An appropriate post - filtering for GSC beamformer*

Nguyen Thi Huyen Chau[1,†], Quan Trong The[2,*,†]

[1] Thang Long University, Hanoi, Vietnam
[2] Post and Telecommunication Institute of Technology, Hanoi, Vietnam

Abstract
Nowadays, the using of microphone array (MA) technology has been common installed in almost acoustic equipment, such as smart phone, voice-controlled device, teleconference system, surveillance equipment, cochlear implant, hearing aids for processing the noisy mixture and enhancing the speech quality, the speech intelligibility and perceptual metric listener. Generalized sidelobe canceller (GSC) is one the most effective beamforming method for extracting the desired target speaker while suppressing the interference, third-party talker and other signal from uncertain direction without speech distortion. However, because of the complex noisy environment, the inaccurate estimation of preferred steering vector, the displacement of MA geometry, the different microphone quality, the error of sampling rate or the rapidly changed environmental factor, the GSC beamformer's performance often corrupted. The existence of speech distortion or musical noise decreases the perceptual metric listener, the speech intelligibility. In this contribution, the author proposed efficient post-filtering for eliminating the noise level with an acceptable satisfied speech quality to enhance the GSC beamformer's output signal. The conducted simulation shows the effectiveness of removing the noise level to 11.7 dB and improving the signal-to-noise (SNR) ratio from 12.5 to 13.4 dB. The author's proposed post-filtering can deal other complicated problems.

Keywords
microphone array, beamforming, speech enhancement, post - Filtering, the signal-to-noise ratio

1. Introduction

Speech enhancement plays an important role in numerous acoustic speech applications, such as, teleconference system, hearing aids, voice - controlled devices, smart-phone, cochlear implant for extracting the desired target speaker while attenuating the different incoming signal from other signals. Nowadays, the utilizing of microphone array technology has been popular, due to its convenience of removing noise and enhancing speech at the same time. MA beamforming often uses the spatial information of preferred steering vector, the designed configuration of MA geometry, the properties of surrounding environment to obtain high directional beampattern towards the sound source. The significant benefit of using MA beamforming is the ability of incorporating single - channel algorithm, pre - processing, post - Filtering for enhancing the digital signal processing system's evaluation.

Microphone array can be categorized into two groups: the fixed beamformer and adaptive beam-former. Fixed beamformer often uses the prior spatial diversity to generate constant beamformer's coefficients. Delay and sum (DAS) [1-2] beamformer is well-known distinct representation of fixed beamformer. DAS beamformer's weights based on the direction of arrival of useful signal to the MA. However, in non-stationary noise, complex and adverse environment, DAS's beamformer does not work well. Therefore, adaptive beamformer with differential microphone array (DIF) [3-5], minimum variance distortionless response (MVDR) [6-8], linearly constrained distortionless response (LCMV) [9-10] and GSC beamformer. These technique based on

*CITI'2025: 3rd International Workshop on Computer Information Technologies in Industry 4.0, June 11–12, 2025, Ternopil, Ukraine

[1*] Corresponding author.
[†] These authors contributed equally.
huyenchau@thanglong.edu.vn (N.T.H. Chau); theqt@ptit.edu.vn (Q.T. The)
0000-0003-4091-0271 (N.T.H Chau); 0000-0002-2456-9498 (Q.T. The)

constrained criteria of preserving the clean speech data a specified direction, minimizing the total output noise power or combining with other optimum criteria.

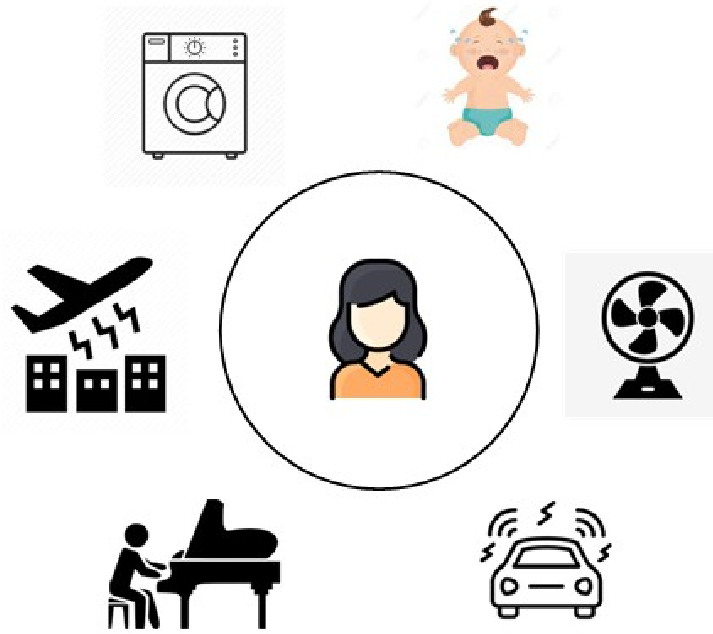

Figure 1: The annoying environment with various types of sound.

