Software Implementation of B-Format Encoding and Decoding

Angelo Farina (*), Emanuele Ugolotti (**)

(*) Dipartimento di Ingegneria Industriale, Università di Parma,
Via delle Scienze - 43100 PARMA - tel. +39 521 905854 - fax +39 521 905705
E-MAIL: farina@pcfarina.eng.unipr.it - HTTP://pcfarina.eng.unipr.it

(**) ASK Automotive Industries, via Fratelli Cervi n. 79, 42100 Reggio Emilia
tel. +39 0522 388311 - fax. 0522 388499 - E-MAIL: tec_ask@sirnet.it

Abstract

New software tools are presented, which enable a standard PC equipped with a multichannel sound board to be used both for the creation of B-format signals, and for their playback over a suitable array of loudspeakers.

The encoding of B-format signals is obtained by convolution of each original soundtrack with a proper B-format impulse response, obtained from measurement in existing spaces or from computation by a room acoustics software.

The decoding is made thanks to a small dedicated program, which can be set to emulate a standard Ambisonics decoder (with fixed-gain shelf filters), or a new, wider class of decoders based on the recent theory about Energetic Analysis of Sound Fields. The decoder includes convolution with inverse filters which compensate for the irregularities of the loudspeakers.

The system was developed mainly as an analysis tool and as a support for subjective listening tests in the automotive field and in the room acoustics field.

1. Introduction

In many cases it is advisable to capture completely the three-dimensional sound field at a listening point, and subsequently to reproduce it as faithfully as possible in a suitable environment. This should in principle reproduce perfectly also the spatial effects, making it possible to recreate the direction of provenience of the wave-fronts.

Among various systems which have been proposed for obtaining this complete 3D recording and reproduction, the Ambisonics method has gained little popularity, despite its straightforward mathematical elegance and its demand for a little number of channels (four) for encoding the complete information.

In the author's opinion, this was caused not only by the well-known commercial problems, but also from the fact that this system was always regarded as an alternative to the standard stereo distribution path. Instead, the basics of the method can be used as the starting point of an objective measurement method for characterising the listening environments, through the measurement of B-format (4-channels) impulse responses. After the measurements, these B-format impulse responses can easily be applied by convolution to existing standard stereo signals: the result is a B-format sound track which emulates the playback of the original signals in the space where the impulse responses were measured.

For listening at such reconstructed B-format signals, a proper decoder is required. The original Ambisonics decoder was first analysed, and it was concluded that his decoding principle is not consistent with the quantities contained in the B-format signals, producing evident artefacts (sound out-of-phase coming from loudspeakers on the opposite side of the

virtual sound source). So a new decoding principle has been developed, based on recent enhancements to the classical sound intensity theory. This decoding scheme was implemented in a set of convolution filters, which are applied to the B-format signals for obtaining the speaker feeds. The creation of these decoding filters has been automated, starting from impulse response measurements taken on the reproduction array: this way also the loudspeaker response is individually equalised, and the system is automatically configured for any kind, number and position of loudspeakers. Obviously, the more regular is the shape of the array, the most stable is the 3D reconstruction of the sound field.

The whole encoding/decoding system has been implemented in a set of software tools, which allow for the measurement of B-format impulse responses making use of a Soundfield microphone or even of a single omnidirectional microphone (sequentially placed in 7 close positions). Other software tools allow for the convolution of the original signals (stereo or multi-track) with the B-format impulse responses, for the preparation of the decoding impulse responses, and for the real-time playback of the B-format signals over a multi-channel sound board. The software tools run on a standard low-cost PC: no expensive DSP board or digital workstation is needed.

The system has been tested for automotive applications, as a substitute of the actual measurement and listening system, which is based on the binaural technology. It resulted that the B-format reproductions are less sensitive to the equalisation of the reproduction system and to the position of the listener, and the spatial effect reconstruction is stable and more natural than with the binaural system. In the next future the system will be used for blind subjective comparisons, with the goal of investigate the relationship between objective parameters and subjective preference. This will be done both for car sound systems and for concert halls.

2. Measurement of B-format impulse responses

The measurement of B-format impulse responses is not very different from the measurement of standard impulse responses with a single omnidirectional microphone. The only evident difference is in the number of microphonic channels.

A standard Soundfield microphone outputs 4 channels, denoted W (omnidirectional) and X, Y and Z ("figure-of-eight pressure responses" along the three Cartesian axes, which are proportional to the sound particle velocity components). So the simplest thing is to employ an already-existing impulse response measurement system, such as MLSSA, TEF or AURORA, and repeatedly measure the impulse response from each of the 4 microphone channels. Alternatively, a new multi-channel impulse response measurement tool has been developed, as presented in [1]: thanks to a 4-channels sound board, the complete B-format impulse response can be measured during a single excitation of the system with a repeated MLS sequence.

The B-format impulse response can be stored as 4 separate files (as it is common, for example, for the MLSSA format, which does not support multi-channel data), as two stereo waveforms, or as a single, 4-channels waveform file. The latter solution is still too advanced nowadays, because most sound editing and convolution software do not recognise 4-channels .WAV files. A choice must be done about the problem caused by the –3 dB amplitude reduction of channel W compared to X,Y,Z: in fact, the Soundfield microphone is built with such a gain offset, and traditionally all the B-format recordings were made with it. So, for compatibility with the past, it was chosen to maintain this different gain. At the decoder stage, the W channel is re-amplified by + 3 dB, so that the subsequent math is not affected by this problem.

The above technique is effective and has no troubles: the only thing required is a Soundfield microphone and a good-quality, 4-channels sound board. These tools, nevertheless, are not actually widely diffused about sound engineers and recording companies.

So an alternative, low-cost implementation has been developed, as described in [2]. A single omnidirectional microphone can be used, together with a standard sound board: 7 impulse response measurements are separately made, moving the microphone from the reference position (R) to other 6 closely spaced positions, placed along the three Cartesian axes in both directions (X+, X-, Y+, Y-, Z+, Z-), as shown in fig. 1.

This way, the figure-of-eight responses along the three axes can be computed with a numerical technique quite similar to the one implemented in pressure-difference sound intensity probes.

In the following, the mathematical theory for making this computation is first described, and an application example is presented.

Let we look at a single pair of microphone positions placed along the X-axis at a relative distance **d**, as in fig. 2. Assume that the sound wave, travelling at speed **c**, is coming with an angle ϑ with the x-axis. Following the well-known theory of Sound Intensity probes [3], the required cosine-weighted pressure at the microphone #1 is given by:

$$p_1(\tau)\cdot\cos(\theta) = \int\left[\frac{p_1(t)-p_2(t)}{d}\cdot c\right].dt \qquad (1)$$

It is clear from (1) that the cosine-weighted pressure is proportional to the sound particle velocity component u_x, being $\rho\cdot c$ (air's impedance, ≈ 415 rayls) the ratio between the pressure and the particle velocity.

