AURALIZATION SOFTWARE FOR THE EVALUATION OF THE RESULTS OBTAINED BY A PYRAMID TRACING CODE: RESULTS OF SUBJECTIVE LISTENING TESTS

Angelo Farina

Dipartimento di Ingegneria Industriale
University of Parma
Via delle Scienze - Parma (Italy)

SUMMARY

This paper deals with the evaluation of the numerical previsions obtained by a newly developed Pyramid Tracing computer program, named Ramsete [1,2,3,4]: as the results are in form of impulse responses, these can be compared with experimental impulse response measurements taken in the actual spaces that are being numerically modelled [5].

Another way to compare the numerical results with the experimental ones is by listening to musical pieces, spoken text, or other signals convoluted with the (pressure) impulse responses. This is possible if the energy impulse responses computed by Ramsete are converted in pressure impulse responses, by an original algorithm described here. This is the core of a new Auralization program, named Aurora.

Aurora is also equipped with a fast convolution software [6], running on a PC and not requiring any DSP board, that can process standard WAV files of unlimited length with binaural impulse responses up to 200000 taps (these can be both experimental or numerical). A standard 16 bit low cost PC audio board and high quality headphones are then used to submit the convoluted signals to judging panels (actually these are students of Parma's University), that have to fill up a questionnaire.

With proper statistical analysis of the questionnaires, the validity of the simulation software has been assessed: although part of the subjects were still able to distinguish between signals convoluted with experimental and numerical impulse responses, they gave on average the same subjective judgement regarding reverberation, bass and treble balance and spatial impression. It means that the simulated responses are not perfectly identical to the experimental ones, but they reproduce correctly the most important acoustic effects.

The research is now going on, searching to obtain reasonably good reproductions also through loudspeakers. On the other hand, the Pyramid Tracing algorithm is being updated, taking into account also diffusing surfaces and extended sources, and the convolver is continuously improved toward the goal of real-time computation.

CONVERSION OF ENERGY IMPULSE RESPONSES IN PRESSURE IMPULSE RESPONSES

Ramsete (as many other Ray Tracing or Beam Tracing programs) produces energy impulse responses of unlimited length, sampled at intervals typically of 10, 5 or 1 ms; a separate response is computed for each octave band (31.5 to 16000 Hz). To obtain reasonably good binaural simulations, two receivers must be located at the sides of a sound diffracting sphere, approximating two ears, as it is shown in fig. 1. In this way, the response in one receiving point is a couple of arrays, each of them with 10 columns (the frequency bands) and some hundreds or thousands of rows (the time intervals).

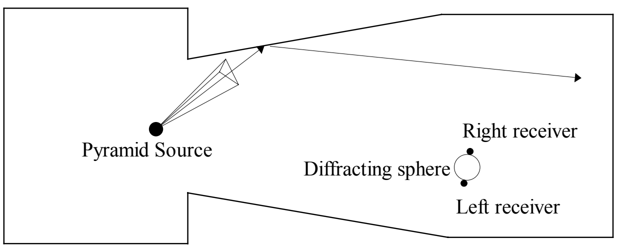

Fig. 1 - Pyramid Tracing Scheme with simulated binaural receiver.

It is necessary to convert each of these two arrays in a single pressure response, with a bandwidth of 20 Hz to 20 kHz, sampled at 44.1 or (better) 48 kHz. Usually these impulse responses are stored in standard WAV format, or in proprietary Impulse Response formats as the TIM format of the MLSSA system. Powerful software tools already exist, enabling any kind of mathematical manipulation of these pressure response files.

The conversion process begin with the generation of a sample of white noise with proper length and sampling rate, covering the impulse response duration. Ten copies of this white noise signal are made, and each of them is “modulated” with an amplitude envelope obtained by the square root of the energy impulse response in the corresponding frequency band. After this, each modulated white signal is passed through a 6-pole IIR digital octave filter, centred on the corresponding frequency. The ten filtered signals are then summed together, obtaining a wide-band pressure response. Eventually, the result is normalised and converted to 16-bit integer.

Although this process can be criticised in many ways, it is very fast and produces impulse responses that exhibit the same energy/time curves, the same reverberation times, and almost exactly the same objective acoustical parameters as the original energy responses. Obviously this is not enough to be accepted as realistic for audible simulations: a subjective evaluation need to be made for this.

