Improved Leakage Signal Blocking Methods for Two Channel Generalized Sidelobe Canceller

Kihyeon Kim* · Hanseok Ko*

ABSTRACT

The two-channel Generalized Sidelobe Canceller (GSC) scheme suffers from the presence of leakage signal in the reference channel. The leakage signal is caused by the dissimilar impulse responses between microphones, and different paths from speech source to microphones. Such leakage is detrimental to speech enhancement of the GSC since the desired reference signal becomes corrupted. In order to suppress the signal leakage, two matrix injection methods are proposed. In the first method, a simple gain compensation matrix is used. In the second, a projection matrix for reducing the error between the actual and the ideal primary and reference signals, is used. This paper describes the performance degradation resulting from leakage, and proposes effective methods to resolve the problem. Representative experiments were conducted to demonstrate the effectiveness of the proposed methods on recorded speech and noise in an actual automobile environment.

Keywords: Two channel GSC, Leakage signal, Signal blocking method, Gain ratio compensation matrix, Projection matrix on ideal signal subspace

1. Introduction

Environmental noise stands out as the main driver in performance degradation of automatic speech recognition systems. In order to address that problem, many research efforts have been mobilized in the past few decades. Particularly for speech enhancement objectives, the single channel noise cancellation method has surfaced as a prominent method, due to its advantages of small computational load and simple realization [1][2]. However, it requires exact estimation of noise components. That is, single channel algorithms can be

* Dept. Of Electronics & Computer Engineering, Korea University

effective in terms of stationary noise, only when sufficient information regarding noise is provided.

Therefore a considerable number of studies have been conducted on multi-channel speech enhancement, which uses various speech and noise signal information from several microphones. The GSC is the most popular approach among these algorithms [3]. The GSC cancels spatial sidelobes, except for the desired direction, and enhances speech signal by weighting input noise components adaptively. It demonstrates good performance even when using a small number of microphones [4].

In the GSC, the adaptive filter only requires input noise components. For this, the blocking matrix blocks the input speech components and provides noise components only to the adaptive filter. However, the blocking matrix of the classical GSC always includes speech leakage in its output, in the actual environment [5].

In this paper, two improved signal-blocking methods are proposed to solve the problem, and the effective implementation of GSC through dual channel based experimental results, is investigated. In Section 2, the basic concept and problems of classical GSC are considered. Section 3 suggests two improved signal-blocking methods of GSC in the time domain. The remaining sections present the experimental results and conclusion.

2. Classical 2-channel GSC

Figure 1 describes the scheme of 2-channel GSC in the time domain [4].

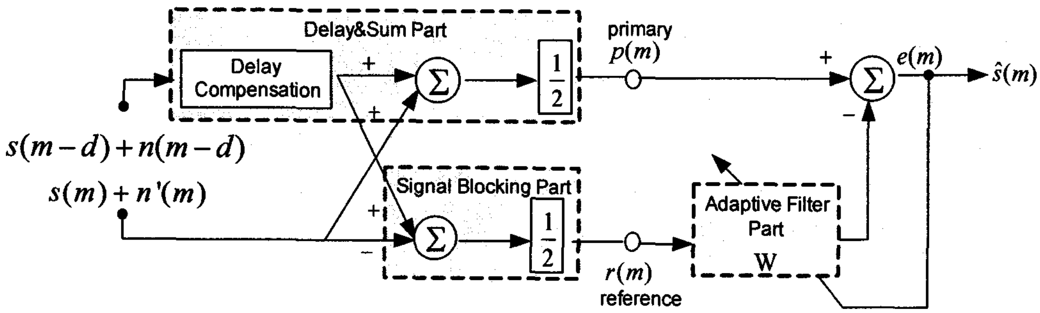

Figure 1. Two channel GSC in the time domain

The GSC is composed of three main components: the delay & sum component, the signal blocking component and the adaptive filter component. Assuming that the time delay, d is compensated, the following expression maps the relationship between the primary reference signal and input.

