"SIPARIO SOUNDS": A NEW OMNIDIRECTIONAL LOUDSPEAKER FOR MIMO AURALISATION

Angelo Farina[1] Lamberto Tronchin[2] Antonella Bevilacqua[3]
Enrico Armelloni[3] Francesca Merli[4] Marco Dolci[5]

[1] Department of Engineering and Architecture, University of Parma, Italy
[2] Department of Architecture, University of Bologna, Italy
[3] CIDEA, University of Parma, Italy
[4] CIRI – Edilizia e Costruzioni, University of Bologna, Italy
[5] Gruppo CSA SpA, Via al Torrente 22, Rimini, Italy
lamberto.tronchin@unibo.it

ABSTRACT

SIPARIO is a project recently funded by the Italian region Emilia Romagna. It aims to virtually reconstruct both 3D audio and 360° video of real performances by making recordings and undertaking acoustic measurements inside historical theatres and concert halls spread all over the Europe. One of the most relevant topics of this project is the development of a new spherical sound source, equipped with 32 individually-controlled loudspeakers [1], designed for measuring acoustic characteristics of these performing arts places. This loudspeaker array is capable of simulating the directivity of any dynamic sound source. The loudspeaker can work also like a traditional omnidirectional sound source for measuring standard acoustical parameters, hence it can be used in room acoustics tests in accordance with ISO3382 standard. Finally, it can create super-directive beams of various shapes, with arbitrary directivity patterns such as spherical harmonics (High Order Ambisonics) or cardioids (Spatial PCM Sampling). The Exponential Sine Sweep (ESS) signal has been employed in order to measure the room impulse response (RIR). It is also possible to employ a Multiple Input Multiple Output (MIMO) approach [2] by feeding sequentially the ESS signal to each transducer of the loudspeaker array, while a spherical microphone array is employed for recording. In this paper it is explained how to compute the matrix of FIR filters necessary to process such a MIMO RIR, and to synthesize arbitrary polar patterns for both source and receiver. Moreover, future developments of the new sound source are briefly described in Section 7, including the applications for the specific project.

1. INTRODUCTION

In room acoustics very often the main purpose of the experimental analysis is the measurements of a RIR. This kind of measurement needs usually a sound source which is perfectly omnidirectional. However, when the aim of the acoustic measurements is to analyze the RIR of a sound source having a complex directivity pattern (e.g. human voice or specific musical instruments), the standard equipment (omnidirectional loudspeaker such as a traditional dodecahedron) is insufficient. Moreover, in relation to performing arts places, where the radiation polar patterns of singers and musicians changes continuously over time, recording a mono anechoic signal with a single microphone is absolutely not suitable for the purpose of auralization, and a sphere of microphones surrounding the source is necessary for recording the sound radiated in every direction and to be used for auralization [3].

The MIMO RIR provides the filter matrix necessary for virtually reconstructing the sound field perceived at listener's position, being generated by a sound source with arbitrary and time-changing directivity [4].

Thus, a virtual spherical loudspeaker array has been developed as a sound source for the MIMO room impulse response measurements. The main subject of this paper is the characterization of this spherical speaker.

2. SIPARIO PROJECT

The SIPARIO project aims to reconstruct a virtual reality playback of artistic performances executed in historical Italian theatres. Considering halls and auditoria as an extension of musical instruments or singer's voice, SIPARIO will give the possibility to listen to the effect of these venues, particularly for places where the access is limited for preserving their integrity, not excluding non-linearities as in musical instruments [5,6].

One of the objectives of the project is to recreate a panoramic soundscape and an immersive video experience where the listener can select his position, look at and listen around (e.g. standing on stage, sitting in one of the balconies or from privileged positions). Virtual listening experiences, capable to respond to the listener's movement, would be a different way to attract a wide audience, not only limited to experts, musicians or acousticians.

The capability of enjoying musical performances through virtual reality systems assumed recently a great importance because of the COVID-19 pandemic, which caused the impossibility to assist personally to concerts and other public events.

Measuring MIMO matrices of RIRs requires the usage of spherical arrays for both microphone and speaker. Despite the high resolution microphone arrays have already been available since 10 years approximately, the utilization of spherical loudspeaker arrays is still experimental. This paper describes the virtual loudspeaker array called "Sipario Sounds" and its acoustical performances.

