

A maximum likelihood direction of arrival estimation method for open-sphere microphone arrays in the spherical harmonic domain (L)

Yuxiang Hu,^{a)} Jing Lu,^{b)} and Xiaojun Qiu^{a)}

Key Laboratory of Modern Acoustics, Institute of Acoustics, Nanjing University, Nanjing 210093, China

(Received 2 April 2015; revised 25 June 2015; accepted 4 July 2015; Published online 11 August 2015)

Open-sphere microphone arrays are preferred over rigid-sphere arrays when minimal interaction between array and the measured sound field is required. However, open-sphere arrays suffer from poor robustness at null frequencies of the spherical Bessel function. This letter proposes a maximum likelihood method for direction of arrival estimation in the spherical harmonic domain, which avoids the division of the spherical Bessel function and can be used at arbitrary frequencies. Furthermore, the method can be easily extended to wideband implementation. Simulation and experiment results demonstrate the superiority of the proposed method over the commonly used methods in open-sphere configurations. © 2015 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4926907>]

[NX]

Pages: 791–794

I. INTRODUCTION

Spherical microphone arrays have been an active area of research in recent years because of their advantage of three-dimensional symmetric configuration. Among their applications, direction of arrival (DOA) estimation is of particular importance in room impulse response analysis,¹ acoustic source localization,² and room geometry inference.^{3,4} The Cramér-Rao bound for DOA estimation in the spherical harmonic (SH) domain has been derived⁵ and various DOA estimation methods based on spherical arrays have been proposed.^{6–8} A comprehensive summary and comparison of these methods can be found in Ref. 2. Most of these methods suffer from the numerical ill-conditioning problem at nodal frequencies of the spherical Bessel function, which is usually unavoidable in open-sphere configuration. The delay-and-sum (DS) method is robust but the performance is limited by low spatial resolution. Therefore, the rigid-sphere configuration is utilized in most cases because of its robust numerical formulation.

The open-sphere configuration, where microphones are arranged on an open, or a virtual sphere, is more desirable in many circumstances when the minimum interference between the array and the sound field is requested or the array with large radius is required.^{9,10} A dual-sphere configuration, which improves spatial resolution and can be used to avoid ill-conditioning problem, was proposed and implemented in plane-wave decomposition (PWD) and room impulse response analysis.^{9,11} However, it requires twice as many microphones as that of a single sphere. Rafaely¹⁰ further investigated the robustness of the array in which microphones are arranged in the volume of a spherical shell, where high robustness can be achieved without increasing the number of the microphones but the installation and implementation of this configuration is complicated.

In this letter, a maximum likelihood DOA estimation (MLE) strategy is proposed in the SH domain to overcome the

open-sphere ill-conditioning problem,^{12,13} which can effectively solve the ill-conditioning problem with a single-layer open-sphere configuration. Furthermore, the proposed method can be extended to wideband DOA estimation directly, which is superior over the current commonly used methods where a cumbersome frequency smoothing technique has to be used.²

II. METHOD

A. Signal model in 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 a plane wave with $\Psi_s = (\theta_s, \phi_s)$ being the DOA and $s(k)$ being its amplitude, where k denotes the wave number. The Q element spherical microphone array is distributed uniformly on an open-sphere with a radius of a centered at the origin of the coordinate system, and $\Omega_q = (\theta_q, \phi_q)$ is the angle position of the q th microphone.¹⁴

The sound pressure of the q th microphone for the incident plane wave can be expressed as⁶

$$p(k, \Omega_q) = s(k)e^{i\mathbf{k}_s^T \mathbf{r}_q} \approx \sum_{n=0}^N \sum_{m=-n}^n b_n(k) Y_{n,m}^*(\Psi_s) Y_{n,m}(\Omega_q) s(k), \quad (1)$$

where $\mathbf{k}_s = -k(\cos\phi_s \sin\theta_s, \sin\phi_s \sin\theta_s, \cos\theta_s)^T$ and $\mathbf{r}_q = a(\cos\theta_q, \sin\phi_q, \cos\theta_q)^T$ 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 $(N+1)^2 \leq Q$. The superscript $(*)$ denotes complex conjugation. For open sphere, $b_n(k) = 4\pi i^n j_n(ka)$, where $i = (-1)^{1/2}$, and $j_n()$ is the spherical Bessel function of order n .⁸ Equation (1) can be expressed in matrix form as

$$p(k, \Omega_q) \approx \mathbf{y}^T(\Omega_q) \mathbf{B}(k) \mathbf{y}^*(\Psi_s) s(k), \quad (2)$$

with

^{a)}Also at School of Electrical and Computer Engineering, RMIT University, Melbourne 3000, Australia.

