Examining the role of context on the recognition of walking sounds

Turchet, Luca; Nordahl, Rolf; Serafin, Stefania

Published in:
Proceedings of Sound and Music Computing Conference

Publication date:
2010

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

Examining the role of context in the recognition of walking sounds

Luca Turchet
Medialogy Department
Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, DK
tur@media.aau.dk

Rolf Nordahl
Medialogy Department
Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, DK
rn@media.aau.dk

Stefania Serafin
Medialogy Department
Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, DK
sts@media.aau.dk

ABSTRACT

In this paper, we present an experiment whose goal was to recognize the role of contextual information in the recognition of environmental sounds.

Forty three subjects participated to a between-subjects experiment where they were asked to walk on a limited area in a laboratory, while the illusion of walking on different surfaces was simulated, with and without an accompanying soundscape. Results show that, in some conditions, adding a soundscape significantly improves surfaces' recognition.

1. INTRODUCTION

When exploring a place by walking, at least two categories of sounds can be identified: the persons own footsteps and the surrounding soundscape. In the movie industry, footsteps sounds represent important elements. Chion writes of footstep sounds as being rich in what he refers to as *materializing sound indices* – those features that can lend concreteness and materiality to what is on-screen, or contrarily, make it seem abstracted and unreal [1]. Studies on soundscape originated with the work of R. Murray Schafer [2]. Among other ideas, Schafer proposed soundwalks as empirical methods for identifying a soundscape for a specific location. In a soundwalk people are supposed to move in a specific location, noticing all the environmental sounds heard. Schafer claimed that each place has a soundmark, i.e., sounds which one identifies a place with. The idea of experiencing a place by listening has been recently further developed by Blesser and Salter [3]. By synthesizing technical, aesthetical and humanistic considerations, the authors describe the field of aural architecture and its importance in everyday life.

In the field of virtual reality, studies have recently shown how the addition of auditory cues could lead to measurable enhancement in the feeling of presence. Results are available on sound delivery methods [4, 5] or sound quality [6, 5]. Recently, the role of self-sound to enhance sense of presence in virtual environments has been investigated. By combining different kinds of auditory feedback consisting of interactive footsteps sounds created by ego-motion with static soundscapes, it was shown how motion in virtual reality is significantly enhanced when moving sound sources and ego-motion are rendered [7].

In [8, 9, 10], a system to synthesize in real-time the sound of footsteps on different materials was presented. The system was composed of a set of four contact microphones, a multichannel soundcard, a set of headphones and a laptop. The microphones detected real footsteps sounds from users, from which the ground reaction force (GRF) was estimated. Such GRF was used to control a sound synthesis engine based on physical models.

This interactive system was evaluated in a between-subjects experiment, where it was compared to a recognition task including recorded and synthesized offline sounds. Results showed that subjects were able to recognize most of the synthesized surfaces with high accuracy. Similar accuracy was also noticed in the recognition of real recorded footsteps sounds, which was an indication of the success of the proposed algorithms and their control [9].

In this paper, we are interested in understanding whether the addition of a soundscape enhances the recognition of the simulated surfaces. Our hypothesis is that context plays an important role in the recognition of the material a person is stepping upon. In order to test such hypothesis, we designed different soundscapes, described in the following section.

2. SOUNDSCAPE DESIGN

The soundscapes of the following environments were built:

1. A beach and seaside during the summer
2. A courtyard of a farm in the countryside
3. A ski slope
4. A forest
5. A park during the fall

Such soundscapes were designed according to the indications given by subjects answering to a questionnaire. Precisely, ten subjects, chosen among those not performing the experiment, were asked to imagine which sounds could occur in the above mentioned environments.

Subjects were asked the following question: "Imagine that you are right now in a forest: which sounds do you think you would hear?" In this particular environment,

subjects indicated sounds like trees, birds, different animals. Among the answers provided, we chose those which were stated by more than one subject, and collected a corresponding sound material using appropriate recordings of real sounds.