3. MICROPHONE ARRAYS

Starting from an array of M "real microphones", we want to synthesize V "virtual microphones" with specific directivity and aiming. The following figure and equation show a generic signal processor, which uses a matrix of linear and time-invariant FIR filters for doing such a processing, usually called "filter and sum beamforming".

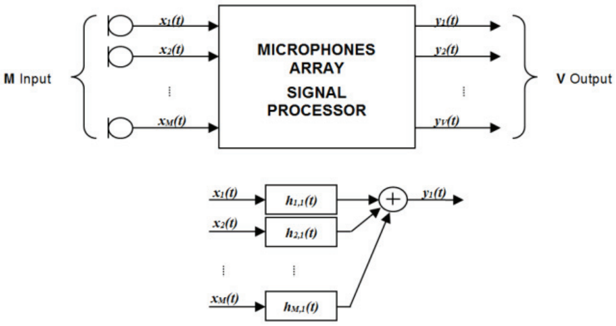

Figure 1: Representation of the microphone array signal processor

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

Where * is the convolution operator.
A technique to find the filtering coefficients, $h_{m,v}$, was described extensively in [7]. This technique is applicable to any microphone array by using the numeric inversion of the characterization impulse responses of the microphones measured in an anechoic chamber or simulated by a variety of methods (closed form mathematics, FEM, etc.).

In particular, the impulse response of the M-channel microphone array has been captured using incident wavefronts from a large number of characterization directions D (362 directions with nearly uniform distribution, that constitutes an adequate spatial oversampling for a 32-channel microphone array like the EM32 Eigenmike®). Each measurement direction determines a row of a matrix $c[n]$ of size $D\times M$.

Figure 2: EM32 Eigenmike® (by MH-Acoustics)

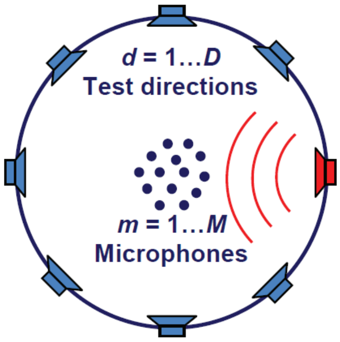

Figure 3: Microphone array characterization

$$c[n] = \begin{bmatrix} c_{1,1} & c_{1,1} & \cdots & c_{1,m} & \cdots & c_{1,M} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,m} & \cdots & c_{2,M} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{d,1} & c_{d,2} & \cdots & c_{d,m} & \cdots & c_{d,M} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{D,1} & c_{D,2} & \cdots & c_{D,m} & \cdots & c_{D,M} \end{bmatrix} \qquad (2)$$

Therefore, it is necessary to define the target directivity Q having dimensions $D\times V$ which synthetizes, for each of the V virtual microphones, the gain in each direction D employed for the array characterization.
These gains could be arbitrarily defined; however, in [5] virtual microphones with cardioid of n^{th} order were used:

$$Q_n(\vartheta,\varphi) = [0.5 + 0.5\cdot\cos(\vartheta)\cdot\cos(\varphi)]^n \qquad (3)$$

A set of 32 4th-order cardioids pointing in the same directions as the capsules of the EM32 Eigenmike® has been used for generating an SPS-32 "P-format" signal [8]. An alternative is to define the target directivities according to the spherical harmonics functions (Ambisonics format). In this case a set of third-order Ambix directivities was chosen, with ACN channel ordering and SN3D amplitude scaling [9], as shown in Figure 4.

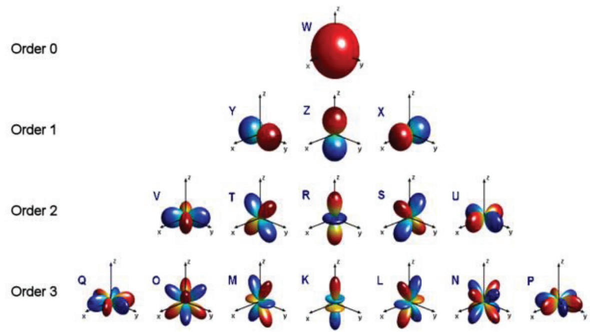

Figure 4: Directivity of spherical harmonics up to 3rd order

The set of filters h can be derived directly from a set of measurements; for this, the C matrix has to be numerically inverted by using a technique called Tikhonov

regularization, first introduced for multichannel audio applications by Kirkeby et al. [10].
A Discrete Fourier Transform (DFT) is employed for passing into the frequency domain. In this way the overdetermined linear system can be evaluated separately for each frequency index k.

