UNIVERSITÀ DI PARMA

UNIVERSITA' DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN
INGEGNERIA INDUSTRIALE

CICLO XXXII

Recording, Analysis and Playback of Spatial Sound Field using Novel Design Methods of Transducer Arrays

Coordinatore:
Chiar.mo Prof. GIANNI ROYER CARFAGNI

Tutore:
Chiar.mo Prof. ANGELO FARINA
Co-tutore:
Chiar.mo Prof. VILLE PULKKI

Dottorando: DANIEL PINARDI

Anni Accademici 2016/2017 – 2017/2018 – 2018/2019

a Mamma e Papà

e

a Laura

Abstract

Nowadays, a growing interest in the recording and reproduction of spatial audio has been observed. With virtual and augmented reality technologies spreading fast thanks to entertainment and video game industries, also the professional opportunities in the field of engineering are evolving. However, despite many microphone arrays are reaching the market, most of them is not optimized for engineering or diagnostic use and remains mainly confined to voice and music recordings.

In this thesis, the design of two new systems for recording and analysing the spatial distribution of sound energy, employing arrays of transducers and cameras, is discussed. Both acoustic and visual spatial information is recorded and combined together to produce static and dynamic colour maps, with a specially designed software and employing Ambisonics and Spatial PCM Sampling (SPS), two common spatial audio formats, for signals processing.

The first solution consists in a microphone array made of 32 capsules and a circular array of eight cameras, optimized for low frequencies. The size of the array is designed accordingly to the frequency range of interest for automotive Noise, Vibration & Harshness (NVH) applications. The second system is an underwater probe with four hydrophones and a panoramic camera, with which it is possible to monitor the effects of underwater noise produced by human activities on marine species.

Finite Elements Method (FEM) simulations have been used to calculate the array response, thus deriving the filtering matrix and performing theoretical evaluation of the spatial performance. Field tests of the proposed solutions are presented in comparison with the current state-of-the-art equipment.

The faithful reproduction of the spatial sound field arouses equally interest. Hence, a method to playback panoramic video with spatial audio is presented, making use of Virtual Reality (VR) technology, spatial audio, individualized Head Related Transfer Functions (HRTFs) and personalized headphones equalization.

The work in its entirety presents a complete methodology for recording, analysing and reproducing the spatial information of soundscapes.

Summary

Index of figures

Index of tables

List of Abbreviations

ACN	Ambisonics Channel Number
ANC	Active Noise Cancelling
BEM	Boundary Elements Method
dB	[decibel]
DoA	Direction of Arrival
DTFT	Discrete Time Fourier Transform
EM	Eigenmike32™
EMIB	Eigenmike32™ Microphone Interface Box
ENC	Engine Noise Cancelling
ESS	Exponential Sine Sweep
FEM	Finite Elements Method
FFT	Fast Fourier Transform
FOA	First Order Ambisonics
FoV	Field of View
fps	Frame per Second
fs	Sampling Frequency
HMD(s)	Head Mounted Display(s)
HOA	Higher Order Ambisonics
HRTFs	Head Related Transfer Functions
HSA	Head-Shaped Array
IFFT	Inverse Fast Fourier Transform
IV	Intensity Vector
IR(s)	Impulse Response(s)
LD	Level Difference
MEMS	Micro Electro-Mechanical System
MuSiC	MUltiple SIgnal Classification
MVDR	Minimum Variance Distortion-less Response
NVH	Noise, Vibrations & Harshness
PML	Perfectly Matched Layer
PSD	Power Spectral Density
PWD	Plane Wave Decomposition
RNC	Road Noise Cancelling
SH	Spherical Harmonics
SC	Spatial Correlation
SF	Speaker Feed
SF2BIN	Speaker Feed to Binaural
SH	Spherical Harmonics
SH2BIN	Spherical Harmonics to Binaural
SH2SF	Spherical Harmonics to Speaker Feed
SIR	Spatial Impulse Response
SLM	Sound Level Meter
SPL	Sound Pressure Level
SPS	Spatial PCM Sampling
VR	Virtual Reality
WNG	White Noise Gain
WFS	Wave Field Synthesis

1. Introduction

The recording and reproduction of spatial properties of sound field is becoming more and more important mainly thanks to videogames and cinema industries; with a growing number of immersive applications, the professional market is benefiting from the development of this new technology too. Low cost solutions for panoramic video and spatial audio reproduction are now widely employed: enhanced telepresence and teleconferencing, remote maintenance and assistance or even lessons and training sessions. Such applications make use of wired or wireless Head-Mounted Displays (HMDs) to reproduce a 360° visual scene and to process the spatial information of the sound field for reconstructing the auditory sensation, keeping the natural perception of the original sound scene.

Despite a large number of microphone arrays for recording spatial audio came to the market in the last years, most of them are not suitable for scientific use. Commercial products, generally built for music recordings, do not provide a satisfactory spatial resolution and often lack of a video recording system. Indeed, professional opportunities in the field of engineering offered by these new technologies are large: more precise evaluation of environmental impact of noise sources, performance analysis and effectiveness of active and passive systems for noise reduction, advanced design for architectural acoustics, just to cite some examples.

This is the situation “in air”, but in case of underwater applications, the lack of professional hydrophone arrays is even greater. It is a pity, because single channel hydrophones can measure only one scalar value, the sound pressure, a physical quantity perceived by mammals, as humans. However, a large number of marine species are sensitive to another quantity, the particle velocity, which can be measured only with an array of hydrophones, as they are not provided with a membrane auditory system. Unfortunately, this means that most of the previous studies and thresholds set for underwater noise pollution are wrong.

Existing software for analysing the sound energy distribution by means of colour mapping miss of some capabilities, which instead are fundamental to perform a scientific diagnosis. First, when analysing a noise recording a constant scaling must be adopted: in fact, otherwise, the information of the relative amplitudes is lost and all contributions are considered equal, which is in general not true. In addition, a proper calibration of the system is necessary in order to show the real sound pressure level recorded. Some existing solutions are not capable to handle a background image to plot the colour map on, which makes the whole analysis useless. Some others instead can superimpose the colour map on a static picture, but cannot use a video for the background, which is instead necessary if moving sources are present in the sound scene or around it.

The capability to record the video together the with spatial audio is needed also for a complete and faithful reproduction: 360° panoramic video with a colour map of the sound pressure level superimposed can be created for being played on visors, making it possible to associate visual and auditory perceptions. Moreover, the natural perception of being inside the environment is kept unchanged by processing the spatial audio with individualized sets of HRTFs.

In this thesis, the design of two microphone arrays specifically built for recording and analysing the spatial information of sound field will be discussed. The first solution features a microphone array of 32 capsules and an array of cameras to record panoramic video. The size of the array is designed accordingly to the frequency range of interest for automotive NVH applications, $50\ Hz - 2\ kHz$. It was decided to design the array having the size similar to a human head, hence the name "Head-Shaped Array" (HSA). This combines the advantages of working well at low frequencies and being compact enough to employ it easily also inside small environment, such as car cockpits. The second system is made of four hydrophones and a panoramic video recording system encapsulated inside an underwater case: this new probe can record the spatial information underwater and it is very useful to analyse the environmental impact of human generated noise on marine species.

A technique to derive conversion filters with FEM simulations, performed in COMSOL Multiphysics, will be presented and results will be discussed with theoretical evaluation and field test.

A complete set of software has been coded in Matlab for processing signals and generating the colour maps, supporting both spatial formats, Ambisonics and SPS: colour map image with background picture, colour map video with background picture and colour map video with background video can be produced.

A cross-correlation filtering tool, which now is not available in any other existing solution, has been coded. If a proper hardware setup is employed, it is possible to record some additional signals together with the microphone array. In the post-processing, a cross-correlation analysis between the SPS format and these additional signals is performed: the result is a filtered colour map, where only the noise coherent with the chosen reference is plotted. This is very useful when studying the individual contribution of each source if more than one is emitting noise at the same time, as it happens usually in the real world.

In conclusion, a solution to measure, process and employ individualized HRTFs for playing back 360° panoramic video with spatial audio on visors or wired HMDs will be presented, making it possible to overcame the limitations of generalized HRTFs and test different sets of HRTFs in order to choose the most suitable for each individual.

2. Recording of the acoustic spatial information

Despite most acousticians still work with single channel microphones and measure only the sound pressure $[Pa]$, a scalar quantity, there is another vector quantity that should not be ignored: the particle velocity $[m/s]$. Both describe some characteristics of the sound field and for this reason, they are also known as "field quantities". The first one tell us how strong is the sound, i.e. the amplitude of the wave, whereas the second one carries the spatial information and tell us the direction of arrival (DoA). Thanks to this information, it is possible to localize the sound source.

An omnidirectional microphone measures the pressure in one point of the space, whilst the response of a pressure gradient microphone, that is a velocity microphone, corresponds to the difference of pressure in two points of the space [1]. With a variable-pattern microphone, different mixtures of pressure and velocity can be obtained, by combining signals opportunely. In Table 1, most common combinations are shown.

Pressure P	Particle velocity G	Polar pattern
1	0	Omnidirectional
> 0.5	< 0.5	Sub-cardioid
0.5	0.5	Cardioid
0.33	0.66	Super-cardioid
0.25	0.75	Hyper-cardioid
0	1	Figure-of-eight

Table 1: Common microphone polar patterns

This leads to the conclusion that to know the complete spatial information in a point of the space at least four signals are required: the pressure and the three Cartesian components of the particle velocity. At the same time, it means that to record the complete spatial information in a point of the space at least four microphones are necessary and a proper combination of the four channels is required to get the sound pressure P and the particle velocity components PV_x, PV_y, PV_z.

2.1. Microphone arrays

A microphone array is a system capable of recording the spatial information of the sound field, by employing more than one transducer. First examples date back to the fifties, when couples of omnidirectional microphones were used for stereophonic recordings [2], [3], [4]. Different techniques were developed, such as XY or ORTF, but none of them was capable to record the full spatial information, because only two

microphones were used and the listener could perceive the DoA of sounds only in one plane.

From that time, the technology developed enormously: four-channel arrays were built firstly and recently, thanks to the digital circuitry, several arrays with much more capsules reached the market. The first four-channel array, the Soundfield™ microphone, was built by Gerzon and Craven in 1975 [5]; in Figure 1 (left) the Soundfield™ microphone is shown together with some other arrays of four capsules: DPA4™ (centre) and Sennheiser Ambeo™ (right).

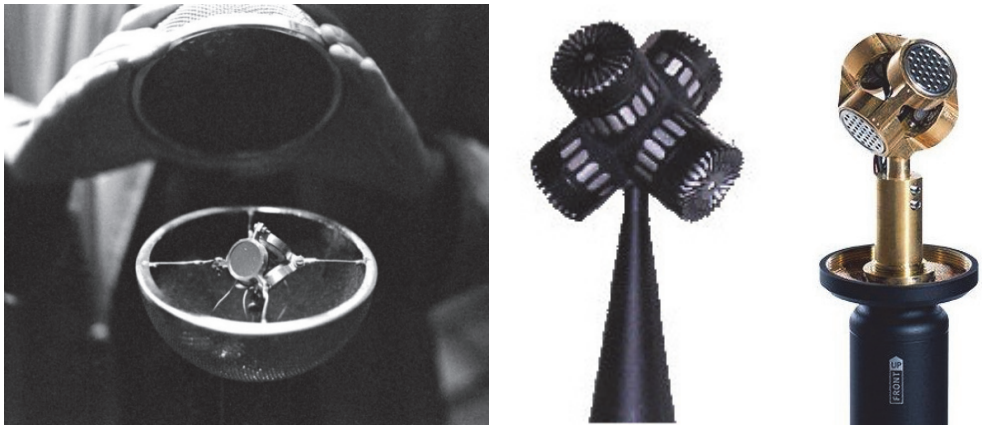


Figure 1: Soundfield™ microphone (left), DPA4™ (middle) and Sennheiser Ambeo™ (right)

Figure 2 shows three spherical microphone arrays the author worked with, which have more than four capsules: the Zylia™, with 19 MEMS transducers, the Eigenmike32™, with 32 capsules and the Bruel&Kjaer array, which integrates 50 capsules and 12 lenses for video recording.

Figure 2: Zylia™ microphone (left), Eigenmike32™ (middle) and Bruel&Kjaer™ array (right)

Whatever is the array, all of them employ the capsules to sample the sound pressure in a certain number of points of the space, producing a multi-channel signal called “A-format”. Then, the spatial information contained into the A-format is encoded: depending on the output format adopted, the result will be a different

formulation of the spatial information, with some useful properties that permit to process and manipulate it, i.e. rotations, in a fast and easy way. In the following paragraphs, the two most common spatial formats will be briefly explained: Ambisonics, or "B-format" (paragraph 2.1.1, page 5) and Spatial PCM Sampling (SPS), or "P-format" (paragraph 2.1.2, page 6).

2.1.1. Ambisonics Theory

The Ambisonics theory has been developed in 1975, thanks to the pioneering work of the British scientist Michael Gerzon, [6], [7]. Ambisonics is a method for representing the complete acoustical spatial information in one point of the space with a reduced number of signals, calculated from the real signals captured by the capsules of the array at the recording point. These are conceptually coincident virtual microphones with directivity patterns corresponding to a set of Spherical Harmonics (SH), mathematical functions of various orders with orthonormal properties; in [8] an explicit formulation of the SH functions up to order five has been made available. Hence, the target of this theory consists in obtaining the SH expansion starting from the signals captured by the capsules, an operation that goes under the name of *Ambisonics encoding* or "A-2-B" format conversion. The order of the Ambisonics expansions is related to the number of channels of the array by the equation:

$$N_{ch} = (o+1)^2, \tag{1}$$

where o is the Ambisonics order, from 0 to n.

The Soundfield™ microphone, capable of recording four signals, was therefore a First Order Ambisonics (FOA) array. The Ambisonics expansion of order 1 is composed by four SH, commonly named W, X, Y and Z, which correspond respectively to the omnidirectional pressure and the three Cartesian components of the particle velocity vector.

Within the following forty years, the limit of FOA has been surpassed by the Higher Order Ambisonics (HOA) arrays: having a larger number of capsules, it is possible to obtain an Ambisonics expansion of higher order, reaching in this way a better spatial accuracy. On the other side, a more advanced electronic is required to place a lot of capsules over a small surface and a much higher computation power is needed to perform the encoding, being the number of input and output very large. Note that the encoding operation for the Soundfield™ microphone was done by an analog circuit.

In Figure 3, the 3D plots of SH directivity patterns are shown up to the order 4. The numeration, starting from zero and increasing from left to right, follows the Ambisonics Channel Number (ACN) format, adopted by the current Ambisonics standard AmbiX [9].

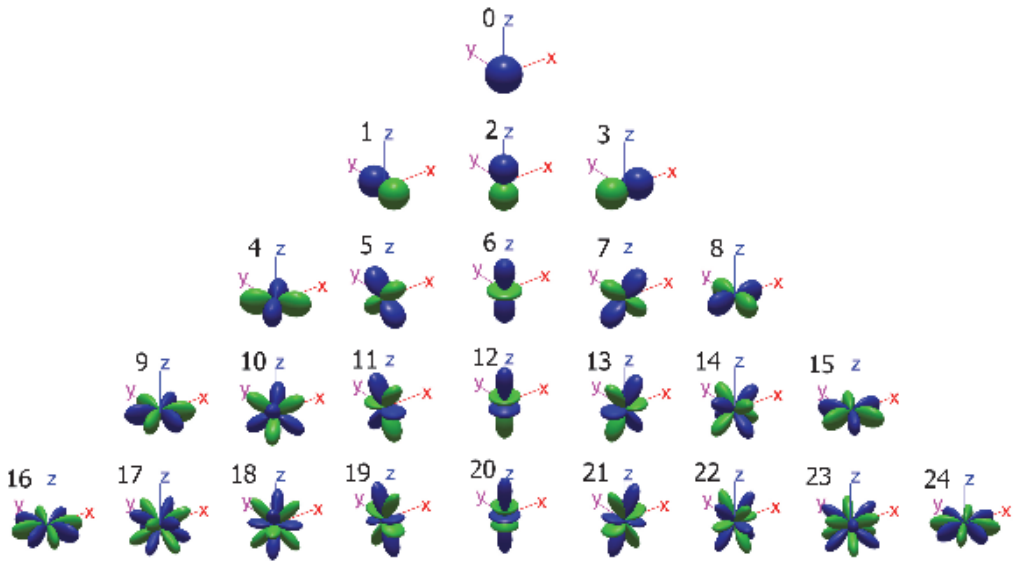


Figure 3: Spherical Harmonics directivity patterns up to order 4

2.1.2. Spatial PCM Sampling Theory

The SPS format [10] was developed as the spatial equivalent to Pulse Code Modulation (PCM) sampling of a time-domain waveform, as Ambisonics is the spatial equivalent of its Fourier representation. The encoding operation in this case is referred to A-2-P format conversion.

Instead of using SH, SPS employs many unidirectional microphones, covering uniformly the surface of a sphere, aiming to capture the complete spatial information. The polar patterns of these virtual microphones are high order Cardioids, Super Cardioids or Hyper Cardioids without any side or rear lobes, defined by the following formula:

$$Q(\vartheta) = [P + \mathrm{G} \cdot \cos(\vartheta)]^O, \quad (2)$$

where P and G are respectively the percentages of pressure and velocity that define the type of virtual microphone (Table 1), ϑ is the angle between the direction of aiming of the virtual microphone and the DoA of the sound wave and O is the order. Figure 4 shows the 2D polar plots of a virtual cardioid ($P = G = 0.5$) of various order from 0 to 10 (left) and a 3D polar plot of a virtual microphone of type super-cardioid ($P = 1/3, G = 2/3$) and order 16 (right).

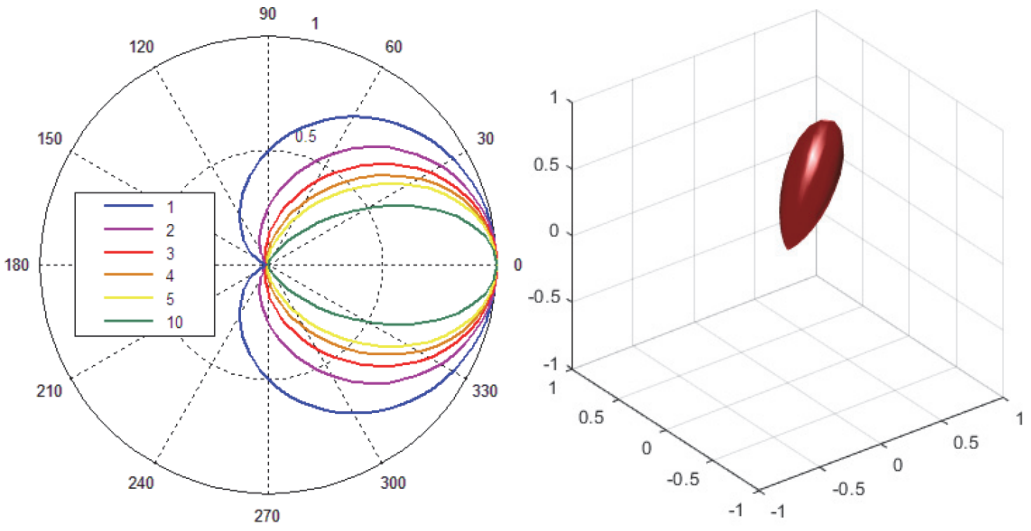


Figure 4: 2D polar plot of virtual cardioids of various order (left) and 3D plot of a virtual super-cardioid of order 16 (right)

2.2. The spatial filtering matrix

The encoding operation consists in a matrix filtering described by:

$$y_v(t) = \sum_{m=1}^{M} x_m(t) * h_{m,v}(t), \quad (3)$$

for a system of m input and v output, where $h_{m,v}$ is the filtering matrix. The scheme of the filtering processor can be thus described as in Figure 5, while Figure 6 shows the scheme for a single channel.

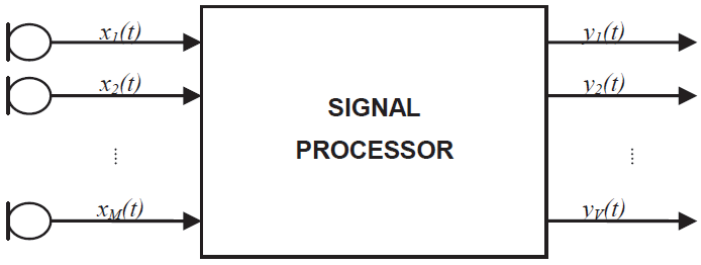


Figure 5: Filtering processor scheme

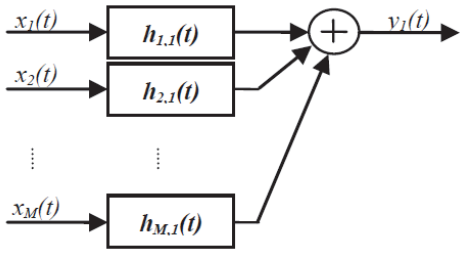


Figure 6: Filtering scheme for a single channel output

Besides the implementation of the convolution engine, the most important aspect of the encoding process is the filtering matrix itself. Two ingredients are required to calculate it: the array response and an inversion method and in turn, there are different ways to get the array response and several inversion methods.

The array response can be calculated numerically solving a theoretical model (paragraph 2.2.3, page 26), measured (paragraph 2.2.4, page 27) or simulated (paragraph 2.2.5, page 29). Some of the inversion methods employed to get the filtering matrix from an array response are briefly described in the following paragraph and their performances will be compared in 2.2.2.3.

2.2.1. Methods for the array response inversion

Three inversion methods of an array response are described. Two of them have been found in literature and can produce filters only for Ambisonics encoding, the third one has been developed at the University of Parma and it is suitable for both Ambisonics and SPS encoding. The array response can be either numerically calculated, measured or simulated.

1. *Least-squares solution in space domain with Tikhonov regularization*

Cited from [11]: "Filters are generated as a least-squares solution with a constraint on noise amplification, using Tikhonov regularization. The method formulates the least-squares problem in the space domain, using the directional measurements of the array response [12]".

2. *Least-squares solution in SH domain with Tikhonov regularization*

Cited from [11]: "Filters are generated as a least-squares solution with a constraint on noise amplification, using Tikhonov regularization. The method formulates the least-squares problem in the spherical harmonic domain, by expressing the array response to an order-limited series of SH coefficients [13]".

3. *Kirkeby inversion with frequency-dependent regularization*

This is the method developed at University of Parma, based on the adoption of a frequency-dependent regularization parameter [14] and the Kirkeby inversion [15], whose equation is:

$$\|H[k]\|_{M\times V} = inv(\|C[k]\|^*_{M\times D} \cdot \|C[k]\|_{D\times M} + \beta[k] \cdot \|I\|_{M\times M}) \cdot \left(\|C[k]\|^*_{M\times D} \cdot \|A\|_{D\times V} \cdot e^{-j\pi(k-1)}\right), \quad (4)$$

where k is the frequency index, $\|C[k]\|$ is the array response, $\|A\|$ is the directivity matrix imposed as target function, $\beta[k]$ is the frequency-dependent regularization parameter and $\|H[k]\|$ is the resulting filtering matrix for an array of M input (the capsules of the array) and V output (ultra-directive virtual microphones or SH).

Depending on the definition of the directivity target matrix $\|A\|$, SPS od Ambisonics filtering can be achieved: in the first case, coefficients are described by equations reported in [8], otherwise by equation (2).

2.2.1.1. *Improvement of the frequency-dependent parameter β*

The frequency-dependent parameter $\beta[k]$ (Figure 7) is defined by the following constraints:

- β_{ol}: low out-band value;
- β_{oh}: high out-band value;
- β_i: in-band value;
- f_{l1}: starting frequency of low transition band;
- f_{l2}: ending frequency of low transition band;
- f_{h1}: starting frequency of high transition band;
- f_{h2}: ending frequency of high transition band.

The four values of the transition frequencies can be calculated also with the central frequencies of the two transition bands and a value that defines their width in octave, as follow:

- f_{cl}: central frequency of low transition band;
- f_{ch}: central frequency of high transition band;
- tbw_{oct}: octave transition band width.

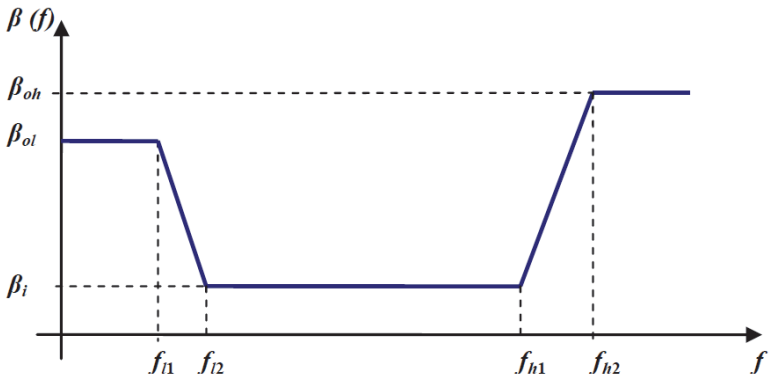


Figure 7: Frequency-dependent regularization parameter $\beta(k)$

A further tweaking of this parameter has been introduced: instead of keeping the in-band value constant, it has been optimized in function of the maximum noise amplification, or *White Noise Gain (WNG)*, referenced in [12] and calculated as:

$$WNG[k] = \max(\mathrm{eig}(\|H[k]\|^* \cdot \|H[k]\|). \quad (5)$$

Before calculating the filtering matrix $\|H[k]\|$, a threshold, e.g. WNG_{max} $[dB]$, has to be set to limit the maximum noise gain amplification. The filtering matrix is

calculated a first time, employing the β parameter with constant value at all frequencies and equal to:

$$\beta = 1/(2 \cdot \alpha), \tag{6}$$

with:

$$\alpha = 10^{(WNG_{max}/20)}. \tag{7}$$

In this way, $WNG(k)$ is calculated as a function of the frequency; in Figure 8 it is shown the result obtained by inverting a simulated response of the EM, with $WNG_{max} = 20\ dB$. Note that the matrix $\|C[k]\|$ was simulated in the frequency range $11.71875\ Hz - 1417.96875\ Hz$ and the filter $\|H[k]\|$ was calculated in the range $f_l = 20\ Hz - f_h = 1.4\ kHz$.

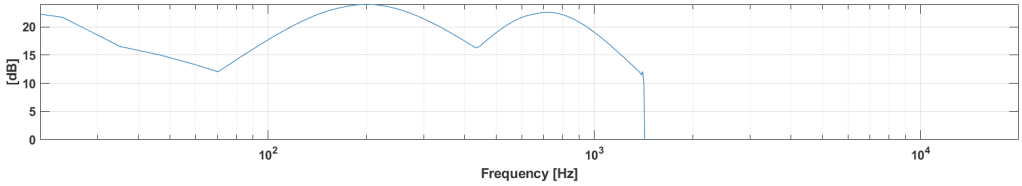


Figure 8: $WNG(k)$, Kirkeby inversion with non-optimized $\beta(k)$

As it is possible to see, even with the imposed threshold $WNG_{max} = 20\ dB$, values are in the range $12\ dB - 24\ dB$, with large variations.

Values of $WNG(k)$ are stored and the filtering matrix is calculated a second time, employing values previously stored of $WNG(k)$. In this way, a new parameter $\beta[k]$ is calculated, as a function of the frequency:

$$\beta[k] = \beta \cdot corrIndex[k], \tag{8}$$

with:

$$corrIndex[k] = 10^{((WNG(k) - WNG_{max})/corrFactor)}, \tag{9}$$

where β is defined by equations (4) and (5) and the $corrFactor$ is a value that define the amount of correction and requires to be set manually.

The low and high out-band values are instead as follow:

$$\beta_{(1:f_l)} = \beta[f_l] \tag{10}$$

$$\beta_{(f_h:end)} = 1 \tag{11}$$

Figure 9 shows an example of the new optimized parameter $\beta[k]$, calculated with $WNG_{max} = 20\ dB$ and $corrFactor = 10$.

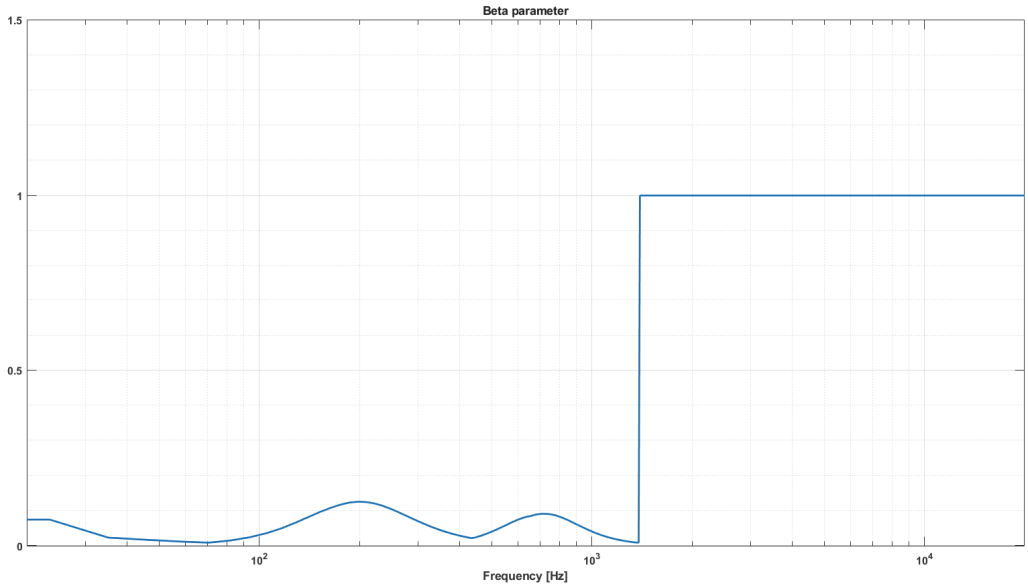


Figure 9: Optimized frequency-dependent regularization parameter $\beta(k)$

Figure 10, Figure 11 and Figure 12 show the results of the correction, with different values of the correction factor.

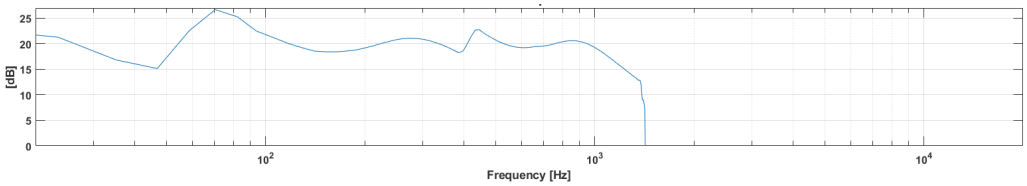


Figure 10: $WNG(k)$, Kirkeby inversion with optimized $\beta(k)$, $corrFactor = 10$

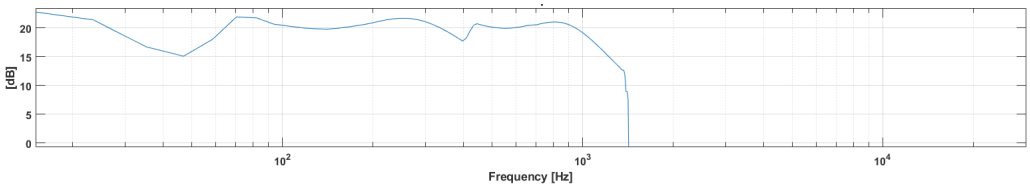


Figure 11: $WNG(k)$, Kirkeby inversion with optimized $\beta(k)$, $corrFactor = 15$

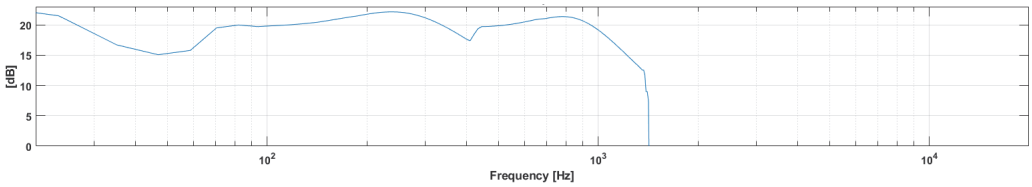


Figure 12: $WNG(k)$, Kirkeby inversion with optimized $\beta(k)$, $corrFactor = 20$

In Figure 10 ($corrFactor = 10$) values are in the range $15.1\ dB - 26.7\ dB$ and variations are still evident, in Figure 11 ($corrFactor = 15$), values are in the range $15.1\ dB - 21.9\ dB$ with reduced variations and in Figure 12 ($corrFactor = 20$), values are in the range $15.1\ dB - 22.1\ dB$ with minimum variations.

2.2.2. Theoretical analysis of spatial performances

The method developed at University of Parma to calculate the filtering matrix by inverting anechoic measurements, which is indeed very effective, presented an unsatisfactory methodology for evaluating the results. The assessment that filters were calculated correctly was done by looking at the 3D plot of the virtual microphone directivity patterns at various octave bands. This procedure allows evaluating correctly the result but it is time consuming, especially in case of SPS format, for which the number of virtual microphones can be very large. Comparative tests were all but not trivial due to the large number of figures to analyse. Finally, the discretization of the frequency limits for various Ambisonics orders was quite poor, as only octave bands were evaluated. Therefore, a solution for fast evaluating beamforming performance has been coded, following a metrics introduced in [12].

2.2.2.1. Ambisonics spatial performances evaluation

An implementation of the evaluation metrics of spatial performances for Ambisonics format has been found in [11]: filters are compared to the ideal SH components, by means of two parameters as a function of the frequency:

- *Spatial Correlation (SC)* between each ideal and reconstructed SH;
- *Level Difference (LD)* of the mean spatial power between ideal and reconstructed SH.

Each Ambisonics order is perfectly reconstructed if:

$$SC_{ideal} = 1 \quad (12)$$

$$LD_{ideal} = -0.5\ dB \quad (13)$$

The two curves define the frequency limits for each order: usually LD for the starting frequency and SC for the ending frequency. A threshold for each parameter is required to identify precisely these frequencies; in this text, it will be always referred to the following values, empirically chosen:

$$SC_{threshold} = 0.95 \quad (14)$$

$$LD_{threshold} = -0.5\ dB \quad (15)$$

This metrics can be applied to any type of array response and inversion method, making it possible to compare quickly different geometries of the array and the various technique to calculate the filtering matrix.

As an example, a response of the EM (Figure 2, middle) measured inside an anechoic room has been inverted by means of method 1 (paragraph 2.2.1) and $WNG_{max} = 20\ dB$. A filtering matrix for encoding Ambisonics 4th order has been generated, with filters of length $4096\ samples$. The spatial performance evaluation is shown in Figure 13. Considering the thresholds defined in (14) and (15), frequency limits for each order can be defined as in Table 2.

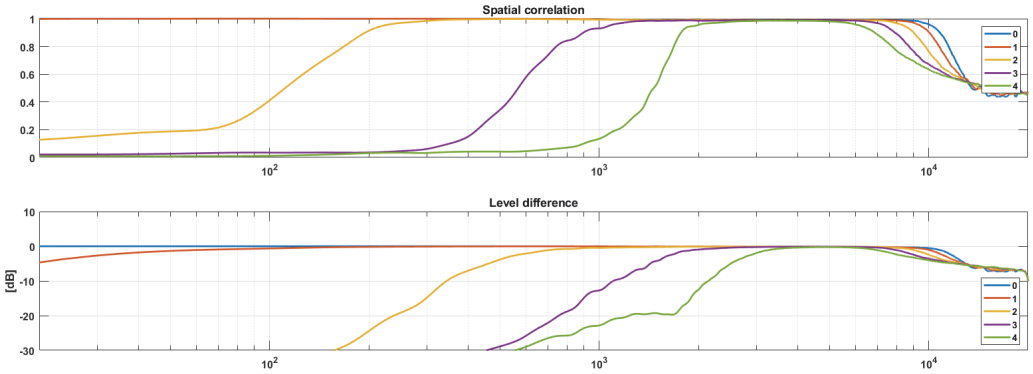


Figure 13: Spatial Correlation and Level Difference, measured array response

	EM, measured response	
Ambisonics order	**Freq. start [Hz]**	**Freq. stop [kHz]**
1	120	9.4
2	900	8.4
3	2400	7.3
4	3700	6.2

Table 2: EM, frequency limits for Ambisonics orders 1-2-3-4, measured response

2.2.2.2. SPS spatial performance evaluation

The inversion method developed at University of Parma can be employed to generate the P-format encoding matrix, defining a proper target matrix for the directivity of the virtual microphones.

It has been developed a method to evaluate the real directivity obtained for each virtual microphone in function of the frequency: in fact, there is not a one-to-one relation between the target and the result. Ideal virtual microphones having directivity closer to the ones obtained are calculated for each directions at all frequencies.

As an example, the same response of the EM measured in anechoic room and employed in the previous paragraph has been used to calculate the SPS filtering matrix by means of the Kirkeby inversion, with $WNG_{max} = 20\ dB$ and without applying any optimization of the parameter $\beta[k]$ in function of $WNG[k]$. The chosen target for the directivity matrix $\|A\|_{D\times V}$ of equation (4) is a set of 32 super-cardioid microphones of order 16, having the ideal directivity showed by the 3D plot of Figure 4 (page 7) and aiming in the direction of the capsules. Figure 14 shows the 3D plot of the directivity obtained at various octave bands, while the histogram plot of Figure 15 gives the information of the real directivity obtained in function of the frequency, in terms of type and order of virtual microphone. Above a certain threshold, as it happens also for the Ambisonics encoding, the array no longer provides beamforming capabilities: note

in fact the sharp change of the directivity pattern between the octaves at 8 kHz and 16 kHz (Figure 14) and the loss of directivity above 13 kHz (Figure 15).

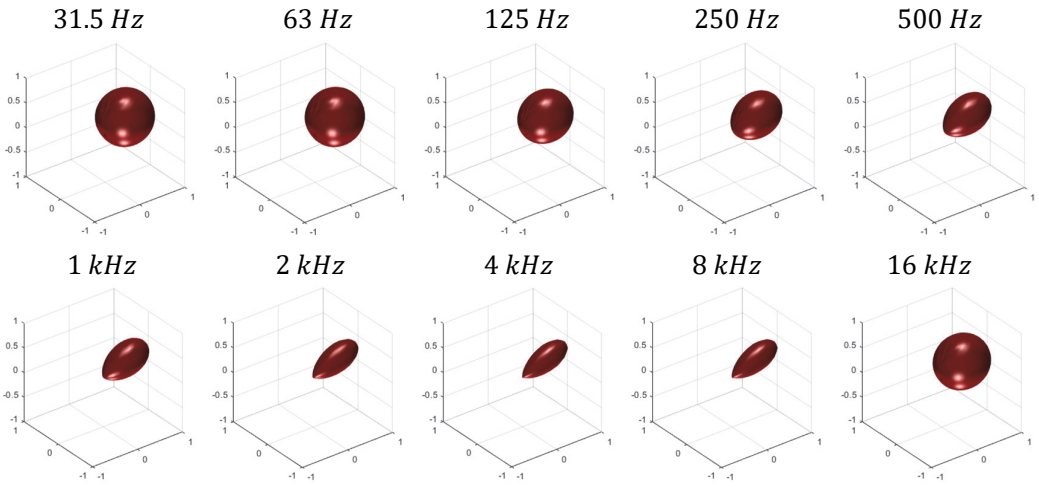


Figure 14: SPS directivity at various octave bands, given a super-cardioid of order 16 as target

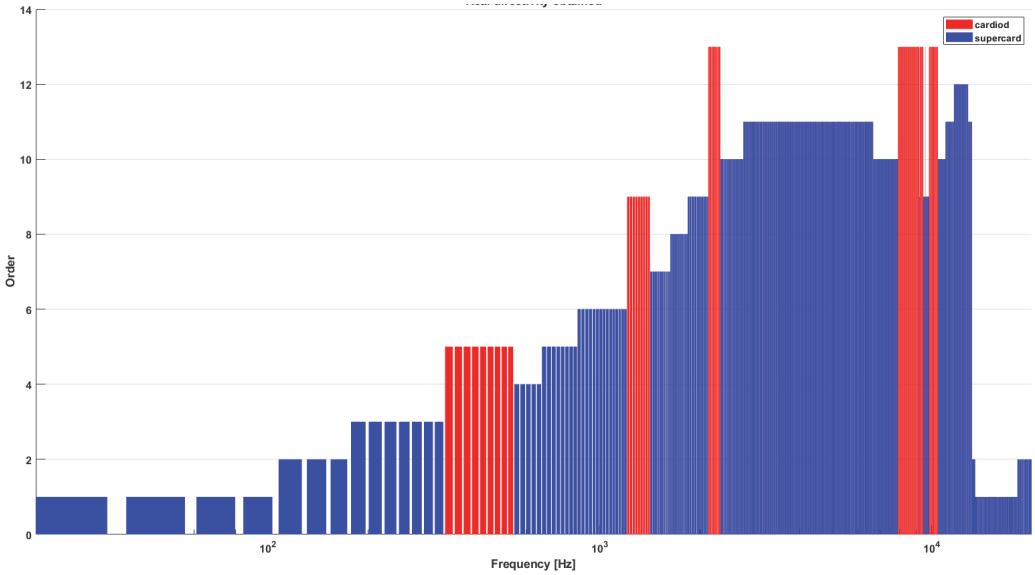


Figure 15: SPS directivity as a function of the frequency

In Figure 15, one can note that the target is never reached: the highest directivity values are close to cardioid of order 13 and super-cardioid of order 12, which are indeed quite similar; in the lower part of the spectrum the beamforming is not beyond super-cardioids of order one, two and three.

Following a metrics similar to the one proposed in [12] and discussed in the previous paragraph, the spatial correlation between the directivity of the target and the directivity of the virtual microphones encoded has been studied; again, optimal value should be one. The result is shown in Figure 16. Note that the value decreases rapidly

going towards lower frequencies, as the directivity is very different with respect to the target.

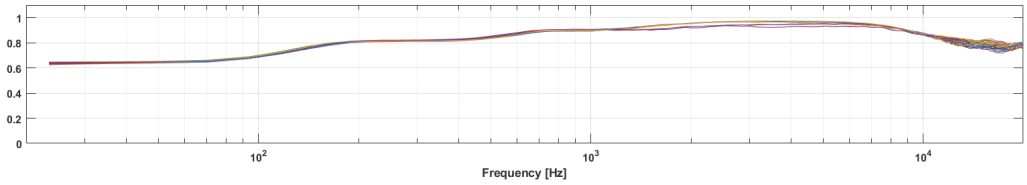


Figure 16: Spatial Correlation between target and encoded virtual microphones

The spatial correlation can be evaluated also between the directivity of the virtual microphones encoded and the directivity of the ideal virtual microphones most similar to the firsts, which in turn are the ones showed in Figure 15. The result is presented in Figure 17 and, as expected, the correlation is in this case very high at all frequencies.

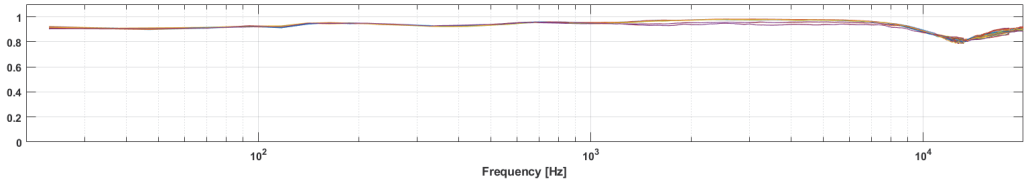


Figure 17: Spatial Correlation between encoded and ideal virtual microphones

The last aspect considered is the uniformity of spatial sampling: ideally, the sum of all virtual microphones should give back the unit sphere at all frequencies. Figure 18 shows at various octave bands the sum of all the 32 virtual microphones of this SPS set.

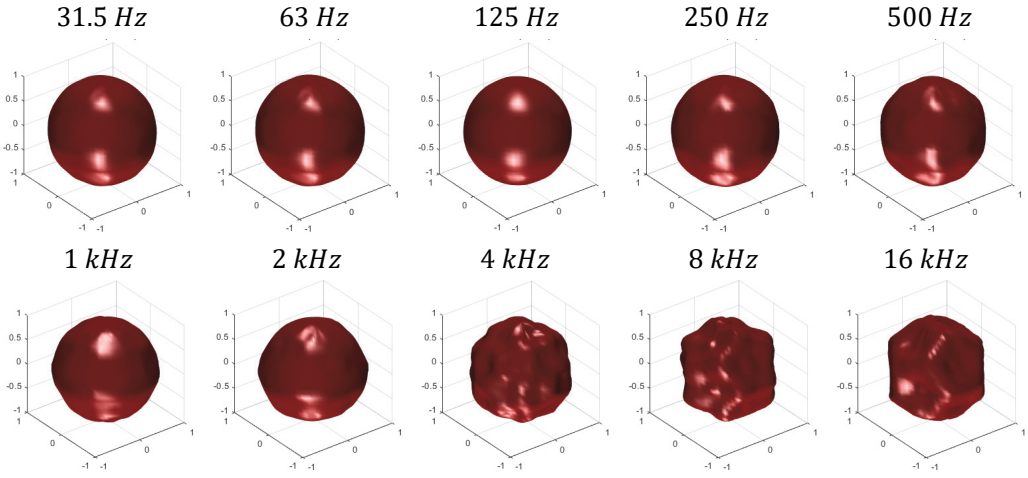


Figure 18: Sum of all virtual microphones at various octave bands

Note that the spatial sampling is quite uniform at least up to 2 kHz and only in the octaves at 4 kHz and 8 kHz there are some distortions. However, it is noted that the energy is not constant at different octaves, therefore an equalization is required. Figure 19 shows a correction curve calculated for the example above, plotted in the range 20 $Hz - 13$ kHz and with the 1 kHz frequency, which is the calibration frequency for

the microphones, imposed at 0 dB. A specific correction curve should be calculated for each case.

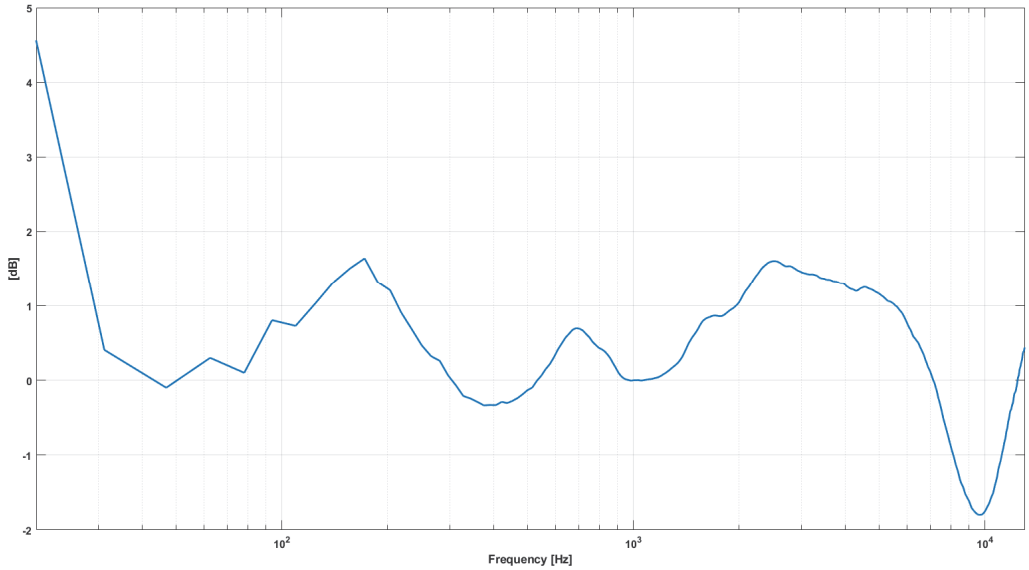


Figure 19: Energy correction curve for virtual microphones of SPS format

The last aspect studied is the spatial distribution of the virtual microphones, which is one of the two parameters required to define the target matrix $\|A\|$ of the Kirkeby inversion (the other one is the directivity of the virtual microphones). In the following, some consideration will be discussed about the geometry of the SPS format.

A first approach consists in defining the direction of the virtual microphones coincident with the capsules of the array. Thus, a one-to-one correspondence is guaranteed between the recorded and encoded signals, but the number of virtual microphones is forced to be equal to the number of capsules.

