A PRELIMINARY STUDY ON SOUND DELIVERY METHODS FOR FOOTSTEP SOUNDS

Luca Turchet, *

Medialogy Section
Department of Architecture, Design
and Media Technology
Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, DK
tur@create.aau.dk

Stefania Serafin,

Medialogy Section
Department of Architecture, Design
and Media Technology
Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, DK
sts@create.aau.dk

ABSTRACT

In this paper, we describe a sound delivery method for footstep sounds, investigating whether subjects prefer static rendering versus dynamic. In this case, dynamic means that the sound delivery method simulates footsteps following the subject. An experiment was run in order to assess subjects' preferences regarding the sound delivery methods. Results show that static rendering is not significantly preferred to dynamic rendering, but subjects disliked rendering where footstep sounds followed a trajectory different from the one they were walking along.

1. INTRODUCTION

Procedural sound synthesis is becoming a successful approach to simulate interactive sounds in virtual environments and computer games [1, 2]. One important category of sounds produced by action of subjects navigating in an environment is the sound of footsteps.

Recently, several algorithms have been proposed to simulate walking sounds. One of the pioneers in this field is Perry Cook, who proposed a collection of physically informed stochastic models (PhiSM) simulating several everyday sonic events [3]. Among such algorithms the sounds of people walking on different surfaces were simulated [4]. A similar algorithm was also proposed in [5], where physically informed models simulate several stochastic surfaces.

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

The research just described does not take into consideration the ability of footstep sounds to be rendered in a 3D space. Sound rendering for virtual environments has reached a level of sophistication that it is possible to render in realtime most of the phenomena which appear in the real world [7].

In this study, we are interested in investigating how subjects react to different kinds of sound rendering algorithms, which follow the user or propose different confusing trajectories.

The results presented in this paper are part of the Natural Interactive Walking (NIW) FET-Open project[1], whose goal is to provide closed-loop interaction paradigms enabling the transfer of skills that have been previously learned in everyday tasks associated to walking. In the NIW project, several walking scenarios are simulated in a multimodal context, where especially audition and haptic play an important role.

* This work was supported by the NIW project

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

2. SYSTEM ARCHITECTURE

The system we adopted for the experiments consists of a a motion capture system (MoCap)[2], a soundcard[3], eight loudspeakers, two sandals with pressure sensors embedded in, and two computers. Figure 1 shows a schematic representation of the overall architecture developed.

Such system was placed in an acoustically isolated laboratory which consisted of a control room and a bigger room where the setup was installed and where the experiments were performed. The control room was 5.45 m large, 2 m long, and 2.85 m high, and it was used by the experimenters providing the stimuli and collecting the experimental results. It hosted two desktop computers.

The first computer run the motion capture software, while the second run the footstep sounds synthesis engine (see section 2.1). The two computers were connected through an ethernet cable and communicate by means of the UDP protocol. The data relative to the motion capture system were sent from the first to the second computer which processed them in order to control the sound engine.

The experiment room was 5.45 m large, 5.55 m long, and 2.85 m high. A transparent glass divided the two rooms, so it was possible for the experimenters to see the users performing the assigned task. The two rooms were connected by means of a talkback system.

The user locomotion was tracked by an Optitrack motion capture system[4], composed by 16 infrared cameras[5]. The cameras were placed in a configuration optimized for the tracking of the head position. In order to achieve this goal, markers were placed on the top of the head using a bicycle helmet. The walking area available to the users for the purposes of the experiments consisted of a rectangle 2.5 x 2.6 m which corresponded to the area fully seen by the infrared cameras. The perimeter of such rectangle was indicated on the floor by means of scotch tape strips (see figure 2)

Users were also tracked by using the pressure sensors embedded in a pair of shoes. Specifically, a pair of light-weight san-

