An example of adding spatial impression to recorded music: signal convolution with binaural impulse responses

Angelo Farina
Dipartimento Ingegneria Industriale
Università di Parma - Via delle Scienze - PARMA

Abstract

The following paper presents the results obtained in a recently growing-up Digital Signal Processing branch: the reconstruction of temporal and spatial characteristics of the sound field in a musical space, starting from monophonic anechoic (dry) digital music recordings, and applying FIR filtering to obtain binaural (stereo) tracks (this process is briefly called *auralisation)*.

The FIR filters can be experimentally derived binaural impulse responses, obtained from a correlation process between the signal emitted through a loudspeaker on the stage of the theatre and the signals received through two binaural microphones, placed at the ear channel entrance of a dummy head. Alternatively, the impulse responses can be generated by a computer simulation of the sound field inside a room, through a wide family of Ray Tracing or Beam Tracing models.

The convolution process has been implemented by two different algorithm: time domain *true convolution* and frequency domain *select-save*; both runs on the Parma's CM-2 parallel supercomputer, and the second also on any Unix or MS-DOS machine.

The convoluted signals can be listened by earphones, and in this case no further processing is required, except an optional equalization. In the case of loudspeaker reproduction, however, it is necessary to process the signals in such a way to eliminate the crosstalk between opposite channels. This processing is easy if the loudspeakers are placed in a small, semi-anechoic chamber.

The results of the convolution process have been compared by direct listening with binaural recordings of the same music pieces, emitted through a loudspeaker in the same theatres, and recorded through the dummy head connected to a DAT digital recorder. Such a comparison shows that the experimental impulse responses perfectly reconstruct the spatial and temporal effects of the room, while the computer simulated ones are very effective in the simulation of reverberation and echoes, but are at some extent innatural in the spatial impression and don't permit a correct localization of the sound source.

The comparison between earphones and loudspeakers points out the superiority of the latters, because they deliver a better naturality, particularly for spatial impression.

One of the most interesting applications of this auralisation system is the capability to point out by direct listening the effects of proposed acoustic corrections: this is particularly useful in the restoration of ancient theatres and churches, as in this case it is necessary to convince civil authorities in charge of the protection of monumental buildings of the benefits obtainable with the insertion of modern materials.

1. Brief history of Auralisation

Auralisation is the process of filtering a monophonic anechoic signal in such a way to reproduce at the ears of the listener the psychoacoustic feeling of an acoustic space, including reverberation, single echoes, frequency colouring, and spatial impression.

This technique was pioneered by the Gottingen Group (Schroeder, Gottlob and Ando) [1,2] in the 70's, using however strong simplifications of the structure of the sound field, because the digital filtering process, at that time, was a very slow task.

More recently, Lehnert and Blauert [3] developed a dummy head recording technique that incorporates extensive DSP processing of the signal prior of the headphone reproduction, including limited capacity of increasing reverberation and adding single echoes, plus frequency domain parametric equalization. Vian and Martin [4] used impulse responses obtained by a computerized beam-tracing model of the room as FIR filters, performing convolution not in real time on a mainframe computer, with subsequent headphone reproduction.

Eventually, a dedicated DSP was developped (Lake DSP) [5], capable of real time auralisation through extensive frequency domain processing. Once programmed with two FIRs, this unit can be operated also detached from any computer.

Recently, Kuttruf [6] has studied the comparative behaviour of Ray Tracing and Image Source computer models, for the production of impulse responses suitable for auralisation. Dalenback [7] et al. employed the Lake DSP to obtain real-time auralisation of anechoic music with impulse responses obtained from a Ray Tracing model, processed to become "binaural" with experimental angle-dependent data obtained by the Lehnert & Blauert dummy head.

2. Statement of the work

Instead of using DSPs to perform convolution, in this work general purpose computers are used, both to prepare the impulse responses, to convolve them with the music, and to manage Input/Output of the audio signals. The computers range from low-cost MS-DOS PCs to a massively parallel supercomputer, passing through Unix workstations. Unexpectedly, the performances of these systems are quite similar, and the fastest processing has been achieved on the PC!

On the other hand, the massively parallel computer (CM-2) is capable to run in reasonable times both direct time-domain *true convolution* and indirect frequency domain *select-save* algorithms.

The final judgement of any auralisation system can only be given *by ears*, actually listening at the signals produced. Furthermore, the *naturality* of the sound can be assessed only by comparison with the actual sound field in a concert hall.

For these reasons, in the first phase it was preferred to use **experimental impulse responses** as FIR filters, although a new computer program is already available (based on a recently developed *pyramid tracing* algorithm, decribed in section 6), capable of predicting impulse responses for arbitrary shaped acoustic spaces, with very short computation times.

The experimental impulse responses also contain the response of loudspeaker and microphones: these could in principle be removed by a deconvolution process, but it was preferred to leave them, both in the impulse responses and in the "live" music recordings. In such a way, the result of the convolution process should be directly comparable with the digital recordings of the same music piece, emitted through the same loudspeaker, placed in the concert hall, and sampled through the binaural microphones of the dummy head. Both of these recordings (digital convolution and direct recordings) should in principle be capable to evocate in the listener the same psychoacoustic effects as if he was in the concert hall.

