Causal System Identification based Compensation for Reverberation-Robust DOA Estimation

Li He
China Mobile Research
Beijing, China
heliyjy@chinamobile.com

Wei Xue
JD Tech Group
Beijing, China
xuewei27@jd.com

***Abstract*—One big challenge of direction of arrival (DOA) estimation is reverberation, since the spatial cues of the source and its reflections cannot be well distinguished. In this paper we propose a DOA compensation factor based on the mean squared error (MSE) of system identification (SI), and used it to adjust the DOA cost function values of conventional methods, such that the robustness of DOA estimation to reverberation can be improved. By analysing the relationship between the beamforming outputs towards different DOAs and the signal captured by the microphone, we show that the observed signal can be more accurately predicted by the signal from beamforming towards the direct-path, than from other directions. This is indicated by a small MSE of SI, and is exploited to calculate a DOA compensation factor. The proposed compensation factor can be easily applied to conventional DOA estimation methods, regardless of the number of microphones and the type of algorithms. Experiments show that the proposed method outperforms baseline methods in various reverberant conditions.**

***Index Terms*—direction of arrival, reverberation, beamforming, system identification, compensation factor**

I. Introduction

Direction of arrival (DOA) estimation using a microphone array has wide applications in systems such as surveillance [1], hands-free devices [2], and automatic speech recognition (ASR) [3]. Such information is prerequisite for speech processing algorithms including beamforming based speech enhancement [4]–[6], dereverberation [7]–[9] and acoustic mapping [10], [11].

Over the past decades, generalized cross correlation (GCC) [12]–[15], steered beamformer response power (SRP) [16]–[21] and multiple signal classification (MUSIC) [22]–[28] based algorithms, are the most popular for DOA estimation, which generally find the DOA which has the best cost function in the spatial spectrum. A problem of these conventional methods is that they rely on the ideal anechoic assumption that only consider the direct path, which makes the performance degrade in reverberation when the signals from DOAs other than the target direction also present.

Many approaches are proposed to cope with the reverberation problem. In [29]–[31], time-frequency (TF) bins dominated by the direct-path signal are identified and used to compute the DOA cost function. More accurate signal models are also used, for example, in [32]–[36]. Specifically, [32], [33] proposed an expectation maximization (EM) based method which decompose the observed signal into the anechoic and reverberation parts; [34], [35] identified the multichannel room impulse responses (RIRs) for DOA estimation; [36] estimates a direct-path signal cross-correlation (DPCC) using the convolutive transfer function approximation. Although more realistic signal models are adopted, these methods usually use energy based criterion to design the cost function or optimization problem, and could suffer from performance degradation when closing to the strong reflector, in which case the energy of reflection is comparable to the direct-path signal.

In this paper to overcome the above problem, we additionally exploit the causality between the beamformed signal and the signal captured by the microphone array for single source DOA estimation in reverberation. The single-channel beamformed signals towards different hypothesized DOAs are dominated by either the direct-path signal or reflection. We show that when conducting system identification (SI) to estimate the transfer function between the beamformed signal and the microphone observation, the observed reverberant signal can be more accurately predicted when the beamformed signal is dominated by the direct-path signal component. Thus, the SI problem is causal when DOA used beamforming orients towards the target source, and yields a small mean squared error (MSE), which is actually the loss function of SI. A new compensation factor is thus proposed based on the MSE of SI, and is used to improve the robustness of DOA estimation.

II. Signal Model

We consider a reverberant environment with an M-element microphone array and a single target source. Assuming that the analysis window is larger than the RIR length, in the short-time Fourier transform (STFT) domain, the m-th microphone signal $Y_m(t,f)$ in the time frame t and frequency bin f is expressed as [7]:

$$\begin{aligned} Y_m(t,f) =& A_{m,d}(f)S(t,f) + A_{m,r}(f)S(t,f) + V_m(t,f) \\ =& X_{m,d}(t,f) + X_{m,r}(t,f) + V_m(t,f), \end{aligned} \tag{1}$$

where we decompose the STFT-domain RIR into the direct-path component $A_{m,d}(f)$ and late component $A_{m,r}(f)$. $X_{m,d}(t,f)$ and $X_{m,r}(t,f)$ are direct-path signal and reverberation, respectively. $V_m(t,f)$ denotes the additive noise at the m-th microphone, which is assumed to be uncorrelated with the source signal $S(t,f)$.

