

Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays

Dmitry N. Zotkin*, Ramani Duraiswami, Nail A. Gumerov

Perceptual Interfaces and Reality Laboratory

Institute for Advanced Computer Studies (UMIACS)

University of Maryland, College Park, MD 20742 USA

dz@umiacs.umd.edu, ramani@umiacs.umd.edu, gumerov@umiacs.umd.edu

Phone: (301)405-8753

Fax: (301)405-6707

Abstract

Spherical and cylindrical microphone arrays offer a number of attractive properties such as direction-independent acoustic behavior and ability to reconstruct the sound field in the vicinity of the array. Beamforming and scene analysis for such arrays is typically done using sound field representation in terms of orthogonal basis functions (spherical/cylindrical harmonics). In this paper, an alternative sound field representation in terms of plane waves is described, and a method for estimating it directly from measurements at microphones is proposed. It is shown that representing a field as a collection of plane waves arriving from various directions simplifies source localization, beamforming, and spatial audio playback. A comparison of the new method with the well-known spherical harmonics based beamforming algorithm is done, and it is shown that both algorithms can be expressed in the same framework but with weights computed differently. It is also shown that the proposed method can be extended to cylindrical arrays. A number of features important for the design and operation of spherical microphone arrays in real applications are revealed. Results indicate that it is possible to reconstruct the sound scene up to order p with p^2 microphones spherical array.

Index Terms

EDICS Categories: AUD-SMCA, AUD-LMAP.

I. INTRODUCTION

Particular configurations of microphones in a microphone array allow for elegant mathematical formulation of relevant signal processing algorithms. One such configuration is to place the microphones on the surface of a virtual sphere [4]; the spherical configuration leads naturally to an elegant mathematical framework based on elementary solutions of Helmholtz equation in spherical coordinates (i.e., spherical harmonics) and developed recently in a number of publications. Due to the 3-D symmetry of such a configuration, the array beampattern is independent of the steering direction and the spatial structure of the acoustic field can be captured without distortion. However, the free-field configuration is subject to numerical problems at certain frequencies [5]. An alternative configuration that does not have these problems is to place the microphones on the surface of a real, usually sound-hard sphere [6]; an additional benefit in this case is that the presence of a scattering object widens the useful frequency band of the array [7]. An overview of the practical design principles for spherical arrays was presented in [8]. A configuration mathematically quite similar to the spherical array is a cylindrical array; here, microphones are placed on the surface of a virtual or a real cylinder in a plane parallel to the base, and the scene analysis is done in terms of circular harmonics [9].

A desired operation on the acoustic field captured by a microphone array is to decompose it into components arriving from various directions. This decomposition is used in many practical applications such as sound source localization, signal enhancement for a direction of interest (beamforming), and spatial playback of captured auditory scenes [10]. The acoustic field in the vicinity of the array can be represented in various functional bases. The traditional representation, and the one in which source localization and beamforming is usually done, is based on spherical/circular harmonics. A modal spherical beamformer is proposed in [6]; the idea is to decompose both the sound field and the desired beamforming pattern into spherical harmonics; then, the beamforming weights can be found simply by dividing one by the other. An alternative and equally complete basis is a collection of plane waves [1], which is the basis used in the work presented here. An intuitive motivation for the work is to note that when a sound scene is represented in a plane wave basis, it, by the very nature of the representation, consists of components that arrive from various directions; hence, source localization and beamforming can be done by the mere act of converting the sound scene into a plane-wave representation.

An excellent review of signal processing with spherical/cylindrical arrays is done in the book [11], where various waveform estimation and parameter estimation problems are addressed in terms of spherical/cylindrical harmonic representations (“eigenbeams”, or modes). In the current paper, the same estimation problems are considered in terms of the plane wave representation. A somewhat similar approach is taken in [12], where a framework for performing plane-wave decomposition using spherical convolution

is presented. The framework in that paper uses a two-step decomposition process based on computation of spherical Fourier coefficients and assumes a continuous pressure-sensitive microphone array surface, which is not realizable in practice. An extension to practical arrays is done in [13], where simulations and an experimental evaluation of the ability of a spherical microphone array to analyze reverberant sound fields are presented. It is noted in [13] that the modal beamformer [6] is actually performing a decomposition into plane waves if the desired beampattern is a (truncated) Dirac delta function; hence, [13] is essentially a re-formulation and a very thorough analysis of [6] for the case of a special beampattern. In particular, both [6] and [13] replace integration over sphere surface by a quadrature over microphone positions, and those quadrature points must satisfy the discrete orthonormality equation for spherical harmonics. The number of microphones necessary for implementing the quadrature is a concern for practical implementation; in [8], it is shown that exact quadrature of order p requires $2p^2$ microphones arranged quite inconveniently (so-called Gaussian quadrature) or $4p^2$ microphones in more or less general arrangement. Manufacturing of an array that would use the Gaussian quadrature points is quite difficult because of the dense microphone arrangement near poles, and experiments in [13] use one microphone on the surface of a sound-hard sphere moved sequentially to all positions in Gaussian quadrature in order to capture a sound scene that is artificially repeated many times. Another weakness of [6] and [13] is the numerical instability of the beamforming weights, which are inversely proportional to the fast-decaying spherical mode strength.

In this paper, several contributions are made. First, an alternative derivation of the spherical beamforming algorithm of [6] and [13] is presented using Gegenbauer plane wave expansion, and it is shown that the algorithm can be implemented in one step as a product of a weight matrix and a vector of measured microphone potentials. Second, it is shown experimentally that a set of so-called Fliege points [14] provides a very good quadrature approximation with only p^2 points that are well-distributed over the sphere surface [15]. Third, a novel method for obtaining the plane-wave decomposition directly from the signals measured at microphones is presented. It is based on computing the decomposition coefficients via minimum least-squares fitting and can also be implemented as a matrix-vector product. The performance of the method is evaluated and compared with the performance of beamforming-based decomposition under realistic operating conditions such as in the case of an array with a finite number of microphones, in the existence of environmental noise, and in the presence of aliasing effects using both real and synthetic data. Finally, it is shown that the proposed method can be applied to a cylindrical array as well with minor notational changes in the equations. Simulated and experimental results with spherical, hemispherical, and cylindrical microphone arrays are presented.

II. BACKGROUND

Basis Functions: In a three-dimensional space with no acoustic sources enclosed, acoustic wave propagation at a wavenumber k is governed by the Helmholtz equation

$$\nabla^2 \psi(k, \mathbf{r}) + k^2 \psi(k, \mathbf{r}) = 0, \quad (1)$$

where $\psi(k, \mathbf{r})$ is the Fourier transform of the pressure, which is proportional to the velocity potential and will be loosely referred to as a potential in this paper. Solutions of the Helmholtz equation can be expanded as a series of spherical basis functions – the regular $R_n^m(k, \mathbf{r})$ in finite regions and the singular $S_n^m(k, \mathbf{r})$ in infinite regions:

$$R_n^m(k, \mathbf{r}) = j_n(kr)Y_n^m(\theta, \varphi); \quad S_n^m(k, \mathbf{r}) = h_n(kr)Y_n^m(\theta, \varphi), \quad (2)$$

where (r, θ, φ) are spherical coordinates of the radius vector \mathbf{r} , $j_n(kr)$ and $h_n(kr)$ are the spherical Bessel and spherical Hankel functions, respectively, and $Y_n^m(\theta, \varphi)$ are the orthonormal spherical harmonics. For later use, define also $J_n(kr)$ and $H_n(kr)$ to be the regular Bessel and Hankel functions, respectively. Note that the complex conjugation of $Y_n^m(\theta, \varphi)$ is equivalent to $Y_n^{-m}(\theta, \varphi)$ and that

$$\sum_{m=-n}^n Y_n^m(\theta_1, \varphi_1) Y_n^{-m}(\theta_2, \varphi_2) = \frac{2n+1}{4\pi} P_n(\mathbf{r}_1 \cdot \mathbf{r}_2) \quad (3)$$

(addition theorem for spherical harmonics), where $P_n(\mu)$ is the Legendre polynomial of degree n .

Any regular acoustic field $\Phi(k, \mathbf{r})$ in a region that does not contain sources can be represented as an infinite sum of regular functions with some complex coefficients $C_n^m(k)$ as

$$\Phi(k, \mathbf{r}) = \sum_{n=0}^{\infty} \sum_{m=-n}^n C_n^m(k) R_n^m(k, \mathbf{r}). \quad (4)$$

In practice, the outer summation is truncated:

$$\Phi(k, \mathbf{r}) = \sum_{n=0}^{p-1} \sum_{m=-n}^n C_n^m(k) R_n^m(k, \mathbf{r}). \quad (5)$$

The truncation number p depends on k and on the radius D of the region in which the approximation (5) is used to represent the field [16]. The truncated series (5) is called the (spatially) bandlimited representation, as spatial modes of order higher than p are not used. However, the spatial and temporal frequencies are interrelated; the upper limit on temporal frequency provides a lower limit on the wavelength and thus on spatial acoustic potential variations; therefore p is tied to k . It is shown in [16] that p^* determined as

$$p^* = \frac{ekD - 1}{2} \quad (6)$$

provides approximation consistent with typical signal quantization error due to series being truncated at p^* terms rather than at infinity. A more elaborate error analysis, which allows determination of p^* for given desired truncation error, is also available in [16].

