

An improved MVDR beamformer's noise reduction based on post - filtering

Nguyen Thi Huyen Chau¹, Quan Trong The², Nguyen Minh Thang³

¹Faculty of Information Technology, Thang Long University, Hanoi, Vietnam

²Faculty of Information Security, Posts and Telecommunications Institute of Technologies (PTIT), Hanoi, Vietnam

³Faculty of Information Technology, Ho Chi Minh City University of Economics and Finance, Ho Chi Minh City, Vietnam

Article Info

Article history:

Received May 22, 2025

Revised Jul 27, 2026

Accepted Aug 19, 2026

Keywords:

Microphone array beamforming

Minimum variance distortionless

Noise reduction

Post-filtering

Signal-to-noise ratio

ABSTRACT

Microphone-array beamforming is widely used to enhance target speech in noisy acoustic environments by exploiting spatial information from multiple sensors. The minimum variance distortionless response (MVDR) beamformer can effectively attenuate interference while preserving a signal arriving from a specified direction; however, its performance may deteriorate in practical recording conditions because of microphone mismatch, array displacement, steering-vector estimation errors, sampling-rate variations, and residual environmental noise. This study proposes a post-filtering technique to improve the output of an MVDR beamformer by combining Wiener-based spectral attenuation with spatial information obtained from the phase difference between two microphone signals. The proposed filter incorporates direction-of-arrival information to construct an additional spectral gain for suppressing residual noise while retaining the desired speech component. The method was evaluated using a dual-microphone array in a realistic indoor environment containing background and interfering noise. Experimental results show that the proposed approach increases the signal-to-noise ratio (SNR) from 10.8 dB to 12.7 dB and reduces the residual noise level to approximately 11.2 dB. NIST STNR and WADA SNR measurements further demonstrate substantial improvement compared with conventional MVDR processing. The results confirm the potential of spatially informed post-filtering for robust multi-channel speech enhancement.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Quan Trong The

Faculty of Information Security

Posts and Telecommunications Institute of Technologies (PTIT)

Hanoi, Vietnam

Email: theqt@ptit.edu.vn

1. INTRODUCTION

In real-life scenario, automatic speech recognition are significantly degraded by the sound of transport vehicle, television, washing machine, mechanical equipment, aircraft with several types of noise, non-directional noise, coherent, incoherent, diffuse noise field or diffuse noise field. Consequently, these factors has seriously corrupts the speech intelligibility, different objective acoustic measurement, the perceptual metric listener and speech quality of the recorded signals. Therefore, the essential in almost acoustic speech applications is addressing the problem of noise reduction and improving the speech quality. To deal with this task, speech enhancement or separation is usually implemented as a front-end to remove annoying and complex noise. Single-channel approach often based on subtraction method, which uses the criteria about stationary

noise. Therefore, in several non-stationary noise or in the complex recording situation with non-directional noise, diffuse noise field, single-channel algorithm causes the speech distortion of desired target speaker. To avoid this drawback, microphone array (MA) [1]-[3] technology can be used to achieve advantage of noise suppression, dereverberant and speech enhancement simultaneously with perspective results without speech distortion or musical noise. MA technology exploits the designed MA geometry to steer frequency response at specified direction while removing signal from other directions.

With the advantage of high spatial diversity, high directivity factor and the ability of incorporating single-channel approach, MA technology installed into multi-channel system, such as: smartphone, acoustic instrument, hearing aid, cochlear implant, human-machine interface, teleconference system to improving the speech quality of processed signal. The fixed beamformer with delay-and-sum (DAS) beamformer use assumed impinging incident of speaker to form the weights, which are applied to MA signals. Adaptive beamformer incorporate the prior information, the surrounding acoustic properties, the transfer function, the designed array configuration, the characteristics of background noise to obtain optimum beamformer's coefficient in preserving the desired talker while eliminating interference, background noise. The most useful representations of adaptive beamformer is differential microphone (DIF) [4]-[7], generalized sidelobe canceller (GSC) [8]-[11], linearly constrained minimum variance (LCMV) [12]-[15], minimum variance distortionless response (MVDR) [16]-[23].

MVDR beamformer has capability of concerning the steerable beampattern at sound source, interest useful speech. However, the performance often interfered by complex and adverse environment. There are many efforts attempt to increase the overall evaluation under realistic scenario.

Yamaoka *et al.* [16] introducing exploiting MVDR method for calculating spectral estimation based on the Kronecker product to achieve better evaluation under various recording scenarios. This article exploited the spectral estimation filters for achieving the more robust of speech enhancement than the traditional beamformer in realistic experiment.

