Dual-Channel Speech Enhancement Based On Speech Presence Probability

Quan Trong The

Abstract*: This paper introduces technology to improve sound quality, which serves the needs of media and entertainment. Major challenging problem in the speech processing applications like mobile phones, hands-free phones, car communication, teleconference systems, hearing aids, voice coders, automatic speech recognition and forensics etc., is to eliminate the background noise. Speech enhancement algorithms are widely used for these applications in order to remove the noise from degraded speech in the noisy environment. Hence, the conventional noise reduction methods introduce more residual noise and speech distortion. So, it has been found that the noise reduction process is more effective to improve the speech quality but it affects the intelligibility of the clean speech signal. In this paper, we introduce a new model of coherence-based noise reduction method for the complex noise environment in which a target speech coexists with a coherent noise around. From the coherence model, the information of speech presence probability is added to better track noise variation accurately; and during the speech presence and speech absent period, adaptive coherence-based method is adjusted. The performance of suggested method is evaluated in condition of diffuse and real street noise, and it improves the speech signal quality less speech distortion and residual noise.

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

I. INTRODUCTION

Significant knowledge about microphone arrays has been gained from years of intense research and product development [1-3]. Speech acquisition, speech enhancement, speech recognition, surveillance, warfare with large number of microphone to reduce noise and save target speaker. Now microphone array are widely applied in almost technology industries. The main advantage of multi-microphone is their ability using the additional spatial information: direction of arrival (DOA), coherence noise field, geometry of microphone arrays for pre-processing or pre-filtering to suppress background noise. The effectiveness of performance increases corresponding the more number of microphone or more valuable spatial priori information. The speech quality is improved by reliable processing algorithm, increasing number of microphones. Dual-microphone is one of the most Basic, simple form of microphone array, due to lower cost, demand on calculation time.

Coherence based methods are known as easily implementation of dual channels methods [4-6]. These methods are based that the coherence between noise, which received on microphones, different with coherence between target speakers. Coherence base method exploits the characteristic to enhance noisy signal

II. COHERENCE-BASED METHODS

In dual-microphone system, the two noisy signals $x_1(t), x_2(t)$ at the microphones can be modelled as following equations:

$$x_1(t) = s_1(t) + n_1(t) \qquad (1)$$

$$x_2(t) = s_2(t) + n_2(t) \qquad (2)$$

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:

$$X_1(f,k) = S_1(f,k) + N_1(t) \qquad (3)$$

$$X_2(f,k) = S_2(f,k) + N_2(t) \qquad (4)$$

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

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

$$\Gamma_{X_1X_2}(f,k) = \frac{P_{X_1X_2}(f,k)}{\sqrt{P_{X_1X_1}(f,k)P_{X_2X_2}(f,k)}} \qquad (5)$$

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

$$P_{X_iX_i}(f,k) = \alpha P_{X_iX_i}(f,k-1) + (1-\alpha)X_i(f,k)X_i^*(f,k) \qquad (6)$$

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

In [6], the authors proposed the following gain function:

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

Where $L(f,k)$ is a function that depends on the estimated SNR at frequency bin f.

Revised Manuscript Received on January 06, 2020.

***** Correspondence Author

Quan Trong The, Department of Information Technologies and Programming, University ITMO, Saint-Petersburg, Russian Federation. Email: quantrongthe1984@gmail.com

III. SPEECH PRESENCE PROBABILITY AND THE PROPOSED METHOD

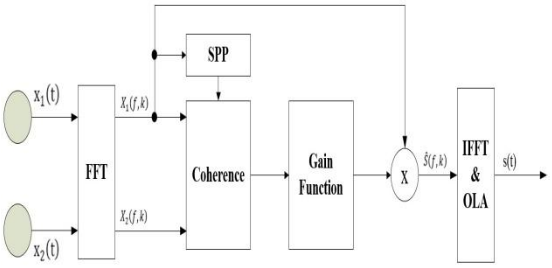

Fig. 2. Block diagram of the proposed method

Many algorithms in speech signal processing require the information whether speech is present or not. Speech enhancement algorithms aim at reducing nonstationary, nonstationary and unwanted interference while ensuring undistortion of target speaker. So one of the most important of coherence-based filter is a precisely information of speech presence or absence. The performance of filter depends on estimation algorithm of speech presence probability. In [7-8], the authors analyzed the minimum mean square error (MMSE) based spectral noise power estimator and presented a soft speech presence probability (SPP). The coherence-based proposed method exploited $SPP(f,k)$ frame-by-frame, as:

