

Experimental 3D Sound Field Analysis with a Microphone Array

Mathieu Guillaume¹, and Yves Grenier¹

¹*GET - Télécom Paris, Paris, France*

Correspondence should be addressed to Mathieu Guillaume (mathieu.guillaume@enst.fr)

ABSTRACT

This article deals with spatial analysis of real 3D sound fields, and presents an optimal beamforming algorithm minimizing the power of side lobes, used to improve the spatial representation of sound fields. Optimal tap vectors are computed for a set of incidence directions mapping the whole sphere. The robustness of the algorithm to sensor-positioning errors and to sensor noise is tested with real data, acquired using microphone and loudspeaker arrays. It is shown that there is good agreement between the real and the corresponding simulated data regarding the incidence direction. Moreover, early reflections, providing new peaks in the representation depending on the location of secondary sources are also clearly visible.

1. INTRODUCTION

This article deals with the problem of experimental 3D sound field analysis. Sound field analysis is the discipline concerned with the knowledge of the parameters associated to a specific sound field model. Among the different possibilities of sound field descriptions, a vast amount of literature is dedicated to sound field analysis based on spherical harmonics. The articles of Poletti [10] and Daniel [2] constitute excellent reviews in this domain. Besides spherical harmonics, other ways to represent the sound field inside a given region of space exist, such as the Kirchhoff-Helmholtz integral equation [14], or the plane wave decomposition [14].

The objective of experimental sound field analysis is to estimate the parameters of the chosen sound field description from real data acquired by a microphone array. Several articles deal with the estimation of the coefficients of the spherical harmonic decomposition using a spherical microphone array [11, 6, 1, 7]. These articles attempt to orthogonalize the spherical harmonics basis truncated to a specific order observed only at discrete positions located on a sphere of a particular radius. Other strategies use more general array geometries with the same objective in mind [5].

In this article, the sound field description used is the plane-wave decomposition [3, 4]. The equivalence between spherical harmonic and plane wave decompositions has already been investigated earlier [9]. The sound

field analysis scheme described in this article consists in estimating the values of the spatial Fourier transform of the sound field. At a given frequency, the interesting zone to be analyzed in the wave vector domain is the sphere defined by the dispersion relationship [3]. In a previous article [3], we have proposed a beamforming algorithm which uses the nice properties of generalized prolate spheroidal wave sequences [13] to focus the power of the spatial filter in a particular direction of incidence. In this article, we want to test the behavior of this algorithm confronted to real data acquired by a microphone array, and particularly the robustness of the algorithm to slight position errors and to sensor noise.

This article first describes the electroacoustic chain used for data acquisition, as well as the calibration procedures used to correct the time-frequency responses of the electroacoustic devices and to automatically estimate the positions of loudspeakers and microphones. In section 3, the main results of the beamforming algorithm are recalled. This beamforming strategy is then used to compute a set of spatial filters aiming to estimate the spatial Fourier transform of the sound field, thus giving a cartography of it at a specified frequency. In section 4, this sound field analysis scheme is used on real data and is compared to simulated data. Moreover, the performance of this strategy are highlighted by comparing sound field analysis using optimized spatial filters to uniform spatial filters. Finally, some concluding remarks and perspec-

tives are made.

2. MATERIAL DESCRIPTION

2.1. Electroacoustic chain description

The electroacoustic chain used for the sound field measurement was constituted of:

- 16 loudspeakers Tannoy System 600. They have been placed in an anechoic chamber. 12 loudspeakers were mounted on a circular aluminum frame and 4 were laid down on the ground. See figure 1 for more explicit details on the configuration.
- 3 six-channels power amplifiers Yamaha CM 6150.
- 8 omnidirectional microphones Schoeps MK 2, amplified by 4 stereo pre-amplifiers Behringer Ultragain Pro.
- The reference signals sent to the loudspeakers and the recorded signals are synchronized using a 24-track hard disk recorder Tascam MX-2424.

