

# A Speech Enhancement Based on Minimum Variance Distortionless Response Beamformer in Adverse Recording Scenario

QuanTrong The

Post and Telecommunication Institute of Technology, Hanoi, Vietnam

## Abstract

In many hands-free speech communication systems such as surveillance device, voice controlled – equipment, teleconference system, mobile phones, smart - home, hearing aid, the captured speech signal often degraded due to the existence of third - party talker, unwanted interference, noise, such that the speech enhancement technique are required to enhance the speech quality, speech intelligibility and perceptual listener. The microphone array (MA) technology has been commonly applied to various types of acoustic applications, because of the exploiting the priori information of MA distribution, the designed configuration of geometry, the characteristics of surrounding environment to obtain better noise reduction and speech enhancement at the same time. By using MA, the problem of sound source localization, estimation of direction of arrival of interest useful signal, the steered beampattern toward the desired target speaker, the suppressing of total all background noise are easy resolved. In MA beamforming technique, Minimum Variance Distortionless Response (MVDR) beamformer is one of the most useful methods for extracting the desired talker at certain direction while eliminating all background noise with speech distortion. However, the ideal performance of MVDR beamformer is usually corrupted in realistic recording situations due to the microphone mismatches, the different microphone sensitivities, the displacement of MA configuration. In this article, the author proposed an accurate calculation of steering vector to improve MVDR beamformer's evaluation in complex and annoying environment. The numerical result has confirmed the effectiveness of the author's suggested method in increasing the speech quality from 10.8 (dB) to 12.2 (dB) and reducing the speech distortion to 5.2 (dB). The superiority of this method can be integrated into a multi-channel system to achieve sustainable signal processing.

## Keywords

Microphone array, minimum variance distortionless response, beamforming, speech enhancement, steering vector, speech quality, the signal-to-noise ratio.


## 1. Introduction


Nowadays, the using of hearing aids, smartphones, voice – controlled devices, cellular communication, teleconferencing equipment, smart-vehicle present new complex challenges

---

ITTAP'2024: 4th International Workshop on Information Technologies: Theoretical and Applied Problems, October 23–25, 2024, Ternopil, Ukraine, Opole, Poland

\*Corresponding author:

 theqt@ptit.edu.vn

 0000-0002-2456-9558



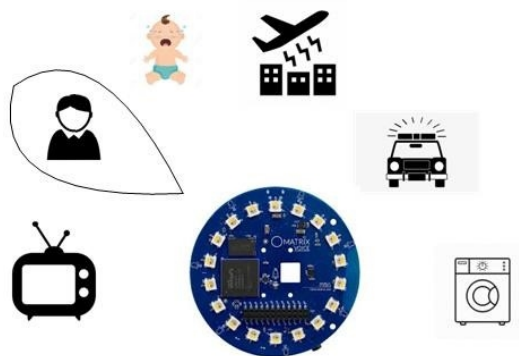
© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

for speech enhancement algorithms in recovering the original clean speech component while suppressing background noise without speech distortion. The single channel, which is often based on spectral subtraction, owns the simplicity of performance and easily installed in almost acoustic instruments. However, this approach can perfectly work in stationary environments. In the non-stationary situations and rapidly changing characteristics, this method causes speech distortion or corrupted signal. Therefore, MA technology has been installed for overcoming this drawback. MA beamforming uses the priori information to achieve better noise reduction and speech enhancement at the same time.