GSC beamformer is one of the most useful beamforming, which concerns the steerable beampattern at specified direction and use an adaptive noise canceller to obtain the clean speech data. GSC beamformer outperforms in coherent - diff use noise fi eld and owns the easy computing. Unfortunately, in realistic recording environment, the error of sampling, the displacement of MA, the inaccurate estimation of environmental properties, the error of sampling rate, the different sensitivities of sensors usually decrease the GSC beamformer's performance. There is numerous research and work attempt dealing this drawback.

Wang J et al [11] proposed controlling the smoothing parameter of the adaptive interference canceller (AIC) to improve the interference suppression and block the target speech distortion. The author's approach based on a time - varying Gaussian distribution under maximum likelihood algorithm.

The numerical results has confirmed the robustness of GSC beamformer in various types of recording environment.

Kim S et al [12] used a phase - error filter (PEF) for increasing multi-channel signal processing system. This approach improved the fixed beamformer of GSC structure and the coefficients adaptively updated by PEF. The conducted experiment shows better promising results with perceptual performance and intelligibility score, noise reduction.

In [13], the author exploited phase difference from received array signals to estimate the target-to-non-target directional signal ratio to adaptively control AIC. The demonstrated experiment has verified the problem of reducing residual noise in comparison with conventional GSC beamformer, PEF filter under different conditions.

In [14], Ali R et al proposed using an external microphone in incorporating with an existing local microphone array for obtaining noise reduction and speech enhancement, which applied in voice - controlled device, hearing aids, teleconference system and cochlear implant. This combination allows decreasing noise level and increasing the speech quality in comparison with traditional GSC beamformer.

Zohourian M et al [15] investigated maximum likelihood technique and target presence probability to determine a common spectral postfi lter, which leads to larger speech enhancement and improvements of desired target signal in realistic experiment.

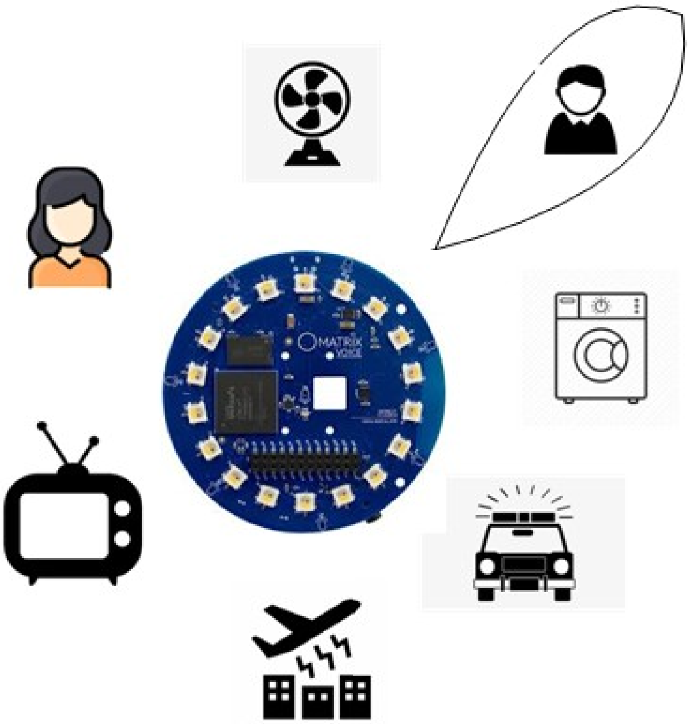
Figure 2: The designed beampattern on the specified direction of desired speaker.

Priyanka S et al [16] studied Least Mean Square (LMS), Normalized LMS and Recursive Least Square (RLS) to incorporate with GSC beamformer to evaluate the overall performance under various types of noise. These adaptive algorithms allow increasing the noise reduction and speech enhancement in real-life recording scenario. The numerical simulation validated the author's suggested approach.

Jiang Q et al [17] presented a new adaptively controlling for GSC beamformer. This method uses the output of superdirective beamformer and blocking matrix to compute the signal-to-interference (SIR) ratio in a certain frequency range. The ratio was applied to update the adaptive noise canceller (ANC) to achieve the robust beamformer's evaluation under noisy condition.

However, these works cannot address the realistic recording scenario, because of the undetermined reasons, the imperfect estimation of necessary parameter. In this contribution, the author proposed an efficient post - Filtering for overcoming this drawback for further increasing the speech enhancement and noise reduction. The suggested approach based on the priori information of direction of arrival of interest useful signal to obtain the post - Filtering.

The rest of this contribution is organized as following ways: Section II describes the GSC's structure in frequency domain, the author's suggested post - Filtering is presented in section III with using the prior direction of arrival and section IV illustrates the conducted experiment in real-life recording environment, section V concludes the work.

