Audio Engineering Society

Convention Express Paper 399

Presented at the AES 159th Convention

2025 October 23-25, Long Beach, CA, USA

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Impulse response Synthesis in simulations with simple models

Paolo Martignon[1], Mario Di Cola[1]

[1] *Contralto Audio, Via Frentana, 17 66043 Casoli (CH) - Italia*

Correspondence should be addressed to Paolo Martignon (p.martignon@contralto-audio.com)

ABSTRACT

One of the potential applications of simulators for architectural acoustics is auralization, i.e., the creation of impulse responses of the simulated room (RIR, Room Impulse Response) with spatial content, such as Binaural (BRIR) or Ambisonics RIR. One of the challenges with ray-based simulations is achieving a sufficient temporal density of ray impacts on the receiver to ensure good quality in the simulated RIR. In many cases, the impacts may be adequate for an accurate simulation of the ETC (Energy Time Curve), but not for a perceptually convincing quality of the RIR in its final waveform. A method is proposed to synthesize RIRs with perceptual plausibility, even when only a limited number of impacts is available, while preserving the ETC behaviour.

1 Introduction

Geometry-based acoustics methods—such as Ray Tracing, Pyramid or Cone Tracing, and Particle Tracing—are widely adopted in most simulation tools for the mid-to-high frequency range, often in combination with wave-equation solvers for the low-frequency domain (hybrid models). A recurring challenge of these methods lies in achieving a sufficient temporal density of ray impacts at the receiver, which is essential to ensure high-quality simulated room impulse responses (RIRs). While the number of impacts may often be adequate for an accurate estimation of the Energy Time Curve (ETC), it is frequently insufficient to produce RIRs with perceptually convincing waveform quality.

For instance, in the Pyramid Tracing method implemented by Prof. Angelo Farina in his software *Ramsete*, corrective amplitude factors are introduced to compensate for a systematic underestimation of ray impacts after a critical time threshold. This adjustment restores the accuracy of the ETC but does not help in yielding perceptually plausible RIRs. More generally, with other ray-based methods, simplified modeling of diffusion (which always interacts, to varying degrees, with specular reflection) or a limited number of rays or particles can lead to similar shortcomings.

A commonly adopted approach in many simulation tools is to reconstruct the RIR by summing Dirac delta impulses, appropriately weighted and equalized according to the simulated spectral content of each impact and positioned at their simulated arrival times. However, this technique is prone to excessive temporal gaps when, as mentioned before, the number of evaluated impacts is too scarse, resulting in an unnaturally sounding RIR. The method proposed here addresses this issue by replacing individual impulses with pre-shaped signal "grains," selectively applied, only where needed, to fill such gaps.

The paper describes the method used for calculating the impacts, derived by a (slightly simplified)

implementation of Farina's *Ramsete* propagation algorithm. Then, the "white grains" technique for RIR synthesis is outlined, together with the method used for its binaural rendering.

2 Acoustic propagation

As mentioned in the introduction, the paper proposes a method to synthesize a RIR from a list of impacts on the receivers that describe the acoustic response of the room. Each impact is supposed to carry the following information: arrival time, amplitude spectrum in given octave fraction resolution, arrival direction. This description, that from now on will be addressed like "reflectogram", is retrieved in our simulator from an algorithm of "cone tracing", as a simplification of the "pyramid tracing" used in the software *Ramsete*, which detailed description can be found in [1, 2]. Here's a synthetic description of the simulation model and reflectogram creation procedure.

There are three basic elements:

- one (or more) point source featuring 3D directivity
- a 3D indoor environment made of surfaces, each of which featuring a spectral absorption coefficient (Alpha) and a scattering coefficient (Gamma)(this latter coefficient can be treated as surface dependent or global)
- one (or more) point receiver.

A number N of even distributed rays are sent from the source, each carrying information about their emission direction, covered distance and spectral energy reduction due to the history of wall bounces absorption/scattering.

