24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

LOCALIZATION OF MULTIPLE ACOUSTIC SOURCES USING A MAXIMUM LIKELIHOOD METHOD IN THE SPHERICAL HARMONIC DOMAIN

Yuxiang Hu and Jing Lu

Key Lab of Modern Acoustics, Institute of Acoustics, Nanjing University, Nanjing, China 210093
email: dg1523051@nju.edu.cn

Localization of multiple sources have been studied 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. The beamformer-based methods suffer from spatial resolution limitations, while subspace-based methods, such as multiple signal classification (MUSIC), suffer from performance degradation in noisy environment. This paper proposes a multiple source localization approach based on the maximum likelihood source localization method in the spherical harmonic domain, with implementation of an efficient alternating projection procedure based on sequential iterative search of signal source locations. The proposed method avoids the division of the spherical Bessel function, which makes it suitable for both the rigid-sphere and the open-sphere configurations. Simulation results show that the proposed method has a significant superiority over the commonly used methods, including the MUSIC and the PWD methods. Experiments in a normal listening room validate the effectiveness of the proposed method.
Keywords: source localization, maximum likelihood, spherical microphone array

1. Introduction

Spherical microphone arrays have the benefit of a rotationally symmetric spatial directivity because of its three-dimension symmetric configuration. This makes spherical microphone arrays 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]-[4] and subspace-based [5]-[6]. Beamformer-based methods, such as the plane-wave decomposition (PWD) [4] and the minimum variance distortionless response (MVDR) [3], although considered simple, suffer from spatial resolution limitations. Therefore, these methods are unable to localize spatially adjacent sources. Subspace-based methods, such as the multiple signal classification (MUSIC) [5], provide a high spatial resolution. However, for the subspace-based methods, the number of sources must be known a prior in order to obtain accurate results. Besides, subspace-based methods suffer from severe performance degradation when the signal-to-noise (SNR) falls below a certain threshold [7]. Therefore, subspace is not suitable for the environment with low SNR. Furthermore, in order to improve the robustness of DOA estimation of coherent sources, focusing matrices and frequency smoothing (FS) techniques have to be employed [6].

A maximum likelihood DOA estimation method in the spherical harmonic domain (SHMLE) is proposed in Ref. 8, which is an attractive alternative DOA estimation method in the spherical harmonic domain circumventing the problems mentioned above. The SHMLE method provides higher

spatial resolution and stable localization performance in the low SNR environment. Besides, the wide-band implementation of the SHMLE method is very straightforward.

It should be noted that the SHMLE method proposed in Ref. 8 only considers the one source situation, and the performance in localizing multiple sources is yet to be analysed. In this paper, the SHMLE method is extended to estimate the DOA of multiple sources. Furthermore, the grid search method used in Ref. 8 is not suitable for the multiple source situation due to its high computational burden. To solve this problem, an efficient nonlinear optimization algorithm with implementation of the alternating projection method is introduced for the localization of multiple sources, which reduces the computational burden significantly. The performance of the proposed method is compared with the PWD and the MUSIC methods. Experiments using a 32-element spherical microphone array validate the superiority of the proposed method.

2. Method

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 L plane waves with $\Psi_l = (\theta_l, \phi_l)$ (l = 1, 2, …, 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 $\Omega_q = (\theta_q, \phi_q)$ is the angle position of the q-th microphone [9].

The sound pressure of the q-th microphone for the incident wave can be expressed as [10]

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

where $\mathbf{k}_l = -k(\cos\phi_l\sin\theta_l, \sin\phi_l\sin\theta_l, \cos\theta_l)^{\mathrm{T}}$ and $\mathbf{r}_q = a(\cos\phi_q\sin\theta_q, \sin\phi_q\sin\theta_q, \cos\theta_q)^{\mathrm{T}}$ denote the wave vector of the plane wave and 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. $b_n(k)$ is a function of array configuration [10]. Equation (1) can be expressed in matrix form as

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

with

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

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

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

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

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

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

where the superscript ($^{\mathrm{T}}$) denotes the transpose.