$$L(f,k) = \frac{SPP(f,k)}{1-SPP(f,k)} \tag{8}$$

As we know that, $SPP(f,k)$ in the range 0.1 ÷ 0.9. When the speech absent, the signals in two microphone contain only noise, so in these frames, $SPP(f,k)$ reaches the minimum value, and function $L(f,k)$ tends to a very small value which equal approximately 0.01, the transfer function $H(f,k)$ also equal one, and we received absolutely noise reduction at the output signal because $Y_{out}(f,k) \approx X_1(f,k) - X_2(f,k)$. When the speech present, $SPP(f,k))$ reaches the larger value, function $L(f,k)$ also tends to larger, but transfer function $H(f,k)$ tends to zero, because $\left|\Gamma_{X_1X_2}(f,k)\right| < 1$; so at the output signal the desired speech still remaining, $Y_{out}(f,k) \approx X_1(f,k)$.

IV. EXPERIMENTS AND RESULTS

In this section, the author apply the suggested method (Coh-SPP) to speech enhancement problem and evaluate its performance. The objective measures of speech quality NIST STNR, WADA SNR [9] used to estimate and evaluate the improvement of suggested algorithm. Two observed signals were segmented of 512 samples. The sampling frequency is 16kHz, Hamming window, 50% overlap of frames, and smoothing parameter $\alpha = 0.5$ when calculated auto-spectral, cross-spectral density by recursive formulation.

A. Experiment in reverberant room

The purpose of the experiment was to compare the effectiveness of the suppression of diffuse noise in the room. The scheme of the experiment is shown in Fig. 3.

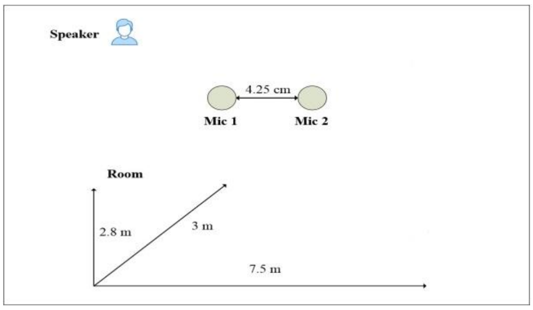

Fig. 3. The scheme of experiments

The experiment was conducted in an office space, T_{60} = 320*ms*. Noise was generated through the speaker at a distance about 5*m* from the microphone array. The result is demonstrated in Fig 4.

Noise reduction is about 20 ÷ 30 (dB).

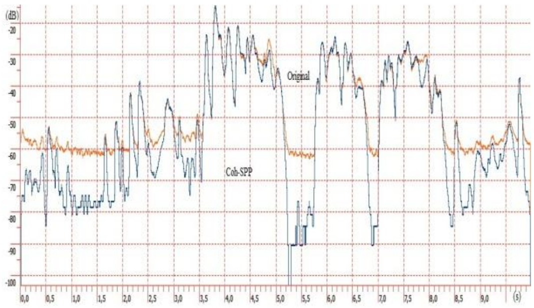

Fig. 4. Energy of original signal and processed signal by Coh-SPP

B. Experiments in diffuse noise filed

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 omnidirectional microphones spaced by 8.6 cm in public environments.

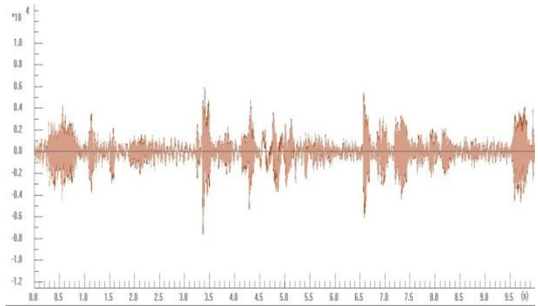

Fig. 5. Amplitude of original signal.

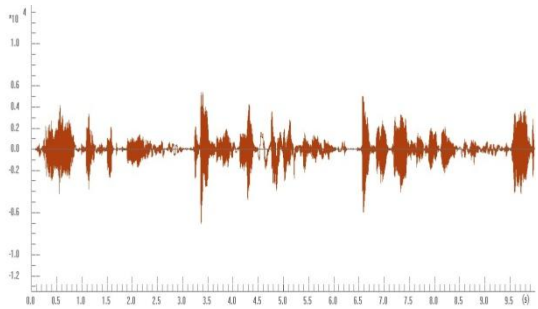

Fig. 6. Amplitude of processed signal by Coh-SPP.

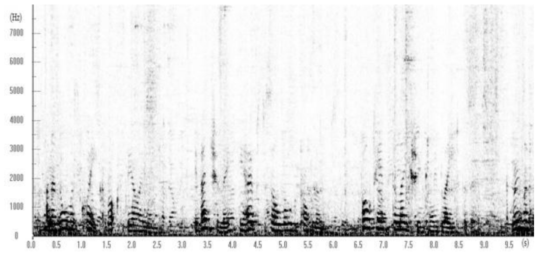

Fig. 7. Spectrogram of original signal.

