Reducing the Speech Distortion at GSC Beamformer's Output Signal

Nguyen Duy Phuong
Faculty of Information Technology
Posts and Telecommunications Institute of Technologies
Hanoi, Vietnam
phuongnd@ptit.edu.vn

Quan Trong The
Lab Blockchain, Faculty of Information Security
Posts and Telecommunications Institute of Technologies
Hanoi, Vietnam
theqt@ptit.edu.vn

***Abstract*—Due to the high directional beampattern, microphone array (MA) technology has been implemented in almost all acoustic equipment to extract the target speaker and eliminate background noise, interference or third-party talkers at the same time. MA beamforming is an efficient solution to address the coherent problem of speech enhancement's spectral subtraction in a complex and adverse environment. Generalized sidelobe canceller (GSC) beamformer own the high directivity factor and noise reduction in preserving the speech component and suppressing background noise. Because of many reasons, the GSC beamformer's output signal is often corrupted and speech quality is not satisfied by the listener. In this article, the author suggests using an effective gain function to reduce speech distortion and enhance the signal-to-noise (SNR) ratio. The numerical simulation has shown the advantage of the author's direction approach in increasing the SNR from 7.3 to 10.3 dB and decreasing the speech distortion to 4.8 dB. The described technique can be integrated with the environment's characteristic and more spatial information to further improve the GSC beamformer's performance in other speech applications.**

***Index Terms*—microphone array, beamforming, speech enhancement, noise reduction, generalized sidelobe canceler.**

I. Introduction

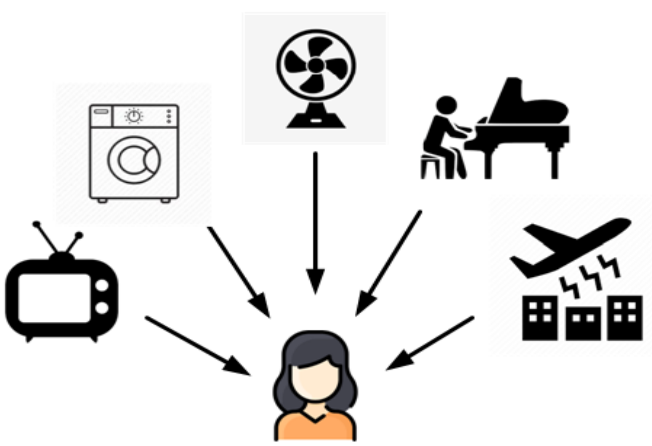


Fig. 1. There are many noise sources significantly on human satisfactory.

Nowadays, the necessity of communication between humans leads to the use of various types of speech applications, which record the clean speech data and transfer it through the network. However, in a realistic recording scenario, the existence of third-party talkers, interference, non-directional noise and other noise sources, as in Fig. 1. To ensure the speech quality, speech enhancement is an essential problem in almost all acoustic equipment for recovering the clean speech data while suppressing all background noise. Single-channel approach often based on spectral subtraction, which usually works well in condition of stationary noise. However, in complex and adverse recording situations, the rapidly changed noise environment causes the speech distortion and musical noise.

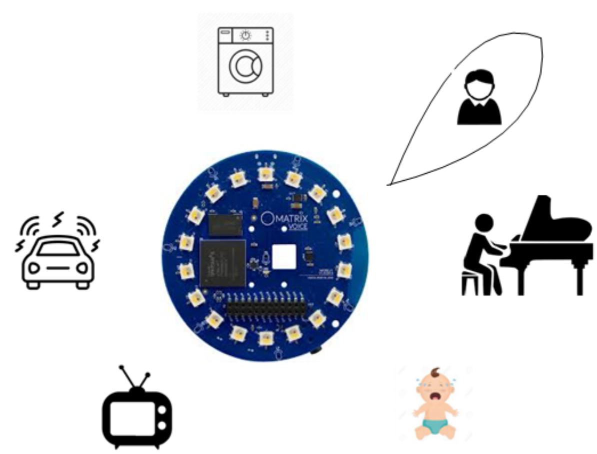


Fig. 2. The using of MA beamforming for recovering the target speaker while removing background noise.