At every bounce, the ray is specular reflected on the impacted surfaces and the carried energy spectrum is updated, as well as the distance covered, and its path goes on this way.

At every bounce, a check is made to decide if the bounced ray is impacting the receiver: if yes, an event is registered in the reflectogram, with arrival time given by the covered distance, and amplitude spectrum influenced by bounces history, original emission direction (based on source directivity), spherical divergence and air absorption.

At every bounce, despite the specular reflected ray impacting the receiver or not, a scattered contribution is added to the reflectogram, with amplitude spectrum handled by the scattering coefficient Gamma.

The mechanism and energy balance between specular reflections and scattering are implemented strictly following the concept and the equations proposed by Farina in [3].

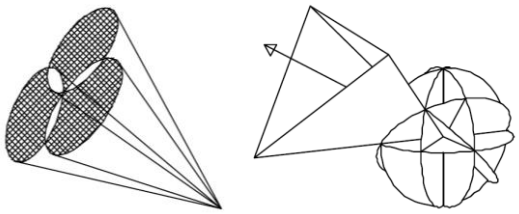


Figure 1. Pyramid Tracing (right) against Cone Tracing (left). Pyramids have no intersections.

In *Ramsete* Pyramid Tracing method, rays are considered to be axis of pyramids, and the impact check is done controlling the receiver point position with respect to the pyramid, internal (impact) or external (no impact). This optimizes the complementarity among rays, minimizing the risk of zero or multiple impacts on a receiver where just one impact should be there (see figure 1). For now, we are considering ray to be just cones, which is much simpler in ray-receiver impact check; it is less accurate (may partially mistake the first reflection structure in the reflectogram) but is not supposed to make us loose generality concerning the topic treated in this paper. The direct field contribution is calculated separately and is not prone to these ambiguities.

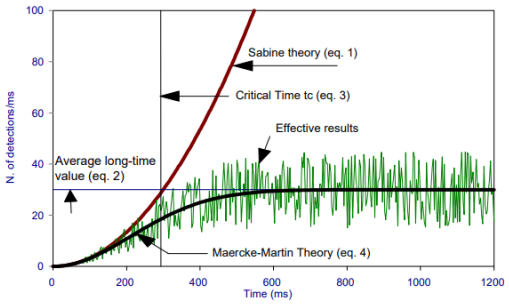


Figure 2. Time density saturation at critical time in Pyramid (and Cone) Tracing.

A key feature of Ramsete algorithm is that the considered pyramids (or cones in our simplified version, without loss of generality) increase their width while propagating, up to a point where it is comparable with the enclosure dimensions. At that point, called critical time, all the bounced rays always hit the receiver, automatically creating the effect of a diffused field, but saturating the time density of impacts to a constant value, depending on the numbers of rays N (see Figure 2). In the real world, such density should instead grow exponentially in time. Ramsete introduces a useful amplitude

correction factor, which, under certain assumptions, adequately compensates for the lack of energy deriving from the lack of impacts. We implement the same correction in our simulation. As mentioned earlier, this allows for a good estimation of the ETC (with the time bins method explained at paragraph 4), but the information needed for synthesizing a well sounding RIR may happen to be insufficient.

3 From reflectogram to RIR using the "white grains" technique

Once a reflectogram is defined, a commonly adopted and intuitive approach is to reconstruct the RIR by summing Dirac delta impulses, appropriately weighted and equalized according to the simulated spectral content of each impact, and positioned at their simulated arrival times. If the reflectogram is poor, this technique is prone to excessive temporal gaps, resulting in an unnatural sound for the RIR. The method proposed here addresses this issue by replacing individual impulses with pre-shaped signal "grains," selectively applied to fill such gaps only where needed. These grains have variable lengths, ranging from simple impulses (see. Fig.3) to noise bursts (see fig.4); but, independently on their length, they are individually scaled to have a well-defined total energy content, allowing them to be combined (via weighted summation) in a way that preserves the simulated spectral properties of each impact.