$$\overset{D\times M}{\|c[n]\|} * \overset{M\times V}{\|h[n]\|} = \overset{D\times V}{\|q\|}\cdot\delta[n] \overset{DFT}{\Rightarrow} \overset{D\times M}{\|C[k]\|}\cdot\overset{M\times V}{\|H[k]\|} = \overset{D\times V}{\|Q\|}\cdot e^{-j\pi k} \quad (4)$$

The solution could be obtained as reported in the following equation, which is the solution using the Kirkeby method.

$$\overset{M\times V}{\|H[k]\|} = \frac{\overset{M\times D}{\|C[k]\|^{H}}\cdot\overset{D\times V}{\|Q\|}\cdot e^{-j\pi k}}{\overset{M\times D}{\|C[k]\|^{H}}\cdot\overset{D\times M}{\|C[k]\|} + \beta[k]\cdot\overset{M\times M}{\|I\|}} \quad (5)$$

Where H is the Hermitian operator (conjugate transpose).
A casualization term (a pure delay equal to half the filter length) $e^{-j\pi k}$ has been added to the numerator as well as a frequency-dependent regularization term $\beta[k]$ to the denominator.
Finally, the filter matrix H determined in frequency domain is transformed into time domain in order to obtain the matrix of FIR filters coefficients, which converts the M raw microphone signals into the V virtual microphones.

$$\overset{M\times V}{\|h_{MA}[n]\|} = IDFT\left[\overset{M\times V}{\|H[k]\|}\right] \quad (6)$$

4. LOUDSPEAKER ARRAYS

A prototype spherical array loudspeaker has been built by Lorenzo Chiesi at the University of Parma, Italy [1]. It consists of a wooden sphere having a 200mm diameter, equipped with 32 2" drivers (RCF MB2N101), as shown in Figure 5.

Figure 5: Prototype spherical array loudspeaker [1]

Albeit this unit is equipped with a 32-channels power amplifier, when performing MIMO measurements the test signal needs to be emitted by just one of the 32 loudspeakers, and then the measurement has to be repeated feeding another loudspeaker, 32 times.
As such, this loudspeaker array can be practical when an arbitrary directivity must be synthesized in live applications. To measure a MIMO RIR it is also possible to use a sphere equipped with just one, larger and better loudspeaker, shown in Figure 6. In order to create a 32-channels virtual loudspeaker array, it should be rotated in 32 uniformly spaced directions using a two-axes turntable, as shown in Figure 7.

Figure 6: Virtual loudspeaker array "Sipario Sounds"

Figure 7: The single sphere loudspeaker on the two-axes turntable, inside Teatro Regio of Parma

As in the previous discussion related to the microphone array, it is possible to create radiation patterns of arbitrary shape by filtering the signals feeding the S individual loudspeakers (or loudspeaker positions). Again, a matrix of FIR filters generates W virtual sources having arbitrary directivity and aiming.

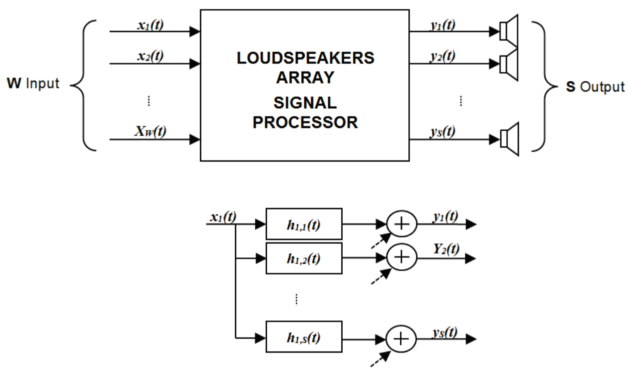

Figure 8: Representation of the loudspeaker array signal processor

Formula (1) has been modified into formula (7) below.

$$y_s(t) = h_{w,s}(t) * x_w(t) \quad s = 1..S \quad (7)$$

The loudspeaker array is characterized by means of anechoic impulse response measurements, rotating the system on a two axis turntable and measuring the impulse response $c_{d,s}$ of each loudspeaker s in a large number D of directions. The result, for each loudspeaker, would be the determination of one column of a matrix $C[n]$ having dimensions $D \times S$.
For the prototype speaker employed here, S = 32 loudspeakers and D = 362 directions, both almost uniformly spread on the surface of a sphere.

