ICSE 2010

PROCEEDINGS OF THE SECOND INTERNATIONAL CONFERENCE ON SCIENCE AND ENGINEERING

Volume - 1

Electronics
Electrical Power
Information Technology
Engg: Physics

Sedona Hotel, Yangon, Myanmar
December 2-3, 2010

PROCEEDINGS OF THE SECOND INTERNATIONAL CONFERENCE ON SCIENCE AND ENGINEERING

Organized by
Ministry of Science and Technology

DECEMBER 2-3, 2010
SEDONA HOTEL, YANGON, MYANMAR

ELECTRONIC ENGINEERING

A Frequency-selective Noise Spectral Subtraction Approach

Thiri Thandar Aung[1], N. Tangsangiumvisai[2], A. Nishihara[3]

[1,2]*Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Phyathai Road, Pathumwan, Bangkok 10330, THAILAND*
[1]suuthiri@gmail.com, [2]Nisachon.T@chula.ac.th
[3]*The Center for Research and Development of Educational Technology, Tokyo Institute of Technology, Ookayama, Meguro-ku, Tokyo, 152-8552 JAPAN*
[3]aki@cradle.titech.ac.jp

***Abstract* - In this paper, a new approach for frequency-selective noise spectral subtraction is proposed, based on the Magnitude Spectral Subtraction (MSS) technique and the two-channel Generalized Sidelobe Canceller (GSC) for noise spectrum estimation. A comparison of the proposed and the conventional noise spectral subtraction techniques is given, based on computer simulations, in terms of noise attenuation performance and the ability to avoid speech distortion. Simulation results clarify that less amount of speech distortion is obtained whereas higher noise attenuation can be achieved, as compared to the conventional noise spectral subtraction technique.**

***Keywords* - noise reduction, speech enhancement, spectral subtraction, adaptive beamforming, etc.**

I. INTRODUCTION

Several Noise Reduction (NR) techniques have been proposed to eliminate the additive background noise in order to improve speech intelligibility [1]-[5]. Depending on the number of microphones in the system, the NR techniques are then categorized into 2 groups; i.e. single-channel and multi-channel NR techniques. The frequency-domain single-channel NR approaches, such as the Spectral Subtraction (SS) technique [2], are normally employed, as compared to the time-domain ones, due to their low computational complexity. However, by increasing the amount of noise spectral subtraction, the speech distortion is perceived as 'musical noise' [3]. It is suggested in [6] that the Magnitude SS (MSS) technique introduces less amount of musical noise than the Power SS (PSS) one. Thus it is chosen in this paper to use MSS for noise reduction.

By using spatial information, the multi-channel NR techniques have been introduced to obtain better estimate of the noise and interference signals than the single-channel NR techniques [4]. As a result, improved Noise Attenuation (NA) performance is obtained when the more accurate estimate of the noise spectrum is achieved. Recently, the two-channel Generalised Sidelobe Canceller (2chGSC) algorithm is proposed to be employed together with the NR techniques in order to remove the interfering signals [8].

In order to control the amount of speech distortion while achieving sufficient level of noise reduction, it is therefore proposed in this paper to employ the variable and frequency-selective noise subtraction parameter for the MSS technique, which is employed together with the 2chGSC algorithm. The performance of the 2chGSC+MSS algorithm that employed the proposed scheme for variable subtraction is compared under various acoustic environments to that using fix parameter for noise subtraction.

This paper is organized as follows. In section 2, the 2chGSC+MSS algorithm is described. The proposed noise subtraction parameter to be employed in the 2chGSC+MSS algorithm is presented in Section 3. Simulation results are given in Section 4, followed by the conclusions in Section 5.

II. THE TWO-CHANNEL NOISE REDUCTION ALGORITHM

The two-channel NR technique employing Generalized Sidelobe Canceller has been proposed in [7] to estimate the noise spectrum so that noise reduction can be obtained sufficiently. A block diagram of the 2chGSC algorithm is given in Fig. 1.

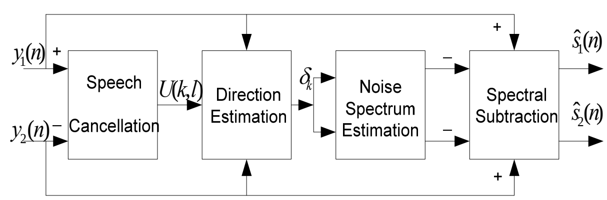

Fig. 1 A block diagram of two-channel NR algorithm

The two noisy speech signals, which are obtained from the first microphone and second microphone, is modeled as a sum of clean speech signal $s(n)$ and the m^{th} interfering noise signals, $b(n)$, as represented by

$$y_i(n)=s(n)+\sum_{m=1}^{M} b_m\big(n-(i-1)\delta_m\big),\ i=1,2 \qquad (1)$$

where i is the number of microphones, M is the number of noise sources and δ_m is the time delay of the *m-th* interfering noise source. The two noisy signals are then transformed into the frequency domain using Short-time Fourier Transform

(STFT). The speech spectrum is cancelled by subtracting the noisy spectrum obtained in the second microphone from that obtained in the first microphone to get the noise-only spectrum.

