Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2007, Article ID 94267, 15 pages
doi:10.1155/2007/94267

Research Article

Sound Field Analysis Based on Analytical Beamforming

M. Guillaume and Y. Grenier

Département Traitement du Signal et des Images (TSI), École Nationale Supérieure des Télécommunications, CNRS-UMR-5141 LTCI, 46 rue Barrault, 75634 Paris Cedex 13, France

Received 1 May 2006; Revised 4 August 2006; Accepted 13 August 2006

Recommended by Christof Faller

The plane wave decomposition is an efficient analysis tool for multidimensional fields, particularly well fitted to the description of sound fields, whether these ones are continuous or discrete, obtained by a microphone array. In this article, a beamforming algorithm is presented in order to estimate the plane wave decomposition of the initial sound field. Our algorithm aims at deriving a spatial filter which preserves only the sound field component coming from a single direction and rejects the others. The originality of our approach is that the criterion uses a continuous instead of a discrete set of incidence directions to derive the tap vector. Then, a spatial filter bank is used to perform a global analysis of sound fields. The efficiency of our approach and its robustness to sensor noise and position errors are demonstrated through simulations. Finally, the influence of microphone directivity characteristics is also investigated.

Copyright © 2007 M. Guillaume and Y. Grenier. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Directional analysis of sound fields is determinant in domains such as the study of vibrating structures, source localization, and applications dedicated to the control of sound fields, like wave field synthesis [1, 2], sound systems based on spherical harmonics [3], and vector-base amplitude panning [4]. In the particular case of 3D audio systems, the aim is to give the listener the impression of a realistic acoustic environment, which supposes that one is able to capture accurately a particular hall acoustics by the measure. For this purpose, microphone arrays are deployed in practice and some signal processing is applied in order to extract parameters to provide a spatial description of sound fields. Recent works have considered the case of spherical microphone arrays to estimate the spherical harmonic decomposition of the sound field to a limited order [5–8].

Another possible spatial description of sound fields is the plane wave decomposition, and beamforming can be used to estimate it. Beamforming is a versatile approach to spatial filtering [9]. Indeed, elementary beamforming consists in steering the sensor array in a particular direction, so that the corresponding spatial filter only preserves the sound field component coming from this direction and rejects the others. For this purpose, frequency beamforming techniques are well indicated. Firstly, the Fourier transforms of the time signals recorded by the microphones are computed. Then, at each frequency, the Fourier transforms of the microphone signals are weighted by a set of coefficients, constituting the tap vector. The tap vector is optimized in order that the response of the spatial filter approximates optimally a reference response. Generally, "optimally" means to minimize the mean square error between the effective and the reference responses on a discrete set of incidence directions [10–12]. For this kind of beamforming, the choice of the discrete set of incidence directions used for the definition of the mean square error norm is of crucial importance. In this article, a more difficult path has been chosen to optimize the tap vector, but it enables to circumvent this problem: the tap vector is still computed in order that the corresponding spatial filter only preserves the sound field component coming from a particular incidence direction, but the criterion implemented to achieve this objective is evaluated on a *continuous* set of incidence directions spanning the whole solid angle instead of a discrete set of incidence directions. This approach has been enabled by combining some results of linear acoustics theory and the efficiency of representation of nonuniformly space-sampled sound fields by the plane wave decomposition.

In previous works, we have already used the plane wave decomposition to describe the spatial behavior of sound

fields. In a first article, a method was given to derive optimal analysis windows weighting the measured microphone signals for bidimensional arrays [13]. Then, the analysis performance was further improved using generalized prolate spheroidal wave sequences to estimate the plane wave decomposition for a particular wave vector [14] in the case of tridimensional microphone arrays. In this article, the presentation of this sound field analysis approach is made clearer and more complete, by introducing a better description of the measured sound field. Moreover, a novelty is the use of a regularization procedure and the study of the robustness of the analysis to sensor noise, microphone error positions, and microphone directivity characteristics.

In Section 2, the plane wave decomposition is introduced, and the decomposition of the measured sound field is linked to that of the initial sound field. In Section 3, the detailed procedure implemented to compute the optimal tap vector used for beamforming is derived, and a regularization procedure used to increase the robustness of the analysis is presented. Then, several array configurations are compared. At Section 4, the use of regularization is validated through simulations concerning the influence of sensor noise and microphone error positions between the reference and the deployed array. Finally, the influence of microphone directivity characteristics is also investigated.

2. MULTIDIMENSIONAL FIELDS DESCRIPTION

In this section, the definition of the plane wave decomposition is first recalled. Then, it is employed to derive general forms of solutions to the inhomogeneous wave equation. At the end of this section, the plane wave decomposition is also used to model the measured sound field, and the corresponding decomposition is linked to that of the initial continuous space-time sound field.

2.1. The plane wave decomposition

The notations $\mathbf{k} = [k_x, k_y, k_z]$ and $\mathbf{r} = [x, y, z]$ in Cartesian coordinates or $\mathbf{k} = [k, \phi_k, \theta_k]$ and $\mathbf{r} = [r, \phi_r, \theta_r]$ in spherical coordinates will be used throughout this article. The quadridimensional Fourier transform [15] of the field $p(\mathbf{r}, t)$, also known as the plane wave decomposition since the atoms of the decomposition are the plane waves $e^{i(\mathbf{k}\cdot\mathbf{r}+\omega t)}$ for all $(\mathbf{k}, \omega)$ in $\mathbb{R}^4$, is defined by the relation

$$P(\mathbf{k}, \omega) = \iiiint_{(\mathbf{r},t)\in\mathbb{R}^4} p(\mathbf{r}, t) e^{-\mathrm{i}(\mathbf{k}\cdot\mathbf{r}+\omega t)} \,\mathrm{d}^3\,\mathbf{r}\,\mathrm{d}\,t. \tag{1}$$

The inverse quadridimensional Fourier transform enables to recover $p(\mathbf{r}, t)$ from its Fourier transform $P(\mathbf{k}, \omega)$. It is defined by the following relation

$$p(\mathbf{r}, t) = \frac{1}{(2\pi)^4} \iiiint_{(\mathbf{k},\omega)\in\mathbb{R}^4} P(\mathbf{k}, \omega) e^{i(\mathbf{k}\cdot\mathbf{r}+\omega t)} \,\mathrm{d}^3\,\mathbf{k}\,\mathrm{d}\,\omega. \tag{2}$$

The synthesis operator defined at (2) is able to synthesize any sound field, whether it is far field or near field, granted that the integration is performed for $(\mathbf{k}, \omega)$ in $\mathbb{R}^4$.

2.2. The wave equation

Acoustic fields are ruled by the wave equation:

$$\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} = -q(\mathbf{r}, t), \tag{3}$$

where $q(\mathbf{r}, t)$ is a source term. Additional initial and boundary conditions are required to ensure the existence and the uniqueness of the acoustic pressure field [16]. From the equivalence between boundary conditions and source term, we can say that the solution exists and is unique if the source term is known for every point of space $\mathbf{r}$ and every time instant t.

The Fourier transform of the inhomogeneous wave equation (3) yields

$$\left(|\mathbf{k}|^2 - \frac{\omega^2}{c^2}\right) P(\mathbf{k}, \omega) = Q(\mathbf{k}, \omega). \tag{4}$$

The acoustic pressure field is analytically given by the formula:

$$p(\mathbf{r}, t) = \frac{1}{(2\pi)^4} \iiiint_{(\mathbf{k},\omega)\in\mathbb{R}^4} \frac{Q(\mathbf{k}, \omega)}{|\mathbf{k}|^2 - \omega^2/c^2} e^{i(\mathbf{k}\cdot\mathbf{r}+\omega t)} \,\mathrm{d}^3\,\mathbf{k}\,\mathrm{d}\,\omega. \tag{5}$$

From (5), it can be deduced that the plane wave decomposition of the acoustic pressure field is likely to have singularities in the region of the frequency-wave vector domain $(\omega, \mathbf{k})$ when the dispersion relationship $\omega^2 - c^2|\mathbf{k}|^2 = 0$ is satisfied.

2.3. Measured sound field description

The microphone array has M_{mic} microphones, located at positions $\mathbf{r_m}$. In the following, we will assume that the sensors used are perfect omnidirectional microphones, so that the signal measured by the mth microphone—denoted by $p_{\mathrm{meas},m}(t)$ afterward—exactly corresponds to the value of the initial sound field $p(\mathbf{r_m}, t)$ at the microphone position. This is a simplification of the overall measurement process. A more precise formula for the sound field measured by a microphone array is established in Algorithm 1 at (11). When using perfect omnidirectional microphones, this equation resumes to:

$$p_{\mathrm{meas}}(\mathbf{r}, t) = \sum_{m=1}^{M_{\mathrm{mic}}} p(\mathbf{r_m}, t)\delta(\mathbf{r} - \mathbf{r_m}). \tag{6}$$

This equation is analogous to that modeling time signals $s(t)$ sampled at instants t_m,

$$s_{\mathrm{sam}}(t) = \sum_{m=1}^{M} s(t)\delta(t - t_m). \tag{7}$$

The electric signal measured by a microphone can be viewed as a continuous beamforming output signal [9], because the microphone is performing a local sort of spatial filtering by integrating the sound field on the whole surface of its membrane. This could be modeled by the following equation:

$$p_{\text{meas},m}(t) = (p *_4 h_{\text{mic},m})(\mathbf{r_m}, t), \tag{8}$$

where $*_4$ denotes the quadridimensional convolution product and $h_{\text{mic},m}$ is the space-time impulse response of the mth microphone. To interpret the previous equation, let us consider the convolution product $p *_4 h_{\text{mic},m}$ globally and not only at the position $\mathbf{r_m}$. Its Fourier transform is given by

$$P_{\text{glo}}(\mathbf{k}, \omega) = P(\mathbf{k}, \omega) \cdot H_{\text{mic},m}(\mathbf{k}, \omega). \tag{9}$$

The Fourier transform of the impulse response $H_{\text{mic},m}$ provides information on the frequency and the wave number bandwidth of the microphone, and also on the directivity characteristics of the mth microphone. Granted that the frequency component of the impulse response is dependent on electronics and that the wave vector component is dependent on the microphone geometry, the microphone impulse response could be fairly assumed to be separable:

$$H_{\text{mic},m}(\mathbf{k}, \omega) = K(\mathbf{k}) \cdot \Omega(\omega). \tag{10}$$

For an ideal omnidirectional microphone, $\Omega(\omega) = 1$, for all $|\omega| < \omega_{\max}$, $K(\mathbf{k}) = 1$, for all $|\mathbf{k}| < \omega_{\max}/c$ and 0 elsewhere. For a gradient microphone oriented along axis $\mathbf{r}_{\text{mic}}$, the directivity function is $K(\mathbf{k}) = \cos(\mathbf{k}, \mathbf{r}_{\text{mic}})$, for all $|\mathbf{k}| < \omega_{\max}/c$ and 0 elsewhere, where $(\mathbf{k}, \mathbf{r}_{\text{mic}})$ is the angle between vectors $\mathbf{k}$ and $\mathbf{r}_{\text{mic}}$.
The sound field measured by the microphone array could be modeled as

$$p_{\text{meas}}(\mathbf{r}, t) = \sum_{m=1}^{M_{\text{mic}}} p_{\text{meas},m}(t) \cdot \delta(\mathbf{r} - \mathbf{r_m}). \tag{11}$$

Algorithm 1: Digression on the measurement model.

