Audio Engineering Society

Convention Paper

Presented at the 134th Convention
2013 May 4–7 Rome, Italy

Recording and playback techniques employed for the Urban Sounds project

Angelo Farina[1], Andrea Capra[1], Alberto Amendola[1] and Simone Campanini[1]

[1] Industrial Engineering Department, Università di Parma, Via delle Scienze 181/A, 43124 Parma, Italy
farina@unipr.it

ABSTRACT

The "Urban Sounds" project, born from a cooperation of the Industrial Engineering Department - University of Parma with the municipal institution La Casa della Musica, aims to record characteristic soundscapes in the town of Parma with a dual purpose: delivering to posterity an archive of recorded sound fields to document Parma in 2012, employing advanced 3D surround recording techniques, and creation of a "musical" Ambisonics composition for leading the audience through a virtual tour of the town.

The archive includes recordings of various "soundscapes", such as streets, squares, schools, churches, meeting places, public parks, train station, airport and everything was considered peculiar of the town. This paper details the advanced digital sound processing techniques employed for recording, processing and playback.

1. INTRODUCTION

The urban landscape is constantly changing: habits and customs of the people, transportation, residential areas and industrial ones change themselves. From a "visual" point of view, the media that allow us to keep track of the changes (photos and videos) have been extensively used in the past 150 years, but it's not the same for the "acoustic" point of view. Almost all sounds and noises of earlier eras have been lost, with the small exception of "events" which were considered worth to be recorded. These usually do not cover "urban soundscapes", and are limited to "performances" of artists or political events.

The idea behind this research, a collaboration between the Industrial Engineering Department of the University of Parma and the municipal institution "Casa della Musica", is the sampling of the sounds of the city, by means of a state-of-art 3D audio recording system. The aim is dual: delivering to posterity an archive of recorded sound fields to document Parma in 2012 with advanced 3D surround recording techniques and creation of a "musical" surround composition for leading the audience through a virtual tour of the town. In this paper, in addition to the description of the project and its implementation, we will discuss the possibilities of surround playback offered by different processing techniques of the 32 raw signals coming from the microphone probe.

2. HARDWARE AND SOFTWARE

The Department of Industrial Engineering, University of Parma has been involved in research into immersive 3D audio since many years, focusing both on the recording of sound fields using microphone arrays and reproduction of such recordings using advanced surround sound systems. Important collaborations with the RAI Research Centre (Turin) and with La Casa della Musica (Parma) permitted to setup facilities and equipment which allow for state-of-the-art multichannel recording and playback. In this chapter we describe all the equipment and the signal processing used for this research project.

2.1. The recording system

The probe chosen for this research is the Eigenmike® microphone array, produced by mhAcoustics [1]. As shown in Figure 1, the Eigenmike® is a sphere of aluminium (the radius is 42 mm) with 32 high quality capsules placed on its surface; microphones, pre-amplifiers and A/D converters are packed inside the sphere and all the signals are delivered to the audio interface through a digital CAT-6 cable, employing the A-net Ethernet-based protocol.

Figure 1 Eigenmike® 32-capsules microphone array

The audio interface is an EMIB Firewire interface. It provides to the user two analogue headphones outputs, one ADAT output and the world clock ports for syncing with external hardware.

The probe is not originally equipped with a windshield, fundamental accessory for outdoor recordings, so we designed, built and tested one, taking care of leaving sufficient air around the sphere for cooling the powerful electronics contained within the sphere.

Figure 2 Custom Eigenmike® windshield

We used a Mac Book Pro 13” as recording machine, connected via firewire with the EMIB interface.

Plogue Bidule [5] was the software chosen for recording, as it is one of the few capable of recording a single file containing 32 channels sampled at 48 kHz, 24 bit, either in WAV or W64 formats (the latter allows for very long recordings, exceeding the 2Gb limit of the WAV format).

The gain of the microphone preamplifiers embedded inside the probe was controlled by means of a custom Python application that generates a proper MIDI message and sends it to the Eignemike®.

