

Roberto A. Tenenbaum

Emeritus Member, ABCM

tenenbaum@iprj.uerj.br

Mech. Eng. Dept., Instituto Politécnico

Universidade Estadual do Rio de Janeiro - UERJ

28630-050, Nova Friburgo, RJ, Brazil

Thiago S. Camilo

thiagosc@petrobras.com.br;

Engenharia de Logística, Logística de Gás

Natural, PETROBRÁS

Julio Cesar B. Torres

juliotorres@ufrj.br

Graph. Eng. Dept., Escola Politécnica

Universidade Federal do Rio de Janeiro - UFRJ

21949-900 Rio de Janeiro, RJ, Brazil

Samir N. Y. Gerges

Emeritus Member, ABCM

samir@emc.ufsc.br

Mech. Eng. Dept

Universidade Federal de Santa Catarina - UFSC

Hybrid Method for Numerical Simulation of Room Acoustics with Auralization: Part 1 – Theoretical and Numerical Aspects

A new hybrid simulation method for room acoustics is presented. It is grounded on two well known numerical simulation techniques for room acoustics. The hybrid method takes geometric acoustics for granted and uses an improved version of the classical ray-tracing technique for computing the specular reflections and a slightly modified energy transition method for simulating the diffuse reflections. The impulse responses (IRs) are then computed by superposition of the specular and diffuse ones. The reverberant part of the IRs gains a much more realistic aspect than the one obtained from either ones method alone. Both the ray-tracing technique and the energy transition method are reviewed and some improved features are discussed. The sound source and the receiver modeling are presented, including a new method for simulating the head related transfer functions (HRTFs), and based on the wavelet technique, which provides a shorter time computation of the binaural impulse responses for auralization purposes. In a companion paper, entitled Hybrid method for numerical simulation of room acoustics: Part 2 – Validation of the computational code RAIOS 3, a validation of the method in an international inter-comparison with other softwares and measurements data is presented, showing the efficiency and the accuracy of the proposed hybrid method over non-hybrid ones.

Keywords: computer modeling, room acoustics, hybrid method, auralization

Introduction

Due to the advance of computer technologies and the improvement of programming languages computers are, nowadays, capable to simulate a large variety of sound wave propagation effects. Numerical methods transport the physical reality to the computer realm and are one of the powerful design tools, available for architects, engineers, acousticians and so on. The most popular techniques to simulate room acoustics are the ray-tracing method (Kulowski, 1984; Embrechts, 1982), the image or virtual source method (Dance *et al.*, 1997), the cone-tracing or pyramid-tracing method (Farina, 1995; Stephenson, 1996; Farina, 2000), the free path distribution method (Vorländer, 2000b), and the energy transition method (Kruzins *et al.*, 1982; Alarcão *et al.*, 2000).

The acoustics of a room involves, simultaneously, distinct sound propagation phenomena, such as reflection, absorption, diffusion, diffraction, damping and interference, among others. Numerical simulation using a technique that considers only part of the phenomena will not represent the correct behavior of the room acoustics. Numerical experiments combining different phenomena, called hybrid methods, have been published (Vorländer, 1989) and are a main trend in this area. In addition, the numerical simulation should include adequate techniques to deal with the sound source emission, boundary room surfaces and sound reception modeling, among others, including the directional characteristics of the human hearing if one of the required outputs is the room auralization (Blauert, 1997).

Numerical simulation should be correctly applied to the physical reality and, in the other hand, must be able to run under the available computational capacity. The sound source modeling, for example,

should incorporate the power spectrum density, but usually the available information about spectral density is discretized in octave bands. The scattering phenomenon — which sometimes is still called as diffusion — has received more attention and research since subjective experiments pointed out a large influence of this phenomenon in the acoustical quality of a room (Davies *et al.*, 2001; D'Antonio *et al.*, 1998a; D'Antonio *et al.*, 1998b). A lot of research have been done to understand and to model the scattering phenomenon (Haan *et al.*, 1993; Torres *et al.*, 2001; D'Antonio *et al.*, 2000). Nowadays, two international standards are available to define the scattering and the diffusion phenomena. The scattering is defined in ISO/WD 17487 (2001) where the scattering coefficient specifies the energy fraction spread over directions not reflected specularly. The scattering coefficient is based on the work of Vorländer and Mommertz (2000a) and has been widely used in room acoustics. The diffusion coefficient is defined by the international standard AES-4ID-2001 (2001) and states how uniform are the distribution of the spread energy (Gomes, 2002).

In this work, the simulation of sound propagation in a closed space is based on a new hybrid method, where the specular reflections are obtained by an improved ray-tracing technique and the diffuse reflections are computed by using a slightly modified energy transition method (Camilo, 2003). The main goal of combining these two methods is to generate a good estimation of the impulse response (IR) of a room, for a source-receiver position combination, from which the main acoustical parameters can be obtained as well as the auralization. Auralization is the process of rendering sound for acoustic simulation (Møller, 1992). In this context, the head-related transfer functions (HRTFs) are the most important issue when concerning receptor modeling for realistic simulation of human spatial hearing (Blauert, 1997). Besides the method employed to obtain the acoustical behavior of a room, any auralization system must include HRTF modeling. Such model must simulate the influence of the hearing subject in the acoustic environment.

The HRTF data consists of a set of directional transfer functions, between sound sources located in the space around the listener and the ear canals of a human (or dummy) head in an anechoic environment. Much research has been done in order to measure the HRTF or its time-domain counterpart, the head-related impulse response (HRIR) (Mellert *et al.*, 1977). The latest available HRTF data sets were obtained by the Massachusetts Institute of Technology (Martin *et al.*, 1995) and the University of California (Thompson *et al.*, 2001). The first laboratory measured these functions using a dummy head for large and small pinnae. The second one measured those over one hundred different subjects and dummy heads, for both large and small pinnae, in free-field conditions.

For auralization purposes, one of the main aspects to be considered is the huge data bank for the HRTFs and the heavy convolution process involved. A new method to model the HRTFs based on wavelet transforms and sparse filters, which reduces significantly the computational effort, will be discussed here.

The adequate superposition of the ray-tracing and energy-transition impulse responses leads to a global IR that provides reliable values for the room acoustics parameters and also a realistic auralization. The theory that underlines the proposed hybrid method is presented and some numerical aspects are discussed in the next sections. The software that implements the hybrid method, named RAIOS 3, is also commented.

Computer simulation of room acoustics, not being a trivial task because of the great number of phenomena to be taken into account, must, above all, be validated by reliable measurements in a variety of rooms with distinct acoustical characteristics. For the same room, different algorithms may lead to unlike results. Of course, looking to the impulse responses themselves, not too much information about the differences can be recognized; but the values obtained for the acoustical parameters computed from them — like T_{30} , EDT, C_{80} , and LF, and so on — pinpoints the differences. In a companion paper, *Hybrid method for numerical simulation of room acoustics: Part 2 – Validation of the computational code RAIOS 3*, an international inter-comparison among 21 teams working with 9 different program codes for room acoustics simulation and a validation with measurement data will be reported (Bork, 2005).

Ray-Tracing Method

In this classical method, the sound wave is modeled by geometrical acoustics — essentially the same model for geometrical optics — where the sound wave is assumed to propagate in straight sound rays (Kuttruff, 2000) emitted from the sound source. Each one of these rays carries information about the power spectrum and the traveled sound wave path. Propagating in a straight line, the ray is submitted to the dissipation effects caused by the air viscosity and by the room contour surfaces, which include phenomena such as surface absorption and specular reflection (Snell law). The part of the incident sound wave that is reflected in directions different from the specular one is called diffuse reflection but, actually, corresponds to a scattering phenomenon.

