Sound & Vibration 2026, 60(5), 4413.
doi: https://doi.org/10.59400/sv4413

Article

A practical approach for improving the robustness of MVDR beamformers

Nguyen Thi Huyen Chau[1], **Quan Trong The**[2,*], **Pham Van Ha**[3]

[1] Faculty of Information Technology, Thang Long University, Hanoi 100000, Vietnam

[2] Blockchain Lab, Faculty of Information Security, Posts and Telecommunications Institute of Technology (PTIT), Hanoi 100000, Vietnam

[3] School of Information & Communications Technology, Hanoi University of Industry, Hanoi 100000, Vietnam

*** Corresponding author:** Quan Trong The, theqt@ptit.edu.vn

CITATION

Chau NTH, The QT, Ha PV. A practical approach for improving the robustness of MVDR beamformers. Sound & Vibration. 2026; 60(5): 4413. https://doi.org/10.59400/sv4413

ARTICLE INFO

Received: 19 May 2026
Revised: 21 June 2026
Accepted: 26 June 2026
Available online: 24 July 2026

Abstract: Microphone array (MA) owns the convenience of alleviating the background noise field, interference, and third-party speakers while preserving the original speech component with a high directivity index perspective. MA beamforming utilizes prior spatial information about the direction of arrival of useful signals, the characteristics of the surrounding noise field, the designed geometry of MA, and the obtained parameters after processing received array signals to achieve the advantages of speech enhancement and noise reduction. Minimum Variance Distortionless Response (MVDR) beamformer has the capability of attenuating the surrounding noise field, interference, or third-party talker while saving the original speech component of the desired talker at a specified location. However, under realistic recording scenarios, due to the complex and annoying situation, the movement of the talker during conversation, the error of internal settings for capturing the noisy mixture, the error of sampling frequency, the different time of starting recording of two microphones, the overall effectiveness of the MVDR beamformer is often degraded because of speech distortion and musical noise. In this article, the author proposed an efficient method for enhancing the robustness of the MVDR beamformer under complex and annoying situations. The numerical simulations have shown that the speech distortion was reduced to 5 dB, the musical and residual noise were suppressed to 15.2 dB, and the speech quality was increased from 12.1 to 12.9 dB.

Keywords: microphone array; beamforming; noise reduction; speech enhancement; the signal-to-noise ratio; minimum variance distortionless response

1. Introduction

In several speech applications, such as, teleconference systems, surveillance devices, voice-controlled equipment, smartphones, smart homes, hearing aids, the demand for noise reduction and recovering the original speech component is an essential requirement for satisfying the listener. The single-channel approach often utilizes spectral subtraction (SS) in stationary noise environments, which yields promising signals with high quality. However, as in **Figure 1**, under non-stationary situations, because of undetermined reasons, the annoying complex scenario, the existence of numerous noise sources, and the internal electrical noise of equipment, SS causes speech distortion due to the estimation of noise power not being exact under complex situations.

Figure 1. The complexity with the presence of various types of noise sources in the human living environment.

Therefore, the use of MA is commonly implemented in almost all speech equipment. The MA beamformer exploits prior spatial information about the preferred impinging angle of the target speaker, desired talker, or clean speech data [1–3], the coherence of observed MA signals [4–6], the characteristics of the recorded scenario and background noise, the designed geometry of MA for performing noise suppression and preserving the speech component with high quality or different perceptual metrics for listeners at the same time [7–9], as in **Figure 2**. As a result, the effectiveness of the beamformer, the noise level reduction, the speech intelligibility, objective and subjective measurement, the signal-to-noise ratio, and speech quality of the beamformer's evaluation have increased by preserving clean speech data while attenuating background noise, interference, or a third-party speaker.

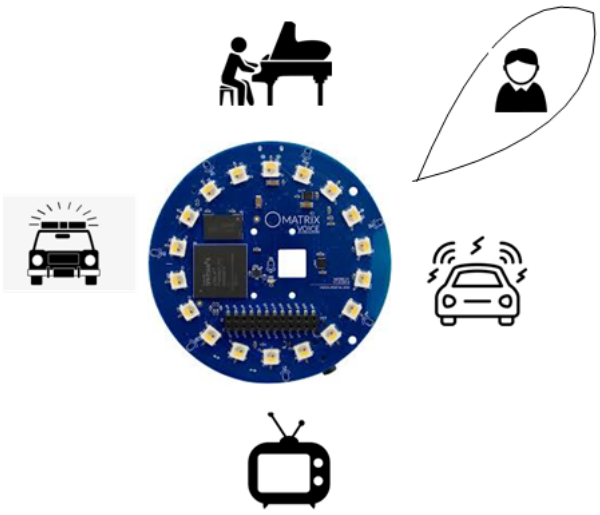

Figure 2. The scheme of MA beamforming to mitigate the effect of complex conditions while saving the target sound.

Beamforming technique is a process which generates an appropriate spatial filtering on a certain designed MA geometry. A beamformer can extract the original useful information from a noisy mixture of array signals by applying a constrained criterion without speech distortion. Array antenna system captures the signal from different and various locations and directs the only useful signal while attenuating interference from different directions by implementing a MA beamformer. In realistic scenarios, the effectiveness of voice-controlled equipment, automatic speech recognition, speech distant acquisition, teleconference system, television, smart speaker, far-field voice interaction; beamforming method own more strength in enhancing the overall evaluation in comparison to single channel processing, due to MA utilizes the prior information about assumed direction of arrival, the designed MA geometry, the characteristics of surrounding environment, the type of noise for obtaining improvement of noise suppression and speech enhancement at the same time. Beamforming techniques have become a perspective solution, which serves at the front-end of almost acoustic speech applications, as in **Figure 3**.

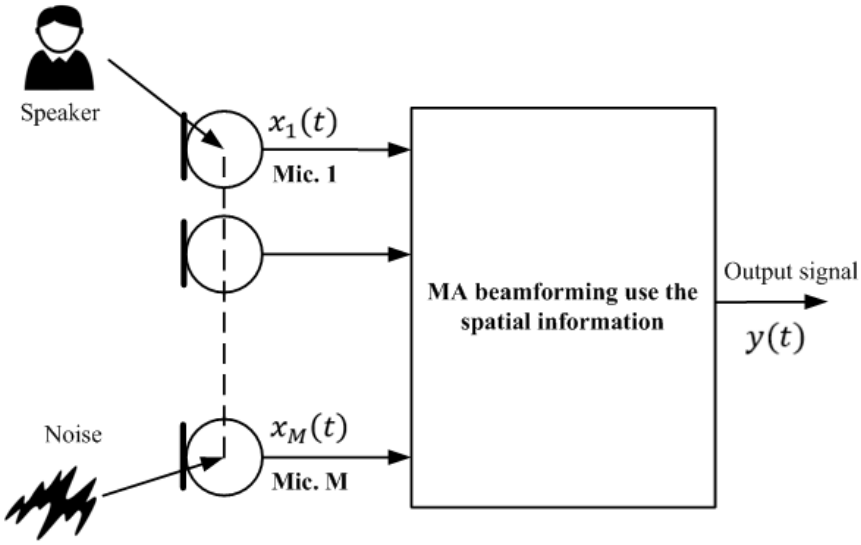

Figure 3. The scheme of the MA beamformer to recover the speech component while attenuating the background noise field.