III. Proposed Method

In this section, we propose a compensation factor for different DOAs based on the MSE of SI, which takes the beamformed signals from different DOAs as inputs, and predicts the microphone observation. Based on the causality analysis of SI, we show that the reverberant observation can be more accurately predicted by the direct-path signal than reverberation, which results in a smaller MSE value. Further, the factor is derived based on the MSE, and is used to adjust the DOA cost function of conventional methods.

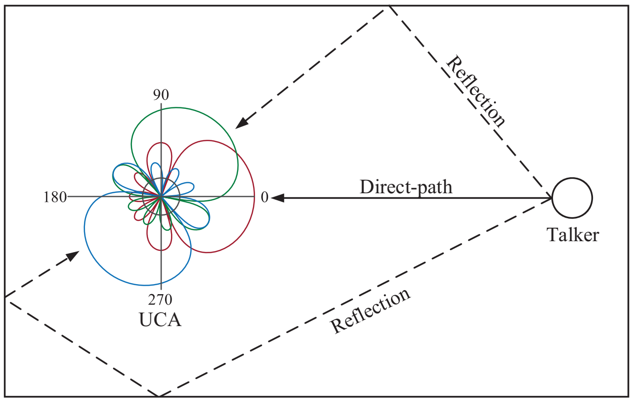

Fig. 1. Schematic illustration of uniform circular array (UCA) based beamforming towards different hypothesized DOAs.

A. Beamforming

As shown in Fig. 1, the reverberant signal can be viewed as the combination of the direct-path signal and source signal replicas coming from different mirror locations in the room [7]. To differentiate these signals for further analysis, the beamforming towards each hypothesized DOA is firstly performed to yield a single-channel enhanced signal, in which either the direct-path signal or the reverberation is dominated.

To compromise between the beam directivity and computational complexity, a fixed regularized superdirective beamformer [37] is adopted. The beamformed signal $Z(t,f,\theta)$ towards a hypothesized DOA θ is calculated as:

$$Z(t,f,\theta) = \mathbf{W}^H(f,\theta)\mathbf{Y}(t,f), \tag{2}$$

where $\mathbf{W}(f,\theta)$ denotes the beamformer, which uses the anechoic steering vector and its expression is omitted here for simplicity. $\mathbf{Y}(t,f) = [Y_1(t,f), Y_2(t,f), ..., Y_M(t,f)]^T$ is the multichannel signal vector.

Now we temporarily ignore the noise term in (1) and will discuss it later, then (2) is further decomposed as:

$$\begin{aligned} Z(t,f,\theta) &= \mathbf{W}^H(f,\theta)[\mathbf{X}_d(t,f)+\mathbf{X}_r(t,f)] \\ &= \hat{X}_{1,d}(t,f,\theta)+\hat{X}_{1,r}(t,f,\theta), \end{aligned} \tag{3}$$

where $\mathbf{X}_d(t,f) = [X_{1,d}(t,f), X_{2,d}(t,f), ..., X_{M,d}(t,f)]^T$, and $\hat{X}_{1,d}(t,f,\theta) = \mathbf{W}^H(f,\theta)\mathbf{X}_d(t,f)$ is the direct-path signal estimation of the first channel. $\mathbf{X}_r(t,f)$ and $\hat{X}_{1,r}(t,f,\theta)$ are defined similarly to $\mathbf{X}_d(t,f)$ and $\hat{X}_{1,d}(t,f,\theta)$, and denote reverberation signal components. Due to the optimization goal of the beamformer, generally we can expect that the $\hat{X}_{1,d}(t,f,\theta)$ reaches the highest dominance in $Z(t,f,\theta)$ when θ is the DOA of the source.

