Direction of arrival estimation of Multiple acoustic sources using a maximum likelihood method in the spherical harmonic domain

Yuxiang Hu [1], Jing Lu [1,*] and Xiaojun Qiu [2]

[1] Key Lab of Modern Acoustics, Institute of Acoustics, Nanjing University, Nanjing 210093, China; dg1523051@smail.nju.edu.cn (Y. H.)
[2] Centre for Audio, Acoustics and Vibration, Faculty of Engineering and IT, University of Technology Sydney, NSW 2007, Australia; xiaojun.qiu@uts.edu.au
* Correspondence: lujing@nju.edu.cn; Tel.: +86-025-8359-3373

Academic Editor: name
Received: date; Accepted: date; Published: date

Abstract: Direction of arrival estimation (DOA) of multiple acoustic sources has been used for a wide range of applications, including room geometry inference, source separation and speech enhancement. The beamformer-based and subspace-based methods are most commonly used for spherical microphone arrays; however, the former suffers from spatial resolution limitations, while the later suffers from performance degradation in noisy environment. This letter proposes a multiple source localization approach based on the maximum likelihood method in the spherical harmonic domain and implements an efficient sequential iterative search of maxima on the cost function in the spherical harmonic domain. The proposed method avoids the division of the spherical Bessel function, which makes it suitable for both rigid-sphere and open-sphere configurations. Simulation results show that the proposed method has a significant superiority over the commonly used frequency smoothing multiple signal classification method. Experiments in a normal listening room and a reverberation room validate the effectiveness of the proposed method.

Keywords: multiple source localization; maximum likelihood; spherical harmonic domain; alternating projection

1. Introduction

The rotationally symmetric spatial directivity makes the spherical microphone array an appealing structure in many audio applications, among which the acoustic source localization, or the direction of arrival (DOA) estimation, plays an important role in speech enhancement [1], room impulse response analysis [2], and room geometry inference [3].

Various DOA estimation methods have been proposed, which can be generally classified as beamformer-based [2-5] and subspace-based [6-7]. The beamformer-based methods, such as those based on plane-wave decomposition (PWD) [4] and the minimum variance distortionless response (MVDR) beamformer [3], have the benefit of straightforward implementation, but suffer from low spatial resolution. The subspace-based methods, such as the multiple signal classification (MUSIC) [6], provide a high spatial resolution; however, they suffer from severe performance degradation when the signal-to-noise ratio (SNR) is low [8]. In order to improve the robustness of the DOA estimation of coherent sources, wideband expansion based on focusing matrices or frequency smoothing (FS) techniques has to be employed [7].

We proposed a maximum likelihood DOA estimation method in the spherical harmonic domain (SHMLE) recently, which is an attractive alternative DOA estimation method with advantages of high spatial resolution, strong robustness and straightforward wideband implementation [9]. The proposed SHMLE method only considered one source situation, while two or more sources often need to be localized in many practical applications. In this letter, the SHMLE method is extended to estimate the DOA of multiple sources. Generally speaking, the DOAs can be determined by searching maxima on the maximum likelihood (ML) cost function. However, the commonly used grid search method is only effective in finding the global maximum, which restricts its applicability in one source situation. To achieve effective DOA estimation of multiple sources, an efficient sequential iterative search method is introduced in the spherical harmonic (SH) domain. Experiments using a 32-element spherical microphone array validate the feasibility and superiority of the proposed method.

2. Methods

2.1. Signal model in the spherical harmonic domain

The standard spherical coordinate system is utilized with r, θ and φ representing the radius, the elevation angle and the azimuth, respectively. The sound field is assumed to be composed of plane waves from L sources with $\Psi_l = (\theta_l, \varphi_l)$ ($l = 1, 2, \ldots, L$) being the DOA of the l-th plane wave and $s_l(k)$ being its amplitude, where k denotes the wave number. The Q element spherical microphone array is distributed uniformly on a sphere with a radius of a centred at the origin of the coordinate system, and $\Omega_q = (\theta_q, \varphi_q)$ is the angle position of the q-th microphone [10].

