THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA Vol.27, No.1E 2008. 3. pp. 17~24

Application of Block On-Line Blind Source Separation to Acoustic Echo Cancellation

Duong Q.K. Ngoc*, Chul Park*, Seung-Hyon Nam*
*Department of Electronic Engineering, Paichai University
(Received January 4, 2008; accepted February 27, 2008)

Abstract

Blind speech separation (BSS) is well-known as a powerful technique for speech enhancement in many real world environments. In this paper, we propose a new application of BSS - acoustic echo cancellation (AEC) in a car environment. For this purpose, we develop a block-online BSS algorithm which provides robust separation than a batch version in changing environments with moving speakers. Simulation results using real world recordings show that the block-online BSS algorithm is very robust to speaker movement. When combined with AEC, simulation results using real audio recording in a car confirm the expectation that BSS improves double talk detection and echo suppression.

Keywords: *blind source separation; acoustic echo cancellation; double talk detection; echo return loss enhancement.*

I. Introduction

Acoustic echo cancellation (AEC) is an essential part of a hands-free communication system to eliminate acoustic echo picked up by a microphone. AEC is normally employed in a car where a driver can make a hands-free communication with a far-end talker while driving. We assume that a voice from the far-end side is played back by many speakers (normally more than four) installed in a car. In a car, there exist strong consistent background noise and the loudspeaker-enclosure microphone (LEM) environment is constantly changing. All these make AEC operation difficult. Furthermore, during double-talk period, the adaptive filter of AEC is not adapted to maintain stability of the AEC and echo may not be well-suppressed. Moreover, hands-free communication may suffer from interfering signals such as background noise and voices of other passengers which cannot be removed in the AEC.

In order to solve these difficulties, in this paper, we propose to combine blind source separation (BSS) with AEC to separate the echo and local signals. The conceptual block diagram of the AEC combined with BSS is shown in Fig. 1. Block-online BSS works on two microphone signals and separate input signals into local driver's voice and the interfering signals. The interfering signal may be far-end echo or local interfering signals. The output channel that contains the local driver's voice is then fed into the AEC. The residual echo remains in the separated output is then removed further in the AEC.

Blind source separation is a technique to separate original signals from a set of mixtures without any information on the originals signals or a mixing system except that original signals are statistically independent each other. One practical method for convolutive BSS is multichannel blind deconvolution (MBD). MBD is a time-domain technique that does not suffer from channel permutation since each separated

Corresponding author: Seung-Hyon Nam (shnam@pcu.ac.kr)
439-6 Doma-dong, Seo-ku, Daejeon 302-735, Korea
Dept. of Electronic Engineering, Paichai University

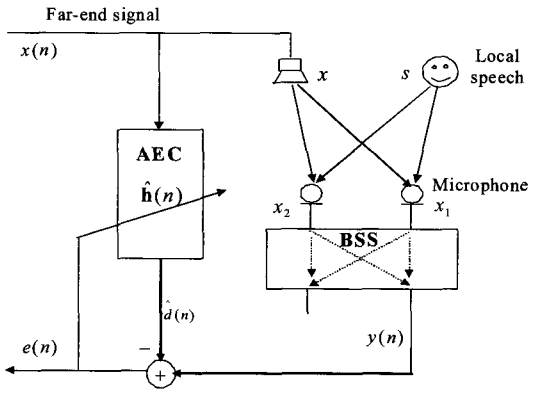

Fig. 1. Block diagram of AEC combines with BSS.

signal appears at the channel closest to each original source. However, many existing MBD algorithms converge slowly and separation performance is not good enough. Furthermore, the separated outputs are often whitened. Recently, MBD with spectral normalization, which provides fast and improved separation while relieving the whitening problem significantly, has been proposed [1-3].

II. Block Online BSS

The BSS algorithm chosen in this work is a batch off-line version. The detail description of the algorithm is given in [2,3]. It can be implemented in the frequency domain using FFT for reduced computation.

Let X and u be mixed input and separated (or unmixed) output vectors, respectively. In addition, define y=f(u), where f is the time-domain element -wise nonlinear function [2]. Then the update rule for the unmixing filter matrix W can be described by the following simplified conceptual expression in the frequency domain.

$$\Delta \mathbf{W}^f(b) = \left\{\mathbf{I} - \mathbf{\Lambda}_y^{-1}(b)\mathbf{y}^f(b)\left(\mathbf{u}^f(b)\right)^H \mathbf{\Lambda}_u^{-1}(b)\right\}\mathbf{W}^f(b) \quad (1)$$

Here, b is the block index, f denotes the frequency quantity, $\Lambda_Y(b)$ and $\Lambda_U(b)$ are diagonal matrices that have $\sqrt{\mathbf{P}_{y_i}(b)}$ and $\sqrt{\mathbf{P}_{u_i}(b)}$ as diagonal elements, respectively. Further, I denotes the identity matrix. It is important to notice that the unmixing filters used here are right-sided. Refer to [3] for some interesting and important issues on filter shape.

