The Natural Interactive Walking Project and Emergence of Its Results in Research on Rhythmic Walking Interaction and the Role of Footsteps in Affecting Body Ownership

Justyna Maculewicz, Erik Sikström and Stefania Serafin

Architecture, Design and Media Technology Department, Aalborg University Copenhagen, 2450, Copenhagen, Denmark
{jma, es, sts}@create.aau.dk
http://media.aau.dk/smc

Abstract. In this chapter we describe how the results of the Natural Interactive Project, which was funded within the 7th Framework Programme and ended in 2011, started several research directions concerning the role of auditory and haptic feedback in footstep simulations. We chose elements of the project which are interesting in a broader context of interactive walking with audio and haptic feedback to present and discuss the developed systems for gait analysis and feedback presentation, but also, what is even more interesting to show how it influence humans behavior and perception. We hope also to open a discussion on why we actually can manipulate our behavior and show the importance of explaining it from the neurological perspective. We start with a general introduction, moving on to more specific parts of the project, that are followed by the results of the research which were conducted after project's termination but based on its results.

1 Introduction

Walking is an activity that plays an important part in our daily lives. In addition to being a natural means of transportation, walking is also characterized by the resulting sound, which can provide rich information about the surrounding and a walker. The study of the human perception of locomotion sounds has addressed several properties of the walking sound source. The sound of footsteps conveys information about walker's gender [1, 2], posture [3], emotions [2], the hardness and size of their shoe sole[2], and the ground material on which they are stepping [4]. It was proven that sounds of footsteps convey both temporal and spatial information about locomotion [5].

Auditory feedback has also strength to change our behavior. Studies show that interactive auditory feedback produced by walkers affects walking pace. In the studies of [6, 7] individuals were provided with footstep sounds simulating different surface materials, interactively generated using a sound synthesis engine [8]. Results show that subjects' walking speed changed as a function of the simulated ground material.

From the clinical perspective, sensory feedback and cueing in walking received an increased attention. It is well known that sensory feedback have a positive effect on gait in patients with the neurological disorders, among which is also Parkinson's disease

SikstrÃűm E., Maculewicz J. and Serafin S.
The Natural Interactive Walking Project and Emergence of Its Results in Research on Rhythmic Walking Interaction and the Role of Footsteps in Affecting Body Ownership.
DOI: 10.5220/0006162100030024
In *European Project Space on Intelligent Systems, Pattern Recognition and Biomedical Systems (EPS Lisbon 2015)*, pages 3-24
ISBN: 978-989-758-095-6

(PD) [9–14]. Rhythmic (metronome-like) auditory cues have been found to produce gait improvement in several studies [9–15]. External rhythms presented by auditory cues may improve gait characteristics [13–15], but also be used to identify deficits in gait adaptability [16].

Research on sensory feedback while walking is also important in the ares of virtual augmented realities.

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. Studies on auditory feedback in VR are focused on sound delivery methods [17, 18], sound quantity and quality of auditory versus visual information [19], 3D sound [20, 21] and enhancement of self-motion and presence in virtual environments [22–24].

Within the study of human perception of walking sounds researchers have focused on topics such as gender identification [1], posture recognition [3], emotional experiences of different types of shoe sound (based on the material of the sole) on different floor types (carpet and ceramic tiles) [25], and walking pace depending on various types of synthesized audio feedback of steps on various ground textures [26].

2 The Objectives of NIW Project

The NIW project contributed to scientific knowledge in two key areas. First it reinforced the understanding of how our feet interact with surfaces on which we walk. Second, informed the design of such interactions, by forging links with recent advances in the haptics of direct manipulation and in locomotion in real-world environments. The created methods have potential to impact a wide range of future applications that have been prominent in recently funded research within Europe and North America. Examples include floor-based navigational aids for airports or railway stations, guidance systems for the visually impaired, augmented reality training systems for search and rescue, interactive entertainment, and physical rehabilitation.

The NIW project proceeded from the hypothesis that walking, by enabling rich interactions with floor surfaces, consistently conveys enactive information that manifests itself predominantly through haptic and auditory cues. Vision was regarded as playing an integrative role linking locomotion to obstacle avoidance, navigation, balance, and the understanding of details occurring at ground level. The ecological information was obtained from interaction with ground surfaces allows us to navigate and orient during everyday tasks in unfamiliar environments, by means of the invariant ecological meaning that we have learned through prior experience with walking tasks.

At the moment of the project execution, research indicated that the human haptic and auditory sensory channels are particularly sensitive to material properties explored during walking [4], and earlier studies have demonstrated strong links between the physical attributes of the relevant sounding objects and the auditory percepts they generate [27]. The project intention was to select, among these attributes, those which evoke most salient perceptual cues in subjects.

Physically based sound synthesis models are capable of representing sustained and transient interactions between objects of different forms and material types, and such methods were used in the NIW project in order to model and synthesize the sonic effects of basic interactions between feet and ground materials, including impacts, friction, or the rolling of loose materials.

The two objectives which guided the project are:

1. The production of a set of foot-floor multimodal interaction methods, for the virtual rendering of ground attributes, whose perceptual saliency has been validated
2. The synthesis of an immersive floor installation displaying a scenario of ground attributes and floor events on which to perform walking tasks, designed in an effort to become of interest in areas such as rehabilitation and entertainment.

The results of the research within the NIW project can be enclosed in the three milestones:

– Design, engineering, and prototyping of floor interaction technologies
– A validated set of ecological foot-based interaction methods, paradigms and prototypes, and designs for interactive scenarios using these paradigms
– Integration and usability testing of floor interaction technologies in immersive scenarios.

The forthcoming sections will focus on research initiated at Aalborg University and continued after the termination of the project.

3 Synthesis of Footsteps Sounds

3.1 Microphone-based Model

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).
Our goal in developing the footsteps sounds synthesis engine was to synthesize a footstep sound on different kinds of materials starting from a signal in the audio domain containing a generic footstep sound on a whatever material. Our approach to achieve this goal consisted of removing the contribution of the resonator, keeping the exciter and considering 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 were summed in order to have a complete footstep sound.

In order to simulate the footsteps sounds on different types of materials, the ground reaction force estimated with this technique was used to control various sound synthesis algorithms based on physical models, simulating both solid and aggregate surfaces [28, 29]. The proposed footsteps synthesizer was implemented in the Max/MSP sound synthesis and multimedia real-time platform.[1]

Below we present an introduction to developed physically based sound synthesis engine that is able to simulate the sounds of walking on different surfaces. This introduction is an excision from [8]. Figure 2 presents the setup that was used for testing designed models of feedback delivery. We developed a physically based sound synthesis engine that is able to simulate the sounds of walking on different surfaces. Acoustic

[1] www.cycling74.com

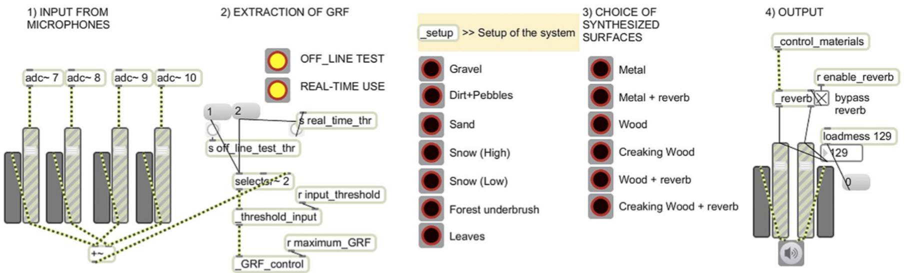

Fig. 1. A screenshot of the graphical user interface for the developed sound synthesis engine. [8].