2.2. Signal processing

All the collected recordings are 32-channels files, each channel containing the signal of the correspondent capsule. From such sampled sound field it is possible to derive a number of virtual microphones, shaping the polar pattern and pointing to any direction. The processing technique here used for deriving virtual microphones was developed by the RAI Research Center in Turin and by AIDA, a spinoff of the University of Parma [4].

No theory is assumed for computing the filters: they are derived directly from a set of measurements. The characterization of the array is based on a matrix of measured anechoic impulse responses, obtained with the sound source placed at a large number D of positions all around the probe. A matrix of measured impulse response coefficients is formed and the matrix has to be numerically inverted (usually employing some

approximate techniques, such as Least Squares plus regularization); in this way the output of the virtual microphone is maximally close to the ideal response prescribed. This method also inherently corrects for transducer deviations and acoustical artefacts (shielding, diffraction, reflection, etc.).

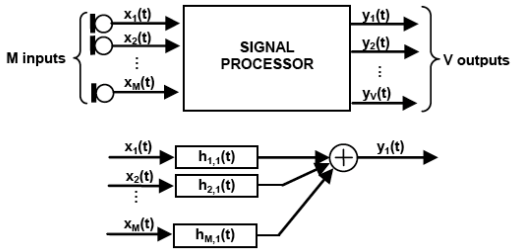

Figure 3 Scheme of the processing for deriving virtual microphones from an array of transducers.

All the processing here summarized can be performed using a Matlab software, using as input data the number of desired virtual microphones, their polar patterns and their pointing direction. The output is a MxV matrix of FIR filters ($h_{m,v}$), where M is the number of capsules (32) and V is the number of desired virtual microphones.
We generated three sets of virtual microphones, according to the following different approaches: High Order Ambisonics (HOA), Direct Synthesis of Virtual Microphones (DSVM) and Spatial PCM Sampling (SPS).

2.2.1. High Order Ambisonics

In this case the number of "virtual microphones" being synthesized is 16 because our goal is the conversion of our raw recordings to 3rd Order Ambisonics signals. Typically, each FIR filter is 2048 samples long (at 48kHz sampling rate). Each harmonic, thus, requires to sum the results of the convolution of 32 input channels with "his" 16 FIR filters. And for getting all the required 16 Ambisonic outputs, we need to convolve-and-sum over a matrix of 32x16 FIR filters, each of 2048 samples.

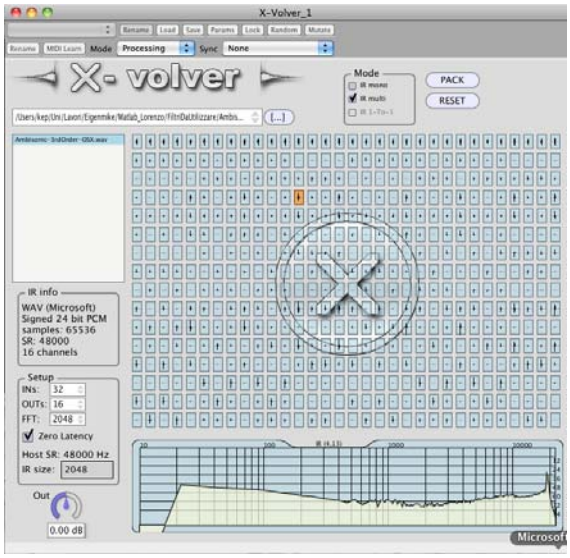

Figure 4 X-Volver and 32x16 FIR matrix

For performing these massive multichannel filtering operations, a special VST plugin was developed, called X-volver, and running either on Mac or Win32 platforms; this plugin is freely available in [7]. Figure 4 shows the X-volver plugin: a 32x16 filter matrix is being employed for converting the signal coming from the 32-capsules spherical microphone array to the 16 3rd order Ambisonic signals.