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

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

where

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

$\mathbf{p}(k,\boldsymbol{\Omega}) = [p(k,\Omega_1), p(k,\Omega_2), \ldots, p(k,\Omega_Q)]^{\mathrm{T}}$ 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 additive sensor noise added to the system. The noise is assumed to be complex Gaussian, to be uncorrelated with the signal, to have zero mean, and for simplicity, to

be spatially white with a covariance matrix $\mathbf{R}_{\nu}(k)=\sigma_{\nu}^{2}\mathbf{I}_{Q}$, 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 $(N+1)^2 \le Q$) [8]

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

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

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

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

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

and the covariance matrix

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

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 $\mathbf{\Theta}=\left[\mathbf{\Psi}^{T},\mathbf{S}^{T},\sigma_{n}^{2}\right]^{\mathrm{T}}$ as the vector of all the 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 paper, $\mathbf{\Psi}$, s and σ_{ν}^{2} are assumed to be deterministic and unknown, while the observed data $\mathbf{p}_{\mathbf{nm}}$ is considered random [11]. The likelihood function of $\mathbf{p}_{\mathbf{nm}}$ given $\mathbf{\Theta}$ in the SH domain can be expressed as [8], [11]

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

where $\mathbf{V}_{\mathbf{nm}}(k,\Psi)=\mathbf{B}(k)\mathbf{Y}^{\mathrm{H}}(\mathbf{\Psi})$ and $|\cdot|$ denotes the matrix determinant. The solution to Eq. (16) is given by [8]

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

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

$$\hat{\boldsymbol{\Psi}} = \arg\max_{\boldsymbol{\Psi}} J\left(\boldsymbol{\Psi}\right). \tag{19}$$

The narrowband SHMLE only requires the solution of Eq. (19) at a specific k, which can be described as

$$\hat{\boldsymbol{\Psi}} = \arg\max_{\boldsymbol{\Psi}} \left(-20\log_{10} \left\| \mathbf{p}_{\mathbf{nm}}\left(k\right) - \mathbf{V}_{\mathbf{nm}}\left(k,\boldsymbol{\Psi}\right) \mathbf{V}_{\mathbf{nm}}\left(k,\boldsymbol{\Psi}\right)^{\dagger} \mathbf{p}_{\mathbf{nm}}\left(k\right) \right\| \right) \tag{20}$$

Although it is simple, the narrowband solution suffers from performance degradation in the presence of coherent noise, e.g., the early reflections of the source signal [6]. Fortunately, the SHMLE has the remarkable benefit of easy implementation of wideband DOA as described in Eq. (17)-(19). This is superior over the existing methods, which usually require a quite cumbersome frequency smoothing (FS) technique to realize wideband DOA [6]. Compared with the maximum likelihood method proposed in Ref. 12, the division of $b_n(k)$ is avoided as shown in Eqs. (17)-(20), which makes the method proposed in this paper a better choice for open-sphere arrays.

2.3 DOA estimation of multiple sources

For the one source situation, Eq. (19) can be solved using grid search method. However, for the multiple source case, an exhaustive multidimensional grid search is computationally prohibitive. A nonlinear optimization method is applied in this paper with implementation of the alternating projection method [13]. The alternating projection approach breaks the multidimensional parameter search into a sequence of single source parameter search and yields fast convergence rate. DOA estimation of two-sources is described as follows, and it can be easily extended to scenarios with more sources.

Alternating Projection Algorithm in spherical harmonic domain:

Step 1) Estimate the location of the stronger source s_1 on a signal source grid

$$\Psi_{s_1}^{(0)} = \arg\max_{\Psi_{s_1}} J\left(\Psi_{s_1}\right). \tag{21}$$

Step 2) Estimate the location of the weaker source s_2 on a single source grid under the assumption of a two source model while keep the first source location estimate from step 1) constant:

$$\Psi_{s_2}^{(0)} = \arg\max_{\Psi_{s_2}} J\left(\left[\Psi_{s_1}^{(0)}, \Psi_{s_2}\right]\right). \tag{22}$$

For $i = 1, \ldots,$ repeat step 3) and 4) until the localization results of s_1 and s_2 between adjacent iterations are smaller than 0.001°.

