ADVANCED 3-D AUDIO IN AN INTERACTIVE MULTIMEDIA ENVIRONMENT

Tapio Lokki[1], Riitta Väänänen[2], Lauri Savioja[1], and Jyri Huopaniemi[2]

Helsinki University of Technology
[1]Laboratory of Telecommunications Software and Multimedia
P.O.Box 1100, FIN-02015 HUT, Finland
[2]Laboratory of Acoustics and Audio Signal Processing
P.O.Box 3000, FIN-02015 HUT, Finland
`http://www.tcm.hut.fi/Research/DIVA/`
Tapio.Lokki@hut.fi, Riitta.Vaananen@hut.fi, Lauri.Savioja@hut.fi, Jyri.Huopaniemi@hut.fi

ABSTRACT

We present a digital signal processing (DSP) implementation of a virtual, interactive acoustic environment, which is a part of an audiovisual presentation in a digital interactive virtual acoustic research project. The process of modeling the sound propagation from a sound source to the listener is referred to as auralization, and it is described how this process is done by digital signal processing. Main emphasis is on the implementation of dynamic direct sound and early reflections, which is a part of the signal processing that changes along with the listener movements in the virtual environment. Many practical issues have to be taken into consideration when time-variant processing is done to produce an impression of naturally responding acoustic environment.

1. INTRODUCTION

This paper presents a filter structure which is a part of the Virtual Auditory Environment (VAE) system implemented in Digital Interactive Virtual Acoustics (DIVA) research project [1, 2]. The aim of the complete DIVA system is a virtual concert performance, where animated musicians play their virtual instruments in a virtual environment. The DIVA system covers sound synthesis, room acoustics modeling, and 3-D sound positioning, combined with synchronized animation lead by a human conductor.

In this paper, we first describe the characteristics of room response which are taken into account in creating a computer-modeled room response. Then, we give an overview of the auralization DSP structure. By auralization we mean modeling and reproduction of the room acoustics, spatial hearing, and sound source directivity. We separately describe the production of the dynamically changing direct sound and early reflections, and the time-invariant late reverberation. Finally, the paper is concluded with suggestions about application areas of the presented acoustic environment modeling system and further work to be done in our project.

2. NATURAL SOUNDING INTERACTIVE VIRTUAL AUDITORY ENVIRONMENT

The listening environment strongly affects the perceived sound. In a room, for example, the perceived sound contains not only the sound coming directly from the source, but also the sound reflected from the surfaces, e.g., the walls and the furniture inside the room. The reflections give information about the acoustics of the room, e.g., its size, shape, and reverberation time.

Each reflection is characterized by the direction, gain, the delay with respect to the direct sound, and frequency modification caused by air and surface material absorption. The gain, delay and frequency modification can be analyzed from the room impulse response. The directions of individual reflections can not be identified, but they are important for perceiving the spatial impression of the room.

To obtain a natural sounding virtual environment, the reflections and their above mentioned properties have to be created. In the process of modeling room acoustics the impulse response is often divided into three parts: The direct sound, early reflections and late reverberation. These parts are illustrated in Figure 1. The perceived direct sound and early reflections give information about the geometry of the room, as well as the locations and orientations of the sound source and the listener. The late reverberation is characterized by an increased density of reflections, and exponentially decaying energy response. The late reverberation is often considered nearly diffuse. In a diffuse sound field the reflections come from all directions with the same probability, and their amplitudes are randomly distributed. Under these assumptions

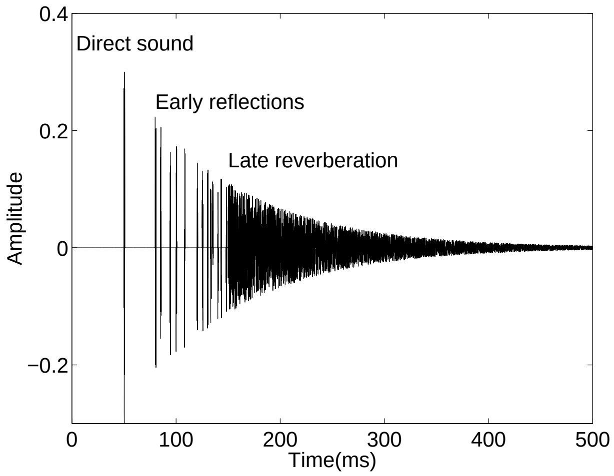

Figure 1. An artificially created room impulse response.