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 [30]. 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. A comprehensive collection of footstep sounds was implemented. Metal and wood were implemented as solid surfaces. 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. Gravel, sand, snow, forest underbrush, dry leaves, pebbles, and high grass are the materials which were implemented as aggregate sounds. The simulated metal, wood, and creaking wood surfaces were further 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. 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 the walkers 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. The amplitude of the different components

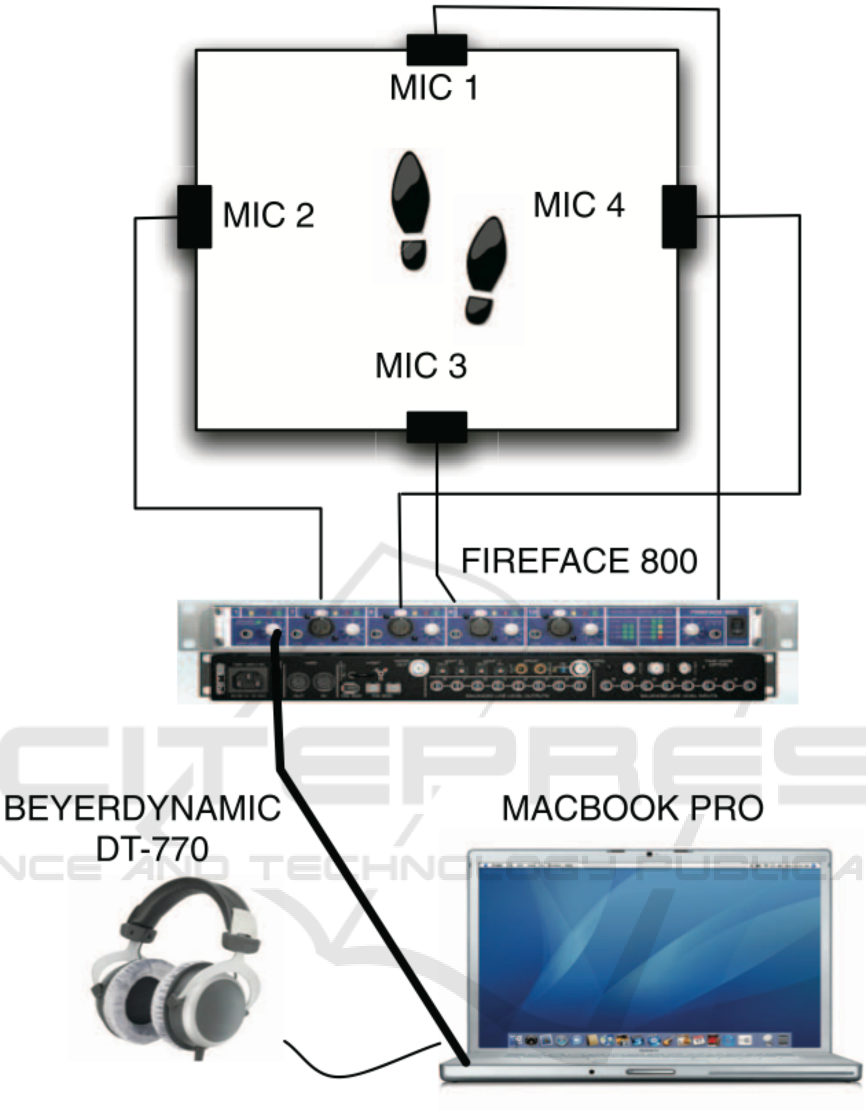

Fig. 2. Hardware components of the developed system: microphones, multichannel soundcard, laptop, and headphones [8].

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.

3.2 The MoCap-based System

In this section we address the problem of calculating the GRF from the data tracked by means of a Motion Capture System (MoCap) in order to provide a real-time control of the footsteps synthesizer. Our goal was to develop a system which could satisfy the requirements of shoe independence, fidelity in the accuracy of the feet movements, and free navigation.

Figure 3 shows a schematic representation of the overall architecture developed. This system is placed in an acoustically isolated laboratory and consists of a MoCap[2], two soundcards[3], sixteen loudspeakers[4], and two computers. The first computer runs the motion capture software[5], while the second runs the audio synthesis engine. The two computers are connected through an ethernet cable and communicate by means of the UDP protocol. The data relative to the MoCap are sent from the first to the second computer which processes them in order to control the sound engine.
The MoCap is composed by 16 infrared cameras[6] which are placed in a configuration optimized for the tracking of the feet. In order to achieve this goal, two sets of markers are placed on each shoe worn by the subjects, in correspondence to the heel and to the toe respectively.
Concerning the auditory feedback, the sounds are delivered through a set of sixteen loudspeakers or through headphones.

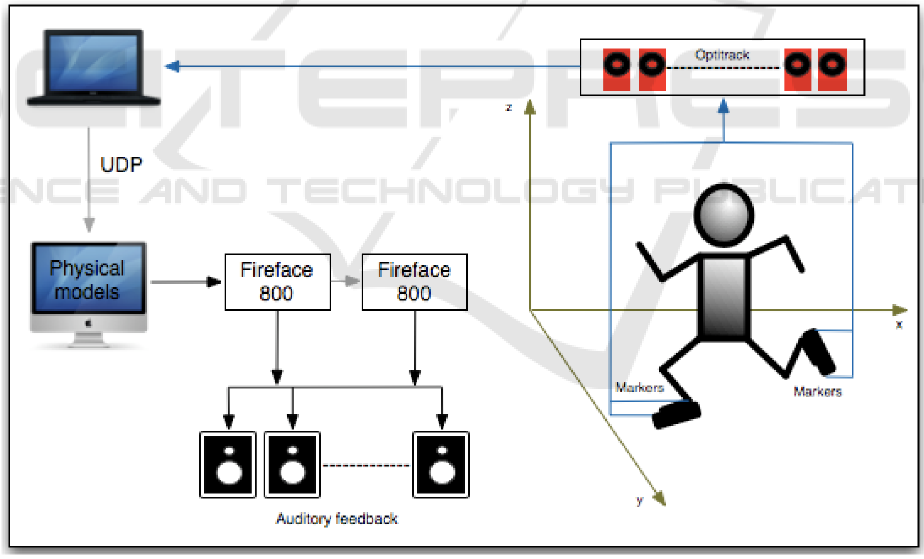

Fig. 3. A block diagram of the developed system and the used reference coordinates system.

