Impact Factor 6.1

Journal of Cyber Security

ISSN:2096-1146

Scopus

DOI

Google Scholar

More Information
www.journalcybersecurity.com

Crossref doi Google Scholar scopus

An Improvement of MVDR Beamformer under Adverse Situation

Quan Trong The - ORCID: 0000-0002-2456-9598 - Lab Blockchain, Posts and Telecommunications Institute of Technologies (PTIT), Hanoi, Vietnam

Nguyen Thi Huyen Chau - ORCID: 0000-0003-4091-0271 - Thang Long University, Hanoi, Vietnam

Abstract

Nowadays, the use of microphone array beamformer has been widely commonly due to its convenience of steering the beampattern on the specified sound location and attenuating the background noise field. Microphone array beamforming own the capability of suppressing interference, third-party speakers and different adverse noise fields with high directivity index of extracting the clean speech data without speech distortion. Minimum Variance Distortionless Response beamformer bases on the constrained criteria of minimizing the total output noise power while saving the target talker by ensurign the beampattern equals one at certain direction. However, under realistic recording environments, due to the movement of speaker during conversations, the error of start time recording, the different sensitivities of microphones, the error of sampling frequency, the internal electrical acoustic equipments, the inaccurate distribution of microphones, the overall beamformer's performance often degraded. The unacceptable surrounding noise level or speech distortion corrupt the speech intelligibility of output signal. In the article, the author proposed two-stage - based method for adaptively updating the smoothing parameters for increasing the beamformer's evaluation in real-life scenarios. The numerical simulation has shown the improvement of reducing the speech distortion to 5 dB, removing the background noise level to 12.9 dB and increasing the speech quality in the term of signal-to-noise ratio from 5.3 to 8.8 dB.

I. Introduction

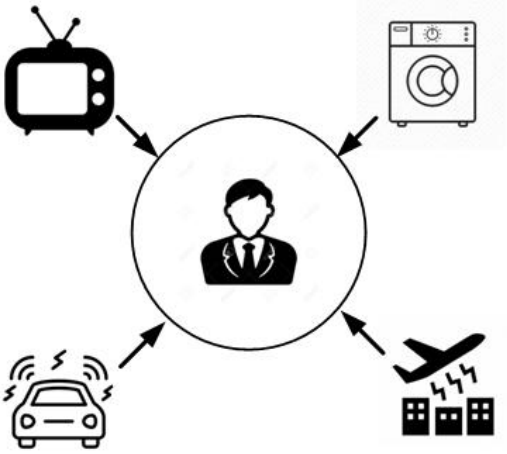


Figure 1: The adverse environment around human – life

Microphone array (MA) [1-2] beamforming with perspective noise reduction and speech enhancement is becoming more popular. The necessary of exploting MA technology is given by Figure 1.

Its highlight capability is concerning the pattern at certain location of desired target talker and significantly mitigating the background noise field. MA beamformer own high directivity index and ability of incorporating single-channel method, measured coherence between received array signals, the spatial prior characteristic of recorded environment for obtaining advantage of speech enhancement and noise suppression at the same time. The scheme of MA beamformer's working to recover the clean speech data from noisy mixture is described in Figure 2.

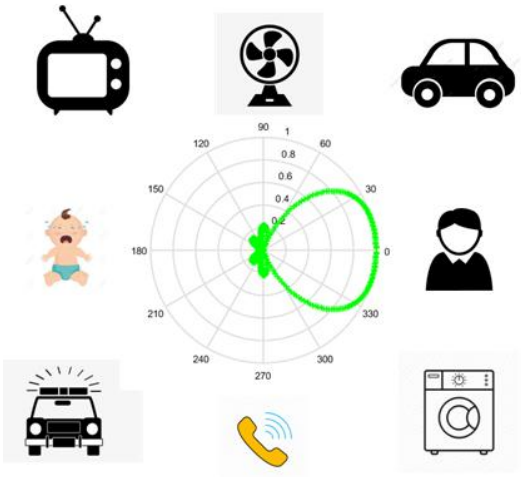


Figure 2: The steerable pattern at specified desired talker

The implementation of MA beamforming technique in the frequency domain is given by Figure 3.