3. Hardware

Three hardware systems have been used in this work:

1) Impulse response measurement system (fig. 1);
2) Music reproduction & recording system (fig. 2);
3) Digital filtering and convolution system (fig. 3).

System 1) and 2) share the same sound source (a dodechaedron omnidirectional loudspeaker) and receivers (dummy head with binaural microphones): the first gives as result two 64k points impulse responses (Left amd Right), saved on a PC Hard Disk in the MLSSA .TIM format (IEEE/Microsoft float). The second gives a DAT tape, carrying the recorded stereo tracks.

System 3) can be implemented on various platform: they can be quite complicated, because the digital anechoic music and the convolved results need to be transferred back and forth between computers and a DAT recorder. Digital interfaces of audio equipment (CDs and DATs) conforms to well known standards (AES-EBU, SPDIF), but these interfaces are actually not implemented on Unix or MS-Dos computers, with very few exceptions. For this reason, the digital I/O was performed with a Silicon Graphics Indigo workstation, available at the University of Bologna, that is equipped with SPDIF input and output.

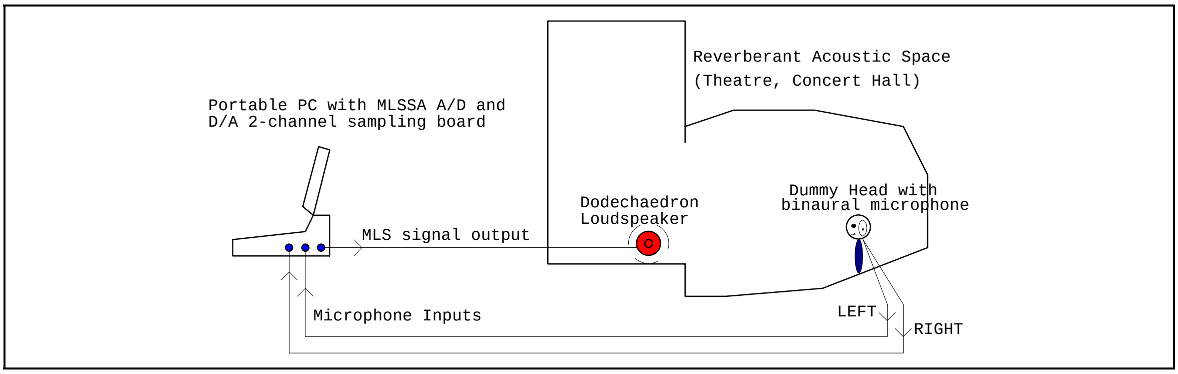

Fig. 1 - Impulse Response Measurement System by MLS cross-correlation.

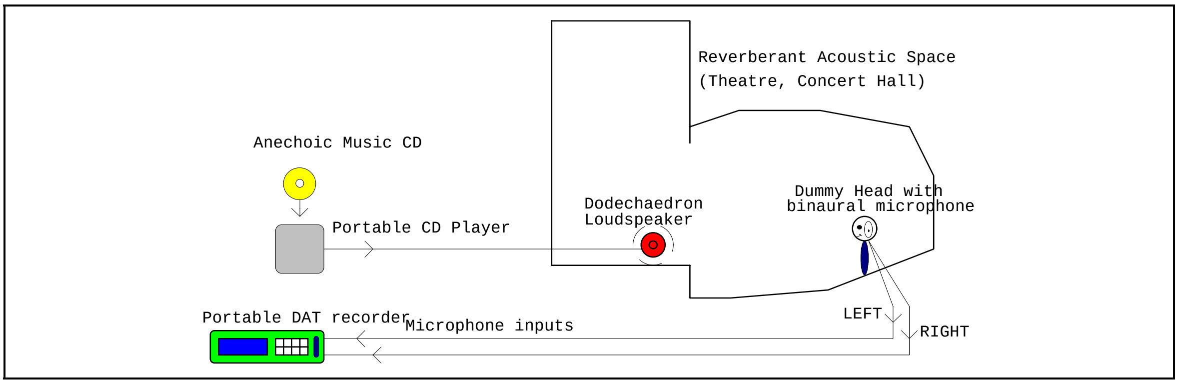

Fig. 2 - Music reproduction and recording sistem.

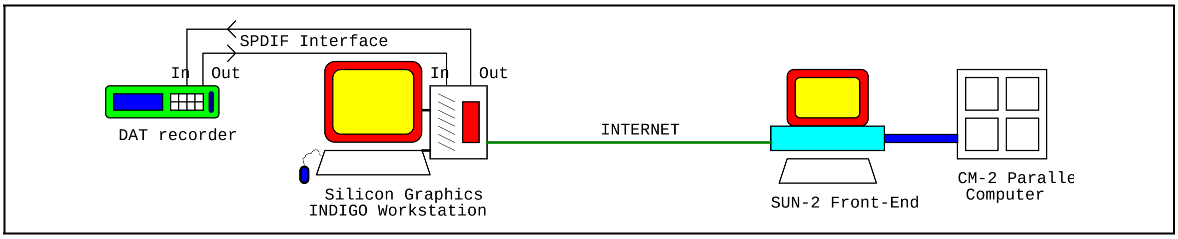

Fig. 3 - Digital filtering and convolution system for the CM-2.

