An Improved MVDR Filter
Using Speech Presence Probability

Quan Trong The[1][0000−0002−2456−9598]

[1] University ITMO, St.Petersburg, Russia
`quantrongthe@itmo.ru`

Abstract. This paper describes an improved minimum variance distortionless response filter in a two-microphone speech enhancement system. Dual-microphone system, which is one of the most basically form of microphone array, has potentially ability of easy implementation, low cost of computation, exploiting of a priori spatial information. The proposed algorithm uses a current estimation information of target speech activity for calculating more precisely auto and cross power spectral densities. Due to the disadvantage of the conventional algorithm is still existing speech distortion, the author introduces a new adaptive technique signal processing, which is suitable for dual-microphone system. The proposal technique evaluated in noisy environments and compared with the conventional algorithm. The results show the reduction of target speech suppression about 8dB, and the quality of estimated speech in the term of the signal-to-noise ratio increased about 15.2 (dB). The enhanced performance provided that suggested algorithm can be incorporated into multi microphone signal processing system. Furthermore, speech presence probability intends to combine with various pre-filtering, post-filtering technique to obtain a certain of noise reduction. The rest of paper is organized as follow. In the Section 2, the scenario of dual-microphone system and combination with speech presence probability are introduced. Section 3 includes the experiments, discussion of significant achievement of the efficient proposed technique. Finally, Section 4 gives the future of the above algorithm's development in different condition of noise.

Keywords: noise reduction, microphone array, dual-microphone, minimum variance distortion response, speech presence probability

1 Introduction

Almost single-channel algorithm aim using spectral subtraction at reduce background noise while maintaining useful speech component. In many speech application, which associated with human life, such as speech coding, communication

system, distant conference require a high speech quality without speech distortion or delay. In real environment, the target speaker always interfered by coherent, incoherent, diffuse noise and others unwanted acoustic sound. Due to highly noisy environments, the recorded signals can be corrupted and it's speech intelligibility is affected. With single algorithm approach, the limitation is existing of original speech suppression and musical noise. Microphone array [1] has been studied in many research articles. Since the crucial spatial information has been exploited, many unresolved problem include attenuation desired signal, residual noise can be easily removed. Microphone array signal processing give us more advantages than mono system. In such scenario, the most important factor is the spatial diversity, which obtained by geometry distribution of microphones. The diversity is combined with some appropriate signal processing techniques to improve captured signals, which contain unknown interferences and additional noises.

Minimum Variance Distortionless Response (MVDR) [2-6] is the most effective algorithm in term of noise reduction while saving the target speaker. MVDR filter processes the input diversity, that is the direction of arrival (DOA) of useful signal, and based on a constraint condition of minimization total output power noise and unaffectedness on desired signal.

However, in real application; due to the rapidly change of undetermined type of noise or complex surroundings, MVDR filter's evaluation has the limitation. In this paper, that author proposes to incorporate the speech presence probability [7-8] (SPP) into the MVDR filter to reduce speech distortion and increase the speech quality of suggested algorithm. Objective measure used for comparing to the conventional MVDR filter. The promising preliminary results provided that, the suggested algorithm can be considered as pre-filtering method in various complex equipments.

2   Combination of MVDR filter and Speech Presence Probability

In this section, the author presented signal processing principles of combination between conventional MVDR filter and speech presence probability.

A dual - microphone system (MA2), in which placed two omnidirectional microphones, is the basic form of microphone array signal processing. The received diversity such as coherence between two noisy signals, direction of arrival, phase difference, power level difference, are easily need to processed to achieve significant noise reduction in compared to single-channel method. The scheme of digital signal processing by MA2 show in Fig 1. The desired target speaker source in the same workplace with dual microphone and relates to axis MA2 an angle θ_s.