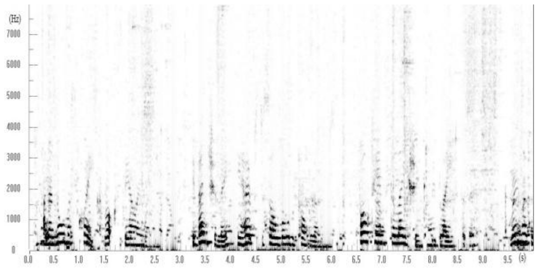

Fig. 8. Spectrogram of processed signal by Coh-SPP.

Fig 5, 6, 7, 8 show amplitude and spectrogram of original, processed signal by Coh-SPP. From Table 1, Coh-SPP increases the quality of signal after filtering.

Table I Ratio Signal-To-Noise Snr (Db) In Diffuse Noise

Method Estimation	Original signal	Coh-SPP
NIST STNR	11.0	14.5
WADA SNR	5.5	23.0

C. Experiments in street noise

Experimental results prove the ability of the algorithm to suppress nonstationary interference while maintaining the speech signal of the target speaker. Spectrogram and energy are shown in Fig 9, 10, 11. The ratio signal-to-noise SNR improved to 31.8 ÷ 47.5 (dB).

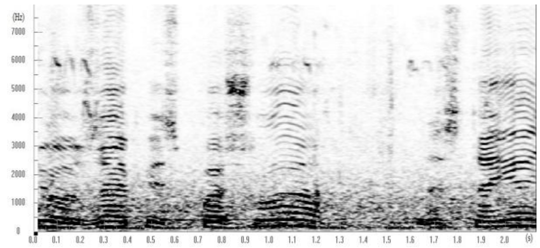

Fig. 9. Spectrogram of original signal.

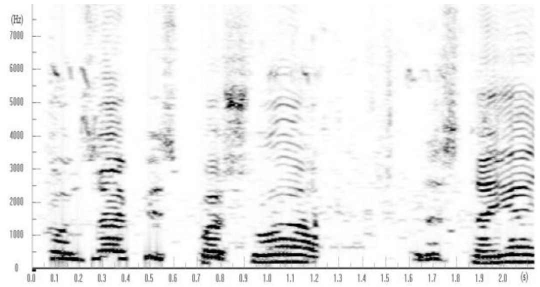

Fig. 10. Spectrogram of processed signal by Coh-SPP.

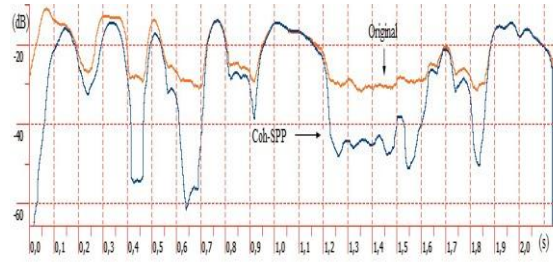

Fig. 11. Energy of original and processed signal by Coh-SPP.

Table Ii Ratio Signal-To-Noise Snr (Db) In Street Noise

Method Estimation	Original signal	Coh-SPP
NIST STNR	15	31.8
WADA SNR	9.9	47.5

V. CONCLUSION

In this paper, the author have proposed a robust approach of coherence based method, that only requires the a priori knowledge of speech presence probability. Simulation results demonstrate that the proposed approach is efficiently suppress background noise and remain the whole speech signal. Digital speech applications such as mobile phones, hearing aids, etc., must be robust to acoustical background noise and reverberation. For this reason, such devices are often equipped with noise reduction/speech enhancement algorithms. The proposal method deals with problem of saving target speaker while suppressing diffuse noise in condition of reverberant room, street. The speech presence probability based above approach is appropriate for multi-microphone system to enhance sound quality.

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. Le Bouquin R,. Faucon G. "Using the Coherence Function for Noise Reduction", IEE Proc. 139(3): 276-280, June 1992.
5. Allen J.B., Berkley D.A,. Blauert J. 'Multi microphone signal processing technique to remove room reverberation from speech signals', J. Acoust. Soc. of Amer., 62(4):912–915, 1977.
6. Yousefian N., Kokkinakis K., Loizou PC. "A coherence-based algorithm for noise reduction in dual-microphone applications", 2010 18th European Signal Processing Conference, 1904-1908.
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/.
10. http://sisec2011.wiki.irisa.fr/tiki-index79f3.html?page=Twochannel+mixtures+of+speech+and+real-world+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.