This computation is easier in the frequency domain. So the two pressure impulse responses measured at the two microphones are first FFT-transformed, then eqn. 1 is applied. Finally, the result is backward transformed to time domain through an inverse FFT. The process is repeated for each pair of microphone positions: along each axis, three pairs can be considered, two with spacing **d** and one with spacing **2·d**. The average of the first two is employed for estimating properly the higher frequencies, and the third one is better for low frequencies. The three computations are merged together in the frequency domain, through proper cross-over filters, and a single IFFT is then performed.

The above computation was implemented in a dedicated program, which reads the 7 omnidirectional impulse responses, and writes directly the 4 B-format impulse responses. The results are thus 4 separate mono .WAV files, because a proper flag for identifying 4-channels B-format .WAV file has not been standardized yet and most waveform editing programs cannot manage properly 4-channels files.

Fig. 3 shows a typical B-format Impulse Response, measured in the S.Maria del Fiore church in Florence. Note the higher background noise on the velocity channels at the end of the decay, due to the use of the pressure-difference integration technique: the IRs measured with a Soundfield microphone, instead, have velocity responses almost as quiet as the pressure channel W.

The use of experimental B-format impulse responses (with the Soundfield microphone) is by far the method which produces more realistic results: in a direct comparison experiment, the sound emitted in a concert hall through a dodechaedron loudspeaker was recorded through a Soundfield microphone. Afterwards, the B-format impulse response was measured through the same loudspeaker and microphone, and this IR was used for convolution with the original signal. A subjective comparison between the convolved signal and the true direct recording

revealed that they were almost identical, except for the background noise, which was higher in the true recording.

3. Computation (synthesis) of B-format impulse responses

Sometimes making B-format IR measurements is not the preferred choice, because a certain sonic ambience is required, and no hall with similar acoustics is available. Until wide collections of experimental B-format IRs are not available, the only possible alternative is to use artificial (synthetic) impulse responses. These could in principle be built "by hand", working with a waveform editor such as CoolEdit: the direct wave can be panned through the XYZ channels for giving 3D localisation, some early reflections can be added from different directions, and a reverberant tail can be appended with various degrees of correlation between the channels for emulating more-or-less diffuse sound fields.

This "by hand" construction requires nevertheless a deep knowledge of the mathematical structure of a B-format IR. Only a very little number of researchers around the world have gained such a knowledge by analysing experimental B-format IRs, and employing the Sound Intensity theory for understanding the phase relationships between pressure (W) and particle velocity components (XYZ). It must be noted that here we are speaking about instantaneous quantities, not the well-known time-averaged quantities which are defined in standard acoustics books [3]: although some work of Stanzial e Schiffer has been published [4], the energetic analysis of sound fields in the time domain is still an advanced research topic.

So actually the better way for preparing synthetic B-format impulse responses is to employ a general-purpose room acoustics program. Nowadays there are dozens of such programs, based on various algorithms such as Ray Tracing, Image Sources, Conical Beam Tracing, Triangular Beam Tracing (Pyramid Tracing), and so on. The author developed in past years one of these numerical codes, named Ramsete [5,6,7], based on Pyramid Tracing, which can be used for the purpose of creating synthetic B-format impulse responses; obviously, almost any other of the above mentioned programs can be used also, provided that at each receiver point not only the energetic time history is saved, but also the direction of provenience of each energy arrival.

With a room acoustics software, the room is first described through a 3D CAD tool. At each surface, proper acoustical properties are assigned, such as the (frequency dependent) absorption and diffusion coefficients. Then the sound sources and the receivers are introduced, as shown in fig. 4, and a separate impulse response is computed connecting each source with each receiver.

Typically, these impulse responses are octave-band energetic echograms. They have to be transformed in wide-band pressure (W) and particle velocity (XYZ) IRs through proper algorithms. This task is particularly easy for the programs who do not employ an hybrid scheme with different treatment for the early and late part of the impulse response. For example, in Ramsete at each energy arrival on a receiver, 13 numbers are stored: the energy in each of 10 octave bands, plus the three Cartesian components of the versor of the sound ray.

The omnidirectional, wide-band impulse response (W) is obtained by following these steps:

1) For each octave band, an echogram is built, placing a dirac's delta function at the arrival time of each energy contribution, with an amplitude equal to the square root of the energy. If two reflections arrive at the same sample, their amplitude is summed.
2) The echogram is passed through a 6-poles octave-band filter, centered on the proper frequency.
3) The result is accumulated over the ten frequency bands.

The other three channels (W, X, Y) are obtained with the same technique, but the amplitude is first multiplied by the Cartesian component of the energy vector on the corresponding axis: this means that some reflections can have negative amplitude, and when two of them arrive within the same sample, their amplitude is summed algebraically.

The above technique is not completely satisfactory: in fact, in the late part of the tail the number of energy arrivals per millisecond is significantly lower than in nature, and this causes a certain "roughness" of the tail, in comparison with experimental IRs. Furthermore, it can never happen that two out-of-phase reflections cancel out on the W channel, as all the contributions to the W channel are assumed to be positive: this means that in the reverberant tail the absolute value of the W samples are always greater than X,Y and Z, whilst in experimental impulse response this can be not true, as demonstrated by fig. 3. So the reverberant tail is always more diffuse than with experimental IRs. More research is still required for improving the reconstruction of B-format impulse responses starting from energetic echograms.

Fig. 5 shows an example of a synthetic B-format impulse response obtained with the Ramsete program in the geometric case depicted in fig. 4.

4. Encoding of B-format signals by convolution

Provided that a set of B-format impulse responses has been measured or computed for a given receiver position in an acoustic space and several different sound sources positions, it is possible to "place" a sound track in the virtual sound space simply by convolving the original (dry) signal with the proper B-format IR. Adding the results of the convolution of different sound tracks with IRs relative to different source positions, a complete "soundscape" can be created: for example, it is possible to "place" in their proper positions the single instruments of a "virtual orchestra", starting from multi-miked, multitrack studio recordings (which are almost perfectly anechoic) or from separately synthesised MIDI sequences.

Even when a standard stereo soundtrack is all what is available, it is possible to place a virtual pair of stereo loudspeakers in the virtual space, and reconstruct properly the room's ambience in such conditions. As in this case a certain degree of reverberation is already included in the original soundtrack, probably it is better to use B-format IRs not containing too much reverberation. As they are, anyway, standard WAV files, it is always possible to modify these B-format impulse responses, reducing the reverberation by applying a properly-shaped fade-out.

