Fast and controllable box-shaped room impulse response algorithm

Ernesto Accolti† Federico Miyara‡

Laboratorio de Acústica y Electroacústica, Facultad de Electrónica, Universidad Nacional de Rosario. Riobamba 245 bis, Rosario, Argentina

† *E-mail: eaccolti@fceia.unr.edu.ar*

‡ *E-mail: fmiyara@fceia.unr.edu.ar*

***Abstract*— A quite realistic room impulse response (RIR) algorithm has been developed. The underlying model is based on physical properties of RIRs. Some approximations and simplifications based on physical and psychoacoustic principles allow a fast calculation with negligible errors in energy or psychoacoustic parameters. Early reverberation (ER) is calculated by means of the image source method (ISM) until a transition time is reached when the sound field may be considered random. Based on statistical acoustics theory, late reverberation (LR) is modelled from gaussian noise taking into account the frequency-dependent energy decay curve and an ISM-based, smoothed peak density growth. To accomplish a realistic energy balance between ER and LR they are compared during an overlapping interval.**

***Keywords*— Room Impulse Response, Box Shaped Rooms, Architectural Acoustics, Acoustics.**

1. INTRODUCTION

There are several models for computing acoustic room impulse responses (RIRs) The selection of the most appropriate one depends on the application and is the result of a compromise between accuracy and computational cost. RIR`s allow to compute aural simulations as well as some acoustical parameters.

This paper introduces a fast but realistic model using a small amount of information. The RIRs obtained will be used for auralisation. Emphasis is in realism rather than accuracy. The room response should be acceptable for non-specialised listeners even if it does not correspond exactly to the properties of surface materials such as absorption coefficients and their frequency dependence. It is intended for use as a complement of a routine for the controlled combination of sounds [1].

1.1. Image Source Method Models

The image source method (ISM) applied to box-shaped rooms for computing RIRs is well known. Figure 1 shows a two dimensional virtual source model that is easily extended to the three dimensions of the room. The real receiver receives acoustic rays from the real source and the image sources located at the virtual rooms. The number of rays usually depends on certain criterion being reached.

With the layout of Fig 1, assuming that each source emits a ray starting at the same initial time $t = 0$, the reflection pattern at a receiver position can be obtained. The model allows for an attenuation factor that depends on the reflection coefficient of each wall through which it goes. The reflection coefficient R is a complex value that depends on the incidence angle θ and could be expressed, for bulk materials and plane waves, in terms of the specific acoustic impedance Z_a of the wall

$$R(\theta) = \frac{Z_a - Z_f}{Z_a + Z_f} \tag{1}$$

where $Z_f = \rho_0 c/\cos(\theta)$ is the impedance of the air for plane waves on infinite panels, being ρ_0 the air density, c the sound velocity and θ the incidence angle. Sometimes it is preferable to assign the incidence angle dependence to the panel impedance, but in the line we are following it is interesting to state the dependence as if it were of the field (i.e., air).

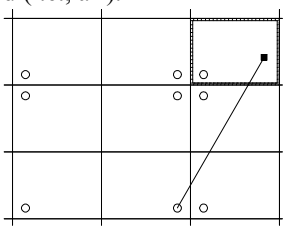

Figure 1: Image sound sources of a rectangular room. White Circles: Real and image sources. Black Square: Receiver.

Furthermore, there are models for R when the materials cannot be treated as bulk, as well as for spherical waves and finite panels.

Rindel [8] developed an empirical approximation for calculating Z_f for a finite dimensions panel of S area and P perimeter introducing the parameter $e = 4P/S$:

$$Z_f = \rho_0 c\left[\left(\cos^2\theta - \frac{1.2\pi}{ke}\right)^2 + \left(\frac{1.44\pi^3}{k^2e^2}\right) + \left(\frac{16\pi^4}{k^8e^8}\right)\right]^{-1/4} \tag{2}$$

where k is the wave number.

However, Z_a data are not as easily found in the literature as random-incidence absorption coefficients. We may approximate R by its random-incidence value R_r, obtained from the random-incidence absorption coefficient α:

$$R_r = \sqrt{1 - \alpha} \tag{3}$$

In this case the incidence angle and the reflection pattern are ignored, but this random value approximates well the 60º incidence reflection coefficient value R_{60}. Furthermore Rindel used a simplified form of the incidence-angle-dependent reflection coefficient $R(\theta)$ as in

Eq. (1) determining Z_a from R_{60}, the reflecting panel dimensions and a direct inspection comparison of the material impedance with that of the air, assuming the approximation for locally reacting and real impedance materials as

$$Z_a = Z_{f60}\frac{1 \pm R_{60}}{1 \mp R_{60}} \tag{4}$$

where the upper signs are used when materials are hard and heavy such as those whose impedance is larger than $2\rho_0 c$, and when the impedance is smaller the lower signs should be used.