Source Modeling

Virtual directional sound sources can be obtained by emitting a large number of rays over all directions around the point where the center of sound source is located and filtering the power spectrum according to its directional characteristics. Many models have been proposed to obtain the maximum of homogeneity in the angular distribution of emitted rays, since there exist no closed solution, of course, for dividing a sphere in more than 20 exactly equal spaced directions. A method that proposes a randomly generated algorithm to provide ray directions was presented in (Embrechts, 1982).

Another method, dividing a sphere in octants, the octants in six triangles and each triangle being divided recursively, was published in (Tenenbaum *et al.*, 1992). Figure 1 illustrates the technique.

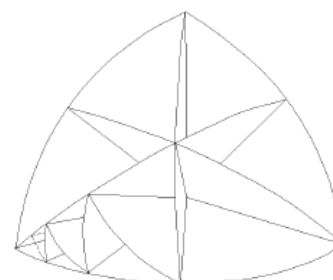


Figure 1. Subdivision of the octant into triangles of maximum area.

A source modeling that presents better results, having the homogeneity criterion in mind, is based on the regular icosahedrons. Each triangular face is divided in four equilateral triangles and the process continues recursively. Figure 2 illustrates the technique. After n subdivisions the number of rays, N_R , takes the simple form

$$N_R = 2 + 10 \cdot 4^n \quad (1)$$

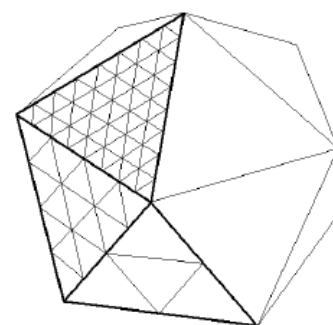


Figure 2. Subdivision of icosahedra surfaces, for ($n = 0, 1, 2, 3$).

Figure 3 shows the subdivision of one of the triangular faces for $n = 6$. It can be observed some rarefaction around the spherical triangle center, which indicates a certain lack of homogeneity.

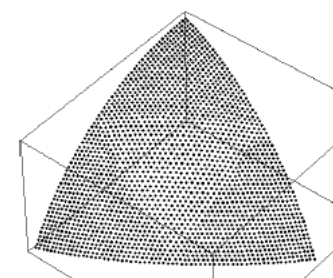


Figure 3. Source distribution for $n = 6$, and $N_R = 40,962$. See the rarefaction at the center.

Best results, having homogeneity criterion in mind, is obtained by geodesic subdivision of the regular icosahedrons. Each one of the original triangular faces is subdivided recursively into new regular triangles, as illustrated in Fig. 4.

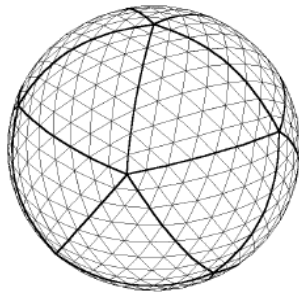


Figure 4. Geodesic triangles subdivision (n = 8).

The recursive algorithm is described by Lewer (1993). After n subdivisions the number of rays is given by

$$N_R = 2 + 10 \cdot n^2 \tag{2}$$

Figure 5 presents one of the triangles of the icosahedrons after 65 divisions leading to 42,252 directions, very close to an exact homogeneous distribution. The obtained homogeneity is better than the one of the previous methods and the number of rays obtained by Eq. (2) is better controlled than the number obtained by Eq. (1).

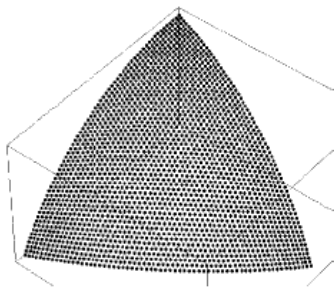


Figure 5. Source distribution for 65th order. Better homogeneity than cited models.

Modeling the Reflections

The model proposed by Vorländer and Mommertz (2000a) for the scattering provided by a surface is the basis of the international standard ISO/WD 17487 (2001). This model includes the scattering coefficient δ that quantifies the energy fraction reflected in a direction different from the specular, varying from 0 to 1, i.e., from a purely specular reflection to a totally diffuse one, respectively. However, it is not recommended to apply the ray-tracing method to simulate the spread of the reflected energy, since for each incident ray a very large number of directions would be emitted from the incidence point, leading to an explosive computational effort, representing too much usage of memory and cpu time. In practice, the application of the ray-tracing method is restricted to the specular reflections, assuming that all diffuse energy is also concentrated in the same ray, i.e., making $\delta = 0$. Here, the energy reduction due to the scattering effect is computed as follows.

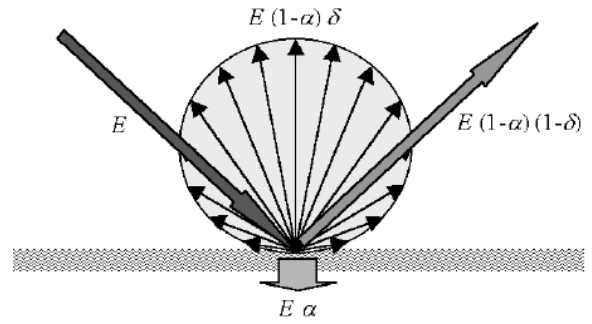


Figure 6. Energy distribution between absorbed, specularly reflected, and scattered rays.

In Fig. 6, E is the incident energy on the surface, α is its absorption coefficient, and δ is the scattering coefficient. So, the absorbed energy is $E\alpha$, the scattered energy is $E(1 - \alpha)\delta$, and the reflected energy is $E(1 - \alpha)(1 - \delta)$. The scattered energy is stored, together with the respective reflection time, for the diffusive part of the algorithm. In order to obtain an impulse response and the sound pressure distribution, the energy E_N carried by each ray after N reflections is given by

$$E_N = \frac{E_F}{N_R} D_{\theta\phi} e^{-\gamma d} \prod_i^N (1 - \alpha_i)(1 - \delta_i), \tag{3}$$

where: E_F is the total energy emitted by the sound source; N_R is the number of rays; $D_{\theta\phi}$ is the source directivity, a function of the spherical coordinates of azimuth θ and elevation ϕ ; γ is the air absorption coefficient; d is the distance path traveled by the ray; α_i is the absorption coefficient at surface i ; and δ_i is the scattering coefficient at the same surface. The parameters α_i and δ_i are considered as octave band dependents.

The specularly reflected ray direction can easily be computed using the simple vector relationship

$$\mathbf{r}_r = \mathbf{r}_i - 2(\mathbf{r}_i \cdot \mathbf{n})\mathbf{n}, \tag{4}$$

where the subscripts i and r states for incident and reflected, respectively, and \mathbf{n} is the normal to the boundary surface, pointing to the inside.

Surfaces Modeling

Many mathematical approaches can be applied to model the boundary surfaces. The simplest one and the most efficient to reduce the computational complexity are based on a non-regular surface partition into triangles, as shown in Fig. 7. It must be found the values of λ that satisfies

$$P_0 = \sum_{i=1}^3 \lambda_i P_i, \tag{5}$$

where P_0 is the intersection point of the ray with the surface and $P_i = (x_i, y_i)$; $i = 1, 2, 3$ are the triangle vertex. P_0 is an internal point if

$$\sum_{i=1}^3 \lambda_i \leq 1 \quad \text{and} \quad \lambda_i \geq 0. \tag{6}$$

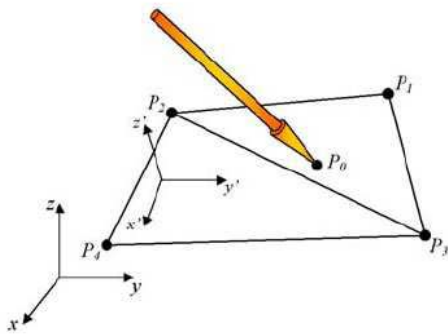


Figure 7. Non-regular division of a surface using triangles.