To test the effectiveness of this methodology, the response of the EM has been inverted twice with Kirkeby method, $WNG_{max} = 20\,dB$, length of the filters $4096\,samples$, without optimizing the regularization parameter β. Two different target matrices $\|A\|$ have been defined for the inversions, in both cases producing 32 virtual microphones of type super cardioid of order 16, but with a different set of directions. In the first case, virtual microphones aim in the same direction of the capsules of the array while in the second case, directions correspond to a nearly-uniform grid of 32 points (see paragraph 2.2.5.3, page 34).

The spatial performance of the filters (Figure 20 and Figure 21) have been studied following the metrics previously explained. The sum of all the virtual microphones at various octave bands is shown (Figure 22 and Figure 23) and finally the directivity obtained in function of the frequency for the two cases is presented (Figure 24). In Figure 21, it is possible to note a curve clearly different from the others: it is the curve of the virtual microphone aiming the South Pole, a direction that is present only in the nearly-uniform grid of 32 points. In that direction, the body of the array acts as a shield, limiting the beamforming.

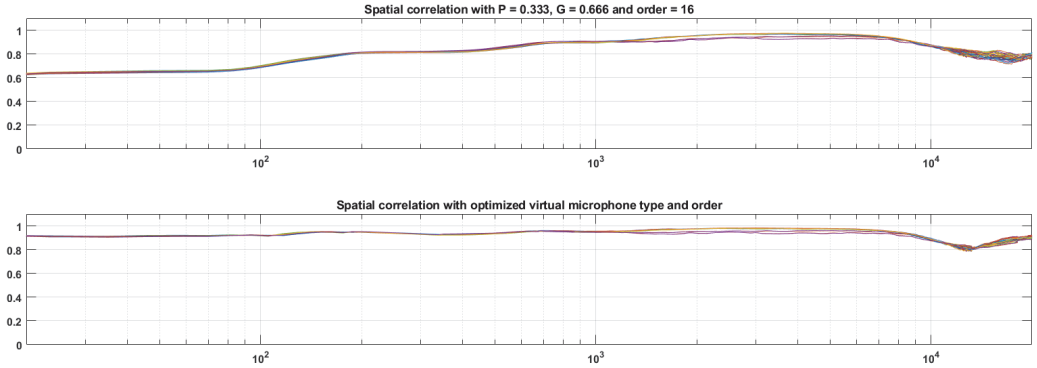


Figure 20: EM, spatial performance of SPS, virtual microphone directions equal to capsules

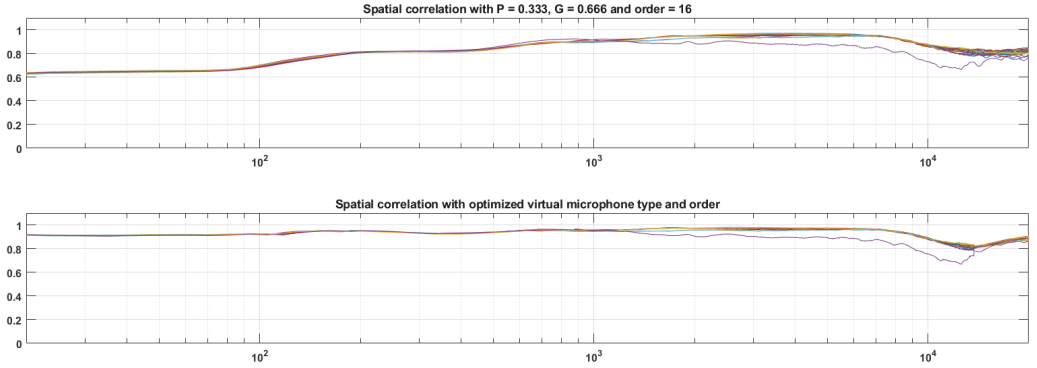


Figure 21: EM, spatial performance of SPS, virtual microphone directions equal to nearly-uniform grid

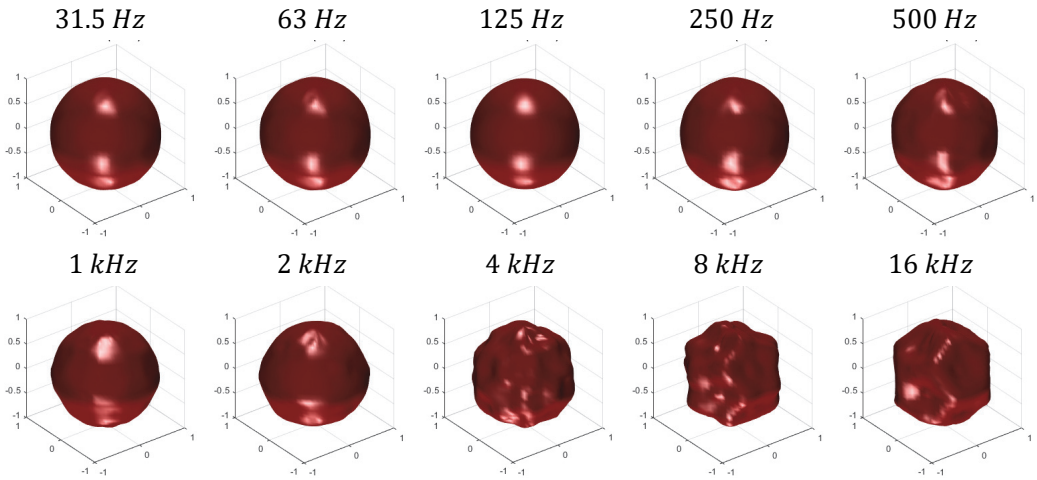


Figure 22: EM, sum of all virtual microphones of the SPS format, directions equal to capsules

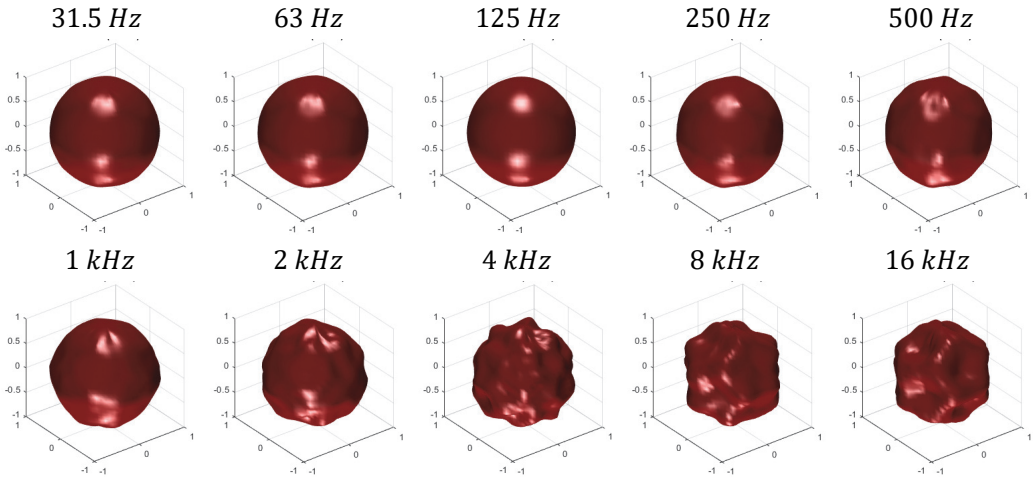


Figure 23: EM, sum of all virtual microphones of the SPS format, directions equal to nearly-uniform grid

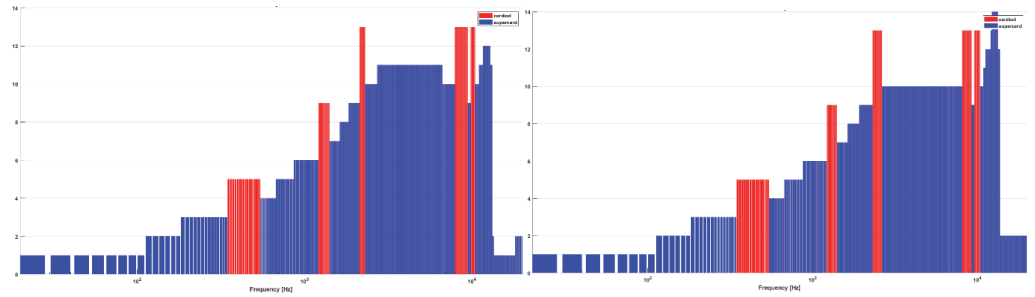


Figure 24: EM, directivity of the virtual microphones of the SPS format, directions equal to capsules (left) and directions equal to nearly-uniform grid (right)

In Figure 24, one can note that the filtering matrix obtained with the virtual microphones aiming in the same direction of the capsules behaves better at high frequency: in the range $2.3\ kHz - 8.2\ kHz$ a higher directivity is reached. The uniform grid instead ensure the uniformity of the spatial sampling: i.e. the sum of the virtual microphone in the $2\ kHz$ octave band (Figure 22) appears flattened at the bottom. In addition, it must be said that the EM is a spherical array with capsules covering the surface very uniformly, otherwise there could be much more difference, as shown in the following.

An analogue analysis has been done for the HSA (see paragraph 2.3), a non-spherical array with non-uniform distribution of the capsules over the surface. The simulated response (see paragraph 2.2.5) has been inverted in the frequency range $20\ Hz - 3.5\ kHz$, with similar parameters employed in the previous example. Virtual microphones of type super cardioid of order eight have been set for the target matrix. Spatial performances are shown in Figure 25 and Figure 26 for the two cases: virtual microphone directions equal to capsule directions and virtual microphone directions equal to the nearly-uniform grid of 32 points. Figure 27 and Figure 28 show the sum of all the virtual microphones at various octave bands (up to $2\ kHz$), whilst in Figure 29 the directivity obtained in function of the frequency for the two cases is presented.

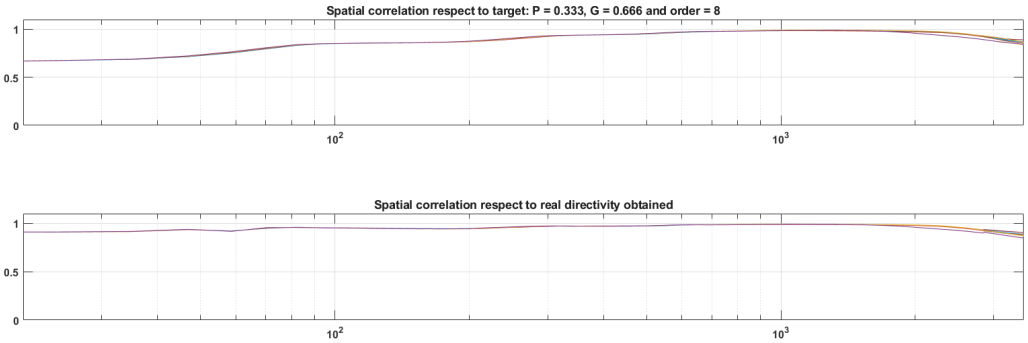


Figure 25: HSA, spatial performance of SPS format, virtual microphone directions equal to capsules

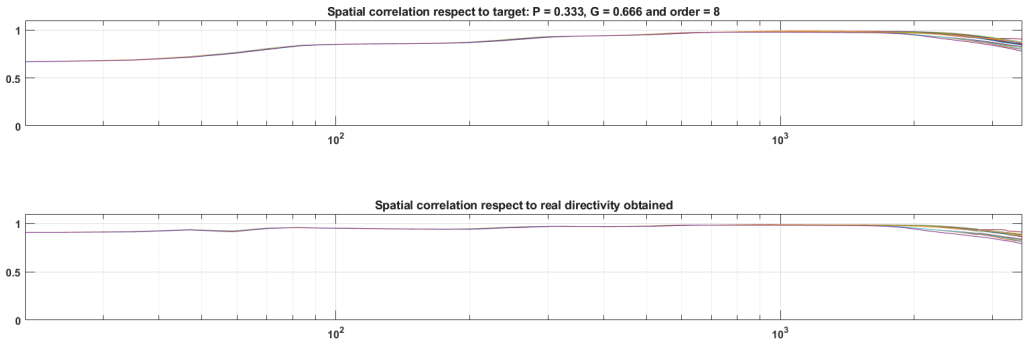


Figure 26: HSA, spatial performance of SPS format, virtual microphone directions equal to nearly-uniform grid

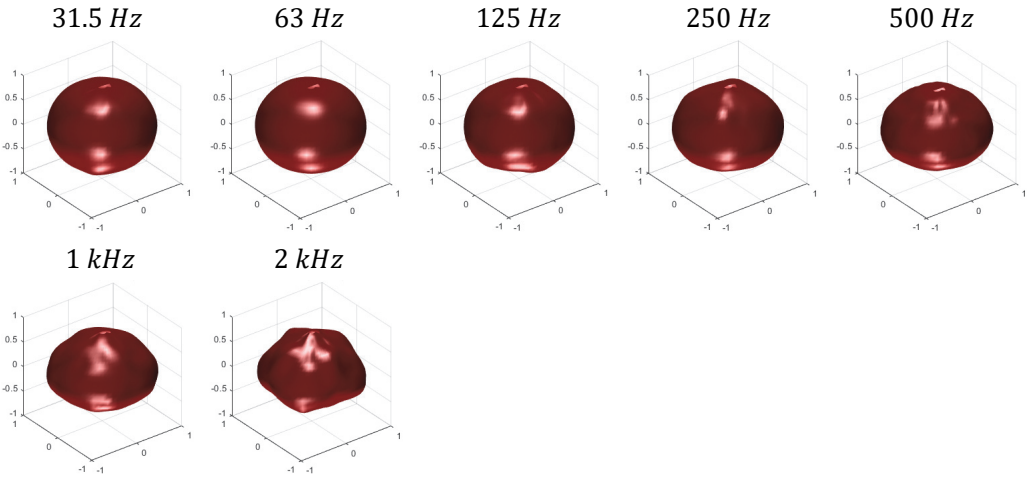


Figure 27: HSA, sum of all virtual microphones of the SPS format, directions equal to capsules

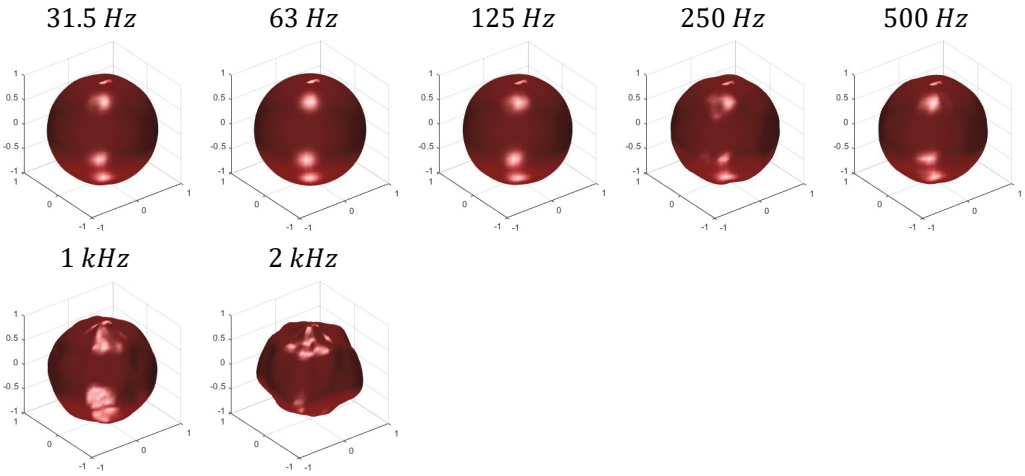


Figure 28: HSA, sum of all virtual microphones of the SPS format, directions equal to nearly-uniform grid

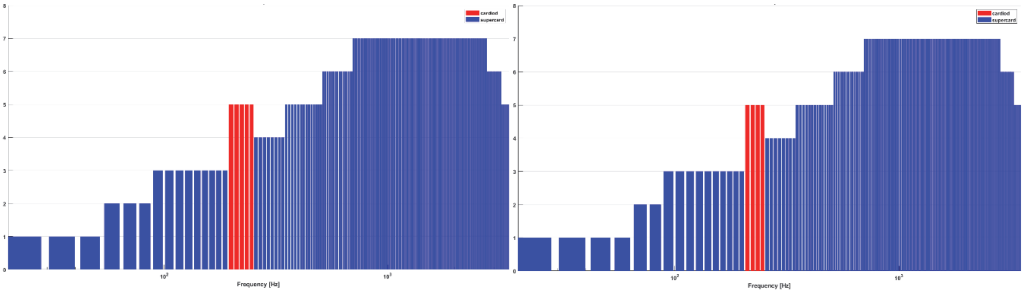


Figure 29: HSA, directivity of the virtual microphones of the SPS format, directions equal to capsules (left) and directions equal to nearly-uniform grid (right)

The difference in terms of maximum directivity is almost negligible: the target defined with virtual microphone directions coincident to the capsule directions produced a minimum improvement at low frequency (Figure 29); by comparing Figure 25 and Figure 26 we note in the second one a slight tendency of the curves to open at high frequency. A part from that, the real difference consist in the uniformity of the spatial sampling: the distribution of the capsules is asymmetric over the surface, and the surface itself is not spherical. As consequence, when virtual microphone directions are coincident with the capsules the sum of the virtual microphone is considerably distorted (Figure 27). Therefore, in this case, there is not a preferable solution, but it depends on how the array is mounted and used: if it is mounted on a mannequin torso (i.e. Figure 78, right), the first solution can be preferable, if instead it is mounted on a microphone stand (i.e. Figure 78, left), the nearly-uniform target could be better.

Therefore, we can conclude by saying that best performance in terms of directivity are provided by virtual microphones having directions coincident with directions of the capsules of the array. Nevertheless, this condition do not ensure a uniform sampling, unless the array is spherical and the capsules cover its surface uniformly. If the array is not spherical nor the distribution of the capsules over the surface is uniform, the previous target can be employed or not, depending on how the array is mounted and the purpose of the recording.

2.2.2.3. Comparison of the inversion methods

The three inversion methods described in paragraph 2.2.1 have been compared by inverting the same array response with identical parameters for all of them and then looking at the results in terms of spatial correlation, level difference and noise gain amplification of the filters.

The response employed for this comparison is the anechoic measurement of the EM already used previously. A filtering matrix for encoding Ambisonics 4th order has been calculated, with length of the filters 4096 $samples$ and $WNG_{max} = 20\ dB$.

In Figure 30, it is showed the frequency-dependent β parameter calculated with the following parameters:

- $\beta_{ol} = 1$;
- $\beta_{oh} = 1$;
- $\beta_i = 0.05$;
- $f_{cl} = 20\ Hz$;
- $f_{ch} = 13500\ Hz$;
- $tbw_{oct} = 0.3$

The value $\beta_i = 0.05$, assumed in the range $20\ Hz - 12\ kHz$, is calculated with equations (6) and (7), in function of $WNG_{max} = 20\ dB$.

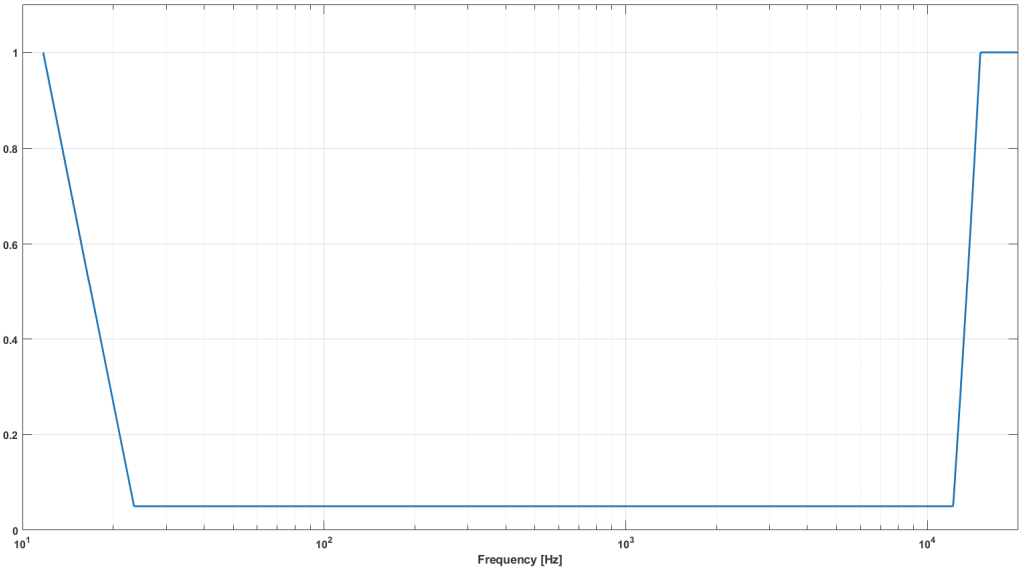


Figure 30: Frequency-dependent regularization parameter $\beta(k)$

Results are showed in Figure 31, Figure 32 and Figure 33 respectively for the method 1, 2 and 3.

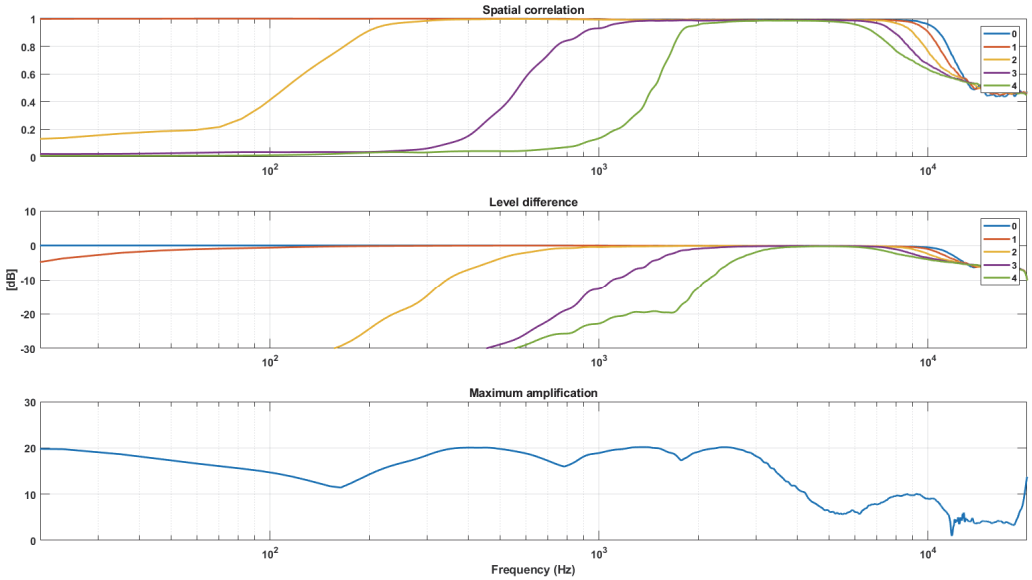


Figure 31: EM, spatial performances with least-squares solution in space domain and Tikhonov regularization

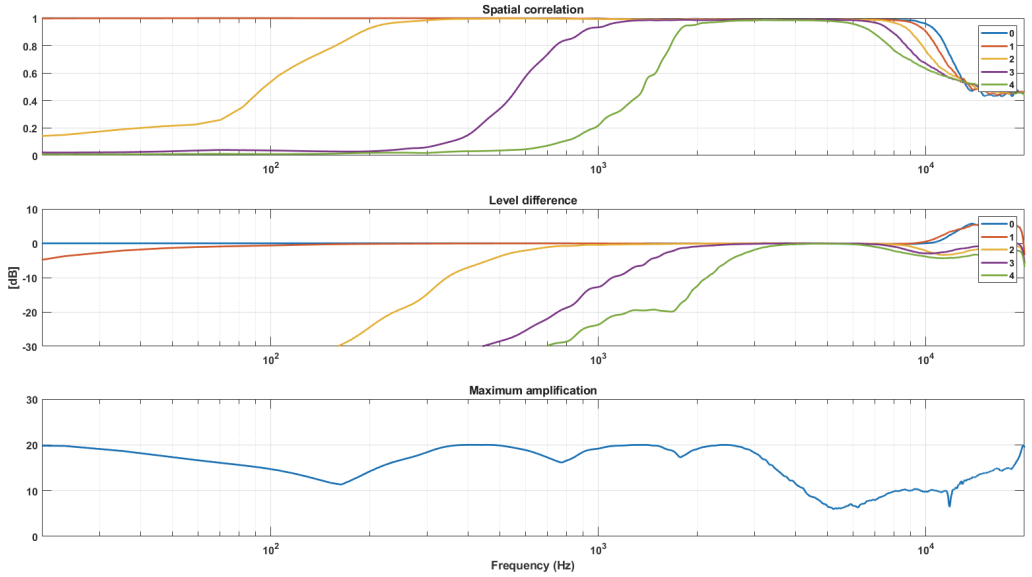


Figure 32: EM, spatial performances with least-squares solution in SH domain and Tikhonov regularization

It is possible to note that results provided by the methods 1 and 2 are very similar in terms of SC and LD, but the method 2 presents a higher level of noise amplification at high frequency, above 10 kHz. A fact reflected also by the curves of LD, which tend to increase instead of decrease beyond 10 kHz.

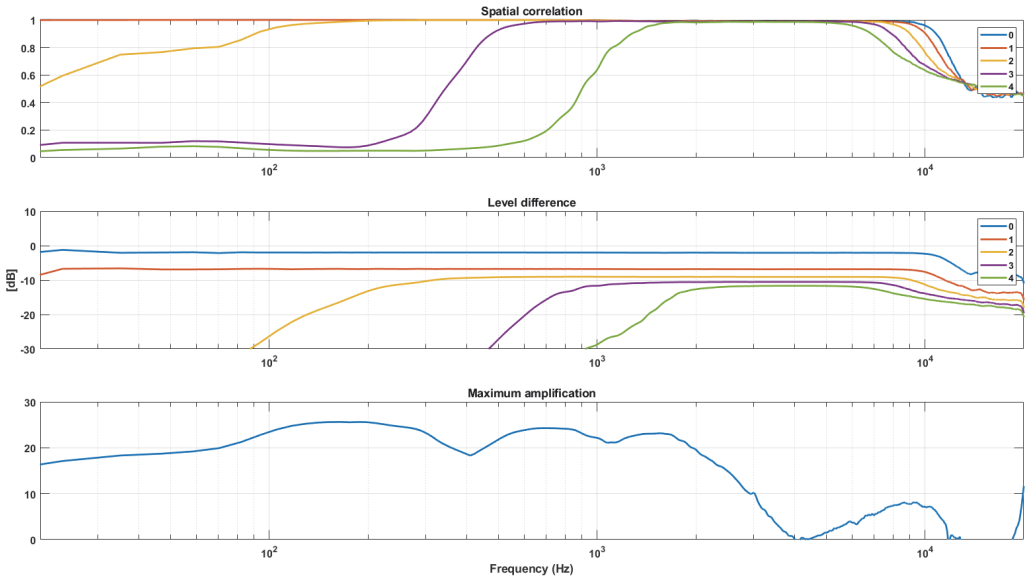


Figure 33: EM, spatial performances with Kirkeby inversion and $\beta_i = 0.05$

The spatial performances of the filter calculated with method 3 (Kirkeby) are instead very different respect to the other two. The lower frequency limit for each order is considerably extended downward, whilst no differences are found at high frequency. The other aspect is that the curves of LD are not matched in the region of the spectrum where they are flat. This method provides also the lowest noise amplification at very low and medium-high frequencies, below $50\ Hz$ and above $2.5\ kHz$ in this case. It must be said, indeed, that in the central part of the spectrum the noise amplification is higher respect the other methods and the given constraints is not respected, with three peaks of $25.6\ dB$, $24.3\ dB$ and $23.1\ dB$. In order not to have any peak above $20\ dB$ and make a fair comparison, the inversion has been performed a second time, setting manually the value $\beta_i = 0.18$, which correspond to $WNG_{max} = 8.9\ dB$. The results is showed in Figure 34. Note that, even if now the noise amplification level do not exceed the given constraint, spatial performance at low frequencies are still improved.

The frequency limits for each order for the three inversion methods are summarized in Table 3, calculated with thresholds defined in 2.2.2.1. Note that in case of method 3, as the curves of LD are not matched, the threshold has been calculated by subtracting $0.5\ dB$ from the value assumed in the flat region.

In Figure 35, the PSD of the three filters are shown in red, blue and black for methods 1, 2 and 3, respectively; to keep the figure legible, only the first 16 signals, corresponding to Ambisonics 3rd order, have been analysed.

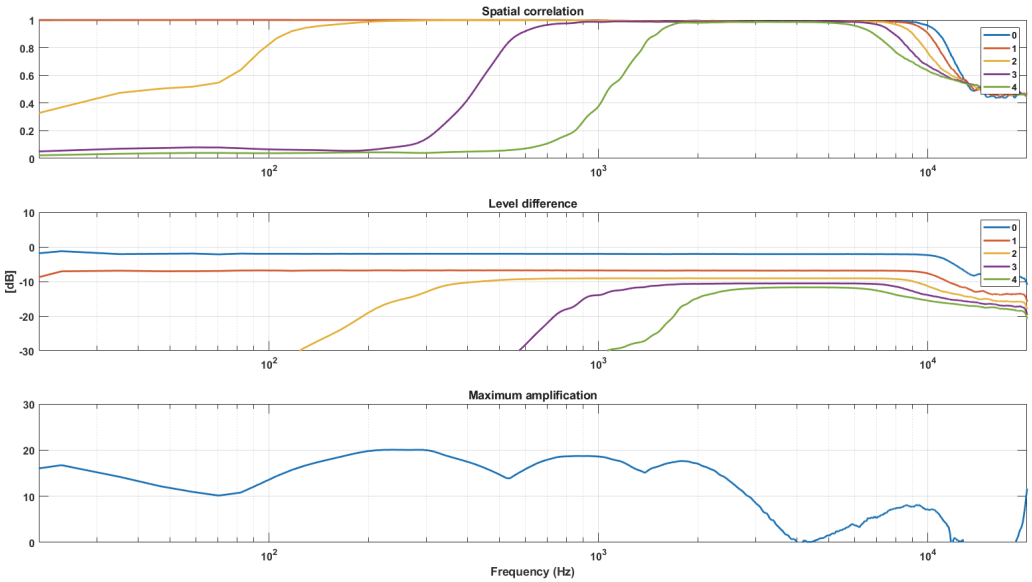


Figure 34: EM, spatial performances with Kirkeby inversion and $\beta_i = 0.18$

	Inversion method 1		Inversion method 2		Inversion method 3	
Ambisonics order	**Fr. start [Hz]**	**Fr. stop [kHz]**	**Fr. start [Hz]**	**Fr. stop [kHz]**	**Fr. start [Hz]**	**Fr. stop [kHz]**
1	120	9.4	120	9.4	25	9.4
2	900	8.4	900	8.4	370	8.4
3	2400	7.3	2400	7.3	1240	7.3
4	3700	6.1	3700	6.1	2300	6.1

Table 3: EM, frequency limits for Ambisonics orders 1-2-3-4, various inversion methods

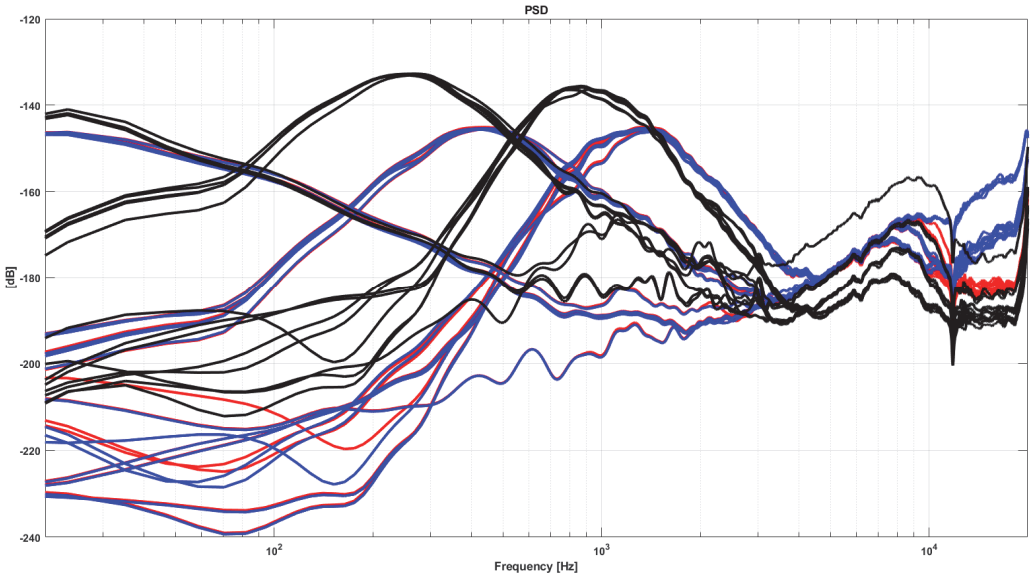


Figure 35: EM, PSD of filters, inversion method 1(red), 2 (blue) and 3 (black)

Finally, the array response has been inverted with Kirkeby method and the β parameter optimized in function of the WNG threshold. The optimized parameter is shown in Figure 36 and spatial performance are presented in Figure 37.

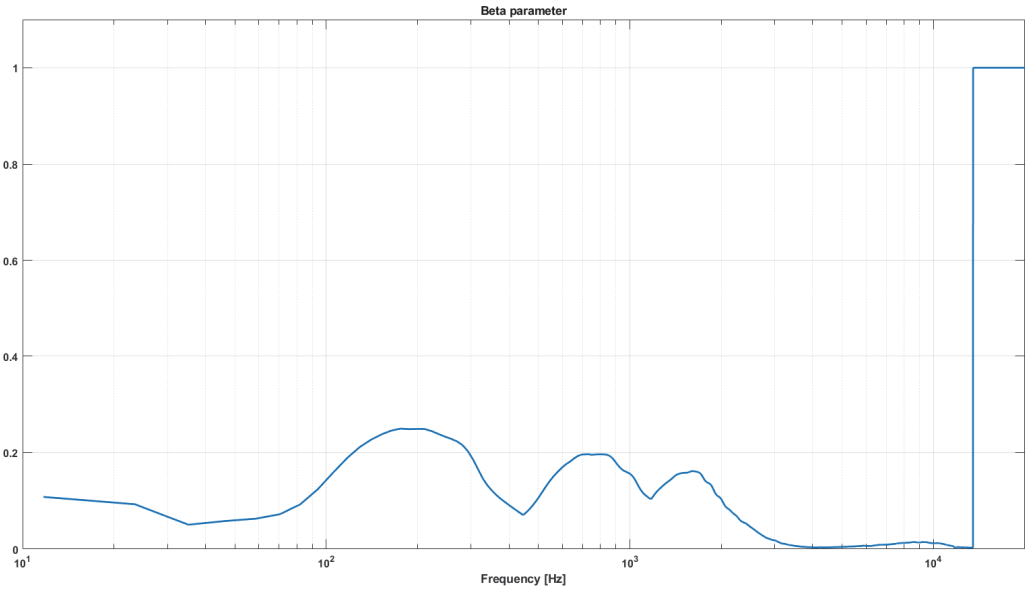


Figure 36: EM, optimized regularization parameter $\beta(k)$

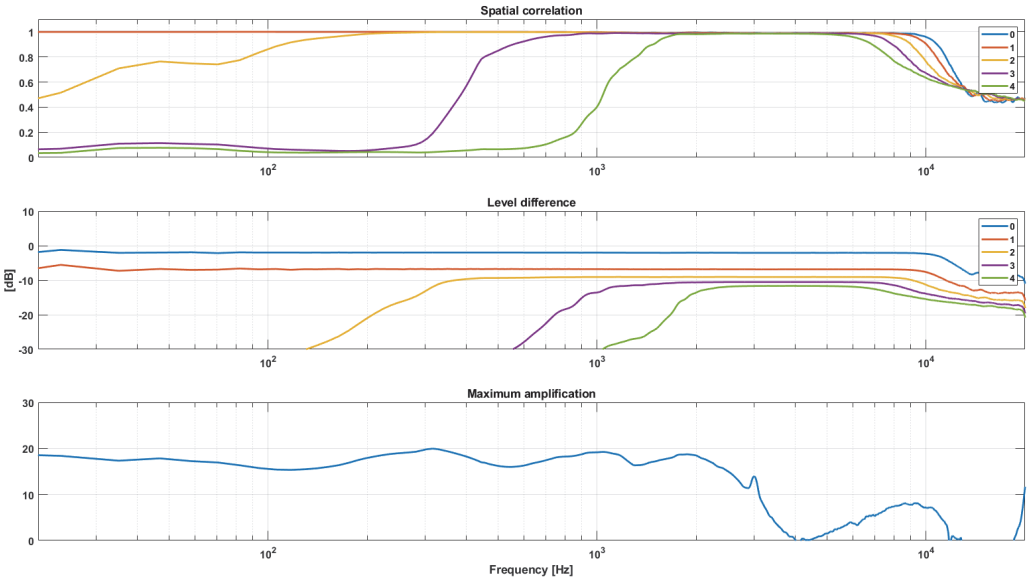


Figure 37: EM, spatial performances with Kirkeby inversion and optimized $\beta(k)$

Again, frequency limits for each order have been defined with previous thresholds; they are summarized in Table 4, in comparison with the ones obtained without the optimization of the regularization parameter β. The analysis is concluded with the PSD of the two filters (Figure 38). Note that the optimization of the regularization parameter produced further lowering of the starting frequency for each order, despite the WNG never exceeds the imposed limit of 20 dB.

	Kirkeby Inversion β not optimized		**Kirkeby Inversion β optimized**	
Ambisonics order	**Freq. start [Hz]**	**Freq. stop [kHz]**	**Freq. start [Hz]**	**Freq. stop [kHz]**
1	25	9.4	20	9.4
2	370	8.4	360	8.4
3	1240	7.3	1200	7.3
4	2300	6.1	2250	6.1

Table 4: EM, frequency limits for Ambisonics orders 1-2-3-4, Kirkeby Inversion

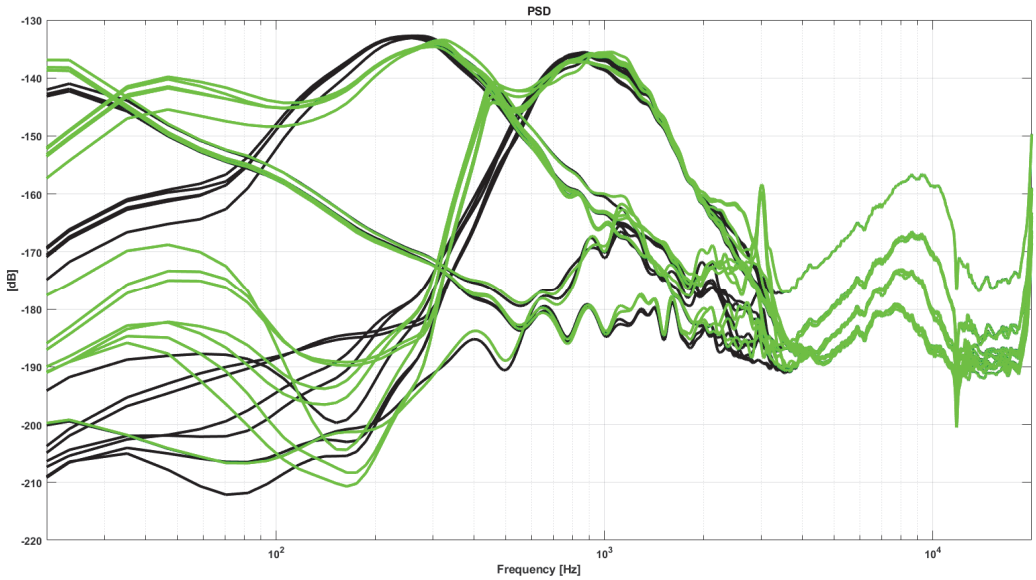


Figure 38: EM, PSD of filters, Kirkeby inversion, $\beta(k)$ non-optimized (black) and optimized (green)

2.2.3. Theoretical solution and numerical array response

The theoretical approach is possible whenever an analytical solution of the physical problem has been found. In this case, the phenomenon can be reduced to a sound wave impinging on the surface of the array and being diffracted.

The analytical solution is available in literature for a plane wave diffracted by a rigid sphere [16], but in the real world things can be much more complex. The plane wave is a first simplification, as the sound wave emitted by point sources, such as loudspeakers or the mouth of people, is spherical. However, after a certain distance, spherical waves can be approximated by plane waves. The shape of the array, which must be spherical, gives another constraint: there is not a general solution for arrays of any shape and each case should be solved specifically, a task that becomes immediately too complex even for a small variation of the geometry. Lastly, the surface of the array is considered infinitely rigid, so no deformation or energy

dissipation is taken into account. A numerical implementation of this solution has been found in [11].

To conclude, three theoretical methods for deriving the Ambisonics encoding matrix are briefly explained below. They relies on different models of the array, defined by the following parameters: radius of the array, number and positions of the capsules. A more exhaustive explanation can be found in [11].

1. *Plain SH transform*

It is the simplest approach: filters are calculated by means of the pseudo-inverse applied to the matrix of SH evaluated in the positions of the capsules. Parameters required: array radius, number and position of capsules.

2. *Radial response inversion with Tikhonov regularization*

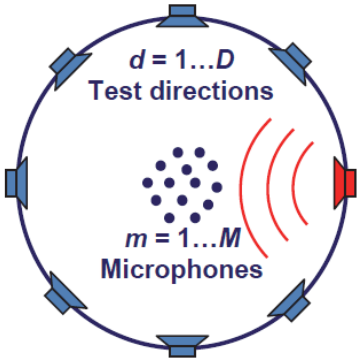
Cited from [11]: "Filters are generated from an inversion of the radial components of the response, neglecting spatial aliasing effects and non-ideal arrangements of the microphones; one filter is shared by all SH signals of the same order. The single-channel inversion problem is done with a constraint on the maximum allowed noise amplification with the Tikhonov regularization [12], [17]". Parameters required: array radius, number of capsules.

3. *Radial response inversion with soft-limiting*

Cited from [11]: "Filters are generated from an inversion of the radial components of the response, neglecting spatial aliasing effects and non-ideal arrangements of the microphones; one filter is shared by all SH signals of the same order in this case. The single-channel inversion problem is done applying a limit to the maximum noise amplification allowed: the limiting follows the approach of [18]". Parameters required: array radius, number of capsules.

2.2.4. Measured array response

The advantage of this method is the possibility to be employed with arrays of any shape, but it requires a complex setup and the availability of an anechoic room. The equipment needed for the measurement is a studio monitor loudspeaker with a very flat frequency response and a two-axis turntable (Figure 40). In addition, the system employed to record the signals coming from the array (A-format) must be provided at least with one output channel, so that the measurement signal can be played synchronously with the recording, preserving the correct time of flight and phase-match between all the measurements. Figure 39 summarizes the measurement method: having the array mounted on a two-axis turntable is the same than having a spherical loudspeakers rig surrounding the array itself, hence when the test signal is played by the loudspeaker, one line of the Spatial Impulse Response (SIR) matrix is obtained.



$$\left\| c[n] \right\|_{D\times M} = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,m} & \cdots & c_{1,M} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,m} & \cdots & c_{2,M} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{d,1} & c_{d,2} & \cdots & c_{d,m} & \cdots & c_{d,M} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{D,1} & c_{D,2} & \cdots & c_{D,m} & \cdots & c_{D,M} \end{bmatrix}$$

Figure 39: Spatial Impulse Response measurement scheme

Once the array is measured, one of the inversion methods described in 2.2.1 can be applied to get the filtering matrix for A-2-B or A-2-P format conversion.

Figure 40: SIR measurement equipment, a two-axis turntable and a studio monitor inside an anechoic room

2.2.5. Simulated array response

This method overcomes together the disadvantages of both anechoic measurement and theoretical solution: it is possible to obtain the response of array, hence its filtering matrix and subsequently the performance evaluation, for geometry of any shape and without having to realize and measure a prototype. Theoretical solution is in fact available only for spherical geometry and building, assembling and measuring each sample is expensive, time consuming and quite impractical, if not impossible in case of underwater array. Therefore, a methodology to perform FEM simulation has been developed in COMSOL Multiphysics. The output is close to the anechoic measurement: rather than recording the pressure in time domain, the spectra of sound pressure is evaluated at microphone positions for all the directions tested.

The simulated array response is then inverted to calculate the filtering matrix. This operation can be done directly from the frequency domain, where the simulation result is, or converting the latter back to time domain. All simulations have been performed with plane waves, leaving the spherical wave radiation for further development.

2.2.5.1. Modeling

The study started with a very simple geometry, a sphere of 40 mm diameter and four capsules placed in the vertex of a tetrahedron. The model employed the PML (Figure 41), a widely used solution to minimize reflections of outgoing waves. Initially, only the central frequencies of the octave bands were simulated. A set of test directions has been defined in the model as *specific combinations* of a *parametric sweep*. An exhaustive explanation of the simulation grid is provided in 2.2.5.3, page 34.

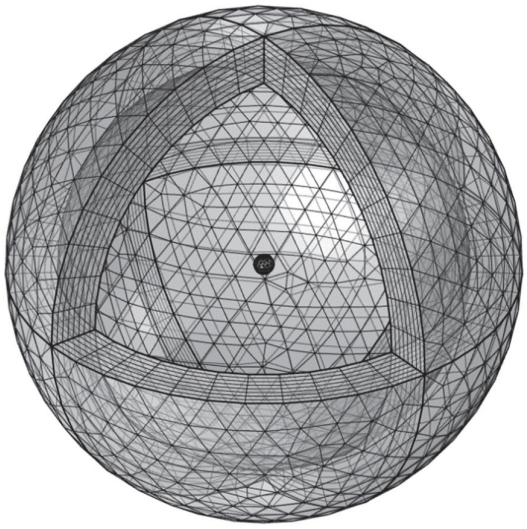


Figure 41: Meshed model of a sphere with four capsules inside a PML

A FOA filtering matrix has been synthesized at the simulated frequencies and the 3D plots of the virtual microphones have been produced (Figure 42), to ensure they were reconstructed properly.

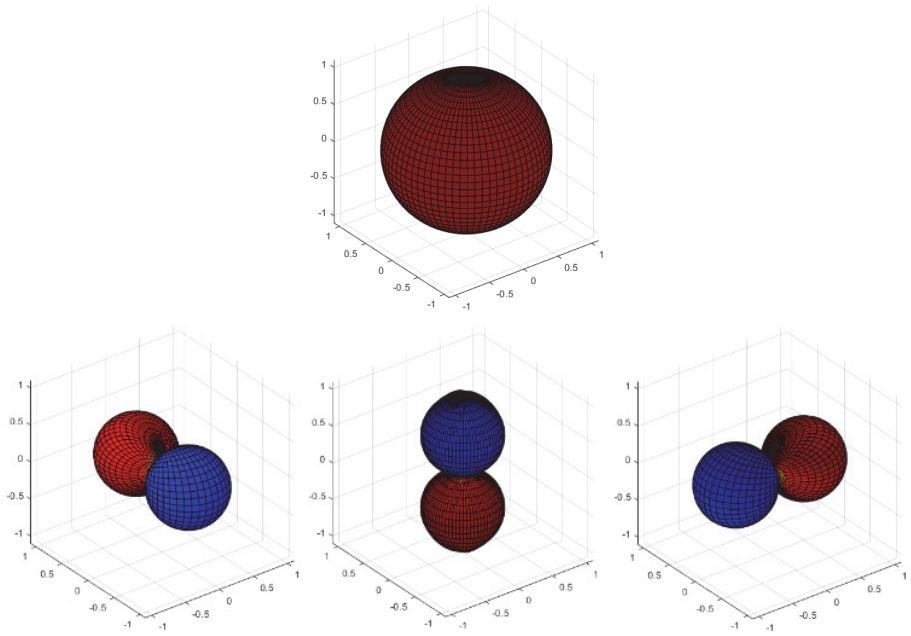


Figure 42: 3D plot of the directivity patterns of FOA virtual microphones

Subsequently several improvements have been introduced with the aim of speeding up the solution and improving the results. The PML has been substituted by different method to define the wave radiation field (Figure 43). A part from being much faster, with this method it is also very practical to switch between plane waves and spherical waves just changing a property of the incident pressure field, a great advantage for further developments. In addition, the parametric definitions of the testing directions has been improved with an auxiliary sweep, again of specific combinations of azimuth and elevation. With same parameters for the mesh and employing the exterior radius of the PML for the radius of the radiation field, the calculation time reduced dramatically of a factor 40. Again, the validity of numerical results has been checked by looking at the 3D plots of the virtual microphones, which remained identical to the ones of Figure 42.