The MA beamformer includes fixed and adaptive beamformers, which are designed and applied for processing received array signals with a specified purpose. Fixed beamformer, which the coefficients are based on the impinging angle of the helpful signal, with a common sufficient delay-and-sum (DAS) beamformer. DAS is very sufficient in removing surrounding noncoherent noise but inefficient in nonstationary with complex scenarios and reverberant noise fields. Another widely installed beamformer is the adaptive beamformer, which analyzes the condition of recording and the properties of background noise for performing beamformer. Adaptive beamformer with Minimum Variance Distortionless Response (MVDR) beamformer, Generalized Sidelobe Canceller (GSC) beamformer, Linearly Constraint Minimum Variance (LCMV) beamformer, or Differential Microphone Array (DIF) beamformer. These above beamformers are usually based on optimum criteria for extracting useful clean speech data at specified locations, which are contained in representations of

steering vectors and removing the background noise field by considering a steerable pattern.

GSC beamformer's structure contains 3 parts: (i) the fixed beamformer, which steers the pattern at the target speaker; (ii) the blocking matrix suppresses the speech component and passes only the noise component; and (ii) the adaptive noise canceller preserves the desired talker while reducing the noise level by applying adaptive signal processing algorithms. LCMV is an efficient beamformer that uses constrained criteria of frequency response in different directions while alleviating total noise power at the output signal. DIF operates on MA signals and implements a subtraction signal with a designed time delay to achieve high spatial diversity in the direction of speech and a null-beam pattern at the noise source. In addition, the MA beamformer can be integrated with different signal processing algorithms to improve the effectiveness of speech enhancement.

MVDR beamformer aims at alleviating the noise power, the third-party speaker, and the annoying interference while warranting that the original signal is not distorted in a certain direction. The representation of the MVDR beamformer is formulated by minimizing the total noise power at the output and the value of the beampattern is equal "1" at the direction of the talker. The MVDR beamformer is parameterized by the observed spatial covariance matrix of recorded signals and related transfer function, which describes the relation between the speech source and microphones. The transfer function or steering vector projects the original signal at a specified location related to the configuration of the microphone system, and thus an accurate and precise steering vector is a potential problem in almost all speech applications.

MVDR beamformer owns high spatial diversity with perspective noise reduction and speech enhancement. Its main purpose is minimizing the effects of third-party speakers at other directions, interference, or noisy environments while saving the helpful talker without speech distortion or maximum signal-to-noise and other perceptual metrics. In comparison with single-channel methods or different techniques, the MVDR beamformer provides better noise suppression and speech enhancement.

However, the MVDR beamformer is very sensitive to the exact steering vector and complex environment. In addition, due to the microphone mismatches, the internal electrical noise, the different times of recording between microphones, the annoying recording situation, the presence of undetermined noise sources, the expected results may not be satisfied, the overall speech quality, perceptual acoustic factors, and MVDR beamformer's evaluation often degrade. The distortion and musical noise decrease speech intelligibility and other perceptual metrics for listeners.

There are numerous works and research studies to overcome this drawback.

In As'ad et al. [10], the author designed a perspective beamforming algorithm for binaural hearing aids. This approach is based on the selection of a monaural–binaural MVDR beamformer at each time–frequency bin, which depends on an efficient target activity detection system. The experiments demonstrated the advantage of increasing robustness in comparison to both monaural-binaural MVDR beamformers.

Yamaoka et al. [11] used time-frequency (TF) binary masking, which requires effective target-active periods and interferer-wise-active periods for generating an

efficient spectral mask. This paper conducted experiments under realistic complex noisy conditions with the existence of different types of noise sources. The author's method has the advantage of enhancing the MVDR beamformer's performance regardless of the presence of different noise sources.

Ali et al. [12] analyzed the effects of sound location, designed geometry, room properties, microphone characteristics, or transfer function on beamformers. This paper introduced integrating the traditional MVDR beamformer with LCMV to obtain more benefit in speech enhancement and noise reduction.

Lagacé et al [13] proposed using ego-noise reduction as a preprocessing step, which allowed resolving the problem of unseen internal electrical and mechanical interference, which often occur during operating of equipment. The experimental results have shown the improvement of signal-to-distortion ratio and word error rate.

Yadav et al. [14] presented acoustic vector sensors (AVS), which contained spatial information about the impinging angle of sound propagation, velocity, and characteristics of frequency. This work provided insight into the evaluation of the MVDR beamformer for practical situations.

The spectral properties of fractional lower order covariance (FLOC) are studied for improving the robustness in terms of noise reduction and speech enhancement of the MVDR beamformer in Song [15]. The experimental simulations have demonstrated the higher stability of reducing steering vector mismatches under realistic recording situations.

Chaudhari et al. [16] aimed to improve the stability of the MVDR beamformer by using a diagonal loading factor, which is added to the second diagonal of the covariance matrix. The contribution addressed the problem of estimating the adaptive loading factor, which is based on the snapshots of MA signals. An analysis of the tradeoff between robustness and stability was performed.

Wang et al. [17] addressed the sensitivities of space-time MVDR beamformers with frequency errors by combining broadband spatial spectrum estimation computation. The method utilized beam response in any arbitrary order frequency and integrated gain constraint for achieving the optimal weight of beamformer. The demonstrated experiments show the effective evaluation of reducing the complexity of calculation and increasing the speech enhancement of the MVDR beamformer.

Erdim et al. [18] suggested using the covariance matrix taper (CMT) for extending the beampattern notches for suppressing multiple interferences, annoying noise sources, and different talkers from other directions. This approach increased the MVDR beamformer's evaluation in complicated scenarios by removing the effect of several undetermined noise sources.

Huang et al. [19] proposed methods for finding an optimum steering vector by formulating the beamformer output power maximization and double-sided norm perturbation constraint. The author simulated the experiments with improved performance in terms of signal-to-noise-plus-interference ratio (SINR).

Unfortunately, under realistic conditions with the presence of transport vehicles, washing machines, televisions, fans, or third-party speakers, the moving head of the talker, the internal electrical noise, the different sampling frequency, the microphone

mismatches, the error of calculation about the steering vector, the displacement of the MA distribution, the microphone mismatches, the different times of starting recording of microphones, the moving head of the talker, the different types of noise fields, these mentioned approaches cannot deal with the complicated drawback of the MVDR beamformer. As a result, this article proposes a two-stage speech enhancement for the MVDR beamformer with the aim of suppressing musical/residual noise fields, reducing speech distortion while increasing the speech quality by comparing the SNR between MA signals, the MVDR beamformer's output signal, and the suggested technique.

At the first stage, the paper's approach attempts to use the properties of beam patterns for estimating accurate time delay between microphones. Consequently, at the second stage, the author uses prior information about the direction of arrival of interest useful clean speech data for computing the temporal signal-to-noise (SNR) ratio to obtain the formulation of coherence noise. To improve the capability of suppressing noise at low frequencies, the coherence between noise was modified. Finally, the suggested method allowed achieving an efficient steering vector and optimum coefficients of the MVDR beamformer.

