International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-3, January 2020

A Speech Enhancement MVDR Algorithm in Diffuse Noise Field

Quan Trong The

Abstract**:** ***The minimum variance distortionless response (MVDR) beamformer often used in speech application for separating sound source and suppressing ambient, coherent, stationary and non-stationary noise in real complex environment. This paper deals problem speech enhancement in diffuse noise field by using a modified MVDR, which incorporates speech presence probability to estimate auto and cross power spectral densities. This combination gives the advantage of saving target speaker while suppressing background noise. A efficiency post-filtering, which is a function depends on speech presence probability, used for increasing the quality of filtered signal. The performance evaluation demonstrates the ability of proposal algorithm when compared to conventional MVDR.***

Keywords**:** ***speech presence probability, noise reduction, dual-microphone, speech enhancement, post-filtering, filter, microphone array, minimum variance distortionless response.***

I. INTRODUCTION

Perceptual quality and intelligibility of many speech applications, such as speech recognition, biometric system, voice - controlled command, hearing aids, can be seriously degraded by additive noise, unwanted interference, stationary or non-stationary noise. To avoid this problem, several spatial information based multi-microphone have proposed. Microphone array is a processing signal technique, that use a priori knowledge of direction of arrival (DOA) of interest signal, or coherence function of noise field, to reduce background noise and separate one or many sources of target speaker. Microphone array processing [1-3] has received much attention for distant speech recognition, speech enhancement, dereverberation, cock-tail party, due to the ability using spatial information about sound sources and effective processing algorithms, which ensures undistortion of useful signal. Minimum Variance Distortionless Response (MVDR) [4-6] is one of the most popular digital signal processing, which used in microphone-array system. MVDR use the steering vector, that contains DOA of target speaker, and suppress the rest of noise. However, in diffuse noise field, it is always difficult to remove noise without speech distortion.

In this paper, the author proposes to use the speech presence probability (SPP) [7-8] into a modified formulation of auto and cross power spectral densities and post-filtering.

Revised Manuscript Received on January 30, 2020.
***** Correspondence Author
Quan Trong The*, Department of Information Technologies and Programming Faculty, University ITMO, Saint-Petersburg, Russian Federation. Email: quantrongthe1984@gmail.com

This combined algorithm was tested on simulated room with reverberation, mixture of speech and real-world back ground and compared to several known microphone array post-processing algorithms.

II. MVDR ALGORITHM AND SPEECH PRESENCE PROBABILITY

As we know that, dual microphone (MA2) have the significant advantage overcome single-channel algorithm, that spatial information, coherence of noise field, coherence of useful speech, direction of arrival (DOA) based processing MA2 algorithm. Using the geometry distribution, model of signal processing by MA2 represented in Fig. 1. as:

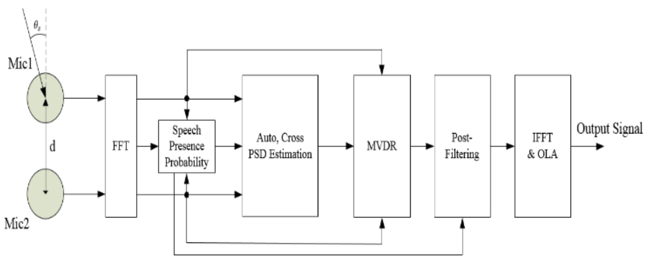

Fig. 1. The scheme of combination MVDR Filter and SPP.

Signal $\boldsymbol{X}(f,k) = [X_1(f,k) \quad X_2(f,k)]^T, \boldsymbol{N}(f,k) = [N_1(f,k) \quad N_2(f,k)]^T$ hereafter denoted as microphone signals, noise signals which captured on dual-microphone system respectively. Steering vector $\boldsymbol{V}(f,k) = [e^{j\Phi_s} \quad e^{-j\Phi_s}]^T$ contains the information of direction of arrival of target speaker. $\Phi_s = \pi f \tau_0 cos(\theta_s)$ is phase shifts between microphones, c is the sound speed (343(m/s)), $\tau_0 = d/c$ is the sound delay between the microphones. Model of signals is defined as:

$$\boldsymbol{X}(f,k) = S(f,k)\boldsymbol{V}(f,k + \boldsymbol{N}(f,k) \quad (1)$$

