A FLEXIBLE FRAMEWORK FOR PARAMETRIC AURALIZATION

Tapio Lokki and Lauri Savioja (Tapio.Lokki@hut.fi and Lauri.Savioja@hut.fi)
Telecommunications Software and Multimedia Laboratory, Helsinki University of Technology
P.O. Box 5400, FIN-02015 HUT, FINLAND

Abstract: In this paper we present a modular and flexible auralization framework. Auralization methods are to be used with room acoustics modeling techniques that rely on sound field decomposition in which each reflection is modeled with an elementary wave. As an example implementation, we review the DIVA auralization system. In addition, diffraction modeling with image source method is discussed. Finally, we briefly present the results of evaluation of the quality of our auralization system.

Keywords: auralization, room acoustic modeling, diffraction, evaluation of auralization quality

NOTATION

c	speed of sound in air
$h_d(t)$	impulse response of an edge
l	edge point to receiver distance
m	source to edge point distance
R	position of receiver
S	position of source
t	time
z	length of an edge
á	angle relative to source (in Cartesian coordinates)
â$_{\pm\pm}$	directivity function
è$_{\mathrm{R}}$	angle relative to receiver (in cylindrical coordinates)
è$_{\mathrm{S}}$	angle relative to source (in cylindrical coordinates)
è$_{\mathrm{w}}$	wedge angle (in cylindrical coordinates)
ã	angle relative to receiver (in Cartesian coordinates)
ä	unit impulse
í	wedge index

1. INTRODUCTION

Computational modeling of room acoustics has emphasized the research on auralization methods during last decade. Auralization, as defined by Kleiner et al. [1], is understood as convolution of measured or modeled binaural room impulse responses with dry audio signal. With auralization, 3D models of designed rooms and spaces can be listened and auralization has been found as an intuitive way to demonstrate acoustical design for non-acousticians who are not familiar with objective room acoustical parameters. Another application area is found in virtual reality simulations and computer games in which the traditional approach cannot be used, since they require interactive auralization. In these applications auralization methods that enable dynamic room acoustic modeling and auralization in real time have to be utilized.

In this paper auralization is understood to consider both room acoustic modeling and actual convolution process. Usually, the whole binaural impulse responses are computed before the actual convolution process. That enables the use of most know room acoustic modeling methods. Both ray-based (ray-tracing, cone tracing, and image source method, etc.) and wave-based (FEM, BEM, finite difference, etc.) methods are currently applied and often the goal is to model sound propagation as accurately as possible in the space under study. The convolution process is realized either in the time domain using long finite impulse response (FIR) filters or in the frequency domain by multiplying spectra of responses and signals. The actual convolution can be done in real-time, although it is seldom needed. This approach is mainly used in room acoustics prediction programs.

The other way to implement auralization is aimed at dynamic and interactive applications in which the ultimate accuracy of room acoustic modeling is not always needed, since the goal usually is to render plausible, “good enough”, spatial audio for a certain application. The room acoustic modeling methods applied with interactive applications are simplified and optimization is often performed from the perceptual point of view.

The aim of this paper is to present a generic signal processing framework for auralization purposes in dynamic and interactive applications. The framework can be utilized with room acoustic prediction methods that rely on the sound field decomposition [2]. In this concept the entire sound field is divided into elementary wave fronts that are computed and auralized separately. In other words the binaural impulse responses are not explicitly formed, instead each reflection is defined with a set of parameters, which controls audio signal processing. The main benefit of such sound field decomposition is that every single early reflection can be processed separately enabling dynamic time-varying rendering. In addition, based on the computational capacity, the accuracy of auralization can be refined. As computers get faster more and more early reflections can be rendered accurately, diffraction can be included to modeling, more complex geometries can be used, etc.

2. ROOM ACOUSTIC MODELING WITH SOUND FIELD DECOMPOSITION CONCEPT

In room acoustic modeling the propagation of sound waves in a space is studied. The modeling can be divided into two subparts: modeling of the propagation in a medium such as air and modeling of reflections from boundaries of a space. The modeling of wave propagation is quite straightforward. In a free space each sound source emits a spherical wave front, i.e., an elementary wave that propagates homogeneously in all directions. Amplitude of sound is inversely proportional to the distance from the sound source.

Modeling of reflections is more challenging. In each reflection a new wave front is created, and the reflections can be modeled as new sound sources. Therefore it is possible to reduce the model such that recursively in each reflection new sound sources are created. Finally, we have only sound source and all the reflections are replaced with secondary sources. In accurate modeling every reflection is further separated into specular, diffuse, and diffraction parts [2].