The analysed spectrum is divided into p uniform subbands. The speech cancelled spectrum in the p^{th} subband is given in terms of the integrated noise spectrum in the p^{th} subband, $N_p(k,l)$, as

$$U(\tilde{k},l) = 2jN_p(\tilde{k},l)e^{-j(\frac{\tilde{k}\delta_p}{2})}sin\left[\frac{\tilde{k}\delta_p}{2}\right] \tag{2}$$

where $k^u{}_{p-1} \le \tilde{k} < k^u{}_p$ is the discrete frequency components between the adjacent p^{th} subband for $p = 1, 2, .., P$ when P is the number of subband and $k^u{}_p$ is the maximum frequency component in each p^{th} subband. The discrete frequency component is $k = 1, 2, ..., K$ when employing K-point STFT.

Assuming that the speech signal and the interfering noise signals are uncorrelated, the Direction of Arrival (DoA) of the virtual noise source, $\hat{\delta}_p$, of the p^{th} subband in the first microphone can be calculated as

$$\hat{\delta}'_p = 2arg_n max\left[IFFT\left[\frac{U(\tilde{k},l)X_1^*(\tilde{k},l)}{|U(\tilde{k},l)||X_1^*(\tilde{k},l)|}\right]\right] \tag{3}$$

Similarly, the same procedure is taken to find $\hat{\delta}_p$ in the second microphone [7].

The integrated noise spectrum in the p^{th} subband is then obtained to be

$$|\hat{N}_{p,i}(\tilde{k},l)| = \begin{cases} U(\tilde{k},l)H_{p,i}(\tilde{k},l), & \left|sin(\tilde{k}\hat{\delta}_p/2)\right| \ge \varepsilon \\ \gamma|\hat{N}_{p,i}^{pre}(\tilde{k},l)|^2 + (1-\gamma)E[|\hat{N}_{p,i}(\tilde{k},l)|^2 \,||X_i(\tilde{k},l)|^2], & otherwise \end{cases} \tag{4}$$

where $0 < \gamma < 1$ is the forgetting factor controlling the rate of noise reduction. The noise compensator is given by

$$H_{p,i}(\tilde{k},l) = \frac{e^{j\left[\frac{\tilde{k}\hat{\delta}_p}{2}\right]}}{2j\,sin\left[\frac{\tilde{k}\hat{\delta}_p}{2}\right]} \tag{5}$$

whereas $\hat{N}_{p,i}^{pre}(\tilde{k},l)$ is the estimation of the noise spectrum in the previous frame. The noise spectrum in each microphone is calculated over the entire frequency region as

$$\tilde{N}_i(k,l) = \sum_{p=1}^{p}\tilde{N}_{p,i}(\tilde{k},l) \qquad i = 1,2 \tag{6}$$

Finally, the enhanced speech spectrum at each microphone is obtained by subtracting the estimated noise spectrum from the noisy speech spectrum in each channel. The combination of the 2chGSC algorithm with the MSS technique to obtain two-channel NR algorithm is demonstrated as shown in Fig. 2.

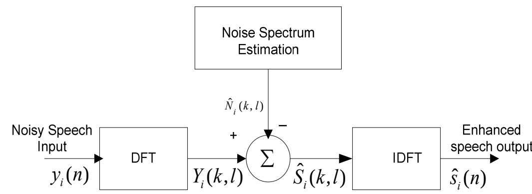

Fig. 2 A block diagram of the SS algorithm

The estimated enhanced speech spectrum, $|\hat{S}_i(k,l)|$, is obtained as given by

$$|\hat{S}_i(k,l)| = \begin{cases} |Y(k,l)| - \alpha|B(k,l)|, & |Y(k,l)| \ge \alpha|B(k,l)| \\ \rho|Y(k,l)|, & otherwise \end{cases} \tag{7}$$

where α is the noise subtraction parameter and ρ is the noise floor parameter. The estimate of $\hat{S}_i(k,l)$ contains both magnitude and phase is shown below as

$$\hat{S}(k,l) = |\hat{S}(k,l)|\, e^{j\theta_y(k,l)} \tag{8}$$

where $\theta_y(k,l)$ is the phase of the noisy spectrum in each microphone. Then $\hat{s}_i(n)$ is obtained by performing Inverse STFT. The enhanced speech signal from the first microphone ($i = 1$) is used in this paper.

