A Two Microphone-Based Approach for Speech Enhancement in Adverse Environments

Kai Li, Yanmeng Guo, Qiang Fu, Yonghong Yan
Key Laboratory of Speech Acoustics and Content Understanding,
Institute of Acoustics, Chinese Academy of Sciences, China

***Abstract*—In this paper, we present a two microphone-based speech enhancement algorithm using spatial information-based spectral amplitude estimation in adverse environments. A target presence probability estimator based on Bayes detection rules using both phase difference and magnitude squared coherence is proposed for soft decision. Experimental results show that the proposed algorithm achieves better noise suppression and lower signal distortion over the comparative dual-channel algorithms.**

I. Introduction

Microphone array has been widely used to improve the performance of speech communication and automatic speech recognition (ASR) systems in adverse noise environments. Among them, dual-channel speech enhancement methods are receiving more attention recently from industries for mobile headsets, hearing aids, or other compact distant-talking speech communication and recognition applications because of their low cost implementation and acceptable performance. Numerous dual channel algorithms have already been proposed, such as the adaptive-based method [1], coherence-based method [2], and the phase difference-based method [3] [4]. These methods show good performance in specific noise field. However, in real environment, the performance of these algorithms deteriorate when diffuse noises coexist with directional noises.

In this paper, we consider the problem of speech enhancement when both directional noises and diffuse noises coexist in adverse environments. Ephraim and Malah derived a well-known estimator which minimizes the mean-square error of the short-term log-spectral amplitude (MMSE-LSA) [5]. However, it is inefficient at attenuating interfering speech components, since it is not able to differentiate such components from the target speech components. Using spatial information can overcome this problem. Phase difference represents the direction of arrival of signals and shows good performance in coherent noise field. Magnitude squared coherence (MSC) represents the acoustical noise field information and shows good target detection performance in diffuse noise field. These two kinds of spatial information share complementary advantages. In order to use these spatial information of signals and acoustical environments to improve the performance of traditional spectral amplitude estimator based speech enhancement, we reformulate statistical MMSE-LSA algorithm and propose a dual channel speech enhancement method as shown in Fig.1, which mainly consists of three parts: a two-channel gain function, a spatial information controlled soft decision noise estimator and a target presence probability estimator.

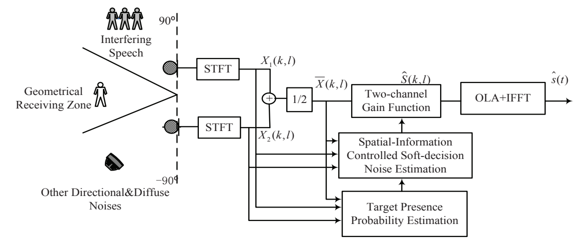

Fig. 1. Block diagram of the proposed speech enhancement algorithm.

II. The Proposed Dual-Channel Speech Enhancement Algorithm

A. Dual-Channel Spectral Amplitude Estimation

Denote $A = |S(k,l)|$ as the clean speech spectral amplitude, and $\hat{A}$ as its estimate. The gain function of MMSE-LSA estimator [5] which minimizes $E\{(\log A - \log \hat{A})^2\}$ can be simplified by a function of $a\ priori$ SNR and $a\ posterior$ SNR [6], so the LSA estimator can be written as follows:

$$\hat{A} = \exp\{E[\log A|(\overline{X})]\} = G(\xi(k,l), \gamma(k,l))|\overline{X}| \tag{1}$$

where $\xi(k,l) = \lambda_s(k,l)/\lambda_n(k,l)$ and $\gamma(k,l) = |\overline{X}|^2/\lambda_n(k,l)$ is the a priori and a posteriori SNR respectively. $\overline{X}$ denotes the averaged two channel observed spectral components. $\lambda_s(k,l)$ and $\lambda_n(k,l)$ are the short-term spectrum of speech and noise signals in frame l and frequency bin k respectively.

In order to use the spatial information of signals and acoustical environments to improve the performance of traditional spectral amplitude estimator. We derive a dual-channel log-spectral amplitude estimator based on spatial information using both phase difference $\Delta\Psi$ and magnitude squared coherence Γ, so that they can benefit complementary advantages when both directional noises and diffuse noises coexist. Without loss of generality we omit k and l in the following equations for simplicity.

$$\begin{aligned} \hat{A} &= \exp\{E[\log A|(\overline{X}, \Delta\Psi, \Gamma)]\} \\ &= G(\xi(\Delta\Psi, \Gamma), \gamma(\Delta\Psi, \Gamma))|\overline{X}| \end{aligned} \tag{2}$$

The $a\ priori$ and $a\ posteriori$ SNRs are mainly dependent on the dual-channel noise spectrum estimation which will be described in II-B.

B. Spatial Information Controlled Noise Spectrum Estimation

Two hypotheses are defined here:

(a) $H_1(k,l)$: speech is present and from the geometrical receiving zone, as shown in Fig.1;

(b) $H_0(k,l)$: speech is absent or present but outside the geometrical receiving zone;

We incorporate spatial information into soft decision based noise estimation [6] and propose a spatial information controlled soft decision method. One of the important parts of the two-channel noise estimation is $P[H_1|\overline{X},\Delta\Psi,\Gamma]$, which is a dual channel target presence probability estimator.

