



POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Characterisation of an Ambisonics Reproduction Room

TESI DI LAUREA MAGISTRALE IN
MUSIC AND ACOUSTIC ENGINEERING - INGEGNERIA DELL' IN-
FORMAZIONE

Author: **Sofia Parrinelli**

Student ID: 991735

Advisor: Prof. Fabio Antonacci

Co-advisors: Enrico Zoboli, Dott. Mirco Pezzoli

Academic Year: 2023-2024

Abstract

The complexity of a sound field holds many mathematical and engineering challenges for its accurate reconstruction.

Ambisonics is a recording and playback technique capable of representing sound fields through decomposition into a set of spherical harmonics, thanks to a determined mathematical discretisation procedure.

The accurate reproduction of the sound field is confined to a central *sweet spot*, an area that is centrally located within a treated acoustic site. The optimal reproduction system must exhibit a configuration that can surround the *sweet spot*, embracing it as far as possible into the reproduced sound field.

However, empirical test cases challenge this limitation, indicating that even in listening rooms without a perfect anechoic behaviour and with non ideal speaker array layout the representation of the sound field could be sufficiently valid and reliable.

This thesis aims to investigate the quality of the spherical harmonic reproduction within a dedicated spatial configuration which is appropriately designed for audio listening and compatible to host Ambisonics sound system. The objective is to illustrate a method for the assessment of the analyzed site overall quality and propose metrics for the spherical harmonics reconstruction evaluation.

Various factors have been analysed such as the structural room characteristics, the amount of loudspeakers utilized, the listening position placement, the use of equalisation hence, the influence of all these factors on the reproduction of a diffuse field.

The acoustic parameters of the room have been derived and the measured spherical harmonics have been evaluated in relation to the ideal spherical harmonics by means of the proposed specific metrics. These metrics are established for the evaluation of the microphone encoding while, however, their use with regard to the evaluation of audio reproduction is experimental. In the context of an optimal acoustic performance of the room, the proposed metrics have proven to be effective and useful for the assessment of spherical harmonics reproduction rendering within the site.

Keywords: Ambisonics, surround system, spherical harmonics, acoustic parameters.

Abstract in lingua italiana

La complessità di un campo sonoro presenta molte sfide matematiche e ingegneristiche per una sua ricostruzione accurata e soddisfacente.

Ambisonics è un metodo di registrazione e riproduzione in grado di rappresentare i campi sonori attraverso la decomposizione in armoniche sferiche.

La riproduzione accurata del campo sonoro è limitata ad uno *sweet spot*, area che si trova in posizione centrale all'interno di un ambiente acusticamente trattato. Il sistema di riproduzione ottimale deve possedere una configurazione che possa circondare lo *sweet spot* in modo da immergearlo il più possibile nel campo sonoro riprodotto. Tuttavia, l'evidenza empirica mette in discussione questo limite, suggerendo che anche in ambienti d'ascolto dal comportamento non anecoico e con una disposizione delle casse diversa da quella ideale, la rappresentazione del campo sonoro potrebbe essere piuttosto valida.

Questo lavoro di tesi indaga sulla qualità della riproduzione delle armoniche sferiche all'interno di una configurazione spaziale dedicata, la quale è compatibile per l'ascolto e la rappresentazione audio tramite sistema Ambisonics. L'obiettivo è mostrare un metodo per la valutazione della qualità complessiva del sito studiato e proporre una metrica per la valutazione degli armonici sferici ricostruiti.

Diversi sono i fattori analizzati, come l'influenza delle caratteristiche della stanza, della quantità di altoparlanti, della posizione di ascolto e dell'utilizzo di un'equalizzazione sulla riproduzione di un campo diffuso. I parametri acustici della stanza sono stati ricavati e le armoniche sferiche misurate sono state confrontate con le armoniche sferiche ideali tramite metriche assodate per la valutazione dell'encoding microfonico ma sperimentali per quanto riguarda la valutazione della riproduzione audio. Nel contesto di ottime proprietà acustiche della stanza, le metriche proposte sono risultate efficaci anche per una resa della performance di riproduzione degli armonici sferici all'interno del sito.

Parole chiave: Ambisonics, audio immersivo, armoniche sferiche, parametri acustici.

Contents

Abstract	i
Abstract in lingua italiana	iii
Contents	v
Introduction	1
1 Background	3
1.1 Multichannel Audio Representation: a Touch of History	3
1.2 Ambisonics	4
1.3 Encoding Process	8
1.4 Decoding Process	9
1.5 Omnidirectional Microphone	17
1.6 Microphone Array	17
1.7 Obtaining the Impulse Response: Techniques and Applications	19
1.8 Room Acoustic Notions	22
1.8.1 Acoustical Parameters	24
1.9 Ambisonics Listening Spaces: Examples	27
1.10 DANTE Technology	28
1.11 Adobe Audition	29
1.12 Smaart	30
1.13 Room EQ Wizard	30
2 State of the Art	31
2.1 Metrics for Listening Reproduction Systems Assessment	31
2.2 Used Metrics for the Evaluation	32
2.2.1 Spatial Correlation	32
2.2.2 Level Difference	33

2.3 Eigenmike64 [®]	34
3 Listening Room Acoustic Characterization	37
3.1 Analysis Procedure	37
3.2 Results	45
4 Listening Room Performance Measurements	57
4.1 Introduction and Methodologies	57
4.2 Room Equalization	61
4.3 Spatial Correlation and Level Difference Results	64
4.4 Graphical Representations of the Sound Field	71
4.4.1 Spherical Harmonic Visualization	71
4.4.2 Pressure Measured on the Microphone	75
4.4.3 Decoding of the Extracted Spherical Harmonics	78
5 Conclusions	85
5.1 Limitations	86
5.2 Future Developments	86
Bibliography	89
List of Figures	97
List of Tables	101
List of Symbols	104
Acknowledgements	107

Introduction

From the second half of the 19th century onwards, following Edison's studies on recording, many engineers conducted research on the fidelity of sound reproduction and its enhancement in terms of quality. The interest in spatial audio systems and psychoacoustic correlations peaked in 1998/1999, thanks to the earlier studies by M. Gerzon, P. Fellgett and D. H. Cooper, dating back to the 1970s and 1980s dating back to the 1970s and 1980s [24, 41].

Achieving a truly immersive sound field necessitates replicating the complexity of spatial excitation and the intricate features of real-world sound environments, therefore the capture and recreation of such immersive sound fields pose significant challenges. Currently, the achievable levels of sound fidelity are extremely high, giving one or more listeners the impression of being immersed in a realistic, yet virtual, sound environment. This is due to the fact that technologies related to microphones and speakers are now capable of ensuring excellent sound reproduction and the fidelity of contemporary recording equipment has reached remarkable levels, almost approaching the boundaries of the human auditory system's perceptual limits.

Spatial audio unveils information about both active and passive acoustic objects in our surroundings, offering a sense of presence and awareness. The primary objective of a surround sound system is, indeed, essentially to deliver spatial experiences to listeners. The ability to design an optimal immersive sound field holds diverse applications such as virtual reality, entertainment applications, augmented reality, music reproduction, artistic performances, simulations.

The assessment of virtual audio is an issue that involves mathematical, practical, and psychological dimensions. It becomes evident that addressing this challenge comprehensively requires a holistic approach, considering all these elements. Temporal properties, such as fidelity in reproduction concerning frequency and dynamic range, and spatial properties, as accuracy in reproducing the sound field in terms of the directionality of its components, including the spatial characteristics of the room and its acoustic consequences, are crucial elements to consider in sound reproduction. In addition, temporal properties can, of course, be influenced by spatial properties such as absorption or reflections.

This thesis was conducted in conjunction with ASK Industries, the company responsible for the immersive audio reproduction room being studied, so-called Virtual Room, which was designed in-house by their research team for automotive audio research purposes.

The ultimate objective is to offer guidance for assessing spaces devoted to immersive listening and compatible with Ambisonics, subsequently present potentially improvement strategies. This thesis focuses on a numerical assessment, evaluating mathematical parameters, rather than relying on a subjective and psychoacoustic evaluation.

The whole project can be divided into two main sections. The first part is devoted to the acoustic characterization of the site through extrapolation of acoustic parameters useful to provide a qualitative assessment of the case study and to evaluate whether the requirements for an Ambisonics reproduction are provided. Once established that the acoustical requirements for effective immersive sound reproduction are fulfilled, the second step of the analysis proposes a method for the evaluation of the reproduction of the spherical harmonics within the site. Level Difference and Spatial Correlation are two metrics proposed in [60] for the evaluation of the encoding quality of a microphone array. In this study, the mentioned metrics are used to assess the reproduction system, encompassing both the room itself and its speaker array. Two listening positions were analyzed to compare the quality in a central position and in a lateral position, in order to attain an idea of the *sweet spot* confinements. The presence or absence of a proper tuning led to further considerations. An adequate equalization, in fact, has proven to be crucial for the enhancement of the performance captured by the proposed metrics Level Difference and Spatial Correlation. The thesis is organized in the following chapters:

Chapter 1 gives an overview on the theoretical topics that this work touches, such as Ambisonics basic concepts and with it decoding process, encoding process, considerations on microphones array and reproduction sites. Room acoustic notions are given in this first chapter and a brief overview on the utilized tools is included.

Chapter 2 describes the metrics Spatial Correlation and Level Difference and introduces the microphone array em64 Eigenmike®.

In *Chapter 3* the acoustic analysis of the room and the evaluation of the acoustic parameters are illustrated. Furthermore, the measurement procedure is defined and discussed.

Chapter 4 elucidates the processing, the set-up and the results of the spherical harmonic reproduction analysis. Here, the metrics described in *Chapter 2* are extracted and the considerations for different listening positions and tuning are carried out.

Chapter 5 draws conclusions. Here, considerations on the obtained results, limitation of the proposed approach as well as suggestion of future development are depicted.

1 | Background

1.1. Multichannel Audio Representation: a Touch of History

At the outset of the exploration into sound reproduction, mono was the only playback system known, hence everything was recorded and played back through a single channel. However, a single channel could not provide the spatiality, a sense of sound envelope, or of depth. Huge developments occur with the acquisition of two channels, thanks to Blumlein theorisations (1933). The goal during the decades has always been and will continue to be in future to maximize the spatial rendering and to allow more audiences to listen.

Multichannel reproduction, where stereophony is its simplest form made of two channels, manages each channel with its associated speaker, which constitutes an independent sound source. Multiple sound sources can be managed to create the illusion of phantom sources and virtual origins of the sound source.

An important treatise from 1953, [73], describes, from the perspective of the period, the research and objectives achieved in the field and suggests the use of two channels as the best compromise between economic, technological feasibility of sound reproduction systems of that era. On the other hand, Vermeulen expresses in [79]:

"Although stereophonic reproduction can give a sufficiently accurate imitation of an orchestra, it is necessary to imitate also the wall reflections of the concert hall, in order that the reproduction may be musically satisfactory. This can be done by means of several loudspeakers, distributed over the listening room, to which the signal is fed with different time-lags."

Since the 1970s, quadraphonic arrangement has been used by setting four channels and four speakers positioned at the corners of the square surrounding the listener. This configuration did not provide satisfactory psychoacoustic results and did not guarantee an accurate sound image reproduction. The use of an appropriate decoding matrix led Dolby to develop Dolby Surround: one channel to the front-left, one in the front-center, and one

to the front-right, with a single surround channel used to drive several speakers arranged next to and behind the audience. Reverberation and other ambient peculiarities are directed into the surround channel, as well as various sound effects, enveloping the audience.

The term *Surround System* refers to multichannel audio systems which are characterized by the presence of five or more channels, the related loudspeakers are strategically positioned around the listener to create a fully immersive audio experience.

For an insight on the history of multi-channel systems development it is suggested the reading of [77].

1.2. Ambisonics

The basic concepts behind this particular type of *Surround System* were initially developed by Duane Cooper in 1972 [24] in Illinois. Concurrently, Peter Fellgett and Michael Gerzon developed their concept of Ambisonics [41] in England. Peter Craven played a significant role in designing the required microphone and, with the support of the NRDC (National Research Development Corporation), the team in the UK marketed the Ambisonics recording/playback system. For further historical insights refer to [13].

Ambisonics is a sound field system based on spherical harmonic expansion as discretization of the sound field.

The theory of Ambisonics incorporates optimal decoder theorems for a faithfully reproduction of the sound field, taking into account psychoacoustic criteria based on optimizing the reconstructed phase and energy vectors [17, 40]. Ambisonics technique guarantees valid reproducibility within the so-called *sweet spot*, an area where the spatial audio reproduction is optimized for the immersive listening and accurate sound localization. Within this *sweet spot*, fidelity and coherence are maintained throughout the sound field reproduction process.

The directivity information of the first order in a single point consists in four components that constitute the *B-Format*: the omnidirective signal pattern (called W and considered as the zero-th order) and the bidirectional, figure of eight, aligned with the three cartesian axes (called respectively X, Y, Z). The first order ensures the correct reproduction of the sound field within a small listening area due to its low spatial resolution limits.

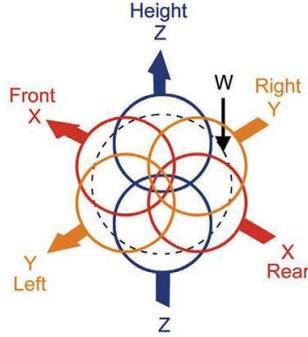


Figure 1.1: B-Format of the first order Ambisonic, picture from: [5]

The prerequisite necessary to recreate an ideal diffuse field, because of its intrinsic nature, is the decorrelation of the reproduced signal.

High directional correlation, that still characterize the first order Ambisonics, could not guarantee a sufficient resolution for the mapping of diffuseness, spatiotemporality and depth.

The need is, at this point, the utilization of higher order Ambisonics (HOA), these can improve exponentially the reproduction or recording resolution at the risk of increasing the technological complexity. HOA correspond, mathematically, to the angular portion of the solution to the wave equation [20]. HOA are, therefore, capable of improving the directional resolution and refining the mapping of non correlated fields, ensuring higher spatial depth quality of perception and helping in the spatial enlargement of the *sweet spot*.

HOA perform a sort of spherical Fourier Transform of all acoustic events that are present around the reference point, leading to the spherical harmonics signals.

As explained in [26, 27], the spherical harmonic decomposition of the acoustic pressure field consists in transforming the wave equation into the spherical coordinate system, characterized of azimuth θ , elevation ϕ and radius distance r for each point from the center point O . The formulation leads to the following Fourier-Bessel decomposition.

From the wave equation, in cartesian coordinates, where $p(x, y, z, t)$ is the pressure in a point placed in (x, y, z) at the time instant t , c is the speed of sound:

$$\nabla^2 p(x, y, z, t) = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} p(x, y, z, t) \quad , \quad (1.1)$$

the separation of the time varying component $e^{i\omega t}$ from the wave equation gives the Helmholtz wave equation, where k is the wave number:

$$\nabla^2 p(x, y, z) + k^2 p(x, y, z) = 0 \quad . \quad (1.2)$$

By utilizing the spherical coordinate system:

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} p \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} p \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} p = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} p(r, \theta, \phi, t) \quad , \quad (1.3)$$

The Fourier-Bessel decomposition can be obtained by extracting the spatial components from 1.3:

$$p(r, \theta, \phi) = \sum_{n=0}^{\infty} 4\pi i^n j_n(kr) \sum_{-n \leq m \leq n} B_{nm} Y_{nm}(\theta, \phi) \quad . \quad (1.4)$$

with k referred to the wave number $k = \frac{2\pi f}{c}$, n the Ambisonics order and m its degree. B_{nm} are the resulting spatial components, frequency dependent coefficients that represent the sound field within the sphere region centred on the origin point. Y_{nm} are the spherical harmonics. Ultimately, $j_n(kr)$ are the spherical Bessel functions of the first kind, whose curves depicts the contribution to the sound field as a function of the distance r from the center O .

The Ambisonics coefficients of order n are generated by considering up to the n -th order derivatives of the sound field at the central point. Physically, the first component, which can be considered zero-th order $B_{0,0}$ (omnidirectional channel W), represent the pressure field. Its spatial derivative information, i.e. pressure gradient and acoustic velocity (correlated one to the other following the Euler equation), are given by the successive order (X, Y, Z channels). At second and higher orders, Ambisonics adds information about higher order derivatives of the pressure field.

In practice, the summation is done, for the encoding process, up to a finite number of order N , leading to a number of spatial components equal to $(N + 1)^2$ if the problem is addressed in three dimensions.

For example, recalling the first order Ambisonics (including with this wording the zero-th order), as already seen in figure 1.1, will be composed by the parameters:

- W= $B_{0,0}$ omnidirective component: pressure field element.
- X= $B_{1,1}$ bidirective component: pressure gradient and acoustic velocity information.
- Y= $B_{1,-1}$ bidirective component: pressure gradient and acoustic velocity information.
- Z= $B_{1,0}$ bidirective component: pressure gradient and acoustic velocity information.

The spherical harmonics are basis functions defined on the unit sphere as:

$$Y_{nm}(\theta, \phi) = \sqrt{\frac{(2n+1)}{2} \frac{(n-m)!}{(n+m)!}} P_{nm} \cos(\theta) e^{im\phi} \quad . \quad (1.5)$$

Spherical harmonics exhibit the following orthonormality property:

$$\int_0^{2\pi} d\phi \int_0^\pi Y_{nm}(\theta, \phi) Y_{n'm'}^*(\theta, \phi) \sin(\theta) d\theta = \delta_{nn'} \delta_{mm'} \quad . \quad (1.6)$$

Where the integration is over the unit sphere and δ is the Kronecker delta function, see [81] for a complete analysis. P_{nm} are the associated Legendre functions having order n and degree m .

Each Ambisonics order $n > 1$ includes, therefore, $2n+1$ spherical harmonics, leading to the total number of $(N+1)^2$ as already expressed.

The spatial representation of the spherical harmonic is shown in the following image:

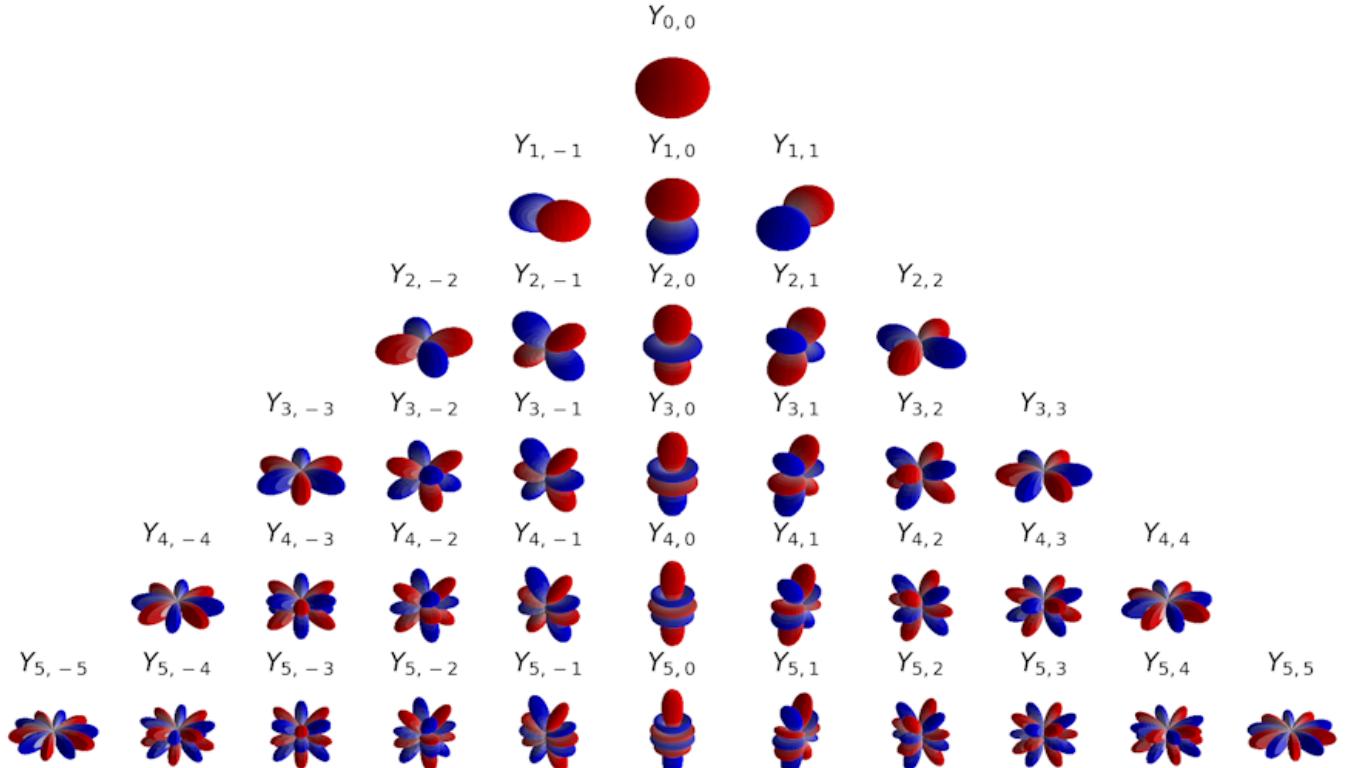


Figure 1.2: Spherical harmonic 3-D representation up to 5-th order. Picture taken from [31].

The natural sound scene will be, therefore, decomposed by the encoding, with appropriate

microphones, while the reassembly from spherical harmonics (thus discretised sound field) into the new sound field generated by the reproduction system is handled by decoding.

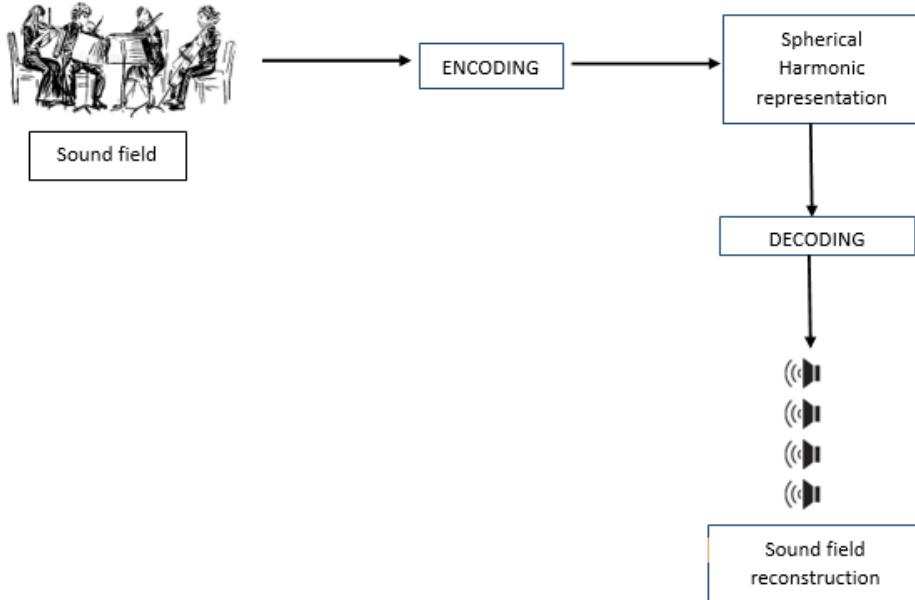


Figure 1.3: Brief Ambisonics system scheme.

1.3. Encoding Process

The encoding process is characterized by equations that allow the composition of virtual sound scenes produced by a series of virtual sources located in the space around the microphone.

For the first order, the encoding of the sound field in the four channels W, X, Y and Z is based on the *B – format* and the directivity patterns depicted in figure 1.1.

The spherical harmonic decomposition of the sound field is performed over the virtual sound sources (far field sources as well as near field sources) and this process provides the encoding equations. The encoding equations are, thus, the expression of $B_{nm}(\omega)$, frequency dependent spatial components, as function of the positions of the signal and of the virtual source. If the source is assumed to be far enough, the contribution can be intended as a plane wave and the encoding process the following:

$$\mathbf{B}(\omega) = \mathbf{S}(\omega) \cdot \mathbf{Y} \quad , \quad (1.7)$$

being:

$$\mathbf{B}(\omega) = [B_{0,0}(\omega), \quad B_{-1,1}(\omega), \quad \dots, \quad B_{NM}(\omega)] \quad , \quad (1.8)$$

$$\mathbf{S}(\omega) = [S^\omega(\theta_1, \phi_1), \quad S^\omega(\theta_2, \phi_2), \quad \dots, \quad S^\omega(\theta_U, \phi_U)] \quad , \quad (1.9)$$

while:

$$\mathbf{Y} = \begin{bmatrix} Y_{0,0}(\theta_1\phi_1) & Y_{1,-1}(\theta_1\phi_1) & Y_{1,0}(\theta_1\phi_1) & \cdots & Y_{NM}(\theta_1\phi_1) \\ Y_{0,0}(\theta_2\phi_2) & Y_{1,-1}(\theta_2\phi_2) & Y_{1,0}(\theta_2\phi_2) & \cdots & Y_{NM}(\theta_2\phi_2) \\ \vdots & \vdots & \ddots & & \vdots \\ Y_{0,0}(\theta_U\phi_U) & Y_{1,-1}(\theta_U\phi_U) & Y_{1,0}(\theta_U\phi_U) & \cdots & Y_{NM}(\theta_U\phi_U) \end{bmatrix} . \quad (1.10)$$

The encoding equation of a plane wave coming from (θ_u, ϕ_u) , $1 \leq u \leq U$ can be expressed as a simple process of weighting the signal $\mathbf{S}(\omega)$ with factors that depend on the wave incidence which are, essentially, spherical harmonic functions.

The extraction of the spatial components for general field sources (not only plane waves) is possible through a matrix of filters that consider a frequency dependent factor that models the near field effect and therefore the wave front curvature.

Optimized microphones or microphone arrays for the B-Format recording should have a specified capsules arrangement. Generally, microphone capsules should be arranged as closely spaced as possible in order to ensure quality at high frequency and maximize the aliasing frequency. These aspects will be discussed in the following chapters.

1.4. Decoding Process

The decoder has the aim to match the continuous directivity patterns that has to be recreated with the associated channel, for the first order W, X, Y, Z, and mapping the signal s_l in the proper discretized directions given by the loudspeakers positions θ_l, δ_l .

$$s_l(t) = w_lW(t) + x_lX(t) + y_lY(t) + z_lZ(t) \quad . \quad (1.11)$$

Where the parameters w_l, x_l, y_l, z_l define the decoding.

The Ambisonics mapping is, therefore, the distribution of the signals to the loudspeaker array by utilizing proper weights obtained from the discretization of functions of each Ambisonics order. The goal is, therefore, the recomposition of the Ambisonics components at the centre of the loudspeaker array, being this area the listening position. The decoder reconstructs the sound field through the combination of presumed plane waves, assuming

that the loudspeakers act as far-field sources and the layout is a concentric speaker array with equidistance from a center listening point. This process involves a matrix operation, specifically combining the signals with real weighting gains:

$$\mathbf{S}(\omega) = \mathbf{D}(\omega) \cdot \mathbf{B}(\omega) , \quad (1.12)$$

with $\mathbf{S}(\omega)$ the component, for the specify frequency, of the vector of emitted signals, L being the number of loudspeakers:

$$\mathbf{S}(\omega) = [S^\omega(\theta_1, \phi_1), \quad S^\omega(\theta_2, \phi_2), \quad \dots, \quad S^\omega(\theta_L, \phi_L)] , \quad (1.13)$$

$\mathbf{B}(\omega)$, as already specified, is the vector of Ambisonics components $B_{nm}(\omega)$, which is frequency dependant, that are intended to be recomposed:

$$\mathbf{B}(\omega) = [B_{0,0}(\omega), \quad B_{-1,1}(\omega), \quad \dots, \quad B_{NM}(\omega)] . \quad (1.14)$$

$\mathbf{D}(\omega)$ is the decoding matrix, its dimensions are $L \times (N + 1)^2$:

$$\mathbf{D} = \begin{bmatrix} d_{0,0}^1 & d_{1,-1}^1 & d_{1,0}^1 & \cdots & d_{NM}^1 \\ d_{0,0}^2 & d_{1,-1}^2 & d_{1,0}^2 & \cdots & d_{NM}^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ d_{0,0}^L & d_{1,-1}^L & d_{1,0}^L & \cdots & d_{NM}^L \end{bmatrix} ,$$

this matrix either can be frequency dependant, this possibility depends on the chosen method for the decoding matrix entries computation. In fact, the decoding matrix can be defined in various ways and different methodologies are available [58, 84].

In the case regular polygon and polyhedra loudspeaker arrays are provided, more straightforward computations can be followed in order to design the proper decoding matrix.

A primary method consist in firstly sampling the spherical harmonics at the speaker directions θ_l, ϕ_l up to $l = L$, therefore get its pseudo-inverse to find the basic decoding matrix.

$$\mathbf{D} = \mathbf{Y}^\dagger = \mathbf{Y}^T (\mathbf{Y} \cdot \mathbf{Y}^T)^{-1} , \quad (1.15)$$

where \mathbf{Y} is the matrix composed by the spherical harmonics:

$$\mathbf{Y} = \begin{bmatrix} Y_{0,0}(\theta_1\phi_1) & Y_{1,-1}(\theta_1\phi_1) & Y_{1,0}(\theta_1\phi_1) & \cdots & Y_{NM}(\theta_1\phi_1) \\ Y_{0,0}(\theta_2\phi_2) & Y_{1,-1}(\theta_2\phi_2) & Y_{1,0}(\theta_2\phi_2) & \cdots & Y_{NM}(\theta_2\phi_2) \\ \vdots & \vdots & \ddots & & \vdots \\ Y_{0,0}(\theta_L\phi_L) & Y_{1,-1}(\theta_L\phi_L) & Y_{1,0}(\theta_L\phi_L) & \cdots & Y_{NM}(\theta_L\phi_L) \end{bmatrix}.$$

The Sampling Decoder is another traditional design approach which simply consist in mapping the spherical harmonics into the desired space discretization.

$$\mathbf{D} = \frac{4\pi}{L} \mathbf{Y} \quad (1.16)$$

The matrix $\mathbf{D}(\omega)$ is, therefore, composed by gains that aims to recompose the encoded sound field at the centre listening position basing on the type of speaker layout and the decoder strategy applied, as explained in [20, 87].

If the arrangement of the array deviates severely from an ideal speaker layout, these basic methods might lead to an imperfect solution, hence perfect reconstruction cannot be guaranteed across the entire rendering volume.

The most simple decoders seen, i.e. 1.15 and 1.16, assume that the loudspeakers produce plane waves. In literature, other decoding methods are discussed to overcome this assumption. In particular, in [27] the authors introduced a Near Field Compensation considering the radial aspects.

