ROBUST ONLINE DIRECTION OF ARRIVAL ESTIMATION USING LOW DIMENSIONAL SPHERICAL HARMONIC FEATURES

Vishnuvardhan Varanasi and Rajesh Hegde

Indian Institute of Technology, Kanpur
Email: {vishnuv, rhegde}@iitk.ac.in

ABSTRACT

Signal processing in spherical harmonic domain has the ability to decouple frequency dependent and location dependent components of the signal received. A method for low dimensional spherical harmonic feature extraction is proposed in this work for DOA estimation in noisy and reverberant environments. The features are extracted using frequency smoothing and a transformation which makes them frequency and signal invariant. Additionally an online manifold regularization framework is explored which utilizes the proposed spherical harmonic features to compute real time DOA estimates. This framework minimizes an instantaneous risk function and finds an inverse mapping function that maps spherical harmonic features to the DOA estimate. Performance of the proposed DOA estimation method is then compared with DOA estimates obtained from features such as generalized cross correlation and relative transfer function in a semi-supervised manifold regularization framework. Experimental results on DOA estimation in terms of root mean square error and probability of resolution indicate a reasonable improvement in the localization performance along with significant reduction in feature dimension.

Index Terms— Online Learning, Spherical Microphone Array, Acoustic Source Localization, Spherical Harmonics

1. INTRODUCTION

DOA Estimation is a challenging task especially in presence of noise and reverberation. Various applications of acoustic source localization include Distant Automatic Speech Recognition [1] , Music Information Retreival [2] , Automatic Camera Steering [3, 4] and hence it has been an active area of research. Also, Spherical Microphone Array (SMA) [5] has received significant attention of researchers in the recent days. SMA captures spherical variation of acoustic field [6] with spherical harmonics. In this work, we address the problem of DOA estimation using spherical harmonic features. A wide range of DOA estimation algorithms are developed in the past few decades. Most of them can be broadly categorized as follows. Subspace based methods such as MUltiple SIgnal Classification(MUSIC), Spherical Harmonic MUSIC (SH-MUSIC) [7, 8], Steered beamforming methods such as Steered Response Power with Phase Tranform (SRP-PHAT) [9], Time Delay of Arrival (TDOA) based approaches [10], Learning based approaches with features such as Generalized Cross Correlation (GCC) [11], Relative Transfer Function (RTF) [12] and Sparsity based methods [13].

In adverse environments, above mentioned methods suffer from inaccurate DOA estimates or high computational complexity. Learning based approaches provide better performance in adverse environments. However, features based on GCC and RTF are high dimensional and hence learning becomes computationally complex. The contributions of this work are two fold. First, it proposes novel low dimensional spherical harmonic features which are robust to noise and reverberation. Secondly, an online manifold regularization framework that can utilize the low dimensional nature of these features to compute real time DOA estimates is developed.

The rest of the paper is organized as follows. Section 2 introduces data model in spherical harmonic domain and extraction of low dimensional and signal invariant spherical harmonic features. Section 4 evaluates performance of different features with the proposed spherical harmonic features(SHF) using Online Manifold Regularization(OMR). Section 5 concludes the paper.

2. LOW DIMENSIONAL SPHERICAL HARMONIC FEATURES FOR DOA ESTIMATION

In this section, computation of low dimensional spherical harmonic features for DOA estimation is described. The data model in spherical harmonics domain is subject to frequency smoothing and then transformed to obtain signal invariant spherical harmonic features. The low dimensionality and robustness of these features is also described herein.

2.1. Data Model in Spherical Harmonic Domain

Consider an acoustic scene where L point sources are located along with a spherical microphone array. The location of sources is indicated by $\Psi_i = (\theta_i, \phi_i) \quad i = 1, 2, ...L$ where θ and ϕ are the elevation and azimuth angles respectively. Location of microphones is given by $\Omega_i = (\theta_i, \phi_i) \quad i = 1, 2, ...I$. The following data model gives the sound pressure observed in frequency domain with I microphones and L sources.

$$\mathbf{p}(k) = \mathbf{V}(k, \mathbf{\Psi})\mathbf{s}(k) \; + \; \mathbf{n}(k) \tag{1}$$

where k is wave number, $\mathbf{s}(k)$ is source strength vector and $\mathbf{n}(k)$ is additive zero mean white gaussian noise and $\mathbf{V}(k, \mathbf{\Psi})$ is steering matrix.

