DETECTION AND LOCALIZATION OF MULTIPLE WIDEBAND ACOUSTIC SOURCES BASED ON WAVEFIELD DECOMPOSITION USING SPHERICAL APERTURES

H. Teutsch

Multimedia Technologies Research Department
Avaya Labs
233 Mt Airy Rd, Basking Ridge, NJ 07920, USA
teutsch@research.avayalabs.com

W. Kellermann

Multimedia Communications and Signal Processing
University of Erlangen-Nuremberg
Cauerstr. 7, 91058 Erlangen, Germany
wk@LNT.de

ABSTRACT

This paper discusses novel methods for detecting and localizing multiple wideband acoustic sources using spherical apertures. In contrast to traditional methods the techniques presented here are not based on processing the output of individual microphones directly. Instead, the microphone signals are used to decompose the wavefield into its spherical harmonics which are subsequently used as a basis for novel frequency-independent multiple-source localization and detection methods.

Index Terms— Spherical microphone arrays, wavefield decomposition, spherical harmonics, source localization, source detection

1. INTRODUCTION

Spherical microphone arrays offer an ideal tool for capturing and analyzing three-dimensional wavefields. Consequently, researchers have been spending considerable effort to advance this relatively new and emerging technology [1, 2, 3, 4], thereby focusing mostly on the design and analysis of spherical microphone arrays as well as on beamforming. In all of the above references, the analysis of wavefields by spherical microphone arrays is not based on processing the sensor signals directly but rather on an orthonormal decomposition of the wavefield into spherical harmonics. In addition to the elegance of the mathematical framework, one of the main advantages of performing the analysis in the new 'modal' domain is the fact that the frequency-dependent components are decoupled from the angular-dependent components. This advantage over traditional array processing algorithms allows for a "fresh look" on related array processing problems such as source localization and detection. The number and position of wideband acoustic sources are of interest for, e.g. acoustic surveillance and for applications based on wavefield analysis and beamforming as a preprocessing step. In [5], the concept of using spherical harmonics for the localization of multiple wideband acoustic sources is introduced. This paper extends the analysis and provides simulations demonstrating its estimation performance. In addition, a source detection method is discussed.

The remainder of this paper is organized as follows. Section 2 describes the wavefield decomposition process using spherical apertures. Sections 3 and 4 detail the localization and detection methods, respectively. Finally, conclusions are drawn in Section 5. As a notational convention throughout this paper, all symbols in normal typeset denote scalar quantities while symbols and underlined symbols typeset in boldface are vector and matrix quantities, respectively.

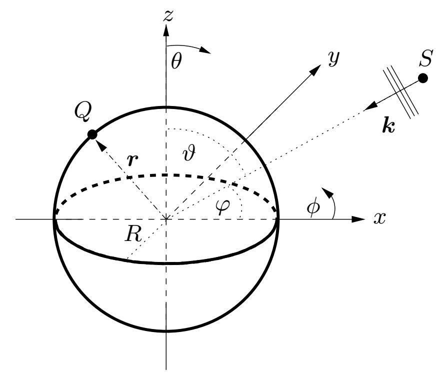

Fig. 1. *Planar wave front impinging on a spherical aperture.*

2. WAVEFIELD DECOMPOSITION

The model of a planar wave front impinging on a spherical aperture of radius R is shown in Fig. 1. A farfield source, S, generates a plane wave with wavenumber vector $\boldsymbol{k}$ at the observation point, Q, on the surface of the sphere.