In our case, the sampling of sound fields is made in the space domain. Using a well-known property of the multidimensional Dirac delta function, the measured sound field can be interpreted as the product between the initial sound field and another function, characterizing the sampling lattice:

$$p_{\text{meas}}(\mathbf{r}, t) = p(\mathbf{r}, t) \cdot \sum_{m=1}^{M_{\text{mic}}} \delta(\mathbf{r} - \mathbf{r_m}) \mathbf{1}(t). \tag{12}$$

In this equation $\mathbf{1}(t)$ stands for the function, whose value is 1 for all time instants t.

The quadridimensional Fourier transform of the measured sound field is

$$P_{\text{meas}}(\mathbf{k}, \omega) = P(\mathbf{k}, \omega) *_4 \left(\delta(\omega) \sum_{m=1}^{M_{\text{mic}}} e^{-\mathrm{i}\,\mathbf{k}\cdot\mathbf{r_m}} \right), \tag{13}$$

where $*_4$ is the symbol used for the four-dimensional convolution product.

The frequency component of the measured sound field is not distorted compared to that of the original sound field. On the other hand, the wave vector component is distorted by the convolution with the spatial characteristic function of the microphone array $\sum_{m=1}^{M_{\text{mic}}} e^{-\mathrm{i}\,\mathbf{k}\cdot\mathbf{r_m}}$. Thus, the measured sound field, which is discrete, no longer verifies the wave equation.

The number of microphones used in the array is always insufficient to enable conditions for the perfect reconstruction of sound fields compared to the well-known background of the sampling theory of time signals. Thus, the analysis of sound fields could only be approximated in practice. All what can be done is reducing the distortion introduced by the spatial sampling process.

3. BEAMFORMING FOR THE ESTIMATION OF THE PLANE WAVE DECOMPOSITION

Some signal processing on the measured data can be implemented in order to estimate the plane wave decomposition of the initial sound field, denoted as $\widehat{P}(\mathbf{k}, \omega)$ thereafter. We will only be interested in estimating this decomposition on the domain defined by the dispersion relationship $\omega^2 - c^2|\mathbf{k}|^2 = 0$, because this is the area of the frequency-wave vector domain for which the Fourier transform of the initial sound field $P(\mathbf{k}, \omega)$ is likely to have singularities. It seems that the restriction of the Fourier domain $(\mathbf{k}, \omega)$ in $\mathbb{R}^4$ to that defined by the dispersion relationship $\omega^2 - c^2|\mathbf{k}|^2 = 0$—a cone in four dimensions—is in agreement with the study performed in [17], which investigates the problem of sampling and reconstruction of the plenacoustic function when it is observed in the space domain on a line, on a plane, or in the whole space domain.

The method that we take as a reference afterward directly estimates the plane wave decomposition from (13), by computing the quadridimensional Fourier transform of the measured sound field. In practice, the Fourier transform for the time variable is firstly computed for every microphone signal, using the discrete Fourier transform, to obtain $p_\omega(\mathbf{r_m}, \omega_r)$, for a set of pulsations $(\omega_r)_{r\in[1,N_r]}$. The spatial Fourier transform is secondly computed digitally using

$$\widehat{P}(\mathbf{k}, \omega_r) = \sum_{m=1}^{M_{\text{mic}}} p_\omega(\mathbf{r_m}, \omega_r) e^{-\mathrm{i}\,\mathbf{k}\cdot\mathbf{r_m}}. \tag{14}$$

This reference method is far from being the most efficient. More degrees of freedom are required in order to achieve a better estimation of the plane wave decomposition. This can be done using frequency beamforming techniques. In this case, the first step of the signal processing remains identical: the Fourier transform of the measured

signal is computed using the discrete Fourier transform to obtain $p_\omega(\mathbf{r_m}, \omega_r)$ for a set of pulsations $(\omega_r)_{r\in[1,N_r]}$. Then, for each pulsation ω_r, and for a particular wave vector $\mathbf{k_0}$, we use a dedicated tap vector $\underline{\mathbf{w}}(\mathbf{k_0}, \omega_r) = [w_1(\mathbf{k_0}, \omega_r), \ldots, w_{M_{\text{mic}}}(\mathbf{k_0}, \omega_r)]^T$ to weight the spatial samples:

$$\hat{P}(\mathbf{k_0}, \omega_r) = \sum_{m=1}^{M_{\text{mic}}} w_m(\mathbf{k_0}, \omega_r)\, p_\omega(\mathbf{r_m}, \omega_r) e^{-\mathrm{i}\,\mathbf{k_0}\cdot\mathbf{r_m}}. \tag{15}$$

Thus, the reference method is retrieved by applying uniform weights $w_m = 1$. The objective of next sections is to provide a criterion to compute an optimal tap vector $\underline{\mathbf{w}}(\mathbf{k_0}, \omega_r)$.

3.1. Spatial filter and spatial aliasing

Equation (15) gives the method to compute digitally the estimation of the plane wave decomposition for a given pulsation ω_r and wave vector $\mathbf{k_0}$, but does not provide a method to compute the associated weights. For this purpose, we start from (12), equivalent to (6), but we introduce the weights w_m which differentiate the analyzed sound field from the measured sound field. The expression of the analyzed sound field is defined as

$$p_{\text{ana}}(\mathbf{r}, t) = p(\mathbf{r}, t) \cdot \sum_{m=1}^{M_{\text{mic}}} w_m \delta(\mathbf{r} - \mathbf{r_m})\mathbf{1}(t). \tag{16}$$

The quadridimensional Fourier transform of the previous equation is

$$P_{\text{ana}}(\mathbf{k}, \omega) = P(\mathbf{k}, \omega) *_4 \left(\delta(\omega) \sum_{m=1}^{M_{\text{mic}}} w_m e^{-\mathrm{i}\,\mathbf{k}\cdot\mathbf{r_m}}\right). \tag{17}$$

Let us explicit this convolution product. The convolution with $\delta(\omega)$ is omitted because convolving with the Dirac delta function is identity:

$$P_{\text{ana}}(\mathbf{k}, \omega) = \frac{1}{(2\pi)^3} \iiint_{\mathbf{k_1}\in\mathbb{R}^3} P(\mathbf{k_1}, \omega) \cdot \left(\sum_{m=1}^{M_{\text{mic}}} w_m e^{-\mathrm{i}(\mathbf{k}-\mathbf{k_1})\cdot\mathbf{r_m}}\right) \mathrm{d}^3\,\mathbf{k_1}. \tag{18}$$

With the previous equation, the analyzed sound field is still dependent of the wave vector $\mathbf{k}$, whereas the output of a frequency beamforming technique has to be a number. This requires to evaluate (18) for a specific wave vector $\mathbf{k}$. Granted that we want to design a good estimator of the spatial Fourier transform for a given wave vector $\mathbf{k_0}$ at a given pulsation ω_r, we choose the output signal of the beamformer to be that obtained by evaluating (18) for wave vector $\mathbf{k_0}$ and pulsation ω_r, according to (15),

$$\hat{P}(\mathbf{k_0}, \omega_r) \triangleq P_{\text{ana}}(\mathbf{k_0}, \omega_r). \tag{19}$$

The estimation of the Fourier transform $\hat{P}(\mathbf{k_0}, \omega_r)$ introduced at (19) is computed using the spatial filter defined as

$$h(\mathbf{k}) = \left(\sum_{m=1}^{M_{\text{mic}}} w_m e^{i(\mathbf{k}-\mathbf{k_0})\cdot\mathbf{r_m}}\right). \tag{20}$$

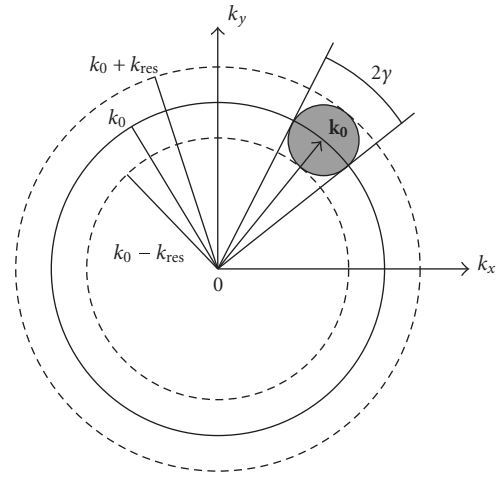

Figure 1: Slice of the 3D representation illustrating the optimization procedure: the power of the spatial filter is maximized in the sphere centered on $\mathbf{k_0}$ (gray disk) and minimized elsewhere in the spherical crown included between radii $k_0 - k_{\text{res}}$ and $k_0 + k_{\text{res}}$.

If it was perfect, then the response of the beamformer (19) to an input plane wave $e^{i(\mathbf{k}\cdot\mathbf{r}+\omega t)}$ should be null for every plane wave except for the plane wave of interest $e^{i(\mathbf{k_0}\cdot\mathbf{r}+\omega_r t)}$,

$$(2\pi)^4 \delta(\omega - \omega_r)\delta(\mathbf{k} - \mathbf{k_0}). \tag{21}$$

In fact, the response of the ideal beamformer is nothing else than the Fourier transform of the concerned plane wave. However, the effective response of the beamformer to an input plane wave $e^{i(\mathbf{k}\cdot\mathbf{r}+\omega t)}$ is

$$2\pi\delta(\omega - \omega_r)h(\mathbf{k}). \tag{22}$$

Thus, combining the last two equations, we can say that an ideal beamformer has to achieve the identity

$$h(\mathbf{k}) = (2\pi)^3 \delta(\mathbf{k} - \mathbf{k_0}). \tag{23}$$

Spatial aliasing occurs as soon as the response of the spatial filter differs from this ideal response. Unfortunately, the response of the corresponding spatial filter in the space domain is $e^{\mathrm{i}\,\mathbf{k_0}\cdot\mathbf{r}}$, requiring the observation of the sound field on the whole space domain. Thus, it is impossible in practice with a finite number of microphones that the response of the spatial filter—(20)—should be that of (21), so that spatial aliasing inevitably occurs.