Sphere Scattering: The solution for the problem of sound scattering by a sound-hard sphere was first given in [17]. Many authors list the solution for the case of finite-distance source [18] and for arbitrary regular acoustic field [16]. Assume that the plane wave $e^{i\mathbf{k}\cdot\mathbf{r}}$ is propagating in the direction $\mathbf{s} = (1, \theta, \varphi)$ and is scattered by a rigid sphere of radius a placed at the origin. The potential $\psi(\mathbf{s}, \mathbf{s}')$ created at an arbitrary observation point $\mathbf{s}' = (r', \theta', \varphi')$ is given by

$$\psi(k, \mathbf{s}, \mathbf{s}') = 4\pi \sum_{n=0}^{\infty} i^n b_n(ka, kr') \sum_{m=-n}^n Y_n^m(\mathbf{s}) Y_n^{-m}(\mathbf{s}') \quad (7)$$

($r' \geq a$). The coefficient $b_n(ka, kr')$ is known as the spherical mode strength [6]:

$$b_n(ka, kr') = j_n(kr') - \frac{h'_n(ka)}{j'_n(ka)} h_n(kr'). \quad (8)$$

Note that if the observation point is located on the surface of the sphere, one can use the Wronskian for Bessel functions [16] to simplify equations (7) and (8) to equivalent forms that are easier to compute

$$b_n(ka, ka) = \frac{i}{(ka)^2} \frac{1}{h'_n(ka)}, \quad (9)$$

$$\psi(k, \mathbf{s}, \mathbf{s}') = \frac{i}{(ka)^2} \sum_{n=0}^{\infty} \frac{i^n (2n+1) P_n(\mathbf{s} \cdot \mathbf{s}')}{h'_n(ka)}. \quad (10)$$

Note that \mathbf{s} in these equations is the direction of wave propagation (not direction of arrival) and $\psi(k, \mathbf{s}, \mathbf{s}')$ is largest when \mathbf{s} and \mathbf{s}' are opposing and the wave impinges on the sphere at the location of microphone (i.e., $\theta' = \pi - \theta$ and $\varphi' = \pi + \varphi$). Some authors (e.g., [18]) use \mathbf{s} as the direction towards the source, in which case equations become slightly different.

If an incident field is arbitrary and is described by equation (4), the potential for the microphone at \mathbf{s}' on the surface of the sound-hard sphere is [16]

$$\psi(k, \mathbf{s}') = \frac{i}{(ka)^2} \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{C_n^m(k) Y_n^m(\mathbf{s}')}{h'_n(ka)}. \quad (11)$$

Plane-Wave Expansion: Any regular acoustic field $\Phi(k, \mathbf{r})$ in a region can also be represented as a superposition of plane waves with each plane wave weighted by $\mu(k, \mathbf{s})$:

$$\Phi(k, \mathbf{r}) = \frac{1}{4\pi} \int_{S_u} \mu(k, \mathbf{s}) e^{i\mathbf{k}\cdot\mathbf{r}} dS(\mathbf{s}). \quad (12)$$

Note here that this representation requires quadrature over the sphere, i.e. over all possible directions \mathbf{s} ; $\mu(k, \mathbf{s})$ is known as the signature function [16] and fully specifies the field in the region. The multipole and the plane-wave representations can be converted to each other via the Gegenbauer expansion [21]:

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^n i^n Y_n^{-m}(\mathbf{s}) R_n^m(k, \mathbf{r}), \quad (13)$$

$$R_n^m(k, \mathbf{r}) = \frac{1}{4\pi} i^{-n} \int_{S_u} e^{i\mathbf{k}\cdot\mathbf{r}} Y_n^m(\mathbf{s}) dS(\mathbf{s}).$$

The signature function and the expansion coefficients are related as follows:

$$\begin{aligned}\mu(k, \mathbf{s}) &= \sum_{n=0}^{\infty} \sum_{m=-n}^n i^{-n} C_n^m(k) Y_n^m(\mathbf{s}), \\ C_n^m(k) &= i^n \int_{S_u} \mu(k, \mathbf{s}) Y_n^{-m}(\mathbf{s}) dS(\mathbf{s}).\end{aligned}\tag{14}$$

In practice, integration is replaced with summation over L quadrature points with quadrature weights $v(\mathbf{s}_l)$. The grid should be chosen appropriately to preserve the orthonormality of spherical harmonics when the integration over the surface is replaced by the summation over the grid [22]. The Gaussian quadrature was used in [13]. The procedure for supplying an alternate set of points proposed by Fliege [14] turns out to generate orthonormality-preserving grids [1] with equal quadrature weights. As such, in this paper Fliege grids are used for simulation (and the physical microphone arrays built for the experiments described below follow Fliege grids). Thus, in the discrete case equation (13) becomes

$$\begin{aligned}e^{i\mathbf{k}\mathbf{r}\cdot\mathbf{s}} &= 4\pi \sum_{n=0}^{p-1} \sum_{m=-n}^n i^n Y_n^{-m}(\mathbf{s}) R_n^m(k, \mathbf{r}), \\ R_n^m(k, \mathbf{r}) &= \frac{i^{-n}}{4\pi} \sum_{l=1}^L v(\mathbf{s}_l) Y_n^m(\mathbf{s}_l) e^{i\mathbf{k}\mathbf{r}\cdot\mathbf{s}_l}.\end{aligned}\tag{15}$$

These equations stipulate how the plane wave can be represented in the form (4) and correspondingly how a spherical mode can be represented in the form (12). Representation (15) is spatially bandlimited to $O(p^2)$, and p should be properly chosen to make the truncation error small.

Acoustic Image Principle: On a physical spherical microphone array, part of the sphere area is lost for cabling output. As such, microphones that would be positioned in that area are missing, which disrupts orthonormality of spherical harmonics. However, microphones on a *hemispherical* surface can easily be arranged in accordance with some regular (e.g. Fliege) grid. If a hemispherical array is then placed on an infinite reflective surface (such as a sufficiently large table), the resulting configuration can be treated as a spherical array with twice as many microphones [19] as follows.

Assume that an acoustic source is placed on one side of an infinite sound-hard plane. The acoustic image principle [20] states that at any point on the same side of the plane the acoustic potential is the sum of the potentials created by the source and by its reflection in the plane. Accordingly, in the case of the hemispherical microphone array mounted on an infinite rigid plane the potential $\psi_h(k, \mathbf{s}, \mathbf{s}')$ created at point \mathbf{s}' by a plane wave propagating in direction \mathbf{s} is given by summing up potentials due to two (original and reflected) plane waves given by equation (10):

$$\psi_h(k, \mathbf{s}, \mathbf{s}') = \psi(k, \mathbf{s}, \mathbf{s}') + \psi(k, \tilde{\mathbf{s}}, \mathbf{s}'),\tag{16}$$

where $\tilde{\mathbf{s}}$ is \mathbf{s} reflected in the array base plane; specifically, if $\mathbf{s} = (\theta, \varphi)$, then $\tilde{\mathbf{s}} = (\pi - \theta, \varphi)$. Because of the same image principle, the potential at each image microphone created by reflecting the corresponding real array microphone in the base plane is equal to the potential at the real one. Thus, the number of microphones is essentially doubled. Of course, with a hemispherical array any measured set of potentials is symmetric around the horizontal plane and the solved acoustic scene is consequentially symmetric as well. Care should be taken to disregard the reflected part of the scene ($\theta > \pi/2$).

Cylinder scattering: The decomposition framework described herein is also applicable to the cylindrical array. In this case, the wave propagation direction \mathbf{s} is determined by single angle φ and observation point \mathbf{s}' is defined by (r', φ') in cylindrical coordinates. The potential $\psi_c(k, \mathbf{s}, \mathbf{s}')$ at \mathbf{s}' due to that plane wave (“c” stands for cylinder) is

$$\psi_c(k, \mathbf{s}, \mathbf{s}') = \sum_{n=-\infty}^{\infty} i^n B_n(ka, kr') e^{-in(\varphi-\varphi')}, \quad (17)$$

where a is the array radius. $B_n(ka, kr')$ is a cylindrical mode strength and is given by

$$B_n(ka, kr') = J_n(kr') - \frac{H'_n(ka)}{J'_n(ka)} H(kr'). \quad (18)$$

When $r' = a$ (i.e., the point is located on the cylinder surface), $B_n(ka, ka)$ simplifies to

$$B_n(ka, ka) = \frac{2i}{\pi ka} \frac{1}{H'_n(ka)} \quad (19)$$

and $\psi_c(k, \mathbf{s}, \mathbf{s}')$ becomes dependent only on angles φ and φ' :

$$\psi_c(k, \mathbf{s}, \mathbf{s}') = \psi_c(k, \varphi, \varphi') = \frac{2i}{\pi ka} \sum_{n=-\infty}^{\infty} \frac{i^n}{H'_n(ka)} e^{-in(\varphi-\varphi')}. \quad (20)$$

Note the great similarity in the equations between the spherical and cylindrical cases. In practice, the summation is truncated to $2p - 1$ terms. Numerical simulations show a behavior of the truncation error similar to the spherical array case.

III. ACOUSTIC SCENE DECOMPOSITION

Common framework: The goal of scene analysis is to decompose the scene into pieces arriving from various directions. A beamformer allows one to pick up the signal arriving from a given direction; hence, the traditional way of performing such a decomposition is to perform a number of beamforming operations with a (truncated) Dirac delta function as a desired beam pattern [6]. The decomposition obtained in this way is referred to as the beamforming decomposition. An alternative method, and the one that is explored in this paper, is to find, through solving a system of linear equations, the set of plane waves that “best explains” the observed potential distribution at the microphones. The decomposition obtained is this way is referred to as the least-squares decomposition. Once the decomposition is obtained by either method, it

can be used for various purposes, including source localization using steered response power, auralization of a component coming from a specific direction (i.e., beamforming), visualization of acoustic energy distribution in space, and further analysis of specific scene components.

