WESPAC IX 2006
The 9th Western Pacific Acoustics Conference
Seoul, Korea, June 26-28, 2006

ROOM IMPULSE RESPONSES AS TEMPORAL AND SPATIAL FILTERS

Angelo Farina
University of Parma
Industrial Engineering Dept., Parco Area delle Scienze 181/A, 43100 Parma, ITALY
E-mail: farina@unipr.it

ABSTRACT

Nowadays it is possible to measure room impulse responses inside famous theatres and concert halls, and use them as filters for applying high quality reverberation to recordings and soundtracks. However, also spatial information can be measured and replicated this way, once the concept of omnidirectional source and receiver has been deleted.

The paper will focus on the extension of the current measurement and rendering techniques, making use of arrays of sources and microphones, allowing for a compact and elegant representation of the spatial transfer function of a room, which preserves the reciprocity principle.

KEYWORDS: Concert Hall Acoustics, Measurement Techniques

INTRODUCTION

The concept of impulse response is nowadays widely accepted as a physical-mathematical model of the behavior of a linear, time-invariant system, characterized with just one input port and one output port.

In acoustics, this concept is usually applied to the study of sound propagation from an emission point and a receiver point, located within the same environment.

Nevertheless, this technique is usually implemented by means of an omnidirectional sound source, and by an omnidirectional receiver (pressure microphone). This way any spatial information is lost, both on the emission pattern of real sources, and on the direction of arrival of the wavefronts arriving on the receiver.

In the past it was attempted to obtain partially some spatial information by means of directive transducers (both sources and receivers). But this happened without a rational basis, with just one significant exception, represented by the Ambisonics method derived by Gerzon in the seventies [1].

Recently, advanced impulse-response measurement techniques have been developed [2], capable of performances significantly better than previous methods; furthermore, it is now possible to build, at reasonable costs, multichannel sound systems making use of large arrays of loudspeakers and microphones.

Only very recently a method for characterizing the emission directivity of sound sources has been proposed, employing the same mathematical basis already employed for characterizing the directivity of microphones. More specifically, this method was proposed by Dave Malham in 2003 [3], and it employs an expansion of the directivity of a point sound source by means of 1st-order spherical harmonics (O-format signal).

We are proposing now to extend and generalize this approach: both the sound source and the receiver can be spatially characterized by means of an expansion in a series of spherical harmonics, stopping the expansion to a reasonably-high order (3^{rd}, 4^{th} or even 5^{th} order).

This way, a complete characterization of the spatial transfer function between the emission and receiver points is obtained.

IMPULSE RESPONSE MEASUREMENTS

When spatial information is neglected (i.e., both source and receivers are point and omnidirectional), the whole information about the room's transfer function is contained in its impulse response, under the common hypothesis that the acoustics of a room is a linear, time-invariant system.

This includes both time-domain effects (echoes, discrete reflections, statistical reverberant tail) and frequency-domain effects (frequency response, frequency-dependent reverberation).

The following figure shows how a room can be seen, under these hypotheses, as a single-input, single-output "black box".

The system employed for making impulse response measurements is conceptually described in fig. 1. A computer generates a special test signal, which passes through an audio power amplifier and is emitted through a loudspeaker placed inside the theatre. The signal reverberates inside the room, and is captured by a microphone. After proper preamplification, this microphonic signal is digitalized by the same computer which was generating the test signal.