[2] from Naturalpoint with software Tracking Tools 2.0

[3] FireFace 800 soundcard: http://www.rme-audio.de/en_products_fireface_800.php

[4] http://naturalpoint.com/optitrack/

[5] OptiTrack FLEX:V100R2

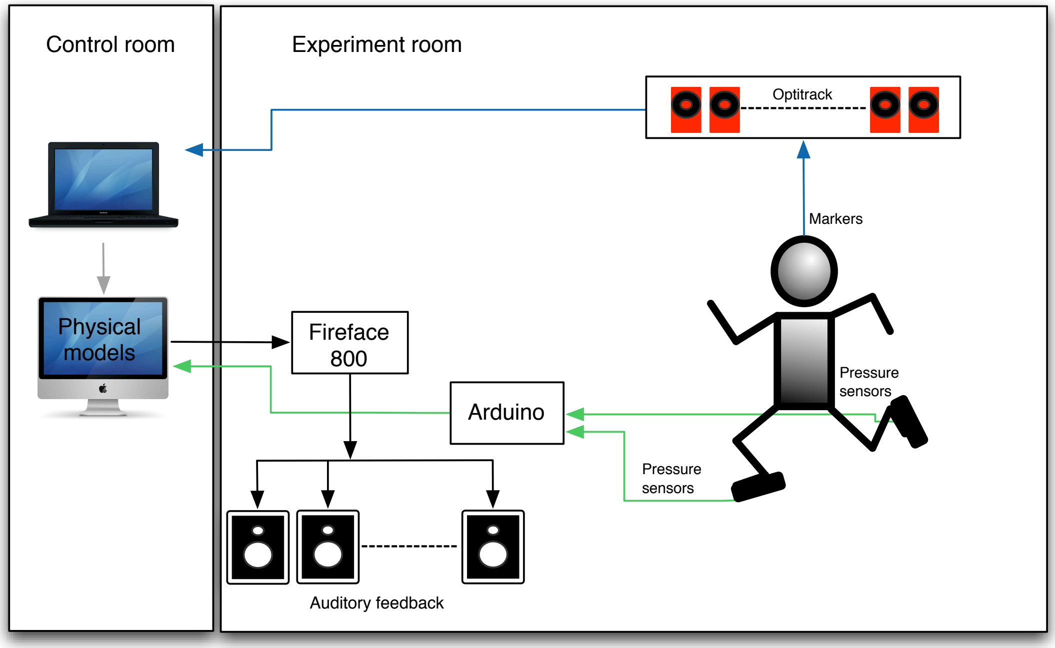

Figure 1: *A block diagram of the architecture.*

dals was used (Model Arpenaz-50, Decathlon, Villeneuve d'Ascq, France). The sole had two FSR pressure sensors[6] whose aim was to detect the pressure force of the feet during the locomotion of a subject wearing the shoes. The two sensors were placed in correspondence to the heel and toe respectively in each shoe. The analogue values of each of these sensors were digitalized by means of an Arduino Diecimila board[7] and were used to drive the audio synthesis.

The configuration of the eight loudspeakers is illustrated in Figure 2. In detail, the loudspeakers[8] were placed on the ground at the vertices and at the middle point of the sides of the rectangular floor. During the experiments the loudspeakers were hidden from view using acoustically transparent curtains.

2.1. Footstep sounds synthesis engine

In previous research, we proposed a sound synthesis engine able to simulate footstep sounds on aggregate and solid surfaces [8]. Such engine is based on physical models which are driven by a signal, in the audio domain, expressing the ground reaction force (GRF), i.e., the reaction force supplied by the ground at every step. In our simulations the GRF corresponds to the amplitude envelope extracted from an audio signal containing a footstep sound.

The engine can operate both offline and in real-time. The two approaches differ in the way the input GRF is generated. Concerning the realtime work, various systems for the generation of such input have been developed and tested [8, 9, 10]. In the proposed experiments, the footstep sounds synthesis is driven interactively during the locomotion of the subject wearing the shoes. The description of the control algorithms based on the analysis of the values of the pressure sensors coming from the shoes can be found in [11]).

The sound synthesis algorithms were implemented in C++ as external libraries for the Max/MSP[9] sound synthesis and multimedia real-time platform.

2.2. Sound delivery methods

We implemented and tested two different types of approaches for the delivery of the footstep sounds through the loudspeakers: static and dynamic diffusion.

