANOVA, repeated measures, with the factors emotion and ground

[2] `http://www.puredata.org`

[3] `http://www.speech.kth.se/music/papers/2010_RB_ISon`

was conducted on the participants' values separately for each of the five features. Since the limited length of this paper, in the following we present the effects only for factors heel-to-toe Inter-Onset-Interval and heel-to-heel IOI as revealed by analysis. We also discuss the mean values for each of these factors and how they relate to results in previous studies on emotion expression in music performance.

3.1. Heel-to-toe IOI

Results show that there was a significant main effect of emotion for heel-to-toe IOI for the right foot, $F(1.45, 21.746) = 4.275$, $p = .038$, but not for the left foot. There was not significant main effect for factor ground. This suggests that subjects used different heel-to-toe IOI of the right foot for each emotional intention, and that their choices were independent from the particular ground (see Figures 3 and 4)).

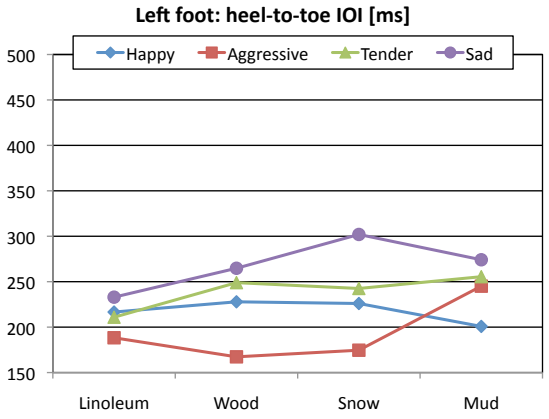


Figure 3: *Left foot mean heel-to-toe Inter-Onset-Interval [ms] for the four grounds, across all subjects, and for each emotion.*

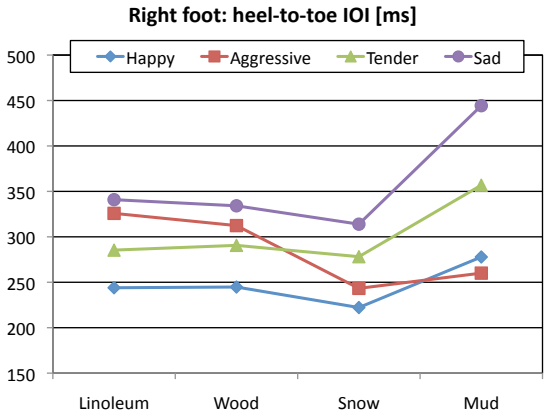


Figure 4: *Right foot mean heel-to-toe Inter-Onset-Interval [ms] for the four grounds, across all subjects, and for each emotion.*

3.2. Heel-to-heel IOI

The results show that there was a significant main effect of emotion, $F(1.361, 20.419) = 6.734$, $p = .011$, but not for factor ground. Also for this parameter, results suggest that subjects used different heel-to-heel IOI for different emotions, and that their choices were independent from the particular ground. Nevertheless, by looking to the mean values, it emerges that subjects had a tendency to change their behavior when the footstep sound changed (see Figure 5). In particular it seems that subjects had a tendency to walk slower, when the footstep sound was that of a muddy ground, and faster, when the footstep sound was that of a iced snow, independently from the emotion.

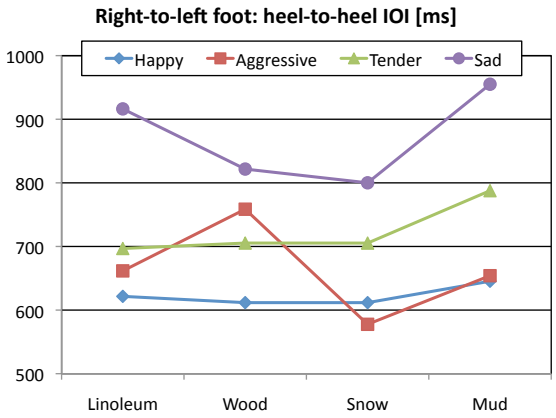


Figure 5: *Mean heel-to-heel Inter-Onset-Interval [ms], between right and left foot, for the four grounds, across all subjects, and for each emotion.*

3.3. Discussion

In general it has been observed the same tendency as in music performance, sad walking was slower than tender and happy walking. Aggressive walking had a more varied speed, and this could reflect previous results in music performance where tempo was not an important parameter for the communication of anger [6]. Results show that walkers used a more active walking style (faster pace) when the sound of the walking surface was characterized by an higher spectral centroid (e.g. iced snow), and a less active style (slower pace) when the spectral centroid was low (e.g. muddy ground). Harder texture sounds lead to more aggressive walking patterns while softer ones to more tender and sad walking styles. It can be noticed that the average value of heel-to-toe IOI for the right foot was larger than the corresponding one for the left foot (see 3 and 4), and that it had a trend similar to that of heet-to-heel IOI. This could suggest that subjects control the walking speed with the right foot.

The preliminary results reported above reflect the answers given by the subjects at the end of the experimental session. Subjects were asked the following final question: "Do you think there is a relationship between the sound produced by the shoes and the way you were walking during the experiment?", nine subjects answers *Yes*, six *No*, and one subject *I don't know*. To the follow-up question "What did influence you most?", ten subjects answered that they were mostly influenced by the emotion they were asked to express; two subjects found they were most influenced by the sound textures; two subjects were influenced by the test settings, the type of shoes, and the cables attached to them; one subject was more influenced by the instructions he was given; one subject was influenced by the dynamic of the steps.

4. CONCLUSIONS

Even if not statistically significant, it has been found that walking patterns can be influenced by the sound of the ground. In par-

ticular there are sounds that make people walk faster or slower independently from the emotional intention of the person. These results could be taken into account when designing the sonification of footstep sounds in virtual reality environments, such as new interactive floors where the sound feedback of footstep sounds can influence the behavior of users, or in rehabilitation and therapy applications when the control of the walking style of a client is a desired goal. For example it could be that subjects can be induced to walk with a faster or slower pace depending on the sound feedback: walking speed is often slower in persons with stroke and it could be modulated using sound as an alternative for example to methods using visual feedback [14].

The experiment was indeed conducted in an unnatural setting, and the nature of walking was influenced by the presence of cables connecting the shoes to the Arduino board. This has certainly introduced a bias in the data collected. This problem will probably be solved with future wireless versions of the shoes.

5. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Commission's Seventh Framework Programme under FET-Open grant agreement 222107 NIW - Natural Interactive Walking, from the European Network of Excellence MINET, and from KTH School of Computer Science and Communication, Dept. of Speech Music and Hearing.

6. REFERENCES

[1] Rolf I. Godøy and M. Leman, Eds., *Musical Gestures: Sound, Movement, and Meaning*, Routledge, 2009.

[2] Yon Visell, Federico Fontana, Bruno L. Giordano, Rolf Nordahl, Stefania Serafin, and Roberto Bresin, "Sound design and perception in walking interactions," *International Journal of Human-Computer Studies*, vol. 67, no. 11, pp. 947–959, jul 2009, Special issue on Sonic Interaction Design - SI: Sonic Interaction Design.

[3] B. Giordano and R. Bresin, "Walking and playing: What's the origin of emotional expressiveness in music?," in *ICMPC9 - 9th International Conference on Music Perception & Cognition*, M. Baroni, A. R. Addessi, R. Caterina, and M. Costa, Eds., Bologna, aug 2006, p. 149, Bonomia University Press (abstract), abstract only.

[4] Patrik N. Juslin and John A. Sloboda, Eds., *Music and emotion: theory and research*, Oxford University Press, Oxford (UK), 2001.

[5] Roberto Bresin and Anders Friberg, "Emotion rendering in music: range and characteristic values of seven musical variables," *CORTEX*, submitted.

[6] Tuomas Eerola, Anders Friberg, and Roberto Bresin, "Emotion perception in music: importance and additive effects of seven musical factors," *CORTEX*, submitted.

[7] T. Hermann and A. Hunt, "Guest editors' introduction: An introduction to interactive sonification," *Multimedia, IEEE*, vol. 12, no. 2, pp. 20 – 24, april-june 2005.

[8] A. Gabrielsson and P.N. Juslin, "Emotional expression in music," in *Handbook of Affective Sciences*, R.J. Davidson, K.R. Scherer, and H. Hill Goldsmith, Eds., pp. 503–534. Oxford University Press, New York, first edition, 2002.

[9] Yon Visell, Alvin Law, Jessica Ip, Rishi Rajalingham, Severin Smith, Jeremy R. Cooperstock, Gianpaolo Borin, Marco Civolani, Federico Fontana, Pietro Polotti, Rolf Nordahl, Stefania Serafin, and Luca Turchet, "Contact-based sensing methods for walking interactions," Deliverable 3.1, NIW project, 2009.

[10] V. Hayward, G. Dietz, A. Berrezag, N. Ouartiand, Y. Visell, and J. Cooperstock, "Haptic device engineering for walking interactions," Deliverable 2.1, NIW project, 2009.

[11] Stefano Delle Monache, Delphine Devallez, Carlo Drioli, Federico Fontana, Stefano Papetti, Pietro Polotti, and Davide Rocchesso, "Sound synthesis tools for sound design," Deliverable 2.2, CLOSED project, 2008.

[12] Antonio De Sena, Carlo Drioli, Federico Fontana, Stefano Papetti, Rolf Nordahl, Stefania Serafin, Luca Turchet, Yon Visell, Alvin Law, Jessica Ip, Rishi Rajalingham, Severin Smith, and Jeremy R. Cooperstock, "Auditory and haptic augmentation of floor surfaces," Deliverable 4.1, NIW project, 2009.

[13] Roberto Bresin, Stefano Delle Monache, Federico Fontana, Stefano Papetti, Pietro Polotti, and Yon Visell, "Auditory feedback through continuous control of crumpling sound synthesis," in *Proceedings of Sonic Interaction Design: Sound, Information and Experience. A CHI 2008 Workshop organized by COST Action IC0601*, Florence, Italy, apr 2008, pp. 23–28.

[14] Anouk Lamontagne, Joyce Fung, Bradford McFadyen, and Jocelyn Faubert, "Modulation of walking speed by changing optic flow in persons with stroke," *Journal of NeuroEngineering and Rehabilitation*, vol. 4, no. 1, pp. 22, 2007.

Audio-tactile Display of Ground Properties Using Interactive Shoes

Stefano Papetti[1], Federico Fontana[2], Marco Civolani[1],
Amir Berrezag[3], and Vincent Hayward[3]

[1] Università di Verona, Department of Computer Science
strada Le Grazie, 15 – 37134 Verona, Italy
{stefano.papetti,marco.civolani}@univr.it
[2] Università di Udine, Department of Mathematics and Computer Science
via delle Scienze, 206 – 33100 Udine, Italy
federico.fontana@uniud.it
[3] UPMC Universitè de Paris 06, Institut des Systèmes Intelligents et de Robotique
4 place Jussieu – 75005, Paris, France
amir.berrezag@isir.upmc.fr,hayward@cim.mcgill.ca

Abstract. We describe an audio-tactile stimulation system that can be worn and that is capable of providing the sensation of walking over grounds of different type. The system includes miniature loudspeakers and broadband vibrotactile transducers embedded in the soles. The system is particularly effective at suggesting grounds that have granular or crumpling properties. By offering a broad spectrum of floor augmentations with moderate technological requirements, the proposed prototype represents a solution that can be easily replicated in the research laboratory. This paper documents the design and features of the diverse components that characterize the prototype in detail, as well as its current limits.

Keywords: Interactive shoes, foot-based interfaces.

1 Introduction

As a recent complement to the design and engineering of machine interfaces for the study of human locomotion, balance and equilibrium in walking [1], researchers in human-computer interaction have begun to address questions concerning the interactive display to humans of psychophysical cues at ground level. Initially concentrating on the visual modality, with a focus on interaction scenarios ranging mainly between performing arts and gaming, researchers have then recognized the importance of non-visual ground cues as means to enrich the interaction with floors through the feet [2]. As a result, some radically novel paradigms have emerged enabling users to experience multimodal floor augmentations that – in addition to the mentioned application fields – are expected to play roles also in rehabilitation, critical labor environment simulations and for navigation aids for both normally gifted and impaired people.

R. Nordahl et al. (Eds.): HAID 2010, LNCS 6306, pp. 117–128, 2010.
© Springer-Verlag Berlin Heidelberg 2010

Among such paradigms, those which are grounded on an *ecological* approach to interaction design appear to be especially promising [3]. For its strong potential to result into "natural" interactions that furthermore do not need any specific training or cultural probing, this approach has received specific attention especially by designers of non visual displays, in which implicit sonic and vibrotactile signals can be set to operate at the periphery of the focus of attention [4]. Notable results that can be ascribed to this approach include active floor installations using vibrotactile devices, capable of conveying realistic sensations of snow- and ice-covered grounds to users walking over them [5].

Floor-based setups offer virtually unlimited physical space where to locate sensors and actuators. Furthermore, the networking of different physical components just requires to connect them together via a wired communication infrastructure. Power connections are generally not a problem as well for non-mobile interfaces, such as active floors. Conversely, the realization of an infrastructure of this kind poses serious technical questions when the same interaction paradigm is exported to a mobile interface, that is, a pair of shoes.

In the mobile case basically all physical components need to be tailored in order to minimize size, weight and power consumption, meanwhile guaranteeing an acceptable performance of the interface. Moreover they must be robust, since they are moved around by users engaged in walking or running tasks. Under such working conditions, the measurement of the force exerted by the foot over a sole, the real-time computations which are needed to generate an output from time-varying force data, and the consequent display of realistic sound and vibrations from the shoes by means of active components become more difficult to be realized and kept stable across time.

In the following of this paper, the current development state of our project on interactive shoes is detailed so as to provide the reader with an in-depth vision on their design and engineering. Section 2 describes the concept. Section 3 forwards to a parallel publication in these proceedings, on the accurate measurement and analog-to-digital conversion of force data using cheap sensors and processing. Section 4 presents the software that is employed to generate feedback in real-time. Section 5 illustrates the characteristics and positioning of the haptic actuators and loudspeakers used for providing feedback. Finally, Section 6 briefly reports about the performances and current limits of the interface, furthermore outlining ongoing and future work.

2 Design Concept

During everyday walking tasks we are continuously exposed to changes in the floor properties. The perception of level and inclination allows walkers to maintain correct equilibrium and posture. In parallel, auditory and tactile ground cues provide significant detail about the floor characteristics [3]. By influencing gait and walking gestures, these cues determine our level of familiarity and confidence with ground surfaces. Blind persons make intensive use of such cues during their everyday way-finding and landmarking tasks across familiar environments.

Our interactive shoes aim at simulating changes in ground surface, by augmenting otherwise neutral (i.e., flat and homogeneous) floors. Augmented reality is receiving increasing recognition by interaction designers, for its potential to enrich our surrounding environments with additional information. As opposed to substitution, augmentation can smoothly alert of changing conditions and, if the warning messages are carefully designed, it can support user's decisions through the presentation of implicit (especially non-visual) signals [4]. Actuated shoes represent a significant instantiation of this concept, as they can display audio-tactile ground cues for purposes ranging from support to way-finding and landmarking, as mentioned before, up to rehabilitation, entertainment and simulations of immersive reality.

In a preliminary work, we have prototyped a pair of interactive shoes provided with force sensors and small loudspeakers, capable of generating auditory feedback in real-time by foot pressure data acquired during a walking task [6]. This prototype was able to concentrate all computational and power resources inside a backpack that could be worn by users. The interface, hence, could be made strictly mobile, and consequently it allowed total freedom of movement to users who could navigate, even for a long time, across an environment. We chose rubber clogs as they could be easily "hacked" by cutting, grooving, drilling their body. Thanks to these shoes we could test initial design ideas, especially concerning the choice and location of different types of sensors. On the other hand, the range of feet fitting with a single pair of clogs was limited to few sizes. This fact excluded a number of potential users from testing the prototype.

One important lesson that we learned from this preliminary prototype was that, as somehow expected already at the conceptual stage, substitution is far more difficult than augmentation. In other words, it is not easy to "cancel" the floor upon which one is walking meanwhile providing alternative ground surface properties. On the other hand, it is relatively easier to add simulated properties to the real floor by superimposing a layer of virtual material to it. So, for instance, turning a wooden floor into a marble surface would in principle require to mask the resonances coming from the wood, a goal that is clearly hard to achieve. As opposed to this, covering a "dry" material such as concrete with virtual gravel, dry leaves, or snow, is at reach of current feedback design techniques [2].

The recent introduction of vibrotactile actuators has resulted in a new prototype (see Fig. 1), that has dramatically improved the realism of the simulations. Furthermore, we have switched from clogs to sandals, as they can fit with a larger range of foot sizes. Thanks to a better positioning of the sensors – refer to Section 3 – we got satisfactory force detections by fastening feet sized between 38 and 44 (Italian standard scale) through the three buckles every sandal is provided with.

On the other hand, vibrotactile actuators are more demanding than small loudspeakers in terms of power consumption. For this reason a wired connection had to be planned in the current setup, to feed the actuators with high amplitude signals provided by a couple of power amplifiers. Holding this physical constraint, it was logical to locate all the physical components that do not need to reside

Fig. 1. Current shoe-based interface prototype

on the shoes (i.e., acquisition board, computer and output signal interface) off the wearable part of the interface.

A schematic of the components forming the prototype is illustrated in Fig. 2. Every shoe is provided with two force sensors, one small loudspeaker and two haptic actuators, all depicted within the rectangle in dashed line. The next sections detail such components, step by step.

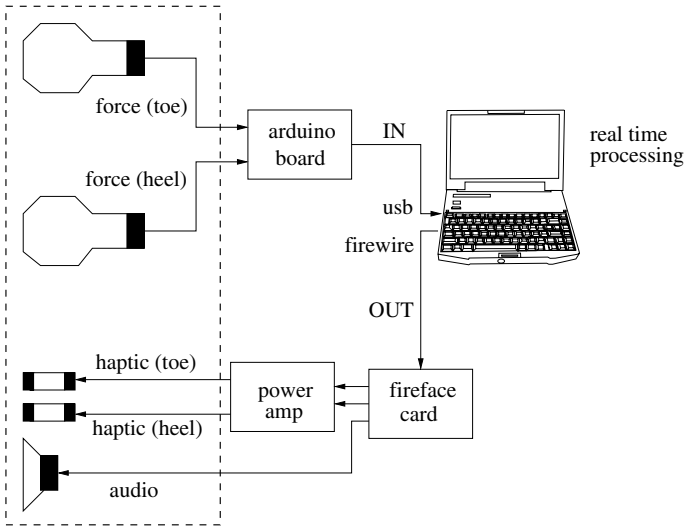


Fig. 2. Illustration of the prototype components. Sensors and actuators of one shoe are surrounded by the rectangle in dashed line.

3 Data Acquisition

The current implementation of the acquisition system presents a number of novelties that needed to expand this section into a self-contained paper, to which the reader is forwarded [7].

4 Real-Time Synthesis and Control

Interactive sounds are dynamic, informative, contextual, and occasionally bringing emotional content [8]. For this reason, sound design tools and systems should offer extensive control over the design process, generation, and interactive manipulation of sound.

Modern sound synthesis techniques, with their inherent parametric control, are able to handle continuous sound feedback in interactive contexts dynamically and effectively. In particular, physically-based synthesis models describe sonic interactions between resonating objects, and compute the resulting vibrations. Such vibrations are usually described in terms of signals accounting for local displacement and velocity of one or more objects. These signals can be directly sent to a loudspeaker. At the same time, they can drive a haptic device. Overall, physically-based models allow to maintain a tight coherence of the multimodal (in our case tactile and auditory) feedback.

For our specific purposes we made use of an open source software product called Sound Design Toolkit (SDT)[1] which is made of a set of physically-founded tools for designing, synthesizing and manipulating ecological sounds in real time. SDT consists of a library of external objects (*externals*) and programs (*patches*) for the real-time DSP environments Max/MSP and Pure Data (Pd). In particular, each external represents a physically-based or -inspired algorithm for sound synthesis or control, while the SDT patches combine those externals into complete control and sound models.

Below, first a brief description of the models providing audio-haptic feedback is given. Afterwards, an explanation is provided on how such models are controlled by making use of the data coming from the force sensors.

4.1 Physical Models of Contact Events

Contacts between solid bodies form a large class of sonic phenomena in everyday environments, and it has been shown that many contact interactions can be successfully simulated by using a flexible one-dimensional impact or friction model. Regarding this, short acoustic events like impacts can strongly gain or change in expressive content when set, for example, in an appropriate temporal sequence [9].

[1] The SDT is freely available from the following SVN repository: `https://svn.sme-ccppd.org/svn/sobs/SoundDesignTools/`

Soft Impact. The SDT *soft impact* model allows to synthesize the sound of impact on a soft surface, or of soft impact between two surfaces. The soft impact algorithm [10] exploits a rather simplified yet effective approach. Indeed the algorithm is physically-inspired, but it mainly focuses on the actual acoustic result: no actual interaction between objects is simulated, instead the algorithm exploits a filtered noise burst – representing a force signal – to excite a modal resonator [11]. The rationale behind the algorithm can be qualitatively justified considering that non-sharp contacts can be reduced to dense sequences of micro-impacts, thus in a sense discretizing the surfaces of the interacting objects as multiple contact areas. Also, the use of specifically filtered noise signals can be motivated considering that such micro-impacts can have a quasi-random character.

The available model parameters allow for full control of the modal resonator, an ADSR envelope (*attack time*, *decay time*, *sustain gain*, *sustain time*, *release time*), and the *cut frequencies* of two filters (respectively, high- and low-pass) which process the noise burst.

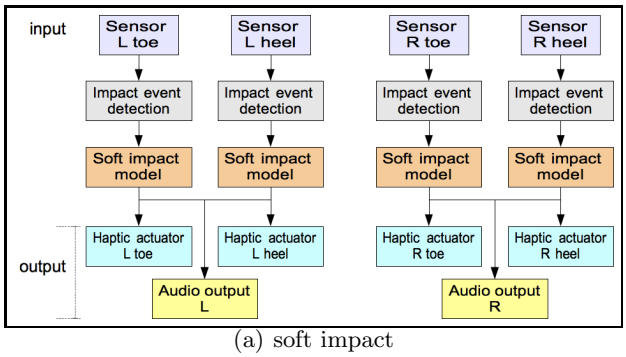
The *soft impact* model has been used to simulate the contact between a shoe and homogeneous floors or wet grounds, in particular providing two separate envelopes corresponding respectively to the heel and the toe. The aim was to add resonances to neutral floors in order to change the perceived ground material: for instance, a floor made of concrete can be augmented so as to resonate and vibrate like a wooden floor, or as a marshy ground.

Crumpling. The SDT *continuous crumpling* model [12,13] is the result of an *ad hoc* control layer superimposed to a low-level impact model. Similarly to the soft impact model, the crumpling algorithm does not actually model physical contacts between solid objects but, rather, time sequences of crumpling events, represented by groups of impact events. These sequences provide data that drive the evolution across time of the impact model parameters.

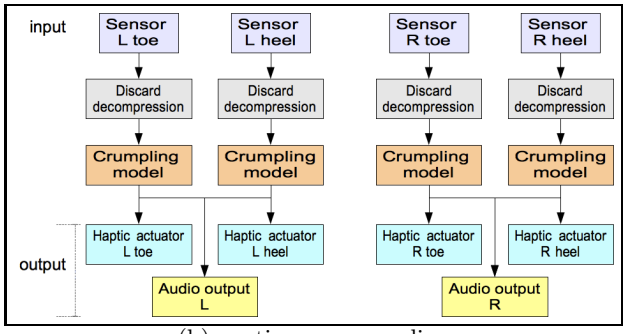
Both the temporal distribution of crumpling events and their own power follow stochastic laws which are derived from physics [14]. Such laws govern 1) the energy dissipation occurring during an impact, and 2) the temporal distribution of adjacent events. Each phenomenon exposes a characteristic parameter, resulting in the control of the average interval between events and the average power of impacts, respectively.

As for the actual implementation of the algorithm, the user is provided with several physically meaningful parameters, which allow to set: the applied *force* giving rise to crumpling events and being proportional to their average power, and the *resistance* put up by the material being crumpled, corresponding to the granularity of the latter.

In the perspective of simulating virtual aggregate grounds, the resistance parameter allows control of the compactness of the ground (the lower the resistance, the smoother and more uniform the sequence of crackling events), while the force parameter – being proportional to the energy of the micro-impact events – can be mapped directly to the pressure exerted by the foot on the ground.



(a) soft impact



(b) continuous crumpling

Fig. 3. Diagram explaining the use of the *soft impact* and *continuous crumpling* models. The diagrams display connections, data conditioning and signal outputs driving both the audio and the haptic feedback.

These models have already been successfully exploited for synthesizing sounds and haptic feedback in simulations of walking on aggregate grounds [15], furthermore they have been adapted for simulating the sound of walking on icy snow, creaking floors, brushwood.

4.2 Data Conditioning and Synthesis Control

The high-rate data stream coming from the Arduino accounts for the force signals measured at each foot's toe and heel. These four signals are received by a Pd patch communicating with the Arduino, and then routed to the SDT model in use. Moreover, the same patch can save the force signals as text files (especially useful for e.g. offline analysis), and later reproduce such files in a way to control the SDT models with prerecorded data.

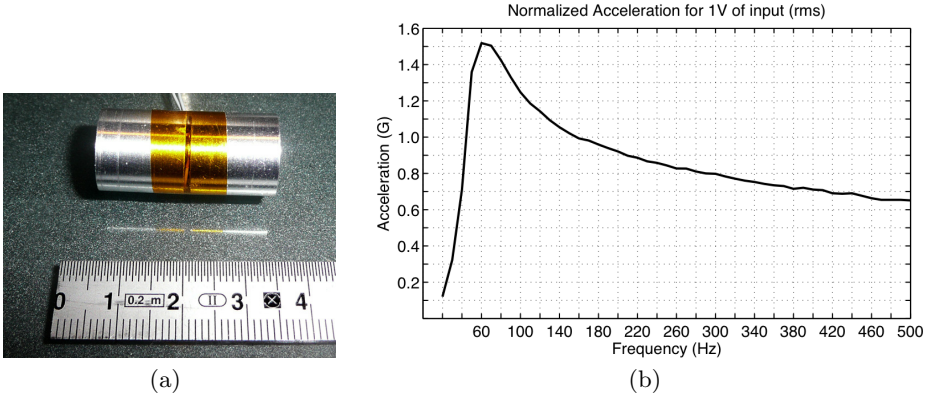
As for the *soft impact* model, the force signals coming from the sensors are used to trigger four individual noise bursts which excite a resonating object modeling the ground. More in detail, a simple algorithm is used which detects impact events occurring at both the heel and the toe. As soon as such an event is

detected, the current force value is used to set the amplitude of the corresponding noise burst: the larger this amplitude, the more energetic the micro-impacts on the resonating object. See the diagram in Fig. 3(a) for an explanation of the use of the *soft impact* model.

Thanks to its external interface, the *continuous crumpling* continuously maps gestures into force parameters: in fact, the model reacts only to variations in the applied force, this way filtering out constant components that do not reflect active interactions. In this regard, the force signals coming from the four sensors have been mapped to the force parameters of four separate instances of the *crumpling* model. By considering that aggregate grounds dynamically respond to a foot falling on them or scraping over them, it was hypothesized that the model should provide energy proportionally to the changes in the force, accounting for corresponding variations of the foot compression. To this end, a gate function has been employed which filters out negative variations of the force, thus excluding feedback when the foot depresses the ground. See the diagram in Fig. 3(b) for an explanation of the use of the *continuous crumpling* model.

5 Vibrotactile and Audio Feedback

Vibrotactile feedback is produced by two vibrotactile transducers embedded in the front and the rear of the shoe sole respectively (Fig. 4(a)) [16] (Haptuator, Tactile Labs Inc., Deux-Montagnes, Qc, Canada). Two cavities were made in the soles to accommodate these broadband vibrotactile actuators. 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. They were bonded in place to ensure good transmission of the vibrations inside the soles. When activated, vibrations propagated well in the light, stiff foam. An improved type of such actuators is currently being deployed. This new type uses



(a) (b)

Fig. 4. Haptic actuator: (a) Architecture. (b) Characteristic response (from TactileLabs Inc.).

multiple magnets instead of just one, in a configuration designed to concentrate the magnetic flux on a small region. They also benefit from a new suspension design that expends the low-end response significantly and boost the vibratory power despite being of 10 mm in diameter. Such a small size makes it possible to embed them in a variety of objects. Due to their design and structure they can be immersed in the inside of a sole, meanwhile they are able to support the weight of a person with a very little vertical deflection, yet free to vibrate in the horizontal direction because of its anisotropic structure.

In addition to vibrations, each shoe emits sounds from one Goobay Soundball Mobile battery loudspeaker mounted on the top buckle (see Fig. 1). These devices are provided with on-board micro-amplifiers, hence they can be connected directly to the audio card. As any small, low-power loudspeaker device, they exhibit unavoidable performance limits both in the emitted sound pressure level (2.4 W RMS) and low frequency cutoff (about 200 Hz).

The former limit does not pose problems as far as an ecological loudness level of the walking sounds is set. Large levels can be obtained at the cost of audible distortion and faster discharge time of the battery.

The latter limit has perceptual implications instead, as walking interactions can give rise to acoustic energy also in the low frequency. This energy, however, is the result of resonances that are consequence of slowly decaying, large wavelength vibrations occurring in certain floors when they are excited by a shoe impacting over them—think of a floor made with a layer of large wooden bars or panels, for instance. In this sense, forcing a sonic shoe to reproduce low frequencies has no ecological meaning. Due to the aforementioned physical mechanism, it is in fact the floor that should display unlocalized low-frequency sounds on a large area while mimicking the dissipation of mechanical energy that has been transferred when someone walks over it.

5.1 Low and High Frequency Routing

For their efficiency in the high frequency band the small loudspeakers radiate acoustic waves that, by defining a shortest path to the listener's ears and for their strong directivity, create a neat localization of the sound source in correspondence of the shoes. By arriving at the ears later than such waves, any other auditory stimulus is aggregated to the same source location by the listeners due to the known *precedence* effect.

As opposed to small loudspeakers, the haptic actuators generate components in the low frequency. A look to Fig. 4(b) in fact shows that the response of the actuators lies approximately above 50% of the 60 Hz peak value in the range 50-300 Hz. Part of the mechanical energy that they emit, in the form of vibrations across this range, is transmitted to the floor through the shoe sole. Although not comparable with the vibrations of a floor surface that naturally resonates at those frequencies, this energy propagates across the ground and can be heard in the proximity of the walking area. In conclusion, the haptic actuators mitigate the absence in the interface of mid-range loudspeakers and woofers, capable of adding sound energy in the low frequency band.

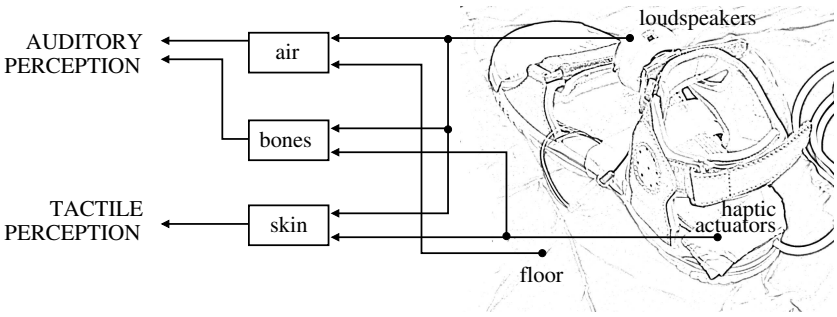


Fig. 5. Information paths connecting the actuated shoes to the auditory and tactile perception

In addition to this effect, a fraction of the vibrational energy that reaches the foot is probably transmitted to the auditory nerve through bone conduction. The set of potentially active information paths connecting the actuated shoes to our perceptual system are illustrated in Fig. 5.

To gain control of this information we have included some filters, all at software level, before routing the digital signal to the various channels in Fig. 2. Specifically, they include high-pass filters that cut off components below 20 Hz otherwise incoming to the output devices, and a smooth resonator that equalizes the response in Fig. 4(b) in the range 50-200 Hz, meanwhile cutting off the frequency components that lie outside the active range of the haptic actuator.

By significantly reducing the energy of the signal that is off the admissible band of the respective devices, these stages contribute to minimize the distortion artifacts and spurious frequency components otherwise introducing unpredictable effects in both the haptic actuators and small loudspeakers.

A systematic inspection of the contribution to listening and touch of the diverse sources of audio and vibrotactile signal, illustrated in Fig. 5, is far from being straightforward. If, on the one hand, their individual effects have been clearly experienced by anybody who informally tested the prototype, on the other hand a quantitative analysis of the perceptual impact of the vibrational energy propagating through the skin, in connection with the acoustic energy propagating through the air and conducted by the bones, would be worth opening another research chapter. In fact, not only the auditory and tactile percepts are difficult to be isolated in presence of a spatially concentrated audio and haptic stimulation, but also the coherence of the resulting (acousto-tactile) multimodal percept is probably conditioned by—perhaps even illusory—cross-modal effects occurring between the two modalities [17]. All this complexity at least does not affect the localization of sounds, for the precedence effect mentioned at the beginning of this section.

6 Conclusions and Future Work

One of the most important achievements of the proposed interface resides in the low latency of the feedback. Due to the excellent features of the audio card and its drivers, the efficiency of the real-time synthesis is not destroyed by bottlenecks encountered at the I/O stages of the system. More in general, this performance has come out as a result of trying many configurations among various operating and hardware systems, and by repeatedly refining the firmware of the acquisition board and the patches in Pd. Measurements made using an oscilloscope connected to the loudspeaker output, and triggered by an input from the audio card simulating a change in force signal, showed that this latency amounts to about 16 ms using the crumpling model.

The analysis of force data is not yet ready to resolve elaborate foot gestures and locomotion tasks that differ from simple walking. Improvements can be made on this analysis, but a general solution to the accurate detection of force during foot movement is not behind the corner if cheap sensors are employed. For instance, sensors like those we used in our prototype saturate at few hundred Newtons, corresponding to some tens of kilos. Hence, they cannot measure changes in the force occurring when users keep standing on the toes or the heel while moving their feet.

The cable connecting the shoes to the amplifiers represents a technological limit of the prototype, that will not be solved unless a new generation of power-efficient haptic devices becomes available.

In spite of these limits, the performance of the prototype is by all means encouraging. Ongoing research deals with the measurement of the skin displacement elicited by the haptic actuators, and with perceptual experimentation on a specific cross-modal effect induced by the multimodal feedback.

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. Iwata, H.: Haptic interface. In: Sears, A., Jacko, J.A. (eds.) The Human-Computer Interaction Handbook, 2nd edn. Lawrence Erlbaum Assoc., New York (2008)
2. Visell, Y., Fontana, F., Giordano, B., Nordahl, R., Serafin, S., Bresin, R.: Sound design and perception in walking interactions. Int. J. Human-Computer Studies, 947–959 (2009)
3. Giordano, B.L., McAdams, S., Visell, Y., Cooperstock, J.R., Yao, H., Hayward, V.: Non-visual identification of walking grounds. In: Proc. of Acoustics 2008 in J. Acoust. Soc. Am., vol. 123(5), p. 3412 (2008)
4. Norman, D.: The Design of Future Things. Basic Books, New York (2007)

5. Visell, Y., Cooperstock, J.: Design of a vibrotactile device via a rigid surface. In: Proc. of the IEEE Haptics Symposium, Waltham, MA (2010)
6. Papetti, S., Fontana, F., Civolani, M.: A shoe-based interface for ecological ground augmentation. In: Proc. 4th Int. Haptic and Auditory Interaction Design Workshop, Dresden, Germany, vol. 2 (2009)
7. Civolani, M., Fontana, F., Papetti, S.: Efficient acquisition of force data in interactive shoe designs. In: Proc. 5th Int. Haptic and Auditory Interaction Design Workshop (2010); Elsewhere in these proceedings
8. Bresin, R., de Witt, A., Papetti, S., Civolani, M., Fontana, F.: Expressive sonification of footstep sounds. In: Bresin, R., Hermann, T., Hunt, A. (eds.) Proc. of the Interaction Sonification workshop (ISon) 2010, KTH, Stockholm, Sweden (2010)
9. Rocchesso, D., Fontana, F. (eds.): The Sounding Object. Mondo Estremo (2003), http://www.soundobject.org/
10. Papetti, S.: Sound modeling issues in interactive sonification: from basic contact events to synthesis and manipulation tools. PhD thesis, University of Verona, Italy (2010)
11. Adrien, J.M.: The missing link: Modal synthesis, pp. 269–297 (1991)
12. Fontana, F., Bresin, R.: Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. In: Proc. Colloquium on Musical Informatics, Florence, Italy, pp. 109–114 (2003)
13. Bresin, R., Delle Monache, S., Fontana, F., Papetti, S., Polotti, P., Visell, Y.: Auditory feedback from continuous control of crumpling sound synthesis. In: CHI 2008 Workshop on Sonic Interaction Design, Florence, Italy. ACM, New York (2008)
14. Houle, P.A., Sethna, J.P.: Acoustic emission from crumpling paper. Physical Review E 54, 278–283 (1996)
15. Visell, Y., Cooperstock, J., Giordano, B.L., Franinovic, K., Law, A., McAdams, S., Jathal, K., Fontana, F.: A vibrotactile device for display of virtual ground materials in walking. In: Ferre, M. (ed.) EuroHaptics 2008. LNCS, vol. 5024, pp. 420–426. Springer, Heidelberg (2008)
16. Hayward, V., Dietz, G., Berrezag, A., Visell, N.O.Y., Cooperstock, J.: Haptic device engineering for walking interactions. Deliverable 2.1, NIW project (2009), http://www.niwproject.eu
17. Altinsoy, E.: Auditory-Tactile interaction in Virtual Environments. Shaker Verlag, Aachen (2006), http://www.ias.et.tu-dresden.de/akustik/Mitarbeiter/Altinsoy/data/15.pdf

Efficient Acquisition of Force Data in Interactive Shoe Designs

Marco Civolani[1], Federico Fontana[2], and Stefano Papetti[1]

[1] Università di Verona, Department of Computer Science
strada Le Grazie, 15 – 37134 Verona, Italy
{marco.civolani,stefano.papetti}@univr.it
[2] Università di Udine, Department of Mathematics and Computer Science
via delle Scienze, 206 – 33100 Udine, Italy
federico.fontana@uniud.it

Abstract. A four-channel sensing system is proposed for the capture of force data from the feet during walking tasks. Developed for an instrumented shoe design prototype, the system solves general issues of latency of the response, accuracy of the data, and robustness of the transmission of digital signals to the host computer. Such issues are often left partially unanswered by solutions for which compactness, accessibility and cost are taken into primary consideration. By adopting widely used force sensing (Interlink) and analog-to-digital conversion and pre-processing (Arduino) components, the proposed system is expected to raise interest among interaction designers of interfaces, in which the reliable and sufficiently broadband acquisition of force signals is desired.

Keywords: Force sensing, closed-loop interfaces.

1 Introduction

Assessment of human gait has been a well known issue in biomechanics and biomedical engineering. Early experiments in measuring and evaluating forces under the foot date back to the end of the 19th century (Beely, 1882; Momburg, 1908). Since that time, various techniques and methods have been implemented with the same purpose, involving the use of different kinds of floor-based electromechanic transducers [1,2,3].

Within this specific research topic, we are currently developing an instrumented shoe design that virtually reproduces ground surfaces, by interactively augmenting otherwise neutral (i.e., flat and homogeneous) floors [16]. By means of appropriate force sensing, real time signal processing, and final displaying of multimodal (audio and tactile) cues through portable loudspeakers and haptic actuators underfoot, this design is expected to serve purposes such as navigation in functional spaces, support to physical rehabilitation, and entertainment.

While planning the design of a sensing system for such shoes, we wanted to reach an acceptable trade off between accuracy of the recorded data and accessibility and cost of the technology. Furthermore, the same system had to deliver a

R. Nordahl et al. (Eds.): HAID 2010, LNCS 6306, pp. 129–138, 2010.
© Springer-Verlag Berlin Heidelberg 2010

continuous data flow affording tight close-loop interaction with walkers through the real time processing of such data. Recently, small and light force sensing devices have been developed, allowing for within the shoe, online force measurements. In parallel, the exploitation of powerful integrated electronic devices has led to the design of portable and wearable data acquisition systems [4,5,6,7].

Unfortunately, there is a general lack of simple yet reliable wearable force measurement systems. It is true that biomechanical and biomedical researchers and engineers can choose commercial solutions providing integrated hardware and software platforms for accurate underfoot force measurements and gait analysis (see for example the Tekscan product line). Besides their quality, hardly these solutions can be adapted for prototyping novel, flexible concepts of foot-floor interfaces. Furthermore, integrated technologies like these do not always guarantee low-level accessibility to raw data, nor they specify exact figures of latency for the transmission of the related signals.

For these reasons, we have worked on the in-depth optimization of an architecture based on popular hardware in the interaction design field, made by Interlink[1] and Arduino[2] and overall costing few tens of dollars. Its performance depends on the number and characteristics of the sensors and the specifications of the acquisition board: together, they set the dynamic range and the band of the acquired information. Besides its applicability to instrumented shoes, the same architecture is of potential interest in all situations where accurate force data must be acquired and processed in real time.

This paper describes in detail its design and operation inside our prototype.

2 Sensors

Two Force Sensing Resistors (FSR) have been inserted between a sandal and a removable sole, forming an additional layer on top of the sandal itself (Fig. 1). Interlink equipment was chosen, for its versatility and popularity among interaction designers [8]. In this sort of "sandwich" configuration, the two sensors detect forces respectively in correspondence of the toe and the heel.

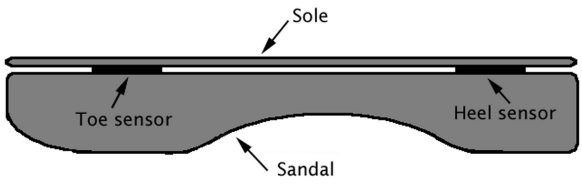


Fig. 1. Mechanical assembly of the shoe. For both toe and heel the Interlink FSR model 402 has been used.

[1] http://www.interlinkelectronics.com/
[2] http://arduino.cc/

Each FSR is connected in series with a fixed resistor. Together, they realize a voltage divider. The four dividers, two for each shoe, are finally connected to the first four analog inputs of an Arduino Duemilanove board, another quite popular device. Fig. 2 illustrates this connection.

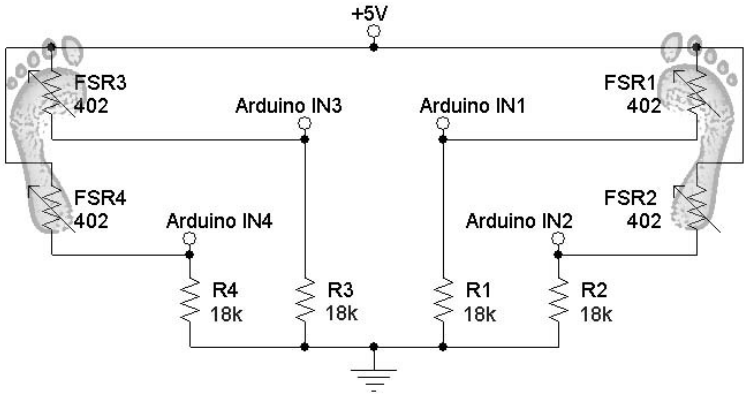


Fig. 2. Schematic of the conditioning circuit. The 5 V and ground pin are accessible from the Arduino board.

A previous version of the prototype used different sensor models and resistance values in the voltage divider [9]. In the following we will see that optimizing these two components leads to substantial improvements.

2.1 Characterization of the FSR

Interaction designers know that Interlink force sensors have good response in front of rapid and large changes in the applied force [10], and they are also especially robust [11]. Thus, they are suitable for detecting forces such as those typically exerted by the foot during gait. For their small size, the model 400 and model 402 are the only Interlink products that can find place inside a shoe without risk of breaking due to excessive mechanical deformation. Although quite similar, the model 400 overall exhibits a different behavior compared to the 402 [12]: in our previous prototype, a good balance between front and rear sensitivity had been achieved by placing a model 400 under the heel and a 402 under the toe.

While changing shoe model during the development of the current prototype, we found that the model 400 reaches mechanical saturation much faster than the 402. This overall unbalanced the acquisition of the data. Clearly, the foot pressure had a different distribution in the new sole.

Measurements have then been made to characterize their behavior. The same measurements furthermore proved useful to characterize the nonlinear behavior of the overall system, caused by the electrical coupling between the FSR and the

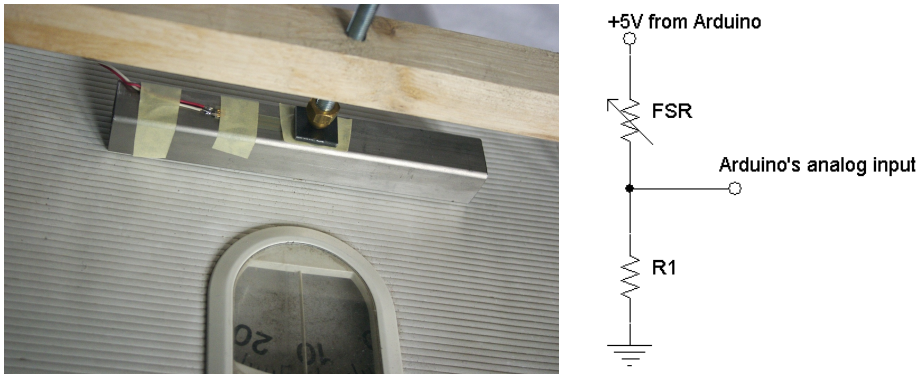


Fig. 3. Characterization of the FSR. (left) Mechanical setup. (right) Voltage divider for the acquisition of measured forces.

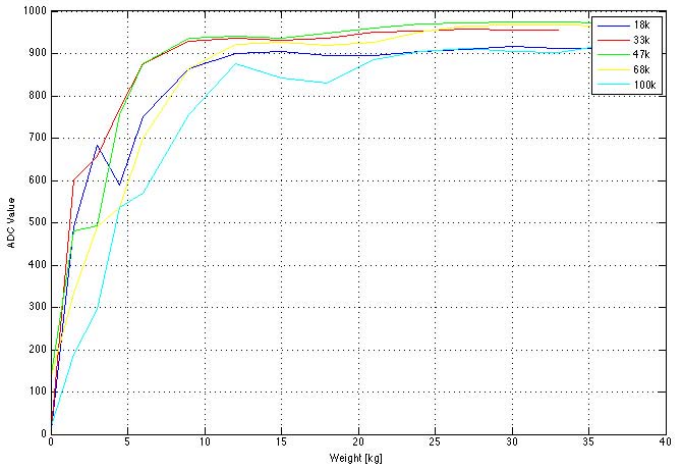
analog-to-digital converter (ADC) in the Atmel microcontroller on-board the Arduino [13].

A manual wooden press has been built, and mounted onto a weighing scale. This press had an interchangeable termination, providing mechanical matching with the active area of the sensor. With this simple setup, we could read the value of the force exerted by the press over the sensor. Fig. 3 (left) shows the setup.

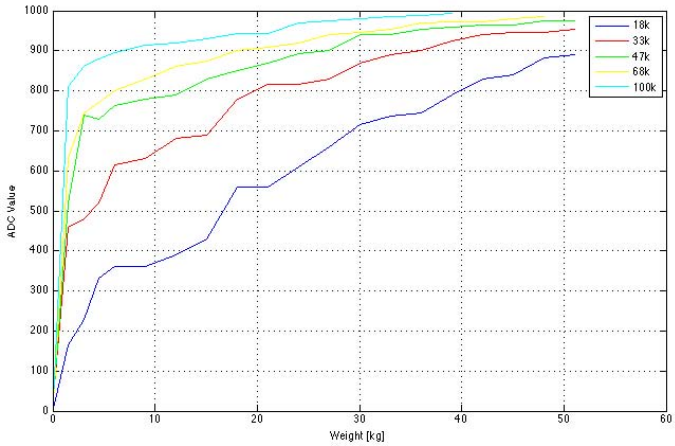
The voltage divider containing the FSR was connected to the Arduino analog input as shown in Fig. 3 (right), then pressed. As a result of this simple test, several sets of curves have been obtained for each sensor mapping force into ADC values, for changing values of the resistance R1 in the voltage divider. In practice, each curve represents the relationship between the applied force and the ADC output (ranging from 0 to 1023) for a specific value of R1. Fig. 4 displays the corresponding plots, respectively for sensors model 400 (a) and 402 (b). Their inspection shows that the system is more linear when using the model 402.

All measurements were made by choosing standard resistance values for R1 below 100 kΩ. Smaller values result in a smoother response that furthermore reduces the dynamic range, particularly with the model 402. This means that when the FSR reaches mechanical saturation, the voltage across R1 is below the maximum input voltage of the ADC (5 V). Conversely, values amounting to more than 100 kΩ introduce a strong nonlinearity. Choosing R1 = 18 kΩ results in a good compromise between range and linearity of the system.

In conclusion, the model 402 has been preferred for all sensing points and the resistance in the voltage dividers has been set to 18 kΩ: these choices have substantially improved the sensitivity of our instrumented shoes. Besides our specific application case, a simple measurement experience like the one described here can reward each time an FSR must be coupled to an ADC in an acquisition system of interest.



(a) Model 400.



(b) Model 402.

Fig. 4. Force/ADC maps for changing values of R1

3 Firmware

A good firmware program can enable the cheap hardware on-board the Arduino to perform as an effective front-end for data acquisition from the sensors. A reliable analog-to-digital conversion along with a low-latency transmission of the acquired data are in fact necessary to achieve this goal, when designing a real time interactive system providing instantaneous feedback. The essential tasks of the Arduino are illustrated in Fig. 5.

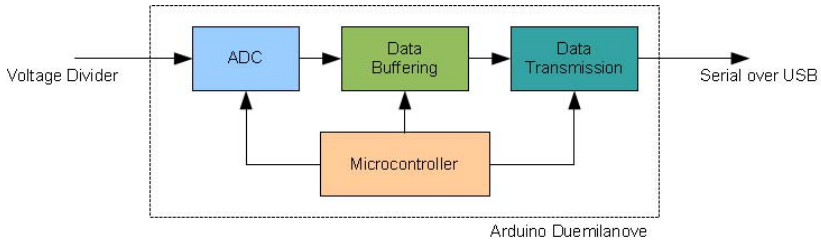


Fig. 5. Blocks of the data acquisition module

The firmware employed in the previous prototype of the interactive shoes included an application program, that repeatedly read the four values coming from the respective analog inputs through a traditional polling procedure nested inside a main loop [9]. Then, data were filtered in such a way that a packet containing the values was sent only if two subsequent samples differed at least by a given threshold. Concerning the serial transmission, we sent to the host packets made of four ADC values separated by a space character. Each packet ended with a separator (end of line) character.

Both the polling sequence and the transmission routine have been written using functions and procedure calls provided by the Arduino SDK. The use of this framework allows to write code very easily, thanks to the good level of abstraction provided by the embedded C++ API. Unfortunately the compiled program is often not efficient enough, especially if running on applications requiring a constantly low latency during the acquisition of uniform data flows.

In our case, the unconditioned use of the API affected the reliability of the analog-to-digital conversion and the constancy of the transmission rate of the serial connection over USB from the Arduino to the host. In fact, the previous firmware used the `analogRead` function, which, as a normal practice in acquisition procedures of this kind, was called for every channel during polling. As a consequence, the sampling frequency could not be kept stable: the main loop in fact can be interrupted at any moment, and it takes an unpredictable time for the system to return to the application program.

The transmission was managed using the `Serial.print` function, which prints out data to the serial port in the form of ASCII characters. In particular, every number digit was printed using the respective character. Once again, the API-based solution is quite standard and easy to be coded, but poorly performing. For instance, transmitting four 10-bit values requires, in the worst case (i.e., all the four values greater than 1000), $8 \cdot 4 \cdot 4 + 8 \cdot 4 = 160$ bits.

Interrupts can introduce heavy jitter. In the Arduino this problem is a consequence of the weak control of the low level structures inside the microcontroller through the API. In essence, the Arduino is not provided with resources capable to handle both processes in parallel:

- the ATMega168 is a single-thread machine, thus when the CPU is handling the ADC interrupt it cannot (among other things) transmit data over the serial connection;
- the Arduino USB connection is controlled by a FTDI FT232RL chip, which *emulates* a serial RS-232 connection over a USB line.

A consequence of the first point is that sampling with an exceedingly high frequency is unsafe since some values would be inevitably lost. The second point deals with the buffers that are present on the transmitter (inside the FTDI chip) and the receiver (in the USB controller driver, run by the operating system of the host computer). These buffers are governed by independent schedulers and procedures [14], and a low-level debugging of their mutual activity is a hard task.

As it often happens in this kind of architectures, the best performance is achieved by writing part of the code in assembly language at the cost of decreased readability and portability. A solution in between, that can be adopted with AVR-compliant microcontrollers like the ATMega168, is to write programs encoding API calls and AVR Libc instructions[3] together.

In the current firmware, we have set the ADC to work in *free-running* mode. Under this mode, the ADC interrupts the program only when the acquisition on the selected channel is completed. The sampling frequency can be selected by setting the ADC *prescaler* bits (ADPS) in the ADCSRA register. In free-running mode, a single conversion takes 13.5 CPU cycles, thus the sampling frequency turns out to be equal to $Fs = (16/13.5)/P$ MHz, where P represents the value of the prescaler.

Furthermore, a custom transmission protocol has been implemented [15]: binary data are transmitted instead of ASCII values using the `Serial.write` function. Every packet is formed by two bytes, B_{MSB} and B_{LSB}, such that their juxtaposition is $B_{MSB}\,B_{LSB}$ = 1CCCCVVV 0VVVVVVV. In this structure, the four bits denoted with C encode the channel number (16 channels are allowed), whereas the ten bits denoted with V encode the measured value thus guaranteeing sufficient accuracy for the acquired force. In the end, each sensor is assigned to a different channel. Hence, sending four values (one for each channel) using this protocol occupies a constant packet size, equal to $8 \cdot 2 \cdot 4 = 64$ bits.

3.1 Uniform Sampling of Force Data

We opted for a reliable, although not necessarily optimal assignment of the system variables in the new firmware, by empirically determining the transmission rates of in absence of drop-outs.

An oscilloscope was connected to a digital pin of the board. The pin was set at the beginning of the ADC routine, and was reset at the end of the same routine. The ADC routine called the transmission procedure only when the transmission buffer was full.

[3] `http://www.nongnu.org/avr-libc/`

When sending a cluster of bytes (i.e., a buffer) with the `Serial.write` function, the Atmel performs a serial transmission of data to the FTDI chip's buffer. For what we said above, data are sent on the USB line according to specific handshaking and buffering policies that in principle vary with the transmitter/receiver protocol.

The observations made on the oscilloscope uncovered that the serial transmission from the Atmel to the FTDI chip requires a large amount of CPU time. The values coming from the ADC which fall into this bottleneck are lost. For this reason we set the prescaler to $P = 128$, even if lower values are allowed.

By setting the buffer size to 2 bytes we obtained a regular sequence of impulses from the probed pin, testifying uniform sampling. In practice, this buffer size ensured that the transmission time is shorter than the sampling time. At this point, a linear relationship between the transmission rate and sampling frequency can be figured out. Table 1 lists possible choices complying with this relationship.

The values showed in Table 1 must be divided by the number of input channels, if a polling procedure is implemented. In our application, involving four channels, the sampling frequency per channel is $F_s = 5882/4 \approx 1470$ Hz. Considering that the FSR's have a response time of about 2 ms [12], a latency that is certainly smaller than any human response to psychophysical cue changes, the obtained sampling frequency is well above twice as much the Nyquist limit of $1/0.002 = 500$ Hz.

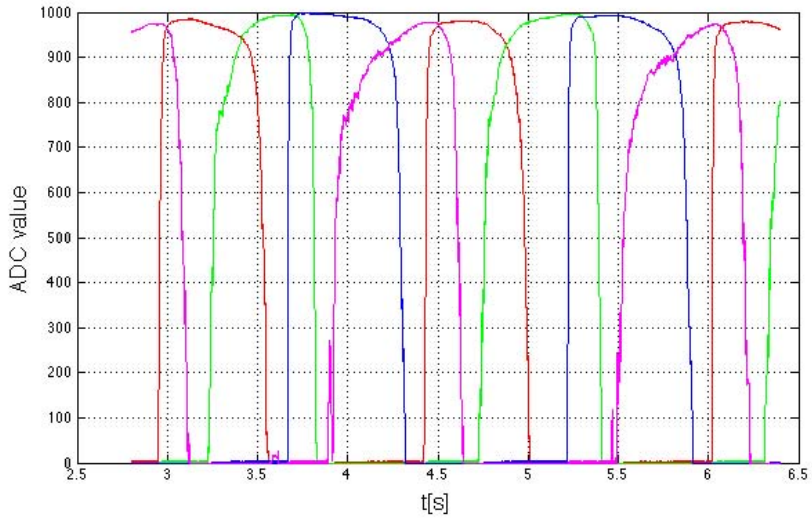
Table 1. Relationship between sampling frequency (F_s) and transmission rate (TXrate). Values per single channel. Prescaler set to 128.

TXrate [baud]	9600	19200	38400	57600	115200
T_s [ms]	2.1	1.04	0.52	0.36	0.17
F_s [Hz]	476	961	1923	2778	5882

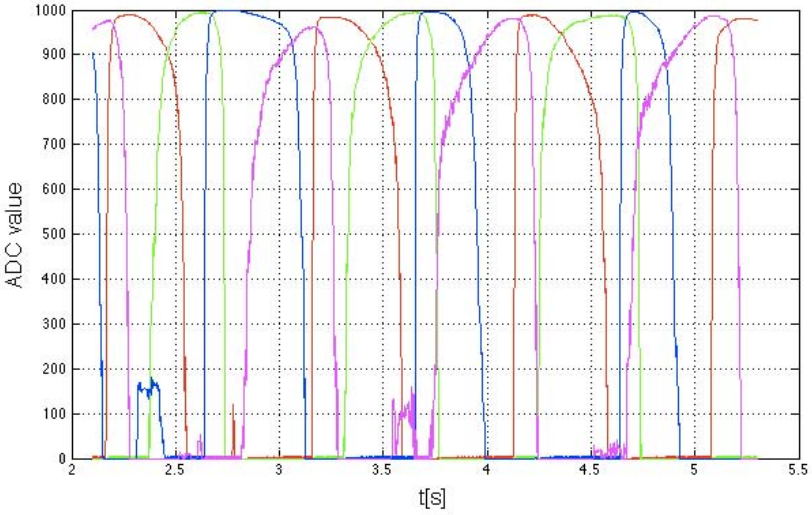
4 Results and Conclusions

Fig. 6 shows some plots of force signals recorded using the proposed system. These plots provide evidence of a substantial absence of noise and jitter in the data. If perhaps not accurate enough for applications where extremely high resolution is mandatory, such data is suitable for most interaction design and other applications.

In our case the force signals are sent to the host computer for further processing [16]. Still, for many other acquisition processes requiring similar performance and cost, we think that designers and others can make profitable use of the solutions and tests described in this work.



(a) Slow walk.



(b) Fast walk.

Fig. 6. Force plots. Red = left heel, green = left toe, blue = right heel, magenta = right toe.

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. Stott, J.R.R., Hutton, W.C., Stokes, I.A.F.: Forces Under the Foot. Journal of Bone and Joint Surgery 55-B, 335–345 (1973)
2. Manley, M.T., Solomon, E.: The Clinical Assessment of the Normal and Abnormal Foot During Locomotion. Journal of Prosthetics and Orthotics 3, 10–110 (1979)
3. Soames, R.W., Blake, C.D., Stott, J.R.R., Goodbody, A., Brewerton, D.A.: Measurement of pressure under the foot during function. Journal of Medical and Biological Engineering and Computing 20, 489–495 (1982)
4. Pollard, J.P., Quesne, L.P.L., Tappin, J.W.: Forces Under the Foot. Journal of Biomedical Engineering 5, 37–40 (1983)
5. Zhu, H., Maalej, N., Webster, J.G., Tompkings, W.J., Bach-Y-Rita, P., Wertsch, J.J.: An Umbilical Data-Acquisition System for Measuring Pressures Between the Foot and Shoe. IEEE Transactions on Biomedical Engineering 37, 908–911 (1990)
6. Faivre, A., Dahan, M., Parratte, B., Monnier, G.: Instrumented shoes for pathological gait assessment. Mechanics Research Communications 31, 627–632 (2004)
7. Morris, S.J.: A Shoe-Integrated Sensor System for Wireless Gait Analysis and Real-Time Therapeutic Feedback. PhD thesis, Massachusetts Institute of Technology (2004)
8. Miranda, E., Wanderley, M.: New Digital Musical Instruments: Control and Interaction Beyond the Keyboard. AR Editions (2006)
9. Papetti, S., Fontana, F., Civolani, M.: A shoe-based interface for ecological ground augmentation. In: Proc. 4th Int. Haptic and Auditory Interaction Design Workshop, Dresden, Germany, vol. 2 (2009)
10. Hollinger, A., Wanderley, M.M.: Evaluation of commercial force-sensing resistors. Unpublished report 1, Input Devices and Music Interaction Laboratory (IDMIL), Music Technology Schulich - School of Music, McGill University Montreal, QC, Canada (2006), http://www.idmil.org/publications
11. Vecchi, F., Freschi, C., Micera, S., Sabatini, A., Dario, P., Sacchetti, R.: Experimental evaluation of two commercial force sensors for applications in biomechanics and motor control. In: Proceedings of the 5th Annual Conference of the International Functional Electrical Stimulation Society, Aalborg, Denmark, p. 44 (2000)
12. Interlink: Force sensing resistor integration guide and evaluation parts catalog. Datasheet 90-45632 Rev. D, Interlink Electronics, 546 Flynn Road, Camarillo, CA 93012, USA (2009), http://www.interlinkelectronics.com/library/media/papers/pdf/fsrguide.pdf
13. Atmel: Atmega168 datasheet. Datasheet 2545RAVR07/09, Atmel Corporation, 2325 Orchard Parkway, San Jose, CA 95131, USA (2009), http://www.atmel.com/dyn/products/datasheets_v2.asp?family_id=607
14. FTDI: Data throughput, latency and handshaking. Application note AN232B-04 Rev. 1.1, Future Technology Devices International Ltd., Seaward Place, Centurion Business Park, Glasgow, G41 1HH, UK (2006), http://www.ftdichip.com/Documents/AppNotes/AN232B-04_DataLatencyFlow.pdf
15. Visell, Y., Law, A., Ip, J., Rajalingham, R., Smith, S., Cooperstock, J.R., Borin, G., Civolani, M., Fontana, F., Polotti, P., Nordahl, R., Serafin, S., Turchet, L.: Contact-based sensing methods for walking interactions. Deliverable 3.1, NIW project (2009), http://www.niwproject.eu
16. Papetti, S., Fontana, F., Civolani, M., Berrezag, A., Hayward, V.: Audio-tactile display of ground properties using interactive shoes. In: Proc. 5th Int. Haptic and Auditory Interaction Design Workshop (2010); Elsewhere in these proceedings

PHYSICS-BASED AND SPIKE-GUIDED TOOLS FOR SOUND DESIGN

Kamil Adiloğlu,
METISS group IRISA-INRIA
Rennes Cedex, France
NI group, Berlin Institute of Technology
Berlin, Germany
kamil.adiloglu@irisa.fr

Carlo Drioli, Pietro Polotti,
University of Verona,
Verona, Italy
carlo.drioli@univr.it
pietro.polotti@univr.it

Davide Rocchesso, Stefano delle Monache,
IUAV University of Venice
Venice, Italy
roc@iuav.it
stefano.dellemonache@gmail.com

ABSTRACT

In this paper we present graphical tools and parameters search algorithms for the timbre space exploration and design of complex sounds generated by physical modeling synthesis. The tools are built around a sparse representation of sounds based on Gammatone functions and provide the designer with both a graphical and an auditory insight. The auditory representation of a number of reference sounds, located as landmarks in a 2D sound design space, provides the designer with an effective aid to direct his search for new sounds. The sonic landmarks can either be synthetic sounds chosen by the user or be automatically derived by using clever parameter search and clustering algorithms. The proposed probabilistic method in this paper makes use of the sparse representations to model the distance between sparsely represented sounds. A subsequent optimization model minimizes those distances to estimate the optimal parameters, which generate the landmark sounds on the given auditory landscape.

1. INTRODUCTION

The state-of-the-art sound synthesis techniques offer a high degree of naturalness and expressiveness. Sometimes, however, this comes at a price of a relevant complexity in both the control of real time performance, and the parametric tuning required in the sound design process. One such case is physical modeling, in which the process that produces the sound is represented by means of simplified equations of the underlying physical laws. As a result, the parameters involved in the modeling become in principle easy to understand, because they have real counterparts, and easy to control, because our sensorial experience mediates the action-reaction patterns to which they relate [1]. Nonetheless, the most accurate and expressive models available today are often described in terms of detailed physical relations, which makes the parameters inaccessible or unfeasible to non specialists and, anyhow, results in complex control mapping issues. Moreover, the nonlinear nature of the phenomena under observation may sometimes lead to numerical schemes that do not always reflect the behavior of the real systems in the whole parameters space. These considerations motivate the search for new tools to aid the parameter tuning of synthesis tools in general. Such tools are of particular interest in the case of physical modeling audio synthesis.

In this paper, we propose a graphical tool equipped with clever parameter optimization algorithms for the timbre space exploration and interactive design of complex sounds by physical modeling synthesis. We exploit a sparse auditory representation of sounds based on Gammatone functions. This representation provides the designer with a graphical and an auditory insight that may be used in place of, or combined with, the set of low-level physical parameters of the models. A supervised framework makes use of the sparse representations to model the distance between two sounds. These distances are incorporated in a subsequent probabilistic estimation method based on the expected loss principle to select the optimal parameter values to support the sonic landmark paradigm. The graphical interface incorporates the auditory representation of the landmarks located in a 2D sound design space, and provides an effective aid to direct the search along the paths that lie in the proximity of the most inspiring sonic landmarks.

The use of terminology and metaphors referring to the environment and landscapes has a rather old tradition in the field of sounds perception, especially when referred to the perception of ecological and everyday sounds. In the late 70's, the Canadian composer R. Murray Schafer introduced the term "soundscape", defined as the auditory equivalent to landscape [2], and Barry Truax published his Handbook for Acoustic Ecology [3]. The term soundscape perception is also used in a scientific context to characterize how inhabitants perceive, experience and appraise their sonic environment.

Our use of the terms "sonic landscape" and "sonic landmark" however refers to a particular organization of sounds in a 2D sonic space. In a related work, Momeni and Wessel proposed a tool to construct and access multidimensional parameter spaces through 2D (or 3D) interfaces [4]. The interpolation between a set of examples was obtained by a mixture of gaussian kernels whose weight and width could be adjusted. The user could position the example objects in arbitrary positions and visual feedback was implicit in the system. Although similar in some extent, our proposed control layer relies on auditory representations of sonic landmarks, and it

is especially suited to exploring wildly non-linear sonic spaces.

We rely on a sound synthesis framework based on a class of physical models for everyday sounds, which includes low level events (impacts and friction) and high-level events (bouncing, breaking, rolling, crumpling). This Physically-based Sound Design Tools package (SDT from now on) has being developed and supported within a number of EU funded research projects on audio synthesis and sound design (Sounding Objetcs (SOb), Closing the Loop of Sound Evaluation and Design (CLOSED), Natural Interactive Walking (NIW))[5, 6].

The paper is organized as follows: first, a description of the sound synthesis engine and of the spike-based auditory representation is given in Section 2; Section 3 describes the interpolation sound space and the graphical tool proposed to assist the sound designer. In this section, some sound design examples obtained with the tool are also illustrated. The subsequent section introduces the clever parameter search methods. After presenting some optimization results, the paper concludes in Section 5 with final remarks and future issues.

2. DESIGN AND IMPLEMENTATION OF THE SPIKE-BASED PARAMETER INTERPOLATION TOOL

The interactive interpolation tool presented is organized as a client-server distributed application, in which the client side hosts the user graphical front-end based on the auditory spike sound representation, and the server side hosts the SDT audio synthesis engine. In the following we provide some details on the sound synthesis algorithm used to generate the reference and the new sounds, on the spike analysis framework used to graphically represent the sounds, and the properties of the interpolation in the 2D space.

2.1. The sound synthesis engine

The sound synthesis chosen to illustrate our design tool is a physical modeling implementation of friction sounds synthesis, included in the aforementioned SDT package.

The scope of the SDT is to provide a platform of sound synthesis tools that interaction designers can exploit in their sketching activities and that can be run on common real time software such as Max/MSP and Pd. The aim is also to provide the patches with a set of side tools to easily manage projects and that can be of help when working with acquisition boards and sensors. Most of the models contained in the SDT, aimed at reproducing sounds from solid objects interaction, are structured as two resonating objects interacting by means of a contact model. The friction model specifically referred to here relies on a description of the average behavior of a multitude of micro-contacts made by hypothetical bristles extending from each of two sliding surfaces. A modal decomposition is adopted for both interacting objects, leading to a first parametric subset including mode frequencies, decay factors and gains. The remaining low-level parameters are related to the interaction mechanisms and to the interaction force specification. Furthermore, the friction model needs to be activated by some stochastic model controlling the temporal behavior of the external rubbing force and the pressure on the sliding object (called the rubber). To gain an insight of the phenomenological role of the low-level physical parameters of the friction model, and of what a sound designer can be asked to deal with, a description is given in Table 1. Further details on the friction model can be found in [5], Chap. 8.

Sym.	Physical Description	Phenomenological Description
σ_0	bristle stiffness	affects the evolution of mode lock-in
σ_1	bristle dissipation	affects the sound bandwidth
σ_2	viscous friction	affects the speed of timbre evolution and pitch
σ_3	noise coefficient	affects the perceived surface roughness
μ_d	dynamic friction coeff.	high values reduce the sound bandwidth
μ_s	static friction coeff.	affects the smoothness of sound attack
v_s	Stribeck velocity	affects the smoothness of sound attack
f_N	pressure on rubber (normal force)	high values give rougher and louder sounds
f_P	external rubbing force (parallel force)	affects the perceived speed of rubber
ω_i,g_i, τ_i	frequency, gain, and decay of i-th resonator mode	affects perceived material, shape and dimension of the interacting objects

Table 1: A phenomenological parameter guide to friction model.

2.2. Spike representation

In a classical signal representation approach, overlapping discrete blocks of signal are used. However, in this method, in particular for non-stationary signals, a small shift of a block can cause a large change in the representation, depending on where an acoustic event falls within the block. A sparse, shiftable representation method based on atom-like filter functions can solve the problem. Hence, following [7], a sound signal $x(t)$ can be approximated as a linear combination of K filter functions, the so-called spikes. In this study, we used the Gammatone functions $\gamma_{f_k,t_k}(t)$ with amplitudes a_k and residual $\epsilon_{K+1}(t)$:

$$x(t) \quad = \quad \sum_{k=1}^{K} a_k \gamma_{f_k,t_k}(t) + \epsilon_{K+1}(t). \qquad (1)$$

A Gammatone function is defined by its center frequency f_k and its filter order. In a Gammatone filter bank, the filter center frequencies are distributed across the frequency axis in proportion to their bandwidth. The shape of the magnitude characteristics of the fourth order Gammatone filter approximates the magnitude characteristics of the human auditory filter in a proper way [8]. Hence, these filter functions have a biological background, that makes them more promising for perceptually relevant tasks, and optimal jointly from a computational and representational point of view, with respect to any other orthogonal or redundant time-frequency representation. Furthermore, it has been shown that the gammatone filters are highly efficient for natural sounds including the environmental sounds. Even though we have a synthesis model and not natural recordings, the physical modeling paradigm to model environmental sounds (in our case the friction model) encouraged us to use the gammatone filters as the basis of our decomposition. We used the implementation within Slaney's auditory toolbox [9] for the gammatone filterbank, which generates filters with the center frequencies and bandwidths calculated depending on the equivalent rectangular bandwidth (ERB) model [10].

Each spike $s_k = (t_k, f_k, a_k)$ is composed of the temporal offset t_k, the center frequency f_k of the corresponding Gammatone filter and the amplitude a_k.

By varying K within Equation 1, the SNR values of the spike code, and correspondingly the sparsity of the representation, can be changed. Increasing K increases the SNR of the spike code. Small K (high sparsity) decreases the SNR.

In Figure 1, we show the spike code of the simplest sound generated by the SDT algorithms, i.e. an impact sound. In this

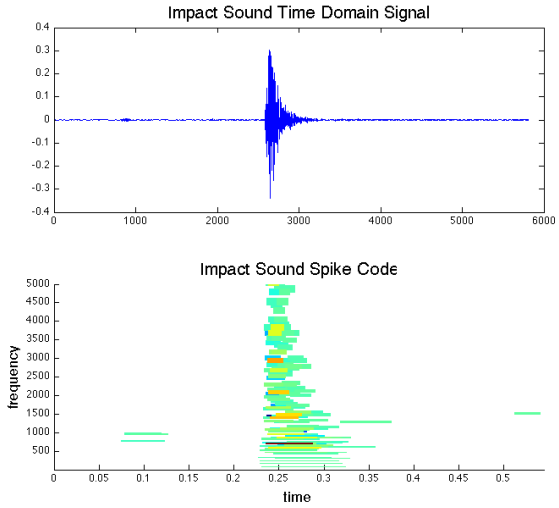


Figure 1: The sound wave and the spike code of an impact sound. This picture indicates how the spike code captures the skeleton of the sound.

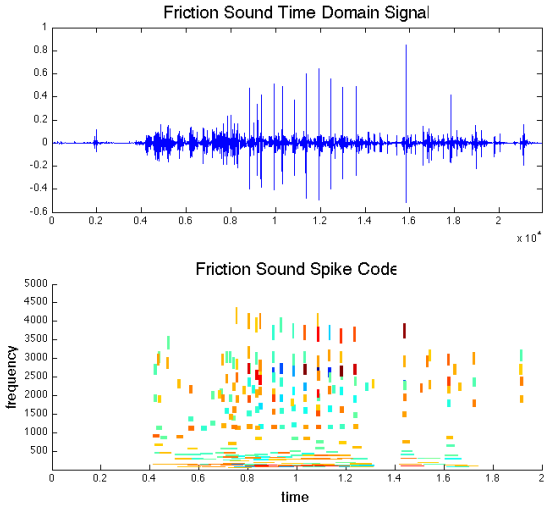


Figure 2: The sound wave and the spike code of a real friction sound. Similarly to the impact sound, in this picture we indicate how the spike code captures the skeleton of the friction sound.

example, a Gammatone filter bank of 256 filters is used for generating the spike code consisting of $K = 32$ spikes. Note that salient areas in the wave of this sound are coded with more spikes than other areas. In Figure 2, we show a real friction sound. Since the friction sound is much longer than the previous example with an impact sound, we used the same filterbank but used 256spikes to code the friction sound. These two examples indicate that the spike representation captures the signal characteristics properly.

2.2.1. *Dissimilarity Between Spike-Coded Sounds*

The spike code representation offers a totally new paradigm in observing similarities between sounds. This representation enables to solve the problem of detecting similarities by calculating these similarities without destroying the original structure of the pattern [11]. Basically, the distance between two spikes of different spike codes is computed as step. In the following step, the total distance between the given spike codes is calculated by taking the minimum of the sums of spike distances of all possible pairwise assignments. Intuitively, this method measures the minimal effort to transform one spike code into the other in terms of the single spike distance.

The distance $d_s(s, s')$ between two spikes $s = (t, f, a)$ and $s' = (t', f', a')$ is composed of three individual distances, namely the distance $d_t(t, t')$ between time offsets, $d_f(f, f')$ between center frequencies, and $d_a(a, a')$ between amplitudes:

$$d_s(s, s') = \tau d_t(t, t') + \phi d_f(f, f') + (1 - \tau - \phi) d_a(a, a'), \quad (2)$$

with $\tau, \phi \geq 0, \tau + \phi \leq 1$. Parameters τ and ϕ allow to emphasize either the temporal, spectral, or amplitude aspect, while guaranteeing that the weights sum up to 1.

The temporal distance is calculated linearly, simply by taking the absolute value of the time difference between the time offsets of the corresponding spikes. In order to calculate the frequency distance, we consider logarithmic frequencies, according to the Weber-Fechner law. The amplitude difference is obtained by dividing each amplitude by the maximal amplitude of both sounds, thereby considering the volume differences as well.

Given the spike codes $\mathbf{s} = \{s_1, \ldots, s_K\}$ and $\mathbf{s}' = \{s'_1, \ldots, s'_K\}$ of two sounds and given the distance measure $d_s(s_i, s'_j)$ for two spikes $s_i \in \mathbf{s}, s'_j \in \mathbf{s}'$, we can define a dissimilarity $D(\mathbf{s}, \mathbf{s}')$ between two sounds by establishing a bijection between $\mathbf{s}$ and $\mathbf{s}'$. For a permutation $\mu \in S_K$ we assign to each spike in $\mathbf{s}$ exactly one spike in $\mathbf{s}'$, so that $s_i \to s'_{\mu(i)}$. Then we define the dissimilarity between two spike-coded sounds as the normalized sum of the distances between the K corresponding spikes:

$$D(\mathbf{s}, \mathbf{s}') = \frac{1}{K} \sum_{i=1}^{K} d_s(s_i, s'_{\mu(i)}), \quad (3)$$

with μ minimizing the dissimilarity:

$$\mu = \text{argmin}_{\mu \in S_K} D(\mathbf{s}, \mathbf{s}'). \quad (4)$$

In order to minimize μ in Equation (4), we propose a method, which is based on the alignment of the spikes to minimize the total distance. This method stems from the graph theory, and is called the Hungarian algorithm.

2.2.2. *Graph Matching*

Instead of solving the problem of minimizing the assignment μ in Equation 4 directly, we consider the spike codes of two given sounds as two point graphs combined in a bipartite graph. In this constellation, the vertices are the spikes $s_i \in \mathbf{s}, s'_j \in \mathbf{s}'$ of two spike-coded sounds $\mathbf{s}, \mathbf{s}'$, where each sound corresponds to a disjoint subgraph of the bipartite graph, and the weights w_{ij} (similarities) between them are derived from the distance d_s : $w_{ij} = 1 - d_s(s_i, s'_j)$ with $\mu(i) = j$. This consideration converts the problem into a combinatorial one of finding a perfect matching of the weights in a bipartite graph by the Hungarian algorithm [12].

We consider a matching to be a subset of the edges of the given graph containing each vertex only once. In a perfect matching, every vertex within the graph is adjacent to one and only one edge. The Kuhn-Munkres Theorem [12] guarantees the convergence of the algorithm to a perfect matching.

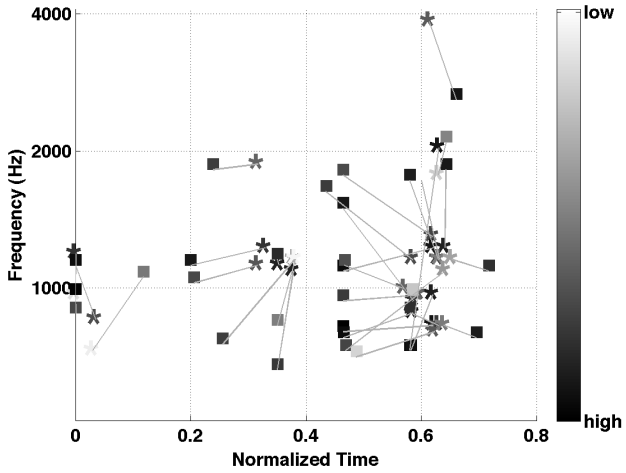


Figure 3: This figure shows how the Hungarian algorithm matches individual spikes of two spike coded pouring liquid sounds to minimise the distance between them. The spikes of the sounds are indicated by a different symbol ('*','■'). The light intensities of these symbols encode the amplitudes of the spikes.

Figure 3 shows an example about the matching results of two spike coded sounds performed by the Hungarian algorithm. This plot indicates the matching results between two sounds of the same sound class, which exemplifies that the distances between spike patterns of same sound classes are in general close to each other.

3. THE SONIC SPACE, SONIC LANDMARKS AND GUI

The sonic landscape is organized as a 2D space, where reference sounds (sonic landmarks) are located. The organization of the reference sounds may rely on perceptual criteria, on multidimensional scaling, they can be derived on a statistical basis (e.g., clustering), or may be arbitrarily decided by the user.

The sonic landmarks in the sonic space form a 2D scatter points set. The sound designer may choose to generate a new sound in the vicinity of a set of sonic landmarks inspired by their sonic properties. Depending on the coordinates of the new position indicated by the user, the new set of synthesis parameters is generated through an interpolation scheme accounting for the neighboring sonic landmarks. Convenient interpolation schemes for this class of problems are often based on a preliminary triangulation step, aimed at creating a grid of triangles connecting the scatter points together. One possibility, which is used here, is the Delaunay triangulation. A new interpolated point is then computed by first finding the three vertices of the triangle in which the point is confined, and then applying a linear interpolation formula based on the equations of the plane defined by the three vertices.

Figure 4 gives an example of such interpolation for a small set of data points (circles) and a scalar z parameter. The interpolated data are depicted as stars.

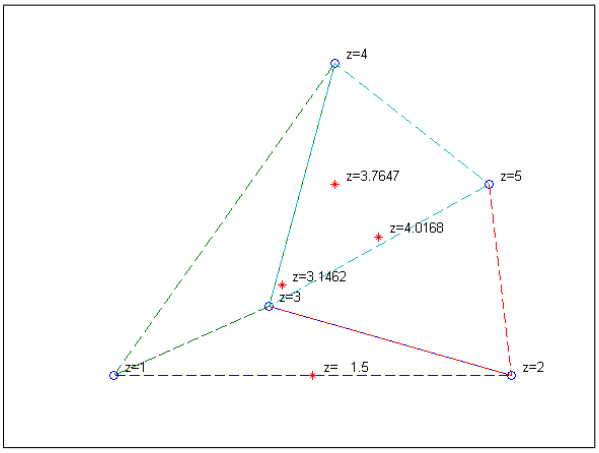


Figure 4: Scatter points interpolation: Delaunay triangulation (dashed lines) on a set of data points (circles), and four interpolated points (stars).

A graphical user front-end (GUI) was implemented in Matlab and is shown in Figure 5. Its main frame represents the 2D sonic space with the sonic landmarks located in pre-defined spots (the figures show a configuration, where the landmarks are at the vertices of a rectangle). The location of the reference sounds can be arbitrarily decided and is specified in a configuration file loaded at the beginning of each session. Once the landmark sounds have been loaded and the auditory representation computed, the user can perform a number of actions, including creating a new sound by pointing to a position in the proximity of the sonic landmarks with desirable characteristics, listening to landmarks and new sounds by clicking on the spike plot, deleting newly generated sounds which are not of interest for the user, saving the parametric setting of a new sound as presets in the XML-based format used by the SDT GUI, and checking network connectivity with the Max/MSP or Pd server running the SDT sound synthesis engine.

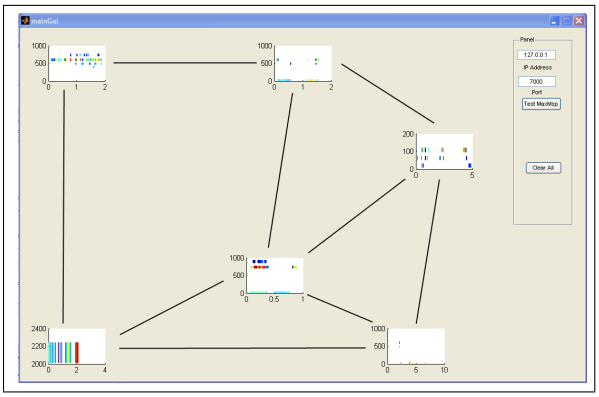


Figure 5: Matlab GUI. Spike representation of four sounds (sonic landmarks) representative of the entire sound space generated by means of the SDT friction model. By clicking in the graphical area around a spike representation, it is possible to listen to the corresponding sound. On the left, a control panel is provided for the communication with the SDT model in Max/MSP

The spike-based GUI is connected through an OSC network layer to a Max/MSP or Pd server running the SDT sound synthesis engine. When the user confirms the position of a new sound to be generated, the GUI starts the communication with the SDT server, proceeding through the following steps: first, the new interpolated parameters are sent to the server, which updates the controller values; when done, the SDT synthesis engine is started and an audio file is generated; finally, the GUI loads the audio file, generates the new auditory representation, and plots the result in the 2D sonic space.

The results of a sound design experiment with six reference sounds (the sonic landmarks) organized in the 2D space are shown in Figure 6 on the left. The spike representation of the reference sounds, and the Delaunay triangulation generated with this configuration, are shown. Six new sounds were generated by choosing six interpolation positions (depicted as red stars). The synthesis sounds resulting from the interpolated parameters, converted into the spike-based representation, are represented in Figure 6 on the right.

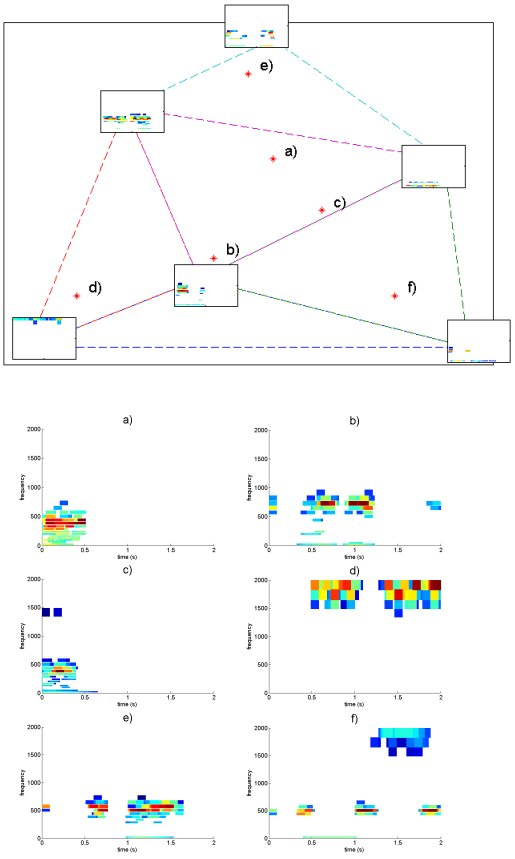


Figure 6: Interpolation example: the new sounds a), b) and c) are generated by interpolation of points in the same triangle; the new sounds d), e) and f) are located in three different triangles.

Perceptually, the results match the expectations, and the dynamic and spectral characteristics are recognized as actually deriving from the characteristics of the neighboring landmarks [1].

[1]The audio files corresponding to the sound landmarks and to the interpolated sounds are available for download at `http://mordente.sci.univr.it/~carlodrioli/Dafx10/Experiments.htm`

4. PARAMETERS SEARCH ALGORITHMS

The sonic landmarks are the representatives of a physical sound synthesis model. They not only show the diversity of the sounds, which can be generated by this model, but can also be used as starting points to generate new sounds, which are mainly interpolations of several of them (See Section 3).

Therefore defining the sonic landmarks constitute a very useful step in the sound design process using a synthesis model. As mentioned earlier (Section 3), they can be real sounds or results of some statistical methods like the MDS or clustering. For the landmarks, defined by using such methods, we do not know the input parameter values of the physical sound synthesis model, which would generate them. The ad-hoc approach the sound designers generally utilize is the trial and error method. This generally yields sub-optimal solutions and takes a long time. In this paper, we propose a Bayesian framework for the parameter optimization of a physical sound model problem for a given sonic landmark for which the input parameter values are unknown. From this point on, we call the landmarks as reference sounds. For this purpose, we make use of the spike code and the corresponding distance measure presented in Section 2.2. The reference sound and each candidate sound are coded by using the spike code. Subsequently, the distance between the spike codes of these two sounds is calculated. We use Gaussian Processes [13] [14] to model this distance function. We combine this model with a clever optimization method to select the next candidate sound, which is expected to improve the quality of the candidate sounds by minimizing the distance function modeled by the Gaussian process. In each iteration, the real distance between the reference and candidate is calculated. By using this distance, the Gaussian process is trained once again with the real distance values calculated so far. The newly trained Gaussian process is incorporated in the optimization method to select the next candidate sound. This procedure is repeated as long as the distance between the reference and the candidate is above a certain threshold value.

4.1. Gaussian Processes

A Gaussian process (GP) is a stochastic process containing any finite number of random variables having a joint Gaussian distribution. A GP is defined by its mean $m(x_i)$ and covariance functions $k(x_i, x_j)$. So, we write $f(x_i) \sim GP(m(x_i), k(x_i, x_j))$.

We use a zero mean prior GP with squared exponential covariance function given by

$$k(x_i, x_j) = \sigma_f^2 exp\{-\frac{1}{2}(x_i - x_j)^T M^{-1}(x_i - x_j)\}, \quad (5)$$

where σ_f^2 is the variance of the function we model and M is a diagonal matrix of the length scales of the Gaussian components.

In real world applications, it is generally not possible to obtain the real function values $f(x)$ but their noisy versions $y = f(x)+\epsilon$. Pursuing this idea, we assume that our observations are noisy. We assume additive noise ϵ with zero mean and variance represented by σ_n^2. Therefore our final covariance function is written as

$$cov(x_i, x_j) = k(x_i, x_j) + \sigma_n^2 \delta_{ij}, \quad (6)$$

where δ_{ij} is the Kronecker delta, which is one if $i = j$, otherwise zero.

The goal of modeling the function f by using the GPs is to be able to make predictions about the data points, that have not been observed yet by simply incorporating the data points observed so far.

In order to be able to make predictions about the function values of the unobserved data points, we need to make use of the observed data points and get to the posterior distribution.

The function values corresponding to the data points can be sampled from the joint distribution by evaluating the posterior mean and covariance functions. Hence, for each new data point, the posterior distribution is re-calculated and the function value for the new data point is sampled from this distribution. However, we do not know the optimal values for the so-called hyper parameters of the posterior distribution.

The hyper parameters of the posterior distribution of the function values are given by $\theta = \{\sigma_f, M_{ii}, \sigma_n\}$, where M_{ii} represents the length scales of the Gaussian components and $i = 1 \dots D$ (See Equation 5). In order to determine the optimal hyper parameter values, we compute the probability of the data given the hyper parameters, which we call the marginal likelihood.

Therefore, after each new data point, the marginal likelihood is maximized to model the function values regarding the new data points in a better way. Subsequently, the posterior distribution is utilized to estimate the function value of the new data point. Note that our aim is not to estimate the function value as close as possible to the real function value, but to find the data point iteratively, which minimizes this value. Therefore, we need a clever strategy to select the next data point. However, it is still unclear how we select the new data points. For this purpose, we define the concept of expected improvement using the GP we described here.

4.2. The Expected Improvement Function

The GP we defined in the previous section models the distance function in a proper way. However, we do not want to model the distance function at first place, but rather try to find its minimum with respect to a given reference. In particular, if the evaluation of the function is a time consuming operation, a clever sampling method is needed. Therefore, we need a selection criterion to sample the data space cleverly to come closer to this minimum as fast as possible. Osborne et. al. [15] suggest a probabilistic method for this sampling problem, in particular for the minimization of such time consuming functions.

Imagine that we have only one iteration left to improve the quality of our estimation. In our model, the best estimation is the closest sound generated by the physical sound model given the reference sound. Suppose that $(\mathbf{X_0}, \mathbf{f_0})$ are the data points and their corresponding function values evaluated so far. The best estimation among these function values is defined as $\eta = \min \mathbf{f_0}$. Hence, the evaluation of the last data point $(\mathbf{x}, f)$ could improve the quality of our estimation in the following manner:

$$\lambda(f) = \begin{cases} f; & f < \eta \\ \eta; & f \geq \eta \end{cases} \tag{7}$$

Hence, the expected improvement [15] is simply $\min(f, \eta)$ either the previous minimum or the function value of the last data point. Given our GP over the function, which we try to minimize, we can define the total expected improvement given a data point x as follows:

$$\Lambda(\mathbf{x}|\theta) = \int \lambda(f) p(f|\mathbf{x}, \theta) df, \tag{8}$$

where we use the predictive distribution of the function values. After some elementary calculations, the total expected improvement function becomes the following form:

$$\Lambda(\mathbf{x}|\theta) = \eta + (\mu - \eta)\Phi(\eta; \mu, C) - C\mathcal{N}(\eta; \mu, C), \tag{9}$$

where Φ and $\mathcal{N}$ are the cumulative probability distribution function (cdf) and the probability density function (pdf) of the function values respectively. The mean μ and the covariance C are the mean and the covariance of the posterior GP defined in the previous section.

Note that the expected improvement function given in Equation 9 decreases firstly when the mean value μ becomes lower than the current minimum η, and secondly when the covariance C becomes larger. The first case implies exploitation and the second case exploration. In other words, the expected improvement function exploits the region, where the current minimum is as well as it explores other regions with high covariance. Consequently, this function defines a balance between these two concepts, which is also known as the exploitation / exploration dilemma.

The data point for which the expected improvement function is the lowest is the next data point for the next evaluation. Hence, we need to minimize this function. Fortunately, the expected improvement function given in Equation 9 is easy to evaluate and differentiable infinitely many times. Therefore, there are many methods to minimize this function with respect to the data point x by incorporating its gradient and Hessian. Here we used Matlab optimization toolbox to minimize the expected improvement function. Figure 7 shows an iteration of the expected improvement scenario. As one can see, the expected improvement function (line with the diamonds at the bottom of the figure) is lower for the regions, where the covariance of the GP is large as well as for the regions, where the current minimum is. So the expected improvement function indeed balances the exploration of the unvisited regions with the exploitation of the observed regions, which potentially contain the global minimum.

4.3. Tests and Results

The parameter optimization model has been tested with the synthesized sounds at first, which were synthesized by the same physical model. For this purpose, we made use of the preset parameter values. In the physical modeling paradigm, different parameter combinations can yield similar results. This means that there are many local minima for a given reference sound. Therefore, we wanted to see whether the optimization model can reveal the parameter proximity by using the spike based similarity measure. The experiments performed show that the model can indeed approximate the parameter values, which have been used for generating the reference sound.

The physical model we wanted to optimize is a highly complex model with a large quantity of input parameters. These parameters not only control the spectral content of the generated sounds but also their temporal content. Furthermore, because of the stochastic nature of the model, the temporal content is not static, but it changes randomly in a certain range without diverging from the group of sounds which the model is supposed to generate. In other words, the physical model generates not exactly the same sound

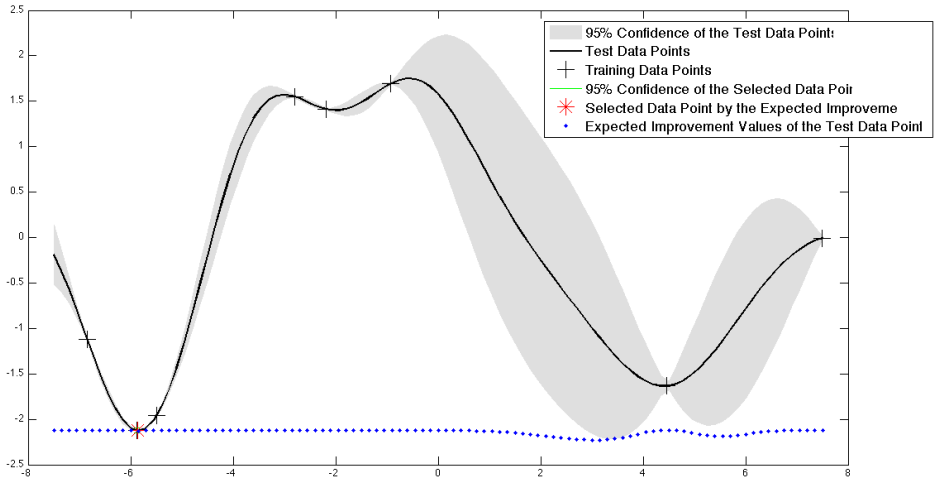


Figure 7: One iteration performed using the expected improvement function is shown. The line with diamonds shows the expected improvement values of the test data points. The asterisk on this line is the data point selected by minimizing the expected improvement function. The crosses are the training data points observed so far. The solid line is the test data points, which are evaluated with respect to the current GP estimation of the function to be modeled. The shaded area around this line is the confidence interval of the GP.

for the same parameters. Since the spike based similarity measure captures the temporal similarity as well as the spectral similarity, these random changes in the temporal content of the generated sounds were a problem. Therefore, we restricted the temporal diversity of the sounds at first and concentrated ourselves on the spectral content. Including the parameters controlling the temporal behaviour of the sound into the optimization procedure could be a possibility. Unfortunately, these parameters are not scalars, but mainly mappings between two functions. We simply kept the temporal pattern of the sounds to be generated by the model constant. By doing this, we made the model generate very similar sounds for the same parameter values, but not exactly the same sounds.

Since the physical model has a large number of input parameters, we wanted to increase the complexity of the optimization model gradually. Thus, we did not cover the parameter domain completely but started with a smaller number of parameters to optimize. For these tests, the other parameters were kept constant. Hence, we started with optimizing one parameter only, namely the frequency. Afterwards, we performed 3D optimization with the parameters frequency, gain and decay. Finally, we performed six dimensional optimization with the additional parameters bristle stiffness, bristle viscosity and surface roughness. Several results of these optimization experiments are shown in Table 2.

In the results of the three dimensional parameters, the order of the parameters is given as frequency, decay, gain. For the six dimensional results, the first three parameters are like in the three dimensional results. The last three parameters are bristle stiffness, bristle viscosity, surface roughness. All the parameters shown in Table 2 are normalized to the interval $[0 : 128]$. As one can see, experiments repeated with the same reference sound yield similar results.

For the last results in Table 2, we show the progress in Figure 8 by plotting the spike codes of the sounds suggested by the

Real Values	Estimated Values
65-12-105	66-59-105
65-12-105	68-64-56
50-54.5-50-36.6-65.8-0	50.5-83.2-38.4-23.9-66.5-42.9
50-54.5-50-36.6-65.8-0	49.5-45.5-63.3-7.7-77.9-0
50-54.5-50-36.6-65.8-0	46.8-55.3-55.7-28.8-71.4-0
100-61.5-100-64.2-89.4-83.7	100-66.6-100-66.6-100-100
100-61.5-100-64.2-89.4-83.7	100-53.5-100- 71.3-100-100
0.8-95-0-50-11.5-50	0-100-0-100-26.4-55.7
34.15-96.75-70-50-52.5-76	0-100-100-46.5-100-0

Table 2: This table shows three and six dimensional optimization results. The column on the left hand side shows the parameter values of the given sonic landmark. The second column on the right hand side shows the estimated parameter values.

expected improvement function. The fifth sound is the optimal approximation found by the method. The sixth sound next to the offered optimal is the sound landmark for which we run the optimization. As one can see, the similarity to the sound landmark increases gradually form iteration to iteration. However, please recall that the similarity of the suggested sounds does not become larger after each iteration, because the algorithm balances the exploration with the exploitation. While exploring some unobserved regions, the suggested sound can be less similar to the sound landmark than the most similar sound detected so far. However in the end, the detected optimal is likely to be the global optimal because of this reason.

In the near future, these results will be extended with real world sounds. Testing the optimization model against some preset values still shows the capability of the model to some extent. However testing it against real world sounds is mandatory to observe the real performance of the model and to see its pros and cons com-

pletely. For a real sound, there will not be an exact match of the parameters. Therefore, the evaluation of the model for a real world sound should be done in another way. Human judgements can be used like in a psychoacoustic experiment. The SNR between the real world sound and the estimated sound can give an objective criterion about the quality of the estimated sound.

In order to test the model, the temporal pattern of the real world sound should be mapped to the temporal parameters of the model. This should be done by determining the temporal pattern of the real world sound by means of proper methods as envelope-follower, pitch trackers and so on beforehand and train the model again only for the spectral content.

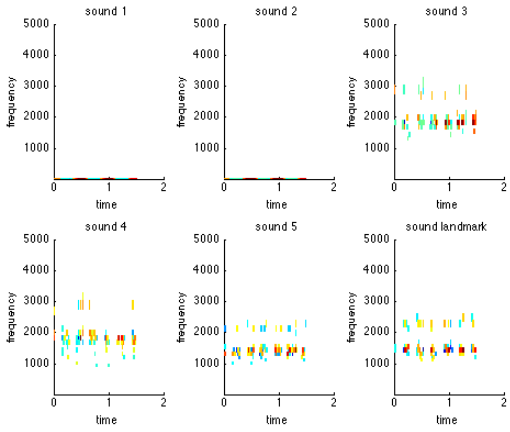


Figure 8: This figure shows the progress of one optimization experiment. The spike codes of the shown sounds 1-5 are the suggested sounds from the expected improvement function in the consecutive iterations. The last sound is the sound landmark.

When the framework will be complete, it will be possible for the designer to pick up an environmental sound or to produce an imitation that could be translated into a sonic landmark, automatically chosen as the closest synthetic realization that can be produced by the given sound model.

5. CONCLUSIONS

The initial results show that the visualization platform in combination with a probabilistic framework helps the sound designer to automatically navigate through the soundscape and to rapidly find the desired sound.

The experiments we performed so far contain preset sounds of the physical model itself. Nonetheless, these results give an idea about the capabilities of the optimization framework we propose in this paper. Furthermore, in the near future, the model will be tested against real world sounds to see the simulation and optimization capacity of the model in much wider sonic contexts.

6. ACKNOWLEDGEMENTS

This work is part of the research carried out within two EU funded project: CLOSED (Closing the Loop of Sound Evaluation and Design), FP6-NEST-PATH no. 29085, and NIW (Natural Interactive Walking), FP7-ICT-2007 FET Open no. 222107.

7. REFERENCES

[1] Cumhur Erkut, Vesa Välimäki, Matti Karjalainen, and Henri Penttinen, ," in *Physics-based Sound Synthesis*, Pietro Polotti and Davide Rocchesso, Eds., vol. Sound to Sense, Sense to Sound. A State of the Art in Sound and Music Computing, chapter 8, pp. 303–343. Logos Verlag, Berlin, 2008.

[2] Vermont: Destiny Books Rochester, Ed., *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester, Vermont: Destiny Books, 1994.

[3] Berry Truax, *Handbook for Acoustic Ecology*, ARC Publications, 1978.

[4] A. Momeni and D. Wessel, "Characterizing and controlling musical material intuitively with graphical models," in *Proceedings of the New Interfaces for Musical Expression Conference*, Montreal, Canada, 2003.

[5] Davide Rocchesso and Federico Fontana, Eds., *The Sounding Object*. Mondo Estremo, 2003.

[6] Stefano Delle Monache et Al., "Sound design tool: Algorithms for ecologically-founded sound synthesis," *submitted to ACM SIGCHI Designing Interactive Systems DIS2010 Conference*, 2010.

[7] E. Smith and M. S. Lewicki, "Efficient coding of time-relative structure using spikes," *Neural Computation*, vol. 17, pp. 19–45, 2006.

[8] R. D. Patterson and J. Holdsworth, "A functional model of neural activity patterns and auditory images," *Advances in Speech, Hearing and Language Processing*, vol. 3, pp. 547–563, 1996.

[9] M. Slaney, "A matlab toolbox for auditory modeling work," *Interval Research Corporation*, 1998.

[10] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, vol. 47, pp. 103–138, 1990.

[11] K. Adiloglu, R. Annies, E. Wahlen, H. Purwins, and K. Obermayer, "A graphical representation and dissimilarity measure for basic everyday sound events," *IEEE Transactions on Audio, Speech and Language Processing*, 2010, submitted.

[12] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, 1955.

[13] C. E. Rasmussen, "Gaussian processes in machine learning," in *Advanced Lectures on Machine Learning, Revised Lectures*, O. Bousquet, U. von Luxburg, and G. Rätsch, Eds., pp. 63–71. Springer Verlag, Heidelberg, Germany, 2004.

[14] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA, USA, 2006.

[15] M. A. Osborne, R. Garnett, and S. J. Roberts, "Gaussian processes for global optimization," in *3rd International Conference on Learning and Intelligent Optimization (LION3)*, Trento, Italy, 2009.

A Toolkit for Explorations in Sonic Interaction Design

Stefano Delle Monache
IUAV - University of Venice
Venezia, Italy
stefano.dellemonache@gmail.com

Pietro Polotti
University of Verona
Verona, Italy
pietro.polotti@univr.it

Davide Rocchesso
IUAV - University of Venice
Venezia, Italy
roc@iuav.it

ABSTRACT

Physics-based sound synthesis represents a promising paradigm for the design of a veridical and effective continuous feedback in augmented everyday contexts. In this paper, we introduce the Sound Design Toolkit (SDT), a software package available as a complete front-end application, providing a palette of virtual lutheries and foley pits, that can be exploited in sonic interaction design research and education. In particular, the package includes polyphonic features and connectivity to multiple external devices and sensors in order to facilitate the embedding of sonic attributes in interactive artifacts. The present release represents an initial version towards an effective and usable tool for sonic interaction designers.

Keywords

Sonic interaction design, physics-based sound synthesis, user interfaces.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Auditory (non-speech) feedback*

1. INTRODUCTION

One might assert that the core of interaction design research is the relationship between action, on one side, and function and meaning on the other. In the last decades, research in interaction design developed a multiplicity of theoretical approaches, methods and tools to steer interactive artifacts towards the naturalness of interaction we are used to in everyday environments. Recently, major discussions in the research community arose with respect to the reliability of research methodologies for interaction design.

In particular, the debate concerns the relationship between design and science, questioning the points of convergence and divergence of design practices as compared to scientific approaches, in order to highlight fruitful implications for interaction design research [5]. The research community correctly agrees that HCI research aimed at supporting design should be grounded in design practice, theory and philosophy. That is, the design activities should be carried out *as research*, and not *for research* "pursued primarily in service of a theoretical concern" [18]. In the same spirit, we take Stolterman's assertion, according to whom "it is possible to predict the potential success of new approaches, methods, and tools based on how designerly they are" [34].

Sonic interaction design (SID) is positioned in between auditory display, interaction design, ubiquitous computing and interactive arts. In SID, the role of sound as natural carrier of information is exploited as effective means to establish continuous negotiation with interactive artifacts [21, 30]. Some clear everyday life examples are the vacuum cleaner, wherein sound conveys a strong sense of power and control over the suction, and the use of rhythm in activities such as hammering, rubbing with sand paper, or cutting vegetables on the chopping board. Therefore, research in SID relies on knowledge in ecological acoustics, everyday sound perception and categorization, sound and music computing and 1) investigates the iterative process of design for sound creation [6, 35], 2) considers evaluation as a fundamental stage of sound design [22], 3) applies different approaches and methods to ideas prototyping [17, 29, 27, 20], 4) aims at developing design-oriented sound synthesis tools for complex interactive applications. In fact, there's a strong need to provide designers with a set of skills and tools that allow them to consciously design the acoustic behaviour of future artifacts.

In this paper, we introduce the Sound Design Toolkit (SDT), a software package consisting in a set of physics-based sound synthesis models, especially addressed to SID research and education. As a work in progress, the development of the sound synthesis engines, graphical user interfaces (GUI) and control layers advances together with the realization of interactive workbenches, and the evaluation of the system in workshop settings. These activities serve as *tools for reflection*, in order to collect a useful repertoire of design ideas and concepts [31]. In this way, SID as research advances together with research in sound and music computing.

The paper has the following structure: in Section 2, we consider why a physics-based approach can be suitable in interaction design; Section 3 describes the philosophy underlying the SDT, its features, the GUI design and its implementation; in Section 4, some interactive applications are briefly treated, where the SDT is employed to generate consistent continuous sound feedback; in Section 5, we draw our conclusions.

2. PHYSICS-BASED SOUND SYNTHESIS: FROM VOLATILITY TO PHYSICALITY.

In sound and music computing research, physics-based sound synthesis is indeed not a novelty anymore. In recent years, physical models have been applied to non-musical, everyday sounds, too [7, 10, 28]. According to this approach, sound generation can

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
AM'10, September 15–17, 2010, Piteå, Sweden.
Copyright © 2010 ACM 978-1-4503-0046-9/10/09…$10.00.

be described in terms of physical events, as they occur in everyday life. Most of the experiences take inspiration from the musical field, mainly responding to the needs of composers and musicians who want to approach musical synthesis in terms of simulation of the physics of acoustic instruments [36]. The idea is that if a virtual instrument is designed in such a way that its behaviour responds much like the actual acoustic instrument, the synthesized sound will be natural and expressive in a performance context. Given the physical consistency of the models, it is straightforward to map their control parameters to continuous physical interactions, and to describe resonators and their interaction modalities by means of physical and geometric properties. Therefore, the intrinsically natural behaviour of the physics-based sound models can potentially provide an embodied experience, since the computer-generated sound feedback is energetically consistent with the action [4, 11, 14].

Nonetheless, the creation of a functional, continuous sound feedback requires different approaches and tools, that have to be grounded on design practice. The question is about how to make a product sensible to manipulative actions, not as an "intelligent" or knowledgeable subject, but as an object capable of feeding the stimuli back in a dialogic form. In everyday environments, this is especially important for those objects that are designed to provide multiple functions, thus changing their identity over time. In particular, this is the case of those objects integrating electronic technologies and computational elements, as in home automation, automotive applications, consumer electronics and entertainment. In these contexts, sound plays a crucial role to characterize the identity of an object and the experience of use.

Being designers mainly confident with vision and visual thinking, current sketching approaches largely involve forms of visual storytelling [6] and the sonic dimension is mainly approached by intuition and creativity. Designers are often unaware of the auditory domain, of its complexity and potential, and they largely ignore sound processing and synthesis methods. Sounds are time-bounded, volatile, and difficult to outline without an appropriate vocabulary. A designer is hardly able to describe sound attributes such as spectrum, transients, etc. . It would be even more complicated to ask a designer to "guess" how to produce the sound he or she imagines, even with the most advanced off-the-shelf software. Indeed, it is quite difficult to sketch expressive sound feedback starting from scratch: sample-based processing requires a sort of sculptural attitude, while abstract sound synthesis techniques need quite complex control layers, though some interesting tools for the generation of environmental sounds have been recently released[1].

A physics-based approach results to be closer to designers thinking: sounds can be directly described in terms of configurations, geometries, materials and their properties, dynamics of gestures to the point that allows a certain immediate "visualization" and association of the sketched sound with its interactive context, not to mention an intuitive sensory connection with touch. We argue that a physical approach to sound can potentially encourage a certain operational physicality that is much needed in design practice. This approach strongly complements the established method of Foley, a common design practice for audiovisual media [7, 37, 28, 2]. Like foley artists, sound designers are provided with a palette of virtual sounding objects that can be explored and combined, to create compound, elaborated sound events.

One may oppose this assertion suggesting that designers do not have a deep understanding of mechanics and fluid dynamics, which seems to be needed to make sense of exotic parameters such as the Stribeck velocity or the Reynolds number. That's precisely why there is the strong need of an interpretative layer that maps physical descriptions to a more naïve physics [32]. Most of this paper is dedicated to proposing one such layer. Finally, the increasing availability of sensors, actuators and low-cost embeddable CPUs makes physics-based sound synthesis a promising paradigm for the design of interactive continuous sound feedback in everyday contexts.

3. IMPLEMENTATION AND DESCRIPTION

The SDT[2] complies with multi-platform requirements of the European project CLOSED [35], and currently runs in Max/MSP environment[3], though the provided source code can be compiled for Pure Data (Pd)[4], as well. The provided physics-based sound models, developed and implemented as externals and patches, can be easily coupled with physical objects and are computationally affordable for real-time applications on ordinary hardware. The sound models algorithms are developed according to three main points: 1) auditory perceptual relevance; 2) *cartoonification*, i.e. simplification of the underlying physics and emphasizing of its most relevant aspects in order to increase both computational efficiency and perceptual sharpness; 3) parametric temporal control ensuring appropriate, natural and expressive articulations of sound event sequences [28].

As shown in figure 1, most of the provided sound algorithms follow the modular structure "resonator–interactor–resonator", representing the interaction between two resonating objects. The resonators are modeled through the modal [1] or the digital waveguide [33] techniques. The resulting timbral richness – even when using very basic resonators – is due to the choice of non-linear interaction models. Models of impact [24] and friction [13] have been implemented and exploited as basic sonic events underlying complex sound phenomena. Rolling, bouncing, breaking and crumpling sounds are modeled through complex temporal patterns on top of the impact model, while the friction model is used to simulate rubbing, squeaking and braking sounds. Furthermore, different liquid-related sounds – as burbling, dripping, pouring and frying – have been implemented making use of a bubble model [37].

Hence, by means of SDT, sonically augmented (sounding) objects

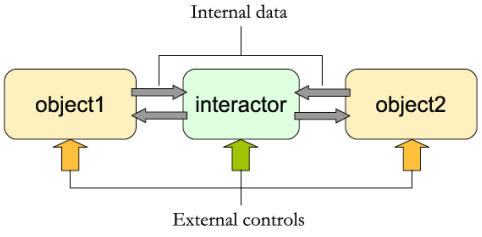


Figure 1: The modular structure implements a feedback communication between interaction and resonator models.

[1]See for example the "Sounds of nature" VST plugins, developed by xoxos (www.xoxos.net), the "Tapestrea" developed by the Sound Lab of Princeton University (http://taps.cs.princeton.edu/, the "Tassman" software by Applied Acoustics System (http://www.applied-acoustics.com/tassman/overview/) or Andy Farnell's Pd patches for everyday-sound synthesis (www.obiwannabe.co.uk).

[2]The SDT-0.5.0 can be freely downloaded at http://www.soundobject.org/BasicSID/Gamelunch/Resources.html.

[3]www.cycling74.com.

[4]www.puredata.info.

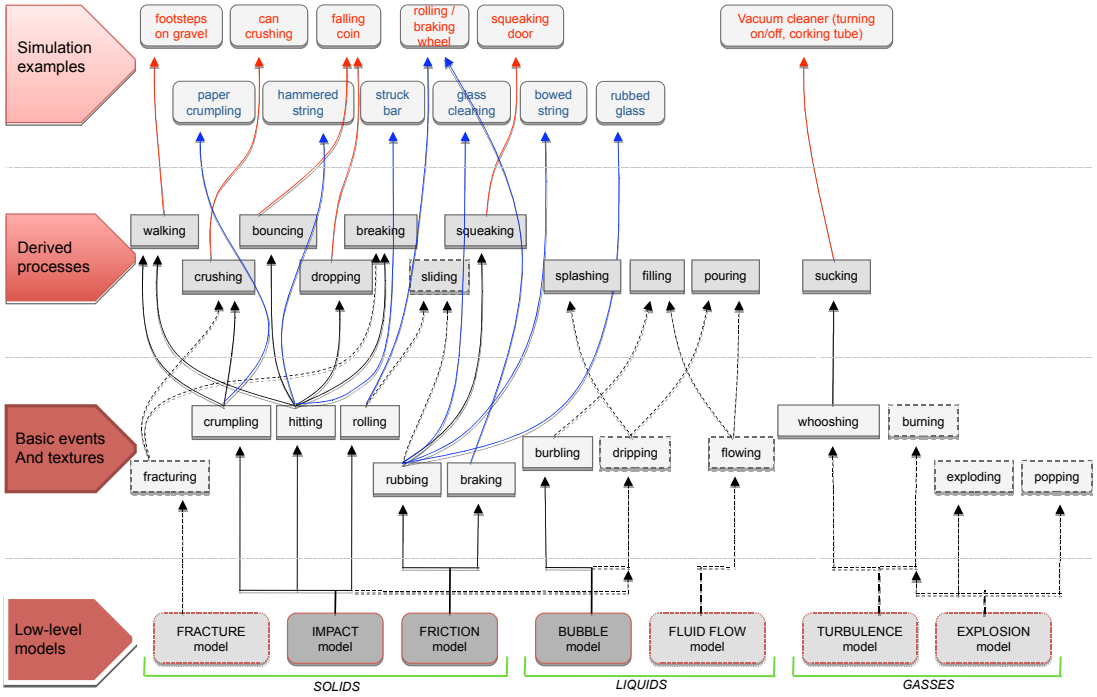


Figure 2: The proposed taxonomy of everyday sounds.

can be created with an expressive acoustic behaviour in the sense of ecological hearing introduced by Gaver [19].

Our main objective is to provide a tool that progressively can 1) grow away from a musical metaphor; 2) respond to the requirements of design thinking; 3) be sufficiently immediate and easy to use in the sketching and prototyping stages of the design process; 4) represent a valuable platform in SID activities.

Given these guidelines, the main features of the current release of the SDT are:

- a high level, general template of available sound models, organized according to a taxonomy of everyday sounds;
- the clustering of the control parameters into high and low level GUIs according to their more or less meaningful and immediate effects, thus providing a clear monitoring of the sound models;
- an intuitive understanding of the control parameters, facilitated by a common sense and perceptually-based naming and ranging of the physical parameters;
- the possibility to allocate multiple instances (polyphony) of the sound models in order to facilitate the design of compound sound events;
- a comfortable integration of external devices allowing an interactive control of the sound models.

3.1 Navigating the SDT package

Figure 2 shows the proposed taxonomy: the hierarchy is established from low-level sound events to more complex, patterned or compound processes. The corresponding low-level sound models are presented at the bottom of the graph, while the second level shows basic events and sound textures straightly derivable from them. Processes that can be related to temporal patterns of basic events and textures are presented in the third level. Lastly, the top level contains several examples of the implemented simulations, while dashed connections represent expected dependencies for the simulations yet to be developed.

When the SDT application is started, a front-end shows up displaying the available sound models among low-level basic events and sound textures, while the derived processes are obtained through their combinations. Furthermore, such a taxonomy provides the designer with an initial palette of sound events to browse in the early stage of sound sketching.

3.2 User-centered features

Often, even the sketching stage of a sound design process requires the preparation and collection of a large amount of code, with consequent issues of redundancy, version control, necessity of cumbersome tweaks, latency issues, problems in software communication with sensing devices, exponential proliferation of control maps, and so forth.

The aim of the SDT is to provide interaction designers with a sound design application that runs on common real-time environments, such as Max/MSP and Pd, and that can be easily exploited in their activities. For this purpose, the overall framework of the SDT is designed in order to automatically rebuild the audio and control connections, where needed. In practice, it is possible to manage multiple instances of the same sound model within a single working session, without recourse to any editing of Max/MSP patches. In figure 3, three independent impact sound models are loaded.

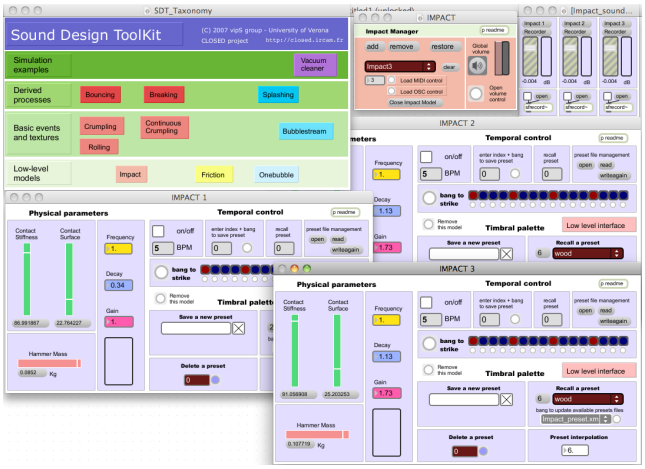


Figure 3: Three independent impact models.

In the model manager patch, the sound model instances are added indexically and can be browsed, removed or restored, at user's choice. Furthermore, volume controls and audio recording tools are automatically generated and grouped in a separate window, providing a global mixer.

Programming effective control maps is indeed a harsh task. Control maps give shape to the interaction, and represent the core of the coupling between action and function [25]. Thus, major efforts were undertaken in order to supply an environment sufficiently flexible, wherein various combinations of sensors data/physical parameters mapping can be easily accessed and tested.

A preliminary and extensive work on the sound models led to a selection of those physical parameters endowed with a major effectiveness in shaping sonic behaviours. In general, these parameters have been chosen as interactively controllable dimensions, and displayed in the high-level interface of the sound models.
Independent MIDI and OSC communication patches for each instance of a sound model can be loaded in the manager patch. As shown in figure 4, each MIDI/OSC control patch is provided with its own mapping setup section. A sensor data input/output matrix allows to choose a primary mapping configuration , *one-to-one* and *one-to-many*. It is possible to edit and save, as Max/MSP patch, each specific interactive parameter map. Newer maps are automatically listed and can be recalled in the drop-down menu. Such a feature makes it possible to collect a set of different solutions that can be easily combined at user's choice. In practice, the user has an auditory analogue of the board used by designer to rapidly compare a large number of drawn sketches. Finally, figure 5 shows the modular organization of the software package in folders and subfolders, providing a general and clear framework of sub-patches that can be accessed in view of further development and updates.

Far from being exhaustive, this high level organization of the SDT helps solving a set of bottlenecks encountered when making massive use of sound models: arrangement of patches, control

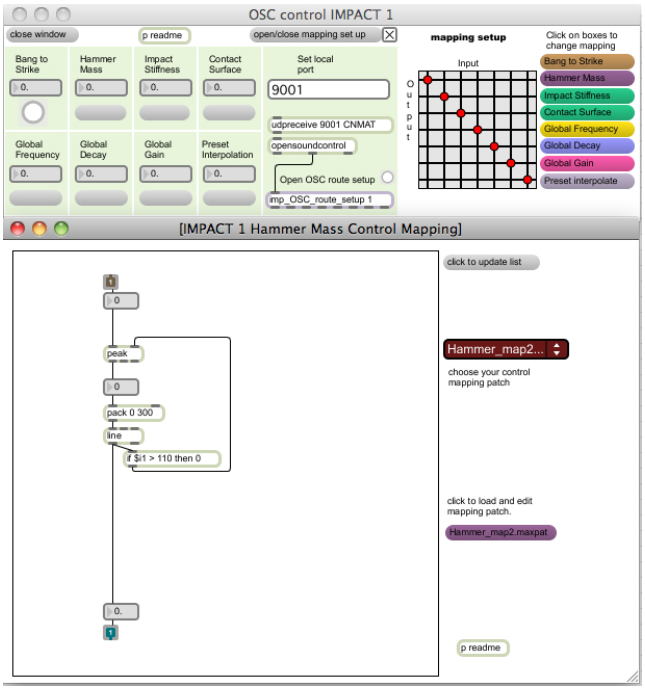


Figure 4: OSC control patch of instance 1 of the impact model. The red bulleted matrix manages the routing of incoming OSC signals, while the colored boxes allow to access single parameter mapping patches, edit, save and recall them. In this figure, the Hammer mass control mapping is shown as an example.

through a variety of sensors, sound mixing and so forth. Certainly, this framework doesn't relieve from having a basic knowledge of software programming, but it facilitates a direct access and use of the physics-based sound models and their capabilities in an interactive context. Also this is a work in progress: experiences of use are collected, in order to improve the "plug & play" characteristics of the SDT environment, investigate and exploit the sonic aesthetics of the physics-based sound synthesis paradigm.

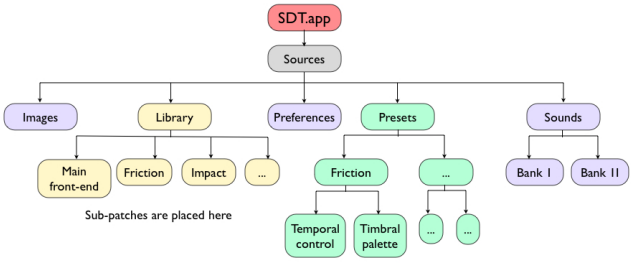


Figure 5: The SDT folders tree: Images, sub-patches, general preferences, presets and sound banks are respectively located in the pertinent subfolders. The Library folder is further split into as many sub-folders as the number of available sound models, and it contains all patches and abstractions. In the Presets folder, sub-folders that refer to each sound models contain separate lower level folders where Temporal control and Timbral palette presets are saved as XML file. Finally, in the Sound folder, samples can be recorded as AIFF files.

3.3 Clarity and intuitiveness

Physics-based sound synthesis generally requires the user to tune large sets of variables, and their mutual relationships are often difficult to perceive at a glance, especially in real time situations, where an immediacy of intervention is needed. In order to ease the access to sound models, the available control parameters are assigned either to low- or high-level interfaces. The current GUI provides, in fact, a functional hierarchy of parameters: the intelligibility of the most effective control parameters is facilitated in the high level interface, though their accessibility is retained in the low-level interface too. As shown in figure 6, the high-level interfaces are organized in three main sections: 1) physical parameters; 2) temporal control; 3) timbral palette.

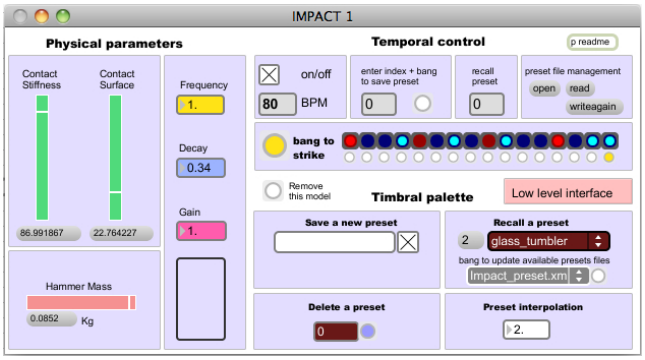


Figure 6: High-level interface of the impact model: physical parameters, temporal control and timbral palette sections.

The *physical parameters* section encompasses the most direct and effective parameters: for instance, in the case of the impact model, it includes the contact stiffness and the shape of the contact surface (parameters controlling interaction in figure 1), the hammer mass (parameter controlling resonator 1 in figure 1), and global factors of frequency, decay time and gain (parameters controlling resonator 2 in figure 1).

The *temporal control* section provides an "off line" sequencing tool that allow to simulate gestures or patterns. Indeed, creating interesting and effective timbres is important but not sufficient. In continuous interaction the efficaciousness of the sound feedback largely depends on how the articulation of the sound properties is coupled with the human action over time. As an example, in the friction model this section allows to control the temporal patterns of the rubbing velocity and exerted pressure. The coupling of the two forces strongly affects the expressiveness of the resulting gesture. Typical examples are rubbing on a glass harmonica, or drawing a bow across a violin string.

The *timbral palette* section allows to manage, store, recall, delete or interpolate configurations of sound parameters. The interpolation of reference sonic states (i.e. of parameter values) is one of the interactively controllable dimensions, as well. In previous works, the exploration of this feature revealed to be an effective and expressive means to convey information [12, 29]. Configurations are saved as presets in XML files, thus allowing them to be read, modified or generated from other XML-compatible software. Finally, the low-level interface gives access to all the available controls of the model, for further tuning and refining the sound (see figure 7). For example, it is possible to activate, deactivate or mute each resonant mode and singularly modify its frequency, decay time and gain.

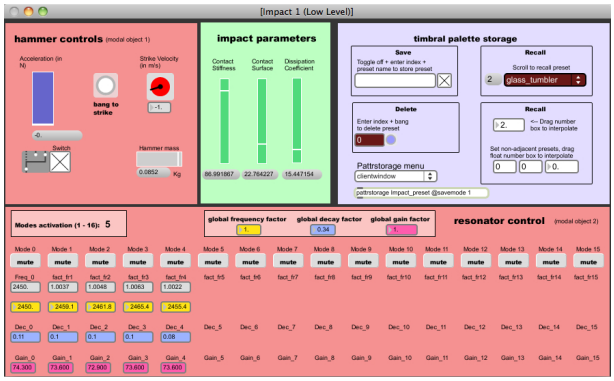


Figure 7: Impact model: the low-level interface gives access to all the available parameters.

4. EMBODYING THE SDT

The SDT has been extensively used in various interactive installations, workshops and research activities [29]. In the Gamelunch installation [9, 26], an interactive dining table, mock-ups and design experiences have been sketched around well defined themes and constraints, with the aim of exploring the interplay between the sensory channels, testing and improving the sound models, and finally developing a logic and a vocabulary for the SDT. The interactive environment includes a) a table with embedded sensors, acquisition boards, loudspeakers; b) bowls and dishes; c) graspable sensor-augmented objects such as a water jug, a tray, cutlery and bottles. Continuous interactions, such as cutting, piercing, pouring, stirring, are captured in order to drive the control parameters of some SDT sound models and elaborate feedbacks with immediate, non-symbolic and pre-attentional meaning. Following a basic design approach [3, 15, 16, 23], all the realizations have been systematized as exercises around specific themes. In particular, three themes have been investigated: a) continuous sound feedback for mechanical connections; b) supportive and expressive feedback for cyclic continuous actions; c) contradictory feedback for continuous action [29]. In general, the provided sound feedback stressed the richness of the physics-based sound models, to the point that some users that experienced the Gamelunch tried to override the design space, challenging expressive misuses of the augmented objects. The experiences and activities carried out around the sonically augmented dining scenario have been collected and published on the website: `http://www.soundobject.org/BasicSID/`. Recently, the SDT was successfully employed in a new project about gesture sonification, out of the context for which it was originally conceived[5]. Finally, a recent study explored the connections between narrative discourse, aesthetic attributes and sonic dynamic properties of interactive commodities. The investigation focused on how specific expressive qualities, that can be distilled from a discussion on film sound cases, could be transferred to the possibilities afforded by the physics-based approach in the SDT. The chosen narratives transferred well to the SDT implementation, making use of only two sound models, very simple control maps, and without sound processing. Results showed the effectiveness of the SDT in founding an appropriate, meaningful gesture-sound relationship, and constructing a semantic that related strongly to the designer's intention [8].

[5] `http://www.visualsonic.eu`.

5. CONCLUSION AND FUTURE WORK

The Sound Design Toolkit (SDT) is a set of perceptually-oriented and physically-consistent tools for sound synthesis aimed at the next generation of sound designers. The SDT consists of several sound synthesis algorithms implemented as Max/MSP externals and patches, and following a certain taxonomy of everyday sounds. The possibility to map the control parameters of the models to continuous physical interactions, and their correspondence to the physical and geometric properties of the simulated objects, make the SDT an effective tool for designing expressive sound feedback in real-time applications. The SDT has recently been revised and refined in an iterative process aimed at improving its effectiveness and usability in interactive contexts. To this end, the development proceeded in parallel with the realization of several sonically augmented everyday objects which were used for testing the sound models, while investigating the expressiveness of continuous sound feedback in the user-object interaction loop.

Future work will include the publishing of the SDT in various web repositories (e.g., `www.maxobjects.com`), thus collecting feedback and suggestions from a larger user base, in perspective of improvements and additions. Further realizations of sonic interactive everyday objects will help broadening the scope of the SDT. Moreover, all these experiences are being progressively systematized in order to collect a set of basic tutorials and exercises in sonic interaction design.

6. ACKNOWLEDGEMENTS

This work is part of the research carried out within two EU funded project: CLOSED (Closing the Loop of Sound Evaluation and Design), FP6-NEST-PATH no. 29085, and NIW (Natural Interactive Walking), FP7-ICT-2007 FET Open no. 222107.

7. REFERENCES

[1] J.-M. Adrien. The missing link: modal synthesis. In G. De Poli, A. Piccialli, and C. Roads, editors, *Representations of musical signals*, pages 269–298. MIT Press, Cambridge, MA, USA, 1991.

[2] V. T. Ament. *The Foley Grail*. Focal Press, 2009.

[3] G. Anceschi. Basic Design, fondamenta del design. In G. Anceschi, M. Botta, and M. A. Garito, editors, *L'ambiente dell'apprendimento – Web design e processi cognitivi*, pages 57–67. McGraw Hill, Milano, Italia, 2006.

[4] N. Armstrong. *An enactive approach to digital musical instrument design*. PhD thesis, Princeton University, 2006.

[5] C. Bartneck. Notes on design and science in the HCI community. *Design Issues*, 25(2):46–61, 2009.

[6] B. Buxton. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann, 2007.

[7] P. R. Cook. *Real Sound Synthesis for Interactive Applications*. A. K. Peters, Ltd., Natick, MA, USA, 2002.

[8] S. Delle Monache, D. Hug, and C. Erkut. Basic exploration of narration and performativity for sounding interactive commodities. In *Proceedings of the 5th International Workshop on Haptic and Audio Interaction Design (HAID)*, Copenhagen, Denmark, 2010.

[9] S. Delle Monache, P. Polotti, S. Papetti, and D. Rocchesso. Sonically augmented found objects. In *Proc. New Interfaces for Musical Expression*, pages 154–157, Genova, Italy, 2008.

[10] K. v. d. Doel. Physically based models for liquid sounds. *ACM Trans. Appl. Percept.*, 2(4):534–546, 2005.

[11] P. Dourish. *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press, Cambridge, Mass., 2001.

[12] C. Drioli, P. Polotti, D. Rocchesso, S. Delle Monache, K. Adiloglu, R. Anniés, and K. Obermayer. Auditory representations as landmarks in the sound design space. In *Proc. Conference on Sound and Music Computing*, pages 315–320, Port, Portugal, july 2009.

[13] P. Dupont, V. Hayward, B. Armstrong, and F. Altpeter. Single state elasto-plastic friction models. *IEEE Trans. Automat. Contr.*, 47(5):787–792, 2002.

[14] G. Essl and S. O'Modhrain. An enactive approach to the design of new tangible musical instruments. *Organised Sound*, 11(3):285–296, 2006.

[15] A. Findeli. Rethinking design education for the 21st century: Theoretical, methodological, and ethical discussion. *Design Issues*, 17(1):5–17, 2001.

[16] K. Franinovic. Toward basic interaction design. *Elisava Temes de Disseny Journal*, 2009.

[17] K. Franinovic and Y. Visell. Strategies for sonic interaction design: from context to basic design. In *Proceedings of the 14th International Conference on Auditory Display*, Paris, France, 2008. inproceedings.

[18] W. Gaver, J. Bowers, T. Kerridge, A. Boucher, and N. Jarvis. Anatomy of a failure: how we knew when our design went wrong, and what we learned from it. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 2213–2222, New York, NY, USA, 2009. ACM.

[19] W. W. Gaver. What in the world do we hear? an ecological approach to auditory event perception. *Ecological Psychology*, 5:1–29, 1993.

[20] D. Hug. Investigating narrative and performative sound design strategies for interactive commodities. In S. Ystad, M. Aramaki, R. Kronland-Martinet, and K. Jensen, editors, *Auditory Display - 6th International Symposium, CMMR/ICAD 2009, Copenhagen, Denmark, May 18-22, 2009, Revised Papers*, volume 5954 of *Lecture Notes in Computer Science*. Springer, 2010.

[21] A. Jylhä and C. Erkut. A hand clap interface for sonic interaction with the computer. In *CHI EA '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 3175–3180, New York, NY, USA, 2009. ACM.

[22] G. Lemaitre, O. Houix, Y. Visell, K. Franinovic, N. Misdariis, and P. S. P. Toward the design and evaluation of continuous sound in tangible interfaces: The spinotron. *International Journal of Human-Computer Studies*, 67, 2009. Special issue on Sonic Interaction Design, to appear.

[23] E. Lupton and J. C. Phillips. *Graphic Design: The New Basics*. Princeton Architectural Press, 2008.

[24] D. W. Marhefka and D. E. Orin. A compliant contact model with nonlinear damping for simulation of robotic systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 29(6):566–572, 1999.

[25] E. R. Miranda and M. Wanderley. *New Digital Musical Instruments: Control And Interaction Beyond the Keyboard (Computer Music and Digital Audio Series)*. A-R Editions, Inc., 2006.

[26] P. Polotti, S. Delle Monache, S. Papetti, and D. Rocchesso. Gamelunch: forging a dining experience through sound. In *CHI '08: CHI '08 extended abstracts on Human factors in computing systems*, pages 2281–2286, New York, NY, USA,

2008. ACM.

[27] M. Rinott and I. Ekman. Using vocal sketching for designing sonic interactions. In *to be published in DIS 2010: Proceedings of the Designing Interactive Systems Conference*, 2010.

