Audio Engineering Society

Conference Paper

Presented at the Conference on
Sound Field Control
2016 July 18–20, Guildford, UK

Measuring Spatial MIMO Impulse Responses in Rooms Employing Spherical Transducer Arrays

Angelo Farina[1], Lorenzo Chiesi[2]

[1] *Dipartimento di Ingegneria Industriale, Università di Parma, Parco Area delle Scienze 181/A - 43124 Parma - ITALY*

Correspondence should be addressed to Angelo Farina (angelo.farina@unipr.it)

The paper presents a new measurement method aimed to characterise completely the sound propagation form a point to another point inside a room, taking into account the directionality of the source and of the receiver. The method makes use of two spherical arrays of transducers, almost uniformly scattered on the surface of a rigid sphere, which can synthesize arbitrary polar patterns. The paper describes the beamforming method employed for synthesizing the required polar patterns over a wide frequency range, and how to process the results in various ways: graphical mapping of sound reflections inside the room, reconstructing the trajectories of such reflections and auralization, when for the first time both source and receiver can be allowed to freely rotate during the test.

1 Introduction

The paper presents the development of a new measurement technique for room acoustics, based on a spatial MIMO approach [1]. In practice, we employ a spherical array of transducers both for the sound source and for the receiver. Both are currently spherical arrays with 32 transducers, and this allows for the complete spatial control of the sound propagation in the room during the measurement of the impulse response.

Employing a suitable matrix of FIR filters, superdirective beamforming is performed both at the source and at the receiver, creating a large number of prescribed polar patterns in emission and in reception [2]. The measurements are performed in "raw mode", that is simply feeding signal to one emitting transducer, while capturing the signal from all 32 microphones. This is repeated for all 32 emitting transducers, creating a huge matrix of 32x32 "raw" room impulse responses.

This way, it is also possible to employ a spherical source equipped with just one directive loudspeaker, which is rotated in 32 different directions by means of a specially−built two−axes turntable. We did also build a true 32−loudspeakers spherical source, which allows to feed many or all of them simultaneously [3]. This paper provides some comparative data between the two variants of the source. The beamforming filter matrixes can later be applied to the massive set of raw data, with several different possible approaches:

1) High Order Ambisonics [4] both at the source and at the receiver, creating a MIMO matrix of spherical harmonics functions, typically 16x16 (at orders up to 3rd).
2) SPS (Spatial PCM Sampling) [5], that is radiating sound in just one of the 32 directions, by means of a superdirective beamformer aimed in the direction of each of the 32 loudspeakers, and synthesizing 32 ultradirective microphones aimed at the same directions of the real 32 capsules
3) Emulating the radiation pattern of a given musical instrument (violin, trumpet, cello, piano, etc.) at the source, and a binaural receiver looking in various directions at the receiver.
4) Any possible combination of the above described approaches (for example, an Ambisonics source and a binaural receiver)

Particularly important in terms of sound field control is the SPS approach [5]. An huge SPS data set contains the complete spatial information, which can be analysed with the goal of reconstructing the spatial transfer function between the two points.

The result of the analysis can be represented by colour dynamic maps, showing how the sound propagates in space form the source to the receiver while the time is passing.

Furthermore, the analysis can reconstruct the trajectories of single discrete reflections, locating the reflection points on the boundaries, for reflections up to 3rd order, and with simultaneous arrival of up to two wavefronts.

Finally the spatial MIMO matrix of impulse response can be employed for auralization, with almost any kind of reconstruction methods, such as binaural, High Order Ambisonics and Spatial PCM Sampling, and making it possible to simulate a sound source with arbitrary directivity, and the effect of rotating the source and/or the receiver in realtime during the auralization.

A demo session will be set up accompanying the presentation of this paper, where these auralization capabilities will be exploited, providing a degree of realism, which was not obtainable with traditional auralization methods, where the time−variant spatial transfer function is not properly reconstructed. The point is that the radiation pattern of any sound source is not time−invariant (except for loudspeakers).

This is due to two factors: musicians and singers move during their performance, and their radiation polar patterns change over time depending on the note they are playing or singing. Representing the radiation of a real source by means of a static radiation pattern and a single, mono anechoic signal recorded with a single microphone in a reference position is absolutely unrealistic, as the spatial transfer function between source and receiver will be "frozen". In reality, as the radiation pattern is continuously changing, the transfer function between source and receiver will also change continuously.

But if we record the anechoic soundtrack employing a suitably large number of microphones all around the source, they will capture this spatial variation of radiation. Hence, feeding all the channels of such a multichannel anechoic recordings as the inputs of the MIMO IR matrix, will recreate at the listener the effect of this variability. The improvement obtained with the new auralization method will be made audible to the listeners during the demo session.

2 Processing Of Microphone Arrays

The processing technique to derive from the signals of an array of *M* microphones a number of "virtual microphones" with specific directivity is called beamforming. The following figure and equation shows a generic processor where filters coefficients must be determined.