Obviously, the PC version of the system is simpler: a 66 MHz 486 DX-2 PC has been equipped with a low cost 16 bit A/D and D/A audio board (Microsoft Sound System). The I/O can be made with analog signals, with a slight performance degradation, that however is hardly heard after the processing.

4. Software for Convolution

The convolution of a continous input signal x(t) with a linear filter characterized by an impulse response h(t) yields an output signals y(t) by the well-known convolution integral:

$$y(\tau) = x(\tau) \otimes h(\tau) = \int_0^\infty x(\tau - t) \cdot h(t) \cdot dt \qquad (1)$$

When the input signal and the impulse response are digitally sampled (t = i·Dt) and the impulse response has finite lenght N, such an integral reduces to a summation:

$$y(i) = \sum_{j=0}^{N-1} x(i-j) \cdot h(j) \qquad (2)$$

The sum of N products must be carried out for each sampled datum, resulting into an enormous number of multiplications and sums! These computations need to made with float arithmetic, to avoid overflow and excessive numerical noise. For these reasons, the real time *direct true convolution* is actually restricted to impulse response lenghts of a few hundreths points, while a satisfactory description of a typical concert hall transfer function requires at least N=65536 points (at 44.1 kHz sampling rate).

However, the convolution task can be significantly simplified performing FFTs and IFFTs, because the time-domain convolution reduces to simple multiplication, in the frequency domain, between the complex Fourier spectra of the input signal and of the impulse response. As the FFT algorithm inherently suppose the analyzed segment of signal to be periodic, a straightforward implementation of the Frequency Domain processing produces unsatisfactory results: the periodicity caused by FFTs must be removed from the output sequence.

This can be done with two algorithms, called *overlap-add* and *select-save* [8]. In this work the second one has been implemented. The following flow chart explain the process:

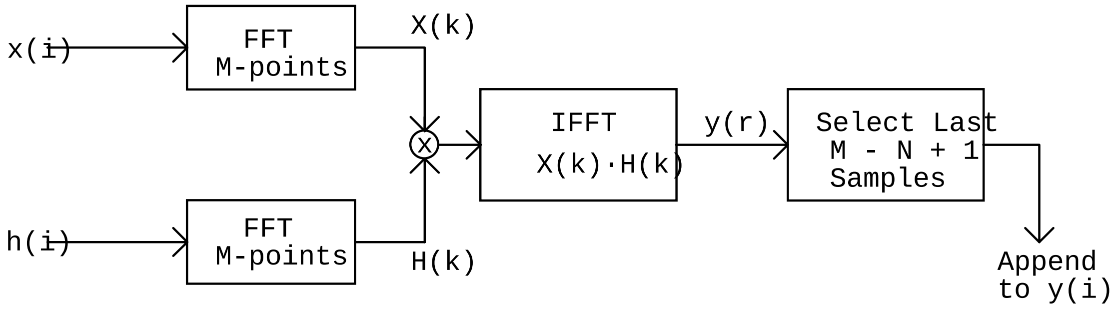

Fig. 4 - SELECT-SAVE Flow Chart

As the process outputs only M-N+1 convolved data, the input window of M points must be shifted to right over the input sequence of exactly M+N-1 points, before performing the convolution of the subsequent segment.

The tradeoff is that FFTs of lenght M>N are required. Typically, a factor of 4 (M=4·N) gives the better efficiency to the select-save algorithm: if N is 65536 (2^{16}), one need to perform FFTs over data segments of lenght M=65536·4=262144 points! This require a very large memory allocation, typically 1 Mbyte, for storing the input sequence or the output signal. The overall memory requirement for the whole select-save algorithm is thus several Mbytes! This is the reason for which this algorithm cannot be easily implemented on widely diffused low cost DSP boards, that are not equipped with such a storage capability.

In principle, the select-save algorithm largely reduces the number of float multiplications required to perform the convolution. Each FFT or IFFT requires $M \cdot \log_2(M)$ multiplications: a couple of FFT and IFFT produces however 3/4·M new output data, and so the number of multiplications for each output datum is about 50, instead of 65536.

On the other hand, FFTs are very dangerous operations: the data flow is segmented, and each segment is processed separately. Numerical oddities can affect in different manner separate segments, and the computational "noise" (that is audible when the signal amplitude is low) changes from segment to segment. For these reasons, the sound quality of signals filtered with true convolution should be better than with the select-save, provided that the larger number of multiplations don't introduce additional numerical noise.

Furthermore, the true convolution is a process that is well suited for parallel coding and processing: allocating a linear shape of N processors, all the N multiplications can be carried out simultaneously; then the sum of the N results can be achieved very efficiently through the C* "+=" assignement of the parallel data to a scalar variable.

Four different convolution codes have been developed:

- CM-CONV performs true convolution in time domain on the CM-2 computer (in C*);
- CM-SEL performs select-save convolution on the CM-2, using parallel FFT (in C paris);
- SUN-SEL performs select-save convolution on the SUN frontend (in C ansi).
- PC-SEL performs select-save convolution on a 486 PC (in MS Fortran Powerstation).

5. Processing to avoid crosstalk in loudspeaker reproduction.

Fig. 5 shows what happens when a couple of loudspeakers is placed in front of the listener's head: the signal coming from each loudspeaker reaches both the ears, and so at the ear channel entrance both Left and Right channels arrive mixed [9].