[28] D. Rocchesso, R. Bresin, and M. Fernström. Sounding objects. *IEEE Multimedia*, 10(2):42–52, april 2003.

[29] D. Rocchesso, P. Polotti, and S. Delle Monache. Designing continuous sonic interaction. *International Journal of Design*, 3(3), December 2009.

[30] D. Rocchesso, S. Serafin, F. Behrendt, N. Bernardini, R. Bresin, G. Eckel, K. Franinovic, T. Hermann, S. Pauletto, P. Susini, and Y. Visell. Sonic interaction design: sound, information and experience. In *CHI '08: CHI '08 extended abstracts on Human factors in computing systems*, pages 3969–3972, New York, NY, USA, 2008. ACM.

[31] D. A. Schön. *The Reflective Practitioner*. Basic Books, London, UK, 1983.

[32] B. Smith and R. Casati. Naive physics: An essay in ontology. *Philosphical Psychology*, 7(2):225–244, 1994.

[33] J. O. Smith. Principles of digital waveguide models of musical instruments. In M. Kahrs and K. Brandenburg, editors, *Applications of Digital Signal Processing to Audio and Acoustics*, pages 417–466. Kluwer Academic Publishers, 1998.

[34] E. Stolterman. The nature of design practice and implications for interaction design research. *International Journal of Design*, 2(1):55–65, April 2008.

[35] P. Susini, N. Misdariis, G. Lemaitre, D. Rocchesso, P. Polotti, K. Franinovic, Y. Visell, and K. Obermayer. Closing the loop of sound evaluation and design. In *Proceedings of the 2nd ISCA/DEGA Tutorial and Research Workshop on Perceptual Quality of Systems*, Berlin, 2006.

[36] V. Valimaki, J. Pakarinen, C. Erkut, and M. Karjalainen. Discrete-time modelling of musical instruments. *Reports on Progress in Physics*, 69(1):1–78, January 2006.

[37] K. van den Doel, P. G. Kry, and D. K. Pai. Foleyautomatic: physically-based sound effects for interactive simulation and animation. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 537–544, New York, NY, USA, 2001. ACM.

Preliminary Experiment Combining Virtual Reality Haptic Shoes and Audio Synthesis

Rolf Nordahl[1], Amir Berrezag[2], Smilen Dimitrov[1],
Luca Turchet[1], Vincent Hayward[2], and Stefania Serafin[1]

[1] Aalborg University Copenhagen, Medialogy, Lautrupvang 15, 2750
Ballerup, Denmark
{rn,sd,tur,sts}@media.aau.dk
[2] UPMC Univ Paris 06, Institut des Systèmes Intelligents et de Robotique, 4 place Jussieu,
75005 Paris, France
{amir.berrezag,vincent.hayward}@isir.upmc.fr

Abstract. We describe a system that provides combined auditory and haptic sensations to simulate walking on different grounds. It uses a physical model that drives haptic transducers embedded in sandals and headphones. The model represents walking interactions with solid surfaces that can creak, or be covered with crumpling material. In a preliminary discrimination experiment, 15 participants were asked to recognize four different surfaces in a list of sixteen possibilities and under three different conditions, haptics only, audition only and combined haptic-audition. The results indicate that subjects are able to recognize most of the stimuli in the audition only condition, and some of the material properties such as hardness in the haptics only condition. The combination of auditory and haptic cues did not improve recognition significantly.

Keywords: physical models, walking sounds, audio-haptic interaction.

1 Introduction

Multimodality is an increasingly common feature of interactive systems. Whilst most studies focus on the interaction between vision and audition or between vision and touch, interaction between touch and audition is also strong because of two sources of sensory information have high temporal resolution. The perception literature contains many reports of audio-tactile interaction effects, see [1,2,3,4] for examples and surveys, and there has been studies directed at leveraging audiotactile to enhance interaction with virtual worlds [5,6,7,8,9,10].

We typically spend a great amount of our waking hours interacting with the world through our feet, performing simultaneous auditory and haptic probing. However, most studies, so far, both from the perception and from the virtual reality literature, have focused on the hands. A notable exception is the work of Giordano et al., who showed that the feet were also effective at probing the world with discriminative touch, with and without access to auditory information. Their results suggested that integration of foot-haptic and auditory information does follow simple integration rules [11].

Almost all haptic device development research is directed at stimulating the hand but what about the foot? Similarly to research on haptic devices for the hand, the approaches

A.M.L. Kappers et al. (Eds.): EuroHaptics 2010, Part II, LNCS 6192, pp. 123–129, 2010.
© Springer-Verlag Berlin Heidelberg 2010

to stimulate the foot broadly follow two directions [12]. The force-feedback option has been explored considering either rolling or ground-referenced devices, e.g., [13,14]. This approach typically involves considerable engineering challenges and cost. The vibrotactile option has also been explored. There are two possible ways to provide the signal. Actuators can be embedded in the floor, see [15] for a survey, or they can be embedded in worn shoes. In the later case, however, to our knowledge, there has been no attempt up until now to aim for any kind of reproduction fidelity needed for a virtual reality simulation. Only signaling functions were considered [16,17].

We describe a preliminary study carried out with haptic feedback sandals that employed a newly introduced broadband vibrotactile transducer. The transducers as well as the headphone worn by the participants were driven by the same physical model. The model, presented in greater detail elsewhere [18], is able to represent, for virtual reality purposes, the kind of interaction that one might expect from stepping on solid surfaces that may or may not present the characteristics of a creaking material or be covered by crumpling objects such as dry leaves. Within the auditory modality, this model was already shown to enable good discrimination among these materials.

Here, we used the same model to drive the haptic simulations in order to investigate whether the experience of a virtual world may be enhanced by providing haptic feedback through the feet. In the present study, the participants passively received sensory information through touch and audition which is not the condition in which the system is intended to be used ultimately. Even though sensorimotor coupling was inexistent, as if another person did the walking for the participants, interesting results were obtained.

2 Simulation Hardware and Software

2.1 Haptic Hardware

A pair of light-weight sandals size 43 was procured (Model Arpenaz-50, Decathlon, Villeneuve d'Ascq, France). This particular model has light, stiff foam soles that are easy to gouge and fashion. Two cavities were made in the thickness of the sole to accommodate two 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. As indicated in Fig. 1, one actuator was placed under the heel of the wearer and the other under the ball of the foot. They were bonded in place to ensure good transmission of the vibrations inside the soles. When activated, vibrations propagated far in the light, stiff foam. In the present configuration, these two actuators were driven by the same signal but could be activated separately to emphasize, for instance, the front or back activation, to strike a balance, or to realize other effects such as modulating different, back-front signals during heel-toe movements.

The sole has force sensors intended to pick the foot-floor interaction force in order to drive the audio and haptic synthesis. They were not used in the present study. As far as auditory feedback is concerned, it was delivered through closed headphones (DT 770, beyerdynamic, Heilbronn, Germany).

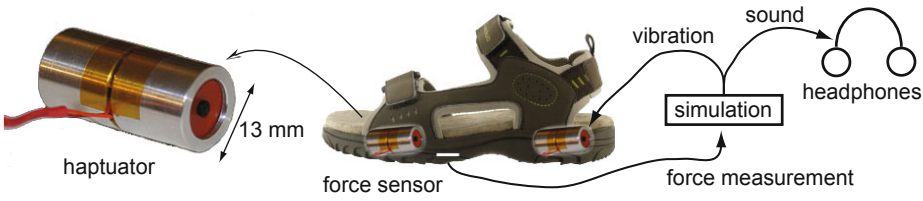


Fig. 1. System (one shoe shown). Left: recoil-type actuation from Tactile Labs Inc. The moving parts are protected by an alumimum enclosure able to bear the weight of a person. Middle: approximate location of the actuators in the sandal. Right: system diagram showing the interconnections. Here the force signal was not used.

2.2 Audio-Haptic Simulation

This model and its discretization is described elsewhere in detail [19]. The model has been recently adapted to the audio simulation of footsteps [18]. Here, we used the same model to drive the haptic and the audio synthesis. It is briefly recalled below.

A footstep sound may be considered to cause multiple micro-impacts between a sole, i.e., an *exciter*, and a floor, i.e., a *resonator*. Such interaction can be either discrete, as in the case of walking on a solid surface, or continuous, as in the case of a foot sliding across the floor.

In the simulation of discrete impacts, the excitation is brief and has an unbiased frequency response. The interaction is modelled by a Hunt-Crossley-type interaction where the force, f, between two bodies, combines hardening elasticity and a dissipation term [20]. Let x represent contact interpenetration and $\alpha > 1$ be a coefficient used to shape the nonlinear hardening, the special model form we used is

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

The model described was discretized as proposed in [21].

If the interaction called for slip, we adopted a model where the relationship between relative velocity v of the bodies in contact and friction force f is governed by a differential equation rather than a static map [22]. Considering that friction results from a large number of microscopic damped elastic bonds with an average deflection z, a viscous term, $\sigma_2 v$, and a noise term, $\sigma_3 w$, to represent roughness, we have

$$f(z,\dot{z},v,w) = \sigma_0 z + \sigma_1 \dot{z} + \sigma_2 v + \sigma_3 w.$$

The force specified by these models is applied to a virtual mass which produces a displacement signal that is then processed by a linear shaping filter intended to represent the resonator.

Stochastic parameterization is employed to simulate particle interactions thereby avoiding to model each of many particles explicitely. 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.

We used this approach to model both solid and aggregate surfaces. A solid surface is represented by an impact and a slide. The impact model alone was used to recreate the sound and the feel produced when walking on wood. The friction model was tuned to simulate walking on creaking wood. To simulate walking on aggregate grounds, we used a physically informed sonic models (PhiSM) algorithm [23]. The synthesis was tuned to simulate snow and forest underbrush.

These algorithms were implemented as an extension to the Max/MSP platform[1] to drive both the auditory and haptic feedback.

3 Preliminary Evaluation

We conducted a within-subjects experiment whose goal was to assess the ability of subjects to recognize the surfaces they were exposed to using auditory and haptic stimuli, and a combination of both. The experiment entailed asking three groups of fifteen participants to passively experience the stimuli described in the previous section. The first group received the haptic stimuli only, the second the audio only, and the third the combined stimuli. The synthesis was tuned to evoke four different surfaces: wood, creaking wood, snow, and underbrush. Participants had to select in a list of sixteen different materials one that matched best their experience. They also rated the realism and quality of their experience on a seven-point Likert scale.

Procedure: The participants were asked to wear the sandals, and the headphones described in the previous section, and to sit in a chair. In the condition with haptics only they wore earplugs and sound protection headsets. Participants were asked to recognize the stimuli they were exposed to. They were given a list of sixteen options, made of fifteen materials: wood, creaking wood, underbrush, snow, frozen snow, beach sand, gravel, metal, high grass, dry leaves, concrete, dirt, puddles, water, and carpet plus an additional "I don't know" option. The materials were chosen in order to cover a large set of solid and aggregate materials, which could represent realistic walking surfaces with characteristics similar to the ones simulated. The reason to add the "I don't know" option was in order to allow subjects to express their complete uncertainty in the recognition of materials, which would not have been possible using forced choices. Each of the four simulated surfaces were presented twice in a randomized order. When presented one of the four stimuli, participants had to match it to one in the list and rated the realism and quality of the simulations. At the conclusion of the experiment, participants were asked to leave comments.

Participants: The forty five volunteers (students and faculty at the Engineering college of Copenhagen; 31 men and 14 women; average age =24.5, sd=4.6) were randomly assigned to one of the three groups (audio only, haptic only or audio-haptic) for a total of 15 participants per condition. None reported hearing problems or other sensory impairments. In order for the size of sandals not to affect performance, subjects wore shoes sizes from 41 and 45 (as mentioned before, the sandals were size 43).

[1] www.cycling74.com

Results and discussion: Table 1 shows the confusion matrices produced by the three groups.

Table 1. Confusion matrices with: haptics, audio, and combined haptics and audio

		WD	CW	SW	UB	—	FS	BS	GL	MT	HG	DL	CC	DR	PD	WT	CP
haptics	WD	6	2				4	1	2	5		2	6	2			
	CW		11	2	4	5	5	2	1								
	SW		1	13		3	3	2	4		1	2				1	
	UB	1		2	4		10	2	8			1		2			
audio	WD	11	1	1	1	11	1		1	1			1			1	
	CW	1	27				2										
	SW			23			3	2				2					
	UB			2	20	1	2		1		1	3					
haptics and audio	WD	13	1	1		9	1			1			3	1			
	CW		29					1									
	SW			23			2		3						2		
	UB			2	16		6		4			2					

Legend:

WD	wood	CW	creaking wood
SW	snow	UB	underbrush
—	don't know	FS	Frozen snow
BS	beach sand	GL	Gravel
MT	metal	HG	High grass
DL	dry leaves	CC	concrete
DR	dirt	PD	puddles
WT	Water	CP	carpet

From the results, it can be noticed that haptic cues alone gave the subjects the possiblity to recognize surfaces categories but with poor fine discrimination. This is the case for the wood simulation that was easily confused with metal, concrete or dirt. A solid surface was not confused with aggregates such as snow, underbrush, grass or with soft surfaces such as puddles or carpet. Recognition rates with audio stimuli were much higher as seen in the corresponding dominant diagonal. The audio simulation of wood still caused some confusion as seen in the numerous 'don't know' answers that were not present with the haptic experience. The small number of confusions was mostly among the aggregates. When the two stimuli (identical at the signal level) were delivered simultaneously, performance was not necessarily better as if conflict was created; such phenomenon is noticeable during the presentation of simulated underbrush. What is quite interesting is that a similar phenomenon occurs during the multimodal identification of *real* materials [11] (full report forthcoming). In both haptic and auditory modality, the friction simulation was an important cue which facilitated the recognition of creaking wood.

Table 2 shows the degree to which participants judged the realism and quality of the experience. The degree of realism was calculated by looking only at that data from correct

Table 2. Average realism and quality scores from a seven-point Likert scale

	realism				quality			
	wood	creaking	snow	underbrush	wood	creaking	snow	underbrush
haptics	3.3	4.3	4.8	3.2	2.8	4.0	4.2	4.3
audio	1.8	3.5	5.2	5.5	3.2	3.8	5.1	4.9
combined	3.3	3.6	5.3	4.6	4.0	4.7	5.2	4.6

answers, i.e., when the surfaces were correctly recognized. As far as the quality judgement is concerned, the data was based on all the answers different from 'don't know'.

Overall, the aggregate surfaces were considered as more realistic and with a higher quality than the solid surfaces. In all conditions, the addition of the creaking sound to the simulation of wood increased both the rated quality and the realism.

4 Conclusions and Future Work

We described a system able to simulate the auditory and haptic sensation of walking on different materials and presented the results of a preliminary surface recognition experiment. This experiment was conducted under three different conditions: auditory feedback, haptic feedback, and both.

By presenting the stimuli to the participants passively sitting in a chair, we introduced a high degree of control on the stimulation. However, this method of delivery is highly contrived since it eliminates the tight sensorimotor coupling that is natural during walking and foot interaction. It is true for the auditory channel, but even more so for the haptic channel. In spite of these drastically constrained conditions, performance was surprisingly good.

We are currently running follow up experiments allowing subjects to walk in a controlled laboratory, where their steps are tracked and used to drive the simulation. We believe introducing a higher level of interactivity will significantly enhance the recognition rates as well as the perceived quality and realism of the simulation.

Acknowledgments

The research leading to these results has received funding from the European Community 's Seventh Framework Programme under FET-Open grant agreement n. 222107 NIW - Natural Interactive Walking.

References

1. Lederman, S.: Auditory texture perception. Perception (1979)
2. Jousmaki, V., Hari, R.: Parchment-skin illusion: sound-biased touch. Current Biology 8(6), R190–R191 (1998)
3. Shimojo, S., Shams, L.: Sensory modalities are not separate modalities: plasticity and interactions. Current Opinion in Neurobiology 11(4), 505–509 (2001)

4. Bresciani, J.P., Ernst, M.O., Drewing, K., Bouyer, G., Maury, V., Kheddar, A.: Feeling what you hear: Auditory signals can modulate tactile tap perception. Experimental Brain Research 162, 172–180 (2005)
5. Ramstein, C., Hayward, V.: The pantograph: A large workspace haptic device for a multimodal human-computer interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2004, ACM/SIGCHI Companion-4/94, pp. 57–58 (1994)
6. DiFranco, D., Beauregard, G.L., 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 (1997)
7. DiFilippo, D., Pai, D.K.: Contact interaction with integrated audio and haptics. In: Proceedings of the International Conference on Auditory Display, ICAD (2000)
8. Pai, D., Doel, K., James, D., Lang, J., Lloyd, J., Richmond, J., Yau, S.: Scanning physical interaction behavior of 3d objects. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, pp. 87–96 (2001)
9. Sreng, J., Bergez, F., Legarrec, J., Lécuyer, A., Andriot, C.: Using an event-based approach to improve the multimodal rendering of 6dof virtual contact. In: Proceedings of ACM Symposium on Virtual Reality Software and Technology (ACM VRST), pp. 173–179 (2007)
10. Avanzini, F., Crosato, P.: Integrating physically based sound models in a multimodal rendering architecture. The Journal of Visualization and Computer Animation 17(3-4), 411–419 (2006)
11. Giordano, B.L., Mcadams, S., Visell, Y., Cooperstock, J., Yao, H.Y., Hayward, V.: Non-visual identification of walking grounds. Journal of the Acoustical Society of America 123(5), 3412 (2008)
12. Hayward, V., MacLean, K.E.: Do it yourself haptics, part-i. IEEE Robotics and Automation Magazine 14(4), 88–104 (2007)
13. Iwata, H., Yano, H., Tomioka, H.: Powered shoes. In: ACM SIGGRAPH 2006 Emerging technologies, p. 28 (2006)
14. Schmidt, H., Hesse, S., Bernhardt, R., Krüger, J.: Hapticwalker—a novel haptic foot device. ACM Transactions on Applied Perception 2(2), 166–180 (2005)
15. Visell, Y., Law, A., Cooperstock, J.R.: Touch is everywhere: Floor surfaces as ambient haptic interfaces. IEEE Transactions on Haptics 2, 148–159 (2009)
16. Fu, X., Li, D.: Haptic shoes: representing information by vibration. In: Proceedings of the 2005 Asia-Pacific Symposium on Information Visualisation, pp. 47–50 (2005)
17. Magana, M., Velazquez, R.: On-shoe tactile display. In: IEEE International Workshop on Haptic Audio Visual Environments and Games (HAVE 2008), pp. 114–119 (2008)
18. Nordahl, R., Serafin, S., Turchet, L.: Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In: Proc. IEEE VR 2010 (2010)
19. Avanzini, F., Serafin, S., Rocchesso, D.: Interactive simulation of rigid body interaction with friction-induced sound generation. IEEE Transactions on Speech and Audio Processing 13(5 Part 2), 1073–1081 (2005)
20. Hunt, K.H., Crossley, F.R.E.: Coefficient of restitution interpreted as damping in vibroimpact. ASME Journal of Applied Mechanics 42(2), 440–445 (1975)
21. Avanzini, F., Rocchesso, D.: Modeling collision sounds: Non-linear contact force. In: Proc. COST-G6 Conf. Digital Audio Effects (DAFx 2001), pp. 61–66 (2001)
22. Dupont, P., Hayward, V., Armstrong, B., Altpeter, F.: Single state elastoplastic friction models. IEEE Transactions on Automatic Control 47(5), 787–792 (2002)
23. Cook, P.: Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. Computer Music Journal 21(3), 38–49 (1997)

Sound Synthesis and Evaluation of Interactive Footsteps for Virtual Reality Applications

Rolf Nordahl * Stefania Serafin† Luca Turchet‡

Medialogy, Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, DK

ABSTRACT

A system to synthesize in real-time the sound of footsteps on different materials is presented. The system is based on microphones which allow the user to interact with his own footwear. This solution distinguishes our system from previous efforts that require specific shoes enhanced with sensors. The microphones detect real footsteps sounds from users, from which the ground reaction force (GRF) is estimated. Such GRF is used to control a sound synthesis engine based on physical models. Evaluations of the system in terms of sound validity and fidelity of interaction are described.

Keywords: sound synthesis, physical models, footsteps sounds, auditory perception

Index Terms: H.5.5 [Information Systems]: Information Interfaces and Presentation—Sound and Music Computing; H.5.2 [Information Systems]: Information Interfaces and Presentation—User Interfaces

1 INTRODUCTION

The development of efficient yet accurate simulation algorithms, together with improvements in hardware technology, has boosted the research on auditory display and physically based sound models for virtual environments (VEs)[27, 23, 7].

The addition of auditory cues and their importance in enhancing the sense of immersion and presence is a recognized fact in virtual environment research and development. Most prior work in this area has focused on sound delivery methods [25, 24], sound quantity and quality of auditory versus visual information [4] and 3D sound [11, 28]. Recent studies have investigated the role of auditory cues in enhancing self-motion and presence in virtual environments [17, 15, 26].

Self-generated sounds have been often used as enhancements to VEs and first-person 3D computer games – particularly in the form of footstep sounds accompanying self-motion or the presence of other virtual humans.

Such sounds are used to produce embodiment and a sense of weight with the overall goal of heightening the sense of "realness" to the character or person. Usually such sounds are taken from sound libraries or recorded by *Foley* artists who put shoes in their hands and interact with different materials to simulate the act of walking.

Recently, several physics based algorithms have been proposed to simulate the sounds of walking. One of the pioneers in this field is Perry Cook, who proposed a collection of physically informed stochastic models (PhiSM) simulating several everyday sonic events [5]. Among such algorithms the sounds of people walking on different surfaces were simulated [6]. A similar algorithm was also proposed in [10], where physically informed models reproduced several aggregate surfaces. Procedural sound synthesis of walking has also been recently described in [9].

Previous work on interactive footwear, such as the research performed by Paradiso and coworkers [21, 3], consisted of designing shoes augmented with sensors used to control footsteps sounds. A smaller number of examples, such as recent work of Nordahl [20] and Law et al. [18], have even aimed to provide multimodal cues linked to footsteps events in such environments.

In this paper, we are particularly interested in developing a solution which requires a minimum amount of sensing technology and is shoe independent, which means that subjects can keep their own footwear while using the system. This creates several advantages from the interaction side: users do not need to wear ad-hoc designed shoes, whose wearability is decreased by the addition of several sensors.

We propose an interactive system which enables a designer to synthesize in real-time footsteps sounds of different materials. We describe the results of experiments whose goal is to test the degree of realism of the system, the ability of subjects to recognize the virtual material they are walking on, and the fidelity of interaction. The ultimate goal is to integrate this system in the simulation of multimodal virtual environments where the act of walking plays an important role.

2 THE SOUND SYNTHESIS ENGINE

We developed a physically based sound synthesis engine able to simulate the sounds of walking on different surfaces. Acoustic and vibrational signatures of locomotion are the result of more elementary physical interactions, including impacts, friction, or fracture events, between objects with certain material properties (hardness, density, etc.) and shapes. The decomposition of complex everyday sound phenomena in terms of more elementary ones has been an organizing idea in auditory display research during recent decades [12]. In our simulations, we draw a primary distinction between solid and aggregate ground surfaces, the latter being assumed to possess a granular structure, such as that of gravel.

2.1 Solid surfaces

Sonic interactions between solid surfaces have been extensively investigated, and results are available which describe the relationship between physical and perceptual parameters of objects in contact [16, 27]. Such sounds are typically short in duration, with a sharp temporal onset and relatively rapid decay.

A common approach to synthesizing such sounds is based on a lumped source-filter model, in which an impulsive excitation $s(t)$, modeling the physics of contact, is passed through a linear filter $h(t)$, modeling the response of the vibrating object as $y(t) = s(t) \star h(t)$.

Modal synthesis [1] is one widely adopted implementation of this idea. In this synthesis technique, the response model $h(t)$ is decomposed in terms of the resonant frequencies f_i of the vibrating object, also known as the modes of the object. The re-

*e-mail: rn@media.aau.dk
†e-mail:sts@media.aau.dk
‡e-mail:tur@media.aau.dk

IEEE Virtual Reality 2010
20 - 24 March, Waltham, Massachusetts, USA
978-1-4244-6236-0/10/$26.00 ©2010 IEEE

sponse is modeled as a bank of filters with impulse response $h(t) = \sum_i a_i e^{-b_i t} \sin(2\pi f_i t)$, where a_i represent the amplitudes of the modes, b_i the decay rates of the modes, and f_i the frequencies of the modes.

A footstep sound can be considered as the result of multiple micro-impact sounds between a shoe and a floor. The set of such micro-events can be thought as the result of the interaction between an *exciter* and a *resonator*. The exciter is represented by the interaction between shoe and ground. Such interaction can be either continuous, as in the case of a foot sliding across the floor, or discrete, as in the case of walking on a solid surface.

To simulate such scenarios, both an impact and friction model were implemented.

In the impact model, the excitation corresponding to each impact $s(t)$ is assumed to possess a short temporal extent and an unbiased frequency response. Such excitation consists of a discrete-time model of the force f between the two bodies, dependent on additional parameters governing the elasticity of the materials, their velocity of impact $\dot{x}$, and masses:

$$\mathrm{f}(\mathrm{x},\dot{x}) = \begin{cases} -k\mathrm{x}^{\alpha} - \lambda\, x^{\alpha}\, \dot{x} & \text{if } x > 0 \\ 0 & x \leq 0 \end{cases}$$

where α depends on the local geometry around the contact surface, and x stands for the compression of the exciter (when $x > 0$ the two objects are in contact) [2].

In the friction model we adopted a dynamic model, where the relationship between relative velocity v of the bodies in contact and friction force f is represented through a differential equation rather than static mapping. Assuming that friction results from a large number of microscopic elastic bonds, called bristles in [8], the v-to-f relationship is expressed as:

$$f(z,\dot{z},v,w) = \sigma_0 z + \sigma_1 \dot{z} + \sigma_2 v + \sigma_3 w$$

where z is the average bristle deflection, the coefficient σ_0 is the bristle stiffness, σ_1 the bristle damping, and the term $\sigma_2 v$ accounts for linear viscous friction. The fourth component $\sigma_3 w$ relates to surface roughness, and is simulated as fractal noise.

2.2 Aggregate surfaces

To synthesize aggregate surfaces, we implemented the physically informed sonic models (PhiSM) algorithm [5].

This model simulates particle interactions by using a stochastic parameterization. This means that the different particles do not have to be modeled explicitly, but only the probability that particles will create some noise is simulated. For many particle systems, this phenomenon is well taken into account by using a simple Poisson distribution, where the sound probability is constant at each time step, giving rise to an exponential probability waiting time between events.

2.3 Implementation

Using the algorithms described in the previous sections, we implemented a comprehensive collection of footstep sounds. As solid surfaces, we implemented metal and wood. In these materials, the impact model was used to simulate the act of walking, while the friction model was used to simulate the sound of creaking wood.

As aggregate surfaces, we implemented gravel, sand, snow, forest underbrush, dry leaves, pebbles and high grass. The simulated metal, wood and creaking wood surfaces were furthermore enhanced by using some reverberation. The role of reverberation is discussed in the testing section.

The sound synthesis algorithms were implemented in C++ as external libraries for the Max/MSP sound synthesis and multimedia real-time platform.[1] To enable compatibility with the Pure Data platform, [2] the algorithms were implemented using Flext. [3]

In our simulations, designers have access to a sonic palette making it possible to manipulate all such parameters, including material properties. One of the challenges in implementing the sounds of different surfaces was to find suitable combinations of parameters which provided a realistic simulation. For each simulated surface, recorded sounds were analyzed according to their combinations of events, and each subevent was simulated independently. As an example, the sound produced while walking on dry leaves is a combination of granular sounds with long duration both at low and high frequencies, and noticeable random sounds with not very high density that give to the whole sound a crunchy aspect. These different components were simulated with several stochastic models having the same density, duration, frequency and number of colliding objects.

The amplitude of the different components were also appropriately weighed, according to the same contribution present in the corresponding real sounds. Finally, a scaling factor for the subcomponents volumes gives to the whole sound an appropriate volume, in order to recreate a similar sound level which it would happen during a real footstep on each particular material.

A pilot test was run to ascertain that such a global volume plays an important role in the judgments concerning the sounds' realism and in the recognition of the surface material. Indeed, wrong settings for such a parameter can cause wrong recognitions.

3 Controlling the sound synthesis engine

The developed sound synthesis engine is controlled as following: users are asked to walk inside an area delimited by four microphones placed on the floor in a square configuration. Specifically, we used four Shure BETA 91,[4] high performance condenser microphones with a tailored frequency response designed specifically for kick drums and other bass instruments. The microphones' features made them a good candidate for the purpose of capturing footsteps sounds. In the interaction between a foot and a sole, the exciter is usually called ground reaction force (GRF), i.e., the reaction force supplied by the ground at every step. The aim of the phase of analysis has been that of extracting the GRF from the acoustic waveform. The real footsteps sounds produced are detected by the microphone, and their GRF extracted and used to control the temporal evolution of the synthetic footsteps. An example of a footstep sound and its corresponding GRF is shown in Figure 1.

4 Experiment

We conducted different experiments whose goal is to investigate the ability of subjects to recognize the different walking sounds they were exposed to. The study of human perception of locomotion sounds has addressed several properties of walking sound sources: the gender [19, 13] and posture of a walker [22], the emotions of a walker [13], the hardness and size of the shoe sole [13], and the ground material [14].

Such studies have been concerned only with recognition of sounds in an off-line scenario, where subjects were asked to listen to some sounds and classify them. In this experiment, we are interested in having subjects classify sounds both off-line, but also in an active settings, i.e., by using the developed interactive system. One of our hypotheses is that the recognition when using the interactive system is higher than in the off-line setup.

Moreover, we conducted an experiment using recorded real sounds in order to compare their recognition rate with that of the developed synthesized sounds.

[1] www.cycling74.com

[2] www.puredata.org

[3] http://puredata.info/Members/thomas/flext

[4] http://www.shure.com/

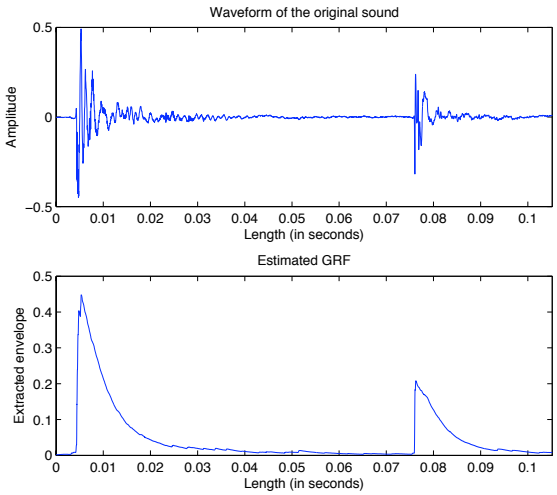


Figure 1: A footstep sound (top) and the corresponding calculated GRF (bottom).

4.1 Methods

Three kinds of experiments were conducted:

1. experiment 1: recognition of footsteps sounds generated in real time by the subjects.

2. Experiment 2: recognition of synthesized recorded footsteps sounds.

3. Experiment 3: recognition of real recorded footsteps sounds.

The sounds provided during experiment 1 were synthesized sounds generated in real time while subjects were walking using the interactive system described in the previous section.
The sounds provided during experiment 2 consisted of recordings of footsteps sounds generated by the use of the interactive system. The sounds provided during experiment 3 consisted of recordings of real footsteps sounds on different surfaces. Such sounds were chosen among those available on the Hollywood Edge sound effects library.[5] Each sound in experiment 2 and 3, composed of several footsteps, had duration of about 7 seconds.

Participants were exposed to 26 trials in experiment 1 and 2, and 30 trials in experiment 3. During experiments 1 and 2, 13 stimuli were presented twice in randomized order. The stimuli consisted of footsteps sounds on the following surfaces: beach sand, gravel, dirt plus pebbles (like in a country road), snow (in particular deep snow), high grass, forest underbrush (a forest floor composed by dirt, leaves and branches breaking), dry leaves, wood, creaking wood and metal. To simulate room characteristics, footsteps sounds on wood, creaking wood and metal were enhanced adding a certain amount of reverberation.

In experiment 3, fifteen stimuli were presented twice in randomized order. They consisted of the previous mentioned sounds without the reverberated ones, more footsteps sounds on carpet, concrete, frozen snow, puddles and water.

4.1.1 Participants

Fourtyfive participants were divided in three groups (n = 15) to perform the three between-subjects experiments. The three groups were composed respectively of 6 men and 9 women, aged between 19 and 29 (mean=22.13,standard deviation=2.47), 8 men and 7 women, aged between 20 and 35 (mean=22.73,standard deviation=4.01), and 10 men and 5 women, aged between 20 and 29 (mean=23.13,standard deviation=2.39). All participants reported normal hearing conditions. All participants were naive with respect to the experimental setup and to the purpose of the experiment.

During experiment 1 the shoes used by subjects were sneakers, trainers, boots and other kinds of shoes with rubber soil.
The participants took in average about 24, 15 and 12 minutes for experiments 1, 2 and 3 respectively.

4.1.2 Setup

Figure 2: A subject performing experiment 1.

All experiments were carried out in an acoustically isolated laboratory where three setups were installed. The setup for experiment 1 consisted of the interactive system, and the participants were asked to use it in order to generate the footsteps sounds in real time (see Figure 2).
The setup for experiments 2 and 3 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.

[5] www.hollywoodedge.com/

	Correct answers	Wrong answers	I don't know	Realism	Quality
Beach Sand	50.	36.67	13.33	5.3	5.15
Gravel	70.	26.66	3.33	5.57	5.57
Dirt pebbles	10.	86.67	3.33	5.75	4.73
Snow	80.	16.67	3.33	5.81	5.67
High Grass	0.	83.33	16.67	-	5.4
Forest Underbrush	63.33	33.33	3.33	4.84	4.7
Dry Leaves	40.	60.	0.	4.75	4.96
Wood	46.67	20.	33.33	3.93	4.55
Creaking Wood	93.33	6.67	0.	5.16	5.17
Metal	80.	13.33	6.67	3.33	4.77
Wood plus Reverb	20.	70.	10.	3.83	4.76
Creaking Wood plus Reverb	93.33	3.33	3.33	4.93	5.17
Metal plus Reverb	83.33	10.	6.67	3.6	5.23

Table 1: Results of experiment 1: recognition (in percentage) of the surfaces with the interactive system.

4.1.3 Task

During experiment 1 the participants were asked to wear a pair of headphones and to walk in the area delimited by the microphones. They were given the list of different surfaces to be held in one hand, presented as non-forced alternate choice.

During the act of walking they listened simultaneously to footsteps sounds on a different surface according to the stimulus presented. The task consisted of answering by voice the following three questions after the presentation of the stimulus:

1. Which surface do you think you are walking on? For each stimulus choose an answer in the following list: 1) beach sand, 2) gravel, 3) dirt plus pebbles, 4) snow, 5) high grass, 6) forest underbrush, 7) dry leaves, 8) wood, 9) creaking wood, 10) metal, 11) carpet, 12) concrete, 13) frozen snow, 14) puddles and water, 15) I don't know.

2. How close to real life is the sound in comparison with the surface you think it is? Evaluate the degree of realism on a scale from 1 to 7 (1=low realism, 7=high realism).

3. Evaluate the quality of the sound on a scale from 1 to 7 (1=low quality, 7=high quality).

At the end of the experiment, subjects were asked some questions concerning the naturalness of the interaction with the system and to comment on its usability and possible integration in a virtual reality environment.

The task in experiment 2 was similar to experiment 1. However, subjects were sitting on a chair, listening to the sounds through headphones and interacting with the interface mentioned in section 4.1.2. The task in experiment 3 was similar to experiment 2, but 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. At the end of the experiments 2 and 3 the subjects were also given the opportunity to leave an open comment on their experience interacting with the system.
The list included a range of materials wider than those presented in experiment 1 and 2 (see section 4.1). Conversely, in experiment 3 all the materials in the list were presented. The subjects were informed that they could choose the same material more than one time and that they were not forced to choose all the materials in the list. In addition for experiment 1, they could use the interactive system as much as they wanted before giving an answer. Likewise for experiments 2 and 3 they could listen to the sounds as much as they wanted. When passed to the next stimulus they could not change the answer to the previous stimuli.

4.2 Results

The collected answers were analyzed and compared between the three experiments. Results concerning the percentage of correct answers in experiment 1 and 2 are illustrated in Figure 3, while the comparison between the two experiments in terms of realism and quality of the sounds is showed in Figure 4. The degree of realism was calculated only looking at data from correct answers, i.e., when the surface was correctly recognised.

The first noticeable element emerging from both figures is that almost always the use of the interactive system gave rise to a better recognition of the surfaces and a higher evaluation of realism and quality of the proposed sounds, rather than the recorded sounds.
In both experiments the footsteps sounds on snow, creaking wood (with and without reverb), gravel and metal (with reverberation) were correctly recognized with high percentage, while the recognition of the surfaces dirt plus pebbles, high grass and wood (with reverberation) turned out to be wrong most of the times. Regarding the recognition of the other surfaces, good results were found for beach sand and forest underbrush, while correct recognition for dry leaves and wood (without reverberation) were under 50%.

All percentages were higher in experiment 1, although an in-depth analysis shows significant difference only for dry leaves (χ^2 = 4.1761, df = 1, p-value = 0.041) and metal (χ^2 = 4.6886, df = 1, p-value = 0.03036).

An analysis performed on the wrong answers reveals that in average subjects tended to classify erroneously a surface as another belonging to a same category (e.g., wood-concrete, snow-frozen snow, dry leaves-forest underbrush) rather than to different categories (e.g., wood-water, wood-gravel, metal-dry leaves).
Moreover, results show that the addition of the reverberation to the sounds gave rise to better recognitions for metal, and worse for wood plus reverberation, which was perceived most of the times as concrete (not tangible differences were found for the creaking wood). As concerns the comparisons between reverberated and not reverberated sounds in terms of realism negligible differences were found, while in terms of quality the reverberated sounds led to light higher evaluations.

Results of the third experiment are illustrated in Figure 5. Recognition of recorded sounds was quite good in average, with a better performance for the solids and liquids surfaces compared to the aggregate ones. In particular metal, wood, creaking wood, concrete, frozen snow, water and gravel show very high percentages, while the sound of the high grass confirms the negative trend already emerged in the previous experiments. All the other materials present percentages over 50%, with the exception of dry leaves, dirt plus pebbles and forest underbrush, as the data in Table 3 show.

	Correct answers	Wrong answers	I don't know	Realism	Quality
Beach Sand	36.67	63.33	0.	4.18	4.77
Gravel	66.67	33.33	0.	4.9	4.57
Dirt pebbles	10.	90.	0.	4.71	4.37
Snow	83.33	13.33	3.33	5.17	5.14
High Grass	3.33	93.33	3.33	5.	4.3
Forest Underbrush	36.67	60.	3.33	4.22	4.27
Dry Leaves	13.33	86.67	0.	4.5	3.87
Wood	36.67	33.33	30.	4.58	4.
Creaking Wood	76.67	13.33	10.	3.17	3.82
Metal	50.	16.67	33.33	2.93	3.29
Wood plus Reverb	6.67	73.33	20.	6.	4.04
Creaking Wood plus Reverb	76.67	13.33	10.	3.35	3.89
Metal plus Reverb	66.67	6.67	26.67	3.3	4.3

Table 2: Results of experiment 2: recognition (in percentage) of the surfaces with the recorded synthesized sounds.

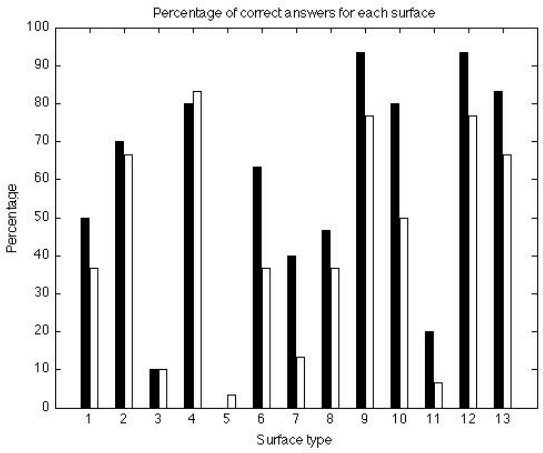


Figure 3: Comparison of the percentages of correct answers for each surface in experiment 1 (black) and 2 (white). Surface type from left to right: 1-beach sand, 2-gravel, 3-dirt pebbles, 4-snow, 5-high grass, 6-forest underbrush, 7-dry leaves, 8-wood, 9-creaking wood, 10-metal, 11-wood plus reverberation, 12-creaking wood plus reverberation, 13-metal plus reverberation. Notice that the missing element in column 5 indicates the fact that none of the subjects was able to recognize the high-grass in experiment 1.

The degree of certainty in the answers seems to be on average consistent with the percentage of correctness (even if there are some exceptions, as for the footsteps on puddles, which were erroneously classified as footsteps through the water).

What emerges from these results is the ability of the subjects in distinguishing materials in the same category for solid surfaces, and their difficulties in the recognition of aggregate surfaces (aspect also confirmed by the comments of the participants). Indeed the analysis of the wrong answers for aggregate surfaces confirms the tendency already showed in the previous experiments, in classifying erroneously a surface as another belonging to a same category.

From the comparison with the results of the recognition of the surfaces presented in the previous two experiments, one can note that the percentage of correct answers for the same surfaces is higher for the experiment 3, with the exception of the snow in both experiment 1 and 2, and of forest underbrush and creaking wood (with reverberation) in experiment 1. Moreover similar percentages were found for beach sand in experiment 1 and 3, and the percentage of gravel is high in the same experiments. Finally the very low percentages for the high grass in the three experiments confirm that this is a sound difficult to recognize.

The final questions of experiment 1 (evaluated on a seven-point Likert scale) show that subjects judged the interaction with the system quite natural (mean= 5.6), and that they felt quite normal (mean= 5.33) and a little bit constrained (mean 2.9) during the act of walking. Indeed, subjects commented on the need of a wider area to walk and of a wireless headphones set.

Finally, regarding the "I do not know" answers the percentage was higher in experiment 2 (10.77%) rather than experiment 1 (7.95%), and lower (3.11%) for the experiment 3. Tables 1,2, 3 show in details the results of experiment 1, 2 and 3 respectively.

5 Discussions

A footstep sound is extremely dependent both on the kind of shoes a person is wearing and on the kind of floor the person is walking on. All sounds were synthesized assuming that the shoes hitting the floor had a solid sole. This aspect is extremely important in the simulation of solid floors. As a matter of fact, when interacting with virtual wood and metal more than one participant commented of having the sensation of wearing a different kind of shoe. More precisely, they commented that they felt like they were wearing a shoe with a solid sole. This indicates the ability of auditory feedback to affect perception of material.

homogeneous ones: as proof of the good success our design we

	Correct answers	Wrong answers	I don't know	Degree of certainty
Beach Sand	53.34	46.67	0.	4.5
Gravel	86.67	13.33	0.	5.07
Dirt pebbles	43.33	56.67	0.	4.8
Snow	63.33	36.67	0.	5.93
Frozen Snow	70.	16.67	13.33	5.66
High Grass	10.	83.33	6.67	3.5
Forest Underbrush	36.67	60.	3.33	4.14
Dry Leaves	56.67	43.33	0.	4.77
Concrete	70.	23.33	6.67	4.43
Wood	80.	16.67	3.33	5.17
Creaking Wood	90.	6.67	3.33	6.14
Metal	96.67	3.33	0.	6.33
Carpet	66.67	23.33	10.	4.78
Puddles	33.33	66.67	0.	5.8
Water	86.67	13.33	0.	6.57

Table 3: Results of experiment 3: recognition (in percentage) of the surfaces with the recorded real sounds.

found that for the sound of the wood and metal floors more than one participant commented that he/she felt like wearing a different kind of shoe, and for the precision with a solid soil.

In general, the use of the interactive system facilitated the recognition task, and the sound quality of the system was perceived as higher.

One peculiar element of the interactive system is the lack of haptic feedback which is present when walking in the real world and is an important element in the perception of a surface. This lack will be compensated in future implementations of the system, where haptic feedback will be integrated. Some subjects also commented on the importance of visual feedback, which would have obviously helped in the recognition task.

6 Conclusions and future work

In this paper, we introduced a real-time footsteps synthesizer controlled by the user, which works independently from the footwear. This is a feature that distinguishes our prototype from other systems developed with similar goals.

The system was tested in a between-subjects experiment, where it was compared to a recognition task including recorded and synthesized offline sounds. Results show that subjects are able to recognize most of the synthesized surfaces using the interactive system with high accuracy. Similar accuracy can be noticed in the recognition of real recorded footsteps sounds, which is an indication of the success of the proposed algorithms and their control.

The developed system is ready to be integrated in computer games and interactive installations where a user can navigate. The simulations proposed, however, reproduce the act of walking on a flat surface.

On the other hand, real life scenarios include also uphill and downhill movements whose footsteps sounds differ significantly from those produced while walking on a flat surface. Such situations can be incorporated in our synthesis engine, by modifying different parameters of the corresponding sounds such as amplitude and temporal variations.

In future work, we indeed plan to utilize the system in multimodal environments, and include haptic and visual feedback, to understand the role of the different sensorial modalities to enhance sense of immersion and presence in scenarios where walking plays an important role.

7 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. [6]

References

[1] J. Adrien. *The missing link: Modal synthesis*. MIT Press Cambridge, MA, USA, 1991.

[2] F. Avanzini and D. Rocchesso. Modeling collision sounds: Non-linear contact force. In *Proc. COST-G6 Conf. Digital Audio Effects (DAFx-01)*, pages 61–66, 2001.

[3] A. Benbasat, S. Morris, and J. Paradiso. A wireless modular sensor architecture and its application in on-shoe gait analysis. In *Sensors, 2003. Proceedings of IEEE*, volume 2, 2003.

[4] P. Chueng and P. Marsden. Designing Auditory Spaces to Support Sense of Place: The Role of Expectation. In *CSCW Workshop: The Role of Place in Shaping Virtual Community*. Citeseer, 2002.

[5] P. Cook. Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. *Computer Music Journal*, 21(3):38–49, 1997.

[6] 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.

[7] P. Cook. *Real sound synthesis for interactive applications*. AK Peters, Ltd., 2002.

[8] P. Dupont, V. Hayward, B. Armstrong, and F. Altpeter. Single state elastoplastic friction models. *IEEE Transactions on Automatic Control*, 47(5):787–792, 2002.

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

[10] 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.

[11] J. Freeman and J. Lessiter. Hear there & everywhere: the effects of multi-channel audio on presence. *Proceedings of ICAD 2001*, pages 231–234, 2001.

[12] W. Gaver. What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology*, 5(1):1–29, 1993.

[13] B. Giordano and S. Mcadams. Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates. *The Journal of the Acoustical Society of America*, 119:1171, 2006.

[14] B. Giordano, S. Mcadams, Y. Visell, J. Cooperstock, H. Yao, and V. Hayward. Non-visual identification of walking grounds. *Journal of the Acoustical Society of America*, 123(5):3412, 2008.

[6] www.niwproject.eu

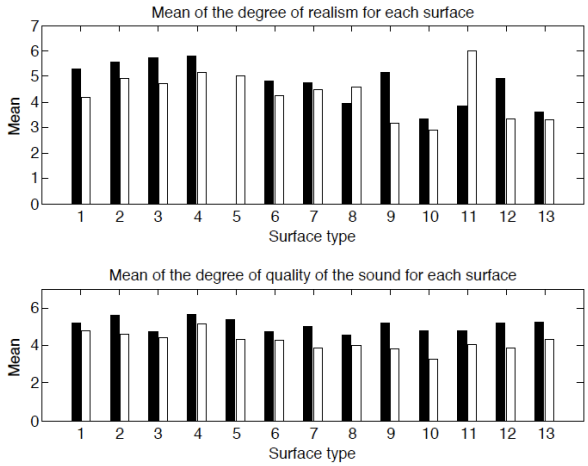


Figure 4: Comparison of the mean of the degree of realism (top) and quality of the sound (bottom) for each surface in experiment 1 (black) and 2 (white). Surface type from left to right: 1-beach sand, 2-gravel, 3-dirt pebbles, 4-snow, 5-high grass, 6-forest underbrush, 7-dry leaves, 8-wood, 9-creaking wood, 10-metal, 11-wood plus reverberation, 12-creaking wood plus reverberation, 13-metal plus reverberation.

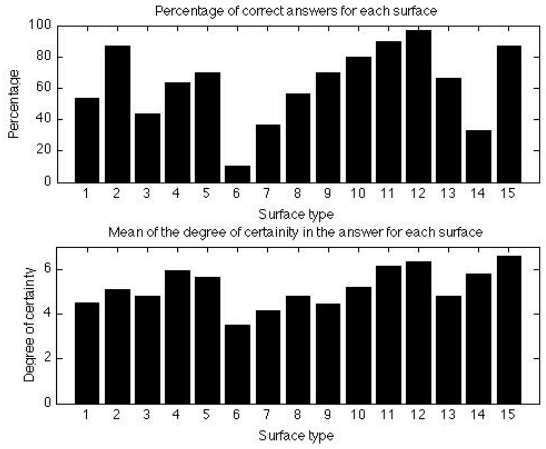


Figure 5: Percentages of correct answers (top) and mean of the degree of certainity in the answer for each surface in experiment 3. Surface type from left to right: 1-beach sand, 2-gravel, 3-dirt pebbles, 4-snow, 5-frozen snow, 6-high grass, 7-forest underbrush, 8-dry leaves, 9-concrete, 10-wood, 11-creaking wood, 12-metal, 13-carpet, 14-puddles and 15-water.

[15] B. Kapralos, D. Zikovitz, M. Jenkin, and L. Harris. Auditory cues in the perception of self-motion. In *116th AES convention*. Citeseer, 2004.

[16] R. Klatzky, D. Pai, and E. Krotkov. Perception of material from contact sounds. *Presence: Teleoperators & Virtual Environments*, 9(4):399–410, 2000.

[17] P. Larsson, D. Västfjäll, and M. Kleiner. Perception of self-motion and presence in auditory virtual environments. In *Proceedings of seventh annual workshop presence*, pages 252–258, 2004.

[18] A. Law, B. Peck, Y. Visell, P. Kry, and J. Cooperstock. A Multimodal Floor-space for Experiencing Material Deformation Underfoot in Virtual Reality. In *IEEE International Workshop on Haptic Audio visual Environments and Games, 2008. HAVE 2008*, pages 126–131, 2008.

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

[20] R. Nordahl. Increasing the motion of users in photorealistic virtual environments by utilizing auditory rendering of the environment and ego-motion. *Proceedings of Presence*, pages 57–62, 2006.

[21] J. Paradiso, K. Hsiao, and E. Hu. Interactive music for instrumented dancing shoes. In *Proc. of the 1999 International Computer Music Conference*, pages 453–456, 1999.

[22] R. Pastore, J. Flint, J. Gaston, and M. Solomon. Auditory event perception: The source-perception loop for posture in human gait. *Perception and Psychophysics*, 70(1):13, 2008.

[23] N. Raghuvanshi and M. Lin. Interactive sound synthesis for large scale environments. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games*, pages 101–108. ACM New York, NY, USA, 2006.

[24] R. Sanders Jr. *The effect of sound delivery methods on a users sense of presence in a virtual environment*. PhD thesis, NAVAL POSTGRADUATE SCHOOL, 2002.

[25] R. Storms and M. Zyda. Interactions in perceived quality of auditory-visual displays. *Presence: Teleoperators & Virtual Environments*, 9(6):557–580, 2000.

[26] A. Väljamäe, P. Larsson, D. Västfjäll, and M. Kleiner. Travelling without moving: Auditory scene cues for translational self-motion. In *Proceedings of ICAD05*, 2005.

[27] K. Van Den Doel, P. Kry, and D. Pai. FoleyAutomatic: physically-based sound effects for interactive simulation and animation. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 537–544. ACM New York, NY, USA, 2001.

[28] D. Västfjäll. The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments. *CyberPsychology & Behavior*, 6(2):181–188, 2003.

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.

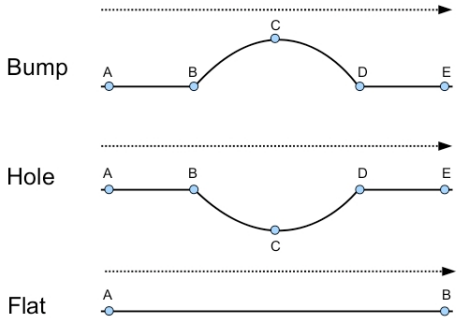


Figure 1: *The three types of surfaces modeled.*

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).

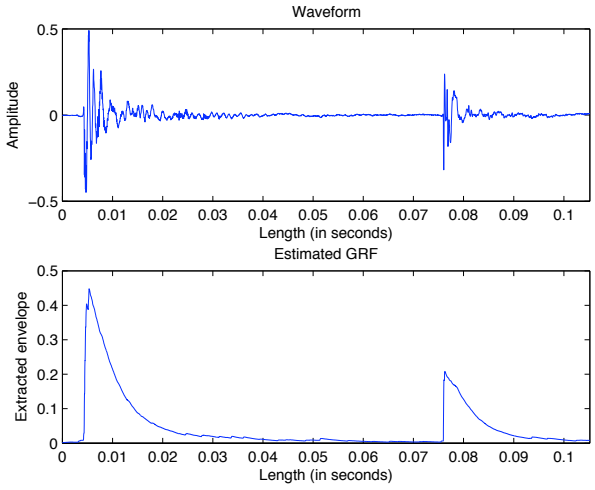


Figure 2: *Waveform of the used footstep on wood (top) and its relative extracted GRF (bottom).*

[2] www.hollywoodedge.com/

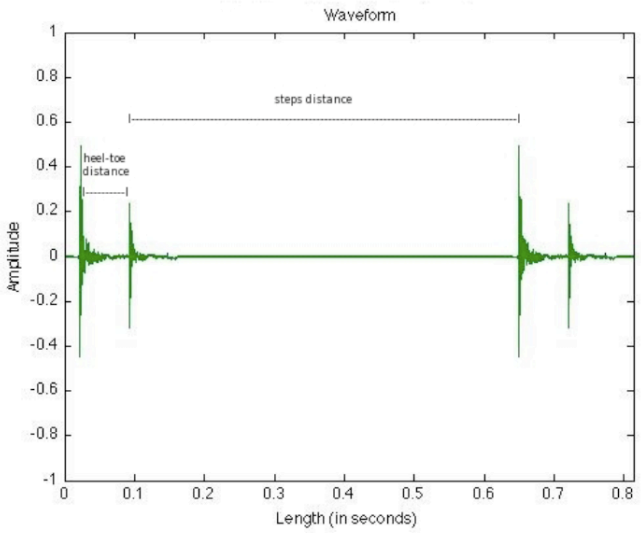


Figure 3: *Temporal distances between (named "steps distance" in the Figure), and within (named "heel-toe distance" in the Figure) footsteps.*

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 $\longrightarrow$ 1150	-	69 (fixed)
bump_2_step	16	19 = 4+6+5+4	100	550 $\longrightarrow$ 1150	-	69 (fixed)
hole_1_step	18	31 = 4+12+11+4	-50	750 $\longrightarrow$ 150	-	69 (fixed)
hole_2_step	11	19 = 4+6+5+4	-100	750 $\longrightarrow$ 150	-	69 (fixed)
bump_1_h_t	24	31 = 4+12+11+4	-	550 (fixed)	30	0 $\longrightarrow$ 360 (+ 69)
bump_2_h_t	14	19 = 4+6+5+4	-	550 (fixed)	60	0 $\longrightarrow$ 360 (+ 69)
hole_1_h_t	24	31 = 4+12+11+4	-	550 (fixed)	-20	240 $\longrightarrow$ 0 (+ 69)
hole_2_h_t	13	19 = 4+6+5+4	-	550 (fixed)	-40	240 $\longrightarrow$ 0 (+ 69)
bump_1_comb	32	31 = 4+12+11+4	50	550 $\longrightarrow$ 1150	30	0 $\longrightarrow$ 360 (+ 69)
bump_2_comb	19	19 = 4+6+5+4	100	550 $\longrightarrow$ 1150	60	0 $\longrightarrow$ 360 (+ 69)
hole_1_comb	22	31 = 4+12+11+4	-50	750 $\longrightarrow$ 150	-20	240 $\longrightarrow$ 0 (+ 69)
hole_2_comb	15	19 = 4+6+5+4	-100	750 $\longrightarrow$ 150	-40	240 $\longrightarrow$ 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

Identification of virtual grounds using virtual reality haptic shoes and sound synthesis

Stefania Serafin, Luca Turchet, Rolf Nordahl and Smilen Dimitrov
Medialogy, Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, Denmark
sts,tur,rn,sd@media.aau.dk

Amir Berrezag and Vincent Hayward
UPMC Univ Paris 06, Institut des Systèmes Intelligents et de Robotique,
4 place Jussieu,75005 Paris, France
amir.berrezag,vincent.hayward@isir.upmc.fr

Abstract

We describe a system which simulates in real-time the auditory and haptic sensation of walking on different surfaces. The system is based on physical models, that drive both the haptic and audio synthesizers, and a pair of shoes enhanced with sensors and actuators. In a discrimination experiment, subjects were asked to recognize the different simulated surfaces they were exposed to with uni-modal (auditory or haptic) and bi-modal (auditory and haptic) cues. Results show that subjects perform the recognition task better when using auditory feedback versus haptic feedback. The combination of auditory and haptic feedback only in some conditions significantly enhances recognition.

Keywords: walking simulation, physical models, audio-haptic interaction

1 INTRODUCTION

Research on multimodality has been focused on the interaction between vision and other modalities. However, lately the interest in investigating the interaction between other senses, such as touch and audition, has grown. This has been facilitated by the rapid development of haptic devices, together with efficient and accurate simulation algorithms. Several studies have indeed investigated audio-tactile phenomena such as the multimodal recognition of textures [10] and stiffness, both using physical stimuli [5] and simulations based on physical models [1].

The cited studies have focused on the interaction between touch and audition in hand-based interactions. To our knowledge, the interaction of auditory and haptic feedback in foot-based devices is still an unexplored topic. A notable exception is the work of Giordano et al., who showed that the feet were also effective at probing the world with discriminative touch, with and without access to auditory information. Their results suggested that integration of foot-haptic and auditory information does follow simple integration rules [8].

In previous research, we described a system able to recreate the auditory and haptic sensation of walking on different materials and presented the results of a preliminary surface recognition experiment [11]. This experiment was conducted under three different conditions: auditory feedback, haptic feedback, and both.

By presenting the stimuli to the participants passively sitting in a chair, we introduced a high degree of control on the stimulation. However, this method of delivery is highly contrived since it eliminates the tight sensorimotor coupling that is natural during walking and foot interaction. It is true for the auditory channel, but even more so for the haptic channel. In spite of these drastically constrained conditions, performance was surprisingly good.

In this paper, we extend such work by allowing subjects to walk in a controlled laboratory, where their steps are tracked and used to drive the simulation. We investigate whether introducing a higher level of

interactivity will significantly enhance the recognition rates as well as the perceived quality and realism of the simulation.

2 SIMULATION SOFTWARE

We developed a physically based synthesis engine able to simulate the auditory and haptic sensation of walking on different surfaces. Acoustic and vibrational signatures of locomotion are the result of more elementary physical interactions, including impacts, friction, or fracture events, between objects with certain material properties (hardness, density, etc.) and shapes. The decomposition of complex everyday sound phenomena in terms of more elementary ones has been an organizing idea in auditory display research during recent decades [7]. In our simulations, we draw a primary distinction between solid and aggregate ground surfaces, the latter being assumed to possess a granular structure, such as that of gravel.

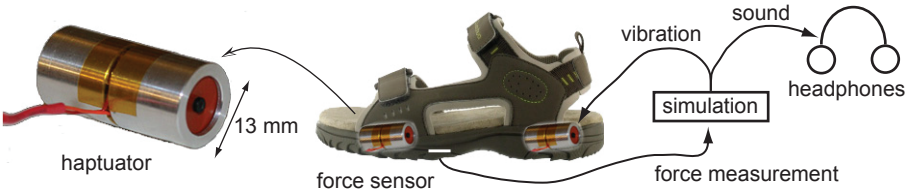


Figure 1: System (one shoe shown). Left: recoil-type actuation from Tactile Labs Inc. The moving parts are protected by an alumimum enclosure able to bear the weight of a person. Middle: approximate location of the actuators in the sandal. Right: system diagram showing the interconnections.

Figure 2: A picture of one pressure sensor and two actuators embedded in the shoes.

The impact model and its discretization are described elsewhere in detail [3]. The model has been recently adapted to the audio simulation of footsteps [12]. Here, we used the same model to drive the haptic and the audio synthesis. It is briefly recalled below.

A footstep sound may be considered to cause multiple micro-impacts between a sole, i.e., an *exciter*, and a floor, i.e., a *resonator*. Such interaction can be either discrete, as in the case of walking on a solid surface, or continuous, as in the case of a foot sliding across the floor.

In the simulation of discrete impacts, the excitation is brief and has an unbiased frequency response. The interaction is modelled by a Hunt-Crossley-type interaction where the force, f, between two bodies, combines hardening elasticity and a dissipation term [9]. Let x represent contact interpenetration and $\alpha > 1$ be a coefficient used to shape the nonlinear hardening, the special model form we used is

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

The model described was discretized as proposed in [2].

If the interaction called for slip, we adopted a model where the relationship between relative velocity v of the bodies in contact and friction force f is governed by a differential equation rather than a static map [6]. Considering that friction results from a large number of microscopic damped elastic bonds with an average deflection z, a viscous term, $\sigma_2 v$, and a noise term, $\sigma_3 w$, to represent roughness, we have

$$f(z, \dot{z}, v, w) = \sigma_0 z + \sigma_1 \dot{z} + \sigma_2 v + \sigma_3 w.$$

The force specified by these models is applied to a virtual mass which produces a displacement signal that is then processed by a linear shaping filter intended to represent the resonator.

Stochastic parameterization is employed to simulate particle interactions thereby avoiding to model each of many particles explicitely. 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.

We used this approach to model both solid and aggregate surfaces. A solid surface is represented by an impact and a slide. The impact model alone was used to recreate the sound and the feel produced when walking on wood. The friction model was tuned to simulate walking on creaking wood. To simulate walking on aggregate grounds, we used a physically informed sonic models (PhiSM) algorithm [4]. The synthesis was tuned to simulate snow and forest underbrush.

These algorithms were implemented as an extension to the Max/MSP platform[1] to drive both the auditory and haptic feedback.

3 Simulation Hardware

In order to provide both audio and haptic feedback, haptic shoes enhanced with pressure sensors have been developed. A pair of light-weight sandals was procured (Model Arpenaz-50, Decathlon, Villeneuve d'Ascq, France). This particular model has light, stiff foam soles that are easy to gouge and fashion. Four cavities were made in the tickness of 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. As indicated in Fig. 1 and Fig. 2, two actuators were placed under the heel of the wearer and the other two under the ball of the foot. There were bonded in place to ensure good transmission of the vibrations inside the soles. When activated, vibrations propagated far in the light, stiff foam. In the present configuration, the four actuators were driven by the same signal but could be activated separately to emphasize, for instance, the front or back activation, to strick a balance, or to realize other effects such as modulating different, back-front signals during heel-toe movements.

The sole has two FSR pressure sensors[2] whose aim was to detect the pressure force of the feet during the locomotion of a subject wearing the shoes. The two sensors were placed in correspondence to the heel and toe respectively in each shoe. The analogue values of each of these sensors were digitized by means of an Arduino Diecimila board[3] and were used to drive the audio and haptic synthesis.

A cable came out from each shoe, with the function of transporting the signals of the pressure sensors and of the actuators. Such cables were about 5 meters long, and they were connected through DB9 connectors to two 4TP (twisted pair) cables: one 4TP cable carried the sensor signals to a breakout board-which contained trimmers, that formed voltage dividers with the FSRs, which then interfaced to an Arduino board. The other 4TP cable carried the actuator signals from a pair of Pyle Pro PCA1[4] mini 2X15 W stereo amplifiers, driven by outputs from a FireFace 800 soundcard.[5] Each stereo amplifier handled 4 actuators found on a single shoe, and each output channel of the amplifier drove two actuators connected in parallel. The PC handled the Arduino through a USB connection, and the FireFace soundcard through a FireWire connection.

[1] www.cycling74.com
[2] I.E.E. SS-U-N-S-00039
[3] http://arduino.cc/
[4] http://www.pyleaudio.com/manuals/PCA1.pdf
[5] http://www.rme-audio.com/english/firewire/ff800.htm

4 CONTROLLING THE SYNTHESIS ENGINE

The audio-haptic synthesis engine is controlled by a ground reaction force (GRF), i.e., a force exerted by the ground on a body in contact with it. We first created a database of GRFs, by extracting the amplitude envelopes from different audio files of recorded footsteps, chosen among those available on the Hollywood Edge sound effects library.[6] Figure 3 shows the waveform of one of the footstep sounds used and its corresponding extracted GRF.

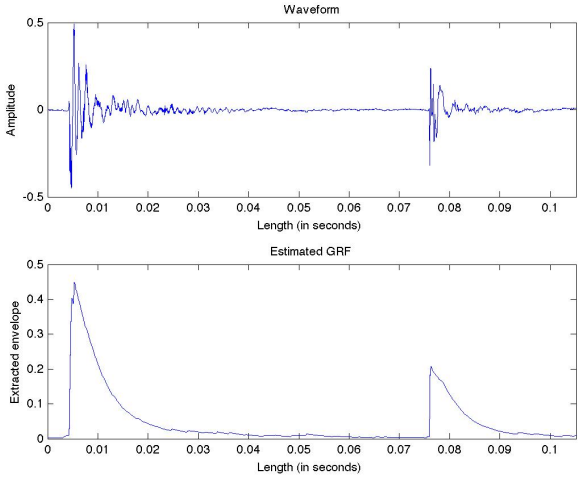


Figure 3: Time domain waveform of a footsteps sound (top) and its corresponding extracted GRF (bottom).

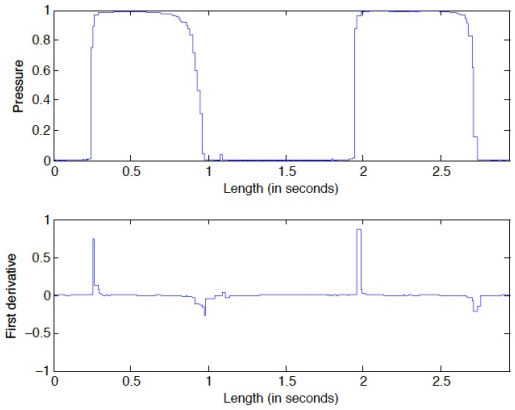


Figure 4: Waveform of a characteristic signal captured by the pressure sensors (top), and its derivative (bottom).

We then used the data captured by the pressure sensors and calculated their first time derivatives, which are related to the intensity with which the foot hits the ground. An example of data captured from the pressure sensors and its corresponding derivative is shown in Figure 4. The values of the first derivative were used as control for triggering the footsteps synthesizer, some GRFs corresponding to heel or toe

[6] www.hollywoodedge.com/

according to the activated sensor. Five types of heel and toes audio files were used for the purpose, and randomly chosen at the moment of triggering, giving rise to 25 possible combinations.

Each time the value of the first derivative became bigger than a threshold, the GRF corresponding to the activated sensor was triggered into the engine. More precisely we checked only positive changes in the derivative value, since we were interested in the generation of the sound when the step hits – and not when it leaves – the ground. Other thresholds, both on the signals and on their first derivatives, were used in order to handle some boundary conditions, like the standing of the subject, with the aim of controlling the generation of sound. Such thresholds were set in a phase of calibration of the system, which had to take into account the different weights of the subjects wearing the shoes, in order to have an average value suitable for all the possible cases.

5 Evaluating the system

We conducted a between-subjects experiment whose goal was to investigate the ability of subjects to recognize the different simulated auditory and haptic stimuli they were exposed to. The experiment consisted of three conditions: in the first condition subjects were exposed to auditory feedback only, in the second condition to haptic feedback only, and in the third condition to a combination of audio and haptic feedback.

The sounds provided in the first condition were synthesized sounds generated in real time while subjects were walking using the interactive system described in the previous section.
The same stimuli were provided in the second condition by using the haptic shoes which provided haptic feedback in real-time. In the third condition, a combination of auditory and haptic stimuli was provided.

5.1 Setup

All experiments were carried out in an acoustically isolated laboratory. The walking area was approximately 18 square meters, delimited by the walls of the laboratory. The setup for the first and third condition of the experiment consisted of the pair of sandals described in section 2, an Arduino board, a Fireface soundcard, a laptop and a set of headphones (Sennheiser HD 650, http://www.sennheiser.com).

In the first condition with audio only, the haptic actuators were not used, and the pressure sensors were driving the synthesis engine only to provide auditory feedback.
In the second condition (haptics only), participants were asked to wear earplugs and sound protection headsets instead of headphones, in order to block any audible feedback produced by the actuators. Indeed such sounds were not in the same quality range of the sounds designed to be conveyed through the headphones, and they could have biased the judgments of the participants.

In order to facilitate the navigation of the subjects, the wires coming out from the shoes in all setups, as well as the wires connecting the headphones to the soundcard, were linked to a bumbag or to snaplinks attached to trousers (see Fig. 5).

5.2 Participants

Thirty participants were divided in three groups (n = 10) to perform the between-subjects experiment. The three groups were composed respectively of 7 men and 3 women, aged between 20 and 35 (average age=24.6, sd=4.67), 9 men and 1 woman, aged between 20 and 31 (average age=23.4, sd=3.23), and 7 men and 3 women, aged between 21 and 25 (average age=22.7, sd=1.07). 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 in average about 10, 13 and 11 minutes for condition 1, 2 and 3 respectively.

5.3 Task

During condition 1 and 3 the participants were asked to wear the pair of sandals and the headphones described in sections 2 and 5.1, and to walk in the laboratory. During condition 2 they were asked to wear the pair of sandals, earplugs and sound protection headsets, and to walk in the laboratory.

Participants were exposed to 16 trials in each condition. During the experiment, 8 stimuli were presented twice in randomized order. The stimuli consisted of audio and haptic simulations of footsteps sounds

Figure 5: A subject performing the experiment in the audio-haptic condition.

on the following surfaces: beach sand, gravel, snow (in particular deep snow), forest underbrush (a forest floor composed by dirt, leaves and branches breaking), dry leaves, wood, creaking wood and metal.

Participants were given a list of different surfaces to be held in one hand, presented as non-forced alternate choice. Such list included a range of materials wider than those presented in experiments.

During the act of walking they listened simultaneously to footsteps sounds and/or vibrations on a different surface according to the stimulus presented. The task consisted of answering by voice the following three questions after the presentation of the stimulus:

1. Which surface do you think you are walking on? For each stimulus choose an answer in the following list: 1) beach sand, 2) gravel, 3) dirt plus pebbles, 4) snow, 5) high grass, 6) forest underbrush, 7) dry leaves, 8) wood, 9) creaking wood, 10) metal, 11) carpet, 12) concrete, 13) frozen snow, 14) puddles, 5) water, 16) I don't know.

2. How close to real life is the sound in comparison with the surface you think it is? Evaluate the degree of realism on a scale from 1 to 7 (1=low realism, 7=high realism).

3. Evaluate the quality of the sound on a scale from 1 to 7 (1=low quality, 7=high quality).

The participants were informed that they could choose the same material more than one time and that they were not forced to choose all the materials in the list. In addition, they could use the interactive system as much as they wanted before giving an answer. When passed to the next stimulus they could not change the answer to the previous stimuli.

6 Results and discussions

The collected answers were analyzed and compared between the three conditions. Results are summarized in the confusion matrices reported in Table 1,2 and 3. Results confirm that auditory modality is dominant on the haptic modality and that the haptic task was more difficult than the other two.

Table 1: Confusion matrices with audio condition

	BS	GL	SW	UB	DL	WD	CW	MT	FS	CC	HG	DR	PD	WT	CP	—
BS	6	4	3					3			3				1	
GL		8		5	2						1	1	1	1		1
SW	1		15						3		1					
UB	1	4		6		1			6							2
DL	1	5		5	5				1		1					2
WD						10	2	1	1	2						4
CW							19									1
MT								13					1			6

Legend:	WD	wood	CW	creaking wood	SW	snow	UB	underbrush
	—	don't know	FS	Frozen snow	BS	beach sand	GL	Gravel
	MT	metal	HG	High grass	DL	dry leaves	CC	concrete
	DR	dirt	PD	puddles	WT	Water	CP	carpet

Table 2: Confusion matrices with haptic condition

	BS	GL	SW	UB	DL	WD	CW	MT	FS	CC	HG	DR	PD	WT	CP	—
BS	3	2	5	2			1		1			3				3
GL	5	1	3	1	1		1				2	2	1		1	2
SW	4		7		1		2		5							1
UB	1	3	5			2	1		1			3				4
DL	1	2	1		3				1			2			4	6
WD		2	2			6	3	1		1						5
CW			2			1	7		3	1	1			2		3
MT					1		5	7				2			1	4

Legend:	WD	wood	CW	creaking wood	SW	snow	UB	underbrush
	—	don't know	FS	Frozen snow	BS	beach sand	GL	Gravel
	MT	metal	HG	High grass	DL	dry leaves	CC	concrete
	DR	dirt	PD	puddles	WT	Water	CP	carpet

Table 3: Confusion matrices with audio-haptic condition

	BS	GL	SW	UB	DL	WD	CW	MT	FS	CC	HG	DR	PD	WT	CP	—
BS	8	4	4		1				2		1					
GL		18		2												
SW			13						6							1
UB		1	2	8	1				6			2				
DL	2	5		3	5		1		1		1	2				
WD						18				1						1
CW							20									
MT								14								6

Legend:

WD	wood	CW	creaking wood	SW	snow	UB	underbrush
—	don't know	FS	Frozen snow	BS	beach sand	GL	Gravel
MT	metal	HG	High grass	DL	dry leaves	CC	concrete
DR	dirt	PD	puddles	WT	Water	CP	carpet

Table 4: Average realism scores from a seven-point Likert scale.

	BS	GL	SW	UB	DL	WD	CW	MT
audio	5	5.1875	5.5	3.6667	5.1667	4.2	4.4211	3.1538
haptics	5	5	6	-	3.6667	4	4.1429	4.8571
combined	4.125	5.2778	5.4615	5.25	3.6	2.8333	3.95	3.0714

Table 5: Average quality scores from a seven-point Likert scale.

	BS	GL	SW	UB	DL	WD	CW	MT
audio	5.05	4.9722	5.275	4.7222	5.2222	4.5625	4.4211	4.5
haptics	3.8824	4.2222	4.7895	3.75	3.8571	3.8667	3.6875	3.9375
combined	4.5	5.15	5.6316	4.7	4.85	3.3684	3.7	4.0714

The addition of the haptic to the audio modality seemed to help the recognition only in few cases. Indeed percentages are on average slightly higher in condition 3 rather than condition 1, although an in-depth analysis shows significant difference only for gravel (χ^2 = 8.9011, df = 1, p-value = 0.00285) and wood (χ^2 = 5.8333, df = 1, p-value = 0.01573).

In the audio-haptic condition subjects recognized quite well most of the modeled surfaces. Very high percentages were found for gravel, wood, creaking wood.

An analysis performed on the wrong answers of conditions 1 and 3 reveals that on average subjects tended to classify erroneously a surface as another belonging to a same category (e.g., wood-creaking wood-concrete, snow-frozen snow, dry leaves-forest underbrush-dirt) rather than to different categories (e.g., wood-water, wood-gravel, metal-dry leaves). In particular considering the category snow-frozen snow, snow was recognized with a percentage of 90% in the audio condition and of 95% in the audio-haptic condition.

Moreover, what emerges from these results is the ability of the subjects in distinguishing materials in the same category for solid surfaces, and their difficulties in the recognition of aggregate surfaces (aspect also confirmed by the comments of the participants). The same tendencies were found in another experiment we conducted using the same physical models controlled by another system [12] as well as in a previous experiment with audio-haptic stimuli, where subjects were receiving the stimuli passively [11].

Subjects were able to distinguish at haptic level solid surfaces with a percentage of 53%, and were able to distinguish aggregate surfaces with a percentage of 74%.

As regards the percentage of the "I don't know" answers, it was higher for condition 2 (17.5%), rather condition 1 (10 %) and lower in condition 3 (5%) and this is an indication of the difficulty of the task proposed. Condition 2 took longer (13 minutes) rather than condition 1 (9 minutes and 30 seconds) and 3 (11 minutes). This is also an indication of the difficulty of the task proposed.

Table 4 and 5 report results on the perceived quality and realism of the simulation. The degree of realism was calculated by looking only at that data from correct answers, i.e., when the surfaces were correctly recognized. As far as the quality judgement is concerned, the data was based on all the answers different from "I don't know". The results are similar to those presented in [11], even if in the previous studies a lower number of actuators was present (two for each shoe instead of four) and the experiment was performed in passive conditions. Moreover, results show that the combination of auditory and haptic stimuli does not always enhance realism and perceived quality of the simulation.

7 Conclusion

In this paper, we introduced a real-time footsteps synthesizer able to provide audio and haptic feedback, and which is controlled by the user during the act of walking by means of shoes embedded with sensors and actuators. The system was tested in a between-subjects experiment.

Results show that subjects are able to recognize most of the synthesized surfaces using the interactive system with high accuracy. Similar accuracy can be noticed in the recognition of real recorded footsteps sounds, which is an indication of the success of the proposed algorithms and their control.

The developed system is ready to be integrated in computer games and interactive installations where a user can navigate.

In future work, we indeed plan to utilize the system in multimodal environments, and include visual feedback, to understand the role of the different sensorial modalities to enhance sense of immersion and presence in scenarios where walking plays an important role.

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.[7]

[7] www.niwproject.eu

REFERENCES

[1] Avanzini, F. and Crosato, P. (2006). Haptic-auditory rendering and perception of contact stiffness. *Lecture Notes in Computer Science*, 4129:24.

[2] Avanzini, F. and Rocchesso, D. (2001). Modeling collision sounds: Non-linear contact force. In *Proc. COST-G6 Conf. Digital Audio Effects (DAFx-01)*, pages 61–66.

[3] Avanzini, F., Serafin, S., and Rocchesso, D. (2005). Interactive simulation of rigid body interaction with friction-induced sound generation. *IEEE transactions on speech and audio processing*, 13(5 Part 2):1073–1081.

[4] Cook, P. (1997). Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. *Computer Music Journal*, 21(3):38–49.

[5] DiFranco, D., Beauregard, G. L., and Srinivasan, M. (1997). The effect of auditory cues on the haptic perception of stiffness in virtual environments. *Proceedings of the ASME Dynamic Systems and Control Division.*

[6] Dupont, P., Hayward, V., Armstrong, B., and Altpeter, F. (2002). Single state elastoplastic friction models. *IEEE Transactions on Automatic Control*, 47(5):787–792.

[7] Gaver, W. (1993). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology*, 5(1):1–29.

[8] Giordano, B. L., Mcadams, S., Visell, Y., Cooperstock, J., Yao, H. Y., and Hayward, V. (2008). Non-visual identification of walking grounds. *Journal of the Acoustical Society of America*, 123(5):3412–3412.

[9] Hunt, K. H. and Crossley, F. R. E. (1975). Coefficient of restitution interpreted as damping in vibroimpact. *ASME Journal of Applied Mechanics*, 42(2):440–445.

[10] Lederman, S. (1979). Auditory texture perception. *Perception*.

[11] Nordahl, R., Berrezag, A., Dimitrov, S., Turchet, L., Hayward, V., and Serafin, S. (2010a). Preliminary experiment combining virtual reality haptic shoes and audio synthesis. In *Proc. Eurohaptics*.

[12] Nordahl, R., Serafin, S., and Turchet, L. (2010b). Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In *Proc. IEEE VR 2010*.

Influence of auditory and visual feedback for perceiving walking over bumps and holes in desktop VR

Luca Turchet* AAU Copenhagen, Denmark

Maud Marchal† INSA/INRIA Rennes, France

Anatole Lécuyer‡ INRIA Rennes, France

Stefania Serafin§ AAU Copenhagen, Denmark

Rolf Nordahl¶ AAU Copenhagen, Denmark

Abstract

In this paper, we present an experiment whose goal is to investigate the role of sound and vision in the recognition of different surface profiles in a walking scenario. Fifteen subjects participated to two within-subjects experiments where they were asked to interact with a desktop system simulating bumps, holes and flat surfaces by means of audio, visual and audio-visual cues. Results of the first experiment show that participants are able to successfully identify the surface profiles provided through the proposed audio-visual techniques. Results of a second experiment in which conflictual audio-visual stimuli were presented, reveal that for some of the proposed visual effects the visual feedback is dominant on the auditory one, while for the others the role of dominance is inverted.

CR Categories: I.3.5 [Computer Graphics]: Physics Based Modeling—; I.6.5 [Simulation and Modeling]: Model Development— [H.5.5]: Information Interfaces and Presentation (e.g. HCI)—Sound and Music Computing.

Keywords: Bump and hole, pseudo-haptic, design, audio-visual.

1 Introduction

In most current virtual reality setups, users are restricted to walk on a flat surface. Only few locomotion interfaces are able to render uneven grounds, but they have the disadvantages of being costly and cumbersome [Hollerbach et al. 2003; Iwata et al. 2001]. Recently, research has shown that it is possible to simulate the act of walking on unflat surfaces only using visual cues [Marchal et al. 2010]. These results are a development of previous research on pseudo-haptic simulation [Lécuyer et al. 2004]. The main idea of the research described in [Lécuyer et al. 2004] was to investigate whether it was possible to simulate a bump or a hole by only using visual feedback.

In the sound synthesis community, all previous research on walking sounds has focused on flat surfaces [Cook 2002; Fontana and Bresin 2003; Farnell 2007; Miner and Caudell 2005]. Recently, a sound synthesis engine able to simulate footsteps sounds on aggregate and solid surfaces has been presented [Nordahl et al. 2010; Turchet et al. 2010]. Such sound engine is based on physical models synthesis and it works both offline and in realtime. In all the different surfaces simulated, the system energy parameter of the engine is controlled by a ground reaction force (GRF), i.e., the reaction force supplied by the ground at every step. During the offline working of the engine, such force is estimated from recordings on real footsteps sounds.

In [Serafin et al. 2010], researchers used such engine in order to run a preliminary experiment whose goal was to assess the role of temporal aspects in sonically simulating the act of walking on a bump or a hole. In particular, they investigated whether the timing between heel and toe and the timing between footsteps affected perception of walking on unflat surfaces. Results showed that it is possible to simulate a bump or a hole only by using temporal information.

In this paper, we are interested in understanding whether the addition of sounds to the visual feedback enhances the pseudo-haptic feedback for the simulation of unflat surfaces. Our hypothesis is that sound plays an important role in the recognition of the surface profile and that it can be dominant on the visual modality according to the visual stimulus presented.

*e-mail:tur@media.aau.dk

†e-mail:maud.marchal@irisa.fr

‡e-mail:anatole.lecuyer@irisa.fr

§e-mail:sts@media.aau.dk

¶e-mail:rn@media.aau.dk

2 Visual and auditory simulation

2.1 Visual feedback

The visual techniques used to simulate the act of walking over bumps and holes were the same proposed in [Marchal et al. 2010]: a straightforward modification of the camera's height (H), a modification of the camera's navigation velocity (V), a modification of the camera's orientation (O), and the combination of the three effects (HOV). Our simulations used a known mathematical profile, the Gaussian one, which was used for the simulations of both holes and bumps.

Since the velocity technique was used for the generation of the auditory feedback we briefly recall it here. The velocity effect is based on the variation of the camera velocity. Thus, the camera velocity is decreased when the user is going up and increased when the user is going down. We used a different algorithm for the ascending and descending cases. The algorithm computes the ratio R_{Velocity} applied between the real user velocity and the virtual camera velocity. The camera velocity is then modified following the equation:

$$\text{Velocity}^t = \text{Velocity}^{t-1} \,.\, R^t_{\text{Velocity}} \tag{1}$$

- Ascending case:

$$R^t_{\text{Velocity}} = exp(-R_{\text{AscendingV}} \,.\, \alpha^t) \tag{2}$$

where α is the tangent angle of the Gaussian curve and $R_{\text{AscendingV}}$ is a constant.

- Descending case:
This algorithm is designed to give a run up for a while after the bump or at the beginning of the hole. At time t, the ratio is updated regarding the difference between the user height in scene at times $t-1$ and t :

$$R^t_{\text{Velocity}} = R^{t-1}_{\text{Velocity}} + \Delta_{\text{Height}} \,.\, R_{\text{DescendingV}} \tag{3}$$

where the ratio $R_{\text{DescendingV}}$ is a constant. When the subject reaches the end of the descent, his speed is at a maximum.

If he is walking in a hole, then he starts to go up and his speed value will be given by the ascending algorithm. If the subject is on a bump, he will reach the plane ground after the bump. His speed ratio R_{Velocity} will start decreasing at 0.1 unit per second, until another bump/hole is reached or the ratio is back to normal.

2.2 Auditory feedback

As concerns the sounds, the goal was to add an auditory feedback consistent with the images provided at the screen, but since we were interested just in a proof of concept no interaction between the visual and sound engines was created.
The technique adopted to render bumps and holes at auditory level has been the placement of footsteps sounds at different temporal intervals, taking as reference the work of the visual engine set with the velocity effect. In order to create a coherent mapping between the space travelled by the visual rendering along the surface and the temporal distances between steps, we chose to divide the surface profile in equal parts covering the same amount of space, and to place in those points a step (see Figure 1). Subsequently the temporal intervals to reach each point located on the curve have been calculated according to the equations utilized for the velocity effect by the visual engine (see section 2.1). Precisely, such calculations have been performed using the same parameters defining the Gaussian curve (height and sigma), but modifying the values of the R factors regulating the lows for the ascending and descending case because of the necessity of synchronizing properly the two systems.

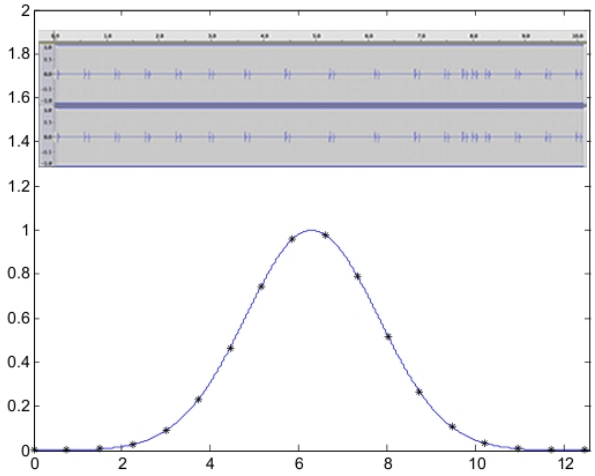


Figure 1: *Gaussian profile used for a bump simulation with the indication of the points in which the footstep sounds occur, and stereo waveform of the corresponding sound file.*

The input files for the sound synthesis engine (see section 1) resulting from such calculations, have been generated by means of MATLAB placing at different temporal patterns a single footstep sound (the same footstep was used to create the different stimuli presented to the subjects). The use of the same footstep sound 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 sound used was a recording of a real footstep on concrete. Such sound was chosen among those available in the Hollywood Edge sound effects library.[1]
For the purpose of this experiment, two types of surfaces, gravel and wood, were chosen. The reason for choosing two materials was to assess whether the surface type affected the quality of the results.

[1] www.hollywoodedge.com/

3 Experiment

We conducted two within-subjects experiments using three different kinds of stimuli: visual only, audio only, and audio visual.

1. Experiment 1: recognition of bumps, holes and flat surfaces by means of visual, audio and (coherent) audio-visual stimuli.
2. Experiment 2: recognition of bumps and holes in presence of incoherent audio-visual stimuli.

The images provided during the experiments were generated by the visual engine, while the sounds consisted of footsteps sounds generated by the offline use of the sound synthesis engine.
The goal of the first experiment was to investigate the ability of subjects to recognize the different walking surface profiles they were exposed to. One of our hypotheses was that the recognition would have improved using the bimodal stimuli rather than the single modalities alone.
The goal of the second experiment was to determine the role of dominance of the two modalities involved, in presence of incoherent stimuli. Our hypothesis was that the visual modality would have been dominant on the auditory modality for the H, O and HOV effects, while for the V effect the role of dominance would have been held by the auditory modality.

3.1 Participants

The experiments were performed by 15 participants, 12 men and 3 women, aged between 22 and 29 (mean=26.2,standard deviation=2.1662). All participants had normal or corrected vision and 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 40 minutes to complete the two experiments and to fill the questionnaire.

3.2 Task

During experiment 1 participants were given the list of three different surfaces (bump, hole, flat), presented as forced alternate choice, while during experiment 2 the possible choices were only two, i.e., bump or hole, in order to understand which modality was dominant. The task consisted of recognizing to which surface the walk corresponded after the presentation of the stimulus. Each stimulus was presented only one time before giving the answer, and when moving to the next stimulus participants could not change the answer to the previous stimuli.
After both experiments, a preference questionnaire was proposed in which participants had to grade from 1 (low appreciation) to 7 (high appreciation) three further effects (a bump provided at audio, visual and audio-visual level) according to 3 subjective criteria: efficiency of the simulation, realism of the simulation, and global appreciation of the effect. In the questionnaire we did not mentioned the words audio, visual or audio-visual, we simply called them condition 1, 2 and 3, in order to leave to the participants the task of creating the proper mapping.
At the end of the questionnaire, participants were also given the opportunity to leave an open comment on their experience interacting with the system.

3.3 Setup

All experiments were carried out in an acoustically isolated laboratory where the setup for the experiment was installed. It con-

sisted of a simple graphical user interface with which the participants were asked to interact. The interaction consisted of pressing keys on the keyboard in order to start each trial and to give the corresponding answer.
The virtual environment was a simple corridor with given dimensions (height=3.0m, length=12.6m, width=2.0m). There was a part in the center of the corridor where the height could be modified during the experiments: the user could experience either on a bump, a hole or a plane. To symbolize this variable part of the corridor, a transparent cube was represented on the ground with a height of 0.5m and a surface of 6.6mx2m, as illustrated in Figure 2. The variable height of the ground was not visible in order to exclude visual cues from the scene.

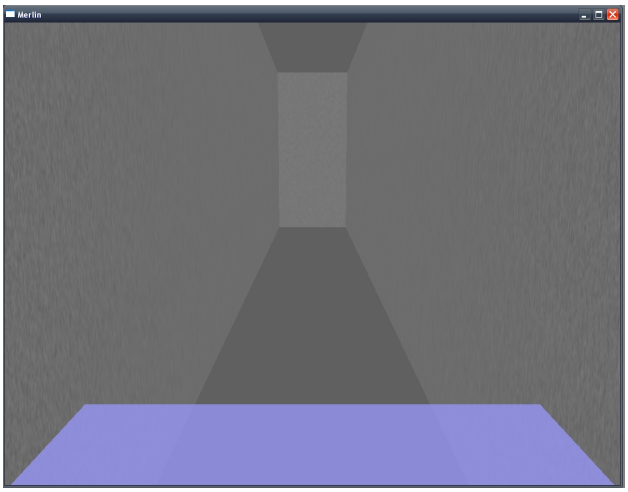


Figure 2: *A screenshot at the beginning of a trial.*

The computer utilized for the experiment had clock speed equal to 3.40 GHz, 2 GB of memory, it run the Windows XP operative system, and it was equipped with the SigmaTel Sound Card. Sounds were provided to the user by means of a set of headphones.[2]
During the experiments subjects were sitting on a chair, listening to the sounds through headphones and interacting with the interface.

3.4 Experimental plan

Participants were exposed to 105 trials in experiment 1, where 15 stimuli were repeated seven times. The stimuli consisted of bumps, holes and flat surfaces rendered by means of sounds (footsteps on two types of materials, gravel and wood), images (using the HOV technique), and the combinations of images and sounds.
During experiment 2 participants were exposed to 48 trials, where 8 audio-visual incoherent stimuli were repeated six times. In detail, only bumps ad holes were rendered: when a bump was presented at visual level a hole was provided at auditory level, and vice versa. The techniques used to render the bumps and holes at visual level were Height, Orientation, Velocity and HOV, while the sounds provided were the same presented in experiment 1 using wood as material.
The temporal durations of the trials were 10.5 seconds for bumps, 9 seconds for holes, and 9.5 seconds for the flat surfaces.

4 Results

For each participant, the percentage of correct answers was estimated for the different experimental conditions. Results are shown in tables 1, 2.
Concerning experiment 1, the first noticeable element emerging from results is the high percentages of correct answers for all the three modalities. The presentation of flat surfaces led to better performances rather than bumps and holes. The addition of audio stimuli did not produce higher percentages for the recognition of the surfaces rather than the visual only modality (which is already very high), and an in depth analysis performed by means of the Kruskal-Wallis test revealed that all the differences are not statistically significant.
Moreover no substantial differences have been found between stimuli involving footsteps sounds on wood and those involving footsteps sounds on gravel.

		Bump	Hole	Flat	%Correct answers
Visual	**Bump**	98	7	0	93.33
	Hole	7	98	0	93.33
	Flat	0	0	105	100
Audio 1	**Bump**	96	8	1	91.42
	Hole	8	97	0	92.38
	Flat	6	0	99	94.28
Audio 2	**Bump**	96	5	4	91.42
	Hole	4	98	3	93.33
	Flat	3	1	101	96.19
Audio-Visual 1	**Bump**	98	7	0	93.33
	Hole	8	97	0	92.38
	Flat	0	0	105	100
Audio-Visual 2	**Bump**	99	5	1	94.28
	Hole	8	97	0	92.38
	Flat	1	0	104	99.04

Table 1: *Confusion matrix illustrating the results for experiment 1. Legend: Visual: the HOV effect; Audio 1: sounds of footsteps on gravel; Audio 2: sounds of footsteps on wood; Audio-Visual 1: sounds of footsteps on gravel plus the HOV effect; Audio-Visual 2: sounds of footsteps on wood plus the HOV effect.*

As regards experiment 2, results show clearly that in presence of audio-visual conflicts audio is dominated by vision when H and O effects are presented. Conversely, vision is dominated by audio when V and HOV effects are presented. In particular the highest role of dominance has been found for audio stimuli respect visual stimuli provided by means of the Velocity effect.

	% Visual dominance	% Audio dominance
Bump H	85.55	14.45
Hole H	81.11	18.89
Bump O	72.22	27.78
Hole O	71.11	28.89
Bump V	8.88	91.12
Hole V	3.33	96.67
Bump HOV	30	70
Hole HOV	26.66	73.34

Table 2: *Table illustrating results for experiment 2. Legend: H: the Height effect; O: the Orientation effect; V: the Velocity effect; HOV: the combination of the three effects.*

Table 3 shows the results concerning the grades obtained by the three different modalities for each of the subjective criteria. Results clearly show better evaluations in the audio-visual modality (condition 3) rather than the audio (condition 1) and visual (condition 2) modalities alone. In particular an in depth analysis by means of t-test reveals that such differences are statistically significant for the efficiency criterium for condition 3 respect to condi-

[2]Sennheiser HD 600, http://www.sennheiser.com

tion 1 (p=0.01644), for the realism criterium for condition 3 respect condition 2 (p=0.02156), and for the appreciation criterium for condition 3 respect to condition 1 (p=0.02047) and 2 (p=0.006438). No significant differences in the criteria investigated have been found between condition 1 and 2.

	Audio		**Visual**		**Audio-Visual**	
	μ	σ	μ	σ	μ	σ
Efficiency	4.5333	1.7265	5.3333	1.3452	5.8667	0.9904
Realism	4.5333	1.5523	4.1333	1.5976	5.5333	1.5523
Appreciation	4.4667	1.7674	4.4667	1.3020	5.8667	1.3020

Table 3: *Table illustrating the answers of the questionnaire (average scores from a seven-point Likert scale, and relative standard deviation).*

5 General discussion

Results of experiment 1 show that the simulated surface profiles were recognized with high percentages in all the three modalities, and this is a confirm of the success of the techniques proposed in [Marchal et al. 2010] and [Serafin et al. 2010]. Although the addition of audio stimuli did not produce significant higher percentages rather than the visual only modality, the subjective questionnaire revealed, for the criteria investigated, a clear preference of the bimodal stimuli respect to the stimuli presented in the single modalities. In addition no significant differences were found between stimuli involving footsteps sounds on wood and those involving footsteps sounds on gravel and this is a further confirm of the results illustrated in [Serafin et al. 2010].
In case of conflictual audio-visual stimuli the relationships of dominance between modalities turned out to be very clear, according to the visual effect adopted. Our hypotheses were confirmed for all the effects with the exception of HOV for which, surprisingly, the auditory modality is dominant on the visual one. Presumably an explanation lies in the fact that the V effect was present in the HOV condition, and it is well known in the field of multimodal perception that the temporal resolution of the auditory system is better than the one of the visual system [Stein and Meredith 1993]. Our results confirm such behaviour of the perceptive system.
What emerges from an in depth analysis of results of experiment 2 is that participants on average were consistent in their choices: indeed in case of conflicts, they tended to choose always the same modality as guide (according to the visual effect presented).
Some of the subjects reported in their comments that in experiment 2 they noticed conflicts between what they heard and what they saw at screen. In particular one of them explicitly reported that in case of conflicts he was driven in his decision by the visual stimulus since not seeing the feet of the person walking in the animation, he imagined that the person would have been able to place the steps in an arbitrary way.

6 Conclusion and future work

In this paper, we described an experiment whose goal is to assess whether the combination of auditory and visual information affects the perception of walking on a hole or a bump using a desktop system. Results show that the surface profiles simulated by means of audio, visual and audio-visual stimuli were accurately recognized in all the three modalities, but the subjective questionnaire revealed a clear preference of the bimodal stimuli.
As concerns the auditory stimuli, they were created only varying temporal parameters of footsteps, such as the distance between steps. Other important aspects, such as amplitude and spectral information was not modified. These parameters are object of future investigations. 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 realism of the simulated experience. In future works we plan to integrate the visual system with the sound synthesis engine in order to have an audio-visual interactive simulation of the proposed surface profiles, both with a desktop system and an immersive VR configuration with HMD.

Acknowledgements

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. [3] This publication is supported by COST Action IC0601.

References

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

FARNELL, A. 2007. Marching onwards: procedural synthetic footsteps for video games and animation. *Proceedings of the Pure Data Convention*.

FONTANA, F., AND BRESIN, R. 2003. Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. *Proc. Colloquium on Musical Informatics*, 109–114.

HOLLERBACH, J., CHECCACCI, D., NOMA, H., YANAGIDA, Y., AND TETSUTANI, N. 2003. Simulating side slopes on locomotion interfaces using torso forces. In *Haptic Symposium*, Citeseer, 91–98.

IWATA, H., YANO, H., AND NAKAIZUMI, F. 2001. Gait master: A versatile locomotion interface for uneven virtual terrain. *vr*, 131.

LÉCUYER, A., BURKHARDT, J., AND ETIENNE, L. 2004. 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*, ACM, 246.

MARCHAL, M., LECUYER, A., CIRIO, G., BONNET, L., AND EMILY, M. 2010. Walking Up and Down in Immersive Virtual Worlds: Novel Interactive Techniques Based on Visual Feedback. In *Proceedings of IEEE Symposium on 3D User Interface*.

MINER, N., AND CAUDELL, T. 2005. Using wavelets to synthesize stochastic-based sounds for immersive virtual environments. *ACM Transactions on Applied Perception 2*, 4, 521–528.

NORDAHL, R., SERAFIN, S., AND TURCHET, L. 2010. Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In *Proceedings of IEEE Virtual Reality*.

SERAFIN, S., TURCHET, L., AND NORDAHL, R. 2010. Do you hear a bump or a hole? An experiment on temporal aspects in footsteps recognition. In *Proceedings of Digital Audio Effects Conference*.

STEIN, B. E., AND MEREDITH, M. A. 1993. *The Merging of the Senses*, 1st ed. MIT Press.

TURCHET, L., SERAFIN, S., DIMITROV, S., AND NORDAHL, R. 2010. Physically based sound synthesis and control of footsteps sounds. In *Proceedings of Digital Audio Effects Conference*.

[3] www.niwproject.eu

Proc. of MMSP'10 2010 IEEE International Worshop on Multimedia Signal Processing, Oct 4-6, 2010, Saint Malo, France, pp. 269-273

Audio-haptic physically-based simulation of walking on different grounds

Luca Turchet [#1], Rolf Nordahl [#4], Stefania Serafin [#2], Amir Berrezag [*6], Smilen Dimitrov [#3], Vincent Hayward [*5]

[#] *Aalborg University Copenhagen, Medialogy*
Lautrupvang 15, 2750 Ballerup, Denmark
[1] tur, [2] sts, [3] sd, [4] rn@media.aau.dk

[*] UPMC *Univ Paris 06, Institut des Systèmes Intelligents et de Robotique*
4 place Jussieu, 75005 Paris, France
[5] vincent.hayward, [6] amir.berrezag@isir.upmc.fr

***Abstract*—We describe a system which simulates in real-time the auditory and haptic sensations of walking on different surfaces. The system is based on a pair of sandals enhanced with pressure sensors and actuators. The pressure sensors detect the interaction force during walking, and control several physically based synthesis algorithms, which drive both the auditory and haptic feedback. The different hardware and software components of the system are described, together with possible uses and possibilities for improvements in future design iterations.**

I. Introduction

In the development of multimodal environments such as games and virtual reality installations, walking sounds are often used to produce physicality and a sense of action, with the overall goal of heightening the realism of a character or person. Such sounds are usually obtained from sound libraries or recorded by *Foley* artists, who often wear shoes in their hands and interact with different materials to simulate the act of walking on different surfaces. Such process is rather time-consuming, and can nowadays be complemented by approaches which take advantage of the developments of technology. As a matter of fact, recently the production of walking sounds has been approached algorithmically, and several solutions have been proposed to simulate the sounds produced during the interaction between a foot and a surface. One of the pioneers in this direction is Perry Cook, who proposed a collection of physically informed stochastic models (PhiSM), which simulate several everyday sonic events [4]. Among such algorithms the sounds of people walking on different surfaces were simulated [5]. A similar algorithm was also proposed in [9], where physically informed models simulate several stochastic surfaces. Procedural sound synthesis of walking has also been recently described in [8].

To our knowledge, the topic of audio-haptic simulation of the act of walking has not been investigated yet. However, lately the interest in investigating the interaction between touch and audition has grown. The possibility of investigating the interaction between auditory and haptic feedback has been facilitated by the rapid progress of haptic technology, together with the development of efficient and accurate simulation algorithms. Several studies have indeed investigated the multimodal recognition of textures [13] and stiffness, both using recorded stimuli [6] and simulations based on physical models [1].

The cited studies have focused on the interaction between touch and audition in hand based interactions. On the other hand, the interaction of auditory and haptic feedback in foot has not been studied, except by Giordano et al., who showed that the feet were also effective at probing the world with discriminative touch, with and without access to auditory information. Their results suggested that integration of foot-haptic and auditory information does follow simple integration rules [11].

In previous research, we described a system able to simulate the auditory and haptic sensation of walking on different materials and presented the results of a preliminary surface recognition experiment [14]. This experiment was conducted under three different conditions: auditory feedback, haptic feedback, and both.

By presenting the stimuli to the participants passively sitting in a chair, we introduced a high degree of control on the stimulation. However, this method of delivery is highly contrived since it eliminates the tight sensorimotor coupling that is natural during walking and foot interaction. It is true for the auditory channel, but even more so for the haptic channel. In spite of these drastically constrained conditions, performance was surprisingly good.

In this paper, we extend the developed technology by allowing subjects to walk in a controlled laboratory, where their steps are tracked and used to drive the simulation. We believe that introducing a higher level of interactivity significantly enhances the perceived quality and realism of the simulation.

The results presented in this paper are part of the Natural

MMSP'10, October 4-6, 2010, Saint-Malo, France.
???-?-????-????-?/10/$??.?? *©2010 IEEE.*

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 feedback play an important role.

II. Simulation Software

We developed a physically based synthesis engine able to simulate the auditory and haptic sensation of walking on different surfaces. Acoustic and vibrational signatures of locomotion are the result of more elementary physical interactions, including impacts, friction, or fracture events, between objects with certain material properties (hardness, density, etc.) and shapes. The decomposition of complex everyday sound phenomena in terms of more elementary ones has been an organizing idea in auditory display research during recent decades [10]. In our simulations, we draw a primary distinction between solid and aggregate ground surfaces, the latter being assumed to possess a granular structure, such as that of gravel.

Fig. 2. A picture of one pressure sensor and two actuators embedded in the shoes.

This model and its discretization are described elsewhere in detail [3]. The model has been recently adapted to the audio simulation of footsteps [15]. Here, we used the same model to drive the haptic and the audio synthesis. It is briefly recalled below.

A footstep sound may be considered to cause multiple micro-impacts between a sole, i.e., an *exciter*, and a floor, i.e., a *resonator*. Such interaction can be either discrete, as in the case of walking on a solid surface, or continuous, as in the case of a foot sliding across the floor.

In the simulation of discrete impacts, the excitation is brief and has an unbiased frequency response. The interaction is modelled by a Hunt-Crossley-type interaction where the force, f, between two bodies, combines hardening elasticity and a dissipation term [12]. Let x represent contact interpenetration and $\alpha > 1$ be a coefficient used to shape the nonlinear hardening, the special model form we used is

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

The model described was discretized as proposed in [2].

If the interaction called for slip, we adopted a model where the relationship between relative velocity v of the bodies in contact and friction force f is governed by a differential equation rather than a static map [7]. Considering that friction results from a large number of microscopic damped elastic bonds with an average deflection z, a viscous term, $\sigma_2 v$, and a noise term, $\sigma_3 w$, to represent roughness, we have

$$f(z, \dot{z}, v, w) = \sigma_0 z + \sigma_1 \dot{z} + \sigma_2 v + \sigma_3 w.$$

The force specified by these models is applied to a virtual mass which produces a displacement signal that is then processed by a linear shaping filter intended to represent the resonator. A solid surface is represented by an impact and a slide. The impact model alone was used to recreate the sound and the feel produced when walking on wood. The friction model was tuned to simulate walking on creaking wood.

To simulate walking on aggregate grounds, we used a physically informed sonic models (PhiSM) algorithm [4]. Stochastic parameterization is employed to simulate particle interactions thereby avoiding to model each of many particles explicitely. 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.

III. Simulation Hardware

In order to provide both audio and haptic feedback, haptic shoes enhanced with pressure sensors have been developed. A pair of light-weight sandals was procured (Model Arpenaz-50, Decathlon, Villeneuve d'Ascq, France). This particular model has light, stiff foam soles that are easy to gouge and fashion. Four cavities were made in the tickness of 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. As indicated in Figure 1 and Figure 2, two actuators were placed under the heel of the wearer and the other two under the ball of the foot. There were bonded in place to ensure good transmission of the vibrations inside the soles. When activated, vibrations propagated far in the light, stiff foam. In the present configuration, the four actuators were driven by the same signal but could be activated separately to emphasize, for instance, the front or back activation, to strick a balance, or to realize other effects such

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

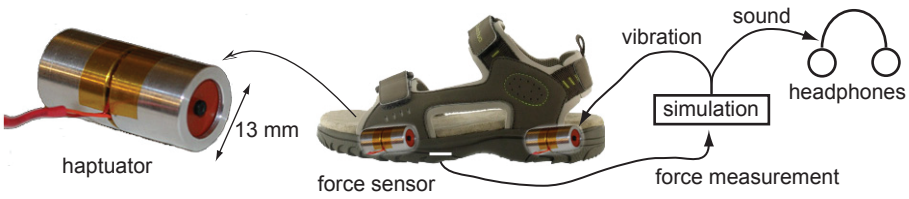


Fig. 1. System (one shoe shown). Left: recoil-type actuation from Tactile Labs Inc. The moving parts are protected by an alumimum enclosure able to bear the weight of a person. Middle: approximate location of the actuators in the sandal. Right: system diagram showing the interconnections.

as modulating different, back-front signals during heel-toe movements.

The sole has two force sensitive resistors (FSRs) pressure sensors[2] whose aim is to detect the pressure force of the feet during the locomotion of a subject wearing the shoes. The two sensors are placed in correspondence to the heel and toe respectively in each shoe.

The analogue values of each of these sensors are digitized by means of an Arduino Diecimila board[3] and used to drive the audio and haptic synthesis.

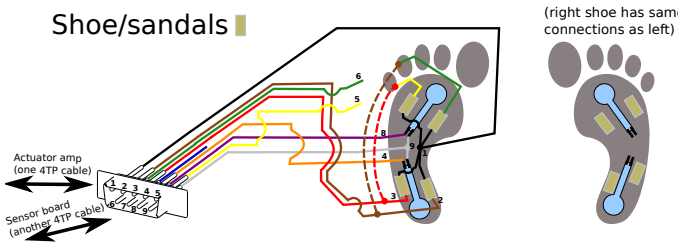


Fig. 3. Schematic representation of the cabling required to run sensors and actuators for one shoe.

A cable exits from each shoe, with the function of transporting the signals for the pressure sensors and for the actuators. Such cables are about 5 meters long, and they are connected through DB9 connectors to two 4TP (twisted pair) cables. One 4TP cable carries the sensor signals to a breakout board which contains trimmers, that form voltage dividers with the FSRs, which then interfaces to an Arduino board. The other 4TP cable carries the actuator signals from a pair of Pyle Pro PCA1[4] mini 2X15 W stereo amplifiers, driven by outputs from a FireFace 800 soundcard.[5] Each stereo amplifier handles 4 actuators found on a single shoe, each output channel of the amplifier driving two actuators connected in parallel. The PC handles the Arduino through a USB connection, and the FireFace soundcard through a FireWire connection. The connection among the different elements of the system is illustrated in Figure 4.

IV. Implementation

Using the algorithms just described we implemented a comprehensive collection of footstep sounds.

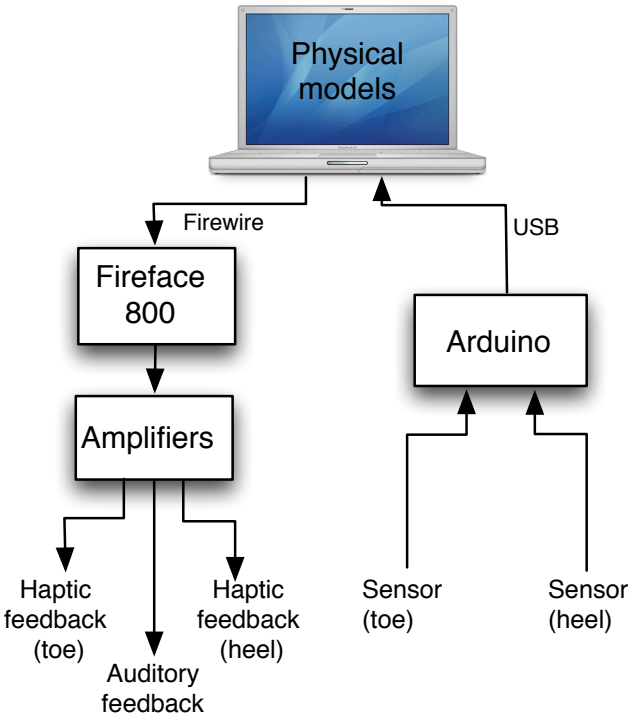


Fig. 4. Diagram illustrating the different hardware components of the system, together with their connections to the PC. The representation is for one shoe.

The sound synthesis algorithms were implemented in C++ as external libraries for the Max/MSP sound synthesis and multimedia real-time platform.[6] To enable compatibility with the Pure Data platform, [7] the algorithms were implemented using Flext.[8] One of the challenges in implementing the sounds of different surfaces was to find the suitable combinations of parameters and their range of variations which provided a realistic simulation. In our simulations, designers have access to a sonic palette making it possible to manipulate all such parameters, including material properties.

[2] I.E.E. SS-U-N-S-00039
[3] http://arduino.cc/
[4] http://www.pyleaudio.com/manuals/PCA1.pdf
[5] http://www.rme-audio.com/english/firewire/ff800.htm

[6] www.cycling74.com
[7] www.puredata.org
[8] http://puredata.info/Members/thomas/flext

The synthesized sounds have also been enhanced with reverberation algorithms. For this purpose we experimented with two approaches. The first was an algorithmically generated reverb working in real time, implemented as external for Max/MSP, called gigaverb∼. [9] The second made use of the technique of convolving a signal with a impulse response; such an approach was possible in real time, thanks to an external object allowing convolution with zero latency. [10]

The best results in sound quality were found using the second approach, which allowed to render more realistically the sizes of various kinds of indoor environments according to the impulse response chosen.

V. Controlling the Engine

The footsteps synthesizer has been designed and implemented to be controlled by a unique input parameter, that is the ground reaction force (GRF). The real-time extrapolation of a GRF from the signals coming from the sensors turned out to be not the right choice for the control of the sound synthesis engine because of the features of the signal itself. For that reason we opted for a solution based on recorded GRF files and on a system of thresholds applied both on the signals and on their first derivatives. In particular, we used the values of the first derivative as control for triggering, into the footsteps synthesizer, some GRFs corresponding to heel or toe according to the activated sensor.

While a subject walks there is a variation of the values of the pressure sensors in correspondence to each step. Such variation is the basis for obtaining first time derivatives of the sensors signals, which remain related to the intensity with which the foot hits the ground. Each time the value of the first derivative becomes bigger than a threshold, the GRF corresponding to the activated sensor is triggered into the engine. More precisely we checked only positive changes in the derivative value, since we were interested in the generation of the sound when the step hits – and not when it leaves – the ground. Other thresholds, both on the signals and on their first derivatives, were used in order to handle some boundary conditions, like the standing of the subject, with the aim of controlling the generation of sound. Such thresholds are set in a phase of calibration of the system, which has to take into account the different weights of the subjects wearing the shoes, in order to have an average value suitable for all the possible cases.

The GRFs triggered have been created by extracting the amplitude envelope from audio files of recorded footsteps on concrete, divided into heel and toe components. Such sounds were chosen among those available on the Hollywood Edge sound effects library. [11] Five types of heel and toes audio files were used and randomly chosen at the moment of the triggering, giving rise to 25 possible combinations. Such behavior has been adopted in order to not have always the same GRF as input of the engine, and this allows to have differences in the generated sounds at every step, increasing thus the degree of realism of the walking experience.

The input GRF controls directly all or some of the parameters of the various algorithms, as well as the range of variation of the amplitudes of both the subcomponents and global sound. Figure 5 shows the waveform of one of the footstep sounds used and its corresponding extracted GRF. The GRF was calculated from the input signal extracting its amplitude envelope. To perform envelope extraction we used a simple non-linear low-pass filter proposed by Cook in [16]:

$$e(n) = (1 - b(n))|x(n)| + b(n)e(n-1)$$

where

$$b = \begin{cases} b_{up} & \text{if } |x(n)| > e(n-1) \\ b_{down} & \text{otherwise} \end{cases}$$

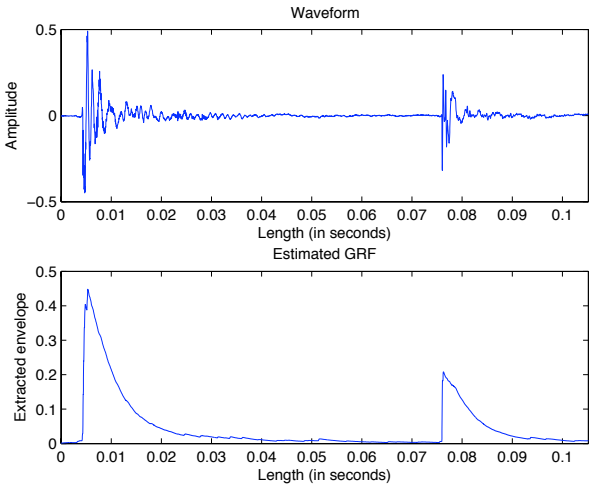


Fig. 5. *Time domain waveform of the used footstep on wood (top) and its relative extracted GRF (bottom).*

Figure 6 shows on top a waveform of a characteristic signal captured by a pressure sensor, and its derivative on the bottom. It is possible to notice the spikes on the derivative, which were used as time information for triggering a corresponding GRF.

VI. Evaluation of the System

The system has been evaluated by using it both offline and interactively. The complete results of this evaluation are described in [14], [17]. Here, we limit ourselves to summarize the interesting results, which can help in the development of the next shoe prototype. Overall, results show that subjects are able to recognize most of the synthesized surfaces with high accuracy. Results moreover confirm that auditory modality is dominant on the haptic

[9] Available at http://www.akustische-kunst.org
[10] http://www-users.york.ac.uk/ ajh508/index.html
[11] www.hollywoodedge.com/.

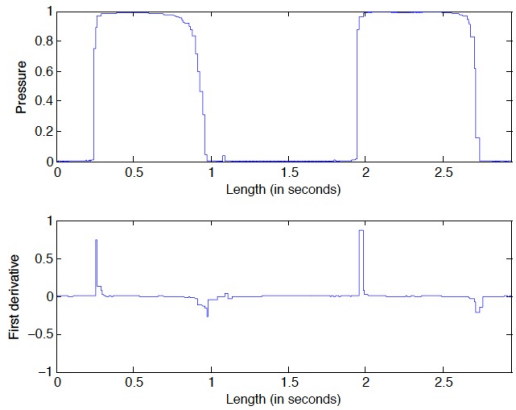


Fig. 6. Waveform of a characteristic signal captured by the pressure sensors (top), and its derivative (bottom).

modality and that the haptic task was more difficult than the other two. This can be due to the low sensitivity of the foot when exposed to haptic signals. Such results can also be improved by the design of more powerful actuators, which are currently under development. It is interesting to notice that similar accuracy can be observed in the recognition of real recorded footsteps sounds, which is an indication of the success of the proposed algorithms and their control.

VII. Conclusion

In this paper, we introduced a real-time footsteps synthesizer able to provide audio and haptic feedback, and which is controlled by the user during the act of walking by means of shoes embedded with sensors and actuators.

The developed system is ready to be integrated in computer games and interactive installations where a user can navigate. In future work, we indeed plan to utilize the system in multimodal environments including visual feedback, to understand the role of the different sensorial modalities to enhance sense of immersion and presence in scenarios where the act of walking plays an important role.

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. [12]

References

[1] F. Avanzini and P. Crosato. Haptic-auditory rendering and perception of contact stiffness. *Lecture Notes in Computer Science*, 4129:24, 2006.

[2] F. Avanzini and D. Rocchesso. Modeling collision sounds: Non-linear contact force. In *Proc. COST-G6 Conf. Digital Audio Effects (DAFx-01)*, pages 61–66, 2001.

[3] F. Avanzini, S. Serafin, and D. Rocchesso. Interactive simulation of rigid body interaction with friction-induced sound generation. *IEEE transactions on speech and audio processing*, 13(5 Part 2):1073–1081, 2005.

[4] P. Cook. Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. *Computer Music Journal*, 21(3):38–49, 1997.

[5] 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.

[6] D. DiFranco, G. L. Beauregard, and M. Srinivasan. The effect of auditory cues on the haptic perception of stiffness in virtual environments. *Proceedings of the ASME Dynamic Systems and Control Division.*, 1997.

[7] P. Dupont, V. Hayward, B. Armstrong, and F. Altpeter. Single state elastoplastic friction models. *IEEE Transactions on Automatic Control*, 47(5):787–792, 2002.

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

[9] 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.

[10] W. Gaver. What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology*, 5(1):1–29, 1993.

[11] B. L. Giordano, S. Mcadams, Y. Visell, J. Cooperstock, H. Y. Yao, and V. Hayward. Non-visual identification of walking grounds. *Journal of the Acoustical Society of America*, 123(5):3412–3412, 2008.

[12] K. H. Hunt and F. R. E. Crossley. Coefficient of restitution interpreted as damping in vibroimpact. *ASME Journal of Applied Mechanics*, 42(2):440–445, 1975.

[13] S. Lederman. Auditory texture perception. *Perception*, 1979.

[14] R. Nordahl, A. Berrezag, S. Dimitrov, L. Turchet, V. Hayward, and S. Serafin. Preliminary experiment combining virtual reality haptic shoes and audio synthesis. In *Proc. Eurohaptics*, 2010.

[15] R. Nordahl, S. Serafin, and L. Turchet. Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In *Proc. IEEE VR 2010*, 2010.

[16] L. Peltola, C. Erkut, P. Cook, and V. Valimaki. Synthesis of hand clapping sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(3):1021–1029, 2007.

[17] 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. In *Proc. Eurohaptics symposium on Haptics and Audio-visual environments*, 2010.

[12] www.niwproject.eu

Conflicting Audio-haptic Feedback in Physically Based Simulation of Walking Sounds

Luca Turchet, Stefania Serafin, Smilen Dimitrov, and Rolf Nordahl

Medialogy, Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, DK
{tur,sts,sd,rn}@media.aau.dk

Abstract. We describe an audio-haptic experiment conducted using a system which simulates in real-time the auditory and haptic sensation of walking on different surfaces. The system is based on physical models, that drive both the haptic and audio synthesizers, and a pair of shoes enhanced with sensors and actuators. Such experiment was run to examine the ability of subjects to recognize the different surfaces with both coherent and incoherent audio-haptic stimuli. Results show that in this kind of tasks the auditory modality is dominant on the haptic one.

1 Introduction

While several studies have investigated the interaction between touch and audition in hand based interactions, to our knowledge, the interaction of auditory and haptic feedback in foot based devices is still an unexplored topic.

A notable exception is the work of Giordano et al., who showed that the feet were also effective at probing the world with discriminative touch, with and without access to auditory information. Their results suggested that integration of foot-haptic and auditory information does follow simple integration rules [1].

In previous research, we described a system able to simulate the auditory and haptic sensation of walking on different materials and presented the results of a preliminary surface recognition experiment [2]. This experiment was conducted under three different conditions: auditory feedback, haptic feedback, and audio-haptic feedback. By presenting the stimuli to the participants passively sitting in a chair, we introduced a high degree of control on the stimulation. However, this method of delivery is highly contrived since it eliminates the tight sensorimotor coupling that is natural during walking and foot interaction. It is true for the auditory channel, but even more so for the haptic channel. In spite of these drastically constrained conditions, performance was surprisingly good.

In particular, the results indicated that subjects were able to recognize most of the stimuli in the audition only condition, and some of the material properties such as hardness in the haptics only condition. Nevertheless, the combination of auditory and haptic cues did not significantly improve recognition.

In a successive research we extended that work improving the developed technology which allowed subjects to walk in a controlled laboratory, where their

R. Nordahl et al. (Eds.): HAID 2010, LNCS 6306, pp. 97–106, 2010.
© Springer-Verlag Berlin Heidelberg 2010

steps were tracked and used to drive the simulation [4]. Overall, results showed that subjects were able to recognize most of the synthesized surfaces with high accuracy. Results moreover confirmed that auditory modality is dominant on the haptic modality and that the haptic task was more difficult than the other two. Indeed such results showed that subjects performed the recognition task better when using auditory feedback versus haptic feedback, and that the combination of auditory and haptic feedback only in some conditions significantly enhanced the recognition.

Starting from those results, in this paper we investigate in a deeper way the role of dominance of the two modalities involved by means of a preliminary discrimination experiment. In particular, while in previous research we focused on providing coherent stimuli in the auditory and haptic modality, here we provide conflicting stimuli, to understand which modality is dominant.

The results presented in this paper are 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 feedback play an important role.

2 Simulation Hardware and Software

We developed a system which simulates in real-time the auditory and haptic sensation of walking on different surfaces. A schematic representation of this system is shown in Figure 1. In order to provide both audio and haptic feedback, haptic shoes enhanced with pressure sensors have been developed. The way pressure sensors and actuators are embedded in the sandals can be seen in Figure 2, and a picture of a user wearing the shoes is shown in Figure 3. A complete description of such system and of all its components is given elsewhere in detail [5].

The hardware allows to control in real-time of a sound synthesis engine based on physical models. Such engine is illustrated in our previous research [3,7,6]. The same physical models have been used to drive the haptic and the audio synthesis.

As mentioned in section 1, the system has been evaluated by using it both offline and interactively. The complete results of this evaluation are described in [2,4].

3 Experiment

We conducted an experiment whose goal was to investigate the role of dominance of the audio and haptic modalities during the use of our walking system. Subjects were asked to interact with the system and to recognize the different walking sounds and vibrations they were exposed to.

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

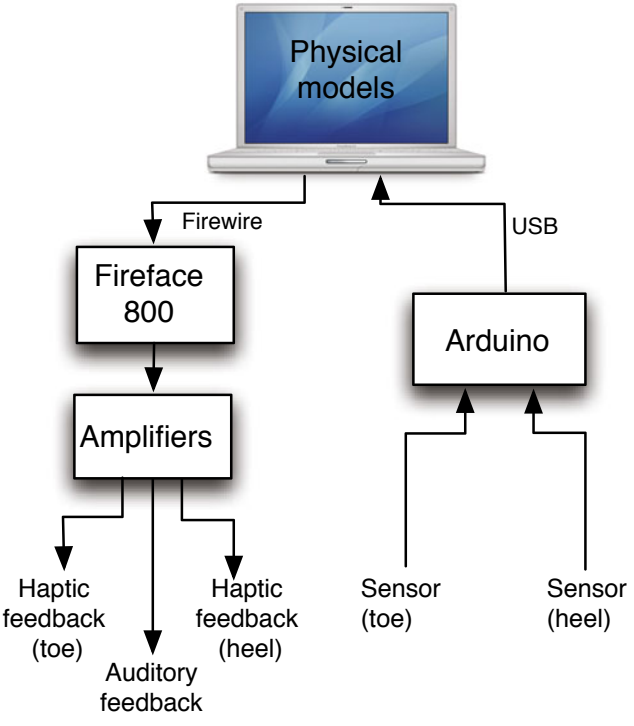


Fig. 1. Diagram illustrating the different hardware components of the system, together with their connections to the PC. The representation is for one shoe.

The experiment consisted of both coherent and incoherent audio-haptic stimuli. In presence of coherent stimuli the same surface material was presented both at audio and haptic level. Instead the provided incoherent stimuli consisted of different surface materials; in particular when at audio level a solid surface was presented, at haptic level an aggregate surface was modeled, and viceversa.

One of our hypotheses was that the audio modality would have dominated the haptic one. Another was that the recognition would have slightly improved using coherent stimuli rather than the incoherent ones. Similarly we hypothesized higher evaluations in terms of realism and quality in presence of coherent stimuli.

3.1 Participants

Ten participants, 7 male and 3 female, aged between 20 and 38 (mean = 25.81, standard deviation = 5.77), were involved in the experiment.

All participants reported normal hearing conditions and all of them were naive with respect to the experimental setup and to the purpose of the experiment.

The participants took on average about 11 minutes to complete the experiment.

Fig. 2. A picture of one pressure sensor and two actuators embedded in the shoes

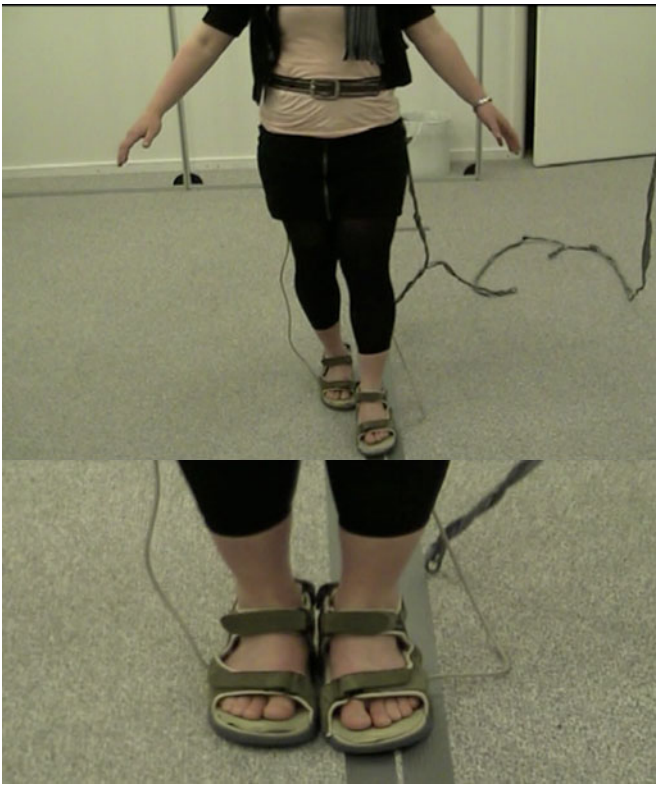


Fig. 3. Top: A user wearing the sandals enhanced with sensors and actuators. Bottom: the sandals.

3.2 Setup

The experiment was carried out in an acoustically isolated laboratory. The walking area was approximately 18 square meters, delimited by the walls of the laboratory.

The setup consisted of the pair of sandals mentioned in section 2, an Arduino board, a Fireface soundcard, a laptop and a set of headphones[2]. In order to facilitate the navigation of the subjects, the wires coming out from the shoes in all setups, as well as the wires connecting the headphones to the soundcard, were linked to a bumbag or to snaplinks attached to trousers.

3.3 Task

During the experiment participants were asked to wear the pair of sandals and the headphones described in sections 2 and 3.2, and to walk in the laboratory.

During the act of walking they listened simultaneously to footsteps sounds and vibrations on a different surface according to the stimulus presented. The task consisted of answering, by writing on a paper, the following three questions after the presentation of the stimulus:

1. Which surface do you think you are walking on? For each stimulus choose a material.
2. How close to real life is the sound in comparison with the surface you think it is? Evaluate the degree of realism on a scale from 1 to 7 (1=low realism, 7=high realism).
3. Evaluate the quality of the sound on a scale from 1 to 7 (1=low quality, 7=high quality).

As opposed to our previous research, participants were not provided with a forced list of possible choices. This was due to the fact that we wanted subjects to be somehow creative in their recognition of the surface, without guessing from a predefined list.

Subjects were informed that they could use the interactive system as much as they wanted before giving an answer. They were also told that they could choose the same material more than one time. When passed to the next stimulus they could not change the answer to the previous stimuli.

At the conclusion of the experiment, participants were asked some questions concerning the naturalness of the interaction with the system and to comment on its usability and possible integration in a virtual reality environment. In particular the questionnaire was the following:

- Imagine that this is part of a system used to navigate in a computer game, answer to the following questions:
  1. How natural is the interaction? Evaluate on a scale from 1 to 7 (1=little natural, 7=very natural)

[2] Sennheiser HD 600, http://www.sennheiser.com

2. How normal do you feel during the act of walking? Evaluate on a scale from 1 to 7 (1=little normal, 7=very normal)
3. How constrained do you feel during the act of walking? Evaluate on a scale from 1 to 7 (1=little constrained, 7=very constrained)

In addition they were also given the opportunity to leave an open comment on their experience interacting with the system.

3.4 Experimental Plan

Participants were exposed to 12 trials consisting of 4 coherent stimuli and 8 incoherent stimuli. The 12 audio-haptic stimuli were presented once in randomized order.

The modeled surfaces were 4 (2 solid and 2 aggregate): wood, metal, snow and gravel. In presence of incoherent stimuli the conflict was rendered providing on the one hand one of the solid surfaces by means of auditory feedback, while at haptic level one of the aggregate surfaces was presented. On the other hand, analogously, another set of incoherent stimuli consisted of aggregate surfaces at auditory level while a solid surface was presented by means of haptic feedback.

4 Results and Discussion

Table 1 shows the confusion matrix resulting from the experiment. In such table are illustrated the participants answers gathered according to three material categories: solid, aggregate and liquid.

The first noticeable element emerging from the table is that none of the participants classified as liquid the simulated surfaces. Moreover it is very evident that in presence of conflicts the auditory modality is strongly dominant on the haptic one. This result is more clearly illustrated in table 2, and it seems to be confirmed by the answers concerning the names of the chosen materials (see table 3). Indeed, participants had the tendency to answer with the same names chosen when a same material was presented at auditory level both in presence of coherent and incoherent stimuli. In other words they were driven in their choice by the auditory feedback.

This does not mean that they completely ignored the haptic part of the bimodal stimulus; indeed in the comments two participants reported that they noticed that only some of the haptic stimuli where appropriate for the sound they were listening. Moreover results show that in two cases participants seemed to be driven in their choices, at least partially, by the haptic feedback; this is the case of the audio-haptic stimulus metal-snow for which one participant surprisingly gave the answer “mud” which can be considered appropriate for the haptic stimulus. Concerning the stimulus snow-wood one participant chose the answer “carpet” which seems to consider both the two components of the bimodal stimulus since the sound of snow could be interpreted as the sound on carpet in presence of an haptic feedback expressing a solid surface like wood. Both the participants, like all the others, were asked, at the conclusion of the experiment,

Table 1. Confusion matrix of experiment

Stimulus		Answer			
Audio	**Haptic**	**Solid**	**Aggregate**	**Liquid**	**I don't know**
Wood	Wood	7	1		2
Wood	Snow	7			3
Wood	Gravel	6	1		3
Metal	Metal	9			1
Metal	Snow	6	1		3
Metal	Gravel	8			2
Gravel	Gravel		9		1
Gravel	Wood		10		
Gravel	Metal		7		3
Snow	Snow		10		
Snow	Wood		9	1	
Snow	Metal		9		1

Table 2. Percentages of dominance of the auditory and haptic modalities in presence of incoherent stimuli

Stimulus		Dominance	
Audio	**Haptic**	**% Audio**	**% Haptic**
Wood	Snow	70	0
Wood	Gravel	60	0
Metal	Snow	60	10
Metal	Gravel	80	0
Gravel	Wood	100	0
Gravel	Metal	70	0
Snow	Wood	90	10
Snow	Metal	90	0

to explain their answers and they confirmed their choices. In particular all the participants were asked at the end of the experiment to classify their answers as belonging to the categories of solid, aggregate or liquid materials.

For the stimulus wood-gravel one participant reported the answer "not solid plastic" which could be addressed as haptic dominance, but the same participant chose the same answer also for the coherent stimulus wood-wood and for this reason we did not consider such answer as haptic dominance in table 2.

Regarding the percentages of "I don't know" answers, although the number is low, it is an indication of the difficulty of the proposed task. This fact was also confirmed in numerous comments left by participants who on average reported that the task was very difficult. Although the comparison between coherent and incoherent stimuli with the same auditory stimulus do not reveal any statistically significant difference (confirming from another point of view the dominance of the auditory feedback), the percentages of "I don't know" answers is on average higher for the incoherent stimuli.

Table 3. Names of the materials chosen for each audio-haptic stimulus. In bold the choices which seem to be driven by the haptic feedback.

Stimulus		**Answer**
Audio	**Haptic**	**Names of chosen materials**
Wood	Wood	wood, concrete, plastic, not solid plastic
Wood	Snow	wood, concrete, plastic, gum
Wood	Gravel	wood, concrete, plastic, not solid plastic
Metal	Metal	metal, iron, steel, wood, glass
Metal	Snow	metal, iron, steel, wood, glass, **mud**
Metal	Gravel	metal, iron, steel, wood, glass, plastic
Gravel	Gravel	gravel, little stones, sand
Gravel	Wood	gravel, little stones, sand
Gravel	Metal	gravel, little stones, sand
Snow	Snow	snow, ice, gravel, sand, leaves, not solid plastic, paper
Snow	Wood	snow, ice, gravel, sand, leaves, not solid plastic, **carpet**
Snow	Metal	snow, ice, gravel, sand, not solid plastic

Table 4. Average realism and quality scores from a seven-point Likert scale and relative standard deviation

Stimulus		**Realism**		**Quality**	
Audio	**Haptic**	μ	σ	μ	σ
Wood	Wood	2.75	1.488	3.25	1.2817
Wood	Snow	2.7143	1.496	3.2857	1.2536
Wood	Gravel	2.5714	1.8127	2.7143	1.3801
Metal	Metal	2.8889	1.6159	4	2
Metal	Snow	3.8571	1.496	3.5714	1.1339
Metal	Gravel	2.625	1.3025	2.875	1.5526
Gravel	Gravel	3.2222	1.0929	4	1
Gravel	Wood	4	0.9428	4.4	0.9661
Gravel	Metal	3.7143	1.8898	4.1429	1.4639
Snow	Snow	3.7	1.567	3.8	1.3984
Snow	Wood	3.6	1.4298	3.7	1.567
Snow	Metal	4	1.5	4.1111	1.453

Table 5. Questionnaire results. Average scores from a seven-point Likert scale and relative standard deviation.

