AN INTERACTIVE AND REAL TIME 3D AURALIZATION SYSTEM FOR ROOM ACOUSTICS (THE AURALIAS PROJECT)

Lionel Bos and Jean-Jacques Embrechts

University of Liege, Dept. Electrical Engineering and Computer Science, INTELSIG group,
Sart-Tilman B28, B-4000 Liege 1 (Belgium)
bos@montefiore.ulg.ac.be, jjembrechts@ulg.ac.be

ABSTRACT

The AURALIAS project (www.auralias.be) aims at developing an interactive and real-time auralization system for room acoustics. The user is immersed into a virtual sound field, looking at an image of the virtual room and interacting with it through an appropriate interface. Pre-computed 3D Directional Room Impulse Responses (DRIRs) are first generated by a ray tracing software. They are then convolved with the anechoic signals using frequency block segmented convolution. The filtered signals are finally distributed either to a VBAP (6 loudspeakers at the moment) or an HRTF sound field reproduction module.
The first prototype allows the auralization of multiple isotropic sound sources, the displacement of the virtual listener and the modification of his head's orientation. Further developments will offer the support for moving sources, mobile sweet spot and real-time acoustical or geometrical modifications of the virtual room.

Index Terms— 3D audio, auralization, room acoustics, real-time signal processing.

1. INTRODUCTION

Auralization is "the technique of creating audible sound files from (simulated, measured, or synthesized) data" [1]. In the field of room acoustics predictions, the simulated data consist in room impulse responses (RIR) or directional room impulse response (DRIR) [2].

The audio reproduction system used for creating these virtual sound fields can be headphones, including the effect of HRTFs (head-related transfer functions), or a set of loudspeakers arranged in a specific environment, like an anechoic room or an immersion studio [1,3].

At the beginning, auralization systems were mainly developed conjointly with room acoustics programs. From the 90's, the most famous of them were equipped with an auralization module based on headphones' reproduction. But, as the computers became more and more powerful, the feeling of immersion was improved by the coupling of the auralization with the visualization of the virtual room, allowing the displacement of the listener and the sound source [1,4]. This enhanced complexity requires *real-time auralization*, including very fast convolution algorithms between the RIRs and the input anechoic signals.

2. THE AURALIAS RESEARCH PROJECT (2007-2010) [5]

AURALIAS aims at developing an interactive tool for architects, acousticians and more generally all people (including the final user) involved in a room acoustics project. AURALIAS is a collaboration between three research teams:

- the INTELSIG group of the university of Liege, specialized in applied acoustics, audio signal processing and image processing (for the tracking of the user: see later),
- the LISA group of the free university of Brussels, specialized in computer science, image synthesis and analysis,
- the LUCID group of the university of Liege, specialized in architecture and human-machine interaction.

The system must be able to provide an immersion of the "listener" in a virtual 3D sound field, while he's looking at an image of the virtual room and interacting with it through an appropriate interface.

It can be a single-user application, but it has more preferably been thought as a collaborative experience between a restricted number of persons (for example, between the acoustician and the architect). In this respect, the immersion studio is presently equipped with six loudspeakers in a horizontal plane (the stereo pair at ± 30 deg. from the frontal viewing direction and four more loudspeakers every 60 deg., in a nearly circular arrangement: see fig. 1). This horizontal structure should be completed with some loudspeakers in elevation, in front of the users. All these loudspeakers are fed with their own auralized signal computed by the VBAP technique [6].

The users of the studio are placed in the middle of the loudspeakers' circle and they look at the large screen in front of them, on which an image of the virtual room is projected: see figure 2. They can interact with this projected view, for example by displacing their virtual viewing position in the room, which is also the virtual listener's position. The auralized signals are then changed accordingly, in real-time.

