An Increased Robust Speech Enhancement of Superdirective Beamformer

Quan Trong The[1], Ninh Thi Thu Trang[1]

[1]Faculty of Information Security, Post and Telecommunication Institute of Technology (PTIT), Hanoi, Vietnam

Abstract

Microphone array (MA) technology commonly implemented in almost all acoustic equipments, such as, hearing aid, surveillance devices, smart phone, cochlear implant, voice - controlled device, teleconferencing system for extracting the desired target speaker while suppressing background noise, annoying recording scenario without speech distortion. MA beamformer use the priori spatial information about the designed distribution of MA, the characteristics of environment and the ability of combination with single-channel approach to obtain a steerable beampattern in a specified direction of sound source for recovering the clean speech data from the noisy mixture with high directivity index and high satisfactory speech quality. A superdirective (SDB) beamformer is one of the most helpful beamforming methods for preserving the original speech component in the diffuse noise field. SDB beamformer applied to numerous speech applications because of its advantages and easy implementation. Under realistic recording scenario, due to the complex environment, the inaccurate estimation of preferred direction of arrival (DoA) of useful signal, the error of sampling rate, the different speech sensitivity, the overall SDB beamformer's performance often corrupted. The speech distortion, musical noise, the remained noisy component usually exists and degrades the speech quality of the final output signal. In this contribution, the author proposed method for increasing robust speech enhancement of SDB beamformer in adverse environments. The numerical simulation has confirmed the effectiveness of the author's suggested technique in improving the speech quality by removing musical noise, background noise at SDB beamformer's output signal.

Keywords

Microphone array, speech enhancement, speech quality, noise reduction, beamforming.

1. Introduction

Nowadays, the using of MA technology has been popular and significantly increases the overall speech enhancement of various types of speech applications. MA use the spatial information about designed geometry, the preferred steering vector, the characteristic of surrounding noise, the properties of background noise, the coherence of captured MA signals to obtain high - directional beampattern towards the sound source while suppressing interference, third - party talker, as in Figure 1. MA beamforming, which owns the advantage of noise reduction and speech enhancement simultaneously, allows achieving the original speech component without distortion. Compared with single - channel approach, MA beamforming has the flexible working with different recording scenarios, such as, coherent/incoherent, diffuse and other complex situations.

MA beamforming technique can be categorized into two groups: the fixed beamformer and adaptive beamformer. The fixed beamformer bases on the prior information of incident angle of helpful signal to achiveve the beamformer's coefficient. Fixed beamformer includes Delay - and - sum (DAS) beamformer [1], which often used in almost digital signal processing system. Adaptive beamformer exploit the properties of recorded data, the rapid changed environmental factors, the constrained criteria of minimizing the total output noise power for recovering the clean speech data. Differential microphone array (DIF) [2-3], Minimum Variance Distortionless Response (MVDR) [4-6], Linearly Constrained Minimum Variance (LCMV) [7-8], Generalized sidelobe canceller (GSC) [9-10] and SDB Beamformer [11-19] is the most helpful beamforming technique, which common installed into multi-channel signal processing system to achieve the original speech data while removing total noise with high speech quality, perceptual metric listener.

[1]CMIS-2025: Eighth International Workshop on Computer Modeling and Intelligent Systems, May 5, 2025, Zaporizhzhia, Ukraine
theqt@ptit.edu.vn (Quan Trong The); trangntt2@ptit.edu.vn (Ninh Thi Thu Trang);
0000-0002-2456-9598 (Quan Trong The); 0009-0009-5214-6144 (Ninh Thi Thu Trang);

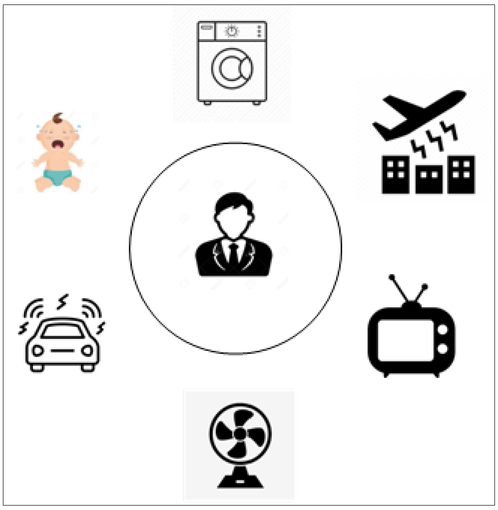

Figure 1: The complex and annoying environment around the human - life

As in Figure 2, SDB beamformer uses the properties of diffuse noise field to extract the desired talker at certain location with high directivity index. However, due to the complex and annoying environment, the error of estimation the preferred steering vector, the displacement of MA distribution, the overall SDB beamformer's performance often decreased. Consequently, speech distortion, musical noise still be challenging problem. There are numerous efforts were studied for dealing this task. The scheme of MA beamforming is given by Figure 3.