A modern laptop, equipped with at least an Intel i5 processor, can easily perform such filtering in real-time, either during the recording or during playback.

2.2.2. Direct synthesis of virtual microphones

The idea is to feed every loudspeaker of the playback system with a virtual high order cardioid focused on the direction of the correspondent speaker. As in the case of HOA, we have a desired number V of "virtual microphones" that leads to the creation of a 32xV FIR matrix (each filter is typically 2048 samples long). For example, if eight surround speakers are placed on the vertex of an octagon around the listener, 8 cardioids (3rd order) can be synthesized. with elevation of 0° and azimuth varying between 0° to 315° in 45° steps.

The strong point of this technique is the creation of microphone signals perfectly in phase that does not need further processing (such as a decoder) but which can be successfully delivered directly to the speakers.

2.2.3. Spatial PCM Sampling

We recently developed an alternative technology, which does not rely anymore on spherical harmonics as the "intermediate format", but instead employs simply a number of highly-directive cardioids, covering uniformly the surface of a sphere, for capturing the complete spatial information [6]. In practice, this is the equivalent (in a spherical space) to the representation of a waveform (in time) as a sequence of impulses (PCM, pulse code modulation). Conversely, the Ambisonics approach can be regarded as the equivalent (in a spherical space) to the representation of a waveform (in time) as the superposition of a number of sinusoids (Fourier transform).

We call this new approach Spatial PCM Sampling (SPS); for implementing it in practice, we start with the signal coming from a spherical microphone array (actually, an Eigenmike®). The signals coming form the 32 capsules are filtered by means of a matrix of 32x32 FIR filters, which synthesize 32 virtual high order cardioids, pointing in the same directions as the 32 capsules. So the Eigenmike® is employed as a superdirective beamformer. The 32 superdirective virtual microphones perform a spatial PCM sampling, as each of them can be thought as having a directivity pattern approximating a "spatial Dirac's Delta function". Figure 5 compares the standard PCM representation of a waveform in time with the "spatial PCM".

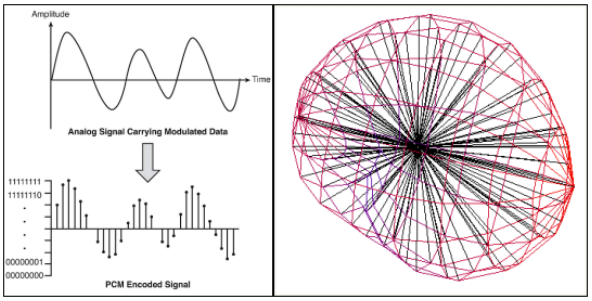

Figure 5 PCM sampling of a waveform in time (left) and of a balloon in space (right)

Once the SPS signals have been obtained (either by recording or by synthesis), it is possible to manipulate them quite easily, performing standard operations such as rotation, stretching, zooming, etc. Instead of performing these operations in the spherical harmonics domain, we can now perform them directly in the spatial PCM domain: hence, the most general transformation assumes the form of a 2D spatial FIR filter.

Finally, it is possible to process the SPS signals for feeding the loudspeakers in a playback system. The math for designing these "decoding" filters is substantially identical to the math employed for encoding the SPS signals.

This is obtained by reprocessing the SPS recording: a new set of virtual microphones is extracted, one feeding each loudspeaker of the playback array. The directivity and aiming of each of these virtual microphones is obtained by solving a linear equation system, imposing that the SPS signals re-recorded placing the Eigenmike® probe at the centre of the playback system are maximally similar to the original recorded SPS signals.

This approach, which is not Ambisonics-based, also corrects inherently for deviations from ideality of the loudspeakers employed, both in terms of magnitude/phase response, and in terms of placement/aiming/shielding. In this approach there is no requirement for the loudspeakers to be equidistant from the listener, so they can be conveniently placed along the walls and in the corners of the room. For feeding a N-loudspeakers array with our 32-channels SPS signals, we need to create a "decoding matrix" of 32xN FIR filters, with substantially the same mathematical approach employed for deriving the "encoding matrix" of 32x32 FIR filters, already described in chapter 2.2.