For static diffusion we intend that the footstep sound, generated interactively during the locomotion of the user wearing the shoes, is diffused simultaneously to the eight loudspeakers, and with the same amplitude in each loudspeaker.

Conversely, during the dynamic diffusion the user position was tracked by the MoCap and it was used to diffuse the footsteps sound according to a sound diffusion algorithm based on ambisonics. Specifically, to achieve the dynamism we used the ambisonic tools for Max/MSP[10] which allow to move virtual sound sources along trajectories defined on a tridimensional space [12]. Such algorithm was set in order to place under the user feet the virtual sound source containing the footstep sounds. In this way the sound followed the user trajectories during his/her locomotion, and therefore the eight loudspeakers delivered the footstep sounds with different amplitudes. As an example, in reference to figure 2, when the user position was near the loudspeakers 1 and 2, the effect resulting from the dynamic diffusion was that the sound was mostly delivered through these two loudspeakers while the loudspeakers 5 and 6, placed on the opposite sides, did not deliver any sound.

[6] I.E.E. SS-U-N-S-00039
[7] http://arduino.cc/
[8] Dynaudio BM5A speakers: http://www.dynaudioacoustics.com/
[9] http://cycling74.com/
[10] Available at http://www.icst.net/research/projects/ambisonics-tools/

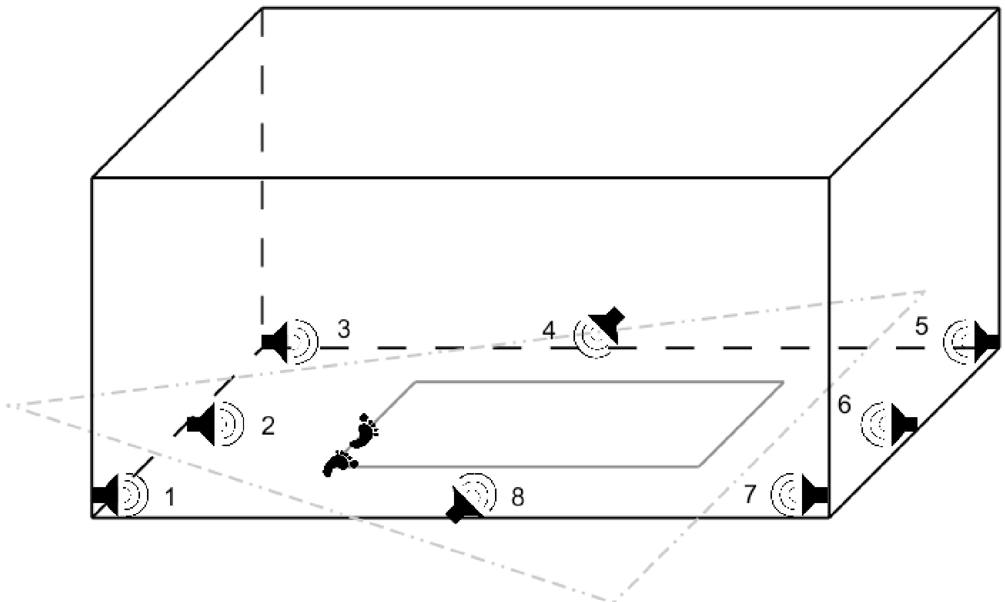

Figure 2: *Loudspeakers configuration used in this study. In dark grey the the perimeter of the rectangle delimiting the walking area (indicated on the floor by means of scotch tape strips). In light grey the triangular trajectory for the delivery of the distractors during the dynamic diffusion condition in experiment 1.*

In addition to the static and dynamic diffusion, during the experiments we also delivered the sounds using a second type of dynamic diffusion where the sound delivery was incoherent with the user position. In detail, in this configuration we chose to deliver the sound as coming from the side opposite to the user position. As an example, in reference to figure 2, when the user position was near the loudspeakers 1 and 2, the effect resulting from the dynamic diffusion was that the sound was mostly delivered through loudspeakers 5 and 6, while the loudspeakers 1 and 2 did not deliver any sound.