2. GSC beamformer

Consider a microphone array comprising M microphones, which captures a target desired sound at specified direction in a noisy reverberant recording scenario. The representation of received array signals $y(k,n)=[Y_1(k,n)\ldots Y_M(k,n)]^T$ can be described in the short-time Fourier transform (STFT) domain as [18]:

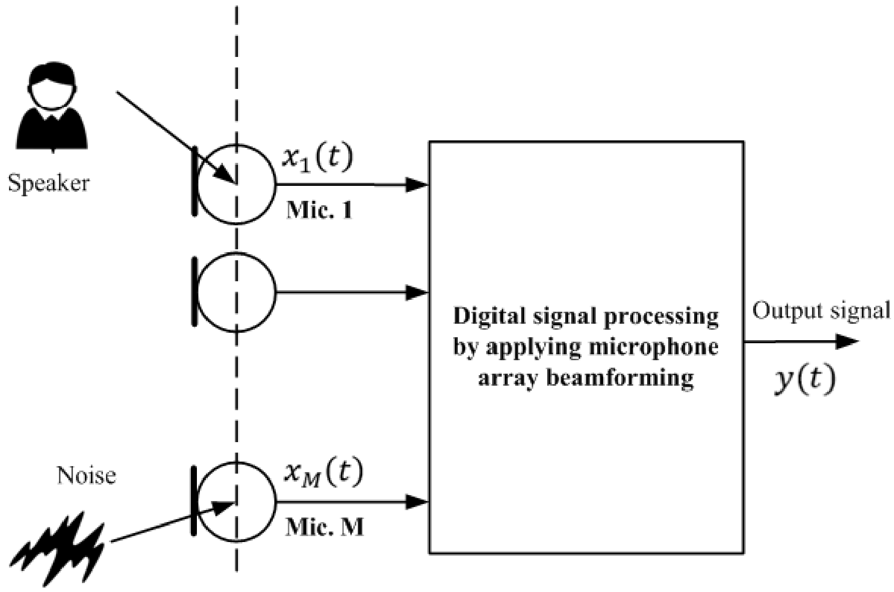

Figure 3: The scheme of principal working of microphone array.

$$y(k,n)=x(k,n)+v(k,n) \tag{1}$$

$$\text{¿} g_r(k,n)X_r(k,n)+v(k,n) \tag{2}$$

where k is the frequency index, n is the current frame, the superscript $\square^T$ denotes the transpose operator, $v(k,n)$ means the noise signal and $x(k,n)$ is the target speech component, $X_r(k,n)$ is the clean speech signal at the selected reference microphone, respectively.

The steering vector $g_r(k,n)$ with respect to the $r-th$ microphone can be calculated as the following equation [9]:

$$g_r(k,n)=\left[\frac{G_1(k,n)}{G_r(k,n)},\ldots,\frac{G_M(k.n)}{G_r(k,n)}\right]^T \tag{3}$$

where $G_i(k,n)$ denote the acoustic transfer function from the sound source to the t-th microphone. We can define that $g_r(k,n)\approx g_r(n)$, based on assumption that the environment is slowly time - varying.

Note that each frequency bin is treated independently, and the author will omit the frequency index k for brevity. The necessary problem of speech enhancement is finding an optimum coefficient obtaining the processed signal, which approximately the original speech component. We recover the desired speaker by the GSC beamformer as:

$$\widehat{X}(n)=W_{GSC}^H(n)y(n) \tag{4}$$

where $W_{GSC}(n)=[W_1(n),\ldots,W_M(n)]^T$ is a filter length M. The traditional GSC beamforming technique retains the desired talker undistorted while suppressing the background noise, interference and surrounding noise. Therefore, the formulation of $W_{GSC}(n)$ can be expressed as:

$$W_{GSC}(n)=W_q-BW_a(n) \tag{5}$$

where $W_a(n) \epsilon C^{(M-1)x1}$, $B \epsilon C^{Mx(M-1)}$, $w_q \epsilon C^{(M-1)}$. The fixed beamformer W_q concerns the steerable beampattern at specified direction and form a main signal $Y_s(n)=W_q^H(n)\,y(n)$ for the desired speech component. A matched filter $W_q(n)=\frac{g_r}{g_r^H g_r}$ is commonly adopted due to its robustness under various types of noisy conditions [19]. The blocking matrix used for completely alleviating the target signal and provide perfect noise references $u(n)=B\,y(n)$ for the interference signal, which can be achieved by spanning the left null-space of g_r, $B^H g_r=0$ and then steers the null-beampattern at the direction of target speech.

The Adaptive Interference Canceller often deploys un constrained adaptive filter $W_a(n)$ to block the residual noise in $Y_s(n)$ with the reference signal $u(n)$.

In practical recording situation, the robust speech enhancement of GSC beamformer is indispensable due to the microphone mismatches, the displacement of MA geometry, the different microphone quality, the inaccurate estimation of preferred steering vector, the error of sampling rate or the blocking matrix imperfect block the target speech signal, the reference signal contains target speech leakage. Consequently, the remained noise component at GSC beamformer's output signal or speech distortion seriously degrade the performance of signal processing system.

3. The author's suggested technique

In this section, the author proposed using an efficient post - Filtering, which based on the prior spatial information of preferred direction of interest useful target talker, the phase difference of two mounted microphone signals for achieving the necessary formulation.