Gerzon in [17], introduced two parameters: r_v and r_e , well described in [51]. Being G_l the loudspeaker gain associated to the loudspeaker l :

- r_v is the velocity localization vector as sum of the signals from the speakers. It predicts low-frequency localization almost perfectly.

$$r_v \hat{u}_v = Re \frac{\sum_{l=1}^L G_l \hat{u}_l}{\sum_{l=1}^L G_l} \quad (1.17)$$

with L number of loudspeakers, \hat{u}_l the vector connecting the listening central point with the speaker direction and $\hat{u}_v = (\cos \theta_v, \sin \theta_v)$.

- r_e is the energy localization vector as sum of the squares of all the speakers signals summed. Its value represents the concentration of source energy in the direction θ_e

and predicts quite well mid-frequency localization.

$$r_e \hat{u}_e = Re \frac{\sum_{l=1}^L (G_l G_l^*) \hat{u}_l}{\sum_{l=1}^L (G_l G_l^*)} \quad (1.18)$$

with $\hat{u}_e = (\cos \theta_e, \sin \theta_e)$.

Maximizing the value of the energy localization vector and getting it to point in the right direction is the crux of the decoder design problem. The goals of a proper decoding system that takes into consideration psychoacoustics needs are:

- Obtain constant amplitude gains for all source directions.
- Obtain constant energy gains for all source directions.
- Obtain a correct reproduced wavefront direction and velocity (r_v) at low frequencies.
- Obtain a maximum concentration of energy in the source direction (r_e) at high frequencies.
- Matching high and low frequency perceived directions.

Proper loudspeaker gains G_l , associated to the decoder matrix, that characterize the decoding design, can help pleasing as much as possible these needs, see Max-rE and All-Round Ambisonic Decoding (AllRaD), see [61, 85] for a proper insight. AllRaD is probably the most adaptable decoder technique available [84] and matches the properties of vector-base amplitude panning with Ambisonics advantages. AllRaD firstly applies a sampling decoder to a very high resoluted t -design source layout in order to guarantee optimal virtual values of r_e and r_v . As a second step, AllRaD redistributes these virtual signals, computed through the sampling decoder, to the real loudspeaker array through a VBAP matrix[88].

$$\mathbf{D} = \mathbf{G} \frac{4\pi}{L'} \mathbf{Y} \quad , \quad (1.19)$$

considering \mathbf{Y} built on L' imaginary sources redistributed by the VBAP matrix $\mathbf{G} = [G_1, \dots, G_l, \dots, G_L]$, well described in [69].

Loudspeaker Array Configuration

Achieving accurate audio reproduction quality of a surround system requires a non-reverberant site while obtaining a physically accurate reproduction across the entire audio bandwidth, especially over an extensive listening area, requires an exceedingly large number of loudspeakers and Ambisonics order.

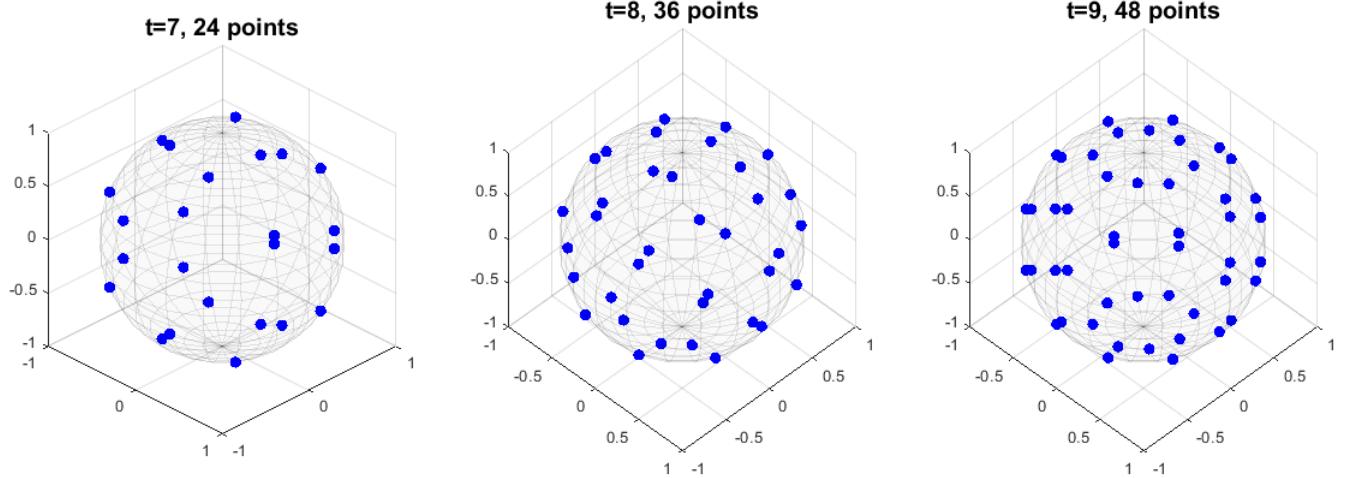


Figure 1.4: T-design configurations from [11].

A t -design, as expressed in [86], represents a distinctive form of spatial discretization designed to optimize loudspeaker arrangements for Ambisonics. The approach involves approximating a unit sphere using a finite set of points.

This approximation ensures that the integral over the sphere of a polynomial of degree t equals the average value of the same polynomial evaluated at the t -design set of points [28, 42, 44]. This characteristic establishes an equal-weight quadrature rule on the sphere, making t -designs particularly advantageous for Ambisonics applications, [66, 86].

For a set of points $p = [J_1, \dots, J_P]$ on a unit sphere S , any polynomial f having degree at most equal to t has to satisfy the following equation:

$$\int_S f(x) dS = \frac{1}{P} \sum_{p=1}^P f(J_p) \quad (1.20)$$

The spherical t -design are not unique, therefore, for a number of points P the maximum degree t has to be found.

In a t -design with L speakers, $t \geq 2N + 1$ should be used since the integral evaluates the product of the spherical function with spherical harmonics of up to order N and the optimal positioning of speakers enables accurate reproduction of Ambisonics panning

functions. By using a t -design with a sufficient number of speakers L , the total number of required output channels remains determined by $(N + 1)^2$. However, the specific arrangement of speakers in a t -design enhances the efficiency and accuracy of Ambisonics reproduction. A discretization with a t -design where $t \geq 2N + 1$ guarantees the independence of the energy vector from the panning direction and ensures constancy in the energy centroid (the Cartesian center of mass of the energetically weighted unit sphere), this is demonstrated in [85].

A conventional Ambisonics decoding is straightforward only with optimal loudspeaker configurations, achieving direction-independent energy, energy spread, estimated phantom source loudness, and width.

Decoders like AllRaP/AllRaD [85–87] are more versatile, working for arbitrary loudspeaker arrangements, aiming to create phantom sources with stable loudness and adjustable width.

In practical playback situations, mounting speakers exactly at the locations of a t -design is often impractical. Compromises regarding the region covered by loudspeakers and their actual positions may be necessary. In such cases, careful Ambisonic decoder design becomes necessary, as ensured by AllRaD. The common approach of decoders like AllRaD, as already expressed in 1.19, is to decode to a virtual layout of regular loudspeakers, and then decode this layout of virtual loudspeakers to the real layout using technique as vector-base amplitude panning (VBAP) [85, 88].

Concluding this theoretical overview section, the majority of traditional Ambisonic decoder designs for all orders $n \leq N$, necessitate a minimum of

$$L = (N + 1)^2 \quad (1.21)$$

loudspeakers evenly distributed to cover a spherical space. In principle, AllRaP and AllRaD possess the capability of Ambisonic playback for higher N on a wide range of loudspeaker configurations, even when the number of loudspeakers is limited and covers only a portion of the surrounding directions.

Considerations on Arrival Time Difference

The time difference between the arrival of loudspeaker signals has been shown to be an important factor in the perception of spatial images due to psychoacoustic factors, this is well discussed and studied in articles such as [23, 52, 57, 75].

When two identical stimuli are presented with a short time delay from different speakers, experiments showed that the sound image shifts toward the leading sound as the time delay

increases. As a result, the signal that arrives first is prioritized, and the image is localized from the direction of the first sound. These perceptual phenomena are collectively referred to as the *Precedence Effect* (or *Law of the First Wavefront* [21], *Haas Effect* [43]). Increasing the time delay further, both signals may be heard as different sound elements. The point at which this occurs is called the echo threshold.

In multi-loudspeaker reproduction, the *sweet spot* is approximately located at the same distance from all speakers. Obviously, without a proper equalization, even though the listener stands within the *sweet spot*, signals emitted by different speakers can arrive at different times, strengths, and directions. The impact of the *Precedence Effect* depends on these factors, especially on Arrival Time Difference (ATD).

As explained in [23], by setting the ATD from 0 to 1 ms, the perceived position of the phantom image will gradually shift toward the position of the loudspeaker that is leading in time. This is the region of summing localisation, where the contributions of the speakers are integrated.

From 1 ms to at least 5 ms, the *Precedence Effect* becomes increasingly predominant, hence the contribution from the lagging speaker is largely suppressed. In this region of localization dominance the direction of the leading signal dominates the perception.

After 5 milliseconds, the transition region starts to appear. Here the listener starts hearing audio events distinctly on each speaker. This last range is called the echo threshold.

Evaluations on Ambisonics Order Limitations

Exact reproduction of a diffuse sound field requires that expansion 1.4 is satisfied for all orders and all modes up to infinity even though most of the power of the sound field within the chosen reproduction region is contained in the first orders. Then the field could be well reproduced by equating only the terms referred up to a defined order N .

Assuming that the field $p(r, k)$, on a sphere of radius r , is approximated by a series $\hat{p}(r, k)$ up to order N , by truncating the infinite series, the normalized truncation error is defined in [80] as:

$$\epsilon_N = \frac{\int |p(r, k) - \hat{p}(r, k)(r, k)|^2 d\hat{r}}{\int |p(r, k)|^2 d\hat{r}} \quad (1.22)$$

Integrating over the unit sphere region, with k being the wave number.

The following image depicts the trend of the normalized truncation error as function of the product of the wavenumber k and the sphere radius from [80] which is proven to be equal to:

$$\epsilon_N(kr) = 1 - \sum_{n=0}^N (2n+1)(j_n(kr))^2$$

Where $p(r, k)$ is the field produced by a plane-wave source.

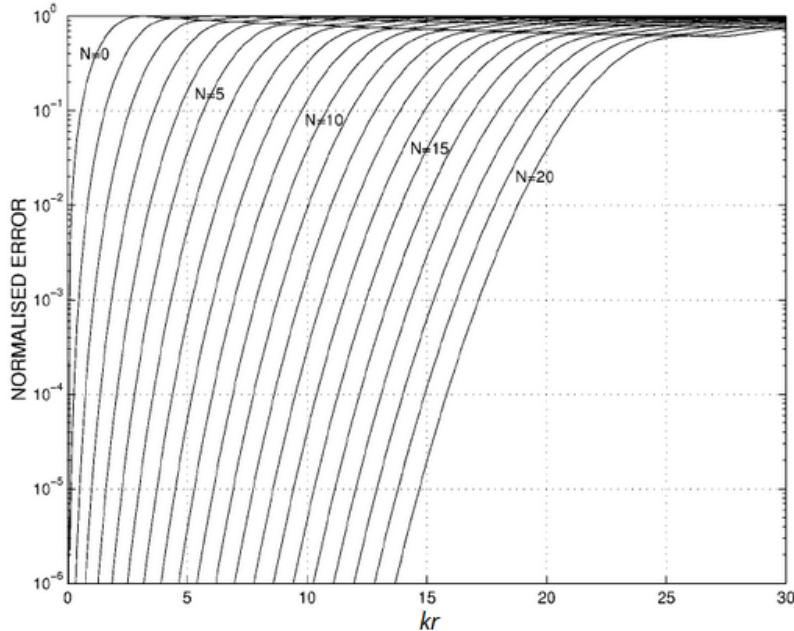


Figure 1.5: From [80] p. 4: Normalized truncation error as a function of kr for various reproduction orders N .

Therefore, for a given order, the higher the operating frequency, the smaller the reproduction sphere, and for any order n , the error decreases monotonically below a certain frequency.

If the order expansion is adequate enough to depict the plane wave field on the circumference of a sphere of a certain radius, it implies that the field is accurately replicated at every point inside the sphere. This affirmation aligns with the Kirchhoff–Helmholtz theorem, asserting that the sound field at any given point within a source-free volume can be entirely determined by the sound pressure and pressure gradient on the continuous surface enclosing that volume.

In general, looking at the graph, the following simple rule of thumb can be used to determine the reproduction order:

$$N = kr$$

Ensuring an error of around 4%, this should be sufficient for a good quality reproduction.

1.5. Omnidirectional Microphone

One of the microphones used for the current study is an omnidirectional microphone designed with high-precision piezoelectric technology, enabling exceptional sensitivity to acoustic pressure, and its board is a printed circuit.

The omnidirectionality of the microphone ensures a uniform frequency response from all directions, a crucial feature for measurements within the surrounding audio site.

The capsule of the microphone is composed of piezoelectric materials. These materials have the property of generating an electric charge when subjected to mechanical deformations, such as the acoustic pressure caused by sound waves. When sound waves impact the microphone capsule, the piezoelectric materials within the capsule undergo mechanical deformation in response to the variation in acoustic pressure. This mechanical deformation generates an electric charge proportional to the intensity of the sound.

The microphone capsule incorporates a condenser structure with two plates (one movable and one fixed) separated by air or a dielectric. Inside this capsule, piezoelectric materials are present.

The printed circuit houses and connects the electronic components necessary for amplifying the signal from the capsule. The produced signal is then transmitted through the circuit to be processed and amplified.

The combination of these technologies enables the microphone to provide high sensitivity, extended frequency response, and the ability to capture subtle sonic details.

1.6. Microphone Array

A microphone array is a device capable of simultaneously capturing an acoustic field in more spatial locations.

Typically, it consists of a collection of multiple microphones in a particular layout arrangement and a beamformer. The primary function of the beamformer is to integrate the signals from the individual microphones, thereby, basing on the computed beamforming algorithm, deriving a representation of the perceived acoustic field.

Specific directions of the acoustic field can be enhanced and others attenuated thanks to the beamformer, enabling the achievement of more intricate directivity patterns beyond the conventional configurations: Bidirectional, Cardioid, Supercardioid, and Hypercardioid and its implementation can lead to the acquisition of data for sound field representation systems as Ambisonics.

The layouts of the microphone capsules can be various, as a three dimensional layout. In this particular arrangement the capsules can be placed all equidistant from a center

point, therefore in a spherical layout. The sphere dimension and the number of capsules characterize the the quality of the sound field capture.

All the capsules ensures an equi-sensibility to the arriving sound waves.

One of the initial prototypes for a spherical microphone was the tetrahedral microphone, pioneered by Gerzon and Craven, commonly referred to as the *SoundField* microphone [30]. This microphone consisted of four capsules strategically positioned at the corners of a tetrahedron, forming a virtual sphere. *SoundField* microphones use an array configuration known as the A-format. The A-format employs four sub-cardioid capsules, i.e. a polar response that is slightly more omni than cardioid mounted on the surface of a tetrahedron. Basic operation processing of the A-format leads to the B-format components.

- $X = 0.5((LF-LB) + (RF-RB))$
- $Y = 0.5((LF-RB) - (RF-LB))$
- $Z = 0.5((LF-LB) + (RB-RF))$
- $W = 0.5(LF+LB+RF+RB)$

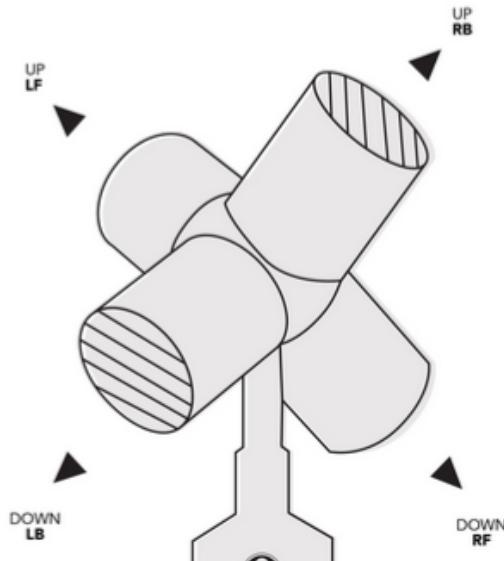


Figure 1.6: *SoundField* microphone structure, picture from [30].

The four capsules can pick up all the data needed for a basic first-order representation of the sound field.

1.7. Obtaining the Impulse Response: Techniques and Applications

The whole acoustical study within the listening room has been carried out by the extraction and the analysis of impulse responses, i.e. the pressure at a given receiver position with respect to an impulse or proper signal coming from a particular source position.

The impulse response contains all the information that depicts how the room behaves when exposed to the sound impulse in that particular measurement position.

In room acoustics studies, the accurate measurement of the impulse response is fundamental, since acoustical parameters are derived from it.

There are various possible techniques in order to measure the impulse response which have been developed for different purposes and configurations, employing various test signals and types of post processing. Citing some of them, the MLS (Maximum Length Sequence) method [65], in particular IRS variant (inverted repeated sequence), seems to be the most useful method in the case of occupied room or exterior sites due to its strong immunity to the possible incoming noise. The Time-Stretched Pulses technique is based on a time expansion and compression technique of an impulsive signal, [19] with the aim to increase the amount of sound power emitted for a fixed magnitude without increasing the nonlinearities.

The recording of natural pulses is the original definition of a room impulse response and will be described in few paragraphs.

A well described and depth analysis of the different measurement method can be found in [74].

Sinusoidal Sweep Excitation

The Time Delay Spectrometry method was invented in 1967 by Richard Heyser [45]. For practical and convenient reasons (by the need to set up a frequency range validity in the whole frequency range in one signal, with respect to the old methods lacks, see [33]), linearly swept sine wave signals are utilized:

$$s(t) = \sin(q \cdot t^2) \quad (1.23)$$

where q is the so-called *sweep rate* in $\frac{\text{Hz}}{\text{second}}$. The swept sine wave begins at frequency $f = 0$, and "sweeps" linearly upward toward higher frequencies.

In the linear sweep, the instantaneous frequency $f(t)$ varies linearly with time:

$$f(t) = f_0 + qt \quad (1.24)$$

and the energy is constant for all frequencies.

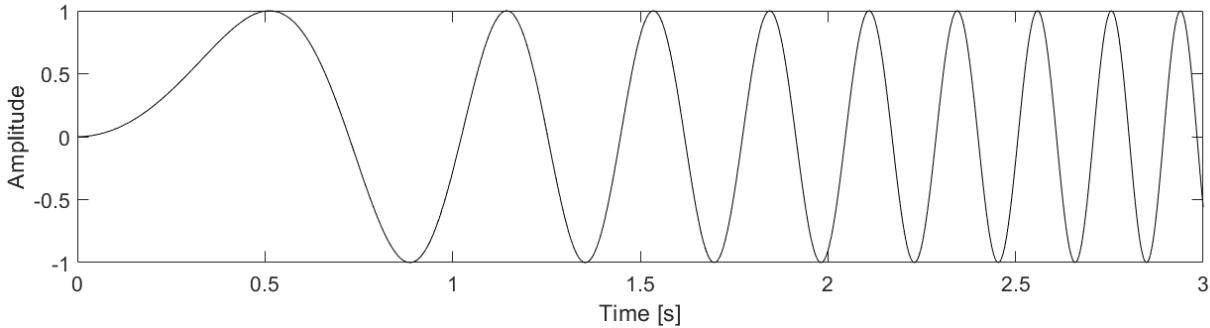


Figure 1.7: Sinusoidal sweep time trend, $q = 6$.

The noise level can be reduced by lengthening the sweep signal. The distortion level can be reduced by decreasing the volume but this can increase the background noise contaminating the result. Hence, a compromise level has to be set for every measurement site [62].

A linear sweep has a “white” spectrum, and consequently provides a signal-to-noise ratio which is not sufficiently good enough at low frequencies. For this reason, often the measurement is split in two sections, with a slower sweep rate for the first low frequency part, followed by a second section with a higher sweep rate covering medium and high frequencies.

Being the loudspeaker often subjected to non-linear phenomena, and the subsequent propagation inside the room is not perfectly time-invariant, the useful part of the impulse response has to be extracted: the interesting part to measure is the impulse response of the linear system $h(t)$, removing the artifacts caused by noise, non-linear behavior of the loudspeaker and time-variance.

The method developed by Farina, see [32], is based on an exponential sweep test signal and provides a satisfactory answer to these problems: the noise rejection properly high and non-linear effects result perfectly separated from the linear response.

The mathematical definition of the exponential sine sweep signal is the following:

$$s(t) = \sin \left[\frac{\omega_1 \cdot T}{\ln \left(\frac{\omega_2}{\omega_1} \right)} \cdot \left(e^{\frac{t}{T} \cdot \ln \left(\frac{\omega_2}{\omega_1} \right)} - 1 \right) \right] \quad (1.25)$$

With ω_1 starting frequency, ω_2 ending frequency in T seconds.

The sine sweep signal characteristics to be defined are, therefore, the range of frequency to excite, the sweep duration and the duration of silence to be inserted after the sweep.

After recording the signal, the post-processing has to be done to obtain the linear system's impulse response $h(t)$.

The output signal has to be, then, convolved with a proper filtered signal.

The convolution can be easily computed by multiplying the signals in frequency domain in order to limit the computational time.

$$h(t) = y(t) \otimes f(t) \quad (1.26)$$

↓

$$H(\omega) = Y(\omega) \cdot F(\omega) \quad (1.27)$$

Where $f(t)$ is the time-reversal of the excitation signal $s(t)$.

It is noted that, being $x(t)$ the sine sweep delivered to the speaker:

$$x(t) \otimes f(t) = \delta(t) \quad (1.28)$$

The obtained signal $h(t)$ in (1.26) is the desired impulse response, which ideally should be a perfect Dirac's delta but in reality it contains the room characteristics and information, spuriousness due to reflections, distortions, or any noise, so conditions due to the inherent properties of the speaker and the room.

If the exponential sweep with fade-in and fade-out is utilized, the obtained impulse response often shows a significant pre-ringing just before the direct sound.

The way of controlling this type of pre-ringing is to create a proper filter, and to apply it to the measured IR. This filter is called 'Kirkeby filter' thanks to its developer Ole Kirkeby [53]. The sweep signal $s(t)$ is converted into the frequency domain and the computation of its inverse through Kirkeby algorithm is done:

$$F(\omega) = \frac{S(\omega)^*}{S(\omega)^* S(\omega) + \nu(\omega)} \quad (1.29)$$

Where $\nu(f)$ is a frequency dependent regularization parameter.

When the filter obtained is convolved with the measured signal, the result will be the optimal impulse response.

By utilizing a proper windowing, the required portion of the impulse response can be extracted, discarding noise and non linearity.

In terms of sweep design, it may be useful to adjust the spectrum of the sweep to offset variations in the measured spectrum, such as the non-flat frequency behaviour of the source.

Balloon Popping Excitation

The excitation of an enclosed space through an impulsive noise is a used method for probing the acoustics of a space [39].

By exciting the system with a "natural" impulsive signal as close as possible to a Dirac impulse δ , the recording of the room's response to this produced sound represents its impulse response without the need of a further post processing.

In particular, the pop of a balloon can excite the site in a wide frequency range generating a relatively uniform radiation pattern without include the speaker intrinsic characteristics. The purpose should be, then, the evaluation limited to the physical space itself and not on eventual lacks of the reproduction system.

1.8. Room Acoustic Notions

Modal and Diffuse Domains

For the acoustic study of a room, it is important to distinguish two types of situations in which the acoustic analysis must be delineated following the different characteristics of sound propagation and behaviour.

The Schroeder frequency is the theoretical limit separating these two regions: the modal region and the static region.

Falling within the modal region, the environment is defined as *acoustically small*, here the resonance modes of the room dominate and greatly influence the frequency response, developing valleys and peaks on its magnitude.

The listening quality results highly compromised and suggestible to small shifts of the listening point or source and the reproduction does not remain faithful due to the unbalanced frequency response of the room, [37].

The generation of standing waves is verified on the basis of specific relationships between the considered frequency and room dimensions. This results in nodal points, where the amplitude will be minimum, and antinodal points, where the amplitude will be maximum.

The choice of speaker position and listening area is, therefore, of paramount importance in order to avoid nodal points and so, energy gaps.

The overall impulse response of the room is, in fact, implied by the response of the speakers, including their specific output characteristics, and the response of the room behaviour.

In *acoustically large rooms*, on the other hand, modes will only be present in very low frequencies, even below the audible threshold and the recreated sound field can be considered diffuse. A diffuse field consists of an infinite number of plane waves propagating in space which are decorrelated to each other and evenly distributed in all directions.

The Schroeder frequency f_c depends not only on the room volume V but also on its average reverberation time RT , thus on all its acoustic characterisations and geometric structure.

$$f_c = 2000 \sqrt{\frac{RT}{V}} \quad (1.30)$$

The denotations of *acoustically large room* and *acoustically small room* do not refer to the room size but to the factors listed above.

If the walls of the room are of equal size and parallel in pairs, the resonances will be particularly well defined while in asymmetrical rooms the modes continue to exist but the particular geometry provides the necessary scattering to highly limit them.

Room modes arise, therefore, for low frequencies that interact with the size of the site. Very high frequencies with a short wavelength are not related to the objects in the room. In the modal region the field cannot be considered diffuse, modal analysis is necessary, on the other hand, in the static region the diffuse field is ideally achieved, therefore a homogeneous and isotropic sound field.

As it is not the purpose of this study to delve into these topics, [54, 72] are recommended for further information.

Considerations on Low-Reverberation Chamber

Recording studio, listening spaces and home studios come under the *small room* classification, where modes occurs.

An intensive acoustic treatment should be therefore applied in order to get the required sound spread qualities. The addition of absorbing materials can reduce the amount of energy of resonant modes, dissipating sound energy rather than reflecting it.

However, even with the use of absorbing materials, it is quite difficult to completely eliminate the totality of resonant modes within a room. This is particularly true in the lower frequencies, where modes can have very large wavelengths and can be more difficult to

control with absorbing materials.

In a low-reverberant room, the sound absorption level is high, therefore the reflections of sound waves are minimised. This leads to a significant reduction of the resonant modes energy, especially at higher frequencies. However, even in well treated rooms is unlikely to completely eliminate all low frequencies resonant modes.

Non-symmetrical geometries, non-parallel walls and the design of the structure considering various walls dimensions helps much the scattering of the sound waves, distributing the energy more evenly within the room, eliminating direct reflection.

Some of the reflected energy is scattered in many directions instead of being reflected directly back into the room. This scattering, indeed, reduces the possibility of direct reflections and helps reduce the creation of standing waves and resonant modes.

Resuming, acoustic scattering can help achieve an even spread of sound within a space, ensuring that acoustic energy is distributed equally in all directions rather than being concentrated in a few specific areas.

The drier the room acoustically, approaching an anechoic situation, the more similar the behavior of the acoustic field will be to a free field (the ideal field that anechoic chambers aim to approximate).

As soon as the chamber is sufficiently treated, with a low reverberation time and limited resonance modes, the manipulation of the sound field to be reconstructed can enjoy an extended range of possibilities.

A spherical array of speakers should be inserted in the correct and studied positions, i.e. not in correspondence with any nodal points still present and in an appropriate global configuration, the central position turns out to be an excellent listening point for multi-channel audio systems.

To sum up, the elimination of acoustic interference by means of an appropriate sound absorption study and space management with a studied layout of speaker arrays are the basis for an immersive surround system. The ideal would be the construction, through the array, of a listening area approaching a diffuse behaviour which, as explained in 1.2, is the necessary basis for a good Ambisonics system.

1.8.1. Acoustical Parameters

A proper study on the necessary acoustic parameters has been carried out in order to assess the acoustical quality of the room under study.

Room acoustic investigates how physical boundaries cause alterations of an original acoustic signal, influencing the way it is perceived by the listener.

The room acoustic parameters are useful for describing the possible acoustic alterations

present in a closed space as well as give a feedback on the quality and characteristics of this space. These parameters can help matching the subjective impression to the physical measurable qualities.

The ISO standard lists provide definitions and measurement method of the acknowledged acoustical parameters. ISO 3382-1 [9], for example, contains guidance for the parameters extracted in the analysis.

Standards, though, do not give clues on how to interpret the obtained values, nor about the target values the room should have basing on the purpose, nor a room classification guidance [38].

Energy Decay Curve

The energy decay curve EDC or Schroeder curve is defined as:

$$EDC_t = 10 \cdot \log \left(\frac{E_t^\infty}{E_0^\infty} \right) \quad (1.31)$$

Where the energy is computed from the Schroeder's backwards integration, being $h(\tau)$ the measured impulse response properly filtered in octave bands:

$$E(t) = \int_{t_0}^{\infty} h^2(\tau) d\tau = \int_{-\infty}^{t_0} h^2(\tau) d(-\tau) \quad (1.32)$$

Due to the impulse finite length t_F :

$$E(t) \simeq E_{t_0}^{t_F} = \int_{t_F}^{t_0} h^2(\tau) d(-\tau) \quad (1.33)$$

The energy decay curve gives the graphical representation of the sound pressure level decay along time, this decay starts at 0 dB in correspondence of the detected peak of the impulse response in time domain, therefore in correspondence of the direct sound.

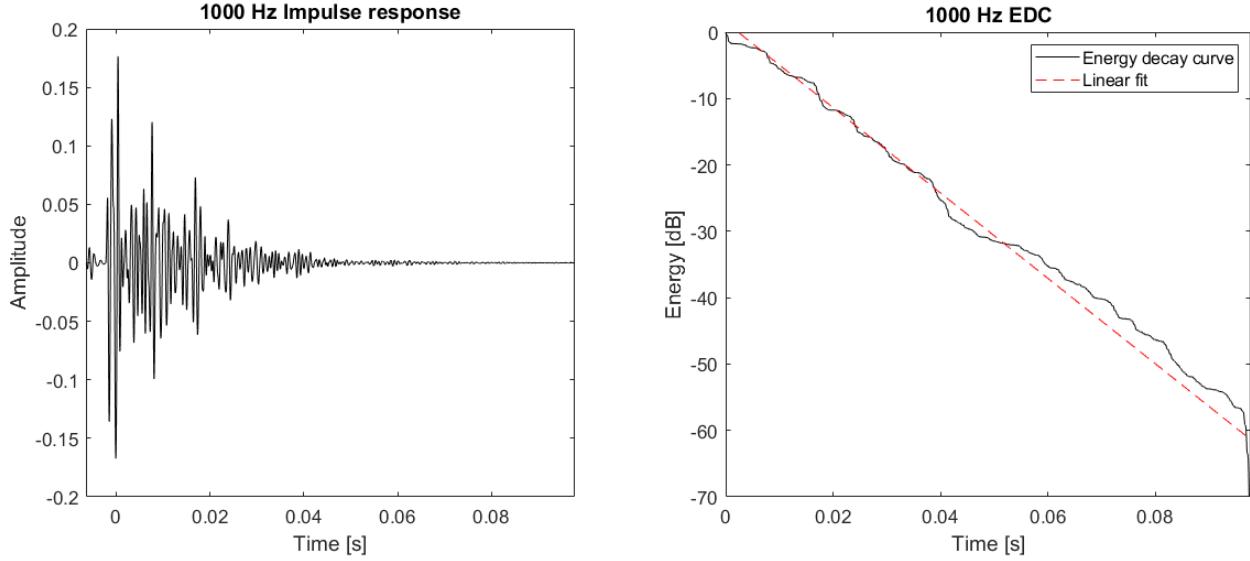


Figure 1.8: Octave-band filtered Impulse response and related Energy Decay Curve at 1000Hz.