Step 3) Using nonlinear optimization method to search the location of the first source while keeping the estimate of the second source location from the previous iteration constant:

$$\Psi_{s_1}^{(i)} = \arg\max_{\Psi_{s_1}} J\left(\left[\Psi_{s_1}, \Psi_{s_2}^{(i-1)}\right]\right). \tag{23}$$

Step 4) Using nonlinear optimization method to search the location of the second source while keeping the estimate of the first source location from step 3) constant:

$$\Psi_{s_2}^{(i)} = \arg\max_{\Psi_{s_2}} J\left(\left[\Psi_{s_1}^{(i)}, \Psi_{s_2}\right]\right). \tag{24}$$

In step 1) and 2), a rough grid search provides sufficient initial source location information. In step 3) and 4), we use the Quasi-Newton (QN) method with Broyden-Fletcher-Goldfarb-Shanno algorithm [14], and the QN method is available in MATLAB as in the *fminunc* function.

3. Simulation and experiments

The Eigenmike® [15] microphone array model, as depicted in Fig. 1 with $Q = 32$ microphones arranged uniformly on a sphere with radius $a = 4.2$ cm, is used in both simulations and experiments. The performance of the wideband SHMLE method is compared with the commonly used FSPWD [6] and FSMUSIC [2] method with frequency range $ka \in [2.5\ 3.5]$.

Figure 1: Eigenmike microphone and two sound sources.

3.1 Anechoic simulation

Figure 2 depicts the simulated wideband localization results of FSPWD, PSMUSIC and SHMLE. The sound source is placed at $(\theta_l, \phi_l) = (90^o, 180^o)$. The source signal is white Gaussian noise sampled at f_s = 16 kHz, and a frame of 1024 samples has been extracted from the recording. The SNR is set at 20 dB. The DOA of the sound source is denoted by a solid black circle in all these figures. It can be seen that the peaks in the acoustic maps matches the DOA of the real sound sources. The spatial resolution of FSMUSIC and SHMLE are much better than that of FSPWD.

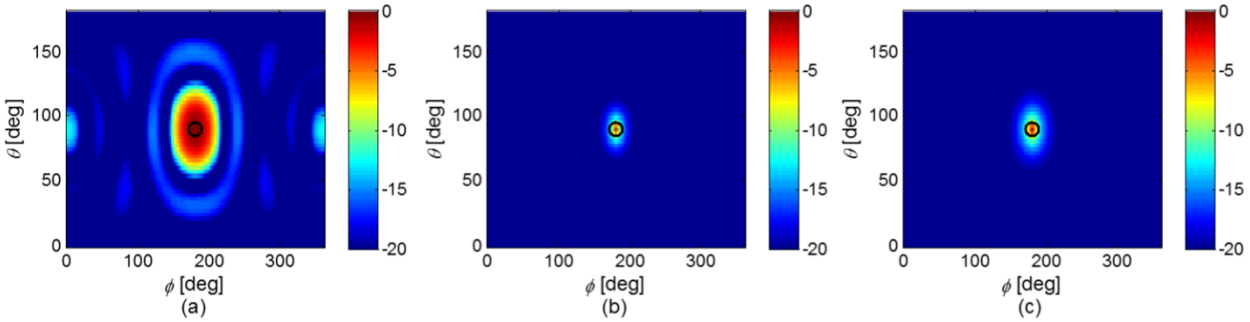

Figure 2: Simulated wideband localization results of (a) FSPWD (b) FSMUSIC and (c) SHMLE.

3.2 Localization performance versus SNR in a reverberation room

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

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

Figure 3 depicts the localization RMSE of FSPWD, PSMUSIC and SHMLE in a reverberant room. In this simulation, the room dimensions are 4×6×3 m^3, the microphone array is located at [2.5, 3, 1.5] m and the speakers are placed 1.0 m away from the array center with 100 uniform distributed directions. 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. 16. The reverberation time is set at 0.3 s and the SNR varies from −10 dB to 20 dB. When the SNR is larger than 0 dB, the RMSE of these three methods are nearly the same. It can also

be seen that the performance of FSMUSHC breaks down when SNR is lower than −2 dB. At low SNR situation, the SHMLE has the best performance.