The optimal solution is need to extract the useful signal and suppress noise. And the enhanced signal is obtained by equation (2):

$$\hat{S}(f,k) = \boldsymbol{W}^H(f,k)\boldsymbol{X}(f,k) \quad (2)$$

Where $\boldsymbol{W}(f,k)$ is vector of coefficients, and $(\ )^H$ denotes the transpose operator.
MVDR filter uses a constraint that minimizes the total output of noise while keeps target speaker unaffected. This leads to the solution in (3)

$$\boldsymbol{W}(f,k) = \frac{\boldsymbol{P}_{NN}^{-1}(f,k)\boldsymbol{V}(f,k)}{\boldsymbol{V}^{H}(f,k)\boldsymbol{P}_{NN}^{-1}(f,k)\boldsymbol{V}(f,k)} \quad (3)$$

In real environment, the matrix of observed signal is used instead of matrix of noise signal. So vector of coefficients can be expressed as:

$$\boldsymbol{W}(f,k) = \frac{\boldsymbol{P}_{XX}^{-1}(f,k)\boldsymbol{V}(f,k)}{\boldsymbol{V}^{H}(f,k)\boldsymbol{P}_{XX}^{-1}(f,k)\boldsymbol{V}(f,k)} \quad (4)$$

With matrix $\boldsymbol{P}_{XX}(f,k)$ calculated as follows:

$$\boldsymbol{P}_{XX}(f,k) = \begin{bmatrix} P_{X_1X_1}(f,k) * 1.001 & P_{X_1X_2}(f,k) \\ P_{X_2X_1}(f,k) & P_{X_2X_2}(f,k) * 1.001 \end{bmatrix} \quad (5)$$

The auto and cross spectral power densities of observed signal, $P_{X_iX_j}(f,k), i,j = 1,2$ is a smoothed – spectral; which computed by a recursive formulation as:

$$P_{X_iX_j}(f,k) = \gamma P_{X_iX_j}(f,k-1) + (1-\gamma)X_i(f,k)X_j^*(f,k) \quad (6)$$

$\gamma \in \{0,1\}$ is a constant smoothing parameter.

A. Modified MVDR Algorithm

Many algorithms in speech signal processing require the information whether speech is present or not. The author's modified MVDR algorithm aim at cancelling diffuse noise field while keeping the target speech signal unaffected. The suggested algorithm based on an estimation of speech presence probability. The author proposed a modified formulation matrix, which used in equation (5). At first, the PSD $P_{X_1X_1}(f,k)$ and CPSD $P_{X_2X_1}(f,k)$ estimated recursively with current $SPP(f,k)$ as follows:

$$P_{X_1X_1}(f,k) = SPP(f,k)P_{X_1X_1}(f,k-1) + \left(1 - SPP(f,k)\right)X_1(f,k)X_1^*(f,k) \quad (7)$$

$$P_{X_2X_1}(f,k) = SPP(f,k)P_{X_2X_1}(f,k-1) + \left(1 - SPP(f,k)\right)X_2(f,k)X_1^*(f,k) \quad (8)$$

But, PSD $P_{X_2X_2}(f,k)$ and CPSD $P_{X_2X_1}(f,k)$ used only instantaneous value of $X_1(f,k)$ and $X_2(f,k)$:

$$P_{X_2X_2}(f,k) = X_2(f,k)X_2^*(f,k) \quad (9)$$

$$P_{X_1X_2}(f,k) = X_1(f,k)X_2^*(f,k) \quad (10)$$

With a new structure of matrix spectral power observed signal, MVDR filter ensures save target speaker without speech distortionless and suppress background noise.