We're after here denote the distance between microphones is d, the sound speed is c $(343m/s)$, $\tau_0 = d/c$ is the sound delay. General algorithm is considered in frequency-domain with current frame k, frequency f, desired signal $X(f, k)$, two noisy signals were recorded $Y_1(f, k), Y_2(f, k)$, and additive noise
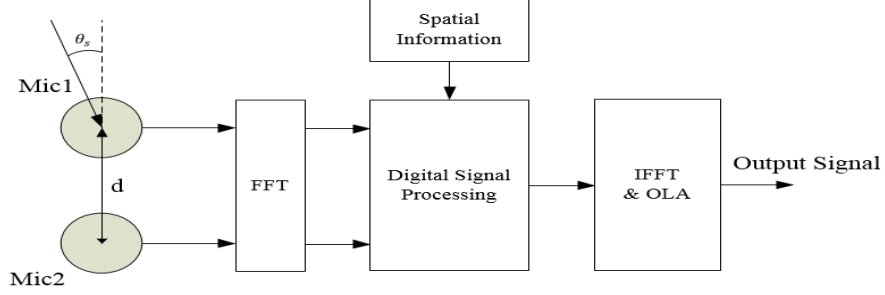
Fig. 1. The scheme of MA2.

$N_1(f, k), N_2(f, k)$. In the form of vector, the representation of short-term Fourier transform can be expressed as:

$$Y_1(f, k) = X(f, k)e^{j\Phi_s} + V_1(f, k) \tag{1}$$

$$Y_2(f, k) = X(f, k)e^{-j\Phi_s} + V_2(f, k) \tag{2}$$

Let's start with $\boldsymbol{Y}(f, k) = [Y_1(f, k) \quad Y_2(f, k)]^T$, $\boldsymbol{V}(f, k) = [V_1(f, k) \quad V_2(f, k)]^T$ and $\boldsymbol{D}(f, \theta_s) = [e^{j\Phi_s} \quad e^{-j\Phi_s}]^T$ with $()^T$ indicates transpose operator, and $\boldsymbol{D}(f, \theta_s)$ is phase shift vector, where $\Phi_s = \pi f \tau_0 cos(\theta_s)$. The equation (1-2) can be rewritten as:

$$\boldsymbol{Y}(f, k) = X(f, k)\boldsymbol{D}(f, \theta_s) + \boldsymbol{V}(f, k) \tag{3}$$

All signal processing algorithm aim finding an optimal solution $\boldsymbol{W}(f, k)$ at ensuring noise reduction and maintaining target speaker. The output signal is obtained by multiplying the coefficients of solution with vector input signals $\boldsymbol{Y}(f, k)$. The estimated signal $\hat{X}(f, k)$ given by:

$$\hat{X}(f, k) = \boldsymbol{W}^H(f, k)\boldsymbol{Y}(f, k) \tag{4}$$

where $()^H$ is the symbol of Hermitian conjugation.

With inverse short-term Fourier and add-overlap, the output signal is transformed into time domain.

The purpose of dual-microphone system is to extract the interest signal. Exploiting of priori diversity is main advantage of MVDR filter. MVDR ensures minimization of the total noise power, while maintaining the undistorted desired signal from given determined direction θ_s. The constraint problem leads to the optimal solution, which can be expressed in form of vector coefficients as follows:

$$\boldsymbol{W}(f, k) = \frac{\boldsymbol{P}_{VV}^{-1}(f, k)\boldsymbol{D}_s(f)}{\boldsymbol{D}_s^H(f)\boldsymbol{P}_{VV}^{-1}(f, k)\boldsymbol{D}_s(f)} \tag{5}$$

where $\boldsymbol{P}_{VV}(f, k)$ is a cross spectral matrix of noise signals, $\boldsymbol{P}_{VV}(f, k) = E\{\boldsymbol{V}(f, k)\boldsymbol{V}^*(f, k)\}$.

Unfortunately, it's always difficult to calculate spectral matrix $\boldsymbol{P}_{VV}(f,k)$, so spectral matrix of observed signals used instead of. The cross spectral matrix of observed signals: $\boldsymbol{P}_{YY}(f,k) = E\{\boldsymbol{Y}(f,k)\boldsymbol{Y}^*(f,k)\}$.

