A Coherence Based Dual-Microphone Speech Enhancement in Diffuse Noise Field

Quan Trong The

Abstract*: *In recent years, the problem of improving the quality of noisy signal has always been a dilemma, challenging scientists; especially in diffuse noise conditions. There has been great and growing interest in dual-microphone processing for hands-free speech enhancement. In this paper, the author addresses a improvement of coherence based filter in diffuse noise field. Explicit information about speech presence or absence is necessary in many speech processing applications. The proposal algorithm uses a soft speech presence probability as an optimal parameter to process noisy signal in diffuse noise condition. The evaluation proves that, speech application dual-microphone can be incorporated with this efficiency algorithm for pre-processing.

Keywords*: *coherence function, speech presence probability, microphone array, dual-microphone, diffuse noise, filter, speech enhancement.

I. INTRODUCTION

An interesting and important problem related to spatial audio is capture and reproduction of arbitrary acoustic fields. When a human listens to an audio scene, a multitude of factors are extracted by the brain from the audio streams, including the number of competing foreground sources, their directions, environmental characteristics, presence of background sources, etc. Nowadays, hands-free technology has been rapidly increasing in communication systems. It allows a natural form of communication, as if the communication partner was right next to you, the major problem in these systems is the addition of background noise. In this paper implemented beamforming methods for noise reduction in hands-free communication. A beamformer is an array of microphones, which can do spatial filtering. A microphone array can be used to infer directionality from sampled spatial variations of the acoustic field [1-3].

Dual-microphone, which is one of the most simple type microphone array, used very popular due to their easily implementation, low cost. Coherence function [4-6] based speech enhancement algorithms have been developed for many years. A reliable speech presence probability (SPP) estimator is important to many frequency domain speech enhancement algorithms. It is known that a good estimate of SPP can be obtained by having a smooth a-posteriori signal to noise ratio (SNR) function, which can be achieved by reducing the noise variance when estimating the speech power spectrum. In this paper, the author proposed a new coherence based algorithm, which use an estimation of speech presence probability (SPP) [7-8] to obtain target speaker, while ensuring suppress diffuse noise field. The results provide that, the suggested algorithm is suitable for diffuse noise conditions.

Revised Manuscript Received on January 15, 2020
***** Correspondence Author
Quan Trong The*, Information Technologies and Programming Faculty, ITMO University, Saint-Petersburg, Russian-Federation. Email: quantrongthe1984@gmail.com.

II. OVERVIEW OF COHERENCE FUNCTION

Microphone array is a signal processing technique, that provides the capability of spatial filtering. Dual microphone is the most basic form, due to configuration, the cost of performance and simply implement digital signal processing algorithms. The noise reduction systems play an important role in communications. In this section, an overview of the common signal processing model of a dual-microphone system based coherence function is studied.

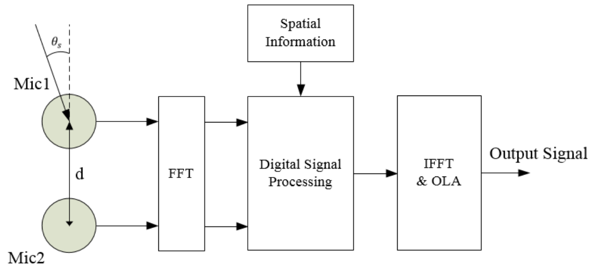

Fig. 1. Digital signal processing of dual-microphone system.

In a noisy environment, dual-microphone receive speech and noise signals; and the signal, $x_1(t), x_2(t)$ at the microphones, which can be expressed as:

$$x_1(t) = s_1(t) + n_1(t) \tag{1}$$

$$x_2(t) = s_2(t) + n_2(t) \tag{2}$$

where $s_1(t), s_2(t)$ is the speech signal at microphones, and $n_1(t), n_2(t)$ are additive noises. In the short-time Fourier transform (STFT) domain, the signal model (1-2) for the $k-th$ frame is:

$$X_1(f,k) = S_1(f,k) + N_1(f,k) \tag{3}$$

$$X_2(f,k) = S_2(f,k) + N_2(f,k) \tag{4}$$

Where $X_1(f,k), X_2(f,k)$ is the STFT of the $x_1(t), x_2(t)$, $S_1(f,k), S_2(f,k), N_1(f,k), N_2(f,k)$ are the STFT of $s_1(t), s_2(t), n_1(t), n_2(t)$respectively for the k −th frame and f −th frequency bin.