$$\begin{bmatrix} p(m) \\ r(m) \end{bmatrix} = X^T \begin{bmatrix} ch1(m) \\ ch2(m) \end{bmatrix} = X^T A^T \begin{bmatrix} s(m) \\ n(m) \\ n'(m) \end{bmatrix}; \tag{1}$$

$$X^T = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}, \; A^T = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Notations m, $s(m)$ and $n(m)$, $n'(m)$ are the time index, speech and noise signals, respectively. $ch1$, $ch2$ are the time delay compensated input signals of channel 1 and channel 2, respectively. p, r are the primary and reference signal, respectively.

The delay and sum component synchronizes and sums two channel inputs. This induces the primary signal to have speech and an average of two input noise components, n, n' of input. The signal-blocking component inhibits the speech component in the reference, by subtracting one input from the other. Eventually, it only provides the difference between two noise components to the adaptive filter. The adaptive filter component estimates the noise component of the primary signal by weighting the reference signal adaptively. In the final step, this estimated noise is subtracted from the primary signal, to obtain enhanced speech. The adaptive filter uses the Least Mean Square (LMS) algorithm to estimate filter weights, as shown in the following equation [3][6].

$$\hat{\mathrm{W}} = \arg\min_{\mathrm{W}} E[e^2(m)] \tag{2}$$

, where e is the error signal between the primary and output of the adaptive filter, W is its weights array.

This criterion has two assumptions. The first is that the speech and the noise component of the input are not correlated. The second is that two input noise components demonstrate correlation. Based on these assumptions, the equation can be changed as follows.

$$\begin{aligned} \hat{\mathrm{W}} &= \arg\min_{\mathrm{W}} E[(s + n_p - \mathrm{W} * n_r)^2] \\ &= \arg\min_{\mathrm{W}} E[(n_p - \mathrm{W} * n_r)^2] \end{aligned} \tag{3}$$

, where n_p, n_r are the noise component of the primary and the reference signal, respectively.

3. Leakage signal blocking methods

The GSC demonstrates good performance under only two conditions. First, the input of the adaptive filter does not include the speech component. The GSC outputs the distorted speech with much echo, by subtracting the estimate of the speech component from the primary signal, unless the input of the adaptive filter only has a noise component. Secondly, noise components of the primary and the reference signal demonstrate correlation. Non-correlative characteristic of these signals forces LMS to only have a 0's array as its solution. That is, the adaptive filter never plays the role of reducing the noise component of the primary signal.

However, in an actual environment, the classical GSC suffers from the leakage signal caused by non-identical impulse responses between microphones and different paths from speech source to microphones. It is assumed that the channel input is the following equation:

$$\begin{bmatrix} ch1(m) \\ ch2(m) \end{bmatrix} = A^T \begin{bmatrix} s(m) \\ n(m) \\ n'(m) \end{bmatrix}; \quad A^T = \begin{bmatrix} 1 & 1 & 0 \\ \alpha(i) & 0 & 1 \end{bmatrix}, \tag{4}$$

the reference signal has the leakage as equation (5). (Block " " means the leakage, the coefficient of the speech component in the reference signal.)

$$\begin{bmatrix} p(m) \\ r(m) \end{bmatrix} = B^T \begin{bmatrix} s(m) \\ n(m) \\ n'(m) \end{bmatrix}; B^T = X^T A^T,$$
$$B^T = \begin{bmatrix} (1+\alpha(i))/2 & 1/2 & 1/2 \\ "(1-\alpha(i))/2" & 1/2 & -1/2 \end{bmatrix} \tag{5}$$

, where i is the frame index grouping m, α is the gain ratio of the speech component of channel 2 to that of channel 1. It makes the adaptive filter have the leakage and output the signal partially adapted to the speech. This causes distortion of speech and degrades SNR improvement of the GSC.