Any square-integrable function on the sphere can be expanded into a series of spherical harmonics, see e.g. [6]. One such function is the plane wave measured on the surface of the sphere which can therefore be represented as (see e.g. [1])

$$e^{i\boldsymbol{k}^T\boldsymbol{r}}\Big|_{r=R} = 4\pi \sum_{n=0}^{\infty} i^n j_n(kR) \sum_{m=-n}^{n} Y_n^m(\theta,\phi) Y_n^m(\vartheta,\varphi)^*, \quad (1)$$

where $i^2 = -1$, $j_n(kR)$ is the n-th order spherical Bessel function, where $k = ||\boldsymbol{k}||$ is the wavenumber, and where the asterisk, '$*$', denotes complex conjugation. Furthermore,

$$Y_n^m(\theta,\phi) \triangleq \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}}\, P_n^m(\cos\theta)\, e^{im\phi}, \quad (2)$$

are the spherical harmonics of order n and degree m with $P_n^m(\cos\theta)$ being defined as the associated Legendre polynomial [6]. Applying the spherical Fourier transform [7] to Eq. (1), i.e. calculating the plane wave response of a pressure sensitive spherical aperture, results in the Fourier series expansion coefficients for a plane wave as [8]

$$F_n^m(kR,\varphi,\vartheta) = \sqrt{4\pi} i^n j_n(kR) Y_n^m(\vartheta,\varphi)^*. \quad (3)$$

Note that the frequency- and angular-dependent Fourier coefficients in Eq. (3) are commonly referred to as 'eigenbeams' [1].

3. EIGEN-BEAM BASED SOURCE LOCALIZATION

In this section it is shown how the eigenbeams in Eq. (3) can be used to estimate the direction of arrival (DOA) of multiple plane waves impinging on the surface of the sphere.

The algorithm is based on a recurrence relation for associated Legendre polynomials [6], i.e.

$$2m\cot\vartheta P_n^m(\cos\vartheta) = (m-n-1)(n+m)P_n^{m-1}(\cos\vartheta) - P_n^{m+1}(\cos\vartheta). \tag{4}$$

Note that this particular recurrence relation is based on a fixed order n of the associated Legendre polynomials.

To arrive at an algorithm that can be implemented on a digital signal processor, the infinite series in Eq. (1) needs to be truncated to a finite $n = N$. As a result, one obtains $2N+1$ eigenbeams for each order N. Now, let $\boldsymbol{F}(kR, \vartheta, \varphi)$ define a manifold vector for fixed order $n = N$ as

$$\boldsymbol{F}(kR, \vartheta, \varphi) = [F_N^{-N}, F_N^{-N+1}, \ldots, F_N^0, \ldots, F_N^N]^T, \tag{5}$$

where the superscript $'T'$ denotes transposition. Note that here and in the following functional dependencies have been dropped for notational convenience.

Now, three vectors of length $2N-1$ are extracted from the vector in Eq. (5) as

$$\boldsymbol{F}^{(l)}(\vartheta, \varphi) \triangleq \underline{\boldsymbol{\Delta}}^{(l)}\underline{\boldsymbol{D}}_0\boldsymbol{F}(\vartheta, \varphi), \quad l = [-1, 0, 1], \tag{6}$$

where $\underline{\boldsymbol{\Delta}}^{(-1)}$, $\underline{\boldsymbol{\Delta}}^{(0)}$, and $\underline{\boldsymbol{\Delta}}^{(1)}$ extract the first, middle, and last $2N-1$ elements from $\underline{\boldsymbol{D}}_0\boldsymbol{F}(\vartheta, \varphi)$ with

$$\underline{\boldsymbol{D}}_0 = \text{diag}\{(-1)^N, \ldots, (-1)^0, 1, \ldots, 1^N\}. \tag{7}$$