The author's post - Filtering is given by:

$$thoPF(n)=\frac{\sigma_x^2(n)}{\sigma_x^2(n)+\sigma_v^2(n)} \tag{6}$$

where $\sigma_x(n)$, $\sigma_v(n)$ is the variance of speech and noise, respectively.

Based on prior steering vector $g_r(n)$ and observed covariance matrix of received array signals, the covariance of speech component and noise can be yield as:

$$\sigma_x^2(n)=\frac{1}{g_r^H(n)\,\Phi_{yy}^{-1}(n)\,g_r(n)} \tag{7}$$

And

$$\sigma_v^2(n)=W_{GSC}^H(n)\,\Phi_{vv}(n)\,W_{GSC}(n) \tag{8}$$

where $\Phi_{yy}(n)$, $\Phi_{vv}(n)$ is covariance matrix of observed microphone array signals and noise.

$\Phi_{yy}(n)=E\{y^H(n)\,y(n)\}$ based on the received array signals, and $\Phi_{vv}(n)$ is computed at the frame, in which exists the only noise - frame.

In numerous practical recording situation, the information about the noise is not always available, and calculation of $\Phi_{vv}(n)$ still challenge task in almost acoustic equipment.

The author suggested exploiting the spectral masking $G^{BM}(n)=1-\exp\left(j\phi_{ij}^{norm}\right)$ and $G^{DS}(n)=0.5+0.5*\exp\left(j\phi_{ij}^{norm}\right)$ [20], $\phi_{ij}^{norm}=\frac{\psi.c}{k.d}$, ψ is average phase difference of $M-1$ pairs, between the microphone i-th and $i+1$ th, $i\epsilon\{1\ldots M-1\}$, $c=343(m/s)$ means the sound speed propagation in the fresh air, d the range between two mounted microphones.

The covariance matrix of noise can be determined as:

$$\Phi_{vv}(n)=\frac{G^{BM}(n)}{G^{DSB}(n)+G^{BM}(n)}\times\Phi_{yy}(n) \tag{9}$$

The advantage of the author's proposed approach is utilizing the designed MA configuration, the impinging angle of interest speaker relative to the MA to compute post - Filtering for further removing the musical noise, residual noise and increasing the speech quality.

And the enhanced signal can be derived as:

$$\widehat{X}_{thoPF}(n)=\widehat{X}(n)\times thoPF(n) \tag{10}$$

In the next section, the author illustrates experiment to verify the effectiveness of post - Filtering.

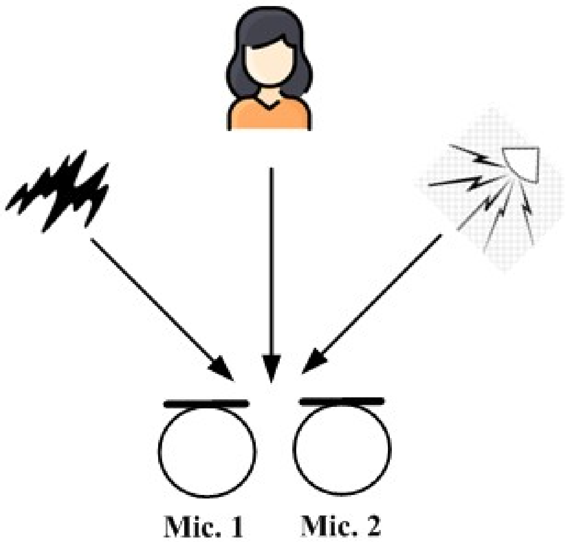

Figure 4: The conducted experiment in realistic living room under adverse environment.

4. Experiments

In this section, the author illustrates an experiment to verify the advantage of proposed post - Filtering (adpF) in comparison with the conventional GSC beamformer (cGSCbe). An objective measurement [21] is applied to calculate the speech quality between the observed microphone array signals, the processed signal by cGSbe, adpF. A talker stand at distance $L=5(m)$, the preferred direction of arrival of useful signal is $\theta_s=90(deg)$ relatives to the axis of dual - microphone system (DMA2), the range between two mounted microphones is $d=5(cm)$. The experiment is conducted in living room with size $3.5x5.2x4.0(m)$. The scheme of experiment is shown in Figure 4.