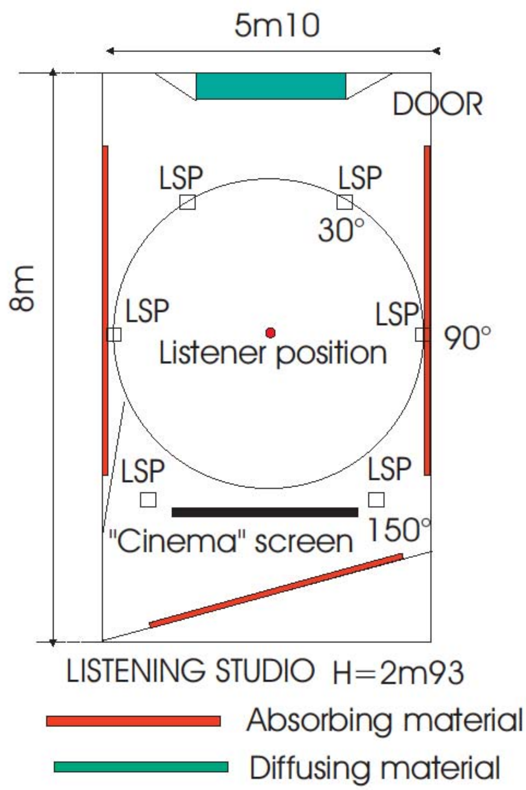

Figure 1: Plane view of the immersion studio with the position of the six loudspeakers (LSP).

As the system is conceived for professional applications, especially for architects and acousticians, it is intended to keep the quality of auralization as high as possible. In this respect, an accurate localization of the direct sound and the early specular contributions is of course an important issue, but also the rendering of all directional characteristics of the sound field (see next section: the room acoustics model). Directional impulse responses are therefore computed around the listener and each of them is distributed to the appropriate loudspeakers for reproduction. This allows for the auralization of rooms with special geometries, such as long disproportionate rooms in which flutter echoes in certain directions can dominate the late part of the RIR.

Finally, interaction with the general system is provided by an interface, which must be as intuitive and user-friendly as possible.

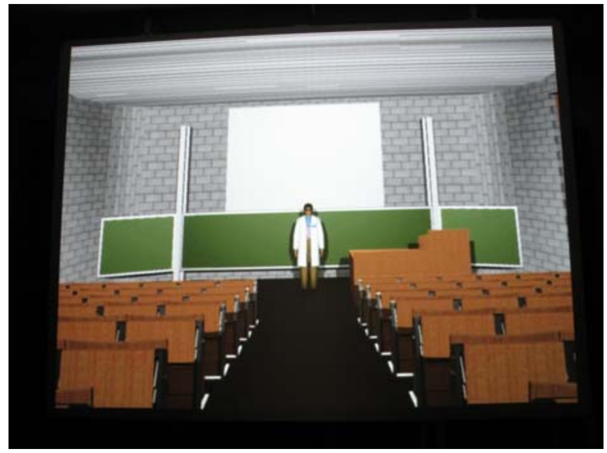

Figure 2: View of the screen in the immersion studio. The users are looking to an image of the virtual room. The front loudspeakers (not apparent on this figure) are situated immediately on the left and right of the screen.

3. THE ROOM ACOUSTICS MODEL

The room acoustics model used in AURALIAS to compute the (D)RIRs is a ray-tracing program developed by the Applied Acoustics' team of the university of Liege [7,8]. This program includes most options offered by modern room acoustics programs (including wideband diffusion [8]) and it has been recently updated to compute directional echograms [2].

In this respect, each receiving position in the modelled room is surrounded by a transparent spherical receptor, divided into a given number of solid angles covering the entire sphere. Usual choices are 6 (up-down, left-right, and front-rear) or 26 solid angles (extension of 45 x 45 deg.). Each solid angle of each receptor records the energy of the incident sound rays, which leads to an echogram per solid angle. This process is repeated for each source position, which finally gives one echogram per solid angle, per frequency (octave) band and per source position, for each receptor.

Finally, these directional echograms (see figure 3) are converted into directional RIRs. Echograms (which are energy-time responses) give the impulse responses' envelope, and the conversion algorithm is adding phase information through "an adequate fine structure representing the actual reflections statistics" [1, p.225].

