EFFICIENT GEOMETRY-BASED SOUND REVERBERATION

Augusto Sarti *Stefano Tubaro*
Dip. di Elettronica e Informazione – Politecnico di Milano,
Piazza Leonardo Da Vinci 32, 20133 Milano, Italy
e-mail: Augusto.Sarti/Stefano.Tubaro@polimi.it

ABSTRACT

In this paper we propose a novel approach to sound reverberation based on the the geometric analysis of the acoustic environment, which allows the listener to freely move within it. The method is based on a combination of a tapped delay line for early reverberation, designed through beam tracing; and a wave digital network for late reverberation, designed through path tracing. The method is efficient enough to enable a real-time sound rendering for Virtal Reality applications, on low-cost PC platforms.

1 Introduction

When the goal of a sound spatialization technique is not just to obtain plausible reverberations, but to achieve an audio-realistic rendering of the impact of a given environment, the computational complexity may easily become an issue. Advanced rendering techniques are, in fact, aimed at creating a sense of presence by enabling a certain auditory comprehension of the proportions and the geometry of the surrounding space. In order to achieve this goal, sound rendering techniques are usually based either on numerical methods (finite elements, boundary elements and finite differences) or geometrical methods (image source, path tracing, beam tracing and radiosity). However, only some of the geometrical methods seem adequate for low-cost implementations. In this Paper we propose a novel geometrical approach to sound spatialization that is suitable for simulating the acoustic behavior of large environments of complex shape, whose reflectors and sources are fixed while the receiver is free to move. The resulting numerical structure has the advantage of satisfying a number of requirements and needs: low computational cost, infinite impulse response, moving receiver, zero latency, both specular and lambertian reflection. The overall approach is based on a preliminary geometrical analysis of the spatialization environment, which produces the parameters for the auralization process (see Fig. 1).

The auralization block is made of two functional blocks (see Fig. 2), the former is a tapped delay line, which accounts for early reflections, and the latter is a

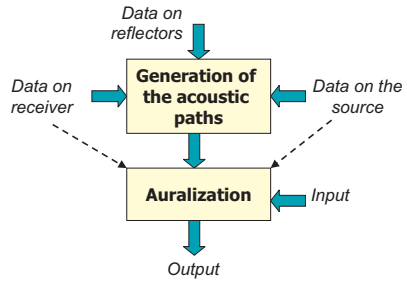

Figure 1: Overall approach to geometry-based sound spatialization. The generation of the acoustic paths is based on a geometric analysis of the environment. The auralization process uses the results of this analysis to render the sound correctly.

Waveguide Digital Network (WDN), which models late reverberation. This choice arises from the fact that the two individual blocks are suitable for a correct simulation of rather complementary aspects of the environment's acoustics, therefore they are best used in a joint and specialized fashion.

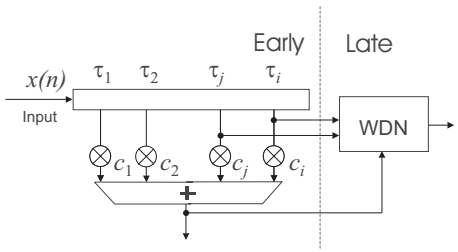

Figure 2: Structure of the our spatialization algorithm. The coefficients c account for the attenuation due to distance.

The conceptually simplest way to simulate specular reflections in a virtual environment in a geometric fashion is through tapped delay lines [3]. Such solutions, however, are characterized by modest computational requirements only for early reflections. In fact, if we look at the echogram (a graphical representation of the temporal distribution of the echoes) associated to an envi-

ronment of arbitrary polyhedral shape, we notice that its density grows with the square power of the time [6]. In particular, if we used the image source method [1] to compute the structure of the tapped delay line, the number of virtual sources

$$N_{\text{vir}} = \frac{n_w}{n_w - 2}\left[(n_w - 1)^{\epsilon} - 1\right] \quad ,$$

needed to account for early reflections would increase exponentially with the maximum order of reflection [6] ϵ, while the number of visible sources would be

$$N_{\text{vis}} = \frac{4\pi\,(ct_{\max})^3}{3V} \quad , \tag{1}$$

where n_w is the number of walls, V the room's volume and $t_{\max}$ the actual duration of the impulse response. Notice that N_{vis} grows with the third power of $t_{\max}$ and the ratio $\frac{N_{\text{vir}}}{N_{\text{vis}}}$ increases with ϵ, which makes this technique very inefficient. This complexity limitation forces us to use tapped delay lines only for the initial transient (early reverberations), while for the late reverberation, we need to adopt an alternative solution.