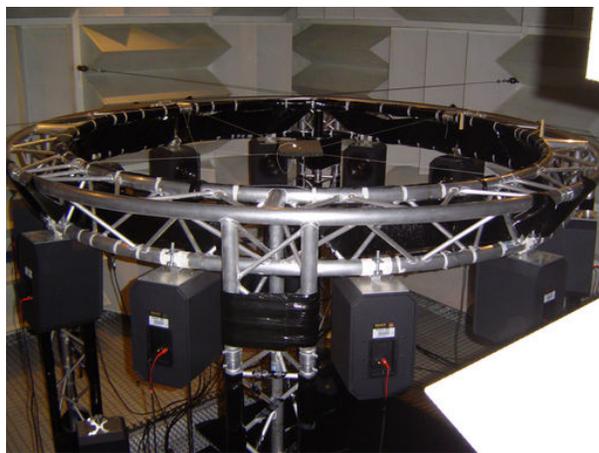


Fig. 1: Loudspeaker array used: circular array of radius 1.1 m with 12 loudspeakers regularly spaced, and 4 other loudspeakers at a different height (partially visible on this photo).

The 8 microphones were combined in a linear logarithmically-spaced microphone array, visible on figure 2. The microphones are located at x-axis reference coordinates [0.025, 0.050, 0.075, 0.119, 0.200, 0.336, 0.565, 0.950] m.

This elementary microphone array was rotated 12 times to generate 8 circular arrays with logarithmically-spaced radii, as visible on figure 3.

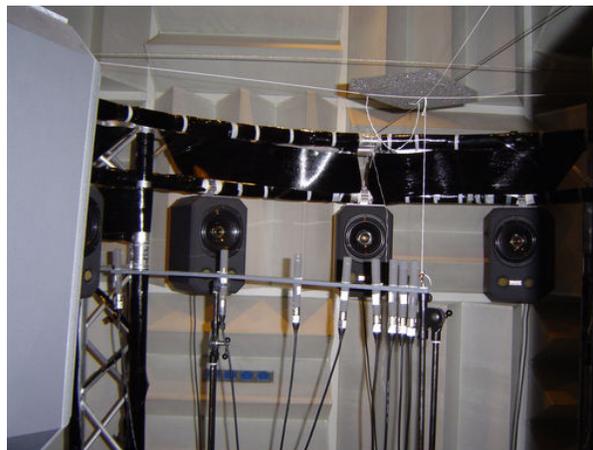


Fig. 2: Microphone sub-array used: linear array of 8 microphones logarithmically spaced from 2.5 cm to 95 cm.

The impulse responses were sequentially measured for each position of the elementary array, between each loudspeaker and each microphone using sweeps signals [8]. They have been measured twice for each position of the elementary antenna to prevent some potential errors due to presence of impulse noise for instance.

Calibration of the electroacoustic chain has been performed. The calibration is not absolute, but relative to one microphone, one pre-amp channel and one input of the hard disk recorder, taken as a reference. Firstly, the eight microphones were used to measure the same impulse response using the same pre-amp and sequencer channel, in order to only correct the frequency response of microphones. It appears that the response of the microphones were identical in the whole frequency band to ± 1 dB. The differences measured at high frequencies were not considered significant because of the difficulty to position the microphones exactly at the same position. Indeed, the small position errors generate differences in the measured impulse responses mainly at high frequencies where the involved wavelengths become comparable to the magnitude of error positions. Secondly, the pre-amplifiers have been calibrated using only one microphone and one input of the hard disk recorder. The frequency responses of the pre-amplifiers were identical except for a global gain. These gain differences occur even if we have taken care of using notched poten-

tiometers with the same graduation. Thirdly, the channels of the hard disk recorder have also been calibrated. They exhibit a perfectly flat response in the range of audio signals, but show different time delays according to the channel used. This has certainly to be linked to internal recording synchronization of the device. Finally, the global calibration consists only in adjusting the time delays and the gains between the different channels: it does not constitute an overwhelming challenge, but the correction deserved to be set up, in order to have coherent signals before applying the sound field analysis algorithm.

