Noise reduction by using more reference signals in GSC beamformer

Quan Trong The[1], Nguyen Thi Huyen Chau[2] and Nguyen Manh Son[3]

[1]Lab Blockchain, Faculty of Information Security, Posts and Telecommunications Institute of Technologies (PTIT), 122 Hoang Quoc Viet Str., Nghia Do Ward, Hanoi, 100000, Vietnam

[2]Faculty of Information Technology, Thang Long University, Nghiem Xuan Yem Str., Dai Kim Ward, Hanoi, 100000, Vietnam

[3]Faculty of Information Technology, Posts and Telecommunications Institute of Technologies (PTIT), 122 Hoang Quoc Viet Str., Nghia Do Ward, Hanoi, 100000, Vietnam

Abstract

Nowadays, the use of microphone array (MA) technology has been popular, due to its high spatial diversity, high directivity factor and the steered beampattern towards the specified sound location. MA beamforming uses the spatial information about the direction of arrival of a helpful signal, the designed geometry of MA distribution, the properties of surrounding noise, background noise to achieve noise reduction and speech enhancement at the same time. Generalized Sidelobe Canceller (GSC) beamformer is one of the most useful techniques for minimizing the total noise power while preserving the clean speech data at a certain direction without speech distortion. However, under complex recording scenarios, the overall MVDR's evaluation often degraded and speech quality, speech intelligibility corrupted because of the moving head of speaker, the different microphone sensitivities, the error of sampling rate, the displacement of MA. In this article, we introduced a modified steering vector to increase the robustness of GSC beamformer in realistic recording situations. This approach is based on the preferred steering vector to achieve more reference signal to improve GSC beamformer's performance in complex recording scenarios. The demonstrated experiments have confirmed the advantage of the proposed method in reducing musical noise, noise level at GSC beamformer's output signal. The conducted experiment shows the improved signal-to-noise ratio (SNR) from 7.4 to 8.8 (dB). The suggested technique can be implemented in a multi-channel system for dealing other complicated problems, such as reverberation, speech recognition.

Keywords

generalized sidelobe canceller, beamforming, signal-to-noise ratio, robust, speech enhancement

1. Introduction

In numerous acoustics equipment, for example, hearing aids, listening devices, voice-controlled devices, the captured speech signals are usually corrupted by unwanted acoustic factors, such as, third-party talker, interference, diffuse noise field, incoherent noise field, which often corrupts the speech quality. In the single-channel approach, the popular spectral subtraction method is often applied due to its easy implementation. However, in a rapidly changing noisy environment, the properties of background noise can not be determined or calculated. Therefore, the speech distortion or musical noise often occurs. As in figure 1, there are various types of noise, which significantly affects the speech intelligibility.

MA beamforming [1, 2] owns the capability of enhanced speech enhancement and noise reduction at the same time without speech distortion. MA beamformer often uses the prior spatial information, such as, the designed MA geometry, the characteristics of surrounding noise, the coherence of observed MA signals, the preferred steering vector of target signal to the MA distribution to obtain an improved signal processing. Consequently, MA beamformer utilize different methods and techniques to increase performance in a realistic recording environment.

As in figure 2, MA beamforming technique concerns the steerable beampattern at specified direction on the desired speaker. The beampattern suppresses the other signals from different directions

CS&SE@SW 2025: 8th Workshop for Young Scientists in Computer Science & Software Engineering, November 21, 2025, Kryvyi Rih, Ukraine
https://doi.org/10.55056/cssesw2025/paper11
theqt@ptit.edu.vn (Q. T. The); huyenchau@thanglong.edu.vn (N. T. H. Chau); sonnm@ptit.edu.vn (N. M. Son)
0000-0002-2456-9598 (Q. T. The); 0000-0003-4091-0271 (N. T. H. Chau); 0009-0007-4005-5646 (N. M. Son)

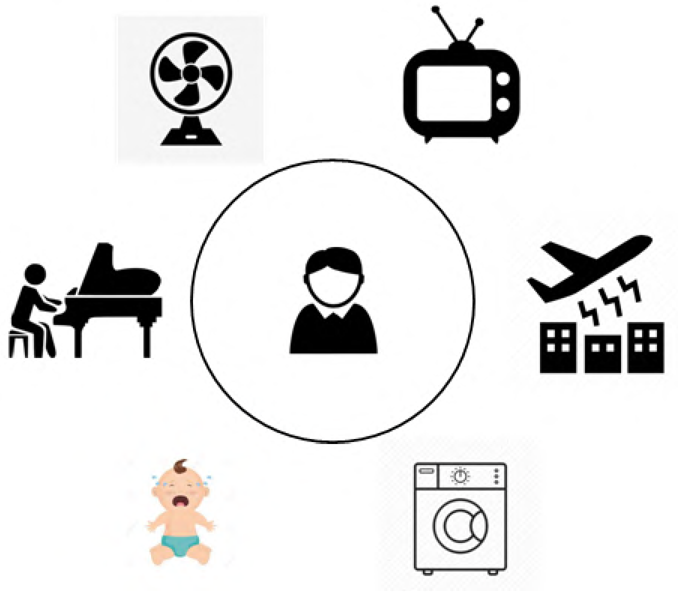


Figure 1: There are various types of noise seriously affected on human-life.

while preserving the only speech component at a certain location. Beamformer has the advantage of incorporating pre-processing, post-filtering or different methods for improving the speech enhancement.