In Tuladhar and Buck [17], time-frequency-bin-wise switching (TFS), which served as efficient voice activity detection (VAD) to switch implemented beamformer for processing array signals. The advantage of this approach is suppressing several particular interferences, non-directional noise sources by using MVDR beamformer. The numerical simulation has shown the improvement of speech enhancement in speech quality, perceptual metric listener underdetermined situations.

As'ad *et al.* [18] proposed the unit circle rectification (UCR) method for satisfying a property of the ensemble MVDR beamformer. This technique incorporated the sample matrix inversion with MVDR beamformer to attenuating the effect of limited snapshots. Consequently, UCR satisfied the criteria of MVDR beamformer ensemble. The simulated results has shown the effectiveness of UCR in reducing the white surrounding noise field, interference than diagonal loading. The demonstrated experiment show the advantage of the author's suggested technique in improving the speech quality of MVDR beamformer's output signal.

Lagac *et al.* [19] proposed selection between monaural and binaural MVDR beamformer at each considered time-frequency bin. The decision depends on an accurate target activity detection system, which based on a binaural beamforming system for cancelling target source. The described approach applied binaural beamforming as speech presence probability in demanding of blocking the target speech component. The simulation results show the robustness of calculating the transfer function, related steering vector of impinging angle between existing algorithms.

Park *et al.* [20], the authors proposed using principal component analysis (PCA) to adaptively adjust covariance matrix to reduce ego-noise by applying MVDR beamformer. The promising result has confirmed the effectiveness of reducing the word error rate by 55% and enhancing the signal-to-distortion ratio by up to 10 dB under various types of recording scenarios.

Son *et al.* [21] used beamspace approximation, which use a combination of the spatial fast-Fourier transform (FFT) to obtain linear-logarithmic complexity to enhance constructing and inverting the array covariance matrix. The conducted experiments studied the MVDR beamformer's tradeoffs between load computation and signal-plus-noise ratio (SINR) to examine the superiority for large array system in realistic environment.

Zhang *et al.* [22] calculated new frequency estimation with considering a variable parameter for computing the augmented covariance matrix to resolves a significant degradation of MVDR beamformer's evaluation. The improved performance is derived through computer simulations.

Jiang *et al.* [23], addressed the problem of source separation from the noisy mixture by using deep learning MVDR framework with recurrent neural networks for resolving the problem of matrix inversion and

eigenvalue decomposition. The promising result is removing the residual noise field while keeping the target speech undistorted. The numerical results have confirmed the superiority with several metrics and automatic speech recognition under complex situations.

Martin [24] presented multistage forward/backward for reducing the complexity of sample covariance matrix's (SCM) inversion. The obtained advantage of this method is making the original complex-value of SCM progressively become real-valued block diagonal matrix. The numerical simulation demonstrates that the suggested technique own high computational robustness and speech quality in comparison with existing MVDR beamformer in extracting the desired target speaker at specified location while minimizing the total noise power.

In practical acoustic environments, the performance of existing beamforming techniques can be significantly affected by the unpredictable characteristics of the surrounding noise and various interfering factors. In particular, changes in microphone-array geometry, inaccuracies in the estimation of the desired steering vector, sampling-frequency errors, variations in microphone sensitivity, speaker movement during speech, adverse acoustic conditions, and microphone mismatch may reduce the effectiveness of the MVDR beamformer. These practical imperfections can result in residual or musical noise, insufficient noise suppression, and speech distortion, ultimately leading to a decline in the perceived and measured quality of the enhanced speech. Consequently, the existing approaches may not adequately address all of these challenges under real-world operating conditions. To overcome these limitations, this study proposes an efficient post-filtering approach designed to further suppress residual background and musical noise after MVDR processing. The proposed method aims to enhance objective speech-quality measures, particularly STNR and speech intelligibility, thereby improving the robustness of speech enhancement in realistic acoustic environments.

2. MVDR BEAMFORMER

In this section, the author describes the principal working of MVDR beamformer in the frequency-domain. MVDR beamformer based on the constrained criteria of minimizing the total output noise power at the output while preserving the original speech component at specified direction without speech distortion [16]-[19]. MVDR beamformer owns the high spatial diversity at certain sound source.

In general case, the author uses dual - microphone system (DMA2) to illustrate the principal working of MVDR beamformer. At the current frame k , current frequency f , $\omega = 2\pi f$, the observed microphone array signals $Y_1(j\omega, k)$, $Y_2(j\omega, k)$ in the short-time Fourier transform can be represented as the following equations,

$$Y_1(j\omega, k) = S(j\omega, k)e^{j\Phi_s} + N_1(j\omega, k) \quad (1)$$

$$Y_2(j\omega, k) = S(j\omega, k)e^{-j\Phi_s} + N_2(j\omega, k) \quad (2)$$

where $\Phi_s = \pi f \tau_0 \cos(\theta_s)$, $\tau_0 = \frac{d}{c}$, θ_s is the incident angle of interest useful signal to the axis of DMA2, d is the range between two mounted microphones, c is sound speed propagation in the air and $c = 343(\frac{m}{s})$. $N_1(j\omega, k)$, $N_2(j\omega, k)$ means the background noise, interference and competitive speaker, or non - directional noise, $S(j\omega, k)$ is the desired target data.