B. DOA Compensation Factor

1) Causal system identification: Now we identify the transfer function $H(t,f,\theta)$ between the beamformed signal and the reverberant signal. The optimization problem can be defined based on SI which minimizes the MSE as follows:

$$H(t,f,\theta) = \underset{H(t,f,\theta)}{\arg\min}\, E\{|e(t,f,\theta)|^2\}, \tag{4}$$

where $e(t,f,\theta) \triangleq Y_1(t,f) - H^*(t,f,\theta)Z(t,f,\theta)$ denotes the error signal of the SI problem and $[\ ]^*$ is the conjugation. We note that $H(t,f,\theta)$ is different with the steering vector used for beamforming in (2).

According to (1) and (3), since the noise term has been temporarily ignored here, we have:

$$\begin{aligned} e(t,f,\theta) =& X_{1,d}(t,f)+X_{1,\mathrm{r}}(t,f)- \\ & H^*(t,f,\theta)[\hat{X}_{1,d}(t,f,\theta)+\hat{X}_{1,\mathrm{r}}(t,f,\theta)] \\ =& G(f)X_{1,d}(t,f)-H^*(t,f,\theta)\hat{X}_{1,d}(t,f,\theta)- \\ & H^*(t,f,\theta)\hat{X}_{1,\mathrm{r}}(t,f,\theta), \end{aligned} \tag{5}$$

where $G(f){=}1+A_{1,r}(f)/A_{1,d}(f)$. We note that theoretically the reverberant component $X_{1,\mathrm{r}}(t,f)$ in the observed signal can be causally and linearly predicted by the direct-path signal $X_{1,\mathrm{d}}(t,f)$, since the direct-path signal comes earlier than the reverberation, and the transfer function is the time-delayed and attenuated version of the RIR.

Substitute (5) into (4), the MSE is given by:

$$\begin{aligned} & E\left\{|e(t,f,\theta)|^2\right\} \\ &= E\left\{\left|G(f)X_{1,d}(t,f)-H^*(t,f,\theta)\hat{X}_{1,d}(t,f,\theta)\right|^2\right\}+ \\ & \quad E\left\{\left|H^*(t,f,\theta)\hat{X}_{1,\mathrm{r}}(t,f,\theta)\right|^2\right\}+\epsilon(t,f), \end{aligned} \tag{6}$$

where $\epsilon(t,f)$ is the cross-correlation related term. In such case we hope to use the direct-path signal component in the signal obtained by multichannel beamforming, to predict the reverberant observation in the first channel. When the identified transfer function converges to the optimal solution, we have $E\left\{\left|G(f)X_{1,d}(t,f)-H^*(t,f,\theta)\hat{X}_{1,d}(t,f,\theta)\right|^2\right\} \to 0$. Hence, MSE is biased by $E\left\{\left|H^*(t,f,\theta)\hat{X}_{1,\mathrm{r}}(t,f,\theta)\right|^2\right\}$, which means strong reflection-path components $\hat{X}_r(t,f,\theta)$ will result in large MSE bias.

By using the beamformer which utilizes the anechoic steering vector, we aim to estimate the direct-path signal in the reference channel. Thus if the beamformer steers towards the source, we obtain the highest dominance of the direct-path components $\hat{X}_d(t,f,\theta)$, whereas minimum reflection-path components $\hat{X}_r(t,f,\theta)$ in $Z(t,f,\theta)$, such that the MSE is the smallest and the identified system $H(t,f,\theta)$ is causal. In other words, compared with the reflection, the observed signal can be better predicted by the direct-path signal, which results in a smaller MSE, thus the MSE can be regarded as an indicator of the causality of $H(t,f,\theta)$, showing whether the beamformed signal is from the direct path.