For determining the filters, we start form a set of measurements of the loudspeaker's impulse responses, performed placing our 32-capsules microphone array at the centre of the listening room (in the "sweep spot position", where the head of the listener should be). The conditions to be imposed for finding the values of filters are that the SPS signals captured by the microphone array, if placed in the centre of the listening room, are identical to the "original" SPS signals.

2.3. The playback systems

The two advanced surround systems used for this research are all available at Casa della Musica (Parma): one is a Wave Field Synthesis room and the second one is a HOA 3D audio system.

2.3.1. WFS Room

In Casa del Suono museum (managed by Casa della Musica) a special Wave Field Synthesis room named “Sala Bianca” (Figure 6) is available [11]. The "Sala Bianca" is designed for 30 seats (7.5 by 4.5 meters and 4.5m high) and equipped with 189 loudspeakers forming a complete crown around the room. These loudspeakers, not visible, are embedded in the walls, just above the ear’s eight: the height of the ring is a compromise between typical ear height for a seated and a standing audience. This room employs a high number of loudspeakers, power amplifiers, D/A converters, all interfaced with state-of-the-art computers. The design of the audio system and control software was made by Fons Adriaensen [12].

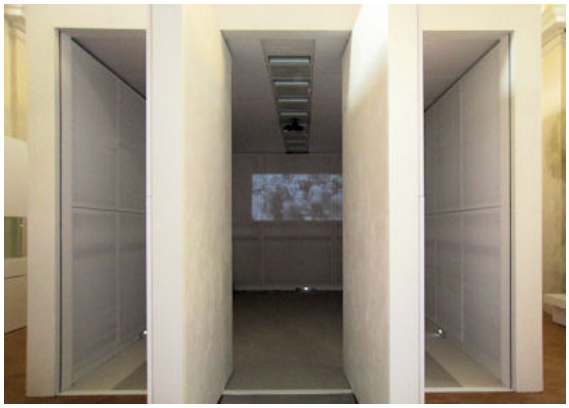

Figure 6 WFS room named “Sala Bianca”, equipped with a 189 speakers ring

The usable frequency range is 50Hz to 20kHz, and sound quality is remarkably good for a design of this size. Except for the space taken by the speakers, the walls are completely covered by sound absorbing material, leading to reasonably good 'dead' acoustics.

2.3.2. HOA Ambisonic Room

A treated HOA 3D room is also available at Casa della Musica. The audio system is made of 16 speakers (Turbosound Impact 50), 8 placed on a regular octagon in the horizontal plane (3rd order), eight speakers are placed at +45° and -45° of elevation (1st order). Everything is controlled by a desktop pc with RME Hammerfall audio interface and completed by Apogee AD-16x digital-to-analog converter and two QSC CX168 amplifiers.

Figure 7 Listening room in Casa della Musica (Parma)

2.4. A first comparison between different rendering techniques

At the time when this paper was written, just one comparative listening test was completed, involving a small number of highly qualified people (sound engineers and musicians). The listeners were given a lot of time for performing the test, they were allowed to listen several times to each sound sample and to switch at will. The test was performed inside the HOA Ambisonic room. The Eigenmike® recordings used for the comparison were essentially two: an orchestra recording and a urban soundscape recorded for the Urban Sounds project (next chapter for details). The compared techniques are the three mentioned in the previous chapter:

- High Order Ambisonics
- Direct synthesis of virtual microphones
- Spatial PCM Sampling

A through analysis shows the approach #2 (Virtual Microphones) produces better localisation when employing 4th-order cardioids, at the price of a less-enveloping and less-natural sound (with the 4th order cardioids the playback loses definition and depth at low frequencies). When the virtual microphones are employed for synthesising 3rd-order cardioids, the localisation results are substantially the same as HOA, which appears preferable being slightly more enveloping and more natural.