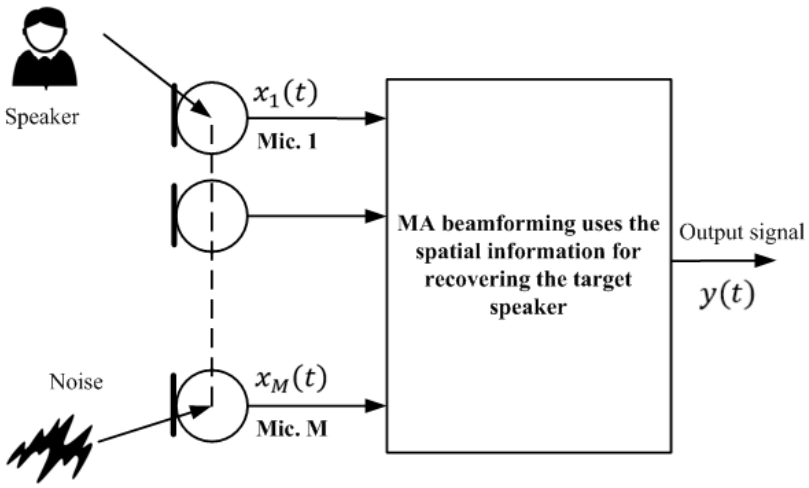


Figure 3: The implementation of microphone array beamforming with the aim of extracting the desired target speech component

Minimum Variance Distortionless Response (MVDR) beamformer is an efficient solution which has been integrated into various types of speech applications, such as, voice-controlled device, hearing aids, smart phone, television, smart intelligent vehicle, portable mobile phone, cochlear implant. MVDR technique based on the optimum criteria of attenuating surrounding noise field by steering the beampattern at target speaker. However, in realistic scenario, due to the error of sampling rate, the moving speaker during conversation,the existence of unwanted acoustic factors, the complex of environment of coherent/incoherent and diffuse noise field, the presence of different talker, the internal electrical noise, the overall MVDR beamformer's evaluation usually degraded.

In [3], Chaudhari K introduced using an adaptive loading factor, which estimated by considering actual snapshots for addressing the problem of poor suppression and larger deviations in beamformer's evaluation. The numerical simulation has confirmed the effectiveness of the proposed technique in improving the robust of MVDR beamformer.

Zhang Z et al [4] proposed applying matrix inversion and eigenvalue for training and verifying deep learning MVDR beamformer. The obtained experiments has demonstrated the advantage of improving several objective perceptual metric listener with perspective results.

Ali R et al [5] described an integration of MVDR beamformer and Linearly Constrained Minimum Variance (LCMV) for reducing the drawbacks when using priori spatial information, such as, relative transfer function, sound source location, environmental characteristics, the designed geometry of MA. The experimental results have shown the benefits of increasing the speech quality in the term of noise reduction and speech enhancement.

In this article, the author proposed an efficient technique for increasing the robust of MVDR beamformer with reducing speech distortion, suppressing surrounding noise level and improving speech quality of beamformer's output signal.

II. MVDR Beamformer

In this section, the author use dual-microphone system (DMA2) for illustrating the signal processing procedure of MVDR beamformer in the frequency-domain. DMA2 is the most useful configuration of MA, which has been installed into numerous speech applications for extracting the desired talker while suppressing interference and background noise. The scheme of MVDR beamformer is given by Figure 4.

With the assumed impinging angle θ_s to the axis of DMA2, at current considered frequency f and frame k, the received array signals $X_1(f,k), X_2(f,k)$ can be represented as the following equations in short-time Fourier transform:

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

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

where $N_1(f,k)$, $N_2(f,k)$ are the addtive noise, third-party speaker, coherent/incoherent or diffuse noise field, directional/non-directional noise source, $S(f,k)$ means the orignal speech component, $\Phi_s = \pi f \tau_0 cos(\theta_s)$ is the phase delay, $\tau_0 = d/c, c = 343(m/s)$ is the sound speed propagation in the air and d is the range between microphones.