Expanding the steering matrix using spherical harmonics [14, 15], Equation 1 can be rewritten as

$$\mathbf{p}(k) = \mathbf{Y}(\mathbf{\Omega})\mathbf{B}(kr)\mathbf{Y}^{\mathbf{H}}(\mathbf{\Psi})\mathbf{s}(k) + \mathbf{n}(k) \tag{2}$$

$\mathbf{Y}(\mathbf{\Omega}) \in \mathbb{C}^{(N+1)^2 \times I}$,$\mathbf{Y}(\mathbf{\Psi}) \in \mathbb{C}^{(N+1)^2 \times L}$ are the spherical harmonics matrices with angular positions corresponding to microphones and sources respectively. $\mathbf{B}(kr)$ is the mode strength matrix corresponding to the wave number k and SMA of radius r.

The matrix formulation of orthogonality of spherical harmonics for uniform sampling or nearly uniform sampling [16] can be stated as

$$\frac{4\pi}{I}\mathbf{Y}^{\mathbf{H}}(\mathbf{\Omega})\mathbf{Y}(\mathbf{\Omega}) = \mathbf{I} \tag{3}$$

Consider a reverberant environment with R significant reflections with a_i as the reflection coefficient of the i^{th} reflector. The transformation between spatial and spherical harmonics domain can be done using spherical harmonics matrix as

$$\mathbf{p_{nm}}(k) = \mathbf{Y^H}(\mathbf{\Omega})\mathbf{p}(k) \tag{4}$$

where $\mathbf{p_{nm}}(k)$ is the observation vector in spherical harmonics domain. From Equations 2 and 3, data model takes the following form [17].

$$\mathbf{p_{nm}}(k) = \frac{I}{4\pi}\mathbf{B}(kr)\mathbf{Y^H}(\mathbf{\Psi})\,\mathbf{a}\,s(k) + \mathbf{Y^H}(\mathbf{\Omega})\mathbf{n}(k) \tag{5}$$

where $\mathbf{a} = [a_1, a_2, ...a_R]^T$. The signal term in Equation 5 becomes independent of frequency by left multiplying Equation 5 with $\mathbf{B^{-1}}(kr)$. For notational convenience lets use the following representation.

$$\mathbf{x_{nm}}(k) = \frac{4\pi}{I}\mathbf{B^{-1}}(kr)\mathbf{p_{nm}}(k) \tag{6}$$

From Equations 5 and 6, the final data model takes the following form.

$$\mathbf{x_{nm}}(k) = \mathbf{Y^H}(\mathbf{\Psi})\,\mathbf{a}\,s(k) + \frac{4\pi}{I}\mathbf{B^{-1}}(kr)\mathbf{Y^H}(\mathbf{\Omega})\mathbf{n}(k) \tag{7}$$

This data model in spherical harmonics domain is used for extraction of robust and low dimensional features as explained in the following section.

2.2. Extraction of Low Dimensional Signal Invariant Spherical Harmonic Features

In earlier work on extraction of features for learning algorithms, RTF and GCC are widely used. These features are high in dimension and this dimensionality increases quadratically with increase in number of microphones. Their performance reduces in noisy and reverberant environments. So, new features are obtained from the spherical harmonics data model proposed in this work. These features are low dimensional and robust. In order to obtain these low dimensional features that are signal invariant, the data model in Equation 7 is considered.