The convolution is a quite heavy computational task if performed in the time domain. In fact, the convolution of a long sequence x(i) (the original signal) with an impulse response h(t) having a length of N samples, is mathematically described by:

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

And thus it requires N multiplications and sums for each processed sample. As a typical impulse response length, at a sampling frequency of 44.1 kHz, is between 100,000 and 300,000 points, the number of multiplications per second is of the order of 10,000 billions…. Too much even for the fastest digital computers or DSP units. Furthermore, this has to be repeated four times (the 4 channels of the B-format), and for the number of soundtracks to be convolved with different IRs.

Fortunately, the processing can be done with a ridiculous computational effort in the frequency domain, employing the well-known algorithms called Select-and-Save and Overlap-

and-Save [8]. Fig. 6 shows the flow diagram of the select-save algorithm, which was employed in this work for implementing the B-format encoder.

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.

For obtaining this, FFTs of lenght M>N are required. Typically, a factor of 4 (M=4·N) gives the better efficiency to the select-save algorithm. As the process outputs only M-N+1 periodicity-free 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.

With such an algorithm, even a low-cost PC is capable of performing these computations in real-time, if the IR is not too long. A 200 MHz Pentium Pro, for example, has a real-time limit N_{max} around 65,000 points for the convolution of a single mono track with a 4-channels B-format IRs, producing a 4-channels output. This without an optimised FFT routine and working with floating-point math. Some preliminary tests on the new Pentium II – 300, making use of the optimised routine FFT-W, show that N_{max} is increased to 300,000 points, and the use of the MMX engine for integer computations has yet to be explored. This means that with latest generation CPUs, it will be possible to encode in real time 4 – 8 soundtracks with very long IRs, without the need of expensive DSP-based boards or workstations.

The only problem with the convolution approach is that the sound engineer has no real-time control on the process: there are not knobs to turn or sliders to move, the only choice is about the IRs to convolve. After the convolution, it is always possible to “pan” by traditional mixers between different versions of the same soundtrack (for example for emulating a moving source, panning between the convolutions with a series of IRs taken with closely spaced source positions), but it is clear that analogic encoders are much more easy to use and give the possibility to add some artistic effect to the surround mix.

No dedicated convolver was needed for performing the B-format encoding: the standard convolver already presented in [9] was used inside the CoolEdit PRO multi-track environment.

5. Decoding of B-format signals with Ambisonics-like methods

Traditionally, B-format signals are passed through an Ambisonics decoder, which delivers the proper feeds to an array of loudspeakers [10]. The basic 4-loudspeakers system allows for the decoding of only the three channels WXY, in the horizontal plane, whilst a full 3D decoder makes use also of the Z channel, but requires a larger loudspeaker array (typically a cuboid with 8 loudspeakers, or even more).

Let we first introduce the basic decoding equation of an Ambisonics B-format decoder, which computes the feed F_i for a generic loudspeaker placed in a regular array (all the loudspeakers are at the same distance from the center of the array, and the angles between the loudspeakers are always the same). Assume that the loudspeaker is placed in the space along a vector which starts in the center of the array (where the origin of the reference system is located), and let we call α, β and γ the angles between this vector and the three main axes x, y and z, as shown in fig. 7. The proper feed for the loudspeaker is a weighted sum of the 4 channels W, X, Y and Z:

$$F_i = \frac{1}{2} \cdot \left[G_1 \cdot W + G_2 \cdot \left(X \cdot \cos(\alpha) + Y \cdot \cos(\beta) + Z \cdot \cos(\gamma) \right) \right] \quad (3)$$

The gains G_1 and G_2 were originally delivered by proper analog shelf filters, for adjusting the soundfield reconstruction to two different psychoacoustics mechanisms, one holding for the low frequencies (below 500 Hz), and the other for higher frequencies. These gains are also influenced by the number N of loudspeakers in the array.

If the array is not regular, it is necessary to introduce different gains for each loudspeaker, instead of employing the same gains for all them. This topic was recently investigated by Moorer [11], who developed a mathematical formulation (derived from the original Ambisonics scheme) which allows for the computation of the gain matrix for any number and positions of a simmetrical 2D speaker array. The Moorer formulation, anyway, reduces simply to $G_1=G_2=1$ when the array is regular, and thus evidently it does not take into account the psychoacoustic shelf filtering. Furthermore, the case of complete 3D reconstruction is not covered, although certainly this is feasible with a simple extension of the Moorer equations.

For the following, only the cases regarding almost regular speaker arrays are considered. Even in this case, the proper values of the gains were never published clearly, and probably different manufacturers of hardware decoders were using different values. Jean-Marc Jot from IRCAM suggests these values for horizontal-only (without Z) decoding on regular arrays:

Version	Name	Author	G_1	G_2	$\Gamma=G_2/G_1$
a)	Concert Hall, all frequencies	D. Malham	$\sqrt{\frac{8}{3 \cdot N}}$	$\sqrt{\frac{8}{3 \cdot N}}$	1
b)	Studio, high frequencies	M. Gerzon	$\sqrt{\frac{8}{4 \cdot N}}$	$\sqrt{\frac{8}{2 \cdot N}}$	$\sqrt{2}$
c)	Studio, low frequencies	M. Gerzon	$\sqrt{\frac{8}{6 \cdot N}}$	$\sqrt{\frac{8 \cdot 2}{3 \cdot N}}$	2
d)	Studio, very low frequencies	J.M. Jot	$\sqrt{\frac{8}{2 \cdot N^2}}$	$\sqrt{\frac{8 \cdot 2}{N^2}}$	2

The first three cases are normalised so that the average sound density (proportional to the sum of the squares of the signal amplitudes coming from all the loudspeakers) is constant; the fourth equation is normalised in the hypothesis that the pressure values are summing (coherent sum, valid only at very low frequencies and at the centre of the array).

Apart from the normalisation, what happens is that the ratio $\boldsymbol{\Gamma}$ between G_2 and G_1 is varying between 1 and 2. Another valid source of information are the Gerzon's patents on the Ambisonics technology: the U.S. version of them is now freely available on the Internet. In U.S. pat. N. 3,997,725, these values are reported for a complete full-sphere 3D decoder:

Frequency	G_1	G_2	$\Gamma=G_2/G_1$
> 500 Hz	$\sqrt{2}$	$\sqrt{2}$	1
< 500 Hz	1	$\sqrt{3}$	$\sqrt{3}$

The above values are not consistent from an energetic point of view, as the total energy density at high frequency is 20% greater than at low frequency (probably this effect is not

audible, and probably it is compensated by the fact that at low frequency the sum tends to happen more in pressure than in energy). Again, what appears important is the ratio Γ between G_2 and G_1, which ranges between 1 and 1.732.