With the definition, $\mathbf{Y}(j\omega, k) = [Y_1(j\omega, k) \ Y_2(j\omega, k)]^T$, $\mathbf{N}(j\omega, k) = [N_1(j\omega, k) \ N_2(j\omega, k)]^T$, $\mathbf{h}(j\omega, \theta_s) = [e^{j\Phi_s} \ e^{-j\Phi_s}]^T$, the (1), (2) can be rewritten in the following way,

$$\mathbf{Y}(j\omega, k) = S(j\omega, k)\mathbf{h}(j\omega, \theta_s) + \mathbf{N}(j\omega, k) \quad (3)$$

The essential problem of speech enhancement is finding an optimum coefficient $\mathbf{W}(j\omega, k)$ to obtain the final signal $\hat{S}(j\omega, k)$ approximately the original clean speech data.

$$\hat{S}(j\omega, k) = \mathbf{W}^H(j\omega, k) * \mathbf{Y}(j\omega, k) \quad (4)$$

where H is conjugate transposition. The mathematical formulation of MVDR beamformer can be expressed as,

$$\begin{aligned} \min_{\mathbf{W}(j\omega, k, \theta_s)} \quad & \mathbf{W}^H(j\omega, k, \theta_s) \Phi_{NN}(j\omega, k) \mathbf{W}(j\omega, k, \theta_s) \\ \text{s.t.} \quad & \mathbf{W}^H(j\omega, k, \theta_s) \mathbf{h}(j\omega, \theta_s) = 1 \end{aligned} \quad (5)$$

with $\Phi_{NN}(j\omega, k) = E\{\mathbf{N}^H(j\omega, k)\mathbf{N}(j\omega, k)\}$ mean the captured noisy covariance matrix.

Hence, the MVDR beamformer's filter coefficients can be derived as,

$$\mathbf{W}(j\omega, k, \theta_s) = \frac{\Phi_{NN}^{-1}(j\omega, k)\mathbf{h}(j\omega, \theta_s)}{\mathbf{h}^H(j\omega, \theta_s)\Phi_{NN}^{-1}(j\omega, k)\mathbf{h}(j\omega, \theta_s)} \quad (6)$$

However, in realistic recording scenario, the information about noise is not always available, therefore, we use the observed microphone array instead of. So, the final optimum MVDR beamformer's weights can be computed as,

$$\mathbf{W}_{MVDR}(j\omega, k, \theta_s) = \frac{\Phi_{YY}^{-1}(j\omega, k)\mathbf{h}(j\omega, \theta_s)}{\mathbf{h}^H(j\omega, \theta_s)\Phi_{YY}^{-1}(j\omega, k)\mathbf{h}(j\omega, \theta_s)} \quad (7)$$

where $\Phi_{YY}(j\omega, k) = E\{\mathbf{Y}^H(j\omega, k)\mathbf{Y}(j\omega, k)\}$ is the spectral covariance matrix of captured MA signals.

The auto - cross power spectral densities (PSD) between two received MA signals are recursively calculated by equations,

$$P_{Y_i Y_i}(j\omega, k) = (1 - \alpha)P_{Y_i Y_i}(j\omega, k - 1) + \alpha Y_i^*(j\omega, k)X_i(j\omega, k) \quad (8)$$

$$P_{Y_i Y_j}(j\omega, k) = (1 - \alpha)P_{Y_i Y_j}(j\omega, k - 1) + \alpha Y_i^*(j\omega, k)Y_j(j\omega, k) \quad (9)$$

The output signal of MVDR beamformer is derived as,

$$y_{MVDR}(j\omega, k) = \mathbf{W}_{MVDR}^H(j\omega, k, \theta_s) \times \mathbf{Y}(j\omega, k) \quad (10)$$

Due to many reasons, such as, the complex and annoying environment, the error of MA distribution, the error of sampling rate, the difference of MA quality, or imprecise estimation of preferred steering vector, the performance of MVDR beamformer usually degraded. The requirement of an efficient signal processing technique is an essential part of numerous speech applications. Consequently, in the next section, the author proposed an effective post - filtering for suppressing the musical noise, remained noise at the MVDR beamformer's output to increase the speech quality, speech intelligibility.