	μ	σ
Naturalness	3.5	1.6499
Normality	4	1.5635
Constriction	4.2	1.3166

Table 4 shows the degree to which participants judged the realism and the quality of the experience. Such parameters were calculated by looking only at the answers different from "I dont know". Contrary to our hypotheses we did not find higher evaluations of these parameters for the coherent stimuli compared to the incoherent one. Surprisingly for some stimuli the evaluations are even higher for the incoherent stimuli. Anyways an in depth statistical analysis performed with the t-test revealed that all these differences are not significative.

Finally, as concerns the questionnaire conducted at the conclusion of the experiment, results in table 5 show that that subjects judged the interaction with the system not too much natural (mean = 3.5), and that they felt quite normal (mean = 4) but at the same time quite constrained (mean = 4.2) during the act of walking.

Indeed, more than one subject commented on the need of a wireless system able to convey vibrations to the shoes and sounds to the headphones set.

5 Conclusion and Future Work

In this paper, we describe an experiment conducted with a real-time footsteps synthesizer able to provide audio and haptic feedback, and which is controlled by the user during the act of walking by means of shoes embedded with sensors and actuators.

In the experiment, both coherent and incoherent audio-haptic stimuli were provided. Results confirm that auditory modality is dominant on the haptic one. This can be due to the low sensitivity of the foot when exposed to haptic signals.

The developed system is ready to be integrated in computer games and interactive installations where a user can navigate.

In future work, we indeed plan to utilize the system in multimodal environments, and include visual feedback, to understand the role of the different sensorial modalities to enhance sense of immersion and presence in scenarios where walking plays an important role.

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.[3] The authors wish also to thank Amir Berrezag and Vincent Hayward who provided the shoes used in the experiments.

References

1. Giordano, B.L., Mcadams, S., Visell, Y., Cooperstock, J., Yao, H.Y., Hayward, V.: Non-visual identification of walking grounds. Journal of the Acoustical Society of America 123(5), 3412–3412 (2008)

[3] www.niwproject.eu

2. Nordahl, R., Berrezag, A., Dimitrov, S., Turchet, L., Hayward, V., Serafin, S.: Preliminary experiment combining virtual reality haptic shoes and audio synthesis. In: Proc. Eurohaptics (2010)
3. Nordahl, R., Serafin, S., Turchet, L.: Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In: Proc. IEEE VR 2010 (2010)
4. 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: Proc. of Eurohaptics Symposium on Haptics and Audio-Visual Environments (2010)
5. Turchet, L., Nordahl, R., Berrezag, A., Dimitrov, S., Hayward, V., Serafin, S.: Audio-haptic physically based simulation of walking sounds. In: Proc. of IEEE International Workshop on Multimedia Signal Processing (2010)
6. Turchet, L., Nordahl, R., Serafin, S.: Examining the role of context in the recognition of walking sounds. In: Proc. of Sound and Music Computing Conference (2010)
7. 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)

PHYSICALLY BASED SOUND SYNTHESIS AND CONTROL OF FOOTSTEPS SOUNDS

Luca Turchet, Stefania Serafin, Smilen Dimitrov, Rolf Nordahl

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

ABSTRACT

We describe a system to synthesize in real-time footsteps sounds. The sound engine is based on physical models and physically inspired models reproducing the act of walking on several surfaces. To control the real-time engine, three solutions are proposed. The first two solutions are based on floor microphones, while the third one is based on shoes enhanced with sensors. The different solutions proposed are discussed in the paper.

1. INTRODUCTION

Footsteps sounds represent important elements in movies and computer games. Such sounds are used to produce embodiment and a sense of weight with the overall goal of heightening the sense of "realness" to the character or person. Usually such sounds are obtained from sound libraries or recorded by *Foley* artists. Such artists wear shoes in their hands and interact with different materials to simulate the act of walking.

Recently, several algorithms have been proposed to simulate the sounds of walking. One of the pioneers in this field is Perry Cook, who proposed a collection of physically informed stochastic models (PhiSM) simulating several everyday sonic events [1]. Among such algorithms the sounds of people walking on different surfaces were simulated [2]. A similar algorithm was also proposed in [3], where physically informed models simulate several stochastic surfaces.

Procedural sound synthesis of walking has also been recently described in [4]. The characteristic events of a footstep sounds were reproduced by simulating the so-called ground reaction force (GRF), i.e., the reaction force supplied by the ground at every step

The results presented in this paper are 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 feedback play an important role. In this paper, we describe a sound synthesis engine developed in the context of the NIW project. Different solutions to control the engine are also described and discussed.

2. THE SOUND SYNTHESIS ENGINE

We developed a physically based sound synthesis engine able to simulate the sounds of walking on different surfaces. Acoustic and vibrational signatures of locomotion are the result of more elementary physical interactions, including impacts, friction, or fracture events, between objects with certain material properties (hardness, density, etc.) and shapes. The decomposition of complex everyday sound phenomena in terms of more elementary ones has been an organizing idea in auditory display research during recent decades [5]. In our simulations, we draw a primary distinction between solid and aggregate ground surfaces, the latter being assumed to possess a granular structure, such as that of gravel.

A footstep sound can be considered as the result of multiple micro-impact sounds between a shoe and a floor. The set of such micro-events can be thought as the result of the interaction between an *exciter* and a *resonator*. In mechanics such exciter is usually called ground reaction force. Section 3 illustrates how such force has been calculated starting from three kinds of input signals. The estimated GRF has been used to control different sound synthesis algorithms, which reproduce solid and aggregate surfaces.

2.1. Solid surfaces

Sonic interactions between solid surfaces have been extensively investigated, and results are available which describe the relationship between physical and perceptual parameters of objects in contact [6, 7]. Such sounds are typically short in duration, with a sharp temporal onset and relatively rapid decay.

A common approach to synthesize such sounds is based on a lumped source-filter model, in which an impulsive excitation $s(t)$, modelling the physics of contact, is passed through a linear filter $h(t)$, modelling the response of the vibrating object as $y(t) = s(t) \star h(t)$.

Modal synthesis [8] is one widely adopted implementation of this idea. In this synthesis technique, the response model $h(t)$ is decomposed in terms of the resonant frequencies f_i of the vibrating object, also known as the modes of the object. The response is modelled as a bank of filters with impulse response $h(t) = \sum_i a_i e^{-b_i t} \sin(2\pi f_i t)$, where a_i represent the amplitudes of the modes, b_i the decay rates of the modes, and f_i the frequencies of the modes.

In our situation, the simulation of the interaction between a shoe and a floor is obtained by decomposing the resulting sound into an exciter and a resonator. Such interaction can be either continuous, as in the case of a foot sliding across the floor, or discrete, as in the case of walking on a solid surface. To simulate such scenarios, both an impact and friction model were implemented.

In the impact model, the excitation corresponding to each impact $s(t)$ is assumed to possess a short temporal extent and an unbiased frequency response. Such excitation consists of a discrete-time model of the force f between the two bodies, dependent

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

on additional parameters governing the elasticity of the materials, their velocity of impact $\dot{x}$, and masses:

$$f(x,\dot{x}) = \begin{cases} -kx^{\alpha} - \lambda x^{\alpha}\dot{x} & \text{if } x > 0 \\ 0 & x \leq 0 \end{cases}$$

where α depends on the local geometry around the contact surface, and x stands for the compression of the exciter (when $x > 0$ the two objects are in contact) [9].

In the friction model we adopted a dynamic model, where the relationship between relative velocity v of the bodies in contact and friction force f is represented through a differential equation rather than static mapping. Assuming that friction results from a large number of microscopic elastic bonds, called bristles in [10], the v-to-f relationship is expressed as:

$$f(z,\dot{z},v,w) = \sigma_0 z + \sigma_1 \dot{z} + \sigma_2 v + \sigma_3 w$$

where z is the average bristle deflection, the coefficient σ_0 is the bristle stiffness, σ_1 the bristle damping, and the term $\sigma_2 v$ accounts for linear viscous friction. The fourth component $\sigma_3 w$ relates to surface roughness, and is simulated as fractal noise.

2.2. Aggregate surfaces

To synthesize aggregate surfaces, we implemented the physically informed sonic models (PhiSM) algorithm [1].

The PhiSM simulates particle interactions by using a stochastic parametrization. This means that the different particles do not have to be modelled explicitly, but only the probability that particles will create some noise is simulated. For many particle systems, this phenomenon is well taken into account by using a simple Poisson distribution, where the sound probability is constant at each time step, giving rise to an exponential probability waiting time between events.

The continuous crumpling model is based on the impact model, on top of which a statistics of temporal impact events is superimposed.

2.3. Sound design

In order to investigate how the combination of parameters in the different basic models described in the previous section affects the perception of material, a generalized footsteps synthesizer has been built adopting the following design approach.

A footstep sounds is dependent on the kind of shoes the subject wears and obviously the kind of surface the user is walking on. We designed and synthesized the different sounds assuming that the shoes hitting the floor had a solid sole. This aspect is more important in the simulation of solid floors rather than of the non homogeneous ones.

The algorithms to synthesize such sounds have been developed combining a spectral analysis of recordings from real footsteps with some ad-hoc manipulations of the control parameters of the different algorithms. The starting point has been the listening of the recordings of real footsteps sounds, in order to extrapolate the main features characterizing the sound of the footsteps on each surface, giving particular attention to those simply recognizable to a first listening. Indeed, the different components clearly noticeable in each sound, i.e., different subevents that characterize the sound itself, have been took into consideration, with the aim of simulating independently them and their evolution in the time, and subsequently of combining them appropriately in order to construct the wanted global sound.

As an example, the sound produced while walking on dry leaves is a combination of granular sounds with long duration both at low and high frequencies, and noticeable random sounds with not very high density that give to the whole sound a crunchy aspect. Another example is the sound of walking on gravel, composed by the contribution of the sounds of stones of different dimensions, which when colliding give rise to different random sounds with different features.

The amplitude of the different components were also appropriately weighed, according to the same contribution present in the corresponding real sounds. Finally, a scaling factor for the sub-components volumes gives to the whole sound an appropriate volume, in order to recreate a similar sound level which it would happen during a real footstep on each particular material.

2.4. Implementation

Using the algorithms described in Sections 2.1 and 2.2, as well as the sound design paradigms illustrated in Section 2.3, we implemented a comprehensive collection of footstep sounds.

The sound synthesis algorithms were implemented in C++ as external libraries for the Max/MSP sound synthesis and multimedia real-time platform.[2] To enable compatibility with the Pure Data platform, [3] the algorithms were implemented using Flext.[4] A screenshot of the final graphical user interface can be seen in Figure 1.

The footsteps synthesizer has been designed and implemented to be controlled by an unique input parameter, that is the ground reaction force. Such input parameter has been normalized in order to be a time-varying continuous curve ranging on a scale [0,1]. To this purpose a setup control has been developed in order to detect a plausible maximum input value to be scaled to 1 (and all the values bigger than such a maximum are clipped to 1). Moreover, a threshold to eliminate any background noise was placed in order to set to 0 any incoming value under it. In this way the algorithms developed work independently from the system which the GRF is detected with.

The working of the algorithms is also based on the control of both the end of each step and the end of the footstep sound, produced by the algorithms themselves. Such informations are extrapolated from the detected input GRF, thanks to a system of thresholds.[5] Each time that a step is finished, a new combination of some parameters and some amplitudes is calculated for the synthesis of the next step, thanks to random numbers varying in appropriate ranges. Such a behavior allows to increase noticeably the degree of realism of the proposed sounds (as in real life, the sound of each step is different from the previous).

Finally, the input GRF controls directly all or some of the parameters of the various algorithms, as well as the range of variation of the amplitudes of both the subcomponents and global sound. One of the challenges in implementing the sounds of different surfaces was to find the suitable combinations of parameters and their

[2] www.cycling74.com

[3] www.puredata.org

[4] http://puredata.info/Members/thomas/flext

[5] For efficiency in computational load the end of the sound produced by the synthesis algorithms is set to 400 ms after the detected end of the step. Indeed on average the duration of the sound on the various materials does not last after such a temporal interval. In this way only the control on the end is performed.

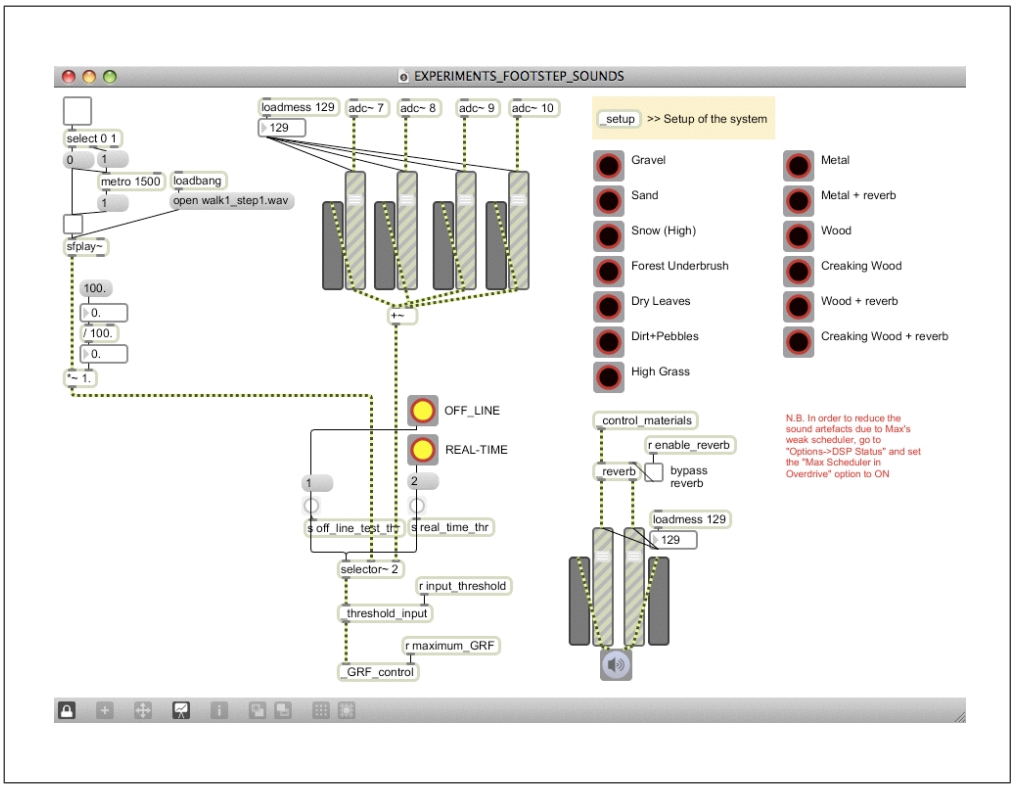


Figure 1: The sound synthesis engine.

range of variations which provided a realistic simulation. In our simulations, designers have access to a sonic palette making it possible to manipulate all such parameters, including material properties.

The synthesized sounds have also been enhanced with reverberation algorithms. For this purpose we experimented with two approaches. The first was an algorithmically generated reverb working in real time, implemented as external for Max/MSP, called gigaverb∼. [6] The second made use of the technique of convolving a signal with a impulse response; such an approach was possible in real time, thanks to an external object allowing convolution with zero latency. [7]

The best results in the sound quality were found using the second approach, which allowed to render more realistically the sizes of various kinds of indoor environments according to the impulse response chosen.

The following solid and aggregate surfaces have been simulated and tested [11, 12, 13]. [8]

Wood and creaking wood: the sound of footsteps on wood has been synthesized by means of the impact model. In particular the GRF controls only the impact force parameter, while all the other parameters do not change their status.
Some creaking sounds have also been simulated, in order to realistically synthesize the typical sounds of footsteps on parquet floors. Such sounds have been generated by means of the friction model, controlling with random ramps the external rubbing force and pressure on rubber parameters. The ranges of variation and duration of such ramps have been set by means of random numbers, which are calculated each time the step is detected as finished. The simulation of creaking sounds enhances the realism of the sound of walking on wood, because their frequency, amplitude and duration change at every step. Such creaking sounds are also randomly generated, as it happens in real life.

Metal: the sound of footsteps on metal has been synthesized by means of the impact model. In particular the GRF controls only the impact force parameter, while all the other parameters do not change their status. Various kinds of metal can be synthesized thanks to the model.

Deep and low snow: the footsteps sound on snow has been synthesized by means of one PhISM model and one crumpling model in order to simulate the two subcomponents of the sound produced when the foot drops into the snow and the snow breaks under the foot respectively. Deep and low snow have been developed thanks to different settings of the parameters of both the models.
The incoming input GRF controls both the system energy parameter of the PhISM model and the volume of the crumpling model. In this way higher the GRF and higher is the amplitude of the sound produced, as it happens in the reality when the foot drops into the snow with various intensity.
As concerns the PhISM model the GRF is mapped in the system decay parameter, and when the sound produced by the algorithms is detected as finished, a new random number is calculated to con-

[6] Available at http://www.akustische-kunst.org
[7] http://www-users.york.ac.uk/ ajh508/index.html
[8] Related sound examples can be found at http://www.niwproject.eu.

trol the sound decay parameter. As regards the crumpling model the GRF is mapped into the range of variation of the impact force parameter, while when the sound produced by the algorithms is detected as finished, new random numbers are calculated to control the density of the crumpling effects, and the parameters resistance, contact surface (α) and decay of mode number 0.
All the other parameters for both the models do not change their status once set.

Gravel: the footstep sound on various kinds of gravel has been simulated by means of the combination of one, two or three PhISM models. This kind of sound is composed by the contribution of the sounds of stones of different dimensions, which when colliding give rise to different random sounds with different features. It is possible to reproduce such a complex sound by setting appropriately the parameters of the models in order to simulate one, two or three kinds of distinct collisions between stones of the same dimension.
The system energy parameter of each model is controlled by the incoming input GRF. In this way a higher GRF creates a higher amplitude of the sound produced. In addition to this, the GRF is mapped to the system decay parameter, and this allows to simulate the degree of dispersion of the little stones from the foot.
Finally, the number of colliding objects parameter is calculated randomly each time the step is finished, while all the other parameters do not change their status.

Beach sand: the sound of the footsteps on beach sand has been developed by means of a single PhISM model. The GRF is mapped to the sound decay parameter while all the other parameters do not change their status. The global volume for this sound is low as also in the reality such a sound is cushioned, and such a feature is held also for high values of the GRF controlling the system energy parameter.

Forest underbrush: three PhISM models working in parallel are used to synthesize the complex sound of footsteps on forest floor. The first model simulates the fall of the foot on a dirt floor, the second the trampling on the underbrush of the forest (e.g. leaves), and the third the breaking of little branches under the foot. The volumes of the three models change at every step thanks to random calculations, so their balance varies continuously during the walk. In particular a further control on the volume of the third model allow to simulate the random breaking of the branches, which does not happen at every step neither with the same intensity.
The GRF controls the system decay parameter of the first and second model, while the detection of the end of the sound produced by the algorithms activates new random calculations controlling the number of colliding objects for the three models and the system decay parameter for the third model. All the other parameters of the three models do not change their status once set.

Dry Leaves: the footstep sound on dry leaves sound is a combination of granular sounds with long duration both at low and high frequencies, and noticeable random sounds with not very high density that gives to the whole sound a crunchy aspect. These three components have been reproduced with as many PhISMs with the same characteristics of density, duration, frequency and number of colliding objects. The first and second models simulate two layers of leaves at high and low frequencies respectively, while the third model implements the crunchy sounds.
The GRF controls the system decay parameter of the first and second model, while the detection of the end of the sound produced by the algorithms activates a new random calculation for the control of the number of colliding objects of the third model.
All the other parameters of the three models do not change their status once set.

Dirt plus pebbles: three PhISM models working in parallel are used to synthesize the sound of footsteps on a country road with dirt and some pebbles. The first model simulates the fall of the foot on a dirt floor, while the second and the third the trampling on some pebbles of two different kinds.
The volumes of the second and the third models change at every step by means of random calculations, so the balance of the contribution to the sound of the three models varies continuously during the walk.
The GRF controls the system decay parameter of the three model, while the detection of the end of the sound produced by the algorithms activates new random calculations controlling the number of colliding objects for the three models. All the other parameters of the three models do not change their status once set.

High grass: three PhISM models working in parallel are used to synthesize the sound of footsteps on high grass (about 30 cm). The first and second models simulates the trampling on two different grass bundles, while the third adds some crunchy sounds. The detection of the end of the sound produced by the algorithms activates two new random calculations for the control of the number of colliding objects of the second model and the volume of the third model (not at every step the crunchy sounds are present).

3. CONTROLLING THE ENGINE

Three different systems have been developed and tested in order to control the engine in real time. All of them share the same technique to estimate the GRF from the input. The type of input used is a signal in the audio domain, and the setups differ principally for the typology of the audio input they provide.

The GRF has been calculated from the input signal extracting its amplitude envelope, i.e. finding the ground reaction force from the acoustic waveform. To perform envelope extraction we used a simple non-linear low-pass filter proposed by Cook in [14]:

$$e(n) = (1 - b(n))|x(n)| + b(n)e(n-1)$$

where

$$b = \begin{cases} b_{up} & \text{if } |x(n)| > e(n-1) \\ b_{down} & \text{otherwise} \end{cases}$$

Fig. 2 shows the envelope extracted form a recorded footstep sound on a concrete floor, the sub-events heel/toe that can be found within it.

3.1. The first system: floor microphones

In the first solution we adopted a set of four microphones placed on the floor.

In particular we used the Shure BETA 91[9], a high performance condenser microphone with a tailored frequency response designed specifically for kick drums and other bass instruments. Its features made it a good candidate for our purpose of capturing the footsteps sounds. The configuration we followed consisted of the placement of the microphones along the vertices of an ideal square on the floor at 1.5 meters distance from each others, delimiting the area inside which a subject could walk (see Figure 3).

[9] http://www.shure.com/

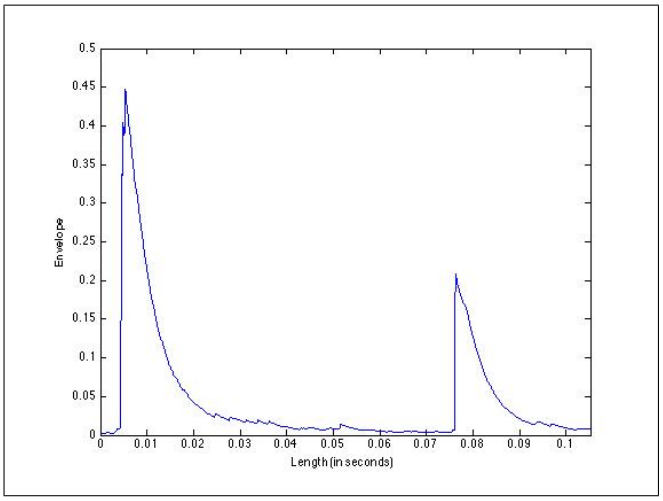


Figure 2: Amplitude envelope (GRF) extracted form a recorded footstep sound on a concrete floor.

The real footsteps sounds produced by a subject are detected by the microphone, and their GRF extracted and used to control the temporal evolution of the synthetic footsteps. The synthesized sounds are finally conveyed to the user by means of headphones.

3.2. The second system: tangible acoustics interfaces (TAI-CHI)

The second system consisted of a medium density fiberboard (MDF), 2.5 x 2 m in size and 1 cm thick. The board had four accelerometers (Knowles BU-21771) mounted on top, placed at the middle of each side of the board (Figure 4); such a placement of multiple sensors could eventually allow for tracking of position of where a footstep occurred [15]. Each accelerometer was connected to a dedicated preamplifier; the cables from all preamplifiers are collected in a small breakout box, which both provides power and routes the sensor signals further to soundcard inputs for A/D conversion. Standard stereo microphone (Ceam Cavi Li-YCY) cables were used as preamplifier cables. Each preamplifier consisted of a small power stabiliser based on a (National Semiconductor) 7805, as well as a simple non-inverting op-amp amplifier based on a STMicroelectronics TS921 rail-to-rail operational amplifier, advertised for audio range use.

The main intent behind the use of accelerometers, was to obtain an audio range signal of foot-surface interaction, while ideally suppressing the influence of all other acoustic sources of interference in the testing environment, by relying on the accelerometers capturing only the sound signal that had propagated through the board. The initial version of the TAI-CHI board consisted simply of the Knowles accelerometers mounted on the board, connected directly to microphone cables that were collected at the breakout box. The experience with the initial TAI-CHI version showed that the signal arriving at the soundcard was too noisy to be usable with the thresholding algorithms in use.

To remedy this, first a series of measurements was conducted, which determined that placement of the sensors on the top of the board gives better results than mounting them on the side. Power sources were limited to either a battery or lab power source to eliminate possible problems with mains hum, and higher power supply

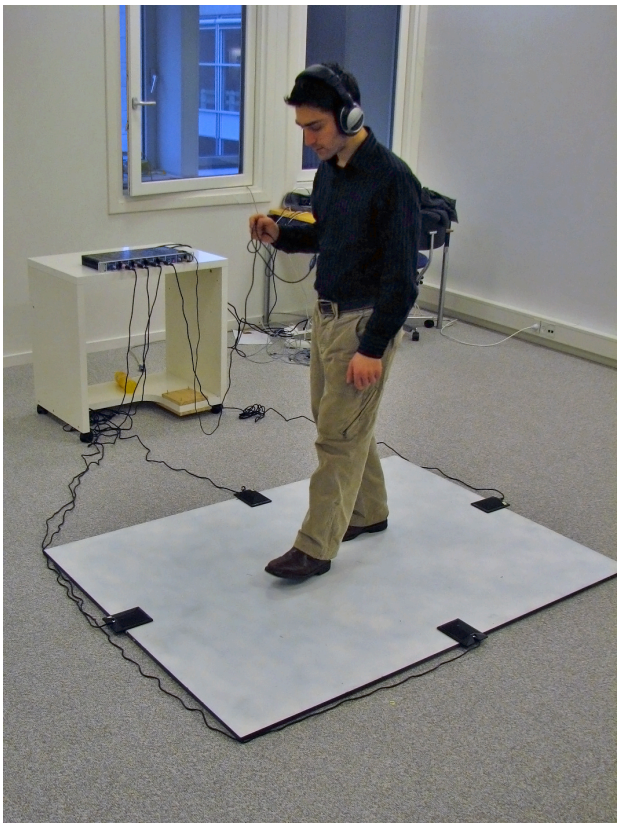


Figure 3: A subject using the first described system: floor microphones.

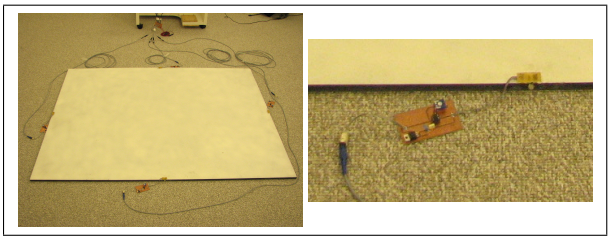


Figure 4: TAI-CHI board with accelerometer sensors (left), closeup of a single accelerometer sensor (mounted with Tack-All) and amplifier (right)

voltages were tried, in hope to increase SNR sufficiently for the relatively limited cable length (10m).

When all of that failed, a simple preamplifier was implemented, with the dual purpose of stabilizing the power supplied to the sensor, as well as buffering the sensor output signal before it goes on to the cable. In addition, different materials were tried as binding agents between materials: initially, Dantex Tack-All (a putty-like adhesive) was used for mounting, since it also allows for easy removal of the sensors. Subsequently, a thermal plastic that can melt in hot water marketed as Polymorph Plastic was tried as a binding agent - and was found to have lesser dampening effects on the sig-

nal than Tack-All, even though the plastic was used to encapsulate the sensors completely while mounting (and only small amounts of Tack-All were used for the same purpose, leaving most of the sensor uncovered).

Whereas the implementation of the preamplifiers and use of polymorph plastic for mounting, did indeed bring up the signal to levels usable for the threshold algorithm, at the same time, the increased sensitivity defeated the original intent of the use of accelerometers: to isolate the sound of foot-board interaction from other acoustic sources in the environment. In fact, in this state, the system was able to pick up hand claps, that could subsequently trigger the thresholding algorithms. Materials such as glass wool were then attempted as isolation materials placed over the accelerometers, but they tended to either dampen the signal too much, or not enough.

Due to the difficulties involved with finding, managing and fine-tuning all important parameters, until the system became usable with the thresholding algorithms in use, a different approach from using a TAI-CHI accelerometer-based system for acquisition of footsteps was attempted: the use of pressure sensitive resistors mounted in the shoes, in order to deliver a footstep trigger signal.

3.3. The third system: haptic shoes enhanced with pressure sensors

The third control solution consisted of a pair of light-weight sandals enhanced with FSR pressure sensors.[10] Two sensors were placed in correspondence to the heel and toe respectively in each shoe. Their aim was to detect the pressure force of the feet during the locomotion of a subject wearing the shoes. The analogue values of each of the four sensors were digitalized by means of an Arduino Diecimila board.[11]

The extrapolation of a useful GRF from the signals coming from the sensors turned out to be not the right choice for the control of the sound synthesis engine because of the features of the signal itself. For that reason we opted for a solution based on recorded GRF files and on a system of thresholds applied both on the signals and on their first derivatives. In particular we used the values of the first derivative as control for triggering, into the footsteps synthesizer, some GRFs corresponding to heel or toe according to the activated sensor.

While a subject walks there is a variation of the values of the pressure sensors in correspondence to each step. Such variation is the basis for obtaining first time derivatives of the sensors signals, which remain related to the intensity with which the foot hits the ground. Each time the value of the first derivative becomes bigger than a threshold, the GRF corresponding to the activated sensor is triggered into the engine. More precisely we checked only positive changes in the derivative value, since we were interested in the generation of the sound when the step hits (and not when it leaves) the ground. Other thresholds, both on the signals and on their first derivatives, were used in order to handle some boundary conditions, like the standing of the subject, with the aim of controlling the generation of sound. Such thresholds are set in a phase of calibration of the system, which has to take into account the different weights of the subjects wearing the shoes, in order to have an average value suitable for all the possible cases.

The GRFs triggered have been created by extracting the amplitude envelope from audio files of recorded footsteps on concrete, subdivided in the heel and toe parts. Five types of heel and toes audio files were used and randomly chosen at the moment of the triggering, giving rise to 25 possible combinations. Such behavior has been adopted in order to not have always the same GRF as input of the engine, and this allows to have differences in the generated sounds at every step, increasing thus the degree of realism of the walking experience.

Finally, the synthesized footstep sounds can be delivered to the subject both trough headphones and loudspeakers.

4. ADVANTAGES AND DISADVANTAGES OF THE SYSTEMS

In this section the three systems able to control interactively the sound engine during the act of walking of a subject are compared. These three control solutions share the same technique for the estimate of the GRF but differ for the type of audio input provided in real time. In the second system the input is the sound of real footsteps captured by microphones placed on the floor. The input in the third system is the signal corresponding to the vibrations generated by the footsteps and propagated along a board. In the fourth system the input is a set of prerecorded GRFs corresponding to the parts of the footsteps related to the heel and to the toe.

Hereinafter the three interactive systems are compared in terms of portability, easiness of setup, wearability, navigation, sound quality, sensing capabilities and integration in VR environments.

Portabilty. Both the first and third systems are easily portable. The second system consists of four microphones, a soundcard,[12] a laptop and a set of headphones. The third system consists of a pair of shoes enhanced with sensors, an Arduino board, soundcard, a laptop and a set of headphones. The second system consists of four piezoelectric sensors, an amplifier, a MDF, a soundcard, a laptop and a set of headphones; it is not easily portable because of the big dimensions of the MDF, as well as the quantity and the weight of the things to carry, and moreover the piezoelectric sensors are very delicate.

Easiness of setup. At hardware level the easiest system to setup is the first, since it is easier to place microphones on the ground rather than attaching the piezoelectric sensors on the MDF, or setup the shoes enhanced with sensors. At software level all the three systems require a phase of setup in which the global parameters and thresholds have to be calibrated. The first and second system require the microphone and the piezoelectric sensors respectively, to send data to the soundcard which sends it to the Max/MSP environment, while to setup the third system it is necessary to ensure that the Arduino is receiving data from the sensors and sending them to the Max/MSP environment.

Sound quality. The sound quality of the systems depends on the quality of the sound synthesis algorithms, the mapping between the sensors data and the algorithms as well as the audio delivery methods used. As concerns the quality of the synthesized sounds, good results in recognition tasks have been obtained in our previous studies[11, 12, 13].

As regards the audio delivery method, the sounds in the first system must be delivered through headphones. This is due to the fact that the surrounding sonic environment needs to be relatively

[10] I.E.E. SS-U-N-S-00039

[11] http://arduino.cc/

[12] For all the systems we used the Fireface 800 sound card, http://www.rme-audio.com/english/firewire/ff800.htm.

quiet, since the microphones should pick up only the real footsteps sounds. The second system shares the same requirement, since external noise put into vibration the MDF causing unwanted input signals. On the other hand, the third system is exempt from this problem and loudspeakers can be used in place of the headphones.

Wearability. The first and second hardware configuration allow users to wear their own footwear. On the other hand the shoe enhanced system requires users to wear a specific size of footwear, so it does not have the shoe independence requirement which characterizes the other two systems. Anyway the shoes developed are sandals comfortable and light.

In the first and second system the sound must be delivered through headphones, while in the third system it can be also delivered through loudspeakers, so the user is required to wear the designed shoes but not the headphones.

Navigation. The floor microphones and piezoelectric sensor require the user to navigate in a specific location delimited by the space inside the microphones or by the MDF dimension respectively, and in both the cases the synthesized sound must be delivered through headphones, whose wire connects them to the soundcard.

On the other hand, the area delimited by the shoe enhanced system is limited by both the length of the wires coming out from the shoes and the length of the wires connecting the headphones to the soundcard. To make more easy the navigation of the subjects, the shoes wires have been linked to a bumbag or to snaplinks attached to trousers.

Sensing capabilities. The use of the microphones on the ground turned out to be the best solution concerning the accuracy of the detection of the GRF corresponding to the movements of the subjects. On the other hand the approach with the MDF revealed that the piezoelectric sensors did not work at the same level of accuracy than the microphones, since some of the differences in the footsteps dynamics were not detected in high precision. The third system instead generates interactively the sounds not taking into account the exact step movement made by the subject, therefore losing the mapping concerning the dynamics suffers of a lack of realism.

As concerns the quality of the input signal for the sound engine, the floor microphones approach works for any type of indoor solid floor but the detection of the real footstep sounds noticeably get worse in presence of carpeted floors because the resulting sound is damped. Good results have been found also using as floor big pieces of cardboard fixed on the ground. The second system instead works only by means of the MDF, while the third system works independently from the floor on which the subject tramples on, and the input signal depends only on the triggering behaviour.

The latency problem is absent in all the systems, so the delivery of the sound to the user happens in real time with the movements of his/her feet.

Integration in VR environments. All the systems have been developed at software level as extension to the Max/MSP sound synthesis engine. The platform can be easily combined with several interfaces and different software packages. A protocol which has been shown to be suitable for integration purposes is the Open Sound Control protocol. [13]

The three systems can be integrated with visual feedback, to simulate different multimodal environments. The haptic integration can be provided only by means of the haptic shoes. The use of haptic shoes used revealed that the haptic integration does not work for the first and second systems because such shoes produce a noise while working; such a noise is detected by the microphones and piezoelectric sensors and constitutes a not negligible input error for the sound synthesis engine.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed different solutions to synthesize footsteps sounds in real-time. The solutions are based both on floor microphones and shoes enhanced with sensors. The shoes enhanced with sensors have the advantage of a higher number of sensing capabilities and do not require the environment to be acoustically isolated. On the other hand, they require users to be able to wear a particular size of shoes. The floor microphones have the advantage that users can wear their own footwear. However, they require a quiet environment to be used, and they have limited sensing capabilities. Overall, the three different interactive systems described showed to be suitable as a floor based interaction device to navigate virtual environments.

We are currently enhancing the systems with haptic and visual feedback, to simulate different multimodal environments.

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. [14]

7. REFERENCES

[1] P.R. Cook, "Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds," *Computer Music Journal*, vol. 21, no. 3, pp. 38–49, 1997.

[2] 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*, pp. 73–78, 2002.

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

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

[5] W.W. Gaver, "What in the world do we hear?: An ecological approach to auditory event perception," *Ecological Psychology*, vol. 5, no. 1, pp. 1–29, 1993.

[6] R.L. Klatzky, D.K. Pai, and E.P. Krotkov, "Perception of material from contact sounds," *Presence: Teleoperators & Virtual Environments*, vol. 9, no. 4, pp. 399–410, 2000.

[7] K. Van Den Doel, P.G. Kry, and D.K. Pai, "FoleyAutomatic: physically-based sound effects for interactive simulation and animation," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM New York, NY, USA, 2001, pp. 537–544.

[13] http://opensoundcontrol.org/

[14] www.niwproject.eu

[8] J.M. Adrien, *The missing link: Modal synthesis*, MIT Press Cambridge, MA, USA, 1991.

[9] F. Avanzini and D. Rocchesso, "Modeling collision sounds: Non-linear contact force," in *Proc. COST-G6 Conf. Digital Audio Effects (DAFx-01)*, 2001, pp. 61–66.

[10] P. Dupont, V. Hayward, B. Armstrong, and F. Altpeter, "Single state elastoplastic friction models," *IEEE Transactions on Automatic Control*, vol. 47, no. 5, pp. 787–792, 2002.

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

[12] L. Turchet, S. Serafin, and R. Nordahl, "Examining the role of context in the recognition of walking sounds," in *Proceedings of Sound and Music Computing Conference*, 2010.

[13] 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," in *Proceedings of Eurohaptics symposium on Haptic and Audio-Visual Stimuli: Enhancing Experiences and Interaction*, 2010.

[14] L. Peltola, C. Erkut, P.R. Cook, and V. Valimaki, "Synthesis of hand clapping sounds," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1021–1029, 2007.

[15] M. Fabiani, *Development of a tangible human-machine interface exploiting in-solid vibrational signals acquired by multiple sensors*.

Examining the role of context in the recognition of walking sounds

Luca Turchet
Medialogy Department
Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, DK
tur@media.aau.dk

Rolf Nordahl
Medialogy Department
Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, DK
rn@media.aau.dk

Stefania Serafin
Medialogy Department
Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, DK
sts@media.aau.dk

ABSTRACT

In this paper, we present an experiment whose goal was to recognize the role of contextual information in the recognition of environmental sounds.

Forty three subjects participated to a between-subjects experiment where they were asked to walk on a limited area in a laboratory, while the illusion of walking on different surfaces was simulated, with and without an accompanying soundscape. Results show that, in some conditions, adding a soundscape significantly improves surfaces' recognition.

1. INTRODUCTION

When exploring a place by walking, at least two categories of sounds can be identified: the persons own footsteps and the surrounding soundscape. In the movie industry, footsteps sounds represent important elements. Chion writes 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]. Studies on soundscape originated with the work of R. Murray Schafer [2]. Among other ideas, Schafer proposed soundwalks as empirical methods for identifying a soundscape for a specific location. In a soundwalk people are supposed to move in a specific location, noticing all the environmental sounds heard. Schafer claimed that each place has a soundmark, i.e., sounds which one identifies a place with. The idea of experiencing a place by listening has been recently further developed by Blesser and Salter [3]. By synthesizing technical, aesthetical and humanistic considerations, the authors describe the field of aural architecture and its importance in everyday life.

In the field of virtual reality, studies have recently shown how the addition of auditory cues could lead to measurable enhancement in the feeling of presence. Results are available on sound delivery methods [4, 5] or sound quality [6, 5]. Recently, the role of self-sound to enhance sense of presence in virtual environments has been investigated. By combining different kinds of auditory feedback consisting of interactive footsteps sounds created by ego-motion with static soundscapes, it was shown how motion in virtual reality is significantly enhanced when moving sound sources and ego-motion are rendered [7].

In [8, 9, 10], a system to synthesize in real-time the sound of footsteps on different materials was presented. The system was composed of a set of four contact microphones, a multichannel soundcard, a set of headphones and a laptop. The microphones detected real footsteps sounds from users, from which the ground reaction force (GRF) was estimated. Such GRF was used to control a sound synthesis engine based on physical models.

This interactive system was evaluated in a between-subjects experiment, where it was compared to a recognition task including recorded and synthesized offline sounds. Results showed that subjects were able to recognize most of the synthesized surfaces with high accuracy. Similar accuracy was also noticed in the recognition of real recorded footsteps sounds, which was an indication of the success of the proposed algorithms and their control [9].

In this paper, we are interested in understanding whether the addition of a soundscape enhances the recognition of the simulated surfaces. Our hypothesis is that context plays an important role in the recognition of the material a person is stepping upon. In order to test such hypothesis, we designed different soundscapes, described in the following section.

2. SOUNDSCAPE DESIGN

The soundscapes of the following environments were built:

1. A beach and seaside during the summer
2. A courtyard of a farm in the countryside
3. A ski slope
4. A forest
5. A park during the fall

Such soundscapes were designed according to the indications given by subjects answering to a questionnaire. Precisely, ten subjects, chosen among those not performing the experiment, were asked to imagine which sounds could occur in the above mentioned environments.

Subjects were asked the following question: "Imagine that you are right now in a forest: which sounds do you think you would hear?" In this particular environment,

Copyright: ©2010 Luca Turchet et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

subjects indicated sounds like trees, birds, different animals. Among the answers provided, we chose those which were stated by more than one subject, and collected a corresponding sound material using appropriate recordings of real sounds.

Such sounds were chosen among those available both on the Hollywood Edge sound effects library [1] and on the Freesound.org website. [2]
The chosen sounds were opportunely edited and assembled using the sound editor Adobe Audition 3. Soundscapes were designed with the goal of providing a clear idea of the designed environment already from the first seconds.

3. EXPERIMENT

We conducted an experiment whose goal was to investigate the ability of subjects to recognize the different walking sounds they were exposed to in three conditions: without soundscapes, with coherent soundscapes and with incoherent soundscapes.
The footsteps sounds provided during the three conditions were synthesized sounds generated in real time while subjects were walking using the interactive system described in the previous section. The soundscapes were audio files played in background independently from the subjects movements. The volumes of both footsteps and soundscapes were set by empirical investigation.

One of our hypotheses was that the recognition would have improved using coherent soundscapes rather than the conditions with no soundscapes and with incoherent soundscapes. Similarly we hypothesized higher evaluations in terms of realism and quality in presence of coherent soundscapes.

3.1 Methods

A between-subject experiment with the following three conditions was conducted:

1. Condition 1: footsteps sounds without soundscapes.
2. Condition 2: footsteps sounds with coherent soundscapes.
3. Condition 3: footsteps sounds with incoherent soundscapes.

Participants were exposed to 10 trials in conditions 1 and 2, and 12 trials in condition 3.

During conditions 1 and 2, 5 stimuli were presented twice in randomized order. The stimuli in condition 1 consisted of footsteps sounds on the following surfaces: beach sand, gravel, snow (in particular deep snow), forest underbrush (a forest floor composed by dirt, leaves and branches breaking), dry leaves. In condition 2 the stimuli consisted of the same footsteps sounds provided in condition 1 with in addition the corresponding coherent soundscape mentioned in section 2.

[1] www.hollywoodedge.com/
[2] www.freesound.org/

During condition 3, 6 stimuli were presented twice in randomized order. The stimuli consisted of footsteps sounds on the surfaces beach sand, snow, forest underbrush, with in addition an incoherent soundscape. As an example in presence of the footstep sound on beach sand the provided soundscapes corresponded to those of footstep sounds on snow (i.e., the ski slope) and on forest underbrush (i.e., the forest environment).

3.1.1 Participants

Forty three participants were divided in three groups to perform the three conditions in a between-subjects experiments (n = 15, n = 15 and n = 13 respectively). The three groups were composed respectively of 11 men and 4 women, aged between 21 and 28 (mean = 23.67, standard deviation = 2.12), 8 men and 7 women, aged between 19 and 38 (mean = 24.67, standard deviation = 5.97), and 6 men and 7 women, aged between 21 and 30 (mean = 24, standard deviation = 3.1). All participants reported normal hearing conditions. All participants were naive with respect to the experimental setup and to the purpose of the experiment.

During the experiment the shoes used by subjects were sneakers, trainers, boots and other kinds of shoes with rubber soil.
The participants took on average about 11, 13 and 16 minutes for conditions 1, 2 and 3 respectively.

3.1.2 Setup

The experiment was carried out in an acoustically isolated laboratory where the setup was installed (see Fig. 2). Participants were asked to walk inside an area delimited by four microphones placed in a square configuration on a medium density fiberboard (MDF). [3] Specifically, we used four Shure BETA 91, [4] high performance condenser microphones with a tailored frequency response designed specifically for kick drums and other bass instruments. The microphones' features made them a good candidate for the purpose of capturing footsteps sounds.
The MDF was adopted in place of the carpeted floor of the laboratory in order to improve the quality of the input signal.

The floor microphones were connected to a soundcard, [5] which in turn was connected to a laptop running the sound synthesis engine. Finally the synthesized sounds, as well as the soundscapes, were provided to the user by means of a set of headphones. [6]

3.1.3 Task

During the experiment the participants were asked to wear a pair of headphones and to walk on the MDF in the area delimited by the microphones. They were given a list of different surfaces to be held in one hand, presented as non-forced alternate choice. Such list included a range of materials wider than those presented in experiment. During

[3] 2.5 x 2 m in size and 1 cm thick.
[4] http://www.shure.com/
[5] We used the Fireface 800 sound card, http://www.rme-audio.com/english/firewire/ff800.htm.
[6] Beyerdynamic DT-770, http://www.beyerdynamic.de/

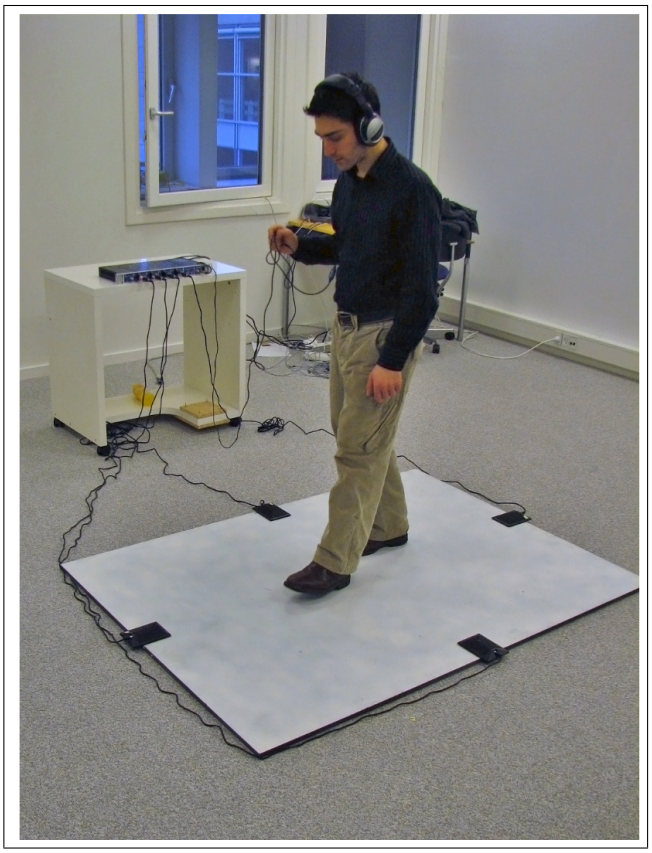


Figure 2. A subject using the interactive footsteps synthesizer. The four contact microphones are clearly noticeable.

the act of walking they listened simultaneously to footsteps sounds on a different surface according to the stimuli presented. The task, common to the three conditions, consisted of answering by voice the following three questions after the presentation of the stimulus:

1. Which surface do you think you are walking on? For each stimulus choose an answer in the following list: 1) beach sand, 2) gravel, 3) dirt, 4) snow, 5) high grass, 6) forest underbrush, 7) dry leaves, 8) wood, 9) creaking wood, 10) metal, 11) carpet, 12) concrete, 13) frozen snow, 14) puddles, 15) water, 16) I don't know.

2. How close to real life is the sound in comparison with the surface you think it is? Evaluate the degree of realism on a scale from 1 to 7 (1 = low realism, 7 = high realism).

3. Evaluate the quality of the sound on a scale from 1 to 7 (1 = low quality, 7 = high quality).

In condition 2 and 3, participants were also asked to recognize what was the environment they were walking.
They were informed that they could choose the same material more than once and that they were not forced to choose all the materials in the list. In addition they could use the interactive system as much as they wanted before giving an answer. When passed to the next stimulus they could not change the answer to the previous stimuli.
At the conclusion of the experiment, participants were also given the opportunity to leave an open comment on their experience interacting with the system.

3.2 Results

The collected answers were analyzed and compared between the three conditions. Results are shown in tables 1, 2 and 3.

The first noticeable element emerging from the three tables is that the use of the interactive system in the condition of coherent soundscapes gave rise to a better recognition of the surfaces and a higher evaluation of realism and quality of the proposed sounds, rather than the conditions with no soundscapes and with incoherent soundscapes. Concerning the percentages of correct answers, they are higher for condition 2 compared to condition 1, for each surface, and the analysis by means of a chi-square test reveals that such differences are statistically significant for beach sand ($p = 0.005515$) and forest underbrush ($p = 0.01904$).
It is particularly interesting to notice that overall adding a soundscape enhances the recognition factor, and this is especially noticeable for those situations where the recognition was rather low without a soundscape.

Similarly the percentages of correct answers are higher for condition 2 compared to condition 3, for each surface, in particular the differences are statistically significant for beach sand ($p = 6.232e-07$), snow ($p = 0.01439$) and forest underbrush ($p = 0.001637$).
Furthermore, the percentages of correct answers are higher for condition 1 compared to condition 3, for each surface, but the differences are not statistically significant.

The analysis of the wrong answers reveals that in all the experiments none of the presented aggregate surfaces was recognized as a solid surface. This means that all subjects were able to identify at least the nature of the surface, which was an expected feature of the simulations. An observation from the subjects performing the experiment was that many subjects perceived the simulated sounds as very similar, and therefore hard to recognize and distinguish from the list provided.

It is interesting to examine what happens when subjects are exposed to soundscapes which are incoherent, as shown in Table 3. In this situation, we consider as correct the answer provided when subjects recognize the surface they are walking on, and not the soundscape. As it can be noticed, the percentage of correct answers is quite low. As expected, adding an incoherent soundscape creates a stronger context which often confuses the subjects. This can be observed, for example, in the case of beach sand footsteps which were rendered together with a forest soundscape and a ski slope soundscape. The recognition rate is higher in the first case than in the second, where several subjects confused sand with snow. The subjects' answers for the three conditions are outlined in the confusion matrices shown in Table 4, 5, 6 respectively. Such matrices show information concerning actual classifications performed by the subjects. From the matrices, it can be noticed how the subjects' recognition varies from condi-

tion 1 to condition 2. As an example, the first row of the matrix illustrated the number of subjects which recognized the beach sand surface, with (Table 4) and without (Table 5) a soundscape. The role of the soundscape to enhance the recognition is clearly noticeable.

Table 6 illustrates the confusion matrix for condition 3, i.e., when incoherent soundscapes are presented to the subjects. In this situation, it is clearly noticeable how the nature of the soundscape plays an important role. Moreover, it can be noticed how the incoherent soundscape is in most situation predominant, in the sense that subjects tend to judge the surface they are stepping upon more listening to the soundscape than listening to the actual surface. On the other hand, even if subjects are not able to recognize the surface they are stepping upon, they never confuse its nature, in the sense that they never select a solid surface when exposed to an aggregate one.

In addition, Tables 1, 2 and 3 show the degree to which participants judged the realism and quality of the experience. The degree of realism was calculated by looking only at that data from correct answers, i.e., when the surfaces were correctly recognized. This choice was performed since we were interested in understanding whether the simulation of specific surfaces recognized by the subjects was satisfactory.

As far as the quality judgement is concerned, the data was based on all the answers different from "I don't know".

The mean of realism is higher for condition 2 compared to condition 1 for each surface with the exception of beach sand (which is almost equal). The analysis by means of a t-test reveals that such differences are statistically significant for snow (p = 0.01055) and forest underbrush (p = 0.005595).
Analogously, the mean of realism is higher for condition 2 compared to condition 3 for each surface with the exception of beach sand (which is almost equal). In particular, the differences are statistically significant for beach sand (p = 0.002568) and snow (p = 0.001938).
Moreover, the mean of realism is higher for condition 1 compared to condition 3 for each surface with the exception of forest underbrush, which is minor. Such differences are statistically significant for beach sand (p = 0.001302), and for forest underbrush (p = 0.03438), which, as said, is greater for experiment 3.

As regards the mean of quality, it is higher for condition 2 compared to condition 1, with statistically significant differences for all the surfaces with the exception of dry leaves: beach sand (p = 0.009619), gravel (p = 0.02169), snow (p = 0.0006874) and forest underbrush (p = 0.02198).
The mean of quality is higher for condition 2 compared to condition 3 for each surface, and in particular the differences are statistically significant for beach sand (p = 0.006187), for snow (p = 9.596e-05).
Furthermore, the mean of quality is similar for condition 1 compared to condition 3, with the exception of forest underbrush for which it is higher in condition 3 compared to condition 1, with statistically significant differences (p = 0.03204).

	% Correct answers	% Wrong answers	% "I don't know"	Realism	Quality
Beach Sand	50.	46.67	3.33	5.2	4.7241
Gravel	83.33	6.67	10.	5.2	4.6296
Snow	73.33	26.67	0.	5.2955	5.1167
Forest Underbrush	40.	50.	10.	3.5	4.1923
Dry Leaves	16.67	63.33	20.	4.4	3.9167

Table 1. Results of condition 1: recognition of the surfaces without soundscapes.

	% Correct answers	%Wrong answers	% "I don't know"	Realism	Quality	% Correct soundscape
Beach Sand	86.67	10.	3.33	5.1481	5.5172	93.33
Gravel	86.67	13.33	0.	5.3077	5.4	86.67
Snow	80.	20.	0.	6.1667	6.0667	83.33
Forest Underbrush	73.33	26.67	0.	4.9091	5.0333	100.
Dry Leaves	30.	70.	0.	4.4444	4.5	96.67

Table 2. Results of condition 2: recognition of the surfaces with coherent soundscapes.

The comparison about the percentages of "I don't know" answers reveals that for each surface they are higher for condition 1 compared to condition 2, and for condition 3 compared to condition 2. In addition, they are higher for condition 3 compared to condition 1, for each surface with the exception of forest underbrush (which is minor).

As regards the percentages of correct answers about the soundscapes presented, they are higher for condition 2 compared to condition 3, and in particular the differences are statistically significant for the ski slope soundscape (p = 0.0003945).

Overall, subjects observed that soundscapes play an important role in recognition of the surfaces, precisely for their ability to create a context. Especially in terms of conflicting cues, as it was the case in condition 3, subjects were trying to identify the strongest cues, i.e. the element which had the strongest recognition factor. Sometimes the subjects found this task quite hard to complete, and this is why the percentage of "I don't know" answers is higher in condition 3 as opposed to condition 2.

When leaving a comment, several subjects observed that the recognition of snow was extremely realistic. This observation is also confirmed by the high degree of realism (mean = 5.3) and quality (mean = 5.1) with which such surface was rated.

On the other hand, for some subjects the concept of dry leaves was rather confusing, and this is also confirmed by the low recognition rate of such surface.

Overall, this experiment represents a strong indication

Material	Soundscape	% Correct answers	% Wrong answers	% No idea	Realism	Quality	% Correct soundscape
Beach Sand	Forest	38.46	57.7	3.84	4	5.16	88.46
Beach Sand	Ski slope	15.38	69.24	15.38	3.75	4.3182	38.46
Snow	Forest	50	42.31	7.69	5.2857	5.3542	96.15
Snow	Beach	50	46.16	3.84	4.7692	5	88.46
Forest Underbrush	Beach	38.46	53.85	7.69	4.8	4.9583	65.38
Forest Underbrush	Ski slope	30.76	65.4	3.84	4.3	4.6087	46.15

Table 3. Results of condition 3: recognition of the surfaces with incoherent soundscapes.

	BS	GL	SW	UB	DL	HG	DR	FS	WD	CW	MT	CC	PD	WT	CP	—
BS	15	2	6	2	2	2										1
GL		25			1								1			3
SW			22	1	1			6								
UB			4	12			1	10								3
DL		9		4	5			6								6

Legend:

WD	wood	CW	creaking wood	SW	snow	UB	underbrush
—	don't know	FS	Frozen snow	BS	beach sand	GL	Gravel
MT	metal	HG	High grass	DL	dry leaves	CC	concrete
DR	dirt	PD	puddles	WT	Water	CP	carpet

Table 4. Confusion matrix of condition 1.

	BS	GL	SW	UB	DL	HG	DR	FS	WD	CW	MT	CC	PD	WT	CP	—
BS	26	1	1	1												1
GL		26			3	1										
SW	2		24					4								
UB		1	2	22	1			4								
DL	6			9	9			6								

Legend:

WD	wood	CW	creaking wood	SW	snow	UB	underbrush
—	don't know	FS	Frozen snow	BS	beach sand	GL	Gravel
MT	metal	HG	High grass	DL	dry leaves	CC	concrete
DR	dirt	PD	puddles	WT	Water	CP	carpet

Table 5. Confusion matrix of condition 2.

	Soundscape	BS	GL	SW	UB	DL	HG	DR	FS	WD	CW	MT	CC	PD	WT	CP	—
BS	Forest	10		2	5	3	1	1	3								1
BS	Ski slope	4	1	5		1	5	3	1					1	1		4
SW	Forest		5	13	1	1			4								2
SW	Beach	2	6	13	2				2								1
UB	Beach		3	4	10	2			5								2
UB	Ski slope		1	1	8	3	1	2	8							1	1

Legend:

WD	wood	CW	creaking wood	SW	snow	UB	underbrush
—	don't know	FS	Frozen snow	BS	beach sand	GL	Gravel
MT	metal	HG	High grass	DL	dry leaves	CC	concrete
DR	dirt	PD	puddles	WT	Water	CP	carpet

Table 6. Confusion matrix of condition 3.

of the importance of context in the recognition of a virtual auditory place, where self sounds created by users' footsteps and soundscapes are combined. Further investigations are needed to enhance the realism of the simulated soundscape, in particular by having the auditory cues changing according to the motion of the subject in the space.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we describe an experiment whose goal is to understand the role of soundscapes in creating a sense of place and context when designing a virtual walking experience. In this particular experiment, the user was not able to interact with the soundscapes, which were made of mere soundtracks. The results described are an interesting starting point for further investigations on the role of environmental sounds to create a sense of place. While walking an acting in an environment, a person is exposed to her own self-sounds as well as the sounds of the place. This paper presents a preliminary investigation of the role of these different elements both taken in isolation and combined. Further investigations are needed to gain a better understanding of the cognitive factors involved when subjects are exposed to different sound events, especially when a situation of semantic incongruence is present.

We are also planning to design the soundscapes in a multichannel environment, where moving sound sources are present, and the location of the different sound events depends on the location of the subjects. We are also planning to enhance the simulations with visual and haptic feedback. This will allow us to investigate in depth the role of sound to create a sense of place in unimodal and multimodal environments.

5. 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. [7]

6. REFERENCES

[1] M. Chion, C. Gorbman, and W. Murch, *Audio-vision: sound on screen*. Columbia Univ Pr, 1994.

[2] R. Schafer, *The tuning of the world*. Alfred A. Knopf, 1977.

[3] B. Blesser and L. Salter, *Spaces speak, are you listening?: experiencing aural architecture*. MIT Press, 2006.

[4] R. Storms and M. Zyda, "Interactions in perceived quality of auditory-visual displays," *Presence: Teleoperators & Virtual Environments*, vol. 9, no. 6, pp. 557–580, 2000.

[5] R. Sanders Jr, *The effect of sound delivery methods on a users sense of presence in a virtual environment*. PhD thesis, NAVAL POSTGRADUATE SCHOOL, 2002.

[7] www.niwproject.eu

[6] P. Chueng and P. Marsden, "Designing Auditory Spaces to Support Sense of Place: The Role of Expectation," in *CSCW Workshop: The Role of Place in Shaping Virtual Community*, Citeseer, 2002.

[7] R. Nordahl, "Increasing the motion of users in photorealistic virtual environments by utilizing auditory rendering of the environment and ego-motion," *Proceedings of Presence*, pp. 57–62, 2006.

[8] L. Turchet, S. Serafin, S. Dimitrov, and R. Nordahl, "Physically based sound synthesis and control of footsteps sounds," in *Proceedings of Digital Audio Effects Conference*, 2010.

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

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

Interaction Capture in Immersive Virtual Environments via an Intelligent Floor Surface

Yon Visell, Alvin Law, Jessica Ip, Severin Smith, Jeremy R. Cooperstock
McGill University, Centre for Intelligent Machines and CIRMMT, Montreal, Canada

ABSTRACT

We present techniques to enable users to interact on foot with simulated natural ground surfaces, such as soil or ice, in immersive virtual environments. Position and force estimates from in-floor force sensors are used to synthesize plausible auditory and vibrotactile feedback in response. Relevant rendering techniques are discussed in the context of walking on a virtual frozen pond.

1 INTRODUCTION

Sensations accompanying walking on natural ground surfaces in real world environments (sand in the desert, or snow in winter) are multimodal and highly evocative of the settings in which they occur [8]. Limited prior research has addressed foot-based interaction with VR and AR environments. Arguably, one reason has been the lack of efficient techniques for capturing foot-floor contact interactions via distributed floor areas. Here, we present a novel solution using a network of instrumented floor tiles, and methods for capturing foot-floor contact interactions so as to render multimodal responses of virtual ground materials.

Devices for enabling omnidirectional in-place locomotion in VEs exist [2], but are complex and costly. Most prior work on tactile interaction with floor surfaces utilizes surface sensing arrays [4, 6] for applications such as person tracking, activity tracking, or musical performance. Commercial entities such as Gesturetek and Reactrix have developed floor interfaces for dynamic visual displays of liquid or solid objects. In such cases, video sensing is often used, but it provides no direct information about contact forces. Such information is needed for simulating highly contact-dependent interactions with virtual materials. Further references on this topic are available in a related publication [9].

2 DISTRIBUTED FLOOR INTERFACE

The floor interface consists of a 6×6 array of rigid tiles covering a 3.5 sq. m area. Each tile is instrumented with four force sensors and a vibrotactile (VT) actuator. The floor is coated in projection paint, and a pair of overhead video projectors is used for visual display. The design of the actuated tile interface is described in detail elsewhere [7, 9]. It consists of a rigid plate $30.5 \times 30.5 \times 2$ cm in size, supported on vibration mounts, and coupled to a voice coil actuator. Actuators are driven by audio signals that are generated by a computer and amplified. The device has a passband from approximately 50 to 750 Hz, and can reproduce the largest forces required for interaction with virtual ground surfaces (i.e., more than 30 N across the passband). We sense normal forces below the vibration mounts using resistive force sensors (Interlink model 402 FSR). This data is conditioned, amplified, and digitized via a 32-channel, 16-bit acquisition board, and relayed over Ethernet. An array of six small form factor computers is used for force data processing and audio-vibrotactile (VT) synthesis. Each is responsible for synthesizing VT feedback from six tiles, ensuring a palpable response to force input can be provided at low latencies. Data is forwarded to a separate visual rendering server.

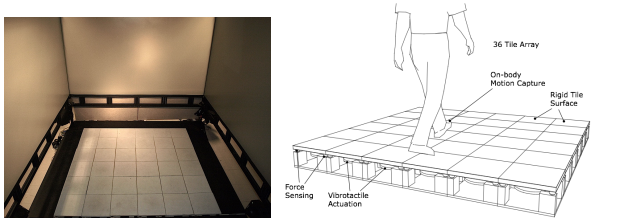


Figure 1: Left: Distributed floor interface situated within an immersive, rear projected VE simulator. Right: illustration showing sensing and actuating components.

3 INTRINSIC CONTACT SENSING

Intrinsic contact sensing resolves locations and forces of surface contact between objects using internal force and torque measurements [1], and can provide contact position estimates with fewer sensors than are needed for surface-based sensing. This method has previously been applied to problems in robotic manipulation. The idea is to represent a foot pressure distribution $p_R(\mathbf{x})$ by a point $\mathbf{x}_c$, known as the contact centroid, such that a normal force F_c that gives rise to the same intrinsic force measurements as $p_R(\mathbf{x})$ does [1]. For a single floor tile (Fig. 2), with force sensor locations $\mathbf{x}_j$ where internal force measurements f_j are taken, $\mathbf{x}_c$ and the force $\mathbf{F}_c = (0,0,F_c)$ can be recovered from the measurements $\mathbf{F}_j = (0,0,f_j)$ by solving the force and torque equilibrium equations,

$$\sum_{j=1}^{4} f_j + F_c + f_p = 0 \qquad (1)$$

$$\sum_{j=1}^{4} \mathbf{x}_j \times \mathbf{F}_j + \mathbf{x}_c \times \mathbf{F}_c + \mathbf{x}_p \times \mathbf{F}_p = 0\,. \qquad (2)$$

$\mathbf{F}_p = (0,0,f_p)$ is the weight of the the plate and actuator at the tile's center $\mathbf{x}_p$. Solving the three nontrivial scalar equalities (1,2) yields:

$$F_c = \sum_{i=1}^{4} f_i - f_p, \quad \mathbf{x}_c = \frac{1}{F_c}\left(\sum_{i=1}^{4} (\mathbf{x}_i - \mathbf{x}_p) f_i + f_c \mathbf{x}_p\right) \qquad (3)$$

This result can be readily extended to cases in which $p_R(\mathbf{x})$ overlaps several tiles [9]. Our device was empirically determined to be able to localize contacts with a typical accuracy of 2 cm, a number that compares favorably to the dimensions of an adult shoe. Further details are given in [9].

4 CHANNELING MATERIAL INTERACTIONS

We have implemented a demonstration consisting of a virtual frozen pond that users may walk on, producing patterns of surface cracks that are rendered via audio, visual, and VT channels (Fig. 3).

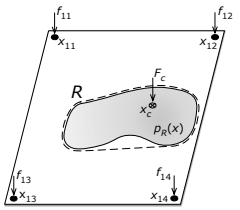


Figure 2: Contact centroid $\mathbf{x}_c$ from pressure distribution $p_R(\mathbf{x})$.

4.1 Audio-tactile rendering

Audio and VT display channels provide feedback accompanying the fracture of the virtual ice sheet underfoot. The two are derived from a local fracture mechanics model. Fracture events are described via an event time t_i and energy loss E_i. Figure 4 illustrates the continuum model and a simple mechanical analog used for synthesis. In the stuck state, the surface has stiffness $K = k_1 + k_2$ and is governed by:

$$F(t) = m\ddot{x} + b\dot{x} + K(x - x_0), \quad x_0 = k_2\xi(t)/K \tag{4}$$

where $\xi(t)$ represents the net plastic displacement up to time t. A Mohr-Coulomb yield criterion is applied to determine slip onset: When the force F_ξ on the plastic unit exceeds a threshold F_0 (either constant or stochastic), a slip event with incremental displacement $\Delta\xi(t)$, and energy loss ΔW, representing the inelastic work of fracture growth, is generated. Slip displacements are rendered as impulsive transients, by exciting a bank of modal oscillators with impulse response $s(t) = \sum_i a_i e^{-b_i t}\sin(2\pi f_i t)$, determined by amplitudes a_i, decay rates b_i, and frequencies f_i. An independent response is rendered in parallel for each tile.

4.2 Visual animation and control

Surface cracks are often animated by simulating the inelastic evolution of a distributed stress state [5, 3]. Here, we adopt a simpler simulation fusing the local temporal crack-growth model given above with heuristics for spatial pattern growth. This efficient method allows us to the stringent real-time requirements for audio-VT feedback. The contact centroid $\mathbf{x}_c$ provides a local summary of the spatial stress injected by the foot. A fracture pattern consists of a collection of crack fronts, defined by linear sequences of node positions, $\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_n$. Fronts originate at seed locations $\mathbf{p} = \mathbf{c}_0$. The fracture is rendered as line primitives $\ell_k = (\mathbf{c}_k - \mathbf{c}_{k-1})$ on the ice sheet (Fig. 5). Seed locations $\mathbf{p}$ are determined by foot-floor contact. Our method is mesh-free, and the seed locations are unconstrained. A crack event initiated by the audio-tactile process at time t_i with energy $E(t_i)$ results in the creation of a new seed or the growth of fractures from an existing one. A new seed $\mathbf{p}$ is formed at the location of the dominant contact centroid $\mathbf{x}_c$ if no existing seed lies within distance Δp. The new seed $\mathbf{p}$ is created with a random number N_c of latent crack fronts, $\mathbf{c}_0^1, \mathbf{c}_0^2, \ldots \mathbf{c}_0^{N_c}$. We sample N_c uniformly in $2, 3, \ldots 6$. A crack front propagates from a seed $\mathbf{p}$ nearest to $\mathbf{x}_c$. With probability $1/N_c$ the jth crack front of $\mathbf{p}$ is extended. Its growth is determined by a propagation vector $\mathbf{d}_m^j$ such that $\mathbf{c}_m^j = \mathbf{c}_{m-1}^j + \mathbf{d}_m^j$. We take $\mathbf{d}_m^j = \alpha E \hat{\mathbf{n}}_m^j$, where E is the crack energy, α is a global growth rate parameter, and $\hat{\mathbf{n}}_m^j$ is the direction. Since we lack information about the principal stress directions at the front, we propagate in a direction $\hat{\mathbf{n}}_m^j = \hat{\mathbf{n}}_{m-1}^j + \beta\hat{\mathbf{t}}$, where $\beta \sim N(\beta; 0, \sigma)$ is a Gaussian random variable and $\hat{\mathbf{t}} = \hat{\mathbf{n}}^j \times \hat{\mathbf{u}}$, where $\mathbf{u}$ is the upward surface normal (i.e., $\mathbf{t}$ is a unit vector tangent to $\hat{\mathbf{n}}^j$). The initial directions at $\mathbf{p}$ are spaced equally on the circle.

5 Conclusion

This paper described techniques for interaction with virtual ground material simulations using a distributed, multimodal floor interface. The methods are low in cost and complexity, and accessible to multiple users without body-worn markers or equipment. Despite these promising results, there are several aspects in which this system can be improved, including: tile sensing accuracy, tile array density, the use of multi-tile force data, and more accurate physical simulation of ground materials. Nonetheless, we believe that these methods hold promise for improving presence in immersive VR and AR environments.

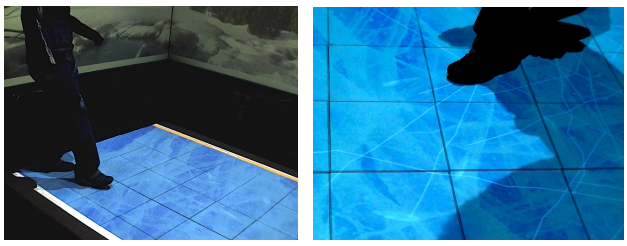


Figure 3: Still images from the frozen pond scenario.

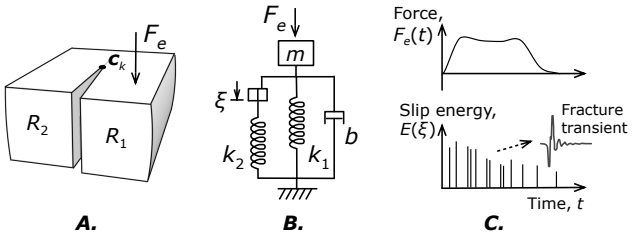


Figure 4: *A.* Behavior at the crack front $\mathbf{c}_k$ is modeled via a fracture mechanics approach. A visco-elasto-plastic body undergoes shear sliding fracture. *B.* A simple mechanical analog. *C.* Each slip event is rendered as an impulsive transient.

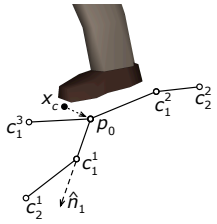


Figure 5: Crack patterns, graphs of lines between nodes, $\mathbf{c}_i$.

Acknowledgements

The authors acknowledge support from the MDEIE of Quebec for the EU FP7 project NIW (ICT-FET no. 222107).

References

[1] A. Bicchi, J. Salisbury, and D. Brock. Contact sensing from force measurements. *The International Journal of Robotics Research*, 12(3):249, 1993.

[2] J. Hollerbach. Locomotion interfaces and rendering. In M. Lin and M. Otaduy, editors, *Haptic Rendering: Foundations, Algorithms and Applications*. A K Peters, Ltd, 2008.

[3] J. O'Brien and J. Hodgins. Graphical modeling and animation of brittle fracture. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 137–146. ACM Press/Addison-Wesley Publishing Co. New York, NY, USA, 1999.

[4] J. Paradiso, C. Abler, K.-y. Hsiao, and M. Reynolds. The magic carpet: physical sensing for immersive environments. In *CHI '97: CHI '97 extended abstracts on Human factors in computing systems*, pages 277–278, New York, NY, USA, 1997. ACM.

[5] M. Pauly, R. Keiser, B. Adams, P. Dutré, M. Gross, and L. Guibas. Meshless animation of fracturing solids. *Proceedings of ACM Siggraph 2005*, 24(3):957–964, 2005.

[6] A. Schmidt, M. Strohbach, K. v. Laerhoven, A. Friday, and H.-W. Gellersen. Context acquisition based on load sensing. In *UbiComp '02: Proceedings of the 4th international conference on Ubiquitous Computing*, pages 333–350, London, UK, 2002. Springer-Verlag.

[7] Y. Visell and J. Cooperstock. Optimized Design of a Vibrotactile Display via a Rigid Surface. In *Proc. of IEEE Haptics Symposium*, 2010.

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

[9] Y. Visell, A. Law, S. Smith, J. Ip, R. Rajalingham, and J. Cooperstock. Contact sensing and interaction techniques for a distributed multimodal floor display. In *Proc. of IEEE 3DUI*, 2010.

Design of a Vibrotactile Display via a Rigid Surface

Yon Visell and Jeremy R. Cooperstock
McGill University, Centre for Intelligent Machines and CIRMMT, Montreal, Canada

ABSTRACT

This paper describes the analysis, optimized redesign and evaluation of a high fidelity vibrotactile interface integrated in a rigid surface. The main application of the embodiment described here is vibrotactile display of virtual ground surface material properties for immersive environments, although the design principles are general. The device consists of a light, composite plate mounted on an elastic suspension, with integrated force sensors. It is actuated by a single voice coil motor. The structural dynamics of the device were optimized, within constraints imposed by the requirements of user interaction, and corrected via digital inverse filtering, in order to enable accurate simulation of virtual ground materials. Measurements of the resulting display demonstrate that it is capable of accurately reproducing forces of more than 40 N across a usable frequency band from 50 Hz to 750 Hz.

Keywords: Vibrotactile display design, Vibrotactile rendering, Foot interfaces

1 INTRODUCTION

Vibrotactile display devices consist of palpable interfaces that are capable of vibrating at frequencies salient to human tactile perception, but are not necessarily capable of static force display [35]. Advantages of such displays include their low cost, power efficiency, and capability of rendering transient or textural effects accurately, at high temporal resolution, when suitable actuators are used.

1.1 Vibrotactile augmentation of floor surfaces

The work presented here concerns the design of a vibrotactile display device integrated in a rigid panel. The specific embodiment involved is an augmented floor panel designed to enable vibrotactile interaction with ground surfaces. Potential applications of such a device include the simulation of ground textures for virtual and augmented reality simulation [39] or telepresence (e.g., for remote planetary simulation), the rendering of abstract effects or other ecological cues for rehabilitation, or the presentation of tactile feedback to accompany the operation of virtual foot controls, control surfaces, or other interfaces [40]. This device constitutes a redesign and optimization of an interface that was introduced in earlier work by the authors [36]. The goal was to systematically improve its fidelity. One motivation for doing so is to avoid artifacts that can affect realism and prevent careful control of stimuli in experiments involving human vibrotactile perception. The main artifacts alleviated by the redesign are due to selective variations in the device frequency response within the range of human vibrotactile sensitivity. We have also attempted to provide enough documentation that the device, or others like it, may be produced by researchers or practitioners interested in vibrotactile display via rigid surfaces. The sections that follow present the device concept, its mechanical and electronic structure, the optimizations that were undertaken to improve its dynamic response, and an analysis of the results.

1.2 Background

The fidelity of a haptic device depends on the selection and arrangement of electronic components, and on its structural design, as has long been emphasized in the research literature on force-feedback displays. Many of the analogous questions for the design of vibrotactile (VT) displays have received less attention. The vibrotactile augmentation of touch surfaces has been widely investigated for HCI applications [24, 11, 23], although design issues affecting their perceptual transparency have often been neglected. A few examples demonstrating optimized vibrotactile device engineering [41] or software compensation of their dynamics [18] do, nonetheless, exist. Issues related to sensor and actuator selection are covered in recent reviews of haptic hardware prototyping [13], while the extensive literature on acoustic and vibration engineering provide the knowledge needed for structural dynamics optimization [22, 7, 10].

We are not aware of any prior work on the systematic design of vibrotactile floor panels for haptic interaction. Passive floor-based vibrotactile actuation has been used to present low frequency information in audiovisual display applications, for special effects (e.g., vehicle rumble), in immersive cinema or VR settings [33]. The fidelity requirements that must be met by an *interactive* haptic display are, all things being otherwise equal, higher, since its users are able to actively sample its response to actions of the feet.

Many designs for vibrotactile-augmented shoes or insoles have been proposed [19, 27, 25], but nearly all have been designed for symbolic information display (e.g., directional indicators or encoded messages) or biofeedback stimuli, rather than for presenting virtual ecological stimuli. State of the art haptic force feedback interfaces for walking [14] are capable of preserving kinesthetic degrees of freedom but are restricted to a small fraction of the bandwidth that our device achieves. As noted below, a wide bandwidth interface is advantageous for reproducing the rich variety of high-frequency force information that is generated during walking in natural environments [28].

2 DISPLAY CONCEPT AND MOTIVATION

The interface of the device (Fig. 5) consists of a rigid plate that supplies vibrations in response to forces supplied by a user's foot, via the shoe. The total normal force $F(t)$ applied to the plate by a user is measured. It can be assumed to consist of two components: isolated transients with high frequency content, generated by foot impacts with the plate, and low-frequency forces generated by active human motions, limited in bandwidth to no more than 10 Hz [3, 37]. Vibrotactile feedback is assumed to be constrained, due to actuator limitations, to frequencies greater than a minimum value on the order of 20 Hz.

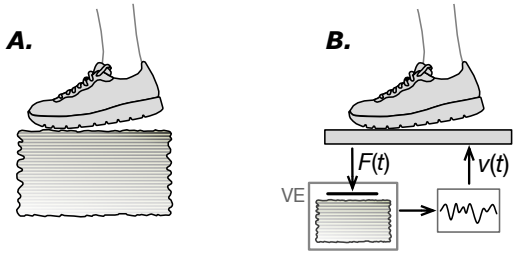


Figure 1: (*A.*) The experience of stepping onto a deformable object or surface is (*B.*) simulated via a rigid vibrotactile interface.

A haptic simulation (Fig. 1) provides feedback approximating the vibration response felt during interaction with a virtual object. Our rendering algorithms are of admittance type, computing velocities of displacement in response to forces applied to the virtual

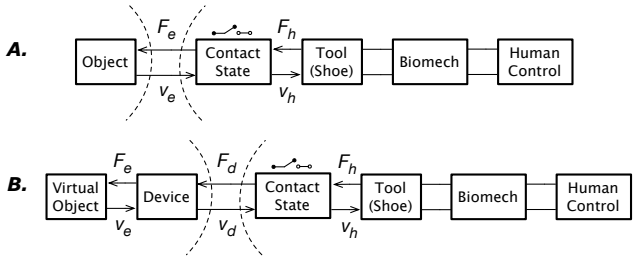


Figure 2: The haptic display concept. (*A.*) A real system being simulated involves interaction with an object via a tool (the shoe) in a user's grasp. The user exerts a force $F_h(t)$ and experiences a vibration $v_h(t)$. (*B.*) The display mediates interaction with the virtual environment. The dashed lines demarcate system components associated with the object, or its haptic presentation, from those of the user.

object. An example of such an algorithm is described in Sec. 2.2. As with most vibrotactile displays, vibrations are presented in open loop fashion. During the course of interaction, user coupling to the device interface is expected to vary from no contact to full-foot contact with large forces. In the virtual environment, interaction is mediated by an idealized proxy that is constrained to a constant region on the virtual object surface, at which force is applied. In these respects, the contact state is entirely modeled in the real world (Fig. 2). In other words, user-supplied forces (zero or otherwise), as measured by the device, are continuously mapped to the simulation, and there is no modeling of contact variation within the virtual environment.

2.1 Transparency

In order to ensure perceptual transparency of the display, the admittance of the virtual object it presents should resemble as closely as possible that of its real counterpart. As illustrated in Figure 2, this requires that the (velocity) feedback generated from forces applied to the virtual object closely match those produced by interaction with the real one, and that the device transfer function $H_d(s)$ be as close to unity as possible. Here, we refer to the free (unloaded) response of the display, corresponding to components to the left of the dashed lines in Fig. 2. When coupled to a user, the acceleration response of the interface will be modified due to the state of contact between the user and plate, and the biomechanical dynamics of the user (measurements are presented in Sec. 5). These variations are beneficial in allowing our device to mimic natural variations in the response of a real object that is being simulated. The dependence of the virtual object response on the force exerted by the user is incorporated explicitly, because our haptic rendering models are force driven. This contrasts with the case of impedance based displays, where only position (or velocity) is measured. In order to ensure transparent open-loop playback of haptic events in settings like the latter, the device's response requires grip-dependent corrections that account for changes in the display admittance due to grip or arm dynamics [9]. Further discussion is provided in Sec. 4.5.

2.2 Application to normal force texture simulation

An example application motivating the design reported here is the display of normal force textures, in the form of high frequency vibrations simulating the feel of stepping onto natural ground materials [36]. When a shoe steps onto porous materials like soil or snow, it is subjected to interactions that can include: viscoelastic components, describing the recoverable deformation of the volume of the ground surrounding the contact interface; transient shock components, from the impact of toe and heel against the ground; and plastic components from the collapse of air pockets, brittle structures or granular force chains, resulting in unrecoverable deformation [6, 29]. Combinations of such effects give rise to low-frequency forces and high frequency, texture-like vibrations that are characteristic of human walking on different surfaces [28]. Figure 3 presents an example of force and vibration data acquired from one footstep on a gravel surface. Because the vibration signature is continuously coupled to the force input over time in such examples, there is no straightforward way to convincingly use recorded footstep vibrations for vibrotactile rendering, although more flexible granular sound-synthesis methods could be used [1, 5]. For the modeling of simpler interactions, involving impulsive contact with solid materials, recorded transient playback techniques could be used [18].

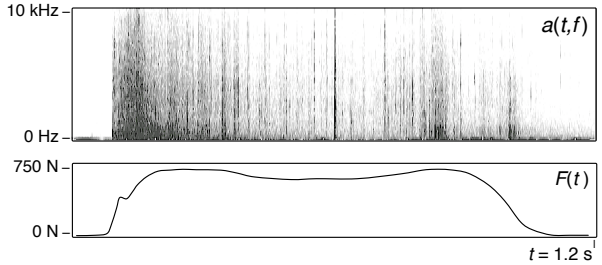


Figure 3: Vibration spectrogram $a(t,f)$ and normal force $F(t)$ measured from one footstep onto rock gravel (Authors' recording). Note the discrete, broadband impact events evidenced by vertical lines in the spectrogram.

We briefly summarize one approach we have taken to the haptic synthesis of interaction with such surfaces. It is based on fracture mechanics, which has also proved useful for modeling other types of haptic interaction involving damage [12, 20]. Figure 4 illustrates the continuum model and a simple mechanical analog used for synthesis. In the stuck state, the surface has stiffness $K = k_1 + k_2$ and is governed by:

$$F(t) = m\ddot{x} + b\dot{x} + K(x - x_0), \quad x_0 = k_2\xi(t)/K \tag{1}$$

where $\xi(t)$ represents the net plastic displacement up to time t. A Mohr-Coulomb yield criterion is applied to determine slip onset: When the force on the plastic unit exceeds a threshold value (which may be constant or noise-dependent), a slip event generates an incremental displacement $\Delta\xi(t)$, along with an energy loss of ΔW representing the inelastic work of fracture growth. Slip displacements are rendered as discrete transients, using an event-based approach [18]. High frequency components of such transient events are known to depend in detail on the materials and forces of interaction, and we model some of these dependencies when synthesizing the transients [38].

A high fidelity vibrotactile display such as that presented here is useful for reproducing such phenomena without artifacts.

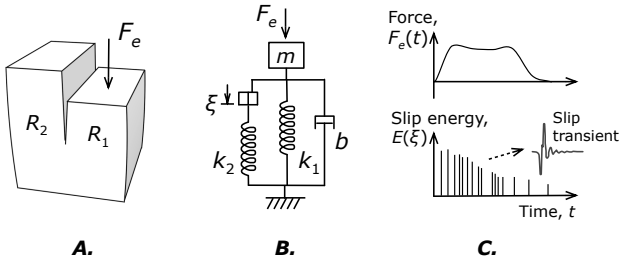


Figure 4: Normal force texture synthesis. *A.* A fracture mechanics approach is adopted. A visco-elasto-plastic body undergoes shear sliding fracture due to applied force F_e. *B.* A simple mechanical analog for the generation of slip events $\xi(t)$ in response to F_e. *C.* For vibrotactile display, each slip event is rendered as an impulsive transient using an event-based approach.

3 COMPONENTS

3.1 Mechanical Structure

The mechanical design is broadly similar to that of the original device, but accounts better for both the static and dynamic performance requirements for the display. The top plate provides an interface to the body, which in the case of our device is assumed to consist of a foot wearing a shoe. Statically, the device must resist bending when loaded vertically by a force of several hundred Newtons. The rigid deflection of the plate under this load must be minimized subject to the constraint that the plate be able to vibrate freely. This trade-off is analyzed in Section 4.3.

The top plate consists of a commercially manufactured aluminum honeycomb sandwich panel component (Museum Services Corp model 0513 SSP) with dimensions $30.4 \times 30.4 \times 2.54$ cm and a weight of 400 g. The panel has aluminum facings with a thickness of 0.08 cm. This material was selected for its high bending stiffness to weight ratio. The panel sides are closed with a basswood frame to eliminate acoustic emissions or deformations that otherwise result from small deflections of the honeycomb at the edges of the panel. The plate is supported by cylindrical SBR rubber elastic elements positioned as shown in Fig. 5. In dynamic or multi-tile configurations, a retaining socket surrounding the elastic support (not present in the figure) is used to keep the plate from changing position. The actuator is mounted via an aluminum bracket bonded to the center underside of the plate.

3.2 Sensing

In order to render an interactive response from a virtual ground surface using the kinds of models we have developed (Sec. 2.2), it is necessary to capture the normal force applied to the tile by the user's foot. Positioning the force sensors beneath the plate is feasible, since the bandwidth of the force applied to the plate by the user is limited. In the design presented here, the sensors are furthermore positioned beneath the elastic suspension of the device, so as to better isolate them from the actuators.

Force sensing is performed via four load cell force transducers (Measurement Systems model FX19) located below the vibration mount located under each corner of the plate. Although the cost for outfitting a single-plate device with these sensors is not prohibitive, many of the applications we have in mind consist of two dimensional $m \times n$ arrays of tiles, requiring a number $N = 4mn$ of sensors. As a result, in a second configuration, four low-cost resistive force sensors (Interlink model 402 FSR) are used in place of load cells. After conditioning, the response of these sensors to an applied force is nonlinear, and varies up to 25% from part to part (according to manufacturer ratings). Consequently, a measurement and subsequent linearization and force calibration of each is performed, using a calibrated load cell force sensor (details are provided in a separate publication [39]). After such a calibration, a linear response accurate to within 5% can be obtained using low cost parts.

3.3 Actuation

The tile is actuated by a single Lorentz force type inertial motor (Clark Synthesis model TST429) with a nominal impedance of 6 Ohms. The actuator is coupled to the tile by a 1.25 cm diameter threaded rod interfacing with an aluminum bracket, as shown in Fig. 5. The actuator has a usable bandwidth of about 25 Hz to 20 kHz, and is capable of driving the plate above strongly enough to quickly produce numbness in the region of the foot that is in contact with the tile.

3.4 Electronics

Analog data from the force sensors is conditioned, amplified, and digitized via a custom acquisition board, based on an Altera FPGA, with 16-bit analog-to-digital converters. Data from each sensor is sampled at a rate of 1 kHz and transported to a host computer over UDP via the board's 10 Mbps Ethernet interface. Digital to analog conversion of the signal driving the actuator is performed using a low noise 24-bit, 96 kHz audio interface (Edirol model FA-101), and amplification is performed using a compact, class-D audio amplifier based on the Tripath TK2050. The amplifier is rated as capable of providing 100 W to a nominal 4 Ohm actuator impedance.

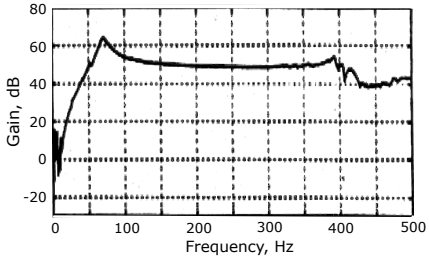


Figure 6: Typical acceleration frequency response of the Clark Synthesis TST family of actuators (manufacturer supplied data, reproduced with permission).

4 DYNAMIC RESPONSE

The main factors affecting the dynamic response of the device are the actuator characteristics, the dynamics of the rigid plate, and that of the elastic suspension. In a suitable regime, one may regard these as lumped linear systems, with respective Laplace transform domain transfer functions $H_a(s), H_p(s)$ and $H_s(s)$. The corresponding model device transfer function can be given by $H_d(s) = H_a(s)H_p(s)H_s(s)$. However, the spatial configuration of these elements relative to each other also contributes to the structural dynamics of the display.

4.1 Actuator Response

Fig. 6 shows a typical amplitude frequency response for this family of actuators. The data was acquired from an accelerometer attached to the mounting bold, with the actuator attached to a rigid wooden surface. The shape of this nominal response affects that of our device (Sec. 5).

4.2 Plate Response

The contribution of the interface plate to the dynamic response of the device can be predicted in terms of the vibrational characteristics of the plate, and the coupling between plate and actuator.

4.2.1 Free response

In the case of a homogeneous, isotropic plate, the free vibrations of interest are governed by a partial differential equation for bending wave displacements. In the thin plate approximation, it is [10]:

$$\rho h \frac{\partial^2 z}{\partial t^2} + D\nabla^4 z = 0, \quad D = \frac{Eh^3}{12(1-\nu^2)} \tag{2}$$

Here, $z = z(\mathbf{r}, t)$ is displacement, $\mathbf{r} = (x, y)$ is a position on the tile surface, t is time, E and ν are the Young's modulus and Poisson's ratio of the plate material, ρ is its mass density, and h is the plate thickness, D is known as the flexural stiffness, and we neglect effects of damping.

Below the first resonant frequency of the plate, f_1, the vibrational response is well approximated by that of an ideal, rigid mass coupled to an elastic suspension (Sec. 4.3 below). Since we seek a usable display bandwidth with an upper limit of 1 kHz, we want to ensure that $f_1 > 1$ kHz. For a rectangular, homogeneous, isotropic plate, the frequencies of the normal modes of bending oscillation

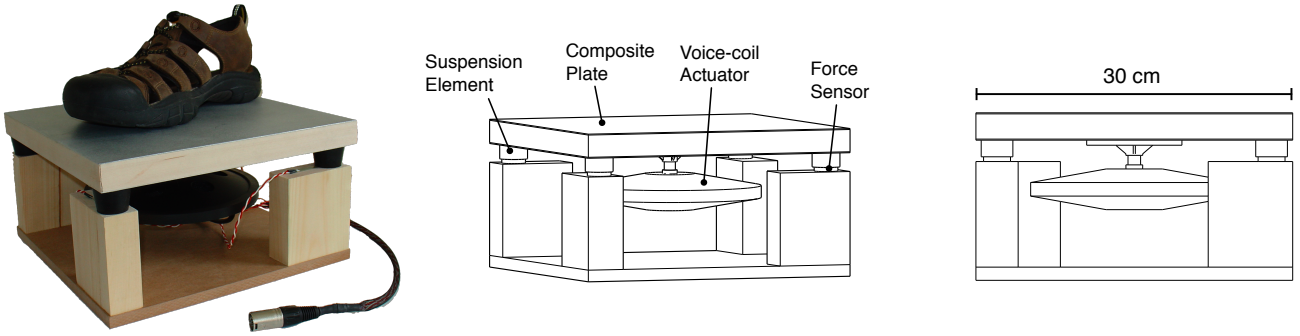


Figure 5: Vibrotactile floor interface hardware for a single tile unit. Left: Photo with large mens' shoe, showing representative size. The model shown is based on the low-cost force sensing resistor option. The cable in the foreground interfaces the sensors with the data acquisition unit. Middle: View showing main components. Right: Side view with top dimension.

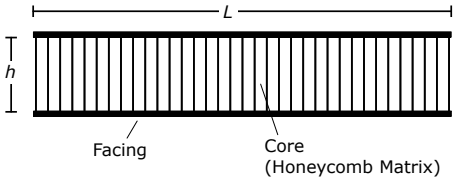


Figure 7: A composite sandwich plate, shown in profile (side view). It consists of a stiff, solid facing material bonded (typically laminated) to a lightweight core, such as a metal or kevlar honeycomb matrix.

are of the form [10]:

$$f(k) = \frac{hk^2}{2\pi}\sqrt{\frac{D}{\rho h}}, \quad k \propto L^{-1} \tag{3}$$

Here, k is the magnitude of the wavenumber vector $\mathbf{k} = (k_x, k_y)$ of the oscillation, and L^{-1} is the inverse length of the plate, and x and y are the directions tangent to the plate surface parallel to its edges. For pure simply supported boundary conditions (which are never achieved in practice), $k_x = m\pi/L$ and $k_y = n\pi/L$, where m and n are positive (non-zero) integers [10]. The lowest frequency mode is $f_1 = f(|\mathbf{k}_1|)$, with $\mathbf{k}_1 = (\pi/L, \pi/L)$. The amplitudes of the normal modes are

$$\eta_{mn}(\mathbf{r}) = \sin(k_x x)\sin(k_y y) \tag{4}$$

After inserting the form of D from Eq. 2 into Eq. 3, Poisson's ratio ν contributes a factor $(1-\nu^2)^{-1/2}$ to $f(k)$. This factor lies in a range from 1.0 to 1.16 for solid materials such as are considered here, so in a first approximation we may ignore it. In order to maximize the frequency $f(k)$ for any given value of the wavenumber magnitude k, and given plate geometric properties L and h, the plate material should be selected to possess a large Young's modulus E and low mass density ρ. Solid, isotropic materials such as metals can achieve a high stiffness, but their mass density is typically commensurately higher, which is a limitation. In addition, their mass grows linearly with thickness, making them less efficient to actuate. In practice, for a solid plywood plate (light enough to be efficiently actuated) with dimensions 30.4 cm × 30.4 cm × 3.75 cm, one finds $f_1 \approx 400$ Hz, which falls short of the design target.

4.2.2 Composite plate

A class of structures that achieves higher stiffness-to-mass ratios than is possible with uniform solids is that of composite sandwiches. Such a panel is formed via the use of thin layers of stiff solid material bonded to a lightweight core. Fig. 7 illustrates a composite sandwich of the type used in our device. The core consists of an aluminum honeycomb matrix, while the facing materials consist of aluminum sheet. The bending stiffness of such a material is approximately given by $E \approx \sqrt{3}E_f t_f/h$ [15], where E_f is the Young's modulus of the facing material, t_f is the thickness of the facing material, and h is the core thickness (Fig. 7). To a first approximation, the frequencies of the normal modes of bending oscillation of such a composite sandwich plate can be obtained[1] substituting this expression for E into Eq. 3. The resulting frequencies are:

$$f(k) = \frac{k^2}{2\pi}\sqrt{\frac{E_f t_f h}{12\rho}}, \quad k \propto L^{-1} \tag{5}$$

Again, the admissible values of k depend on the boundary conditions. This equation (which is valid only for thin facings $t_f \ll h$) depends on the stiffness E_f of the facing material and the average mass density of the entire panel. For our device, assuming idealized boundary conditions, the minimum value of k is $\pi\sqrt{2}/L$ (see Sec. 4.4). The other factors are given by: $h = 2.5$ cm, $t_f = 0.083$ cm, $E_f \approx 60$ GPa, and $\rho \approx 170$ kg / m^3. For these values, Eq. (5) yields $f_1 \approx 845$ Hz. This overestimates the measured value of f_1 for our device by approximately 10% (Sec. 5).

4.2.3 Actuator coupling

The actuated plate is driven by a surface force (pressure) distribution $F(\mathbf{r},t)$, and the resulting governing equation possesses a driving term:

$$\rho h\frac{\partial^2 z}{\partial t^2} + D\nabla^4 z = F(\mathbf{r},t) \tag{6}$$

For our device, the latter can be modeled as a $F(\mathbf{r},t) = F(t)\phi(\mathbf{r})$, where $F(t)$ is the actuator force amplitude, and $\phi(\mathbf{r})$ approximates a spatial Dirac delta function $\delta(\mathbf{r}-\mathbf{r}_0)$ centered at the tile midpoint. Generalizing slightly, one can consider the case of N independent point actuators at locations $\mathbf{r}_i$, in which case $F(\mathbf{r}) = \sum_{i=1}^{N} F_i(t)\delta(\mathbf{r}-\mathbf{r}_i)$, where F_i is the force signal from the ith actuator.

An arbitrary displacement $z(\mathbf{r},t)$ can be expanded in the normal modes $\eta_{mn}(\mathbf{r})$ of vibration, yielding modal coordinates $Z_{mn}(t)$ defined by:

$$z(\mathbf{r},t) = \sum_{m,n\geq 0} Z_{mn}(t)\eta_{mn}(\mathbf{r}), \quad \text{where} \tag{7}$$

$$Z_{mn}(t) = \int z(\mathbf{r},t)\eta_{mn}(\mathbf{r})d^2\mathbf{r} \tag{8}$$

[1] The authors did not find this expression for $f(k)$ in the research literature, but approximations like it are presumably well-known.

In these coordinates, the physical equation is [22]:

$$\left(\frac{\partial^2}{\partial t^2}+\omega_{mn}^2\right)Z_{mn}(t)=\sum_i \eta_{mn}(\mathbf{r}_i)F_i(t) \tag{9}$$

where ω_{mn} is the angular frequency of mode (m,n). Modes having nodes at the actuator location, $\eta_{mn}(\mathbf{r}_i)=0$, satisfy the homogeneous form of Eq. (9), so are not excited by the corresponding actuator signal, and instead contribute antiresonances to the device response. As shown in Sec. 5, the bandwidth of our device is essentially limited by the antiresonance near 775 Hz for $(m,n)=(1,1)$.

4.3 Elastic Suspension

The actuator is designed to supply a force only along the direction in which it is attached, so we are primarily interested in normal modes of oscillation in the z-direction. To a first approximation, the elastic (SBR rubber) suspension may be treated as a linear, lumped element with stiffness K in the z-direction, coupled to a rigid plate with total mass M (Fig. 8). For small displacements, this stiffness is given by $K=EA/h$, where E is the Young's modulus of the suspension element, A is its vertical surface area, and h is its height. When the plate is not subjected to load from a foot, the mass $M=M_0$ is due to the plate and actuator. For the device described here, $M_0=2.7$ kg.

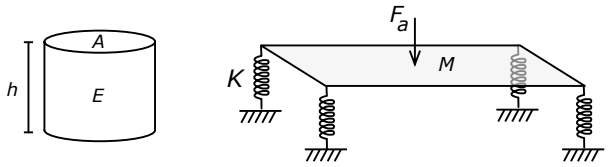


Figure 8: Left: Geometry of one elastic suspension element with height h, cross-sectional area A, and elastic modulus E. Right: Lumped model of the tile-suspension system. F_a is the actuator supplied force, $K=EA/h$ is the stiffness of a suspension element, and M is the combined mass of the actuator and tile.

Due to the symmetric placement of the actuator, there is a single dominant normal mode of oscillation, in which all suspension elements are driven in-phase in the vertical direction. The resonant frequency of this mode in the unloaded condition is given by $f_0=(2\pi)^{-1}\sqrt{4K/M_0}=(2\pi)^{-1}\sqrt{4EA/(hM_0)}$. Above this frequency, the gain of the transfer function factor due to the suspension, $H_{susp}(f)$, is expected to be approximately constant.

A softer suspension leads to a larger response bandwidth (i.e., one with a lower frequency extent) but also to a higher static deflection δz under a load F from a foot. The latter two are related by $\delta z=F/K=Fh/(EA)$. As nominal design specifications, we aimed to select E,A and h so that $f_0\le 50$ Hz, so that the passband would overlap most of the frequency range of sensitivity of the FA II (Pacinian) tactile mechanoreceptors in the foot. The latter closely resembles the range for those in the hand [16, 34]. In addition, we aimed to achieve a worst-case static deflection $\delta z\le 5$ mm under a load of $F=1000$ N, corresponding to a large human balanced at one corner. Both f_0 and δz depend only on the ratio A/h. For convenience, we choose the cross sectional area A of the suspension element to match that of the force sensor. To choose the remaining parameters, we perform a simultaneous grid search on h and E to minimize f_0 and δz. Figure 9 illustrates the dependence of f_0 and δz on E for a representative set of selected parameter values, i.e., $A=8$ cm^2 and $h=2.5$ cm. We selected SBR rubber vibration mounts with geometry and Young's modulus $E=2.0$ MPa consistent with this optimization. As shown in the figure, the resulting system is expected to achieve $f_0\approx 49$ Hz and $\delta z\approx 4.5$ mm.

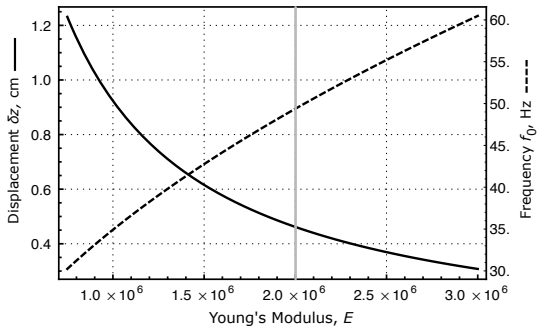


Figure 9: Resonant supension frequency f_0 and static deflection δz vs. Young's modulus for the system model with $A=8$ cm^2 and $h=2.5$ cm. The value $E=2$ MPa (grey vertical line) satisfies $f_0<50$ Hz and $\delta z<5$ mm.

4.4 Simulation

The vibrational characteristics of the tile-actuator system were simulated using finite element method (FEM) analysis. A three-dimensional geometric model of the device was designed, incorporating the sandwich panel, actuator (modeled as a homogeneous cylindrical mass), actuator bracket and connecting rod, elastic elements, and supporting structure. Material properties were assigned approximating those of the device itself, with the core honeycomb matrix replaced by a homogeneous solid with an equivalent density. Although this model entails a number of approximations, the qualitative results were expected to be correct.

4.4.1 Eigenfrequency analysis

At high frequencies, the resonant modes of the plate itself are observed. The first two are shown in Fig. 10. The lowest frequency mode has wavenumber $\mathbf{k}=(1,1)$, while the next highest resonance shown appears to correspond to a plate-induced mixture of the $(0,2)$ and $(2,0)$ modes. Their frequencies, 896 Hz and 1032 Hz, lie at a ratio of about 1.15:1.

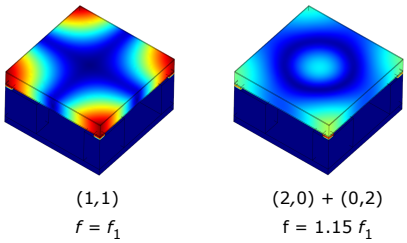


Figure 10: Visualization of the first two resonant modes of the device due to the bending mode vibrations of the plate as identified in the FEM analysis. The ring mode (right side) arises from the mixing of the plate modes indexed by the integers shown [10]. The frequencies are 896 Hz and 1032 Hz.

At low frequencies, a series of six resonances at frequencies of 28, 45, 51, and 77 Hz is observed, corresponding to deformations of the elastic support elements, in agreement with the lumped model of Sec. 4.3. The number of resonances exactly matches the six degrees of freedom of perturbation of the rigid tile. The deformation shapes recovered from the FEM analysis show that the 45 Hz resonance corresponds to oscillation in the direction normal to the plate surface; it is expected to dominate when the display is driven by the actuator.

4.4.2 Frequency response simulation

The FEM simulation of the frequency response was performed with a sinusoidal driving signal originating at the actuator. The z-axis

acceleration was measured at several points on the surface of the plate, with the results shown in Fig. 11. As determined above, the cross-shaped mode gives rise to an antiresonance, due to the actuator location. The magnitude effect of these resonances, and to a secondary degree their frequency, depends substantially on the measuring point.

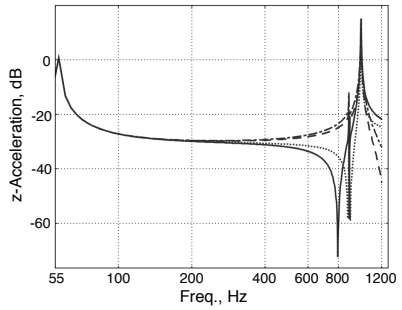


Figure 11: Result of a FEM analysis of the acceleration response of the device as measured at four different points on the plate surface. At frequencies below the first resonant mode, the device response is approximately constant over the surface of the plate.

4.5 Modeling user-supplied loads

We have so far discussed the response of the device without accounting for effect of to its user. According to the haptic interaction model adopted here (Sec. 2) it is correct to ignore the influence of this load, as transparency requires that the free (unloaded) display transfer function approach unity. However, in other circumstances it may be desirable to compensate for variations in the display response due to user contact. Examples where this might be the case include passive vibrotactile stimulation for tactile communication or psychophysical experiments, or for use with impedance-based rendering algorithms or simulations that explicitly model user contact conditions in the virtual environment. At the frequencies of interest, a user may be modeled as a passive viscoelastic mechanical system. Figure 12 shows a simple link-segment biomechanical model and equivalent mechanical network coupled to the display. Physiologically, biomechanical properties such as leg joint stiffness cannot be regarded as static or linear during movement, as muscular activation regularly modulates the dynamic properties of the lower limbs during activities on foot [26, 21, 8]. However, as a typical leg stiffness value one may take $k = 10$ kN/m [26]. Upon adopting a simple, single mass-spring-damper model with this value, and assuming an effective mass of 50 kg, one finds that a user-supplied load should suppress frequencies in the vicinity of 200-250 Hz. This prediction is consistent with the load-dependent frequency response measurements we report below (Sec. 5.1). Similar grip-dependent effects are observed to affect manually operated haptic [17, 9] and vibrotactile [41] devices.

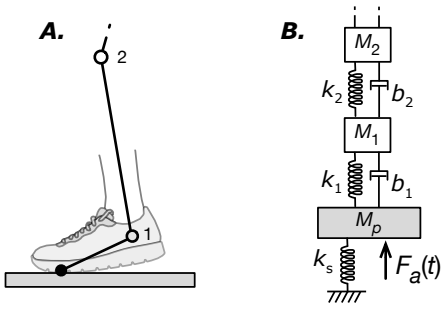


Figure 12: (*A.*) A simplified model of a user-supplied load can be provided in the form of a link-segment model or (*B.*) its equivalent mechanical network, coupled to the display device. $F_a(t)$ is force applied by the actuator.

5 Evaluation

As noted in Sec. 2.1, our assessment of the transparency of this device is based on the extent to which the unloaded transfer function of the display approximates unity over the range of perceptually salient frequencies. We focus here on the flatness of the magnitude frequency response. The vibrotactile sense is sensitive to phase differences [2], and it is possible that a vibrotactile display may contribute enough phase distortion to affect perception, particularly where transient signal components are concerned. However knowledge about vibrotactile sensitivity to temporal phase distortion is incomplete, and, as far as we are aware, phase linearity has never been used to evaluate vibrotactile displays. It is nonetheless an important consideration in the related field of loudspeaker design.

5.1 Frequency response measurement

The magnitude frequency response of the device was measured by driving the plate with the actuator using the chirp method, with a slow sinusoidal frequency sweep (rate of 100 Hz / s). Measurement was performed using a piezoelectric accelerometer (AKG model CP-411) bonded to the top surface of the plate, as described in the caption of Fig. 13. Frequency response measurements were taken for several different foot-plate contact conditions, while the foot was wearing a rubber soled shoe. As noted above, these contact conditions modify the impedance of the display, altering its response. The results are shown in Fig. 13.

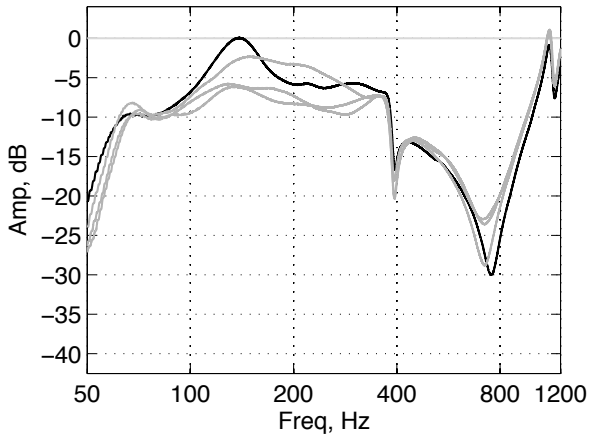


Figure 13: The measured log magnitude frequency response of the display. Measurements were taken at a point equidistant from the plate center and edge, on a line through the center, 15 degrees from the diagonal. The free response is shown with a black line, and other foot-floor contact conditions with varying load applied via the foot are shown in gray.

The variation in the magnitude frequency response below 80 Hz is, following the analysis above, likely to be attributed to a combination of the resonant modes of the vibration suspension and the actuator response (Fig. 6). The antiresonance around 775 Hz and the resonance near 1100 Hz are due to the two lowest frequency normal modes of the plate (Fig. 10) identified above. Variation in the range from 70 Hz to 700 Hz can be attributed in part to the actuator response. The latter (Fig. 6) includes a resonance near 70 Hz and additional coloration above 380 Hz, likely accounting for the smaller notch seen in the measurements. Through additional measurements with a calibrated accelerometer (Analog devices model

ADXL 320), it was determined that a force of 40 N could be transmitted at all frequencies between 50 and 750 Hz. Another pertinent quantity, nonlinear distortion, was estimated from measurements at 300 Hz to be slightly more than 5% (mean absolute percent error) up to a force of 30 N.

5.2 Digital correction

It is possible to improve the nominal frequency response $H_d(f)$ of the display by filtering the actuator signal $F(t)$ via a suitably designed linear, time-invariant corrective filter, H_c. Here, we concentrate on the magnitude only: H_c is designed so that the corrected device frequency response $H_{d'}(f) = H_d(f)H_c(f) \approx g$ in the band of interest, where g is a constant gain factor.

Since such a filter lacks any spatial dependency, correction is most effective below the first resonant frequency of the device, f_1, because above it the device transfer function varies across the surface of the plate (Fig. 11). As a result, correction is primarily useful for compensating factors, such as actuator characteristics, that can be treated as lumped parameter system elements.

We designed the inverse filter H_c to equalize the device response in the frequency range from $f = 50$ Hz to 750 Hz. It was implemented digitally as an IIR filter of order N, which was estimated using the least p-th norm optimization method [32]. Figure 14 shows a comparison of the original (free) frequency response of the device with responses corrected by filters of order $N = 10$ and 14. In the latter case, the response is flat in a passband with -10 dB roll off near 50 Hz and 750 Hz.

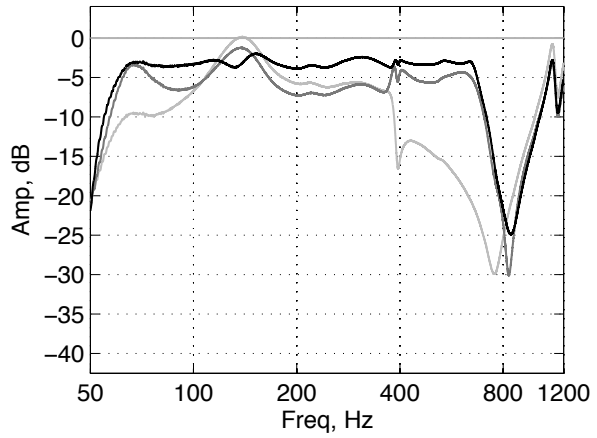


Figure 14: Measured device response without foot contact, uncorrected (lightest gray) and with correction by digital IIR filters of order $N = 10$ (medium gray) and $N = 14$ (black).

6 Conclusion

We have presented a vibrotactile display device integrated in a rigid surface, consisting of an actuated and instrumented floor tile. The analysis considered factors affecting the bandwidth of the device, including the response of the elastic suspension and rigid plate, the actuator coupling, and the static deflection of the suspension under load by a human foot. The plate was implemented as a lightweight composite sandwich panel constructed from aluminum honeycomb. A digital filter was designed to compensate for artifacts in the measured frequency response. The device allows for accurate reproduction of frequencies between about 50 and 750 Hz.

The device is simple, and designed to be easily reproducible or adaptable to the demands of various applications or research tasks [37, 38]. Formulae such as Eq. (5) indicate how the response characteristics of the display can be expected to scale with system dimensions.

Despite these results, a number of areas can be identified in which the device might be improved. They include:

- *Plate material selection:* As presented in the analysis above, a composite panel similar to that used here, but constructed from a honeycomb core with larger core thickness h, and facing thickness t_f would achieve a usable bandwidth that extends to higher frequencies. The first resonant frequency of the plate scales as $\sqrt{t_f h}$ in the thin plate approximation (which will require corrections if the plate or facings are too thick).

- *Actuator design:* The large voice coil actuator used in the present device introduces significant coloration in the frequency response of the device, much of which can be equalized. A better response may be achieved through the use of smaller, more efficient voice coil motors coupled to the structure in a spatial configuration chosen to optimize the device transfer function.

- *Vibration control:* Active structural vibration control strategies exist that could improve the fidelity of such a display using arrays of surface mounted sensors and actuators, together with closed-loop controllers [4, 30, 31]. Such techniques might be used to achieve accurately controlled vibrational responses under a wider array of contact loads.

In ongoing research in our lab, the device is being utilized to study the rendering and display of virtual ground surface properties, related aspects of haptic perception, and the integration of such display components in multimodal virtual and augmented reality environments [39].

Acknowledgements

We thank Vincent Hayward and Guillaume Millet for useful discussions, and the anonymous reviewers for their helpful comments. The authors gratefully acknowledge support from the MDEIE of Quebec for the EU FP7 project NIW (no. 222107). The first author acknowledges support from the ESF COST Action on Sonic Interaction Design (no. IC0601).

References

[1] S. Barrass and M. Adcock. Interactive granular synthesis of haptic contact sounds. In *AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio*, 2002.

[2] S. Bensmaia and M. Hollins. Complex tactile waveform discrimination. *The Journal of the Acoustical Society of America*, 108:1236, 2000.

[3] G. Burdea and M. Akay. *Force and touch feedback for virtual reality*. Wiley New York, 1996.

[4] K. Chandrashekhara and A. Agarwal. Active vibration control of laminated composite plates using piezoelectric devices: a finite element approach. *Journal of Intelligent Material Systems and Structures*, 4(4):496, 1993.

[5] A. Crossan, J. Williamson, and R. Murray-Smith. Haptic granular synthesis: Targeting, visualisation and texturing. In *Proc. Intl Symposium on Non-visual & Multimodal Visualization*, pages 527–532, 2004.

[6] S. Dixon and A. Cooke. Shoe-Surface Interaction in Tennis. *Biomedical engineering principles in sports*, page 125, 2004.

[7] F. Fahy. *Foundations of engineering acoustics*. Academic Press, 2001.

[8] C. Farley and O. Gonzalez. Leg stiffness and stride frequency in human running. *Journal of Biomechanics*, 29(2):181–186, 1996.

[9] J. Fiene, K. Kuchenbecker, and G. Niemeyer. Event-based haptics with grip force compensation. In *Proc IEEE Haptic Symposium*, 2006.

[10] N. Fletcher and T. Rossing. *The physics of musical instruments*. Springer Verlag, 1998.

[11] M. Fukumoto and T. Sugimura. Active click: tactile feedback for touch panels. In *CHI'01 extended abstracts*, pages 121–122. ACM, 2001.

[12] V. Hayward. Physically-based haptic synthesis. In M. Lin and M. Otaduy, editors, *Haptic Rendering: Foundations, Algorithms and Applications*. AK Peters, Ltd, 2007.

[13] V. Hayward and K. Maclean. Do It Yourself Haptics: Part I. *IEEE Robotics & Automation Magazine*, page 89, 2007.

[14] J. Hollerbach. Locomotion interfaces and rendering. *Haptic Rendering: Foundations, Algorithms and Applications. AK Peters, Ltd*, 2008.

[15] J. Kee Paik, A. Thayamballi, and G. Sung Kim. The strength characteristics of aluminum honeycomb sandwich panels. *Thin-walled structures*, 35(3):205–231, 1999.

[16] P. M. Kennedy and J. T. Inglis. Distribution and behaviour of glabrous cutaneous receptors in the human foot sole. *The Journal of Physiology*, 583(3), 2002.

[17] T. A. Kern. Modeling the User. *Engineering Haptic Devices (T. A. Kern, Ed.), Springer-Verlag*, 2009.

[18] K. Kuchenbecker, J. Fiene, and G. Niemeyer. Improving contact realism through event-based haptic feedback. *IEEE Transactions on Visualization and Computer Graphics*, 12(2):219–230, 2006.

[19] M. Magana and R. Velazquez. On-shoe tactile display. In *IEEE Intl Workshop on Haptic Audiovisual Environments and Games*, 2008.

[20] M. Mahvash and V. Hayward. Haptic rendering of cutting: A fracture mechanics approach. *Haptics-e*, 2(3):1–12, 2001.

[21] D. Marigold and A. Patla. Adapting locomotion to different surface compliances: neuromuscular responses and changes in movement dynamics. *Journal of neurophysiology*, 94(3):1733, 2005.

[22] L. Meirovitch. *Dynamics and control of structures*. John Wiley & Sons, 1990.

[23] A. Nashel and S. Razzaque. Tactile virtual buttons for mobile devices. In *Proceedings of CHI*, pages 854–855. ACM, 2003.

[24] I. Poupyrev, S. Maruyama, and J. Rekimoto. Ambient touch: designing tactile interfaces for handheld devices. In *Proceedings of UIST*. ACM, 2002.

[25] A. Priplata, J. Niemi, J. Harry, L. Lipsitz, and J. Collins. Vibrating insoles and balance control in elderly people. *The Lancet*, 362(9390):1123–1124, 2003.

[26] S. Rapoport, J. Mizrahi, E. Kimmel, O. Verbitsky, and E. Isakov. Constant and Variable Stiffness and Damping of the Leg Joints in Human Hopping. *Journal of Biomechanical Engineering*, 125, 2001.

[27] A. Rovers and H. Van Essen. Guidelines for haptic interpersonal communication applications: an exploration of foot interaction styles. *Virtual Reality*, 9(2):177–191, 2006.

[28] J. Sabatier and A. Ekimov. A Review of Human Signatures in Urban Environments Using Seismic and Acoustic Methods. In *2008 IEEE Conference on Technologies for Homeland Security*, pages 215–220, 2008.

[29] V. Stiles, I. James, S. Dixon, and I. Guisasola. Natural Turf Surfaces: The Case for Continued Research. *Sports Medicine*, 39(1):65, 2009.

[30] U. Stobener and L. Gaul. Modal vibration control for PVDF coated plates. *Journal of Intelligent Material Systems and Structures*, 11(4):283, 2000.

[31] N. Tanaka and T. Sanada. Modal control of a rectangular plate using smart sensors and smart actuators. *Smart Materials and Structures*, 16(1):36–46, 2007.

[32] J. Taylor and Q. Huang. *CRC handbook of electrical filters*. CRC, 1997.

[33] H. Tramberend, F. Hasenbrink, G. Eckel, U. Lechner, and M. Goebel. CyberStage – An Advanced Virtual Environment. *ERCIM News*, 31, October 1997.

[34] M. Trulsson. Microreceptive afferents in the human sural nerve. *Exp Brain Res*, 137:111–116, 2001.

[35] Y. Visell. Tactile sensory substitution: Models for enaction in HCI. *Interacting with Computers*, In press, 2008.

[36] Y. Visell, J. Cooperstock, B. Giordano, K. Franinovic, A. Law, S. McAdams, K. Jathal, and F. Fontana. A Vibrotactile Device for Display of Virtual Ground Materials in Walking. In *Proceedings of Eurohaptics*, 2008.

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

[38] Y. Visell, A. Law, and J. Cooperstock. Touch Is Everywhere: Floor Surfaces as Ambient Haptic Interfaces. *IEEE Transactions on Haptics*, 2009.

[39] Y. Visell, A. Law, J. Ip, S. Smith, and J. R. Cooperstock. Interaction Capture in Immersive Virtual Environments via an Intelligent Floor Surface. In *Proc. of IEEE Virtual Reality*, 2010 (To appear).

[40] Y. Visell, S. Smith, A. Law, R. Rajalingham, and J. R. Cooperstock. Contact Sensing and Interaction Techniques for a Distributed, Multimodal Floor Display. In *Proc. of IEEE 3D User Interfaces*, 2010 (To appear).

[41] H. Yao, V. Hayward, and R. Ellis. A tactile enhancement instrument for minimally invasive surgery. *Computer Aided Surgery*, 10(4):233–239, 2005.

Contact Sensing and Interaction Techniques for a Distributed, Multimodal Floor Display

Yon Visell, Severin Smith, Alvin Law, Rishi Rajalingham, Jeremy R. Cooperstock
McGill University, Montreal, Canada

ABSTRACT

This paper presents a novel interface and set of techniques enabling users to interact via the feet with augmented floor surfaces. The interface consists of an array of instrumented floor tiles distributed over an area of several square meters. Intrinsic force sensing is used to capture foot-floor contact at resolutions as fine as 1 cm, for use with floor-based multimodal touch surface interfaces. We present the results of a preliminary evaluation of the usability of such a display.

Keywords: Foot interaction, Multimodal display

1 INTRODUCTION

Arguably, one factor that has limited the use of foot-based interaction for computationally augmented environments is the lack of efficient interfaces and interaction techniques capable of capturing touch via the feet over a distributed display. In this paper, we present the design of an interface based on a distributed network of low-cost, rigid floor tile components, with integrated sensing and actuation capabilities. In order to make good use of this system, we draw on contact sensing techniques that are able to capture foot-floor interactions with much finer resolution than is achieved if the tile is regarded as the smallest relevant spatial unit.

2 FOOT INPUT IN HUMAN-COMPUTER INTERACTION

Examples of the use of foot-controlled input in HCI, interactive arts and video gaming date at least as early as Amiga's Joypad (1983) [13]. Pearson and Weiser later introduced a foot input device for a desktop PC [10]. Despite the sustained interest in touch screens for the hands, less research has addressed the design and usability of similar interfaces for the feet. Companies such as Gesturetek and Reactrix have developed interactive floor-based visual displays using video sensing technology, but such sensors provide no direct information about foot-floor contact forces and positions. Such information is arguably essential for rendering interactions with virtual objects or controls. Floor-based multimodal (visual, auditory, tactile) information displays have only recently begun to be investigated [14]. In the domain of immersive virtual environments (VEs), devices for enabling omnidirectional in-place locomotion in VEs exist [4], but are complex and costly. Lower cost methods for navigation and interaction in VEs, such as the shoe-based Step WIM interface of LaViola et al. [7], require special apparel and provide limited feedback. Most prior work on tactile interaction with floor surfaces utilizes surface sensing arrays [9, 12], for applications such as person tracking, activity tracking, or musical performance. Although similar sensing interfaces are now commercially available, costs remain high. Further differences between tactile sensing and our approach are noted in Sec. 4.

3 SYSTEM DESCRIPTION AND COMPONENTS

The floor interface (Fig. 1) consists of a square array of 36 rigid tiles, each of which is instrumented with force sensors (four per tile) and a vibrotactile actuator. The floor is coated in gray projection paint. A pair of overhead video projectors is used for visual display, in order to reduce the impact of shadows cast by users. The tiles are rigid, composite plates with dimensions $30.5 \times 30.5 \times 2$ cm, supported by elastic vibration mounts, and coupled to a vibrotactile (VT) actuator (Clark Synthesis, model TST229) beneath the plate. Actuator signals are generated on personal computers, output via digital audio interfaces, and amplified. The floor tile display achieves a VT passband from 50 Hz to 750 Hz, and is capable of reproducing the largest forces needed for interaction with virtual ground surface objects or properties (more than 30 N across the indicated frequency band).

Normal forces are sensed at locations below the corner vibration supports of each tile using a total of four resistive force sensors (Interlink model 402 FSR). Analog data from the force sensors is conditioned, amplified, and digitized via a 32-channel, 16-bit data acquisition board. Each sensor is sampled at a rate of up to 1 kHz transmitted over a low-latency Ethernet link. An array of six small form factor computers is used for force data processing and audio-VT rendering. A separate, networked server is responsible for rendering visual feedback and managing user input.

4 INTRINSIC CONTACT SENSING

Intrinsic contact based sensing aims to resolve the locations of contact, the forces at the interface, and the moment about the contact normals using internal force and torque measurements [2]. It is assumed to involve contact between a rigid apparatus and an object (here, a foot). This approach has mainly been applied to problems in robotic manipulation, but we have adapted it to foot-ground interaction sensing. It can be viewed as an efficient alternative to sensing via dense surface mounted arrays, as it requires far fewer sensors. The method resolves the contact centroid $\mathbf{x}_c$ associated with a pressure distribution $p_R(\mathbf{x})$ that is distributed over an area R. $\mathbf{x}_c$ is a contact point such that a point force F_c that gives rise to the same intrinsic force measurements as $p_R(\mathbf{x})$ does [2]. The sensing problem is simple to formulate for a single floor tile (Fig. 2), with force sensor locations $\mathbf{x}_j$ where internal force measurements f_j are taken and j indexes the tile sensors. We ignore friction effects for the moment. The contact centroid $\mathbf{x}_c$ and normal force $\mathbf{F}_c = (0,0,F_c)$ can be recovered from scalar force measurements $\mathbf{F}_j = (0,0,f_j)$ via force and torque equilibrium equations,

$$\sum_{j=1}^{4} f_j + F_c + f_p = 0 \quad (1)$$

$$\sum_{j=1}^{4} \mathbf{x}_j \times \mathbf{F}_j + \mathbf{x}_c \times \mathbf{F}_c + \mathbf{x}_p \times \mathbf{F}_p = 0\,. \quad (2)$$

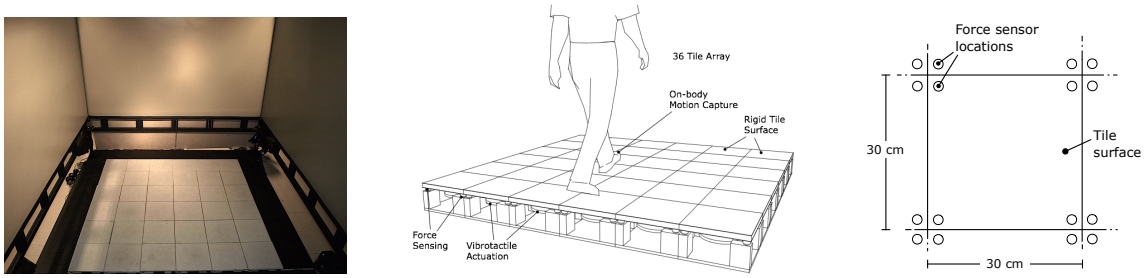


Figure 1: Left: The floor interface is situated within an immersive, rear projected virtual environment simulator. Middle: Sensing and actuating components are integrated beneath the floor. Right: View from above showing sensor locations.

$\mathbf{F}_p = (0,0,f_p)$ is the weight of the the plate and actuator at the tile's center $\mathbf{x}_p$. The three nontrivial scalar equalities (1, 2) can be solved for contact centroid parameters:

$$F_c = \sum_{i=1}^{4} f_i - f_p, \quad \mathbf{x}_c = \frac{1}{F_c}\left(\sum_{i=1}^{4}(\mathbf{x}_i - \mathbf{x}_p)f_i + f_c\mathbf{x}_p\right) \qquad (3)$$

The contact centroid lies within the convex hull of the contact area (dashed line, Fig. 2) at the centroid of the pressure distribution [2], and thus provides a concise summary of the foot-floor contact locus, but not about shape or orientation. When the foot-floor contact area overlaps two or more tiles, the pressure centroid $\mathbf{x}_R$ for the entire contact area can be computed from contact centroids $\mathbf{x}_{ck}$ for each tile (computed from Eq. (3)). It is given by the weighted average $\mathbf{x}_c = w_1\mathbf{x}_{c1} + w_2\mathbf{x}_{c2}$, where $w_k = F_i/F$. The domain-independence of this result thus makes it possible to track these points as they move across tile boundaries. Figure 3 represents results of real measurements of 50 estimated contact positions determined by the method of Eq. (3), using a single calibrated floor tile. Despite distortion near tile edges, contacts were localized with a typical accuracy of 1.5 cm, and worst-case values of ≈ 3 cm, smaller than the linear dimensions of the tile (30 cm) or the typical width of a shoe.

5 Applications: Floor Touch UIs

These sensing methods can be employed to implement virtual floor-based touch interfaces. One set of examples we have created consist of array of standard UI widgets to be controlled with the feet (Fig. 4). Input is based on a multi-touch screen metaphor mediated by a set of interaction points (cursor locations), which are defined as the contact centroids $\mathbf{x}_c$ with the largest forces. Force thresholds associated with a control are used to determine selection. The controls provide positive tactile feedback supplied by the actuators, in the form of synthesized click-like transient vibrations or sliding (friction) vibrations.

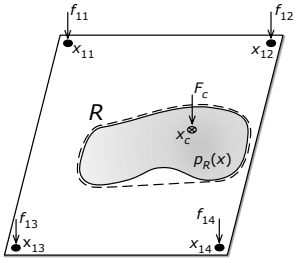


Figure 2: A normal force distribution $p_R(\mathbf{x})$ and associated contact centroid position $\mathbf{x}_c$.

Interface design toolkit Interface design is facilitated by a software layer and network protocol that abstracts the hardware systems, which are accessed over a local Ethernet network, and connects them to the user interface. This software layer processes the sensor data to extract interaction points, and provides them with IDs that persist throughout contact. Second, it allows to remotely cue and present VT feedback localized to the area defined by each interface object on the floor. The protocol design is based in part on the TUIO protocol for table-top touch interfaces [6].

5.1 Preliminary User Evaluation

A question we encountered when beginning to design such touch-surface applications concerned the appropriate size of virtual controls. The required size depends on factors including sensing limitations, users' motor abilities, target parameters, and feedback modalities, as has been extensively studied and modeled in the HCI literature [8, 5]. The size appropriate for touch screen controls has been shown to depend on the interaction technique adopted. Precision control strategies can enable single pixel accuracy in finger-based touch screen interaction [1, 11], and related techniques may prove effective for use with the feet. Limited research has addressed floor interfaces (Sec. 5), so we focused here on a basic task requiring the selection of controls presented at various locations and sizes to a stationary user.

Human movement research has investigated foot movement control in diverse settings. Visually guided targeting with the foot has been found to be effectively modeled by a similar version of Fitts' law as is employed for modeling hand movements, with an execution time about twice as long for a similar hand movement [3]. However, the present, preliminary, investigation addresses a situation in which usability is manifestly co-determined by both operator and device limitations, providing a window on both.

Apparatus The apparatus is the floor interface presented above. Due to the floor size, the sensor calibration was less accurate by a factor of two than that which yielded the position estimates noted above, but was sufficient for interaction points to be

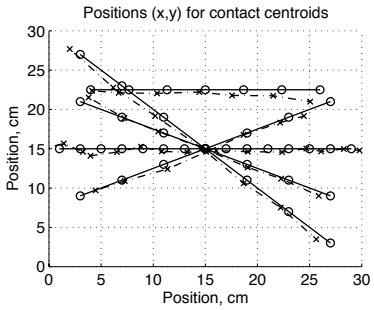


Figure 3: Results of 50 measurements comparing true normal force positions (circles) with contact centroid estimates (Xs).

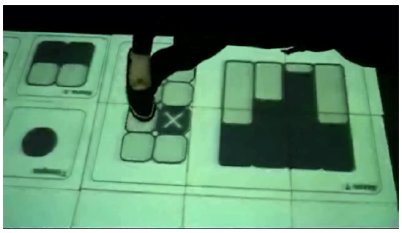


Figure 4: Still image from video of a user interacting with floor-based interface widgets.

effectively tracked over extended distances on the floor, as shown in the supplementary video.

Stimuli and method The stimuli consist of round virtual buttons to be selected by users, who began each trial with their feet in locations marked by white rectangles. Users could activate a button by pressing it in a way that resulted in a contact centroid within the area of the button exceeding a force threshold of about 35 N. The buttons ranged in diameter from 4.5 to 16.5 cm, and were presented at four distances, on lines radiating from between their feet, oriented at one of two angles, as shown in Fig. 5. Upon selection, the buttons provided visual feedback in the form of a 20 cm white disc centered in place of the original appearance. All buttons provided the same feedback. Only the buttons and foot locations were displayed. No audio or VT feedback was provided.

Hypothesis We expected user performance to improve with target size and degrade with target distance. Interaction between target distance and width might be anticipated, but we do not attempt to validate a model. We expected good performance for targets that are at least as wide as the foot.

Participants Eight participants, ranging in age between 21 and 38, volunteered for this study. All of them were research staff or students in the Faculty of Engineering at McGill University.

Procedure Participants wore their own shoes during the experiment, and selected targets with their preferred, dominant foot. They were instructed to activate the buttons precisely and quickly. The non-preferred foot was not constrained, but participants were required to return both feet to the two rectangular regions shown in Fig. 5 between stimuli. Most chose to leave their non-preferred foot in place throughout each session.

The experiment began with a practice period lasting three minutes, followed by the main experiment. The latter consisted of two sessions of 12 minutes, with a short pause in between. A total of 240 stimuli were presented to each participant. Stimuli were presented in sequential, randomized order. Each button appeared and remained visible and active for two seconds during which users were able to select it. A three second pause followed, after which

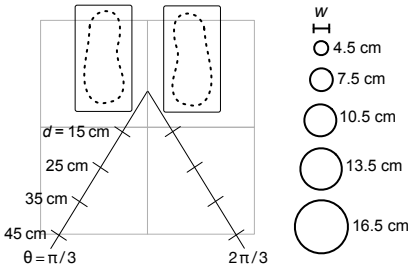


Figure 5: Configuration and stimuli from the experiment.

