INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Cramér-Rao Lower Bound for DOA Estimation with an Array of Directional Microphones in Reverberant Environments

Weiguang Chen, Cheng Xue, Xionghu Zhong

College of Computer Science and Electronic Engineering, Hunan University, China
{cwg, chengxue, and xzhong}@hnu.edu.cn

Abstract

Existing direction-of-arrival (DOA) estimation methods usually assume that signals are received by an array of omnidirectional microphones. The performance can be seriously degraded due to heavy reverberation and noise. In this paper, DOA estimation using an array with directional microphones is considered. As the signal response varies over different DOAs, the magnitude information as well as the phase information can be employed to estimate the DOA. We first introduce the spherically isotropic noise field using directional microphones. The Cramér-Rao Lower Bound (CRLB) for DOA estimation is then derived and compared with that using omnidirectional microphones under different signal-to-reverberation ratio (SRR) environments. In addition, we extend existing steered response power (SRP), minimum variance distortionless response (MVDR) and multiple signal classification (MUSIC) estimators for the DOA estimation using directional microphone arrays. Both CRLB Analysis and DOA estimation show that better DOA estimation performance can be achieved by using a directional microphone array.

Index Terms: DOA estimation, directional microphone, microphone array, Cramér-Rao lower bound, reverberation.

1. Introduction

Direction-of-arrival (DOA) estimation is essentially important in microphone array signal processing and has a broad range of applications such as acoustic source localization and tracking [1, 2, 3, 4], beamforming [5], source separation [6] and camera steering [7]. Traditional estimation methods including SRP [8], MVDR [9] and MUSIC [10] algorithms overwhelmingly rely on the assumption that an array of omnidirectional microphones are employed. DOA estimation using a directional microphone array is rarely studied. In addition, it is well known that the performance of DOA estimation can be significantly degraded when reverberation and noise are present.

In order to reduce the DOA estimation error in the presence of reverberation and noise, various algorithms have been developed in the past [11, 12, 13, 14]. The phase transform (PHAT) weighting was proposed to improve the robustness of SRP method, known as SRP-PHAT algorithm [15]. The direct path dominance (DPD) test [16] is developed to identify the time-frequency (TF) bins which are dominated by the direct path and a DOA estimation method is proposed based on the selected TF bins, resulting in a considerable performance improvement according to the study on a spherical array. In [11], the DPD test approach is extended to arbitrary arrays for source localization. Further, a modified weighted prediction error method is introduced to reduce the performance degradation due to reverberation in the middle frequency range [12]. In [14], the DOAs of both direct path and early reflections are estimated. Two iterative nonlinear least squares (NLS) based localization methods are used so that the early reflections can be subtracted from the received signals. In [13], reverberation is modeled as a diffusive field and an expectation-maximization (EM) algorithm is employed to cluster the TF bins due to an acoustic source and estimate the DOA. Although better performance can be achieved, the EM based method is computationally expensive. It is worth mentioning that all these methods are developed for the omnidirectional microphone arrays.

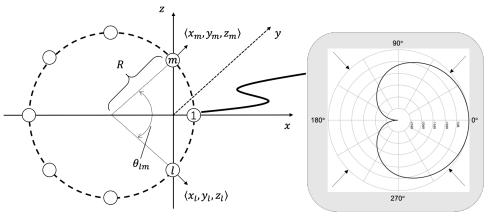


Figure 1: *A uniform circular array with cardioid microphones.*

Recently, directional microphones have been studied for DOA estimation and shown advantages over omnidirectional microphones. Directional microphones have different magnitude response over various DOAs, as shown in Fig. 1. Hence, the magnitude information can also be used in DOA estimation. In [17], a DOA estimator is proposed to alleviate the spatial aliasing using directional microphones. A uniform circular array with directional microphones is also used in the design of differential beamforming and it is shown that better performance can be achieved compared to omnidirectional microphone arrays [18]. In this paper, we focus on the DOA estimation using directional microphone arrays under reverberant environments. The reverberation is modelled as a diffusive noise field. The cross power spectral density (PSD) matrix of the received signals is constructed to derive a closed-form CRLB expression through the inverse of the Fisher information matrix (FIM). The CRLBs of omnidirectional and directional microphone arrays are analyzed under different simulated experiments. We also extend three conventional DOA estimators, i.e., SRP, MVDR and MUSIC, to the directional microphone arrays, and conducted experiments to compare the DOA estimation performance. To better understand the behavior of directional microphone array in different reverberant environments, we also evaluate the spatial spectrum pattern in addition to the root mean square error (RMSE) of DOA estimates.