Finally, in presence of the dynamic diffusion, the delay due to the MoCap was negligible for the purposes of the experiments.

3. DESCRIPTION OF THE EXPERIMENTS

We performed two experiments in order to assess subjects' reaction to the sound delivery methods. During the experiments participants were asked to wear the sandals, and the cycling helmet mentioned in section 2 and walk in the laboratory according to the tasks of the two experiments.

3.1. Experiment 1

The task of the first experiment consisted on walking in circular way along the perimeter of the walking area (i.e. the rectangle indicated on the floor by means of scotch tape strips). During their walk they produced interactively footstep sounds which were delivered through the loudspeakers according to the following six conditions:

- static diffusion
- coherent dynamic diffusion
- incoherent dynamic diffusion
- static diffusion plus static distractors
- coherent dynamic diffusion plus dynamic distractors
- incoherent dynamic diffusion plus dynamic distractors

The three methods explained in section 2.2 were presented with and without distractors. Such distractors consisted of footstep sounds of a virtual person walking in the same room. Specifically, in presence of static diffusion the distractors were delivered statically (i.e., with the same volume in all the loudspeakers), while in the dynamic diffusion condition they were delivered dynamically following a triangular trajectory (see figure 2).

Participants were exposed to twelve trials, where the six conditions were presented twice in randomized order. Each trial lasted one minute.

The sound engine was set in order to synthesize footstep sounds on two different kinds of materials: wood and forest underbrush. Each condition was presented with both wood and forest underbrush. The reason for choosing two materials was to assess whether the surface type affected the quality of the results. In this particular situation, a solid and an aggregate surface were chosen.

The distractors were presented using the same surface chosen for the participants' walks. In order to keep the distinction between distractors and participants' footstep sounds simple, distractors were presented with a lower volume, a small change in the timber, and with a moderately quick gait.

After the presentation of each stimulus participants were required to evaluate on a seven-point Likert scale the following questions:

- How well could you localize under your feet the footstep sounds you produced?
- How well did the sounds of your footsteps follow your position in the room?

- How much did your walk in the virtual environment seem consistent with your walk in the real world?
- How natural did your interaction with the environment seem?
- To what degree did you feel confused or disoriented while walking?

The goal of this experiment was to compare the proposed diffusion methods. The incoherent dynamic diffusion was included in the experiment in order to assess if any difference in the participants evaluations between coherent dynamic and static diffusion, was not due only to the fact that the source was moving, but that it was moving coherently with the user position. In order to make a consistent comparison, the sound synthesis engine was set with appropriate volumes for footstep sounds delivered through the static and the dynamic diffusion, i.e. there was no big volume difference between the two sound delivery methods. The distractors were used to assess if the same differences in the participants evaluations of the three diffusion methods were found both in presence and in absence of distractors.

Our hypotheses were that the coherent dynamic condition would have got better results rather than the others (in particular the static one), that the incoherent dynamic condition would have been evaluated as the worst, and that the use of distractors would have worsen the participants evaluations in comparison with the case in which the distractors were not presented.

3.2. Experiment 2

Starting from the results of the first experiment we designed a second experiment in order to investigate in a deeper way the users' perception of the static and coherent dynamic diffusions. The task of such experiment consisted on walking freely inside the walking area. Participants were exposed to fourteen trials, where seven surface materials were presented in randomized order according to the two delivery methods (static and coherent dynamic diffusion). The seven surface materials, five aggregate and two solid, were gravel, sand, snow, dry leaves, forest underbrush, wood and metal. Each trial lasted one minute. After the presentation of each stimulus participants were required to evaluate on a seven-point Likert scale the same questions presented in the first experiment.

The goal of this experiment was to assess whether participants showed a preference for one of the two proposed methods when exploring the virtual environment by walking freely (and a without predefined trajectory like in experiment 1). Furthermore we were interested in assessing whether the type of used surface could affect the quality of the results.

3.3. Results of experiment 1

The first experiment was performed by thirteen subjects, 10 males and 3 females, aged between 21 and 38 (mean=24, standard deviation=4.51), who took on average about 17 minutes to complete it. Results are illustrated in figure 3. Various ANOVA (with and without repeated measures) were performed in order to assess if the differences found in the results were significative. All post-hoc analyses were performed using the least significant difference (LSD) test with Bonferroni's correction.