The optimization problem is solved by the frequency-domain normalized least mean square algorithm [38], and $H(t,f,\theta)$ is updated iteratively as:

$$H(t+1,f,\theta) = H(t,f,\theta) + \gamma\frac{e^*(t,f,\theta)Z(t,f,\theta)}{\sigma_z(t,f,\theta)}, \quad (7)$$

with γ being the step size,

$$\sigma_z(t,f,\theta) = \rho\sigma_z(t-1,f,\theta) + (1-\rho)|Z(t,f,\theta)|^2, \quad (8)$$

and $\rho \in (0,1]$ being a constant smoothing factor.

In practice, to improve the robustness to noise, the update is active only when the speech presence probability (SPP) exceeds a threshold β, and the SPP is estimated by [39].

In Fig. 2, we show an example of the MSEs using the beamformed signal towards the direct path and a reflection path, respectively. The observed signal can be more accurately predicted by the beamformed signal from the source than reflection, which results in a smaller MSE.

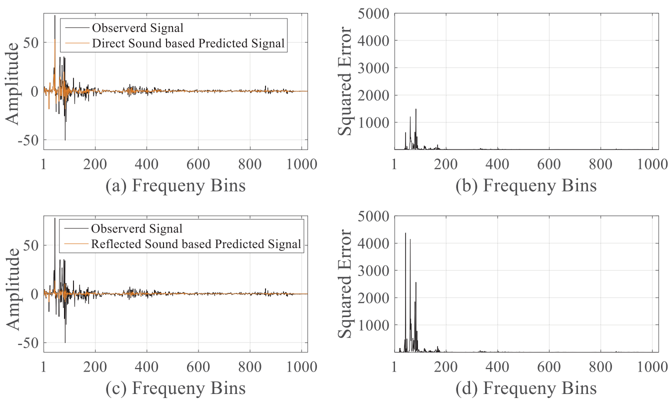

Fig. 2. The frequency-domain presentation of observed signal, signal predicted using the beamformed signal from the source (a) and reflection (c). (b) and (d) represent corresponding squared error of (a) and (c), respectively.

2) Hypothesized DOA based compensation factor: Based on the above analysis, we propose a DOA compensation factor $Q(t,\theta)$ as:

$$Q(t,\theta) = \left|\frac{\min\limits_{\theta\in\Omega_\theta}[\xi(t,\theta)]}{\xi(t,\theta)}\right|^2, \quad (9)$$

where Ω_θ is the set of all hypothesized DOAs. $\xi(t,\theta) = \sum\limits_{f\in N_b}|e(t,f,\theta)|^2$ denotes the broadband MSE summed over all frequencies, and N_b is the set of all frequency bins. The $Q(t,\theta)$ calculated in (9) is normalized to the range of $(0,1]$, and the larger the value, the greater the probability that corresponding direction is the direct path. To reduce the fluctuation of MSE, the $\xi(t,\theta)$ is updated by recursive smoothing, as:

$$\tilde{\xi}(t,\theta)=\alpha\tilde{\xi}(t-1,\theta)+(1-\alpha)\xi(t,\theta), \quad (10)$$

where $\alpha \in (0,1]$ is a smoothing factor, and $\tilde{\xi}(t,\theta)$ presents the smoothed broadband MSE, which is uesd in (9).

It should be noted that, although the causality can help to indicate the correctness of direct-path DOA hypothesis, it mainly focuses on discriminating the direct path and reflection, and is not considering to improve the spatial discrimination for adjacent DOAs, which is commonly done by conventional DOA methods. Therefore, in this paper, (9) is only used as a compensation factor for an existing DOA cost function, instead of as an independent new DOA cost function.

C. DOA Estimation

The estimated compensation factor can be directly used by the various DOA estimation methods, by:

$$\theta(t) = \arg\max_\theta[Q(t,\theta)J(t,\theta)], \quad (11)$$

where $J(t,\theta)$ denotes the cost function of the conventional method, and $\theta(t)$ is the estimated DOA in frame t. By weighting the compensation factor to DOA cost function, the peak value corresponding to the direct sound is highlighted, whereas the peak value is weakened when the enhanced signal is from reflection.