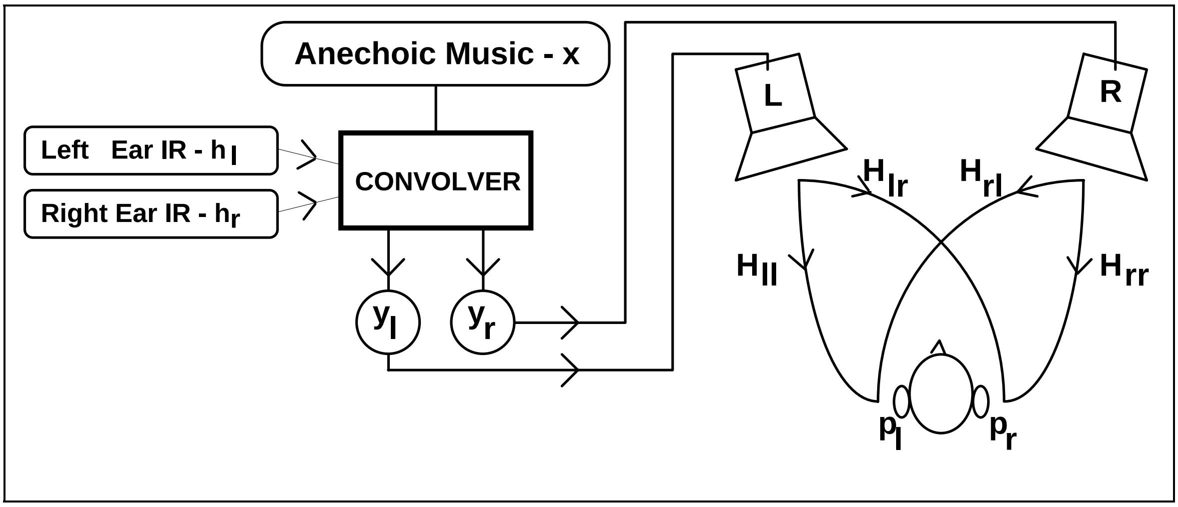

Fig. 5 - Flow Diagram of signals arriving to ears through loudspeakers.

Furthermore, the Head Related Transfer Function is already inserted in the IRs used for convolution, but the signal coming from loudspeaker newly interferes with the listener's head, so the head filtering is applied twice!

For these reasons, a more involved processing is needed. In mathematical notation, the signals arriving at the Left and Right ear channel entrance of the listener's head can be described as:

$$\begin{aligned} p_l &= y_l \otimes h_{ll} + y_r \otimes h_{rl} \\ p_r &= y_r \otimes h_{rr} + y_l \otimes h_{lr} \end{aligned} \qquad (3)$$

Noting that the stereo signal y_l and y_r were derived from a single mono input x, through convolution with the two binaural IRs h_l and h_r, and passing to the frequency domain by FFT, the previous relations can be rewritten as:

$$\begin{aligned} P_L &= X \cdot \left(H_L \cdot H_{LL} + H_R \cdot H_{RL}\right) \\ P_R &= X \cdot \left(H_R \cdot H_{RR} + H_L \cdot H_{LR}\right) \end{aligned} \qquad (4)$$

Now, let we substitute two adequate IRs (h_l' and h_r') instead of the original ones: they must be capable to make the terms in parenthesis equal to the (wanted) IRs H_L and H_R. After a few passages, we get:

$$\begin{aligned} H_L' &= \frac{H_{RR} \cdot H_L - H_{RL} \cdot H_R}{H_{LL} \cdot H_{RR} - H_{LR} \cdot H_{RL}} \\ H_R' &= \frac{H_{LL} \cdot H_R - H_{LR} \cdot H_L}{H_{LL} \cdot H_{RR} - H_{LR} \cdot H_{RL}} \end{aligned} \qquad (5)$$

These simple relations permit to evaluate the wanted IRs, performing the calculations in frequency domain and returning to time domain (if required) by FFT.

One need to measure the four IRs related to loudspeakers and receiving head, whose effect is so completely removed from the audio playback. This means that not only the crosstalk is eliminated, but also unwanted frequency filtering due to loudspeaker response. Furthermore, the head-related transfer function is eliminated too, avoiding the double-filtering stated above.

In theory this processing could also eliminate reflections on the room's walls, and this way a special listening environment should not be necessary. However, the experience made shows that is unlikely that this happen: the presence of room reflections cause comb filtering in the frequency domain, and this can produce numerical difficulties in the processing, giving modified IRs with strong peaks that destroy the spectral balance.

Actually two solutions have been found to this limitation: if the listening must be made in a not-absorbing room (i.e. for public demonstrations), the above computations are made extracting only the anechoic portion of the 4 experimental IRs. This way the digital crosstalk compensation does not take into account the presence of room reflections, but they are still present, and cause a certain degree of crosstalk and spectral alterations.

Alternatively, a small anechoic chamber was employed: also if it is not perfectly anechoic (particularly in the low frequency range), it is death enough to avoid numerical oddities in the filtering process. In this case the 4 experimental IRs are not truncated, but their lenght is reduced applying a semi-gaussian window, to trim out the noise following the sound decay.