The first noticeable thing is the difference between results of the two surface materials, wood and forest underbrush, for what concerns the dynamic coherent condition and its comparison with the static condition, in the case of absence of distractors. Indeed while for the wood material the differences between such two conditions are negligible for all the investigated parameters, instead for the forest underbrush material the differences are noticeable and significant (p-value = 0.000064). In presence of distractors such behavior is not hold neither for wood nor for forest, and all the differences are not significative.

A trend common to the two materials is that always the dynamic incoherent condition, both in presence and in absence of distractors, gave rise to lower evaluations in terms of localization, following, consistency and naturalness, and higher evaluations for what concerns the disorientation. In detail, for both materials, significance has been found concerning the differences between the dynamic coherent and dynamic incoherent conditions (for wood: p-value = 0.000285 and p-value = 0.005136, for forest: p-value $<$ 0.000001 and p-value = 0.043566, for the cases with and without distractors respectively), and between the static and dynamic incoherent conditions (for wood: p-value = 0.004283 and p-value = 0.002424, for forest: p-value $<$ 0.000001 both for the cases with and without distractors).

For both materials, as regards the parameters localization, following, consistency and naturalness, almost always the evaluations in absence of distractors are higher than when the distractors are present (the opposite behavior coherently happens for what concerns the disorientation parameter). This is more evident for the forest underbrush material, and indeed the difference between these two conditions is significant (p-value = 0.03817), while for wood is not.

In particular for both materials the disorientation is higher in presence of distractors rather than in absence, but significant difference between these two conditions was found only for the forest underbrush material (p-value = 0.01923). The condition dynamic incoherent with distractors was evaluated as the most disorienting in both materials; conversely, for the forest underbrush material only, the dynamic coherent condition was evaluated as the less disorienting.

As previously said, at global level the dynamic coherent condition gave rise to significant better results than the static one for what concerns the forest material in absence of distractors. In addition a successive analysis for each of the investigated parameters revealed significant difference between the two conditions only for the naturalness parameter (p-value = 0.013238).

3.4. Results of experiment 2

The second experiment was performed by ten subjects, 8 males and 2 females, aged between 19 and 37 (mean=28.8, standard deviation=5.63), who took on average about 17 minutes to complete it.

Results are illustrated in figure 4. As it is possible to notice, participants did not show any preference for one of the two methods. Evaluations of the investigated items of the questionnaire were very similar between the two methods for all the surfaces (all the differences are not statistically significant).

However it is possible to notice that the participant answers to the questionnaire items were not always similar for each surface material. In particular it is possible to observe that the metal surface on average produced the lower scores for the localization, naturalness and consistency items, and the higher scores for the disorientation item.