Therefore, exploiting MA beamforming [1-2] has attracted numerous scholars and engineers to explore beampattern for noise reduction and speech enhancement at the same time as in Fig. 2. Fig. 3 shows the principal working of MA beamformer to extract the desired target speaker while removing surrounding noise. GSC beamformer is an efficient solution for steering the designed beampattern at specified direction and applies adaptive noise canceler (ANC) to obtain clean speech data.

However, because of the error of sampling rate, the imprecise MA configuration, the error of determining steering vector, the difference of microphone sensitivities, the complex and complicated environmental factors, the microphone mismatches, the GSC beamformer often degraded. There are many efforts to overcome this above drawback to improve the beamformer's evaluation in various types of situations.

In [3], some adaptive algorithms for speech enhancement were studied for incorporating with GSC beamformers to obtain better results in real-life recording scenarios. These algorithms include Recursive Least Square (RLS), Least Mean Square (LMS), Normalized LMS. The numerical simulation showed the effectiveness of the author's approach.

Jiang Q et al [4] used a superdirective beamformer's output to calculate the coherence and determine the signal-to-noise ratio (SIR) to control the updating parameter of adaptive noise canceler in a complex environment. The numerical simulation has confirmed the advantage of the proposed method in reducing different noise interferences at different locations and increasing the robustness of GSC beamformer in noisy scenarios.

Barnov A [5] presented a modified RLS algorithm to improve the robustness of ANC's performance in a realistic recording environment. The author's technique includes an amount of diagonal loading to enhance adaptive filtering. The experiment operated a recorded signal in a vehicle to demonstrate the effectiveness of this direction research.

In [6], frequency - domain adaptive blocking matrix was described with a prospective voice activity detection, which is based on multiple main lobes to adaptively adjust updating parameters according to the changed environment. The conducted experiment showed better suppression of speech leakage in GSC structure.

Chang D et al [7] proposed a method for computing the time-varying direction of arrival of interesting useful signals. The method used an augmented Kalman filter to determine accurate coefficients and alleviate the unwanted influence of surrounding noise. The simulation evaluation revealed that the author's suggested approach has better speech quality in the terms of SNR and convergence rate in comparison with traditional beamformer.

In [8], a systematic configuration for designing a multiple constrained BM for removing total speech components is demonstrated. The author's idea is based on the analysis of blocking matrix and signal leakage to achieve effective sparse BM and eigen-space BM. The obtained results have demonstrated the advantage of the suggested sparse BM for blocking speech leakage and improving the GSC beamformer's evaluation in various noisy conditions.

The listed research is often illustrated in laboratory conditions without a complex and annoying environment and can not properly address the speech distortion or musical noise, which often occurs in GSC beamformer. Therefore, in this paper, the author introduced an efficient gain function to overcome speech distortion and improve the speech quality.

II. GENERALIZED SIDELOBE CANCELLER BEAMFORMER

In this section, the author introduces the principle working of GSC beamformer's structure with 03 parts. The fixed beamformer (FBF) steers the designed beampattern towards the sound source location. The blocking matrix (BM) blocks the speech component from a noisy mixture of received array signals to achieve the only noise component. The final part of GSC beamformer is an adaptive noise canceler (ANS), which extracts the original clean speech data from FBF's output by using the BM's output as reference signal. GSC beamformer is an efficient solution for preserving the target talker while suppressing surrounding noise, third-party speakers from other directions, interference or non-directional noise. In this article, the author uses the Wiener filter as promising ANC to illustrate beamformer's performance.