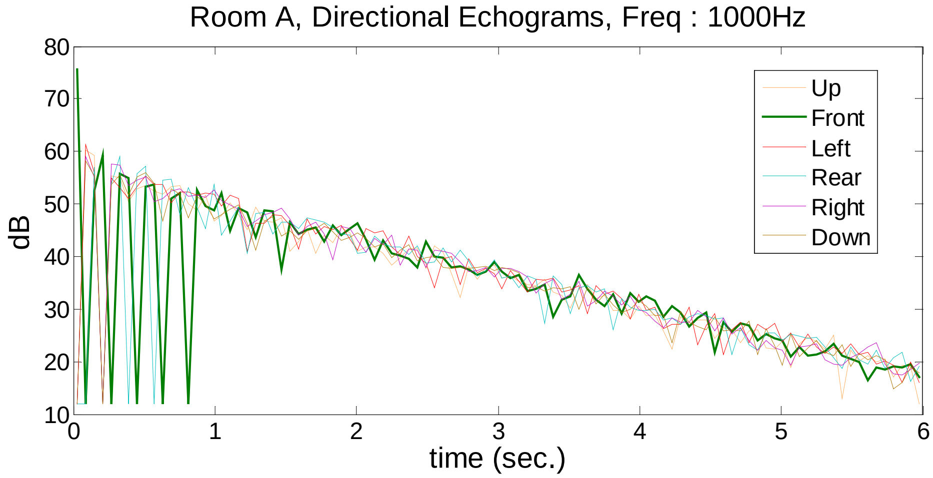

Figure 3: Directional echograms computed in Room A [2,8] and in the 1 kHz octave-band. All reflections are specular.

To improve localization cues, each specular reflection is computed individually, and their amplitude and delay are recorded at each receptor position. Their contribution is therefore removed from the directional echograms, while they are used to derive specific "filters" or impulse responses which can be precisely oriented (in azimuth and elevation) around the virtual listener.

Initially, the ray-tracing program was not conceived for real-time applications. However, as real-time interaction is an important issue for AURALIAS, the algorithm is presently updated to allow for fast modifications of some characteristics of the virtual room:

- the displacement of the sources and the virtual listener;
- the real-time modification of the surfaces' acoustical properties and of some geometrical parameters.

4. THE AURALIZATION FILTER

The auralization filter deals both with sound reverberation and sound spatialization.

4.1. Convolution and reverberation

The reverberation part is currently implemented with a frequency block segmented convolution based on the overlap-add method. According to this method, the filter response is split into several uniform blocks. These blocks are then transformed into the frequency domain by the FFT method, multiplied with a block of input samples (previously converted in the frequency domain too), and finally transformed back to the time domain. This method is less computational intensive than the direct "temporal" convolution, but it introduces an input-output delay at least equal to the length one block. Reducing this latency by using smaller blocks leads to increase the computational load. To deal with this dilemma, the method can be improved by splitting the filter response into non uniform block lengths, using smaller blocks at the beginning of the filter to reduce the latency and larger blocks at the end to keep the computational load acceptable[9, 10]. It's also possible to obtain a zero delay convolution by convolving the first block in the time domain [10].

However, it appears nowadays that the computational load associated to a specific reverberation algorithm may not be the main criteria for real time applications any more. Indeed, the multi-core architecture of modern Central Processor Units (CPU) allows for parallel tasks and offers more and more powerful instructions. These methods have then to be also judged on their ability to take advantage of recent CPUs architectures. This remains of course a large field of investigations, but it seems quite obvious that it's easier to parallelize a simplest algorithm than a more complicated one!

4.2. Spatialization

The sound spatialization is achieved by using the Vector Based Amplitude Panning (VBAP) method [6, 11]. This method allows creating virtual sound sources by distributing the same monophonic signal on several loudspeakers. The position of a virtual sound source then depends on the amplitude of the signal applied to each loudspeaker. As previously mentioned in this paper, a two dimensional spatialization is currently obtained with a 6 loudspeakers configuration and is about to be completed soon to obtain a three dimensional restitution.