2. Problem Formulation

2.1. Signal model

Assume that a single source in the far-field and an array with M directional microphones are considered. The received signal in short time Fourier transform (STFT) domain can be addressed

http://dx.doi.org/10.21437/Interspeech.2021-2267

as

$$\mathbf{y}(t,k) = \mathbf{a}(\phi,k)s(t,k) + \mathbf{r}(t,k) + \mathbf{n}(t,k), \quad (1)$$

where $\mathbf{y}(t,k)$ denotes the received signal in the t-th frame and k-th frequency bin, and $\mathbf{a}(\phi,k)$ is the steering vector on DOA ϕ, and $s(t,k)$ denotes the source signal, and $\mathbf{r}(t,k)$ denotes the reverberant signal and $\mathbf{n}(t,k)$ is the additive noise. To simplify the analysis, the source signal is modeled as a zero-mean Gaussian Process for the CRLB analysis, given by

$$s(t,k) \sim \mathcal{N}(0, \eta_s(t,k)), \quad (2)$$

where $\eta_s(t,k)$ is the power spectral density (PSD) of the source signal. The reverberation and the noise component are assumed to be uncorrelated and modelled by zero-mean multivariate Gaussian process, given as

$$\mathbf{r}(t,k) \sim \mathcal{N}(0, \eta_r(t,k)\mathbf{\Gamma}(k)), \quad (3)$$

$$\mathbf{n}(t,k) \sim \mathcal{N}(0, \eta_n(t,k)\mathbf{I}), \quad (4)$$

where $\eta_r(t,k)$ and $\eta_n(t,k)$ are the variance denoting the power of the reverberation and noise respectively, and $\mathbf{I}$ is an $M \times M$ identity matrix, and $\mathbf{\Gamma}(k)$ denotes the normalized PSD matrix of the reverberation. Note that $\eta_r(t,k)$ denotes the power of reverberation while $\mathbf{\Gamma}(k)$ captures the spatial coherence.

2.2. Isotropic noise modeling

Assuming that the sound field of reverberation are homogeneous and isotropic, the matrix $\mathbf{\Gamma}(k)$ is the average cross spectral density for all spherical directions [19, 20], defined as

$$\mathbf{\Gamma}(k) = \frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi} \mathbf{a}(\omega,\theta,\phi)\mathbf{a}^H(\omega,\theta,\phi)\sin\theta \mathrm{d}\theta\mathrm{d}\phi, \quad (5)$$

where $\omega = 2\pi f k/K$ is the angular frequency with K denoting the number of frequency bins. Considering a circular array as shown in Fig. 1, the steering vector $\mathbf{a}(\omega,\theta,\phi)$ is given by

$$\mathbf{a}(\omega,\theta,\phi) = [a_1, a_2, \cdots, a_m, \cdots, a_M]^T, \quad (6)$$

where $\theta \in [0,\pi]$ and $\phi \in [0, 2\pi)$ are the elevation and azimuth angles respectively. Each element in (6) is given as

$$a_m = T_m(\theta,\phi)e^{j\frac{\omega R}{c}\cos\psi_m}, \quad (7)$$

where $T_m(\theta,\phi)$ is the directional pattern for the m-th first order directional microphone and ψ_m is the angle between the incident wave and the axis of the directional microphone which can be expressed as [21]

$$T_m(\theta,\phi) = \alpha_m + (1-\alpha_m)\cos\psi_m, \quad (8)$$

$$\cos\psi_m = x_m\cos\phi\sin\theta + y_m\sin\phi\sin\theta + z_m\cos\theta, \quad (9)$$

$$x_m = \cos\phi_m\sin\theta_m, \quad (10)$$

$$y_m = \sin\phi_m\sin\theta_m, \quad (11)$$

$$z_m = \cos\theta_m, \quad (12)$$

where $\alpha_m \in [0,1]$ defines the directivity of the m-th directional microphone and (x_m, y_m, z_m) denotes the unit orientation vector of the m-th directional microphone. Fig. 1 gives an illustration of cardioid microphones with $\alpha_m = 0.5$. It can be observed that magnitude response is evenly distributed on the circle and the microphone has the maximum magnitude response at 0° (along positive x-axis) and minimum response at 180° (along negative x-axis).