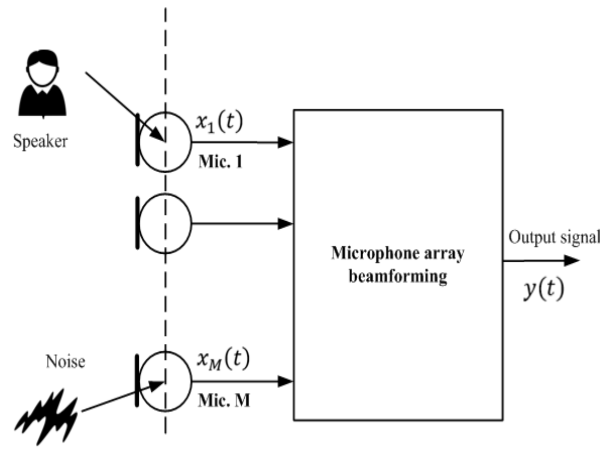


Fig. 3. The scheme of MA beamforming to obtain the desired target talker.

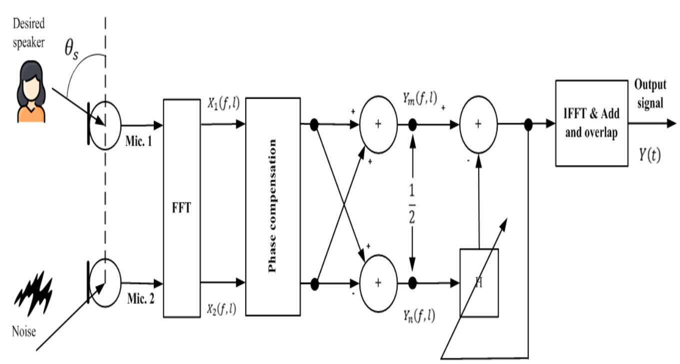


Fig. 4. The principal working of GSC beamformer's in the frequency domain.

In the general case, the author uses a dual-microphone system (DMA2) to demonstrate the scheme of GSC beamformer in the frequency domain. At the considered frame l, frequency f, the observed MA signals, $X_1(f,l), X_2(f,l)$ can be formulated as:

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

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

where $S(f,l)$ denotes the original speech component, $N_1(f,l), N_2(f,l)$ is the additive noise at two microphones, respectively. $\Phi_s = \pi f \tau_0 cos(\theta_s)$, θ_s is preferred direction of arrival of useful signal relative to the axis of DMA2, $\tau_0 = \frac{d}{c}$, d means the range between microphones, c is the sound speed propagation in the fresh air and $c = 343(m/s)$.

The FBF beamformer's output is named as the main signal $Y_m(f,l)$ and the output of the BM block is reference signal $Y_n(f,l)$. These signal can be derived after using phase alignment as:

$$Y_m(f,l) = \frac{X_1(f,l)e^{-j\Phi_s} + X_2(f,l)e^{j\Phi_s}}{2} \tag{3}$$

$$Y_n(f,l) = \frac{X_1(f,l)e^{-j\Phi_s} - X_2(f,l)e^{j\Phi_s}}{2} \tag{4}$$

The Wiener filter plays an important role in recovering the clean speech data from $Y_m(f,l)$ by using $Y_n(f,l)$ as a reference signal. The Wiener filter computes the auto and cross power spectral densities between main signal and reference signal as:

$$P_{YmYn}(f,l) = (1-\alpha)P_{YmYn}(f,l-1) + \alpha Y_m(f,l)Y_n^*(f,l) \tag{5}$$

$$P_{YnYn}(f,l) = (1-\alpha)P_{YnYn}(f,l-1) + \alpha Y_n(f,l)Y_n^*(f,l) \tag{6}$$

with a smoothing parameter α in range of 0 to 1.

Hence, the coefficients of Wiener filter is defined as:

$$H_{GSC}(f,l) = \frac{P_{YmYn}(f,l)}{P_{YnYn}(f,l)} \tag{7}$$

The beamformer output signal is derived by:

$$Y(f,l) = Y_m(f,l) - Y_n(f,l) * H_{GSC}(f,l) \tag{8}$$

In real-life situations, due to numerous reasons, the GSC beamformer's output signal often corrupted. The remaining noise and speech distortion often occur and degrade the speech intelligibility of the output signal. In the next section, the author proposed an appropriate gain function to obtain the desired target speaker.