It can be shown that by considering I plane waves impinging on the spherical aperture, the recurrence relation, Eq. (4), can be expressed as

$$\underline{\boldsymbol{D}}_1\underline{\boldsymbol{F}}^{(0)} = \underline{\boldsymbol{D}}_2\underline{\boldsymbol{F}}^{(-1)}\underline{\boldsymbol{\Phi}} + \underline{\boldsymbol{D}}_3\underline{\boldsymbol{F}}^{(1)}\underline{\boldsymbol{\Phi}}^*, \tag{8}$$

where

$$\underline{\boldsymbol{F}}^{(l)} = [\boldsymbol{F}^{(l)}(\vartheta_1, \varphi_1) \mid \boldsymbol{F}^{(l)}(\vartheta_2, \varphi_2) \mid \ldots \mid \boldsymbol{F}^{(l)}(\vartheta_I, \varphi_I)], \tag{9}$$

with $l = [-1, 0, 1]$ and

$$\boldsymbol{\Phi} = \text{diag}\{\mu_1, \ldots, \mu_I\}, \tag{10}$$

where

$$\mu_\iota = \tan\vartheta_\iota \cdot e^{-i\varphi_\iota}, \ \iota = 1, 2, \ldots, I. \tag{11}$$

Also, with integer $-(N-1) \leq \nu \leq (N-1)$

$$\begin{aligned}\underline{\boldsymbol{D}}_1 &= 2\,\text{diag}\{|\nu| \cdot a_N^\nu\},\\ \underline{\boldsymbol{D}}_2 &= \text{diag}\{(\nu - N - 1)\cdot(N+\nu)\cdot a_N^{\nu-1}\},\\ \underline{\boldsymbol{D}}_3 &= \text{diag}\{a_N^{-(N-2)}, \ldots, a_N^0, -a_N^1, a_N^2, a_N^3, \ldots, a_N^N\},\end{aligned} \tag{12}$$

where $a_N^m = [(2N+1)/(4\pi)\cdot(N-m)!/(N+m)!]^{-1/2}$.

The remainder of the algorithm is very similar to the standard ESPRIT multiple-source localization method introduced in [9].

There, a signal subspace matrix is estimated by extracting the I principal eigenvalues from the power spectral density matrix [8]

$$\underline{\boldsymbol{S}}_{\boldsymbol{XX}} = \underline{\boldsymbol{F}}\underline{\boldsymbol{S}}_{\boldsymbol{SS}}\underline{\boldsymbol{F}}^H + \underline{\boldsymbol{S}}_{\boldsymbol{NN}}, \tag{13}$$

where $\underline{\boldsymbol{S}}_{\boldsymbol{NN}}$ is the matrix containing the noise power spectral densities. The matrix containing the signal power spectral densities, $\underline{\boldsymbol{S}}_{\boldsymbol{SS}}$, comprises both the desired sources as well as strong directional interferers. The superscript $'H'$ denotes the Hermitian operation.

Since, similar to the conventional array manifold matrix for sensor arrays, the eigenbeam manifold matrix

$$\underline{\boldsymbol{F}} \triangleq [\boldsymbol{F}(\vartheta_1, \varphi_1) \mid \boldsymbol{F}(\vartheta_2, \varphi_2) \mid \ldots \mid \boldsymbol{F}(\vartheta_I, \varphi_I)], \tag{14}$$

is related to the signal subspace matrix $\underline{\boldsymbol{U}}_S$ by a non-singular matrix $\underline{\boldsymbol{T}}$, it follows that $\underline{\boldsymbol{U}}_S^{(l)} = \underline{\boldsymbol{\Delta}}^{(l)}\underline{\boldsymbol{U}}_S, l = [-1, 0, 1]$.

The power spectral density matrix, Eq. (13), is estimated based on the decomposed output (spherical eigenbeams) of the spherical aperture. Equation (8) can then be expressed as

$$\underline{\boldsymbol{D}}_1\underline{\boldsymbol{U}}_S^{(0)} = \underline{\boldsymbol{E}}\begin{bmatrix}\underline{\boldsymbol{\Psi}}^T\\ \underline{\boldsymbol{\Psi}}^H\end{bmatrix}, \tag{15}$$

where,

$$\begin{aligned}\underline{\boldsymbol{E}} &= [\underline{\boldsymbol{D}}_2\underline{\boldsymbol{U}}_S^{(-1)} \mid \underline{\boldsymbol{D}}_3\underline{\boldsymbol{U}}_S^{(1)}], &(16)\\ \underline{\boldsymbol{\Psi}} &= \underline{\boldsymbol{T}}^{-1}\underline{\boldsymbol{\Phi}}\underline{\boldsymbol{T}}. &(17)\end{aligned}$$