[2]Optitrack: http://naturalpoint.com/optitrack/
[3]FireFace 800 soundcard: http://www.rme-audio.com
[4]Dynaudio BM5A: http://www.dynaudioacoustics.com
[5]Tracking Tools 2.0
[6]OptiTrack FLEX:V100R2

3.3 Wireless Shoes

In this section we address the problem of calculating the GRF from the data tracked by sensors placed directly on the walker shoes, with the goal of providing a real-time control of the footsteps synthesizer using in addition a wireless trasmission.
The setup for the developed shoe-integrated sensors system is illustrated in Figure 4. Such system is composed by a laptop, a wireless data acquisition system (DAQ), and a pair of sandals each of which is equipped with two force sensing resistors[7] and two 3-axes accelerometers[8]. More in detail the two FSR sensors were placed under the insole in correspondence to the heel and toe respectively. Their aim was to detect the pressure force of the feet during the locomotion of the walker. The two accelerometers instead were fixed inside the shoes. Two cavities were made in the thickness of the sole to accommodate them in correspondence to the heel and toe respectively. In order to better fix the accelerometers to the shoes the two cavities containing them were filled with glue.
The analog sensor values were transmitted to the laptop by means of a portable and wearable DAQ.

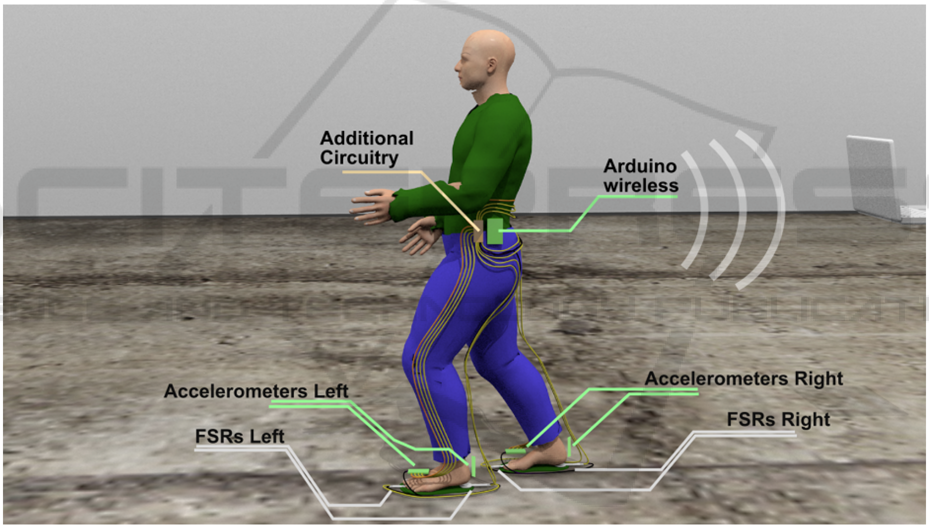

Fig. 4. Setup for the wireless shoes system: the user wears the sensor enhanced shoes and the wireless data acquisition system.

The wireless DAQ consists of three boards: an Arduino MEGA 2560 board[9], a custom analog preamplification board, and a Watterott RedFly[10] wireless shield. In the nomenclature of the Arduino community, a “shield” is a printed circuit board (PCB) that

[7]FSR: I.E.E. SS-U-N-S-00039
[8]ADXL325: http://www.analog.com
[9]http://arduino.cc
[10]http://www.watterott.net/projects/redfly-shield

matches the layout of the I/O pins on a given Arduino board, allowing that a shield can be "stacked onto" the Arduino board, with stable mechanical and electrical connections.

All three boards are stacked together. In this way the wireless DAQ system can be easily put together in a single box, to which a battery can be attached. This results in a standalone, portable device that can be attached to the user's clothes, allowing greater freedom of movement for the user.

Since each foot carries two FSRs and two 3-axis accelerometers, which together provide 8 analog channels of data, the system demands capability to process 16 analog inputs in total. That is precisely the number of analog inputs offered by an Arduino MEGA 2560, whose role is to sample the input channels, and format and pass the data to the wireless RedFly shield. The analog preamplification board is a collection of four quad rail-to-rail operational amplifier chips (STmicroelectronics TS924), providing 16 voltage followers for input buffering of the 16 analog signals, as well as of four trimmers, to complete the voltage divider of the FSR sensors; and connectors. The Watterott RedFly shield is based on a Redpine Signals RS9110-N-11-22 WLAN interface chipset, and communicates with the Arduino through serial (UART) at 230400 baud. Preliminary measurements show that the entire wireless DAQ stack consumes about 200 mA with a 9V power supply, therefore we chose a power supply of 9V as the battery format.

3.4 Discussion

At software level, both the proposed systems are based on the triggering of the GRFs, and this solution was adopted rather than creating the signal in real-time. Indeed one could think to generate a signal approximating the shapes of the GRFs.

For instance the contributions of the heel strike, (as well as the one of the toe), could be approximated by means of a signal created by a rapid exponential function with a certain maximum peak followed by a negative exponential function with a certain decay time. Nevertheless it was not possible to create such signal in real time for two reasons. First the synthesis engine needs to be controlled by a signal in the auditory domain having a sample rate of 44100 Hz, but the data coming from both the systems arrive with a frequency much lower, 1000 Hz. Therefore mapping the data coming from the tracking devices to an auditory signal will result in a step function not usable for the purposes of creating a proper GRF to control the synthesis engine. Secondly the computational load to perform this operation added to the one of the algorithms for the sound synthesis would be too high, with the consequent decrease of the system performances mostly in terms of latency.

Hereinafter we discuss the advantages and disadvantages of the two developed systems in terms of portability, easiness of setup, wearability, navigation, sensing capabilities, sound quality, and integration in VR environments.

Portability. The MoCap based system is not portable as it requires to carry all the components of the architecture discussed in section 3.2. Conversely the wireless shoe system is easily portable.

Easiness of Setup. At hardware level, while the MoCap based system is not easy to setup since it consists of many components, instead the wireless shoe system does not show any difficulty in the set up process. Both the systems at software level require an initial phase in which global parameters and thresholds of the proposed techniques have to be calibrated, but such calibration is however simple and quick.

Wearability. The MoCap based system allows users to wear their own footwear, and in addition no wires linked to the user are involved. The only technology required to be worn by the users consists of the four sets of markers which have to be attached to the shoes by means scotch tape. However, they are very light and therefore their presence is not noticeable, and in addition they do not constitute an obstacle for the user walk. Conversely the wireless shoe system is not shoe-independent since the users are required to wear the developed sandals. In addition walkers need to carry the box containing the Arduino board which is attached at the trousers, but its presence is not noticeable.

Navigation. In both the systems the user is free to navigate as no wires are involved. However the walking area in the case of the MoCap based system is delimited by the coverage angle of the infrared cameras, while the wireless shoes can be used in a wider area, also outdoor.