III. The Author's Proposed Method

The author's idea is exploiting the spatial information to form an efficient gain function to gain the GSC beamformer's output signal as approximately desired target speaker.

In [9], the prior information about the direction of arrival of interest talker allows deriving the spectral masking $G_l^{DSB}(f) = 0.5 + 0.5 * exp(j\phi_{12,l}^{norm}(f))$ and $G_l^{BM}(f) = 1 - exp(j\phi_{12,l}^{norm}(f))$, where $\phi_{12,l}^{norm}(f) = \frac{\phi_{12,l}.c}{f.d}$, $\phi_{12,l}$ is the phase difference between $X_1(f,l), X_2(f,l)$.

Based on the observed phase difference and the $G_l^{DSB}(f), G_l^{BM}(f)$, the author proposed computing *a priori* $SNR_1(f,l)$ as:

$$SNR_1(f,l) = \frac{G_l^{DSB}(f)}{G_l^{BM}(f)} \tag{9}$$

And an efficient gain function $GF_1(f,l)$, which based on *a priori* $SNR_1(f,l)$ as:

$$GF_1(f,l) = \frac{SNR_1(f,l)}{SNR_1(f,l)+1} \tag{10}$$

In addition, the definition of coherence between $X_1(f,l), X_2(f,l)$ can be computed as [10]:

$$\Gamma_{X_1X_2}(f,l) = \frac{SNR_2(f,l)}{1+SNR_2(f,l)}e^{j2\Phi_s} + \frac{1}{1+SNR_2(f,l)}\Gamma_N(f) \tag{11}$$

where $\Gamma_N(f) = 1$ in the coherent noise field, and $\Gamma_N(f) = \frac{sin(\omega\tau_0)}{\omega\tau_0}$ in diffuse noise field, $\omega = 2\pi f$.

$\Gamma_{X_1X_2}(f,l) = \frac{P_{X_1X_2}(f,l)}{\sqrt{P_{X_1X_1}(f,l)*P_{X_2X_2}(f,l)}}$, $P_{X_iX_i}(f,l), P_{X_iX_j}(f,l),\ i,j \in \{1,2\}$ recursively determined according considered frame l:

$$P_{X_iX_j}(f,l) = (1-\beta)P_{X_iX_j}(f,l-1) + \beta X_i(u,v)X_j^*(f,l) \tag{12}$$

where β is the smoothing parameter in range $\{0...1\}$. If we denote $GF_2(f,l) = \frac{SNR_2(f,l)}{SNR_2(f,l)+1}$, the equation (11) can be rewritten as the following way:

$$\Gamma_{X_1X_2}(f,l) = GF_2(f,l)e^{j2\Phi_s} + (1-GF_2(f,l))\Gamma_N(f) \tag{13}$$

And:

$$GF_2(f,l) = \frac{\Gamma_{X_1X_2}(f,l) - \Gamma_N(f)}{e^{j2\Phi_s} - \Gamma_N(f)} \tag{14}$$

In complex and adverse environment, due to the presence of coherent, incoherent noise, diffuse noise and non-directional noise, the author proposed using the formulation of $\Gamma_N(f) = \frac{sin(\omega\tau_0)}{\omega\tau_0(1+\frac{\sigma_n^2}{P_{nn}(f)})}$ [11] with σ_n^2 is uncorrelated noise variance and $P_{nn}(f)$ means noise power spectral of considered recording scenario.

