The Use of T-F Masking in Dual-Microphone System

Quan Trong The
Information Technologies and Programming Faculty
University ITMO
St.Petersburg, Russia Federation
quantrongthe1984@gmail.com

2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus) | 978-1-6654-0476-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/ElConRus51938.2021.9396338

***Abstract*—In a variety of speech applications, the use of microphone array for extracting desired signal and suppressing interference noise is common widely. In particular, the effectiveness, caused by exploiting a priori spatial information and adaptive signal processing algorithm, allows saving the useful target signal and attenuating background noise. MVDR is one of the most promising filters that are used in speech equipment. MVDR cancels noise and preserves signal, which come from a certain direction. However, the interference and reverberation are not negligible, the degraded performance is not avoidable. In this paper, the author proposed a combination T-F masking with MVDR filter for increasing the robustness performance in scenario with complex noise. With an estimation of T-F masking, the resulted experiments are validated the suggested method, which is suitable for incorporating into multimicrophone system**

Keywords— microphone array; dual-microphone system; MVDR; noise reduction; target speaker;, speech applications

INTRODUCTION

Beamforming technology for speech applications is considered in various of microphone array signal processing with exploiting spatial information for obtaining high direction target speaker in comparison with conventional single channel approach. A lots of important applications, which uses microphone array to attenuate interfering signal, eliminate complex background noise, track the desired target speaker are getting common widely in real-life. Such as hearing aids, that always incorporates with superdirectivity beamformer; speech enhancement, speech recognition, which combine an improved signal processing algorithm and the use of a priori spatial information of direction of arrival or characteristic surrounding environments. The main advantage of microphone array is suppressing unwanted interference signal and saving desired signal, which is the promising purpose in almost speech applications.

Dual-microphone system is one the most basic element of microphone array, causes by it's compact and convenient. In such equipment, like mobile phone, two microphones are mounted and uses all required available parameters to process the array received signals for delivering the highest gain direction, highest interference reduction. The author considered dual-microphone system for developing the proposed method due to it's low cost computation and easily implementation for acquiring target speech signals.

Basic adaptive microphone array signal processing are especially perspective approach in terms of preserving original useful signal while suppressing incoherence, coherence, diffuse or complex background noise. They are based on beamformer, such as delay-and-sum (DAS), differential microphone array (DIF), generalizer sidelobe canceller (GSC), minimum variance distortionless response (MVDR). The adaptive beamforming MVDR aims at extracting the desired target signal at a certain direction while ensuring the noise power at the output signal is minimizes. One of the key parameter, which determines direction of arrival, is the steering vector.

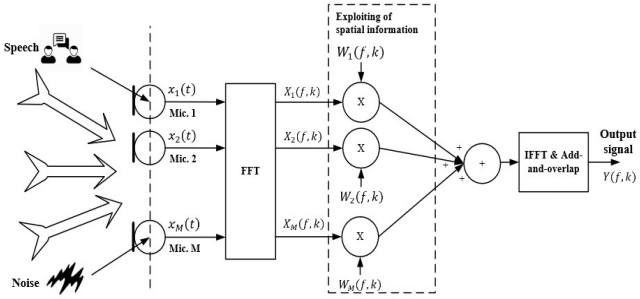

Fig. 1. The scheme of microphone array.

However, the performance of classic beamforming MVDR is degraded, due to many factors, such as the error in microphone arrangements, error of steering vector, different microphone sensitivities. These reason lead to corrupt microphone array signals and reduce the effectiveness and capability of incorporated algorithm.

MVDR beamforming is effective in coherence noise and very sensitive with steering vector. The presence of error of direction of target speech can affect speech quality of final resulted signal. Due to error of determined parameters, speech distortion in the output performance still occurs. At aiming speech distortion reduction, the author proposed an combination MVDR with using speech presence probability in [11]. However, this method works well in scenario with coherence noise. In complex noisy environment, this approach doesn't achieve acceptable result.

In this paper, the author introduces a development of previous work. T-F masking technology is used for suppressing incoherence or interference noise, and other coherence noise will be eliminated by performance MVDR algorithm. T-F masking is a promising technology for suppressing complex noise and increasing robustness of MVDR filter for obtaining desired speech signal and reducing the power level of noise. The T-F masking has various type spectral mask. In this work, CGMM Expectation-Maximization (EM) algorithm is considered for deriving necessary spectral mask.