Our approach to the modeling of late reverberation is based on a Waveguide Digital Network (WDN) [8], which is a numerical structure that implements IIR filters through an interconnection of two-directional delay lines through scattering matrices. A key aspect of our solution to the spatialization problem is in the synthesis of the two algorithmic blocks, which is done using a fully geometrical approach. In fact, we use a beam-tracing approach [2] to synthesize a *dynamical* tapped delay line for the modeling of early reflections with moving receivers. The geometric analysis of the environment for the synthesis of the WDN that models the late reverberation is based on a path-tracing algorithm. In fact, we generate a scattering matrix for each wall node and assign appropriate lengths to the delay lines on the basis of the propagation geometry.

2 Early reflections

As already said above, early reflections ($\epsilon \leq \epsilon_m = 2$, typically) are modeled with a tapped delay line [3] with dynamic structure [4]. In order to account for both delay and attenuation associated to wave propagation, each tap is connected to a delay line of $\frac{L}{c} f_c$ samples and weighted by a factor $\frac{1}{L}$, where f_c is the sampling frequency, c the speed of sound and L is the length of the acoustic path. As the receiver is free to move, and L is an integer, an appropriate interpolator filter needs to be introduced (see [4]) in order to enable continuous variations of the length of the delay line without introducing sound artifacts.

The construction of a tapped delay line that correctly models early reflections in an environment of assigned shape, requires its preliminary geometric analyisis. In order to do so, we adopted a method that generalizes the image source method to compute the first reflections, as it works with *beams* rather than rays [2]. In fact, early reverberation are still characterized by spherical waves reflected by flat surface elements (see Fig. 3). The first step is to analyze geometric information on source and reflectors using a beam tracing algorithm (see Fig. 4). Using this beam tree and the information on the receiver, we can then generate the acoustic paths by checks visibility conditions through a simple beam-tree lookup process. This also allows the receiver to moves, as it is quite easy to update length and topology of the delay lines according to the "visible" beams. The TDL-based auralization process uses this information to process sounds.

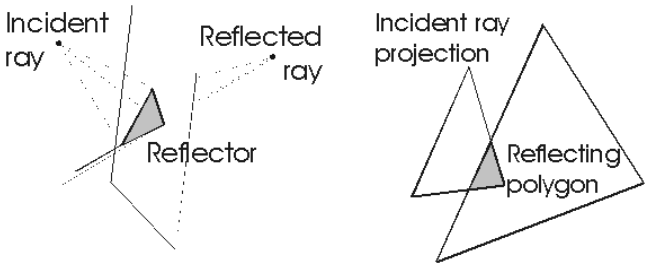

Figure 3: Pyramidal beams: early reverberations are modeled as the reflection of spherical waves on triangular surface elements.

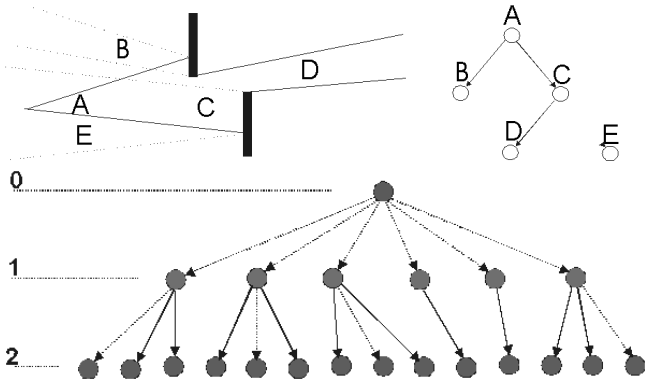

Figure 4: Beam tree construction process.

3 Late reverberation

Eq. 1 shows another interesting problem related to the distribution of the echoes: their density increases with the square power of the time while their energy decreases due to distance and reflection losses. The computational cost associated with their modeling is the same in both cases of direct path and early reflections, but their contribution to the impulse response tends to dim down rather quickly. Our approach to sound rendering, on the other hand, keeps the transient (early reflections) separate from late reverberations, which are simulated

by a Waveguide Digital Network (WDN) [8]. The samples processed by the WDN are those that correspond to the reflections of order $\epsilon > \epsilon_m$ and the WDN's output is mixed with the output of the tapped delay line (see Fig. 2).

Notice that, adopting a WDN structure means assuming that late reverberation are caused by planar waves. This, however, is a very reasonable assumption as a spherical wave that has already been reflected $\epsilon > \epsilon_m$ times, has traveled a long way already. The corresponding beam will thus be described as a generalized cylinder rather than a generalized cone.

