Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

Virtusound – a real time auralization system

D. Alarcão; D. Santos; J. L. Bento Coelho

CAPS, DEEC, IST, TULisbon, Portugal
diogo.alarcao@ist.utl.pt

PACS: 43.55.KA, 43.55.HY

ABSTRACT

A real time room acoustics auralization system for arbitrarily shaped rooms was developed. *Virtusound* uses an accelerated image source method and a time-dependent radiosity method for the computation of the binaural room impulse response. The total room impulse response is obtained by summing the individual responses from each method. Accelerating techniques and algorithmic improvements are used in *Virtusound* allowing real time auralization through binaural technology. The current implementation of *Virtusound* is done on a standard multi-core personal computer..

INTRODUCTION

Auralization is the process of making audible the acoustics of complex virtual architectural spaces in a realistic and accurate manner. This paper presents a newly developed and implemented real-time room acoustics simulation and auralization platform, the *Virtusound* system.

Real time auralization systems have been developed by other authors, see for example [1, 2], but *Virtusound* has its own unique features, the more important ones being capable of simulating in real time both specular and diffuse reflections, and being implemented on a single multi-threaded PC, so no network communications are necessary.

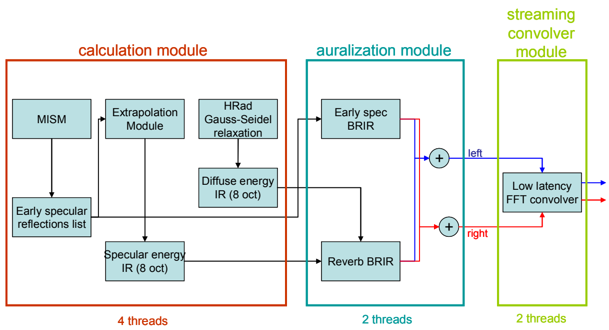
For most practical purposes in room acoustics it is only necessary to obtain the prediction of how the sound energy propagates inside the enclosed space.

In addition, under the simplifying assumption that the enclosure's walls reflect sound energy as a mixture of specularly and diffusely reflected components, it can be shown that solutions for the propagation of sound energy inside enclosures can be achieved by a combined method [3, 5]. This new combined method resorts to an extended mirror image source method solving for the propagation of the specularly reflected sound energy components and to a time-dependent hierarchical radiosity approach in order to solve for the propagation of the diffusely reflected sound energy components.

This combined method was used for the development of the *Virtusound* system by the Acoustics and Noise Control Group, at CAPS-IST, TULisbon.

SYSTEM ARQUITECTURE

Figure 1 shows the block diagram of the complete *Virtusound* system.



Source: (Author, 2009)

Figure. 1. Block diagram of the *Virtusound* system

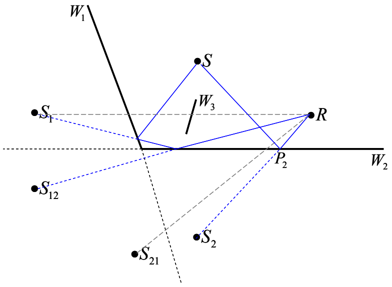
As can be seen, the developed system is composed by three main modules: Calculation, Auralization and Streaming.

This system is being currently implemented on an *Intel I7 Extreme* processor platform, running on eight threads

DESCRIPTION OF THE SYSTEM

Calculation Module

The algorithm for solving the specularly reflected components was implemented by using an extended image source method, while the algorithm for solving the diffusely reflected components is based on a Hierarchical Radiosity [4] algorithm that allows the error thresholds to be defined by the user in order to ensure maximum accuracy of the solution and low computation times.



Source: (Author, 2005)

Figure. 2. Extended Image Source Method.

Extended Image Source Method

Owing to the exponential growth of the number of potential image sources with reflection order, the exact calculation of the mirror image sources is performed only until some given maximum geometrical reflection order, typically 3 to 5.