**Figure 1:** The complex and annoying environment effect on the human life.

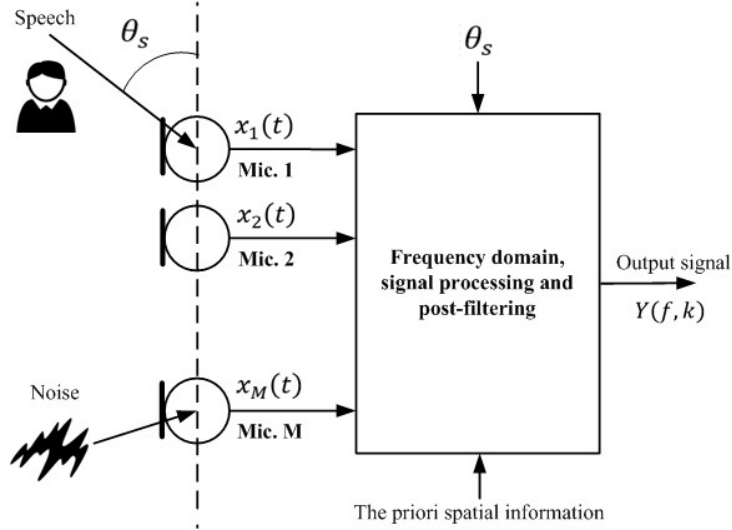
The beamforming exploits the designed MA distribution, the direction of arrival of speech, the properties of surrounding noise environments, the constrained formulation of signal processing, the additive spectral mask, post – filtering, preprocessing to process the captured signals for obtaining the highest quality of speech enhancement. The MA beamformer can be categorized into two groups: the fixed beamformer with delay and sum DAS [1-4], the adaptive beamformer with differential microphone array DIF [5-8], generalized sidelobe canceller [9-12], minimum variance distortionless response [13-16], linear constrained minimum variance LCMV [16-20].



**Figure 2:** The implementation of microphone array beamforming

The use of adaptive beamforming technique depends on the particular configuration and the purpose of the signal processing system on saving the target talker while minimizing the effect

of background noise. MVDR may be in the commonly utilized beamforming designs for numerous speech applications, which minimize the noise power at the output while maintaining the original speech component. In more complicated situations, where several sources, non-directional noise, competing talkers, it was applied to separate sound sources without speech distortion.



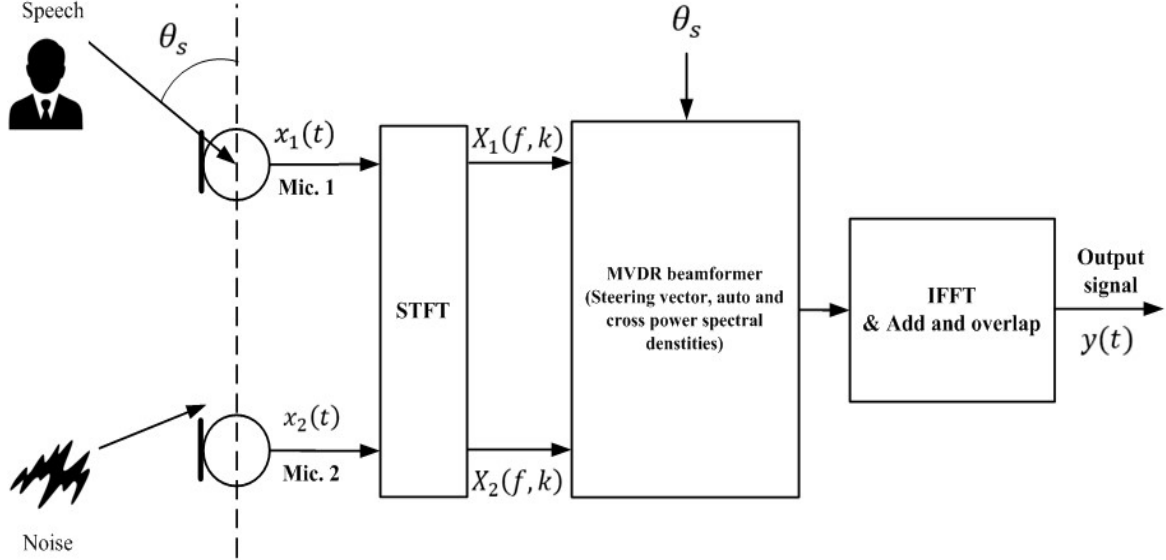
**Figure 3:** The implementation of MA beamforming to extract the desired speech

Although theoretical the MVDR provides optimal spatial diversity, when the interfering source locations are constantly changing, the problem of speech enhancement is more challenging due to the computation and source localization. Assuming the direction of arrival (DoA) is known, the MVDR beamformer estimates the desired signal while minimizing the variance of noise component at the output of beamformer. In practice, the DoA of the target desired signal is not calculated exactly, which significantly degrades the performance of. A lot of devoted research has been done to improve the robustness of MVDR beamformer by extending the region where the source can be detected, determined according to the characteristics of environment. Nevertheless, even assuming perfect source localization (SSL), the fact that microphone sensors may have distinct, impulsive response, different directional gain and another level of uncertainty that the MVDR beamformer is not able handle all scenarios well.