Matrix $\boldsymbol{P}_{YY}(f,k)$ can be computed as:

$$\boldsymbol{P}_{YY}(f,k) = \left\{ \begin{matrix} P_{Y_1Y_1}(f,k)*1.001 & P_{Y_1Y_2}(f,k) \\ P_{Y_2Y_1}(f,k) & P_{Y_2Y_2}(f,k)*1.001 \end{matrix} \right\} \qquad (6)$$

where $P_{Y_iY_i}(f,k), P_{Y_iY_j}(f,k)$ are the smoothed cross-spectra:

$$P_{Y_iY_j}(f,k) = \alpha P_{X_iX_j}(f,k-1) + (1-\alpha)Y_i^*(f,k)Y_j(f,k) \qquad i,j \in \{1,2\} \quad (7)$$

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

So in conventional MVDR fitler, the coefficients become:

$$\boldsymbol{W}(f,k) = \frac{\boldsymbol{P}_{YY}^{-1}(f,k)\boldsymbol{D}_s(f)}{\boldsymbol{D}_s^H(f)\boldsymbol{P}_{YY}^{-1}(f,k)\boldsymbol{D}_s(f)} \qquad (8)$$

In practical implementations, the target speaker may not stay precisely, the captured signals can be influenced by unwanted interference and can not give accuracy direction of arrival of useful signal; furthermore, the different sensitivities, spatial location, frequency response, mismatch of microphones or errors of calculation of steering vector can negative affect on remaining desired speech at the output of system. This produces may led to both poor interference reduction and target speech distortion, and hence cause performance degradation [2].

Requirement of knowledge of speech presence probability is an essential information to estimate and control the updating rate.
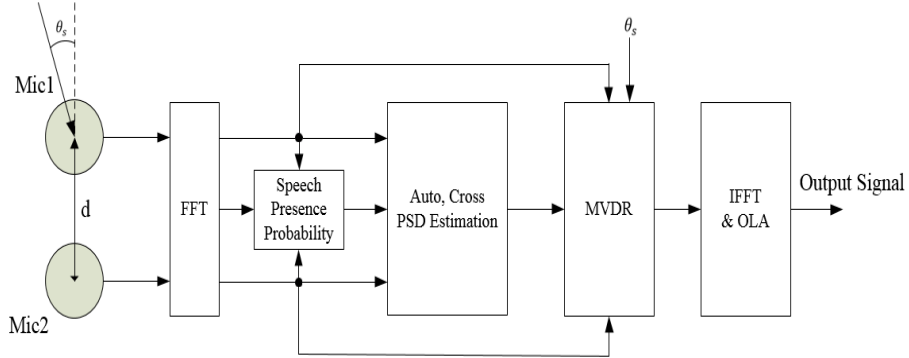

Fig. 2. The scheme of combination.

The author proposed an current estimation of speech presence probability to adjust the auto and cross power spectral densities of observed signals.

$$P_{X_i X_j}(f,k) = SPP(f,k)P_{X_i X_j}(f,k-1)+(1-SPP(f,k))X_i^*(f,k)X_j(f,k) \quad i,j \in \{1,2\}$$
$$(9)$$

The new adaptive suggested algorithm ensures accuracy, exactly and immediately calculating according to the presence or absence of speech components. This approach leads to decrease the speech distortion when compared to conventional MVDR filter. Experiments have confirmed the effectiveness of the proposed solution.

3   Experiments and results

In this section, the suggested algorithm (MVDR-SPP) is performed to deal speech enhancement and reduce speech distortion problem in an anechoeic chamber. Dual microphone were placed on a table at the center of room, distance between two microphones was set $5(cm)$; a speaker stood at distance $2(m)$ from dual-microphone. The purpose of the experiment was to test the MVDR-SPP algorithm on real signals and verify the improvement of reducing speech suppression when compared to conventional MVDR filter (MVDR-CONV). The objective measure NIST STNR [9] used to measure the signal-to-noise ratio (SNR). The scheme of the experiment is shown in Fig. 3. Two noisy recorded signals was sampled at sampling rate $16(kHz)$. For calculating PSD estimation, these necessary parameters: 512 point FFT, a Hamming window, overlap 50% were set.
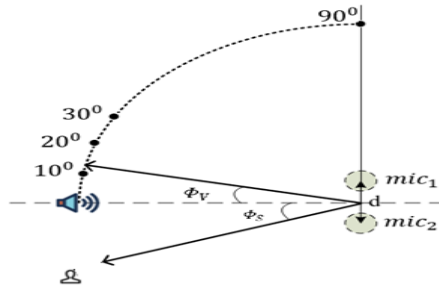

Fig. 3. The scheme of experiments

The target direction was set in the direction of the speaker ($\Phi_s = -30^0$).