The sound pressure of the q-th microphone for the incident waves can be expressed as [11]

$$p\left(k,\Omega_q\right)=\sum_{l=1}^{L}s_l\left(k\right)e^{\mathrm{i}\mathbf{k}_l^{\mathrm{T}}\mathbf{r}_q}\approx\sum_{l=1}^{L}s_l\left(k\right)\sum_{n=0}^{N}\sum_{m=-n}^{n}b_n\left(k\right)Y_{n,m}^{*}\left(\Psi_l\right)Y_{n,m}\left(\Omega_q\right), \tag{1}$$

where $\mathbf{k}_l = -k(\cos\varphi_l\sin\theta_l, \sin\varphi_l\sin\theta_l, \cos\theta_l)^{\mathrm{T}}$ and $\mathbf{r}_q = a(\cos\varphi_q\sin\theta_q, \sin\varphi_q\sin\theta_q, \cos\theta_q)^{\mathrm{T}}$ denote the wave vector of the lth plane wave and the position of the q-th microphone in the Cartesian coordinate. $Y_{n,m}$ is the spherical harmonic of order n and degree m, N is the highest order number for the plane wave decomposition and satisfies $(N+1)^2 < Q$. The superscript (*) denotes complex conjugation, and $b_n(k)$ is a function of array configuration [11]. Equation (1) can be expressed in matrix form as

$$p\left(k,\Omega_q\right)\approx\mathbf{y}^{\mathrm{T}}\left(\Omega_q\right)\mathbf{B}\left(k\right)\mathbf{Y}^{\mathrm{H}}\left(\mathbf{\Psi}\right)\mathbf{s}\left(k\right), \tag{2}$$

with

$$\mathbf{y}\left(\Omega_q\right)=\left[Y_{0,0}\left(\Omega_q\right),\ Y_{1,-1}\left(\Omega_q\right),\ Y_{1,0}\left(\Omega_q\right),\ Y_{1,1}\left(\Omega_q\right),\ \ldots,\ Y_{N,N}\left(\Omega_q\right)\right]^{\mathrm{T}}, \tag{3}$$

$$\mathbf{y}\left(\Psi_l\right)=\left[Y_{0,0}\left(\Psi_l\right),\ Y_{1,-1}\left(\Psi_l\right),\ Y_{1,0}\left(\Psi_l\right),\ Y_{1,1}\left(\Psi_l\right),\ \ldots,\ Y_{N,N}\left(\Psi_l\right)\right]^{\mathrm{T}}, \tag{4}$$

$$\mathbf{Y}\left(\mathbf{\Psi}\right)=\left[\mathbf{y}\left(\Psi_1\right),\mathbf{y}\left(\Psi_2\right),\ldots,\mathbf{y}\left(\Psi_L\right)\right]^{\mathrm{T}}, \tag{5}$$

$$\mathbf{B}(k)=diag\{b_0(k),\ b_1(k),\ b_1(k),\ b_1(k),\ \ldots,\ b_N(k)\}, \tag{6}$$

$$\mathbf{s}\left(k\right)=\left[s_1\left(k\right),s_2\left(k\right),\ldots,s_L\left(k\right)\right]^{\mathrm{T}}, \tag{7}$$

$$\mathbf{\Psi}=\left[\Psi_1,\Psi_2,\ldots,\Psi_L\right], \tag{8}$$

where the superscript (T) denotes the transpose.

In the presence of additive noise, the sound pressure at all Q microphones can be expressed as

$$\mathbf{p}\left(k,\mathbf{\Omega}\right)\approx\mathbf{Y}\left(\mathbf{\Omega}\right)\mathbf{B}\left(k\right)\mathbf{Y}^{\mathrm{H}}\left(\mathbf{\Psi}\right)\mathbf{s}\left(k\right)+\mathbf{v}\left(k\right), \tag{9}$$

where

$$\mathbf{Y}(\boldsymbol{\Omega})=\left[\mathbf{y}(\Omega_1),\mathbf{y}(\Omega_2),\ldots,\mathbf{y}(\Omega_Q)\right]^{\mathrm{T}},\tag{10}$$

$\mathbf{p}(k,\boldsymbol{\Omega}) = [p(k,\Omega_1), p(k,\Omega_2), \ldots, p(k,\Omega_Q)]^{\mathrm{T}}$ is the vector of the sound pressure of Q microphones, and $\mathbf{v}(k) = [\nu_1(k), \nu_2(k), \ldots, \nu_Q(k)]^{\mathrm{T}}$ is the vector of the additive sensor noise added to the system. The uncorrelated noise is assumed to be zero mean complex Gaussian and, for simplicity, be spatially white with a covariance matrix $\mathbf{R}_{\mathrm{v}}(k)=\sigma_{\nu}^2\mathbf{I}_Q$, where σ_{ν}^2 is the unknown noise variance and $\mathbf{I}_Q$ is the identity matrix of order $Q\times Q$.