For the first order differential microphones, the cross spectral density for an isotropic field is formulated as [19]

$$\begin{aligned}\mathbf{\Gamma}_{lm}(k) &= \frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi} T_l(\theta,\phi)T_m(\theta,\phi)e^{j\frac{\omega d_{lm}}{c}\cos\theta}\sin\theta \mathrm{d}\theta\mathrm{d}\phi \\ &= \frac{\alpha_l\alpha_m\sin(gd_{lm})}{gd_{lm}} + [\sin(gd_{lm}) - gd_{lm}\cos(gd_{lm})] \\ &\times \frac{(1-\alpha_l)(1-\alpha_m)(x_lx_m + y_ly_m)}{(gd_{lm})^3} \\ &+ \frac{(1-\alpha_l)(1-\alpha_m)z_lz_m}{(gd_{lm})^3}\left[(gd_{lm})^2\sin(gd_{lm})\right. \\ &+2(gd_{lm})\cos(gd_{lm}) - 2\sin(gd_{lm})] \\ &- j\left[gd_{lm}\sin(gd_{lm}) - (gd_{lm})^2\cos(gd_{lm})\right] \\ &\times \left[\frac{(1-\alpha_l)\alpha_mz_l}{(gd_{lm})^3} + \frac{(1-\alpha_m)\alpha_lz_m}{(gd_{lm})^3}\right], \end{aligned} \quad (13)$$

where $g = \omega/c$ and $l, m \in \{1, 2, \cdots, M\}$ are the indexes of microphones. However, it assumes that the array is linear and the microphones are located along the z-axis. According to (5), the cross spectral density for circular array with directional microphones is computed as

$$\begin{aligned}\mathbf{\Gamma}_{lm}(k) = \frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi} &T_l(\theta,\phi)T_m(\theta,\phi) \\ &e^{j\frac{\omega R}{c}(\cos\psi_l - \cos\psi_m)}\sin\theta \mathrm{d}\theta\mathrm{d}\phi, \end{aligned} \quad (14)$$

where R denotes the radius. Instead of using a more complex equation (14) to calculate the PSD matrix of circular arrays, we can compute it by the formula in (13). Let us consider a circular array in the x-z plane as shown in Fig. 1. Suppose that the l-th and m-th microphones are located at z-axis and θ_{lm} denotes the angle between them. Then we can easily apply (13) to the circular microphone array. Note that $y_m = y_l = 0$, $x_m = x_l = \cos(\theta_{lm}/2)$ and $z_m = -z_l = \sin(\theta_{lm}/2)$. Given the same directivity parameter α for all microphones, the cross spectral density between the l-th and m-th microphones can be simplified as

$$\begin{aligned}\mathbf{\Gamma}_{lm}(k) = &\frac{\alpha^2\sin(gd_{lm})}{gd_{lm}} + \frac{(1-\alpha)^2x_lx_m}{(gd_{lm})^3} \\ &\times [\sin(gd_{lm}) - gd_{lm}\cos(gd_{lm})] \\ &+ \frac{(1-\alpha)^2z_lz_m}{(gd_{lm})^3}\left[(gd_{lm})^2\sin(gd_{lm})\right. \\ &\left.+2(gd_{lm})\cos(gd_{lm}) - 2\sin(gd_{lm})\right]. \end{aligned} \quad (15)$$

3. Derivation of the CRLB

The CRLB provides a lower bound on the covariance matrix of any unbiased estimator. The CRLB for DOA estimation has been well studied with omni-directional microphones [22, 23]. However, investigation with directional microphones in reverberant environments remains untapped. Considering the signal model in (1), the parameter vector is defined as

$$\mathbf{\Omega} = [\eta_s, \eta_r, \eta_n, \phi]^T. \quad (16)$$

Substituting (2), (3) and (4) into (1), the received signal is modeled as

$$\mathbf{y}(t,k) \sim \mathcal{N}(0, \mathbf{\Phi}_\mathbf{y}) \quad (17)$$

where

$$\mathbf{\Phi}_\mathbf{y} = \eta_s\mathbf{a}(\phi,k)\mathbf{a}^H(\phi,k) + \eta_r\mathbf{\Gamma}(k) + \eta_n\mathbf{I}. \quad (18)$$