Assume that microphones in the array are arranged on an ‘‘M-grid’’ (‘‘microphone’’) over L_q directions \mathbf{s}'_q . The goal is to measure the potentials $\Psi(k, \mathbf{s}'_q)$ at those microphones and to recover the (complex) magnitudes of scene components $\lambda(k, \mathbf{s}_j)$ over L_j directions \mathbf{s}_j comprising an ‘‘S-grid’’ (‘‘source’’). Here and later, index q iterates over microphone positions and index j iterates over a set of plane wave directions. Also, for simplicity of derivations, it is assumed that M-grid has uniform quadrature weights. Denote by Λ the $L_j \times 1$ vector of unknown $\lambda(k, \mathbf{s}_j)$, by W the $L_j \times L_q$ matrix of weights, and by Ψ the $L_q \times 1$ vector of measured potentials $\Psi(k, \mathbf{s}'_q)$. In these terms, the goal is to form matrix W consisting of weights $w(k, \mathbf{s}_j, \mathbf{s}'_q)$ and to compute $\Lambda = W\Psi$ in one step. Note that everything is done here in the frequency domain. In practice, time-domain signals are recorded at microphones, and directional scene components should be produced as waveforms in time domain; conversion between real time-domain signals and frequency domain is covered in Section IV.

Beamforming (BF) Decomposition [6]: A beamformer allows one to pick up the signal arriving from a given direction, and a number of such independent beamformers can be used to look simultaneously in L_j directions. In this case, the weight $w(k, \mathbf{s}_j, \mathbf{s}'_q)$ is expressed as

$$w(k, \mathbf{s}_j, \mathbf{s}'_q) = \sum_{n=0}^{p-1} \frac{1}{i^n b_n(ka, ka)} \sum_{m=-n}^n Y_n^{-m}(\theta_j, \varphi_j) Y_n^m(\theta'_q, \varphi'_q), \quad (21)$$

Using equation (3) and equation (9), this can be simplified to

$$w(k, \mathbf{s}_j, \mathbf{s}'_q) = -i \frac{(ka)^2}{4\pi} \sum_{n=0}^{p-1} i^{-n} h'_n(ka) (2n+1) P_n(\mathbf{s}_j \cdot \mathbf{s}_q). \quad (22)$$

Alternative Derivation ([1] equation (22)): The weight matrix W relating Ψ to Λ can be computed directly using Gegenbauer expansion of a plane wave. First, note that by definition of plane-wave field representation the signature function $\mu(k, \mathbf{s})$ represents weights assigned to individual plane waves that compose the scene; hence, the desired $\lambda(k, \mathbf{s})$ is identical to the signature function $\mu(k, \mathbf{s})$. It follows from equation (11) that

$$C_n^m(k) = -i(ka)^2 h'_n(ka) \int_{S_u} \psi(k, \mathbf{s}') Y_n^{-m}(\mathbf{s}') dS(\mathbf{s}'). \quad (23)$$

The expression for $\mu(k, \mathbf{s})$ from equation (14) then becomes

$$\begin{aligned} \mu(k, \mathbf{s}) &= -i(ka)^2 \sum_{n=0}^{\infty} i^{-n} h'_n(ka) \times \\ &\quad \times \int_{S_u} \psi(k, \mathbf{s}') \sum_{m=-n}^n Y_n^m(\mathbf{s}) Y_n^{-m}(\mathbf{s}') dS(\mathbf{s}') \end{aligned} \quad (24)$$

and is further reduced to

$$\mu(k, \mathbf{s}) = -i \frac{(ka)^2}{4\pi} \sum_{n=0}^{\infty} (2n+1) i^{-n} h'_n(ka) \int_{S_u} \psi(k, \mathbf{s}') P_n(\mathbf{s} \cdot \mathbf{s}') dS(\mathbf{s}') \quad (25)$$

using the addition theorem for spherical harmonics. In the discrete case, the summation is truncated to p terms and the integral is replaced by quadrature over points \mathbf{s}'_q with $\mu(k, \mathbf{s}_j)$ becoming

$$\mu(k, \mathbf{s}_j) = -i \frac{(ka)^2}{4\pi} \sum_{n=0}^{p-1} (2n+1) i^{-n} h'_n(ka) \sum_{q=1}^{L_q} \Psi(k, \mathbf{s}'_q) P_n(\mathbf{s}_j \cdot \mathbf{s}'_q). \quad (26)$$

Regrouping gives the W matrix as

$$w(k, \mathbf{s}_j, \mathbf{s}'_q) = -i \frac{(ka)^2}{4\pi} \sum_{n=0}^{p-1} (2n+1) i^{-n} h'_n(ka) P_n(\mathbf{s}_j \cdot \mathbf{s}'_q), \quad (27)$$

which in fact is exactly equivalent to the weights of Meyer-Elko beamformer except that the use of the Wronskian has simplified the expression for $b_n(ka, kr')$. This derivation, in fact, shows that the BF decomposition can be viewed as an act of sampling the plane-wave representation signature function corresponding to a specific direction.

Least-squares (LS) Decomposition ([1] equation (19)): This method is essentially based on computing Λ that best explains the observed Ψ . The potential at each microphone is just the sum of the potentials created by all oncoming plane waves; thus, given Λ , one can compute Ψ as

$$\Psi = F\Lambda, \quad (28)$$

where F is $L_q \times L_j$ matrix with entries $F(k, \mathbf{s}'_q, \mathbf{s}_j)$:

$$F(k, \mathbf{s}'_q, \mathbf{s}_j) = \frac{i}{(ka)^2} \sum_{n=0}^{p-1} \frac{i^n (2n+1) P_n(\mathbf{s}_j \cdot \mathbf{s}'_q)}{h'_n(ka)}, \quad (29)$$

which is just equation (10) truncated to p terms. This linear system can be solved for Λ . If $L_j = L_q$, then $W = F^{-1}$, and if $L_j < L_q$, then the system is overdetermined and is solved in the least squares sense so that $W = (F^T F)^{-1} F^T$. The LS method can be thought of as a way to perform simultaneous separation of a scene into a collection of directional components for pre-determined set of directions, much in the same way as a number of sequential beamforming operations covering the same set of directions, but in parallel and with different formulation of the separation matrix.

Hemispherical Array: In case of hemispherical array, the matrix $F(k, \mathbf{s}'_q, \mathbf{s}_j)$ used in the LS decomposition should be replaced with the matrix $F_h(k, \mathbf{s}'_q, \mathbf{s}_j)$:

$$F_h(k, \mathbf{s}'_q, \mathbf{s}_j) = F(k, \mathbf{s}'_q, \mathbf{s}_j) + F(k, \mathbf{s}'_q, \tilde{\mathbf{s}}_j) \quad (30)$$

as in equation (16). In addition, the S-grid should cover only the upper hemisphere (the other one contains image sources). The BF decomposition requires no modifications for use with the hemispherical array except that the set of look directions covering only the upper hemisphere should be used as well.

Cylindrical array: In the cylindrical array case and assuming the same notation as before, the matrix W for the BF decomposition is given by [11]

$$w_c(k, \varphi_j, \varphi'_q) = 2\pi \sum_{n=-(p-1)}^{p-1} i^{-n} B_n^{-1}(ka) e^{-in(\varphi_j - \varphi'_q)}, \quad (31)$$

which can be simplified using Equation (19) to

$$w_c(k, \varphi_j, \varphi'_q) = -i\pi^2 ka \sum_{n=-(p-1)}^{p-1} i^{-n} H'_n(ka) e^{-in(\varphi_j - \varphi'_q)}. \quad (32)$$

For LS decomposition, the linear system $\Psi = F\Lambda$ is formed and is solved $\Lambda = W\Psi$ in the same manner as for the spherical array case, where F elements are of form $F_c(k, \varphi'_q, \varphi_j)$ given by Equation (20) truncated to $2p - 1$ terms:

$$F_c(k, \varphi'_q, \varphi_j) = \frac{2i}{\pi ka} \sum_{n=-(p-1)}^{p-1} \frac{i^n}{H'_n(ka)} e^{-in(\varphi_j - \varphi'_q)}. \quad (33)$$

IV. PRACTICAL IMPLEMENTATION

Assume that the time-domain signal recorded at a pressure-sensitive microphone located at \mathbf{s}'_q is $x_q(t)$ and that the sampling frequency is f_s ; denote $k_s = 2\pi f_s/c$, where c is the sound speed. For the block of the signal of length N , compute the Fourier transform at each microphone; the result has $N/2$ complex Fourier coefficients at wavenumbers $k_s/N, 2k_s/N, \dots, k_s/2$ (assume that the DC offset for the block is zero). The decomposition is performed separately at each wavenumber k ; note that the matrix W is different for different k . The decomposition coefficients $\lambda(k, \mathbf{s}_j)$ are computed by doing the matrix vector-product $\Lambda = W\Psi$ multiplication as described above; the potential $\Psi(k, \mathbf{s}'_q)$ is simply the Fourier coefficient for microphone at \mathbf{s}'_q at wavenumber k . Then, the output time-domain waveform $y_j(t)$ for the direction \mathbf{s}_j is obtained by assuming that the computed $\lambda(k, \mathbf{s}_j)$ is the Fourier coefficient of the output signal at wavenumber k and performing inverse Fourier transform of the set of $\lambda(k, \mathbf{s}_j)$.

In practice, time-overlapped smoothly fading windows are used to eliminate windowing artifacts such as clicks occurring on the window boundary. Also, the practical spherical/cylindrical array has a limited useful frequency band, which is determined from above by spatial aliasing and from below by array size and the equipment noise floor (if the acoustic wave length is substantially larger than the array size, the difference in potentials at different microphones is quite small and may be too small to detect in presence of quantization or in electronic components noise). In this case, the computations are done only for wavenumbers corresponding to useful frequency band and the rest of the $\lambda(k, \mathbf{s}_j)$ coefficients are zeroed out for the inverse Fourier transform.