The article is organized as: Section 2 introduces the principal working of the MVDR beamformer in the frequency domain. The author's idea is presented in Section 3. The experiment demonstrated in Section 4 with significant improvement of the MVDR beamformer's evaluation. The conclusion and the author's future work are in Section 5.

2. Minimum variance distortionless response beamformer

This section demonstrates the evaluation of the MVDR beamformer using a dual-microphone system, which is commonly installed into various types of acoustic equipment, such as cochlear implants, voice-controlled devices, security systems, hearing aids, video teleconferencing systems, smartphones, surveillance devices, and biometric systems. The author utilizes a dual-microphone system to describe the configuration of the MVDR beamformer in **Figure 4**.

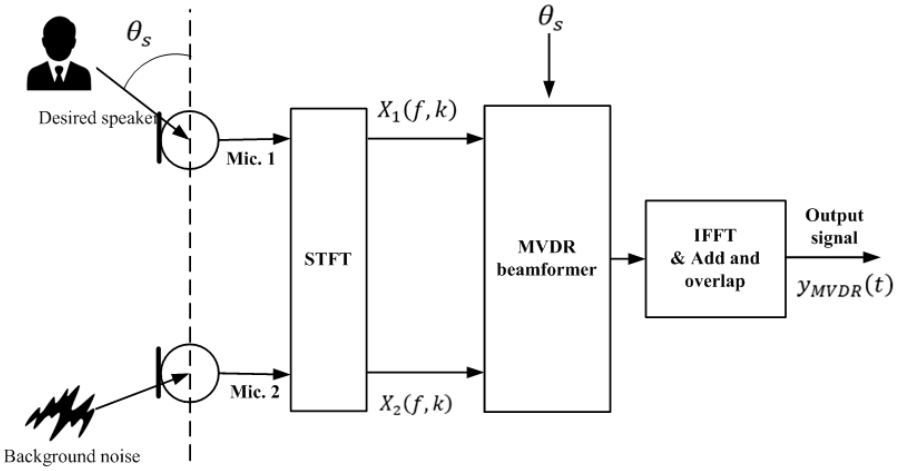

Figure 4. The model of MVDR beamformer for processing MA signals in the frequency domain.

At the current considered frame k, current frequency f, the preferred impinging incident angle of the target speaker is θ_s, the representation of received array signals

can be expressed as:

$$\begin{cases} X_1(f,k) = S(f,k)e^{j\Phi_s} + N_1(f,k) \\ X_2(f,k) = S(f,k)e^{-j\Phi_s} + N_2(f,k) \end{cases} \tag{1}$$

where $S(f,k)$ means the helpful speech component of the target speaker; $\Phi_s = \pi f \tau_0 cos\left(\theta_s\right)$ is the phase delay between the sound source and microphones, $\tau_0 = \frac{d}{c}$, d is the inter-distance of microphones, $c = 343$ (m/s) is the sound speed propagation in fresh air; $N_1(f,k)$, $N_2(f,k)$ is the annoying environment, additive noise or interference.

We can use the definitions $\boldsymbol{X}(f,k) = \begin{bmatrix} X_1(f,k) & X_2(f,k) \end{bmatrix}^T$, $\boldsymbol{N}(f,k) = \begin{bmatrix} N_1(f,k) & N_2(f,k) \end{bmatrix}^T$ and steering vector $\boldsymbol{D}_s\left(f,\theta_s\right) = \begin{bmatrix} e^{j\Phi_s} & e^{-j\Phi_s} \end{bmatrix}^T$, Equations (1) and (2) can be rewritten as:

$$\boldsymbol{X}(f,k) = S(f,k)\boldsymbol{D}_s\left(f,\theta_s\right) + \boldsymbol{N}(f,k) \tag{2}$$

The problem of most speech applications is determining the necessary coefficients $\boldsymbol{W}(f,k)$ for obtaining the original speech component at a known sound location while eliminating the surrounding noise field, annoying acoustic factors from different directions, interference, or a third-party talker in complicated situations. The promising signal $\widehat{S}(f,k) = \boldsymbol{W}^H(f,k)\boldsymbol{X}(f,k)$ is approximately the clean speech data. With different MA beamformers, the constrained criteria for recovering the pure speech component are expressed with different equations. With MVDR beamformer, the expression can be formulated as:

$$\begin{aligned} &\underset{\boldsymbol{W}(f,k)}{min} \boldsymbol{W}^H(f,k)\boldsymbol{\Phi}_{NN}(f,k)\boldsymbol{W}(f,k) \\ &st\ \boldsymbol{W}^H(f,k)\boldsymbol{D}_s\left(f,\theta_s\right) = 1 \end{aligned} \tag{3}$$

where $\boldsymbol{\Phi}_{NN}(f,k) = E\left\{\boldsymbol{N}^H(f,k)\boldsymbol{N}(f,k)\right\}$ is the covariance matrix of noise.

MVDR beamformer aims at preserving the speech data with specified direction, warranting the value of the beampattern equal to "1" and minimizing the total output noise power while reducing the effect of background noise.

The optimum MVDR beamformer's solution can be defined as:

$$\boldsymbol{W}_{MVDR}(f,k) = \frac{\boldsymbol{\Phi}_{NN}^{-1}(f,k)\boldsymbol{D}_s\left(f,\theta_s\right)}{\boldsymbol{D}_S^H\left(f,\theta_s\right)\boldsymbol{\Phi}_{NN}^{-1}(f,k)\boldsymbol{D}_s\left(f,\theta_s\right)} \tag{4}$$

However, under realistic recording situations, due to the complexity of environments, the influence of complicated acoustic factors, the presence of undetermined noise sources, the task of computing information about noise is still a difficult challenge. Therefore, the observed microphone array signals are applied for calculating the MVDR beamformer's weights.

$$\boldsymbol{W}_{MVDR}(f,k) = \frac{\boldsymbol{\Phi}_{XX}^{-1}(f,k)\boldsymbol{D}_s\left(f,\theta_s\right)}{\boldsymbol{D}_s^H\left(f,\theta_s\right)\boldsymbol{\Phi}_{XX}^{-1}(f,k)\boldsymbol{D}_s\left(f,\theta_s\right)} \tag{5}$$

where $\boldsymbol{\Phi}_{XX}(f,k) = E\left\{\boldsymbol{X}^H(f,k)\boldsymbol{X}(f,k)\right\}$.

The auto-cross power spectral densities of $X_1(f,k)$, $X_2(f,k)$ can be calculated by applying the recursive equations:

$$P_{X_iX_i}(f,k) = \alpha P_{X_iX_i}(f,k-1) + (1-\alpha)\left|X_i(f,k)\right|^2 \tag{6}$$

$$P_{X_iX_j}(f,k) = \alpha P_{X_iX_j}(f,k-1) + (1-\alpha)X_i^*(f,k)X_j(f,k) \tag{7}$$

where $i,j = 1,2$ and α is the smoothing parameter in the range $\{0\ldots1\}$. In our work, $\alpha = 0.1$.