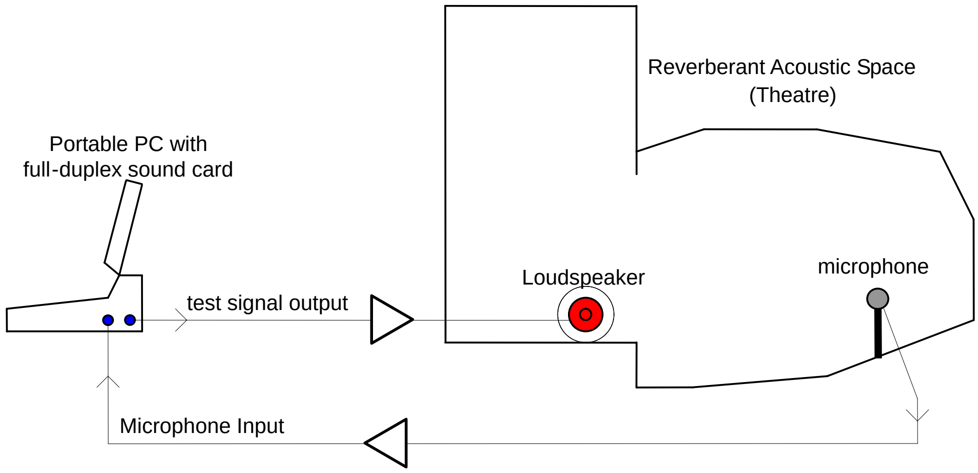


Fig. 1 – schematic diagram of the measurement system

A first approximation to the above system is a "black box", conceptually described as a Linear, Time Invariant System, with added some noise to the output, as shown in fig. 2.

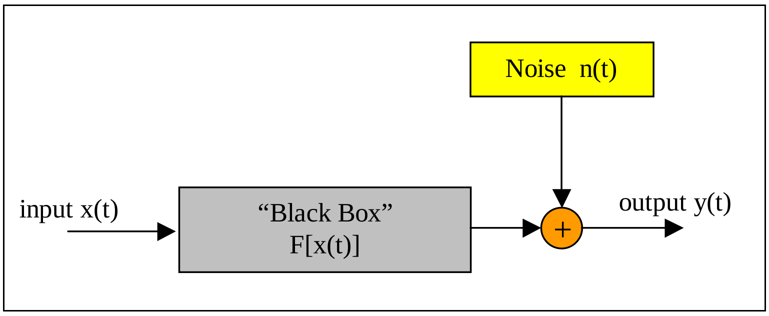


Fig. 2 – A basic input/output system

In reality, the loudspeaker is often subjected to not-linear phenomena, and the subsequent propagation inside the theatre is not perfectly time-invariant.

The quantity which we are interested to measure is the impulse response of the linear system h(t), removing the artifacts caused by noise, not-linear behavior of the loudspeaker and time-variance.

The method chosen, based on an exponential sweep test signal with aperiodic deconvolution, provides a good answer to three above problems: the noise rejection is better than with an MLS signal of the same length, not-linear effects are perfectly separated from the linear response, and the usage of a single, long sweep (with no synchronous averaging) avoids any trouble in case the system has some time variance.

The mathematical definition of the test signal is as follows:

$$x(t) = \sin\left[\frac{\omega_1 \cdot T}{\ln\left(\frac{\omega_2}{\omega_1}\right)} \cdot \left(e^{\frac{t}{T}\cdot\ln\left(\frac{\omega_2}{\omega_1}\right)} - 1\right)\right]$$

This is a sweep which starts at angular frequency ω_1, ends at angular frequency ω_2, taking T seconds.

When this signal, which has constant amplitude and is followed by some seconds of silence, is played through the loudspeaker, and the room response is recorded through the microphone, the resulting signal exhibit the effects of the reverberation of the room (which "spreads" horizontally the sweep signal), of the noise (appearing mainly at low frequencies) and of the not-linear distortion.

These "distorted" harmonic components appear as straight lines, above the "main line" which corresponds with the linear response of the system. Fig. 3 shows both the signal emitted and the signal re-recorded through the microphone.

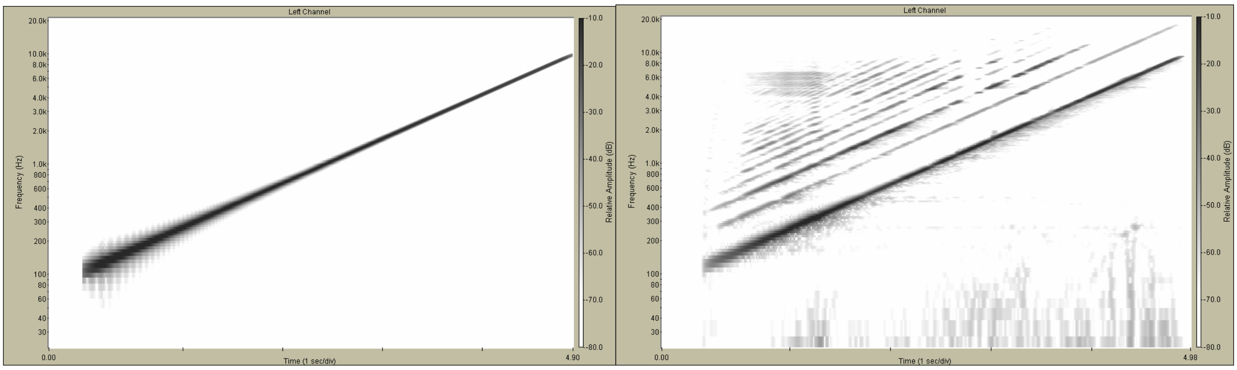


Fig. 4 – sonograph of the test signal x(t) and of the response signal y(t)

Now the output signal y(t) has been recorded, and it is time to post-process it, for extracting the linear system's impulse response h(t).