3. THE AUTHOR'S PROPOSED POST - FILTERING

The author's approach is modifying the Wiener filter for suppressing the noise level at MVDR beamformer's output signal. The tradition Wiener filter $H_{Wiener}(j\omega, k)$ can be calculated as the following equation,

$$H_{Wiener}(j\omega, k) = \frac{\sigma_s^2(j\omega, k)}{\sigma_s^2(j\omega, k) + \sigma_n^2(j\omega, k)} \quad (11)$$

where $\sigma_s^2(j\omega, k) = \frac{1}{\mathbf{h}^H(j\omega, \theta_s)\Phi_{YY}^{-1}(j\omega, k)\mathbf{h}(j\omega, \theta_s)}$ is variance of original clean speech component at the specified direction θ_s and $\sigma_n^2(j\omega, k)$ is the covariance of noise at output of MVDR beamformer and can be computed by minimum statistic (MS) [25].

Due to the complex and rapidly changed recording scenario, the effectiveness of Wiener filter often degraded and there is still exists musical noise, remained noise component at the MVDR beamformer's output. The author proposed modifying Wiener filtering with incorporating the prior information of the phase difference between two captured microphone array signals $Y_1(j\omega, k)$, $Y_2(j\omega, k)$.

Ellis [26], the prior information about the direction of arrival (DOA) of interest useful speech component allows deriving the spectral masking $G_k^{BM}(j\omega) = 1 - \exp(j\phi_{12,k}^{norm}(j\omega))$ and $G_k^{DSB}(j\omega) = 0.5 + 0.5 * \exp(j\phi_{12,k}^{norm}(j\omega))$ and where $\phi_{12,k}^{norm}(j\omega) = \frac{\phi_{12,k} - c}{f \cdot d}$, $\phi_{12,k}$ is the phase difference between $Y_1(j\omega, k)$, $Y_2(j\omega, k)$.

The modified Wiener filter is given by,

$$\hat{H}_{Wiener}(j\omega, k) = \beta \frac{\sigma_s^2(j\omega, k)}{\sigma_s^2(j\omega, k) + \sigma_{nMVDR}^2(j\omega, k)} + (1 - \beta) \frac{G_k^{DSB}(j\omega)}{G_k^{DSB}(j\omega) + G_k^{BM}(j\omega)} \quad (12)$$

where β is an appropriate parameter, which in $\{0..1\}$. In the experiment, the value of $\beta = 0.35$ was based on our heuristic knowledge.

The processed signal by applying modified Wiener filter is computed as,

$$\hat{y}_{MVDR}(j\omega, k) = y_{MVDR}(j\omega, k) * \hat{H}_{Wiener}(j\omega, k) \quad (13)$$

By adding the spectral masking gain, which based on the prior of DOA of interest useful signal, the modified Wiener filter is suitable under annoying acoustic environment to removing the remained background noise to increase the speech quality, perceptual metric listener.

The appealing properties of described technique is utilizing the gain function, which based on preferred DOA of interest helpful signal. The method exploits the gain of superdirective beamformer and blocking matrix for determining the gain function. In practical scenario, the success of Wiener filter usually degraded due to the complicated problem of estimating covariance of speech $\sigma_s^2(j\omega, k)$ and $\sigma_n^2(j\omega, k)$. These necessities parameters significantly depends on the microphone sensitivities, the homogeneity of recording situation, the time of simultaneous capture, the sampling frequency. Hence, the estimation of covariance of speech and noise usually not exact under complex environment with presence several undetermined noise sources. This additive gain function allows increasing the capability of suppressing the remained noise level. The exploiting of additive post-filtering will improve the overall performance of MVDR beamformer. In the next section, the author performed an experiment in realistic situation to verify the effectiveness of the mentioned post - filtering.

4. EXPERIMENTS AND DISCUSSION

In this section, the author will illustrate the advantage of the author's proposed method in enhancing the speech quality and reducing the remained noise level. The experiment was performed in living room with the existence of washing machine, interference and annoying background noise. The author uses dual - microphone system (DMA2) to capture the original speech component of talker. The range between two mounted microphones is $d = 5(cm)$, the distance from talker to the axis of DMA2 is $L = 3(m)$, the DOA of speaker is $\theta_s = 90(deg)$.

An objective measurement was implemented for calculating the signal-to-noise ratio (SNR) for comparing the speech quality between the captured MA signals, the processed signal by tradition MVDR beamformer (dioMVDR) and the author's suggested post - filtering (arsgpF). For capturing the talker's speech component, the sampling frequency $F_s = 16kHz$, the overlap 50% were used.

The observed MA signals is shown in Figure 1. For performing MVDR beamformer, these parameters $nFFT = 512$, the smoothing parameter $\alpha = 0.1$. The obtained MVDR beamformer's output signal can be derived in Figure 2.