Also, note that Equations (10) and (20) assume that \mathbf{s} (φ , respectively) is the direction of propagation, not the direction towards the source, and that this notation is kept through the rest of the equations in this

paper. Accordingly if one wishes to obtain the sound field component arriving from directions \hat{s}_j (i.e., beamform in the direction \hat{s}_j), then he/she should use the *opposite* direction ($\theta_j = \pi - \hat{\theta}_j$, $\varphi_j = \pi + \hat{\varphi}_j$) in computing the weight matrix W .

V. SIMULATION SETUP

Spherical array: A simulation of source localization using steered response power [23] with both BF and LS decompositions was performed. The array radius a was set to 0.106 m. The M-grid was set to be 64-point Fliege grid [1] [14], referred to as “64F” from now on. The spatial Nyquist criteria states that that intermicrophone distance should be less than half the wavelength in order to avoid spatial aliasing; the Nyquist frequency for the simulated array was about 3.85 kHz.

A legitimate question is how to construct the “best” S-grid for the plane-wave decomposition for a given array configuration. The S-grid should cover the space well and provide good quadrature. The reasonable choice is to use the same grid as M-grid; setting $L_j = L_q$ in fact also assures no information will be lost during decomposition (in other words, the number of degrees of freedom in the scene recording and in the decomposed scene representation are the same). In this work, two S-grids were tested: first was identical to the microphone grid (“64F”) and second was a 49-point Fliege grid (“49F”), which also covers the whole space evenly, provides good quadrature, and has smaller number of points and hence lesser computational load but decreased accuracy.

Simulations were performed for random source directions at frequencies from 0.5 to 6 kHz with 0.5 kHz step size. For each frequency, each source direction, and each noise variance, the potentials at microphone locations were computed using equation (10) with $p = p_m$. Then, each potential was synthetically corrupted by Gaussian noise with zero mean and given variance, and either BF or LS decomposition was applied to solve for $\lambda(k, \mathbf{s})$ over the S-grid using $p = p_s$. In the plots, p_m and p_s are given in terms of p^* computed as prescribed by equation (6) (note that p^* is different for different k and therefore for different frequencies). The direction in S-grid corresponding to the largest magnitude of $\lambda(k, \mathbf{s})$ was taken to be the detected direction of the source. The error measure was the angular difference, in radians, between the “true” source direction and the detected direction. Root mean square error (RMSE) was computed over 1024 trials (i.e., 1024 source directions) for each frequency, each noise variance, and each method. In addition, the noise tolerance of algorithms was tested with three fixed noise variance values of 0, 0.5, and 1.0. Note that computed potentials typically have magnitude around one, and one can think of the measurements being corrupted with noise of magnitude of zero, 50%, and 100% of the measurements themselves.

In addition, a simulation experiment was performed for an acoustic scene consisting of *two* plane waves arriving from different random directions. The goal of this experiment was to see whether the spatial resolution of two methods differs. Each direction in the pair was selected randomly and independently

of the other one; thus, it is possible that in some pairs the directions were too close to each other in order to be successfully resolved. The potential at each microphone was set to be the sum of potentials generated by two plane waves (of equal magnitudes) arriving from their respective directions. After computing the signature function, a simple repeated gradient ascent algorithm was used to find the two largest isolated peaks in the S-grid, and those two peaks were taken to be the detected directions for the sources. Localization error was then computed for each source and averaged over two sources. Note that two detected directions D_1 and D_2 can be matched to two ground-truth source directions S_1 and S_2 in two ways (D_1 to S_1 and D_2 to S_2 or vice versa); the matching giving the lower overall localization error was chosen. The experiment was also repeated 1024 times, and RMSE was computed.

Hemispherical array: Simulations with the hemispherical array show substantially the same findings as for the spherical array case presumably because the underlying theory is very similar. They are not reported here for lack of additional information.

Cylindrical array: An infinite sound-hard cylindrical array of radius 0.101 m with 16 equally-spaced microphones was simulated for localization experiments. The estimated spatial aliasing frequency was 4.3 kHz and simulations were done for frequencies from 0.5 to 8.0 kHz with 0.5 kHz step. Simulated plane wave was impinging on the array from a random direction and the potentials at the microphones were computed using Equation (20) truncated at $2p - 1$ terms with $p = p_m$. These potentials were fed either to BF decomposition or to LS decomposition with $p = p_s$ to compute the coefficients $\lambda(k, \mathbf{s})$. The S-grid was identical to the M-grid. The direction maximizing $|\lambda(k, \mathbf{s})|$ was selected as simulated localization direction. The error was defined as an angular distance between the simulated wave direction and the localization direction. Simulations were repeated 1024 times, and RMSE was computed.

VI. EXPERIMENTAL SETUP

Several sets of experiments with real spherical, hemispherical, and cylindrical arrays were performed in a large (approximately $5.5 \times 4.9 \times 2.75$ m) office room. Most of the room wall and ceiling area was covered with acoustic absorbing foam (“egg crate” foam) to reduce reverberation. A typical computer speaker (Harman/Kardon) was used to produce test signals. Two test signals were used: a continuous sine wave of a given frequency (ranging from 0.5 to 6 kHz with 0.5 kHz step size) and a 2.46 ms long upsweep (chirp) signal containing the same frequencies. For the sine wave signal, a CD containing 12 test signals was made, and the test signals were played via a CD player connected to the speaker and recorded digitally at 12 bits via two 32-channel NI PCI-6071E data acquisition boards. For the chirp signal, the signal was output to the speaker via the analog output channel of one of the data acquisition boards. The analog output and analog input subsystems of both boards were synchronized to run off a hardware clock common for both boards to ensure repeatability of the experiment and to allow for time averaging.



Fig. 1. Spherical (left), hemispherical (middle), and cylindrical (right) microphone arrays used. The hemispherical array is mounted on a circular table.

Spherical array experiment: The 60-microphone spherical array, made out of a hollow plastic lampshade of radius 0.101 m with wall thickness of about 1.5 mm, was placed in the center of the room on a tripod. The array is shown in Figure 1 on the left. The arrangement of the microphones in the array followed 64-point Fliege grid [14] with nodes 12, 24, 29, and 37 removed due to the need to accommodate the cable outlet. The resulting grid constituted an M-grid for the spherical array experiments, referred to as “60F” later on. The Nyquist frequency for the array was about 3.85 kHz. Four positions in the room, all at a distance of about 1.5 m from the array, were chosen to place the source. The positions were selected to roughly encircle the array in azimuth and to represent a variety of elevations. The angular coordinates of each position were determined by visually projecting them onto the spherical array surface. The loudspeaker was placed at each of those positions and the test signals were played and recorded. A 2 s recording was done for each of the sine wave signals. The chirp signal was repeated 10 times with one-second pauses between chirps to minimize reverberant noise. The recorded signal was then time-averaged over these 10 trials. The recorded signal (both sine wave and chirp) was then windowed with 5.0 ms rectangular or Hann window and the potential at each frequency of interest (0.5 to 6 kHz with 0.5 kHz step size) was computed via Fourier transform. The potentials were used as inputs to the two solution methods. The direction with the largest $\lambda(k, s)$ magnitude in the S-grid was taken to be the detected source direction. Two S-grids were used: “60F” and “49F”.

Hemispherical array experiment: The 64-microphone hemispherical array, made out of a half of a bowling ball of radius equal to 0.109 m, was mounted in the center of a circular table of radius equal to 0.457 m. The array is shown in Figure 1 in the middle. The microphone locations for the array were obtained as described in [15], forming a “64H” grid. The Nyquist frequency for the array is about 4.66 kHz. The table with the array was placed in the center of the room. Four random positions in the upper hemisphere of the array were chosen for source placement, covering various azimuths and elevations as well. The rest of the experimental setup is the same as for the spherical array experiment.

As the S-grid for the hemispherical array shall cover only upper hemisphere, S-grid different from the spherical array setup had to be used. The first S-grid was “64H” (same as the microphone grid for the hemispherical array). The second S-grid was the 121-point Fliege grid modified to remove all points having negative z , resulting in a 62-point “62H” grid.

Cylindrical array: The 16-microphone cylindrical array was made from a large empty stainless steel canister with the radius of 0.113 m and the height of 0.279 m. The microphones were mounted on half-height line at regular intervals covering the circumference of the array, forming a “16C” grid. The array is shown in Figure 1 on the right. The Nyquist frequency for the array is about 3.9 kHz. The array was placed in the center of the room on a tripod as shown in the picture. Five random positions were chosen for source placement, all in horizontal plane in elevation and roughly encircling the array in azimuth. The test signals used in the setup are the same as for the spherical array case except that they span wider frequency range (up to 8 kHz). The decomposition grid (S-grid) for the cylindrical array was chosen to be the same as the microphone grid (i.e., the decomposition is done over the same directions in which measurements are done).

VII. RESULTS

In the following subsections, the results obtained in the simulations and in the experiments are described. Those include varying the truncation number, adding regularization, and changing test signal, signal windowing function, and noise variance.

A. General noise tolerance

Somewhat surprisingly, in experiments it was observed that the performance degradation of both algorithms at highest noise magnitude is not significant. That means that isotropic noise with magnitude comparable to the signal (i.e., the SNR is about zero dB) does not substantially interfere with localization. The reason for that is not clear; the significant redundancy provided by 64 available measurements is likely playing a role. To avoid clutter in the plots, the results related to noise tolerance are not presented beyond this paragraph, and all plots are made for zero-noise case. A future work is planned to assess noise and reverberation tolerance in more details.