Coherence function of two signals $X_1(f,k)$ and $X_2(f,k)$ is expressed as follow:

$$\Gamma_{X_1X_2}(f,k) = \frac{P_{X_1X_2}(f,k)}{\sqrt{P_{X_1X_1}(f,k)P_{2X_2}(f,k)}} \quad (5)$$

In the equation (5), $P_{X_1X_1}(f,k)$, $P_{X_2X_2}(f,k)$ and $P_{X_1X_2}(f,k)$ are the Power Spectrum Density (PSD) of $X_1(f,k)$, $X_2(f,k)$ and Cross Power Spectrum Density (CPSD) of $X_1(f,k)$ and $X_2(f,k)$,respectively. The CPSD and PSD can be estimated recursively as:

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

With $i,j \in \{1,2\}$. Where γ is a smoothing factor in the range [0,1].

For obtaining enhanced signal, in [6] a method of using the signal-to-noise ratio has been proposed in coherent noise environment.

$$H(f,k) = 1 - \left|\Gamma_{X_1X_2}(f,k)\right|^{L(f,k)} \quad (7)$$

where function $L(f,k)$ depends on ratio signal-to-noise.

III. THE PROPSED ALGORITHM

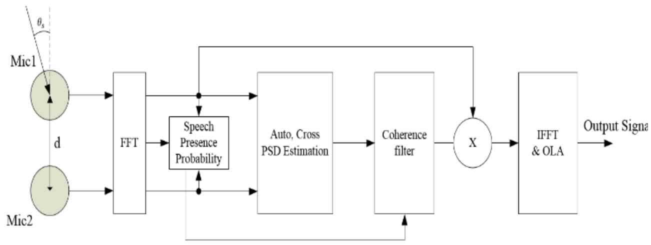

Fig. 2. The proposal algorithm.

In condition of diffuse noise field, the author introduces a new adaptive algorithm, which also exploits coherence function and speech presence probability, to estimate target speaker. The proposal coherence filter is expressed as:

$$H(f,k) = \left(1 - \left|\Gamma_{X_1X_2}(f,k)\right|\right)^{(1-SPP(f,k))} \quad (8)$$

IV. EXPERIMENTS AND RESULTS

The author used the two-channel mixtures of speech and real-world back ground in SiSEC 2010 noisy speech dataset [10]. Background noise signals were recorded via a pair of omni-directional microphones spaced by 8.6 cm in public environments. The objective measure NIST STNR, WADA SNR [9] is used to estimate the quality of original and processed signal by suggested algorithm. In this section, the author apply the suggested method to the speech enhancement problem and evaluate its performance. All two signals were sampled with frequency 16 kHz. For the PSD estimation the author used a 512 point FFT, a Hamming window, and an overlap of 256 samples, $\gamma = 0.5$. The purpose of the experiment was to test the proposed algorithm on real signals.

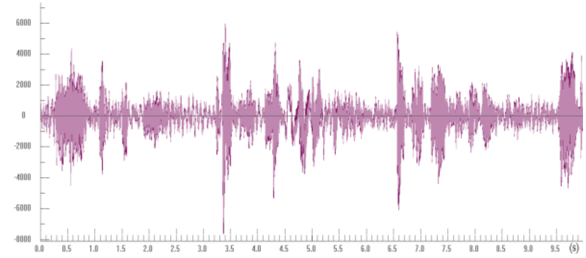

Fig. 3. Amplitude of original signal.