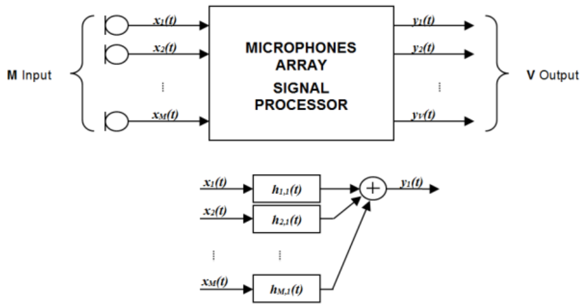

Figure 1: Scheme of the mics array signal processor

$$y_v(t) = \sum_{m=1}^{M} x_m(t) * h_{m,v}(t) \tag{1}$$

To find the filtering coefficients $h_{m,v}$ the authors developed a technique applicable to any microphone array, based on the numeric inversion of characterization impulse responses of the microphones measured in an anechoic chamber or simulated by a variety of methods (closed form mathematics, FEM...), as described in [2].

In particular, the impulse response of the *M*-channels microphone array is captured, subject to incident plane waves from a large number of characterization directions *D* (usually 362 directions with nearly uniform distribution constitute an adequate spatial oversampling for a 32 channels microphones array like the Eigenmike™). Each measurement determines a row of the matrix $c[n]$ of size DxM.

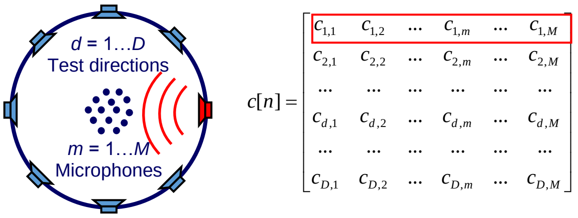

Figure 2: Microphones array characterization

Then a matrix of target directivities Q is defined, with dimensions DxV: it describes for each of the V virtual microphones to synthesize the desired gain in each of the D directions of measurement.
For describing the target directivity pattern, it is convenient to use a mathematical definition; for example the directivity of a cardioid of n^{th} order is:

$$Q_n(\vartheta,\varphi)=\left[0.5+0.5\cdot\cos(\vartheta)\cdot\cos(\varphi)\right]^n \qquad (2)$$

Nevertheless, this approach can also be employed for creating virtual microphones with much more complex patterns, such as high order harmonics or even strange patterns, which cannot be described analytically: it is simply required to specify a gain value for each of the D directions.
The matrix of filters to be determined must satisfy the following linear equation system, which can be solved numerically by means of the LS method, once transformed into the frequency domain, so that a solution can be evaluated separately for each frequency:

$$\overset{D\times M}{\|c[n]\|} * \overset{M\times V}{\|h[n]\|} = \overset{D\times V}{\|Q\|}\cdot\delta[n] \overset{DFT}{\Rightarrow} \overset{D\times M}{\|C[k]\|}\cdot\overset{M\times V}{\|H[k]\|} = \overset{D\times V}{\|Q\|} \qquad (3)$$

The following equation provides the solution of the LS problem with the addition of a casualization term to the numerator and a regularization term to the denominator.

$$\overset{M\times V}{\|H[k]\|} = \frac{\overset{M\times D}{\|C[k]\|^*}\cdot\overset{D\times V}{\|Q\|}\cdot e^{-j\pi k}}{\overset{M\times D}{\|C[k]\|^*}\cdot\overset{D\times M}{\|C[k]\|}+\beta[k]\cdot\overset{M\times M}{\|I\|}} \qquad (4)$$

The filters thus determined in frequency domain can be transformed back to time domain, to obtain a matrix of FIR coefficients working as a microphone beamformer for V virtual microphones.

$$\overset{M\times V}{\|h_{\mathrm{Receiver}}[n]\|} = IDFT\left[\overset{M\times V}{\|H[k]\|}\right] \qquad (5)$$

3 Processing of Loudspeaker Arrays

Similarly to what described for microphone arrays, it is possible to define the following processing system capable of generating driving signals for a loudspeaker array made of S transducers, creating the signals of W virtual sources with arbitrarily defined directivity.

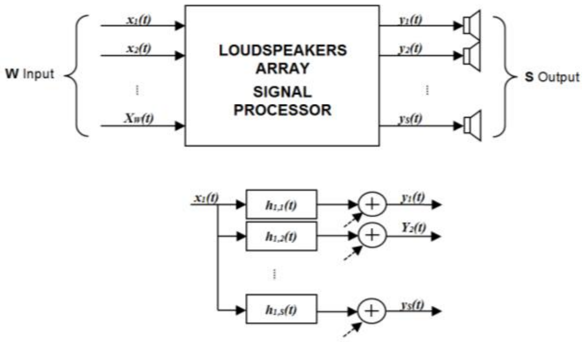

Figure 3: Scheme of the signal processor for a loudspeaker array

$$y_s(t)=\sum_{w=1}^{W} x_w(t) * h_{w,s}(t) \qquad (6)$$

In this case the impulse response of each speaker in the array must be measured with a microphone located in each of the D directions of characterization, thus determining a column of the matrix $c[n]$ of size DxS. Also in this case, considering a 32-loudspeakers array, the nearly uniform distribution of 362 directions is adequate.