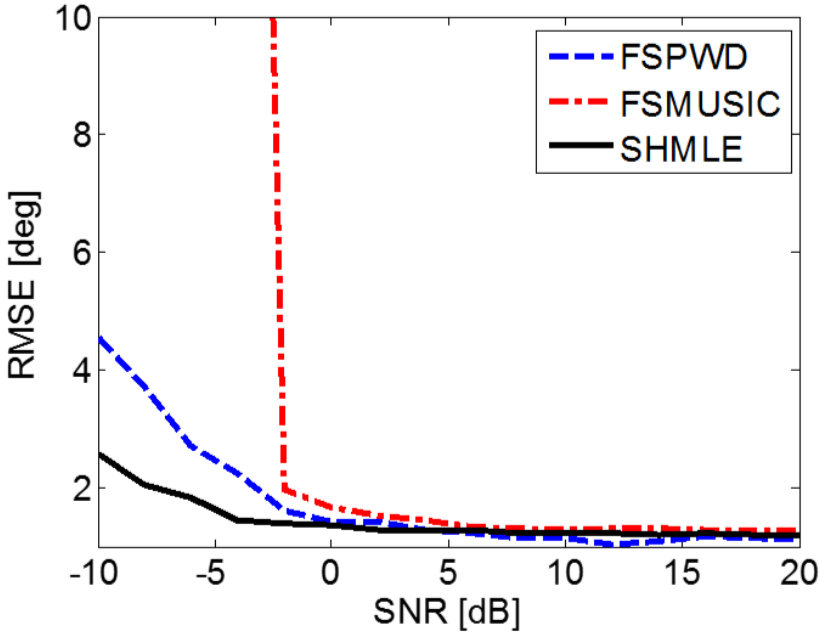

Figure 3: Localization RMSE of (a) FSPWD (b) FSMUSIC and (c) SHMLE in a reverberant room.

3.3 Multiple-source experiments in a listening room

The experiments of DOA estimation of two sources are carried out in a listening room with background noise less than 30 dBA as depicted in Fig. 1. The room dimensions are 5×8×4 m^3. The reverberation time is around 0.3 s. In the experiments, the microphone array is located at [2.5 3 1.5] m. Sound sources incident from eight different directions with a distance of 1.5 m, with elevation and azimuth angles shown in Table I. In each experiment, the sound sources consist of s_1 and one of the other seven sources, and s_1 is stronger than the other sources.

Table 1: Eight sound sources incident direction

Source	s_1	s_2	s_3	s_4
(θ, ϕ)	(91.5°, 81.3°)	(94.2°, 259.1°)	(93.7°, 207.5°)	(93.1°, 178.1°)
Source	s_5	s_6	s_7	s_8
(θ, ϕ)	(92.5°, 138.7°)	(92.3°, 120.2°)	(92.2°, 108.1°)	(91.8°, 100.8°)

Figure 4 depicts the localization results for two sources case using FSPWD and FSMUSIC methods. It can be seen that when the separation angle between the two sources is larger than 60°, both FSPWD and FSMUSIC can distinguish them as depicted in Fig 4(a) and (d). When the separation angle between sources is around 40°, the FSPWD can only locate the stronger source, while FSMUSIC can distinguish both sources, as depicted in Fig. 4(b) and (e). When the separation angle between sources is smaller than 20°, both methods can only locate the stronger source while fail to identify the weaker one, as depicted in Fig 4(c) and (f). It should be noted that $\mathbf{V}_{\mathbf{nm}}(k,\Psi)$ in Eq. (17) contains the steering vector of all sources. Therefore, acoustic maps as depicted in Fig. (4) are not suitable for multiple sources SHMLE.

