SOUND FIELD ANALYSIS WITH A TWO-DIMENSIONAL MICROPHONE ARRAY

M. Guillaume, and Y. Grenier

École Nationale Supérieure des Télécommunications
Département TSI. 46 Rue Barrault
75634 Paris Cedex 13, France

ABSTRACT

Sound field analysis is performed from data recorded by a microphone array in order to extract some spatial information about the sound field. In this article, the assets of the multidimensional Fourier transform are emphasized in order to perform this analysis step. Indeed, the finite number of microphones inevitably introduces spatial aliasing, which is linked to the spectral broadening due to the discrete and finite nature of the analysis window. First, given the geometry of the array, the hypothesis that sound fields satisfie the dispersion relation, obtained from the wave equation, is used to design an optimal analysis window, which minimizes spatial aliasing according to a criterion inspired by prolate spheroidal spectral analysis. Then, several array configurations are compared to analyze sound fields with large frequency bandwidth.

1. INTRODUCTION

Several models are used to accurately describe a sound field inside a given domain: in Wave Field Synthesis, the sound field is described by the Kirchhoff-Helmholtz integral equation [1], whereas it is described as an expansion onto an orthogonal basis, such as the spherical harmonics, in Ambisonics [2] or in modal reconstruction [3]. In practice, estimating the parameters of the model by the mean of a microphone array during a recording session is a great challenge to be solved. Often, this challenge is circumvented by doing several dry sound recordings of the sound scene, where the microphones are located in the nearfield of the instruments, and then considering them as point sources with variable directivity. In this case, the real instruments and the room are replaced by physical models, which lie upon virtual acoustics. Reverberation is treated separately. Nevertheless, some sound reproduction systems exist, which aim at recreating a sound field by using its property of reciprocity [4], and there is a vast litterature on array processing, of whom some dedicated applications to sound field analysis and synthesis [5].

In this article, the sound field is modelled as an expansion onto the plane wave basis. The analysis operator is the multidimensional Fourier transform, giving the parameters of the expansion, and requiring the ideal knowledge of the sound field on the domain of interest. The synthesis operator is the inverse Fourier transform, giving the value of the sound field at any time instant, and at any location, and requiring the ideal knowledge of the parameters. For practical purposes, an exact estimation of the parameters is impossible, mainly because of the finite number of microphones. Thus, a perfect analysis of the sound field is impossible. Nevertheless, some processing could be made to perform an approximated sound field analysis, which is the topic of this paper.

corresponding author : mathieu.guillaume@enst.fr

In the first part of this article, the multidimensional Fourier transform and the physical background concerning the propagation of sound fields are introduced. In the second part, it is explained why a perfect analysis is unfeasible in practice, due to the spatial aliasing introduced by the discreteness and finiteness of the microphone array. It is also explained that the sampled version of the sound field is linked to its original version by convolution with a finite discrete analysis window, and a method is given to design this analysis window in order to achieve a good tradeoff between resolution and good focalisation of the energy inside the main lobe of the window. In the next part, the demonstration of the efficiency of this approach is emphasized, and several array geometries are compared. Finally, some concluding remarks and perspectives of this research are made.

2. SOUND FIELD MODEL

2.1. Multidimensional Fourier transform

Consider a well-behaved four-dimensionnal field $p(\mathbf{r},t)$, for which the Fourier transform exists, which is defined by:

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

It can be viewed as an *analysis operator*, which decomposes the sound field into plane waves. The associated *synthesis operator*, used to reconstruct the sound field from its analysis parameters, is nothing else than the inverse Fourier transform:

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

2.2. Physical background

The ear is sentitive to acoustic pressure variations by the mean of the ear drum, so that the analysis of the acoustic pressure field seems relevant in the domain of accurate sound field reproduction over an extended area. The acoustic pressure field obeys the wave equation, which takes the following form in a domain empty of sources [6]:

$$\nabla^2 p(\mathbf{r},t) - \frac{1}{c^2}\frac{\partial^2 p(\mathbf{r},t)}{\partial t^2} = 0 \qquad (3)$$