For the uniformly spatial sampling configuration used in this letter, the following orthogonal relation holds (note that $(N+1)^2 \le Q$) [10]

$$\frac{4\pi}{Q}\mathbf{Y}^{\mathrm{H}}(\boldsymbol{\Omega})\mathbf{Y}(\boldsymbol{\Omega})=\mathbf{I}_{(N+1)^2}.\tag{11}$$

The SH transform can be carried out by multiplying both sides of Eq. (9) from the left with $\frac{4\pi}{Q}\mathbf{Y}^{\mathrm{H}}(\boldsymbol{\Omega})$, which yields

$$\mathbf{p}_{\mathbf{nm}}(k)\approx\mathbf{B}(k)\mathbf{Y}^{\mathrm{H}}(\boldsymbol{\Psi})\mathbf{s}(k)+\mathbf{v}_{\mathbf{nm}}(k),\tag{12}$$

where $\mathbf{p}_{\mathbf{nm}}(k)$ is a vector containing $(N+1)^2$ SH domain coefficients, i.e.,

$$\mathbf{p}_{\mathbf{nm}}(k)=\left[p_{0,0}(k),\ p_{1,-1}(k),\ p_{1,0}(k),\ p_{1,1}(k),\ \ldots,\ p_{N,N}(k)\right]^{\mathrm{T}}.\tag{13}$$

The second term on the right side of Eq. (12) is the noise expressed in the SH domain, i.e. $\mathbf{v}_{\mathbf{nm}}(k)=\frac{4\pi}{Q}\mathbf{Y}^{\mathrm{H}}(\Omega)\mathbf{v}(k)$, with the mean

$$E\left[\mathbf{v}_{\mathbf{nm}}(k)\right]=\frac{4\pi}{Q}\mathbf{Y}^{\mathrm{H}}(\boldsymbol{\Omega})E\left[\mathbf{v}(k)\right]=\mathbf{0},\tag{14}$$

and the covariance matrix

$$\mathbf{R}_{\mathbf{nm}}(k)=E\left[\frac{4\pi}{Q}\mathbf{Y}^{\mathrm{H}}(\boldsymbol{\Omega})\mathbf{v}(k)\mathbf{v}^{\mathrm{H}}(k)\mathbf{Y}(\boldsymbol{\Omega})\frac{4\pi}{Q}\right]=\frac{4\pi}{Q}\cdot\sigma_{\nu}^2\mathbf{I}_{(N+1)^2},\tag{15}$$

where $E(\cdot)$ denotes the statistical expectation. Apparently, the noise model in the SH domain is also zero-mean complex Gaussian.

2.2. Sound source localization in the spherical harmonic domain

Define $\boldsymbol{\Theta}=\left[\boldsymbol{\Psi}^T,\mathbf{S}^T,\sigma_n^2\right]^{\mathrm{T}}$ as the vector of all unknown parameters, where $\mathbf{S}=\left[\mathbf{s}(k_{\min})^{\mathrm{T}},\ldots,\mathbf{s}(k_{\max})^{\mathrm{T}}\right]^{\mathrm{T}}$ contains the amplitudes of the source signals with k_{min} and k_{max} representing the minimum and maximum wave numbers and satisfying $ka \le N$. Throughout this letter, $\boldsymbol{\Psi}$, s and σ_v^2 are assumed to be deterministic and unknown, while the observed data $\mathbf{p}_{\mathbf{nm}}$ is considered to be random [12]. The likelihood function of $\mathbf{p}_{\mathbf{nm}}$ given $\boldsymbol{\Theta}$ in the SH domain can be expressed as [9,12]

$$f(\mathbf{p}_{\mathbf{nm}};\boldsymbol{\Theta})=\frac{\exp\left\{-\sum_{k=k_{min}}^{k_{max}}\left[\mathbf{p}_{\mathbf{nm}}(k)-\mathbf{V}_{\mathbf{nm}}(k,\boldsymbol{\Psi})\mathbf{s}(k)\right]^{\mathrm{H}}\mathbf{R}_{\mathbf{nm}}^{-1}\left[\mathbf{p}_{\mathbf{nm}}(k)-\mathbf{V}_{\mathbf{nm}}(k,\boldsymbol{\Psi})\mathbf{s}(k)\right]\right\}}{\left(\pi^{(N+1)^2}\left|\mathbf{R}_{\mathbf{nm}}\right|\right)^{k_{max}-k_{min}}},\tag{16}$$

where $\mathbf{V}_{\mathbf{nm}}(k,\Psi) = \mathbf{B}(k)\mathbf{Y}^{\mathrm{H}}(\boldsymbol{\Psi})$, and $|\cdot|$ denotes the matrix determinant. The solution to Eq. (16) is given by [9]

$$\hat{\mathbf{\Psi}} = \arg\min_{\mathbf{\Psi}} \sum_{k=k_{min}}^{k_{max}} \left\| \mathbf{p}_{\mathbf{nm}}(k) - \mathbf{V}_{\mathbf{nm}}(k,\mathbf{\Psi})\mathbf{V}_{\mathbf{nm}}(k,\mathbf{\Psi})^{\dagger}\, \mathbf{p}_{\mathbf{nm}}(k) \right\|^2, \tag{17}$$

where $(\cdot)^{\dagger}$ denotes pseudo-inverse operation.