Table 2 shows the RMSE of FSPWD, FSMUSIC and SHMLE for a10 s recorded data. It can be seen that the RMSE of these three method are close. For the stronger source, the RMSE of SHMLE is smaller than that of the FSPWD and FSMUSIC, which is mainly caused by that the grid search resolution of FSPWD and FSMUSIC is 1°. The FSPWD method can only locate the stronger sources

when the separation angle between the two sources is smaller than 60°. When the separation angle between the two sources is close to 20°, the FSMUSIC can only locate the stronger one, while the SHMLE can distinguish both sources. It is clear that the spatial resolution of the SHMLE is better than that of the FSPWD and FSMUSIC.

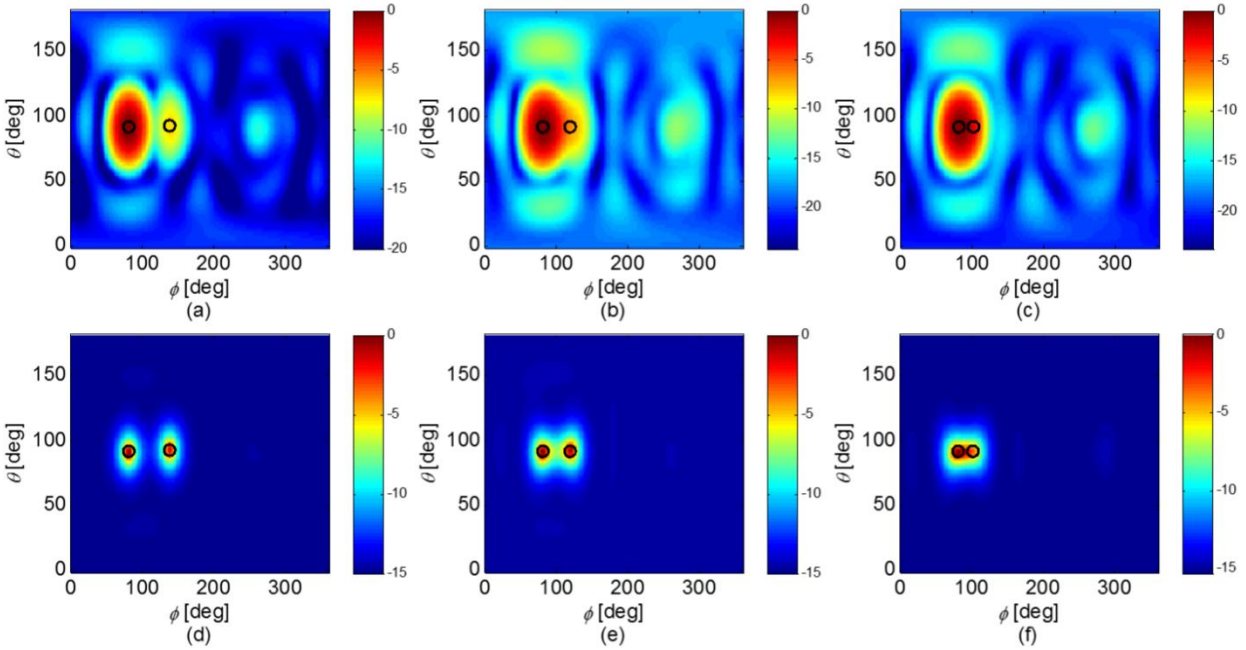

Figure 4: Wideband localization results for two sources case. (a) FSPWD, s1 and s5, (b) FSPWD, s1 and s6, (c) FSPWD, s1 and s8, (d) FSMUSIC, s1 and s5, (e) FSMUSIC, s1 and s6, (f) FSMUSIC, s1 and s8.

