An Adaptive Post - Filtering - Based Generalized Sidelobe Canceller Beamformer's Speech Enhancement

QuanTrong The

Post and Telecommunication Institute of Technology, Hanoi, Vietnam

Abstract

Speech signal acquisition from distant and separation sound source is a challenging problem in almost speech applications, especially in complex and annoying environments in presence of competing talker, transport vehicle, interference, non - directional noise. For several applications like speech recognition, teleconferencing system, smart phone, voice - controlled equipment, hands - free human - machine interface, the high signal - to - noise ratio (SNR), the high speech quality - intelligibility and perceptual listener is prerequisite to achieve an acceptable result from any algorithm trying to recover the clean speech component from the mixture of speech - noise. Due to the rapidly changing acoustic characteristics and varying the location of talker respect to the captured microphone, the single - channel approach or fixed beamformer do not deliver sufficient performance. Adaptive beamformer, which the filter's coefficients are tracked, updated according to the changed recording situations for preserving the clean speech component while suppressing the total background noise. Generalized Sidelobe Canceller (GSC) beamformer is one of the most useful microphone arrays (MA) beamforming for both eliminating surrounding noise and interference while saving the desired target speaker at certain direction. However, in realistic recording scenario, the GSC beamformer's performance often decreased due to the microphone mismatches, the error of sampling rate, the inaccurate estimation of preferred steering vector, the displacement of MA, the different MA sensitivities. In this article, the author suggested using an additive post - Filtering, which removes noise level at the output of GSC beamformer and increases the speech quality. The obtained simulation has confirmed the effectiveness of the proposed post - Filtering can be integrated into a multi-channel system.

Keywords

Generalized sidelobe canceller beamformer, post - Filtering, signal-to-noise ratio, speech enhancement, speech quality.

1. Introduction

Speech is the most prominent and primary part of interaction between human - to - human and human – to – machine communication in several fields such as speech acquisition, speech recognition, surveillance devices, smart-home, voice - controlled equipment, hearing aids, speaker identification, teleconference system, mobile phone.

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

* Corresponding author:
theqt@ptit.edu.vn
0000-0002-2456-9558

Nowadays, numerous types of interfering signals, non - directional noise, incoherent noise, diffuse noise, complicated environment degrade the speech quality, speech intelligibility and make the listening task difficult for listener and decrease the reliability of present-day speech communication, as in Fig. 1. Therefore, to achieve near – transparent the original clean speech data, noise suppression, improve speech quality, enhance the signal-to-noise ratio or increase the corrupted speech is one of the main types of research over last few decades.

Figure 1: The complex and adverse environment

Minimizing the total output noise power, decreasing the degree of speech distortion of observed signal and improving one or more perceptual aspects of speech is main research [1-2, 30]. The classification of speech enhancement techniques depends on the number of microphones, which are used for capturing, receiving and collecting speech, such as single, dual or multi-channel. The single - channel is often based on the spectral subtraction method, which leads to speech distortion in rapidly changing environments. Therefore, microphone array technology has been commonly used for solving complex problems of speech enhancement in annoying and adverse recording scenarios. The scheme of MA beamformer's work is given on Fig. 2. The MA beamforming can be categorized into two groups: fixed beamformer with delay and sum DAS [3-6], adaptive beamformer with differential microphone array DIF [7-10], minimum variance distortionless response MVDR [11-14], linearly constraint minimum variance LCMV [15-18], generalized sidelobe canceller GSC [19-22]. The adaptive beamformer changes the optimum filter's coefficients according to the changing surrounding environment to obtain better noise reduction and improve the speech quality.

