

PLANE-WAVE DECOMPOSITION OF A SOUND SCENE USING A CYLINDRICAL MICROPHONE ARRAY

Dmitry N. Zotkin and Ramani Duraiswami

Perceptual Interfaces and Reality Lab
Institute for Advanced Computer Studies (UMIACS)
University of Maryland, College Park
College Park, MD 20742

ABSTRACT

The analysis for microphone arrays formed by mounting microphones on a sound-hard spherical or cylindrical baffle is typically performed using a decomposition of the sound field in terms of orthogonal basis functions. An alternative representation in terms of plane waves and a method for obtaining the coefficients of such a representation directly from measurements was proposed recently for the case of a spherical array. It was shown that representing the field as a collection of plane waves arriving from various directions simplifies both source localization and beamforming. In this paper, these results are extended to the case of the cylindrical array. Similarly to the spherical array case, localization and beamforming based on plane-wave decomposition perform as well as the traditional orthogonal function based methods while being numerically more stable. Both simulated and experimental results are presented.

Index Terms— Acoustic fields, circular arrays, array signal processing, acoustic position measurement.

1. INTRODUCTION

Arrangements of microphones in free-field and baffled spherical/cylindrical configurations are being studied for applications in acoustic scene analysis. These arrangements allow for elegant formulation of the signal processing algorithms involved, specifically, for the analysis of the sound field in terms of spherical/circular harmonics, which are the elementary solutions of the Helmholtz equation in spherical/cylindrical coordinates [1] [2]. It is known that the free-field array solution has singularities at certain frequencies [3] and that they can be avoided by mounting microphones on a solid baffle; an additional benefit of the baffled case is that the presence of the scattering object increases the useful frequency band of the array [4]. An excellent review of both spherical and cylindrical arrays can be found in [5].

It is often desired to decompose the sound field into components arriving from various directions. Such an analysis is used in many practical applications such as sound localization, signal enhancement for a direction of interest, and spatial playback of captured auditory scenes [6]. Such a decomposition can be done using beamforming operation with the (truncated) Dirac delta function as the desired beampattern both for the spherical array [7] [8] and for the cylindrical array [2]. An alternative approach is to note that in the plane-wave basis [9] a sound field is represented as a weighted collection of plane waves arriving from various directions; therefore, *the mere act of converting a sound field into the plane-wave basis decomposes it into directional components*. In a recent paper [10], this was done directly from the measurements at the microphones using

a minimum least-squares (LS) procedure. It was shown that the LS method provides substantially the same beamforming and localization abilities while having a simpler mathematical formulation. The LS procedure is especially suitable for the case where *all* directional components of the field are to be computed (e.g., for immersive audio rendering application), as opposed to the case where one is interested in listening selectively in *one* direction. In this work, the LS method is extended to the case of sound-hard cylindrical array.

2. BACKGROUND

Spherical array case: Consider a microphone located on the surface of a sound-hard sphere of radius a at a point described by direction $\mathbf{s}' = (\theta', \varphi')$. Let the Fourier transform of the acoustic pressure at the microphone be $\psi(k, \mathbf{s}')$. The latter will be loosely referred to as a “potential” in this paper; the actual velocity potential is proportional to this quantity. The potential $\psi(k, \mathbf{s}, \mathbf{s}')$ due to a plane wave with a wavenumber k propagating in the direction $\mathbf{s} = (\theta, \varphi)$ is given by

$$\psi(k, \mathbf{s}, \mathbf{s}') = 4\pi \sum_{n=0}^{\infty} i^n b_n(ka) \sum_{m=-n}^n Y_n^m(\theta, \varphi) Y_n^m(\theta', \varphi'), \quad (1)$$

which can be simplified to

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

using the addition theorem for spherical harmonics. In the equations, $Y_n^m(\theta, \varphi)$ are the orthonormal spherical harmonics; $P_n(\mathbf{s} \cdot \mathbf{s}')$ are the associated Legendre polynomials; and $J_n(ka)$, $H_n(ka)$, $j_n(ka)$, and $h_n(ka)$ (all used below) are Bessel, Hankel, spherical Bessel, and spherical Hankel functions, respectively. $b_n(ka)$ is known as a spherical mode strength and is given by

$$b_n(ka) = j_n(ka) - \frac{j_n'(ka)}{h_n'(ka)} h_n(ka). \quad (3)$$

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 the microphone (i.e., $\theta' = \pi - \theta$ and $\varphi' = \pi + \varphi$).