From the Schroeder curve all the main acoustical parameters can be computed.

Reverberation Time

Reverberation time, T_{60} is defined in [9] as:

“the duration required for the space-averaged sound energy density in an enclosure to decrease by 60 dB after the source emission has stopped”.

Therefore T_{60} is the value of EDC where the slope reaches -60 dB.

Reverberation time can range from 0.1 seconds, or less, in anechoic chambers, to 10 seconds or more in large spaces. The more useful parameters for the current analysis are:

- Early decay time EDT or T_{10} : time where EDC reaches -10 dB
- T_{20} : amount of time where EDC moves from -5 dB to -25 dB

Clarity

The measure of clarity C_{80} denotes the value of EDC at $t_e = 80$ ms, while C_{50} denotes the value of EDC at $t_e = 50$ ms. C_{80} is used when the purpose is the study of room

acoustical quality for music reproduction sources, while C_{50} assesses the influence of room acoustics in speech intelligibility. In this case the C_{50} will be taken into account because of the particular properties of the room under study.

$$C_{50} = 10 \log_{10} \frac{\int_0^{50ms} h^2(t) dt}{\int_{50ms}^{\infty} h^2(t) dt} \quad (1.34)$$

1.9. Ambisonics Listening Spaces: Examples

The current section briefly illustrates three examples of notable Ambisonics audio playback sites: the *Espace de Projection* in Paris, located at IRCAM, *Sonosfera®* in Pesaro's Civic Museum and IEM-Cube hosted in the Institute of Electronic Music and Acoustics University of Music and Applied Arts of Graz.

The *Espace de Projection* is a medium-sized performance hall created, under the supervision of Pierre Boulez, by the architects Renzo Piano, Richard Rogers and acoustician M. Peutzand, inaugurated in 1978 and recently renovated. The focal point of this site is its adaptability, offering extensive flexibility in shape, size, and acoustic characteristics. Reverberation time can range from 0.4 to 4 seconds. The hall hosts technological equipment for Wave Field Synthesis and Ambisonics. For the Ambisonics reproduction a 75-speaker dome is provided. This loudspeaker disposition comprises four linear horizontal arrays and a box-shaped array that encompasses the walls and ceiling. The horizontal arrays are utilized for 2D Higher-Order Ambisonics up to 128-th order, while the box-shaped array is specifically designed for 3D HOA up to order 9.

Sonosfera® is located in Pesaro and offers a mobile technological amphitheater, designed by David Monacchi, equipped with an array of 45 custom-built loudspeakers uniformly positioned in spherical layout. The site benefits from high acoustic sound-absorbing properties and has been specifically designed for the spherical reconstruction of HOA field recordings.



Figure 1.9: Sonosfera® graphical rendering from [14].

The IEM-Cube is a 120 m² soundproofed concert hall that provides a 49 channel sound system with over 80 loudspeakers. Inside the hall, a hemispherical array of 24 loudspeakers is installed, this allows reproduction of immersive sound fields through Ambisonics. The 24 loudspeakers are aligned in three rings, the lower ring presents twelve speakers in order to achieve a maximum localization in the horizontal plane. The middle ring consists of eight loudspeakers while four speakers are located in the top ring. For a proper insight the following readings are suggested [15, 82].



Figure 1.10: IEM-Cube, [7].

1.10. DANTE Technology

Developed by Audinate, Dante, acronym of Digital Audio Network Through Ethernet, stands out as a leading solution for digital audio distribution over Ethernet networks.

Audio networking plays a central role in modern audio systems, facilitating the seamless transmission of digital audio signals between various devices. Traditional analog audio connections are being replaced by digital alternatives, offering improved flexibility, scalability, and efficiency. In this evolving landscape, Dante has emerged.

Audinate is the driving force behind the development of DANTE technology. The company specializes in audio networking solutions and has played a crucial role in shaping the way audio is transmitted, processed, and managed in professional environments.

DANTE enables the high-performance distribution of digital audio signals operating on standard Ethernet networks. DANTE has many key features as:

- DANTE's ability to maintain low-latency transmission ensures minimal delay between input and output.
- DANTE networks are inherently scalable, allowing users to easily expand their audio systems by adding more devices. This scalability is particularly advantageous in

large installations where numerous audio channels need to be managed.

- DANTE is designed to be interoperable with various audio devices, being perfectly suited for live events, recordings, broadcasts, and more.
- DANTE allows simultaneous transmission of multiple digital audio signals and different devices within the same network can receive a channel as the carrier of the audio signal.

Equipment needed for using DANTE:

- Devices that need to be connected to each other (DANTE-compatible).
- A switch to act as a "bridge" for connections.
- Cables to connect the devices (Cat 5E or higher).

With the free software 'DANTE® Controller,' it is possible to determine from which specified DANTE enabled device to which other specified DANTE enabled receiver the audio signal should be routed across the network.

1.11. Adobe Audition

Adobe Audition is a multichannel audio management, analysis, and recording software that allows optimal real-time signal visualization and frequency response. The Aurora package [2] ensures a set of modules that enable analysis and operations on signals such as:

- Linear Convolution: Allows the convolution of two signals.
- Deconvolve IRS Signal: Facilitates the deconvolution of Impulse Response Signal.
- Generate SWEEP Signal: Generates a SWEEP signal for various applications.
- Acoustical Parameters: - Enables the calculation of Acoustical Parameters defined in the ISO standard 3382 across 10 octave bands. For reverberation time estimation, the squared impulse response is backward-integrated, with an optional noise-removal algorithm.
- Inverse Filter: Implements an inverse filter operation by time reversing the original signal.
- Flatten Spectrum for Equalization: Flattens the spectrum for equalization purposes.

- Kirkeby's Inverse Filter: Utilizes Kirkeby's algorithm in order to compute the inverse filter

1.12. Smaart

Smaart is an advanced software designed for real-time audio management and analysis, with a particular focus on achieving instant frequency response through the convolution process of the measurement with the stimulus signal.

Smaart offers a platform for multichannel audio signal management, analysis, and recording. One of Smart's distinctive features is its ability to perform real-time analysis, allowing users to obtain instant responses while manipulating signals, moving the microphone position or modify obstacles disposition real time.

Real-time convolution with pink or white noise is an effective method for measuring and visualizing the frequency response of a system. This real time process provides an instantaneous representation of sound transmission across the frequency spectrum in that particular measurement position.

1.13. Room EQ Wizard

REW (Room EQ Wizard) is free software for room acoustic analysis and loudspeaker measurement. Among the useful tools REW can offer: measuring SPL, impedance, analyze frequency and impulse responses, measuring distortion, visualization of useful plots as decay plots, waterfalls, spectrograms and energy-time curves and many other applications.

REW offers the possibility to automatically adjusting the settings of parametric equalisers adjusting responses to match a target curve.

The equalizer tool is used to determine the proper filters to apply to a given response in order to reach an ideal target properly selected.

The tool offers many equalizer types that are characterized by different settings as resolution, number of allowable filters, and other adjustment ranges.

2 | State of the Art

2.1. Metrics for Listening Reproduction Systems Assessment

Over the years, various research experiments have been conducted proposing a method for the evaluation of immersive audio reproduction systems and different evaluation metrics have been employed. Metrics for the assessment of speech transmission quality are well described by the procedure ITU-T P.800 [48], while for audio it is possible to refer to ITU-R Rec. BS.1534-3 [50] (from which is designed the WebMUSHRA platform), BS.1116-3, [46], and P.1310 [47] for the spatial audio quality evaluation. All the mentioned and cited metrics involve subjective evaluation, hence through a sample of listeners which can be more or less experienced.

Approaching to a more objective metric, POLQA [49] metric provides an evaluation by predicting speech quality through digital speech signal analysis. POLQA is a full-reference algorithm, thus necessitates the original signal, and analyzes the reference signal with respect to the test one sample-by-sample, after performing a temporal alignment of both. The Perceptual Evaluation of Audio Quality (PEAQ) [76] is, instead, a standardized algorithm with the purpose of propose an objective measure of the perceived audio quality. The algorithm simulates the perceptual properties of the human ear and integrates multiple model output variables to constitute the metric. PEAQ characterizes the perceived audio quality similarly to how sample of listeners would rate it in a listening test following ITU-R BS.1116 ([46]), therefore modeling its mean opinion scores, ranging from 1 (bad) to 5 (excellent).

The literature offers many subjective evaluations methods, [18, 70, 71] describing the various set up possibilities and the different different guidelines to follow. Examples of procedures include listeners utilizing pointers to accurately indicate the perceived source direction as well as reporting the source's origin direction through graphical interfaces representing the area. A subjective experiment can be, of course, considered valid if a sample of careful listeners, without specific hearing problems, is recruited in the experiment. It's worth noting that listening cues differs from person to person, and the reported

quality may be subjected to significant generalizations due to our psychoacoustic interpretation and, therefore, distortions. Thus, the conclusion up to now is that there is not a widely recognized objective and mathematical metric for evaluating an immersive audio reproduction, except through algorithms formulated on generalized subjective cases ([76]) that assume as reference values subjective experiments results.

The AMBIQUAL metric, proposed in [64], provides a method for the objective evaluation of audio listening and localization, the used algorithm is the ViSQOLAudio and the procedure consists in comparing B-format signals obtained with the encoding. AMBIQUAL can predict the perceived quality and the spatial localization accuracy by computing a signal comparison directly from the B-format Ambisonic components. The spectro-temporal trends are compared for each ambisonic channel (spherical harmonic component), measuring the similarity between the reference and the test audio signal through intensity level differences.

2.2. Used Metrics for the Evaluation

Two metrics were taken in consideration in this thesis work in order to evaluate the spherical harmonic recreation in the studied site: Spatial Correlation and Level Difference.

These order-dependent performance metrics are introduced in previous researches to assess the quality of the encoding process in Ambisonics microphones [60, 67, 68].

The microphone under investigation was, in these researches, subjected to a test signal coming from a singular source, a properly well equalized speaker (presenting a flat frequency response). The microphone, or, alternatively, the speaker, underwent spatial displacements facilitated by a precision turning and rotating table covering a grid of directions sufficiently wide to ensure an omnidirectional excitation.

The parameters in question subsequently reported the validity of the spherical harmonics encoding process under the assumption of an ideal omnidirectional excitation.

2.2.1. Spatial Correlation

The Spatial Correlation is a metric that characterizes the similarity between the generated patterns and the ideal spherical harmonic patterns.

For each order and degree, the ideal spherical harmonics set \mathbf{Y} , obtained following 1.5, is compared to the measured spherical harmonics $\hat{\mathbf{Y}}$ functions obtained from the encoded sound field.

$$SC_{nm}(\omega) = \frac{\sum_{l=1}^L \left(\hat{Y}_{nm}^\omega(\theta_l, \phi_l) \right)^T \cdot Y_{nm}(\theta_l, \phi_l)}{\sum_{l=1}^L \left(\sqrt{\left(\hat{Y}_{nm}^\omega(\theta_l, \phi_l) \right)^T \cdot \hat{Y}_{nm}^\omega(\theta_l, \phi_l)} \cdot \sqrt{(Y_{nm}(\theta_l, \phi_l))^T \cdot Y_{nm}(\theta_l, \phi_l)} \right)} . \quad (2.1)$$

The index (θ_l, ϕ_l) corresponds to each azimuth and elevation of the individual loudspeaker l , and ω denotes the frequency for the reason that the measured spherical harmonic will be frequency dependent.

After averaging the metrics across all the d directions, further averaging is then applied among the spherical harmonics of index m belonging to the same Ambisonics order n , obtaining a parameter which is only order and frequency dependant:

$$SC_n(\omega) = \frac{1}{2n+1} \sum_{m=-n}^n |SC_{nm}(\omega)| \quad . \quad (2.2)$$

Spatial Correlation varies in the range of $[0, 1]$, with the ideal value being 1 when each Ambisonics order is perfectly reconstructed.

2.2.2. Level Difference

The Level Difference is calculated as the disparity between the mean level across all directions, which is the total energy sum of the encoded response over all points, and the corresponding ideal components:

$$LD_{nm}(\omega) = \frac{1}{L} \sum_{l=1}^L \frac{|Y_{nm}(\theta_l, \phi_l)|^2}{\hat{Y}_{nm}^\omega(\theta_l, \phi_l) \cdot \hat{Y}_{nm}^\omega(\theta_l, \phi_l)^*} \quad . \quad (2.3)$$

Level Difference values varies in the range of $[-\infty, +\infty]$ dB, with the ideal value being 0 dB when each Ambisonics order is perfectly reconstructed.

Again, as function of the Ambisonics order throught the averaging:

$$LD_n(\omega) = -10 \log \left[\frac{1}{2n+1} \sum_{m=-n}^n LD_{nm}(\omega) \right] \quad . \quad (2.4)$$

2.3. Eigenmike64[®]

em64 Eigenmike[®] is a spherical microphone array implemented by MH Acoustics [10] and made of 64 omnidirectional electret microphones mounted on a rigid spherical baffle that contains inside all the electronic components, see [3]. The array is able to decompose the sound field into spherical harmonics up to the sixth order HOA. The microphone is powered by PoE (Power-over-Ethernet).

The em64 Eigenmike[®] performs 24-bit A/D conversion and microphone calibration takes place internally. The resulting 64 calibrated microphone signals can be accessed on a Mac or Windows PC through Dante's virtual sound card.

All the eigenbeams are computed such that they have an ideally flat magnitude frequency response, up to the spatial Nyquist frequency. It is crucial to emphasize that eigenbeams of higher orders are constrained to narrower operating frequency ranges. This restriction is implemented to minimize the generation of system noise during the encoding process, with the aim of controlling the maximum level of such noise.

The possible selectable beampatterns are: omnidirectional, hypercardioid, supercardioid, cardioid and dipole.

Below some technical characteristics:

- maximum SPL 130 dB
- frequency response from 20 Hz to 20k Hz
- diameter of the sphere of 8.6 cm



Figure 2.1: em64 Eigenmike[®] microphone, picture from [4].

Spatial Aliasing

A phenomenon influenced by the physical spacing between the individual capsules of a microphone array is the spatial aliasing. This spacing determines an upper frequency limit for the signals recorded that has to be considered in the analysis of HOA signals. A greater distance d between sensors establishes a more restrictive frequency limit beyond which spatial aliasing occurs. Therefore, in order to avoid aliasing the distance d has to be:

$$d < \frac{\lambda_{min}}{2} \quad , \quad (2.5)$$

being λ_{min} the minimum wavelength that corresponds to the maximum frequency f_{max} .

$$\frac{c_0}{f_{max}} > 2d \longrightarrow f_{max} < \frac{c_0}{2d} \quad . \quad (2.6)$$

By the way, it is common to define the upper frequency limit as follows:

$$f_{max} < \frac{c_0}{4 \cdot d} \quad . \quad (2.7)$$

Therefore, for the em64 Eigenmike[®], considering the speed of sound $c_0 = 343 \frac{m}{s}$ and the distance d between the capsules equal to 0.025 meters:

$$f_{max} = 3430 Hz \quad .$$

Range of Validity

The forthcoming analyses in this thesis work will require as a prerequisite the consideration of the encoding properties of the em64 array as ideal.

It is possible to rely on the consultable datasheet [59] of the Eigenmike[®] em32 to get an idea of the validity ranges for each order of the device.

It's crucial to highlight that eigenbeams of higher orders are confined to narrower operating frequency ranges. This limitation is implemented to minimize the potential self-noise generated by the system during the encoding process. The table provided below outlines the minimum operating frequency for each order.

Eigenbeam Order, n	Cutoff Frequency
0,1	30 Hz
2	400 Hz
3	1000 Hz
4	1800 Hz

Table 2.1: Table from p.12 of [59].

In the absence of a suitable datasheet for the em 64 model, future evaluations should be made considering similar behaviour and reliability to the em 32 model, being the microphones in both models electret microphones.

3 | Listening Room Acoustic Characterization

3.1. Analysis Procedure

Preliminary Considerations

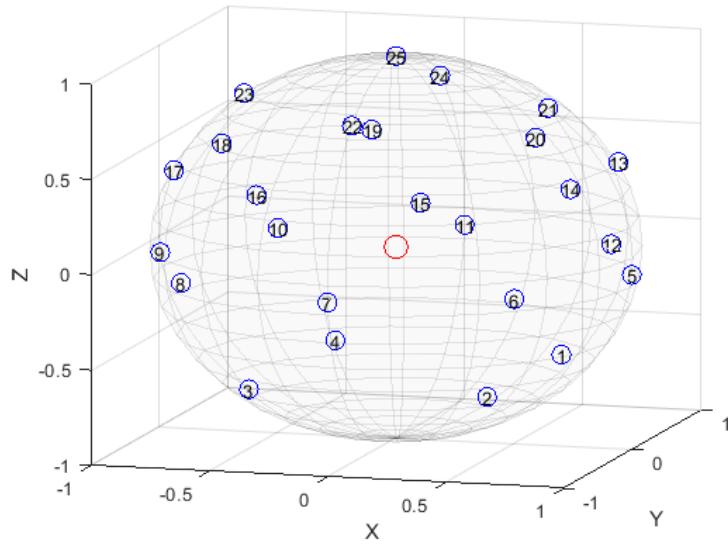
The site to be studied is devoted to simulations and tuning projects for automotive purposes and was acoustically specially designed. The only structural parallels present are the ceiling and floor, vertical walls provides a geometrical structure designed specifically to restrict resonant modes.

- Maximum width: 3.54 m
- Maximum depth: 4.45 m
- Height: 3 m

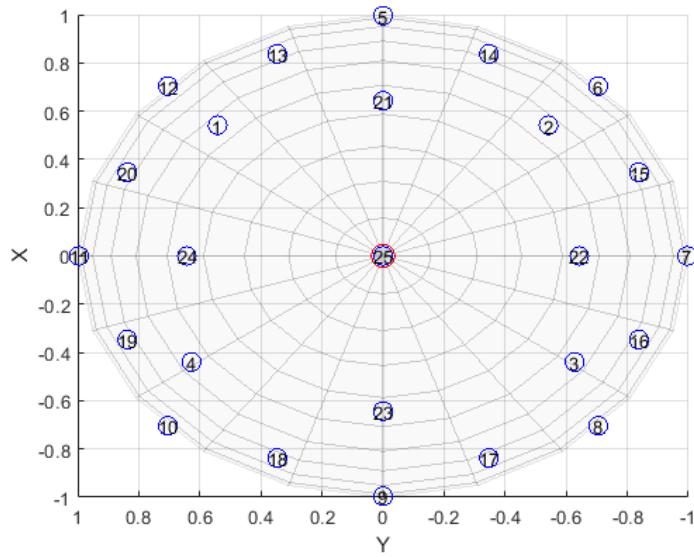
The walls are covered with specially designed sound-absorbing panels.

The spherical array of speakers consists of 25 tweeters, 25 woofers and two subwoofers, see figures 3.1. The latter are positioned under the seat, which is centred in the middle of the spherical array of speakers.

It is noted that the array layout is not a t -design configuration, although the surrounding space seems to be, at sight view, well covered and discretised by the speakers positions. A screen is placed in front of the listening position for the utility of the room and a car seat is positioned in the listening position, as shown in figure 3.2.



(a) Speaker array lateral view.



(b) Speaker array top view.

Figure 3.1: Array layout views. The listening position is depicted in red, the TV screen is just above speaker number five.

The vertical subdivision of the speaker is the following:

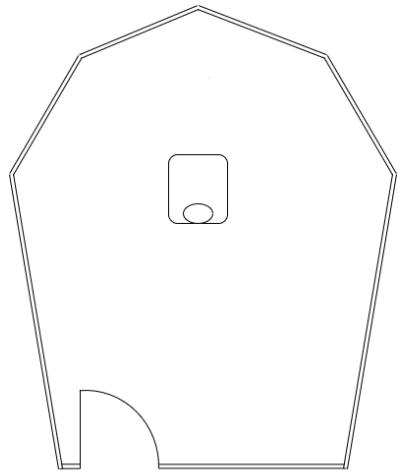
- First level, located on the floor: speakers 1, 2, 3, 4.
- Second level, just below listener shoulder's height: speakers n. 5, 6, 7, 8, 9, 10, 11, 12.
- Third level, above listener head's height: speakers n. 13, 14, 15, 16, 17, 18, 19, 20.

- Fourth level, almost ceiling level: speakers n. 21, 22, 23, 24
- Fifth level, ceiling height: speaker n. 25.

The subwoofers are not considered in the Ambisonics analysis due to frequency range intrinsic limitations. The speaker array is capable of reproducing up to 4th-order Ambisonics by using a common decoder, as detailed in 1.21.



(a) Virtual Room.



(b) Indicative and simplified floor plan of the Virtual Room.

Figure 3.2

Loudspeakers Compensation

The first practical step is the achievement of an optimal frequency response of the loudspeakers. The aim is to reach the appropriate starting point in order to analyze the response of the room.

In order to do that, a flat response of the loudspeakers is necessary.

The loudspeaker exhibits non completely flat behaviour in the spectrum when subjected to an exponential sweep signal, the goal, thus, is to stimulate it by using an suitably pre-distorted signal that is able to compensate the irregularities.

The recording are carried out in a dead room, sufficiently far away and isolated from unwanted reflection. The following distances are reported:

- 1.9 meters from the first obstacle
- 1.5 meters from the floor
- 1.5 meters from the ceiling

- 1.5 meters: distance microphone to the loudspeaker

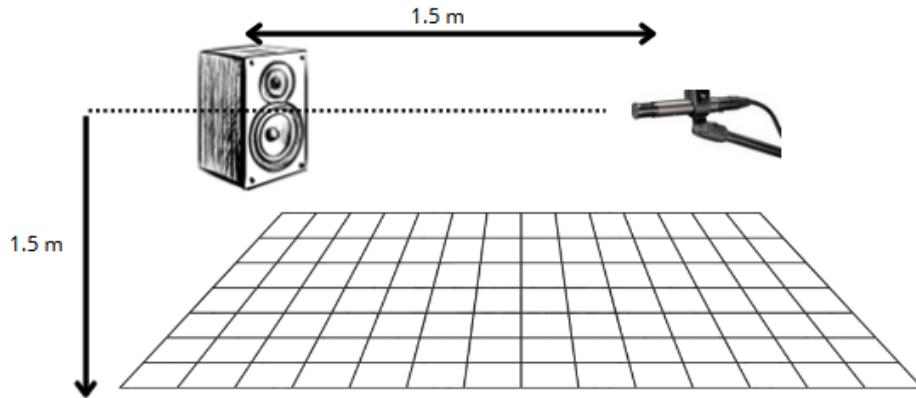


Figure 3.3: Measurement spatial setup

The measurements are recorded by an omnidirectional microphone.

In the following graph the set up is depicted:

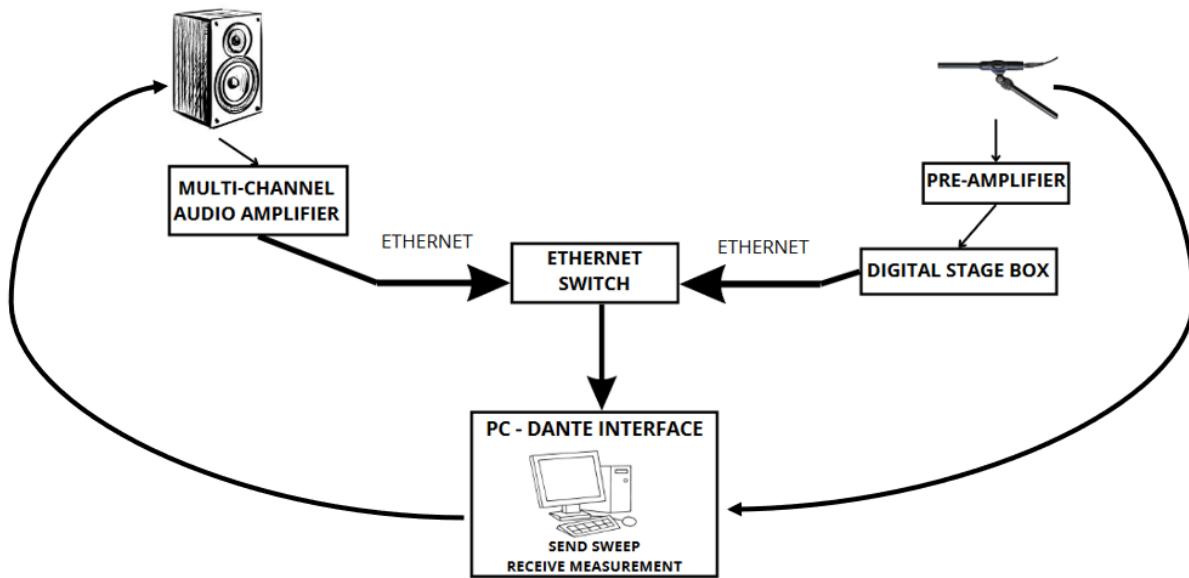


Figure 3.4: Measurement device setup.

The loudspeaker's crossover point, marking the transition between the tweeter and woofer, is set at 2500 Hz. The woofer operates within the frequency range starting from 40 Hz and a Butterworth filter with a slope of 12 dB per octave is applied.

The exponential sweep used to stimulate the loudspeaker lasts for 3 seconds plus 1 seconds

of silence, spanning the frequency range of 22-22000 Hz and includes both a fade-in and a fade-out.

As an initial step, the inverse of the sine sweep is obtained through a Kirkeby inversion [32], as described in equation (1.23). The loudspeaker impulse response can now be derived by convolving the recorded sine sweep with the inverse of the initial sine sweep. Then, The Kirkeby inverse method has been employed once more to derive the inverse of this impulse response.

By convolving the inverse of the impulse response with the initial sweep, a pre-distorted sweep signal is extracted.

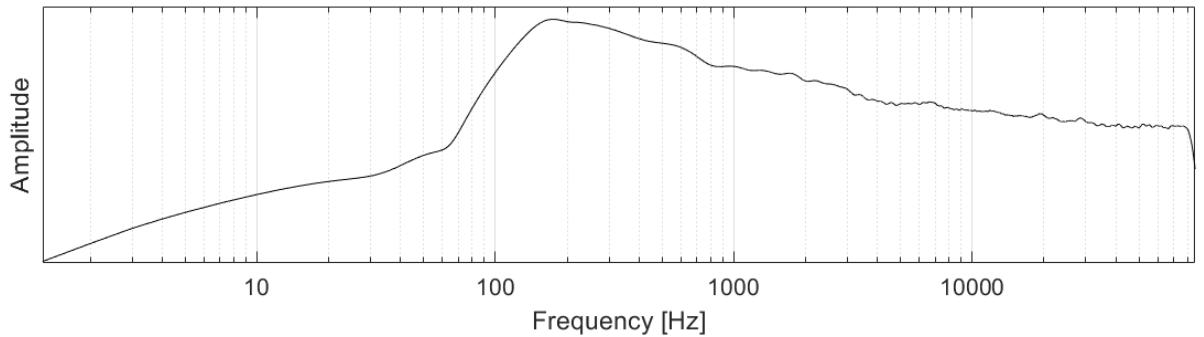


Figure 3.5: Spectrum of the computed pre-distorted exponential sine sweep.

In the following image the response spectrums are shown with no calibration performed in order to only depict the frequency dependent trend, therefore the amplitude axes will not consider any unit of measurement.

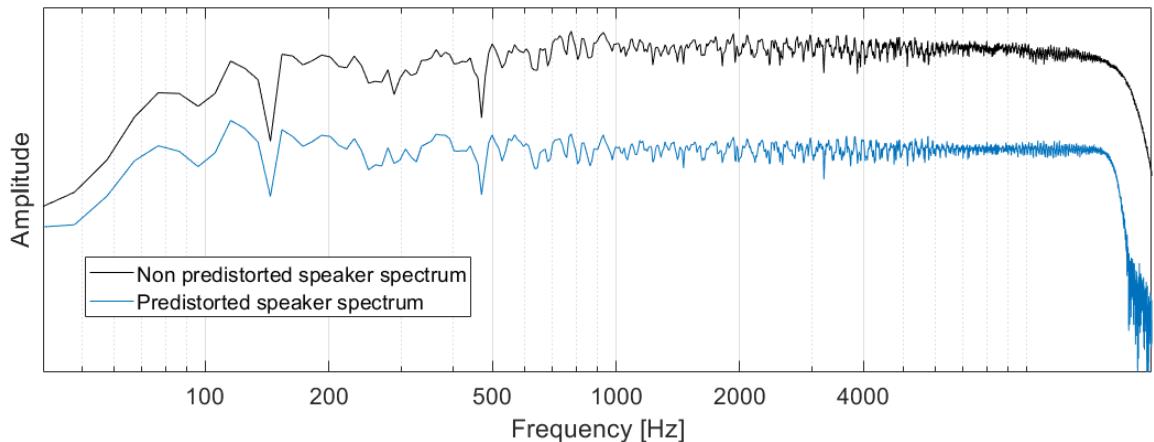


Figure 3.6: Comparison between the spectrum of the original impulse response with the pre-distorted one.

This signal guarantees a flat response when used to stimulate the loudspeaker. Although the loudspeaker already exhibits a relatively flat response, the pre-distorted sweep ensures a balanced stimulation that is, in this case, necessary.

The analysis done for one loudspeaker is considered applicable for each one of the other 24 present in the virtual room being all the speakers the same brand and model.

Sweep Measurement Procedures

The same omnidirectional microphone employed in the previously procedure is utilized for the measurement inside the Virtual Room.

For each set-up the microphone was positioned at the driver's location at ear height.

The objective is to analyze the spectrum of the impulse response achieved by linearly summing all the individual impulse responses generated through the sequential excitation of each loudspeaker, therefore find the acoustical parameters: reverberation time, clarity and early decay time.

The ultimate objective is to propose potential solutions for optimizing the acoustic arrangement, establish an acoustical parameter-based description of the environment.

As done for the compensation of the speakers, the crossover frequency was configured at 2500 Hz and woofer operates starting from 40 Hz, utilizing a Butterworth filter with a slope of 12 dB per octave.

The pre-distorted sweep was used for all the following measurements.

The responses obtained from the individual excitation of each loudspeaker are initially aligned in time domain to account for any potential irregularities in their placement and are then summed together.