The Fourier transform of the above equation (3) yields:

$$\left(|\mathbf{k}|^2 - \frac{\omega^2}{c^2}\right) P(\mathbf{k},\omega) = 0 \qquad (4)$$

Nontrivial solutions to the homogeneous wave equation are obtained only if the *dispersion relation* $\omega^2 = c^2|\mathbf{k}|^2$ is verified. This

means that only a subset of the analysis parameters $P(\mathbf{k},\omega)$, verifying the dispersion relation, is required to correctly represent a given sound field. The assumption that audio signals are bandlimited to $|\omega| < \Omega$ further restricts the set of analysis parameters.

3. PARAMETER ESTIMATION

3.1. Spatial aliasing

The sound field is observed only on a set of N measure points located by $\mathbf{r}_n$, which are the positions of ideal omnidirectional microphones. The spatial analysis of the sound field is only possible by means of its sampled weighted version:

$$p_{\text{ana}}(\mathbf{r},t) = \sum_{n=1}^{N} w_n \delta(\mathbf{r}-\mathbf{r}_n)\, p(\mathbf{r},t) = W(\mathbf{r},t)\,.p(\mathbf{r},t) \qquad (5)$$

The time sampling of signals is not problematic, because the number of samples is usually sufficient to avoid aliasing in the time domain, so that it will be ignored for the rest of this document.

The product of two fields is transformed into a convolution product in the dual domain by the Fourier transform, so that the sampled weighted sound field is linked to the initial one by the relation:

$$P_{\text{ana}}(\mathbf{k},\omega) = W(\mathbf{k},\omega) * P(\mathbf{k},\omega) \qquad (6)$$

where $*$ denotes the convolution product symbol. The analysis window is discrete and finite in space so that its Fourier transform has a main lobe, limiting the resolution of the analysis, and also an infinity of secondary lobes in the $\mathbf{k}-$space.

The more convenient way to study the sampled sound field is to use the following discrete inner product:

$$< f|g >= \int_{t=-\infty}^{+\infty} \left[\sum_{n=1}^{N} w_n f(\mathbf{r}_n,t)\, \overline{w_n g(\mathbf{r}_n,t)} \right] \mathrm{d}t \qquad (7)$$

where the subscript $\bar{}$ denotes complex conjugation.

Nevertheless, the plane waves $e^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}+\omega t)}$ no longer form an orthogonal set for this discrete inner product (7), whereas the orthogonality property was satisfied with the continuous inner product — Fourier transform theorem. The consequence is that spatial aliasing inevitably occurs.

In this paper, we are interested in two-dimensional arrays in the plane (Oxy), so that the Fourier transform of the analysis window is independent of the k_z component of the vector $\mathbf{k}$. This means that any spectral information of the continuous sound field located at a specific value of k_z is copied to all values of k_z due to equation (6). The initial spectrum at pulsation ω was supposed to be non null only on a sphere of radius $|\mathbf{k}| = \omega/c$. For two-dimensional projection of sound fields, this states that the spectrum is non null only inside a circle of radius $|\mathbf{k}_r| = |k_x\mathbf{u_x} + k_y\mathbf{u_y}| = \omega/c$. Nevertheless, the location of spectral peaks gives direction of incidence of sources up to the ambiguity between up and down, so that a tri-dimensional spatial analysis is nearly achieved by two-dimensional arrays.

An example of Fourier transform of a two-dimensionnal sampled plane wave, using the antenna displayed at the right bottom of figure 1, with $k_x = k_y = 0$, is plotted on the left of figure 2, which is to be compared to $\delta(\mathbf{k})$, the Fourier transform of the continuous plane wave. A multitude of high-level side lobes is visible, which is to be avoided for the purpose of sound field analysis.