Since this check is very simple and fast, dividing all boundary surfaces into triangles implies that the computational effort is strongly reduced, even using many triangles for each plane. This is a very simple artifice that results in high computing time saving, being an improvement of the ray-tracing technique.

Receiver Modeling

The modeling technique used to simulate how a ray is captured by the receiver has a decisive importance in the method accuracy. In the real world, a wave front intercepts the microphone surface, considered very small, but in the ray-tracing model a ray vector must find a larger surface. This inversion in the mathematical treatment requires some care. Almost all models found in the literature for the receiver are based on a volumetric expansion around the point under consideration. In the reception algorithms, this expansion is usually spherical and the expression for the computation of the energy E_R retained by a spherical receiver is (Farina, 1995; Ballesteros et al., 1992)

$$E_R = \frac{L}{2R} E, \tag{7}$$

where L is the length of the ray portion crossing the receiver, R is the receiver radius, and E is the ray energy. If simultaneous rays cross the receiver, then an average is performed (Farina, 1995).

In spite of being widely used, Eq. (7) can furnish rather inaccurate results. This can be confirmed with a simple example, shown in Fig. 8, where two spherical concentric receivers R_1 and R_2 are being reached simultaneously by uniformly distributed acoustic rays, represented by the dots. Receiver R_2 is being crossed by nine rays while receiver R_1 captures only one. Using Eq. (7), it is easy to verify that the average energy retained by receiver R_1 , with radius a , is around 1.26 times the average energy retained by receiver R_2 , with radius $3a$.

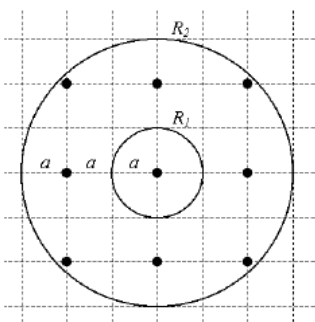


Figure 8. Example of two spherical receivers with radius a and $3a$. The dots represent acoustic rays.

A better alternative is to convert the sound energy to intensity. Instead of a solid, a circular plate is considered. The reception disk rotates around its center so that the acoustical ray is always orthogonal to it. The intensity at the receiver, I_R , at an instant t is

$$I_R = \sum_i E_i / \pi r^2, \tag{8}$$

where $\sum_i E_i$ is the sum of the powers of all incident rays at the same instant, and r is the reception disk geometric radius. With this simple stratagem, which poses no computational problem, the situation exemplified by Fig. 8, now two disks with radius a and $3a$, the same acoustic intensity is obtained for both receivers. This is another simple but efficient improvement to the ray-tracing method, introduced by our algorithm.

Final Considerations on Ray-Tracing Method

Studies made with musicians and experienced listeners demonstrated that the use of diffusers in a room increases the mutual hearing, the self hearing, the articulation, and the sound quality (Davies et al., 2001), being a desirable room feature, especially if music is concerned. So, it is definitely mandatory, for good prediction results, to have a good model for the scattering phenomenon at the boundary surfaces (Torres et al., 2001). The ray-tracing method is quite good to predict specular reflections, but may not present good results when the goal is the performance of auralization (Dalenbäck et al., 1994). This fact can be clearly understood by taking a look at Fig. 6: Each specular reflection generates a great number of scattered directions. If new acoustic rays are associated to the scattered directions, the exponential rays number increase will virtually explode the method. In the next sections, the general structure of a slightly modified energy transition method, which furnishes better results for highly diffusive rooms, is presented.

Energy Transition Method

The energy transition method (Alarcão et al., 2000; Dalenbäck et al., 1994; Bento Coelho et al., 2000) — a variation of the so-called random walk approach (Kruzins et al., 1982), or the radiosity method (Lam, 1996) — is based on the exchange of acoustic energy between the source and the surfaces of a room and among the surfaces themselves, assuming that this exchange happens at a regular time interval, equal to the room characteristic time,

$$\tau = \frac{4V}{cS}, \tag{9}$$

where V is the room volume, c is the sound propagation speed, and S is the global boundary surface. This interval is also called the room transition time and corresponds to the time spent by the sound wave to cover the free mean path,

$$l_m = c\tau = \frac{4V}{S}. \tag{10}$$

In other words, the room is assumed (as proposed by Sabine) as totally diffuse.

Source Modeling

The central idea of this method is that all surfaces irradiate sound to the remaining ones proportionally to the solid angle of the last with regard to the geometric center of the first. Also, the source irradiates sound to the boundary room surfaces proportionally to the

solid angle of the surface as seen from the source center, as illustrated in Fig. 9.

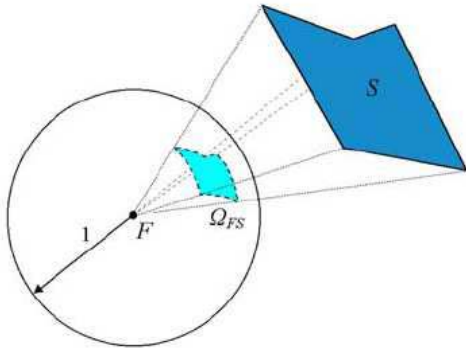


Figure 9. Solid angle Ω_{FS} that the sound source center F “sees” a surface S.

Therefore, the sound source distributes energy to the room surfaces and each surface irradiates its energy to the other ones, successively. The energy received by a surface, E_S , from the sound source is given by

$$E_s = E_F \frac{\Omega_{FS}}{4\pi} D_{\theta\theta} e^{-\gamma d}, \tag{11}$$

where E_F is the source energy and Ω_{FS} is the solid angle of the surface S related to sound source center F.

This is a statistical model for the sound propagation and once the energy is emitted from the sound source, the distribution over n surfaces can be represented by a line matrix in the form $\mathbf{E}_0 = (E_{01}, E_{02}, \dots, E_{0n})$, where E_{0i} is the initial energy received by the i th surface.

Surface Modeling

In the modified version of the energy transition method, the surfaces are divided into triangular elements — as much as the desired precision requires — so that the solid angles are computed with much more accuracy. Figure 10 illustrates the surface partition.

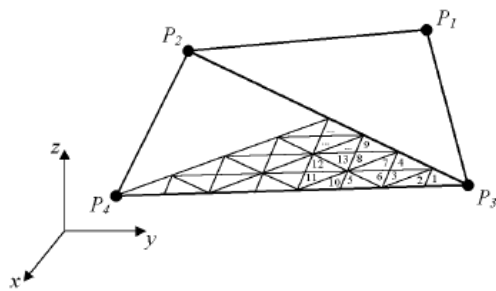


Figure 10. Partition of a surface into triangular elements.

When the surface's subdivision process finishes, the energy influence of a surface element to the other ones must be computed. Each surface element reflects the received energy to all visible elements in the room, emitted from its geometrical center. The algorithm used in the classical technique computes the solid angles without dividing the surfaces by, essentially, the scan-line method (Fonseca *et al.*, 1998). This algorithm obtains a very good precision with a long computer time. Using the surfaces partition technique,

the obtained values are not so precise, but they are sufficiently accurate with low computer time consuming.

The energy transition $\Psi_{i,j}$ from the i th element to other visible j th element is given by

$$\Psi_{i,j} = (1 - \alpha_i) \frac{\Omega_{ij}}{\pi} \sin \beta e^{-\gamma d_{ij}}, \tag{12}$$

where Ω_{ij} is the solid angle of the surface element j relative to the surface element i , β is the elevation angle of element j relative to the plane of element i , and d_{ij} is the distance between the elements. Arranging each energy element Ψ_{ij} in a square matrix Ψ , called energy transition matrix, the energy distribution over the surfaces after k iterations, can be obtained by

$$\mathbf{E}_k = \mathbf{E}_0 \Psi^k. \tag{13}$$

Hybrid Method

The proposed hybrid method uses a combination of the ray-tracing method — with the circular plate receiver, described in Section 2.4, and the non-regular surface partition, described in Section 2.3 — with the energy transition method — with the partition of surfaces into triangular elements, as described in Section 3.2 — to simulate the most important phenomena of sound propagation in rooms: air and surface absorption and specular and diffuse reflections. Usually, the ray-tracing method is very accurate to calculate the specular reflection, mainly the first reflections of a room (Vorländer, 1989). However, the method does not provide a precise simulation of the reverberant tail and does not furnish good prediction when the room is highly diffuse (Suh *et al.*, 1999). On the other hand, the energy transition method models with good accuracy the sound field in diffusive rooms, being efficient to simulate the reverberant tail. As a rule of thumb, the two methods have their accuracy varying roughly with the room average scattering coefficient, $\bar{\delta}$, in opposite directions, as illustrated in Fig. 11.