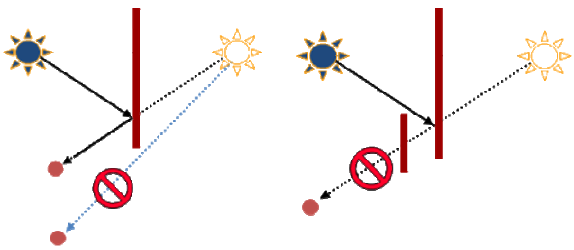
Number of potential image sources:

$$M + M(M-1) + M(M-1)^2 + \ldots + M(M-1)^{K-1} = \frac{M(M-1)^K - 1}{M-2} \approx (M-1)^K \quad (1)$$

where M is number of room walls and K the reflection order.

Accelerated Image Source Method

In order to obtain the list of correct mirror image sources some validity and visibility tests need to be performed.



Source: (Author, 2005)

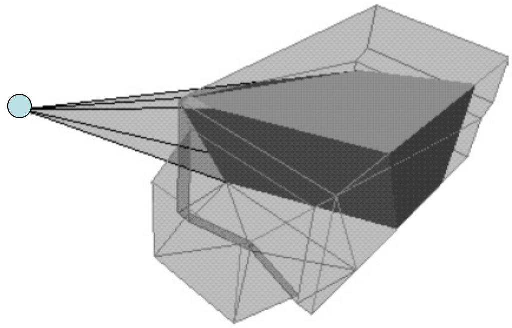
Figure. 3. Validity and visibility tests.

To do this in the least time possible, four accelerating techniques are used:

•*Back-Face Culling*, a straightforward implementation that consists on taking into account the orientation of the input polygons according to their inward-orientated normals. Then, mirroring takes place only in the case that the sources (be it the original sound source or the constructed mirror images) are facing the inward face of some polygon.

•*Impossible Wall Combinations* [6, 7], a technique that permits recognizing a priori non-physical mirror images representing impossible combinations of input polygon pairs. The original geometric representation of the enclosure under study is used in order to build a list of impossible polygons combinations, meaning that sound reflected by some particular polygon can not reach directly other polygons. The information stored in a pre-process phase in the complete list of impossible reflections combinations is then used during the geometric construction phase of the potential mirror images in order to discard immediately these combinations. With this accelerating technique, one spares greatly on the costly validity and visibility tests.

•*View-Frustum* [8] allows a third type of branch to cut from the complete tree of potential mirror images. These branches represent mirror images, which are constructed in a valid manner, but which are not visible from any position within the enclosure and therefore do not take part in the room's impulse response.



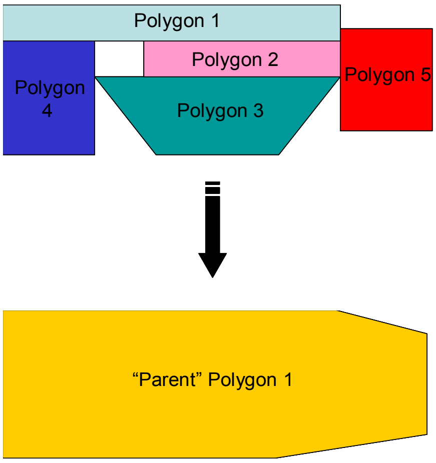
Source: (Author, 2005)

Figure 4. View frustum for discarding subsequent higher reflection order mirror sources

This accelerating technique was used in a slightly different manner in [6], while in [7] the authors suggest the use of radiation angles, which in practical terms fulfills the same purpose. Start with some mirror image and construct its view frustum as indicated in the example of Figure 1. In practice, one uses the fact that valid reflection surfaces for descendants of an image source (yielding higher reflection order mirror sources) are the ones whose inward-orientated faces are seen within the view frustum, whose apex coincides with the mirror image. This accelerating technique yields a good performance increase in the MISM since it allows, especially for high reflection orders, to cut immediately many branches of the tree of potential images.

•*Clustering of Input Polygons* uses the fact that in the majority of the enclosures found in practice all the input polygons (which can be in great number) normally lie on a small set of tri-dimensional planes. Therefore, in a pre-processing phase, a clustering algorithm builds up a single "parent" polygon in a similar way as the so-called convex hull of a set of points is constructed. This clustering algorithm greatly reduces the complexity of the problem, because the number of wall combinations can be greatly reduced in a wide variety of practical cases, with the apparent consequence that the number of potential mirror images is enormously reduced due to the reduced basis of the exponential law.