Other aspect to take into account is sound diffusion. In the most commonly used materials the direction of major energy reflection is the specular one but there is energy reflected in other directions also. A ray that strikes a wall can be seen as a constant incident angle plane wave front, and then a reflection pattern could be constructed for each incidence angle. This fact, combined with the phase of R, will cause a ray to have a more wide impulse response in comparison with an ideal Dirac's delta depending on the reflection pattern for each incidence angle of each wall that the ray has struck. In an ideally diffusive field this information would affect the absorption coefficient for random incidence α_r as an average for all the possible incidence angles, but the phase information is lost as well as the information about the panel dimensions dependence.

The attenuation for each ray would be expressed in terms of α_r, and could be attained as a Finite Impulse Response (FIR) filter for each ray. The FIR representation instead of an Infinite Impulse Response filter (IIR) would benefit the model because it does not need to filter an impulsive signal before adding it to that of the other rays since its impulse response is the same as its coefficients. In other words, the FIR representation of each ray can be directly summed with an adequate delay.

There are models that use IIR filters, for example nested allpass filters (called reverberators), and have interesting applications in sound recording but they do not provide high correlation with physical characteristics of the room such as frequency dependence of reverberation time. Some modifications of this kind of model can incorporate these physical characteristics but in a stchocastic way. This is interesting for modelling the late reverberation.

One simple methodology for designing this kind of FIRs is the frequency sampling technique [7]. Then each ray's impulse response can be achieved by interpolating a continuous curve in the frequency domain and then sampling it at N points that will fit the frequency resolution F_s/N.

On the other hand, a statistical model is derived from the ISM by considering the time density of rays, which grows as the square of time (see [4]). If M is the number of rays, the ray density per sample $\Delta M/\Delta n$ is obtained from the time density by replacing $t = n/F_s$, yielding

$$\frac{\Delta M}{\Delta n} = \frac{4\pi c^3 n^2}{V F_s{}^3} \tag{5}$$

where t is the continuous time; n is the sample number; F_s is the sampling rate and V is the room volume.

Integrating Eq. (5) over the discrete time n it is shown that the total number of rays grows as the cube of time and so does the computational cost.

The diffuse pressure decay factor, given by

$$m_{f_i} = e^{-\frac{6{,}91}{T_i}\frac{n}{F_s}} \tag{6}$$

where T_i is the reverberation time for the i-th frequency band (centred at f_i), can be used as an envelope for white noise.

1.2. Structure of RIRs

RIRs can be separated into direct sound (DS), early reverberation (ER) and late reverberation (LR) as shown in Fig. 2.

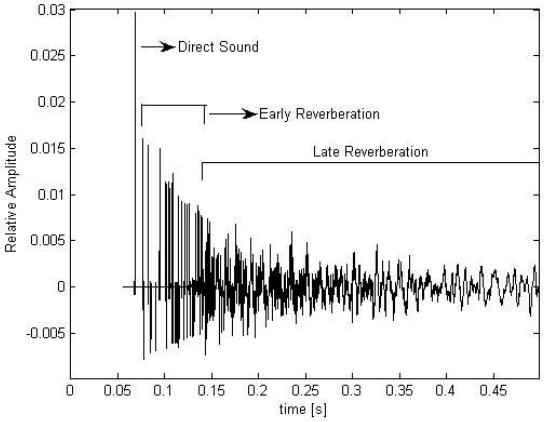

Figure 2: Modelled RIR and its parts.

If specular rays are assumed, i.e., surfaces are not diffusive, the DS is attenuated only by divergence and air attenuation. In contrast, the ER and LR also include the attenuation caused by sound absorption of the walls in which each ray reflects. Furthermore, since specular reflections are assumed, only the incidence angle-dependent reflection coefficient is needed.

Many models dissociate these three parts since the mix levels ratio between DS and ER plus LR gives the sensation of source distance. However, DS and ER can be treated as a single block if the source distance is given, assuming DS reflects in *zero* walls.

LR is the reverberation after the transition time t_L, where a diffuse sound field begins to be achieved, until it is inaudible. This t_L can be defined as the time when the minimum value of the correlation between the whole RIR signal and the RIR signal from a time dependent parameter through the end of the RIR [3]. So defined, t_L cannot be calculated before the whole RIR is computed or measured. To circumvent this difficulty, Hidaka derived empirically the average t_L as the 8 % of the reverberation time (RT) at the 500 Hz octave band for 24

symphony halls, 13 chamber music halls and 22 opera houses, finding a correlation of 0,76 between both parameters. Hidaka's literature research in that paper reveals that the LR can be modelled by a random phase model because psychoacoustic impressions do not discriminate the microscopic part of LR and that only the gross spectral and temporal characteristics are meaningful.