It must be noted that a value of $\Gamma = 1$ means that no anti-phase signal comes out from the loudspeakers placed in the opposite direction from the original sound provenience, and this ensures that the surround effect is consistent in a wide listening area.

Recently, Thierry Leconte suggested to have not simply two fixed gains, one at low frequencies and the other at high frequencies, but a continuously-varying ratio Γ between the two gains, with progressive transition between the low frequency and the high frequency behaviour.

Due to the difficulty to find the proper gains, and to the above suggestion, the software decoder here described was implemented so that a variable ratio Γ can be graphically introduced by means of a sort of 1/3 octave graphic band equaliser, as shown in fig. 8. The gains G_1 and G_2 are then found imposing the energetic equivalence, which yields:

$$G_1 = \sqrt{\frac{8}{\left(2+\Gamma^2\right)\cdot N}} \qquad ; \qquad G_2 = G_1 \cdot \Gamma \tag{4}$$

for 2D, horizontal-only decoding, and

$$G_1 = \sqrt{\frac{8}{\left(3+\Gamma^2\right)\cdot N}} \qquad ; \qquad G_2 = G_1 \cdot \Gamma \tag{5}$$

for 3D, full-sphere decoding. If in the lower graphic equaliser a not-flat response is introduced, the gains are altered correspondingly at each frequency, so that an overall response correction is performed.

In the actual not-real-time version of the decoder, a 2-columns data table, containing the values of G_1 and G_2 for each of the 31 frequency bands, is saved in an editable ASCII file (called the “shelf filter” file), which is required for proper working of the software decoder. If no file is introduced, the decoder assumes a value of 1 for the ratio ρ at every frequency.

When a real-time version of the software decoder will be realised, this double graphic equaliser control will be available for real-time adjustment during the playback.

6. Self-adjustment of the decoder based on B-format IR measurements.

The location of loudspeakers in space and the assignment of the reproduction channels to them were usually made “by hand”. A novel feature of the software decoder is its ability to automatically adjust itself, based on a simple and straightforward measurement of a B-format impulse response performed in the center of the reproduction array.

Making use of the multichannel impulse response measurement system already described in [1], an automated B-format measurement is made exciting subsequently each loudspeaker in the array: for example, a basic 2D array of just 4 loudspeakers gives 4 B-format impulse responses.

Analyzing each of these IRs, the software is capable to detect the effective director cosines [$\cos(\alpha)$, $\cos(\beta)$, $\cos(\gamma)$] of each loudspeaker, which are required in the decoding formula (3): they are easily obtained, being proportional to the amplitude of the direct wave

recorded on the X, Y and Z channels. The proper sign is found by comparison with the W channel direct wave.

But these measurements are useful also for another, very important task: the W impulse response is used for the creation of an inverse filter (which is simply another impulse response), which is designed so that the frequency response and the absolute delay of each loudspeaker are perfectly equalised. This means that the array can be built with speakers of different kinds and makers, placed at distances form the center not exactly equal, and even at not very exact angular positions: the self-adjustment procedure takes care of all these mismatches, and compensates as much as possible for them.

It must be noted, anyway, that the best results are obtained employing matched, high-quality loudspeakers, carefully placed in a regular array.

For the design of the inverse filters different options are possible. More details are presented in [9] and [12]. If the goal is simply to give a frequency equalisation, then the Neely and Allen approach produces short and stable minimum-phase filters, but the phase is uncontrolled, and this can be a problem for an Ambisonics array. In principle, the Mourjopoulos least-squares inversion is optimal, as it can linearise also the phase response, and even the room's reverberation is deconvolved out. In practice, anyway, the room's reverberation removal can work only in a very little space region around the measurement microphone, and it resulted that a complete Mourjopoulos inversion usually cause intolerable artifacts at listening positions far from the array's center.

A very promising technique is the Kirkeby's inversion based on the regularisation technique [12], applied to the direct wave only: it is very fast to compute, it controls also the phase, but it does not attempt to correct for the room's reverberation, making the inverse filter reasonably short and quite robust for other listening positions.

Anyway, the creation of these inverse filters is a topic about which the research is not finished yet. Actually the creation of the inverse filters has to be done "by hand" (by means of the software tools presented in [9]), so that it is possible to adjust the various parameters which are required by each inversion technique: the software decoder simply asks for the inverse filter of each playback channel, and this is convolved with the filters created by the decoder itself, producing a final filter network as shown in fig. 9.

Fig. 10 shows the user's interface of the decoder program: it can be seen that it does not asks for the input B-format files. In fact, the decoder has the simple goal of producing the matrix of decoding filters, which is a set of 4xN impulse responses, with pre-defined names; for example, DEC_X_08.WAV is the X filter for the loudspeaker # 8. The decoder simply asks for the location of the B-format impulse responses (and their inverses) employed for the self-adjustment, which have to be named consistently; for example, IMP_W_04.WAV is the impulse response measured on channel W of the Soundfield microphone with loudspeaker #4, and INV_W_04.WAV is its inverse filter.

7. Real-time convolution player

After the complete set of 4xN filters has been created, it is possible to employ it for playing B-format signals. The player is again simply a convolver, which works accordingly to the scheme of fig. 9. Each of the input channels (WXYZ) is convolved with a separate set of N impulse responses. Then the 4 results for each loudspeaker are simply summed together, and sent to the proper output channel.

Being in this case the filters usually quite short (typically 1024 to 4096 points), all these convolutions and sums can be done in real time through the select-save algorithm already described. The computation speed is further increased by the fact that the initial FFT on the 4

input channels can be used for all the processing, and furthermore the sum between the components can be made, for each output channel, before the IFFT is performed. So at each FFT frame only 4 direct FFT and N inverse FFT are required (the multiplications and sums are substantially negligible on the total number of operations). Even a 90 MHz Pentium can do this task for 4-channels playback, and a 200 MHz Pentium Pro has computing power widely in excess for an 8-channels playback (probably the power is enough even for an higher number of playback channels, but the author does not have a suitable soundboard, neither so many good loudspeakers….). This means that the convolution player can be used on a large installed base of low power computers, provided that low-cost multichannel sound boards are installed.

In any case, the convolution player can also be used for batch production of already-decoded waveform files. Although this is usually not very useful, in some cases it is advisable to prepare the pre-decoded files in advance, so that they can simply be played (with a multi-track software such as CoolEditPro or SAW) by the final user. This way, for example, it is easy to prepare a 5.1 compatible soundtrack, to be played through a standard horizontal-only loudspeaker setup for home-theater applications.

