Different Techniques for Measuring Spatial Sound Properties of Auditoria: a Review

A Bevilacqua¹, F Merli²

¹Department of Engineering & Architecture, University of Parma, Via delle Scienze, 43124 Parma Italy
²Department of Architecture, University of Bologna, Via dell’Università 50, 47521 Cesena, Italy

¹antonella.bevilacqua@unipr.it

Abstract. The interest in sound spatialization is strongly increased throughout the last decade. Sound spatialization, or as it is commonly named auralization, is very important for both the design phase of any listening room and the virtual reproduction of a 3D sound field [1]. The international standard ISO 3382:2009 explains the different techniques of how to measure spatial parameters (e.g. LE, LF, IACC) and one of the methods that can be performed for a binaural measurement is the utilization of a dummy head. Nowadays room impulse response (RIR) is often measured by using multichannel transducers working independently each other. This technique is very useful to virtually reproduce a 3D soundscape, corresponding to the same sound perception as it is found inside the architectural volume where the measurements have been performed [2]. In this paper, an alternative procedure of measuring and reconstructing the auralization characteristics is presented. Furthermore, this methodology has been compared with other techniques of 3D sound reproduction. This paper treats the opportunity to improve the quality of soundscape reproduction other than the analysis of acoustical parameters required by standards.

1. Introduction

When Gran Teatro La Fenice of Venice burned during the night of 29th January 1996, one of the most important opera houses in the world seemed suddenly to be disappeared. Its acoustical behaviour, however, was partially saved because several sets of measurements had been performed just two months before the fire, employing a binaural impulse response (IR) technique [3].

In the meanwhile, pushed by the sorrowful news, and following the original idea of Gerzon it was proposed to the acousticians to collect as many as possible IRs measurements inside any historical theatres and concert halls, in order to both assess their acoustical behavior and to preserve them to the posterity [4]. Several people proposed to collect a wide collection of IR measurements, and several new techniques were developed.

This paper is mainly focused on the investigation of these IR measurements utilized with the purpose of faithfully recreating historical performing arts architecture throughout a virtual reconstruction of a 3D sound field [5].

The discussed methodology has already been extensively described in [6], where all the standard acoustical measurement techniques have been incorporated and merged in a singular approach: three different types of microphones (i.e. a binaural dummy head, two cardioids in ORTF configuration and a Soundfield microphone) have been mounted on a rotating table, thus that a set of IRs can be measured at each angular position. Fig. 1 shows a scheme of such microphone setup.
The results of this set of measurements have been compared with the existing methods of measurements employed for concert halls (i.e. binaural, B-format and Wave Field Synthesis), but it is preferred for allowing the derivation of standard surround formats such as Optimum Cardioid Triangle (OCT) and INA. In addition, by using this approach it is possible to change the selection of methodology ranging from the Binaural Room Scanning method [7] to the Wave Field Synthesis (WFS) [8], passing through the Ambiophonics hybrid approach [8] and the Poletti high-order circular microphones [9].

This paper discusses briefly the results, and more in detail the process of how the IRs have been analysed to achieve the auralization, facilitating the immersive listening experience of a real soundfield.

2. Equipment and Methodology

2.1. Measurement of multichannel impulse responses

This section describes in detail the measurement techniques, the equipment (i.e. hardware and software) and the procedure employed for a selecting list of various performing arts spaces [10]. Although most of these items are already widely analysed in other papers and studies, a resumption of such concepts and a combination of them in a singular and coherent approach try to illustrate a general methodology from which all investigated multichannel formats can be derived [11].

2.1.1. Test signal and deconvolution

The excitation-deconvolution technique employed for the IR measurement is the exponential sine sweep (ESS), as initially suggested by Farina [12]. A good compromise between frequency range to be measured, length of the sweep and signal-to-noise ratio (SNR) has been reached by choosing the following parameters.

| Table 1. Setting of the sound signal |
|-------------------------------------|
| Start frequency | 22 Hz |
| End frequency   | 22 kHz |
| Length of sweep | 15 s   |
| Silence between consecutive sweeps | 10 s   |
| Sweep type      | Exponential |

The unusual length of the silence between consecutive sweeps is owed to the traveling time of the rotating table to move the microphone to different angulations and to be ready for the following...
measurement. The rotation is triggered by a specific pulsive signal, automatically generated during the silence gap and active on the second channel of the sound card.