Cylindrical array case: In the case of the baffled cylindrical array, the problem becomes one-dimensional, the microphone is located at the point described by direction φ' , and the wave propagation direction is φ . The potential $\psi(k, \varphi, \varphi')$ becomes

$$\psi(k, \varphi, \varphi') = \sum_{n=-\infty}^{\infty} i^n B_n(ka) e^{-in(\varphi - \varphi')}, \quad (4)$$

where $B_n(ka)$ is a cylindrical mode strength and is given by

$$B_n(ka) = J_n(ka) - \frac{J'_n(ka)}{H'_n(ka)} H_n(ka). \quad (5)$$

Note the similarity of equations to the spherical array case. As in the previous case, $\psi(k, \varphi, \varphi')$ is largest when $\varphi' = \pi + \varphi$.

3. BEAMFORMING (BF) DECOMPOSITION

One way to decompose the acoustic field into components arriving from various directions is to perform a number of beamforming operations. The decomposition obtained in this way is referred to as the beamforming (BF) decomposition.

Spherical array case: Assume that the direction of interest is $\mathbf{s}_0 = (\theta_0, \varphi_0)$. The desired beampattern is the delta function in the space of directions $\delta(\mathbf{s} - \mathbf{s}_0)$, which can be expanded in terms of spherical harmonics as

$$\delta(\mathbf{s} - \mathbf{s}_0) = \sum_{n=0}^{\infty} \sum_{m=-n}^n Y_n^m(\theta, \varphi) Y_n^m(\theta_0, \varphi_0) \quad (6)$$

and the beamformer weights $w(\mathbf{s}_0, \mathbf{s}')$ required to implement this beampattern are [7] [8]

$$w(k, \mathbf{s}_0, \mathbf{s}') = 4\pi \sum_{n=0}^{\infty} \frac{1}{i^n b_n(ka)} \sum_{m=-n}^n Y_n^m(\theta_0, \varphi_0) Y_n^m(\theta', \varphi'), \quad (7)$$

which can be simplified again [9] to

$$w(k, \mathbf{s}_0, \mathbf{s}') = \sum_{n=0}^{\infty} \frac{(2n+1)P_n(\mathbf{s}_0 \cdot \mathbf{s}')}{i^n b_n(ka)}. \quad (8)$$

Cylindrical array case: Here the direction of interest is φ_0 and

$$\delta(\varphi - \varphi_0) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in(\varphi - \varphi_0)}. \quad (9)$$

Therefore, the beamformer weights $w(\varphi_0, \varphi')$ are simply [5]

$$w(k, \varphi_0, \varphi') = 2\pi \sum_{n=-\infty}^{\infty} \frac{1}{i^n B_n(ka)} e^{-in(\varphi_0 - \varphi')}. \quad (10)$$

Series Truncation: It is known [5] [7] that $b_n(x)$ and $B_n(x)$ tend to zero exponentially for fixed x as $n \rightarrow \infty$; therefore, the sums in Equations (2) and (4) converge, and in fact the convergence is very fast once $n > x$. Hence, these expressions can be truncated:

$$\psi(k, \mathbf{s}, \mathbf{s}') \approx \sum_{n=0}^{p-1} (2n+1) i^n b_n(ka) P_n(\mathbf{s} \cdot \mathbf{s}'), \quad (11)$$

$$\psi(k, \varphi, \varphi') \approx \sum_{n=-(p-1)}^{p-1} i^n B_n(ka) e^{-in(\varphi - \varphi')}. \quad (12)$$

The truncation number p depends on the required error tolerance. In [9], error bounds are derived for the spherical array case, and it is shown that the truncation error in Equation (11) is negligible for $p = p^* = (eka - 1)/2$; however, p can be set arbitrarily high without causing numerical problems. Numerical simulations show similar behavior in cylindrical array case.