3.2. Optimal analysis window

Granted that the vector $\mathbf{k}$ of a plane wave is linked to the incidence direction, the broadening of the spectrum due to the convolution by the Fourier transform of the analysis window has two effects: the first one is that two plane waves with too near wavevectors $\mathbf{k_{r_1}}$ and $\mathbf{k_{r_2}}$ can not be separated by this analysis because of the lack of resolution, and the second one is that the presence of too high-level side lobes in the analysis window can hide some other spectral information or be wrongly interpretated as the presence of new spectral peaks in the Fourier transform. This observation motivates the choice of an analysis window which focus the power mainly in its main lobe, which spread indicates the resolution of the analysis. It is a well-known problem in the field of spectral analysis, particularly the one based on prolate spheroidal windows [7].

According to these observations, the criterion used to optimize the performance of the spatial analysis at pulsation ω is the maximization of the following rate [8]:

$$\lambda = \frac{\iint_{\mathbf{k_r}\in\mathcal{C}_{k_{\text{res}}(\omega)}} |W(\mathbf{k_r})|^2 \mathrm{d}^2\mathbf{k_r}}{\iint_{\mathbf{k_r}\in\mathcal{C}_{k_{\text{tot}}(\omega)}} |W(\mathbf{k_r})|^2 \mathrm{d}^2\mathbf{k_r}} \qquad (8)$$

where $\mathcal{C}_{k_{\text{res}}}$ (respectively $\mathcal{C}_{k_{\text{tot}}}$) is the circle of radius $|\mathbf{k_r}| = k_{\text{res}}(\omega)$ (respectively $|\mathbf{k_r}| = k_{\text{tot}}(\omega)$). The choice of $k_{\text{res}}(\omega)$ and $k_{\text{tot}}(\omega)$ is discussed further below.

The expression of λ can be reformulated into a matrix form:

$$\begin{aligned} \lambda &= \frac{\underline{\boldsymbol{w}}^H \boldsymbol{\mathcal{T}}_{\text{res}} \underline{\boldsymbol{w}}}{\underline{\boldsymbol{w}}^H \boldsymbol{\mathcal{T}}_{\text{tot}} \underline{\boldsymbol{w}}} \quad \text{with : } \underline{\boldsymbol{w}} = [w_1,\ldots,w_N]^T \\ \boldsymbol{\mathcal{T}}_{\text{res}}(m,n) &= \frac{2\pi k_{\text{res}}}{r_{\text{mn}}} \mathrm{J}_1(k_{\text{res}} r_{\text{mn}}) \\ \boldsymbol{\mathcal{T}}_{\text{tot}}(m,n) &= \frac{2\pi k_{\text{tot}}}{r_{\text{mn}}} \mathrm{J}_1(k_{\text{tot}} r_{\text{mn}}) \\ r_{\text{mn}} &= \left[(x_m-x_n)^2 + (y_m-y_n)^2\right]^{1/2} \end{aligned} \qquad (9)$$

where $\mathrm{J}_1(x)$ is the first order Bessel function.

The maximization of λ is equivalent to the resolution of the generalized eigenvalue problem [8]:

$$\boldsymbol{\mathcal{T}}_{\text{res}} \underline{\boldsymbol{w}}_i = \lambda_i \boldsymbol{\mathcal{T}}_{\text{tot}} \underline{\boldsymbol{w}}_i \qquad (10)$$

There are N generalized eigenvectors $\underline{\boldsymbol{w}}_i$, but only the one corresponding to the maximum eigenvalue is retained: this is the optimal analysis window for a given pulsation ω and for a fixed array geometry.

The thinner the resolution is, the lower the corresponding eigenvalues λ_i are. The antenna dimensions condition the best achievable resolution. In the same way, the bigger the radius k_{tot} is, the lower the corresponding eigenvalues are. So there is a tradeoff between resolution and efficiency of the weighting for the choice of $k_{\text{tot}}(\omega)$ and $k_{\text{res}}(\omega)$.