Define the cost function as

$$J(\mathbf{\Psi}) = -10\log_{10}\left(\sum_{k=k_{min}}^{k_{max}} \left\| \mathbf{p}_{\mathbf{nm}}(k) - \mathbf{V}_{\mathbf{nm}}(k,\mathbf{\Psi})\mathbf{V}_{\mathbf{nm}}(k,\mathbf{\Psi})^{\dagger}\, \mathbf{p}_{\mathbf{nm}}(k) \right\|^2 \right), \tag{18}$$

then the wideband estimator can be described as

$$\hat{\mathbf{\Psi}} = \arg\max_{\mathbf{\Psi}} J(\mathbf{\Psi}). \tag{19}$$

The SHMLE has the remarkable benefit of easy wideband implementation as described in Eqs. (17)-(19). This is superior over the other methods in the spherical harmonic domain, which usually require a quite cumbersome frequency smoothing (FS) technique to realize wideband DOA [7]. Compared with the maximum likelihood method in Ref. 13, the division of $b_n(k)$ is avoided, which makes the method proposed in this letter suitable for both rigid-sphere and open-sphere arrays.

2.3. DOA estimation of multiple sources

For one source situation, Eq. (19) can be solved using the grid search method. For P grid points and L sources situation, the computational load of Eq. (19) is $O(P^L)$, which is computationally prohibitive. Moreover, effective discrimination of the multiple maxima in the cost function is very difficult even if repetitive traversal is feasible. To alleviate these problems, a nonlinear optimization algorithm is applied in the SH domain with implementation of the alternating projection method [14]. The alternating projection method avoids the multidimensional search by estimating the location of one source sequentially while fixing the estimates of other source locations from the previous iteration.

For nonlinear optimization methods, the initial locations of the sound sources is critical to reach the global maximum. In this letter, the simplified gird search method is adopted to find initial locations, and the procedure of the method is described as follows.

(1) Estimate the location of the first source s_1 on a single source grid with

$$\Psi_1^{(0)} = \arg\max_{\Psi_1} J(\Psi_1). \tag{20}$$

(2) For $l = 2, \ldots, L$, estimate the location of the lth source s_l, assuming locations of the first $l-1$ sources are fixed by using

$$\Psi_l^{(0)} = \arg\max_{\Psi_2} J\left(\left[\mathbf{\Psi}_{l-1}^{(0)}, \Psi_l\right]\right), \tag{21}$$

where $\mathbf{\Psi}_{l-1}^{(0)}$ denotes the initial locations of first $l-1$ sources, i.e.

$$\mathbf{\Psi}_{l-1}^{(0)} = \left[\Psi_1^{(0)}, \Psi_2^{(0)}, \cdots, \Psi_{l-1}^{(0)}\right]. \tag{22}$$

In steps (1) and (2), Eq. (19) only needs to be calculated $P \times L$ times and the effective initial location information can be obtained. In some cases, this initialization process is not necessary since a good initial location estimate is available, for example, from the estimate of the previous data for slowly moving sources.

After initialization, the accurate locations can be estimated using a nonlinear optimization algorithm with implementation of the alternating projection method [14]. The location of Ψ_l at the $(i+1)$th iteration can be estimated by solving the one-dimensional maximization problem

$$\Psi_l^{(i+1)} = \arg\max_{\Psi_l} J\left(\left[\ \Psi_l, \mathbf{\Psi}_{\mathbf{s}}^{(i)},\right]\right), \tag{23}$$

where $\mathbf{\Psi}_{\mathbf{s}}^{(i)}$ denotes the estimated locations of other $L-1$ sources, i.e.

$$\mathbf{\Psi}_{\mathbf{s}}^{(i)} = \left[\Psi_1^{(i+1)}, \cdots, \Psi_{l-1}^{(i+1)}, \Psi_{l+1}^{(i)}, \cdots, \Psi_L^{(i)}\right]. \tag{24}$$

In Eq. (23), the Quasi-Newton (QN) method with Broyden-Fletcher-Goldfarb-Shanno algorithm [15] is used, and the QN method is available in MATLAB as in the *fminunc* function. For the beamformer-based and subspace-based methods, the localization results are acoustic maps [7]. An extra algorithm is needed to identify the location of the sound sources from the acoustic maps. On the contrary, the method proposed in this letter can automatically provide the localization results, as described in Eq. (23). Furthermore, the localization precision of the proposed method can be extremely high, as will be demonstrated in the following simulations and experiments.