2. THE MODEL

The proposed model separates the RIR into two blocks: first, the ER plus the DS, and second, ER to LR transition plus LR. Before computing the amplitude of these blocks, the arrival time t_i of each ray i is calculated (using ISM and approximating by the closest discrete time n_i) until time t_c, represented by a sample n_c, from which onwards the probability of a ray arriving at any sample is 1. This occurs when the ray density $\Delta M/\Delta n = 1$, i.e. after sample

$$n_c = \sqrt{\frac{V F_s^{\,3}}{4\pi c^3}} \tag{7}$$

A sorted look-up table [6] allows calculating the reflection time pattern in a faster way ignoring image sources that arrive after a certain cutoff time. This reflection pattern is used in both blocks and is studied in the next two subsections.

A digital filter set is used in both blocks. These filters are designed by the overlap-add FFT method using a half window size overlap. The window size is 1024 samples in the first block and 4096 in the second. The filters were designed to compensate their inherent group delays.

2.1. Direct Sound and Early Reverberation

The attenuation of each ray is approximated by summing octave-band-filtered impulses weighted by the combined reflection coefficients of the walls in which the ray is reflected. The low-frequency and high frequency ends of the audible range are filtered by shelving filters instead of the band-pass ones. In Fig. 3 the filtered impulses corresponding to the five upper bands are shown.

This method is faster than that of the window method for FIR filter design because each filtered impulse can be pre-calculated and re-used for each ray doing just one multiplication and one addition for each frequency band and each sample of the filtered impulse. Furthermore it has been checked that when no weighting is applied before summation, the original impulse is acceptably well reconstructed. The reflection coefficients are approximated for each incidence angle by Rindel's model.

Figure 4 shows the impulse response of a ray and Fig. 5 depicts its frequency response.

The frequency response presents a staircase shape because of the filtered impulse technique used. This is not important since the difference with a smoother response cannot be easily perceived.

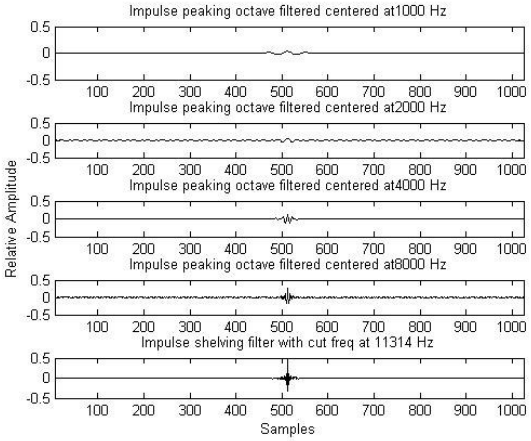

Figure 3: Set of filtered impulses.

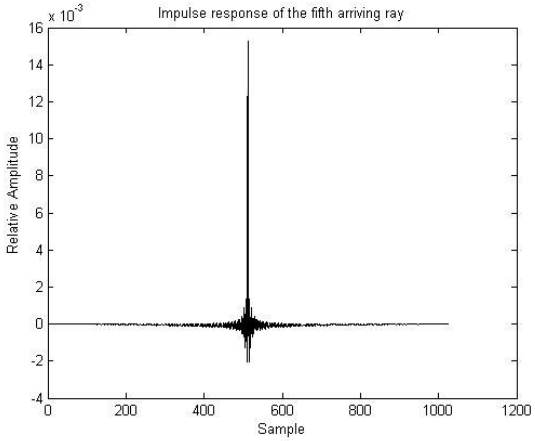

Figure 4: Example of a ray impulse response.

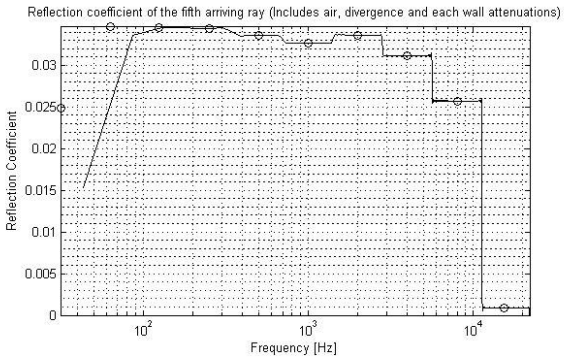

Figure 5: Frequency response of ray from Fig. 4. Line: Fourier Transform (2048 points) of the modelled impulse response. Circles: Combined reflection coefficients of the walls in which the ray is reflected.