By solving Eq. (15) in a total least-squares sense [10], an estimate for $\underline{\boldsymbol{\Psi}}$ can be obtained. Finally, by realizing that the complex eigenvalues of $\underline{\boldsymbol{\Psi}}$ are the entries of $\underline{\boldsymbol{\Phi}}$, the azimuth of the impinging plane waves, $\varphi_\iota, \iota = 1, 2, \ldots, I$, can be readily identified by the phase value of these eigenvalues. Similarly, the DOA in elevation, ϑ_ι, simply correspond to the inverse tangent of the magnitude of the eigenvalues, see Eq. (11). Several observations are of interest. Firstly, the number of spherical harmonics to be extracted from the wavefield must satisfy the relation $N \geq I + 1$. Secondly, source localization using eigenbeams obtained by wavefield decomposition along the surface of a spherical aperture is *inherently frequency-independent* which exhibits a significant advantage over 'traditional' ESPRIT and other subspace-based multiple-source localization algorithms. This result is a direct consequence of the fact that the frequency-dependent terms in $\underline{\boldsymbol{F}}$, i.e. $j_N(kR)$, cancel in Eq. (8).

For a real-world implementation, of course, one needs to sample the surface of the sphere at discrete microphone positions. Many sampling schemes have been recently proposed in the literature, see e.g. [1, 3, 4, 8, 11]. In the following only one particular sampling scheme is considered, i.e. the so-called 't-design' [11] method employing $M = 32$ microphones.

Figure 2 exemplifies the variance of the estimation error for one specific source-sensor-room scenario of the two-source (uncorrelated bandlimited white noise) DOA estimation algorithm in azimuth and elevation, respectively. Note that only the results for source 1 are reproduced. The results for the second source are very similar and provide no new insight. It is obvious that the variance of the estimator for spatially uncorrelated noise fields is significantly larger than the ideal estimator, i.e. the Cramer-Rao lower bound (CRLB) for spherical apertures [8]. Before interpreting the results in Fig. 2 it should be noted that the CRLB is really only meaningful for unbiased estimators in spatially uncorrelated white noise. A major contributor to the offset in estimation performance between the CRLB and the proposed method is the process of spherical sampling which results in a biased estimate due to the introduction of significant aliasing to the recorded spherical harmonics, which is not covered by the ideal model. Also depicted in Fig. 2 is the estimation error due to a varying amount of reverberation in a simulated room of size $(L \times W \times H) = (8 \times 8 \times 3)$ m modeled by a wall reflection coefficient β utilizing the image method [12]. It can be deduced

