19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

NEW METHOD FOR AURALIZING THE RESULTS OF ROOM ACOUSTICS SIMULATIONS

PACS: 43.55.Cs

Angelo Farina[1], Simone Fontana[1], Simone Campanini[1]
[1]Laboratorio di Acustica ed Elettroacustica, Parma, farina@pcfarina.eng.unipr.it

ABSTRACT

The results of most room acoustics simulation codes are usually energetic reflectograms with limited time resolution (usually 1ms or worst) and frequency resolution (typically octave bands, in a few cases 1/3 octave bands). When these results are employed for auralization, it is necessary to derive a set of high resolution impulse responses, providing narrow time/frequency detail, and containing spatial information (Binaural or Ambisonics).

A novel process for creating these hi-res impulse responses is described, making use of sine-bursts synthesis (a sort of wavelet synthesis). For each octave band and each time slot, a number of sinusoidal bursts are generated. The time-domain envelope of these bursts is properly shaped and overlapped, so that they merge smoothly. Their amplitude is modulated in such a way that the computed energetic parameters (T60, Clarity, etc.) are perfectly reconstructed.

For binaural rendering, each tone burst is convolved with the HRTF corresponding to the arrival direction. For Ambisonics rendering, each tome burst is copied on all the channels, with gains derived by the Ambisonics encoding formulas corresponding to the direction-of-arrival.

Blind listening tests performed in a special listening room demonstrated that the new auralization technique outperforms the previously employed one, which was based on octave-band filtered noise bursts instead of multiple wavelets.

IMPULSE RESPONSE SYNTHESIS

The problem of the impulse response (IR) synthesis appears every time we want to judge by ours own ears the result of a room acoustic simulation.

Usually ray or beam tracing software packages model the sound field propagation by applying the geometric approximation. The user must define at least one sound source, one receiver and the acoustical properties of every internal surface of the room; then, the algorithm simulates the power irradiated by the source and annotates, instant by instant, the portion intercepted by the receiver.

The output of the simulation process is a table containing the energy values captured by the receiver for every time instant, at every frequency (usually in octave bands); this table is called energy IR or echogram. In Ramsete [1], the output is a table of ten echograms: every column corresponds to an octave band. It's impossible to use directly an echogram for an audio convolution operation and this for many reasons: first of all the sampling time normally adopted for acoustic simulation is too big, typically from 1 ms to 10 ms. Another question is the absence of phase information in the simulation and, consequentially, in the echogram. Finally, we get 10 octave-band echograms, and we need to merge them in a single, wide-band impulse response.

The synthesis process of an audible IR, therefore, needs two interpolations, in the time domain and in the frequency domain (we have only 10 echogram-samples for the entire audio band). In more practical words, when we reduce the sampling time to more useful values, i.e. 22.67 μs for 44.1 kHz sampling rate, a zero sequence appears between two samples in the echogram (fig. 1), a sort of vacuum that must be filled. A simple interpolation is not sufficient because the phase don't exist: it must be re-created.

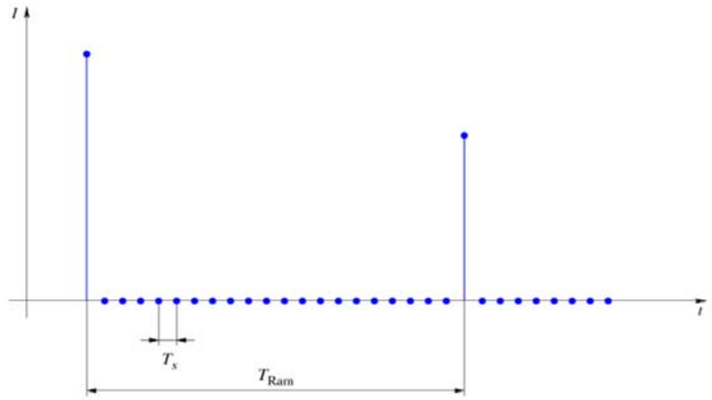

Figure 1 - Oversampling effect: zero-sequence appearance between two echogram samples

The Farina's software AudioConverter uses band filtered white noise bursts: these are wide band signals that covers the whole frequency range. AudioConverter is robust and reliable in the acoustical parameters calculations, but white-noise-IR presents at least two limits: a "synthesized" sound and a very irregular(fretted) state in the transient, when the IR is used for Acoustic Quality Test. Both the problems are originated by the discontinuities of white noise.