The choice of setting established for the test signal was determined in order to get good quality of IR measurements, given a wide frequency span and a good dynamic range (approximately 90 dB) [13]. In addition, the test signal can be considered immune from any background noise eventually appearing during the measurement recordings.

2.1.2. Sound source
An omnidirectional sound source is usually preferred for RIR measurements. Although a common omnidirectional sound source employed for acoustics measurements does not correspond to the effective directivity patterns existing in the real world (e.g. musical instruments or human voice) [14], the utilization of it is indicated by current standards (i.e. ISO 3382:2012). The omnidirectionality is preferred because it avoids strange room effects, as they can appear with highly directive loudspeakers (e.g. abnormal bounce of energy creating echoes or focalization of reflections in undesirable directions).

A customized, ultra-compact dodecahedron loudspeaker was built specifically for the purpose of this research study. It has been equipped by 12 full-range drivers installed on a small size enclosure with a diameter of approximately 200 mm. Given its limited dimensions, this outfit is not capable to produce a significant acoustical power under 120 Hz. As such, a subwoofer incorporated inside the cylindrical transportation case was added in order to power the low-frequency range. The other necessary equipment consists of a power amplifier (i.e. 300 W RMS) installed at the base of the loudspeaker. This compact dodecahedron has been produced by LookLine, a small Italian firm specialized in high quality and customized sound systems.

2.1.3. Microphones
Three different microphonic probes were employed for this experiment:

- Two high-quality cardioids in ORTF configuration (i.e. Neumann K-140, spaced 180mm and inclined by 110° on the horizontal plane);
- A binaural dummy head (i.e. Neumann KU-100);
- A B-format 4-channel pressure-velocity probe (i.e. Soundfield ST-250).

All these microphones were installed onto a rotating table in such a way that the main axis of rotation crossed the center of the dummy head passing through the intersection point of the main axes of the two cardioids (which were mounted just above the dummy head). Alternatively, the Soundfield microphone was displaced at 1m distant from the central axis of rotation, in front of the dummy head. The angular step was established to be 10°, having 36 measurements to complete a whole rotation on the horizontal plane [15, 16].

2.2. Methods for auralization
A basic method to realize the auralization is the convolution. The IRs can be employed as very long FIR filters, to be applied to dry (or anechoic) musical or speech recordings. The convolution is a very efficient filtering technique, particularly resourceful if implemented with specific algorithms on fast processors, as clearly demonstrated in [17].

In the following sections, it is described how to create sets of IRs more suitable for a convolver. To accomplish this purpose, it is possible to adapt the IR results to any of the currently available formats suitable for multichannel reproduction, in order to rebuild as faithfully as possible the spatial attributes of the real sound field [18].

2.2.1. ORTF-stereo impulse response
This is considered the most basic analysis processing, which can reproduce a standard stereo auralization. The process is based on the elaboration of monoaural recordings, possibly undertaken in
anechoic or dry rooms, that should be executed for each section of the orchestra or for each singer in
order to keep separated the frequency ranges [19].

Each monoaural recording has to be convolved with a specific IR stereo, as obtained by doing RIR
measurements by using cardioid microphones in ORTF configuration. Basically, each of these IR
results should be identified by considering the exact position of the sound source inside the performing
arts space [20].

At the end the results obtained by a convolution of all the dry recordings can be included in a single
stereo output file, which is suitable to reproduce the elaborated signals in a stereo modality (i.e. as
perceived by using 2 loudspeakers).

2.2.2. Binaural impulse response (binaural room scanning)
The binaural approach is very similar to the ORTF-stereo method, already explained in paragraph
2.2.1, but with the difference of employing binaural IRs. In this way, the result of the convolution is a
2-channels file, which would be suitable for headphone reproduction [21].

However, two methods can be potentially adopted in order to improve the quality of surround
effects: a cross-talk cancellation can be added for a loudspeaker reproduction; and a head-tracking
sensor that would be preferred for a headphone reproduction, becoming useful to drive a real-time
convolver by switching the IRs being convolved as soon as the listener rotates his head.