2.2. Automatic position calibration

The array processing used in the beamforming algorithm described at section 3 is very sensitive to error positions, so that an automatic position calibration procedure is greatly appreciated. The procedure described in [12] was used to automatically calibrate the position of loudspeakers and microphones. It consists in first estimating the positions of the loudspeakers and microphones from the knowledge of the time of flight matrix by a classical multidimensional scaling process. Then, this first estimation is refined by an iterative Levenberg-Marquardt procedure to compute the maximum-likelihood estimation of the loudspeakers and microphones positions. This automatic position calibration only requires to attach additional microphones in the neighborhood of each loudspeaker. The estimated positions are displayed on figure 3. It confirms that the array deployed in practice is in the main trend conform to the reference one, that is circular arrays with logarithmically-spaced radii. But it also shows slight differences compared to the reference one, useful to be taken into account in order to have good performance for the sound field analysis.

3. SOUND FIELD ANALYSIS

The initial sound field $p(\mathbf{r}, t)$ is sampled by a microphone array in the space domain at the microphone positions located by \mathbf{r}_m , $m \in [1, \dots, M_{\text{mic}}]$. If the microphones are omnidirectional, the sampled sound field could be modeled as [3]:

$$p_{\text{sam}}(\mathbf{r}, t) = p(\mathbf{r}, t) \cdot \sum_{m=1}^{M_{\text{mic}}} \delta(\mathbf{r} - \mathbf{r}_m) \quad (1)$$

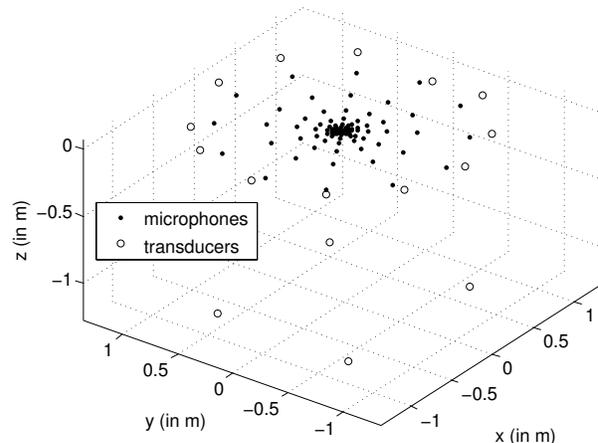


Fig. 3: Result of the automatic position calibration procedure: estimated loudspeakers and microphones positions.

It is more appropriate to work in the frequency domain for the purpose of beamforming, because sound fields satisfy the propagation equation. Thus, they satisfy the dispersion relationship $k = \frac{\omega}{c}$ where k is the wavenumber, norm of the wave vector \mathbf{k} , which is to space what frequency is to time. The sound field to be analyzed is weighted by a tap vector $\underline{w}(\omega) = [w_1(\omega), \dots, w_{M_{\text{mic}}}(\omega)]^T$ to enable a more flexible analysis [3]:

$$p_{\text{ana}}(\mathbf{r}, \omega) = p(\mathbf{r}, \omega) \cdot \sum_{m=1}^{M_{\text{mic}}} w_m(\omega) \delta(\mathbf{r} - \mathbf{r}_m) \quad (2)$$

The spatial Fourier transform of the sound field to be analyzed is given by the following equation:

$$p_{\text{ana}}(\mathbf{k}, \omega) = p(\mathbf{k}, \omega) *_3 \sum_{m=1}^{M_{\text{mic}}} w_m(\omega) e^{-i\mathbf{k} \cdot \mathbf{r}_m} \quad (3)$$

$$\triangleq p(\mathbf{k}, \omega) *_3 h(\mathbf{k}, \omega) \quad (4)$$

where $*_3$ denotes three-dimensional spatial convolution product, and h is the response of the spatial filter.

3.1. Beamforming algorithm

The beamforming algorithm has been presented in details in [3]. A brief summary is given here. The beamforming algorithm aims at focusing the power of the spatial filter $h(\mathbf{k}, \omega)$ in a given direction (ϕ_0, θ_0) where ϕ_0

is the azimuth and θ_0 is the colatitude. Let us state the spatial convolution product of eq. (3):

$$p_{\text{ana}}(\mathbf{k}, \omega) = \iiint_{\mathbf{k}_1 \in \mathbb{R}^3} p(\mathbf{k}_1, \omega) h(\mathbf{k} - \mathbf{k}_1, \omega) d^3 \mathbf{k}_1 \quad (5)$$