Because of the microphone mismatches, the difference between microphone gain – sensitivities, the inaccurate MA distribution, the error of computing about the DoA and the undetermined characteristics of background environments significantly decrease the MVDR beamformer’s performance. In this paper, the author proposed a new approach for adaptively estimating the steering vector to enhance speech enhancement.

The rest of this paper is organized as follows. The first section introduces the problem of speech enhancement by using MA technology. The second section describes the principal evaluation of MVDR beamforming in the frequency domain. The author’s suggested method was presented in section III. The experiments were conducted in section IV with perspective numerical results. Section V concludes and the author’s work in the future.

## 2. Minimum Variance Distortionless Response Beamformer



**Figure 4:** The scheme of MVDR beamformer in frequency - domain

In this section, the author describes the principle of working of MA in the frequency domain. In general case, the dual - microphone system (DMA2) was illustrated to understand the problem of signal processing by MVDR beamformer.

At the current frequency - frame  $(f, k)$ , the observed MA signals on two microphones:  $X_1(f, k), X_2(f, k)$  can be presented as:

$$X_1(f, k) = S(f, k)e^{j\phi_s} + N_1(f, k) \quad (1)$$

$$X_2(f, k) = S(f, k)e^{-j\phi_s} + N_2(f, k) \quad (2)$$

Where  $S(f, k)$  is the original clean speech signal,  $N_1(f, k), N_2(f, k)$  is the additive noise, which significantly degrades on the speech quality.  $\Phi_s = \pi f \tau_0 \cos(\theta_s)$ ,  $\theta_s$  is the direction of arrival of interest useful signal relative to the axis of DMA2,  $\tau_0 = d/c$  is the time delay,  $d$  is the range between two mounted microphones,  $c = 343$  (m/s) is sound speed propagation in the air.

With the definition:  $\mathbf{X}(f, k) = [X_1(f, k) X_2(f, k)]^T$ ,  $\mathbf{N}(f, k) = [N_1(f, k) N_2(f, k)]^T$ ,  $\mathbf{D}_s(f, \theta_s) = [e^{j\phi_s} e^{-j\phi_s}]^T$ , these equations (1) - (2) can be rewritten as:

$$\mathbf{X}(f, k) = S(f, k)\mathbf{D}_s(f, \theta_s) + \mathbf{N}(f, k) \quad (3)$$

The requirement of speech enhancement is finding an optimum filter's coefficients  $\mathbf{W}(f, k)$ , which ensures the estimated signal  $S(f, k) = \mathbf{W}^H(f, k)\mathbf{X}(f, k)$  approximately the original clean speech.

MVDR beamformer based on the constrained criteria of minimum the total out noise power while preserving the speech component without speech distortion. The formulation of MVDR beamformer can be expressed as the following way:

$$\frac{\min}{W(f,k)} W^H(f,k) \Phi_{NN}(f,k) W(f,k) \text{ st } W^H(f,k) D_s(f, \theta_s) = 1 \quad (4)$$

With  $\Phi_{NN}(f, k)$  is the spectral covariance matrix of noise.

From the constrained problem of MVDR beamformer, the optimum coefficients were defined as:

$$W_{MVDR}(f, k) = \frac{\Phi_{NN}^{-1}(f, k) D_s(f, \theta_s)}{D_s^H(f, \theta_s) \Phi_{NN}^{-1}(f, k) D_s(f, \theta_s)} \quad (5)$$

Unfortunately, the priori information about noisy environment is not always available, so the captured MA signals were used instead of. Finally, the optimum MVDR beamformer's coefficients were derived as:

$$W_{MVDR}(f, k) = \frac{\Phi_{XX}^{-1}(f, k) D_s(f, \theta_s)}{D_s^H(f, \theta_s) \Phi_{XX}^{-1}(f, k) D_s(f, \theta_s)} \quad (6)$$

Where  $\Phi_{XX}^{-1}(f, k) = E\{X^H(f, k) X(f, k)\} = \begin{matrix} E\{|X_1(f, k)|^2\} & E\{X_1^i(f, k) X_2(f, k)\} \\ E\{X_2^i(f, k) X_1(f, k)\} & E\{|X_2(f, k)|^2\} \end{matrix}$

$(\cdot)^H$  is conjugate operator.