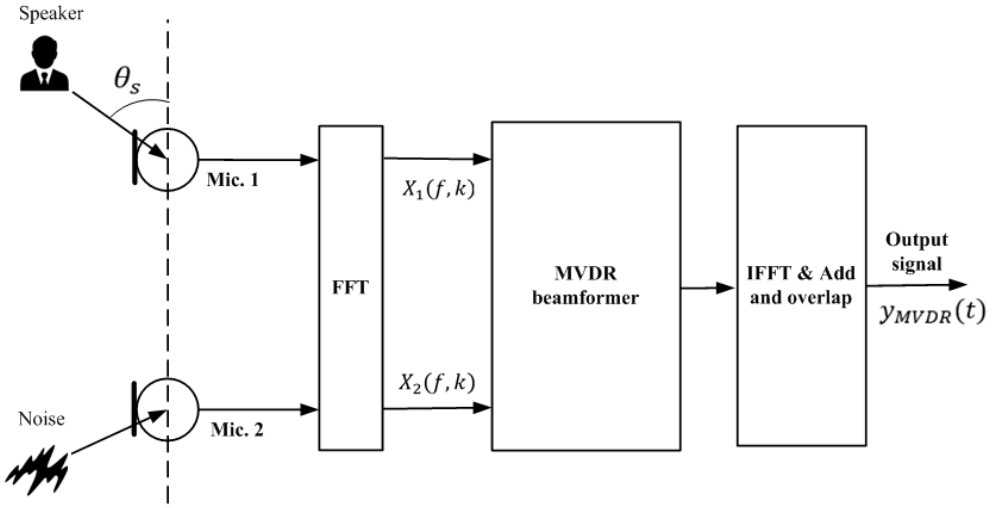


Figure 4: The structure of MVDR beamformer in the frequency - domain

With the defined symbols $\boldsymbol{X}(f,k) = [X_1(f,k) \quad X_2(f,k)]^T$, $\boldsymbol{D}_s(f,\theta_s) = [e^{j\Phi_s} \quad e^{-j\Phi_s}]^T$, $\boldsymbol{N}(f,k) = [N_1(f,k) \quad N_2(f,k)]^T$, the equation (1-2) can be expressed as:

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

The essential core of almost speech applications is calculating the optimum coefficients $\boldsymbol{W}(f,k)$ for preserving the desired target talker and attenuating the significant effect of surrounding noise field. The final result is obtaining an approximate signal $\hat{S}(f,k) \approx S(f,k)$.

The constrained criteria of MVDR beamformer can be described as the following formulation:

$$\underset{\boldsymbol{W}(f,k,\theta_s)}{min} \boldsymbol{W}^H(f,k,\theta_s)\boldsymbol{\Phi}_{NN}(f,k)\boldsymbol{W}(f,k,\theta_s) \; st \; \boldsymbol{W}^H(f,k,\theta_s)\boldsymbol{D}_s(f,\theta_s) = 1 \tag{4}$$

From equation (..), the expression of MVDR beamformer can be derived as the equation:

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

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

Unfortunately, under realistic recording situations, due to the complexity of environment, the task of computing noisy information still be difficult problem in almost acoustic equipments. Hence, the covariance matrix of received array signals is applied for computing the weights of MVDR beamformer.

$$\boldsymbol{W}_{MVDR}(f,k,\theta_s) = \frac{\boldsymbol{\Phi}_{XX}^{-1}(f,k)\boldsymbol{D}_s(f,\theta_s)}{\boldsymbol{D}_S^H(f,\theta_s)\boldsymbol{\Phi}_{XX}^{-1}(f,k)\boldsymbol{D}_s(f,\theta_s)} \tag{6}$$

where covaraince matrix of received array signals $\boldsymbol{\Phi}_{XX}(f,k) = E\{\boldsymbol{X}(f,k)\boldsymbol{X}^H(f,k)\}$.

$$\boldsymbol{\Phi}_{XX}(f,k) = \begin{Bmatrix} P_{X_1X_1}(f,k) & P_{X_1X_2}(f,k) \\ P_{X_2X_1}(f,k) & P_{X_2X_2}(f,k) \end{Bmatrix} \tag{7}$$

The auto and cross power spectral densities (PSD) of $X_1(f,k), X_2(f,k)$ can be defined as:

$$P_{X_iX_i}(f,k) = \alpha P_{X_iX_i}(f,k-1) + (1-\alpha)E\{|X_i(f,k)|^2\} \tag{8}$$

$$P_{X_iX_j}(f,k) = \alpha P_{X_iX_j}(f,k) + (1-\alpha)X_i^*(f,k)X_j(f,k) \tag{9}$$

with $i,j = \{1..2\}$ and α is the smoothing parameter in the range $\{0 \dots 1\}$.

In realistic situation with presence of complex and annoying environment, the error of MA configuration, the existence of various types of acoustic equipments, the complex and anooying scenario, MVDR beamformer's output signal usually corrupted. In the next section, an effective method, which based on two stage of estimating the exact time delay and frequency smoothing parameter for improving the robustness of beamformer's performance in adverse scenario.