One concern of the proposed method might be the computational complexity since besides computing the DOA cost function of conventional methods, additional beamforming and SI are performed for each DOA. In fact, as the fixed beamformer $\mathbf{W}(f,\theta)$ is used whose coefficients are computed offline, and the SI is recursively updated, for each DOA, the proposed compensation factor only involves the computation of a) obtain $Z(t,f,\theta)$ from (2), b) update $H(t,f,\theta)$ from (7), and c) compute the broadband MSE and the compensation factor from (9).

Fig. 3 shows an example of the calculated compensation factor and pseudo spectrum of conventional DOA estimation methods when closing to a strong reflector. It can be observed that under the reverberant condition, the calculated compensation factor can help DOA estimation methods overcome the effect of the reverberation.

IV. Experiment

In this section, the well-known (SRP-PHAT) [40] and broadband MUSIC [22] methods are used for comparison.

A. Experimental setup and evaluation

We simultaneously record the real data using three identical six-element uniform circular microphone arrays (UCAs) with a radius of 4.25 cm in a $9\times6\times3$ m^3 meeting room. The target speaker is seated at $(3,3,1.2)$ m, and the microphone arrays are located at $(4,3,1.2)$ m, $(5,3,1.2)$ m and $(6,3,1.2)$ m, which means the target source is 1 m, 2 m and 3 m away from the array center, respectively. The reverberation time of the room is approximately 450 ms. For different source-array distances, the direct-to-reverberant ratios (DRRs) [7] are different, which changes the difficulty of DOA estimation. We also evaluate the performance as a function of microphone

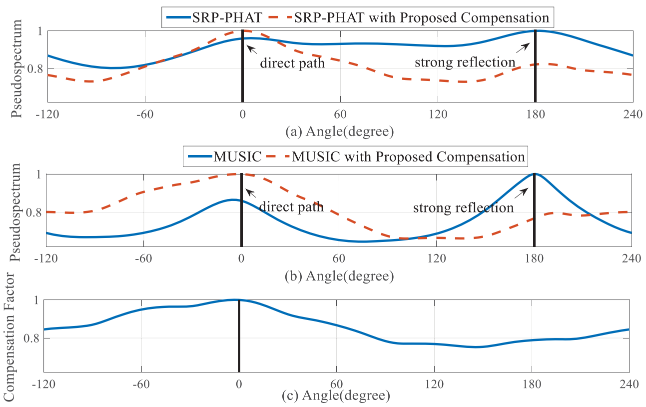

Fig. 3. The pseudo spectrum of SRP-PHAT and compensated SRP-PHAT (a), the pseudo spectrum MUSIC and compensated MUSIC (b), corresponding compensation factor (c). Direct-path DOA angle is 0 marked with vertical solid line. Reflection-path DOA angle is 180 marked with vertical dotted line. Reverberant environment: T60 is 600ms + 20dB SNR white Gaussian noise + uniform circular array (UCA) closet to a strong reflector.

number (six-microphone and dual-microphone) since the proposed method relies on beamforming. In the dual-microphone case, two microphones closet to the target source are used.

We asked twenty speakers to read randomly selected texts for signal recording under 16 kHz sampling rate, which results in about 35 minutes data for testing. The STFT analysis window is 2048 samples Hamming window with 75% overlap. Due to the diffused background noise, the signal-to-noise ratios (SNRs) were around 24 dB, 22 dB, 20.5 dB, corresponding to 1 m, 2 m and 3 m source-array distances.