The auto - cross power spectral densities can be calculated as the recursive formulation:

$$P_{X_i X_i}(f, k) = \alpha P_{X_i X_i}(f, k-1) + (1-\alpha) X_i^i(f, k) X_i^{\square}(f, k) \quad (7)$$

$$P_{X_i X_j}(f, k) = \alpha P_{X_i X_j}(f, k) + (1-\alpha) X_i^i(f, k) X_j^{\square}(f, k) \quad (8)$$

Where  $\alpha$  is the smoothing parameter, which in the range  $\{0... 1\}$ .

In realistic recording situations, due to the microphone mismatches, the differences of microphone sensitivities, the error of estimation of preferred steering vector, the displacement of MA distribution, the imprecise of sampling frequency, the moving head of speaker, the overall MVDR beamformer's performance usually corrupted. In the next section, the author proposed a new method for determining accurate steering vectors, which seriously affects signal processing by beamforming technique.

### 3. The proposed method of estimating the steering vector

Steering vector  $\mathbf{D}_s(f, \theta_s)$  play an important role in MVDR beamformer. However, the preferred steering vector often changed in the frame with the presence of speech component. Therefore, the author proposed a recursive formulation for estimating steering vectors as:

$$\mathbf{D}_s(f, k, \theta_s) = \beta \mathbf{D}_s(f, k-1, \theta_s) + (1 - \beta) \mathbf{D}_s^{GEVD}(f, k) \quad (9)$$

With smoothing parameter  $\beta$  and the initial steering vector  $\mathbf{D}_s(f, 0, \theta_s) = \mathbf{D}_s(f, \theta_s)$ .

The steering vector in frame – speech  $\mathbf{D}^{GEVD}_s(f, k)$  is computed by generalized eigenvalue decomposition as the following way:

$$\mathbf{D}^{GEVD}_s(f, k) = \mathbf{R}_{NN}(f, k) P\{\mathbf{R}_{NN}^{-1}(f, k) \mathbf{R}_{SS}(f, k)\} \quad (10)$$

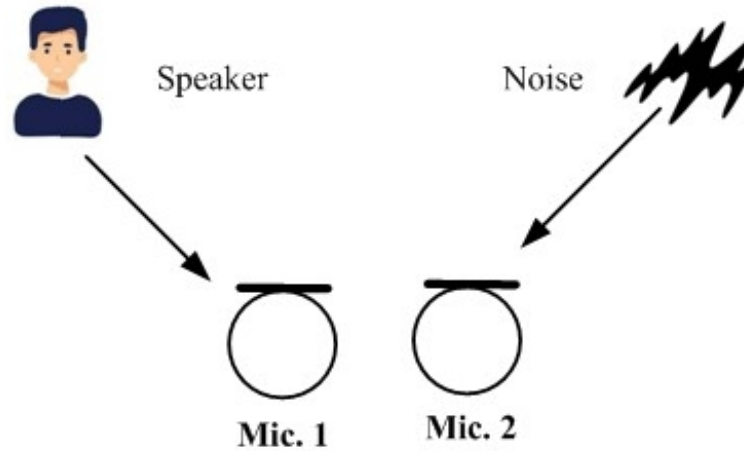
Where  $P\{\mathbf{R}_{NN}^{-1} \mathbf{R}_{SS}\}$  extracts the principal eigenvector by applying generalized eigenvalues decomposition to the spectral matrix of noise and speech.