Spherical harmonic coefficients of the observations in Equation 7 are a function of direction of arrival of the signal and the source strength defined as follows.

$$\mathbf{x_{nm}}(k) = f(\Psi, s(k)) \tag{8}$$

The objective here is to find a transformation $\mathbb{T}$ that operates on $\mathbf{x_{nm}}(k)$ to give feature vector $\mathbf{y_{nm}}$ such that the feature is a function of only direction of arrival, thus making it signal invariant.

$$\mathbf{y_{nm}} = \mathbb{T}(\mathbf{x_{nm}}(k)) \quad subject\ to \quad \mathbf{y_{nm}} = \tilde{f}\,(\Psi) \tag{9}$$

This transformation normalizes $\mathbf{x_{nm}}(k)$ by the spherical harmonic coefficient $x_{00}(k)$, since $x_{00}(k)$ is dependent only on the signal strength $s(k)$ and not on the DOA yielding $\mathbf{\tilde{x}_{nm}(k)}$ which is signal invariant. It must be noted that the dependence of $\mathbf{\tilde{x}_{nm}(k)}$ on noise can be reduced by taking an Expectation of $\mathbf{\tilde{x}_{nm}}(k)$. The methodology for obtaining the transformation $\mathbb{T}$ is discussed in the ensuing sections.

2.2.1. Signal Invariance

Considering x_{00} in Equation 7, it can be written as

$$x_{00}(k) = \sum_{i=1}^{R} Y_0^0(\psi_i)\; a_i\; s(k) + \frac{4\pi}{I}\sum_{i=1}^{I} Y_0^0(\Omega_i)\; n(k) \tag{10}$$

The spherical harmonic basis functions of order n and degree m can be defined as

$$Y_n^m(\theta,\phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m\,(cos\;\theta)\,e^{im\phi} \tag{11}$$

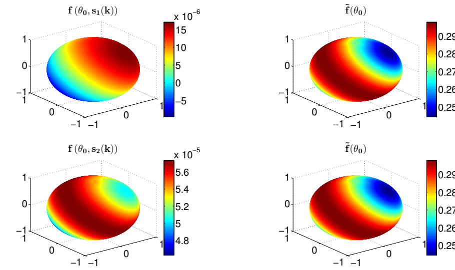

Fig. 1. Spherical field variation of two signals arriving from same direction with different signal strengths(left column). Spherical field variation after feature extraction showing that the features are invariant to the signal strengths(right column).

For the case where m and n are zero,

$$Y_0^0(\psi) = \sqrt{\frac{1}{4\pi}} \tag{12}$$

From Equations 10 and 12, we have

$$x_{00}(k) = \frac{1}{\sqrt{4\pi}}\left[\left(\sum_{i=1}^{R} a_i\right)s(k) + \frac{4\pi}{I}\left(\sum_{i=1}^{I} n_i(k)\right)\right] \tag{13}$$

As I increases, the second term tends to zero as fourier transform of zero mean gaussian noise is also zero mean gaussian. Thus, $x_{00}(k)$ is just a scalar multiple of the signal strength i.e $x_{00} = \alpha\ .\ s(k)$ where α is some constant. Hence, $x_{00}(k)$ can be used to remove the influence of the signal strength on the observations. From Equation 7, we obtain signal invariant features $\mathbf{\tilde{x}_{nm}}(k)$ as

$$\mathbf{\tilde{x}_{nm}(k)} = \frac{\mathbf{x_{nm}}(k)}{x_{00}(k)} = \frac{1}{\alpha}\;\mathbf{Y^H}(\psi)\;\mathbf{a} + \frac{\mathbf{z_{nm}}(k)}{\alpha\; s(k)} \tag{14}$$

where $\mathbf{z_{nm}}(k) = \frac{4\pi}{I}\mathbf{B^{-1}}(kr)\mathbf{Y^H}(\mathbf{\Omega})\mathbf{n}(k)$

2.2.2. Robustness to noise

It may be noted from Equation 14 that $\mathbf{\tilde{x}_{nm}}(k)$ contains two terms. The first has information on the DOA and the second corresponds to noise. An Expectation operator is now applied on Equation 14 to obtain $\mathbf{y_{nm}}$. For different wave number k, the first term in Equation 14 remains the same since it depends only on the DOA. On the other hand, the second term which is a realization of noise approaches asymptotically to zero since noise is assumed to be WGN. The transformation $\mathbb{T}$ is now complete and yields $\mathbf{y_{nm}}$ as follows.

$$\mathbf{y_{nm}} = \mathbf{E_k}(\mathbf{\tilde{x}_{nm}}(k)) = \mathbf{E_k}\left[\frac{\mathbf{x_{nm}}(k)}{x_{00}(k)}\right] \tag{15}$$

The spherical harmonic features $\mathbf{y_{nm}}$ are thus obtained using a transformation $\mathbb{T}$ as discussed in the aforementioned sections.