The rest of paper if organized as following: Section 2 describes the conventional MVDR filter and proposed the

 978-1-6654-0476-1/21/$31.00 ©2021 IEEE

author's previous work. Section 3 presents the improved method for dealing implementation MVDR filter in complex environment, which introduces the scheme of signal processing system. Section 4 contains experiments to demonstrate the capability, the effectiveness and comparison between the promising method and previous work.

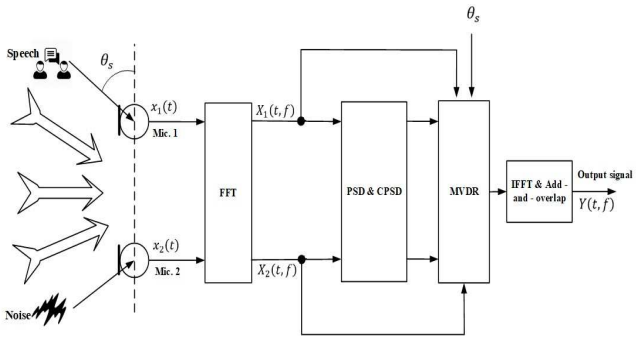

Fig. 2. The MVDR filter.

I. MVDR FILTER

Let's $S(t,f)$ is the representation of original target speaker at current time t and frequency bin f. With a certain distance between two microphone d, sound speech $c = 340(m/s)$, direction of arrival θ_s, the steering vector is $\boldsymbol{d}_s(f,\theta_s) = [e^{j\Phi_s} \quad e^{-j\Phi_s}]^T$, where $\Phi_s = \pi f \tau_0 cos(\theta_s), \tau_0 = d/c$. The observed microphone array signals $\boldsymbol{X}(t,f) = [X_1(t,f) \quad X_2(t,f)]^T$, and additive noises $\boldsymbol{N}(t,f) = [N_1(t,f) \quad N_2(t,f)]^T$ on each microphone.

In the frequency domain, the entire of signal processing microphone array has the following form:

$$\boldsymbol{X}(t,f) = S(t,f)\boldsymbol{d}_s(f,\theta_s) + \boldsymbol{N}(t,f) \tag{1}$$

The optimal solution MVDR algorithm, that based on the constrained criteria of minimization the total noise power without speech distortion:

$$\boldsymbol{W}(t,f) = \frac{\boldsymbol{\Phi}_{XX}^{-1}(t,f)\boldsymbol{d}_s(f,\theta_s)}{\boldsymbol{d}_s^H(f,\theta_s)\boldsymbol{\Phi}_{XX}^{-1}(t,f)\boldsymbol{d}_s(f,\theta_s)} \tag{2}$$

Where $\boldsymbol{\Phi}_{XX}(t,f)$ is cross power spectral densities matrix of array received signals.

The cross spectral matrix of recorded microphone signals: $\boldsymbol{\Phi}_{XX}(t,f) = E\{\boldsymbol{X}(t,f)\boldsymbol{X}^*(t,f)\}$ can be determined as:

$$\boldsymbol{\Phi}_{XX}(t,f) = \begin{Bmatrix} P_{X_1X_1}(t,f) * 1.001 & P_{X_1X_2}(t,f) \\ P_{X_2X_1}(t,f) & P_{X_2X_2}(t,f) * 1.001 \end{Bmatrix} \tag{3}$$

Where $P_{X_iX_j}(t,f), P_{X_iX_i}(t,f), i,j \in \{1,2\}$ defined as:

$$P_{X_iX_j}(f,k) = (1-\alpha)P_{X_iX_j}(f,k-1) + \alpha X_i^*(f,k)X_j(f,k), i,j \in \{1,2\} \tag{4}$$

where α is the smoothing parameter, which in the range $\{0 \dots 1\}$

The promising signal at the output is:

$$Y(t,f) = \boldsymbol{W}^H(t,f)\boldsymbol{X}(t,f) \tag{5}$$

In previous work [11], the author suggested an combination with speech presence probability (SPP) [8-9] for accurate calculating power spectral densities to reduce speech distortion as:

$$P_{X_iX_j}(f,k) = \big(1 - SPP(t,f)\big)P_{X_iX_j}(f,k-1) + SPP(t,f)X_i^*(f,k)X_j(f,k), i,j \in \{1,2\} \tag{6}$$

Unfortunately, in complex environment, where presence of interference or unwanted signals, this method only ensures saving target useful signal and eliminate a little noise power. At the aiming improved noise reduction and gain direction, the author proposed the use of T-F masking for suppressing incoherence and complex noise at the microphone array recorded signals, and then implement the suggested method [11].