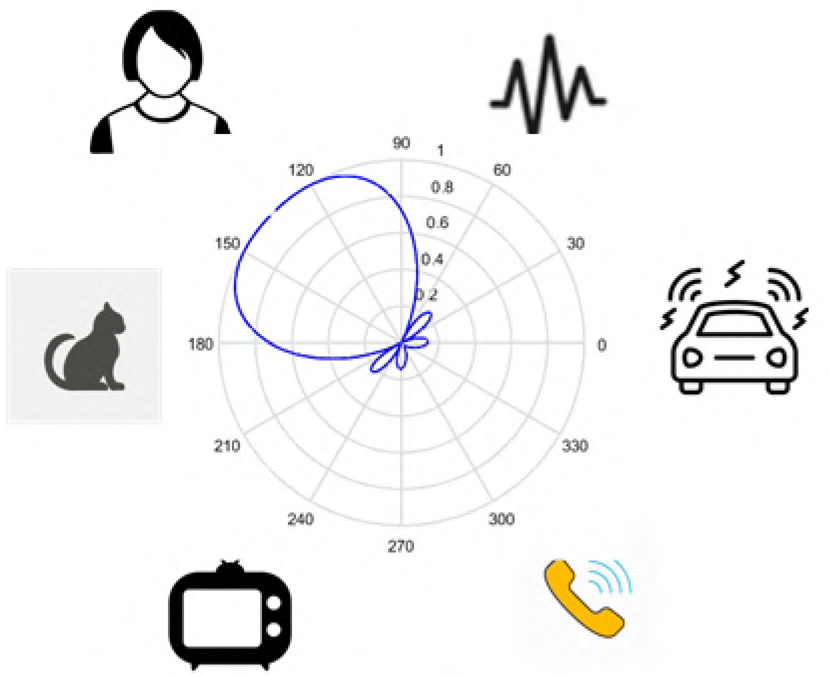


Figure 2: The promising beampattern concerns on target talker while eliminating surrounding noise.

A well-known multi-channel approach is GSC beamformer, which is an efficient solution for extracting the clean speech data while removing the background noise. GSC beamformer often installed into various types of acoustic equipment. However, in a realistic recording scenario, the error of microphone sensitivities, the error of sampling rate, the moving head of speaker during conversation, the inaccurate MA distribution, the complex and annoying situation, which often decreases the evaluation of GSC beamformer. There many efforts attempt to enhance the beamformer's performance in different scenarios.

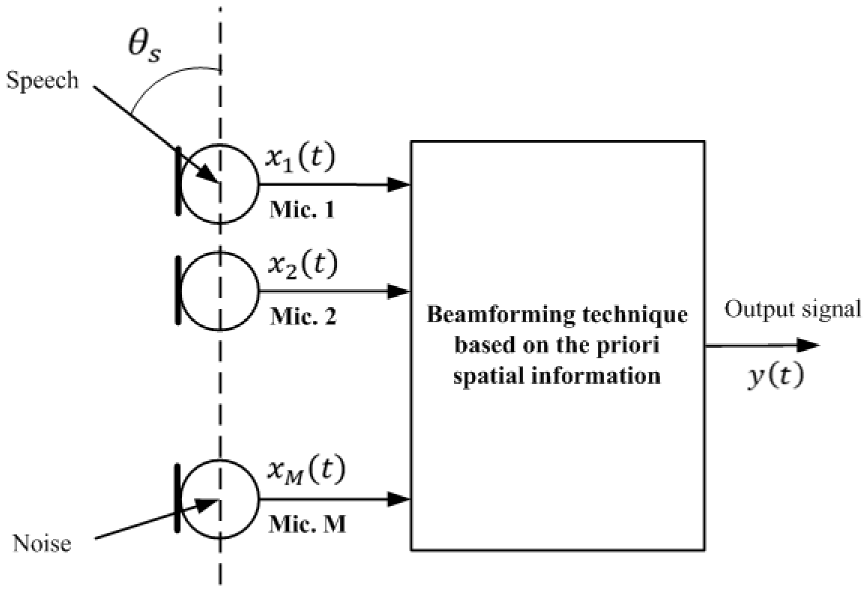


Figure 3: The structure of MA beamforming to process array received signals.

In [3], the authors present different adaptive algorithms to enhance the speech quality of recorded signals in a complex and noisy environment. The numerical simulation was conducted with Least Mean Square (LMS), Normalized LMS and Recursive Least Square (RLS) to increase the speech enhancement and noise suppression in various types of noise. The obtained signal-to-noise ratio (SNR), perceptual evaluation of speech quality (PESQ) and Log Likelihood Ratio (LLR) showed the effectiveness of the author's approach.