The accelerated image source method with the clustering algorithm uses a simplified representation of the enclosure given by all the "parent" polygons (which are convex) for the geometric construction of the potential images. Only when a valid and visible mirror image is found, in relation to this set of "parent" polygons, does one need the complete set of original input polygons for determining which of them actually are responsible for the construction of the determined mirror image.



Source: (Author, 2005)
Figure 5. Example of clustering of input polygons into one single "parent" polygon

This determination is not a very costly operation, since one need only to determine inside which of the input polygons the intersection points are located. It is obvious that a valid and visible image relative to the reduced set of "parent" polygons can be an invalid or invisible mirror image afterwards, relative to the finer set of input polygons. However, these cases constitute normally a minority, since the "parent" polygons are constructed in a way similar to the convex hull of a set of points, which is the convex set with smallest area that contains all the original input points.

Next, the pseudo-code of the accelerated image source method is presented by using the above mentioned accelerating techniques. The procedure is a recursive one, thereby traversing one branch of the potential mirror images until the maximum reflection order is reached. The nomenclature of the variables and functions is self-descriptive.

```
calculateimages(r,W_i){
  if(length(r.poly) < MAXORDER){
    if(notbackface(r.coord,W_i)){
      if(insideviewfrustum(W_i,r.coord,last(r.poly))){
        ++n;
        r_n.coord = r.coord - 2(dp(W_i) + r.coord.normal(W_i))normal(W_i);
        append(W_i to r_n.poly);
        store(descendent[r] = r_n; predecessor[r_n] = r);
        if(imagevisible(receiver,r_n)){
          append(r_n to visibleimageslist)
        }
        for(j = 1 to M){
          calculateimages(r_n,W_j)
        }
      }
    }
  }
}
```

In Table 1 we indicate some calculation times for the implemented accelerated image source method for different rooms. Calculation times are for reflection order 4. As can be seen, the calculation times are very small, suitable for real time auralization purposes.

Table 1 - Calculation times for the accelerated image source method

Room	Polygons	Calc Time [ms]
Simple	26	1
Congress Centre	55	15
PTB Music Studio	77	20

Source: (Author, 2009)

Extension to High Orders of Reflection

Information supplied by the accelerated image source method is then used for a statistical and deterministic extrapolation to higher reflection orders in order to obtain the reverberation tail of the room's specular impulse response.The extrapolation step resorts to the following parameters, obtained during the exact geometrical construction phase:

- Number of visible images per reflection order: ***N(K)***
- Mean distance receiver - visible images of order K: $\bar{D}(K)$
- Standard deviation of the distances of order K: $\sigma(K)$
- Mean reflection coefficient for reflections of order K: $\bar{\rho}(K)$
- Standard deviation of reflection coefficients of order K: ***g(K)***
- Mean diffusivity coefficient for reflections of order K: $\bar{\Delta}(K)$
- Standard deviation of the diffusivity coefficients of order K: $\eta(K)$

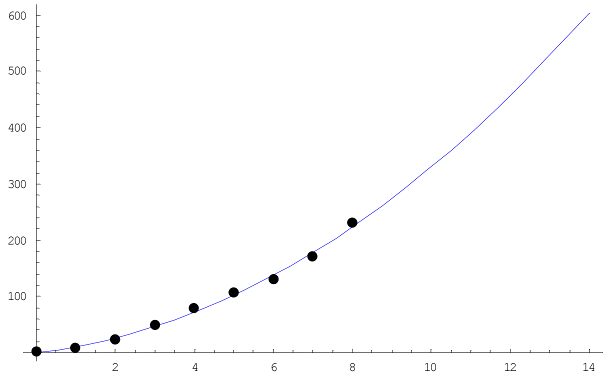
The following functional relationship for the number of visible images per reflection order is adopted for the extrapolation step:

$$\lim_{K\to\infty} N(K) = a + bK^2 \quad (2)$$

where the coefficients *a* and *b* are obtained by a least-squares fit to the data obtained by the accelerated image source algorithm, ***N(K)***.