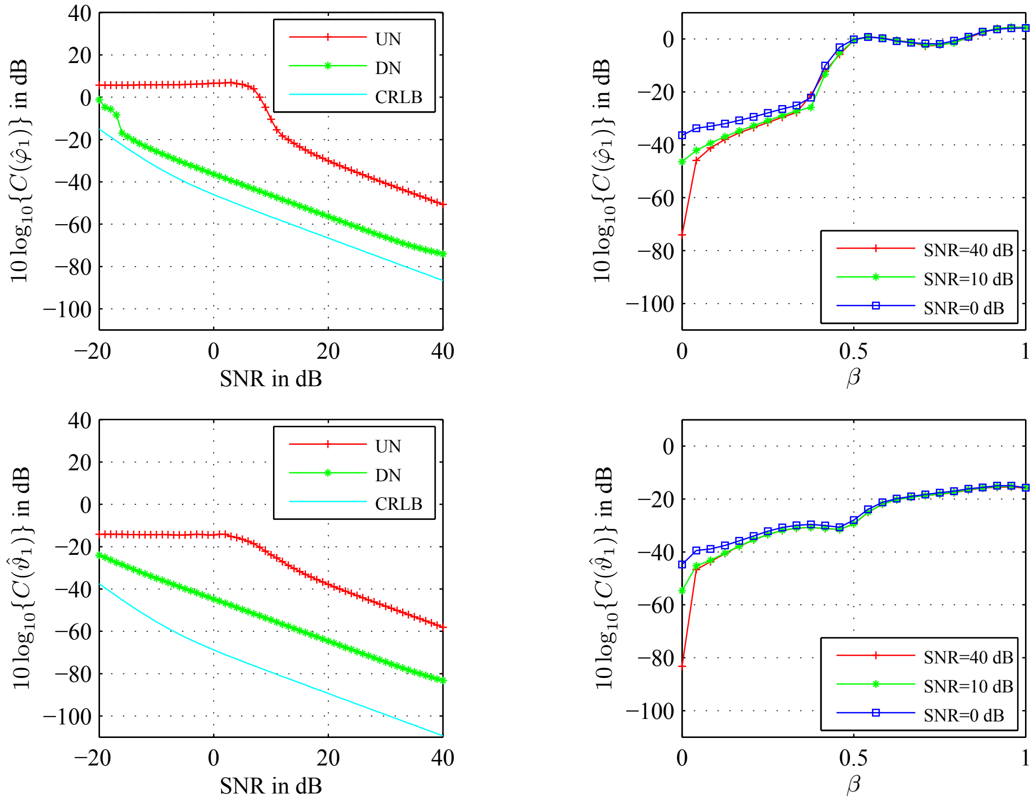

Fig. 2. *Estimation error variance, C, for two incident plane waves – uncorrelated bandlimited white noise – of equal power where $\boldsymbol{\varphi} = [\pi/6, \pi/2]^T$ (upper row) and $\boldsymbol{\vartheta} = [\pi/9, 5\pi/18]^T$ (lower row) [source one shown]. $R = 0.04$ m, $1\,kHz < f < 3\,kHz$, $M = 32$, $N = 3$, blocklength: 1024 samples, 300 independent trial runs. 'UN' and 'DN' denote uncorrelated and spherically diffuse white noise, respectively.*

that while the estimation algorithm does work satisfactorily for low to moderate SNR and small β, algorithms that incorporate a reverberated signal model should be considered for future developments. Future work will also include an in-depth performance analysis of the estimator. Note that the cotangent in Eq. (3) prohibits a source to be located at $\vartheta = \pi/2$. This problem can be alleviated by rotating the individual eigenbeams appropriately [1].

4. EIGEN-BEAM BASED SOURCE DETECTION

It has been tacitly assumed in Section 3 that the number of sources, I, necessary for estimating the signal subspace, $\underline{\boldsymbol{U}}_S$, is known a-priori. However, obtaining this a-priori knowledge is often unrealistic in practice. This section describes a method for source detection, i.e. providing an estimate for the number of active sources.

It follows for the superposition of plane waves on spherical aperture with Eq. (1) and $\boldsymbol{k}^T\boldsymbol{r} = kr[\sin\theta\sin\vartheta\cos(\phi-\varphi)+\cos\theta\cos\vartheta]$ that

$$P(kR,\theta,\phi,\boldsymbol{\vartheta},\boldsymbol{\varphi}) = \sum_{\iota=1}^{I} e^{ikR[\sin\theta\sin\vartheta_\iota\cos(\phi-\varphi_\iota)+\cos\theta\cos\vartheta_\iota]}$$
$$= 4\pi\sum_{\iota=1}^{I}\sum_{n=0}^{\infty} i^n j_n(kR)\sum_{m=-n}^{n} Y_n^m(\theta,\phi)Y_n^m(\vartheta_\iota,\varphi_\iota)^*, \tag{18}$$

where the symbols $\boldsymbol{\vartheta}$ and $\boldsymbol{\varphi}$ in Eq. (18), denoting the angular dependencies in elevation and azimuth, respectively, are the DOAs of I impinging plane waves, $\boldsymbol{\vartheta} = [\vartheta_1,\vartheta_2,\ldots,\vartheta_I]^T$ and $\boldsymbol{\varphi} = [\varphi_1,\varphi_2,\ldots,\varphi_I]^T$.