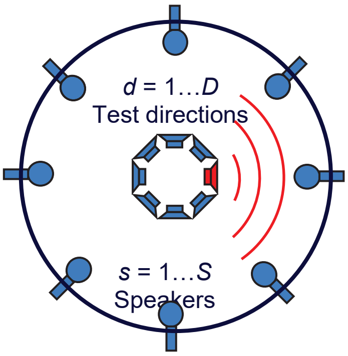

Figure 9: Speakers array characterization

$$c[n] = \begin{bmatrix} c_{1,1} & c_{1,1} & \cdots & \boxed{c_{1,s}} & \cdots & c_{1,S} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,s} & \cdots & c_{2,S} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{d,1} & c_{d,2} & \cdots & c_{d,s} & \cdots & c_{d,S} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{D,1} & c_{D,2} & \cdots & c_{D,s} & \cdots & c_{D,S} \end{bmatrix} \quad (8)$$

For every W virtual source directivity the following step is to determine the matrix giving the target directivity Q for each direction. As in the case of virtual microphones, it is therefore possible to convert to a frequency domain, in order to simplify the convolutions.

$$\overset{D\times S}{\|c[n]\|} * \overset{S\times W}{\|h[n]\|} = \overset{D\times W}{\|q\|} \cdot \delta[n] \overset{DFT}{\Rightarrow} \overset{D\times S}{\|C[k]\|} \cdot \overset{S\times W}{\|H[k]\|} = \overset{D\times W}{\|Q\|} \cdot e^{-j\pi k} \quad (9)$$

The over-determined linear equation system (9) is solved with the Kirkeby approach [10]:

$$\overset{S\times W}{\|H[k]\|} = \frac{\overset{S\times D}{\|C[k]\|^{H}} \cdot \overset{D\times W}{\|Q\|} \cdot e^{-j\pi k}}{\overset{S\times D}{\|C[k]\|^{H}} \cdot \overset{D\times S}{\|C[k]\|} + \beta[k] \cdot \overset{S\times S}{\|I\|}} \quad (10)$$

The resulting FIR filter matrix H is transposed in order to preserve the conventional operation direction (from input space of W directive sources to the output space of S speaker feeds), and converted back to time domain. The following equation summarizes this result:

$$\overset{W\times S}{\|h_{SA}[n]\|} = IDFT\left[\overset{S\times W}{\|H[k]\|}\right]^{T} \quad (11)$$

5. PERFORMANCES OF THE VIRTUAL LOUDSPEAKER ARRAY

This Section deals with the acoustical performances and beamforming capabilities of the virtual loudspeaker array, as shown in Figures 6 and 7. The prototype array, as shown in Figure 5, was already extensively described in [1].
The measurements were performed inside the anechoic chamber kindly made available by IRCAM, in Paris, with the setup shown in Figure 10.

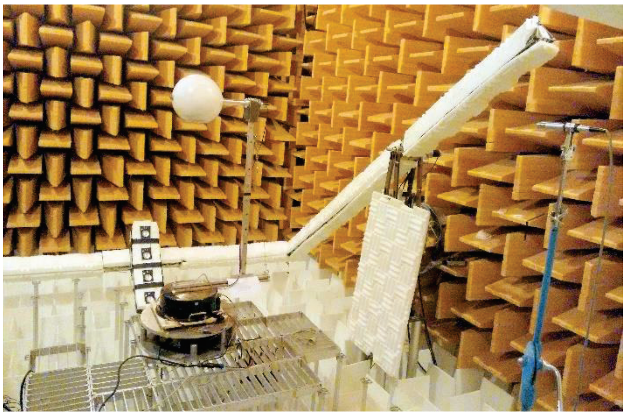

Figure 10: Anechoic measurements at IRCAM, Paris

First of all, the on-axis frequency response and directivity of the single loudspeaker mounted on the spherical shell has been measured in the horizontal plane, as shown in Figures 11 and 12.