We define the output signal of the beamformer steered in the direction (ϕ_0, θ_0) as:

$$\begin{aligned} p_{(\phi_0, \theta_0)}(\omega) &\triangleq p_{\text{ana}}(\mathbf{k}_0, \omega) \\ &= \iiint_{\mathbf{k} \in \mathbb{R}^3} p(\mathbf{k}, \omega) \cdot \\ &\quad h_{(\phi_0, \theta_0)}(\mathbf{k}_0 - \mathbf{k}, \omega) d^3 \mathbf{k} \end{aligned} \quad (6)$$

where $\mathbf{k}_0 = [\frac{\omega}{c}, \phi_0, \theta_0]$ in spherical coordinates. The updated notation of the spatial filter $h_{(\phi_0, \theta_0)}$ indicates that it is dependent on the incidence direction. The area of the wave vector domain which deserves to be studied at pulsation ω for the spatial Fourier transform $p(\mathbf{k}, \omega)$ of the sound field is the sphere defined by the dispersion relationship $|\mathbf{k}| = \frac{\omega}{c}$. The spatial filter $h_{(\phi_0, \theta_0)}$ is optimized so that it focuses its power in the neighborhood of \mathbf{k}_0 compared to the global power of the spatial filter in the neighborhood of the sphere defined by the dispersion relationship. In mathematical terms, this leads to the following optimization procedure:

$$\begin{aligned} \underline{w}(\omega) = & \max_{[w_1, \dots, w_{M_{\text{mic}}}] \in \mathbb{C}^{M_{\text{mic}}}} \\ & \frac{\iiint_{\mathbf{k} \in \mathcal{S}(\mathbf{k}_0, k_{\text{res}})} |h_{(\phi_0, \theta_0)}(\mathbf{k}_0 - \mathbf{k})|^2 d^3 \mathbf{k}}{\iiint_{\mathbf{k} \in C(\mathbf{0}, \frac{\omega}{c} - k_{\text{res}}, \frac{\omega}{c} + k_{\text{res}})} |h_{(\phi_0, \theta_0)}(\mathbf{k}_0 - \mathbf{k})|^2 d^3 \mathbf{k}} \end{aligned} \quad (8)$$

In this equation, $\mathcal{S}(\mathbf{k}_0, k_{\text{res}})$ indicates a sphere of center \mathbf{k}_0 with radius k_{res} , and $C(\mathbf{0}, \frac{\omega}{c} - k_{\text{res}}, \frac{\omega}{c} + k_{\text{res}})$ is the interior of the domain delimited by the two spheres of center $\mathbf{0}$ with radii $\frac{\omega}{c} - k_{\text{res}}$ and $\frac{\omega}{c} + k_{\text{res}}$. Thus, the philosophy of the criterion (8) is that the corresponding spatial filter $h_{(\phi_0, \theta_0)}$ minimizes the influence of potential interference signals located outside the sphere $\mathcal{S}(\mathbf{k}_0, k_{\text{res}})$.

The details of the computation of the optimal tap vector $\underline{w}(\omega)$ could be found in [3].

At frequency $f = 3932$ Hz, for the incidence direction defined by azimuth $\phi_0 = 0$ dg and elevation $90 - \theta_0 =$

44 dg, we have represented the spatial Fourier transform of the corresponding spatial filter $h_{(\phi_0, \theta_0)}$ computed on the sphere of radius $k = 2\pi f/c = 72 \text{ m}^{-1}$ with a uniform and an optimized tap vector on figure 4. Granted that the microphone array is almost bi-dimensional, the antenna cannot distinguish two different plane waves symmetrical with respect to the antenna plane. This explains the presence of two main lobes in the response of the spatial filter. The lobes are not perfectly symmetrical because of the slight position errors in the deployed microphone array.

The power focusing ratio and the white noise gain [11] of the microphone array have been averaged on a set of incidence directions spanning the whole sphere and the result has been plotted along frequency on figure 5. It is seen that the power focusing ration is improved by almost a factor 2 until 7 kHz using optimal tap vectors compared to uniform tap vectors. The wavelength corresponding to this frequency is 5 cm. It coincides with the diameter of the smallest circular array used in the global array. Concerning the amplification of noise, we use a regularization parameter in the beamforming algorithm, so that the noise is only amplified at the most by +5 dB compared to uniform weighting.