What is done, is to convolve the output signal with a proper filtering impulse response f(t), defined mathematically in such a way that:

$$h(t) = y(t) \otimes f(t)$$

The tricks here are two:

to implement the convolution aperiodically, for avoiding that the resulting impulse response folds back from the end to the beginning of the time frame (which would cause the harmonic distortion products to contaminate the linear response)

to employ the Time Reversal Mirror approach for creating the inverse filter f(t)

In practice, f(t) is simply the time-reversal of the test signal x(t). This makes the inverse filter very long, and consequently the above convolution operation is very "heavy" in terms of number of computations and memory accesses required (on modern processors, memory accesses are the slower operation, up to 100 times slower than multiplications).

However, the author developed a fast and efficient convolution technique, which allows for computing the above convolution in a time which is significantly shorter than the length of the signal. [4]

It must also be taken into account the fact that the test signal has not a white (flat) spectrum: due to the fact that the instantaneous frequency sweeps slowly at low frequencies, and much faster at high frequencies, the resulting spectrum is pink (falling down by -3 dB/octave in a Fourier spectrum). Of course, the inverse filter must compensate for this: a proper amplitude modulation is consequently applied to the reversed sweep signal, so that its amplitude is now increasing by +3 dB/octave, as shown in fig. 5.

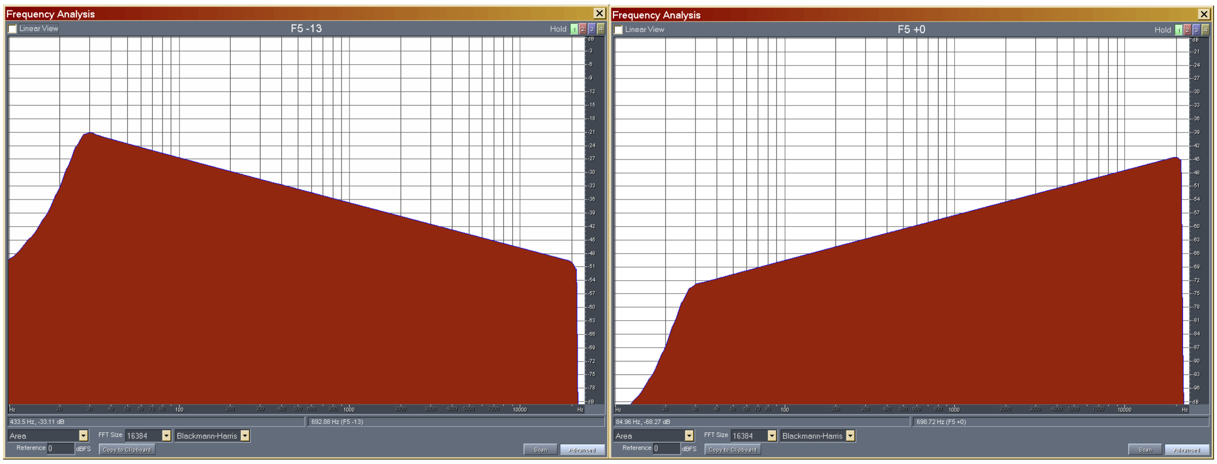


Fig. 5 – Fourier spectrum of the test signal (left) and of the inverse filter (right)

When the output signal y(t) is convolved with the inverse filter f(t), the linear response packs up to an almost perfect impulse response, with a delay equal to the length of the test signal. But also the harmonic distortion responses do pack at precise time delay, occurring earlier than the linear response. The aperiodic deconvolution technique avoids that these anticipatory response folds back inside the time window, contaminating the late part of the impulse response.

Fig. 8 shows a typical result after the convolution with the inverse filter has been applied.