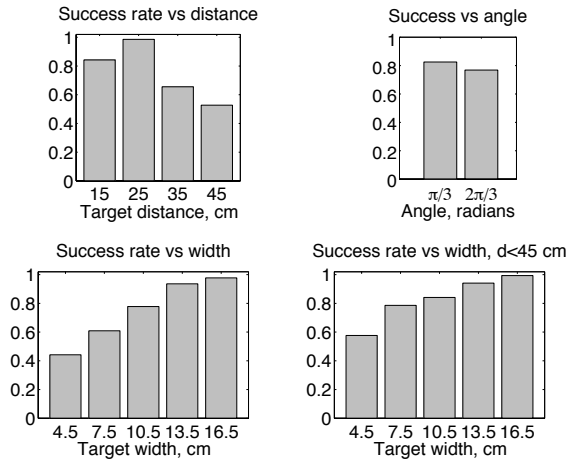


Figure 6: Top Left: Successful target selection rate vs. distance, averaged across the other conditions. Top Right: Success rate vs. angle of presentation (measured away from preferred foot). Lower Left: Successful target selection rate vs. button width, averaged across the other conditions. Lower Right: The same measure, excluding the farthest targets.

the next button appeared. The success of selection and time required were recorded. Participants completed a response questionnaire and provided verbal comments afterward.

Analysis Summaries of the success frequencies are presented in Figure 6. Using a logistic regression analysis, we determined that the main factors of width w, distance d, and bearing angle θ significantly affected success of selection ($p < 0.001$). The fitted logit is $z = 1.4 + 0.071w - 0.062d - 0.6\theta$ (with t-values > 7.8). θ is in radians, increasing away from the preferred foot; d and w are measured in cm. The model correctly predicts 86% of the responses.

Discussion Users selected larger targets within the allotted two-second interval at a higher rate of success than smaller ones. Performance with the largest was very high (98%), and that for the smallest was low (44%). Small targets pose two potential problems. First, as with other touch surface displays, targets can be occluded by the foot during selection. This problem appeared to be mitigated because targets could be seen before selection, while during selection they were projected on the top of the foot. Second, limitations on precise control can arise from factors such as shoe width, human motor abilities, and sensor positioning errors. Six out of eight participants reported finding a strategy to activate the small buttons by using a feature of the shoe or changing the applied force. Software interaction techniques for improving precise touch screen control are known [1, 11], and we intend to investigate these in future work. Nearby targets (distances of D =15 to 25 cm) were selected at a higher rate. However, performance was better at 25 cm than at the nearest distance of 15 cm (98.5% vs. 84%, with $p < 0.001$ using Fisher's exact test, two-tailed). One likely reason is that if an interface element is beneath a standing user, it can be occluded from view by the body, or present a difficult viewing angle. Due to such effects we would not expect selection time T to follow a Fitts' Law relation, $T = a + b\log_2(D/W)$, but this was not tested here. Although a mobile user may be able to avoid visibility problems, they seem to be an important consideration. For our device, position sensing is most accurate near the centers of the tiles, as indicated in the preceding section. This was noticed by users of the system, two

of whom volunteered that they had learned to better activate small buttons that were close to edges by pressing them off-center. In ongoing work, we are developing algorithms for correcting such positioning distortions. Participants consistently reported difficulty in selecting targets that were oriented away from their active, selecting foot, however the effect of bearing angle on performance was small (Fig. 6). It is possible that these responses were indicative of a larger motor effort. Neck fatigue was most frequently cited by participants as a source of discomfort.

Future work Although these results are suggestive, further work is needed in order to characterize the usability aspects of this display, and others like it. A greater understanding of factors such as control element size, display scale, motor abilities, modalities, and other aspects salient to the use of such a device will certainly be needed.

One notable question not addressed by this study concerns the interplay between users' movements on foot and their interactions with the touch surface. A novel aspect is that, implicitly, both feet are involved, due to requirements of movement and of maintaining balance. In everyday actions, like striking a soccer ball, weight is often shifted onto one foot, which specifies an anchored location, while the opposite is used to perform an action. Floor interfaces that involve movement may thus be expected to have something of the flavor of bimanual interaction in HCI, a connection we intend to explore in future work.

5.2 Floor UIs: Potential application space

Virtual floor controls could be advantageous in some areas of man-machine interaction in which foot interfaces are common, such as manufacturing, mass transportation, or dentistry. Certain applications areas, such as those related to pedestrian navigation or map-based visualization, may emerge as particularly salient. In domains such as medicine, documented problems with existing foot controls might be overcome [15]. Other fields, including entertainment, gaming, and marketing, were noted in the introduction.

6 Conclusions

This paper presented interaction techniques based on intrinsic contact based force sensing via a novel distributed floor interface. Such foot-floor contact information is not usually available through optical sensing channels such as motion capture. The system is low in cost and complexity, and can be employed by multiple simultaneous users, without any specialized apparel. In addition, this paper demonstrates the integration of these interaction techniques within multimodal displays implementing virtual ground surface simulations or floor-based control interfaces. Despite the promising nature of these results, there are several respects in which the present system might be improved or extended:

– Our system senses three DOF per tile (the normal force f_c and position $\mathbf{x}_c$). To solve the sensing problem required assuming frictionless contact (Sec. 4). A future interface for sensing the full six rigid DOF per tile (via additional sensors) would remove this assumption.

– A floor interface with a denser array of tiles would be capable of capturing more information about foot-ground contact shape.

– During multi-tile foot-floor contact, a contact-based sensing approach results in clusters of contact centroids. New techniques are needed in order to acquire the information arising from such features.

– Interaction points can be followed only as long as foot-floor contact is sustained. Methods for tracking users' feet between gestures or actions would be desirable.

– As noted more extensively in Sec. 5.1, above, further work is needed in many areas of usability in order to develop design guidelines and strategies for floor-based interfaces.

It is nonetheless hoped that this contribution succeeds in highlighting the potential of floor-based touch surfaces.

Acknowledgements

The authors acknowledge support from the MDEIE of Quebec for the EU FP7 project NIW (no. 222107).

References

[1] P. Albinsson and S. Zhai. High precision touch screen interaction. In *Proc of the ACM SIGCHI conference on Human factors in computing systems*, pages 105–112. ACM New York, NY, USA, 2003.

[2] A. Bicchi, J. Salisbury, and D. Brock. Contact sensing from force measurements. *The Int J of Robotics Research*, 12(3):249, 1993.

[3] E. Hoffmann. A comparison of hand and foot movement times. *Ergonomics*, 34(4):397, 1991.

[4] J. Hollerbach. Locomotion interfaces and rendering. *Haptic Rendering: Foundations, Algorithms and Applications. AK Peters, Ltd*, 2008.

[5] J. Jacko and A. Sears. *The human-computer interaction handbook: fundamentals, evolving technologies, and emerging applications*. Lawrence Erlbaum Assoc Inc, 2003.

[6] M. Kaltenbrunner, T. Bovermann, R. Bencina, and E. Costanza. Tuio: A protocol for table-top tangible user interfaces. In *Proc of Gesture Workshop 2005*. Gesture Workshop, 2005.

[7] J. LaViola Jr, D. Feliz, D. Keefe, and R. Zeleznik. Hands-free multi-scale navigation in virtual environments. In *Proc of the 2001 symposium on Interactive 3D graphics*, pages 9–15. ACM New York, NY, USA, 2001.

[8] I. MacKenzie. Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 7(1):91–139, 1992.

[9] J. Paradiso, C. Abler, K.-y. Hsiao, and M. Reynolds. The magic carpet: physical sensing for immersive environments. In *CHI '97: CHI '97 extended abstracts on Human factors in computing systems*, pages 277–278, New York, NY, USA, 1997. ACM.

[10] G. Pearson and M. Weiser. Of moles and men: the design of foot controls for workstations. In *Proc of the ACM SIGCHI conf on Human factors in Comp Sys*, pages 333–339. ACM New York, NY, USA, 1986.

[11] R. L. Potter, L. J. Weldon, and B. Shneiderman. Improving the accuracy of touch screens: an experimental evaluation of three strategies. In *CHI '88: Proc of the SIGCHI conference on Human factors in computing systems*, pages 27–32, New York, NY, USA, 1988. ACM.

[12] A. Schmidt, M. Strohbach, K. v. Laerhoven, A. Friday, and H.-W. Gellersen. Context acquisition based on load sensing. In *UbiComp '02: Proc of the 4th international conference on Ubiquitous Computing*, pages 333–350, London, UK, 2002. Springer-Verlag.

[13] J. Sinclair, P. Hingston, and M. Masek. Considerations for the design of exergames. In *Proc of the 5th Int ACM Conf on Comp Graphics and Interactive Techniques, Southeast Asia*, page 295, 2007.

[14] Y. Visell, A. Law, and J. Cooperstock. Touch Is Everywhere: Floor Surfaces as Ambient Haptic Interfaces. *IEEE Transactions on Haptics*, 2009.

[15] L. Wauben, M. van Veelen, D. Gossot, and R. Goossens. Application of ergonomic guidelines during minimally invasive surgery: a questionnaire survey of 284 surgeons. *Surgical endoscopy*, 20(8):1268–1274, 2006.

Touch Is Everywhere: Floor Surfaces as Ambient Haptic Interfaces

Yon Visell, Alvin Law, and Jeremy R. Cooperstock

Abstract—Floor surfaces are notable for the diverse roles that they play in our negotiation of everyday environments. Haptic communication via floor surfaces could enhance or enable many computer-supported activities that involve movement on foot. In this paper, we discuss potential applications of such interfaces in everyday environments and present a haptically augmented floor component through which several interaction methods are being evaluated. We describe two approaches to the design of structured vibrotactile signals for this device. The first is centered on a musical phrase metaphor, as employed in prior work on tactile display. The second is based upon the synthesis of rhythmic patterns of virtual physical impact transients. We report on an experiment in which participants were able to identify communication units that were constructed from these signals and displayed via a floor interface at well above chance levels. The results support the feasibility of tactile information display via such interfaces and provide further indications as to how to effectively design vibrotactile signals for them.

Index Terms—Vibrotactile interfaces, ubiquitous computing, haptic communication design, floor surfaces.

1 Introduction

Tactile feedback has received growing attention as a means of enhancing or enabling information display in diverse computing applications. As human-computer interaction has extended beyond the desktop computing paradigm, and into every other domain of human activity, tactile display has grown in importance. This is attributable, in part, to its ability to overcome the sensory overload and attentional demands that arise in complex, multitasking environments.

Two new paradigms that have emerged within this context are those of mobile or wearable information appliances, and of ambient computing. While significant attention has been devoted to the opportunities for tactile display to enhance mobile applications, less has been given to haptic interaction with computationally augmented environments. Nonetheless, the fundamental role that floor surfaces play in our haptic negotiation of everyday environments suggests that they hold significant potential for active tactile information display.

The design of haptic information for ground surfaces has a long history, as is evidenced in urban environments. Haptic markers are commonly used to indicate locations or paths of interest to visually impaired people. Similarly, they are employed to emphasize low-lying features, such as subway stairs, that need to be highlighted even to sighted individuals. Many everyday ground surfaces could also be profitably augmented with active tactile information displays. Such displays might find roles that are complementary to those that have been explored in the mobile computing domain. Some simple end-user scenarios may be helpful to guide the discussion.

- A visually impaired pedestrian is traveling to an appointment on foot and by public transportation. Reaching a noisy urban crosswalk, she is able to infer, via a vibrotactile cue received near the curb, that the crossing signal is red. Later, she receives a cue indicating that it is safe to cross (Fig. 1). She enters the subway and is able to feel, via haptic feedback supplied near the platform edge, that the train has not arrived. Once it arrives, similar cues indicate the locations of the train doors. She reaches the building of an office, and locates the elevator. While ascending to her destination, she receives a vibrotactile cue from the elevator floor, indicating the floor number that has been reached, and instantly knows when to disembark.
- An elderly person taking a shower in his home receives subtle vibrotactile feedback to his feet, unconsciously aiding him in maintaining his balance in response to sensed shifts in his center of mass and pressure.
- A rescue team is engaged in an augmented reality training simulation that aids them in learning to evaluate and respond to the changing conditions of a structure during an emergency. They receive realistic cues through the response of a haptically actuated floor delivering signals appropriate to the material and local stability of the ground surface in the simulation.

1.1 Potential Advantages

Floor-based haptic communication interfaces may share many positive features that are characteristic of other kinds of haptic communication.

• *The authors are with the Centre for Intelligent Machines, McGill University, 3480 University St., Room 424, Montreal H2A 3A7, Canada. E-mail: yon@cim.mcgill.ca, vinnylaw@gmail.com, jer@cim.mcgill.ca.*

Manuscript received 16 Nov. 2008; revised 4 May 2009; accepted 21 May 2009; published online 10 July 2009.
Recommended for acceptance by K. Kahol, V. Hayward, and S. Brewster.
For information on obtaining reprints of this article, please send e-mail to: toh@computer.org, and reference IEEECS Log Number THSI-2008-11-0093.
Digital Object Identifier no. 10.1109/ToH.2009.31.

1939-1412/09/$25.00 © 2009 IEEE Published by the IEEE CS, RAS, & CES
Authorized licensed use limited to: McGill University. Downloaded on March 08,2010 at 14:24:42 EST from IEEE Xplore. Restrictions apply.

Fig. 1. A possible end-user scenario: pedestrians receive vibrotactile cues via the ground surface (shown as shading around the feet), indicating the location of the crossing and state of its signal.

- *Complementarity to other modalities:* Such displays can function well even when visual attention is occupied, or when the environment is noisy.
- *Attentional salience:* Tactile displays can be an effective means of directing attention to a significant event or location, in the presence of task load or sensory distraction.
- *Discreetness:* Information received through such a display need only be apparent to its user.

Haptic communication via floor interfaces may, in addition, offer some specific advantages, such as those of:

- *Ubiquity and unintrusiveness:* Reflecting the near universal presence of foot-floor contact in human environments.
- *Acceptability:* While people may be averse to touching certain objects in public spaces such as restrooms, floor surfaces are broadly acceptable to touch, via the shoes.
- *Ready accessibility:* Floor-based interfaces need not require users to possess or wear any special technology, making them accessible to a wider range of users.
- *Lack of specific demands on visual or manual search:* An interface might be positioned at a location (such as a pedestrian crossing) in such a way that little visual or manual search is required to locate it. This can be an advantage for users with visual impairments.
- *Location, navigation, and event salience:* Such interfaces may be well suited for applications that involve pedestrian navigation, wayfinding, or location-based information display.
- *Locomotion-specific effects:* As noted below, a floor might be readily designed to actively enhance specific sensorimotor aspects of walking.

2 Background

While the literature on haptic interaction design for the feet is quite limited, much of what is known about haptic communication via other areas of the body may be readily extended to the case of interest. In addition, in large part through research conducted over the last two decades, more is now known about tactile sensation in the feet. This background knowledge can, by extension, inform the design of floor-based interfaces, and it also raises issues that are uniquely applicable to haptic communication design for floor surfaces.

2.1 Tactile Sensation and Psychophysics of the Foot

The tactile sensory physiology and psychophysics of the foot have been the subject of considerable research, if to a lesser extent than in the case of the hand. The foot has, since the mid-20th century, been acknowledged as one of the most sensitive parts of the body to vibrotactile stimulation [42], [40]. The sensory physiology of the sole of the foot is similar to that of the glabrous skin of the hand, including the same types of tactile mechanoreceptors as are present in the latter [34], namely, fast-adapting type I and II receptors (FA I, FA II) and slow-adapting type I and II receptors (SA I, SA II). Their nervous responses largely mirror those of receptors in the hand [17], although some differences in peripheral vibrotactile information coding have been identified (see, e.g., Ribot-Ciscar et al. [26]). With regard to the receptor spatial distribution, Kennedy and Inglis found those in glabrous skin of the foot sole to be relatively widely distributed, with (in contrast to the hand) little preferential accumulation in the toe areas [17]. Receptive fields were found to be larger than in the hand by a factor of about three. Physiological activation thresholds were determined to be higher, on average—by a factor of approximately eight, in the case of FA II receptors [17]. This latter finding agrees with earlier observations of higher psychophysical thresholds to vibrotactile stimulation of the foot compared to the hand [40]. It has been suggested that this is due, in part, to biomechanical differences between the skin of the hands and of the feet, and possibly to mechanoreceptor properties [41]. The ball and arch of the foot have been found to be the areas most sensitive to vibrotactile stimulation [14].

Sensitivity has also been assessed for populations of different ages, with elderly people demonstrating elevated thresholds for vibrotactile stimulation at FA II mediated frequencies [14] (i.e., those most often targeted by vibrotactile displays [37]). Thus, age is a significant factor in haptic interaction design for floor interfaces. As in other areas of haptic communication, such differences may be compensated by learning on the part of users, or by plasticity effects, whereby repeated exposure over time has been found to improve vibrotactile discrimination [7].

Distinctive functional characteristics of the foot relative to the hand include the reduced prehensile dexterity of the former (which is reflected in the kinds of activities in which it is involved), and the fact that static and dynamic forces on the feet during stance are higher and more sustained than those in the hand (i.e., on the order of 100-1,000 Newtons in the former case). Thus, while the thresholds measured in the studies cited above were assessed as subjects were lying down or otherwise off their feet, when individuals are walking, those thresholds may be higher, due to adaptation effects resulting from the large forces involved [8]. As in the

Authorized licensed use limited to: McGill University. Downloaded on March 08,2010 at 14:24:42 EST from IEEE Xplore. Restrictions apply.

case of the hand, most of the receptor types of the foot are simultaneously active during normal motor activities, unlike the more segregated responses that are observed to accompany simpler cutaneous stimulation by static probes, vibrators, or electrodes [17], [34]. Ribot-Ciscar et al. observed that vibrotactile stimulation of the foot can lead to a transformation of physiological messages potentially leading to the overestimation of static forces through coactivation of SA I afferents [26]. As a result, the application of extrinsic vibrotactile stimulation can result in unintended behavioral modifications affecting posture and gait. Ribot-Ciscar et al., among others, have previously identified similar, manual sensorimotor impairments affecting workers exposed to high-amplitude vibrations [25]. Roll and Gilhodes have investigated various proprioceptive illusions that can be induced by vibrotactile stimulation [28].

Humans on foot are implicitly engaged in a sensorimotor task (e.g., quiet stance or normal walking). The cutaneous tactile channels addressed by the types of interface we describe here are active in the peripheral regulation of balance and locomotion through reflexes coordinating stimuli felt through the feet to muscles in the leg and foot (see, e.g., Tax et al. [32], Zehr et al. [43]). During locomotion, the coupling of motor reflexes to cutaneous stimulation has been found to depend on both stimulus properties and on the instantaneous gait phase at the time of stimulation. Thus, if a display is meant to be accessed during locomotion, careful consideration is warranted of stimulus design and the timing of presentation. Such entanglements of tactile sensation in the feet and motor behavior suggest that constraints on the design of actuated floor displays are needed in order to avoid adverse effects on gait and stance. At the same time, certain applications might exploit such effects. For example, a person entering a dangerous area might stop more quickly if cued by a suitable vibrotactile warning signal from the floor.

Finally, there has been much recent interest in the observation that it is possible to enhance sensation in the feet, and thereby postural and gait control, by providing subthreshold noise to the foot soles [23]. This effect is seen as significant for elderly populations, and for others with peripheral neuropathies. Sensory aids of this type could be implemented through the kinds of display described in this paper, either as a central or complementary feature.

2.2 Prior Research on Interfaces for the Feet

While there has been little research on the design of haptically actuated floor surfaces, much may be learned from past work in areas such as the passive haptic design of ground surfaces, tactile feedback in foot-based human-computer interaction, and locomotion interfaces for virtual environments.

2.2.1 Haptic Design of Ground Surfaces

Public transit areas, such as urban sidewalks, pose special risks to pedestrians with visual impairments. This is partly due to the fact that they cannot make use of visual cues or signs that are the most common means of marking hazards (e.g., at intersections). Tactile ground surface indicators consist of regularly textured areas of ground, in the form of patterns of raised domes, bars, or other bumps, arranged on the sidewalk to mark significant paths or points of safety concern [29]. While international specifications for such markers remain to be established, they must be clearly identifiable, without being obtrusive. When higher than about 5 mm, they have been found to pose risks for stumbling or falling [19]. Alternative means of demarcating floor areas have been proposed to remedy this. Kobayashi et al. investigated the discrimination of floor areas by elasticity, in order to determine the feasibility of using this type of material variation as a substitute for ground surface indicators [18]. An active vibrotactile cue supplied through the ground could provide another approach.

2.2.2 Haptic Locomotion Interfaces

One area of recent research has concerned the engineering of locomotion interfaces for virtual environments, as recently reviewed by Iwata [15] and Hollerbach [12]. However, this research has predominantly focused on the challenging problems of high-fidelity force-reflecting haptic interaction for omnidirectional virtual walking experiences. The display of vibrotactile information underfoot for the purpose of increasing immersion during locomotion in virtual environments has only recently begun to be addressed [20].

2.2.3 Tactile Communication with the Feet

Vibrotactile displays for presenting information-bearing stimuli to the soles of the feet have been little explored to date. Shoes for this purpose have, for example, been investigated for information conveyance via nonintrusive or hands-free interfaces [31], [36]. Rovers et al. found that users were able to identify several families of haptic icons, consisting of moving patterns on the foot sole presented through an array of small vibration motors in the sole of a shoe-like apparatus [31].

Despite the limited research that has addressed the feet, there is ample evidence that information can be transmitted via diverse body surfaces, devices, encodings, and under many different conditions; extensive reviews of related considerations from the literature in these areas are provided in [7], [37], [21]. Beginning in the 1960s, Geldard and others systematically studied the use of tactile displays for communicating symbolic information via different parts of the body. Later, research on sensory substitution aimed at conveying information about shape, spatial configuration, or environmental conditions near a user of a distributed tactile display; such displays were designed for body parts such as the tongue, forehead, thigh, or back [37]. Basic guidelines for stimulation by vibrotactile feedback are now being formulated [16]. Trends in recent research aim at uncovering central capacities for, and limitations on, tactile information processing [7], and at establishing a foundation for the design of large sets of structured vibrotactile messages, based on perceptual and usability criteria [21]. Although such guidelines necessarily depend on the display device, application, and user community addressed, basic strategies have been successfully applied to many different interfaces and sensory modalities. We profit from prior research on musically inspired tactile communication design discussed in Section 3.7.

Authorized licensed use limited to: McGill University. Downloaded on March 08,2010 at 14:24:42 EST from IEEE Xplore. Restrictions apply.

2.2.4 Human-Machine and Human-Computer Interfaces

There are many control interfaces for machine operation by foot (car accelerator pedals, dental equipment, and sewing machines), and somewhat fewer for human-computer interaction (foot-controlled computer mice, sensing floors, and shoes). Few of these have profited from active haptic feedback. Systems providing haptic warning cues via an automobile's accelerator pedal have been researched for many years as a means of improving driving safety [22], and implementations have now reached the market (e.g., Infiniti's Distance Control Assist). Ferber et al. studied haptic communication during a human control task conducted on a haptically augmented stair climbing machine. They found some simple haptic cues supplied to the feet via an exercise machine to be effective at aiding participants in maintaining a target exercise level. The cues consisted of regularly spaced tapping sequences encouraging the person exercising to exert more effort when he or she slowed down [6]. Rovan and Hayward developed vibrotactile interfaces for furnishing additional feedback during computer music performance, including augmented floor tiles and in-shoe stimulators [30]. While they reported the feedback to be subjectively effective at conveying spatial and temporal information, no systematic evaluation was performed. We are not aware of any prior work on the design of vibrotactile information displays for pedestrians via the actuation of floor surfaces.

3 Ambient Haptic Floors: Display Design

Since ambient haptic floor displays have only recently been developed, the potential design space has been little explored to date. In this section, we highlight roles that such interfaces might profitably take on in future applications and discuss design issues, including prototyping methods. We present a prototype floor platform and describe several design strategies that we have implemented for it.

3.1 Roles for Haptic Communication

Although application requirements depend on factors that are difficult to anticipate outside of a case-based discussion, a number of possible roles for a haptic floor interface can be identified.

Location-based information display. Due, in part, to the commonality of foot-ground contact, such a device may be suited for providing notifications about a current or future event associated with a location (e.g., the arrival of a bus at its stop), or of informing about an ongoing process.

Display of navigation-salient information. Such a device could be highly suited for demarcating a location or region of interest to pedestrians (e.g., a crosswalk location, or an in-store promotion), a direction or directions of interest, or for indicating a pedestrian path or passage (e.g., the route to a nearby emergency exit).

Functional augmentation of features or devices. In interactive settings, floors as interfaces may prove useful for supplementing the functionality of existing architectural features (e.g., steps, entrances, stairs, and tactile ground surface indicators) or interfaces (e.g., foot switches, haptic locomotion interfaces, and exercise machines [6]).

Creative, entertainment, or emotional communication. They may also be used for diversion, comfort, entertainment, or in the context of artistic creation, for example, by silently providing a common tempo to musical performers, or by communicating information to a computer music performer [30].

Sensorimotor aids. Cutaneous tactile stimulation via the soles of the feet can be designed to improve locomotion during rehabilitation or normal activities [24], [10]. It could also be used to provoke reflexes meant to avert a pedestrian from stepping in the path of danger.

Virtual reality display. Augmented floor surfaces can be used for simulating natural ground surface features (e.g., earth, sand, and stone) for virtual reality or augmented reality simulations that involve movement on foot (see [38] and Section 3.5).

3.2 Physical Interaction Design Considerations

Good design demands consideration of the device's physical configuration and other factors affecting the interaction. Some relevant considerations include: Whether the interface is to be integrated with an existing feature or device (e.g., an area of tactile ground surface indicators), or if it is to constitute a new artifact, such as a custom-built floor platform; Whether the interaction is to be afforded to a single person or to a group; The required spatial resolution; The level of interactivity that is needed; The level of independence of stimulation to each foot. Such considerations can only be addressed in detail through a particular design.

3.3 Prototyping Technologies and Methods

Many of the technologies needed for prototyping vibrotactile floor displays are readily available and require comparatively little engineering to deploy. The necessary electronic components consist of vibrotactile transducers, amplifiers, and signal generation circuitry (typically some form of computer with associated digital to analog converter).

Among the distinguishing technical requirements are the relatively higher power demands, owing to both the elevated vibrotactile sensory thresholds in the feet relative to those in the hand, and to the requirement of actuating a floor surface bearing a large mass.

Among other commercially available actuation technologies, one suitable choice is the linear voice coil motor. It consists of a metallic mass suspended on an elastic structure between the coils of an electromagnet, which, when driven, exerts a force directly against a structure to which it is attached (see the recent review article [9] for a discussion). An advantage of these actuators is that they provide independent control over amplitude and frequency, which is not the case for simpler devices, such as the eccentric mass motors commonly used in mobile phones. A second advantage is that the entire actuator may be concealed within the device. Several commercially available voice coil actuators exist that may be suitable for prototyping floor surface applications such as those described here.[1] The main technical requirements are that they possess a usable frequency bandwidth overlapping the domain of sensitivity

1. Examples include the IBEAM VT200 (Sonic Immersion LLC, Hürth, Germany) and the TST239 (Clark Synthesis, Inc., Littleton, Colorado).

Authorized licensed use limited to: McGill University. Downloaded on March 08,2010 at 14:24:42 EST from IEEE Xplore. Restrictions apply.

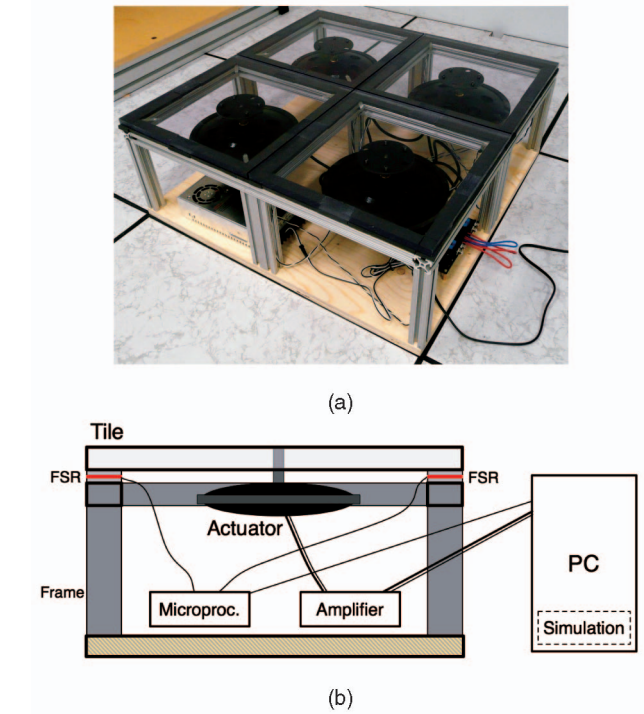


Fig. 2. (a) Interactive, four-tile prototype described in the text. (b) Diagram of same, with PC and rendering software.

of FA II receptors in the feet and are capable of supplying sufficient power to actuate the surface. Such actuators can readily be driven with normal audio amplifiers.

Some technical simplifications are made possible by presenting vibrotactile stimuli via a rigid, actuated floor surface. First, this approach facilitates the protection of all electronic components beneath the surface of the floor (see Section 3.4). Second, the size of the surface being actuated means that if it is driven at frequencies extending into the audible band, usable auditory feedback can be supplied through the same interface, allowing for cross-modal information display. Since vibrations above about 1,000 Hz are not felt, higher frequency signal components may, in part, be shaped so as to improve the accompanying auditory feedback. While we have found this feedback usable in practice, there are significant limitations arising from structural resonances and emission characteristics of the foot-floor system.

In some cases, a stimulus may be supplied through the floor in a way that is conditional upon, or computed as a function of, forces supplied by a pedestrian stepping onto that area of the floor. For example, a floor surface may be configured to provide a response resembling a (virtual) foot switch, controlled by foot pressure. Interactivity may be enabled via an array of force sensors embedded in the floor. A detailed discussion is omitted here. One example is described below.

We would like to address a few concerns that may arise in the mind of the reader regarding such interactions. First, despite the coupling between actuator and sensor, the forces supplied by the former are often negligible, because the magnitudes involved (on the order of a few Newtons or less) are much smaller than the low-frequency forces

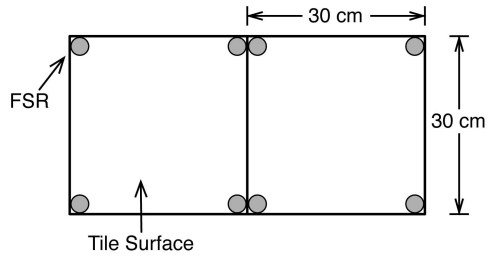


Fig. 3. A 2×2 tile prototype seen from above. Force sensing resistors are positioned as shown.

exerted by pedestrians against the ground (i.e., as large as 1,000 Newtons). Second, although actuator size or other constraints may dictate a coarse spatial arrangement of actuators (distributed with a spacing of one or several feet in distance) this need not necessarily be apparent to a user. A stimulus is generally felt through each area of the foot in proportion to the force it exerts on the vibrating surface. Thus, even though the spatial density of actuators may be very coarse, the resulting experience is that the stimulus originates at the locus of foot-floor contact. The effect is similar to that used to present virtual tactile buttons on touchscreen displays. For similar reasons, boundaries between individual tiles are less noticeable than might be expected, provided similar responses are supplied by adjacent tiles.

3.4 Prototype Vibrotactile Device

The device used in our work has evolved through a series of earlier prototypes [38] to pedestrians standing or walking upon it. It is assumed that they are wearing their accustomed footwear, and no special equipment is required to be worn. In addition to the actuators, force sensing capabilities have been integrated to enable interactivity, along the lines described above. We have previously used this interface to interactively synthesize vibrotactile signatures similar to those that would normally be generated by walking on natural materials, such as snow or gravel [38].

The device is simple and robust in construction. A prototype is shown in Fig. 2. The four tiles are rigid polycarbonate, of dimensions $30.5 \text{ cm} \times 30.5 \text{ cm} \times 1.25$ cm. They rest on an aluminum substructure. A force sensing resistor (Interlink model 403) encased in foam rubber 0.5 cm thick is positioned under each of the four corners of each tile (Fig. 3).

A vibrotactile actuator is rigidly attached to the underside of each tile via a steel mounting bracket. In applications that are interactive, data from the force sensors are digitized and transmitted via a serial USB link to a personal computer that interactively generates the vibrotactile signals. The actuators are inertial voice coil motors (Clark Synthesis model TST-239). They are capable of driving the floor surface with a greater power (by a factor of about four) than is required for the tiles used here. However, they are also used to prototype applications in which much larger tile surfaces are driven.

Authorized licensed use limited to: McGill University. Downloaded on March 08,2010 at 14:24:42 EST from IEEE Xplore. Restrictions apply.

3.5 Temporal Interaction Design Considerations

A floor-based haptic communication display, such as that described above, can also be characterized in terms of temporal properties of the stimuli involved, and the user interactions, if any, that generate them. Conceptually, one can group these qualities into those describing short-time properties of the vibrotactile signals being displayed, the manner in which they are patterned on longer time scales, and the type of interactivity afforded by the display and application.

For the purpose of this paper, *short-time* features are those related to time scales (smaller than about 200 ms) over which the stimulus may be thought to be relatively stationary or unitary. Such signals can be thought of as building blocks from which more complex stimuli can be built. They may be characterized by properties related to their frequency content, temporal extent, temporal amplitude envelope, or modulation that is applied to any of these. Alternatively, they might, as discussed below, be designed by analogy to a physical event, such as an impact transient (IT), with properties resembling hardness or resonance.

Low-level signals can be profitably organized into structured *temporal patterns*. A motivation for doing so is to improve identifiability or to increase perceptual information content in a set of stimuli. The frequency bandwidth and temporal resolution of the tactile channel limit the amount of information that can be encoded via short-time features. In recent research on vibrotactile communication, musically inspired structures (motifs or rhythms) have been profitably used to design larger sets of vibrotactile icons that are perceptually well distinguished from each other [3], [35], [33], [21]. An alternative approach, discussed further below, is to arrange low-level signals to resemble the temporal pattern of a physical process, such as bouncing or breaking (for an example from an auditory perspective, see [39]).

Finally, properties related to the *interaction* itself refer to the manner in which vibrotactile signals are generated in response to user actions, as captured through sensors of the device. Here, one can also allow for the possibility that the device is noninteractive. Properties of the interaction are determined by the sensing method used, and the way in which the resulting data are used to control the synthesis of the vibrotactile signal. According to the nature of this control mapping, the interaction may be characterized as *discrete* or *continuous*. A continuous interaction gesture can be described as *effective* (in the gestural typology suggested by Cadoz [4]) if the vibrotactile signal is generated in a way that energetically increases with the energy of users' actions. A related notion is that stimuli may be synthesized as if generated by a virtual physical system, such as a bed of gravel, which is meant to be stepped upon.

3.6 Three Interaction Cases

To illustrate these design considerations, we present three different interaction case examples (Fig. 4) developed for the interface described in Section 3.4.

1. A floor surface displays structured vibrotactile signals composed of abstract elements (such as a sum of sinusoids with a certain temporal envelope) resembling musical notes and arranged in a rhythmic pattern (i.e., a *tacton*).

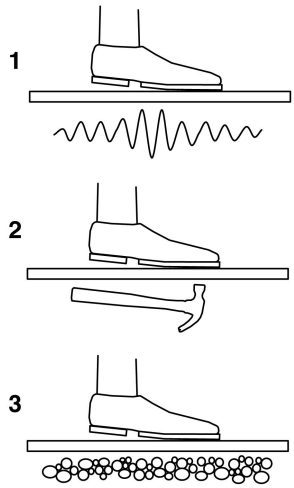


Fig. 4. Representations of three interaction paradigms. **1.** A vibrotactile message designed using waveform-level properties (i.e., a *tacton*). **2.** One derived from a model of a physical process, such as an object tapping on the tile. **3.** A virtual physical material interaction, such as a step onto a bed of gravel.

2. Such a surface presents stimuli that resemble the vibrational signature resulting from a physical object, such as a hammer, tapping with a rhythmic pattern on the underside of the tile. These representationally coded signals may be thought of as *tactile icons*.
3. A tile surface responds to the applied force exerted by a footstep with a signature approximating that of a natural ground material, such as gravel or snow. A computer simulation is responsible for rendering this feedback in a way that is highly dependent upon the measured force profile with which the step is executed in time [38].

Case 3 has been explored in an earlier publication by the authors [38]. Methods for the design of stimuli for cases 1 and 2 are described in detail in Section 3.7. They are evaluated in the experiments described in Section 4.

3.7 Vibrotactile Stimulus Design

Methods for the presentation of structured vibrotactile stimuli corresponding to case examples 2 and 3 of Section 3.6 were prototyped. The approach is based, in part, on a musical phrase metaphor that has been evaluated positively in recent literature on vibrotactile display [33], [3], [35]. In it, a stimulus is encoded in a rhythmic phrase, characterized by a set of notes each having a certain onset time, duration, and amplitude. A phrase is constrained to consist of two or more repetitions of a musical bar. The time domain of a single bar is quantized into 24 unit steps. A tempo is set, determining the total duration of a bar (in seconds). Notes can begin at any step and possess durations given by integer numbers of steps. Only a single note is allowed to play at any time. The amplitude of each is specified at its onset. Notes within the pattern of a given vibrotactile signal differ only in their amplitudes and durations, so that all possess the same short-time parameters (e.g., frequency and roughness). This limits the number of parameters that must be specified for each

Authorized licensed use limited to: McGill University. Downloaded on March 08,2010 at 14:24:42 EST from IEEE Xplore. Restrictions apply.

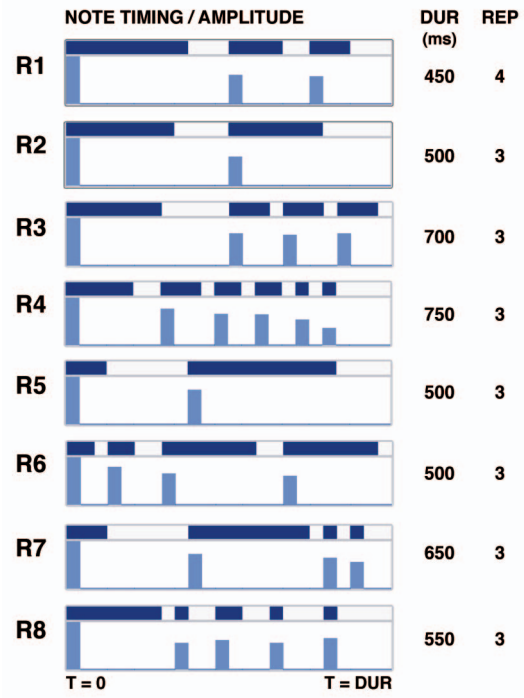


Fig. 5. The eight rhythms (R1-R8) from which both sets of stimuli were built. Time runs left to right. The bold horizontal lines show the temporal extent of notes. The vertical bars coincident with the onsets indicate amplitude. Bar duration and number of repetitions are shown.

stimulus. The algorithms used to synthesize the note-level signals are described in the sections that follow.

The vibrotactile signals used in the experiment described in Section 4 were built from a set of eight rhythmic patterns generated using this method. The corresponding bars are shown in Fig. 5. The duration of each bar is between 450 and 750 ms, and each is repeated a fixed number of times (either three or four). As shown in the figure, an accented note marks the beginning of each bar, serving as an "anchor" for the pattern [33]. The waveform of a stimulus from the experimental set, corresponding to rhythm R1 from Fig. 5, is shown in Fig. 7.

3.7.1 Waveshaping Synthesis

The first approach we adopted to the design of short-time stimuli is based on the specification of basic signal properties affecting the frequency content, duration, amplitude modulation, and amplitude temporal envelope of the signal, as prescribed in case example 1 of Section 3.6.

A basic harmonic signal $s(t)$ is composed of a sum of a fundamental sinusoidal component $s_0(t) = \sin(\omega_0 t)$ at angular frequency ω_0, and N harmonic components at frequencies $\omega_k = k\omega_0$. The desired waveform can be efficiently generated by a standard technique from audio synthesis, known as Chebychev waveshaping [1]. In it, $s_0(t)$ is passed through a static nonlinear transfer function $w(x) = \sum_k a_k T_k(x)$, where a_k are the desired harmonic amplitudes and $T_k(x)$ is the kth Chebychev polynomial. The result gives the desired harmonic signal $s(t)$ as

$$s(t) = w(s_0(t)) = \sum_k a_k T_k(s_0(t)) \tag{1}$$

$$= \sum_{k=0}^{N-1} a_k \sin(\omega_k t). \tag{2}$$

Using this method, it is simple to design a waveform whose energy lies within the target frequency band, centered on approximately 250 Hz.

The complete short-time waveform $y(t)$ is obtained by multiplying the harmonic waveform $s(t)$ by an amplitude envelope function $e(t)$ with desired starting time t_s, duration T, amplitude A, and attack and decay times t_a and t_d. Finally, it is subjected to form of roughening, via amplitude modulation with a signal $r(t) = 1 + m_r \sin(2\pi f_r t)$ of frequency f_r and modulation depth m_r. That is, $y(t) = s(t)\, e(t)\, r(t)$.

o generate the stimuli used in the experiment (Section 4), $N = 10$ harmonics were used, with fundamental frequency f_0 between 30 and 70 Hz, and a modulation frequency of $f_r = 25$ Hz. The harmonic amplitudes a_k and the roughness m_r, were set heuristically by the designer, as was the base frequency f_0. Waveforms used in the experiments are shown in Fig. 6a.

3.7.2 Synthesis of Virtual Impact Transients

The second method used to synthesize short-time stimulus components is based on virtual impact events, consisting of transient forces generated by the collision of two virtual elements: an exciting, but nonresonant object, termed the

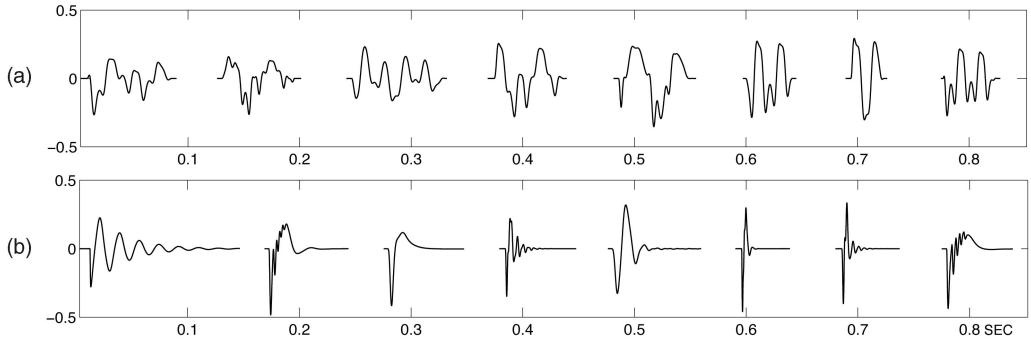


Fig. 6. Waveforms used to generate stimuli used in the evaluation, ordered as reported on in the experiment. (a) Those generated using waveshaping synthesis. (b) Those generated using virtual impact transient model. The final stimulus sets result upon superimposing the rhythmic patterns described in Section 3.7.

Authorized licensed use limited to: McGill University. Downloaded on March 08,2010 at 14:24:42 EST from IEEE Xplore. Restrictions apply.

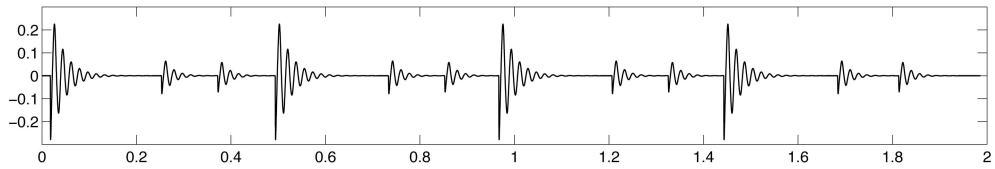


Fig. 7. Example of a stimulus used in the experiment, resulting from the model settings that produced the first "impact" waveform of Fig. 6 in combination with the first rhythmic pattern of Fig. 5.

hammer, and a resonant object, termed the sounding object [27]. The impact force $F(t)$ is generated by an efficient, physically based synthesis model obtained from a simplified version of Hertz' law, known as the Hunt-Crossley model [13]. It is obtained by numerically integrating the differential equation

$$F(t) = kx(t)^{\alpha} - \lambda x(t)^{\alpha}\dot{x}(t). \tag{3}$$

Here, $x(t)$ is the compression displacement and $\dot{x}(t)$ is the compression velocity. The impact force has parameters governing stiffness k, dissipation λ, and contact shape α. This force is provided as an input to the resonant sounding object. The latter is modeled as a set of N resonant filters operating in parallel, with a combined impulse response given by

$$y(t) = \sum_{i=1}^{N} a_i e^{-b_i t} \sin(\omega_i t), \tag{4}$$

and determined by a set of parameters governing the modal amplitudes a_i, decay rates b_i, and resonant angular frequencies ω_i. An impact event is synthesized by initializing (3) with the desired velocity v_I of impact, and subsequently, integrating the composite system in time. See [27] for a more detailed discussion, including an overview of the numerical implementation we have used, which is based on the open source Sound Design Toolkit from the University of Verona.

The impact event transients used in the evaluation part of this paper were synthesized with two resonant modes ($N = 2$). The transients themselves exhibit more complex frequency spectra than this might suggest, due to dynamic coupling of the nonlinear impact model with the modal resonant object [27]. Exemplary waveforms synthesized for the evaluation using this method are shown in the bottom of Fig. 6.

4 Evaluation

An evaluation of the methods for vibrotactile information display via floor surfaces that were described above has been conducted. The aim was to determine whether vibrotactile signals designed in the two ways described could be distinguished when presented via a floor surface. Two sets of eight signals were designed, one based on the waveshaping method and the other based on the virtual impact transient method. (We refer to these as sets WS and IT, respectively.) Both sets were designed using the same eight underlying rhythmic patterns (shown in Fig. 5) as noted in Section 3.7.

The vibrotactile stimulus sets differed, therefore, only in the note-level signals from which they were composed. Each tacton from the waveshaping synthesis (WS) set was assigned distinct values of the fundamental frequency f_0, ten harmonic amplitudes a_k, and amplitude modulation depth m_r. Stimuli from the impact transient set were each designed with different values of the hardness k, contact shape α, and of the frequencies ω_i and decay factors b_i of the two modes. The latter were chosen so that the frequency spectrum of the stimulus had most of its energy near 250 Hz.

Another significant difference between the two sets of stimuli is that the notes from the waveshaping synthesis set differ in duration according to the length of the note in the bar, while in the impact transient set, the amplitude envelope of each note decays exponentially as determined by the modal decay factors b_i. However, because the relevant decay times were generally short, as can be seen in row 2 of Fig. 6, the durations were shorter, in nearly all cases, than in the WS set, and they were independent of the note duration specified in the rhythmic pattern.

The experiment assessed both the rate of correct identification and learning for the two sets. In descriptive comments received during pretesting, we observed a tendency for people to describe the IT stimulus set in terms of familiar impact events. As a result, we hypothesized that these stimuli might prove more recognizable, despite the smaller number of parameters that were used to distinguish them, and the relative lack of note duration information in the patterns.

4.1 Methodology

A total of 24 people aged between 20 and 39 years took part in the experiment. All gave their informed consent in agreement with university ethics guidelines. Twelve of them were male and 12 were female. Fourteen of them were university students. The experiment was designed with a single independent factor, resulting in each group of 12 participants being presented with the task of identifying stimuli from one set. A between-groups design was used because the rhythmic content in the two stimulus sets is identical, so participants exposed to one set would be expected to perform better than otherwise expected with the second set.

Each participant was given a pair of hard soled men's dress shoes in his or her size to wear during the experiment. Apart from size differences, all the shoes were identical. The amplitude of vibration of the tile was adjusted as needed for each participant, depending on his or her ability to detect a reference vibrotactile noise signal. At the beginning of the session, participants received instructions, together with an explanation and demonstration of the experimental interface

Authorized licensed use limited to: McGill University. Downloaded on March 08,2010 at 14:24:42 EST from IEEE Xplore. Restrictions apply.

and the operation of the display. At all subsequent stages (except during pauses), participants were required to stand on the actuated floor surface, and to wear closed-ear headphones playing pink noise at a loudness sufficient to mask the (generally low-level) sounds produced by the apparatus. The software application used during the experiment ran on a personal computer, implementing the respective design method from Section 3.7. In addition to the floor tile, the interface consisted of a graphical user interface with numbered buttons, one for each icon, presented on a computer monitor. Participants selected items in the graphical user interface using a mouse.

The experiment was based on absolute identification of the eight stimuli from the respective set, with a unique correct response required for each (the numerical identifier of the icon, ranging from 1 to 8). The same stimuli were used for all participants within a given group, but they were presented in random order in each session of the experiment, with a different ordering presented to each participant.

After an introduction to the device and interaction method, participants were given five minutes of self-guided learning. During this time, they could select a numerical identifier and be presented with the stimulus corresponding to their selection. The rest of the experiment was divided into six sessions. During each session, all stimuli were presented twice. Thus, overall, each participant was asked to identify every vibrotactile stimulus from their set a total of 12 times. Each session took less than about four minutes to be completed. Participants were allowed a short break between sessions, but in most cases, preferred to continue so as to avoid forgetting the learned associations. Participants were presented with stimuli sequentially. At each presentation, they could press a button to play the stimulus up to four times before supplying a response. Feedback, in the form of the correct stimulus ID, was provided after each response was given. As in previous studies [11], the reason for providing feedback was to facilitate the assessment of recognition after learning and rate of learning throughout the experiment.

4.2 Results and Discussion

A log of the stimuli and responses was recorded by the application throughout the experiments. Participants were also interviewed following the experiment.

The stimulus sets each included eight items, near the limit of what participants might be expected to retain in working memory. Training was limited, consisting of a maximum of, on average, 20 reinforced presentations of each stimulus at the end of the final session. The stimuli were not assigned any mnemonic, other than an arbitrary numerical index, that could be used to remember them. Although introducing a semantic association to each stimulus (for example, the name of a meaningful event) can be used to improve recall, good design of display and stimuli is a prerequisite to good performance [33]. Since our intention was to assess the suitability of the display methods behind these two stimulus sets, we wished to avoid introducing any external effects that might be associated to the choice of semantic labels.

The mean rate of correct identification after six sessions of enforced learning, averaged between all participants (including both stimulus sets) was 61 percent, with a standard deviation of 21 percent. Chance performance would correspond to 12.5 percent. Postlearning identification rates for participants in the impact transient group and for those in the waveshaping synthesis group are shown in Fig. 8. Four of the 12 participants in the IT group were able to achieve 80 percent or better correct identification after less than 20 minutes of enforced learning. One of the 12 participants in the WS group was able to reach this level of performance.

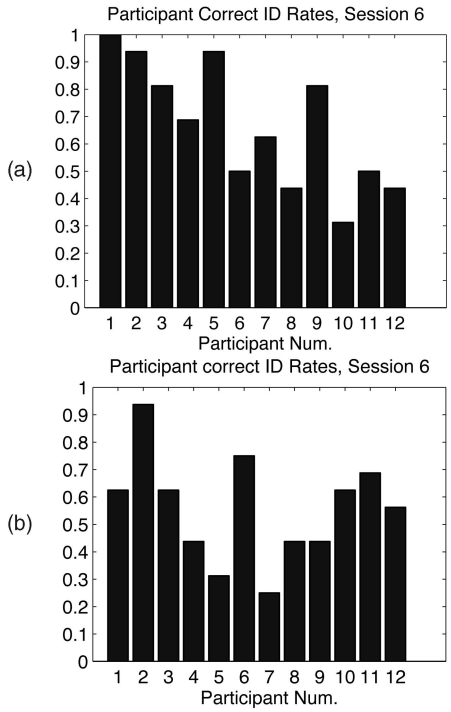


Fig. 8. (a) Correct identification rate for each of the 12 participants in group assigned to stimulus set IT, after six sessions of assessment with enforced learning. (b) Similar rates for the WS set group.

The results obtained appear to be comparable to published results on absolute identification of vibrotactile stimuli via manual interfaces after short periods of learning. For example, Enriquez et al. reported average identification performance of 73 percent (versus the expected chance performance rate of 33 percent) after an average of 20 minutes of learning [5].

The confusion patterns for the stimuli in each group of the experiment are shown in Fig. 9, averaged between all sessions and participants. The least confused stimulus was identified at an average rate of 80 percent, while the most confused stimulus was identified at an average rate of only 25 percent. A close comparison of the confusions within each stimulus set with the stimulus properties themselves did not reveal any easily discernible feature of the rhythms or the short-time waveforms that caused them to be confused. Nonetheless, the confusion patterns for the two sets appear to be relatively distinguished from one another, which suggests that the short-time features were perceived differently. Performance varied considerably between participants as the experiment proceeded, with a few showing consistently high improvement between sessions, while

Authorized licensed use limited to: McGill University. Downloaded on March 08,2010 at 14:24:42 EST from IEEE Xplore. Restrictions apply.

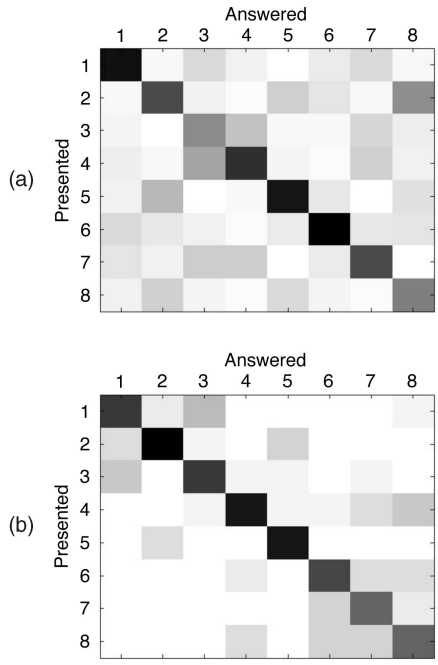


Fig. 9. (a) Confusion pattern within the set of eight WS stimuli in session 6, averaged between participants. (b) The same data for the set of eight IT stimuli.

others showed nearly none. A summary of the mean correct identification rate using each icon set, averaged between all sessions, is provided in Fig. 11. Mean performance on the impact transient set was higher than on the waveshaping synthesis set, but the level of significance is marginal (ANOVA 1-way $p = 0.06$). Mean correct identification rates in each enforced learning session are shown in Fig. 10. The mean correct identification rate after each of the first two training sessions was significantly higher for the IT set ($p = 0.01$ and 0.02, respectively), but thereafter, the results lacked significance ($p > 0.1$ in all cases).

The IT group approached its peak performance with far fewer presentations of the stimuli than was the case for the

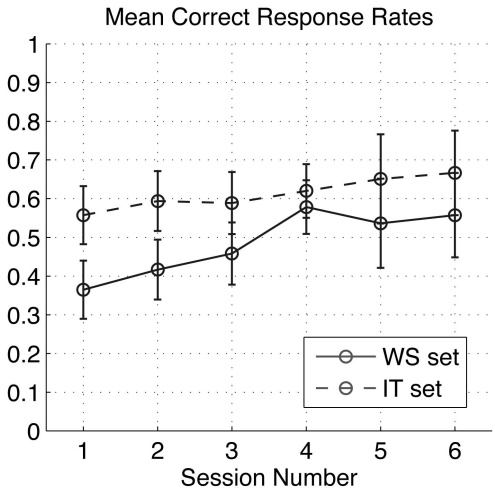


Fig. 10. Mean correct identification rates in each session. The error bars show 95 percent confidence intervals.

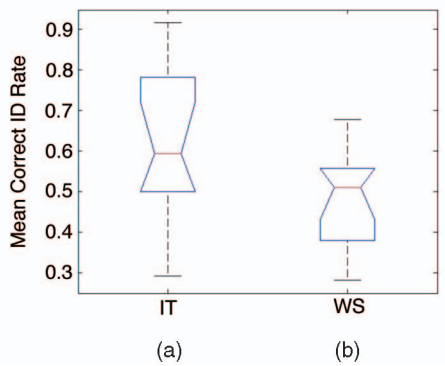


Fig. 11. Box plot summary of the mean correct identification rate for the (a) IT and (b) WS stimulus sets.

WS group. Concretely, after session 1, the IT group achieved more than 80 percent of the performance it reached in session 6, a rate that the WS group would not achieve until session 3. This suggests that stimuli similar to the IT set might be advantageous for applications in which little training would be expected, although further investigation of this is needed. Examples of such applications could include displays in public spaces, such as crosswalk indicators.

4.2.1 Subjective Ratings and Comments

Participants completed questionnaires after the experiment, rating the task difficulty, subjective properties of the stimuli, providing descriptive feedback regarding the strategies they adopted. Participants were asked to rate the difficulty of the first task session and the last session on a scale from 1 (very easy) to 5 (very hard). As expected, the first task session was rated as significantly harder than the last session (mean 3.96 versus 3.04, with $p = 0.04$).

An analysis of the rated difficulty of the first session (Anova one-way), with the stimulus set type as the independent factor, revealed the WS set was perceived as initially harder than the IT set (mean 4.4 versus 3.5, with $p = 0.02$). However, there were no statistically significant differences between the rated difficulties of the stimulus sets during the last session. This echoes what was seen in the performance data: the IT stimulus set was more identifiable than the WS set at the beginning of the experiment, but the difference faded before the end of the experiment. Based on participant comments, cognitive fatigue may have been a factor here.

The study also aimed to assess the level of comfort to users of a vibrotactile display employing these stimuli. Participants were asked to rate the vibrotactile stimuli on a five-point scale from 5 (comfortable) to 1 (uncomfortable). The average rating among all participants was 3.5 with a standard deviation of 1.1. No participant rated the stimuli as uncomfortable, and no significant differences in comfort ratings were found between the two groups. Some of them noted becoming uncomfortable with standing in the same place and posture during each session.

During subsequent interviews, several participants reported that the association part of the task, which required them to learn the numerical indices of the stimuli, was the hardest part of the experiment, and that they experienced

Authorized licensed use limited to: McGill University. Downloaded on March 08,2010 at 14:24:42 EST from IEEE Xplore. Restrictions apply.

cognitive fatigue as the experiment wore on. Some participants suggested the task may have been facilitated by the addition of semantic information or a nonnumerical mnemonic symbol (e.g., an animal name). As discussed above, this observation may hold potential for aiding the association part of the task, and, while it would certainly make sense for an application designer to consider, it does not necessarily speak of the design of the stimuli themselves.

4.2.2 Description of the Cause of the Stimulus

In a separate study, 10 people who did not participate in the forgoing experiment were presented with the 16 stimuli from the sets used above, in random sequential order, and were asked to provide, using their own words, a short label describing what caused the vibration. Sensory conditions were the same as in the first experiment. The most common response was "unknown" (about 18 percent of IT set labels and 40 percent of WS set labels). One hundred and thirteen other responses were obtained, 72 of which were unique, if often very similar (e.g., "door knock" and "angry door knock"). Among these labels, most were related to familiar impact events, including: "door knock," "horse gallop," "hammer," "drum," "foot falls overhead," and "tapping." To give one example, the second IT stimulus was described by seven of 10 participants as caused by a "knock" or "knocking," and by another two as originating from a "hammer." There appeared to be a greater tendency for IT set stimuli to be attributed to such events than was the case for WS set stimuli (mean 58 percent versus 18 percent, with $p = 0.0002$). Here, however, we used our own judgement in deciding whether a label referred to an impact. Several labels were associated only to the WS set, including: "bass guitar," "cellphone," "machinery," and "car start."

While it seems likely that there are differences in causal interpretation between the two groups of stimuli, more research is clearly needed. It is tempting to suggest that such differences, if established, might be linked to the effectiveness of the stimuli at communicating information. However, analogous research in the auditory display domain, comparing the effectiveness of melodically designed "earcons" with representational "auditory icons," has shown that this issue is far from simple or straightforward [2].

5 Conclusion

Floor surfaces already play roles that intimately link tactile sensation in the feet to our everyday activities. The use of such surfaces as displays for active haptic communication appears to be a valuable idea, and one that naturally extends accepted areas of practice in the design of haptic information for ground surfaces. We have highlighted the range of roles that such an interface might play in future applications, and have pointed to a few potential end user applications that illustrate these roles.

Distinguishing advantages of these displays include their applicability to a wide range of settings and environments, their ready accessibility to anyone on foot, and their potential for seamlessly presenting dynamic information linked to a specific location, navigational, or locomotor task.

We have outlined design considerations for haptic communication through floor surfaces and detailed a device and set of interaction studies that have been developed to embody these. The technological feasibility of such devices is high and the cost of prototyping them is low.

A characterization in terms of temporal interaction properties was used to organize different possible interaction designs presented through the floor interface, and three demonstrations that we have developed were described and situated within it. Stimulus design methods based on signal-level specifications were highlighted, by close analogy to prior work on vibrotactile information display. Design methods based on physical metaphors, such as virtual impact events or structured materials, were also presented.

The experimental results support the identifiability of vibrotactile messages conveyed through floor surfaces and point toward the adaptation of design guidelines developed for other modes of haptic display to this setting. They also suggest that physical metaphors may hold promise for improving the design of stimuli for such an information display, although further research is needed on this subject.

Acknowledgments

The authors gratefully acknowledge support from the MDEIE of Quebec for the EU FP7 project NIW (no. 222107). The first author acknowledges support from the EU FP6 project CLOSED (no. 29085) and the ESF COST Action on Sonic Interaction Design (no. IC0601).

References

[1] D. Arfib, "Digital Synthesis of Complex Spectra by Means of Multiplication of Non-Linear Distorted Sine Waves," *J. Audio Eng. Soc.*, vol. 27, no. 10, pp. 757-779, 1979.

[2] S.A. Brewster, "Non-Speech Auditory Output," *The Human-Computer Interaction Handbook*, A. Sears and J.A. Jacko, eds., second ed., Lawrence Erlbaum Assoc., 2008.

[3] L.M. Brown, S.A. Brewster, and H.C. Purchase, "Tactile Crescendos and Sforzandos: Applying Musical Techniques to Tactile Icon Design," *Proc. ACM 2006 Conf. Human Factors in Computing Systems*, 2006.

[4] C. Cadoz, "Instrumental Gesture and Musical Composition," *Proc. Int'l Computer Music Conf. (ICMC)*, 1988.

[5] M. Enriquez, K.E. MacLean, and C. Chita, "Haptic Phonemes: Basic Building Blocks of Haptic Communication," *Proc. Eighth Int'l Conf. Multimodal Interfaces*, 2006.

[6] A.R. Ferber, M. Peshkin, and J.E. Colgate, "Using Haptic Communications with the Leg to Maintain Exercise Intensity," *Proc. 16th IEEE Int'l Conf. Robot and Human Interactive Comm. (RO-MAN)*, 2007.

[7] A. Gallace, H.Z. Tan, and C. Spence, "The Body Surface as a Communication System: The State of the Art After 50 Years," *Presence: Teleoperators and Virtual Environments*, vol. 16, no. 6, pp. 655-676, 2007.

[8] G.A. Gescheider and R.T. Verrillo, "Vibrotactile Frequency Characteristics as Determined by Adaptation and Masking Procedures," *Sensory Functions of the Skin of Humans*, D.R. Kenshalo, ed., Plenum, 1978.

[9] V. Hayward and K. Maclean, "Do It Yourself Haptics, Part I," *IEEE Robotics and Automation Magazine*, vol. 14, no. 4, pp. 88-104, Dec. 2007.

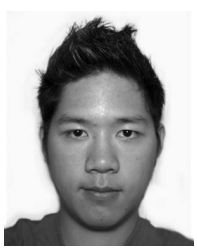
[10] J. Hijmans, J. Geertzen, B. Schokker, and K. Postema, "Development of Vibrating Insoles," *Int'l J. Rehabilitation Research*, vol. 30, no. 4, pp. 343-345, 2007.

[11] E. Hoggan and S. Brewster, "Designing Audio and Tactile Crossmodal Icons for Mobile Devices," *Proc. ACM Int'l Conf. Multimodal Interfaces (ICMI)*, 2007.

Authorized licensed use limited to: McGill University. Downloaded on March 08,2010 at 14:24:42 EST from IEEE Xplore. Restrictions apply.

[12] J. Hollerbach, "Locomotion Interfaces and Rendering," *Haptic Rendering: Foundations, Algorithms and Applications,* M. Lin and M. Otaduy, eds., A K Peters, Ltd., 2008.
[13] K.H. Hunt and F.R.E. Crossley, "Coefficient of Restitution Interpreted as Damping in Vibroimpact," *ASME J. Applied Mechanics,* vol. 42, no. 2, pp. 440-445, 1975.
[14] J. Inglis, P. Kennedy, C. Wells, and R. Chua, "The Role of Cutaneous Receptors in the Foot," *Advances in Experimental Medicine and Biology,* vol. 508, pp. 111-117, 2002.
[15] H. Iwata, "Haptic Interface," *The Human-Computer Interaction Handbook,* A. Sears and J.A. Jacko, eds., second ed., Lawrence Erlbaum Assoc., 2008.
[16] L.A. Jones and N.B. Sarter, "Tactile Displays: Guidance for Their Design and Application," *Human Factors,* vol. 50, no. 1, pp. 90-111, 2008.
[17] P.M. Kennedy and J.T. Inglis, "Distribution and Behaviour of Glabrous Cutaneous Receptors in the Human Foot Sole," *J. Physiology,* vol. 583, no. 3, pp. 995-1002, 2002.
[18] Y. Kobayashi, R. Osaka, T. Hara, and H. Fujimoto, "How Accurately People Can Discriminate the Differences of Floor Materials with Various Elasticities," *IEEE Trans. Neural Systems and Rehabilitation Eng.,* vol. 16, no. 1, pp. 99-105, Feb. 2008.
[19] Y. Kobayashi, T. Takashima, M. Hayashi, and H. Fujimoto, "Gait Analysis of People Walking on Tactile Ground Surface Indicators," *IEEE Trans. Neural Systems and Rehabilitation Eng.,* vol. 13, no. 1, pp. 53-59, Mar. 2005.
[20] A.W. Law, B.V. Peck, Y. Visell, P.G. Kry, and J.R. Cooperstock, "A Multi-Modal Floor-Space for Displaying Material Deformation Underfoot in Virtual Reality," *Proc. IEEE Int'l Workshop Haptic Audio Visual Environments and Their Applications,* 2008.
[21] K.E. MacLean, "Foundations of Transparency in Tactile Information Design," *IEEE Trans. Haptics,* vol. 1, no. 2, pp. 84-95, July-Dec. 2008.
[22] *Generic Intelligent Driver Support,* J.A. Michon, ed. CRC Press, 1993.
[23] A. Priplata, J. Niemi, M. Salen, J. Harry, L.A. Lipsitz, and J.J. Collins, "Noise-Enhanced Human Balance Control," *Physical Rev. Letters,* vol. 89, no. 23, pp. 238101-1-238101-4, 2002.
[24] A.A. Priplata, B.L. Patritti, J.B. Niemi, R. Hughes, D.C. Gravelle, L.A. Lipsitz, A. Veves, J. Stein, P. Bonato, and J.J. Collins, "Noise-Enhanced Balance Control in Patients with Diabetes and Patients with Stroke," *Annals of Neurology,* vol. 59, no. 1, pp. 4-12, 2006.
[25] E. Ribot-Ciscar, J.P. Roll, and G.M. Gauthier, "Comparative Effects of Whole Body Vibration on Sensorimotor Performance Achieved with a Mini-Stick and a Macro-Stick in Force and Position Control Modes," *Aviation, Space, and Environmental Medicine,* vol. 57, pp. 792-799, 1986.
[26] E. Ribot-Ciscar, J.P. Vedel, and J.P. Roll, "Vibration Sensitivity of Slowly and Rapidly Adapting Cutaneous Mechanoreceptors in the Human Foot and Leg," *Neuroscience Letters,* vol. 104, pp. 130-135, 1989.
[27] *The Sounding Object,* D. Rocchesso and F. Fontana, eds. Edizioni di Mondo Estremo, 2003.
[28] J.P. Roll and J.C. Gilhodes, "Proprioceptive Sensory Codes Mediating Movement Trajectory Perception: Human Hand Vibration-Induced Drawing Illusions," *Canadian J. Physiology and Pharmacology,* vol. 73, no. 2, pp. 295-304, 1995.
[29] T. Rosburg, "Tactile Ground Surface Indicators in Public Places," *Human Haptic Perception,* M. Grunwald, ed., Birkhäuser, 2008.
[30] J. Rovan and V. Hayward, "Typology of Tactile Sounds and Their Synthesis in Gesture-Driven Computer Music Performance," *Trends in Gestural Control of Music,* M. Wanderley and M. Battier, eds., Editions IRCAM, 2000.
[31] A.F. Rovers and H. van Essen, "Guidelines for Haptic Interpersonal Communication Applications," *Virtual Reality,* vol. 9, pp. 177-191, 2006.
[32] A. Tax, B. van Wezel, and V. Dietz, "Bipedal Reflex Coordination to Tactile Stimulation of the Sural Nerve During Human Running," *J. Neurophysiology,* vol. 73, no. 5, pp. 1947-1964, 1995.
[33] D. Ternes and K.E. MacLean, "Designing Large Sets of Haptic Icons with Rhythm," *Proc. Eurohaptics,* 2008.
[34] M. Trulsson, "Microreceptive Afferents in the Human Sural Nerve," *Experimental Brain Research,* vol. 137, pp. 111-116, 2001.
[35] J.B.F. van Erp and M.M.A. Spape, "Distilling the Underlying Dimensions of Tactile Melodies," *Proc. Eurohaptics,* 2003.
[36] R. Velazquez, "On-Shoe Tactile Display," *Proc. IEEE Int'l Workshop Haptic Audio Visual Environments and Their Applications,* 2008.
[37] Y. Visell, "Tactile Sensory Substitution: Models for Enaction in HCI," *Interacting with Computers,* vol. 21, pp. 38-53, 2009.
[38] Y. Visell, J. Cooperstock, B.L. Giordano, K. Franinovic, A. Law, S. McAdams, K. Jathal, and F. Fontana, "A Vibrotactile Device for Display of Virtual Ground Materials in Walking," *Proc. Eurohaptics,* 2008.
[39] W.H. Warren and R.R. Verbrugge, "Auditory Perception of Breaking and Bouncing Events: A Case Study in Ecological Acoustics," *J. Experimental Psychology: Human Perception and Performance,* vol. 10, no. 5, pp. 704-712, 1984.
[40] S. Weinstein, "Intensive and Extensive Aspects of Tactile Sensitivity as a Function of Body Part, Sex, and Laterality," *The Skin Senses,* D.R. Kenshalo, ed., Thomas, 1968.
[41] C. Wells, L.M. Ward, R. Chua, and J.T. Inglis, "Regional Variation and Changes with Ageing in Vibrotactile Sensitivity in the Human Footsole," *J. Gerontology A: Biological Sciences and Medical Sciences,* vol. 58, no. 8, pp. B680-B686, 2003.
[42] A. Wilska, "On the Vibrational Sensitivity in Different Regions of the Body Surface," *Acta Physiologica Scandinavica,* vol. 31, pp. 285-289, 1954.
[43] E.P. Zehr, R.B. Stein, and T. Komiyama, "Function of Sural Nerve Reflexes during Human Walking," *J. Physiology,* vol. 507, pp. 305-314, 1998.

Yon Visell received the BA degree in physics from Wesleyan University and the MA degree in physics from the University of Texas at Austin in 1999. Currently, he is working toward the PhD degree in electrical and computer engineering at the Centre for Intelligent Machines, McGill University, Montreal, Canada. His research concerns physically-based haptic synthesis and interface design for vibrotactile display. He is a Canadian national delegate to the ESF COST Action on Sonic Interaction Design, and was a local coordinator for the European project CLOSED at the University of Applied Arts, Zurich. For two years, he was at Loquendo, San Francisco, California, conducting research on automatic speech recognition, and was engaged in sonar signal processing at ARL, Austin, Texas. Subsequently, he was employed at Ableton AG, Berlin, Germany, where he led audio signal processing development for an award-winning professional music production suite. In 2004, he co-founded the art technology organization Zero-Th, Pula, Croatia.



Alvin Law received the BASc degree in engineering physics in 2007 from the University of British Columbia, Vancouver, Canada. Since late 2007, he has been working toward the MASc degree at McGill University, in the area of haptic interaction through augmented ground surfaces. In 2004, he was a research assistant at the UBC SSP Laboratory on the Solar Canopy project. The following summer, he was employed at ATR Institute International in Japan on the Robovie project. During 2006, he was a research assistant at the National University of Singapore studying electromagnetic metamaterials.

Jeremy R. Cooperstock received the PhD degree in electrical and computer engineering from the University of Toronto in 1996. He is currently an associate professor in the Department of Electrical and Computer Engineering, a member of the Centre for Intelligent Machines, and a founding member of CIRMMT at McGill University. He directs the Shared Reality Laboratory and leads the technical development of the Ultravideoconferencing system, for which he was recognized with an award for Most Innovative Use of New Technology from ACM/IEEE Supercomputing and a Distinction Award from the Audio Engineering Society. His research interests focus on computer mediation to facilitate high-fidelity human communication and the underlying technologies that support this goal. He chairs the Audio Engineering Society (AES) Technical Committee on Network Audio Systems and is an associate editor of the *Journal of the AES*.

Authorized licensed use limited to: McGill University. Downloaded on March 08,2010 at 14:24:42 EST from IEEE Xplore. Restrictions apply.

YEAR 3

Six Degrees-of-Freedom Haptic Interaction with Fluids

Gabriel Cirio, Maud Marchal, Sébastien Hillaire, and Anatole Lécuyer

Abstract—We often interact with fluids in our daily life, either through tools such as when holding a glass of water or directly with our body when we swim or we wash our hands. Multimodal interactions with virtual fluids would greatly improve the simulations realism, particularly through haptic interaction. However, achieving realistic, stable and real-time force feedback from fluids is particularly challenging. In this work, we propose a novel approach that allows real-time 6 Degrees of Freedom (DoF) haptic interaction with fluids of variable viscosity. Our haptic rendering technique, based on a Smoothed-Particle Hydrodynamics physical model, provides a realistic haptic feedback through physically-based forces. 6DoF haptic interaction with fluids is made possible thanks to a new coupling scheme and a unified particle model, allowing the use of arbitrary-shaped rigid bodies. Particularly, fluid containers can be created to hold fluid and hence transmit to the user force feedback coming from fluid stirring, pouring, shaking and scooping, to name a few. Moreover, we adapted an existing visual rendering algorithm to meet the frame rate requirements of the haptic algorithms. We evaluate and illustrate the main features of our approach through different scenarios, highlighting the 6DoF haptic feedback and the use of containers.

Index Terms—6DoF haptic interaction, computational fluid dynamics, smoothed-particle hydrodynamics, rigid bodies

1 Introduction

Virtual Reality technologies intend to immerse users inside a 3D synthetic world simulated in real-time by a computer. In a typical virtual reality setup, the user is provided with stimulations and sensory feedback on multiple senses such as vision, audition and/or touch (or haptic sense) by means of advanced displays. As an example, in surgical simulations, medical trainees can be provided with visual and haptic feedback that enable to better perceive contacts between the manipulated tools and the surrounding organs.

The computation of haptic feedback in real-time, also known as "haptic rendering", is still an open and active research area. Surprisingly, haptic interaction with fluids has been scarcely studied. Indeed, most current haptic simulations involve only rigid or deformable bodies. However, we often interact with fluids in our daily life, either through tools such as when holding a glass of water or directly with our body when we swim or we wash our hands. Fluids are also found in many applications such as for industrial or medical manipulations - involving for instance blood flow and natural liquids. Enabling haptic feedback in the interaction with fluids, besides allowing more realistic simulations, would enable a wide range of novel simulation scenarios and applications.

The haptic interactive simulation of fluids is particularly challenging, especially to achieve realistic, stable and real-time force-feedback using physically-based models. To simulate interactions between fluids and rigid bodies within haptic rendering, previous studies have proposed pre-computed ad-hoc algorithms [1], approaches featuring only 3 Degrees of Freedom (DoF) and non-viscous fluids [2], or implementations restricted to simple object shapes and small amounts of fluid [3]. Thus, as for today, there is a lack of models and rendering techniques handling complex 6DoF haptic interactions with viscous fluids in real-time.

In this work, we propose a novel approach for real-time 6DoF haptic interaction with viscous fluids through arbitrary shaped rigid bodies and 6DoF haptic devices. It represents a significant leap forward in interaction possibilities compared to previous work on haptic interaction with fluids. It is the first approach to allow a full 6DoF haptic interaction with viscous fluids. Until now, real-time interaction was restricted to simply "swiping" the surface of a fluid volume with simple objects, dramatically reducing the interaction possibilities when compared to real-life conditions. In this work, we show how many rich and complex fluid manipulations are now easily achieved, with stable force and torque outputs. Interaction is no longer limited to 3 degrees of freedom: not only torque feedback plays a major role in providing a compelling feeling of realism, it also opens an entirely new horizon of interaction techniques in Virtual Reality. Among these are the use of concave containers to hold fluid and hence transmit to the user force feedback coming, for example, from fluid stirring, pouring, shaking and scooping, as well as the inertia of the fluid inside the container.

This work also introduces the Smoothed-Particle Hydrodynamics (SPH) model into the fluid haptic simulation, and, to the best of our knowledge, to the haptic realm in general, yielding positive results. Many issues

- *G. Cirio, M. Marchal, S. Hillaire and A. Lécuyer are with INRIA Rennes, France.*
  E-mail: {gcirio, mmarchal, shillaire, alecuyer}@irisa.fr
- *S. Hillaire is also with Orange Labs, Rennes, France.*

of major concern for the haptic community are avoided in our work through the use of the SPH model, such as the generation of smooth and stable haptic forces, the transparent handling of multiple contact points, the parallelizability of computations and their scalability, the seamless use of arbitrary-shaped objects, and the computation of a fully dynamic world. By showing that haptic forces computed through the SPH model are particularly well suited for the interaction with fluids, haptic interaction with other media can now be imagined for the SPH model.

Hence, our major contributions are:

- **A novel 6DoF approach for the haptic interaction with viscous fluids**,
- **The introduction of the SPH model for haptic fluid computations**

In addition, other technical contributions of our paper concern:

- a generic scheme for the haptic coupling of any number of devices and rigid bodies,
- the GPU acceleration of the underlying fluid and rigid body model to achieve real-time speeds,
- the adaptation of an existing graphic rendering algorithm to meet high frame rate requirements

The paper is organized as follows: section 2 provides a summary of related work. In section 3 we describe our novel haptic rendering technique, initially coupled to a single particle and with 3DoF haptic feedback. The algorithm is extended in section 4 for 6DoF haptic rendering through the use of a unified particle model for rigid body interaction with fluids. Section 5 describes how we adapt a screen space fluid graphic rendering algorithm to meet the frame rate requirements of our rendering technique. We evaluate and illustrate our approach in section 6, highlighting its main features and interaction possibilities through different scenarios. Particularly, fluid containers can be created to hold fluid and hence transmit to the user force feedback coming, for example, from fluid stirring, pouring, shaking and scooping. Strong forces such as fluid resistance and weight can be captured, as well as light forces like the inertia of the fluid inside the container. We discuss our results and the limitations of our technique in section 7, before concluding.

2 Related Work

2.1 Real-Time Fluid Simulation

Two main physical approaches based on Navier-Stokes equations can be distinguished [4]: the Eulerian approach, where space is subdivided in a grid and the different physical quantities are computed for each grid cell, and the Lagrangian approach, where the fluid is treated as a system of independent particles, each with a position and a velocity. Although each approach has been extensively studied, we focus on the Lagrangian approach, since the simulated fluid is not bounded by a grid, is faster to compute and better preserves small-scale details, allowing richer interactions scenarios, which is especially interesting for haptic feedback. On the other hand, compared to Eulerian approaches, they have a lower order spatial accuracy and some interesting behaviors can disappear in low-density regions. For a detailed description of the existing approaches for the simulation of fluids, we refer the reader to [4].

Smoothed-Particle Hydrodynamics [5] (SPH) techniques are very popular among Lagrangian simulations. With SPH, each particle gets its different physical quantities from an interpolation of its neighboring particles quantities. Specific kernels are used to obtain a smooth interpolation of values through the weighted contribution of neighbors according to their relative distance. From these quantities, the Navier-Stokes equations are used to compute the interaction forces between particles, and thus a new state for each particle. SPH simulations of fluids have been shown to be stable and fast enough to achieve interactive rates [6] [7].

Simulating physically-based fluids is a computationally expensive task, a critical factor for real-time applications. In order to accelerate the computation of the simulations, research has focused on the implementation of the SPH algorithm on GPU, being naturally parallel. Early attempts [8] [9] were not fully GPU-based. In [10], the entire simulation runs on GPU, by using a texture representation of a grid space subdivision for neighbor computation purposes. In [11], values are accumulated by computing the contribution of each particle *to* its neighbors, instead of computing the contribution *from* its neighbors. This results in faster computations, at the cost of a huge memory consumption, achieving interactive frame rates with up to 130,000 simulated particles.

2.2 SPH Fluid-Rigid Body Interaction

Realistic rigid-fluid interactions are required to create elaborate scenes, especially when dealing with arbitrary-shaped objects. In SPH simulations, collisions between fluid particles and rigid bodies are usually handled through penalty-based methods, by introducing forces opposite to the particle movement at the contact point [12]. These forces are derived from a penalty boundary potential in [13]. Lennard-Jones attraction and repulsion forces are used in [14] together with fluid viscosity forces to model the interaction between fluid particles and deformable polygonal meshes. Another approach to obtain solid-fluid interactions is to represent rigid bodies as a set of particles, as initially introduced in the Eulerian realm [15]. In an unified approach [16] [7], rigid body particles are simulated as fluid particles, using the same pressure and viscosity as between fluid particles. The mass and density values of each solid object particle are set according to the object properties. With the same particle-based representation of rigid-bodies, a more complex and accurate method is proposed in [17], called direct forcing. Control forces and

constraint equations are used in a predictor-corrector fashion, allowing different slip conditions and enforcing non-penetration.

With the exception of [14], the aforementioned algorithms have not been used with a real-time implementation.

2.3 Visual Rendering of SPH Fluids

Many techniques have been proposed to render fluids simulated through SPH. To this aim, an optimized marching cube method is proposed in [18]. However, the result looks blobby, and requires a finite grid representation of the fluid (whereas, compared to Eulerian fluids, an SPH simulation does not). To overcome this problem, the fluid mesh is computed in screen space [19]. To this aim, the front fluid surface depth is rendered from the point of view. This depth buffer is then smoothed using a binomial filter to attenuate the irregular look of the fluid. Finally, a mesh representing the fluid is generated in screen space using a marching square algorithm. This method is efficient but still requires an expensive mesh generation.

Recent methods proposed to render fluids in screen space without generating any meshes. In [20], per-pixel normals of the fluid surface in view space are computed using directly the fluid depth buffer, representing the front fluid surface. Thus, per-pixel lighting as well as environment reflection could be computed by rendering a single full screen quad. A similar method has been proposed in [21], where the fluid depth buffer is smoothed out by minimizing its curvature and the overall fluid appearance is improved using high frequency noise. Moreover, fluid thickness is taken into account in order to simulate light extinction through the medium. With a precomputed SPH fluid simulation, this method achieves real time performances when rendering 64,000 particles.

2.4 6DoF Haptic Rendering

There is a long history of research on 6DoF Haptic interaction with complex objects. It is a challenging task, mainly due to cost of collision detection computations. Attempts made by using convex primitives [22] or NURBS representations [23] greatly limit the number, size and complexity of the virtual interactive objects. Voxel-based haptic rendering [24], improved in [25] and [26], was the first technique to allow efficient 6DoF haptic rendering without shape or complexity constraints. Simple penalty forces are computed between the point-sampled dynamic tool and the voxelized static virtual environment. A sensation-preserving simplification algorithm [27] trades accuracy for speed through multiple levels of detail while preserving the interaction forces. A fast contact point tracking algorithm based on spatialized hierarchies is used in [28] for approximating the collision detection. It is combined with a slower algorithm for exact computations, and allows the use of moderately sized moving objects at haptic rates. In [29], the authors extend the god-object 3DoF technique [30], using continuous collision detection and a constraint-based approach, enforcing non-penetration and improving stiffness perception.

These algorithms are efficient when interacting with rigid and often static obstacles of the virtual environment. 6DoF haptic interaction with deformable models has also been addressed in the past [31] [32]. Fluids, however, have the particularity of being highly deformable and dynamic, with their shape varying at each time-step, often with unconnected components (such as drops). Hence, other haptic feedback algorithms and coupling schemes are required and used in the haptic interaction with fluids.

2.5 Haptic Rendering of Fluids

Recent work has explored the possibilities of haptic interaction with fluids. Hence, different techniques were used to lower the computational demands of the simulation in order to achieve acceptable frame rates for haptic interaction. An example is the pre-computation of the non-linear interaction forces between fluid and rigid bodies, as described in [1]. These forces are added to the linear forces computed in real-time, hence obtaining very high frame rates (around 500Hz), since the main computations are performed offline. However, since force models are ad-hoc, it dramatically limits the number of different interaction scenarios that can be simulated. The method is illustrated through a canoe simulation. Another approach is the extension of a 2D technique to 3D space, as shown in [33]. The fluid surface is modeled as a 15x15 mass-spring network, and a stack of 2D fluid layers is below the surface. The layer deformations produced by the haptic probe are transmitted to the other layers according to their depth and density. Since the algorithm is intended for low-end computers with a small computing power, there is a trade-off between the realism of the simulation and its speed, as well as a limited number of degrees of freedom.

Other techniques are based on existing Eulerian physically-based simulations of fluids. In [3], the authors propose a 6DoF interaction technique with an Eulerian viscous fluid. The two-way rigid-fluid interaction is achieved by discretizing the boundary of the haptic probe into the simulation grid. Forces and torques exerted on the probe are computed by summing the contributions of each edge of the probe boundary. The authors illustrate the technique through a painting application, with the fluid defined locally around the brush, hence allowing the painter to feel the paint, and achieving 40 to 70 frames per second (fps). However, interaction possibilities are limited due to the high computational cost of Eulerian approaches, requiring the use of simple probe shapes and small amounts of fluid. Another haptic interaction implementation with an Eulerian fluid has been recently shown in [2], where

the simulation is accelerated through GPU. The user is allowed to interact with smoke via a spherical probe at 30 to 75 fps. However, the viscosity of the fluid is not taken into account, and the implementation is limited to 3DoF with simple object shapes, considerably reducing the interaction possibilities.

Our fluid haptic rendering technique is based on Smoothed-Particle Hydrodynamics [5] for haptic and fluid simulation [6], and it is the first attempt to bring the SPH model to the haptic fluid realm. Other 6DoF techniques provide high rate algorithms, but are often limited to the interaction with static objects and a reduced number of contact points, which precludes their use with a fluid simulation. The different components of our approach are shown in Figure 1. Using a simple and efficient GPU implementation, with performances close to those reported in [11], our physical simulation reaches framerates suitable for haptic interaction with large volumes of fluid. In order to achieve 6DoF feedback, we improve a unified particle-based approach [7] by bringing it to high speeds. We use this force computation approach combined with a Virtual Coupling mechanism [34] to provide 6DoF haptic coupling. Visual rendering is made possible by adapting a fluid graphic rendering algorithm [21] based on meshless screen space rendering to meet the frame rate requirements of our rendering technique.

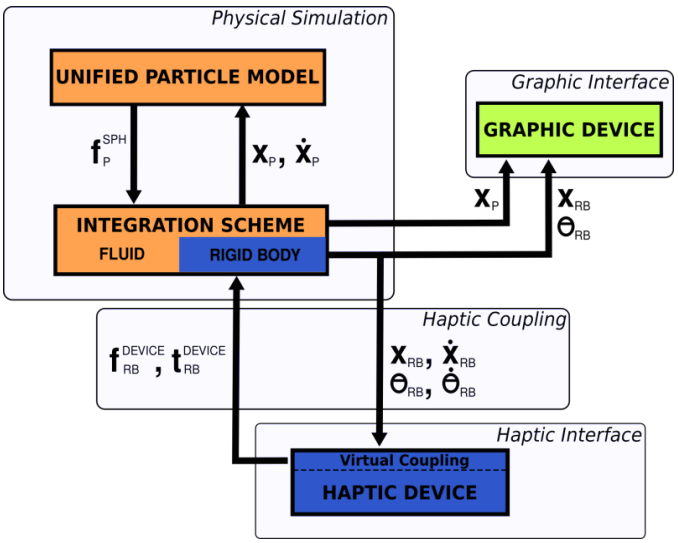


Fig. 1. Simulation overview. The unified particle model computes the different interaction forces between SPH particles P, $\mathbf{f}_P^{SPH}$, including haptic feedback forces. Then, rigid body dynamics are applied to the coupled rigid body RB, taking into account the force $\mathbf{f}_{RB}^{DEVICE}$ and torque $\mathbf{t}_{RB}^{DEVICE}$ generated by the 6DoF haptic device through Virtual Coupling. The new linear and angular velocity ($\dot{X}_{RB}, \dot{\theta}_{RB}$) and position ($X_{RB}, \theta_{RB}$) are sent to the haptic device.

3 Haptic Rendering of Fluids

In this section, we describe our novel haptic rendering technique based on a Smoothed-Particle Hydrodynamics physical model [6]. We first recall the relevant equations involved in the SPH simulation of fluids, for the computation of the physical quantities (density) and the related forces (pressure and viscosity forces). Then, we describe our approach with the computation of SPH haptic forces and a haptic rendering algorithm integrated to the physical simulation, implemented entirely on GPU.

3.1 SPH Simulation of Fluids

An SPH fluid simulation is based on particles carrying different physical properties, such as mass and viscosity, discretizing the fluid volume. These particles have a smoothing radius, a spatial distance defining a neighborhood around them. Physical quantities, such as density and interaction forces, can be computed for each particle through the weighted sum of the relevant properties or quantities of the particles inside its neighborhood. The weight of the contributions of a neighbor particle depends on its distance to the treated particle, and is defined in a function called smoothing kernel.

The smoothed quantity $A(\mathbf{x})$ at any position $\mathbf{x}$ in space is computed through the general formula:

$$A(\mathbf{x}) = \sum_i A_i V_i W(\mathbf{x} - \mathbf{x_i}, h) \quad (1)$$

where A_i is the discrete quantity $A(\mathbf{x_i})$ sampled for neighboring particle i at position $\mathbf{x_i}$, V_i is the volume of i, and W is the smoothing kernel of support h, where particles farther than h are not taken into account.

Hence, the density ρ of particle p can be computed through:

$$\rho_p(\mathbf{x_p}) = \sum_i m_i W(\mathbf{x_p} - \mathbf{x_i}, h) \quad (2)$$

where m_i is the mass of neighboring particle i. We recall $V_i = m_i/\rho_i$, with ρ_i being the density of particle i.

The motion of fluids is driven by the Navier-Stokes equations, with the following formulation (after dropping mass conservation and convective terms, inherent to a Lagrangian simulation) :

$$\rho \frac{D\mathbf{v}}{Dt} = -\nabla P + \rho \mathbf{g} + \mu \nabla^2 \mathbf{v} \quad (3)$$

where P is the pressure, $\mathbf{v}$ the velocity, μ the viscosity coefficient and $\mathbf{g}$ the gravity field. ∇ and ∇^2 are respectively the gradient and Laplacian of the physical quantities. This formulation leads to 3 distinct forces:

$$\rho \mathbf{a} = \mathbf{f}^{pressure} + \mathbf{f}^{gravity} + \mathbf{f}^{viscosity} \quad (4)$$

with $\mathbf{a}$ being the particle acceleration. Pressure forces are computed from pressure quantities, and viscosity forces are computed from velocities:

$$\mathbf{f}_p^{pressure}(\mathbf{x_p}) = -V_p \sum_i V_i \frac{P_p + P_i}{2} \nabla W(\mathbf{x_p} - \mathbf{x_i}, h) \quad (5)$$

$$\mathbf{f}_p^{viscosity}(\mathbf{x_p}) = \mu V_p \sum_i V_i(\mathbf{v}_i - \mathbf{v}_p)\nabla^2 W(\mathbf{x_p} - \mathbf{x_i}, h) \quad (6)$$

The symmetrization of the force computations and the kernel functions follow those proposed in [6]. The pressure computation follows the modified ideal gas state equation proposed in [35]:

$$P = k(\rho - \rho^0) \quad (7)$$

where k is a gas constant that depends on the temperature, and ρ^0 the rest density.

Therefore, the fluid simulation is done in two consecutive steps. First, the density is computed for each particle, and the pressure is derived from it. Second, after computing pressure and viscosity forces, acceleration, velocity and position are deduced through a Leap-Frog integration scheme [36].

Parameters such as the viscosity can be changed to obtain different fluid behaviors, from smoke (no viscosity) to honey (very high viscosity).

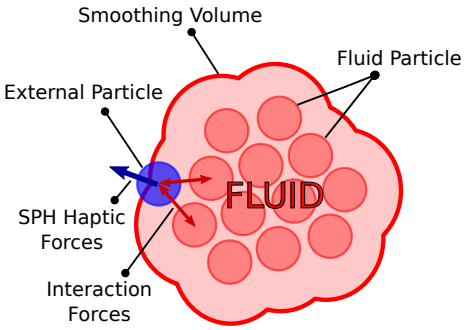


Fig. 2. Smoothing Volume and SPH haptic forces. The Smoothing Volume of a volume of fluid is defined by the sum of the smoothing volumes of each particle belonging to the fluid. When an external particle enters the Smoothing Volume, SPH haptic forces are seamlessly computed between the external particle and the fluid particles whose smoothing volume contains the external particle.

3.2 SPH Haptic Rendering

We introduce a novel haptic rendering algorithm based on Smoothed-Particle Hydrodynamics. Our method allows to compute force-feedback when interacting with fluid through a single fluid particle. Our technique can be classified in the group of distance-field-based haptic rendering techniques, and can therefore be used in simulations where exact surface representation is not required [37]. This is the case for fluids, and particularly SPH fluids, where the exact boundary of the fluid is not explicitly defined: boundary particles are not the boundaries of the fluid, but rather the last points in space that provide information about the fluid boundaries.

3.2.1 Smoothing Volume and SPH Haptic Forces

In the SPH model, forces are computed when a particle enters the smoothing volume of another particle, which is defined as a sphere with the smoothing radius as radius and the particle position as center. The sum of the smoothing volumes of the particles of an entity defines the Smoothing Volume of the entity, behaving like a force field volume around the entity. The Smoothing Volume is illustrated in Figure 2. Any external particle (not belonging to the entity) inside the entity Smoothing Volume will trigger the computation of forces exerted on both the entity and the external particle.

The forces generated when an external particle is inside the Smoothing Volume of an entity are SPH haptic forces, noted $\mathbf{f}^{haptic}$. They repel the external particle from the entity, and the entity from the external particle in a 2-way coupling scenario. Following the computation of quantities in the SPH model, $\mathbf{f}^{haptic}$ has the general form of Equation 1.

The profile of the haptic rendering is closely related to the choice of $\mathbf{f}^{haptic}$. When interacting with fluid through a particle, or in other words when the haptic device is coupled to a particle, an intuitive choice is to treat the coupled particle as a fluid particle. The coupled particle is submitted to the same physical forces as any other fluid particle. Since these forces are then rendered through the haptic device, they provide a direct way to interact with the fluid volume. Hence, our SPH haptic forces, when coupled to a particle, are the sum of pressure and viscosity forces:

$$\mathbf{f}_p^{haptic}(\mathbf{x_p}) = \mathbf{f}_p^{pressure}(\mathbf{x_p}) + \mathbf{f}_p^{viscosity}(\mathbf{x_p}) \quad (8)$$

3.2.2 Haptic Rendering Algorithm

Algorithm 1 shows the different steps of the computation of haptic feedback when interacting with fluid through a single particle coupled to a haptic device. It can be broken down to 4 distinct steps:

- **density computation (lines 1-3)**. A density quantity is computed for each particle, according to the neighboring particles distance and mass.
- **SPH haptic force computation (lines 4-12)**. According to the type of particle (a flag specifying fluid or external), the algorithm computes either fluid interaction forces from section 3.1, or SPH haptic forces from section 3.2.1.
- **fluid particle integration (lines 13-16)**. After adding external forces (such as gravity), a new position and velocity is computed for each fluid particle by integrating forces over the simulation time step.
- **coupled particle integration and haptic coupling (lines 17-22)**. After adding external forces (such as gravity), particles coupled to a haptic device follow a different integration process, taking into account forces from the haptic interface, and sending back their new position and velocity.

3.3 GPU Implementation

The entire physical model is implemented on GPU using the CUDA framework [38]. It is based on Green's imple-

Algorithm 1 Haptic rendering algorithm for a single particle coupling: the different steps of the computation of haptic feedback from the interaction between fluid and external particles are illustrated. Each external particle is coupled to a haptic device.

```
for all particles in fluid do
    compute density (Eq. 2)
end for
for all particles in fluid do
    for all neighboring particles do
        if neighbor is coupled to a haptic device then
            compute SPH haptic forces (Eq. 8)
        else
            compute fluid forces (Eqs. 5 and 6)
        end if
    end for
end for
for all particles in fluid do
    add external forces (gravity)
    integrate (new position and velocity)
end for
for all particles coupled to a haptic device do
    add external forces (gravity)
    add force from haptic interface
    integrate (new position and velocity)
    send new position and velocity to haptic interface
end for
```

mentation of a CUDA particle simulator [39], modified to compute the SPH Navier-Stokes equations.

3.3.1 Optimizations

The increase in performance compared to a CPU implementation comes from the parallelization of the density, force and integration computations. Knowing that an SPH particle only interacts with other particles inside its smoothing radius, the 3D interactive space is subdivided in a regular grid, restricting the search for neighbors to the 26 grid cells around the particle cell and the particle cell itself. Major optimizations regarding memory access and data structures [38] are:

- the sorting of the particles by cell in GPU memory in order to ensure a coalesced read, hence optimizing memory access;
- the binding of the sorted data arrays to textures, and the use of texture lookups to benefit from texture data caching;
- the use of an OpenGL Vertex Array Object for the particle position array, allowing the graphic rendering algorithm to access position data without copying it back to CPU memory.

3.3.2 Update rate

In theory, the graphic rendering frame rate and the update rate of the haptic device could and should be independent. Graphics should be rendered at 24 fps for a comfortable visualization, while the haptic update rate, depending on the computation time of the fluid simulation loop, requires a higher frequency. However, using an all-GPU design to achieve plausible haptic interaction rates requires the use of the GPU for simulation computations *and* for graphic rendering. In current implementations of graphic drivers, graphic rendering is a GPU blocking task, hence the simulation and the graphic rendering on the GPU are not parallelizable. Hence, using a different frequency for graphic update would cause periodic drops in the haptic update rate, introducing artifacts in the haptic rendering. Since every simulation loop has to be rendered, we needed to optimize our graphic rendering algorithm to achieve reasonable haptic update rates, as detailed in section 5. A possible way to dissociate the graphic from the simulation loop would be the use of two GPUs, with one doing simulation computations and the other focusing on graphic rendering. The position data required for the graphic rendering could be copied from the simulation GPU to the other through asynchronous data transfer [38], avoiding a copy lock on the simulation GPU. Moreover, the advent of multi-core GPUs would make dissociation even simpler by removing the need of data transfer between GPUs [40] [41].

4 6DoF Haptic Rendering of Fluids through Rigid Body Interaction

In the previous section, we showed how to generate force feedback when interacting with fluid through a single particle. In this section, we extend our haptic rendering technique to allow the interaction between a rigid body and the fluid, hence producing a 6DoF haptic feedback when coupled to a 6DoF haptic device. We describe our rigid body model, the computation of rigid body dynamics using SPH particles, and the underlying haptic coupling scheme.

4.1 Unified Particle Model

In order to achieve 6DoF haptic interaction with fluids, we required a rigid body model that could fit in the SPH simulation with a minimum impact regarding computation time, and that would allow a flexible coupling mechanism. We propose a real-time approach based on [7] and improved for haptic interaction: a unified particle model allowing the seamless real-time interaction between fluid and arbitrary-shaped rigid bodies.

Rigid bodies can be simply and efficiently modeled with the same SPH particles used in the fluid simulation. This allows to use the SPH model for the computation of forces between fluid and rigid body particles, removing the need of additional collision detection algorithms. Moreover, since interactions are computed between particles, the overall shape of the rigid body is not important. Hence, the unified particle model allows the seamless use of arbitrary-shaped rigid bodies, including concave objects.

Rigid body polygonal meshes are sampled into a set of particles covering the surface of the mesh. Figure 3 shows a bowl mesh and its corresponding particle sampling. Several mesh sampling techniques exist in the literature. We highlight the work presented in [42], for an offline simulation, where the particle sampling is achieved in a preprocessing step using a distance field, and the random placement of particles inside the distance field. Particles are then subjected to an attraction constraint to the zero-isosurface and a repulsion constraint against other particles, achieving an efficient sampling of the surface when reaching a convergence criteria. The survey and the improvement of other techniques are, however, beyond the scope of this paper.

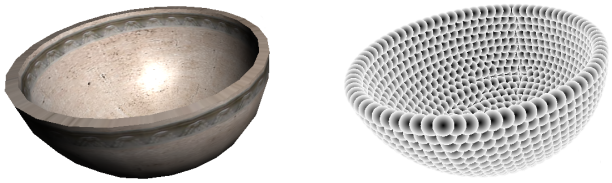


Fig. 3. Rigid body particle sampling. The mesh bowl (left) is converted into a set of particles (right).

The SPH fluid simulation and the haptic rendering algorithm are adapted to allow the simulation of rigid bodies and their interaction with the fluid. These changes are:

- rigid body particles interact with fluid particles through the SPH haptic forces $\mathbf{f}^{haptic}$;
- rigid body particles are given constant densities, since for rigid bodies these quantities are constant throughout the simulation;
- when computing the density of fluid particles, rigid body particles are omitted so that fluid densities are only computed with fluid particles;
- rigid bodies follow rigid body dynamics, as described in section 4.2.

The choice of an efficient $\mathbf{f}^{haptic}$ for a rigid body coupling case is not necessarily the same as in single particle coupling. In the previous section, the coupled particle could be treated as a fluid particle, and hence $\mathbf{f}^{haptic}$ was set to follow fluid forces. When dealing with rigid bodies, previous work has used different forces and mechanisms to simulate a solid-fluid interaction, as described in section 2.2. Among these methods are simple penalty forces [12], Lennard-Jones forces [14] and unified fluid-rigid body forces [16] [7].

In a unified, parallel and time-critical framework, unified fluid-rigid body forces fit well for computation time reasons. Using the same interaction forces as in a fluid-fluid case improves the gain of parallel computation, while providing a reasonable amount of control over the forces through density and viscosity values. These values are set by the user per rigid body, hence allowing different behaviors for each rigid body. However, since Navier-Stokes equations are not physically meant for rigid bodies, density and viscosity values cannot be looked up in the literature, but need to be chosen empirically. This is particularly true for viscosity forces, where higher solid-fluid viscosity values are required to achieve the desired viscous and sticky effects.

We used the pressure forces as for fluid particle interaction, and symmetrized the viscosity forces as in [43] to account for the possibility of having different viscosity values:

$$\mathbf{f}_p^{visc}(\mathbf{x_p}) = V_p \sum_i V_i \frac{\mu_i + \mu_j}{2}(\mathbf{v}_i - \mathbf{v}_p)\nabla^2 W(\mathbf{x_p} - \mathbf{x_i}, h) \quad (9)$$

4.2 Rigid Body Dynamics

With rigid bodies, position and velocity values are no longer computed independently for each particle. SPH haptic forces due to the interaction with the fluid are summed and applied at the center of mass cm of the rigid body, as shown in Figure 4.

$$\mathbf{F}_{cm}^{haptic} = \sum_{i_{body}} \mathbf{f}_{i_{body}}^{haptic} \quad (10)$$

Moreover, torques are also applied to the center of mass due to the same SPH haptic forces exerted on the particles of the rigid body:

$$\tau_{cm}^{haptic} = \sum_{i_{body}} \tau_{i_{body}}^{haptic} \quad (11)$$

with:

$$\tau_{i_{body}}^{haptic} = (\mathbf{x}_{i_{body}} - \mathbf{x}_{cm}) \times \mathbf{f}_{i_{body}}^{haptic} \quad (12)$$

Analogously to the linear movement, the angular acceleration α_{cm} of the center of mass can be computed through the use of rotational dynamics:

$$\alpha_{cm} = I^{-1}\tau_{cm}^{haptic} \quad (13)$$

where I is the moment of inertia of the rigid body.

The angular velocity ω_{cm} and position of the center of mass are then computed using a Leap-Frog integration scheme [36]. The angular velocity is reported to each particle of the body according to their position with respect to the center of mass:

$$\mathbf{v}_{i_{body}} = \mathbf{v}_{cm} + \omega_{cm} \times (\mathbf{x}_{i_{body}} - \mathbf{x}_{cm}) \quad (14)$$

4.3 6DoF Haptic Coupling Scheme

When coupling a rigid body to a 6DoF haptic device, we have to take into account the new rigid body dynamics. Algorithm 2 shows the new rigid body dynamics step, replacing the external particle integration step from Algorithm 1 (lines 17-22), and performing the 6DoF haptic coupling.

Figure 1 summarizes the different steps of the 6DoF haptic rendering technique, from the unified SPH simulation to the 6DoF haptic loop. Following our unified

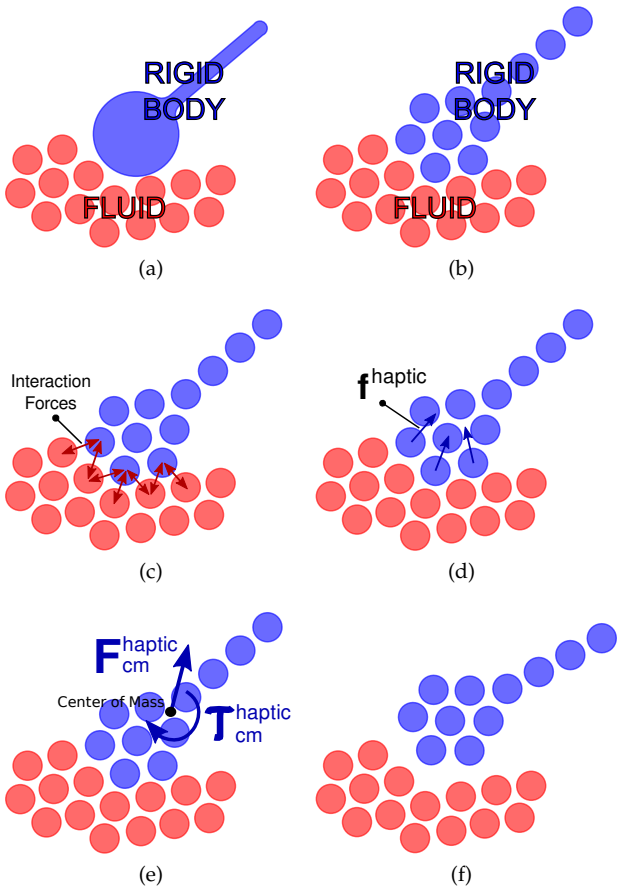


Fig. 4. Illustration of the computation of forces acting on a rigid body. A spoon and volume of fluid (a) are modeled with SPH particles (b). Fluid particles act on the spoon particles (c), generating an SPH haptic force $\mathbf{f}^{haptic}$ per spoon particle (d). These forces are summed resulting in a total force and a total torque applied at the center of mass of the spoon (e). Spoon particles are updated according to the new position and velocity of the spoon(f).

particle model, SPH haptic forces are computed along other forces in the SPH simulation. In the rigid body dynamics step, all forces $\mathbf{f}^{haptic}$ exerted on the coupled rigid body are summed to obtain a total force and a total torque (lines 1-5 of Algorithm 2). The force and torque feedback coming from the haptic device are added to the coupled rigid body (line 7 of Algorithm 2). Then, the new position and velocity are computed by integrating forces over the simulation time step. They are sent to the haptic device, closing the haptic loop in admittance mode [37].

Rigid bodies are also simulated entirely on the GPU. Rigid body particles are stored in the same arrays as fluid particles, since forces are computed for every particle in the simulation. However, they are grouped at the beginning of the arrays in order to ensure a coalesced memory access during rigid body dynamics computations. The sum of SPH haptic forces into a total force following Equation 12 is computed on the GPU by a 2-step tree-based parallel reduction [44], with sequential addressing. In a first CUDA kernel, the force array is partially summed within blocks of maximum thread size, with results stored in global memory. The second CUDA kernel sums the intermediate results into a final force value. Lines 6 to 8 of Algorithm 2 are executed on a single CUDA thread within the previous kernel, since there was no substantial gain in copying the data and executing the instructions on the CPU.

Algorithm 2 6DoF haptic rendering algorithm: rigid body dynamics step and haptic coupling with a 6DoF haptic device.

```
1:  for all rigid bodies do
2:    for all particles in the rigid body do
3:      add force exerted on the particle to total force
4:      add torque from the particle to total torque
5:    end for
6:    add external forces and torques (gravity)
7:    add force and torque from haptic device
8:    integrate (new rigid body position and velocity)
9:    send new position and velocity to haptic device
10:   for all particles in the rigid body do
11:     update particle position and velocity
12:   end for
13: end for
```

This coupling scheme allows the seamless haptic coupling with any rigid body, provided dimensions and masses are compatible with the device span and its maximum efforts. It also allows the computation of N different haptic couplings on the same scene, as well as between N devices and the same rigid body (as holding a bucket with two hands).

4.4 Virtual Coupling

A Virtual Coupling mechanism [34] is introduced between the haptic device and the coupled rigid body, creating a viscoelastic link between them. This mechanism allows the separation of the impedance of the haptic device from the impedance of the virtual environment. It reconciles a high update rate haptic device with a lower rate simulation, leading to an increase of stability [37].

5 Visual Fluid Rendering

As explained in section 3.3.2, the haptic loop needs high frequency updates, which required us to design a visual rendering method with performance over quality in mind. Our visual rendering is close to the one proposed in [21], with reduced computation time. We propose a three step approach: (1) compute per pixel fluid data (front and back fluid volume depth from view), (2) smooth the front surface depth in screen space using a fast bilateral filter and (3) compose the final frame.

5.1 Computing Per-Pixel Fluid Data

In order to render fluids, our method needs the front and back depth of the fluid volume for each pixel of

the view point. To this aim, we render each particle as a sphere. For the sake of performance, we do not render spheres using polygons but as point splats [21]. The back depth of the fluid volume is obtained using a reversed depth test. In order to obtain a connected fluid surface, point splats radius is 1.2 times larger than simulated particles radius. To accelerate this step, this buffer can be computed at a lower resolution than the screen. In our case, we used half the screen resolution.

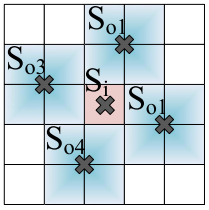


Fig. 5. The bilateral sampling kernel used to smooth the front surface of the fluid.

Using the raw depth values for the front fluid surface makes it appear blobby. To smooth out depth values, we could have used a bilateral Gaussian filter but this is a non separable blur filter. The algorithm proposed in [21] is very efficient but requires too many iterations to be effective. Thus, we propose to use a fast bilateral blur filter taking advantage of the hardware bilinear filtering (Figure 5). First, the depth is sampled for the inner pixel S_i. Then, four outer depths S_o are sampled. Thanks to hardware bilinear filtering, each of these outer samples represents the average of four depth values. The bilateral filtering is achieved by weighting outer depth samples according to the depth difference to the depth sampled at S_i. An outer sample is used only if the four averaged depths are from the fluid surface. We give a weight of 2.0 for S_i and 4.0 for S_o samples. This method allows us to average 17 depth values using only 5 texture samples. This bilateral filter is of lower quality than the filter proposed in [21] and can result in plateaus of equal depth. However, we can obtain a smooth fluid surface in few iterations. After preliminary testings we set the number of iterations to four.

5.2 Fluid Compositing

The fluid volume is finally composed with the virtual scene during a single full-screen pass using the virtual scene color S_c and linear depth S_d buffers, as well as the fluid data buffer. As in [21], the distortion of the scene color Sc perceived through the fluid is approximated as a texture look-up in screen-space with a displacement vector corresponding to the normal of the front fluid interface scaled by the fluid depth and refraction intensity. The final pixel color $col(p)$ is given by:

$$col(p) = \texttt{lerp}(\mathbf{F_c}, \mathbf{S_c}(\mathbf{p} + \beta \times T(p) \times \mathbf{n}.xy), \exp^{-T(p)*\mathbf{F_e}}) + \mathbf{k_s}(\mathbf{n} \bullet \mathbf{h})^{\alpha} \quad (15)$$

Fluid thickness $T(p)$ is first computed for each pixel p using $T(p) = \min(F_bD(p), S_d(p)) - \min(F_fD(p), S_d(p))$ where F_fD and F_bD are respectively the front and back depth of the fluid volume. In the case $T(p) = 0$, we simply copy the scene color. Otherwise, if $T(p) > 0$, we compute the final pixel color using Equation 15 where $\mathbf{F_c}$ is the fluid color, $\mathbf{F_e}$ is the wavelength-dependent color extinction coefficients of the fluid, β represents the amount of background color distortion simulating a refractive index, $\mathbf{h}$ the Blinn half-angle vector, $\mathbf{k_s}$ the specular color and α the specular exponent. The normal n is computed using finite difference on the front fluid surface depth [21].

The proposed optimizations of the technique described in [21] makes the approach suitable for real-time applications. These optimizations can thus be considered as a trade-off between performance and quality.

6 Evaluation

In this section, we evaluate our approach in terms of computation time and haptic rendering (forces and torques). The quality of our real-time visual rendering is also discussed.

6.1 Hardware Setup

The evaluation scenarios were carried out using two Virtuose 6DoF force-feedback devices from Haption (Soulge-sur-Ouette, France). Figure 12 shows both devices in use. The simulations were run on a laptop computer with a Core 2 Extreme X7900 processor at 2.8GHz, 4GB of RAM memory, and a Nvidia Quadro FX 3600M GPU with 512MB of graphic memory.

6.2 Computation Time

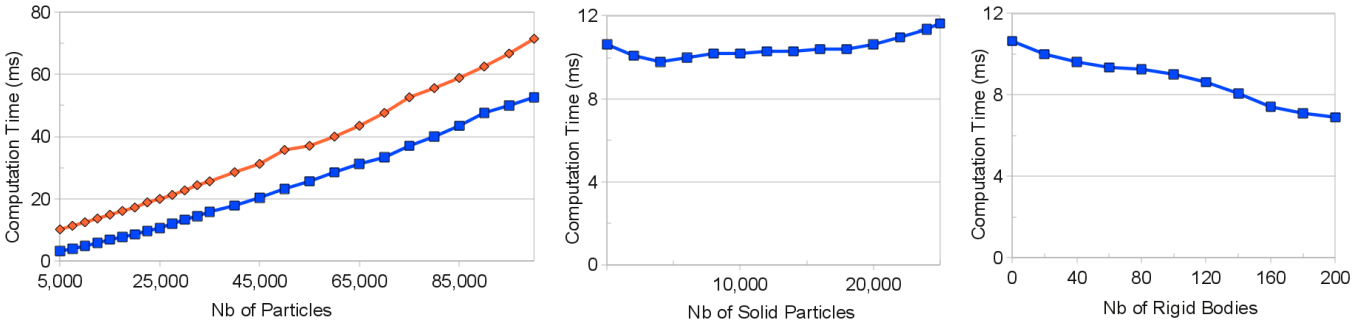
6.2.1 Fluid Simulation Performance

Ideally, for rigid body simulations, update rates should be close to 1000Hz [37]. However, forces due to fluids do not change rapidly. Hence, in order to achieve a smooth and stable haptic rendering, the update rate can be considerably lowered, and rates close to 70Hz have been reported to be satisfactory [3] [2].

Figure 6a shows the performance of our fluid simulation with respect to the number of fluid particles. We measured the computation time of the simulation of a 1x1m fluid pool with and without the graphic rendering.

Under 50ms (20 Hz), we consider that the simulation is not adequate for real time applications. With around 2,300 particles, the simulation and the graphic rendering take the same amount of time, while with 100,000 particles the simulation is around 4 times slower. Hence, the bottleneck of our simulation, when increasing the number of particles, is the physical simulation, while the graphic rendering remains efficient with a much lower drop in frame rate.

At 70 Hz, our implementation can thus simulate 32,000 fluid particles with the aforementioned hardware configuration, thus allowing the interaction with a considerable volume of fluid.



(a) Performance of the fluid simulation (in blue, squares) and the fluid simulation including graphic rendering (in red, diamonds) with respect to the number of fluid particles

(b) Performance of the simulation of a pool of fluid and a rigid body sphere with respect to the size of the rigid body. The number of total particles is constant.

(c) Performance of the simulation of rigid bodies with constant total number of particles with respect to the number of rigid bodies.

Fig. 6. Performance evaluation of our simulation algorithms

6.2.2 *Unified Particle Model Performance*

Figure 6b shows the computation time required to simulate a pool of fluid and a rigid body sphere with respect to the size of the rigid body. As the size of the rigid body increases, the size of the pool decreases to maintain a constant number of particles. The total number of particles is set to 25,000 since it allows a good frame rate for haptic rendering while being enough to simulate detailed objects and large amounts of fluid.

Up to 5,000 rigid body particles, the computation time decreases, due to the skipping of the density computation for rigid body particles. Beyond 5,000 particles, the computation time starts increasing again, since updating the rigid body particle positions from the newly computed values for the center of mass becomes more time consuming. In the GPU, each rigid body is treated in a single CUDA block in order to share the common rigid body values. When dealing with several rigid bodies of smaller size, which is more common than a scene with a single rigid body, the simulation becomes much more efficient, as shown in Figure 6c. From the single 25,000 particles rigid body sphere, we made several non-contacting rigid bodies spheres of equal size. When increasing the number of rigid bodies, with a constant number of total particles, we can see how computation time decreases due to the use of more CUDA blocks with a lower number of threads, making the particle update more parallel and efficient. Overall, the overhead due to rigid body simulation remains small, since in the worst-case scenario of a single rigid body the simulation takes less than 10% more time to compute than a pool of fluid with the same number of particles.

6.3 Graphic Rendering

Figure 7 shows the result of our technique (left) and of the technique proposed in [21] (right) for a fluid volume made of 32768 particles, rendered at a resolution of 1024x768. The method described in [21] requires 60 passes (GPU computation time of 45ms for a final framerate of 20fps) to result in a smooth fluid surface. However, our method only requires 4 passes

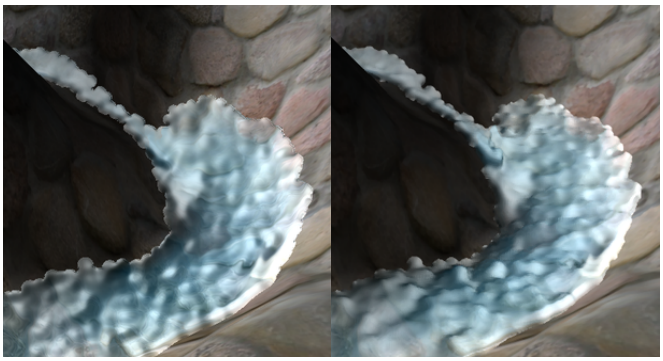


Fig. 7. Comparison of our graphic rendering method (left) with the original version [21] (right) for a volume of fluid.

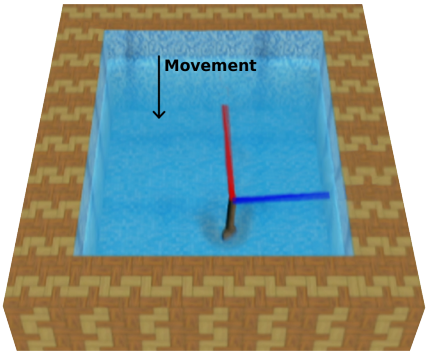
(GPU computation time of 8ms for a final framerate of 90fps). Our method being designed primarily for speed over accuracy, some artifacts are visible at the edges of the volume whereas, using the original approach, the edges are preserved. Moreover, our fluid rendering exhibits plateaus of equal depth producing a surface that seems more flat. However, our method is faster than the original method [21] and accurate enough for our use-cases.

6.4 Example Scenarios

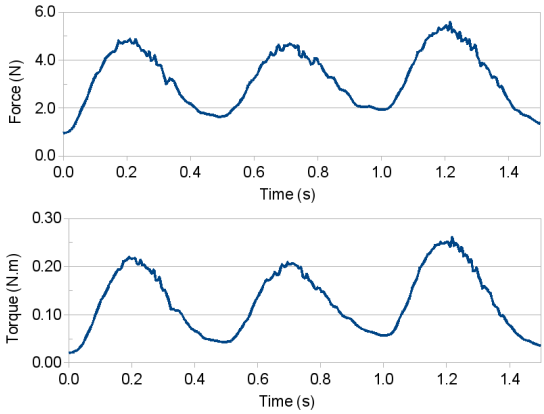
Throughout the following examples, we highlight some interesting interaction scenarios to illustrate the various possibilities of our approach. We recorded the efforts exerted on the haptic device, and provide plots with the magnitude of the efforts with respect to time, in order to have a visual feedback on the haptic rendering.

6.4.1 *6DoF Interaction*

This scenario was designed to highlight the 6DoF capabilities of our technique. The possibility to perceive torques is a main feature of our work, since these torques were very limited or not possible at all in prior existing work. The user exerts a forward and backward movement with a rigid body spoon on a pool of low



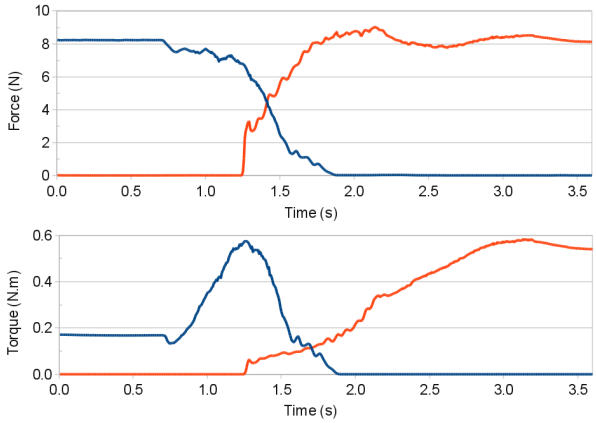
(a) Force (in red) and torque (in blue) exerted on the spoon. The force is opposite to the spoon movement. The torque shows a clock-wise effort in accordance to the spoon movement.



(b) Plot of the force (top) and the torque (bottom) generated by the forward, backward and forward movement of the spoon on the fluid.

Fig. 8. 6DoF haptic interaction: scenario illustrating 6DoF interaction possibilities, with a user holding a virtual spoon through a haptic device and using it to stir a pool of fluid. The user feels the force exerted on the spoon, as well as the corresponding torque, which plays a major role in the haptic perception of the stirring movement.

(a) Fluid is being poured from the bowl into the pan, making the fluid mass shift between both rigid body containers. The containers are hand held by the user through two 6DoF haptic devices.



(b) Plot of the force (top) and the torque (bottom) of the bowl (in blue/dark) and the pan (in red/light) during the pouring motion of some fluid. The fluid mass shift is clearly visible in the force plot, while the torque plot illustrates how the bowl is first tilted (torque peak) and then emptied (torque decrease).

Fig. 9. Container interaction: scenario illustrating the interaction with fluid through rigid body containers. A bowl and a pan, each coupled to a 6DoF haptic device, are used as containers. The fluid held by the bowl is poured into the pan.

viscosity fluid. Figure 8a shows the forces and torques involved in the stirring of the fluid with the spoon. We can notice the force (in red) going opposite to the back-to-front movement of the spoon. The torque (in blue) is generated through a rotation around an axis orthogonal to the linear movement. These visual cues match the plotting of the force and the torque over the duration of the movement (3 forward and backward movements) in Figure 8b.

6.4.2 *Container Interaction*

In this scenario, we illustrate the use of rigid bodies as containers to interact with the fluid. We used two 6DoF Virtuose haptic devices. The first device (left hand) is coupled to a bowl, while the second device (right hand) is coupled to a pan. The fluid is poured from the bowl into the pan, as shown in Figure 9a. Figure 9b shows the torques involved in the pouring movement of the bowl, and how weight shifts from the bowl to the pan.

6.4.3 *Variable Viscosity*

An important feature of our approach is the possibility to interact with fluids of different viscosity, and hence "feel" these different viscosities through the haptic devices. We designed a scenario to capture the forces exerted on a rigid body container being spun around its vertical axis. The container is full of fluid, and initially at rest. The glass is spun with constant speed, kept rotating, and then stopped suddenly. The fluid has a viscosity of 2 Pa.s, while four different viscosities are used for

the rigid body container: 2, 10, 20, 40 Pa.s. The increase in viscosity leads to higher forces exerted on the haptic device. Figure 10 shows the torques generated in each case, clearly showing each of the 4 phases (rest, start, spin, stop) of the movement and each of the 4 viscosities (from low to high: blue, red, yellow, green).

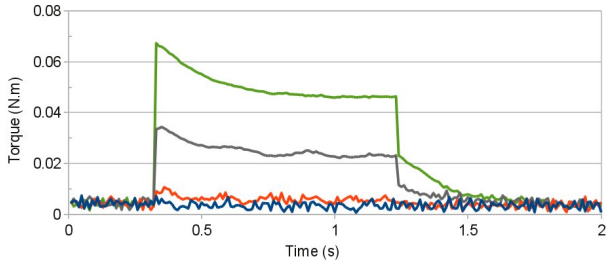


Fig. 10. Plot of the torques exerted on a rigid body container full of fluid (2 Pa.s viscosity) during a spinning movement. The container has different viscosities: 2 Pa.s (blue), 10 Pa.s (red), 20 Pa.s (yellow), 40 Pa.s (green).

6.4.4 Bimanual coupling on the same rigid body

Our haptic coupling and rendering mechanism allows the 6DoF haptic coupling between one or multiple haptic device and any rigid body, as well as between N devices and the same rigid body. This last feature of our approach is illustrated through a bucket full of low viscosity fluid held with two hands. The virtual bucket is coupled to the two 6DoF Virtuose haptic devices, as holding the bucket by each of its handles. The bucket is tilted to the left, to the right, and then to the left again, each time spilling some fluid. Figure 11 shows the force and torque plot for each haptic device during the tilting movements of the bucket.

6.5 A complete use-case

To showcase the main features of our approach, we developed a complete use-case: a virtual cooking simulator. It is a training and entertaining 2-handed interactive application that simulates the preparation process of a crepe. The user holds two virtual objects, a bowl and a pan, through two 6DoF Virtuose haptic devices from Haption, as shown in Figure 12. The simulation guides the user through all the steps required to prepare a crepe (Figure 13): from the stirring and pouring of the batter to the spreading of different toppings on top of the crepe. By preparing virtual crepes, users can experience 6DoF haptic interaction with fluids of varying viscosity. It is one of the many promising applications that can be designed around 6DoF haptic interaction with fluids.

7 Discussion

Table 1 compares our technique in terms of features with previous techniques allowing real-time haptic interaction with fluids, highlighting the important aspects of an interaction scenario. The following discussion is structured around these different features.

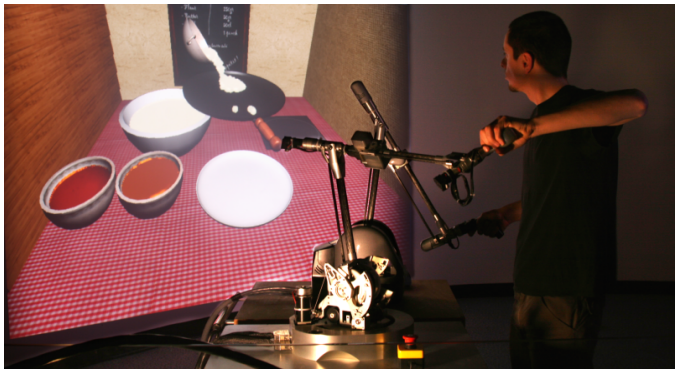


Fig. 12. The virtual crepe preparation simulator: the user pours some batter (a high viscosity fluid) from the bowl and into the pan. He feels the forces and torques from the pouring movement, as well as the weight shifting between his hands as the batter goes from the bowl to the pan.

A main feature of our work is the use of a Lagrangian SPH-based technique for haptic interaction with fluids, with many advantages in a real-time application: it is fast to compute (inherent mass conservation and no advection), it provides freedom regarding the scene (scalability, not bounded to a grid), and provides an intuitive way for haptic coupling. In the SPH model, the inherent smoothing of the SPH haptic forces, and the lack of discrete contact points, ensure that there are no discontinuities in the magnitudes of the resulting forces and torques, providing a stable and smooth haptic feedback.

Regarding the performance of our algorithms, we could run our scenarios at higher frame rates than the other techniques surveyed in section 2.5, even with more complex scenes. The only exception is the work presented in [1], which is not entirely computed in real-time, limiting the interaction possibilities. The higher performance is achieved through the use of the SPH model and our efficient GPU implementation, and performing better than [9] and [10] (12 and 1.7 times faster, respectively), and close to [11] (1.2 times slower), which focus exclusively on the optimization of a particle-based simulation of fluids. The different force and torque plots of section 6.4 provide a visual representation of the force feedback. The haptic forces and torques exerted on the device match the scenarios and the different interactions that are visually perceived, showing that the haptic rendering is faithful and realistic.

The disadvantages of our technique are common to distance-field-based haptic rendering techniques. The use of a smoothing volume makes the interaction approximate, hence focusing the use of the technique on applications where exact surface representation is not required. As mentioned earlier, this approach is very well suited for the haptic interaction with fluids due to their approximate boundary.

Other approaches for solid-fluid interaction, such as direct forcing with constraint equations, provide

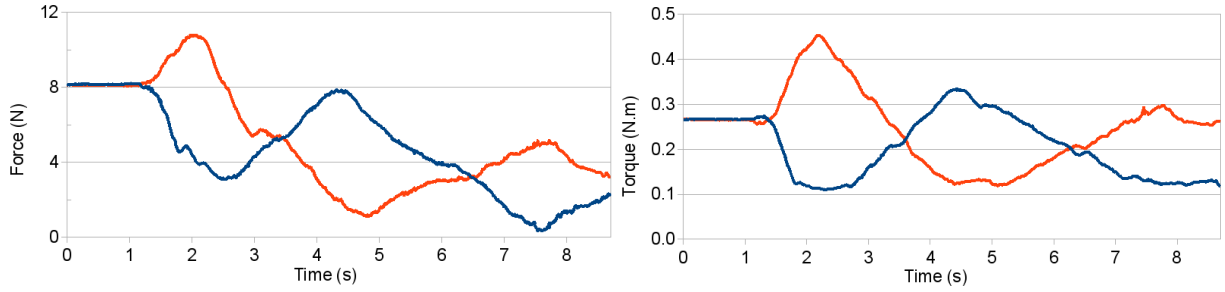


Fig. 11. Bimanual coupling on the same rigid body. Plot of the force (left) and the torque (right) generated by fluid inside a bucket coupled to two 6DoF haptic devices (left hand in red/light, right hand in blue/dark). The peaks and valleys correspond to the tilting movements, increasing the force and the torque on the side to which the bucket is tilted. Forces and torques progressively decrease as fluid is spilled.

Method	Physical Model	DOF	Real-time Computations	Performance[1]	GPU Acceleration	Arbit.-Shaped Rigid Bodies	Variable Viscosity
[33]	Eulerian	3	✓	30 Hz			
[1]	-	6		500 Hz			
[3]	Eulerian	6	✓	40 - 70 Hz			
[2]	Eulerian	3	✓	30 - 75 Hz	✓		
Our method	Lagrangian	6	✓	60 - 120 Hz	✓	✓	✓

TABLE 1
Comparison of our approach with previous work on real-time haptic interaction with fluids.

more control on boundary conditions and enforce non-penetration for colliding particles. However, there is a computational cost with up to 3 collision detection steps for 2-way coupling scenarios, and some restrictions on the number of interacting bodies. Regarding the haptic coupling, the use of a Virtual Coupling mechanism improves the stability of the haptic feedback, but inevitably affects the transparency of the haptic coupling due to the introduction of a viscoelastic link between the haptic device and the rigid body. Further investigation will assess the impact of using a Virtual Coupling mechanism on the perception of fluid haptic feedback.

Although we did not run into stability issues, using explicit integration schemes means that unconditional stability is not guaranteed. Moreover, if a particle moves more than a kernel-size in one time step it could pass through other particles without collision, no matter the solid-fluid interaction algorithm. However, this can only happen with very fast particles (above 10 $m.s^{-1}$ with our current simulation parameters), one order of magnitude faster than the highest speed of a fluid particle in our simulation scenarios. Further investigation is required to assess these issues, but they can only be solved with expensive implicit integration schemes, which would make computations overly expensive given the high number of particles and contact points.

Regarding its benefits, our haptic interaction technique provides real-time 6DoF force-feedback, which was only possible in [1] with pre-computed ad-hoc forces. Moreover, our unified particle model allows the seamless use of complex rigid bodies of arbitrary shape for the 6DoF haptic interaction with the fluid, including concave rigid body containers that can hold the fluid or parts of it. It allows an entire range of scenarios and applications, such as using complex utensils like a mixer to stir a soup, or odd shaped probes to explore the cavities of human body vessels. Other advantages, not illustrated in this work but inherent to the SPH model, allow the design of rich and compelling scenes: an entirely dynamic virtual world and topology-changing rigid bodies without impacting the computation time.

We believe many applications could be designed based on our approach. Such applications span from the medical field (organic fluids like blood) to industrial scenarios (painting, manipulating dangerous fluids) and gaming or entertainment simulations (water and mud in natural scenes, water sports).

8 Conclusion

In this work, we proposed a novel approach for 6DoF haptic interaction with viscous fluids based on a Smoothed-Particle Hydrodynamics physical model. It allows real-time 6DoF haptic interaction with fluids of variable viscosity through arbitrary shaped rigid bodies. Through a novel haptic rendering technique, we compute SPH haptic forces to produce a smooth haptic interaction. Thanks to a unified particle model, rigid bodies can interact with fluids and provide 6DoF haptic feedback. We designed different example scenarios to illustrate and evaluate some of the interaction possibilities

1. Regarding performance, fair comparison is difficult, since existing haptic fluid interaction techniques are based on different physical models and were computed on different hardware. We provide these performance values for information purposes only.

Fig. 13. New interaction possibilities: a bowl coupled to the 6DoF haptic device is used to stir the batter (left), and the same bowl pours maple syrup on the crepe (right).

offered by our technique. The evaluation of the technique showed the efficiency of the rendering algorithm in scenes rendered at 70Hz with up to 32,000 particles.

This novel approach proves that fluids are now readily available for complex 6DoF haptic interactions, and opens an entirely new horizon of applications and interaction techniques in Virtual Reality.

Acknowledgments

This work was supported by the European Community under FP7 FET-Open grant agreement n°222107 NIW - Natural Interactive Walking.

References

[1] Y. Dobashi, M. Sato, S. Hasegawa, T. Yamamoto, M. Kato, and T. Nishita, "A fluid resistance map method for real-time haptic interaction with fluids," *Proc. ACM Symp. Virtual Reality Software and Technology*, 2006.

[2] M. Yang, J. Lu, Z. Zhou, A. Safonova, and K. Kuchenbecker, "A GPU-Based approach for Real-Time haptic rendering of 3D fluids," *Proc. ACM SIGGRAPH Asia Sketches*, Dec. 2009.

[3] W. Baxter and M. C. Lin, "Haptic interaction with fluid media," *Proc. Graphics Interface*, pp. 81–88, 2004.

[4] R. Bridson and M. Muller-Fischer, "Fluid simulation" *ACM SIGGRAPH courses*, 2007.

[5] J. J. Monaghan, "Smoothed particle hydrodynamics," *Annual Review of Astronomy and Astrophysics*, vol. 30, no. 1, pp. 543–574, Sep. 1992.