A little problem exists about the overall frequency response. Often Ramsete simulations make use of a theoretical omnidirectional source (OMNI), having a sound power level of 110 dB in each octave band. It is also possible to use sound source files emulating many famous loudspeakers, with proper directivity balloons: also in this case the computation is conducted as the sound source was fed with pink noise, because this make simpler to look at the resulting spectra in term of overall frequency response. On the other hand, a pressure impulse response that don’t filter any frequency is inherently white, not pink. The pink-to-white conversion is automatically accomplished by the procedure described above.

But there are cases in which the Ramsete simulation is conducted with sound source data that are already white: for example if the sound source file was built by MLSSA measurements on a loudspeaker. In this case the above procedure introduces an artefact, giving a 3 dB/octave increase of the response with frequency: this can be removed setting an appropriate flag during the conversion process (“force pink”).

Another case is a numerical simulation conducted with an OMNI source, that is to be compared with experimental measurements obtained with a sound source not perfectly flat in the whole frequency range. The conversion software allows for this, enabling an octave-band equaliser to adjust each octave band prior of the summing. If the equaliser is adjusted by an octave analysis of the experimental impulse response (probably limited to the direct wave), this already takes into account the “whiteness” of the spectrum, and so it is required again to set the “force pink” flag to avoid a 3 dB/octave positive slope in the computed response.

FAST CONVOLUTION BY FREQUENCY DOMAIN PROCESSING

Although specialised hardware tools to perform continuous delayed convolution do exist since some years [7], they are still very expensive and don’t allow easily to digitally transfer the input and output signals on a PC; these devices are using Frequency Domain Processing with large blocks of data, resulting in a delay that is 3-4 times larger than the impulse response length. Time domain processing DSP boards, on the other hand, are capable of time domain convolution with no delay, but are limited to a few thousands of filter taps (in the better cases).

The Aurora system employs a very different approach: both the input and output data files are stored on the hard disk in standard WAV format and, once convolution is performed, comparative tests can easily be conducted with just a “point and click” delay.

Convolution is performed through the well known “select-save” algorithm [8]: details and performances of the convolution software were already published in [6]. Now the program has been extended to longer impulse responses (up to 200000 taps) and speeded up a lot.

The actual implementation of Aurora gives these computing times (in s) on a i486 DX2-66 PC, with a mono input signal of three different lengths and two (binaural) impulse responses of 5 increasing lengths:

Imp.Resp. length ->	16 kpoints	32 kpoints	64 kpoints	128 kpoints	179593 points
input 10 s, 44.1 kHz	61.85	79.0	108.2	160.4	209.6
input 20 s, 44.1 kHz	108.4	125.1	171.6	300.7	344.8
input 30 s, 44.1 kHz	156.3	183.4	244.6	359.7	488.8

Usually there is no matter in employing samples longer than 30s for comparison tests. It must also be noted that the program takes 10-20 s to initialise, performing the FFTs on the impulse responses; then the select-save loop begins. The I/O time due to disk access is always lower than 10% of the total computation time. The table shows that the effective CPU time spent in the select-save loop ranges from 5 to 15 times the input sample duration, and thus for real-time delayed operation a large speed improvement of the system would be required.

Some tests have shown a speed improvement of a factor 5 employing a Pentium-90, so the real time delayed processing can actually be obtained on such a machine for impulse responses of 16 kpoints at 44 kHz, or 32 kpoints at 22.05 kHz. Multiprocessor (Dual Pentium) machines have still not be tested, as this require a special multi-thread version of the software, that is under development now.

SUBJECTIVE TESTS TO VALIDATE THE SIMULATION SOFTWARE

The proper working of the convolution software was previously verified, comparing convoluted signals with direct recordings taken during loudspeaker playback of anechoic music in some theatres [6]. On the other hand, Ramsete was accurately tested, showing a good agreement between the computed objective energetic parameters and the experimental ones [3,4,5]. Subjective tests are thus required to validate the Aurora's energy-to-pressure conversion algorithm, that is the most delicate point of a computerised audible simulation based on geometrical acoustics assumptions.

Two enclosed spaces were employed for this test: the first is a church in Foligno (I), the second is a large sport arena in Modena (I). The same spaces have been used for objective parameters comparison [5]. Figure 2 shows a perspective view of the two rooms. Two receivers were chosen in the church, one very near the source (with high direct/reverberant ratio) and one in the far field. A single receiver was chosen in the sport arena, as the sound field was very similar everywhere.