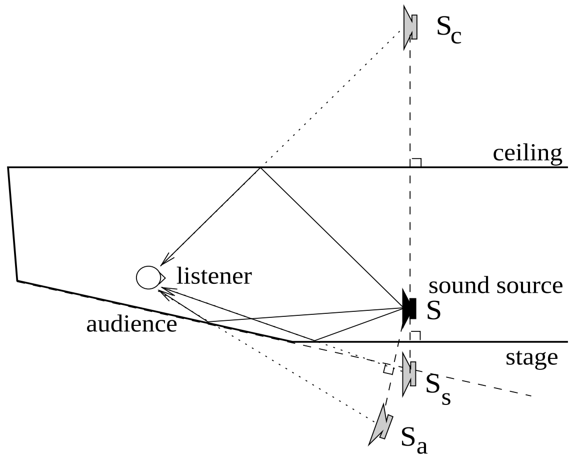

Figure 2. In the image-source method the sound source S is mirrored with respect to surfaces to produce image sources S_s, S_c and S_a which represent the corresponding reflection paths.

the late reverberation does not have to be modeled as individual reflections with directional information.

The impulse response in a room is defined between two points; the sound source and the listener. This means that when the listener moves in a room, the properties of the reflections change depending on his/her location. In an interactive virtual auditory environment the changes in reflections have to be modeled.

3. OVERVIEW OF THE AURALIZATION FILTER STRUCTURE

In the DIVA system an image-source method [3, 4] is used to obtain the locations and orientations of a sound source and image sources that represent reflections from the surfaces of the modeled space. To find image sources the sound source is specularly reflected against the surfaces in the room. An example of three image sources is in Figure 2. Depending on the room geometry and listener position only a small set of image sources is visible corresponding to valid reflection paths. A detailed presentation of the image-source method is out of the scope of this paper, and more about our implementation is presented in, e.g., [1, 2].

The results of the image-source calculation are used as the auralization parameters for the direct sound and each early reflection. These parameters are the location and orientation of the visible image source with respect to the listener location and orientation, and the set of filter coefficients which describe the effect of the surface material in reflections. With these parameters direct sound and early reflections of the modeled space are generated.

4. FILTER IMPLEMENTATIONS FOR DIRECT SOUND AND EARLY REFLECTIONS

Figure 3 illustrates the filter structure used for the auralization implementation. It consists of two blocks, the first of which is time variant; it generates the direc sound and early reflections and it changes along with the listener movements. The second block is the time-invariant late reverberation unit which is explained in section 5. In Figure 3 the blocks $T_{0\ldots N}(z)$ contain directivity filters, material and air absorption filters and a distance dependent gain factor (according to $1/r$-law). The blocks $F_{0\ldots N}(z)$ represent the directional filtering, which is done either by HRTF filtering (discussed below) or by multichannel reproduction, e.g., VBAP that can be used for creating a true 3-D sound field around the listener [5]. The gain $b(r)$ in the input of the late reverberation block is dependent on the distance between the sound source and the listener. It adjusts the level of the direct sound and early reflections fed to the late reverberation block so that the level of the late reverberation remains approximately constant everywhere in modeled space.

The directivity of sound sources is implemented by low-order IIR filters as proposed in [6]. The low-order IIR filters for material absorption are designed by fitting the filters to the magnitude responses that are obtained from absorption coefficients of each material. The method is described in [7]. The air absorption is modeled with first-order IIR filters that are fitted to the magnitude responses calculated analytically as proposed in [7].

In two-channel reproduction, the directional information of the direct sound and early reflections, is

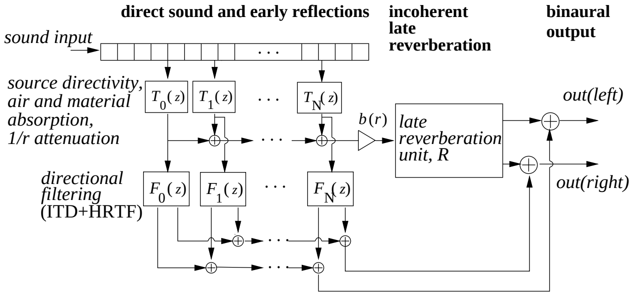

Figure 3. Block diagram of the auralization filter structure.

obtained by head-related transfer function (HRTF) filtering. The HRTF filters are used for simulating the frequency modification that is caused by the effect of the listener's pinnae, head and the upper part of the body. The different modeling techniques for HRTF filters are discussed in [8]. According to [8] the modeling of spatial hearing based on measured HRTFs is done with 30 tap minimum-phase FIR filters from which interaural time delays (ITD) are distinguished and implemented with a short delay line.