2.2.3. Low Dimensionality

Dimensionality of the feature vector plays a major role in determining computational complexity of any machine learning algorithm. The lower the feature vector dimension, lower is the complexity of the algorithm. As RTF and GCC features are defined for a pair of microphones, their dimensionality represented by $\mathbb{D}$ is as follows.

$$\mathbb{D}(RTF/GCC) = (\mathbb{D}(RTF/GCC)\;per\;pair) \times {}^{I}C_2 \tag{16}$$

where

$$ {}^{n}C_k = \frac{n!}{k!(n-k)!} \tag{17} $$

For the proposed spherical harmonic features, the feature dimensionality is governed by $I = \beta(N+1)^2$ where $\beta > 1$ and β depends on the sampling method[18, 19] and N is the order of the spherical microphone array. Feature dimensionality for GCC, RTF for a pair of microphones are as indicated in [11, 12]. Using Equation 16, Table 1 illustrates the dimensionality of proposed feature compared to the existing features. For better accuracy at multiple frequencies, higher number of microphones is preferred. As can be seen from the Table 1, with more than ten times the number of microphones, we are able to get down the dimensionality by ten times.

Feature Type	Microphones dependency	Theoretical Dimensionality	I	Feature Dimension
RTF	Quadratic	$300 \times {}^{I}C_2$	2	300
GCC	Quadratic	$20 \times {}^{I}C_2$	8	560
SHF	Linear	$(N+1)^2 \leq I$	50	49

Table 1. Comparison of dimensionality of various features and their dependence on the number of microphones.

3. ONLINE MANIFOLD REGULARIZATION USING SPHERICAL HARMONIC FEATURES FOR DOA ESTIMATION

Online learning has not been investigated for DOA estimation since high dimensional feature vectors obtained from GCC,RTF makes online learning computationally complex. Since low dimensional spherical harmonic features are used in this work, an online learning framework is developed. Online learning [20] constructs a sequence of functions $f_1, f_2......f_T$. With each incoming feature y_t, algorithm tries to find an estimate of function,f_t so that the instantaneous risk function is minimized. Existence of the label ϕ_t(DOA) at time instant t is indicated by the binary(0 or 1) function,$\delta(\phi_t)$. We assume that this function lies in a Reproducible Kernel Hilbert Space (RKHS) which is associated with a unique kernel function that evaluates each function in the space by an inner product. An essential requirement of the kernel is the notion of locality which can be defined as

$$ \begin{aligned} &for \quad \|y_i - y_j\| << \epsilon_k, \quad k(y_i, y_j) \mapsto 1 \\ &for \quad \|y_i - y_j\| >> \epsilon_k, \quad k(y_i, y_j) \mapsto 0 \end{aligned} \tag{18} $$

for some value of ϵ_k. A common choice of kernel function that follows the notion of locality is a Gaussian kernel with variance ϵ_k^2 defined as

$$ k(y_i, y_j) = exp\left\{-\frac{\|y_i - y_j\|^2}{\epsilon_k^2}\right\} \tag{19} $$

Therefore the definition of RKHS with gaussian kernel can be defined as a set of functions as

$$ H_k = \left\{ f | f(.) = \sum_{i=1}^{N} a_i k_{h_i}(.) ; a_i \,\epsilon\, R, h_i \,\epsilon\, M \right\} \tag{20} $$

Before proceeding to online manifold regularization, a batch semi-supervised regularized risk function can be framed as

$$ J(f) = \frac{1}{l} \sum_{t=1}^{T} \delta(\phi_t) c(f(y_t), \phi_t) + \gamma_K \|f\|_{H_K}^2 + \gamma_M \|f\|_M^2 \tag{21} $$

where c is a cost function, γ_K and γ_M are regularization parameters, $\|.\|_{H_k}^2$ and $\|.\|_M^2$ are norms defined with respect to RKHS and manifold M respectively. Since online learning algorithm has access to only a single instance at a time, the batch risk function is replaced by an instantaneous risk as follows.