Sensing Capabilities. As regards the MoCap based system the functioning of the proposed technique is strictly dependent on the quality of the MoCap system utilized. The requirements for the optimal real-time work of the proposed method are a low latency and a good level of accuracy. The latency problem is the most relevant since the delivery of the sound to the user must be perfectly synchronized with the movements of his/her feet in order to result into a credible closed-loop interaction. For a realistic rendering the latency should be at maximum 15 milliseconds, therefore the current latency of 40 milliseconds is too much for the practical use of the proposed method. However this limit could be lowered by improving the MoCap technology since the latency is due in most part to it. Concerning the accuracy, a high precision MoCap system allows a better tracking of the user gestures which have to be mapped to GRFs for the subsequent generation of the sounds. In general, the overall computational load of this technique is high but if it is divided into two computers, one for the data acquisition and the other for the sound generation, it results acceptable.
As concerns the wireless shoes the total latency is acceptable for the rendering of a realistic interaction and the use of the accelerometers allows to achieve a mapping between the feet movements and the dynamics in footstep sounds similar to the one obtainable by using the microphones system described in [31].

Sound Quality. The sound quality of the system depends on the quality of the sound synthesis algorithms, on the sensing capabilities of the tracking devices, as well as on 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 [32, 33]. In addition, an highly accurate MoCap system, as well as good FSRs and accelerometers, allow to detect the footsteps dynamics with high precision therefore enhancing the degree of realism of the interaction. Concerning the audio delivery method, both headphones and loudspeakers can be used.

Integration in VR Environments. Both the system have been developed at software level as extension to the Max/MSP platform, which can be easily combined with several interfaces and different software packages. Both the system allow the simultaneous coexistence of interactively generated footsteps sounds and of soundscapes pro-

vided by means of the surround sound system.
The architecture of the MoCap based system can be integrated with visual feedback using for example a head mounted display to simulate different multimodal environments. However using the MoCap based system it is not possible to provide the haptic feedback by means of the haptic shoes developed in previous research [34] since such shoes are not involved and their use will result in a non wireless and not shoe-independent system.

It is possible to extend the wireless shoes system embedding some actuators in the sandals in order to provide the haptic feedback. For this purpose another wireless device receiving the haptic signals to must be also involved. Nevertheless, the latency for the round-trip wireless communication would be much higher.

4 The AAU Research after the NIW Project

4.1 Multi Sensory Research on Rhythmic Walking

Due to the successful implementation of physical models into design of ecological feedback and interest in rhythmic walking interaction with auditory and haptic feedback, we continued exploring these areas of research. We are specifically interested in the influence of auditory and haptic ecological feedback and cues on rhythmic walking stability, perceived naturalness of feedback and synchronization ease with presented cues. From our hitherto research emerged several effects, which are interesting in a context of gait rehabilitation, exercise and entertainment. Until now we have been testing gravel and wood sound as ecological feedback and a tone as a non-ecological sound. Results show that when we ask people to walk in their preferred pace, they have the slowest pace with gravel feedback, then wood and a tone motivates to the fastest walking [35]. To test this effect even further we added soundscape sounds which are congruent and incongruent with the sounds of the footsteps. The preliminary analysis shows that feedback sounds can manipulate participants pace even more than footsteps sounds alone [36]. When people are asked to synchronize with above-mentioned rhythmic sounds their results are similar with a slight worse performance with gravel cues [7]. Even though this feedback produces the highest synchronizing error it is perceived as the one, which is the easiest to follow [35]. In the same study [35] we also investigated the influence of feedback in the haptic modality, in rhythmic walking stimulation. We have seen that haptic stimulation is not efficient in rhythmic cueing, bur it might help to improve the naturalness of the walking experience (Fig. 5, 6). In order to understand these results. we turn to neurological data in search of an explanation. The results of our preliminary exploratory encephalographic (EEG) experiment suggest that synchronization with non-ecological sounds requires more attention and in synchronization with ecological sounds is involved a social component of synchronizing with another person. Synchronizing with the pace, which is similar to the natural walking, also requires less attention. The analysis of the EEG data is ongoing [37].

Many different ways of feedback delivery to the user were presented. We can see that different types of auditory feedback can be crucial in all the aspects mentioned before. Many behavioral effects were observed while presenting feedback through mentioned applications. We believe that there is a need now to understand why feedback or

Fig. 5. Experimental setup used in the studies following the NIW project. A microphone was placed below a stepper to detect person's steps. An actuator was located under the top layer of a stepper [7].

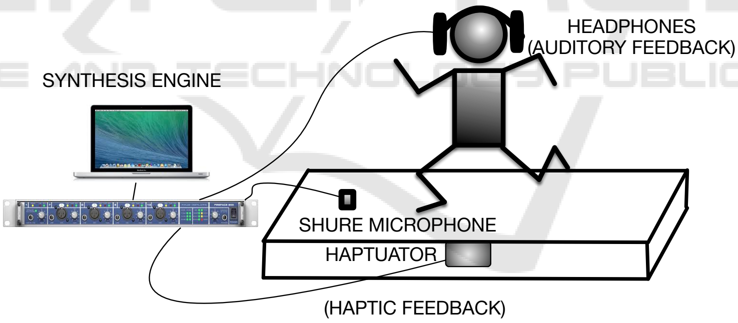

Fig. 6. A visualization of the setup used in the studies following the NIW project. It clarifies the placement of the actuator. We can see that feedback was presented through headphones connected to synthesis engine describe in the previous section via Fireface 800 sound card [35].

cues can manipulate our behavior. Deeper understanding is needed to gain basic knowledge about the neural bases of altered behavior, which will help to build more efficient and precise feedback systems and also to design feedback signal in a way they could be the most efficient for specific need.

4.2 Footstep Sounds and Virtual Body Awareness

An experiment was designed and conducted with the aim of investigating if and how the user's perception of the weight of a first person avatar could be manipulated by the means of altering the auditory feedback. Inspired by the approach used by Li et al. [1], using very basic audio filter configurations, a similar methodology was adopted. Instead of using one floor type, two floors with different acoustic properties were used in the experiment.

Experiment. In order to investigate whether it would be possible to manipulate a user's perception of how heavy their avatar could be by manipulating the audio feedback generated from interactive walking in place input, an experiment was set up involving an immersive virtual reality setup with full body motion tracking and real-time audio feedback. The environment featured a neutral starting area with a green floor and two corridors equipped with different floor materials from which the walking sounds would be evaluated; concrete tiles or wooden planks. The area with the green floor were used for training and transitions. In the experiment, subjects were asked to perform six walks (explained further under the experiment procedure sub-section) while orally giving estimates of the perceived weight of the virtual avatar, as well as rating the perceived suitability of the sounds as sound effects for the ground types that they just walked over. For each of the walks, the audio feedback from the interactive footsteps was manipulated with audio filters that came in three different configurations (A, B and C, explained further in the audio implementation section). Thus, the six walks were divided into three walks on concrete featuring filters A, B and C (one at a time), while three walks were on the wooden floor, also featuring filters A, B and C. As a part of the experiment the virtual avatar body was also available in two sizes (see Avatar and movement controls). The hypothesis was formulated so that the filters would bias the subjects into making different weight estimate depending on the filter type. It was also expected that the filters that had lower center frequencies would be estimated as being heavier.