All the IRs required for crosstalk compensation can be easily measured through the same system already introduced for measurement in musical spaces (see fig. 1)

6. Derivation of Impulse Responses from a Computer Model

Various techniques are available to obtain the impulse responses of a space for music: the experimental ones can be obtained with:

- Gun shots recorded on a DAT tape;
- Real-Time deconvolution of an MLS signal, as shown in sect. 3;
- Off-Time deconvolution of an MLS signal, initially recorded on a DAT tape.

The last two method produces almost exactly the same results, while the first one suffers of poor noise rejection and not uniform spectral and spatial emission of the gun impulse, but is inexpensive and very fast [10].

On the other hand, the impulse responses can be derived by a computer model. After various experiences with Finite Elements and Ray Tracing codes [11,12], the author developed a new algorithm, conceptually similar to the recent *beam tracing* pioneered by Martin [13] and Naylor [14], but with further refinements due to the elimination of any pseudo-random approximation.

Many of the recent computer models are conceptually derived from traditional Ray Tracing, whose main problem was to require a large number of rays to adequately model highly reverberant spaces, with reasonably little receivers. The first improvement was the Cone Tracing, in which the energy of each ray is spread in cones: but this lead to another problem, because the (spherical) sound source cannot be perfectly subdivided in adiacent cones, as they don't cover all its surface.

The Beam Tracing overcomes to this problem, overlapping adiacent cones, and weighting the sound energy with a Gaussian distribution, in such a way that the average sound energy of the overlapped cones is constant on the spherical source's surface. But in this way, the model is still a semi-stochastic one, and the sound field is forced to become "Sabinian" in the late part of the impulse response, independently of the room's shape.

The Pyramid Tracing, subdividing the spherical source in adiacent triangular pyramids, does not have any overlapping problem, and the detection of each sound path can be performed deterministically; the subdivision of the sphere in any number of triangles is achieved with the algorithm of Tenenbaum et al. [15]. Furthermore, no Sabinian queue is attached to the impulse response, that this way can take into account, for example, the effect of coupled spaces with different average absorption, yielding to double-slope queues.

The progressive loss of late reflections, common to all the diverging tracers, is corrected in this case by a multiplication of the received energy for the ratio between the expected temporal density of reflections and the (nearly constant) density obtained from the pyramid tracing, instead of the method suggested by Lewers [16], based on diffuse-field assumptions.

The Pyramid Tracing algorithm is the core of a new general purpose acoustic software, named RAMSETE, that includes:

- a dedicated Windows-based 3D surface CAD;
- an Impulse Response post processor;
- a graphic post-processor capable of contour and colour mapping of the acoustic parameters and visual rendering of the room geometry;
- a sound source manager capable of direct interfacing with data files produced by real time analyzers, and Sound Power calculations according to ISO 3744;
- a material manager and data base, including absorption and insulation data on hundreths of materials.

The impulse responses produced by the Ramsete software are inherently monophonic, also if the information of the provenience of each ray is recorded. The process of converting such data in two binaural responses is actually the only part of the systems that don't performs satisfactorly, because it weights the energy arriving to both ears with a very simple model, based on the cosine of the angle between the ray and the ear-to-ear line. A more sophisticated approach, as that described in [7], shall been implemented in future. Alternatively, it is possible to process an already-stereo anechoic track with a single mono IR, obtaining however a very poor spatial impression.

7. Experiments

7.1 Performance of the convolution software.

Two anechoic music samples (688.107 points long) were chosen for these experiments:

- MOZART: Overture "Le Nozze di Figaro", bars 1-18, duration = 16"
- BRAHMS: 1st mov. Symphony No. 4 in e minor, Op.98, bars 354-362, duration = 17"

These samples come from a Denon CD titled "Anechoic Orchestral Music Recording" (PG-6006): all the material herein recorded is taken from PCM 24 bits digital masters, sampled at the Minoo Civic Hall in Osaka (Japan), where an anechoic chamber was builded on the stage.

The two samples were first transferred digitally, through optic fiber SPDIF interface, on a DAT tape. Then the signals were inputted to the Indigo's SPDIF coax interface, saving them both in .RAW format (2's complement 16 bit integers) and MS-Windows .WAV format. From there the data files were transferred by Internet to the Sun front-end of the CM-2 computer and to the 486 PC. Then the format of .RAW data files was changed from integer to float, through a format conversion utility made for the circumstance.

The first sample ("FIGARO") was convolved with a pair of experimental impulse responses measured at the Teatro Comunale di Ferrara, a famous opera house that has been recently adapted for simphonic music performances, by building a special wooden stage enclosure.

The second sample ("BRAHMS") was convolved with a pair of experimental impulse response coming from an auditoriom room of the Engineering Department of Parma.

In parallel to the experimental determination of impulse responses, the same two anechoic samples were re-recorded on a DAT through the dummy head's microphones, while they were emitted through the omnidirectional loudspeaker driven directly from the CD player, using the system described in sect. 3. Also this material was transferred in computer audio files, through the same (quite complex) chain.

The anechoic samples were convolved employing the four convolution codes.

