SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

EXECUTIVE SUMMARY OF THE THESIS

Characterisation of an Ambisonics Reproduction Room

LAUREA MAGISTRALE IN MUSIC AND ACOUSTIC ENGINEERING - INGEGNERIA DELL' INFORMAZIONE

Author: SOFIA PARRINELLI

Advisor: PROF. FABIO ANTONACCI

Co-advisor: ENRICO ZOBOLI, DOTT. MIRCO PEZZOLI

Academic year: 2022-2023

1. Introduction

Achieving a truly immersive sound field necessitates replicating the complexity of a sound event and this procedure poses significant challenges. Currently, the achievable levels of sound fidelity are extremely high, providing one or more listeners the impression of being immersed in a realistic sound environment. The ability to design an optimal immersive sound field holds diverse applications such as virtual reality, entertainment applications, augmented reality, music reproduction, artistic performances and simulations. This thesis, conducted in conjunction with ASK Industries, aims to guide the assessment and improvement of a room dedicated to immersive listening compatible with Ambisonics. This thesis will focus on a numerical assessment rather than relying on a subjective 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. The second step of the analysis is the evaluation of the reproduction of the spherical harmonics within the site.

Level Difference and Spatial Correlation are two metrics proposed in [5] for the evaluation of the encoding quality of a microphone array. In this study, the mentioned metrics are tested for the assessment of the reproduction system, encompassing both the room itself and its speaker array. These metrics, in fact, have never been tested for audio reproduction except than in this thesis project.

2. Ambisonics Basics

Ambisonics is a surround system based on spherical harmonic expansion as discretization of the sound field. This discretization consists in transforming the wave equation into the spherical coordinate system, characterized by pressure p defined at azimuth θ, elevation ϕ and radius distance r for each point from a central point O, leading to the Fourier-Bessel decomposition:

$$p(r, \theta, \phi) = \sum_{n=0}^{\infty} 4\pi i^n j_n(kr) \sum_{-n \leq m \leq n} B_{nm} Y_{nm}(\theta, \phi), \quad (1)$$

with k referred to the wave number, 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 centered 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 depict the contribution to the sound field as a function of the distance r from the center O. These components provide information about the pressure

field ($B_{0,0}$ component) and its derivatives at different orders. In practice, the summation will be done, for the encoding process, up to a finite number of order N, leading to a number of spatial components equal to $(N+1)^2$.
The spherical harmonics are basis functions defined on the unit sphere as:

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

P_{nm} are the associated Legendre functions having order n and degree m.

2.1. Spatial Encoding and Decoding

The spherical harmonic decomposition of the sound field is performed over the virtual sound sources through microphone arrays. This process provides the encoding equations, these are expressions of B_{nm}, frequency dependent spatial components. If the source $\mathbf{S}(\omega)$ is assumed to be far enough, the contribution can be intended as a plane wave and the encoding process the following:

$$\mathbf{B}(\omega) = \mathbf{S}(\omega) \cdot \mathbf{Y}, \quad (3)$$

being:

$$\mathbf{B} = [B_{0,0}, B_{1,-1}(\omega), ..., B_{n,m}(\omega), ..., B_{NM}(\omega)], \quad (4)$$

$$\mathbf{S}(\omega) = [S^{\omega}(\theta_1,\phi_1), ..., S^{\omega}(\theta_u,\phi_u), ..., S^{\omega}(\theta_U,\phi_U)], \quad (5)$$

$\mathbf{Y}$ is the set of spherical harmonic functions $Y_{nm}(\theta_u,\phi_u)$ with $n < N, -m \leq n \leq m$ and (θ_u,ϕ_u) the sources directions.
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 the wavefront curvature.
The natural sound scene will be, therefore, decomposed by the encoding, with appropriate microphones, while the reassembly from spherical harmonics into the new sound field, generated by the reproduction system, is handled by decoding. The decoder aims of recreating the acoustic event at the center of the loudspeaker array, being this area the listening position called *sweet spot*. The Ambisonics mapping is, therefore, the distribution of the signals to the loudspeaker array through a matrix operation process, specifically combining the signals with real weighting gains:

$$\mathbf{S}(\omega) = \mathbf{D}(\omega) \cdot \mathbf{B}(\omega), \quad (6)$$

with $\mathbf{S}(\omega)$ the component, for the specified frequency, of the vector of emitted signals:

$$\mathbf{S}(\omega) = [S^{\omega}(\theta_1,\phi_1), ..., S^{\omega}(\theta_l,\phi_l), ..., S^{\omega}(\theta_L,\phi_L)], \quad (7)$$

