Recording natural sound sources and implementing them in virtual acoustic scenes

Frank Wefers, Sönke Pelzer, Michael Vorländer
Institute of Technical Acoustics, RWTH Aachen University
Neustrasse 50, 52066 Aachen, Germany
{fwe, spe, mvo}@akustik.rwth-aachen.de

Abstract

The creator of a virtual scene is faced with the task to obtain audio signals for the sound sources. These can be recorded or synthesized. While synthesis instantly delivers anechoic source material, as needed for auralization, it might not sound convincing. Recordings allow capturing the 'real sound' of an object including all nuances. Unfortunately, not all real-world objects can be taken into a lab and be recorded under anechoic conditions.

In this publication we consider a method of recording natural sound sources under real-world conditions and afterwards implementing them into virtual acoustic scenarios. The objects of study are an array of water fountains and an electric sliding door. We recorded these objects on-site, using several microphone setups. For the fountains, a linear microphone array was additionally used also to capture the directivity. We present how the recorded signals can then be mapped to a suitable setup of virtual sound sources. Concerning performance, it is desirable to only use as much virtual sound sources as necessary. We present the results of listening experiments, which were carried out to determine the required minimum number of virtual sound sources for a given distance from listener to object and source width angle.

Introduction

This work was motivated by the creation of a comprehensive demo scene for the real-time auralization system developed at ITA [1]. It allows simulating complex scenes with interconnected rooms and can also simulate the sound transmission through doors and walls. An extended demo scene is being created, which features a virtual model of the complete Eurogress building, including inner rooms and outer surroundings. Our objective is to create realistically sounding virtual scenes, featuring a high level of acoustic details. Building such comprehensive virtual scenes, with lots a individual sound sources, is a challenge concerning the sound design. We address the following question: Is it feasible for auralization to record sound sources on-site, even under non-optimal rather echoic conditions? How must the recorded signals be processed and how can they be implemented into virtual acoustic scenes, maintaining sound source parameters, like the sound width and directivity? Is this method an alternative to sound synthesis, delivering more realistic sounds?

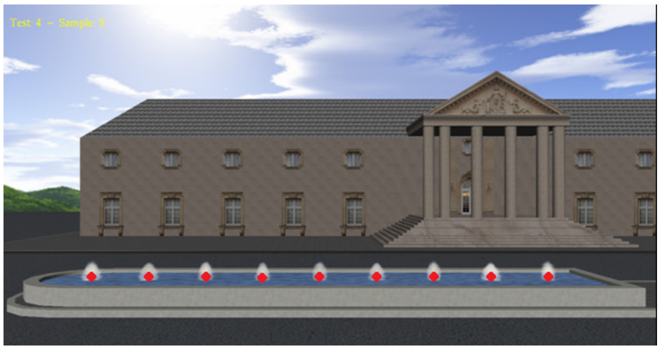

Figure 1: Virtual model of the array of water fountains in front of the Aachen casino as used for the listening test. Red points mark the virtual sound sources in the setups.

On-site recordings

The fountains are located in the outside area of Eurogress, in front of the Aachen casino. Two water basins each feature nine water fountains, all of equal shape and height of approximately 0.5m. Within each basin there are nine fountains arranged in a line. The spacing in between the fountains is 2.5m. Background noise at daytime is significant (main street in close proximity). Therefore, all recordings had to be carried out in the night. The acoustic conditions at the fountains are incomparable to free-field conditions, due the reflections on the surroundings.

The recording was done using a line array of 21 microphones. This allowed not only recording the sound, but also measuring an approximate of the fountains directivity. The array spans a height of 3.81m, allowing to measure elevation angles within [0°, 60°] with a resolution of 3°. The array was positioned 2.20m in distance from the fountains center. This turned out to be a good compromise between far-field conditions and high direct sound levels for reasonable signal-to-background-noise ratios. The recordings sound convincing and bring not only the noise character of the splashing water, but also subtle nuances like bubbling and drop sounds. These nuances make a sound synthesis challenging.

Postprocessing

From the 21 recorded signals we extracted time intervals of 60s without disturbances. An energetic directivity in elevation direction and third-octave resolution was calculated from the measurements. We assumed no azimuthal dependency, due to the rotation symmetry. Auralization was later carried out by using a representative signal recorded at 20° and the relative directivity with respect to 20° elevation. We found that even looping a period of 12s is sufficient for the auralization of the fountains, so that users do not perceive any periodicity.