Implementation. The virtual environment was implemented in Unity 3D[11] while audio feedback was implemented in Pure Data[12]. Network communication between the two platforms were managed using UDP protocol.

Avatar and Movement Controls. The avatar used for the experiment consisted of a full male body. The body was animated using inverse kinematics animation and data acquired from a motion capture system. The avatar also had a walking in place ability that allowed the subjects to generate forward translation from the user performing stepping in place movements. The model of the avatar body was also available in two versions where one (a copy of the original model) had been modified to have an upper body with a larger body mass with thicker arms and a torso with a bigger belly and chest (see Fig. 7). A calibration procedure also allowed the avatar to be scaled to fit a subject's own height and length of limbs.

[11] http://www.unity3d.com
[12] http://puredata.info

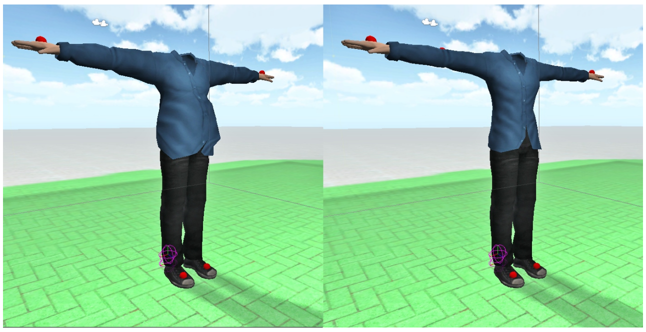

Fig. 7. The big and the small avatar bodies.

Audio Implementation. A skilled sound designer would probably utilize a combination of DSP effects and editing techniques, when asked to make a certain footstep sequence sound heavier or lighter. We have for the sake of this experiment chosen to resort to only using one method for changing the sounds for the sake of experimental control. The audio feedback was made up out of two sets of Foley recordings of footsteps, one for the concrete surface sounds and one for the wooden surface sounds. The recordings were made in an anechoic chamber using a Neumann u87ai microphone[13] at a close distance and an RME Fireface 800 audio interface[14]. Each set contained seven variations of steps on either a concrete tile or wooden planks (a wooden pallet). The playback was triggered by the walking in place script in the Unity implementation and cues were sent via UDP to the Pure Data patch that would play the sounds. The footstep sounds were played back at a random order, including the avoiding of consecutive repetitions of individual samples. Three different filters were applied to the recordings using the Equalization plug-in in Audacity[15] with the graphic EQ setting with the B-spline interpolation method selected. These filters were rendered into copies of the concrete and wood sets of the Foley recordings, generating a total of 6x7 audio files. The filters were set up in peak-shapes (similarly to Li et al. [1]), amplifying center frequencies of either 80Hz (hereafter labelled "filter"), 160Hz (labelled "filter B"), or 315Hz (labelled "filter C") with +20 dB, as shown on the slider of the interface. The two adjacent sliders were set to be amplifying their respective frequency areas with +10 dB (see Fig. 8). According to the frequency response curve presented in the plug-in, the actual amplification of the center frequency was lower than +20 dB and a bit more spread out (for an overview of all three filters, see Fig. 9). Finally all of the audio files were normalized to the same output level. The footstep sounds were then presented to the subjects through headphones (see section Equipment and Facilities), at approximately 65 dB (measured from the headphone using an AZ instruments AZ8922 digital sound level meter). Which files would be played was determined by a ground surface detection script in Unity that identifies the texture that the avatar is positioned above. The filters would be selected manually using keyboard input.

[13]https://www.neumann.com
[14]http://www.rme-audio.de
[15]http://audacity.sourceforge.net/

For the green areas in the virtual environment, a neutral type of footstep sound was used consisting out of a short “blip” sound, a sinus tone following an exponential attack and decay envelope (the ead∼ object in Pure Data) with a 5 ms attack and 40 ms decay.

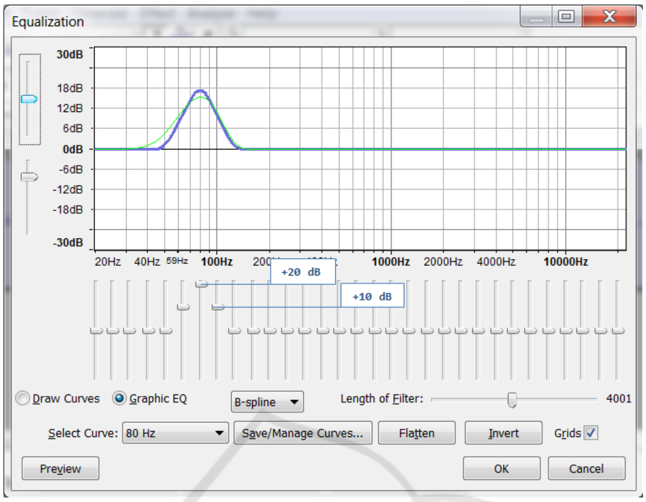

Fig. 8. Filter settings for filter A.

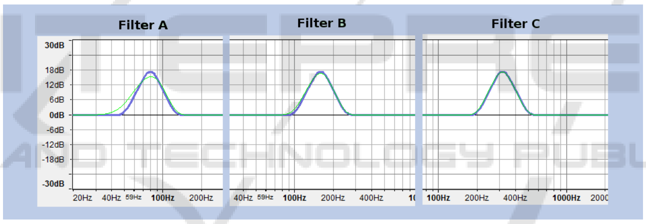

Fig. 9. Overview of the filters.

Equipment and Facilities. The virtual environment, audio and motion tracking software were running on one Windows 7 PC computer (Intel i7-4470K 3.5GHz CPU, 16 GB RAM and an Nvidia GTX 780 graphics card). The head mounted display was a nVisor SX with a FOV of 60 degrees with a screen resolution 1280x1024 pixel in each eye. The audio were delivered through an RME Fireface 800 with a pair of Sennheiser HD570 headphones. The motion-tracking was done with a Naturalpoint optitrack motion-tracking system with 11 cameras of the model V100:R2 and with 10 3-point trackables attached on the subjects feet, knees, hip, elbows hands and on the head mounted display.

Discussion. Changing the size of the avatar did not make the subjects provide significantly different evaluations in the weight estimates or suitability ratings. This could possibly have to do with the subject did not pay much attention to the appearance of their virtual bodies, even though they were given an opportunity to see it during the calibration process and the training session. It may also have to do with technical limitations of the implementation and the hardware such as the head mounted display's rather narrow field of view.