In some way, the effective response of the beamformer has to approximate the ideal one: it has to be maximal for $\mathbf{k} = \mathbf{k_0}$ and minimal elsewhere. We can further improve what elsewhere means when the fields analyzed are sound fields: at pulsation ω_r, the interesting area of the wave vector domain is the sphere of radius $|\mathbf{k}| = \omega_r/c$. Granted that we want to estimate $P(\mathbf{k_0}, \omega_r)$, a good strategy consists in focusing the power of the spatial filter in the neighborhood of the wave vector $\mathbf{k_0}$ and minimizing the global power of the spatial filter on the sphere defined by the dispersion relationship (see Figure 1). The tap vector optimizing the estimation of

the Fourier transform for wave vector $\mathbf{k_0}$ and pulsation ω_r is the solution of the following equation:

$$\underline{\mathbf{w}}(\mathbf{k_0}, \omega_r) = \max_{[w_1,\ldots,w_{M_{\text{mic}}}]\in\mathbb{C}^{M_{\text{mic}}}} \frac{\iiint_{\mathbf{k}\in\mathcal{S}(\mathbf{k_0},k_{\text{res}})} |h(\mathbf{k})|^2 \,\mathrm{d}^3\mathbf{k}}{\iiint_{\mathbf{k}\in\mathcal{C}(\mathbf{0},k_0-k_{\text{res}},k_0+k_{\text{res}})} |h(\mathbf{k})|^2 \,\mathrm{d}^3\mathbf{k}}. \quad (24)$$

In this equation, $\mathcal{S}(\mathbf{k_0}, k_{\text{res}})$ indicates the sphere with center $\mathbf{k_0}$ and of radius k_{res}, while $\mathcal{C}(\mathbf{0}, k_0 - k_{\text{res}}, k_0 + k_{\text{res}})$ indicates the interior domain delimited by the two spheres with center $\mathbf{0}$ and with radii $k_0 - k_{\text{res}}$ and $k_0 + k_{\text{res}}$, respectively.

Before going through the details of the computation of the tap vector solution of (24), we will explain why this tap vector is a good candidate for the weights of the spatial filter $h(\mathbf{k})$. The response of the spatial filter (20) is constituted of a main lobe and also from many side lobes. The tap vector solution is such that it focuses the maximum of the power of its main lobe inside the sphere of resolution $\mathcal{S}(\mathbf{k_0}, k_{\text{res}})$ while attempting to place side lobes with the minimum of power inside the spherical crown $\mathcal{C}(\mathbf{0}, k_0 - k_{\text{res}}, k_0 + k_{\text{res}})$. To summarize, the tap vector solution of (24) is the one minimizing spatial aliasing, regardless of the microphone array geometry.

With the remarks made at the last paragraph, k_{res} in (24) appears as a key parameter to control the resolution of the analysis. It is linked to the angular resolution by the means of the relation

$$\gamma = \arcsin\frac{k_{\text{res}}}{k_0}. \quad (25)$$

The next paragraph deals with the computation of the two integrals of (24).

3.2. Tap vector computation

This section deals with the problem of the tap vector computation, and differentiates our approach from traditional approaches: rather than optimizing the tap vector over a discrete set of incidence directions, such as in [10–12], the optimization is applied over a continuous set of directions. As we will see, this optimization can be formulated analytically by using the development of a plane wave into spherical harmonics.

3.2.1. Kernels computation

We begin by expanding the numerator of (24):

$$\iiint_{\mathbf{k}\in\mathcal{S}(\mathbf{k_0},k_{\text{res}})} |h(\mathbf{k})|^2 \,\mathrm{d}^3\mathbf{k} = \iiint_{\mathbf{k}\in\mathcal{S}(\mathbf{k_0},k_{\text{res}})} \sum_{m=1}^{M_{\text{mic}}}\sum_{n=1}^{M_{\text{mic}}} \overline{w_m} e^{-i(\mathbf{k}-\mathbf{k_0})\cdot(\mathbf{r_m}-\mathbf{r_n})} w_n \,\mathrm{d}^3\mathbf{k}. \quad (26)$$

The weights, independent of the integration variable $\mathbf{k}$, can be put aside from the integral. Moreover, we change the integration variable to be $\mathbf{k} - \mathbf{k_0}$ instead of $\mathbf{k}$, resuming the previous equation to

$$\begin{aligned}\iiint_{\mathbf{k}\in\mathcal{S}(\mathbf{k_0},k_{\text{res}})} |h(\mathbf{k})|^2 \,\mathrm{d}^3\mathbf{k} &= \sum_{m=1}^{M_{\text{mic}}}\sum_{n=1}^{M_{\text{mic}}} \overline{w_m}\left[\iiint_{\mathbf{k}\in\mathcal{S}(\mathbf{0},k_{\text{res}})} e^{-i\mathbf{k}\cdot(\mathbf{r_m}-\mathbf{r_n})}\,\mathrm{d}^3\mathbf{k}\right] w_n \\ &= \underline{\mathbf{w}}^H \mathcal{T}_{\text{res}} \underline{\mathbf{w}}.\end{aligned} \quad (27)$$

The resolution kernel matrix $\mathcal{T}_{\text{res}}$ is defined by its elementary term

$$\left(\mathcal{T}_{\text{res}}\right)_{(m,n)} = \iiint_{\mathbf{k}\in\mathcal{S}(\mathbf{0},k_{\text{res}})} e^{-\mathrm{i}\,\mathbf{k}\cdot(\mathbf{r_m}-\mathbf{r_n})}\,\mathrm{d}^3\mathbf{k}. \quad (28)$$

Secondly, we continue by expanding the denominator of (24),

$$\begin{aligned}\iiint_{\mathbf{k}\in\mathcal{C}(\mathbf{0},k_0-k_{\text{res}},k_0+k_{\text{res}})} |h(\mathbf{k})|^2 \,\mathrm{d}^3\mathbf{k} &= \sum_{m=1}^{M_{\text{mic}}}\sum_{n=1}^{M_{\text{mic}}} \overline{w_m} w_n \\ &\quad\times\left[e^{i\mathbf{k_0}\cdot(\mathbf{r_m}-\mathbf{r_n})}\iiint_{\mathbf{k}\in\mathcal{C}(\mathbf{0},k_0-k_{\text{res}},k_0+k_{\text{res}})} e^{-\mathrm{i}\,\mathbf{k}\cdot(\mathbf{r_m}-\mathbf{r_n})}\,\mathrm{d}^3\mathbf{k}\right] \\ &= \underline{\mathbf{w}}^H \mathcal{T}_{\text{opt}} \underline{\mathbf{w}}.\end{aligned} \quad (29)$$

The optimization kernel matrix $\mathcal{T}_{\text{opt}}$ is defined by its elementary term

$$\begin{aligned}\left(\mathcal{T}_{\text{opt}}\right)_{(m,n)} &= e^{\mathrm{i}\,\mathbf{k_0}\cdot(\mathbf{r_m}-\mathbf{r_n})} \\ &\quad\cdot\left[\iiint_{\mathbf{k}\in\mathcal{S}(\mathbf{0},k_0+k_{\text{res}})} e^{-\mathrm{i}\,\mathbf{k}\cdot(\mathbf{r_m}-\mathbf{r_n})}\,\mathrm{d}^3\mathbf{k}\right. \\ &\quad\left. - \iiint_{\mathbf{k}\in\mathcal{S}(\mathbf{0},k_0-k_{\text{res}})} e^{-\mathrm{i}\,\mathbf{k}\cdot(\mathbf{r_m}-\mathbf{r_n})}\,\mathrm{d}^3\mathbf{k}\right].\end{aligned} \quad (30)$$

To evaluate the optimization and resolution kernel matrices, it is necessary to be able to compute the following integral:

$$\iiint_{\mathbf{k}\in\mathcal{S}(\mathbf{0},K)} e^{\mathrm{i}\,\mathbf{k}\cdot\mathbf{r}}\,\mathrm{d}^3\mathbf{k}. \quad (31)$$

Granted that the integration domain is a sphere, we express the above integral using the elementary volume described in the spherical coordinate system

$$\mathrm{d}^3\mathbf{k} = k^2\,\mathrm{d}\,k\sin\theta_k\;\mathrm{d}\,\theta_k\,\mathrm{d}\,\phi_k, \quad (32)$$

where $[k, \phi_k, \theta_k]$ indicate the radius, azimuth, and colatitude in the spherical coordinate system. For this purpose, we use the series development of a plane wave into spherical harmonics:

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(-\mathrm{i})^l \mathrm{j}_l(kr)\overline{\mathrm{Y}_l^m}(\phi_r,\theta_r)\mathrm{Y}_l^m(\phi_k,\theta_k). \quad (33)$$

Introducing (33) into (31) yields

$$\begin{aligned}\iiint_{\mathbf{k}\in\mathcal{S}(\mathbf{0},K)} e^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\,\mathrm{d}^3\mathbf{k} &= 4\pi\int_{k=0}^{K}\mathrm{j}_l(kr)k^2\,\mathrm{d}k\cdot\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(-\mathrm{i})^l \\ &\quad\cdot\int_{\phi_k=0}^{2\pi}\int_{\theta_k=0}^{\pi}\overline{\mathrm{Y}_l^m}(\phi_r,\theta_r)\mathrm{Y}_l^m(\phi_k,\theta_k)\,\mathrm{d}\phi_k\sin\theta_k\,\mathrm{d}\theta_k.\end{aligned} \tag{34}$$

From the orthogonality property of the spherical harmonics, only the term with $l = m = 0$ is nonnull. The integral finally resumes to

$$\begin{aligned}\iiint_{\mathbf{k}\in\mathcal{S}(\mathbf{0},K)} e^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\,\mathrm{d}^3\mathbf{k} &= 4\pi\int_{k=0}^{K}\mathrm{j}_0(kr)k^2\,dk \\ &= \frac{4}{3}\pi K^3\cdot 3\left[\frac{\sin(Kr)}{(Kr)^3}-\frac{\cos(Kr)}{(Kr)^2}\right] \\ &\triangleq \frac{4}{3}\pi K^3\mathrm{jinc}(Kr).\end{aligned} \tag{35}$$