Barnov et al. [4] proposed a technique for controlling the white noise gain (WNG) to achieve a spatially robust GSC beamformer in an annoying recording scenario. The numerical simulation has verified the advantage of the described method on overcoming the mismatch between the preferred related transfer function (RTF) and actual calculated RTF while remaining acceptable speech quality in the term of WNG.

Jiang et al. [5] introduced using the signal-to-interference ratio (SINR) between the output of upper branch and lower branch to adaptively adjust the smoothing parameter of GSC beamformer at a certain frequency range. The experiment results showed the robustness of the beamformer in reducing different noise interference at different directions and achieving better speech enhancement.

Zohourian and Martin [6] proposed a GSC beamformer with maximum likelihood technique for addressing the preservation of spatial cues in separation speaker problems. The method provided an estimated desired talker as target presence probability to compute spectral filters. The numerical results have shown the effectiveness in reducing background noise and other interferences in realistic recording situations.

Ali et al. [7] suggested exploiting an external microphone to combine with GSC beamformer to benefit for noise suppression, which is critical for numerous acoustic speech applications. The experiment results have shown the ability of improving the noise reduction suppression.

Li et al. [8] proposed implementing variable step size least mean square algorithm to control the step in GSC beamformer for decreasing the influence of diffuse noise field and adverse recording noisy condition. The author's suggested method has shown the efficient performance of the proposed technique in a diffuse noise environment.

Based on the assumption that the speech component is distributed by time-varying Gaussian model, Wang et al. [9] used a robust update rule for adaptive interference canceller to alleviate the desired target speech distortion in a realistic recording environment.

However, the above-mentioned research usually works well in laboratory conditions. In realistic recording scenario, because of the adverse environment, the different microphone sensitivities, the error of sampling frequency, the microphone mismatches, the error of microphone array configuration, the displacement of standing speaker to MA, the inaccurate estimation of preferred steering vector, the moving head of talker, the existence of surrounding transport vehicle, the overall GCS beamformer often degraded. The speech distortion and remaining noise cause unsatisfied perceptual metric listeners. In this contribution, we presents a prospective approach to increase GSC beamformer's performance in suppressing the remaining noise at the output signal.

2. Generalized sidelobe canceller beamformer

The traditional GSC beamformer contains three parts. The upper branch often steers the designed beampattern at specified direction to achieve the mixture of clean speech data and background noise. The upper branch usually applies a Delay-and-Sum beamformer to take into account the main signal. The lower branch is used for extracting the only noise component, which plays an important role in adaptive noise reduction algorithms. We uses a blocking matrix to block the speech component. The output of the blocking matrix is a reference signal. The essential part of GSC beamformer is adaptive noise canceller, which recovers the original speech component from the main signal by using Wiener filter, which uses the reference signal as noise, to suppress interference signal and competitive talker. The scheme of GSC beamformer is described in figure 4.

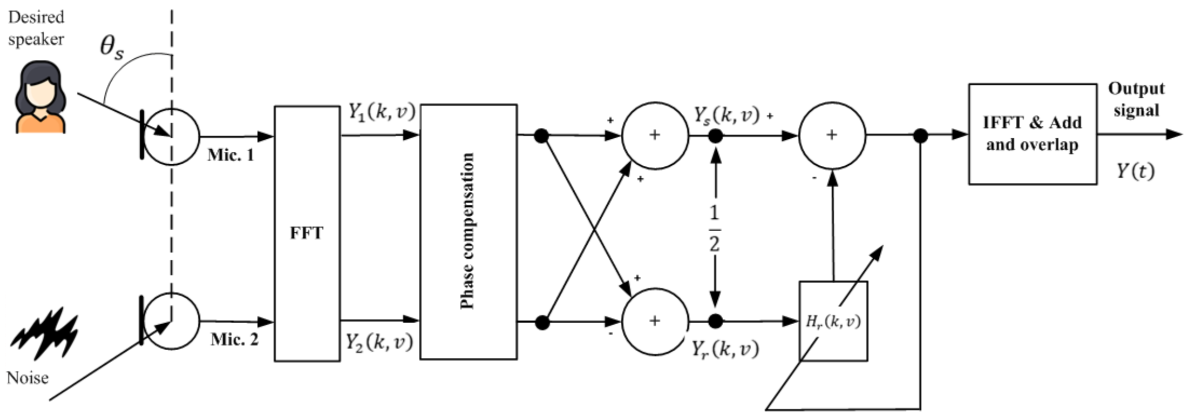


Figure 4: The structure of GSC beamformer in the frequency-domain.

Two received array signals $Y_1(k,v), Y_2(k,v)$ at current frame k, frequency v can be formulated as:

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

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

where d is the range between two microphones, $c = 343(m/s)$ means the sound speed propagation in the fresh air, $\tau_0 = d/c$ denotes the time delay of the interest signal to the microphones, θ_s represents the impinging angle of useful signal to the axis of DM2, $N_1(k,v), N_2(k,v)$ is unwanted noise, non-directional noise, interference or third-party speaker, $S(k,v)$ is the clean speech data.