B. Simulation results, spherical array

Dependence on the truncation number: In this section, the plots of results for simulated spherical “64F” microphone array and for “49F” and “64F” S-grids are presented. In each figure, decomposition method and S-grid used are shown in the plot legend. In all plots presented, $p_m = 10p^*$ to accurately

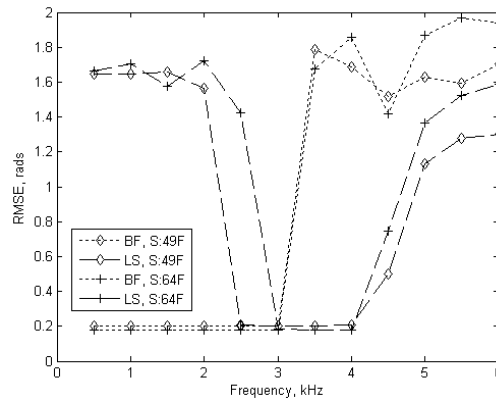


Fig. 2. Simulation, $p_m = 10p^*$, $p_s = p^*$. S-grid used and decomposition method are annotated in the plot (BF: beamforming decomposition, LS: LS decomposition).

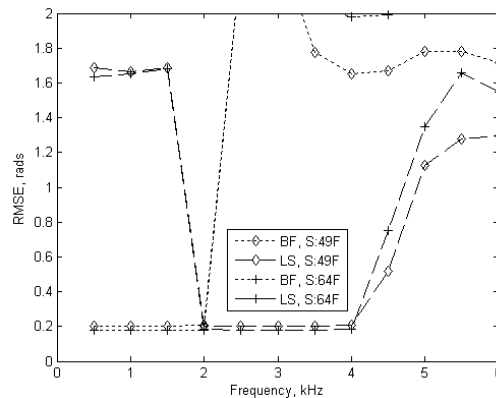


Fig. 3. Simulation, $p_m = 10p^*$, $p_s = \frac{3}{2}p^*$. See Figure 2 legend for abbreviations.

reflect the potentials that would be observed in real system (where $p_m = \infty$). Figure 2 shows the RMSE plots obtained with $p_s = p^*$. The breakdown of error into azimuth and elevation components shows that they are substantially equal (plots not presented for reasons of space) due to full 3-D symmetry of the setup. The error lower bound is not zero because of the discrete grid nature. For the 49-point S-grid BF decomposition shows good behavior for up to 3 kHz, whereas LS decomposition does exactly the opposite – localization is random up to 2.5 kHz and is perfect above that. When the frequency is increased beyond the spatial aliasing limit (above approximately 4 kHz), the LS decomposition performance gracefully degrades. For the 64-point S-grid BF decomposition shows the same behavior (because it computes the signature function for each direction in S-grid directly from the potential measurements) and the LS decomposition operating range starts at slightly higher frequency.

The next two plots show the RMSE behavior for two methods when p_s is changed. In Figure 3 the case of $p_m = 10p^*$ and $p_s = \frac{3}{2}p^*$ is plotted, and Figure 4 shows the RMSE plots obtained for $p_m = 10p^*$

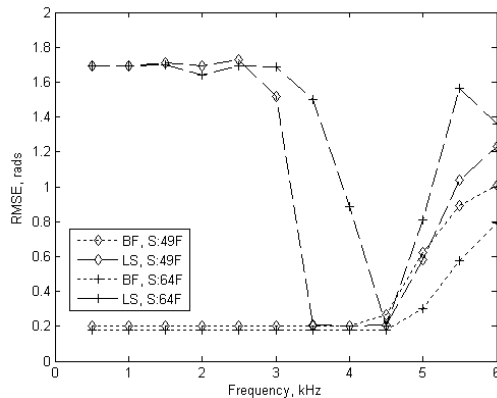


Fig. 4. Simulation, $p_m = 10p^*$, $p_s = \frac{3}{4}p^*$. See Figure 2 legend for abbreviations.

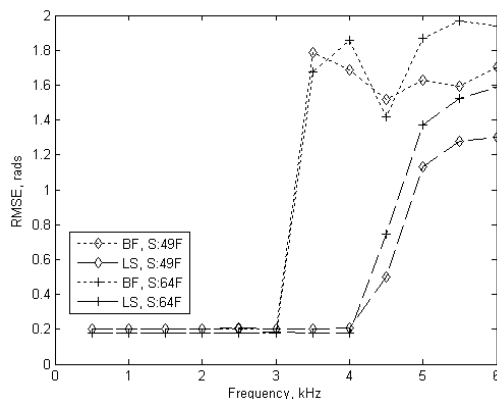


Fig. 5. Simulation, $p_m = 10p^*$, $p_s = p^*$, regularization $\varepsilon = 10^{-2}$. See Figure 2 legend for abbreviations.

and $p_s = \frac{3}{4}p^*$ (in the latter case, care should be taken to ensure p_s stays above zero at low frequencies (e.g., by enforcing $p_s \geq 1$ constraint)). By comparing these plots with Figure 2, it can be noted that the higher the p_s is, the lower is the frequency at which two events happen: a) BF decomposition breaks down and b) LS decomposition starts to work. If p_s is increased further (to $2p^*$) or decreased further (to $\frac{1}{2}p^*$), the same trends are observed (corresponding plots are not shown here for space reasons). Note that with $p_s = \frac{3}{4}p^*$ BF decomposition is successfully operating up to the spatial frequency limit of the array and the error increases very gradually above this limit. On the other hand, LS decomposition working range seems to decrease as p_s is decreased.

Regularization: Analysis of the matrix F (equation (29)) shows that it is poorly conditioned (but is not singular) in the low frequency region where the LS method shows high localization error and that the conditioning improves when p_s is increased, which is consistent with the plots above. Therefore, a regularization term was added in an attempt to improve algorithm's performance so that when $L_j = L_q$, $W = (F + \varepsilon I)^{-1}$ and when $L_j < L_q$, $W = (F^T F + \varepsilon I)^{-1} F^T$, where I is the identity matrix and ε is the

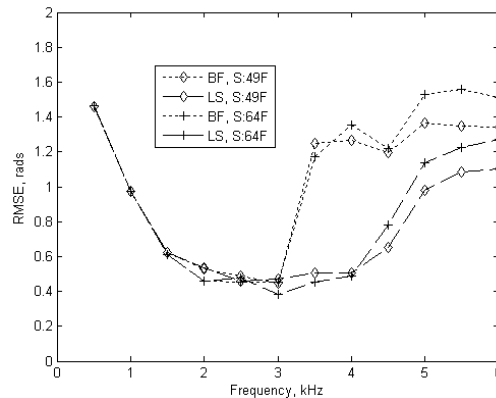


Fig. 6. Simulation, $p_m = 10p^*$, $p_s = p^*$, $\varepsilon = 10^{-2}$, simultaneous localization of two plane wave sources. See Figure 2 legend for abbreviations.

regularization constant. Figure 5 shows the obtained RMSE plots with $p_m = 10p^*$, $p_s = p^*$, and $\varepsilon = 10^{-2}$. It can be seen that inclusion of the regularization terms allows for successful LS method application over the whole operating range of array (in fact, error plots of BF and LS decompositions are identical up to about 2.5 kHz). Also, the particular value of ε ranging from 10^{-7} to 1.0 only marginally influences the results.

Spatial resolution comparison: In this simulation, two plane waves are presented to the array from different directions, and successful localization of both sources is sought. After solving for M using either BF or LS decomposition, two largest peaks in $\lambda(k, s_j)$ magnitude over S-grid were found, localization error was computed for both sources, and the error absolute value was geometrically averaged. In Figure 6, the localization RMSE is shown. It is seen that the localization is significantly hampered at low frequencies. Examination of the actual $\lambda(k, s)$ values plotted over the S-grid (not shown here for the reasons of space) for various frequencies reveals that the width of the peak created on the S-grid by a source is large for low frequencies so that there is more chance of two peaks merging and appearing as one broad peak. (Alternatively, this can be explained via the well-known fact that the width of the beam pattern for the spherical array is very large at low frequencies). Above approximately 1.5 kHz, two waves are resolved successfully in most of the trials. For the sake of consistency, in these plots $p_m = 10p^*$, $p_s = p^*$, and $\varepsilon = 10^{-2}$; if p_s is reduced for BF decomposition as described above, the low-error frequency regions of BF and LS decompositions coincide. The lowest observed error is still higher than for one source case because directions of two plane waves are chosen randomly and therefore could be closer to each other than the distance between points in S-grid. A plot of the likelihood of resolving two sources versus the angular distance between those shows results consistent with the above explanations and is not shown for the reasons of space.

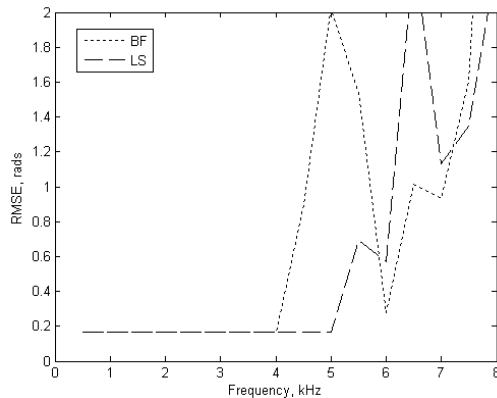


Fig. 7. Simulation, cylindrical array, $p_m = 10p^*$, $p_s = p^*$, regularization $\varepsilon = 10^{-2}$.