The following table shows the Figaro's computation times (in s) for the 4 codes, for different lenghts of the convolved impulse response:

IR lenght	CM-CONV		CM-SEL		SUN-SEL		PC-SEL	
N	Total time	CPU Time	Total Time	CPU Time	Total Time	CPU Time	Total Time	CPU Time
256	696	680	--------	--------	881	854	62.8	34.6
512	698	684	--------	--------	950	923	59.4	33.4
1024	693	678	--------	--------	1027	999	58.9	33.2
2048	689	674	1126	4.99	1100	1070	63.0	35.9
4096	694	678	1157	4.89	1188	1153	63.8	37.5
8192	705	684	1105	6.16	1319	1275	67.5	41.4
16384	749	718	1155	7.54	1480	1383	72.9	45.6
32768	851	799	1111	7.90	1667	1519	102.9	72.4
65536	1090	965	1370	9.88	1892	1676	113.6	77.2

The same data are also presented in graphical form:

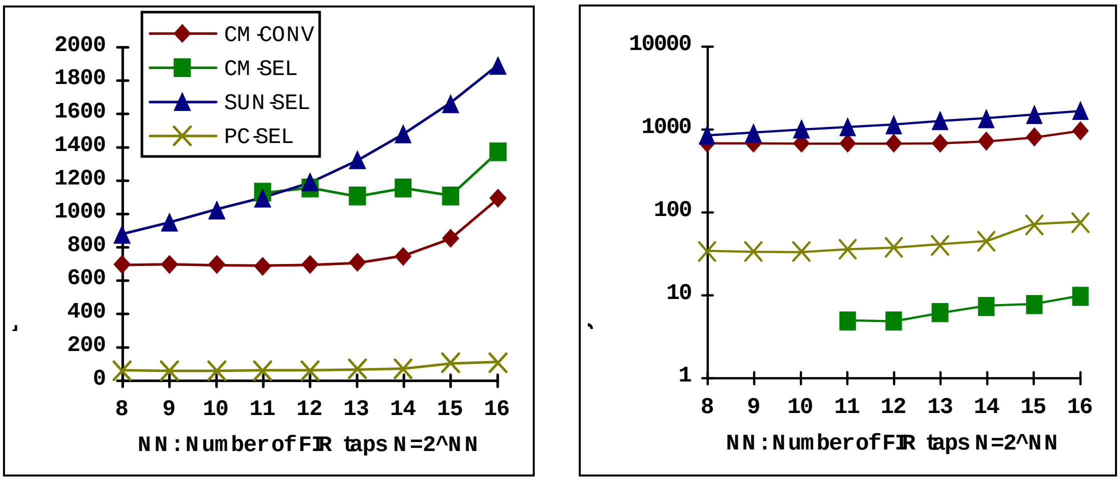

Fig. 6 - Graphs of computation times

7.2. Discussion of the performance data

It can be seen that the CPU times for the CM-SEL code are effectively very low, in good accordance with the previded ones: on the CM-2 also with N=65536 the computation time is less than the actual sample duration (17 s), making it possible, in principle, real time processing. However, the total run times remain larger than the CM-CONV code, and unexpectedtly also than the PC-SEL: this is due to I/O limitations of the front-end, that is very slow managing Mbytes of data back and forth to the CM-2; the parallel machine is very fast computing FFTs, but the larger data flow causes an overall reduction in performance against the direct time-domain convolution. In CM-CONV, infact, the parallel engine works almost all the time, crunching Gflops, while the I/O between the front-end and the CM-2 remains to a minimum.

The SUN-SEL code is always the slower one: this is due to the bad math performance of the SUN's CPU while performing FFTs (in fact the CPU time is very near the total time).

The PC-SEL has been an enjoyable surprise: actually with a low cost PC the total computation time is about 1/10 of the better performance obtained on the CM-2! Furthermore, this code directly reads and write a .WAV file, that can be immediately played for listening, without tedious data transfer on varios digital links (Internet and AES-EBU). As a further advantage, PC-SEL is a stereo convolver, producing a mono-to-stereo full binauralisation with a single processing, and with an increase in computation time of less than 15%. It must be noted that the results presented above are all for a mono convolution, and that the codes for CM-2 and SUN requires a separate run for the second ear, with a doubling of computation times.

The results show that real-time convolution is actually beyond the capacity of the system: regarding the CM-2 this is due mainly to bottle-necks in the disk access and in I/O limitations of the front-end, because the 8192 CPUs have power enough to obtain real time processing in the frequency domain. With the PC a speed increase of a factor 10 should be required to obtain real-time processing. Something can be done refining the algorithm, also because the new 32-bit MS Fortran compiler has been employed only for three months, and many optimization techniques are probably to be discovered.

However, waiting for the next generation of general-purpose CPUs, actually the system gives reasonably good performances for off-line processing, and can be used to make subjective tests on acoustic quality in concert halls.

7.3 Headphone listening

Eventually, the convolved stereo data files were retransformed in integer format, placed again on the Indigo computer, and transferred through the SPDIF digital output interface to the DAT recorder, storing in sequence on the tape the original anechoic signal, the CM-2 convolved one (true time domain convolution), the PC convolved (select-save, outputted to the DAT through the D/A PC board), and the music signal recorded "live" in the rooms.

Listening by headphones at such a DAT tape gives a direct feeling of how good the digital processing can be. The convolved signal is almost identical to the "live" recording, except for the background noise, that affects only the latter. The segmentation effects of the select-save algorithm can be heard only in the silence following the music, and only from sharp-eared musicophils.