The output of upper branch and lower can be expressed as the following equations:

$$Y_s(k,v) = \frac{Y_1(k,v)e^{-j\Phi_s} + Y_2(k,v)e^{j\Phi_s}}{2} \tag{3}$$

$$Y_r(k,v) = \frac{Y_1(k,v)e^{-j\Phi_s} - Y_2(k,v)e^{j\Phi_s}}{2} \tag{4}$$

The adaptive noise canceller (ANC) is an essential part of GSC beamformer for extracting the desired target speaker while suppressing surrounding noise and other interference. The Wiener filter is often implemented for removing background noise.

Wiener filter based on the auto - cross power spectral densities (PSD) between the main signal and reference signal as:

$$P_{YsYr}(k,v) = (1-\alpha)P_{YsYr}(k,v-1) + \alpha Y_s(k,v)Y_r^*(k,v) \tag{5}$$

$$P_{YrYr}(k,v) = (1-\alpha)P_{YrYr}(k,v-1) + \alpha Y_r(k,v)Y_r^*(k,v) \tag{6}$$

where α denotes an appropriate smoothing parameter in range $0..1$.

So, the adaptive noise canceler's filter is defined as:

$$H_r(k,v) = \frac{P_{YsYr}(k,v)}{P_{YrYr}(k,v)} \tag{7}$$

The promising final GSC beamformer's output can be derived as:

$$Y_{GSC}(k,v) = Y_s(k,v) - Y_r(k,v) * H_r(k,v) \tag{8}$$

However, in realistic recording scenario, due to the complex environment, the undetermined moving head of talker, the difference of microphone sensitivities, the error of sampling rate, the imprecise MA configuration, the existence of competitive talker, the inaccurate estimation of transfer function, the overall GSC beamformer's evaluation often corrupted. The remaining noise, speech distortion reduce the speech quality and speech intelligibility. Therefore, in this next section, the authors propose an effective approach for eliminating the noise level of GSC beamformer's output signal to enhance the overall signal-to-noise ratio.

3. Our approach

Our idea is to generate more reference signals to improve the effectiveness of Wiener filters in reducing the remaining noise. From the described formulation of received array signals, we can easily calculate the additive reference signal with following equations:

$$Y_p(k,v) = \frac{Y_1(k,v)e^{-j2\Phi_s} - Y_2(k,v)}{2} \tag{9}$$

and:

$$Y_q(k,v) = \frac{Y_2(k,v)e^{j2\Phi_s} - Y_1(k,v)}{2} \tag{10}$$

Because of complex and annoying recording situations, the inaccurate preferred steering vector, the existence of different noise, such as, coherent noise, diffuse noise, incoherent noise, the output of blocking matrix can not contain the only coherent noise to remove total noise component at the final output signal.

By using the prior information of the preferred direction of arrival of interesting useful signals, we uses more additive reference signals to increase the GSC beamformer's performance. The auto and cross PSD between $Y_s(k,v)$ and $Y_p(k,v), Y_q(k,v)$ can be determined as:

$$P_{YsYp}(k,v) = (1-\alpha)P_{YsYp}(k,v-1) + \alpha Y_s(k,v)Y_p^*(k,v) \tag{11}$$

$$P_{YpYp}(k,v) = (1-\alpha)P_{YpYp}(k,v-1) + \alpha Y_p(k,v)Y_p^*(k,v) \tag{12}$$

and:

$$P_{YsYq}(k,v) = (1-\alpha)P_{YsYq}(k,v-1) + \alpha Y_s(k,v)Y_q^*(k,v) \tag{13}$$

$$P_{YqYq}(k,v) = (1-\alpha)P_{YqYq}(k,v-1) + \alpha Y_q(k,v)Y_q^*(k,v) \tag{14}$$

The adaptive noise canceller $H_p(k,v), H_q(k,v)$ can be computed as:

$$H_p(k,v) = \frac{P_{YsYp}(k,v)}{P_{YpYp}(k,v)} \tag{15}$$

$$H_q(k,v) = \frac{P_{Y_sY_q}(k,v)}{P_{Y_qY_q}(k,v)} \tag{16}$$

By applying the more additive reference signal, the final output signal can be derived as:

$$\hat{Y}_{GSC}(k,v) = Y_s(k,v) - Y_r(k,v) * H_r(k,v) - Y_p(k,v) * H_p(k,v) - Y_q(k,v) * H_q(k,v) \tag{17}$$

With using the more additive reference signal, the GSC beamformer's performance can be increased by suppressing the remaining noise at the final output signal. We will conduct an experiment to verify the effectiveness of the suggested technique in a realistic environment with more additive reference signals.

4. Experiments

In this section, we will demonstrate an experiment with a dual-microphone system (DMA2) to verify the advantage of the described method in suppressing the noise level. The experiment was conducted in the living room $(4x3.5x5(m))$ with a DMA2, a stand speaker at the distance $L = 5(m)$, the distance between two mounted microphones is $d = 5(cm)$. The speaker stands at $\theta_s = 90(deg)$. The entire conversation includes the background noise, non-directional noise or third-party talkers.