For the rectangular enclosure:

$$N(K) = 2 + 4K^2 \quad (3)$$



Source: (Author, 2009)
Figure 6. Example of least squares fit of the quadratic functional to experimental number of correct mirror sources obtained up to order 8 – Auditorium of the IST congress centre

After that, and based on the Information supplied by the accelerated image source method, we need to estimate a "Mean distance receiver - image" and the related "Standard deviation" for each statistical order, that it's possible applying a linear regression:

$$y = \alpha + \beta x \quad (4)$$

where x and y are the statistical order and the corresponding "Mean distance receiver - image" or the correspoding "Standard deviation". α and β are obtained with the information supplied by the accelerated image source method, $\bar{D}(K)$ and $\sigma(K)$

The distribution used for extrapolating the distances between each statistical image and the receiver in function of reflection order *K* is a Gaussian-distribution:

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5}$$

where μ and σ^2 are the mean and the variance of the distribution, previously estimated according to equation (4) to each order K. So $N(K)$ reflections, as given by equation (2), are generated according to this distribution for every higher reflection order.

The mean reflection coefficient $\overline{\rho}(K)$, and the corresponding standard deviation $g(K)$, for reflections of order K are calculated from the registered collision frequencies with the walls, $f(W_j,K)$:

$$\overline{\rho}(K)=\sum_{j=1}^{M}\left[\frac{f(W_j,K)}{\sum_{i=1}^{M}f(W_i,K)}\right]\rho(W_j)$$

$$g(K)=\overline{\rho^2(K)}-\left(\overline{\rho}(K)\right)^2$$

and the mean specularity coefficient, $\overline{1-\Delta(K)}$, and the corresponding standard deviation $\eta(K)$, for reflections of order K, are also calculated from the registered walls' collision frequencies:

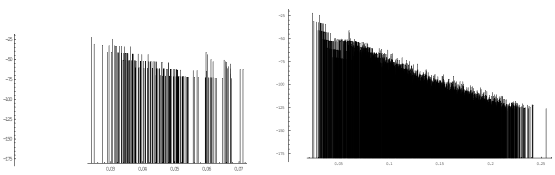
$$\overline{1-\Delta(K)}=\sum_{j=1}^{M}\left[\frac{f(W_j,K)}{\sum_{i=1}^{M}f(W_i,K)}\right](1-\Delta(W_j))$$

$$\eta(K)=\overline{(1-\Delta(K))^2}-\left(\overline{1-\Delta(K)}\right)^2$$

The room impulse response from specular reflections is thereby calculated from:

$$I_{specular}(t)=\sum_{K=0}^{\text{maxgeoorder}}\sum_{i=1}^{N(K)}\frac{\Pi}{4\pi d_i^2}\rho_i e^{-md_i}(1-\Delta_i)\delta\left(t-\frac{d_i}{c}\right)$$
$$+\sum_{K=\text{maxgeoorder}+1}^{\text{maxorder}}\sum_{i=1}^{N(K)}\frac{\Pi}{4\pi D_i(K)^2}\overline{\rho}(K)e^{-mD_i(K)}\overline{1-\Delta(K)}\delta\left(t-\frac{D_i(K)}{c}\right)$$

where "maxgeoorder" means the maximum reflection order for which the exact geometrical calculation of the visible images is done.



K=6 K=45

Source: (Author, 2010)

Figure 7. Room echograms from specular reflections. Left: up to order 6. Right: up to order 45

Hierarchical Time Dependent Radiosity Method

Solving for the diffuse reflected components by a finite element approach, raises two problems:

Problem 1: For the method of images it is required that the walls constituting the geometry of the enclosure be as large as possible, whereas for the finite element approach one desires that the input polygons constitute a mesh of several smaller "patches".

Problem 2: If M initial walls are split into n patches, then the number of energy links, i.e. the approximate number of form factors, will be proportional $O(n^2)$, and therefore the calculation effort is high.