The output signal of the MVDR beamformer can be defined as:

$$Y_{MVDR}(f,k) = \boldsymbol{W}_{MVDR}^H(f,k)\boldsymbol{X}(f,k) \tag{8}$$

Unfortunately, in real-life acoustic situations, due to the error of transfer function, the internal electrical noise of acoustic equipment, the movement of speaker, the inaccurate calculation of transferring function, the displacement of MA geometry, the different time of starting record, the different microphone sensitivities, the existence of complex noise environment, MA beamformer's performance is often corrupted. The large amount of ambient noise, musical noise, residual noise, or speech distortion causes the degraded MVDR beamformer's evaluation under practical situations. In the next section, the author will introduce an effective method for decreasing the speech distortion and suppressing the remaining noise level of the MVDR beamformer's output signal with perspective results of noise suppression and speech enhancement.

3. The author's proposed approach

The author's proposed approach includes the precise calculation of time delay τ_s and then determining the matrix coherence of noise for enhancing the MVDR beamformer's evaluation and effectiveness of noise suppression under complicated recording situations as in **Figure 5**.

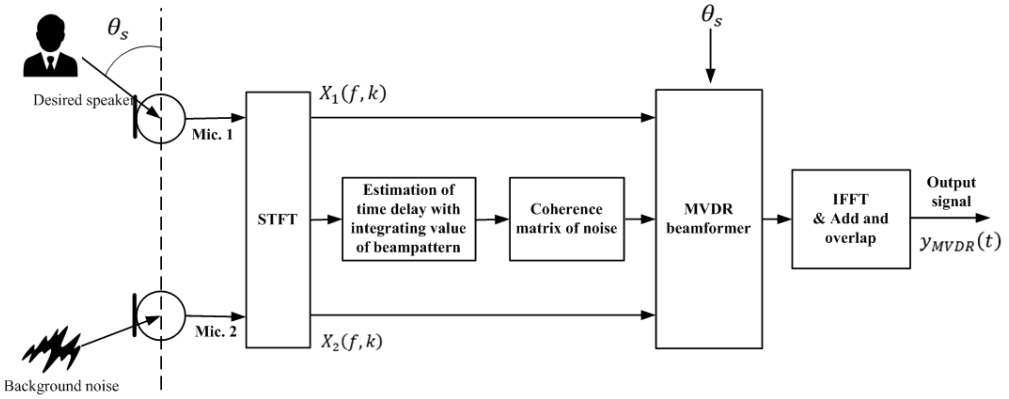

Figure 5. Two-stage-based on the suggested technique for enhancing the robustness of the MVDR beamformer.

At the first stage, the necessary time delay allows obtaining the precise steering vector or transfer function, which indicates where the beampattern concerns to save the

clean speech data in the MVDR beamformer. With GCC-PHAT [20], the time delay τ_s yields as:

$$\tau_s = \underset{\tau}{max}\ \frac{X_1(f,k)X_2^*(f,k)}{|X_1(f,k)X_2(f,k)|}e^{-j\pi f\tau} \tag{9}$$

with τ is an estimated value of time delay. However, due to a rapidly changing environment, the moving talker during conversation, the complex conditions, the error of sampling frequency, the microphone mismatches, the effect of surrounding noise field, the effectiveness of (9) can be seriously degraded. The author's idea is utilizing the value of beampattern for increasing the exact computation of time delay τ_s.

The value of the pattern of a beamformer is defined by using a delay-and-sum beamformer as:

$$B(f) = \frac{\left|X_1(f,k)e^{-j\pi f\frac{\tau}{2}} + X_2(f,k)e^{j\pi f\frac{\tau}{2}}\right|}{|X_1(f,k)|} \tag{10}$$

with assumption, $|X_1(f,k)| \approx |X_2(f,k)|$, the beam pattern can be achieved as:

$$B(f) = \left|e^{j\angle(X_1(f,k))}e^{-j\pi f\frac{\tau}{2}} + e^{j\angle(X_2(f,k))}e^{j\pi f\frac{\tau}{2}}\right| \tag{11}$$

Consequently, this paper modifies the formulation, which is used for calculating the precise time delay as the described expression:

$$\tau_s = \underset{\tau}{max}\left(\frac{X_1(f,k)X_2^*(f,k)}{|X_1(f,k)X_2(f,k)|}e^{-j\pi f\tau} + \left|e^{j\angle(X_1(f,k))}e^{-j\pi f\frac{\tau}{2}} + e^{j\angle(X_2(f,k))}e^{j\pi f\frac{\tau}{2}}\right|\right) \tag{12}$$

With precise time delay τ_s, we will receive a more robust steering vector, $\boldsymbol{D}_s(f,\theta_s) = \begin{bmatrix} e^{j\Phi_s} & e^{-j\Phi_s}\end{bmatrix}^T$, $\Phi_s = \pi f\tau_s$.

GCC-PHAT measured the temporal estimation of time delay, but it is very sensitive to wall, ceiling, and floor reflections and is not suitable for broadband signals, such as speech. Consequently, the article applied another measurement for beampattern, which is based on a DAS beamformer for enhancing the calculation of time delay τ_s.

From Equation (4), we can easily realize that the effectiveness of the MVDR beamformer usually depends on the precise covariance matrix of noise, $\boldsymbol{\Phi}_{NN}(f,k)$.

$$\boldsymbol{\Phi}_{NN}(f,k) = \sigma_n^2(f,k)\begin{bmatrix} 1 & \Gamma_{N_1N_2}(f,k) \\ \Gamma_{N_2N_1}(f,k) & 1\end{bmatrix} \tag{13}$$

where $\sigma_n^2(f,k)$ is the variance of noise components, and $\Gamma_{N_1N_2}(f,k)$, $\Gamma_{N_2N_1}(f,k)$ is the coherence between two noise parts at microphone 1 and 2.

Consequently, from (13) and (4), the coefficient of the MVDR beamformer can be achieved as:

$$\boldsymbol{W}_{MVDR}(f,k) = \frac{\begin{bmatrix} 1 & \Gamma_{N_1N_2}(f,k) \\ \Gamma_{N_2N_1}(f,k) & 1\end{bmatrix}^{-1}\boldsymbol{D}_s(f,\theta_s)}{\boldsymbol{D}_s^H(f,\theta_s)\begin{bmatrix} 1 & \Gamma_{N_1N_2}(f,k) \\ \Gamma_{N_2N_1}(f,k) & 1\end{bmatrix}^{-1}\boldsymbol{D}_s(f,\theta_s)} \tag{14}$$

and:

$$\boldsymbol{W}_{MVDR}(f,k)=\frac{\begin{bmatrix}1 & -\Gamma_{N_1N_2}(f,k)\\ -\Gamma_{N_2N_1}(f,k) & 1\end{bmatrix}\boldsymbol{D}_s(f,\theta_s)}{\boldsymbol{D}_s^H(f,\theta_s)\begin{bmatrix}1 & -\Gamma_{N_1N_2}(f,k)\\ -\Gamma_{N_2N_1}(f,k) & 1\end{bmatrix}\boldsymbol{D}_s(f,\theta_s)} \tag{15}$$

Now, the problem is finding an optimum expression of coherence $\Gamma_{N_1N_2}(f,k)$, which contains rapidly changed complex environment.