B. Post-Filtering

The OMLSA estimator proposed in [9] computes the gain $G_{H0}(f,k)$ and $G_{H1}(f,k)$. The author's suggested post-filtering expressed as:

$$H(f,k) = G_{H1}^{\frac{SPP(f,k)}{1+SPP(f,k)}} G_{H0}^{1-SPP(f,k)} \quad (12)$$

in experiments with reverberant room.

$$H(f,k) = G_{H0}^{1-SPP(f,k)} \quad (13)$$

$$i,j \in \{1,2\}$$

In condition of diffuse noise field.

III. EXPERIMENTS AND RESULTS

The objective measures of speech quality NIST STNR [10] were used to estimate and evaluate the improvement of suggested algorithm and compared to the conventional MVDR (MVDR-CONV). In this section, the author apply the suggested method (MVDR-Enhance) to speech enhancement problem and evaluate its performance. Two signals was segmented of 512 samples. The sampling frequency is 16 kHz, Hamming window, 50% overlap of frames, and smoothing parameter $\gamma = 0.5$ when calculated auto-spectral, cross-spectral density by recursive formulation.

A. Experiment in a reverberant room

For comparing the sustainability and effectiveness of suggested modified MVDR filter, the author evaluated experiment in the room with condition diffuse noise. Speaker, dual-microphone, which at the center of room, three dimensions of room are shown in Fig.2. With some speech enhancement dual-microphone algorithm [11]: DAS (Delay and Sum), DIF (Differential), Generalized Sidelobe Canceller (GSC); the proposal algorithm reaches better noise reduction. The results shown in Table I.

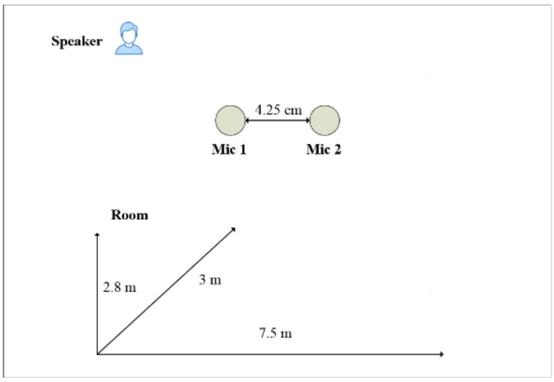

Fig. 2. The scheme of experiments

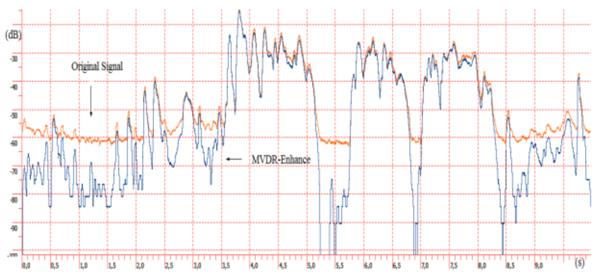

Fig. 3. Energy of original and processed signal by MVDR-Enhance.

Table I: Signal-to-noise ratio (dB)

DAS	GSC	Hypercardioid	Cardioid	MVDR-Enhance
1.5	3.5	5.5	6.5	17.5

B. Experiment in diffuse noise field

The author used the two-channel mixtures of speech and real-world back ground in SiSEC 2010 noisy speech dataset [12]. Background noise signals were recorded via a pair of omnidirectional microphones spaced by 8.6 cm in public environments

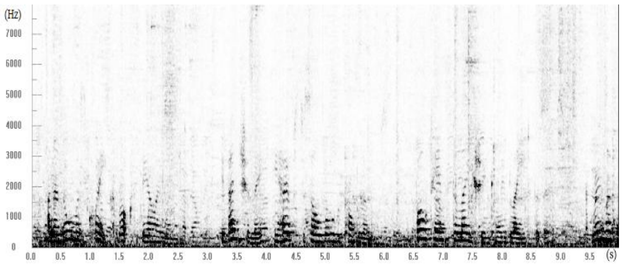

Fig. 4. Spectrogram of original signal.