The purpose of this section is comparing the effectiveness of the proposed technique (MoRe) with traditional GSC beamformer (traGSCbe) in suppressing noise level and increasing the speech quality in the term of signal-to-noise ratio. An objective measurement [10] was used for computing promising results.

For recording the conversation, these parameters were used: $nFFT = 512$, overlap 50%, frequency sampling $Fs = 16kHz$. The waveform and spectrum of observed MA signals can be shown in figure 6.

With optimum smoothing parameter $\alpha = 0.1$,the processed signal by implementing GSC beamformer is illustrated in figure 7.

From figure 7, we can see the musical noise and remaining noise decreased the speech quality of GSC beamformer's output signal. The existence of a certain noise level and musical noise does not satisfy the perceptual listener. The promising result by using MoRe is presented in figure 8.

Figure 9 compared the energy between microphone array signals, the processed signals by traGSCbe and MoRe. From Fig , we can see that the desired target speech component was preserved while background noise of beamformer's output signal was reduced. As a result, the measured SNR was increased from 7.4 to 8.8 (dB), and the noise level was decreased to 9.8 (dB).

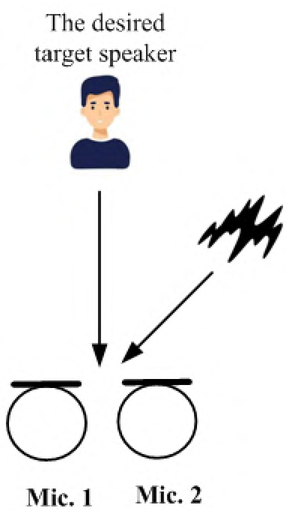


Figure 5: The demonstrated experiment with dual-microphone system.

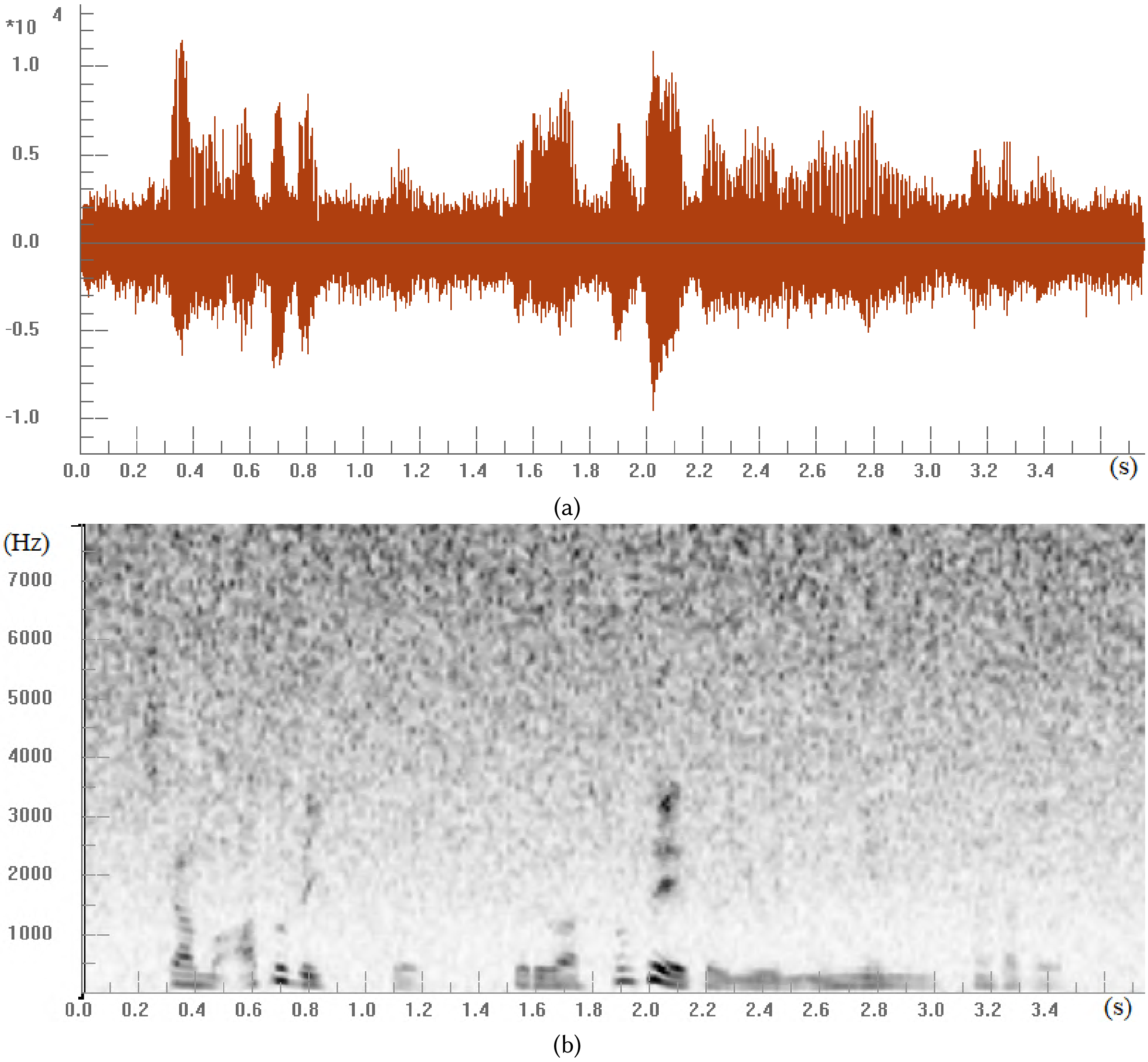


Figure 6: The waveform (a) and spectrogram (b) of microphone array signals.