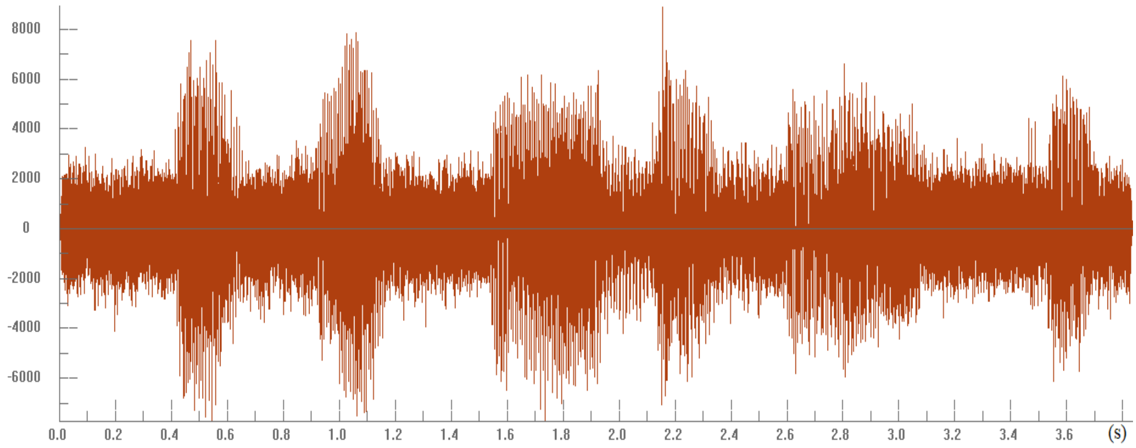

Figure 5: The original received array signals.

For capturing the original microphone array signals, these parameters were used: frequency sampling $Fs = 16\,kHz$, Hanning window, $nFFT = 512$, overlap $50\,\%$. The received array signals can be shown in Figure 5 and Figure 6.

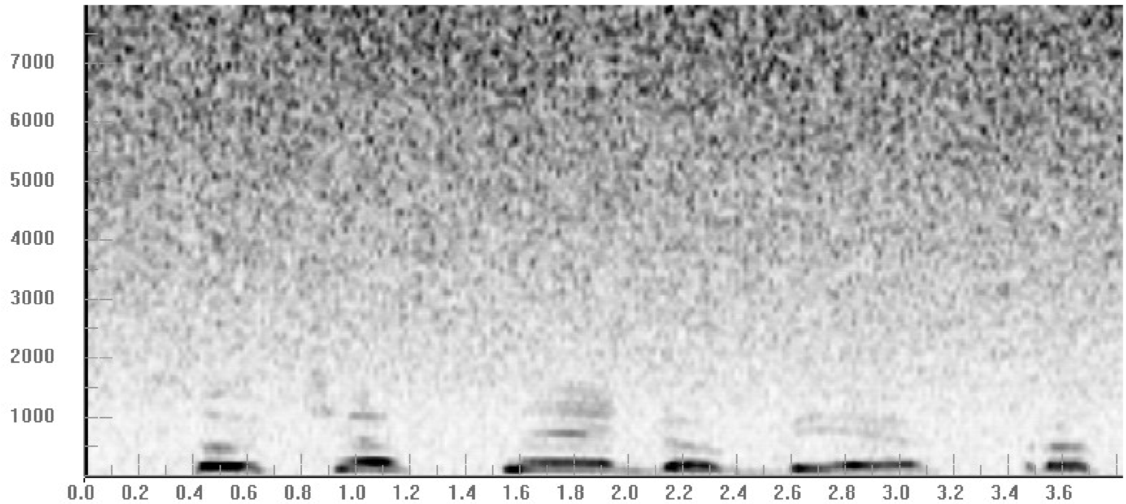

Figure 6: The spectrogram of captured microphone array signals.

With suitable smoothing parameter $\alpha = 0.1$, the GSC beamformer's output signal can be derived:

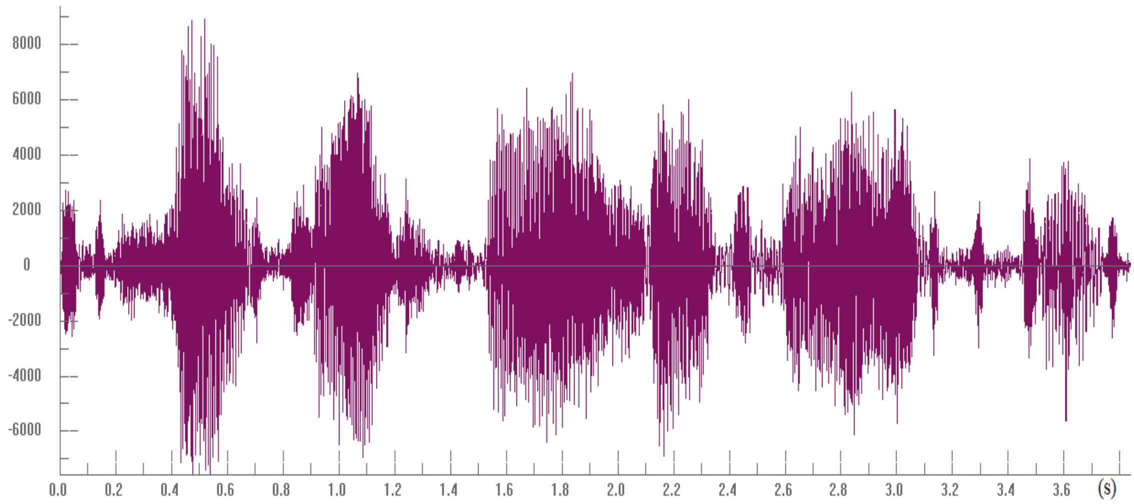

Figure 7: The waveform of processed signal by cGSCbe.