Some of these secondary sources are not visible to a listening point due to occlusion by surfaces. For this reason validity of each source is verified with a visibility check. One of the most commonly applied room acoustic modeling techniques, which enables the computation of specular reflections, is the image source method [3,4]. It can also be extended to handle diffraction [5,6].

Figure 1 illustrates the concept of sound field decomposition. Each reflection from a wall is replaced with an image source and each corner (except convex rectangular corners) is replaced with an edge source. All of these secondary sources emit wave fronts that are shown inside the geometry. Diffuse reflections are not considered in this visualization. Indeed, diffuse reflections are not trivial to implement with the image-source method [2], although in most cases diffuse reflections are an important part of the sound field.

In actual rendering the effect of each source is composed to produce the final sound field in the listening positions. With the concept of image sources each elementary wave can be easily filtered with frequency dependent acoustic phenomena such as sound source directivity, distance delay and attenuation, air and material absorption. All spatial sound reproduction methods, such as binaural reproduction for headphones or loudspeakers as well as multi-channel systems such as the vector base amplitude panning (VBAP) [7] and Ambisonics [8] can be easily applied, since with sound field decomposition each elementary wave can be panned to correct direction and a true 3D sound field is produced for a listener.

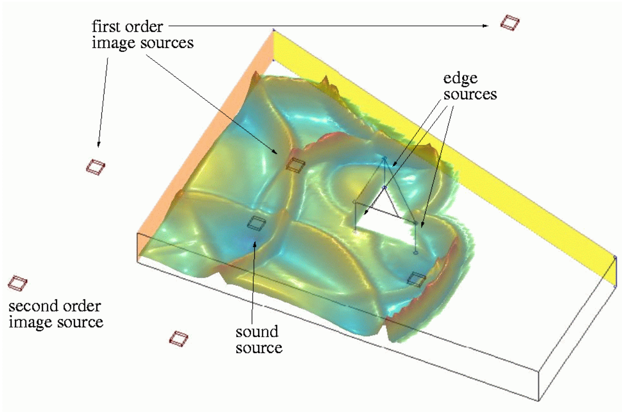

Fig. 1. An example of the sound field decomposition into elementary waves modeled with the image-source method. The illustration is done by computing an impulse response in each pixel and by plotting the time moment of 680^{th} sample that corresponds to 14.2 ms in time.

The next section presents a general auralization framework with which a decomposed sound field can be rendered audible. As an example implementation we briefly describe the DIVA auralization system [9].

3. A FLEXIBLE AURALIZATION FRAMEWORK

Rendering of each elementary wave, emitted by each secondary source, is conceptually copying, delaying, and filtering of sound signal of the sound source. Such delaying of signal can be easily implemented with a long delay line containing several outputs. In dynamic situation (where the delay time of direct sound and reflections change) the pick-up points from the long delay line have to be implemented with fractional delays [10] to guarantee smooth and continuous output. The proper interpolation with constant updates implements also the Doppler effect, which is a desired feature in several virtual reality applications.

As mentioned earlier, such acoustic phenomena as sound source directivity, distance attenuation, air and material absorption have to be modeled to get naturally sounding rendering. Each of these phenomena can be implemented separately with low order digital filters [9]. These filters can be attached to each pick-up point. In Fig. 2, where a schematic drawing of proposed signal processing chain is illustrated, the filter blocks $T_{0...N}(z)$ implement above-mentioned phenomena. In addition to direction independent filtering each secondary source is panned to correct direction. The filtering required for panning depends on the reproduction method. The proposed framework allows both binaural (illustrated in Fig. 2) and multi-channel rendering.

With modern computers the secondary sources can be computed only for early reflections, and late reverberation part has to be produced separately. Late reverberation can be implemented with efficient recursive algorithms or by direct convolution of some modeled response. The recursive algorithms usually produce an ideal response (exponentially decaying noise), but with convolution any kind of response can be rendered. Usually, late reverberation part is treated as time and place invariant response.