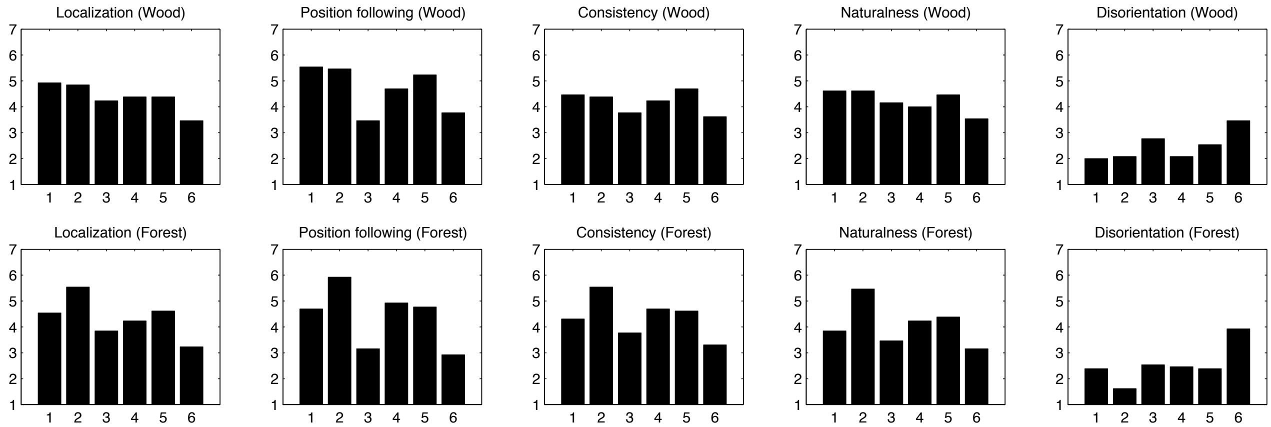

Figure 3: *Results of the first experiment. Conditions are indicated on x-axis: 1- static diffusion, 2- coherent dynamic diffusion, 3- incoherent dynamic diffusion, 4- static diffusion plus static distractors, 5- coherent dynamic diffusion plus dynamic distractors, 6- incoherent dynamic diffusion plus dynamic distractors.*

4. GENERAL DISCUSSION

The first noticeable element emerging from the results of the first experiment is the different behavior found for the coherent dynamic condition in the two simulated materials, in the case of absence of distractors. For forest underbrush such condition seems to play an important role since it got the best evaluations among all the investigated parameters, while for wood it got evaluations very similar to those of the static condition. So our hypothesis that people preferred the dynamic coherent condition to the static one was only partially confirmed.

Instead the hypothesis that the dynamic incoherent condition would have given the worst evaluations was confirmed, and found statistically significant for both materials and both in presence and absence of distractors.

It is therefore possible to conclude that users can perceive very well that their interaction with the virtual environment is not realistic nor natural when the source is not moving coherently with their position. This is an indication of the success of our simulations. The hypothesis concerning the distractors was confirmed: for both materials, almost always the evaluations in absence of distractors are better than when the distractors are present, although significant differences were found only for forest. In addition, the evaluations of the disorientation parameter were higher in presence of distractors (but significant only for forest). This indicates that the use of distractors, i.e., walking sounds evoking the presence of another person walking in the same room as the subject, is likely to influence the perception of footstep sounds associated with the subject.

Concerning the second experiment, results were clear: participants' evaluations did not differ for the two proposed methods, and this is an indication that the two methods could both be used in a virtual environment to deliver interactively generated footsteps sounds. However, other tests should be conducted in order to confirm this and assess more in detail other eventual differences in the perception of the two methods.

5. CONCLUSION AND FUTURE WORK

In this paper we have described an experiment whose goal was to assess the importance of surround sound rendering in simulating footstep sounds for virtual environments. Results show that static delivery method is not significantly preferred to the (coherent) dynamic one, and that participants disliked the renderings where footstep sounds followed a trajectory different from the one they were walking along.

In future experiments we will investigate in a deeper way the differences between the static and dynamic diffusion methods, as well as other parameters related to sound rendering, such as the role of reverberation and the role of amplitude.

We also plan to integrate the proposed footstep sounds renderings in an audio-haptic-visual environment, to design and evaluate different multimodal experiences based on walking.

6. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme under FET-Open grant agreement 222107 NIW - Natural Interactive Walking.[11]

7. REFERENCES

[1] P.R. Cook, *Real sound synthesis for interactive applications*, AK Peters, Ltd., 2002.

[2] A. Farnell, *Designing Sound*, MIT Press, 2010.

[3] P.R. Cook, "Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds," *Computer Music Journal*, vol. 21, no. 3, pp. 38–49, 1997.