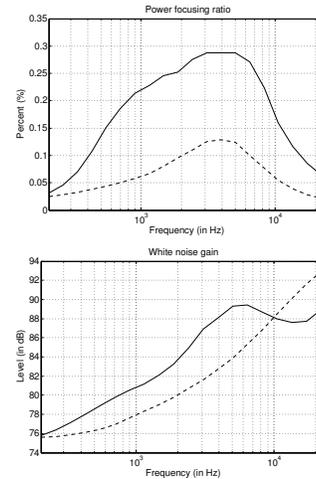


Fig. 5: Mean power focusing ratio (top) and white noise gain (bottom) of the antenna versus frequency. Uniform (dashed) and optimized (solid) tap vectors

3.2. Global sound field analysis scheme

In the previous paragraph, we have developed a beam-

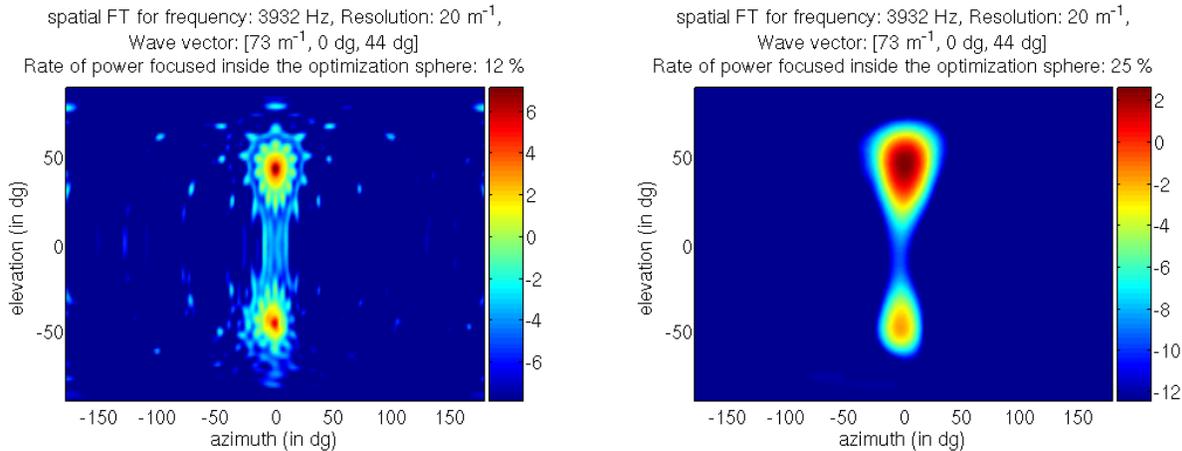


Fig. 4: Spatial filter response for uniform tap vector (left) and optimized tap vector (right).

forming algorithm which filters the incident sound field from a given direction (ϕ_0, θ_0) at a pulsation ω . The output of the spatial filter (6) could be considered as an estimator of the spatial Fourier transform evaluated for the wave vector \mathbf{k}_0 , that is $\hat{p}(\mathbf{k}_0, \omega)$. The sound field analysis step at a pulsation ω consists in computing the estimated values of the spatial Fourier transform $\hat{p}(\mathbf{k}_0, \omega)$ for several wave vectors \mathbf{k}_0 spanning the sphere defined by the dispersion relationship. Thus, it requires the computation of as many spatial filters $h_{(\phi_0, \theta_0)}$ as the number of wave vectors used in this spherical mesh.

Then, there exists three different manners to represent a cartography associated to the computed values $\hat{p}(\mathbf{k}_0, \omega_0)$:

- For each wave vector of the spherical mesh, $|\hat{p}(\mathbf{k}_0, \omega_0)|$ corresponds to a point of the sphere with same azimuth and elevation than \mathbf{k}_0 . The amplitude is given by a color map in decibel scale. This representation is used at the bottom of figures 6 and 7. The sphere is viewed from two angles to have a more global view of the estimated spectrum.
- The spectrum $|\hat{p}(\mathbf{k}_0, \omega_0)|$ is plotted on a flattened sphere, using the same color map as in the previous case. Examples of this representation are given at the top left of figures 6 and 7.
- The last method is analog to directivity diagrams, extended to three dimensions. The representation is defined as $r(\phi, \theta) = |\hat{p}(\frac{\omega_0}{c}, \phi, \theta, \omega_0)|$. For a specific

angle (ϕ, θ) , the radius is equal to the module of the estimated spatial Fourier transform $|\hat{p}(\frac{\omega_0}{c}, \phi, \theta)|$. Examples of this representation can be viewed at the top right of figures 6 and 7.