The CRLB can be obtained from the inverse of the FIM. Given (17), the (l, m)-th element of the FIM is given by

$$[\mathbf{FIM}]_{lm} = N\mathbf{Tr}\left[\mathbf{\Phi_y^{-1}}\frac{\partial \mathbf{\Phi_y}}{\partial \mathbf{\Omega}_l}\mathbf{\Phi_y^{-1}}\frac{\partial \mathbf{\Phi_y}}{\partial \mathbf{\Omega}_m}\right], \tag{19}$$

where N is the number of snapshots, and $\mathbf{\Omega}_l, \mathbf{\Omega}_m \in \mathbf{\Omega}$, and $\mathbf{Tr}[\cdot]$ is the trace of the matrix. By using the matrix inversion Lemma, $\mathbf{\Phi}_y^{-1}$ is calculated as

$$\mathbf{\Phi_y^{-1}} = \mathbf{\Psi}^{-1} - \mathbf{\Psi}^{-1}\mathbf{a}(\phi,k)\mathbf{a}^H(\phi,k)\mathbf{\Psi}^{-1} \times \frac{1}{\eta_s^{-1} + \mathbf{a}^H(\phi,k)\mathbf{\Psi}^{-1}\mathbf{a}(\phi,k)}, \tag{20}$$

where $\mathbf{\Psi} = \eta_r\mathbf{\Gamma}(k) + \eta_n\mathbf{I}$. After deriving the partial derivative $\frac{\partial \mathbf{\Phi_y}}{\partial \mathbf{\Omega}_m}$ in (19) for each parameter, the expressions for the FIM blocks are summarized as:

$$[\mathbf{FIM}]_{ss} = N\mathbf{Tr}\left[\mathbf{\Phi_y^{-1}}\mathbf{a}\mathbf{a}^H\mathbf{\Phi_y^{-1}}\mathbf{a}\mathbf{a}^H\right], \tag{21}$$

$$[\mathbf{FIM}]_{rr} = N\mathbf{Tr}\left[\mathbf{\Phi_y^{-1}}\mathbf{\Gamma}\mathbf{\Phi_y^{-1}}\mathbf{\Gamma}\right], \tag{22}$$

$$[\mathbf{FIM}]_{nn} = N\mathbf{Tr}\left[\mathbf{\Phi_y^{-1}}\mathbf{I}\mathbf{\Phi_y^{-1}}\mathbf{I}\right], \tag{23}$$

$$[\mathbf{FIM}]_{\theta\theta} = N\mathbf{Tr}\left[\mathbf{\Phi_y^{-1}}\frac{\partial \mathbf{a}\mathbf{a}^H}{\partial\theta}\mathbf{\Phi_y^{-1}}\frac{\partial \mathbf{a}\mathbf{a}^H}{\partial\theta}\right], \tag{24}$$

$$[\mathbf{FIM}]_{rs} = [\mathbf{FIM}]_{sr} = N\mathbf{Tr}\left[\mathbf{\Phi_y^{-1}}\mathbf{\Gamma}\mathbf{\Phi_y^{-1}}\mathbf{a}\mathbf{a}^H\right], \tag{25}$$

$$[\mathbf{FIM}]_{s\theta} = [\mathbf{FIM}]_{\theta s} = N\mathbf{Tr}\left[\mathbf{\Phi_y^{-1}}\mathbf{a}\mathbf{a}^H\mathbf{\Phi_y^{-1}}\frac{\partial \mathbf{a}\mathbf{a}^H}{\partial\theta}\right], \tag{26}$$

$$[\mathbf{FIM}]_{sn} = [\mathbf{FIM}]_{ns} = N\mathbf{Tr}\left[\mathbf{\Phi_y^{-1}}\mathbf{a}\mathbf{a}^H\mathbf{\Phi_y^{-1}}\mathbf{I}\right], \tag{27}$$