[6] M. Muller, D. Charypar, and M. Gross, "Particle-based fluid simulation for interactive applications," *Proc. ACM SIGGRAPH/Eurographics Symp. Computer Animation*, 2003.

[7] B. Solenthaler, J. Schlfli, and R. Pajarola, "A unified particle model for fluid-solid interactions," *Comput. Animat. Virtual Worlds*, vol. 18, no. 1, pp. 69–82, 2007.

[8] T. Amada, M. Imura, Y. Yasumuro, Y. Manabe, and K. Chihara, "Particle-Based fluid simulation on GPU," *Proc. GP2 Workshop*, 2004.

[9] K. Hegeman, N. Carr, and G. Miller, "Particle-Based fluid simulation on the GPU," *LNCS*, pp. 228–235, 2006.

[10] T. Harada, S. Koshizuka, and Y. Kawaguchi, "Smoothed particle hydrodynamics on GPUs," in *Proc. Computer Graphics Int'l*, 2007.

[11] Y. Zhang, B. Solenthaler, and R. Pajarola, "Adaptive sampling and rendering of fluids on the GPU," *Proc. Symp. Point-Based Graphics*, Aug. 2008.

[12] J. J. Monaghan, "Smoothed particle hydrodynamics," *Reports on Progress in Physics*, vol. 68, no. 8, pp. 1703–1759, 2005.

[13] J. Bonet and S. Kulasegaram, "A simplified approach to enhance the performance of smooth particle hydrodynamics methods," *Applied Mathematics and Computation*, vol. 126, no. 2-3, pp. 133–155, Mar. 2002.

[14] M. Muller, S. Schirm, M. Teschner, B. Heidelberger, and M. Gross, "Interaction of fluids with deformable solids," *Comput. Animat. Virtual Worlds*, vol. 15, no. 3-4, pp. 159–171, 2004.

[15] M. Carlson, P. J. Mucha, and G. Turk, "Rigid fluid: animating the interplay between rigid bodies and fluid," *ACM SIGGRAPH Papers*, 2004.

[16] R. Keiser, B. Adams, P. Dutre, L. Guibas, and M. Pauly, "Multiresolution Particle-Based fluids," ETH Zurich, Tech. Rep., 2006.

[17] M. Becker, H. Tessendorf, and M. Teschner, "Direct forcing for lagrangian Rigid-Fluid coupling," *IEEE Trans. Visualization and Computer Graphics*, vol. 15, no. 3, pp. 493–503, 2009.

[18] I. D. Rosenberg and K. Birdwell, "Real-time particle isosurface extraction", *Proc. ACM Symp. on Interactive 3D Graphics and Games*, 2008.

[19] M. Muller, S. Schirm, and S. Duthaler, "Screen space meshes," *Proc. ACM SIGGRAPH/Eurographics Symp. Computer Animation*, 2007.

[20] H. Cords and O. Staadt, "Instant liquids," *Poster Proc. ACM Siggraph/Eurographics Symp. Computer Animation*, 2008.

[21] W. J. van der Laan, S. Green, and M. Sainz, "Screen space fluid rendering with curvature flow," *Proc. ACM Symp. Interactive 3D Graphics and Games*, 2009.

[22] A. Gregory, A. Mascarenhas, S. Ehmann, M. Lin, and D. Manocha, "Six degree-of-freedom haptic display of polygonal models," *Proc. IEEE Conf. Visualization*, 2000.

[23] D. D. Nelson and E. Cohen, "Optimization-Based virtual surface contact manipulation at force control rates," *Proc. IEEE Virtual Reality*, 2000.

[24] W. A. McNeely, K. D. Puterbaugh, and J. J. Troy, "Six degree-of-freedom haptic rendering using voxel sampling," *ACM SIGGRAPH Papers*, 1999.

[25] M. Wan and W. A. McNeely, "Quasi-Static approximation for 6 Degrees-of-Freedom haptic rendering," *Proc. IEEE Visualization*, 2003.

[26] W. A. McNeely, K. D. Puterbaugh, and J. J. Troy, "Advances in voxel-based 6-DOF haptic rendering," *ACM SIGGRAPH Courses*, 2005.

[27] M. A. Otaduy and M. C. Lin, "Sensation preserving simplification for haptic rendering," *ACM SIGGRAPH Papers*, 2003.

[28] D. E. Johnson, P. Willemsen, and E. Cohen, "Six Degree-of-Freedom haptic rendering using spatialized normal cone search," *IEEE Trans. Visualization and Computer Graphics*, vol. 11, no. 6, pp. 661–670, 2005.

[29] M. Ortega, S. Redon, and S. Coquillart, "A six Degree-of-Freedom God-Object method for haptic display of rigid bodies with surface properties," *IEEE Trans. Visualization and Computer Graphics*, vol. 13, no. 3, pp. 458–469, 2007.

[30] C. Zilles and J. Salisbury, "A constraint-based god-object method for haptic display," *Proc. IEEE/RSJ Intelligent Robots and Systems*, 1995.

[31] S. Cotin, H. Delingette, and N. Ayache, "Real-Time elastic deformations of soft tissues for surgery simulation," *IEEE Trans. Visualization and Computer Graphics*, vol. 5, no. 1, pp. 62–73, 1999.

[32] J. Barbic and D. James, "Time-critical distributed contact for 6-DoF haptic rendering of adaptively sampled reduced deformable models," *Proc. ACM SIGGRAPH/Eurographics Symp. Computer Animation*, 2007.

[33] J. Mora and W. Lee, "Real-Time 3D fluid interaction with a haptic user interface," *Proc. IEEE 3D User Interfaces*, 2008.

[34] J. Colgate, M. Stanley, and J. Brown, "Issues in the haptic display of tool use," *Proc. IEEE/RSJ Intelligent Robots and Systems*, 1995.

[35] M. Desbrun and M. Cani, "Smoothed particles: a new paradigm for animating highly deformable bodies," *Proc. Eurographics Workshop on Computer Animation and Simulation*, 1996.

[36] C. Pozrikidis, *Numerical Computation in Science and Engineering*, illustrated edition ed, Oxford University Press, USA, Apr. 1998.

[37] M. C. Lin and M. Otaduy, Eds., *Haptic Rendering: Foundations, Algorithms and Applications*, A. K. Peters, Jul. 2008.

[38] NVIDIA, "NVIDIA CUDA programming guide 2.0," 2008.

[39] S. Green, "CUDA particles," Jun. 2008.

[40] C. R. Johns and D. A. Brokenshire, "Introduction to the cell broadband engine architecture,", *IBM J. Res. Dev.*, vol. 51, no. 5, pp. 503–519, 2007.

[41] L. Seiler, D. Carmean, E. Sprangle, T. Forsyth, M. Abrash, P. Dubey, S. Junkins, A. Lake, J. Sugerman, R. Cavin, R. Espasa, E. Grochowski, T. Juan, and P. Hanrahan, "Larrabee: a many-core

x86 architecture for visual computing," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–15, 2008.
[42] N. Bell, Y. Yu, and P. J. Mucha, "Particle-based simulation of granular materials," *Proc. ACM SIGGRAPH/Eurographics Symp. Computer Animation*, 2005.
[43] M. Muller, B. Solenthaler, R. Keiser, and M. Gross, "Particle-based fluid-fluid interaction," *Proc. ACM SIGGRAPH/Eurographics Symp. on Computer Animation*, 2005.
[44] M. Harris, "Optimizing parallel reduction in CUDA," Jun. 2008.

Gabriel Cirio is a PhD student at the French National Research Institute for Computer Science and Control (INRIA) in Rennes, France. His main research interests include virtual reality, physically-based modeling and haptic rendering.

Maud Marchal is an Associate Professor in the Computer Science Department at INSA (Engineer school) in Rennes, France. She received her PhD in Computer Science from the University Joseph Fourier in Grenoble, France in 2006. Her main research interests include physical simulation, biomechanics, haptic interfaces, 3D interaction and virtual reality.

Sébastien Hillaire is a PhD student at the French National Research Institute for Computer Science and Control (INRIA) and France Telecom Orange Labs in Rennes, France. His main research interests include virtual reality, gaze-tracking, visual attention models and real-time high-quality rendering on GPUs.

Anatole Lécuyer is a Research Scientist at the French National Research Institute for Computer Science and Control (INRIA) in Rennes, France. He received his PhD in computer science from the University of Paris XI. His main research interests include virtual reality, 3D interaction, haptic interfaces, and brain-computer interfaces. He promotes a perception-based approach for the design of virtual environments and is currently an associate editor of the ACM Transactions on Applied Perception.

"Shoes-Your-Style": Changing Sound of Footsteps to Create New Walking Experiences

Anatole Lécuyer[1], Maud Marchal[1], Adrien Hamelin[1], David Wolinski[1],
Federico Fontana[2], Marco Civolani[2], Stefano Papetti[3], Stefania Serafin[4]

[1]INRIA/IRISA, Campus de Beaulieu, F35042 Rennes Cedex, France {firstname.lastname@inria.fr}
[2]University of Udine, Via delle Scienze, 206, I-33100 Udine, Italy {firstname.lastname@uniud.it}
[3]University of Verona, Strada le Grazie 15, I-37134 Verona Italy {firstname.lastname@univr.it}
[4]Aalborg University Copenhagen, Lautrupvang 15, 2750 Ballerup, Denmark {sts@media.aau.dk}

ABSTRACT
With "Shoes-Your-Style" we introduce the entertaining possibility to change auditory feedback and sound of footsteps when walking indoor and outdoor, to alter the nature (or "style") of shoes and contacts with the ground. This approach requires two components: a step detector and a sound generator. Artificial sounds of footsteps can be generated and tuned by the system, or by the user, to simulate any type of shoes (high-toe, sneakers, flip-flops, etc) or ground (marble, snow, mud, gravel, etc), generate artificial walking events (collisions, cracking objects, etc) or even unrealistic sounds (noises, words, music, etc). Such new kind of feedback should pave the way to creative interactive walking experiences. Potential applications range from entertainment and multimedia, to sport training and medical rehabilitation and reeducation.

Categories and Subject Descriptors
H.5.1 **[Multimedia Information Systems]**: *Audio input/output;* H.5.2 **[User Interfaces]**: *Auditory (non-speech) feedback.*

General Terms
Design, Human Factors.

Keywords
Shoes, auditory feedback, footstep, sound, walking.

1. INTRODUCTION
Footwear has considerably evolved in the long history of humans. Shoes have always been considered as a way to improve and support safety, protection and comfort of human locomotion. They have also been soon after considered as an esthetic product revealing the style but also other social messages of their owner.

More recently, shoes have evolved towards high technology, and began to incorporate electronic and computerized systems [2] as well as sensing devices. As an example, they can embed pedometers to gather input about human walk, such as for sport training.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Conference'10, Month 1–2, 2010, City, State, Country.
Copyright 2010 ACM 1-58113-000-0/00/0010…$10.00.

The current appearance of shoes is mostly *visual* including shape and colors, but also occasional lights and visual effects; and *auditory*, due to the sounds generated at each footstep. Sound of footsteps is highly dependent on the type of shoe-ground contact and of the materials involved, but it can also reveal important properties of human walker [1].

Indeed, when we walk, the main source of feedback and exchange with our surrounding is often the sound of contacts between our feet and the ground. It is a very strong and varied source of explicit or implicit information. Sound of footsteps informs about the type of shoes that we are wearing, the type of ground that we are walking over, and, potentially, about many other events that can occur during the walk (cracking objects, unexpected events, etc). It can also inform about the physical or mental state of the walker, such as his/her fatigue, stress or hurriedness.

In this paper we introduce the possibility to modify auditory feedback corresponding to sounds of footsteps when walking indoor and outdoor, for implicit or explicit feedback and control of human walk. In this approach, that we called "Shoes-Your-Style", the auditory feedback corresponding to the shoe-ground contact (or footstep sound) is manipulated at will by the user and/or the system. Sound of footsteps is therefore considered here at the basis of a new and dynamic feedback during indoor and outdoor walk for different kinds of mobile interactive experiences.

2. RELATED WORK: FROM AUGMENTED SHOES TO VIRTUAL GROUNDS
The idea of augmenting shoes has been around for quite some time already. Famous modifications, already available today in many footwear shops, involve simple luminescent laces that glow in the dark, sneakers with led-blinking soles and capable of sensing foot-floor contact, "heating" shoes, "massaging" shoes, "self-lacing" shoes, or even shoes that eliminate smells.

Sounds have already been incorporated, such as with GaitSpots pair of additional "squeakers" but providing only a basic "squeaking" auditory when in contact with the floor or a ball. GaitSpots system is commercialized for reinforcing gait training skills or for heel strike. Other "gadget" shoes embed audio or CD players and loudspeakers to play music at the level of the shoes.

However, all these auditory feedback are rather limited and uncorrelated from user's activity.

Recent works led to more sophisticated shoe-based applications. The "Nike+" sport kit consists in an accelerometer placed in (or on) Nike shoes, and a receiver attached to an iPod or iPhone. The sensor is able to detect each step that a runner takes, and from this information the runner's speed and the distance he has traveled can be calculated. The dedicated iPhone application stores statistics of the exercises. The "CabBoots" [2] are even more complex shoes embedding contact sensors and tactile feedback, and connected to a pedestrian navigation system. The shoes can then guide the user during his walk along predefined paths.

More recently, Adidas made shoes with Augmented Reality codes (visual patterns) which enable the consumer to play interactive games on-line with a webcam. The shoes can become controllers and input devices for various applications and videogames [3] [6]. Actually, feet-based interactive floor technologies [5] are already available for entertainment and videogames such as with Konami's "Dance-Dance-Revolution-Stage", or Nintendo's "Wii-Fit". However, all these interfaces have been not concerned with the sound modality as a mean to convey prominent floor cues when walking outdoor or indoor.

3. "SHOES-YOUR-STYLE" CONCEPT

3.1 Initial idea and extensions

The initial idea of "Shoes-your-style" is to give a walker the novel experience that he is potentially walking with shoes other than his/her. To do so, whenever a step is detected, a sound is played, related to the kind of shoes to be simulated.

Beyond this idea, the purpose of our approach is to introduce new possibilities of mobile interaction, relying on new entertaining walking sensations, and be able to share it with the people around. Even if it is possible to use this concept in videogames and virtual environments, the primary goal is here to do it in the real world during indoor and outdoor walk. With Shoes-Your-Style, it becomes possible to give sensations of walking over different grounds, and to suggest walking in other places, and/or to step on very unusual reliefs or objects.

3.2 Technical description

Technically speaking, the Shoes-Your-Style approach requires two main components: a step detector (input), and a sound generator (output). It is meant to be a wearable system for large-scale use. This global scheme is illustrated on Figure 1. Sounds are generated and played by a simulation/computer in response to the detection of steps. A controller enables to select and interactively set up the parameters of the simulated sound at will.

The most natural implementation of this concept probably consists in embedding all the input/output components directly inside the shoes of the user. This option will be described in details in the next section. However, many options exist considering that the input/output components could also be located on other parts of the body or in other equipments such as in Smartphones. Indeed, novel generations of Smartphones include high-quality audio feedback and accelerometers which could serve respectively to play the walking sounds and to analyze body motion and detect steps. The input and output components could also be split and located at different places. For instance, walking sounds could be played by a Smartphone while a pedometer could detect steps, such as in the existing Nike+ application for iPhone, for which sneakers with force sensors can communicate with a Smartphone and a pedometer simulation.

In the next section we detail one implementation of this concept which involves a all-in-one peripheral: the "Sonic Shoes". Future work could consider testing other solutions based on accelerometers for tracking steps for instance.

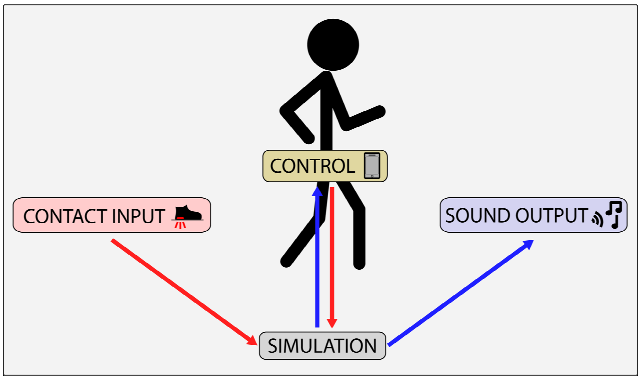


Figure 1: "Shoes-Your-Style" schematic diagram.

4. THE "SONIC-SHOES"

4.1 Input/output shoes

Our "Sonic-Shoes" are displayed on Figure 3. On each shoe there is a pressure sensor embedded both at the back (heel) (Fig 2:1) and at the front (toe) (Fig 2:2) with an additional sole on top (Fig 2:3). There is also one speaker per shoe which is attached to the middle strap (Fig 2:4). The outgoing cables are sent to the parallel port (Fig 2:5) of a controller box.

In the controller box, there is an "Arduino Duemilanove" micro-controller which is wired to the parallel port cables coming from the shoes and to a jack port for incoming sound. This box is then connected to the computer through USB and a double headed jack cable (the computer is here the main source of the sound).

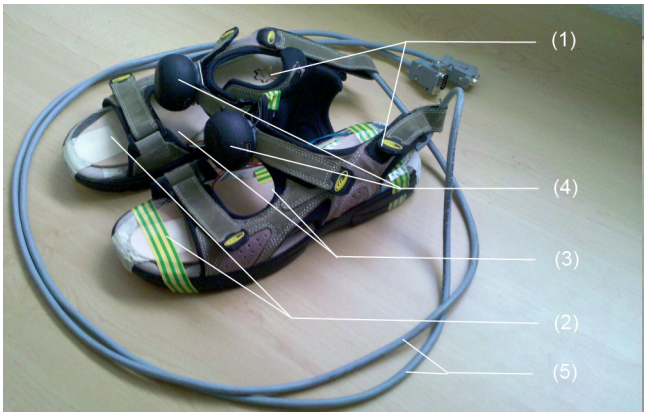


Figure 2: The Sonic Shoes.

4.2 Software components

Our software application is then capable of displaying all the data received from the sensors (Fig 3:1), playing different sets of sounds (Fig 3:2) creating/modifying/using XML configuration files (Fig 3:3) and communicating with a smart phone which then becomes a remote controller (Fig 3:4). Our software architecture is composed of different independent modules for handling either the different ways of playing a set of sounds, a GUI, a WiFi server for Smartphone communication, a USB communication module and an XML parser. The application is based on Qt (http://qt.nokia.com/) and FMOD (http://www.fmod.org/).

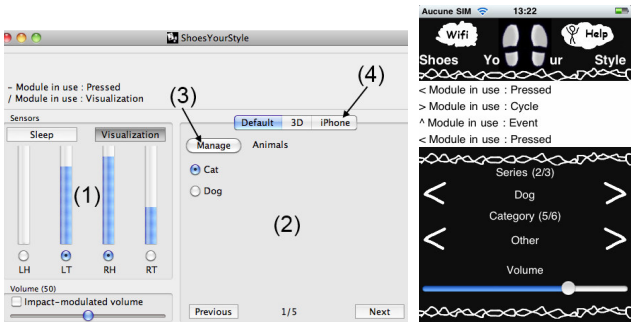


Figure 3: Software components
(Left) Main application running on computer,
(Right) Smartphone application and user interface.

Figure 3-Right also depicts the iPhone application which lets the user control the computer program through WiFi, i.e., choose the sounds played and the volume. It is also capable of displaying sensor information: shoes icons on top of the screen change colors when the corresponding sensor is active. The white area beneath the shoes (Figure 3-Right) is a console which echoes the main program running on the computer.

4.3 System in use

The whole system as worn by the user is displayed in Figure 4. It consists of the Sonic-Shoes (Fig 4:1) and the controller box and computer both placed in a backpack (Fig 4:2). Additionally, there is the iPhone (Fig 4:3) connected to the computer through WiFi used as a remote control.

For implementation details, the computer used was a MacBook-Pro running MacOS X 10.6.4 with an i7 2,55GHz processor, 4GB RAM and integrated sound. The phone used was an iPhone 3G. With this system, we could obtain a response time of the Sonic-Shoes (time between a step detection and the auditory feedback generation) smaller than 20 ms.

5. WALKING SOUNDS

Sounds are emitted by the Sonic-Shoes when a step is detected. In our current implementation, steps are detected by using only the force sensors located at the back of the shoes. A sound is therefore played each time a contact occurs between the heel and the ground.

All the sounds used were derived from audio files containing multiple sounds of footsteps from open-source sound libraries (www.freesounds.org). Each sound was manually worked out using Audacity software.

We could propose and test various kinds of sound, which are listed and briefly described hereafter.

5.1 Shoes sounds

The first category of sounds corresponds to different kinds of shoes potentially worn by the user. Among them we can stress: *high-toes* (in an echoing corridor), *clogs*, *spurs* (evoking "cow-boy boots"), and *flip-flops* (the sound is played here when the foot is taken off the ground as if the flip-flops is slapping on the heel).

5.2 Ground sounds

The second category corresponds to different types of relief the user is walking over. Among them we can first stress "aggregate" grounds: *gravel*, *leaves* (steps in a forest during autumn, when the leaves lay dry on the ground), *ice* (steps over a small amount of ice, thus a crackling sound), or *snow* (steps in a huge amount of fresh snow). Then we find "solid" grounds: *wooden parquet* floor, *metal* (metallic catwalk), or *concrete*. Such sounds can also evoke a whole place and being in a different location such as with echoing walking sounds in a *forest* or in a *cave*.

5.3 Event sounds

The third category comprises different sounds of walking over animate or inanimate objects, and other unexpected events. We can evoke for example walking over the tail of a *cat* (meowing suddenly), or over a deforming *can*. We can also play the sounds of a dog's *bark*, a *shooting*, an *explosion* or *thunderbolts*. As an example, these sounds can be randomly generated by the system at one step to surprise the user.

5.4 Unusual or abstract sounds

The fourth category corresponds to the simulation of very unusual, and sometimes poetic, sounds. This is the case for instance with walking over *water* (with steps in a small amount of water, such as in a water patch), or with *beeps* or strange *noises.*

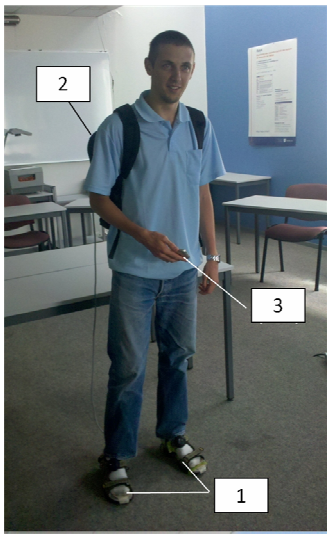


Figure 4: The system as worn by the user.

5.5 Sequences and cycles

The fifth category corresponds to the playing of different sounds at successive steps, in sequence or cycle. As an example, we can

use a sequence of steps to pronounce different syllabuses or a sentence such as *"I-am-your-fa-ther"* (the famous quote from Stars War's movie and Darth Vader character), or a music play such as the basic *"do-re-mi-fa-sol-la-si-do"* tune.

These categories and the corresponding examples are helpful to illustrate, in a non-exhaustive way, the richness and variety of sounds that can be used. They can be easily extended with other sounds and other approaches, using for instance more semantics and other parameters of the walk.

6. DISCUSSION

We believe Shoes-Your-Style concept has various kinds of possible application and extension. The first, and major, application concerns the entertainment field, and the design of mobile and entertaining "augmented walking experiences". Shoes-Your-Style could spread by extending existing technologies (e.g., Nike products) and, later, apply to Smartphone technologies.

Shoes-Your-Style might also support fitness and training shoe-based applications that could include auditory feedback related to steps, movements, or performance of runners. Shoes-Your-Style could also apply for medical applications and treatment of balance disorders or motor deficiencies. Indeed, this new source of sensory feedback might influence and alter the real walk of the user and induce specific walks. But future work is now needed to design the corresponding auditory feedback, and assessing the potential impact of this technology.

As a result, we think that Shoes-Your-Style concept could open an interesting path of research on feet-based mobile interaction with auditory feedback, and the study of novel kinds of walking experiences.

Future work could first consist in conducting a series of experiments to evaluate the design of our system and the potential applications of Shoes-Your-Style. We could also study and design a Smartphone version of Shoes-Your-Style based on inertial sensors and loudspeakers of an iPhone for instance. This solution could be compared to the shoes-based solution.

Besides, we could also study the design of other kinds of sound and other walking experiences based on auditory feedback. For instance, we could propose to use the force sensors located at the level of the toes for other kinds of rendering and sensations. We could also give the possibility to record sounds using a microphone (such as the one embedded in a Smartphone) to display them later with the system. The volume of the sound could be set proportional to the intensity of the contact or to the detected pressure. Physical simulation of sounds [4] could also be incorporated to increase the realism of played sounds.

7. CONCLUSION

We have introduced the "Shoes-Your-Style" approach, meant to enhance the walking experience by displaying artificial sounds of footsteps during mobile indoor and outdoor walks.

This concept is illustrated with "Sonic-Shoes" which embed both pressure sensors (to detect foot-floor contacts) and loudspeakers (to display sounds of footsteps) within a "all-in-one" system.

We could propose and test various kinds of sound, such as sounds of other shoes, of varied grounds, or of unexpected events. The applications envisioned for such new feet-based interaction scheme belong to various fields ranging from multimedia and entertainment, to sport training and medical rehabilitation/reeducation.

8. ACKNOWLEDGMENTS

This work was supported by the European Community under FP7 FET-Open grant agreement 222107 NIW - Natural Interactive Walking.

9. REFERENCES

[1] Bresin, R., de Witt, A., Papetti, S., Civolani, M., and Fontana, F. 2010. Expressive sonification of footstep sounds. In *Proceedings of Interactive Sonification Workshop*.

[2] Frey, M. 2006. CabBoots: shoes with an integrated navigation system. In *Proceedings of Ars-Electronica*.

[3] Nordahl, R., Berrezag, A., Dimitrov, S., Turchet, L., Hayward, V., and Serafin, S. 2010. Preliminary experiment combining virtual reality haptic shoes and sound synthesis. In *Proceedings of Eurohaptics.*

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

[5] Paradiso, J., Abler, C., Hsiao, K.Y., and Reynolds, M. 1997. The magic carpet: physical sensing for immersive environments. In *Proceedings of ACM CHI Conference*.

[6] Rovers, A.F. and Van Essen, H.A. 2005. FootIO - design and evaluation of a device to enable foot interaction over a computer network. In *Proceedings of World Haptics Conference*.

"Tap, Squeeze and Stir" the Virtual World: Touching the Different States of Matter Through 6DoF Haptic Interaction

Gabriel Cirio * INRIA Rennes
Maud Marchal * INSA/INRIA Rennes
Aurélien LeGentil * INRIA Rennes
Anatole Lécuyer * INRIA Rennes

ABSTRACT

Haptic interaction with virtual objects is a major concern in the virtual reality field. There are many physically-based efficient models that enable the simulation of a specific type of media, e.g. fluid volumes, deformable and rigid bodies. However, combining these often heterogeneous algorithms in the same virtual scene in order to simulate and interact with different types of media can be a complex task. In this paper, we propose the first haptic rendering technique for the simulation and the interaction with multistate media, namely fluids, deformable bodies and rigid bodies, in real-time and with 6DoF haptic feedback. Based on the Smoothed-Particle Hydrodynamics (SPH) physical model for all three types of media, our method avoids the complexity of dealing with different algorithms and their coupling. We achieve high update rates while simulating a physically-based virtual world governed by fluid and elasticity theories, and show how to render interaction forces and torques through a 6DoF haptic device.

Index Terms: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Physically based modeling; H.5.1 [Information Interfaces and Presentation]: User Interfaces—Haptics I/O; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, Augmented, and Virtual Realities

1 INTRODUCTION

The computer graphics community has developed many different physically-based models to recreate materials with which we interact on a daily basis: smoke, fire, water, clay, plastic and metal, to name a few. We are used to see these simulations in movies or video games, but what about touching them? Haptic rendering algorithms compute force feedback in real-time, and allow to feel the objects in a virtual scene through a haptic device. Due to the high complexity of physical simulations and the high update rate required for haptics, these algorithms often focus on enabling the haptic interaction with a single type of material: either rigid, deformable or fluid media. However, even the most common activity in real life often involves interacting with more than one type of media: walking on pavement, parquet and puddles, having a meal on a plate with meat and a beverage, washing our hands by touching the faucet, the soap and the water. In order to bodily interact with such virtual scenes, we currently need to use different algorithms for each type of media, and set up coupling interfaces between these heterogeneous models. This supposes to increase the complexity and the computational cost of an already highly complex and time consuming simulation.

In this work, we propose a single model with a single haptic rendering technique to simulate and interact simultaneously with rigid, deformable and fluid media in real-time with 6DoF haptic feedback. This novel approach is the first to tackle the problem of physically-based multistate haptic rendering. Based on a Smoothed-Particle Hydrodynamics (SPH) model brought to high update rates through a GPU implementation, we are able to simulate fluid volumes, deformable and rigid bodies in real-time. Using a novel coupling scheme that takes advantage of the SPH model, we show how to render interaction forces and torques through a 6DoF haptic device. We measure the performance of our technique, and collect user feedback during preliminary testing.

*e-mail: {gcirio, mmarchal, alegentil, alecuyer}@inria.fr

2 PREVIOUS WORK

Previous work on haptic interaction with deformable bodies [2] or fluids [4] only allow the interaction with rigid bodies on top of another state. The CORDIS-ANIMA [3] simulation framework was the first to simulate different states with haptic feedback, although using a mass-spring formalism and not the physically-based approach of continuum mechanics. To the best of our knowledge, there is no haptic rendering technique that provides haptic feedback for fluid, deformable and rigid states of matter in the same simulation. Recent work on physically-based simulations has used the Smoothed-Particle Hydrodynamics model for the simulation of multiple states of matter in a unified framework. However, update rates are usually not real-time, or fall short for haptic interaction. In [10], deformable bodies are modeled in real-time with SPH and the Moving Least Squares (MLS) algorithm to compute the elastic forces. The approach is improved in [7], allowing solid, deformable and fluid animation and interaction in a unified approach. A fully SPH-based approach is presented in [12], later corrected in [1] for a rotationally invariant formulation. Although these approaches allow a certain degree of interaction, they are not fast enough and do not solve the haptic rendering issue.

Our approach is the first to achieve multistate haptic rendering. Previous haptic rendering techniques were focused on a single state, e.g. fluid, deformable or solid, making it difficult to have different types of media coexist in the same simulation while providing convincing haptic feedback. Inspired by the work presented in [4], and based on the SPH [8] model, we are able to simulate simultaneously fluid volumes, deformable bodies and rigid objects in a physically-based manner and at high update rates. We improve the work presented in [9] for fluids, and also improve [12] and [1] for deformable bodies by bringing them to the required speeds, and enable the haptic coupling with 6DoF haptic devices.

3 PHYSICALLY-BASED SIMULATION FOR HAPTIC RENDERING OF MULTISTATE MEDIA

The seamless cohabitation of different theoretical models for different states of matter within the same virtual world is made possible thanks to the use of a single common computational approach: the SPH model. With all three states of matter simulated with the same data structures and the same algorithmic, the coupling between the different media becomes a straightforward task, and is included in the simulation. Figure 1 illustrates this principle. Any virtual object of the scene can be either fluid, deformable or rigid, and the state is hidden inside the model. The interactions between the different states are also computed seamlessly based on the same model. In the end, a virtual scene is populated by defining the state of each virtual object together with its corresponding physical properties

(stiffness for deformable bodies, viscosity for fluids). Users can interact with the virtual world through a haptic device using the interaction forces, without ever dealing with the models behind the virtual objects.

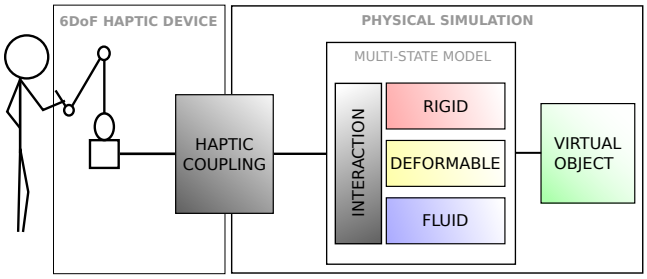


Figure 1: Global view of our approach: a user seamlessly interacts with a scene populated with virtual objects of different states of matter.

In this section, we describe the foundations of our physically-based simulation algorithms.

3.1 SPH-based Physical Simulation

The SPH model [8] computes the smoothed quantity $A_i(\mathbf{x})$ of a particle i at any position $\mathbf{x}$ in space through the general formula:

$$A_i(\mathbf{x}) = \sum_j A_j V_j W(\mathbf{x} - \mathbf{x}_j, h) \tag{1}$$

where A_j is the discrete quantity $A(\mathbf{x}_j)$ sampled for neighboring particle j at position $\mathbf{x}_j$, V_j is the volume of j, and W is the smoothing kernel of support h, where particles farther than h are not taken into account.

3.1.1 Fluid and Rigid Body Model

The motion of fluids is driven by the Navier-Stokes equations. Using the SPH model as presented in [9], pressure and viscosity forces are computed at each time step:

$$\mathbf{f}_i^{pressure}(\mathbf{x}_i) = -V_i \sum_j V_j \frac{P_i + P_j}{2} \nabla W(\mathbf{x}_i - \mathbf{x}_j, h) \tag{2}$$

$$\mathbf{f}_i^{viscosity}(\mathbf{x}_i) = \mu V_i \sum_j V_j (\mathbf{v}_j - \mathbf{v}_i) \nabla^2 W(\mathbf{x}_i - \mathbf{x}_j, h) \tag{3}$$

where P is the pressure, $\mathbf{v}$ the velocity, μ the viscosity coefficient and $\mathbf{g}$ the gravity field. ∇ and ∇^2 are respectively the gradient and Laplacian of the physical quantities. Parameters such as the viscosity can be changed to obtain different fluid behaviors, from smoke (no viscosity) to honey (very high viscosity).

Rigid bodies are modeled with the same SPH particles used in the fluid simulation, creating a unified particle model as described in [12]. This simple rigid body model has the main advantage of being meshless: since the overall shape of the rigid body is not important, it allows the seamless use of arbitrary-shaped rigid bodies, including concave objects. Moreover, it provides a gain in efficiency by computing a single model, since it makes the approach unified with fluid and deformable bodies. Additional collision detection techniques are avoided, since collision computations are included in the neighbor search.

3.1.2 Deformable Body Model

In order to simulate a deformable body with SPH governed by the elasticity theory, we compute elastic forces as in [12] using the improvements proposed in [1] to make the computations rotationally invariant by using a corotational approach. For more details, we refer the reader to the respective articles, as well as to [10].

The elastic force exerted on a particle i is computed from the gradient of the locally rotated displacement field, $\nabla\bar{\mathbf{u}}_i$. Using an SPH formulation, this gradient is defined as:

$$\nabla\bar{\mathbf{u}}_i = \sum_j V_i^0 \bar{\mathbf{u}}_{ji} \nabla W(\mathbf{x}_i^0 - \mathbf{x}_j^0, h)^T \tag{4}$$

where $\mathbf{x}_i^0$ and $\mathbf{x}_j^0$ are the initial positions (in the undeformed state) of particles i and j respectively, and $\bar{\mathbf{u}}_{ji}$ is the locally rotated displacement difference between neighboring particles j and i. Then, the strain ε can be computed from $\nabla\bar{\mathbf{u}}_i$, using a Green-Saint-Venant strain tensor forumulation.

The elastic force $\mathbf{f}_{ji}$ exerted on a neighboring particle j by particles i is defined as the negative gradient of strain energy with respect to displacement, and can be computed as:

$$\mathbf{f}_{ji} = -2V_i^0 (\mathbf{I} + \nabla\bar{\mathbf{u}}_i^T)\sigma_i V_j^0 \nabla W(\mathbf{x}_i^0 - \mathbf{x}_j^0, h) \tag{5}$$

where V_i^0 is the initial volume of particle i and σ is the stress. σ is computed from the strain ε through the linear relation $\sigma = \mathbf{C}\varepsilon$, where $\mathbf{C}$ is defined by only two parameters, the Young modulus and the Poisson ratio.

The elastic force $\mathbf{f}_i^{elasticity}$ exerted on particle i is symmetrized and computed as:

$$\mathbf{f}_i^{elasticity} = \sum_j \frac{-\mathbf{R}_i \mathbf{f}_{ji} + \mathbf{R}_j \mathbf{f}_{ij}}{2} \tag{6}$$

where $\mathbf{R}$ is the rotation matrix computed through the corotational approach.

3.1.3 Interaction Forces

In a unified, parallel and time-critical framework, unified interaction forces fit well for computation time reasons. Using the same interaction forces than in a fluid-fluid case, namely pressure and viscosity forces, improves the gain of parallel computation, while providing a reasonable amount of control over the forces through density and viscosity values. Moreover, by using the SPH model for the computation of interaction forces between fluid, deformable and rigid body particles, it removes the need of additional collision detection algorithms.

3.2 Integration and Simulation Loop

The different models previously described are run simultaneously under the same SPH framework, improving the efficiency of the simulation compared to having three heterogeneous models. Hence, the simulation loop runs in four consecutive steps:

1. Compute corotational matrices $\mathbf{R}_i$ for each particle belonging to a deformable body;
2. Compute new density and elastic properties;
3. Compute new internal (between particles of the same entity) and interaction (between particles of different entities) forces;
4. Integrate (compute new velocity and position) through a Leap-Frog integration scheme for fluids and deformable bodies. For rigid bodies, integrate by applying rigid body dynamics.

The entire physical model is implemented on GPU using the CUDA framework, extending our previous implementation [4] to support deformable bodies.

4 Multimodal Rendering

4.1 6DoF Haptic Rendering

In the previous section we showed how to simulate different states (fluid, deformable and rigid bodies) and their interactions with the same SPH model. Our GPU implementation allows to increase the update rate to speeds enabling haptic interaction. In this section, we detail our coupling scheme for fluid, deformable and rigid body interaction through a 6DoF haptic device in order to obtain force feedback out of a multiple state simulation.

We base our force feedback computations on interaction forces and on a SPH rigid body as proxy. By using an SPH proxy, we avoid additional computations for haptic coupling: the proxy is a rigid body of the virtual scene, and it interacts with all other media around it through the unified interaction forces described in section 3.1.3. When the proxy interacts with different media on the scene, the simulation computes the corresponding interaction forces as any other body from the scene. In this way, the user can feel the virtual world through an arbitrary-shaped rigid body.

Once the interactions forces are computed on the proxy, they are fed to the rigid body dynamics algorithm to perform the haptic coupling between the proxy and the 6DoF haptic device. In the rigid body dynamics step, all forces $\mathbf{f}^{haptic}$ exerted on the proxy rigid body are summed to obtain a total force and a total torque. The force and torque feedback coming from the haptic device are added to the rigid body. Then, the new position and velocity are computed by integrating forces over the simulation time step. They are sent to the haptic device, closing the haptic loop in admittance mode. A Virtual Coupling mechanism [5] is introduced between the haptic device and the rigid body, reconciling a high update rate haptic device with a lower rate simulation, leading to an increase of stability.

The resulting 6DoF haptic coupling scheme is unified, in the sense that it allows the interaction with different media (fluid, deformable and solid) without distinction, since the haptic forces are computed in a unified way.

4.2 Graphic Rendering

As for the physical simulation, we use a unified visual rendering algorithm for fluid and deformable bodies. Previous work [4] has shown how to render fluids with a screen-space technique based on [13], with a trade-off between quality and performance. The three-step approach first computes per pixel fluid data (front and back fluid volume depth from view), then smoothes the front surface depth in screen space using a fast bilateral filter, and finally composes the frame.

We follow the same procedure to render deformable bodies. However, since most bodies are not transparent, as opposed to fluids, we do not handle transparency for deformable bodies, which leads to an increase in performance. In order to achieve a correct occlusion between the different media, we perform a depth test in the fragment shader.

5 Performance

In this section we measure the performance of our haptic rendering technique, in terms of computation time and user feedback through the subjective appreciation of users interacting with different media.

The simulations were carried out using a Virtuose 6DoF force-feedback device from Haption [1], and a laptop computer with a Core 2 Extreme X7900 processor at 2.8GHz, 4GB of RAM memory, and a Nvidia Quadro FX 3600M GPU with 512MB of graphic memory.

5.1 Computation Time

We measured the computation time of three scenes with an increasing number of particles. Each scene has either a fluid volume, a deformable body or a rigid body, all of the same size and shape (a cuboid with a square base of 0.5m side length and a variable height according to the number of particles). We only measured the physical simulation computation time, without taking into account graphic rendering. The simulation ranges from 2,000 to 30,000 particles. Figure 2 plots the measurements for the three different types of media.

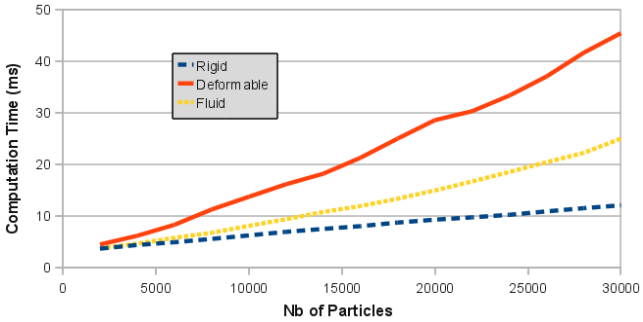


Figure 2: Computation time for scenes made of either a rigid body (dashed curve), a deformable body (dotted curve) or a fluid volume (continuous curve), according to the number of particles.

The higher complexity of the elastic theory implementation compared to the fluid theory and the rigid body dynamics is clearly visible: when the number of particles increases, the computation time for deformable bodies increases roughly 2 times faster than for fluids, and 6 times faster than for rigid bodies.

A determinant factor of our approach is its computation time, since haptic feedback requires update rates higher than the usual 30Hz of real-time visual rendering. It has been shown that, for fluids, update rates of 70Hz are satisfactory for haptic rendering, since forces due to fluids do not change rapidly [4]. Our implementation allows to simulate large volumes of fluid (around 20,000 particles) above 70Hz, as shown in Figure 2. In the case of deformable bodies, low update rate simulations (30-50Hz) use interpolation [11], and faster models rely extensively on pre-computed data and small deformation approaches [6], or limit the complexity of the virtual objects [2]. Our model is an efficient trade-off between these approaches, with a fast simulation with complex deformable bodies (150Hz for 5,000 particles) and a virtual coupling mechanism for a smooth haptic rendering. For rigid bodies, our implementation can achieve 230Hz for arbitrary-shaped bodies of 5,000 particles. While falling short of the 1kHz update rate required for optimal haptic rendering, the first users did not find this problematic.

5.2 User feedback

We asked several users, unfamiliar with haptic devices, to interact with different types of media through our novel multistate approach in order to have a preliminary qualitative feedback. Although using haptic feedback alone users could not efficiently recognize the different states of matter, the addition of haptic feedback on top of visual feedback increased the the degree of realism and the general appreciation of the users. Their feedback was very positive regarding the combination of both modalities.

These preliminary tests showed that deformable bodies could benefit from an increase in update rate. High stiffness values sometimes lead to instabilities that are quickly felt through the haptic interface. At low stiffness values this problem is less frequent, or not present at all. The framerate of the simulation condition the integration step, due to our explicit integration scheme. Increasing the frequency of the simulation would allow the use of a smaller integration step, leading to an increase in stability and the simulation of stiffer materials. Rigid bodies would improve in perceived stiffness as well, even though feedback was positive for this media.

[1] Soulge-sur-Ouette, France

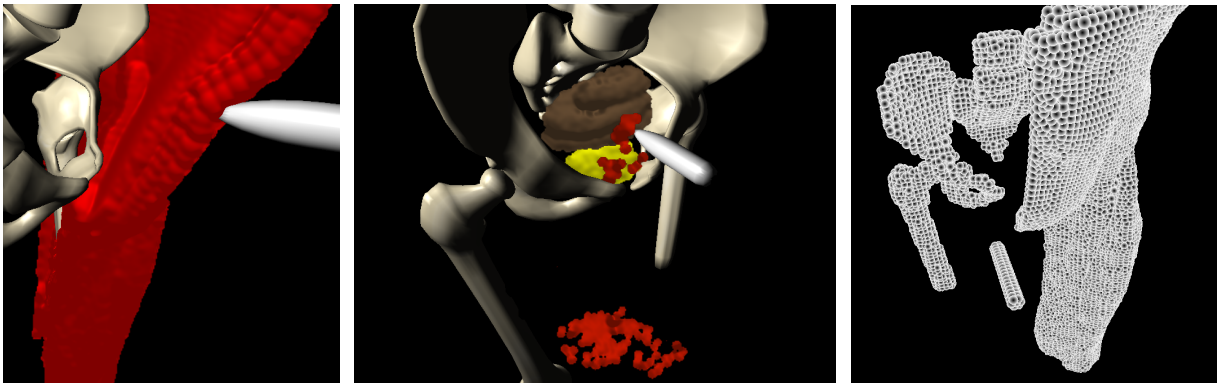


Figure 3: 6DoF haptic interaction with multiple states of matter in an illustration of a medical simulation scenario. Left: the user examines a muscle. Middle: blood pours from the intestine when the user penetrates it with the probe. Right: rendering of the particles composing the scene.

A possible optimization would be the decoupling between haptic and graphic rendering, with different update rates according to the requirements. As both use the GPU with blocking tasks, a multiple GPU architecture would allow such a separation, with a direct impact on the fps of the simulation.

Taken together, our approach efficiently fulfills its objectives, by allowing the simultaneous interaction with different states of matter in the same virtual scene with haptic feedback, and improving the interaction experience as highlighted by the preliminary users.

6 Application Scenarios

Our multistate haptic rendering technique can be applied to different fields, with haptic feedback enhancing the interaction experience. The medical field can substantially benefit from our approach, since surgeons are in constant interaction with fluids (such as blood), deformable bodies (organs, tissues) and rigid bodies (bones). We designed a medical simulation scene as an illustration of potential application, as shown in Figure 3, with these three states. The user can interact with the different body parts through a rigid body probe, and experience haptic feedback through a 6DoF haptic device. Where interactive medical simulations of this kind were restricted to deformable bodies only, we enable haptic feedback on scenes with complex rigid and deformable objects, together with volumes of fluid, without dealing with different models and their coupling.

Other potential application areas include the entertainment field (immersive epxloration of virtual worlds), training scenarios (manipulations of dangerous materials) and industrial simulations (virtual prototyping).

7 Conclusion

This paper introduces the first multistate haptic rendering approach, allowing haptic interaction simultaneously with media in fluid, deformable and rigid states. Based on the Smoothed-Particle Hydrodynamics (SPH) physical model for all three types of media, our method avoids the complexity of dealing with different algorithms and their coupling. We achieve high update rates while simulating a physically-based virtual world governed by fluid and elasticity theories, and show the multimodal rendering of the interaction, with forces and torques displayed through a 6DoF haptic device and with an optimized visual rendering algorithm. Different users interacted with the 3 types of media using our approach and a 6DoF haptic device, and stated that our technique improves the overall interaction experience.

Future work will focus on decoupling graphic and haptic rendering loops for an increased simulation performance. We will then conduct a perceptual evaluation in order to assess the capability of users to recognize the different states of matter.

Acknowledgements

This work was supported by the European Community under FP7 FET-Open grant agreement n°222107 NIW - Natural Interactive Walking.

References

[1] M. Becker, M. Ihmsen, and M. Teschner. Corotated SPH for deformable solids. In *Eurographics* Workshop on Natural Phenomena Proceedings, 2009.

[2] G. Burdea, G. Patounakis, V. Popescu, and R. Weiss. Virtual reality training for the diagnosis of prostate cancer. In *IEEE VR Proceedings*, p.190–197, 1998.

[3] C. Cadoz, A. Luciani, and J.L. Florens. CORDIS-ANIMA: Modeling and Simulation System for Sound and image Synthesisthe General Formalism. *Computer Music Journal*, 17(3), 1993.

[4] G. Cirio, M. Marchal, S. Hillaire, and A. Lécuyer. 6 Degrees-of-Freedom Haptic Interaction with Fluids. *IEEE Transactions on Visualization and Computer Graphics*, to appear.

[5] J. Colgate, M. Stanley, and J. Brown. Issues in the haptic display of tool use. In *IEEE/RSJ Intelligent Robots and Systems Proceedings*, vol.3, p.3140, 1995.

[6] D. L. James and D. K. Pai. A unified treatment of elastostatic contact simulation for real time haptics. In *ACM SIGGRAPH 2005 Courses*, p.141, 2005.

[7] R. Keiser, B. Adams, D. Gasser, P. Bazzi, P. Dutre, and M. Gross. A unified lagrangian approach to solid-fluid animation. In *Eurographics/IEEE Point-Based Graphics Symposium Proceedings*, p.125–148, 2005.

[8] J. J. Monaghan. Smoothed particle hydrodynamics. *Annual Review of Astronomy and Astrophysics*, 30(1):543–574, 1992.

[9] M. Muller, D. Charypar, and M. Gross. Particle-based fluid simulation for interactive applications. In *ACM SIGGRAPH/Eurographics Symposium on Computer Animation Proceedings*, p.154–159, 2003.

[10] M. Muller, S. Schirm, M. Teschner, B. Heidelberger, and M. Gross. Interaction of fluids with deformable solids. *Computer Animation and Virtual Worlds*, 15(3-4):159–171, 2004.

[11] G. Saupin, C. Duriez, and S. Cotin. Contact model for haptic medical simulations. In *International Symposium on Biomedical Simulation Proceedings*, Springer-Verlag, p.157–165, 2008.

[12] B. Solenthaler, J. Schlfli, and R. Pajarola. A unified particle model for fluid-solid interactions. *Comput. Animat. Virtual Worlds*, 18(1):69–82, 2007.

[13] W. J. van der Laan, S. Green, and M. Sainz. Screen space fluid rendering with curvature flow. In *ACM I3D Proceedings*, p.91–98, 2009.

The Virtual Crepe Factory: 6DoF Haptic Interaction with Fluids

Gabriel Cirio
INRIA Rennes*

Maud Marchal
INSA/INRIA Rennes*

Sebastien Hillaire
INRIA Rennes/Orange Labs*

Anatole Lécuyer
INRIA Rennes*

Figure 1: *The Virtual Crepe Factory allows users to interact with different viscous fluids to achieve a crepe preparation recipe. The user follows the different steps of the crepe preparation process (stirring and scooping the dough, pouring the dough into the pan, flipping the crepe, spreading the toppings) and feels corresponding force and torque feedback through two haptic devices, one in each hand.*

Abstract. The Virtual Crepe Factory illustrates our novel approach for 6DoF haptic interaction with fluids. It showcases a 2-handed interactive haptic scenario: a recipe consisting in using different types of fluid in order to make a special pancake also known as "crepe". The scenario guides the user through all the steps required to prepare a crepe: from the stirring and pouring of the dough to the spreading of different toppings, without forgetting the challenging flipping of the crepe. With the Virtual Crepe Factory, users can experience for the first time 6DoF haptic interactions with fluids of varying viscosity. Our novel approach is based on a Smoothed-Particle Hydrodynamics (SPH) physically-based simulation.

Introduction. Fluids are present in many applications such as for industrial or medical manipulations - involving for instance blood flow and natural liquids. The haptic simulation of fluids is particularly challenging, especially to achieve realistic, stable and real-time force feedback. Previous methods for haptic fluid interaction [Baxter and Lin 2004] [Dobashi et al. 2006] were limited to 3DoF, simple objects or pre-computed forces. Thus, as for today, there is a lack of haptic rendering techniques handling complex interactions with viscous fluids. Our approach allows real-time 6DoF haptic interaction with fluids of variable viscosity, through arbitrary shaped rigid bodies and 6DoF haptic devices. Particularly, fluid containers can be created to hold fluid, hence transmiting to the user strong forces such as fluid resistance and weight, as well as light forces like the inertia of the fluid inside the container.

Virtual Crepe Preparation Scenario. As shown in Figure 2, the user holds a 6DoF haptic device in each hand. The scenario can be divided in 4 distinct consecutives stages. The first stage requires the user to stir and scoop the dough from a bowl on the table with his scooping object. It enables the user to feel the torque generated by the resistance from a highly viscous fluid opposing his stirring movement, and the weight and inertia of the fluid inside the hand-held object. In the second stage, the content of the scoop is poured into the pan held with the second hand. The user feels the weight transferring from one hand to the other. The third stage is the solidification and flipping of the crepe. The user feels how mass shifts slow down as the viscosity of the dough raises with the solidification process. When the crepe is ready, it becomes a deformable object, and the user can try to flip the crepe by throwing it into the air, illustrating the underlying phase-change capabilities of our model. The fourth and last stage is the spreading of toppings on the crepe, which combines all the previous haptic interaction possibilities by making the user scoop, pour and spread a low viscosity fluid (maple sirup) and a high viscosity fluid (Brittany's salt-butter caramel).

*e-mail: {gcirio,mmarchal,shillaire,alecuyer}@inria.fr

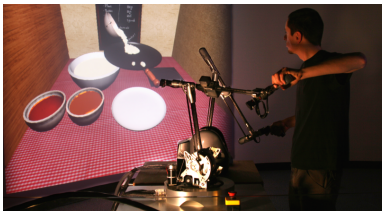


Figure 2: *Setup: Bimanual interaction with two 6-DoF haptic devices (Virtuose, Haption). Left hand for virtual bowl manipulation, Right hand for virtual pan manipulation.*

Technical Description. Our approach is based on an SPH physically-based simulation. We use a unified model to achieve a real-time simulation by using SPH particles for both fluids and rigid bodies. Our approach allows real-time simulation of solid-fluid interactions with arbitrary-shaped rigid bodies and fluids of different viscosities. Furthermore, we use a novel 6DoF haptic rendering based on SPH, providing a soft and continuous haptic feedback particularly well suited for the haptic interaction with fluids. This allows the seamless 6DoF haptic coupling between a haptic device and any rigid body of the virtual scene, but also between N haptic devices and the same rigid body (as holding a bucket with two hands). Our method is implemented on GPU and preliminary testings with standard PC led to 170Hz with 25,000 particles.

Vision. The Virtual Crepe Factory proves that fluids are now ready for complex 6DoF haptic interactions, and opens a new horizon of applications in Virtual Reality. Applications span from the medical field (organic fluids) to industrial scenarios (painting, manipulating dangerous fluids) and entertainment simulations (natural scenes, water sports). We challenge the SIGGRAPH audience: are you a good crepe maker... or a bad one?

Acknowledgements

This work was supported by the European Community under FP7 FET-Open grant n°222107 NIW - Natural Interactive Walking.

References

BAXTER, W., AND LIN, M. C. 2004. Haptic interaction with fluid media. In *Proceedings of Graphics Interface*.

DOBASHI, Y., SATO, M., HASEGAWA, S., YAMAMOTO, T., KATO, M., AND NISHITA, T. A fluid resistance map method for real-time haptic interaction with fluids. In *ACM VRST'06*.

Joyman: an immersive and entertaining interface for virtual locomotion

Julien Pettré* INRIA-Rennes
Orianne Siret INRIA-Rennes
Maud Marchal INSA / INRIA-Rennes
Anatole Lécuyer INRIA-Rennes

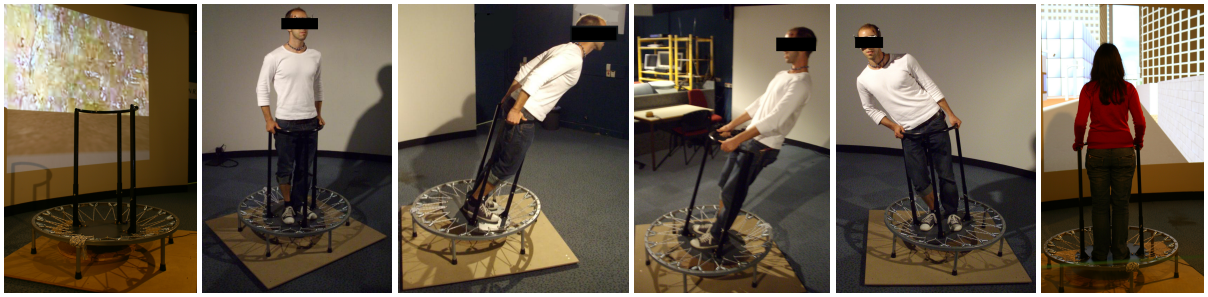


Figure 1: ***Left:*** *The Joyman interface.* ***Center (4 images):*** *Based on a simple mechanical design, the device allows a user to stand upon it and to lean in any direction without falling.* ***Right:*** *The inclination of the board of the device is transformed into a virtual velocity vector: this enables us to immerse users into environment navigation scenarios.*

Abstract We propose to demonstrate a novel interface called Joyman, designed for immersive locomotion in virtual environments. The interface is based on the metaphor of a "human-scale joystick". The device has a simple mechanical design that allows a user to indicate his virtual navigation intentions by leaning in the corresponding direction. It is often said that walking is equivalent to constantly falling ahead. The Joyman follows this assertion: whereas many previous interfaces preserve or stimulate the users proprioception, the Joyman aims at preserving equilibrioception in order to improve the feeling of immersion during virtual locomotion tasks. Thus, the interface could be applied in various virtual reality applications such as urban or architectural project reviews or training, as well as in videogames.

CR Categories: I.3.1 [COMPUTER GRAPHICS]: Hardware Architecture—Input Devices;

Keywords: 3D interface, device, locomotion, virtual worlds

Principle Navigation is one of the fundamental tasks needed for 3D interaction with Virtual Environments. However, it is still a challenge to allow a user endlessly walking in a large virtual world in an immersive manner when the physical real workspace is limited. Many interfaces and techniques were proposed. The Joyman is one of the most recent, as detailed in [Marchal et al. 2011]. While the interface is a low-cost one, it is also based on an innovative principle. Whereas most of previous interfaces proposed an immersive experience of walking in virtual worlds by preserving human proprioceptive senses, the Joyman exploits the human equilibrioception to drive the virtual locomotion. The Joyman is very intuitively used by a novice and the virtual locomotion is initiated identically to real one. The interface is based on the metaphor of a "human-scale joystick": the user leans in the direction of the desired motion (e.g., forward to start walking ahead). In a way, this experience is close to the one of using a Segway or skying.

The Joyman device The main objective of the Joyman is to provide realistic locomotion trajectories in a virtual environment. The interface meets this objective by combining two roles. It first allows users to lean safely beyond the point of falling, as illustrated in Figure 1. Falling is prevented by an adjustable barrier, rigidly fixed to a board upon which the user is standing. The board is finally articulated to a platform using springs which allows the board to tilt when the user is leaning. The second role of the Joyman interface is to transform the measured board inclination into a virtual velocity. This is the functionality of a control law that takes into account the basic constraints of the human locomotion (bounded velocities, accelerations and collinearity of the velocity vector with the tangent to the trajectory). The Joyman is often wrongly compared to a force platform because equilibrioception is involved when using such a device as well. However, users lean widely beyond the point of falling using the Joyman, which is not possible to do using a force platform and which deeply change the felt sensations.

Scenario The robustness of the Joyman enables to easily and safely demonstrate it to a large audience. We propose a basic scenario where a large-scale environment (e.g., small city) is explored. The scene can be visualized through a HMD or by projecting it to a screen as displayed in the companion video. The video shows that the proposed demonstration scenario is ready, nevertheless, we are currently working on extensions, new scenes, new control modes and new scenarios (skying, flying, etc.) that we expect to be available for the conference. The user can exploit his own equilibrioception to control his virtual locomotion through the different streets and buildings. By using the Joyman, we can also maintain a high level of immersion compared to handheld devices (e.g, joysticks). We have demonstrated the Joyman to be intuitively used and we believe this can provide attendees with a novel and fun experience. Various VR applications but also real applications could be envisaged when navigating in a 3D world. Thus, the Joyman interface can be used for example for videogames, rehabilitation, training tasks or virtual visits.

A patent for the interface presented in this submission has been filed on November 19, 2010 under the number FR10/595551.

References

MARCHAL, M., PETTRÉ, J., AND LÉCUYER, A. 2011. Joyman: a human-scale joystick for navigating in virtual worlds. In *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI 2011)*.

*e-mail: julien.pettre@inria.fr

Joint Virtual Reality Conference of EuroVR - EGVE (2011)
R. Blach, S. Coquillart, M. D'Cruz, A. Steed, and G. Welch (Editors)

Short Paper: Comparing Virtual Trajectories Made in Slalom Using Walking-In-Place and Joystick Techniques

Léo Terziman[†1,2,3], Maud Marchal[†1,2], Franck Multon[†2,4], Bruno Arnaldi[†1,2] and Anatole Lécuyer[†2]

[1]Institut National des Sciences Appliquées, Rennes, France
[2]Institut National de Recherche en Informatique et Automatique, Rennes, France
[3]Direction Générale de l'Armement, France
[4]Université de Rennes 2, France

Abstract

In this paper we analyze and compare the trajectories made in a Virtual Environment with two different navigation techniques. The first is a standard joystick technique and the second is the Walking-In-Place (WIP) technique. We propose a spatial and temporal analysis of the trajectories produced with both techniques during a virtual slalom task. We found that trajectories and users' behaviors are very different accross the two conditions. Our results notably show that with the WIP technique the users turned more often and navigated more sequentially, i.e. waited to cross obstacles before changing their direction. However, the users were also able to modulate their speed more precisely with the WIP. These results could be used to optimize the design and future implementations of WIP techniques. Our analysis could also become the basis of a future framework to compare other navigation techniques.

Categories and Subject Descriptors (according to ACM CCS): Computer Graphics [I.3.7]: Three-Dimensional Graphics and Realism—Virtual Reality; Computer Graphics [I.3.6]: Methodology and Techniques—Interaction Techniques

1. Introduction

In Virtual Reality (VR) applications, navigation is one of the fundamental 3D interaction tasks [BKLP05]. The ability to walk is one of the most basic and essential need to complete other more complex tasks within Virtual Environments (VE) (Figure 1). However, most of the time, the working space is limited and real walking cannot be used. Various techniques have been developed so far, based on specific devices such as treadmills or on software metaphors.

The Walking-In-Place (WIP) technique [SUS95] has been introduced to enable a real physical walking movement during virtual navigations. With WIP, the user can move infinitely inside the VE by consciously walking in place in the real world. The motions of the user's body are tracked and analyzed to be converted into inputs for the locomotion simulation into the VE. The first implementations were based on a neural network to detect the footsteps of the user [SUS95]. More recent implementations use frequency analysis [FWW08] or biomechanical state automaton [WWB10]. Another recent WIP implementation proposes to track the user's head motions with a webcam [TME*10].

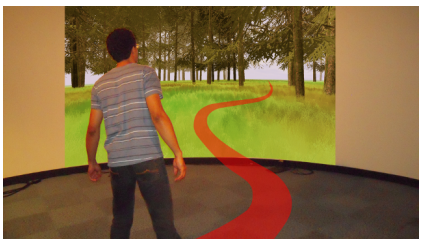


Figure 1: *Slalom navigation in a Virtual Environment.*

† firstname.lastname@inria.fr

© The Eurographics Association 2011.

Most of the evaluations of the WIP technique are based on subjective questionnaire to quantify the presence or potential motion sickness induced by this technique [RSS*02] [UAW*99]. Whitton et al. [WCF*05] have performed a quantitative study to compare the WIP technique to natural walk and joystick-based First-Person Shooter (FPS) techniques, in terms of task performance and kinematics of the task criteria, but only for straight lines. Other studies compared the trajectories produced by different interaction techniques, but only on qualitative criteria [ZLB*05]. However, the navigation techniques are usually used in complex navigations, including turns and changes of speed. To the authors' best knowledge, the WIP has never been evaluated on such kind of trajectories so far. In this paper, we analyse and compare the trajectories produced by the WIP with ones produced by joystick on slalom paths.

2. Experimental Data

The trajectories of 2 navigation techniques are compared: the WIP technique and a standard navigation using a joystick. We chose to implement the WIP technique proposed in [TME*10] using a webcam-based tracking (25 Hz capture), while the joystick was based on a classical analogical implementation. For the WIP technique, the advance speed was computed based on the lateral amplitude and frequency of the user's head oscillations. Both techniques were calibrated to provide the same average speed during the navigations. Because of the limited Degrees of Freedom of the tracking system, the orientation of the head on the roll axis was used to turn. For both techniques, the turns were based on a rate control law of $45^\circ/s$. The WIP implementation, as well as the experimental apparatus used are described with more details in [TME*10]. During the experiment, the participants had to navigate in the VE through 2 different paths composed of eight 3m x 3m gates regularly placed composing a slalom. The instruction received by the participants was to cross the gates in the correct order.

Population: Twelve participants (10 males and 2 females) aged from 22 to 35 ($Mean(M) = 25.4$, $Standard\ Deviation(SD) = 3.35$) performed the experiment. All participants were used to VE but were naïve with respect to the WIP technique and the experimental setup.
Apparatus: The evaluation was performed within a 3D virtual environment without any contextual cues, except for gates that the user had to navigate through. A fog effect was added to mask the distant gates, allowing to perceive only the 2 or 3 closest gates. A texture on the ground provided visual flow information during the navigation. Participants were at a distance of 1.5 m in front of a 1.72 m large and 1.24 m height back-projected screen (with a physical field of view of 60° horizontally and 45° vertically). The resulting image had a resolution of 1600×1200 pixels.
Experimental plan: The participants were exposed to 2 blocks of 12 trials each: one block for each of the experimental conditions. Half of the participants started with the joystick condition, while the other half started with the WIP.
Collected Data: The trajectories of the participants in the VE were recorded at a frequency of 60 Hz for every trial. The orientation of the virtual camera was also recorded.

3. Trajectories Analysis

The proposed evaluation is based on both spatial and temporal analysis. Moreover, the spatial analysis focuses on both macroscopic and microscopic aspects of trajectories.

3.1. Spatial Analysis

The produced trajectories are time parametric functions. However, some relevant information can be extracted from the shape of the trajectories, independently of time.

Distance to Shortest Trajectory: To characterize the differences between the two conditions, we computed for each participant the mean difference between the produced trajectories and the shortest trajectory for each condition. The shortest trajectory was defined as the trajectory composed of straight lines between the inner post of each gate. We conducted a one-way repeated measure ANOVA on the technique used (joystick and WIP). The ANOVA revealed a significant dependency between the distance to the shortest trajectory and the interaction technique used ($F(1,11) = 16.94$, $p = 0.002$). The area between the trajectory and the shortest trajectory is significantly smaller with the joystick technique ($M = 86.55\ m^2$) compared to the WIP technique ($M = 110.93\ m^2$).

Mean Trajectory: To further stress the differences between the two techniques, we computed the mean trajectory for all participants for each set of conditions for the two paths. We re-sampled the trajectories with samples every $1m$ in the forward direction. The new samples were computed as the mean of all the values in the sample interval. A sample of the mean trajectories for the joystick and WIP conditions on the first path is displayed in Figure 2. The standard deviations of the mean trajectories represented in Figure 2 stress higher variability of the trajectories produced with the WIP technique. Moreover, the behaviour during the crossing of the gates seems to be different.

For each gate of the 2 mean trajectories, we computed the signed distance of the inflexion point to the closest gate, as well as the distance to the center of the gate when the users crossed the gates. For both criteria, we conducted a one-way repeated measure ANOVA on the technique used (joystick and WIP). We found a significant effect of the technique on the position of the inflexion points ($F(1,13) = 17.33$, $p = 0.001$). With the WIP technique the inflexion point is located after the gate ($M = 0.14\ m$), while it is located before with the joystick technique ($M = -0.42\ m$). We also

© The Eurographics Association 2011.

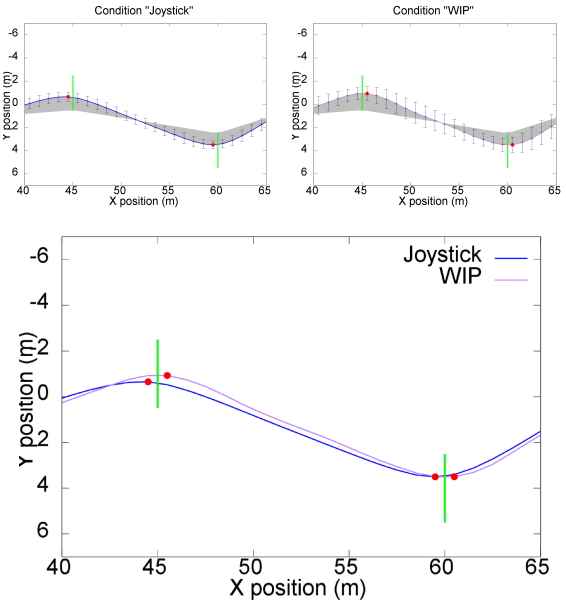


Figure 2: *Samples of the mean trajectories of WIP and joystick. The error boxes represent the standard deviations. Gates are represented in green, the distance to the shortest trajectory in gray and the inflexion points in red.*

found a significant effect of the technique on the distance to the center of the gate during the crossings ($F(1,13) = 34.75$, $p < 0.001$). Thus, the user crossed the gates closer to their centers with the WIP technique ($M = 0.38$ m) compared to the joystick ($M = 0.63$ m). In other words, with the WIP technique, the participants turned after crossing the gates, while they turned before with the joystick.

Curvature of the Trajectory: In order to further stress the differences in the behavior during the turns, we focused our analysis on the curvature of the trajectories. We computed the curvature at a microscopic level of trajectories parametrically defined as plane curves over time. This elicits a high difference in magnitude of curvature peaks between WIP and joystick. The Figure 3 illustrates this difference for two trials of the 6^{th} participant. The signal made of the curvatures of the trajectory over time is composed of impulses. In order to count and detect impulses, we used a first order continuous-time low-pass filter with a cut-off frequency of 10 Hz to filter the computed curvatures, and set a detection threshold for impulses at 0.02 m^{-1}.

To characterize the properties of the curvature for the different conditions, we computed 3 different criteria based on the detected impulses: (1) the number of impulses per trial, (2) the length of the impulses (in seconds) and (3) the area of each impulse. For each criterion, we computed the mean for each participant and we conducted a one-way repeated

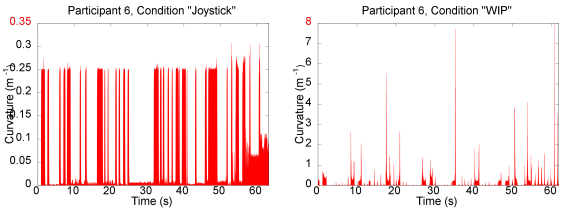


Figure 3: *Curvatures of trajectories for two trials of the 6^{th} participant.*

measure ANOVA on the technique used (joystick and WIP). The analysis revealed a significant dependency to the technique used for all 3 criteria. The results are summarized in Table 1.

Criterion	$F(1,11)$	p
Number of impulses	30.27	< 0.001
Length of impulses	13.43	0.004
Area of impulses	6.77	0.025

Table 1: *Results of the ANOVA on the curvature criteria.*

The participants significantly turned more often with the WIP technique ($M = 1108.9$, $SD = 284.6$) than with the joystick ($M = 608.9$, $SD = 77.2$). However, they turned significantly for a shortest amount of time with the WIP ($M = 0.33$ s, $SD = 0.08$ and $M = 0.45$ s, $SD = 0.07$ respectively). Finally, the area of impulses revealed that the participants significantly turned more with the WIP technique ($M = 0.19$ $s.m^{-1}$, $SD = 0.14$) than with the joystick ($M = 0.08$ $s.m^{-1}$, $SD = 0.02$).

3.2. Temporal Analysis

Speed: We computed the advance speed of the participants depending on the experimental conditions. We filtered and re-sampled the data using the same method as with the computation of the mean trajectories. To characterize the speed behaviour during the trajectories, we computed the average speed for all participants at each gate. We conducted a one-way repeated measure ANOVA on the technique used. We found a significant effect of the technique used on the speed at each gate ($F(1,13) = 195.46$, $p < 0.001$). The users were faster with the WIP technique ($M = 2.48$ $m.s^{-1}$, $SD = 0.1$) compared to the joystick technique ($M = 2.10$ $m.s^{-1}$, $SD = 0.007$). These results suggest that participants tended to cross the gates at full speed with the WIP technique.

Smoothness: The maximum smoothness model [PHA*07] was used as another criterion. According to this model, natural motions tend to be as smooth as possible to

© The Eurographics Association 2011.

minimize the energy involved. For each participant, we computed the mean value of the integrated absolute value of the jerk over the time for each trajectory. Greater jerk corresponds to a smaller smoothness of the trajectory and thus can be considered as less natural. A one-way repeated measure ANOVA on the technique used (joystick and WIP) revealed a significant dependency between the cumulated jerk and the technique used ($F(1,11) = 59.56$, $p < 0.001$). The trajectories are smoother with the joystick technique ($M = 10.7\ m.s^{-2}$, $SD = 1.6$) compared to the WIP technique ($M = 197.5\ m.s^{-2}$, $SD = 84.0$). Thus, the trajectories produced by the joystick seem to be closer to realistic natural walking trajectories compared to WIP, according to the smoothness model [PHA*07].

4. Discussion

Our results show strong differencies between trajectories produced by the WIP and the joystick, according to different criteria. The characteristics of the curvatures highlight more frequent and tighter changes in direction with the WIP technique. It seems that the users were not able to predict their future trajectory accurately with the WIP technique, resulting in a continuous adaptation of the advance direction. The position of the inflexion points and the distance to the center of the gates at each gate are consistent with this hypothesis. The implementation of the turns with the WIP might explain those results. Indeed, if the user's head oscillations were too important, turns could be triggered inappropriately.

Moreover, our results suggest that the trajectories produced with the WIP technique are less "efficient" in terms of traveled distances as well as naturalness compared to those made using the joystick. This difference could be explained by the difference in training of the participants. Indeed, all of them were proficient with the joystick but were using WIP for the first time. The control law for turns can be another reason, as the participants were apparently less able to choose their direction accurately with the WIP technique.

Finally, the curvature of the trajectories is clearly more important with the WIP technique. The participants tended to follow tighter turns. However, the underlying implementation of the turns for both joystick and WIP conditions were strictly identical. Thus, the participants decreased their speed during the turns with the WIP. The users seem more able to modulate their speed precisely with the WIP, which provided more varied navigation patterns. These results could provide guidelines for future implementations of turns when designing a WIP technique.

5. Conclusion

In this paper, we compared a WIP technique to the joystick interface. When observing slalom trajectories produced with both techniques, we found that the navigation strategy of the user was different with the WIP technique compared to the joystick. The participants had more difficulty to anticipate their trajectory with the WIP. We also found that the speed during the turns decreased and the user modulated their speed more precisely with the WIP technique. The trajectories produced by the WIP had more jerk than those produced by the joystick and thus were less likely to feel natural for the users. However, the WIP provided a better control of advance speed while the joystick was more precise for controlling direction.

Taken together, our results provide insights on the behavior of users when navigating using WIP vs. joystick. They could inspire guidelines for future design and implementations of WIP techniques, notably concerning the handling of turns. Last, a future framework to compare 3D navigation techniques could be built based on the proposed analysis.

Acknowledgment: The authors would like to thank Dr Géry Casiez for his help on the data analysis.

References

[BKLP05] BOWMAN D. A., KRUIJFF E., LAVIOLA J. J., POUPYREV I.: *3D User Interfaces: Theory and Practice*. Addison-Wesley Boston (MA), 2005. 1

[FWW08] FEASEL J., WHITTON M. C., WENDT J.: Llcm-wip: Low-latency, continuous-motion walking-in-place. In *IEEE Symposium on 3D User Interfaces* (2008), pp. 97 –104. 1

[PHA*07] PHAM Q., HICHEUR H., ARECHAVALETA G., LAUMOND J., BERTHOZ A.: The formation of trajectories during goal-oriented locomotion in humans. II. a maximum smoothness model. *Eu. J. of Neuroscience 26*, 8 (2007), 2391–2403. 3, 4

[RSS*02] RAZZAQUE S., SWAPP D., SLATER M., WHITTON M. C., STEED A.: Redirected walking in place. In *Proceedings of the EGVE workshop on Virtual Environments* (2002), pp. 123–130. 2

[SUS95] SLATER M., USOH M., STEED A.: Taking steps: The influence of a walking technique on presence in virtual reality. *ACM Trans. on Computer-Human Interaction 2*, 3 (1995), 201–219. 1

[TME*10] TERZIMAN L., MARCHAL M., EMILY M., MULTON F., ARNALDI B., LÉCUYER A.: Shake-your-head: Revisiting walking-in-place for desktop virtual reality. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (2010), pp. 27–34. 1, 2

[UAW*99] USOH M., ARTHUR K., WHITTON M. C., BASTOS R., STEED A., SLATER M., BROOKS JR. F. P.: Walking > walking-in-place > flying, in virtual environments. In *Proceedings of SIGGRAPH* (1999), pp. 359–364. 2

[WCF*05] WHITTON M. C., COHN J., FEASEL J., ZIMMONS P., RAZZAQUE S., POULTON S., MCLEOD B., JR F. B.: Comparing VE locomotion interfaces. In *Proceedings of IEEE Virtual Reality* (2005), pp. 123–130. 2

[WWB10] WENDT J., WHITTON M. C., BROOKS F.: Gud wip: Gait-understanding-driven walking-in-place. In *Proceedings of IEEE Virtual Reality* (2010), pp. 51 –58. 1

[ZLB*05] ZANBAKA C., LOK B., BABU S., ULINSKI A., HODGES L.: Comparison of path visualizations and cognitive measures relative to travel technique in a virtual environment. *IEEE Trans. on Visualization and Computer Graphics 11* (2005), 694–705. 2

© The Eurographics Association 2011.

Joyman: a Human-Scale Joystick for Navigating in Virtual Worlds

Maud Marchal*
INSA/INRIA Rennes, France

Julien Pettré *
INRIA Rennes, France

Anatole Lécuyer *
INRIA Rennes, France

ABSTRACT

In this paper, we propose a novel interface called Joyman, designed for immersive locomotion in virtual environments. Whereas many previous interfaces preserve or stimulate the users proprioception, the Joyman aims at preserving equilibrioception in order to improve the feeling of immersion during virtual locomotion tasks. The proposed interface is based on the metaphor of a human-scale joystick. The device has a simple mechanical design that allows a user to indicate his virtual navigation intentions by leaning accordingly. We also propose a control law inspired by the biomechanics of the human locomotion to transform the measured leaning angle into a walking direction and speed - i.e., a virtual velocity vector. A preliminary evaluation was conducted in order to evaluate the advantages and drawbacks of the proposed interface and to better draw the future expectations of such a device.

Index Terms: H.5.1 [INFORMATION INTERFACES AND PRESENTATION]: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.5.2 [INFORMATION INTERFACES AND PRESENTATION]: User Interfaces—Input devices and strategies

1 INTRODUCTION

Navigation is one of the fundamental tasks needed for 3D interaction with Virtual Environments [3]. The possibility to walk inside the virtual environment (VE) is necessary in many applications of Virtual Reality (VR) such as for virtual visits (review of architectural and urban projects), training tasks or videogames. However, it is still a challenge to allow a user endlessly walking a virtual world in an immersive manner when the physical real workspace is limited. The difference between the dimensions of the real and virtual worlds prevent from tracking users and directly transposing the walking motion into the virtual world.

In order to provide VR users with realistic sensations of walking while keeping them in their limited workspace, numerous types of VR interfaces have been proposed so far. Thus, several types of peripheral devices have been proposed. Simple manual devices such as keyboards, mices or joysticks are widely available. They allow users to indicate their locomotion wills into the virtual world but only involve the users' hands and arms whereas real locomotion involves all of the human limbs. Thus, they cannot be satisfactory used in the frame of immersive applications. Locomotion interfaces made of sophisticated treadmills and tracking systems have been designed to improve the walking sensations in VR [8, 15]. They enable users to endlessly walk while their global displacement is mechanically compensated. Such systems remain however highly expensive and cumbersome and necessarily distort the vestibular sensory feedback - i.e., equilibrioception - due to motion compensation. Finally, software solutions have been proposed to enhance the sensations of walking in VR. Some techniques [11, 12, 4] allows the user to really walk in the real workspace without reaching its limits but do not prevent break of immersion or need a big real workspace.

*e-mail: {maud.marchal,julien.pettre,anatole.lecuyer}@inria.fr

Other techniques such as the Walking-In-Place paradigm [14] preserve both the immersion and the real somatosensory feedback as users virtually walk by performing real walking motions. But the absence of real global displacement neglect equilibrioception for the benefit of proprioception.

In this paper, we explore the use of a novel interface for navigating into virtual worlds that tends to preserve equilibrioception in place of proprioception. It is often said that walking is constantly falling ahead. It is also known that head trajectory precedes the global locomotion trajectory. This justifies our work direction towards the design of a device involving the vestibular system that stands in the human head. Our second objective is to study affordable devices and propose a new interface that does not involve either active mechanical devices or complex tracking systems. The concept of our new interface, called Joyman, is illustrated in Figure 1. The proposed interface is based on the metaphor of a human-scale joystick. The device has a simple mechanical design that allows a user to indicate his virtual navigation intentions by leaning accordingly.

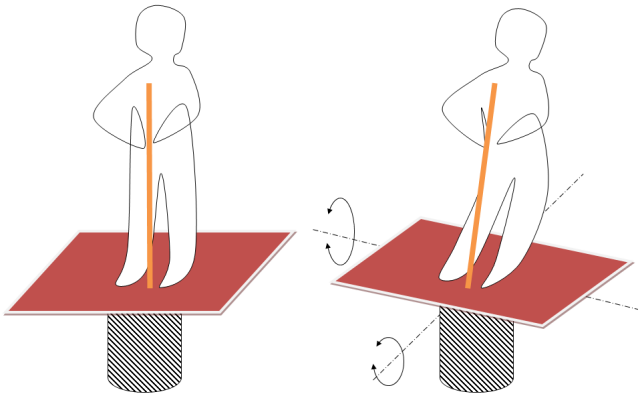


Figure 1: Principle of the Joyman interface for walking into virtual worlds. Following the example of joystick and standing on an articulated platform, users indicates their desired walking direction by leaning.

The main contributions of our paper are the following:

- **A novel and original interface involving equilibrioception to enhance immersion** in virtual environments when performing locomotion tasks. The new interface is based on a mechanical design that meets the requirements of the limited real workspace while preserving the locomotion sensations. It does not involve either active mechanical devices or complex tracking systems.

- **A combination of a new peripheral with a control law** inspired by the non-holonomic nature of the human walk. The law transforms the device orientation into locomotion velocities in order to compose a complete locomotion interface. This law express attainable turning velocities as a function of the tangential one: it has been observed that the faster the walk the lower the attainable turning speed.

The remaining of the paper is organized as follows: Section 2 provides an overview of existing interfaces and discuss their advantages and drawbacks. In Section 3, we introduce the description of our novel locomotion interface. We also present the associated control law. Finally, Section 4 details the preliminary experimental evaluation of our locomotion interface before ending with discussion and perspectives for the Joyman interface.

2 Related Work

In order to provide VR users with realistic sensations of locomotion while keeping them in their limited workspace and without any break in the immersion, numerous types of VR interfaces have been proposed so far. The most straightforward solution probably consists in walking naturally in the real world, using for instance Head Mounted Display (HMD) together with a tracking system which measures head position for updating virtual camera position and orientation in the virtual environment. However, this solution can rarely be used due to the limited physical workspace in the real world, as well as limitations of the tracking devices.

To overcome this problem, numerous "locomotion interfaces" have been designed. The locomotion interfaces keep the user within the limits of the physical workspace while allowing to navigate infinitely in the virtual world. A classical hardware approach consists in using a treadmill which can be either unidirectional (such as the Sarcos Treadport [15]) or omni-directional (such as the Torus Treadmill [8]). Other locomotion interfaces can be based on feet-platforms [7], walking inside a rotating sphere [5] or a set of moving tiles [6], standing on a Cybercarpet [13]. These locomotion interfaces have the advantage of providing realistic walking sensations, involving proprioceptive information. However, the current locomotion interfaces are often too expensive, cumbersome, or complex for being integrated easily in both VR applications.

Besides, classical input devices such as mice, keyboards and joysticks can also be used within interaction metaphors to navigate from one virtual place to another [3]. However, such interaction techniques, relying exclusively on manipulated input devices, do not provide any proprioceptive or vestibular feedback of walking and lead to a poor sensation of locomotion in the VE.

Thus, numerous interaction techniques have been developed to provide software-based navigation capabilities in VR. Software solutions have been proposed in order to keep the user inside the real workspace limits while he is really wallking, such as the World in Miniature technique [11], the Magic Barrier Tape [4] or the Freeze-Backup technique [18]. However these techniques induce breaks in the feeling of immersion during the navigation in a virtual environment. Redirected Walking Technique [12] or Motion Compression [10] preserve the immersion by forcing the user into walking in a curved path in the real world while walking in a straight line in the virtual world. However, these techniques require large workspaces, can be confusing when doing unpredictable or quick changes of direction, and may require distracting events.

Other solutions have been proposed to enhance the sensation of walking in virtual worlds without modifying the user position in the real environment. An efficient solution consists in playing with visual feedback and adding artificial camera motions [9]. The Walking-In-Place technique has been introduced by Slater et al. [14] to enable a real physical walking movement and an efficient navigation technique in 3D virtual environments. In this interaction paradigm, the user is walking in place in the real world, allowing proprioceptive feedback while keeping him in the real workspace. However, the absence of real global displacement neglects equilibrioception.

In this paper, we would like to explore the possibility of an interface that aims at preserving equilibrioception in place of proprioception. The ChairIO project [2] also implies equilibrioception but the user is seated while he is navigating in the virtual world, decreasing the immersion, especially for locomotion tasks. Our new interface is based on the metaphor of a human-scale joystick. Some devices have already been built [17] but they are not designed for enhancing immersion in VE, especially as they are not combined with a control law allowing to reproduce locomotion tasks. The main principles and the design of our new interface, called Joyman, are detailed in the next section.

3 Description of the Joyman Interface

The Joyman is a new interface to navigate into virtual worlds. One main objective of this interface is to go towards realistic locomotion trajectories in the virtual environment. The Joyman meets this objective by combining two components: a peripheral device allowing users to indicate the desired direction of locomotion and a control law which transforms the device state into a virtual velocity vector. The mechanical design of the device is based on the metaphor of a human-scale joystick and mainly consists of a board upon which the user stands. The inclination of this board can be changed by the user. Thus, the user indicates the desired locomotion direction by tilting the platform in the corresponding direction. The mechanical design of the platform allows the user to change the platform inclination by leaning and prevents him from falling, whereas repealing forces tend to maintain the platform horizontal. The control law transforms the device state (pitch and roll angles) into a virtual locomotion velocity vector. The proposed law ensures that humanly feasible velocities are achieved during virtual locomotion. Particularly, the tangential velocity is bounded and depends on the desired angular velocity. These two components are detailed below.

3.1 Device Mechanical Design

The proposed mechanical design of the interface is proposed in Figure 2.

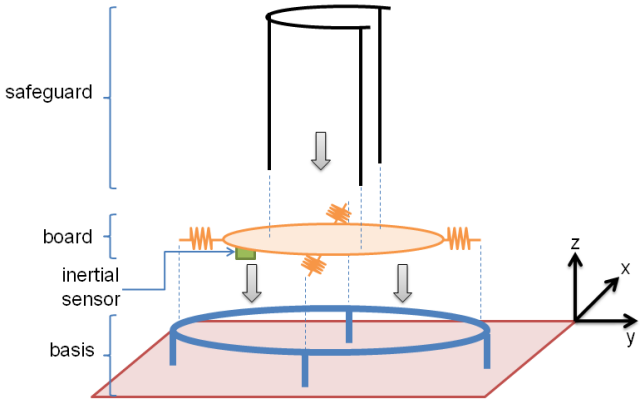


Figure 2: Schematic representation of the main elements composing our peripheral device. The board and the basis are articulated using springs. Springs generate repealing forces that tend to maintain the board in the horizontal plane, however, the board can be freely oriented around the two horizontal axis (pitch and roll) with limited range. Users act on the device by standing on it, and apply forces in order to tilt the board by leaning. The board inclination is measured using an inertial sensor. The safeguard prevents users from falling when leaning.

The key principle of the Joyman is to let the user indicate the desired direction of his virtual locomotion by leaning in the corresponding direction. In other words, our objective is to let a user control his locomotion from his equilibrioception. This main functionality stressed at first the mechanical design of the Joyman device.

The main properties of our design are :

- the user should be able to stand on the platform, which is the natural position in the task of locomotion;
- the user should feel perfectly safe when using the platform which, as a result, should prevent him from falling;
- the user should be able to lean over the limit of his own balance in order to increase the amplitude of vestibular sensations;
- the device should tend to bring the user back in vertical position to avoid highly tiring manœuvers during its use.

We also carefully considered other criteria during the device design process: the most important was to build an affordable interface. Exploring the equilibrioception in place of proprioception to control oneself locomotion in the virtual world is a promising direction to lower these costs. However, we did not want to lower the costs by degrading the feeling of immersion provided by the interface. We aim at building a new interface that does not involve any active mechanical device in its movement and does not imply the use of sophisticated and complex tracking devices. We meet the objectives listed above based on a simple mechanical device as illustrated in Figure 2.
The device is made of 4 components:

The basis supports the whole device. It is made of a flat square platform lying on the ground and of a circle shaped steel frame linked to the board by legs. The legs and the platform are fixed together and the board is large enough to prevent the device from tipping over.

The board is the element supporting users. It is circle shaped and linked to the frame by springs. As a result, the board is free relatively to the frame within limited range of motion. We can approximately but reasonably consider that the board is articulated to the frame by two degrees of freedom which are rotations around the x and y axis as illustrated in Figure 2.

The safeguard is rigidly fixed to the board and prevent users from falling.

The inertial sensor is rigidly attached to the board and acquires the current orientation of the board.

Figure 3 displays the peripheral device built in our lab following the simple mechanical design described above.

Figure 3: Illustration of the new peripheral device.

The basis is made of a square woodcut platform. The steel frame is the one of a trampoline where the springs have also been retrieved (1 m diameter). The basis is a machined plastic plate (0.55 m diameter) equipped with 18 hooks on its rim. A rigid rope is tied up between the hooks and the springs attach the board to the frame. Finally, the safeguard is made of welded iron tubes. Its height can be adapted to each user (height between 0.8 m and 1.2 m). Building costs did not exceed 500$ in spite of the fact that this device is a unique prototype excluding the inertial sensor.

Illustrations of the peripheral device in use are provided in Figure 4. The user can stand straight on the platform and has to lean in any direction to start the locomotion in the virtual world. The control law that allows to navigate in a VE with the new interface is described in the next paragraph.

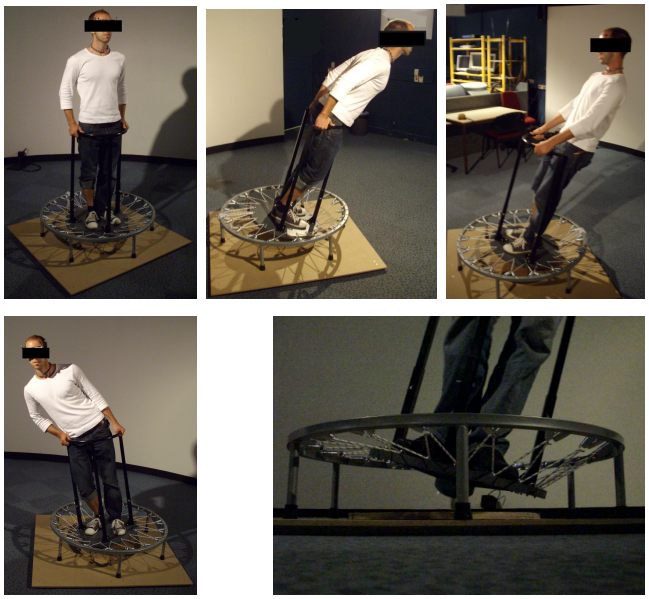


Figure 4: Illustration of the peripheral device in use. Users can stand straight on the board, repealing forces make easy keeping the platform inclination. Users have to lean to start locomotion in virtual worlds. All the possible leaning directions are displayed.

3.2 Virtual Locomotion Control

The main steps of the virtual locomotion control are summarized in Figure 5. The control is composed of two elements detailed in the following paragraph: a locomotion model and a control law.

3.2.1 Locomotion model

The locomotion model allows to translate the position and orientation of the platform into a virtual motion in the VE. We exploited experimental observations of the walking human trajectory to build our model [1].

We model the position P of the user in the virtual space as an oriented point moving in a horizontal plane (see Figure 5):

$$P = \begin{bmatrix} x \\ y \\ \theta \end{bmatrix} \quad (1)$$

The virtual motion is velocity controlled. We assume the virtual trajectory is non-holonomic, which means that the velocity vector orientation and the body orientation are always identical. The non-holonomic nature of walking human trajectory has been experimentally observed in [1].

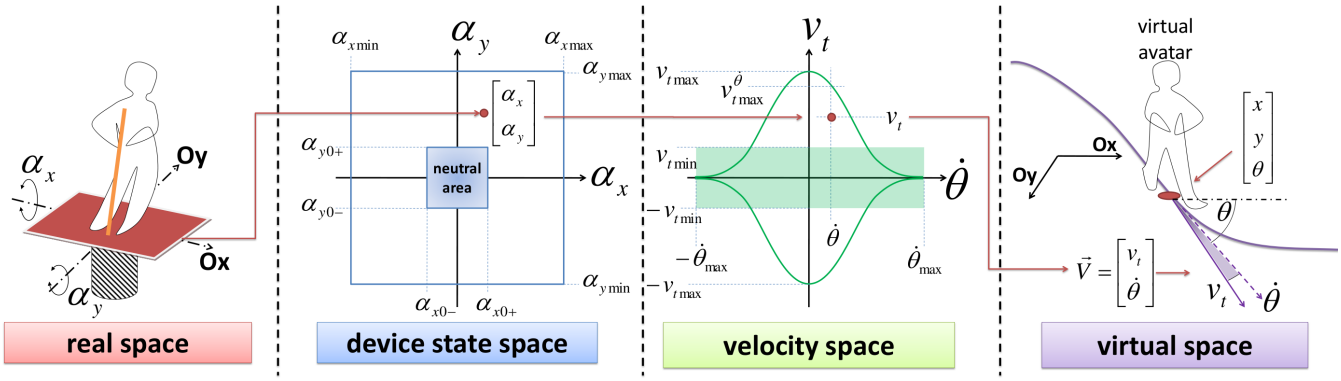


Figure 5: Summary of the different steps for the virtual locomotion control. From the position and orientation of the Joyman, we can compute the virtual velocity vector by using our locomotion model and control law.

Non-holonomy constraint allows us to decompose the velocity vector V as follows:

$$V = v_t . \begin{bmatrix} cos(\theta) \\ sin(\theta) \end{bmatrix} \tag{2}$$

Such a decomposition allows us to independently control the tangential speed v_t and the orientation θ. Human tangential velocity is limited. We denote $v_{t_{max}}$ the maximum tangential velocity bound. By default, we set this bound at $v_{t_{max}} = 1.4\ m.s^{-1}$. It has equally been observed that very slow walking velocities are never reached: such velocities are humanly feasible but not used in practice. We thus define $v_{t_{min}} = 0.6\ m.s^{-1}$. It has been however observed that changing orientation θ affects the amplitude of v_t during human locomotion [19]: humans decelerate when turning. In order to take into account such an observation, we define $v^{\dot{\theta}}_{t_{max}}$ the bound of the tangential velocity knowing the current turning velocity $\dot{\theta}$. $v^{\dot{\theta}}_{t_{max}}$ is defined as follows:

$$\begin{aligned} v^{\dot{\theta}}_{t_{max}} &= a.v_{t_{max}}.e^{-b.\frac{\dot{\theta}}{c}} \\ v^{\dot{\theta}}_{t_{min}} &= a.v_{t_{min}}.e^{-b.\frac{\dot{\theta}}{c}} \end{aligned} \tag{3}$$

where a, b and c are parameters. As a result, we model the reachable tangential velocity $v^{\dot{\theta}}_{t_{max}}$ as a Gaussian function of the current turning velocity $\dot{\theta}$. The higher $\dot{\theta}$, the lower the tangential velocity bound. By default, we arbitrarily choose: $a = 1.07$, $b = 0.5$ and $c = 0.7$. Such values match the experimental observations provided in [19]. Finally, the absolute value of the angular velocity is also bounded to $\dot{\theta}_{max}$. We arbitrarily choose $\dot{\theta}_{max} = 1\ rad.s^{-1}$.

3.2.2 Control Law

The control law allows users to modify the virtual velocity vector V by standing on the device and leaning. The modification of the platform orientation affects the state of the device s_d. s_d is defined by the orientation of the board α_x and α_y relatively to the two horizontal axis $\vec{O}x$ and $\vec{O}y$, as measured by the inertial sensor:

$$s_d = \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} \tag{4}$$

We neglect the orientation around the Oz axis. During calibration stage, we ask the user to stand on the platform and to successively firmly lean towards all the cardinal directions. We estimate the reachable bounds of the board orientation by averaging the reached orientation over a short period of time, $\alpha_{x_{min}}$, $\alpha_{x_{max}}$, $\alpha_{y_{min}}$ and $\alpha_{y_{max}}$.

We also ask the user to stand straight on the platform and define a neutral area bounded by $\alpha_{x_{0+}}$, $\alpha_{x_{0-}}$, $\alpha_{y_{0+}}$ and $\alpha_{y_{0-}}$.

We want the user to control his tangential velocity v_t by leaning in forward or backward direction, i.e., by playing on α_x, whilst angular velocity $\dot{\theta}$ is controlled by leaning on the sides, i.e., by playing on the α_y value. The angular velocity is controlled as follows:

$$\begin{cases} \dot{\theta} = \dot{\theta}_{max}.\frac{\alpha_{y_{max}} - \alpha_y}{\alpha_{y_{max}} - \alpha_{y_{0+}}} & \text{if } \alpha_y > \alpha_{y_{0+}} \\ \dot{\theta} = \dot{\theta}_{max}.\frac{\alpha_y - \alpha_{y_{min}}}{\alpha_{y_{0-}} - \alpha_{y_{min}}} & \text{if } \alpha_y < \alpha_{y_{0-}} \\ \dot{\theta} = 0 & \text{otherwise} \end{cases} \tag{5}$$

which allows to deduce $v^{\dot{\theta}}_{t_{max}}$ and $v^{\dot{\theta}}_{t_{min}}$ according to equation 3 and finally v_t:

$$\begin{cases} v_t = v^{\dot{\theta}}_{t_{min}} + (v^{\dot{\theta}}_{t_{max}} - v^{\dot{\theta}}_{t_{min}}).\frac{\alpha_{x_{max}} - \alpha_x}{\alpha_{x_{max}} - \alpha_{x_{0+}}} & \text{if } \alpha_x > \alpha_{x_{0+}} \\ v_t = -v^{\dot{\theta}}_{t_{min}} + (v^{\dot{\theta}}_{t_{max}} - v^{\dot{\theta}}_{t_{min}}).\frac{\alpha_x - \alpha_{x_{min}}}{\alpha_{x_{0-}} - \alpha_{x_{min}}} & \text{if } \alpha_x < \alpha_{x_{0-}} \\ v_t = 0 & \text{otherwise} \end{cases} \tag{6}$$

At each time step n, the virtual velocity vector $V_n = [v_{t_n}, \dot{\theta}_n]$ is thus deduced from s_d, the state of the device. Before updating the simulation accordingly, we check that no unbelievable acceleration is performed. Thus, given the previous velocity vector V_{n-1}, we finally compute the effective velocity vector V_f:

$$V_f = V_{n-1} + \lfloor \frac{V_n - V_{n-1}}{\Delta t} \rceil \tag{7}$$

where Δt the simulation time step, and $\lfloor\rceil$ a function that truncates the velocity vector variation so that the absolute tangential acceleration does not exceed $1\ m.s^{-2}$ and the angular one does not exceed $1\ rad.s^{-2}$.

3.3 Summary of Joyman Interface

The Joyman interface is designed for immersive virtual locomotion into virtual worlds. Our main aim was to explore the possibility of an interface that tends to preserve the equilibrioception in place of proprioception, in contrary to many current interfaces. The mechanical design of our platform is relatively simple. We showed how the proposed device allows a user to safely perform exaggerated leaning motion, over the limit of balance, in order to indicate his virtual navigation wills. We here remind our long-term objective to be able to perform realistic locomotion trajectories into the virtual world. What can we expect from the Joyman interface?

Immersion The device does not limit technical choices about visual or audio feedback. Typically, the device can be directly used in any immersive environment. More important, the initiation of the locomotion is made by leaning ahead. This corresponds to the real motion to initiate walking, which is considered to be a constant loose of balance and succession of falling ahead. We expect this interaction mode, where vestibular sensory system stimulation is preserved, to be greatly immersive.

Realism We expect that the dynamics of the required manœuvers to operate the Joyman are close to the ones of real walking motions. As an example, using the Joyman, users perform a left turn following a right turn by changing the platform orientation toward its opposite using their whole body. Involved inertias and frictions prevent from performing this change of state immediately: this reproduces the walking behavior during which the inclination of the body is naturally adapted to face centrifuge force. We clearly do not expect users to achieve efficient navigation (in terms of task completion time) using the Joyman.

4 Preliminary Evaluation

As a preliminary evaluation, we propose to compare our interface to the joystick, a classical peripheral often used in VR applications. The joystick can be considered as one of the most performant interface to achieve efficient navigation in terms of task completion time. Thus, the first objective of our evaluation was to quantify the loss of performances of our interface compared to the joystick (our interface involves namely the whole body compared to the joystick where only hands and arms are used). The experiments were conducted using 3D VE displayed either on a screen or on a Head-Mounted Display (HMD). We investigated the effectiveness of our interface to travel complex paths composed of different gates placed in the VE. The second objective of our evaluation was to verify if we can obtain a more immersive navigation with our interface compared to classical peripherals already used in VE, ie. the joystick. Thus, a subjective questionnaire was proposed to the participants to evaluate their subjective preferences in terms of quality of the VE navigation.

4.1 Method

4.1.1 Virtual Environment

The evaluation was performed within a 3D virtual environment without any contextual cues. The only landmarks were the gates that the user had to navigate through. A fog effect was added to mask the distant gates, allowing to perceive only the 2 or 3 closest gates. A texture on the ground provided useful visual flow information during the navigation. The scene was normally illuminated, and no shadows were drawn as illustrated in Figure 6.

4.1.2 Sample Population

Sixteen participants aged from 19 to 28 (Mean= 23.4, SD=2.4) performed the experiment. One participant was left-handed and none of them had known perception disorders. All participants were naïve with respect to the proposed interface, as well as to the experimental setup and purpose of the experiment.

4.1.3 Experimental Conditions

The experiment was carried out using two visual conditions: HMD or a screen. With these two different visual conditions, we would like to evaluate if the display manner (fixed screen or screen moving with the body for the HMD) has an influence on the user sensations. The two visual conditions and associated experimental configurations are illustrated in Figure 7.

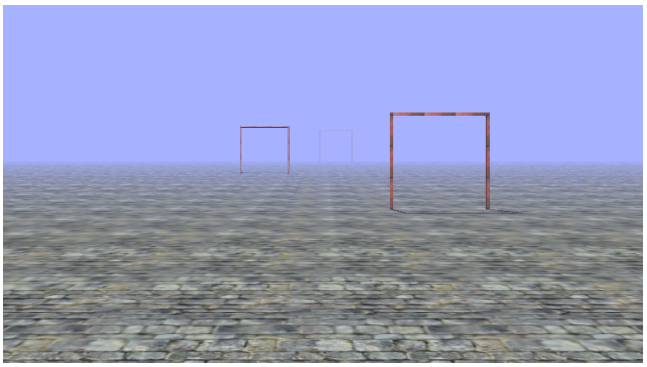
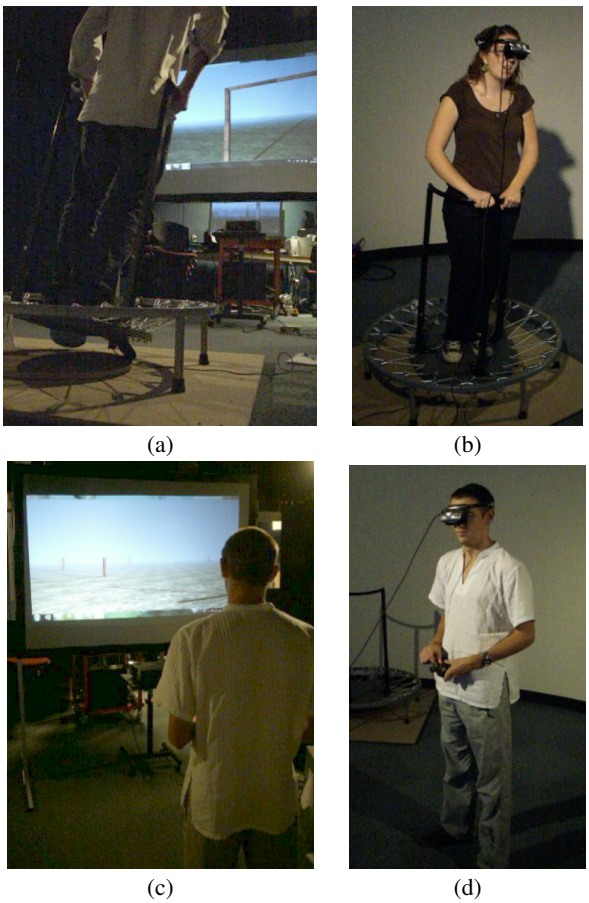


Figure 6: Snapshot of the virtual environment used during the evaluation. The different gates represent the path that the user had to navigate through.



(a) (b) (c) (d)

Figure 7: Description of the experimental setup for the 4 configurations: the virtual scene was visually perceived through HMD (b and d) or through a projection screen (a and b); locomotion was controlled using the Joyman (a and b) or a joystick (c and d).

For the screen condition, the luminosity was controlled in the room using two projectors. The participants were at a distance of 2m in front of a $1.72m$ large and $1.24m$ height back-projected screen (physical field of view of 60 degrees horizontally and 45 degrees vertically). The resulting image had a resolution of 1600×1200 pixels. We used monoscopic rendering, with a frame rate at $60Hz$. The projector used was a DepthQ Stereoscopic. For the HMD condition, we used the eMagin Z800 Head Mounted Display, at 60 Hz (physical field of view of 40 degrees diagonally). The video image was generated without using any head tracking sensor. The participants were standing either on the Joyman or on the ground to complete the task.

4.1.4 Experimental Apparatus

In our experiment, the goal was to compare our interface to the joystick, a classical interface device widely used in VR applications. In addition, we used two different visual conditions (HMD and a screen). During the experiment, the participants had to navigate in the VE through 2 different paths composed of 8 gates each. The gates were 3 meters by 3 meters wide and were disposed in order to form a slalom.

The participants were able to choose at the beginning of each block of conditions when they wanted to launch the experiment. At any time during the experiment, they had the possibility to make a break. After each trial, the participants were automatically teleported to the beginning of the next trial. A black screen displayed 2s notified the beginning of the new trial.

4.1.5 Experimental Plan

The participants were exposed to 4 blocks of 8 trials each: one block for each of the possible combinations of the experimental conditions. The combinations were the following: (1) Joyman, using a HMD; (2) Joyman, using the screen; (3) Joystick, using a HMD; (4) Joystick, using the screen.

The participants were split in 4 groups equally composed of 4 people each. Two groups started with the Joyman configuration and the 2 others with the Joystick configuration. We counter-balanced the conditions between them, meaning that the group starting with the HMD condition in the Joyman configuration starts with the screen condition in the Joystick configuration. For each of the 8 trials of a block, the participants had to navigate through the 2 paths (4 for each, in a random order).

4.2 Results

4.2.1 Collected Data

In our evaluation, we measured for each participant the task completion for each trial and the percentage of success for the different gates. A subjective questionnaire was also proposed to evaluate 8 criteria.

4.2.2 Performances: Task Completion Time and Accuracy

For each participant, the task completion time of each trial was measured for the different experimental conditions. A repeated two-way ANOVA was performed on the two different interfaces and the two visual conditions. The ANOVA accounting for the visual conditions and the task completion time revealed a significant effect ($F(1,526) = 11.63$, p-value< 0.001). A significant effect was also found for the interfaces and the task completion time ($F(1,526) = 54.47$, p-value< 0.001). As expected, the results reveal that the joystick was better than the Joyman in terms of speed of the navigation. The mean value for the completion time of a path was $187s$ (standard deviation=$13s$) for the joystick and $321s$ (standard deviation=$89s$) for the Joyman.

A specific analysis was developed to study the learning effect of the two configurations (Joyman and Joystick). A linear model where all conditions are mixed was fitted to explain the relation between the task completion times and the trial number. For the Joyman configuration, it revealed that the slope of the linear regression was significantly lower than zero (p-value < 0.000001), reflecting a significant decrease in the task completion time as the number of trials increases. The same analysis, where the first trial was removed, showed that the slope was not significantly different from zero anymore (for both visual conditions). Figure 8 illustrates the task completion time of the different trials for the Joyman configuration.

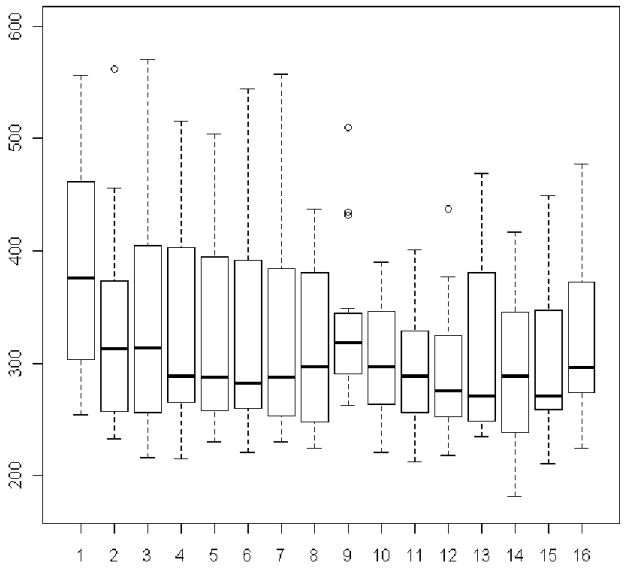


Figure 8: Task completion time (in seconds) for the Joyman configuration (16 trials). The first trial was left to illustrate the learning effect. Each boxplot is delimited by the quartile (25% quantile and 75% quantile) of the distribution of the condition over the individuals. The median is also represented for each trial.

For each participant and for each trial, the percentage of errors for the different paths was measured. There was no error at the end, for both configurations (Joyman and Joystick) and both visual conditions.

4.2.3 Subjective Questionnaire

A preference questionnaire was proposed in which participants had to grade from 1 (low appreciation) to 7 (high appreciation) the two configurations (Joyman and Joystick) according to 8 subjective criteria : (a) Fun, (b) Intuitive, (c) Acccuracy, (d) Presence, (e) Rotation realism, (f) Fatigue, (g) Cybersickness and (i) Global appreciation. The grade 7 for Fatigue and Cybersickness respectively means that the interface does not induce any fatigue and does not imply any cybersickness feeling. Figure 9 shows the results concerning the grades (Likert-scale) obtained by the two different interfaces for each of the subjective criteria.

After performing an ANOVA on the two different conditions, we found a significant effect for 6 criteria: Fun ($F(1,30) = 17.77$, $p < 0.001$), Intuitive ($F(1,30) = 21.25$, p-value< 0.001), Accuracy ($F(1,30) = 23.52$, p-value< 0.001), Presence ($F(1,30) = 13.35$, p-value< 0.001), Rotation realism ($F(1,30) = 6.63$, p-value$= 0.015$) and Fatigue ($F(1,30) = 87.51$, p-value< 0.001). In particular, our new platform was better ranked for Fun, Presence and Rotation realism. No significant effect was found for Cybersickness ($F(1,30) = 4.01$, p-value=0.054) and Global appreciation ($F(1,30) = 0.69$, p-value$= 0.411$).

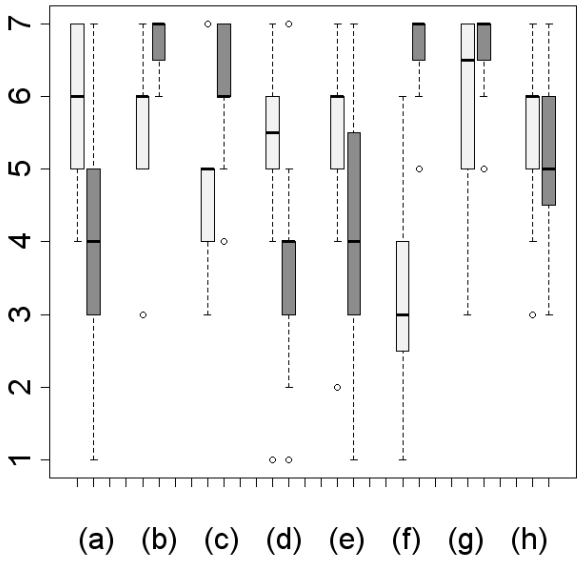


Figure 9: Results for the subjective questionnaire for the two different configurations (Joyman and Joystick) with respect to a Likert-scale grading. The light color corresponds to the Joyman condition and the dark color corresponds to the joystick condition. The subjective criteria are (a) Fun, (b) Intuitive, (c) Accuracy, (d) Presence, (e) Rotation realism, (f) Fatigue, (g) Cybersickness (h) Global appreciation. Each boxplot is delimited by the quartile (25% quantile and 75% quantile) of the distribution of the condition over the individuals. The median is also represented for each condition.

5 Discussion and Perspectives

As a global conclusion of the preliminary evaluation of the Joyman, we can state that the feeling of immersion in the virtual world is significantly improved - in comparison with traditional joystick-based techniques - at the cost of some easiness of use. The Joyman is still at a early stage of development, however, first results are promising and open a large set of possible directions to improve usability and the level of realism of virtual navigation. This section discuss as exhaustively as possible these directions as well as future work to meet our objectives: an easy, intuitive, immersive interface allowing realistic locomotion in virtual worlds.

Interface Calibration The navigation with the Joyman implies the whole body and, as expected, the task completion times were higher for our interface compared to the joystick where only the arms and hands are involved. The use of the whole body, implying more movements, is also a reason for the lower rating for the Accuracy criterion in the subjective questionnaire. However, the results concerning the evaluation of our interface in terms of performances are encouraging as the participants always succeeded to complete the navigation task during the evaluation. One way of improvement for the performances of our interface could be the interface calibration. The control law directly transforms the angle of the interface platform into virtual walking velocities. 3 types of parameters were proposed to design our control law: those define the active angles area into the device state space ($\alpha_{x_{min}}$, $\alpha_{x_{max}}$, $\alpha_{y_{min}}$ and $\alpha_{y_{max}}$, $\alpha_{x_{0+}}$, $\alpha_{x_{0-}}$, $\alpha_{y_{0+}}$ and $\alpha_{y_{0-}}$), those controlling the reachable virtual velocities ($\dot{\theta}_{max}$, $v_{t_{min}}$ and $v_{t_{max}}$) and finally those controlling the dynamics of the relation between tangential and angular velocities (a, b, and c, see Equation 3). Ideally, the first type of parameters should be calibrated for each user. Indeed, given his size, weight and strength, each user provides relatively different efforts to reach a same given platform inclination angle (i.e., a given walking velocity). The slight fatigue reported by participants in the results of our experiments confirms the need for individual calibration. Currently, calibration process consists in recording leaning motions as well as neutral positions to define the bounds of the effective zone for the platform. We made such a calibration before experiments, but kept a unique one all along our evaluation in order to have the same behavior of our interface for all participants. For VR applications, we could envisage to explore various calibration techniques as well, such as by displaying a virtual environment with a moving point-of-view and to ask users to apply the effort on the platform they believe to be corresponding to the motion.

Mechanical design We presented in this paper the first prototype of our new interface. The results of the subjective questionaire suggest that the participants enjoyed the navigation with the Joyman. They namely gave a higher rating for the Fun criterion but also for Presence criterion and Rotation Realism criterion. Specifically, the higher rate given to this last criterion confirms that our mechanical design was appreciated for locomotion tasks in the virtual environment. A way of improvement for our mechanical design could concern the linkage between the platform and the basis. The proposed prototype implements this linkage using an inextensible rope and a set of springs (see Figure 4) as we wanted to have a simple mechanical design. However, we could envisage to modify the linkage to increase the possibility of movements of our interface. Ideally (mechanically speaking), the linkage could be a 2 rotational degrees of freedom, one with a restoring force proportional to the platform inclination.

Control Law Experimental evaluation reveals that the proposed control law is intuitively grasped by users: establishing a relation between linear and angular velocities seems to be naturally accepted by users, and is consistent with observations of human locomotion trajectories. Future work will deal with potential modifications of the control law, for example by using a law that can lower vestibular and visual sensory conflicts.

The device state vector is currently two dimensional: the two angles that describe the platform inclination. With some experience, it appears that the device can be fully and accurately controlled by involving the lower body only, it is not even required to hang on to the barrier (except for moving backward in the current state of the device). Such a property opens interesting perspectives and makes possible to increase the dimension of the device state space. Possible extensions are numerous, and immediately within reach if the VR system is equipped with tracking abilities: hands remain free to achieve secondary actions in the virtual worlds (grasping, touching, pointing tasks, etc.), view direction and locomotion control can be decomposed, etc. Nevertheless, we observed during experiments that most of the participants intuitively attempts to control their locomotion also by moving their upper body as illustrated in Figure 10, in spite of the inefficiency of such motions to significantly increase the inclination of the platform. We however could use this input (i.e., the orientation of the spine relatively to the hips) to control navigation based on an holonomic locomotion model: as opposed to the non-holonomic one, lateral velocities are allowed in addition to tangential and angular ones by removing the constraint imposed by equation 2 (lateral velocities here correspond to sidesteps).

Immersion and sensory feedback Evaluation showed that the Joyman provides a satisfying level of immersion into the virtual world. In addition, the Presence criterion was better ranked for our interface, compared to the joystick. As futur work, we could envisage to improve the level of immersion by adding other interaction techniques. Thus, the visual perception of motion could be improved by adding camera motions like in [16] for example. Thus,

Figure 10: 4 Snapshots extracted from a video sequence acquired during evaluation experiments. We could observe that participants intuitively attempt to use their upper body to control their virtual locomotion in spite of the inefficiency of such actions (note that the platform inclination is almost unchanged from picture to picture). However, this observation provides interesting direction to improve virtual motion controllability, and eventually control side steps from such inputs.

having oscillating view point could reinforce the accuracy of the perception of the traveled distance by reproducing the natural oscillations of the head during human locomotion. This could also lower the feeling that locomotion is too slow as reported by participants in the subjective questionnaire.

6 Conclusion

In this paper, we presented a novel interface to control locomotion into a virtual world while remaining globally static in the real space. The Joyman interface is composed of a new peripheral device and a dedicated control law which transforms the device state into a virtual locomotion velocity vector. Our main contributions are (1) to explore the users ability to exploit equilibrioception to control their virtual locomotion and (2) to maintain a high level of immersion compared to handheld devices (e.g., joysticks).

After a preliminary evaluation, we obtained promising results as the users enjoyed the navigation with our new interface. Presence and realism in the virtual rotations were also underlined. The evaluation also opened several future work directions to improve and extend our interface. Various VR but also real applications could be envisaged, when navigating in a 3D world. Our interface could be used for example for videogames, rehabilitation, training tasks or virtual visits.

A patent for the interface presented in this paper has been filed on November 19, 2010 under the number FR10/595551.

References

[1] G. Arechavaleta, J.-P. Laumond, H. Hicheur, and A. Berthoz. The nonholonomic nature of human locomotion: a modeling study. In *Proceedings of IEEE International Conference on Biomedical Robotics and Biomechatronics*, pages 158–163, 2006.

[2] S. Beckhaus, K. Blom, and M. Haringer. *ChairIO - the Chair-Based Interface*. Ed. Magerkurth und Rötzler, Shaker Verlag, 2007.

[3] D. Bowman, E. Kruijff, J. LaViola, and I. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison-Wesley Boston (MA), 2005.

[4] G. Cirio, M. Marchal, T. Regia-Corte, and A. Lécuyer. The magic barrier tape: a novel metaphor for infinite navigation in virtual worlds with a restricted walking workspace. In *Proceedings of the ACM Symposium on VRST*, pages 155–162, 2009.

[5] K. Fernandes, V. Raja, and J. Eyre. Cybersphere: the fully immersive spherical projection system. *Communications of the ACM*, 46(9):146, 2003.

[6] H. Iwata, H. Yano, H. Fukushima, and H. Noma. Circulafloor: A locomotion interface using circulation of movable tiles. In *Proceedings of IEEE Virtual Reality Conference*, pages 223–230, 2005.

[7] H. Iwata, H. Yano, and F. Nakaizumi. Gait master: A versatile locomotion interface for uneven virtual terrain. In *Proceedings of IEEE Virtual Reality Conference*, pages 131–137, 2001.

[8] H. Iwata and Y. Yoshida. Path reproduction tests using a torus treadmill. *Presence: Teleoperators and Virtual Environments*, 8(6):587–597, 1999.

[9] A. Lécuyer, J.-M. Burkhardt, J.-M. Henaff, and S. Donikian. Camera motions improve the sensation of walking in virtual environments. In *Proceedings of IEEE Virtual Reality Conference*, pages 11–18, 2006.

[10] N. Nitzsche, U. Hanebeck, and G. Schmidt. Motion compression for telepresent walking in large target environments. *Presence: Teleoperators and Virtual Environments*, 13(1):44–60, 2004.

[11] R. Pausch, T. Burnette, D. Brockway, and M. E. Weiblen. Navigation and locomotion in virtual worlds via flight into hand-held miniatures. In *Proceedings of SIGGRAPH'95*, pages 399–400, 1995.

[12] S. Razzaque, Z. Kohn, and M. C. Whitton. Redirected walking. In *Proceedings of Eurographics*, 2001.

[13] M. Schwaiger, T. Thummel, and H. Ulbricht. A 2d-motion platform: The cybercarpet. In *Proceedings of Second Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 415–420, 2007.

[14] M. Slater, M. Usoh, and A. Steed. Taking steps: The influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction*, 2(3):201–219, 1995.

[15] K. Stanney and al. *Handbook of Virtual Environments: Design, Implementation, and Applications*. Lawrence Erlbaum Associates, 2002.

[16] L. Terziman, A. Lécuyer, S. Hillaire, and J. M. Wiener. Can camera motions improve the perception of traveled distance in virtual environments? In *Proceedings of IEEE Virtual Reality Conference*, pages 131–134, 2009.

[17] J. Torchinsky. Giant atari 2600 joystick. `http://www.vgg.com/jason/jdt/projects.html`, 2006. [Online; accessed 6-January-2011].

[18] B. Williams, G. Narasimham, B. Rump, T. McNamara, T. Carr, J. Rieser, and B. Bodenheimer. Exploring large virtual environments with an hmd when physical space is limited. In *Proceedings of the ACM Symposium on Applied Perception in Graphics and Visualization*, pages 41–48, 2007.

[19] Y. Zhang, J. Pettré, Q. Peng, and S. Donikian. Data based steering of virtual human using a velocity-space approach. In *Proceedings of Motion in Games*, Lecture Notes in Computer Science 5884, pages 170–181. Springer, 2009.

MODELING OF LOWER LIMBS FOR VIBROTACTILE COMPENSATION

Martin J.-D. Otis, Guillaume Millet, Stephane Beniak, and Jeremy R. Cooperstock
McGill University, Montreal, Canada

SUMMARY

A rehabilitation device with a vibrotactile actuator is designed for analysis of slipping behaviour on different surfaces. However, the coupling between a subject's lower limbs and the device reduces the accuracy of the ground simulation. A real-time algorithm is developed, based on an FFT filter-bank and a rheological limb model, which compensates for lower limb effects on vibrotactile signals.

INTRODUCTION

The aging of our population motivates the study and analysis of lower leg slipping behaviour on different surfaces, characterized by their coefficients of friction. Our specific interest involves experimental studies to understand how vibration influences human perception of friction on different surfaces. To facilitate such work, we developed a device, illustrated in Fig. 1, that includes a vibrotactile actuator to generate the vibration felt under the foot when walking on various materials, such as ice, snow or sand. The design involves tradeoffs in terms of the simulation of friction and fidelity of the subject's experience.

Of particular importance for perception studies, a mechanically transparent device is required to maximize the accuracy of the synthesized ground properties. However, coupling between user impedance of the lower limbs and the dynamics of the mechanical device can interfere with its transparency, thereby reducing the accuracy of the simulation. Transparency in haptics is defined by the capacity to cover the range of our perception, by producing vibration of high fidelity over a wide bandwidth of frequencies. To compensate for user effects on vibrotactile signals, we present a real-time algorithm based on a polyphase network implementation of an FTT analysis-synthesis filter-bank and rheological modeling of lower limbs.

A previous study demonstrated that we could compensate for the structural properties of the tile, obtaining a flat response in the desired frequency bandwidth [1]. A 14-order FIR filter was designed for this static compensation. By inverting this filter, and applying it to all subsequent outputs of the actuator, we obtain a significantly corrected frequency response. In order to improve the results obtained with this filter and avoid issues of latency and stability, it is also possible to use a uniform and warped low delay filter-bank equalizer such as in speech enhancement applications [2]. However, this static compensation was found to be insufficient, prompting our research to explore the alternative of a dynamic compensation system, which extends the usable frequency range of the vibrotactile signal used for simulating ground properties from 40 to 1024 Hz.

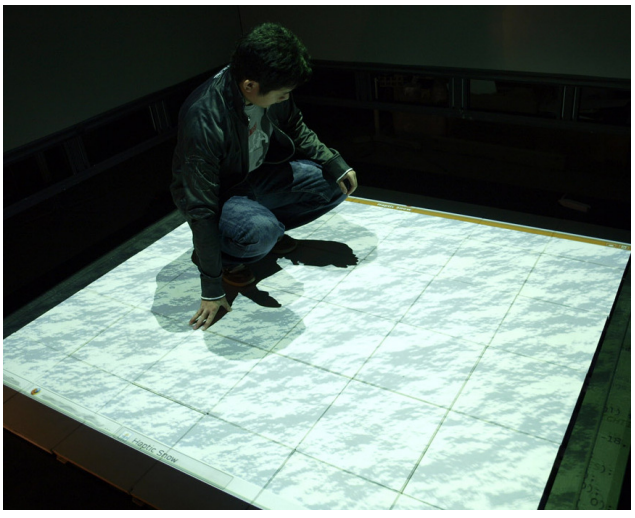


Figure 1: Snow surface simulation in our virtual environment.

MECHANICAL DESIGN

The mechanical design has been presented in previous work [1]. Our initial prototype, which adds a variable-friction mechanism to enable the reduction of the coefficient of friction, is shown in Fig. 2. A Clark Synthesis vibrotactile actuator (model TST429-Platinum) and an accelerometer (model ADXL 305) are mounted under the tile. Signal generation and acquisition is performed with a National Instruments USB NI-6218. A SURE Electronics audio amplifier is used to drive the actuator. Four force sensors are located at each corner of the tile under an elastomer to monitor the user action. All the mechanical parts, which are used for the simulation of coefficient of friction absorb spectral energy different from that of the tile, thus justifying the design of a compensation algorithm.

LOWER LIMBS MODELING

A polyphase network implementation of an FTT analysis-synthesis filter-bank [2] is first used to equalize the response of the unloaded device in a static configuration. A second-order model of the skin and the leg Z_c is used for the lower limbs, and is placed in cascade with the filter bank as described in Fig. 3. This model was used for predicting the kinematics and dynamics of a walker in [3]. Also, this model comes from two studies on the leg and skin parameters given respectively in [4] and [5]. The studies described the skin and the leg with two second order systems. For simplicity, the upper leg is modeled with a first order low-pass filter and the skin, along with the shoe is simplified with a resonant second order high pass filter. This model, described by the following equations, and illustrated in Fig. 2, reduces the number of finger parameters compared to the Kern model (in grasping application):

$$H_l = \frac{k_3}{k_3 + b_3\, s}, \tag{1}$$

$$H_s = \frac{g_2\, s^2}{\frac{k_2}{m_2} + s^2 + \frac{b_2\, s}{m_2}}, \tag{2}$$

$$H_u = H_l + H_s \rightarrow Z_c = H_u^{-1}. \tag{3}$$

In the above, g_2 comes from the sensor gains and is fixed at 0.005. The three values, b_3/k_3, k_2/m_2 and b_2/m_2, must be evaluated within an optimization procedure. It can be noted that the elastomer suspension element under the tile acts like a notch filter with a resonant frequency near 20 Hz, which is very near the characteristic of the leg. This elastomer and the leg are then simplified in one low-pass model with the parameter b_3/k_3.

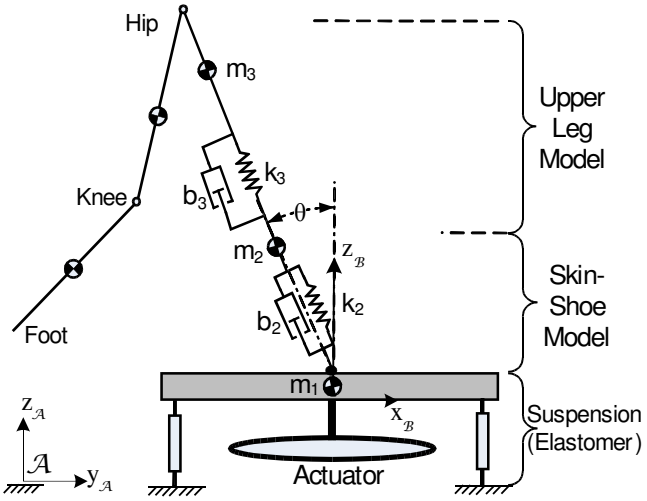


Figure 2: Rheologic modeling of the user and the device

ALGORITHM IMPLEMENTATION

The filter bank adaptively adjusts the gain with weighted spectrum coefficients for each sub-band filter as a function of the static configuration of the tile. The second component dynamically corrects the frequency response of the moving user with the rheological model as shown in Fig. 3. The main challenge here is to simplify the computations involved in order to minimize hardware costs while maintaining a reasonable level of fidelity. To this end, we can do away entirely with the real-time correction algorithm and instead provide a preset equalizer for any specific user based on his or her precomputed parameters. Thus, by studying a variety of individuals and constructing a representative database of mappings from parameters to corresponding correction equalizers, the entire correction system can function via a simple lookup table, interpolating any parameter sets in between those from known users [6].

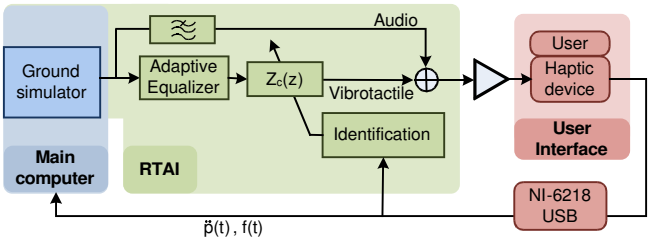


Figure 3: Vibrotactile compensation algorithm

RESULTS

A sweep frequency response is achieved over the bandwidth range of 40 to 1024 Hz, which corresponds to the maximal vibrotactile response of the skin [5]. Under 40 Hz, the device does not exhibit a significant response. The accelerometer provides the magnitude $\mathbf{A}$ of the vibrations. The unloaded response gives the spectral weighting $\mathbf{w}$ for the equalizer, which is computed iteratively with equations (4) and (5):

$$\mathbf{a} = \mathbf{A} - \bar{\mathbf{A}} \quad (4)$$

$$\mathbf{w}_k = \mathbf{w}_{k-1}\left(\frac{\bar{\mathbf{a}}}{\mathbf{a}}\right) \quad (5)$$

where $\bar{\mathbf{A}}$ and $\bar{\mathbf{a}}$ are the mean values. Fig. 4 shows these frequency responses for the unloaded device with the spectral weighting applied by the filter bank equalizer. Two or three frequency sweeps are required to obtain a reasonably accurate, flat frequency response over the desired bandwidth since moderate intensity of vibration threshold ranges from 0.45 to 3 dB [7]. For each pass, the deviation from a flat response decreases from 15.3, 0.90 and 0.28 dB, respectively, with the final error likely due in large part to accelerometer noise.

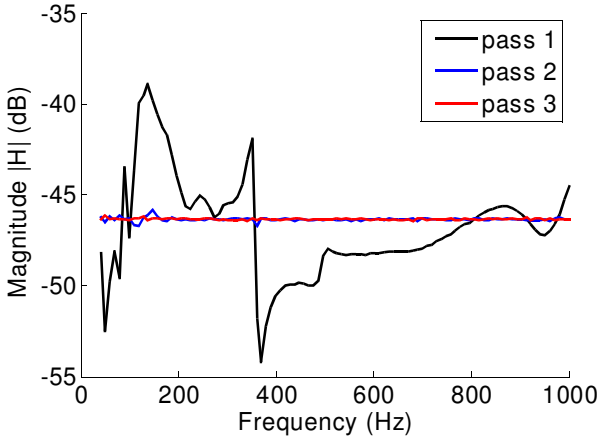


Figure 4: Frequency sweep of the unloaded device with the filter bank equalizer

Ten participants, between 20 and 35 years of age, were then asked to stand on the device with two different types of shoes: compliant and rigid sole. As seen in Fig. 5, the effect of the type of sole is important. The soft sole allows the tile to vibrate more freely and can cause it to enter resonance in mid-range frequencies, between 400 and 600 Hz. In contrast, the rigid sole constrains the tile movement over a wider bandwidth as the weight of the user increases.

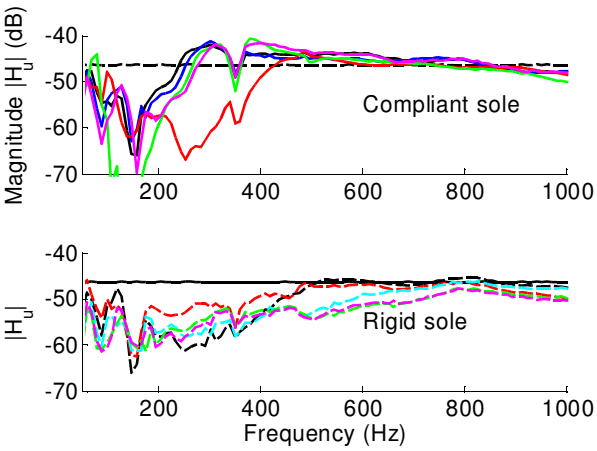


Figure 5: Frequency sweeps for different participants wearing soft-soled (compliant) shoes (upper graph) and rigid soles (lower graph).

The rheological model presented in Equation (3) is then fitted with a trust-region-reflective algorithm for each magnitude of the frequency response, for which the results are shown in Fig. 6. Upper and lower bonds on each parameters (b_3/k_3 , k_2/m_2 and b_2/m_2) are used to avoid divergence or meaningless solutions. Table 1 shows the mean and standard deviation for the three parameters for the participants. It should be noted that the cutoff frequency of the leg is lower than 20 Hz and our device can only perform adequately over 40 Hz. Also, the noise below 200 Hz is a characteristic of the actuator coupled with the mechanical components and is also observed in the unloaded response of Fig. 4. This behaviour cannot be compensated for with a non-adaptive filter-bank equalizer. Nevertheless, this first experiment shows that the model can be used to compensate for the effect of the limbs, whatever the respective sole, on the spectral magnitude.

The final step is to utilize the measured force applied by the user and the accelerometer data to implement an identification algorithm that compensates for the model in real-time. Similar online estimation of human impedance was previously investigated for human-robot interaction stability for movement of the leg [8] and the vibrotactile signal applied to fingers during grasping [6,9]. In this manner, the transfer function of the limb can be inverted and used as described in Fig. 3.