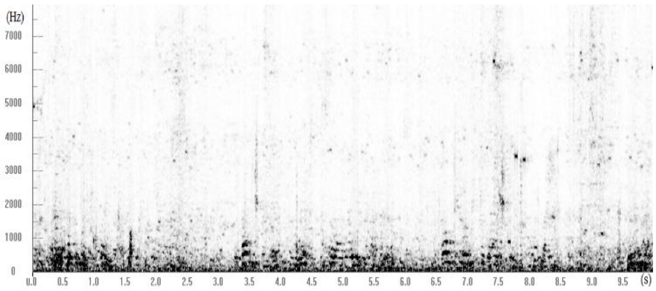

Fig. 5. Spectrogram of processed signal by MVDR-CONV.

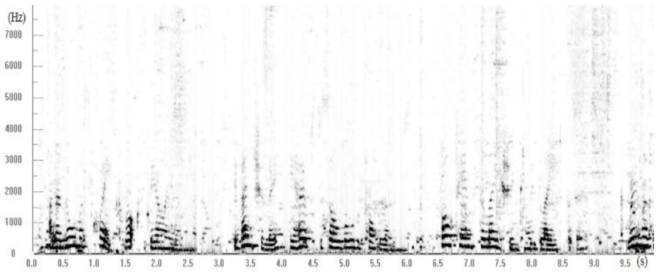

Fig. 6. Spectrogram of processed signal by MVDR-Enhance.

MVDR-CONV not only improves the quality but also degrades the signal in condition of diffuse noise field. MVDR-Enhance can save the target speaker and obviously suppress background noise as demonstrated in Figure 4, 5, 6. The effectiveness of proposed algorithm shown Table II through evaluation of signal-to-noise ratio.

Table II: Signal-to-noise ratio (dB)

Estimation Method	Original signal	MVDR-CONV	MVDR-Enhance
NIST STNR	11.0	8.8	13.5
WADA SNR	5.5	5.6	19.6

IV. CONCLUSIONS

With the improvement in microphone manufacturing technology, the use of microphone array is necessary for human life and has a lot of potential for development. The proposal algorithm, which performed in diffuse noisy acoustic environments, provided increasing of perceptual speech quality than other conventional speech processing algorithm and can be used as fronted application of hearing aids, noise reduction.

REFERENCES

1. Brandstein M. and Ward D. (Eds.). Microphone Arrays: Signal Processing Techniques and Applications, Springer, 2001.
2. Benesty J. and Chen J. Study and Design of Differential Microphone Arrays, Springer, 2013.
3. Benesty J., Chen J., Pan C. Fundamentals of Differential Microphone Arrays, Springer, 2016.
4. Griffiths L.J., Jim C.W. An alternative approach to linearly constrained adaptive beamforming. IEEE Trans. Antennas Propag.,vol.ASSP-30, no.1, pp.27–34, Jan. 1982.
5. 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., Vol.47, No.10, pp.2677–2684, Oct. 1999.
6. Herbordt W., Kellermann W. Analysis of blocking matrices for generalized sidelobe cancellers for non-stationary broadband signals. IEEE International Conference on Acoustics, Speech, and Sig-nal Processing, vol.4, pp.IV-4187, May 2002.
7. Gerkmann T. Unbiased MMSE-Based Noise Power Estimation with Low Complexity and Low Tracking Delay, IEEE TASL, 2012.
8. Gerkmann T., Hendriks R. Noise Power Estimation Based on the Probability of Speech Presence, WASPAA 2011.
9. The objective measures of speech quality. https://labrosa.ee.columbia.edu/projects/snreval/.
10. SiSEC 2010 noisy speech database. http://sisec2011.wiki.irisa.fr/tikiindex79f3.html?page=Two-channel+mixtures+of+speech+and+realworld+background+noise

AUTHORS PROFILE

Quan Trong The is a graduate student within the Computer Science program at Le Quy Don Technical University in 2011. He also graduated with an MS in Information Systems at the Posts and Telecommunications Institute of Technology in 2015. Now he is Ph. D student at National Research ITMO University, Russian Federation. His research interests include Speech Enhancement, Microphone Array.