^{b)}Electronic mail: lujing@nju.edu.cn

$$\mathbf{y}(\Omega_q) = [Y_{0,0}(\Omega_q), Y_{1,-1}(\Omega_q), Y_{1,0}(\Omega_q), Y_{1,1}(\Omega_q), \dots, Y_{N,N}(\Omega_q)]^T, \quad (3)$$

$$\mathbf{y}(\Psi_s) = [Y_{0,0}(\Psi_s), Y_{1,-1}(\Psi_s), Y_{1,0}(\Psi_s), Y_{1,1}(\Psi_s), \dots, Y_{N,N}(\Psi_s)]^T, \quad (4)$$

$$\mathbf{B}(k) = \text{diag}\{b_0(k), b_1(k), b_1(k), b_1(k), \dots, b_N(k)\}, \quad (5)$$

where the superscript $(^T)$ denotes the transpose.

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

$$\mathbf{p}(k, \Omega) \approx \mathbf{Y}(\Omega)\mathbf{B}(k)\mathbf{y}^*(\Psi_s)s(k) + \mathbf{v}(k), \quad (6)$$

where

$$\mathbf{Y}(\Omega) = [\mathbf{y}(\Omega_1), \mathbf{y}(\Omega_2), \dots, \mathbf{y}(\Omega_Q)]^T, \quad (7)$$

$\mathbf{p}(k, \Omega) = [p(k, \Omega_1), p(k, \Omega_2), \dots, p(k, \Omega_Q)]^T$ is the vector of the sound pressure of Q [note that $(N+1)^2 \leq Q$] microphones, and $\mathbf{v}(k) = [\nu_1(k), \nu_2(k), \dots, \nu_Q(k)]^T$ 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 $\mathbf{R}_v(k) = \sigma_v^2 \mathbf{I}_Q$, where σ_v^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 holds¹⁴

$$\frac{4\pi}{Q} \mathbf{Y}^H(\Omega) \mathbf{Y}(\Omega) = \mathbf{I}_{(N+1)^2}. \quad (8)$$

The SH transform can be carried out by multiplying both sides of Eq. (6) from the left by $(4\pi/Q)\mathbf{Y}^H(\Omega)$, which yields²

$$\mathbf{p}_{\text{nm}}(k) \approx \mathbf{B}(k)\mathbf{y}^*(\Psi_s)s(k) + \mathbf{v}_{\text{nm}}(k), \quad (9)$$

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

$$\mathbf{p}_{\text{nm}}(k) = [p_{0,0}(k), p_{1,-1}(k), p_{1,0}(k), p_{1,1}(k), \dots, p_{N,N}(k)]^T. \quad (10)$$

The second term on the right side of Eq. (9) is the noise expressed in the SH domain, i.e., $\mathbf{v}_{\text{nm}}(k) = (4\pi/Q)\mathbf{Y}^H(\Omega)\mathbf{v}(k)$, with the mean

$$E[\mathbf{v}_{\text{nm}}(k)] = \frac{4\pi}{Q} \mathbf{Y}^H(\Omega) E[\mathbf{v}(k)] = \mathbf{0} \quad (11)$$

and the covariance matrix

$$\begin{aligned} \mathbf{R}_{\text{nm}}(k) &= E \left[\frac{4\pi}{Q} \mathbf{Y}^H(\Omega) \mathbf{v}(k) \mathbf{v}^H(k) \mathbf{Y}(\Omega) \frac{4\pi}{Q} \right] \\ &= \frac{4\pi}{Q} \sigma_v^2 \mathbf{I}_{(N+1)^2}, \end{aligned} \quad (12)$$