In order to solve the problem, this paper proposes two transformation matrices, focusing on the improvement of signal blocking in the time domain.

3.1 Gain ratio compensation matrix

The gain ratio α is caused by the difference between the amplitude of the speech component in channel 1 and that in channel 2. It can be compensated by multiplying the matrix C by X^T, producing p, r signals.

$$D = \begin{bmatrix} 1/2 & 1/2\alpha \\ 1/2 & -1/2\alpha \end{bmatrix}; D = X^T C, \qquad (6)$$
$$C = \begin{bmatrix} 1 & 0 \\ 0 & 1/\alpha \end{bmatrix}$$

The matrix D is a gain compensation matrix and produces p, r signals instead of the matrix X^T. D matrix makes the $B^T(2,1)$ component be 0 and eliminates the leakage of the reference signal as equation (7), if estimation of α is correct.

$$\begin{bmatrix} p \\ r \end{bmatrix} = D \begin{bmatrix} ch1 \\ ch2 \end{bmatrix} = B^T \begin{bmatrix} s \\ n \\ n' \end{bmatrix}; \qquad (7)$$
$$B^T = DA^T = \begin{bmatrix} 1 & 1/2 & 1/2\alpha \\ "0" & 1/2 & -1/2\alpha \end{bmatrix}$$

3.2 Projection matrix on ideal signal subspace

In equation (1), assuming that A^T is of the equation (4), which considers a real environment, and B^T is the subspace to produce ideal p, r signals, the new matrix $\hat{X}^T$ can be described by the following equation.

$$\begin{bmatrix} p \\ r \end{bmatrix} = \hat{X}^T \begin{bmatrix} ch1 \\ ch2 \end{bmatrix} = \hat{X}^T A^T \begin{bmatrix} s \\ n \\ n' \end{bmatrix} = B^T \begin{bmatrix} s \\ n \\ n' \end{bmatrix}; \qquad (8)$$
$$A^T = \begin{bmatrix} 1 & 1 & 0 \\ \alpha & 0 & 1 \end{bmatrix}, B^T = \begin{bmatrix} 1 & 1/2 & 1/2 \\ 0 & 1/2 & -1/2 \end{bmatrix}$$

Unless the speech signal and noises are 0, the over-determined equation (9) can be established from the equation (8).

$$A\hat{X} = B \qquad (9)$$

and the solution $\hat{X}$ is obtained by the projection matrix P_{proj}.

$$\hat{X} = P_{proj}B; \quad P_{proj} = (A^T A)^{-1} A^T \tag{10}$$

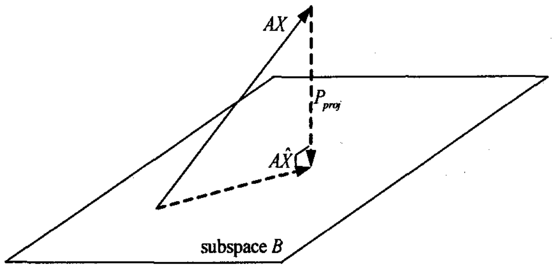

Figure 2. Projection on the ideal primary and reference signal subspace

Compared with that of the previous gain ratio, the compensation matrix purely focuses on minimization of the leakage, the matrix $\hat{X}^T$ maintains the correlative characteristic between noise components of p, r signals, while minimizing the leakage by producing them close to the ideal case. The proposed matrix $\hat{X}^T$ is called the projection matrix P, on the ideal signal subspace. This reduces the $B^T(2,1)$ component by multiplying $1/(\alpha^2+2)$ if the estimated α is correct, as in equation (11). That is, the leakage is attenuated below half and p, r signals become close to the theoretical ideal signals.