In between the weight estimates of the filters, the results indicate a small effect that at least partially supports the experimental hypothesis. The effect is not entirely consistent, in that filter A for the concrete surface was given lighter estimates than filter B in the big body group and in the combined groups analysis. Similarly for the wood surface context, the estimates for filter B were not significantly different from filter A or C in any variant of the analysis and in the big body group there were no significant differences among the estimates for any of the filters. Perhaps further investigations with a greater number of filter configurations, including a greater variety of filter characteristics, could give more detailed information regarding weight estimates in this kind of context.

The suitability ratings were among the filters only significantly different in one case, where in the combined groups filter A given a significantly higher rating than filter C. This could be interpreted that filter C had a negative impact of the audio quality of the footstep samples, as sound effects for the concrete tiles. This information could be useful and should be taken into account for those who plan to use automated Foley sounds in virtual reality applications. Interestingly, the suitability ratings for the wooden floor were very similar for all the three filter settings, suggesting a higher tolerance associated with this material for the type of manipulations used here, than the concrete surface material.

There are also a couple of issues of technical nature since the implementation is still at a prototype stage (and likely not representable of the capacity of some current or future commercial systems) that could have had an impact on the subjects' performances during the experiment. The triggering of the footstep sounds were suffering from a latency that was noticeable during fast walking and occasionally footstep sounds would be triggered more than once per step. There were also an issue in that while the subjects were walking in one of the directions, many times the footsteps would only trigger for one of the two feet. When this happened, the subjects were asked to turn around and walk in the opposite direction in order to get sounds from the steps from both feet and to get the time they needed with a representative and working feedback. These limitations could have made the virtual reality experience less believable and it would be interesting to see if a system of higher quality would yield a different experience in terms of weight estimates and suitability judgements.

Since the experiment used a walking in place-type of method for generating translation in the virtual environment and for triggering the audio feedback, the resulting interaction cannot be considered entirely similar to real walking. This may of course also have had an impact on the subjects' experiences, but it was necessary to employ this type of technique since the motion tracking area in the facilities was quite small.

The audio methods for manipulating the feedback could be more elaborate and advanced (as mentioned in the audio implementation section) and further investigations should consider involving other approaches such as pitch shifting and layering of several sound effect components (also including creaking, cracking and crunching sounds that may belong to the ground surface types).

4.3 Footsteps Sounds and Presentation Formats

When evaluating a sound design for virtual environment, the context where it is to be implemented might have an influence on how it may be perceived. We performed an experiment comparing three presentation formats (audio only, video with audio and an interactive immersive VR format) and their influences on a sound design evaluation task concerning footstep sounds. The evaluation involved estimating the perceived weight of a virtual avatar seen from a first person perspective, as well as the suitability of the sound effect relative to the context.

We investigated the possible influence of the presentation format, here as in immersive VR and two less (technologically) immersive, seated and less interactive desktop formats, on the evaluations of a footstep sound effect with a specific purpose. This purpose is to provide auditory feedback for a walking in place interaction in VR and to describe the avatar's weight. Previously we have seen that the perceived weight of a VR avatar [38], as well as of the actual user [39], might be manipulated by changing the sonic feedback from the walking interactions.

Our aim is to support the process of sound design evaluations, by bringing understanding to what role the context of the presentation may have. As previous research provide hints [40–50], that presentation formats may potentially bring about a slightly different experience of the sound being evaluated. More information on this topic would be important for sound designers when developing audio content for VR simulations and for researchers doing experiments on audio feedback in the these contexts.

Experiment. As immersive VR is potentially able to influence subjective judgements of audio quality, we here perform a study complementing a previous study conducted in a full immersive virtual reality (IVR) setting (full body avatar, gesture controlled locomotion and audio feedback) with two less immersive presentation formats. The formats were:

- An immersive VR condition (VR), using a head mounted display (HMD), motion tracking of feet, knees, hip, hands, arms and head with gesture controlled audio feedback triggered by walking in place input (which is also used for locomotion)
- An audiovisual condition (Video), produced utilizing a motion capture and screen recordings from the VR setup, presented in full screen mode on a laptop screen
- An audio only condition (Audio), using only the audio track from the above mentioned screen recording

The VR condition was also fully interactive, requiring the user to use their whole body to control the avatar and navigate inside the environment, while the Video and Audio conditions were passive with the subject sitting down at a desk. For each of the walks, the audio feedback from the interactive footsteps were manipulated with audio filters in three different peak shaped configurations with center frequencies at either 80 Hz (filter A), 160 Hz (filter B) or 315 Hz (filter C). The filters were all applied to both of the two floor materials. Using 9 point Likert scales, the subjects were asked to estimate the weight of the avatar and the suitability of the footstep sounds as sound effects for the two ground materials. Both a concrete and a wooden surface were evaluated by each subjects, but in one presentation format only per subject.

Hypothesis:

- H_1: The degree of immersiveness and interaction in the presentation formats will have an influence on the participants ratings of the weight estimates
- H_2: The degree of immersiveness and interaction in the presentation formats will have an influence on the participants ratings of the suitability estimates
- H_3: The same patterns observed in the VR presentation format in regards to weight estimates will also be observed in the Audio and Video conditions

In the VR condition 26 subjects participated, while 19 subjects participated in each of the two other conditions. In the VR condition the test population had a mean age of 25 (M = 25.36, SD = 5.2) and of which 7 were female. The test population of the Audio and Video groups, consisted of University staff and students of which 10 were females, had an average age of 29 years (M = 29.12, SD = 7.86). One of these reported a slight hearing impairment, and one reported having tinnitus.

Implementation. The virtual environment was implemented using Unity3D[16] while the audio feedback was implemented using Pure Data[17]. For the Video condition, a video player was implemented using vvvv[18]. Network communication between the different software platforms were managed using UDP protocol.

All software in the VR implementation were running in a Windows 7 PC computer (Intel i7-4479K 3.5GHz CPU, 16 GB ram, with an Nvidia GTX780 graphics card). In the Video and Audio conditions, all software were running on a Dell ProBook 6460b laptop (Intel i5-2410M 2.3 GHz, 4 GB ram, with Intel HD Graphics 3000), also with Windows 7.

The audio feedback implementation used for the VR condition in this experiment was identical to the one presented in the previously described study (see 4.2 Footstep Sounds and Virtual Body Awareness - implementation).