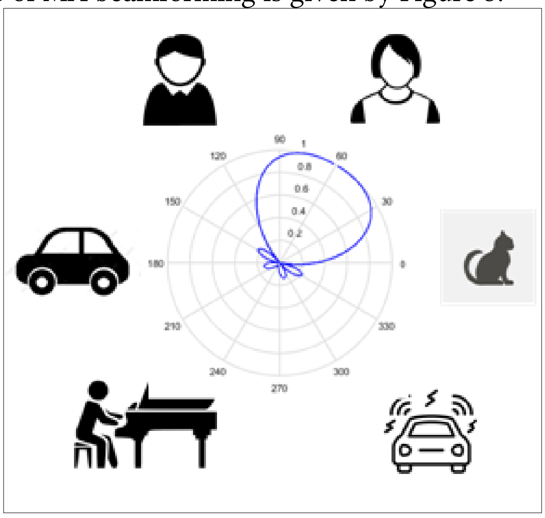

Figure 2: The complex and annoying environment around the human - life

Atkins A [11] proposed an efficient designed beamformer with trade-off between high directivity and low white noise amplification. The numerical simulations demonstrate controlled tuning of various gain properties of speech enhancement and increase the overall SDB evaluation in realistic situations.

Berkun R [12] uses a tunable regularization parameter, which addresses the problem of directivity factor (DF) and white noise gain (WNG) of SDB beamformer. The conducted experiment has confirmed the effectiveness of the author's suggested approach in adjusting DF and WNG.

In [13], the author suggested scheme combination of DAS beamformer and regularized SDB beamformer to reach high directivity factors while suppressing background noise. This direction research derives analytic closed-form expressions of the beamformer gain with controlling WNG or DF.

Wang J [14] suggested method for designing the fractional - order SDB beamformer to obtain the high maximum directivity factor in the spherical and isotropic noise field. The promising result has verified the advantage of this approach.

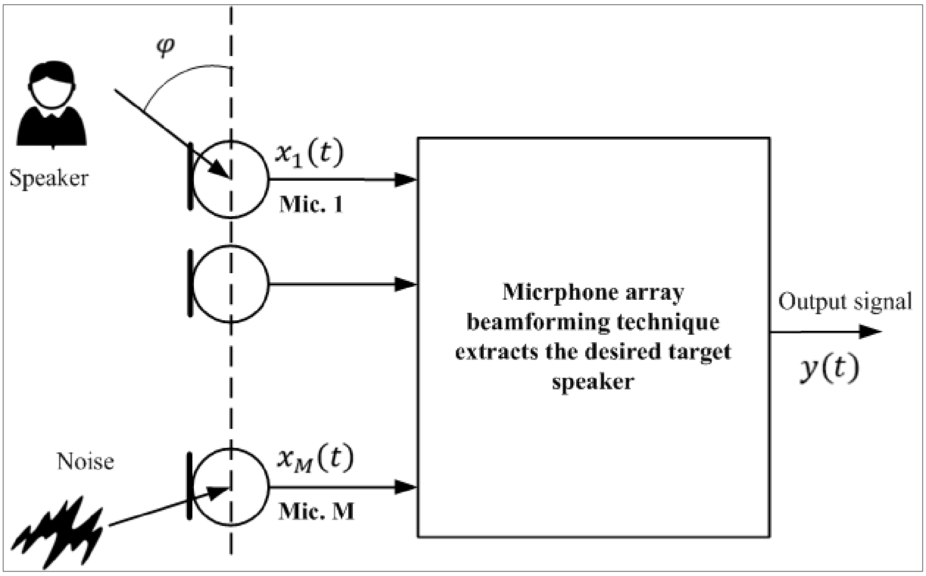

Figure 3: The principal working of microphone array beamforming

Yang X [15] addressed the problem of imperfections, such as adverse and self-noise, sensor mismatches by applying a joint optimization approach with efficient post - Filtering and robust SDB beamformer, which based on constrained WNG or utilizes diagonal loading technique. Experimental results show the improved robustness of SDB beamformer under various recording scenarios.

Gong P [16] presented an optimization of problem for designing beamformer by utilizing constrained criteria of WNG and least - squares. The author uses alternative direction penalty method (ADPM) algorithm to solve and achieve robust signal processing system of SDB beamformer.

In [17], Chen X studied the robustness of SDB beamformer and derived optimum different solutions to resolve array imperfections. The author approach utilized quadratic eigen value (QEP) to obtain the maximum possible DF and constrained WNG.

A data-driven approach, which controls WNG threshold for obtaining robustness speech enhancement, was presented [18]. The suggested method outperformed speech enhancement, noise reduction, high fidelity of the desired acoustic signal in comparison with traditional SDB beamformer.