3. Simulations and experiments

In this section, the performance of the proposed method, i.e., the multiple source SHMLE (MS-SHMLE), is investigated and compared to the FSMUSIC method [7], which has the benefits of high spatial resolution and easy implementation. The Eigenmike® [16] microphone array model, with $Q = 32$ microphones arranged uniformly on a sphere with radius $a = 4.2$ cm (depicted in Fig. 1), was used in both simulations and experiments. Only two source cases were considered in simulations and experiments, and the proposed method can be easily utilized to other scenarios with more sources as described in Sec. 2.3. The source signals are independent white Gaussian noise sampled at a sampling rate of $f_s = 16$ kHz, and a frame of 1024 samples is extracted from the recordings. The localization frequency range is $ka \in [2.5\ 3.5]$. The grid resolution of FSMUSIC is 1°.

(**a**) (**b**)

Figure 1. Eigenmike® and two sound sources in (a) normal listening room (b) reverberation room.

Root mean squared error (RMSE) is used to assess the performance of the localization results, which is defined as

$$\mathrm{RMSE} = \sqrt{E\left\{\left(\mathbf{\Psi} - \hat{\mathbf{\Psi}}\right)\left(\mathbf{\Psi} - \hat{\mathbf{\Psi}}\right)^{\mathrm{T}}\right\}}. \tag{25}$$

3.1. Simulations with different SNR

Figure 2 depicts the RMSE of the FSMUSIC and the MS-SHMLE as a function of SNR in a room with reverberation time of 0.3 s. In this simulation, the room dimensions are 6×7×5 m^3, the microphone array is located at [3, 2.5, 1.5] m and the speakers are placed 1.0 m away from the array center. Sound sources incident from directions of (90°, 180°) and (90°, 120°). The RMSE is averaged over 100 different trials. The room impulse responses between the sound sources and the microphones positioned on the rigid sphere are simulated using the method proposed in Ref. 17. It can be seen that the RMSE of the MS-SHMLE is better than that of FSMUSIC especially for the low SNR situations, and FSMUSIC fails to present meaningful DOAs when the SNR is lower than 2 dB.

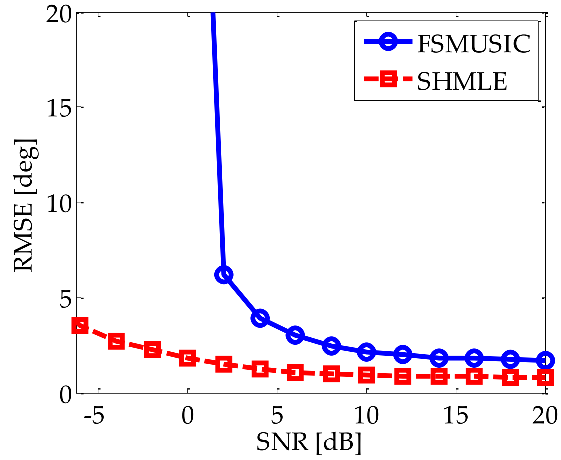


Figure 2. Localization RMSE of the FSMUSIC and the MS-SHMLE versus SNR in a room with reverberation time of 0.3 s

3.2. Simulations under different reverberation time

When the reverberation time is higher than 1.5 s, the method proposed in Ref. 17 is not suitable to simulate the room impulse responses because of its high computational burden and memory requirement. Therefore, an open-sphere array configuration is used in this simulation and the room impulse responses are simulated using the method proposed in Ref. 18. When then open-sphere configuration is used, the FSMUSIC method suffers from ill-conditioning around the zeros of the spherical Bessel function, and a mitigation method proposed in Ref. 19 is utilized.

Figure 3 depicts the localization RMSE of the FSMUSIC and the MS-SHMLE as a function of the reverberation time T_{60}. In this simulation, the SNR is fixed at 15 dB. It can be seen that the RMSE of the MS-SHMLE is better than that of FSMUSIC under different reverberation time.