The video was captured using FRAPS with 60 FPS frame rate (although the frame rate of the IVR implementation was not the same as this) and the same screen resolution as used in the head mounted display (1280*1024 pixels per eye) but with one camera (the right) deactivated (since stereoscopic presentation was not used in this experiment), resulting in a resolution 1280*1024 pixels. The audio was captured at stereo 44.1 kHz, 1411 kbps. In order to make the video files smaller (the original size was over 6 Gb per file) further compression was applied using VLC[19], but same resolution with a data rate of 6000 kbps at 30 frames/second and with the audio quality remaining the same as in the original video capture. The interface used for the test was developed using Pure

[16]www.unity3d.com
[17]puredata.info
[18]vvvv.org
[19]www.videolan.org/vlc/

Data and vvvv. The Pure Data patch contained the graphical user interface presented on an external monitor (Lenovo L200pwD), the data logging functions and networking components for communicating with the vvvv patch. The vvvv patch held the functions necessary for streaming the video files from disk with audio and presenting on a 14 inch laptop monitor (Dell Probook 6460b) in fullscreen mode. The audio was presented through a pair of Sennheiser HD570 headphones connected to an RME Fireface 800 audio interface at the same approximate 65 dB level as in the VR condition.

Experiment Procedure. The participants partaking in the VR condition were the same as in the in the experiment described in section 4.2 Footstep Sounds and Virtual Body Awareness.

For the Video and Audio conditions, the experiment procedure was shorter than the VR condition. The participants were seated in by a desk in a quiet office space and received written instructions and a graphical user interface consisting of buttons (for triggering the stimuli) and sliders (for the likert scales). Once the participants had read through the instructions they were allowed to begin the evaluations, listening and watching (depending on the condition) to the pre-recorded walk as many times as they needed and giving the weight and suitability ratings, before continuing on to the next. The question had the same formulation as in the VR experiment. They were not allowed to go back and change their ratings, or deviate from the presentation order which was randomized before each session.

Discussion. The degree of support for our first hypothesis is not very strong. There were three out of 18 comparisons with significant differences for both weight estimates and suitability ratings. In concrete there were the three significant cases and for the footsteps on the wooden material there were no differences at all between the different presentation formats for weight estimates and the suitability ratings. Despite this, the Audio condition rendered the most significant differences between the filters, as if the effect of the different filters were more pronounced in that context. That might seem obvious as the Audio only condition offers the least amount of distraction, but it might also have been that the Video and the VR conditions were not distractive enough. An action packed computer game presented in the VR condition might have yielded a different result.

The suitability for two of the filters (b and c) were also given significantly higher ratings in the VR condition than in Audio once (ConcB) and in Video twice (ConcB and ConcC). This may hint that the VR presentation format may make users more tolerant to poor sound effects.

As various interactive tasks may have different demands on the attentional resources of the user. In order to learn more about how demanding the employed tasks are, some kind of instrument should be applied, such as the NASA task load index (NASA-TLX) [51] for measuring efforts required. This information could then be taken into account when studying results from audio quality judgements provided by users of VR simulations with varying levels of interaction complexity (such as wandering around and exploring, performing simple tasks or complicated tasks that must be completed within a limited amount of time).

What we also can see from the analysis is that the effect of the audio filters follow similar patterns in all presentation formats, with lighter estimates for higher center frequencies. This is especially pronounced for the wood material and less prominent for the concrete.

The results hint that a presentation format with only audio would yield a more pronounced effect in a comparison of different sound designs than in a presentation accompanied by visual feedback or as in an immersive VR format. These findings are somewhat coherent with previous research that suggests that tasks such as computer games may change the user's experience of a sound design or even reduce the user's ability to detect impairments in sound quality.

5 Conclusions

In this chapter we introduced several research directions related to walking, specifically on simulating audio and haptic sensation of walking, simulating walking in virtual reality and walking as a rhythmic activity. We presented ways of audio and haptic feedback generation in a form of footsteps natural and unnatural sounds. It was shown how different types of feedback can influence our behavior and perception. Our plans for the near future is to broaden the explanation of how can we actually can manipulate these from the neurological perspective to build more efficient, precise, and goal-directed feedback systems.

References

1. Li, X., Logan, R.J., Pastore, R.E.: Perception of acoustic source characteristics: Walking sounds. The Journal of the Acoustical Society of America 90 (1991) 3036–3049
2. Giordano, B., Bresin, R.: Walking and playing: Whats the origin of emotional expressiveness in music. In: Proc. Int. Conf. Music Perception and Cognition. (2006)
3. Pastore, R.E., Flint, J.D., Gaston, J.R., Solomon, M.J.: Auditory event perception: The sourceperception loop for posture in human gait. Perception & psychophysics 70 (2008) 13–29
4. Giordano, B. L., McAdams, S., Visell, Y., Cooperstock, J., Yao, H. Y., Hayward, V.: Non-visual identification of walking grounds. The Journal of the Acoustical Society of America 123 (2008) 3412–3412
5. Young, W., Rodger, M., Craig, C.M.: Perceiving and reenacting spatiotemporal characteristics of walking sounds. Journal of Experimental Psychology: Human Perception and Performance 39 (2013) 464
6. Turchet, L., Serafin, S., Cesari, P.: Walking pace affected by interactive sounds simulating stepping on different terrains. ACM Transactions on Applied Perception (TAP) 10 (2013) 23
7. Maculewicz, J., Jylha, A., Serafin, S., Erkut, C.: The effects of ecological auditory feedback on rhythmic walking interaction. MultiMedia, IEEE 22 (2015) 24–31
8. Nordahl, R., Turchet, L., Serafin, S.: Sound synthesis and evaluation of interactive footsteps and environmental sounds rendering for virtual reality applications. Visualization and Computer Graphics, IEEE Transactions on 17 (2011) 1234–1244
9. De Dreu, M., Van Der Wilk, A., Poppe, E., Kwakkel, G., Van Wegen, E.: Rehabilitation, exercise therapy and music in patients with parkinson's disease: a meta-analysis of the effects of music-based movement therapy on walking ability, balance and quality of life. Parkinsonism & related disorders 18 (2012) S114–S119