Huang G [19] proposed diagonalization of the noise pseudo-coherence matrix of the desired/noise signals and Fourier matrix to properly select the dimension of subspace and more flexible WNG and DF than the conventional regularized SDB beamformer.

However, these above approaches performed in laboratory condition with proper environmental factors. Due to the complex and annoying recording situation, the displacement of MA geometry, the inaccurate estimation of preferred steering vector, the error of sampling frequency, the microphone mismatches, the different microphone sensitivities, SDB beamformer's evaluation usually degraded. In this paper, the author introduces modifying covariance matrix of complex diffuse noise field and enhanced steering vector to obtain more robustness of SDB beamformer. The author's method improves the real-time SDB beamformer's performance without processing large database.

2. Superdirective beamformer

In this section, the principal working of SDB beamformer will be presented. The author uses dual - microphone array (DMA2) system to describe the representation of observed MA signals in Figure 4. In the currently considered frame k, the frequency f, the received array signals $X_1(f,k), X_2(f,k)$ can be formulated in the short-time Fourier transform (STFT) as:

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

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

Where $S(f,k)$ is the original speech component, $N_1(f,k), N_2(f,k)$ is the additive noise, $\Phi_s = \pi f \tau_0 \cos(\theta_s)$, θ_s is the preferred steering vector of the interest useful signal to axis of DMA2, $\tau_0 = d/c$, d is the distance between two mounted microphones, $c = 343(m/s)$ is sound speed propagation in the fresh air.

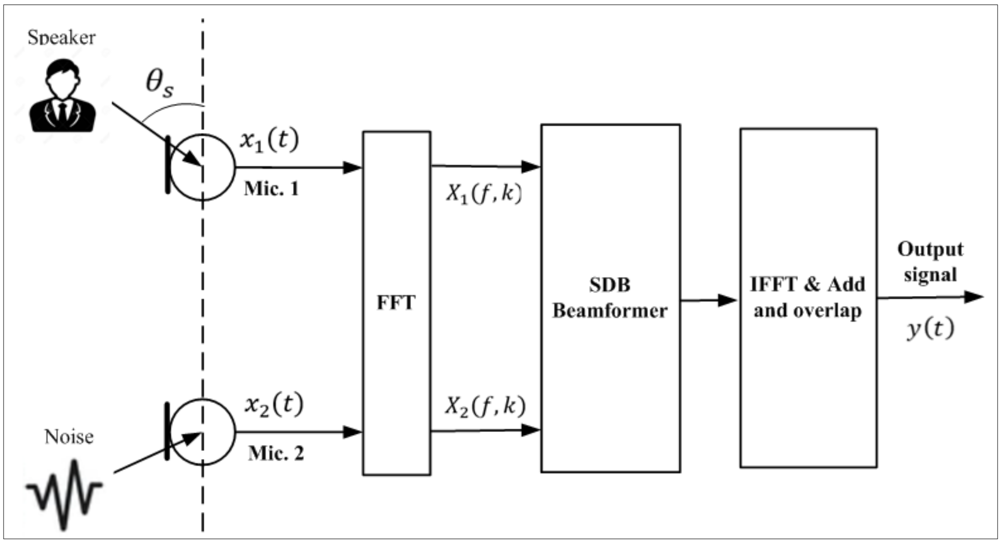

Figure 4: The scheme of SDB beamformer in the frequency - domain

If we denote $X(f,k) = \left[X_1(f,k) \quad X_2(f,k)\right]^T$, $N(f,k) = \left[N_1(f,k) \quad N_2(f,k)\right]^T$, $D_s(f,\theta_s) = \left[e^{j\Phi_s} \quad e^{-j\Phi_s}\right]^T$, T is transpose operator, the equation (1)-(2) can be rewritten as:

$$X(f,k) = S(f,s)D_s(f,\theta_s) + N(f,k) \tag{3}$$

The essential core problem of speech enhancement is determining an optimum coefficient $W(f,k)$ for obtaining approximate clean speech data.

$$\hat{S}(f,k) = W^H(f,k)X(f,k) \tag{4}$$

where H is conjugate operator.

And $\hat{S}(f,k) \approx S(f,k)$.

In the diffuse noise field, SDB beamformer's weight can be computed as:

$$W_{SDB}(f,k) = \frac{\Gamma^{-1}(f)D_s(f,\theta_s)}{D_s^H(f,\theta_s)\Gamma^{-1}(f)D_s(f,\theta_s)} \tag{5}$$

$$\Gamma(f) = \begin{bmatrix} 1 & \dfrac{\sin(\omega\tau_0)}{\omega\tau_0} \\ \dfrac{\sin(\omega\tau_0)}{\omega\tau_0} & 1 \end{bmatrix} \tag{6}$$

Where $\omega = 2\pi f$.