Accordingly, the summations in computation of beamforming weights (Equations (8) and (10)) are also truncated to p and $2p - 1$

terms, respectively. However, the numerical weakness of the beamforming algorithm is that the beamforming weight is inversely proportional to mode strength, which rapidly approaches zero when $n > ka$; therefore, one needs to be extremely careful in choosing p so as not to cause amplification of white noise. In practice, the equipment (microphone/amplifier/ADC) noise floor, along with mode strength magnitude plots such as Figure 1 of [7], are used in determining p for computing weights. Trying to use a higher p in search for larger signal separation results in a total loss of desired signal as the 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).

Matrix formulation: If a set of L_j directions \mathbf{s}_j is known, the BF decomposition can be written as a matrix-vector product at each wavenumber k . Assume that L_i microphones are placed at directions \mathbf{s}'_i , that the potential measured at the microphone at \mathbf{s}'_i is $\Psi(k, \mathbf{s}'_i)$, and that an $L_i \times 1$ column vector Ψ consists of values $\Psi(k, \mathbf{s}'_i)$. The desired matrix-vector product formulation can be written as $\Lambda = W\Psi$, where a matrix of weights W is simply an $L_j \times L_i$ matrix of values $w(k, \mathbf{s}_j, \mathbf{s}'_i)$. The resulting $L_j \times 1$ vector Λ consists of values $\mu(k, \mathbf{s}_j)$ that represent complex magnitude of the scene component corresponding to the direction \mathbf{s}_j at wavenumber k . Practical details of packing time-domain signals into a vector Ψ and transforming back to time domain from vector Λ are given in Section 5.

4. LEAST-SQUARES (LS) DECOMPOSITION

An alternative algorithm for the spherical array case was introduced in [10]. It is based on finding the set of plane wave magnitudes $\mu(k, \mathbf{s}_j)$ that best explain the set of observed potentials $\Psi(k, \mathbf{s}'_i)$. If a unit magnitude plane wave is present with direction of propagation \mathbf{s}_j , then the potential at microphone \mathbf{s}'_i is given by Equation (11). If multiple plane waves are present with magnitudes $\mu(k, \mathbf{s}_j)$, then due to linearity the potential at the microphone is simply a sum of potentials due to all those plane waves. In matrix form, $\Psi = F\Lambda$, where F is an $L_i \times L_j$ matrix of values $\psi(k, \mathbf{s}_j, \mathbf{s}'_i)$. This is a linear system of equations and can be solved for Λ , giving again the simple one-step matrix-vector product algorithm for computing $\Lambda = W\Psi$, where for $L_j \leq L_i$ the solution is given by least-squares fitting with Tikhonov regularization:

$$W = (F^T F + \varepsilon I)^{-1} F^T, \quad (13)$$

where I is the identity matrix and ε is a regularization parameter (this also covers the case of unique solution when $L_j = L_i$; when $L_j > L_i$, the system is underdetermined and there is an infinite number of solutions). Regularization is necessary because even though the computation of the elements of matrix F is not subject to numerical problems, the matrix itself can be ill-conditioned at lower frequencies and computation of the (pseudo)inverse of F can be unstable [10]. The regularization parameter ε is related to the condition number of F . It was shown in [10] that the specific value of ε within reasonable bounds (from 10^{-7} to 10^{-2} relative to the elements of Ψ) only marginally influences the decomposition results.

In this paper, a similar algorithm is proposed and tested for the cylindrical array. Assume that L_i microphones are placed at directions φ'_i and that there are L_j directions φ_j that form the basis for the plane-wave decomposition. The potential at the microphone located at φ'_i due to a unit magnitude plane wave traveling in the direction φ_j is given by Equation (12), and the potential due to many plane waves is the sum of the potentials of these plane waves weighted by their magnitudes $\mu(k, \varphi_j)$. In matrix form, $\Psi = F\Lambda$, with notational conventions being the same as for the spherical array case. The linear system is solved $\Lambda = W\Psi$ either exactly when $L_j = L_i$

or approximately in a least-squares sense when $L_j < L_i$; both cases are covered with W given by Equation (13).