III. The author's proposed method

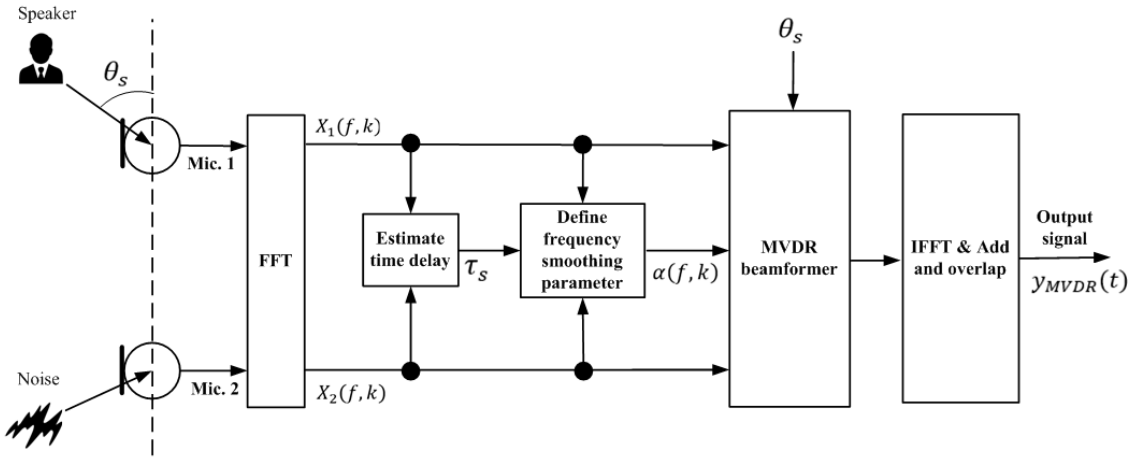


Figure 5: The author's proposed method

The author's idea is determining the precise time delay by modifying the formulation of GCC-PHAT:

$$\tau_s = \max_{\tau} \frac{X_1(f,k)X_2^*(f,k)}{|X_1(f,k)X_2^*(f,k)|} e^{j2\pi f\tau} \tag{10}$$

In rapidly changed scenario, the important task is specifying whether exists the clean speech data or noise. At the frame with speech, the speech presence probability [6] is essenttial

information for computing the necessary acoustic factors. Therefore, the author used it for modifying GCC-PHAT [7] to achieve robust of time delay as the following expression:

$$\tau_s = \underset{\tau}{max} \frac{X_1(f,k)X_2^*(f,k)}{|X_1(f,k)X_2^*(f,k)|^{SPP(f,k)}} e^{j2\pi f\tau} \quad (11)$$

After that, with phase delay $\Phi_s = \pi f \tau_s$, the suggested technique based on the observed coherence signals for taking into the account of frequency smoothing parameter. The expression of coherence between $X_1(f,k)$, $X_2(f,k)$ can be derived as [8]:

$$\Gamma_{X_1X_2}(f,k) = \frac{SNR(f,k)}{1+SNR(f,k)} e^{j2\Phi_s} + \frac{1}{1+SNR(f,k)} \Gamma_N(f) \quad (12)$$

where $\Gamma_{X_1X_2}(f,k) = \frac{P_{X_1X_2}(f,k)}{\sqrt{P_{X_1X_1}(f,k) \times P_{X_2X_2}(f,k)}}$ and $\Gamma_N(f)$ means the coherence of noise field. In coherent noise field, $\Gamma_N(f) = 1$ and in diffuse nois field $\Gamma_N(f) = \frac{sin(\omega\tau_0)}{\omega\tau_0}$, $\omega = 2\pi f$.

With symbol $\rho(f,k) = \frac{SNR(f,k)}{1+SNR(f,k)}$, the equation (12) can be rewritten as:

$$\Gamma_{X_1X_2}(f,k) = \rho(f,k)e^{j2\Phi_s} + \big(1-\rho(f,k)\big)\Gamma_N(f) \quad (13)$$

And:

$$\rho(f,k) = \frac{\Gamma_{X_1X_2}(f,k) - \Gamma_N(f)}{e^{j2\Phi_s} - \Gamma_N(f)} \quad (14)$$

In real-life speech application, with presence of undetermined noise sources or adverse situation, the expression of coherence of noise field will be updated according to the complex of environment as [9]:

$$\Gamma_N(f) = \frac{sin(\omega\tau_0)}{\omega\tau_0 \left(1 + \big(1 - SPP(f,k)\big)\frac{\gamma_n}{P_{nn}}\right)} \quad (15)$$

with γ_n is uncorrelated noise component and P_{nn} is spectral noise floor.

The frequency smothing parameter is changed as $\alpha(f,k) = \rho(f,k)$. With the frequency smoothing parameter, the covariance matrix of MA signals is adaptively updating and tracking to the speech presence probability under annoying situations. In section IV, an experiments was performend for verifying the effectiveness of described method in reducing speech distortion and suppressing background noise field.

IV. Experiments and discussion

The purpose of this section is verifying the advantages of described method (desme) in reducing the musical noise, speech distortion and improving the speech enhancement in comparison to traditional MVDR beamformer. The experiments is conducted in real-life

scenarios with presence of third-party speaker, the sound of television, the fan, the washing machine and other different non-directional noise sources. The scheme of illustrated simulation has been described in the Figure 6.