Such sounds were chosen among those available both on the Hollywood Edge sound effects library [1] and on the Freesound.org website. [2]
The chosen sounds were opportunely edited and assembled using the sound editor Adobe Audition 3. Soundscapes were designed with the goal of providing a clear idea of the designed environment already from the first seconds.

3. EXPERIMENT

We conducted an experiment whose goal was to investigate the ability of subjects to recognize the different walking sounds they were exposed to in three conditions: without soundscapes, with coherent soundscapes and with incoherent soundscapes.
The footsteps sounds provided during the three conditions were synthesized sounds generated in real time while subjects were walking using the interactive system described in the previous section. The soundscapes were audio files played in background independently from the subjects movements. The volumes of both footsteps and soundscapes were set by empirical investigation.

One of our hypotheses was that the recognition would have improved using coherent soundscapes rather than the conditions with no soundscapes and with incoherent soundscapes. Similarly we hypothesized higher evaluations in terms of realism and quality in presence of coherent soundscapes.

3.1 Methods

A between-subject experiment with the following three conditions was conducted:

1. Condition 1: footsteps sounds without soundscapes.
2. Condition 2: footsteps sounds with coherent soundscapes.
3. Condition 3: footsteps sounds with incoherent soundscapes.

Participants were exposed to 10 trials in conditions 1 and 2, and 12 trials in condition 3.

During conditions 1 and 2, 5 stimuli were presented twice in randomized order. The stimuli in condition 1 consisted of footsteps sounds on the following surfaces: beach sand, gravel, snow (in particular deep snow), forest underbrush (a forest floor composed by dirt, leaves and branches breaking), dry leaves. In condition 2 the stimuli consisted of the same footsteps sounds provided in condition 1 with in addition the corresponding coherent soundscape mentioned in section 2.

During condition 3, 6 stimuli were presented twice in randomized order. The stimuli consisted of footsteps sounds on the surfaces beach sand, snow, forest underbrush, with in addition an incoherent soundscape. As an example in presence of the footstep sound on beach sand the provided soundscapes corresponded to those of footstep sounds on snow (i.e., the ski slope) and on forest underbrush (i.e., the forest environment).

3.1.1 Participants

Forty three participants were divided in three groups to perform the three conditions in a between-subjects experiments (n = 15, n = 15 and n = 13 respectively). The three groups were composed respectively of 11 men and 4 women, aged between 21 and 28 (mean = 23.67, standard deviation = 2.12), 8 men and 7 women, aged between 19 and 38 (mean = 24.67, standard deviation = 5.97), and 6 men and 7 women, aged between 21 and 30 (mean = 24, standard deviation = 3.1). All participants reported normal hearing conditions. All participants were naive with respect to the experimental setup and to the purpose of the experiment.

During the experiment the shoes used by subjects were sneakers, trainers, boots and other kinds of shoes with rubber soil.
The participants took on average about 11, 13 and 16 minutes for conditions 1, 2 and 3 respectively.

3.1.2 Setup

The experiment was carried out in an acoustically isolated laboratory where the setup was installed (see Fig. 2). Participants were asked to walk inside an area delimited by four microphones placed in a square configuration on a medium density fiberboard (MDF). [3] Specifically, we used four Shure BETA 91, [4] high performance condenser microphones with a tailored frequency response designed specifically for kick drums and other bass instruments. The microphones' features made them a good candidate for the purpose of capturing footsteps sounds.
The MDF was adopted in place of the carpeted floor of the laboratory in order to improve the quality of the input signal.