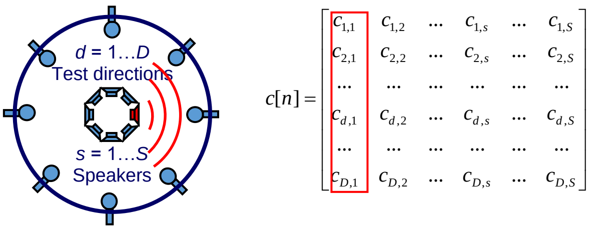

Figure 4: Loudspeaker array characterization

Again, the target directivity Q must be defined for every direction D, creating a matrix with dimensions DxW, describing for each of the W virtual sources to synthesize the desired directivity in each of the D directions of characterization.
Similarly to the case of virtual microphones, it is therefore possible to define the following linear equation system, and its solution H, in the frequency domain.

$$\overset{D\times S}{\|c[n]\|} * \overset{S\times W}{\|h[n]\|} = \overset{D\times W}{\|Q\|}\cdot\delta[n] \overset{DFT}{\Rightarrow} \overset{D\times M}{\|C[k]\|}\cdot\overset{M\times V}{\|H[k]\|} = \overset{D\times V}{\|Q\|} \qquad (7)$$

$$\overset{S\times W}{\|H[k]\|} = \frac{\overset{S\times D}{\|C[k]\|^*}\cdot\overset{D\times W}{\|Q\|}\cdot e^{-j\pi k}}{\overset{S\times D}{\|C[k]\|^*}\cdot\overset{D\times S}{\|C[k]\|}+\beta[k]\cdot\overset{S\times S}{\|I\|}} \qquad (8)$$

When converting back to time domain, the only trick consists in performing the transposition of the filter matrix, to preserve the conventional operation direction (from input space of W directive sources to the output space of S speaker feeds):

$$\overset{W\times S}{\left\| h_{Source}[n] \right\|} = IDFT\left[\overset{S\times W}{\left\| H[k] \right\|} \right]^T \qquad (9)$$

4 "Virtual" Loudspeaker Array

The goal we had in building a loudspeaker array was to use it as a source in MIMO (Multiple Input Multiple Output) room impulse response measurements [1]. In this application the loudspeakers of the array are not used simultaneously but one at a time. It was therefore possible to avoid the construction of a real array containing 32 transducers realizing instead a sphere with a single loudspeaker that can be oriented in any direction by means of a two axes positioner robot, while keeping the centre of the sphere always in the same point in space (with an accuracy of +/- 5mm).
The sphere used is made of plexiglass and has a diameter of 25cm. A wide-band, 80W, 5 inch. loudspeaker (CIARE PM135ND) was placed inside it, together with some sound-absorbing material and a counterweight to allow for smooth rotation. The positioning system consists of an old B&K turntable revamped with stepper motor and an ad-hoc controller capable of driving also the second axis. The entire system is powered and controlled trough an Ethernet link, also carrying Power over Ethernet (PoE), with a control software written in Matlab.
The measurement is performed by sequentially pointing the sphere in 32 different directions almost equispaced (the same geometry as the microphone capsules of an Eigenmike™), simulating an array of 32 loudspeakers.

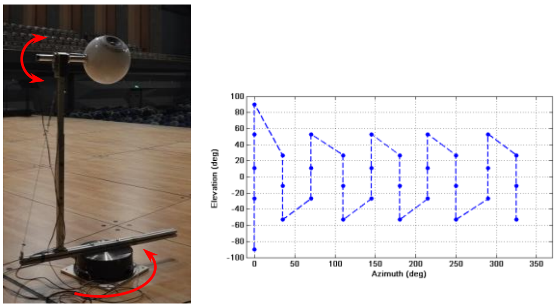

Figure 5: The "simulated" speaker array

4.1 Calibration of the virtual loudspeaker array

Taking advantage of the symmetry properties of the sphere around the speaker's axis, the system characterization was reduced to the measurement in an anechoic chamber of 181 impulse responses around the device, along a single circumference. The first measurement is performed along the loudspeaker axis, then the upper positioner axis is moved by 1° each time, rotating the speaker upwards.

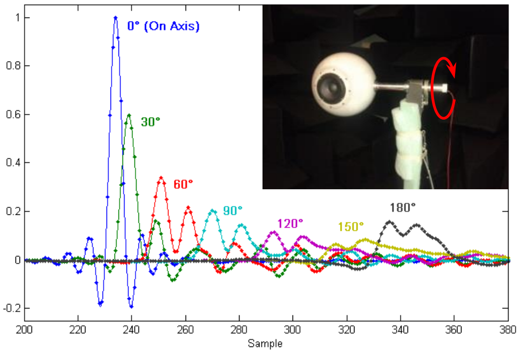

Figure 6: Measured impulse responses at various angles from the loudspeaker axis

As shown in figure 6, during the rotation the time of arrival of the first wavefront changes, as the distance of the loudspeaker from the microphone is changing. The C matrix necessary for the synthesis of the source beamformer filters can be generated by properly selecting some of the 181 measured impulse responses.
Each cell of the matrix in fact is indexed by a particular direction of characterization d and by a specific speaker s. Therefore for each cell it is possible to calculate the angular distance between the axis of the loudspeaker s and the direction of characterization d, then the impulse response measured at the nearest angle is chosen.
The following figures show the results of the synthesis of an omnidirectional virtual speaker prove that the system responds correctly in the band from 30Hz to 30kHz. At higher frequencies, the beaming effect of the speakers affect the uniformity of the directivity pattern.