The jinc function is analog to the jinc function in optics, which appears when dealing with the computation of the Fourier transform of a circular aperture. The jinc function is the tridimensional Fourier transform of a spherical domain. It tends to 1 when its argument tends to 0. From these results, the expression of the resolution and optimization kernels becomes, using the notation $\mathbf{r_m}-\mathbf{r_n} = [r_{mn},\phi_{r_{mn}},\theta_{r_{mn}}]$ in spherical coordinates,

$$(\mathcal{T}_{\mathrm{res}})_{(m,n)} = \frac{4}{3}\pi k_{\mathrm{res}}^3\mathrm{jinc}(k_{\mathrm{res}}r_{mn}) \tag{36}$$

$$\begin{aligned}(\mathcal{T}_{\mathrm{opt}})_{(m,n)} &= e^{\mathrm{i}\mathbf{k_0}\cdot(\mathbf{r_m}-\mathbf{r_n})} \\ &\quad\times\left[\frac{4}{3}\pi(k_0+k_{\mathrm{res}})^3\mathrm{jinc}[(k_0+k_{\mathrm{res}})r_{mn}]\right. \\ &\quad\left.-\frac{4}{3}\pi(k_0-k_{\mathrm{res}})^3\mathrm{jinc}[(k_0-k_{\mathrm{res}})r_{mn}]\right].\end{aligned} \tag{37}$$

Finally, the criterion (24) could be expressed into matrix form:

$$\underline{\mathbf{w}}(\mathbf{k_0},\omega_r) = \max_{[w_1,\ldots,w_{M_{\mathrm{mic}}}]\in\mathbb{C}^{M_{\mathrm{mic}}}}\frac{\underline{\mathbf{w}}^H\mathcal{T}_{\mathrm{res}}\underline{\mathbf{w}}}{\underline{\mathbf{w}}^H\mathcal{T}_{\mathrm{opt}}\underline{\mathbf{w}}}. \tag{38}$$

The optimal tap vector which maximizes (38) is also the eigenvector corresponding to the most powerful eigenvalue of the generalized eigenvalue problem of (39), as stated by Bronez in a work on spectral estimation of irregularly sampled multidimensional processes by generalized prolate spheroidal sequences [18]. The principle is the same in our approach, which only differentiates from [18] by a different choice of kernels: in [18], the fields were supposed band-limited inside a parallelepiped, while we suppose fields band-limited inside a sphere,

$$\mathcal{T}_{\mathrm{res}}\underline{\mathbf{w}}(\mathbf{k_0},\omega_r) = \sigma\mathcal{T}_{\mathrm{opt}}\underline{\mathbf{w}}(\mathbf{k_0},\omega_r). \tag{39}$$

This gives a method to compute the optimal tap vector. The performance of this tap vector is characterized by the power focusing ratio

$$\mathrm{PFR} = \frac{\underline{\mathbf{w}}^H\mathcal{T}_{\mathrm{res}}\underline{\mathbf{w}}}{\underline{\mathbf{w}}^H\mathcal{T}_{\mathrm{opt}}\underline{\mathbf{w}}}. \tag{40}$$

It gives the amount of power focused in the resolution sphere compared to the power in the neighborhood—in the spherical crown—of the sphere defined by the dispersion relationship (see Figure 1).

The tap vector is undetermined to a complex coefficient, so that an amplitude and phase normalization are applied. The amplitude normalization is made, so that the power inside the resolution sphere is unitary $\underline{\mathbf{w}}^H\mathcal{T}_{\mathrm{res}}\underline{\mathbf{w}} = 1$. The phase normalization is made, so that the sum of the weights $\sum_{m=1}^{M_{\mathrm{mic}}} w_m$ is a real number: thus none phase distortion is introduced by the spatial filter for wave vector $\mathbf{k_0}$, as seen in (20).

3.2.2. Regularization

Beamforming algorithms could be prone to noise amplification, mainly at low frequencies. Generally, the amplification of noise is characterized by the white noise gain [8]. This criterion has to be modified in the context of nonuniform multidimensional sampling. If sound fields are supposed to be band-limited in the wave vector domain inside the sphere of radius $|\mathbf{k}| = k_{\max} = \omega_{\max}/c$, and if the noise spectral density is assumed to be flat inside this sphere, then the noise amplification is characterized by the power of the spatial filter inside this sphere. Using an analogous reasoning as that used to compute the power of the spatial filter inside the optimization zone (29), the expression of the white noise gain (WNG) is

$$\mathrm{WNG} = \underline{\mathbf{w}}^H\mathcal{T}_{\mathrm{noi}}\underline{\mathbf{w}}, \tag{41}$$

$$(\mathcal{T}_{\mathrm{noi}})_{(m,n)} \triangleq e^{i\mathbf{k_0}\cdot(\mathbf{r_m}-\mathbf{r_n})}\frac{4}{3}\pi k_{\max}^3\mathrm{jinc}(k_{\max}r_{mn}). \tag{42}$$

$\mathcal{T}_{\mathrm{noi}}$ is the noise kernel matrix. Equation (41) computes the power of the spatial filter $h(\mathbf{k})$ inside the sphere of radius $|\mathbf{k}| = k_{\max}$.

It is possible to reduce the white noise gain during the tap vector computation procedure by adding a regularization step. The criterion (38) is updated in the following manner:

$$\underline{\mathbf{w}}(\mathbf{k_0},\omega_r) = \max_{[w_1,\ldots,w_{M_{\mathrm{mic}}}]\in\mathbb{C}^{M_{\mathrm{mic}}}}\frac{\underline{\mathbf{w}}^H\mathcal{T}_{\mathrm{res}}\underline{\mathbf{w}}}{\underline{\mathbf{w}}^H\left[(1-\lambda)\mathcal{T}_{\mathrm{opt}}+\lambda\mathcal{T}_{\mathrm{noi}}\right]\underline{\mathbf{w}}}. \tag{43}$$

The optimal tap vector of the regularized criterion is the eigenvector corresponding to the most powerful eigenvalue of the generalized eigenvalue problem:

$$\mathcal{T}_{\mathrm{res}}\underline{\mathbf{w}}(\mathbf{k_0},\omega_r) = \sigma\left[(1-\lambda)\mathcal{T}_{\mathrm{opt}}-\lambda\mathcal{T}_{\mathrm{noi}}\right]\underline{\mathbf{w}}(\mathbf{k_0},\omega_r). \tag{44}$$

The white noise gain depends on the value of the regularization parameter λ: increasing values of the regularization

parameter from 0 to 1 decreases the white noise gain and unfortunately also decreases the power focusing ratio. A tradeoff between the power focusing ratio and the white noise gain must be made.

The power focusing ratio and the white noise gain are displayed on Figure 2 for several values of the regularization parameter $\lambda = [10^{-10}, 10^{-8}, 10^{-7}, 10^{-6}]$. Moreover, the power focusing ratio and the white noise gain using uniform tap vectors—reference method—are also represented. The PFR and WNG represented have been averaged on a set of wave vectors $(\mathbf{k_n})_{n\in[1,N_k]}$ at each pulsation ω_r. Figure 2 has been obtained using the "icodt" geometry for the microphone array, which is described in Section 3.3.

The best PFR corresponds to $\lambda = 0$ (no regularization) but using these tap vectors amplifies the sensor noise of 40 dB at low frequencies and approximately 20–25 dB in the mid frequencies. The figure confirms that the WNG decreases when the value of the regularization parameter increases. The value of $\lambda = 10^{-7}$ achieves a good tradeoff between the power focusing ratio and the white noise gain. It is this value of the regularization parameter which we will be referring to thereafter when indicating that we are using a regularized analysis.

3.3. Array geometry optimization

The two global parameters having an impact on the quality of beamforming are the choice of the tap vector weights and the location of the microphones. In Section 3.1, we have optimized the weights of the tap vector regardless of the microphone array geometry. In this section, the problem of the optimization of the microphone array is addressed. The use of 1D microphone arrays to perform a 3D sound field analysis is the worst configuration because it introduces a strong form of spatial aliasing. Indeed, if the antenna is located on the (Ox) axis, the antenna is only able to analyze the k_x component in the wave vector domain. If the parameter k_x of a plane wave is correctly estimated, it nonetheless leaves an indetermination: all couples of parameters (k_y, k_z) satisfying $k_y^2 + k_z^2 = (\omega^2/c^2) - k_x^2$ are possible solutions for the two remaining components of the wave vector $\mathbf{k}$: this is a phenomenon comparable to that of the cone of confusion appearing in the estimation of the incidence direction from the knowledge of interaural time delays (ITDs). The use of 2D microphone arrays reduces spatial aliasing. Indeed, if the antenna is located in the (Oxy) plane, it enables to analyze the k_x and k_y components in the wave vector domain. Thus, if the parameters k_x and k_y of an incoming plane wave are correctly estimated, the two possible solutions for the last parameter k_z are $\pm\sqrt{(\omega^2/c^2) - k_x^2 - k_y^2}$: the ambiguity lies in the confusion between up and down. The use of 3D microphone arrays enables to remove this form of spatial aliasing.

The other form of spatial aliasing is due to the spacing between microphones. Using uniform spacing between microphones enables to perform a correct sound field analysis until the Nyquist rate, that is, at least two samples per wavelength. Above the frequency corresponding to this wavelength, there is another form of strong aliasing due to the apparition of

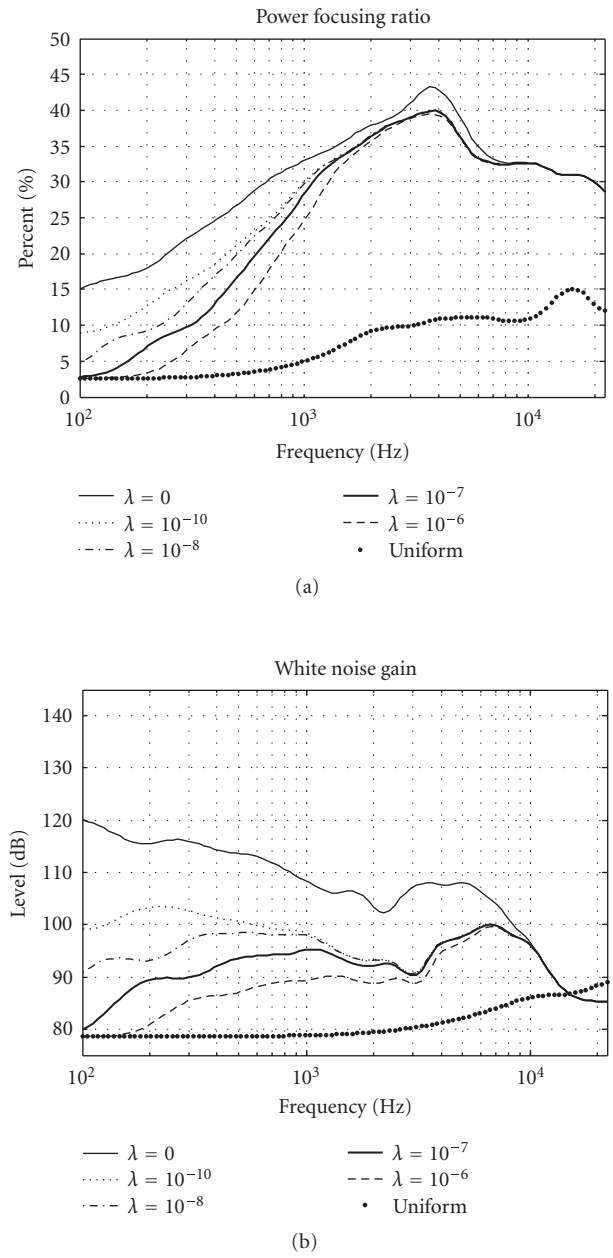

Figure 2: Power focusing ratio (PFR) and white noise gain (WNG) for several values of the regularization parameter λ and for uniform weighting.

replica—it can be interpreted as side lobes with power comparable to that of the main lobe—in the spatial spectrum, degrading substantially the power focusing ratio. The use of nonuniform spacing, and especially logarithmic spacing, attenuate these replicas. The use of nonuniform microphone arrays has already been emphasized in [13] for 2D microphone arrays: compared to uniform arrays, such as cross- or circular arrays, they enable to analyze the sound field in a large frequency band using the same number of microphones.