For improved spectrum visualization, the *smoothSpectrum function* provided by the Matlab toolbox IoSR was employed. This function mitigates the spectral fluctuations by applying a $\frac{1}{6}$ -octave smoothing using a Gaussian window with a standard deviation of $f/6$, where f represent the central frequency.

The headrest has been removed in order to avoid sound absorption or reflections and any potential obstruction of the microphone's reception.

All the responses have been accurately calibrated, and the magnitude of the spectrum has been adjusted to ensure the correct values.

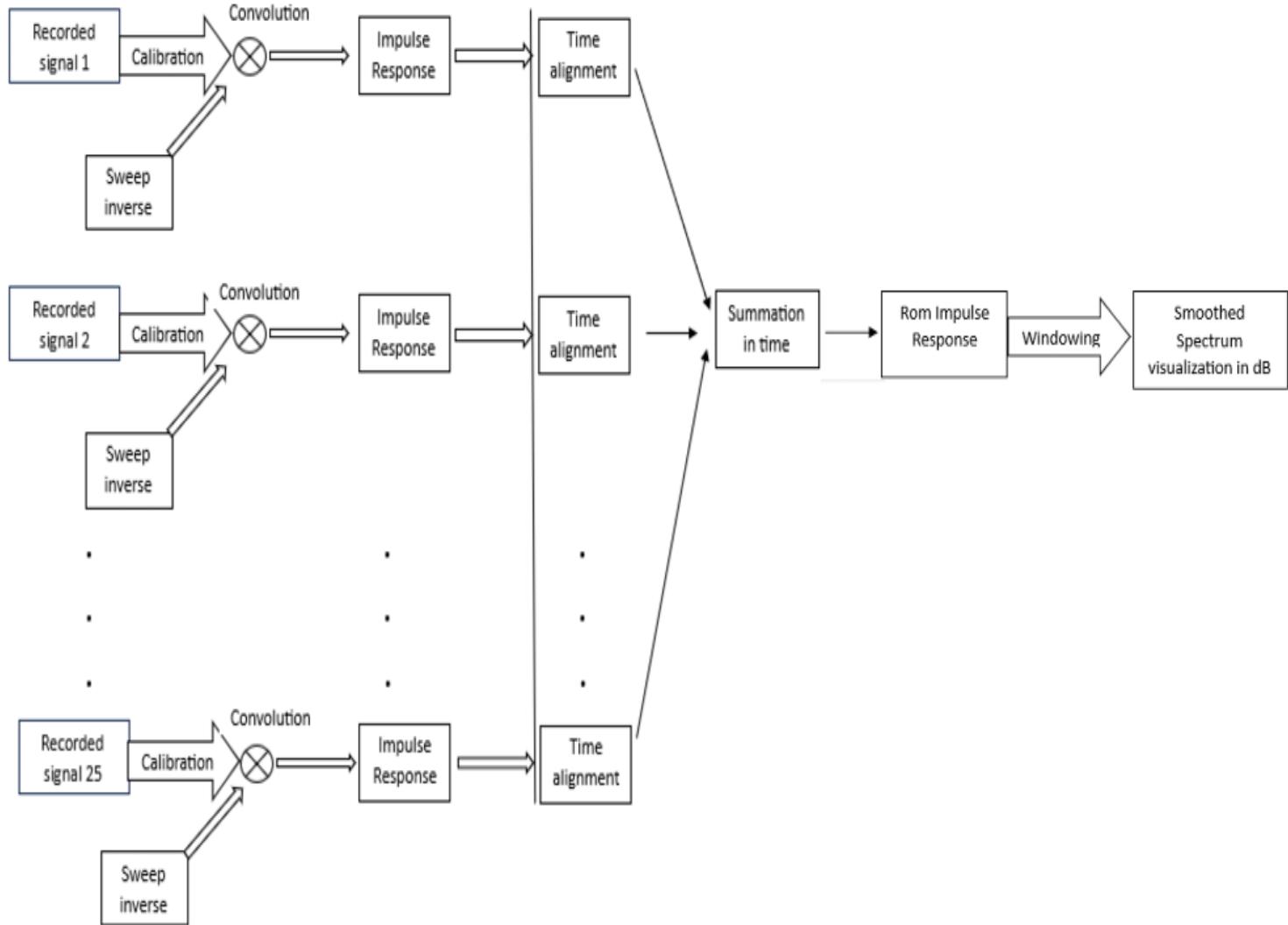


Figure 3.7: Procedure scheme for the analysis computation.

In the context of assessing room configurations for optimal functionality, various set-ups were experimented to determine the most effective arrangement. The goal is to identify and eliminate any reflections that could potentially impede the proper functionality of the system.

Five set-ups were tested utilizing polyester fiber panels 2.5 cm and 5 cm thick:

- Original configuration.
- Cover TV with 5 cm thick panel.
- Without TV.
- Without TV and door covered by 5 cm and 2.5 cm thick panel.

- Without the seat.

Subwoofers

During this analysis, two subwoofers are placed beneath the seat. A non-predistorted 15 seconds sine sweep, 10-10000 Hz, 0.2 fade in, was sent to the subwoofers for the examination of them inside the site. The microphone was positioned, as always, at the driver's seat.

It is pointed out that the subwoofers response will not be consider in the assessment of spherical harmonic reproduction, being its frequency response out of the frequency evaluation range.

Balloons Popping Measurements Procedure

In order to study the room itself without influences from the speakers, balloons were popped in 3 different positions using the two configurations mentioned in the previous section: with the TV and door present, and then covered.

Three positions of explosion were tested in order to evaluate the presence of resonances and reflection nearby three different points:

- Centered: 70 cm upon the microphone in driver position.
- Back: 1.20 m distant from the microphone in the door direction.
- Front: 1.40 m distant from the microphone, in the TV direction.

Considerations for the Acoustical Parameters Extrapolation

The signal under study was properly cut and analysed by using the toolbox provided by IoSR: iosr.acoustic.irStats [8]. The utilized function returns the necessary parameters by employing a method based on ISO 3382- 1:2009.

This involves using reverse cumulative trapezoidal integration to estimate the decay curve and a linear least-square fit to determine the slope between 0 dB and -20 dB for the T20 calculation.

In the following image the energy decay curve and its impulse response is shown for the 62.5 Hz octave band component extracted from the room impulse response:

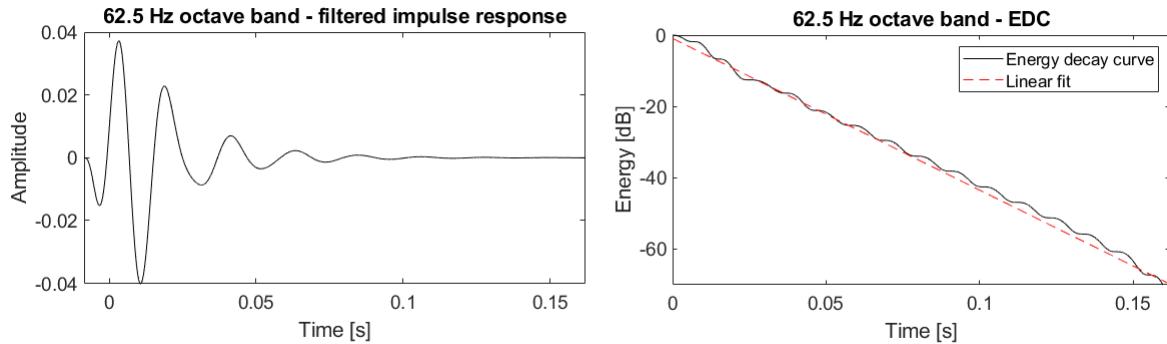


Figure 3.8: Example of filtered impulse response in 62.5 Hz octave band and its related energy decay curve.

Therefore, the signal has been appropriately truncated prior to analysis specifically, just before the onset of the direct sound and immediately after surpassing the noise threshold. Consequently, there is no need for a noise compensation algorithm, and the decay curve is accurately detected. The parameters extracted are:

- T20 (Reverberation Time)
- Speech clarity index: C_{50}
- Early Decay Time

3.2. Results

Sweep Measurements

In all the configurations assessed with the sweep signal, a decay of the magnitude of the spectrum around 1600 Hz was noted.

In the following image it is possible to see the spectrum of the various configurations tested. Speakers were activated individually, and the resulting impulse responses from each speaker were extrapolated, see 1.7. The obtained impulses were synchronized with each other to delete any potential time delays between speakers, which will be addressed in subsequent section, and summed all together.

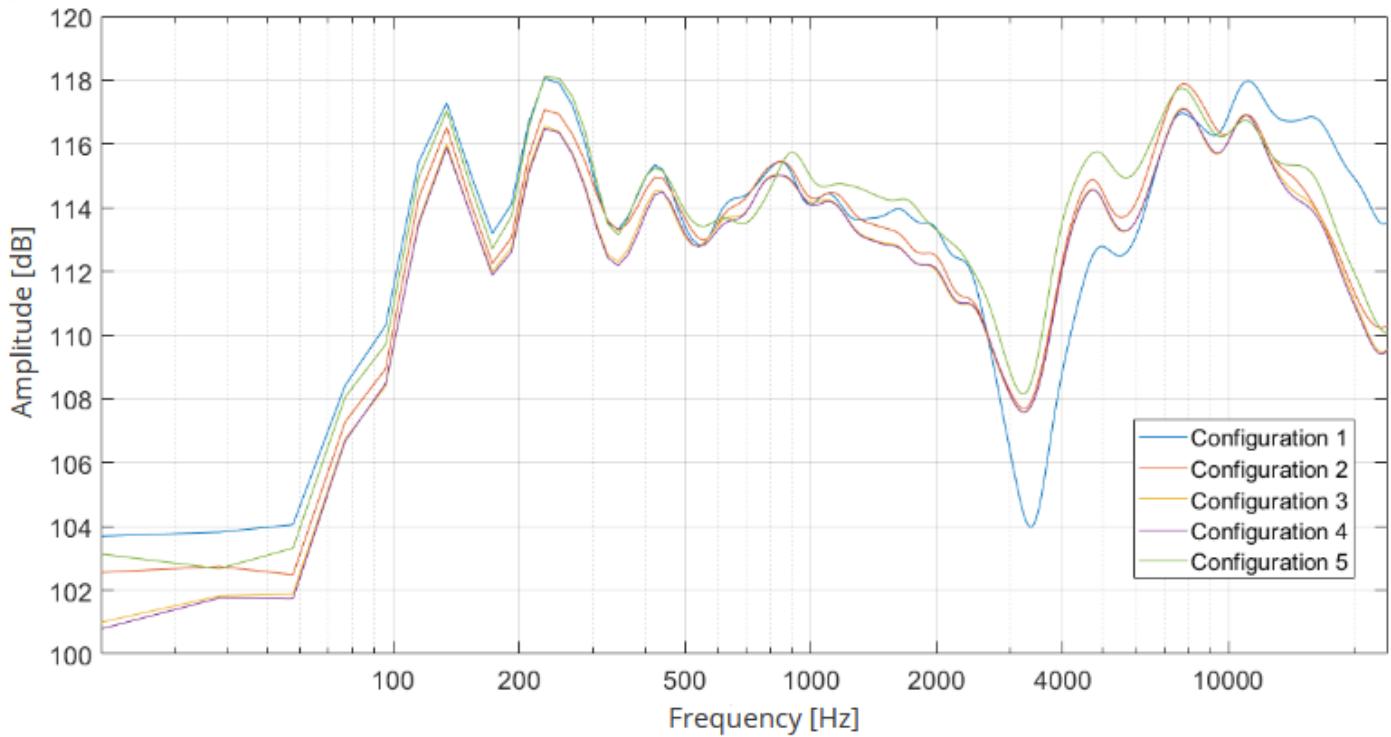


Figure 3.9: Spectra comparison. Configuration 1: original configuration. Configuration 2: TV covered. Configuration 3: TV removed, Configuration 4: TV removed and door covered. Configuration 5: seat removed.

Further analyses were conducted with the aim of examining the issue and identifying its root cause. Consequently, the direct responses from two speakers, identified as problematic in all configurations (speakers 7 and 9), were analyzed.

The microphone was placed at a distance of 80 cm, centered between the tweeter and woofer, testing both with pre-distortion and without, with and without a grille.

The analyses were further deepened using Smaart, 1.12, capable of displaying impulse response and its real-time spectrum.

It was deduced that the issue persisted even in the case of direct sound, ruling out a specific room reflection, or a mistake in the pre-distortion, as the cause. Since the response of speaker n. 19 was correct in the 1650 Hz zone, it was observed that the difference between this speaker and the others lay in the fact that the tweeter of speaker 19 was the only one with a non-inverted phase. The phases of the tweeters of the other speakers had been inverted in a previous use of the room.

The decay was not due, therefore, to a reflection or absorption of the specific frequency but to a misconfiguration of the tweeter phases due to previous use of the room.

Once the phases of all speakers have been corrected, the following final results were obtained, by testing the pre-distorted sweep with the TV and door presence and then their covering through an absorbing polyester fiber panel 5 cm thick for the TV and 2,5 cm thick for the door.

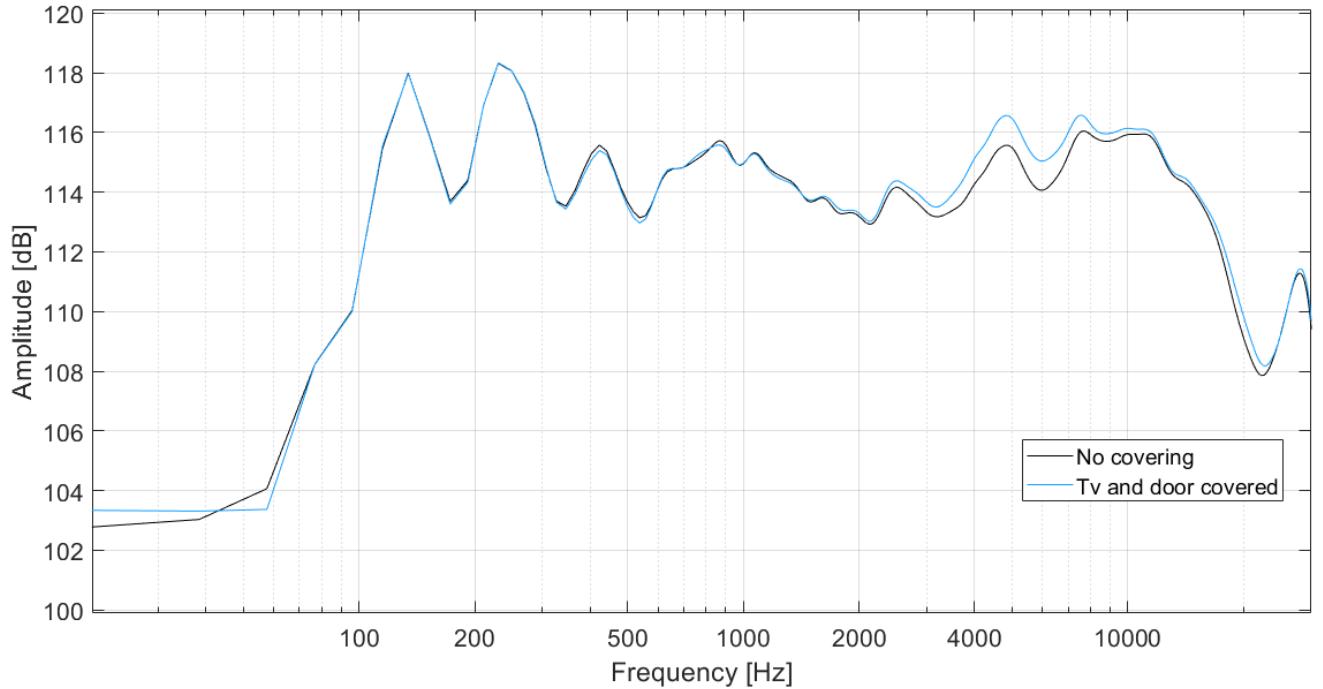


Figure 3.10: Spectra comparison. In blue: original configuration. In orange: TV and door covered by absorbing polyester fiber panel 5 cm thick for the TV and 2,5 cm thick for the door.

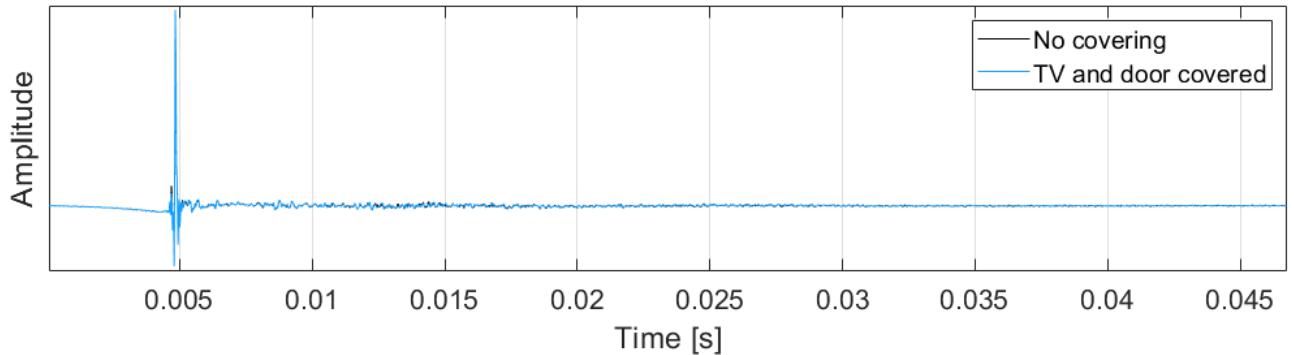


Figure 3.11: Time trends comparison. In black: original configuration. In blue: TV and door covered by absorbing polyester fiber panel 5 cm thick for the TV and 2,5 cm thick for the door.

The resulting response is much more linear in frequency. A gain was applied to loudspeaker n. 13 of +4 dB and to loudspeaker n. 25 of -4 dB to both tweeter and woofer to compensate loss of balance in the volume, the whole equalization will be done in subsequent section.

The previous problem is completely fixed and it can be observed that the two configurations do not lead to great differences in the time (the two impulse responses in time domain overlap in figure 3.11) and frequency responses of the room.

The following image illustrates the spectra of individual impulses for each speaker in the original configuration.

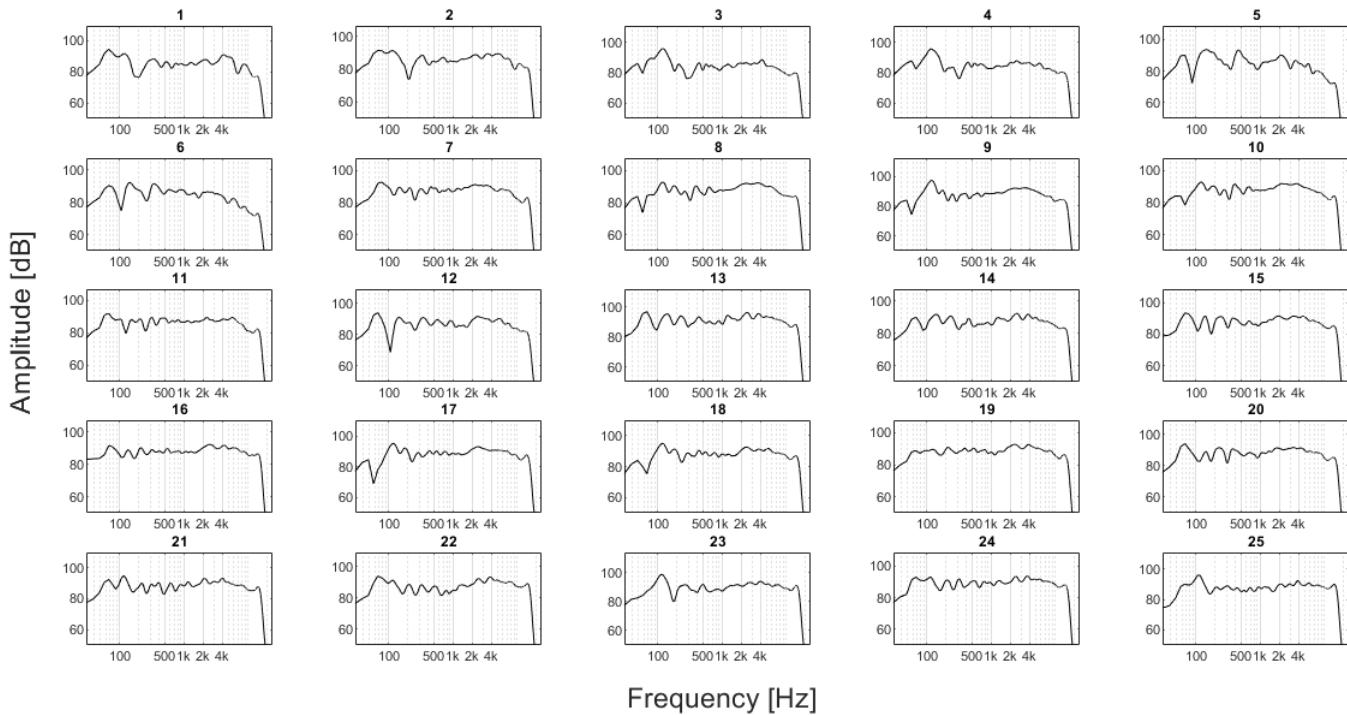


Figure 3.12: Spectra of all the single impulse responses with the first configuration (original one). Above each image, the number of the related excited speaker is depicted.

Subwoofers

The impulse response generated by the two subwoofers, obtained once again from the sum of the single responses aligned together, is depicted below:

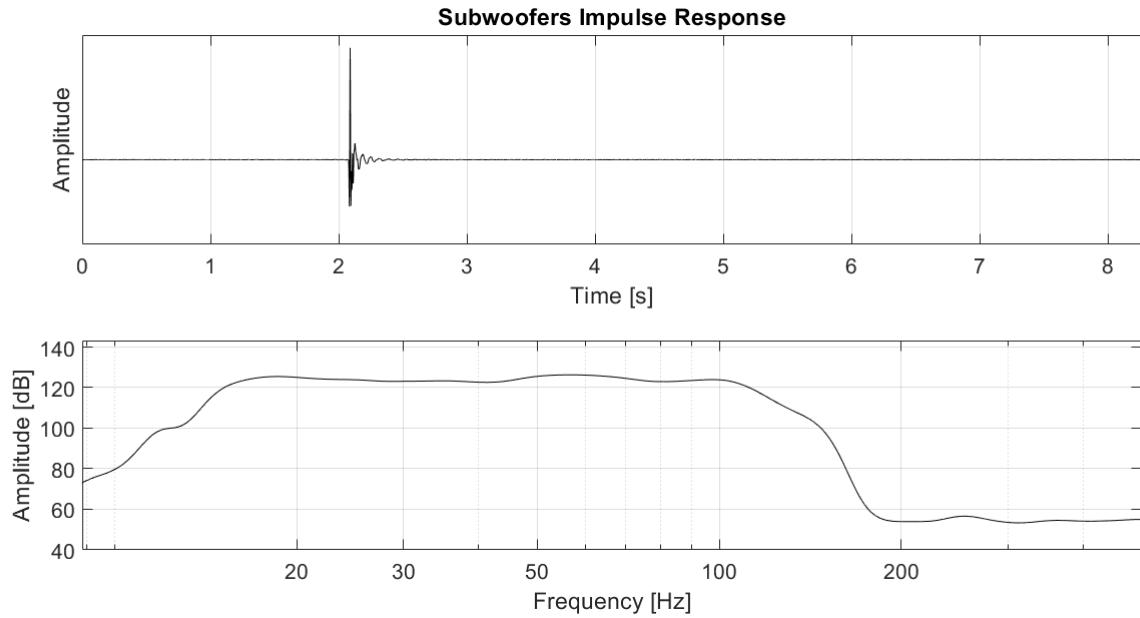


Figure 3.13: Impulse obtained by linearly summing the individual subwoofers response. Time domain and spectrum.

The spectrum exhibits adequately flat frequency behaviour within the desired range. Consequently, there is no need for a pre-distortion or any tuning to address any irregularities.

Balloon Measurements

The following image depicts the room impulse response generated by the balloon popping:

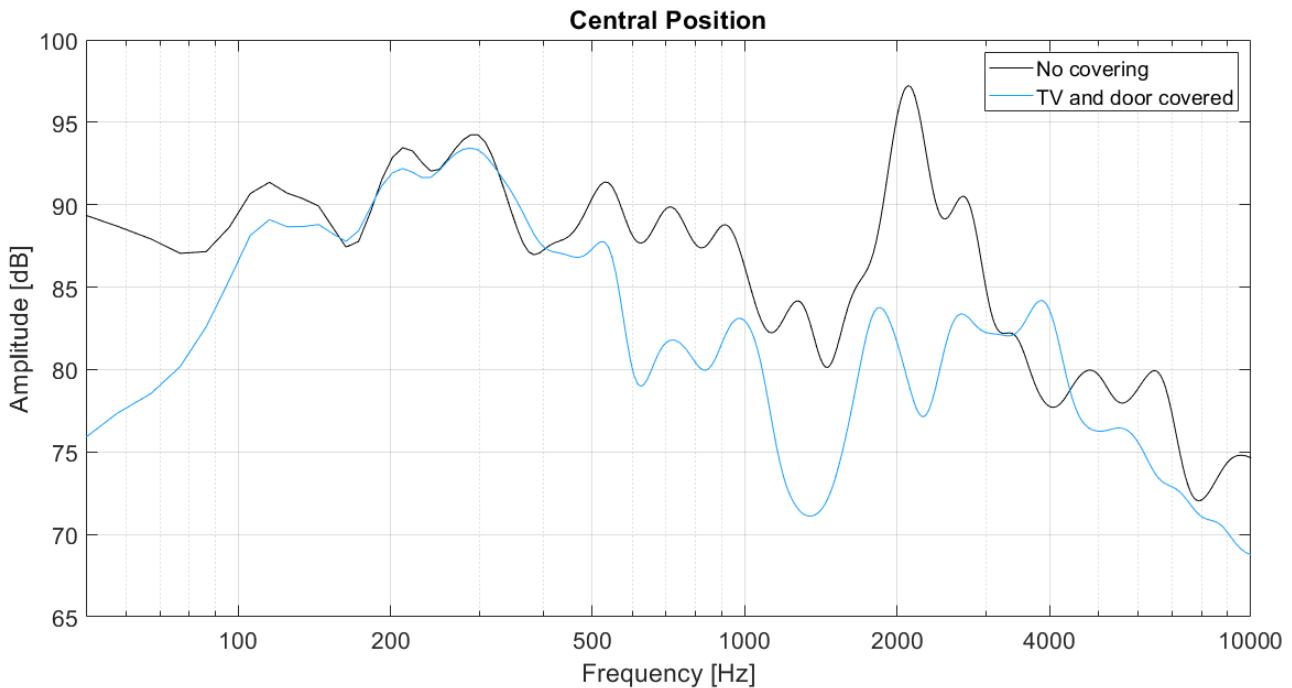


Figure 3.14: Balloon popping spectra measured from the central position.

In figure 3.14, the emphasis should be directed towards comparing the two configurations rather than delving into the details of the frequency trends, in fact, the spectrum of the balloon explosion itself may not exhibit a flat profile and the analysis of the measured spectrum may lead to erroneous conclusions.

Upon bursting the balloon, a sharp and bright sound emanating from the ceiling was distinctly noticeable. This effect can be attributed to the absence of any covering, allowing for additional reflections over time:

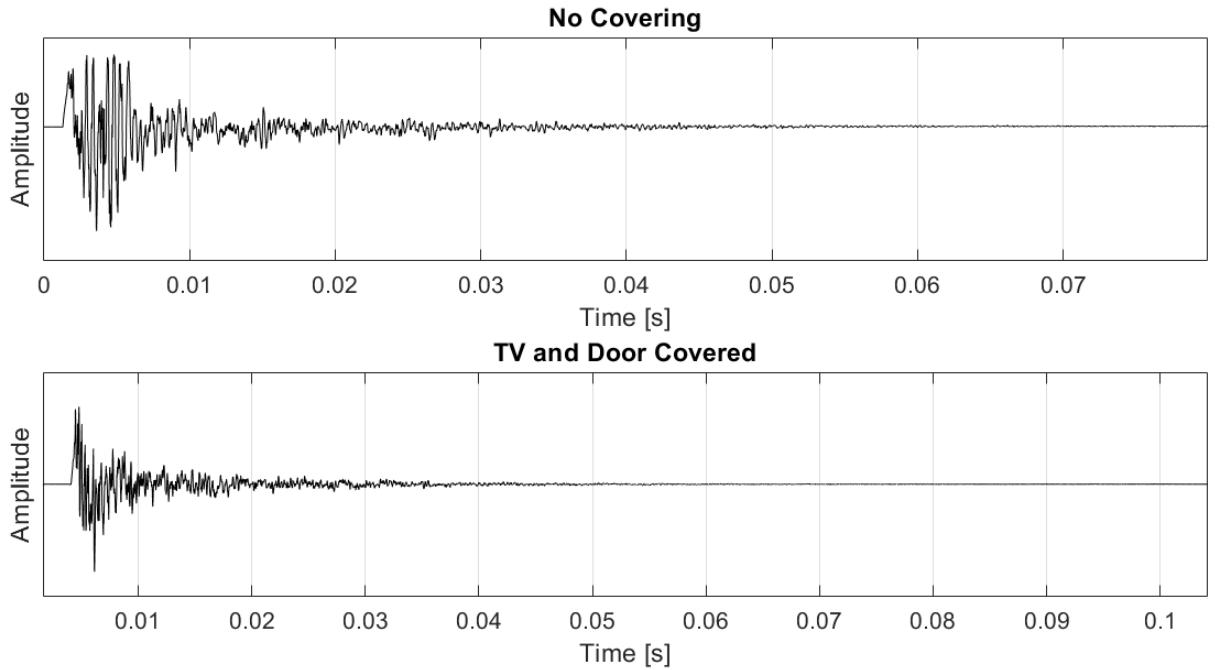


Figure 3.15: Balloon popping time trend measured from the central position.

These reflections can be traced back to the metal structure supporting the upper loudspeaker, potentially causing the peak just above 2000 Hz in the absence of any covering. Therefore, it is plausible that the panel above the TV is responsible for absorbing these reflections.

Acoustical Parameters

In the following bar graphs the detected acoustic parameters are displayed. The values are displayed in octave bands: the octave-band filters are calculated according to ANSI S1.1-1986 and IEC standards [16].

The parameters obtained through sweep excitation and through balloon excitation are compared, for each configuration, in order to have a more accurate and completed feedback of the acoustic performances.