Method #3 provided the worst results, due to the heavy mathematical computations involved, causing some audible artefacts (the most noticeable is pre-ringing).

All these results were pointed out with both sound samples, although they resulted more evident for the orchestral one. Despite this sub-optimal result, the authors see the Spatial PCM Sampling as a very promising technique and these results will be employed for improving the processing methods in the near future.

In particular, the filter synthesis method needs to be revised, forcing the generation of quasi-minimum-phase filters, for avoiding pre-ringing.

3. "URBAN SOUNDS" PROJECT

3.1. The places

Parma is a small ancient town, established in the Roman era. During the ages the town was enriched with monuments, squares, parks, buildings that make this centre a destination for tourists from all over the world. For this reason it was not simple to define the most relevant and characteristic places. We defined a first set of 30 locations, postponing to future recording sessions the many other places not sampled yet. In this set we sampled the train station, the airport, two public parks, several squares, a bridge, an highway, an outdoor market and some indoor public places as a commercial centre, a swimming pool, an underground parking, a school canteen and an Italian opera theatre.
For all these places a panoramic photo was taken in the same position of the microphone (an example in Figure 8).

Figure 8 Example of a panoramic photo of "Piazza del Duomo", one of the most important squares of Parma

3.2. The virtual tour

The Virtual Tour is comparable to a musical composition: there are sounds that should be placed in a certain order, adjusted in level and mixed with the right timing. The idea behind the mix is to lead the listener through a well-defined route (Figure 9), starting from the train station and ending to the airport of the town.

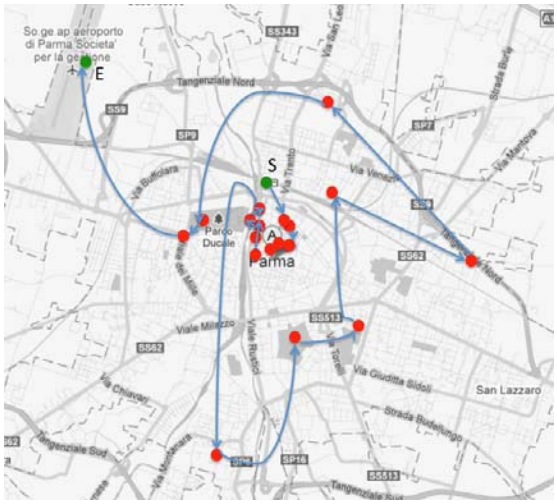

Figure 9 The Virtual Tour of Parma is a route from S (train station) to E (airport)

In this case all the mix was performed converting the "raw" 32-channels recordings to 3rd Order Ambisonics signals (16-channels B-format). The choice of Ambisonics was dictated by the availability of powerful and free software for handling and mixing the sound field.
The mix was monitored and created in the HOA Ambisonic Room (Figure 7) on Linux Ubuntu using the open source DAW "Ardour" [8] and Ambdec [9] for the decoding of the Ambisonic signals.
The composition is accompanied by a "musical" soundscape, a common thread that binds all the urban soundscapes: this musical piece is the result of a recording session with Bloom, a interactive-generative iPad application developed by Brian Eno [10]. The sound coming from iPad is stereo but a special spatialization was performed for rendering the sound in Ambisonic format. Ardour, AMB-Plugins [9] and Zita-Rev [9] were used for achieving an engaging surround sound that was mixed to the recorded soundscapes.

3.2.1. Permanent installation

The Virtual Tour is now available at Sala Bianca (WFS Room). In this case the Wave Field Syntehsis system is used for simulating eight "virtual" speakers placed on a regular octagon at a distance of 15 meters from the centre of the room. Such a distance is useful for loosing the perception of the proximity effect, leading to the listener almost plane waves. The reproduction of 3rd order Ambisonic compositions through a WFS

simulation of the speakers ring gives better performances than a real ring. Unfortunately the system is a planar ring but the lack of speakers above the head is almost compensated by the detailed reconstruction of the sound field in the horizontal plane.