$$ J_t(f) = \frac{T}{l} \delta(\phi_t) c(f(y_t), \phi_t) + \gamma_K \|f\|_{H_k}^2 + \gamma_M \|f\|_M^2 \tag{22} $$

The function f_{t+1} can now be found by minimizing $J_t(f)$ as

$$ f_{t+1} = \arg\min_f \; J_t(f) \tag{23} $$

The online algorithm performs this minimization by using gradient descent to obtain f_{t+1} as

$$ f_{t+1} = f_t - \eta_t \left. \frac{\partial J_t(f)}{\partial f} \right|_{f_t} \tag{24} $$

Finding f_{t+1} translates to finding the coefficients $a_i^{(t+1)}$ where $i = 1, ...t$. These coefficients can be computed in two steps using Equations 22 and 24.

$$ \begin{aligned} a_i^{(t+1)} = (1 - \eta_t \gamma_K) a_i^{(t)} - 2\eta_t \gamma_M (f_t(y_i) - f_t(y_t)) w_{it} \\ for\ i = 1, 2, 3...., t-1 \end{aligned} \tag{25} $$

Further, a_i at t^{th} index at time instant $t+1$ can be computed as

$$ a_t^{(t+1)} = 2\eta_t \gamma_M (f_t(y_i) - f_t(y_t)) w_{it} - \eta_t \frac{T}{l} \delta(y_t) c'(f(y_t, \phi_t)) \tag{26} $$

Once the coefficients are found at time instant $t+1$, DOA for feature vector y_{t+1} can be found using the definition of function f_{t+1} as defined below.

$$ f_{t+1}(y_{t+1}) = \sum_{i=1}^{t} a_i^{(t+1)} k(y_i, y_{t+1}) \tag{27} $$

Online manifold regularization can be summarized as follows. At each time instant, a feature vector y_t is received. Using Equations 25 and 26, an new estimate of the function, f_{t+1} is obtained. DOA at time instant $t+1$ can be estimated by using function, f_{t+1} as indicated in Equation 27.

4. PERFORMANCE EVALUATION

The performance of the proposed low dimensional spherical harmonic features in an online manifold regularization framework is evaluated by conducting experiments on DOA estimation for varying T_{60} and noise. A comparative performance analysis is done with the Semi-Supervised Learning (SSL) approach [12].

4.1. Experiments on DOA Estimation

In this work, we utilize speech data from GRID corpus [21] for experiments. Experimental study is conducted by simulating a room of dimensions $6 \times 6 \times 6\ m$ with various boundary reflection coefficients corresponding to different T_{60}. A rigid spherical microphone array of radius 15 cm is placed at center of the room and the source is placed at different azimuthal angles and a fixed elevation. Since SMA is used, DOA estimation can be easily extended to elevation as well. Impulse responses from source at different azimuthal angles to microphones on the sphere are computed using SMIR generator [22]. The objective is to recover the unknown azimuth angle of incoming signal using a training set that consists of both labeled and unlabeled samples. Label ratio i.e ratio of number of labeled samples to total number of samples in semi-supervised and online algorithm implementation is 0.4.

4.2. Experimental Results

Different quality measures such as cumulative Root Mean Square Error(RMSE) [7] and Probability of Resolution are used to evaluate the performance of the proposed online manifold regularization method. Results are also compared to semi-supervised learning methods.

4.2.1. Cumulative RMSE Analysis

The cumulative RMSE is defined as follows.

$$RMSE = \frac{1}{LT}\sum_{tr=1}^{T}\sum_{l=1}^{L}\left[\left(\phi_l - \hat{\phi}_l^{(tr)}\right)^2\right] \tag{28}$$

where tr is the trial number among T trials and l indicates the source location among L possible different locations.