In this section, we will focus on the study of 3D microphone arrays, and several geometries will be compared using

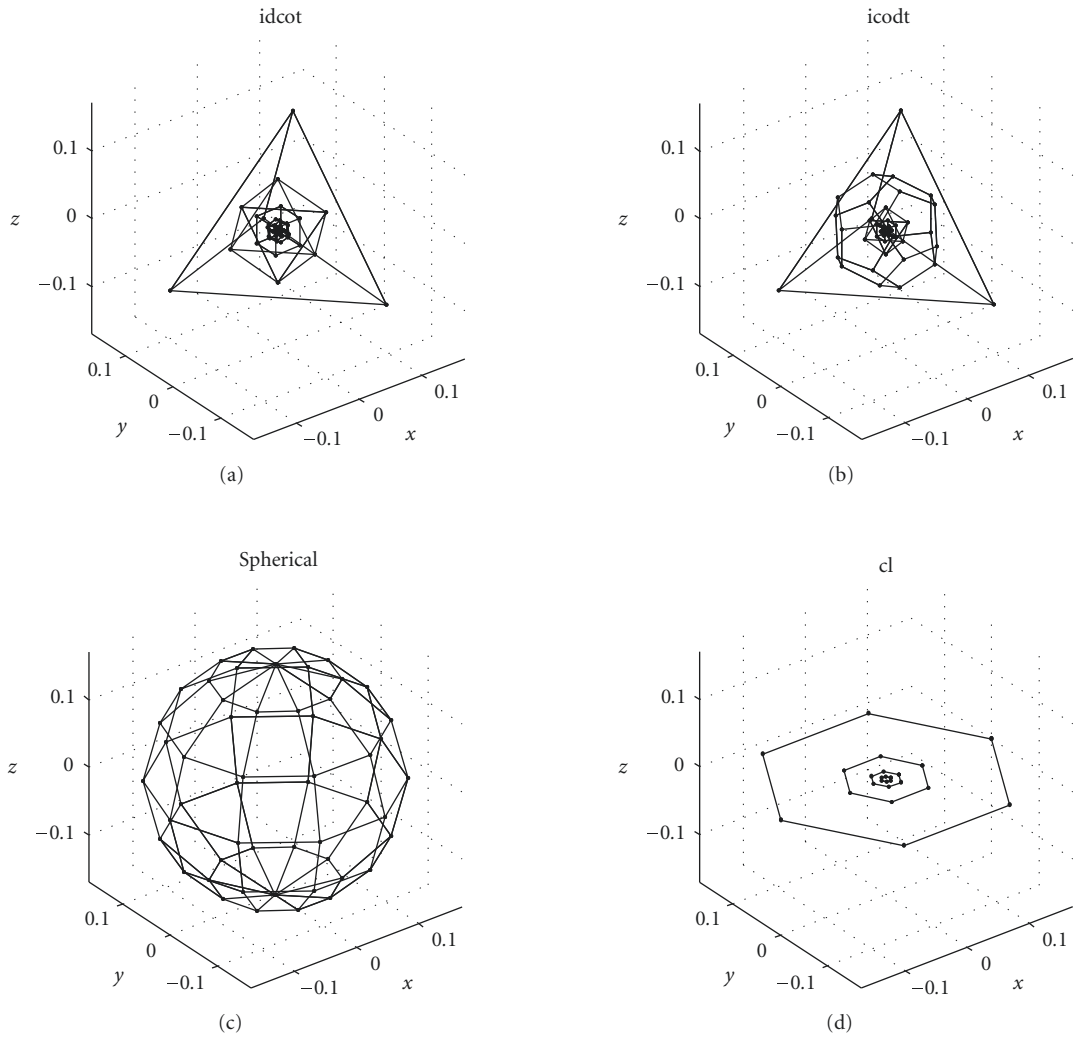

Figure 3: Microphone array geometries used for comparison: logarithmically spaced radii spherical array "idcot" (a) and "icodt" (b), regular spherical array (c) and double-height logarithmically spaced radii circular array (d).

the criteria of the power focusing ratio and white noise gain. The array geometries tested in simulation share common characteristics: they are all inscribed in a sphere of radius 0.17 m, and the number of microphones used is 50 ± 1 microphones. Here are the descriptions of the geometries used, shown on Figure 3.

(i) A spherical array of radius 0.17 cm using a uniform mesh using 10 microphones for the azimuth variable, and 7 microphones for the elevation variable. Thus, the array is constituted of 52 microphones (the two poles are counted only once).

(ii) Four circular arrays constituted of 6 microphones regularly spaced, with radii logarithmically spaced from 0.007 m to 0.17 m, and another microphone at the center of these circles. This subarray is duplicated twice in the planes defined by their equations $z = \pm 0.0025$ m. The global array is thus a "double-height logarithmically spaced radii circular array" made up with 50 microphones. The acronym used in the legend for this array is "cl."

(iii) Two arrays constituted of several Platonic solids: the tetrahedron, the octahedron, the cube, the icosahedron, and the dodecahedron which, respectively, have 4, 6, 8, 12, and 20 vertexes. These Platonic solids are inscribed in spheres with radii logarithmically spaced between 0.007 m and 0.17 m. The first array uses the order icosahedron, dodecahedron, cube, octahedron, and tetrahedron ("idcot" in the legends thereafter), while the second uses the order icosahedron, cube, octahedron, dodecahedron and tetrahedron ("icodt" in the legends) for increasing values of the radius. Finally, a last microphone is positioned at the origin. These two antennas are made up with 51 elements.

(iv) The last array uses a randomly distributed configuration of microphones ("random" in the legends). These microphones are uniformly distributed for

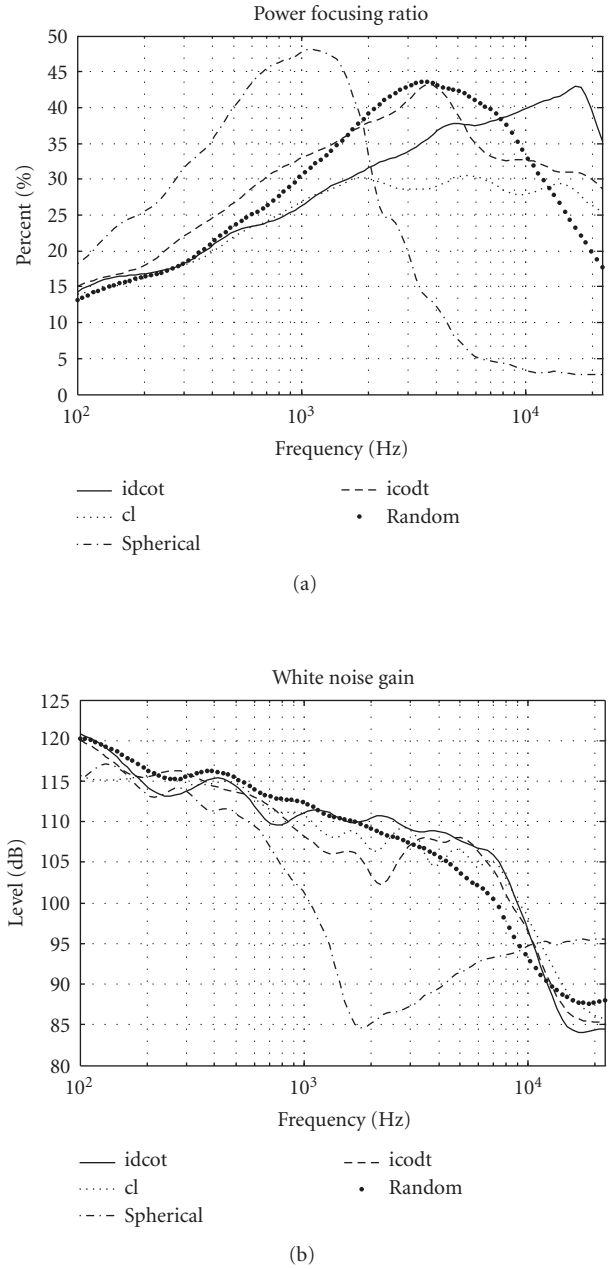

Figure 4: (a) Power focusing ratio (PFR) and (b) white noise gain (WNG) of several microphone arrays.

the azimuth and elevation variable, while it is the logarithm of the radial variable which is uniformly distributed. This array has also 51 microphones.

The power focusing ratios and the corresponding white noise gains of these 5 types of arrays are represented on Figure 4, using optimal nonregularized tap vectors. It is seen that the spherical array is well dedicated to the analysis of sound fields in the band of frequency around 1 kHz. At this frequency, the wavelength is 0.34 m, corresponding to the diameter of the spherical array. The power focusing ratio is largely lower for higher frequencies, because the microphone array does not sufficiently have closer microphones. This default is avoided by using the other kinds of microphone arrays, which have good performance on the whole frequency bandwidth of sound fields. Concerning the two Platonic arrays, the maximum power focusing ratio happens at the frequency corresponding to the wavelength 1.3 R, where R is the radius of the dodecahedron, namely 3.3 kHz for the "icodt" antenna, and 16 kHz for the "idcot" antenna. The distance 1.3 R is the mean distance between one vertex of the dodecahedron and the others. The random array is a little less efficient than the "icodt" array, in particular at high frequencies. The double-height logarithmically spaced radii circular array—quasi-bidimensional—is less efficient than true tridimensional arrays. Concerning the white noise gain, the logarithmic arrays present similar behaviors, the "icodt" having a slightly better trend. The minimum white noise gain of the spherical array happens at 1.7 kHz which corresponds approximately to the wavelength equal to the mean distance between microphones.