![Figure 2. Two loudspeakers angled +/- 10°, fed by a cross-talk cancellation filtering system](image)

2.2.3. B-format impulse responses (Ambisonics)
For this type of auralization, each dry mono source should be convolved with a B-format IR. In other
words, a 4-channel B-format output can be obtained after mixing all these convolutions.

A reproduction of a B-format signal, resulting from a specific array type of loudspeaker, requires an
Ambisonics decoder to compute the input data for each speaker [22].

A creation of a software-based decoder has been pioneered by Farina [16, 23] and then furthermore
finalized by other researchers at the University of York, who released a set of VST plugins that allow
to be manipulating and decoding B-format signals over various loudspeaker rigs.

2.2.4. ITU 5.1 surround (from selected B-format impulse response)
The principal concepts of the ITU 5.1 rendering is the configuration of the microphones selected for
RIR measurements, which drive the main 5 loudspeakers. For each available setup, it is possible to
select a subset of 5 over 36 positions corresponding to where the Soundfield microphone was installed,
possibly closer to the intended positions of the selected setup [24]. Then, a mono IR can be extracted
from a B-format IR measured in each of these 5 designated positions by using Visual Virtual
Microphone software, started to be developed by David McGriffy and advanced by following studies
resulting into a multichannel output [25].

At the end, each mono dry source can be convolved with the 5-channel IR derived from the
corresponding sound source position, and the results of all these convolutions can be mixed in a 5-
channel track as a collection of final output data, which is suitable to be reproduced by a ITU
loudspeaker rig [26].
2.2.5. *Mark Poletti’s high-directivity virtual microphones*

Considering two cardioid microphones, given in ORTF configuration as already discussed into the Introduction, during a complete rotation they describe also a small circumference along their position more in proximity to the rotational axis.

To explain the process in simple words, looking at just one of the two cardioid microphones, the samples gathered during a complete rotation are 36 IRs. A set of various-orders coincident microphones can be derived from this measured data, considering that the microphones ideally should be placed in the centre of the rotation axis, in order to simulate a modified version of the Poletti’s theory [9].

The aim of this method is to establish a class of multileaf-shaped horizontal directivity patterns resulting from various orders. The order 0 is considered as omnidirectional; the order 1 is composed of two crossed figure-of-eight microphones (like an Ambisonics disposed only on a horizontal plane); then order 2 and 3 can be also calculated, having directivity patterns related to the cosine of respectively twice and three times the angle.

From this kind of analysis, an advanced high-order Ambisonics decoder (working only for a horizontal plane) can be fed. An alternative way of employing these signals resulting from a high-order Ambisonics is to drive a standard 5.1 ITU array by synthesizing 5 asymmetrical directivity patterns [27].

2.2.6. *Circular Wave Field Synthesis (WFS) approach*

The 36 B-format measurements undertaken along a 1m-radius circumference are exactly the set of data required to apply the WFS method.

Huygens principles are applied to the realization of this method [28]. The WFS is applied to a certain number of microphones installed inside a dedicated listening area along a specific trajectory of a circumference. As such, the cycle of expansion and shrinking can only be performed onto the horizontal plane. This is one of the ways to derive the feeds for a specific loudspeaker array, which can be applied to a listening room of a medium-sized by starting with a dimension of 1m-radius array [29]. Furthermore, it is possible to stretch the array in such a way that the loudspeakers can be arranged along 4-linear arrays instead of being disposed by using a circular array.

2.2.7. *Hybrid methods (Ambiophonics)*

The Ambiophonics method is a hybrid solution that aims to mask the defects of two basic systems: a cross-talk cancellation to be reproduced for binaural signals emitted by closely-spaced loudspeakers (i.e. Stereo Dipole); and a 3D surround driven by a convolution of oriented virtual microphones [30].

A typical Ambiophonics array can be realized by installing a stereo dipole mounted frontally, with the addition of eight loudspeakers placed along the main three axial directions that surround the whole outfit, as shown in the Figure 3.

The theory for deriving the signals for these loudspeakers has been already presented in the previous sections, and the assembly of the whole system has been extensively described in [31].