4. EXPERIMENTAL RESULTS

The spectrum of a real loudspeaker was estimated at frequency $f = 1034$ Hz using uniform tap vectors first. It is represented on figure 6 using the three different manners introduced in the previous paragraph. The spectrum is constituted of a main lobe in the direction $az = 270$ dg, $el = 0$ dg, and also of several significant side lobes. The representation of the spectrum uses a decibel scale of 15 dB extent. The position of the main lobe agrees very well with the estimated position of the transducer during the automatic position calibration step ($az = 270$ dg, $el = 3$ dg).

The same spectrum is plotted on figure 7, using a set of optimal tap vectors instead of uniform tap vectors. We can see that the spectrum is more localized in the wave vector domain and that side lobes almost disappear using a range of dynamics of 15 dB. Thus, the estimation of the spectrum is enhanced compared to the one using uniform tap vectors.

Sound field analysis using optimal tap vectors can exhibit the presence of early reflections, whereas the analysis based on uniform tap vectors failed to render so subtle details. This is shown on figure 8. At the top of the figure

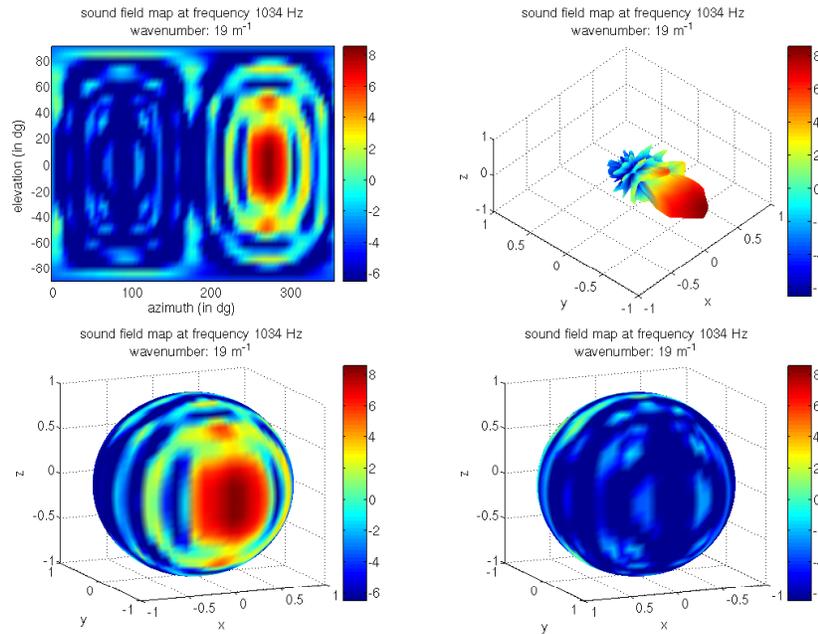


Fig. 6: Different sound field cartographies for the loudspeaker located by $[r = 1.1 \text{ m}, az = 270 \text{ dg}, el = 3 \text{ dg}]$ by uniform tap vectors. Flattened sphere (top left), directivity-like diagram (top right), sphere from first view angle (bottom left), sphere from second view angle (bottom right).

is represented the cartography of the real loudspeaker using either uniform or optimal tap vectors, whereas at the bottom of the figure is represented the cartography of the sound field corresponding to a monopole source located at the estimated position of the loudspeaker analyzed, using either uniform and tap vectors. The cartography of the sound field using uniform tap vectors seems noisier than the cartography of the simulated loudspeaker. On the other hand, the cartography of the real transducer using optimal tap vectors exhibits clearly new lobes corresponding to early reflections. These reflections are those generated by the loudspeakers located at the opposite of the circular array (see Fig. 1). We can see that these new lobes are not present on the cartography of the simulated loudspeaker.