At pulsation ω, the spectral information of the continuous sound field is contained inside the circle of radius $k_{lim}(\omega) = \omega/c$. Assuming that all spectral information is contained in the main lobe of the analysis window, the region of interest for the sampled sound field in the $\mathbf{k}-$space is a circle of radius $k_{lim} + k_{\text{res}}$. To avoid spatial aliasing in these conditions, the condition $k_{\text{tot}} > 2k_{lim} + k_{\text{res}}$ has to be verified, from equation (6). A good choice for the resolution is to divide the global area in the $\mathbf{k}-$domain into N equal parts, each one corresponding ideally to the information brougth by one microphone:

$$k_{\text{tot}}^2 = N k_{\text{res}}^2 \qquad (11)$$

4. ARRAY GEOMETRY

In the previous section, the analysis window was optimized whereas the geometry has been fixed. In the present section, several geometries are compared, such as linear or logarithmic cross arrays, circular arrays, and logarithmically-spaced circular arrays. These geometries are compared in similar conditions: the arrays have the same characterisctic dimensions $L_x = L_y = 0.5\mathrm{m}$, and the same number of elements N equal to 49. The array geometries are represented on figure 2. At the top left corner is the linearly-spaced cross array, which elements are uniformly spaced by $4\mathrm{cm}$. At the top right corner is the logaritmically-spaced cross array, for which each branch contains 12 sensors logarithmically spaced between $1\mathrm{cm}$ to $50\mathrm{cm}$, plus a sensor at the center of the array. At the bottom left is the circular array, containing sensors angularly spaced by $2\pi/49$ radians. And at the bottom right is the logarithmically-spaced circular array, which is constituted of 6 circular arrays of 8 sensors each, and which radii are logarithmically spaced from $1\mathrm{cm}$ to $50\mathrm{cm}$, plus a sensor at the center. Random arrays have also been tested, with a radial distribution following a normal distribution, and the angular distribution following a uniform distribution between 0 and 2π.

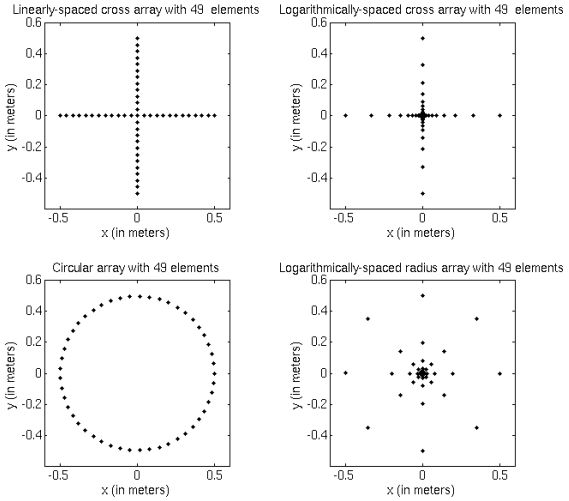


Fig. 1. Array geometries used for performance tests. Linear and logarithmic cross arrays at the top, circular array at the left bottom, and logarithmically-spaced radii circular array at the right bottom.

4.1. Comparison between uniform and optimal analysis windows

In this part, the uniform analysis window and the discrete generalized prolate sequence window (DGPSW) are compared, in the case of the logarithmically-spaced circular array, for a frequency $f = 3862\mathrm{Hz}$. for this frequency, $k_{\mathrm{tot}} = 4\pi f/c + k_{\mathrm{res}} = 166\mathrm{m}^{-1}$, with $k_{\mathrm{res}} = 23.8\mathrm{m}^{-1}$, from equation (11). The two-dimensional Fourier transforms of the analysis windows were computed and are displayed on figure 2. It is seen than the uniform analysis window has a better resolution than the prolate one. On the other hand, only 9.3 percent of power is focused inside the circle of radius $k_{\mathrm{res}}(\omega)$ compared to the entire energy comprised into the displayed circle. The generalized discrete prolate spheroidal window focuses 41 percent of the power at the frequency of study in the circle. Globally, it is seen than the DGPSW is more localized in the wavevector domain than the uniform window. These conclusions hold for all frequencies of interest, and it is the purpose of the next paragraph to study the frequency performance of the different types of array introduced at part 4.