In Figure 4, 5; the amplitude and spectrogram of original were demonstrated. NIST STNR measured signal-to-noise ratio was $-0.1(dB)$. After using proposal algorithm, amplitude and spectrogram of estimated signal shown in Figure 6, 7.

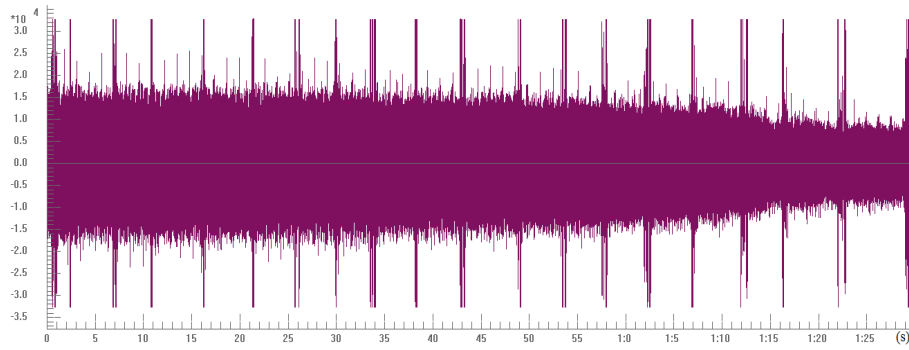Figure 8 shows RMS between original and processed signal by MVDR-SPP.

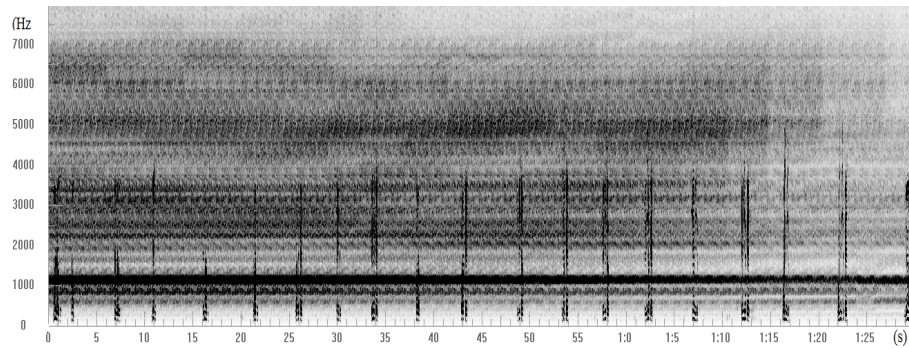**Fig. 4.** Amplitude of original signal.

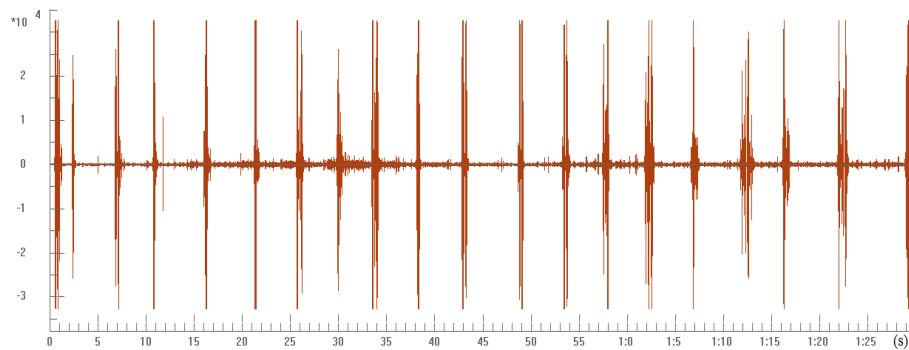**Fig. 5.** Spectrogram of original signal.

Fig. 6. Amplitude of processed signal.