From this observation the idea to substitute the noise with another signal having lower discontinuity, in particular the choice was the sum of sinusoidal bursts, 12 for every octave-band and for every time-slot of the energetic echogram table; the reverberating field stochastic behavior was preserved by assignment of casual phase to every burst generated. Beside, to avoid the appearance of a discontinuity in the joint point of two consecutive time-slots, the bursts are made longer than the original time slots, and were partially overlapped and windowed with a raised cosine flat-top window.

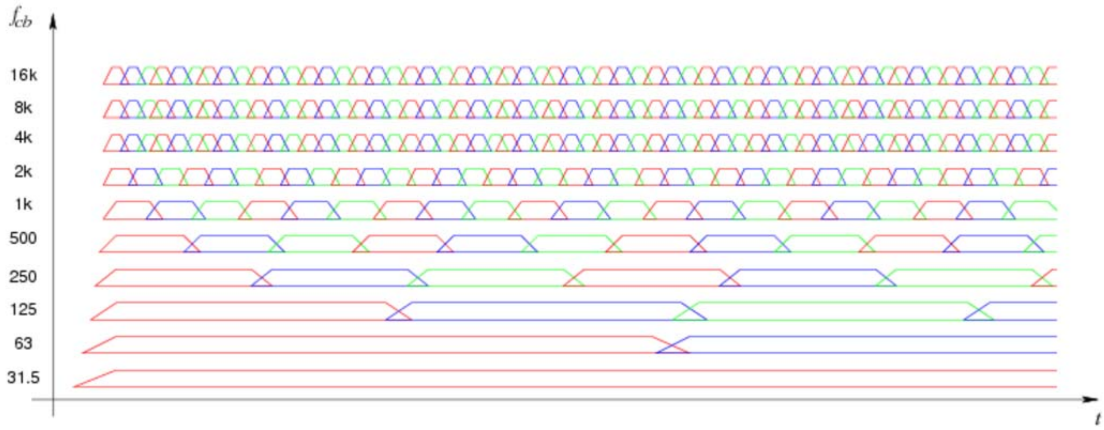

Figure 2 - IR building scheme: every row correspond to one echograms interpolated with bursts (the 'trapeziums') obtained by the sum of 12 sinusoidal signals. The desired IR is computed summing all rows.

Because the sinusoidal signals characteristics, it was absolutely fundamental to build variable length bursts, where the length is a band function: 100 ms for the 31.5 Hz band, 50 ms for the 63 Hz, etc. But in this way, the bursts are much longer than the time slots of the echogram: a statistically reasonable solution, thus, was to take the mean value of the square root of all the time slots covered by the burst length (fig. 2).

BINAURAL IMPULSE RESPONSE SYNTHESIS

When we want to achieve spatial room auralization [2], (Binaural or Ambisonics) we have to produce an IR containing spatial information, based on some geometrical information: the wave propagation direction and the exact flight time are the most important; with these data we can reconstruct the phase of every reflection in a credible way. Ramsete can save all this information for the first 15-20 reflection orders in a text file where we find also the octave-band energies for every reflection. Thanks to this option, it was possible to extend the synthesis algorithm functionalities to spatial rendering, in Binaural (stereo, 2 channels) and Ambisonic (B-format, 4 channels). In the present article we treat only the first, but the concepts are absolutely similar in both cases.

The adopted technique consists splitting the IR in two parts: an early part and a reverberant tail. The early part covers the first 80 ms from the direct wave, and contains most of the spatial information which can be decoded by human brain; the tail is not critical for these purposes and can be computed with the previously described algorithm, simply taking care to create two uncorrelated reverberant tails for the two ears. Then, we can focus the attention to the IR head synthesis, which follows these steps:

1. Each early reflection is modeled by a Dirac's delta signal.
2. This delta is filtered in 10 octave bands, and each of them is given an amplitude corresponding with the square root of the energy values detected by the receiver.
3. The amplitude-corrected filtered signals are summed back in a single wide-band signal, which now has the proper spectrum.
4. Computation of azimuth and elevation of the early reflection, and selection of the proper HRTF.
5. The processed delta is convolved with the HRTF, creating the signals at the two ears.
6. Sum of all signals obtained from all the early reflections for each of the two ears.

We decided to use the IRCAM's Listen project HRTF sets, because of their high quality.

Nevertheless, the IR's heads sound appearance wasn't fully natural, obviously due to the ideal delta modeling: for this reason we have added the possibility to import a user impulse in wav format.