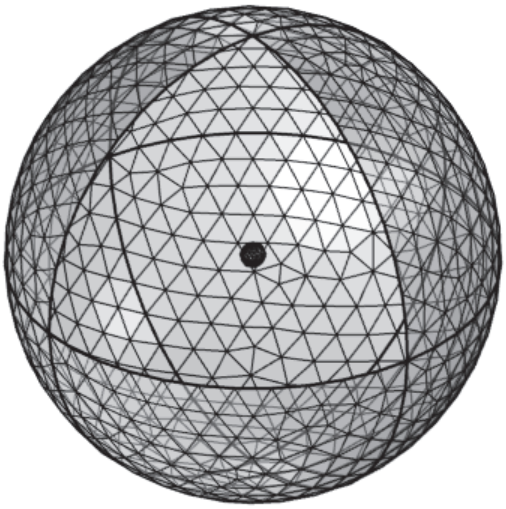


Figure 43: Meshed model of a sphere with four capsules inside a Wave Radiation Field

Initially, simulations have been done employing the pressure-acoustics module without any multiphysics coupling. Hence, the surface of the array is considered infinitely rigid and only the material of the domain (e.g. air or water) has to be defined. In this condition, only the diffraction of the plane wave over the surface of the geometry is taken into account, as for the numerical solution described in paragraph 2.2.3. Then, multiphysics coupling between acoustic and solid domains has been introduced and evaluated (paragraph 2.2.5.4, page 39).

2.2.5.2. Validation of the method

The correct reconstruction of the directivity patterns of FOA virtual microphones (Figure 42) provides an effective validation of the model. However, the simulation has been checked also against the numerical solution of the theoretical model explained in 2.2.3, employing the metrics described in 2.2.2.1.

The numerical solution of plane wave diffracted by a rigid sphere has been calculated for the small sphere with four microphones showed in 2.2.5.1., with a test grid of 62 directions, corresponding to the spherical design T-10 (see next paragraph for details regarding spherical T-design geometries).

The simulation has been calculated in the range $20\ Hz - 7\ kHz$ with same test grid and a frequency resolution $df = fs/nfft = 48000/2048 = 23.4375\ Hz$. The upper frequency limit has been chosen in order to limit the computation time.

Spatial performances have been evaluated by inverting the numerically calculated response and the simulated response (converted from frequency to time domain), with method 1 (described in paragraph 2.2.1), $WNG_{max} = 20\ dB$ and length of the filters $2048\ samples$. Results are presented in Figure 44 and Figure 45. As the simulation

has been solved up to 7 kHz, this is the upper frequency showed in both figures. One can note that results are identical.

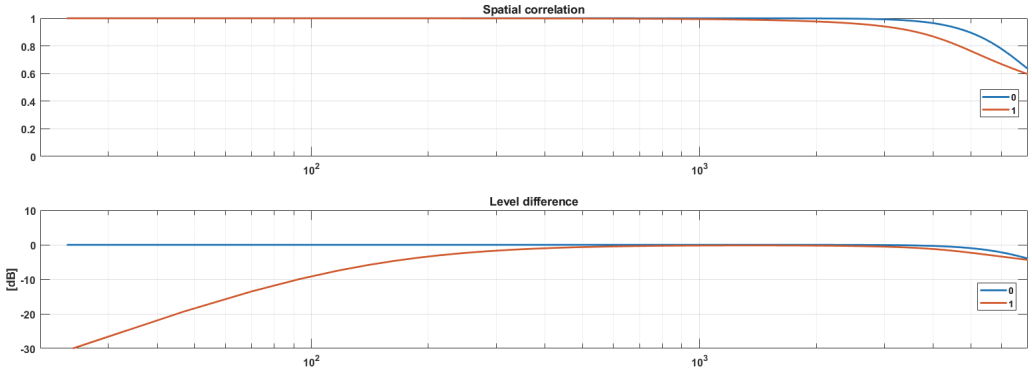


Figure 44: Spatial performance of the test sphere with four microphones, numerical response

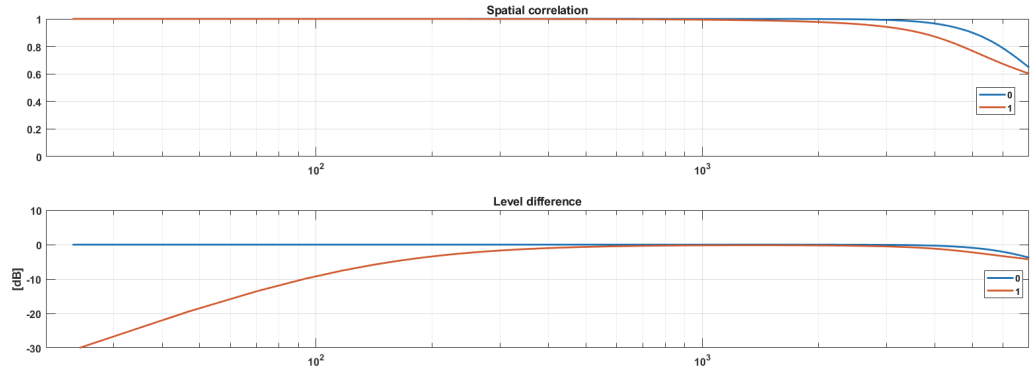


Figure 45: Spatial performance of the test sphere with four microphones, simulated response

The same verification has been done also for the EM, which exhibits a more complex geometry. The geometry has been designed in a 3D CAD software and a model has been set up with the improvements described in the previous section (Figure 46). In this case, the simulation has been computed up to 3.5 kHz to limit the computation time required, as the array is bigger, thus the mesh requires much more elements and, in addition, it is a HOA array, therefore also the number of testing directions has to be larger. A test grid of 242 directions, corresponding to the spherical design T-21 (see following paragraph) has been used. Again, the maximum frequency showed in the figure corresponds to the highest simulated frequency.

Both numerical and simulated solutions have been inverted with method 1, $WNG_{max} = 20\ dB$ and length of the filters 2048 $samples$. LD curves (Figure 48) are identical, whilst SC curves (Figure 47) are slightly different: the simulated model is not perfectly spherical due to the presence of the lower body of the array.

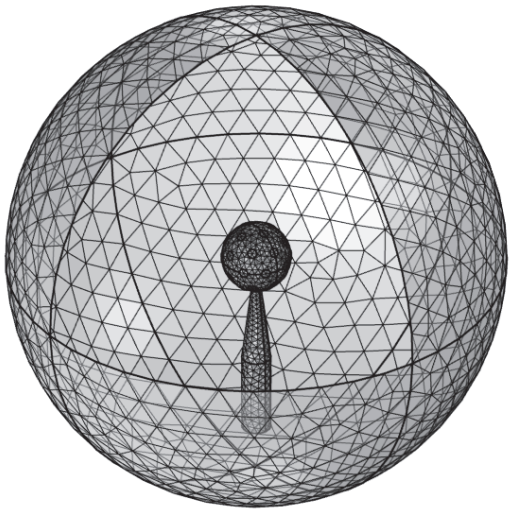


Figure 46: Meshed model of the Eigenmike32™

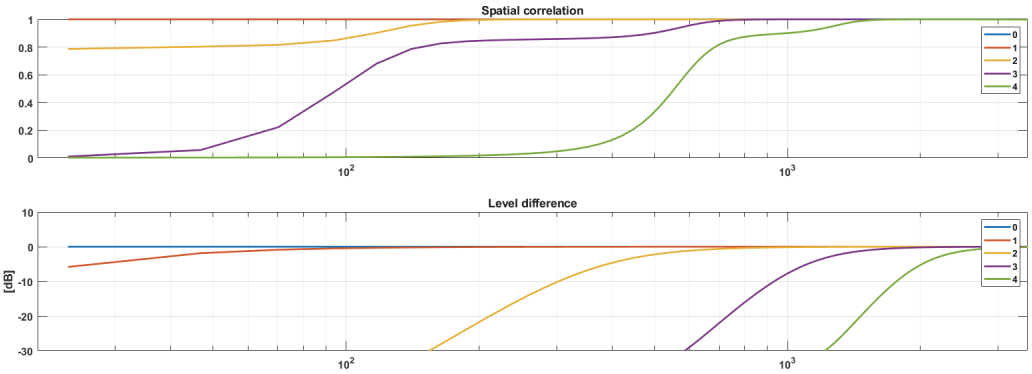


Figure 47: EM, spatial performance of theoretical response, Ambisonics order 4th

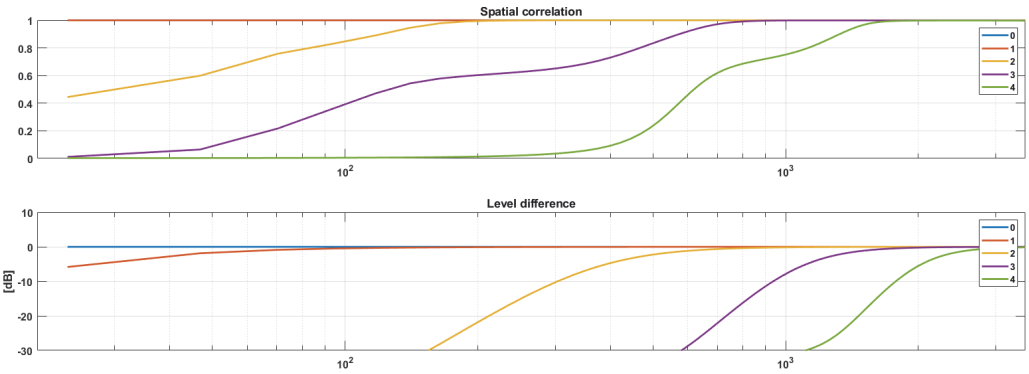


Figure 48: EM, spatial performance of simulated response, Ambisonics order 4th

2.2.5.3. Simulation grid: nearly-uniform balloon and T-design geometry

The grid of testing directions is another important aspect to be considered: in fact, a reduction of the number of directions to test will result in a reduction of the computation time required.

The technique developed at University of Parma for measuring the arrays in anechoic room is based on a *nearly-uniform grid* [19]. T-design geometries [20] up to order 21 have been tested against four nearly-uniform grids, respectively with 362, 241, 122 and 60 directions and two monospaced grid of 666 and 171 points. T-design and nearly-uniform geometries have been slightly modified, adding two points in correspondence of the poles, which are very useful to improve the colour maps, as explained in the next chapter. Theoretical model and simulations have been solved employing these grids to investigate the geometry which provides best result with the minimum number of points to encode various Ambisonics orders.

The FOA case (sphere of $40mm$ diameter with four microphones in the vertices of the tetrahedron) has been solved numerically and the response inverted with method 1 (paragraph 2.2.1), $WNG_{max} = 20\ dB$ and length of the filters $2048\ sample$. The following grids have been tested for this case: monospaced with 171 points (Figure 49, left), nearly-uniform with 122 points (Figure 49, right), nearly-uniform with 62 points (Figure 50, left) and T-10 design with 60 points (Figure 50, right).

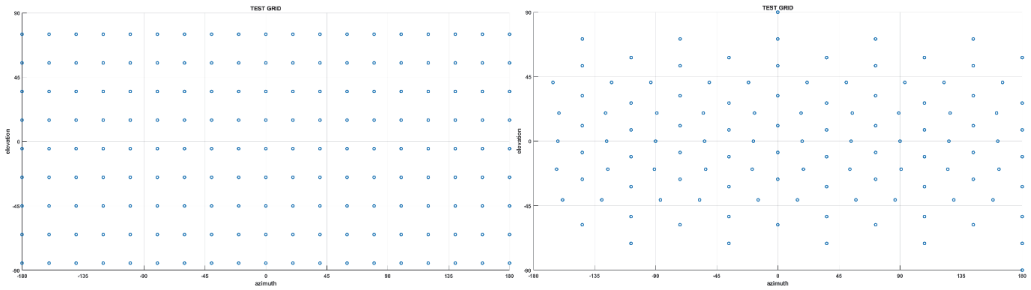


Figure 49: Monospaced grid with 171 points (left) and nearly-uniform grid with 122 points (right)

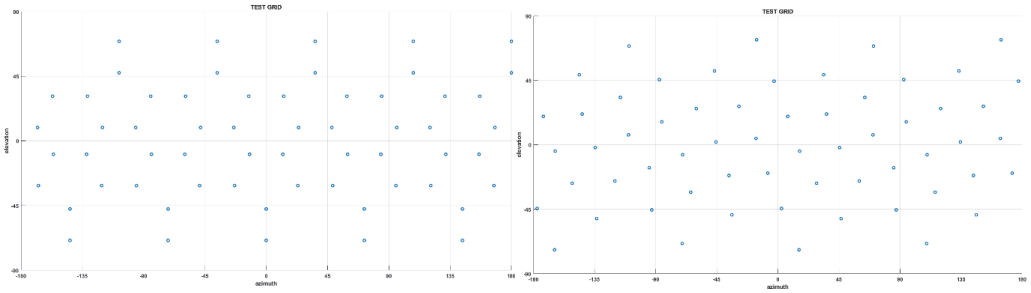


Figure 50: Nearly-uniform grid with 60 points (left) and T-10 design with 60 points (right)

Results are presented in the following in terms of SC and LD for each case. Figure 52 shows results for the nearly-uniform grids with 122 and 60 points and T-10 design with 62 points, as they are substantially identical. It appears from Figure 51 that

monospaced grid should be avoided, as spatial performance observed is equal or lower despite the high number of points. The nearly-uniform grid with 122 points can be discarded too, having twice the number of points of the two remaining grids, T-10 design and nearly-uniform with 60 points, but providing the same performance. The latter instead have almost the same amount of points and provide similar spatial performance, therefore a further analysis is required.

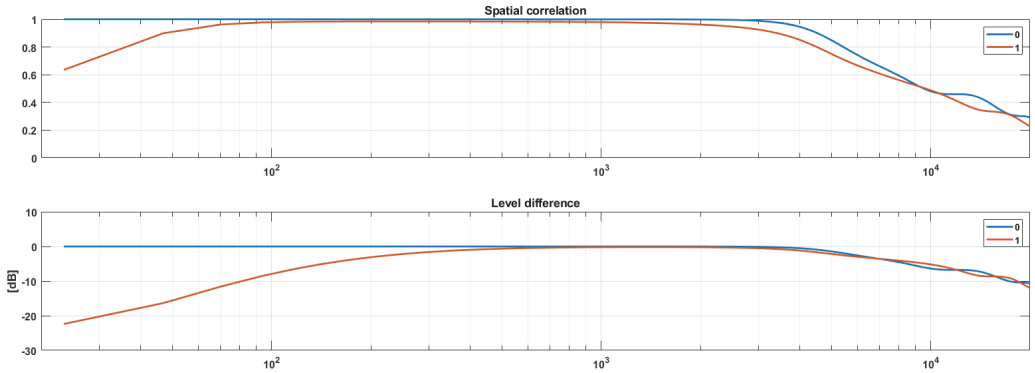


Figure 51: Evaluation of monospaced grid, 171 points, 1st order Ambisonics

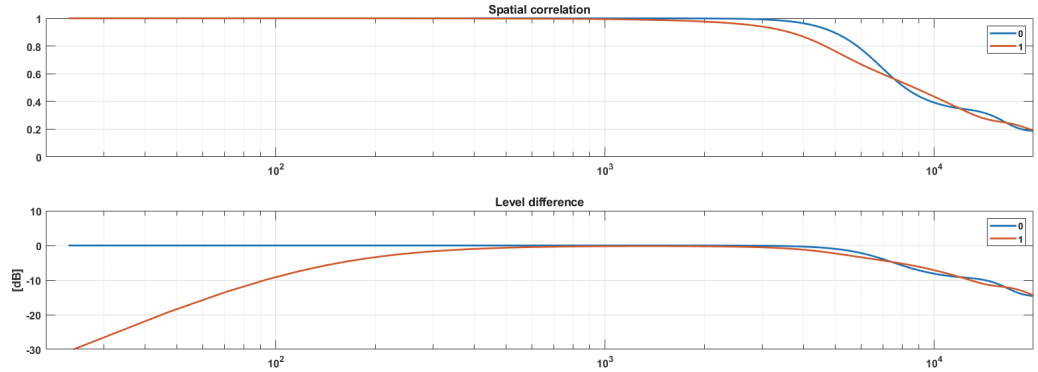


Figure 52: Evaluation of nearly-uniform grids, 122 and 60 points, and T-10 grid, 62 points, 1st order Ambisonics

Two FEM simulations of the sphere with four microphones (model is shown in Figure 43) have been computed with the nearly-uniform grid with 60 directions and T-10 design grid. In order to reduce the computation time at minimum, only the central frequencies of the octave bands up to 2 kHz have been simulated. The 3D plot of the directivity patterns of the four virtual microphones of FOA have been produced and compared.

Figure 53 shows the comparison of the Y component at 1 kHz. It is possible to note that the nearly-uniform grid with the 62 points produced a visible distortion of the polar pattern which instead is not given by the T10 design, that in conclusion is preferable.

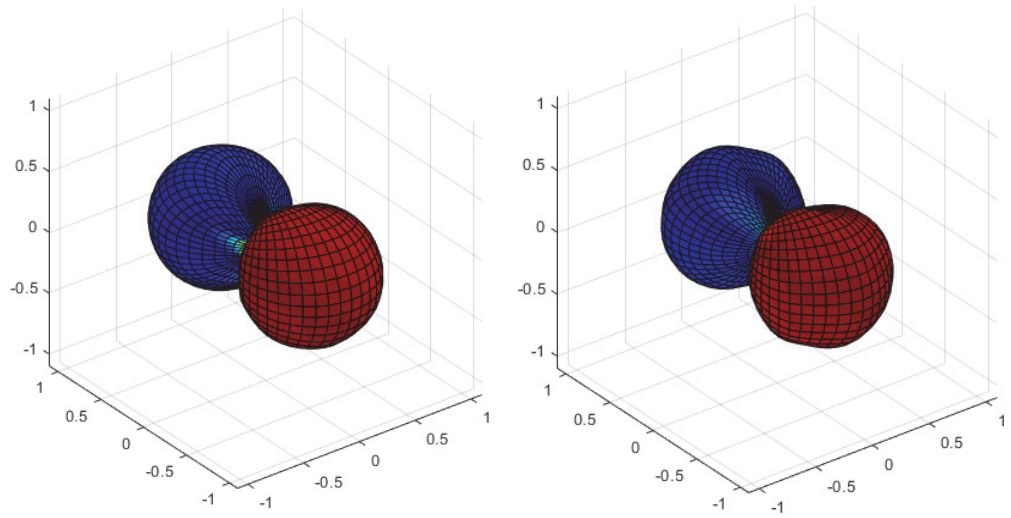


Figure 53: FOA, Y component at 1 *kHz, simulation grids: T-10 (left), nearly-uniform 62 (right)*

A HOA case has been investigated too, by calculating the numerical solutions for the geometry of the EM, again employing inversion method 1, $WNG_{max} = 20\ dB$, length of the filters 2048 $samples$ and Ambisonics up to order four. This time the tested grids are monospaced with 666 points (Figure 54, left), nearly-uniform with 362 points (Figure 54, right), nearly-uniform with 241 points (Figure 55, left) and T-21 design with 240 points (Figure 55, right).

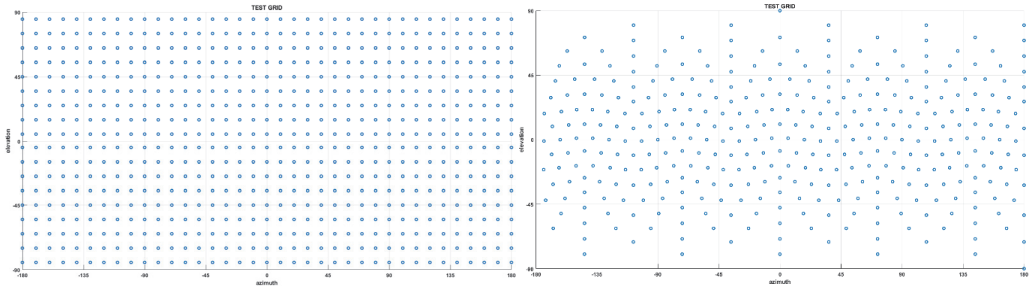


Figure 54: Monospaced grid with 666 points (left) and nearly-uniform grid with 362 points (right)

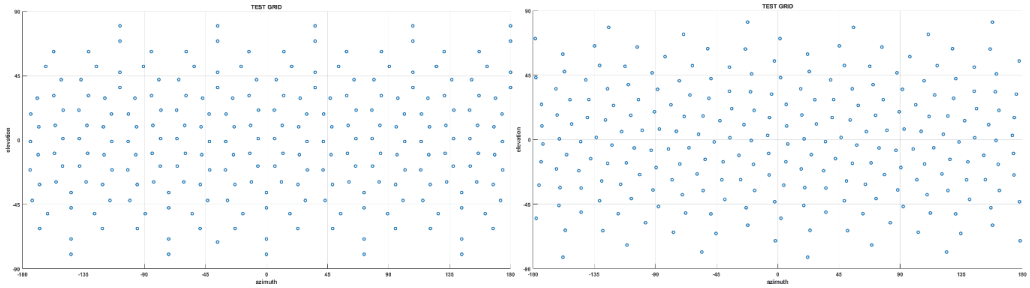


Figure 55: Nearly-uniform grid with 241 points (left) and T21 design with 240 points (right)

Results are presented in the following figures for each grid, in terms of SC and LD. Even if the frequency limits for the various orders, calculated with thresholds defined in (12) and (13), are similar, the analysis of these results allows to identify the

T-21 geometry as the preferable solution, as it provides the best compromise between number of points and spatial performance of the filters.

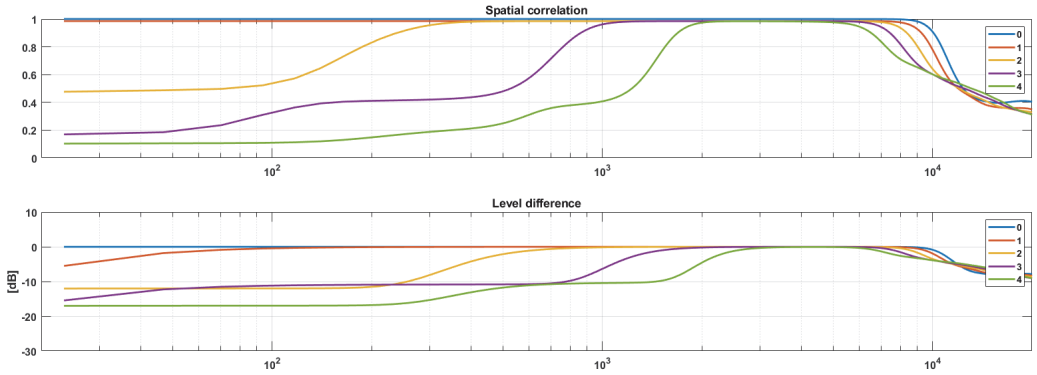


Figure 56: Evaluation of monospaced grid, 666 points, 4th order Ambisonics

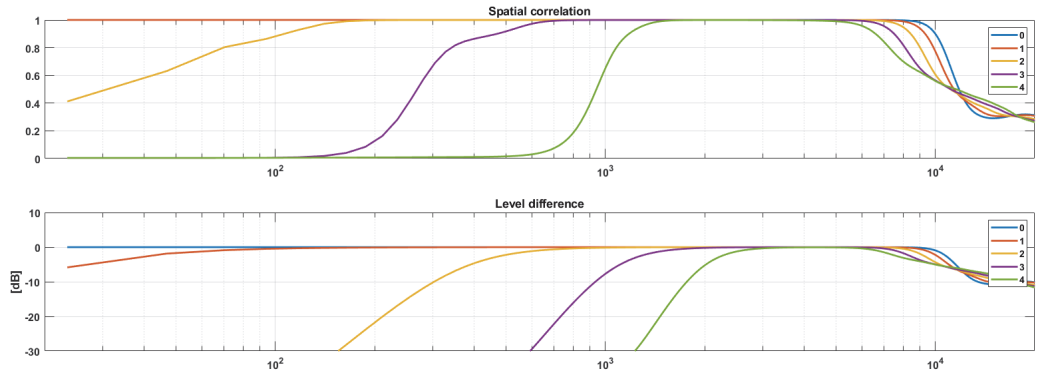


Figure 57: Evaluation of nearly-uniform grid, 362 points, 4th order Ambisonics

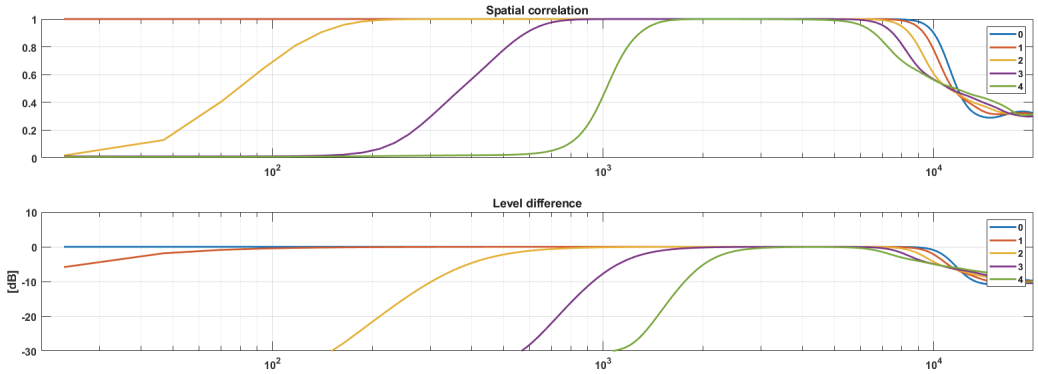


Figure 58: Evaluation of nearly-uniform grid, 241 points, 4th order Ambisonics

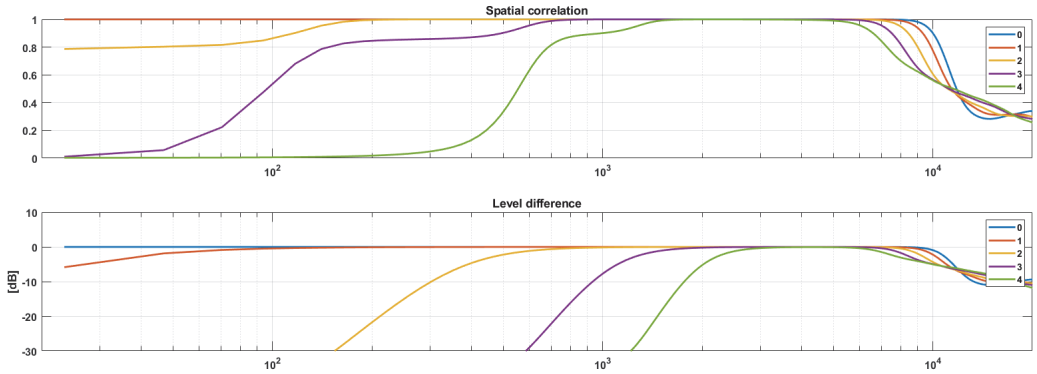


Figure 59: Evaluation of T-21 grid, 242 points, 4th order Ambisonics

In order to confirm the results of this analysis, FEM simulations of each case have been compared, employing the model of Figure 46 and calculating the solutions for the central frequencies of the octave bands up to 2 kHz. Figure 60 shows the 3D plot of the directivity patterns of the first 16 virtual microphones (Ambisonics 3rd order) at 1 kHz.

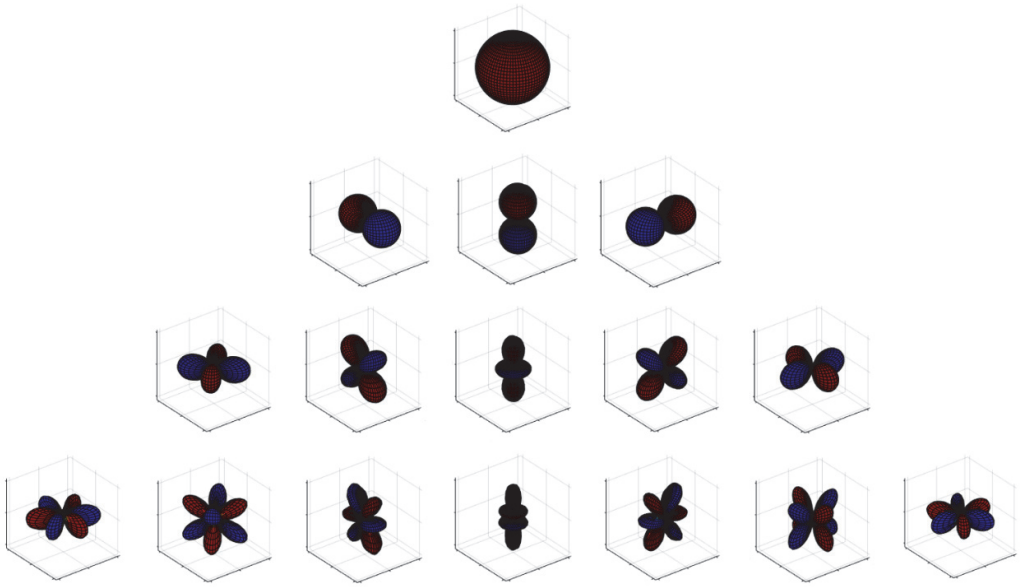


Figure 60: 3D plot of the Ambisonics 3rd order SH, simulated with a T-21 grid of directions

Several distortions has been observed by employing nearly-uniform grid with 241 directions. The behaviour of the monospace grid with 666 points, nearly-uniform grid of 362 points and T-21 design is almost identical at higher frequencies (above 500 Hz). However, the T-design distribution presents better performances at low frequencies and the minimum number of points with respect to the others. Therefore, it is the best solution for simulating or measuring microphone arrays for Ambisonics up to order four. To conclude, an analogue analysis has been performed for a second order microphone, testing monospace grid with 300 points, nearly-uniform grids with 362 and 241 points and T-15 design. All results are summarized in Table 5.

Ambisonics order	Min. T-design grid	Min. nearly uniform grid
1	T-10 – 60 points	122 points
2	T-15 – 120 points	362 points
3	T-21 – 240 points	362 points

Table 5: Minimum T-design and nearly-uniform geometries for each Ambisonics order

2.2.5.4. Multiphysics coupling: Acoustic – Structure

The "Solid Mechanics" physic has been introduced in the model: the multiphysics coupling between *pressure-acoustics* and *solid-mechanics* physics takes into account the interaction of the pressure field with the surface of the array, being the latter not infinitely rigid. The behaviour of materials is often frequency-dependent and it is quite difficult to find information on this relation. For this reason, it is desirable that the influence of the mechanic coupling would be negligible. Each case should be evaluated individually, as it depends on the materials of the array and of the domain.

As an example, three simulations of the EM have been solved in the range $11.71875\ Hz - 1417.96875\ Hz$, employing the model of Figure 46 and a frequency resolution $df = fs/nfft = 48000/4096 = 11.71875\ Hz$. In the first case, the solid-mechanics physic was not introduced, thus considering the surface of the array infinitely rigid: in this condition, the simulation is very similar to the theoretical model. In the second case, multiphysics coupling between pressure-acoustics and solid-mechanics physics has been introduced, employing aluminium for both the frame and the capsules, with COMSOL built-in properties. The third case has been solved again with the multiphysics coupling and employing rubber for the material of the array. Built-in properties of NBR (25-acrylonitrile and 75-butadiene) have been used for the rubber: density $\rho = 1000\ kg/m^3$, Young's modulus $E = 4e6\ Pa$ and Poisson's ratio $\nu = 0.48$. In both cases, no damping has been used for the materials.

Filters have been produced for the three cases with inversion method 1 (paragraph 2.2.1) and their frequency responses have been compared through the PSD, which are showed superimposed (Figure 61) in the frequency range $20\ Hz - 1.4\ kHz$, respectively in red, blue and black for cases one, two and three. One can note that the solid-mechanics physic is negligible for the aluminium, as expected in air.

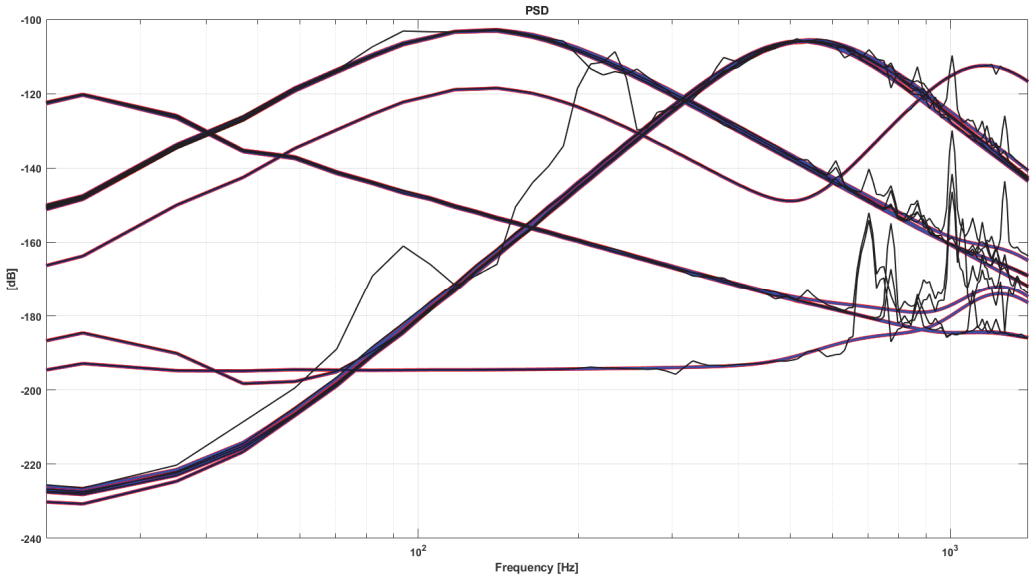


Figure 61: Evaluation of the acoustic-mechanics coupling

2.2.5.5. On-axis measured filter

It has proved that FEM simulations can overcame the need to measure the array inside an anechoic room. However, spatial performance of the filtering matrix calculated with simulations can improve if some information from a measurement are taken into account, as discussed in the following.

The simulation of the EM has been inverted to produce a SPS filtering matrix: 36 virtual microphones with direction defined by T-8 design, directivity of type super-cardioid of order eight, filters of length 4096 $samples$, $WNG_{max} = 20\ dB$, Kirkeby inversion without optimizing regularization parameter β.

The evaluation of the spatial performance presented in 2.2.2.2 was performed by convolving the filtering matrix with the same response of the array, whether calculated, measured or simulated, employed to encode the filtering matrix itself and then results were compared to the ideal directivity of the target. Now instead, the evaluation of the spatial performance is done by convolving the filtering matrix encoded from the simulated response with a measured response of the array. Figure 62 presents the result in terms of SC and LD, Figure 63 shows the real directivity obtained. Note that at lower frequencies the directivity is quite poor.

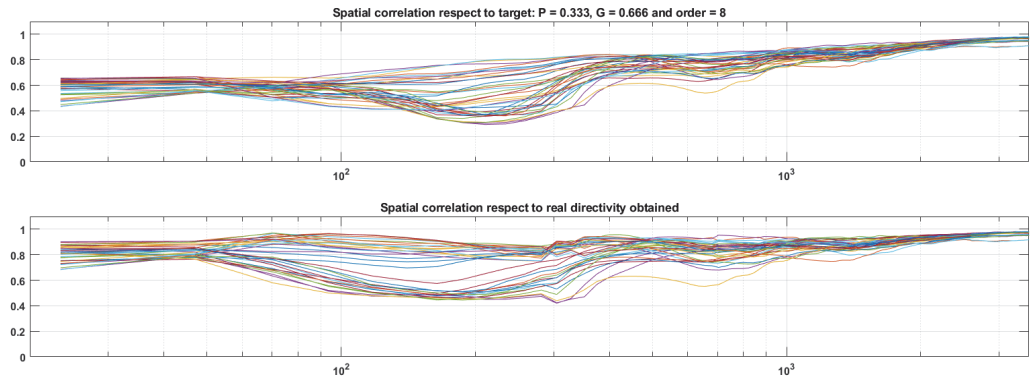


Figure 62: EM, spatial performance of a SPS filtering matrix

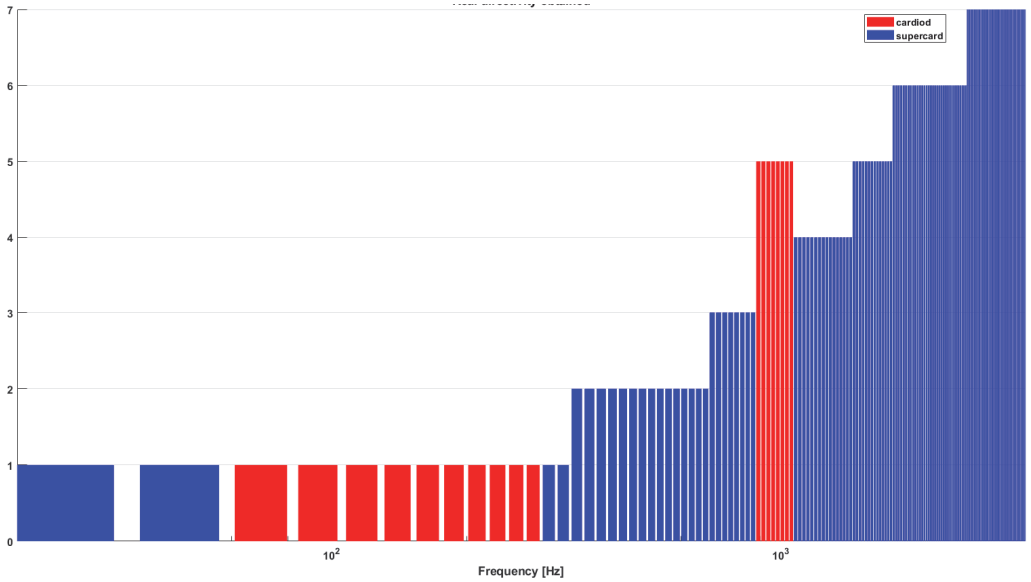


Figure 63: EM, directivity in function of the frequency of a SPS filtering matrix

The method developed to improve this result consists in employing an anechoic measurement of the capsules of the array, made on-axis with respect to the loudspeaker. The main advantage of the simulation technique, that is not requiring an anechoic chamber and a complex measuring equipment, is lost. However, the time required for the measurement of the capsules on-axis is much shorter respect to a complete measurement of the SIR.

Once the IRs of the capsules are measured on-axis respect to the loudspeaker, they must be realigned at sample. This can be done with the following sequence of operations:

- Up-sampling of the measured impulse responses (i.e. by a factor 32);
- Discrete-time Hilbert transform [21];
- Realignment of all the signals to the maximum absolute peak of the Hilbert transform;
- Down-sampling of the realigned impulse responses by the same factor of the up-sampling.

Finally, their inverse filters are calculated. To evaluate them, they are convolved with the corresponding direct filter. Figure 64 shows the result of this convolution: in time domain, it is almost a perfect Dirac delta function, therefore in frequency domain the response is flat, in this case in the range $30\ Hz - 13\ kHz$.

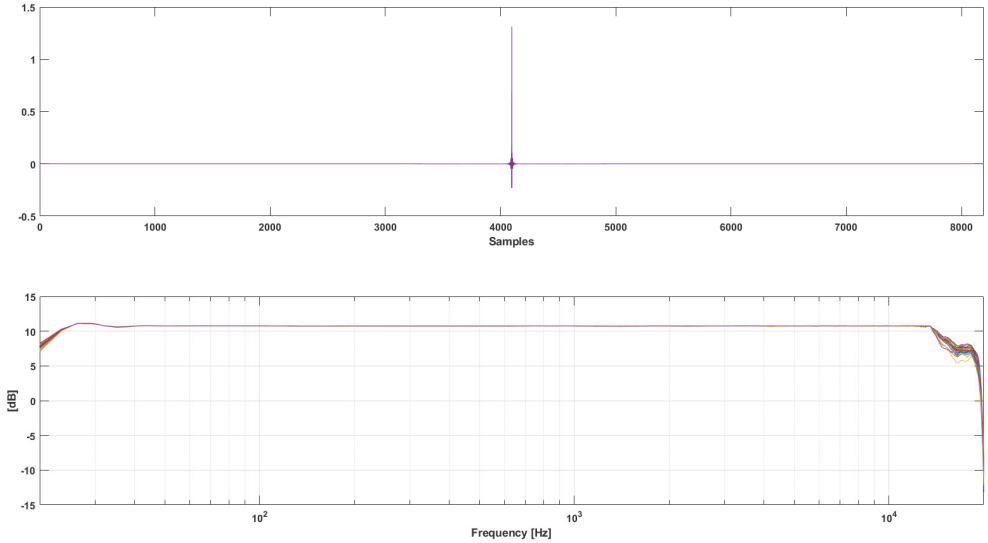


Figure 64: Evaluation of the inverse filters of IRs of the capsules of an array measured on-axis

The inverse filters of the IRs measured on-axis are convolved with the SPS filtering matrix. In this way, when the filtering matrix is applied to a recording of the array, the response of the capsules is corrected. Spatial performance of the "corrected" SPS filtering matrix are evaluated, by convolution with the response of the array measured in anechoic room (Figure 65 and Figure 66). Note that spatial performance improved considerably at low frequency but decreased at high frequency. In fact, the on-axis response is measured with capsules already mounted on the array, causing the diffracted field to be taken into account twice, once with the anechoic measurement on-axis and once with the simulation. This effect becomes relevant at high frequency, whilst at low frequencies the wavelength is long enough to avoid it. In reverse, the phase information gets more and more importance lowering the frequency, providing a consistent improvement of the beamforming capability.

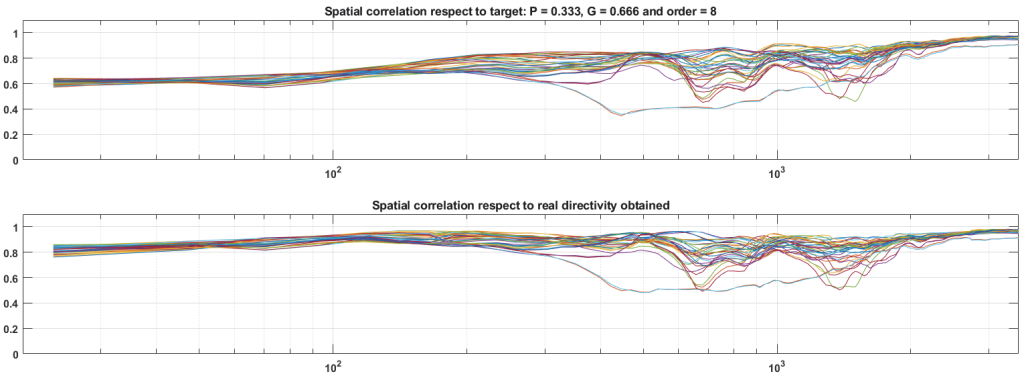


Figure 65: EM, spatial performance of a SPS filtering matrix corrected with on-axis response

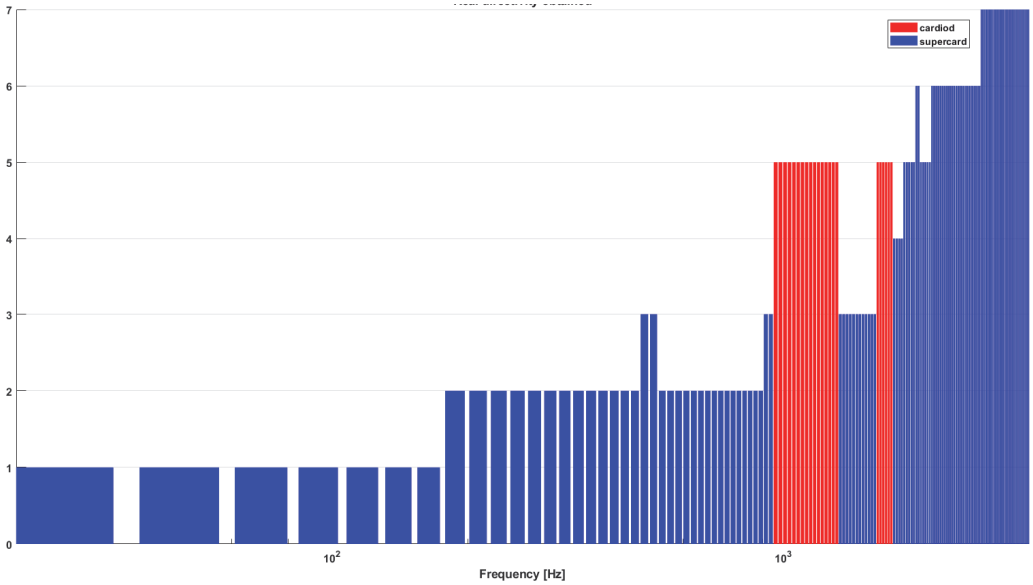


Figure 66: EM, directivity in function of the frequency of a SPS filtering matrix corrected with on-axis response

To overcome this problem, there are two possibilities. The first one, preferable but difficult in practice, consists in measuring the capsules dismounted from the array, so that the diffracted field is not present at all. In this way, the phase information can be employed at all frequencies. The second solution consists in measuring the capsules when already mounted on the array and then filtering the responses, so that the phase information is introduced only at low frequencies.

In this case, the second solution has been adopted: the filtering matrix has been corrected with a low-pass filtered version at $650\,Hz$ of the inverse filters of the capsules measured on-axis. Results are showed in Figure 67 and Figure 68. Note that the spatial performance at low frequencies is now improved without any reduction of spatial performance at high frequencies, proving the effectiveness of this method.

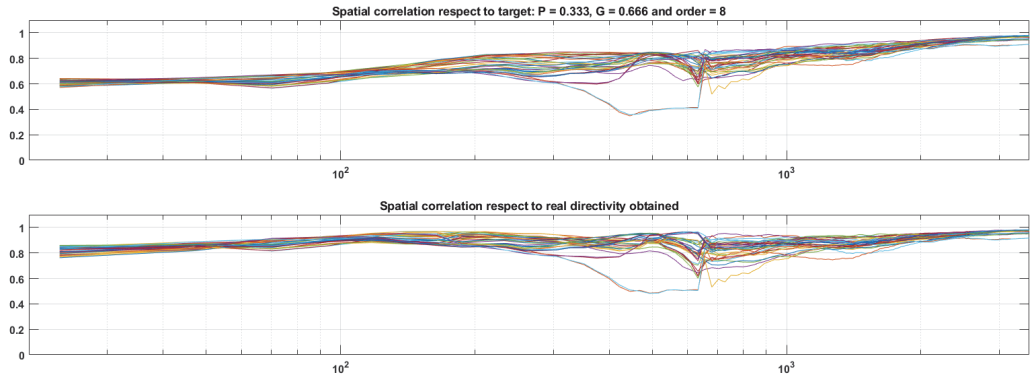


Figure 67: EM, spatial performance of a SPS filtering matrix corrected with pre-filtered on-axis response

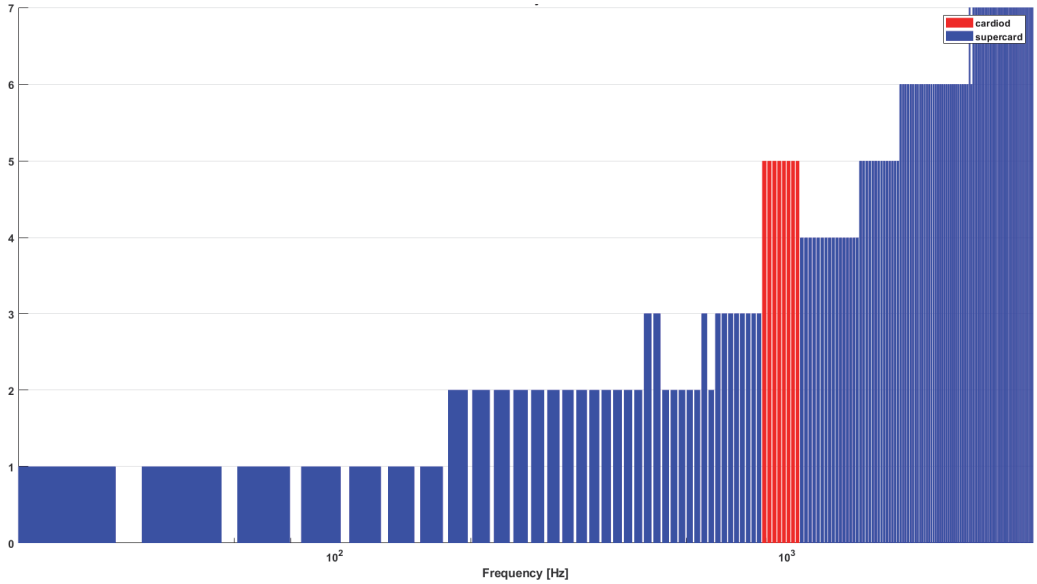


Figure 68: EM, directivity in function of the frequency of a SPS filtering matrix corrected with pre-filtered on-axis response

2.2.5.6. BEM simulation

Boundary Elements Method is a different way to perform simulations: the solution is calculated on the surfaces of the model, the only entities where mesh is defined, by solving the equations of the chosen physics. Depending on the frequency range and the dimension of the model, and consequently on the number of elements of the mesh, BEM can be slower or much faster than FEM. For this reason, a comparative test is very helpful to define the frequency under which it is convenient to use FEM and above which it is convenient to use BEM.