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

B. Sound source localization in spherical harmonic domain

In this section, the wideband and narrowband MLE will be derived in the SH domain. Define $\Theta = [\Psi_s, \mathbf{s}^T, \sigma_v^2]^T$ as the vector of all the unknown parameters, where $\mathbf{s} = [s(k_{\min}), \dots, s(k_{\max})]^T$ contains the amplitudes of the source signals with k_{\min} and k_{\max} representing the minimum and maximum wave numbers and satisfying $ka \leq N$. Throughout this letter, Ψ_s , \mathbf{s} , and σ_v^2 are assumed to be deterministic and unknown, while the observed data \mathbf{p}_{nm} are considered random.^{13,15} The likelihood function of Θ in the SH domain can be expressed as,^{12,13}

$$f(\mathbf{p}_{\text{nm}}; \Theta) = \frac{\exp \left\{ - \sum_{k=k_{\min}}^{k_{\max}} [\mathbf{p}_{\text{nm}}(k) - \mathbf{d}_{\text{nm}}(k, \Psi)s(k)]^H \mathbf{R}_{\text{nm}}^{-1} [\mathbf{p}_{\text{nm}}(k) - \mathbf{d}_{\text{nm}}(k, \Psi)s(k)] \right\}}{\left(\pi^{(N+1)^2} |\mathbf{R}_{\text{nm}}| \right)^{k_{\max} - k_{\min}}}, \quad (13)$$

where $\mathbf{d}_{\text{nm}}(k, \Psi) = \mathbf{B}(k)\mathbf{y}^*(\Psi)$ and $|\cdot|$ denotes the matrix determinant. The log-likelihood function $L(\mathbf{p}_{\text{nm}}; \Theta)$, considering (12) and neglecting all the constant terms, is given by

$$\begin{aligned} L(\mathbf{p}_{\text{nm}}; \Theta) &= -(N+1)^2(k_{\max} - k_{\min}) \log \sigma_v^2 \\ &\quad - \frac{Q}{4\pi\sigma_v^2} \sum_{k=k_{\min}}^{k_{\max}} \|\mathbf{p}_{\text{nm}}(k) - \mathbf{d}_{\text{nm}}(k, \Psi)s(k)\|^2, \end{aligned} \quad (14)$$

where $\|\cdot\|$ denotes the Euclidean norm.

Fixing Ψ and \mathbf{s} , and then maximizing (14) with respect to σ_v^2 yields

$$\begin{aligned} \hat{\sigma}_v^2 &= \frac{Q}{4\pi(N+1)^2(k_{\max} - k_{\min})} \\ &\quad \times \sum_{k=k_{\min}}^{k_{\max}} \|\mathbf{p}_{\text{nm}}(k) - \mathbf{d}_{\text{nm}}(k, \Psi)s(k)\|^2. \end{aligned} \quad (15)$$

Substituting Eqs. (9) and (15) into Eq. (14), neglecting the constant terms and considering the monotonicity of the logarithm function, the MLE for Ψ_s and \mathbf{s} can be obtained as

$$(\hat{\Psi}_s, \hat{\mathbf{s}}) = \arg \min_{\Psi_s, \mathbf{s}} \sum_{k=k_{\min}}^{k_{\max}} \|\mathbf{p}_{\text{nm}}(k) - \mathbf{d}_{\text{nm}}(k, \Psi)s(k)\|^2. \quad (16)$$

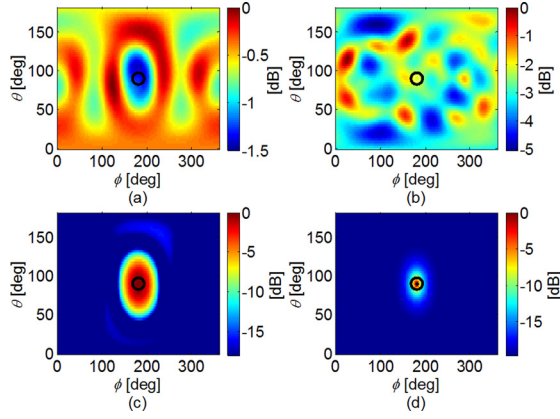


FIG. 1. (Color online) Narrowband DOA estimation results ($N=4$, $a=13.9$ cm, $Q=32$) of (a) PWD, (b) MUSIC, (c) DS, and (d) SHMLE at $ka=3.142$, where the value of $b_0(k)$ is close to zero.