The floor microphones were connected to a soundcard, [5] which in turn was connected to a laptop running the sound synthesis engine. Finally the synthesized sounds, as well as the soundscapes, were provided to the user by means of a set of headphones. [6]

3.1.3 Task

During the experiment the participants were asked to wear a pair of headphones and to walk on the MDF in the area delimited by the microphones. They were given a list of different surfaces to be held in one hand, presented as non-forced alternate choice. Such list included a range of materials wider than those presented in experiment. During

[1] www.hollywoodedge.com/
[2] www.freesound.org/
[3] 2.5 x 2 m in size and 1 cm thick.
[4] http://www.shure.com/
[5] We used the Fireface 800 sound card, http://www.rme-audio.com/english/firewire/ff800.htm.
[6] Beyerdynamic DT-770, http://www.beyerdynamic.de/

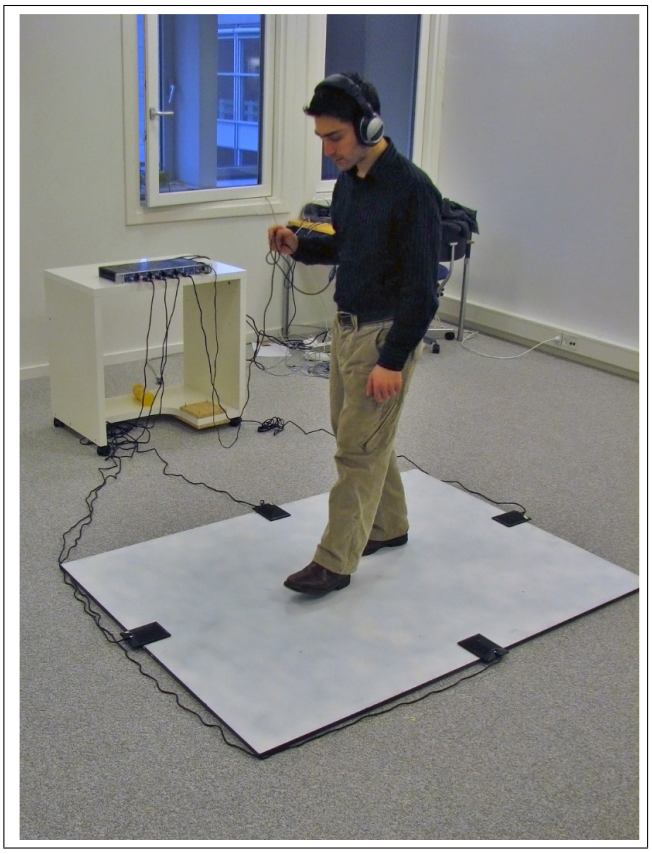

Figure 2. A subject using the interactive footsteps synthesizer. The four contact microphones are clearly noticeable.

the act of walking they listened simultaneously to footsteps sounds on a different surface according to the stimuli presented. The task, common to the three conditions, consisted of answering by voice the following three questions after the presentation of the stimulus:

1. Which surface do you think you are walking on? For each stimulus choose an answer in the following list: 1) beach sand, 2) gravel, 3) dirt, 4) snow, 5) high grass, 6) forest underbrush, 7) dry leaves, 8) wood, 9) creaking wood, 10) metal, 11) carpet, 12) concrete, 13) frozen snow, 14) puddles, 15) water, 16) I don't know.

2. How close to real life is the sound in comparison with the surface you think it is? Evaluate the degree of realism on a scale from 1 to 7 (1 = low realism, 7 = high realism).

3. Evaluate the quality of the sound on a scale from 1 to 7 (1 = low quality, 7 = high quality).

In condition 2 and 3, participants were also asked to recognize what was the environment they were walking.
They were informed that they could choose the same material more than once and that they were not forced to choose all the materials in the list. In addition they could use the interactive system as much as they wanted before giving an answer. When passed to the next stimulus they could not change the answer to the previous stimuli.

At the conclusion of the experiment, participants were also given the opportunity to leave an open comment on their experience interacting with the system.

3.2 Results

The collected answers were analyzed and compared between the three conditions. Results are shown in tables 1, 2 and 3.

The first noticeable element emerging from the three tables is that the use of the interactive system in the condition of coherent soundscapes gave rise to a better recognition of the surfaces and a higher evaluation of realism and quality of the proposed sounds, rather than the conditions with no soundscapes and with incoherent soundscapes. Concerning the percentages of correct answers, they are higher for condition 2 compared to condition 1, for each surface, and the analysis by means of a chi-square test reveals that such differences are statistically significant for beach sand ($p = 0.005515$) and forest underbrush ($p = 0.01904$).
It is particularly interesting to notice that overall adding a soundscape enhances the recognition factor, and this is especially noticeable for those situations where the recognition was rather low without a soundscape.