A major difference between the LS algorithm and the BF algorithm is numerical stability. In BF algorithm, the weights are given by Equation (8) (sphere) or Equation (10) (cylinder). As discussed before, these equations diverge due to the exponential growth of $b_n^{-1}(ka)$ ($B_n^{-1}(ka)$, respectively) term. In contrast, the equations used to compute matrix F in LS algorithm converge and in fact converge very fast. Therefore, the truncation number can be set arbitrarily high. Furthermore, it can be hypothesized that the LS algorithm could provide a decomposition that is more physically correct because the truncation number can be set higher and the wave scattering can be modeled more accurately.

5. PRACTICAL IMPLEMENTATION

Assume that the time-domain signal at a pressure-sensitive microphone located at \mathbf{s}'_i is $x_i(t)$ 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 coefficients for 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 , as W is different for different k . The decomposition coefficients $\mu(k, \mathbf{s}_j)$ are computed with the matrix vector-product $\Lambda = W\Psi$ as described above. The potential $\Psi(k, \mathbf{s}'_i)$ is simply the Fourier coefficient for the microphone at \mathbf{s}'_i at wavenumber k . Then, the output time-domain signal $y_j(t)$ for the direction \mathbf{s}_j is obtained by assuming that the computed $\mu(k, \mathbf{s}_j)$ is the Fourier coefficient of the output signal at wavenumber k and performing inverse Fourier transform of the set of $\mu(k, \mathbf{s}_j)$.

In practice, overlapping smoothly fading windows are used to eliminate windowing artifacts such as clicks occurring on the frame 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 the array size and equipment noise floor (if the acoustic wave length is substantially larger than the array size, the difference in potentials at different microphones is small and may be masked by electronic noise or by quantization effects). In this case, the computations are done only for the wavenumbers corresponding to the useful frequency band and the remaining $\mu(k, \mathbf{s}_j)$ coefficients are zeroed out for inverse Fourier transform.

Also note that Equations (2), (4), (8), and (10) assume that \mathbf{s}_0 (φ_0 , respectively) is the direction of propagation, not the direction towards the source. Accordingly if one wishes to decompose the sound field into components arriving from directions $\tilde{\mathbf{s}}_j$ (i.e., beamform in the direction $\tilde{\mathbf{s}}_j$), the *opposite* direction ($\theta_j = \pi - \theta_j$, $\varphi_j = \pi + \tilde{\varphi}_j$) should be used in computing the weight matrix W .

6. EXPERIMENTAL SETUP

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. A plane wave was impinging at the array from a random direction φ_w . The microphone potentials were computed according to Equation (12) with some truncation number p (denoted as p_m later on). Then, these potentials were fed either to BF procedure or to LS procedure to compute the coefficients $\mu(k, \varphi_j)$ using *different* p (denoted as p_s later on). The set of plane-wave directions φ_j was chosen to be the same as the set of microphone directions φ'_i . Setting $L_j = L_i$ is in fact reasonable as no information would be lost during the decomposition (in other words, the number of degrees of freedom in the recording itself and in the plane-wave representation are the same). The direction φ_l for which $|\mu(k, \varphi_j)|$ was maximal was selected as