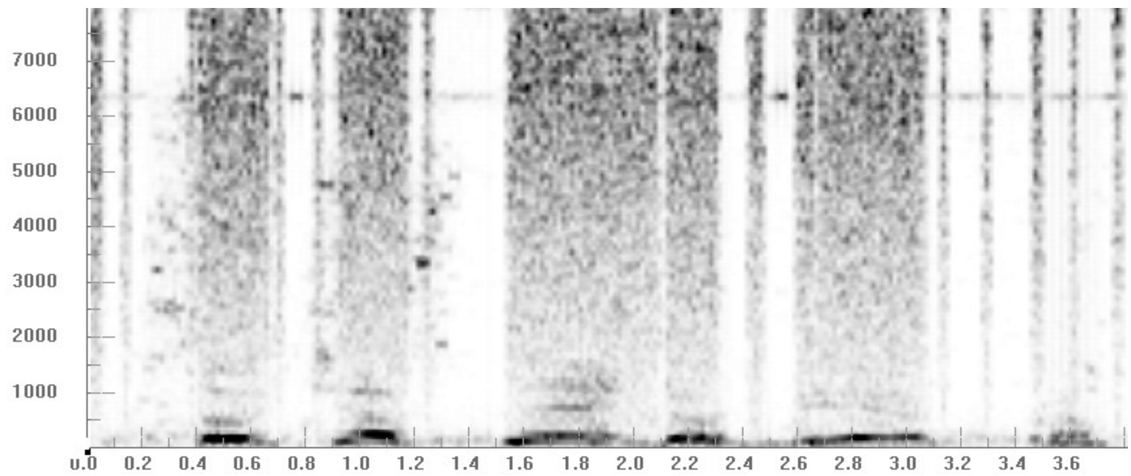

Figure 8: The spectrogram of processed signal by cGSCbe.

Because of the complex and annoying recording environment, the microphone mismatches, the moving head of speaker, the inaccurate estimation of steering vector, GSC beamformer's performance often corrupted. The musical noise, residual noise still be a challenging problem. In Figure 8, the musical noise seriously affects the output signal. Musical noise occurs due to the heterogeneous environment, microphone mismatches or error of sampling rate at high-frequency band. Therefore, the necessary of post - Filtering to overcome this drawback is an essential core in almost acoustic problems. The efficient author's suggested technique suppresses musical noise has been illustrated in Figure 9 and Figure 10.

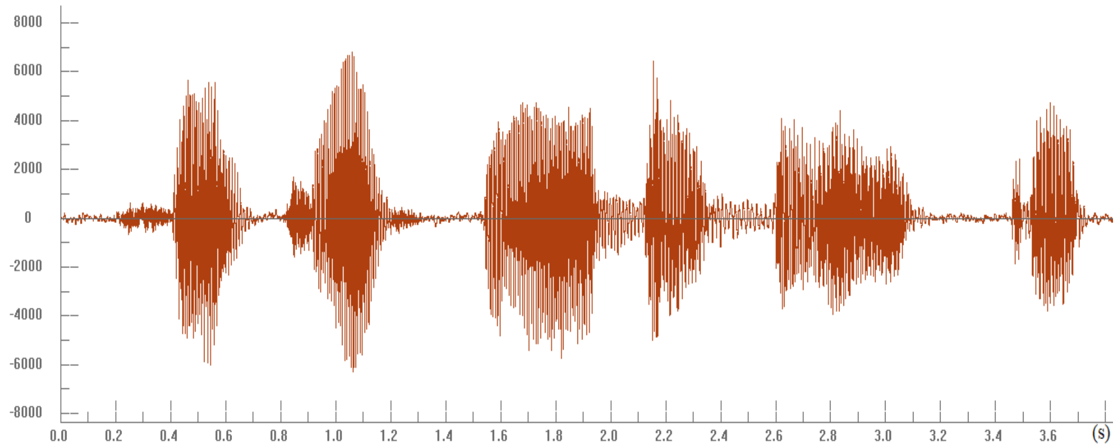

Figure 9: The processed signal by applying adpF.

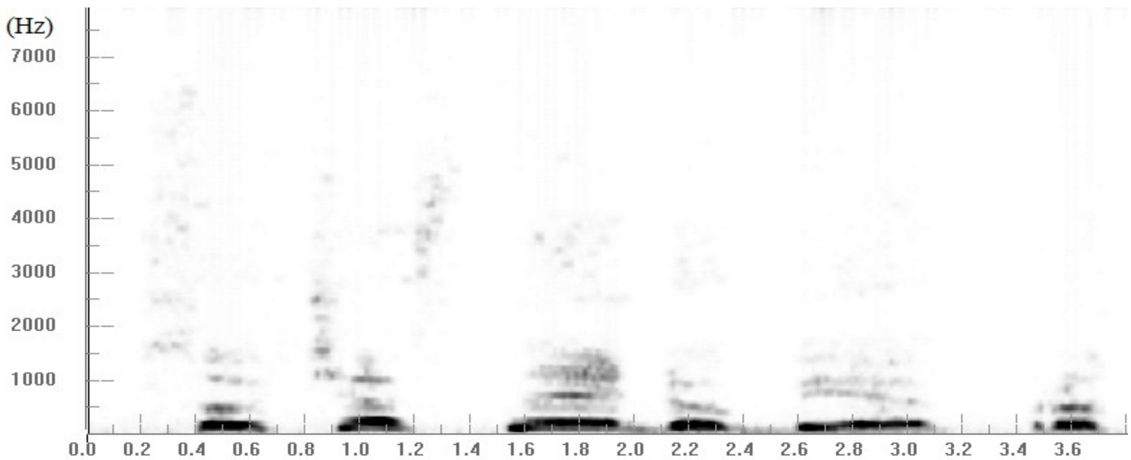

Figure 10: The spectrogram of promising signal by utilizing adpF.