The network topology and the scattering coefficients of a WDN are computed by sampling the reflectors at particular fixed points named "wall nodes". Each wall node corresponds to a scattering matrix, while the digital lines connected to it simulate the propagation of planar waves. Scattering coefficients of a wall node are computed by taking into account the acoustic characteristics of the wall, and are organized into a square matrix. The values of the samples injected into the N delay lines connected to a generic node k can be represented by a vector of the form $\mathbf{P}_k^- = \left[\begin{array}{cccc} p_{1k}^- & p_{2k}^- & \dots & p_{Nk}^- \end{array} \right]^T$. The input/output relationship $\mathbf{P}^- = \mathbf{A}\mathbf{P}^+$ expresses the outgoing pressure vector $\mathbf{P}^-$ as a function of the incoming pressure vector $\mathbf{P}^+$ through the nodal scattering matrix $\mathbf{A}$.

Before injecting audio samples into the numerical structure, the WDN needs to be correctly modeled. This can be done through a geometrical analysis of the environment based on a path-tracing algorithm. In fact, we generate a scattering matrix for each node and assign appropriate lengths to the delay lines on the basis of the propagation geometry.

In order to model the WDN, all the beams corresponding to the maximum reflection order ϵ_m are extracted from the beam-tree and projected onto each wall element k. A visibility test is conducted for each wall node in order to determine which virtual source will feed samples to that node. At this point, the reflected beam is treated as a cylindrical one and a path tracing approach can be adopted to characterize the following reflections.

Let us consider a generic planar wave going from the wall node i to wall node k (in the direction $\mathbf{s}_{ki}$). We would like to select, among the other available wall nodes, the one that lies in the most "specular position" with respect to the considered incident wave. This will the node j such that the direction of propagation $\mathbf{s}_{jk}$ is the closest to $\mathbf{s}_{ki} - 2\left(\mathbf{s}_{ki} \cdot \mathbf{n}\right)\mathbf{n}$, where $\mathbf{n}$ is the normal to the reflecting wall described by node k. Notice that $\cos\theta = \mathbf{s}_{ki} \cdot \mathbf{n}$, defines the angle of incidence of the acoustic path. If Z is the normalized acoustic impedance of the wall, the scattering matrix coefficient $\left(r_{ij}\right)_k$ can be computed as

$$\left(r_{ij}\right)_k = \frac{Z\cos\theta - 1}{Z\cos\theta + 1} \quad , \tag{2}$$

while the other elements of the same column of the scattering matrix $\mathbf{A}_k$ will be set to zero.

As a general rule, the greater the number of wall nodes, the better the approximation. Once determined the node in *quasi-specular* position and the corresponding reflection coefficient $\left(r_{ij}\right)_k$, we can iterate the procedure for the node j (next reflection of the path). Indeed, the iteration will continue until we find a path where a delay line already exists. The receiver R can now be connected to a wall node following the same procedure described above.

It is worth noticing that the energy of the impulse response does not increase with the number of wall nodes chosen for the modeling. In fact, with a greater number of nodes, we will have thinner beams, which reduces the likelyhood for the receiver to be invested by a beam.

4 Examples of application

In order to test the effectiveness of the proposed sound rendering solution, we modeled a variety of environments of different sizes and shapes. We tested the results with both static and moving receivers with very realistic results. The echograms reported in this Section correspond to the church-shaped environment of Fig. 5. Fig. 6 shows the sonogram corresponding to a simulation based on the image-source method with a maximum order of reflection $\epsilon_M = 8$ (order of reflection below which the acoustic path is still audible). Fig. 7 shows the simulation results with different numbers of wall nodes and order of reflection ϵ_M. Notice that, given a number of wall nodes n, the computational cost remains approximately the same regardless of the shape of the environment. Moreover, the subjective quality does not depend significantly on n, which allows us to use just one node per wall. On the other hand, the subjective quality improves significantly with the maximum order of reflection ϵ_M, but a reasonable choice for good perceptive results is $\epsilon_M = 5$ or 6. In fact, higher values would only increase the computational complexity. In all the considered cases, however, the computational cost was kept under or about 1000 multiplications per sample with a good perceptual quality, making the structure of Fig. 2 adequate for real-time, low-cost, sound rendering applications.

Notice that the average energy of the arriving echoes is the same in the simulation as in the ideal case. In fact, although their amplitude may look different, there is a certain difference in the density of the arriving echoes.

5 Conclusions

In this paper a new 3D sound rendering technique based on geometrical acoustics was proposed. Early reflections were estimated by the classic Image Method, while

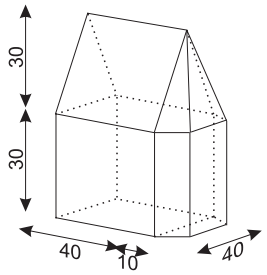

Figure 5: Geometric structure of the test environment (all lengths are in meters).