5. CONCLUSION AND PERSPECTIVES

In this article, we have used a beamforming algorithm on real data acquired by a microphone array. The sound field analysis of real loudspeakers agrees very well with the sound field analysis of simulated loudspeakers, positioned at the same locations than the real ones. The

only differences between real and simulated loudspeakers concern the presence of early reflections in the case of real transducers. These early reflections go unnoticed when using uniform tap vectors so that optimal tap vectors based on generalized prolate spheroidal wave sequences represent an appreciable contribution to the domain of sound field analysis.

It has been shown that the microphone array, made up with several circular arrays with logarithmically-spaced radii, achieves a good power focusing ratio over a large frequency band, until 7 kHz. The wavelength corresponding to this frequency is 5 cm: it coincides with the minimal diameter of the circular arrays used during the measurement campaign. Moreover, the power focusing ratio using optimal tap vectors is improved by a factor 2 compared to the use of uniform tap vectors while keeping the amplification of noise to a satisfying level.

The main characteristics of the sound field analysis scheme described in this article are the resolution and the dynamics range. It has been shown that optimal tap vectors achieve good performance regarding the power focusing ratio when the resolution expected is fixed to a reasonable level. Moreover, the sound field cartography

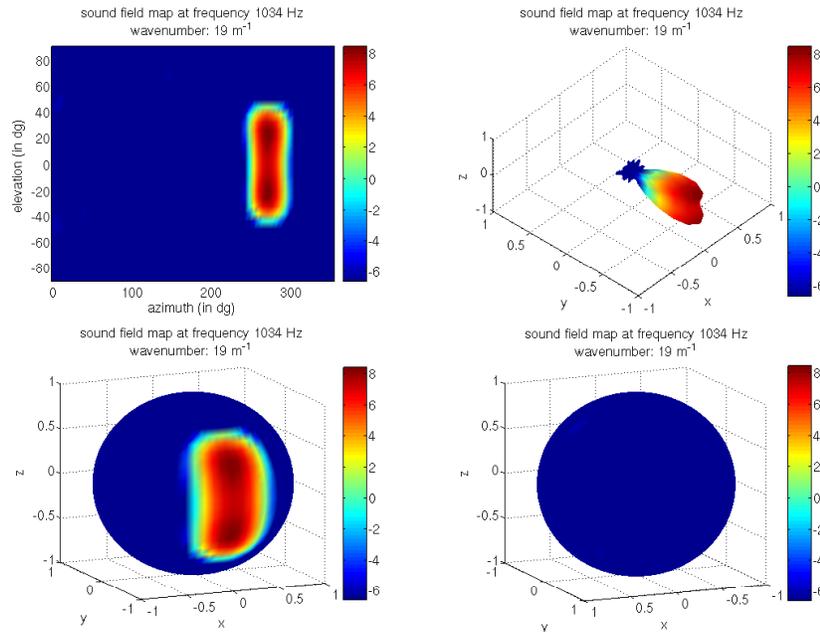


Fig. 7: Different sound field cartographies for the loudspeaker located by $[r = 1.1 \text{ m}, az = 270 \text{ dg}, el = 3 \text{ dg}]$ by optimal tap vectors. Flattened sphere (top left), directivity-like diagram (top right), sphere from first view angle (bottom left), sphere from second view angle (bottom right).

uses a larger extent of dynamics using optimal tap vectors compared to uniform tap vectors, thus enabling the detection of early reflections.

The results obtained in this article can be further improved in different manners:

- The use of three-dimensional array geometries could remove the ambiguity between up and down relative to the use of two-dimensional arrays, thus enabling a further improvement of the power focusing ratio.
- Granted that the direct sound component is correctly estimated in an anechoic chamber, it would be interesting to study real reverberant rooms with this sound field analysis scheme.
- Currently, the range of dynamics of 15 dB is sufficient to characterize the direct sound component and the first reflections, but seems insufficient to characterize the late reverberation. This range of dynamics could be improved either by studying

other microphone array geometries, or by increasing the number of microphones used, or by lowering the expectancies concerning the resolution of the acoustics imaging.