$$\begin{gathered}\begin{bmatrix} p(m) \\ r(m) \end{bmatrix} = P\begin{bmatrix} ch1(m) \\ ch2(m) \end{bmatrix} = B^T\begin{bmatrix} s(m) \\ n(m) \\ n'(m) \end{bmatrix}; \; P = \hat{X}^T, \\ B^T = PA^T = \begin{bmatrix} \frac{2\alpha^2+\alpha+3}{2(\alpha^2+2)} & \frac{\alpha^2-\alpha+3}{2(\alpha^2+2)} & \frac{\alpha+2}{2(\alpha^2+2)} \\ \frac{1-\alpha}{2}\frac{1}{(\alpha^2+2)} & \frac{\alpha^2-\alpha+3}{2(\alpha^2+2)} & \frac{-(\alpha+2)}{2(\alpha^2+2)} \end{bmatrix},\end{gathered} \tag{11}$$

4. Experiments

In the experiments, we use actual recorded speech and noise data of 11 KHz, 16 bit PCM. Two microphones are located at a distance of 15 cm, in order to record the datum. Input data has various SNR ratios by synthesizing clean speech and car noise signals. In order to estimate α, the ratio of Root Mean Square (RMS) of channel 1 to channel 2, signal is applied at every frame using 10 ms (110 samples). The adaptive filter uses the

Normalized Least Mean Square (NLMS) algorithm [6].

In Table 1, the experiment presents levels of the leakage in the reference signal according to kinds of matrix producing p, r signal.

Table 1. Levels of the leakage according to transformation matrices

Transformation matrix producing p, r signals	X^T	P	D
Level of leakage (RMS)	0.0082	0.0080	0.0078

Levels of the leakage in table 1 are RMS values of the reference signal produced only by the clean speech input. In both cases, using the gain ratio compensation matrix D and the projection matrix P on the ideal signal subspace, the level of the leakage is not diminished up to the expectation compared with the original matrix X^T. This means that the use of the RMS ratio is not sufficient to estimate α correctly. However, it is important to note that the leakage is actually decreased by the simple estimation of α and the proposed matrices.

As mentioned previously, the correlative characteristic between noise components of the primary and the reference signal is an important condition when estimating weights of the adaptive filter. It determines the performance of noise reduction of the GSC. Figure 3 presents the coherence between noise components of p, r according to transformation matrices.

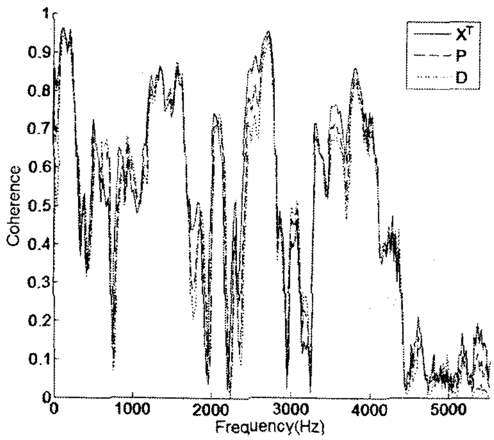

Figure 3. Coherences between noise components of the primary and reference signal according to transformation matrices. (SNR of input : 0 dB)

Figure 3 demonstrates that the coherence between noise components of p, r decreases in the order of X^T , P , D . The original transformation matrix X^T preserves the correlative characteristic between noise components of p, r sufficiently, because it simply plays the role of sum and subtraction of the input. The coherence of P is greater than D . The matrix P can keep coherence, by reducing the error between the actual and the ideal primary and reference signals. However, the matrix D focuses on elimination of the leakage without consideration of noise distortion.

The following figure presents energy envelopes of the GSC output when the GSC actual car noise is its input.