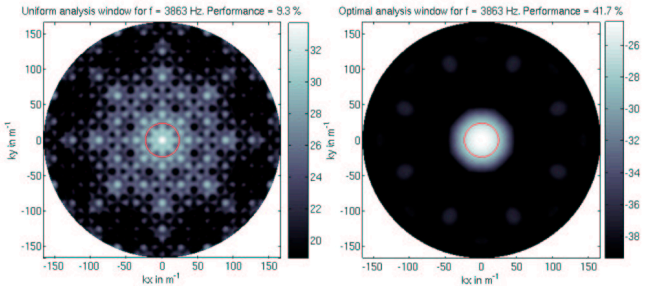


Fig. 2. 2D- spectrum of a sampled plane wave with $k_x = k_y = 0$ using a uniform analysis window at the top, and an optimal analysis window, maximizing the energy inside the circle, at the bottom.

4.2. Comparison of performance between arrays

The geometries presented in the beginning of this part were compared using optimal analysis windows, the design of which has been presented in section 3.2. The results are represented on figure 3, where the values of the maximum eigenvalue λ_{max} from equations (8) and (10) have been plotted along frequency. It is seen that the regularly-spaced arrays have efficient performance only in a restricted domain of frequencies, namely the low frequencies. At high frequencies, the sensors are too spaced, compared to the wavelength $\nu = c/f$, so that no spatial information could be efficiently retrieved. The critical value, 3400Hz for the linearly-spaced cross array, is to be linked to the spacing between sensors. Indeed, the corresponding wavelength is $\nu = 340/3400 = 0.1\mathrm{m}$ which is nearly the Shannon limit —two points per wavelength at critical sampling— which would have occured in the case of a uniform two-dimensional sampling. The "typical" performance of a random array is also good at low frequency and decreases with frequency. On the other hand, logarithmically-spaced arrays achieve good performance on the entire band of frequencies being studied. Indeed, sensors being located at the center of the array are very near, thus enabling analysis of the high wavenumbers. To study the spatial spectrum at low frequencies, one needs to further space the sensors, because all the signals would be quasi-identical if the sensors were too near, comparing to the wavelength being studied.

4.3. Examples

Some examples of sound fields are now analyzed. In the first one, the sound field is the sum of two plane waves of wavenumber $k = 96\mathrm{m}^{-1}$ corresponding to a frequency of 5.2kHz, with equal amplitudes. The plane waves are described by their wavevector $\mathbf{k} = [k, \phi, \theta]$, in spherical coordinates, where k is the wavenumber, ϕ the azimuth angle, and θ the elevation angle. The first plane wave has a wavevector $\mathbf{k_1} = [k, \pi/4, \pi/8]$, and the second one $\mathbf{k_2} = [k, -\pi/2, -\pi/8]$. Note that the plane waves are tri-dimensional, with a non null elevation angle. Their projection on the two-dimensional plane of observation is $\mathbf{k}' = [k\cos\theta, \phi]$ in cylindrical coordinates, yielding $\mathbf{k_1'} = [89, \pi/4]$ and $\mathbf{k_2'} = [89, -\pi/2]$. These two peaks are seen on the two-dimensional spectrum of the sampled

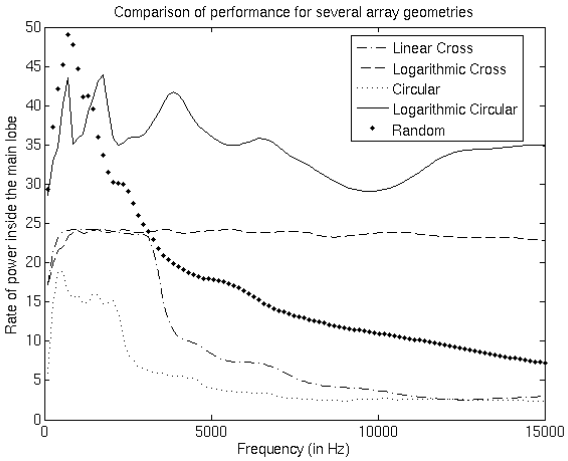


Fig. 3. Performance of optimal weightings for several array geometries versus frequency