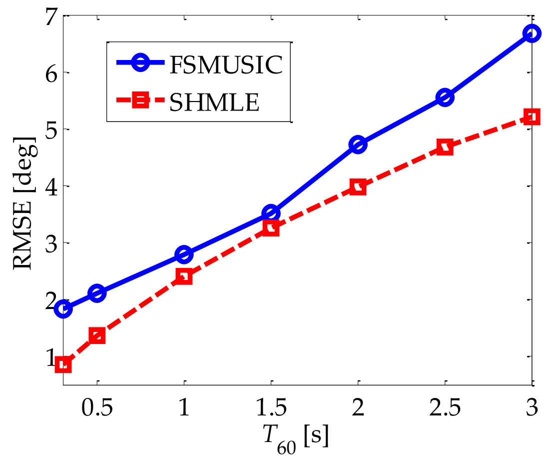


Figure 3. Localization RMSE of the FSMUSIC and the MS-SHMLE versus reverberation time with an SNR of 15 dB.

3.3. Two-source experiments in a listening room

The experiments for DOA estimation of two sources were carried out in a listening room with background noise less than 30 dBA as depicted in Fig. 1(a). The room dimensions are 5×8×4 m^3 and the reverberation time is around 0.3 s. The microphone array was located at [2.5, 3, 1.5] m. Two sound sources were placed 1.5 m away from the array with 7 different angle differences $\Delta\varphi$. The sound pressure level (SPL) difference of the two sources at the array center is around 2 dB.

Figure 4 depicts the localization results for two sources case using the FSMUSIC methods. The DOA of the sound source is denoted by a solid black circle in all these figures. It can be found that FSMUSIC can distinguish both sources when the angle difference between the two sources is larger than 20°, as depicted in Fig 4(a) and (b). When the separation angle between sources is close to 20°, FSMUSIC fails to identify two sources, as depicted in Fig 4(c), because the local maxima of the weaker source is totally merged into the main lobe of the stronger one. It should be noted that $\mathbf{V}_{nm}(k,\Psi)$ in Eq. (17) contains the steering vector of all sources. Therefore, the acoustic maps depicted in Fig. 4 are not suitable for the MS-SHMLE method.

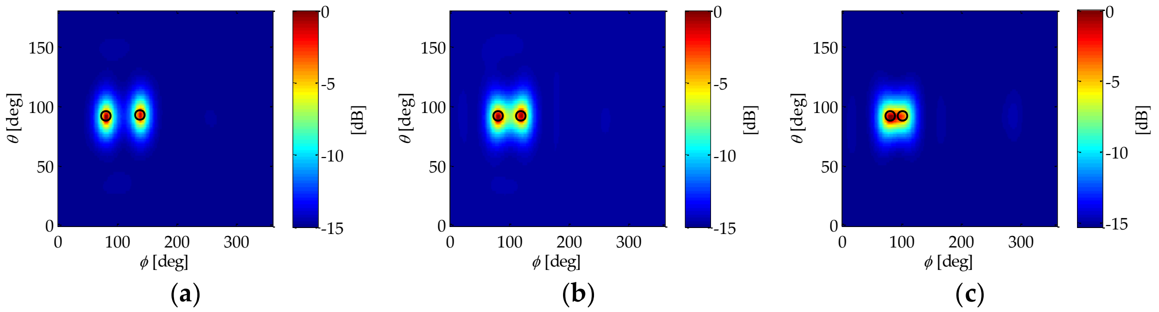


Figure 4. Localization results for two sources case using the FSMUSIC method in a listening room with (a) $\Delta\varphi = 60°$, (b) $\Delta\varphi = 40°$ and (c) $\Delta\varphi = 20°$.

Table 1 shows the RMSE of the FSMUSIC and the MS-SHMLE for a 10 s recorded data. Although the RMSE of these two methods are close, the FSMUSIC can only locate the stronger one when the separation angle between the two sources is close to 20°, while the MS-SHMLE can distinguish both sources. The average RMSE of the MS-SHMLE is comparatively lower than FSMUSIC, which is in consistence with the simulations results.