$\mathbf{D}(\omega)$ is the decoding matrix, its dimensions are L x $(N+1)^2$, L being the number of speakers). This matrix either can be frequency dependant, this possibility depends on the chosen method for computing the elements of the decoding matrix. In fact, the decoding matrix can be defined in various ways: in case regular polygonal loudspeaker arrays are provided, more straightforward computations can be followed in order to design the proper decoding matrix, while advanced methods can guarantee more adaptability and evenly sound distribution on a large region [3]. The majority of traditional Ambisonic decoder designed for all orders $n \leq N$, necessitate a minimum of

$$L = (N+1)^2 \quad (8)$$

loudspeakers evenly distributed to cover a spherical space. Additionally, a non-reverberant site is required to achieve accurate audio reproduction quality of a surround system.

3. Approach

The room to be studied is acoustically properly treated and its geometrical structure has been designed specifically to restrict resonant modes providing non-parallel walls. The spherical array of speakers consists of 25 tweeters, 25 woofers and two subwoofers. These latter are positioned under the seat, which is centered in the middle of the spherical array of speakers.

3.1. Room Acoustic Characterization

The objective is to analyze the spectrum of the room impulse response (IR), achieved by linearly summing all the individual IRs generated through the sequential excitation of each loudspeaker, finding the useful acoustical parameters. A pre-distorted exponential sine sweep was used for the measurements and recorded through an omnidirectional measurement microphone placed in the central position. The predistortion leads to the exclusive evaluation of the room itself, ensuring a balanced frequency excitation. Five room configurations were tested, involving the placement of absorbing panels to cover the TV screen and the door (two primary sources of potential reflections), to determine the most effective arrangement. A second analysis was conducted to study the room itself without any influence from the speakers, for the aim balloons were popped and the IR was recorded with an omnidirection microphone. The signal under study was properly cut and the necessary parameters, by employing the method elucidated in ISO 3382-1:2009, have been extracted using reverse cumulative trapezoidal integration to estimate the decay curve. The extracted parameters are: Reverberation Time ($T20$), Speech clarity index (C_{50}), Early Decay Time (EDT).

3.2. Spherical Harmonics Reproduction Assessment

Procedure

A key characteristic of a diffuse field is 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.
The proposed metrics aim to comprehensively evaluate the spatial configuration, encompassing both the playback array and the room structure, for the recreation of a diffuse sound space. Issues such as reflections, speaker configuration limitations, and system failures can result in inadequate mapping of spherical harmonics.
The microphone used for these measurements, Eigenmike64®, has to be considered ideal within the frequency ranges described in [4].
Two positions were tested in order to conduct the evaluation even in a decentralized position: central position and 40 centimeters left of center position, both at ear's height. A 10 seconds long exponential sine sweep was employed. The entire spherical woofer - tweeter array was employed, exciting the speakers one at a time while the subwoofers were not employed in this evaluation.
The encoding filters provided with the microphone array, which are specifically designed to retrieve the spherical harmonic coefficients, are stored in the encoding matrix $\mathbf{H}$ composed by the set of $H^{\omega}_{nm,j}$. The IR matrix was recorded and stored as matrix $\mathbf{S}(\omega)$. For the extraction of measured spherical harmonics $\hat{\mathbf{Y}}(\omega)$, each IR has been convolved with its proper filter in frequency domain and the summation for each microphone in time domain has been performed, as depicted in [2].

$$\hat{Y}^{\omega}_{nm}(\theta_l, \phi_l) = \sum_{j=1}^{J} S^{\omega}_{j}(\theta_l, \phi_l) \cdot H^{\omega}_{nm,j}, \tag{9}$$

here w denotes the frequency, j the capsule number, up to the amount of capsules J, and (θ_l, ϕ_l) the source direction up to $l = L$. The dimension of the obtained matrix is, therefore:

$$\hat{\mathbf{Y}}(\omega) = 25 \text{ x } 49 \text{ spherical signals } \hat{Y}^{\omega}_{nm}(\theta_l, \phi_l),$$

considering 25 sound sources directions and 49 spherical harmonics.
Two equalisations were carried out for the two measurement positions. 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: $\hat{Y}^{\omega}_{0,0}(\theta_l, \phi_l)$.