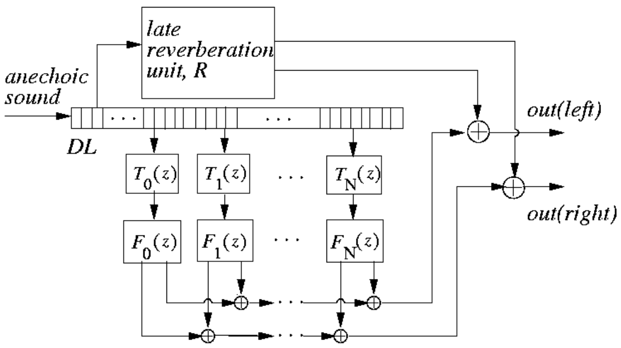

Fig. 2. Signal processing structure for auralization based on sound field decomposition.

3.1. The DIVA Auralization system

As an example of auralization system that is based on sound field composition we present the DIVA auralization system developed since 1994 at the Helsinki University of Technology (HUT). Our aim has been to create a system that produces a perceptually authentic rendering of a modeled space. In our system the modeling is divided into two parts. The first part is time- and place-variant containing modeling of the direct sound and early reflections. The image source method is applied for this purpose. The second part is for rendering late reverberation that is assumed to be diffuse and its rendering parameters do not change as a function of time or place.

In the DIVA auralization system the image-source calculation provides the auralization parameters, which are finally converted to signal processing parameters. The reason for this two level process is the fact, that in dynamic rendering the auralization parameters do not need to be updated for every audio sample. However, the signal processing parameters have to be defined on a sample by sample basis. In the DIVA auralization system this is achieved by interpolating the signal processing parameters between the updates of auralization parameters.

The image source method implemented in the DIVA auralization system gives the following parameters for each image source:

- order of reflection,
- orientation (azimuth and elevation angles) of sound source,
- distance from the listener,
- incoming direction of sound (azimuth and elevation angle in relation to the listener),
- set of filter coefficients describing the material properties in reflections,
- required parameters for calculation of response from a diffracting edge in the case of an edge source.

The parameters of late reverberation are pre-calculated based either on measurements or results of room acoustic modeling. By this technique we can tune the reverberation time and some other essential features of late reverberation according to the properties of the space.

The signal processing structure utilized in the DIVA auralization system is depicted in Fig. 2. It contains a long delay line *DL*, which is fed with anechoic sound to be processed. The distance of the image source from the listener defines the pick-up point to the filter block $T_k(z)$, where k=0,1,2,..,N is the identifier of the image source (k=0 corresponds to the direct sound). Blocks $T_{0...N}(z)$ modify sound signal with the sound source directivity filters, distance dependent gains, air absorption filters and material filters (not for the direct sound). The incoming direction of the sound is defined with blocks $F_{0...N}(z)$ containing directional filtering or panning depending on the reproduction method. The superimposed outputs of the filters $F_{0...N}(z)$ are finally summed with the outputs of the late reverberation unit R which is a complex recursive algorithm [11].

3.2. Modeling of Diffraction

The most recent advancement in the DIVA auralization system is the diffraction modeling, and it is described in more detail in this section. Svensson et al. [5] have derived a mathematical solution for calculating the impulse response for an edge of a finite length. The impulse response is calculated from the source to the listening position through the edge. With this analytical solution the edge diffraction is modeled to the DIVA auralization system. The auralization parameters for an edge source are:

- wedge angle è_w,
- position of source S ,
- position of receiver R,
- start and end point of the edge z_0 and z_1,
- normal vector $\boldsymbol{n}$ of a surface.

With this data for each edge, the impulse response is calculated with the following equations [5,12]:

$$h_d\left(t\right) = -\frac{n}{4p}\int_{z0}^{z1} d\left(t - \frac{m+l}{c}\right)\frac{b_{++} + b_{+-} + b_{-+} + b_{--}}{ml}\,dz$$

$$b_{\pm\pm} = \frac{\sin\left[n\left(p \pm q_S \pm q_R\right)\right]}{\cosh\left(n\cosh^{-1}\frac{1+\sin a\,\sin g}{\cos a\,\cos g}\right) - \cos\left[n\left(p \pm q_S \pm q_R\right)\right]}$$

An example of a finite wedge is depicted in Fig. 3 to illustrate the variables. In addition, c is speed of sound, í=ð/ è_w is the wedge index, m is the source-to-edge point distance, and l is the edge point-to-receiver distance. The integration range is between the two end points of a finite edge.