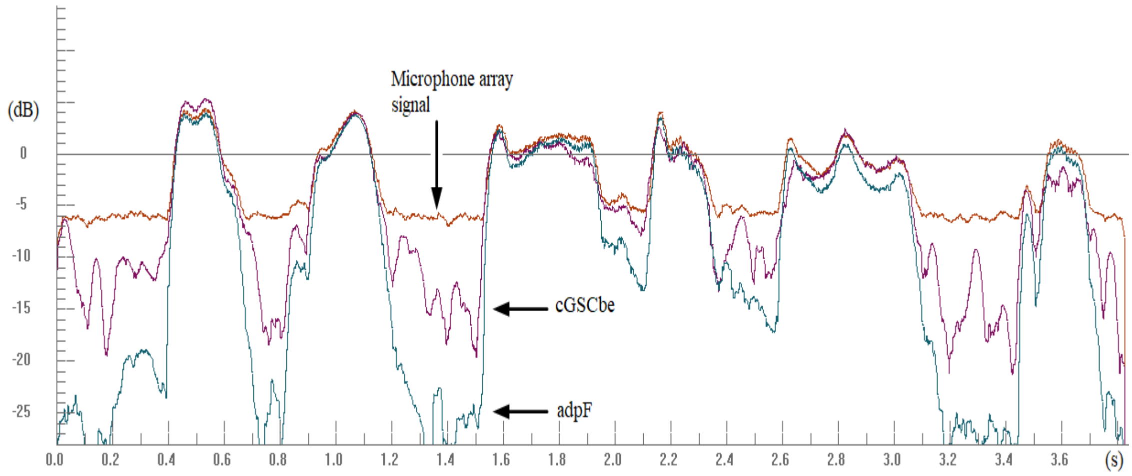

Figure 11: The comparison energy between the observed MA signals, the processed signal by cGSCbe, adpF.

Figure 11 and Table 1 describe the comparison energy between the observed microphone array, the obtained signals by cGSCbe, adpF.

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

Method Estimation	Microphone array signal	cGSCbe	adpF
NIST STNR	8.2	15.4	22.8
WADA SNR	2.3	10.1	22.6

The effectiveness of the author's proposed method in reducing musical noise, residual background noise to 11.7 dB and improving the perceptual metric listener, speech intelligibility, speech quality from 12.5 to 13.4 dB. The above post - Filtering not only recover the speech component but also removes musical noise, background noise. The advantage of the author's post - Filtering is using the accurate estimation of variance of speech and variance of noise, which was calculated by applying spectral gain. The author's approach is utilizing the prior information of impinging incident angle of direction of arrival of interest signal to form an additive post - Filtering for alleviating the musical noise, background noise. The presented post - Filtering owns the characteristic of rapid changed environmental factors and exactly computes variance of speech and noise component according to the considered frame. This approach ensures preserving the original clean speech date while suppressing background noise. In the future, the author will continue investigating the described above method for multi-channel processing system.

Conclusion

GSC beamformer is commonly installed in almost acoustic equipment, due to its high spatial diversity, high directivity index and easy implementation. In realistic recording environment, due to the microphone mismatches, the different microphone sensitivities and other undetermined reasons, GSC beamformer's evaluation usually degrades. The existence of musical noise or unacceptable residual noise corrupts the speech quality of processed signal. In this paper, the author illustrates an efficient post - Filtering to suppress the remained musical noise, which occurs at GSC beamformer's output due to the complex and adverse recording scenario. Consequently, the effectiveness of the author's suggested technique is confirmed by decreasing the background noise to 11.7 dB and increasing the speech quality in the term of signal - to - noise ratio from 12.5 to 13.4 dB. The numerical results has confirmed the effectiveness of the author's post - Filtering in reducing noise component and improving the speech quality, perceptual metric listener and speech intelligibility. The numerical simulation has confirmed the effectiveness of the author's suggested method in improving the robust speech enhancement. The above proposed approach can be integrated into multi-channel system.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] Y. Zeng and R. C. Hendriks, "Distributed Delay and Sum Beamformer for Speech Enhancement via Randomized Gossip," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 1, pp. 260-273, Jan. 2014, doi: 10.1109/TASLP.2013.2290861.