In figure 3 we report one of the binaural impulses responses obtained with the sine burst method.

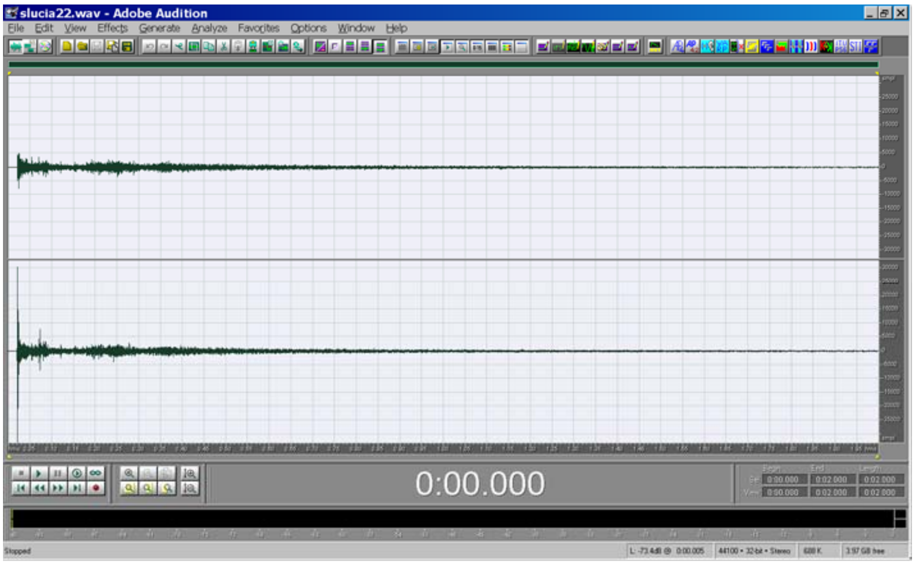

Figure 3 – Binaural IR synthesized with the sine burst technique.

LISTENING SYSTEM

Binaural signals are designed for headphone reproduction, but they can also be presented with loudspeakers, by means of crosstalk cancellation. Crosstalk cancellation (CTC) is the filtering process that allows the signals from the left and right loudspeaker to feed only the ipselateral ear, without reaching the contralateral ear. In this way the right channel is provided to the right ear, and so it is also for the left ear, exactly as in a headphone system. These systems are called Transaural™ systems.
One of the problems of CTC is the high sensitivity of cancellation to the listening position. CTC is normally designed for one position and little movements of the listener can quickly deteriorate transaural system performances. Recently a new system, called stereo dipole ([4]), has been proposed to solve this problem.

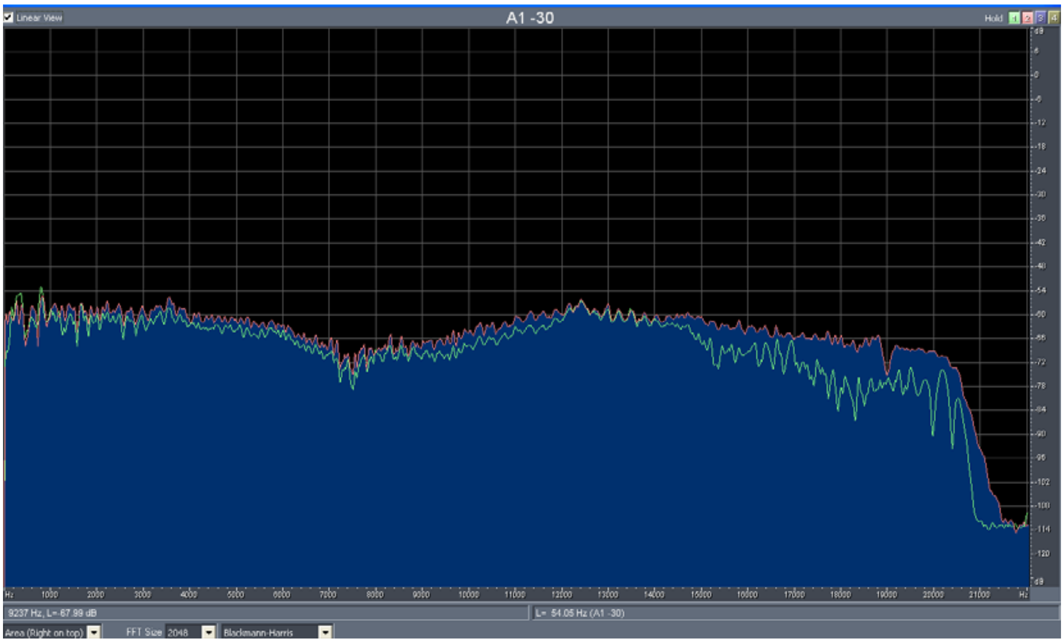

Figure 4 – Ipselateral transfer function unprocessed (red) e processed (green)