$$[\mathbf{FIM}]_{r\theta} = [\mathbf{FIM}]_{\theta r} = N\mathbf{Tr}\left[\mathbf{\Phi_y^{-1}}\mathbf{\Gamma}\mathbf{\Phi_y^{-1}}\frac{\partial \mathbf{a}\mathbf{a}^H}{\partial\theta}\right], \tag{28}$$

$$[\mathbf{FIM}]_{rn} = [\mathbf{FIM}]_{nr} = N\mathbf{Tr}\left[\mathbf{\Phi_y^{-1}}\mathbf{\Gamma}\mathbf{\Phi_y^{-1}}\mathbf{I}\right], \tag{29}$$

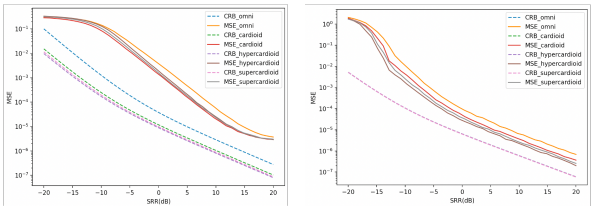
$$[\mathbf{FIM}]_{n\theta} = [\mathbf{FIM}]_{\theta n} = N\mathbf{Tr}\left[\mathbf{\Phi_y^{-1}}\mathbf{\Phi_y^{-1}}\frac{\partial \mathbf{a}\mathbf{a}^H}{\partial\theta}\right]. \tag{30}$$

Note that only the CRLB for ϕ is of interest in this paper. The CRLB corresponding to θ is

$$\text{CRLB}_\theta = \mathbf{FIM}_{\theta\theta}^{-1}. \tag{31}$$

4. DOA Estimation

In addition to the theoretical analysis on CRLB, we also extend three widely used DOA estimation methods, i.e., SRP, MVDR and MUSIC, to directional microphone arrays. The spatial spectra of these methods are respectively given as 1) SRP: $P_\theta(\omega) = \sum_k \left(\sum_l \sum_m W_{lm}\left(a_l\mathbf{y}_l(t,k)\right)\left(a_m\mathbf{y}_m(t,k)\right)^*\right)$ where $(\cdot)^*$ denotes complex conjugation, and W_{lm} is a weighting function given by $1/|\mathbf{y}_l(t,k)\mathbf{y}_m(t,k)^*|$ when phase transform is applied; 2) MVDR: $P_\theta(\omega) = 1/(\mathbf{a}^H\mathrm{R}_{yy}^{-1}\mathbf{a})$ where R_{yy} is the correlation matrix of observed signals and $(\cdot)^H$ denotes Hermitian transposition; and 3) MUSIC: $P_\theta(\omega) = 1/(\mathbf{a}^H\mathrm{U}_N\mathrm{U}_N^H\mathbf{a})$ where U_N is the matrix of eigenvectors that span noise subspace. After constructing the spatial spectra, these methods estimate the source DOA by identify the largest peak in the spectrum. Note that with directional microphone arrays, the magnitude information is also modeled in the steering vector. In order to study the DOA estimation performance comprehensively, four different metrics are employed.



(a) Linear array (b) Circular array

Figure 2: *CRLB and MSE of omnidirectional and directional microphone as function of SRR.*

Root mean square error (RMSE) is used to measure the accuracy of DOA estimators and calculated as

$$\text{RMSE}_\theta = \sqrt{\frac{1}{P}\sum_{l=1}^{P}\left(\hat{\theta}_p - \theta_{p,truth}\right)^2}, \tag{32}$$

where P is the number of simulation runs and $\theta_{p,truth}$ denotes the ground truth direction, and $\hat{\theta}_p$ denotes the estimated DOA of the p-th simulation run.

Sidelobe uncertainty (SU) measures the difference between main lobe and the largest sidelobe, defined as [24]

$$\Delta_{\text{dB}} = 10\log_{10}\left[1 + 10^{\left(\frac{ML-SL}{10}\right)}\right] \tag{33}$$

where ML and SL denote main lobe and sidelobe level in dB respectively.