As in Yousefian et al. [21], the coherence between MA signals represents the existence of phase delay Φ_s and the coherence of noise.

$$\Gamma_{X_1X_2}(f,k)=\frac{SNR(f,k)}{1+SNR(f,k)}e^{j2\Phi_s}+\frac{1}{1+SNR(f,k)}\Gamma_{N_1N_2}(f,k) \tag{16}$$

where:

$$\Gamma_{X_1X_2}(f,k)=\frac{P_{X_1X_2}(f,k)}{\sqrt{P_{X_1X_1}(f,k)\times P_{X_2X_2}(f,k)}} \tag{17}$$

and $SNR(f,k)=\frac{\sigma_s^2(f,k)}{\sigma_n^2(f,k)}$ is the temporal signal-to-noise ratio, $\sigma_s^2(f,k)$ is the variance of speech. With the precise time delay τ_s, which is determined at the first stage, the temporal $SNR(f,k)$ yields as:

$$SNR(f,k)=\frac{1+cos\left(\Phi_{12,k}^{norm}(f)\right)}{4\left(1-cos\left(\Phi_{12,k}^{norm}(f)\right)\right)} \tag{18}$$

where $\Phi_{12,k}^{norm}(f)=\frac{\angle(X_1(f,k))-\angle(X_2(f,k))}{2\pi f\tau_s}$, $\angle$ is the operator that takes the angle.

From (16), the coherence $\Gamma_{N_1N_2}(f,k)$ expressed as:

$$\Gamma_{N_1N_2}(f,k)=\Gamma_{X_1X_2}(f,k)\left(1+SNR(f,k)\right)-e^{j2\Phi_s}SNR(f,k) \tag{19}$$

In a similar way, the formulation of $\Gamma_{N_2N_1}(f,k)$ can be determined as:

$$\Gamma_{N_2N_1}(f,k)=\Gamma_{X_2X_1}(f,k)\left(1+SNR(f,k)\right)-e^{-j2\Phi_s}SNR(f,k) \tag{20}$$

Under real-life practical speech application, this paper modified the formulation of coherence of noise field by multiplying with $\frac{sin(2\pi f\tau_s)}{2\pi f\tau_s}$ [22] as below formulations:

$$\widehat{\Gamma}_{N_1N_2}(f,k)=\Gamma_{N_1N_2}(f,k)\times\frac{sin\left(2\pi f\tau_s\right)}{2\pi f\tau_s} \tag{21}$$

$$\widehat{\Gamma}_{N_2N_1}(f,k)=\Gamma_{N_2N_1}(f,k)\times\frac{sin\left(2\pi f\tau_s\right)}{2\pi f\tau_s} \tag{22}$$

The coherence matrix of noise in the MVDR beamformer is modified with $\frac{sin(2\pi f\tau_s)}{2\pi f\tau_s}$ for cancelling interference and noise suppression.

4. Experiments and discussions

In this section, the author will perform an experiment to illustrate the effectiveness of the proposed method (CovMatNoi) in comparison with traditional MVDR beamformer, delay and sum (DAS) [23] beamformer, diagonal-loading beamformer (DL) [16] and coherence-based dual-microphone noise reduction (CohNoi) [24] while preserving clean speech data, saving the value of the beampattern at a specified direction while attenuating the surrounding noise field or the effect of different noisy acoustic factors. The experiment was conducted in a living room (3 × 4 × 4.5 (m)) under conditions with the existence of the sound of a television, a washing machine, a fan, and other different undetermined noise sources. The configuration of the experiment shown in **Figure 6**.

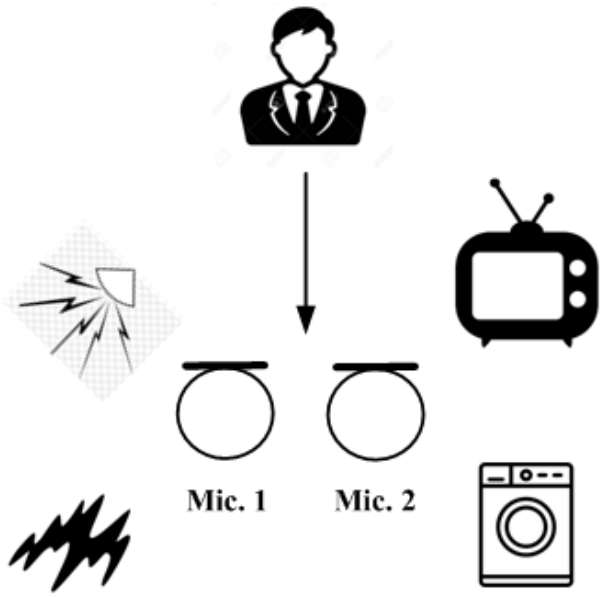

Figure 6. The illustrated experiment was performed in a realistic living room.

For recording the target speaker, the experiment applied DMA2, which has the compact size, the flexible configuration of microphone and easy integration with different signal processing techniques. The simulation with a speaker, who stands at a distance L = 3.5 (m) from the axis of DMA2, which was used for capturing the conversation of the desired talker, and the sound of surrounding noise. The author applied frequency sampling Fs = 16 kHz for storing the received array signal with overlap 50%. The preferred direction of the helpful signal is $\theta_s = 90\,(\text{deg})$. To compare the speech quality between MA signals, the processed signal using a traditional MVDR beamformer and CovMatNoi, this paper used an objective measurement [25] for calculating the obtained SNR. The speaker read the newspaper and DMA2 recorded the conversion in 3–5 min.

MA signals can be shown in **Figure 7** with waveform and spectrogram.

From our heuristic knowledge, $\alpha = 0.1$, $nFFT = 512$ are appropriate smoothing parameters for implementing the MVDR beamformer. The promising signal can be derived in **Figure 8** with original speech data and noise reduction.