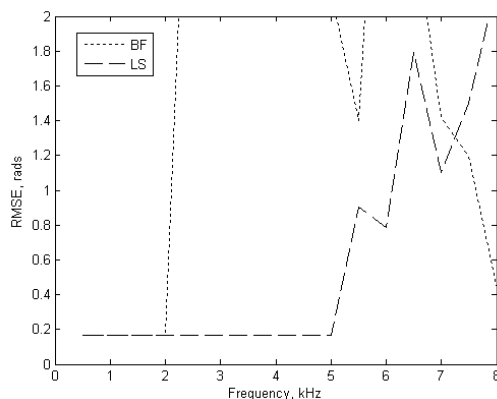


Fig. 8. Simulation, cylindrical array, $p_m = 10p^*$, $p_s = 3p^*/2$, regularization $\varepsilon = 10^{-2}$.

C. Simulation results, cylindrical array

Simulations quite similar to the spherical array case were run for the 16-microphone cylindrical array. In all plots presented in this section, $p_m = 10p^*$ as well and regularization with $\varepsilon = 10^{-2}$ is used for LS algorithm. Figure 7 shows the RMSE for both methods when $p_s = p^*$. It can be seen that the BF algorithm performance starts to degrade at about 4.5 kHz and the LS algorithm works fine up to about 5.5 kHz. The latter number actually exceeds the spatial aliasing limit of 4.3 kHz. Poor BF performance at higher frequencies is consistent with results on spherical array.

The case of $p_s = 3p^*/2$ is shown in Figure 8. It can be seen that the BF method breaks down at about 2.5 kHz and the LS results are unchanged. If p_s is increased further (plot not shown), the same effect becomes more pronounced.

Another simulated experiment was done with $p_s = 3p^*/4$. The corresponding error plots are shown in Figure 9. It can be seen that the operational range of BF method is widened and is in fact the same now as the operational range for LS method, with error increasing gradually about 5.5 kHz. Overall, the

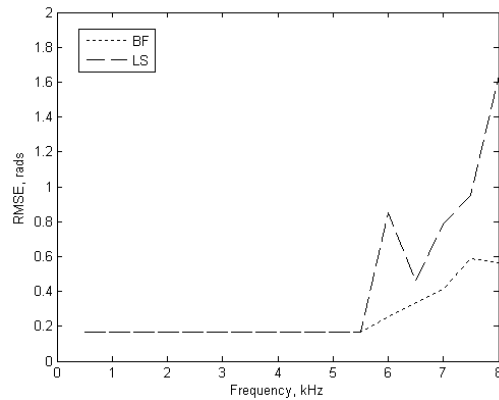


Fig. 9. Simulation, cylindrical array, $p_m = 10p^*$, $p_s = 3p^*/4$, regularization $\varepsilon = 10^{-2}$.

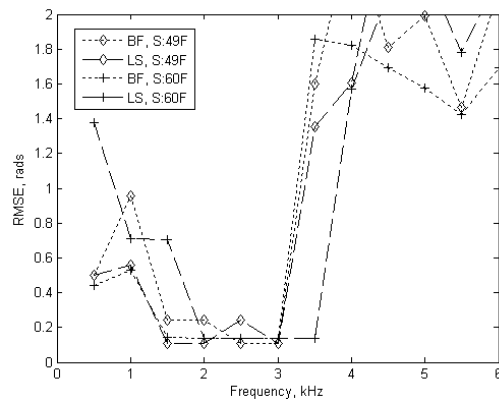


Fig. 10. Spherical microphone array experiment, sine wave signal, data frame includes signal onset.

obtained behavior is quite consistent with the same observed in simulations with the spherical array.

D. Experimental results, spherical array

The experiment with the real “60F” spherical array was designed and carried out as described earlier. No substantial differences were observed in the results when the signal windowing function (rectangular or Hann) was changed and when ε was varied within the same range as in simulations, suggesting that regularization has “binary” effect (unless ε is set to unreasonably high value). Therefore, only the results obtained with rectangular windowing function are presented below for “49F” and “60F” S-grids. In all plots below, $p_s = p^*$ and $\varepsilon = 10^{-2}$.

Sine wave signal: Figure 10 demonstrates the localization performance obtained when the data frame was selected to include the start of the wave signal so that reverberation effects are minimized. The operating frequency of the array appears to be from about 1.5 to about 3.5 kHz, and acceptable results are obtained below 1.5 kHz. The BF decomposition exhibits earlier performance degradation similar to that observed in simulations. In Figure 11, a data frame is selected in the middle of the sine wave signal

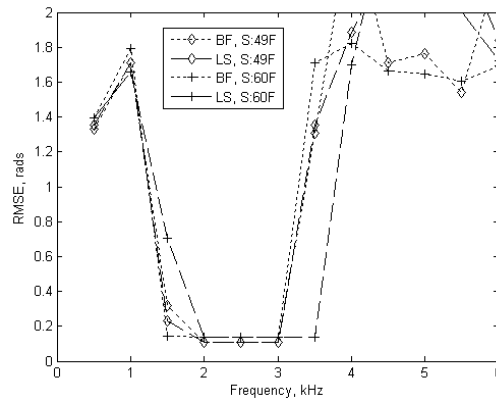


Fig. 11. Spherical microphone array experiment, sine wave signal, data frame is selected in the middle of the signal.

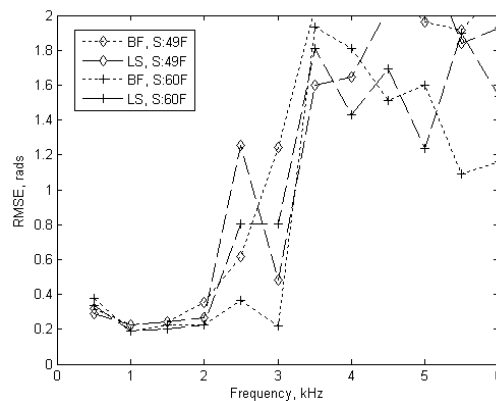


Fig. 12. Spherical microphone array experiment, chirp signal.

playback so that presumably more reverberation is present. A difference is indeed seen at low frequencies, where the algorithm is more influenced by noise (and reverberation) due to the signal wavelength being comparable to the array size.

Chirp signal: Figure 12 shows localization performance for the chirp signal repeated 10 times with 1 s pauses and time-averaged. In this plot, no significant difference is observed between BF and LS methods. Note the low error in the low-frequency range; this is consistent with earlier plots showing increase in localization accuracy in low-frequency range when there is no reverberation mixed in the signal. In addition, the experiment was repeated 16 times for a total of 160 repetitions of the chirp signal time-averaged in an attempt to increase signal-to-noise ratio. This produced no difference in results compared to Figure 12.

E. Experimental results, hemispherical array

The same experiments were repeated with the hemispherical array made of half a bowling ball and mounted on a circular table. Again, the particular windowing of the signal did not appear to change the

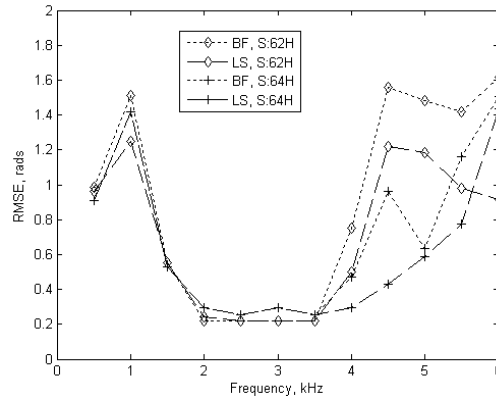


Fig. 13. Hemispherical microphone array experiment, sine wave signal, data frame includes signal onset.

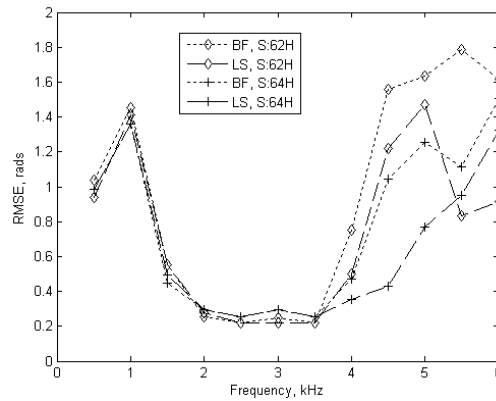


Fig. 14. Hemispherical microphone array experiment, sine wave signal, data frame is selected in the middle of the signal.

results significantly, so the results obtained with rectangular windowing are reported. In all plots below, $p_s = p^*$ and $\varepsilon = 10^{-2}$. Results with “62H” and with “64H” S-grids are presented.

Sine wave signal: Similar to the experiment with the spherical array, localization of the sine wave signal was performed for two data frames, first for the data frame that included the start of the signal (Figure 13) and second for the data frame in the middle of the signal (Figure 14). Presumably the acoustic field in the second case was corrupted by reverberation to some extent. The general structure of the plots is similar to the ones seen for the spherical array. The range of good localization is extended to higher frequencies (up to about 4.5 kHz). Unlike the spherical array, localization at low frequencies is always poor (this is also seen in the RMSE plot for the chirp signal below).

Chirp signal: This experiment was done with the chirp signal time-averaged over 10 repetitions. Figure 15 presents the localization RMSE versus frequency for the chirp signal. The range of good localization is consistent with the sine wave signal, with high error in the low frequency range. The experiment was also repeated 16 times for a total of 160 repetitions of the chirp signal time-averaged, which produced

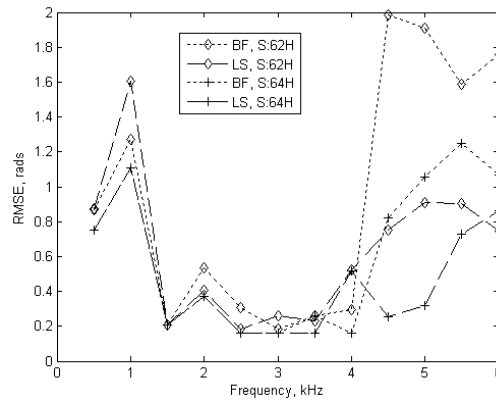


Fig. 15. Hemispherical microphone array experiment, chirp signal.