SDB beamformer especially effective in diffuse noise field with high directional beampattern towards the sound source. Because of the complex and annoying environment, the displacement of MA distribution, the inaccurate estimation of incident angle of the impinging helpful signal, the microphone mismatches, the different microphone quality, the error of sampling frequency, the moving head of speaker during a conversation, the existence of non-directional noise or undetermined reason, SDB beamformer's performance often degraded. Consequently, the musical noise or speech distortion occurs and significantly effects speech quality.

For overcoming this drawback, in the next section, the author proposed an effective modified SDB beamformer for decreasing the speech distortion, musical noise, noise level and increasing the perceptual metric listener and speech intelligibility.

3. The author's proposed method

The author's ideal is enhancing the accurate estimation of steering vector $D_s(f,\theta_s)$ and modifying the covariance matrix $\Gamma(f)$, which according to the rapidly changing environmental factors.

Due to the complex environment, the steering vector can be expressed in the way:

$$D_s(f,\theta_s) = \left[e^{j\Phi_s} \quad (1-\beta)e^{-j\Phi_s} \right]^T \tag{7}$$

where β presents the distortion of clean speech data at second microphone in comparison with first microphone and $0 < \beta < 1$. In this paper, the author uses the information of standard deviation of diagonal loading of noisy covariance matrix, and $\beta = \mathrm{std}(\mathrm{diag}(\Psi_{NN}(f,k))$ [20], where $\Psi_{NN}(f,k) = E\left\{ N^H(f,k)N(f,k) \right\}$ is the covariance matrix of noise. β leads to 0 and β presents the error of representation of speech component at the observed microphone array signals. The noisy covariance matrix can be computed by using Voice Activity Detection (VAD) [21] for determining whether frame is the only noise.

As we know that, the coherence between two points in diffuse noise field can be formulated as $\Gamma_{X_1X_2}(f) = \dfrac{\sin(2\pi f \tau_0)}{2\pi f \tau_0}$. The author proposed using the speech presence probability $SPP(f,k)$ [22] to adjust $\Gamma_{X_1X_2}(f)$ as the following equation:

$$\Gamma_{X_1X_2}(f) = \frac{\sin(2\pi f \tau_0)}{\left(1+\left(1-SPP(f,k)\right)\dfrac{\sigma_n^2}{P_{nn}(f,k)}\right)2\pi f \tau_0} e^{j2\Phi_s} \tag{8}$$

where σ_n^2 is covariance of uncorrelated noise, $P_{nn}(f,k)$ is spectral density of diffuse noise field [23].

The equation (7) contains the steering vector and ensure the robustness of the coherence between two points in diffuse noise field. Therefore, the coherence matrix can be modified as the following equation:

$$\Gamma(f) = \begin{bmatrix} 1 & \Gamma_{X_1X_2}(f) \\ \Gamma_{X_1X_2}(f) & 1 \end{bmatrix} \tag{9}$$

With the improved steering vector and modified coherence matrix in complex diffuse noise field according to the rapidly changing recording scenario, the author's proposed technique can be applied for adaptively achieving the optimum coefficient.

4. Experiments

In this section, the author aims at demonstrating the effectiveness of the suggested technique (SDB-sgt) in enhancing SDB beamformer's robustness speech enhancement with increasing the speech quality in the term of the signal - to - noise ratio, reducing musical noise and noise level at SDB beamformer's output signal. The experiment was conducted in living room with the existence of interference, non - directional noise, third - party talker, washing machine and other undetermined sources. A talker stands at distance $L = 4(m)$, the incident angle of helpful signal to the axis of dual - microphone system (DMA2) is $\theta_s = 90(\deg)$, the distance between two mounted microphones is $d = 5(cm)$. For capturing the clean speech data, these parameters were set: $nFFT = 512$, the sampling frequency $Fs = 16kHz$, overlap 50%. An objective measurement [24] was implemented for calculating the speech quality in the term of signal-to-noise (SNR) ratio. The scheme of demonstrated experiment is shown in Figure 5.

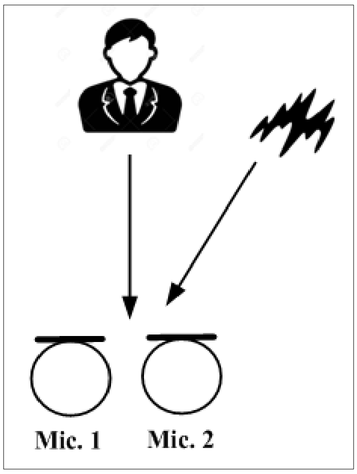

Figure 5: The conducted experiment to illustrate the advantage of the author's proposed method
The waveform of microphone array signals is given in Figure 6 and Figure 7.