Table 2: RMSE of FSPWD, FSMUSIC and SHMLE for two sources case

	RMSE of FSPWD		RMSE of FSMUSIC		RMSE of SHMLE	
	strong	weak	strong	weak	strong	weak
s_1 & s_2	0.62°	1.88°	0.73°	1.19°	0.32°	1.66°
s_1 & s_3	0.59°	1.52°	0.59°	0.85°	0.32°	0.91°
s_1 & s_4	0.64°	1.03°	0.76°	0.89°	0.42°	0.90°
s_1 & s_5	0.60°	1.20°	0.77°	0.71°	0.31°	0.79°
s_1 & s_6	0.87°	-	0.83°	0.78°	0.34°	1.05°
s_1 & s_7	1.65°	-	1.08°	1.65°	0.44°	1.76°
s_1 & s_8	2.32°	-	1.66°	-	0.63°	1.54°

4. Conclusion

This paper proposes a maximum likelihood multiple source localization method in the spherical harmonic domain. The maximum likelihood strategy is used and the nonlinear optimization algorithm is applied to reduce the computational burden. The proposed method is suitable for both the open sphere and the rigid sphere array. From the simulations and experiments on a 32-microphone model, it can be seen that the proposed SHMLE method has very good spatial resolution and can distinguish two sources with less than 20° angle difference. Furthermore, the performance is stable in low SNR environment, circumventing the problem faced by the subspace-based method.

Acknowledgment

This work was supported by National Natural Science Foundation of China (Grants No. 11374156 and No. 11474163) and the Fundamental Research Funds for the Central University.

REFERENCES

1 Kumatani, K., McDonough, J. and Raj, B. Microphone array processing for distant speech recognition: From close-talking microphones to far-field sensors, *IEEE Signal Processing Magazine*, **29** (6), 127-140, (2012).

2 Khaykin, D. and Rafaely, B. Acoustic analysis by spherical microphone array processing of room impulse responses, *The Journal of the Acoustical Society of America*, **132**(1), 261-270, (2012).

3 Mabande, E., Kowalczyk, K., Sun, H. and Kellermann, W. Room geometry inference based on spherical microphone array eigenbeam processing, *The Journal of the Acoustical Society of America*, **134**(4), 2773-2789, (2013).

4 Park, M. and Rafaely, B. Sound-field analysis by plane-wave decomposition using spherical microphone array, *The Journal of the Acoustical Society of America*, **118**(5), 3094-3103, (2005).

5 Nadiri, O. and Rafaely, B. Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **22**(10), 1494-1505, (2014).

6 Sun, H., Mabande, E., Kowalczyk, K., and Kellermann, W. Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing, *The Journal of the Acoustical Society of America*, **131**(4), 2828-2840, (2012).

7 Mestre, X. and Lagunas, M. Á. Modified subspace algorithms for DoA estimation with large arrays, *IEEE Transactions on Signal Processing*, **56**(2), 598-614, (2008).

8 Hu, Y., Lu, J. and Qiu, X. A maximum likelihood direction of arrival estimation method for open-sphere microphone arrays in the spherical harmonic domain, *The Journal of the Acoustical Society of America*, **138**(2), 791-794, (2015).

9 Rafaely, B. Analysis and design of spherical microphone arrays, *IEEE Transactions on speech and audio processing*, **13**(1), 135-143, (2005).

10 Rafaely, B., *Fundamentals of spherical array processing*, Springer, (2015).

11 Chen, C. E., Lorenzelli, F., Hudson, R. E., and Yao, K. Maximum likelihood DOA estimation of multiple wideband sources in the presence of nonuniform sensor noise. *EURASIP Journal on Advances in Signal Processing*, 1-12, (2007).

12 Tervo, S. and Politis, A. Direction of arrival estimation of reflections from room impulse responses using a spherical microphone array, *IEEE/ACM Transactions on Audio, Speech and Language Processing*, **23**(10), 1539-1551, (2015).

13 Chen, J. C., Hudson, R. E. and Yao, K. Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field, *IEEE transactions on Signal Processing*, **50**(8), 1843-1854, (2002).

14 Dennis, Jr, J. E. and Mor é, J. J. Quasi-Newton methods, motivation and theory, *SIAM review*, **19**(1), 46-89, (1977).

15 Acoustics, M. H. EM32 Eigenmike microphone array release notes (v17. 0), *25 Summit Ave, Summit, NJ 07901, USA*, (2013).

16 Jarrett, D. P., Habets, E. A. P., Thomas, M. R. P., and Naylor, P. A. Rigid sphere room impulse response simulation: Algorithm and applications, *The Journal of the Acoustical Society of America*, **132**(3), 1462-1472, (2012).