Based on the priori spatial information, the author suggested a hybrid combination between $GF_1(f,l)$ and $GF_2(f,l)$ with an speech presence probability $spp(f,l)$ [12] as the following equation:

$$GF(f,l) = spp(f,l)GF_1(f,l) + (1-spp(f,l))GF_2(f,l) \tag{15}$$

And the final GSC beamformer's output signal can be obtained as:

$$Y_{GF}(f,l) = GF(f,l) \times Y(f,l) \tag{16}$$

In the next section, the author will demonstrate the improvement of the author's proposed approach in reducing the speech distortion and increasing the speech quality, speech intelligibility.

IV. EXPERIMENTS

The purpose of this section is to demonstrate the effectiveness of the author's method (hybridGF) in comparison the speech quality of observed microphone array signals, the processed signals by traditional GSC beamformer (tdtGSC) and hybridGF in realistic recording scenarios. The simulated experiment was conducted in the living room (5x4x3.5 m) with target stand speakers at distance $L = 2(m)$ and non-directional noise, third-part talker, interference.

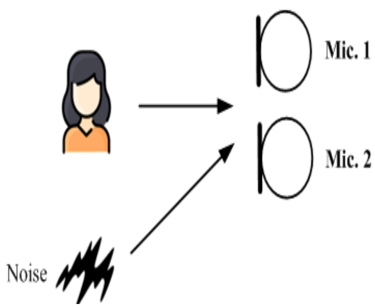
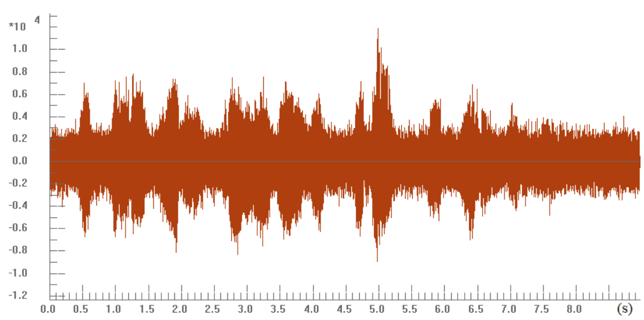
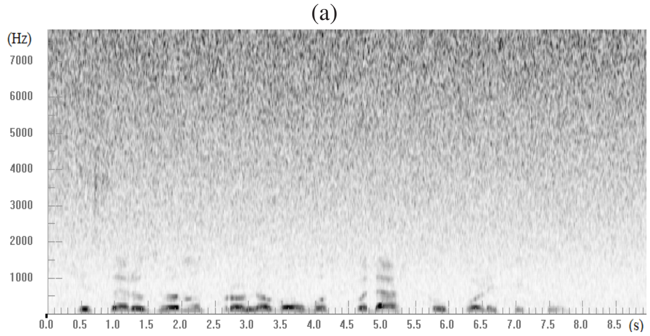


Fig. 5. The simulated experiment in living room.

The author used a dual-microphone system with the range between two mounted microphones is $d = 5(cm)$, the direction of arrival of interest signal is $\theta_s = 90(deg)$. These parameters $nFFT = 512$, frequency sampling $Fs = 16kHz$, overlap 50% were set for capturing the clean speech data. An objective measurement [13] was applied for computing the signal-to-noise ratio (SNR). The waveform and spectrum of observed MA signals is given by Fig. 6.



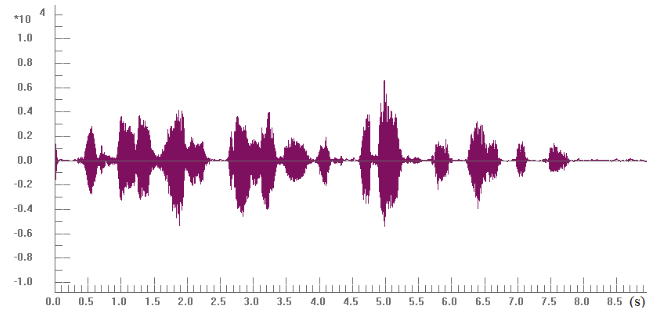
(a)