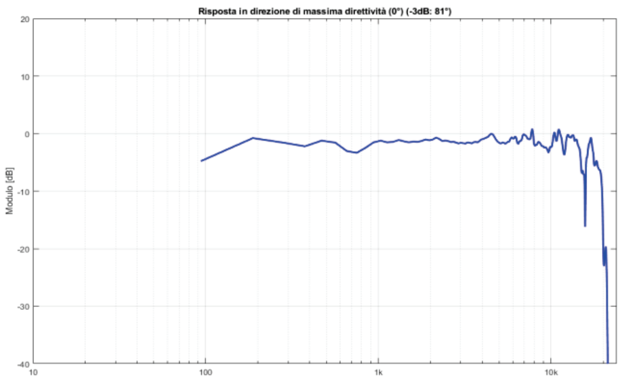

Figure 11: on-axis frequency response at 1W, 1m, single loudspeaker

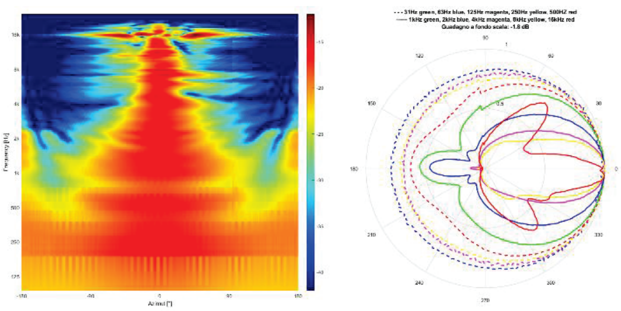

Figure 12: Directivity of the single loudspeaker

Then a complete measurement set has been completed on the whole sphere with 362 directions, resulting in the measured matrix of impulse responses C (eq. 8). The

beamforming matrices for SPS-32 and 3OA (16 ch.) have been calculated by using eq. 10 and 11.
These beamforming filter matrices have been applied to the horizontal data set, as indicated in Figure 12, to show the results.
Figures 13 and 14 show the resulting directivity patterns for a 4th order cardioid (SPS-32) and for the 2nd-order 4-leafed pattern U (see Figure 4).

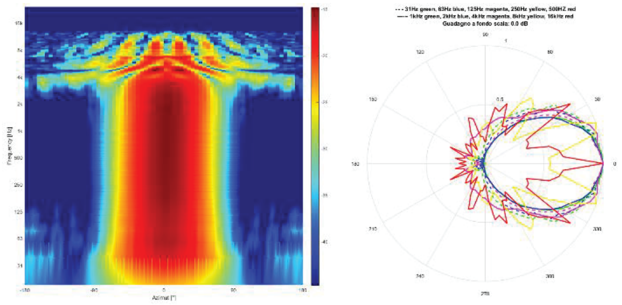

Figure 13: Directivity of a 4th order cardioid source

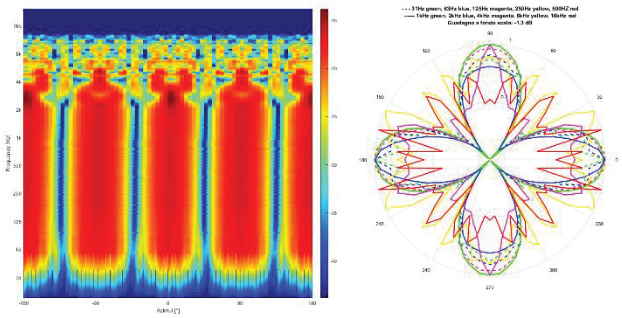

Figure 14: Directivity of HOA component "U"

Finally, Figure 15 shows the on-axis frequency response of the virtual source having a 4th-order cardioid directivity.

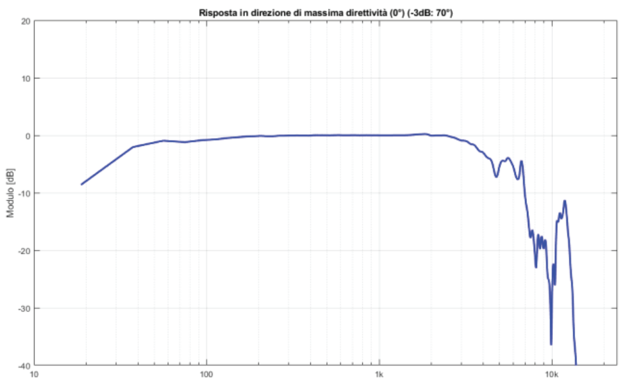

Figure 15: on-axis frequency response of a virtual source with 4th order cardioid directivity

Analyzing these results, it can be seen how the resulting virtual loudspeaker array can control the directivity of the radiated sound in the frequency range of 63 Hz to 4 kHz (slightly wider for the cardioid, and narrower for spherical harmonics patterns). Also the resulting frequency response is reasonably flat within such a frequency range.
This frequency range is considered good enough to perform acoustical measurements and to analyze performing arts places by using objective acoustical parameters and by visualizing reflections and echoes generated by the sound source.
This frequency range, instead, is not large enough for high quality auralization, which requires a full bandwidth between 20 Hz and 20 kHz.