Kurtosis (KU) is used to measure the tails of a distribution, defined as $g_2 = \mu_4/\delta^4 - 3$ where μ_4 is the fourth central moment and δ is the standard deviation [25]. In this paper, we regard the spatial spectra around the main lobe (i.e. $[\theta_{\text{main}} - 30^\circ, \theta_{\text{main}} + 30^\circ]$) as a distribution and the kurtosis on the left and right sides of main lobe are averaged. The higher the kurtosis is, the more obvious the peak presents.

Directivity index (DI) is defined as the ratio between the power of the desired direction and the power averaged over all directions [26]. It is usually used to measure the quality of a beamformer. We introduce it here to measure the spatial spectra generated by DOA estimators, computed as

$$\text{DI} = 10\log_{10}\left(\frac{\frac{1}{2\theta_{th}}\int_{\theta_s-\theta_{th}}^{\theta_s+\theta_{th}} S(\theta)d\theta}{\frac{1}{2\pi}\int_0^{2\pi} S(\theta)d\theta}\right), \tag{34}$$

where $S(\theta)$ is the spatial spectra, θ denotes the azimuth angle, θ_s is the desired direction and $\theta_{th} = 10^\circ$.

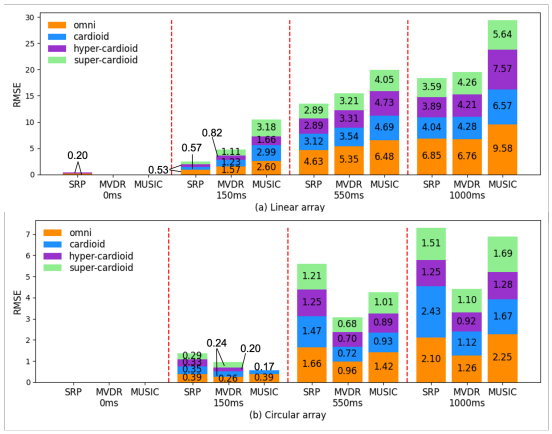


Figure 3: *RMSE comparison with different type microphones.*

5. Performance evaluation

We consider an 8 microphone circular array with a radius of 8cm and a 4 microphone linear array with a spacing of 8cm. The directional microphones point outwards along the radius in the circular array and toward broadside in the linear array. For the circular array, the range of the target DOA is $[0^\circ, 360^\circ]$, and for the linear array, the range is $[-60^\circ, 60^\circ]$. Three first-order microphones, i.e., cardioid ($\alpha = 0.5$), hyper-cardioid ($\alpha = 0.33$) and super-cardioid ($\alpha = 0.41$), are used in both CRLB studies and simulated experiments.

CRLB analysis. In our first experiment, the CRLBs of different microphones types are studied. The SRR = $10\log_{10}(\frac{\eta_s}{\eta_r})$ varies from -20 dB to 20dB while other parameters are fixed. For each SRR, 200 Monte Carlo (MC) experiments are performed. M microphone signals with frequency $f = 1000$Hz are synthetically created with N i.i.d. snapshots. The reverberation is simulated as diffusive noise using PSD matrix $\mathbf{\Gamma}(k)$ in (13) and (15), for linear array and circular array respectively. The MUSIC method is used to estimate the DOA and the search grid is set to 0.01°. The MSE is computed and compared with the averaged CRLB for all DOAs. Fig. 2 shows the CRLBs and the MSEs estimates versus different SRRs. For linear array, the CRLBs of directional microphones are lower than that of omnidirectional microphones. The MSEs are also decreased as the directional microphones are used.