III. THE PROPOSED NOISE SPECTRAL SUBTRACTION PARAMETER

In the conventional SS technique, the noise subtraction parameter α is fixed across the whole frequency spectrum. If the parameter α is to be increased for further noise subtraction, large α, however, introduces speech distortion. On the other hand, by reducing the α factor to preserve speech quality, the additive noise source is not sufficiently eliminated.

It is therefore proposed in this paper that the parameter α to be frequency dependent. High value of α is chosen when the speech spectral components are more dominant than the noise spectral components, i.e. in the low-frequency region. In order to preserve speech spectral components, the value of α is chosen, on the other hand, to be very small when the noise spectral level is high, i.e. in the high-frequency region. Therefore the noise subtraction parameter, α, in eq.(7) is chosen to be dependent on the frequency components of the difference between the noise. For each l^{th} analysis frame, the variable noise subtraction parameter is obtained and updated by

$$\alpha(k,l+1) = \mu\alpha(k,l) + \gamma e(k,l) \quad (9)$$

where $0<\mu<1$, $\gamma > 0$, k represents the frequency components. The difference between the noisy signal and the enhanced speech signal as given by

$$e(k,l) = y(k,l) - \hat{s}(k,l) \quad (10)$$

IV. EXPERIMENTS AND RESULTS

The performance of the 2chGSC+MSS algorithm that employed the proposed scheme for variable subtraction was compared under various acoustic environments to that using fix parameter for noise subtraction.

A. Experimental conditions

In this paper, two noise acoustic environments: one-noise-source and three-noise-source conditions were evaluated. To model the effect of noises on both microphones, Additive White Gaussian Noise (AWGN) were employed in the experiments. Various input SNR levels were investigated (5-25dB).

The length of the signal was 25ms and it was sampled at 8kHz. The frame shift was 12.5ms and the window function was Hamming. The analyzed spectrum was divided into 100 subbands (P=100). By employing the sampling rate of 8kHz, the bandwidth of each subband was therefore equal to 40Hz.

B. Objective Evaluation

The output signal-to-noise ratio (SNR) was used to measure the noise attenuation performance. It is defined as

$$OutputSNR(dB) = 10 \times log_{10}\left\{\frac{\sum_{n=0}^{N} s^2(n)}{\sum_{n=0}^{N}(s(n)-\hat{s}(n))^2}\right\} \quad (11)$$

where N is the number of samples. When the output SNR is greater than input SNR, the NR algorithm gives improved noise attenuation performance. On the other hand, there will still be some interfering noise on the enhanced speech signal when the output SNR is less than input SNR.

The other measurement was Log-spectral Distance (LSD) which is the difference between the log spectrum of clean speech signal and that of the noisy signal. LSD is used to measure the distortion of the desired speech, given by

$$LSD(dB) = \frac{10}{j} \times \sum_{l=0}^{J-1}\left\{log_{10}\left(\left|\hat{S}(k,l)\right|^2\right) - log_{10}(|S(k,l)|^2)\right\} \quad (12)$$

where J is the number of speech activity frames and $S(k,l)$ and $\hat{S}(k,l)$ are the clean speech spectrum and the enhanced speech spectrum for the l^{th} analysis frame. A low LSD level indicates the low level of speech distortion.

C. Simulation results

Experimental results in two acoustic environments are demonstrated in this subsection. By considering at the one-noise-source environment, it is shown in Fig3 that the NA performance of the 2chGSC+MSS technique with the proposed α is better than the one with fixed α. In addition, from Fig. 4, the former one has lower LSD level than the latter one. Therefore, 2chGSC+MSS with proposed α introduces low level of speech distortion than the one with fixed α.

The NA performance of these investigated algorithms in the three-noise-source case can also confirm with the one-noise-source case, i.e. the NR technique with proposed α gives higher output SNR than the one with fixed α as shown in Fig. 5. It is interesting to notice that the output SNR for both acoustic environments when using the proposed α are the same for all cases. For speech distortion, it can be shown in Fig. 6 that the proposed α yields the lower amount of speech distortion, as compared to the fixed α.