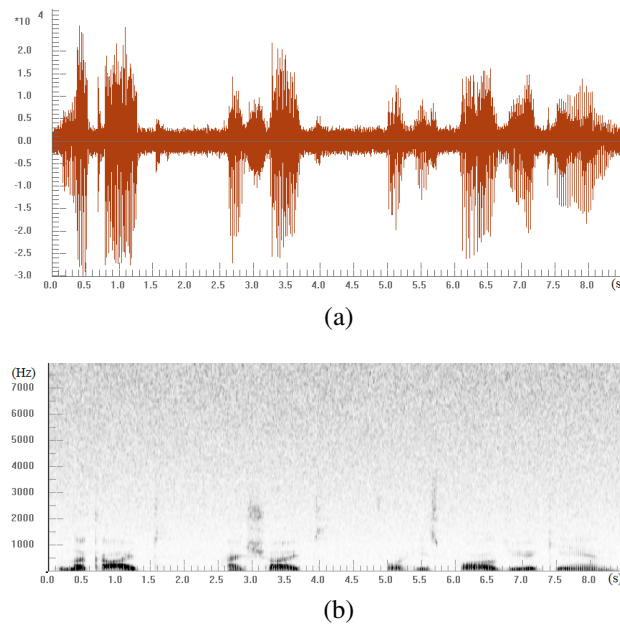


Figure 1. The waveform (a) and spectrogram (b) of captured MA signals

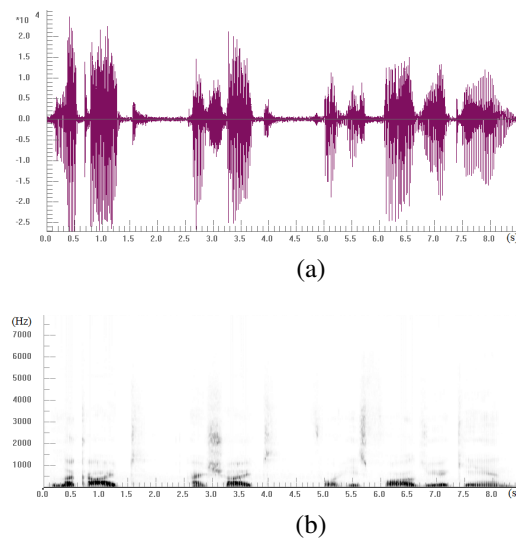


Figure 2. The promising signal of MVDR beamformer's performance: waveform (a) and spectrogram (b)

In practical and acoustically complex environments, the performance of an MVDR beamformer can be significantly degraded by various uncertainties. These may include microphone mismatch, inaccurate estimation of the desired steering vector, changes in the microphone array configuration, variations in microphone sensitivity, and sampling-rate errors. Such imperfections can result in speech distortion and residual noise in the beamformer output. To address these issues, the author introduces an additive post-filtering approach designed to further attenuate residual noise while maintaining the integrity of the desired speech signal. After applying the proposed arsgpF method, the resulting enhanced signal is presented in Figure 3.

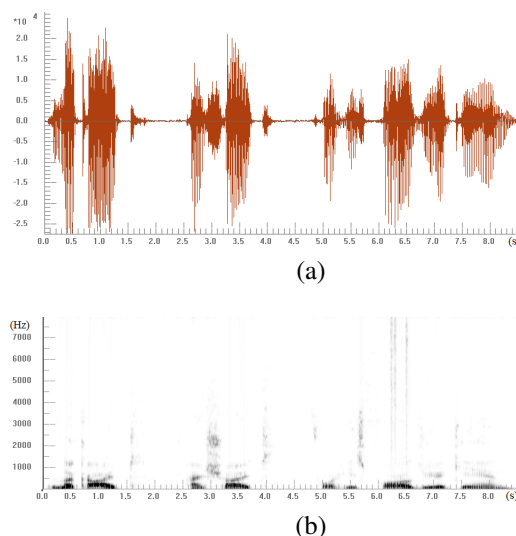


Figure 3. The perspective results by applying arsgpF: waveform (a) and spectrogram (b)

The compared energy between MA signals, the processed signal by dioMVDR, arsgpF is shown in Figure 4, arsgpF allows removing the noise level to 11.2 dB and consequently, the signal-to-noise was increased from 10.8 to 12.7 dB. The SNR (dB) as shown in Table 1.