With Bayes rules and the assumption of the statistical independence of $\overline{X}$, $\Delta\Psi$ and Γ, the target presence probability can be derived by

$$P(H_1|\overline{X},\Delta\Psi,\Gamma) = [1+\Lambda^{-1}\frac{1-P(H_1|\Delta\Psi,\Gamma)}{P(H_1|\Delta\Psi,\Gamma)}]^{-1} \tag{3}$$

The likelihood ratio Λ can be expressed with the $a\ priori$ and $a\ posteriori$ SNRs as [6]:

$$\Lambda = \frac{1}{1+\xi(\Delta\Psi,\Gamma)}\exp[\frac{\xi(\Delta\Psi,\Gamma)}{\xi(\Delta\Psi,\Gamma)+1}\gamma(\Delta\Psi,\Gamma)] \tag{4}$$

The $a\ priori$ target presence probability can be defined as:

$$P(H_1|\Delta\Psi,\Gamma) = P(H_1|\Delta\Psi)P(H_1|\Gamma) \tag{5}$$

III. Experimental Evaluations

The experiment was carried out in a real room of $4.1\text{m}\times 4\text{m}\times 3\text{m}$ with a reverberation time (RT_{60}) of 300ms. The inter-microphone distance is 5cm and the system is implemented under a sampling rate of 8kHz. The predefined geometrical receiving zone is $\pm 20^\circ$ using phase difference information. Five male and five female data, which are randomly selected from TIMIT database are used as target speech signals. Besides the background diffuse noises, a background music noise and a competing speaker are located at 60° and 90° of the array respectively. All the sound sources are 1m away from the microphone array. Dual-channel GSC algorithm in [1], PEF algorithm in [3] and PDSAE algorithm proposed in [4] are used for comparison.

The results of the above methods are evaluated in Noise Reduction (NR), Log-spectral Distance (LSD), and subjective study of speech waveforms and spectrograms, as shown in Fig.2 and Fig.3. We can see that, compared with the other three algorithms, our proposed algorithm shows considerable improvement in terms of noise reduction and LSD over various SNR conditions.

IV. Conclusions

We have described a dual-channel speech enhancement method using spatial information-based speech estimator and spatial information controlled soft decision noise spectrum estimator. We reformulated statistical MMSE-LSA algorithm, in which spatial information including both phase difference and magnitude squared coherence between channels is incorporated so that they can share complementary advantages. The performance of the proposed algorithm is verified in some experiments and shows better evaluation results than other competitive two channel speech enhancement algorithms.

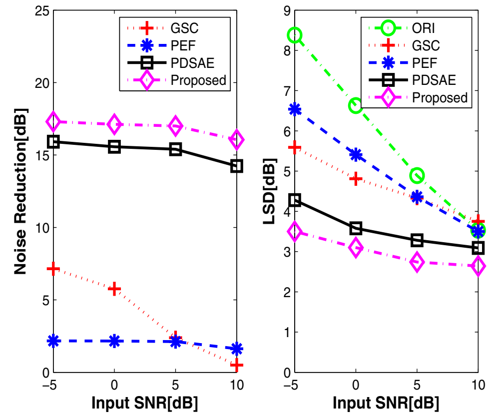

Fig. 2. Performance comparisons under real room experiment configuration.

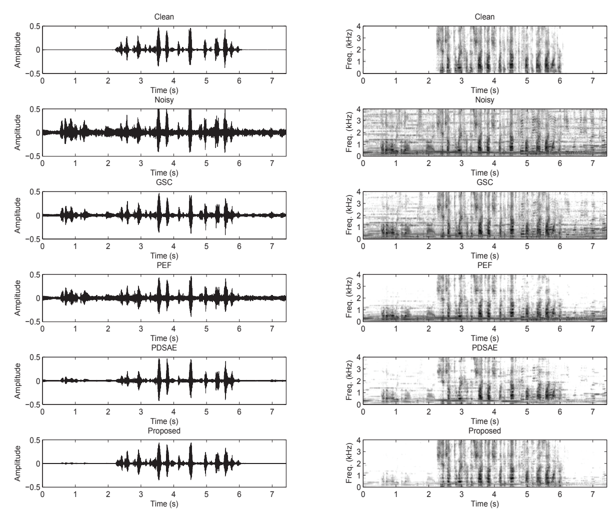

Fig. 3. Speech waveforms and spectrograms. Noise conditions: real room environment with music noise at 60° and interfering speech at 90°, SNR of 5dB.

References

[1] J. Chen, K. Phua, L. Shue, and H. Sun, "Performance evaluation of adaptive dual microphone systems," *Speech Commun.*, vol. 51, no. 12, pp. 1180–1193, 2009.

[2] R. Le Bouquin, A. A. Azirani, and G. Faucon, "Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 5, pp. 484–487, 1997.

[3] Parham Aarabi and Guangji Shi, "Phase-based dual-microphone robust speech enhancement," *IEEE Trans. on Syst. Man. Cyb*, vol. 34, no. 4, pp. 1763–1773, 2004.

[4] K. Kim, S. Y. Jeong, J. H. Jeong, K. C. Oh, and J. Kim, "Dual channel noise reduction method using phase difference-based spectral amplitude estimation," in *Proc. ICASSP-2010*, 2010, pp. 217–220.

[5] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 33, no. 2, pp. 443–445, 1985.

[6] N. S. Kim and J. H. Chang, "Spectral enhancement based on global soft decision," *IEEE Signal Process. Lett.*, vol. 7, no. 5, pp. 108–110, 2000.