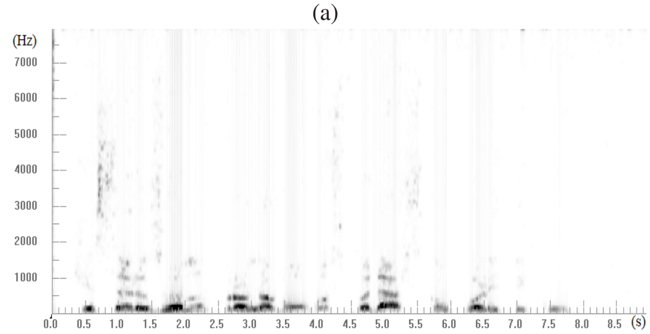
(b)

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

By applying the tdtGSC, the processed signals is given by Fig. 7.



(a)



(b)

Fig. 7. The waveform (a) and spectrogram (b) of processed signal by tdtGSC

Because of the complex and adverse environment, the error of preferred steering vector, the moving head of talker, the different microphone sensitivities, GSC beamformer's evaluation often corrupted. The speech distortion usually occurs and makes speech quality unsatisfactory perceptual metric.

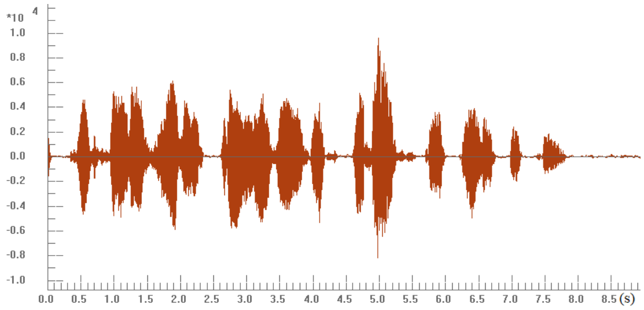
By using hybridGF, the obtained signal is shown in Fig. 8.

Using the proper gain function, the speech enhancement of GSC beamformer is increased. The SNR of tdtGF is increased from 7.3 to 10.3 dB, the speech distortion was reduced to 4.8 dB. The promising result was depicted in Fig and Table 1.

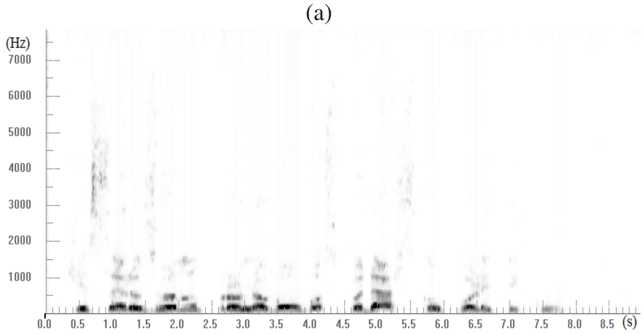
TABLE I
THE SIGNAL-TO-NOISE RATIO (DB)

Method Estimation	Microphone array signal	tdtGSC	hybridGF
NIST STNR	7.5	15.2	22.5
WADA SNR	1.9	9.8	20.1

In a realistic recording environment, the error of estimation of steering vector, the inaccurate MA configuration, the moving talker, the undetermined reasons, the complex and annoying acoustic factors, the difference of microphone quality, the microphone mismatches, the overall GSC beamformer's evaluation often seriously affected. GSC beamformer is very sensitive with the incoherent noise, the imprecise environmental acoustic factor. In this experiment, tdtGSC has shown the ability of removing background noise and preserving the clean speech data. But the speech distortion still existed. Therefore, the necessary requirement is an effective gain function, which



(a)



(b)