The proposed hybrid method applies the improved ray-tracing to model the specular reflections and the modified transition method to simulate the diffuse reflections. The final impulse response is obtained by just superposing both results.

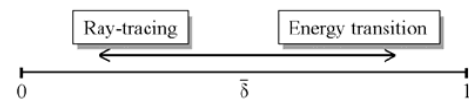


Figure 11. Accuracy dependence with the average room scattering coefficient.

Specular Reflection Processing

The impulse response generated by the proposed method is obtained in two steps. First, the specular reflections are processed and the necessary information to the diffuse processing (second phase) is stored. Each ray emitted from the sound source is submitted to the effects produced by its incidence on a surface element (subdivision). The energy spectrum of the incident ray is modified by the absorption and spreading characteristics of the surface element. The ray energy after N reflections is given by Eq. (3).

The accumulated energy in each surface element up to, say, the reverberation time, T_{60} , roughly estimated previously (by Eyring formula, for instance), is organized in a rectangular matrix \mathbf{E}_D with dimension $n \times T_{60}$.

$$E_D = \begin{pmatrix} E_{1,1} & E_{1,2} & E_{1,3} & \dots & E_{1,T60-2} & E_{1,T60-1} & E_{1,T60} \\ E_{2,1} & E_{2,2} & E_{2,3} & \dots & E_{2,T60-2} & E_{2,T60-1} & E_{2,T60} \\ E_{3,1} & E_{3,2} & E_{3,3} & \dots & E_{3,T60-2} & E_{3,T60-1} & E_{3,T60} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ E_{n-2,1} & E_{n-2,2} & E_{n-2,3} & \dots & E_{n-2,T60-2} & E_{n-2,T60-1} & E_{n-2,T60} \\ E_{n-1,1} & E_{n-1,2} & E_{n-1,3} & \dots & E_{n-1,T60-2} & E_{n-1,T60-1} & E_{n-1,T60} \\ E_{n,1} & E_{n,2} & E_{n,3} & \dots & E_{n,T60-2} & E_{n,T60-1} & E_{n,T60} \end{pmatrix} \quad (14)$$

where n is the total number of surface elements in the room.

For each specular reflection, the matrix E_D is updated in the respective position, according to the number of the element and to the incidence time. At the end of the ray-tracing computation, when all rays are processed, the matrix E_D represents an initial temporal and spatial distribution of the energy held in the surfaces. This energy will be reflected by the surface elements as diffuse energy. However, in this method, it's considered that the diffuse energy will not suffer new specular reflections. Therefore, by the end of the first step — specular reflection — a genuine specular impulse response $h_s(t)$ is obtained, by adding the lines of E_D .

Diffuse Processing

The second step of the proposed method consists on the diffuse reflections processing. Using the initial diffuse energy matrix E_D , obtained at the first step, the energy transitions between the surface elements are computed. The process starts emitting the energy of all n surface elements at the first column ($t = 1$) of E_D to all visible elements of the room, i.e., those with positive solid angles, according to Eq. (12). Since the matrix E_D is organized such that each column represents the energy of the elements at a discrete time, the energy from an element at column t will propagate to the right in the matrix and to different columns, according to the propagation time between the elements. Figure 12 illustrates the element $E_{3,1}$ of the matrix furnishing energy to other elements that have non-negative solid angles with regard to the 'source' element.

The energy transition between an element i , radiated at time t , $E_{i,t}$, to another element j that receives the energy at time $t + \tau$, $E_{j,t+\tau}$ where $\tau = d_{ij}/c$, is updated by the recursive relationship

$$E_{j,t+\tau} \leftarrow E_{j,t+\tau} + E_{i,t} \cdot \Psi_{i,j}, \quad \text{for } i = 1, 2, \dots \quad (15)$$

where Ψ_{ij} is the energy transition, given by Eq. (12).

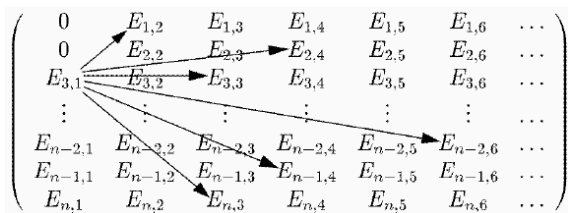


Figure 12. Energy transition in the diffuse matrix.

Changing successively the matrix E_D from time $t = 1$ to, say, T_{60} and using the iteration process described in Eq. (15), the diffuse impulse response, $h_d(t)$, is obtained by adding the accumulated energy at each time column. The global room impulse response, $h(t)$, will be obtained then by superposing the specular and diffuse impulse responses, as

$$h(t) = h_s(t) + h_d(t). \quad (16)$$

The energy transition from each surface element to the receiver must, of course, be computed. At each time interval, all visible surface elements from the receiver location will contribute with some energy to it, given by

$$I_R(t+t') = \frac{E_{i,t}}{\pi d_{iR}^2} \sin \beta, \quad (17)$$

where $I_R(t+t')$ is the intensity in the receiver at instant $t+t'$, with t' given by $t' = d_{iR}/c$ and d_{iR} being the distance between the center of element i and the receiver center.

Auralization

The HRTF modeling is one of the most important issues for a realistic human hearing simulation. In systems with real-time auralization output or real-time updating, such as video-games, cinema sound effects and even in some room acoustic simulators, the complexity of such functions needs to be drastically reduced both in time and in frequency resolution. This reduction results in a simplified simulation of the acoustical environment, limited to simple effects of source positioning or crude reverberation effects. Such limitation is due to the large number of directions and the resolution (time and frequency) required to produce the human sound impression and, consequently, the perception of a realistic acoustical environment. The main problem for direct use of HRTF data resides in the large number of functions and in their large lengths. Human hearing has an angular resolution, called localization blur, which varies from 5 to 20 degrees, depending on the direction (Blauert, 1997). To keep this precision, a set of HRTFs must have approximately 1,400 functions, one for each direction and for each ear. In the auralization system proposed here, it was taken the MIT Media Lab database (Martin *et al.*, 1995), consisting of 710 directional transfer functions for each ear, stored in a temporal data bank (Torres *et al.*, 2004). Each one of these functions corresponds to a FIR filter that, after applying the reduction process explained below in Subsection 5.2, has approximately 30 coefficients for each direction.

The HRTFs combined with the acoustic rays are used to produce binaural impulse responses (BIRs), which can be convolved with an anechoic signal to generate a binaural sound, which should correspond to that one heard in the real environment. However, if the number of coefficients used to generate these BIRs is not reduced, an auralization system becomes extremely time consuming, due to the large number of incoming acoustic rays and since each ray must be convolved with the corresponding HRTF.

Some new approaches have been proposed in order to reduce the representative data of HRTFs (Kistler *et al.*, 1992; Haneda *et al.*, 1999; Keränen *et al.*, 2001; Georgiou *et al.*, 1999), but the use of such models in systems with auralization still has many drawbacks. One of them is the HRTF implementation which, even if greatly compressed, requires a large computational cost to uncompress or recover the data. Another negative aspect is that the reduction in the number of coefficients may remove some essential information, necessary for a realistic perception of the source direction, and in the auralization case, produces an unrealistic acoustical environment simulation.