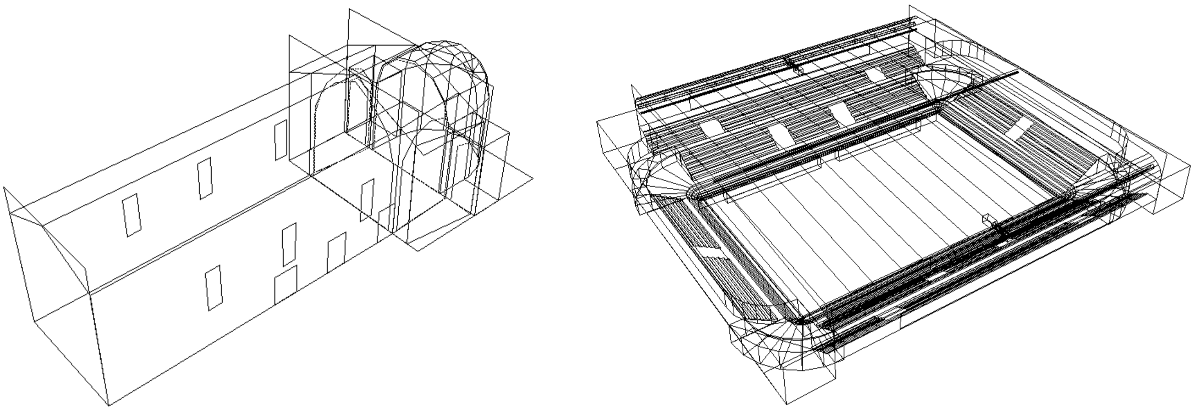

Fig. 2 - perspective view of the two rooms

Two anechoic samples were used for the test: a music piece of Mozart ("Le Nozze di Figaro", Ouverture. bars 1-18) and a poem of Leopardi ("Il Sabato del Villaggio"). The samples were digitally transferred from CD to hard disk through a Toshiba CD-ROM drive. They were convoluted both with the numerically simulated impulse responses and with the experimental ones. Thus 6 pairs of samples were obtained, in which the presentation order was randomly shuffled. To obtain comparison data, two "control groups" each of other 6 pairs of samples were mixed with the "true comparisons" set: 6 "really equal" pairs (obtained playing twice the same sample) and 6 "really different" pairs (obtained with random associations of samples coming from different rooms).

14 subjects were asked to listen to the 18 pairs, filling up for each pair the following questionnaire:

Pair no. .... Are the two samples A and B equal? yes ☐ no ☐

If Your response is no, explain why:

	a lot (-2)	slightly (-1)	no difference (0)	slightly (+1)	a lot (+2)	
A is more reverberant	☐	☐	☐	☐	☐	B is more reverberant
A has pronounced bass	☐	☐	☐	☐	☐	B has pronounced bass
A has pronounced treble	☐	☐	☐	☐	☐	B has pronounced treble
A has wide spatial impression	☐	☐	☐	☐	☐	B has wide spatial impression

The following table summarises the results of the first question (percentage of "equality"):

Ramsete/Experimental comparison	Truly equal samples	Truly different samples
23.0 %	71.8 %	2.5 %

These percentages show that Ramsete's simulations are actually fooling one subject over four. But let we look at what are the average differences and the standard deviations which are reported from the subjects, both in tabular and graphical format:

	Ramsete/Exp. comparison	Truly equal samples	Truly different samples
more reverberant	--0.06 ± 0.92	-0.03 ± 0.58	-0.36 ± 1.38
pronounced bass	-0.06 ± 0.77	-0.10 ± 0.41	+0.24 ± 1.30
pronounced treble	0.00 ± 0.72	+0.04 ± 0.41	+0.35 ± 1.07
wide spatial impression	+0.24 ± 0.87	+0.06 ± 0.40	+0.64 ± 1.22