As an example, the model of the EM has been solved with both methods, without multiphysics, for a single direction at various frequencies: 1 kHz, 2 kHz, 3 kHz and 4 kHz. In Table 6, the time required for each simulation is shown.

	Time [s]	
Frequency [kHz]	**BEM**	**FEM**
1000	92	7
2000	93	18
3000	94	82
4000	99	495

Table 6: FEM and BEM, comparison of the simulation time

The time required to solve the BEM model, almost constant at all frequencies, is greater than the time required for FEM solution up to 2 kHz. At 3 kHz the two

methods are comparable but at 4 kHz the time of FEM simulation literally explodes, making BEM preferable.

Nevertheless, BEM method will not be employed in the following discussion. The new microphone array is optimized for low frequencies, hence simulations will be performed in the range $20\ Hz - 3.5\ kHz$, which is not high enough to require the transition to the BEM model. Instead, the FEM simulation of new hydrophone array is quite fast: the material of the domain is water and the wavelength inside water is almost 4.5 times respect to the air. Therefore, the number of elements of the mesh remains quite low to have an acceptable computation time.

2.2.6. Comparison of calculated, measured and simulated response

Finally, the three different types of responses have been compared, employing the EM:

- Numerically calculated response, T-21 grid (240 points), length of the response $2048\ samples$;
- Anechoic measurement, nearly-uniform grid (362 points), length of the response $1024\ samples$;
- Simulated response, T-21 grid (240 points), $nfft = 4096$ and frequency resolution $df = \frac{fs}{nfft} = \frac{48000}{4096} \cong 11.72\ Hz$.

All of them have been inverted with Kirkeby method (paragraph 2.2.1), $WNG_{max} = 20\ dB$ and equalization of frequency-dependent parameter β in function of the maximum noise amplification. Filtering matrices for encoding Ambisonics up to order four with length of the filters $4096\ samples$ have been produced and spatial performances have been evaluated (Figure 69, Figure 70 and Figure 71). Then, frequency limits for each Ambisonics order have been obtained for the three cases (Table 7), employing thresholds defined in (12) and (13). Finally, the frequency responses of the filters have been compared by means of the PSD (Figure 72). By comparing Table 7 and Table 2, one can note that the spatial performance of the filters obtained with Kirkeby inversion are better than the one obtained with inversion method 1.

The behaviour of the theoretical and simulated responses is identical at low frequencies: in fact, the main difference between the two methods, which is the body of the array, is irrelevant at such a low frequency. Conversely, the frequency range where this characteristic would be very impactful is not reached with the simulation.

It is possible to note that, in general, the results obtained with the measurement are slightly better.

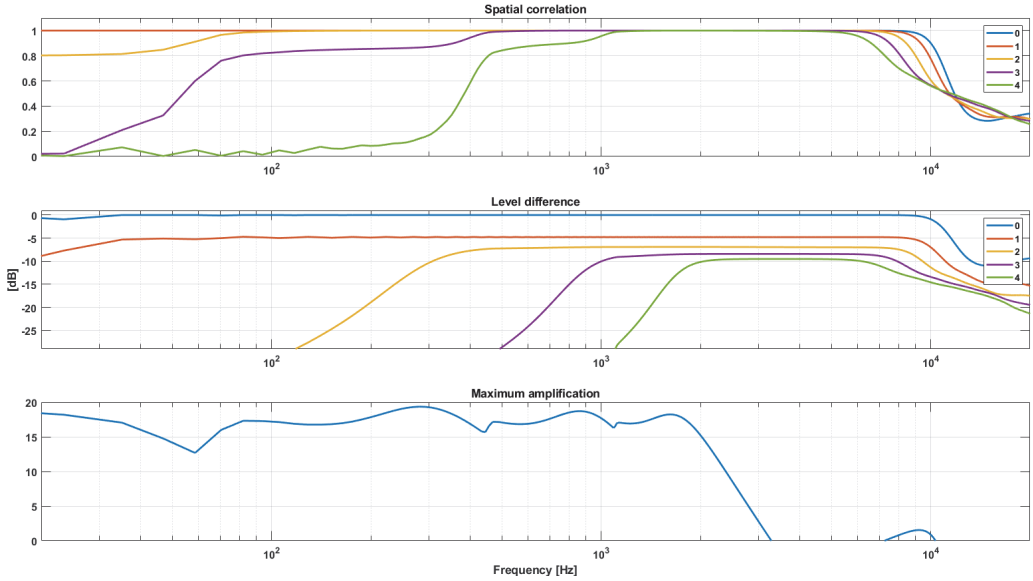


Figure 69: EM, spatial performance evaluation, numerical response, Ambisonics 4th order

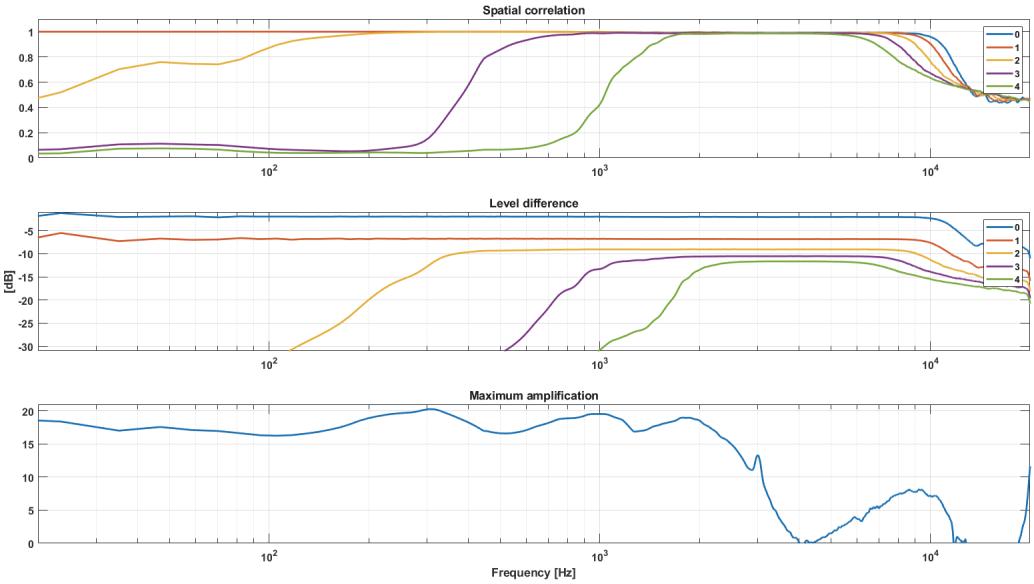


Figure 70: EM, spatial performance evaluation, measured response, Ambisonics 4th order

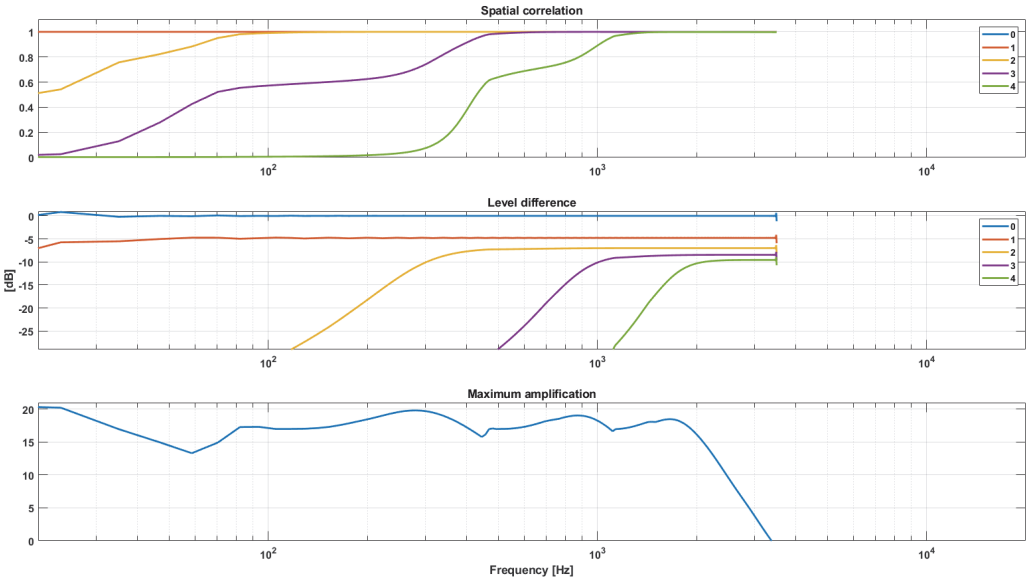


Figure 71: EM, spatial performance evaluation, simulated response, Ambisonics 4th order

	Eigenmike Numerical response		Eigenmike Measured response		Eigenmike Simulated response	
Ambisonics order	**Fr. start [Hz]**	**Fr. stop [kHz]**	**Fr. start [Hz]**	**Fr. stop [kHz]**	**Freq. start [Hz]**	**Fr. stop [kHz]**
1	30	8.8	20	9.4	30	> 3.5
2	430	7.9	360	8.4	430	> 3.5
3	1250	7.0	1200	7.3	1250	> 3.5
4	2050	6.1	2250	6.2	2050	> 3.5

Table 7: EM, frequency limits for Ambisonics orders 1-2-3-4, various array responses

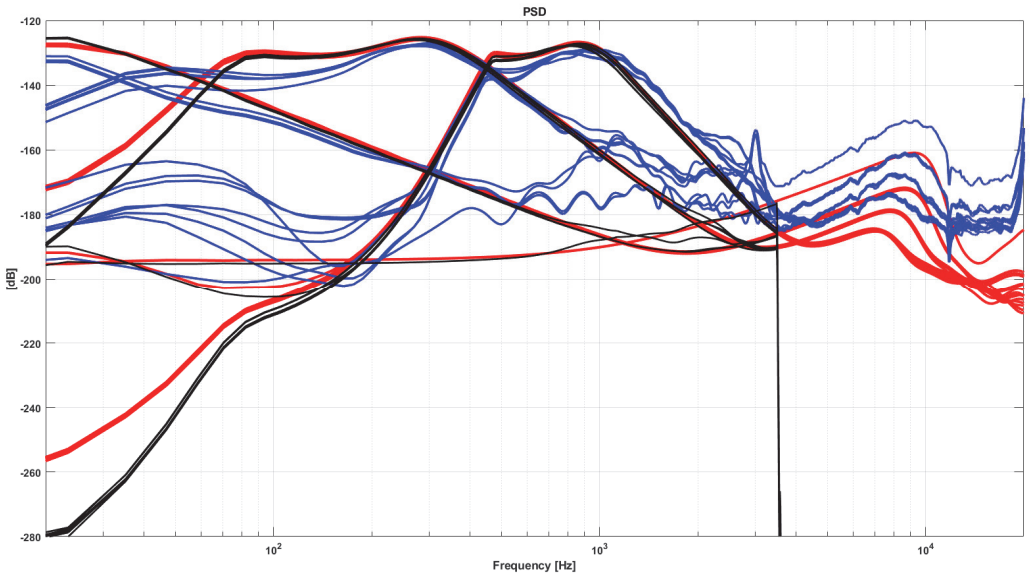


Figure 72: EM, PSD of the filters with numerical (red), measured (blue) and simulated (black) responses

2.3. Design of a microphone array optimized for low frequency

The new microphone array has been designed in order to meet the following requirements:

- Working frequency range of beamforming extended toward low frequencies, making it suitable for NVH applications;
- Capability to encode at least third order Ambisonics to ensure great spatial accuracy at higher frequency;
- Integrating a high quality video recording system to produce 360° panoramic video for virtual reality reproduction with visors and HMDs;
- Relatively small size, so that it can be placed easily inside small environments, such as car cockpits;
- Possibility to mount it on a microphone stands or in a dummy torso;
- Possibility to record synchronously some additional signals to be used for enhanced analysis with cross-correlation filtering.

A large number of channels is required to encode HOA signals. For this reason, the electronic circuitry of an Eigenmike32™ has been employed: in a previous collaboration between University of Parma and RAI – Radiotelevisione Italiana, an EM microphone was disassembled and reassembled in a cylindrical array [22], keeping the preamplifiers and ADC in a dedicated casing, detachable from the rest of the cylinder (Figure 73). The electrical interfacing between the electronic and the microphones is operated by two IDC female connectors with 36 contacts each one, 32 of which are employed, 16 for grounds and 16 for signals. In this way, by sharing the same electronics, it has been possible to build and test a large number of arrays of various shape in the years, such as planar arrays [22], all having up to 32 channels. Employing the same system and providing the cables connected to the microphones with a couple of IDC male connectors, a 32 channels array has been built.

Figure 73: EM electronics for arrays up to 32 channels

This solution is quite practical also to record a set of additional references together with the array. The electronic of the EM in fact must be connected to the Eigenmike Microphone Interface Box (EMIB), which communicates with the computer through a FireWire bus. This bus, which provides also the power supply, supports the daisy chain connection, allowing to connect to the EMIB an additional sound card. The Core Audio driver provided by the iOS environment allows to create an *aggregate device*: the two audio cards are managed by the system as one interface, with input and output channels equal to the sum of input and output channels of the two interfaces. When an aggregate device is employed, the same clock must be shared between the two sound cards, to ensure a synchronous sampling.

A MetricHalo 2882, provided with eight microphonic input, has been used for the purpose: the EMIB has been set as master, the MatricHalo as slave and the clock has been shared via Word-Clock, which is supported by both interfaces. Forty channels have been recorded synchronously with an aggregate device and the software Plogue Bidule, the first 32 sampled by the EMIB, thus coming from the array, the last 8 sampled by the MetricHalo and used for additional microphones and accelerometers [23].

Regarding the panoramic video recording system, nowadays several low cost solutions are available on the market, such as Samsung Gear 360, Ricoh Theta V or Vuze XR. All these equipments are made of two hemispherical coincident lenses, which have the advantage of reducing the stitching errors and can be employed next to a microphone array (Figure 117). However, this solution entails the introduction of an offset between acoustic and visual recording centres, resulting in a mismatch between the two recordings. A correction can be applied to the image, but it is not trivial and often not qualitatively acceptable. Due to this limitation, these compact solutions cannot be employed and a system with distributed cameras must be designed. The chosen solution is made of a planar ring of eight GoPro Hero Session 4; a first prototype is shown in Figure 74.

Figure 74: GoPro Hero Session 4 (left) and the prototype of ring for 8 GoPro cameras (right)

These cameras can record full HD resolution video ($1920p \times 1080p$) at 60 fps with a Field of View (FoV) of $72.22°$ on the vertical axis and $122.64°$ on the horizontal axis. They can be mounted tilted by 90° to enhance the vertical coverage. Two small shadow circles are still present at the top and bottom of the stitching sphere, which result in two black bands of $(180 - 122.64)/2 = 28.68°$ in the equirectangular

projection. Such a mounting reduces the superimposition angle between two consecutive shootings, which is $(72.22 \cdot 8 - 360)/8 = 27.22°$, still enough for an acceptable stitching. The stitching operation can be done with third party software, i.e. Kolor Autopano Video in combination with Kolor Autopano Giga.

A cylindrical housing has been designed to keep the cameras in the correct position (Figure 75), maximizing the superimposition of the shootings to improve the stitching. Two sets of closing grid have been realized, the second one not provided with the holes for the cameras, so that the array can be used also without them.

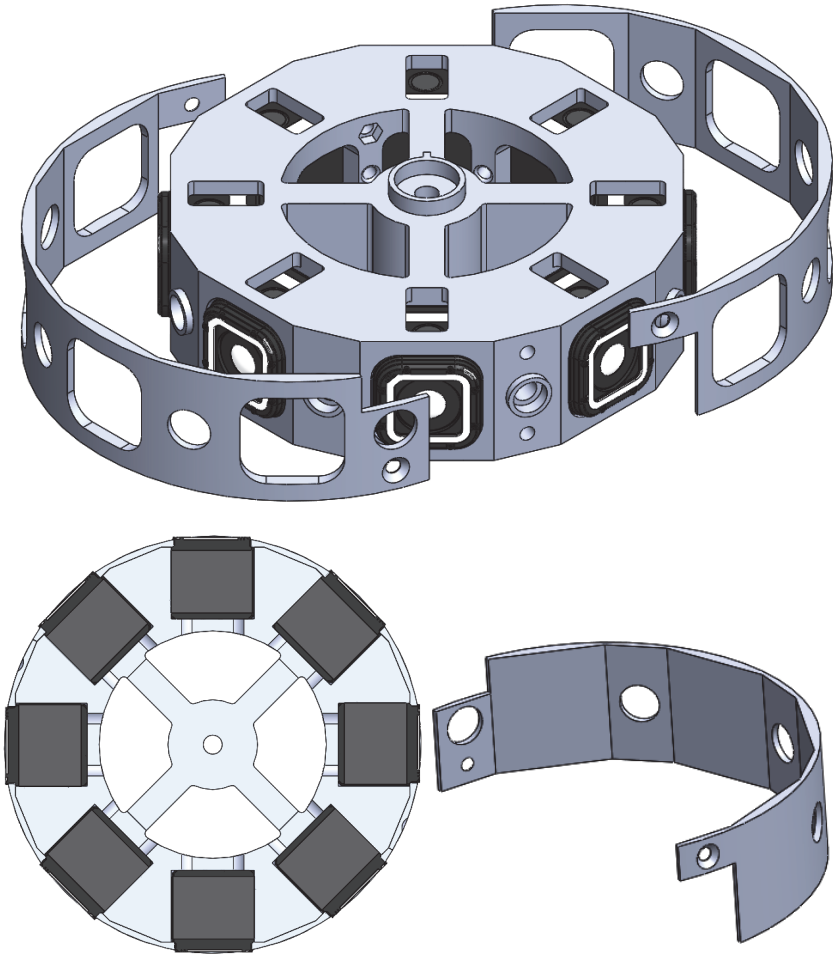


Figure 75: Ring of 8 GoPro, overall view (above), view from above (below, left) and the closing system for usage without cameras (below, right)

The size of the array must be designed accordingly to the frequency range of interest. For automotive NVH applications, it is necessary to analyse a frequency range from $50\ Hz$ to $2\ kHz$. According to the guidelines provided by Boaz Rafaely [24], it was decided to design a microphone array having a size similar to a human head (Figure 76, above), hence the name "Head-Shaped Array" (HSA). The cylindrical housing has been closed above and below with two hemispheres: the diameter of the array resulted to be $186mm$ with a total height of $235mm$. The lower part of the array is provided with a $3/8"$ thread to mount it on standard microphone stands (Figure 76,

below, left). A sort of “neck” of 100 mm diameter can also be attached to place the array in a dummy torso, a very common solution for automotive measurements (Figure 76, below, right).

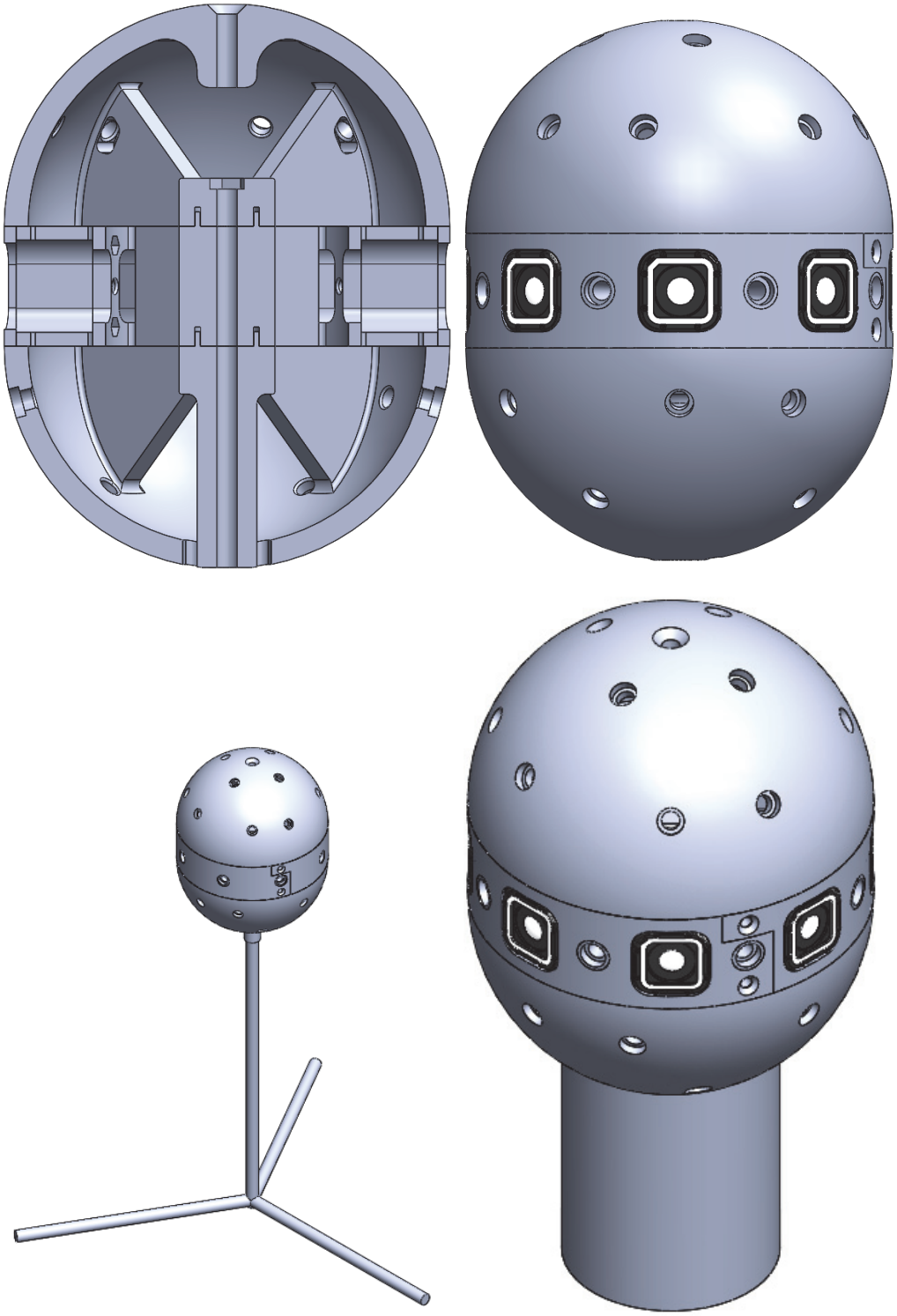


Figure 76: HSA, section view (above, left), front view (above, right), microphone stand mounting (below, left) and dummy torso mounting (below, right)

A $3/8"$ thread is present also on the top of the upper shell: it would be necessary to measure the array in an anechoic room with the two-axis turntable and it can also be used to mount an additional panoramic video recording system, such as a compact 360° camera.

The array is provided with 32 capsules, arranged as follow:

- 8 capsules are placed in a central, horizontal ring, each one between two cameras;
- Above and below the central ring there are other 2 rings, respectively of 8 and 4 capsules, for a total of five different values of elevation;
- Along the azimuthal plane, all capsules are equally spaced by an angle of $360/32 = 11.25°$, to maximize the horizontal spatial sampling;
- The rings above and below the middle plane are not symmetric. To avoid the shielding effect of the neck and dummy torso, the two rings of microphones in the lower shell have been moved upward. The elevation angles are as follow: first ring above, $+30°$, second ring above, $+65°$, first ring below, $-15°$, second ring below, $-45°$.

Capsules employed are Primo EM172, characterized by an omnidirectional response, a maximum SPL of $119dB$, $14dB(A)$ of electrical background noise and the frequency response of Figure 77.

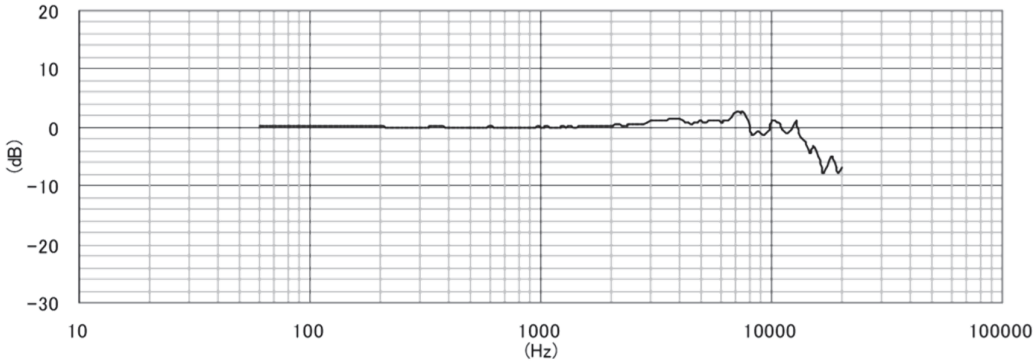


Figure 77: Frequency response of Primo EM172 capsule

All parts have been 3D-printed in ULTEM 9085 with an industrial 3D printer. Figure 78 shows the array, once assembled, mounted on a microphone stand (left) or on a dummy torso (right).

Figure 78: HSA, microphone stand mounting (left) and dummy torso mounting (right)

The matrices for A-2-B and A-2-P conversions have been calculated with a FEM simulation. The solid-mechanics coupling has been tested by comparing two coarse simulations in the range $20\ Hz - 3.5\ kHz$ with $df = fs/nfft = 48000/2048 = 23.4375\ Hz$.

Aluminium has been set as material for the capsules with COMSOL built-in characteristics, while ULTEM 9085 properties have been retrieved by manufacturer datasheet and are as follow: density $\rho = 1340\ [kg/m^3]$, Poisson's ratio $\nu = 0.33\ [adm]$ and Young's modulus $E = 2.2e9\ [Pa]$.

The simulations have been processed to get Ambisonics 3rd order filters by means of Kirkeby inversion and the model of Figure 79. Superimposition of the PSD of the filters obtained by the two simulations are showed in Figure 80, in blue and red respectively with and without multiphysics coupling between pressure-acoustics and solid-mechanics.

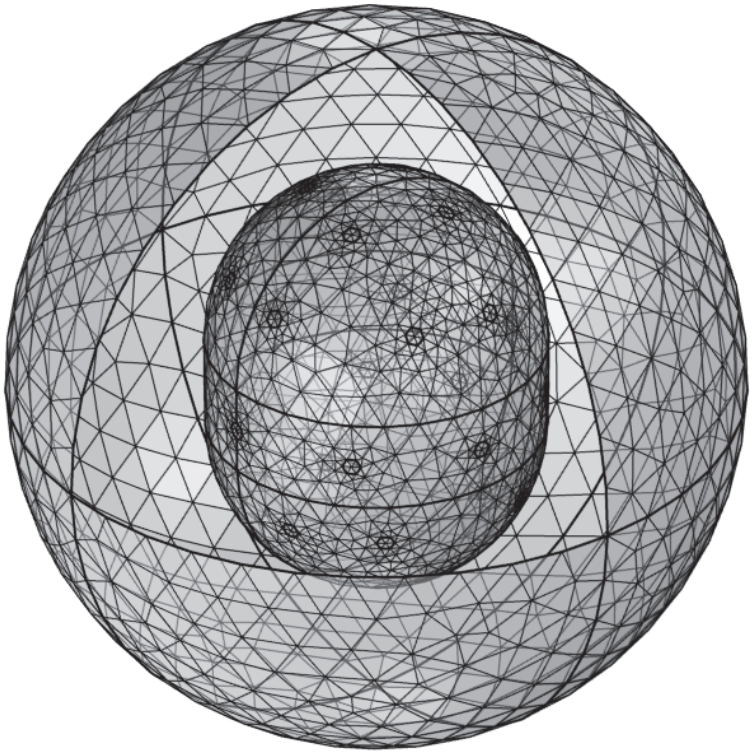


Figure 79: FEM model of the HSA

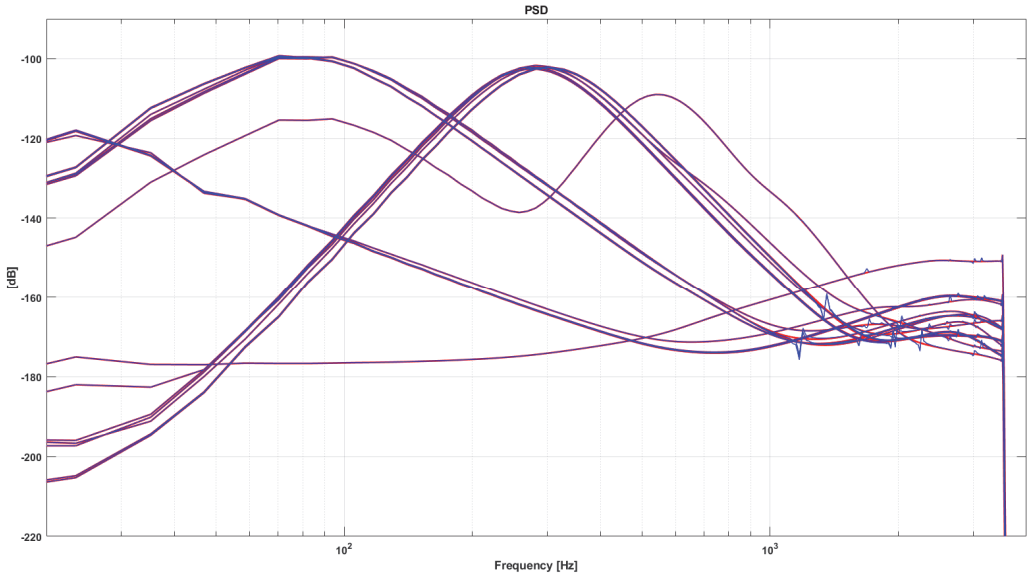


Figure 80: PSD of filters, multiphysics coupling evaluation: with solid-mechanics (blue) and without solid-mechanics (red)

Note that filters are almost identical a part minimum differences above 1 kHz. For this reason, the multiphysics coupling was not used.

2.3.1. Performance evaluation

A first evaluation of the performance of the array has been studied with a simplified geometry: a sphere of radius equal to the radius of the cameras housing, that is 93 mm, and placing the capsules in the same directions of the real array. In Table 8, directions of the capsules are reported in degrees for the first eight capsules, and then the scheme is repeated four times along the azimuth.

N°	Azimuth [°]	Elevation [°]
1	0.00	0
2	11.25	30
3	22.50	-15
4	33.75	65
5	45.00	0
6	56.25	-15
7	67.50	30
8	78.75	-45

Table 8: HSA, capsules directions

The numerical solution has been calculated with a T-21 grid and then the parameters for evaluating spatial performance have been produced with the inversion method 1 (paragraph 2.2.1), $WNG_{max} = 20\ dB$, Ambisonics up to order four and length of the filters 4096 $samples$. SC and LD are shown in Figure 81 for the simplified HSA and in Figure 82 for the EM. Numerical values of the frequency limits for each Ambisonics order, calculated with thresholds defined in (12) and (13), are summarized in Table 9.

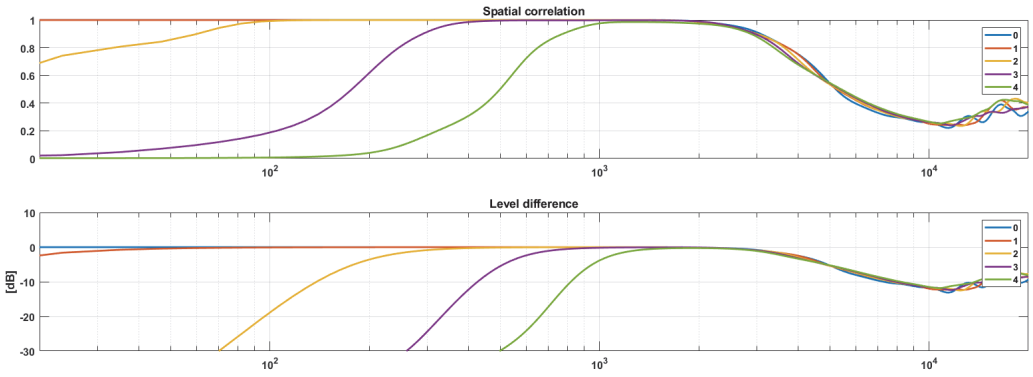


Figure 81: HSA, spatial performance of the spherical model, theoretical response

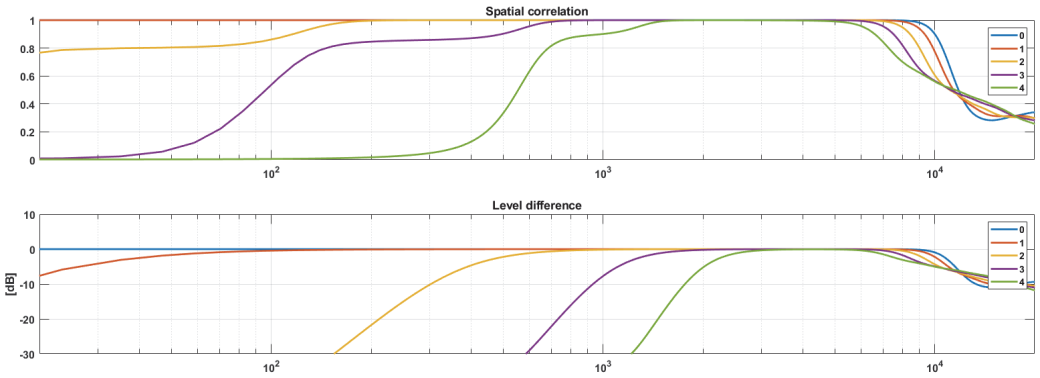


Figure 82: EM, spatial performance, theoretical response

	HSA (spherical model) Numerical solution		**EM Numerical solution**	
Ambisonics order	**Freq. start [Hz]**	**Freq. stop [kHz]**	**Freq. start [Hz]**	**Freq. stop [kHz]**
1	45	2.4	95	8.9
2	340	2.4	750	7.9
3	810	2.4	1730	7.0
4	1470	2.4	2850	6.0

Table 9: HSA (spherical model) and EM, spatial performance comparison

Note that the starting frequencies of all the Ambisonics orders for the simplified HSA are about half respect to the EM, validating the improvement of the new array in the low frequency range. The 2nd order Ambisonics is now available below $400Hz$, which is around the upper working limit of wideband ANC systems while the 3rd order Ambisonics is available under $800Hz$, which is around the upper working limit of narrowband ANC systems. In addition, this analysis provides an important information about the upper frequency to set in the FEM model for the simulation.

Here below it is showed also the result that would have been achieved with the radius of the new HSA simplified as a sphere ($93\ mm$) and keeping the directions of the capsules as in the EM (Figure 83). Table 10 shows the comparison between the two simplified geometries of the HSA. It is clear that capsules positioning has a great role in determining the spatial performances: the improvement obtained is quite relevant, particularly for the upper limit.

This result is of great importance: the new array is built mainly for low frequencies and even if the transducer placement is suboptimal the spatial performance in the range of interest are still fine. In any case, it was not possible to place capsules in the same directions as the EM due to mechanical constraints (a part from the fact that the real array is not spherical at all).

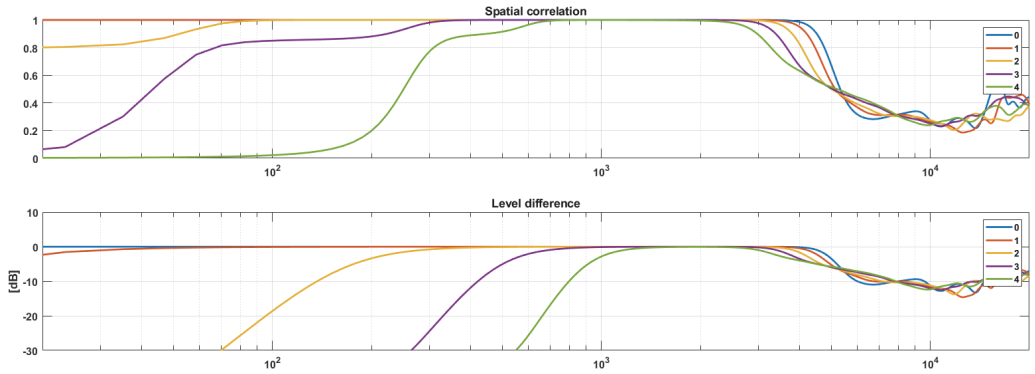


Figure 83: HSA, spatial performance of a model with EM capsule directions, theoretical response

	HSA (spherical model) Numerical solution		HSA (spherical model, EM capsule directions) Numerical solution	
Ambisonics order	**Freq. start [Hz]**	**Freq. stop [kHz]**	**Freq. start [Hz]**	**Freq. stop [kHz]**
1	45	2.4	45	4.0
2	340	2.4	340	3.6
3	810	2.4	780	3.2
4	1470	2.4	1290	2.7

Table 10: HSA, spatial performance comparison between spherical model and spherical model with EM capsule directions

The simulation of the real array has been performed with the model of Figure 79, resolution $df = fs/nfft = 48000/4096 \cong 11.72\,Hz$, frequency range $20\,Hz - 3.5\,kHz$ and without solid-mechanics coupling. Spatial performance of the filtering matrix for Ambisonics up to order four is shown in Figure 84, while in Table 11 a comparison of the numerical results obtained in the three cases is presented.

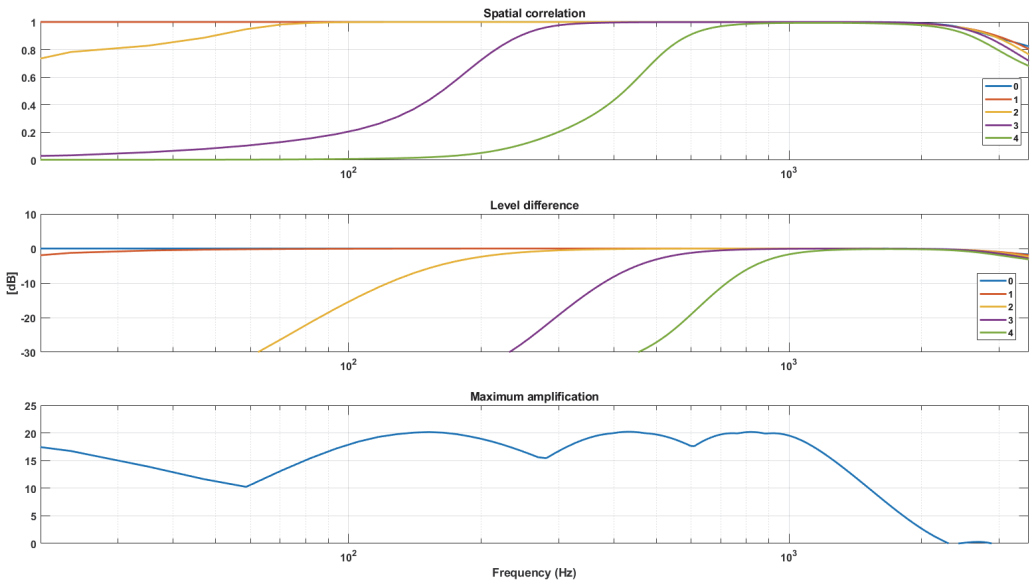


Figure 84: HSA, spatial performance evaluation, simulated response

	HSA Simulation		HSA (spherical model) Numerical solution		HSA (spherical model, EM capsule directions) Numerical solution	
Ambisonics order	**Freq. start [Hz]**	**Freq. stop [kHz]**	**Freq. start [Hz]**	**Freq. stop [kHz]**	**Freq. start [Hz]**	**Freq. stop [kHz]**
1	40	2.5	45	2.4	45	4.0
2	300	2.5	340	2.4	340	3.6
3	710	2.5	810	2.4	780	3.2
4	1220	2.3	1470	2.4	1290	2.7

Table 11: HSA, spatial performance comparison between simulated array and numerical solution of simplified models

The result of the simulation has been processed also with the Kirkeby inversion method (paragraph 2.2.1), which should provide the best performance; it has been used a threshold $WNG_{max} = 20\,dB$, Ambisonics order 4th, length of the filters 4096 $samples$ and a proper optimization of the regularization parameter β (Figure 85). Evaluation of the spatial performance of the filters is presented in Figure 86. In Table 12, frequency limits for each Ambisonics order are reported in comparison with the limits identified with inversion method 1: it is possible to note a consistent improvement.

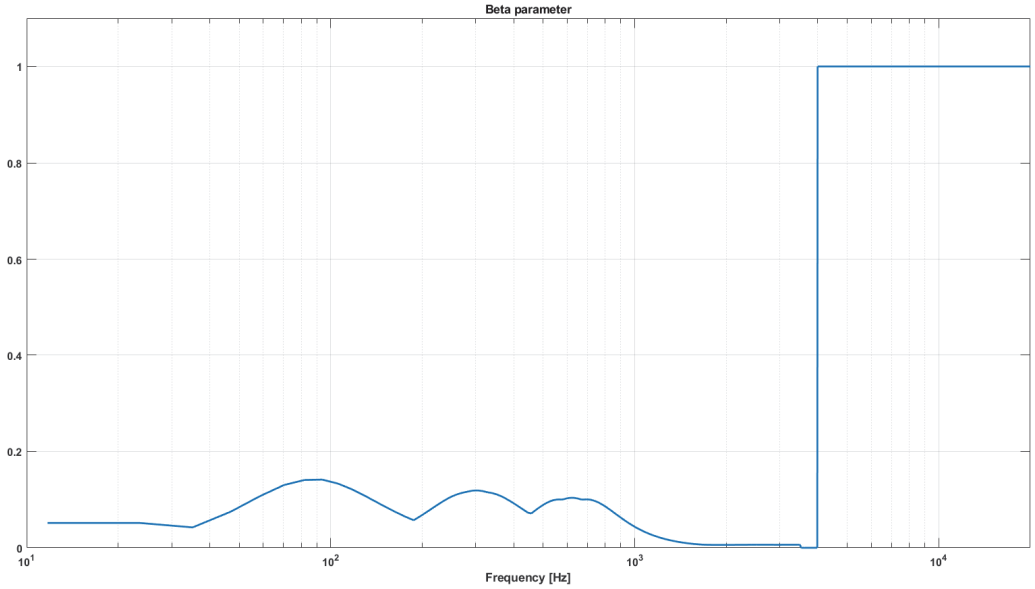


Figure 85: HSA, optimized regularization parameter $\beta(k)$

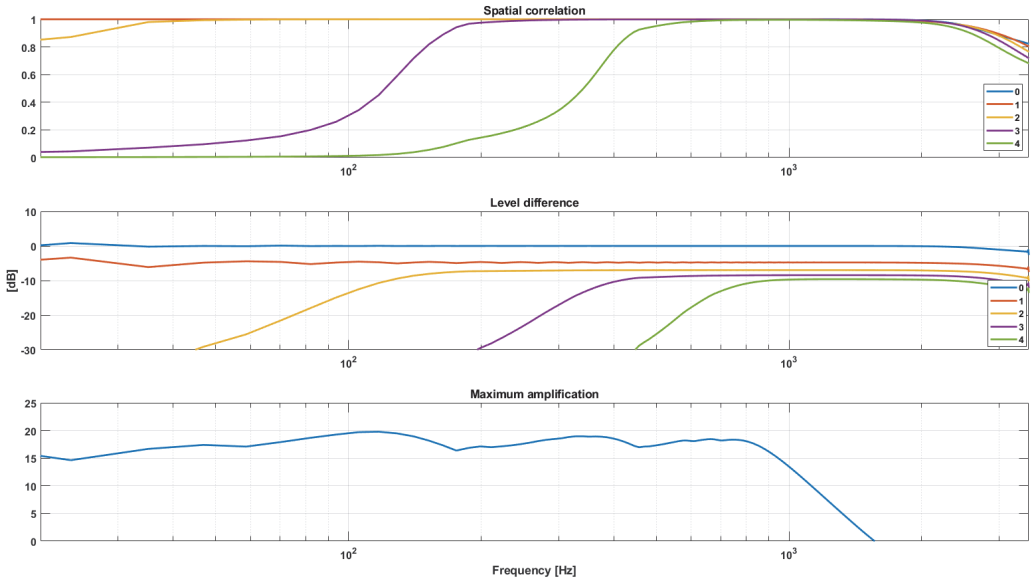


Figure 86: HSA, spatial performance evaluation, simulated response and Kirkeby inversion

	HSA Simulation Inversion method 1		HSA Simulation Inversion method 3	
Ambisonics order	**Freq. start [Hz]**	**Freq. stop [Hz]**	**Freq. start [Hz]**	**Freq. stop [kHz]**
1	40	2.5	20	2.5
2	300	2.5	170	2.5
3	710	2.5	510	2.5
4	1220	2.3	880	2.3

Table 12: HSA, spatial performance comparison between inversion methods 1 and 3

In conclusion, the spatial performances have been studied also for the SPS format, in comparison with the EM. In both cases, simulations have been processed with Kirkeby inversion, $WNG_{max} = 20\,dB$, length of the filters $4096\,samples$ and a proper optimization of the regularization parameter β. Two different targets have been imposed to the inversion: the first is a set of 32 virtual microphones having the directivity of super-cardioid of 16th order and aiming in the directions of the nearly-uniform grid with 32 points, the second in a set of 32 virtual microphones aiming in the direction of the capsules. Therefore, one target is the same for both arrays. Figure 87 and Figure 88 show the spatial performances of SPS format for the two arrays with the nearly-uniform grid. In Figure 89 and Figure 90, it is presented the directivity in function of the frequency for this case.