4.1. Interpolation of time-varying parameters

In a dynamic system where the listener or the sound source can move in the modeled acoustic space, the parameters which change according to the user movements have to be updated so frequently that the perceived auditory environment corresponses to the movements. Furthermore, interpolation of varying parameters is needed to ensure that no discontinuities in the sound can be perceived. The interactive movements can be taken into account in the auralization if the processing of direct sound and each early reflection is done individually.

The parameters of each reflection are updated at a rate considerably smaller (20 Hz in our system) than the sampling rate of the source sound. The sound source directivity, air absorption and HRTF filter coefficients are changed when the update occurs. Both the IIR and FIR filter coefficients can be changed without disturbing effects if the variations of the filter coefficients are small enough. To obtain such small variations between updates a dense grid of filter coefficients is computed in advance. The stability of interpolated IIR filters is guaranteed if the filters are constructed of second-order sections [9].

The other time-varying parameters namely, distance attenuation gain, propagation delay, and ITD have to be interpolated at every sample between two updates. For distance attenuation gain the interpolation used is a linear interpolation. For propagation delay and ITD, first-order FIR fractional delay between two samples is used to obtain smooth and continuous output. Also higher order interpolation can be used, but first-order is found to be good enough in our system. To minimize the lowpass filtering caused by linear interpolation [10], the fractional delays are applied only when the listener moves. When the listener is not moving, the sample closest to the exact delay is used.

5. LATE REVERBERATION

The late reverberant field in auralization systems is often produced by recursive digital filter structures [11, 12, 13]. The aims in producing artificial late reverberation are:

1. Producing an exponentially decaying impulse response with a dense pattern of reflections, to avoid fluttering in the reverberation.
2. A high modal density of the response. In other words there should be enough resonances in the frequency domain, and no particular resonances should be emphasized more than others, to the extent that would be perceived as coloration in the response to listened sound.
3. In order to simulate natural air absorption, the reverberation time has to decrease smoothly as a function of frequency.
4. To attain a good spatial impression of the sound, and to avoid in-head localization, the late reverberation at both ears of the listener should be incoherent [14].

We have used a recursive digital filter structure consisting of parallel feedback delay loops. The reverberator structure is depicted in Figure 4 [15]. A single loop consists of a delay line, a lowpass filter $H_n(z)$, and a comb-allpass filter $A_n(z)$. The lowpass filters implement and define the frequency dependent reverberation time [11, 13], and the comb-allpass filters added in context with each delay line increase the reflection density as a function of time [15]. The outputs of the parallel delay loops are summed and fed back to the inputs of the delay lines, multiplied by a gain β. The structure is a special case of a feedback delay network, studied by, e.g., Jot [13], and Rocchesso and Smith [16]. A dense impulse response, and incoherent reverberation from different delay lines to different reproduction channels can be obtained by this structure. The quality of the reverberation depends on the number of delay loops, which is chosen according to the available computational capacity.

6. APPLICATIONS AND FUTURE WORK

The proposed filter structure can be used, for example, for demonstrating architectural acoustics by cre-

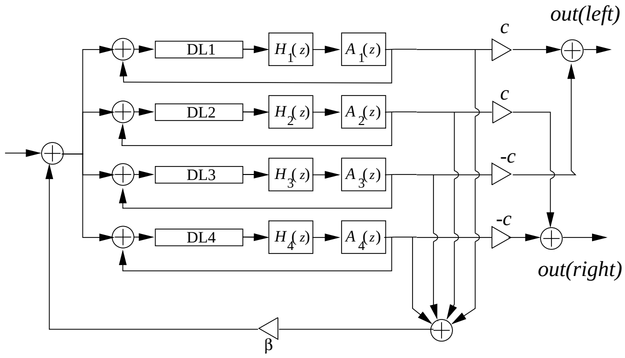

Figure 4. The reverberator filter structure used in the DIVA system to produce incoherent reverberation to separate reproduction channels.

ating a simulated acoustic environment in the design stage of a concert hall. It can also be used to bring further realism to any virtual reality system by creating natural sounding three dimensional acoustic environment around the listener. Also, this kind of a system is helpful for composers who wish to add spatial effects to natural or synthetic music.