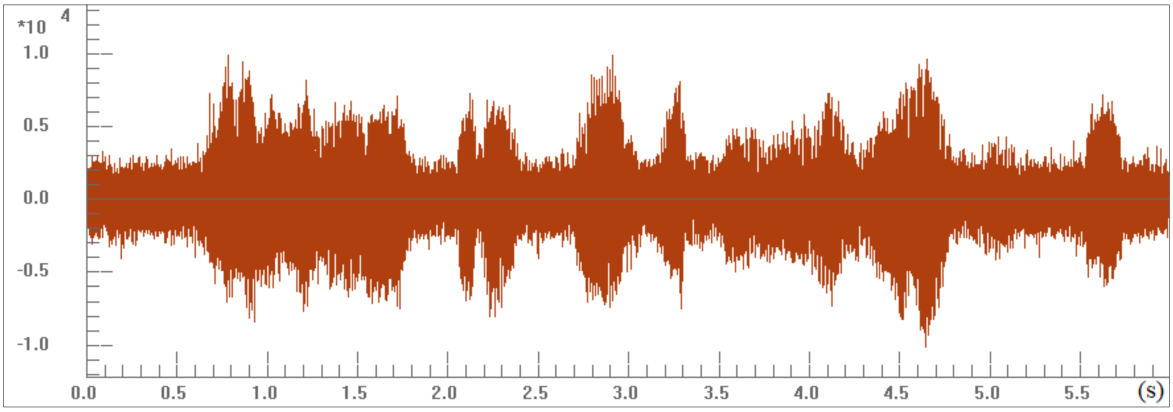

Figure 6: The waveform of received array signals

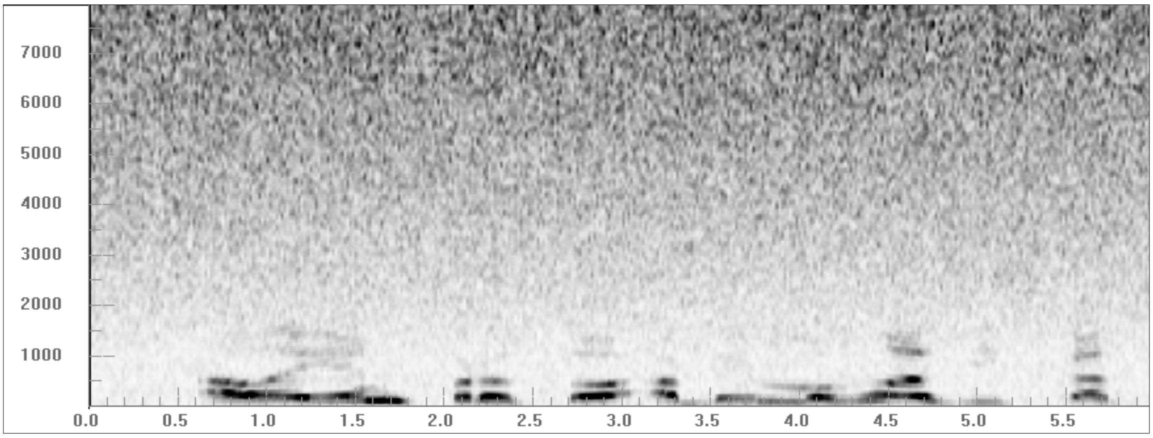

Figure 7: Spectrogram of received array signals

For further signal processing, smoothing parameter $\alpha = 0.1$ was used for implementing SDB beamformer to extract the target speaker. The obtained results are given in Figure 8 and Figure 9.