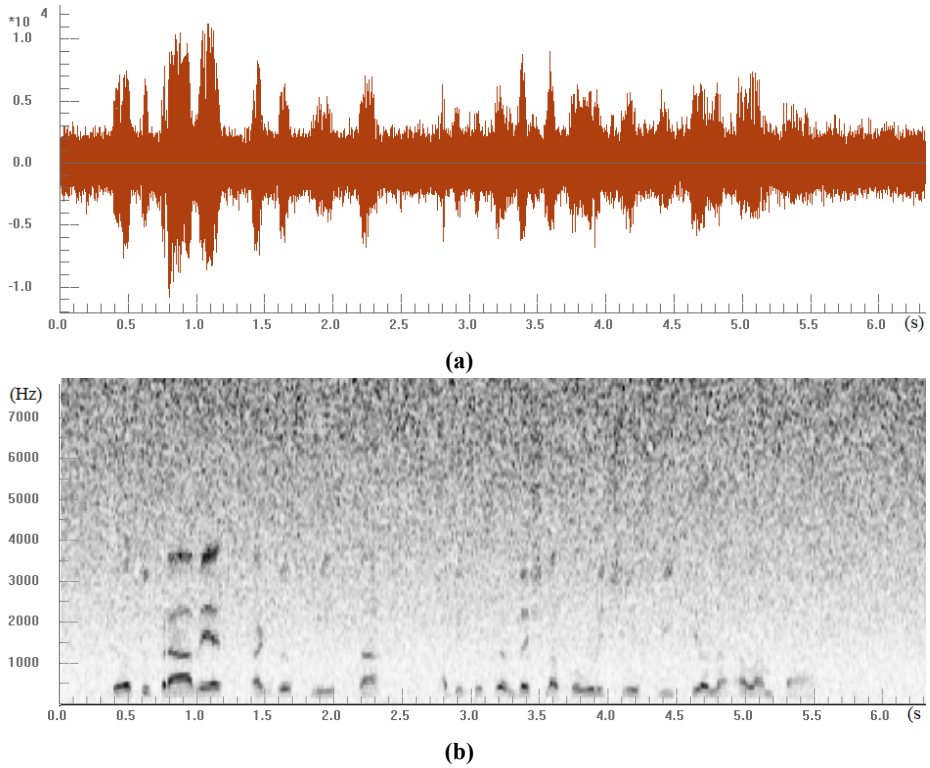

Figure 7. **(a)** The waveform; **(b)** Spectrogram of MA signals.

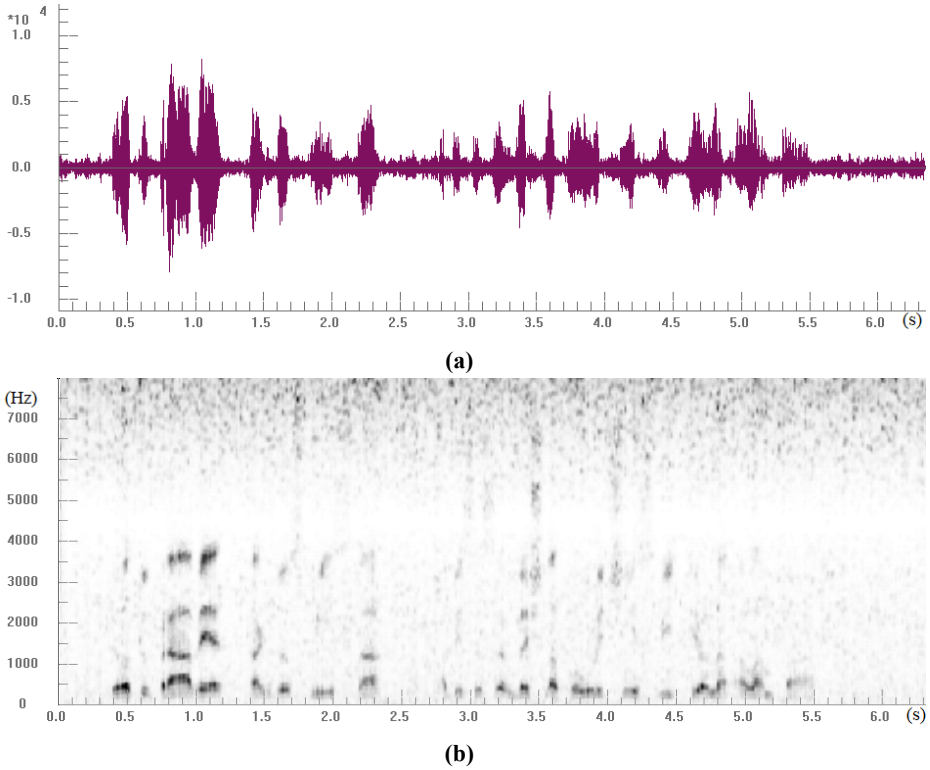

Figure 8. The MVDR beamformer's output signal: **(a)** Waveform; **(b)** Spectrogram.

However, due to the complexity of environmental recording, the error of sampling frequency, the internal electrical noise of acoustic equipment, the inaccurate computation of steering vector, the microphones mismatches, the different time of starting capturing between microphones, the movement of speaker, the displacement of microphones, the performance of MVDR beamformer is usually degraded. The information about the

transferring function or observed MA signals, which contain the speech component, corrupted the beamformer's evaluation. The speech component was distorted, and a large musical noise component degraded the intelligibility and the speech quality. The author's proposed method has fixed this drawback by reducing speech distortion and increasing speech quality and speech intelligibility in terms of SNR.

The promising signal, which is obtained by applying CovMatNoi, is given by **Figure 9**.

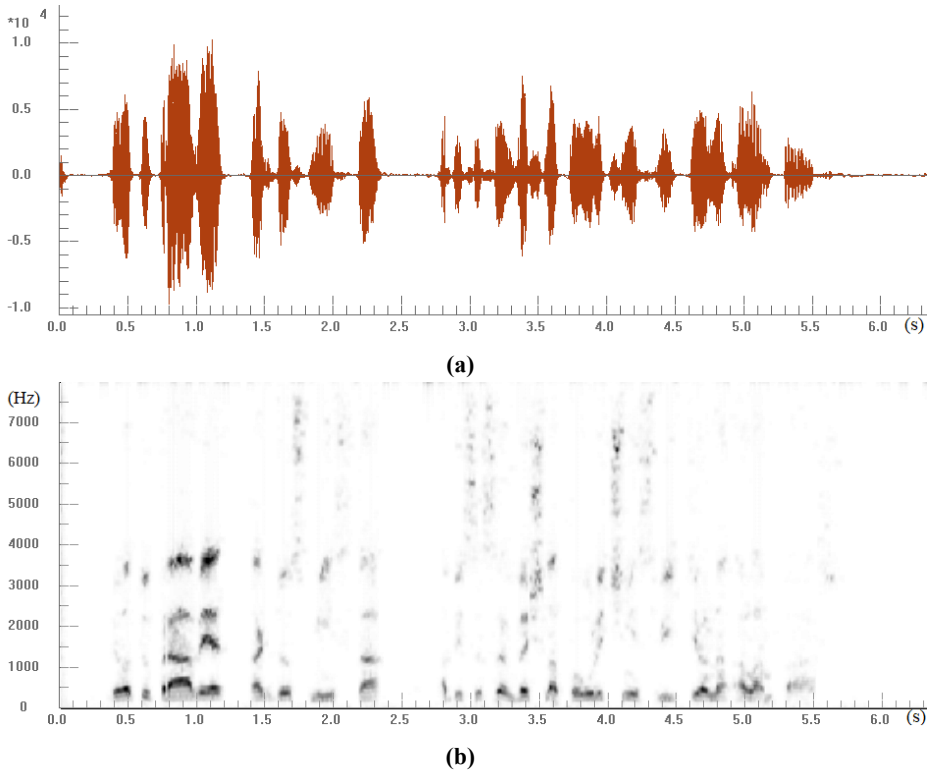

Figure 9. The promising signal by applying CovMatNoi: **(a)** Waveform; **(b)** Spectrogram.

Figure 10 compared the energy between MA signals, the processed signals by implementing a traditional MVDR beamformer, and the author's suggested technique. The perspective results are demonstrated by **Figure 10**, where speech distortion was reduced to 5 dB, the background noise level was removed to 15.2 dB, and speech quality in terms of SNR was improved from 12.1 to 12.9 dB, as in **Table 1**.