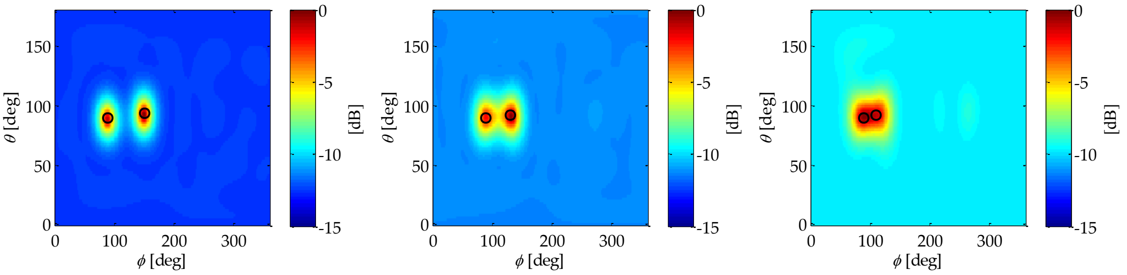
Table 1. RMSE of the FSMUSIC and the MS-SHMLE for two sources case in a listening room

Angle difference	RMSE of FSMUSIC			RMSE of MS-SHMLE		
	Strong	weak	total	strong	weak	total
180 °	0.73 °	1.19 °	0.99 °	0.32 °	1.66 °	1.20 °
120 °	0.59 °	0.85 °	0.73 °	0.32 °	0.91 °	0.68 °
90 °	0.76 °	0.89 °	0.83 °	0.42 °	0.90 °	0.70 °
60 °	0.77 °	0.71 °	0.74 °	0.32 °	0.79 °	0.60 °
40 °	0.83 °	0.78 °	0.81 °	0.34 °	1.05 °	0.78 °
30 °	1.08 °	1.65 °	1.39 °	0.44 °	1.76 °	1.28 °
20 °	1.66 °	-	-	0.63 °	1.54 °	1.18 °

3.4. Two-source experiments in a reverberation room

To further validate the robustness of the proposed algorithm in high reverberant environments, the experiments for DOA estimation of two sources were also carried out in a reverberation room as depicted in Fig. 1(b). The room dimensions are 5.9×7.35×5.22 m^3. The reverberation time is around 3 s at frequency range $ka \in [2.5\ 3.5]$. In the experiments, the microphone array was located at [3 2.5 1.5] m, and the sound sources were placed 1.5 m away from the array.

Figure 5 depicts the localization results for the two sources case using the FSMUSIC method. Similar to the results in the listening room, when the angle difference between the two sources is larger than 20°, FSMUSIC can distinguish both sources as depicted in Fig 5(a) and (b). When the separation angle between sources is close to 20°, FSMUSIC can only locate the stronger source while fail to identify the weaker one, as depicted in Fig 5(c).



(a) (b) (c)

Figure 5. Localization results for two sources case using the FSMUSIC method in a listening room with (a) $\Delta\varphi = 60°$, (b) $\Delta\varphi = 40°$ and (c) $\Delta\varphi = 20°$.

Table 2 shows the RMSE of the FSMUSIC and the MS-SHMLE for a10 s recorded data. It can be seen that the localization RMSE in the reverberation room is higher than that in the listening room. The RMSE of these two methods increase significantly when the angle difference between the two sources is close to or lower than 30°. When the separation angle between the two sources is close to 20°, the FSMUSIC can only locate the stronger one, while the MS-SHMLE can distinguish both sources. The superiority of the MS-SHMLE in high reverberant environment coincides well with the simulations presented in Sec. 3.2.

Table 2. RMSE of the FSMUSIC and the MS-SHMLE for the two sources case in a reverberation room

Angle difference	RMSE of FSMUSIC			RMSE of MS-SHMLE		
	strong	weak	Total	Strong	weak	total
180°	1.38°	1.38°	1.38°	1.42°	1.22°	1.32°
120°	1.25°	1.17°	1.21°	1.30°	1.19°	1.25°
90°	1.49°	1.17°	1.34°	1.40°	1.07°	1.24°
60°	1.54°	1.20°	1.38°	1.29°	0.98°	1.15°
40°	1.53°	1.34°	1.44°	1.73°	1.42°	1.59°
30°	4.22°	3.46°	3.86°	1.99°	2.02°	2.00°
20°	4.57°	-	-	3.56°	4.02°	3.80°

4. Conclusion

This letter proposes a multiple source localization method in the spherical harmonic domain using the maximum likelihood strategy. To avoid high-dimensional grid search with extremely high computational burden, a nonlinear optimization algorithm with implementation of the alternating projection method is introduced, leading to an efficient MS-SHMLE method. The proposed method avoids the division of the spherical Bessel function, which makes it suitable for both rigid-sphere and open-sphere configurations. Simulations and experiments on a 32-microphone model demonstrate that the proposed MS-SHMLE method has very good spatial resolution and can distinguish two sources with 20° angle difference in both normal listening room and reverberation room. Furthermore, the performance is stable in low SNR environment, circumventing the problem faced by the subspace-based method.