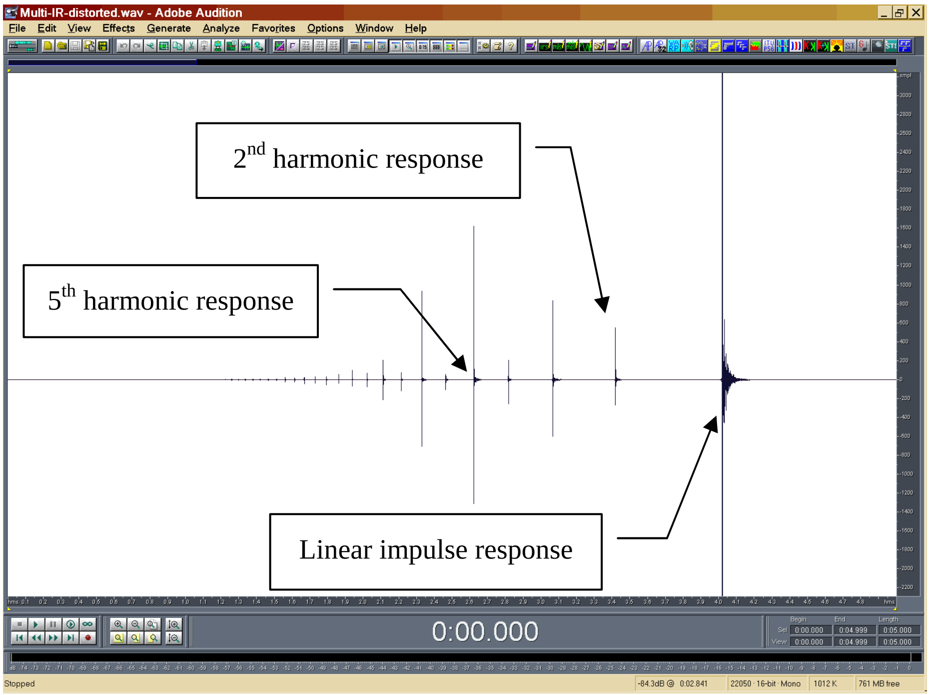


Fig. 6 – output signal y(t) convolved with the inverse filter f(t)

At this point, applying a suitable time window it is possible to extract just the portion required, containing only the linear response and discarding the distortion products.

The advantage of the new technique above the traditional MLS method can be shown easily, repeating the measurement in the same conditions and with the very same equipment. Fig. 9 shows this comparison in the case of a measurement made in an highly reverberant space (a church).

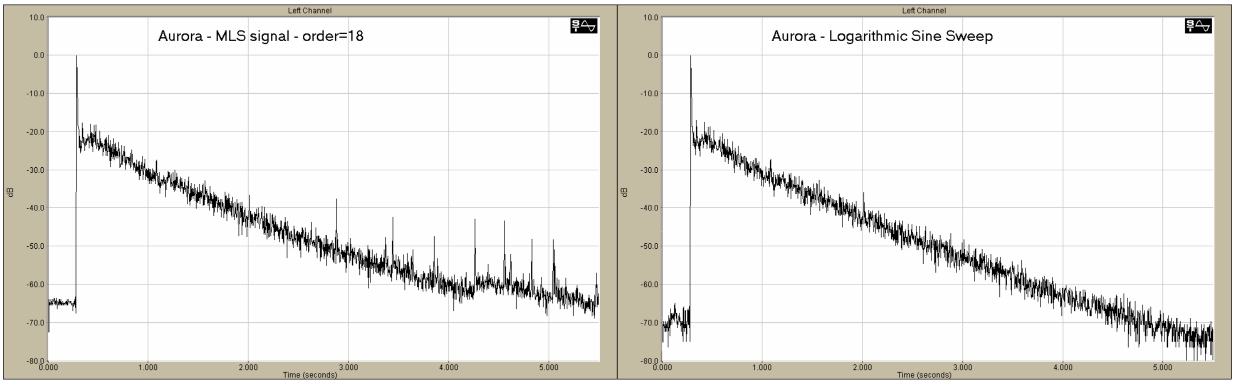


Fig. 7 – comparison between MLS and sine sweep measurements

It is easy to see how the exponential sine sweep method produces better S/N ratio, and the disappearance of those nasty peaks which contaminate the late part of the MLS responses, actually caused by the slew rate limitation of the power amplifier and loudspeaker employed for the measurements, which produce severe harmonic distortion.

