Spatio-Spectral Masking for Spherical Array Beamforming

Uri Abend Boaz Rafaely
Ben-Gurion University of the Negev
Department of Electrical and Computer Engineering
Beer-Sheva 84105, Israel

***Abstract*—Beamforming using spherical arrays has become increasingly popular in recent years. However, the performance of beamforming algorithms is greatly affected by the limited number of sensors. This work offers a novel approach based on pre-processing of the spatial data in order to better separate the signal from noise, thus improving beamforming performance. The method involves transformation of the data to the spatio-spectral domain, using the spatially-localized spherical Fourier transform, followed by masking. The masking function is defined using a-priori knowledge of signal to noise ratio. The performance of the proposed algorithm is then evaluated using a simulation study, showing improvement over conventional spatial filtering.**

I. Introduction

One of the most popular methods of array processing is beamforming, where the known array geometry is utilized in order to realize spatial filtering by a weight and sum operation on the sensor outputs. Beamformer design is usually performed by optimizing a pre-determined cost function, chosen according to the desired performance and application. Typical design objectives include signal enhancement, noise reduction and de-reverberation. Many beamformers have been developed over the years; three of the most popular are the delay and sum beamformer, maximizing the Signal to Noise Ratio (SNR) [1], the Dolph-Chebychev beamformer that provides minimum side lobes for a given main lobe width [2] and the Maximum Directivity (MD) beamformer that maximizes the Directivity Factor (DF) of the beam pattern [3]. While fixed beamformers work well for stationary environments, to overcome non-stationary noise and disturbances, adaptive spatial filtering have been proposed, such as the minimum variance distortionless response (MVDR) [4] and the linearly-constrained minimum-variance (LCMV) [5].

The beamformers described above can be applied to arrays with various microphone configurations. Spherical arrays, in particular, have become increasingly popular due to the symmetry of the sphere and the Spherical Fourier Transform (SFT), allowing a design that is independent of the sensors location [3].

While beamformers with various configurations are used successfully in many applications, their spatial filtering performance is limited by spatial resolution. Therefore, for some challenging acoustic conditions or when the array is composed of a small number of microphones, performance is unsatisfactory. The design of microphone arrays with improved signal-enhancement performance is therefore still a great challenge.

In this work, we propose a pre-processing method aimed at improving the separation of a desired signal from a noisy background environment. This approach is based on the finding that different spatial components of a plane-wave sound field have different Spherical Harmonics (SH) spectra, allowing better separation, when compared to space-domain filtering. Based on available signal to noise ratio values, a spatio-spectral mask is applied to enhance the desired signal [6]. This approach is similar to the one used in speech signal processing, where the use of masks facilitates speaker separation in the short-time Fourier transform (STFT) domain [7].

In order to transform the signal to the spatio-spectral domain, the Spatially Localized Spherical Fourier Transform (SLSFT) is used on the plane-wave density signal of the sound field. The SLSFT is an analysis tool used in many fields of science and engineering, such as planetary topography and gravity [8], [9], [10]. Similar to the STFT, this transform produces a joint spatio-spectral distribution of a function on the sphere. The transform is performed by multiplying a function of interest by a spatially localized window function. In this work it is shown that improved separation between signal and noise can be achieved using the SLSFT compared to spatial filtering.

This paper is organized as follows: section II presents the necessary mathematical background on spherical array processing and the SLSFT. Section III presents the suggested method. Section IV presents the simulation study and section V presents conclusions and suggestions for future work.

II. Signal Processing on the Sphere

This section presents the mathematical background required for the development of the proposed method in the following section.

A. Beamforming in the Spherical Harmonics Domain

Consider an acoustical pressure field $p(\theta, \phi)$ defined in spherical coordinates. The elevation angle, $\theta \in (0, \pi)$, is measured from the positive z-axis and the azimuth angle, $\phi \in (0, 2\pi)$, is measured from the x-axis towards the y-axis. In this paper we assume narrow-band signals and far-field excitation, allowing representation of a source as an arriving plane-wave. $p(\theta, \phi)$ is now sampled by a spherical array comprised of Q microphones at locations $\{(\theta_i, \phi_i)\}_{i=1}^{Q}$. Assuming the function $p(\theta, \phi)$ has a maximal SH order of N, and that the number of samples Q is sufficiently large, the discrete SFT can be employed to obtain $\mathbf{p}_{nm}$ [3]. the spherical Fourier coefficients of the function are given by