Obviously the "Ambisonics effect" will happen only if this soundtrack is played exactly on the same setup in which the self-adjustment of the decoder was made: on different 5.1 systems the inverse filters will be mismatched with the loudspeakers, and the results could be unpredictable (no experiment in this sense was made yet).

The user's interface of the convolution player is shown in fig. 11.

8. Conclusion and future work

Four software tools have been described: they allows for the measurement of B-format impulse responses without the need of a Soundfield microphone, for the convolution of these IRs with original soundtracks, producing a B-format mix, for the preparation of the decoder filters for a given loudspeaker array and for the real-time playback of the B-format signal over it.

At the time of writing, these software tools are yet at an early development stage: for example the first one is still a character-based batch program, the second is not yet optimised for speed on modern CPUs, the third does not include the automatic inverse filter computation, and does not interact in real-time with the real-time player.

Nevertheless the sound comes out, and the effect is really great: the author does not have an analog encoder/decoder to compare with, so the only available comparison is with a traditional "transaural" system based on 2-channels binaural techniques [13]. In a direct comparison experiment, it was evident that the B-format system gives stable and robust image localisation in a wide listening area, independently from the orientation of the listener, while the binaural systems works only for a very precise position and orientation of the listener's head. The adoption of the stereo dipole technique, as suggested in [12], improved only slightly the image stability with the binaural technique: the B-format system is superior both in delivering proper localisation cues and for the naturality of the sound.

In the future the research will prosecute on various fronts: new techniques for designing optimal inverse filters will be evaluated, the software will be translated from Fortran to C++ for improving speed and responsivity to the user, an easiest user interface will be added. The optimal frequency dependence of the ρ ratio between shelf-filter gains will be explored through subjective preference tests.

Another possible front is the research of a different decoding principle, not based simply on a linear combination of the 4 channels WXYZ, so that when the sound is strongly oriented in a particular direction (i.e., when the Sound Intensity Level L_i approaches the Sound Density

Level L_d), no signal is emitted from loudspeakers not located near that direction. This could fix one of the weak points of the original Ambisonics and derived matrix decoders, the incapability of reproducing sound fields with a $L_d - L_i$ value lower than 1.76 dB for 2D-decoding and 3 dB for 3D-decoding.

In any case, the performances of the system will be evaluated by subjective localisation and preference tests, in comparison with the binaural setup already in use.

9. References

[1] Farina A., Ugolotti E. – "Automatic Measurement System for Car Audio Applications" – Pre-Prints of the 104th AES Convention, Amsterdam, 16-19 may 1998.

[2] Farina A., Tronchin L., "3D Impulse Response measurements on S.Maria del Fiore Church, Florence, Italy" – Proc. of ICA98, International Conference on Acoustics, Seattle (WA) 20-26 June 1998.

[3] Fahy F.J. - *Sound intensity* - Elsevier Applied Science, London, 1989

[4] G. Schiffrer and D.Stanzial, "Energetic Properties of Acoustic Fields", J. Acoust. Soc. Am. 96, pp. 3645-3653, 1994.

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

[6] Farina A., "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, Computational Mechanics Publications, Southampton (GB) 1995.

[7] Farina A., "Verification of the accuracy of the Pyramid Tracing algorithm by comparison with experimental measurements by objective parameters" - ICA95 (International Conference on Acoustics), Trondheim (Norway) 26-30 June 1995.

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

[9] A. Farina, F. Righini, "Software implementation of an MLS analyzer, with tools for convolution, auralization and inverse filtering" - Pre-prints of the 103rd AES Convention, New York, 26-29 September 1997.

[10] Gerzon M., "Ambisonics in Multichannel Broadcasting and Video" - *Journal of Audio Engineering Society*, Vol. 33, Number 11 pp. 859 (1985).

[11] Moorer J.A., "Music recording in the age of multi-channel" - Pre-prints of the 103rd AES Convention, New York, 26-29 September 1997.

[12] Kahana Y, Nelson P.A., Kirkeby O., Hamada H., "Objective and subjective assessment of systems for the production of virtual acoustic images for multiple listeners" - Pre-prints of the 103rd AES Convention, New York, 26-29 September 1997.

[13] A. Farina, E. Ugolotti, "Subjective comparison of different car audio systems by the auralization technique" - Pre-prints of the 103rd AES Convention, New York, 26-29 September 1997.

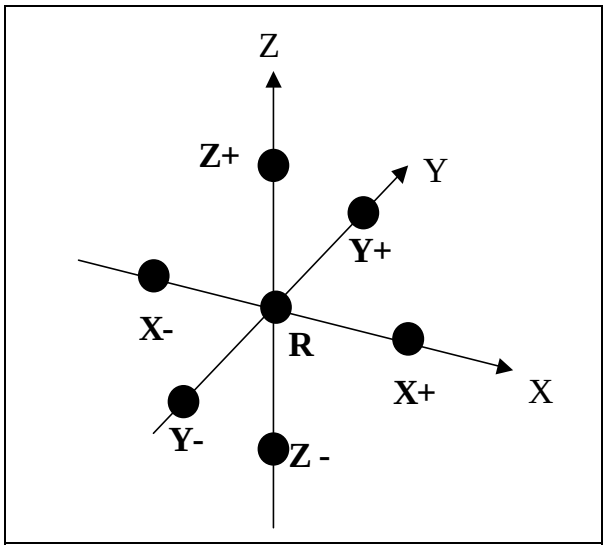

Fig. 1 – Layout of the 7 microphone positions

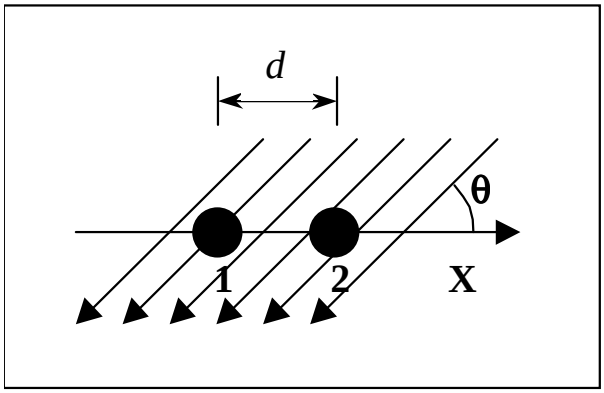

Fig. 2 – Scheme of the pressure-gradient computation by a 2-microphones probe.