HRTF Modeling Using Wavelets

In this section it will be described how the HRTF can be modeled using a multi-stage decomposition structure with wavelets and sparse filters (Vaidyanathan *et al.*, 1993; Torres *et al.*, 2002). The discrete wavelet transform in the Z domain can be obtained by properly choosing the coefficients of prototype filters $H^0(z)$ and $H^1(z)$, and cascading them as in the structure (Vaidyanathan *et al.*, 1993) shown in Fig. 13a.

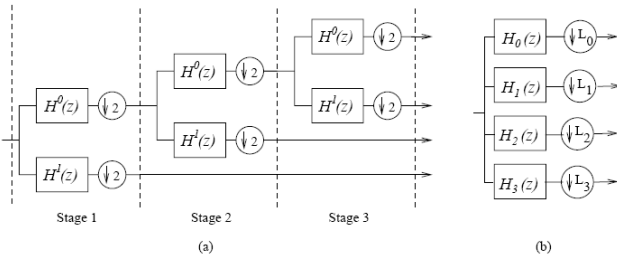


Figure 13. Wavelet implementation (a) by multi-stage decomposition structure and (b) by the equivalent filter bank.

In such case, the corresponding equivalent analysis filters $H_m(z)$ of Fig. 13b are related to the two-band low-pass and high-pass analysis filters, $H^0(z)$ and $H^1(z)$ of Fig. 13a, by

$$H_0(z) = \prod_{k=0}^{J-1} H^0(z^{2^k}), \quad (18)$$

and

$$H_m(z) = H^1(z^{2^{J-m}}) \prod_{k=0}^{J-1-m} H^0(z^{2^k}), \quad (19)$$

for $m = 1, \dots, M-1$, where $M = J+1$ is the number of sub-bands and J is the number of stages of the multi-resolution decomposition.

The structure shown in Fig. 14 is able to model any Finite Impulse Response (FIR) system (Torres *et al.*, 2000). Since the HRIRs are functions with finite time decay, they can be viewed as FIR systems. In this approach, the HRIR (FIR) coefficients are modeled by a set of sparse filters $G_m(z^{L_m})$, for $m = 0, 1, \dots, M-1$, which process the outputs of the wavelet transform. Sparse filters are FIR filters whose nonzero coefficients are separated by $L_m - 1$ zeros, where L_m is the sparsity factor of the sub-band m . Their frequency responses are repeated L_m times over the normalized frequency range from 0 to 2π . The sparsity factors L_m of the sub-filters correspond to the decimation factors of the equivalent filter bank, Fig. 14b, which are given by

$$L_m = \begin{cases} 2^J, & m = 0 \\ 2^{J-m+1}, & m = 1, \dots, J \end{cases} \quad (20)$$

The impulse response $r(n)$ of the resulting structure can be obtained by adding the convolutions of the analysis filters $h_m(n)$ with the sparse filters $g_m(n)$, as

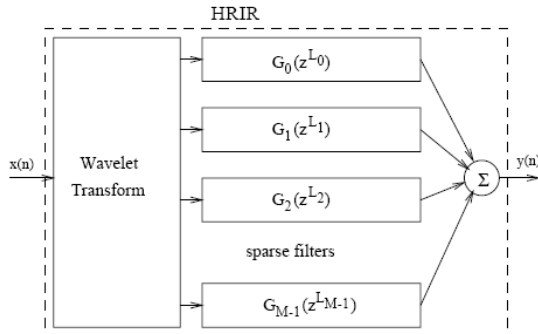


Figure 14. Sparse Structure for implementing a FIR system.

$$r(n) = \sum_{m=0}^{M-1} h_m(n) * g_m(n) \quad (21)$$

The wavelet implementation depends on two elements: The wavelet family and the cascading arrangement of those filters. The wavelet family is a set of low-pass and high-pass filters with the same properties, used in the decomposition structure of Fig. 13. The wavelet used in the auralization method is the Daubechies (1990) of length 4. The decomposition structure applied to the auralization is such that the low-pass frequencies are decomposed over each stage, resulting in a better resolution at lower frequencies, and the frequency components over 11 kHz are concentrated in only one sub-band. This structure is more appropriate since the human hearing system bandwidths are roughly logarithmic (with high spectral resolution at low frequencies and low spectral resolution at high frequencies).

Although arbitrarily fine low-frequency resolution can be obtained by increasing the number of decomposition stages, there is a corresponding increase in the computational load. Furthermore, it is not necessary to have extremely fine low-frequency resolution for HRTF modeling, because its behavior at low frequencies is fairly simple, with the phase response (which leads to the interaural time difference) being the dominant factor (Wightman *et al.*, 1992). Based on the above comments, the structure used to model the HRTF in the auralization process was selected with five sub-bands (Fig. 13 with $J = 4$).

Reduction of the HRTF Model

In order to reduce the size of the model, we first investigate the lengths of the HRIRs provided by MIT. Each of them has originally 512 coefficients. Removing the initial delay, their lengths reduce to approximately 450 samples, depending of the source direction. This length is still very large to be used in an efficient system and can be shortened to 100 samples, using an energy criterion (Torres *et al.*, 2004). Using the first 100 coefficients (after removing initial null values), more than 90% of the total impulse response energy is preserved. The worst case occurs for azimuth angles between 250° and 300° , where the decay of the impulse response amplitude is the slowest. In such case, using the first 100 coefficients, 90% to 95% of the impulse response energy is kept. For the directions in the same side of the sound source, almost 99% of the energy is held.

The criterion used to define how many coefficients are important in each sub-band is based on the HRTF frequency resolution properties (magnitude and phase), energy sub-band losses and human hearing characteristics (Kulkarni *et al.*, 1998). The HRIR energy $E(\phi, \theta)$, where ϕ and θ are the elevation and the azimuth angles, respectively, is given by

$$E(\phi, \theta) = \sum_{n=0}^{N-1} p_{\phi, \theta}^2(n), \quad (22)$$

where N is the length of the HRIR $p_{\phi, \theta}(n)$.

The HRIR energy per sub-band is given by

$$E_m(\phi, \theta) = \sum_{k=0}^{K_m-1} g_{m,k}^2(\phi, \theta), \quad (23)$$

where K_m is the number of sparse coefficients of sub-band m .

The energy contributions of the sparse sub-filters coefficients to the total HRIR energy, obtained varying the number of coefficients from 1 to K_m in Eq. (23), are shown in Fig. 15, using the right ear measurements and the Daub4 prototype for direction $\phi = 0^\circ$ and $\theta = 90^\circ$.

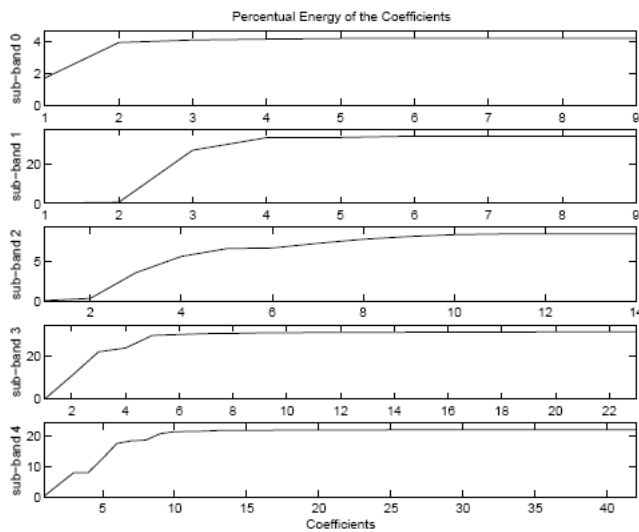


Figure 15. Energy contribution of the sparse coefficients with Daub4 prototype for 0° of elevation and 90° of azimuth.

