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

Serafin, Stefania; Turchet, Luca; Nordahl, Rolf

Published in:
Proceedings of the 2009 Audio Mostly Conference

Publication date:
2009

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

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.

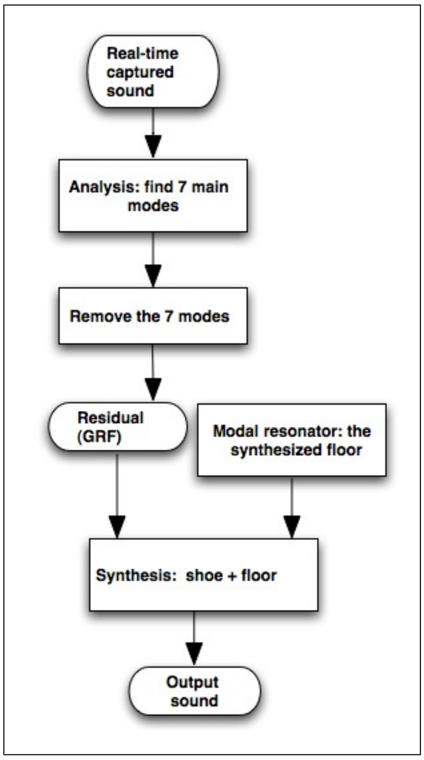

Figure 2: Block diagram of the analysis-synthesis algorithm used.

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 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.

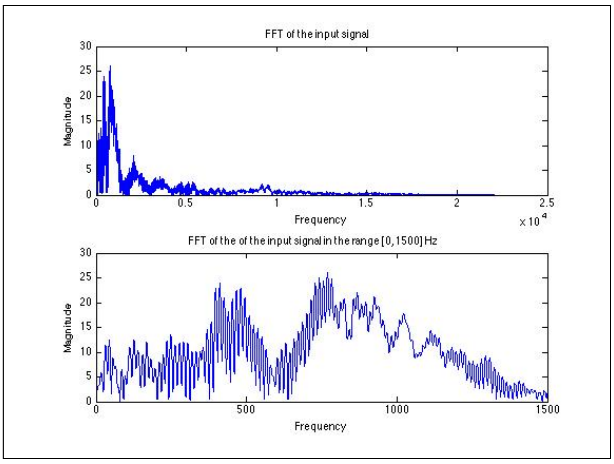

Figure 3: FFT of a sound of footstep on a concrete floor.

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/

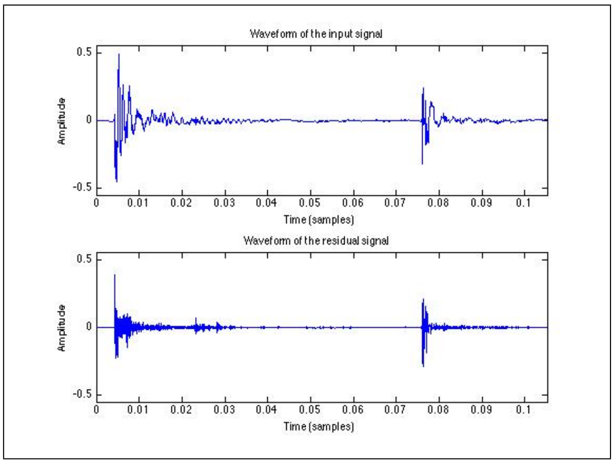

Figure 4: Time domain representation of a recorded footstep sound (top) and the corresponding residual (bottom). Notice how the two main temporal components of the footstep sound, i.e., heel and toe events, can be easily noticeable.

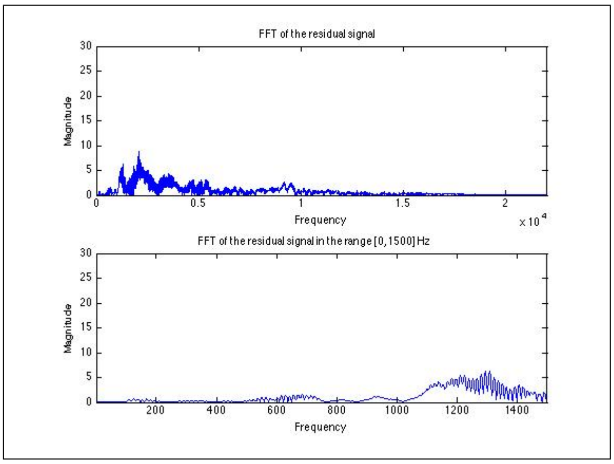

Figure 5: FFT of the residual sound analyzed in Figure 3.

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.

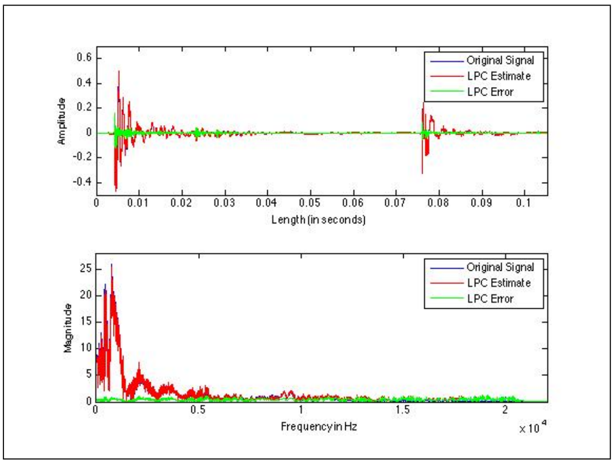

Figure 6: Original footstep sound on a concrete floor, its LPC estimate and the relative error (residual).