[2] P. K. Chodingala, S. S. Chaturvedi, A. T. Patil and H. A. Patil, "Robustness of DAS Beamformer Over MVDR for Replay Attack Detection On Voice Assistants," 2022 IEEE International Conference on Signal Processing and Communications (SPCOM), angalore, India, 2022, pp. 1-5, doi: 10.1109/SPCOM55316.2022.9840757.

[3] 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," 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.

[4] H. Pei, K. Chen, G. Huang, J. Jin, J. Benesty and J. Chen, "On the Design of Low-Rank Differential Beamformers with Nonuniform Linear Microphone Arrays," ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP), Hyderabad, India, 2025, pp. 1-5, doi: 10.1109/ICASSP49660.2025.10889353.

[5] X. Wang, G. Huang, I. Cohen, J. Benesty and J. Chen, "Robust Steerable Differential Beamformers with Null Constraints for Concentric Circular Microphone Arrays," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 4465-4469, doi: 0.1109/ICASSP39728.2021.9414119.

[6] D. Wang and C. Bao, "Multi-channel Speech Enhancement Based on the MVDR Beamformer and Postfilter," 2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Macau, China, 2020, pp. 1-5, doi: 10.1109/ICSPCC50002.2020.9259489.

[7] Y. Liang, C. Bao and J. Zhou, "An Implementaion of the CNN-Based MVDR Beamforming For Speech Enhancement," 2021 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Xi'an, China, 2021, pp. 1-5, doi: 0.1109/ICSPCC52875.2021.9564817.

[8] C. Chang, Z. Yang and J. Chen, "Plug-and-Play MVDR Beamforming for Speech Separation," ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea, Republic of, 2024, pp. 1346-1350, doi: 0.1109/ICASSP48485.2024.10445739.

[9] A. Malek et al., "Speaker extraction using LCMV beamformer with DNN-based SPP and RTF identifi cation scheme," 2017 25th European Signal Processing Conference EUSIPCO), Kos, Greece, 2017, pp. 2274-2278, doi: 10.23919/EUSIPCO.2017.8081615.

[10] S. Markovich-Golan, S. Gannot and W. Kellermann, "Combined LCMV-TRINICON Beamforming for Separating Multiple Speech Sources in Noisy and Reverberant Environments," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 2, pp. 320-332, Feb. 2017, doi: 10.1109/TASLP.2016.2633806.

[11] J. Wang, F. Yang, J. Guo and J. Yang, "Robust Adaptation Control for Generalized Sidelobe Canceller with Time-Varying Gaussian Source Model," 2023 31st European Signal Processing

Conference (EUSIPCO), Helsinki, Finland, 2023, pp. 16-20, doi: 10.23919/EUSIPCO58844.2023.10289801.
[12] S. M. Kim and H. K. Kim, "Multi-Microphone Target Signal Enhancement Using Generalized Sidelobe Canceller Controlled by Phase Error Filter," in IEEE Sensors ournal, vol. 16, no. 21, pp. 7566-7567, Nov.1, 2016, doi: 10.1109/JSEN.2016.2602375.
[13] S. M. Kim, "Hearing Aid Speech Enhancement Using Phase Difference-Controlled Dual - Microphone Generalized Sidelobe Canceller," in IEEE Access, vol. 7, pp. 130663-130671, 2019, doi: 10.1109/ACCESS.2019.2940047.
[14] R. Ali, T. van Waterschoot and M. Moonen, "Generalised Sidelobe Canceller for Noise Reduction in Hearing Devices Using an External Microphone," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018, pp. 521-525, doi: 10.1109/ICASSP.2018.8461366.
[15] M. Zohourian and R. Martin, "GSC-Based Binaural Speaker Separation Preserving Spatial Cues," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018, pp. 516-520, doi: 10.1109/ICASSP.2018.8462658.
[16] S. S. Priyanka and T. K. Kumar, "GSC Beamforming using Diff erent Adaptive Algorithmsfor Speech Enhancement," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-6, doi: 10.1109/ICCCNT45670.2019.8944415.
[17] Q. Jiang, Y. Zhou and H. Liu, "A Robust GSC for Microphone Array Using Coherence and Signal-to-Interference Ratio," 2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Ajman, United Arab Emirates, 2019, pp. 1-5, doi: 10.1109/ISSPIT47144.2019.9001775.
[18] O. Hoshuyama and A. Sugiyama, “Robust adaptive beamforming,” in Microphone Arrays, M. Brandstein and D. Ward, Eds. Springer, 2001, ch. 5, pp. 87−106.
[19] S. Gannot, D. Burshtein, and E. Weinstein, “Signal enhancement using beamforming and nonstationarity with applications to speech,” IEEE Trans. Signal Process., vol. 49, no. 8, pp. 1614 - 1626, Aug. 2001.
[20] S. M. Kim, "Hearing Aid Speech Enhancement Using Phase Diff erence-Controlled Dual-Microphone Generalized Sidelobe Canceller," in IEEE Access, vol. 7, pp. 130663-130671, 2019, doi: 10.1109/ACCESS.2019.2940047.
[21] SNRVAD. [Online]. Available: https://labrosa.ee.columbia.edu/projects/snreval/.