As a conclusion on the array geometry optimization, we can say that good array geometries combine both a domain with a high density of microphones, well dedicated to the study of small wavelengths—high frequencies—and also some distant microphones, dedicated to the to study of large wavelengths—low frequencies. To obtain a significant power focusing ratio in the low frequencies without amplifying too much the noise, some distant microphones are required. Thus, the use of logarithmically spaced microphones for the radial variable and uniformly spaced for the angular variables gives satisfactory results. In practice, the array geometry "icodt" has been retained for the following simulations.

4. SOUND FIELD ANALYSIS

In this section, we propose to detail a signal processing modulus able to perform a global sound field analysis from data recorded by a microphone array. This sound field analysis modulus uses the implementation of the beamformer presented at Section 3 to perform the spatial filtering required to achieve the spatial analysis. Here are the tasks sequentially carried out by the sound field analysis modulus.

(i) First, the Fourier transforms of the microphone data are computed using the FFT.

(ii) Then, at each pulsation ω_r, we use a spherical mesh of the sphere defined by the dispersion relationship $k = \omega_r/c$. For each wave vector $\mathbf{k_n}$ of this spherical mesh, we use the optimal tap vectors $\underline{\mathbf{w}}(\mathbf{k_n}, \omega_r)$ computed from Section 3.2 to estimate the Fourier transform of the initial sound field $\widehat{P}(\mathbf{k_n}, \omega_r)$.

(iii) Finally, we represent the cartography of the sound field at a given frequency on a flattened sphere, with azimuth on the x-axis and elevation on the y-axis. The modulus of the estimated Fourier transform is displayed using a colored-dB scale with 15 dB of dynamics.

All sound field cartographies represented in this section have been computed from simulated data for the microphone array. A source in free field emits a low-pass filtered Dirac delta impulse $\hat{\delta}$, so that the formula used to compute the signal recorded by a microphone of the array is

$$s_{\text{mic}}(t) = \frac{\hat{\delta}(t - ||\mathbf{r_m} - \mathbf{r_s}||/c)}{||\mathbf{r_m} - \mathbf{r_s}||}, \tag{45}$$

where $\mathbf{r_s}$ and $\mathbf{r_m}$, respectively, indicate the position of the source and the microphone.

The low-pass filtered Dirac delta impulse is a sinc function multiplied by a Kaiser-Bessel window [19],

$$\hat{\delta}(t) = \text{sinc}(2f_{\max}t) \cdot \begin{cases} \dfrac{\text{I}_0\left(\alpha\sqrt{1 - t^2/T^2}\right)}{\text{I}_0(\alpha)} & \text{if } |t| \le T, \\ 0 & \text{if } |t| > T, \end{cases} \tag{46}$$

with $f_{\max} = 20\,\text{kHz}$, $\alpha = 12$, to have a relative side lobe attenuation of 90 dB, and $T = 963\,\mu\text{s}$. It is the same simulation method as in [17].

4.1. Sound field cartographies

Two examples of sound field cartographies are represented on Figure 5. The initial source is located at [$r = 1\,\text{m}$, $az = 148\,\text{dg}$, $el = 0\,\text{dg}$] in spherical coordinates. The sound field cartography has been represented at the frequency $f = 2756\,\text{Hz}$ using either uniform tap vectors or optimal tap vectors. The optimal tap vectors have been computed for an angular resolution (25) of 23.5 dg.

In both cases, there is a maximum of power for the incidence direction of the source, that is, for $az = 148\,\text{dg}$ and $el = 0\,\text{dg}$. But the sound field obtained using uniform tap vectors is very blurred: the source is not well localized using the 15-dB extent of dynamics. On the other hand, the source is well localized using optimal tap vectors: there are no other visible side lobes, meaning that their amplitude is below 15 dB compared to the main lobe. We verify on the sound field cartography computed with optimal vectors that the angular resolution of the analysis is approximately 25 dg in this case, corresponding to the value of k_{res} fixed during the optimal tap vectors computation procedure. For this resolution, the average power focusing ratio is 35% compared to 10% using uniform tap vectors at 2756 Hz. Smaller resolutions would have led to a smaller power focusing ratio, and larger resolutions would have led to higher a power focusing ratios.

4.2. Influence of sensor noise and position errors

Two factors degrading the quality of the sound field analysis are the sensor noise generated mainly by the electronic part of the global electro-acoustic chain used in the microphone array and the errors of position between the reference array and the ad hoc deployed array. The sensor noise impairs the analysis mainly at low frequencies, where the amplification of noise is likely to be important. The position errors degrade the analysis mainly at high frequencies, where the magnitude

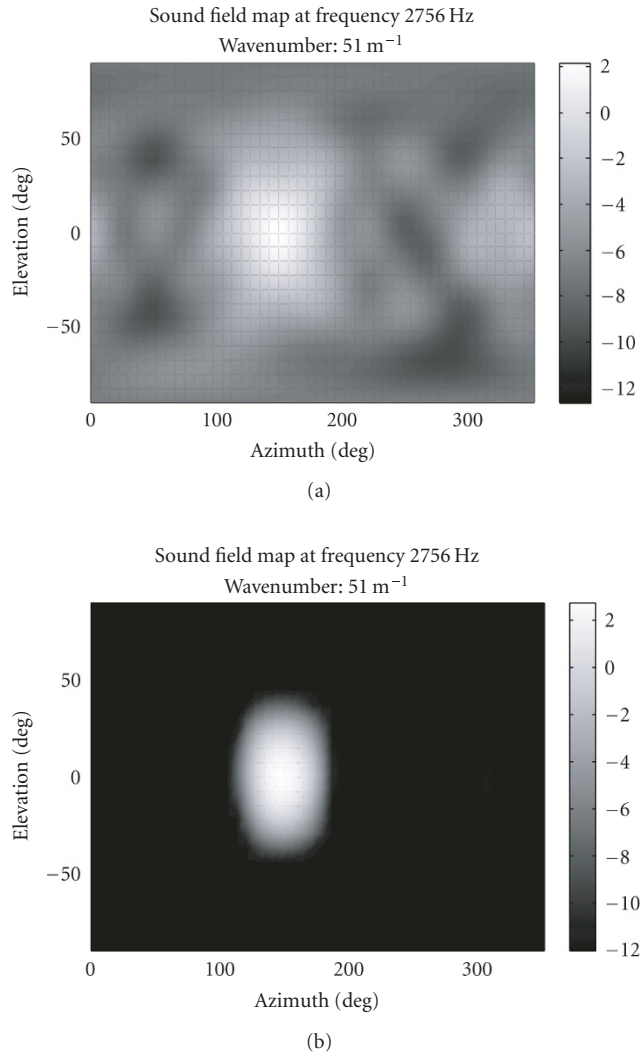

Figure 5: Sound field cartographies for a point source located at [$r = 1\,\text{m}$, $az = 148\,\text{dg}$, $el = 0\,\text{dg}$], at frequency 2756 Hz, using uniform tap vectors (top) or optimal tap vectors (bottom).

of the position errors becomes comparable with the wavelengths analyzed. In this paragraph, we will investigate these two considerations using simulations and will show that the use of regularization improves the robustness of the analysis to these two factors.

We are first considering the case of sensor noise. To highlight its influence, we are considering the analysis of a point source located at [$r = 1.5\,\text{m}$, $az = 52\,\text{dg}$, $el = -46\,\text{dg}$] in spherical coordinates at frequency $f = 345\,\text{Hz}$. The sound field cartographies obtained are represented on Figure 6, using either a regularized or nonregularized analyzer. On this figure, the cartography of the sound field is represented on the left, while the cartography of the noise is represented on the right. The initial data recorded by the microphone array were corrupted by an additive white noise, with signal-to-noise ratio equal to 30 dB. The regularized analysis is represented at the top of Figure 6, while the nonregularized analysis is represented at the bottom. It is seen that the

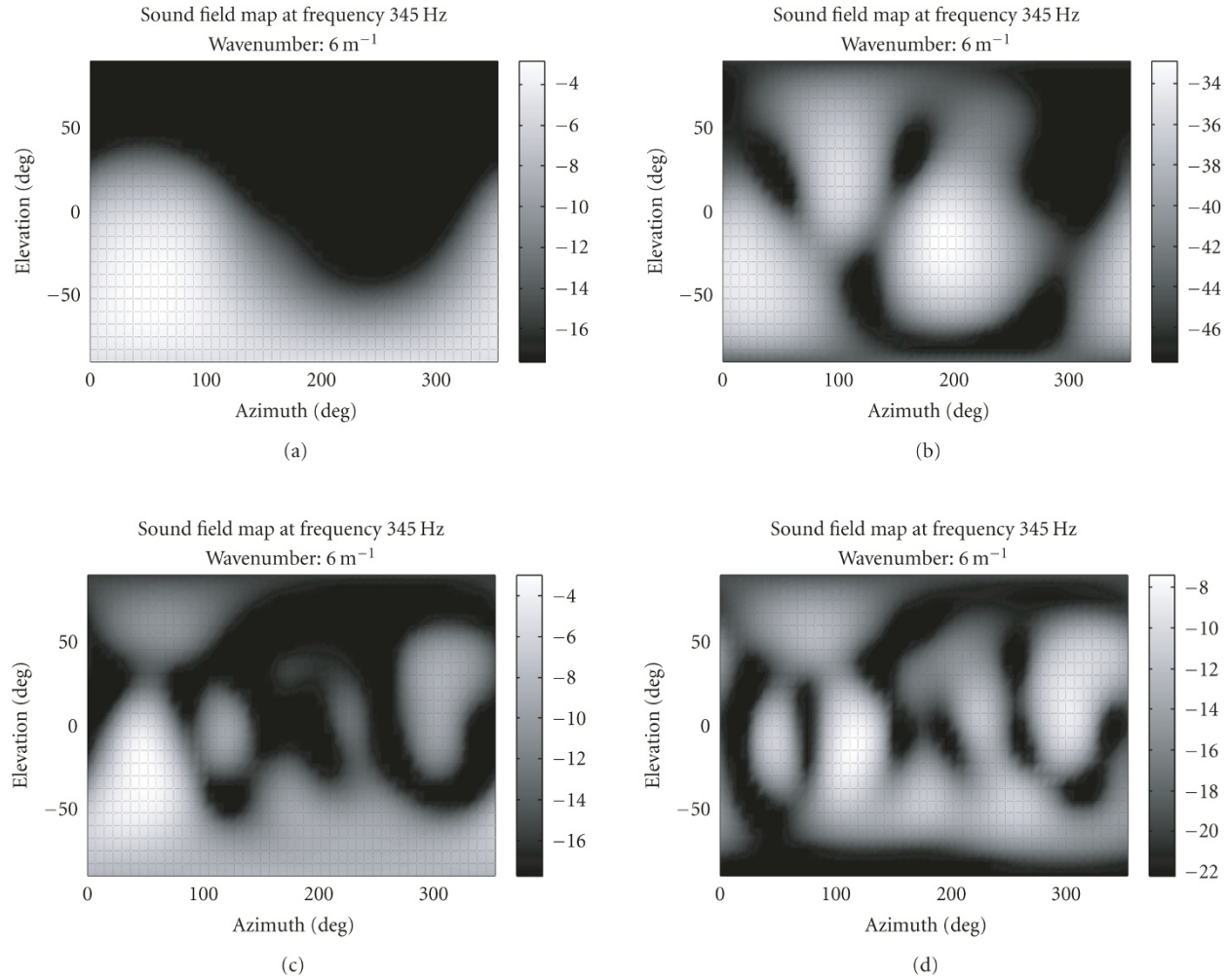

Figure 6: Influence of sensor noise on sound field cartographies for a point source located at [r = 1.5 m, az = 52 dg, el = −46 dg], for frequency 345 Hz, with initial SNR = 30 dB. Sound field regularized (a), error sound field regularized (b), sound field nonregularized (c), and error sound field nonregularized (d).