From Fig. 15, we verify that after a certain number of coefficients, the introduction of another coefficient in the energy calculus does not contribute effectively to the HRIR sub-band energy. Based on that, the energy contributions of the sparse sub-filters coefficients in each sub-band were computed for all HRTFs of the data bank and the “relevant” coefficients were selected. The “relevant” coefficients will be kept in the reduced model and the remaining ones will be removed (not processed).

These windowing processes are applied to the HRTFs for all directions. Since the windowing intervals are slightly different between directions, we can not ensure that all HRIR models have the same energy loss. It can be higher or lower, depending on the direction of the HRTF which is being modeled, but the mean loss does not vary too much. The mean energy loss selected to reduce the number of coefficients of the modeled HRTFs was 10%.

These coefficients intervals are applied to the HRIR, for both ears, at direction $\phi = 0^\circ$ and $\theta = 90^\circ$ and for prototype Daub4. The respective magnitude and phase frequency responses are shown in Figs. 16 to 19. In these figures, the original HRTFs frequency responses are compared to the reduced ones. Notice that, for the same number of coefficients, the energy loss in direct form is always higher than using the wavelet model.

The Auralization System

The auralization is obtained by processing an anechoic sound signal $x(n)$ through the structure presented in Fig. 20.

The signal $x(n)$ is processed in two different branches. The upper branch implements the sound waves first reflections, which reach the receiver at a given position. These reflections are responsible for the sound direction perception and are influenced by the sound waves which achieve the receiver up to about 150 ms after the direct sound (this choice can be, of course, modified, allowing the selection of the first reflections to be used). The lower branch simulates the reverberant tail of the binaural impulse response, which occurs after the first reflections. No directivity is applied to this branch since, after a given time — say, 150 ms — the sound arrives statistically with the same probability from all directions.

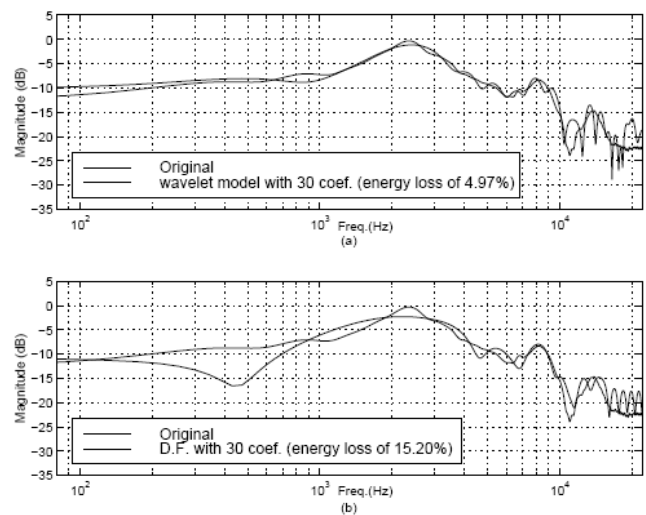


Figure 16. Magnitude comparison between the reduced and original frequency responses for Daub4 prototype ($\phi = 0^\circ$, $\theta = 90^\circ$, left ear).

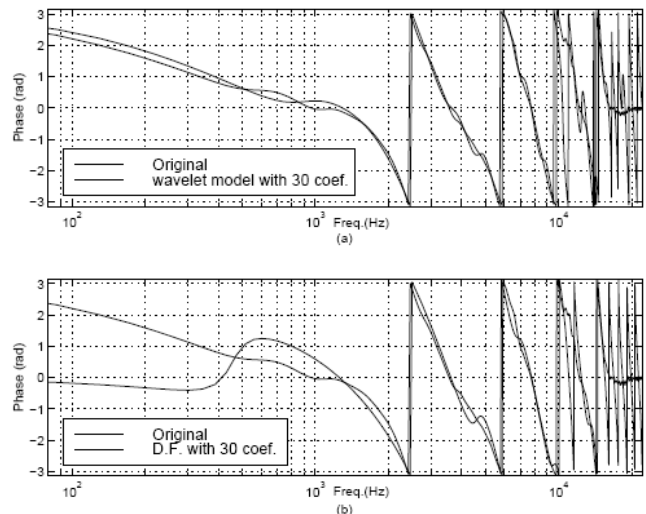


Figure 17. Phase comparison between the reduced and original frequency responses for Daub4 prototype ($\phi = 0^\circ$, $\theta = 90^\circ$, left ear).

In the directional processing, the anechoic signal is decomposed in sub-bands by the wavelet transform. This “spitted” signal is sent to different blocks, named regions, which represents the sound waves received in a given region of the space around the listener. Each region can be made wider or smaller as needed, by specifying a range for elevation and azimuth angles. The directivity of the region is given by the average of all HRTF coefficients (wavelet model) within in this space. Notice that, as wider is the region, poorer directionality is obtained. On the other hand, the region can be made as thin as a single direction, and the mean HRTF will be exactly the HRTF of the direction. However, a large number of regions would be necessary and the processing time will be extremely long. So, in this proposed auralization system, a compromise between the number of regions and processing time must be established according to the accuracy of the directional part of the impulse response and to the computational resources available.

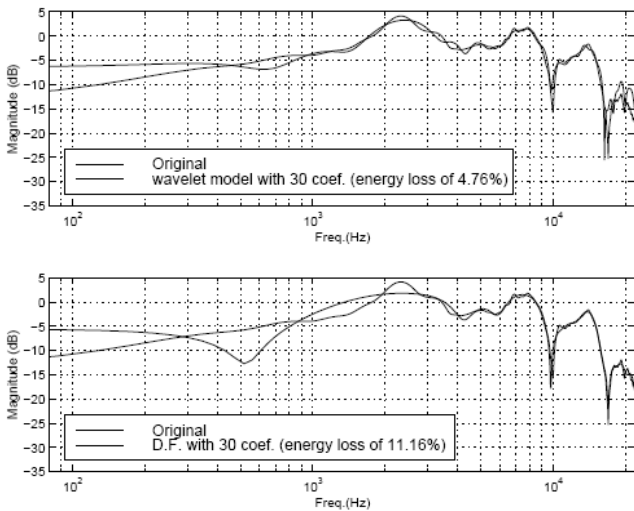


Figure 18. Magnitude comparison between the reduced and original frequency responses for Daub4 prototype ($\phi = 0^\circ$, $\theta = 90^\circ$, right ear).

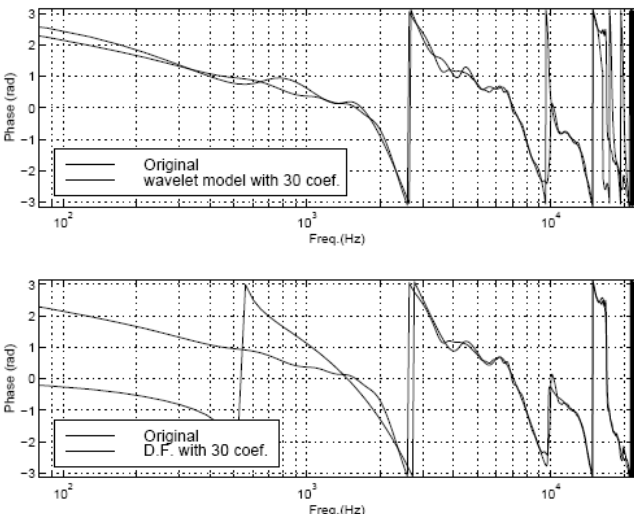


Figure 19. Phase comparison between the reduced and original frequency responses for Daub4 prototype ($\phi = 0^\circ$, $\theta = 90^\circ$, right ear).

Inside each region, the influence of the directionality and the sound waves must be joined to generate a directional impression of the incoming sound from a region in space. The sound wave influence is given by the energy contribution of each ray which reaches the receiver at every instant of time. This energy is computed in frequency bands according to the spectrum division produced by the wavelet transform. For each sub-band, the rays contribution creates short impulse decays which are processed by the coefficients of the representative HRTF of the region, as show in Fig. 21.