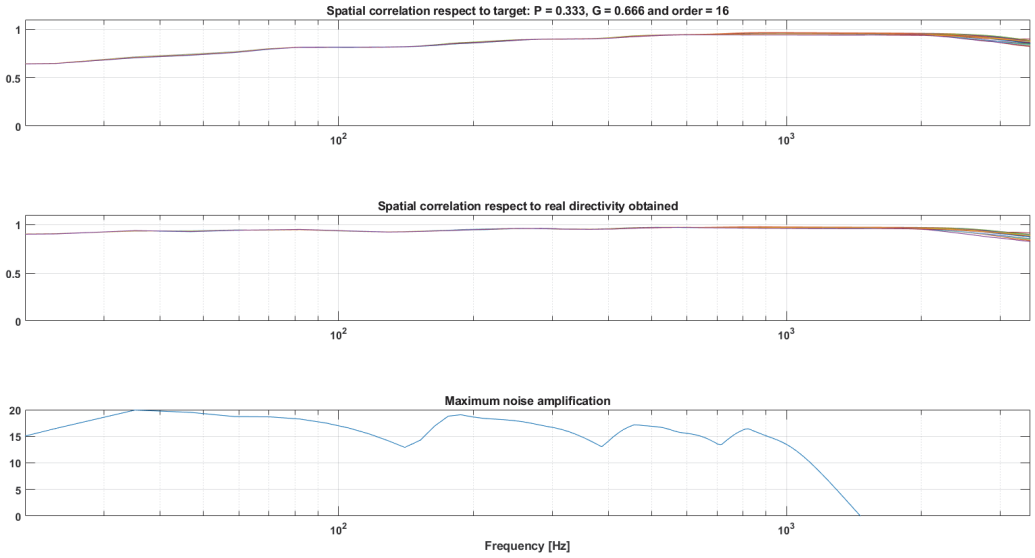


Figure 87: HSA, spatial performance of SPS format, nearly-uniform grid target

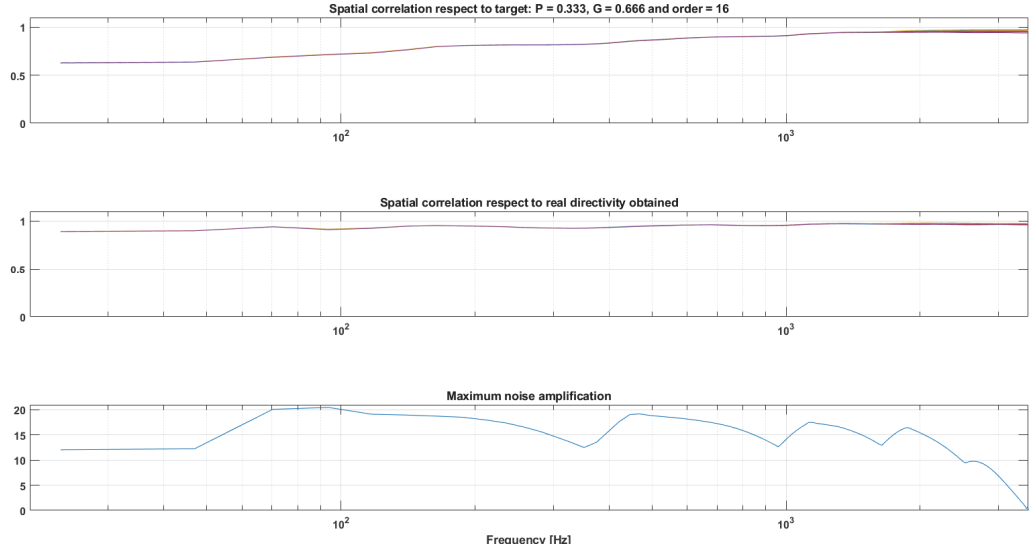


Figure 88: EM, spatial performance of SPS format, nearly-uniform grid target

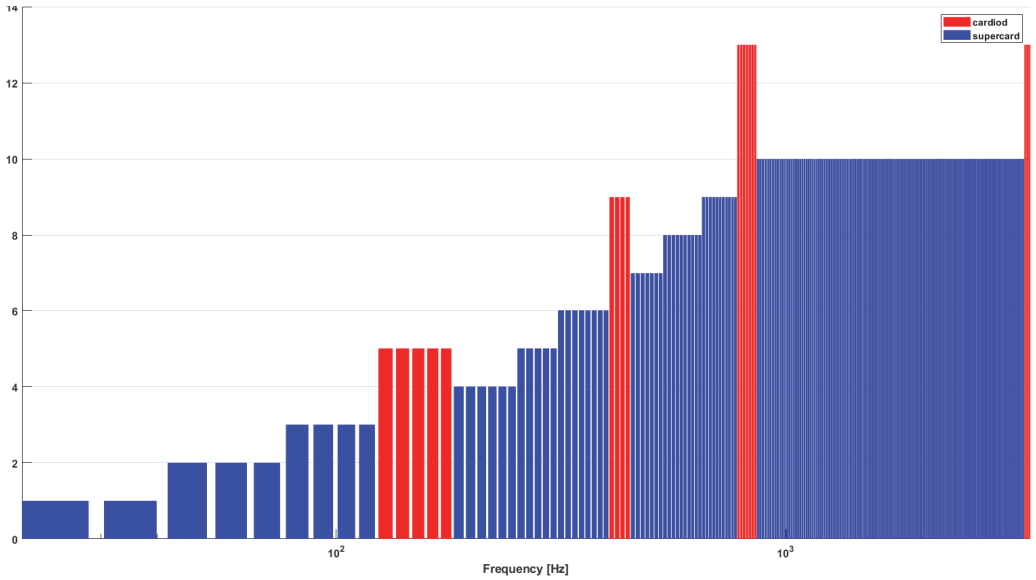


Figure 89: HSA, SPS directivity in function of frequency, nearly-uniform grid target

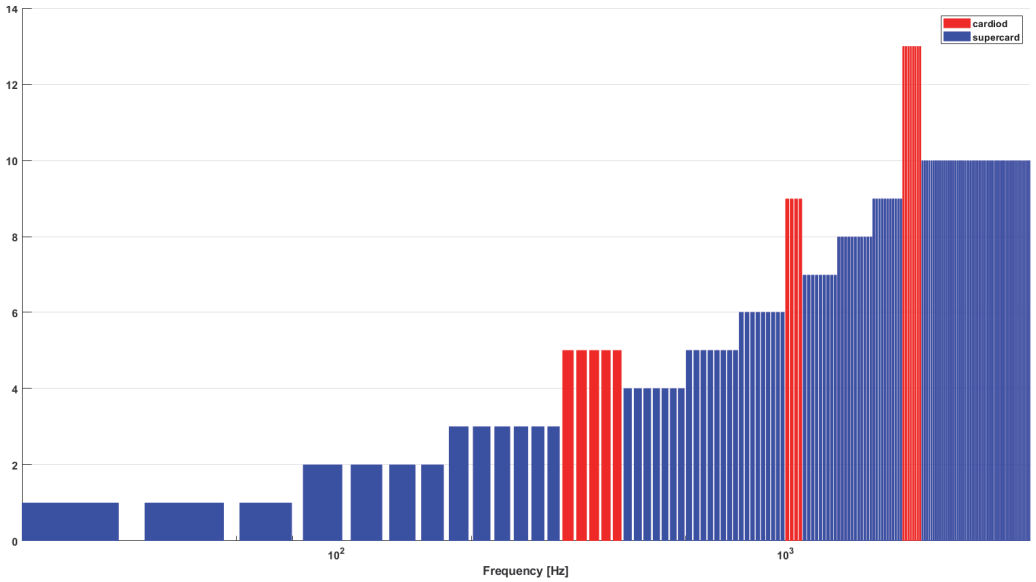


Figure 90: EM, SPS directivity in function of frequency, nearly-uniform grid target

Figure 91 and Figure 92 show the spatial performances of the SPS format for the two arrays with the virtual microphone directions pointing in the same direction of the capsules of each array. In Figure 93 and Figure 94, it is presented the directivity in function of the frequency for this case.

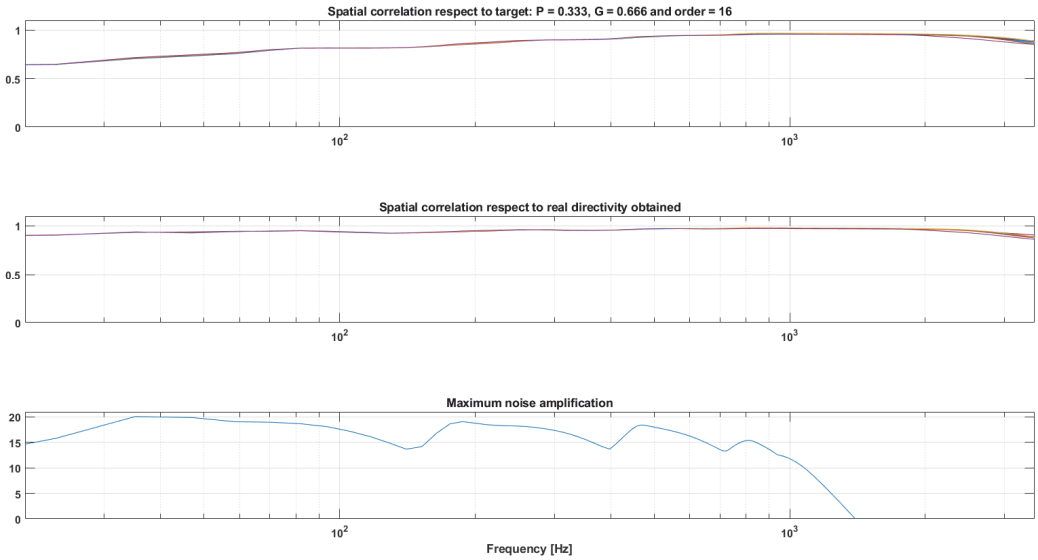


Figure 91: HSA, spatial performance of SPS format, virtual microphones aiming in the direction of the capsules

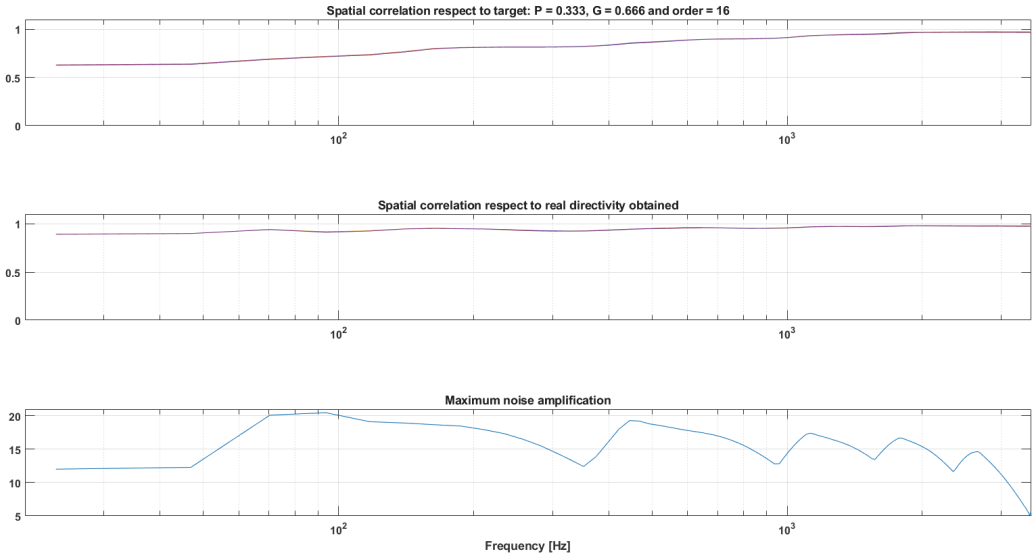


Figure 92: EM, spatial performance of SPS format, virtual microphones aiming in the direction of the capsules

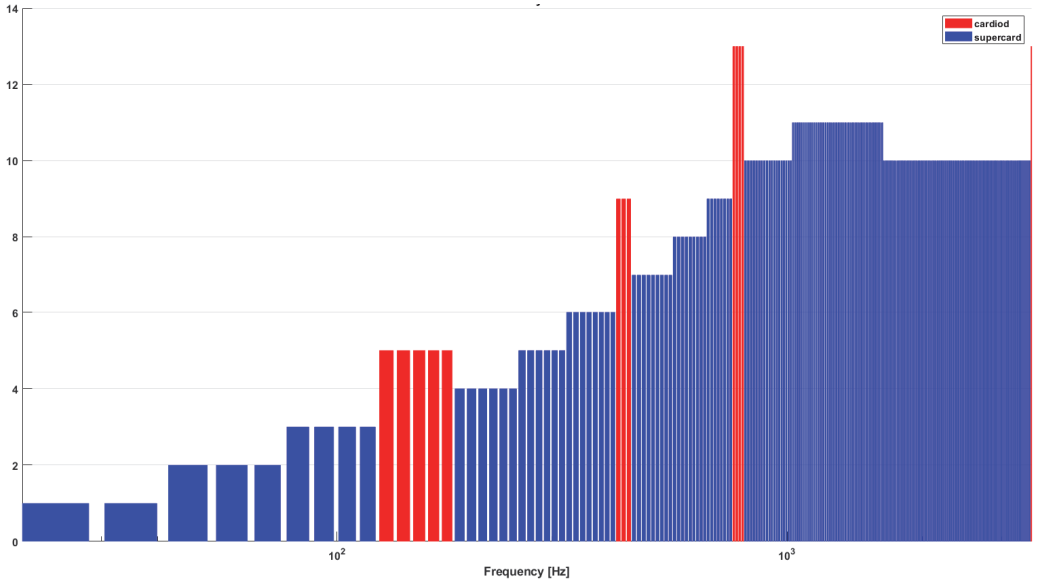


Figure 93: HSA, SPS directivity in function of frequency, virtual microphones aiming in the direction of the capsules

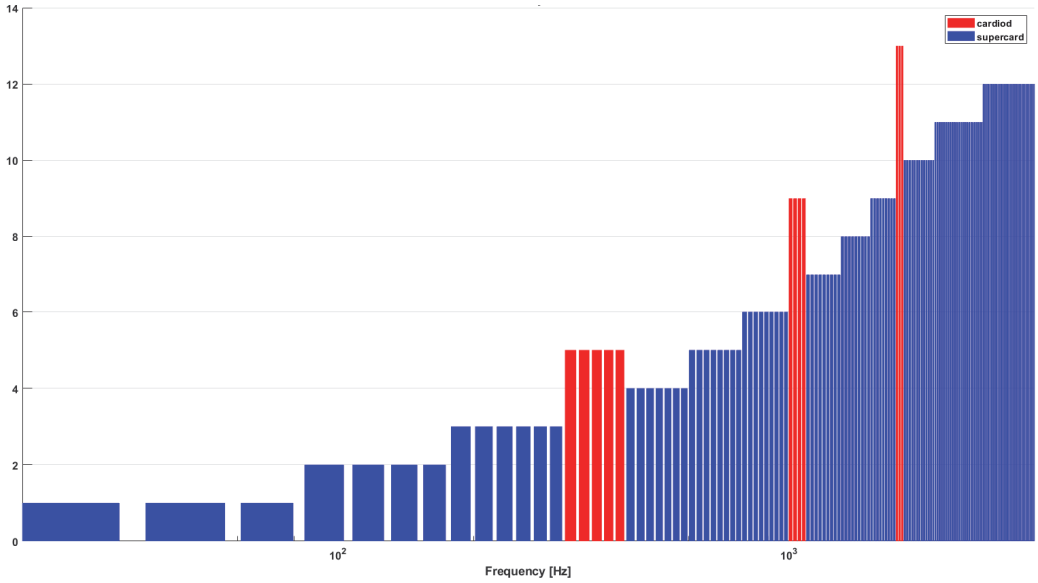


Figure 94: EM, SPS directivity in function of frequency, virtual microphones aiming in the direction of the capsules

By comparing Figure 89 with Figure 90 and Figure 93 with Figure 94, it is possible to note that spatial performance of the HSA are much better in the low frequency range respect to the EM. Contrariwise, by comparing Figure 89 with Figure 93 and Figure 90 with Figure 94 it is possible to note that a better directivity of the virtual microphones is achieved at high frequency if the directions of the virtual microphones are coincident with the direction of the capsules. In this case, for example, first differences are found for the HSA above 300 Hz but a consistent improvement of the directivity is relevant only above 1 kHz. For the EM the improvement is identified above 1.5 kHz.

2.4. Design of a hydrophone array

The new hydrophone array has been designed in order to meet the following requirements:

- Capability of recording Ambisonics first order;
- Integrating a high quality video recording system to produce 360° panoramic video for VR reproduction with HMDs;
- Employing one system to record audio and video together;
- Relatively small size, so that it can be handled easily and safely underwater, even by a single scuba diver;
- Possibility to mount it on a microphone stands or on a handle.

Ambisonics underwater systems already exist, and University of Parma itself developed its own one in the past [25]. Then, it is not that hard to add an underwater panoramic video recording system: 360° camera provided of a specific underwater case are available on the market. Audio and video recordings are not synchronized but it is possible to re-align them at sample, in post-processing. However, this solution involves the usage of two systems, increasing the possibility of failure of the recording session, which is unfortunately not so infrequent when operating underwater.

The main innovation introduced consists in the development of an integrated system for underwater audio-video recordings, capable of recording at the same time the panoramic video and the spatial audio without doubling the systems. To accomplish this task, a Ricoh Theta V (Figure 95, left) has been considered the best choice, for several reasons. It is a panoramic camera capable of recording high quality video with 4k resolution; an underwater case resistant up to 30 m is already available on the market (Figure 95, middle); it is capable of recording also the first order Ambisonics spatial audio in air, thanks to an additional external microphone, called TA-1 (Figure 95, right).

Figure 95: Ricoh Theta V (left), underwater case (middle), external microphone TA-1 (right)

The external microphone TA-1 is connected to the camera with a special 3.5 mm jack provided with six electrical contacts as showed below (Figure 96, left), where:

- GND: ground
- RBU: right back up
- LBD: left back down
- RFD: right front down
- LFU: left front up

The microphone positioning RBU, LBD, RFD and LFU is a standard for the first order Ambisonics arrays. The external ground ring permits to the camera to recognize the presence of the external microphone and activate it, providing the supply.

The connector has been dismounted from the TA-1 and four couples of wires have been soldered with RCA connectors as terminals (Figure 96, right), in order to make the plug-in operation as easy and fast as possible.

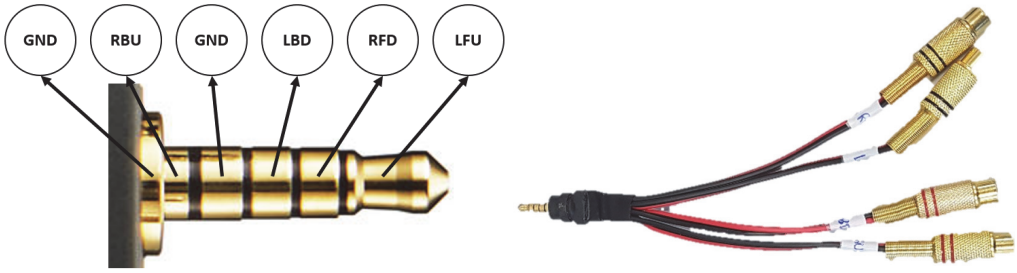


Figure 96: TA-1 connector, electrical contact (left) and rewiring (right)

Four Aquarian Audio hydrophones type H1c (Figure 97) have been employed: omnidirectional response, frequency range $20\,Hz - 4\,kHz$, phantom powered, operational depth up to 80 m, mounting thread $NPT\ \ 1/4"$. The hydrophone cable has been terminated with a male RCA connector.

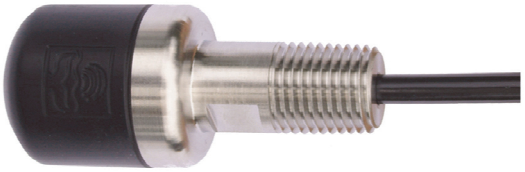


Figure 97: Aquarian Audio H1c hydrophone

An aluminium case has been designed and built and the four hydrophones have been mounted on it (Figure 98, left). The underwater case of the camera has been modified and mounted in an aluminium plug provided with a double O-ring (Figure 98, middle), designed for a static outer pressure of 100 bar, more than ten times the maximum pressure the hydrophones can tolerate. The bottom part of the plug is machined with the profile of the Ricoh Theta V, so that it can be easily mounted and unmounted, and a finger with two screws keep the camera in the correct position when mounted (Figure 98, right).

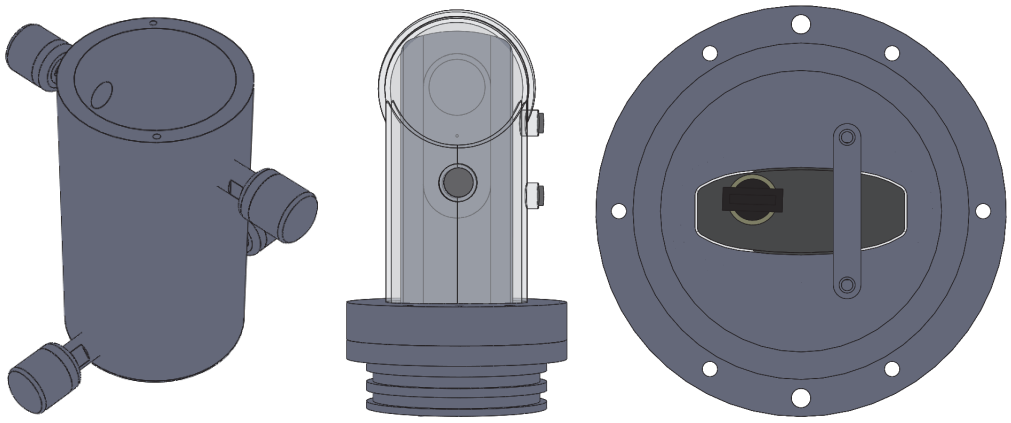


Figure 98: Underwater array design

The inner diameter is $73\ mm$, chosen so that a ZOOM H2 recorder (Figure 99, left) can be placed inside it and connected to the hydrophones. For this alternative assembly, a different plug has been designed and built, without any underwater recording system but provided with a standard thread $UNC\ 1/4"$, so that an external underwater case can be mounted on it (Figure 99, right). When the system is mounted with the ZOOM H2, the capability of recording directly a panoramic video with spatial audio is lost, but it is gained the possibility to record long sessions of underwater noise, up to $24\ h$ at $80\ m$ depth (the limit is now given by the hydrophones and not by the underwater case, which is not present). These two characteristics together make the system ideal for monitoring the environmental impact of underwater noise generated by human activities [26], [27].

Figure 99: ZOOM H2 (left) and the system assembled for underwater noise monitoring (right)

In the following images, the array is shown mounted on a tripod in the two possible configurations: with integrated panoramic video recording system (Figure 100, left) or with the ZOOM H2 recorder (Figure 100, right).

Figure 100: The new underwater array, with panoramic video recording system (left) and with ZOOM H2 recorder (right)

When a recording is taken with the Ricoh Theta V, a digital filtering is automatically applied by the camera: this is the A-2-B format conversion, which is valid for the original TA-1 microphone operating in air and it is completely wrong for the hydrophones. The problem has been solved again with the inversion method developed at University of Parma: the filtering matrix applied by the camera (Figure 101) has been recorded and inverted with the Kirkeby formula, giving the following target:

$$\|A\|_{4\times4} = \left\|\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}\right\|, \tag{16}$$

In this way, the inverse filter (Figure 102) is obtained. The convolution between the filtering matrix and the inverse filter gives back the target, which is a diagonal matrix of Dirac delta functions (Figure 103).

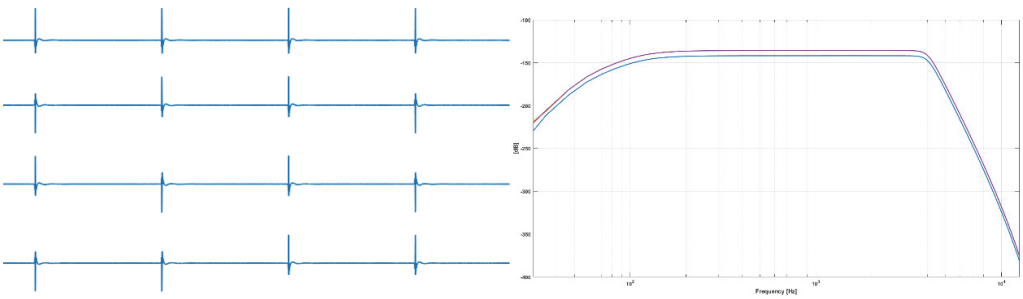


Figure 101: TA-1, default filter matrix, time domain (left) and frequency domain (right)

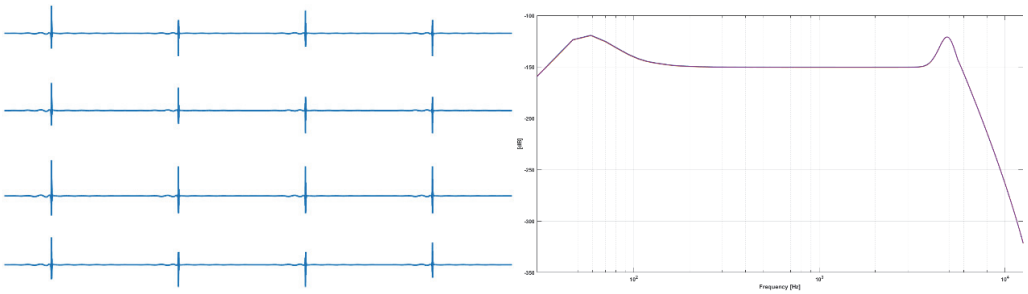


Figure 102: TA-1, inverse filter matrix, time domain (left) and frequency domain (right)

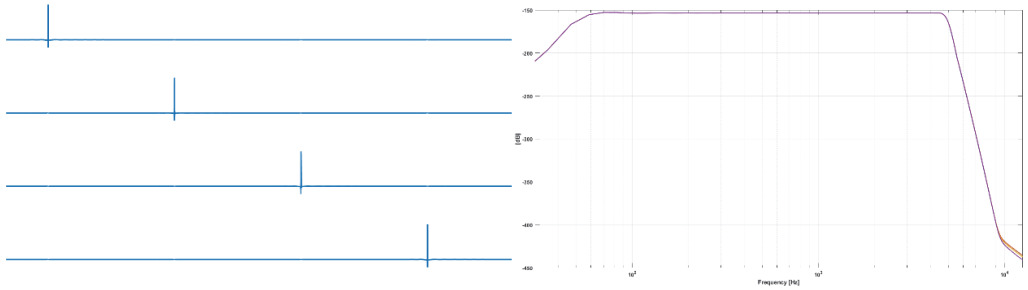


Figure 103: TA-1, Dirac delta diagonal matrix, time domain (left) and frequency domain (right)

Before computing the simulation to get the correct A-2-B format matrix, the influence of the solid-mechanics coupling has been studied. Three simulations have been solved with the model of Figure 104, $df = fs/nfft = 48000/2048 = 23.4375\ Hz$ and respectively considering the solid-mechanics coupling with damping (red), solid-mechanics coupling without damping (blue) and without solid-mechanics coupling (black).

Materials have been set using built-in properties, as follow: aluminium for cylinder body, top cap and hydrophone stems, Poly Methyl MethAcrylate (PMMA) for the underwater case and water for the domain. The rubber of the hydrophone heads, which is the part sensitive to pressure, has been manually defined: density $\rho = 1500\ kg/m^3$, Poisson's ratio $\nu = 0.4$ and Young's modulus $E = 6e8\ Pa$. The damping has been defined for the rubber and for the PMMA as an *isotropic loss factor*: $\eta_{PMMA} = 0.05$ and $\eta_{rubber} = 0.1$.

Filters for Ambisonics 1st order have been obtained with inversion method 1 (paragraph 2.2.1), $WNG_{max} = 20\ dB$ and length of the filters 4096 $samples$. The PSD of the filters are showed superimposed in Figure 105. Due to the presence of a resonance in the frequency range $400\ Hz - 600\ Hz$, the array response has been calculated with solid-mechanics coupling.

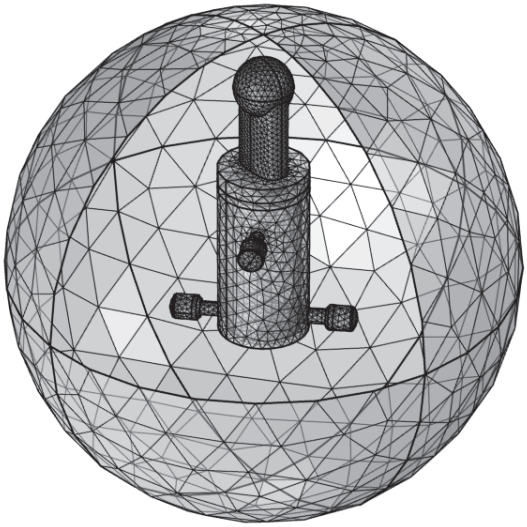


Figure 104: FEM model of the underwater panoramic audio-video recording system

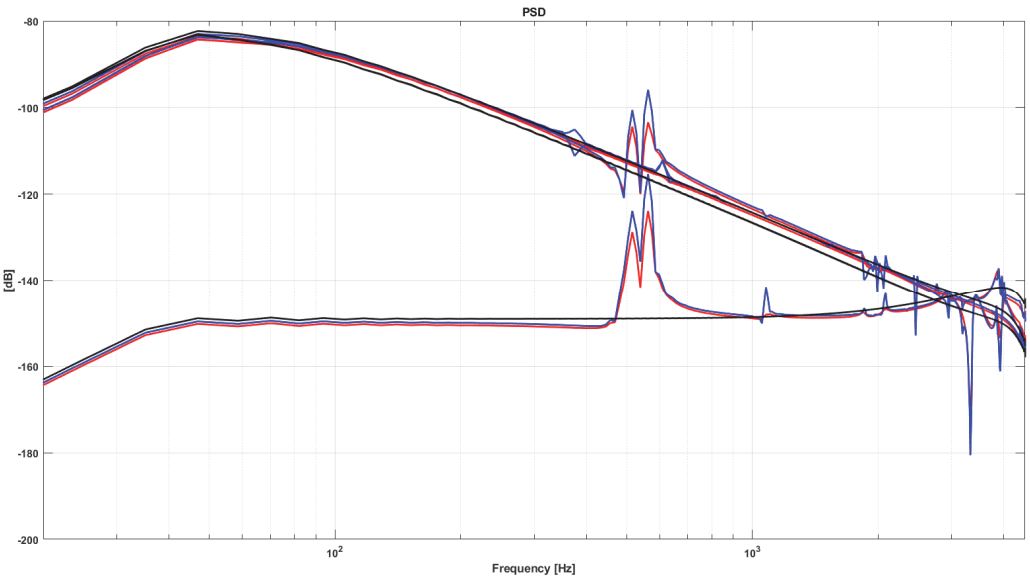


Figure 105: Underwater system, PSD comparison of the filters, damped solid-mechanics (red), undamped solid-mechanics (blue), without solid-mechanics (black)

2.4.1. Performance evaluation

In this case, two different simulations have been done to get the A-2-B format conversion matrix, one for each possible mounting of the probe. The model of the system with an integrated audio-video recording solution is the one of Figure 104, whilst in Figure 106 it is shown the model of the system for underwater noise monitoring, which do not provide an integrated panoramic video recording solution. Both simulations have been solved with $df = fs/nfft = 48000/4096 \cong 11.72\ Hz$ and, being a FOA probe, the number of testing directions has been limited to a T-

design 10 with 62 directions (north and south poles added). Then filters for Ambisonics 1st order have been obtained with Kirkeby inversion (paragraph 2.2.1), $WNG_{max} = 20\ dB$, length of the filters 4096 $samples$ and a proper optimization of regularization parameter β.

Spatial performances (Figure 107) resulted to be identical for the two filters, as the underwater case is transparent to the sound propagation, being the wavelength $\lambda_{min} = c/f_{max} = 1520/4500 \cong 0.34\ m$. Frequency limits for the Ambisonics first order are summarized in Table 13. Figure 108 shows the comparison of the directivity patterns of the four virtual microphones at the central frequencies of two octave bands, 1 kHz and 4 kHz. It is possible to note clearly the distortion of the lobes in the latter, which in fact is out of the usable range.

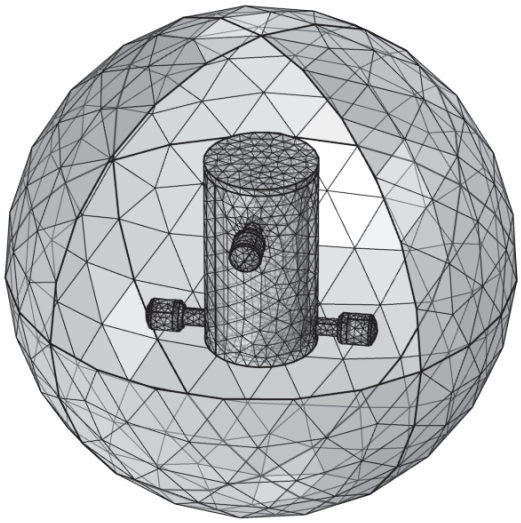


Figure 106: FEM model of the underwater noise monitoring system

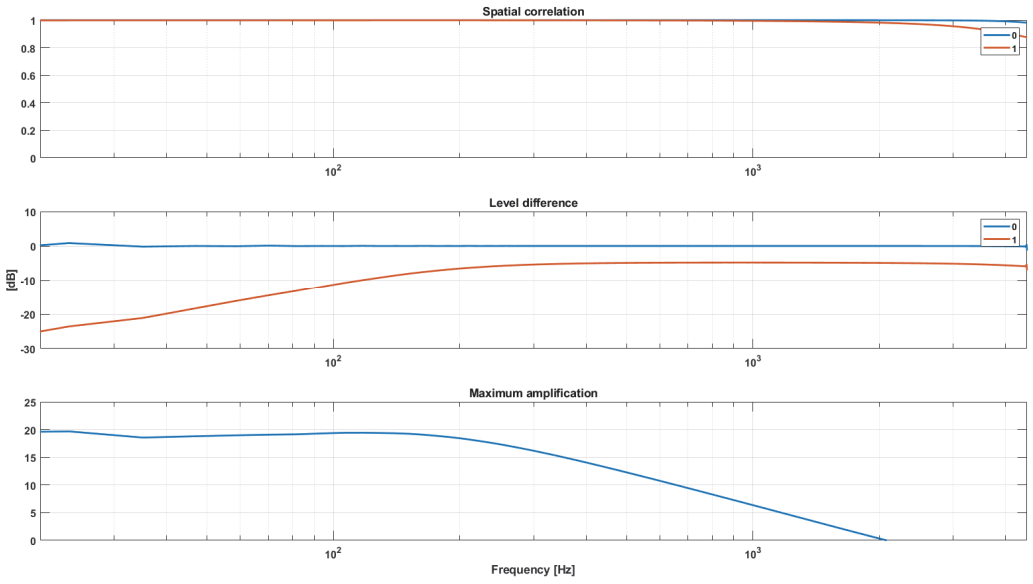


Figure 107: Underwater system, spatial performance evaluation

	Underwater audio-video recording system	
Ambisonics order	**Freq. start [Hz]**	**Freq. stop [Hz]**
1	290	3.1

Table 13: New underwater system, frequency range for Ambisonics 1st order

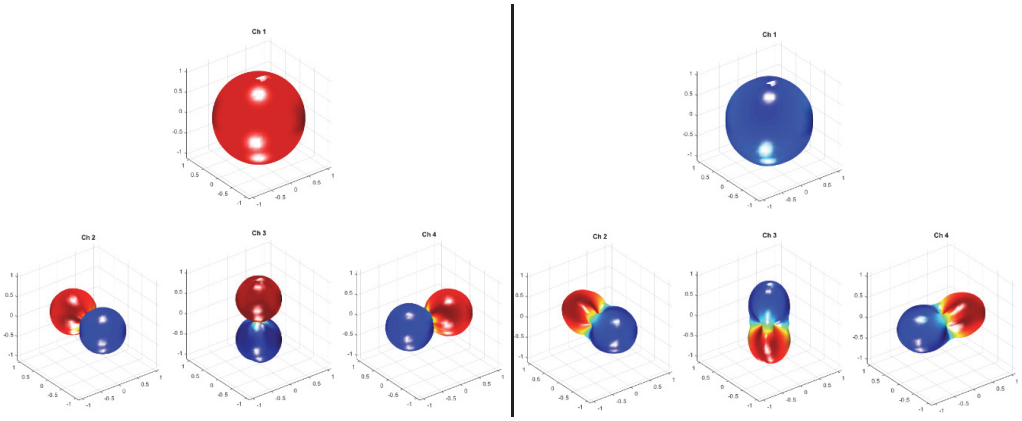


Figure 108: Underwater system, 1st order virtual microphone directivity patterns at 1 kHz (left) and 4 kHz (right)

2.4.2. Comparison with existing probe

As mentioned before, another hydrophones array for underwater noise monitoring was already existing [25], capable of recording long measurement sessions but not integrating any panoramic video system. The main disadvantage of the first prototype consisted in the presence of four cables, each $20\ m$ long, connecting the hydrophones to the recorder. They have been removed in the new probe, as the recorder and the cables are inside the cylindrical body.

Similarly to the new probe, a simulation of the old prototype has been performed (Figure 109) in the frequency range $20\ Hz - 4.5\ kHz$, without solid-mechanics coupling. The result has been processed with Kirkeby inversion method and optimization of the regularization parameter, to get a filtering matrix for Ambisonics first order. Spatial performances have been evaluated (Figure 110) to define the working range of the beamforming (Table 14).

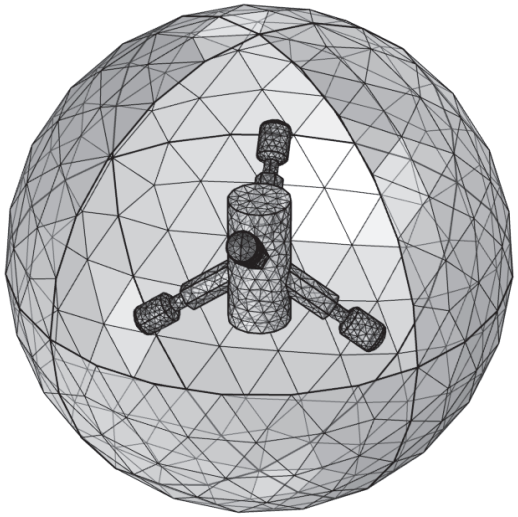


Figure 109: First prototype of the underwater system, meshed model

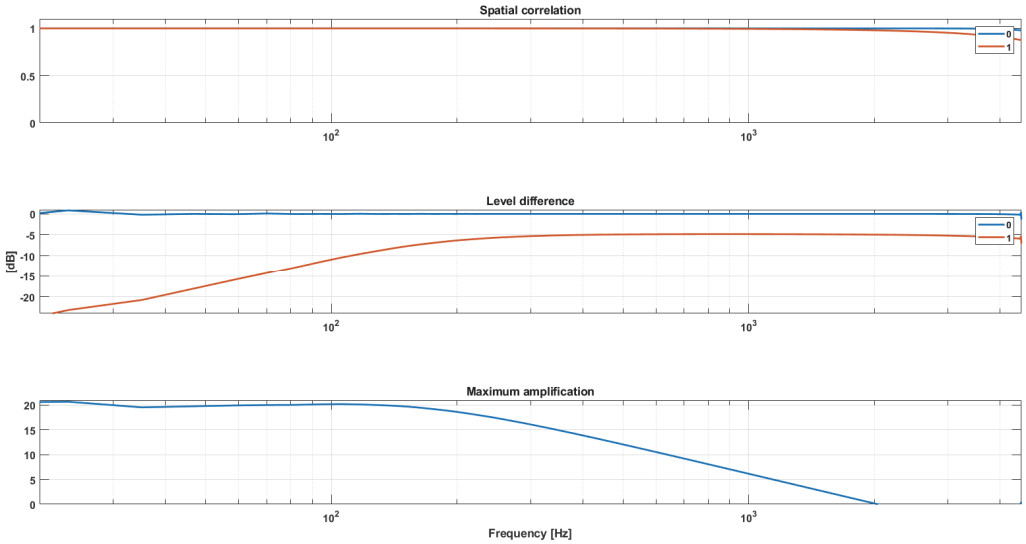


Figure 110: First prototype of the underwater system, spatial performances evaluation

	Underwater audio-video recording system	
Ambisonics order	**Freq. start [Hz]**	**Freq. stop [Hz]**
1	290	3.1

Table 14: First prototype of underwater system, frequency range for Ambisonics 1st order

The result is the same of the new probe, which means that more functionalities have been added, without losing the performance of the previous prototype.

3. Spatial information analysis

The spatial information recorded with arrays of microphones and hydrophones can be employed for enhanced sound field analysis through the Sound Colour Mapping technique [28]. The information of the spatial distribution of sound energy is graphically represented with colour maps superimposed on a background image, making it possible to identify noise sources, reflections, leakages, sound propagation paths, with a wide range of applications, such as [23], [29], [30] and [31].

Many solutions already exist for this task, each with its pros and cons. A couple of these software, the ones the author worked with mostly, are referenced in [32] and [33], with the second one well documented in [34].

The main advantage of these solutions is their capability of working real-time, an aspect of great importance because it allows to use the array as an acoustic camera. However, most of them seem focused on the identification of the DoA, neglecting an essential feature for noise analysis, that is showing the real values of calibrated SPL. By applying a normalization, the capability to detect the DoA of sounds seems to improve but the information related to the energy is completely lost. Hence, every contribution is considered equal, which is profoundly incorrect.

A proper visualization of the map is equally important. As an example, a map is shown in Figure 111: sounds are identified with spots of different dimensions in function of the amplitude and different colours in function of the frequency (discretization of frequency can be done for example with octave or third octave bands). In this way, all the information related to a sound source is provided in a single map: its position and the energy at various frequencies. Nevertheless, if more than one source or many reflections are present at the same time, the visualization becomes very confused. In addition, in case of noise recording this method would be useless because the information of the total energy is lost.

Moreover, none of the existing software is actually capable of producing a colour map video superimposed to a background image. Nor existing software is capable to support a background video: if boundary conditions of the sound field are not stationary during the recording, their effects on the colour map cannot be identified. This occurs, i.e., when recording inside a car cockpit while running on the road and another car passes in the opposite direction. It could be difficult to explain the spot that suddenly appears and disappears if a panoramic picture of the car is employed.

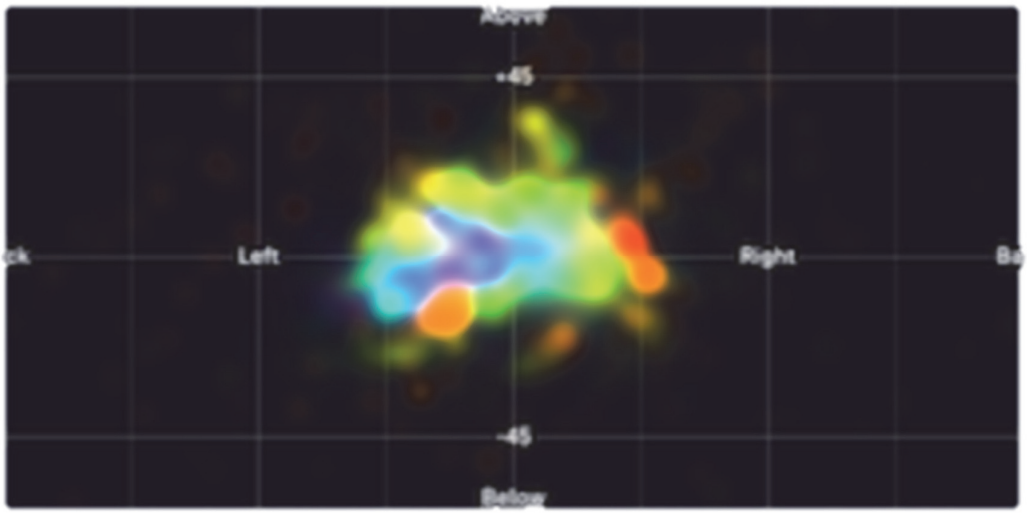


Figure 111: An example of sound colour mapping

3.1. Development of a sound colour mapping software

The new software, developed in Matlab, is based on a previous work described in [19], [35] and [36]. The architecture of the software is quite simple and can be divided in two main steps: the encoding and the analysis.

The encoding is performed by means of the convolution, calculated in frequency domain, between the multi-channel file recorded by the array (A-format) and the filtering matrix (see paragraph 2.2) for B-format or P-format conversion.

Then, the encoded format is processed, to get one of the three possible outputs implemented:

- Static colour map over a background image;
- Video colour map over a background image;
- Video colour map over a background video.

In case of Ambisonics processing, the encoding can be unnecessary. In fact, some arrays provide directly the B-format, without the need to perform the conversion. In those cases, two requirements must be respected: the recorded file format has to be a ".wav" and the Ambisonics format has to follow the AmbiX standard.

Background image or video can be taken with any panoramic recording system, such as the ring of GoPro (paragraph 2.3), the Ricoh Theta V (paragraph 2.4) or others (Samsung Gear 360, Vuze XR to cite some of the system the author worked with).

The visualization style chosen for the colour map is the same for Ambisonics and SPS implementation. It consists in associating the amount of energy to the colour, where red is the highest value and blue is the lowest. A threshold can be set to plot only values above a certain limit: values under the threshold result in a transparent portion of the colour map. Isolevel lines can be plotted, in case with the numeric value associated.

The colour map is superimposed on the background with selectable percentage of transparency and the background can be in colour or black and white. The aspect ratio of the map is always 2:1, with the azimuth in the range $-\pi \sim \pi$ and the elevation in the range $-\pi/2 \sim \pi/2$. If the panoramic background has a different aspect ratio, black bands are automatically added horizontally or vertically, depending if the aspect ratio is $\lessgtr$ 2:1.

The recording can be trimmed after being imported and an A-weight filtering can be applied, with respect of the current standard.

3.1.1. Ambisonics implementation

Several algorithms have been found in [11] for Ambisonics sound mapping and have been implemented:

- PWD, Plain Wave Decomposition;
- IV, Intensity Vector;
- MVDR, Multiple Variance Distortionless Response;
- MUSIC, MUltiple SIgnal Classification;

The processing is performed in frequency domain. The one-sided PSD of the B-format signal is obtained with the Welch's Overlapped Segment Averaging Spectral Estimation: each signal is divided into small chunks (the length should be multiple of 2), overlapped by a factor greater then 0.5 (typically 0.75), windowed (usually with a Hann window) and then the DTFT is computed by means of the FFT algorithm for each chunk. All chunks are then averaged. The PSD is one-side, so given a certain number $nfft$ of bins for the FFT, only the first $[(nfft/2) + 1]$ are taken, which is possible without loss of information, as the signals are reals.

After having calculated the PSD for all the signals of the B-format, it is possible to apply a filtering, which is easy and computationally light, as performed directly in the frequency domain. Therefore, the frequency limits for each Ambisonics order, as explained in 2.2.2.1., are applied. It is also possible to apply some other filtering to obtain a series of different analysis: unfiltered map (default), low-pass filtered map, band-pass filtered map and octave bands filtered maps can be calculated. Higher is the $nfft$ value employed for the DTFT, more precise will be the filtering.

Then, the covariance matrix is computed:

$$C_{B \times B} = cov(PSD_{\frac{nfft}{2}+1 \times B}), \tag{17}$$

where $PSD_{\frac{nfft}{2}+1 \times B}$ is the filtered Power Spectral Density calculated for a B-format signal having a number B of SH with $nfft$ bins of the DTFT.

The DoA estimation is calculated by each algorithm with the covariance matrix $C_{B \times B}$ at a dense grid of directions (see paragraphs 2.2.2.2 and 2.2.5.3 for details related to the grid of directions).

Before DoA matrix is plotted, the calibrated values of SPL are retrieved:

$$\|SPL\| = \frac{|||DoA|||}{\max(|||DoA|||)} \cdot rms, \tag{18}$$

with

$$rms = \frac{\sum_{f=1}^{nfft} p(f)}{2}, \tag{19}$$

and

$$p(f) = real(PSD(f))^2 + imag(PSD(f))^2, \tag{20}$$

Note that the last equation consists in the application of the Parseval's Theorem, which ensures the conservation of the energy.

Finally, the matrix of SPL values is converted in dB:

$$\|SPL[dB]\| = 10 \cdot \log_{10}\|SPL\|. \tag{21}$$

3.1.2. SPS implementation

The processing for SPS format is similar to Ambisonics. The PSD is calculated for all the signals of the P-format, than values of SPL are calculated as follow:

$$\|SPL_{P\times 1}[dB]\| = 10 \cdot \log_{10}\|rms_{P\times 1}\|, \tag{22}$$

where P is the number of virtual microphones of the SPS format and rms are the calibrated values of sound pressure in $[Pa]$, calculated with equations (17) and (18) for each of the P signals.