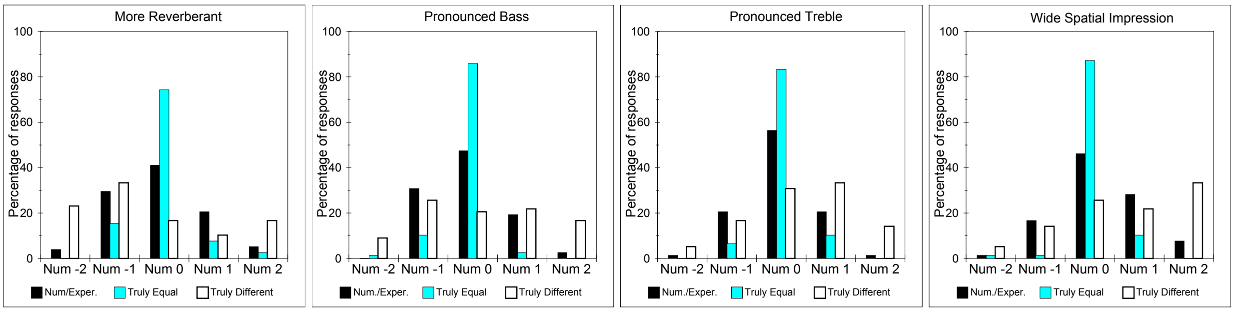

Fig. 3 - Distribution histograms of the four questions.

Surprisingly the Ramsete/Aurora's simulations are giving almost correct impression of the four subjective parameters, as they are judged on average nearly equal to the experimental responses. A greater difference is shown from the "truly different" control group (as obvious), while the "truly equal" group shows average values similar to the numerical/experimental comparison. This result, that could seem strange, is probably due to the fact that differences in "truly equal" responses are found by the most unreliable subjects, while differences between numerical and experimental samples are found by the most sharp-eared listeners. The standard deviation of the numerical/experimental comparison is greater than that of the "truly equal" pairs, while the standard deviation of the "truly different" group is still greater.

CONCLUSIONS

The subjective results show that the simulated responses are not perfectly indistinguishable from the experimental ones: a trained listener can easily identify the difference in a comparison test. On the other hand, these differences are actually very little, as the main subjective parameters are judged equal. Asking the subjects to specify the true nature of the difference, most of them were responding "I feel a slight difference, but I am not able to identify its cause".

Probably a wider subjective experiment is needed, with a larger number of subjects and including some other rooms with different acoustic behaviour. Also a better questionnaire can now be arranged, based on the improvements suggested from many subjects.

Nevertheless the Aurora system has shown its capability of producing realistic binaural simulations, that make it possible to appreciate even slight modifications in the numerical model of a room. Usually the goal of an auralization system is to provide an audible feedback to the designer, or to demonstrate the effects of proposed acoustic treatments of the room [9]: these capabilities are now available with a low-cost PC and no additional specialised hardware, with computation times short enough to perform dozens of consecutive tests, and with no practical limitation about the sampling frequency and the length of both the convoluted signals and the impulse responses.

REFERENCES

[1] Farina A., Cocchi A. - "Correzione acustica di ex-chiese riadattate per utilizzo concertistico: un esempio di progettazione di interventi non-sabiniani con l'ausilio del calcolatore" - Proceedings of *Acoustics and Recovery of Spaces for Music* - Ferrara, 27-28 October 1993.

[2] Farina, A. -"RAMSETE - a new Pyramid Tracer for medium and large scale acoustic problems", *Proc. of Euro-Noise 95*, Lyon, France 21-23 march 1995.

[3] Farina, A. & Maffei, L. - "Sound Propagation Outdoor: Comparison between Numerical Previsions and Experimental Results" - COMACO95, *Proc. of Int. Conf. on Comput. Acoustics and its Environmental Applications*, Southampton, England, 1995, Computational Mechanics Publications, Southampton 1995.

[4] Farina, A.- "Pyramid Tracing vs. Ray Tracing for the simulation of sound propagation in large rooms" - COMACO95, *Proc. of Int. Conf. on Comput. Acoustics and its Environmental Applications*, Southampton, England, 1995, Computational Mechanics Publications, Southampton 1995.

[5] Farina A. - "Verification of the accuracy of the Pyramid Tracing algorithm by comparison with experimental measurements of objective acoustic parameters" - Proc. of ICA95, Trondheim, Norway, 26-30 June 1995.

[6] Farina A. - "An example of adding spatial impression to recorded music: signal convolution with binaural impulse responses" - Proc. of "Acoustics and Recovery of Spaces for Music", Ferrara (Italy) 27-28 Oct. 1993

[7] Connoly B. - "A User Guide for the Lake FDP-1 plus" - Lake DSP Pty. Ltd, Maroubra (Australia) - Sept. 1992

[8] Oppheneim A.V., Schafer R. - "Digital Signal Processing" - Prentice Hall, Englewood Cliffs, NJ 1975, p. 242

[9] Kleiner M., Dalenback B., Svensson P. - "Auralization - An Overview" - AES Journ., vol. 41 N.11, p.861-875, November 1993.