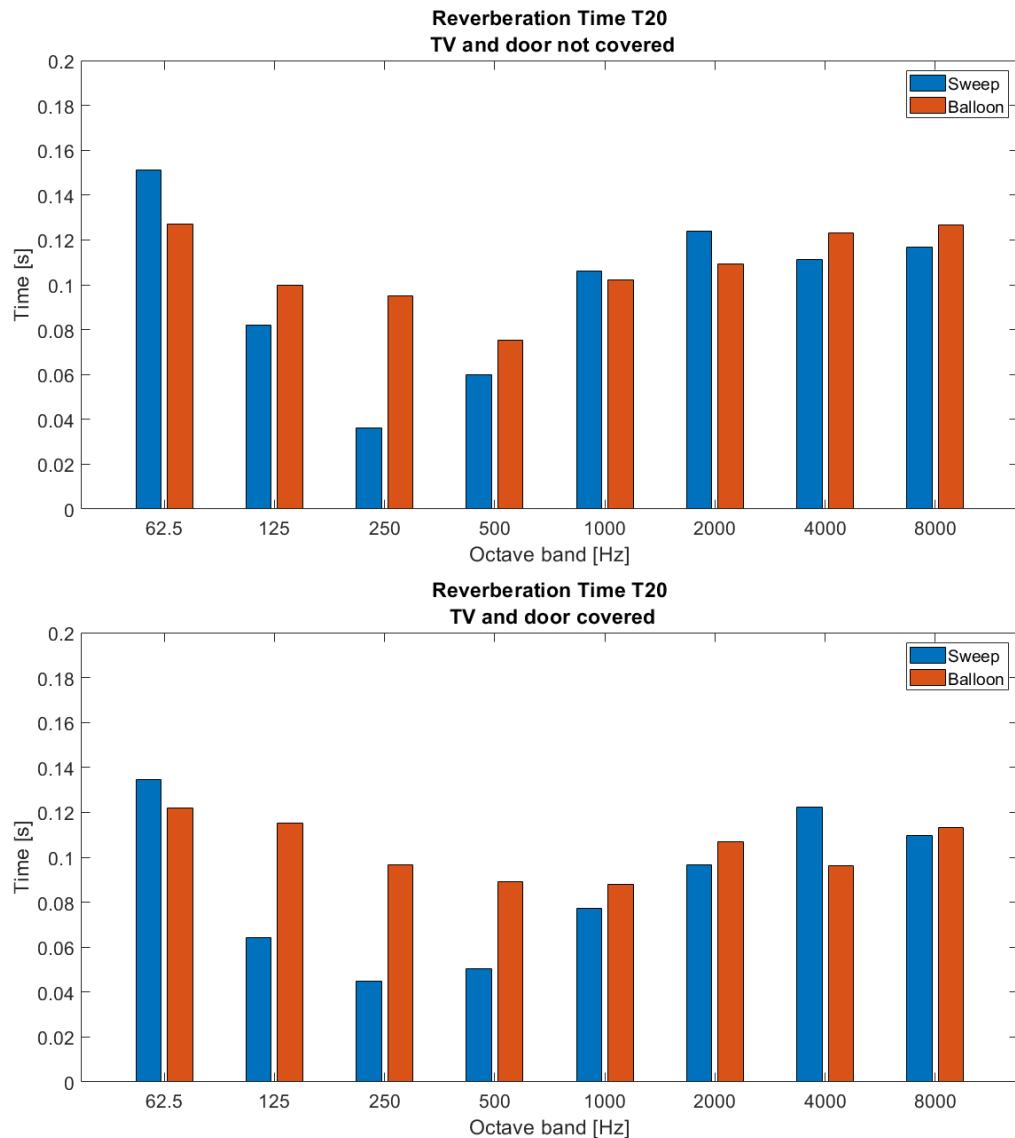


Figure 3.16: Center position acoustic results: T20

It is possible to observe that the T20 values are optimal even without the absorbing cover.

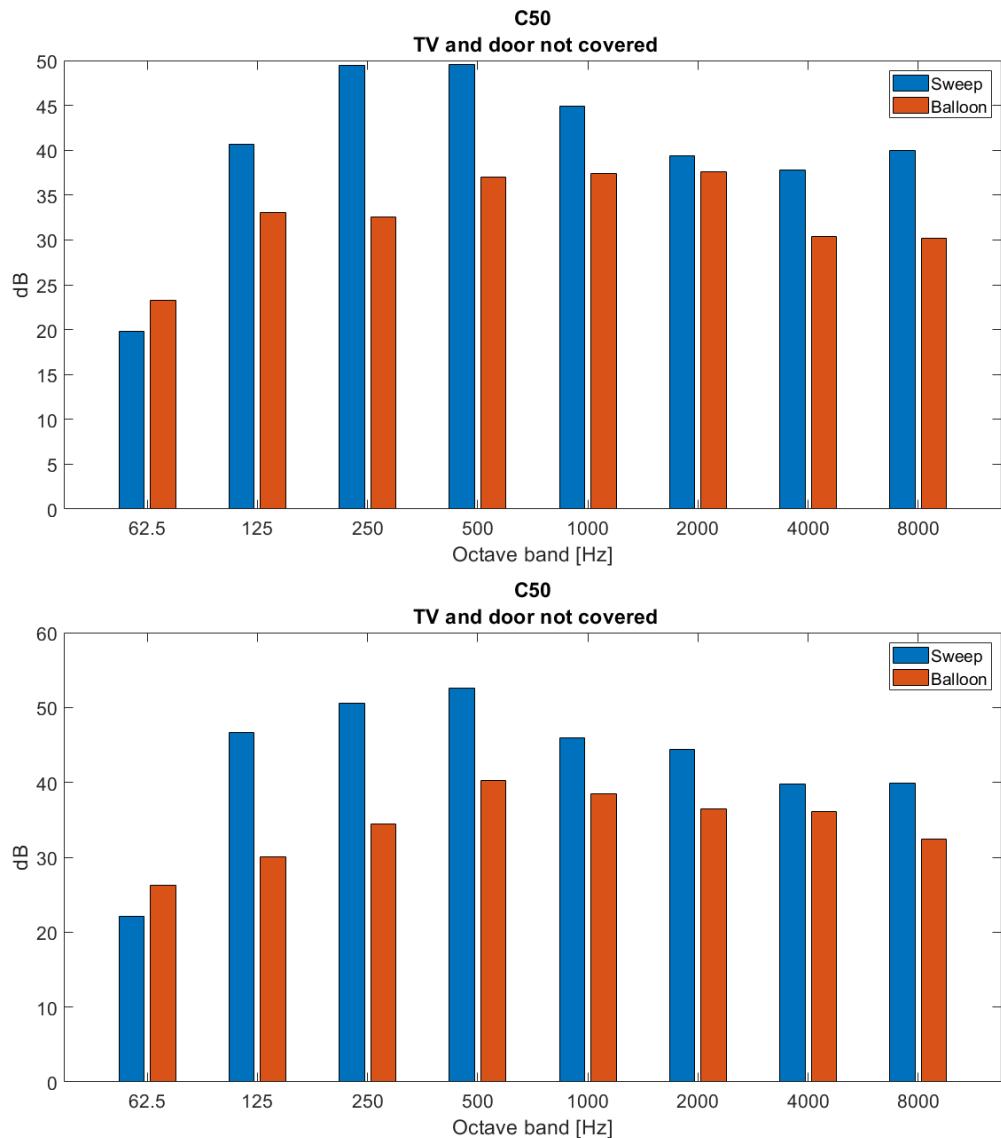


Figure 3.17: Center position acoustic results: C50

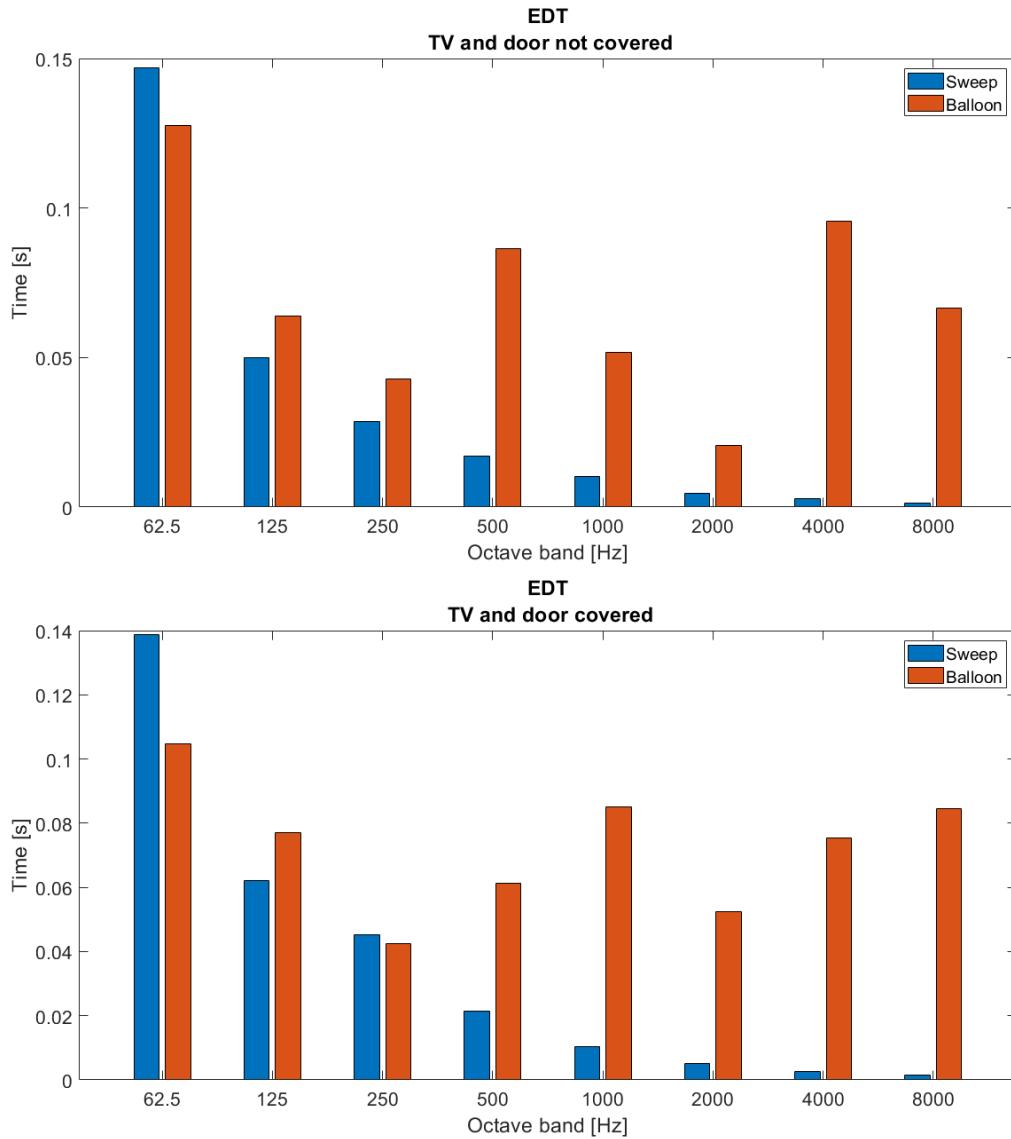


Figure 3.18: Center position acoustic results: Early Decay Time

Focusing on the comparison between cover set up and not cover one in the central position the conclusive observation can be effectuated: the optimal acoustic parameters are effectively achieved in both situations. Therefore, the decision is not to alter any room predisposition and leave it as is.

In this table the resulting values for 31 Hz octave band are presented when the room in its original configuration (with no covering applied) is excited by the sweep from the subwoofers:

T20	C50	EDT
0.2 s	10 dB	0.34 s

Table 3.1: Acoustic parameters obtained through subwoofers excitation for 31 Hz octave band.

The room has been confirmed to possess heightened semi-anechoic characteristics, distinguished by exceptional clarity and minimal reverberation. This is guaranteed by efficient sound absorption achieved through a well-designed ambient projection.

4 | Listening Room Performance Measurements

4.1. Introduction and Methodologies

The overall goal of this chapter analysis is to obtain two parameters serving as indicators of error and quality in the representation of the sound field using spherical harmonic components: Level Difference and Spatial Correlation.

The parameters, reported in 2.2, are used in literature for the assessment of the encoding process, therefore for the evaluation of the reconstruction of the spherical harmonic that a proper microphone can guarantee.

The quality of the recreation in a physical space of sound environments is usually evaluated through subjective tests on localization and perceive sound characteristics. Objective metrics exist for speech ([12]) and audio quality ([13]) but no objective metrics are agreed upon for spatial audio quality evaluation.

This thesis proposes, at this point of the analysis, a method of evaluating spherical harmonics as representations of the diffuse sound field recreated within the room.

A diffuse sound field can be represented as a combination of all spherical harmonics, being the key characteristic of a diffuse field the uniformity of the acoustic energy distribution. In terms of spherical harmonics, this means that all of them should be present approximately uniformly in the field, in particular the first ones, whose ensure the main contribute of the omnidirectional component.

However, in a real-world situation, the distribution of spherical harmonics could be influenced by the geometry of the sound sources, ambient reflections and so on.

The proposed metrics aim to assess, therefore, the spatial configuration, which includes the playback array and the site structure, as a room for the recreation of a diffuse sound space. Reflections, limited speaker configuration, system failures can lead to inadequate mapping of spherical harmonics.

The microphone used for these measurements, Eigenmike64®, is to be regarded as ideal

in the frequency ranges described in 2.3.

Measurements Set Up

Two positions were tested in order to conduct the evaluation even outside the center of the *sweet spot*:

- Center position, at ear's height.
- 40 centimeters left of center position, at ear's height.



Figure 4.1: The Eigenmike64[®] in central position.

Setting details are listed below:

- An exponential sine sweep 10 seconds plus 2 silence seconds long was employed.
- Crossover point between woofer and tweeter was set to 2500 Hz
- A high-pass filter at 80 Hz was applied to the woofer, filtering out low frequencies.
- The subwoofers, as previously mentioned, were not employed in this evaluation.
- The entire spherical woofer - tweeter array was employed, exciting the speakers one at a time.
- The headrest has been removed to limit close-up obstacles.

Extraction of the Measured Spherical Harmonic Coefficients and Ideal Spherical Harmonic Coefficients

The given encoding filters of the Eigenmike[®], specifically crafted to retrieve the spherical harmonic coefficients, are stored in the encoding matrix \mathbf{H} . This filter matrix comprises 64 filters for each spherical harmonic, in the case of the 64-capsule Eigenmike, which extends up to the 6th order and encompassing a total of 49 spherical harmonics.

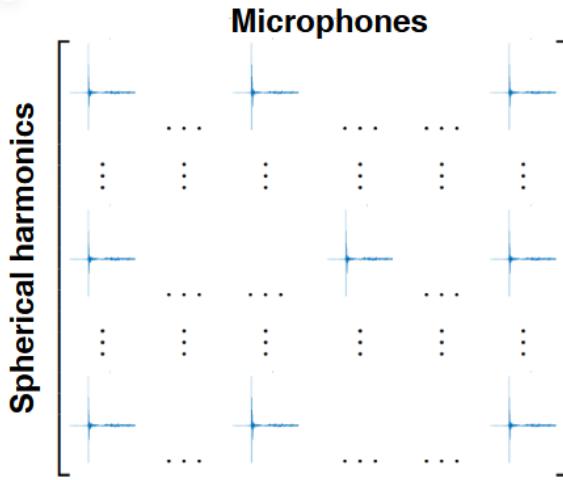


Figure 4.2: Filter matrix \mathbf{H} .

$$\mathbf{H} = 49 \times 64 \text{ signals}$$

Through the convolution of each signal, generated via the individual speaker sweeps, with the inverse of the excitation sweep, the impulse response matrix was recorded and stored as matrix S .

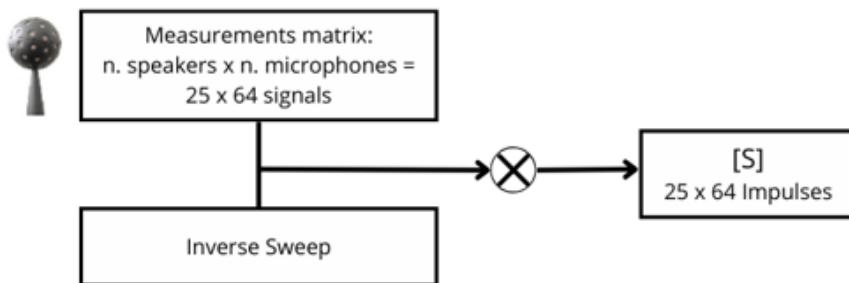


Figure 4.3: Impulse Response matrix extraction.

Each impulse response is convoluted with its proper filter in frequency domain and the summation for each microphone in time domain is performed, as depicted in [67] page 2 and in [34].

$$\hat{Y}_{nm}^{\omega}(\theta_l, \phi_l) = \sum_{j=1}^J S_j^{\omega}(\theta_l, \phi_l) \cdot H_{nm,j}^{\omega} , \quad (4.1)$$

here ω denotes the frequency, (θ_l, ϕ_l) the source $l \leq L$ origin direction and j the capsule number, up to the amount of capsules J .

The dimension of the obtained matrix is, therefore:

$$\hat{\mathbf{Y}}(\omega) = \begin{bmatrix} \hat{Y}_{0,0}^{\omega}(\theta_1\phi_1) & \hat{Y}_{1,-1}^{\omega}(\theta_1\phi_1) & \hat{Y}_{1,0}^{\omega}(\theta_1\phi_1) & \dots & \hat{Y}_{NM}^{\omega}(\theta_1\phi_1) \\ \hat{Y}_{0,0}^{\omega}(\theta_2\phi_2) & \hat{Y}_{1,-1}^{\omega}(\theta_2\phi_2) & \hat{Y}_{1,0}^{\omega}(\theta_2\phi_2) & \dots & \hat{Y}_{NM}^{\omega}(\theta_2\phi_2) \\ \vdots & \vdots & \ddots & & \vdots \\ \hat{Y}_{0,0}^{\omega}(\theta_L\phi_L) & \hat{Y}_{1,-1}^{\omega}(\theta_L\phi_L) & \hat{Y}_{1,0}^{\omega}(\theta_L\phi_L) & \dots & \hat{Y}_{NM}^{\omega}(\theta_L\phi_L) \end{bmatrix} ,$$

$$\hat{\mathbf{Y}}(\omega) = 25 \times 49 \text{ spherical signals} .$$

Considering 25 sound sources directions and 49 spherical harmonics.

In an ideal scenario, implying a perfect reconstruction, the $\hat{\mathbf{Y}}(\omega)$ matrix would exhibit frequency independence, resulting in $\hat{\mathbf{Y}}(\omega) = \mathbf{Y}$ for all directions d across all frequencies.

The Matlab Toolbox 'Spherical Harmonic Transform Library', [12], developed by Archontis Politis that performs the complex ideal spherical harmonic coefficients calculation was utilized, the already seen equation 1.5 is again depicted for convenience:

$$Y_{nm}(\theta_l, \phi_l) = \sqrt{\frac{(2n+1)}{2} \frac{(n-m)!}{(n+m)!}} P_{nm} \cos(\theta_l) e^{im\phi_l} ,$$

the dimension of the obtained frequency independent matrix, for 25 directions and 6 orders is, therefore:

$$\mathbf{Y} = 25 \times 49 .$$

And, again, θ and ϕ are the angles of each direction l , respectively azimuth and elevation; m is the degree of the spherical harmonic while n is the order of it in the range $[-m \leq n \leq +m]$. P_{nm} are the associated Legendre polynomials.

4.2. Room Equalization

Two equalisations were carried out, one for the central listening position and one for the position shifted to the left by 40 cm.

The aim is to assess whether and how much the measured spherical harmonics improve qualitatively with the original system in the absence of tuning and with appropriate tuning.

The first spherical harmonic, therefore the omnidirectional component, for each source direction was analyzed for the tuning parameters extraction: $\hat{Y}_{0,0}^{\omega}(\theta_l, \phi_l)$.

Time Alignment

As explained in 1.4, the time alignment of each speaker is an important factor that influences the quality of the surround system and source localization. Therefore, in light of this, it is essential to align the speakers in time during the procedure of room tuning. It is noted that each speaker's tweeter and woofer appear to be already aligned.

The remaining step is to align the speakers for a the listening position within the *sweet spot*.

The ADT values between the most distant speaker from each measurement position and the remaining speakers are saved and input into the system provided by the amplifier for adjustment.

Speaker number	Sample delay to apply	Speaker number	Sample delay to apply
1	6	1	48
2	0	2	0
3	10	3	2
4	13	4	71
5	41	5	62
6	44	6	28
7	47	7	18
8	61	8	46
9	73	9	89
10	60	10	120
11	47	11	122
12	47	12	104
13	39	13	67
14	38	14	29
15	55	15	19
16	57	16	26
17	61	17	53
18	53	18	94
19	66	19	128
20	66	20	121
21	57	21	58
22	67	22	39
23	77	23	80
24	70	24	113
25	119	25	115

Table 4.1: Delay table, central position.

Table 4.2: Delay table, lateral position.

Frequency Response Magnitude Equalization

Specific filters FIR (Finite Impulse Response) have been applied to each speaker in order to equalise the spectrum of the reproduction system in that exact spatial configuration with respect to the precise listening position. REW Equalizer Tool 1.13 was employed for the FIR extraction, the procedure details are listed below:

- Generic Equalize provided by REW was employed.
- Target type set to full range speaker. Hence, a single FIR was created for each

speaker and individually sent to the relative woofer and tweeter.

- Low frequency cut-off set to 80 Hz (consistently with the cross-over applied) with a slope of 12 dB/oct
- Target Levels were set equals for all the FIRs
- Match range: 20 - 20k Hz
- Individual boost for each filter composing the FIR was set to 6 dB
- Overall max boosts for each FIR were set to 0 dB

The Generic Equaliser supports a full range of filters and filter settings. For the peaking filters the bandwidth in Hz between the half gain points is given by:

$$\text{Bandwidth} = \frac{\text{centre frequency}}{Q}$$

The Generic setting allows 20 parametric filters. The adjustment ranges are and other details can be found in [6].

The following example is the obtained FIR for the speaker number 6 in By measuring by placing the microphone array 40 cm left.

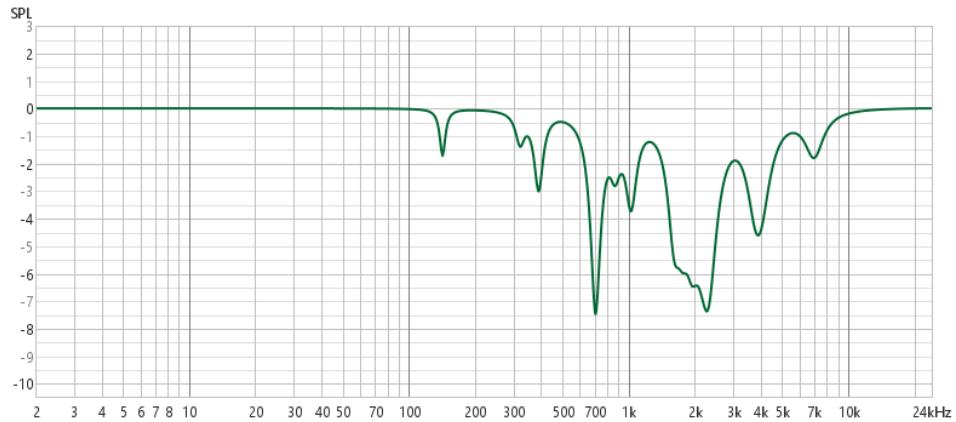


Figure 4.4: FIR obtained for the speaker number 6 measuring in 40 cm left position from central point.

The filters composing the depicted FIR are:

Type	Central Frequency (Hz)	Gain (dB)	Q
PK	142.5	-1.70	14.141
PK	321.0	-1.10	8.766
PK	390.0	-2.80	8.791
PK	708.0	-7.10	7.738
PK	869.0	-1.50	7.364
PK	1028	-3.00	7.088
PK	1611	-3.30	6.267
PK	1763	-2.20	6.106
PK	1947	-2.70	5.821
PK	2144	-1.20	5.026
PK	2313	-5.40	4.471
PK	3762	-1.60	4.585
PK	3994	-2.90	3.782
PK	6994	-1.60	3.650

Table 4.3: FIR parameters for speaker 6 measuring in 40 cm left position from central point.

4.3. Spatial Correlation and Level Difference Results

It is noticed that the magnitudes of the values to be compared are out of coherence due to gain applied and their incompatibility ranges. Therefore a normalization is applied to Y and \hat{Y} in order to map the values in the complex unit sphere and compare them properly:

$$Y_{norm} = \frac{Y}{\max(Y)} \quad , \quad (4.2)$$

$$\hat{Y}_{norm} = \frac{\hat{Y}}{\max(\hat{Y})} \quad . \quad (4.3)$$

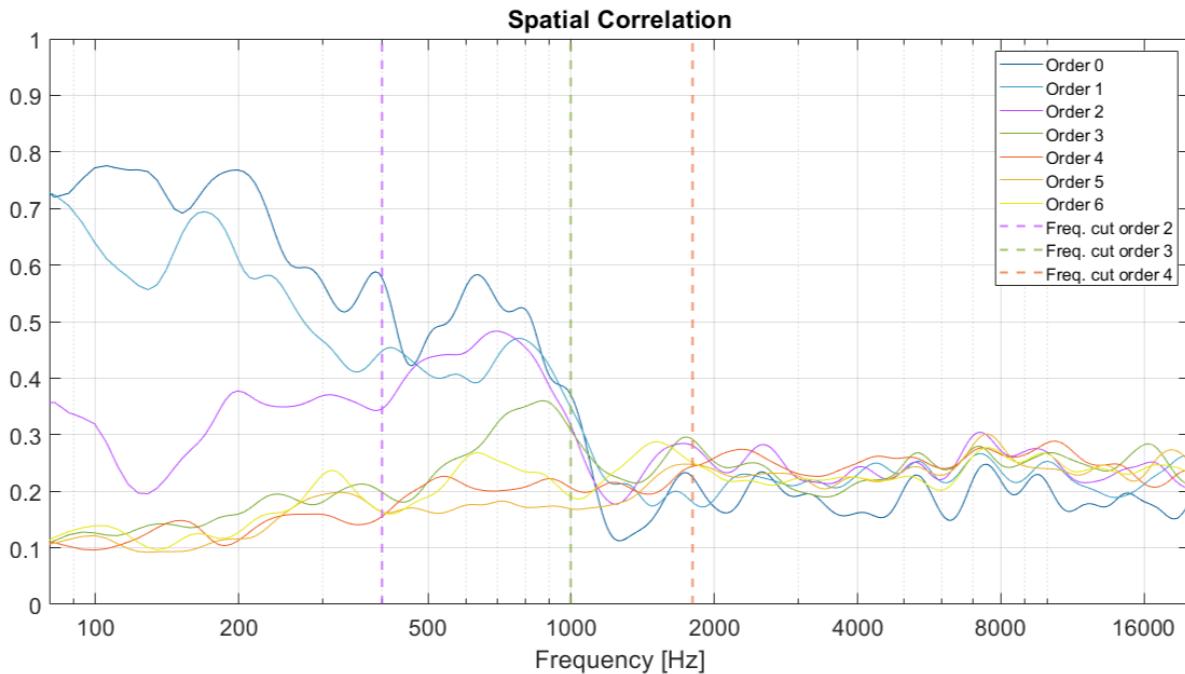
The terms Y_{norm} and \hat{Y}_{norm} are, therefore substitutes for Y and \hat{Y} respectively in formulas 2.1 and 2.3.

Spatial Correlation is depicted, assessing the similarity between the generated patterns and the ideal spherical harmonic patterns.

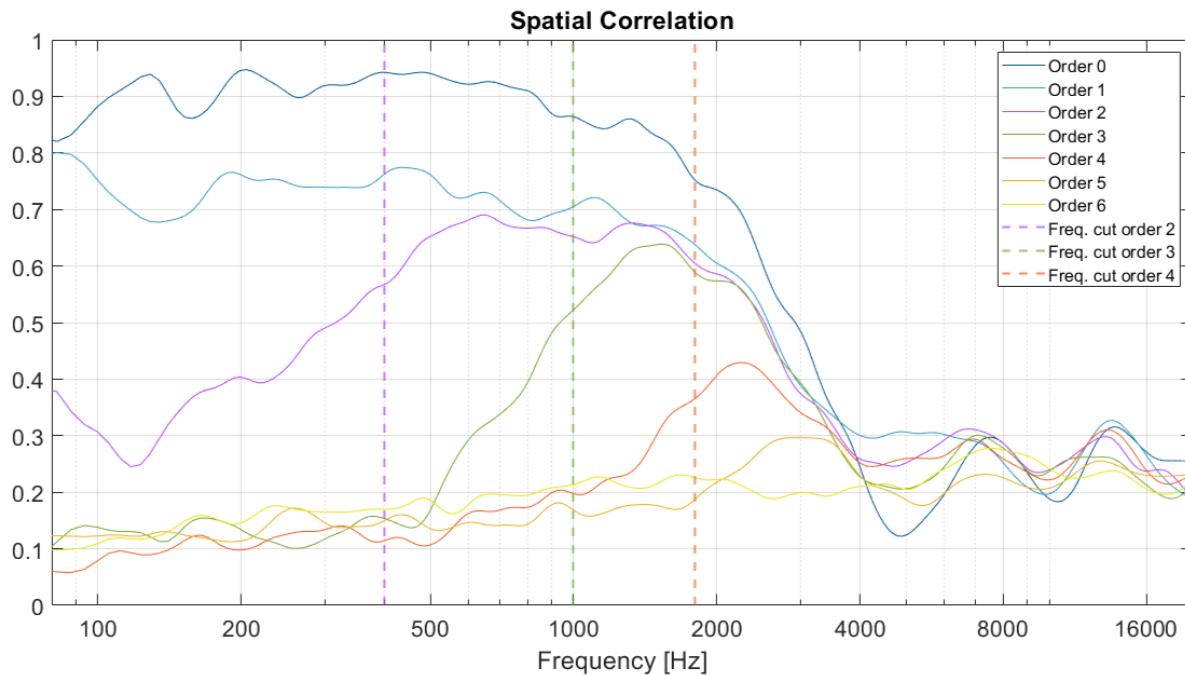
Each order's behavior up to the sixth one can be observed; however, it is important to note that the Virtual Room is designed to provide sound reconstruction only up to the

fourth order.

The vertical lines highlights the cut off frequencies of the microphone array, as described in 2.3, therefore, for that particular order, the trend should be evaluated for frequencies higher than the cut off.