The 3-D sound modeling in general is useful in multimedia applications, e.g., in audiovisual user interfaces, and computer games. It could also be used for processing of sound for commercial multichannel reproduction systems, e.g., in movie and home theater systems.

The auralization part of the DIVA system consists of different parts which all affect the observed quality of the audio output. Therefore listening tests should be carried out to tell which parts of the system mostly require enhancements. The factors that should be tested include, e.g., the number of auralized image sources, the accuracy of absorption material filtering, and the directivity modeling of the image sources.

7. REFERENCES

[1] T. Takala, R. Hänninen, V. Välimäki, L. Savioja, J. Huopaniemi, T. Huotilainen, and M. Karjalainen. "An integrated system for virtual audio reality," *Audio Eng. Soc. 100th Conv.*, Copenhagen, Denmark, May 1996, preprint 4229 (M4).

[2] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen. "Virtual environment simulation - Advances in the DIVA project," *Proc. Int. Conf. Auditory Display (ICAD'97)*, Palo Alto, California, USA, Nov. 1997, pp. 43-46.

[3] J. B. Allen and D. A. Berkley. "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, No. 4, pp. 943-950, Apr. 1979.

[4] J. Borish. "Extension of the image model to arbitrary polyhedra," *J. Acoust. Soc. Am.*, vol. 75, No. 6, pp. 1827-1836, Jun. 1984.

[5] V. Pulkki. "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, Vol. 45, No. 6, pp. 456-466, Jun. 1997.

[6] M. Karjalainen, J. Huopaniemi, and V. Välimäki, "Direction-dependent physical modeling of musical instruments," *Proc. of the 15th I.C.A.*, Trondheim, Norway, Jun. 1995, pp. 451-455.

[7] J. Huopaniemi, L. Savioja, and M. Karjalainen. "Modeling of reflections and air absorption in acoustical spaces - A digital filter design approach," *IEEE 1997 Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk, New Paltz, New York, Oct. 1997.

[8] J. Huopaniemi and M. Karjalainen, "Review of digital filter design and implementation methods for 3-D sound," *Audio Eng. Soc. 102th Conv.*, Munich, Germany, Mar. 1997, preprint 4461.

[9] J.-M. Jot, V. Larcher, and O. Warusfel. "Digital signal processing issues in the context of binaural and transaural stereophony,"*Audio Eng. Soc. 98th Conv.*, Paris, France, Feb. 1995, preprint 3980.

[10] T. Laakso, V. Välimäki, M. Karjalainen, and U. Laine. "Splitting the unit delay - tools for fractional delay filter design," *IEEE Signal Processing Magazine*, vol. 13, No. 1, pp. 30-60, Jan. 1996.

[11] M. R. Schroeder. "Natural-sounding artificial reverberation," *J. Audio Eng. Soc.*, vol. 10, No. 3, pp. 219-223, 1962.

[12] J. A. Moorer. "About this reverberation business," *Comp. Music J.*, vol. 3, No. 2, pp. 13-28, 1979.

[13] J.-M. Jot. "Etude et realisation d'un spatialisateur de sons par modeles physique et perceptifs," *PhD thesis, l'Ecole Nationale Superieure des Telecommunications,* Sep. 1992, Telecom Paris 92 E 019, 221 p.

[14] M. Barron. "The subjective effects of first reflections in concert halls - Need for lateral reflections," *J. Sound and Vibration*, vol. 15, pp. 211-232, 1971.

[15] R. Väänänen, V. Välimäki, J. Huopaniemi, and M. Karjalainen. "Efficient and parametric reverberator for room acoustics modeling," *The Proc. of ICMC'97*, Thessaloniki, Greece, Sep. 1997, pp. 200-203.

[16] D. Rocchesso and J. O. Smith. "Circulant and elliptic feedback delay networks for artificial reverberation," *IEEE Trans. Speech and Audio Processing*, vol. 5, No. 1, pp. 51-63, Jan. 1997.