6. MEASUREMENTS OF IMPULSE RESPONSES AND PROCESSING

MIMO measurements can be performed by employing the proposed spherical speaker array and a spherical 32-channel microphone array (i.e. EM32 Eigenmike®) and by feeding the speaker with a suitable test signal, an ESS type signal for instance [11, 12].
An explanation of how the signals are processed is briefly introduced.
The computer sends the ESS signal to the amplifier connected with the loudspeaker through one of the analog outputs of the EMIB Firewire interface.
The EM32 Eigenmike® acquires the 32 microphone signals synchronously with the playback.
The measurement is repeated 32 times by moving the loudspeaker to the next aiming direction with the automated turntable.
After the convolution of the recorded signals with the inverse sweep and a proper time windowing, a full MIMO matrix is obtained having size $S\times M$, where $S = 32$ speakers and $M = 32$ microphones.

$$\overset{S\times M}{\|IR_{MIMO}[n]\|} = \begin{bmatrix} ir_{1,1} & ir_{1,2} & \dots & ir_{1,m} & \dots & ir_{1,M} \\ ir_{2,1} & ir_{2,2} & \dots & ir_{2,m} & \dots & ir_{2,M} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ ir_{s,1} & ir_{s,2} & \dots & ir_{s,m} & \dots & ir_{s,M} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ ir_{S,1} & ir_{S,2} & \dots & ir_{S,m} & \dots & ir_{S,M} \end{bmatrix} \quad (12)$$

This is called the "raw" MIMO IR matrix.
Appropriate beamforming filters for both source and receiver can now be applied to this raw matrix. In this way a new beam-formed MIMO IR matrix is obtained, able to provide a set of specific directivities for both the sound source and the receiver:

$$\overset{W\times V}{\|IR_{MIMO-BF}[n]\|} = \overset{W\times S}{\|h_{SA}[n]\|} * \overset{S\times M}{\|IR_{MIMO}[n]\|} * \overset{M\times V}{\|h_{MA}[n]\|} \quad (11)$$

In particular, this procedure can be used for generating different types of directivity patterns, for both the source and the receiver. Two typical cases are given below:

- HOA 3rd order for both source and receiver, resulting in a 16x16 Ambisonics IR matrix;
- SPS (Spatial PCM Sampling), synthesizing 32 directive beams at the source and 32 directive virtual microphones at the receiver, in both cases with 4th order cardioid directivity, uniformly covering a whole spherical surface area.

The HOA matrix is ideal for real-time auralization, as initially suggested in [4].

The SPS matrix, instead, is perfect for mapping room reflections and analyzing multiple reflection paths, as described in [2].
The exploitation of these possibilities, applied to the theaters measured within the SIPARIO project, will be the subject of following papers.
In accordance with the ISO 3382:2012 standard [13], impulse responses measurements in musical performance rooms (i.e. theatres, auditoria, concert halls, TV studios) should be undertaken by employing an omnidirectional sound source and an omnidirectional microphone. This can be easily obtained by selecting just the first row and first column of the Ambisonics MIMO matrix. In this way it is possible to have an omnidirectional directivity pattern for both the source and the microphone, a good S/N ratio, and a better omnidirectional pattern than what was previously possible by using a dodecahedron loudspeaker and a standard measurement-grade Omni microphone.

7. CONCLUSIONS AND FURTHER RESEARCH

Alongside the aim of faithfully capturing the physical and acoustic characteristics of a room, one of the goals of SIPARIO project is to perform the auralization of these venues.
This can be performed on low-cost personal virtual reality systems (head-mounted displays plus headphones), as shown in Figure 16.

Figure 16: Representation of acoustic virtual reality. Photo by Shutterstock® Studio, 2018

The auralization has the primary goal of providing an immersive listening experience to people that have never attended such live performances, in such a way to be virtually present in a theatre using inexpensive devices, such as a smartphone, a set of headphones and a Google Cardboard visor.
Furthermore, the construction of a high quality listening room and the availability of HOA MIMO IR matrices will make it possible to perform a real-time auralization of live music by giving the performer the same acoustical feedback as when playing or singing in a real theatre.
The previously described system makes it possible to measure these MIMO IR matrices, which can be used to calculate standard acoustical parameters and to perform advanced auralization.
However, the system could be improved by:

- extending the frequency response of the source;
- increasing the spatial resolution of the microphone array.