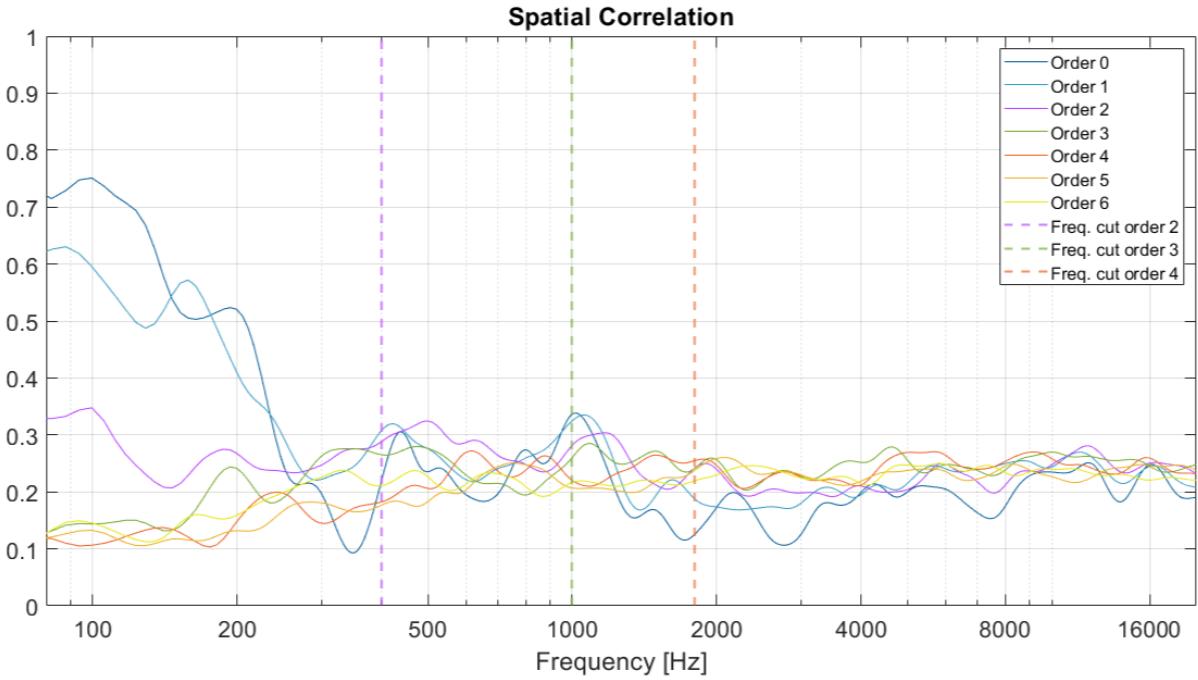


(a) No equalization applied.

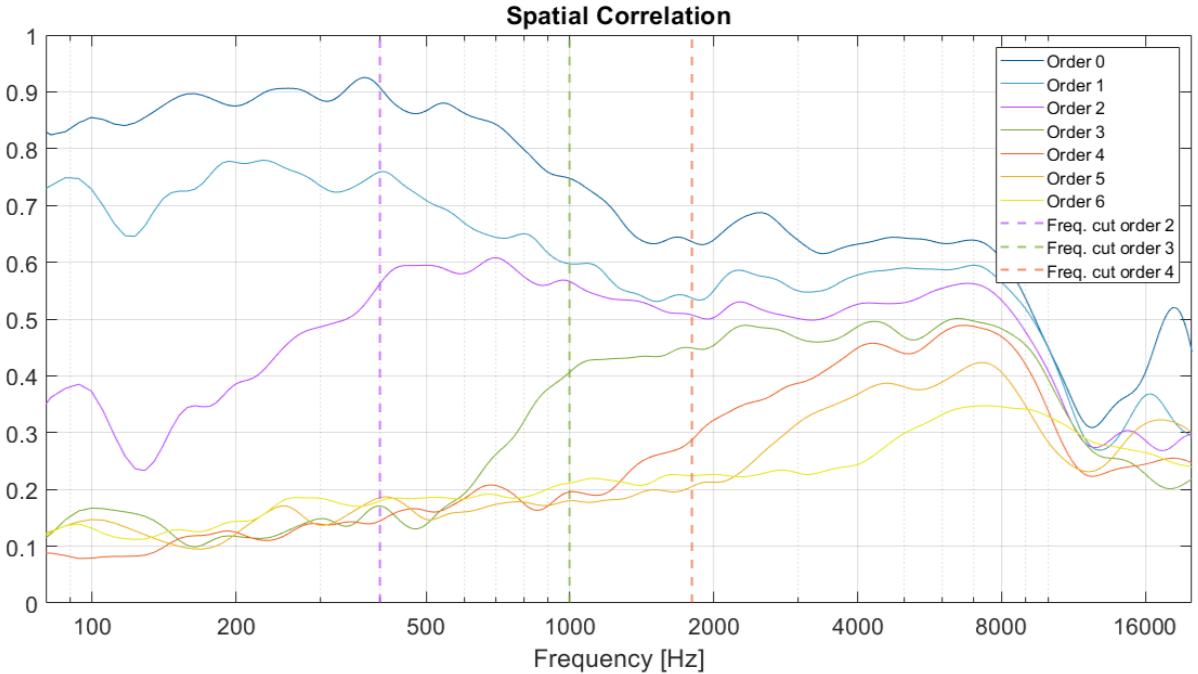


(b) Proper equalization applied.

Figure 4.5: Spatial Correlation for each order measuring in central position. A 1/6 octave band smoothing is applied.



(a) No equalization applied.



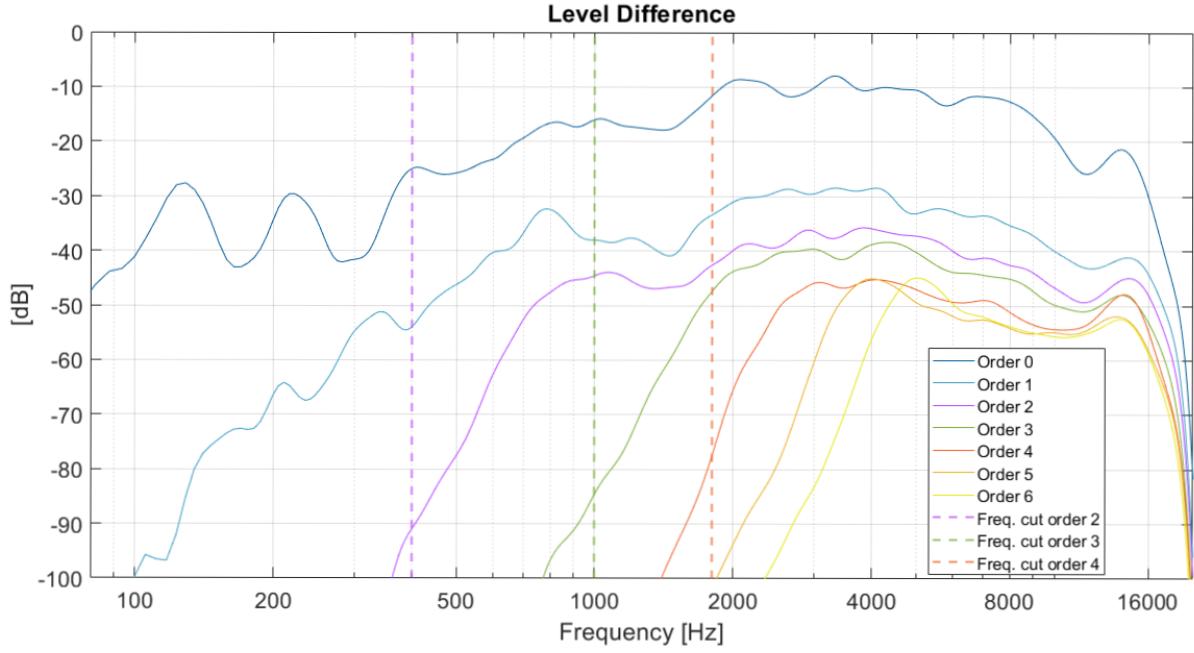
(b) Proper equalization applied.

Figure 4.6: Spatial Correlation for each order measuring in lateral position. A 1/6 octave band smoothing is applied.

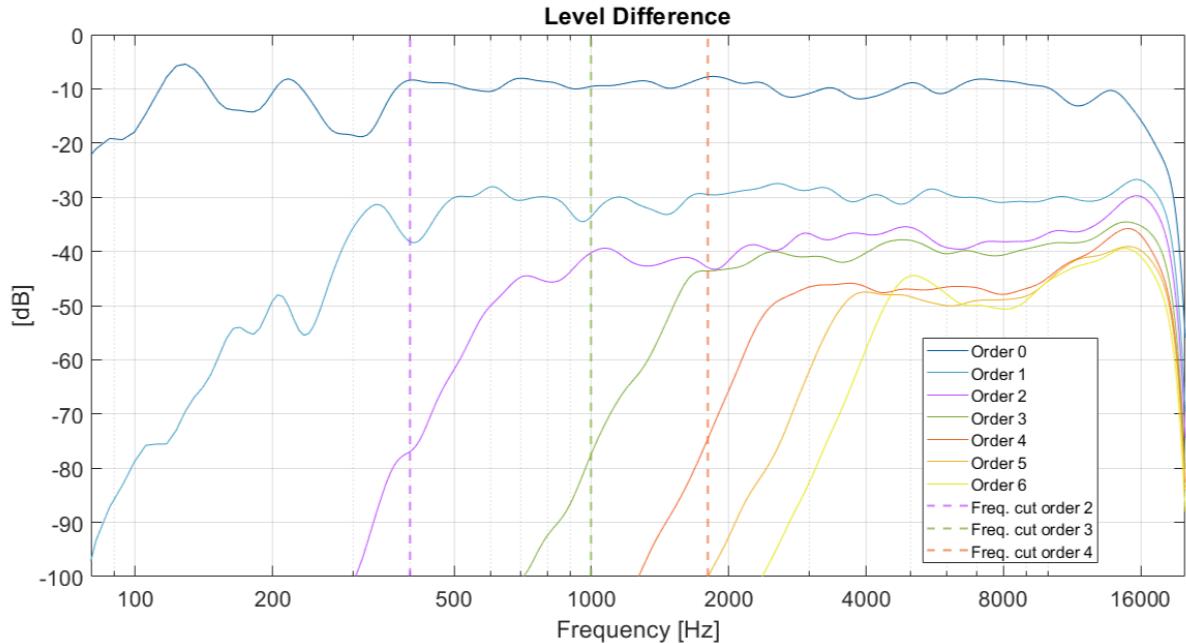
It is possible to observe that, comparing the tested cases with no equalization, the central position provides a slightly better performance for the first four orders up to 1000 Hz. The

introduction of the equalizations results in an overall enhancement, the improvement of the performance will most probably depend on the accuracy in the equalization process.

The Level Difference is depicted for each test case and order, as the disparity between the energy of the encoded response over the sources directions and the corresponding ideal components.

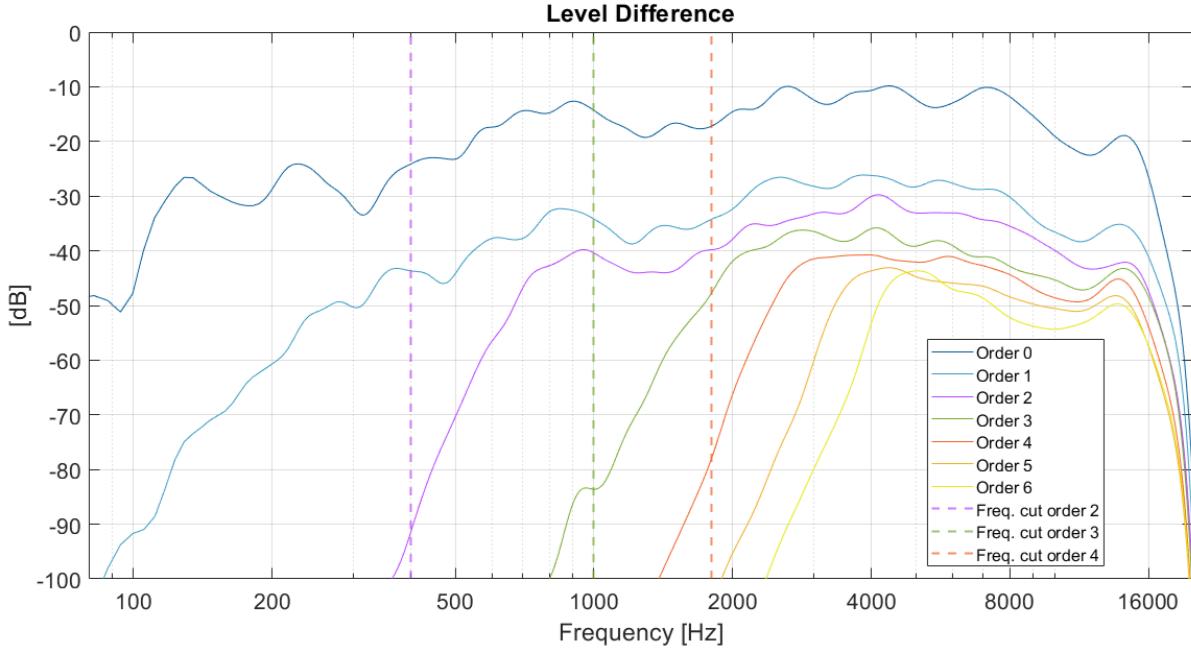


(a) No equalization applied.

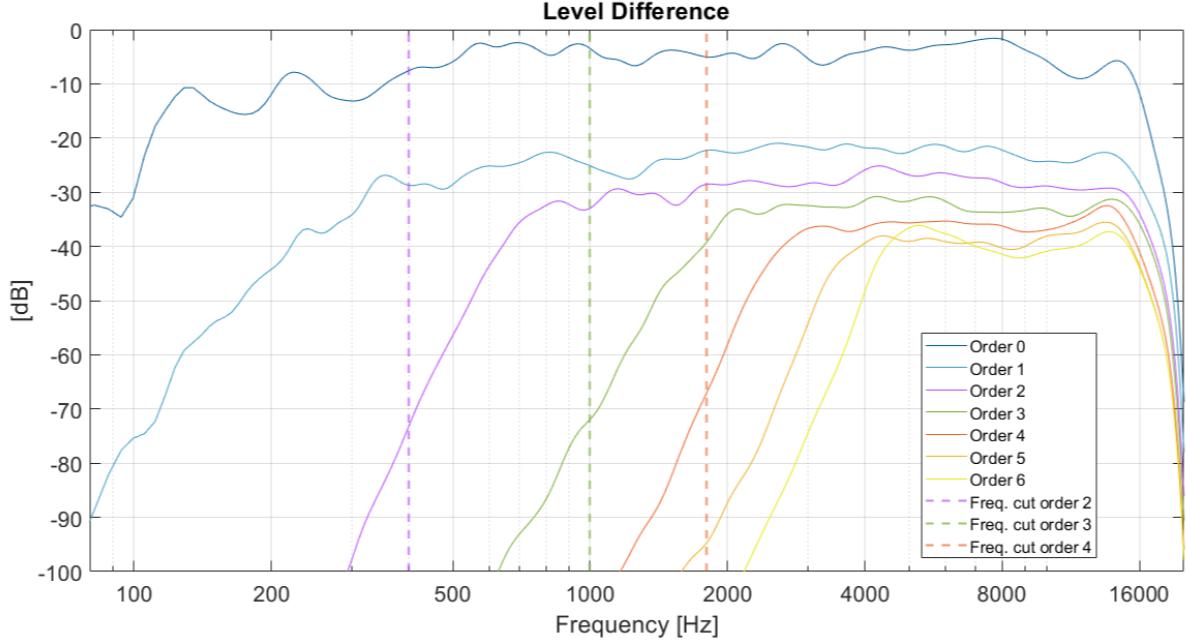


(b) Proper equalization applied.

Figure 4.7: Level Difference for each order measuring in central position. A 1/6 octave band smoothing is applied.



(a) No equalization applied.



(b) Proper equalization applied.

Figure 4.8: Level Difference for each order measuring in lateral position. A 1/6 octave band smoothing is applied.

The same observations done for Spatial Correlation can be effectuated for the Level Difference performance. The room provides, as already affirmed, a reconstruction up to order 4, however it is interesting to observe that even in higher orders enhancement occur. The

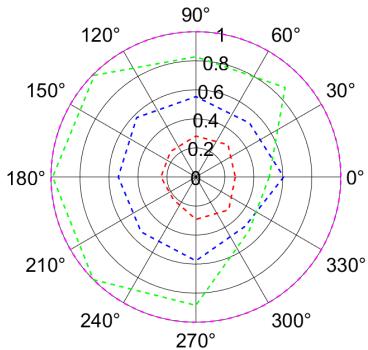
validity ranges of the microphone array are noticeable by looking at the trends of individual orders. The importance of an accurate equalization is certainly essential, especially the precision of the time alignment with respect to the listening position and a proper volume balance.

4.4. Graphical Representations of the Sound Field

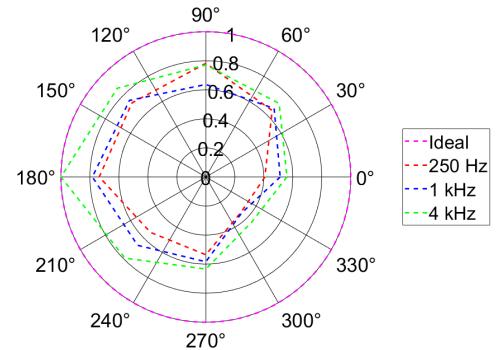
4.4.1. Spherical Harmonic Visualization

The current section provides a rendering of the measured spherical harmonics directivity. Polar patterns are displayed for octave bands, and the comparison with the directivity of ideal spherical harmonics can be made.

The polar plot enables visualization of the horizontal section of the described field, hence, the values of directivity returned by the sound sources located in the larger circumference of the hemispherical array (at the level where the elevation ϕ_l is the closest to zero), therefore from the values obtained by speakers 5, 6, 7, 8, 9, 10, 11, 12. This is the reason why the measured directivities are displayed roughly, with low sampling.

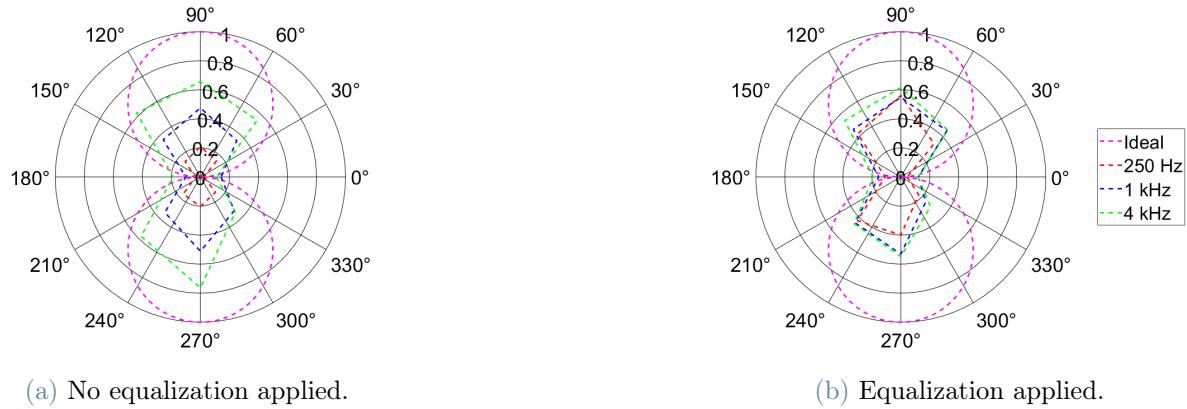
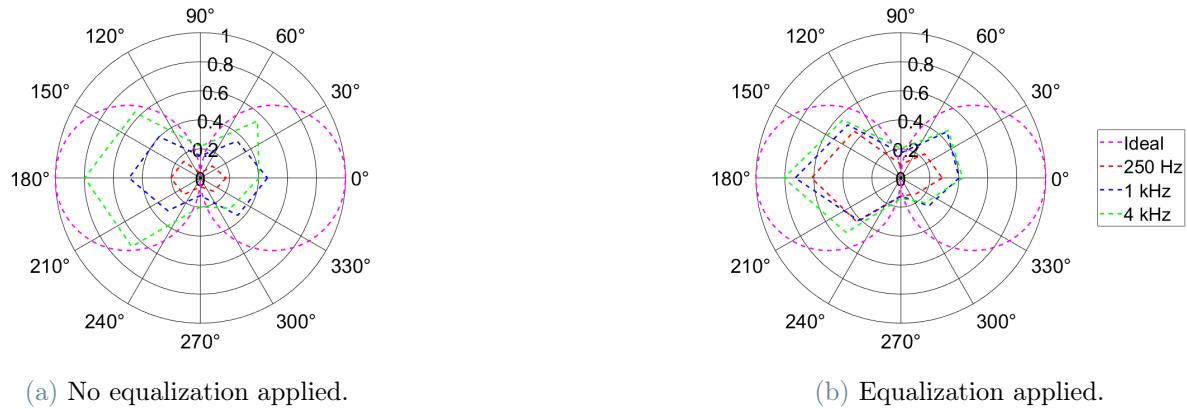
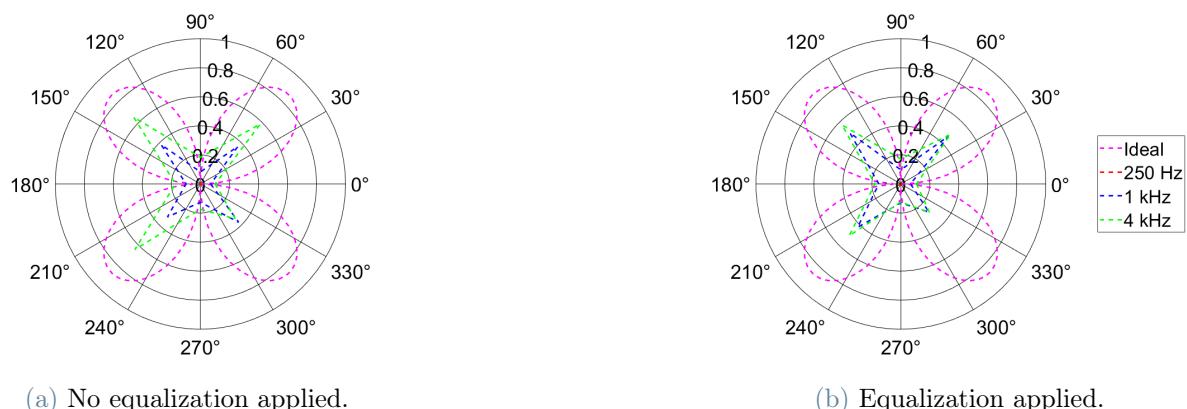


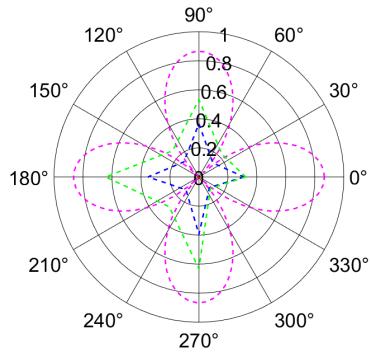
(a) No equalization applied.



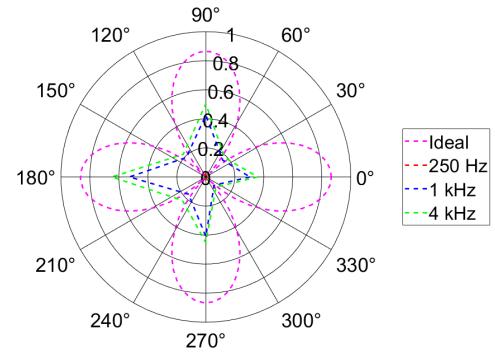
(b) Equalization applied.

Figure 4.9: \hat{Y}_{00} . Central measurements.

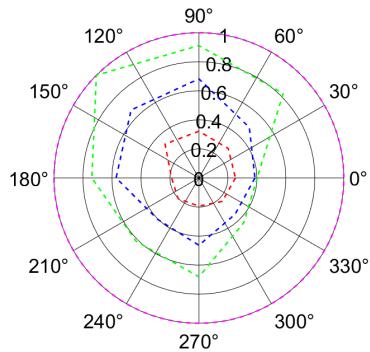
Figure 4.10: \hat{Y}_{-11} . Central measurements.Figure 4.11: \hat{Y}_{11} . Central measurements.Figure 4.12: \hat{Y}_{-22} . Central measurements.



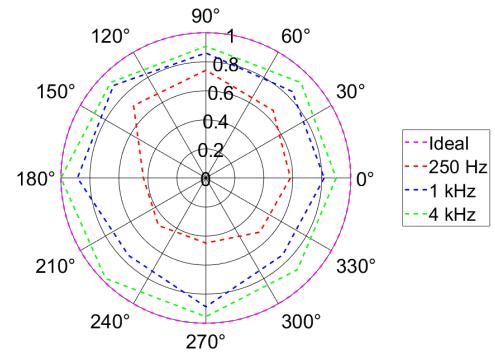
(a) No equalization applied.



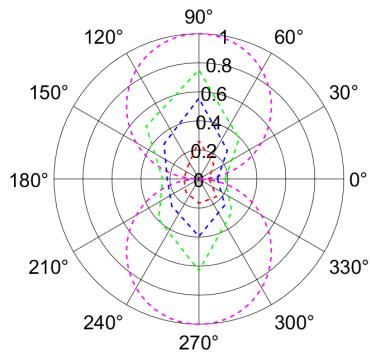
(b) Equalization applied.

Figure 4.13: \hat{Y}_{22} . Central measurements.

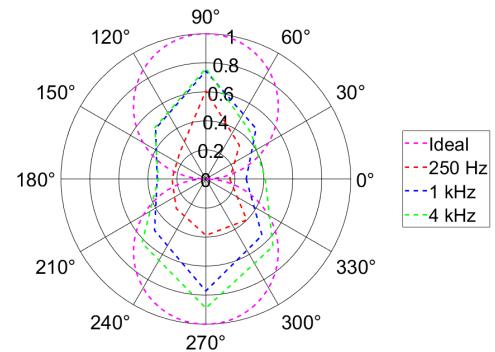
(a) No equalization applied.



(b) Equalization applied.

Figure 4.14: \hat{Y}_{00} . Left measurements.

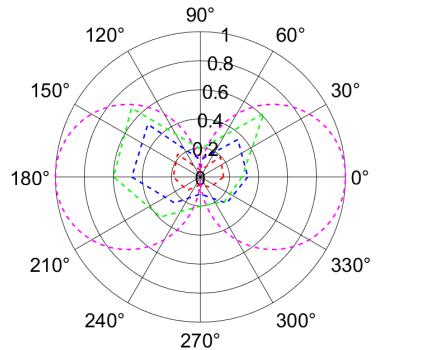
(a) No equalization applied.



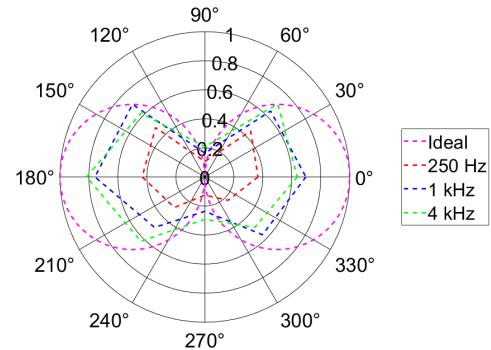
(b) Equalization applied.

Figure 4.15: \hat{Y}_{-11} . Left measurements.

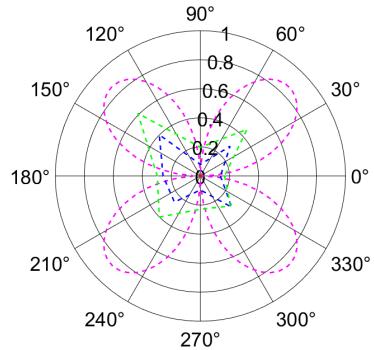
4 | Listening Room Performance Measurements



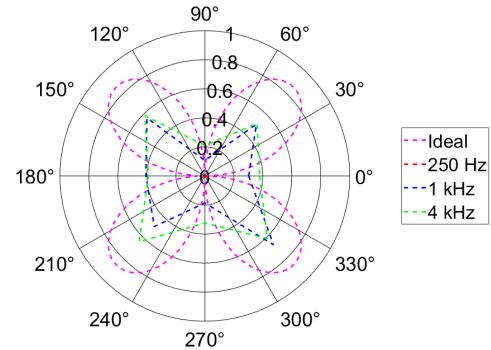
(a) No equalization applied.



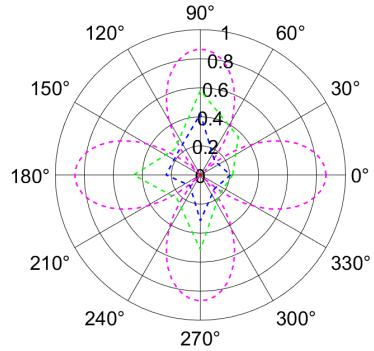
(b) Equalization applied.

Figure 4.16: \hat{Y}_{11} . Left measurements.

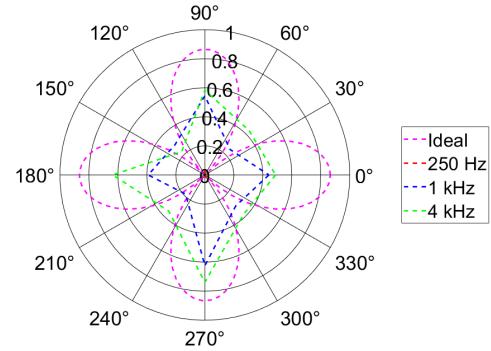
(a) No equalization applied.



(b) Equalization applied.

Figure 4.17: \hat{Y}_{-22} . Left measurements.

(a) No equalization applied.



(b) Equalization applied.

Figure 4.18: \hat{Y}_{22} . Left measurements.

From these patterns the main enhancement can be observe, comparing the non equalized directivities with the equalized ones, in a greater coherence achieved across different octave

bands. The more pronounced directivity provided by higher frequency is confirmed and it is interesting to observe how this increases even for lower frequencies by applying the equalization.

4.4.2. Pressure Measured on the Microphone

The images in this paragraph depict the pressure perceived by each of the 64 capsules of the microphone array.

The contributions of each speaker have been properly summed for each microphone, thus obtaining 64 room impulse responses.