$$\mathbf{p}_{nm} = [p_{00}, p_{1(-1)}, p_{10}, p_{11}, ..., p_{NN}]^T \tag{1}$$

Utilizing the assumption of far-field, the coefficients can be written as

$$p_{nm} = b_n(kr) a_{nm}(k) \tag{2}$$

with $b_n(kr)$ being a function dependent on the array configuration, $a_{nm}(k)$ being the plane-wave density [11], k being the wave number and r being the radius of the spherical array. Under some assumptions given in [12], the values of $b_n(kr)$ are non-zero, allowing the extraction of the plane-wave density from the acoustical pressure. Following the plane-wave decomposition method, beamforming can be employed. The beamformer's output, denoted y, can now be given by

$$y = \mathbf{w}_{nm}^H \mathbf{a}_{nm} \tag{3}$$

with $\mathbf{w}_{nm}$ being a vector of beamforming weights and $\mathbf{a}_{nm}$ being the plane-wave density function. Both have the structure given in Eq. (1).

B. Spatially Localized Spherical Fourier Transform

The Spatially Localized Spherical Fourier Transform (SLSFT) is an analytical tool for examining functions on the sphere. The transform is given by multiplying a function with a spatial window function and performing SFT, thus enabling analysis of specific parts of the function. The SLSFT of a function $f(\theta, \phi)$ is given by [8]

$$g(\theta, \phi, n, m) = \int_0^{2\pi} \int_0^{\pi} f(\theta', \phi') [\Lambda(\theta, \phi, 0) h(\theta', \phi')] \times Y_n^{m^*}(\theta', \phi') \sin\theta' d\theta' d\phi' \tag{4}$$

with $Y_n^m(\theta, \phi)$ being the normalized Spherical Harmonics function given in [3], $\Lambda(\alpha, \beta, \gamma)$ is the rotation operator given in [3], and $h(\theta, \phi)$ is the window function. In most cases, the window function would be axis-symmetrical, meaning that it is only a function of the elevation $h(\theta, \phi) = h(\theta)$. This reduces the rotation operation to a two-dimensional one [3].

The Inverse Spatially Localized Spherical Fourier Transfrom (ISLSFT) is given by integration over space [8]

$$\int_0^{2\pi} \int_0^{\pi} g(\theta, \phi, n, m) \sin\theta d\theta d\phi = \sqrt{4\pi} h_{00} f_{nm} \tag{5}$$

The choice of window is important for the performance of the SLSFT [8]. In practice, the window is rotated to a discrete set of Q' points $\{(\theta_j, \phi_j)\}_{j=1}^{Q'}$ and the resulting transform is given by - $g(\theta_j, \phi_j, n, m)$.

III. Soft Masking in the Spatio-Spectral Domain

Consider a spherical microphone array measuring the acoustical pressure in a sound field. The sound field is excited by a source, represented by a unit amplitude plane-wave arriving at an angle (θ_0, ϕ_0), and noise, represented by a diffuse sound field. The pressure is sampled at the microphones to receive $p(\theta_i, \phi_i)$, with $\{(\theta_i, \phi_i)\}_{i=1}^{Q}$ being the sampling points on the surface of the sphere. The pressure is then transformed to the SH domain using the SFT and the plane-wave density $\mathbf{a}_{nm}$ is extracted, as shown in section II.

$\mathbf{a}_{nm}$ can now be represented as the sum

$$\mathbf{a}_{nm} = \mathbf{s}_{nm} + \mathbf{v}_{nm} \tag{6}$$

with the entries of vector $\mathbf{s}_{nm}$ given by $s_{nm} = Y_n^{m^*}(\theta_0, \phi_0)$ and $\mathbf{v}_{nm}$ being a white Gaussian noise.

This is a conventional setting for a beamforming problem, where the aim can be the extraction of the amplitude of $\mathbf{s}_{nm}$, for instance. Most methods would use some known information about $\mathbf{s}_{nm}$ and $\mathbf{v}_{nm}$, i.e direction of arrival or other statistics, to reduce the energy of the undesired signal $\mathbf{v}_{nm}$ while keeping $\mathbf{s}_{nm}$ undistorted.