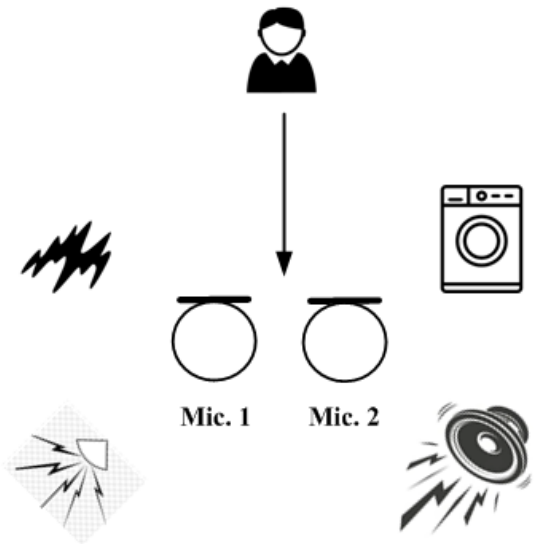
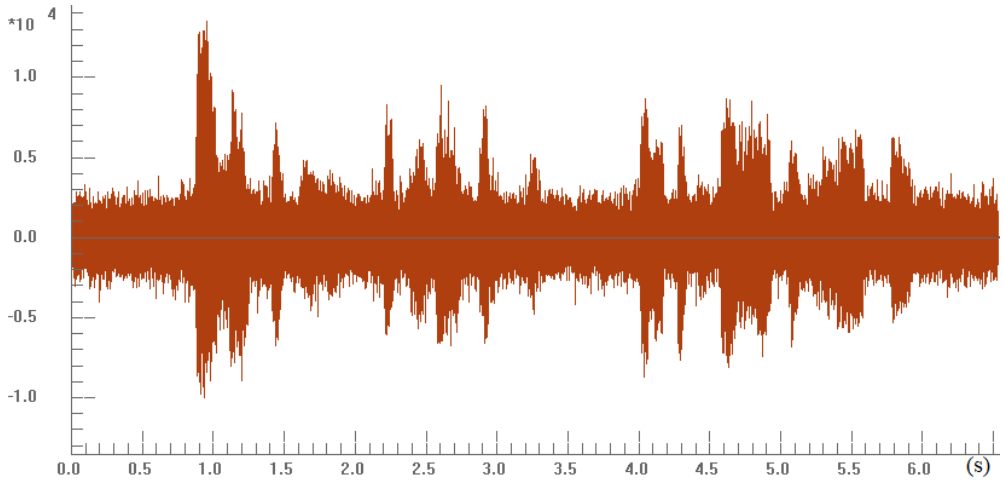


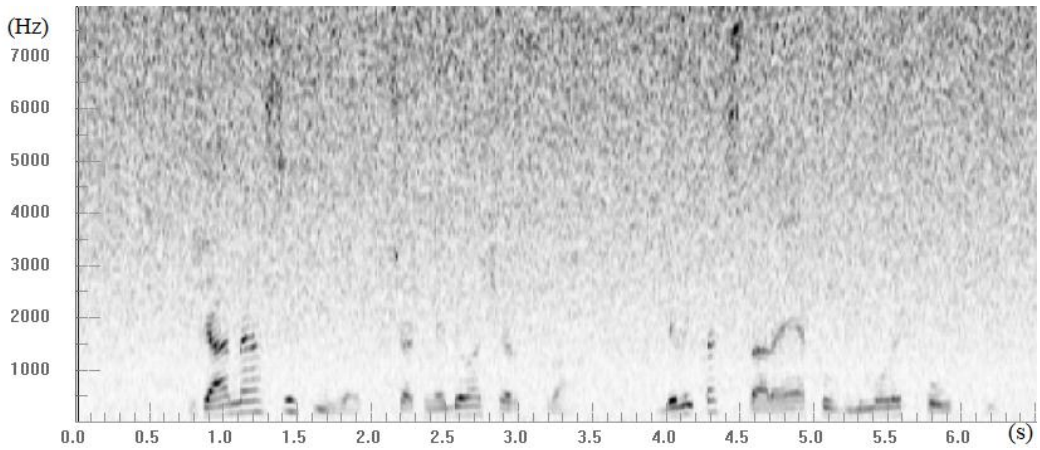
Figure 6: The conducted experiment in living room

The distance from a stand speaker to DMA2 is $L = 3.5(m)$. The size of living room is $3x4x4.5(m)$. The preferred direction of useful signal is $\theta_s = 90(deg)$, the range between two mounted microphones is $d = 4.25\ (cm)$. The author used an objective measurement [10] for comparing the obtained SNR between MA signals, the processed signals by conventional MVDR beamformer and desme. In addition, the curve of these signals's energy will be drawn for visual analyzing the effectiveness of suggested techniques. For recording the noisy mixture, these parameters are used: overlap 50%, frequency rate $Fs = 16\ kHz$.

The waveform and spectrogram of MA signals is shown by Figure 7.