The diffraction is implemented to the DIVA system by using edge sources in addition to image sources. These edge sources contain one filter more than the image sources and this filter implements the impulse response of an edge. Edge sources implement the diffraction phenomenon as an impulse response in one point [12], but in real life diffraction sources are not point-like. Sound passes the edge through all points along the edge. However, most of the energy is concentrated on the least-time point of the edge. Based on this, the simplification to a point-like secondary source is not too severe. In addition, the diffraction image source, being a point source, can be panned to the direction the least-time point indicates as proposed by Torres et al. [13]. The same principle holds for the sound source directivity, since from the viewpoint of an edge most of the sound energy from the actual source radiates towards the least-time point of the edge. The situations where this simplification could be most audible would be long edges that are close to a listener.

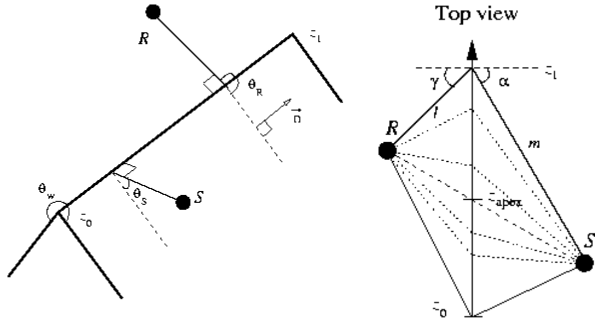

Fig. 3. Geometry of a finite wegde. The positions of source S and receiver R are indicated in cylindrical coordinates. On the right, sound paths via edge points z_0 and z_1 are indicated by solid lines, the least-time sound path via the apex point z_{apex} is depicted with dashed line and some other sound paths are illustrated with dotted lines.

In the current implementation the edge diffraction filters are designed between image source calculation and auralization processes. As such, our implementation is not practical for real-time use, but dynamic off-line rendering is straightforward.

4. PERCEPTUAL EVALUATION OF THE DIVA SYSTEM

In the design and implementation of the DIVA auralization system we have pursued towards an ultimate goal of an authentic auralization in which a listener is unable to distinguish a simulated sound from a recorded sound. For this reason our system has been evaluated by both objective and subjective means. The main emphasis on this section is on the subjective case, but first the objective approach is briefly reviewed. In both cases the careful analysis is performed with a model of one lecture hall.

The objective evaluation has been based on calculation of room acoustic attributes such as reverberation time (T20), early decay time (EDT) and clarity (C50). These attributes have been obtained both from the simulation results and from the corresponding measured impulse responses. In general, the results show that above 400Hz the attributes coincide quite well. However, below that there are some minor defects in modeling, for example, the auralizations are less reverberant than the recordings on that frequency range.

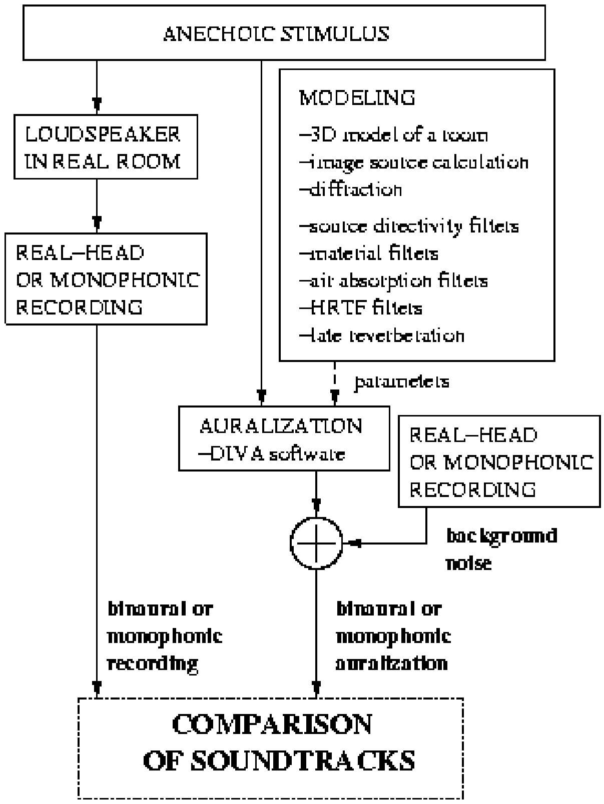

Fig. 4. A framework for perceptual evaluation of auralization systems.

4.1. Evaluation framework

The perceptual evaluation of auralization quality was based on the framework illustrated in Fig. 4 [14]. The evaluation was performed comparing recorded and auralized soundtracks. The recordings made in the studied lecture room were considered as reference signals.

To find out subjective perceptual differences between the recorded and the auralized soundtracks several listening tests have been carried out.

Different listening test methods have been tried out due to the reason that no recommended listening test methodology for testing the auralization quality exists. Finally, we utilized the method called double-blind triple stimulus with hidden reference, including interval scales [15].