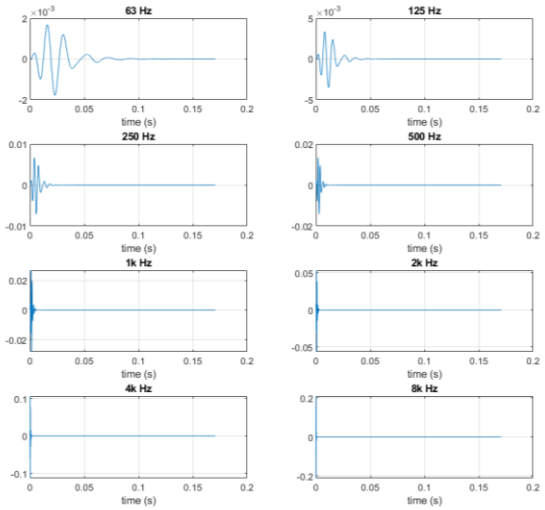


Figure 3: Impulse derived grains used for direct field and primary reflections; they are just the band filtered version of a Dirac Delta.

The grains database includes multiple randomized variants, even among equal length grains, generated using pseudorandom noise with different seeds, to avoid possible perceptual artifacts caused by repetitive sequences. Depending on arrival time and frequency band, different grain types are selected: longer grains are applied only where strictly necessary—namely, in the late reverberant tail and not in the early reflections, nor at sufficiently low frequencies. Preserving sharp impulses in early reflections and low-frequency bands, is crucial for simulating deterministic phenomena such as resonances and comb-filtering, consistent with the fidelity of the propagation and architectural models.

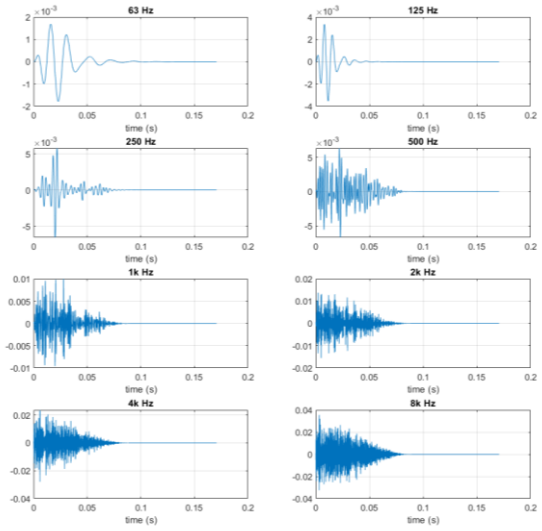


Figure 4: One example of grain set used for later contributions. Compared to the impulse derived set, they have increased length only for the mid-high range.

Having these "white grains" base available, already divided in frequency bands, the new approach for RIR synthesis can be summarized, for each impact, in the following three steps:

1. Chose a suitable "grains" set depending on the arrival time
2. Perform a weighted sum of the chosen grains set over the whole audio band, basing on the amplitude spectrum of the current impact
3. Sum the obtained broadband grain to the RIR at the impact arrival time.

For 1) we specifically split the RIR into three parts: direct field plus first reflections, transition and reverb. In the first part, the only impulses derived grains set

is used (the one shown in Figure 3); for the transition part, grains derived from gradually longer bursts (at mid-high frequencies) are chosen, proportionally to the specific arrival time; in the third part, maximally long grains are constantly used, but always renewed at every impact in terms of the "seed" used for burst random noise generation (several variations of the set shown in Figure 4).