maximal value of the estimated spatial Fourier transform of the sound field is −3 dB, while the maximal value for the noise is −33 dB in the regularized case. Thus, the analysis using regularized tap vectors keeps approximately constant the signal-to-noise ratio in the frequency-wave vector domain compared to the time-space domain. On the other hand, the maximal value for the noise using nonregularized tap vectors is −8 dB, a difference of 25 dB with the regularized case, which corresponds to the difference between the two curves on Figure 2 at frequency 345 Hz. Thus, the sound field using regularized tap vectors keeps the same quality because the extent of dynamics used for representation is only 15 dB while the SNR is 30 dB. On the other hand, the noise is amplified when using nonregularized tap vectors, and this effect becomes visible at the bottom left of Figure 6. Thus, it is desirable to use the regularization to limit the degradation due to the presence of sensor noise.

We will now investigate the effects of position errors on the sound field analysis. For this purpose, position errors are assumed to create an additional noise on the microphones. This noise is defined as the difference between the signal really measured and the one that would have been measured if the microphone was located at the right place. On Figure 7, the corresponding signal-to-noise ratio is represented along frequency for several values of the uncertainty in the positionings: ±0.5, ±1.5, ±2.5, and ±5 mm. It is seen that the global trend is in $1/f$ because the slope of the curves is approximately −20 dB/dec. At a given frequency, the error is also 20 dB higher when the error increases by a factor 10 (see the curves related to 1 and 10 mm). Thus, the noise generated by position error affects the performance of the analysis at high frequencies. When the SNR is about 0 dB, the slope of the curve is no more 20 dB/dec, because the error cannot be superior to the signal power in mean. The boundary between these two parts of the curve happens approximately when the uncertainty is equal to the quarter of the wavelength, that is, 17 kHz for the uncertainty ±5 mm, this is slightly visible on Figure 6.

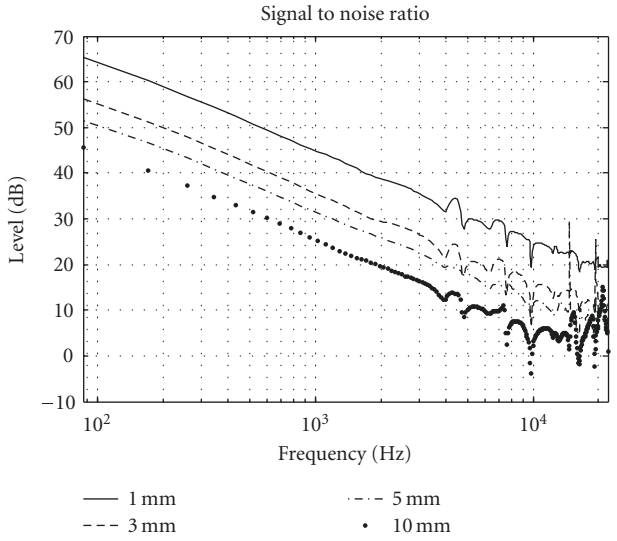

Figure 7: Signal-to-noise ratio versus frequency due to position errors.

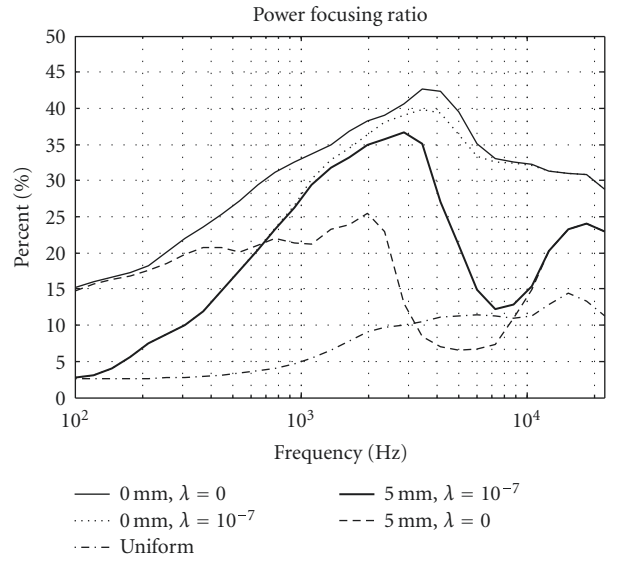

Figure 8: Influence of position errors on the power focusing ratio (PFR), with or without regularization. The position errors are uniformly distributed between $[-2.5, 2.5]$ mm.

The position errors induce a fall of the power focusing ratio. This is shown on Figure 8. The power focusing ratio of the analyzer has been computed using the tap vectors computed for the reference microphone array, but used on the real deployed array, with uncertainty of position of ± 2.5 mm, in two cases: using regularized or nonregularized tap vectors. The reference PFR (without position errors) have also been displayed in three cases: using regularized, nonregularized or uniform tap vectors. Once again, the use of regularization improves the robustness of the analysis: it is seen that the PFR obtained using regularized tap vectors is always superior to the PFR obtained using uniform tap vectors, contrary to the PFR obtained using nonregularized tap vectors. The difference between the reference PFR and the real one using regularized tap vectors is small until 3.4 kHz, compared to a few hundreds Hz using nonregularized tap vectors.

To conclude on the influence of position errors, a sound field cartography is represented on Figure 9 at $f = 3618$ Hz for a source located at $[r = 1\,\text{m},\ az = 270\,\text{dg},\ el = 31\,\text{dg}]$, with or without the use of regularization, for an uncertainty of ± 5 mm in the microphone positions. It is seen that the sound source is well resolved in the regularized case and not resolved in the other case.

4.3. *Influence of the directivity of microphones*

In the development made at Section 3, the microphones were assumed to be perfectly omnidirectional in order that (12) holds. If the directivity of the microphone differs from this ideal one, we can still compute by simulation the measured microphone signals using (8). Thus we can study the influence of the microphone directivity on the sound field analysis when still using the same analyzer as in the case of omnidirectional microphones.

Two examples of sound field cartographies are represented on Figures 10 and 11 for a source with coordinates $[r = 1\,\text{m},\ az = 180\,\text{dg},\ el = -58\,\text{dg}]$, at the frequencies of 689 Hz and 6202 Hz. The sound field cartographies are represented in two cases: using either omnidirectional microphones (top of Figures 10 and 11) or cardioid microphones oriented to the origin, towards the exterior (bottom of Figures 10 and 11).

The sound field cartographies at these two frequencies (and also for other frequencies) are noisier using cardioid microphones instead of omnidirectional microphones, but the direction of incidence is still correctly estimated. These cartographies have been computed using the regularized analyzer. Thus, we can say that it is better to use omnidirectional microphones with the analysis presented in this article.

5. FUTURE WORK

The main focus of this article has been to present a new method to perform spatial filtering: analytical beamforming. Throughout this article, we have mentioned three methods to estimate the plane wave decomposition: using uniform tap vectors in (15), which is the method we took as a reference, using a tap vector optimized over a discrete set of incidence directions, as in [10–12], or using the tap vector optimized over a continuous set of incidence directions, solution of (24). The comparison of the two last methods of optimization is a sufficiently important task, requiring some extra research effort to dedicate a future complete article. The corresponding spatial filters have to be compared with regard to their performance—using criteria such as the power focusing ratio and the white noise gain—and their computation complexity. Moreover, a crucial point in the discrete approach is the wave vector mesh used for the optimization procedure,

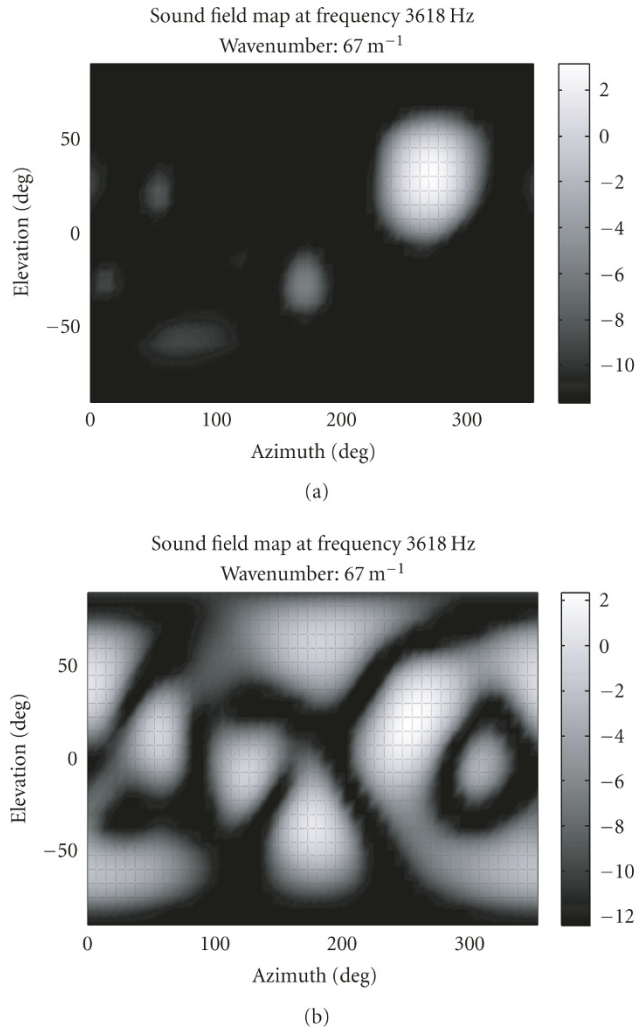

Figure 9: Sound field map obtained with measured data (mean position error of 5 mm) with regularized tap vectors (a) or nonregularized (b) tap vectors at frequency 3618 Hz for a source with coordinates [$r = 1$ m, $az = 270$ dg, $el = 31$ dg].