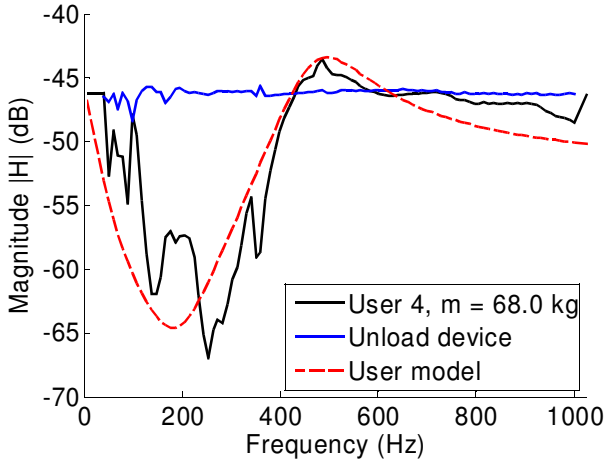


Figure 6: Frequency analysis with the user model

Table 1: Model parameters

Parameter	***Mean***	***Standard deviation***
b_3/k_3	0.0037	0.0016
k_2/m_2	5.8x10^6	4.1x10^6
b_2/m_2	2402	1.5x10^3

CONCLUSION AND FUTURE WORKS

We presented a simple compensation algorithm for haptic transparency of our vibrotactile device, which synthesizes vibrations corresponding to different types of grounds with variable friction properties. Using a filter-bank equalizer and a rheological model allows for compensation of the effects of the device, the lower limb, and the shoe.

The table-based dynamic compensation presented here is only an approximation. More elaborate techniques, such as the filter-bank adaptive equalizer, exploiting short-time DFT/IDFT could be used to reduce the error of the model. Our future work is concerned with a strategy for adaptively changing the spectral weights in real time as the user walks on the device. Further investigation must be conducted on the dynamic effect of the mechanism itself. At present, the position of the contact point changes the apparent stiffness and damping at this point modifies the frequency response of the tile. In addition, acceleration must be evaluated at the contact point since the movement of the tile differs depending on where the force constraint is applied.

ACKNOWLEDGEMENTS

This research was supported by a grant from the MDEIE of Quebec for the European 7th Framework Programme project Natural Interactive Walking (NIW, no. 222107). The authors gratefully acknowledge this support and thank Yon Visell for useful discussions.

REFERENCES

[1] Y. Visell and J. R. Cooperstock, "Design of a Vibrotactile Display via a Rigid Surface," *Haptics Symposium*, IEEE, pp. 133-140, March 2010.

[2] Heinrich W. Lollmann, and Peter Vary, "Uniform and warped low delay filter-banks for speech enhancement," *Speech Communication*, Elsevier, vol. 49 (no. 7,8), pp. 574-587, 2007.

[3] Marcus G. Pandy and Necip Berme, "Synthesis of human walking: A planar model for single support," *Journal of Biomechanics*, vol. 21 (no. 12), pp. 1053 - 1060, 1988.

[4] M. H. Pope, H. Broman and T. Hansson, "Impact response of the standing subject - a feasibility study," *Clinical Biomechanics*, vol. 4 (no. 4), pp. 195-200, 1989.

[5] Lundstrom, R., "Local vibrations-mechanical impedance of the human hand's glabrous skin," *Journal of Biomechanics*, vol. 17 (no. 2), pp. 137-144, 1984.

[6] J. Fiene, K. J. Kuchenbecker and G. Niemeyer, "Event-based haptic tapping with grip force compensation," *Haptics Symposium*, IEEE, pp. 117-123, March 2006.

[7] J. C. Craig, "Difference threshold for intensity of tactile stimuli," *Perception and Psychophysics*, vol. 11, pp. 150–152, 1972.

[8] Y. Tanaka, T. Onishi, T. Tsuji, N. Yamada, Y. Takeda and I. Masamori, "Analysis and modeling of human impedance properties for designing a human-machine control system," *ICRA*, IEEE, pp. 3627-3632, April 2007.

[9] T. A. Kern, S. Klages, T. Meiss and R. Werthschützky, "Study of the Influence of Varying Diameter and Grasp-Forces on Mechanical Impedance for the Grasp of Cylindrical Objects," *Eurohaptics Conference*, pp. 113–118, June 2006.

INITIAL DEVELOPMENT OF A VARIABLE-FRICTION FLOOR SURFACE

Guillaume Millet, Martin J.-D. Otis, Gary Chaw, Jeremy R. Cooperstock
Centre for Intelligent Machines, McGill University

ABSTRACT

This paper investigates the design of a variable-friction floor device for clinical and rehabilitation applications. Several designs based on rolling elements were proposed and investigated. Coefficients of static friction were measured to compare capabilities of floor tiles made of ball transfer units or covered with PTFE. The measurements showed that while a device based on rolling elements is limited in supported footwear and in simulating heel strike, and can lead to a complex friction-variation system, it can simulate friction coefficients as low as ice.

INTRODUCTION

Falls are a major cause of serious injuries and even deaths. One in three Canadians aged 65 and over will fall at least once every year, causing up to 88% of injuries in this age group [1]. One of the most hazardous conditions is ice where the very top layer is water. A number of human-centred approaches have been used to estimate slipping and falling hazards and risks [2] and to analyze biomechanics [3] and motor patterns [4] when anticipating or walking on slippery floors. One of the challenges of biomechanical studies is reproducing the unexpected nature of real-life slipping accidents [3], which is impeded by the lack of variable-friction floor devices.

In the context of walking, friction is the force resisting the relative motion between the floor surface and a walker's shoe. Reducing friction is usually accomplished by one of the three following methods. One is by using low-friction materials, such as high-density polyethylene (HDPE) or polytetrafluoroethylene (PTFE). Another is by changing dry friction into lubricated friction, that is, placing a lubricant such as oil, water, or grease between the two surfaces. A third way is to change sliding friction into a much smaller type of rolling friction, with devices such as rolling-element bearings. Friction is commonly quantified with Coulomb's model; examples of coefficients of static friction are 0.04 for PTFE/PTFE contact, 0.02 for ice/steel, or under 0.01 for rolling elements.

Designing a floor device able to generate a wide range of friction coefficients when people walk on it poses two interdependent issues: how to achieve very low friction under load and how to increase friction in a controllable manner. Lubricated friction is not considered here for practical reasons. The very low friction of rolling encouraged us to investigate rolling-based designs of variable-friction floor devices.

1D VARIABLE-FRICTION FLOOR TILE

Rolling objects are frequently used for transport, in particular for moving heavy loads efficiently. A common example is the roller conveyor systems used in material handling and packaging industries. Their capability to provide both very low friction under the weight of a human and a relatively simple way to vary their apparent friction make them a good candidate for a variable-friction floor surface. Indeed, rollers, as 1D-rotating devices, can be borne by bearings on other surfaces than the one in contact with the load. For example, rollers are usually borne by ball bearings located inside, on the sides. Hence, the external surface of rollers can be in contact with both feet above and a braking mechanism underneath.

To ensure a certain spatial resolution in the friction simulation, an array of sets of rollers is necessary, where each set forms an independent variable-friction floor tile. Figure 1 illustrates the design of such a tile, consisting of thirteen ⌀2.2 x 25 cm standard conveyor rollers each separated by a 0.8 mm gap. The frame that bears the rollers' shafts is mounted on four load cells that are set under each corner so as

to measure normal forces; a load cell on each side allows measurement of tangential forces. The measure of foot-applied forces can be used for biomechanical analyses but also as inputs for closed-loop control of the friction.

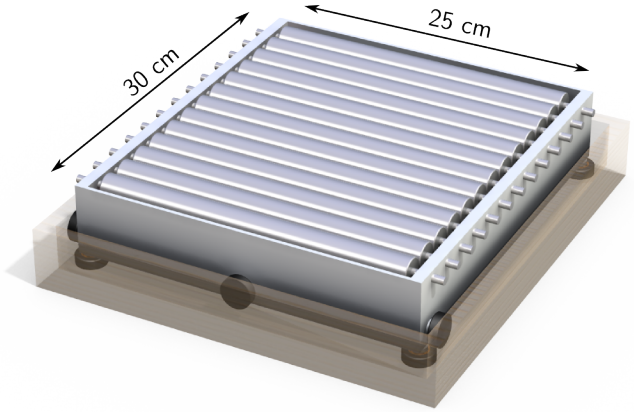


Figure 1: CAD view of a 1D variable-friction floor tile made of conveyor rollers

The variation of friction is realized by a braking mechanism under the rollers. Figure 2 shows an example design in which a brake lining **a** is pressed against rollers **b** above. A lever-arm system is used to apply sufficient force on the brake lining. The plate **c** hinges on the axis **d** and compresses the foam insert **e**, which behaves like an angular spring between the plate and the lining. The axis of rotation is both in the tangent plane of the rollers and in the upper plane of the lining. In this way, the contact surface between the rollers and the lining is distributed over the entire width of the lining and, assuming that the foam has a linear stiffness curve, the force applied by the brake lining on the rollers is the same along the width of the lining. This helps distributes the wear and, as a result, maintains the friction properties of the device over time.

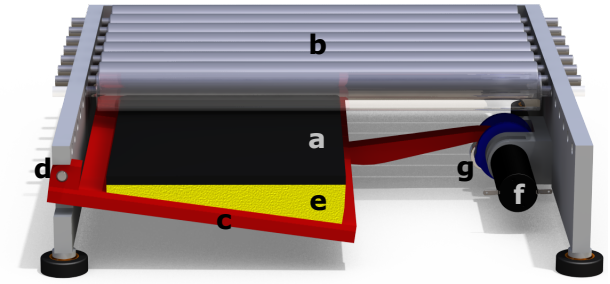


Figure 2: Overview of the braking mechanism

Actuation of the brake is achieved with a conventional ⌀22 mm DC gear motor **f** driving an off-centre bearing **g** that moves the arm of the plate **c**. As can be seen in Figure 3, the maximal braking force is achieved when the eccentric **g** is at the top of its circular trajectory, which corresponds to very little torque for the motor since the load at this position is borne by the hinge of the eccentric disk. This design results in a compact, simple, and relatively low-cost solution.

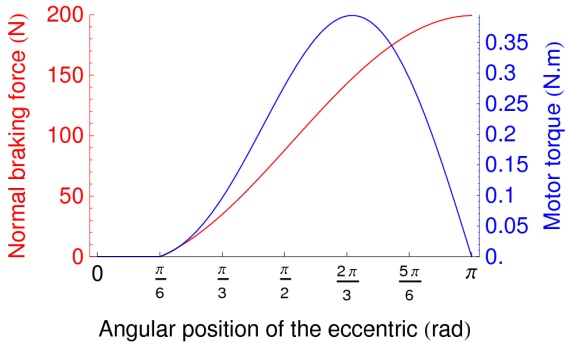


Figure 3: Braking force applied on the rollers and associated motor torque, according to the angular position of the eccentric

Limitations

Although this novel brake mechanism could be effective for altering levels of friction, it suffers from several issues that make it ill-suited for applications in mobility and biomechanics studies. First, the friction variation is restricted to one dimension; in other words, the friction display is *anisotropic*, whereas slippery floors are usually equally slippery in all directions. This difference is likely to induce a different gait from that used on an equally slippery floor. Nevertheless, cases exist where friction anisotropy can be a desirable feature, for example, to simulate the effect of a roller or ice skate.

A second issue relates to the floor shape. Rollers form an *uneven surface* that does not simulate natural floors and would have significant perceptual consequences, mainly if the walker is barefoot and to a lesser extent with shoes. Adding a belt around the rollers may reduce this issue to a certain degree, as the junctions between tiles would nevertheless remain uneven. However, this could significantly impact the fidelity of the friction simulation since it would increase the inertia of moving elements under the foot, as all the belt and rollers would

have to move simultaneously. Another way to flatten the surface is to decrease the roller diameter. Custom rollers using needle roller bearings could lower the diameter to 10 mm.

Another potential issue related to the use of rolling elements is the *vibrotactile noise* emitted by the rollers when turning, which could have perceptual consequences and influence balance control [5]. It could also interfere with vibrotactile actuation if the latter was implemented in the system, as in our previous work [6].

A fourth concern is the *visibility* of the rollers. This provides a visible cue of friction anisotropy and, as a result, is likely to influence people to adjust their gait and locomotion strategies [3]. However, the rollers could be covered with a suitable paint and concealed with video projection techniques.

In summary, the design presented here raises potential issues with respect to several modalities: kinesthetic (friction anisotropy), tactile (shape and vibrotactile), and visual. While the tactile and visual issues can be compensated to a certain extent, the friction anisotropy remains a core limitation of a roller-based design. This led us to investigate the design of an omnidirectional variable-friction device.

2D VARIABLE-FRICTION FLOOR DEVICE

Figure 4: CAD view of an omnidirectional low-friction floor tile, made of ball transfer units

One way to vary friction in both horizontal directions is to use omnidirectional rolling devices, such as ball transfer units (BTUs). Their principle is identical to that of a trackball. As illustrated in Figure 4, a BTU involves a large load-bearing ball supported by many small balls encapsulated in a hemispherical cup. They are commonly used in an array to move objects with minimum effort in any direction, e.g., in cargo and baggage handling applications. However, unlike 1D-rotating devices such as rollers, the large ball in BTUs is borne on the same surface as the one in contact with the load. This precludes the design of a braking mechanism that is in contact with the ball from underneath, as with rollers.

Different types of BTUs exist according to the fixing method. The appropriate type for our purpose should minimize the outside diameter of the frame compared to the ball size, in order to maximize the array density of units and reduce the uneven-surface issue. These factors suggest the use of bolt or glide[1] unit. This type of units is simple and inexpensive, and its large ball exposure can be advantageous in the design of a braking mechanism. Commercial glide units contain a ⌀25 mm ball, allowing for construction of an array of 81 units on a 30 x 30 cm tile (see Figure 4) or of 99 units in a hexagonal arrangement. In order to avoid jamming at heel strike, a cover plate above the BTUs' frames would flatten the surface, so that the balls protrude only 2 mm out of the plate.

The friction variation of a BTU-based floor tile could be accomplished in two different ways. First, braking friction could be applied by pressing a rubbing element against each ball, either on its exposed part or on the side, through a hole drilled in the frame of the BTUs. Both options present severe constraints due to the very limited space available for such elements. The second method, raising braking pins between the BTUs that rub against the shoe, may be more reliable, as this does not require any contact with the balls themselves. Different levels of friction could thus be obtained by controlling the number, the size, the height, and/or the material of the raised pins.

This method of friction variation could be implemented in another type of design, without BTUs but with low-friction materials instead. However, BTUs are expected to provide less friction. To verify this, we constructed two prototypes of slippery tiles, one with BTUs (see Figure 5) and another covered with PTFE (see Figure 6). We then compared their coefficients of static friction μ, calculated from empirical measurements of the friction angle φ such as $\mu = \tan \varphi$, as measured by the apparatus illustrated in Figure 6. The tile **a** was tilted by a lin-

1 http://www.skf.com/files/600570.pdf

ear actuator **b** (D-Box Odyssey). Its orientation and the movement of shoes were measured by motion capture (OptiTrack) on four people of weight between 50 and 80 kg. The shoes were covered with PTFE sheets for the PTFE tile and with aluminum plates for the BTU tile, as discussed below.

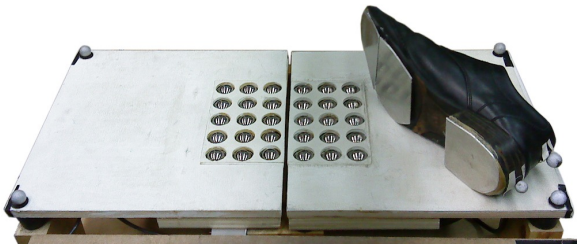


Figure 5: Prototype of slippery tiles with BTUs, made from the vibrotactile floor tiles described in reference [6]

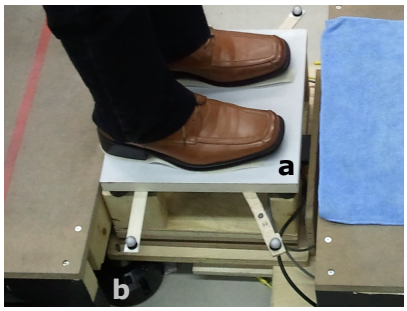


Figure 6: Setup for measuring friction angles

Coefficients of 0.11±0.01 and 0.025±0.005 were measured for the PTFE and BTU tiles, respectively. Although BTUs provide a lower coefficient of static friction, close that of ice, they present some important limitations.

Limitations using BTUs

The spherical surface of BTUs creates a *high contact pressure* between the balls and the walker, which deforms the sole of the foot or the shoe, in particular with soft soles. This can result in the sole, whether barefoot or rubber, contacting the cover plate, which in turn produces friction during walking.

Another constraint, closely related to the issue of uneven surface discussed above, is the space between the BTUs due to their frame. This prevents simulation of a realistic heel strike if the foot is highly inclined, such as during long strides. This limitation could be reduced with a denser array of units and a rounded heel edge. Smaller commercial BTUs, known as miniature BTUs, contain a ⌀5 mm ball, but their frame is larger compared to that of glide units. To avoid the heel contacting the frame between the balls, it would need to be rounded sufficiently.

Visibility of balls is harder to compensate than rollers. Steel balls are reflective and cannot be covered with paint, while nonreflecting balls made out of nylon cannot bear sufficient load. Nevertheless, if the entire floor consists of variable-friction BTU tiles, slippery conditions cannot be guessed solely by vision.

CONCLUSION

We proposed and discussed several designs of rolling-based variable-friction floor devices. Ball transfer units allow friction coefficients as low as that of ice but impose some limitations on footwear and for simulating heel strike. Future work includes investigation of another promising approach to vary friction by combining low- and high-friction surfaces in a controllable manner.

ACKNOWLEDGEMENTS

This research was supported by a grant from the MDEIE of Quebec for the European 7th Framework Programme project Natural Interactive Walking (NIW, no. 222107). The authors gratefully acknowledge this support and thank Yon Visell, Jennifer Campos, Yue Li, and Jen Hsu for their comments.

REFERENCES

[1] V. Scott, L. Wagar, and S. Elliott. "Falls & related injuries among older canadians: fall-related hospitalizations & intervention initiatives." Technical report, Public Health Agency of Canada, 2010.

[2] R. Grönqvist et al., "Human-centred approaches in slipperiness measurement," *Ergonomics*, vol. 44(13), pp. 1167–1199, 2001.

[3] R. Cham and M. S. Redfern, "Changes in gait when anticipating slippery floors," *Gait & Posture*, vol. 15(2), pp. 159–171, 2002.

[4] G. Cappellini, Y. P. Ivanenko, N. Dominici, R. E. Poppele, and F. Lacquaniti, "Motor Patterns During Walking on a Slippery Walkway," *J. Neurophysiology*, vol. 103, pp. 746–760, 2010.

[5] R. Roll, A. Kavounoudias, and J. Roll, "Cutaneous afferents from human plantar sole contribute to body posture awareness," *Neuroreport*, vol. 13(15), pp. 1957–1961, 2002.

[6] Y. Visell, J. Cooperstock, "Design of a Vibrotactile Display via a Rigid Surface," *Proceedings of the IEEE Haptics Symposium*, 2010.

AN EXPLORATION ON THE INFLUENCE OF VIBROTACTILE CUES DURING DIGITAL PIANO PLAYING

Federico Fontana
Marco Civolani
Università di Udine
Department of Mathematics
and Computer Science
via delle Scienze, 206
33100 Udine, Italy
federico.fontana@uniud.it

Stefano Papetti
Valentina del Bello
Università di Verona
Department of Computer Science
strada Le Grazie, 15
37134 Verona, Italy
stefano.papetti@univr.it

Balázs Bank
Budapest University of Technology
and Economics
Department of Measurement
and Information Systems
H-1521 Budapest, Hungary
bank@mit.bme.hu

ABSTRACT

An exploratory experiment was carried out in which subjects with different musical skills were asked to play a digital piano keyboard, first by following a specific key sequence and style of execution, and then performing freely. Judgments of perceived sound quality were recorded in three different settings, including standard use of the digital piano with its own internal loudspeakers, and conversely use of the same keyboard for controlling a physics-based piano sound synthesis model running on a laptop in real time. Through its audio card, the laptop drove a couple of external loudspeakers, and occasionally a couple of shakers screwed to the bottom of the keyboard. The experiment showed that subjects prefer the combination of sonic and vibrotactile feedback provided by the synthesis model when playing the key sequences, whereas they promote the quality of the original instrument when performing free. Even if springing out of a preliminary evaluation, these results were in good accordance with the development stage of the synthesis software at the time of the experiment. They suggest that vibrotactile feedback modifies, and potentially improves the performer's experience when playing on a digital piano keyboard.

1. INTRODUCTION

For its versatility and diffusion in diverse musical styles, with the advent of electro-mechanics, electronics, and finally digital technology, the piano has been progressively re-designed and engineered in different forms mainly to make its portability easier. Although sounding quite different, siblings such as the Clavinet and Rhodes electric pianos were initially able to keep a certain flavor of the original instrument, and then to conquer their own niche in contemporary music. In the meantime the early digital piano keyboards had begun to revolutionize the musical instrument market, by making piano performances possible on small stages and at home, where it would be otherwise unpractical or impossible to set up the original instrument.

With the advance of technology, digital pianos have progressively increased in sound accuracy and fidelity of their keyboard's response. Current flagship products exhibit sounds and key mechanics that satisfy the performer completely, once taken into account the relatively minor cost, size and weight of the digital instrument compared to its mechanical counterpart. On the other hand, the issue of vibrotactile feedback has been still left largely unexplored, despite playing a fundamental role in the performance on a real piano.

In spite of the current psychological and applied research trend toward a more systematic inclusion of vibrotactile devices in musical interfaces [1, 2, 3], also thanks to the notable decrease in costs of the related actuation technologies, we found only few studies on the topic of vibrations in the piano. Among these studies [4] there are a computational solution for improving the vibrotactile feedback provided by the upright piano through adaptation of the keybed impedance toward the characteristic values of the grand piano [5], and signs of research activity advertised in the CCRMA web pages, on tactile feedback design applied to the keyboards based on previous research made by Chafe [6].

Performers traditionally experience vibrotactile feedback in digital pianos only as a by-product of the resident loudspeaker system. By transmitting vibrations across the instrument body during the sound reproduction, the loudspeakers are in fact responsible of providing some related cues to the player. Arguably, such cues cannot achieve the intensity nor resemble the quality of those originating from a real piano keyboard, when the soundboard resonates under the action of the strings.

At least one flagship product, the Yamaha AvantGrand digital piano (see www.avant-grand.com), augments sounds by means of an active vibration system. By enabling transducers located under the keyboard and behind the music stand, it promises to engage users in a full-body sensory experience during playing [7, 8]. Indeed, the multimodal perception of harmonic components across the whole human body has been recognized to increase the engagement and sense of presence in users [9].

Copyright: ©2011 Federico Fontana et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Figure 1. Clavinova YDP-113.

In the case of the Yamaha product, all its design solutions concur to form an extremely faithful reproduction of the original instrument, including its visual appearance. Due to unavoidable contingencies, in our experiment we rather made use of a less sophisticated musical interface, in both visual and non visual sense. More precisely:

- we used a Yamaha Clavinova YDP-113, an inexpensive digital piano (see Figure 1) that provides internal loudspeakers, but also offers MIDI master keyboard functionalities;
- in alternative to the Clavinova sounds, we synthesized sonic and vibrotactile feedback in real time by means of a physics-based piano sound model running on a laptop.

Vision did not play an integrative role in the multimodal scene. In fact, due to the appearance of the Clavinova-based setup, subjects were constantly aware that they were *not* playing a real piano. Using this setup we were able to provide the subjects with different sounds, with and without vibrotactile feedback.

The results of our tests overall suggest that the subjective judgments on sound quality were influenced by the vibrotactile feedback. Furthermore, as shown in the following of this paper, an analysis of the individual judgments indicates that the tactile modality can improve the auditory perception of digital piano sounds. However promising, such results call for a more robust and systematic validation that is expected to become object of future research.

2. EXPERIMENT

The experimental hypothesis was that the vibrotactile feedback coming from the instrument keyboard had influence on the perceived quality of piano sounds.

2.1 Subjects

Nine subjects voluntarily participated in the experiment. Three of them were pianists, four of them were other instrument players, and two of them were non-musicians.

Figure 2. Monacor Carpower BR 25 shaker.

2.2 Setup and configurations

Instead of explicitly being involved in a vibrotactile evaluation task, subjects were asked to rate the sound quality associated to two different digital piano settings: in the former the Clavinova worked with its internal loudspeakers; in the latter, the Clavinova controlled a synthesis software running in real time on a Core 2 Duo Dell Latitude E6400 laptop, in its turn driving a pair of Genelec 2029BR external loudspeakers along with a pair of Monacor Carpower BR 25 shakers.

The shakers (one is shown in Figure 2) were screwed to the bottom of the Clavinova. They are able to transmit mechanical power to the body they are in contact with. As a side effect they also generate some sound, amounting to few dB of intensity level that adds to the loudspeaker emission.

In spite of a claimed active band in the range 30-300 Hz, we measured that the Monacor shakers in practice work up to a few kHz, hence covering sufficiently well the entire vibrotactile perceptual band in correspondence of the finger, in particular including its higher sensitivity region centered around 250 Hz [10].

As noted by Bank [11], the active range of the shaker is sufficient to excite all the components that can be perceived by the palm [12] during normal piano playing. Figure 3 shows examples of such thresholds in dashed lines for the notes C_2, C_4, C_6, and the C chord $\sharp 2$.

The software running on the laptop implemented a recently developed physics-based model [13] for the synthesis of piano sounds. The model was configured to compute two sound signals in correspondence of the left and right part of the soundboard. These signals formed the output for the Genelec loudspeakers and, equivalently, for the shakers.

In their own admission, the professional "golden ears" working on its fine-tuning, at the time of the experiment the model was not yet well balanced in the higher octave range. Furthermore we did not apply any amplitude, nor spectral equalization to the signals feeding the shakers: such manipulations are needed to simulate the vibrotactile response of specific piano keyboards such as those investigated by Bank [11]. For our purpose, we just tuned the intensity level of the shakers based on the subjective impression of two expert piano players who helped realize

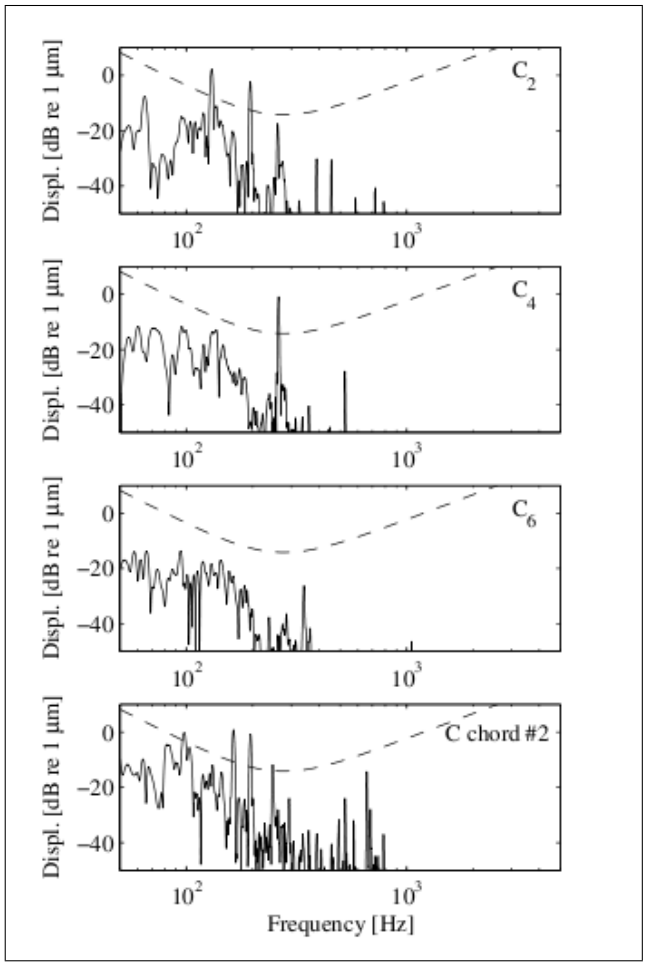


Figure 3. Spectra of the tactile response of a Bösendorfer grand piano measured at the keyboard (solid line), along with perceptual thresholds for the palm from Verrillo [12] (dashed line). Examples given for notes C_2, C_4, C_6, and the C chord ♯2, all played at *forte* level.

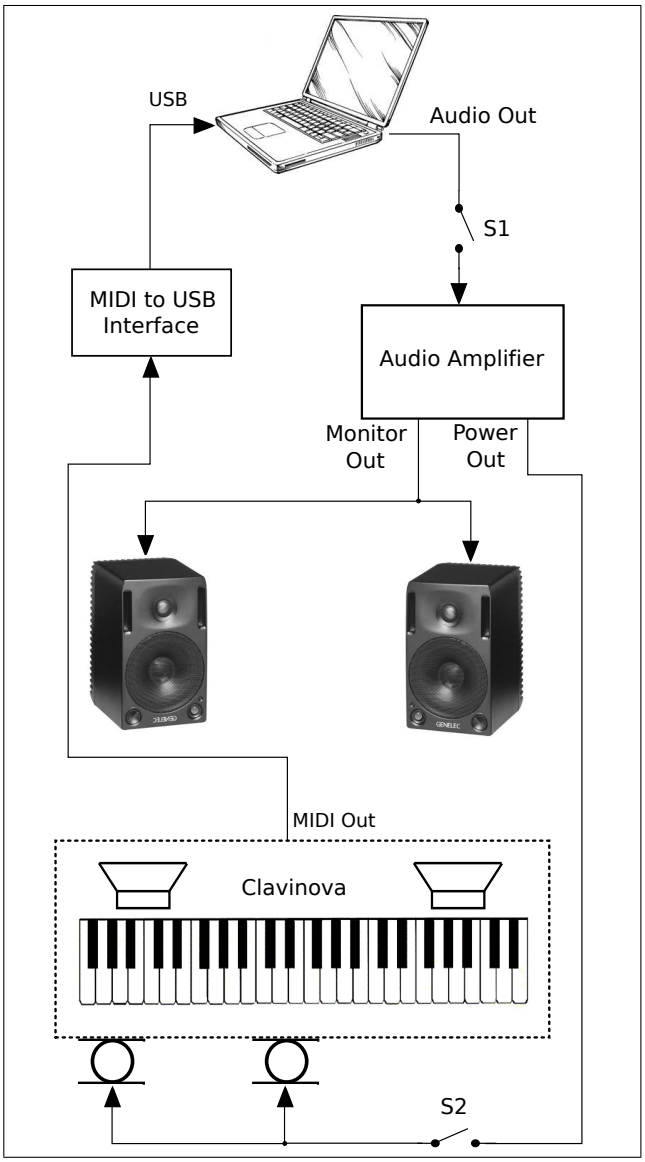


Figure 4. Experimental setup.

the experiment.

Figure 4 illustrates the experimental setup. Powered by a Pioneer A-225 stereo amplifier, the two shakers were respectively positioned under the leftmost and mid part of the keyboard. In this way they emitted energy in correspondence of the lower octaves, whose keys are mostly responsible of producing vibrations falling within the tactile perceptual band (see Figure 3).

Subjects were exposed to three possible experimental configurations—refer also to the positions of the switches S1 and S2 in Figure 4:

C: Clavinova only (Clavinova speakers on; S1 off);

M: physical model with Genelec (Clavinova speakers off; S1 on; S2 off);

MS: physical model with Genelec and shakers (Clavinova speakers off; S1 on; S2 on).

The intensity levels were equalized so as to minimize the overall loudness changes across the three configurations, i.e., by setting the level of C midway M and MS. As a result, we measured Sound Pressure Levels (SPL) between 69 dB(A) (configuration M) and 71.5 dB(A) (configuration MS), with configuration C lying between the two. At any moment, by operating on the amplifier and by turning up or down the main volume of the Clavinova to predefined levels, the experimenter could easily switch between the different configurations. Subjects were not informed about the presence of the shakers below the keyboard, nor obviously about the fact that these devices could be switched on and off during the experiment.

2.3 Task

Subjects were initially asked to play the keys F, G, A, B of the lower four octaves (numbered 1, 2, 3 and 4 starting from the left side) in both directions. In more detail, they had to perform four tasks: playing i) an ascending *staccato* using only their right forefinger, ii) an ascending *legato* using at each octave the first four fingers of their right hand, iii) a descending *staccato* using only their right forefinger, and iv) a descending *legato* using at each octave the first four fingers of their right hand. The ascending pat-

terns were performed by moving upward across the keyboard, conversely the descending patterns were performed starting from the fourth octave down to the first one. The sixteen hot keys were marked with a red pencil to help unpracticed piano players accomplish the task without effort.

In addition to the tasks explained above, subjects with self-reported sufficient ability to play the piano were invited to perform freely on the instrument, for instance by playing one of their preferred songs.

2.4 Method

The four tasks were repeated three times across the different configurations C, M and MS, for a total of $4 \cdot 3 \cdot 3 = 36$ randomized *short* trials for each subject. This part of the experiment took about 35 minutes. Additional 15-20 minutes were required by the pianists to accomplish the free performance in the three configurations C, M and MS, summing to three additional *long* trials for these subjects.

At the end of every short trial, subjects marked the perceived sound quality on a scale ranging from 1 (very low) to 7 (very high). Long trials were instead judged qualitatively by the pianists. At the end of the experiment, the subjective skill in playing the piano was rated (1 to 7) along with the difficulty in performing the task (low, medium, high, very high). Only one subject reported a medium difficulty in performing the task, all the rest of the group otherwise rated the task difficulty to be low.

3. RESULTS

The aggregation of the judgments given by all the subjects during the short trials provides three sets of $4 \cdot 3 \cdot 9 = 108$ ratings, each set corresponding to a respective configuration. The mean values respectively amount to 3.741 for C, 3.981 for M, and 4.185 for MS (last row of Table 1).

A repeated-measures ANOVA conducted on the subjects' average ratings under the three configurations states that the mean values are significantly different at a 2% significance level: $F(2, 8) = 3.44$, p $= 0.017$. A similar analysis, made by restricting the attention to couples of such sets, shows that the difference is significant for both M and MS ($F(1, 8) = 15.46$, p < 0.001) and C and MS ($F(1, 8) = 3.44$, p < 0.017), wheras it is not significant between C and M ($F(1, 8) = 1.89$, p $= 0.19$). Furthermore, t-tests pairing the subjective judgments across the different conditions show p-values respectively equal to $p_{\text{C}\leftrightarrow\text{M}} = 0.176$, $p_{\text{C}\leftrightarrow\text{MS}} = 0.015$, and $p_{\text{M}\leftrightarrow\text{MS}} = 0.256$, with obvious meaning of the subscripts of p.

Concerning the free performance, the three pianists respectively opted for playing a pop song by the Beatles, an improvised jazz tune, and a *preludio* by Bach. All of them had a strong preference for the C configuration.

4. DISCUSSION

The different judgments existing between the single key patterns and free performance were almost certainly affected by the limited accuracy of the physical model in the higher octaves. It seems clear that as soon as the pianists heard a degradation of the sound quality, the vibrotactile modality lost any significance in their subjective judgment, and consequently the physics-based model was downgraded in their judgments.

Mean Values			
Subject	**C**	**M**	**MS**
Pianists			
2	4.000	5.000	5.167
5	3.833	5.000	5.250
8	4.084	2.333	2.083
Other Musicians			
1	4.083	5.250	5.083
4	2.833	3.500	2.917
7	5.667	4.167	4.333
9	3.750	3.750	4.250
Non Musicians			
3	3.000	3.500	4.250
6	2.417	3.333	4.333
Aggregate			
All	**3.741**	**3.981**	**4.185**

Table 1. Mean values for the different configurations, by subject plus aggregate.

Conversely, the results obtained by judgments on single notes ranging in the lower four octaves reveal that the vibrotactile feedback adds discrimination in the otherwise not significantly different judgment of the physics-based sound against the Clavinova samples.

However encouraging, the result on the aggregate data becomes less dramatic if reported on an individual basis. In fact, if we analyze the significance of the different configurations subject-by-subject then we discover that, once taken individually, subjects tend to grade the configurations mainly by their sound, and rarely the difference between M and MS gains significance.

Figure 5 shows, using double-sided arrows, subject-by-subject significance of the differences between configurations, obtained by computing t-values of the corresponding paired data at 5% significance level. The aggregate result presented in Section 3 is shown as well, at the bottom of the same figure.

From this figure, along with a look at the subjective mean values listed in Table 1, it can be observed that pianists seem to fairly weigh the vibrotactile modality, and rather base their judgments on robust decisions informed by the auditory modality—indeed, the question to the subjects was exactly that of rating the quality of the sound. As we move toward less specialized listeners, i.e. other instrument players, the corresponding mean values and related significances become more variate, including those of two subjects who do not appreciate any significant difference among the experimental configurations. Finally, non-experienced subjects exhibit less common judgments, including one subject who finds significant differences between M and MS.

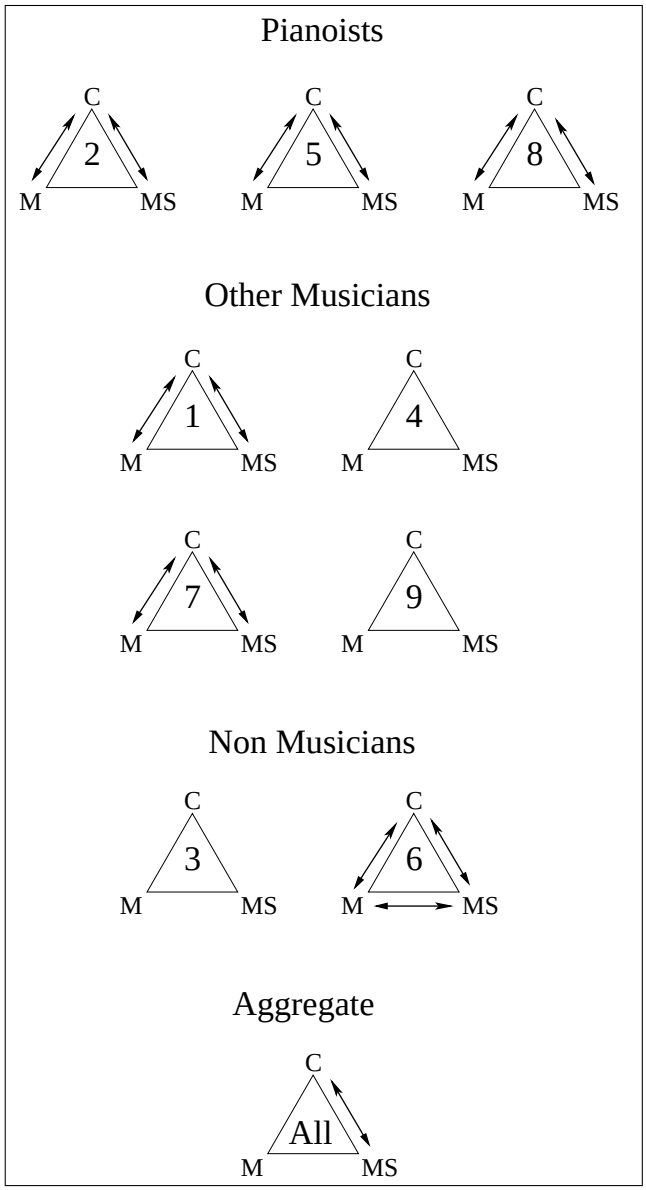


Figure 5. Significance of judgments, by subject plus aggregate. Double-sided arrows connect configurations whose different judgment is significant at 5% level, according to t-tests pairing the corresponding data. Subject number written at the center of the corresponding triangle.

Even if the number of subjects populating the different categories is too scarce to signify anything, a possibly relevant case record can be inferred from the subject-by-subject analysis. Specifically there may be a trend, potentially indicating a decreasing importance of the vibrotactile feedback for subjects who have more familiarity with musical sound evaluations, with minimal significance of the tactile modality for pianists. On the other hand, the results of non-experienced subjects may suggest that physically-consistent vibrotactile feedback can help unpracticed users enter into contact with musical interfaces such as keyboards, whose complexity of use is well known by practitioners.

In summary, the proposed experiment represents a first, far from being exhaustive attempt to understand the importance of vibrotactile cues in digital piano playing. Furthermore, it shows limits in setup and methodology. Contingent difficulties for the experimenter in recruiting a sufficient number of subjects with different musical skill levels, together with the impossibility (due to limits in manpower) to prepare two acoustically different physics-based piano models having comparable sound quality, prevented to include a control session in which subjects could compare two models without vibrotactile feedback. Were musicians biased by the assumption that the vibration of a digital piano keyboard could not be changed during the tasks, or was the vibrotactile feedback equalized too roughly to elicit a definite sensation of quality improvement in pianists, are just two among the many questions that cannot be answered using our data.

5. CONCLUSIONS

In an experiment where subjects, after playing a digital piano, had to rate two different instrument models, we investigated the salience of vibrotactile cues as potential enhancers of the sound quality. Specifically, the test was made by switching on and off two shakers while subjects were performing a task with the latter model. Conversely, the former was played without any modification of its vibrotactile feedback.

Results say that, overall, the inclusion of the vibrotactile modality adds a significant improvement to the quality of the sound of the latter model. An analysis conducted subject-by-subject suggests that differences exist among the individual judgments on the same models, without a specific preference for the configuration enabling the vibrotactile feedback.

A classification of the subjects based on their knowledge of musical instruments, specifically the piano, was postulated to investigate musical skill as a possible predictor of preference. In the limits of the low number of subjects forming the three resulting classes, the subjective analyses indicate that pianists may be only weakly (albeit not significantly) influenced by the vibrotactile augmentation when making judgments on sound quality, whereas other musicians and non-musicians may be influenced more. Nevertheless, this conclusion is purely tentative at this stage of the experimentation.

Encouraged by this experience, we are planning to follow up with a more robust setup and overall experimental design. Concerning the setup, we will operate an equalization of the signals from the shakers meanwhile providing all the conditions for comparing configurations, in which the control of the auditory and vibrotactile feedback will be completely independent and orthogonal. Concerning the experimental methodology, new subjective tasks will be designed allowing for the extraction of more reliable figures of sound quality and realism of the overall experience.

Acknowledgments

The research leading to these results has received partial funding from the Joint Project E-PHASE, participated by the University of Verona and Viscount SpA, and from the

European Community's Seventh Framework Programme under FET-Open grant agreement 222107 NIW - Natural Interactive Walking.

6. REFERENCES

[1] A. Askenfelt and E. V. Jansson, "On vibration and finger touch in stringed instrument playing," *Music Perception*, vol. 9, no. 3, pp. 311–350, 1992.

[2] A. Galembo and A. Askenfelt, "Quality assessment of musical instruments - effects of multimodality," in *5th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM5)*, Hannover, Germany, 2003.

[3] E. R. Miranda and M. M. Wanderley, *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard*. Middleton, WI: A-R Editions, 2006.

[4] W. Goebl and C. Palmer, "Tactile feedback and timing accuracy in piano performance," *Experimental Brain Research*, vol. 186, no. 3, pp. 471–479, 2008.

[5] M. Keane, "Improving the upright piano," *Acoustics Australia*, vol. 35, no. 1, pp. 11–15, Apr. 2007.

[6] C. Chafe, "Tactile audio feedback," in *Proc. Int. Computer Music Conf.*, Tokio, Japan, Sep. 10-15 1993, pp. 76–79.

[7] E. Guizzo, "Keyboard maestro," *IEEE Spectrum*, vol. 47, no. 2, pp. 32–33, Feb. 2010.

[8] Yamaha Corp., "Electronic Musical Instrument," 1988, US patent 5,189,242.

[9] M. Altinsoy and S. Merchel, "Cross-modal frequency matching: Sound and whole-body vibration," in *Haptic and Audio Interaction Design*, ser. Lecture Notes in Computer Science, R. Nordahl, S. Serafin, F. Fontana, and S. Brewster, Eds. Springer Berlin / Heidelberg, 2010, vol. 6306, pp. 37–45. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-15841-4_5

[10] R. T. Verrillo, "Psychophysics of vibrotactile stimulation," *J. of the Acoustical Society of America*, vol. 77, no. 1, pp. 225–232, Jan. 1985.

[11] B. Bank, "Vibration of piano keys – measurement results and implementation possibilities," Dept. Comp. Sci., University of Verona, Italy, Tech. Rep., 2008, internal project report.

[12] R. T. Verrillo, "Effect of contactor area on the vibrotactile threshold," *J. of the Acoustical Society of America*, vol. 35, no. 12, pp. 1962–1966, 1963.

[13] B. Bank, S. Zambon, and F. Fontana, "A modal-based real-time piano synthesizer," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 18, no. 4, pp. 809–821, 2010, special Issue on Virtual Analog Audio Effects and Musical Instruments.

Effects of ecological auditory and vibrotactile underfoot feedback on human gait: a preliminary investigation

Stefano Papetti
Dept. Computer Science
University of Verona
stefano.papetti@univr.it

Federico Fontana
Dept. Mathematics and Computer Science
University of Udine
federico.fontana@uniud.it

Marco Civolani
Dept. Mathematics and Computer Science
University of Udine
marco.civolani@uniud.it

ABSTRACT

We describe an experiment where subjects were asked to walk along a predefined path while wearing a pair of instrumented sandals which provide auditory and vibrotactile feedback. Three experimental conditions were set up, corresponding to different feedback: two conditions simulating different ground materials, and one neutral (i.e. control) condition without artificial feedback. Preliminary results indicate that non-visual interactive feedback in a walking task influences, although not significantly, the subjects' gait patterns. Interaction designers may exploit these results in foot interfaces which provide vibrotactile and auditory cues to display virtual ground properties for non-intrusive navigation aid.

Author Keywords

Foot interfaces, Actuated footwear, Navigation aid, Interaction design, Ecological cues, Prototype.

ACM Classification Keywords

H.5.1 Information Interfaces and Presentation: Multimedia Information Systems—*Artificial, augmented, and virtual realities*; H.5.2 Information Interfaces and Presentation: User Interfaces—*Auditory (non-speech) feedback; Haptic I/O*

General Terms

Experimentation, Human Factors, Measurement.

INTRODUCTION

Despite being part of our everyday life, walking is the result of the complex integration of visual, auditory, tactile, vestibular and proprioceptive cues [8]. Concurrently with it, humans are able to carry out several other tasks, thus generally leaving walking to the periphery of the attention and calling it back to the main focus only occasionally, for instance when salient changes in the ground properties occur (e.g. a step or a pothole). Evidences like this indicate that it is possible to deliver informative messages during walking without hampering one's locomotion. In this regard, interaction designers have recently begun to promote virtual floor concepts as means to exchange non-visual information through the feet and the ground [15].

Several studies have investigated the possibility of influencing gait patterns by using rhythmical auditory cues, e.g. for rehabilitation purposes [4, 9], also in addition to visual cues [13], as well as of improving balance control in elderly people [10].

The role of footstep sounds on posture has been studied in [7], while the effect of 3D audio and stereo graphics has been exploited in [12] to redirect walking in a virtual reality environment. Also, the effect of auditory startle on the avoidance of obstacles during walking has been studied in [11].

As for the influence of tactile feedback on locomotion, vibrations underfoot and applied to tendons have been studied respectively in [16] and [14, 3], while changes in sole sensitivity has been investigated in [5].

In what follows we describe a preliminary investigation on the influence on human gait of ecologically-consistent auditory and vibrotactile feedback provided by instrumented footwear. In particular, when used on a quiet solid floor, our footwear prototype can convey non-visual cues displaying different virtual grounds, such as mud or snow.

In a related work [1] the authors have shown that gait patterns, accounting for specific emotional intentions, are affected by ecological auditory and vibrotactile underfoot feedback, albeit not to a statistically significant degree. The results of our experiment support such previous conclusions, overall suggesting that supplying subjects which walk on a neutral, solid floor with non-visual interactive cues representing virtual grounds may have perceptual salience, however not to the point to push them into changing their usual walking style.

Such results may be exploited in navigation aid applications, providing users with informative yet non-intrusive auditory and vibrotactile underfoot cues. For instance, virtual pathways may be defined by simulating different ground materials through non-visual cues, thus helping normally sighted as well as visually-impaired persons keep the right track in complex functional spaces such as airports, railway stations and so on.

EXPERIMENT

Our experimental hypothesis was that human gait can be influenced by providing ecological audio-tactile feedback through the feet.

Subjects

Eight subjects, seven males and one female, of Italian nationality participated in the experiment. Their average age was 22.3 years.

Six subjects were right-handed and also considered their right foot as dominant. One subject was left-handed and one ambidextrous, also with regard to the use of feet. None of them reported any foot- or locomotion-related deficiencies.

Experimental setup

The equipment used in the experiment consisted in a pair of sandals provided with force sensing resistors (Interlink 402 FSR) fixed under the insole (one at the toe and one at the heel) and exciters (Dayton Audio DAEX32) affording audio-tactile feedback, which are embedded in the sole (one at the toe and one at the heel) [6]. Each sandal is connected via a single multi-conductor cable to a backpack holding some battery-powered devices: two small audio amplifiers, a wireless transmitter and two receivers. The sensors provide data via wireless transmission to a host computer, where they are processed and mapped to different real-time sound synthesis engines. The generated audio signals are then sent back wirelessly to the exciters, thus closing the interaction loop.

Thanks to the wearable lightweight interface (sandals plus backpack) and wireless communication, freedom of movement is ensured.

The round-trip latency exhibited by the system – measured as the delay between the onset of an impulse at the force sensors and the arrival of the corresponding feedback signals to the exciters [2] – amounts to about 20 ms, which results in coherence and unity in the multimodal percept.

Stimuli

Three experimental conditions were set up, corresponding to different stimuli: one control condition (N) without additional feedback, and two conditions (S and M) providing artificial audio-tactile feedback. The latter was generated in real-time by driving two different physics-based sound models[1] with the data provided by the force sensors: two ground materials were simulated, respectively icy snow (S) and mud (M).

Procedure

The participants were asked to wear the instrumented sandals and the backpack, and walk along a eight-shaped path in the experiment room. They were informed that a change in the multimodal feedback could occur at each trial. No other instructions were given.

In order to avoid biases due to the room configuration, half of the participants started walking from one of the shorter sides of the rectangular room and the other half from the other side.

[1]Taken from the Sound Design Toolkit: a library for ecologically-funded sound synthesis in Max/MSP and Pd, freely available at `www.soundobject.org/SDT`.

The experiment considered 8 trials for each condition, resulting in 24 trials for each experimental session, which lasted about 15 minutes. The order of the trials was randomized for each subject: before each trial the experimenter switched the feedback among the three conditions N, S and M, accordingly to such order.

In view of the following analysis, the force data provided by the four sensors were recorded separately for each trial, with a resolution of $10\text{bit}@1050\text{Hz}$ per channel.

RESULTS AND DISCUSSION

The recorded force data were analyzed in order to extract the following inter-onset intervals (IOIs) during walk: heel-to-toe intervals between contacts at the heel and the toe, for the left and right foot (respectively labeled h2tL and h2tR); inter-foot heel-to-heel intervals between contacts at the right and left heel (h2h); inter-feet heel-to-toe intervals between contacts at the right heel and left toe (h2tRL). The contacts were detected making use of thresholds and taking into account the slope of the force curves. The IOIs corresponding to the first and last steps in each trial were then discarded, as they introduced unreliable values in the measurement: indeed in such phases the subjects respectively accelerated from stance to a walking pace, and decelerated until stop.

A first one-way repeated measures ANOVA (RM-ANOVA) was performed on the data recorded in the neutral condition N, in order to verify whether the subjects walked with a regular pace when no stimuli were present. The different trials in the neutral condition N were considered as repeated measures for each subject, and the mean IOIs in each trial were used. The respective results for the four IOIs are shown in Table 1. The obtained p-values are very high, meaning

h2tL	F(7,7) = 1.36 , $p = 0.2420$
h2tR	F(7,7) = 0.21 , $p = 0.9814$
h2h	F(7,7) = 0.69 , $p = 0.6788$
h2tRL	F(7,7) = 1.38 , $p = 0.2345$

Table 1. Results from RM-ANOVA for the neutral condition N.

that the subjects walked with a very regular gait in the neutral condition N. In particular the h2tR interval appears extremely regular.

Considering the above results and since the majority of the participants (six over eight) were right-handed, we performed a second RM-ANOVA on such subjects only. As shown in Table 2, the IOIs from these subjects are much more irregular, except for the h2tR interval. These may indicate that

h2tL	F(7,5) = 2.30 , $p = 0.0485$
h2tR	F(7,5) = 0.97 , $p = 0.4713$
h2h	F(7,5) = 2.88 , $p = 0.0173$
h2tRL	F(7,5) = 6.03 , $p = 0.0001$

Table 2. Results from RM-ANOVA for right-handed subjects in the neutral condition N.

only the dominant foot provides reliable data. Also this may

suggest that a confidence value $\alpha = 0.05$ can be assumed in the analysis, thus allowing to isolate the irregularities found for the right-handed subjects in the neutral condition N.

The IOIs in the three conditions N, S and M were also tested using a RM-ANOVA, considering the mean IOIs over all trials in each condition. The results of Table 3 show that the

h2tL	F(2,7) = 2.22 , $p = 0.1452$
h2tR	F(2,7) = 3.09 , $p = 0.0775$
h2h	F(2,7) = 3.01 , $p = 0.0815$
h2tRL	F(2,7) = 2.58 , $p = 0.1108$

Table 3. Results from RM-ANOVA for the conditions N, S and M.

subjects' gait was slightly affected by the audio-tactile feedback provided in the conditions S and M, as the IOIs are more irregular compared to the neutral condition N (see Table 1). This is clearly evident especially considering the result concerning the h2tR and h2h intervals. By assuming a confidence value $\alpha = 0.05$ as suggested above, the results from the h2tR and h2h intervals are indeed close to statistical significance.

However, the relatively high p-values indicate that such effect needs further investigation. In particular the experiment should be repeated with a larger number of subjects, as eight subjects do not guarantee the conditions for performing a proper ANOVA test.

CONCLUSION

In a broader perspective, the lack of a clear statistical significance in the obtained results may have an alternative interpretation. In fact the ability to provide salient non-visual cues underfoot, which do not significantly alter one's walking style, may enable the design of foot interfaces supplying informative, meanwhile non-intrusive messages for guiding users across spaces otherwise difficult to navigate.

Future experiments will aim at finding firm conclusions along this line, through the development of alternative hypotheses and employing a larger number of subjects.

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.

REFERENCES

1. Bresin, R., de Witt, A., Papetti, S., Civolani, M., and Fontana, F. Expressive sonification of footstep sounds. In *Proc. of the Interaction Sonification workshop (ISon) 2010*, R. Bresin, T. Hermann, and A. Hunt, Eds. (KTH, Stockholm, Sweden, 2010).
2. Civolani, M., Fontana, F., and Papetti, S. Efficient acquisition of force data in interactive shoe designs. In *Proc. 5th Int. Workshop on Haptic and Audio Interaction Design (HAID '10)*, Springer (Copenhagen, Denmark, 2010).
3. Courtine, G., Pozzo, T., Lucas, B., and Schieppati, M. Continuous, bilateral achilles' tendon vibration is not detrimental to human walk. *Brain Research Bulletin 55*, 1 (2001), 107 – 115.
4. Howe, T. E., Lövgreen, B., Cody, F. W. J., Ashton, V. J., and Oldham, J. A. Auditory cues can modify the gait of persons with early-stage parkinson's disease: a method for enhancing parkinsonian walking performance? *Clin Rehabil 17*, 4 (2003), 363–7.
5. Nurse, M. A., and Nigg, B. M. The effect of changes in foot sensation on plantar pressure and muscle activity. *Clinical Biomechanics 16*, 9 (2001), 719 – 727.
6. Papetti, S., Civolani, M., and Fontana, F. Rhythm'n'Shoes: a wearable foot tapping interface with audio-tactile feedback. In *Proc. Conf. on New Interfaces for Musical Expression (NIME)* (2011).
7. Pastore, R. E., Flint, J. D., Gaston, J. R., and Solomon, M. J. Auditory event perception: The source-perception loop for posture in human gait. *Perception & Psychophysics 70*, 1 (2008), 13–29.
8. Pelah, A., and Koenderink, J. J. Editorial: Walking in real and virtual environments. *ACM Trans. Appl. Percept. 4*, 1 (2007), 1.
9. Prassas, S., Thaut, M., McIntosh, G., and Rice, R. Effect of auditory rhythmic cuing on gait kinematic parameters of stroke patients. *Gait & Posture 6*, 3 (1997), 218 – 223.
10. Priplata, A. A., Niemi, J. B., Harry, J. D., Lipsitz, L. A., and Collins, J. J. Vibrating insoles and balance control in elderly people. *Lance 362*, 9390 (2003), 1123–1124.
11. Queralt, A., Weerdesteyn, V., Van Duijnhoven, H. J. R., Castellote, J. M., Valls-Sol, J., and Duysens, J. The effects of an auditory startle on obstacle avoidance during walking. *The Journal of Physiology 586*, Pt 18 (2008), 4453–4463.
12. Razzaque, S., Kohn, Z., and Whitton, M. C. Redirected walking. In *Proc. SIGGRAPH/Eurographics* (2001).
13. Suteerawattananon, M., Morris, G. S., Etnyre, B. R., Jankovic, J., and Protas, E. J. Effects of visual and auditory cues on gait in individuals with parkinson's disease. *Journal of the Neurological Sciences 219*, 1-2 (2004), 63 – 69.
14. Verschueren, S. M. P., Swinnen, S. P., Desloovere, K., and Duysens, J. Vibration-induced changes in emg during human locomotion. *Journal of Neurophysiology 89*, 3 (2003), 1299–1307.
15. Visell, Y., Fontana, F., Giordano, B., Nordahl, R., Serafin, S., and Bresin, R. Sound design and perception in walking interactions. *International Journal of Human-Computer Studies 67*, 11 (2009), 947 – 959. Special issue on Sonic Interaction Design (SID).
16. Visell, Y., Giordano, B. L., Millet, G., and Cooperstock, J. R. Vibration influences haptic perception of surface compliance during walking. *PLoS ONE 6*, 3 (03 2011), e17697.

AUDITORY RECOGNITION OF FLOOR SURFACES BY TEMPORAL AND SPECTRAL CUES OF WALKING

Federico Fontana

University of Udine
Dipartimento di Matematica e Informatica
via delle Scienze 208
33100 Udine, Italy
federico.fontana@uniud.it

Fabio Morreale

University of Trento
Dipartimento di Ingegneria
e Scienza dell'Informazione
via Sommarive 14
38123 Povo (TN), Italy
fabio.morreale@disi.unitn.it

Tony Regia-Corte, Anatole Lécuyer, Maud Marchal

INRIA
Campus Universitaire de Beaulieu
35042 Rennes, France
tony.regia-corte@wanadoo.fr
{Anatole.Lecuyer,Maud.Marchal}@inria.fr

ABSTRACT

In a multiple choice auditory experimental task, listeners had to discriminate walks over floors made of concrete, wood, gravel, or dried twigs. Sound stimuli were obtained by mixing temporal and spectral signal components, resulting in hybrid formulations of such materials. In this way we analyzed the saliency of the corresponding cues of time and frequency in the recognition of a specific floor. Results show that listeners differently weigh such cues during recognition, however this tendency is not polarized enough to enable interaction designers to reduce the functionality of a walking sound synthesizer to simple operations made on the temporal or spectral domain depending on the simulated material.

1. INTRODUCTION

Walking sounds play a significant role in everyday listening environments. They in fact convey information at different levels of approaching walkers, and about the surface they traverse. By listening to a temporal sequence of footsteps humans try to identify surface material [1], meanwhile they make inferences on shoe types and walking styles to support decisions on gender as well as about physical, biomechanical, and affective characters of a person [2, 3, 4].

For their importance in the auditory scene, walking sounds have found place in multimodal displays since the early days of interactive simulation. All vintage electronic game players probably remember the iconic use of footstep sounds to render the number and moving speed of the enemies in Space Invaders™, a popular computer game of the late 70's. Nowadays, with increasing computational and memory resources as well as quality of the pc audio hardware, sound designers can make use of rich collections of accurate walking sounds recorded under different conditions—for instance see the related section in sounddogs.com. Furthermore, techniques exist for the interactive synthesis of footstep sounds [5, 6, 7].

Walking sounds lose ecological coherence if they are inaccurately displayed in an interactive context, in terms of audio quality or temporal synchronization. When the feedback message can be completely determined some hundreds of milliseconds before displaying, or when a comparable latency of the response to an input is tolerated, then accurate pre-recorded footstep sounds can be selected from an even huge sample database, and post-processed before reproduction also by devices whose memory and computation power are relatively smaller, such as mobile platforms. Conversely, there are situations in which no more than few tens of milliseconds are allowed to a computer system for delivering interactive sounds. In the most compelling case, feedback must be provided under continuously varying control conditions. Consider, for instance, a virtual environment simulating ground materials having spatially-varying degrees of compliance, such as gravel areas of different average composition or mud of variable thickness covering a solid floor base: In this case, physically based or physically informed synthesis models such as those mentioned before become unavoidable.

Techniques for the rendering of continuous interactions between objects have reached amazing results in the computer graphics field, thanks to efforts in basic, applied, and technological research leading in particular to recent powerful GPU cores. The audio field, however needing similar research effort and dedicated resources to achieve goals of comparable impact, has come to relatively less resounding results and technologies for its ancillary reputation with respect to vision. Furthermore, computer graphics researchers and developers can rely on the ability of a user to focus

MATERIAL	PHYSICAL PROPERTIES	ACOUSTIC PROPERTIES
C , W	Solid	Spectral Cues
G , T	Aggregate	Temporal Cues

Figure 1: *Experimental hypothesis. (C: concrete, W: wood, G: gravel, T: twigs.)*

on a selected part of the visual field. As opposed to this ability, listeners integrate all simultaneous events of an auditory stream into a single temporal process, whose overall control is known to be almost impossible to achieve in everyday listening environments and ecological soundscapes powered using consumer audio reproduction hardware.

Researchers in interactive synthesis look for models providing accurate sounds and direct access to their distinctive parameters, meanwhile parsimonious in terms of needed computational and memory resources. Successful modeling is usually based on applicable psychoacoustic results, helping sound interaction designers to get rid of redundancies otherwise burdening the resulting synthesis engine. In our specific case, we aim at interactively synthesizing realistic walking sounds by means of a simple model, whose computations and continuous control are at reach of current consumer hardware. Our basic hypothesis, elaborated in the next section, is that perceived material properties in walking sounds map to temporal or spectral auditory cues proportionally to the nature and, hence, category of the floor material.

2. EXPERIMENTAL HYPOTHESIS

Footstep sounds range across a number of possibilities depending on the shoe type, nature of the floor, and foot action. We restrict the pool of possible interactions to those generated by a male, walking with normal style upon a flat floor using leather shoes. The only experimental variable then will be the floor material.

We categorize floors into *solid* or *aggregate*. The former, such as concrete, marble, wood, are stiff. The latter, such as gravel, dry leaves, sand, allow relative motion of their constituent units and progressively adapt to the sole profile during the interaction. Now,

- solid materials give rise to short, repeatable impacts having a definite spectral color;
- aggregate materials elicit sequences of tiny impacts having distinctive temporal density, that create a sort of "crumpling", less resonant sound.

We hypothesize that the perception of solid materials is mainly determined by *spectral* cues, conversely the perception of aggregate materials is mainly determined by *temporal* cues. In particular, we experiment using concrete (C) and wooden (W) floors, representative of solid materials, as well as with gravel (G) and dried twigs (T), representative of aggregate materials. Figure 1 illustrates the hypothesis.

3. METHOD

A non-interactive setup was used in the experiment. Although forcing the subjects to perform a passive task, this choice was made to keep control on the stimuli and, more in general, focus on the auditory modality. In particular, by avoiding interactive sonic augmentations in the context of a multimodal walking task, subjects could just listen to footstep sounds without incurring in potential bias, determined by self-locomotion over a room floor having completely different visual and tactile attributes.

3.1. Setup

The experiment was set up in the VIPS laboratory at the Department of Computer Science, University of Verona, Italy. Subjects were sitting in front of a Mac Pro pc running a Java application communicating (via the *pdj* library) with Pure Data, a free software environment for real time audio synthesis also enabling simple visualizations (through the *GEM* library). They listened to the auditory stimuli through a pair of AKG K240 headphones.

3.2. Participants

Thirteen male and three female undergraduate computer science students aged aged 22 to 31 (mean = 24.62, std = 2.55) participated in the experiment. Few of them had some experience in sound processing. All of them reported to usually wear snickers.

At the end of the experiment, every subject completed a subjective questionnaire about the realism and ease of identification of the audio stimuli.

3.3. Procedure

One footstep by a normally walking male wearing leather shoes was repeatedly recorded while he stepped over a tray filled with gravel and, then, dried twigs. Recordings were made inside a silent, normally reverberant room using a Zoom H2 digital hand recorder standing 0.5 m far from the tray. For either material, seven recordings were selected and randomly enqueued to create walking sequences lasting 12 s and containing 13 footsteps. In addition to the in-house recordings, high quality samples of a male walking on concrete and on wooden parquet were downloaded from the commercial database `sounddogs.com`. Using these samples, two further walking sequences were created having the same beat and average Sound Pressure Level as of those based on in-house recordings.

Temporal envelopes were extracted from every sequence, by computing the signal

$$e_M[n] = (1 - b[n])|s_M[n]| + b[n]e_M[n-1] \qquad (1)$$

out of the corresponding sequence s_M, $M \in \mathcal{M} = \{C, W, G, T\}$. (Refer to Figure 1 for the meaning of the C, W, G, and T.) As in previous research on synthetic footsteps, the envelope following parameter $b[n]$ was set to 0.8 when $|s_M[n]| > e_M[n-1]$, and to 0.998 otherwise [5]. By following the input when its magnitude is greater than the envelope, and by in parallel allowing a comparably slow decay of the envelope itself when the same magnitude is smaller, this setting ensures that amplitude peaks are tracked accurately, while leaving spurious peaking components off the envelope signal e_M.

By dividing every sequence s_M by its envelope e_M, we computed signals $u_M = s_M/e_M$ in which the temporal dynamics was

removed. In other words, we manipulated the footstep sequences so to have stationary amplitude along time.

What remained in u_M was a spectral color, that we extracted with a 48th-order inverse LPC filter h_M^{-1} estimated in correspondence of those parts of the signals containing footstep sounds. Using this filter order, if training the model using *one* footstep then we could not detect differences between the original sound and the correspondingly resynthesized footstep. We emphasize that the resulting LPC filter in any case estimated one single transfer function, independently of the number of footsteps taken from the original sequence which informed the model. Since we trained the estimator with the entire sequence, the resynthesized sound had a slightly different color compared to any other footstep belonging to the original sequence.

In the end, for every material M a highly realistic version $\tilde{s}_M$ of the original sequence s_M could be resynthesized by convolving digital white noise w by the "coloring" filter h_M, and then multiplying its output, i.e. the synthetic version $\tilde{u}_M$ of u_M, by the envelope signal e_M:

$$\tilde{s}_M[n] = (w * h_M)[n] \cdot e_M[n] = \tilde{u}_M[n] \cdot e_M[n]. \quad (2)$$

This technique draws ideas from a family of physically-informed models of walking sounds [5, 7]. In the meantime it provides a simpler, more controlled resynthesis process avoiding stochastic generation of patterns as in such models. In our case, the silent parts of the four envelopes were tailored to generate synthetic sequences having identical walking tempos. This simple manipulation ensured seamless mutual exchange of the envelopes among sequences, as explained in the following.

Sixteen stimuli were created by adding twelve *hybrid* resyntheses to the *native* stimuli $\tilde{s}_C$, $\tilde{s}_W$, $\tilde{s}_G$, and $\tilde{s}_T$. Every hybrid stimulus $\tilde{s}_{M_t,M_f}$, $M_t, M_f \in \mathcal{M}$ was defined as to account for the spectral color of material M_f *and* the temporal envelope of material $M_t \neq M_f$:

$$\tilde{s}_{M_t,M_f}[n] = (w * h_{M_f})[n] \cdot e_{M_t}[n] = \tilde{u}_{M_f}[n] \cdot e_{M_t}[n]. \quad (3)$$

For each material M_f, we checked that all hybrid temporal manipulations using $M_t \neq M_f$ did not notably alter the spectral information of $\tilde{s}_{M_f}$, and thus its original color. In fact, an inspection of the spectra $E_M(\omega)$ of the various envelopes made by Fourier-transforming e_M, i.e., $E_M(\omega) = \mathcal{F}\{e_M\}(\omega)$, shows that they all have a comparable spectrum. More precisely, all spectra E_C, E_W, E_G, E_T exhibit similar magnitudes, that are shown in Figure 2 after removing the respective dc component for ease of inspection. This means that the spectral differences in $\tilde{S}_{M_t,M_f}(\omega)$ caused by multiplying $\tilde{u}_{M_f}$ by e_{M_t}, that is,

$$\tilde{S}_{M_t,M_f}(\omega) = \mathcal{F}\{\tilde{u}_{M_f} \cdot e_{M_t}\}(\omega) = (\tilde{U}_{M_f} * E_{M_t})(\omega), \quad (4)$$

are substantially independent of the material, hence almost identical to those introduced in $\tilde{s}_{M_f}$ by its own envelope e_{M_f}.

Symmetrically, the temporal artifacts which are caused by hybridization between two different materials can be considered minor. In fact, because of the LPC design methodology, all filters h_C, h_W, h_G, h_T do transform white noise into a stationary signal independently of the material.

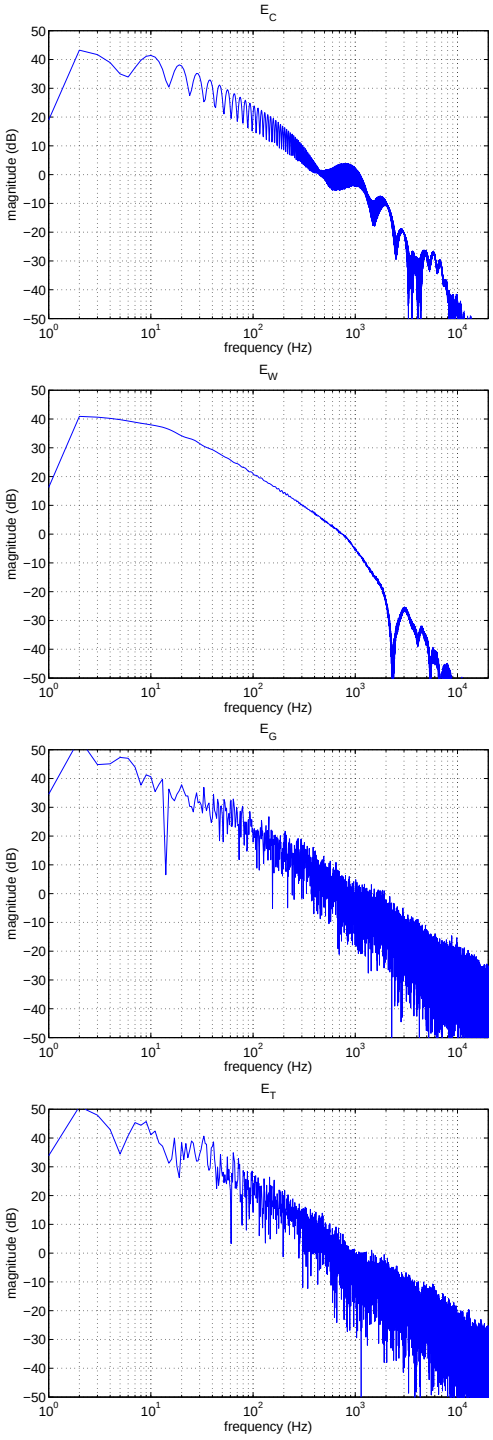


Figure 2: *Magnitude spectra of envelopes E_C, E_W, E_G, and E_T. Respective dc components removed for ease of inspection.*

3.4. Protocol

To become confident with the auditory stimuli and the graphic interface used in the experiment subjects trained for some minutes before starting an individual session, by selecting and playing each one of the original sequences s_C, s_W, s_G, s_T for several times. Figure 3 shows the four graphic buttons exposed by the interface, enabling subjects to play the corresponding original sequences during training.

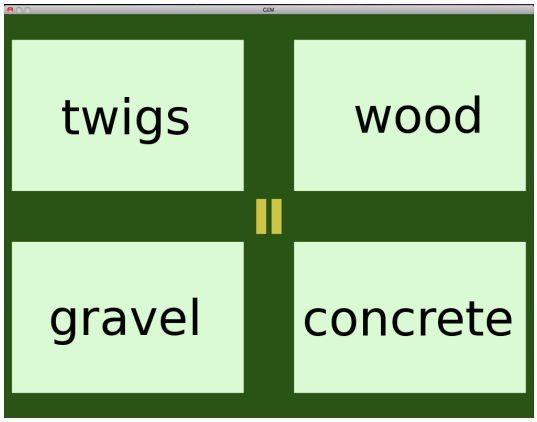


Figure 3: *Translation in English of the graphical interface used in the experiment.*

Each individual session consisted of 192 trials, obtained by randomly playing each one of the sixteen synthetic stimuli for twelve times. At each trial the subject listened to a stimulus, and selected one material by clicking the related button in the interface. When the button was released, the screen froze for two seconds and changed color to inform subjects of the conclusion of the trial. After this short pause, a new trial was performed.

The four buttons randomly switched position at each trial. Subjects could temporarily stop the experiment by clicking the pause icon '||' located in the middle of the screen, whenever they wanted to take a short break among trials. It took approximately 45 minutes for each participant to complete the session.

4. RESULTS

For each participant, percentages of selection for the four materials C, W, G, T were analyzed. We considered the fraction of participants who showed significantly correct (random is 25%) material recognitions from the four synthetic stimuli $\tilde{s}_C$, $\tilde{s}_W$, $\tilde{s}_G$, $\tilde{s}_T$, across the twelve repetitions. The critical value (with $\alpha = 0.05$) of the one-tailed binomial test Bin(12,0.25) is 7 trials (i.e., 58.33%): only the participants with an auditory recognition of the original materials higher than 58.33% were included in the analysis. After this check, 16 participants were considered for the recognition of dried twigs, 15 for gravel, 16 for wood, and 10 for concrete.

The results of the analysis are presented in Figure 4. In these plots, a bar exhibiting a low percentage means that the correspondingly substituted information (either temporal or spatial) is important for the recognition of the original material, represented by the leftmost bar in the same plot. The difference from random percentage (25%) was tested using one-proportion (two-tailed) z tests.

By aggregating the data, we also evaluated the auditory recognition of material categories. Again, this analysis was conducted using data from participants exhibiting an auditory recognition significantly higher than random concerning the two sets of stimuli accounting for the respective categories (24 trials for each category). In this case, the critical value (this time computed by a one-proportion/one-tailed z test to account for the larger number of trials, with $\alpha = 0.05$) is 10 trials, corresponding to 41.67%. All the participants passed the check.

The results of the new analysis are presented in Figure 5. For

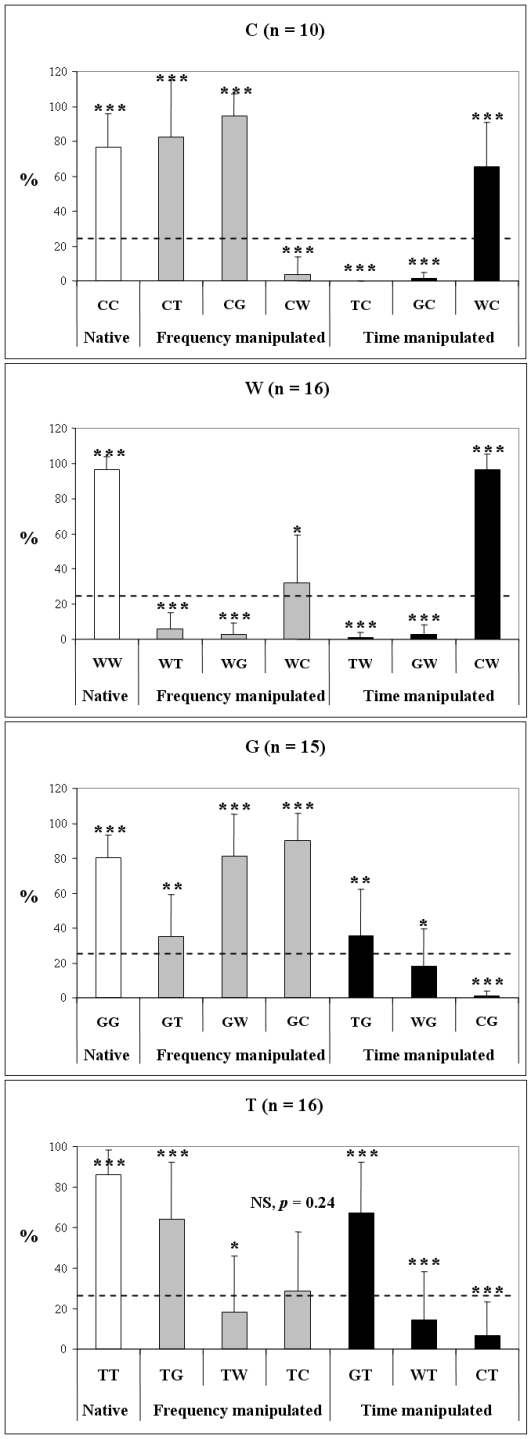


Figure 4: *Mean percentages of selection (lines represent std) for C,W,G,T as a fu nction of the auditory stimulus $\tilde{s}_{M_t,M_f}$. The difference from random selection (line at 25%) was tested using one-proportion (two-tailed) z tests. Note: $*: p < 0.05$, $**: p < 0.01$, $***: p < 0.001$, NS: not significant, n: number of subjects.*

the different percentages of selection, the difference relative to ran-

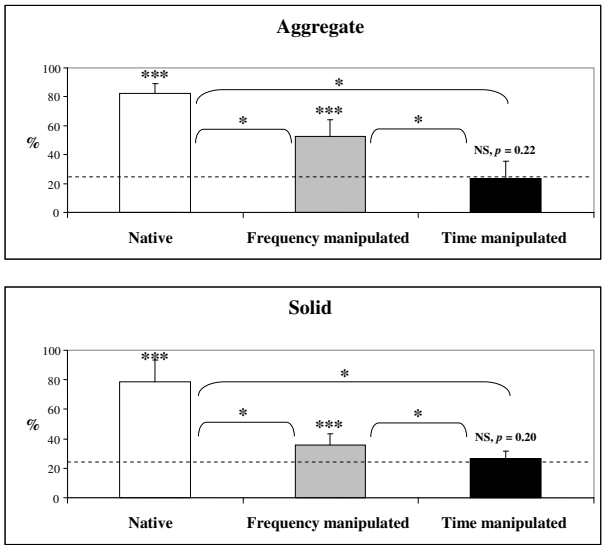


Figure 5: *Mean percentages of selection (lines represent std) for material categories (Aggregate and Solid) as a function of the auditory stimulus $\tilde{s}_{M_t,M_f}$. The difference from random selection (line at 25%) was tested using one-proportion (two-tailed) z tests. The differences between the three audio conditions were tested with two-proportion z tests (two-tailed and Bonferroni-adjusted alpha level with $p = 0.05/3 = 0.0167$). Note: $*: p < 0.05$, $**: p < 0.01$, $***: p < 0.001$, NS: not significant.*

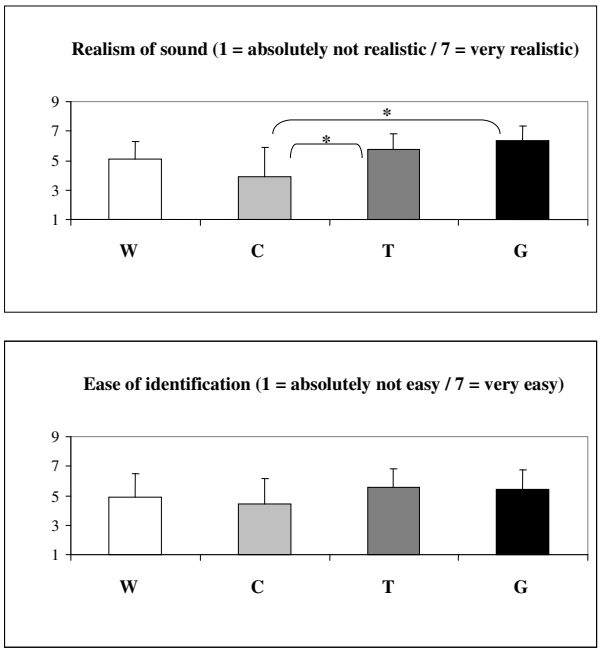


Figure 6: *Mean and standard deviation (lines represent std) of subjective ratings about "realism of sound" and "ease of identification" for the four materials C, W, G, T. Differences between materials were tested with two-proportion z tests (two-tailed and Bonferroni-adjusted alpha level with $p = 0.05/6 = 0.0083$). Only the significant differences are presented, $*: p < 0.05$.*

dom (25%) was tested using one-proportion (two-tailed) z tests. Thus, for the aggregate category, the percentages of selection in native (82.29%) and frequency manipulated (52.78%) conditions were significantly different from random ($z = 25.98$, $p < 0.001$ and $z = 21.77$, $p < 0.001$, respectively). By contrast, the percentage of selection in the time manipulated condition (23.44%) was not significantly different from random ($z = -1.22$, $p = 0.22$). On the other hand, for the solid category, the percentages of selection in native (78.39%) and frequency manipulated (35.59%) conditions were significantly different from random ($z = 24.16$, $p < 0.001$ and $z = 8.30$, $p < 0.001$, respectively). By contrast, the percentage of selection in the time manipulated condition (26.65%) was not significantly different from random ($z = 1.29$, $p = 0.20$). The differences between the three audio conditions were tested with two-proportion (two-tailed) z tests.

A correction for experiment-wise error was realized by using Bonferroni-adjusted alpha level (p divided by the number of tests). Thus, in order to compare the three audio conditions (native, frequency manipulated, and time manipulated), the alpha level was adjusted to $p = 0.05/3 = 0.0167$. For the aggregate category, the analysis showed that the native condition was significantly different from the frequency manipulated ($z = 10.23$, $p < 0.05$) and time manipulated ($z = 20.56$, $p < 0.05$) conditions. The difference between frequency manipulated and time manipulated conditions was significantly different ($z = 14.50$, $p < 0.05$) as well. For the solid category, the analysis indicated that the native condition was significantly different from the frequency manipulated ($z = 14.57$, $p < 0.05$) and time manipulated ($z = 17.96$, $p < 0.05$) conditions. The difference between frequency manipulated and time manipulated conditions was also significantly different ($z = 4.63$, $p < 0.05$).

After the experiment, a questionnaire was proposed in which each participant had to grade from 1 to 7 the four native stimuli according to two subjective criteria: realism, and ease of identification. Figure 6 shows the means and standard deviations of the four native stimuli for each of the subjective criteria. Wilcoxon signed rank (two-tailed) tests with Bonferroni correction showed significant differences only for the realism of sounds: between concrete and dried twigs ($z = -3.28$, $p = 0.001$), and between concrete and gravel ($z = -3.16$, $p = 0.0016$).

5. DISCUSSION

The histograms for concrete and wood in Figure 4 show that subjects tolerate swapping between the temporal features of C and W, both belonging to the solid category, conversely the substitution in the same signals with temporal features extracted from aggregate materials (i.e. G and T) harms the recognition. This result is in favor of the initial hypothesis. The effect of spectral manipulations of C and W is more articulate. In this case the hypothesis is essentially confirmed with wood, whose distinctive color cannot be changed using any other spectrum. In parallel, subjects are tolerant to substitutions in C with spectra from aggregate materials. This greater tolerance may be due to the basic lack of distinct color of concrete floors, especially for listeners who usually wear rubber sole shoes such as snickers (indeed the majority of our sample). The same conclusion finds partial confirmation by Figure 6 which, in the limits of the significance of its data, shows greater

confidence by subjects in recognizing aggregate materials.

The histograms in Figure 4 regarding gravel and twigs partially support the initial hypothesis. Time swaps between G and T are tolerated to a lesser extent compared to solid floors. Like before, substituting the temporal features of solid materials in an aggregate sound is not tolerated. Spectral substitutions are not as destructive as they were for solid materials, especially in the case of gravel. The worst situation is when the spectrum of W is substituted in T, again probably due to the distinct color that wood resonances bring into the sound.

Figure 5 would further support this discussion. In fact, in spite of the low significance of the data from time manipulations (i.e. black bars), it shows that subjects are primarily sensitive to temporal substitutions between solid and aggregate materials. In parallel, spectral changes are more tolerated during the recognition of aggregate material compared to solid floors.

6. CONCLUSIONS

The proposed experiment has confirmed that solid and aggregate floor materials exhibit precise temporal features, that cannot be interchanged while designing accurate walking sounds. Within such respective categories, spectral color represents an important cue for the recognition of solid materials, conversely sounds of aggregate materials seem to tolerate larger artefacts in their spectra.

Current interfaces operating at ground level, capable of conveying augmented sensations of material properties over otherwise neutral floors, exist in form of active tiles [8] and instrumented shoes [9]. Such interfaces sense the force exerted by a walker over the ground, and instantaneously react by providing auditory and vibrotactile feedback to her or his feet through the action of specific actuators.

The proposed model has been conceived as an alternative to existing real-time physics-based synthesizers onboard those interfaces. Once connected to the physical components realizing their input and output stages, it could easily generate interactive walking sounds by controlling at runtime, through the force sensors and depending on the simulated material, the attack time and shape of the temporal envelope signals, and consequently the amplitude of the LPC-filtered noise forming the auditory and, possibly, also the vibrotactile feedback.

7. 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.

8. REFERENCES

[1] B. L. Giordano, S. McAdams, Y. Visell, J. R. Cooperstock, H. Yao, and V. Hayward, "Non-visual identification of walking grounds," in *Proc. of Acoustics'08 in J. Acoust. Soc. Am.*, vol. 123 (5), 2008, p. 3412.

[2] X. Li, R. J. Logan, and R. E. Pastore, "Perception of acoustic source characteristics: Walking sounds," *Journal of the Acoustical Society of America*, vol. 90, no. 6, pp. 3036–3049, 1991.

[3] Y. Visell, F. Fontana, B. Giordano, R. Nordahl, S. Serafin, and R. Bresin, "Sound design and perception in walking interactions," *Int. J. Human-Computer Studies*, no. 67, pp. 947–959, 2009.

[4] R. Bresin, A. de Witt, S. Papetti, M. Civolani, and F. Fontana, "Expressive sonification of footstep sounds," in *Proc. of the Interaction Sonification workshop (ISon) 2010*, R. Bresin, T. Hermann, and A. Hunt, Eds., KTH, Stockholm, Sweden, Apr. 7 2010.

[5] P. R. Cook, "Modeling Bill's gait: Analysis and parametric synthesis of walking sounds," in *Proc. Audio Engineering Society 22 Conference on Virtual, Synthetic and Entertainment Audio*. Espoo, Finland: AES, Jul. 2002.

[6] A. J. Farnell, "Marching onwards – procedural synthetic footsteps for video games and animation," in *pd Convention*, 2007.

[7] L. Turchet, S. Serafin, and R. Nordahl, "Physically based sound synthesis and control of footsteps sounds," in *Proc. Conf. on Digital Audio Effects (DAFX-10)*, Graz, Austria, Sep. 6-10 2010.

[8] Y. Visell, J. Cooperstock, B. L. Giordano, K. Franinovic, A. Law, S. McAdams, K. Jathal, and F. Fontana, "A vibrotactile device for display of virtual ground materials in walking," in *Proc. of Eurohaptics 2008*, 2008.

[9] S. Papetti, M. Civolani, F. Fontana, A. Berrezag, and V. Hayward, "Audio-tactile display of ground properties using interactive shoes," in *Proc. 5th Int. Haptic and Auditory Interaction Design Workshop*, Sep. 2010, pp. 117–128.

Rhythm'n'shoes: a wearable foot tapping interface with audio-tactile feedback

Stefano Papetti
Dept. of Computer Science,
University of Verona
Strada Le Grazie, 15
37134 Verona, Italy
stefano.papetti@univr.it

Marco Civolani
Dept. of Mathematics and
Computer Science,
University of Udine
Via delle Scienze, 206
33100 Udine, Italy
marco.civolani@uniud.it

Federico Fontana
Dept. of Mathematics and
Computer Science,
University of Udine
Via delle Scienze, 206
33100 Udine, Italy
federico.fontana@uniud.it

ABSTRACT

A wearable shoe-based interface is presented, which enables users to play percussive virtual instruments by tapping with their feet. The interface consists of a pair of sandals equipped with four force sensors and actuators affording audio-tactile feedback. The sensors provide data via wireless transmission to a host computer, where they are processed and mapped to a physically-based sound synthesis engine. Since the system provides OSC and MIDI compatibility, alternative electronic instruments can be used as well. The audio signals are then sent back wirelessly to the audio-tactile exciters embedded in the sandals' sole, and optionally to headphones and additional external loudspeakers. The round-trip wireless communication only introduces very small latency (about 10 ms), thus allowing tight timing while playing.

Keywords

interface, footwear, wireless, wearable, mobile, foot tapping, sensors, multimodal feedback, audio, tactile, embodied

1. INTRODUCTION

In many cultures, both music and dance performers make use of foot tapping, from folk fiddlers and street buskers to flamenco and tap dancers. For instance, a fiddler stomping on a pub's wooden floor can cheer on the audience meanwhile supporting his or her own playing by adding a simple percussion part; buskers often include foot drums in their setup to add even complex percussion parts to their guitar playing and singing. Moreover, traditional musical genres exist (e.g., from Ireland, Canada and USA) where players make extensive use of foot percussions (*podo-rhythm*) as main accompaniment. As for dance, foot tapping can have both an expressive and rhythmic function, to the extent that some dance genres are centered on both the musical and gestural performance produced by the dancer's feet.

On the other hand, in everyday life many musicians and music enthusiasts alike find themselves "tapping songs" with their fingers, hands and feet. In particular, the feet are usually available for tapping even while working at one's desk, or reading a newspaper while traveling on a train (co-workers and other passengers permitting). Such tapping may represent the song's main melody, rhythm, or even accurately simulate its percussions part.

The gesture of playing rhythms with the feet offers spontaneity and expressivity, at the same time enabling an embodied experience. In parallel, as musical ability grows, foot tapping becomes a fundamental resource for musicians to stay within tempos and generate rhythms in the case of drumming or percussionism.

Taking inspiration from these observations, and starting from a prototype shoe-based interface we had previously realized for interactive walking purposes [13], we decided to implement a wearable musical interface for foot tapping. We have then developed a mobile shoe-based musical interface called "Rhythm'n'Shoes", that allows users to control percussive sounds generated by potentially any electronic instruments: the system is natively interfaced with a physically-based sound synthesis engine, however it is suitable for use with any instrument compatible with MIDI or OSC protocols. The interface provides the user with audio-tactile feedback through exciters embedded in the sole, and optionally audio-only feedback through headphones and additional external loudspeakers.

A similarly immediate approach to playing rhythms, that avoid the use of virtual drum interfaces (like the Roland V-Drum and Clavia ddrum), but instead takes inspiration from the common experience of hitting the chest or thighs with the hands, is described in [3]. The interface consists of a pair of gloves embedding piezo microphones, that are used as sensing devices.

Several recent studies take into account novel percussion instruments [2], interfaces for percussion tasks [9], and related technological solutions. Some examples are: a physical modeling drum synthesizer and interface [7]; a wireless drumstick interface [15]. Finally, a renowned commercial electronic percussion interface and instrument is the Korg Wavedrum.

Haptic feedback is exploited in [10] to foster rhythm skills while playing a virtual drum kit. In [5] a physically intuitive haptic drumstick is discussed, where the haptic dynamics are altered to assist performers in playing single-handed drum rolls. In a more general context, several researches exist which consider the use of haptic feedback in digital musical interfaces and instruments [6, 12, 11].

2. INTERFACE DESIGN

This section describes the design of the interface in a bottom-up fashion, from the hardware implementation to the software level.

2.1 Hardware

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
NIME'11, 30 May–1 June 2011, Oslo, Norway.
Copyright remains with the author(s).

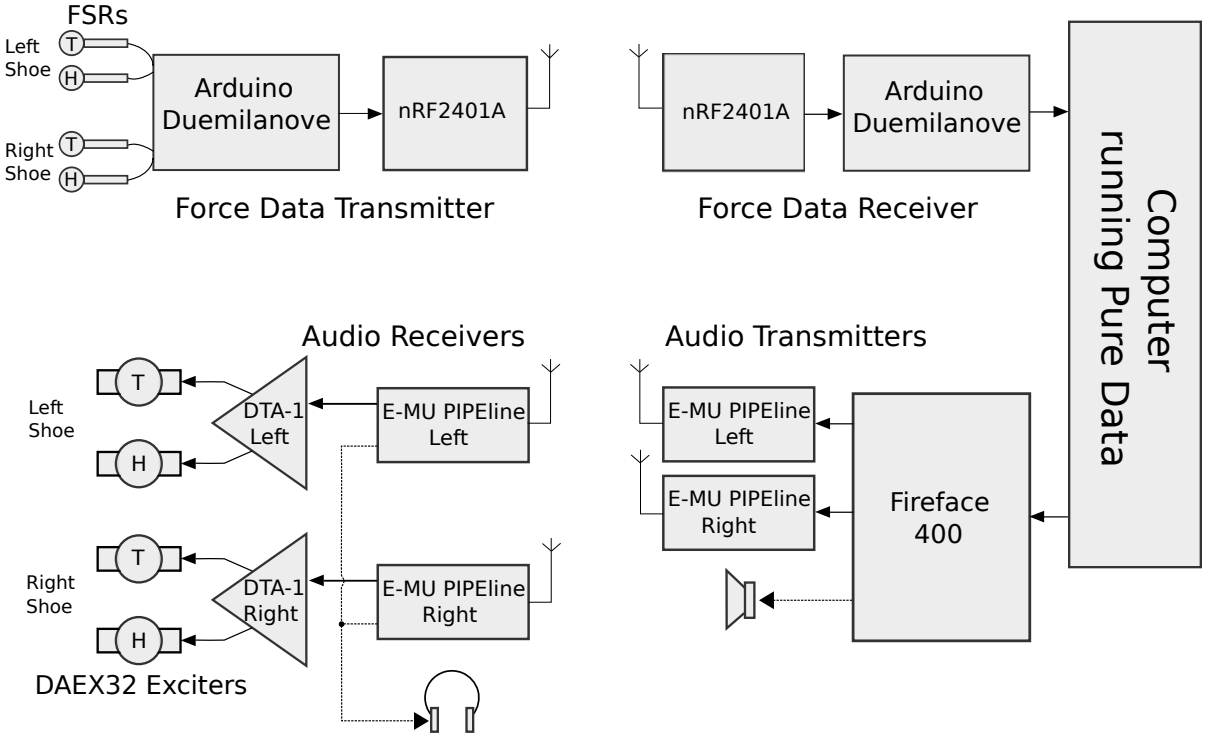


Figure 1: Block diagram representing the hardware setup.

The hardware setup is depicted in Fig. 1. The user wears a pair of sandals each one with a couple of Force Sensing Resistors (Interlink FSR model 402) fixed under the insole: one at the toe and one at the heel. The sensors are connected to the first four analog inputs of the transmitter Arduino board via a voltage divider; force data are sampled using a polling procedure and encapsulated in a two bytes-per-channel custom protocol [8]. The packets are sent over a single-wire serial line to a 2.4 GHz wireless transceiver module[1] based on the nRF2401A chip by Nordic Semiconductor. At the other end, another nRF2401A module receives the data stream realizing a simplex wireless line and routes it to the receiver Arduino which in turns communicates to a personal computer via a USB connection. The computer runs an instance of the Pure Data real-time environment which fetches and processes the stream producing the audio-tactile signals (see section 2.2). Since each shoe embeds two audio-tactile exciters, the software synthesizer must provide four discrete output channels: a multichannel audio interface is thus required. In our case the RME Fireface400 was chosen[2]. The four channels are grouped in two stereo pairs, one for each shoe: each pair is injected into an E-MU PIPEline wireless transceiver. Each E-MU on the transmitter side is coupled to another one on the receiver side, thus a four-audio-channels simplex wireless connection is obtained. The line-level output of each E-MU at the receiver end is injected into a 15 Watts Dayton Audio DTA-1 Class T digital amplifier which drives two Dayton Audio DAEX32 exciters fixed under each sandal, one at the toe and one at the heel, closing the interaction loop.

[1] `http://www.sparkfun.com/products/152`

[2] The number of output channels can be reduced to two if audio-only feedback is needed: each of the two channels could for instance, sonify the corresponding shoe. In this case the stereo signal originally generated for each shoe can be mixed to mono and routed into a two-channels bus. This allows for the use of a stereo audio interface like the ones mounted on laptops.

The whole hardware system has been designed keeping in mind the following main constraints:

1. Low latency;
2. Wearability;
3. Low-cost and easily-available components.

Previous works proved that despite its low cost, the Arduino Duemilanove board can become a high performance front-end for signal acquisition on personal computers [8]. Native AVR instructions together with interrupt calls can be used to write a very efficient firmware, configuring the microcontroller's ADC to sample up to six analog channels with 10 bits resolution each. Moreover, the serial-over-USB interface can be forced to send synchronous data continuously, thus obtaining uniform sampling at the host application level. A measurements campaign performed on a previous prototype [13] showed that the value of sampling frequency per single analog channel depends on the serial data rate and on the number of channels: with 115200bps, $F_s = 5882/channels$ Hz which corresponds to approximatively 1470 Hz per channel using four FSR's. In the aforementioned prototype a single Arduino was directly connected to the host computer via a USB cable and the audio-tactile feedback from the audio interface was provided via wired connection as well. After properly adjusting the audio buffer's size, the latency of the whole chain, from the Arduino's analog inputs to the analog outputs of the audio board was reduced to 15 ms.

Since the new setup has been conceived for live performers, it must provide freedom of movement, thus a wireless system represents a good choice. Once again, overall latency must be kept as low as possible and the interface needs to be wearable (i.e. lightweight and small). Several solutions have been proposed in the past, providing wireless acquisition of control signals for musicians and performers [**?**], [**?**]: almost all of them were based on custom hardware and were

quite expensive (the Kroonde from La Kitchen [**?**] for example, costed 1200 Euros and now is out of production). As previously said, in the present prototype the wireless transmission of force data is managed by two transceiver modules based on the nRF2401A chip: this unit can handle a half-duplex digital radio stream with bit rates up to 1Mbps[3]. In order to keep false detection and data packet loss rates as low as possible, the nRF2401A is able to encapsulate the incoming payload (here represented by the force data) into a 256 bits long packet which contains also a preamble and a CRC; thus, the total amount of data to be transmitted increases but still, the available band is well above the 115200bps required in order to obtain the same performance as the old, wired prototype. The aforementioned firmware has been consequently modified, for allowing the communication between the Arduino and the transceiver on both sides.

2.2 Software

At the software level there are three modules realized in Pure Data (Pd), which are organized in a bottom-up hierarchy:

1. The first layer implements conditioning and analysis on the data stream provided by the FSR sensors, in order to detect tapping events. As soon as a tapping event is detected, this module outputs a measure of the event's energy.

2. The second layer maps the detected events alternatively to MIDI or OSC messages, or directly to the parameters of a sound synthesis engine running in Pd.

3. The third layer implements a physically-based sound synthesis engine which models impact events, that is driven by the detected tapping events.

These three layers are described in detail here below.

2.2.1 Data conditioning and analysis

The force data received by the host computer through the Arduino receiver, is unpacked (TODO) this way obtaining four separate data streams corresponding to the four FSR sensors (left heel and toe, right heel and toe).

These data is then conditioned in order to... TODO up-only? thresholds? offset? mean?

The pre-conditioned force data is then analyzed in order to detect the onset of tapping events and their energy. To this end we made use of a Pd object called `bonk~` which outperforms the standard envelope following technique [14]: `bonk~` utilizes bounded-Q analysis to decompose the incoming signal into frequency bands and computes the power in each of them, then it looks for sharp edges in the spectral envelope of the signal. In this way `bonk~` allows to detect percussive events very accurately. As `bonk~` is meant to analyze audio signals, before being sent to it the preconditioned force data streams are converted into audio signals:

- the Arduino transmitter samples force signals with 10 bit resolution, therefore the received raw data are described by integer numbers in the range 0-1023. However, since audio signals in Pd are described as streams of 32-bit floating point numbers ranging from -1 to 1, the force data values are converted to comply with the range from 0 to 1.

- the force data streams are then oversampled (by using the Pd object `sig~`) from the original sample rate of 1470.5 Hz to Pd's audio sample rate.

[3]The data rate must be either 250kbps or 1Mpbs.

TODO parametri bonk

The object `bonk~` can also be trained to recognize signals based on template matching. In this way, one could potentially track different kinds of tapping events coming from the same input channel, and map them separately. This feature has not yet been exploited in the current version of the interface.

The output provided by `bonk~` consists in a measure of the energy of the detected event. In detail this is calculated as the sum of the square roots of the amplitudes of the frequency bands. The value is normalized so that 100 is an attack of amplitude of about 1, where 1 is also the maximum positive amplitude of an audio signal in Pd.

2.2.2 Mapping

The next step consists in converting the energy values of the detected tapping events in order to comply with the control parameters of different instruments. In particular, a threefold conversion path is implemented, complying with three distinct protocols:

MIDI The energy values are converted into integer values in the range 0-127 to comply with MIDI velocity values. As soon as a tapping event is detected, a MIDI "note on" message is generated and associated with such velocity value. A MIDI "note off" message is produced following each "note on" message, after a settable delay time. The MIDI "channel" and "note number" for each of the four data streams can be assigned to interact with any MIDI instrument, however the default configuration already offers a common drum setup according to the General MIDI standard.

OSC Since OSC-compatible instruments may accept different kind of messages, the user can customize the OSC messages as he/she wishes. The original range of energy values (0-100) can be converted to a range of choice using the Pd object `maxlib/scale`. For a start, similarly to the MIDI case, the default configuration already offers predefined messages for communicating the onset of tapping events and their energy. The latter is expressed as float values from 0 to 100, as originally provided by `bonk~`.

SDT The energy values are converted into physically-consistent velocity values in m/s that can be sent to a physically-based impact model.

The three mappings described above can be chosen alternatively via a switch implemented in Pd.

2.2.3 Sound synthesis

In order to provide a self-sufficient system, we have chosen to include a sound synthesis engine in our prototype, this way transforming the interface in a proper complete instrument.

The sound synthesis engine makes use of a physically-based impact model [4] which is part of a library for Max and Pd called Sound Design Toolkit (SDT).[4] The model simulates a mass (object 1) colliding with a resonator (object 2). The collision is represented by a nonlinear spring with dissipation, while the resonator is described by the modal synthesis paradigm [1]. The main control parameters give access to:

- the mass m (in Kg) of object 1;

- the resonating modes of object 2, namely their frequency $f_{0..n}$ (where n is the number of modes) in Hz, their decay time $t_{0..n}$ in s, and their gain $g_{0..n}$;

[4]Freely available at `http://www.soundobject.org/SDT`.

- the nonlinear spring, namely its nonlinearity exponent α and stiffness k in Kg/N$^{\alpha}$.

Each force data stream is mapped to a separate instance of the impact model, this way resulting in a different sound for each tapping position.

3. THE SYSTEM IN USE

The system is not limited to playing drums or other traditional percussions, but encourages the experimentation of further languages and aesthetics.

Since the system does not require any visual skill and thanks to the provided haptic feedback, it is perfectly suitable to both blind and hearing impaired persons. In particular, as alternative utilizations, the hearing impaired could exploit the interface as a personal stage monitoring system or just a way to feel the music through the feet.

energetic coupling

4. FUTURE DEVELOPMENT

5. CONCLUSIONS

As anticipated in the introduction, the act of playing rhythms with only the feet guarantees spontaneity and expressivity. Moreover, thanks to the audio-tactile feedback, the interface offers a truly embodied interaction even playing an electronic instrument. Additionally, the use of physically-based sound models for generating both the audio and tactile feedback, allows a physically consistent and realistic experience.

The described system is suitable for musical genres where players make use of foot percussions (*podo-rhythm*), as well as for modern performances, where one person can play more than one instrument at a time, or where enhanced control over other electronic instruments is required.

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.

7. REFERENCES

[1] J. M. Adrien. The missing link: Modal synthesis. In G. De Poli, A. Piccialli, and C. Roads, editors, *Representations of Musical Signals*, pages 269–297. MIT Press, Cambridge, MA, 1991.

[2] R. M. Aimi. New expressive percussion instruments. *M.S. Thesis. MIT*, 2002.

[3] M. S. Andresen, M. Bach, and K. R. Kristensen. The lapslapper: feel the beat. In *Proc. of the 5th Int. Conf. on Haptic and Audio Interaction Design (HAID '10)*, pages 160–168, Berlin, Heidelberg, 2010. Springer-Verlag.

[4] F. Avanzini, M. Rath, and D. Rocchesso. Physically-based audio rendering of contact. In *Proc. IEEE Int. Conf. on Multimedia and Expo (ICME '02)*, volume 2, pages 445–448, 2002.

[5] E. Berdahl, W. Verplank, J. O. Smith III, and G. Niemeyer. A physically intuitive haptic drumstick. In *Int. Computer Music Conference*, page 363–366, Copenhagen, Denmark, 2007.

[6] C. Chafe. Tactile Audio Feedback. In *Proc. Int. Computer Music Conf. (ICMC '93)*, Tokyo, Japan, 1993.

[7] K. Chuchacz, S. O'Modhrain, and R. Woods. Physical models and musical controllers: designing a novel electronic percussion instrument. In *Proceedings of the 7th Int. Conf. on New Interfaces for Musical Expression (NIME '07)*, pages 37–40. ACM, 2007.

[8] M. Civolani, F. Fontana, and S. Papetti. Efficient acquisition of force data in interactive shoe designs. In *Proc. 5th Int. Workshop on Haptic and Audio Interaction Design (HAID '10)*, Copenhagen, Denmark, 2010. Springer.

[9] M. Collicutt, C. Casciato, and M. M. Wanderley. From real to virtual: A comparison of input devices for percussion tasks. In *Proc. Int. Conf. on New Interfaces for Musical Expression (NIME '09)*, 2009.

[10] S. Holland, A. J. Bouwer, M. Dalgelish, and T. M. Hurtig. Feeling the beat where it counts: fostering multi-limb rhythm skills with the haptic drum kit. In *Proc. of the 4th Int. Conf. on on Tangible, embedded, and embodied interaction (TEI '10)*, pages 21–28, New York, NY, USA, 2010. ACM.

[11] M. T. Marshall and M. M. Wanderley. Vibrotactile feedback in digital musical instruments. In *Proc. of the 6th Int. Conf. on New Interfaces for Musical Expression (NIME '06)*, pages 226–229, Paris, France, France, 2006. IRCAM — Centre Pompidou.

[12] M. S. O'Modhrain. *Playing by feel: incorporating haptic feedback into computer-based musical instruments*. PhD thesis, Stanford, CA, USA, 2001. AAI3000074.

[13] S. Papetti, F. Fontana, M. Civolani, A. Berrezag, and V. Hayward. Audio-tactile display of ground properties using interactive shoes. In *Proc. 5th Int. Workshop on Haptic and Audio Interaction Design (HAID '10)*, Copenhagen, Denmark, 2010. Springer.

[14] M. Puckette, T. Apel, and D. Zicarelli. Real-time audio analysis tools for Pd and MSP. In *Proc. Int. Computer Music Conf. (ICMC '98)*, pages 109–112, Ann Arbor, Michigan, USA, 1998.

[15] D. Young and I. Fujinaga. Aobachi: A new interface for japanese drumming. In *Proc. Int. Conf. on New Interfaces for Musical Expression (NIME '04)*, 2004.

Numerical Methods for a Nonlinear Impact Model: A Comparative Study With Closed-Form Corrections

Stefano Papetti, Federico Avanzini, and Davide Rocchesso

***Abstract*—A physically based impact model—already known and exploited in the field of sound synthesis—is studied using both analytical tools and numerical simulations. It is shown that the Hamiltonian of a physical system composed of a mass impacting on a wall can be expressed analytically as a function of the mass velocity during contact. Moreover, an efficient and accurate approximation for the mass outbound velocity is presented, which allows to estimate the Hamiltonian at the end of the contact. Analytical results are then compared to numerical simulations obtained by discretizing the system with several numerical methods. It is shown that, for some regions of the parameter space, the trajectories of the discretized systems may significantly drift from the analytically derived curves. Two approaches, based on enforcing numerical energy consistency, are then proposed to improve the accuracy of numerical simulations.**

***Index Terms*—Energy conservation, impact modeling, nonlinear dynamical systems, numerical simulation, physics computing.**

I. Introduction

PHYSICAL models of impacts between objects are ubiquitous in many areas of science and engineering, including robotics [1], haptics [2], computer graphics [3], acoustics [4] and sound synthesis [5]. The phenomenologically plausible and energy-consistent behavior of contacting bodies is especially crucial in simulations of interactions based on sustained or repeated impacts, such as in rolling [6], scraping, or bouncing [7].

The higher the upper limit of the perceptual bandwidth is (and the rendering rate), the more critical the accuracy that real-time numerical simulations can afford; thus, making the problem of impact modeling increasingly complex when moving from graphic, to haptic, to auditory displays. Whereas for most graphic displays it is sufficient to describe an impact in terms of the ratio between outbound and inbound velocities, in haptic display and in sound synthesis the perceived characteristics of the impact depend on how bodies interact during contact. Therefore, more sophisticated impact models and carefully designed discretizations are necessary in audio and haptic contexts.

Manuscript received November 10, 2009; revised June 25, 2010 and November 12, 2010; accepted February 03, 2011. Date of publication February 22, 2011; date of current version July 22, 2011. This work was supported by the European Community's Seventh Framework Program under FET-Open grant agreement 222107 NIW-Natural Interactive Walking. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Richard C. Rose.

S. Papetti is with the Computer Science Department, University of Verona, 37134 Verona, Italy (e-mail: stefano.papetti@univr.it).

F. Avanzini is with the Department of Information Engineering, University of Padova, 35131 Padova, Italy (e-mail: avanzini@dei.unipd.it).

D. Rocchesso is with the Department of Art and Industrial Design, IUAV University of Venice, 30135 Venice, Italy (e-mail: roc@iuav.it).

Digital Object Identifier 10.1109/TASL.2011.2118204

A. Impact Models

The classic starting point is the Hertz model of collision between two spheres, which can be extended to include internal viscosity [8]. The impact force in such models is the sum of a nonlinear elastic term—in the form of a power law of compression—and a dissipative component proportional to the compression velocity—via a second power law of compression. For example, see [9] for the case of two colliding spheres.

In the context of musical acoustics, Stulov proposed a piano hammer model that includes relaxation properties of felt [10]. Other models exist that take plastic deformations into account, thus introducing abrupt direction changes in the force-compression curves at the transition between loading and unloading [11].

Particularly popular is the model by Hunt and Crossley [1], [12]–[15], that generalizes the extended Hertz model by considering a variable exponent that accounts for different contact shapes. In this model—described in detail in Section II—the power laws in the elastic and dissipative term are considered to be equal, thus allowing easier closed-form calculations [16]. Despite not being fully justified in physical terms, the Hunt–Crossley model has been quite successful in some areas of engineering because it allows to derive the phase trajectories in closed form, and because it is sufficiently complex to represent a wide variety of contact phenomena. In our work, we adopted this model and extended the range of the available analytical results.

B. Applications in Acoustic Modeling

Contact models can serve as a basis for developing models of acoustic phenomena. In the context of physically based sound synthesis, the Hunt–Crossley model has been used to develop an impact sound model [5], where a generic resonating object is used in place of a rigid wall.

Other models of more complex acoustic phenomena have been developed based on the very same impact model studied here. As an example, a *bouncing* sound model [7] has been obtained by superimposing a constant force, which simulates gravity, on a plain impact sound model. Also, a *rolling* sound model [6] has been implemented by driving an impact sound model by means of a physically inspired control layer. More precisely, the continuous interaction of a ball rolling on a surface has been modeled as a dense temporal sequence of micro-impacts driven by the geometry of the contacting surfaces, and modulated by the ball's asymmetry.

Accurate and consistent impact modeling is crucial in some audio-haptic rendering applications, such as interactive floors or shoes with vibratory and sonic augmentation [17].

1558-7916/$26.00 © 2011 IEEE

In the context of musical sound synthesis, the piano and other percussive musical instruments have also been modeled by using dissipative impact models [10].

C. Issues With Discrete Time

A wide range of numerical methods can be employed to discretize the interaction of impacting bodies. Given a reference continuous-time system, such as the Hunt–Crossley model, the goal is to obtain numerical quantities that follow the continuous-time trajectories as closely as possible, at an affordable computational cost. Thus, efficiency and accuracy are central issues.

Another important goal is energy consistency, especially in the case of repeated or sustained contacts. A numerical method, albeit being accurate, can introduce spurious oscillations or instabilities if it fails in terms of energy conservation. In this regard, the use of wave-digital structures (e.g., wave digital filters [18]) allows to preserve the passivity of certain continuous-time systems. Another approach is that of energy-based methods, i.e., numerical schemes based on the definition of a numerical energy that is proved to be conserved in lossless conditions [19]. These provably stable discrete-time models can be derived for some continuous-time models, including some nonlinear oscillators [20], but are not available for more general nonlinear contact models.

As for applications which make use of impact/contact models, energy inconsistencies are a recurring issue. In computer graphics, where the constraint of low frame rates makes numerical systems prone to instabilities [3], a typical example is provided by a steady object in resting contact with a rigid floor: when the system does not retain passivity, the object can move upward and bounce [21]. Similar issues are encountered in simulations of haptic contact, where stiffness values are usually limited by requirements on system passivity [22], [23], whereas higher values can cause the system to become unstable, for example oscillating, or reacting actively to the input. In numerical sound synthesis by physical models [24], artifacts and inconsistencies can become audible especially in situations of sustained or repeated contact interactions, as in rolling, sliding, or bouncing.

Instead of aiming at provably stable numerical methods, this study looks at the accuracy of some methods commonly found in physics-based engines [3], and measures their performance in following the theoretical phase trajectories and in reproducing the energy exchanges that occur in the continuous-time impact model.

D. Outline

In this paper, the Hunt–Crossley impact model is first characterized by proving some novel analytical results. Such results are then used as a reference to compare the accuracy of several numerical simulations of the model, obtained by discretizing the continuous-time equations with a number of widely used numerical methods. It is shown that for some regions of the parameter space, the trajectories of the discretized systems may significantly drift from the analytically derived curves. Finally, by exploiting the provided analytical results, two approaches are proposed which improve the accuracy of the numerical simulations, thus restoring their energy consistency.

More in detail, the main contribution of this work can be summarized as follows.

1) Novel approximate closed-form expression for the outbound velocity, and quantification of its relative error. Expression of the total energy (Hamiltonian) as a function of compression velocity (Section II-A).
2) Analysis of the distortion caused by a constant external force on the analytically derived phase portraits. This is of some importance for applying the results to real-world simulations (Section II-B).
3) Comparison of the analytically derived phase portrait and Hamiltonian with those obtained by applying four relevant numerical methods to the Hunt–Crossley impact model. Two critical cases are examined: low dissipation and hard impact (Section III-B).
4) Proposal of two correction methods for numerical models, one based on the analytical dependence of compression on velocity, and one based on a constraint on the outbound velocity (Section IV).

The remainder of the paper is organized as follows. Section II contains the analytical study, Section III-A describes different numerical methods which are used for discretizing the continuous-time system, while in Section III-B the corresponding numerical simulations are compared. Sections IV-A and IV-B show how the provided analytical results can be used to improve the behavior of the numerical simulations. Finally, in Sections IV-C and IV-D the computational cost and accuracy of simulations with and without corrections are compared and evaluated.

II. Impact Model

The Hunt–Crossley impact model [12] is described by the following nonlinear equation for the impact force:

$$f(x,v) = \begin{cases} kx^{\alpha} + \lambda x^{\alpha} v = kx^{\alpha} \cdot (1 + \mu v), & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (1)$$

where x is the *compression*, $v = \dot{x}$ is the *compression velocity*, $\alpha > 1$ is the exponent of a power law and represents the *local shape* of contact surfaces, k is the *stiffness coefficient*, and $0 \leq \lambda \leq k$ is the *damping coefficient*. The mathematically convenient term $\mu\ (= \lambda/k)$ allows to simplify some closed-form calculations. The impact force model thus represents a nonlinear spring of constant k in parallel with a nonlinear damper of constant λ. The term kx^{α} corresponds to the *elastic component*, while $\lambda x^{\alpha} v$ represents the *dissipation* due to internal friction.

Marhefka and Orin [1] made use of the Hunt–Crossley model in order to represent the impact between a single mass m and a comparatively massive object which does not move during collision, therefore considering the system described by the equation

$$ma(t) = -f\left(x(t), v(t)\right) \quad (2)$$

where a is the *acceleration* of the mass m.

In order to simplify the study of the model and avoid details on the geometry of the contact surfaces, one can equivalently consider a point-mass impacting on a wall, where the contact is represented by the Hunt–Crossley model [13]. According to this

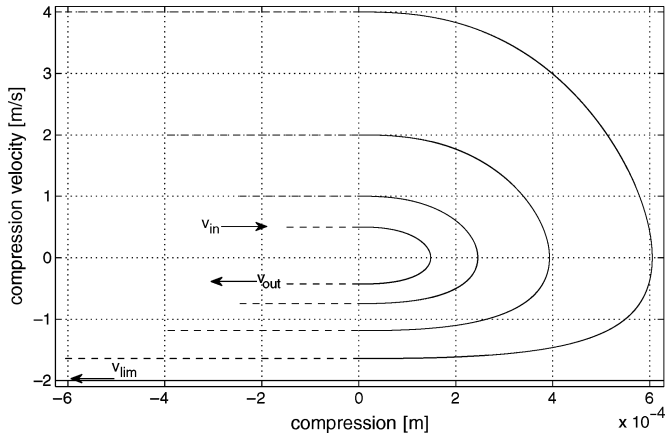


Fig. 1. Phase portraits for varying input velocities: $v_{\rm in}$ =0.5, 1, 2, 4 m/s. Other values of parameters are: $m = 10^{-2}$ Kg, $k = 10^{7}$ N/m$^{\alpha}$, $\mu = 0.5$ s/m, $\alpha = 1.5$. Solid lines represent the mass trajectory during contact; dashed lines represent free motion.

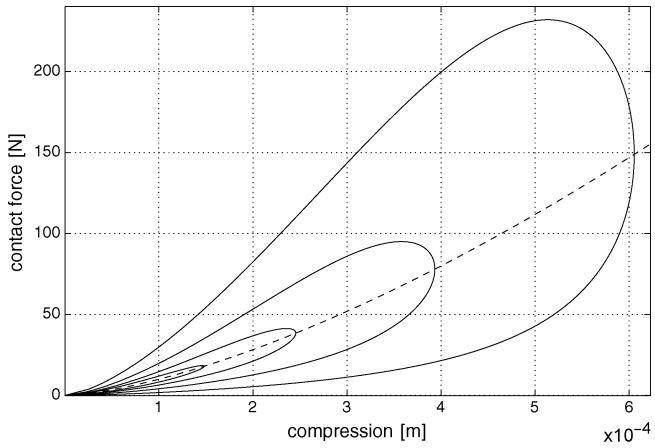


Fig. 2. Compression-force characteristics for varying input velocities: $v_{\rm in}$ =0.5, 1, 2, 4 m/s. Solid lines represent the case when dissipation is taken into account (the values of parameters are the same as in Fig. 1). The dashed line represents the curve kx^{α}, where no dissipation is considered ($\lambda = 0$).

configuration—represented by the system of (2) and (1)—the compression x and the compression velocity v are respectively defined as the position and velocity of the point-mass with respect to the wall (zero reference). The position and velocity are considered positive, respectively, while the mass is in contact with the wall, and when the mass is directed towards the latter.

A. Properties and Analytical Results

Thanks to the simple form of (2), the model can be treated analytically and some of its properties can be inferred. Hereafter, the initial conditions $x(0) = 0$ and $\dot{x}(0) = v_{\rm in}$ are considered, that is to say that the mass hits the wall with velocity $v_{\rm in}$ at time $t = 0$.

1) Compression: It is shown in [1] that from (2) it follows:

$$x(v) = \left[\frac{m(\alpha+1)}{k\mu^2}\cdot\left(-\mu(v-v_{\rm in}) + \log\left|\frac{1+\mu v}{1+\mu v_{\rm in}}\right|\right)\right]^{\frac{1}{\alpha+1}} \tag{3}$$

which can be exploited for plotting the phase portraits on the (x, v) plane shown in Fig. 1. From Fig. 1 it can be inferred that, due to the viscous dissipation occurring during contact, the relation $v(t+dt) < v(t)$ holds, and in particular the output velocity $v_{\rm out}$ is always smaller in magnitude than the corresponding $v_{\rm in}$. Moreover, for increasing $v_{\rm in}$'s, $v_{\rm out}$ converges to the limit value $v_{\rm lim} \triangleq -1/\mu$. The line $v = v_{\rm lim}$ represents the trajectory where the elastic and dissipative terms cancel, and separates two regions of the phase space, each of which is never entered by trajectories started in the other one.

Equation (3) allows to infer the *maximum compression* experienced during contact, which occurs when the compression velocity equals zero:

$$x_{\max} = x(0) = \left[\frac{m(\alpha+1)}{k\mu^2}\cdot\left(\mu v_{\rm in} + \log\left|\frac{1}{1+\mu v_{\rm in}}\right|\right)\right]^{\frac{1}{\alpha+1}}. \tag{4}$$

Finally, by substituting (3) in (2) one can plot the compression-force characteristics during collision, which are shown in Fig. 2. It can be noted that the dissipative term $\lambda x^{\alpha} v$ introduces hysteresis around the curve kx^{α}.

2) Output Velocity: The *restitution coefficient* E is defined as

$$E \triangleq \left|\frac{v_{\rm out}}{v_{\rm in}}\right|. \tag{5}$$

Note that $v_{\rm in}$ and $v_{\rm out}$ correspond to the roots of the right-hand side of (3), that is the points where $x = 0$. As a result, $v_{\rm out}$ can be defined implicitly from (3) as a function of $(\mu, v_{\rm in})$ only:

$$\mu v_{\rm out} - \log|1+\mu v_{\rm out}| = \mu v_{\rm in} - \log|1+\mu v_{\rm in}|. \tag{6}$$

This implies that $\mu v_{\rm out}$ is a function of $\mu v_{\rm in}$ only, and therefore E is also a function of $\mu v_{\rm in}$ only.

Analytical derivations of the dependence $E(\mu v_{\rm in})$ have been classically performed in the limit of small initial velocities and/or small dissipation [12].[1] However, we suggest that a nonlocal approximation $\tilde{v}_{\rm out}$ can be empirically determined as an *ansatz* which fits the curve $E(\mu v_{\rm in})$ in the two limit regions $\mu v_{\rm in} \to 0_+$ and $\mu v_{\rm in} \to +\infty$, thus obtaining:[2]

$$\tilde{v}_{\rm out}(\mu, v_{\rm in}) = v_{\rm lim}\left[1 - \left(\sum_{j=0}^{n} b_j \cdot v_{\rm in}^{j}\right) e^{-2\mu v_{\rm in}}\right] \tag{7}$$

where, in the case $n = 4$, the coefficients b_j are

$$b_0 = 1, \quad b_1 = \mu, \quad b_2 = \frac{2}{3}\mu^2, \quad b_3 = \frac{2}{9}\mu^3, \quad b_4 = \frac{14}{135}\mu^4. \tag{8}$$

From now on, unless specified otherwise, the notation $\tilde{v}_{\rm out}$ will refer to the fourth-order approximation provided by (7) and the coefficients (8). Conventional iterative zero-finding methods applied to (6) can always be used to compute a more precise release velocity at a higher computational cost (see Section IV-C).

Fig. 3 shows the error introduced by the approximate value $\tilde{v}_{\rm out}$, when compared to the corresponding value computed numerically as a zero of (6).

[1]Appendix A provides an example.

[2]See Appendix B for details.

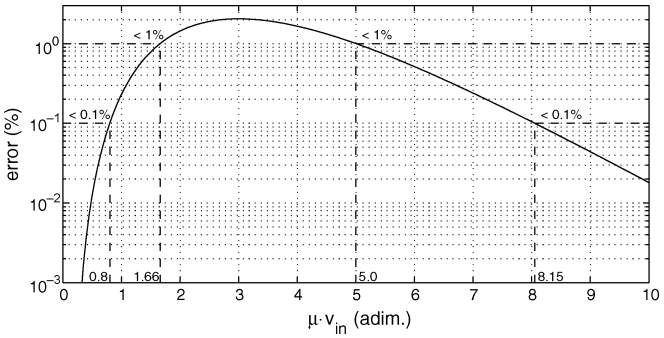


Fig. 3. Log-scaled percentage error of the output velocity approximated by $\bar{v}_{\rm out}$, with respect to the value computed numerically as a zero of (6). Ranges of $\mu v_{\rm in}$ are shown, for which the maximum error is respectively less than 1% and 0.1%.

3) Contact Time: It is shown in [25] that the *contact time* can be expressed as

$$\tau = \left(\frac{m}{k}\right)^{\frac{1}{\alpha+1}} \cdot \left(\frac{\mu^2}{\alpha+1}\right)^{\frac{\alpha}{\alpha+1}} \cdot \int_{v_{\rm out}}^{v_{\rm in}} \frac{1}{(1+\mu v)\left[-\mu(v-v_{\rm in}) + \log\left|\frac{1+\mu v}{1+\mu v_{\rm in}}\right|\right]^{\frac{\alpha}{\alpha+1}}}. \quad (9)$$

Equation (9) states that the contact time τ depends only on μ, the exponent α and the ratio m/k, plus obviously the impact velocity $v_{\rm in}$. Since neither m nor k affect the value of the integral (recall that $v_{\rm out}$ depends only on μ and $v_{\rm in}$), it follows that, given a fixed $v_{\rm in}$, the proportionality $\tau \sim (m/k)^{1/(\alpha+1)}$ holds.

From an auditory point of view the value of the contact time is strongly correlated to the perceived "hardness" of the impact [25], [26]. Namely, as the contact time decreases, the perceived hardness increases. Recalling the power-law dependence above and (1) it follows that, for a fixed mass m, "hard" and "soft" impacts correspond respectively to high and low force values.

4) Energy Properties and Behavior: The energy variation in a mechanical system can be calculated as the work made by the non-conservative forces $f_{\rm nc}$ acting on the system along a certain path $x_1 \rightarrow x_2$

$$\Delta H = \int_{x_1}^{x_2} f_{\rm nc}(x)dx = \int_{t_1}^{t_2} f_{\rm nc}(t)v(t)dt = -\Delta\Lambda \quad (10)$$

where H is the total energy content, known as the *Hamiltonian*, Λ is the energy dissipation, and the second integral is obtained by application of the chain rule ($dx(t) = v(t)dt$) and considering that t_1 and t_2 correspond respectively to the instants when the displacements x_1 and x_2 are reached. The Hamiltonian H is the sum of potential and kinetic energies, hereafter named V and T, respectively,

$$H(t) = V(t) + T(t). \quad (11)$$

With regard to the system represented by (2), T is related to the dynamics of the mass, which is described by the left-hand side of (2), while V is related to the elastic component of the impact force of (1).

In agreement with the last integral in (10), multiplying both sides of (2) by $v(t) = dx/dt$ and time-integrating them, gives

$$\underbrace{\int_0^t ma(s)v(s)ds}_{T(t)-T_0} = -\underbrace{\int_0^t kx(s)^\alpha v(s)ds}_{V(t)} - \underbrace{\int_0^t \lambda x(s)^\alpha v(s)^2 ds}_{\Lambda(t)>0} \quad (12)$$

where the force expression of (1) has been considered in the case $x > 0$ only. The first two integrals in (12) can be solved explicitly, obtaining

$$V(t) = \frac{kx(t)^{\alpha+1}}{\alpha+1}, \quad T(t) = \frac{mv(t)^2}{2}. \quad (13)$$

Considering a system where the mass travels with velocity $v_{\rm in}$ before an impact occurs, then the *initial Hamiltonian* corresponds to the initial kinetic energy

$$H_0 = T_0 = \frac{mv_{\rm in}^2}{2}. \quad (14)$$

From (10) it follows that at each time instant t:

$$H(t) = H_0 - \Lambda(t) \quad (15)$$

and, since $\Lambda(t) > 0$, the following inequalities hold:

$$0 \le H(t+dt) \le H(t). \quad (16)$$

Indicating τ as the instant when an impact ends, the *final Hamiltonian* of the system, that is the energy content right after contact, can be written as

$$H_\tau = T_\tau = \frac{mv_{\rm out}^2}{2}. \quad (17)$$

Also, the total amount of energy dissipation occurred during contact is

$$\Delta H_\tau = H_\tau - H_0 = -\Lambda_\tau. \quad (18)$$

Since the area enclosed by the hysteresis loops shown in Fig. 2 is by definition the work done by the impact force, such area corresponds to ΔH_τ.

As for the rightmost integral in (12), which is non-solvable, an equivalent expression can be obtained by substituting the complementary results for the remaining integrals:

$$\Lambda(t) = \int_0^t \lambda x(s)^\alpha v(s)^2 ds = \frac{m\left(v_{\rm in}^2 - v(t)^2\right)}{2} - \frac{kx(t)^{\alpha+1}}{\alpha+1}. \quad (19)$$

Finally, by substituting (3) in (19) and recalling (15), the following expression in v only is found:

$$H(v) = H_0 - \Lambda(v) = \frac{m}{2}v^2 - \frac{m}{\mu}(v - v_{\rm in}) + \frac{m}{\mu^2}\log\left|\frac{1+\mu v}{1+\mu v_{\rm in}}\right| \quad (20)$$

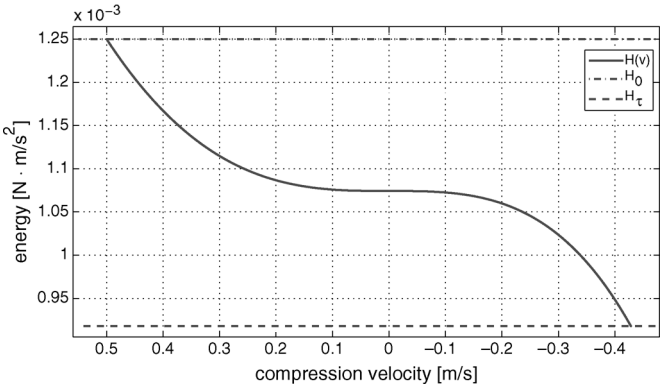


Fig. 4. Compression velocity–Hamiltonian characteristic. The two horizontal lines display $H_0 = T_0$ and $H_\tau = T_\tau$, that is, respectively, the initial and the final Hamiltonian. The values of parameters are the same as in Fig. 1 for $v_{\rm in} = 0.5$ m/s. The compression velocity axis has been inverted, thus allowing to read the graph from left to right.

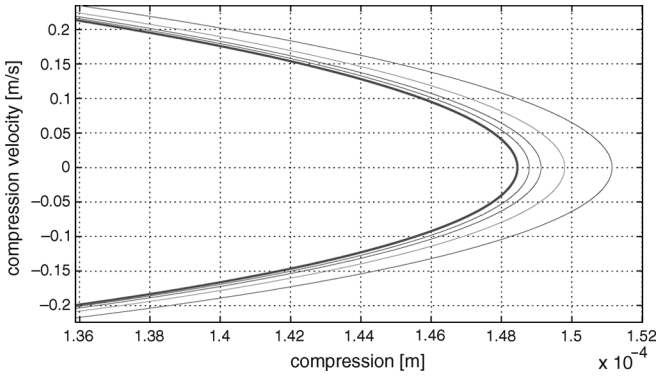


Fig. 5. Detail of phase portraits for different external forces applied: $f_e = m \cdot g \cdot (0.5, 1, 2, 4)$ N, where $g = 9.81$ m/s^2 is the standard gravity. The bold line represents the case where $f_e = 0$. The values of parameters are the same as in Fig. 1 for $v_{\rm in} = 0.5$ m/s. Notice that, due to the lack of analytical results when an external force is present, the phase portraits have been obtained as the result of numerical simulations.

which can be used for plotting the curve shown in Fig. 4.

B. Addition of a Constant External Force

When a constant external force f_e (e.g., the force of gravity) is applied to the mass, (2) has to be rewritten as

$$ma(t) + f_e = -f\left(x(t), v(t)\right). \tag{21}$$

Unfortunately, in this case no closed-form analytical results can be found as those described in Section II-A. In more detail, multiplying both sides of (21) by $v(t)$ and time-integrating them, an equation is found where an unsolvable integral is present, this way preventing to directly obtain explicit-form expressions for $x(v)$, $\tilde{v}_{\rm out}$ and τ.

Rewriting the second-order equation in (21) as a system of first-order equations

$$\begin{cases} \dot{x} = v \\ \dot{v} = -\frac{k}{m}x^\alpha - \frac{\lambda}{m}x^\alpha v - \frac{f_e}{m} \end{cases} \tag{22}$$

the equilibrium point of the system is found to be $(x_{\rm eq}, v_{\rm eq}) = (-f_e^{1/\alpha}/k, 0)$, which corresponds to the *compression offset* in stationary conditions.

As Fig. 5 shows, for positive values of f_e and $v_{\rm in}$, the velocity of the mass during the compression phase is generally greater than in the case when $f_e = 0$. In particular, at the beginning of contact interaction, since f_e is higher than the current impact force f, the compression velocity exceeds $v_{\rm in}$. On the other hand, compared to the case when $f_e = 0$, during the decompression phase the absolute value of the mass velocity v decreases, resulting in lower output velocities. Moreover, the resulting maximum compression is always greater than that calculated by (4).

III. Numerical Simulations

In this section, the continuous-time system described by (2) is discretized by means of several numerical methods, and the resulting numerical systems are studied.

A. Numerical Methods

Different numerical methods were considered, which are commonly used in various fields of applications ranging from computer graphics, to physical simulation of dynamical systems and digital signal processing. Thanks to their low order—which generally results in low computational costs—the chosen methods are particularly suitable for real-time applications:

- the trapezoid rule is popular to translate analog filter structures to discrete-time filters, and it is the basis for wave digital filters;
- Verlet integration is popular in physics-based graphic engines;
- Heun is a second-order method (complexity similar to the previous two) representative of the Runge–Kutta family;
- fourth-order Runge–Kutta is expected to be more accurate, and more expensive.

Following the standard notation in numerical analysis, the integration step is a constant named h $(= 1/F_s)$.

1) 1-Step Adams-Moulton (AM1): Is a A-stable second-order implicit method [27], also known as *bilinear transformation*, or *trapezoid rule*.

Discretizing (2) results in the following equation in state-space form:

$$\begin{bmatrix} x_{n+1} \\ v_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_n \\ v_n \end{bmatrix} + \begin{bmatrix} \frac{h^2}{4m} \\ \frac{h}{2m} \end{bmatrix} [f_{n+1} + f_n] \tag{23}$$

where the expression for the discrete-time force is obtained by replacing the continuous-time variables $x(t)$ and $v(t)$ in (1) with their discrete-time counterparts.

Since the AM1 method is implicit, (23) is also in implicit form, and this is reflected in the instantaneous relationship between $[x_{n+1}\ v_{n+1}]^T$ and f_{n+1}. Unfortunately, since f_{n+1} also has an instantaneous dependence on x_{n+1} and v_{n+1} given by (1), the discrete-time counterpart of the system described by (2) contains a *delay-free loop*, which is not directly computable and—because of the nonlinear dependence $f(x, v)$—needs some special handling in order to be solved. In particular, the *K-method* [28] together with *Newton's method* [27] are used, weighing on the efficiency of the simulation (see Section IV-C).