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

Method estimation	Microphone array signal	traGSCbe	MoRe
NIST STNR	11.2	20.1	27.5
WADA SNR	7.0	18.0	26.8

The appealing properties of the our direction research is exploiting the priori spatial information about the impinging angle of target speaker to the axis of DMA2 to obtain more additive reference signal $Y_p(k, v), Y_q(k, v)$, which are not correlated with $Y_r(k, v)$. The use of three reference signals allows Wiener filter owning the high capability of eliminating musical noise, surrounding noise and remaining noise. The numerical result has confirmed the ability of the proposed method of low computation of calculating additive reference signal, which based on the assumption of preferred steering vector, to improve the GSC beamformer's evaluation in an annoying environment. The described technique can be integrated into various types of acoustic equipment to precess MA signals at real-time mode.

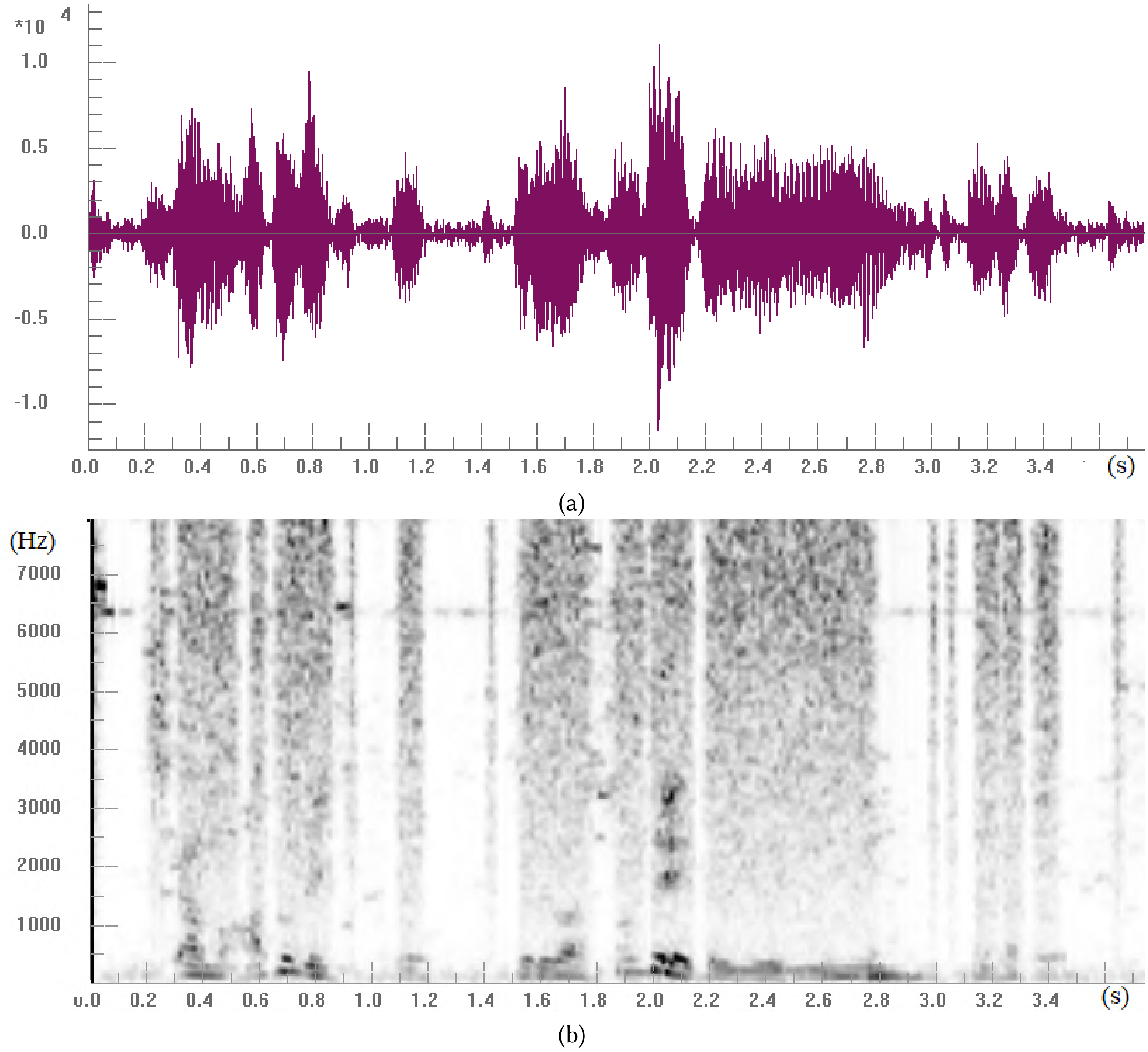