In the lower branch of Fig. (20) the anechoic signal is delayed by the same amount of time defined for the first reflections. Then the signal is processed by two different reverberates with the same reverberation time (T_{60}) whose coefficients are randomly generated to avoid correlation between left and right signals. The reverberation is implemented by a parallel set of comb filters with the same T_{60} . This reverberation time is obtained from the non-directional impulse response given by the rays which arrive after the time defined for the first reflections — approximately 150 ms. The output signals from each branch of the system are combined to generate the binaural signal (left and right signals).

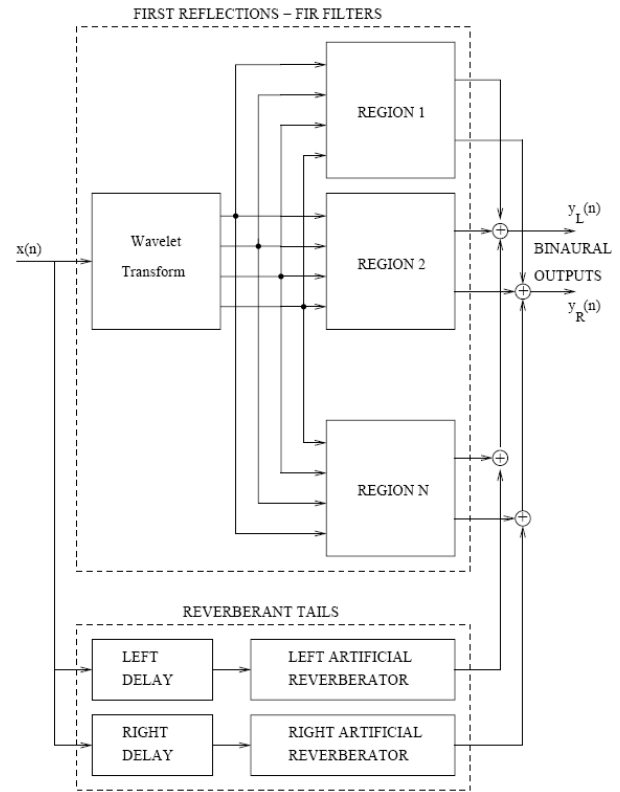


Figure 20. General structure for implementation of auralization.

The auralization implementation involves two steps. First, the rays captured by the receiver are separated according to its arrival time and spatial position around the listener. Rays with arrival time up to the first reflections time, say, 150 ms, are distributed over the regions of Fig. 20 and this is called directional processing. Rays with arrival time greater than the first reflections time are used to compose the reverberant tail for diffuse processing.

In the directional processing, small impulse responses are generated inside each region, according to the subband energy of each ray, since there are few first reflections. In order to generate 3D audio, the directional processing can be done as shown in Fig. 21, i.e., by processing an input signal $x(n)$, sample by sample, through each region structure, or, alternatively, by convolving all subbands of the regions to generate a directional impulse response corresponding to the first reflections.

In the diffuse processing, all rays are added, according to the arrival time, to produce a global impulse response, which starts at 150 ms, for instance. Over this diffuse impulse response, a T_{60} is calculated and a pair of artificial reverberators (binaural) is built to present the same decay.

Therefore, the BIRs implementation can be done by using directly the structure shown in Fig. 20 or by substituting the directional processing part of Fig. 20 by a pair of FIR filters, whose length corresponds to that one given by the first reflections time (150 ms, in our example). This simplified version of the auralization system is shown in Fig. 22.

The scheme presented in Fig. 20 is more flexible than the one presented in Fig. 22. A rotation head movement can be easily implemented simply by interchanging the small impulse responses between regions. The same change, using the scheme of Fig. 22, would require a new convolution of all regions with the wavelet transform.

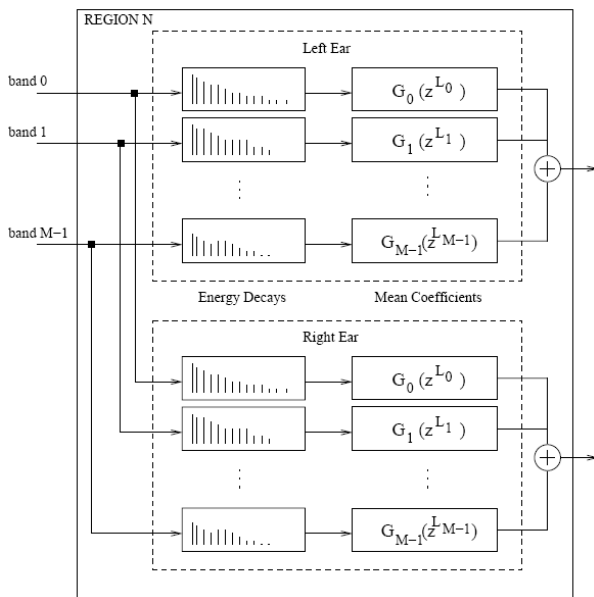


Figure 21. Region implementation for directional sound processing.

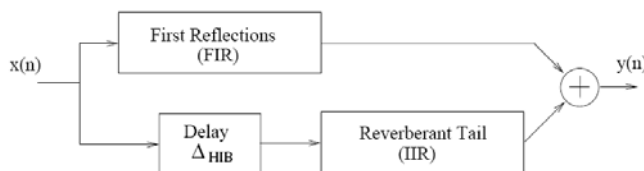


Figure 22. Simplified implementation of BIRs.

Conclusions

A new hybrid method to simulate the sound propagation in rooms was described. The main advantage of the proposed method is that the impulse response is computed combining the ray tracing technique, to obtain the first reflections, and the energy transition method, to generate the reverberant tail. This provides accurate and more realistic impulse responses than the ones obtained from either one of the methods alone.

Some small improvements in the classical techniques were reported and the way they were combined to produce a hybrid method was explained. This combination, as will be shown in the companion validation paper (Tenenbaum *et al.*), results in a room acoustics prediction with very good accuracy, regarding actual measurements.

The HRTF wavelet modeling and the auralization scheme used in the software RAIOS 3 lead to a significant improvement in the computational complexity, since the wavelets are computed once and the computational complexity decreases as the number of directions increases (comparing to the direct form of HRIR). The reduction in the number of coefficients to about 30% of the original HRIR allows a significant reduction in the number of mathematical operations — and, consequently, in the processing time, when the sound is processed by different directions (HRIRs) at the same time.

The efficacy of such model can be confirmed by running the software to simulate controlled conditions, where the results could be compared with precise measurements. This will be done in the companion paper (Tenenbaum *et al.*, 2007), which demonstrates the adequacy of the proposed method to simulate actual rooms, especially the ones with diffusion elements.