2) Verlet: Is a second-order explicit method commonly used in computer graphics [3], video games, and molecular dynamics simulation, where it is typically used for integrating Newton's equation of motion in order to describe the trajectory of moving

particles. The one used here is a variant, called *velocity Verlet*, which provides better handling of the velocity variable and can be seen as a predictor-corrector method.

Discretizing (2) results in the following implementation scheme:

$$\begin{aligned} x_{n+1} &= x_n + hv_n + \frac{h^2}{2}\frac{f_n}{m} \\ v_{n+\frac{1}{2}} &= v_n + \frac{h}{2}\frac{f_n}{m}, \quad \text{predictor} \\ f_{n+1} &= f\left(x_{n+1}, v_{n+\frac{1}{2}}\right) \\ v_{n+1} &= v_{n+\frac{1}{2}} + \frac{h}{2}\frac{f_{n+1}}{m}, \quad \text{corrector.} \end{aligned} \tag{24}$$

It should be noted that this algorithm assumes that f_{n+1} only depends on the predicted velocity $v_{n+(1/2)}$, which clearly gives rise to inaccuracies.

3) Heun: Is a predictor-corrector explicit method [3], [27], with the forward Euler method as predictor and the trapezoid rule as corrector. It can also be seen as a *second-order Runge–Kutta method* (RK2).

Discretizing (2) results in the following implementation scheme:

$$\begin{aligned} \tilde{v}_{n+1} &= v_n + h\frac{f_n}{m}, \quad \text{predictor} \\ x_{n+1} &= x_n + \frac{h}{2}(v_n + \tilde{v}_{n+1}) \\ f_{n+1} &= f(x_{n+1}, \tilde{v}_{n+1}) \\ v_{n+1} &= v_n + \frac{h}{2}\frac{f_n + f_{n+1}}{m}, \quad \text{corrector.} \end{aligned} \tag{25}$$

Again, it should be noted that both x_{n+1} and f_{n+1} only depend on the predicted velocity $\tilde{v}_{n+1}$, and this gives rise to inaccuracies.

4) Fourth-Order Runge–Kutta (RK4): Is an explicit iterative method [3], [27] which is widely used to solve ODEs with improved accuracy.

Discretizing (2) results in the following implementation scheme:

$$\begin{aligned} x_{n+1} &= x_n + \frac{1}{6}(l_1 + 2l_2 + 2l_3 + l_4) \\ v_{n+1} &= v_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \end{aligned} \tag{26a}$$

where

$$\begin{aligned} l_1 &= hv_n, \quad l_2 = h\left(v_n + \frac{k_1}{2}\right) \\ l_3 &= h\left(v_n + \frac{k_2}{2}\right), \quad l_4 = h(v_n + k_3) \\ k_1 &= h\frac{f_n}{m}, \quad k_2 = h\frac{f\left(x_n + \frac{l_1}{2}, v_n + \frac{k_1}{2}\right)}{m} \\ k_3 &= h\frac{f\left(x_n + \frac{l_2}{2}, v_n + \frac{k_2}{2}\right)}{m} \\ k_4 &= h\frac{f(x_n + l_3, v_n + k_3)}{m}. \end{aligned} \tag{26b}$$

It should be noted that, for each sample, both the velocity and the nonlinear force of (1) need to be evaluated four times, therefore strongly affecting the efficiency of the simulation (see Section IV-C).

B. Experimental Results

In order to evaluate the chosen numerical methods, it is useful to compare the behavior of the corresponding simulations against the known analytical results (see Section II-A).

The main references used to quantitatively assess the reliability of a particular numerical method *during* contact are (3) and (20), which express respectively the compression x and the Hamiltonian H as functions of the compression velocity v. The errors in x and H are then measured as the maximum deviations of their discrete-time versions from the respective analytical curves $x(v)$ and $H(v)$, in relation to the entire variation range of the quantities considered (i.e., $x_{\max}$ and ΔH_τ). In detail, taking into account (20) and defining $H^{\text{sim}}(x_n, v_n)$ as the Hamiltonian computed from (13) using compression and velocity values resulting from a numerical simulation,[3] and ΔH_τ as in (18), the *maximum deviation on* H is calculated by means of the following expression:

$$\%\text{dev}(H) = 100 \cdot \left| \frac{\max_n \left\{ H^{\text{sim}}(x_n, v_n) - H(v_n) \right\}}{\Delta H_\tau} \right|. \tag{27}$$

Thus, the deviation is obtained by normalizing the absolute error according to the *range of variation* of the quantities being evaluated, and picking the maximum deviation along such range. It can be viewed as the Chebyshev distance between H^{sim} and H during contact, normalized by ΔH_τ. Such measure has a different meaning compared to ordinary relative errors, since the latter would only account for local deviations from the analytical curves. Similarly to what is described above, considering (3) and defining $x_{\max}$ as in (4), the *maximum deviation on* x is calculated as

$$\%\text{dev}(x) = 100 \cdot \left| \frac{\max_n \left\{ x_n - x(v_n) \right\}}{x_{\max}} \right|. \tag{28}$$

Finally, another indicator which allows to evaluate the accuracy and consistency of the simulations at release from contact is provided by the output velocity computed numerically as a zero of (6), which is used to calculate the relative error $\%\text{err}(v_{\text{out}})$ of its simulated counterpart $v_{\text{out}}^{\text{sim}}$.

Throughout the following example simulations some values of parameters are kept constant: $m = 10^{-2}$ kg, and $F_s = 44.1$ kHz (i.e., a standard audio sample rate).

1) Non-Critical Simulations: In order to verify the numerical implementations, the parameters of the model are set to a "safe" configuration, that is to far-from-extreme values. In this case, contact extends over many samples, thus ensuring that the simulations are influenced only to a negligible extent by the chosen sampling rate (see Section I-C) and should more likely behave as the original continuous-time system.

This is confirmed qualitatively by Fig. 6, where the plots of all such simulations substantially overlap and coincide with the

[3]Not to be confused with the discrete Hamiltonian defined in [24] for finite-difference schemes.

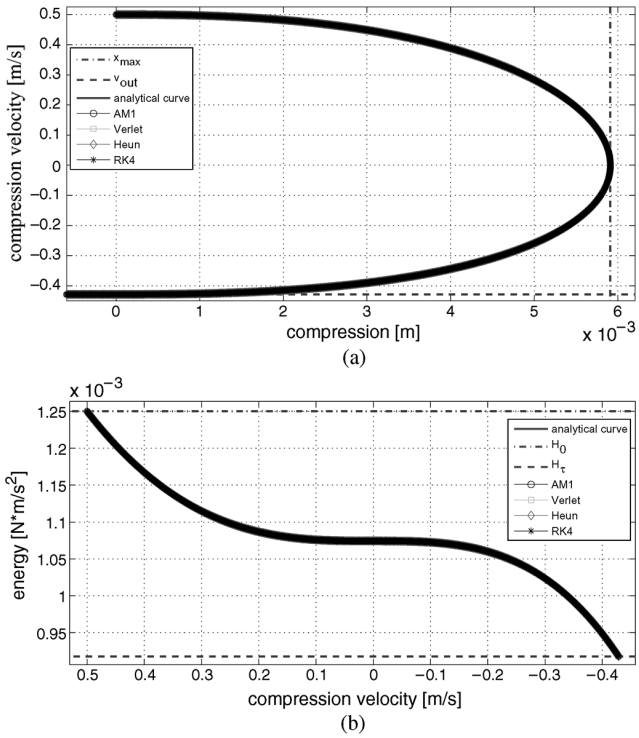


Fig. 6. Qualitative comparison of different methods in noncritical simulations. The values of parameters are $k = 10^3$ N/m$^\alpha$, $\mu = 0.5$ s/m, $\alpha = 1.5$, $v_{\rm in} = 0.5$ m/s. (a) Phase portraits. The two tangent lines represent the maximum compression $x_{\max}$ calculated by (4) and the output velocity $v_{\rm out}$ computed numerically as a zero of (6). (b) Energy behaviors. The two horizontal lines display $H_0 = T_0$ and $H_\tau = T_\tau$, i.e., respectively, the initial and the final Hamiltonian.

TABLE I
SUMMARY OF ERRORS IN NONCRITICAL SIMULATIONS. THE VALUES OF PARAMETERS ARE $k = 10^3$ N/m$^\alpha$, $\mu = 0.5$ s/m, $\alpha = 1.5$, $v_{\rm in} = 0.5$ m/s

errors	AM1	Verlet	Heun	RK4
%dev(x)	0.255	0.018	0.319	0.005
%err($v_{\rm out}$)	$+2\cdot10^{-5}$	$+2\cdot10^{-6}$	$-3\cdot10^{-5}$	$+2\cdot10^{-6}$
%dev(H)	$3\cdot10^{-4}$	0.052	$4\cdot10^{-4}$	$1\cdot10^{-5}$

analytical curves. Moreover, Table I offers a quantitative evaluation of the simulations, showing the errors introduced by the different numerical methods considered.

2) Critical Parameter Regions: It has been found empirically that when $\tau^{\rm sim} \leq 4$ samples, the errors in x, H and $v_{\rm out}$ heavily increase, resulting in an extremely poor reliability of all the simulations. The obvious reason for this behavior is that, since only very few samples of data are available, the numerical systems are totally unable to describe the original continuous-time counterpart. Hence, in the following study, only values of parameters resulting in $\tau^{\rm sim} > 4$ samples are considered.

It has been observed that when $\mu \to 0_+$ and/or as the contact time $\tau^{\rm sim}$ decreases (i.e., for "hard" impacts), the behavior of most numerical implementations tends to become inconsistent with the continuous-time system. Hereafter, the numerical systems are studied for these two critical configurations, respectively, named *case 1* and *case 2*.

Case 1: Low dissipation ($\mu \to 0_+$).

In case of low dissipation, the Hamiltonian of both Verlet- and Heun-discretized systems is prone to oscillations, while

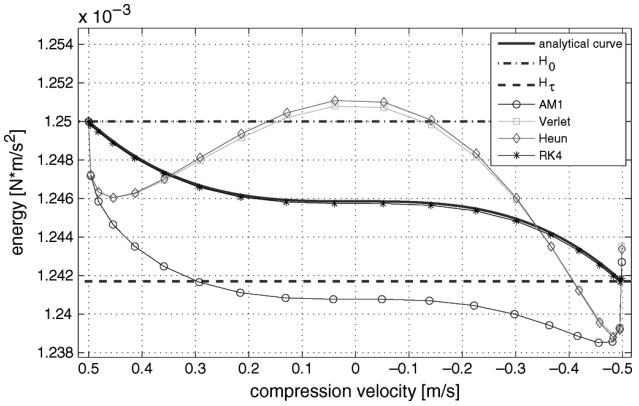


Fig. 7. Comparison of Hamiltonians for different implementations of a simulation example following *case 1*. The values of parameters are $k = 10^7$ N/m$^\alpha$, $\mu = 0.01$ s/m, $\alpha = 1.3$, $v_{\rm in} = 0.5$ m/s. The contact time equals 19 samples.

TABLE II
SUMMARY OF ERRORS IN EXAMPLE SIMULATIONS OF *case 1* AND *case 2*. THE LAST COLUMN SHOWS THE ERROR YIELDED BY THE APPROXIMATE VALUE $\tilde{v}_{\rm out}$ WITH RESPECT TO THE VALUE COMPUTED NUMERICALLY: IT IS WORTH NOTICING THAT IN BOTH CASES THE ERROR IS LOWER THAN THOSE YIELDED BY THE SIMULATIONS. (a) SIMULATIONS FOLLOWING *case 1*. THE VALUES OF PARAMETERS ARE THE SAME AS IN FIG. 7. (b) SIMULATIONS FOLLOWING *case 2*. THE VALUES OF PARAMETERS ARE THE SAME AS IN FIG. 8

(a)

errors	AM1	Verlet	Heun	RK4	$\tilde{v}_{\rm out}$
%dev(x)	1.011	1.083	1.136	0.052	-
%err($v_{\rm out}$)	$+0.039$	$+0.073$	$+0.067$	$+0.006$	$-1\cdot10^{-5}$
%dev(H)	61.302	59.542	63.042	1.427	-

(b)

errors	AM1	Verlet	Heun	RK4	$\tilde{v}_{\rm out}$
%dev(x)	4.381	4.418	19.506	0.412	-
%err($v_{\rm out}$)	$+2.551$	$+0.839$	-4.692	-0.105	-0.013
%dev(H)	7.885	9.475	23.387	0.410	-

contact ends in an inconsistent final energy state: typically, $H_\tau^{\rm Verlet,Heun} > H_\tau$, where H_τ is defined as in (17) and is calculated using values of $v_{\rm out}$ computed numerically as zeros of (6). As for AM1-discretized systems, these generally tend to dissipate too much energy during contact (i.e., $H^{\rm AM1} < H$), while slightly gaining spurious energy as the contact ends (i.e., $H_\tau^{\rm AM1} > H_\tau$). On the other hand, RK4-discretized systems generally behave quite consistently both during and after the contact interaction (i.e., $H^{\rm RK4} \approx H$).

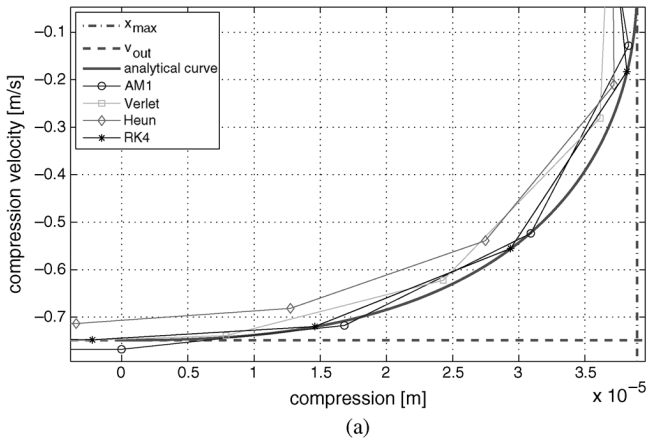
Fig. 7 shows the Hamiltonian of a simulation example with low dissipation ($\mu = 0.01$), while Table II(a) shows the resulting errors.

Case 2: Hard impacts.

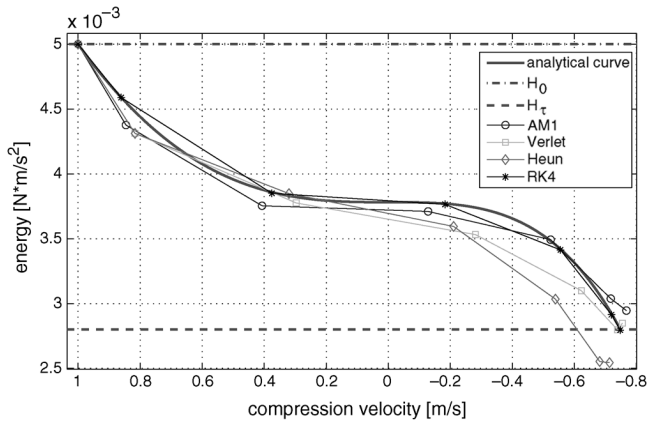
With the exception of RK4-discretized systems, in this case the simulations usually show more spread errors: $x^{\rm sim}$, $H^{\rm sim}$, and $v_{\rm out}^{\rm sim}$ tend to substantially deviate from the respective analytical results.

Fig. 8 shows a hard impact simulation example where the values of parameters are $k = 10^9$ N/m$^\alpha$, $\mu = 0.5$ s/m, $\alpha = 1.5$, $v_{\rm in} = 1$ m/s, while Table II(b) shows the corresponding errors. The resulting contact time τ equals 6 samples.

It is worth noticing that the RK4 method has been proved to behave quite consistently across disparate configurations of



(a)



(b)

Fig. 8. Comparison of phase portraits and Hamiltonians for different implementations of a simulation example following *case 2*. The values of parameters are $k = 10^9$ N/m$^\alpha$, $\mu = 0.5$ s/m, $\alpha = 1.5$, $v_{\rm in} = 1$ m/s. The contact time equals 6 samples. (a) Detail of phase portraits. (b) Hamiltonians.

parameters. Therefore, a highly oversampled RK-discretized system can be taken as a reference, able to provide extremely accurate simulations.

IV. Improved Numerical Simulations

A. Exploitation of Analytical Results

Making use of some of the analytical results obtained in Section II-A, in this section solutions are proposed to fix the inconsistencies pointed out in Section III-B. The aim is to improve the accuracy and reliability of simulations which make use of the impact model under study, in view of their implementation as real-time applications.

1) Hybrid Numerical-Analytical Computation: Instead of computing both the state variables x_n and v_n as the result of a numerical simulation—for example using one of the numerical methods described in Section III-A—this solution consists in computing only the compression velocity v_n *numerically*, then employing it in (3) to calculate the corresponding value of compression $x(v_n)$ *analytically*.

As a result, the corrected numerical system strictly follows the analytical curves $x(v)$ and $H(v)$.

Since the computation is to be made at each sample for the whole duration of the contact interaction, this solution noticeably increases the computational load of the simulation.

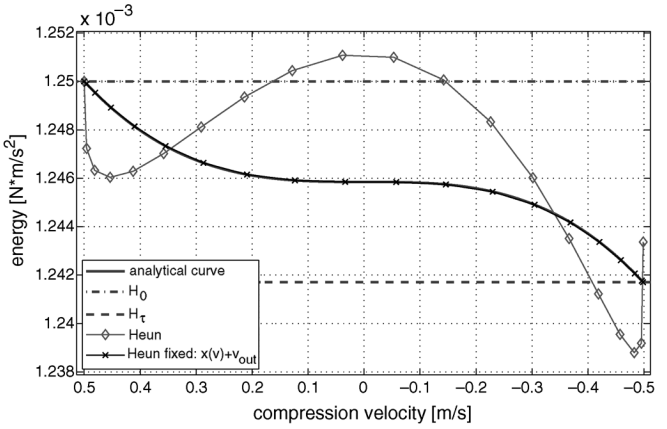


Fig. 9. Comparison of the Hamiltonians of a Heun-based simulation example following *case 1*, with and without corrections. The values of parameters are the same as in Fig. 7.

2) Output Velocity Constraint: The solution above can only be applied while the contact interaction lasts, and therefore the behavior of a numerical system upon the end of the impact is not affected. An additional method is proposed here to control the energy content (i.e., the residual kinetic energy $T_\tau^{\rm sim} = H_\tau^{\rm sim}$) at that time: as soon as x_n takes a non-positive value following an impact (i.e., at the detachment of the two objects), the output velocity can be forced to the approximate value $\tilde{v}_{\rm out}$ given by (7).

Considering that the error introduced by (7) depends solely on the product $\mu v_{\rm in}$ (see Fig. 3), it is advisable to apply the correction only when the product $\mu v_{\rm in}$ corresponds to an acceptable error, or the risk is to even worsen the behavior of the numerical system. However, the use of this conditional correction always implies a tradeoff: on the one hand it guarantees not to introduce errors greater than a chosen maximum, while on the other hand 1) within the excluded range of $\mu v_{\rm in}$, the output velocity is not controlled and therefore depends only on the plain numerical method chosen, 2) within the included intervals of $\mu v_{\rm in}$, the risk is that the correction introduces errors even greater than those provided by the non-corrected numerical system (this is true especially for noncritical parameter regions).

Once the output velocity has been forced to $\tilde{v}_{\rm out}$, the corresponding compression should be set to 0, this way adhering to the compression-force characteristics shown in Fig. 2 by closing their numerical counterpart at $(x = 0, f = 0)$, and ensuring that the final potential energy $V_\tau^{\rm sim}$ is set to zero.

The computation of $\tilde{v}_{\rm out}$ only needs to take place in correspondence to an impact event, and as soon as the impact velocity $v_{\rm in}$ is known.

B. Numerical Simulations With Corrections

In order to test the described corrections, they were applied to the worst behaving simulation examples provided in Section III-B2: Figs. 9 and 10 show a comparison of Heun-based simulations following, respectively, *case 1* and *case 2*, with and without corrections.

1) Improved Energy Behavior: When the hybrid correction described in Section IV-A1 is applied, any simulation strictly adheres to the analytical curves $H(v)$ and $x(v)$ during contact,

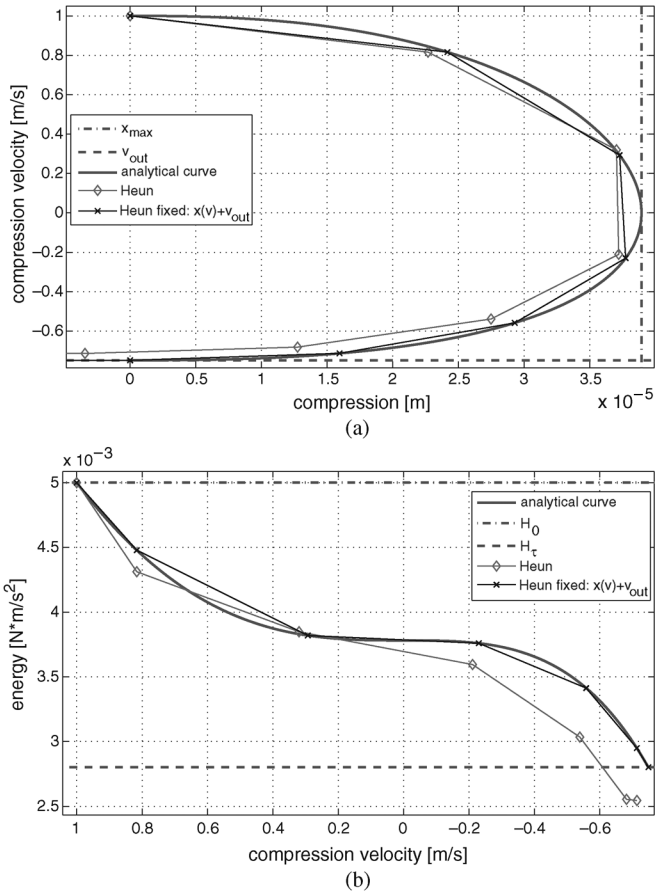


Fig. 10. Comparison of a Heun-based simulation example following *case 2*, with and without corrections. The values of parameters are the same as in Fig. 8. (a) Phase portraits. (b) Hamiltonians.

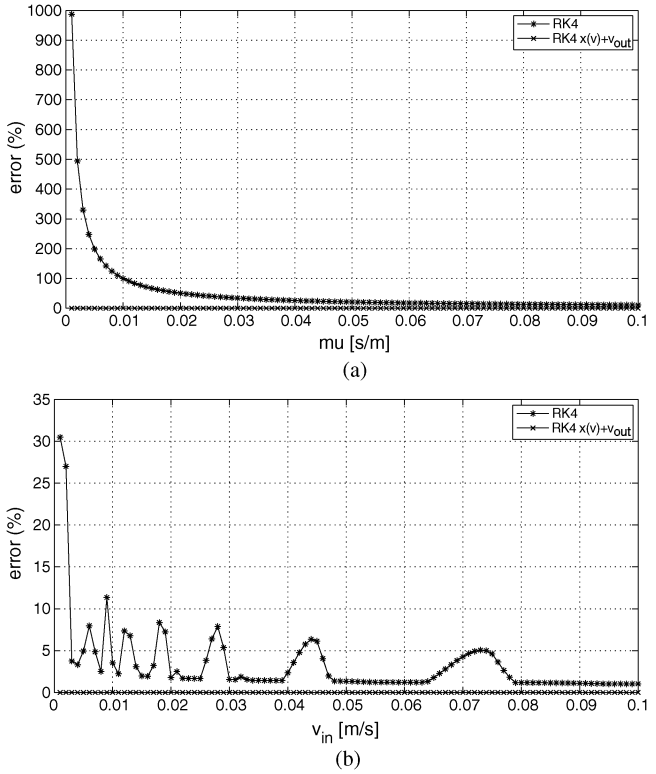


Fig. 11. Trend of error on the Hamiltonian for RK4-based simulations with and without corrections, for small values of μ and $v_{\rm in}$. For the values of $\mu v_{\rm in}$ considered, the output velocity constraint is always applied (see Fig. 3). The values of parameters, where kept constant, are $k = 10^9$ N/m$^\alpha$, $\mu = 0.5$ s/m, $\alpha = 1.5$, $v_{\rm in} = 0.5$ m/s. At each corresponding point of the two parallel plots, the product $\mu v_{\rm in}$ is the same. It can be noted that, when not using any corrections, the weight of μ on the error is clearly greater than that of $v_{\rm in}$. (a) Maximum deviation on H as μ varies. (b) Maximum deviation on H as $v_{\rm in}$ varies.

that is the respective errors, as defined in (27) and (28), are equal to zero.[4]

As for the energy state upon the end of the interaction, the error on H_τ depends either on the error introduced by $\tilde{v}_{\rm out}$ (when the output velocity constraint described in Section IV-A2 is actually applied), or on the error introduced by the plain numerical simulation. As already stated in Section IV-A2, in the first case the maximum error on $v_{\rm out}$ is predictable, and clearly the same goes for the error on H_τ.

Equation (27) allows to quantitatively assess the improvements on the energy consistency of the numerical simulations.

The trend of error on H resulting from simulations with and without corrections is depicted in Fig. 11(a) and 11(b), which show that even the best numerical method among those considered (i.e., RK4) can be improved, especially for critical values of parameters (see Section III-B2) such as $\mu v_{\rm in} \to 0_+$.

2) Sequence of Impacts: In order to better appreciate the importance of the proposed corrections, a sequence of rebounds has been implemented applying a conservative force (e.g., gravity) during free motion only.[5] Thanks to this setup, one can track the accumulation of energy anomalies at each contact interaction.

To this end, the residual energy $H^{\rm sim}_{\tau(i)}$ of numerical simulations after the ith impact was examined and compared to the residual energy $H_{\tau(i)}$ due to the exit velocity $v_{\rm out}(i)$ of the ith rebound, computed numerically as a zero of (6). The corresponding relative error provides a measure of the inaccuracies *accumulated* during the sequence of impacts. In addition, the deviation of $H^{\rm sim}$ with respect to the analytical curve $H(v)$ was measured according to (27) along the whole sequence of impacts.

Table III shows the errors on the residual energy $H^{\rm sim}_{\tau(100)}$ and the maximum deviations of $H^{\rm sim}$ occurred during a sequence of 100 impacts, for simulation examples following *case 1* and *case 2*. Notice that, since in some simulations the energy can strongly oscillate during contact (see Fig. 7), the maximum deviations from $H(v)$ do not necessarily reflect the accumulated errors. The last columns show the errors resulting from simulations corrected as suggested in Section IV-A, where the error threshold for the output velocity constraint is set to 0.1%, i.e., the correction is always applied (see Fig. 3).

In order to better understand the importance of such corrections, Fig. 12 provides a partial plot of the sequence of rebounds for *case 2*, where only the two better performing simulations are depicted.

[4]Not taking into account the inherent errors related to the representation of numbers in the digital domain.

[5]That is, the external force is temporarily suspended during the short contact at each rebound. As shown in Section II-B, when an external force is applied during contact, no closed-form analytical result is available, this way making the corrections provided in Section IV-A unsuitable.

TABLE III
SUMMARY OF ERRORS ON THE RESIDUAL ENERGY $H^{\mathrm{sim}}_{\tau(100)}$ (ACCUMULATED ERROR), AND MAXIMUM DEVIATIONS OF H^{sim} CALCULATED ACCORDING TO (27) ALONG THE WHOLE SUCCESSION OF 100 REBOUNDS. (a) SIMULATION EXAMPLES FOLLOWING *case 1*. THE VALUES OF PARAMETERS ARE THE SAME AS IN FIG. 7, EXCEPT FOR INPUT VELOCITIES DECREASING AT EACH REBOUND. (b) SIMULATION EXAMPLES FOLLOWING *case 2*. THE VALUES OF PARAMETERS ARE THE SAME AS IN FIG. 8, EXCEPT FOR INPUT VELOCITIES DECREASING AT EACH REBOUND

(a)

errors on H	AM1	Verlet	Heun	RK4	corrected sim.
accum. %err	12.022	10.059	147.036	0.907	$< 10^{-7}$
max %dev	72.962	72.107	63.043	14.466	0

(b)

errors on H	AM1	Verlet	Heun	RK4	corrected sim.
accum. %err	24.285	15.780	946.600	2.153	0.001
max %dev	69.156	43.966	27.418	6.255	0

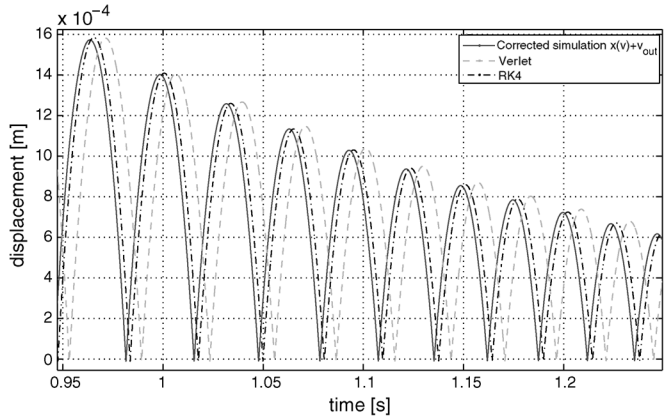


Fig. 12. Sequence of rebounds obtained from Verlet- and RK4-based simulations following *case 2*, compared with the trajectory of a corrected simulation.

From Table III and Fig. 12 it is evident that, even in case of errors apparently negligible for a single impact (see the errors relative to RK4 in Table II), indeed energy inconsistencies accumulate and can become noticeable—if not disastrous—after a certain number of contact interactions.

On the other hand, the errors resulting from the corrected simulations are actually negligible, and they are only due to inaccuracies in the forced exit velocity $\tilde{v}_{\mathrm{out}}$. The effectiveness of such corrections is confirmed comparing the results of a $16\times$ oversampled RK4 simulation used as a reference: in *case 1* the accumulated error on H is 0.001% and the maximum deviation is 0.024%, while in *case 2* they are, respectively, 0.002% and 0.006%.

C. Computational Cost

In this section, the computational costs of both the numerical implementations seen in Section III-A and the corrections suggested in Section IV-A are taken into account. In particular, the cost has been measured as the number of operations (i.e., memory write/read accesses and arithmetical operations) needed to execute an algorithm.

Table IV(a) summarizes the cost of the algorithms implementing the numerical methods. Since the AM1-based implementation makes use of Newton's method, its cost is displayed on two sub-columns: the first one shows the constant cost per sample, while the second one (in italics) shows the cost of a single iteration of Newton's method. Notice that the number of iterations per sample is not predictable.

TABLE IV
NUMBER OF OPERATIONS NEEDED BY THE NUMERICAL METHODS SHOWN IN SECTION III-A AND THE CORRECTIONS DESCRIBED IN SECTION IV-A. THE TOTALS IN BRACKETS ACCOUNT FOR WRITE/READ OPERATIONS. (a) SINCE THE AM1-BASED IMPLEMENTATION MAKES USE OF NEWTON'S METHOD, ITS COST IS DISPLAYED ON TWO SUB-COLUMNS: THE FIRST ONE SHOWS THE COST PER SAMPLE, WHILE THE SECOND ONE (IN ITALICS) SHOWS THE COST OF A SINGLE NEWTON'S METHOD ITERATION. (b) THE LAST COLUMN (IN ITALICS) SHOWS THE COST OF A SINGLE ZERO-FINDING ITERATION ON (6)

(a)

	AM1		Verlet	Heun	RK4
write	6	*10*	5	5	20
read	16	*33*	20	21	72
$+/-$	3	*8*	7	8	22
$\times$	8	*11*	5	5	18
$\div$	0	*1*	3	3	8
bit-shift	0	*0*	2	3	6
exp	0	*1*	1	1	4
log	0	*0*	0	0	0
compare	0	*4*	1	1	4
TOTAL	11 (33)	*25 (68)*	19 (44)	21 (47)	62 (154)

(b)

	hybrid correct.	$\tilde{v}_{\mathrm{out}}$ constraint	v_{out} as zero of (6)
write	1	2	*1*
read	7	4	*3*
$+/-$	6	5	*4*
$\times$	7	6	*3*
$\div$	3	4	*1*
bit-shift	0	1	*0*
exp	0	1	*0*
log	1	0	*1*
compare	1	1	*0*
TOTAL	18 (26)	18 (24)	*9 (12)*

Table IV(b) shows the cost of the corrections. In particular, the columns "hybrid correct." and "$\tilde{v}_{\mathrm{out}}$ constraint" account for straightforward implementations of (3) and (7), (8), respectively, which consist in substituting the continuous-time variables with their numerical counterparts.

For comparison, the last column reports the cost of a single zero-finding iteration on (6). The number of iterations depends on μ and v_{in}, and is in the order of some tens (usually between 15 and 40). Moreover, as μ decreases the number of iterations increases. It is clear that, despite being more precise than the approximate value $\tilde{v}_{\mathrm{out}}$, the value computed numerically as a zero of (6) implies several times the number of operations required by $\tilde{v}_{\mathrm{out}}$.

Since the computational load of simple write/read operations is generally low (if not negligible), two totals for each column are reported: one excluding write/read operations and, in brackets, one accounting for them.

Recalling that the hybrid correction only affects the computational cost during contact, whereas the output velocity constraint is applied at most once per impact event (as shown in Section IV-A2, the output velocity constraint should be applied conditionally), from Table IV one can infer that Verlet- or Heun-based simulations with corrections are roughly three times as efficient as plain RK4-based simulations during free motion, and almost twice as efficient during contact. On the other hand, while the exact computational load of an AM1-based simulation

is not predictable, it can be noted that it already matches the cost of a RK4-based one after two iterations of Newton's method.

D. Evaluation of Methods

The considered implementations can be finally evaluated in the light of the results regarding their accuracy (see Section III-B and IV-B2) and computational load (see Section IV-C).

It can be stated that, among the non-corrected implementations, the best all-round performance is achieved by RK4-discretized systems, although they are quite computationally expensive. On the other hand, a corrected Verlet-discretized system is generally at least as good as a non-corrected RK4 implementation, at a fraction of its computational load.

As for AM1-based implementations, it was shown that they behave quite poorly in the critical regions identified as *case 1* and *case 2*. These poor results in terms of accuracy, together with a generally high (and non-predictable) computational load, set the AM1 method as a hardly recommendable choice. Apparently such conclusion can be quite surprising, especially if one considers that AM1 is the only implicit and A-stable method among those considered. However, the presence of nonlinearities, together with the inaccuracies introduced by the K-method and Newton's method, justify the behavior of AM1-based implementations.

V. Conclusion

A nonlinear physical model of impact with sound synthesis applications has been reviewed, and its properties have been studied using both analytical tools and numerical simulations.

Several numerical realizations have been compared, and their shortcomings with regard to the corresponding analytical results have been pointed out. Special emphasis has been placed on energy consistency.

It has been shown that by exploiting the analytical results provided, the inconsistencies of the numerical realizations can be amended, thus restoring the correct energy state of the simulated systems, during and after contact.

Future research will consider finding a closed-form approximation of the release velocity for the system of (21), where a constant external force is applied, in this way allowing to implement suitable corrections. Even without such a closed-form solution, zero-finding numerical procedures could be profitably used to fix individual impacts. More interestingly for applications in acoustics, solutions will be investigated for extending the corrections to the case where vibrational losses are present, thus being applicable to impacts with resonating objects.

Appendix

A. Polynomial Expansion of the Output Velocity

As mentioned in Section II-A2, Hunt and Crossley [12] suggested that, in the limit $\mu v_{\rm in} \to 0_+$, the restitution coefficient can be approximated by the linear function $\widetilde{E}(\mu v_{\rm in}) = (1 - 2/3\mu v_{\rm in})$. Then, recalling (5), the corresponding output velocity is

$$\tilde{v}_{\rm out}(\mu, v_{\rm in}) = \left(-v_{\rm in} + \frac{2}{3}\mu v_{\rm in}^2\right). \tag{29}$$

This result can be easily verified through a Taylor expansion of the two sides in (6), using the approximation $\log|1+\epsilon| \approx \epsilon - (\epsilon^2/2) + (\epsilon^3/3)$, for $\epsilon \to 0_+$.

The same approach can be used in order to find a polynomial expansion of order n

$$\tilde{v}_{\rm out} = \sum_{j=1}^{n} a_j \cdot v_{\rm in}^j \tag{30}$$

in the limit $\mu v_{\rm in} \to 0_+$. Equation (30) is then substituted into the left-hand side of the Taylor expansion of (6). For the case $n = 4$ one obtains

$$\left[\frac{\mu^2}{2}v_{\rm out}^2 - \frac{\mu^3}{3}v_{\rm out}^3 + \frac{\mu^4}{4}v_{\rm out}^4 - \frac{\mu^5}{5}v_{\rm out}^5\right]_{v_{\rm out}=\sum_1^4 a_j v_{\rm in}^j} = \frac{1}{2}\mu^2 v_{\rm in}^2 - \frac{1}{3}\mu^3 v_{\rm in}^3 + \frac{1}{4}\mu^4 v_{\rm in}^4 - \frac{1}{5}\mu^5 v_{\rm in}^5. \tag{31}$$

The coefficients a_j are then determined by equating the two sides of (31) term by term, leading to the system

$$\begin{cases} \frac{1}{2}\mu^2 a_1^2 = \frac{1}{2}\mu^2 \\ \frac{1}{3}\mu^3 a_1^3 - \mu^2 a_1 a_2 = \frac{1}{3}\mu^2 \\ \frac{1}{4}\mu^4 a_1^4 - \mu^3 a_1^2 a_2 + \mu^2 a_1 a_3 + \frac{1}{2}\mu^2 a_2^2 = \frac{1}{4}\mu^2 \\ \frac{1}{5}\mu^5 a_1^5 - \mu^4 a_1^3 a_2 + \mu^3 a_1 a_2^2 + \mu^3 a_1^2 a_3 - \\ \quad \mu^2 a_1 a_4 - \mu^2 a_2 a_3 = \frac{1}{5}\mu^2 \end{cases} \tag{32}$$

which yields

$$a_1 = -1, \quad a_2 = \frac{2}{3}\mu, \quad a_3 = -\frac{4}{9}\mu^2, \quad a_4 = \frac{44}{135}\mu^3. \tag{33}$$

It follows that, in (30), $v_{\rm out}$ is a function of $(\mu, v_{\rm in})$ only. Equations (30) and (33) result in a slightly better approximation than (29), but still provide a local approximation.

B. Nonlocal Approximation for the Output Velocity

From (6) $v_{\rm out}$ can be written as

$$v_{\rm out} = \frac{1}{\mu}\left[(1+\mu v_{\rm in})e^{-\mu\Delta v} - 1\right] = v_{\rm lim}\left[1 - (1+\mu v_{\rm in})e^{-\mu\Delta v}\right] \tag{34}$$

where $\Delta v = v_{\rm in} - v_{\rm out}$. Equation (34) emphasizes that the convergence $v_{\rm out} \to v_{\rm lim}$ for $v_{\rm in} \to +\infty$ is governed by the fast-decreasing term $e^{-\mu\Delta v}$.

For small $v_{\rm in}$'s, the zeroth order approximation of the exponential term is $e^{-2\mu v_{\rm in}}$; however, it is easy to verify that the equation

$$\tilde{v}_{\rm out} = v_{\rm lim}\left[1 - (1+\mu v_{\rm in})e^{-2\mu v_{\rm in}}\right] \tag{35}$$

does not provide an accurate approximation of $v_{\rm out}$ for small $v_{\rm in}$'s. More precisely, its Taylor expansion around $v_{\rm in} = 0$ only matches the first Taylor coefficient of the expansion (30)

$$\tilde{v}_{\rm out} \underset{v_{\rm in}\to 0_+}{\approx} \sum_{j=1}^{n} \tilde{a}_j \cdot v_{\rm in}^j, \quad \tilde{a}_1 = -1 \quad \text{and} \quad \tilde{a}_j \neq a_j \quad (j > 1). \tag{36}$$

This qualitative discussion justifies to some extent the nonlocal approximation for $v_{\rm out}$ given in (7). Indeed (7) combines the two views—$\mu v_{\rm in} \to 0_+$ and $\mu v_{\rm in} \to +\infty$—on $v_{\rm out}$: the exponential term ensures the convergence $v_{\rm out} \to v_{\rm lim}$ for high values of

μv_{in}, while the polynomial coefficients b_j are determined by imposing that the Taylor expansion of (7) around $v_{\text{in}} = 0$ matches that of (30). Recalling that

$$e^{-2\mu v_{\text{in}}} \underset{v_{\text{in}} \to 0_+}{\approx} \sum_{j=1}^{n} c_j \cdot v_{\text{in}}^j, \quad c_j = \frac{(-1)^j}{j!}(2\mu)^j \tag{37}$$

then one can verify that the lth-order coefficient a_l of the Taylor expansion of (7) is given as $a_l = (1/\mu)\sum_{j=1}^{l} b_j \cdot c_{(l-j)}$. Then the coefficients b_j can be determined recursively as

$$b_0 = 1, \quad b_j = \mu a_j - \sum_{l=0}^{j-1} b_l \cdot c_{(j-l)} \quad (1 \le j \le n). \tag{38}$$

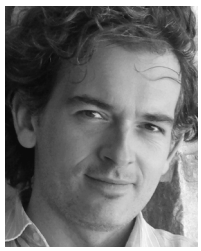
Applying this recursive equation for the case $n = 4$ yields (8).

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their insightful and helpful comments.

REFERENCES

[1] D. W. Marhefka and D. E. Orin, "A compliant contact model with nonlinear damping for simulation of robotic systems," *IEEE Trans. Syst., Man, Cybern. A: Syst. Humans*, vol. 29, no. 6, pp. 566–572, Nov. 1999.
[2] M. Mahvash, V. Hayward, and J. Lloyd, "Haptic rendering of tool contact," in *Proc. Eurohaptics*, 2002, pp. 110–115.
[3] M. Müller, J. Stam, D. James, and N. Thürey, "Real time physics: Class notes," in *Proc. SIGGRAPH '08: ACM SIGGRAPH 2008 Classes*, New York, 2008, pp. 1–90, ACM.
[4] A. Chaigne and J. Kergomard, *Acoustique des Instruments de Musique*. Paris, France: Belin, 2008.
[5] F. Avanzini and D. Rocchesso, "Modeling collision sounds: Non-linear contact force," in *Proc. Int. Conf. Digital Audio Effects (DAFx-01)*, Limerick, Ireland, Dec. 2001, pp. 61–66.
[6] M. Rath and D. Rocchesso, "Continuous sonic feedback from a rolling ball," *IEEE Multimedia*, vol. 12, no. 2, pp. 60–69, Apr./Jun. 2005.
[7] D. Rocchesso and F. Fontana, Eds., The Sounding Object Mondo Estremo, 2003 [Online]. Available: http://www.soundobject.org/
[8] P. Flores, J. P. Claro, and H. M. Lankarani, *Kinematics and Dynamics of Multibody Systems With Imperfect Joints: Models and Case Studies*. New York: Springer, 2008.
[9] G. Kuwabara and K. Kono, "Restitution coefficient in a collision between two spheres," *Jap. J. Appl. Phys.*, vol. 26, no. 8, pp. 1230–1233, 1987.
[10] A. Stulov, "Dynamic behavior and mechanical features of wool felt," *Acta Mechanica*, vol. 169, no. 1, pp. 13–21, 2004.
[11] L. Vu-Quoc and X. Zhang, "An elastoplastic contact force-displacement model in the normal direction: Displacement-driven version," in *Proc.: Math., Phys., Eng. Sci.*, 1999, pp. 4013–4044.
[12] K. H. Hunt and F. R. E. Crossley, "Coefficient of restitution interpreted as damping in vibroimpact," *ASME J. Appl. Mech.*, pp. 440–445, June 1975.
[13] N. Diolaiti, C. Melchiorri, and S. Stramigioli, "Contact impedance estimation for robotic systems," *IEEE Trans. Robotics*, vol. 21, no. 5, pp. 925–935, Oct. 2005.
[14] H. M. Lankarani and P. E. Nikravesh, "A contact force model with hysteresis damping for impact analysis of multibody systems," *J. Mech. Design*, vol. 112, no. 3, pp. 369–376, 1990.
[15] P. Flores, J. Ambrósio, J. Claro, and H. Lankarani, "Influence of the contact-impact force model on the dynamic response of multi-body systems," *Proc. Inst. Mech. Eng.-K*, vol. 220, no. 1, pp. 21–34, 2006.
[16] L. Pust and F. Peterka, "Impact oscillator with Hertz's model of contact," *Meccanica*, vol. 38, no. 1, pp. 99–116, 2003.
[17] S. Papetti, F. Fontana, M. Civolani, A. Berrezag, and V. Hayward, "Audio-tactile display of ground properties using interactive shoes," in *Haptic and Audio Interaction Design*, ser. Lecture Notes in Computer Science, R. Nordahl, S. Serafin, F. Fontana, and S. Brewster, Eds. Berlin/Heidelberg, Germany: Springer, 2010, vol. 6306, pp. 117–128 [Online]. Available: http://dx.doi.org/10.1007/978-3-642-15841-4_13
[18] A. Fettweis, "Wave digital filters: Theory and practice," *Proc. IEEE*, vol. 74, no. 2, pp. 270–327, Feb. 1986.
[19] S. Bilbao, "Robust physical modeling sound synthesis for nonlinear systems," *IEEE Signal Process. Mag.*, vol. 24, no. 2, pp. 32–41, Mar. 2007.
[20] A. Sarti and G. De Poli, "Toward nonlinear wave digital filters," *IEEE Trans. Signal Process.*, vol. 47, no. 6, pp. 1654–1668, Jun. 1999.
[21] I. Millington, *Game Physics Engine Development*. San Mateo, CA: Morgan Kaufmann, 2007.
[22] J. E. Colgate and J. M. Brown, "Factors affecting the Z-width of a haptic display," in *Proc. IEEE Int. Conf. Robot. Autom.*, San Diego, CA, May 1994, pp. 3205–3210.
[23] K. Kuchenbecker, J. Fiene, and G. Niemeyer, "Improving contact realism through event-based haptic feedback," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 2, pp. 219–230, Mar.–Apr. 2006.
[24] S. Bilbao, *Numerical Sound Synthesis*. New York: Wiley, 2009.
[25] F. Avanzini and D. Rocchesso, "Physical modeling of impacts: Theory and experiments on contact time and spectral centroid," in *Proc. Int. Conf. Sound Music Comput. (SMC'04)*, 2004, pp. 287–293.
[26] B. L. Giordano, D. Rocchesso, and S. McAdams, "Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability," *J. Exper. Psychol.: Human Percept. Perf.*, vol. 36, no. 2, pp. 462–476, 2010.
[27] A. Quarteroni, R. Sacco, and F. Saleri, *Numerical Mathematics*, 2nd ed. New York: Springer, 2007.
[28] G. Borin, G. De Poli, and D. Rocchesso, "Elimination of delay-free loops in discrete-time models of nonlinear acoustic systems," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 597–605, Sep. 2000.

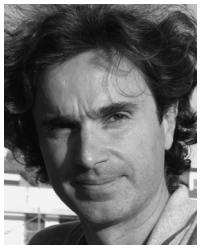


Stefano Papetti received the Laurea degree in computer engineering from the University of Padova, Padua, Italy, in 2006, and the Ph.D. degree in computer science from the University of Verona, Verona, Italy in 2010.

During the Ph.D. degree, he worked as a Visiting Researcher at Medialogy (Department of Architecture, Design and Media Technology), Aalborg University, Copenhagen, Denmark. Since 2006, he has been with the Computer Science Department, University of Verona, where he currently works as Research Associate. His research activity is focused on models and applications for interactive sound synthesis, and on the design and evaluation of auditory and haptic interfaces.

Federico Avanzini received the Laurea degree (*cum laude*) in physics from the University of Milano, Milan, Italy, in 1997 and the Ph.D. degree in information engineering from the University of Padova, Padua, Italy, in 2001 with a research project on sound and voice synthesis by physical modeling.

During the Ph.D. degree, he also worked as a Visiting Researcher at the Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Espoo, Finland. Since 2002, he has been with the Sound and Music Computing Group, Department of Information Engineering, University of Padova, where he is currently an Assistant Professor, teaching courses in computer science and sound and music computing. His main research interests are in the area of sound synthesis and processing, with particular focus on musical sound synthesis, nonspeech sound in multimodal interfaces, voice synthesis, and analysis. He has authored more than 70 publications in peer-reviewed international journals and conferences, and has been involved in numerous EU research projects. He is PI of the EU project DREAM (Culture2007) and of industry-funded projects.



Davide Rocchesso received the Ph.D. degree from the University of Padova, Padua, Italy, in 1996.

From 1998 and 2006, he was with the Computer Science Department, University of Verona, Verona, Italy, as an Assistant and Associate Professor. Since 2006, he has been with the Department of Art and Industrial Design, IUAV University of Venice, Venice, Italy, as an Associate Professor. He has been the coordinator of EU project SOb (the Sounding Object), and local coordinator of the EU project CLOSED (Closing the Loop Of Sound Evaluation and Design) and of the Coordination Action $\mathrm{S2S}^2$ (Sound-to-Sense; Sense-to-Sound). He authored or coauthored over one hundred publications in scientific journals, books, and conferences. His main research interests are sound modelling for interaction design, sound synthesis by physical modeling, and design and evaluation of interactions.

Dr. Rocchesso is currently chairing the COST Action IC-0601 SID (Sonic Interaction Design).

Sound synthesis and evaluation of interactive footsteps and environmental sounds rendering for virtual reality applications

Rolf Nordahl, Luca Turchet, *Student Member, IEEE* and Stefania Serafin

Abstract—We propose a system that affords real-time sound synthesis of footsteps on different materials. The system is based on microphones, which detect real footstep sounds from subjects, from which the ground reaction force (GRF) is estimated. Such GRF is used to control a sound synthesis engine based on physical models. Two experiments were conducted. In the first experiment, the ability of subjects to recognize the surface they were exposed to was assessed. In the second experiment, the sound synthesis engine was enhanced with environmental sounds. Results show that, in some conditions, adding a soundscape significantly improves the recognition of the simulated environment.

Index Terms—sound and music computing, walking, surface simulation, soundscape rendering.

1 Introduction

In the field of virtual reality, the development of efficient yet accurate simulation algorithms, together with improvements in hardware technology, has boosted research on auditory display and physically based sound models for virtual environments (VEs) [39], [32], [9].

The addition of auditory cues and their importance in enhancing the sense of immersion and presence is a recognized fact in virtual environment research and development. Most prior work in this area has focused on sound delivery methods [36], [33], sound quantity and quality of auditory versus visual information [6] and 3D sound [13], [40]. Recent studies have investigated the role of auditory cues in enhancing self-motion and presence in virtual environments [20], [18], [38].

Self-generated sounds have been often used as enhancements to VEs and first-person 3D computer games, particularly in the form of footstep sounds accompanying self-motion or the presence of other virtual humans.

These sounds are used to produce embodiment and a sense of weight with the overall goal of heightening the sense of "realness" to the character or person. Typically these sounds are taken from sound libraries or recorded by *Foley* artists who put shoes in their hands and interact with different materials to simulate the act of walking.

Recently, several physics based algorithms have been proposed to simulate the sounds of walking. One of the pioneers in this field is Perry Cook, who proposed a collection of physically informed stochastic models (PhiSM) simulating several everyday sonic events [7]. Among these algorithms, the sounds of people walking on different surfaces were simulated [8]. A similar algorithm was also proposed in [12], where physically informed models reproduced several aggregate surfaces. Procedural sound synthesis of walking has also been recently described in [11].

When exploring a place by walking, one can identify two main categories of sound: one's own footsteps and the surrounding soundscape. In the movie industry, footstep sounds represent important elements. Chion writes 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 [5]. Studies on soundscapes originated from the work of R. Murray Schafer [34]. Among other ideas, Schafer proposed soundwalks as empirical methods for identifying soundscapes for specific locations. During a soundwalk people are asked to move in a specific location, paying attention to all the environmental sounds heard. Schafer claimed that each place has a soundmark, i.e., sounds which one identifies a place with. The idea of experiencing a place by listening has been recently further developed by Blesser and Salter [4]. By synthesizing technical, aesthetical and humanistic considerations, the authors describe the field of aural architecture and its importance in everyday life.

The study of human perception of locomotion sounds has addressed several properties of walking sound sources: the gender [22], [15] and posture of a walker [29], the emotions of a walker [15], the hardness and size of the shoe sole [15], and the ground

• *R. Nordahl, L.Turchet and S. Serafin are with the Department of Architecture, Design and Media Technology, Aalborg University Copenhagen, Lautrupvang 15, 2750, Ballerup, DK.*
E-mail: rn,tur,sts@media.aau.dk

material [16].

Such studies have been concerned only with recognition of sounds in an off-line scenario, where subjects were asked to listen to some sounds and classify them.

Concerning interactive footwear, previous work such as research performed by Paradiso and coworkers [28], [3] consisted of designing shoes augmented with sensors used to control footstep sounds. A smaller number of examples, such as recent work of Nordahl [23] and Law et al. [21], have even aimed to provide multimodal cues linked to footsteps events in such environments.

In [35], we presented the first prototype of a system able to synthesize in real-time the sound of footsteps on different materials. This interactive system was evaluated in a between-subjects experiment, where it was compared to a recognition task including recorded and synthesized offline sounds. Results showed that subjects were able to recognize most of the synthesized surfaces with high accuracy. Similar accuracy was also noticed in the recognition of real recorded footstep sounds, which was an indication of the success of the proposed algorithms and their control [25].

In this paper, we are interested in understanding whether the addition of a soundscape enhances the recognition of the simulated surfaces. Our hypothesis is that context plays an important role in the recognition of the material a person is stepping upon.

The faithful simulation of auditory cues is often a neglected aspect in virtual reality research. In this paper we both introduce a physically based sound synthesis engine which simulates footstep sounds, and we describe three experiments which were performed in order to achieve a better understanding of the role of virtual footstep sounds and the surrounding soundscape in simulating an auditory space.

We begin by describing in Section 2 the sound synthesis engine developed and in Section 3 its relative control. In Section 4 we then describe the first experiment which assessed the ability of subjects to recognize the synthesized surfaces, and we then present in Section 5 a second experiment where the simulation of footsteps is enhanced by the addition of environmental soundscapes.

2 THE SOUND SYNTHESIS ENGINE

We developed a physically based sound synthesis engine able to simulate the sounds of walking on different surfaces. Acoustic and vibrational signatures of locomotion are the result of more elementary physical interactions, including impacts, friction, or fracture events, between objects with certain material properties (hardness, density, etc.) and shapes. The decomposition of complex everyday sound phenomena in terms of more elementary ones has been a recurring idea in auditory display research during recent decades [14]. In our simulations, we draw a primary distinction between solid and aggregate ground surfaces, the latter being assumed to possess a granular structure, such as that of gravel, snow or sand.

2.1 Solid surfaces

Sonic interactions between solid surfaces have been extensively investigated, and results are available which describe the relationship between physical and perceptual parameters of objects in contact [19], [39]. These sounds are typically short in duration, with a sharp temporal onset and relatively rapid decay. A good physically motivated synthesis for solid objects is modal synthesis [1], [39], where a vibrating object is modeled by a bank of harmonic oscillators excited by and external stimulus. In this synthesis technique, the response model $h(t)$ is decomposed in terms of the resonant frequencies f_i of the vibrating object, also known as the modes of the object. The response is modeled as a bank of filters with impulse response $h(t) = \sum_i a_i e^{-b_i t} \sin(2\pi f_i t)$, where a_i represent the amplitudes of the modes, b_i the decay rates of the modes, and f_i the frequencies of the modes. Frequencies and decay rates of the modes are determined by the geometry and material properties of the object, and the gains of the modes depend on contact location on the object. A footstep sound may be considered to cause multiple micro-impacts between the sole of a shoe, i.e., an *exciter*, and a floor, i.e., a *resonator*. This interaction can be either discrete, as in the case of walking on a solid surface, or continuous, as in the case of a foot sliding across the floor.

In the simulation of discrete impacts, the external excitation is brief and has an unbiased frequency response. The interaction is modelled by a Hunt-Crossley-type interaction where the force, f, between two bodies, combines hardening elasticity and a dissipation term [17]. Let x represent contact interpenetration and $\alpha > 1$ be a coefficient used to shape the nonlinear hardening, the special model form we used is

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

The model described was discretized as proposed in [2].

If the interaction called for slip, we adopted a model where the relationship between relative velocity v of the bodies in contact and friction force f is governed by a differential equation rather than a static map [10]. Considering that friction results from a large number of microscopic damped elastic bonds with an average deflection z, a viscous term, $\sigma_2 v$, and a noise term, $\sigma_3 w$, to represent roughness, we have

$$f(z, \dot{z}, v, w) = \sigma_0 z + \sigma_1 \dot{z} + \sigma_2 v + \sigma_3 w.$$

The force specified by these models is applied to a virtual mass which produces a displacement signal that

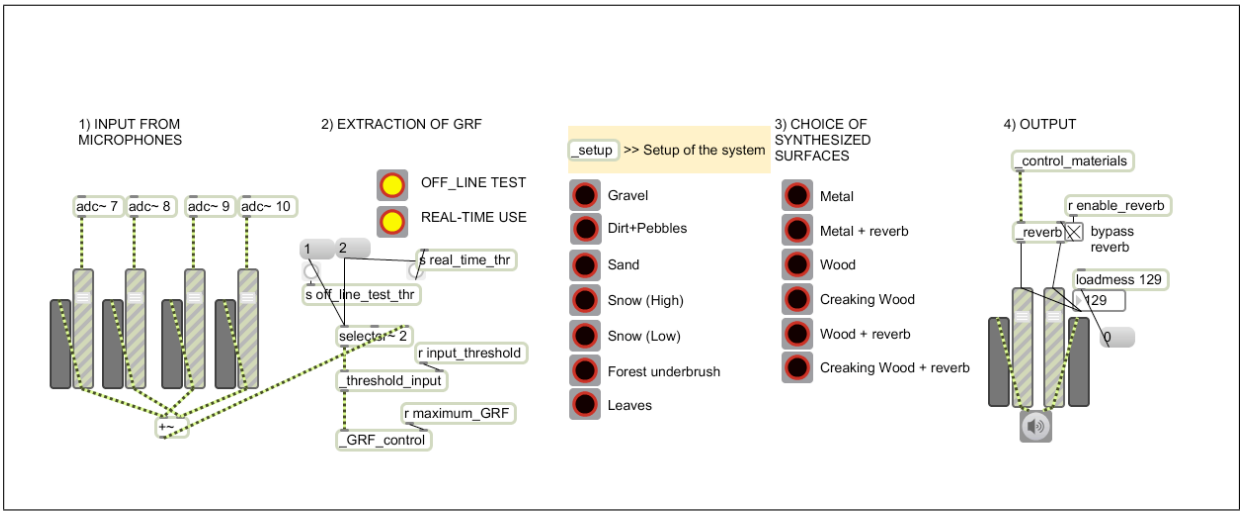


Fig. 1. A screenshot of the graphical user interface for the developed sound synthesis engine.

is then processed by a linear shaping filter intended to represent the resonator.

2.2 Aggregate surfaces

The synthesis algorithms just described are directly applicable only to cases in which the ground surfaces consist of a solid body. To simulate footsteps onto aggregate ground materials such as sand, snow, gravel, a common temporal process based on a stochastic particle model is needed. Different particles do not have to be modeled explicitly, but only the probability that particles will create some noise is simulated.

The simulation is therefore characterized by energies and transition times that depend on the characteristics of the system and the amount of power it absorbs while changing configuration. The simulation dynamically captures macroscopic information about the resulting composite system through time [12].

Specifically, the simulation is performed in terms of the probabilistic distribution of the energies E of the short transients, which can be assumed to follow a power law $p(E) \propto E^{\gamma}$, and a model of the temporal density $N(t)$ of transients as a stationary Poisson process, under the assumption that the inter-transient event times τ are assumed to be independent: $P(\tau) = \lambda e^{\lambda\tau}$ [27]. The value of γ determines the type of noise produced by the process. The parameters γ and λ together determine the macroscopic process dynamics. A simple view of this process is that each transient event consists of a microscopic solid impact with energy E. Thus, in addition, an individual transient can be assumed to possess a resonant response $h(t)$, which is specified in the same way as described above. The resulting parameters characterize each transient event independently of the evolution of the macroscopic system.

Several models of this general type have been developed in order to mimic the sound of a footstep onto aggregate grounds [8], [12], [26].

2.3 Implementation

Using the algorithms described in the previous sections, we implemented a comprehensive collection of footstep sounds. As solid surfaces, we implemented metal and wood. In these materials, the impact model was used to simulate the act of walking, while the friction model was used to simulate the sound of creaking wood.

As aggregate surfaces, we implemented gravel, sand, snow, forest underbrush, dry leaves, pebbles and high grass. The simulated metal, wood and creaking wood surfaces were furthermore enhanced by using some reverberation. Reverberation was implemented by convolving in real-time the footstep sounds with the impulse response recorded in different indoor environments.

The sound synthesis algorithms were implemented in C++ as external libraries for the Max/MSP sound synthesis and multimedia real-time platform.[1] To enable compatibility with the Pure Data platform,[2] the algorithms were implemented using Flext.[3] A screenshot of the final graphical user interface can be seen in Figure 1.

In our simulations, designers have access to a sonic palette making it possible to manipulate all such parameters, including material properties. One of the challenges in implementing the sounds of different surfaces was to find suitable combinations of parameters which provided a realistic simulation. In the synthesis of aggregate materials, parameters such as intensity, arrival times, and impact form a powerful set of independent parametric controls capable of rendering both the process dynamics, which is related to the temporal granularity of the interaction (and linked to the size of the foot, the walking speed, and

1. www.cycling74.com
2. www.puredata.org
3. http://puredata.info/Members/thomas/flext

the walker's weight), and the type of material the aggregate surface is made of. These controls enable the sound designer to choose foot-ground contact sounds from a particularly rich physically-informed palette.

For each simulated surface, recorded sounds were analyzed according to their combinations of events, and each subevent was simulated independently. As an example, the sound produced while walking on dry leaves is a combination of granular sounds with long duration both at low and high frequencies, and noticeable random sounds with not very high density that give to the whole sound a crunchy aspect. These different components were simulated with several aggregate models having the same density, duration, frequency and number of colliding objects.

A pilot test was run to ascertain that global volume plays an important role in the judgments concerning the sounds' realism and in the recognition of the surface material. Indeed, wrong settings for such a parameter can cause wrong recognitions.

The amplitude of the different components were also weighted, according to the same contribution present in the corresponding real sounds. Finally, a scaling factor was applied to the volumes of the different components. This was done in order to recreate a sound level similar to the one happening during a real footstep on each particular material.

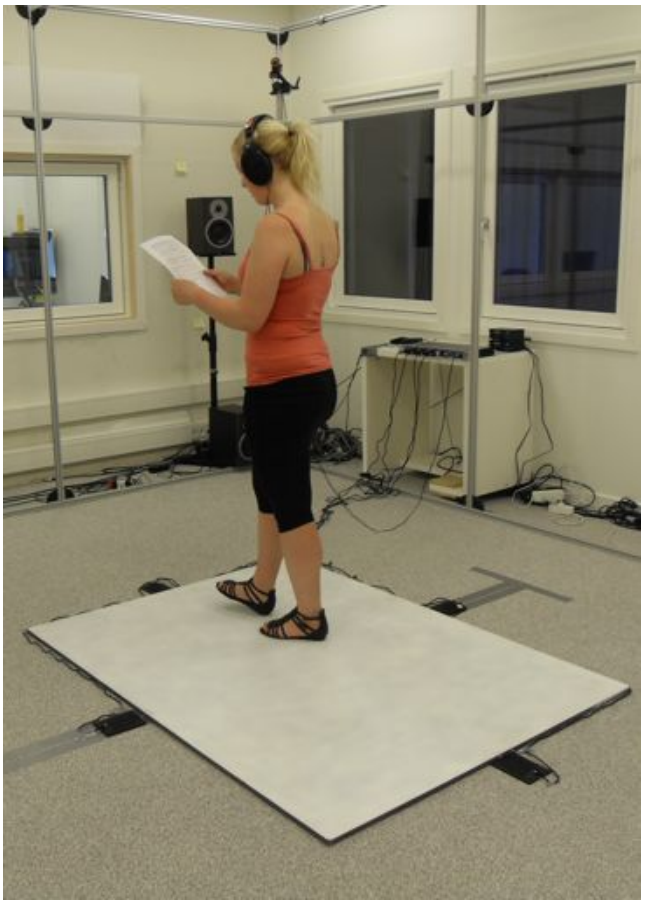


Fig. 2. A subject using the interactive footsteps synthesizer. Notice the MDF board surrounded by the four microphones.

3 Controlling the Sound Synthesis Engine

In the interaction between a foot and a sole, the exciter is usually called ground reaction force (GRF), i.e., the reaction force supplied by the ground at every step. The developed sound synthesis engine is controlled in realtime by extracting the GRF while subjects perform the act of walking. Specifically, as can be seen in Figure 2, subjects are asked to walk on a medium density fiberboard (MDF) which is 2.5 x 2 m in size and 1 cm thick. Four microphones are placed on the sides of the fiberboard in a square configuration. Specifically, we used four Shure BETA 91, high performance condenser microphones with a tailored frequency response designed specifically for kick drums and other bass instruments. The microphones' features made them a good candidate for the purpose of capturing footstep sounds. A multichannel Fireface 800 soundcard receives audio signals from the microphones, and transmits it via Firewire to a Macbook Pro laptop. Here their GRF is estimated from the input signal. In order to perform a more precise estimation, among the four footstep sounds detected by the four microphones, we choose the one which is higher in amplitude.

To calculate the GRF we use a simple non-linear low-pass filter proposed by Cook in [30]:

$$e(n) = (1 - b(n))|x(n)| + b(n)e(n-1)$$

where

$$b = \begin{cases} b_{up} & \text{if } |x(n)| > e(n-1) \\ b_{down} & \text{otherwise} \end{cases}$$

An example of a footstep sound and its corresponding GRF is shown in Figure 3. The extracted GRF is used to control the force parameter of the different sound synthesis engines. For example, as shown in Figure 4, the extracted GRF is used to control in real-time the force parameter of the impact model. Finally the synthesized sounds, as well as the soundscapes, are provided to the user by means of a set of Beyerdynamic DT-770 headphones. Figure 5 shows a schematic representation of the different hardware components used.

Informal tests with several participants during the development stage show that no perceived latency is experienced. Moreover, the amplitude of the synthetic footsteps is high enough to mask the one of the real footsteps, in such a way that subjects perceive only the synthetic ones. The kind of shoes which the subjects are wearing affect the overall shape of the

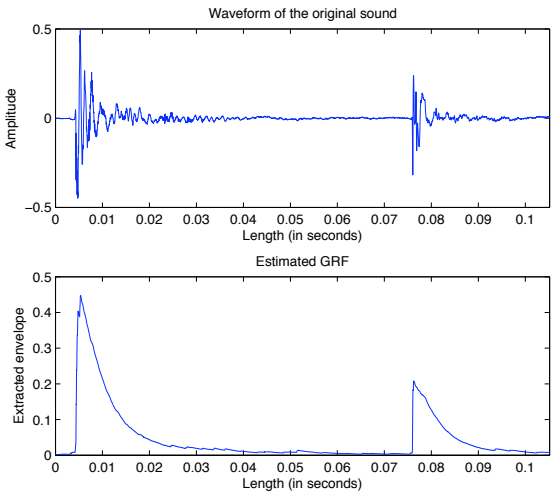


Fig. 3. A footstep sound (top) and the corresponding calculated GRF (bottom).

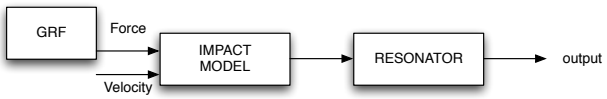


Fig. 4. Block diagram representing the algorithm and control of solid material sounds.

GRF, which is mapped to the amplitude of the sound synthesis engine. Therefore the kind of shoes worn by the subjects is reflected in the synthesis engine. This is is a desired feature of the engine, since it matches the real world experience were different shoes produce different sounds when people are walking on the same surface.

4 Experiment 1

In the first experiment, we investigated the ability of subjects to recognize the different walking sounds they were exposed to. In this experiment, subjects were asked to recognize the sounds in an active setting, i.e., by using the developed interactive system.

4.1 Method

The sounds provided during the experiment were synthesized sounds generated in real time while subjects were walking using the interactive system described in the previous section.

Participants were exposed to 26 trials, where 13 stimuli were presented twice in randomized order. The stimuli consisted of footstep sounds on the following surfaces: beach sand, gravel, dirt plus pebbles (like in a country road), snow (in particular deep snow), high grass, forest underbrush (a forest floor composed by dirt, leaves and branches breaking), dry leaves, wood, creaking wood and metal. To simulate

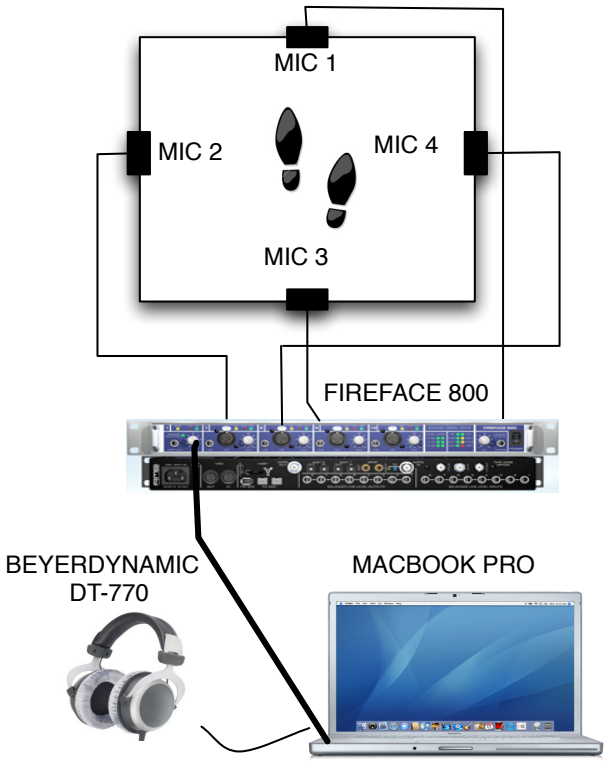


Fig. 5. Hardware components of the system developed: microphones, multichannel soundcard, laptop and headphones.

room characteristics, footstep sounds on wood, creaking wood and metal were enhanced adding a certain amount of reverberation.

4.1.1 Participants

15 participants (6 men and 9 women), aged between 19 and 29 (mean=22.13, standard deviation=2.47), took part to the experiment. All participants reported normal hearing conditions. All participants were naive with respect to the experimental setup and to the purpose of the experiment.

The shoes used by subjects were sneakers, trainers, boots and other kinds of shoes with rubber sole. The participants took on average 24 minutes to complete the experiment.

4.1.2 Setup

All experiments were carried out in an acoustically isolated laboratory where the setups were installed. The setup consisted of the interactive system, and the participants were asked to use it in order to generate the footstep sounds in real time (see Figure 2).

4.1.3 Task

Participants were asked to wear a pair of headphones and to walk in the area delimited by the microphones.

They were given a list of different surfaces to be held in one hand, presented as non-forced alternate choice. The list of surfaces presented to the subjects is outlined in the first row of Table 1. It represents an extended list of the surfaces the subjects were exposed to.

At the end of the experiment, subjects were asked some questions concerning the naturalness of the interaction with the system.

4.2 Results

Table 1 shows the confusion matrix which displays the results of Experiment 1. The matrix shows information concerning actual classifications performed by the subjects. The first row represents the list of materials the subjects could choose from, while the first column represents the actual stimuli subjects were exposed to. The choice of having a wider list of materials was made to avoid as much as possible that subjects simply guessed a choice.

From this Table, it is possible to notice how surfaces such as snow, creaking wood with and without reverberation, gravel and metal with reverberation were correctly recognized in a high number of instances. Recognition of surfaces such as dirt plus pebbles, high grass and wood appeared to be wrong most of the times. An analysis performed on the wrong answers reveals that in average subjects tended to classify erroneously a surface as another belonging to a same category (e.g., wood versus concrete, snow versus frozen snow, dry leaves versus forest underbrush) rather than to different categories (e.g., wood versus water, wood versus gravel, metal versus dry leaves). Moreover, results show that the addition of the reverberation to the sounds gave rise to better recognitions for metal, and worse for wood plus reverberation, which was perceived most of the times as concrete (not tangible differences were found for the creaking wood). Overall, recognition rates are similar to those measured on recorded footstep sounds [25].

5 Experiment 2

The first experiment does not take into account the fact that when walking in a space, either indoor or outdoor, subjects are exposed not only to the sounds of their own footsteps, but also to sounds of the environment. Our hypothesis is that environmental sounds play an important role in the recognition of the surface subjects are stepping upon. We tested our hypothesis by conducting an experiment whose goal was to investigate the ability of subjects to recognize the different walking sounds they were exposed to under three conditions: without an accompanying soundscape, with a coherent soundscape and with an incoherent soundscape. In this experiment, we define as coherent soundscape a soundscape that matches the expectations of the subjects. Specifically, soundscapes of the following environments were designed:

1) A beach and seaside during the summer
2) A courtyard of a farm in the countryside
3) A ski slope
4) A forest
5) A park during the fall

The specific content of the soundscapes was designed according to the indications given by subjects answering to a questionnaire. Precisely, ten subjects, chosen among those not partecipating to the final experiment, were asked to imagine which sounds could occur in the above mentioned environments. Specifically, subjects were asked the following question: "Imagine that you are right now in a forest: which sounds do you think you would hear?" In this particular environment, subjects indicated sounds like trees, birds, different animals. Among the answers provided, we chose those which were stated by more than one subject, and collected a corresponding sound material using recordings of real sounds.

The most occurring named sound events were used as sonic elements to build the final soundscape. The sounds were chosen among those available both on the Hollywood Edge sound effects library[4] and on the Freesound website[5]. The chosen sounds were using the sound editor Adobe Audition 3[6]. Soundscapes were designed with the goal of providing a clear idea of the designed environment already from the first seconds. As an example, in the soundscape of the forest animals characteristic of that environment, such as the cockoo, could be heard, and the density of natural sounds is higher as compared to the soundscape of a park. The footstep sounds provided during the three conditions were synthesized sounds generated in real time while subjects were walking using the interactive system described previously. The soundscapes were audio files played in background independently from the subjects movements. The amplitudes of both footsteps and soundscapes were set by empirical investigation.

5.1 Method

A between-subject experiment with the following three conditions was conducted:

1) Condition 1: footstep sounds without soundscapes.
2) Condition 2: footstep sounds with coherent soundscapes.
3) Condition 3: footstep sounds with incoherent soundscapes.

4. http://www.hollywoodedge.com
5. http://www.freesound.org/
6. http://www.adobe.com/products/audition/

	BS	GL	DR	SW	HG	UB	DL	WD	CW	MT	WR	CR	MR	FS	CC	PD	WT	CP	—
BS	15	2		5			1							2				1	4
GL		21	2			1	1							4					1
DR		1	3	2		6	6		1					10					1
SW				24	1									4					1
HG	2	7	3	1	0	3	7			2									5
UB		1	3	1		19	1							3				1	1
DL	1	3	5			5	12							4					
WD			1	2				14		1						1	1		10
CW								1	28					1					
MT								1		24				1	2				2
WR									3	11	6				7				3
CR												28			1				1
MR								1					25	1	1				2

Legend:
WD wood — CW creaking wood — SW snow — UB underbrush
— don't know — FS Frozen snow — BS beach sand — GL Gravel
MT metal — HG High grass — DL dry leaves — CC concrete
DR dirt — PD puddles — WT Water — CP carpet
WR wood reverb — MR metal reverb — CR creaking+ reverb

TABLE 1
Confusion matrix for experiment 1: recognition of synthesized footstep sounds.

The first condition was meant as a validation of the results of experiment 1, but using a smaller set of surfaces.

Subjects were exposed to ten trials in conditions 1 and 2, where five stimuli were presented twice in randomized order. The stimuli in condition 1 consisted of footstep sounds on the following surfaces: beach sand, gravel, snow (in particular deep snow), forest underbrush (a forest floor composed by dirt, leaves and branches breaking), dry leaves. In condition 2 the stimuli consisted of the same footstep sounds provided in condition 1 with in addition the corresponding coherent soundscape described in section 5.

In condition 3, subjects were exposed to six stimuli presented twice in randomized order. The stimuli consisted of footstep sounds on the surfaces beach sand, snow, forest underbrush, with in addition an incoherent soundscape. As an example of incoherent soundscape, in presence of footstep sounds on beach sand, the provided soundscape corresponded to that of a landscape with snow (i.e., a ski slope) and forest underbrush (i.e., the forest environment). The choice of using sand, snow and forest underbrush as surfaces was justified by the fact that we wanted to select three very distinct outdoor landscapes (beach, mountain and forest).

5.1.1 Participants

Forty three participants were divided in three groups to perform the three conditions in a between-subjects experiment (n = 15, n = 15 and n = 13 respectively). The three groups were composed respectively of 11 men and 4 women, aged between 21 and 28 (mean=23.67, standard deviation=2.12), 8 men and 7 women, aged between 19 and 38 (mean=24.67, standard deviation=5.97), and 6 men and 7 women, aged between 21 and 30 (mean=24, standard deviation=3.1). All participants reported normal hearing conditions. All participants were naive with respect to the experimental setup and to the purpose of the experiment.

During the experiment the shoes used by subjects were sneakers, trainers, boots and other kinds of shoes with rubber sole. The participants took on average 11, 13 and 16 minutes in average for conditions 1, 2 and 3 respectively.

5.1.2 Setup

The experiment was carried out in the same acoustically isolated laboratory as of the previous experiment, where the MDF delimited by four microphones placed in a square configuration was installed (see Figure 2).

5.1.3 Task

During the experiment subjects were asked to wear a pair of headphones and to walk on the MDF in the area delimited by the microphones. They were given a list of different surfaces to be held in one hand, presented as non-forced alternate choice. The list included a range of materials wider than those presented in the experiment. As before, the choice of having a wider list of materials was made to avoid as much as possible that subjects simply guessed a choice.

During the act of walking subjects listened simultaneously to footstep sounds on a different surface according to the stimuli presented. The task, common to the three conditions, consisted of orally answering the following three questions after the presentation of the stimulus:

1) Which surface do you think you are walking on? For each stimulus choose an answer in the following list: 1) beach sand, 2) gravel, 3) dirt, 4) snow, 5) high grass, 6) forest underbrush, 7) dry leaves, 8) wood, 9) creaking wood, 10) metal,

11) carpet, 12) concrete, 13) frozen snow, 14) puddles, 15) water, 16) I don't know.
2) How close to real life is the sound in comparison with the surface you think it is? Evaluate the degree of realism on a scale from 1 to 7 (1=low realism, 7=high realism).
3) Evaluate the quality of the sound on a scale from 1 to 7 (1=low quality, 7=high quality).

In conditions 2 and 3, participants were also asked to recognize what was the environment they were walking on.
They were informed that they could choose the same material more than once and that they were not forced to choose all the materials in the list. In addition, they could use the interactive system as much as they wanted before providing an answer. After moving to the next stimulus, subjects could not change their answers to the previous stimuli.
At the end of the experiment, the subjects were also given the opportunity to leave an open comment on their experience interacting with the system.

5.2 Results

The collected answers were analyzed and compared between the three conditions. Results are shown in Tables 2, 3 and 4.

The first noticeable element emerging from the three tables is that the use of the interactive system in the condition of coherent soundscapes gave rise to a higher recognition rate and a higher evaluation of realism and quality of the proposed sounds, compared to the conditions with no soundscapes and with incoherent soundscapes. Concerning the percentages of correct answers, they are higher for condition 2 compared to condition 1, and this is true for every surface. The analysis by means of a chi-square test reveals that differences are statistically significant for beach sand ($p < 0.01$) and forest underbrush ($p = 0.019$). Especially in the condition where subjects were exposed to a stimulus representing a footstep on a forest underbrush, the soundscape was extremely important in facilitating the recognition. As a matter of fact, the soundscape, contained several elements characteristics of a forest, such as birds singing, which represented important cues for the subjects. It is particularly interesting to notice that overall adding a soundscape enhances the recognition factor, and this is especially noticeable for those situations where the recognition was rather low without a soundscape.

Similarly, the percentages of correct answers are higher for condition 2 compared to condition 3 for each surface. Differences are statistically significant for beach sand ($p < 0.01$), snow ($p = 0.0144$) and forest underbrush ($p < 0.01$). Furthermore, the percentages of correct answers are higher for condition 1 compared to condition 3, for each surface, but the differences are not statistically significant.

The analysis of the wrong answers reveals that in all the experiments none of the presented surfaces was recognized as a solid surface. This means that all subjects were able to identify at least the nature of the surface, which was an expected feature of the simulations. After the experiment, several subjects observed that the simulated footstep sounds were perceived as very similar, and therefore hard to recognize and distinguish from the list provided.

It is interesting to examine what happens when subjects are exposed to soundscapes which are incoherent, as shown in Table 4. In this situation, we consider as correct the answer provided when subjects recognize the surface they are walking on, and not the soundscape. As it can be noticed, the percentage of correct answers is quite low. As expected, adding an incoherent soundscape creates a stronger context which often confuses the subjects. This can be observed, for example, in the case of beach sand footsteps which were rendered together with a forest soundscape and a ski slope soundscape. The recognition rate is higher in the first case than in the second, where several subjects confused sand with snow. The answers for the three conditions are outlined in the confusion matrices shown in Table 5, 6, 7 respectively. As before, the first row represents the list of materials the subjects could choose from, while the first column represent the actual stimuli subjects were exposed to.

From the matrices, it can be noticed how the subjects' recognition varies from condition 1 to condition 2. As an example, the second row of the matrix illustrates the number of subjects which recognized the beach sand surface, with (5) and without (6) a soundscape. The role of the soundscape to enhance the recognition is clearly noticeable.

Table 7 illustrates the confusion matrix for condition 3, i.e., when incoherent soundscapes are presented to the subjects. In this situation, it is clearly noticeable how the nature of the soundscape plays an important role. Moreover, it can be noticed how the incoherent soundscape is in most situation predominant, in the sense that subjects tend to judge the surface they are stepping upon more listening to the soundscape than listening to the actual surface. On the other hand, even if subjects are not able to recognize the surface they are stepping upon, they never confuse its nature, in the sense that they never select a solid surface when exposed to an aggregate one.

In addition, Tables 2, 3 and 4 show the degree to which participants judged the realism and quality of the experience. The degree of realism was calculated by looking only at that data from correct answers, i.e., when the surfaces were correctly recognized. This choice was performed since we were interested in understanding whether the simulation of specific surfaces recognized by the subjects was satisfactory.

As far as the quality judgement is concerned, the data was based on all the answers different from "I

	% Correct answers	% Wrong answers	% "I don't know"	Realism	Quality
Beach Sand	50.00	46.67	3.33	5.20	4.72
Gravel	83.33	6.67	10.00	5.20	4.63
Snow	73.33	26.67	0.00	5.30	5.12
Forest underbrush	40.00	50.00	10.00	3.50	4.19
Dry Leaves	16.67	63.33	20.00	4.40	3.92

TABLE 2
Results of condition 1, experiment 2: recognition of the surfaces without soundscapes.

	% Correct answers	%Wrong answers	% "I don't know"	Realism	Quality	% Correct soundscape
Beach Sand	86.67	10.00	3.33	5.15	5.52	93.33
Gravel	86.67	13.33	0.00	5.31	5.40	86.67
Snow	80.00	20.00	0.00	6.17	6.07	83.33
Forest underbrush	73.33	26.67	0.00	4.91	5.03	100.00
Dry Leaves	30.00	70.00	0.00	4.44	4.50	96.67

TABLE 3
Results of condition 2, experiment 2: recognition of the surfaces with coherent soundscapes.

Material	Soundscape	% Correct answers	% Wrong answers	% No idea	Realism	Quality	% Correct soundscape
Beach Sand	Forest	38.46	57.70	3.84	4.00	5.16	88.46
Beach Sand	Ski slope	15.38	69.24	15.38	3.75	4.32	38.46
Snow	Forest	50.00	42.31	7.69	5.29	5.35	96.15
Snow	Beach	50.00	46.16	3.84	4.77	5.00	88.46
Forest underbrush	Beach	38.46	53.85	7.69	4.80	4.96	65.38
Forest underbrush	Ski slope	30.76	65.40	3.84	4.30	4.61	46.15

TABLE 4
Results of condition 3, experiment 2: recognition of the surfaces with incoherent soundscapes.

don't know".

The mean of realism is higher for condition 2 compared to condition 1 for each surface with the exception of beach sand (which is almost equal). The analysis by means of a t-test reveals that differences are statistically significant for snow ($p = 0.0105$) and forest underbrush ($p < 0.01$). Analogously, the mean of realism is higher for condition 2 compared to condition 3 for each surface with the exception of beach sand (which is almost equal). In particular, the differences are statistically significant for beach sand ($p < 0.01$) and snow ($p < 0.01$). Moreover, the mean of realism is higher for condition 1 compared to condition 3 for each surface with the exception of forest underbrush, which is minor. Differences are statistically significant for beach sand ($p < 0.01$), and for forest underbrush ($p = 0.0344$), which, as said, is greater for experiment 3.

As regards the mean of quality, it is higher for condition 2 compared to condition 1, with statistically significant differences for all the surfaces with the exception of dry leaves: beach sand ($p < 0.01$), gravel ($p = 0.0217$), snow ($p < 0.01$) and forest underbrush ($p = 0.0219$). The mean of quality is higher for condition 2 compared to condition 3 for each surface, and in particular the differences are statistically significant for beach sand ($p < 0.01$), for snow ($p < 0.01$). Furthermore, the mean of quality is similar for condition 1 compared to condition 3, with the exception of forest underbrush for which it is higher in condition 3 compared to condition 1, with statistically significant differences ($p = 0.032$).

The comparison about the percentages of "I don't know" answers reveals that for each surface they are higher for condition 1 compared to condition 2, and for condition 3 compared to condition 2. In addition, they are higher for condition 3 compared to condition 1, for each surface with the exception of forest underbrush (which is minor).

As regards to the percentages of correct answers about the soundscapes presented, they are higher for condition 2 compared to condition 3, and in particular the differences are statistically significant for the ski slope soundscape ($p < 0.01$).

Overall, subjects observed that soundscapes play an important role in recognition of the surfaces, precisely for their ability to create a context. Especially in terms of conflicting cues, as it was the case in condition 3, subjects were trying to identify the strongest cues, i.e., the element which had the strongest recognition

	BS	GL	SW	UB	DL	HG	DR	FS	WD	CW	MT	CC	PD	WT	CP	—
BS	15	2	6	2	2	2										1
GL		25			1								1			3
SW			22	1	1			6								
UB			4	12			1	10								3
DL		9		4	5			6								6

Legend: WD wood, CW creaking wood, SW snow, UB underbrush, — don't know, FS Frozen snow, BS beach sand, GL Gravel, MT metal, HG High grass, DL dry leaves, CC concrete, DR dirt, PD puddles, WT Water, CP carpet

TABLE 5
Confusion matrix for condition 1, experiment 2: recognition of footstep sounds without a soundscape.

	BS	GL	SW	UB	DL	HG	DR	FS	WD	CW	MT	CC	PD	WT	CP	—
BS	26	1	1	1												1
GL		26			3	1										
SW	2		24					4								
UB		1	2	22	1			4								
DL	6			9	9			6								

Legend: WD wood, CW creaking wood, SW snow, UB underbrush, — don't know, FS Frozen snow, BS beach sand, GL Gravel, MT metal, HG High grass, DL dry leaves, CC concrete, DR dirt, PD puddles, WT Water, CP carpet

TABLE 6
Confusion matrix for condition 2, experiment 2: recognition of footstep sounds with a coherent soundscape.

	Soundscape	BS	GL	SW	UB	DL	HG	DR	FS	WD	CW	MT	CC	PD	WT	CP	—
BS	Forest	10		2	5	3	1	1	3								1
BS	Ski slope	4	1	5		1	5	3	1					1	1		4
SW	Forest		5	13	1	1			4								2
SW	Beach	2	6	13	2				2								1
UB	Beach		3	4	10	2			5								2
UB	Ski slope		1	1	8	3	1	2	8							1	1

Legend: WD wood, CW creaking wood, SW snow, UB underbrush, — don't know, FS Frozen snow, BS beach sand, GL Gravel, MT metal, HG High grass, DL dry leaves, CC concrete, DR dirt, PD puddles, WT Water, CP carpet

TABLE 7
Confusion matrix for condition 3, experiment 2: recognition of footstep sounds with an incoherent soundscape.

factor. Sometimes the subjects found this task quite hard to complete, and this is why the percentage of "I don't know" answers is higher in condition 3 as opposed to condition 2.

When leaving a comment, several subjects observed that the recognition of snow was extremely realistic. This observation is also confirmed by the high degree of realism (mean = 5.3) and quality (mean = 5.1) with which the surface was rated.

On the other hand, for some subjects the concept of dry leaves was rather confusing, and this is also confirmed by the low recognition rate of that surface.

Overall, this experiment represents a strong indication of the importance of context in the recognition of a virtual auditory place, where self sounds created by users' footsteps and soundscapes are combined. Further investigations are needed to enhance the realism of the simulated soundscape, in particular by having the auditory cues changing according to the motion of the subject in the space.

6 Conclusions and Future Work

In this paper, we describe a system able to sonically simulate the act of walking on different virtual spaces. Two experiments are also presented, whose goal is to understand the ability of subjects to recognize the surfaces they were exposed to, and the role of soundscape design in creating a sense of place and context when implementing a virtual auditory walking experience.

In the second experiment described, the user was not able to interact with the soundscapes, which were created using pre-composed soundtracks. This slightly reduces the possibility of reproducing a natural experience. As an example, in the real world subjects can affect the surrounding soundscape for example by hitting sound producing objects with

their feet while walking. On the other hand, subjects could interact with the physically simulated sounds of footsteps generated by their own act of walking.

Results indicate the ability of subjects to recognize the virtually simulated surfaces, together with the importance of the addition of environmental sounds in enhancing the recognition. As shown in [31], results are consistent with those obtained when exposing subjects to real surfaces, which is an indication of the quality of the footsteps synthesizer. The experiments also stimulated some interesting observations regarding the interface adopted and the way the sound synthesis algorithms were controlled. As can be seen in Figure 2, the board used is limited in size and does not allow a natural walking experience. In future investigations, we indeed decided to control the sound synthesis algorithms by using shoes enhanced with sensors [24]. This choice was motivated by the fact that we wanted to enhance the simulation with haptic feedback, provided by some actuators placed on the sole of the shoes. The use of haptic feedback on the shoes allows a realistic simulation of the tactile sensation of walking on different surfaces, since this is quite different in solid versus aggregate surfaces, as an example. The switch to shoes enhanced with sensors, however, also allowed a wider walking possibility for the subjects, now limited only by the actual size of the laboratory. By enhancing the simulation of auditory feedback with haptic feedback, experiments with both coherent and incoherent stimuli were performed, in order to understand the role of the auditory and haptic modality in recognizing the virtual surfaces subjects are exposed to. Results show that subjects perform better in the recognition task when exposed to auditory feedback rather than haptic feedback, and the combination of auditory and haptic feedback does not significantly enhance the recognition [24], [37].

Overall, the results described are an interesting starting point for further investigations on the role of self-sounds and environmental sounds to create a sense of place in a virtual environment. While walking an acting in an environment, a person is exposed to her own self-sounds as well as the sounds of the place. This paper presents a preliminary investigation of the role of these different elements both taken in isolation and combined. Further investigations are needed to gain a better understanding of the cognitive factors involved when subjects are exposed to different sound events, especially when a situation of semantic incongruence is present.

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 (www.niwproject.eu).

REFERENCES

[1] J. Adrien. *The missing link: Modal synthesis*. MIT Press Cambridge, MA, USA, 1991.

[2] F. Avanzini and D. Rocchesso. Modeling collision sounds: Non-linear contact force. In *Proc. COST-G6 Conf. Digital Audio Effects (DAFx-01)*, pages 61–66, 2001.

[3] A. Benbasat, S. Morris, and J. Paradiso. A wireless modular sensor architecture and its application in on-shoe gait analysis. In *Sensors, 2003. Proceedings of IEEE*, volume 2, 2003.

[4] B. Blesser and L. Salter. *Spaces speak, are you listening?: experiencing aural architecture*. MIT Press, 2006.

[5] M. Chion and C. Gorbman. *Audio-vision*. Nathan, 1990.

[6] P. Chueng and P. Marsden. Designing Auditory Spaces to Support Sense of Place: The Role of Expectation. In *CSCW Workshop: The Role of Place in Shaping Virtual Community*. Citeseer, 2002.

[7] P. Cook. Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. *Computer Music Journal*, 21(3):38–49, 1997.

[8] 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.

[9] P. Cook. *Real sound synthesis for interactive applications*. AK Peters, Ltd., 2002.

[10] P. Dupont, V. Hayward, B. Armstrong, and F. Altpeter. Single state elastoplastic friction models. *IEEE Transactions on Automatic Control*, 47(5):787–792, 2002.

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

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

[13] J. Freeman and J. Lessiter. Hear there & everywhere: the effects of multi-channel audio on presence. *Proceedings of ICAD 2001*, pages 231–234, 2001.

[14] W. Gaver. What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology*, 5(1):1–29, 1993.

[15] B. Giordano and S. Mcadams. Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates. *The Journal of the Acoustical Society of America*, 119:1171, 2006.

[16] B. Giordano, S. Mcadams, Y. Visell, J. Cooperstock, H. Yao, and V. Hayward. Non-visual identification of walking grounds. *Journal of the Acoustical Society of America*, 123(5):3412, 2008.

[17] K. H. Hunt and F. R. E. Crossley. Coefficient of restitution interpreted as damping in vibroimpact. *ASME Journal of Applied Mechanics*, 42(2):440–445, 1975.

[18] B. Kapralos, D. Zikovitz, M. Jenkin, and L. Harris. Auditory cues in the perception of self-motion. In *116th AES convention*. Citeseer, 2004.

[19] R. Klatzky, D. Pai, and E. Krotkov. Perception of material from contact sounds. *Presence: Teleoperators & Virtual Environments*, 9(4):399–410, 2000.

[20] P. Larsson, D. Västfjäll, and M. Kleiner. Perception of self-motion and presence in auditory virtual environments. In *Proceedings of seventh annual workshop presence*, pages 252–258, 2004.

[21] A. Law, B. Peck, Y. Visell, P. Kry, and J. Cooperstock. A Multimodal Floor-space for Experiencing Material Deformation Underfoot in Virtual Reality. In *IEEE International Workshop on Haptic Audio visual Environments and Games, 2008. HAVE 2008*, pages 126–131, 2008.

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

[23] R. Nordahl. Increasing the motion of users in photorealistic virtual environments by utilizing auditory rendering of the environment and ego-motion. *Proceedings of Presence*, pages 57–62, 2006.

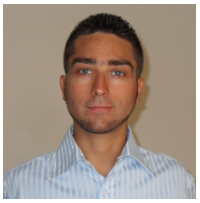
[24] R. Nordahl, A. Berrezag, S. Dimitrov, L. Turchet, V. Hayward, and S. Serafin. Preliminary Experiment Combining Virtual

Reality Haptic Shoes And Audio Synthesis. *Haptics: Generating and Perceiving Tangible Sensations*, pages 123–129, 2010.
[25] R. Nordahl, S. Serafin, and L. Turchet. Sound Synthesis and Evaluation of Interactive Footsteps for Virtual Reality Applications. In *Proc. IEEE Virtual Reality*, 2010.
[26] S. O'Modhrain and G. Essl. Pebblebox and crumblebag: Tactile interfaces for granular synthesis. In *Proc. NIME 2004*, pages 74–79, Hamamatsu, Japan, 2004.
[27] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 2nd edition, 1984.
[28] J. Paradiso, K. Hsiao, and E. Hu. Interactive music for instrumented dancing shoes. In *Proc. of the 1999 International Computer Music Conference*, pages 453–456, 1999.
[29] R. Pastore, J. Flint, J. Gaston, and M. Solomon. Auditory event perception: The source-perception loop for posture in human gait. *Perception and Psychophysics*, 70(1):13, 2008.
[30] L. Peltola, C. Erkut, P. Cook, and V. Valimaki. Synthesis of hand clapping sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(3):1021–1029, 2007.
[31] S. S. R. Nordahl and L. Turchet. Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In *Proceedings of IEEE Virtual Reality*, 2010.
[32] N. Raghuvanshi and M. Lin. Interactive sound synthesis for large scale environments. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games*, pages 101–108. ACM New York, NY, USA, 2006.
[33] R. Sanders Jr. *The effect of sound delivery methods on a users sense of presence in a virtual environment*. PhD thesis, NAVAL POSTGRADUATE SCHOOL, 2002.
[34] R. Schafer. *The tuning of the world*. Alfred A. Knopf, 1977.
[35] S. Serafin, L. Turchet, and R. Nordahl. Extraction of ground reaction forces for real-time synthesis of walking sounds. In *Proc. Audiomostly*, 2009.
[36] R. Storms and M. Zyda. Interactions in perceived quality of auditory-visual displays. *Presence: Teleoperators & Virtual Environments*, 9(6):557–580, 2000.
[37] L. Turchet, S. Serafin, S. Dimitrov, and R. Nordahl. Conflicting Audio-haptic Feedback in Physically Based Simulation of Walking Sounds. In *Haptic and Audio Interaction Design: 5th International Workshop, HAID 2010, Copenhagen, Denmark, September 16-17, 2010, Proceedings*, page 97. Springer, 2010.
[38] A. Väljamäe, P. Larsson, D. Västfjäll, and M. Kleiner. Travelling without moving: Auditory scene cues for translational self-motion. In *Proceedings of ICAD05*, 2005.
[39] K. Van Den Doel, P. Kry, and D. Pai. FoleyAutomatic: physically-based sound effects for interactive simulation and animation. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 537–544. ACM New York, NY, USA, 2001.
[40] D. Västfjäll. The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments. *CyberPsychology & Behavior*, 6(2):181–188, 2003.

Rolf Nordahl Rolf Nordahl, born 1974, is currently assistant professor in Medialogy at Aalborg University Copenhagen. His research interests lie within VR, (Tele)-Presence, Sonic Interaction Design, and developing new methods and evaluation-techniques for VR, Presence and Games.

He is principal investigator for the EU funded project Natural Interactive Walking, and has earlier done seminal work on the EU-project BENOGO (the project was focused on HMD-based photo-realistic VR). Likewise he is a recognized member of the expert panel under the Danish Evaluation Institute as well as being member of various steering committees.

He frequently publishes both journal and conference papers and is also giving talks at international level. Lately Rolf Nordahl was invited to run a special series of lectures at among other places Yale University (Connecticut).



Luca Turchet 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 a 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 Stefania Serafin is Associate Professor in sound modeling at Aalborg University in 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.

AUDITORY FEEDBACK IN A MULTIMODAL BALANCING TASK: WALKING ON A VIRTUAL PLANK

Stefania Serafin
Department of Architecture, Design and Media Technology
Aalborg University Copenhagen
sts@create.aau.dk

Luca Turchet
Department of Architecture, Design and Media Technology
Aalborg University Copenhagen
tur@create.aau.dk

Rolf Nordahl
Department of Architecture, Design and Media Technology
Aalborg University Copenhagen
rn@create.aau.dk

ABSTRACT

We describe a multimodal system which exploits the use of footwear-based interaction in virtual environments. We developed a pair of shoes enhanced with pressure sensors, actuators, and markers. Such shoes control a multichannel surround sound system and drive a physically based sound synthesis engine which simulates the act of walking on different surfaces. We present the system in all its components, and explain its ability to simulate natural interactive walking in virtual environments.

The system was used in an experiment whose goal was to assess the ability of subjects to walk blindfolded on a virtual plank. Results show that subjects perform the task slightly better when they are exposed to haptic feedback as opposed to auditory feedback, although no significant differences are measured. The combination of auditory and haptic feedback does not significantly enhances the task performance.

1. INTRODUCTION

In the academic community, foot-based interactions have mostly been concerned with the engineering of locomotion interfaces for virtual environments [1]. A notable exception is the work of Paradiso and coworkers, who pioneered the development of shoes enhanced with sensors, able to capture 16 different parameters such as pressure, orientation, acceleration [2]. Such shoes were used for entertainment purpose as well as for rehabilitation studies [3]. The company Nike has also developed an accelerometer which can be attached to running shoes and connected to an iPod, in such a way that, when a person runs, the iPod tracks and reports different information.

In this paper we mostly focus on enhancing our awareness of auditory and haptic feedback in foot based devices, topic which is still rather unexplored.

We describe a multimodal interactive space which has been developed with the goal of creating audio-haptic-visual simulations of walking-based interactions. The system requires users to walk around a space wearing a pair of shoes enhanced with sensors and actuators. The position of such shoes is tracked by a motion capture system, and the shoes drive a audio-visual-haptic synthesis engine based on physical models. The idea of enhancing shoes with sensors and actuators is similar to the ones we have been exploring in the context of the Natural Interactive Walking (NIW) FET-Open EU project [1] [7, 8]. The ultimate goal of this project 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.

As a case study of the developed architecture, we describe an experiment where subjects were asked to walk straight on a virtual plank. The use of audio-haptic augmented footwear for navigation has not been extensively explored in the research community. An exception are the Cabboots [4], a pair of actuated boots which provide information concerning the shape of a path.

As another example, recently Takeuchi proposed Gilded Gait, a system which changes the perceived physical texture of the ground [5]. The Gilded Gait system is designed as a pair of insoles with vibrotactile feedback to simulate ground textures.

Recently, it has also been demonstrated that walking straight is a hard task also in the physical world. The research was justified by the common belief that people, when getting lost, tend to walk into circles. Subjects were asked to walk straight in two outdoor environments: a forest and a desert. When subjects were not able to see the sun, they walked in circles. It was suggested that veering from a straight course is the result of accumulating noise in the sensorimotor system, which, without an external directional reference to recalibrate the subjective straight ahead [6]. In this paper, we investigate the ability of subjects to walk straight on a narrow virtual plank, with the help of auditory and visual feedback. The results of this research can be applied to the fields of rehabilitation, navigation in virtual and physical worlds as well as entertainment.

2. THE OVERALL ARCHITECTURE

Figure 1 shows a schematic representation of the overall architecture developed. The architecture consists of a motion capture system (MoCap)(Optitrack by Naturalpoint),

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

Copyright: ©2011 Stefania Serafin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

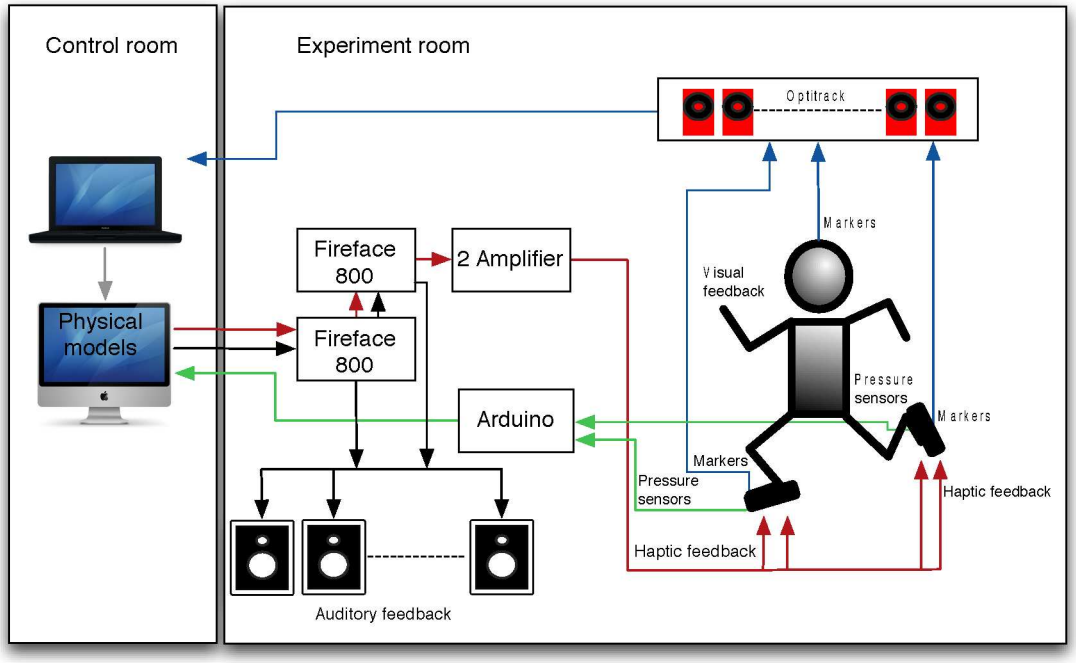


Figure 1. A schematic representation of the multimodal architecture to simulate natural interactive walking.

two soundcards (Fireface 800), twelve loudspeakers (Dynaudio), two amplifiers, two haptic shoes and two computers. Such system is placed in an acoustically isolated laboratory which consists of a control room and a larger interaction room where the setup is installed and where the experiments are performed.

The control room is used by the experimenters providing the stimuli and collecting the results. It hosts two desktop computers. The first computer runs the motion capture software [2], while the second runs the audio-haptic synthesis engine. The two computers are connected through an ethernet cable and communicate via UDP. The coordinates relative to the motion capture system are sent from the first to the second computer which processes them in order to control the sound engine.

A transparent glass divides the two rooms, so it is possible for the experimenters to see the users performing the assigned task.

The two rooms are connected by means of a talkback system. The experiment room is 5.45 m large, 5.55 m long, and 2.85 m high, and the walking area available to the users is about $24m^2$.

3. SIMULATION HARDWARE

3.1 Tracking the user

The user locomotion is tracked by an Optitrack motion capture system [3], composed by 16 infrared cameras [4]. The cameras are placed in a configuration optimized for the tracking of the feet and head position simultaneously. In order to achieve this goal, markers are placed on the top of each shoe worn by the subjects as well as on top of the head.

Users are also tracked by using the pressure sensors embedded in a pair of sandals. Specifically, a pair of lightweight sandals was used (Model Arpenaz-50, Decathlon, Villeneuve d'Ascq, France).

The sole has two FSR pressure sensors [5] whose aim is to detect the pressure force of the feet during the locomotion of a subject wearing the shoes. The two sensors were placed in correspondence to the heel and toe respectively in each shoe. The analogue values of each of these sensors were digitalized by means of an Arduino Diecimila board [6] and were used to drive the audio and haptic synthesis.

3.2 Actuated shoes

In order to provide haptic feedback during the act of walking, a pair of sandals has been recently enhanced with sensors and actuators [9]. The particular model of shoes chosen has light, stiff foam soles that are easy to gouge and fashion. Four cavities were made in the thickness of 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. As indicated in Figure 2, two actuators were placed under the heel of the wearer and the other two under the ball of the foot. These were bonded in place to ensure good transmission of the vibrations inside the soles. When activated, vibrations propagated far in the light, stiff foam. In

[2] Tracking Tools 2.0 by Naturalpoint
[3] http://naturalpoint.com/optitrack/
[4] OptiTrack FLEX:V100R2

[5] I.E.E. SS-U-N-S-00039
[6] http://arduino.cc/

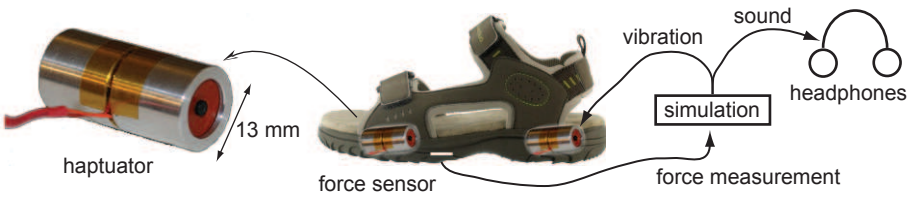


Figure 2. The developed haptic shoes used in this experiment.

the present configuration, the four actuators were driven by the same signal but could be activated separately to emphasize, for instance, the front or back activation, or to realize other effects such as modulating different, back-front signals during heel-toe movements.

A cable exits from each shoe, with the function of transporting the signals of the pressure sensors and for the actuators. Such cables were about 5 meters long, and they were connected [7] to two 4TP (twisted pair) cables: one 4TP cable carries the sensor signals to a breakout board [8], which then interfaces to an Arduino board; the other 4TP cable carries the actuator signals from a pair of Pyle Pro PCA1 [9] mini 2X15 W stereo amplifiers, driven by outputs from a FireFace 800 soundcard. [10]

Each stereo amplifier handles 4 actuators found on a single shoe, each output channel of the amplifier driving two actuators connected in parallel. The PC handles the Arduino through a USB connection, and the FireFace soundcard through a FireWire connection.

In our virtual environment the auditory feedback can be delivered by means of headphones (specifically, Sennheiser HD 650) or a set of 16 channels loudspeakers (Dynaudio BM5A speakers).

4. AUDIO-HAPTIC FEEDBACK

We developed a multimodal synthesis engine able to reproduce auditory and haptic feedback. Auditory feedback is obtained by the combination of a footstep and a soundscape sound synthesis engine. Haptic feedback is provided by means of the haptic shoes previously described. The haptic synthesis is driven by the same engine used for the synthesis of footstep sounds, and is able to simulate the haptic sensation of walking on different surfaces, as illustrated in [9]. The engine for footstep sounds, based on physical models, is able to render the sounds of footsteps both on solid and aggregate surfaces. Several different materials have been simulated, in particular wood, creaking wood, and metal as concerns the solid surfaces, and gravel, snow, sand, dirt, forest underbrush, dry leaves, and high grass as regards the aggregate surfaces. A complete description of such engine in terms of sound design, implementation and control systems is presented in [10]. Using such engine, we implemented a comprehensive collection of footstep sounds. As solid surfaces, we implemented metal, wood, and creaking wood. In these materials, the impact model was used to simulate the act of walking, while the friction model was used to simulate the creaking sounds typical of creaking wood floors. As aggregate surfaces, we implemented gravel, sand, snow, forest underbrush, dry leaves, pebbles and high grass. The simulated metal, wood and creaking wood surfaces were furthermore enhanced by using some reverberation.

To control the audio-haptic footsteps synthesizer in our virtual environment, we use the haptic shoes: the audio-haptic synthesis is driven interactively during the locomotion of the subject wearing the shoes. The description of the control algorithms based on the analysis of the values of the pressure sensors coming from the haptic shoes can be found [9]. Such engine has been extensively tested by means of several audio and audio-haptic experiments and results can be found in [11] [12] [13] [14].

4.1 Movement to sound mapping

Figure 3 shows the dimensions of the path the users were asked to walk on. The mapping between feet movement and delivered auditory feedback was designed as follows:

- zone 1: a narrow band, 15 cm large and 120 cm long, corresponding to the straight direction to be covered. When both the feet stepped inside this zone, a creaking sound corresponding to the ecological sound of stepping on a creaking plank was provided.
- zones 2 and 3: two narrow bands, 10 cm large and 120 cm long, contiguous to zone 1 and placed at its left and right respectively. When one of the feet was stepping in this zone, no feedback was provided to it. However, subject was still able to continue the task by having the foot move to the correct zone 1. This is analogue to the situation of balancing with only one foot.
- zone 4 and 5: the areas contiguous to zone 2 and 3 and placed at its left and right respectively. When both the feet were inside such zone this was considered as failure for the task, and the recording of a long scream of a person falling down was triggered.
- zone 6: the area in front of zones 1, 2 and 3. When one or both the feet were inside this zone this was considered as success for the task, and the recording of a drum roll with an applause was triggered.
- zone 7: the area beyond of zones 1, 2 and 3. When both the feet were inside this zone, no sound was delivered.

[7] through DB9 connectors
[8] containing trimmers, that form voltage dividers with the FSRs
[9] http://www.pyleaudio.com/manuals/PCA1.pdf
[10] http://www.rme-audio.com/english/firewire/ff800.htm

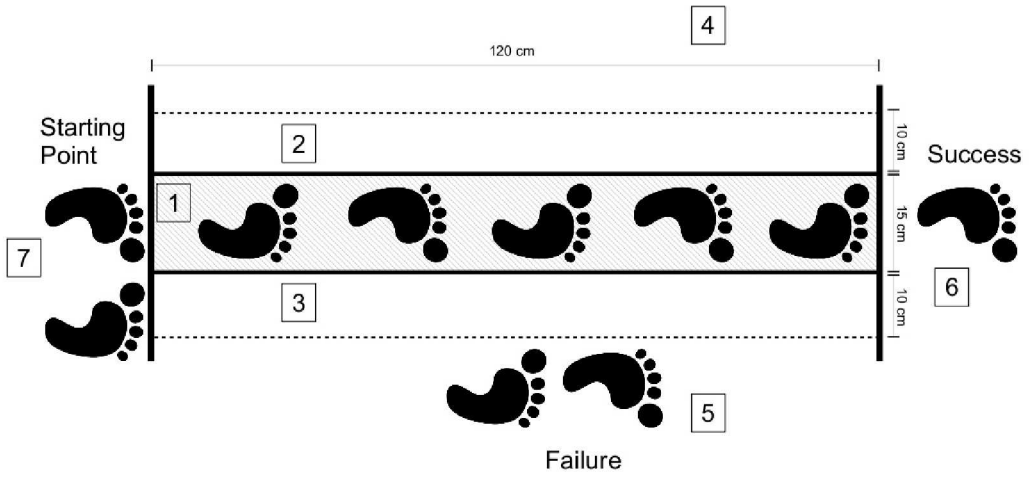


Figure 3. The zones in which the walking space was divided.

5. EXPERIMENT DESIGN

We designed an experiment in order to investigate the role of auditory and haptic feedback in facilitating a balancing task on a virtual plank. In such experiment, we asked subjects to walk straight in order not to virtually fall from the plank. Specifically, subjects were give the following instructions: "Imagine you are walking on a wooden plank. Your task is to walk from one side to the other. Walk slowly and pay attention to the feedback you receive in order to succeed on your task. If your feet are outside of the plank you will fall."

Figure 4 shows a subject performing the experiment. In this particular situation, no visual feedback was provided, and subjects were asked to walk being driven only by auditory and haptic feedback. The same stimuli were provided for the auditory and haptic simulation and designed as follows: when a user is walking on top of the virtual plank, the feet's position is detected by the motion capture system. In this case, the synthesis engine provides as a stimulus the sound and haptic feedback of a creaking wood.

5.1 Participants

The experiment was performed by 15 participants, 14 men and 1 women, aged between 22 and 28 (mean=23.8, standard deviation=1.97). All participants reported normal hearing conditions. The participants took on average 6.8 minutes to complete the experiment.

Subjects were randomly exposed to the four following conditions: auditory feedback, haptic feedback, audio-haptic feedback and no feedback. Each condition was tried twice, given in total eight trials for each subject.

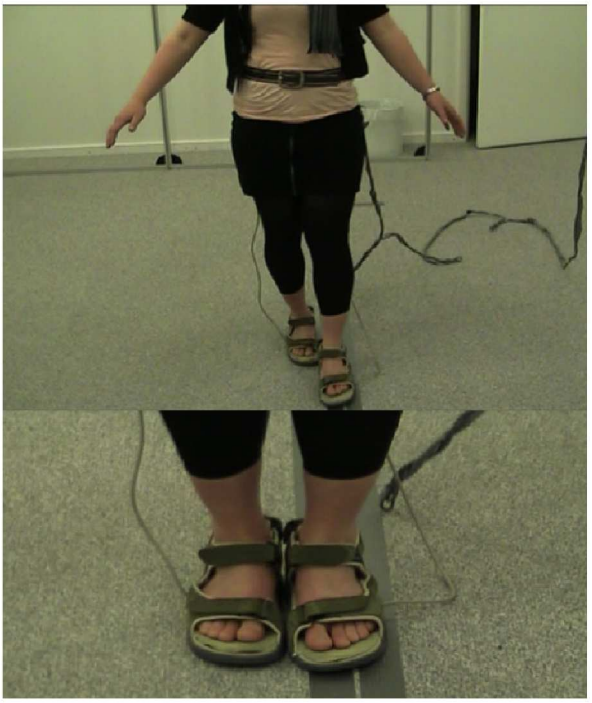


Figure 4. A subject performing the experiment of walking on a virtual plank. The tape on the floor represents the area where the plank is located.

6. RESULTS OF THE EXPERIMENT

Table 1 shows the performance for each subject. The numbers in each row for each conditions indicate whether the subject performed successfully the task ones, twice or never. The results show that feedback helps balance mostly when haptic stimuli are provided. In this case, 46.6% of the tasks

Table 1. Summary of the results of the experiment. The number in each element of the matrix represents the times the task was successful (once, twice or never).

Condition/ Subject number	**Audio (A)**	**Haptic (H)**	**Audio-haptic (AH)**	**No-feedback (N)**
1	2	1	2	1
2	2	1	1	1
3		1		
4	1	1		
5	1		2	
6	1	1	2	1
7		2	1	
8	1			
9	1	1		
10				
11	1	1		
12		2	1	1
13				
14		1	2	1
15	2	2	1	2

were successfully completed. In the case where a combination of auditory and haptic feedback was provided, 43.3% of the tasks were completed. With only auditory feedback, 40% of the tasks were completed, while with no feedback only 26.6%. These results show how feedback slightly helps the balancing task. Haptic feedback performed better than the combination of auditory and haptic. This can be due to the fact that haptic feedback was provided directly to the feet, so the subjects had a closer spatial connection between the path they had to step on and the corresponding feedback.

A post-experimental questionnaire was also performed, where subjects were asked several questions on the ability to freely move in the environment, to adjust to the technology and to which feedback was the most helpful. Indeed, 7 subjects found the haptic feedback to be the most helpful, 6 subjects the auditory feedback and 2 subjects the combination of auditory and haptic feedback. One subject indeed commented that the most useful feedback was when there was background noise (the pink noise used to mask the auditory feedback) and only vibration was provided. All subjects claimed to notice the relationship between actions performed and feedback provided.

The subjects also commented that the feedback did not always match their expectations, since sometimes no feedback was provided. It is hard to assess if this was due to a technical fault of the system (for example, due to faulty tracking from the motion capture system), or to the fact that subjects were experiencing the condition with no feedback. Some subjects understood that the condition without any feedback was done on purpose, others confused it with a bug of the system.

Some subjects also commented on the fact that shoes were not fitting their size. Moreover, some felt disable without the visual feedback. One subject observed that he simply ignored the feedback and walked straight. This is an indication of his unwillingness of suspending his disbelief, and behave in a way similar to how they would behave when walking on a real narrow plank [15].

Overall, observations of most of the subjects showed that they were walking carefully listening and feeling the feedback in order to successfully complete the task. It is hard to assess whether the lack of feedback was the condition subjects were exposed to, the fact that they were outside the plank or a fault of the system. Afterall, previous mentioned research has shown that subjects do not walk straight even when they think they do. Some of the test subjects were noticeably not walking straight, although in the post-experimental questionnaire they commented on a faulty system. Very few understood that the lack of feedback was provided intentionally.

7. CONCLUSION

In this paper, we introduced a multimodal architecture whose goal is to simulate natural interactive walking in virtual environments. We present an experiment which assessed the role of auditory and haptic feedback, together with their combination, in helping subjects to complete the task of walking on a virtual rope.

The experiment provided some indications that haptic feedback at the feet level is more useful than auditory feedback when balancing on a virtual plank. Moreover, most subjects behaved in the virtual world as if they would have behaved in the real world, i.e., by walking slowly and carefully to try not to fall from the plank. More experiments, however, are needed to achieve a better understanding of the role of the different modalities in helping navigation and balance control.

8. ACKNOWLEDGMENT

The research leading to these results has received funding from the European Community's Seventh Framework Program under FET-Open grant agreement 222107 NIW - Natural Interactive Walking. [11] The authors would like to thank Vincent Hayward, Smilen Dimitrov and Amir Berrezag who built the sandals used in this experiment, and Jon Ram Pedersen and Kristina Daniliauskaite who collaborated in preliminary versions of the described experiment.

9. REFERENCES

[1] A. Pelah and J. Koenderink, "Editorial: Walking in real and virtual environments," *ACM Transactions on Applied Perception (TAP)*, vol. 4, no. 1, p. 1, 2007.

[2] J. Paradiso, K. Hsiao, and E. Hu, "Interactive music for instrumented dancing shoes," in *Proc. of the 1999 International Computer Music Conference*, 1999, pp. 453–456.

[3] A. Benbasat, S. Morris, and J. Paradiso, "A wireless modular sensor architecture and its application in on-shoe gait analysis," in *Sensors, 2003. Proceedings of IEEE*, vol. 2, 2003.

[4] M. Frey, "CabBoots: shoes with integrated guidance system," in *Proceedings of the 1st international conference on Tangible and embedded interaction*. ACM, 2007, pp. 245–246.

[5] Y. Takeuchi, "Gilded gait: reshaping the urban experience with augmented footsteps," in *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. ACM, 2010, pp. 185–188.

[6] J. Souman, I. Frissen, M. Sreenivasa, and M. Ernst, "Walking straight into circles," *Current Biology*, vol. 19, no. 18, pp. 1538–1542, 2009.

[7] Y. Visell, F. Fontana, B. Giordano, R. Nordahl, S. Serafin, and R. Bresin, "Sound design and perception in walking interactions," *International journal of human-computer studies*, vol. 67, no. 11, pp. 947–959, 2009.

[8] R. Nordahl, S. Serafin, and L. Turchet, "Sound synthesis and evaluation of interactive footsteps for virtual reality applications," in *Virtual Reality Conference (VR), 2010 IEEE*. IEEE, 2010, pp. 147–153.

[9] L. Turchet, R. Nordahl, A. Berrezag, S. Dimitrov, V. Hayward, and S. Serafin, "Audio-haptic physically based simulation of walking sounds," in *Proc. of IEEE International Workshop on Multimedia Signal Processing*, 2010.

[10] L. Turchet, S. Serafin, S. Dimitrov, and R. Nordahl, "Physically based sound synthesis and control of footsteps sounds," in *Proceedings of Digital Audio Effects Conference*, 2010.

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

[11] R. Nordahl, S. Serafin, and L. Turchet, "Sound synthesis and evaluation of interactive footsteps for virtual reality applications," in *Proc. IEEE VR 2010*, 2010.

[12] R. Nordahl, A. Berrezag, S. Dimitrov, L. Turchet, V. Hayward, and S. Serafin, "Preliminary experiment combining virtual reality haptic shoes and audio synthesis," in *Proc. Eurohaptics*, 2010.

[13] 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," in *Proc. Eurohaptics symposium on Haptics and Audio-visual environments*, 2010.

[14] L. Turchet, R. Nordahl, and S. Serafin, "Examining the role of context in the recognition of walking sound," in *Proc. of Sound and Music Computing Conference*, 2010.

[15] M. Slater, "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, p. 3549, 2009.

An Investigation on Temporal Aspects in the Audio-Haptic Simulation of Footsteps

Luca Turchet and Stefania Serafin

Department of Architecture, Design and Media Technology,
Aalborg University Copenhagen
{tur,sts}@create.aau.dk

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

Keywords: Footstep sounds, physical models, auditory feedback, haptic feedback.

1 Introduction

Previous research on simulating walking sounds using physics based engines has focused on the act of walking on flat surfaces [2,4,3,14,11]. In the virtual reality community, few locomotion interfaces are able to render uneven grounds, and they have the disadvantage of being costly and cumbersome [5,6,8]. Recently, research has shown that it is possible to simulate the act of walking on unflat surfaces by only using visual cues [10]. Three parameters of camera motion 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 look at a video of the simulations. Results show that such visualization techniques successfully simulate bumps and holes located in the ground. These results are a development of previous research on pseudo-haptic simulation [9]. This research was extended by implementing a multimodal (audio-visual) simulation of walking on a bump or a hole [16]. Results in this case showed that the auditory cues reinforce the visual cues when coherent cues are provided in both modalities. When subjects were exposed to conflicting cues, for example by simulating visually the act of walking on a bump and auditorily 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 [21].

In this paper, we are interested in exploring the possibility of implementing such pseudo-haptic feedback from the sonic and haptic point of view. Recently,

A. Vatakis et al. (Eds.): Time and Time Perception 2010, LNAI 6789, pp. 101–115, 2011.
© Springer-Verlag Berlin Heidelberg 2011

we developed a system which can provide combined auditory and haptic sensations that arise while walking on aggregate and solid surfaces. The system is composed of an audio-haptic synthesis engine, and a pair of shoes enhanced with sensors and actuators. Such engine is based on physical models, that drive both the haptic and audio synthesis. A complete description of such system and of all its components is given elsewhere in detail [12], [17].

In a previous study [15], we used the synthesis engine in order to run an experiment whose goal was to assess the role of temporal aspects in sonically simulating the act of walking on a bump or a hole. In particular, we investigated whether the timing between heel and toe and the timing between footsteps affected perception of walking on unflat surfaces. In the experiment the footstep sounds where prerecorded. Results showed that it is possible to simulate a bump or a hole only by using temporal information [15].

Starting from those results, in this paper we are interested in understanding whether at haptic level it is possible to simulate the act of walking on a bump or a hole. Such haptic information is generated by means of the same techniques applied in [15] for the auditory simulation of bumps and holes, i.e., varying the temporal distances between the vibrations corresponding to the heel and to the toe as well as those between footsteps.

2 Simulation Hardware and Software

In this section we briefly describe the system used in the experiments presented in this paper. As mentioned in section 1, the complete description of such system can be found in our previous research [20], [17].

We developed a system which simulates both offline and in real-time the auditory and haptic sensation of walking on different surfaces. Specifically, the sensation of walking on solid surfaces is simulated by using and impact model [7], while to simulate walking on aggregate grounds, we used a physically informed sonic models (PhiSM) algorithm [1].

In order to provide both audio and haptic feedback, haptic shoes enhanced with pressure sensors have been developed. In detail, such shoes are pair of lightweight sandals (Model Arpenaz-50, Decathlon, Villeneuve d'Ascq, France). This particular model has light, stiff foam soles that are easy to gouge and fashion. Four cavities were made in the tickness of 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. As indicated in Fig. 1, two actuators were placed under the heel of the wearer and the other two under the ball of the foot. They were bonded in place to ensure good transmission of the vibrations inside the soles. When activated, vibrations propagated far in the light, stiff foam. The sole has force sensors intended to pick the foot-floor interaction force in order the drive the audio and haptic synthesis. They were not used in the present study.

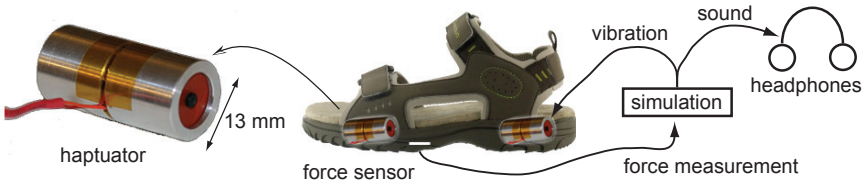


Fig. 1. System (one shoe shown). Left: recoil-type actuation from Tactile Labs Inc. The moving parts are protected by an aluminum enclosure able to bear the weight of a person. Middle: approximate location of the actuators in the sandal. Right: system diagram showing the interconnections. Here the force signal was not used.

The involved hardware provides a control in real-time of an audio-haptic synthesis engine. 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.

As previously mentioned, the engine can work both offline and in realtime. The two approaches differ for the way the input GRF is generated. Concerning the realtime implementation, various systems for the generation of such input have been developed and tested [20], [17], [12], [13], [18], [19]. In the offline implementation, the input signal 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.

In the experiment presented in this paper, 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.[1] Three types of surface profiles have been created starting from the footstep sound generator: bumps, holes, and flat surfaces (see Figure 2). The details about the simulation of such surface profiles are shown in section 3.4. For the purpose of the experiment, 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 one experiment whose goal was to ascertain the participants' ability to recognize if the sounds and the haptic sensations they were exposed to

[1] Hollywood Edge sound effects library: `www.hollywoodedge.com`

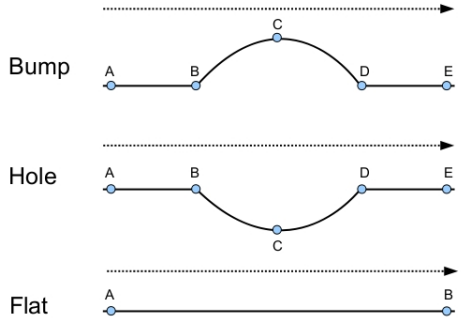


Fig. 2. The three types of surfaces modeled

corresponded to walking on a bump, a hole or a flat surface. Specifically, subjects were exposed to one of the three following conditions:

1. Condition 1: auditory simulation of bumps and holes
2. Condition 2: haptic simulation of bumps and holes
3. Condition 3: audio-haptic simulation of bumps and holes

The results of the experiment related to the first condition have been already published in our previous research [15] and they are illustrated in section 4.1. In this paper we are interested in extending to the haptic level those results achieved with the auditory cues. Indeed the technique used for the simulation of the three types of surface profiles at haptic level is exactly the same proposed for the auditory level. In detail, the audio files used as input to the audio-haptic synthesis engine were created placing at different temporal intervals the footstep sound generator (see section 3.4). The basic 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.

The goal of condition 2 was to assess whether at haptic level it is possible to simulate the act of walking on a bump or a hole by means of this technique, when subjects are not walking but are sitting on a chair and passively receive the haptic stimuli. In this haptic condition, a noise signal masked the auditory input generated by the haptic shoes as result of the activation of the actuators. In particular, participants were presented with a continuous 70 dB spl pink noise over the headphones described in section 3.2. In condition 3, we were interested in understanding whether the addition of haptic feedback to the auditory one enhanced the recognition of the proposed surface profiles as well as the realism of the simulation.

3.1 Participants

Forty-five participants were divided in three groups (n=15) to perform the experiment. The three groups were composed respectively of 11 men and 4 women,

aged between 20 and 29 (mean=23.6, standard deviation=2.84), 11 men and 4 women, aged between 21 and 32 (mean=24.86, standard deviation=3.48) and 11 men and 4 women, aged between 20 and 28 (mean=23.06, standard deviation=2.40). 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, 17 and 21 minutes for experiments 1, 2 and 3 respectively.

3.2 Setup

The experiment was 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 participants were asked to interact, and a spreadsheet to collect their answers. The interface was created using the Max/MSP program[2] and was composed only by buttons to be pressed. Each button was numbered, and by pressing it an audio or audio-haptic stimulus was triggered and conveyed to the user by means of headphones[3] and the haptic shoes respectively. The choice of delivering auditory feedback using headphones was motivated by the fact that we wanted the subjects to be isolated from external noise. 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 the experiment subjects were sitting on a chair, listening to the sounds through headphones and feeling the haptic vibrations through the haptic shoes, 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 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). Furthermore, in conditions 2 and 3 they were asked to evaluate, again on a seven point Likert scale, the degree of realism and the degree of quality of the perceived stimulus. 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.

3.4 Haptic and Auditory Simulation

As described in section 3, the technique used for the simulation of bumps, holes and flat surface consisted of temporal intervals variations. The temporal patterns used were designed to simulate 14 different surface profiles. Specifically 2 flat, 6

[2] Max/MSP: `www.cycling74.com`
[3] Sennheiser HD 600, `http://www.sennheiser.com`

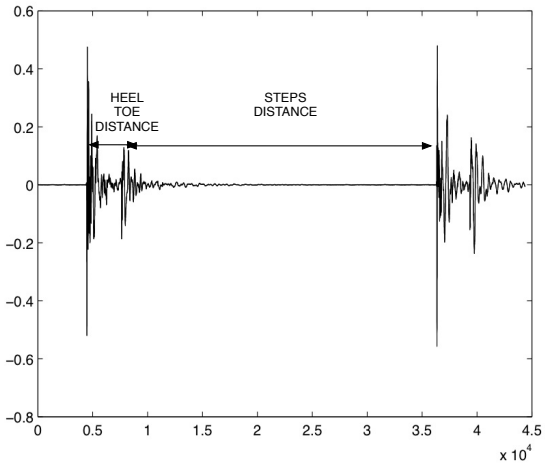


Fig. 3. Temporal distances between (named "steps distance" in the Figure), and within (named "heel-toe distance" in the Figure) footsteps. x-axis: Time (s). y-axis: Amplitude.

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).

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 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. In all the three conditions participants were exposed to 28 trials, where 14 surface profiles were presented twice in randomized order. In condition 1 and 2 (see sections 4.1 and 4.2) the simulated surface profiles are all those presented in table 1, while in condition 3 only 7 of them was used and presented both in audio and audio-haptic conditions (see section 4.3).

The audio-haptic synthesis engine was set in order to synthesize footstep sounds and vibrations on two different kinds of materials: wood and gravel. Each surface profile was presented with both wood and gravel.

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

	Duration (sec.)	Number of steps	Footsteps temporal distance increment (ms.)	Footsteps temporal distance range (ms.)	Heel-toe increment (ms.)	Heel-toe range (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)

The haptic signal was separated in order to activate coherently the actuators placed under the heel and the toe. In our simulation the actuators placed under the heel of the right foot were activated simultaneously with those of the left foot, using the same signal. Analogously, the haptic signal for the toes was conveyed simultaneously to the actuators placed under the toes of both the feet.

4 Results

In this section we show the results for condition 1 (audio condition), condition 2 (haptic condition) and condition 3 (comparison between audio and audio-haptic conditions). In section 4.1 we report the results found in a previous work [15], where the 14 surface profiles described in section 3.4 were used to simulate a bump or a hole only by using auditory cues.

4.1 Results of Condition 1 (Audio Condition)

The results of the condition 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 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

Table 2. Results of condition 1 (audio condition) for the wood surface

	Bump	Hole	Flat	Mean certainty Total	Mean certainty Correct answers	Mean certainty 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 3. Results of condition 1 (audio condition) for the gravel 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

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.

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

4.2 Results of Condition 2 (Haptic Condition)

Tables 4 and 5 illustrate the results of condition 2 for wood and gravel respectively. The structure of these tables is the same of tables 2 and 3, with the addition of the two columns expressing the average scores of the degree of realism and of the degree of quality.

The first noticeable element emerging from both tables is that it is possible to simulate bumps, holes and flat surfaces by means of haptic stimuli using only the variation of the temporal information. Indeed, on average, subjects could successfully recognize the simulated surfaces using only the haptic cues described in section 3.4, both with wood and gravel. As a matter of fact, the percentage of correct answers is quite high in most of the conditions, reaching also 100 % of correct answers in two conditions, while only in three conditions the score is under 50 %.

From a comparison between all the conditions it is possible to notice that on average the stimuli involving only the variations in the heel-toe distance (i.e., those with suffix _h_t) are recognized with the lowest scores in comparison with the other conditions. Such tendency is also confirmed by values in column 5 which expresses the total degree of certainty for each stimulus (both of correct and wrong answers) and therefore it is an indication of the difficulty of the choice. Indeed in both the tables such values are on average lower for the stimuli with suffix _h_t than all the other conditions.

A chi-square test and a t-test were 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, nor in the recognition rate for the different conditions in the same material. Moreover no significant difference was measured between the recognition rate obtained when changing the temporal information between footsteps versus the one obtained when changing the temporal information within footsteps. In addition, the combination of the two temporal information did not significantly enhance the recognition of a bump or a hole. Finally, a t-test performed on the realism and on the quality evaluations, did not reveal any significant difference between the two materials nor among the conditions.