No audible difference in the background noise can be noted between the CM-2 true convolution and the PC select-save, because the first introduce sligthly more digital noise, due to the larger number of math manipulations on each datum, but the output from the PC is performed through a D/A + A/D link, that introduce requantization noise.

If a PC board carrying SPDIF input and output would be available at low cost, this system could be well suited for professional applications.

Actually the test tapes have been submitted, for judgement, to music critics and famous conductors, along with a questionnaire developed for room acoustic comparisons [17].

By the same time, a subjective evaluation test is being conducted, using as test people students of the Engineering Department of Parma. This test is preliminary to what explained in the next section, about loudspeaker reproduction in a miniaturized semi-anechoic chamber: this is because the first responses, obtained from the musicians, tell about "unnaturality" of the headphone listening, and about difficulties in the sound source localization (the well known front-back confusion).

7.4 Loudspeaker reproduction

A special miniaturized anechoic chamber has been used for loudspeaker reproduction, because the digital signal processing can (easily) remove the loudspeaker crosstalk, as explained in section 5, but it is very difficult to remove the reverberation!

A standard true anechoic room is very expensive, and actually at the University of Parma this facility is not available. On the other hand, one of the goals of the whole Ramsete project is to show that enormous money investments are not necessary to obtain high quality musical sound prevision and processing.

Fig. 7 shows the miniaturized anechoic chamber: it has been derived from a commercially avaliable, low cost audiometry seat. Inside the glass fiber moulded shell a 4 way stereo loudspeaker system has been installed, with electronic cross-overs and time and frequency equalization obtained with a Sigtec digital processing unit. This equalization is not absolutely necessary, because, as stated in sect. 5, the digital processing performed on the impulse responses prior the convolution has the capability to equalize out any (linear) frequency distortion or time misalignment of the loudspeakers. However, it resulted that the presence of this separate equalizer gives better performances, mainly due to the reduction of interaction between the cross-talk cancellation and frequency correction.

The convolution process need to be performed separately for loudspeaker reproduction, using the modified impulse responses instead of the original ones. Furthermore, also the music directly recorded in the acoustic spaces needs processing to avoid crosstalk, but in this case, being the signal already stereo, the computing time is about twice that needed for the anechoic music auralisation.

The impulse responses H_{LL}, H_{LR}, H_{RL} and H_{RR}, required for the processing, vary with the position and size of the listeners inside the cab. For this reason, before any listening test, a MLSSA set of 4 measurements is conducted, while the subject is seating, wearing a binaural microphone headset (Sennheiser MKE 2002). These responses are used to produce the two IRs used for the convolution, and the music pieces need to be convolved explicitely for each test subject, also if the pieces prepared for a subject are roughly acceptable for first-order evaluations on other subjects. For this reason, actually no DAT tape has been produced for the loudspeaker reproduction.

The author's personal experience of listening with such a system is enthusiastic, because the front-back confusion is avoided, and the fidelity is as good that one cannot say, for example, if a voice is coming out the loudspeakers or from a person speaking outside the cab, mantaining however a more-or-less accurate source localization.

In the next months this system shall be used for a listening test on a medium number of University students, concerning the capability of distinguish between digitally convoluted music pieces and the same recorded "live", and between different rooms.

Once the digital processing system shall be validated, this facility should be used for the study of the correlation of objective acoustic parameters with subjective preferences, obtained by questionnaires, in parallel with direct subjective interviews of people listening at concerts in famous theaters.

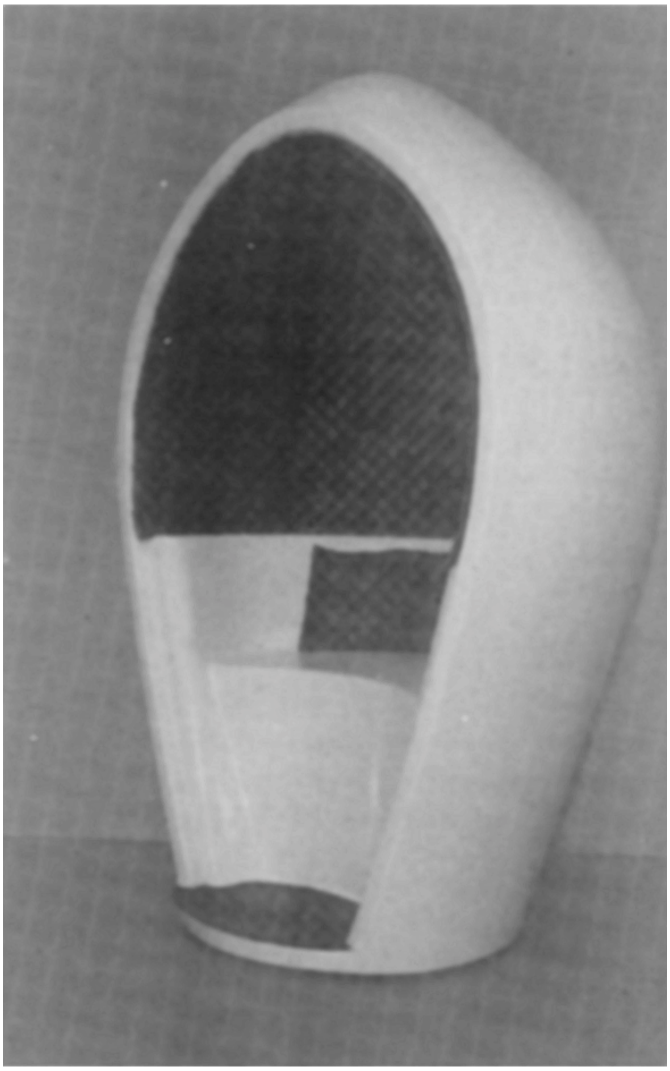

Fig. 7 - Single-Seat Anechoic Chamber (Opened).