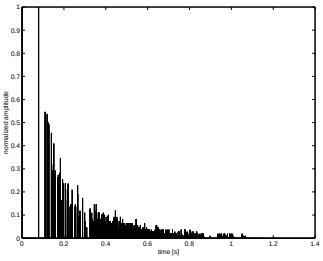

Figure 6: Ideal sonogram computed using the image source method.

late reverberation was simulated by a Wave Digital Network. The listener had the chance to freely move in a virtual environment by means of a low computational cost algorithm that exploits the physical properties of the room. We are currently working on a new version of this method, which allows the sound source to move within the environment, and incorporates the modeling of diffraction.

References

[1] J. Borish, "Extension of the image model to arbitrary polyhedra". *J. of the Ac. Soc. of America*, Vol. 75, No. 6, June 1984, pp 1827-1836.

[2] T. Funkhouser, I. Carlbom, G. Elko, G. Pingali, M. Sondhi, J. West, "A Beam Tracing Approach to Acoustic Modeling for Interactive Virtual Environments". Computer Graphics (SIGGRAPH '98), Orlando, FL, July 1998, pp. 21-32.

[3] W. Gardner, "Reverberation algorithms", in *Applications of digital signal processing to audio and acoustics*, edited by M. Kahrs and K. Brandenburg, Kluwer, 1998, pp.85-131.

[4] T. I. Laakso, V. Välimäki, M. Karjalainen, U. Laine: "Tools for fractional delay filter design". *IEEE Sig. Proc. Magazine*, Jan. 1996, pp. 30-60.

[5] P. Krämer, H. Alphei, A. Kohlrausch, D. Püschel, "Simulation von Nachhallzeit und Energie-Verlauf mit einem neuen nichtstatistischen Computermodell". *Acustica*, Vol. 75, 1992, pp. 233-245.

[6] H. Kuttruff, "Room Acoustics". Elsevier Applied Science, 3^{rd} ed. 1991, p. 92.

[7] H. Kuttruff, "Simulierte Nachhallkurven in Rechteckräumen mit diffusem Schallfeld". *Acustica*, Vol. 25, 1971, pp. 333-342.

[8] J.O. Smith, "Principles of digital waveguide models of musical instruments", in *Applications of digital signal processing to audio and acoustics*, edited by M. Kahrs and K. Brandenburg, Kluwer, 1998, pp. 417-466.

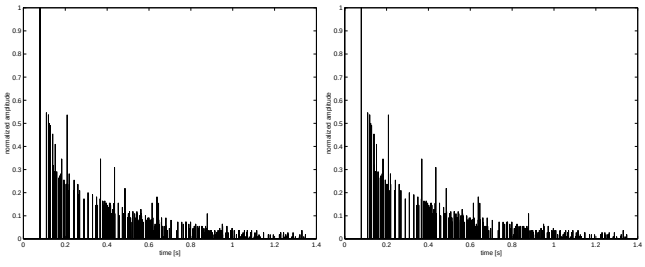

Figure 7: Simulation results with a total of 14 wall nodes. Top: $\epsilon = 2$, with a computational complexity of 404 mult./sample. Bottom: $\epsilon = 3$, with a computational complexity of 695 mult./sample

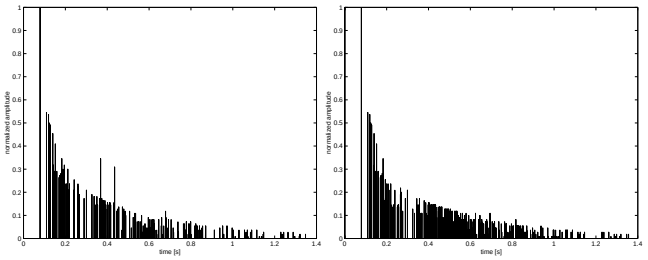

Figure 8: Simulation results with a total of 23 wall nodes. Top: $\epsilon = 2$, with a computational complexity of 594 mult./sample. Bottom: $\epsilon = 3$, with a computational complexity of 1049 mult./samples.

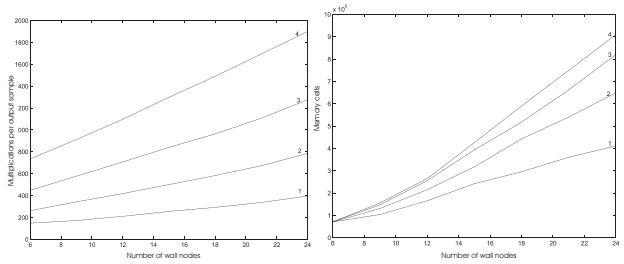

Figure 9: Number of multiplications per sample (left), and memory requirements (right) as a function of the number n wall nodes and of ϵ.