After performing the decomposition step, cf. Section 2, one obtains the Fourier series expansion coefficients, i.e. eigenbeams, for I plane waves due to the superposition principle as

$$G_n^m(kR,\boldsymbol{\vartheta},\boldsymbol{\varphi}) = \sqrt{4\pi}\sum_{\iota=1}^{I} i^n j_n(kR)Y_n^m(\vartheta_\iota,\varphi_\iota)^*. \tag{19}$$

Now, a wavefield *synthesis* operation that superimposes a finite number of eigenbeams is considered. This synthesis operation can be seen as an order-limited inverse spatial Fourier series expansion with respect to a reference point on the sphere (θ_0,ϕ_0). The result, after the sampling and wavefield truncation operations is

$$P(kR,\theta_0,\phi_0,\boldsymbol{\vartheta},\boldsymbol{\varphi}) = 4\pi\sum_{\iota=1}^{I}\sum_{n=0}^{N} i^n j_n(kR)$$
$$\times\sum_{m=-n}^{n} Y_n^m(\theta_0,\phi_0)Y_n^m(\vartheta_\iota,\varphi_\iota)^* + \mathcal{E}_\mathrm{s} + \mathcal{E}_\mathrm{t}. \tag{20}$$

For the sake of clarity, the error due to the truncation, $\mathcal{E}_\mathrm{t}$ and the error due to sampling, $\mathcal{E}_\mathrm{s}$, are assumed to be sufficiently small in

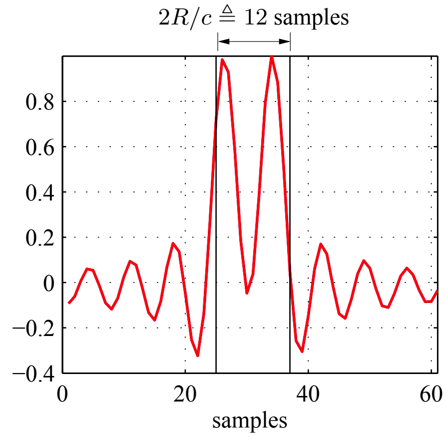

Fig. 3. *Normalized impulse response corresponding to a wavefield containing two equi-power uncorrelated and bandlimited plane waves where* $\boldsymbol{\vartheta} = [\pi/3, 4\pi/9]^T$ *and* $\boldsymbol{\varphi} = [\pi/6, 2\pi/3]^T$.

the following. The interested reader is referred to the literature, e.g. [8], for more details on these errors. With Eq. (18), Eq. (20) can be written as

$$P(kR, \theta_0, \phi_0, \boldsymbol{\vartheta}, \boldsymbol{\varphi}) = \sum_{\iota=1}^{I} e^{ikR[\sin\theta_0 \sin\vartheta_\iota \cos(\phi_0-\varphi_\iota)+\cos\theta_0\cos\vartheta_\iota]}$$

$$\approx 4\pi \sum_{\iota=1}^{I}\sum_{n=0}^{N} i^n j_n(kR) \sum_{m=-n}^{n} Y_n^m(\theta_0,\phi_0) Y_n^m(\vartheta_\iota,\varphi_\iota)^*. \quad (21)$$

In the time-domain, Eq. (21) reads

$$p(t, \boldsymbol{\tau}) = \sum_{\iota=1}^{I} \delta(t - \tau_\iota), \quad (22)$$

where with $\iota = 1, 2, \ldots, I$

$$\tau_\iota = \frac{R}{c}\left[\sin\theta_0 \sin\vartheta_\iota \cos(\phi_0 - \varphi_\iota) + \cos\theta_0 \cos\vartheta_\iota\right]. \quad (23)$$

In other words, by inverse Fourier transforming a re-synthesized wavefield that had been decomposed into its (order-limited) spherical harmonics, a simple source detection scheme now merely involves a count of the peaks in Eq. (23).