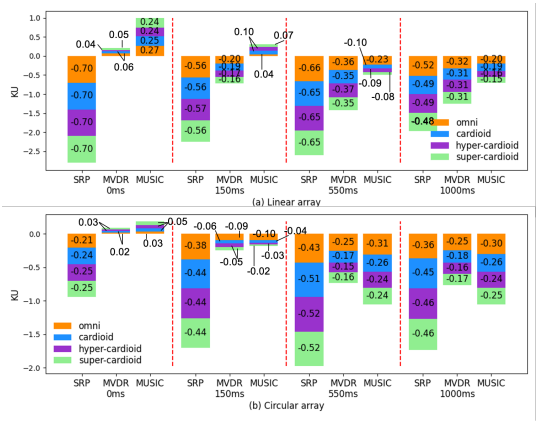


Figure 4: *Kurtosis comparison with different type microphones.*

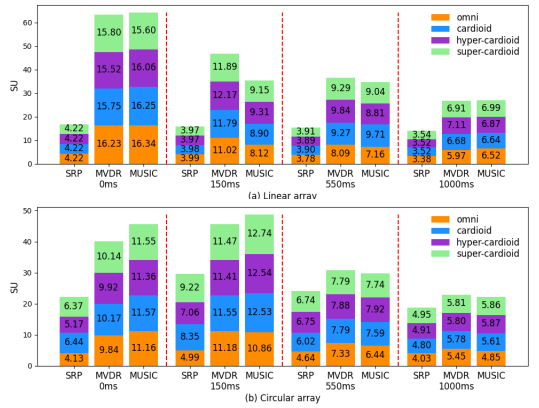


Figure 5: *Sidelobe uncertainty comparison with different type microphones.*

DOA performance. The performance of DOA estimation is examined using simulated reverberation and noisy signals.

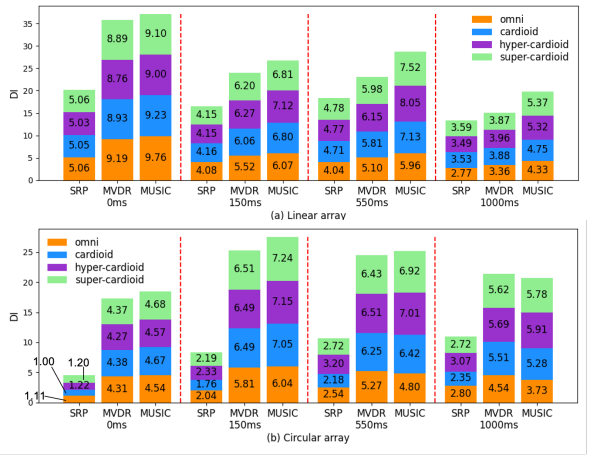


Figure 6: *Directivity index comparison with different type microphones.*

The RIRs are simulated using the image source method via *pyroomacoustics* [27] and the magnitude response is calculated for each image source according to (8). The reverberation time are set to be approximately $\{150, 550, 1000\}$ ms. The sampling frequency is 32kHz and the frame length of STFT is 32ms with 50% overlap. The frequency bins in the range $[500, 4000]$Hz are utilized to perform DOA estimation. White Gaussian noise is added in each channel to generate an average SNR 20dB (SNR $= 10\log_{10}(\frac{\eta_s}{\eta_n})$). The DOAs are generated by 5° spacing, and we thus have 72 possible positions for the circular array and 24 positions for the linear array. Fig. 3(a) shows the comparison of RMSE between directional and omnidirectional linear microphone arrays in terms of four metrics versus SRRs. It shows that a lower RMSE can be obtained using directional microphones. From Fig. 4(a), we can observe that a higher kurtosis can be obtained when applying MUSIC algorithm with directional microphones. Fig. 5(a) and Fig. 6(a) also indicate that a better spatial spectra can be obtained using directional microphones for MVDR and MUSIC methods. Since we use phase transform weighting in SRP method, there is no difference between different microphone types in terms of the spatial spectral pattern. For circular array, the results are as shown in Fig. 3(b), Fig. 4(b), Fig. 5(b) and Fig. 6(b). It can be observed that the directional microphones can improve the robustness of the DOA estimator, both in RMSE of DOA estimation and the pattern of the spatial spectra.

6. Conclusions

In this paper, by modeling reverberation as a spherically isotropic noise, the CRLB for DOA estimation is derived using directional microphones. For linear arrays, a lower CRLB and MSE can be obtained by using directional microphones. For circular arrays, although the CRLBs of different microphones are not much different, the DOA performance can also be improved by combining the magnitude information and the phase information. The simulated experiment results also demonstrate that better performance can be achieved by using directional microphone arrays for DOA estimation.

7. Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 61971186.

8. References

[1] J. Yang, X. Zhong, W. Chen, and W. Wang, "Multiple acoustic source localization in microphone array networks," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 334–347, 2020.

[2] X. Zhong and J. R. Hopgood, "Particle filtering for tdoa based acoustic source tracking: Nonconcurrent multiple talkers," *Signal processing*, vol. 96, pp. 382–394, 2014.