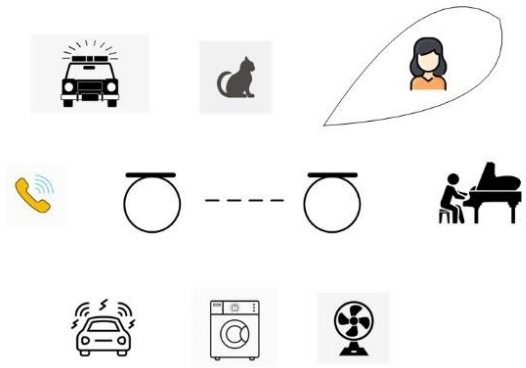

Figure 2: The steered beampattern toward the desired target talker

MA has been widely installed in various types of speech applications because of their great performance with capabilities of sound source localization, steered beampattern at certain direction in Fig. 3. GSC beamformer contains the parts: the fixed beamformer (FBF) for concerning the sound source, the blocking matrix (BM) for removing the speech component and adaptive noise canceller (ANC), which plays important role to extract desired target speaker and remove noise from the output of FBF. Usually, the adaptive blocking matrix (ABM) is applied to reject speech components while passing through noise. In [24], an estimation of signal-to-interference (SIR) ratio for controlling ABM, ANC coefficients. In [25], Hoshuyama a new ABM by using constrained coefficients and norm – constrained adaptive filter was used for ANC filter. Yoon [26] exploited the sound – source presence probability estimated the captured MA signals and voice activity detection into ABM. In [27], Herbordt demonstrated a similar GSC structure in frequency domain. Despite the advantages of these above methods, the accuracy of ABM, ANC filter's coefficients is still complicated challenging, especially under low signal-to-noise situations, non - directional noise or incoherent/diffuse noise field. To overcome this drawback of speech enhancement, in this contribution, the author proposed using an additive post - filtering for saving the speech component while suppressing noise level to achieve better speech quality.

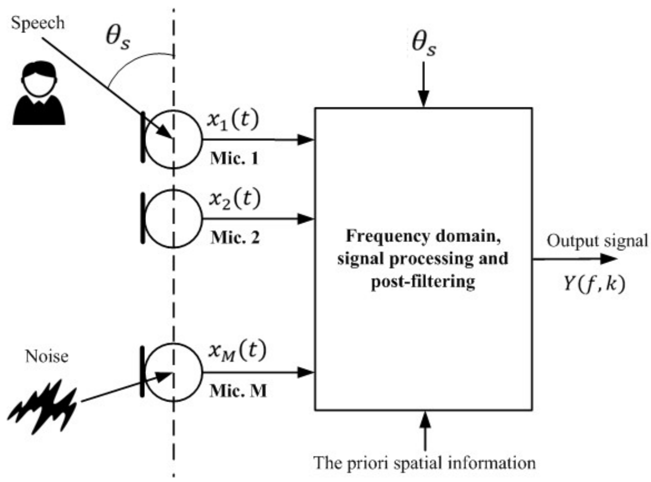

Figure 3: The implementation of microphone array beamforming in the frequency – domain

The rest of this paper is organized as follows: The first section introduces speech enhancement and MA technology. The second section describes the principal working of GSC beamformer and the reasons for degradation performance. The author proposed post – Filtering in the next section. Section IV demonstrates a perspective experiment to illustrate the advantages of suggested techniques. Section V concludes and the author's future work.

2. Generalized Sidelobe Canceller Beamformer

In this section, the author presents the scheme working of GSC beamformer in the frequency – domain. As in Fig. 4, in general case, we will investigate with the dual – microphone array. At the considered frequency f, frame k, the observed microphone array signals $X_1(f, k)$,$X_2(f, k)$ can be formulated as:

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

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

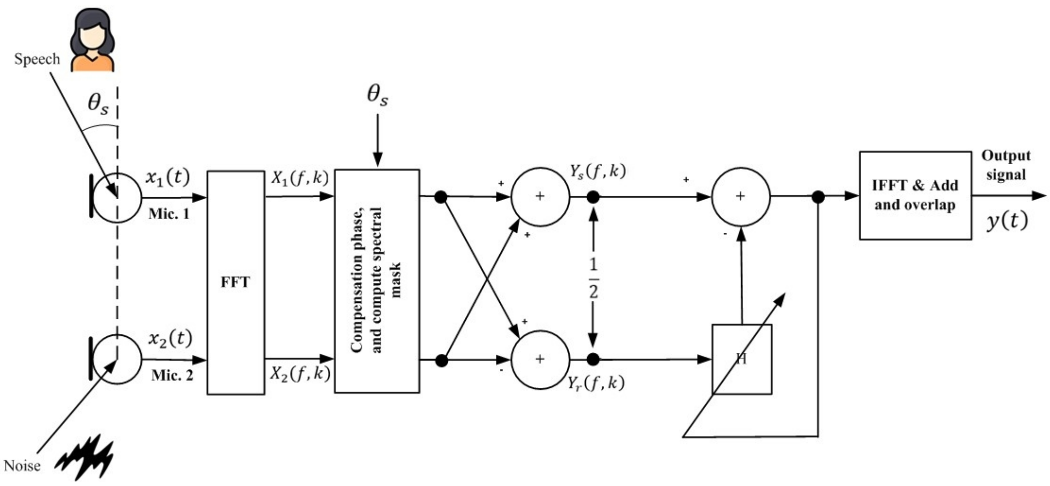

Figure 4: The scheme working of GSC beamformer

Where $S(f, k)$ is the original clean speech, $N_1(f, k)$, $N_2(f, k)$ is the additive noise, interference or surrounding noise in two microphones, $\Phi_s = \pi f \tau_0 cos(\theta_s)$, θ_s is the direction of arrival of target speaker to the axis of DMA2, $\tau_0 = d/c$, d is the range between two microphone, $c = 343$ (m/s) is the sound speed propagation in the air.

The main and reference signal $Y_s(f, k)$,$Y_r(f, k)$ are calculated as the following equations:

$$Y_s(f,k)=\frac{1}{2}(X1(f,k)e^{-j\Phi s}+X2(f,k)e^{j\Phi s}) \quad (3)$$

$$Y_r(f,k)=\frac{1}{2}(X1(f,k)e^{-j\Phi s}+X2(f,k)e^{j\Phi s}) \quad (4)$$

The ANC filter often uses adaptive Wiener filter for alleviating the noise component, which is contained in the main signal $Y_s(f, k)$. The optimum Wiener filter's coefficients can be derived from the below formulations:

$$H_w(f,k)=\frac{E\{Y_s(f,k)Y_r^{*}(f,k)\}}{E\{|Y_r(f,k)|^2\}} \tag{5}$$

The auto – cross power spectral densities of the main and reference signal are determined as the recursive equations:

$$P_{Y_sY_r}(f,k)=\alpha P_{Y_sY_r}(f,k-1)+(1-\alpha)\,Y_s(f,k)Y_r^{*}(f,k) \tag{6}$$

$$P_{Y_rY_r}(f,k)=\alpha P_{Y_rY_r}(f,k-1)+(1-\alpha)\,Y_r(f,k)Y_r^{*}(f,k) \tag{7}$$

The GSC beamformer's output signal yields as:

$$Y_{GSC}(f,k)=Y_s(f,k)-H_w(f,k)\times Y_r(f,k) \tag{8}$$

Unfortunately, in practice, because of the complex and annoying surrounding environment, the imprecise estimation of impinging signal relative to the MA geometry, the displacement of architecture of MA, the different MA sensitivities, undetermined background noise negatively effect on GSC beamformer's evaluation. In several recording situations, the remaining noise component decreases the speech quality and speech intelligibility. In the next section, the author proposed using a new post – Filtering for suppressing noise level, increasing the signal-to-noise ratio.