10. Thaut, M.H., Abiru, M.: Rhythmic auditory stimulation in rehabilitation of movement disorders: a review of current research. (2010)
11. McIntosh, G.C., Brown, S.H., Rice, R.R., Thaut, M.H.: Rhythmic auditory-motor facilitation of gait patterns in patients with parkinson's disease. Journal of Neurology, Neurosurgery & Psychiatry 62 (1997) 22–26
12. Suteerawattananon, M., Morris, G., Etnyre, B., Jankovic, J., Protas, E.: Effects of visual and auditory cues on gait in individuals with parkinson's disease. Journal of the neurological sciences 219 (2004) 63–69
13. Roerdink, M., Lamoth, C.J., Kwakkel, G., Van Wieringen, P.C., Beek, P.J.: Gait coordination after stroke: benefits of acoustically paced treadmill walking. Physical Therapy 87 (2007) 1009–1022
14. Nieuwboer, A., Kwakkel, G., Rochester, L., Jones, D., van Wegen, E., Willems, A.M., Chavret, F., Hetherington, V., Baker, K., Lim, I.: Cueing training in the home improves gait-related mobility in parkinsons disease: the rescue trial. Journal of Neurology, Neurosurgery & Psychiatry 78 (2007) 134–140
15. Thaut, M., Leins, A., Rice, R., Argstatter, H., Kenyon, G., McIntosh, G., Bolay, H., Fetter, M.: Rhythmic auditor y stimulation improves gait more than ndt/bobath training in near-ambulatory patients early poststroke: a single-blind, randomized trial. Neurorehabilitation and neural repair 21 (2007) 455–459
16. Bank, P.J., Roerdink, M., Peper, C.: Comparing the efficacy of metronome beeps and stepping stones to adjust gait: steps to follow! Experimental brain research 209 (2011) 159–169
17. Storms, R., Zyda, M.: Interactions in perceived quality of auditory-visual displays. Presence 9 (2000) 557–580
18. Sanders Jr, R.D.: The effect of sound delivery methods on a users sense of presence in a virtual environment. PhD thesis, Naval Postgraduate School (2002)
19. Chueng, P., Marsden, P.: Designing auditory spaces to support sense of place: the role of expectation. In: CSCW Workshop: The Role of Place in Shaping Virtual Community. Citeseer, Citeseer (2002)
20. Freeman, J., Lessiter, J.: Hear there & everywhere: the effects of multi-channel audio on presence. In: Proceedings of ICAD. (2001) 231–234
21. Västfjäll, D.: The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments. CyberPsychology & Behavior 6 (2003) 181–188
22. Larsson, P., Västfjäll, D., Kleiner, M.: Perception of self-motion and presence in auditory virtual environments. In: Proceedings of Seventh Annual Workshop Presence 2004. (2004) 252–258
23. Kapralos, B., Zikovitz, D., Jenkin, M.R., Harris, L.R.: Auditory cues in the perception of self motion. In: Audio Engineering Society Convention 116, Audio Engineering Society (2004)
24. Väljamäe, A., Larsson, P., Västfjäll, D., Kleiner, M.: Travelling without moving: Auditory scene cues for translational self-motion. In: Proceedings of ICAD05. (2005)
25. Tonetto, P.L.M., Klanovicz, C.P., Spence, P.C.: Modifying action sounds influences peoples emotional responses and bodily sensations. i-Perception 5 (2014) 153–163
26. Bresin, R., de Witt, A., Papetti, S., Civolani, M., Fontana, F.: Expressive sonification of footstep sounds. In: Proceedings of ISon 2010: 3rd Interactive Sonification Workshop. (2010) 51–54
27. Rocchesso, D., Bresin, R., Fernstrom, M.: Sounding objects. MultiMedia, IEEE 10 (2003) 42–52
28. Avanzini, F., Rocchesso, D.: Modeling collision sounds: Non-linear contact force. In: Proc. COST-G6 Conf. Digital Audio Effects (DAFx-01). (2001) 61–66
29. Cook, P.: Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. Computer Music Journal 21 (1997) 38–49

30. Gaver, W.W.: What in the world do we hear?: An ecological approach to auditory event perception. Ecological psychology 5 (1993) 1–29
31. 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) 161–168
32. Nordahl, R., Serafin, S., Turchet, L.: Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In: Proceedings of the IEEE Virtual Reality Conference. (2010) 147–153
33. Turchet, L., Serafin, S., Nordahl, R.: Examining the role of context in the recognition of walking sounds. In: Proceedings of Sound and Music Computing Conference. (2010)
34. Turchet, L., Nordahl, R., Berrezag, A., Dimitrov, S., Hayward, V., Serafin, S.: Audio-haptic physically based simulation of walking on different grounds. In: Proceedings of IEEE International Workshop on Multimedia Signal Processing, IEEE Press (2010) 269–273
35. Maculewicz, J., Cumhur, E., Serafin, S.: An investigation on the impact of auditory and haptic feedback on rhythmic walking interactions. International Journal of Human-Computer Studies 2015submitted.
36. Maculewicz, J., Cumhur, E., Serafin, S.: The influence of soundscapes and footsteps sounds in affecting preferred walking pace. International Conference on Auditory Display 2015submitted.
37. Maculewicz, J., Nowik, A., Serafin, S., Lise, K., Króliczak, G.: The effects of ecological auditory cueing on rhythmic walking interaction: Eeg study. (2015)
38. Sikström, E., de Götzen, A., Serafin, S.: Self-characteristics and sound in immersive virtual reality - estimating avatar weight from footstep sounds. In: Virtual Reality (VR), 2015 IEEE, IEEE (2015)
39. Tajadura-Jiménez, A., Basia, M., Deroy, O., Fairhurst, M., Marquardt, N., Bianchi-Berthouze, N.: As light as your footsteps: altering walking sounds to change perceived body weight, emotional state and gait. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, ACM (2015) 2943–2952
40. Zielinski, S.K., Rumsey, F., Bech, S., De Bruyn, B., Kassier, R.: Computer games and multichannel audio quality-the effect of division of attention between auditory and visual modalities. In: Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality, Audio Engineering Society (2003)
41. Kassier, R., Zielinski, S.K., Rumsey, F.: Computer games and multichannel audio quality part 2'evaluation of time-variant audio degradations under divided and undivided attention. In: Audio Engineering Society Convention 115, Audio Engineering Society (2003)
42. Reiter, U., Weitzel, M.: Influence of interaction on perceived quality in audiovisual applications: evaluation of cross-modal influence. In: Proc. 13th International Conference on Auditory Displays (ICAD), Montreal, Canada. (2007)
43. Rumsey, F., Ward, P., Zielinski, S.K.: Can playing a computer game affect perception of audio-visual synchrony? In: Audio Engineering Society Convention 117, Audio Engineering Society (2004)
44. Reiter, U., Weitzel, M.: Influence of interaction on perceived quality in audio visual applications: subjective assessment with n-back working memory task, ii. In: Audio Engineering Society Convention 122, Audio Engineering Society (2007)
45. Reiter, U.: Toward a salience model for interactive audiovisual applications of moderate complexity. (audio mostly) 101
46. Larsson, P., Vastfjall, D., Kleiner, M.: Ecological acoustics and the multi-modal perception of rooms: real and unreal experiences of auditory-visual virtual environments. (2001)
47. Larsson, P., Västfjäll, D., Kleiner, M.: Auditory-visual interaction in real and virtual rooms. In: Proceedings of the Forum Acusticum, 3rd EAA European Congress on Acoustics, Sevilla, Spain. (2002)

48. Larsson, P., Västfjäll, D., Kleiner, M.: The actor-observer effect in virtual reality presentations. CyberPsychology & Behavior 4 (2001) 239–246
49. Harrison, W.J., Thompson, M.B., Sanderson, P.M.: Multisensory integration with a head-mounted display: background visual motion and sound motion. Human Factors: The Journal of the Human Factors and Ergonomics Society (2010)
50. Thompson, M.B., Sanderson, P.M.: Multisensory integration with a head-mounted display: Sound delivery and self-motion. Human Factors: The Journal of the Human Factors and Ergonomics Society 50 (2008) 789–800
51. Hart, S.G., Staveland, L.E.: Development of nasa-tlx (task load index): Results of empirical and theoretical research. Advances in psychology 52 (1988) 139–183