If the reflection coefficient of each wall is approximated from random incidence absorption coefficient, fewer data are required and they are readily available in the literature for a wide variety of materials.

2.2. Late Reverberation

LR is calculated from random noise because the random phase field can be modelled as the gross spectrum and

temporal characteristics without causing wrong psychoacoustic impressions [3]. A decreasing exponential envelope must then be applied to account for the reverberation decay, as explained in 2.2.2.

The empirical Hidaka model of t_L has been derived for different music rooms which have some common acoustic properties and does not explain all of these rooms. Considering those differences, a larger t_L is used. Instead of 8 % we use 12 % RT_{500Hz} until a better model is available.

2.2.1. Transition block

Since ER ray density is low, LR begins with a transition period from ER during which random values taken from a bimodal distribution are assigned to the discrete times n_i corresponding to the successive ray arrivals obtained from the ISM. Eq. 5 ensures, hence, a smooth transition. The bimodal distribution is accomplished by adding a two-valued distribution and a gaussian distribution. Both modes turn to be quasi-gaussian, as seen in the upper panel of Fig. 6. The standard deviation of the gaussian component models to some extent the free-path variance of the rays.

2.2.2. Ending LR block

From t_c onwards, all samples are assigned random values with a gaussian distribution until an instant when the octave with the longest decay curve reaches a given threshold.

A schematic form of both parts of the noise from which the entire LR is calculated is shown in the Fig. 6.

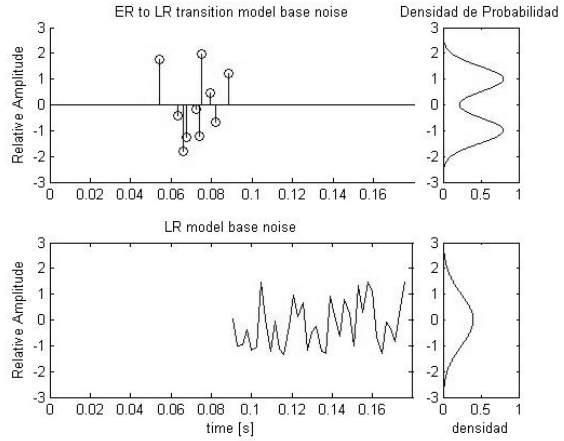

Figure 6: Example of both segments of an LR model before applying an envelope.

In order to prevent an energy leap at t_c, the standard deviation of the transition and ending blocks of the LR must be equal, since energy is proportional to the variance. This is achieved taking

$$m = \sqrt{\sigma_2^{\,2} - \sigma_1^{\,2}} \tag{8}$$

where m is the mode and σ_1 and σ_2 are the standard deviations of the gaussian components of the transition and ending parts. Fig. 7 shows the energy at 30 ms intervals. As can be seen, there is no energy leap at t_c.

It must be noted that this energy evolution is previous to applying any time envelope such as described in next section.

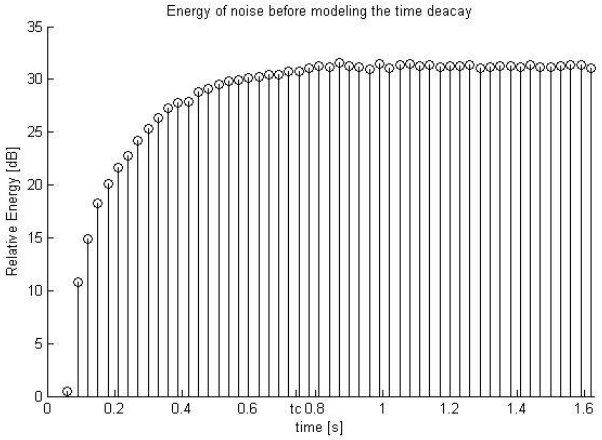

Figure 7: Evolution of LR energy before applying an envelope for modelling time decay

2.2.3. LR Decay

The frequency dependent decay of the entire LR, from t_L to the end of the RIR (including both the transition and ending blocks), is modelled by filtering the noise (Fig. 6) with filters similar to those used in the ER section, then applying an envelope to each one and finally summing all of these filtered and enveloped signals.

The envelope of each filtered noise is given by the statistic model using Eq. (6).

2.3. ER to LR Amplitude Balance

To compute the balance between both blocks of the RIR, the transition time t_L is extended to reach at least six rays more than that necessary to achieve a random field. Then the amplitude of the second part is scaled to reach the maximum value of those six ISM rays for the same rays modelled with the noise, and finally the ISM rays are kept.