In experiments for the proposed method we choose: $\gamma = 0.02$, $\rho = 0.95$, $\alpha = 0.98$, $\beta = 0.3$, $\sigma_z(1, f, \theta) = |Z(1, f, \theta)|^2$, $H(1, f, \theta)$ is initialized as zero. We choose these parameters according to a) γ is the step size for which a small value is chosen, b) ρ and α are the smoothing factors commonly used for speech signal processing, and is normally chosen as larger than 0.9, c) β is the SPP threshold normally chosen within $[0.3, 0.65]$. Two frame-level metrics, Accuracy and Root Mean Square Error (RMSE), are used. We consider the estimation as correct if the absolute error is less than a certain threshold, which is set as $10°$ here. Non-speech frames are identified by an energy-based voice detector and are excluded for evaluation.

B. Experimental results

The results for different source-array distances using UCA and dual-microphone array are depicted in Fig. 4 and Fig. 5, respectively. We can see that the proposed compensation improves conventional DOA estimation for all cases tested.

In Fig. 4, it is shown that for 1 m source-array distance, the proposed method can help SRP-PHAT to get the lowest RMSE and 95.87% accuracy. When increasing the source-array distance, all DOA estimation methods degrade in performance as a result of the decreased DRR and SNR. However, when the source-array distance is 3 m, the proposed method can still improve the accuracies of both baseline methods by nearly 20%, and the RMSE is reduced by $10°$ and $20°$ for SRP-PHAT and MUSIC, respectively. Similar results can be observed in the dual-microphone array case, as shown in Fig. 5. We see that in this case we get worse performance compared with UCA, due to the fact the available spatial information is limited. However, the proposed method can always help DOA estimation methods attain a more accurate estimation in different reverberant conditions.

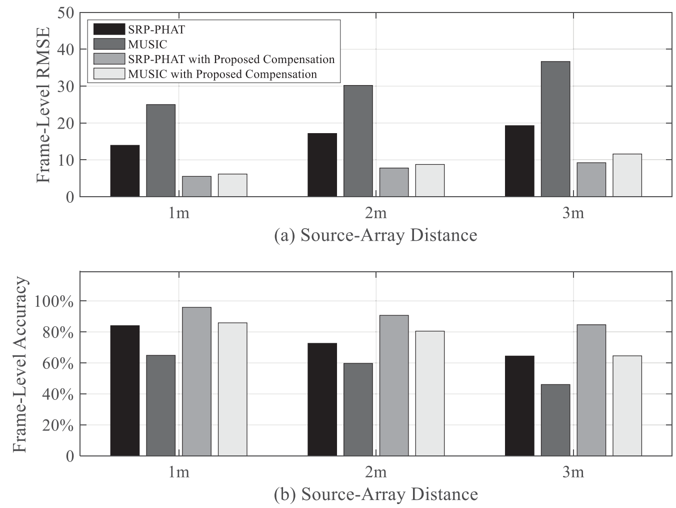

Fig. 4. Frame-level RMSE (a) and Accuracy (b) of different algorithms using uniform circular array (UCA) for different source-array distances.

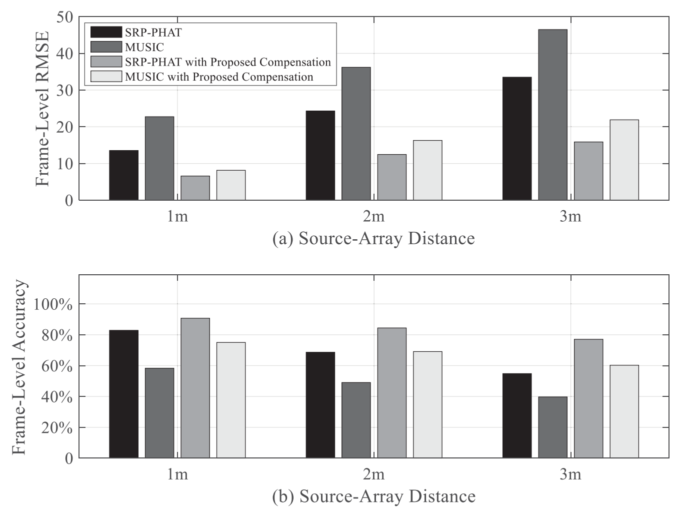

Fig. 5. Frame-level RMSE (a) and Accuracy (b) of different algorithms using dual-microphone array for different source-array distances.