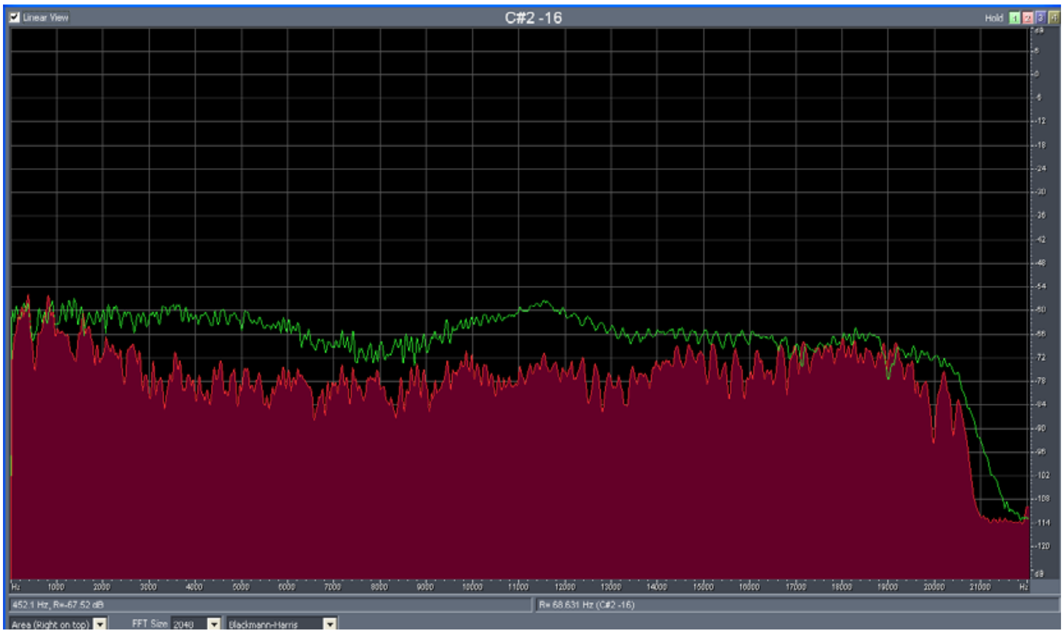

Figure 5 – Cross-path transfer function unprocessed (green line) e processed (red area)

The stereo dipole creation process consists in

1. Binaural impulse response measurement at the listening position. We obtained them using the Neumann Ku-100 Dummy head
2. Inversion of the binaural impulse response 2x2 matrix, to obtain cross-talk cancelling filters
3. Convolution of the input stereo signal with the 2x2 matrix of inverse filters

The inversion of the binaural impulse response matrix reveals particularly critical for timbre reproduction, due to the transient and the dynamics of the equalizer and the sensitivity of the equalization to the listening position; in our stereo dipole implementation we choose to

1. Cancel crosstalk only the direct sound without correcting for the room effect. This leads to shorter and smoother CTC filters.
2. To cancel crosstalk only in the range 1000-12000 Hz, in order to avoid bass boosting involved in CTC with stereo dipole and errors due to high frequencies (very unstable behaviour for little movements of the listening position).
3. Not try to equalize the loudspeaker, in order to leave unaltered the loudspeaker proper sound.

This solution reveals effective for correct timber reproduction, spatial stability and crosstalk cancellation.

The stereo dipole system has been setup in the Casa della Musica laboratory [5], with two loudspeakers spanning a 10 degrees angle in front of the listener; its performances are reported in figure 4 and 5. In figure 4 it is possible to compare the direct path before and after the processing. It can be seen that the signal stays almost unvaried, except for a little loss in high frequencies. In figure 5 it is possible to remark that crosstalk has been reduced of 10 dB in the useful band.