![Figure 3. Ambiophonics 3D](image-url)
2.2.8. Beamforming with Multiple Input Multiple Output (MIMO) method

The MIMO approach was developed in order to synthesize any arbitrary polar patterns without the necessity of executing any rotation of the microphone setup. Spherical array microphones (e.g. em32 Eigenmike®, by mhAcoustics) equipped with a considerable number of n-channel transducers, synthesize virtual microphones with any directivity by using time-invariant FIR filters [32]. The uniformly distribution of the capsules over a spherical surface area allows to increase significantly the number of characterization directions where the incident wavefronts can be captured from. In this way the reproduction of the sound field is not limited to the horizontal plane but can be a combination of all the 3 axes playing back a panoramic soundscape very close to the reality [33]. Despite the acoustical properties of the selected theatres were not measured with this technique, following papers will explain more deeply the functions of the beamforming by using a spherical array type of microphone combined with a loudspeaker of similar characteristics.

3. Results

Twelve worldwide theatres were measured with the methods described in Section 2.2, and Table 2 reports the number of the source and receiver points selected in order to best reconstruct the volume of these listening rooms.

| N. | Theatre                                | N. Sources/Receivers |
|----|----------------------------------------|----------------------|
| 1  | Uhara Hall of Kobe,                    | 2/2                  |
| 2  | Japan - Noh Drama Theater of Kobe      | 2/2                  |
| 3  | Japan - Concert Hall of Kirishima      | 3/3                  |
| 4  | Italy - Greek Theater of Siracusa      | 2/1                  |
| 5  | Italy - Greek-Roman Theater of Taormina| 3/2                  |
| 6  | Italy - Auditorium of Parma            | 3/3                  |
| 7  | Italy - Auditorium of Rome (Sala 700)  | 3/2                  |
| 8  | Italy - Auditorium of Rome (Sala 1200) | 3/3                  |
| 9  | Italy - Auditorium of Rome (Sala 2700) | 3/5                  |
| 10 | Australia - Sydney Opera House (opera theater) | 4/3                |
| 11 | Australia - Sydney Opera House (concert hall) | 3/3               |
| 12 | Australia - Sydney Opera House (studio) | 3/1                 |

Figure 4 below shows a quick comparison of the reverberation time (RT) results obtained after IR measurements inside the theatres listed above. The methods utilized for reconstructing the auralization of the twelve selected theatres are considered fully valid for storing the acoustical properties of any valuable listening rooms, such as concert halls and auditoria [34, 35].
4. Conclusions
This paper explains how it is possible to derive subsets of IRs in order to virtually reproduce acoustic characteristics of any size of room, following the available 3D sound technologies. In particular, in relation to the reconstruction of a sound field from IR measurements performed inside a real place, it is noticeable to say that the auralization is possible to be achieved by employing:

- Two loudspeakers reproducing a standard stereo effect;
- Headphones incorporating head tracking reproducing a binaural perception;
- Loudspeakers mounted in a very proximity (with the insertion of cross-talk canceling filters);
- Regular array of loudspeakers reproducing a 2D (horizontal) or 3D (spherical) Ambisonics;
- Standard microphonic setups (e.g. OCT, INA) reproducing a surround ITU 5.1;
- Mark Poletti’s circular-array method reproducing high directivity patterns;
- Wave Field Synthesis approach (WFS) reproducing a wide-area auralization;
- Any combination of the discussed methods, resulting in any type of surround method (Ambiophonics, Panoramio-phonics, and other derivations).

The beamforming technique, in combination with the MIMO method, would be an innovative technology coming from the development of all the listed techniques, which the authors will show in following papers with detailed measurement results and data compared with the techniques already described [36, 37].

The preservation of these important historical buildings throughout the faithfully reproduction of the physical and acoustic characteristics, not only achieves the target proposed by M. Gerzon but also offers the unique experience of a virtual immersive listening to people that never attended live performances in such beautiful architectural masterpieces [38, 39].

Acknowledgments
The authors would like to thank Prof. Angelo Farina for his strong help during the measurement campaign, and all the students that considerably collaborated during the measurements and the post-processing of measured data.
Conflicts of Interest: The authors declare no conflicts of interest.