8. Application to the restoration of old churches

The first practical application of the Ramsete software, connected to the auralisation system discussed in this paper, was the preparation of comparative music samples to evaluate the effects of different proposed stagings and acoustic corrections in a recently restored monumental church, namely the Basilica of S.Domenico in Foligno, as reported in a separate paper (Cocchi and Farina [18]).

First a simulation of the church in the actual state was conducted; this way it was verified that the numerical simulation of the sound propagation obtained by the numerical Pyramid Tracing model is satisfactory, compared to experimental measurements of Impulse Responses obtained in that room with gun shots and binaural DAT recordings.

Then 5 different stagings and acoustic panel setups have been tested by numerical simulations.

Along with traditional evaluations, based on objective acoustic parameters, the auralisation system has been employed to subjectively compare the acoustic performance of the 6 proposed configurations. In this case, needing only a comparative judgement, the lack of naturality caused by headphone listening of sinthetic impulse response was not a great problem.

On the DAT tape the musical piece convolved with experimental IRs (gun shots) is first recorded. Then the piece convolved with sinthetic IRs follow, relative to the actual status of the church. The comparison of these two pieces gives an idea of the similarity of the sinthetic IRs with the experimental ones: actually it can be stated that they are only similar, but audible differences still persist.

Eventually some samples follow for comparison, obtained by sinthetic IRs relative to different proposed configurations. The differences between these samples are easily perceived, and a judgement can be made about the better one.

References

[1] Schroeder M.R., Gottlob D. Siebrasse K.F. - "Comparative study of european concert halls: correlation of subjective preference with geometric and acoustic parameters" - J.A.S.A., vol.56 p.1195 (1974).

[2] Ando Y. - "Concert Hall Acoustics" - Springer-Verlag, Berlin Heidelberg 1985.

[3] Lehnert H., Blauert J. - "Principles of Binaural Room Simulation" - Applied Acoustics, vol. 35 Nos 3&4 (1992).

[4] Vian J.P., Martin J. - "Binaural Room Acoustics Simulation: Practical Uses and Applications" - Applied Acoustics, vol. 35 Nos 3&4 (1992).

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

[6] Kuttruff K.H. - "Simulation Models for Auralization in Room Acoustics" - Proc. of 14th. ICA, F2-2, Bejing 3-10 september 1992.

[7] Dalenback B.I., Svensson P., Kleiner M. - "Prediction and Auralization based on a combined Image Source/Ray Tracing Model" - Proc. of 14th. ICA, F2-7, Bejing 3-10 september 1992.

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

[9] Moller H. - "Fundamentals of Binaural Technology" - Applied Acoustics vol. 36 no.s 3 & 4, pag. 171 (1992).

[10] Cocchi A., Farina A., Fausti P., Garai M., Semprini G. - "New possibilities in room acoustics measurements: real-time analyzer, DAT, computer: a comparative approach" Proc. of 6th FASE Congress - Zurich 29-31 july 1992.

[11] Farina A., Pompoli R. - "Noise reduction in a Tractor Cab using resonators and active loudspeakers: FEM prediction and experimental verification" - Proc. of 2nd ATA International Conference on Vehicle Comfort - Bologna, 14-16 october 1992.

[12] Farina A., Cocchi A., Garai M., Semprini G. - "Old Churches as concert halls: a non-sabinian approach to optimum design of acoustic correction" - Proc. of 14th. ICA, F5-7, Bejing 3-10 september 1992.

[13] Maercke D., Martin J. - "The prediction of Echograms and Impulse Responses within the Epidaure software" - Applied Acoustics Vol. 38 no.s 2-4, pag. 93 (1993).

[14] Naylor G.M. - "ODEON - Another Hybrid Room Acoustical Model" - Applied Acoustics Vol. 38 no.s 2-4, pag. 131 (1993).

[15] Tenenbaum R., Slama J., Ballesteros M. - "Numerical simulation of room acoustics: a new approach for source modelling" - Proc. of 14th. ICA, F6-11, Bejing 3-10 september 1992.

[16] Lewers T. - "A combined Beam Tracing and Radiant Exchange Computer Model of Room Acoustics" - Applied Acoustics Vol. 38 no.s 2-4, pag. 161 (1993).

[17] Tronchin L. - "Qualità acustica nei teatri: analisi sperimentale e confronto con valutazioni soggettive in alcuni esempi italiani" - Tesi di Laurea - Univ. di Bologna, A.A. 91/92 - Relatore A. Cocchi, Correlatori L. Rocco e A. Farina.

[18] A. Farina, A. Cocchi - "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.