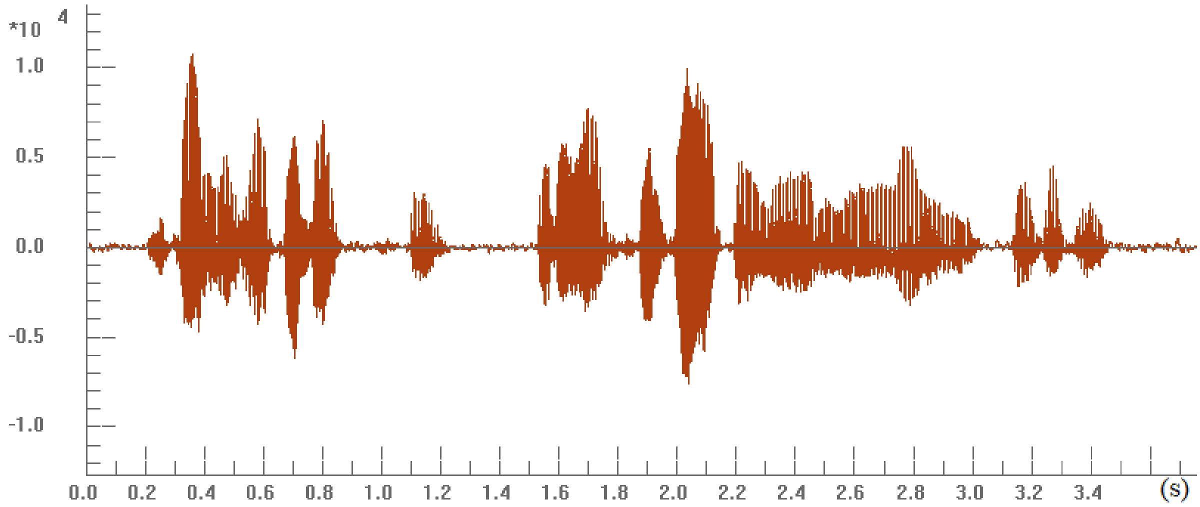
Figure 7: The waveform (a) and spectrogram (b) of processed signal by applying traGSCbe.

5. Conclusion

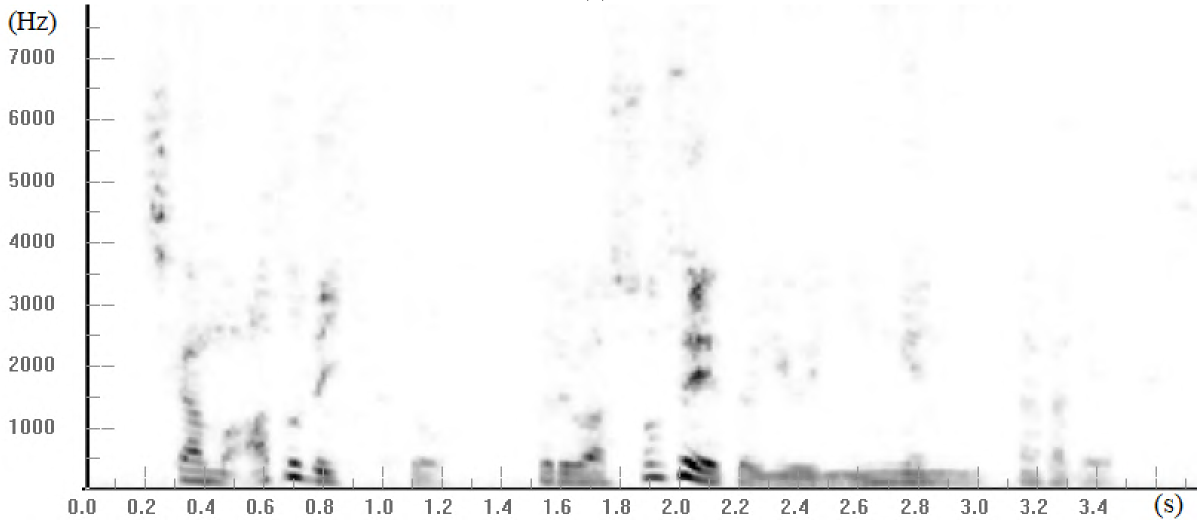
MA beamforming with the advantage of applying the spatial diversity to obtain the original speech data while reducing the background noise, surrounding noise. GSC beamformer is one of the most efficient techniques, which are commonly installed into numerous speech applications. Because of the error of MA geometry configuration, the error of sampling frequency, the different microphone sensitivities, the moving head of talker, interference or third-party talker, the GSC beamformer's performance often corrupted. In this article, we presented an efficient method to increase the number of reference signals to reduce musical noise, background noise to 9.8 dB and improve the speech quality in the term of SNR from 7.4 to 8.8 dB. The described approach can be applied in multi-channel to address more complicated problems.