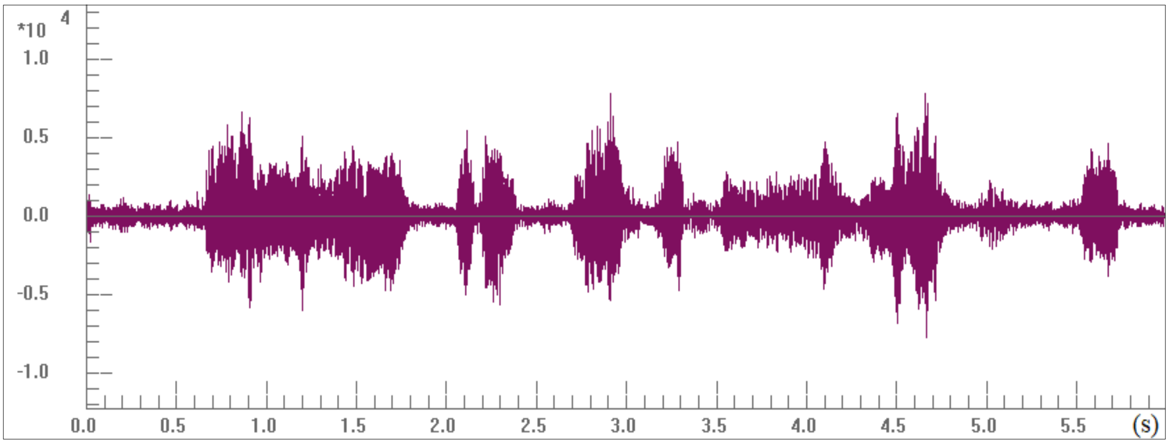

Figure 8: The processed signal by applying SDB - beamformer

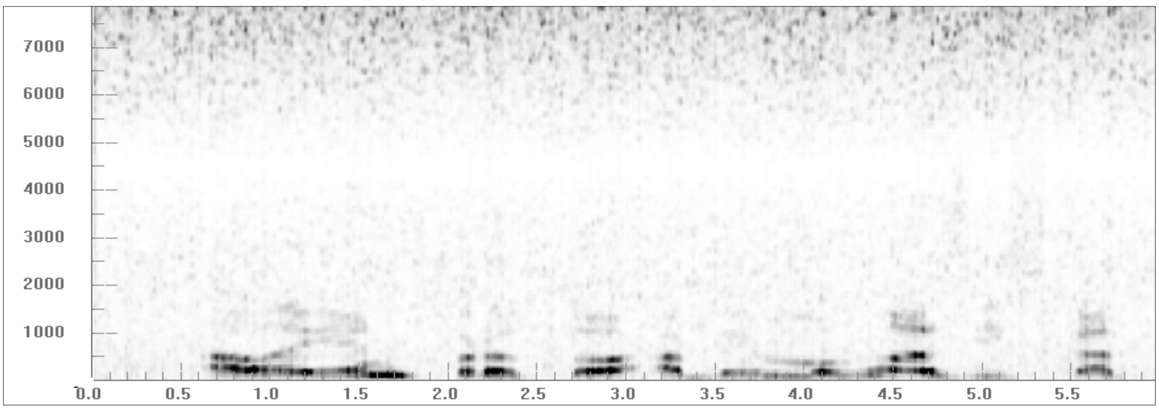

Figure 9: Spectrogram of SDB – beamformer's output signal

Because of the error of sampling rate, the different microphone sensitivities, the displacement of designed microphone distribution, the inaccurate estimation of preferred steering vector, the moving head of speaker, the adverse environment, SDB beamformer's evaluation often degraded. The existence of speech distortion, musical noise, remained noise corrupt the speech quality. Therefore, the author suggested improving the accurate calculation of steering vector and modified the noisy coherence matrix increase SDB beamformer's performance in adverse environment.

The promising result of SDB-sgt is given in Figure 10 and Figure 11. Figure 12 describes the comparison of energy between the observed microphone array signals, the processed signals by SDB-beamformer, SDB-sgt.