It is interesting to note that Eq. (16) is a standard least squares (LS) criterion, which is widely used in beamformer design¹⁶ and sound field control.¹⁷ The equivalence between MLE and LS under Gaussian noise model assumption is well known.¹⁵

Once Ψ is fixed, the estimator for \mathbf{s} in each frequency bin is simply the LS solution

$$\hat{\mathbf{s}}(k) = \mathbf{d}_{\text{nm}}(k, \Psi)^\dagger \mathbf{p}_{\text{nm}}(k), \quad (17)$$

where $(\cdot)^\dagger$ denotes pseudo-inverse operation. Substituting Eq. (17) back into Eq. (16) leads to the estimator for Ψ_s :

$$\hat{\Psi}_s = \arg \min_{\Psi} \sum_{k=k_{\min}}^{k_{\max}} \|\mathbf{p}_{\text{nm}}(k) - \mathbf{d}_{\text{nm}}(k, \Psi) \mathbf{d}_{\text{nm}}(k, \Psi)^\dagger \mathbf{p}_{\text{nm}}(k)\|^2. \quad (18)$$

The estimator can also be described as

$$\hat{\Psi}_s = \arg \max_{\Psi} \left\{ -10 \log_{10} \left(\sum_{k=k_{\min}}^{k_{\max}} \|\mathbf{p}_{\text{nm}}(k) - \mathbf{d}_{\text{nm}}(k, \Psi) \mathbf{d}_{\text{nm}}(k, \Psi)^\dagger \mathbf{p}_{\text{nm}}(k)\|^2 \right) \right\}. \quad (19)$$

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

$$\hat{\Psi}_s = \arg \max_{\Psi} (-20 \log_{10} \|\mathbf{p}_{\text{nm}}(k) - \mathbf{d}_{\text{nm}}(k, \Psi) \mathbf{d}_{\text{nm}}(k, \Psi)^\dagger \mathbf{p}_{\text{nm}}(k)\|). \quad (20)$$

Although it is simple, the narrowband solution suffers from performance degradation in the presence of coherent noise,

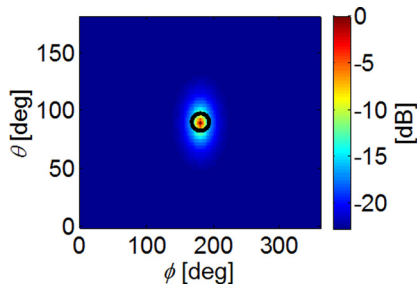


FIG. 2. (Color online) Wideband DOA estimation result ($N=4$, $a=13.9$ cm, $Q=32$) of SHMLE with frequency range $ka \in [3, 4]$.

e.g., the early reflections of the source signal.² Fortunately, the MLE in the SH domain has the remarkable benefit of easy implementation of wideband DOA as described in Eqs. (18) and (19). This is superior over the existing methods, which usually require a quite cumbersome frequency smoothing (FS) technique to realize wideband DOA.²

As pointed out in many references,^{9,10} $b_n(k)$ for open-sphere is very low around zeros of the spherical Bessel function, which leads to the ill-conditioning problem when algorithms such as PWD and multiple signal classification (MUSIC) are used. This problem can be avoided by choosing a proper k for narrowband DOA estimation. However, when these methods (including the DS method) are extended to wideband situation using FS technique, the numerical ill-conditioning problem cannot be avoided, which limits the implementation of the conventional DOA estimation methods for open-sphere arrays.^{2,6,8} On the other hand, the division of $b_n(k)$ is unnecessary as shown in Eqs. (18)–(20), which makes the method proposed in this paper a better choice for open-sphere arrays. Furthermore, the single-sphere configuration is much easier to be implemented compared with the dual-sphere or the spherical-shell microphone arrays.^{9,10}