References

1. Kumatani, K.; McDonough, J.; Raj, B. Microphone array processing for distant speech recognition: From close-talking microphones to far-field sensors. *IEEE Signal Process. Mag.* **2012**, *29*, 127-140, DOI: 10.1109/MSP.2012.2205285.
2. Khaykin, D.; Rafaely, B. Acoustic analysis by spherical microphone array processing of room impulse responses. *J. Acoust. Soc. Am.* **2012**, *132*, 261-270, DOI: http://dx.doi.org/10.1121/1.4726012.
3. Mabande, E.; Kowalczyk, K.; Sun, H.; Kellermann, W. Room geometry inference based on spherical microphone array eigenbeam processing. *J. Acoust. Soc. Am.* **2013**, *134*, 2773-2789, DOI: http://dx.doi.org/10.1121/1.4820895.
4. Park, M.; Rafaely, B. Sound-field analysis by plane-wave decomposition using spherical microphone array. *J. Acoust. Soc. Am.* **2005**, *118*, 3094-3103, DOI: http://dx.doi.org/10.1121/1.2063108.
5. Zhang, L.; Ding, D.; Yang, D.; Wang, J.; Shi, J. Sound Source Localization Using Non-Conformal Surface Sound Field Transformation Based on Spherical Harmonic Wave Decomposition. *Sensors* **2017**, *17*, 1-12, DOI: 10.3390/s17051087.
6. Nadiri, O.; Rafaely, B. Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2014**, *22*, 1494-1505, DOI: 10.1109/TASLP.2014.2337846.
7. Sun, H.; Mabande, E.; Kowalczyk, K.; Kellermann, W. Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing. *J. Acoust. Soc. Am.* **2012**, *131*, 2828-2840, DOI: http://dx.doi.org/10.1121/1.3688476.
8. Mestre, X.; Lagunas, M. Á. Modified subspace algorithms for DoA estimation with large arrays. *IEEE Trans. Signal Process.* **2008**, *56*, 598-614, DOI: 10.1109/TSP.2007.907884.
9. Hu, Y.; Lu, J.; Qiu, X. A maximum likelihood direction of arrival estimation method for open-sphere microphone arrays in the spherical harmonic domain. *J. Acoust. Soc. Am.* **2015**, *138*, 791-794, DOI: http://dx.doi.org/10.1121/1.4926907.
10. Rafaely, B. Analysis and design of spherical microphone arrays. *IEEE Trans. Speech Audio Process.* **2005**, *13*, 135-143, DOI: 10.1109/TSA.2004.839244.
11. Rafaely, B. Fundamentals of spherical array processing. Springer: Berlin, 2015; pp. 57-99, ISBN 978-3-662-45663-7.
12. Chen, C. E.; Lorenzelli, F.; Hudson, R. E.; Yao, K. Maximum likelihood DOA estimation of multiple wideband sources in the presence of nonuniform sensor noise. *EURASIP J. Adv. Signal Process.* **2007**, *2008*, 1-12, DOI: https://doi.org/10.1155/2008/835079.
13. Tervo, S.; Politis, A. Direction of arrival estimation of reflections from room impulse responses using a spherical microphone array. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 1539-1551, DOI: 10.1109/TASLP.2015.2439573.
14. Chen, J. C.; Hudson, R. E.; Yao, K. Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field. *IEEE Trans. Signal Process.* **2002**, *50*, 1843-1854, DOI: 10.1109/TSP.2002.800420.
15. Dennis, Jr, J. E.; Moré, J. J. Quasi-Newton methods, motivation and theory. *SIAM review* **1977**, *19*, 46-89, DOI: https://doi.org/10.1137/1019005.
16. Meyer, J.; Elko, G. A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Orlando, FL, USA, May 2002; pp. 1781-1781, DOI: 10.1109/ICASSP.2002.5744968.
17. Jarrett, D. P.; Habets, E. A. P.; Thomas, M. R. P.; Naylor, P. A. Rigid sphere room impulse response simulation: Algorithm and applications. *J. Acoust. Soc. Am.* **2012**, *132*, 1462-1472, DOI: http://dx.doi.org/10.1121/1.4740497.
18. Allen, J. B.; Berkley, D. A. Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am.* **1979**, *65*, 943–950, DOI: http://dx.doi.org/10.1121/1.382599.
19. Rafaely, B. Bessel nulls recovery in spherical microphone arrays for time-limited signals. *IEEE Trans. Audio Speech Lang. Process.* **2011**, *19*, 2430-2438, DOI: 10.1109/TASL.2011.2136338.