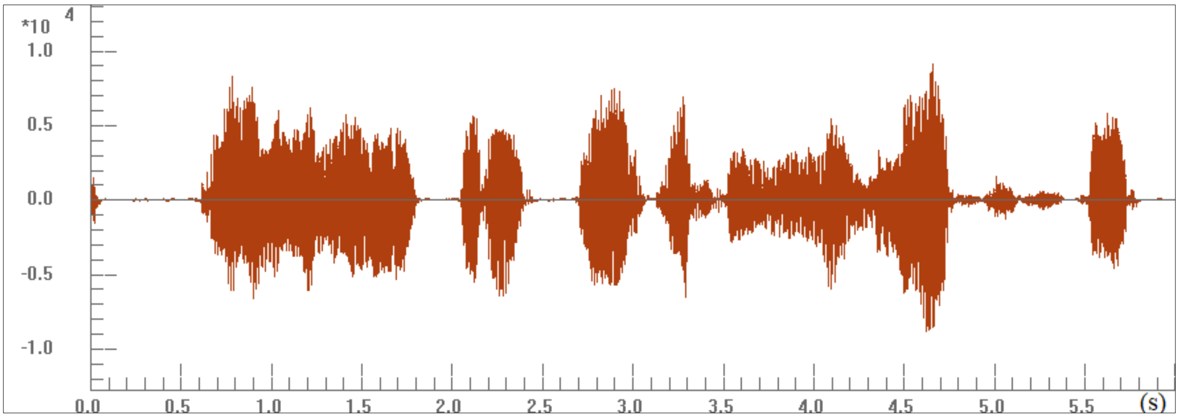

Figure 10: The processed signal by utilizing SDB-sgt

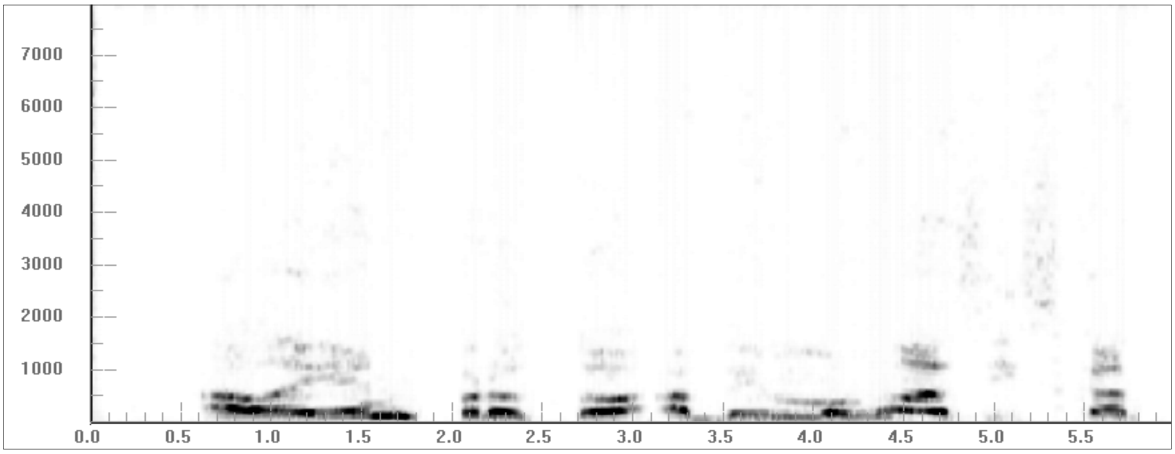

Figure 11: Spectrogram of SDB-sgt's output signal

The advantage of SDB-sgt has been confirmed with the decreasing the musical noise, noise reduction to 9.5 dB and Table 1 shows the increased the SNR from 10.9 to 12.7 dB.

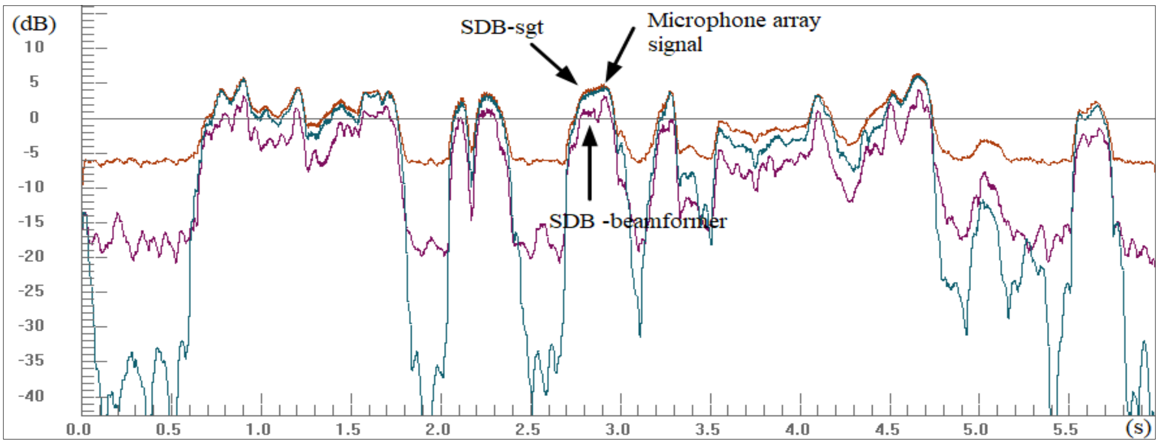

Figure 12: The comparison energy between microphone array signal and processed signals by SDB - beamformer, SDB - sgt

Table 1

The comparison speech quality (dB)

Method	Microphone	SDB-beamformer	SDB-sgt

Estimation	array		
NIST STNR	6.5	19.2	30.1
WADA SNR	3.6	12.7	25.4

The effectiveness of the author's proposed method was illustrated through the numerical simulations. The approach modifies the steering vector and the formulation of coherence between two points sources, which exploit the speech presence probability and standard covariance matrix of observed microphone array signals. The modified covariance matrix and enhanced steering vector adaptively change according to the characteristics of surrounding noise, the complex recording scenario. Therefore, the updating these necessary parameters allows improving the robustness of SDB beamformer in complex diffuse noise field. Steering vector, which contains the direction and source location, plays an important role in SDB beamformer in extracting the desired target speaker. Accurate estimation of steering vector and modified covariance matrix of complex diffuse noise allow increasing SDB beamformer's performance in realistic recording environment.

5. Conclusion

In this contribution, the author proposed an effective modified steering vector and covariance matrix of diffuse noise field in presence of annoying recording scenario. The obtained numerical results have showed the effectiveness of musical noise reduction, noise level suppression and increasing the speech quality in the term of signal-to-noise ratio from 10.9 to 12.7 dB. The author's approach based on adaptive tracking and adjusting the steering vector and covariance matrix according to the simultaneous changing environmental factors. The suggested technique incorporates the speech presence probability to enhance the robustness of computing the necessary parameter for determining SDB beamformer's coefficient. The above method can be integrated into multi-channel system for dealing other complicated problems, such as speech recognition, reverberation.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] F. Prieur, O. M. H. Rindal and A. Austeng, "Signal Coherence and Image Amplitude With the Filtered Delay Multiply and Sum Beamformer," in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 65, no. 7, pp. 1133-1140, July 2018, doi: 10.1109/TUFFC.2018.2831789.