4 Case study with focus on ETC curve

As an application example of the proposed method, we simulated the acoustics of a typical theater shape, with our Cone Tracing algorithm, using 100 rays. Realistic absorption coefficients spectra were attributed to the wall surfaces. Figure 5 shows all the bounces points on the walls: it gives a general idea of the propagation of the bouncing rays and, at the same time, reconstructs the theater shape in a pointillist painter fashion. The two little circles represent the source (close to front wall) and the receiver (on a gallery seat).

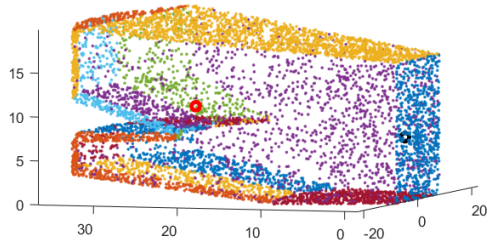


Figure 5: theater shape reconstructed though ray bounces points on walls

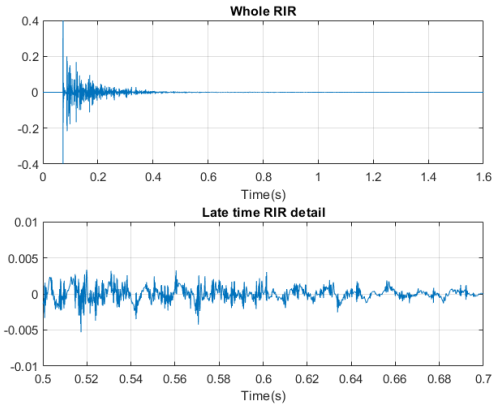


Figure 6: RIR synthesized using the only impulse-derived grains set (shown in Figure 3)

IMPULSE RESPONSE

As a reference for comparison, Figure 6 shows the RIR synthesized using only impulse derived grains set, the ones shown in Figure 3.

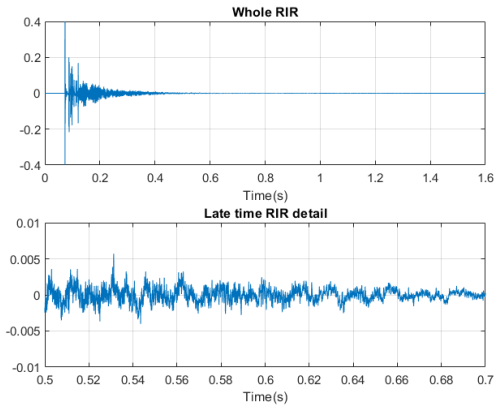


Figure 7: RIR synthesized with "white grains" technique.

The RIR obtained from the white grains technique is shown in Figure 7. The "filling effect" is clearly observable. Listening to the two RIR, even before convolution/auralization, gives a very simple and net result: the one enriched with grains is generally plausible at listening, whilst the other sounds unacceptably "scrambled", "crunchy".

ETC

As mentioned earlier, the philosophy of this approach is to optimize the RIR maintaining unaltered the ETC that arises directly from the simulation (reflectogram). From the synthesized RIR, the ETC can be estimated as its squared value, smoothed on a proper time constant (method A). The ETC from the reflectogram, instead, can be reconstructed dividing the whole response in equal time intervals (bins), and simply adding the energy contribution of each impact to the right bin, depending on its arrival time (method B). We chose a duration of 50ms for both the time constant of method A and the bin of method B, as a very rough and frequency independent estimate of the human ear integration time.
Figure 8 shows the superimposition of the ETCs derived with the two methods, using a RIR synthesized with the only impulse-derived grains set of Figure 3.
Figure 9 shows the same but having used the "white grains" technique for RIR synthesis.
As a result, the ETC does not look to diverge from the former case in a very meaningful way.
Although not shown here, it is intuitive from these charts, and verified by the authors, that even more

negligible changes arise in the derived Schroeder integrals and in the related reverberation times.