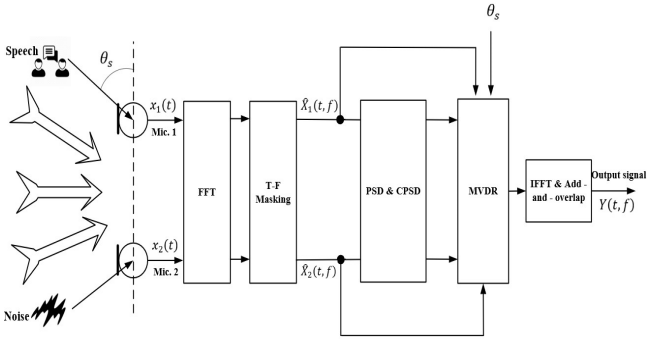

Fig. 3. The proposed method.

II. T-F MASKING

T-F masking in one the most important technology signal processing has many attractive to scholar. T-F masking is spectral mask, which determined from a priori information and filters out microphone array signals. In the author's approach, EM algorithm is considered to calculated an suitable spectral mask. This algorithm determines the posteriori probability noise in each frequency-bin f, each time slot t in a mixture of clean speech, noise. Useful target signal $s(t,f)$ and spatial coherent noise $n_m(t,f)$ according to a following distribution:

$$p\big(n_m(t,f);\phi_{tfm}\big) = \mathcal{N}\big(0,\phi_{tfm}\big) \tag{7}$$

A Gaussian distribution with variance $|n_m(t,f)|^2 = \phi_{tfm}$ and zero mean. The microphone array signals have a complex Gaussian mixture model:

$$p(x(t,f);\lambda) = \sum_{m=1}^{N+1} \alpha_{fm} p(y(t,f)|C(t,f)=l;\lambda)$$
$$p(x(t,f)|C(t,f)=m;\lambda) = \mathcal{N}_c(0, \phi_{tfm}\boldsymbol{B}_{fm}) \quad (8)$$

Where α_{fm} stand by a mixture weight $\left(\sum_{m=1}^{M+1} \alpha_{fm} = 1\right)$, $\boldsymbol{B}_{fm} = \hat{\boldsymbol{h}}_m(f)\hat{\boldsymbol{h}}_m^H(f)$is spatial correlation matrix of noise source m. $C(t,f) = m, m = 1,\dots,M$ acording to source classes, and $C(t,f) = M+1$ denotes to noise class. The important key, the log likelihood can be expressed as the below equation:

$$\mathcal{L}(\lambda) = \sum_t \sum_f log p(x(t,f);\lambda) \quad (9)$$
$$= \sum_t \sum_f log \sum_m \alpha_{fm} \mathcal{N}_c(0, \phi_{ftm} B_{fm})$$

Where $\lambda = \{\lambda_m\} = \{\alpha_{fm}, \phi_{tfm}, \boldsymbol{B}_{fm}\}$ is the parameter set. EM algorithm helps maximum the value of log likelihood. The posteriori probability is defined as $M_m(t,f) = p(C(t,f) = m|x(t,f),\lambda)$.