The quality of auralization has many different aspects and it is multidimensional by nature. Of course, subjects could only judge whether the soundtracks differ or not, but then no information about the nature of differences is achieved. To obtain more information about possible differences, two attributes, namely spatial and timbral differences, have been studied.

The assessment has been an iterative process containing several evaluation rounds. Totally 20 subjects (three females and 17 males) participated in the final listening test. All of them reported normal hearing although this was not verified with audiometric tests. The test was done in a standard listening room and the headphone reproduction method was applied with Sennheiser HD-580 headphones.

The listening task was to compare spatial and timbral differences between the recorded and the auralized soundtracks. Subjects were told to quantify sound source location, size of space, and reverberation when considering spatial differences. Similarly, such attributes as color of sound and frequency content were told to subjects to be listened for judging timbral differences.

4.2. Evaluation Results

The results of the listening tests show that for certain types of signals we have achieved auralizations that are nearly imperceptible from the corresponding recordings. In general, there were no significant differences between the grades given to spatial and timbral properties. Signals having sustained total characteristics such as sound of a clarinet were judged with the best grades. With signals having transients such as a hit of a snare drum the differences were clearly audible but on the average they were evaluated to be plausible and natural sounding.

5. CONCLUSIONS

This paper deals with auralization methods that are mainly design for virtual reality simulations and computer games. The flexible auralization framework that can be applied in interactive applications is presented. As an example implementation the DIVA auralization system is briefly overviewed. In addition, the inclusion of diffraction to the image source method is discussed. Finally, the evaluation of the quality of implemented system is reported.

6. REFERENCES

1. M. Kleiner, B.-I. Dalenbäck, and U.P. Svensson. Auralization – an overview. *J. Audio Eng. Soc.*, 41(11):861–875, Nov. 1993.
2. U.P. Svensson and U.R. Kristiansen. Computational modeling and simulation of acoustic spaces. In *AES 22nd Int. Conf. on Virtual, Synthetic and Entertainment Audio*, pages 11–30, Espoo, Finland, June 15-17 2002.
3. J. B. Allen and D. A. Berkley. Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am.*, 65(4):943–950, 1979.
4. J. Borish. Extension of the image model to arbitrary polyhedra. *J. Acoust. Soc. Am.*, 75(6):1827–1836, 1984.

5. U.P. Svensson, R.I. Fred, and J. Vanderkooy. Analytic secondary source model of edge diffraction impulse responses. *J. Acoust. Soc. Am.*, 106(5):2331–2344, 1999.
6. V. Pulkki, T. Lokki, and L. Savioja. Implementation and visualization of edge diffraction with image-source method. In *the 112nd Audio Engineering Society (AES) Convention*, Munich, Germany, May 10-13 2002. preprint no. 5603.
7. V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc.*, 45(6):456–466, June 1997.
8. M. Gerzon. Periphony: With-height sound reproduction. *J. Audio Eng. Soc.*, 21(1/2):2–10, 1973.
9. L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen. Creating interactive virtual acoustic environments. *J. Audio Eng. Soc.*, 47(9):675–705, 1999.
10. T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine. Splitting the unit delay – tools for fractional delay filter design. *IEEE Signal Processing Magazine*, 13(1):30–60, Jan. 1996.
11. R. Väänänen, V. Välimäki, J. Huopaniemi, and M. Karjalainen. Efficient and parametric reverberator for room acoustics modeling. In *Proc. Int. Computer Music Conf. (ICMC'97)*, pages 200–203, Thessaloniki, Greece, Sept. 1997.
12. T. Lokki, U.P. Svensson, and L. Savioja. An efficient auralization of edge diffraction. In *AES 21st Int. Conf. on Architectural Acoustics and Sound Reinforcement*, pages 166–172, St. Petersburg, Russia, June 1-3 2002.
13. R.R. Torres, U.P. Svensson, and M. Kleiner. Computation of edge diffraction for more accurate room acoustics auralization. *J. Acoust. Soc. Am.*, 109(2):600–610, 2001.
14. T. Lokki, J. Hiipakka, and L. Savioja. A framework for evaluating virtual acoustic environments. In *the 110th Audio Engineering Society (AES) Convention*, Amsterdam, the Netherlands, May 12-15 2001. preprint no. 5317.
15. ITU-R. Recommendation BS.1116-1, Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. International Telecommunication Union Radiocomminication Assembly, 1997.