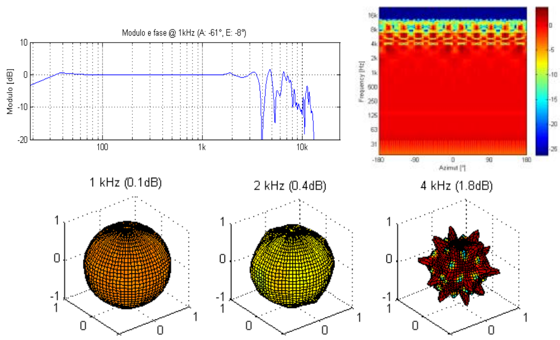

Figure 7: Frequency response, XY plane and 3D directivity of omnidirectional virtual speaker

By comparison, the horizontal directivity of a B&K 4292 dodecahedron is reported, which demonstrates how the uniformity of emission and the bandwidth of the our virtual loudspeaker array are clearly superior to traditional systems, that start beaming already at 1kHz.

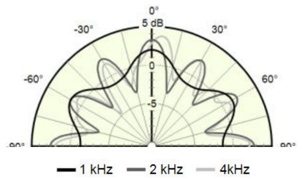

Figure 8: Directivity of a reference dodecahedron

Of course the system can also synthesize virtual speakers with high directivity. In particular, we report the results of the synthesis by imposing a 8*th* order cardioid target function. This type of directivity will be used in the processing illustrated in next chapters.

The diagrams show that the system works in the band from 30-40Hz to 3-4kHz.

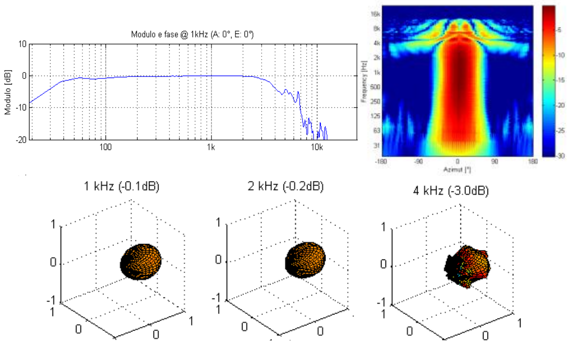

Figure 9: Frequency response, *XY* plane and 3D directivity of 8*th* order cardioid virtual speaker

5 “Real” Loudspeaker Array

As reported in [3], we did also build a “real” loudspeaker array: a wooden sphere with a diameter of 200mm equipped with 32 2” drivers (RCF MB2N101), as shown in fig. 10.

Figure 10 – the 32-speakers array

In this case there is no assurance that the loudspeakers are exactly identical, so it was necessary to perform a much longer measurement procedure inside the anechoic room, rotating the sphere in 362 combinations of azimuth and elevation, and measuring the 32 impulse responses of all the loudspeakers for every position of the source.

Nevertheless, such a large effort was paid by getting a set of beamforming filters capable of significantly exceeding the performances of the “virtual” loudspeaker array presented in the previous chapter, and also the performances of the new “real” loudspeaker array fed with signals which did not take into account individual difference between the transducers, as presented in [3].

First of all we present the results obtained when emulating a perfect omnidirectional source (same target function as shown in Figure 7 for the “virtual” loudspeaker array):

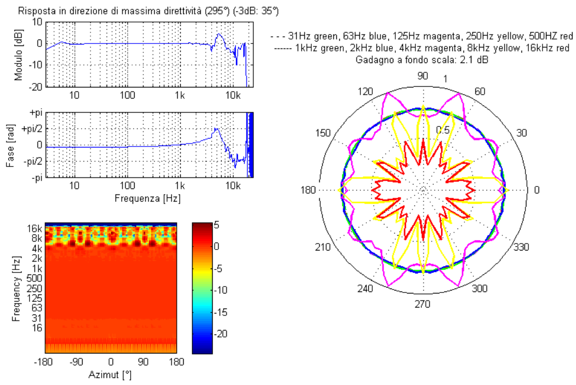

Figure 11: Frequency response, *XY* plane and polar patterns of an omnidirectional source synthesized by means of the “real” loudspeaker array

Then we attempted to create the same 8th-order cardioid we synthesized with the “virtual” loudspeaker array, as shown in figure 9:

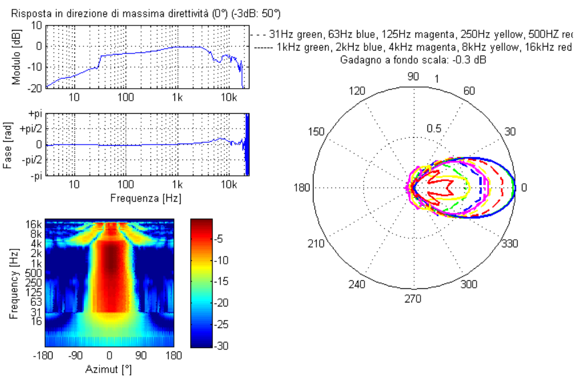

Figure 12: Frequency response, *XY* plane and polar patterns of an 8th-order cardioid synthesized by means of the real loudspeaker array

Of course, an 8th-order cardioid is really too much for a 32-drivers spherical array (its maximum theoretical order lies between 4th and 5th order). Hence we attempted to synthesize a 4th-order cardioid, which is already a very directive sound source, and the results are shown here:

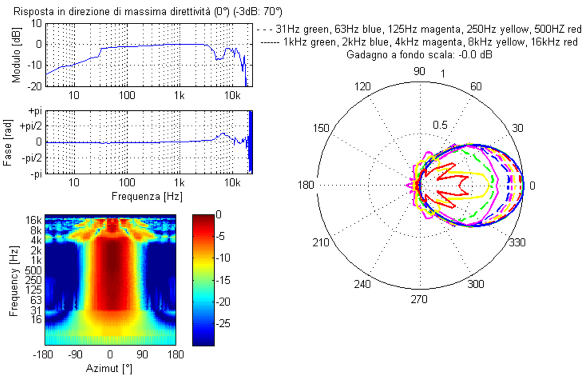

Figure 13: Frequency response, *XY* plane and polar patterns of an 4th-order cardioid synthesized by means of the real loudspeaker array

As shown in previous figures the performances of the new loudspeaker array are very good up to approximately 4 kHz. Above this frequency, the spatial aliasing effects disrupt the capability of getting a correct spatial control of the radiated sound field.
This is shown very well in the following figure, where the obtained beam width is charted for various cardioid orders (1 to 8) vs. frequency:

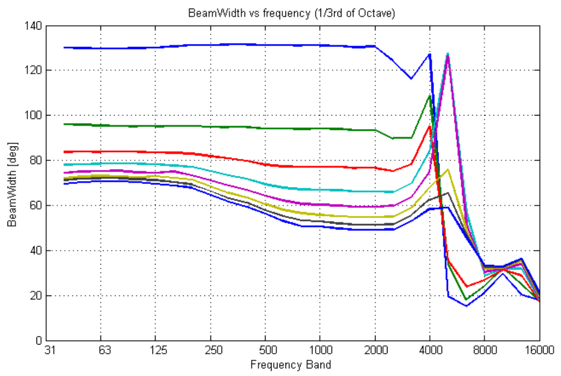

Figure 14. Beam width vs. frequency. "Real" loudspeaker array"

The availability of a "real" spherical loudspeaker array makes it possible to employ it also for realtime emulation of a source with arbitrary directivity pattern (human voice, musical instruments, etc.) as it will be shown during the demo session of this conference.

6 Measurement and Processing of MIMO Impulse Responses

Traditionally the measurement of impulse responses is taken within environments such as theatres, auditoriums, cinemas, in order to calculate the objective acoustic parameters (mainly the reverberation time). Normally the source (a dodecahedron or a firecracker) is placed in one or more points on the stage, where you will find the actors or musicians, and an omnidirectional microphone is placed in correspondence of several seats in the audience area. For each pair of positions a single channel impulse response is recorded.
The same can be done by employing the proposed loudspeaker array and a spherical 32 channels microphones array such as the Eigenmike™. For each pair of source-receiver positions the measuring procedure is performed, radiating sound in each of the 32 directions previously defined (either by rotating the single-driver source to that direction, or feeding just one of the 32 drivers of the "real" loudspeaker array).
For each radiation direction of the source, a multichannel impulse response is captured by means of the ESS (Exponential Sine Sweep) method determining each row of the matrix named IR_{MIMO} with size SxM (S = 32 speakers, M = 32 microphones).

$$\left\| \overset{S\times M}{IR_{MIMO}}[n] \right\| = \begin{bmatrix} ir_{1,1} & ir_{1,2} & \dots & ir_{1,m} & \dots & ir_{1,M} \\ ir_{2,1} & ir_{2,2} & \dots & ir_{2,m} & \dots & ir_{2,M} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ ir_{s,1} & ir_{s,2} & \dots & ir_{s,m} & \dots & ir_{s,M} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ ir_{S,1} & ir_{S,2} & \dots & ir_{S,m} & \dots & ir_{S,M} \end{bmatrix} \qquad (10)$$

Appropriate source beamforming filters and receiver beamforming filters can now be applied to the measured matrix, computing a new matrix with new specific directivity properties of both the source and the receiver.

$$\left\| \overset{W\times V}{IR_{MIMO-BF}} \right\| = \left\| \overset{W\times S}{h_{Source}} \right\| * \left\| \overset{S\times M}{IR_{MIMO}} \right\| * \left\| \overset{M\times V}{h_{\text{Receiver}}} \right\| \qquad (11)$$

In particular, it is useful to synthesize and use the following "symmetric" beamforming matrices, for either the source and the receiver:

- Omnidirectional Beamformer (single channel)
- First Order Ambisonics Beamformer, composed by 4 polar patterns representing spherical harmonics of orders 0, and 1 both for the source and for the receiver.
- HOA (High Order Ambisonics) Beamformer, composed by 16 polar patterns representing spherical harmonics of orders 0, 1, 2 and 3 both for the source and for the receiver.

- SPS (Spatial PCM Sampling) Beamformer composed of 32 patterns with 4th order cardioid directivity directed to 32 directions, uniformly covering the whole solid angle [5] both for the source and for the receiver.
- SPS (Spatial PCM Sampling) Beamformer composed of 122 patterns with 8th order cardioid directivity directed to 122 directions, uniformly covering the whole solid angle both for the source and for the receiver.

With these filter matrixes we can then calculate:

$$IR_{OMNI} = \overset{1\times S}{\left\| h_{Source-OMNI} \right\|} * \overset{S\times M}{\left\| IR_{MIMO} \right\|} * \overset{M\times 1}{\left\| h_{Receiver-OMNI} \right\|} \quad (12)$$

which is the classical single channel impulse response used for the calculation of the acoustic parameters according to ISO 3382.