The P values of pressure are then interpolated over a dense grid of points rescaled over the background image. The grid over which values are interpolated depends on the directions of the virtual microphones and affects considerably the result. It can be better explained with the following example. A recording of $30\ s$ of pink noise has been taken with the EM and encoded into two SPS formats, the first one with 32 virtual microphones aiming in the directions of the capsules, the second one with 122 virtual microphones coincident with the directions of a nearly-uniform grid. It has been shown in 2.2.2.2 that the grid with virtual microphones aiming in the direction of the capsules provides better beamforming performances in function of the frequency. However, the map is distorted in the upper and lower region (Figure 115), because no virtual microphones are defined for the poles. Instead, the map should be "closed" (Figure 116), in the same way of the equirectangular image of the panoramic background, which is obtained by means of an equidistant cylindrical projection (also known as geographic projection), where each pole degenerates on a line. This is well shown by the classic example of the projection of the globe, with the Tissot's indicatrix of deformation (Figure 112).

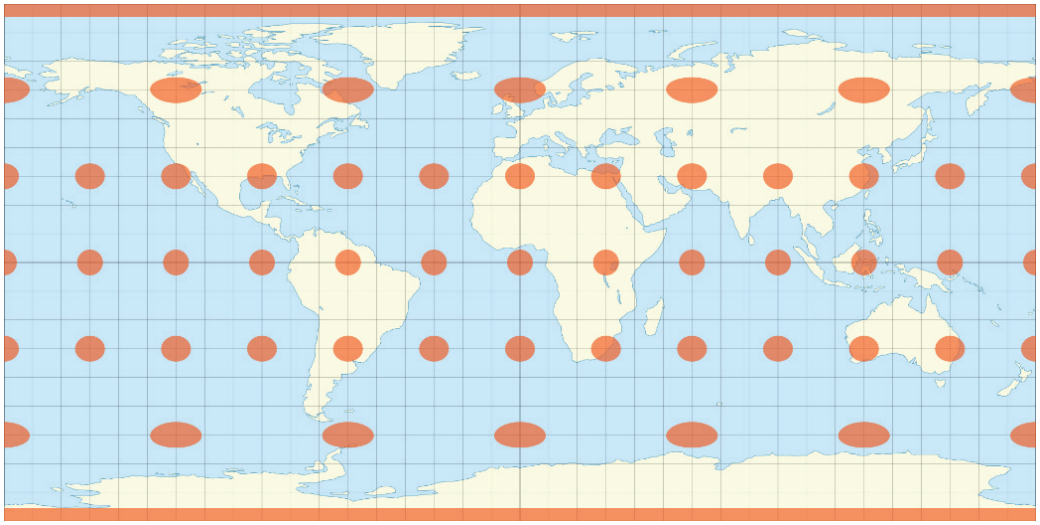


Figure 112: Equidistant cylindrical projection with Tissot's indicatrix of projection

Figure 113 and Figure 114 show the two grids of virtual microphones superimposed with the background picture. For the grid of 122 directions, one can note that virtual microphones pointing in the poles have been replicated along the horizontal plane, accordingly to Figure 112, therefore the total number of points becomes 140.

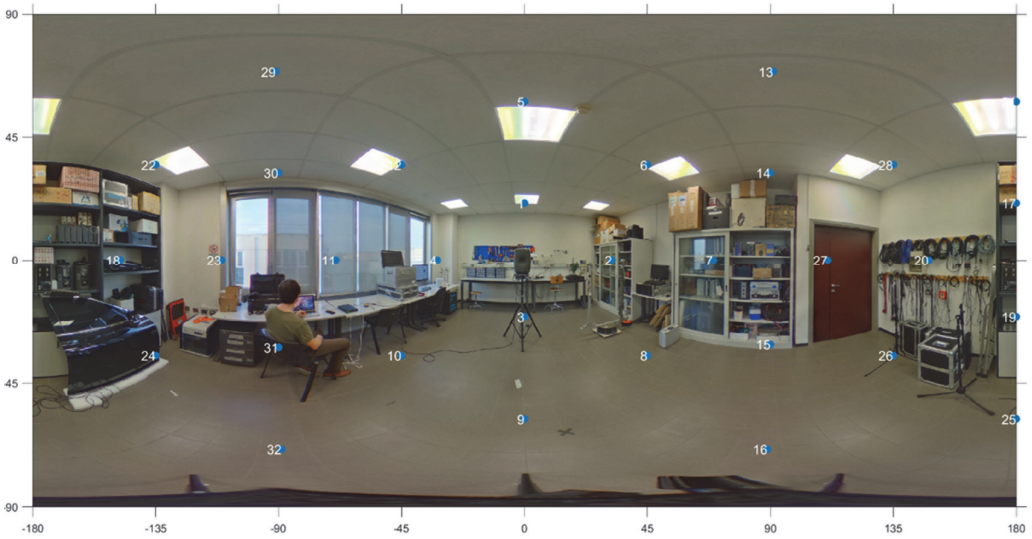


Figure 113: 32 points grid, directions coincident with EM capsules

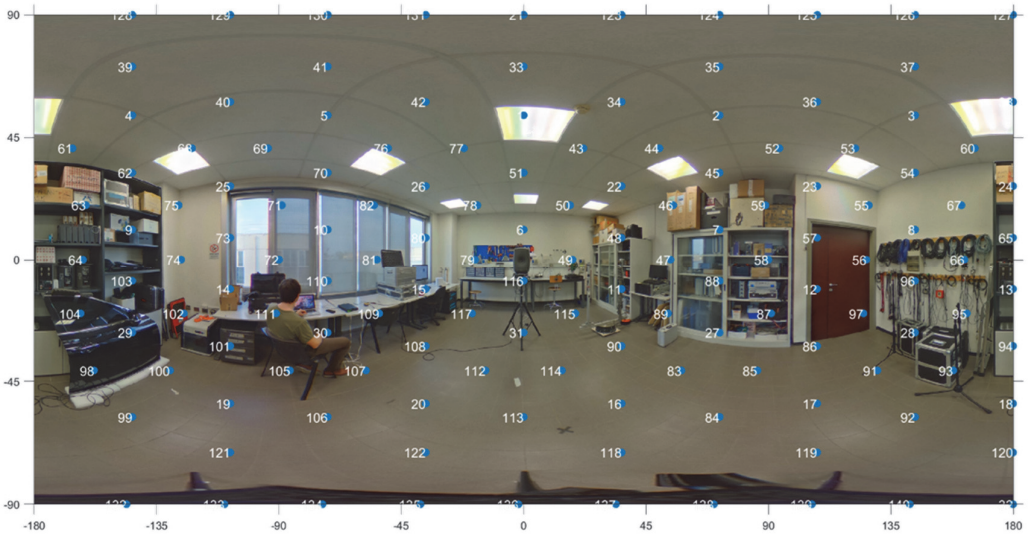


Figure 114: 140 points grid, directions coincident with nearly-uniform grid

In Figure 115 and Figure 116, the colour maps are showed superimposed to a black and white background image with a noise threshold of 15 dB, isolevel lines every 2 dB with associated numeric values and the colour scale legend.

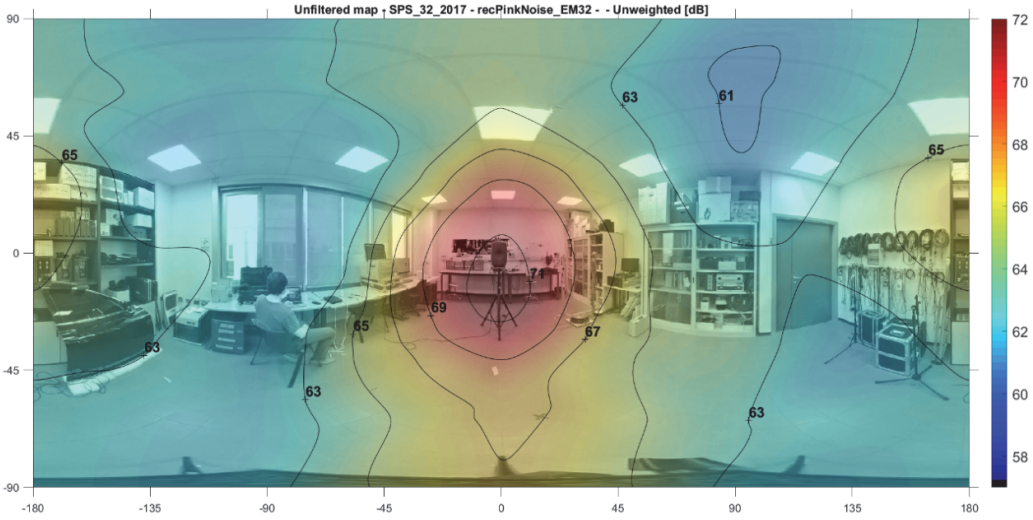


Figure 115: Colour map with 32 points grid

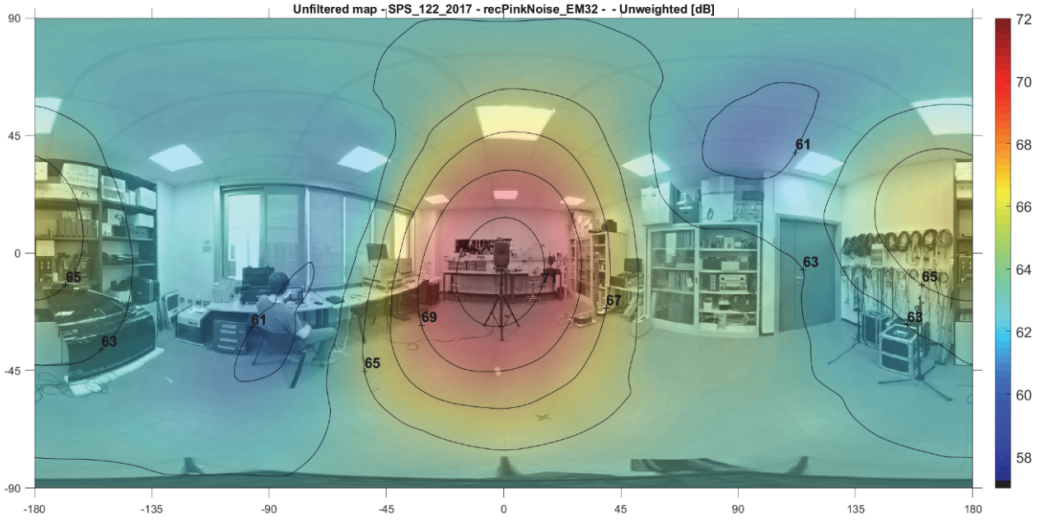


Figure 116: Colour map with 122 points grid

3.1.3. Image correction

Most of the times, as in the previous examples, the optical centre of the video recording system does not coincide with the acoustic centre of the array, which instead is one of the main advantages of the new array presented in 2.3. When the two centres are not coincident, as in Figure 117, some misalignments are generated between the background and the superimposed colour map. A solution has been developed to correct the image.

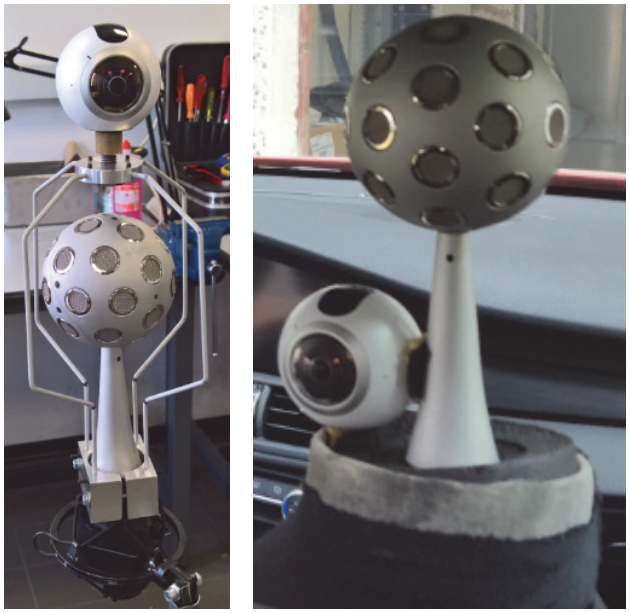


Figure 117: A panoramic camera mounted over the EM (left) or next to it (right)

The correction is based on two parameters: the offset between the two centres, either vertical or horizontal, and the radius of the sphere where the correction is applied, which should be the average distance between the array and the sound sources or obstacles around it. The amount of correction applied at each pixel of the image is calculated as follow:

$$\|C_{v\times h}\| = \frac{\tan^{-1}(R\cdot\sin\|E_{v\times h}\|-Y)}{R\cdot\cos\|E_{v\times h}\|+{}^{\pi}/_{2}}, \tag{23}$$

for the vertical offset and

$$\|C_{v\times h}\| = \frac{\tan^{-1}(R\cdot\sin\|A_{v\times h}\|-H)}{R\cdot\cos\|A_{v\times h}\|+\pi}, \tag{24}$$

for the horizontal offset, where v and h are the number of horizontal and vertical pixels, R is the radius of the sphere, Y and H are the vertical and horizontal offsets and with:

$$E_{v\times h} = \begin{pmatrix} {}^{\pi}/_{2} & \cdots & {}^{\pi}/_{2} \\ \vdots & \ddots & \vdots \\ -{}^{\pi}/_{2} & \cdots & -{}^{\pi}/_{2} \end{pmatrix}_{v\times h}, \tag{25}$$

$$A_{v\times h} = \begin{pmatrix} -\pi & \cdots & \pi \\ \vdots & \ddots & \vdots \\ -\pi & \cdots & \pi \end{pmatrix}_{v\times h}, \tag{26}$$

the matrices of elevation or azimuth values in each pixel.

In this way, the correction is progressive, maximum in the centre and null at the poles, as shown in Figure 118, where a vertical correction has been applied with $Y = 0.15\ m$ and $R = 2.5\ m$.

Figure 118: Example of correction of the vertical offset for a background image

The other correction implemented is the horizontal misalignment between the pointing directions of the array and the camera, caused by a positioning error, which is much easier to compensate being a rigid translation. An example is shown in Figure 119: a horizontal tilt of 45° has been applied to the same colour map of Figure 116.

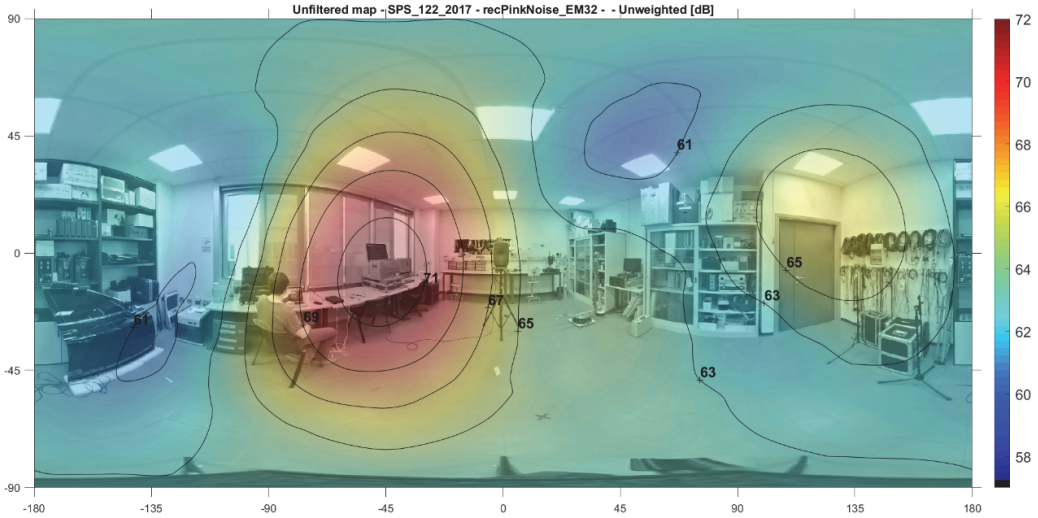


Figure 119: Example of horizontal tilting of the colour map

3.1.4. Calibration

The procedure to calibrate the SPL for a microphone array is described in the following paragraph. Note that both the array and the encoding matrix are calibrated together, therefore a specific calibration factor has to be calculated every time a new encoding matrix is employed.

A test signal is played by a Studio Monitor loudspeaker, having a very flat frequency response such as the Genelec 8351 AMP, inside an anechoic room. The test signal is a pink noise filtered in the 1 kHz octave band.

The SPL is measured at 1 m distance on-axis respect to the loudspeaker with a Sound Level Meter (SLM). A Bruel&Kjaer type 2260 (Figure 120, left) has been used, with capsule type 4189 calibrated with a Bruel&Kjaer type 4230 (Figure 120, right). In this case, the level of the loudspeaker has been set in order to have $SPL_{1m} = 80\ dB$.

The SLM is removed and the test signal is recorded in the same position with the microphone array. The A-format is then converted to B and P formats.

Figure 120: B&K 2260 (left) and B&K 4230 (right)

A first calibration factor k_{sps} is calculated for SPS format, so that the energetic sum of all the SPS signals results equal to the true SPL measured by the SLM:

$$k_{sps} = 10^{\left(\frac{SPL_{1m} - SPL_{tot}}{20}\right)}, \tag{27}$$

with:

$$SPL_{tot} = 20 \cdot \log(\sum SPS_{rms}), \tag{28}$$

where SPS_{rms} are the rms values of all the SPS signals.

A second calibration factor is calculated for Ambisonics, imposing that the level of the channel W, which is the virtual microphone having omnidirectional directivity pattern, equates the SPL measured by the SLM:

$$k_{ambix} = 10^{\left(\frac{SPL_{1m} - SPL_{omni}}{20}\right)}, \tag{29}$$

with:

$$SPL_{omni} = 20 \cdot \log(W_{rms}), \tag{30}$$

where W_{rms} is the rms value of the first signal of the B-format, which is the virtual microphone 0 following the ACN standard numbering.

3.1.5. Cross-correlation analysis

As explained in paragraph 2.3, it is possible to record synchronously some additional signals together with the 32 signals of the A format coming from the array. Employing this set of additional signals, called “references” in the following, it is possible to process the cross-correlation colour maps.

Before solving equation (17), (18) and (20) the PSD is filtered in frequency domain with the transfer function H_1, calculated as follow:

$$H_1(f) = \frac{P_{yx}(f)}{P_{xx}(f)}, \tag{31}$$

where $P_{yx}(f)$ and $P_{xx}(f)$ are respectively the cross-spectrum and the auto-spectrum of the signals x and y. The transfer function $H_1(f)$ must be evaluated between each of

the P signals x of the SPS format and each of the y references. The transfer function is calculated using Welch's Averaged Periodogram [37].

Thanks to this method, it is possible to produce colour maps where only the noise coherent with the references is plotted, which is very useful to study separately the effects produced by several noise sources active together. In example, when recording the noise inside a cockpit with the car running on the road, it is possible to "clean" the colour map with the cross-correlation between the array and a series of accelerometers, mounted on the engine, suspension brackets or panels.

As an example, a recording made with the EM is processed: two loudspeakers are playing two uncorrelated test signals, which are recorded also with two additional microphones, placed one in front of each loudspeaker. The P-122 format is encoded and a colour map is produced (Figure 121) with a threshold of 5 dB and isolevel curves every 0.5 dB.

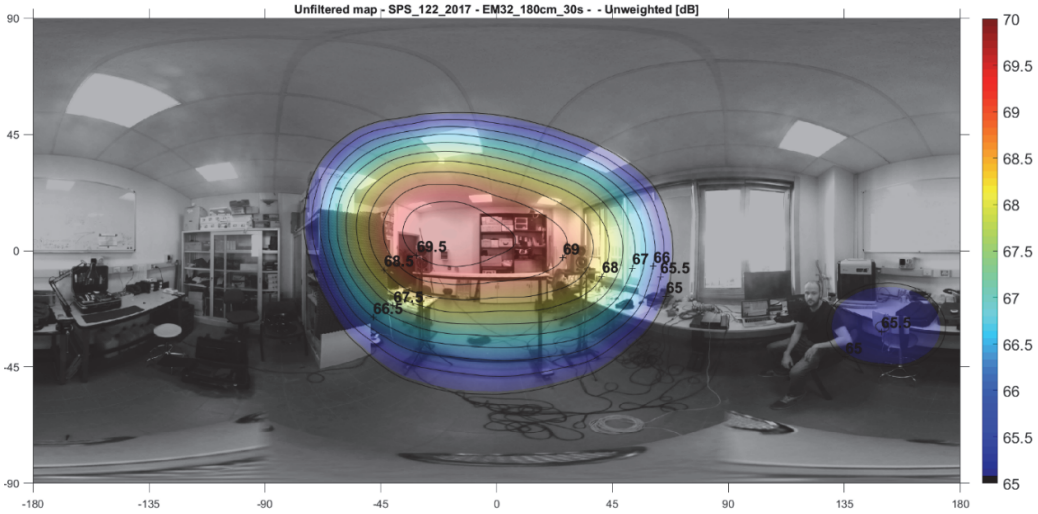


Figure 121: Colour map of two loudspeakers playing together two uncorrelated signals

Then two colour maps are produced with the cross-correlation filtering, the first one employing left microphone as reference (Figure 122), the second one employing right microphone as reference (Figure 123). Note that the recording of the array is not changed and the two loudspeakers were playing together.

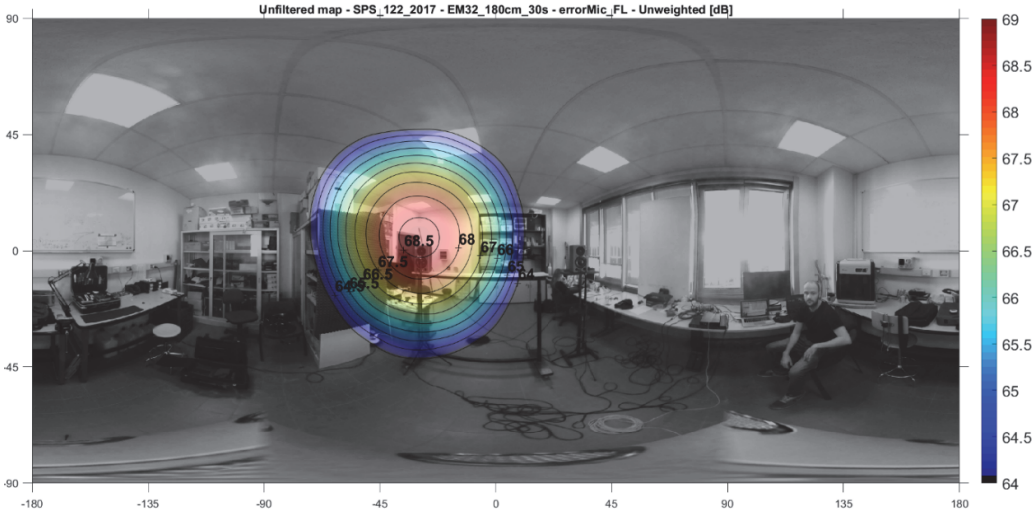


Figure 122: Cross-correlation colour map of two loudspeakers playing together two uncorrelated signals, left reference

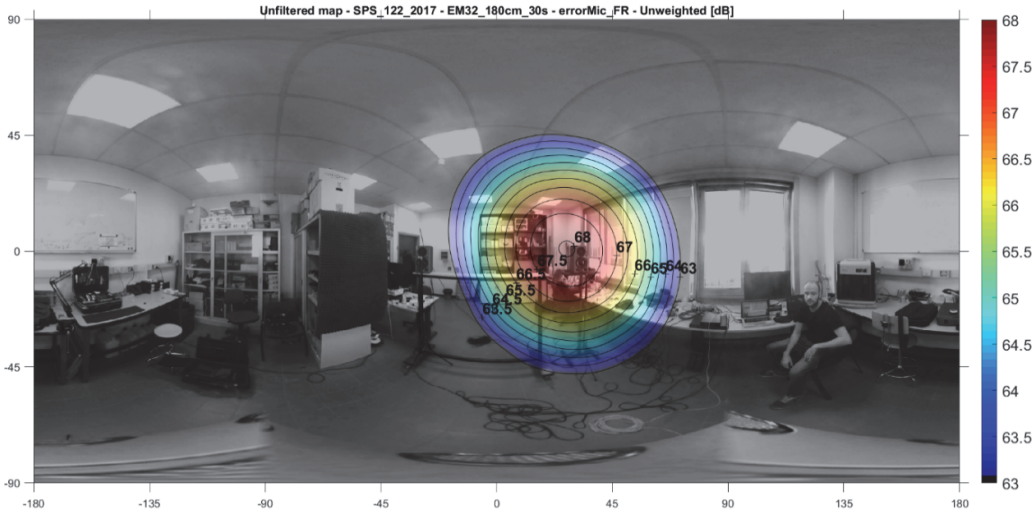


Figure 123: Cross-correlation colour map of two loudspeakers playing together two uncorrelated signals, right reference

Additionally, a total coherence colour map, which is the mixing between the previous two, can be produced (Figure 124). This result substantially corresponds to the map of Figure 121, enhanced by the cross-correlation processing. Note that one loudspeaker was playing slightly louder than the other, as shown also in Figure 121. Finally, a normalized total coherence colour map can be produced (Figure 125): the information of the relative amplitude between the sources is lost, but all the active sources are now equally individualized.

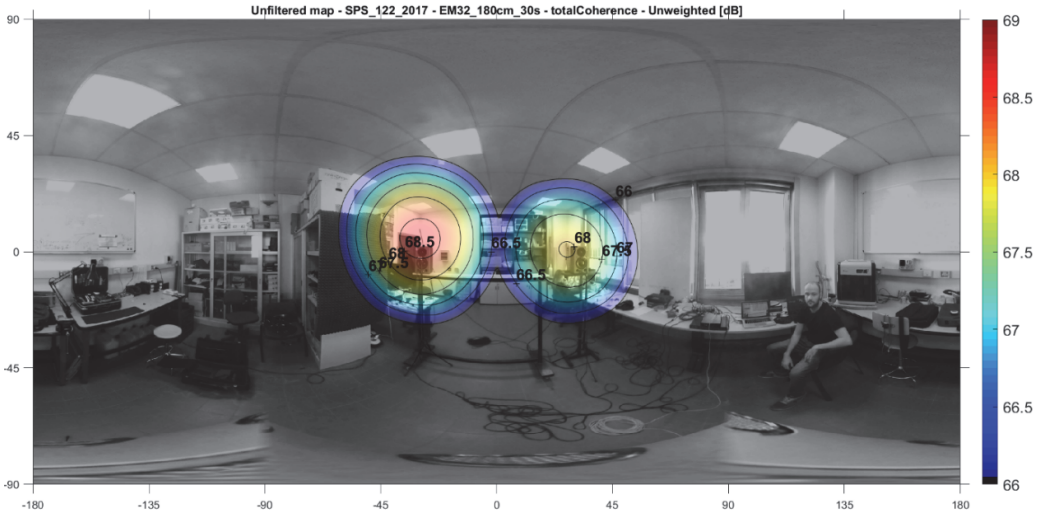


Figure 124: Total coherence map

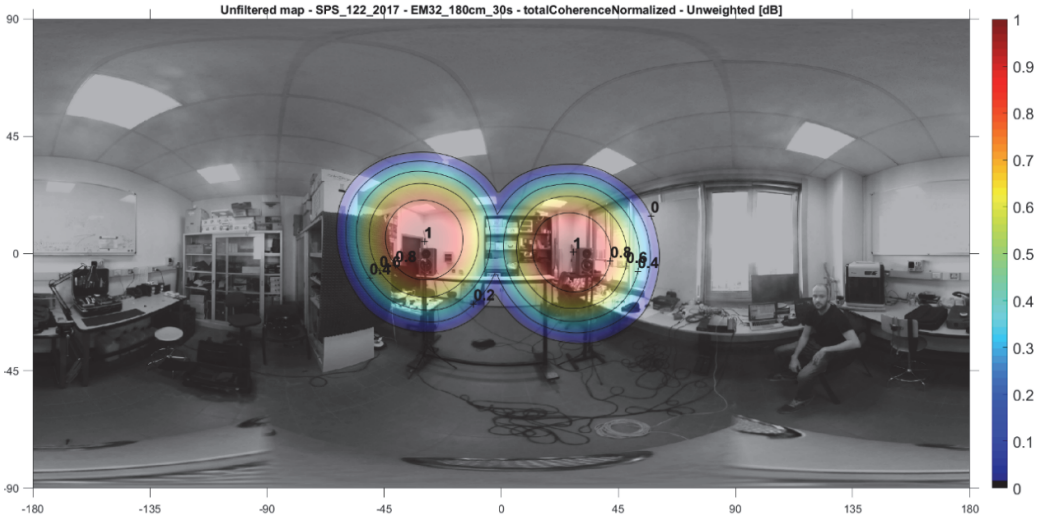


Figure 125: Normalized total coherence map

3.1.6. Dynamic mapping

Two types of dynamic mapping have been implemented: video map over background image and video map over background video.

In both cases, the encoded format is processed in chunk of signals, called $buffer$. The dimension of the buffer and the overlap are related each other, because they define the frame rate of the output video. If the frame rate is imposed (common values are 30 fps or 60 fps) the overlap is:

$$overlap = 1 - \frac{fs}{fps \cdot buffer}, \tag{32}$$

where fs is the sampling frequency of the recorded signal. As an example, with common values the result is: $overlap = 1 - \frac{48000}{30 \cdot 16384} \cong 0.9$.

This is very important to produce VR video, particularly if a background video is employed, to avoid the effect of speeding up or slowing down the result. If instead the output is produced mainly for the analysis, it could be useful to reduce the speed of the output. In this case, the overlap is set close to one, i.e. 0.97.

When a video is employed for the background, it is fundamental to align temporally the array recording with the panoramic video: being in fact two separate systems, these recordings are not synchronised. The realignment is calculated by means of the transfer function (eq. 27) between the array recording and the audio recorded by the video system. Once the transfer function $H_1(f)$ is calculated, the associated IR in time domain is obtained by means of the IFFT; then, the shift between the video and the array recording is given by the delay of the peak of the transfer function filter.

To speed up the processing, the colour map is not generated and superimposed on the background frame by frame. Instead, all the frames of the colour map are calculated first and then generated, filling in black those areas where the energy is below the threshold, hence that should be transparent (Figure 126). The result is stored in a temporary file. Finally, the temporary colour map video is superimposed with transparency on the background video (Figure 127) with the software FFmpeg [38], by means of the following complex filter:

```
'ffmpeg -y -i colormap.mp4 -i background.mp4 -filter_complex "
    ...
'[0:v]setpts=PTS-STARTPTS, scale=' width 'x' height '[top]; '
    ...
'[1:v]setpts=PTS-STARTPTS, scale=' width 'x' height '[bottom]; '
    ...
'[top]split[m][a]; [m][a]alphamerge[keyed]; [bottom] [keyed]
    ...
 overlay=eof_action=endall" ' overlay.mp4
```

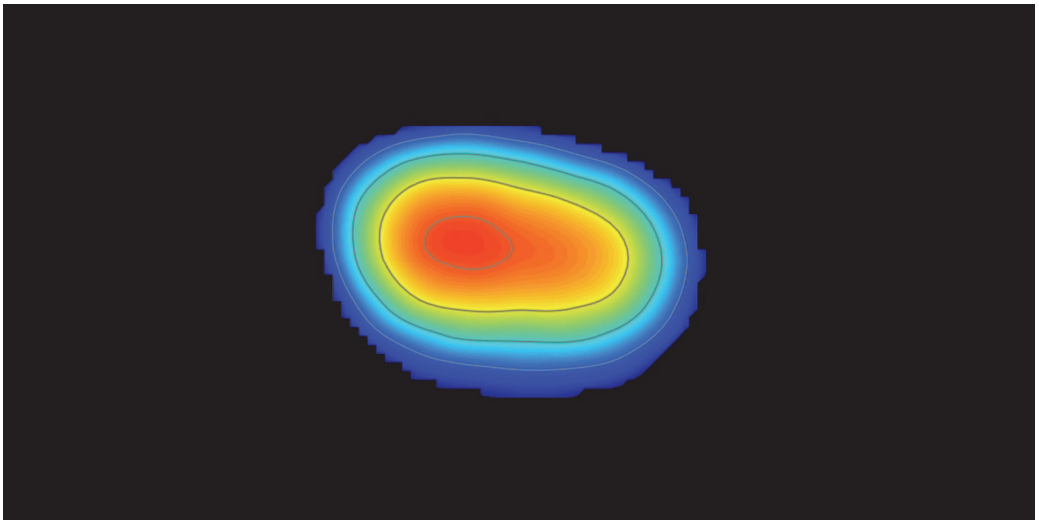


Figure 126: Temporary colour map video with black areas

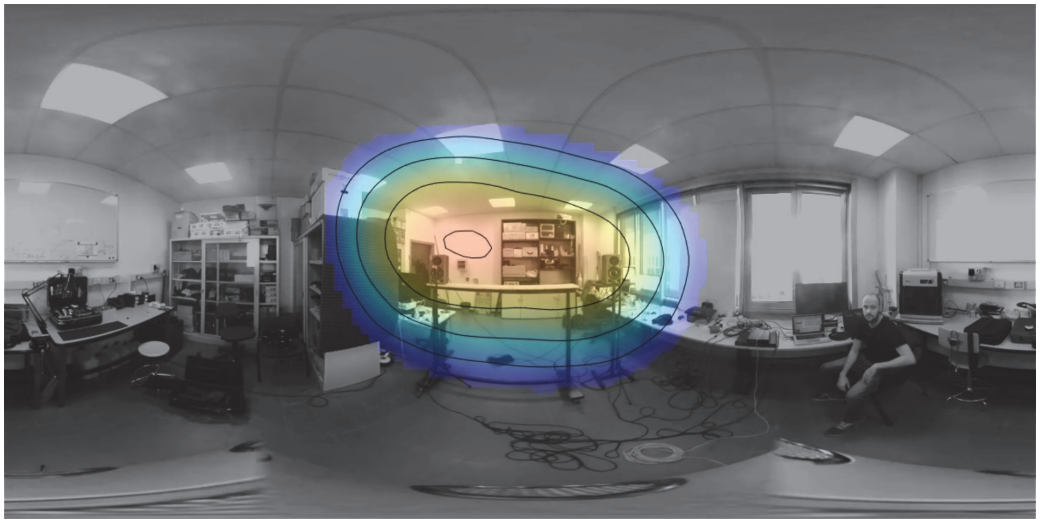


Figure 127: Colour map video superimposed over the background with transparency

When the background is a static image, a background video is produced by repeating the image, with the software FFmpeg:

```
  'ffmpeg -loop 1 -i image.png -c:v libx264 -t ' time ' -pix_fmt
yuv420p backgroundVideo.mp4'
```

If the target is the analysis and not the VR reproduction, it is also possible to add a view of the signal in time domain under the video with a moving pointer (Figure 128).

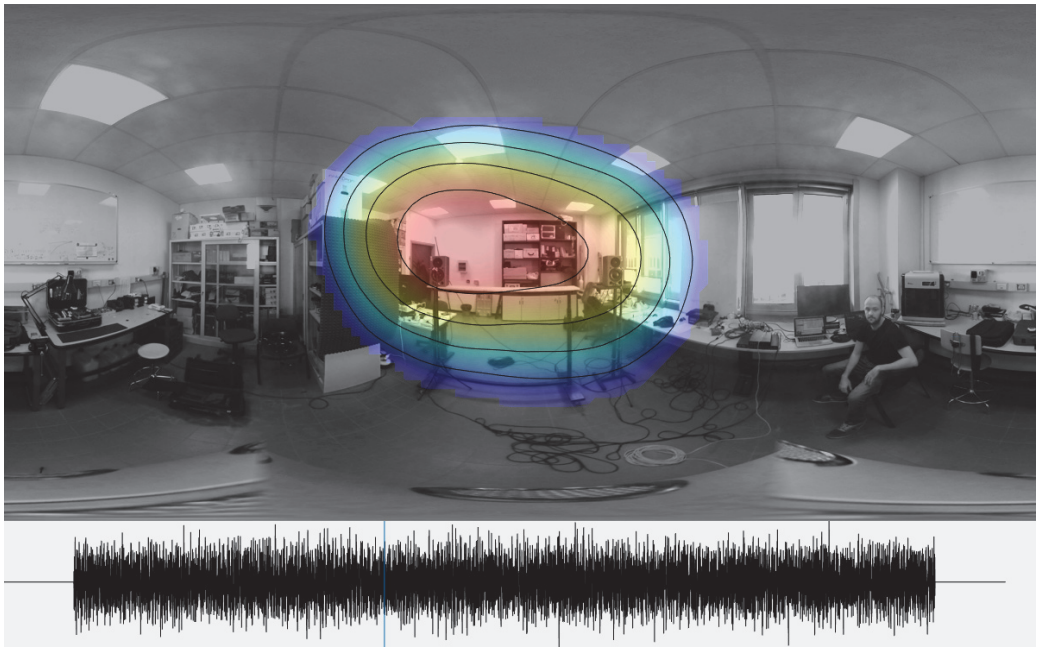


Figure 128: Time signal under the colour map

If the target is analysis, the suggested output format of the video is an .mp4 file, in case with a mono sound track corresponding to the first signal of Ambisonics format, that is an omnidirectional microphone. If instead the target is VR reproduction,

the suggested output format is a ".mov" file, which currently is the only video format capable of integrating a sound track with more than eight channels. It is possible in fact to add a .wav audio track made of 16 channels (Ambisonics third order), 16 bits, 48 kHz. In that case, some metadata are injected at the end of the processing, so that video players can automatically recognise the video format and start to reproduce it in 360° mode, decoding properly also the sound track.

3.2. Application example – Head-Shaped Array

The performances of the new array have been compared to the EM and then it has been successfully employed to evaluate automotive ANC systems [23].

First, frequency limits calculated for each Ambisonics orders from the analysis of spatial performances have been checked and then compared with theoretical results. The usage of this limit is fundamental, particularly if the encoding matrix is calculated with a simulated response, to ensure a correct beamforming. The recording employed is the one described in 3.1.2: a loudspeaker that plays 30 s of pink noise inside the laboratory, recorded by each array one at a time, at 2.5 m distance.

As an example, two colour maps in the range $20\,Hz - 20\,kHz$ have been produced with the signal recorded by the HSA and employing a filtering matrix for Ambisonics up to order four, calculated with the simulated response. In Figure 129, it is presented the result without limiting the Ambisonics orders, while in Figure 130 frequency limits have been introduced. One can note that in the first case, the result is completely wrong.

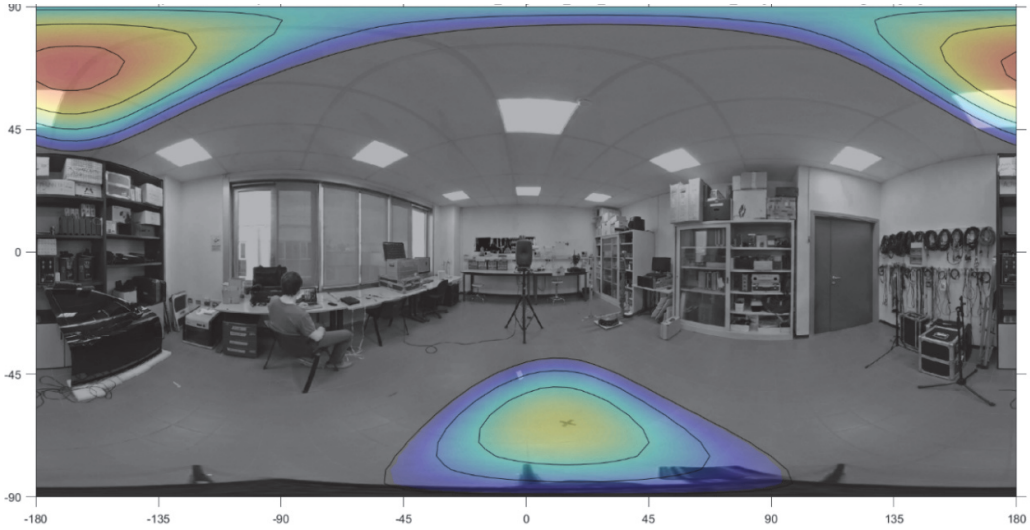


Figure 129: HSA, colour map without limiting Ambisonics orders

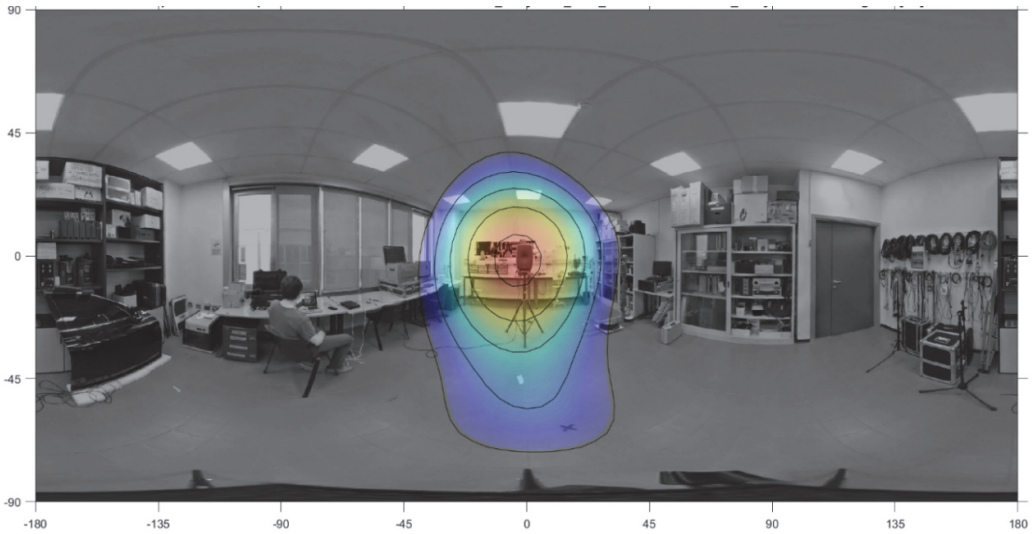


Figure 130: HSA, colour map with Ambisonics orders limited in frequency

The filtering matrices for Ambisonics 4th order encoding obtained with simulations of the two arrays have been tested carefully, by comparing a series of colour maps at various octave bands from 31.5 Hz to 2 kHz. Initially, frequency limits have been set as the ones previously calculated (Table 7 and Table 12) and reported in Table 15. Then, limits have been tuned manually, lowering if possible or raising when necessary. Final results are showed in Table 16.

	EM Simulated response Kirkeby inversion		**HSA Simulated response Kirkeby inversion**	
Ambisonics order	**Freq. start [Hz]**	**Freq. stop [Hz]**	**Freq. start [Hz]**	**Freq. stop [kHz]**
1	30	3.5	20	2.5
2	430	3.5	170	2.5
3	1250	3.5	510	2.5
4	2050	3.5	880	2.3

Table 15: EM and HSA, frequency limits for Ambisonics orders 1-2-3-4 calculated with the spatial performance analysis

	EM Simulated response Kirkeby inversion		**HSA Simulated response Kirkeby inversion**	
Ambisonics order	**Freq. start [Hz]**	**Freq. stop [Hz]**	**Freq. start [Hz]**	**Freq. stop [kHz]**
1	30	3.5	20	2.5
2	550	3.5	260	2.5
3	900	3.5	550	2.5
4	1500	3.5	750	2.3

Table 16: EM and HSA, frequency limits for Ambisonics orders 1-2-3-4 tuned with colour map analysis

In Figure 131 and Figure 132, it is showed the increase of spatial resolution with the EM and HSA in the octave band centred at 500 Hz. The first array, get the benefit only of the second order, while the second one reaches the third. Note that performances of 1st order are comparable, whilst HSA is much more directive already at 2nd order.

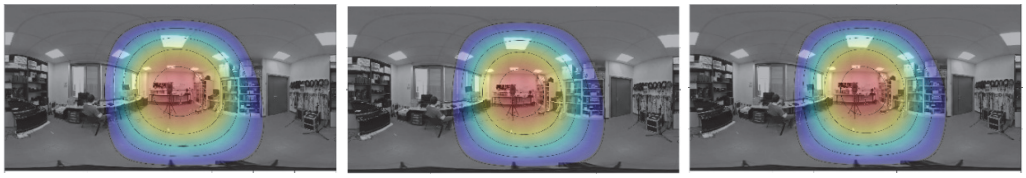


Figure 131: EM, 500 Hz octave band, 1st order (left), 2nd order (middle) and 3rd order (right)

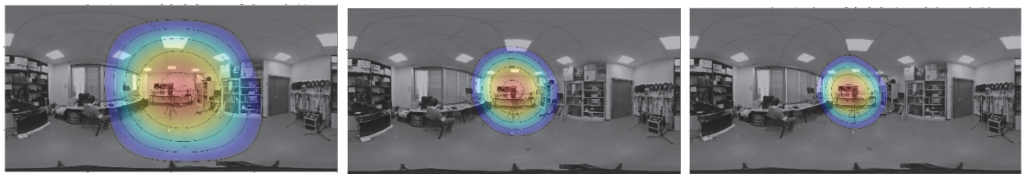


Figure 132: HSA, 500 Hz octave band, 1st order (left), 2nd order (middle) and 3rd order (right)

The benefits of on-axis filtering (described in 2.2.5.5) have been tested with the EM. The filtering matrix previously obtained has been convolved with the inverse filters of IRs of the capsules, measured on-axis respect to the loudspeaker in an anechoic room. The introduction of this filtering made it possible to reduce considerably the limits at each order, as showed in Table 17.

	EM Simulated response Kirkeby inversion		**EM Simulated response Kirkeby inversion On-axis filtering**	
Ambisonics order	**Freq. start [Hz]**	**Freq. stop [Hz]**	**Freq. start [Hz]**	**Freq. stop [kHz]**
1	30	3.5	30	3.5
2	550	3.5	150	3.5
3	900	3.5	780	3.5
4	1500	3.5	1500	3.5

Table 17: EM, comparison of frequency limits for Ambisonics orders 1-2-3-4 with and without on-axis response filtering

Figure 133 shows the octave bands centred at 250 Hz and 500 Hz for which first and second orders are available, in accordance with Table 16. Figure 134 shows the same octave bands, processed with encoding matrix filtered with on-axis response of the capsules, therefore with limit of Table 17. The improvement of the beamforming performance is substantial.

Figure 133: EM, octave bands at 250 Hz and 500 Hz, without on-axis capsule response filtering

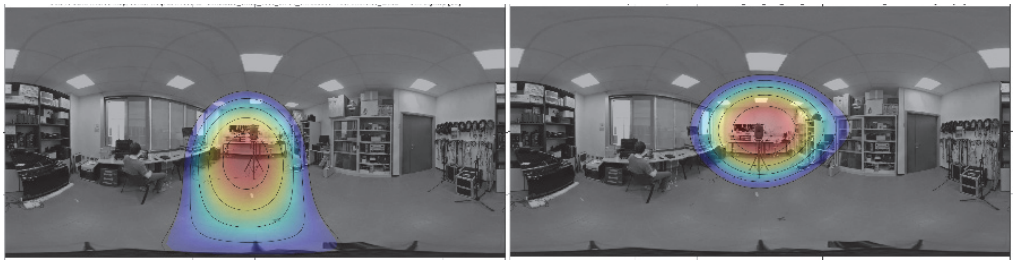


Figure 134: EM, octave bands at 250 Hz and 500 Hz, with on-axis capsule response filtering

To conclude, a similar recording taken at the reduced distance of $1\ m$ from the loudspeaker with both arrays has been processed. Figure 135 shows the EM (left) compared to the HSA (right) in the octave band centred at $31.5\ Hz$. One can note the great improvement of accuracy gained with the latter.

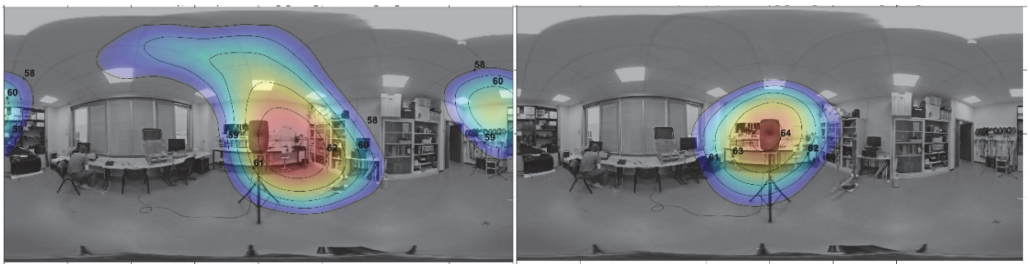


Figure 135: EM (right) and HSA (left), comparison of spatial resolution in the 31.5 Hz octave band

3.2.1. Evaluation of performances of an ENC system

The first field usage of the HSA presented consists in the performance evaluation of an Engine Noise Cancelling (ENC) system, a tonal algorithm that reduces the engine order components, installed on a sedan car.