[2] 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.

[3] X. Zhao, X. Luo, G. Huang, J. Chen and J. Benesty, "Differential Beamforming with Null Constraints for Spherical Microphone Arrays," 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.

[4] X. Wang, J. Benesty, G. Huang and J. Chen, "A Minimum Variance Distortionless Response Spectral Estimator with Kronecker Product Filters," 2022 30th European Signal Processing Conference (EUSIPCO), Belgrade, Serbia, 2022, pp. 2261-2265, doi: 10.23919/EUSIPCO55093.2022.9909584.

[5] Y. Kuwahara and F. Kimihito, "Breast Imaging by Time Reverse Minimum Variance Distortionless Response," 2020 IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting, Montreal, QC, Canada, 2020, pp. 1341-1342, doi: 10.1109/IEEECONF35879.2020.9330085.

[6] S. Athira, S. Sahoo, S. Benny and V. Chandrasekar, "Minimum Variance Distortionless Response based Distributed Beamforming Technique for C-Band Weather Radar," 2023 IEEE India Geoscience and Remote Sensing Symposium (InGARSS), Bangalore, India, 2023, pp. 1-4, doi: 10.1109/InGARSS59135.2023.10490350.

[7] A. Aroudi and S. Doclo, "Cognitive-driven Binaural LCMV Beamformer Using EEG-based Auditory Attention Decoding," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 406-410, doi: 10.1109/ICASSP.2019.8683635.

[8] N. Gößling, E. Hadad, S. Gannot and S. Doclo, "Binaural LCMV Beamforming With Partial Noise Estimation," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2942-2955, 2020, doi: 10.1109/TASLP.2020.3034526.

[9] 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," 2021 IEEE UFFC Latin America Ultrasonics Symposium (LAUS), Gainesville, FL, USA, 2021, pp. 1-4, doi: 10.1109/LAUS53676.2021.9639164.

[10] 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.

[11] A. Atkins, Y. Ben-Hur, I. Cohen and J. Benesty, "Robust superdirective beamformer with optimal regularization," 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC), Xi'an, China, 2016, pp. 1-5, doi: 10.1109/IWAENC.2016.7602963.

[12] R. Berkun, I. Cohen and J. Benesty, "A tunable beamformer for robust superdirective beamforming," 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC), Xi'an, China, 2016, pp. 1-5, doi: 10.1109/IWAENC.2016.7602952.

[13] R. Berkun, I. Cohen and J. Benesty, "Combined Beamformers for Robust Broadband Regularized Superdirective Beamforming," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 5, pp. 877-886, May 2015, doi: 10.1109/TASLP.2015.2410139.

[14] J. Wang, F. Yang and J. Yang, "A General Approach to the Design of the Fractional-Order Superdirective Beamformer," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 70, no. 11, pp. 4291-4295, Nov. 2023, doi: 10.1109/TCSII.2023.3287918.

[15] X. Yang, G. Huang, J. Jin, J. Chen and J. Benesty, "Design and Optimization of Superdirective Beamforming and Post-Filtering for Speech Enhancement," ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 2025, pp. 1-5, doi: 10.1109/ICASSP49660.2025.10890758.

[16] P. Gong, J. Zhang, Y. Wu, L. Yu and L. Li, "Robust Superdirective Beamforming Based on ADPM," 2022 5th International Conference on Information Communication and Signal Processing (ICICSP), Shenzhen, China, 2022, pp. 723-727, doi: 10.1109/ICICSP55539.2022.10050544.

[17] X. Chen, J. Benesty, G. Huang and J. Chen, "On the Robustness of the Superdirective Beamformer," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 838-849, 2021, doi: 10.1109/TASLP.2021.3053410.

[18] H. Pei et al., "Data-Driven White Noise Gain Constrained Robust Superdirective Beamformer for Speech Enhancement," ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 2025, pp. 1-5, doi: 10.1109/ICASSP49660.2025.10889723.

[19] G. Huang, J. Benesty and J. Chen, "Subspace superdirective beamforming with uniform circular microphone arrays," 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC), Xi'an, China, 2016, pp. 1-5, doi: 10.1109/IWAENC.2016.7602915.

[20] 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.

[21] N. Shankar, G. S. Bhat and I. M. S. Panahi, "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.

[22] T. Gerkmann, M. Krawczyk and R. Martin, "Speech presence probability estimation based on temporal cepstrum smoothing," 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA, 2010, pp. 4254-4257, doi: 10.1109/ICASSP.2010.5495677.
[23] 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.
[24] SNRVAD. [Online]. Available: https://labrosa.ee.columbia.edu/projects/snreval/.