5. FEATURES OF THE FIRST PROTOTYPE AND FURTHER DEVELOPMENTS

5.1. Hardware

The first prototype runs on a standard PC under Windows XP 32 bits whose main specifications are:

- CPU : Intel 2.4GHz core 2 quad, Ram :3GB
- GPU : NVIDIA GeForce 8500 GT

A *Toshiba TDP – EX 20* data projector is used for the projection of virtual rooms in front of the users.

Sound is reproduced by six *Far XMD range Digital active three way* loudspeakers, those loudspeakers being driven by an *EDIROL AudioCapture FA – 101* sound card.

The user interface is deported on a *Samsung Q1 Ultra* UMPC. To improve the quality of immersion, this device should be completed (or replaced) in the near future by a joystick.

5.2. Possible actions by the user

The first prototype allows the auralization of an isotropic and monophonic sound source. The user can turn this source on and off, he can control its audio volume and choose its anechoic message. The virtual listener can move to pre-defined positions in the virtual room. Those positions (and that of the source) have been chosen during a previous step of room acoustics simulation, since it's not yet possible to compute long DRIRs in real time. During the auralization session, the adequate DRIRs are then extracted from a filterbank according to the current position of the listener. This also allows to switch between different virtual rooms during the auralization. Finally, it's also possible to choose the orientation of the listener in the horizontal plane and to record the audio output for playback.

5.3. Second prototype

The next prototype is presently under development and it will offer new features concerning the user's interaction in the auralized rooms and both the sound spatialization and reverberation methods.

It will then be possible to add several sound sources and to move them to pre-defined positions. It should also be possible to auralize the user own speech and to partially modify the auralized room on the fly. Room's modifications could affect both the acoustic properties of materials and the geometry of the room.

The current two dimensional VBAP sound spatialization module will be first completed by a three dimensional one and then by some others spatialization methods like headphones HRTF sound reproduction. In order to compare them, the software architecture will enable to switch from one spatialisation mode to another at running time.

As most sound spatialization methods suffer from a quite small sweet spot, a user tracking device will also be implemented to adapt the sound to the real user's position in the immersion studio. It will also be possible to modify the loudspeakers setup to analyse on the fly its effect on the spatialization quality.

The reverberation will also be more deeply investigated by implementing different convolution modules and by allowing the user to switch between them at running time.

6. ACKNOWLEDGEMENTS

AURALIAS is a project funded by the Region of Wallonia (Belgium) under contract n. 616416.

7. REFERENCES

[1] M. Vorländer: Auralization. Springer-Verlag, Berlin, 2008

[2] J.J. Embrechts, N. Werner, S. Lesoinne: Computation of Directional Impulse Responses in Rooms for Better Auralization. Convention paper n.6498 of the 118th convention of the Audio Eng. Soc., Barcelona, 2005

[3] M. Kleiner, B.-I. Dalenbäck, P. Svensson: Auralization-An Overview. J. Audio Eng. Soc. **41** (1993), 861-875

[4] Reference to the DIVA and EVE homepage. URL: http://eve.hut.fi/

[5] Reference to the AURALIAS homepage. URL: http://www.auralias.be

[6] V. Pulkki: Virtual sound source positioning using vector base amplitude panning. Journal of the Audio Engineering Society **45** (1997), 456-466

[7] J.J. Embrechts: Sound field distribution using randomly traced sound ray techniques. Acustica **51** (1982), 288–295

[8] J.J. Embrechts: Broad spectrum diffusion model for room acoustics ray-tracing algorithms. Journal of the Acoustical Society of America **107** (2000), 2068-2081

[9] W. Gardner: Efficient convolution without input-output delay. Journal of the Audio Engineering Society **43** (1995), 127-136

[10] G. Garcia: Optimal Filter Partition for Efficient Convolution with Short Input/Output Delay. Proceedings of the 108th International Conference of the Audio Engineering Society, n.5660, 2002

[11] V. Pulkki, T. Lokki: Creating auditory displays with multiple loudspeakers using VBAP: A case study with DIVA project. International Conference on Auditory Displays, ICAD, Glasgow, 1998