V. Conclusions

In this paper, we propose to use causal SI based compensation factor to perform reverberation-robust DOA estimation. The proposed algorithm can be easily applied to various DOA estimation methods, regardless of the number of microphones and the type of algorithms. The beamforming is performed on the multichannel observed signals to obtain direct-path signal or reflection-path signal. Then by exploiting causality of SI between the enhanced signal and the observed signal, a compensation factor expressed as a function of hypothesized DOA is formulated. Furthermore, the compensation factor assists DOA estimation by weighting the DOA cost function. Experimental results on multichannel recordings in real reverberant environments demonstrate that the proposed algorithm can help different DOA estimation methods achieve lower RMSE and higher estimation accuracy.

REFERENCES

[1] S. Zhao, E.S. Chng, T. Nguyen, H. Hieu, and Li, "A robust real-time sound source localization system for olivia robot," in *Proc. The Asia-Pacific Signal and Information Processing Association (APSIPA) Annual Summit and Conference (ASC)*, 2010.

[2] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 4, pp. 692–730, 2017.

[3] N. Ito, S. Araki, M. Delcroix, and T. Nakatani, "Probabilistic spatial dictionary based online adaptive beamforming for meeting recognition in noisy and reverberant environments," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 681–685.

[4] O. L. Frost and III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[5] E. A. P. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, "New insights into the mvdr beamformer in room acoustics," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 1, pp. 158–170, 2010.

[6] L. He, Y. Zhou, X. Shu, and H. Liu, "Robust MVDR beamformer based on complex gaussian mixture model with phase prior," in *Proc. IEEE Intl. Conf. on Digital Signal Processing (DSP)*, 2018, pp. 1–5.

[7] P. A. Naylor, N. D. Gaubitch, and Eds., *Speech Dereverberation*, Springer, 2010.

[8] H. A. Javed, A. H. Moore, and P. A. Naylor, "Spherical harmonic rake receivers for dereverberation," in *2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2016, pp. 1–5.

[9] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B. H. Juang, "Speech dereverberation based on variance-normalized delayed linear prediction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1717–1731.

[10] S. Zhao, C. Tuna, T. N. T. Nguyen, and D. L. Jones, "Large-region acoustic source mapping using a movable array and sparse covariance fitting," *J. Acoust. Soc. Am.*, vol. 141, no. 1, pp. 357–372.

[11] C. Evers, A. H. Moore, and P. A. Naylor, "Acoustic simultaneous localization and mapping (A-SLAM) of a moving microphone array and its surrounding speakers," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 6–10.

[12] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[13] R. Yaguchi, S. Shiota, N. Ono, and H. Kiya, "Replay attack detection using generalized cross-correlation of stereo signal," in *Proc. European Signal Processing Conf. (EUSIPCO)*, 2019, pp. 1–5.

[14] L. Huang, S. Zhang, D. Zan, X. Zhang, and F. Li, "Speaker localization with smoothing generalized cross correlation based on naive bayes classifier," in *Proc. IEEE Intl. Conf. on Information Communication and Signal Processing (ICICSP)*, 2018, pp. 98–102.

[15] J. Benesty, J. Chen, and Y. Huang, "Time-delay estimation via linear interpolation and cross correlation," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 509–519, 2004.

[16] D. Salvati, C. Drioli, and G. L. Foresti, "Incoherent frequency fusion for broadband steered response power algorithms in noisy environments," *IEEE Signal Process. Lett.*, vol. 21, no. 5, pp. 581–585, 2014.

[17] M. Cobos, M. Garca-Pineda, and M. Arevalillo-Herrez, "Steered response power localization of acoustic passband signals," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 717–721, 2017.