LISTENING TESTS

Listening tests have been carried out on 5 musician/musicologists at Casa della Musica.
Test has been organized as follows: the person is introduced in the listening room and he is asked to listen to binaural tracks, obtained convolving two anechoic tracks of a Denon CD (*Water Music* di *G.F.Handel e My funny valentine*, solo) with binaural impulse responses obtained by noise burst or sine burst techniques. It must be noticed that, while the sine burst method did employ a set of HRTFs which was carefully selected among the Listen database to best match the listener's own ears, the noise burst method did employ the old Kemar HRTFs available on the MIT-Medialab web site.

The binaural responses have been obtained from the results of the Ramsete model of the San Vitale church in Parma. In the model, the source is clearly on the left of the listener, which is placed in the centre of the church. Listeners can switch between the two response types in real time, in order to make the comparison easier, and they can listen to the samples as many times as they want. Subjects have been asked to fill a questionnaire; the evaluation parameters have been formulated according to [6], and can be visualized in figure 6, together with the test results.

RESULTS

Proper statistical analysis of the responses allowed to assess the quality of the auralization obtained from the RIRs computed according to the two methods, for all the listed parameters.

1. **Roundness and hardness**: one of the major problems of synthesized audio signal is surely the absence of softness and the appearance of sounds unpleasantly sharp: the observed tendency is a greater softness in the signals auralized with sine-bursts IR, above all in the song “My funny valentine”. Nevertheless, in both cases we didn't find sounds properly round and soft.

2. **Localization:** the source localization was significantly better in the sine-burst case. Maybe this is even due to the higher quality of the Listen's HRTF sets and to a HRTF customization step performed before the test. Unstable source localization in the noise-burst case have been reported.
3. **High frequency excess**: the testers had individuated a high frequency excess in the sine-burst songs, especially in My funny valentine.
4. **Low frequency excess**: absence of significant differences between the two techniques
5. **Pleasantness:** remarkable difference in the song “My funny valentine”: the testers appreciated the sine-burst IR. Slightly lower difference for the other song.
6. **Sound Realism**: we don't have reconstructed a completely natural perception yet, but the testers indicated some improvement from the noise-burst technique to the sinus-burst one.

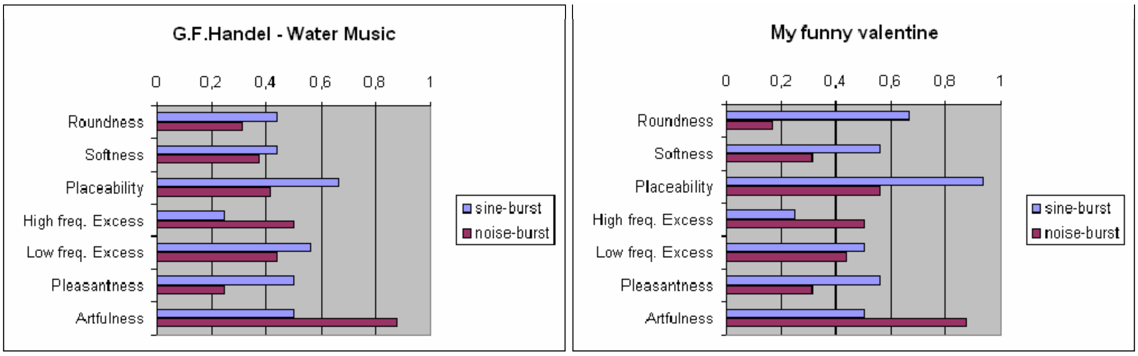

Figure 6 – Test Results: ‘1’ is the best mark

CONCLUSIONS

The present work was aimed to answer whether or not the sine-burst technique outperformed the noise burst technique now used in Ramsete for auralization purposes. Test performed on 5 subjects, using customized binaural auralization and reproduction through stereo dipole showed encouraging results.

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 1995

[2]. M. Kleiner, B.I. Dalenback, P. Svensson, “Auralization: an overview”. Journal of the Audio Engineering Society ,1993.

[3]. http://recherche.ircam.fr/equipes/salles/listen/

[4]. Kirkeby, PA. Nelson, H Hamada “The stereo dipole: A virtual source imaging system using two closely spaced loudspeakers”, Journal of the Audio Engineering Society, 1998

[5]. A. Farina, P. Martignon, A. Azzali, A. Capra “Listening Tests Performed Inside a Virtual Room Acoustic Simulator”, I seminario Música Ciênca e Tecnologia "Acústica Musical", São Paolo do Brasil, 3-5 November 2004

[6]. A. Farina “Acoustic quality of theatres: correlation between experimental measures and subjective evaluations”, Applied Acoustics, Volume 62, Issue 8, Pages 889-1023 (August 2001).