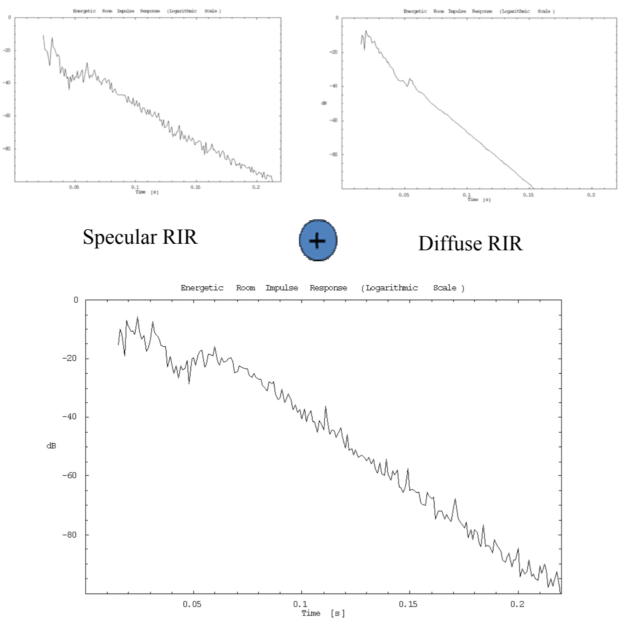
One solution for both problems is to adopt a Multi-Resolution approach through hierarchical linking [4], whereby the patches are generated adaptively from the input walls and the number of links (form factors) is $O(M^2+n)$, therefore permitting a considerable computational saving.

The hierarchical subdivision stop criterion is based in first place on the absolute minimum area of patches (defined by the user) and in second place on a threshold condition for the form factors between pairs of patches (also defined by the user). Form factors are calculated by using the analytic formula of the form factor from a differential area to a parallel disc, and the visibility factor for the occluded form factor is calculated by ray-casting.

The solution of the time-dependent equations is done by a Gauss-Seidel relaxation scheme, using an energy "push-pull" operation throughout each hierarchy of patches. In order to limit the exponential growth of diffuse reflections with increasing reflection orders, a condensing algorithm with an internal sampling rate is used.

At each iteration, energy is gathered at predefined receiving points, calculated through the well known Lambertian cosine law.

The total room impulse response is obtained by the sum of the specular response and the diffuse response.



Total room energy impulse response

Source: (Author, 2010)

Figure 8. Total room energy impulse response.

Auralization module

The auralization system runs independently on asynchronous threads, loading periodically the list of visible images as calculated by the accelerated image sources module, the list of extrapolated images and the energy echograms as calculated by the hierarchical time-dependent radiosity module. The auralization system implements a wide band approach for the treatment of the frequency dependence. The system is currently adapted for rendering sound through binaural technology.

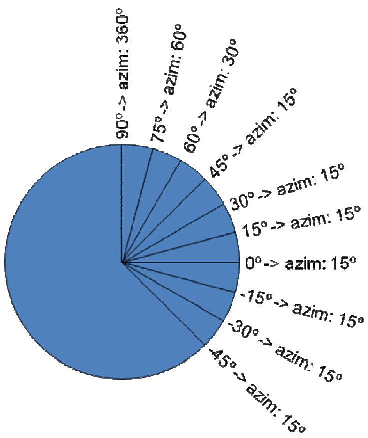
Auralization of the specular components

The list of correct sound images calculated by the accelerated image source method comprises the sources location as well as the list of walls that generated them. The incidence angles regarding the particular receiver point are also recorded in the list, as well as the image source's directivity. With this information, equivalent time invariant linear filters are constructed online by the auralization system. Elements such as source directivity, wall and air absorption are represented through linear filters in logarithmic frequency dependence with an octave band scale with 10 values ranging from 31.5 Hz up to 16 kHz. These filters have their phase determined by a minimum phase reconstruction step conducted through FFT Hilbert Transforms.

In order to obtain the binaural pressure impulse responses, the result of the previous filters have to be convolved with the set of HRIR for each ear corresponding to the correct direction of incidence at the receiver.

The same procedure is done with the extrapolated images, however the linear filters are constructed based on the values obtained for the visible *K* orders and their phase is considered zero. It's also important to refer that the directivity of each extrapolated images is defined by a normal random variate generator between the 187 available HRIR directions.