In practical applications, the priori information of noise and clean speech is not always available. Hence, the author proposed using EM algorithm [22] for calculating the spectral mask, which according to the presence/absence of speech enhancement. Let  $M_s(f, k)$  denote the time – frequency mask for speech and  $M_n(f, k)$  represents the background noise probability. Then, the matrix  $\mathbf{R}_{SS}(f, k), \mathbf{R}_{NN}(f, k)$  yields as the following equations:

$$\begin{aligned} \mathbf{R}_{SS}(f, k) &= \frac{1}{\sum_k M_s(f, k)} \sum_k M_s(f, k) X(f, k), X^H(f, k) \quad (11) \\ \mathbf{R}_{NN}(f, k) &= \frac{1}{\sum_k M_n(f, k)} \sum_k M_n(f, k) X(f, k), X^H(f, k) \quad (12) \end{aligned}$$

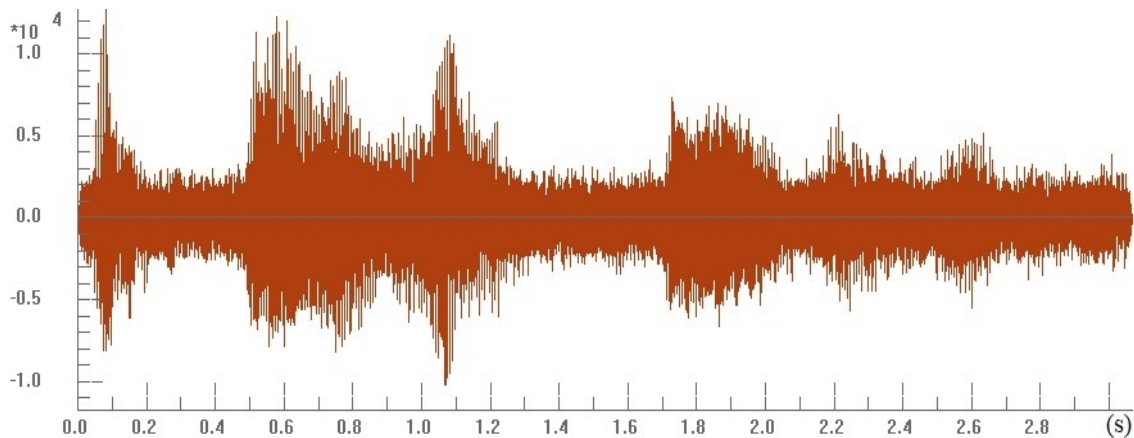
The appealing properties of the author's post – Filtering is tracking, updating the steering vector according to the speech presence probability in the frame – speech, which leads to enhance MVDR beamformer's performance. In the next section, the author demonstrates experiments to confirm the advantages of suggested technique.

## 4. Experiment results



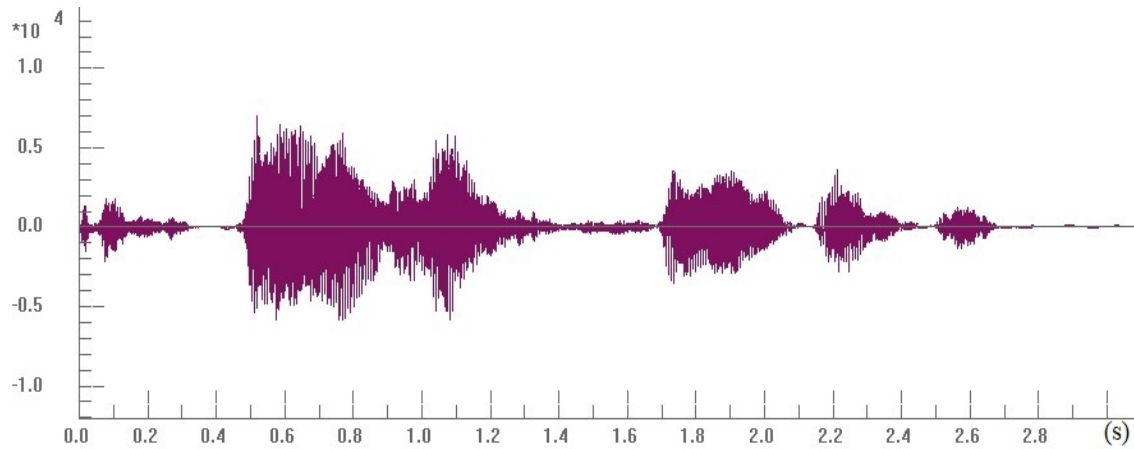
**Figure 5:** The demonstrated experiment with dual - microphone system