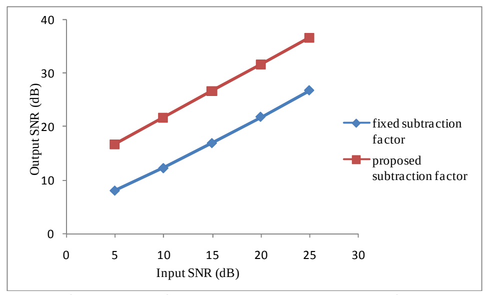

Fig. 3 NA performance of the 2chGSC+MSS with fixed and proposed subtraction factor (α) for one-noise-source environment

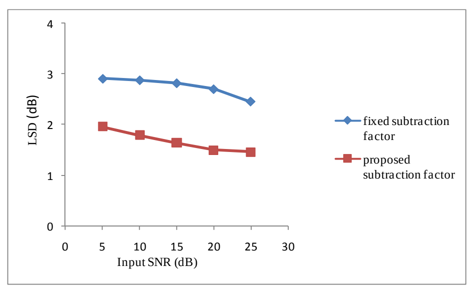

Fig. 4 LSD improvement of the 2chGSC+MSS with fixed and proposed subtraction factor (α) for one-noise-source environment

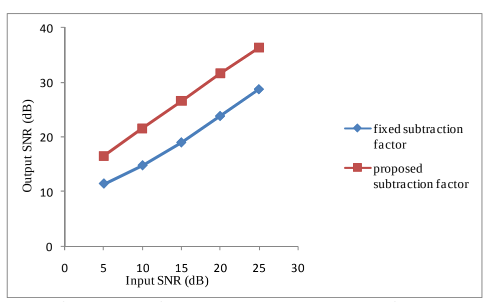

Fig. 5 NA performance of the 2chGSC+MSS with fixed and proposed subtraction factor (α) for three-noise-source environment

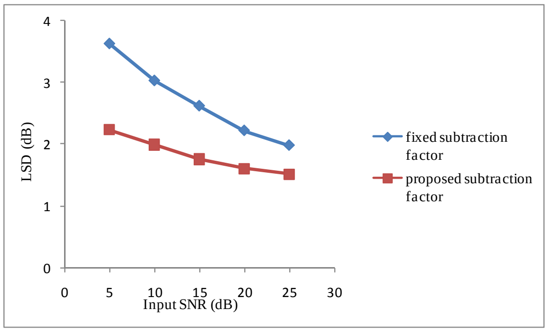

Fig. 6 LSD improvement of the 2chGSC+MSS with fixed and proposed subtraction factor (α) for three-noise-source environment

V. CONCLUSIONS

This paper has proposed a frequency-dependent parameter for noise subtraction to be used in a two-channel Noise Reduction (NR) technique that employs the Generalised Sidelobe Canceller (GSC) and the Magnitude Spectral Subtraction (MSS) method. It has been demonstrated based on computer simulations that the proposed scheme yields better Noise Attenuation performance than the NR technique with fixed subtraction parameter. Furthermore, it has been shown via the LSD plots that lower level of speech distortion is obtained by employing the proposed subtraction parameter, as compared to the conventional one. Hence, the proposed frequency dependent noise subtraction parameter introduces low amount of speech distortion while achieving sufficient level of noise reduction.

ACKNOWLEDGEMENT

This work has been supported by ASEAN University Network/ Southeast Asia Engineering Education Development Network (AUN/SEED-Net).

REFERENCES

[1] J. Li, M. Akagi and Y. Suzuki, "A two-microphone noise reduction method in highly non-stationary multiple-noise-source environments," IEICE Trans. Fundamentals, Vol. E91-A, No. 6, pp. 1337-1346, June 2008.

[2] S.M. McOlash, R.J. Niederjohn, and J.A. Heinen, "A Spectral Subtraction Method for the Enhancement of Speech Corrupted by Non-White, Non-Stationary Noise," Proc. IEEE, Industrial Electronics, Control, and Instrumentation, Vol. 2, No. 20, pp. 872 – 877, November 1995.

[3] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," IEEE Trans. Audio, Speech, Lang. Process., Vol. 14, No. 4, pp. 1218–1234, July 2006.

[4] J. Chen, J. Benesty and Y. Huang, "A Minimum Distortion Noise Reduction Algorithm With Multiple Microphones," IEEE Trans. Audio, Speech, Language Processing, Vol.16, No. 3, pp. 481 – 493, March 2008.

[5] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," IEEE Trans. Speech Audio Processing, Vol. 3, No. 4, pp. 251–266, July 1995.

[6] T. Inoue, Y. Takahashi, H. Saruwatari, K. Shikano, K. Kondo, "Theoretical Analysis of Musical Noise in Generalized Spectral Subtraction: Why should not use Power/Amplitude Subtraction?", Proc. EUSIPCO2010, pp. 994-998, August 2010.

[7] J.Li, M.Akagi and Y.Suzuki, "Extension of the Two-Microphone Noise Reduction Method for Binaural Hearing Aids," Proc. Audio, Language and Image Processing, pp.97-101, July 2008.