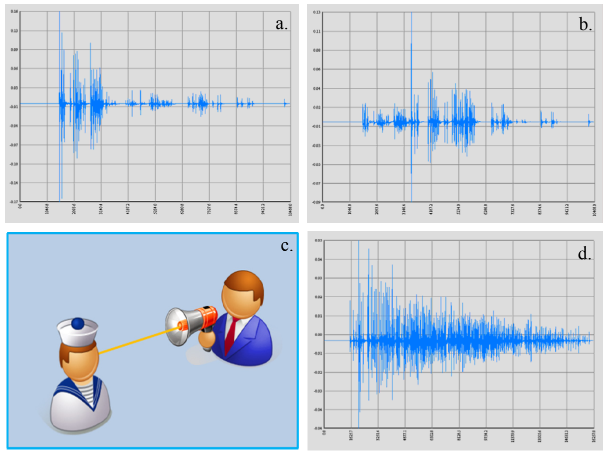
The auralization system uses currently the HRIR database of IRCAM, which were measured at a distance of 2.0 m. This database has a resolution of 15° in the azimuth angle. The vertical resolution consists of 10 elevation angles starting at -45° ending at +90° in 15° steps. The steps per rotation varies from 24 to only 1 (90° elevation). Measurement points are always located at the 15° grid, but with increasing elevation only every second or fourth measurement point is taken into account. As a whole, there are 187 HRIR pairs. Each HRIR data has a sampling rate of 44.1 kHz and 24 bits of quantization, consisting of 8192 samples long, we pre-processed each HRIR data set in order to have only a set of 256 samples.



Source: (Author, 2010)
Figure 9. HRIR Model.

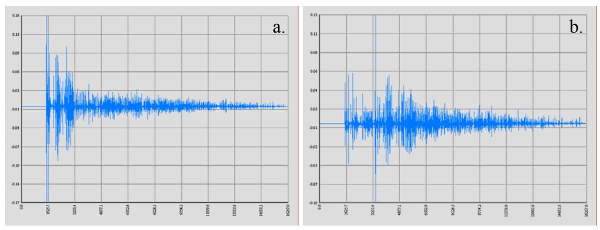
All the convolutions are done in the dual frequency domain using radix-2 FFTs with an overlap-save implementation.

The final result of this stage corresponds to three binaural impulse responses (BRIR) containing all the information about the specularly reflected sound components, two of them using the visible images supplied by the accelerated image source method, corresponding each one to each side of the receiver and the last one using the extrapolated images.



Source: (Author, 2010)
Figure 10. Example of an auralization system result, reflection order=5 for visible images, reflection order=6 till 10 for extrapolated images: a) - visible images BRIR - left side; b) - visible images BRIR - right side; c) - receiver/source related position; d) - extrapolated images BRIR

In order to obtain the complete binaural impulse responses for the specular components it's necessary to combine the BRIR of the visible images with the room reverberation - BRIR of the extrapolated images.



Source: (Author, 2010)
Figure 11. Example of a combined Binaural for the specular components (visible + extrapolated images) a) - BRIR - right side; b) - BRIR - left side.

Auralization of the diffuse components

The result of the hierarchical time-dependent radiosity method consists of an energy echogram for the 10 octave bands. This echogram is considered as an envelope of the squared pressure impulse response for each frequency band.

In order to reconstruct a pressure related impulse response a procedure using a filtered white noise is implemented. In detail, a 10 ms width Gaussian window is applied successively together with an equivalent magnitude filter determined from the calculated energy echograms to the white noise. This procedure is repeated for increasing time steps until the time window slides out of the time interval of interest.

For the binaural simulation two partially correlated responses are created which are equalised by means of an averaged HRIR according to diffuse incidence.

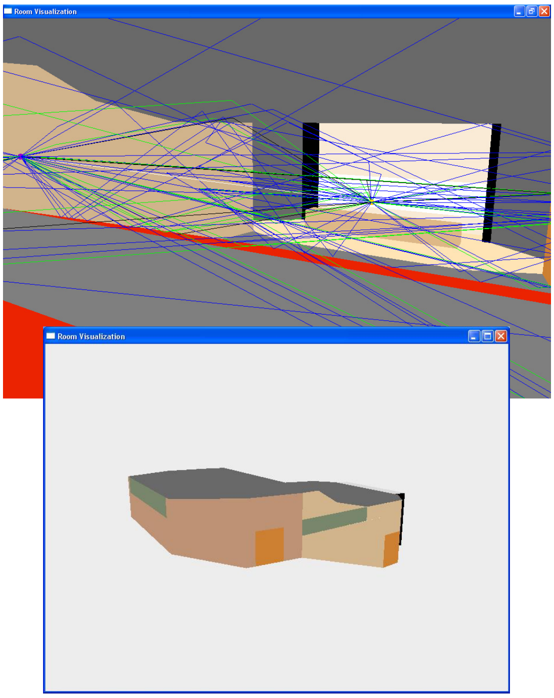
Auralization of all components

The binaural impulse responses for the specular and for the diffuse components are combined into a single, total binaural impulse response set.