This method has nowadays wide usage, and is often employed for measuring high-quality impulse responses which are later employed as numerical filters for applying realistic reverberation and spaciousness during the production of recorded music [5].

DIRECTIVE SOURCES AND RECEIVERS

When we abandon the restriction to omnidirectional sources and receivers, it becomes possible to get also spatial information. A first basic approach is to "sample" the room's spatial response with a number of unidirectional transducers, pointing all around in a number of directions.

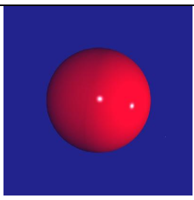
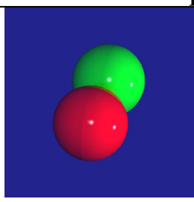
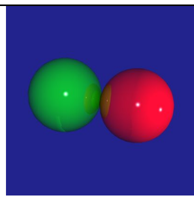
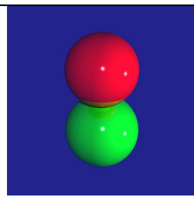
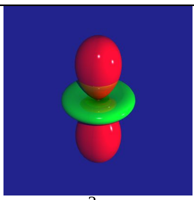
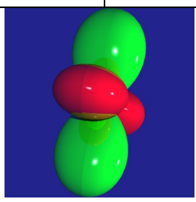
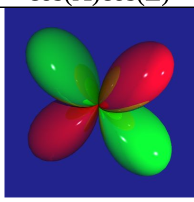
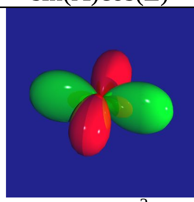
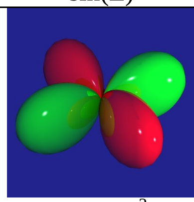
However, such an approach often ends in requiring to repeat a large number of measurements while rotating the transducers in steps, resulting in long measurement times. The approach, furthermore, is not easily scalable: all the measurements need to be performed and analyzed for "covering" uniformly a notional sphere surrounding each transducer.

The approach proposed here is to employ a spherical harmonic expansion of sound field around the source and receiver points. This corresponds to a two-dimensional, spatial Fourier transform, conceptually similar to what is employed in image processing, but working in a spherical coordinate system instead of in a plane Cartesian one.

This approach is the basis of the Ambisonics method [6], initially employed with an expansion limited to 0^{th}-order and 1^{st}-order spherical harmonics around the microphone. Here this concept is extended to higher orders, and adopted for describing both what happens at the source and at the receiver.

For the sake of concision, here we report the mathematical formulas in polar coordinates, as function of the Azimuth angle A and the Elevation angle E, and a pictorial representation for the spherical harmonics of order 0, 1, and 2 – the equations for higher orders are indeed quite common to find.

Table 1 – spherical harmonics up to 3rd order

Order 0	 0.707107		Order 1	 cos(A)cos(E)	 sin(A)cos(E)	 sin(E)
Order 2	 $1.5\sin^2(E)-0.5$	 cos(A)sin(2E)		 sin(A)sin(2E)	 $\cos(2A)\cos^2(E)$	 $\sin(2A)\cos^2(E)$

Unfortunately, "native" loudspeakers or microphones having directivity patterns corresponding to the above spherical harmonic functions are available only for orders 0 and 1 (monopoles and dipoles).

However it is possible to "synthesize" the pattern of a spherical harmonics by combining the signals being fed to, or coming from, a number of individual transducers being part of a closely-spaced transducer array.

The recombination is possible with the following formula:

$$y = \sum_{i=1}^{N} f_i \otimes x_i$$

Where f_i are a set of suitable "matched" FIR filters, designed in such a way to synthesize the required spherical-harmonic pattern. The design of the filtering coefficients can be performed numerically (least-squares approach), starting from a huge number of impulse response measurements made in free field and with a source (or receiver) located in P different polar positions around the transducer array.

The system is solved with the least-squares approximation, imposing the minimization of the total squared error, obtained summing the squares of the deviations between the filtered signals and the theoretical signals v_k:

$$\varepsilon_{tot} = \sum_{k=1}^{P}\left[\sum_{i=1}^{N}\left(f_i \otimes x_{ki}\right) - v_k\right]^2$$

The solution of an overconditioned system requires some sort of regularization. The Nelson-Kirkeby method [7] provides this solution (in frequency domain), which can be adjusted by means of the regularization parameter β:

$$F_i = \frac{X^T \cdot V}{X^T \cdot X + \beta \cdot I}$$

These inverse numerical filters have the advantage that they automatically compensate for the deviation between the responses of the individual transducers, and also for acoustical shielding or diffraction effects due to the mounting structure.

