Measuring impulse responses containing complete spatial information

Angelo Farina, Paolo Martignon, Andrea Capra, Simone Fontana

University of Parma, Industrial Eng. Dept., via delle Scienze 181/A, 43100 PARMA, ITALY

ABSTRACT

Traditional impulse response measurements did capture limited spatial information. Often just omnidirectional sources and microphones are employed.

In some cases it was attempted to get more spatial information employing directive transdudcers: known examples are binaural microphones, figure-of-8 microphones, and directive loudspeakers.

However, these approaches are not scientifically based, and do not provide an easy way to process and visualize the spatial information.

On the other side, psychoacoustics studies demonstrated that "spatial hearing" is one of the dominant factors for the acoustic quality of rooms, particularly for theatres and concert halls.

Of consequence, it is necessarily to reformulate the problem entirely, describing the transfer function between a source and a receiver as a time/space filter. This requires to "sample" the impulse response not only in time, but also in space. This is possible employing spherical harmonics for describing, with a predefined accuracy, the directivity pattern of both source and receiver.

It is possible to build arrays of microphones and of loudspeakers, which, by means of digital filters, can provide the required directive patterns. It can be shown how this makes it possible to extract useful information about the acoustical behavior of the room, and to make high-quality auralization.

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.

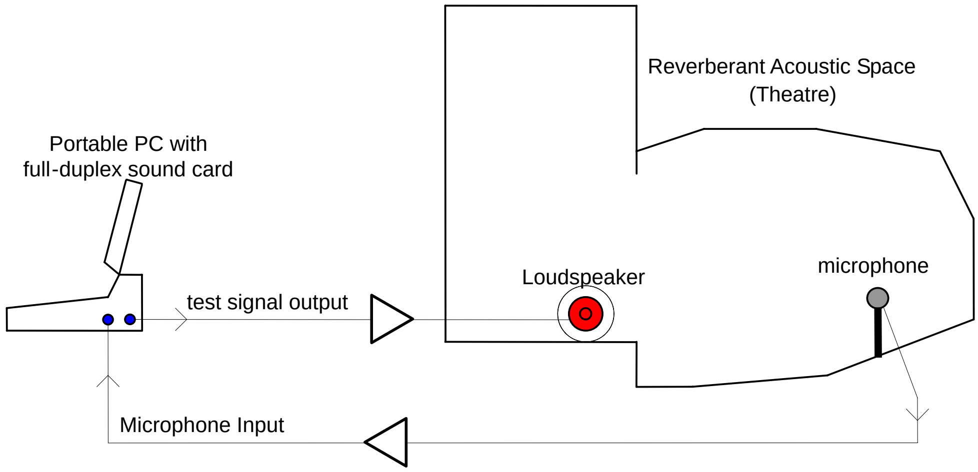


Fig. 1 – schematic diagram of the measurement system

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.

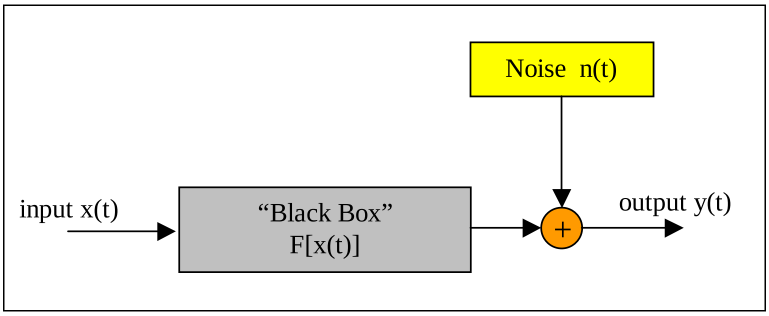


Fig. 2 – A basic input/output system

The usage of proper test signals and deconvolution techniques (exponential sine sweep, aperiodic deconvolution in time domain) make it possible to avoid the problems caused by nonlinearities in the transducers, by the background noise and by the fact that the system is not really time-invariant [2].

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 [4].

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.

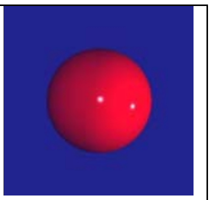
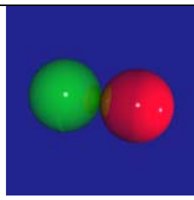
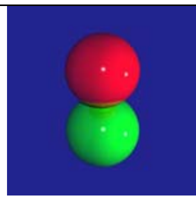
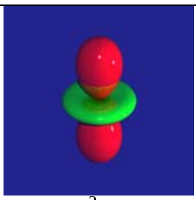
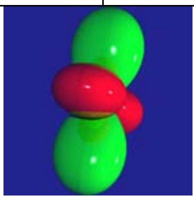
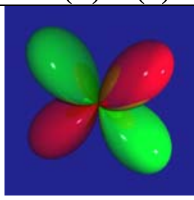
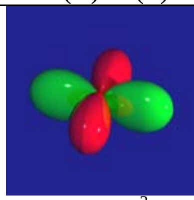
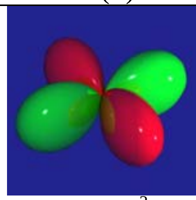
However, such an approach often ends in repeating a large number of measurements while rotating the transducers in steps, resulting in long measurement times. The approach, furthermore, is not easily scalable: all the measurements need to be performed and analyzed for "covering" uniformly a notional 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 [5], initially employed with an expansion limited to 0th-order and 1st-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 harmonic 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 [6] 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 1^{st} 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. Farina, R. Ayalon – "Recording concert hall acoustics for posterity" - 24th AES Conference on Multichannel Audio, Banff, Canada, 26-28 June 2003
5. M.A. Gerzon - "Ambisonics in Multichannel Broadcasting and Video", J. Audio Eng. Soc., vol. 33 no. 11, pp. 859-871 (1985 Nov.)
6. 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).