[18] P. Pertil and E. Cakir, "Robust direction estimation with convolutional neural networks based steered response power," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 6125–6129.

[19] J. P. Dmochowski, J. Benesty, and S. Affes, "A generalized steered response power method for computationally viable source localization," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 8, pp. 2510–2526, 2007.

[20] H. Do and H. F. Silverman, "Stochastic particle filtering: A fast srp-phat single source localization algorithm," in *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2009, pp. 213–216.

[21] K. Wu and A. W. H. Khong, "Acoustic source tracking in reverberant environment using regional steered response power measurement," in *Proc. The Asia-Pacific Signal and Information Processing Association (APSIPA) Annual Summit and Conference (ASC)*, 2013, pp. 1–6.

[22] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[23] D. Zhang, Y. Zhang, G. Zheng, C. Feng, and J. Tang, "Improved DOA estimation algorithm for co-prime linear arrays using root-MUSIC algorithm," *Electron. Lett.*, vol. 53, no. 18, pp. 1277–1279, 2017.

[24] A. P. Mahnaz and G. Sedigheh, "Mℓ1,2-MUSIC algorithm for DOA estimation of coherent sources," *IET Signal Process.*, vol. 11, no. 4, pp. 429–436, 2017.

[25] G. Jiang, X. Mao, and Y. Liu, "Reducing errors for root-MUSIC-based methods in uniform circular arrays," *IET Signal Process.*, vol. 12, no. 1, pp. 31–36, 2018.

[26] J. J. Ermer, J. C. Mosher, M. Huang, and R. M. Leahy, "Paired MEG data set source localization using recursively applied and projected (RAP) MUSIC," *IEEE Trans. Biomedical Engineering*, vol. 47, no. 9, pp. 1248–1260, 2000.

[27] M. L. McCloud and L. L. Scharf, "A new subspace identification algorithm for high-resolution doa estimation," *IEEE Trans. Antennas Propag.*, vol. 50, no. 10, pp. 1382–1390, 2002.

[28] J. P. Dmochowski, J. Benesty, and S. Affes, "Broadband music: Opportunities and challenges for multiple source localization," in *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2007, pp. 18–21.

[29] S. Mohan, M. E. Lockwood, M. L. Kramer, and D. L. Jones, "Localization of multiple acoustic sources with small arrays using a coherence test," *J. Acoust. Soc. Am.*, vol. 123, pp. 2136–2147, 2008.

[30] 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 Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 10, pp. 1494–1505, Oct. 2014.

[31] A. H. Moore, C. Evers, P. A. Naylor, D. L. Alon, and B. Rafaely, "Direction of arrival estimation using pseudo-intensity vectors with direct-path dominance test," in *Proc. European Signal Processing Conf. (EUSIPCO)*, 2015.

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

[33] Y. Dorfan, O. Schwartz, B. Schwartz, E. A. P. Habets, and S. Gannot, "Multiple doa estimation and blind source separation using estimation-maximization," in *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, 2016, pp. 1–5.

[34] Y. Huang and J. Benesty, *Adaptive Signal Processing*, Springer, 2003.

[35] K. Kowalczyk, E. A. P. Habets, W. Kellermann, and P. A. Naylor, "Blind system identification using sparse learning for TDOA estimation of room reflections," *IEEE Signal Process. Lett.*, vol. 20, pp. 653–656, 2013.

[36] W. Xue, Y. Tong, G. Ding, C. Zhang, T. Ma, X. He, and B. Zhou, "Direct-path signal cross-correlation estimation for sound source localization in reverberation," in *Proc. Conf. of Intl. Speech Commun. Assoc. (INTERSPEECH)*, 2019.

[37] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 10, pp. 1365–1376, 1987.

[38] E. Ferrara, "Fast implementations of lms adaptive filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 4, pp. 474–475, 1980.

[39] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 4, pp. 1383–1393, May. 2012.

[40] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Springer, May. 2001.