Metrics Definition

The two metrics utilized for the evaluation of the spherical harmonic recreation in the site are Spatial Correlation (SC) and Level Difference (LD). These performance metrics are order-dependent, established for the evaluation of the microphone encoding [1, 5] while, however, their use with regard to the evaluation of audio reproduction is experimental.
SC 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 2, is compared to the measured spherical harmonics $\hat{\mathbf{Y}}$ functions obtained from the encoded sound field. After averaging the metrics across all the L directions, further averaging is 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:

$$SC_{nm}(\omega) = \sum_{l=1}^{L} \left(\frac{\left(\hat{Y}^{\omega}_{nm}(\theta_l, \phi_l)\right)^T}{\sqrt{\left(\hat{Y}^{\omega}_{nm}(\theta_l, \phi_l)\right)^T \cdot \hat{Y}^{\omega}_{nm}(\theta_l, \phi_l)}} \cdot \frac{Y_{nm}(\theta_l, \phi_l)}{\sqrt{(Y_{nm}(\theta_l, \phi_l))^T \cdot Y_{nm}(\theta_l, \phi_l)}} \right), \tag{10}$$

$$SC_n(\omega) = \frac{1}{2n+1} \sum_{m=-n}^{n} |SC_{nm}(\omega)|, \tag{11}$$

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.

$$LD_{nm}(\omega) = \frac{1}{L} \sum_{l=1}^{L} \frac{|Y_{nm}(\theta_l, \phi_l)|^2}{\hat{Y}^{\omega}_{nm}(\theta_l, \phi_l) \cdot \hat{Y}^{\omega}_{nm}(\theta_l, \phi_l)^*}, \tag{12}$$

$$LD_n(\omega) = -10 \log \left[\frac{1}{2n+1} \sum_{m=-n}^{n} LD_{nm}(\omega) \right], \quad (13)$$

An ideal perfect reconstruction, therefore, will lead to $SC_n(\omega) = 1$ and $LD_n(\omega) = 0$ dB. A normalization is applied to $\mathbf{Y}$ and $\hat{\mathbf{Y}}$ with respect to the maximum value of each respectively in order to map the values in the complex unit sphere and compare them properly.

4. Results

4.1. Room Acoustic Results

In the following bar graphs the detected acoustic parameters are displayed, the sweep excitation results are compared with the balloon excitation ones in order to have a more accurate and complete feedback of the acoustic performances. The conclusive observation to the analysis performed in all the tested configurations is that optimal acoustic parameters are effectively achieved even in the original situation, therefore, the decision is not to alter any room predisposition and leave it as is. For immediacy, just the original configuration results are displayed, here in octave bands. For the 31 Hz octave band the obtained parameters, computed through subwoofers excitation, are: T20 of 0.2 seconds, C50 of 10 dB, and EDT of 0.34 seconds.