In the visualizations the position of each point corresponds to the exact position of the array capsules in em64®, while its orientation is respected too:

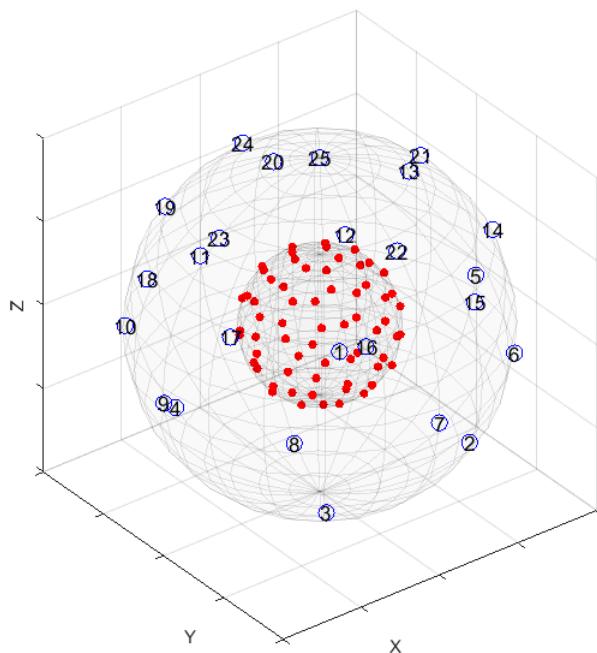
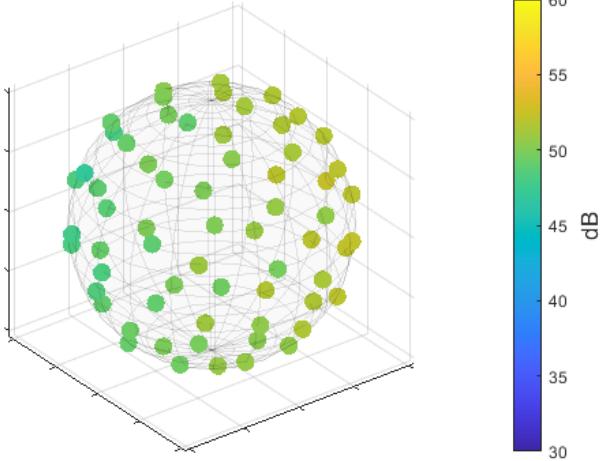


Figure 4.19: Orientation for the next pictures, in red the microphone capsules while the blue circle stands for the positions of the numbered speakers

The visualization of perceived pressure on the microphone array capsuled allows for required comparisons, assessing the homogeneity of the measured field across various case studies.

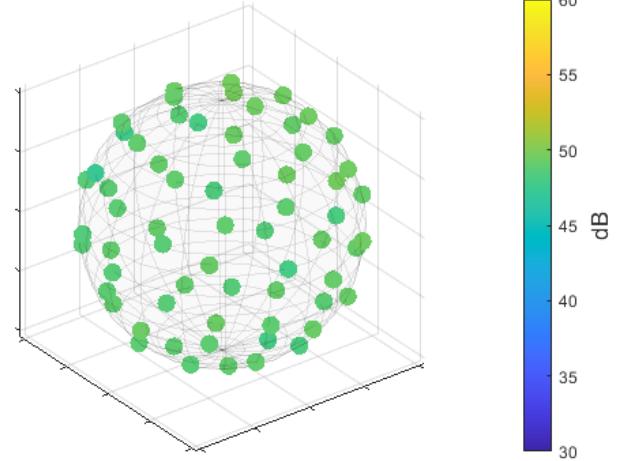
The values are visualized in octave bands:

**Pressure perceived on 64 microphones
500 Hz octave band**



(a) No equalization applied.

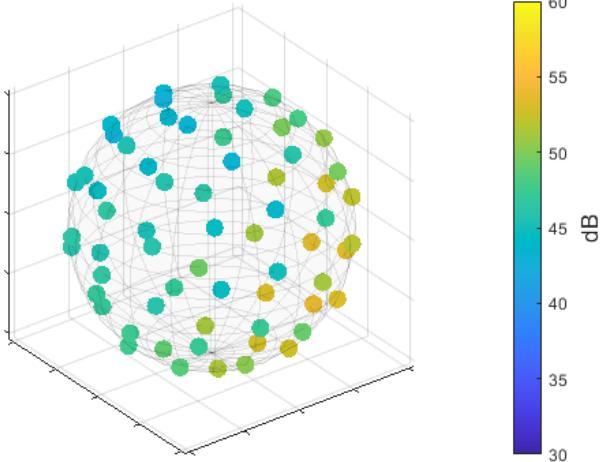
**Pressure perceived on 64 microphones
500 Hz octave band**



(b) Equalization applied.

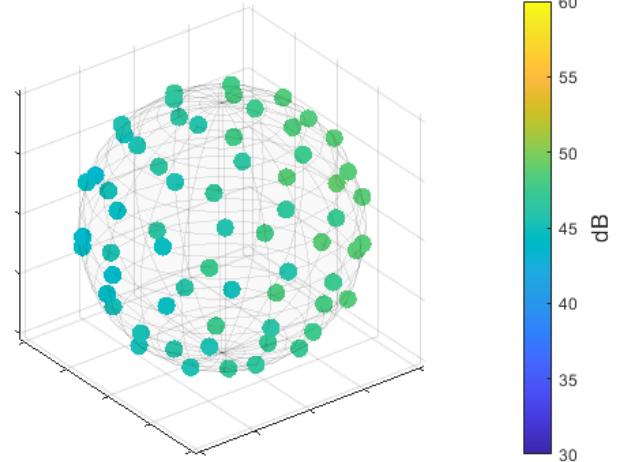
Figure 4.20: Comparison for central position measurements.

**Pressure perceived on 64 microphones
1000 Hz octave band**



(a) No equalization applied.

**Pressure perceived on 64 microphones
1000 Hz octave band**



(b) Equalization applied.

Figure 4.21: Comparison for central position measurements.

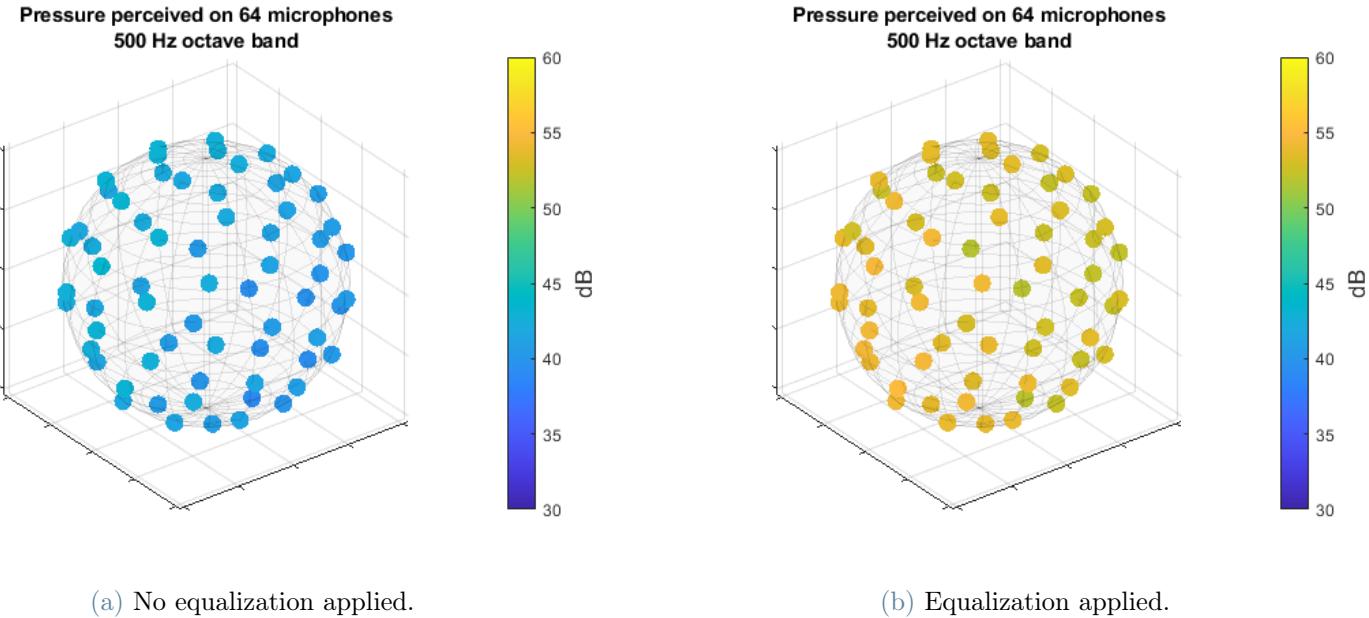


Figure 4.22: Comparison for lateral position measurements.

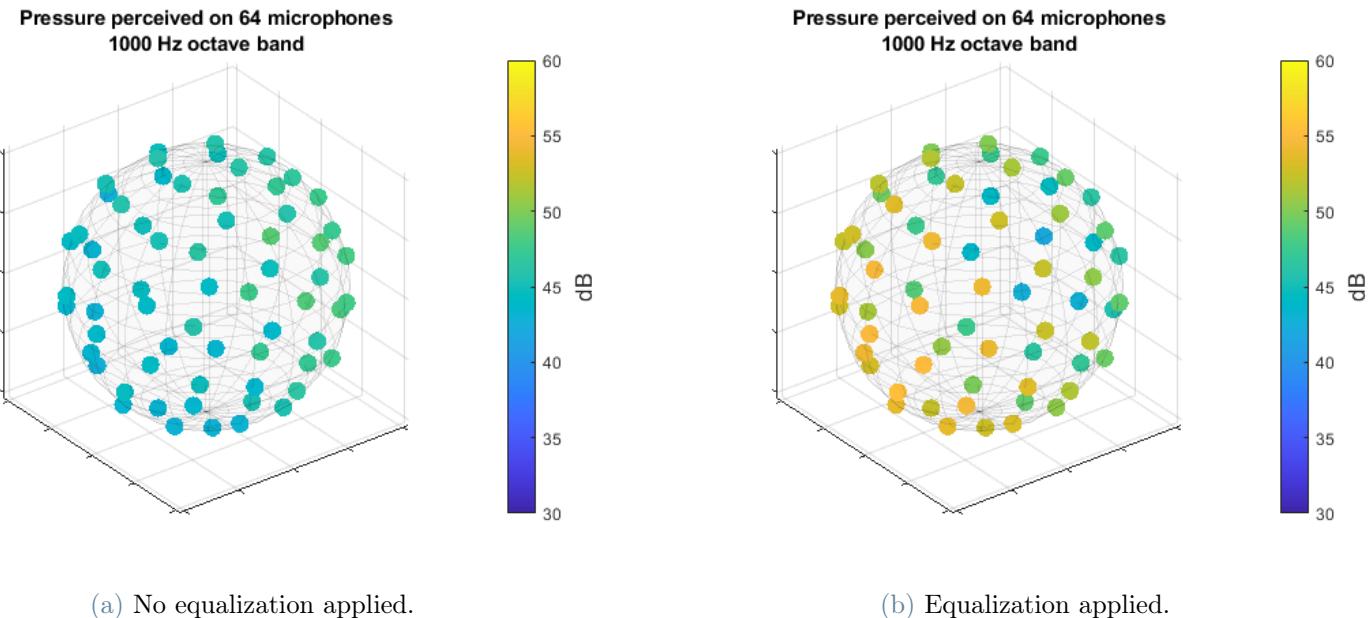


Figure 4.23: Comparison for lateral position measurements.

For the central position improvements on the homogeneity can be observed. On the other hand, the visualisations proposed do not give a sufficiently detailed rendering of the enhancements resulting from the tuning in the tested left measurement position.

The apparent worsening that appears in the case of equalization applied in left positioning is probably implied by the fact that the tuning does not take into account the introduction

of attenuation gains, whose are necessary to suppress the lack of equidistance between the measurement point and the speakers.

For the purpose of this study, the visual feedback proposed in 4.4.1 is preferred.

4.4.3. Decoding of the Extracted Spherical Harmonics

In this section the following procedure is applied for a visualization of the rendering obtained after a decoding process.

This last visualization can be useful, as the one in 4.4.2, to compare the different resulting level of uniformity of the sound field, which, in this case, is reconstructed by a decoder process.

The decoding involves recomposing the field starting from harmonic functions derived from the encoding at the specified measurement position. This entails discretizing the spatial coordinates of the measurement location. Fundamentally, the goal is reconstructing a sound field at the targeted position.

Once the decoding gains are determined for all points in the spatial discretization, a visual representation of the resulting distribution can be generated. This representation offers insight into the reconstructed sound field at the designated position.

Starting from the overall matrix of the spherical harmonics, obtained by summing all the contributions 4.1 from all the individual 25 directions l :

$$\hat{Y}_{TOT_{nm}}^{\omega} = \sum_{l=1}^{25} \hat{Y}_{nm}^{\omega}(\theta_l, \phi_l) \quad . \quad (4.4)$$

The sampling decoder, or projection decoder, is a traditional design approach that simply maps the spherical harmonics into the desired space discretization.

This decoder approach essentially corresponds to a plane-wave decomposition at the direction of each loudspeaker. This decoder behaves as be robust to irregular loudspeaker layouts, but it does not preserve the energy of a source for all directions.

The desired spatial layout is computed by using a t-design with $t = 21$, providing an amount of 240 points.

The decoding matrix is obtained as illustrated in 1.16, here reported considering the current label:

$$\mathbf{D} = \frac{4\pi}{P} \mathbf{Y} \quad , \quad (4.5)$$

where, in this case, \mathbf{Y} corresponds, in extended form, to $Y_{nm}(\theta_p, \phi_p)$, with p up to P , the number of points provided by the space discretization, and θ_p, ϕ_p the respective azimuth and elevation of each point.

The decoding process is computed applying the formula 1.12, where the computed $\hat{\mathbf{Y}}_{\text{TOT}}(\omega) = [\hat{Y}_{TOT_{0,0}}^{\omega}, \hat{Y}_{TOT_{-1,0}}^{\omega}, \dots, \hat{Y}_{TOT_{NM}}^{\omega}]$ corresponds to $\mathbf{B}(\omega)$.

Decoding Ambisonics recordings, captured with a spherical microphone array, requires special consideration for the fact the HOA signals themselves are frequency limited due to the microphone array properties, see 2.3 especially for higher-orders. The use of a single decoding matrix for all mid-high frequencies will not lead to bad results but technically some loss of power and colouration for a source captured from some direction and then decoded can occur, due to the decoding matrix tuned to preserve energy using all HOA signals indistinctly.

A proper solution approach is to use as many decoding matrices as orders frequency ranges. In this case the frequency ranges reported in 2.3 are respected and a fourth order decoding has been computed.

The obtained signals measured in the two positions, decoded up to the fourth order, can be visualized in the octave bands [500 Hz, 1000 Hz, 2000 Hz, 4000 Hz], firstly from the central position measurements and secondly from the lateral position measurements.

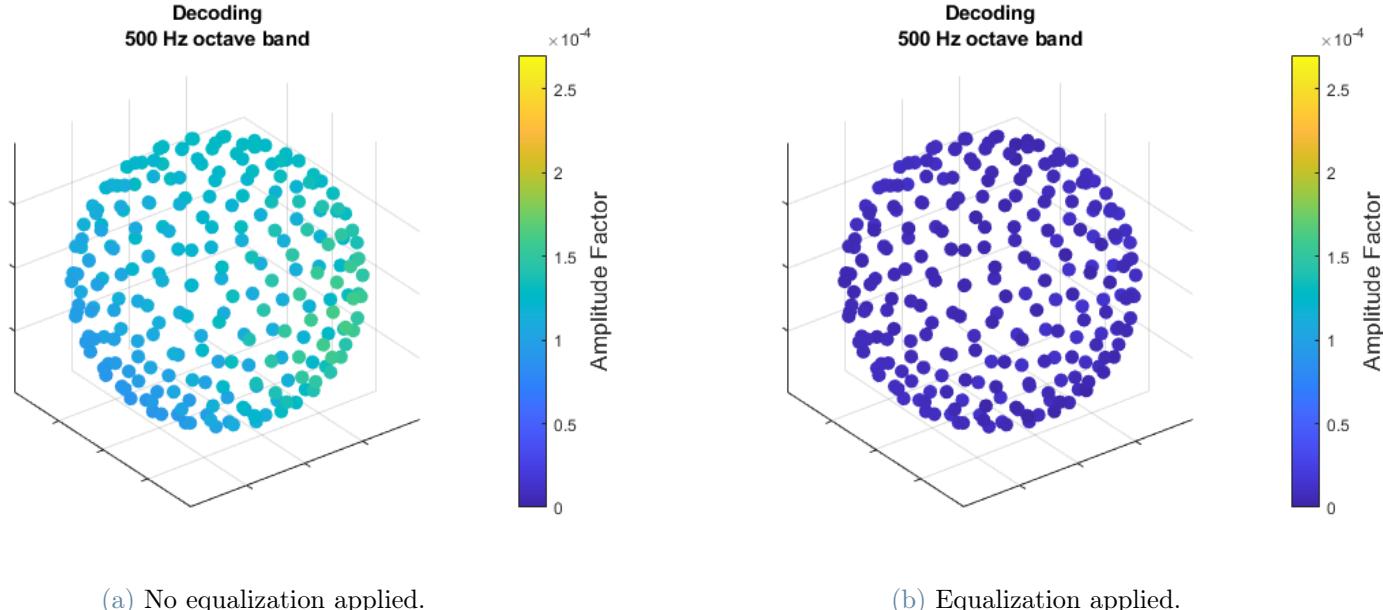


Figure 4.24: 4-th order decoding of the measurements in central position in 240 points t-design ($t=21$).

4| Listening Room Performance Measurements

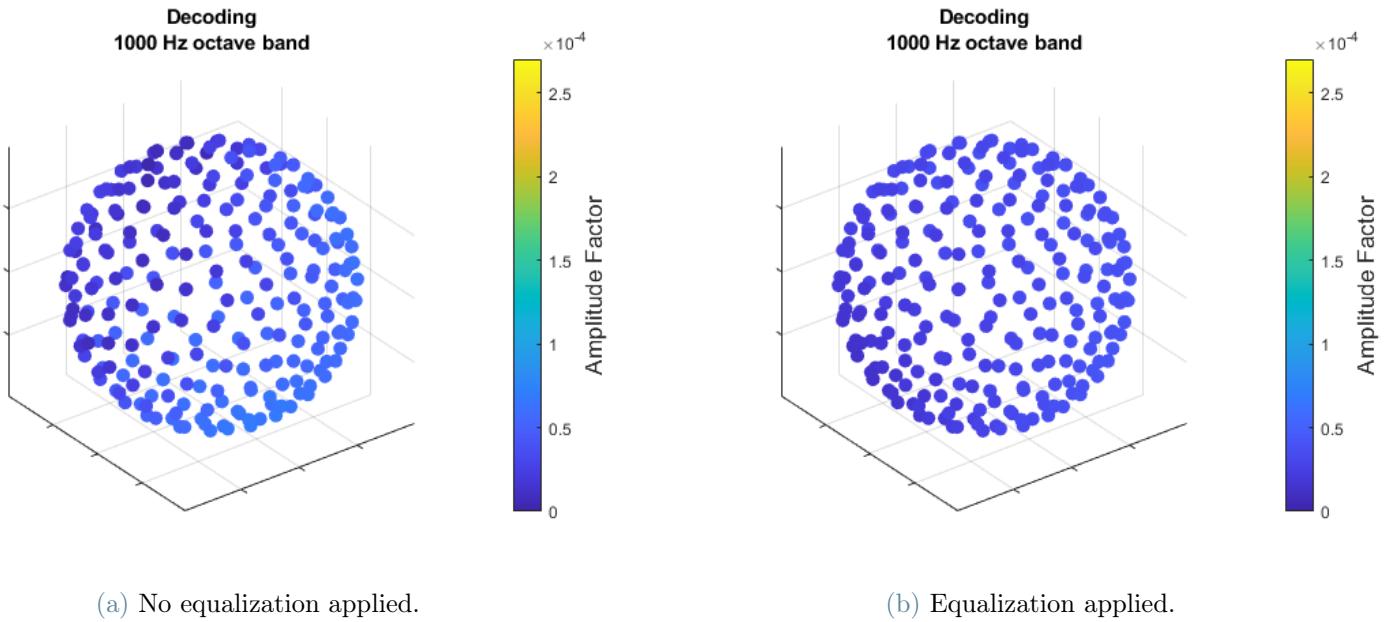


Figure 4.25: 4-th order decoding of the measurements in central position in 240 points t-design ($t=21$).

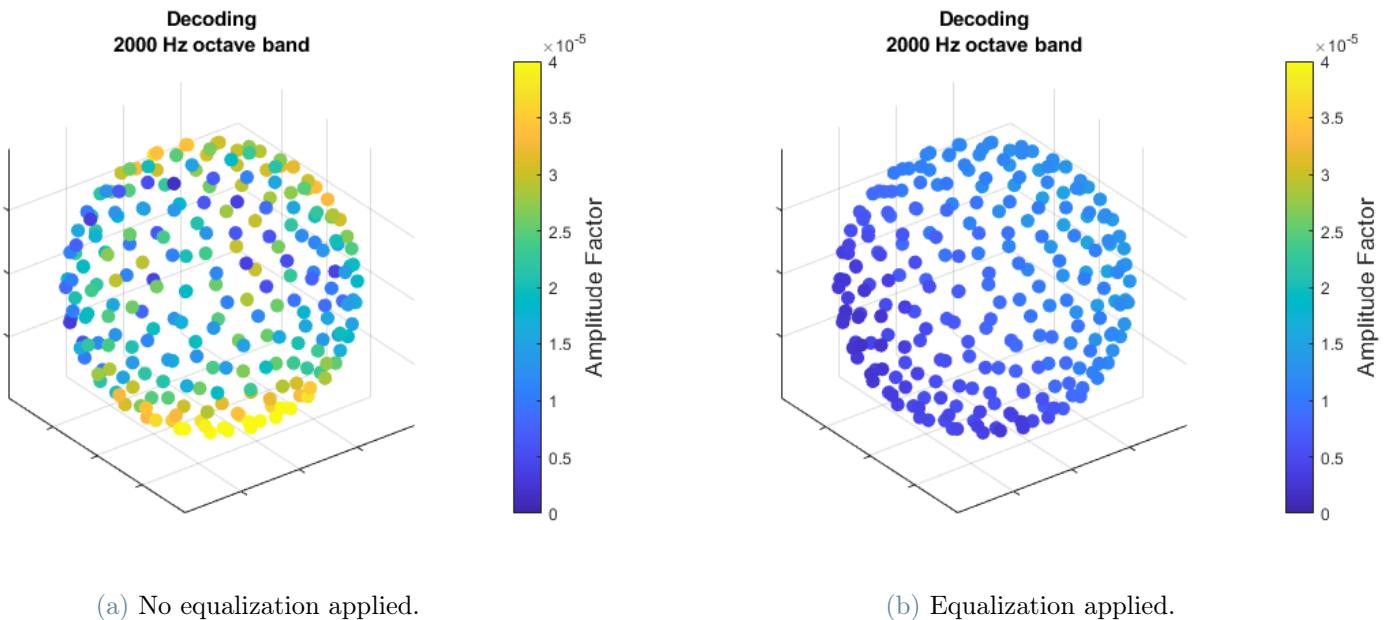


Figure 4.26: 4-th order decoding of the measurements in central position in 240 points t-design ($t=21$).

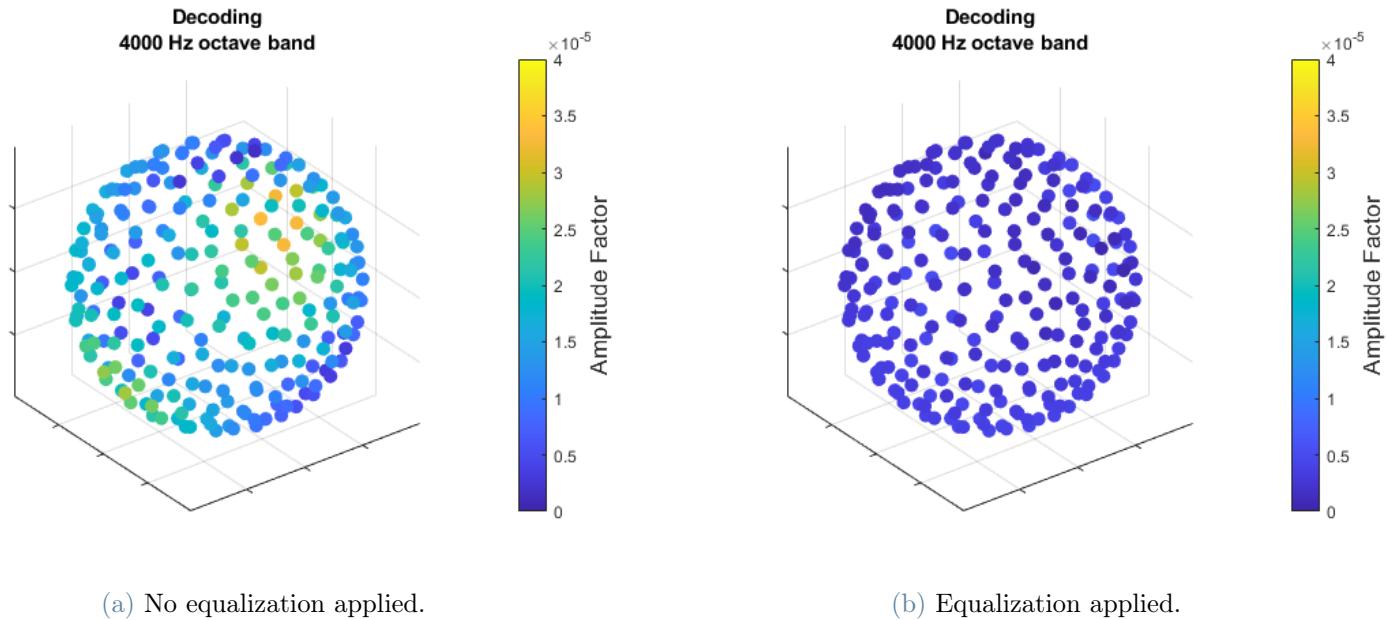


Figure 4.27: 4-th order decoding of the measurements in central position in 240 points t-design ($t=21$).

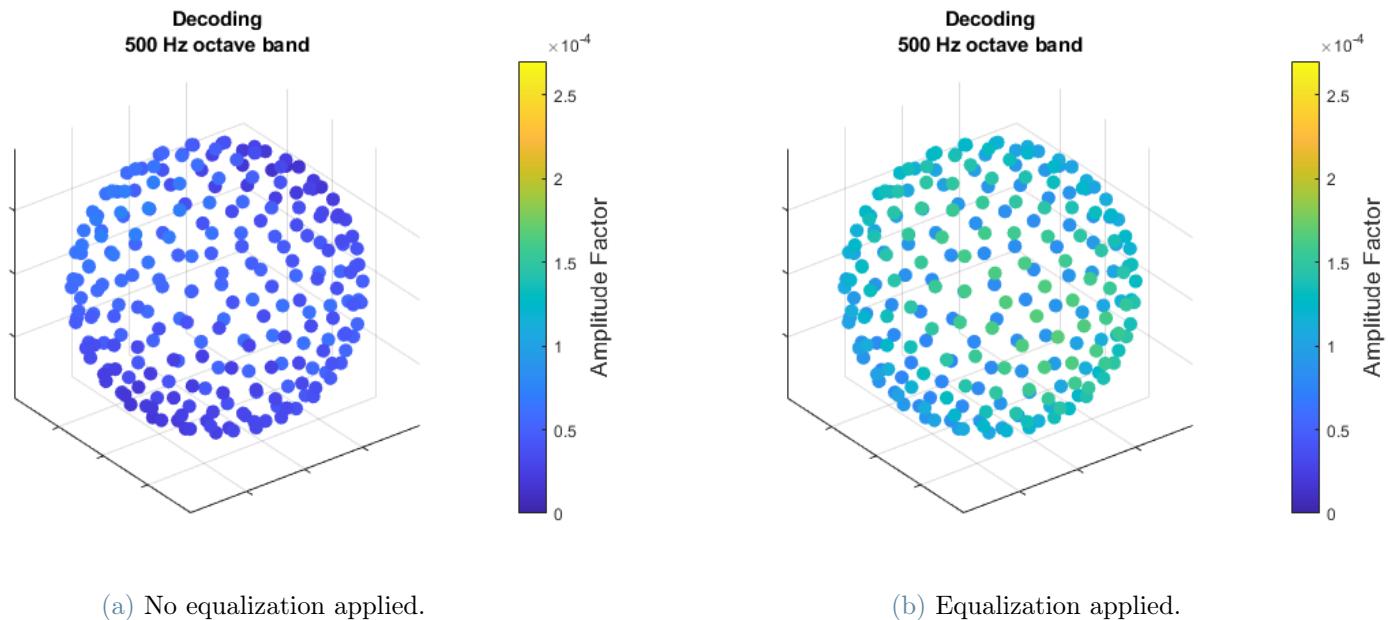


Figure 4.28: 4-th order decoding of the measurements in lateral position in 240 points t-design ($t=21$).

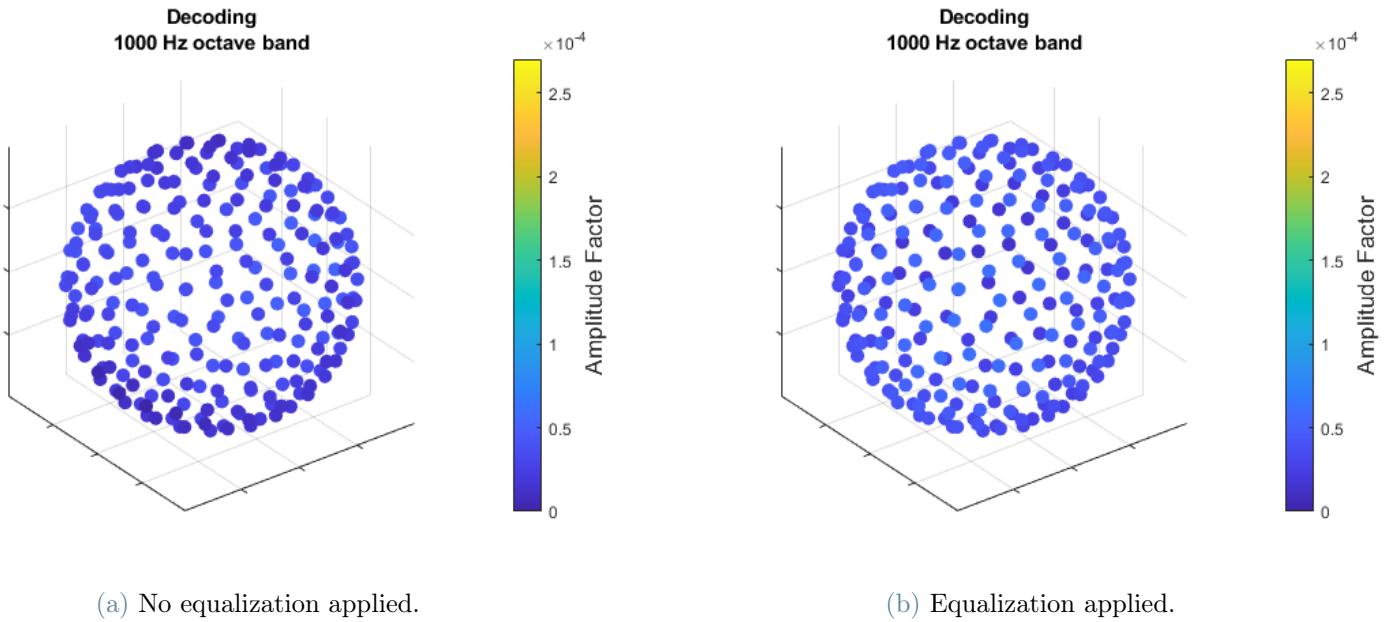


Figure 4.29: 4-th order decoding of the measurements in lateral position in 240 points t-design ($t=21$).