The results shown in Section 5 proof that the virtual loudspeaker array has a very good performance only inside a frequency range between 63 Hz and 4 kHz. This is good enough for computing acoustical parameters or for tracing reflections and echoes, but is not enough for advanced auralization. Above this upper limit, spatial aliasing effects disrupt the capability of getting a correct spatial control of the radiated sound.
The only way to extend the frequency range is by employing a larger number of smaller loudspeakers, although this possibility needs to be evaluated with a potential construction of a new loudspeaker array, alongside with what was already described in [1].

In Section 3 it has been explained how to compute the beamforming filter matrix for a microphone array. The one used till now for MIMO measurements is a spherical 32-channel microphone array produced by MH-acoustics, called EM32 Eigenmike®, available on the market since 10 years. The plan is to replace it with a new spherical microphone array made by the same producer, called EM64 Eigenmike®. As the name suggests, this is equipped of 64 capsules, mounted on a rigid sphere of the same diameter as the original EM32, as shown in Figure 17. This new system is equipped with a digital audio bus, DANTE™, able to manage all the 64 channels on a standard Ethernet cable connected directly to a computer without the need of any audio interface (such as the EMIB which was required by the EM32).

Figure 17: EM64 Eigenmike® (by MH-Acoustics)

Despite not being available yet at the time of writing, and hence not having been tested yet, the expected results of the new EM64 Eigenmike® should provide significantly improved spatial resolution, with HOA signals up to 5th order (36 channels).

8. ACKNOWLEDGMENTS

The Authors wish to thank Lorenzo Chiesi for his constant effort in supporting the realization of this study.

9. REFERENCES

[1] A. Farina, L. Chiesi: "A novel 32-speakers spherical source," *AES 140th Convention Program, e-Brief 258*, Paris, France, June 4-7 2016.

[2] A. Farina, L. Chiesi: "Measuring Spatial MIMO Impulse Reponses in Rooms Employing Spherical Transducer Arrays", *AES Conference on Sound Field Control,* Guilford, UK, July 18-20, 2016.

[3] D D'Orazio: "Anechoic recordings of Italian opera played by orchestra, choir, and soloists", The Journal of the Acoustical Society of America 147, EL157 2020; https://doi.org/10.1121/10.0000739

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

[5] L. Tronchin: "The emulation of nonlinear time-invariant audio systems with memory by means of Volterra series", J. Audio Eng. Soc., 60 (12), 984.996 2012

[6] L. Tronchin, V.L. Coli: "Further investigations in the emulation of nonlinear systems with Volterra series". J. Audio Eng. Soc. 63, 671–683 2015

[7] A. Farina, A. Capra, L. Chiesi, L. Scopece: "A Spherical Microphone Array for Synthesizing Virtual Directive Microphones in Live Broadcasting and in Post Production". *AES 40th International Conference*, Tokyo, Japan. October 8-10, 2010

[8] A. Farina, A. Amendola, L. Chiesi, A. Capra, S. Campanini: "Spatial PCM Sampling: A New Method For Sound Recording And Playback". *AES 52nd International Conference*, Guildford, UK. September 2-4, 2013

[9] C. Nachbar, F. Zotter, E. Deleflie, A. Sontacchi: "ambiX - A Suggested Ambisonics Format," in Proceedings of the 3rd Ambisonics Symposium, Lexington, KY June 2-3, 2011

[10] O. Kirkeby, P.A. Nelson, H. Hamada, F. Orduna-Bustamante: "Fast deconvolution of multichannel systems using regularization", IEEE Transactions on Speech and Audio Processing Volume 6, Issue: 2, Mar 1998.

[11] A. Farina: "Simultaneous measurement of impulse response and distortion with a swept-sine technique" - 108th AES Convention, Paris 18-22 February 2000

[12] A. Farina: "Advancements in impulse response measurements by sine sweeps" - 122th AES Convention, Vienna, Austria, 5-8 May 2007

[13] International Organization for Standardization. *ISO 3382-2:2008 Acoustic-Measurement of room acoustic parameters.* Geneva, Switzerland 2008.