3. The proposed post – Filtering

In [28], the coherence between two microphone signals $\Gamma_{X_1X_2}(f, k)$ is an acoustic parameter, which presents the probability of existence of speech component by an approximate formulation:

$$\Gamma_{X_1X_2}(f,k)\approx\frac{SNR(f,k)}{1+SNR(f,k)}\Gamma_S+\frac{1}{1+SNR(f,k)}\Gamma_n$$

Where $\Gamma_{X_1X_2}(f,k)=\dfrac{P_{X_1X_2}(f,k)}{\sqrt{P_{X_1X_1}(f,k)P_{X_2X_2}(f,k)}}\Gamma_S=e^{j2\Phi_S}$

$$P_{X_iX_i}(f,k)=\beta P_{X_iX_i}(f,k)+(1-\beta)X_i(f,k)X_i^{*}(f,k),i=1,2$$

$$P_{X_iX_j}(f,k)=\beta P_{X_iX_j}(f,k)+(1-\beta)X_i(f,k)X_j^{*}(f,k),i=1,j=2$$

With the smoothing parameter β in the range {0 … 1}.

The author's ideal is using parameter $\frac{SNR(f,k)}{1+SNR(f,k)}$ as an additive post − Filtering for enhancing GSC beamformer's performance.

If we denote $\gamma(f,k)=\frac{SNR(f,k)}{1+SNR(f,k)}$, the equation (9) can be rewritten as:

$$\Gamma_{X_1X_2}(f,k)\approx\gamma(f,k)\Gamma_S+(1-\gamma(f,k))\Gamma_n \qquad (12)$$

With the probability of presence of – speech component, (12) describes an approximate relation between speech and noise. For obtaining the accurate computation, the author proposed using the spectral mask gain for delay and sum $G^{DSB}(f, k)$ and spectral mask for blocking matrix $G^{BM}(f, k)$ [29]. These gain functions relate to the MA beamforming technique for gaining speech components and reject this component to achieve only noise.

$$G^{DSB}(f,k)=\frac{1}{2}\left(1+e^{j\Phi_{12}^{norm}(f,k)}\right) \qquad (13)$$

$$G^{BM}(f,k)=\frac{1}{2}\left(1-e^{j\Phi_{12}^{norm}(f,k)}\right) \qquad (14)$$

Where $\phi_{12}^{norm}(f,k)=\frac{\phi 12(f,k)}{\omega d}$, with $\phi_{12}(f, k)$ is phase difference between two captured microphone array signals.

Therefore, the author suggests a modified formulation (12) to derive an accurate $\gamma(f, k)$ as:

$$\Gamma_{X_1X_2}(f,k)=\gamma(f,k)G^{DSB}(f,k)\Gamma_S+(1-\gamma(f,k))G^{BM}(f,k)\Gamma_n \qquad (15)$$

And:

$$\gamma(f,k)=\frac{\Gamma_{X_1X_2}(f,k)-G^{BM}(f,k)\Gamma_n}{G^{DSB}(f,k)\Gamma_S-G^{BM}(f,k)\Gamma_n} \qquad (16)\ \Gamma$$

With Γ_n = 1 in the coherent noise field, and $\Gamma_n = sin(\omega\tau_{o)/}(\omega\tau_o)$ in diffuse noise field. Finally, the GSC beamformer's output signal is filtered out by applying post − Filtering as:

$$\hat{Y}_{GSM}(f,k)=Y_{GSC}(f,k)\times\gamma(f,k) \qquad (17)$$

In the next section, the author demonstrates a perspective experiment to illustrate the effectiveness of post − Filtering to suppress noise level, which still exist in GSC beamformer's output signal.