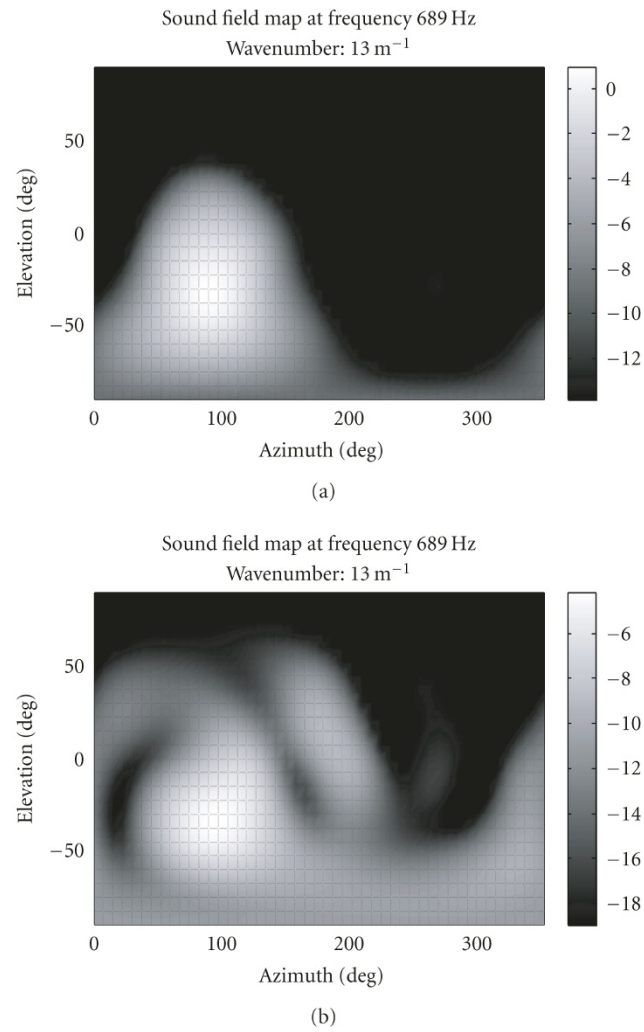

Figure 10: Sound field cartographies for a point source located at [$r = 1$ m, $az = 90$ dg, $el = -31$ dg], at frequency 689 Hz, using omnidirectional (a) or cardioid (b) microphones.

mainly for high wavelengths: indeed, the response of the corresponding spatial filter is likely to diverge apart from the wave vectors used for the optimization if the mesh is too sparse.

6. CONCLUSION

In this article, an analytical beamforming algorithm has been presented. Contrary to traditional beamforming algorithms which compute the coefficients weighting the measures by minimizing the mean square error on a discrete set of incidence directions, our algorithm does not use a discrete but a continuous set of incidence directions for the minimization. Thus, our algorithm avoids potential errors linked to the set of incidence directions used during the computation of the tap vector when using traditional methods.

The strategy used to compute the optimal tap vector for a particular incidence direction is to maximize the power of the sound field coming from the neighborhood of this direction and minimize the power of the sound field coming from other directions. The optimization criterion originally combines some results of linear acoustics theory with the efficiency of the quadridimensional Fourier transform to represent nonuniformly space-sampled fields.

The effectiveness of this algorithm has been demonstrated: it improves substantially the power focusing ratio compared to the reference case using a uniform tap vector. The amplification of noise can be kept to a level comparable to the reference case by using a regularization procedure. A tradeoff between the power focusing ratio and the amplification of noise has to be made. Then, several microphone array setups have been compared. It appears that good array geometries are those combining both a zone with a high density of sensors and also some distant microphones, such as tridimensional microphone arrays with logarithmically spaced microphones.

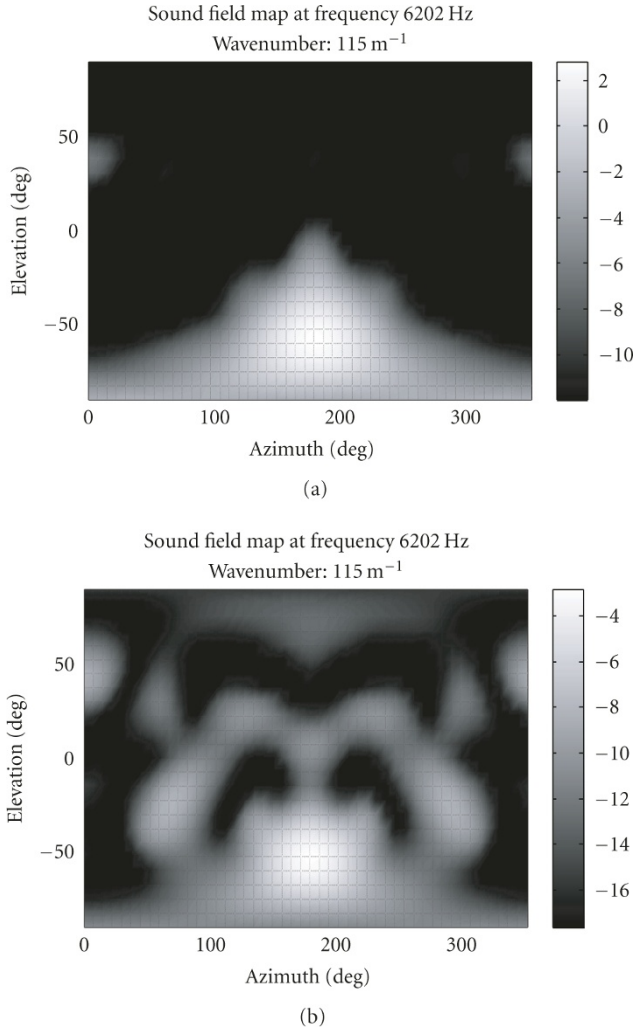

FIGURE 11: Sound field cartographies for a point source located at $[r = 1\,\text{m},\ az = 180\,\text{dg},\ el = -58\,\text{dg}]$, at frequency 6202 Hz, using omnidirectional (a) or cardioid (b) microphones.

Then, the robustness of the analysis to several factors known to degrade the quality of the analysis has been tested. These ones are the sensor noise, the position errors between the reference and the deployed microphone array, and the directivity characteristics of the microphones. The use of regularization is highly recommended and has been validated through simulations concerning the robustness of the analysis to sensor noise and position errors. Concerning the directivity characteristics of the sensors, the analysis is distorted when the directivity differs from the omnidirectional case. This is normal because the microphones were assumed to be omnidirectional for the derivation of the optimal tap vector criterion.

Some further work is needed to take into account more complex directivity characteristics for the optimal tap vector computation step, but the approach presented in this article is already particularly well indicated for sound field analysis dedicated to sound reproduction systems.

REFERENCES

[1] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764–2778, 1993.

[2] D. de Vries and M. M. Boone, "Wave field synthesis and analysis using array technology," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '99)*, pp. 15–18, New Paltz, NY, USA, October 1999.

[3] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *Journal of the Audio Engineering Society*, vol. 53, no. 11, pp. 1004–1025, 2005.

[4] J. Merimaa and V. Pulkki, "Spatial impulse response rendering I: analysis and synthesis," *Journal of the Audio Engineering Society*, vol. 53, no. 12, pp. 1115–1127, 2005.

[5] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '02)*, vol. 2, pp. 1781–1784, Orlando, Fla, USA, May 2002.

[6] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '02)*, vol. 2, pp. 1949–1952, May 2002.

[7] Z. Li, R. Duraiswami, E. Grassi, and L. S. Davis, "Flexible layout and optimal cancellation of the orthonormality error for spherical microphone arrays," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '04)*, vol. 4, pp. 41–44, Montreal, Quebec, Canada, May 2004.

[8] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, 2005.

[9] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[10] L. C. Parra, "Least-squares frequency-invariant beamforming," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '05)*, pp. 102–105, New Paltz, NY, USA, October 2005.

[11] L. C. Parra, "Steerable frequency-invariant beamforming for arbitrary arrays," *Journal of the Acoustical Society of America*, vol. 119, no. 6, pp. 3839–3847, 2006.

[12] S. Yan, "Optimal design of FIR beamformer with frequency invariant patterns," *Applied Acoustics*, vol. 67, no. 6, pp. 511–528, 2006.

[13] M. Guillaume and Y. Grenier, "Sound field analysis with a two-dimensional microphone array," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '06)*, Toulouse, France, May 2006.

[14] M. Guillaume and Y. Grenier, "Sound field analysis based on generalized prolate spheroidal wave sequences," in *120th Convention of the Audio Engineering Society*, Paris, France, May 2006.

[15] E. G. Williams, *Fourier Acoustics*, Academic Press, New York, NY, USA, 1999.

[16] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, McGraw-Hill, New York, NY, USA, 1953.

[17] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3790–3804, 2006.

[18] T. P. Bronez, "Spectral estimation of irregularly sampled multidimensional processes by generalized prolate spheroidal sequences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 12, pp. 1862–1873, 1988.

[19] J. F. Kaiser and R. W. Schafer, "On the use of the I0-sinh window for spectrum analysis," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 1, pp. 105–107, 1980.

M. Guillaume was born in Lille, Nord, France, in 1979. He was student at the Ecole Normale Supérieure de Cachan, where he received his B.S. degree in electronics, and his M.S. degree in acoustics, signal processing and computer science applied to music. He received the Ph.D. degree from Ecole Normale Supérieure des Télécommunications, Paris, in 2006. He is currently working for Trinnov Audio. His research interests are audio signal processing, particularly microphone and loudspeaker arrays, and also mathematics applied to acoustics.

Y. Grenier was born in Ham, Somme, France, in 1950. He received the Ingénieur degree from Ecole Centrale de Paris, in 1972, the Docteur-Ingénieur degree from Ecole Nationale Supérieure des Télécommunications, Paris, in 1977, and the Doctorat d'Etat ès Sciences Physiques, from University of Paris-Sud in 1984. He has been with Ecole Nationale Supérieure des Télécommunications, Paris, since 1977, as Assistant, and since 1984 as Professor. He has been Head of the TSI Department since January 2005. Until 1979, his interests have been in speech recognition, speaker identification, and speaker adaptation of recognition systems. He has then been working on signal modeling, spectral analysis of noisy signals, with applications in speech recognition and synthesis, estimation of nonstationary models, time frequency representations. He is presently interested in audio signal processing (acoustic echo cancellation, noise reduction, signal separation, microphone arrays, loudspeaker arrays). He is a Member of IEEE and AES. He has been the Chairman of, the 10th International Workshop on Acoustic Echo and Noise Control (IWAENC 2006).