[3] X. Zhong and A. B. Premkumar, "Particle filtering approaches for multiple acoustic source detection and 2-d direction of arrival estimation using a single acoustic vector sensor," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4719–4733, 2012.

[4] X. Zhong and J. R. Hopgood, "A time—frequency masking based random finite set particle filtering method for multiple acoustic source detection and tracking," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 23, no. 12, pp. 2356–2370, 2015.

[5] J. Li and P. Stoica, *Robust adaptive beamforming*. Wiley Online Library, 2006.

[6] X. Chen, W. Wang, Y. Wang, X. Zhong, and A. Alinaghi, "Reverberant speech separation with probabilistic time–frequency masking for b-format recordings," *Speech Communication*, vol. 68, pp. 41–54, 2015.

[7] V. Kılıç, X. Zhong, M. Barnard, W. Wang, and J. Kittler, "Audio-visual tracking of a variable number of speakers with a random finite set approach," in *Proc. International Conference on Information Fusion*, 2014, pp. 1–7.

[8] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone arrays*. Springer, 2001, pp. 157–180.

[9] M. Al-Nuaimi, R. Shubair, and K. Al-Midfa, "Direction of arrival estimation in wireless mobile communications using minimum variance distortionless response," in *Proc. International Conference on Innovations in Information Technology*, 2005, pp. 1–5.

[10] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[11] H. Beit-On and B. Rafaely, "Speaker localization using the direct-path dominance test for arbitrary arrays," in *Proc. IEEE International Conference on the Science of Electrical Engineering in Israel*, 2018, pp. 1–4.

[12] H. Wang and J. Lu, "A robust doa estimation method for a linear microphone array under reverberant and noisy environments," *arXiv preprint arXiv:1904.06648*, 2019.

[13] O. Schwartz, Y. Dorfan, E. A. Habets, and S. Gannot, "Multi-speaker doa estimation in reverberation conditions using expectation-maximization," in *Proc. IEEE International Workshop on Acoustic Signal Enhancement*, 2016, pp. 1–5.

[14] J. R. Jensen, J. K. Nielsen, R. Heusdens, and M. G. Christensen, "Doa estimation of audio sources in reverberant environments," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016, pp. 176–180.

[15] J.-M. Valin, F. Michaud, J. Rouat, and D. Létourneau, "Robust sound source localization using a microphone array on a mobile robot," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, 2003, pp. 1228–1233.

[16] O. Nadiri and B. Rafaely, "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*, vol. 22, no. 10, pp. 1494–1505, 2014.

[17] A. Politis, S. Delikaris-Manias, and V. Pulkki, "Direction-of-arrival and diffuseness estimation above spatial aliasing for symmetrical directional microphone arrays," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2015, pp. 6–10.

[18] W. Huang and J. Feng, "Differential beamforming for uniform circular array with directional microphones," in *Proc. INTERSPEECH*, 2020, pp. 71–75.

[19] G. W. Elko, "Spatial coherence functions for differential microphones in isotropic noise fields," in *Microphone Arrays*. Springer, 2001, pp. 61–85.

[20] J. Benesty, I. Cohen, and J. Chen, *Fundamentals of Signal Enhancement and Array Signal Processing*. Wiley Online Library, 2018.

[21] J. Eargle, *The microphone handbook*. Elar Publishing Company, 1982.

[22] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1783–1795, 1990.

[23] P. Stoica and A. Nehorai, "Music, maximum likelihood, and cramer-rao bound: further results and comparisons," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 12, pp. 2140–2150, 1990.

[24] A. Newell and G. Hindman, "Antenna pattern comparison using pattern subtraction and statistical analysis," in *Proc. European Conference on Antennas and Propagation*, 2011, pp. 2537–2540.

[25] P. H. Westfall, "Kurtosis as peakedness, 1905–2014. rip," *The American Statistician*, vol. 68, no. 3, pp. 191–195, 2014.

[26] J. Bitzer and K. U. Simmer, "Superdirective microphone arrays," in *Microphone arrays*. Springer, 2001, pp. 19–38.

[27] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A python package for audio room simulation and array processing algorithms," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018, pp. 351–355.