III. SIMULATIONS AND EXPERIMENTS

The proposed method in this paper is named as spherical harmonic MLE (SHMLE) and compared to the commonly used PWD,⁶ MUSIC,⁸ and DS⁷ methods in the SH domain. An open spherical array of order $N=4$ and radius $a=13.9$ cm, with $Q=32$ microphones arranged in a uniform scheme, was used for all simulations. The sound source is placed at $(\theta_s, \phi_s) = (90^\circ, 180^\circ)$. 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 signal to noise ratio is set at 20 dB. It should be noted that the localization of multiple source using MLE requires a very high computational burden. However, for the single source scenario in this letter, the simple grid search method can be used with the grid resolution of 1° . The distribution of the cost function as shown in Eq. (20) can be plotted to form an acoustic map in which the peak determines the estimated DOA.

Figure 1 depicts the acoustic maps for narrowband PWD, MUSIC, DS and SHMLE methods at $ka=3.142$, where the value of $b_0(k)$ is close to zero. The DOA of the real sound source is denoted by a solid black circle in all these figures. It can be observed that the peaks in the acoustic maps of



FIG. 3. (Color online) Eight-elements open-sphere microphone array.

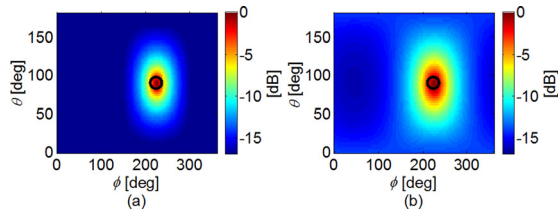


FIG. 4. (Color online) DOA estimation experiment results ($N=1$, $a=13.9$ cm, $Q=8$) of (a) narrowband ($ka=1$) and (b) wideband ($ka \in [0.5\ 1]$) SHMLE method using an 8-elements open-sphere microphone array in the anechoic chamber.

PWD and MUSIC methods failed to match the direction of the sound source because of the null frequency of the spherical Bessel function, while the DS and SHMLE methods can overcome the ill-conditioning problem and achieve correct DOA estimation. It can also be seen that the SHMLE method has significantly higher spatial resolution than the DS method. The localization result for the wideband SHMLE with frequency range $ka \in [3\ 4]$ is depicted in Fig. 2. Similar to the narrowband case, the peaks in the wideband SHMLE acoustic map clearly matches the DOA of the sound source.

An open spherical array of order $N=1$ and radius $a=13.9$ cm, with $Q=8$ microphones arranged uniformly, as depicted in Fig. 3, was used for all experiments, while other conditions are the same as those in the simulations. Figure 4 depicts the acoustic maps in the anechoic chamber for the narrowband and wideband SHMLE. The acoustical signal was played via a loudspeaker, which was located at $(89^\circ, 225^\circ)$ relative to the array. The narrowband frequency corresponds to $ka=1$, and the wideband frequency range is $ka \in [0.5\ 1]$. The experiment results in the anechoic chamber show that the proposed SHMLE method has quite high spatial resolution and localization accuracy. The root-mean-square error (RMSE) of the wideband SHMLE for a 10 s recorded white noise data is 0.58° .

Figure 5 depicts the acoustic maps in the listening room for narrowband and wideband SHMLE. Sound source is located at $(91^\circ, 225^\circ)$ relative to the array. It can be seen that the narrowband SHMLE has higher spatial resolution than the wideband SHMLE, but its estimation result apparently deviates from the true DOA because of the coherent noise caused by reverberation, while the wideband SHMLE result shows superior accuracy. The RMSE of the wideband SHMLE for a 10 s recorded white noise data is 4.43° .

It should be noted that the performance of the MLE method in the frequency domain¹³ is similar to that of the SHMLE method. As most algorithms in the SH domain,⁸ the SHMLE method has the advantage of comparatively lower computational complexity and can be used for various configurations, such as open sphere, rigid sphere and an array with cardioid microphones.