The suggested method is as follows: first we employ the SLSFT as in Eq. (4) on $\mathbf{a}_{nm}$ by setting $f(\theta, \phi) = a(\theta, \phi)$, with $\mathbf{a}_{nm}$ the SFT of $a(\theta, \phi)$, leading to

$$g_a(\theta_j, \phi_j, n, m) = g_s(\theta_j, \phi_i, n, m) + g_v(\theta_j, \phi_j, n, m)$$

$$1 \leq j \leq Q' \;\; ; \; n \leq N + N_h \;\; ; \; -n \leq m \leq n \qquad (7)$$

with N_h being the order of the window function $h(\theta)$ and $g_s(\theta_j, \phi_j, n, m), g_v(\theta_j, \phi_j, n, m)$ the SLSFT of $\mathbf{s}_{nm}$ and $\mathbf{v}_{nm}$, respectively. Next, a masking function $M(\theta_j, \phi_j, n, m)$ is defined. The function in this paper is motivated by the Wiener Filter [13], and assumes a-priori knowledge of the ratio between the desired signal and the total field at each space-SH point:

$$M(\theta_j, \phi_j, n, m) = \\ = \frac{|g_s(\theta_j, \phi_j, n, m)|}{\sqrt{|g_s(\theta_j, \phi_j, n, m)|^2 + |g_v(\theta_j, \phi_j, n, m)|^2}} \qquad (8)$$

The function $g_a(\theta_j, \phi_j, n, m)$ is multiplied by $M(\theta_j, \phi_j, n, m)$, aiming to enhance contributions from $g_s(\theta_j, \phi_j, n, m)$ and attenuate contributions from $g_v(\theta_j, \phi_j, n, m)$.

Now, the inverse transform can be employed, returning to the SH domain and yielding an estimate for the desired signal:

$$\tilde{\mathbf{a}}_{nm} = ISLSFT\{(g_a \times M)\} \qquad (9)$$

At this stage, common beamforming techniques can be used as in Eq. (3) with $\tilde{\mathbf{a}}_{nm}$ replacing $\mathbf{a}_{nm}$. A study of the method's performance will be shown in the next chapter.

To evaluate the performance, two measures are formulated. The first is the SNR defined as

$$\text{SNR} = \frac{||\tilde{\mathbf{s}}_{nm}||^2}{||\tilde{\mathbf{v}}_{nm}||^2} \qquad (10)$$

with $\tilde{\mathbf{s}}_{nm}$,$\tilde{\mathbf{v}}_{nm}$ defined as the operation of the algorithm described in Eq.(9) on $g_s(\theta_j, \phi_j, n, m)$ and $g_v(\theta_j, \phi_j, n, m)$ separately. The second is the distortion defined as

$$\text{Distortion} = \frac{||(\tilde{\mathbf{s}}_{nm} - \mathbf{s}_{nm})||^2}{||\mathbf{s}_{nm}||^2} \qquad (11)$$

The better the estimate of $\mathbf{s}_{nm}$, the better a beamformer will perform in the following processing stage. To gain a good estimation, both performance parameters should be high, that is, a high SNR should be maintained, while preserving the spectral characteristics of the desired signal. The performance of the spatio-spectral masking is compared with a similar masking process in the SH domain. The comparison is made using the same information, given in the SH domain instead of the spatio-spectral domain

$$\tilde{\tilde{a}}_{nm} = a_{nm} \frac{|s_{nm}|}{\sqrt{|s_{nm}|^2 + |v_{nm}|^2}} \qquad (12)$$

The results are given in the next section.

IV. Simulation Study

A sound field $p(\theta, \phi)$ was simulated, comprising of a single plane-wave arriving at a direction $(\theta_0, \phi_0) = (0, 0)$. The sound field was measured at a microphone array with a Gaussian sampling scheme designed for sampling of up to order $N = 10$, setting $Q = 2(N+1)^2 = 242$ [12]. The plane-wave density $\mathbf{a}_{nm}$ is extracted as shown in section II.

A. Spatio-Spectral analysis of a Single Plane Wave

Consider a sound field composed of a desired plane arriving from the north pole, and an interfering plane wave arriving from another direction. $a(\theta, \phi)$ in this case is composed of two plane waves, but around the north pole, the desired signal is dominated by its main lobe, while if the interference arrive sufficiently away from the north pole, it is dominated by its side lobes. In this section, differences between main lobe and side lobes parts of a plane wave are studied. The SLSFT of $a(\theta, \phi)$ is calculated in this case at a single point $(\theta_j, \phi_j) = (0, 0)$ using a window function of a spherical cap:

$$h(\theta) = \begin{cases} 1 & \theta \leq \theta_{cutoff} \\ 0 & else \end{cases} \qquad (13)$$