Similarly the percentages of correct answers are higher for condition 2 compared to condition 3, for each surface, in particular the differences are statistically significant for beach sand ($p = 6.232e\text{-}07$), snow ($p = 0.01439$) and forest underbrush ($p = 0.001637$).
Furthermore, the percentages of correct answers are higher for condition 1 compared to condition 3, for each surface, but the differences are not statistically significant.

The analysis of the wrong answers reveals that in all the experiments none of the presented aggregate surfaces was recognized as a solid surface. This means that all subjects were able to identify at least the nature of the surface, which was an expected feature of the simulations. An observation from the subjects performing the experiment was that many subjects perceived the simulated sounds as very similar, and therefore hard to recognize and distinguish from the list provided.

It is interesting to examine what happens when subjects are exposed to soundscapes which are incoherent, as shown in Table 3. In this situation, we consider as correct the answer provided when subjects recognize the surface they are walking on, and not the soundscape. As it can be noticed, the percentage of correct answers is quite low. As expected, adding an incoherent soundscape creates a stronger context which often confuses the subjects. This can be observed, for example, in the case of beach sand footsteps which were rendered together with a forest soundscape and a ski slope soundscape. The recognition rate is higher in the first case than in the second, where several subjects confused sand with snow. The subjects' answers for the three conditions are outlined in the confusion matrices shown in Table 4, 5, 6 respectively. Such matrices show information concerning actual classifications performed by the subjects. From the matrices, it can be noticed how the subjects' recognition varies from condi-

tion 1 to condition 2. As an example, the first row of the matrix illustrated the number of subjects which recognized the beach sand surface, with (Table 4) and without (Table 5) a soundscape. The role of the soundscape to enhance the recognition is clearly noticeable.

Table 6 illustrates the confusion matrix for condition 3, i.e., when incoherent soundscapes are presented to the subjects. In this situation, it is clearly noticeable how the nature of the soundscape plays an important role. Moreover, it can be noticed how the incoherent soundscape is in most situation predominant, in the sense that subjects tend to judge the surface they are stepping upon more listening to the soundscape than listening to the actual surface. On the other hand, even if subjects are not able to recognize the surface they are stepping upon, they never confuse its nature, in the sense that they never select a solid surface when exposed to an aggregate one.

In addition, Tables 1, 2 and 3 show the degree to which participants judged the realism and quality of the experience. The degree of realism was calculated by looking only at that data from correct answers, i.e., when the surfaces were correctly recognized. This choice was performed since we were interested in understanding whether the simulation of specific surfaces recognized by the subjects was satisfactory.

As far as the quality judgement is concerned, the data was based on all the answers different from "I don't know".

The mean of realism is higher for condition 2 compared to condition 1 for each surface with the exception of beach sand (which is almost equal). The analysis by means of a t-test reveals that such differences are statistically significant for snow (p = 0.01055) and forest underbrush (p = 0.005595).
Analogously, the mean of realism is higher for condition 2 compared to condition 3 for each surface with the exception of beach sand (which is almost equal). In particular, the differences are statistically significant for beach sand (p = 0.002568) and snow (p = 0.001938).
Moreover, the mean of realism is higher for condition 1 compared to condition 3 for each surface with the exception of forest underbrush, which is minor. Such differences are statistically significant for beach sand (p = 0.001302), and for forest underbrush (p = 0.03438), which, as said, is greater for experiment 3.

As regards the mean of quality, it is higher for condition 2 compared to condition 1, with statistically significant differences for all the surfaces with the exception of dry leaves: beach sand (p = 0.009619), gravel (p = 0.02169), snow (p = 0.0006874) and forest underbrush (p = 0.02198).
The mean of quality is higher for condition 2 compared to condition 3 for each surface, and in particular the differences are statistically significant for beach sand (p = 0.006187), for snow (p = 9.596e-05).
Furthermore, the mean of quality is similar for condition 1 compared to condition 3, with the exception of forest underbrush for which it is higher in condition 3 compared to condition 1, with statistically significant differences (p = 0.03204).