IV. CONCLUSIONS

This letter extends the maximum likelihood method for DOA estimation in the SH domain. It is proved that the proposed method can effectively overcome the ill-conditioning problem at nodal frequencies of the spherical Bessel

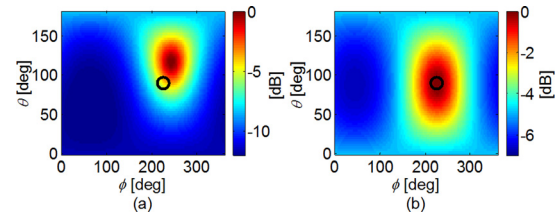


FIG. 5. (Color online) DOA estimation experiment results ($N=1$, $a=13.9$ cm, $Q=8$) of (a) narrowband ($ka=1$) and (b) wideband ($ka \in [0.5\ 1]$) SHMLE method using an 8-elements open-sphere microphone array in the listening room.

function. Therefore, it is a reasonable choice for DOA estimation with open-sphere configuration. Furthermore, it is straightforward to implement the proposed method for wideband signal. The benefit and efficacy of the proposed method is demonstrated by simulations and experiments.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation of China Grant No. 11374156.

- ¹D. Khaykin and B. Rafaely, "Acoustic analysis by spherical microphone array processing of room impulse responses," *J. Acoust. Soc. Am.* **132**, 261–270 (2012).
- ²H. Sun, E. Mabande, K. Kowalczyk, and W. Kellermann, "Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing," *J. Acoust. Soc. Am.* **131**, 2828–2840 (2012).
- ³E. Mabande, K. Kowalczyk, H. Sun, and W. Kellermann, "Room geometry inference based on spherical microphone array eigenbeam processing," *J. Acoust. Soc. Am.* **134**, 2773–2789 (2013).
- ⁴S. Tervo and T. Tossavainen, "3D room geometry estimation from measured impulse responses," in *Proceedings of the ICASSP* (2012), pp. 513–516.
- ⁵L. Kumar and R. M. Hegde, "Stochastic Cramér-Rao bound analysis for DOA estimation in spherical harmonics domain," *IEEE Signal Process. Lett.* **22**, 1030–1034 (2015).
- ⁶M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," *J. Acoust. Soc. Am.* **118**, 3094–3103 (2005).
- ⁷S. Yan, H. Sun, U. P. Svensson, X. Ma, and J. M. Hovem, "Optimal modal beamforming for spherical microphone arrays," *IEEE Trans. Audio Speech Lang. Process.* **19**, 361–371 (2011).
- ⁸B. Rafaely, Y. Peled, M. Agmon, D. Khaykin, and E. Fisher, "Spherical microphone array beamforming," in *Speech Processing in Modern Communication* (Springer, Berlin, 2010), pp. 281–305.
- ⁹I. Balmages and B. Rafaely, "Open-sphere designs for spherical microphone arrays," *IEEE Trans. Audio, Speech, Lang. Process.* **15**, 727–732 (2007).
- ¹⁰B. Rafaely, "The spherical-shell microphone array," *IEEE Trans. Audio Speech Lang. Process.* **16**, 740–747 (2008).
- ¹¹B. Rafaely, I. Balmages, and L. Eger, "High-resolution plane-wave decomposition in an auditorium using a dual-radius scanning spherical microphone array," *J. Acoust. Soc. Am.* **122**, 2661–2668 (2007).
- ¹²I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoust. Speech Signal Process.* **36**, 1553–1560 (1988).
- ¹³C. E. Chen, F. Lorenzelli, R. E. Hudson, and K. Yao, "Maximum likelihood DOA estimation of multiple wideband sources in the presence of nonuniform sensor noise," *EURASIP J. Adv. Signal Process.* **2008**, 835079.
- ¹⁴B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Speech Audio Process.* **13**, 135–143 (2005).
- ¹⁵S. Godsill, "Bayesian computational methods in signal processing," in *Academic Press Library in Signal Processing: Array and Statistical Signal Processing*, edited by R. Chellappa and S. Theodoridis (Academic Press, Amsterdam, 2014), Chap. 4, pp. 143–185.
- ¹⁶S. Doclo and M. Moonen, "Design of far-field and near-field broadband beamformers using eigenfilters," *Signal Process.* **83**, 2641–2673 (2003).
- ¹⁷O. Kirkeby and P. A. Nelson, "Reproduction of plane wave sound fields," *J. Acoust. Soc. Am.* **94**, 2992–3000 (1993).