The most basic of such a closely-spaced transducer array is a spherical array. The following figure shows a source array and a microphone array.

Figure 9 – spherical arrays of loudspeakers (left) and microphones (right)

Once a set of spherical harmonics (in emission or in reception) has been measured, it is possible to recombine them for creating any three-dimensional polar pattern, with an error becoming smaller as the order increases. So it is possible to create the emission directivity pattern of a real musical instrument, or to synthesize the response of an ultra-directive virtual microphone, and to aim them in any direction wanted.

This recombination, again, is trivial: it is just matter of summing the signals coming from each of the spherical harmonics patterns with proper gains. This is already well known with reference to the “receiving” spherical harmonics, which are employed for the reconstruction of a virtual sound field in the high-order Ambisonics method (HOA). The possibilities opened by the measurement of a set of impulse responses which are spatially-expanded in spherical harmonics both at the emission and reception ends is yet to be fully explored.

However, the measurements can be efficiently performed employing a PC equipped with a multichannel sound card. Nowadays a system capable of 24 simultaneous inputs and 24 simultaneous outputs can cost less than 3000 USD, all included. Such a system can be easily employed for performing measurements up to 3rd order (16 harmonics) both in emission and in reception: a sequence of 16 sine sweeps is played, each of them being simultaneously fed with different gains and polarities to the 24 individual loudspeakers being part of the spherical emission array. The signals of the 24 microphones are recorded, and subsequently processed for the deconvolution of the impulse response, and for recomputing the 16 spherical harmonic signals. At the end of the measurement, which takes approximately 8 minutes if 15s-long sweeps are employed, a complete set of 16x16=256 impulse responses are obtained.

This set is a complete characterization of the room impulse response, containing both the time-frequency information, and the spatial information as "seen" both from the source and the receiver. It is therefore possible to derive subsequently, by post-processing the measured set of impulse responses, the virtual impulse response produced by a source having arbitrary directivity and aiming, as captured by a microphone also having arbitrary directivity and aiming.

The data measured also allow for spatial analysis, computation of spatial parameters, pictorial representation of the spatial information as colour maps, and high quality rendering of the recorded spatial information by projection over a suitable three-dimensional sound playback system.

SUMMARY

The method proposed here can be seen as an extension and generalization of the method initially proposed by Gerzon for characterizing the acoustical response o concert halls for the posterity. It removes the limitation of the original approach, which did only deal with omnidirectional sources, and which did analyze the spatial information at the receiver by means of a spherical-harmonics expansion terminated after just the 1st order.

It is expected therefore that, once a collection of these multi-input, multi-output impulse responses will have been measured in a significant number of theatres and concert halls, it will be possible to analyze these data for reaching a deeper understanding of the spatial properties of the sound field, and to assess how these spatial properties affect the human listening perception.

REFERENCES

1. Michael Gerzon - "Recording Concert Hall Acoustics for Posterity", JAES Vol. 23, Number 7 p. 569 (1975)
2. Angelo Farina (2000). Simultaneous measurement of impulse response and distortion with a swept-sine technique. 108th AES Convention, Paris, 18-22 February 2000.
3. Dave Malham - "Spherical Harmonic Coding of Sound Objects - the Ambisonic 'O' Format", Proceedings of the AES 19th International Conference, Schloss Elmau, Germany June 21-24, 2001, pp54-57
4. A. Torger, A. Farina – "Real-time partitioned convolution for Ambiophonics surround sound", 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics - Mohonk Mountain House New Paltz, New York October 21-24, 2001
5. A. Farina, R. Ayalon – "Recording concert hall acoustics for posterity" - 24th AES Conference on Multichannel Audio, Banff, Canada, 26-28 June 2003
6. M.A. Gerzon - "Ambisonics in Multichannel Broadcasting and Video", J. Audio Eng. Soc., vol. 33 no. 11, pp. 859-871 (1985 Nov.)
7. O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution and multichannel systems using regularization," IEEE Trans. Speech Audio Process. 6(2), 189–194 (1998).