4. Experiment results

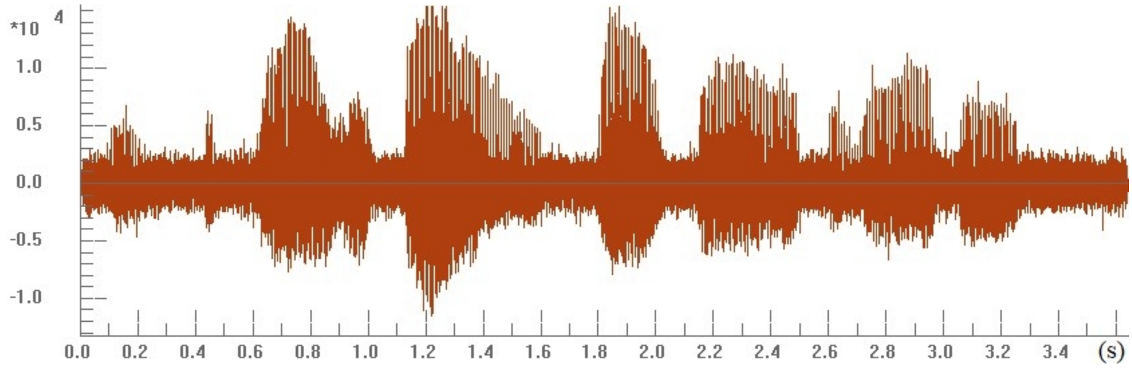

Figure 5: The waveform of captured microphone array signals

The purpose of this experiment is to verify the effectiveness of the author's proposed method in speech enhancement to extract the desired target speaker while suppressing unwanted interference in diffuse noise fields. An objective measurement [23] was used for calculating the signal-to-noise ratio. In this section, a dual – microphone system was used for demonstrating the ability of enhancing speech enhancement after using post – Filtering.
The distance between stand speaker to DMA2 is $\boldsymbol{L} = \mathbf{3.3}\ (\boldsymbol{m})$, the direction of arrival of useful signal is $\boldsymbol{\theta_s} = \mathbf{90}(\boldsymbol{deg})$. For capturing the clean speech and background noise, these parameters were set: the sampling rate is $\boldsymbol{Fs} = \mathbf{16}\ \boldsymbol{kHz}$, the overlap $\mathbf{50}$%. The original microphone array signal was depicted in Figure 5.

For further signal processing, these necessary parameters were set: $nFFT = 512$, smoothing parameter $\alpha = 0.1$, $\beta = 0.1$. With the traditional GSC beamformer, the waveform of the output signal was shown in Figure 6.

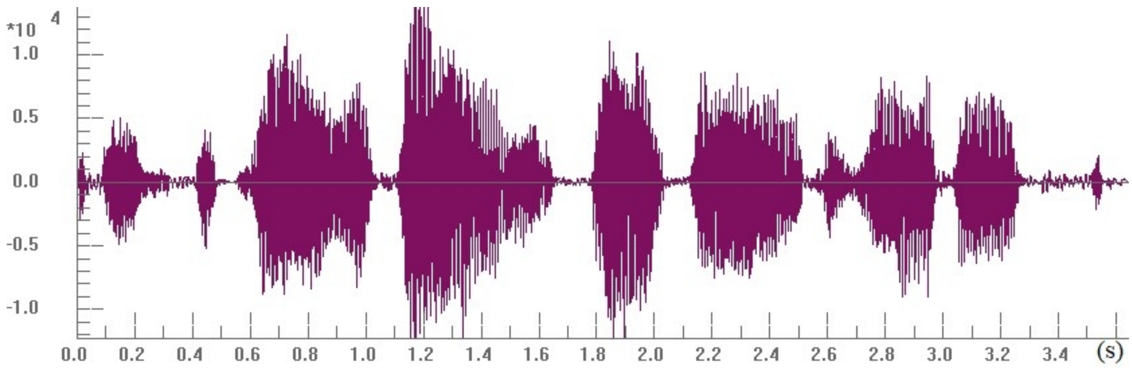

Figure 6: The waveform of the processed by GSC beamformer

In the complex and annoying environment, the GSC beamformer's evaluation often corrupted due to the microphone mismatches, the difference of sensitivities between two microphones, the error of estimation of DoA, the displacement of MA and the rapidly changing of acoustic environment also degrades the speech quality. At the output of GSC beamformer, there are still exits of amount of noise level. For overcoming this drawback, an appropriate post – Filtering was applied for mitigating the remained noise while saving the original speech component. After using the author's suggested post – Filtering, the processed signal can be derived as:

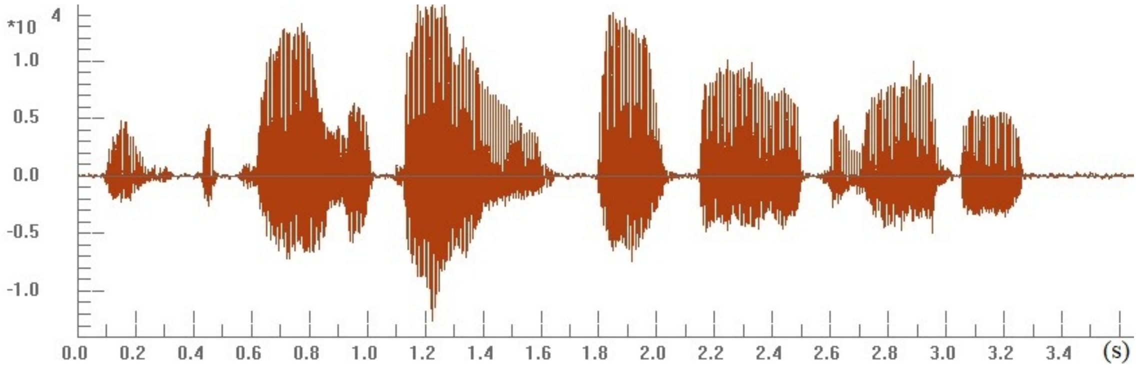

Figure 7: The waveform of the output signal after applying the author's post – Filtering

Table 1 shows the measured SNR between the received MA signals, the processed signals by GSC beamformer and the author's post - Filtering. Figure 8 describes the energy between these signals. As a result, the proposed post - Filtering allows suppressing the noise level to 11 (dB) and increasing the SNR from 11.5 (dB) to 14.3 (dB). From numerical simulations, the advantage of suggested post - Filtering is the capability of improving the obtained speech quality, speech intelligibility, perceptual listener in adverse environment. This method can be integrated into a multi-channel system for solving other complicated problems.

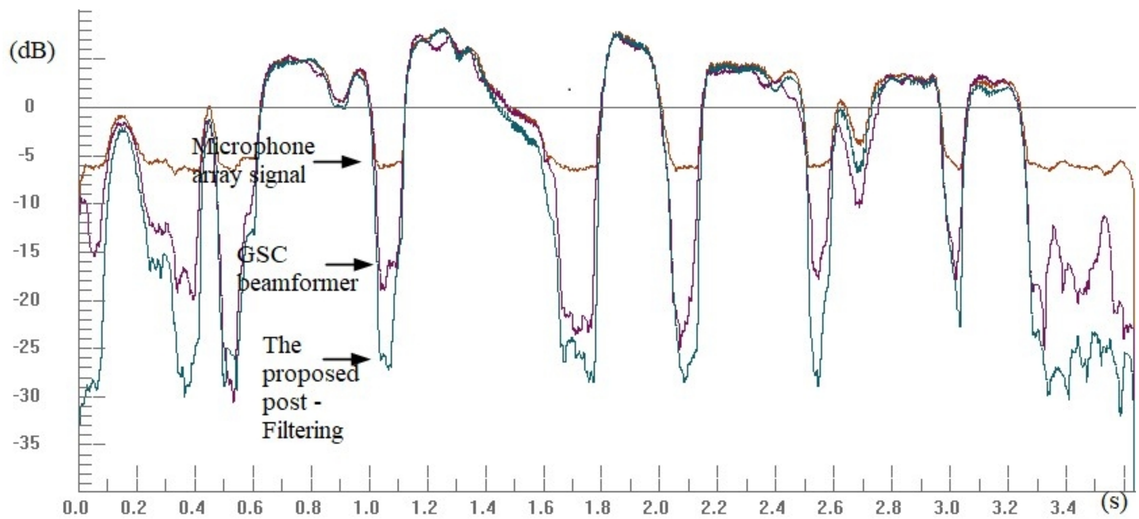

Figure 8: The energy of the microphone array signal, the processes signals by GSC beamformer and the proposed post - Filtering