The adaptive algorithm MVDR SPP allows to suppress nonstationary noise, and prove the ability of algorithm. The noise reduction was about 33.5 dB. The target speaker was remained.
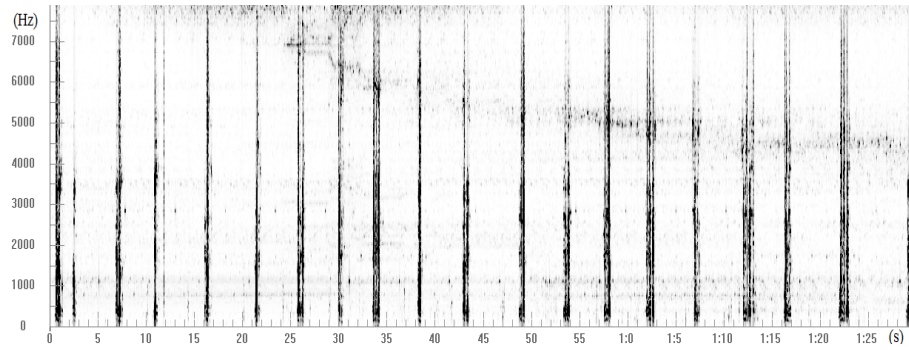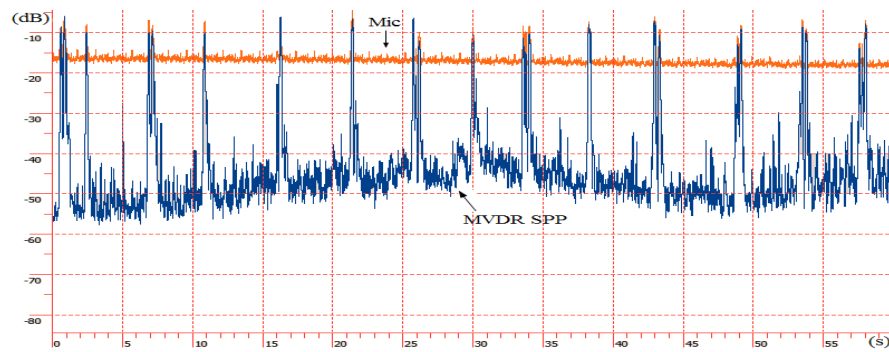
Fig. 7. Spectrogram of processed signal.

Fig. 8. RMS of microphone signal and MVDR SPP output signal ($\Phi_v = 0^0...60^0$).

Fig. 9. Comparison between MVDR SPP and conventional MVDR

From Figure 9; as we can see that algorithm MVDR SPP can save the target speech due to at these frames, the auto and cross spectral were updated according speech presence probability; while conventional MVDR doesn't take in account. The advantage of MVDR SPP is increasing capability of saving speech up to

8dB. The improvement in speech quality presented in Table 1, the increasing of SNR from 26.8 to 42 (dB) provided the capability of suggested algorithm.

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

Method Estimation	Original signal	MVDR-CONV	MVDR-SPP
NIST STNR	4.0	26.8	42

4   Conclusions

This paper addresses the problem of enhancing a speech signal corrupted with additive noise when observations from two microphones are available. The experimental results indicate when the spectral components of the noisy speech changes rapidly, we need an information of speech presence probability to calculate accurately the auto and cross power spectral densities. The algorithm achieves better noise cancellation, less speech distortion, increasing efficiency up to 8 dB and can be used as an efficient front end of speech application. The challenge of time-varying environment is always available, the author continues using other priori spatial diversities to enhance Minimum Variance Distortionless Response filter in different type of noise.

References

1. Brandstein M. and Ward D. (Eds.). Microphone Arrays: Signal Processing Techniques and Applications, Springer, 2001.
2. Ehrenberg L. et al.: Sensitivity Analysis of MVDR and MPDR Beamformers/ IEEE 26-th Convention of Electrical and Electronics Engineers in Israel, 2010, pp. 416-420.
3. Lockwood, M. et al.: Performance of time- and frequency-domain binaural beamformers based on recorded signals from real rooms. J. Acoust. Soc. Am. 115 (1), pp. 379-391, (2004).
4. Stolbov, M., The, Q. Study of MVDR dual-microphone algorithm for speech enhancement in coherent noise presence. Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2019, vol. 19, no.1, pp. 180–183(in Russian).
5. Souden M., Benesty J., Affes S., A study of the LCMV and MVDR noise reduction filters, IEEE Trans.Signal Process., vol. 58, pp. 4925–4935, Sept. 2010.
6. Stolbov, M., Quan Trong The.: Dual-Microphone Speech Enhancement System Attenuating both Coherent and Diffuse Background Noise In: A. A. Salah et al.(Eds.) Proc SPECOM 2019.
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. https://labrosa.ee.columbia.edu/projects/snreval/.