References
[1] Galvez S, Tang Y, Woodcock J, Jackson P, Melchior F, Pike C, Fazi F, Cox T and Hilton A 2018 IEEE Trans. on Multimedia. Vol. 20
[2] Vigeant M, Wang L and Rindel J H 2008 Acta Acustica united with Acustica 94(6)
[3] Tronchin L and Farina A 1997 JAES 45(12)
[4] Gerzon M 1975 JAES 23(7)
[5] Lokki T and Päätynen J 2020 Ac. Science Technology 41(1)
[6] Farina A and Ayalon R 2003 AES Conf. Surr. Sound Techniques Techn. Percep
[7] Postma B N J and Katz B F G 2017 Jour. Ac. Society America 142(5)
[8] Tervo S, Päätynen J, Kuusinen A and Lokki T 2013 AES Jour. Audio Eng. Society 61(1-2)
[9] Poletti M A 2018 JASA 144(3)
[10] Kuunisen A and Lokki T 2015 Jour. Ac. Society America 138 (5)
[11] Caniato M, Favretto S Bettarello F, Schmid C 2018. Acta Acustica united with Acustica 104(6)
[12] Tronchin L 2012 AES: Journal of the Audio Engineering Society 60(12)
[13] Tronchin L 2013 Acta Acustica united with Acustica 99(1)
[14] Postma B N J and Katz B F G 2015 Virtual Reality 19(3-4)
[15] Pulkki V, Pöntynen H and Santala O AES Journ. Audio Eng. Society 67(11)
[16] Farina A and Tronchin L 2005 Ac. Science Technology 26(2)
[17] Tronchin L and Coli V L 2015 AES: Journal of the Audio Engineering Society 63(9)
[18] Badeau R and Plumbley M 2014 IEEE/ACM Trans. Audio Speech Lang Process 22(11)
[19] Tronchin L, Tommasino M C and Fabbri K 2014 International Journal of Sustainable Energy Planning and Management 3
[20] Tronchin L, Manfren M and James PA 2018 Energy 165(A)
[21] Tronchin L, Manfren M, Vodola V 2020 Applied Sciences 10(2) 633
[22] Kaplanis N, Bech Sø, Lokki T, Van Waterschoot T and Holdt Jensen Sø 2019 Journ. Ac. Society America 146(5)
[23] Mc Cormack L, Delikaris-Manias S, Politis A, Pavlidi D, Farina A, Pinardi D and Pulkki V 2019 AES Journ. Audio Eng. Society 67(11)
[24] Remaggi L Jackson P, Coleman P and Parnell T 2018 Proc. AES Conven. Audio Eng. Society
[25] Tronchin L 2013 International Journal of Mechanics 7(3)
[26] Tronchin L and Fabbri K 2017 Energies 10(10) 1621
[27] Hammond B and Jackson P 2018 IEEE Intern. Conf. Ac. Speech Signal Proc.
[28] Kuttruff H 2009 5th Ed. Spon Press
[29] Tronchin L and Knight D J 2018 Building Acoustics 25 (2)
[30] Farina A, Langhoff A and Tronchin L 1998 Journal of New Music Research 27 (4)
[31] Farina A, Amendola A, Chiesi L, Capra A and Campanini S 2013 AES Intern. Conf.
[32] Shimokura R, Tronchin L Cocchi A and Soeta Y 2011 Journal of Sound and Vibration 330 (14)
[33] Farina A and Tronchin L 2013 Acta Acustica United With Acustica 99(1)
[34] Jeon J Y, Kim J H and Ryu J K 2015 J. Acoust. Soc. Am. 137(3)
[35] Caniato M, Gasparella A 2019 Energies 12(8)
[36] Witew I, Behler G and Vorländer M 2005 Ac. Science Technology 26
[37] Tronchin L, Manfren M and Vodola V 2020 Applied Sciences 10(2) 641
[38] Caniato-M, Bettarello F, Marsich-L, Ferluga-A, Sbaizero O and Schmid C 2015 Construction and Building Materials 102(1)
[39] Tronchin L and Knight D J 2016 Int J of Historical Archaeology 20(1)