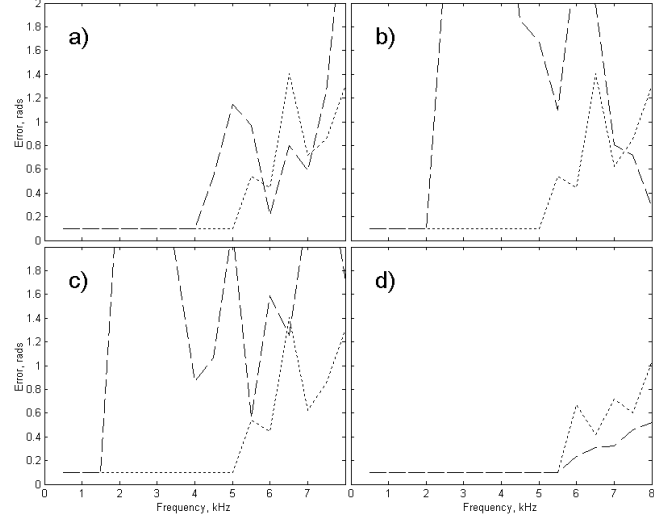


Fig. 1. Localization error for beamforming decomposition (BF, dashed line) and LS decomposition (LS, dotted line). Simulations; in all cases $p_m = 10p^*$ and $\varepsilon = 10^{-2}$. a) $p_s = p^*$. b) $p_s = 3p^*/2$. c) $p_s = 2p^*$. d) $p_s = 3p^*/4$.

the direction in which the algorithm localized the simulated source. The error $|\varphi_l - \varphi_w|$ was averaged over 4096 trials.

In addition, the algorithms were applied to a real acoustic recording obtained from a *spherical* microphone array of radius 0.101 m. 13 microphones roughly encircling the sphere were selected to form a pseudo-cylindrical array. The microphones were located approximately, but not exactly, on the equator of the sphere; some microphones were off the equator by as much as 15 degrees. Also, the angular spacing between microphones was not constant. As such, the operating conditions were far from those assumed in derivations in [5]. In the recording, one person (male) and another person (female) were speaking in turns, located approximately in the direction of microphone #3 and microphone #8 in the array, respectively. The sampling frequency was 39.0625 kHz and the useful array frequency band was from 320 Hz to 3200 Hz.

7. RESULTS AND DISCUSSION

Simulations: A simulated localization experiment was run for various p_m and p_s . Below, the p_m and p_s are expressed in terms of p^* . In all plots presented, $p_m = 10p^*$ to accurately reflect the potentials that would be observed in real system (where $p_m = \infty$). Also in all plots regularization with $\varepsilon = 10^{-2}$ was used in LS algorithm.

Figure 1(a) shows the localization error 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 exceeds the spatial aliasing limit of 4.3 kHz, showing that the degradation of localization and beamforming ability happens rather gradually. Poor BF performance at higher frequencies is due to the above-described divergence of weights. This is confirmed by Figure 1(b) plots, which show the case of $p_s = 3p^*/2$; here, the BF method breaks down at about 2.5 kHz and the LS results are unchanged. If p_s is increased further (Figure 1(c)), the same effect becomes more pronounced.

Another simulated experiment was therefore conducted with p_s decreased to $3p^*/4$. The corresponding error plots are shown in Figure 1(d). It can be seen that the operational range of BF method

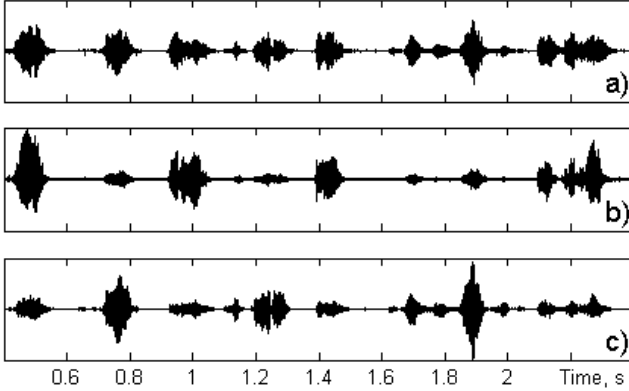


Fig. 2. Sample scene decomposition, 13-element pseudo-cylindrical array, $p_s = ka$. Two speakers located at microphones #3 and #8, respectively, are speaking in turns. a) The raw waveform from microphone #11. b) Plane wave obtained by LS decomposition corresponding to the direction of microphone #3. c) Same for #8.

is widened and is in fact the same now as the operational range for LS method, with error increasing gradually above 5.5 kHz.