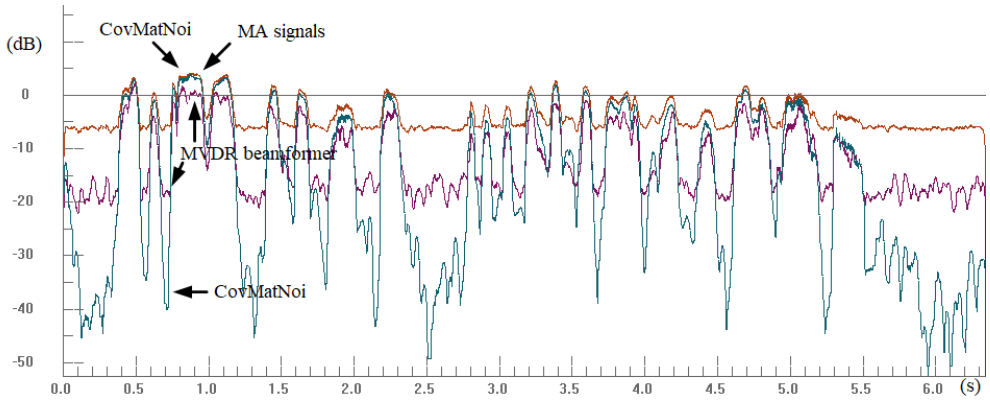

Figure 10. The curve of energy between MA signals and processed signals by the MVDR beamformer and CovMatNoi.

Table 1. The comparison signal-to-noise ratio (dB).

Method estimation	MA signals	DAS beamformer	CohNoi	DL	MVDR beamformer	CovMatNoi
NIST STNR	8.0	11.3	14.6	20.3	18.2	30.3
WADA SNR	9.2	12.1	13.2	17.0	17.6	30.5

In addition, the author performed with the DAS beamformer, DL, and CohNoi. The DAS beamformer attempts to steer the beampattern in a specified direction while containing large noise components at low frequencies. CohNoi does not rely on constrained criteria of minimizing noise power, so the advantage of noise reduction is not like the MVDR beamformer. DL is suitable for adverse noise fields, but without precise steering vector, this method cannot properly outperform in real-life situations.

The promising results by applying CovMatNoi have many advantages in comparison to conventional MVDR beamformers. From the obtained simulated results, this article can derive some assessments such as:

(1) The MA beamformer has the ability to concern the steerable pattern to the specified location. MVDR beamformer has suppressed the surrounding noise field while attempting to preserve the clean speech data with high quality of measured acoustic metrics.

(2) In practical speech applications, due to unwanted reasons, such as microphone mismatches, the error of setting the time record and frequency sampling, the displacement of configured microphone geometry, the movement of the talker during conversation, the existence of several complicated noise sources, and the evaluation of the beamformer, it is usually decreased. Musical noise, large ambient noise, or distortion of clean speech cause an unsatisfied listener.

(3) In a rapidly changing recording scenario, the determined time delay cannot be calculated by applying the PHAT-GCC formulation. So, the author suggested using an additive measurement about the value of the pattern of delay-and-sum beamformer for more precisely computing the exact time delay, which is implemented for representing the steering vector. Besides, the author's proposed method calculated the coherence of the noise field, which is based on the received array signals, the coherence between MA signals, and the temporal SNR. In addition, the formulation of noise coherence is multiplied by $\frac{sin(2\pi f\tau_s)}{2\pi f\tau_s}$ for suppressing the noise component at low frequency and increasing the robustness of the MVDR beamformer.

(4) The advantage of the described method is adaptively computing the time delay and coherence of noise for increasing the MVDR beamformer's performance. The numerical simulation has confirmed that speech distortion was reduced to 5 dB, the musical noise or ambient noise field was suppressed to 15.2 dB, and the overall speech quality of the beamformer's output signal increased from 12.1 to 12.9 dB.

(5) The properties of the author's suggested technique are low computation and easy implementation, which is very suitable for realistic acoustic equipment. The presented method can be applied to various speech applications for addressing different complicated problems, such as, dereverberation, automatic speech recognition and speech enhancement.

(6) The limitation of the author's proposed method is the speech leakage when calculating the coherence of noise. This problem can degrade effectiveness, and there is still the existence of little speech distortion at the MVDR beamformer's output signal.

5. Conclusion

In this article, the author proposed using an efficient method for enhancing the advantage of the MVDR beamformer under adverse and complex noise field conditions. The author's idea is based on GCC-PHAT and additive information about the beam pattern for exactly calculating the time delay τ_s between two microphones. Consequently, the coherence of noise filed $\Gamma_{N_1N_2}(f,k)$, $\Gamma_{N_2N_1}(f,k)$ was computed from the observed coherence of MA signals. Then these values were multiplied with $\frac{sin(\omega\tau_s)}{\omega\tau_s}$ for improving the effectiveness of noise suppression in the low-frequency domain. The conducted experiments confirmed the speech enhancement in reducing musical noise or residual noise component to 15.2 dB, suppressing the speech distortion to 5 dB, and increasing the obtained SNR from 12.1 to 12.9 dB. The presented technique has low computation and is easy to install into various types of speech applications. The mentioned method can be applied to a multi-channel system for resolving many complicated problems, such as, dereverberation, automatic speech recognition, and voice-controlled equipment. In the future, the author will study the characteristics of background noise for further improving this research work.

Author contributions: NTHC: conceptualization, methodology; QTT: formal analysis, data curation; PVH: supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This work was sponsored by Posts and Telecommunications Institute of Technology (PTIT), Hanoi, Vietnam.

Institutional review board statement: Not applicable.

Informed consent statement: Not applicable.

Data availability statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest: The authors declare no conflict of interest.

AI Use Statement: The authors declare that no artificial intelligence (AI) tools were used in the preparation of this manuscript. No AI tools were used for data analysis, interpretation, or generation of scientific content. All outputs were critically reviewed and edited by the authors. The authors take full responsibility for the integrity and accuracy of the work.

References

1. Nnonyelu CJ, Jiang M, Adamopoulou M, et al. A Machine- Learning -based approach to Direction-of-arrival Sectorization using Spherical Microphone Array. In: Proceedings of the 2024 IEEE 13rd Sensor Array

and Multichannel Signal Processing Workshop (SAM); 8–11 July 2024; Corvallis, OR, USA. pp. 1–5. doi: 10.1109/SAM60225.2024.10636592