with θ_{cutoff} being the width of the main lobe, set to be $\frac{\pi}{N} = \frac{\pi}{10}$ [3]. Note that in this case, the window function is not order-limited as it is finite in space. This calculation yields the spectrum of the main lobe denoted $\mathbf{a_1}_{nm}$. The side lobes section can be realized as the remainder of the function outside of the main lobe and is therefore given by $\mathbf{a_2}_{nm} = \mathbf{a}_{nm} - \mathbf{a_1}_{nm}$. The results are shown in figure 1. Although the two functions are not order-limited, because the window function is not order-limited, they are only shown for $0 \leq n \leq N$ as they are equal in magnitude for higher values of n because their sum is an order-limited function. Note the distinct differences in the spectral formation of the main lobe and the side lobes. These difference form the basis for separating a desired signal from interferences originating from the side lobes of other plane-wave components of the sound field.

B. A Single Plane Wave with Additive White Gaussian Noise

In order to evaluate the performance of the proposed algorithm, white Gaussian noise $\mathbf{v}_{nm}$ is added to a unit amplitude plane-wave $\mathbf{s}_{nm}$, both composing $\mathbf{a}_{nm}$. $\mathbf{v}_{nm}$ is modeled in the space domain by a diffuse field comprised of an infinite number of plane-waves

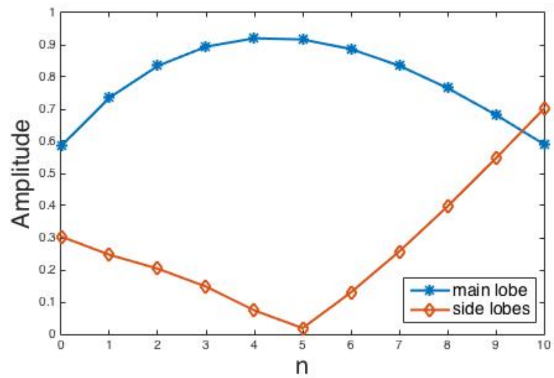

Fig. 1: SFT of a plane-wave's main lobe and side lobes

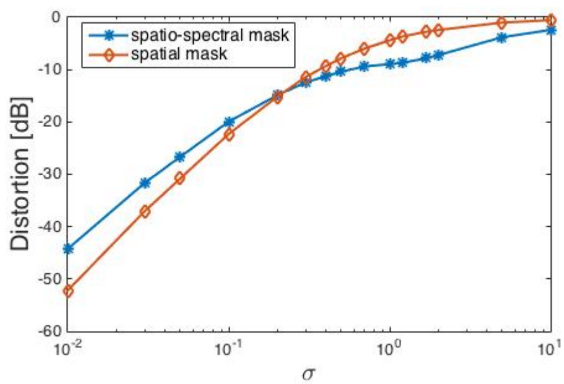

Fig. 2: Distortion as a function of Gaussian noise standard deviation

arriving in all directions with a random phase. The direction of arrival of the desired plane-wave, $\mathbf{s}_{nm}$, was set to $(\theta_0, \phi_0) = (\frac{3\pi}{4}, \frac{3\pi}{2})$. The algorithm described in section III is now employed with the sampling points of the SLSFT chosen according to a nearly uniform sampling scheme with $Q' = 32$ points [12]. A spatially concentrated window is used, designed to maximize the ratio between a function's energy in a polar cap and its total energy, with a cap width of $\frac{\pi}{3}$ and a window order of $N_h = 10$. The method for designing such a window is presented in [9].

Figure 2 shows the distortion for various values of the standard deviation of $\mathbf{v}_{nm}$, σ. While the spatial mask offers better performance for low values, the spatio-spectral mask improves as the value of σ increases. The low values are of less significance as no separation is needed for such low energy of the interference. Figure 3 shows the SNR for various values of σ, with the