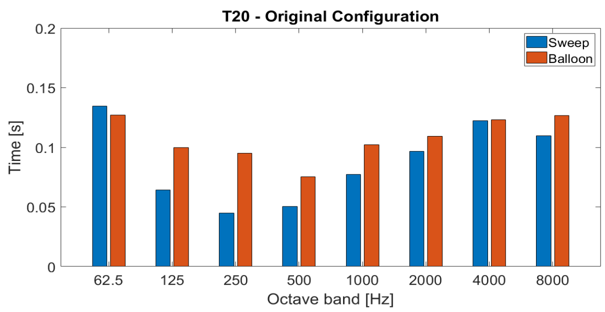


Figure 1: Original configuration: T20

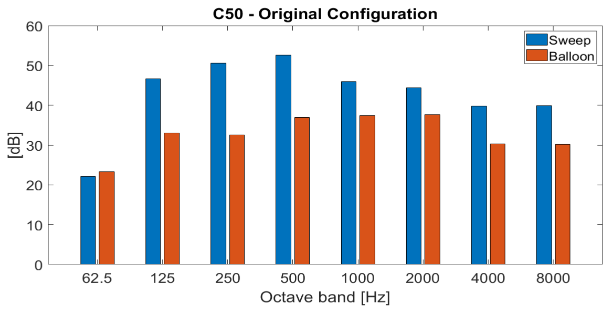
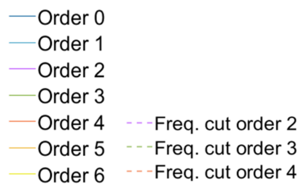


Figure 2: Original configuration: C50

4.2. Spherical Harmonic Reproduction Results

SC and LD are depicted up to order 6, however, it is important to note that the Virtual Room is compatible with sound reconstruction only up to the fourth order. The vertical lines highlight the cutoff frequencies of the microphone array, as illustrated in [4], therefore, for that particular order, the trend should be evaluated for frequencies higher than the cutoff. The legend exposed here is valid for all the following images:



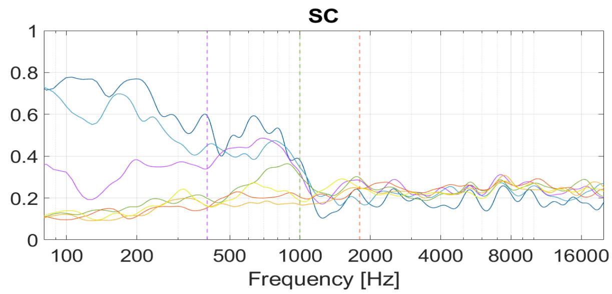


Figure 3: Non equalized central measurement.

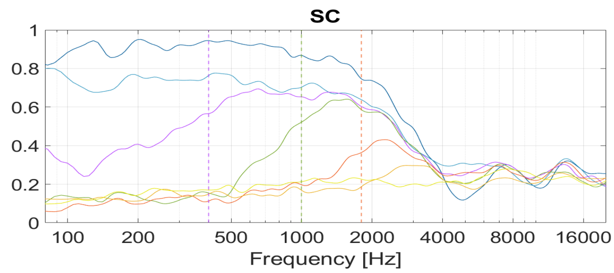


Figure 4: Equalized central measurement.

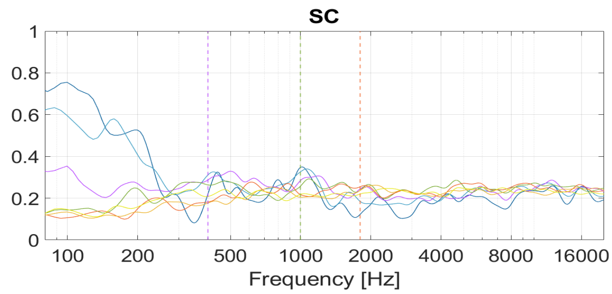


Figure 5: Non equalized left measurement.

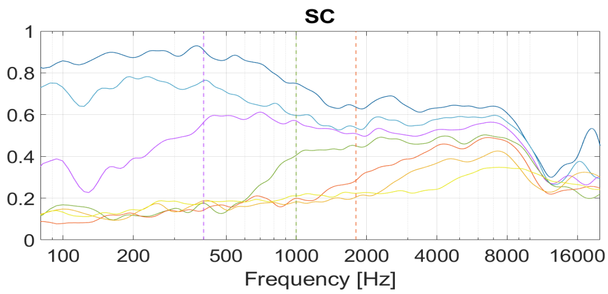


Figure 6: Equalization left measurement.