The measurement has been done with the car stopped, neutral gear, engine at $3500\ rpm$, the array placed in the passenger seat, mounted on a dummy torso (Figure 136, left). Two recordings of thirty second each one have been taken, the first one with the ENC system switched off, the second one with the ENC system switched on. The panoramic image for the background has been taken by employing a Ricoh Theta V

mounted on the dummy torso with an appositely built support. Therefore, the array was removed and the optical centre has been positioned coincident with the acoustic centre of the array manually. The headrest has been dismounted, to ensure the best quality of the background picture in terms of uniformity of the lighting and visibility of the background (Figure 136, right).

Figure 136: HSA mounted on a dummy torso inside a car (left) and panoramic picture of a cockpit (right)

The 32 signals of the A-format have been averaged and PSD have been calculated; in Figure 137, the result is presented for the two recordings, ENC-off (black) and ENC-on (green). Note that system is effective in the frequency range $50\ Hz - 700\ Hz$. The cancelling effect is more evident at the following peaks: $13\ dB$ at $58\ Hz$, $20\ dB$ at $117\ Hz$, $21\ dB$ at $164\ Hz$ and $38.6\ dB$ at $234\ Hz$.

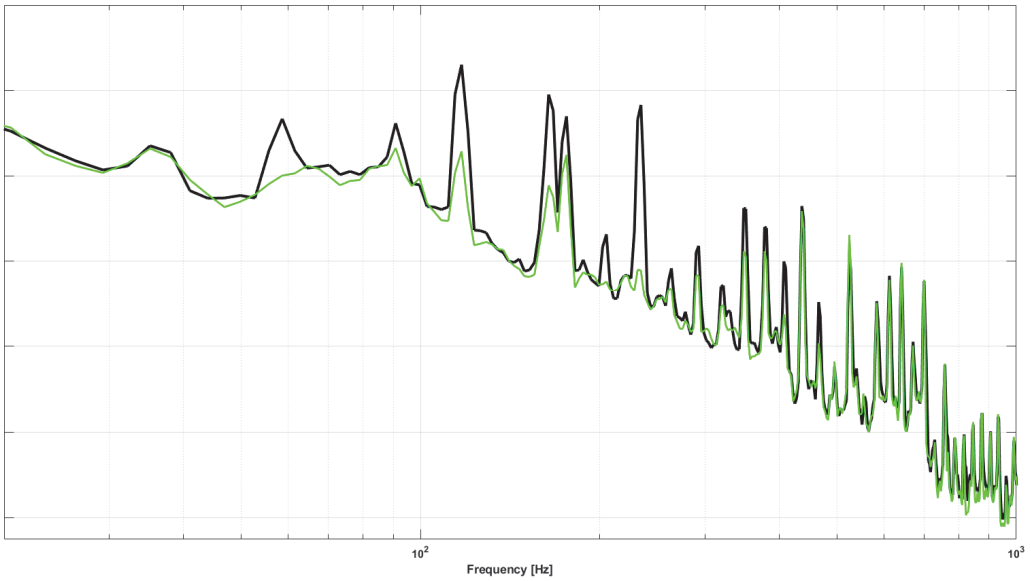


Figure 137: PSD of averaged signals, ENC-off (black) and ENC-on (green)

Colour maps have been produced for each peak with Ambisonics format and PWD method, employing the following band-pass filtering:

- Figure 138 (ENC-off) and Figure 139 (ENC-on), range $52.7\ Hz - 64.4\ Hz$;
- Figure 140 (ENC-off) and Figure 141 (ENC-on), range $111.3\ \text{Hz} - 123\ \text{Hz}$;
- Figure 142 (ENC-off) and Figure 143 (ENC-on), range $152.3\ \text{Hz} - 181.6\ \text{Hz}$;

- Figure 144 (ENC-off) and Figure 145 (ENC-on), range 225.6 Hz – 243.2 Hz.

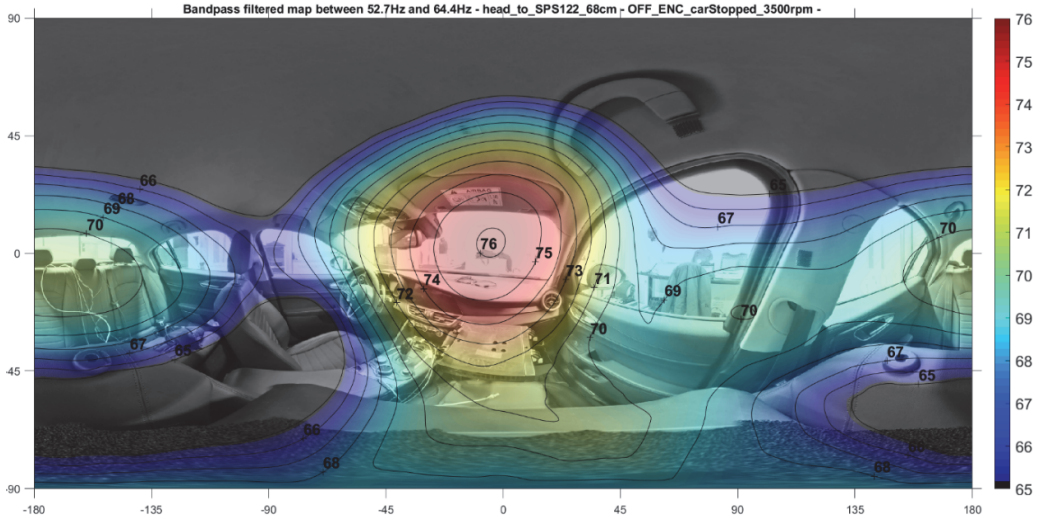


Figure 138: ENC-off, band-pass filtered map, 52.7 Hz – 64.4 Hz

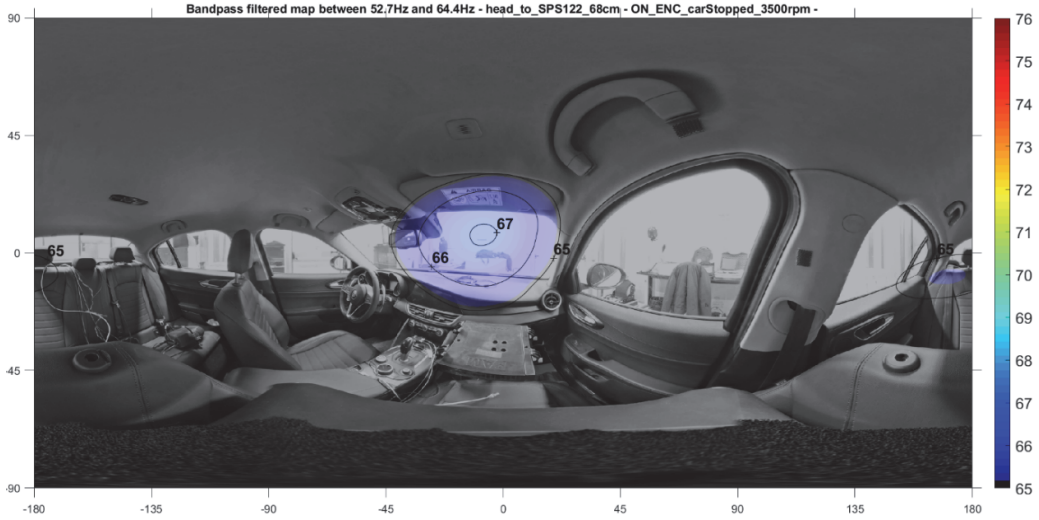


Figure 139: ENC-on, band-pass filtered map, 52.7 Hz – 64.4 Hz

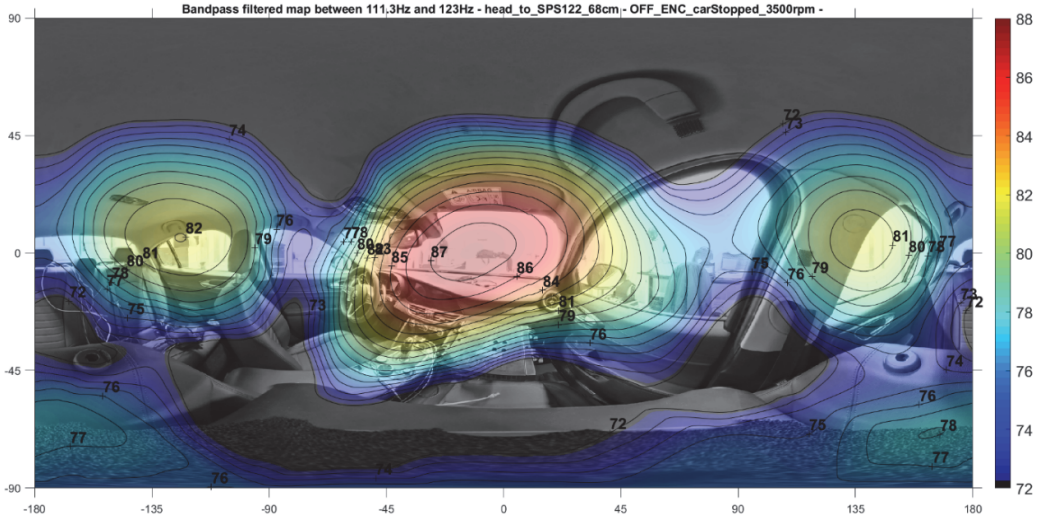


Figure 140: ENC-off, band-pass filtered map, $111.3\ Hz - 123\ Hz$

Figure 141: ENC-on, band-pass filtered map, $111.3\ Hz - 123\ Hz$

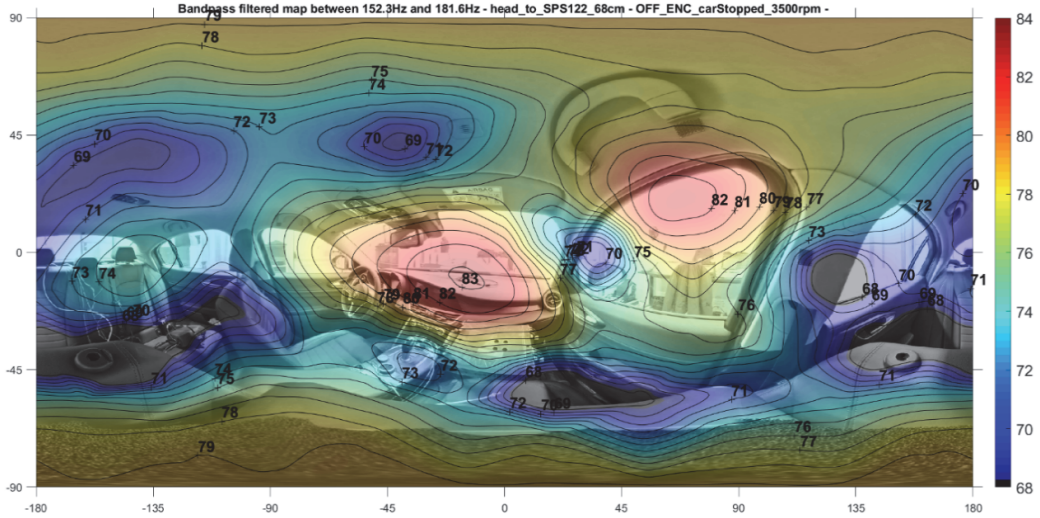


Figure 142: ENC-off, band-pass filtered map, 152.3 *Hz* – 181.6 *Hz*

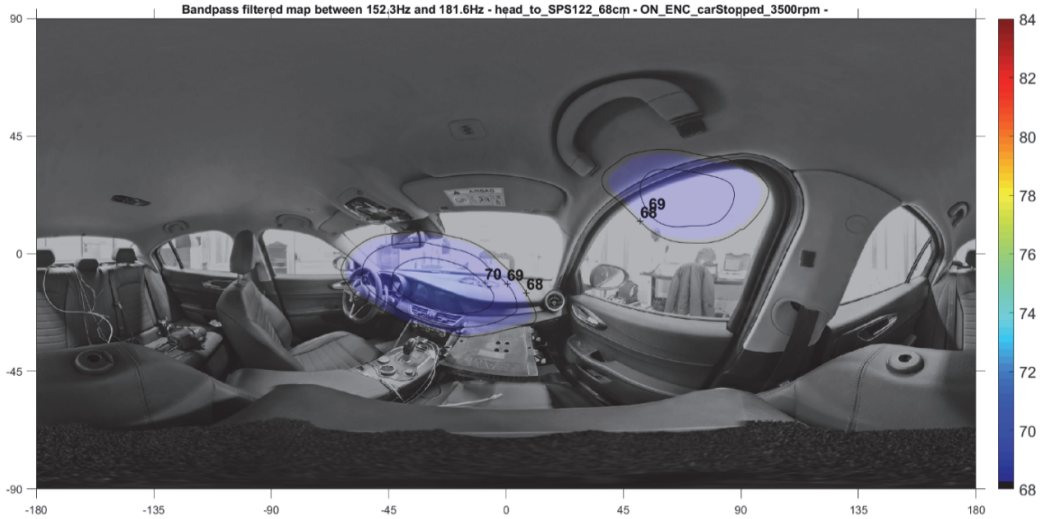


Figure 143: ENC-on, band-pass filtered map, 152.3 *Hz* – 181.6 *Hz*

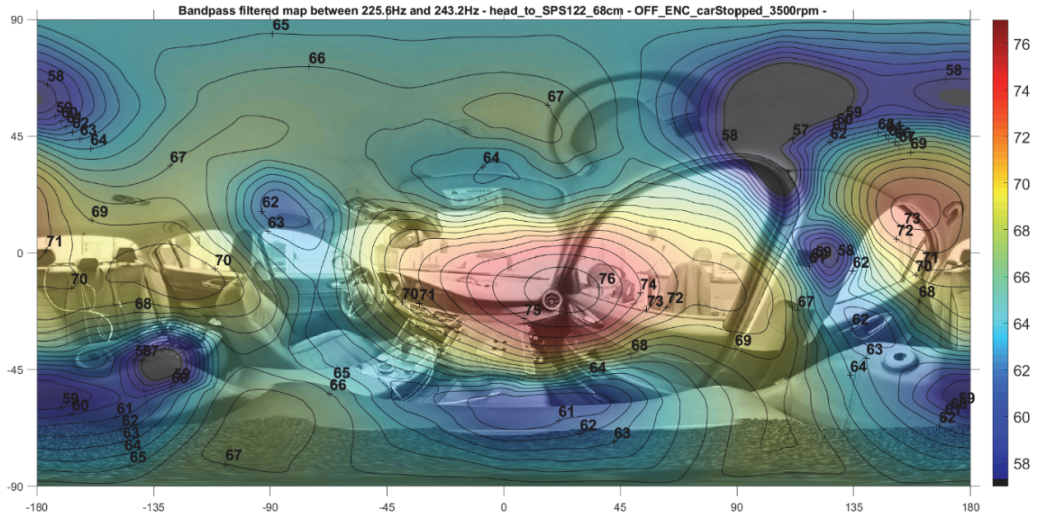


Figure 144: ENC-off, band-pass filtered map, 225.6 Hz – 243.2 Hz

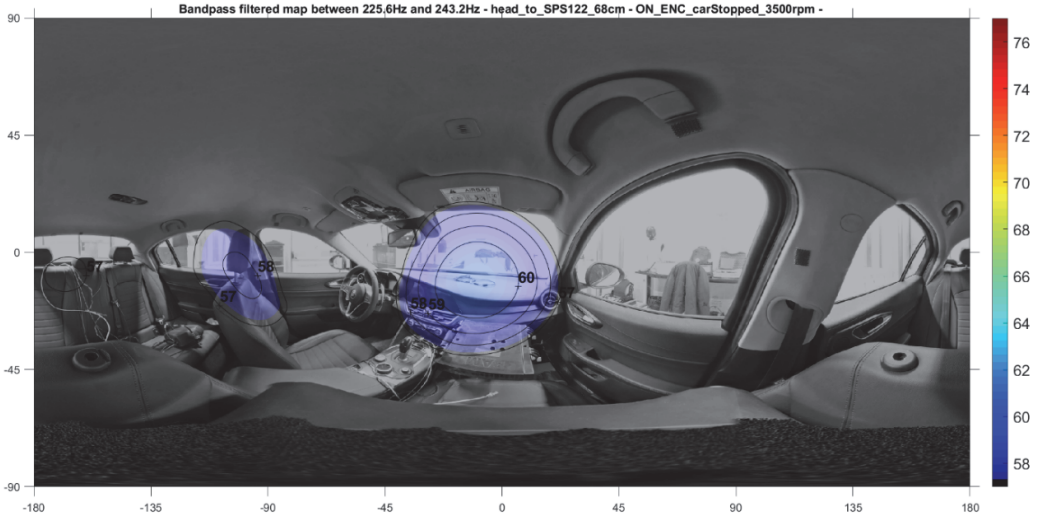


Figure 145: ENC-on, band-pass filtered map, 225.6 Hz – 243.2 Hz

3.2.2. Evaluation of performances of a RNC system

The second field usage presented is the evaluation of the performances of a Road Noise Cancelling (RNC) system, a broadband algorithm that reduces the rolling noise generated by the contact of the wheels with the asphalt, installed in the same car of the previous case.

The measurement has been done with the car running on a straight road with rough asphalt, 4th gear, cruise control at 40 km/h, the array placed in the passenger seat, mounted on a dummy torso (Figure 136, left). Two recordings of thirty second each one have been taken, the first one with the RNC system switched off, the second one with the RNC system switched on. A panoramic video for the background has been

recorded with the ring of GoPro cameras and then stitched with the software Kolor Autopano Video and Kolor Autopano Giga. A frame has been extracted from the video to get the background picture (Figure 146). The headrest was dismounted during the measurement session to ensure the uniformity of the lighting and visibility of the background, which are very important to obtain a good stitching.

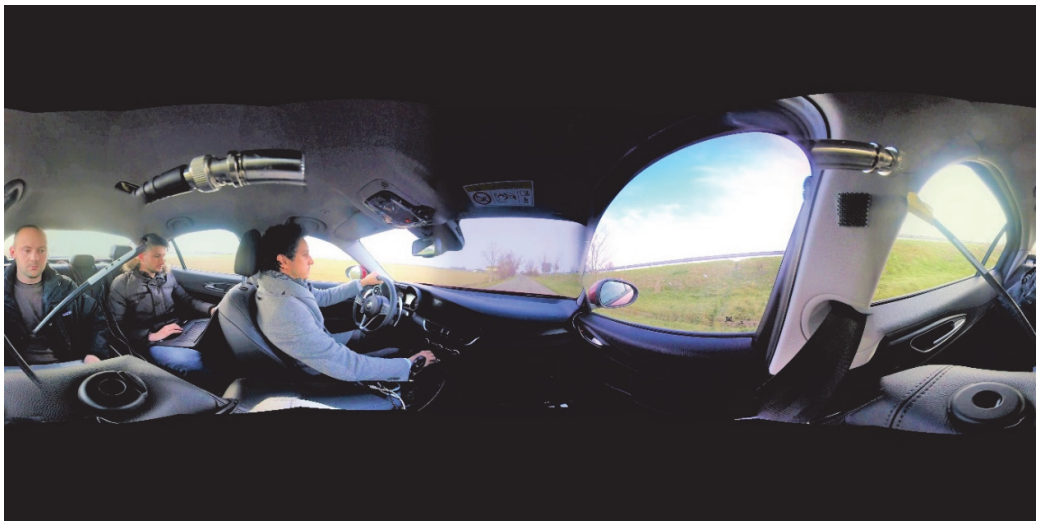


Figure 146: Panoramic picture stitched with the ring of eight GoPro cameras

The 32 signals of the A-format have been averaged and PSD have been calculated; in Figure 147, the result is presented for the two recordings, RNC-off (black) and RNC-on (green). Note that system is effective in the frequency range $20\,Hz - 300\,Hz$. Being a wideband algorithm, the cancelling effect is not concentrated on a few peaks with a great amount of reduction. Conversely, there is a lower amount of reduction in the whole working range.

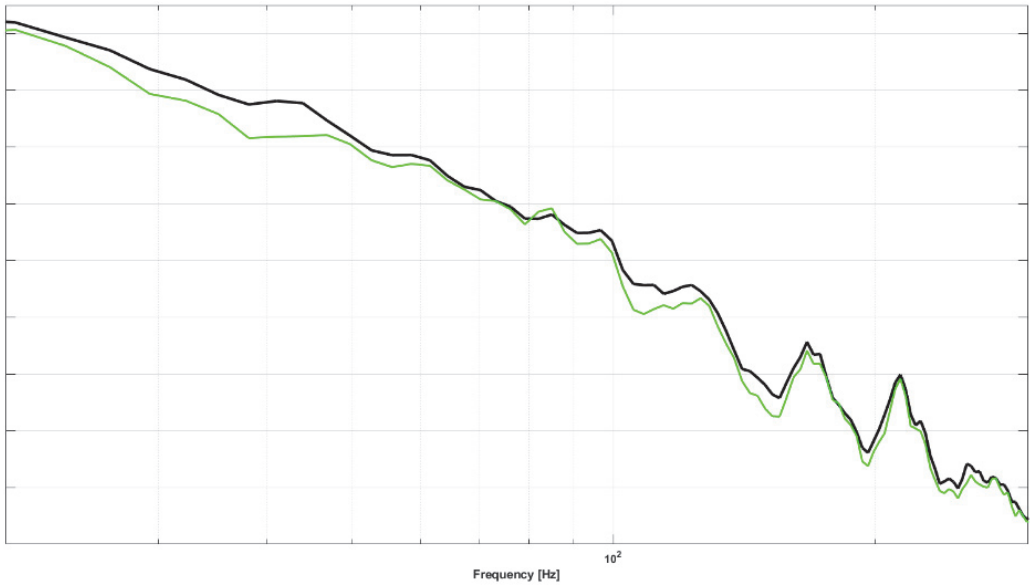


Figure 147: PSD of averaged signals, RNC-off (black) and RNC-on (green)

Colour maps have been produced for each range with Ambisonics format and PWD method, employing the following band-pass filtering:

- Figure 148 (RNC-off) and Figure 149 (RNC-on), range $20\,Hz - 61.5\,Hz$ ($7\,dB$ of max reduction);
- Figure 150 (RNC-off) and Figure 151 (RNC-on), range $87.9\,Hz - 128.9\,Hz$ ($5\,dB$ of max reduction);
- Figure 152 (RNC-off) and Figure 153 (RNC-on), range $140.6\,Hz - 172.9$ ($3.9\,dB$ of max reduction).

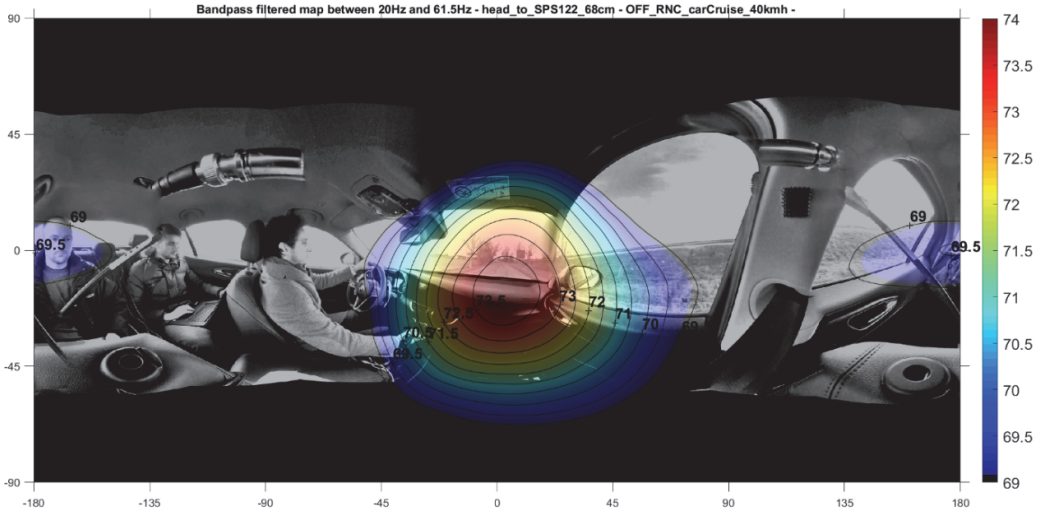


Figure 148: RNC-off, band-pass filtered map, $20\,Hz - 61.5\,Hz$

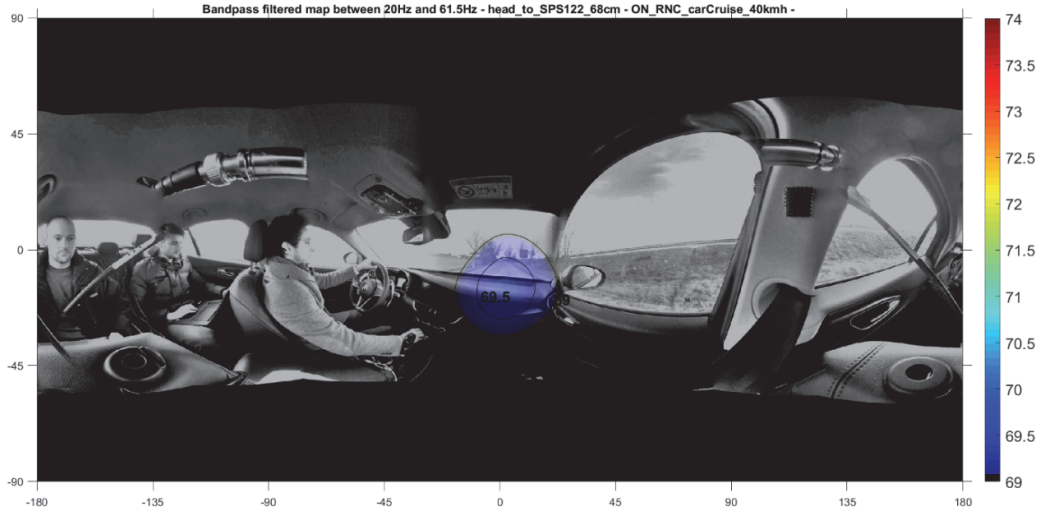


Figure 149: RNC-on, band-pass filtered map, $20\,Hz - 61.5\,Hz$

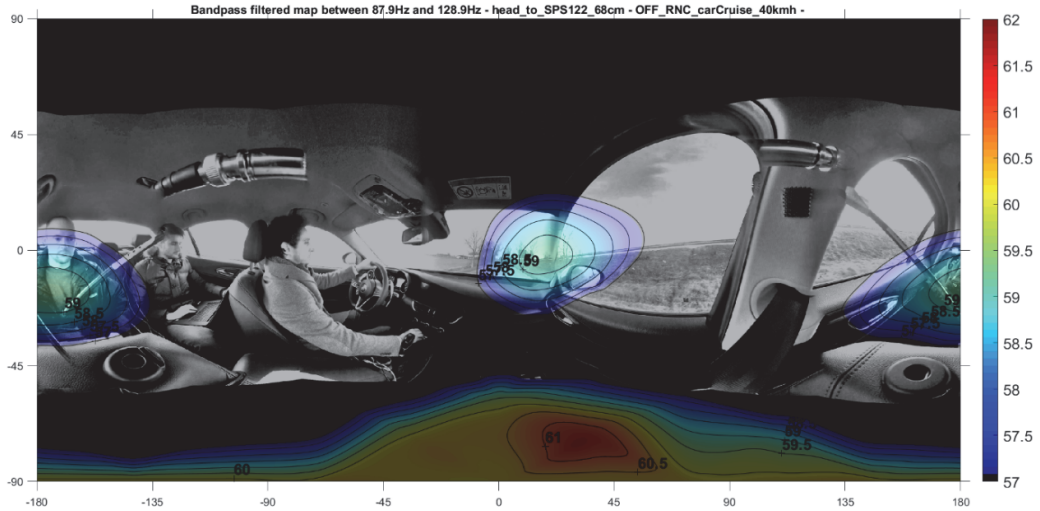


Figure 150: RNC-off, band-pass filtered map, 87.9 Hz – 128.9 Hz

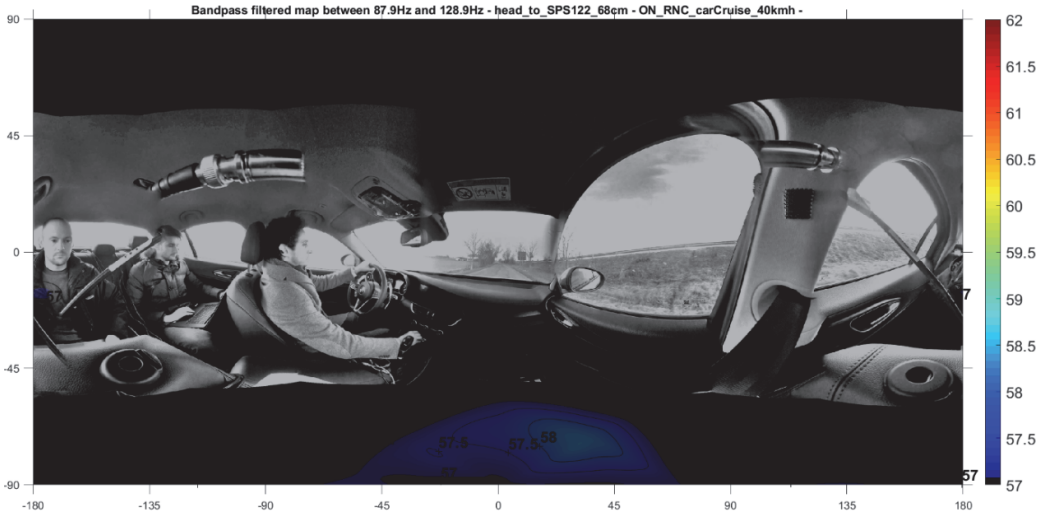


Figure 151: RNC-on, band-pass filtered map, 87.9 Hz – 128.9 Hz

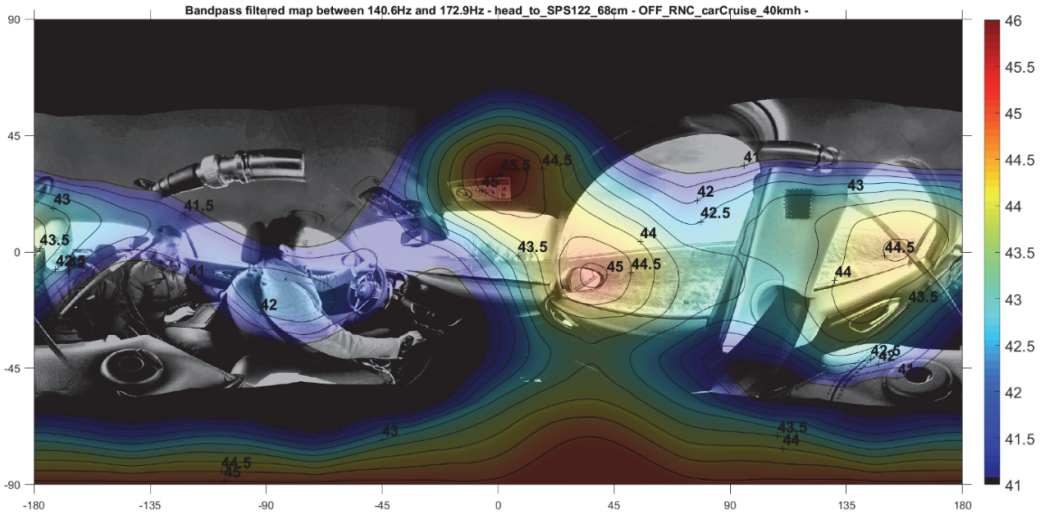


Figure 152: RNC-off, band-pass filtered map, 140.6 *Hz* – 172.9 *Hz*

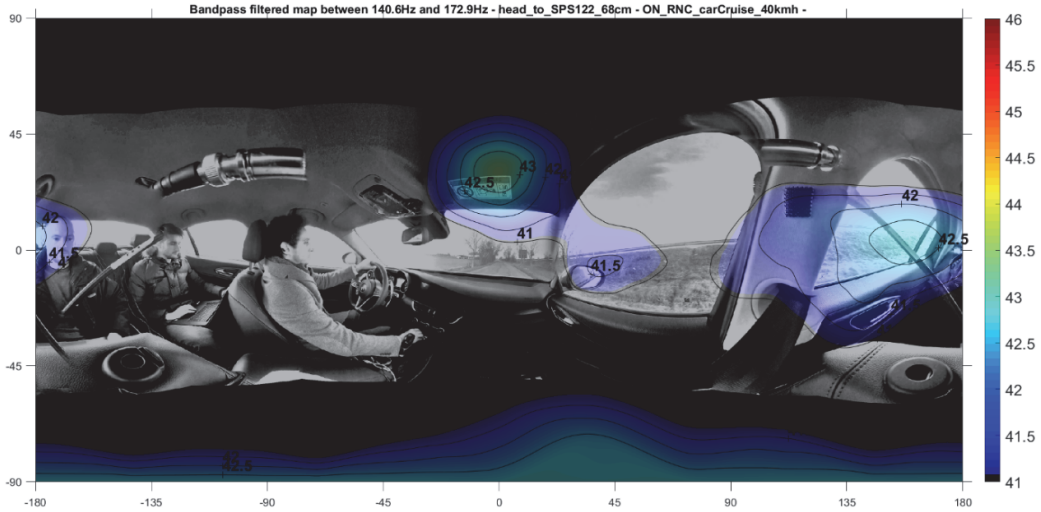


Figure 153: RNC-on, band-pass filtered map, 140.6 *Hz* – 172.9 *Hz*

The last field usage presented is another evaluation of the performances of a RNC system, installed on C-segment car.

This time, the system has been measured with both arrays, the EM and the HSA, mounted one at a time on a mannequin torso positioned on the passenger seat. Test recordings of $50\ s$ have been taken on a straight rough road at $55\ km/h$ with the RNC system initially switched off and then switched on. PSD of the signals recorded in the two conditions are shown in Figure 154, RNC-off (black) and RNC-on (green). Note that system is effective in the frequency range $40\ Hz - 330\ Hz$.

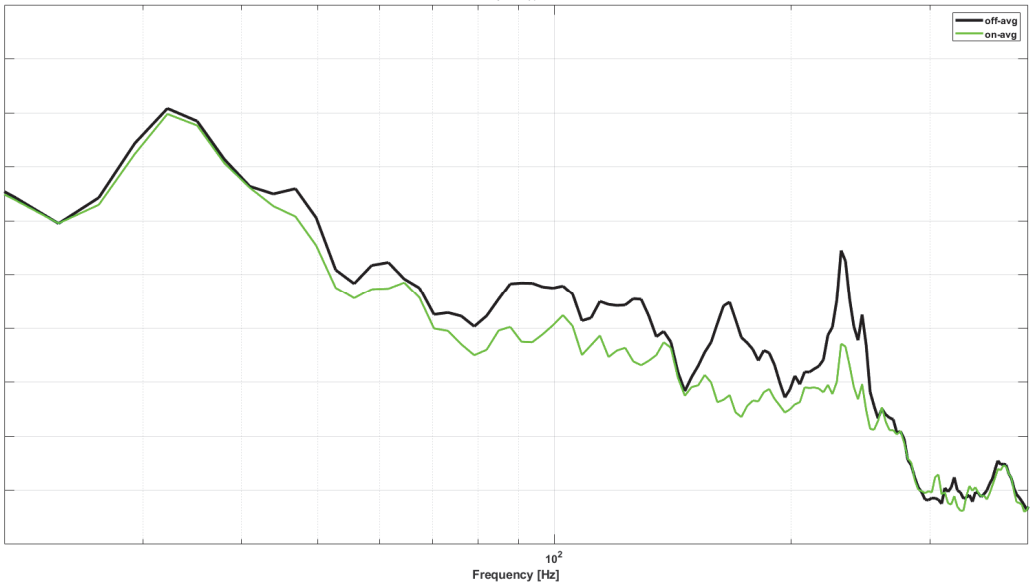


Figure 154: PSD of averaged signals, RNC-off (black) and RNC-on (green)

Colour maps filtered in the range $70\ Hz - 330\ Hz$ have been calculated in the two working conditions for the EM (Figure 155 and Figure 156) and for the HSA (Figure 157 and Figure 158), employing Ambisonics format and PWD method. In the analysed frequency range, the EM provides limited spatial resolution. The HSA, instead, provides more spatial detail, resulting in sharper maps.

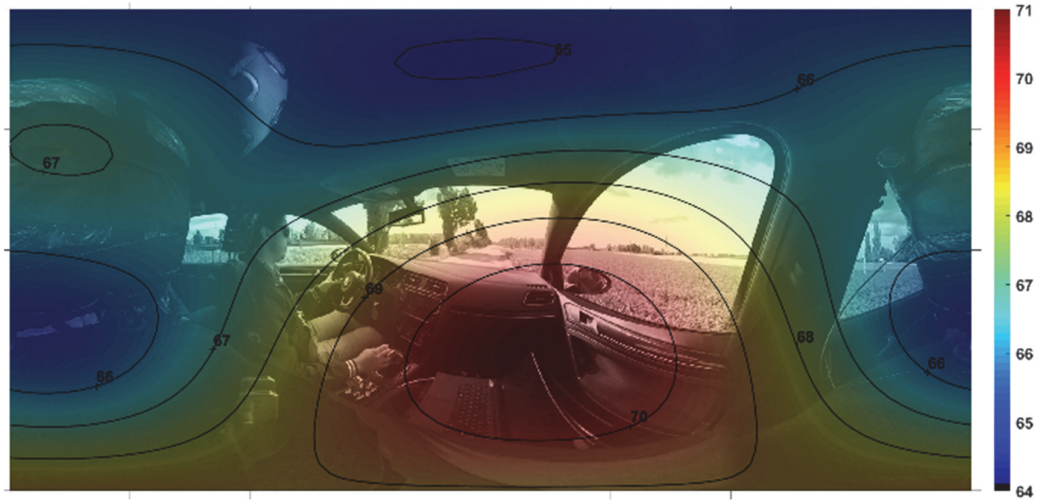


Figure 155: EM, band-pass filtered map, 70 Hz – 330 Hz, RNC-off

Figure 156: EM, band-pass filtered map, 70 Hz – 330 Hz, RNC-on

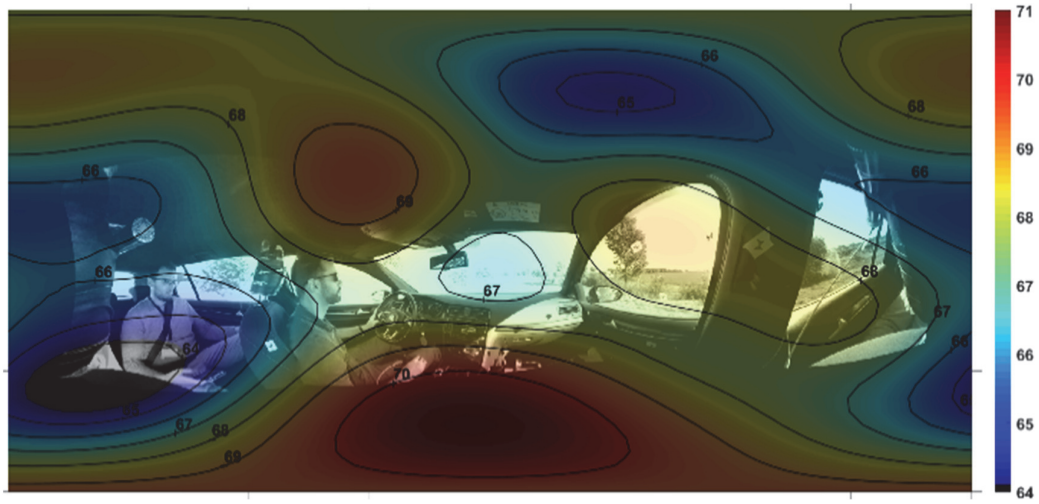


Figure 157: HSA, band-pass filtered map, 70 Hz – 330 Hz, RNC-off

Figure 158: HSA, band-pass filtered map, 70 Hz – 330 Hz, RNC-on

In Figure 157 (HSA, RNC-off), the effects of windows and roof reflections are clearly recognizable. The same details are not visible in Figure 155 (EM, RNC-off), where only a large spot appears in the middle, possibly misleading the sound propagation analysis. Figure 158 (HSA, RNC-on) shows even more precisely the cancellation effectiveness: it is possible to identify the position from which the largest residual energy contribution comes. In contrast, Figure 156 (EM, RNC-on) is not accurate: energy is spread all over the whole map, which does not provide useful information.

3.3. Application example – Underwater probe

To evaluate the beamforming capabilities of the new underwater array, a test dive has been performed, mounting the system on a tripod and placing it on the seabed at a depth of $23\,m$. Then, three different types of recordings have been taken and processed. A correction of the vertical offset has been applied to the background image, as the panoramic camera is mounted over the hydrophones, introducing an offset of $194\,mm$ between the acoustic and optical centres. A radius of $4\,m$ has been set for the correction. The result is shown in Figure 159.

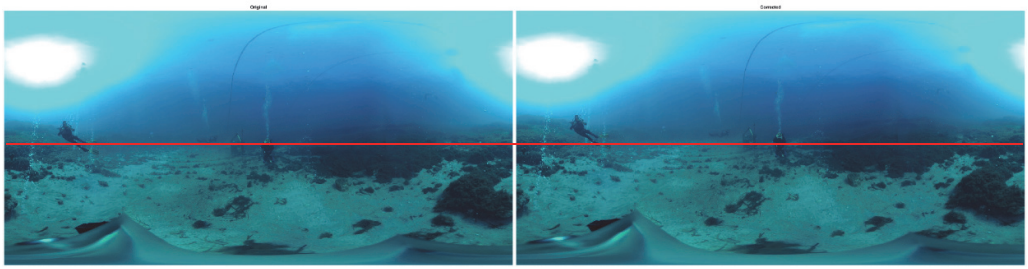


Figure 159: Correction of the offset between acoustic and optical centres

During the first test, three divers have positioned themselves around the system and have emitted with the mouth a constant noise similar to a bellow, one at a time. In Figure 160, it is possible to see the PSD of this type of noise, which exhibits three tonal components at $258\,Hz$, $516\,Hz$ and $774\,Hz$.

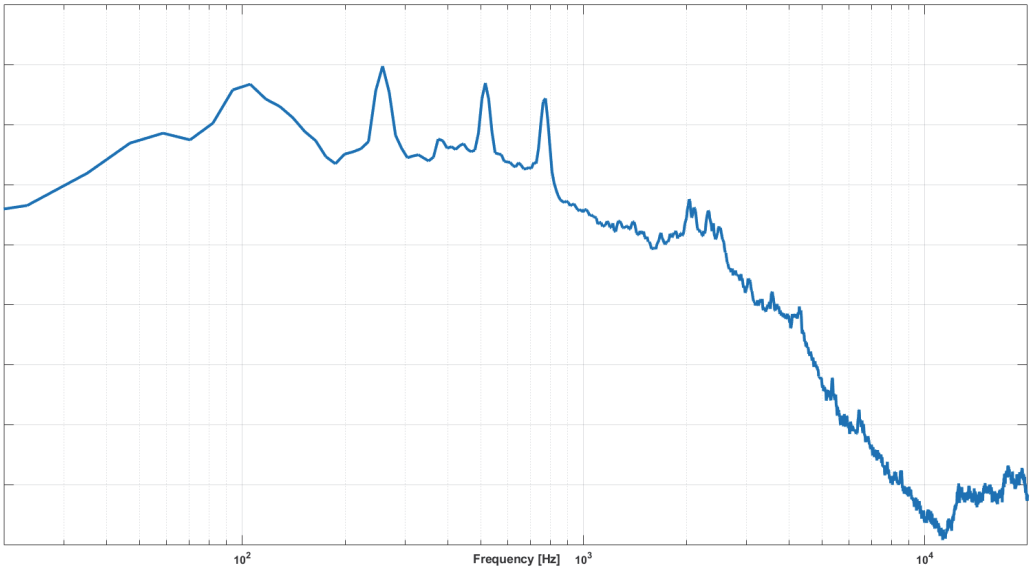


Figure 160: PSD of the underwater test noise emitted by divers with the mouth

The recording has been converted in Ambisonics 1st order and three colour maps have been produced, one for each position of the divers: in front of the array (Figure 161), on the left (Figure 162) and on the right (Figure 163). Colour maps are band-pass filtered in the frequency range $500\,Hz - 800\,Hz$, to include two peaks of the

noise signal. The method employed to generate the colour maps is in this case IV, which does not allow to plot calibrated SPL values but provided the best result. The visualization is normalized in the range $0 - 1$ and the localization of the sources is excellent.

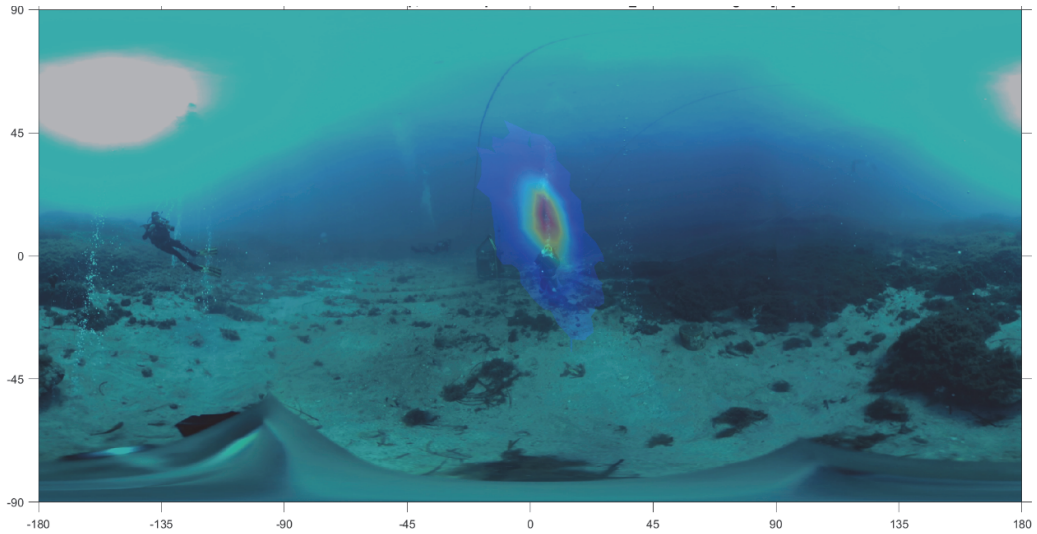


Figure 161: Localization of the diver in front of the underwater array

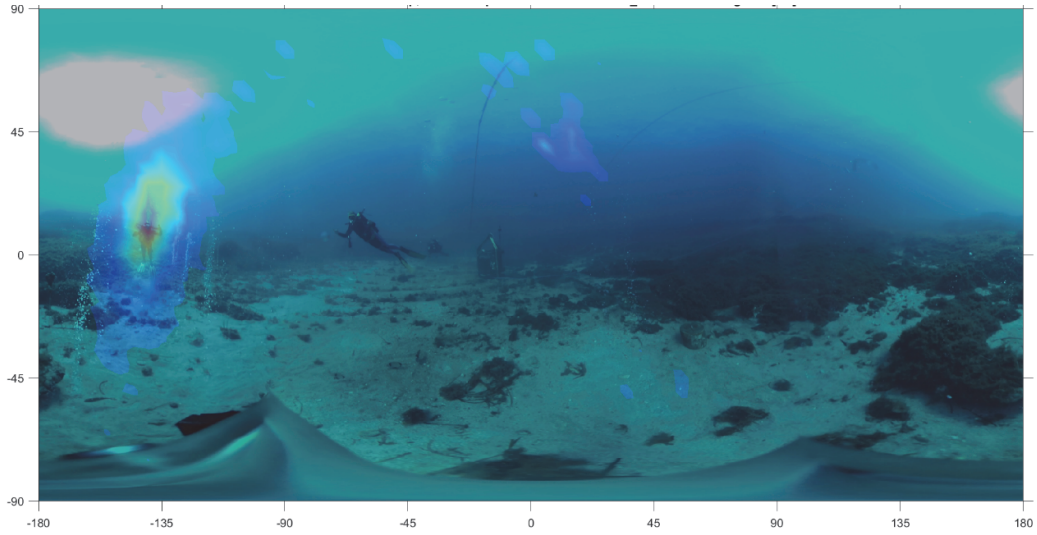


Figure 162: Localization of the diver on the left of the underwater array

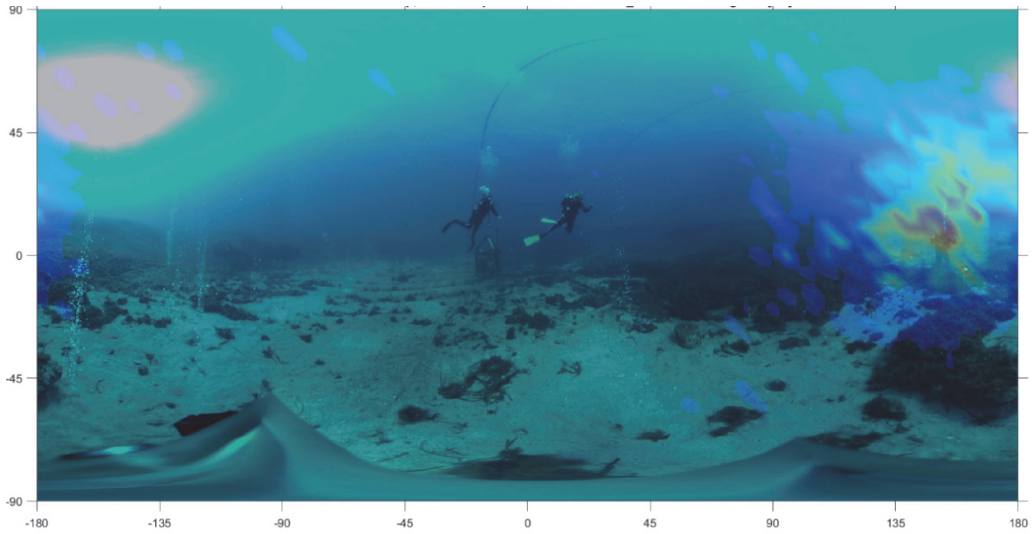


Figure 163: Localization of the diver on the right of the underwater array

The second test has been performed with an impulsive source: one of the diver tapped a small knife against a metal structure located about 10 meters away in front of the array. The spectrum of the signal is in this case almost flat, as shown in Figure 164.