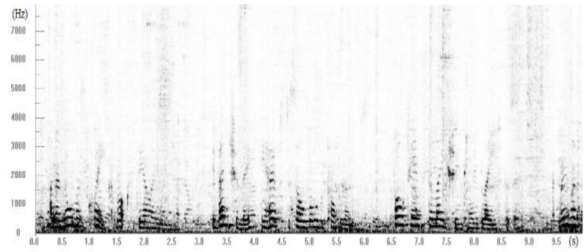

Fig. 4. Spectrogram of original signal.

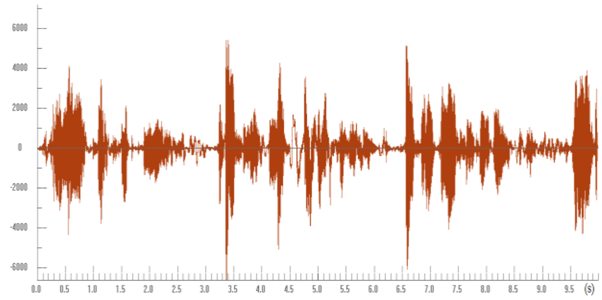

Fig. 5. Amplitude of processed signal.

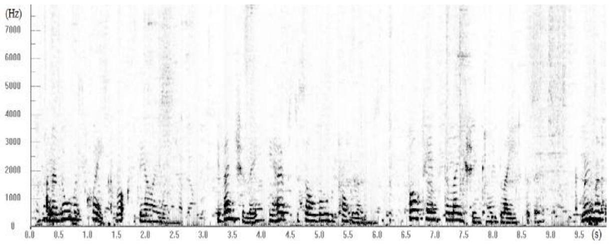

Fig. 6. Spectrogram of processed signal.

Figure 3, 4, 5, 6 demonstrate that the suggested algorithm can save target speaker while reducing background coherent noise. It can be clearly seen that while decreasing the background diffuse noise, algorithm doesn't attenuate the speech components resulting in output signal. There is no musical noise, or speech distortion. At speech frames, when $SPP(f,k) \approx 1$; to $1 - SPP(f,k)$ tends to 0, so coherence filter $H(f,k) \approx 1$; this leads to speech components obviously kept in output signal. In contrast, when $SPP(f,k) \approx 0$, and $(1 - SPP(f,k)) \approx 1$, but $(1 - |\Gamma_{X_1X_2}(f,k)|) < 1$ and tends to 0; so at the noise frame, $H(f,k) \approx 0$, the noisy component suppressed.

The ratio signal-to-noise SNR is improved to 12.3 ÷ 18.8 (dB) as in the Table I.

Table-I: The signal-to-noise ratio SNR (dB)

Estimation method	Signal	
	Original	*The proposed algorithm*
NIST STNR	11.0	12.3
WADA SNR	5.5	18.8

V. CONCLUSION

Dual-microphone beamformer algorithms are promising solutions for noise reduction in hearing aids as they exploit the spatial distribution of the interfering signals and therefore in general lead to less signal distortion than single channel algorithms. In this work, the author have introduced a favorable algorithm, which capable with problem speech enhancement in diffuse noise field. These results sufficient to demonstrate the increasing of signal-to-noise ratio, and ensures the quality of processed signal.

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. Yousefian N., Loizou P. C. A Dual-Microphone Algorithm That Can Cope With Competing-Talker Scenarios. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 21, No. 1, January 2013.
5. Yousefian N., Kokkinakis K., Loizou P.C. A coherence-based algorithm for noise reduction in dual-microphone applications. Proc. Eur. Signal Process. Conf. (EUSIPCO10), Alborg, Denmark, Aug. 2010, pp. 1904–1908.
6. Yousefian N., Loizou P. C. A dual-microphone speech enhancement algorithm based on the coherence function. IEEE Trans. Speech Audio Process., vol. 20, no. 2, pp. 599–609, Feb. 2012.
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 measure: Available: https://labrosa.ee.columbia.edu/projects/snreval/.
10. SiSEC 2010 noisy speech database. Available: http://sisec2011.wiki.irisa.fr/

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.