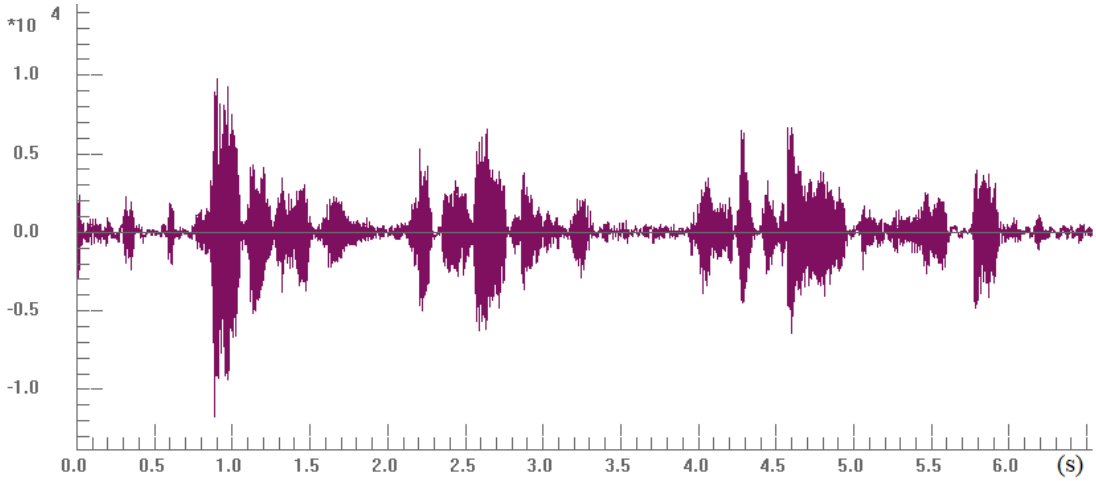
(a)



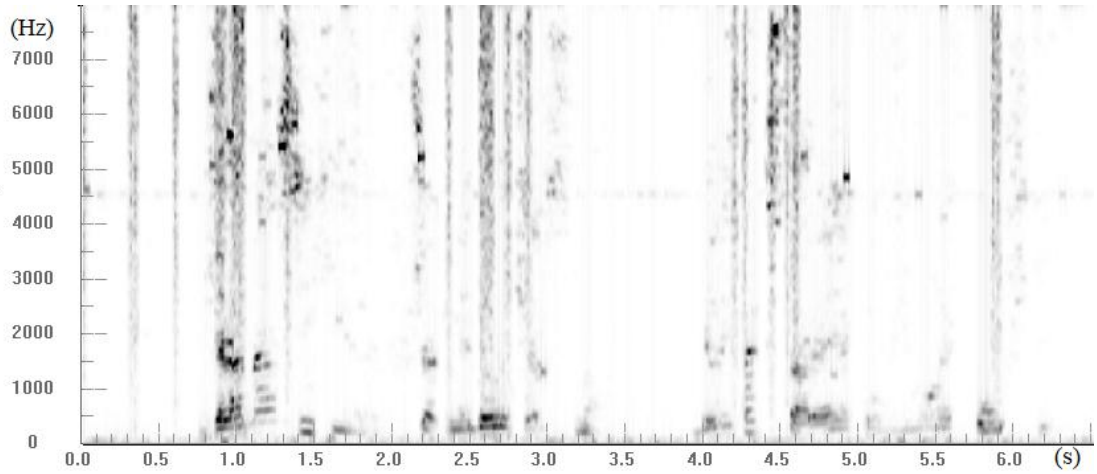
(b)

Figure 7: The waveform (a) and spectrogram (b) of observed microphone array signals

For converting MA signals into frequency-domain, the author applied $nFFT = 512$ and appropriate smoothing parameter $\alpha = 0.1$ for implemeting MVDR beamformer. The output signal can be derived in Figure 8.



(a)

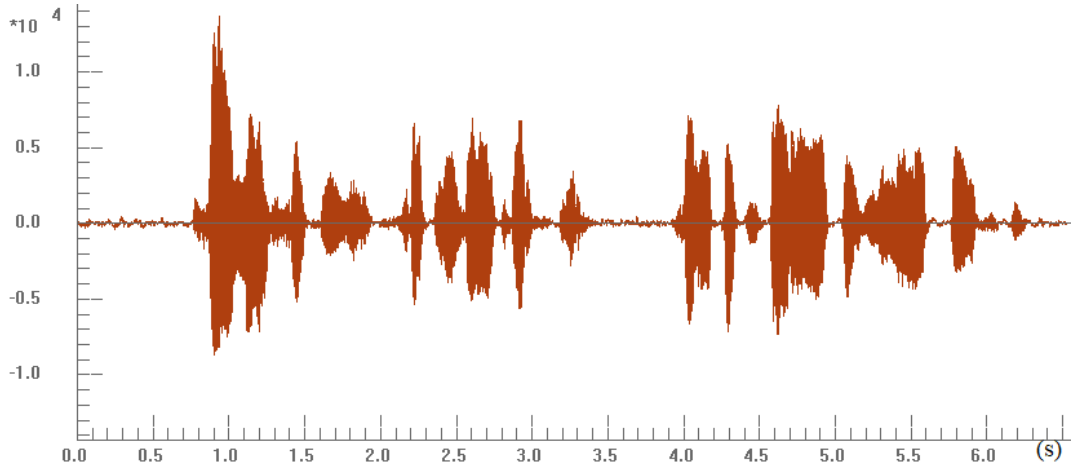


(b)