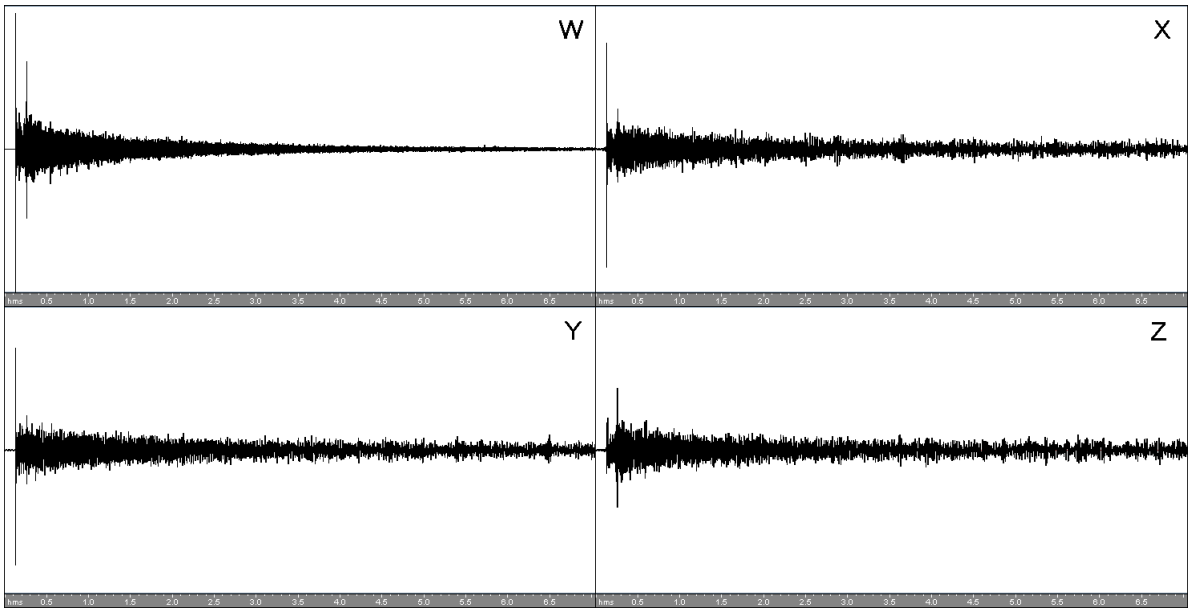

Fig. 3 – B-format impulse response measured in the S.Maria del Fiore Church, Florence

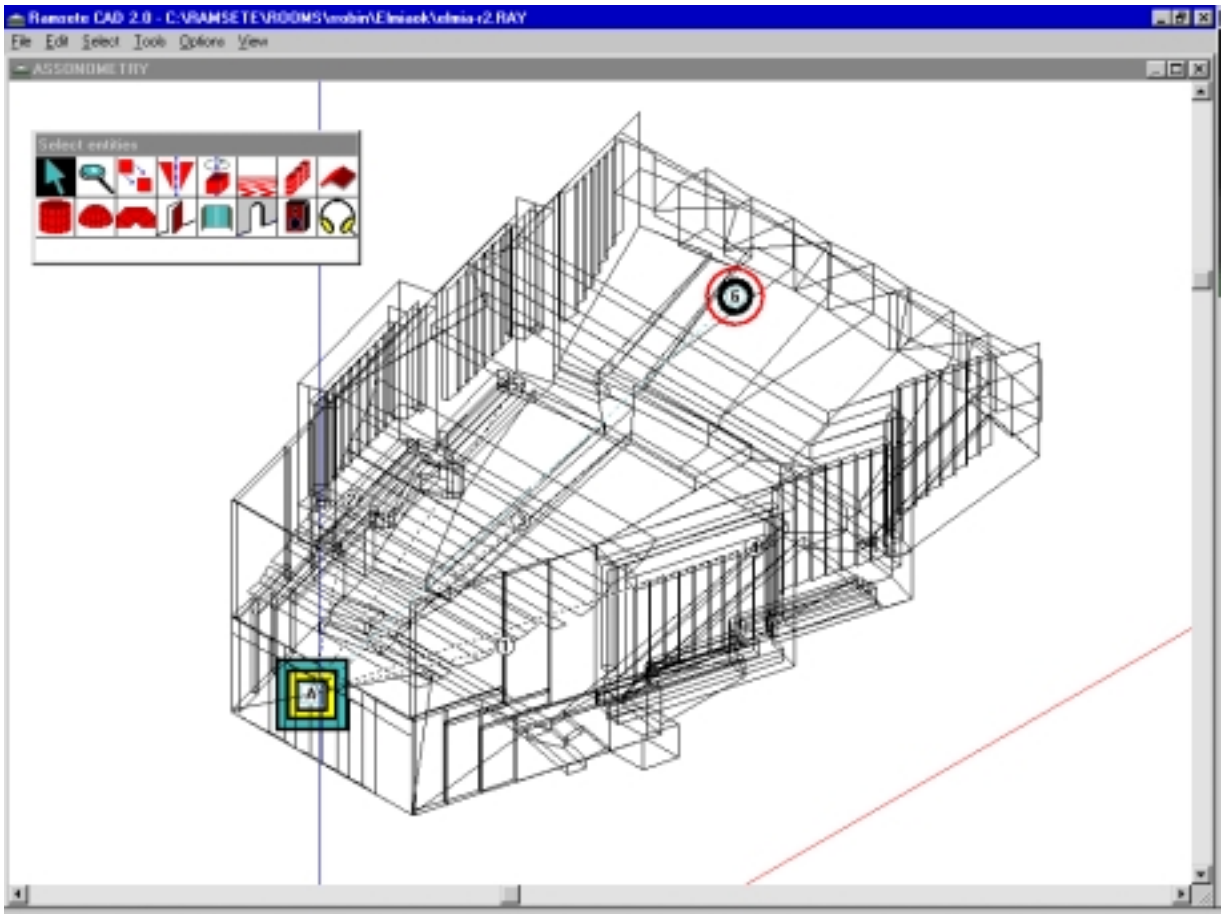

Fig. 4 – definition of the room geometry in Ramsete CAD.