The purpose of the conducted experiment is to compare the effectiveness of the traditional MVDR beamformer (tMVbe) and the author suggested method (ausv) in increasing the speech quality and energy. An objective measurement of SNR [21] was used for computing the obtained processed signals. The experiment was performed in room dimensions about  $9.5 \times 8.0 \times 3.6$  m<sup>3</sup>. The dual – microphone array (DMA2) was used for capturing the original clean speech component, which degraded by noisy environment. The desired speaker stands at distance  $L = 3.5$  (m) to the axis of DMA2 and the direction of arrival of interest signal is  $\theta_s = 60^\circ$  in the condition of existence of noise, interference. The range between two mounted microphones is  $d = 5$ (cm). For recording the mixture of speech and noise, the sampling rate is set  $F_s = 16$  kHz , overlap 50%. The observed MA signals were depicted in Figure 6.



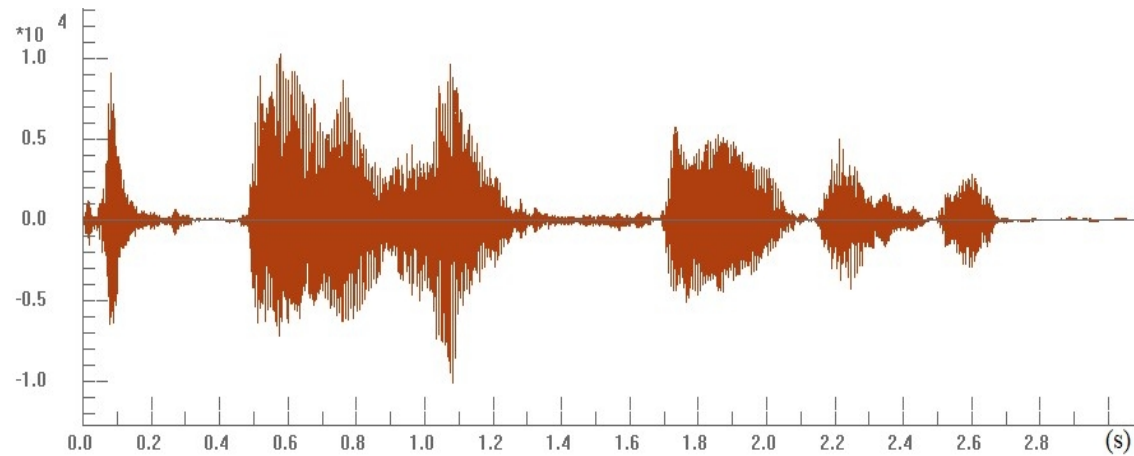
**Figure 6:** The waveform of the captured microphone array signals

$nFFT = 512$ , smoothing parameter  $\alpha = 0.1$  for calculating the auto – cross power spectral densities and  $\beta = 0.92$  to track the changing steering vector. After using the MVDR beamformer, the output signal was derived as Figure 7.



**Figure 7:** The waveform of processed signal by tMVbe

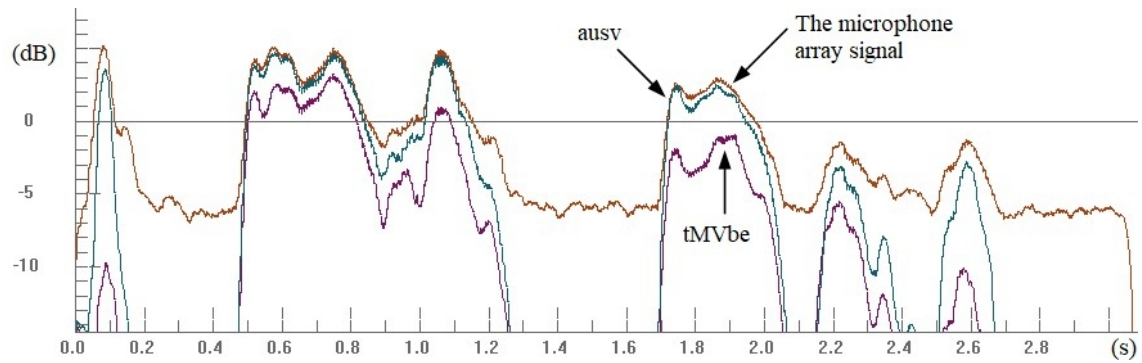
As a result, in the recording scenarios of presence of complex environments, the moving head of talker, the coherent difference sensitivities of microphone arrays, the error of estimation of preferred steering vector significantly affected on the MVDR beamformer's evaluation. Although the noise level was suppressed, the original speech component was not recovered.