At E-step: The value of $M_m(t,f)$ is determined in each time-frequency slot:

$$M_m(t,f) = p(C(t,f) = m|x(t,f),\lambda) = \frac{\alpha_{fm} p(x(t,f)|\lambda_m)}{\sum_m \alpha_{fm} p(x(t,f)|\lambda_m)} \quad (10)$$

At M-step: we calculate and update the parameter set λ as:

$$\phi_{tfm} = \frac{1}{M} x^H(t,f) \boldsymbol{B}_{fm}^{-1} x(t,f)$$
$$\boldsymbol{B}_{fm} = \frac{\sum_t^T \frac{M_m(t,f)}{\beta_{tfm}} x(t,f) x^H(t,f)}{\sum_t^T M_m(t,f)} \quad (11)$$
$$\alpha_{fm} = \frac{1}{T} \sum_t^T M_m(t,f)$$

The spectral mask uses the value $M_m(t,f)$ for suppressing the complicated and incoherence noise in complex condition. The cumulative noise mask is calculated as:

$$M_n(t,f) = \sum_{m=1}^{M} M_m(t,f) \quad (12)$$

And microphone array signals is filtered by T-F masking can be expressed as:

$$\hat{X}_1(t,f) = X_1(t,f) \times M_n(t,f)$$
$$\hat{X}_2(t,f) = X_2(t,f) \times M_n(t,f) \quad (13)$$

IV. Experiments

The purpose if this experiment is test the capability of suggested method for improving performance MVDR filter. The evaluation is implemented with dual-microphone system. The microphone array received signals are sampled at 16 kHz Direction of interest signal is $\theta_s = 0^0$, NFFT = 512, smoothing parameter $\alpha = 0.5$, a Hamming window, overlap 50% are used for calculating PSD and spectral mask. WADA SNR [12] measure the ration signal-to-noise (SNR) after implementing MVDR-SPP [11] and suggested method MVDR-SM-SPP. Distance between two microphone $d = 2.25\ (cm)$.

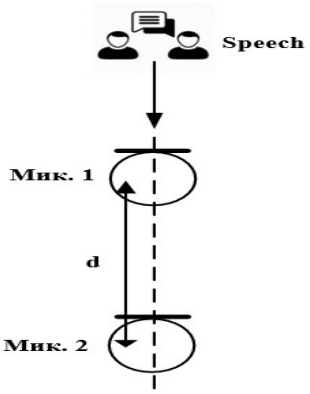

Fig. 4. The scheme of experiment.

The scheme of experiments shown in Figure 4. The amplitude of original signals, which received by dual-microphone system displayed in Figure 5.

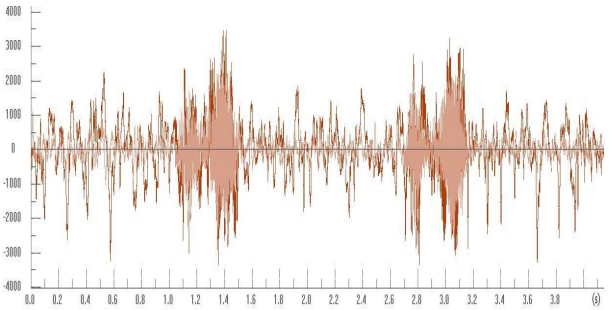

Fig. 5. The amplitude of original signal.

As we can see from Figure 5, in complex noisy condition very low SNR, the component of noise is very difficult to be estimated. This condition affects on performance of MVDR beamformer. The combination of speech presence probability with T-F masking allow extracting desired signal and suppressing background noise, as in Figure 6.

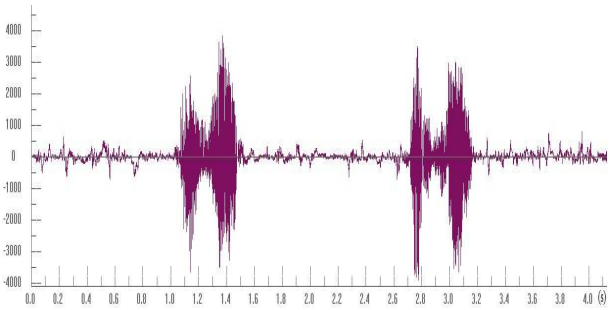

Fig. 6. The output signal by the proposed method.

Figure 7 compared the effectiveness of suggested method (MVDR-EM-SPP) with previous author's method (MVDR-

SPP). MVDR-EM-SPP not only eliminate complicated background noise, also reduce speech distortion in such complex condition surrounding noise. The level of speech distortion is reduced from 3 ÷ 5 (dB). In Table 1, the ratio signal-to-noise (SNR) is measured by WADA SNR.