	% Correct answers	% Wrong answers	% "I don't know"	Realism	Quality
Beach Sand	50.	46.67	3.33	5.2	4.7241
Gravel	83.33	6.67	10.	5.2	4.6296
Snow	73.33	26.67	0.	5.2955	5.1167
Forest Underbrush	40.	50.	10.	3.5	4.1923
Dry Leaves	16.67	63.33	20.	4.4	3.9167

Table 1. Results of condition 1: recognition of the surfaces without soundscapes.

	% Correct answers	%Wrong answers	% "I don't know"	Realism	Quality	% Correct soundscape
Beach Sand	86.67	10.	3.33	5.1481	5.5172	93.33
Gravel	86.67	13.33	0.	5.3077	5.4	86.67
Snow	80.	20.	0.	6.1667	6.0667	83.33
Forest Underbrush	73.33	26.67	0.	4.9091	5.0333	100.
Dry Leaves	30.	70.	0.	4.4444	4.5	96.67

Table 2. Results of condition 2: recognition of the surfaces with coherent soundscapes.

The comparison about the percentages of "I don't know" answers reveals that for each surface they are higher for condition 1 compared to condition 2, and for condition 3 compared to condition 2. In addition, they are higher for condition 3 compared to condition 1, for each surface with the exception of forest underbrush (which is minor).

As regards the percentages of correct answers about the soundscapes presented, they are higher for condition 2 compared to condition 3, and in particular the differences are statistically significant for the ski slope soundscape (p = 0.0003945).

Overall, subjects observed that soundscapes play an important role in recognition of the surfaces, precisely for their ability to create a context. Especially in terms of conflicting cues, as it was the case in condition 3, subjects were trying to identify the strongest cues, i.e. the element which had the strongest recognition factor. Sometimes the subjects found this task quite hard to complete, and this is why the percentage of "I don't know" answers is higher in condition 3 as opposed to condition 2.

When leaving a comment, several subjects observed that the recognition of snow was extremely realistic. This observation is also confirmed by the high degree of realism (mean = 5.3) and quality (mean = 5.1) with which such surface was rated.

On the other hand, for some subjects the concept of dry leaves was rather confusing, and this is also confirmed by the low recognition rate of such surface.

Overall, this experiment represents a strong indication

Material	Soundscape	% Correct answers	% Wrong answers	% No idea	Realism	Quality	% Correct soundscape
Beach Sand	Forest	38.46	57.7	3.84	4	5.16	88.46
Beach Sand	Ski slope	15.38	69.24	15.38	3.75	4.3182	38.46
Snow	Forest	50	42.31	7.69	5.2857	5.3542	96.15
Snow	Beach	50	46.16	3.84	4.7692	5	88.46
Forest Underbrush	Beach	38.46	53.85	7.69	4.8	4.9583	65.38
Forest Underbrush	Ski slope	30.76	65.4	3.84	4.3	4.6087	46.15

Table 3. Results of condition 3: recognition of the surfaces with incoherent soundscapes.

	BS	GL	SW	UB	DL	HG	DR	FS	WD	CW	MT	CC	PD	WT	CP	—
BS	15	2	6	2	2	2										1
GL		25			1								1			3
SW			22	1	1			6								
UB			4	12			1	10								3
DL		9		4	5			6								6

Legend:							
	WD	wood	CW	creaking wood	SW	snow	UB	underbrush
	—	don't know	FS	Frozen snow	BS	beach sand	GL	Gravel
	MT	metal	HG	High grass	DL	dry leaves	CC	concrete
	DR	dirt	PD	puddles	WT	Water	CP	carpet

Table 4. Confusion matrix of condition 1.