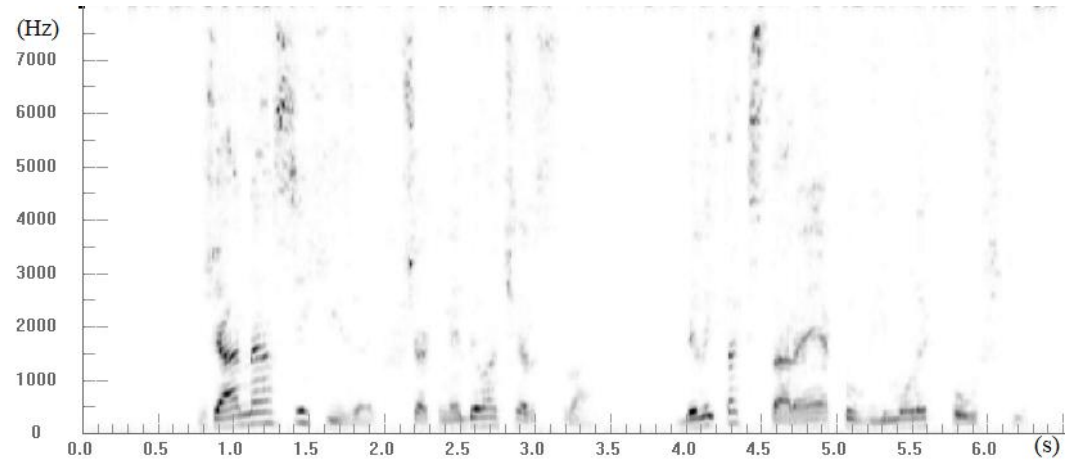
Figure 8: The MVDR beamformer's output signal, (a) waveform and (b) spectrogram

Because of the complexity of recorded environments, the heterogeneous recording environment, the microphone mismatches, the different microphones quality, the inaccurate estimation of tranfer function of target sound source to axis of DMA2, the overall MVDR beamformer's performance often degraded. Hence, the large background noise field or speech distortion is not satisfied the human auditory. Therefore, the author's approach is calculating the exact time delay by modified GCC-PHAT with incorporating speech presence probability at the frame with speech component. With precise measurement of phase delay of useful signal, the author determined smoothing parameter according to the existence of clean speech data in recorded MA signals. Under realistic recording environment, the defined formulation of coherence between two point of noise is modified with speech presence probability. At finally, the smoothing parameter is calculated as frequency-value. This approach allows increasing the beamformer's evaluation and performance with promising results.

The resulting signal yields as:



(a)



(b)

Figure 9: The promising result by applying desme with enhanced in (a) waveform and spectrogram (b)

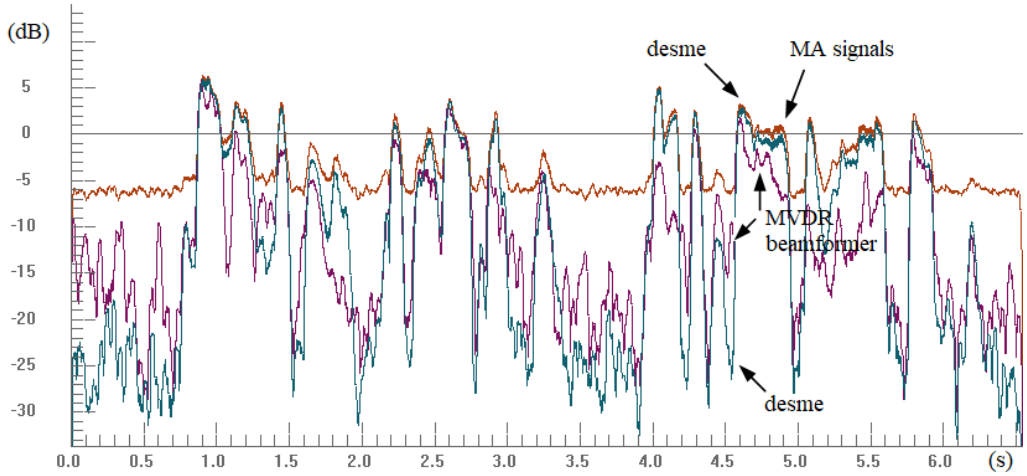


Figure 10: The comparison of energy between MA signals, conventional MVDR beamformer's output singal and obtained result of desme

Figure 10 describes the curve of energy and Table 1 compared obtained SNR between these signals. As we can see that, desme owns the capability of reducing the speech distortion to 5 dB, mitigating the residual or musical noise to 12.9 dB and improving the speech quality in the term of signal-to-noise ratio from 5.3 to 8.8 dB.