Declaration on Generative AI

The authors have not employed any Generative AI tools.



(a)



(b)

Figure 8: The waveform (a) and spectrogram (b) of processed signal by applying MoRe.

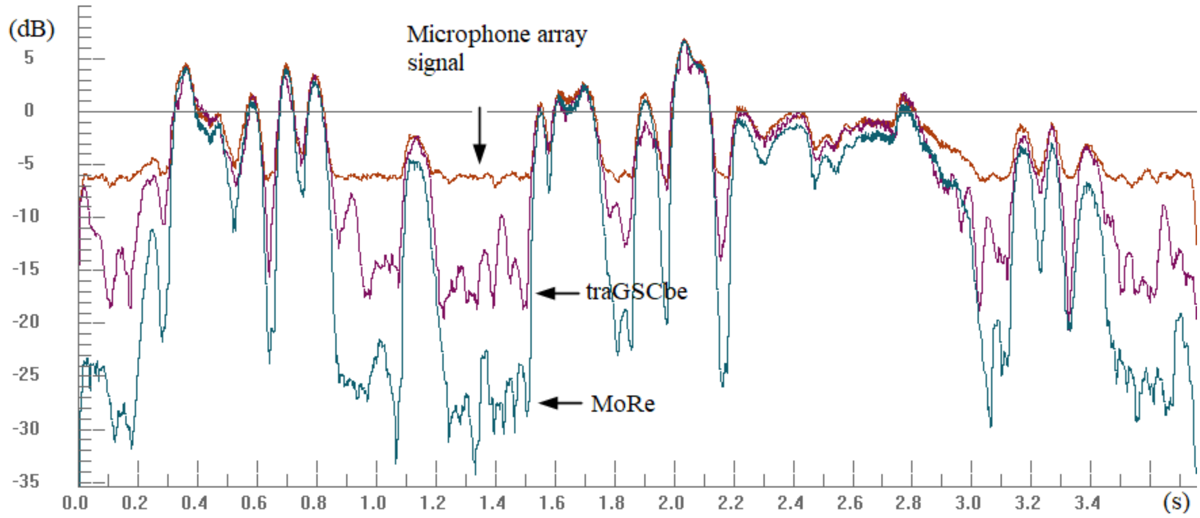


Figure 9: The comparison energy between the original MA signal and processed signals by traGSCbe and MoRe.

References

[1] G. Huang, I. Cohen, J. Benesty, J. Chen, Kronecker Product Beamforming with Multiple Differential Microphone Arrays, in: 2020 IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM), 2020, pp. 1–5. doi:10.1109/SAM48682.2020.9104333.

[2] G. Huang, J. Benesty, J. Chen, I. Cohen, Robust and steerable kronecker product differential beamforming With rectangular microphone arrays, in: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 211–215. doi:10.1109/ICASSP40776.2020.9052988.

[3] S. S. Priyanka, T. K. Kumar, GSC Beamforming using Different Adaptive Algorithms for Speech Enhancement, in: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1–6. doi:10.1109/ICCCNT45670.2019.8944415.

[4] A. Barnov, V. B. Bracha, S. Markovich-Golan, S. Gannot, Spatially Robust GSC Beamforming with Controlled White Noise Gain, in: 2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC), 2018, pp. 231–235. doi:10.1109/IWAENC.2018.8521405.

[5] Q. Jiang, Y. Zhou, H. Liu, A Robust GSC for Microphone Array Using Coherence and Signal-to-Interference Ratio, in: 2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2019, pp. 1–5. doi:10.1109/ISSPIT47144.2019.9001775.

[6] M. Zohourian, R. Martin, GSC-Based Binaural Speaker Separation Preserving Spatial Cues, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 516–520. doi:10.1109/ICASSP.2018.8462658.

[7] R. Ali, T. van Waterschoot, M. Moonen, Generalised Sidelobe Canceller for Noise Reduction in Hearing Devices Using an External Microphone, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 521–525. doi:10.1109/ICASSP.2018.8461366.

[8] S. Li, Q. Liu, W. Wang, Generalized Sidelobe Canceller with Variable Step-Size Least Mean Square Algorithm Controlled by Signal-to-Noise Ratio, in: 2022 5th International Conference on Data Science and Information Technology (DSIT), 2022, pp. 1–6. doi:10.1109/DSIT55514.2022.9943855.

[9] J. Wang, F. Yang, J. Guo, J. Yang, Robust Adaptation Control for Generalized Sidelobe Canceller with Time-Varying Gaussian Source Model, in: 2023 31st European Signal Processing Conference (EUSIPCO), 2023, pp. 16–20. doi:10.23919/EUSIPCO58844.2023.10289801.

[10] D. Ellis, Objective measures of speech quality/SNR, 2011. URL: https://labrosa.ee.columbia.edu/projects/snreval/.