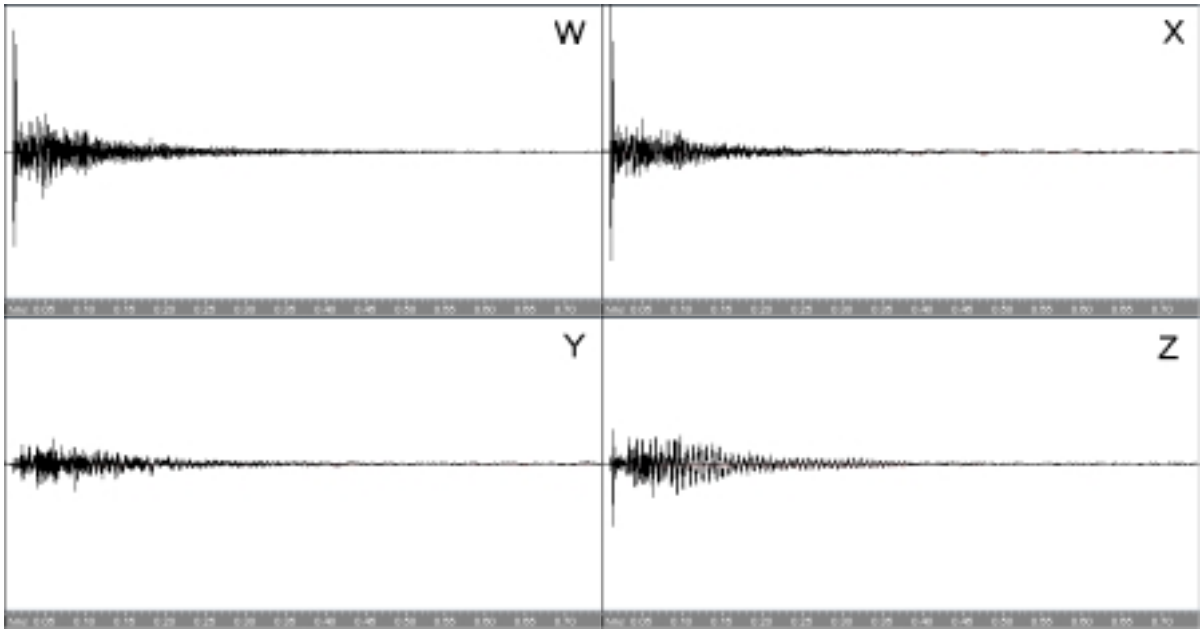

Fig. 5 – B-format impulse response computed with the Ramsete pyramid tracing code.

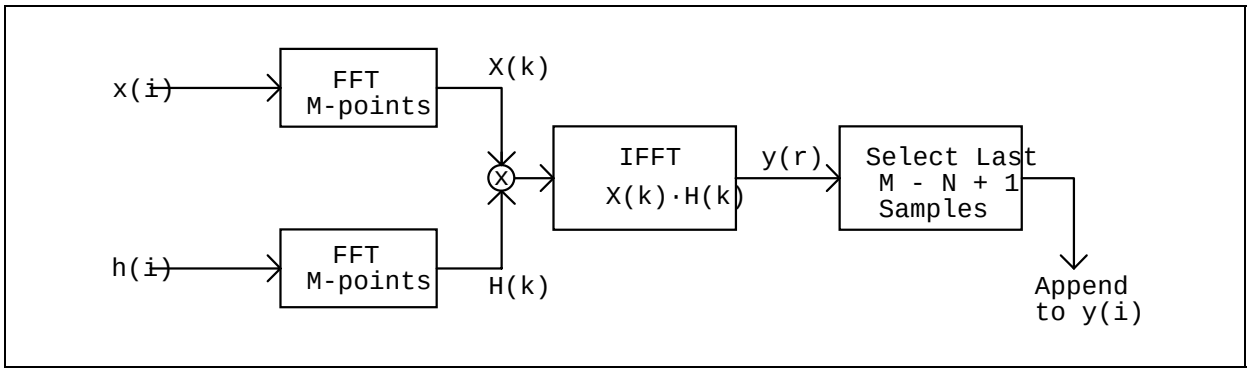

Fig. 6 - SELECT-SAVE Flow Chart

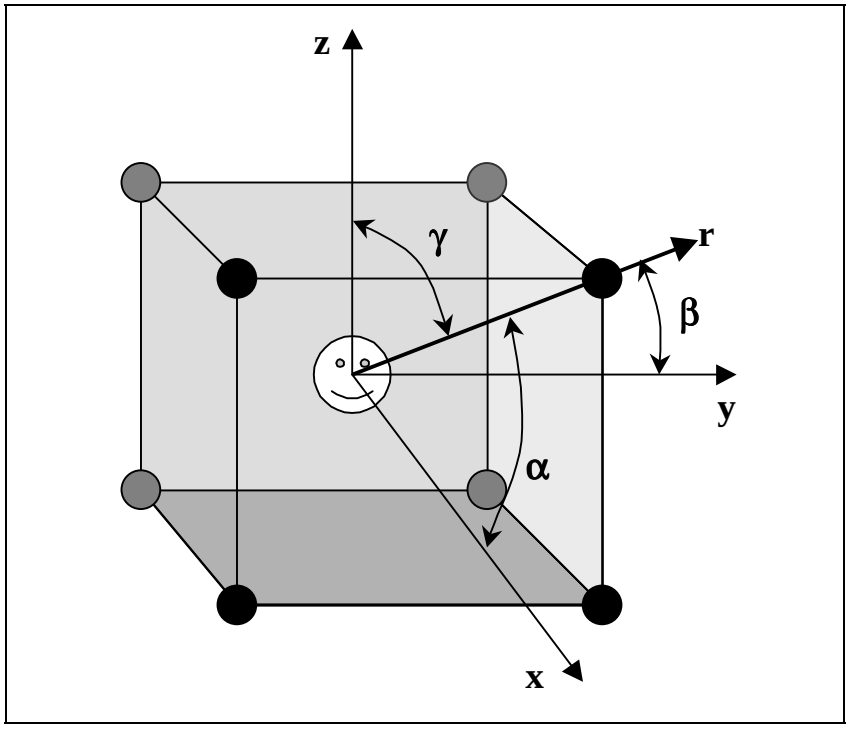

Fig. 7 – Geometry of the speaker array.