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

Methos Estimation	Microphone array signals	GSC beamformer	The proposed post - Filtering
NIST SNR	13.2	19.5	33.8
WADA SNR	6.5	18.1	29.6

5. Conclusion

Nowadays, information and communication technology are developing rapidly with an increasing number of speech applications in human life. Microphone array beamforming is an optimum solution to extract the desired target speech component while suppressing background noise by forming a high directional beampattern toward the sound source. Unfortunately, the underdetermined conditions seriously affect the GSC beamformer's evaluation, that leads to speech distortion in complicated and rapidly changing acoustic environments. In this contribution, the author proposed using an effective post - Filtering for suppressing noise level. The obtained results have confirmed the advantage of the proposed method in removing noise level to 11 dB and enhancing the SNR from 11.5 to 14.3 dB. In the future, the author will investigate the characteristics of coherent noise field, the room reverberation to further enhance the post - Filtering to handle more complicated problems.

References

[1] Loizou P. C. Speech Enhancement: Theory and Practice. Taylor and Francis, London, 2007.

[2] Ephraim Y., Ari H. L., Roberts W. A Brief Survey of Speech Enhancement. 3rd Edition, Electrical Engineering Handbook, CRC, Boca Raton, 2006.

[3] Zeng Y., Hendriks R. C. Distributed delay and sum beamformer for speech enhancement in wireless sensor networks via randomized gossip // Proc 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 2012, pp. 4037-4040, doi: 10.1109/ICASSP.2012.6288804.

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

[5] Zeng Y., Hendriks R. C. Distributed Delay and Sum Beamformer in Regular Networks Based on Synchronous Randomized Gossip // Proc IWAENC 2012; International Workshop on Acoustic Signal Enhancement, Aachen, Germany, 2012, pp. 1-4.

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

[7] Huang G., Benesty J., Chen J., Cohen I. Robust and steerable kronecker product differential beamforming With rectangular microphone arrays // Proc ICASSP 2020 - 2020

IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 211-215, doi: 10.1109/ICASSP40776.2020.9052988.
[8] Wang X., Huang G., Cohen I., Benesty J., Chen J. Robust Steerable Differential Beamformers with Null Constraints for Concentric Circular Microphone Arrays // Proc ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 4465-4469, doi: 10.1109/ICASSP39728.2021.9414119.
[9] Leng X., Chen J., Benesty J. A New Method to Design Steerable First-Order Differential Beamformers. IEEE Signal Processing Letters, vol. 28, pp. 563-567, 2021, doi: 10.1109/LSP.2021.3059533.
[10] Huang G., Wang Y., Benesty J., Cohen I., Chen J. Combined Differential Beamforming With Uniform Linear Microphone Arrays //Proc ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 781-785, doi: 10.1109/ICASSP39728.2021.9414189.
[11] Yamaoka K., Ono N., Makino S., Yamada T. Time-frequency-bin-wise Switching of Minimum Variance Distortionless Response Beamformer for Underdetermined Situations // Proc ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 7908-7912, doi: 10.1109/ICASSP.2019.8683528.
[12] Ali R., Van W. T., Moonen M. Integration of a Priori and Estimated Constraints Into an MVDR Beamformer for Speech Enhancement. IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 27, no. 12, pp. 2288-2300, Dec. 2019, doi: 10.1109/TASLP.2019.2946086.
[13] Q. Trong. A Coherence - Based Post-Filtering for Minimum Variance Distortionless Response Beamformer// Proc 2023 IEEE International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo), Kyiv, Ukraine, 2023, pp. 182-186, doi: 10.1109/UkrMiCo61577.2023.10380331.
[14] Bouchard M., Kamkar P. H. Robust Minimum Variance Distortionless Response Beamformer based on Target Activity Detection in Binaural Hearing Aid Applications // Proc 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Ottawa, ON, Canada, 2019, pp. 1 - 5, doi: 10.1109/GlobalSIP45357.2019.8969387.
[15] Aroudi A., Doclo S. Cognitive-driven Binaural LCMV Beamformer Using EEG - based Auditory Attention Decoding // Proc 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.
[16] Schreibman A., Barnov A., Gendelman A., Tzirkel E. RTF Based LCMV Beamformer with Multiple Reference Microphones // Proc 2020 28th European Signal Processing Conference (EUSIPCO), Amsterdam, Netherlands, 2021, pp. 181-185, doi: 10.23919/Eusipco47968.2020.9287468.
[17] Zhang J., Du J., Dai L. R. Sensor Selection for Relative Acoustic Transfer Function Steered Linearly-Constrained Beamformers. IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 1220-1232, 2021, doi: 10.1109/TASLP.2021.3064399.
[18] Gößling N., Hadad E., Gannot S., Doclo S. Binaural LCMV Beamforming With Partial Noise Estimation. IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2942-2955, 2020, doi: 10.1109/TASLP.2020.3034526.

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

[20] The Q. T., Dung N. T. T. An Improved Implementation of Generalized Sidelobe Canceller Filter in Diffuse Noise Field // Proc 2022 Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), Saint Petersburg, Russian Federation, 2022, pp. 1403-1407, doi: 10.1109/ElConRus54750.2022.9755843.

[21] Dai S., Li M., Abbasi Q. H., Imran M. A. A Fast Blocking Matrix Generating Algorithm for Generalized Sidelobe Canceller Beamformer in High Speed Rail Like Scenario. IEEE Sensors Journal, vol. 21, no. 14, pp. 15775 - 15783, 15 July15, 2021, doi: 10.1109/JSEN.2020.3002699

[22] Middelberg W., Doclo S. Comparison of Generalized Sidelobe Canceller Structures Incorporating External Microphones for Joint Noise and Interferer Reduction. Speech Communication; 14th ITG Conference, online, 2021, pp. 1-5.

[23] https://labrosa.ee.columbia.edu/projects/snreval/.

[24] Hoshuyama. O., Begasse. B., Sugiyama. A., Hirano. A. A Realtime Robust Adaptive Microphone Array Controlled by an SNR Estimation // Proc 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98, Seattle, WA, USA, 15 May 1998; pp. 3605 - 3608.

[25] Hoshuyama. O., Sugiyama. A., Hirano. A. A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters. IEEE Trans. Signal Process. 1999, 47, 2677 - 2684.

[26] Yoon. B., Tashev. I., Malvar. H. Robust adaptive beamforming algorithm using instantaneous direction of arrival with enhanced noise suppression capability // Proc 2007 IEEE International Conference on Acoustics, Speech and Signal Processing, Honolulu, HI, USA, 15–20 April 2007; pp. 133 - 136.

[27] Herbordt. W., Kellermann. W. Computationally efficient frequency-domain robust generalized sidelobe canceller. In Proceedings of the 7th International Workshop on Acoustic Echo and Noise Control (IWAENC), Darmstadt, Germany, 10 -13 September 2001.

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

[29] Kim S. M. Hearing Aid Speech Enhancement Using Phase Difference-Controlled DualMicrophone Generalized Sidelobe Canceller. *IEEE Access*, vol. 7, pp. 130663-130671, 2019, doi: 10.1109/ACCESS.2019.2940047.

[30] https://ceur-ws.org/Vol-3396/paper15.pdf