This final set is then convolved with a real-time streaming 44.1 kHz, 16 bit audio using an audio library called FMOD [9], through a Gardner like low-latency filtering approach. Sound reproduction is done through highfidelity headphones.

USER INTERFACE

Figure 12 shows an example of the user interface that was implemented by the authors. Based on OpenGL [10] code source, this interface gives the users an easy way to try all the options such as changing the position and directivity of the source and the receiver, as well as changing the way one wants to see the room, by using rotation, translation, zoom in and zoom out. An option is also available to see the correct sound paths between the source and the receiver for each visible image order. All these options are available using the keys of the pc keyboard.



Source: (Author, 2010)

Figure 12. Room visualization module showing sound paths between source and receiver – Auditorium of the congress centre of IST

PERFORMANCE EVALUATION

Table 2 gives some data on the performance of the system.

Table 2 - Calculation times for the different tasks of the *Virtusound* system – PTB Music Studio

Task	Calc Time [ms]
MISM early images (order 5)	85
Extrapolation module	0,06
HR Gauss Seidel relaxation	500
Diffuse energy IR at receiver	6
Early specular refl BRIR	5,5
Reverb BRIR	11
FFT Convolver (1024 samples)	10

Source: (Author, 2010)

CONCLUSIONS AND FUTURE WORK

The system described in this paper works in real time, allowing for the movement of the receiver and of the source.

The accelerated image source method permits the real time calculation of all visible images for some source-receiver pair, while the hierarchical time-dependent radiosity method calculates the diffusely reflected sound components through octave band energy echograms.

The auralization system is based on a wide band simulation approach. Elements such as wall and air absorption, source and receiver directivities are taken into account.

Preliminary listening suggest that the auralizations are very convincing, the sound being natural and without colorations.

Future work will access the possibility of using some space partitioning hierarchy for speeding up the accelerated image source method and the hierarchical time dependent radiosity method for rooms with greater complexity.

It is also envisaged to carry out more listening tests with a panel of subjects in order to obtain a better assessment of the auralizations obtained with the *Virtusound* system.

ACKNOWLEDGEMENTS

The work presented here was financially supported by the Portuguese Science and Technology Foundation (Project PTDC/EEA-ELC/80910/2006).

REFERENCES

1 T. Lentz, D. Schröder, M. Vorländer and I. Assenmacher; Virtual Reality System with Integrated Sound Field Simulation and Reproduction. *EURASIP Journal on Advances in Signal Processing,* Vol. 2007, Article ID 70540

2 M. Noisternig, B. Katz, S. Siltanen and L. Savioja; Framework for Real-Time Auralization in Architectural Acoustics. *Acta Acustica united with Acustica, Technical and Applied Papers*, Vol. 94 (2008), 1000 – 1015.

3 D. Alarcão and J.L. Bento Coelho; A Combined Method for Room Acoustics integrating Specular and Diffuse Reflections. *In Proceedings of 18th. International Conference on Acoustics*, Kyoto, Japan, April, 2004.

4 P. Hanrahan, D. Salzman and L. Aupperle; A Rapid Hierarchical Radiosity Algorithm. *Computer Graphics*, 25(4):197-206, 1991.

5 D. Alarcão, Acoustic Modelling for Virtual Spaces; *PhD Thesis*, Instituto Superior Técnico, TULisbon, 2005

6 H. Lee; B. Lee; An efficient algorithm for the image model technique. *Applied Acoustics*, Vol. 24 (1988), 87-115

7 U. R. Kristiansen, A. Krokstad, T. Follestad; Extending the image method to higher-order reflections. *Applied Acoustics*, Vol. 38, n. º 2-4, p. 195-206, 1993

8 Watt, M. Watt; Advanced animation and rendering techniques – Theory and practice. *ACM Press*, Addison-Wesley Publishing Company, 1992

9 http://www.fmod.org

10 http://www.opengl.org