Convolution of anechoic music with binaural impulse responses

Angelo Farina
Dipartimento Ingegneria Industriale
Università di Parma

Abstract

The following paper presents the first results obtained in a recently growing-up Digital Signal Processing branch: the reconstruction of temporal and spatial characteristics of the sound field in a concert hall, starting from monophonic anechoic (dry) digital music recordings, and applying FIR filtering to obtain binaural (stereo) tracks.

The FIR filters employed in this work are 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.

The convolution process has been implemented by two different algorithm: time domain *true convolution* and frequency domain *select-save*; both runs on the CM-2 computer, and the second also on any Unix machine.

The results of the convolution process have been compared by direct headphone 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.

1. Brief history of Auralization

Auralization 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 [5], capable of real time auralization through extensive frequency domain processing. Once programmed with the two FIRs, this unit can be operated also detached from any computer. In the next months, it is advisable that many works shall be conducted using such an hardware system.

2. Statement of the work

In this work the performances of a large, massively parallel computer (CM-2) were tested: it can challenge DSPs in performing convolution tasks, running both direct time-domain *true convolution* and indirect frequency domain *select-save* algorithms.

The final judgement of any auralization 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 this 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), capable of predicting impulse responses for arbitrary shaped acoustic spaces, with very short computation times.

Such 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 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 .TIM format (IEEE/Microsoft float). The second gives a DAT tape, carrying the recorded stereo tracks.

System 3) is quite complicated, because the digital anechoic music and the convolved results need to be transferred back and forth between computers and a DAT recorder. The digital interfaces of the 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 I/O was performed with a Silicon Graphics Indigo workstation, available at the University of Bologna, that is equipped with SPDIF input and output.

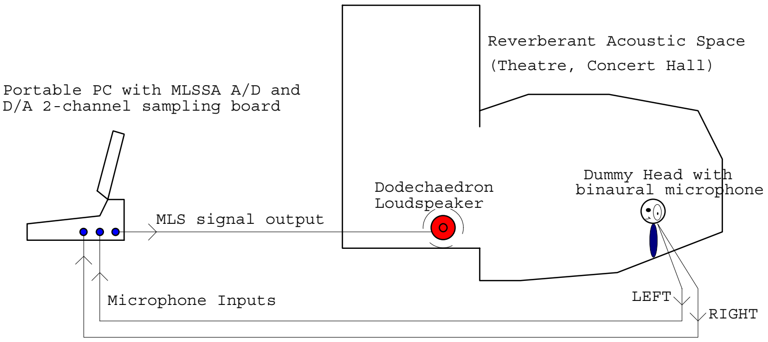

Fig. 1 - Impulse Response Measurement System by MLS cross-correlation.

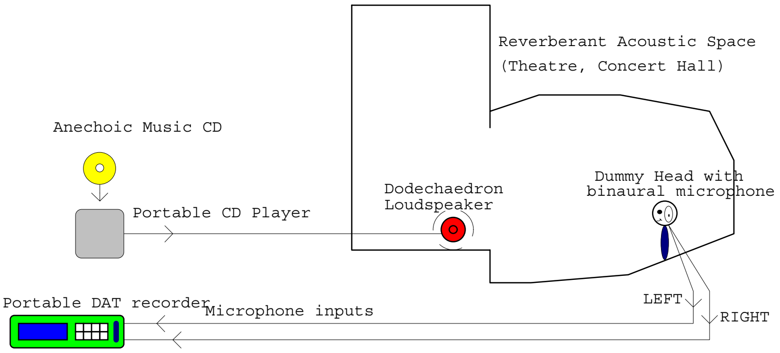

Fig. 2 - Music reproduction and recording sistem.

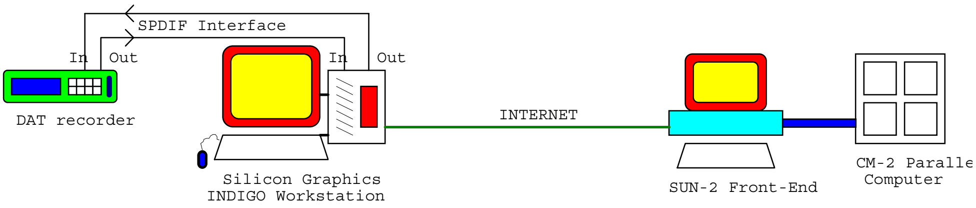

Fig. 3 - Digital filtering and convolution system.

4. Software

The convolution of a continous input signal x(τ) with a linear filter characterized by an impulse response h(τ) yields an output signals y(τ) by the well-known convolution integral; when the input signal and the impulse response are digitally sampled ($\tau = i \cdot \Delta\tau$) and the impulse response has finite lenght N, one obtains:

$$y(\tau) = x(\tau) \otimes h(\tau) = \int_{0}^{\infty} x(\tau - t) \cdot h(t) \cdot dt \quad ; \quad y(i) = \sum_{j=0}^{N-1} x(i-j) \cdot h(j) \qquad (1)$$

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 descriptions 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* [6]. In this work the second one has been implemented. The following flow chart explain the process:

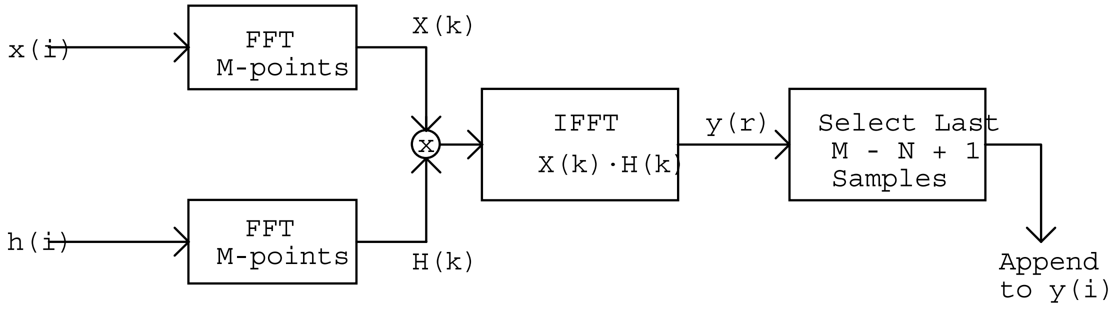

Fig. 4 - SELECT-SAVE Flow Chart

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 spectrum. The overall memory requirement for the whole select-save algorithm is thus several Mbytes!

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.

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.

Three 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 subs (in C paris);
- SUN-SEL performs select-save convolution on the SUN frontend (in C ansi).

5. Experiments

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 SPID coax interface, saving them in .RAW format (2's complement 16 bit integers). From there the data files were transferred by Internet to the Sun front-end of the CM-2 computer. Then the format of 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 Faculty 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.

The anechoic samples were convolved employing the three convolution programs.

The following table shows the computation times (in s) for the three programs, for different lenghts of the convolved impulse response:

IR lenght	CM-CONV		CM-SEL		SUN-SEL	
N	Tot. time	CPU Time	Tot. Time	CPU Time	Tot. Time	CPU Time
256	696.07	680.20	--------	--------	881.54	854.64
512	698.99	684.46	--------	--------	950.38	923.68
1024	693.37	678.93	--------	--------	1027.87	999.83
2048	689.88	674.85	1126.92	4.992	1100.62	1070.43
4096	694.50	678.02	1157.70	4.890	1188.71	1153.59
8192	705.20	684.90	1105.17	6.165	1319.81	1275.25
16384	749.30	718.48	1155.53	7.548	1480.94	1383.12
32768	851.52	799.30	1111.67	7.900	1667.71	1519.19
65536	1090.80	965.81	1370.67	9.886	1892.58	1676.08

The same data are also presented in graphical form:

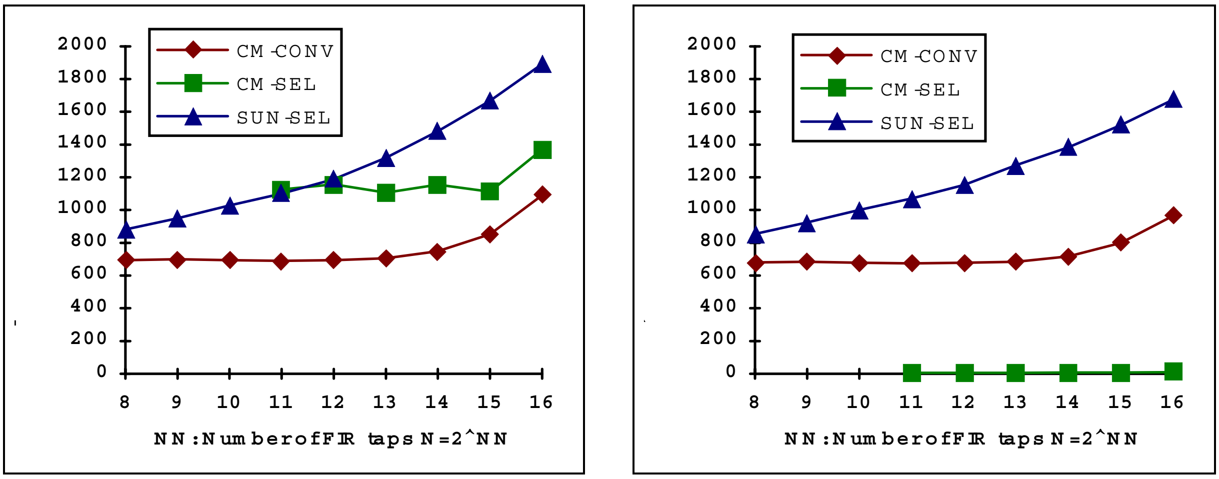

Fig. 5 - Graphs of computation times

6. Discussion of results and conclusions

It can be seen that the CPU times for the CM-SEL code are effectively very low, in good accordance with the previded ones: 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: 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 results show that real-time convolution is actually beyond the capacity of the system, but this is due mainly to bottle-necks in the disk access and in I/O limitations of the front-end. A speed increase of a factor 100 should be required to obtain real-time processing. However, the system gives reasonably good performances for off-line processing, and can be used to make subjective tests on acoustic quality in concert halls.

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 convolved one, and the signal recorded "live" in the rooms.

Listening 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. The CM-CONV code is actually the better one, because it is faster (also if it makes many more float multiplications), and gives a convolved signal more clean.

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