These cylindrical array results are consistent with similar simulations performed with spherical arrays in [10]. In particular, it is observed that the truncation number $p_s = p^*$ considered optimal in [9] is too high for use in real spherical and cylindrical beamforming applications as numerical problems arise even in a simulated setup. Lower truncation numbers such as $p_s = ka$ used by some authors would be more applicable for the BF method. In contrast, with LS method increase of p_s does not lead to localization error increase.

Experimental data: The algorithms were also tested on the data collected from the pseudo-cylindrical array described in Section 6. Figure 2(a) shows the signal collected from the microphone #11 (located at mid-angle between two speakers). This recording has both speakers at approximately the same magnitude. The recording was processed first with $p_s = ka$ and no substantial differences were found in outputs of beamforming decomposition and LS decomposition. Figure 2(b) and (c) show the result of LS decomposition corresponding to the direction of the first and the second speaker, respectively. Then, the recording was processed with $p_s = 2ka$. Figure 3(b) and (c) show the results of beamforming decomposition and LS decomposition for the direction of the second speaker. For reference, plot 3(a) is the same as 2(c) to facilitate comparison between $p_s = ka$ and $p_s = 2ka$. It can be seen that the suppression of unwanted signal is somewhat better in 3(c) compared to 3(a) and that the signal separation is lost in 3(b). This is consistent with the earlier hypothesis that increase in truncation number helps LS decomposition because the physics of the process is modeled better but hurts BF decomposition due to the weights divergence.

8. CONCLUSION AND FUTURE WORK

In this work, plane-wave decomposition analysis was applied to a baffled cylindrical microphone array. It was shown that the localization and beamforming algorithms based on plane-wave decomposition show performance comparable with traditional spherical harmonics based decomposition while not being subject to certain numerical problems associated with the latter. The algorithms will soon be further tested on a prototype cylindrical array that is currently being built. In the future, applications of the algorithms in sound enhancement and remote telepresence are planned.

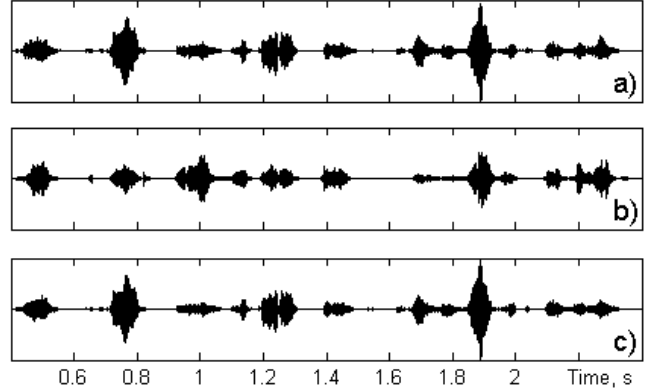


Fig. 3. Sound scene component corresponding to the direction of microphone #8 for varying truncation number. a) $p_s = ka$ (same as figure 2(c)). b) Output of BF decomposition for $p_s = 2ka$. c) Output of LS decomposition for $p_s = 2ka$.

9. REFERENCES

- [1] T. D. Abhayapala and D. B. Ward (2002). "Theory and design of high order sound field microphones using spherical microphone array", Proc. IEEE ICASSP 2002, Orlando, FL, May 2002, vol. 2, pp. 1949-1952.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] H. Teutsch (2007). "Modal array signal processing: principles and applications of acoustic wavefield decomposition", Springer-Verlag, Berlin, Germany.
- [6] Z. Li (2005). "The capture and recreation of 3D auditory scenes", Ph. D. thesis, Department of Computer Science, University of Maryland, College Park.
- [7] 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, May 2002, vol. 2, pp. 1781-1784.
- [8] B. Rafaely (2004). "Plane-wave decomposition of the sound field on a sphere by spherical convolution". Journal Acoustical Society of America, vol. 116(4), pp. 2149-2157.
- [9] 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, October 2005, pp. 150-153.
- [10] 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.