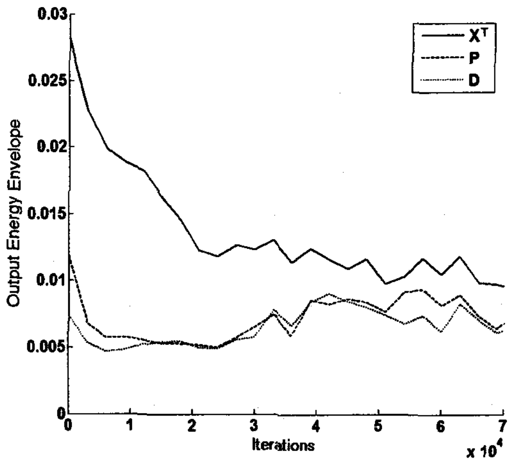

Figure 4. Energy envelopes of the GSC output when the GSC has the only noise as its input.

In Figure 4, the proposed algorithms demonstrate improved performance in terms of noise reduction, by rapidly adapting to the input noise. The gain ratio compensation matrix has the advantage of minimization of leakage and the projection matrix on the ideal subspace has the merits of coherence preservation, while decreasing leakage current.

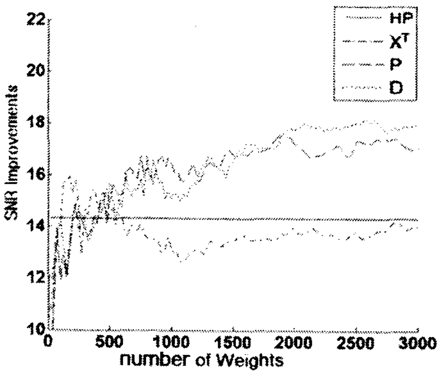

Figure 5. SNR improvements according to the number of weights of the adaptive filter (0 dB input SNR).

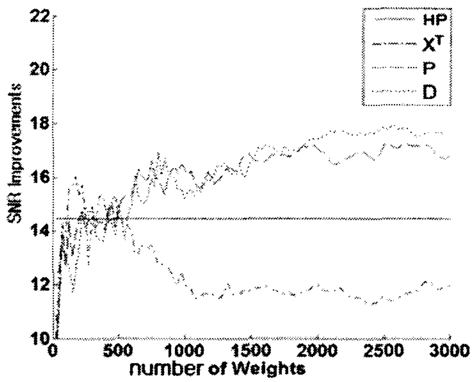

Figure 6. SNR improvements according to the number of weights of the adaptive filter (10 dB input SNR).

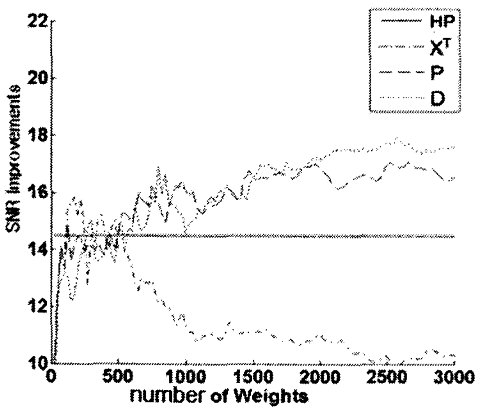

Figure 7. SNR improvements according to the number of weights of the adaptive filter (20 dB input SNR).

The number of weights of the adaptive filter is another important factor to determine the performance of the GSC. Generally, numerous weights help exact adaptation to the noise. However, it requires high computational complexity to estimate these weights. The above figures present SNR improvements, according to the number of weights of the adaptive filter and the input SNR ratio. In figures 5, 6, 7, "HP" represents 1 channel high-pass filter, and is compared with the 2 channel GSC. The line shape of P is similar to that of D in all cases. The proposed algorithms have curves steady at the larger of the weights. 1500 is the approximate optimal number of weights, which are also reflected when stationary. However, in 0~500, X^T demonstrates the best performance and P achieves this in 500~1500, D achieves this over 1500. This result can be interpreted as coherence, the main factor in determining GSC performance at a small number and the leakage at the large number. That is, the lower the number of weights, the greater coherence effects performance, and the greater the number of weights, the correlative characteristic between the reference and the speech component of the primary signal becomes more defined. In the case of X^T, the performance degrades towards higher SNR inputs, even though the system has many weights. This is due to high leakage levels relative to the reference noise.