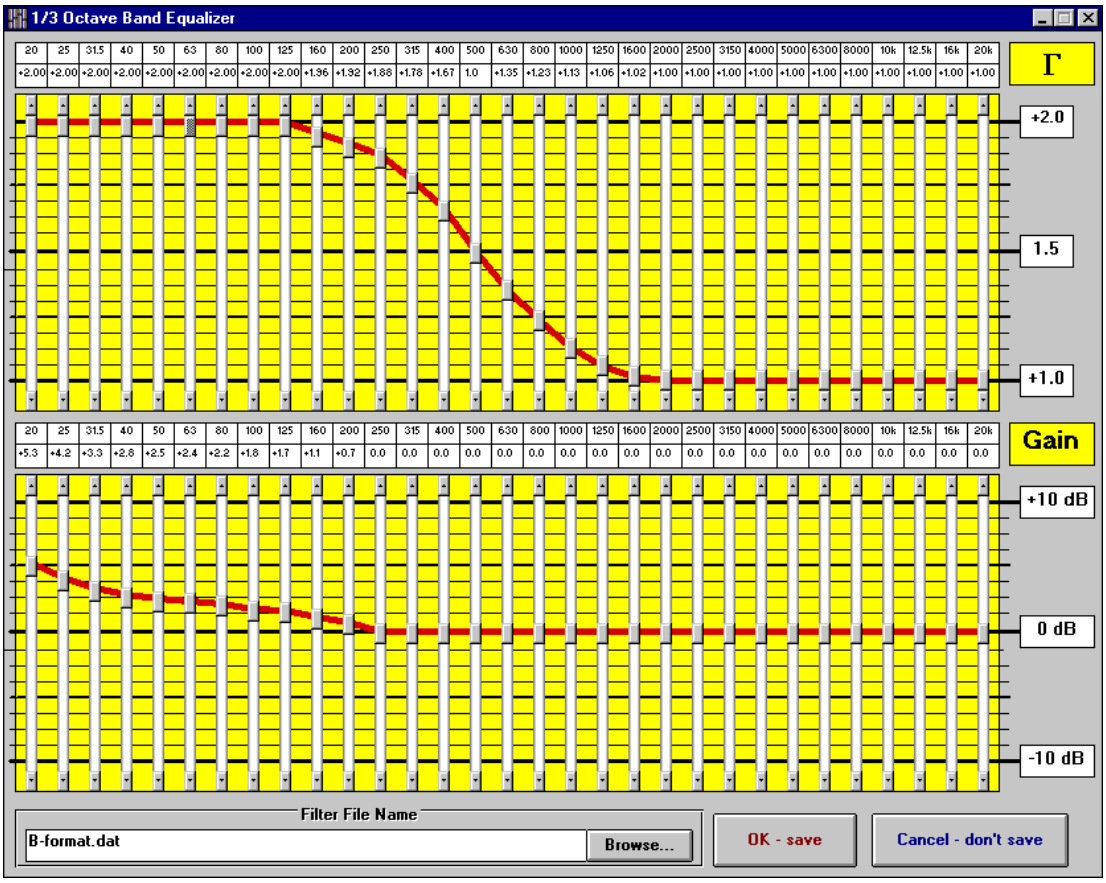

Fig. 8 – Graphic equalizer for setting the ratio Γ and the overall gain.

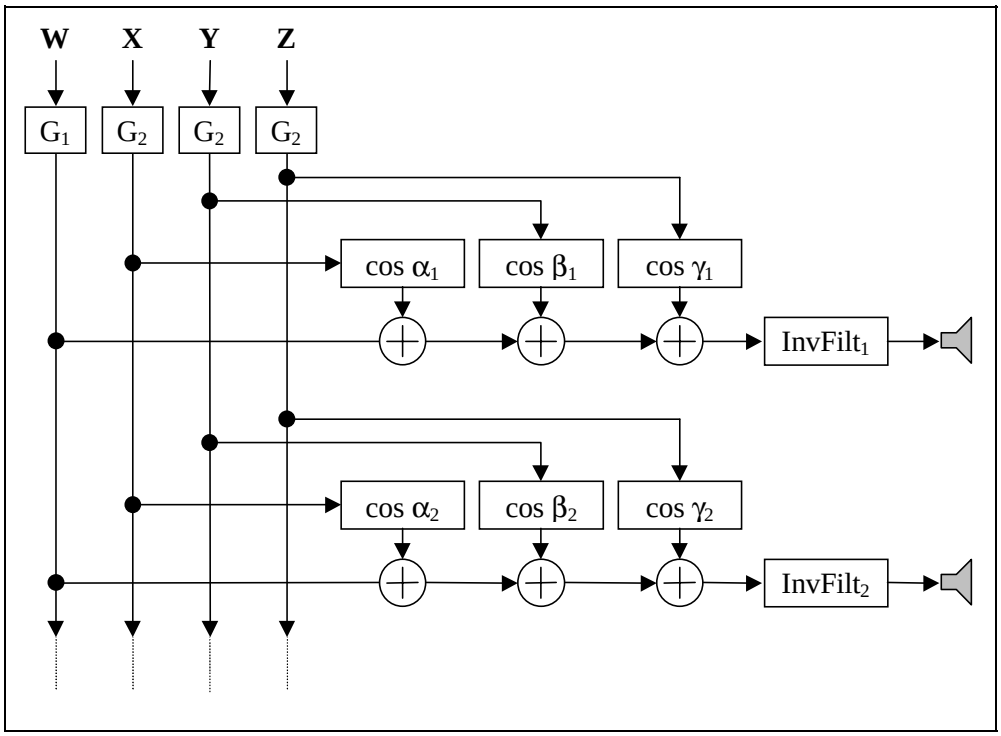

Fig. 9 – Flow diagram of the decoder – only the first two speaker feeds are shown

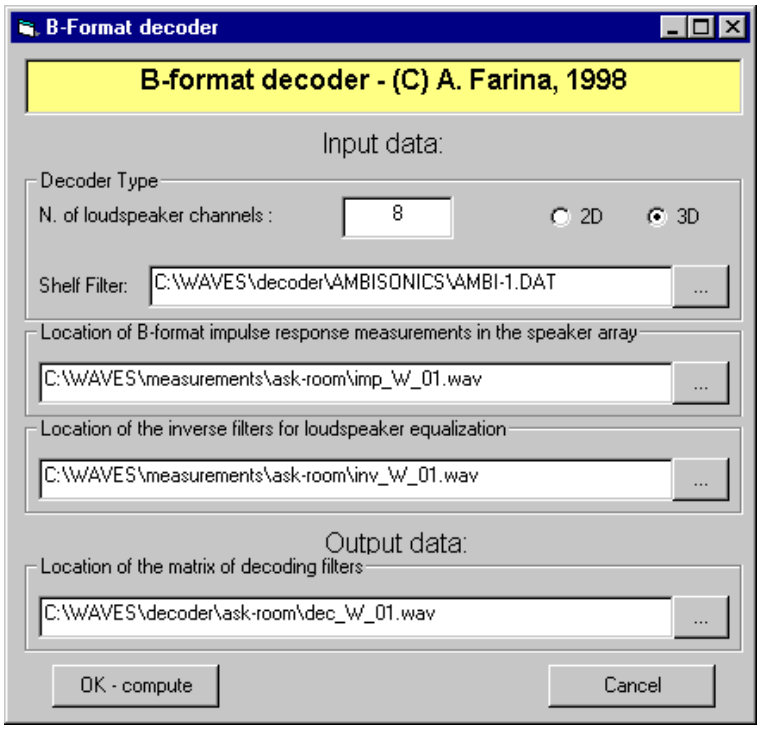

Fig. 10 – User's interface of the software B-format decoder

Fig. 11 – User's interface of the B-Format convolution player