Table 4. Results of condition 2 (haptic condition) for the wood surface

	Bump	Hole	Flat	Mean certainty Total	Mean certainty Correct answers	Mean certainty Wrong answers	% Correct answers	Mean Realism	Mean Quality
flat_1	2		13	5.2	5.3077	4.5	86.66	3.9231	4
flat_2	2		13	5.1333	5.3077	4	86.66	3.6154	4.1333
bump_1_step	13		2	4.8	4.7692	5	86.66	4.0769	4.0667
bump_2_step	15			5.2	5.2		100	3.9333	3.8
hole_1_step	2	11	2	4.6667	4.4	5.2	73.33	3.0909	3.6
hole_2_step	2	13		4.6667	4.6923	4.5	86.66	3.7692	3.9333
bump_1_h_t	7	5	3	3.6667	4.1429	3.25	46.66	3	3.4
bump_2_h_t	9	6		4.4667	4.7778	4	60	3.4444	3.4667
hole_1_h_t	1	8	6	3.8	4.375	3.1429	53.33	3.25	3.2
hole_2_h_t		12	3	3.8667	4.25	2.3333	80	3.75	3.2
bump_1_comb	11	4		4.2667	4.2727	4.25	73.33	3.1818	3.2
bump_2_comb	11	3	1	4.7333	5.1818	3.5	73.33	4.0909	3.7333
hole_1_comb	2	13		4.4	4.3846	4.5	86.66	3.5385	3.8
hole_2_comb	1	12	2	4.6667	5.3333	2	80	3.25	3.2

Table 5. Results of condition 2 (haptic condition) for the gravel surface

	Bump	Hole	Flat	Mean certainty Total	Mean certainty Correct answers	Mean certainty Wrong answers	% Correct answers	Mean Realism	Mean Quality
flat_1			15	4.4667	4.4667		100	3.6	3.8667
flat_2	1	1	13	5.2667	5.6923	2.5	86.66	4.7692	4.3333
bump_1_step	12		3	4.9333	5	4.6667	80	3.9231	3.7333
bump_2_step	11	2	2	4.4667	4.7273	3.75	73.33	3.3636	3.33
hole_1_step		12	3	4.8	4.9167	4.3333	80	3.6667	3.7333
hole_2_step		12	3	5.5333	5.75	4.6667	80	4.25	4.2
bump_1_h_t	8	6	1	4.4	5	3.7143	53.33	2.875	3.5333
bump_2_h_t	11	1	3	3.8667	4.1818	3	73.33	3.7273	3.8667
hole_1_h_t	1	7	7	4.3333	4.2857	4.375	46.66	3.8571	4
hole_2_h_t	3	6	6	3.7333	3.6667	3.7778	40	3.5	3.4667
bump_1_comb	13	1	1	5.2	5.2308	5	86.66	3.7692	4
bump_2_comb	10	4	1	4.4	5.1	3	66.66	3.5	3.5333
hole_1_comb	1	13	1	4.8	4.8462	4.5	86.66	3.3077	3.4667
hole_2_comb		14	1	4.8	4.9286	3	93.33	4.2143	4

4.3 Results of Condition 3 (Audio and Audio-Haptic Condition)

In this experiment we used only a subset of the 14 files used in the previous conditions. This was due in order to avoid that the experiment became too long for the participants, and therefore started answering randomly. In particular, we selected 7 stimuli to be presented both with the modality audio and audio-haptic (suffixes _A and _AH respectively in tables 6 and 7), and both with wood and gravel. Precisely we used those with suffix _2 in table 1, with the exception of the flat stimulus for which we used the one with suffix _1.

The first noticeable element emerging from both tables is that subjects could successfully recognize the simulated surfaces with high precision both at audio and audio-haptic level. An exception is the stimulus hole_h_t_A with wood material for which the score is 40 %, differing from the percentage obtained by the same stimulus in condition 1. Anyways a chi-square test revealed that such difference is not statistically significant so we can state that all the results concerning the audio modality are consistent with those found in condition 1.

Table 6. Results of condition 3 (audio and audio-haptic condition) for the wood surface

	Bump	Hole	Flat	Mean certainty Total	Mean certainty Correct answers	Mean certainty Wrong answers	% Correct answers	Mean Realism	Mean Quality
flat_A			15	6	6	-	100	3.7333	3.6
flat_AH		1	14	6.0667	6.4286	1	93.33	4.9286	4.4667
bump_step_A	11	2	2	4.7333	5.1818	3.5	73.33	3.2727	3.6667
bump_step_AH	12	2	1	5.2667	5.5833	4	80	3.9167	4.6
hole_step_A	1	14		5.6667	5.7143	5	93.33	3.6429	4
hole_step_AH	2	12	1	5.1333	5.6667	3	80	4.75	4.4667
bump_h_t_A	13	2		4.9333	5.3846	2	86.66	3.6154	3.33
bump_h_t_AH	13	2		4.6667	5.2308	1	86.66	4.3846	4.0667
hole_h_t_A	2	6	7	3.7333	4.5	3.2222	40	2.8571	3
hole_h_t_AH	2	11	2	4.0667	4.1818	3.75	73.33	3.6364	4.4
bump_comb_A	12	3		4.2	4.8333	1.6667	80	3.25	3.6667
bump_comb_AH	14	1		5.0667	5.2857	2	93.33	4.2143	4.3333
hole_comb_A	1	12	2	4.9333	5.25	3.6667	80	3.5	3.8
hole_comb_AH	2	13		5.3333	5.5385	4	86.66	4.3846	4.6

Table 7. Results of condition 3 (audio and audio-haptic condition) for the gravel surface

	Bump	Hole	Flat	Mean certainty Total	Mean certainty Correct answers	Mean certainty Wrong answers	% Correct answers	Mean Realism	Mean Quality
flat_A			15	5.3333	5.3333	-	100	3.8	4.2
flat_AH	1	1	13	5.6667	5.8462	4.5	86.66	5.1538	5.2667
bump_step_A	12	2	1	4.5333	4.6667	4	80	3.75	4
bump_step_AH	12	2	1	5.3333	5.5833	4.3333	80	4.3333	4.8667
hole_step_A	1	14		5.1333	5.2857	3	93.33	4.2857	4.7333
hole_step_AH	1	13	1	5.0667	5	5.5	86.66	4.4615	4.9333
bump_h_t_A	12	1	2	4.1333	4.25	3.6667	80	2.9167	3.6667
bump_h_t_AH	12	1	2	5.4	5.25	6	80	4.5833	4.8667
hole_h_t_A		10	5	4.2667	4.4	4	66.66	2.8	3.7333
hole_h_t_AH	3	12		4.6667	4.75	4.3333	80	4.3333	5.3333
bump_comb_A	14	1		5.1333	5.2143	4	93.33	3.9286	4
bump_comb_AH	12	3		5.1333	5.3333	4.3333	80	4.0833	4.6
hole_comb_A	1	13	1	4.9333	5.1538	3.5	86.66	4	4.5333
hole_comb_AH	2	13		5.7333	6.2308	2.5	86.66	4.7692	5.2667

As regards the degree of certainty of the correct answers, most of the times conditions with audio-haptic modality present higher values than the audio modality, Despite an in-depth analysis with t-test did not show any statistically significant difference between the corresponding pairs, the ANOVA analysis revealed that at global level the AH condition gives rise to higher evaluation in the case of gravel material (p-value = 0.0449).

As concerns the degree of realism and the degree of quality, results are always higher for the bimodal stimuli rather than the unimodal ones (see figure 4). The ANOVA analysis shows that such differences are statistically significative, (p-value < 0.0001 and p-value < 0.0001 for realism in the conditions wood and gravel respectively, and p-value = 0.002862 and p-value = 0.003386 for quality in the conditions wood and gravel respectively). In addition an in depth analysis performed with t-test, revealed also significant differences for the pairs of stimuli (audio, audio-haptic) in table 8.

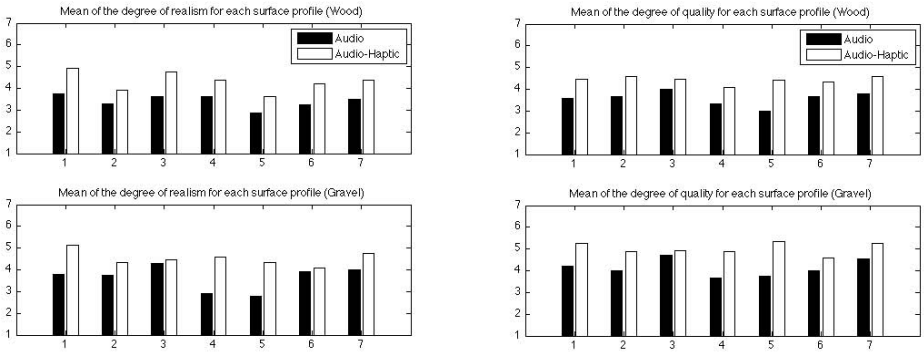


Fig. 4. Average scores of the degree of realism (left) and quality (right), for audio and audio-haptic conditions in condition 3. Surface profile from left to right: 1-flat, 2-bump_step, 3-hole_step, 4-bump_h_t, 5-hole_h_t, 6-bump_comb, 7-hole_comb.

Table 8. Statistically significative differences for average scores of the degree of realism and quality, between audio and audio-haptic conditions in condition 3

Stimulus	**Material**	**P-value realism**	**P-value quality**
flat	wood	0.03381	-
flat	gravel	0.01425	-
hole_h_t	wood	-	0.01834
bump_h_t	gravel	0.001482	0.04558
hole_h_t	gravel	0.02296	0.00732

Considering the total degree of certainty (column 5), it is possible to notice that on average the stimuli involving only the variations in the heel-toe distance present lower values in comparison with the other conditions. This is an indication of a greater difficulty in the participants' choices for these stimuli rather than the others.

Finally, an ANOVA analysis performed on the evaluations of the degree of certainty of correct answers, of realism and of quality, did not reveal any significant difference between the two materials.

5 General Discussion

From a comparison between results of the three condtions it is possible to notice that the use of the proposed techniques allows to simulate the act of walking on a bump, a hole, or a flat surface by passively presenting audio, haptic and audio-haptic stimuli. The most important result is that the haptic modality plays an important role in this kind of tasks. In particular, results of condition 3 show that haptic cues significantly reinforces the auditory cues. Indeed, participants gave rise to higher evaluations in presence of audio-haptic stimuli rather than

the audio-stimuli, for what concerns the realism and the quality of the presented stimuli, as well as they were more certain in giving their correct answers. When the haptic is presented alone it is still possible to simulate quite well the act of walking over a bump, a hole or a flat surface, as results of condition 2 show. From a comparison with the results of condition 1 what emerges is that the auditory modality seems to be dominant on the haptic one in this kind of tasks. A common trend noticed in all the conditions is that the material does not have any particular influence on the participants evaluations, i.e., there is no statistically significant difference in the responses between the presented solid and aggregate surfaces. Another trend more or less common to the three conditions is that on average the stimuli involving only the variations in the heel-toe distance present lower values in comparison with the other conditions for what concerns the total degree of certainty, and this is an indication of a greater difficulty in the participants' choices for these stimuli rather than the others. This also means that participants found more natural the variations in the temporal distance between steps rather than the ones in the temporal distances between heel and toe. In addition, in all the conditions the combination of the two temporal information did not significantly enhance the recognition of a bump or a hole. Subjects were allowed to listen to the simulations as many times as they wished, but in general they listened only once, and this was enough to provide an answer. At the end of the test some informal interviews were performed, where subjects declared that the stimuli were clear. Subjects were also surprised by the ability of the feet to distinguish the simulations.

6 Conclusion and Future Work

In this paper, we described a between subject audio-haptic experiment whose goal was to assess the role of temporal aspects in recognizing whether a person is walking on a flat surface, a bump or a hole.

Results show that varying temporal aspects between footsteps allow to successfully simulating the act of walking on a bump, hole, or flat surface, especially in the auditory modality. Moreover, in the recognition task, haptic cues show to significantly reinforce auditory cues, as the results of condition 3 show. All three conditions were run by using the synthesis engine offline. The reason why we asked subjects to seat was to create an environment similar to a home situation, e.g., when subjects are watching a movie. Indeed, we plan to use the same engine with visual feedback, in order to assess the role of haptic feedback in enhancing the experience of watching an audio-visual scene.

In the future, we are interested in running the same experiments interactively, to understand whether our technique can be used to simulate bumps and holes in a virtual reality setup.

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.[4]. The authors

[4] Natural Interactive Walking Project: `www.niwproject.eu`

would like to thank Professor Vincent Hayward and Amir Berrezag for providing the haptic shoes used in this experiment.

References

1. Cook, P.: Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. Computer Music Journal 21(3), 38–49 (1997)
2. Cook, P.: Modeling Bill's Gait: Analysis and Parametric Synthesis of Walking Sounds. In: Proceedings of the AES 22nd International Conference on Virtual, Synthetic, and Entertainment Audio, pp. 73–78 (2002)
3. Farnell, A.: Marching onwards: procedural synthetic footsteps for video games and animation. In: Proceedings of the Pure Data Convention (2007)
4. Fontana, F., Bresin, R.: Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. In: Proc. Colloquium on Musical Informatics, pp. 109–114 (2003)
5. Hollerbach, J., Checcacci, D., Noma, H., Yanagida, Y., Tetsutani, N.: Simulating side slopes on locomotion interfaces using torso forces. In: Haptic Symposium, Citeseer, pp. 91–98 (2003)
6. Hollerbach, J., Mills, R., Tristano, D., Christensen, R., Thompson, W., Xu, Y.: Torso force feedback realistically simulates slope on treadmill-style locomotion interfaces. The International Journal of Robotics Research 20(12), 939 (2001)
7. Hunt, K.H., Crossley, F.R.E.: Coefficient of restitution interpreted as damping in vibroimpact. ASME Journal of Applied Mechanics 42(2), 440–445 (1975)
8. Iwata, H., Yano, H., Nakaizumi, F.: Gait master: A versatile locomotion interface for uneven virtual terrain. vr, page 131(2001)
9. Lécuyer, A., Burkhardt, J., Etienne, L.: 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, New York (2004)
10. Marchal, M., Lecuyer, A., Cirio, G., Bonnet, L., 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)
11. Miner, N., Caudell, T.: Using wavelets to synthesize stochastic-based sounds for immersive virtual environments. ACM Transactions on Applied Perception 2(4), 521–528 (2005)
12. Nordahl, R., Serafin, S., Turchet, L.: Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In: Proc. IEEE VR (2010)
13. 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: Proc. Eurohaptics Symposium on Haptics and Audio-visual Environments (2010)
14. Serafin, S., Turchet, L., Nordahl, R.: Extraction of ground reaction forces for real-time synthesis of walking sounds. In: Proc. Audiomostly (2009)
15. Serafin, S., Turchet, L., Nordahl, R.: Do You Hear A Bump Or A Hole? An Experiment on Temporal Aspects in the Recognition of Footsteps Sounds (2010)
16. Turchet, L., Marchal, M., Lécuyer, A., Nordahl, R., Serafin, S.: Influence of auditory and visual feedback for perceiving walking over bumps and holes in desktop VR. In: Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology, pp. 139–142. ACM, New York (2010)

17. Turchet, L., Nordahl, R., Berrezag, A., Dimitrov, S., Hayward, V., Serafin, S.: Audio-haptic physically based simulation of walking sounds. In: Proc. of IEEE International Workshop on Multimedia Signal Processing (2010)
18. Turchet, L., Nordahl, R., Serafin, S.: Examining the role of context in the recognition of walking sound. In: Proc. of Sound and Music Computing Conference (2010)
19. Turchet, L., Serafin, S., Dimitrov, S., Nordahl, R.: Conflictual audio-haptic feedback in physically based simulation of walking sounds. In: Proc. of Haptic Audio Interaction Design Conference (2010)
20. 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)
21. Welch, R., Warren, D.: Immediate perceptual response to intersensory discrepancy. Psychological Bulletin 88(3), 638 (1980)

A PRELIMINARY STUDY ON SOUND DELIVERY METHODS FOR FOOTSTEP SOUNDS

Luca Turchet, *

Medialogy Section
Department of Architecture, Design and Media Technology
Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, DK
tur@create.aau.dk

Stefania Serafin,

Medialogy Section
Department of Architecture, Design and Media Technology
Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, DK
sts@create.aau.dk

ABSTRACT

In this paper, we describe a sound delivery method for footstep sounds, investigating whether subjects prefer static rendering versus dynamic. In this case, dynamic means that the sound delivery method simulates footsteps following the subject. An experiment was run in order to assess subjects' preferences regarding the sound delivery methods. Results show that static rendering is not significantly preferred to dynamic rendering, but subjects disliked rendering where footstep sounds followed a trajectory different from the one they were walking along.

1. INTRODUCTION

Procedural sound synthesis is becoming a successful approach to simulate interactive sounds in virtual environments and computer games [1, 2]. One important category of sounds produced by action of subjects navigating in an environment is the sound of footsteps.

Recently, several algorithms have been proposed to simulate walking sounds. One of the pioneers in this field is Perry Cook, who proposed a collection of physically informed stochastic models (PhiSM) simulating several everyday sonic events [3]. Among such algorithms the sounds of people walking on different surfaces were simulated [4]. A similar algorithm was also proposed in [5], where physically informed models simulate several stochastic surfaces.

Recently, in [6] a solution based on granular synthesis was proposed. The characteristic events of footstep sounds were reproduced by simulating the so-called ground reaction force, i.e., the reaction force supplied by the ground at every step.

The research just described does not take into consideration the ability of footstep sounds to be rendered in a 3D space. Sound rendering for virtual environments has reached a level of sophistication that it is possible to render in realtime most of the phenomena which appear in the real world [7].

In this study, we are interested in investigating how subjects react to different kinds of sound rendering algorithms, which follow the user or propose different confusing trajectories.

The results presented in this paper are 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.

2. SYSTEM ARCHITECTURE

The system we adopted for the experiments consists of a a motion capture system (MoCap)[2], a soundcard[3], eight loudspeakers, two sandals with pressure sensors embedded in, and two computers. Figure 1 shows a schematic representation of the overall architecture developed.

Such system was placed in an acoustically isolated laboratory which consisted of a control room and a bigger room where the setup was installed and where the experiments were performed. The control room was 5.45 m large, 2 m long, and 2.85 m high, and it was used by the experimenters providing the stimuli and collecting the experimental results. It hosted two desktop computers.

The first computer run the motion capture software, while the second run the footstep sounds synthesis engine (see section 2.1). The two computers were connected through an ethernet cable and communicate by means of the UDP protocol. The data relative to the motion capture system were sent from the first to the second computer which processed them in order to control the sound engine.

The experiment room was 5.45 m large, 5.55 m long, and 2.85 m high. A transparent glass divided the two rooms, so it was possible for the experimenters to see the users performing the assigned task. The two rooms were connected by means of a talkback system.

The user locomotion was tracked by an Optitrack motion capture system[4], composed by 16 infrared cameras[5]. The cameras were placed in a configuration optimized for the tracking of the head position. In order to achieve this goal, markers were placed on the top of the head using a bicycle helmet. The walking area available to the users for the purposes of the experiments consisted of a rectangle 2.5 x 2.6 m which corresponded to the area fully seen by the infrared cameras. The perimeter of such rectangle was indicated on the floor by means of scotch tape strips (see figure 2)

Users were also tracked by using the pressure sensors embedded in a pair of shoes. Specifically, a pair of light-weight san-

* This work was supported by the NIW project

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

[2] from Naturalpoint with software Tracking Tools 2.0

[3] FireFace 800 soundcard: http://www.rme-audio.de/en_products_fireface_800.php

[4] http://naturalpoint.com/optitrack/

[5] OptiTrack FLEX:V100R2

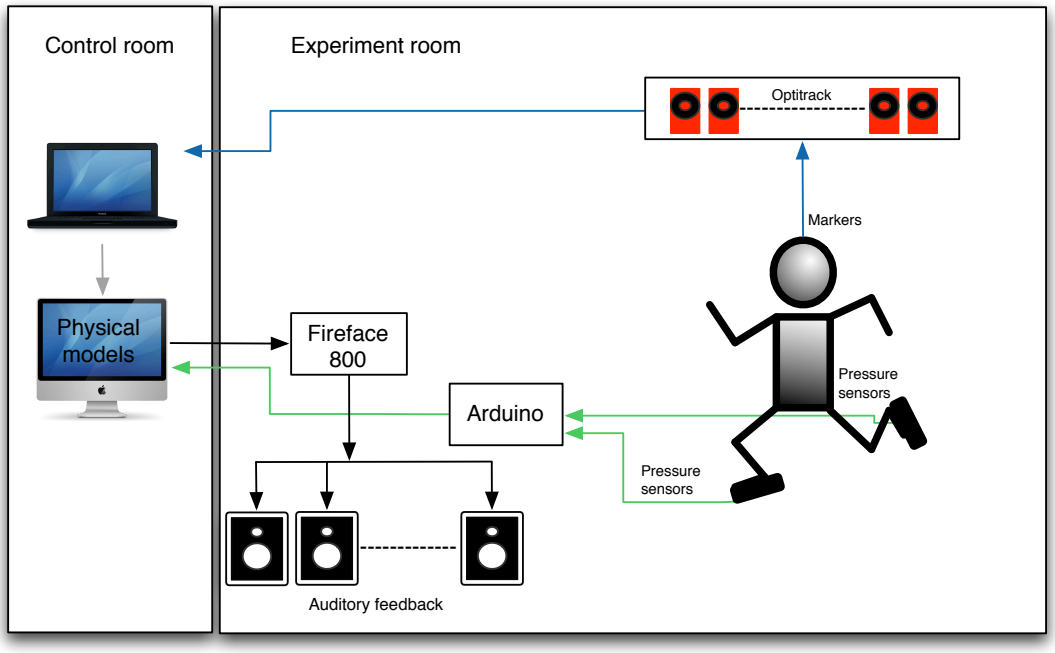


Figure 1: *A block diagram of the architecture.*

dals was used (Model Arpenaz-50, Decathlon, Villeneuve d'Ascq, France). The sole had two FSR pressure sensors[6] whose aim was to detect the pressure force of the feet during the locomotion of a subject wearing the shoes. The two sensors were placed in correspondence to the heel and toe respectively in each shoe. The analogue values of each of these sensors were digitalized by means of an Arduino Diecimila board[7] and were used to drive the audio synthesis.

The configuration of the eight loudspeakers is illustrated in Figure 2. In detail, the loudspeakers[8] were placed on the ground at the vertices and at the middle point of the sides of the rectangular floor. During the experiments the loudspeakers were hidden from view using acoustically transparent curtains.

2.1. Footstep sounds synthesis engine

In previous research, we proposed a sound synthesis engine able to simulate footstep sounds on aggregate and solid surfaces [8]. 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 operate both offline and in real-time. The two approaches differ in the way the input GRF is generated. Concerning the realtime work, various systems for the generation of such input have been developed and tested [8, 9, 10]. In the proposed experiments, the footstep sounds synthesis is driven interactively during the locomotion of the subject wearing the shoes. The description of the control algorithms based on the analysis of the values of the pressure sensors coming from the shoes can be found in [11]).

The sound synthesis algorithms were implemented in C++ as external libraries for the Max/MSP[9] sound synthesis and multimedia real-time platform.

2.2. Sound delivery methods

We implemented and tested two different types of approaches for the delivery of the footstep sounds through the loudspeakers: static and dynamic diffusion.

For static diffusion we intend that the footstep sound, generated interactively during the locomotion of the user wearing the shoes, is diffused simultaneously to the eight loudspeakers, and with the same amplitude in each loudspeaker.

Conversely, during the dynamic diffusion the user position was tracked by the MoCap and it was used to diffuse the footsteps sound according to a sound diffusion algorithm based on ambisonics. Specifically, to achieve the dynamism we used the ambisonic tools for Max/MSP[10] which allow to move virtual sound sources along trajectories defined on a tridimensional space [12]. Such algorithm was set in order to place under the user feet the virtual sound source containing the footstep sounds. In this way the sound followed the user trajectories during his/her locomotion, and therefore the eight loudspeakers delivered the footstep sounds with different amplitudes. As an example, in reference to figure 2, when the user position was near the loudspeakers 1 and 2, the effect resulting from the dynamic diffusion was that the sound was mostly delivered through these two loudspeakers while the loudspeakers 5 and 6, placed on the opposite sides, did not deliver any sound.

[6] I.E.E. SS-U-N-S-00039
[7] http://arduino.cc/
[8] Dynaudio BM5A speakers: http://www.dynaudioacoustics.com/
[9] http://cycling74.com/
[10] Available at http://www.icst.net/research/projects/ambisonics-tools/

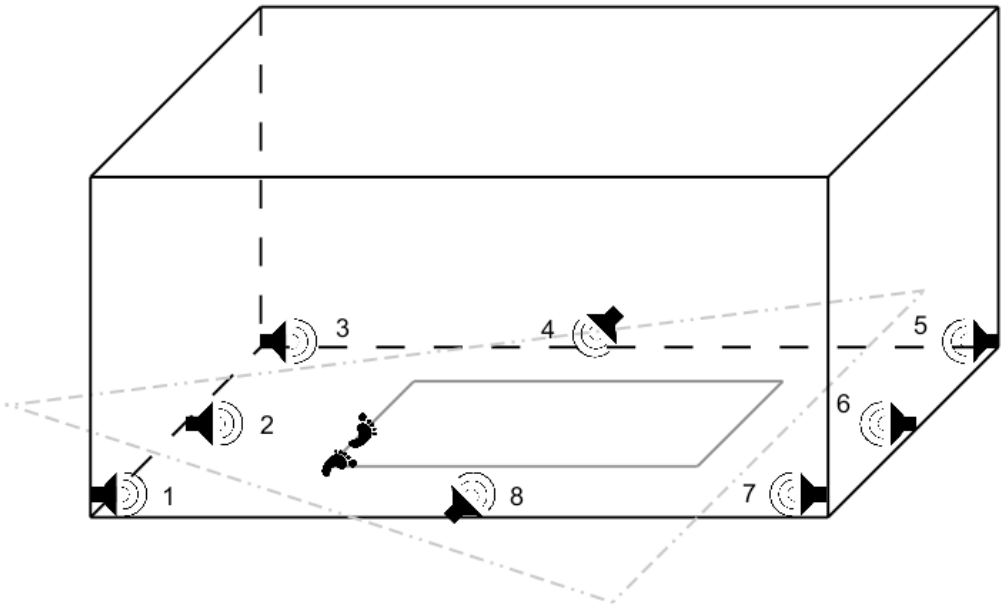


Figure 2: *Loudspeakers configuration used in this study. In dark grey the the perimeter of the rectangle delimiting the walking area (indicated on the floor by means of scotch tape strips). In light grey the triangular trajectory for the delivery of the distractors during the dynamic diffusion condition in experiment 1.*

In addition to the static and dynamic diffusion, during the experiments we also delivered the sounds using a second type of dynamic diffusion where the sound delivery was incoherent with the user position. In detail, in this configuration we chose to deliver the sound as coming from the side opposite to the user position. As an example, in reference to figure 2, when the user position was near the loudspeakers 1 and 2, the effect resulting from the dynamic diffusion was that the sound was mostly delivered through loudspeakers 5 and 6, while the loudspeakers 1 and 2 did not deliver any sound.

Finally, in presence of the dynamic diffusion, the delay due to the MoCap was negligible for the purposes of the experiments.

3. DESCRIPTION OF THE EXPERIMENTS

We performed two experiments in order to assess subjects' reaction to the sound delivery methods. During the experiments participants were asked to wear the sandals, and the cycling helmet mentioned in section 2 and walk in the laboratory according to the tasks of the two experiments.

3.1. Experiment 1

The task of the first experiment consisted on walking in circular way along the perimeter of the walking area (i.e. the rectangle indicated on the floor by means of scotch tape strips). During their walk they produced interactively footstep sounds which were delivered through the loudspeakers according to the following six conditions:

- static diffusion
- coherent dynamic diffusion
- incoherent dynamic diffusion
- static diffusion plus static distractors
- coherent dynamic diffusion plus dynamic distractors
- incoherent dynamic diffusion plus dynamic distractors

The three methods explained in section 2.2 were presented with and without distractors. Such distractors consisted of footstep sounds of a virtual person walking in the same room. Specifically, in presence of static diffusion the distractors were delivered statically (i.e., with the same volume in all the loudspeakers), while in the dynamic diffusion condition they were delivered dynamically following a triangular trajectory (see figure 2).

Participants were exposed to twelve trials, where the six conditions were presented twice in randomized order. Each trial lasted one minute.

The sound engine was set in order to synthesize footstep sounds on two different kinds of materials: wood and forest underbrush. Each condition was presented with both wood and forest underbrush. 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.

The distractors were presented using the same surface chosen for the participants' walks. In order to keep the distinction between distractors and participants' footstep sounds simple, distractors were presented with a lower volume, a small change in the timber, and with a moderately quick gait.

After the presentation of each stimulus participants were required to evaluate on a seven-point Likert scale the following questions:

- How well could you localize under your feet the footstep sounds you produced?
- How well did the sounds of your footsteps follow your position in the room?

- How much did your walk in the virtual environment seem consistent with your walk in the real world?
- How natural did your interaction with the environment seem?
- To what degree did you feel confused or disoriented while walking?

The goal of this experiment was to compare the proposed diffusion methods. The incoherent dynamic diffusion was included in the experiment in order to assess if any difference in the participants evaluations between coherent dynamic and static diffusion, was not due only to the fact that the source was moving, but that it was moving coherently with the user position. In order to make a consistent comparison, the sound synthesis engine was set with appropriate volumes for footstep sounds delivered through the static and the dynamic diffusion, i.e. there was no big volume difference between the two sound delivery methods. The distractors were used to assess if the same differences in the participants evaluations of the three diffusion methods were found both in presence and in absence of distractors.

Our hypotheses were that the coherent dynamic condition would have got better results rather than the others (in particular the static one), that the incoherent dynamic condition would have been evaluated as the worst, and that the use of distractors would have worsen the participants evaluations in comparison with the case in which the distractors were not presented.

3.2. Experiment 2

Starting from the results of the first experiment we designed a second experiment in order to investigate in a deeper way the users' perception of the static and coherent dynamic diffusions. The task of such experiment consisted on walking freely inside the walking area. Participants were exposed to fourteen trials, where seven surface materials were presented in randomized order according to the two delivery methods (static and coherent dynamic diffusion). The seven surface materials, five aggregate and two solid, were gravel, sand, snow, dry leaves, forest underbrush, wood and metal. Each trial lasted one minute. After the presentation of each stimulus participants were required to evaluate on a seven-point Likert scale the same questions presented in the first experiment.

The goal of this experiment was to assess whether participants showed a preference for one of the two proposed methods when exploring the virtual environment by walking freely (and a without predefined trajectory like in experiment 1). Furthermore we were interested in assessing whether the type of used surface could affect the quality of the results.

3.3. Results of experiment 1

The first experiment was performed by thirteen subjects, 10 males and 3 females, aged between 21 and 38 (mean=24, standard deviation=4.51), who took on average about 17 minutes to complete it. Results are illustrated in figure 3. Various ANOVA (with and without repeated measures) were performed in order to assess if the differences found in the results were significative. All post-hoc analyses were performed using the least significant difference (LSD) test with Bonferroni's correction.
The first noticeable thing is the difference between results of the two surface materials, wood and forest underbrush, for what concerns the dynamic coherent condition and its comparison with the static condition, in the case of absence of distractors. Indeed while for the wood material the differences between such two conditions are negligible for all the investigated parameters, instead for the forest underbrush material the differences are noticeable and significant (p-value = 0.000064). In presence of distractors such behavior is not hold neither for wood nor for forest, and all the differences are not significative.

A trend common to the two materials is that always the dynamic incoherent condition, both in presence and in absence of distractors, gave rise to lower evaluations in terms of localization, following, consistency and naturalness, and higher evaluations for what concerns the disorientation. In detail, for both materials, significance has been found concerning the differences between the dynamic coherent and dynamic incoherent conditions (for wood: p-value = 0.000285 and p-value = 0.005136, for forest: p-value $<$ 0.000001 and p-value = 0.043566, for the cases with and without distractors respectively), and between the static and dynamic incoherent conditions (for wood: p-value = 0.004283 and p-value = 0.002424, for forest: p-value $<$ 0.000001 both for the cases with and without distractors).

For both materials, as regards the parameters localization, following, consistency and naturalness, almost always the evaluations in absence of distractors are higher than when the distractors are present (the opposite behavior coherently happens for what concerns the disorientation parameter). This is more evident for the forest underbrush material, and indeed the difference between these two conditions is significant (p-value = 0.03817), while for wood is not.
In particular for both materials the disorientation is higher in presence of distractors rather than in absence, but significant difference between these two conditions was found only for the forest underbrush material (p-value = 0.01923). The condition dynamic incoherent with distractors was evaluated as the most disorienting in both materials; conversely, for the forest underbrush material only, the dynamic coherent condition was evaluated as the less disorienting.

As previously said, at global level the dynamic coherent condition gave rise to significant better results than the static one for what concerns the forest material in absence of distractors. In addition a successive analysis for each of the investigated parameters revealed significant difference between the two conditions only for the naturalness parameter (p-value = 0.013238).

3.4. Results of experiment 2

The second experiment was performed by ten subjects, 8 males and 2 females, aged between 19 and 37 (mean=28.8, standard deviation=5.63), who took on average about 17 minutes to complete it.
Results are illustrated in figure 4. As it is possible to notice, participants did not show any preference for one of the two methods. Evaluations of the investigated items of the questionnaire were very similar between the two methods for all the surfaces (all the differences are not statistically significant).
However it is possible to notice that the participant answers to the questionnaire items were not always similar for each surface material. In particular it is possible to observe that the metal surface on average produced the lower scores for the localization, naturalness and consistency items, and the higher scores for the disorientation item.

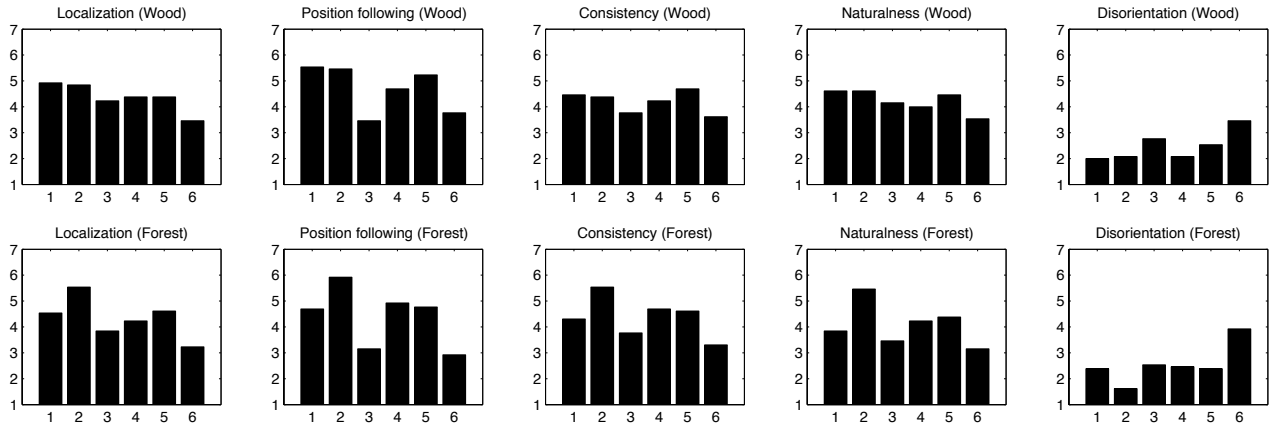


Figure 3: *Results of the first experiment. Conditions are indicated on x-axis: 1- static diffusion, 2- coherent dynamic diffusion, 3- incoherent dynamic diffusion, 4- static diffusion plus static distractors, 5- coherent dynamic diffusion plus dynamic distractors, 6- incoherent dynamic diffusion plus dynamic distractors.*

4. GENERAL DISCUSSION

The first noticeable element emerging from the results of the first experiment is the different behavior found for the coherent dynamic condition in the two simulated materials, in the case of absence of distractors. For forest underbrush such condition seems to play an important role since it got the best evaluations among all the investigated parameters, while for wood it got evaluations very similar to those of the static condition. So our hypothesis that people preferred the dynamic coherent condition to the static one was only partially confirmed.
Instead the hypothesis that the dynamic incoherent condition would have given the worst evaluations was confirmed, and found statistically significant for both materials and both in presence and absence of distractors.
It is therefore possible to conclude that users can perceive very well that their interaction with the virtual environment is not realistic nor natural when the source is not moving coherently with their position. This is an indication of the success of our simulations. The hypothesis concerning the distractors was confirmed: for both materials, almost always the evaluations in absence of distractors are better than when the distractors are present, although significant differences were found only for forest. In addition, the evaluations of the disorientation parameter were higher in presence of distractors (but significant only for forest). This indicates that the use of distractors, i.e., walking sounds evoking the presence of another person walking in the same room as the subject, is likely to influence the perception of footstep sounds associated with the subject.

Concerning the second experiment, results were clear: participants' evaluations did not differ for the two proposed methods, and this is an indication that the two methods could both be used in a virtual environment to deliver interactively generated footsteps sounds. However, other tests should be conducted in order to confirm this and assess more in detail other eventual differences in the perception of the two methods.

5. CONCLUSION AND FUTURE WORK

In this paper we have described an experiment whose goal was to assess the importance of surround sound rendering in simulating footstep sounds for virtual environments. Results show that static delivery method is not significantly preferred to the (coherent) dynamic one, and that participants disliked the renderings where footstep sounds followed a trajectory different from the one they were walking along.

In future experiments we will investigate in a deeper way the differences between the static and dynamic diffusion methods, as well as other parameters related to sound rendering, such as the role of reverberation and the role of amplitude.

We also plan to integrate the proposed footstep sounds renderings in an audio-haptic-visual environment, to design and evaluate different multimodal experiences based on walking.

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.[11]

7. REFERENCES

[1] P.R. Cook, *Real sound synthesis for interactive applications*, AK Peters, Ltd., 2002.

[2] A. Farnell, *Designing Sound*, MIT Press, 2010.

[3] P.R. Cook, "Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds," *Computer Music Journal*, vol. 21, no. 3, pp. 38–49, 1997.

[4] 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*, pp. 73–78, 2002.

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

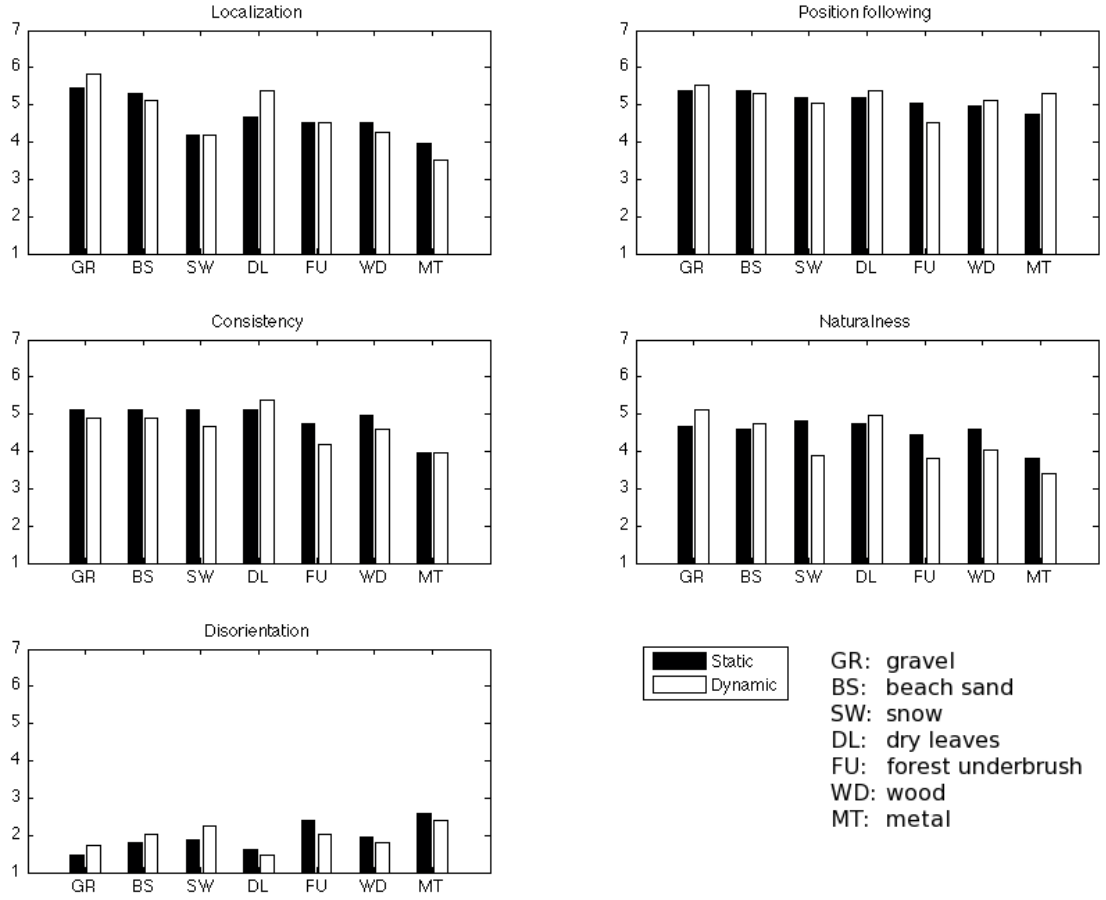


Figure 4: *Results of the second experiment.*

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

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

[7] T. Funkhouser, N. Tsingos, and J.M. Jot, "Survey of methods for modeling sound propagation in interactive virtual environment systems," 2003.

[8] L. Turchet, S. Serafin, S. Dimitrov, and R. Nordahl, "Physically based sound synthesis and control of footsteps sounds," in *Proceedings of Digital Audio Effects Conference*, 2010.

[9] R. Nordahl, S. Serafin, and L. Turchet, "Sound synthesis and evaluation of interactive footsteps for virtual reality applications," in *Proc. IEEE VR 2010*, 2010.

[10] L. Turchet, R. Nordahl, and S. Serafin, "Examining the role of context in the recognition of walking sounds.," in *Proc. of Sound and Music Computing Conference*, 2010.

[11] L. Turchet, R. Nordahl, A. Berrezag, S. Dimitrov, V. Hayward, and S. Serafin, "Audio-haptic physically based simulation of walking sounds.," in *Proc. of IEEE International Workshop on Multimedia Signal Processing*, 2010.

[12] J.C. Schacher and M. Neukom, "Ambisonics Spatialization Tools for Max/MSP," in *Proceedings of the International Computer Music Conference*, 2006.

OPEN ACCESS Freely available online

PLoS one

Vibration Influences Haptic Perception of Surface Compliance During Walking

Yon Visell[1]*, Bruno L. Giordano[2], Guillaume Millet[1], Jeremy R. Cooperstock[1]

1 Centre for Intelligent Machines and CIRMMT, McGill University, Montreal, Quebec, Canada, **2** Department of Music Research and CIRMMT, McGill University, Montreal, Quebec, Canada

Abstract

Background: The haptic perception of ground compliance is used for stable regulation of dynamic posture and the control of locomotion in diverse natural environments. Although rarely investigated in relation to walking, vibrotactile sensory channels are known to be active in the discrimination of material properties of objects and surfaces through touch. This study investigated how the perception of ground surface compliance is altered by plantar vibration feedback.

Methodology/Principal Findings: Subjects walked in shoes over a rigid floor plate that provided plantar vibration feedback, and responded indicating how compliant it felt, either in subjective magnitude or via pairwise comparisons. In one experiment, the compliance of the floor plate was also varied. Results showed that perceived compliance of the plate increased monotonically with vibration feedback intensity, and depended to a lesser extent on the temporal or frequency distribution of the feedback. When both plate stiffness (inverse compliance) and vibration amplitude were manipulated, the effect persisted, with both factors contributing to compliance perception. A significant influence of vibration was observed even for amplitudes close to psychophysical detection thresholds.

Conclusions/Significance: These findings reveal that vibrotactile sensory channels are highly salient to the perception of surface compliance, and suggest that correlations between vibrotactile sensory information and motor activity may be of broader significance for the control of human locomotion than has been previously acknowledged.

Citation: Visell Y, Giordano BL, Millet G, Cooperstock JR (2011) Vibration Influences Haptic Perception of Surface Compliance During Walking. PLoS ONE 6(3): e17697. doi:10.1371/journal.pone.0017697

Editor: Justin Harris, University of Sydney, Australia

Received November 21, 2010; **Accepted** February 8, 2011; **Published** March 25, 2011

Copyright: © 2011 Visell et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported by a grant from the Ministère du Développement économique, innovation et exportation of Quebec (http://www.mdeie.gouv.qc.ca/) for the European 7th Framework Programme project Natural Interactive Walking (NIW, no. 222107). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: yon@cim.mcgill.ca

Introduction

The goal of this study is to measure empirically the role played by vibrotactile sensory information in the perception of ground surfaces during walking. To this end, we focused on a basic property of walking surfaces that is highly salient to locomotion—their mechanical compliance [1,2]. We investigated what influence, if any, vibration feedback to the plantar soles may have on the perception of ground surface compliance during walking.

The perception of ground surfaces is instrumental to enabling us to move easily on foot in diverse natural environments. Human locomotor movements are adapted when stepping onto, off of, or moving over soft, irregular, or slippery surfaces in ways that minimize metabolic costs, reduce impact forces, or stabilize vertical center of mass [1–7]. Compliant ground surfaces, such as sand or soggy grass, perturb locomotion by degrading proprioceptive cues that are acquired via ground contact and by mechanical perturbations due to the compression of material underfoot. Walkers automatically modulate their gait pattern and biomechanics to compensate for such changes in compliance [2].

When haptic sensation in the feet is impaired, as a result of a disease such as diabetes, or through local anesthesia, it can have detrimental effects on locomotion [8–11]. However, knowledge about the influence of different sources of haptic sensory information, such as plantar force or vibromechanical stimuli, on the control of walking is incomplete.

Haptic compliance perception involves discerning the deformability of objects touched with the hand, or of surfaces felt underfoot. Compliance, the inverse of stiffness, is the ratio between displacement and applied force, $C = \Delta x/\Delta F$, and is related to the intrinsic material property of elasticity. Most prior research has investigated compliance perception via manual touch [12–19], but the haptic perceptual system is also able to discriminate walking surfaces of different elasticity [20,21]. Sensitivity is highest when there is direct contact between the surface of the skin and a deformable object. In this setting, cutaneous tactile cues predominate [17]. Conversely, when touch is mediated by a rigid link, such as a stick or a stiff shoe sole, cutaneous force cues are combined with proprioceptive information to form compliance estimates [15,16,19,22]. If cutaneous information is blocked entirely, performance is greatly degraded [16].

While we are not aware of any prior investigation of effects of vibrotactile sensory information on compliance perception, it is well established that high-frequency mechanical vibrations generated during interaction with surfaces via manually tapping or scraping with a probe, or scanning with a finger, can influence the

perception of properties such as hardness and texture [23–28]. For example, amplifying vibrations generated during manual surface scanning, or imposing sinusoidal vibrations, increases perceived surface roughness [29]. Vibrations produced during frictional sliding are indicative of movement [30], and could contribute to compliance perception. On this basis, it could be hypothesized that an amplification of plantar vibration intensity would lead to an increase in the magnitude of compliance estimates, because displacement and compliance are proportional.

Mechanical signals generated during walking on natural ground surfaces constitute a rich source of haptic sensory information [31–33]. The compression of many heterogeneous materials (e.g., wood, snow, gravel) results in inelastic, unrecoverable deformations with energy distributed over a broad frequency band [31]. The pattern of these vibrations is highly correlated with material displacement [34,35], so it is natural to consider them as potential displacement cues. Giordano et al. found that walkers are able to distinguish between the feel of porous and solid ground surfaces, or rock gravel surfaces of different grades, when walking in shoes [36]. When plantar cutaneous input was masked by mechanical vibrations, in the form of synthesized pseudo-random noise (frequency distribution: 50 Hz to 1 kHz), performance was impaired, suggesting that vibrotaction played a significant role. (Further analysis is provided in: Giordano B, Visell Y, Cooperstock JR, Yao HY, Hayward V, and McAdams S (2010) Audiohaptic identification of ground materials during walking, Submitted.)

Relatively few studies have investigated haptic perception with the feet. However, the foot is serially homologous to the hand, and is highly evolved as a sensory instrument. Its perceptual-motor abilities are involved in the regulation of posture and locomotion [37], and in the estimation of ground slipperiness and slant [38–40]. The sensory physiology of the plantar sole is highly developed, with the same type of mechanoreceptor populations as are present in the hand: the fast-adapting (FA) type I and II and slow-adapting (SA) type I and II receptors [41,42]. The sole is highly sensitive to vibration, with FA receptors comprising about 70% of the cutaneous population. Low-frequency forces are sensed by SA receptors [41], and by Golgi organs, muscle spindles, and joint capsule receptors in the muscles, tendons, and joints.

Vibromechanical stimulation of the plantar sole affects both cutaneous receptors and deeper foot and ankle proprioceptors. Such stimuli can result in real or illusory postural effects resembling those due to an increase in local pressure at the same location of the foot sole [43–46]. This could be taken to suggest that amplifying plantar vibration may, by increasing perceived forces, decrease ground compliance estimates—contrary to what is suggested by foot-ground mechanical considerations. However, studies of this type have generally been conducted while subjects stood in place, whereas haptic compliance perception always requires movement. In other experiments on vibration stimulation of the leg muscles or tendons, different effects have been observed to accompany stimulation provided during stance than those induced when it is provided during locomotion. In the former case, it induces whole-body postural tilts (attributed to illusory lengthening of the stimulated muscles), whereas during locomotion it results in modified stepping movements with little overall change in muscle coordination [47–49]. Courtine et al. argued that this reflects the fact that sensory inflow is processed depending on both the body segment where it arises and the performed task [48]. As a result, we questioned whether prior results on postural effects of plantar vibration would apply in our study, in which subjects were actively moving.

Our experiments evaluated the influence of vibrotactile information felt during stepping onto a floor surface on the perceived compliance of the latter. The above-referenced studies involved a diverse range of signal types (noise-like or natural textures, and sinusoidal stimuli), amplitudes, and temporal dependencies. Experiment 1 was designed to investigate effects of vibration feedback on perceived compliance, and to clarify their dependency on time- and frequency-domain stimulus properties. To further determine the extent to which vibrotactile sensory information is combined with cutaneous force and proprioceptive information in the perception of ground compliance, Experiment 2 measured the effect of plantar vibration on compliance perception via a novel apparatus that allowed both the mechanical stiffness of a floor plate and vibration feedback presented through it to be manipulated. Psychophysical amplitude detection thresholds for the stimuli were also measured in order to provide an indication of the relative intensity of the stimuli.

Materials and Methods

Ethics Statement

The experiments were conducted in accordance with McGill University ethics guidelines, and was reviewed and approved by the McGill Research Ethics Board in accordance with the requirements of the McGill University Policy on Ethical Conduct of Research involving Human Subjects and with the Tri-Council Policy Statement: Ethical Conduct For Research Involving Humans.

General Methods

During the experiments, subjects crossed a short walkway incorporating an actuated floor plate that provided vibration feedback in response to forces exerted by the foot. The mechanical stiffness of the plate was manipulated in Experiment 2.

Apparatus

The apparatus consisted of a short walking platform (Fig. 1) permitting subjects to take a single step onto a vibration-actuated floor plate. The plate was actuated by a Lorentz force inertial motor (Clark Synthesis model TST429) rigidly coupled to it from beneath. This plate was used to present walkers with vibration feedback and to present a specified mechanical stiffness to the walker, via a servo controlled mechanism (see Experiment 2).

To ensure that vibromechanical stimuli could be reproduced accurately across a wide range of frequencies, while assuring the stability of the plate under a human walker, we undertook an extensive redesign of our earlier apparatus [50], as fully described in reference [51]. Through measurements, we determined that the device was able to reproduce arbitrary vibrations accurately at forces of more than 40 N within a flat frequency band from 50 to 750 Hz. Within the range of amplitudes used here, vibrations could be presented with a nonlinear waveform distortion of less than 3% (mean absolute percent error, measured at 300 Hz). The device could sense static or transient loads of more than 1000 N supplied by a human foot. Analog data from the force sensors were conditioned, amplified, and digitized via a 16-bit acquisition card (National Instruments model USB-6218). Digital-to-analog conversion of the vibration signal was performed using a 24-bit, 48 kHz audio interface (Edirol model FA-101). The analog signal was then passed through a power amplifier driving the actuator. In order to assess the accurate reproduction of the vibration stimuli, they were independently recorded with a miniature accelerometer permanently attached to the underside of the plate.

General Procedure

During the experiments, stimuli were presented via the plate as subjects stepped on it. They began on one side of the walkway (see

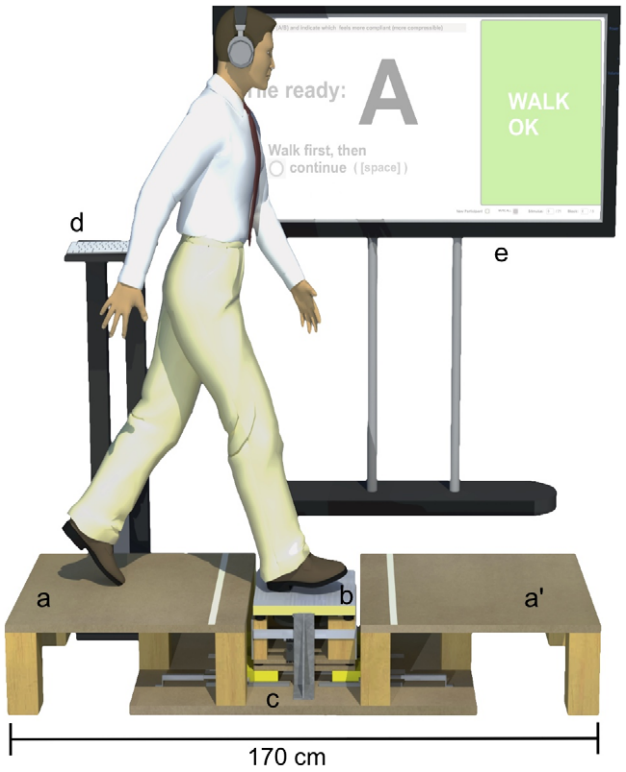


Figure 1. Apparatus for producing compliance and vibration stimuli. Subjects stepped from one side of the platform a onto the vibrating plate b, and onto the opposite platform a′. They then turned, stepped on b again, and returned to a. In Experiment 2, the plate also displaced up to 2 cm in the vertical direction, compressing a volume of EVA foam that was controlled by the linear servomechanism c, to produce the commanded compliance (see Fig. 3). Subjects entered their responses after each trial at the keyboard d and received instructions from the large-screen video monitor e.
doi:10.1371/journal.pone.0017697.g001

Fig. 1), stepped onto the plate with their dominant foot, and proceeded to the opposite side, turned, stepped on the plate again using their dominant foot, returned to the first side, and entered their responses via a computer terminal. Before each experiment, subjects were instructed in the use of the apparatus and interface.

Both experiments took place in a structurally isolated, soundproofed room with a noise-floor rating of PNC20. Subjects wore foam earplugs with an NRR attenuation rating of 33 dB and wireless headphones playing pink noise at a volume sufficient to mask any sounds produced by the vibrating plate and the motors. The non-vibrating walking platforms were isolated from the actuators via cushioning material, eliminating the transmission of vibrations to users before stimulus presentation.

A steady walking pace was enforced via a 1 Hz metronome sound audible above the pink noise. The experiments were conducted at low light levels to allow subjects to focus on what they felt, but sufficient for the walkway to remain visible. However, subjects were asked to avoid looking down at the plate while walking on it, unless necessary to maintain equilibrium, and instead, were instructed to attend to one of the two static visual markers that were positioned at a height of 1.3 m (above foot level), and a distance 1 m from either end of the walkway.

Subjects were required to wear shoes in the experiment in order to avoid directing their attention to the surface properties of the plate. In order to standardize footwear in all experiments, only male subjects were recruited, with North American shoe size between 7 and 12. Each was given an identical model men's hard soled dress shoe in the appropriate size to wear. All subjects reported normal tactile sensation in the feet, with normal walking ability, and were naive with respect to the purpose of the study. They were presented with and signed informed consent forms at the beginning of the experiment and were paid ten dollars (CAD) per hour for their participation upon completion.

Experiment 1

Experiment 1 was based on ratings of subjective compliance. We investigated the perceived compliance of a rigid plate augmented with nine different types of vibration feedback at two amplitude levels, as well as one condition in which no vibration feedback was provided. Subjects walked across each configuration of the plate, and rated its compliance on a continuous scale.

Recruitment. Twenty people participated in the experiment (mean age 24.5 years, STD = 6.9 years, average mass 70.9 kg, STD = 11.2 kg). None participated in any other study on compliance or vibration perception. Other details were as described under "General Methods".

Stimuli. The stimuli consisted of several different types of vibration feedback and one no-vibration reference condition. The stiffness of the plate was held constant, and was set equal to 90 N/mm, the median stiffness value used in Experiment 2. The 18 vibration stimuli were generated by factorial combination of three parameters: amplitude scale A (0.5, 1.0), temporal waveform type $w(t)$ (Sinusoidal, White Noise, Textured Noise), and amplitude envelope $e(t)$ (Constant, Force-Proportional, Dynamic). Each stimulus can be described as an acceleration signal delivered from the plate, having the form $s(t) = GAe(t)w(t)$, where G were stimulus-dependent peak gain factors.

The vibration signals for the nine stimuli resulting from combining the factors "amplitude envelope" and "waveform type" are shown in Fig. 2. Three different waveforms $w(t)$ were used. The first was a sinusoid $w(t) = \sin(2\pi ft)$ with frequency $f = 300$ Hz. The second was a white noise that was band-limited by filtering to remove frequencies above 700 Hz and below 50 Hz. The third was a noise signal intended to resemble the texture felt when a porous material, such as gravel, is compressed. It was obtained by passing an impulsive noise source, consisting of a random impulse train, through a resonant filter. The impulses were identical in amplitude scale, and occurred at times t_i whose time intervals $\Delta t_i = t_i - t_{i-1}$ were sampled from a Poisson stochastic process; the intervals were distributed as $P(\Delta t_i) = \lambda \Delta t_i \exp(-\lambda \Delta t_i)$. The mean event frequency was $\lambda = 0.05$ events/ms. Each impulse was rendered as a 1 ms white noise burst beginning at t_i. The impulse train was passed through a second-order infinite impulse response (IIR) bandpass filter with center frequency $f_c = 300$ Hz and bandwidth 15 Hz. The resulting noise had a rough texture with most energy concentrated in a narrow frequency band at which FA type II mechanoreceptors in the foot sole are most sensitive.

The three different envelopes $e(t)$ specified the amplitude profile of the vibration feedback in response to a footstep with normal force profile $F(t)$. The first was a constant function $e(t) = 1$, and the second was a linear force-proportional envelope $e(t) = F(t)/F_0$, with inverse slope $F_0 = 750$ N. The third was a dynamic envelope $e(t) = \dot{x}(t)$ derived from an admittance-based simulation of a linear, compressible material, where $\dot{x}(t)$ is the time-derivative of the virtual strain $x(t)$ response to $F(t)$, described by

$$\frac{1}{M}F(t) = \ddot{x}(t) + 2\zeta\omega_0\dot{x}(t) + \omega_0^2(x(t) - x_0) \quad (1)$$

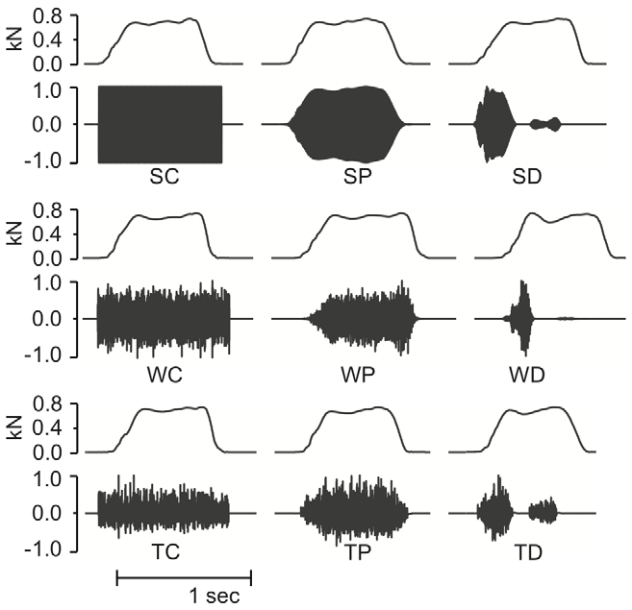


Figure 2. Vibration feedback stimuli. The thin lines (rows 1, 3, and 5) show force profiles from footsteps of one participant onto the plate, and the darker waveforms are the corresponding vibration feedback stimuli. Vibrations could be felt only during foot-plate contact. Stimuli are labeled with a 2-letter string, with the first encoding waveform type (S, W, T = Sinusoidal, White noise, Textured noise), and the second encoding envelope type (C, P, D = Constant, Proportional, Dynamic). The vibration amplitude range was normalized for display purposes.
doi:10.1371/journal.pone.0017697.g002

The envelope $e(t) = \dot{x}(t)$ is obtained in real time as the output of a second-order digital IIR bandpass filter computed by solving (1) using the Laplace transform method and the bilinear transform [52,53]. The filter input was $F(t)/M$ and the output was $\dot{x}(t)$. M could be taken as the mass of a representative volume element, but here it is an arbitrary gain factor. We used frequency $f_0 = w_0/(2\pi) = 12.5$ Hz, and set $\zeta = 1$ for critical damping. This yielded a characteristic envelope response time of $\tau = 80$ ms.

As explained in reference [51], the combination of the dynamic envelope type with the textured noise waveform (i.e., stimulus TD) can be regarded as a simplified micromechanical model for the production of textured vibrations during the compression of a natural, heterogeneous material such as gravel, sand, or snow.

Stimulus intensity equalization. In a pilot study, we observed that the vibration feedback stimuli could significantly increase subjective compliance ratings, and that the effect depended primarily on the stimulus amplitude parameter. However, RMS signal energy and subjective stimulus intensity depended on amplitude, waveform, and envelope type parameters. To ensure that the latter two could be manipulated independently of amplitude, a separate procedure was used, prior to the main part of Experiment 1, to equalize the stimuli with respect to the subjective intensity of vibration. Ten subjects that did not participate in the main experiment were recruited for this equalization experiment (mean age 23.1 years, STD = 6.1 years, average mass 71.1 kg, STD = 10.4 kg), which was based on a two-alternative forced-choice adaptive staircase method. On each trial, subjects walked across two configurations of the plate differing in vibration feedback type and amplitude and reported whether the first or second vibration felt stronger. The order of the two stimuli was random from trial to trial. One of the two, the standard, was always the high-amplitude white noise stimulus (WC1). The other, the comparison, was one of the remaining eight stimulus types parametrized by waveform and envelope (type SC, SP, SD, WP, WD, TC, TP, or TD). The amplitude of the comparison was controlled by a staircase method that tracked the point of subjective equality, i.e., the amplification factor for the comparison stimulus that rendered it as intense as the standard. If the subject indicated that the comparison felt stronger (respectively weaker), then its amplitude was reduced (respectively increased) by one step unit. The step size was initially large (10 dB) and became smaller (3 dB) after two reversals in the direction of the threshold-tracking sequence. Each staircase was run for 12 reversals, and the point of subjective intensity equivalence was calculated as the average between the last 8 reversals. A total of 16 staircases (8 interleaved pairs) were completed by each subject. Other details were as described under "General Procedure".

The results of this procedure were used to assign the values of the gain G of the Experiment 1 stimuli. The stimuli were regarded as equal in subjective intensity, within limitations determined by experiment duration and inter-subject variability. Table 1 reports the measured peak and RMS gain values for the equalized stimuli, which are labeled with a two-letter string, with the first encoding waveform (S, W, T = Sinusoidal, White noise, Textured noise), the second encoding envelope (C, P, D = Constant, Proportional, Dynamic). Values are reported for the high-amplitude ($A = 1.0$) stimuli. Both experiments also included stimuli with $A = 0.5$. Amplitudes were verified by accelerometer measurement while the plate was loaded by a footstep. For the noise stimuli (W and T), since the absolute peak could vary between presentations, a stable measure was obtained as the median of peak amplitudes on a set of 10 ms windows spanning the highest amplitude interval.

Procedure. For each stimulus presentation during the experiment, subjects walked across the plate and rated its compliance using a slider labeled "most compliant" and "least compliant" at the two extremes. Subjects were informed that they might, at times, feel vibrations via their feet, but no further elaboration was given. The first experimental block was a warm-up period in which subjects tried all configurations of the plate that would be presented. In this period, they were instructed to focus on the maximum and minimum compliance within the stimulus set. During the remainder of the experiment, subjects were asked to use the entire range of the slider when rating the stimuli. The 18 equalized stimuli were presented in blocked randomized order, each stimulus being presented once on each of 12 blocks, for a total of 216 trials. The resulting data consisted of twelve compliance ratings per stimulus from each subject. The entire experiment lasted 90 minutes. There was a pause of two minutes between blocks, and a pause of five minutes after the sixth block. Additionally, there was a pause of at least five seconds between stimuli, as subjects entered their responses. Other details were as described under "General Procedure".

Experiment 2

The experiment investigated the extent to which vibration feedback modified perception of the compliance of the floor plate when both vibration amplitude and plate stiffness were manipulated. The resulting data consisted of the proportion of responses in which the comparison was judged more compliant, for each stiffness and amplitude level.

Recruitment. Twenty new subjects participated in the experiment (mean age 23.1 years, STD = 4.1 years, average mass 69.8 kg, STD = 11.2 kg). Other details were as described under "General Methods".

Table 1. Experiments 1 and 2: Peak and RMS amplitudes of plate acceleration, G for the high amplitude $A = 1.0$ vibration stimuli.

		SC	SP	SD	WC	WP	WD	TC TP	TD
Exp. 1	Peak (m/s^2)	2.9	3.7	6.8	17.4	18.4	18.2	3.3 4.6	6.8
	RMS (m/s^2)	1.74	2.4	4.4	10.9	8.5	10.9	1.2 1.75	2.4
Exp. 2	Peak (m/s^2)								0.86
	RMS (m/s^2)								0.29

Stimulus labels: S, W, T = Sinusoidal, White noise, Textured noise waveform; C, P, D = Constant, Proportional, Dynamic envelope.
doi:10.1371/journal.pone.0017697.t001

Apparatus. The plate was integrated with a novel mechanism that allowed it to displace vertically with low friction, and that allowed us to vary the mechanical stiffness of the plate precisely for each stimulus within a range from 40 to 160 N/mm. An automated servomechanism was used to change the amount of surface area of a pair of highly recoverable, 3 cm thick foam pads inserted beneath the plate (Fig. 3).

A calibration procedure made it possible to specify plate stiffness, in values of N/mm, via computer control of foam position, with a mean accuracy of about 1%. Force measurements from the load cells in the apparatus were combined with position measurements from a precise motion capture system (OptiTrack, Model FLEX:V100R2). Calibration was performed using least squares regression fit of force to 60 force-displacement profiles consisting of more than 3000 measurements each (Fig. 4). As illustrated, the force-displacement relationship was approximately linear. Measurements from each force-displacement profile were acquired by loading the plate with a typical footstep, since, due to the finite recovery time of the foam pad, the measured stiffness could depend on the temporal profile of the load.

To ensure that the presented stiffness remained accurate, the calibration process was repeated four times in the course of the experiment, between experimental sessions. The minimum stiffness was 52 N/mm, corresponding to an absolute maximum displacement of approximately 2 cm when a subject walked across it with a maximum downward force of about 1000 N. We avoided using softer settings, as we found that it could otherwise become difficult for subjects to step normally and stably across the plate without the need to look down, which was discouraged (see "General Procedure"). The maximum stiffness was 146 N/mm, which was close to the highest level that could be well controlled. At stiffer settings, the intrinsic compliance of the vibration mounts (location b in Fig. 3) would have a non-negligible influence on the plate compliance in ways that depended on foot location and orientation. In addition, stiffness discrimination underfoot would likely have become less reliable [21].

All motorized movements resulting from compliance changes were performed smoothly to minimize any vibrations due to the mechanism that might otherwise provide information about the foam configuration. For the same reason, the duration of any mechanical reconfiguration was kept constant ($t = 4$ s).

Stimuli. Stimuli were configurations of the floor plate with one of seven different stiffness levels (52, 64, 76, 90, 106, 124 or 146 N/mm) and one of three vibration feedback conditions. The vibrations were the textured, dynamic type (TD). Two different non-zero vibration amplitudes were used, respectively 18 dB and 24 dB lower than the high amplitude ($A = 1.0$) TD stimuli from Experiment 1, as well as a no-vibration condition. These lower amplitudes were selected through pre-testing to ensure that the resulting psychophysical data would be useful. Although Experiment 1 demonstrated that all vibration stimuli could increase perceived compliance, the TD type was selected for further testing because, among those with the highest mean compliance ratings (within one standard error of the mean), they had the shortest duration and one of the lowest RMS amplitudes, limiting the possibility of sensory adaptation during vibration stimulus presentation. The seven stiffnesses and three vibration levels resulted in 21 different comparison stimuli.

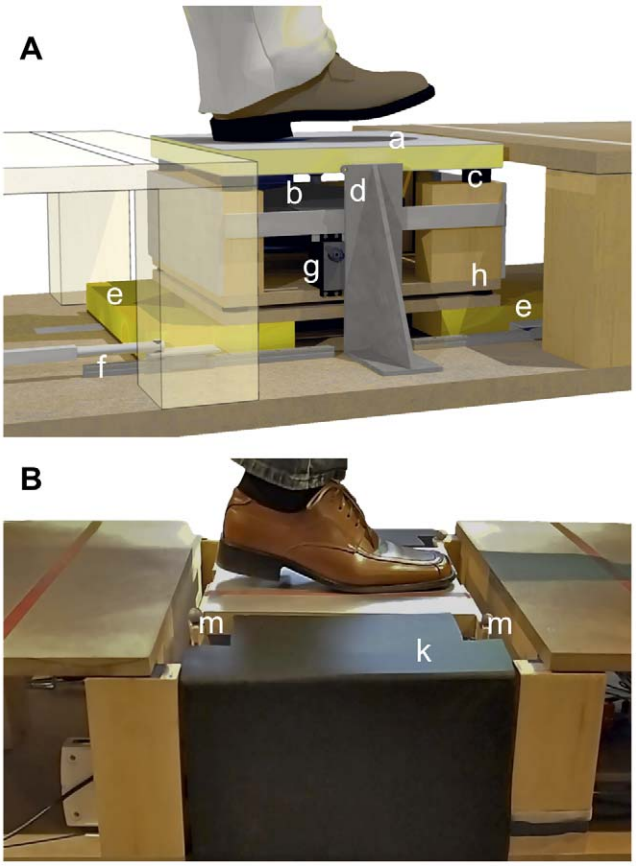


Figure 3. View of the variable compliance mechanism. A. Users stepped onto vibrating plate a, which was driven by vibration actuator b and mounted on suspension c. The plate displaced in the vertical direction, guided by low-noise ball bearing slides d, and compressing a pair of foam inserts e. To produce the commanded compliance, the foam inserts e were positioned by the linear servomechanisms f before each stimulus presentation, while the plate assembly was lifted by servos g. Participant-applied forces were measured by load cells h under four corners of the plate assembly. B. Image of the apparatus and shoe as used in the experiment. Opaque panels k and fabric (not shown) hid the device configuration from subjects' view. Four optical motion capture markers m tracked the displacement of the plate with high precision.
doi:10.1371/journal.pone.0017697.g003

Procedure. The experiment was based on the psychophysical method of constant stimuli, using a two-alternative forced-choice paradigm. Subjects walked across pairs of configurations, one after the other, and responded indicating which felt more compliant. Vibration was added to the comparison stimulus, except in the "no-vibration" condition, and was never added to the standard. The resulting data consisted of the proportion of responses in which each comparison configuration (that is, stiffness and vibration level) was judged more compliant.

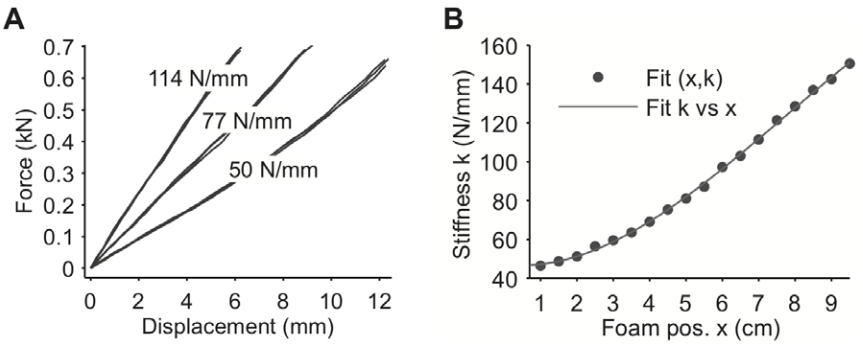


Figure 4. Force-displacement profiles and fits of stiffness vs. foam position used to calibrate the apparatus. A. Examples of three compression profiles are shown (overlaid) for each of three stiffness values. Each calibration was based on sixty such profiles, with more than 3000 data points each. B. Calibration curve fit of stiffness vs. foam position based on measurements at each position and stiffness.
doi:10.1371/journal.pone.0017697.g004

The standard and comparison were presented, as in Experiment 1, in sequential randomized order. Because stiffness was manipulated in the experiment, subjects were required to pause for five seconds between each half of a stimulus pair, in order to provide enough time for the stiffness modification to complete. They were automatically cued to pause and to continue by the software graphical user interface. The timing of this pause was always the same, to avoid giving any indication of the amount of change in compliance. Other details were as described under "General Procedure".

Subjects were told that during the main experiment they would be asked to respond indicating which of the two configurations felt more compliant. They were told that they might, at times, feel vibrations via their feet, but no further elaboration was given. The first experimental block was a warm-up period in which subjects tried six randomly generated stimulus pairs, consisting of random stiffnesses in the range used in the experiment, and vibration feedback of type WC, with a similar intensity to that used in the main experiment, although the type was different.

During the main experiment, all stimulus pairs were presented in each block, in randomized, balanced order. Randomization of stimulus order was independent for each session and each subject, and no blocks were repeated. Subjects were required to leave the apparatus and pause for one minute between blocks, and for four minutes after each third block. They each completed three experimental sessions, comprising a total of twenty blocks. The duration of the first two sessions was 90 minutes and that of the last session was 1 hour. No more than two experimental sessions, separated by at least two hours, were permitted for any subject on a single day. Each subject was presented with each of the 21 pairs a total of 20 times.

Subjects completed a post-experiment questionnaire and interview, which asked whether the vibrations were felt, and what decision strategy was used (see Results).

Psychophysical detection thresholds for the stimuli. Immediately after subjects completed the questionnaire, they participated in a final stage of this experiment, which measured their psychophysical detection thresholds for amplitude. The stimuli consisted of individual configurations of plates, set to the median stiffness level of 90 N/mm, accompanied by vibration feedback of type TD, as used in the main experiment. Amplitude was manipulated independently during the procedure.

The threshold-measurement procedure was based on a single-interval adaptive yes/no staircase method developed by Lecluyse and Meddis for auditory threshold testing [54]. They found, through experiments and simulations, that this procedure yielded similar thresholds to those obtained with two-interval forced choice or maximum likelihood methods, resulted in less variation, and required fewer trials. The latter was a consideration in our experiment, because of the required level of activity and duration, which could lead to fatigue.

Stimuli were presented one at a time, and amplitude was controlled by the staircase procedure. Subjects responded after each presentation indicating whether they felt a vibration from the plate or not. They were instructed to be as sensitive as possible without guessing. When they responded "yes" (resp. "no"), then the amplitude was reduced (resp. increased) by one step unit. The step size was initially large (10 dB) and was reduced to a smaller level (3 dB) after two reversals in direction of the threshold-tracking sequence. For each comparison, one staircase was started at a high amplitude randomly chosen between +8 dB and +15 dB (referenced to the Experiment 2 stimulus amplitude of $0.43\ \mathrm{m/s^2}$) and one at a low amplitude between -8 dB and -15 dB. The two staircases were interleaved, with one of the two randomly selected for presentation on each trial. In order to prevent guessing, 13% of the trials were randomly selected as catch trials, in which no vibration was present. Subjects were warned that if they answered "yes" on a catch trial, both staircases would be re-started. The median number of times that subjects were caught guessing in the experiment was 1.5 (minimum zero, maximum three). In addition, on every 10th trial, subjects were presented with a no-vibration stimulus and were told that no vibration was present, to remind them how that condition felt.

Each staircase was continued until 12 reversals were reached. A total of six staircases (three interleaved pairs) were completed by each subject, and the threshold was calculated as the average between the last eight reversals from all six staircases. Subjects were required to pause for 2 minutes between staircase pairs. The total duration for each subject was approximately 30 minutes.

Results

Vibration Stimulus Factors Influencing Subjective Compliance Judgements

Data from Experiment 1 were analyzed to determine the effect of vibration feedback type on compliance ratings. Mean ratings for all the 19 stimuli are shown in Figure 5. Paired t-tests showed that all 18 vibrating stimuli were perceived as significantly more compliant than the non-vibrating one ($t(19) \geq 7.04$, $p < 0.001$).

We further analyzed compliance ratings in Experiment 1 with a within-subject repeated measures ANOVA, with amplitude, envelope, and waveform as factors. Mean compliance ratings for

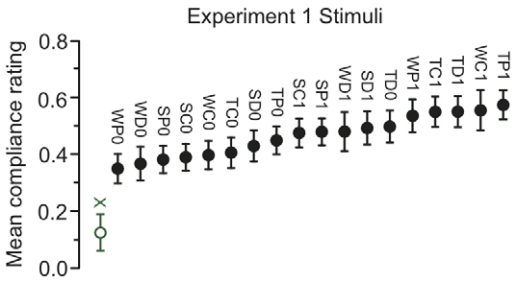


Figure 5. Experiment 1: Subjective compliance ratings for all 19 stimuli, averaged across subjects. A higher value means more compliant (less stiff). "X" labels the no-vibration stimulus. Others are labeled with a 3-letter string encoding waveform, envelope, and amplitude level (0, 1 = Linear amplitude 0.5, 1.0). Error bars = ±1 standard error of the mean (SEM). All vibrating stimuli were significantly more compliant than the no-vibration stimulus.
doi:10.1371/journal.pone.0017697.g005

the values of each factor are shown in Fig. 6. The one-way effect of amplitude was significant ($p<0.001$, $F(2,19)>33.0$, $\eta_p^2 \geq 0.635$), as was that of waveform ($p \leq 0.01$, $F(2,19)>5.24$, $\eta_p^2 \geq 0.216$), but the effect of envelope type was not significant ($F(2,19) \leq 0.057$, $p \leq 0.02$, $\eta_p^2 \leq 0.003$). There were significant effects of all two-way interactions, amplitude × waveform ($p<0.001$, $F>11.5$, $\eta_p^2 \geq 0.378$), amplitude × envelope ($p \leq 0.013$, $F>5.8$, $\eta_p^2 \geq 0.235$), and waveform × envelope ($p<0.001$, $F>9.273$, $\eta_p^2 \geq 0.328$). There was no significant three-way effect ($p>0.07$, $F<2.3$, $\eta_p^2 \leq 0.106$). Based on this analysis, amplitude had the largest influence on compliance judgments, while the effects of waveform and of all two-way interactions were smaller.

Effect of Vibration Feedback on Compliance Perception

The results of Experiment 2 consisted of proportions of responses at which the standard stimulus, a non-vibrating plate with stiffness 90 N/mm, was judged less stiff than a comparison that varied in stiffness and in vibration amplitude. Figure 7 presents the average response proportions. A one-way ANOVA of the response proportions for the factor *amplitude* indicated that vibration significantly decreased stiffness at each stiffness level (see Table 3). Although the variation in responses was larger at higher stiffnesses, the effect of amplitude on subjective compliance was proportionally larger, yielding a higher level of significance. For all subjects, average response proportions at the largest amplitude level ($0.86\ m/s^2$) were higher than in the no-vibration case.

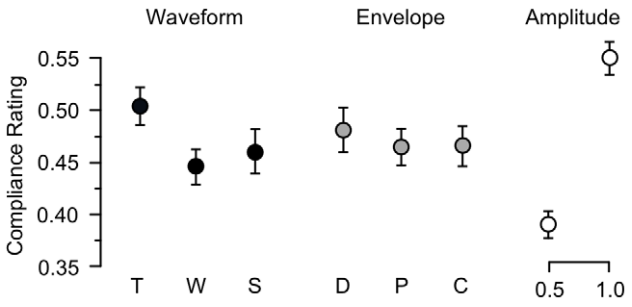


Figure 6. Experiment 1: Average compliance ratings for the three stimulus factors "waveform type", "envelope type", and "amplitude". Stimulus labels are as given in Fig. 5. Error bars: ±1 SEM.
doi:10.1371/journal.pone.0017697.g006

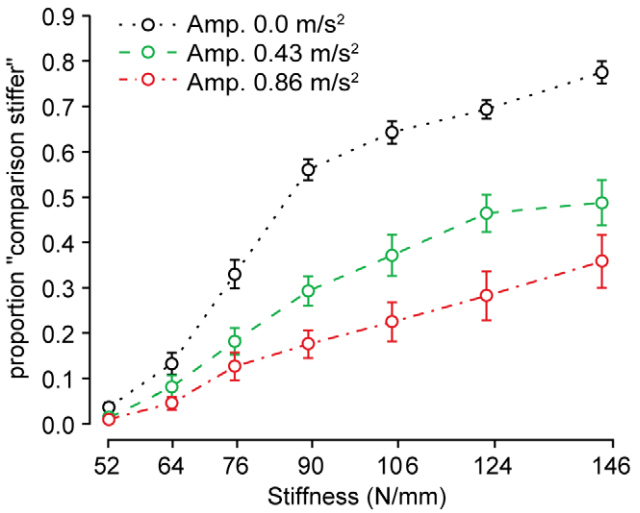


Figure 7. Experiment 2: Mean proportion of comparison stimuli judged stiffer than the standard. A higher proportion implies a judgment of "stiffer". The standard had a stiffness of 90 N/mm, as indicated by the dashed line, and did not present any vibration feedback. Results shown are averaged between all 20 subjects. Error bars: ±1 SEM.
doi:10.1371/journal.pone.0017697.g007

Binary response data at each vibration amplitude level from each of the 20 subjects were fitted to a cumulative normal distribution using a probit regression model. A total of 60 fits were performed. The models explain 86.5% of the variance in the data with Pearson correlation $\rho = 0.93$. The slope, intercept, and point of subjective equivalence (PSE) in stiffness were computed from each fit. Median values in each condition are given in Table 2 and shown in Figure 8. We investigated influences of vibration amplitude on the fit parameters using a nonparametric Friedman test, in order to ensure that the analysis would remain robust to outliers in the data. The latter resulted from a few subjects whose response proportions increased slowly with stiffness in the high-amplitude condition, leading to unusually small slope values, and large PSEs. Vibration amplitude did not significantly affect intercept ($\chi^2(2) = 3.9$, $p>0.14$), indicating that subjects were not biased to indiscriminately answer "softer" when vibration amplitude was higher. PSE increased significantly with amplitude ($\chi^2(2) = 34.9$, $p<10^{-7}$), indicating that the stimulus was perceived as softer when vibration was present, and slope also increased ($\chi^2(2) = 17.2$, $p<0.0002$), indicating that stiffness discrimination performance was impaired in the presence of vibration. Nonparametric repeated-measures tests contrasting all three amplitude levels indicated no significant effect on intercept ($p>0.3$, Bonferroni corrected, BC), but did reveal an effect of amplitude on PSE for all pairings ($p<0.01$, BC). There was a significant effect on slope between the no-vibration and either high- or low-amplitude conditions ($p<0.035$, BC), but not between the low- and high-amplitude vibration conditions ($p>0.34$, BC). Increasing the vibration amplitude level from low to high thus increased the bias in the PSE for stiffness estimation without further decreasing discriminability.

Experiment 2 also measured psychophysical amplitude thresholds for detection of the vibration stimuli, for the same subject pool used in the main part of the experiment, with the plate stiffness set to the median value of 90 N/mm. These thresholds were measured in shoes, and would likely be lower if direct skin contact were involved. The measurements were based on a fast, single

Table 2. Experiment 2: Mean proportions of comparison stimuli that were judged to be stiffer than the standard, as a function of stiffness, with different levels of vibration.

	Response Proportion at Amp. A			one-way ANOVA	
Stiffness	0.0 m/s^2	0.43 m/s^2	0.86 m/s^2	$F(2,57)$	p
52 N/mm	0.034	0.010	0.0068	4.3	0.018
64 —	0.13	0.078	0.042	4.24	0.019
76 —	0.33	0.18	0.13	12.5	$<10^{-4}$
90 —	0.557	0.29	0.17	50.3	$<10^{-12}$
106 —	0.64	0.37	0.22	31.6	$<10^{-9}$
124 —	0.69	0.46	0.28	26.9	$<10^{-8}$
146 —	0.77	0.49	0.36	22.0	$<10^{-7}$

doi:10.1371/journal.pone.0017697.t002

interval yes-no procedure with catch trials [54]. Although Lecluyse and Meddis found this method to yield similar thresholds to those obtained using a two-alternative forced choice task, it could be argued to have led some subjects in our experiment to adopt conservative criteria for responding during the detection staircase, which would yield an overestimate of the thresholds. Figure 9 presents the results of the measurements. The mean threshold was 0.46 m/s^2, with standard deviation 0.18 m/s^2. Measured thresholds for 10 of the 20 subjects were higher than the low-amplitude (0.43 m/s^2) stimulus by more than two standard errors of the mean. For the subgroup of 10 participants with the highest thresholds, we analyzed proportions of responses "more compliant" from the no-vibration (0.0 m/s^2) to the low-amplitude (0.43 m/s^2) vibration condition at the same stiffness value (90 N/mm) as was used in the threshold measurement. A paired two-tailed t-test revealed a significant effect of amplitude on these response proportions (mean 0.41 vs. 0.71 with $p<10^{-4}$, $t(18)>5.2$); see Figure 9. The median PSEs of the psychometric fits for the same subgroup of 10 participants were also significantly higher in the low-amplitude condition than in the no-vibration condition (median 135.2 N/m vs. 99.2 N/m, Friedman $p\leq 0.0005$, $\chi^2=12$), and were close in value to the median PSEs for the complete subject pool. However, the threshold values for the entire subject pool ($n=20$) were not significantly correlated with PSE values in either the low or high vibration amplitude condition, with differences in PSE values between vibration and no-vibration cases, or with mean response proportions at stiffness 90 N/mm (Spearman $|\rho|<0.21$, $p>0.37$ in all cases).

Table 3. Experiment 2: Median values of the PSE, slope, and intercept from per-subject psychometric curve fits at each vibration level.

	Median Value at Peak Amplitude A			Friedman Test	
	0.0 m/s^2	0.43 m/s^2	0.86 m/s^2	$\chi^2(2)$	p
PSE (N/mm)	100.8	134.5	151.6	34.9	$<10^{-7}$
Slope (mm/N)	0.025	0.018	0.018	17.2	<0.0002
Intercept	−2.50	−2.70	−2.77	3.9	>0.14

doi:10.1371/journal.pone.0017697.t003

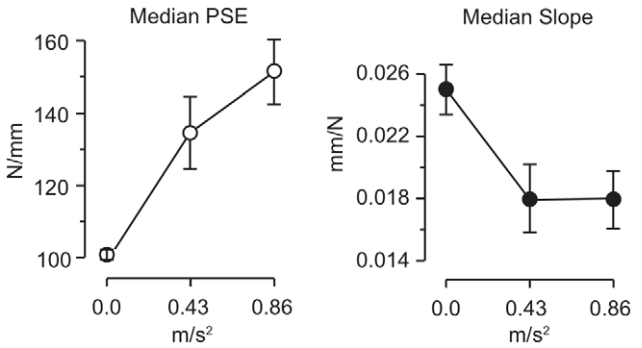


Figure 8. Experiment 2: Median values of the PSE and slope from per-subject psychometric fits at each vibration amplitude level. Error bars: ±1 SEM (outliers excluded).
doi:10.1371/journal.pone.0017697.g008

The post-experiment questionnaire that was completed by Experiment 2 subjects prior to the threshold measurement included the question: "Did you feel any vibration produced by the tile?" Five out of twenty subjects responded "No". A sixth noted that he felt vibration on just 5% of trials, and a seventh reported feeling no vibrations, but did report feeling "a creaking", "like stepping on an old hardwood floor" on some trials. The remaining 13 subjects answered "Yes".

Discussion

These experiments demonstrate that the perceived haptic compliance of a walking surface is increased in the presence of plantar cutaneous vibration feedback. In Experiment 1, we found

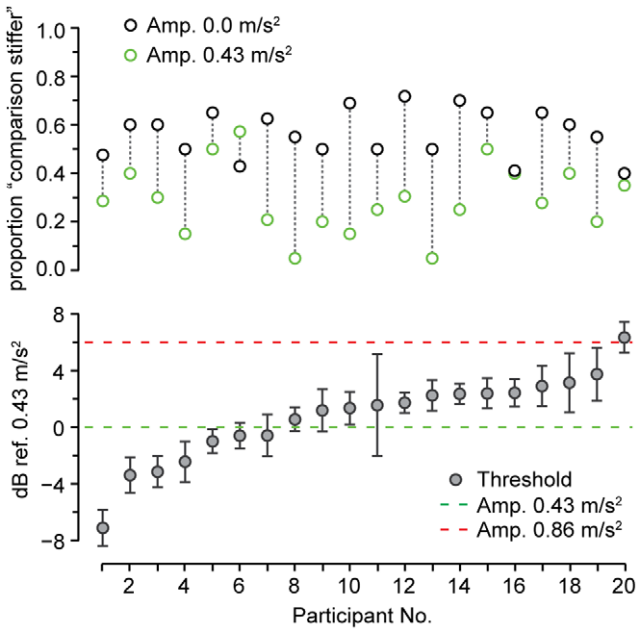


Figure 9. Experiment 2: Psychophysical amplitude detection thresholds and response proportions at stiffness level 90 N/mm. Top: Response proportions for all 20 subjects at two lowest vibration levels, and stiffness level 90 N/mm. Bottom: Amplitude threshold levels, displayed in dB referenced to 0.43 m/s^2. Subjects are sorted in order of increasing threshold (same ordering top and bottom). The dashed lines indicate the amplitudes of the vibrating stimuli used in Experiment 2. Error bars: ±2 SEM.
doi:10.1371/journal.pone.0017697.g009

that the largest effect on perceived compliance was due to vibration amplitude, and that other stimulus factors had a weaker influence. We also found that an increase in perceived compliance could be achieved with types of vibration feedback that differed in waveform, amplitude envelope, or the frequency distribution of their energy.

In Experiment 2, we found that vibromechanical stimuli with peak amplitudes of only 0.43 and $0.86\ \mathrm{m/s^2}$ could elicit an increase in perceived compliance. These levels were 18 dB smaller than those used in Experiment 1. This held at all tested levels of stiffness. A substantial increase in the stiffness of the vibration-augmented plate was required for it to be perceived as having the same stiffness as the non-augmented one, as the vibration feedback produced positive relative shifts in the median PSE values for stiffness of 34% and 50% at the two amplitude levels.

The amplitudes used in Experiment 1 were comparable to those that are experienced during normal walking on natural granular materials such as sand or gravel [31,36], while those used in Experiment 2 were significantly weaker. Through pre-testing for Experiment 2, we determined that higher amplitudes tended to dominate the influence of mechanical stiffness over the range explored. The upper limit of the stiffness range used approached a level at which stiffness perception underfoot is less reliable [21], while the smallest stiffness was near the limit of what we determined subjects could comfortably and safely walk on with this apparatus.

None of the experiments involved training, and the effects observed did not require awareness that vibration feedback was being provided. We can conclude that vibration felt during stepping on a rigid surface is combined with the mechanical stiffness of the surface in the haptic perception of compliance. In addition, the results show that the variation of vibration feedback alone is sufficient to elicit a percept of compliance.

The compliance estimation task adopted in this study resembled prior experiments in which subjects used their hands or arms to estimate the haptic compliance of spring-loaded mechanisms or other objects with non-deformable surfaces [17,19,22]. Based on those results, and on considerations of contact mechanics, it was expected that subjects in our experiments required both force and displacement information (from kinesthetic and tactile channels) in order to judge compliance. In this light, it appears that added vibration feedback results in a modification of force and/or displacement information that increases compliance estimates. As noted earlier, localized vibration stimulation of the foot sole has been shown to have a similar effect on postural control to an increase in force sensation at the same location [43,44], and this could be thought to influence compliance judgments. However, an ideal observer combining force F and displacement x to estimate compliance using the formula $C = x/F$ would produce lower compliance estimates as force sensation is increased. Our results show an opposite tendency, so it appears unlikely that the observed effects on perceived compliance were mediated by increased sensations of applied force. Furthermore, stimuli used in Experiment 2 had a peak amplitude of $0.43\ \mathrm{m/s^2}$, less than 0.5% as large as the smallest amplitude used in the aforementioned studies.(To compare stimuli, we computed accelerations used in experiments by Roll et al. [43,44] from the stimulus properties they reported.) Furthermore, in our experiment, stimuli were felt through a shoe, whereas those used in the aforementioned studies involved direct skin contact.

It might be suggested that the observed results could be attributable to sensory adaptation of SA I afferents due to the vibromechanical stimuli, which could yield a reduction of force estimates. However, in our experiments, exposure times averaged less than 1 second, with at least 5 seconds between presentations, whereas mean adaptation times for SA I afferents are about 10 seconds [55,56]. Additionally, Experiment 2 was based on pairwise comparison of two stimuli, only one of which could include vibrations. Also, the low-amplitude Experiment 2 stimuli were 0.6 dB weaker than the mean psychophysical detection threshold measured for our subject pool. While this might be partly attributable to a tendency of the measurement method used to overestimate the thresholds, it nonetheless appears unlikely that these stimuli could have produced a significant adaptation of SA I responses, even after long exposure times. Furthermore, no subjects reported feeling any desensitization in their feet, and several were unaware of the vibrations. Thus, it appears unlikely that sensory adaptation played a significant role.

Conversely, prior studies have demonstrated that sub-threshold levels of plantar stimulation with vibration noise can enhance cutaneous sensitivity in the foot soles, stabilizing posture [57]. Although such an effect, if present, could have improved haptic force discrimination in our experiment, it would not be expected to influence mean compliance estimates, so does not seem to be able to explain our main results.

Experiments 1 and 2 compared the perceived compliance of plates with and without vibration feedback. A priori, due to this categorical difference, subjects could have responded based on cognitive criteria unrelated to a sensation of compliance. However, there are several reasons we do not believe cognitive effects played an important role. Subjects were consistent in responding that the vibrating plates were more compliant, and no subject inverted this relation. Experiment 2 results did not indicate any tendency on the part of subjects to respond indiscriminately that the vibrating stimulus was "softer" independent of actual compliance. In addition, vibration had a significant influence on compliance at both near-threshold levels (Experiment 2) and at much higher ones (Experiment 1). Furthermore, some subjects in Experiment 2 reported that they were not consciously aware of the presence of vibration feedback. Finally, a few subjects described what they felt in a way that is consistent with the notion of a material being compressed underfoot, and similar responses have been received over the course of numerous demonstrations of the apparatus to naive users.

Taken together, our findings appear to be consistent with the hypothesis that vibration feedback supplied a cue that tended to increase perceived displacement during stepping, due to a sensorimotor contingency similar to that experienced when stepping on a natural material (e.g., snow, gravel) or displacing a mechanism with friction (e.g., a pedal or slider). Assuming this to be the case, and supposing that perceived force was not affected, an ideal observer would infer an increase in compliance that grows linearly with the increased sensation of displacement, due to the relation $C = \Delta x/F$. In this model, a relative increase in estimated displacement of 25.0% and 33.5% in the low- and high-amplitude vibration conditions, respectively, would be required to explain the shifts in median stiffness PSE values measured in Experiment 2.

One counterintuitive finding is that stimuli with amplitude envelopes that were constant could evoke an increased sensation of compliance, contrary to the idea that vibration supplies a force-dependent displacement cue. However in the conditions of this study, vibromechanical energy transmitted to the leg increased with applied force F, due to the increased coupling of foot and plate, even for constant stimuli. As a result, the feedback could appear to have been generated by a stepping action even when there was no explicit relation with applied force, due to the transitive nature of foot-plate contact.

A number of disorders, the most common being diabetes, can impair cutaneous tactile sensation in the feet and have detrimental

effects on locomotion [8–10]. This has led various researchers to investigate relations between sensory impairment and control of balance or locomotion. The present study suggests that vibrotactile sensation may be more involved in the regulation of walking in natural environments than has been acknowledged. One pilot study found that step-synchronized plantar vibration feedback during foot-ground contact may improve locomotion in Parkinson's disease patients, but there were insufficient controls to rule out learning or attentional effects [58]. However, it is plausible that the effects investigated here could play a role in such settings.

Through preliminary body kinematic analyses using motion capture measurements we have found indications of postural modifications during stepping when plantar vibration feedback is supplied. This would be consistent with postural kinesthetic illusions that, absent restraints, result in compensatory body sway [44]. If confirmed, such results might one day prove useful for the development of gait rehabilitation techniques or vibrotactile orthotics. We intend to explore these questions further in future work. Finally, we note that the same compliance illusion seems to be present during interaction via the hands. This is also something we plan to investigate further, in order to situate our results relative to prior literature on manual haptic perception.

Acknowledgments

The authors wish to thank Stephen McAdams for providing access to the Music Perception and Cognition Lab facilities, Gil Zilberstein and Alison Conard for assisting with running the experiments, and the CIRMMT staff for providing technical support. They thank Vincent Hayward and Nizar Ouarti for helpful discussion.

Author Contributions

Conceived and designed the experiments: YV BLG JC. Performed the experiments: YV BLG GM. Contributed reagents/materials/analysis tools: YV GM JRC BLG. Wrote the paper: YV BLG GM JC. Designed the apparatus: YV GM.

References

1. Dixon S, Collop A, Batt M (2000) Surface effects on ground reaction forces and lower extremity kinematics in running. Med Sci Sports Exerc 32: 1919–1926.
2. MacLellan M, Patla A (2006) Adaptations of walking pattern on a compliant surface to regulate dynamic stability. Experimental Brain Research 173: 521–530.
3. Ferris D, Louie M, Farley C (1998) Running in the real world: adjusting leg stiffness for different surfaces. Proceedings of the Royal Society B: Biological Sciences 265: 989.
4. Hardin E, van den Bogert A, Hamill J (2004) Kinematic adaptations during running: effects of footwear, surface, and duration. Medicine & Science in Sports & Exercise 36: 838.
5. Moritz C, Farley C (2004) Passive dynamics change leg mechanics for an unexpected surface during human hopping. Journal of Applied Physiology 97: 1313.
6. Moritz C, Greene S, Farley C (2004) Neuromuscular changes for hopping on a range of damped surfaces. Journal of Applied Physiology 96: 1996.
7. Marigold D, Patla A (2005) Adapting locomotion to different surface compliances: neuromuscular responses and changes in movement dynamics. Journal of neurophysiology 94: 1733.
8. Simmons R, Richardson C, Pozos R (1997) Postural stability of diabetic patients with and without cutaneous sensory deficit in the foot. Diabetes research and clinical practice 36: 153–160.
9. Simoneau G, Ulbrecht J, Derr J, Becker M, Cavanagh P (1994) Postural instability in patients with diabetic sensory neuropathy. Diabetes Care 17: 1411.
10. Mueller M, Minor S, Sahrmann S, Schaaf J, Strube M (1994) Differences in the gait characteristics of patients with diabetes and peripheral neuropathy compared with age-matched controls. Physical Therapy 74: 299.
11. Stål F, Fransson PA, Magnusson M, Karlberg M (2003) Effects of hypothermic anesthesia of the feet on vibration-induced body sway and adaptation. Journal of Vestibular Research 13: 39–52.
12. Harper R, Stevens SS (1964) Subjective hardness of compliant materials. Quarterly Journal of Experimental Psychology 16: 204–215.
13. Scott-Blair G, Coppen F (1940) The subjective judgement of the elastic and plastic properties of soft bodies. British Journal of Psychology 31: 61–79.
14. Freyberger F, Färber B (2006) Compliance discrimination of deformable objects by squeezing with one and two fingers. In: Proc. EuroHaptics volume 2006: 271–276.
15. Tan H, Durlach N, Beauregard G, Srinivasan M (1995) Manual discrimination of compliance using active pinch grasp: The roles of force and work cues. Perception and Psychophysics 57: 495–510.
16. Srinivasan M, LaMotte R (1995) Tactual discrimination of softness. Journal of Neurophysiology 73: 88.
17. LaMotte R (2000) Softness discrimination with a tool. Journal of Neurophysiology 83: 1777.
18. Friedman R, Hester K, Green B, LaMotte R (2008) Magnitude estimation of softness. Experimental brain research 191: 133–142.
19. Bergmann Tiest W, Kappers A (2009) Cues for haptic perception of compliance. IEEE Transactions on Haptics 2: 189–199.
20. Giordano B, Visell Y, Cooperstock JR, Yao H, Hayward V, et al. (2010) Audiohaptic identification of ground materials during walking. (Submitted).
21. Kobayashi Y, Osaka R, Hara T, Fujimoto H (2008) How accurately people can discriminate the differences of floor materials with various elasticities. IEEE Transactions on Neural Systems and Rehabilitation Engineering 16: 99–105.
22. Jones LA, Hunter IW (1990) A perceptual analysis of stiffness. Experimental Brain Research 79: 150–156.
23. Klatzky R, Lederman S (1999) Tactile roughness perception with a rigid link interposed between skin and surface. Perception and Psychophysics 61: 591–607.
24. Hollins M, Risner S (2000) Evidence for the duplex theory of tactile texture perception. Perception & Psychophysics 62: 695–705.
25. Bensmaia S, Hollins M (2003) The vibrations of texture. Somatosens Mot Res 20: 33–43.
26. Bensmaia S, Hollins M (2005) Pacinian representations of fine surface texture. Perception & psychophysics 67: 842.
27. Klatzky R, Lederman S, Hamilton C, Grindley M, Swendsen R (2003) Feeling textures through a probe: Effects of probe and surface geometry and exploratory factors. Perception & Psychophysics 65: 613.
28. Okamura A, Dennerlein J, Howe R (1998) Vibration feedback models for virtual environments. In: 1998 IEEE International Conference on Robotics and Automation, 1998. Proceedings. volume 1.
29. Hollins M, Fox A, Bishop C (2000) Imposed vibration influences perceived tactile smoothness. Perception 29: 1455–1466.
30. Howe R, Cutkosky M (1989) Sensing skin acceleration for slip and texture perception. In: Proc. IEEE International Conference on Robotics and Automation. volume 1: 145–150.
31. Ekimov A, Sabatier J (2006) Vibration and sound signatures of human footsteps in buildings. J Acoust Soc Am 120: 762–768.
32. Galbraith F, Barton M (1970) Ground loading from footsteps. Journal of the Acoustical Society of America 48: 1288–1292.
33. Visell Y, Fontana F, Giordano B, Nordahl R, Serafin S, et al. (2009) Sound design and perception in walking interactions. International Journal of Human-Computer Studies.
34. Herrmann H, Roux S (1990) Statistical models for the fracture of disordered media. North Holland.
35. Alava M, Nukala P, Zapperi S (2006) Statistical models of fracture. Advances in Physics 55: 349–476.
36. Giordano BL, McAdams S, Visell Y, Cooperstock J, Yao H, et al. (2008) Non-visual identification of walking grounds. Journal of the Acoustical Society of America 123: 3412.
37. Meyer PF, Oddsson LIE, Luca CJD (2004) The role of plantar cutaneous sensation in unperturbed stance. Experimental brain research 156: 505–512.
38. Kinsella-Shaw JM, Shaw B, Turvey MT (1992) Perceiving "walk-on-able" slopes. Ecological Psychology 4: 223–239.
39. Donelan JM, Kram R, Kuo AD (2002) Mechanical work for step-to-step transitions is a major determinant of the metabolic cost of human walking. Journal of Experimental Biology 205: 3717–3727.
40. Lockhart T, Woldstad J, Smith J, Ramsey J (2002) Effects of age related sensory degradation on perception of floor slipperiness and associated slip parameters. Safety science 40: 689–703.
41. Trulsson M (2001) Mechanoreceptive afferents in the human sural nerve. Experimental Brain Research 137: 111–116.
42. Kennedy PM, Inglis JT (2002) Distribution and behaviour of glabrous cutaneous receptors in the human foot sole. J Physiol 583.
43. Roll R, Kavounoudias A, Roll J (2002) Cutaneous afferents from human plantar sole contribute to body posture awareness. Neuroreport 13: 1957.
44. Kavounoudias A, Roll R, Roll J (1998) The plantar sole is a 'dynamometric map' for human balance control. Neuroreport 9: 3247.
45. Maurer C, Mergner T, Bolha B, Hlavacka F (2001) Human balance control during cutaneous stimulation of the plantar soles. Neuroscience letters 302: 45–48.

46. Thompson C, Bélanger M, Fung J (2010) Effects of plantar cutaneo-muscular and tendon vibration on posture and balance during quiet and perturbed stance. Human Movement Science: In press.
47. Ivanenko Y, Grasso R, Lacquaniti F (2000) Influence of leg muscle vibration on human walking. Journal of Neurophysiology 84: 1737.
48. Courtine G, Nunzio AD, Schmid M, Beretta M, Schieppati M (2007) Stance- and locomotion-dependent processing of vibration-induced proprioceptive inflow from multiple muscles in humans. Journal of neurophysiology 97: 772.
49. Verschueren S, Swinnen S, Desloovere K, Duysens J (2002) Effects of tendon vibration on the spatiotemporal characteristics of human locomotion. Experimental brain research 143: 231–239.
50. Visell Y, Law A, Cooperstock J (2009) Touch Is Everywhere: Floor Surfaces as Ambient Haptic Interfaces. IEEE Transactions on Haptics 2: 148–159.
51. Visell Y, Cooperstock JR (2010) Design of a vibrotactile display via a rigid surface. In: Proc. of the IEEE Haptics Symposium.
52. Smith JO (2010) Physical audio signal processing. W3K Publishing.
53. Oppenheim A, Schafer R (2009) Discrete-time signal processing. Prentice Hall Signal Processing. pp 1120.
54. Lecluyse W, Meddis R (2009) A simple single-interval adaptive procedure for estimating thresholds in normal and impaired listeners. The Journal of the Acoustical Society of America 126: 2570.
55. Leung Y, Bensmaia S, Hsiao S, Johnson K (2005) Time-course of vibratory adaptation and recovery in cutaneous mechanoreceptive afferents. Journal of neurophysiology 94: 3037.
56. Bensmaia S, Leung Y, Hsiao S, Johnson K (2005) Vibratory adaptation of cutaneous mechanoreceptive afferents. Journal of neurophysiology 94: 3023.
57. Priplata A, Niemi J, Salen M, Harry J, Lipsitz L, et al. (2002) Noise-enhanced human balance control. Physical Review Letters 89: 238101.
58. Novak P, Novak V (2006) Effect of step-synchronized vibration stimulation of soles on gait in Parkinson's disease: a pilot study. Journal of NeuroEngineering and Rehabilitation 3: 9.

Walking on virtual ground: physics, perception, and interface design

Yon Visell

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

March 2011

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

© 2011 Yon Visell

Abstract

The sensorimotor capacities of the foot are crucial to human locomotion in diverse environments, to gathering information about walking surfaces, and to interacting with objects on the ground. Locomotion is increasingly employed to allow users to control and navigate within immersive virtual environments, but, in contrast to the hand, little attention has been given to the rendering of haptic sensations for the feet.

This thesis addresses several challenges motivated by the problem of realizing haptic experiences of walking on virtual ground surfaces. First, a novel family of interfaces is introduced, based on a vibrotactile display integrated in a rigid floor plate. Its structural dynamics and controller have been optimized to ensure its ability to accurately reproduce mechanical vibrations over a wide frequency band, which was instrumental to realizing the perceptual study presented in the second part of the thesis. Distributed arrays of these devices are used to simulate virtual ground surfaces and floor-based multi-touch surfaces, whose usability for human-computer interaction is empirically demonstrated.

The second component of this thesis is an experimental study of the contribution of vibrotactile sensory information to the perception of ground surface compliance. A novel haptic perceptual illusion is demonstrated, in which the apparent compliance of a floor surface is increased by vibrations felt via the plantar sole of the foot. This investigation also revealed the surprising ability of the vibrotactile floor interface to overcome, in part, a core limitation: its inability to display kinesthetic force-displacement information.

The third part of the thesis analyzes texture-like mechanical signals produced through inelastic physical processes in complex, disordered materials like those encountered during walking in many natural terrains. Patterns of fluctuations accompanying sliding friction and fracture processes in quasi-brittle, heterogeneous materials subjected to time-varying loads are characterized using methods from statistical physics. This analysis was used to formulate novel algorithms for the haptic synthesis of high-frequency signatures of fracture processes in fiber composites and compressed granular media.

In conclusion, this thesis presents an innovative hardware interface and techniques for interacting with virtual ground surfaces. It also demonstrates a new haptic perceptual effect that lends justification to the display paradigm adopted here. Finally, it analyzes and models transient, texture-like physical phenomena associated with stepping onto complex, natural ground materials.

Résumé

Les capacités sensori-motrices du pied sont essentielles à la locomotion humaine, à la collecte d'informations sur les surfaces de marche, et à l'interaction avec des objets au sol. La locomotion est de plus en plus utilisée pour interagir et naviguer dans les environnements virtuels immersifs, mais, contrairement à la main, peu d'attention a été accordée au rendu des sensations haptiques pour les pieds.

Cette thèse aborde plusieurs problèmes liés à la réalisation d'expériences haptiques de marche sur des terrains virtuels. Tout d'abord, une nouvelle famille d'interfaces est présentée, fondée sur un dispositif vibrotactile intégré dans un carreau rigide. Sa dynamique structurelle et son contrôleur ont été optimisés pour assurer sa capacité à reproduire fidèlement les vibrations mécaniques dans une large bande de fréquence, ce qui était nécessaire á la réalisation de l'étude de perception présentée en deuxième partie de la thèse. Un pavage de ces dispositifs est utilisé pour simuler des terrains virtuels et des planchers tactiles multi-points, dont l'ergonomie est démontrée de manière empirique.

Le deuxième volet de cette thèse est une étude expérimentale sur la contribution de l'information vibrotactile à la perception de la compliance du sol. Une nouvelle illusion perceptuelle haptique est démontrée, dans laquelle la compliance apparente du sol est augmentée par les vibrations ressenties par la plante du pied. Cette étude a également révélé l'étonnante capacité de l'interface vibrotactile à surmonter, en partie, une limitation intrinsèque : son incapacité à transmettre des informations kinesthésiques force-déplacement.

La troisième partie de la thèse analyse les signaux mécaniques complexes produits par les processus physiques inélastiques dans les matériaux désordonnés tels que ceux rencontrés lors de la marche en terrain naturel. Les modèles de fluctuations accompagnant le frottement de glissement et les processus de fracture dans les matériaux hétérogènes quasi-fragiles soumis aux charges variables sont caractérisés par des méthodes de physique statistique. Cette analyse est utilisée pour formuler de nouveaux algorithmes pour la synthèse haptique des signatures à hautes fréquences des processus de fracture dans les composites de fibres et les materiaux granulaires compressés.

En conclusion, cette thèse présente un dispositif vibrotactile et des techniques novateurs pour interagir avec des terrains virtuels. Elle démontre un nouvel effet perceptuel qui justifie le paradigme d'interaction haptique adopté ici. Enfin, elle analyse et modélise certains phénomènes physiques associés à la marche sur des terrains naturels complexes.

Acknowledgments

This thesis would not have been possible without the support of my colleagues, mentors, and friends. I would like to thank Prof. Jeremy Cooperstock for his supervision, and for his openness and enthusiastic support of this research despite changes in its direction. A number of colleagues at McGill collaborated with me on this work and greatly enriched it. Bruno Giordano, a collaborator on the study presented in Chapter 6, taught me a great deal about perception and experimental rigour. Guillaume Millet was often the one ready to ask questions that weren't being raised. He collaborated on the experiments of Chapter 6, and also provided valuable input on the material simulations presented in Chapter 7. Work by Alvin Law, Severin Smith, and Jessica Ip was instrumental to realizing the multimodal floor interface described in Chapter 5. I would like to thank all of members of the Shared Reality Environments lab for sharing their own energies and realities over the last few years.

I received vital input and encouragement at critical moments from members of my PhD committee, Prof. Vincent Hayward, Prof. Marcelo Wanderley, and Prof. Stephen McAdams. Special thanks to the staff in the Centre for Intelligent Machines and in the Centre for Interdisciplinary Research in Music Media and Technology, especially Cynthia Davidson, for helping everything to run smoothly, and to Don Pavlasek of the mechanical workshop, who provided helpful consulting and assistance with fabrication. Prof. Luc Devroye offered generous advice on the subtleties of random variate generation. External collaborators including Nizar Ouarti, Federico Fontana, Gabriel Cirio, Anatole Lécuyer, and Maud Marchal provided useful and stimulating discussion related to this work.

This research received support from the Quebéc Ministère du Développement économique, de l'Innovation et de l'Exportation (MDEIE) for the European Community 7$^{\text{th}}$ Framework Programme project Natural Interactive Walking (NIW, no. 222107), from the the European Community 6$^{\text{th}}$ Framework Programme project CLOSED (no. 029085), from student research awards from the Centre for Interdisciplinary Research in Music Media and Technology, and from an entrance award and international graduate student fellowships from the Department of Electrical and Computer Engineering.

I thank Karmen Franinović for lending me her support and inspiration, and for being my best collaborator on everything outside of this thesis.

Contents

1 Introduction **1**
1.1 Scope . . . 3
1.2 Overview . . . 3
1.3 Summary of contributions . . . 6

2 Background **9**
2.1 Dynamics and forces in walking . . . 9
2.2 Mechanical interactions and materials . . . 11
2.3 Vibrotactile sensation and perception . . . 11
2.4 Display and interaction techniques . . . 13
2.4.1 Walking in virtual worlds . . . 14
2.4.2 The foot in human-computer interaction . . . 15
2.5 Vibrotactile interfaces and physically based synthesis . . . 16
2.5.1 Hardware . . . 16
2.5.2 Physically based synthesis . . . 18

3 Floor Surfaces as Ambient Haptic Interfaces **21**
3.1 Introduction . . . 23
3.1.1 Potential Advantages . . . 25
3.2 Background . . . 26
3.2.1 Tactile sensation and psychophysics of the foot . . . 26
3.2.2 Prior research on interfaces for the feet . . . 28
3.3 Ambient haptic floors: display design . . . 30
3.3.1 Roles for haptic communication . . . 31
3.3.2 Physical interaction design considerations . . . 32

3.3.3 Prototyping technologies and methods . . . 32
3.3.4 Prototype vibrotactile device . . . 34
3.3.5 Temporal interaction design considerations . . . 35
3.3.6 Three interaction cases . . . 36
3.3.7 Vibrotactile stimulus design . . . 37
3.4 Evaluation . . . 41
3.4.1 Methodology . . . 42
3.4.2 Results and Discussion . . . 43
3.5 Conclusion . . . 48

4 Vibrotactile Display Design 51
4.1 Introduction . . . 53
4.1.1 Vibrotactile augmentation of floor surfaces . . . 53
4.1.2 Background . . . 54
4.2 Display concept and motivation . . . 55
4.2.1 Transparency . . . 56
4.2.2 Application to normal force texture simulation . . . 57
4.3 Components . . . 58
4.3.1 Mechanical Structure . . . 58
4.3.2 Sensing . . . 59
4.3.3 Actuation . . . 60
4.3.4 Electronics . . . 60
4.4 Dynamic Response . . . 61
4.4.1 Actuator Response . . . 61
4.4.2 Plate Response . . . 61
4.4.3 Elastic Suspension . . . 65
4.4.4 Simulation . . . 67
4.4.5 Modeling user-supplied loads . . . 68
4.5 Evaluation . . . 69
4.5.1 Frequency response measurement . . . 70
4.5.2 Digital correction . . . 71
4.6 Conclusion . . . 71

5 Interaction Techniques **75**
5.1 Introduction . . . 77
5.2 Background . . . 77
5.2.1 Floor Touch User Interfaces: Potential application space . . . 78
5.3 System . . . 79
5.4 Contact sensing . . . 80
5.5 Channeling material interactions . . . 83
5.5.1 Non-visual rendering . . . 83
5.5.2 Visual animation and control . . . 85
5.6 Floor Touch Surface Interfaces . . . 86
5.6.1 Preliminary User Evaluation . . . 87
5.7 Conclusions . . . 90

6 Vibration Increases Haptic Perception of Surface Compliance **93**
6.1 Abstract . . . 95
6.2 Introduction . . . 95
6.3 Materials and Methods . . . 98
6.3.1 Ethics Statement . . . 98
6.3.2 General Methods . . . 99
6.3.3 Apparatus . . . 99
6.3.4 General Procedure . . . 99
6.3.5 Experiment 1 . . . 101
6.3.6 Experiment 2 . . . 105
6.4 Results . . . 110
6.4.1 Vibration Stimulus Factors Influencing Subjective Compliance Judgements . . . 110
6.4.2 Effect of Vibration Feedback on Compliance Perception . . . 112
6.5 Discussion . . . 116

7 Simulating Fracture Processes in Heterogeneous Materials **121**
7.1 Introduction . . . 122
7.2 Fracture processes in fiber composites: Fiber Bundle Model . . . 124
7.2.1 Stochastic process formulation . . . 126

7.2.2 Mean damage approximations . . . 129
7.2.3 Numerical simulations . . . 131
7.3 Stress fluctuations in compressed granular media . . . 133
7.3.1 Inverse Fiber Bundle Model . . . 134
7.3.2 Sequential algorithm and mean damage approximation . . . 137
7.3.3 Numerical simulations . . . 139
7.4 Real-time, interactive simulation . . . 140
7.5 Conclusions . . . 143

8 A first passage time problem associated with stick-slip oscillations 145
8.1 Introduction . . . 146
8.2 Stick-slip motion as a first passage time problem . . . 148
8.2.1 Approximate slip time densities . . . 151
8.2.2 Joint first passage densities . . . 153
8.2.3 Upcrossing densities . . . 155
8.3 Stick-slip motion in a Langevin friction model . . . 156
8.3.1 Upcrossing density . . . 157
8.3.2 Results . . . 159
8.4 Conclusion . . . 166

9 Conclusions 167
9.1 Results . . . 167
9.1.1 Devices . . . 167
9.1.2 Information Display and Interaction Techniques . . . 168
9.1.3 Perception . . . 169
9.1.4 Mechanical Interactions . . . 169
9.2 Future Work . . . 170

References 173

List of Figures

2.1 Vibration spectrogram $a(t, f)$ and low-frequency normal foot-ground force $F(t)$ measured at the hard sole of a men's shoe during one footstep of a walker onto rock gravel. 10

3.1 A possible end-user scenario: Pedestrians receive vibrotactile cues via the ground surface, indicating the location of the crossing and state of its signal. 24
3.2 Image and diagram of the interactive, four-tile floor interface. 34
3.3 A 2×2 tile prototype seen from above. Force sensing resistors are positioned as shown. 35
3.4 Pictorial representations of three interaction paradigms. 37
3.5 The eight rhythms (R1 to R8) from which both sets of stimuli were built. . 38
3.6 Waveforms used to generate the vibrotactile stimuli used in the experiment. 40
3.7 Example of a stimulus used in the experiment, resulting from the model settings that produced the first "impact" waveform of Figure 3.6 in combination with the first rhythmic pattern of Figure 3.5. 41
3.8 Correct identification rates for each vibrotactile message set. 44
3.9 Confusion pattern within the set of eight WS stimuli and the set of eight IT stimuli. 44
3.10 Mean correct identification rates in each session. The error bars show 95% confidence intervals. 45
3.11 Box plot summary of the mean correct identification rate for the IT (left side) and WS (right side) stimulus sets. 46

4.1 The experience of stepping onto a deformable object or surface is simulated via a rigid vibrotactile interface. 55

4.2 The haptic display concept. 56
4.3 Vibration spectrogram $a(t, f)$ and normal force $F(t)$ measured from one footstep onto rock gravel. 57
4.4 Normal force texture synthesis. 58
4.5 Vibrotactile floor interface hardware for a single tile unit. 59
4.6 Typical acceleration frequency response of the Clark Synthesis TST family of actuators. 61
4.7 Composite sandwich plate, shown in profile. 63
4.8 The suspension and tile system. 65
4.9 Resonant supension frequency f_0 and static deflection δz vs. Young's modulus. 66
4.10 Visualization of the first two resonant modes of vibration. 67
4.11 FEM analysis of the acceleration response measured at four different points on the plate. 68
4.12 A simplified model of a user supplied load on the device. 69
4.13 The measured log magnitude frequency response of the display. 70
4.14 Corrected and uncorrected device response, without foot contact. 72

5.1 The floor interface, as situated within an immersive, rear projected virtual environment simulator. 79
5.2 Diagrammatic view of the floor interface. 80
5.3 Overview of system components and logical connections between them. . . 80
5.4 The normal force distribution $p_R(\mathbf{x})$ and associated contact centroid position $\mathbf{x}_c$. 82
5.5 Results of 50 measurements comparing true normal force positions (circles) with contact centroid estimates (Xs). 82
5.6 A recorded sequence of contact centroids produced by an individual walking across the floor. 83
5.7 Still images of users interacting with the simulated frozen pond. 83
5.8 Behavior at the crack front $\mathbf{c}_k$ is modeled using a simplified fracture mechanics treatment. 84
5.9 A crack pattern, modeled as a graph of lines between nodes $\mathbf{c}_i$ extending from the seed location p_0. 85
5.10 Still image from video of a user interacting with floor-based interface widgets. 86

5.11 Configuration and stimuli used in the floor touch surface experiment. . . . 88
5.12 Successful target selection rate vs. distance, angle of presentation (measured away from preferred foot), and button width, inclusive and exclusive of the farthest targets. 89

6.1 Apparatus for producing compliance and vibration stimuli. 100
6.2 Vibration feedback stimuli. 102
6.3 View of the variable compliance mechanism. 106
6.4 Force-displacement profiles and fits of stiffness vs. foam position used to calibrate the apparatus. 107
6.5 Experiment 1: Subjective compliance ratings for all 19 stimuli, averaged across subjects. 111
6.6 Experiment 1: Average compliance ratings for the three stimulus factors. . 111
6.7 Experiment 2: Mean proportion of comparison stimuli judged stiffer than the standard. 112
6.8 Experiment 2: Median values of the PSE and slope from per-subject psychometric fits at each vibration amplitude level. 114
6.9 Experiment 2: Psychophysical amplitude detection thresholds and response proportions at 90 N/mm. 115

7.1 Fiber Bundle Models have been successfully used to model fracture physics in quasi-brittle materials including fiber composites such as wood (left), and granular media, such as sand or gravel (right). 123
7.2 Fiber Bundle Model of fracture in a parallel array of N brittle fibers. . . . 125
7.3 The sequential sampling algorithm described equations (7.2-7.4). 128
7.4 Illustration of the mean damage model. 129
7.5 Typical stress-controlled loading responses of the Fiber Bundle Model. . . 131
7.6 The constitutive stress-strain relationship $\sigma(x)$ for models M, $\tilde{M}$, and $\bar{M}$. . 133
7.7 Statistical distribution of times between failure events. 134
7.8 Empirical distribution of the energy E of fracture fluctuations. 135
7.9 Granular flow through a tube can become obstructed by large forces resulting from force chains bridging the walls. 136

7.10 The inverse fiber bundle effectively models a compressed granular medium as a parallel array of fibers, which represent the percolation of one-dimensional force chains through the medium. 136
7.11 Compression results in rapid events through which force chains are restructured, becoming stronger, stiffer and straighter. 137
7.12 Response of the inverse Fiber Bundle Model under ramp loading conditions. 139
7.13 Empirical distribution of times between failure events, Δt and the size of energy fluctuations E. 140
7.14 The response of a bundle of 2000 fibers to a dynamic load, simulating that due to a footstep. 141
7.15 Representative data from the interactive implementation of Fiber Bundle Model M. Data shown corresponds to a model with 350 fibers. Acoustic transfer is simulated by a resonant filter with 10 modes. Left column, Top to Bottom: Force input $F(t)$, strain $x(t)$, threshold $\xi(t)$, and time-frequency spectrogram of acoustic energy of fracture. Right: Fracture event waveform. 142
7.16 The simulation of high-frequency failure processes in localized domains within a distributed medium. 144

8.1 A viscoelastic body sliding across a surface, an example of a system that exhibits frictional stick-slip oscillations. 147
8.2 Phase-space portraits for three model friction laws. 149
8.3 152
8.4 Two phase space trajectories $(x(t), v(t))$ for the stochastic harmonic oscillator system. 160
8.5 The first upcrossing density, $n_1(t)$, which describes the level crossing behavior of the system in the absence of absorption from to slip-to-stick transitions. 161
8.6 Comparison of two different analytic approximations for the first passage time density. 162
8.7 Comparisons of the analytic approximation $\mathcal{F}_g^{(1)}(T)$ with simulation results. 163
8.8 The distribution of stick-slip oscillation periods. 163
8.9 Noise-driven excitation of stick-slip oscillations. 164
8.10 The distribution $p(d)$ of slip displacements: analytical result and Monte Carlo estimate. 165

List of Tables

6.1 Peak and RMS amplitudes of plate acceleration G for the high amplitude $A = 1.0$ vibration stimuli. 105
6.2 Experiment 2: Mean proportions of comparison stimuli that were judged to be stiffer than the standard. 113
6.3 Experiment 2: Median values of the PSE, slope, and intercept from per-subject psychometric curve fits at each vibration level. 114

Chapter 1

Introduction

Our feet are instruments of action and of self-motion, but are also used to perceive our surroundings through the sense of touch. Although tactile perception is finest in the hand, it is also highly developed in the lower extremities. The foot is capable of experiencing most of the range of haptic sensations that are felt via the hand, including kinesthesia (the sense of force and position) and cutaneous sensations including pain, temperature, and vibration. Arguably, the foot is the most often used haptic sensory organ, in the sense that it is nearly always active in touching the ground, regulating balance and locomotion. In this respect, it is one of the most important sources of haptic information about our surroundings.

Haptics plays a distinguished role relative to the other senses, as an important component of it is the bidirectional interaction with our physical surroundings, which, in the case of the foot, largely manifests as self-motion. A diverse variety of ground surfaces and features are encountered during everyday walking activities. Walking and interacting on foot generate a rich variety of multimodal information, much of it available through touch. This information is used to aid motor tasks with the feet, and to identify our surroundings. Nonetheless, the nature of haptic perception in the foot is less understood than in the case of the hand, having mainly been studied in relation to its effects on balance and locomotion.

Haptic interfaces for human-computer interaction have advanced greatly in the last two decades, but lag far behind visual and auditory displays. One reason is the difficulty involved in specifying and displaying haptic stimuli, which consist of forces or displacements distributed on the body surface and the dependence of these signals on body movements during interaction with objects in the environment. The large range and variety of signals

to which the haptic sense is sensitive amplifies this challenge. Existing interfaces are the result of severe tradeoffs concerning what aspects of haptic interaction are perceptually most important, which can be implemented most efficiently, and what paradigms are most useful or most likely to proliferate.

Just as walking is fundamental to the negotiation of natural environments, it is increasingly relevant to human-computer interaction, because immersive interactive systems are becoming more widely used, and because more everyday environments are being augmented with digital technology. Despite the important role played by the feet in numerous human activities, they have received limited attention as a means of interacting with computers.

Although rarely incorporated in early research on virtual environments, walking is increasingly investigated as a means of navigating in immersive virtual spaces, and has been found to have a positive effect on the sense of presence. However, despite remarkable advances in visual display and rendering techniques, little has been done to ensure that the haptic information users obtain about their surroundings through walking agrees with what they experience through the other senses.

The few haptic interfaces for the foot that have been developed have largely been designed to enable locomotion in virtual environments. Existing devices, which are based on similar principles to those used in motorized exercise equipment, have been used to simulate walking, running, turning, or other aspects of locomotion on uneven terrain. However, to date, almost no attention has been given to the simulation of other ground material properties. Indeed, no haptic interface is useful for displaying all possible force signals, and those that simulate kinesthetic aspects of locomotion, which involve large forces and body movements, are ill-suited to reproducing finer textural details that distinguish various natural and man-made ground surfaces.

Vibrotactile interfaces, which are designed to mirror the haptic subsystem that is most sensitive to fast or high-frequency mechanical stimuli, offer a complementary alternative. They are low in cost and complexity, because they are electronically and mechanically simple and can often be operated in open loop control mode. Although unable to simulate kinesthetic information, they have proven effective at rendering object properties, such as surface hardness and texture, that are reflected through high-frequency mechanical vibrations, and are poorly reproduced by most force-feedback devices.

However, little is known about the nature of vibrotactile stimuli that are produced during walking on natural ground surfaces, which include materials that are rarely used

in haptic simulations. Common examples, such as branches, snow, sand, or gravel, are heterogeneous, collective, and disordered, and the complex, inelastic mechanical responses that they exhibit when subjected to loads are responsible for the textural "feel" of walking on them. It appears to be computationally unfeasible to interactively simulate these distributed processes on a microscopic scale, due to the number of elements involved and to their complex organization. Effective statistical models have been developed in the science and engineering literature, but have not yet been applied to haptic rendering.

1.1 Scope

This thesis is motivated by the aim of allowing users to walk on and "feel" natural or man-made ground surfaces that are simulated in a virtual environment. The virtual ground surface is simulated by means of haptic (and, optionally, auditory or visual) feedback provided by a real, instrumented floor. The sensation of touching a virtual surface is produced by modifying the displacement of the floor in response to forces exerted by the foot. As a practical measure, only texture-like, high-frequency temporal components are displayed, because it is far less efficient to reproduce static forces or displacements. Thus, the simulation is based primarily on vibrations presented and felt through a flat, instrumented floor display.

The thesis addresses several problems motivated by this basic scenario. It presents the design of a novel vibrotactile floor interface and uses it to realize virtual ground and touch surface interfaces. It also investigates the contribution of vibrotactile sensory information to the perception of ground surface properties, and demonstrates that vibrotactile feedback can, to some extent, compensate for missing kinesthetic information. This seems to be attributable to the fact that the displacement of many natural materials or mechanisms is accompanied by texture-like vibrations that are produced through inelastic processes like friction and fracture. The manifestation of these processes in some disordered materials, like those found in natural ground surfaces, is analyzed using tools from stochastic physics.

1.2 Overview

This thesis is organized in nine chapters: this introduction, a review of prior literature, three previously published manuscripts, three chapters (not yet submitted for publication),

and a summary.

Chapter 2 provides a survey of prior literature relevant to the realization of experiences walking on, or interacting with, virtual ground surfaces. It contains an overview of relevant aspects of locomotion, a survey of vibrotactile sensation and perception in the feet, a discussion of mechanical interactions in walking on natural and man-made surfaces, a review of prior literature on human-computer interaction via the feet, a discussion of factors affecting the design of vibrotactile interfaces, and a review of methods for the physically based rendering of vibrotactile stimuli. Due to the breadth of the topics involved, further background is reviewed in the relevant sections of the chapters that follow.

Chapter 3 presents applications of a novel vibrotactile floor display to ambient information display in intelligent environments. Potential applications of such displays as navigational aids, non-intrusive information conduits, aids to visually impaired people, entertainment features, and components of immersive virtual environments are discussed. Two methods for the design of vibrotactile messages presented through a floor interface are presented and evaluated in an experiment. The results support the feasibility of ambient vibrotactile information display through floor surfaces, and provide guidelines for the design of usable vibrotactile messages.

Chapter 4 presents the re-engineered design of the vibrotactile floor interface, which was used in other chapters of this thesis to implement and evaluate interaction techniques and rendering algorithms for virtual ground surfaces. The most notable part of this chapter is the consideration given to optimizing the structural dynamics of the device subject to constraints imposed by the need to support static loads from users standing on it. Digital inverse filtering is used to further correct the unloaded device transfer function response, achieving a flat passband from 50 to 750 Hz, and making it well suited for use in the perceptual investigation conducted in Chapter 6.

Chapter 5 investigates new techniques for interacting with computationally augmented floor surfaces. It introduces a distributed, networked version of the vibrotactile floor interface presented in Chapters 3 and 4. Interaction techniques are developed for this distributed display, based on intrinsic contact sensing methods that make it possible to resolve foot-floor contact loci with a resolution of about 2 cm. They are used to realize interactive, multi-

modal simulations of natural walking terrains and multimodal touch surface interfaces for human-computer interaction. A usability evaluation was conducted, yielding preliminary guidelines for the design of such applications.

Chapter 6 presents a study that directly addresses an inherent limitation of the vibrotactile floor interface: its inability to display kinesthetic force-displacement information, hence to reproduce the compliant feel of different ground surfaces. A new haptic perceptual illusion is demonstrated, in which the perceived compliance of a floor surface is increased by vibration feedback felt through it. A series of experiments revealed that the effect scales monotonically with vibration amplitude, and is robust at amplitudes near to the psychophysical detection threshold for the stimuli. The results suggest that vibrotactile channels play a larger role in the control of locomotion than has been previously acknowledged.

Chapter 7 and **Chapter 8** investigate mechanical interactions associated with experiences of walking on disordered, heterogeneous materials like those commonly encountered in natural environments. Although Chapters 4 and 5 presented simple stochastic models for rendering texture-like vibrations associated with the inelastic compression of such materials, here a more fundamental basis for such algorithms is provided, based on statistical physics.

Chapter 7 presents and analyzes a novel methods for simulating failure processes in fiber reinforced composites and granular materials subjected to dynamic loads. An efficient, temporal model for these fracture-like processes is derived based on a stochastic process associated with random spatial variations in the local strain thresholds of a distributed medium.

Chapter 8 analyzes irregular stick-slip motion produced through frictional sliding or compression of rough or disordered surfaces, using methods from statistical physics. A novel first passage time problem is formulated, whose solutions determine the statistical properties of stick-slip motion in a generic class of stochastic friction models. Approximate solutions to this problem are developed here for the first time. They are shown to capture rich, pseudo-periodic motions that are not present in any deterministic limit, but instead correspond to noise activated transient oscillation modes similar to those associated with stochastic resonance phenomena observed in other physical and biological systems.

Chapter 9 concludes this thesis with a summary and a discussion of the main findings and implications for future research.

The novel aspects of this thesis involve hardware interface design, interaction techniques, models of material responses, and perception. The perceptual investigation provided a validation of the ability of the vibrotactile floor surface to simulate a key aspect of different ground surfaces. In addition, it can be regarded as a worst case assessment of the display paradigm proposed here, since it focused on a perceptual dimension that was, a priori, compromised in the original design. The possibility it demonstrated, of using vibrotactile feedback to evoke an illusory sense of compliance, was an unexpected outcome of this thesis.

1.3 Summary of contributions

This thesis contains the following main contributions:

- The design of a novel, high-fidelity vibrotactile interface in a rigid floor surface. The device was analyzed and structurally optimized to ensure a usable passband extending from about 50 to 750 Hz.
- The design of a distributed, touch-sensitive vibrotactile display, and techniques for using it to interact with multimodal simulations of natural ground terrains and touch surface interfaces.
- The demonstration of a new haptic perceptual-motor illusion, in which vibrotactile feedback was revealed to increase the perceived compliance of a rigid surface in a predictable manner.
- An analysis of pseudo-periodic stick-slip motion in a class of stochastic models of sliding friction. A novel first passage time problem was presented, whose solutions were demonstrated to capture the statistical properties of the motion. Noise-induced irregularities in stick-slip motion are characterized as stochastic resonance phenomena, a connection that is, to the best of our knowledge, established here for the first time.

- Time-domain methods for simulating fracture processes in stiff, heterogeneous materials. They are efficient enough to be used for the interactive rendering of vibrotactile and auditory feedback resulting from stress fluctuations in these materials.
- A demonstration of the feasibility of ambient information display with vibrotactile messages communicated through a floor surface, and usability guidelines for the design of such displays.