6. REFERENCES

- [1] Thushara D. Abhayapala and Darren B. Ward. Theory and design of high-order sound field microphones using spherical microphone array. *ICASSP*, 2002.
- [2] Jérôme Daniel and Sébastien Moreau. Further study of sound field coding with high order ambisonics. *116th Convention of The Audio Engineering Society*, May 2004.
- [3] Mathieu Guillaume and Yves Grenier. Sound field analysis based on generalized prolate spheroidal wave sequences. *120th Convention of the Audio Engineering Society*, May 2006.
- [4] Mathieu Guillaume and Yves Grenier. Sound field analysis with a two-dimensional microphone array. *ICASSP*, 2006.

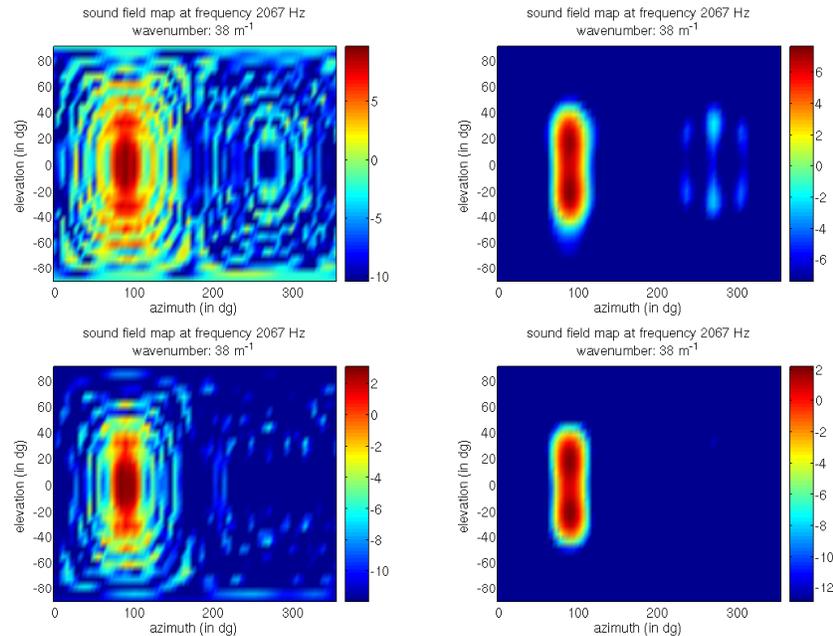


Fig. 8: Sound field cartography of a real loudspeaker (top) and a simulated monopole source located at the same position (bottom) using either uniform tap vectors (left) or optimal tap vectors (right). $f = 2067$ Hz, Estimated position: $[r = 1.11 \text{ m}, az = 91 \text{ dg}, el = 2 \text{ dg}]$.

- [5] Arnaud Laborie, Rémy Bruno, and Sébastien Montoya. A new comprehensive approach of surround sound recording. *114th Convention of The Audio Engineering Society*, March 2003.
- [6] Zhiyun Li, Ramani Duraiswami, Elena Grassi, and Larry S. Davis. Flexible layout and optimal cancellation of the orthonormality error for spherical microphone arrays. *ICASSP*, May 2004.
- [7] Jens Meyer and Gary Elko. A highly scalable microphone array based on an orthonormal decomposition of the soundfield. *ICASSP*, May 2002.
- [8] Swen Müller and Paulo Massarani. Transfer-function measurement with sweeps. *Journal of The Audio Engineering Society*, 49(6):443–471, June 2001.
- [9] Munhum Park and Boaz Rafaely. Sound-field analysis by plane-wave decomposition using spherical microphone array. *Journal of Acoustical Society Of America*, 118(5):3094–3103, November 2005.
- [10] M.A. Poletti. Three-dimensional surround sound systems based on spherical harmonics. *Journal of the Audio Engineering Society*, 53(11):1004–1025, November 2005.
- [11] Boaz Rafaely. Analysis and design of spherical microphone arrays. *IEEE Transactions on Speech and Audio Processing*, 13(1):135–143, January 2005.
- [12] Vikas C. Raykar and Ramani Duraiswami. Automatic position calibration of multiple microphones. *ICASSP*, 2004.
- [13] David J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, September 1982.
- [14] E.G. Williams. *Fourier Acoustics*. Academic Press, 1999.