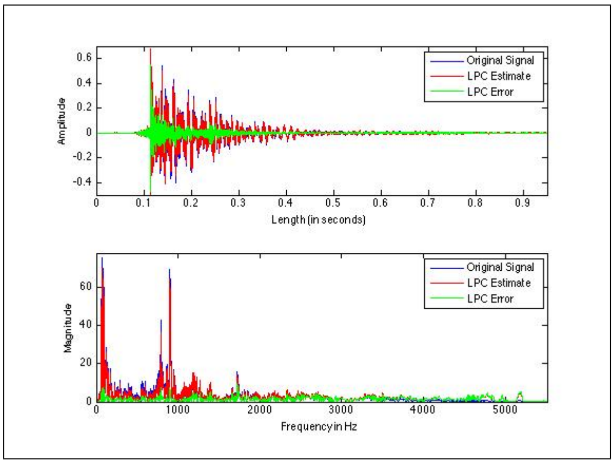

Figure 7: Original footstep sound on a metal floor, its LPC estimate and the relative error (residual).

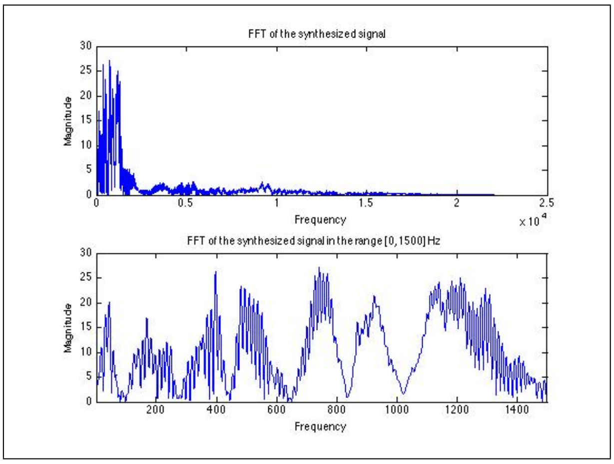

Figure 8: FFT of the synthesized signal.

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.

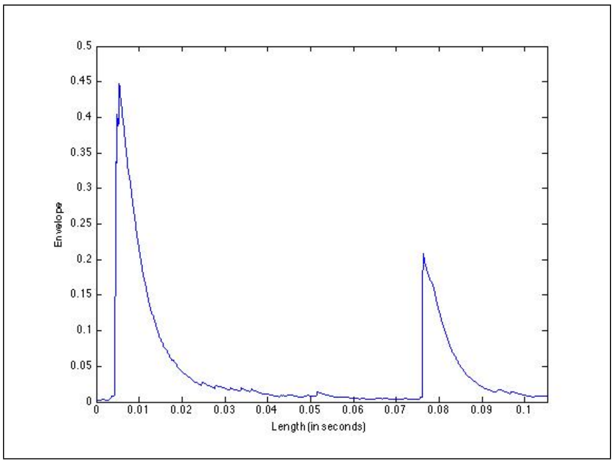

Figure 9: Envelope (GRF) extracted from the sound of Fig. 4.

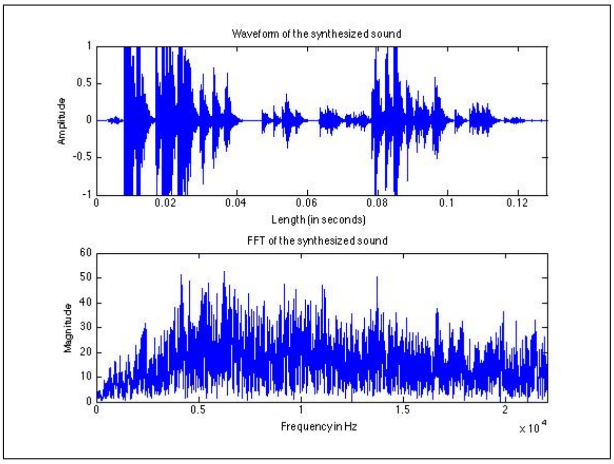

Figure 10: Synthesis of a footstep on little gravel using the GRF to control the PhiSM.

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

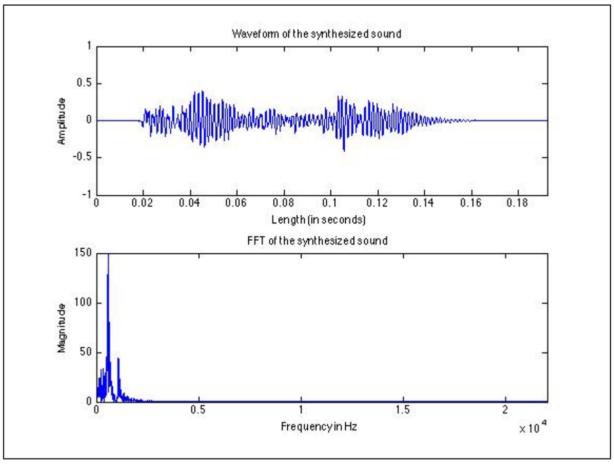

Figure 11: Synthesis of a footstep on wood floor using the GRF to control the algorithm first proposed in [2].

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.