4.2.2. Probability of Resolution Analysis

Probability of resolution within a confidence interval gives a good statistical analysis of DOA estimation. Probability of resolution is defined as follows.

$$\begin{aligned} P_r &= \frac{1}{LT}\sum_{tr=1}^{T}\sum_{l=1}^{L} Pr(|\phi_l - \hat{\phi}_l^{(tr)}| \le \zeta) \\ &= \frac{1}{LT}\sum_{tr=1}^{T}\sum_{l=1}^{L}\left[sgn(\zeta - |\phi_l - \hat{\phi}_l^{(tr)}|)\right] \end{aligned} \tag{29}$$

where $Pr(.)$ denotes the probability of an event, ζ is the confidence interval, tr is the trial number among T trials and l indicates the source location among L possible different locations. $sgn(x)$ is the signum function defined as

$$sgn(x) = \begin{cases} 1 & if \quad x \ge 0 \\ 0 & if \quad x < 0 \end{cases} \tag{30}$$

4.2.3. Experimental Results on DOA Estimation

Experiments on DOA estimation are conducted at various SNR. In Figure 2 and it can be clearly seen that after certain point of time, the localization accuracy improves and approaches zero irrespective of SNR. RMSE in degrees for different reverberation times(T_{60}) is also plotted in Figure 3. It can be noted that performance of online learning algorithm does not vary much when the reverberation times are varied. In Table 2, the proposed Online Manifold Regularization method using spherical harmonic features is also compared to a semi-supervised learning framework which has been used earlier in the context of DOA estimation. In Figure 4, probability of resolution with GCC,RTF, and the proposed spherical harmonic features in a semi-supervised learning framework [12] are compared.

Methods	SNR = 0 dB		SNR = 10 dB		SNR = 20 dB	
	RMSE	P_r	RMSE	P_r	RMSE	P_r
SSL-GCC	12.03	0.21	11.61	0.22	11.06	0.43
SSL-RTF	12.06	0.16	11.86	0.22	10.33	0.41
SSL-SHF	6.91	0.62	4.23	0.78	3.18	0.85
OMR-SHF	2.70	0.535	2.71	0.731	2.32	0.80

Table 2. RMSE for GCC,RTF and SHF features for various DOA estimation methods.

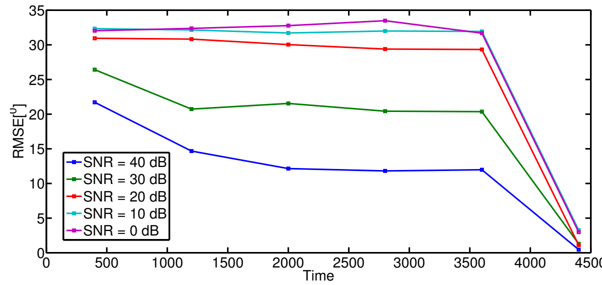

Fig. 2. Variation of RMSE with time at different SNR for Online Manifold Regularization with low dimensional spherical harmonic features.

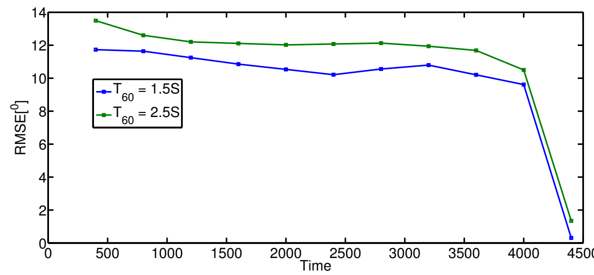

Fig. 3. Variation of RMSE with time at different reverberation times(T_{60}) in case of Online Manifold Regularization with low dimensional spherical harmonic features.

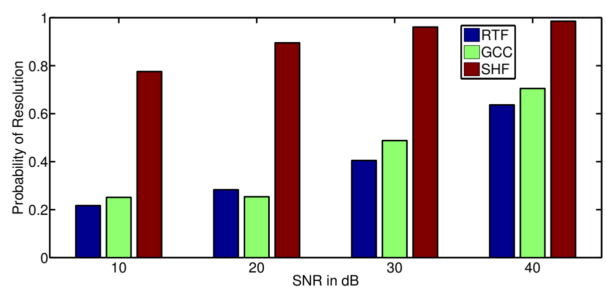

Fig. 4. Illustration of probability of resolution using GCC,RTF and the proposed spherical harmonic features in a semi-supervised learning framework.