Nothing new here. But we can get more information going to an Ambisonics source and receiver:

$$\overset{4\times 4}{\left\| IR_{MIMO-AMBI} \right\|} = \overset{4\times S}{\left\| h_{Source} \right\|} * \overset{S\times M}{\left\| IR_{MIMO} \right\|} * \overset{M\times 4}{\left\| h_{Receiver} \right\|} \quad (13)$$

which represents the spatial transfer function between source and receiver expressed with first order Ambisonics, resulting in a very compact MIMO matrix (just 4x4), which can be easily employed for realtime auralization even on a modest computer, allowing for both the source and the receiver to be easily rotated thanks to the Ambisonics representation.

This is ideal, for example, for VR applications, where the receiver rotation is made by means of an head-tracking system embedded inside a smartphone (as it happens when viewing VR panoramic videos with 1st-order Ambisonics panoramic audio over Youtube, employing a Google Cardboard device.

Of course, more detailed spatial information can be obtained going to higher orders:

$$\overset{16\times 16}{\left\| IR_{MIMO-HOA} \right\|} = \overset{16\times S}{\left\| h_{Source} \right\|} * \overset{S\times M}{\left\| IR_{MIMO} \right\|} * \overset{M\times 16}{\left\| h_{Receiver} \right\|} \quad (14)$$

which represents the spatial transfer function between source and receiver expressed with a high order spherical harmonics MIMO matrix, and hence very well suited for offline auralization to be employed in advanced HOA listening rooms, equipped with 16-24 loudspeakers or more.

For ultimate auralization fidelity we can abandon Ambisonics and go to 32-channels SPS:

$$\overset{32\times 32}{\left\| IR_{MIMO-SPS} \right\|} = \overset{32\times S}{\left\| h_{Source} \right\|} * \overset{S\times M}{\left\| IR_{MIMO} \right\|} * \overset{M\times 32}{\left\| h_{Receiver} \right\|} \quad (15)$$

This is ideal, for example, when the auralization process is fed with a multichannel recording obtained with a number of microphones surrounding the source inside the anechoic room, and when at the playback side the loudspeakers are also arranged almost regularly all around the listening area, so no "decoding" is required and each SPS signal can be fed directly to one loudspeaker.

Finally, we can bring the SPS method to its extreme by synthesizing 122 very directive patterns both for the source and for the receiver:

$$\overset{122\times 122}{\left\| IR_{MIMO-HSPS} \right\|} = \overset{122\times S}{\left\| h_{Source} \right\|} * \overset{S\times M}{\left\| IR_{MIMO} \right\|} * \overset{M\times 122}{\left\| h_{Receiver} \right\|} \quad (16)$$

Which is a convenient representation of the spatial information encoded as a square matrix of impulse responses measured between 122 highly directive virtual sources and 122 highly directive virtual receivers. This high-resolution representation is too massive for realtime auralization, but becomes very useful for analysing the trajectories of sound waves travelling form source to receiver, as it will shown in subchapter 6.2.

6.1 MIMO auralization

As already pointed out, a MIMO impulse response can be employed for auralization in a variety of different scenarios. Here we explain just two of them, but many others are possible.

The first scenario is very basic: the source is a mono anechoic recording, converted to 4-channels B-format by means of a filter matrix emulating the invariant directivity of the source (for example, of a trumpet). The resulting B-format signal goes through a first-order processor (Ambix, by Mathias Kronlachner), allowing for manually rotating the source along the three axes.

The resulting B-format 4-channels signal is convolved with the 4x4 IR matrix of the room to be auralized obtained by equation 13.

Finally, the resulting B-format signal, representing the signal which a Soundfield microphone is capturing at the receiver point inside the room, is passed through a second B-format processor, allowing for the rotation of the sound field, driven by the head-tracker worn by the listener together with a pair of headphones. The signals for the headphones are generated by a B-format-to-binaural converter (3D_Binauralizer by Aristotel Digenis).

The following figure shows this complete auralization system, running in realtime on a laptop inside Plogue Bidule:

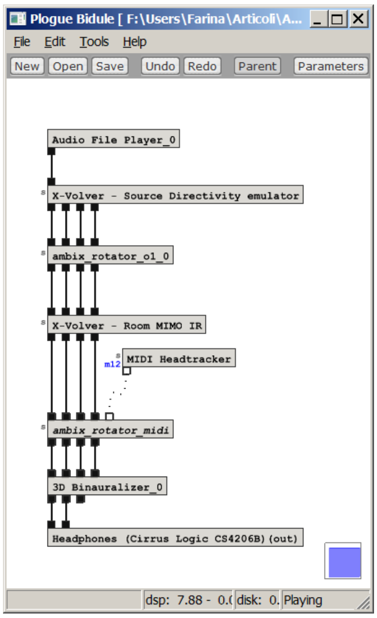

Figure 15: Realtime binaural auralization with headtracking – 1st-order Ambisonics

It must be noted that the CPU load was just 7.88%, despite the fact that the MIMO IR matrix being convolved contains 4x4=16 impulse responses with a length of 54142 samples each, as shown in fig. 16. This auralization method can be used even in the case of a large church, with impulse responses more than 300000 samples long.
With current hardware and software, we are able to run Ambisonics auralizations up to 3rd order both for input and for output (that is, with a 16x16 MIMO matrix).