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

Variable	Speed (rpm)	Power (kW)	
NIST STNR	6.8	13.2	24.0
WADA SNR	1.5	9.4	22.1

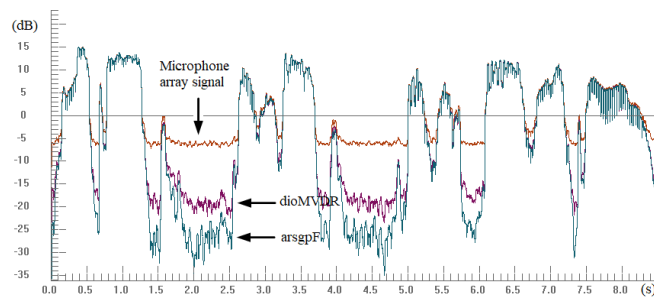


Figure 4. The comparison of energy between MA signals and processed signals by dioMVDR, arsgpF

From the conducted experiments, this paper can conclude some opinions,

- The microphone-array (MA) beamformer can spatially attenuate ambient noise, interfering signals, and unwanted sources arriving from directions other than the desired source, while maintaining the speech component originating from the specified direction. In this experiment, the target speaker was positioned at 90° . As illustrated, the MVDR beamformer effectively retained the desired speech while reducing the contribution of background noise.
- Nevertheless, the performance of the MVDR beamformer may deteriorate under practical acoustic conditions. Several factors can contribute to this degradation, including microphone displacement, internal electronic noise from the recording equipment, room reverberation, and multiple time-varying noise sources. These imperfections may introduce speech distortion and substantial residual noise, resulting in degraded perceptual speech quality. As observed, a considerable amount of residual noise remains in the beamformer output.
- Post-filtering provides an effective approach for further improving the beamformer output. Its primary purpose is to attenuate the remaining noise while retaining the characteristics of the desired speech signal. In this work, the proposed post-filter is derived from the Wiener filtering principle and integrated with a spectral gain function. The gain function incorporates prior information regarding the DOA of the desired speech source. By combining the Wiener-based filtering mechanism with the directional gain function, the proposed approach provides improved noise suppression compared with the beamformer output alone.
- A key feature of the proposed method is the utilization of prior information about the preferred DOA of the target speaker together with the phase difference between signals captured by the microphone array. These characteristics are exploited to construct an effective post-filter capable of suppressing residual background noise while maintaining the integrity of the desired speech. Consequently, the proposed approach reduces ambient noise without introducing significant speech distortion. The results demonstrate the potential of the proposed post-filter for enhancing speech signals in adverse acoustic environments and indicate its applicability to other challenging noise-reduction scenarios.

5. CONCLUSION

This study investigated the degradation of MVDR beamformer performance in practical acoustic environments, where microphone mismatch, array-position errors, inaccuracies in steering-vector estimation, sampling-rate variations, reverberation, and time-varying interference can lead to residual noise and speech distortion. To address these limitations, a modified post-filtering scheme based on Wiener filtering and spatial information derived from the phase difference between microphone signals was developed and applied to the MVDR output. The proposed approach combines the statistical attenuation capability of the Wiener filter with a direction-dependent spectral gain, allowing residual acoustic components to be further suppressed while retaining the speech components associated with the desired DOA. Experimental evaluation using a dual-microphone array in a realistic indoor recording environment demonstrated the potential of the proposed approach for improving the quality of the beamformer output. The results show that the SNR increased from 10.8 dB for the conventional MVDR output to 12.7 dB after post-filtering, while the residual noise level was reduced to approximately 11.2 dB. The objective measurements obtained from the NIST STNR and WADA SNR evaluations likewise demonstrate a substantial improvement after the proposed post-filter was applied, confirming its ability to attenuate residual interference while preserving the principal speech component. An important advantage of the proposed method is that it functions as a complementary processing stage rather than replacing the spatial

filtering mechanism of the MVDR beamformer; consequently, it can potentially be incorporated into existing multi-channel speech-processing systems with limited modification to the beamforming structure. However, the effectiveness of the method remains dependent on the accuracy of the available spatial information and the estimation of the spectral and covariance-related parameters, and its performance may therefore vary under rapidly changing acoustic conditions, significant target-speaker movement, or inaccurate target-direction estimation. Further investigation using different microphone-array configurations, reverberation levels, interference conditions, moving speakers, and larger speech datasets would be valuable for assessing the robustness and generalization of the proposed technique. Future research may also consider adaptive target-direction tracking, more reliable speech and noise statistical estimation, and automatic adjustment of the post-filter weighting parameter to improve operation in dynamic environments. Overall, the results indicate that the integration of spatially informed spectral post-filtering with MVDR beamforming provides a practical approach for mitigating residual noise and improving enhanced speech quality in realistic recording conditions, with potential applications in multi-channel speech enhancement, source separation, and automatic speech recognition.

