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.

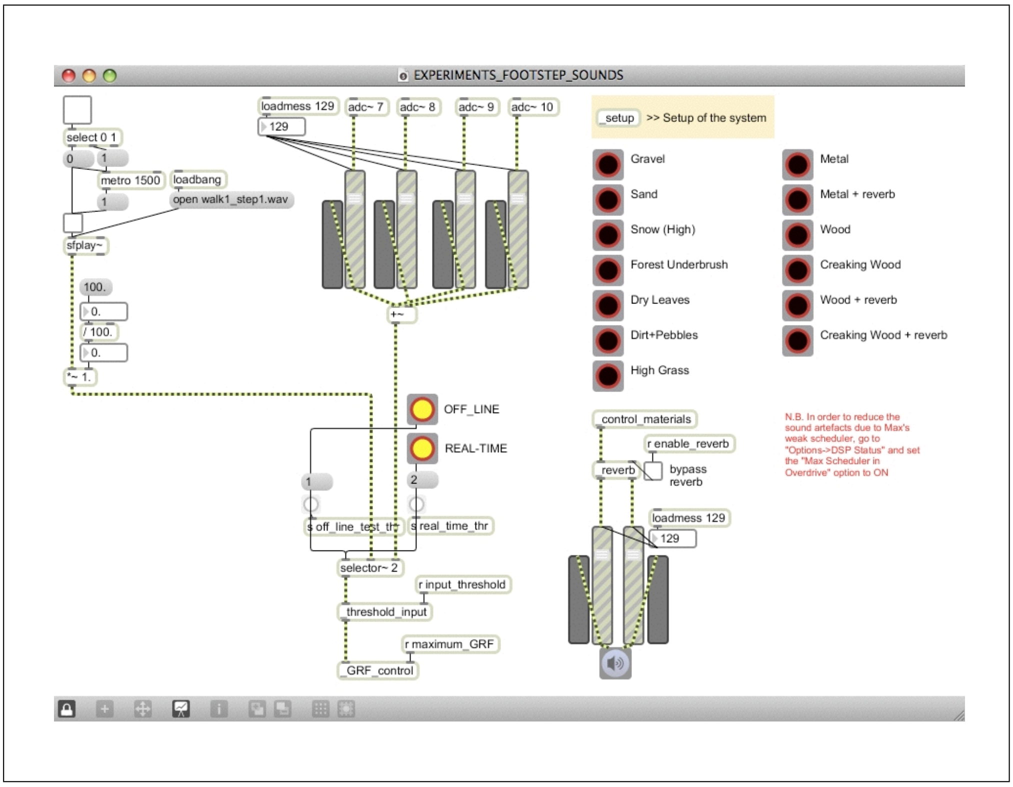

Figure 1: The sound synthesis engine.

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/

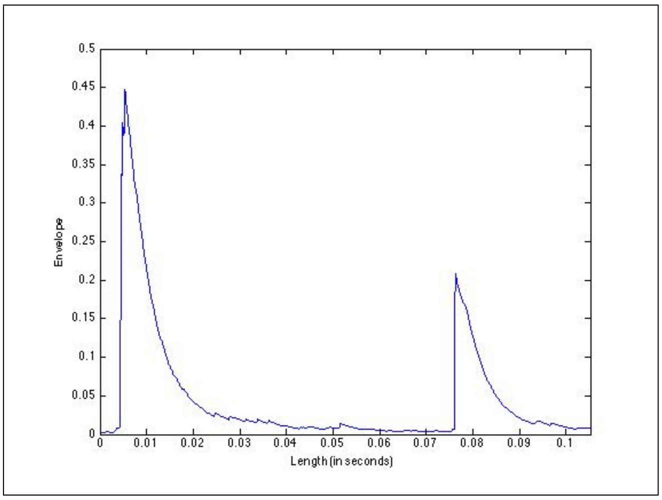

Figure 2: Amplitude envelope (GRF) extracted form a recorded footstep sound on a concrete floor.

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.

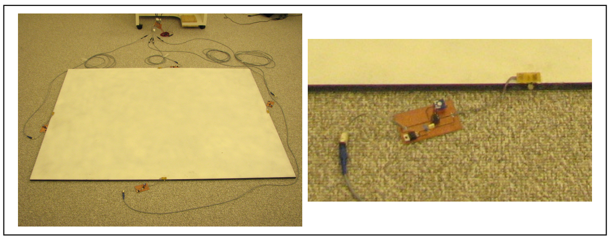

Figure 4: TAI-CHI board with accelerometer sensors (left), closeup of a single accelerometer sensor (mounted with Tack-All) and amplifier (right)

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 signal 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*.