Auralization

We created four auralization setups of the array of fountains. Each auralization setup features a different number of virtual sound sources–from 1, 3, 5 until 9. The single source setup 1 was included, being easily distinguishable. The nine source setup, depicted in figure 1, marks the high-quality case, with 1:1 mapping between fountain and virtual sound source. A crucial point for virtual sound source arrays is that all the sound sources signals need to be uncorrelated; otherwise an undesired directivity is formed by constructive and destructive interference. We obtained quasi-decorrelated signals, but taking time-shifted 12s signals from the 60s recorded signals.

Listening tests

A listening test was conducted for all sound source setups and several distances between fountains and listener. Participants were asked to answer the following question: Which setup sounds more natural to you? We believe that it is reasonable to pose such a high level question. Creators of a scene aim that users perceive it as natural. The important point is that a user does not need to have a reference sound for comparison. The absence of a reference does not spoil the experience, as long as it is perceived natural. Therefore users were *not* presented with a recording of the real fountains before or during the test.

The listening test was carried out in the virtual reality laboratory at ITA. Subjects were presented with an interactive virtual scene, presented on a 3-D projection display of 3.2x2.4m with Polhemus motion tracking. The real-time auralization was performed using binaural synthesis. Participants listened to the binaural stimulus using Sennheiser HD600 circumaural headphones. We allowed listeners to freely move their head, but asked them to stay on the same spot. The test was repeated for several distances between fountains and listener, which cover 3m, 5m, 10m, 20m, 40m corresponding to horizontal spans of 143°, 127°, 90°, 53° and 28°.

The subjects performed paired AB comparisons for all possible pairs of setups in randomized order. Subjects were allowed to listen to the setups as often as they liked. We introduced a small gap of silence between switching of the setups, so that they did not have a change of direct comparison. The statistical analysis was done according to DIN EN 61305-5 [2]. The reliability of the subjects was calculated and subjects with the consistency measures below the Kendall coefficient K<0.6 were excluded from the analysis. From the 20 trained and untrained subjects who participated, from which 15 were reliable. The statistical analysis results in scale factors in the interval [-π/2, +π/2] for each test setup. These factors introduce a rank order on the setups, which represents the perceived naturalness of the setups among the group of participants.

Distance (Span)		**Setup 1** (1 src)	**Setup 2** (3 srcs)	**Setup 3** (5 srcs)	**Setup 4** (9 srcs)
3 m	Scale	0,517	0,838	0,687	0,393
(143°)	Rank	3	**1**	2	4
5 m	Scale	-0,138	0,516	0,635	0,142
(127°)	Rank	4	2	**1**	3
10 m	Scale	-1,056	-0,082	0,393	0,745
(90°)	Rank	4	3	2	**1**
20 m	Scale	-0,589	-0,050	0,364	0,615
(53°)	Rank	4	3	2	**1**
40 m	Scale	-0,476	0,393	0,409	0,332
(28°)	Rank	4	3	**1**	2

Table 2: Results of the interactive listening test

Conclusions

The results of the test are listed in table 1. Standing directly in front of the virtual fountains, within a distance of 3-5m and an angular span of >90°, the majority of subjects chose the solution with three respectively five virtual sources as the most natural one. The consensus among the group of subjects is not strong for these cases. For medium distances in the range of 10-20 m we find a stronger agreement, with significant preference for the nine sources configuration. Still the five sources setup is perceived natural and as good alternative. For a large distance of 40m and a narrow angular span of 28°, we see that users can easily distinguish the single source solution from the others, but do have only a weak agreement along the other setups. A rather small preference for the five sources configuration can be identified.

These results did not match our initial expectations: For small distances we presumed, that there would be a clear choice for the high resolution variant with nine sources. Most subjects reported after the test, that they were clearly able to identify this setup, but they did not consider it the most natural. An important attribute here are the bubbling and splashing sounds within the signals. Considering the crosstalk from the other fountains within the on-site recording, the auralization includes far more than nine isolated fountains, with the effect of blurring out the bubbling and splashing. This is where we presume the reason for the unexpected results. Further investigations could clarify this topic. Nevertheless, we conclude that it is difficult to perform and use on-site recordings for auralizations. Even if this methods results in very realistic sounds, problems arise from the fact that sources cannot be recorded isolated.

References

[1] Schröder, D., Wefers, F., Pelzer, S., Rausch, D., Vorländer, M., Kuhlen, T., Virtual Reality System at RWTH Aachen University, International Symposium on Room Acoustics (ISRA) 2010, Sydney, 2010.

[2] DIN EN 61305-5 Bbl 1:2005-11 Übersetzung IEC / TR 61305-6:2005. Hörprüfung für Lautsprecher - Einzelprüfverfahren und Paarvergleich.