REFERENCES

- [1] Y. Jia, B. Gray, and R. Vaughan, "Measurements of microphone array phase and amplitude behavior towards controllable beamforming," in *2020 IEEE SENSORS*, Rotterdam, Netherlands, 2020, pp. 1-4, doi: 10.1109/SENSORS47125.2020.9278799.
- [2] G. Huang, J. Benesty, J. Chen, and I. Cohen, "Robust and steerable kronecker product differential beamforming with rectangular microphone arrays," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 211-215, doi: 10.1109/ICASSP40776.2020.9052988.
- [3] V. Tourbabin, I. Malka, and E. Tzirkel-Hancock, "Performance of fixed in-car microphone array beamformer under variations in car noise," in *2017 Hands-free Speech Communications and Microphone Arrays (HSCMA)*, San Francisco, CA, USA, 2017, pp. 21-25, doi: 10.1109/HSCMA.2017.7895554.
- [4] K. Zhao, X. Luo, J. Jin, G. Huang, J. Chen, and J. Benesty, "Design of Robust Differential Beamformers with Microphone Arrays of Arbitrary Planar Geometry," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hyderabad, India, 2025, pp. 1-5, doi: 10.1109/ICASSP49660.2025.10888941.
- [5] X. Zhao, X. Luo, G. Huang, J. Chen, and J. Benesty, "Differential Beamforming with Null Constraints for Spherical Microphone Arrays," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seoul, Korea, Republic of, 2024, pp. 776-780, doi: 10.1109/ICASSP48485.2024.10446768.
- [6] X. Luo, J. Jin, G. Huang, J. Chen, and J. Benesty, "Design of Steerable Linear Differential Microphone Arrays With Omnidirectional and Bidirectional Sensors," *IEEE Signal Processing Letters*, vol. 30, pp. 463-467, 2023, doi: 10.1109/LSP.2023.3267969.
- [7] F. M. Ribas *et al.*, "Minimum Variance Generalized Sidelobe Canceller and Eigenspace-based Generalized Sidelobe Canceller Beamformers Combined with Frost Postfilter for Medical Ultrasound Imaging," in *2021 IEEE UFFC Latin America Ultrasonics Symposium (LAUS)*, Gainesville, FL, USA, 2021, pp. 1-4, doi: 10.1109/LAUS53676.2021.9639164.
- [8] J. Wang, F. Yang, J. Guo, and J. Yang, "Robust Adaptation Control for Generalized Sidelobe Canceller with Time-Varying Gaussian Source Model," in *2023 31st European Signal Processing Conference (EUSIPCO)*, Helsinki, Finland, 2023, pp. 16-20, doi: 10.23919/EUSIPCO58844.2023.10289801.
- [9] S. Dai, M. Li, Q. H. Abbasi, and M. A. Imran, "A Fast Blocking Matrix Generating Algorithm for Generalized Sidelobe Canceller Beamformer in High Speed Rail Like Scenario," *IEEE Sensors Journal*, vol. 21, no. 14, pp. 15775-15783, Jul. 15, 2021, doi: 10.1109/JSEN.2020.3002699.
- [10] S. S. Priyanka and T. K. Kumar, "GSC Beamforming using Different Adaptive Algorithms for Speech Enhancement," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, 2019, pp. 1-6, doi: 10.1109/ICCCNT45670.2019.8944415.
- [11] A. Frank, A. Ben-Kish, and I. Cohen, "Constant-Beamwidth Linearly Constrained Minimum Variance Beamformer," in *2022 30th European Signal Processing Conference (EUSIPCO)*, Belgrade, Serbia, 2022, pp. 50-54, doi: 10.23919/EUSIPCO55093.2022.9909899.
- [12] A. M. Narayanan and A. Bertrand, "Group-Utility Metric for Efficient Sensor Selection and Removal in LCMV Beamformers," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 4950-4954, doi: 10.1109/ICASSP40776.2020.9053467.
- [13] V. W. Neo, E. d'Olné, A. H. Moore, and P. A. Naylor, "Fixed Beamformer Design Using Polynomial Eigenvalue Decomposition," in *2022 International Workshop on Acoustic Signal Enhancement (IWAENC)*, Bamberg, Germany, 2022, pp. 1-5, doi: 10.1109/IWAENC53105.2022.9914799.
- [14] J. Zhang, J. Du, and L. -R. Dai, "Sensor Selection for Relative Acoustic Transfer Function Steered Linearly-Constrained Beamformers," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1220-1232, 2021, doi: 10.1109/TASLP.2021.3064399.
- [15] X. Wang, J. Benesty, G. Huang, and J. Chen, "A Minimum Variance Distortionless Response Spectral Estimator with Kronecker Product Filters," in *2022 30th European Signal Processing Conference (EUSIPCO)*, Belgrade, Serbia, 2022, pp. 2261-2265, doi: 10.23919/EUSIPCO55093.2022.9909584.
- [16] K. Yamaoka, N. Ono, S. Makino, and T. Yamada, "Time-frequency-bin-wise switching of minimum variance distortionless response beamformer for underdetermined situations," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 2019, pp. 7908-7912, doi: 10.1109/ICASSP.2019.8683528.
- [17] S. R. Tuladhar and J. R. Buck, "Unit Circle Rectification of the Minimum Variance Distortionless Response Beamformer," *IEEE Journal of Oceanic Engineering*, vol. 45, no. 2, pp. 500-510, Apr. 2020, doi: 10.1109/JOE.2018.2876584.