**Figure 8:** The waveform of processed signal by ausv

A comparison of energy between these signals was illustrated in Figure 9.





**Figure 9:** The comparison energy between microphone array signals and the processed by tMVbe, ausv

By using the author's method, the steering vector was recursively updated frame - by - frame, which provides sustainable robustness beamforming for MVDR beamformer. The obtained result was shown in Figure 8.

Table 1 presents the receives SNR between the MA signals, the processed signal by tMVbe and ausv.

**Table 1**

The signal-to-noise ratio SNR (dB)

Method Estimation	Microphone array signals	tMVbe	ausv
NIST SNR	9.2	15.4	26.2
WADA SNR	7.2	16.1	28.3

As a result, the obtained SNR increased from 10.8 (dB) to 12.2 (dB) in comparison with tMVbe. The speech distortion decreased to 5.2 (dB). The updated steering vector allowed the high directional beampattern adaptively steered toward the sound source according to the rapidly changing of recording environment and ensured alleviation of the surroundings environment. The appealing property of the suggested method is tracking, updating and changing immediately filter's coefficients for extracting the desired target speaker. The experiment has presented the ability of the author's method in reducing speech distortion, improving the MVDR beamformer's performance in adverse environments.

## 5. Conclusion

In this contribution, the author proposed using a new method for computing exactly steering vector, which plays an important role in MVDR beamformer for forming a high directional beampattern towards the sound source while removing the background noise. The appealing property of the suggested method is overcome the heuristic drawback of MVDR beamformer is very sensitive with the direction of arrival of useful speech component, which often decreases the output signal's quality. The numerical result has verified the advantages of the suggested technique for accurately calculating MVDR beamformer's coefficients in complex environments to obtain the original speech component, reduce speech distortion to 5.2 (dB), increase the SNR

from 10.8 (dB) to 12.2 (dB). In the future, the author will investigate the characteristics of diffuse noise field and the effects of reverberation in living room to further enhance the above method.

## References

- [1] Chodingala P. K., Chaturvedi S. S., Patil A. T., Patil H. A. Robustness of DAS Beamformer Over MVDR for Replay Attack Detection On Voice Assistants//Proc 2022 IEEE International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, 2022, pp. 1-5, doi: 10.1109/SPCOM55316.2022.9840757.
- [2] Zeng Y., Hendriks R. C. Distributed Delay and Sum Beamformer for Speech Enhancement via Randomized Gossip. IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 1, pp. 260-273, Jan. 2014, doi: 10.1109/TASLP.2013.2290861.
- [3] Rakotoarisoa I., Fischer J., Valeau V., Marx D., Prax C., Brizzi L.E. Time-domain delay-and-sum beamforming for time-reversal detection of intermittent acoustic sources in flows. J. Acoust. Soc. Am. 136, 2675–2686 (2014). <https://doi.org/10.1121/1.4897402>.
- [4] António L., Ramos L., Holm S., Gudvangen S., Otterlei R. Delay-and-sum beamforming for direction of arrival estimation applied to gunshot acoustics //Proc Proceedings Volume 8019, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense X; 80190U (2011) <https://doi.org/10.1117/12.886833>.
- [5] Huang G., Cohen I., Benesty J., Chen J. Kronecker Product Beamforming with Multiple Differential Microphone Arrays // Proc 2020 IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM), Hangzhou, China, 2020, pp. 1-5, doi: 10.1109/SAM48682.2020.9104333.
- [6] Zhao X., Luo X., Huang G., Chen J., Benesty J. Differential Beamforming with Null Constraints for Spherical Microphone Arrays // Proc 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.
- [7] Luo X., Jin J., Huang G., Chen J., Benesty J. 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.
- [8] Wang X., Huang G., Cohen I., Benesty J., Chen J. Robust Steerable Differential Beamformers with Null Constraints for Concentric Circular Microphone Arrays. Proc ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 4465-4469, doi: 10.1109/ICASSP39728.2021.9414119.
- [9] The Q. T., Huy N. B., Anh P. T. Spectral Mask - Based Technique for Improving Generalized Sidelobe Canceller Beamformer's Evaluation. 2023 Seminar on Signal Processing, Saint Petersburg, Russian Federation, 2023, pp. 106-110, doi: 10.1109/IEEECONF60473.2023.10366094.
- [10] Wang J., Yang F., Guo J., Yang J. Robust Adaptation Control for Generalized Sidelobe Canceller with Time-Varying Gaussian Source Model // Proc 2023 31st European Signal Processing Conference (EUSIPCO), Helsinki, Finland, 2023, pp. 16-20, doi: 10.23919/EUSIPCO58844.2023.10289801.
- [11] Dai S., Li M., Abbasi Q. H., Imran M. A. 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, 15 July 2021, doi: 10.1109/JSEN.2020.3002699.