As an example, Fig. 3 depicts the discretized impulse response $p(t = \nu T, \boldsymbol{\tau})$, where ν is an integer and T is the sampling interval for two equi-power uncorrelated and bandlimited plane waves incident on a continuous aperture in free-field for $\boldsymbol{\vartheta} = [\pi/3, 4\pi/9]^T$ and $\boldsymbol{\varphi} = [\pi/6, 2\pi/3]^T$. The parameters are chosen as $R = 0.04$ m, $N = 5$, $f_s = 1/T = 48$ kHz, and $c = 340$ m/s. The sensor signals were bandlimited to .1 kHz $< f <$ 7 kHz.The reference point on the aperture was chosen to be $(\theta_0, \phi_0) = (\pi/2, 0)$. Also shown are two vertical lines that represent the maximum delays $\tau_{\text{max}} = \pm R/c$, that need to be considered in this detection scheme, cf. Eq. (23). I.e., peaks outside these boundaries can be disregarded. As can be seen, two distinct peaks corresponding to the two sources are clearly visible, which facilitate the detection algorithm for spherical apertures.

In order to be able to track moving sources, an adaptive algorithm is derived in [8].

5. CONCLUSIONS

In this paper, methods for detecting and localizing multiple wideband acoustic sources using spherical apertures based on a decomposition of the wavefield into spherical harmonics has been discussed. It has been shown that the localization of multiple plane waves exhibits acceptable results, especially for sources positioned in low to moderate spherically diffuse noise fields and mildly reverberated environments. For an extension of the presented methods to spherical microphone arrays mounted into rigid spherical baffles, the interested reader is referred to [8]. Future work include a performance comparison using several spherical sampling schemes including a detailed analysis of the source detection method in real acoustic environments and the explicit integration of a more realistic (reverberated) signal model.

6. REFERENCES

[1] J. Meyer and G.W. Elko, “A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield,” in *Proc. IEEE Int. Conf. Acoustics Speech, and Signal Processing*, May 2002, pp. 1781–1784.

[2] T.D. Abhayapala and D.B. Ward, “Theory and design of high order sound field microphones using spherical microphone array,” in *Proc. IEEE Int. Conf. Acoustics Speech, and Signal Processing*, May 2002, pp. 1949–1952.

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

[4] Z. Li and R. Duraiswami, “Flexible and optimal design of spherical microphone arrays for beamforming,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 702–714, February 2007.

[5] H. Teutsch and W. Kellermann, “Eigen-beam processing for direction-of-arrival estimation using spherical apertures,” in *Proc. Joint Workshop on Hands-Free Communication and Microphone Arrays*, March 2005, pp. c–13–c–14.

[6] G. Arfken, *Mathematical Methods for Physicists*, Academic Press, 3rd edition, 1985.

[7] J.R. Driscoll and D.M. Healy Jr., “Computing Fourier transforms and convolutions on the 2-sphere,” *Advances in Applied Mathematics*, vol. 15, pp. 202–250, June 1994.

[8] H. Teutsch, *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*, Springer, 2007.

[9] R. Roy and T. Kailath, “ESPRIT – estimation of signal parameters via rotational invariance techniques,” *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.

[10] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 2nd edition, 1989.

[11] R.H. Hardin and N.J.A. Sloan, “McLaren’s improved snub cube and other new spherical designs in three dimensions,” *Discrete Computational Geometry*, vol. 15, pp. 429–441, 1996.

[12] J. B. Allen and D.A. Berkeley, “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Am*, vol. 65, no. 4, pp. 943–950, April 1979.