- [18] H. As'ad, M. Bouchard, and H. Kamkar-Parsi, "Robust Minimum Variance Distortionless Response Beamformer based on Target Activity Detection in Binaural Hearing Aid Applications," in *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Ottawa, ON, Canada, 2019, pp. 1-5, doi: 10.1109/GlobalSIP45357.2019.8969387.
- [19] P.-O. Lagacé, F. Ferland, and F. Grondin, "Ego-Noise Reduction of a Mobile Robot Using Noise Spatial Covariance Matrix Learning and Minimum Variance Distortionless Response," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Detroit, MI, USA, 2023, pp. 3533-3538, doi: 10.1109/IROS55552.2023.10342193.
- [20] J. -J. Park, G. Rodriguez-Martinez, M. M. Bolding, R. S. Westafer, and V. S. Sathe, "Investigation of Scalable Approximation of the Minimum Variance Distortionless Response Beamformer," in *2024 IEEE International Symposium on Phased Array Systems and Technology (ARRAY)*, Boston, MA, USA, 2024, pp. 1-6, doi: 10.1109/ARRAY58370.2024.10880424.
- [21] K. J. Son, G. S. An, Y. H. Kim, and T. -G. Chang, "Improved "Augmented MVDR Spectrum-Based Frequency Estimation Algorithm" for Unbalanced Power Systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-5, 2021, Art no. 9003205, doi: 10.1109/TIM.2021.3079565.
- [22] Z. Zhang, Y. Xu, M. Yu, S. -X. Zhang, L. Chen, and D. Yu, "ADL-MVDR: All Deep Learning MVDR Beamformer for Target Speech Separation," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, ON, Canada, 2021, pp. 6089-6093, doi: 10.1109/ICASSP39728.2021.9413594.
- [23] S. Jiang, S. Liu, M. Jin, and Z. Lin, "Efficient MVDR Beamformer Employing Multistage Forward/Backward Averaging Scheme," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 61, no. 2, pp. 3963-3976, Apr. 2025, doi: 10.1109/TAES.2024.3494758.
- [24] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 504-512, Jul. 2001, doi: 10.1109/89.928915.
- [25] S. M. Kim, "Hearing Aid Speech Enhancement Using Phase Difference-Controlled Dual-Microphone Generalized Sidelobe Canceller," *IEEE Access*, vol. 7, pp. 130663-130671, 2019, doi: 10.1109/ACCESS.2019.2940047.
- [26] D. P. W. Ellis, "SNREVAL: Speech Noise Ratio and Voice Activity Detection evaluation tools," Laboratory for the Recognition and Organization of Speech and Audio (LabROSA), Columbia University, 2007. [Online]. Available: <https://labrosa.ee.columbia.edu/projects/snreval/>

BIOGRAPHIES OF AUTHORS



Nguyen Thi Huyen Chau    graduated with a Bachelor's degree in Information Technology from the University of Engineering and Technology, Vietnam National University, Hanoi, and a Master's and PhD in Computer Science from Ecole Nationale Supérieure de Mécanique et d'Aérotechnique – ENSMA, France. Dr. Chau has been the Head of the Faculty of Information Technology at Thang Long University (TLU) since 2023. She has also been the Head of the Artificial Intelligence Laboratory, TLU, since 2019. She has supervised 7 master's students, authored or co-authored 25 scientific works, participated in 12 scientific research projects, including 1 completed ministerial-level project and 1 state-level project that is being approved. Her research interests include soft computing, machine learning, intelligent systems, schedulability analysis. She has been leading projects applying artificial intelligence (AI) for screening depression disorder symptoms, integrating AI in music teaching, developing virtual assistants for ATM and CRM devices, developing tools to evaluate learning outcomes and personalize student learning paths at universities. She can be contacted at email huyenchau@thanglong.edu.vn.



Quan Trong The    received his Ph.D. in mathematics from the National Research University ITMO in 2021. He works as a lecturer at Faculty of Information Security of Posts and Telecommunications Institute of Technology, Hanoi, Vietnam. His research interests are related to cybersecurity, speech enhancement, image processing, machine learning. He has authored or coauthored more than 30 publications: 27 proceedings and 03 journals. He can be contacted at email: theqt@ptit.edu.vn.



Nguyen Minh Thang    He received Engineering of Information Technology in 2017 from Ho Chi Minh City University of Technology, Master of Information Technology in 2022 from University of Information Technology, Vietnam National University at Ho Chi Minh City. His key areas of research include information security and applied AI for cybersecurity. He can be contacted at email: thangnm@uef.edu.vn.