References

- AES 4ID 2001, "Aes information document for room acoustics and sound reinforcement systems – characterization and measurement of surface scattering uniformity", *Journal of Audio Engineering Society*, 3(49):149–165, 2001.
- Alarcão, D., Bento Coelho, J.L., Tenenbaum, R.A., "On modeling of room acoustics by a sound energy transition approach", *Proceedings of EEA Symposium on Architectural Acoustics*, Madrid, 2000.
- Ballesteros, M.L., Slama, J.G., Tenenbaum, R.A., "Numerical simulation of sound propagation applied to urban noise control", *Proceedings of Internoise 92*, Toronto, Canada, vol.6, pp. 153–156, 1992.
- Bento Coelho, J.L., Alarcão, D., Almeida, A.M., Abreu, T., Fonseca, N., "Room acoustics design by a sound energy transition approach", *Acta Acustica united with Acustica*, 86(6): 903–910, 2000.
- Blauert, J., "Spatial Hearing", *The MIT Press*, Cambridge, 1997.
- Bork, I., "Report on the 3rd Round Robin on room acoustical computer simulation – Part II: Calculations", *Acta Acustica united with Acustica*, 91(4): 753–763, 2005.
- Camilo T. S. Método híbrido para simulação numérica de acústica de salas: Combinação dos métodos de traçado de raios e transição de energia (in portuguese). In: Master Dissertation, COPPE, UFRJ, Rio de Janeiro, Brazil, 2003.
- Dalenbäck, B-I., Kleiner, M., Svensson, P., "A macroscopic view of diffuse reflection", *Journal of the Audio Engineering Society*, 42(10): 973–987, 1994.
- Dance, S.M., Shield, B.M., "The complete image-source method for the prediction of sound distribution in non-diffuse enclosed spaces", *Journal of Sound and Vibration*, 201(4): 473–489, 1997.
- D'Antonio, P., Cox, T.J., "Difusor application in rooms", *AA*, 60: 113–142, 2000.
- D'Antonio, P., Cox, T.J., "Two decades of sound diffuser design and development. Part 1: Applications and design", *Journal of Audio Engineering Society*, 46(11): 955–976, 1998a.
- D'Antonio, P., Cox, T.J., "Two decades of sound diffuser design and development. Part 2: Prediction, measurement and characterization", *Journal of Audio Engineering Society*, 46(12): 1075–1091, 1998b.
- Daubechies, I., "The wavelet transform, time-frequency localization and signal analysis", *IEEE Trans. Inform. Theory*, 36: 961–1005, 1990.
- Davies, W., Bermond, R., "Influence of diffuse reflections on the playing of musicians", *Proceedings of 17th International Congress on Acoustics*, 4D.10.01, Rome, September 2001.
- Embrechts, J.J., "Randomly traced sound ray techniques", *Acustica*, vol. 51, pp. 285–295, 1982.
- Farina, A., "Pyramid tracing vs. ray tracing for the simulation of sound propagation in large rooms", [online] Available at http://www.spectra.it/dc/fa_01.htm, 2000.
- Farina, A., "RAMSETE – A new pyramid tracer for medium and large scale acoustic problems", *Proceedings of Euro-Noise 95 Conference*, Lyon, March 1995.
- Fonseca, N., Abreu, T., Almeida, A.M., "Fones: Previsão de acústica de salas em espaços não-difusos por uma aproximação de random walk", *Relatório do Instituto Superior Técnico*, Lisboa, 1998.
- Georgiou, P., Kyriakakis, C., "Modeling of head related transfer function for immersive audio using a state-space approach", *Conference Record of the 33th Asilomar Conference on Signals, Systems and Computers*, 1999.
- Gomes, M.H.A., "Determination of the acoustical random incidence scattering coefficient", *D.Sc. Thesis*, UFSC, Florianópolis, SC, Brasil, 2002.
- Haan, C.N., Fricke, F.R., "Surface diffusivity as a measure of the acoustic quality of concert halls", *Proceedings of Australia and New Zealand Architectural Science Association Conference*, pp. 81–90, Sidney, November 1993.
- Haneda, Y., Kaneda, Y., Makino, S., Kitawaki, N., "Common-acoustical-pole and zero modeling of head-related transfer functions", *IEEE Trans. on Speech and Audio Proc.*, 7(2): 188–195, March 1999.
- ISO/WD 17487, "Acoustics – Measurement of the sound scattering properties of surfaces", 2001.
- Keränen, T., Huopaniemi, J., Kärkkäinen, A., "Simplified numerical modeling of individual HRTF", *Proc. International Congress on Acoustics (ICA'2001)*, Rome, Italy, number Paper 4P.38, 2001.
- Kistler, D.J., Wightman, F.L., "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction", *J. Acoust. Soc. Am.*, 91(3): 1637–1647, 1992.

- Kruzins, E., Fricke, F.R., "The prediction of sound fields in non-diffuse spaces by a 'random walk' approach", *Journal of Sound and Vibration*, 81(4): 549–564, 1982.
- Kulkarni, A., Colburn, S., "Role of spectral detail in sound-source localization", *Nature*, 396: 747–749, 1998.
- Kulowski, A., "Algorithmic representation of the ray tracing technique", *Applied Acoustics*, vol.18, pp. 449–469, 1984.
- Kuttruff, K.H., "Room Acoustics", 4 ed. London, Spon Press, 2000.
- Lam, Y.W., "A comparison of three diffuse reflection modeling methods used in room acoustics computer models", *Journal of the Acoustical Society of America*, 100(4): 2181–2192, 1996.
- Lewers, T., "A combined beam tracing and radiant exchange computer model of room acoustics", *Applied Acoustics*, vol.38, pp. 161–178, 1993.
- Martin, K.D., Gardner, W.G., "HRTF measurements of a kemar", *J. Acoust. Soc. Am.*, 97(6):3907–3908, 1995. MIT website: <http://sound.media.mit.edu/KEMAR.html>.
- Mellert, V., Mehrgardt, S., "Transformation characteristics of the external human ear", *J. Acoust. Soc. Am.*, 61(6): 1567–1576, 1977.
- Møller, H., "Fundamentals of binaural technology", *Applied Acoustics*, 36: 171–218, 1992.
- Stephenson, U.M., "Quantized pyramidal beam tracing – A new algorithm for room acoustics and noise immission prognosis", *Acta Acustica united with Acustica*, vol. 82, pp. 517–525, 1996.
- Suh, J.S., Nelson, P.A., "Measurement of transient response of rooms and comparison with geometrical acoustic models", *Journal of the Acoustical Society of America*, 105(4): 2304–2371, 1999.
- Tenenbaum, R.A., Camilo, T.S., Torres, J.C.B., Stutz, L.T., "Hybrid method for numerical simulation of room acoustics: Part 2 – Validation of the computational code RAIOS 3", *J. Brazilian Soc. Mech. Sci. and Eng.*, vol.29, (2), 211–221, 2007.
- Tenenbaum, R.A., Slama, J.G., Ballesteros, M.L., "Numerical simulation of room acoustics: A new approach for source modeling", *Proceedings of 14th International Congress on Acoustics*, F6-11, Beijing, September 1992.
- Thompson, D.M., Algazi, V.R., Duda, R.O., Avendano, C., "The CIPIC HRTF database", *WASSAP '01 (2001 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics)*, October 2001. CIPIC website: <http://interface.cipic.ucdavis.edu/>.
- Torres, J.C.B., Petraglia, M.R., "Performance analysis of adaptive filter structure employing wavelet and sparse sub-filters", *IEE Proceedings – Vision, Image and Signal Processing*, 149(2): 115–119, 2002.
- Torres, J.C.B., Petraglia, M.R., "Performance analysis of an adaptive filter employing wavelets and sparse sub-filters", *EUSIPCO 2000*, vol.2, pp. 997–1001, 2000.
- Torres, J.C.B., Petraglia, M.R., Tenenbaum, R.A., "An efficient wavelet-based HRTF for auralization", *Acta Acustica united with Acustica*, 90(1), January 2004.
- Torres, R.R., Kleiner, M., Svensson, U.P., et al., "Subjective evaluations of scattering in rooms", *Proceedings of 17th International Congress on Acoustics*, 3B.08.02, Rome, 2001.
- Vaidyanathan, P.P., "Multirate Systems and Filter Banks", *Prentice-Hall*, Englewood Cliffs, New Jersey, 1993.
- Vorländer, M., Mommertz, E., "Definition and measurement of random-incidence scattering coefficients", *Applied Acoustics*, vol.60, pp. 187–199, 2000a.
- Vorländer, M., "Room acoustical simulation algorithm based on the free path distribution", *Journal of Sound and Vibration*, 232(1): 129–137, 2000b.
- Vorländer, M., "Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm", *Journal of the Acoustical Society of America*, 86(1): 172–178, 1989.
- Wightman, F.L., Kistler, D.J., "The dominant role of low-frequency interaural time differences in sound localization", *J. Acoust. Soc. Am.*, 91(3): 1648–1661, 1992.