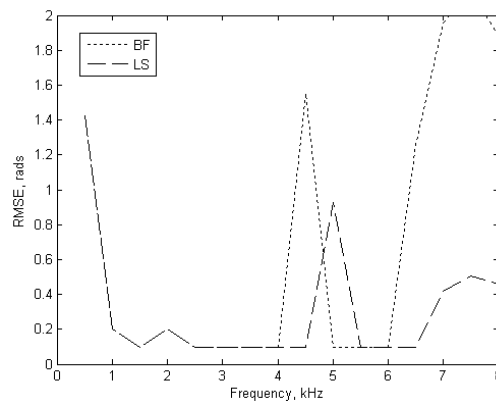


Fig. 16. Cylindrical microphone array experiment, sine wave signal, data frame includes signal onset.

no changes in results compared to averaging over 10 repetitions only.

F. Experimental results, cylindrical array

The same experiments were repeated for the 16-microphone cylindrical array placed approximately in the middle of the room on a tripod. The results obtained with rectangular windowing of the recorded signals in computation of the potentials are reported. No significant changes in results were observed when Hann window was used. In the three plots below, $p_s = p^*$ and $\varepsilon = 10^{-2}$. The decomposition grid (S-grid) coincides with the microphone grid (M-grid).

Sine wave signal: For the sine wave signal experiment, the data frame for processing was selected either to include the onset of the signal or to be from the middle of a one-second long signal burst. The corresponding RMSE plots are shown in Figure 16 and Figure 17, respectively. The trends observed in the plots are similar to those for the spherical and hemispherical arrays. Note the relatively low error at low frequencies in both cases. Note also that the algorithms are working well above the Nyquist frequency and that the working range of the LS method is slightly wider.

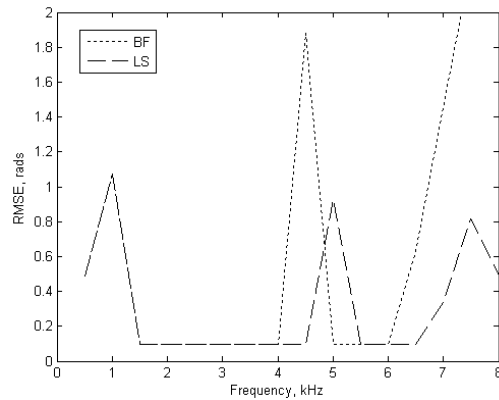


Fig. 17. Cylindrical microphone array experiment, sine wave signal, data frame is selected in the middle of the signal.

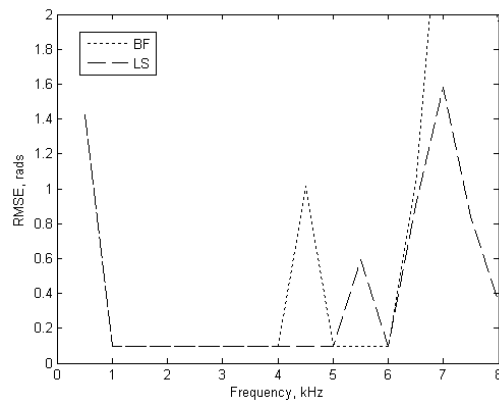


Fig. 18. Cylindrical microphone array experiment, chirp signal.

Chirp signal: In the chirp signal experiment, the test signal is repeated 10 times with one-second pause between the pulses and time-averaged before processing to eliminate reverberation and noise effects. The error behavior seen is substantially the same as for the sine wave signal (Figure 18). The peaks at about 4.5 kHz visible for both the sine wave and the chirp may be due to internal resonances of the hollow array structure. No change in the plot was observed when averaging was done over 160 repeats of the chirp signal instead of 10.

VIII. DISCUSSION

The discussion here is presented in terms of the spherical array, which is more important in practice; however, substantially the same conclusions can be derived for cylindrical array applications.

Numerical stability: To compare two methods, it may be helpful to repeat here the equations governing weight computation. Omitting constant factors and some multipliers, it can be schematically written that

for BF

$$W = \sum_{n=0}^{p-1} \frac{1}{i^n b_n(ka)} P_n(s \cdot s'), \quad (34)$$

whereas for LS $W = F^{-1}$ where F is

$$F = \sum_{n=0}^{p-1} i^n b_n(ka) P_n(s \cdot s'). \quad (35)$$

(for cylindrical array, $B_n(ka)$ would be used instead of $b_n(ka)$). While these two equations are seemingly similar, the methods are quite different in behavior. In BF method, the reciprocal of mode strength is taken for each n independently; this is a significant numerical weakness of the BF decomposition because mode strength rapidly approaches zero when $n > ka$; hence, the beamforming weights diverge as $p \rightarrow \infty$ and one needs to be extremely careful in choosing p so as not to cause amplification of white noise. In practice, the equipment (microphone/amplified/ADC) noise floor, along with mode strength magnitude plots such as Figure 1 of [6], are used in determining p for computing weights. Trying to use higher p in search for larger signal separation results in total loss of desired signal as weights diverge (or, in other words, the system noise is amplified to the point of losing the desired signal). Note that p varies with k (i.e., with frequency).

In contrast, in LS the summation over n is done first and only after that is the whole matrix inverted. Therefore, individual elements in matrix F converge as $p \rightarrow \infty$ and p can be set arbitrarily high without causing numerical problems (in fact, the convergence happens very quickly because of the same exponential decay of $b_n(ka)$ when $n > ka$). However, the opposite problem exist in LS: at low frequencies, the matrix F becomes ill-conditioned because a low frequency plane wave creates only marginal potential differences between microphones. Still, a conclusion can be derived that setting p too large for LS decomposition causes no harm and actually somewhat improves the conditioning of F matrix due to inclusion of higher-order modes, whereas doing the same for BF decomposition causes totally meaningless output due to taking reciprocals of extremely small values. This is confirmed by experimental results. Furthermore, it can be hypothesized that due to the fact that the truncation number can be set arbitrarily high in the LS algorithm without affecting its numerical stability, the decomposition achieved with LS method is more physically correct because the wave scattering is modeled more accurately. However, simulation results show that BF decomposition and LS decomposition are behaving substantially equivalently over the useful array frequency range and either method can successfully extract directional information from the one-source or two-source acoustic field presented to the array. With regard to selection of a particular S-grid to be used for LS decomposition, it can be said that the error plots presented in the paper show no substantial differences between all shown S-grids if the array operating range is limited to

the frequencies below the spatial aliasing frequency. Some differences do occur at the frequencies above the spatial aliasing limit, where error exhibited by BF decomposition appears to be lower.

Simulations versus experiment comparison: The experimental plots for the spherical array show narrower range of good localization performance compared to the simulation plots with the same array. In particular, in simulations good performance is observed up to the Nyquist frequency of the array (about 4 kHz), whereas in real experiments the performance starts to degrade at about 3 or 3.5 kHz. One of the reasons could be the fact that the array is made of a hollow plastic sphere that could absorb sound whereas the theoretical framework assumes sound-hard array support. Another observation is that the presence of reverberation harms the array localization significantly at low frequencies.

The experimental plots for the hemispherical array show higher upper limit due to denser microphone spacing and to the sound-hard sphere used in the construction. However, the localization performance at low frequencies is poor due to the array being mounted on a relatively small table (remember that the theoretical foundation assumes that the array is placed on an infinite, sound-hard plane), which causes deviations from acoustic image principle for sources of longer wavelength. For reference, the frequency of the sound for which the wavelength is equivalent to the table radius is 750 Hz. It can be expected that if the array were mounted directly on a floor or on a wall the localization performance at lower frequencies would be significantly improved.

Spatial resolution: The achievable sound field order, the number of microphones in the array, and the spatial resolution of the array are all related. One way to view this relationship is to consider the implications of representing the arbitrary acoustic scene in a plane-wave basis. If a representation without loss of information and without redundancy is desired, then the number of directions in S-grid should be the same as in M-grid. The number of microphones and the number of decomposition directions are also related to the field order in terms of spherical harmonics; the field of order p has p^2 decomposition coefficients in Equation (5). This gives a rough idea of the relationship between the number of microphones on the sphere, the sound field order, and the highest frequency supported by the microphone array. For example, with $L = 64$ modes of order up to 8 can be resolved, and if $D = a = 0.106$ m, then $k \approx 70$ and $f \approx 4$ kHz. Spatial aliasing occurs above that frequency. The same estimation can be obtained by considering that the beamwidth for a given truncation order is fixed and is approximately equal to the angular intermicrophone spacing [13]; hence, no additional information about the scene could be gained by using S-grid finer than M-grid because the beams thrown would overlap substantially.

A perceived limitation of the LS-method is that the number of directions in S-grid cannot be more than a number of microphones, which seemingly limits the spatial resolution of the method. The BF method technically does not have that limitation, and BF output can be produced for any direction. However,

due to the same beam width argument, the information content of the plane-wave decomposition is equal to that of the original scene even with S-grid being the same as M-grid, and beamforming to a lot of directions can produce more accurate location information but still cannot resolve two sources that are close to each other. If accurate localization is desired with LS-method, interpolation between the positions in S-grid can be performed; this is not done in this paper.