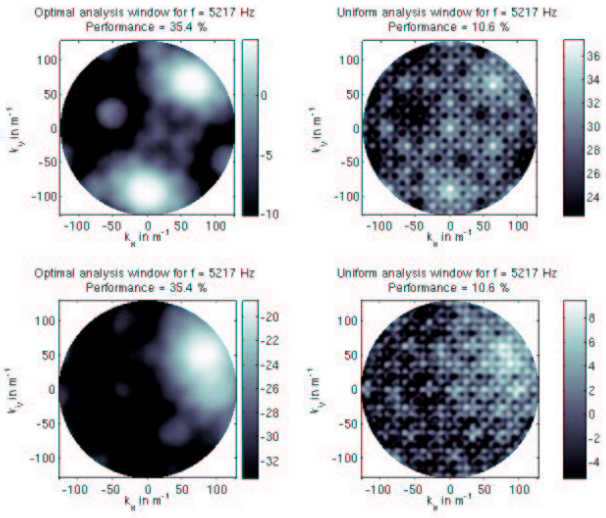


Fig. 4. Examples of Fourier transforms of sampled sound fields. At the top, two plane waves, and at the bottom, two close point sources. On the left using DGPSW, on the right using uniform analysis window

plane waves, represented on figure 4. At the top left is the spectrum of the two sampled plane waves weighted by the optimal DGPSW, and at the top right is the spectrum weighted by a uniform window. It is seen that the two plane waves are resolved in either two cases, but the spectrum is more localized in the case of DGPSW than in the case of uniform analysis window.

In the second example, two monochromatic points sources are simulated, at the same frequency as above, with equal amplitudes. The first one is located at $\mathbf{r_1} = [1.5, 0, 0]$ and the second one at $\mathbf{r_2} = [1, \pi/6, 0]$, still in spherical coordinates. It is seen at the bottom of figure 4 that the wavevector spectrum region excited is broader than with plane waves: this is due to the nearfield effect. Moreover, the component of the spectrum due to the second source is more energetic because the source is nearer than the first source. With the uniform analysis window, it is possible to resolve the two sources, but the spectrum of the two sampled points sources is largely more localized in the wavevector domain by using DGPSW, as seen at the bottom left of figure 4.

5. CONCLUSION

In this paper, a method to perform an approximated analysis of sound fields has been investigated. It is based upon the fact that the sampled version of the sound field is linked to its original version by the convolution of a discrete analysis window. Two parameters condition the performance of this analysis: the choice of the analysis window and the choice of the array geometry. In this article, the analysis window has been designed so that it focuses the maximum of energy inside the main lobe of the window in the wavevector domain, a criterion inspired by the spectral analysis based on prolate spheroidal wave functions. Several geometries have been tested, and it seems that circular arrays with logarithmically-spaced radii achieve the best performance in a large band of frequencies. It is important to emphasize that this approach overcomes the deficiencies of Ambisonics, Wave Field Synthesis or modal reconstruction which require a very rapidly increasing number of microphones when the frequency band becomes larger. It is the best spatial analysis achievable, given the number of microphones of the array. Moreover, this approach enables the analysis of sound recordings performed with real arrays. Some further studies remained to be done about the behaviour of the window design concerning noise influence on the sensors, and also about the influence of sensor location errors on the performance of the method but it is clear that this way is promising.

6. REFERENCES

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

[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] Terence Betlehem and Thushara D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2100–2111, April 2005.

[4] Zhiyun Li, Ramani Duraiswami, and Nail A. Gumerov, "Capture and recreation of higher order 3d sound fields via reciprocity," *International Conference on Audio Display*, 2004.

[5] 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.

[6] P.M. Morse and H. Feshbach, *Methods of Theoretical Physics*, Mc Graw-Hill, 1953.

[7] David Slepian, "Prolate spheroidal wave functions, fourier analysis, and uncertainty—iv: Extensions to many dimensions; generalized prolate spheroidal wave functions," *Bell System Technical Journal*, vol. 43, no. 6, pp. 3009–3058, November 1964.

[8] Thomas 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, December 1988.