Finally, the proposed algorithms provide improved results, by combining a simple high-pass filter into the GSC, as shown in the following figure.

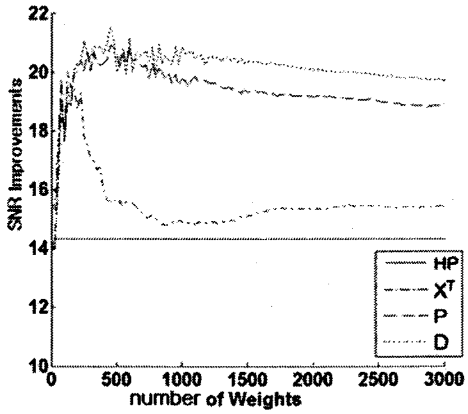

Figure 8. SNR improvements of the proposed algorithms combined with a high-pass filter (0 dB SNR input).

In other cases of input SNR, the curve shapes are similar to that shown in figure 8. The high-pass filter has the threshold frequency of 270 Hz for actual car noises. D demonstrates the best performance, regardless of the number of weights. However, the performance degrades when the number of applied weights is over 1000. This can be interpreted as the best output from the GSC is distorted by the high-pass filter.

5. Conclusion

This paper proposes improved leakage signal blocking methods, and confirms the effectiveness in a two channel GSC. Consequently, the gain ratio compensation matrix D is useful for minimizing the leakage and the projection matrix P on the ideal signal subspace, and keeps the coherence between noise components of p, r signals at a certain level, regardless of the fact that the degree of minimizing leakage is lower than D. The results demonstrate that the optimal algorithm can be selected according to the number of weights. The suggested algorithms demonstrate superior performance by combining a simple high-pass filter.

Acknowledgement

This work was supported by grant No. A17-11-02 from the Korea Institute of Industrial Technology Evaluation & Planning Foundation.

References

[1] Ephraim, Y. 1992. "Statistical-model-based speech enhancement systems," Proc. Of IEEE, Vol.80, No.10, pp. 1526-1555.

[2] Martin, R. 2001. "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Trans. on Speech and Audio processing, Vol.9, No.5, pp. 504-512.

[3] Griffiths, L. J. & Jim, C. W. 1982. "An alternative approach to linearly constrained adaptive beam-forming," IEEE Trans. on Antennas and Propagat., Vol.30, No.1, pp. 27-34.

[4] Compernolle, D. V., Ma, W., Xie, F. & Diest, M. V. 1990. "Speech recognition in noisy environments with the aid of microphone arrays," Speech Comm., Vol.39, No.1-2, pp. 433-442.

[5] Claesson, I. & Nordholm, S. 1992. "A Spatial filtering approach to robust adaptive beamforming," IEEE Trans. on Antennas and Propagat. Vol.40, No.9, pp. 1093-1096.

[6] Simon, H. 2002. *Adaptive Filter Theory*, New Jersey, Prentice Hall, Fourth Edition.

received: January 30, 2006
accepted: March 12, 2006

▲ Kihyeon Kim
Dept. of Electronics Engineering, Korea University,
5Ka-1, Anam-dong, Seongbuk-ku, Seoul, 136-713, Korea
Tel: +82-2-926-2909 Fax: +82-2-3291-2450
H/P : 011-9889-3822
Email: khkim@ispl.korea.ac.kr

▲ Hanseok Ko
Dept. of Electronics Engineering, Korea University,
5Ka-1, Anam-dong, Seongbuk-ku, Seoul, 136-713, Korea
Tel: +82-2-3290-3239 Fax: +82-2-3291-2450
H/P: 011-9001-3239
Email: hsko@korea.ac.kr