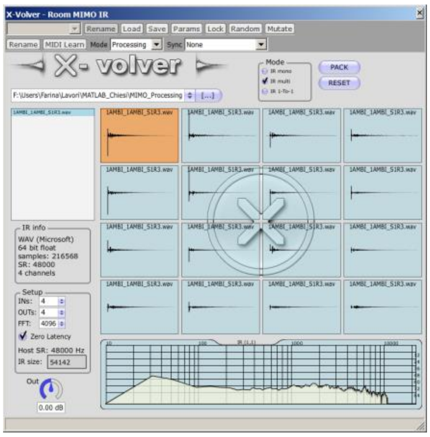

Figure 16: X-volver processing a 4x4 MIMO IR matrix

The second example shown here is employing a large 32x32 SPS MIMO matrix, as defined by equation 15, for feeding the 16-loudspeakers playback system of our 3D listening room. In this case the source is an anechoic recording made employing 8 microphones surrounding a violin inside a small anechoic room, which feed 8 coincident directive virtual sources (slected among the 32 available) pointing in the same directions. Of course, even better results coukld have been be obtained if a complete 32-channels anechoic recording had been made, inside a larger anechoic room, and if our listening room was equipped of a complete set of 32 loudspeakers.

Figure 17: Violin player inside a small anechoic room, surrounded by 8 microphones

Figure 18: Listening room with 16 loudspeakers

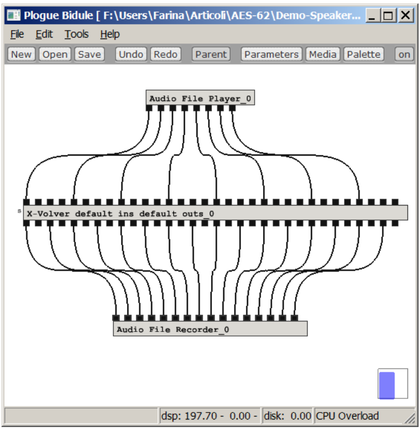

Figure 19: Auralization with an SPS MIMO matrix (32x32)

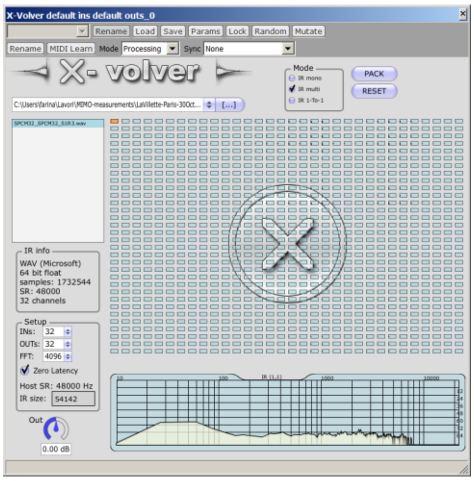

Figure 20: X-volver processing a 32x32 MIMO IR matrix

It can be seen that now the CPU usage is close to 200%, hence the processing cannot be done in realtime. X-volver in this case is processing a matrix of 32x32 IRs (that is, 1024 IRs, as shown in fig. 20), each having a length of 54142 samples. However the auralized soundtrack can be pre-processed and saved to file, and then played over the loudspeaker system. Indeed, a 200% CPU usage means that the processing will be feasible in realtime as soon as the X-volver 2 will be released, employing all the 8 cores of the CPU, instead of just one core as the current X-volver module is doing.

6.2 Reconstruction of trajectories

The MIMO impulse response contains a huge quantity of information about propagation of sound in the environment. The real problem is to display it.

An algorithm that allows to decode and display the propagation of sound represented by rays in a 3D model of the room was developed.

The first step is to identify all the relevant peaks in the omnidirectional impulse response. Each prominent peak corresponds to a sound ray of which it is possible to estimate the path length from the flight time.

For each detected peak, a surrounding window of a few samples (typically 5, when working at 48 kHz) coming from the MIMO-SPS 122x122 impulse response matrix is extracted. The RMS energy of each of these signal chunks is calculated, obtaining a matrix of spatial energy distribution with size 122x122. The indexes of the maximum in this matrix identify the SPS source and receiver directions where the energy radiated form the source is better transferred to the receiver. This pair of directions provide a reasonable approximation of the true directions of emission and reception of the sound ray (better precision can be achieved by taking in consideration more than one cell in the 122x122 energy matrix and applying some form of weighted averaging for computing the directions of emission and reception of the sound ray).

Then, knowing the positions of source and receiver in the room and, for each sound ray:

- the acoustical path distance,
- the direction of emission from source
- the direction of reception at receiver

it becomes possible to determine the trajectory of the direct sound and of single reflection paths, identifying also the impact point (it is the point where the ray going out of the source intersects the ray going into the receiver).

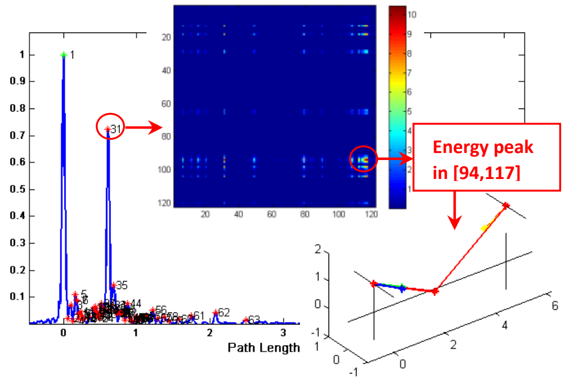

Figure 21: Peak detection, its spatial energy distribution matrix, single reflection model

Knowing the direction of emission, the direction of arrival and the length of the path, a LS method can identify quite accurately the reflection point, and hence the whole trajectory, as shown here:

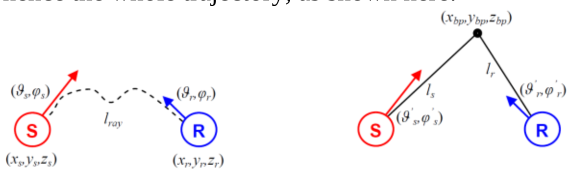

Figure 22: Path identification for a single reflection

If a 3D model of the room is available, it can be used to study the propagation of the sound rays that don’t meet the single reflection model.