2. Itzhak G, Doclo S, Cohen I. Joint Optimization of Microphone Array Geometry and Region-of-Interest Beamforming with Sparse Circular Sector Arrays. In: Proceedings of the 2024 18th International Workshop on Acoustic Signal Enhancement (IWAENC); 9–12 September 2024; Aalborg, Denmark. pp. 135–139. doi: 10.1109/IWAENC61483.2024.10694616
3. Pal S, Yadav SK, Kumar A, et al. Study of Direction of Arrival Estimation with a Differential Microphone Array using MVDR Algorithm. In: Proceedings of the 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT); 24–28 June 2024; Kamand, India. pp. 1–7. doi: 10.1109/ICCCNT61001.2024.10724601
4. Middelberg W, Doclo S. Bias Analysis of Spatial Coherence-Based RTF Vector Estimation for Acoustic Sensor Networks in a Diffuse Sound Field. In: Proceedings of the 2022 International Workshop on Acoustic Signal Enhancement (IWAENC); 5–8 September 2022; Bamberg, Germany. pp. 1–5. doi: 10.1109/IWAENC53105.2022.9914715
5. Zhang F, Pan C, Benesty J, et al. Simplified Maximum SNR Beamformers with Spatial Coherence Matrix Modeling. In: Proceedings of the 2023 31st European Signal Processing Conference (EUSIPCO); 4–8 September 2023; Helsinki, Finland. pp. 6–10. doi: 10.23919/EUSIPCO58844.2023.10289966
6. Zhang X, Jia Y, Jiang J, et al. A Low Frequency Source of Transformer Localization Method Using the Combination of CLEAN algorithm based on Spatial Source Coherence and Multi-signal Classification Algorithm. In: Proceedings of the 2024 9th Asia Conference on Power and Electrical Engineering (ACPEE); 11–13 April 2024; Shanghai, China. pp. 2057–2061. doi: 10.1109/ACPEE60788.2024.10532312
7. Zhao K, Luo X, Jin J, et al. Design of Robust Differential Beamformers with Microphone Arrays of Arbitrary Planar Geometry. In: Proceedings of the ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 6–11 April 2025; Hyderabad, India. pp. 1–5. doi: 10.1109/ICASSP49660.2025.10888941
8. Cao F, Shi J. Geometric optimization design of a five-element microphone array in sound source localization. In: Proceedings of the 2026 11th International Conference on Intelligent Computing and Signal Processing (ICSP); 17–19 April 2026; Hefei, China. pp. 103–107. doi: 10.1109/ICSP69961.2026.11540641
9. Pan C, Chen J, Benesty J. An Approach to Microphone Array Geometry Transform. IEEE Transactions on Audio, Speech and Language Processing. 2025; 33: 869–882. doi: 10.1109/TASLPRO.2025.3539025
10. As'ad H, Bouchard M, Kamkar-Parsi H. Robust Minimum Variance Distortionless Response Beamformer based on Target Activity Detection in Binaural Hearing Aid Applications. In: Proceedings of the 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP); 11–14 November 2019; Ottawa, ON, Canada. pp. 1–5. doi: 10.1109/GlobalSIP45357.2019.8969387
11. Yamaoka K, Ono N, Makino S, et al. Time-frequency-bin-wise Switching of Minimum Variance Distortionless Response Beamformer for Underdetermined Situations. In: Proceedings of the ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 12–17 May 2019; Brighton, UK. pp. 7908–7912. doi: 10.1109/ICASSP.2019.8683528
12. Ali R, Van Waterschoot T, Moonen M. Integration of a Priori and Estimated Constraints Into an MVDR Beamformer for Speech Enhancement. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2019; 27(12): 2288–2300. doi: 10.1109/TASLP.2019.2946086
13. Lagacé PO, Ferland F, Grondin F. Ego-Noise Reduction of a Mobile Robot Using Noise Spatial Covariance Matrix Learning and Minimum Variance Distortionless Response. In: Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 1–5 October 2023; Detroit, MI, USA. pp. 3533–3538. doi: 10.1109/IROS55552.2023.10342193
14. Yadav S, Pal S, Kumar A, et al. Study of MVDR Beamformer for a single Acoustic Vector Sensor. In: Proceedings of the 2023 International Symposium on Ocean Technology (SYMPOL); 13–15 December 2023; Kochi, India. pp. 1–6. doi: 10.1109/SYMPOL59195.2023.10455004
15. Song A. White Noise Array Gain for Minimum Variance Distortionless Response Beamforming With Fractional Lower Order Covariance. IEEE Access. 2018; 6: 71581–71591. doi: 10.1109/ACCESS.2018.2881206
16. Chaudhari K, Sutaone M, Bartakke P. Adaptive Diagonal Loading of MVDR Beamformer For Sustainable Performance In Noisy Conditions. In: Proceedings of the 2020 IEEE Region 10 Symposium (TENSYMP); 5–7 June 2020; Dhaka, Bangladesh. pp. 1144–1147. doi: 10.1109/TENSYMP50017.2020.9230850

17. Wang Y, Xu H, Wang B, et al. Broadband Spatial Spectrum Estimation Based on Space-Time Minimum Variance Distortionless Response and Frequency Derivative Constraints. IEEE Access. 2023; 11: 27955–27962. doi: 10.1109/ACCESS.2023.3258978
18. Erdim S, Buck JR, Gravelle C, et al. Doubly Adaptive Covariance Matrix Taper Universal Beamformer. In: Proceedings of the OCEANS 2022, Hampton Roads; 17–20 October 2022; Hampton Roads, VA, USA. pp. 1–6. doi: 10.1109/OCEANS47191.2022.9977261
19. Huang Y, Zhou M, Vorobyov SA. New Designs on MVDR Robust Adaptive Beamforming Based on Optimal Steering Vector Estimation. IEEE Transactions on Signal Processing. 2019; 67(14): 3624–3638. doi: 10.1109/TSP.2019.2918997
20. Wang J, Qian X, Pan Z, et al. GCC-PHAT with Speech-oriented Attention for Robotic Sound Source Localization. In: Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA); 30 May–5 June 2021; Xi'an, China. pp. 5876–5883. doi: 10.1109/ICRA48506.2021.9561885
21. Yousefian N, Kokkinakis K, Loizou PC. A coherence-based algorithm for noise reduction in dual-microphone applications. In: Proceedings of the 2010 18th European Signal Processing Conference; 23–27 August 2010; Aalborg, Denmark. pp. 1904–1908. Available online: https://www.eurasip.org/Proceedings/Eusipco/Eusipco2010/Contents/papers/1569292239.pdf
22. Erdim S, Buck JR, Berg CJ. Covariance Matrix Tapered Beamformer That Is Universal Over Notch Width. In: Proceedings of the 2022 56th Asilomar Conference on Signals, Systems, and Computers; 31 October–November 2022; Pacific Grove, CA, USA. pp. 1413–1418. doi: 10.1109/IEEECONF56349.2022.10051929
23. Huang GL, Shen CC. Delay-Multiply-and-Sum Beamforming with Transmit Minimum-Variance Estimation in Multi-Angle Plane-Wave Imaging. In: Proceedings of the 2022 IEEE International Ultrasonics Symposium (IUS); 10–13 October 2022; Venice, Italy. pp. 1–4. doi: 10.1109/IUS54386.2022.9958294
24. Chen X, Saleem N, Irfan M, et al. Coherence based Dual Microphone Source Separation in Low SNR Noisy Environments. In: Proceedings of the 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS); 6–8 June 2018; Singapore. pp. 865–869. doi: 10.1109/ICIS.2018.8466495
25. Kim C, Stern RM. Robust signal-to-noise ratio estimation based on waveform amplitude distribution analysis. In: Proceedings of the Interspeech 2008; 22–26 September 2008; Brisbane Australi. pp. 2598–2601. doi: 10.21437/Interspeech.2008-644