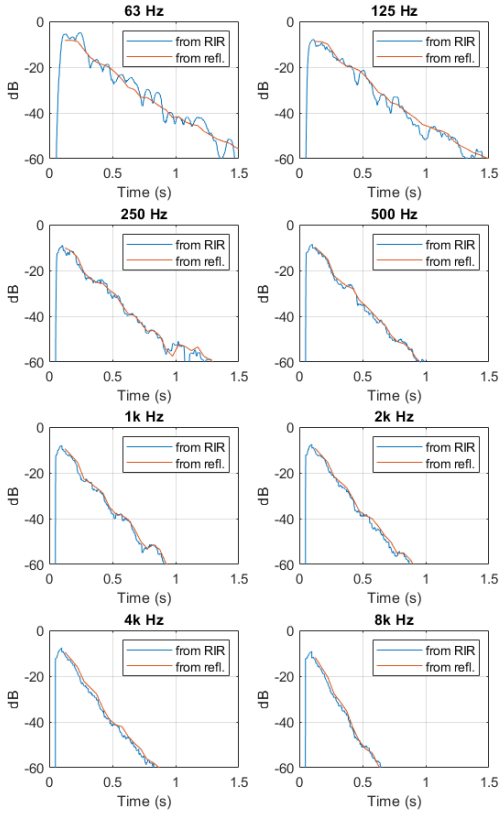


Figure 8: ETC from RIR (method A, blue curve) and from the reflectogram (method B, red curve). RIR is synthesized from the only impulses-derived grains set of Figure 3

PHASE BASED EFFECTS

Both cases show local differences between A and B ETC derivations in the low frequencies: in the 63 and 100 Hz octaves, the synthesized IR ETC oscillates around the base (reflectogram derived) ETC. This is perfectly normal and shows the beats typical of low frequency decays. The phenomenon is identical in the two cases, just because the grains used at low frequencies are always, on purpose, the same impulse-derived version, despite their arrival time, as explained at the previous paragraph and shown in Figures 3 and 4.
If the employed ray-based simulation model was considered still suitable at low frequencies for the simulated environment – namely: big and flat walls and with mostly real impedance – we should consider this beating effect realistic. Within the above assumption - that is a sufficient accuracy of the ray model in the frequency range considered - we could say that other phase-dependent summation effects are possibly realistic. This may be the case of, e.g., the comb filtering effects that can arise from strong isolated first reflections, or the cancellation of a crucial frequency for a subwoofer placed close to an isolated wall, or, under more assumptions and approximations, low frequency stationary waves.
Pulse-derived grains, as opposed to noise burst derived white grains, are used for early times and for low frequency, just to allow the possibly realistic simulation of these phase-based effects.

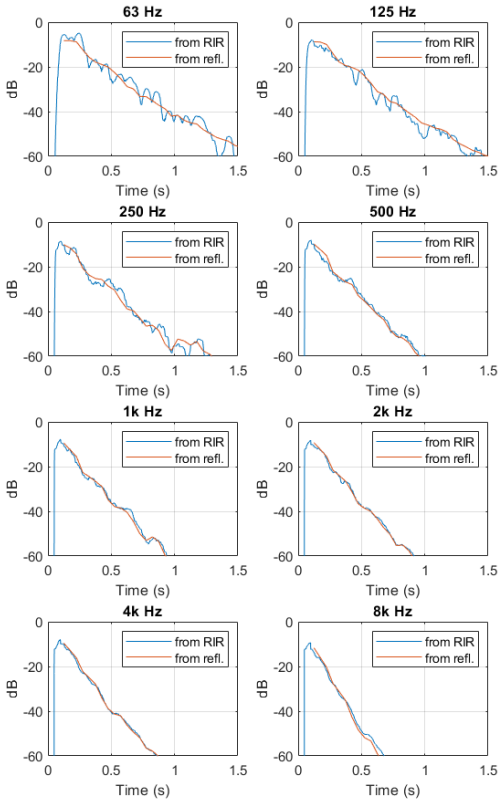


Figure 9: ETC from RIR (method A, blue curve) and from the reflectogram (method B, red curve). RIR is synthesized with the “white grains” technique.