4. CONCLUSIONS

The goal of the "Urban Sounds" project was dual: the creation of an archive of 3D urban soundscapes recorded in Parma and the realization of an immersive virtual tour of the town performed in a WFS room. The composition, object of this paper, is actually available in "Casa del Suono" museum and the soundscapes archive is at disposal of artists and listeners. Everything concerned with the project is collected and displayed in the website www.urbansounds.info, in which the visitor can get a 10 seconds stereo preview of some of the soundscapes recorded.

The "Urban Sounds" project is at its first step, and hence a lot of further work should be done:

- The number of the sampled places should be increased, covering other significant locations in Parma, in order to create a vast archive of urban soundscapes;
- the 3D recordings could be accompanied with a panoramic video, not only with a panoramic photo, for a complete immersive experience of sites where "things happen";
- a permanent installation could be placed in the museum, giving to the visitors the choice of the town places in which they would like to be transported (the idea is a map of the town on a tablet that is used as a remote control).

On the side of signal processing, the future work is all focused on SPS technique:

- the inversion of the audio system IRs matrix has to be optimized for avoiding pre-ringing or other artefacts;
- a larger number of blind listening tests should be performed for a significant comparison between the different 3D audio techniques, and for evaluating objectively the improvements of the SPS method;
- a new set of SPS processing tools should be created for manipulating, rendering and mixing SPS signals, as it is now common practice with HOA signals.

5. ACKNOWLEDGEMENTS

This work was supported by Casa della Musica and by the SITEIA laboratory of the University of Parma. The authors are warmly thankful to Fons Adriaensen for the support in the WFS reproduction step and Francesco Grani for the help in the recording step.

6. REFERENCES

[1] http://www.mhacoustics.com

[2] S. Moreau, J. Daniel, S. Bertet, "*3D sound field recording with High Order Ambisonics - objective measurements and validation of a 4th order spherical microphone.*", presented at 120th AES Convention, Paris, France - May 20-23, 2006.

[3] A. J. Berkhout, D. de Vries and P. Vogel, "*Acoustic control by wave field synthesis*", Journal of the Acoustic Society of America, May 1993, 93(5) pp. 2764-2778.

[4] A. Capra, L. Chiesi, A. Farina, L. Scopece, "A spherical microphone array for synthesizing virtual directive microphones in live broadcasting and in postproduction", presented at 40th AES International Conference "Spatial audio: sense of the sound of space", Tokyo, Japan, October 8 -10 2010.

[5] http://www.plogue.com/products/bidule/

[6] R. Murray Shafer, "The Soundscape", Destiny Books, Rochester, 1977 1994

[7] http://pcfarina.eng.unipr.it/Public/Xvolver/

[8] http://www.ardour.org/

[9] http://kokkinizita.linuxaudio.org/

[10] http://www.generativemusic.com/bloom.html

[11] http://www.casadelsuono.it/

[12] F. Adriaensen, *"The WFS system at La Casa del Suono"*, presented at Linux Audio Conference 2010, Utrecht, The Nederlands.

[13] L. Scopece, A. Farina, A. Capra ,"*360 Degrees Video And Audio Recording And Broadcasting Employing A Parabolic Mirror Camera And A Spherical 32-Capsules Microphone Array*", IBC 2011, Amsterdam, 8-11 September 2011.

[14] Farina, M. Binelli, A. Capra, E. Armelloni, S. Campanini, A. Amendola, "*Recording, Simulation and Reproduction of Spatial Soundfields by Spatial PCM Sampling (SPS)*", International Seminar on Virtual Acoustics, Valencia (Spain), 24-25 November 2011

[15] http://www.kvraudio.com/product/ambisonic_3rd_order_decoder_by_digenis

[16] http://www.brucewiggins.co.uk/?page_id=78