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

[11] Natural Interactive Walking Project: www.niwproject.eu

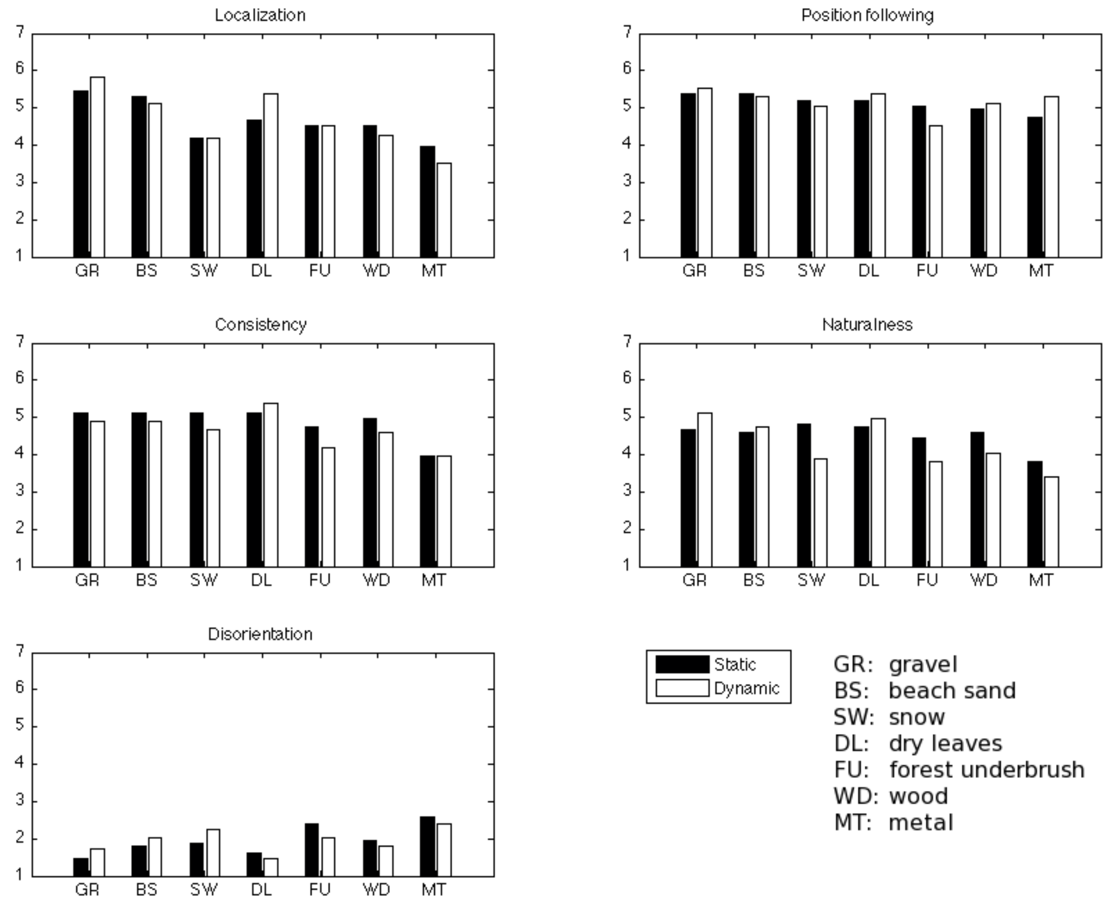

Figure 4: *Results of the second experiment.*

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

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

[7] T. Funkhouser, N. Tsingos, and J.M. Jot, "Survey of methods for modeling sound propagation in interactive virtual environment systems," 2003.

[8] L. Turchet, S. Serafin, S. Dimitrov, and R. Nordahl, "Physically based sound synthesis and control of footsteps sounds," in *Proceedings of Digital Audio Effects Conference*, 2010.

[9] R. Nordahl, S. Serafin, and L. Turchet, "Sound synthesis and evaluation of interactive footsteps for virtual reality applications," in *Proc. IEEE VR 2010*, 2010.

[10] L. Turchet, R. Nordahl, and S. Serafin, "Examining the role of context in the recognition of walking sounds.," in *Proc. of Sound and Music Computing Conference*, 2010.

[11] L. Turchet, R. Nordahl, A. Berrezag, S. Dimitrov, V. Hayward, and S. Serafin, "Audio-haptic physically based simulation of walking sounds.," in *Proc. of IEEE International Workshop on Multimedia Signal Processing*, 2010.

[12] J.C. Schacher and M. Neukom, "Ambisonics Spatialization Tools for Max/MSP," in *Proceedings of the International Computer Music Conference*, 2006.