5 Auralization with spatialized RIR

The described technique, applied to the entire reflectogram, leads to the synthesis of a monophonic RIR. To export a RIR with spatial content, the

reflectogram contributions are instead discriminated by arrival directions. An imaginary sphere, centered on the receiver, is divided into 48 equal area and complementary triangles, each representing an arrival direction interval. A 48 channels audio file is then synthesized, each channel corresponding to an incoming direction interval. To do so, at every impact, a grain is added only to the channel it belongs, basing on its arrival direction.
Each channel may be looked at as the impulse response of a virtual source placed at an angle corresponding to the arrival direction associated to the channel. Commercial VST plugins can be used to: 1) convolve the 48 channels with an anechoic signal, representing what is played by the original source 2) encode a Higher Order Ambisonic (HOA) stream (i.e. 16 channels stream for 3rd order), by entering the 48 virtual source web at the right directions in the encoder plugin 3) rendering of the ambisonics stream to a 2 channel binaural stream via HRTF.
A 3rd order Ambisonic RIR could be directly encoded off-line so to reduce the real-time processing effort.
Several listening tests have been done by the authors, with several different environments, anechoic signals and simulation parameters, while developing the basic method. Blind tests with other subjects will be organized when the simulation technique is more refined.

6 Conclusions

A technique has been developed and tested for ameliorating the synthesis of RIR out of ray based acoustic simulations. The method, based on "white grains", has demonstrated to be able to obtain much more plausible sounding RIR, compared to a simple pulses-based approach, from the same reflectogram; such reflectogram being considered sufficiently accurate for ETC assessment. The technique was outlined in detail focusing on ETC study.
One of the goals of the study was to keep separated two aspects of acoustic simulations and their related needs in terms of effort and complexity (e.g. number of rays, sound diffusion model). The first aspect is ETC, which, alone, allows the calculation of very important perception related parameters like reverberation time, direct to reverberant ratio, Clarity, STI (Speech Transmission Index), Strength and equalization "tilt", and can embody phenomena like presence of echoes or double slopes due to coupled volume. The second aspect is the specific wave shape of the RIR for auralization, which should have the role of translating those perception features for the human ear, without altering the related parameters, and trying to avoid any artifact or unnatural effect.
We also consider that, in general, any simulated RIR will always be much more distant from the real world compared to ETC, because of the "infinite" degrees of freedom involved. At the light of such evaluations, we propose a method which lays halfway between simulation and synthesis: that is, it tries to synthesize a good sounding RIR that entirely reflects the simulated ETC. At the same time, the synthesized RIR aims at reproducing the potentially realistic phase-based effects deriving from deterministic impacts arrival times, at early times and/or low frequencies.
Future work is aimed at investigating the advantages of ray-based acoustics simulations (improvements of our current model), and related auralizations with the new method, using simple architectural models, to take decisions about audio system installations in the pro audio field.

Acknowledgments

A particular thanks goes to Ing. Daniele Ponteggia for the profitable and enlightening conversations on the topic.

References

[1] A.Farina, "RAMSETE - a new Pyramid Tracer for medium and large scale acoustic problems", *Proc. of EURO-NOISE 95 Conference*, Lyon 21-23 march 1995A

[2] A.Farina, "Pyramid Tracing vs. Ray Tracing for the simulation of sound propagation in large rooms", In the volume *"Computational Acoustics and its Environmental Applications",* pp. 109-116, Editor C.A. Brebbia, Computational Mechanics Publications, Southampton (GB) 1995

[3] A. Farina, "Validation of the numerical simulation of the scattered sound field with a geometrical pyramid tracing approach", ACOUSTICS 2000 - IOA Conference, Liverpool (GB) 17-18 April 2000

[4] Michael Vorländer, "Auralization". RWTH Aachen University