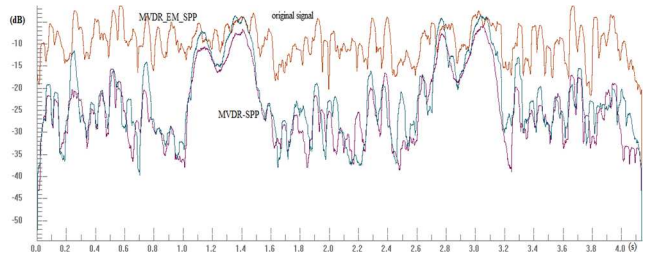

Fig. 7. Comparison energy processed signal by MVDR-SPP and MVDR-EM-SPP.

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

Method Estimation	Original signal	MVDR-SPP	MVDR-EM-SPP
WADA SNR	-0.7	21.0	24.6

The speech quality in term SNR increased from 21.0 to 24.6 (dB). Experiments illustrated that the proposed method allows improving the robustness of MVDR beamformer in condition very complicated and complex environment.

V. CONCLUSION

The purpose is attempting to enhance the performance of MVDR filter with an estimation spectral mask and incorporation with speech presence probability. This work continues developing author's work in study of adaptive MVDR beamformer. The author has discussed the capability of using TF masking for enhanced performance in complex condition. The established technology T-F masking has been proven to be acceptable to MVDR filter, improving the ratio signal-to-noise of the conventional MVDR. In future, the author continues studying others T-F masking for improving speech enhancement in various environments.

REFERENCES

[1] Microphone Arrays / ed. by M. Brandstein, D. Ward. Heidelberg, Germany: Springer-Verlag, 2001. XVIII, 398 p. doi:10.1007/978-3-66204619-7.

[2] Benesty J., Chen J. Huang Y. Microphone Array Signal Processing // Berlin, Germany: Springer-Verlag, 2008.240 p. doi:10.1007/978-3-54078612-2

[3] Benesty J., Chen J., Pan C. Fundaments of Differential Beamforming. // Springer, 2016.122 p. doi:10.1007/978-981-10-1046-0

[4] Benesty J., Cohen I., Chen J. Fundamentals of Signal Enhancement and Array Signal Processing // Wiley, IEEE Press, 2017.440 p.

[5] Lockwood M.E., Jones D. L., Bilger R. C., Lansing C. R., O'Brien W. D., Wheeler B. C., Feng A. S. Performance of time- and frequency domain binaural beamformers based on recorded signals from real rooms // The Journal of the Acoustical Society of America 115, 379 (2004); https://doi.org/10.1121/1.1624064

[6] Sun Z., Li Y., Jiang H., Chen F., Wang Z.. A MVDR-MWF Combined Algorithm for Binaural Hearing Aid System // Proc. 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS), Cleveland, OH, 2018, pp. 1-4, doi: 10.1109/BIOCAS.2018.8584798.

[7] Zhu Y., Fu J., Xu X., Ye Z. Modified Complementary Joint Sparse Representations: A Novel Post-Filtering to MVDR Beamforming // Proc. 2019 IEEE International Workshop on Signal Processing Systems (SiPS), Nanjing, China, 2019, pp. 1-6, doi: 10.1109/SiPS47522.2019.9020522.

[8] Fischer D., Doclo S. Subspace-Based Speech Correlation Vector Estimation for Single-Microphone Multi-Frame MVDR Filtering // Proc. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 856-860, doi: 10.1109/ICASSP40776.2020.9052934.

[9] Fischer D., Doclo S., Habets E. A. P., Gerkmann T. Combined Single-Microphone Wiener and MVDR Filtering based on Speech Interframe Correlations and Speech Presence Probability // Speech Communication; 12. ITG Symposium, Paderborn, Germany, 2016, pp. 1-5.

[10] Cui D., Wang J., Li Z., Li X. A new coherence estimating method: The magnitude squared coherence of smoothing minimum variance distortionless response // Proc. 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Datong, 2016, pp. 1440-1445, doi: 10.1109/CISPBMEI.2016.7852943.

[11] Quan T.T., An Improved MVDR Filter Using Speech Presence Probability // Proc. MICSECS 2019 : The 11th Majorov International Conference on Software Engineering and Computer Systems. Saint Petersburg, December 12-13, 2019.

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