- [12] Middelberg W., Doclo S. Comparison of Generalized Sidelobe Canceller Structures Incorporating External Microphones for Joint Noise and Interferer Reduction. *Speech Communication; 14th ITG Conference*, online, 2021, pp. 1-5.
- [13] Shankar N., Küçük A., Reddy C. K. A., Bhat G. S., Panahi I. M. S. Influence of MVDR beamformer on a Speech Enhancement based Smartphone application for Hearing Aids// *Proc 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 2018*, pp. 417-420, doi: 10.1109/EMBC.2018.8512369.
- [14] Shankar N., Bhat G. S., Panahi I. M. S. Real-time dual-channel speech enhancement by VAD assisted MVDR beamformer for hearing aid applications using smartphone. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 2020*, pp. 952-955, doi: 10.1109/EMBC44109.2020.9175212.
- [15] Wang D., Bao C. Multi-channel Speech Enhancement Based on the MVDR Beamformer and Postfilter // *Proc 2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Macau, China, 2020*, pp. 1-5, doi: 10.1109/ICSPCC50002.2020.9259489.
- [16] Araki S., Ono N., Kinoshita K., Delcroix M. Meeting Recognition with Asynchronous Distributed Microphone Array Using Block-Wise Refinement of Mask-Based MVDR Beamformer // *Proc 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018*, pp. 5694-5698, doi: 10.1109/ICASSP.2018.8462458.
- [17] Hassani A., Bertrand A., Moonen M. LCMV beamforming with subspace projection for multi-speaker speech enhancement // *Proc 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 2016*, pp. 91-95, doi: 10.1109/ICASSP.2016.7471643.
- [18] Xiao J., Pu W., Luo Z. Q., Zhang T. Evaluation of the Penalized Inequality Constrained Minimum Variance Beamformer for Hearing Aids // *Proc 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018*, pp. 3344-3348, doi: 10.1109/ICASSP.2018.8462539.
- [19] Schreibman A., Barnov A., Gendelman A., Tzirkel E. RTF Based LCMV Beamformer with Multiple Reference Microphones // *Proc 2020 28th European Signal Processing Conference (EUSIPCO), Amsterdam, Netherlands, 2021*, pp. 181-185, doi: 10.23919/Eusipco47968.2020.9287468.
- [20] Chazan S. E., Goldberger J., Gannot S. DNN-Based Concurrent Speakers Detector and its Application to Speaker Extraction with LCMV Beamforming // *Proc 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018*, pp. 6712-6716, doi: 10.1109/ICASSP.2018.8462407.
- [21] <https://labrosa.ee.columbia.edu/projects/snreval/>.
- [22] Araki S., Okada M., Higuchi T., Ogawa A., Nakatani T. Spatial correlation model based observation vector clustering and MVDR beamforming for meeting recognition // *Proc 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 2016*, pp. 385-389, doi: 10.1109/ICASSP.2016.7471702.