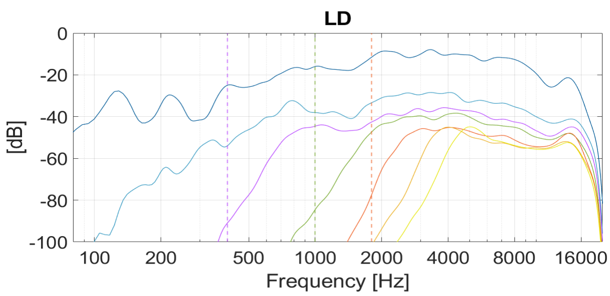


Figure 7: Non equalized central measurement.

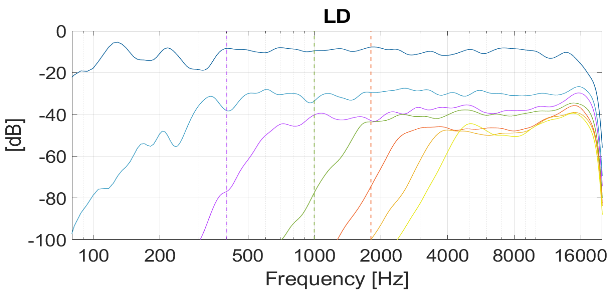


Figure 8: Equalized central measurement.

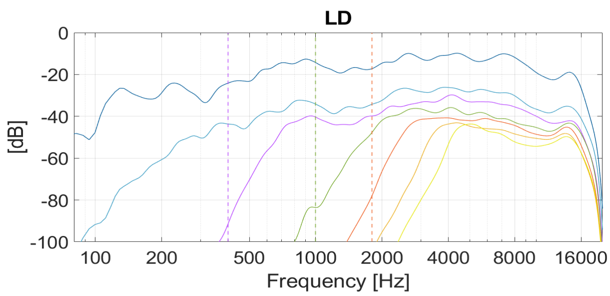


Figure 9: Non equalized left measurement.

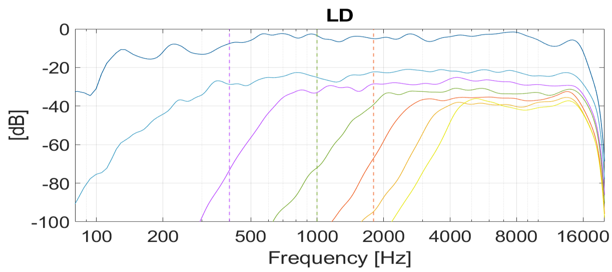


Figure 10: Equalized left measurement.

A rendering of the measured spherical harmonics directivity is proposed. Polar patterns are displayed for octave bands from the central position measurements, enabling the visualization of the horizontal section of the described field and allowing a visual comparison with the directivity of ideal spherical harmonics.

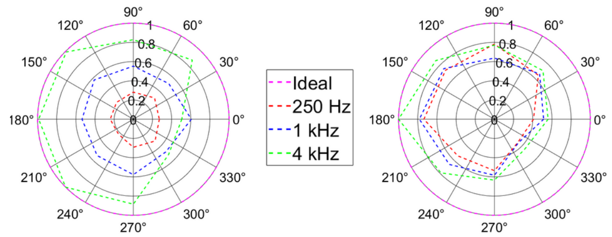


Figure 11: $\hat{Y}_{00}$. Left: no equalized, right: equalized.

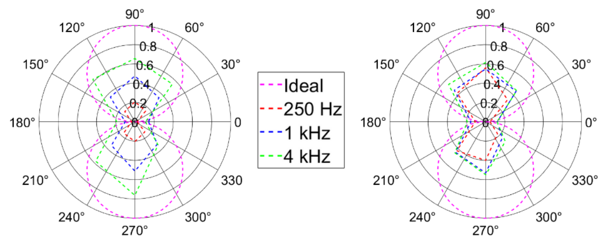


Figure 12: $\hat{Y}_{1-1}$. Left: no equalized, right: equalized.

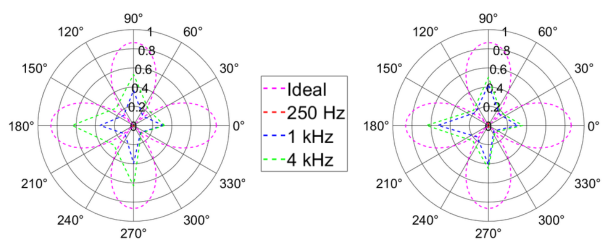


Figure 13: $\hat{Y}_{2-2}$. Left: no equalized, right: equalized.