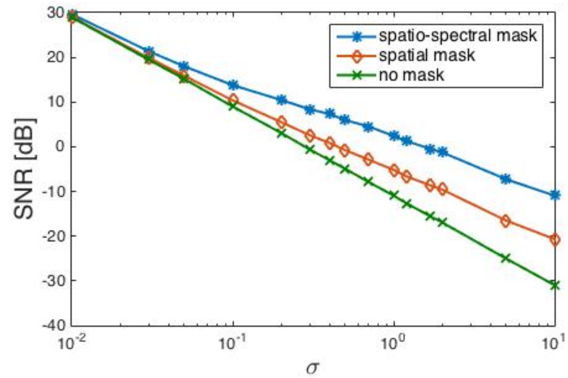

Fig. 3: SNR as a function of Gaussian noise standard deviation

SNR calculated without processing for comparison. Here the performance improves for each value of σ, for high values the difference in SNR reaches a scale of 10 dB, comparing to the spatial mask. The effect of this improvement can be seen in figure 4, showing a computation motivated by the DF [3]. In this calculation, (θ_0, ϕ_0) is the direction of the source's arrival, with the DF given by

$$DF = \frac{|\tilde{a}(\theta_0, \phi_0)|^2}{\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} |\tilde{a}(\theta, \phi)|^2 \sin\theta d\theta d\phi} \qquad (14)$$

This factor measures the amplitude of the peak of a function in comparison to its total energy. For small values of σ, $\mathbf{a}_{nm}$ is computed without distortion, and the DF achieves the value of the MD beamformer, with $(N+1)^2 = 121$. Note the significant improvement for high values of σ; this shows that the original goal, enhancement of $\mathbf{s}_{nm}$ while reducing $\mathbf{v}_{nm}$, has been achieved.

V. Conclusions and Future Work

This paper offers a novel approach, proposing processing of spatial signals as a preliminary stage of beamforming. This method uses the SLSFT to gain better separation of a signal from a spatial field than can be gained by operating only in the spatial or SH domain. Future work on this subject may include performance optimization in order to achieve different beamforming objectives. The optimization can include the choice of the window function, the sampling points of the SLSFT and the spatio-spectral mask.

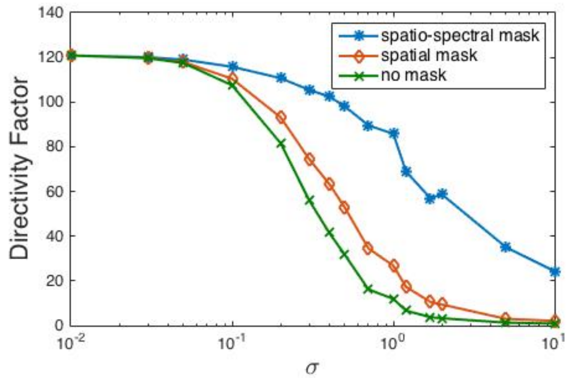

Fig. 4: Directivity factor of processed spatial field

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 609465.

REFERENCES

[1] W. Hansen and J. Woodyard, "A new principle in directional antenna design," *Proceedings of the Institute of Radio Engineers*, vol. 26, no. 3, pp. 333–345, 1938.

[2] C. Dolph, "A current distribution for broadside arrays which optimizes the relationship between beam width and side-lobe level," *Proceedings of the IRE*, vol. 34, no. 6, pp. 335–348, 1946.

[3] B. Rafaely, *Fundamentals of Spherical Array Processing*. Springer, 2015.

[4] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug 1969.

[5] H. L. Van Trees, *Detection, estimation, and modulation theory, optimum array processing*. John Wiley & Sons, 2004.

[6] Z. Khalid, P. Sadeghi, R. Kennedy, S. Durrani *et al.*, "Spatially varying spectral filtering of signals on the unit sphere," *Signal Processing, IEEE Transactions on*, vol. 61, no. 3, pp. 530–544, 2013.

[7] A. M. Reddy and B. Raj, "Soft mask methods for single-channel speaker separation," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 6, pp. 1766–1776, 2007.

[8] Z. Khalid, S. Durrani, P. Sadeghi, R. Kennedy *et al.*, "Spatio-spectral analysis on the sphere using spatially localized spherical harmonics transform," *Signal Processing, IEEE Transactions on*, vol. 60, no. 3, pp. 1487–1492, 2012.

[9] M. A. Wieczorek and F. J. Simons, "Localized spectral analysis on the sphere," *Geophysical Journal International*, vol. 162, no. 3, pp. 655–675, 2005.

[10] M. Simons, S. C. Solomon, and B. H. Hager, "Localization of gravity and topography: Constraints on the tectonics and mantle dynamics of venus," *Geophysical Journal International*, vol. 131, no. 1, pp. 24–44, 1997.

[11] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *The Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2149–2157, 2004.

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

[13] M. Brandstein and D. Ward, *Microphone arrays: signal processing techniques and applications*. Springer Science & Business Media, 2001.