Fig. 8. The waveform (a) and spectrogram (b) of processed signal by hybridGF

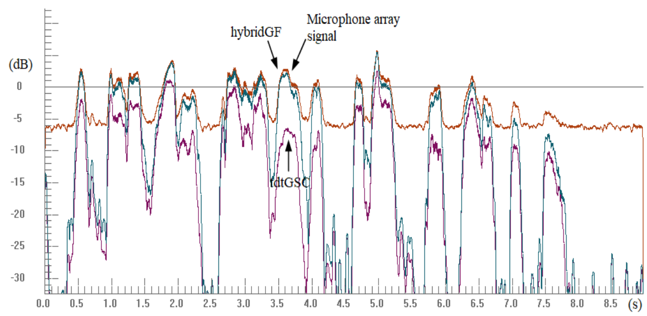


Fig. 9. The energy between the received array signal and processed signals by tdtGSC and hybridGF.

allows achieving the amplitude of the beamformer's output signal approximately the original MA signals.

The author's idea is combining the gain function, which is derived from the prior spatial information, to gain the amplitude of the output signal. The advantage of this direction is exploiting the observed data and has low computation, which is suitable for almost portable acoustic equipment.

V. Conclusion

In this contribution, the author proposed an efficient hybrid gain function for overcoming the drawback of speech distortion in GSC beamformer's performance. The appealing properties of the author's approach is exploiting the prior spatial information to obtain an appropriate combination between multi-channel - based gain function approaches. The numerical simulation in a realistic recording scenario has shown the improvement of reducing the speech distortion to 4.8 dB and increasing the speech quality in the term of signal-to-noise ratio from 7.3 to 10.3 dB. The advantage of the mentioned method is not only recovering the clean speech data, but also is removing the musical noise in comparison between traditional GSC beamformer and suggested technique in adverse situations. This hybrid combination is suitable for almost acoustic speech applications, because of its convenience of incorporating gain functions, which are based on single - multi channel direction. The author's method can be installed into numerous equipment for dealing with other complicated problems, such as speech recognition, speech acquisition.

References

[1] Yuze Sun, Ji Zhang and Xiaopeng Yang, "Design of experimental adaptive beamforming system utilizing microphone array," IET International Radar Conference 2013, Xi'an, 2013, pp. 1-5, doi: 10.1049/cp.2013.0242

[2] Y. Jia, B. Gray and R. Vaughan, "Measurements of microphone array phase and amplitude behavior towards controllable beamforming," 2020 IEEE SENSORS, Rotterdam, Netherlands, 2020, pp. 1-4, doi: 10.1109/SENSORS47125.2020.9278799.

[3] S. S. Priyanka and T. K. Kumar, "GSC Beamforming using Different Adaptive Algorithms for 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.

[4] 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.

[5] A. Barnov, A. Gendelman, A. Schreibman, E. Tzirkel-Hancock and S. Gannot, "A Robust RLS Implementation of the ANC Block in GSC Structures," 2021 29th European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 2021, pp. 261-265, doi: 10.23919/EUSIPCO54536.2021.9615938.

[6] Q. Yang, Y. Zhou, Y. Ma and H. Liu, "A New Frequency-Domain Adaptive Blocking Matrix Design for Generalized Sidelobe Cancellation," 2020 15th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 2020, pp. 109-113, doi: 10.1109/ICSP48669.2020.9321050.

[7] D. -C. Chang and B. -W. Zheng, "Adaptive Generalized Sidelobe Canceler Beamforming With Time-Varying Direction-of-Arrival Estimation for Arrayed Sensors," in IEEE Sensors Journal, vol. 20, no. 8, pp. 4403-4412, 15 April15, 2020, doi: 10.1109/JSEN.2019.2962215.

[8] S. Markovich-Golan, S. Gannot and I. Cohen, "A sparse blocking matrix for multiple constraints GSC beamformer," 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 2012, pp. 197-200, doi: 10.1109/ICASSP.2012.6287851.

[9] 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.

[10] N. Yousefian and P. C. Loizou, "A Dual-Microphone Speech Enhancement Algorithm Based on the Coherence Function," in IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 2, pp. 599-609, Feb. 2012, doi: 10.1109/TASL.2011.2162406.

[11] 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.

[12] 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.

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