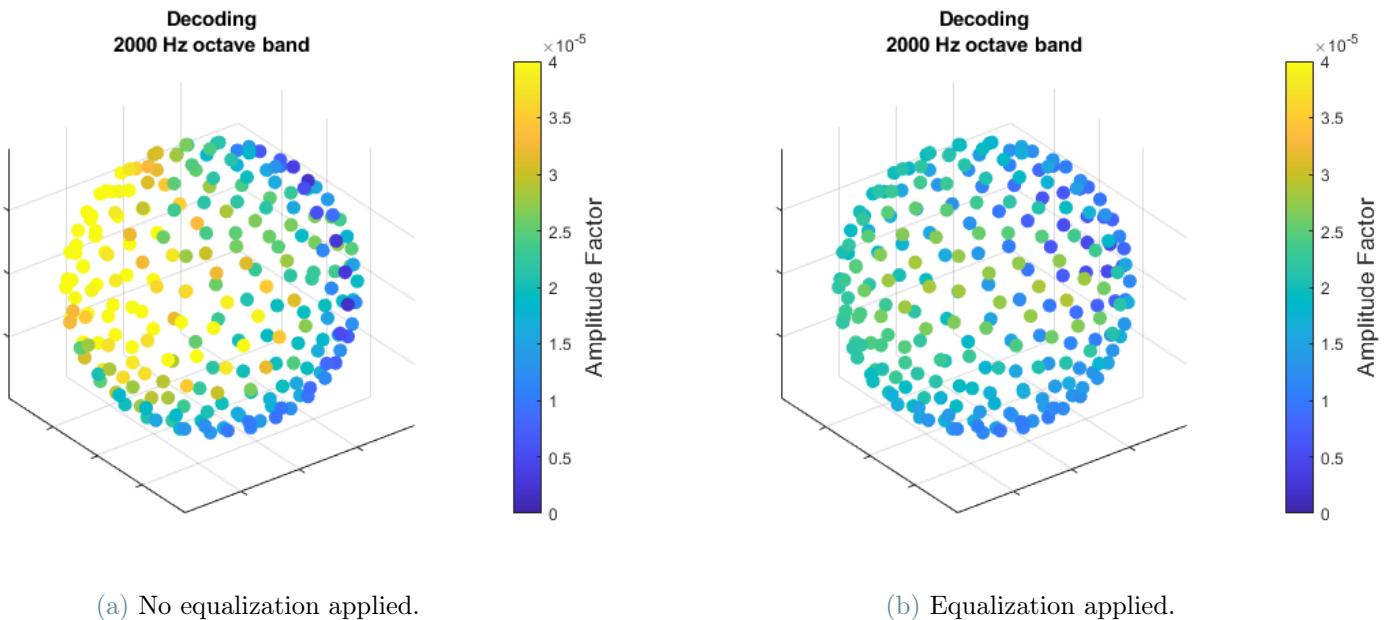


Figure 4.30: 4-th order decoding of the measurements in lateral position in 240 points t-design ($t=21$).

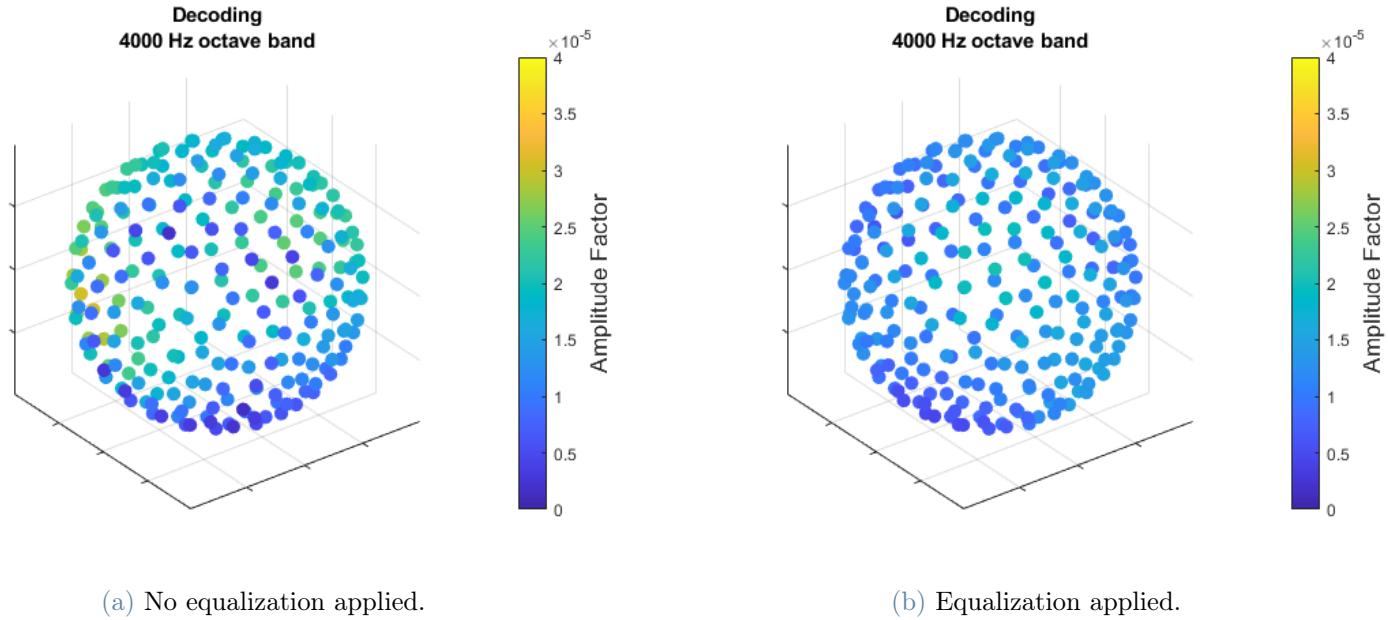


Figure 4.31: 4-th order decoding of the measurements in lateral position in 240 points t-design ($t=21$).

It can be concluded that the visualisations proposed in 4.4.3 and 4.4.2 gives a rendering of the uniformity enhancements led by the tuning. Moreover, the results up to the 1000 Hz octave band appear to depict a relatively uniform scenario even in the test case with no equalization.

5 | Conclusions

This thesis project presents an approach for the characterization and assessment of an room devoted to Ambisonics audio reproduction.

The acoustic performance of the treated room was studied by extrapolating the useful acoustic parameters: reverberation time, clarity index and early decay time, over a wide frequency range. For the acoustic evaluation, the room was excited by an exponential sine sweep specially pre-distorted and sent to the woofers, tweeters and, lastly, an exponential sweep sent to the subwoofers for the lowest frequency analysis.

The room demonstrated to have excellent sound absorption properties for all the frequency range and can definitely be classified as a semi-anechoic room.

Once the acoustic quality of the room was established and its prerequisites for an immersive audio reproduction were judged fulfilled, considerations were made on the available speaker configuration.

The array, consisting of 25 speakers, does not respect a t -design discretisation (which is the ideal one for Ambisonics reproduction) but it's characterised by a configuration that envelops the *sweet spot* spatially quite evenly.

The reproduction of spherical harmonics was tested, then, with the use of Spatial Correlation and Level Difference metrics. The first one aims to characterise the similarity between the generated patterns and ideal spherical harmonic patterns. The latter aims to calculate the disparity between the average level of the patterns in all directions, which is the total sum of the energy of the response encoded at all points, and the corresponding ideal components.

These two metrics have been used in past studies, and described in the literature, for the evaluation of microphone arrays. Therefore, the procedures of these studies was testing the microphone array through triggering it with identical sources (by moving the same speaker) from various directions, in the best case using a numerous amount of points following to a t -design configuration.

These metrics have never been tested for audio reproduction except than in this working project. In this case, therefore, the roles are reversed: the microphone array is intended as ideal while the source environment is evaluated.

Four case studies were tested: central position, in the absence of tuning and applying a tuning to evaluate a possible improvement, lateral position 40 cm left of centre, in the absence of tuning and applying specific tuning for this second measurement position.

The values of both Spatial Correlation and Level Difference result much improved with the use of equalisation, and the performance turns out to be enhanced in the central position than in the off-centre position. The metrics behave, therefore, as desired, proving that they can be used for the Ambisonics evaluation of audio reproduction.

The entire evaluation was conducted without the introduction of a decoder for sine sweeps reproduction. It should be taken into consideration that a properly implemented decoder can greatly enhance the performance of the playback system.

An advanced decoder can ensure constant amplitude gains and constant energy gains for all the speakers, improve low-frequency directivity, enhance localization, and adapt to various speaker array configurations that deviate from the considered ideal configurations by using vector-based techniques associated with Ambisonics.

5.1. Limitations

It is possible to identify some intrinsic limitations present in the applied and in the metric itself:

- The applied equalization does not take into account the necessary gain or attenuation leveling, the equispatiality between the listening point and each individual speaker is not guaranteed, particularly in the lateral measurement position.
- The Level Difference and Spatial Correlation metrics propose a comparison between the measured spherical harmonics generated by a source coming from a direction (θ_l, ϕ_l) and the ideal coefficients calculated in the same direction. It would be interesting and useful to extrapolate a metric that allows a comparison with a more ideal scenario, thus with respect to sources coming from multiple directions placed in ideal positions for the Ambisonics reproduction. Hence, a comparison that takes into account the spatial distribution of the sources would be effectuated.

5.2. Future Developments

In this final section, some proposals as future developments of the project are elucidated:

- The introduction of the decoding can lead to a complete evaluation of the room performance associated to that particular decoding applied. The sweep signal Am-

bisonics components should be sent to the decoder for their reassembly and reproduction. The number of source directions can, therefore, be a multitude, complying with the directionality constraints of the utilized microphone array. The metric proposed could be tested on this comprehensive situation which incorporates the decoding contribution, the assessment will focus, hence, on the enhancement led by the decoding.

- Numerical analysis and simulations of the actual reproduced system can be carried out in order to differentiate the lacks given by of the speaker arrangement with respect to the lacks led by the room structure and eventually by the reproduction system itself. Various layouts and situations can be tested by expanding the virtual loudspeaker arrangement which is intended to be simulated.
- The quantity of tested positions can be augmented in order to have a mapping of the quality evaluated in various listening points and define the confinements of the area that meets a satisfactory Ambisonics reproduction, according to the needs of the use case. The test scenarios can also be extended to untreated acoustically domains and sites designed for a primary purpose different than the audio listening.
- As indicated in the introduction of this thesis, for a comprehensive assessment the perceptual and subjective aspects of audio listening should be taken into account. As a further work an analysis through a sample of listeners is suggested, in order to assess the uniformity of the reproduced sound field and, with the utilization of the the decoder, conducting localization experiments and evaluating the various properties of the immersive system.

Bibliography

- [1] Eigenbeam Data. Specification for Eigenbeams - version 1. *MH Acoustics*.
- [2] Aurora manual. URL <http://pcfarina.eng.unipr.it/aurora/download/Manual-HelpFile/Aurora43Manual.pdf>.
- [3] Eigenmike® microphone. URL <https://mhacoustics.com/products>.
- [4] Eigenmike64® picture. URL <https://www.tegakari.net/en/2022/07/em64-eigenmike-microphone/>.
- [5] B-format picture. URL <https://www.fugato.com/pickett/pantoph.shtml>.
- [6] Room EQ Wizard-equalizer selection. URL https://www.roomeqwizard.com/help/help_en-GB/html/equaliser.html.
- [7] IEM-Cube, Institute of Electronic Music and Acoustics. URL <https://www.kug.ac.at/en/university/campus-and-buildings/iem-institute-17-electronic-music-and-acoustics>.
- [8] Sr:iosr.acoustic.irstats. URL <https://github.com/IoSR-Surrey/MatlabToolbox/blob/master/+iosr/+acoustics/irStats.m>.
- [9] EN ISO 3382-1997. Acoustics. Measurement of Room Acoustic parameters. Part 1: Performance Spaces.
- [10] Eigenmike® microphone, MH acoustic. URL <https://mhacoustics.com/products>.
- [11] T-design reference library. URL <http://neilsloane.com/sphdesigns/>.
- [12] Sr:iosr.acoustic.irstats. URL <http://research.spa.aalto.fi/projects/sht-lib/sht.html>.
- [13] Ambisonic.net. URL <http://www.ambisonic.net/>.
- [14] Sonosfera® article, Comune di Pesaro. URL https://www.comune.pesaro.pu.it/novita-in-comune/dettaglio/news/inaugurazione-della-sonosfera/?tx_

- news_pi1%5Baction%5D=detail&tx_news_pi1%5Bcontroller%5D=Event&cHash=233c3a7d5f90ac658ef86f08da154aa0.
- [15] 3D Sound Field Rendering under Non-idealized Loudspeaker Arrangements. *The Journal of the Acoustical Society of America*, 2008.
 - [16] American National Standard Specification for Octave-band and Fractional-octave-band Analog and Digital Filters. 2009. Acoustic Society of America.
 - [17] J. Abel, M. Kolar, and B. P. Nicholas J. Bryan. Estimating Room Impulse Responses from Recorded Balloon Pops. Audio Engineering Society, San Francisco, CA, 2010.
 - [18] F. B. and Stefan Weinzierl. Audio Quality Assessment for Virtual Reality. In *Sonic Interactions in Virtual Environments*, volume 7, 56, page 145–178. Human–Computer Interaction Series. Springer, Cham, 2022.
 - [19] N. Aoshima. Computer-generated Pulse Signal Applied for Sound Measurement. volume 65, page 1484–1488. 1981.
 - [20] D. Arteaga. Introduction to Ambisonics. 2023.
 - [21] J. Blauert. Localization and the Law of the First Wavefront in the Median Plane. volume 50, page 466–470. Acoustic Society of America, 1971.
 - [22] J. Blauert. *Spatial Hearing The Psychophysics of Human Sound Localization*. The MIT Press, Harvard MA, 1983.
 - [23] A. Brown, D. J. Tollin, and G. C. Stecker. The precedence effect in sound localization. *Journal of the Association for Research in Otolaryngology*, 2014.
 - [24] D. H. Cooper and T. Shiga. Discrete-matrix Multichannel Stereo. volume 2. 1972.
 - [25] J. Daniel. Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format.
 - [26] J. Daniel. Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format. Master’s thesis, France Telecom RD, 2003.
 - [27] J. Daniel and S. Moreau. Further Study of Sound Field Coding with Higher Order Ambisonics. 2004.
 - [28] P. Delsarte, J. M. Goethals, and J. J. Seidel. Spherical Codes and Designs. In *Geometriae Dedicata*, volume 6, pages 363–388. 1977.

- [29] G. Dickins. Soundfield Representation, Reconstruction and Perception. Master of Engineering of the Australian National University, 2003.
- [30] A. Farina. *Soundfield microphone*, . URL https://www.soundfield.com/#/learn/microphone_basics.
- [31] A. Farina. Ambisonics/sps pages, . URL http://pcfarina.eng.unipr.it/Aurora/HOA_ACN_N3D_formulas.html.
- [32] A. Farina. Advancements in Impulse Response Measurements by Sine Sweeps. *122nd Convention Vienna, Austria*, 2007.
- [33] A. Farina. Simultaneous Measurement of Impulse Response and Distortion. *23rd Nordic Sound Symposium*, 2007.
- [34] A. Farina, A. Capra, L. Conti, P. Martignon, and F. Fazi. Measuring Spatial Impulse Responses in Concert Halls and Opera Houses employing a Spherical Microphone Array. *19th International Congress On Acoustics*, 2007.
- [35] F. Fazi. Sound field reproduction, Ph.D. dissertation. ISVR, University of - Southampton, 2010.
- [36] F. Fazi, P. Nelson, and R. Potthast. Analogies and Differences Between three Methods for Soundfield Reproduction.
- [37] M. Ferroni. Evaluation and Psychoacoustic Validation of Techniques for the Analysis of Low Frequency Resonance Modes in Real Small Rooms, 2014.
- [38] A. Fürjes. Room acoustic parameters - what's up? 2021. aQrate Acoustics Ltd.
- [39] M. A. Gerzon. General Metatheory of Auditory Localisation. Audio Engineering Society, Vienna, .
- [40] M. A. Gerzon. Optimum Reproduction Matrices for Multispeaker Stereo. Oxford, UK, .
- [41] M. A. Gerzon. Periphony: With-height Sound Reproduction. 1973.
- [42] M. Gräf and D. Potts. On the Computation of Spherical Designs by a new Optimization Approach based on Fast Spherical Fourier Transforms. volume 119, page 699–724. Numerische Mathematik, 2011.
- [43] H. Haas. The Influence of a Single Echo on the Audibility of Speech. volume 20, page 145–159. Acoustic Society of America, 1949.

- [44] H. Hardin and N. J. A. Sloane. McLaren's Improved Snub Cube and Other New Spherical Designs in Three Dimensions. volume 15, pages 429–441. Discrete and Computational Geometry, 1996.
- [45] R. Heyser. Acoustical Measurements by Time Delay Spectrometry. volume 15, page 370–382. Journal of the Audio Engineering Society, 2014.
- [46] ITU-R Rec. BS.1534-3. Subjective assessment of sound quality. *Int. Telecomm. Union, Geneva*, 2015.
- [47] ITU-R Rec. P.1310. Spatial audio meetings quality. *Int. Telecomm. Union, Geneva*, 2017.
- [48] ITU-R Rec. P.800. Methods for subjective determination of transmission quality. *Int. Telecomm. Union, Geneva*, 1996.
- [49] ITU-R Rec. P.863. Perceptual objective listening quality assessment. *Int. Telecomm. Union, Geneva*, 2014.
- [50] ITU-RBS.1116-3. Methods for the subjective assessment of small impairments in audio systems. *Int. Telecomm. Union, Geneva*, 2015.
- [51] D. Jerome, J. Rault, and J. Polack. Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions. *105th Convention of Audio Engineering Society, San Francisco*, 1998.
- [52] G. S. Kendall. Spatial Perception and Cognition in Multichannel Audio for Electroacoustic Music. *Organised Sound*, 2010.
- [53] O. Kirkeby, P. A. Nelson, and H. Hamada. The "Stereo Dipole" - A Virtual Source Imaging System Using Two Closely Spaced Loudspeakers. volume 46, chapter 5. Journal of Acoustic Engineering Society, 1998.
- [54] E. Kuttruff. *Room Acoustics - sixth edition*. Focal Press, 2017.
- [55] A. Laborie, R. Bruno, and S. Montoya. A New Comprehensive Approach of Surround Sound Recording. Amsterdam, 7 2003. Convention paper 5717.
- [56] R. Lee and E. M. Benjamin. Is My Decoder Ambisonic?
- [57] R. Y. Litovskya, H. S. Colburn, W. A. Yost, and S. J. Guzman. The Precedence Effect. *Acoustical Society of America*, 1999.
- [58] L. Mccormarck, V. Pulkki, A. Politis, O. Scheuregger, and M. Marschall. Higher-Order Spatial Impulse Response Rendering: Investigating the Perceived Effects of

- Spherical Order, Dedicated Diffuse Rendering, and Frequency Resolution. volume 68, chapter 5, page 338–354. Journal of Audio Engineering Society, 2020.
- [59] L. Mh acoustics. Eigenbeam Data Specification for Eigenbeams. Datasheet.
 - [60] S. Moreau, J. Daniel, and S. Bertet. 3D Sound Field Recording with Higher Order Ambisonics – Objective Measurements and Validation of a 4th Order Spherical Microphone. Paris, 2006.
 - [61] D. M. Murillo, F. M. Fazi, and M. Shin. Evaluation of Ambisonics Decoding Methods with Experimental Measurements. Berlin, Germany, 4 2014.
 - [62] S. Müller and P. Massarani. Transfer Function Measurement with Sweeps.
 - [63] S. Müller and P. Massarani. Transfer Function Measurement with Sweeps. volume 49, chapter 6, pages 443–471. JAES, 6 2001.
 - [64] M. Narbutt, J. Skoglund, A. Allen, M. Chinen, D. Barry, and A. Hines. AMBIQUAL: Towards a Quality Metric for Headphone Rendered Compressed Ambisonic Spatial Audio. applied sciences, 2020.
 - [65] A. Novak, F. Rund, and P. Honzík. Impulse Response Measurements using MLS Technique on Nonsynchronous Devices. volume 64, chapter 12, pages 978–987. Journal of the Audio Engineering Society, 2016.
 - [66] D. Pinardi. Spherical t-design for Characterizing the Spatial Response of Microphone Arrays.
 - [67] D. Pinardi and A. Farina. Metrics for Evaluating the Spatial Accuracy of Microphone Arrays.
 - [68] A. Politis and H. Gamper. Comparing Modeled and Measurement-based Spherical Harmonic encoding Filters for Spherical Microphone Arrays. 2017.
 - [69] V. Pulkki. Virtual Sound Source Positioning Using Vector Base Amplitude Panning. volume 45, chapter 6, pages 456 – 466. Journal of Audio Engineering Society, 1997.
 - [70] S.Bertet, E. Parizet, J. Daniel, and O. Warusfel. Investigation on Localisation Accuracy for First and Higher Order Ambisonics Reproduced Sound Sources. Acta Acustica united with Acustica, 2013.
 - [71] S.Bertet, E. Parizet, J. Daniel, and O. Warusfel. Perceptual Evaluation of an Ambisonic Auralization System of Measured 3D Acoustics. volume 7, 56. Acta Acustica, EDP Sciences, 2023.

- [72] M. Schroeder. Schroeder Frequency Revisited. volume 99, page 3240–3241. 1996.
- [73] W. B. Snow. Basic Principles of Stereophonic Sound. volume 61. 1953.
- [74] G.-B. Stan, J.-J. Embrechts, and D. Archambeau. Comparison of Different Impulse Response Measurement Techniques. 2002.
- [75] P. Stitt, S. Bertet, and M. van Walstijn. Off-centre Localisation Performance of Ambisonics and HOA for Large and Small Loudspeaker Array Radii. volume 100. ACTA Acoustica United With Acoustica, 2014.
- [76] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, and C. Colomes. PEAQ - the ITU Standard for Objective Measurement of Perceived Audio Quality. volume 48 1/2, pages 3–29. Journal of the Audio Engineering Society, 2000.
- [77] F. E. Toole. *Sound Reproduction - Loudspeakers and Rooms*. Spon Press, 2009.
- [78] A. Venturi, A. Farina, and L. Tronchin. On the Effects of Pre-processing of Impulse Responses in the Evaluation of Acoustic Parameters on Room Acoustics. ICA Montreal, 2013.
- [79] R. Vermeulen. Stereo-Reverberation. 1958.
- [80] D. B. Ward and T. D. Abhayapala. Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers.
- [81] E. G. Williams. Fourier Acoustics. Sound Radiation and Nearfield Acoustic Holography. 1999.
- [82] J. M. Zm̄olnig, W. Ritsch, and A. Sontacchi. The IEM-Cube, 2003.
- [83] E. Zoboli. Design, construction and test of three-dimensional microphone arrays employing MEMS and A2B technologies. Tesi di laurea magistrale in Music and Acoustic Engineering - Politecnico di Milano, Academic Year: 2021-22.
- [84] F. Zotter. Ambisonics (acoustic Holography and Holophony). IEM, 2017.
- [85] F. Zotter and M. Frank. All-Round Ambisonic Panning and Decoding. Audio Engineering Society.
- [86] F. Zotter and M. Frank. Ambisonic Decoding with Panning-invariant Loudness on Small Layouts (AllRAD2). volume 9943. 2018.
- [87] F. Zotter and M. Frank. Ambisonics. A Practical 3D Audio Theory for Recording,

- Studio Production, Sound Reinforcement, and Virtual Reality. Technical report, Springer International Publishing, 2019.
- [88] F. Zotter, A. Sontacchi, and M. Frank. The Virtual t-design Ambisonics-rig Using VBAPAllRaD. 2010.

List of Figures

1.1	B-Format of the first order Ambisonic, picture from: [5]	5
1.2	Spherical harmonic 3-D representation up to 5-th order. Picture taken from [31].	7
1.3	Brief Ambisonics system scheme.	8
1.4	T-design configurations from [11].	13
1.5	From [80] p. 4: Normalized truncation error as a function of kr for various reproduction orders N .	16
1.6	<i>SoundField</i> microphone structure, picture from [30].	18
1.7	Sinusoidal sweep time trend, $q = 6$.	20
1.8	Octave-band filtered Impulse response and related Energy Decay Curve at 1000Hz.	26
1.9	Sonosfera® graphical rendering from [14].	27
1.10	IEM-Cube, [7].	28
2.1	em64 Eigenmike ® microphone, picture from [4].	34
3.1	Array layout views. The listening position is depicted in red, the TV screen is just above speaker number five.	38
3.2		39
3.3	Measurement spatial setup	40
3.4	Measurement device setup.	40
3.5	Spectrum of the computed pre-distorted exponential sine sweep.	41
3.6	Comparison between the spectrum of the original impulse response with the pre-distorted one.	41
3.7	Procedure scheme for the analysis computation.	43
3.8	Example of filtered impulse response in 62.5 Hz octave band and its related energy decay curve.	45
3.9	Spectra comparison. Configuration 1: original configuration. Configuration 2: TV covered. Configuration 3: TV removed, Configuration 4: TV removed and door covered. Configuration 5: seat removed.	46

3.10 Spectra comparison. In blue: original configuration. In orange: TV and door covered by absorbing polyester fiber panel 5 cm thick for the TV and 2,5 cm thick for the door.	47
3.11 Time trends comparison. In black: original configuration. In blue: TV and door covered by absorbing polyester fiber panel 5 cm thick for the TV and 2,5 cm thick for the door.	47
3.12 Spectra of all the single impulse responses with the first configuration (original one). Above each image, the number of the related excited speaker is depicted.	48
3.13 Impulse obtained by linearly summing the individual subwoofers response. Time domain and spectrum.	49
3.14 Balloon popping spectra measured from the central position.	50
3.15 Balloon popping time trend measured from the central position.	51
3.16 Center position acoustic results: T20	52
3.17 Center position acoustic results: C50	53
3.18 Center position acoustic results: Early Decay Time	54
 4.1 The Eigenmike64 ® in central position.	58
4.2 Filter matrix H	59
4.3 Impulse Response matrix extraction.	59
4.4 FIR obtained for the speaker number 6 measuring in 40 cm left position from central point.	63
4.5 Spatial Correlation for each order measuring in central position. A 1/6 octave band smoothing is applied.	66
4.6 Spatial Correlation for each order measuring in lateral position. A 1/6 octave band smoothing is applied.	67
4.7 Level Difference for each order measuring in central position. A 1/6 octave band smoothing is applied.	69
4.8 Level Difference for each order measuring in lateral position. A 1/6 octave band smoothing is applied.	70
4.9 \hat{Y}_{00} . Central measurements.	71
4.10 \hat{Y}_{-11} . Central measurements.	72
4.11 \hat{Y}_{11} . Central measurements.	72
4.12 \hat{Y}_{-22} . Central measurements.	72
4.13 \hat{Y}_{22} . Central measurements.	73
4.14 \hat{Y}_{00} . Left measurements.	73
4.15 \hat{Y}_{-11} . Left measurements.	73

4.16 \hat{Y}_{11} . Left measurements.	74
4.17 \hat{Y}_{-22} . Left measurements.	74
4.18 \hat{Y}_{22} . Left measurements.	74
4.19 Orientation for the next pictures, in red the microphone capsules while the blue circle stands for the positions of the numbered speakers	75
4.20 Comparison for central position measurements.	76
4.21 Comparison for central position measurements.	76
4.22 Comparison for lateral position measurements.	77
4.23 Comparison for lateral position measurements.	77
4.24 4-th order decoding of the measurements in central position in 240 points t-design (t=21).	79
4.25 4-th order decoding of the measurements in central position in 240 points t-design (t=21).	80
4.26 4-th order decoding of the measurements in central position in 240 points t-design (t=21).	80
4.27 4-th order decoding of the measurements in central position in 240 points t-design (t=21).	81
4.28 4-th order decoding of the measurements in lateral position in 240 points t-design (t=21).	81
4.29 4-th order decoding of the measurements in lateral position in 240 points t-design (t=21).	82
4.30 4-th order decoding of the measurements in lateral position in 240 points t-design (t=21).	82
4.31 4-th order decoding of the measurements in lateral position in 240 points t-design (t=21).	83

List of Tables

2.1	Table from p.12 of [59].	36
3.1	Acoustic parameters obtained through subwoofers excitation for 31 Hz octave band.	55
4.1	Delay table, central position.	62
4.2	Delay table, lateral position.	62
4.3	FIR parameters for speaker 6 measuring in 40 cm left position from central point.	64

List of Symbols

Variable	Description	SI unit
λ	Wavelength	m
ω	Radial frequency	rad
θ	Azimuth	Deg
ϕ	Elevation	Deg
n	Ambisonics order	
m	Ambisonics degree	
N	Truncation Ambisonics order	
P_{mn}	Legendre function	
k	Wave number	
\mathbf{Y}	$Y_{nm}(\theta, \phi)$	Numerical spherical harmonic functions
$\hat{\mathbf{Y}}$	$\hat{Y}_{nm}^\omega(\theta, \phi)$	Measured spherical harmonic functions
$\hat{\mathbf{Y}}_{TOT}(\omega)$	\hat{Y}_{TOTnm}^ω	Measured spherical harmonic coefficients
W	0-th order omnidirectional component in B-format	
X	1-st order bidirective component in B-format	
Y	1-st order bidirective component in B-format	
X	1-st order bidirective component in B-format	
r	Radius, distance from the center	m
$p(x, y, z, t)$	Pressure	Pa
∇^2	Laplacian operator	
L	Number of speakers	
U	Number of source directions	
q	Sweep rate	$\frac{\text{Hz}}{\text{s}}$
Q	Q-factor	
EDC	Energy decay curve	$\frac{\text{dB}}{\text{s}}$
EDT	Early decay time	s

$j_n(kr)$	Spherical Bessel functions of the first kind	
$T20$	Reverberation time	s
$C50$	Clarity index	dB
D	Decoding matrix	
$S(\omega)$	$S^\omega(\theta, \phi)$	Source signals
$B(\omega)$	Spherical harmonic coefficients	
$SC_n(\omega)$	Spatial Correlation	
$LD_n(\omega)$	Level Difference	
H	$H_{nm,j}^\omega$	Encoding filter matrix
f_{max}	Aliasing frequency of Eigenmike64	Hz
d	Microphone array capsule spacing	m
r_e	Energy localization vector	
r_v	Velocity localization vector	
G_l	Loudspeaker gains	
$E(t)$	Response energy	
$\nu(f)$	Frequency regularization parameter for Kirkeby filter	

Acknowledgements