Finally, the rendering obtained in the central position through the sampling decoder process is shown. This visualization can be useful, as the previous one, to compare the different resulting levels of uniformity of the sound field, which, in this case, is reconstructed by a decoder process. The desired spatial layout is computed by using a t-design with $t = 21$, providing an amount of 240 points. Starting from the overall matrix of the spherical harmonics, obtained by summing all the contributions (9) from all the individual 25 directions l:

$$\hat{Y}^{\omega}_{TOT_{nm}} = \sum_{l=1}^{25} \hat{Y}^{\omega}_{nm}(\theta_l, \phi_l), \tag{14}$$

the decoding process is computed applying the formula 6, where the computed

$$\hat{\mathbf{Y}}_{\mathbf{TOT}}(\omega) = [\hat{Y}^{\omega}_{TOT_{0,0}}, \quad \hat{Y}^{\omega}_{TOT_{1,-1}}, \quad \dots, \quad \hat{Y}^{\omega}_{TOT_{NM}}] \tag{15}$$

corresponds to $\mathbf{B}(\omega)$.
In this case, the frequency ranges reported in [4] are respected for the decoding matrix computation and a fourth order decoding has been computed.

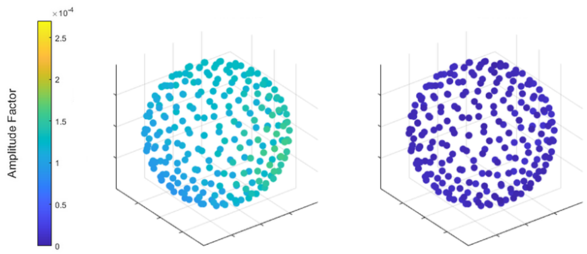


Figure 14: 500 Hz octave band decoding. Left: no equalized, right: equalized.

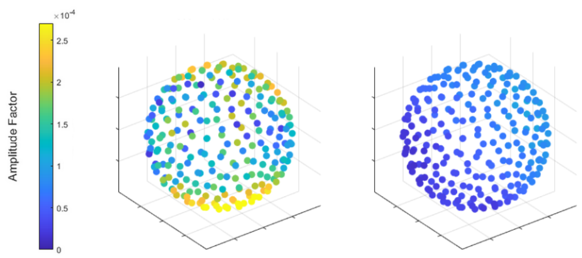


Figure 15: 2000 Hz octave band decoding. Left: no equalized, right equalized.

5. Conclusions

This thesis project presents an approach for the characterization and assessment of a room compatible with Ambisonics audio reproduction.
The acoustic performance of the treated room was studied by extrapolating the useful acoustic parameters over a wide frequency range, demonstrating that the site presents excellent sound absorption properties and can definitely be classified as a semi-anechoic room. Once the room's prerequisites for an immersive audio reproduction were judged fulfilled, the reproduction of spherical harmonics was tested with the use of Spatial Correlation and Level Difference metrics. These two metrics have been used in past studies for the evaluation of microphone arrays, intending the external reproduction system ideal. 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 and lateral position 40 cm left of center, both in the absence of tuning and applying specific tuning for each measurement position to evaluate a possible improvement. The values of both metrics are significantly improved with the use of equalization. The metrics behave, therefore, as desired, proving that they can be used for the Ambisonics evaluation of audio reproduction.

References

[1] H. Gamper A. Politis. Comparing Modeled and Measurement-based Spherical Harmonic encoding Filters for Spherical Microphone Arrays. 2017.

[2] A. Farina D. Pinardi. Metrics for Evaluating the Spatial Accuracy of Microphone Arrays.

[3] M. Frank F. Zotter. All-Round Ambisonic Panning and Decoding. Audio Engineering Society.

[4] LLC Mh acoustics. Eigenbeam Data Specification.

[5] J. Daniel S. Moreau and S. Bertet. 3d Sound Field Recording with Higher Order Ambisonics – Objective Measurements and Validation of a 4th Order Spherical Microphone. Paris, 2006.