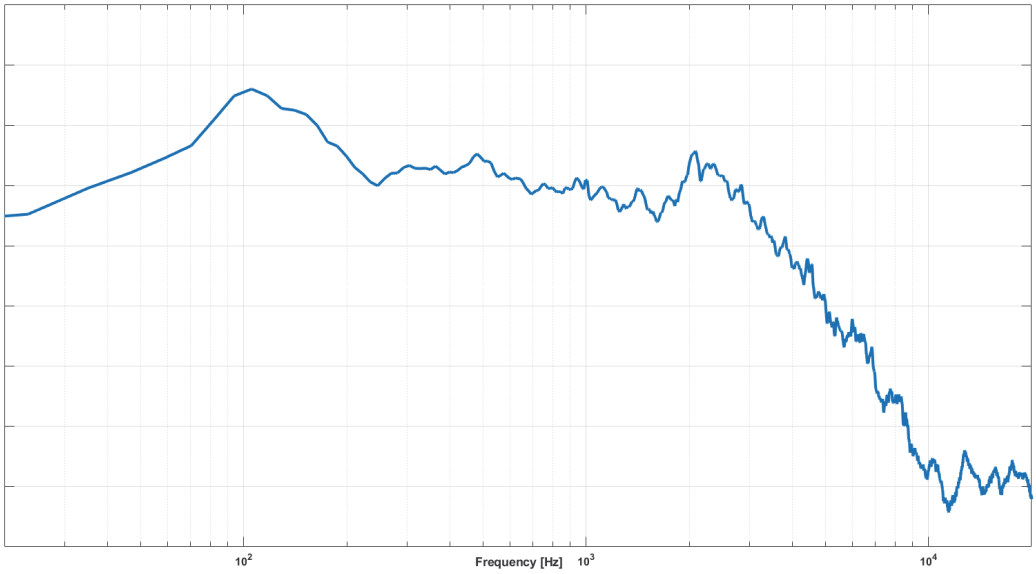


Figure 164: PSD of the underwater impulsive test noise

A colour map filtered in the octave band centred at 1 kHz has been calculated (Figure 165). The localization is excellent also in this case.

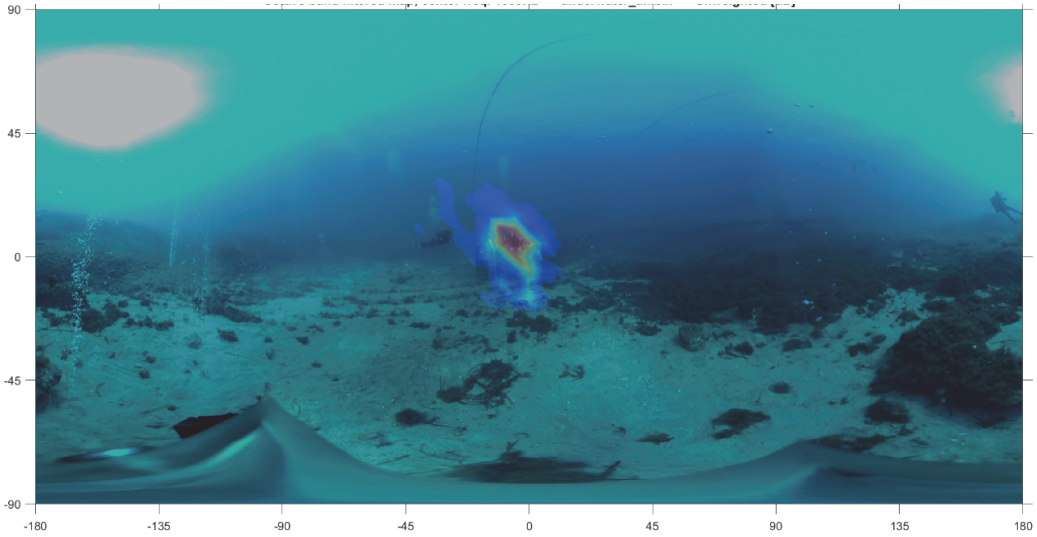


Figure 165: Localization of an impulsive test signal

For the last test, the divers have moved far away from the array to reduce the noise disturbance of the breathing systems and a recording of the background noise has been taken. As it is possible to see also in the previous colour maps, there are columns of bubbles, which is almost pure CO_2, that come out of the seabed. This phenomenon, caused by the underlying volcanic activity, can be observed only in this place, the sea facing the island of Panarea (Messina, Italy). The final goal of the experiment consists in estimating the amount of CO_2 coming out in a certain period, employing the correlation with the noise emitted by the coalescent bubbles. In this preliminary test, the purpose is limited to demonstrate that the new array of hydrophones would be able to perform this monitoring better than a traditional hydrophone.

The PSD of a recording of the background noise is shown in Figure 166.

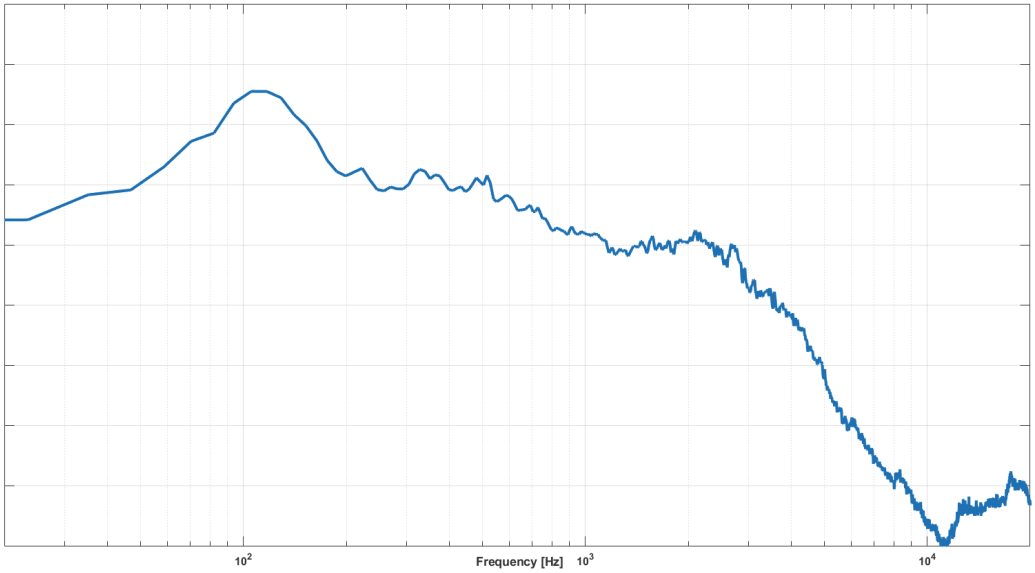


Figure 166: PSD of the underwater background noise caused by CO_2 bubbles

Two colour maps have been produced. The first map (Figure 167) is calculated with PWD method and by filtering the signal in the whole frequency range where beamforming is working properly. Accordingly to the results provided by FEM simulations and described in 2.4.1, a band-pass filter has been defined in the frequency range $300\ Hz - 3\ kHz$. It appear that most of the noise comes from below, which is clearly the expected result.

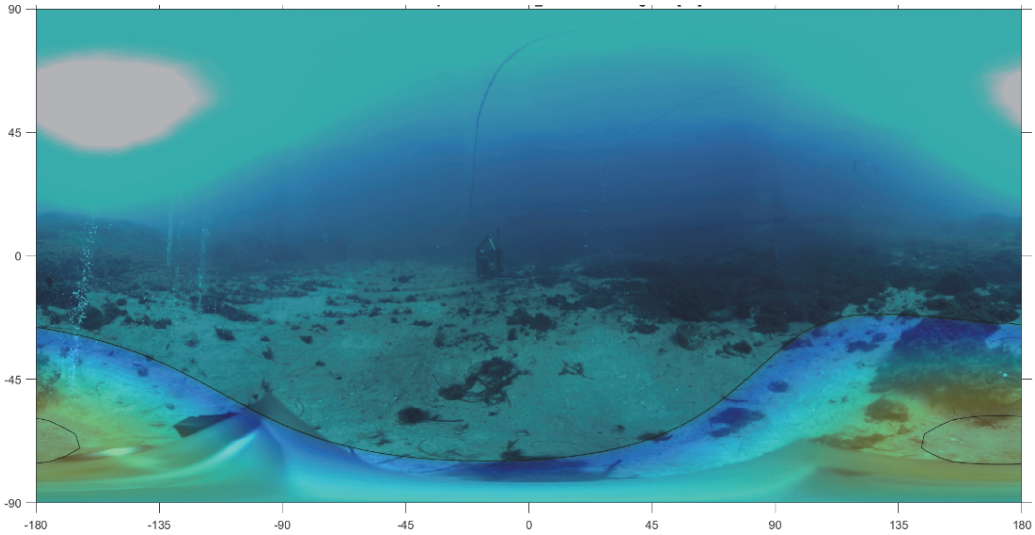


Figure 167: Localization of underwater background noise

The second map (Figure 168) instead has been calculated with IV method in the frequency range $1\ kHz - 3\ kHz$. One can note that the column of bubbles is successfully localized.

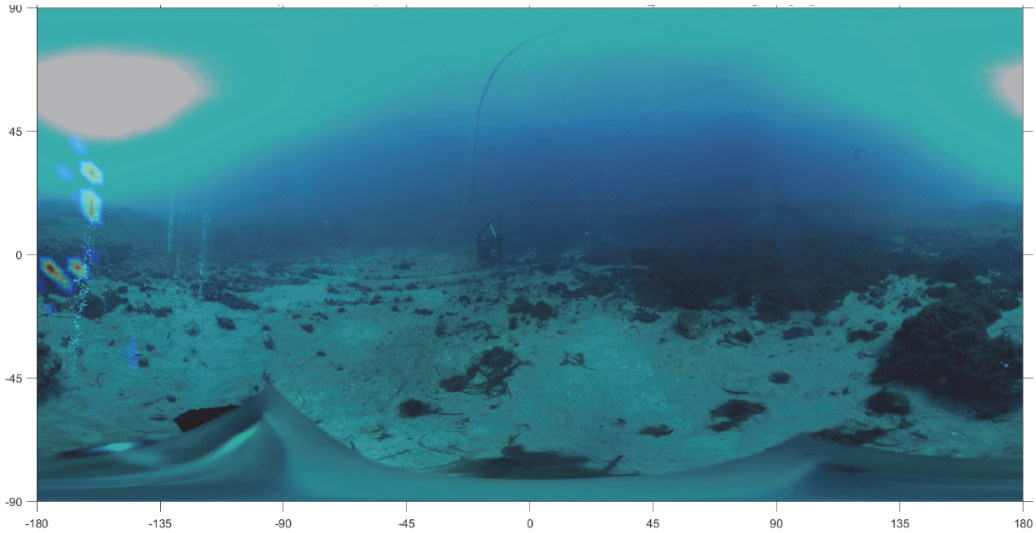


Figure 168: Localization of CO_2 bubbles coming out from the seabed of island of Panarea (Italy)

4. Spatial information reproduction

The capability to reproduce correctly the spatial information is at least important as the capability of making scientific analysis. The range of applications in this case is much larger: virtual and augmented reality entertainment, enhanced telepresence and remote assistance, customer demonstrations, immersive validation tests. However, the development of this work has looked primarily to the last two possibilities.

On one side, demonstrations: customers, for example, can try the experience offered by a new car even from home, thanks to the VR technology. On the other side, test sessions: NVH or audio system engineers of carmaker R&D centres can wear a HMD and perform tests from their office, without the need to take a car in an anechoic garage and troubling the test team. ANC systems or tuning of a sound system can be verified in a way that goes beyond mere quantitative assessment, adding a subjective assessment in a condition very close to the reality.

Of course, for these applications it is of extremely importance to keep the sound field as much as possible identical to the original one: the listening experience cannot be altered in any way, otherwise it lose any validity. In other fields, there is more tolerance: different equalizations, small distortions or other alterations of the sound quality in general are not a problem for telepresence conversation or videogames.

There are substantially two ways to reproduce the sound field, loudspeakers arrays and headphones with binaural processing. In the following, the first solution will be briefly presented, as the work has been mainly done for the second one.

4.1. Loudspeakers arrays

The three dimensional reconstruction of the sound field can rely again on Ambisonics theory, SPS or Wave Field Synthesis (WFS) [39].

The playback systems for Ambisonics and SPS reproduction require a relatively small amount of loudspeakers, most of the times positioned in a spherical geometry, even if it is possible to decode Ambisonics audio over an irregular geometry [40], [41]. WFS synthesis instead requires a very high number of loudspeakers, normally arranged on a horizontal line, in a circle or a square [42], [43].

Some practical examples of Ambisonics and WFS reproduction systems are referenced in [44], [45], [46], [47] and some of them are shown in Figure 169.

Figure 169: ISVR Ambisonics rig, sphere of 40 loudspeakers (above, left), WFS system in Parma, square of 189 loudspeakers (above, right) and an Ambisonics system in Parma, ring of 8 plus cube of 8 loudspeakers and stereo dipole (below)

The most important aspect of the reproduction processing is the decoding of the spatial format over the loudspeaker array. The o output signals for an arrangement of loudspeakers, which are called Speaker Feeds (SF), are obtained by convolving the n input signals of the B or P format with a decoding matrix $\|D\|_{n \times f}$, as follow:

$$sf_o(t) = \sum_{i=1}^{n} y_i(t) * \|D(t)\|_{i \times o}. \quad (33)$$

The decoding matrix can be obtained with a measurement, again employing a microphone array placed in the "sweet spot", which is the place where the head of the listeners must be positioned during the listening session, or alternatively it can be calculated via software. Some decoders the author tested are referenced in [33], [48], [49]. The decoding process for WFS is not covered in this text.

4.2. Headphones reproduction

The most common devices to reproduce the spatial sound field over headphones are visors and tethered HMDs, such as the ones shown in Figure 170. The firsts are portable, wireless, battery powered and relatively low-cost solutions. Of course, their capabilities are limited respect to the others, as the computational power is much lower. Conversely, the second ones are tethered, not portable and more expensive, but much more performing.

Figure 170: A wireless visor (left) and a tethered HMD (right)

The principal disadvantage of the tethered HMDs is that they require a powerful desktop computer with a top-level graphic card, in addition to the presence of the wire, which can limit the movements. They provide video with a refresh rate up to 90 fps, reducing considerably the possibility to suffer of motion sickness, and spatial audio up to 16 channel, which is Ambisonics third order. Currently, visors are instead stuck to a refresh rate of 60 fps (sometimes 72 fps but advanced rendering technique must be employed, such as Fixed Foveated Rendering, to reduce the computational load) and spatial audio with eight channels. There are two different formats of spatial audio with eight channels: MACH1 [50], which is substantially a SPS-8 and TBE, a quasi-second order Ambisonics, where the 6th SH (with ACN numbering) is omitted, therefore reducing the vertical spatialization.

These systems, a part of rendering the video, are fundamental for the spatial audio processing because they are equipped with an accelerometer, which is able to detect the rotations of the head, defined by the set of angles Yaw, Pith and Roll (Y-P-R) or with a quaternion. This is the key information that permits to apply a counter rotation to the encoded audio format, so that head rotations are compensated. This is also one of the reason for which encoded formats are adopted: a part from being capable of describing the spatial information with a relatively small number of signals, in their domain it is quite easy to manipulate the sound field. The counter-rotations matrix must be applied a large number of times each second: in this way, sources do not move with the head, but keep their positions, as it happens in the real world. If a source is perceived in front of us, when turning the head to the right the source will be perceived to our left.

From an acoustic point of view, the main drawback of visors and HMDs of Figure 170 is the presence of inadequate headphones. Most of the time they are of low quality, open ear, hence unable to isolate the user from the outer noise and it is not possible to remove them for replacement. Therefore, a solution without video could be useful, i.e. providing a normal pair of headphones of a head-tracking system. In this way, it would be possible to employ high quality, closed ears headphones, eventually with an Active Noise Reduction system. Of course, in this case an additional software, such a smartphone app, would be required to receive and process the quaternion given by the head-tracker.

4.2.1. The HRTFs

The processing required for headphones reproduction is based on the usage of a set of HRTFs [51].

To measure the HRTFs, a dummy head provided with microphones placed inside the ear canal, such as Neumann KU-100 (Figure 171, right) or Bruel&Kjaer 4128 (Figure 171, left), must be placed in the middle of a loudspeakers system. Then, the impulse response of each loudspeaker is measured with the dummy head. For a system of M loudspeakers, the matrix Speaker Feed to Binaural (SF2BIN) is obtained: $\|SF2BIN\|_{M\times 2}$.

Figure 171: Dummy heads, Neumann KU-100 (left) and B&K type 4128 (right)

Then a decoding matrix has to be either measured or calculated, so that it is possible to convert the SH into the SF. For an Ambisonics format of B signals, it is obtained the matrix: $\|SH2SF\|_{B\times M}$.

By convolving the two matrices, it is possible to realize the binaural reproduction: SH are decoded over the two ears through a system of virtual loudspeakers, which

corresponds to the real system of M loudspeakers where HRTFs have been measured. Hence, it is:

$$\|SH2BIN\|_{B\times 2} = \|SH2SF\|_{B\times M} \cdot \|SF2BIN\|_{M\times 2}. \quad (34)$$

Note that the computational cost for the visor depends only on the number of signals of the B-format. In fact, the decoding, the measurement and the convolution have to be performed only once and this operation is not involved in the real-time processing. Thus, it is desirable to have the highest number of loudspeakers possible for the measurement of the HRTFs, resulting in a better quality of the binaural reproduction as it will be more difficult to perceive the position of the virtual loudspeakers, increasing the naturalness of listening.

4.2.2. Headphones equalization

The headphones equalization is fundamental to ensure the best quality of the reproduction and, more important, to avoid any tonal changes.

The playback system of several visors and HMDs have been measured with a Neumann KU-100 dummy head (Figure 172), pointing out that none of these solutions is equipped with a satisfactory pair of transducers, that is a flat frequency response, sufficiently extended downward. As an example, in Figure 173 the frequency response of three systems is presented: Oculus Go (red), Samsung Odyssey (blue) and Sennheiser HD 4.30 (black). It can be seen that both visors are not capable of reproducing sounds below $100\ Hz$, as Sennheiser headphones do, which also have a flatter frequency response.

Figure 172: Measurement of the headphones of a HMD with a dummy head

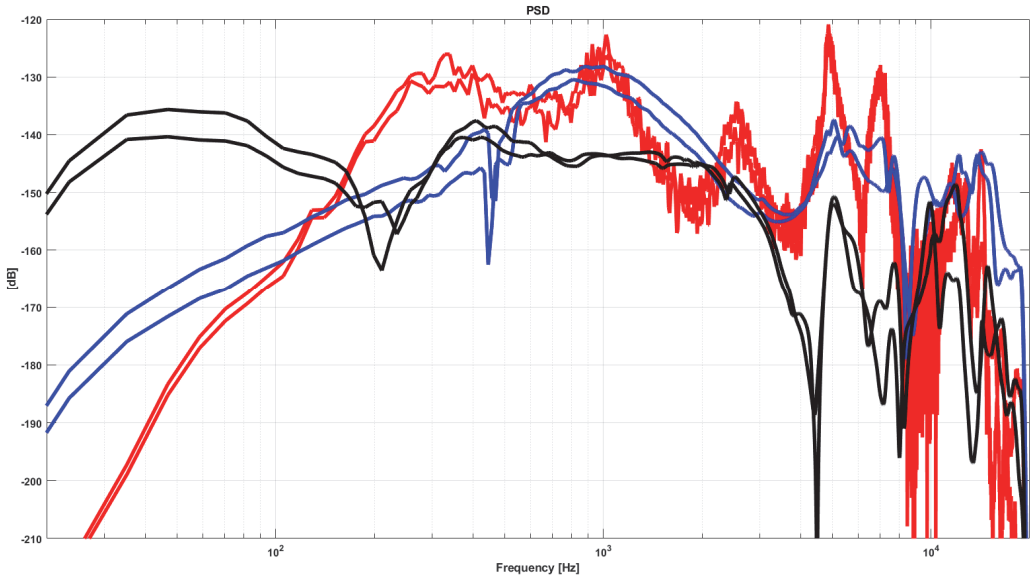


Figure 173: Headphones frequency responses measured with Neumann KU-100, Oculus Go (red), Samsung Odyssey (blue) and Sennheiser HD 4.30 (black)

Inverse filters have been produced for all the systems measured (Figure 174), by means of Kirkeby inversion. By convolving the inverse filters of the headphones with the matrix $\|SH2BIN\|_{B\times 2}$, an equalized binaural filtering matrix is obtained, $\|SH2BIN_{eq}\|_{B\times 2}$.

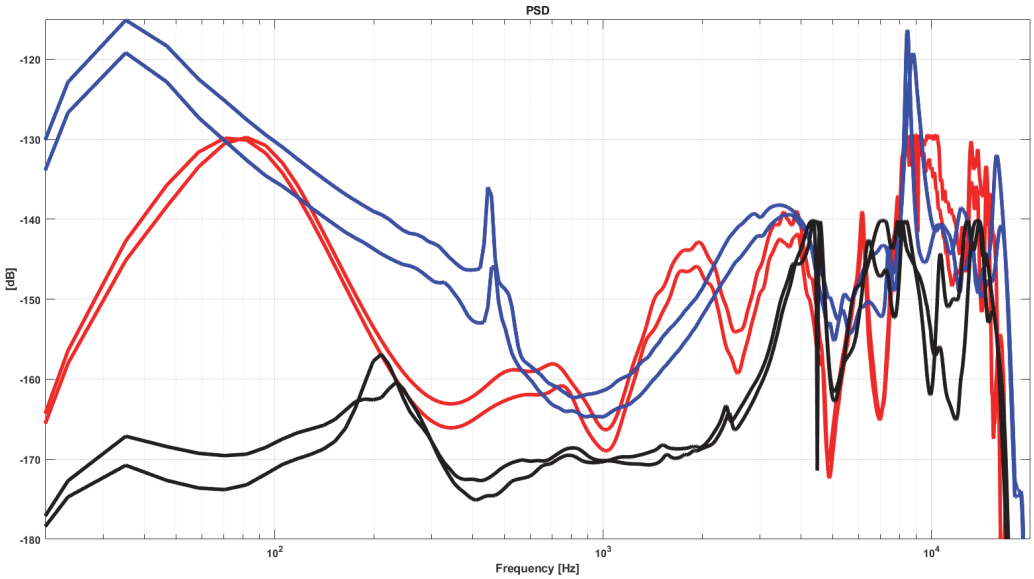


Figure 174: Headphones inverse filters, Oculus Go (red), Samsung Odyssey (blue), Sennheiser HD 4.30 (black)

To conclude, it is pointed out a last problem caused by the usage of contact earphones as the ones shown in Figure 170 or small closed headphones: every time the headphones are worn differently respect to the measurement (i.e. ear fin folded inside or flattened), the equalization and the binaural processing will result worsened or corrupted. For this reason, large closed headphones that allows the ear fins to remain always in the same position are the best choice.

4.2.3. Individualized HRTFs measurement

The usage of individualized HRTFs is widely discussed. It is obvious that a dummy head avoid the needs to measure each individual, which would be clearly not possible. However, the spatialization perceived by each person when using generalized HRTFs is always different and for some people is almost null. This happens because there are many factors occurring together that modify our HRTFs: the dimension and the shape of the head, the dimension and the shape of the ears but also the structure of the torso and the height of the neck, which are responsible of the height perception.

This variance in the perception of the spatial sound field brings some disadvantages, first among all the impossibility to ensure for all people the same experience, as i.e. studied in [52]. In addition, only with individualized HRTFs set it can be reproduced the natural perception of really being inside the original sound field.

For these reasons, several methods have been developed to retrieve an individualized HRTFs set without the classic measurement process [53], [54], [55]. Most of the solutions are based on remote computing. The user is asked to send some pictures of the head and ear fins, and then two possibilities are available: main dimensions of the individual are calculated through image recognition systems and the best matching HRTFs set is selected from a database or a specific set of HRTFs is simulated for that individual [56], [57], [58], [59], [60].

All these methods have the great advantage of being applicable to the large market, but none of them can reach the quality of a measurement based individualised set. Moreover, the pre-equalization of a headset, which is done employing a dummy head, would be made partially ineffective. In fact, it would be lost the matching between the system used for measuring the inverse filters of the headphones and the HRTFs themselves, having individualised only the second one. In addition, a good measurement of HRTFs set is not trivial: a treated, or better, an anechoic room equipped with a loudspeakers rig is needed, together with a special pair of in-ear microphones.

Considering the relatively small amount of people requiring the individualization of HRTFs for the purposes explained previously (internal R&D for subjective tests) and the needs to ensure the best quality possible of the whole process, measured HRTFs have been chosen.

The measurements have been performed in the Ambisonics room shown in Figure 169 (below) which is equipped with 16 loudspeakers arranged over a sphere, half of them forming a horizontal ring and the other half positioned in the vertices of a cube. This placement entails only three different levels of height: the vertices of the cube are at the top and bottom levels, while the ring is in the middle plane. Hence, the third order Ambisonics cannot be completely reconstructed: the 15^{th} SH (with ACN counting) in fact requires four different levels of height for the reconstruction. Two more loudspeakers (Genelec S30D), which are usually employed for a Stereo Dipole system, have been used as subwoofers.

All loudspeakers are pre-equalized: a Bruel&Kjaer omnidirectional microphone (B&K type 4189) is placed in the sweet spot and each loudspeaker is measured with an Exponential Sine Sweep (ESS) [61]. Inverse filters are computed by means of Kirkeby formula (4) and then convolved with the ESS. In this way, a pre-equalized ESS is obtained for each loudspeaker: the delays caused by misalignments and positioning errors are compensated and their frequency responses are made almost perfectly flat.

The HRTFs are measured with a pair of innovative in-ear, open canal microphones, the Exofield (Figure 175), [62]: a couple of MEMS transducers with flat frequency response, small enough to be placed inside the ear canal without obstructing it. The individual whose HRTFs are being measured is seated in the sweet spot wearing the Exofield and the test signals (pre-equalized ESS) are played by the loudspeakers, one after the other, overlapped to reduce the measurement time [63]. This operation provides the matrix $\|SF2BIN\|_{M\times 2}$.

Then, the individual wears the headphones or headset to equalize, keeping the Exofield microphone: the ESS is played by each headphone and recorded by the corresponding microphone. Inverse filters are calculated with (4) and applied to the matrix $\|SF2BIN\|_{M\times 2}$, so that $\left\|SF2BIN_{eq}\right\|_{M\times 2}$ is obtained. Finally, the decoding matrix $\|SH2SF\|_{B\times M}$ is convolved with $\left\|SF2BIN_{eq}\right\|_{M\times 2}$ to get the $\left\|SH2BIN_{eq}\right\|_{B\times 2}$, as in (30).

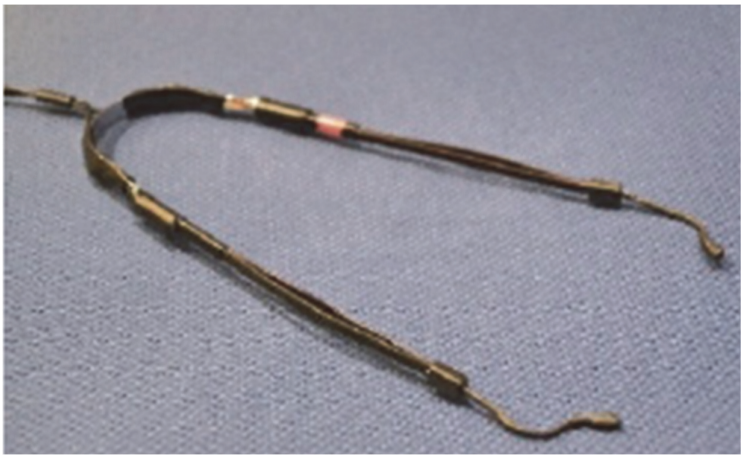


Figure 175: Exofield microphone for HRTFs measurement

4.2.4. Individualized HRTFs implementation for VR reproduction

All the available software capable of reproducing panoramic video with spatial audio on visors and HMDs make use of the binaural approach: B-format is convolved with a built-in matrix $\|SH2BIN\|_{B\times 2}$. Nevertheless, none of them seems to provide

the possibility to switch between different HRTFs set. Two solutions have been developed to solve this problem.

An existing VR video player for HMDs, Vive Cinema, already supporting Ambisonics format up to third order, has been modified [64]. True stereo convolution has been added: most of existing software employ a mono convolution, i.e. right ear, which is then mirrored to get the other side for saving computational power. However, this operation relies on the hypothesis of symmetry of HRTFs, which can be not true for some individuals. Moreover, the possibility to switch between different HRTFs during the reproduction has been introduced, making it possible to compare quickly many sets and chose the most suitable one. All HRTFs are stored in a folder as multi-channel .wav files, $48\ kHz$, $32\ bit$, each channel containing left and right filters in sequence. Filters of different length are supported.

The second solution is for portable visors. A VR video player has been coded with Unity3D and deployed for Android, including Gear VR and Oculus Go API, so that the app can be installed also on smartphones and used with the Samsung Gear headset. The Resonance Audio SDK, developed by Google, has been employed, as it supports Ambisonics up to third order [65]. A solution to change filters coefficients has been developed, so that individualized HRTFs, in case with headphone equalization, can be employed in place of the standard one, which are known as SADIE (Spatial Audio for Domestic Interactive Entertainment) binaural filters, a set of HRTFs measured at the University of York with Neumann KU-100 [66].

Both solutions provided a very effective improvement of the listening experience and are still available for anyone who wants to test individualized HRTFs.

5. Conclusions and future developments

A methodology for a complete study of spatial sound field has been presented, from the measurement to the reproduction, through the analysis. Techniques for designing and characterising microphones arrays have been discussed with particular attention to simulations. Finite Elements Method revealed to be a powerful tool for calculating the response of microphones and hydrophones arrays, whose measurement is often time-consuming, expensive, affected by mechanical and electrical difficulties or electronic limitations and requires special equipment and spaces. A metrics for analysing Spatial PCM Sampling format performances has been coded.

Two prototypes have been designed, built and successfully employed. Filtering matrices obtained with FEM simulations have been tested and used in the field. The new Head-Shaped Array showed good beamforming performance in the low frequency range and allowed to measure and evaluate properly the effectiveness of two Active Noise Control systems for Road Noise Cancelling and Engine Noise Cancelling. The hydrophones probe, which can integrate a panoramic video recording system, proved to be effective and it will be employed in the future projects of underwater noise monitoring, providing reliable results.

A complete suite of software for acoustics holography has been coded to analyse noise recordings, employing Ambisonics and SPS formats. The software can handle also array calibration, cross-correlation filtering and generation of static and dynamic colour maps with picture or video for the background. All these functions have been employed in several fields, from the evaluation of spatial performances of ANC systems inside car cockpits to objective comparison of different microphones arrays. Cross-correlation processing has been used to produce enhanced colour maps for more advanced analysis of the spatial sound field, as well as panoramic videos with spatial audio and superimposed colour map have been created for virtual reality reproduction.

Finally, a complete solution has been implemented to reproduce faithfully the spatial information of sound field making use of the Virtual Reality technology. A fast method to measure individualized HRTFs and equalization filters for headphones has been developed, together with two software for playing panoramic video and Ambisonics spatial audio over visors and HMDs with individualized HRTFs. These systems have been effectively employed to make subjective listening tests and demonstrations.

Despite the powerfulness of these tools, a lot of work is still required to make them as easy to use, affordable and reliable as the professional market requires. On one side, 3D printing helped in reducing the cost of the arrays but on the other side the electronic circuitry still represents a bottleneck. The panoramic video recording system, although it has exhibited very good performance, revealed to be impractical: the stitching procedure is quite slow and a faster solution is desirable.

The next step of FEM simulations is the introduction of spherical radiation, which takes into account the real curvature of the sound wave. It would be a great improvement, as the hypothesis of plane waves is not respected when microphone arrays are employed inside small environments, like the passenger compartment of a car.

Currently, the sound colour mapping software is not working real-time, resulting in long post-processing time and limiting its usage to academic research. The priority future development is certainly a rewriting of the code in a low-level programming language: a VST3 plug-in supported by most Digital Audio Workstation would be the preferable solution.

6. Bibliography

[1] H. F. Olson, "A History of High Quality Studio Microphones," *AES Convention 55,* 1976.

[2] Technical Writer AES Staff, "Stereophonic Recording Techniques: Old Challenges, New Approaches," *JAES,* vol. 54, no. 3, pp. 225-229, 2006.

[3] C. Ceoen, "Comparative Stereophonic Listening Tests," *AES Convention 41,* 1971.

[4] E. R. Madsen, "The Application of Velocity Microphones to Stereophonic Recording," *JAES,* vol. 5, no. 2, pp. 79-85, 1957.

[5] M. A. Gerzon, "The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound," *50th AES Convention,* 1975.

[6] P. Fellfett, "Ambisonics. Part one: general system description," *Studio Sound,* vol. 17, no. 10, p. 60, 1975.

[7] M. A. Gerzon, "Ambisonics. Part two: Studio techniques," *Studio Sound,* vol. 17, no. 10, p. 60, 1975.

[8] A. Farina, "Explicit formulas for High Order Ambisonics," 2017. [Online]. Available: http://www.angelofarina.it/Aurora/HOA_explicit_formulas.htm.

[9] C. Nachbar, F. Zotter and E. Deleflie, "Ambix - A suggested Ambisonics format," *Ambisonics Simposium,* 2011.

[10] A. Farina, A. Amendola, L. Chiesi, A. Capra and S. Campanini, "Spatial PCM Sampling: A New Method for Sound Recording and Playback," *52nd International Conference: Sound Field Control - Engineering and Perception,* 2013.

[11] A. Politis, "Acoustical Spherical Array Processing Library," Department of Signal Processing and Acoustics, Aalto University, Finland, 2016. [Online]. Available: http://research.spa.aalto.fi/projects/spharrayproc-lib/spharrayproc.html#59.

[12] S. Moreau, J. Daniel and S. Bertet, "3D sound field recording with higher order ambisonics-objective measurements and validation of spherical microphone," *120th AES Convention,* 2006.

[13] C. Jin, N. Epain and A. Parthy, "Design, optimization and evaluation of a dual-radius spherical microphone array," *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* vol. 22, no. 1, pp. 193-204, 2014.

[14] A. Farina, A. Capra, L. Chiesi and L. Scopece, "A Spherical Microphone Array for Synthesizing Virtual Directive Microphones in Live Broadcasting and in Post Production," *40th International Conference: Spatial Audio: Sense the Sound of Space,* 2010.

[15] O. Kirkeby and P. A. Nelson, "Digital Filter Design for Inversion Problems in Sound Reproduction," *Journal of Audio Engineering Society,* vol. 47, no. 7/8, pp. 583-595, 1999.

[16] J. W. S. Rayleigh, The Theory of Sound - Volume II, 1945.

[17] A. N. Tikhonov, A. Goncharsky, V. V. Stepanov and A. G. Yagola, Numerical Methods for the Solution of Ill-Posed Problems, Springer, 1995.

[18] B. Bernschutz, C. Porschmann, S. Spors and S. Weinzierl, "Soft-limiting der modalen amplitudenverstarkung bei spharischen mikrofonarrays im plane wave decomposition verfahren," *Proceedings of the 37th Deutsche Jahrestagung fur Akustik (DAGA 2011),* 2011.

[19] L. Chiesi, "Microphone and loudspeaker arrays for room acoustic measurements," 2016. [Online]. Available: http://hdl.handle.net/1889/3196.

[20] R. H. Hardin and N. J. A. Sloane, "McLaren's Improved Snub Cube and Other New Spherical Designs in Three Dimensions," *Discrete and Computational Geometry,* vol. 15, pp. 429-441, 1996.

[21] S. L. Marple, "Computing the Discrete-Time Analytic Signal via FFT," *IEEE Transactions on Signal Processing,* vol. 47, pp. 2600-2603, 1999.

[22] L. Scopece and A. Farina, "Sistema 3DVMS: da dove arriviamo...dove andiamo," *Elettronica e Telecomunicazioni,* vol. 2, pp. 12-19, 2015.

[23] A. Farina, D. Pinardi, M. Binelli, M. Ebri and L. Ebri, "Virtual reality for subjective assessment of sound quality in cars," *144th AES Convention,* 2018.

[24] R. Boaz, Fundamentals of Spherical Array Processing, Springer-Verlag Berlin Heidelberg, 2015.

[25] A. Farina, E. Armelloni and L. Chiesi, "Experimental Evaluation Of The Performances Of A New Pressure-Velocity 3D Probe Based On The Ambisonics Theory," *4th international conference and exhibition on Underwater Acoustic Measurements: Technologies and Results,* 2011.

[26] A. Farina, "Environmental Impact of Underwater Noise: sound pressure and particle velocity," *International meet on sonar systems-sensors,* 2018.

[27] A. Farina, E. Armelloni and D. Pinardi, "The Forgotten Measurement: Sound Pressure and Particle Velocity," *Environment Coastal & Offshore Magazine - Special Issue on Ocean Sound,* 2019.

[28] E. G. Williams, Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography, Academic Press, 1999.

[29] T. Mattioli, A. Farina, E. Armelloni, P. Hameau and M. Díaz-Andreu, "Echoing landscapes: Echolocation and the placement of rock art in the Central Mediterranean," *Journal of Archaeological Science,* vol. 83, pp. 12-25, 2017.

[30] L. Tronchin, A. Venturi, A. Farina and A. Amendola, "Implementing a spherical microphone array to determine 3D sound propagation in the 'Teatro 1763' in Bologna, Italy," *International Symposium on Room Acoustics,* 2013.

[31] M. Binelli, A. Venturi, A. Amendola and A. Farina, "Experimental analysis of spatial properties of the sound field inside a car employing a spherical microphone array," *1330th AES Conference,* 2011.

[32] Blue Ripple Sound, "O3A Core," [Online]. Available: https://www.blueripplesound.com/products/o3a-core.

[33] L. McCormack, “Spatial Audio Real-time Applications (SPARTA),” Aalto University, [Online]. Available: http://research.spa.aalto.fi/projects/sparta_vsts/.

[34] L. McCormack, S. Delikaris-Manias, A. Farina, D. Pinardi and P. Ville, “Real-time conversion of sensor array signals into spherical harmonic signals with applications to spatially localised sub-band sound-field analysis,” *144th AES Convention,* 2018.

[35] A. Farina, M. Binelli, A. Capra, E. Armelloni, S. Campanini and A. Amendola, “Recording, Simulation and Reproduction of Spatial Soundfields by Spatial PCM Sampling (SPS),” *International Seminar on Virtual Acoustics,* 2011.

[36] A. Farina and L. Tronchin, “3D Sound Characterisation in Theatres Employing Microphone Arrays,” *Acta Acustica united with Acustica,* vol. 99, no. 1, pp. 118-125, 2013.

[37] V. Håvard, J. Crowley and G. T. Rocklin, “New Ways of Estimating Frequency Response Functions.,” *Sound and Vibration,* vol. 18, pp. 34-38, 1984.

[38] FFmpeg, “Download,” [Online]. Available: https://ffmpeg.org/download.html.

[39] J. Ahrens, R. Rabenstein and S. Spors, “The Theory of Wave Field Synthesis Revisited,” *124th AES Convention,* 2008.

[40] S. W. Clapp, A. E. Guthrie, J. Braasch and N. Xiang, “Using Ambisonics to Reconstruct Measured Soundfields,” *135th AES Convention,* 2013.

[41] D. Arteaga, “An Ambisonics Decoder for Irregular 3-D Loudspeaker Arrays,” *134th AES Convention,* 2013.

[42] J. Ahrens and S. Spors, “Sound Field Reproduction Using Planar and Linear Arrays of Loudspeakers,” *IEEE Transactions On Audio, Speech, And Language Processing,* vol. 18, no. 8, 2010.

[43] M. M. Boone, “Multi-Actuator Panels (MAPs) as Loudspeaker Arrays for Wave Field Synthesis,” *Journal of the Audio Engineering Society,* vol. 52, no. 7/8, pp. 712-723, 2004.

[44] M. A. Gerzon, “Periphony: With-height Sound Reproduction,” *Journal of the Audio Engineering Society,* vol. 21, pp. 2-10, 1973.

[45] A. Farina and E. Ugolotti, “Subjective comparison between Stereo Dipole and 3D Ambisonics surround systems for automotive applications,” *16th AES Conference,* 1999.

[46] F. Fazi, P. Nelson, J. Christensen and J. Seo, “Surround System Based on Three-Dimensional Sound Field Reconstruction,” *125th AES Convention,* 2008.

[47] A. Capra, A. Farina and F. Grani, ““Urban Sounds”: an acoustical tour of Parma,” *AIA-DAGA International Conference on Acoustics,* 2013.

[48] B. R. Sound, “Rapture3D Advanced Decoder,” [Online]. Available: http://www.blueripplesound.com/products/rapture-3d-advanced.

[49] I. o. E. M. a. Acoustics, “IEM Plug-in Suite,” [Online]. Available: https://plugins.iem.at/download/.

[50] MACH1. [Online]. Available: https://www.mach1.tech/.

[51] M. Binelli, D. Pinardi, T. Nili and A. Farina, “Individualized HRTF for playing VR videos with Ambisonics spatial audio on HMDs,” *AES International Conference on Audio for Virtual and Augmented Reality,* 2018.

[52] D. Poirier-Quinot and B. F. G. Katz, "Impact of HRTF individualization on player performance in a VR shooter game II," *AES Conference on Audio for Virtual and Augmented Reality,* 2018.

[53] A. Kärkkäinen, L. Kärkkäinen and T. Huttunen, "Practical Procedure for Large Scale Personalized Head Related Transfer Function Acquisition," *51st AES International Conference: Loudspeakers and Headphones,* 2013.

[54] F. Shahid, N. Javeri, K. Jain and S. Badhwar, "AI DevOps for Large-Scale HRTF Predition and Evaluation: An End to End Pipeline," *AES International Conference on Audio for Virtual and Augmented Reality,* 2018.

[55] M. Geronazzo, J. Kleimola, E. Sikström, A. de Götzen, S. Serafin and F. Avanzini, "HOBA-VR: HRTF On Demand for Binaural Audio in Immersive Virtual Reality Environments," *144th AES Convention,* 2018.

[56] P. Guillon, R. Nicol and L. Simon, "Head-Related Transfer Functions Reconstruction from Sparse Measurements Considering a Priori Knowledge from Database Analysis: A Pattern Recognition Approach," *125th AES Convention,* 2008.

[57] E. S. Schwenker and G. D. Romigh, "An Evolutionary Algorithm Approach to Customization of Non-Individualized Head Related Transfer Functions," *137th AES Convention,* 2014.

[58] A. Andreopoulou and A. Roginska, "Database Matching of Sparsely Measured Head-Related Transfer Functions," *Journal of the Audio Engineering Society,* vol. 65, no. 7/8, pp. 552-561, 2017.

[59] A. Andreopoulou, A. Roginska and J. P. Bello, "Reduced Representations of HRTF Datasets: A Discriminant Analysis Approach," *135th AES Convention,* 2013.

[60] C. Guezenoc and R. Seguier, "HRTF Individualization: A Survey," *145th AES Convention,* 2018.

[61] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," *108th AES Convention,* 2000.

[62] JVCKENWOOD Corporation, "EXOFIELD® Headphone Technology Replicates the Acoustic Space of a Room," 2018. [Online]. Available: http://pro.jvc.com/pro/pr/2018/ces/JVC_Exofield.html.

[63] P. Balazs, B. Laback and P. Majdak, "Multiple Exponential Sweep Method for Fast Measurement of Head Related Transfer Functions," *122th AES Convention,* 2007.

[64] M. Binelli, "Download," [Online]. Available: www.angelofarina.it/Public/ViveCinema/.

[65] M. Gorzel, A. Allen, I. Kelly, J. Kammerl, A. Gungormusler, H. Yeh and F. Boland, "Efficient Encoding and Decoding of Binaural Sound with Resonance Audio," *AES Conference on Immersive and Interactive Audio,* 2019.

[66] University of York, "SADIE - Spatial Audio for Domestic Interactive Entertainment," 2017. [Online]. Available: https://www.york.ac.uk/sadie-project/GoogleVRSADIE.html.

Acknowledgements

The author would like to express deep gratitude to those who have helped and allowed the work presented in this thesis.

ASK Industries S.p.A. that supported the development and made available effective Active Noise Control systems and the possibility to test them on board of vehicles, with the help of experienced engineers.

Prof. Ville Pulkki and all the people of Aalto University, for their availability and interest in my job.

JVCKENWOOD Corporation for having made available the Exofield technology.

ARGO Tractors S.p.A. who made available the 3D printed parts.

Ed infine, un pensiero ai miei Amici. Quelli con la A maiuscola e di cui non è necessario scrivere il nome. A voi dico grazie. Per esserci da una vita ed esserci stati sempre, anche da lontano, anche da lassù. E soprattutto, per non farmi mai dubitare di perdervi.

Un grande ringraziamento lo rivolgo ai miei colleghi, che mi sopportano da tre anni. In particolare a Marco, da cui ho imparato tanto.

In queste parole, voglio ricordare anche tre persone che mentre scrivo sono due case più in là. Un modello di semplicità, famiglia e affetto.

Ciao ma', ciao pa'. Eccoci qua, alla fine della tesi, un nuovo traguardo raggiunto. Un'occasione per ricordarvi che ogni mio risultato, giorno dopo giorno, arriva da ciò che fate voi, prima di me. Oggi, i complimenti, ve li faccio io.

Il pensiero, il ringraziamento e l'abbraccio più grande sono per Laura. Con te, posso portare avanti la ricerca più bella: essere migliore, ogni giorno.

L'ultima parola è per Angelo, una persona fuori dal comune, a cui va tutta la mia stima ed il mio rispetto.