5. CONCLUSION

A new method for computing low dimensional spherical harmonic features that are robust, highly direction dependent, and signal invariant is proposed in this work. The lower dimensionality of the features ensures that they can be used in a online manifold regularization framework for real time DOA estimation. The experimental results indicate reasonable DOA estimates even in highly noisy and reverberant environment which is motivating enough to utilize the method in teleconferencing applications. Existing learning techniques assume that in a reverberative environment, features correspond to a single speaker. In practice however there can be multiple speakers present at a single time instant. The problem of multiple speaker separation in such contexts will be investigated using time frequency analysis in future work.

6. REFERENCES

[1] Xiong Xiao, Shengkui Zhao, Duc Hoang Ha Nguyen, Xionghu Zhong, Douglas L Jones, Eng-Siong Chng, and Haizhou Li, "The ntu-adsc systems for reverberation challenge 2014," in *Proc. REVERB challenge workshop*, 2014.

[2] Johannes Traa and Paris Smaragdis, "Blind multi-channel source separation by circular-linear statistical modeling of phase differences," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 4320–4324.

[3] Nguyen Trung Hieu Shengkui Zhao, Eng Siong Chng, and Haizhou Li, "A robust real-time sound source localization system for olivia robot," in *2010 APSIPA Annual Summit and Conference*, 2010.

[4] Yiteng Huang, Jacob Benesty, and Gary W Elko, "Passive acoustic source localization for video camera steering," in *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*. IEEE, 2000, vol. 2, pp. II909–II912.

[5] Jens Meyer and Gary Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*. IEEE, 2002, vol. 2, pp. II–1781.

[6] Thushara D Abhayapala and Darren B Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*. IEEE, 2002, vol. 2, pp. II–1949.

[7] Lalan Kumar and Rajesh M Hegde, "Near-field acoustic source localization and beamforming in spherical harmonics domain," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3351–3361, 2016.

[8] Lalan Kumar, Ardhendu Tripathy, and Rajesh M Hegde, "Robust multi-source localization over planar arrays using music-group delay spectrum," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4627–4636, 2014.

[9] Joseph Hector DiBiase, *A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays*, Ph.D. thesis, Brown University, 2000.

[10] Charles Knapp and Glifford Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[11] Xiong Xiao, Shengkui Zhao, Xionghu Zhong, Douglas L Jones, Eng Siong Chng, and Haizhou Li, "A learning-based approach to direction of arrival estimation in noisy and reverberant environments," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 2814–2818.

[12] Bracha Laufer-Goldshtein, Ronen Talmon, and Sharon Gannot, "Semi-supervised sound source localization based on manifold regularization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 8, pp. 1393–1407, 2016.

[13] Kushagra Singhal and Rajesh M Hegde, "A sparse reconstruction method for speech source localization using partial dictionaries over a spherical microphone array.," in *INTERSPEECH*, 2014, pp. 2209–2213.

[14] Nejem Huleihel and Boaz Rafaely, "Spherical array processing for acoustic analysis using room impulse responses and time-domain smoothing," *The Journal of the Acoustical Society of America*, vol. 133, no. 6, pp. 3995–4007, 2013.

[15] Sachin N Kalkur, Sandeep Reddy C, and Rajesh M Hegde, "Joint source localization and separation in spherical harmonic domain using a sparsity based method," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[16] Boaz Rafaely, *Fundamentals of spherical array processing*, vol. 8, Springer, 2015.

[17] Yotam Peled and Boaz Rafaely, "Method for dereverberation and noise reduction using spherical microphone arrays," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 113–116.

[18] Munhum Park and Boaz Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," *The Journal of the Acoustical Society of America*, vol. 118, no. 5, pp. 3094–3103, 2005.

[19] Boaz Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on speech and audio processing*, vol. 13, no. 1, pp. 135–143, 2005.

[20] Andrew B Goldberg, Ming Li, and Xiaojin Zhu, "Online manifold regularization: A new learning setting and empirical study," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2008, pp. 393–407.

[21] Martin Cooke, Jon Barker, Stuart Cunningham, and Xu Shao, "An audio-visual corpus for speech perception and automatic speech recognition," *The Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 2421–2424, 2006.

[22] DP Jarrett, EAP Habets, MRP Thomas, and PA Naylor, "Rigid sphere room impulse response simulation: Algorithm and applications," *The Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 1462–1472, 2012.