Here we see the method used for identifying double reflections:

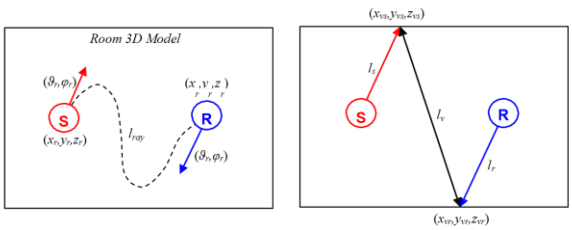

Figure 23: Path identification for a double reflection

In practice, the source ray and the receiver ray are extended until they hit the geometry boundaries, then the reflection points are connected together, if this meets the path length. If it does not meet, then a triple reflection is supposed, as shown here:

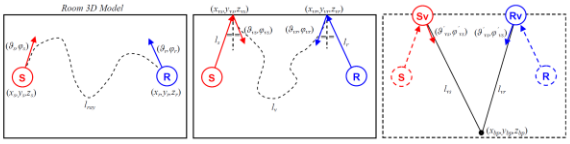

Figure 24: Path identification for a triple reflection

In this case, after extending the source and receiver rays towards room boundaries, a specular reflection is assumed at each impact point: the specular rays are extended again, and if they meet close a room boundary, with the proper path length, then the triple reflection path has been successfully identified.
The result of the trajectory analysis can finally be displayed over the 3D model, colouring each different sound path according to the energy it carries, to make clear which are the most acoustically active surfaces.

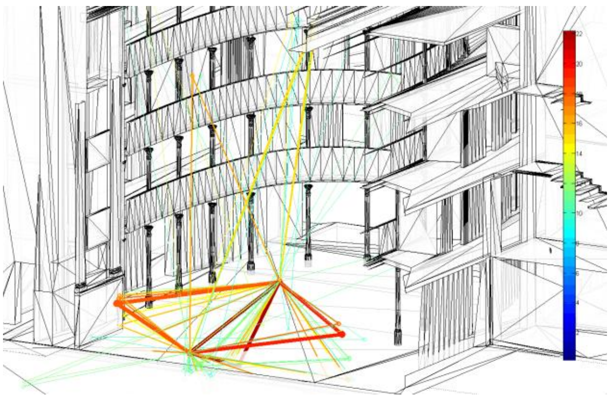

Figure 25: Trajectory analysis up to triple reflections based on MIMO IR measurement

7 Conclusions

This paper has described the usage of spherical transducer arrays, presenting the detail of the method developed for generating suitable filter matrixes, operating the beamforming of both the sound source and the receiver.
Two different spherical sources have been presented, one working as a “virtual” loudspeaker array, the second is a real loudspeaker array composed of 32 small drivers. With these sources, and employing an Eigenmike™ as receiver, a dozen of different theatres have already been measured in France and Italy.
The results of these measurements will be presented in future publications.
The resulting MIMO IRs have shown to be usable for advanced auralization and for analysing the spatial transfer function between source and receiver positions.
Proper software tools have been developed, allowing for massive realtime multichannel convolution of these large IR matrixes and for tracing the trajectories of single rays carrying sound waves from the source to the receiver, including up to three reflections.
These capabilities will be the basis of further research work in the fields of room acoustics, auralization and virtual reality.

Thanksgiving

We want to express here our gratitude to Alberto Amendola, who sadly passed away recently: his brilliant ideas pervade this work, and his fond memory will continue to inspire us.

References

[1] A. Farina, P. Martignon, A. Capra, S. Fontana, “Measuring impulse responses containing complete spatial information”, *22nd AES-UK Conference 2007*, Cambridge, UK, 11-12 April 2007.

[2] Angelo Farina, Andrea Capra, Lorenzo Chiesi, Leonardo Scopece, “A Spherical Microphone Array For Synthesizing Virtual Directive Microphones In Live Broadcasting And In Post Production”, *40th AES Conference*, Tokyo, Japan, 8-10 October 2010.

[3] Angelo Farina, Lorenzo Chiesi, “A novel 32-speakers spherical source”, *E-brief presented at the 140th AES Conference*, Paris, France, 4-7 June 2016.

[4] Bertet, Stéphanie; Daniel, Jérôme; Moreau, Sébastien, “3D Sound Field Recording with Higher Order Ambisonics - Objective Measurements and Validation of Spherical Microphone”, *120th AES Convention*, Paris, France, 20-23 May 2006.

[5] Angelo Farina, Alberto Amendola, Lorenzo Chiesi, Andrea Capra, Simone Campanini, “Spatial Pcm Sampling: A New Method For Sound Recording And Playback”, *AES 52nd International Conference*, Guildford, UK, 2–4 September 2013.