Figure 8 shows the LR and the ER models in the same time line to illustrate how LR is scaled.

3. PRELIMINARY TESTS

Incorporating a head related transfer function (HRTF) for each of the ER rays some initial tests have been done.

Before applying the HRTF, some acoustic parameters have been computed. For instance, the reverberation time (T_{60}) of each octave frequency band matches, for the test room, the expected one from the Eyring statistical model. The major difference is an 8% for the 125 Hz frequency band.

The HRTF was incorporated by convolving measured responses from the public-domain database of the CIPIC of the IEEE [2] with each ray. These data are used to take into account the reflections, attenuations and difractions affecting each ray in the human body

before arriving at the eardrum. This information, with some uncertainty not discussed here, is decoded in the human auditory system and interpreted at the brain as the sound direction. Although each person has an individual HRTF and aural simulation made with anybody else's HRTF would not be proper and would cause differences in predicting the localisation of the sound source, an experiment was conducted for four subjects different from that of the HRTF used and good results have been achieved.

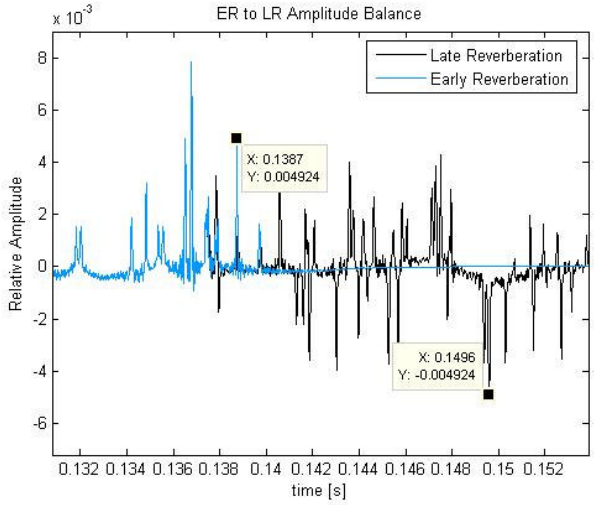

Figure 8: ER to LR amplitude balance.

Even more, it has been shown that the preesence of reverberation adds some error in the subjective detection of sound direction, but the reverberation used in these tests was short (0.25 seg @ 500 Hz).

Several anechoic hand clap recordings processed with one RIR, were presented with audiometric earphones. The subjects had to answer whether the simulation was realistic or not in a Likert five point scale and locate the claps in an interaural coordinate system (elevation and azimuth instead of azimuth and elevation).

For the realism question one subject answered in the best point of the scale and three in the step before. For the sound direction they all answered within a ± 20º margin of the RIR model direction for elevation and azimuth.

4. CONCLUSIONS

A fast and controllable algorithm to calculate box-shaped RIRs has been developed. It incorporates new techniques such as a faster model of the ER in comparison with the window method models and a convenient model for taking into account the LR smooth ray density growing. Additionally, modelling LR from the transition time t_L between ER and LR gives support to the LR model since after this time the sound field could be treated as random incidence.

The preliminary subjective tests give the expected results, making this model stronger for this kind of evaluations.

Future work can extend this model for arbitrary-shaped rooms and incorporate the materials reflection pattern information as well as diffusion for ER.

REFERENCES

[1] Accolti Ernesto and Miyara Federico. Combinación Digital Controlada de Ruidos Diversos. VI Congreso Iberoamericano de Acústica - FIA Bs As 2008.

[2] Algazi V. R., Duda R. O. and Thompson D. M. The CIPIC HRTF database. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, NY 2001.

[3] Hidaka Takayuki, Yamada Yoshinari and Nakagawa Takehiko. New definition of boundary point between early reflections and late reverberation in room impulse responses. J. Acoust. Soc. Am. 122 (1), 326-332, 2007.

[4] Kuttruff, H. Room Acoustics. Spon Press, Fourth Edition, 2000.

[5] Lehmann Eric A. and Johansson Anders M. Prediction of energy decay in room impulse responses simulated with an image-source model. J. Acoust. Soc. Am. 124 (1), 269-277, 2008

[6] McGovern Stephen G. Fast image method for impulse response calculations of box-shaped room. Applied Acoustics 70, 182–189, 2009

[7] Mitra Sanjit. Digital signal processing: A computer based approach. McGraw Hill, 1998.

[8] Rindel J. H. Modelling the Angle-Dependent Pressure Reflection Factor. Applied Acoustics 38, 223-234, 1993.