	BS	GL	SW	UB	DL	HG	DR	FS	WD	CW	MT	CC	PD	WT	CP	—
BS	26	1	1	1												1
GL		26			3	1										
SW	2		24					4								
UB		1	2	22	1			4								
DL	6			9	9			6								

Legend:							
	WD	wood	CW	creaking wood	SW	snow	UB	underbrush
	—	don't know	FS	Frozen snow	BS	beach sand	GL	Gravel
	MT	metal	HG	High grass	DL	dry leaves	CC	concrete
	DR	dirt	PD	puddles	WT	Water	CP	carpet

Table 5. Confusion matrix of condition 2.

	Soundscape	BS	GL	SW	UB	DL	HG	DR	FS	WD	CW	MT	CC	PD	WT	CP	—
BS	Forest	10		2	5	3	1	1	3								1
BS	Ski slope	4	1	5		1	5	3	1					1	1		4
SW	Forest		5	13	1	1			4								2
SW	Beach	2	6	13	2				2								1
UB	Beach		3	4	10	2			5								2
UB	Ski slope		1	1	8	3	1	2	8							1	1

Legend:							
	WD	wood	CW	creaking wood	SW	snow	UB	underbrush
	—	don't know	FS	Frozen snow	BS	beach sand	GL	Gravel
	MT	metal	HG	High grass	DL	dry leaves	CC	concrete
	DR	dirt	PD	puddles	WT	Water	CP	carpet

Table 6. Confusion matrix of condition 3.

of the importance of context in the recognition of a virtual auditory place, where self sounds created by users' footsteps and soundscapes are combined. Further investigations are needed to enhance the realism of the simulated soundscape, in particular by having the auditory cues changing according to the motion of the subject in the space.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we describe an experiment whose goal is to understand the role of soundscapes in creating a sense of place and context when designing a virtual walking experience. In this particular experiment, the user was not able to interact with the soundscapes, which were made of mere soundtracks. The results described are an interesting starting point for further investigations on the role of environmental sounds to create a sense of place. While walking an acting in an environment, a person is exposed to her own self-sounds as well as the sounds of the place. This paper presents a preliminary investigation of the role of these different elements both taken in isolation and combined. Further investigations are needed to gain a better understanding of the cognitive factors involved when subjects are exposed to different sound events, especially when a situation of semantic incongruence is present.

We are also planning to design the soundscapes in a multichannel environment, where moving sound sources are present, and the location of the different sound events depends on the location of the subjects. We are also planning to enhance the simulations with visual and haptic feedback. This will allow us to investigate in depth the role of sound to create a sense of place in unimodal and multimodal environments.

5. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community' s Seventh Framework Programme under FET-Open grant agreement 222107 NIW - Natural Interactive Walking. [7]

6. REFERENCES

[1] M. Chion, C. Gorbman, and W. Murch, *Audio-vision: sound on screen*. Columbia Univ Pr, 1994.

[2] R. Schafer, *The tuning of the world*. Alfred A. Knopf, 1977.

[3] B. Blesser and L. Salter, *Spaces speak, are you listening?: experiencing aural architecture*. MIT Press, 2006.

[4] R. Storms and M. Zyda, "Interactions in perceived quality of auditory-visual displays," *Presence: Teleoperators & Virtual Environments*, vol. 9, no. 6, pp. 557–580, 2000.

[5] R. Sanders Jr, *The effect of sound delivery methods on a users sense of presence in a virtual environment*. PhD thesis, NAVAL POSTGRADUATE SCHOOL, 2002.

[6] P. Chueng and P. Marsden, "Designing Auditory Spaces to Support Sense of Place: The Role of Expectation," in *CSCW Workshop: The Role of Place in Shaping Virtual Community*, Citeseer, 2002.

[7] R. Nordahl, "Increasing the motion of users in photo-realistic virtual environments by utilizing auditory rendering of the environment and ego-motion," *Proceedings of Presence*, pp. 57–62, 2006.

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

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

[10] S. Serafin, L. Turchet, and R. Nordahl, "Extraction of ground reaction forces for real-time synthesis of walking sounds," in *Proc. Audiomostly*, 2009.

[7] www.niwproject.eu