Method Estimation	**Microphone array signal**	**MVDR beamformer**	**desme**
NIST STNR	6.1	14.8	20.1
WADA SNR	7.2	13.6	22.4

Table 1: The obtained SNR between signals

From above Figures and Table 1, this paper can conclude some assessments as:

- MVDR beamformer performed well in complex and annoying environments. The resulting signal is high evaluation with saving the clean speech data while suppressing surrounding noise field.

- Due to many unwanted reasons, such as, the adverse recording environment, the existence of interference, the movement of talker, the internal electrical noise, the microphone mismatches, the error of sampling frequency, beamformer's evaluation usually degraded. As a result, the speech intelligibility of final signal was seriously affected by unacceptable noise level.

- The highligh of the author's suggested method utilized speech presence probability for measuring phase time between two mounted microphones and determining exact value, which indicated the existence of speech component. This value is fequency parameter and is very suitable serving as smoothing parameter in MVDR beamformer.

- The effectiveness is confirming the advantages of described approach in removing background noise, decreasing speech distoriton and enhancing the speech quality.

V. Conclusion

The perspective numerical simulations has confirmed the effectiveness of suggesteda approach in increasing the robust of MVDR beamformer in real-life situations. The appealing properties of the author's proposed method is applying the speech presence probability to precisely estimate time delay and phase delay between two microphones. As a result, the smoothing parameter is choosen from relation between observed coherence of microphones signals and the speech component at each frame. The optimum smoothing parameter contains the spatial information about existence of desired talker under rapidly changing environments. The flexible smoothing parameter allows auto-cross PSD of covariance matrix contains the only noise component. With incorporating the exact steering vector, the desribed technique has shown the ability of concering the beampattern at specified sound source with promising result of reducing speech distortion to 5 dB, suppressing the background noise level to 12.9 dB and enhancing the obtained SNR from 5.3 to 8.8 dB.

References

[1] B. Wang, L. Xu, L. Hu, H. Wang, B. Ruan and Y. Wan, "Spatial Audio Estimation Based on Dual Microphone Arrays," *2025 10th International Conference on Intelligent Computing and Signal Processing (ICSP)*, Xi'an, China, 2025, pp. 101-105, doi: 10.1109/ICSP65755.2025.11087172.

[2] A. A. Gadag, R. Sharma and D. K T, "Beamforming using Different Window Techniques for Near-Field Speech in Anechoic and Reverberant Environment," *2023 26th Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)*, Delhi, India, 2023, pp. 1-5, doi: 10.1109/O-COCOSDA60357.2023.10482942.

[3] K. Chaudhari, M. Sutaone and P. Bartakke, "Adaptive Diagonal Loading of MVDR Beamformer For Sustainable Performance In Noisy Conditions," *2020 IEEE Region 10 Symposium (TENSYMP)*, Dhaka, Bangladesh, 2020, pp. 1144-1147, doi: 10.1109/TENSYMP50017.2020.9230850.

[4] 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," *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.

[5] R. Ali, T. Van Waterschoot and M. Moonen, "Integration of a Priori and Estimated Constraints Into an MVDR Beamformer for Speech Enhancement," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 2288-2300, Dec. 2019, doi: 10.1109/TASLP.2019.2946086.

[6] I. Potamitis, "Estimation of speech presence probability in the field of microphone array," in *IEEE Signal Processing Letters*, vol. 11, no. 12, pp. 956-959, Dec. 2004, doi: 10.1109/LSP.2004.838200.

[7] R. Lee, M. -S. Kang, B. -H. Kim, K. -H. Park, S. Q. Lee and H. -M. Park, "Sound Source Localization Based on GCC-PHAT With Diffuseness Mask in Noisy and Reverberant Environments," in *IEEE Access*, vol. 8, pp. 7373-7382, 2020, doi: 10.1109/ACCESS.2019.2963768.

[8] N. Yousefian, K. Kokkinakis and P. C. Loizou, "A coherence-based algorithm for noise reduction in dual-microphone applications," *2010 18th European Signal Processing Conference*, Aalborg, Denmark, 2010, pp. 1904-1908.

[9] J. Bitzer, K. . -D. Kammeyer and K. U. Simmer, "An alternative implementation of the superdirective beamformer," *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99 (Cat. No.99TH8452)*, New Paltz, NY, USA, 1999, pp. 7-10, doi: 10.1109/ASPAA.1999.810836.

[10] SNRVAD. [Online]. Available: https://labrosa.ee.columbia.edu/projects/snreval/