Truncation number selection: With regard to the truncation number, it should be mentioned that the truncation number $p_s = p^*$ suggested in [16] and considered optimal in [1] is too high for use in real spherical and cylindrical beamforming applications as numerical problems arise even in simulations. Lower truncation number such as $p_s = ka$ used by some authors would be more applicable for BF method. In contrast, with LS method increase of p_s does not lead to localization error increase. Also note that the BF kernel is actually a product of two spherical harmonics of order n (see equation (24)) and therefore has order $2n$. Thus, given 64 microphones in the array, the discrete orthonormality condition is guaranteed to be satisfied only up to the order of about 4, and it is a lucky coincidence that the Fliege point grid has low orthonormality errors at higher p_s .

An important conclusion can be derived regarding the number of microphones necessary for successful recovery of the scene spatial structure up to order p with the spherical array. Earlier work [8] suggested that $2p^2$ is the minimum number necessary for exact integration; however, the experimental results presented here confirm that p^2 microphones arranged over Fliege grid provide adequate quadrature approximation. When less than p^2 microphones are used, spatial aliasing occurs and the scene structure recovery ability inevitably degrades [1].

Spatial aliasing limit: In some RMSE plots presented, localization accuracy is declining quite gradually beyond the spatial aliasing limit. The sharpness of error increase depends on the truncation number and on the particular grid used. This is not surprising since it is known that beamforming pattern degradation happens rather slowly when the frequency is raised above the Nyquist frequency [24]. Hence, proper choice of an S-grid may allow the array designer to go slightly beyond the spatial aliasing limit and still obtain acceptable performance. Note that while points in M-grid shall satisfy the discrete orthonormality property necessary for bridging the gap between equations (25) and (27), there is no such requirement for the points in S-grid. Still, it appears that the regularity of the S-grid plays a role in how fast the localization error increases once the upper frequency limit of the array is reached. From the error plots presented, it appears that a good answer for the “best” S-grid question is to have S-grid identical to M-grid (in other words, to decompose the scene into a set of plane waves in the same directions as the directions of microphone locations on the sphere surface).

Applications: In addition to considered source localization problem, another application of the LS

decomposition is beamforming. It is especially suitable for the case where *all* directional components of the field are to be computed (e.g., for immersive audio rendering application [25]), as opposed to the case where one is interested primarily in listening selectively in *one* direction. It is not clear whether BF or LS decomposition provides better signal separation, either in terms of SNR or perceptually; however, it can be hypothesized that the beamforming performance would be consistent with the localization performance, as localization approach described in this work can be thought of as a number of beamforming operations followed by selecting a component with largest magnitude. Informal listening experiments show that indeed the signal separation achieved between two speakers in real acoustic recording done with a spherical array of Figure 1 is about the same for BF and LS methods for $p_s = 3p^*/4$; when p_s is raised to p^* , the separation appears to be slightly improved for LS method and noise dominates the output signal for BF method. Detailed simulations and experiments are out of the scope of this paper and are planned for the future.

Finally, note that for efficient implementation both BF and LS methods can be realized as a matrix-vector multiplication using the same processing engine because the matrix of weights does not depend on data being processed and can be precomputed for different k and given M-grid and S-grid. A very fast implementation of the engine was recently done on GPU [26] for performing visualization of distribution of acoustic energy in the space by beamforming, in real time, in 8192 directions and plotting the component magnitude as a pixel intensity. The GPU implementation is two orders of magnitude faster than the CPU implementation because the matrix-vector multiplication is ideally suited for highly-parallelized framework of GPU programming.

IX. CONCLUSIONS AND FUTURE WORK

The study presented suggests some principles for the array design and development of plane-wave decomposition algorithms. It was shown that the proposed LS decomposition algorithm can be formulated in the same framework as “classical” spherical harmonics based beamforming algorithm but with differently computed weight matrix. Some advantages of the LS decomposition is simpler computation of the weight matrix and ability to employ higher truncation number than in BF decomposition. Simulated experiments and real experiments were performed with one acoustic source for spherical, hemispherical, and circular microphone arrays. Simulations with two acoustic sources were also performed. Effective frequency bands of the arrays were validated in the experiments, and it was shown that within those no significant performance differences are found between BF and LS decompositions. When the frequency was raised above the Nyquist frequency for the array, the slowest-growing error was observed with LS decomposition using an S-grid identical to the M-grid. In a sense, the obtained results are as good as one can wish for – the effective frequency band is reasonably consistent with theoretical prediction and good localization is demonstrated in the effective band. Future planned work will further evaluate accuracy

and robustness of the two methods and will compare them in the task of capture and reconstruction of arbitrary spatial audio fields for human listeners using spherical microphone arrays.

X. ACKNOWLEDGEMENTS

We would like to thank the U.S. Department of Veterans Affairs for funding this work. We are also thankful to Zhiyun Li and to Elena Grassi for building the equipment involved in the experiments described herein.

REFERENCES

- [1] R. Duraiswami, Z. Li, D. N. Zotkin, E. Grassi, and N. A. Gumerov (2005). "Plane-wave decomposition analysis for the spherical microphone arrays", Proc. IEEE WASPAA 2005, New Paltz, NY, pp. 150-153.
- [2] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov (2008). "Sound field decomposition using spherical microphone arrays", Proc. IEEE ICASSP 2008, Las Vegas, NV, April 2008, pp. 277-280.
- [3] D. N. Zotkin and R. Duraiswami (2009). "Plane-wave decomposition of a sound scene using a cylindrical microphone array", IEEE ICASSP 2009, Taipei, Taiwan, R.O.C., April 2009, in print.
- [4] T. Abhayapala and D. Ward (2002). "Theory and design of high order sound field microphones using spherical microphone array", Proc. IEEE ICASSP 2002, Orlando, FL, vol. 2, pp. 1949-1952.
- [5] T. D. Abhayapala (2008). "Generalized framework for spherical microphone arrays: spatial and frequency decomposition", Proc. IEEE ICASSP 2008, Las Vegas, NV, April 2008, pp. 5268-5271.
- [6] J. Meyer and G. Elko (2002). "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield", Proc. IEEE ICASSP 2002, Orlando, FL, vol. 2, pp. 1781-1784.
- [7] P. Gillett, M. Johnson, and J. Carneal (2008). "Performance benefits of spherical diffracting arrays versus free field arrays", Proc. IEEE ICASSP 2008, Las Vegas, NV, April 2008, pp. 5264-5267.
- [8] B. Rafaely (2005). "Analysis and design of spherical microphone arrays", IEEE Trans. Speech Audio Proc., vol. 13(1), pp. 135-143.
- [9] H. Teutsch and W. Kellerman (2006). "Acoustic source detection and localization based on wavefield decomposition using circular microphone arrays", Journal Acoustical Society of America, vol. 120(5), pp. 2724-2736.
- [10] Z. Li (2005). "The capture and recreation of 3D auditory scenes", Ph. D. thesis, Department of Computer Science, University of Maryland, College Park.
- [11] H. Teutsch (2007). "Modal array signal processing: principles and applications of acoustic wavefield decomposition", Springer-Verlag, Berlin, Germany.
- [12] B. Rafaely (2004). "Plane-wave decomposition of the sound field on a sphere by spherical convolution", J. Acoust. Soc. Am., vol. 116(4), pp. 2149-2157.
- [13] M. Park and B. Rafaely (2005). "Sound-field analysis by plane-wave decomposition using spherical microphone array", J. Acoust. Soc. Am., vol. 118(5), pp. 3094-3103.
- [14] J. Fliege and U. Maier (1999). "The distribution of points on the sphere and corresponding cubature formulae", IMA Journal on Numerical Analysis, vol. 19(2), pp. 317-334.
- [15] Z. Li and R. Duraiswami (2005). "Flexible and optimal design of spherical microphone arrays for beamforming", IEEE Trans. Audio, Speech, and Language Processing, vol. 15(2), pp. 702-714.
- [16] N. A. Gumerov and R. Duraiswami (2005). "Fast multipole methods for the Helmholtz equation in three dimensions", Elsevier, The Netherlands.
- [17] J. W. Strutt (Lord Rayleigh) (1904). "On the acoustic shadow of a sphere", Philos. Trans. R. Soc. London, Ser. A, vol. 203, pp. 87-110.

- [18] R. O. Duda and W. L. Martens (1998). “Range dependence of the response of a spherical head model”, *J. Acoust. Soc. Am.*, vol. 104(5), pp. 3048-3058.
- [19] Z. Li and R. Duraiswami (2005). “Hemispherical microphone array for sound capture and beamforming”, *Proc. IEEE WASPAA 2005*, New Paltz, NY, pp. 106-109.
- [20] P. Morse and K. Ingaard (1968). “Theoretical acoustics”, McGraw Hill, Columbus, OH.
- [21] M. Abramowitz and I. Stegun (1964). “Handbook of mathematical functions”, Govt. Printing Office.
- [22] M. Taylor (1995). “Cubature for the sphere and the discrete spherical harmonic transform”, *SIAM J. Numer. Anal.*, vol. 32(2), pp. 667-670.
- [23] M. Brandstein and D. Ward (Eds.) (2001). “Microphone Arrays”, Springer-Verlag, Berlin, Germany.
- [24] A. O’Donovan, D. N. Zotkin, and R. Duraiswami (2008). “A spherical microphone array based system for immersive audio scene rendering”, *Proc. 14th International Conference on Auditory Displays (ICAD 2008)*, Paris, France, June 2008.
- [25] R. Duraiswami, D. N. Zotkin, Z. Li, E. Grassi, N. A. Gumerov, and L. S. Davis (2005). “High-order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues”, *Proc. AES 119th Conv.*, New York, NY, preprint #6540.
- [26] A. O’Donovan, R. Duraiswami, and N. A. Gumerov (2007). “Real time capture of audio images and their use with video”, *Proc. IEEE WASPAA 2007*, New Paltz, NY, October 2007, pp. 10-13.