The spectral power is computed by the following update rule.

$$\begin{aligned}\mathbf{P}_{y_i}(b) &= (1-\gamma)\mathbf{P}_{y_i}(b-1) + \gamma\left|\mathbf{y}_i^f(b)\right|^2 \\ \mathbf{P}_{u_i}(b) &= (1-\gamma)\mathbf{P}_{u_i}(b-1) + \gamma\left|\mathbf{u}_i^f(b)\right|^2\end{aligned} \quad (2)$$

The practical computation includes linear correlations /convolutions via circular convolutions/correlations in the frequency domain, respectively. Thus, forward /inverse DFT and proper aliasing eliminations are required. The algorithm can be implemented either in the overlap-save or overlap-add methods.

As described in [1], the fast stable convergence of this BSS algorithm comes from the spectral normalization that equalizes the spectral tilt. The algorithm can accommodate a large step size without causing instability so that it can separate the mixed signals successfully within few iterations of a given short data segment. Fast convergence speed is very important for AEC application.

Since BSS should operate in continuous real-time for AEC application, we implement the above BSS algorithm in block-online fashion based on the existing procedures described in [4,5]. The block -online mode of operation is known to provide very robust separation performance in changing environments. Input mixed signals are divided into online and offline blocks. One online block is made of a number of offline blocks (4 blocks in this case). These offline blocks are processed iteratively using the batch BSS described by the update rules (1) and (2). At each offline block m and at each iteration j, the gradient from the offline update is computed. The offline updates are then averaged over K offline block at iteration j.

$$\Delta\tilde{\mathbf{w}}_{pq}^{j-1}(m) = \frac{1}{K}\sum_{n=0}^{L-1}\Delta\tilde{\mathbf{w}}_{pq}^{j-1}(m) \quad (3)$$

where L is the length of the unmixing filter. After each iteration of K off-line blocks, the averaged offline gradients are then added together to get online gradient.

$$\Delta \mathbf{w}_{pq}^{j}(m') = \sum_{n=0}^{L-1} \Delta \tilde{\mathbf{w}}_{pq}^{j-1}(m) \tag{4}$$

where m' is the online index. After $j_{\max}$ iterative learning of K off-line blocks, offline gradient is updated as follows:

$$\tilde{\mathbf{w}}_{pq}^{j}(m) = \tilde{\mathbf{w}}_{pq}^{j-1}(m-1) + \mu_{off} \Delta \mathbf{w}_{pq}^{j-1}(m) \tag{5}$$

In addition, the result of the off-line update is fed into online update

$$\mathbf{w}_{pq}(m') = \lambda \mathbf{w}_{pq}(m'-1) + (1-\lambda) \Delta \tilde{\mathbf{w}}_{pq}^{j_{\max}}(m') \tag{6}$$

where λ controls the time constant of the unmixing filters. Finally, the updated unmixing filter is used as an initial value for the next online update. The block-online BSS algorithm is summarized in Table 1. The main advantage of the block-online algorithm is tracking capability in a changing environment.

III. Frequency-Domain Block AEC

A basic acoustic echo canceller consists of the blocks of band-pass filter, de-correlation filter, voice activity detection, double talk detection, normalized least mean square (NLMS) adaptive filter, non-linear processor, and optional automatic gain control [6-8]. The conceptual block diagram illustrating the operation of the AEC algorithm is shown in Fig. 2. Descriptions on major blocks are given below.

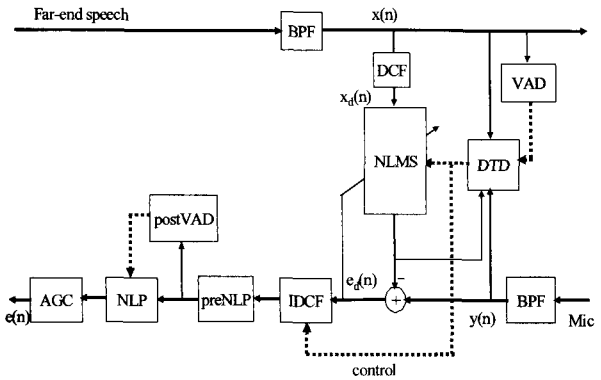

Fig. 2. Block diagram of a basic AEC implementation.

3.1. Decorrelation and inverse decorrelation filters

It is known that a fixed first-order decorrelation filter provides remarkable result [8,10]. The first-order filter can be a simple high-pass filter since speech signals have low-pass characteristics. The formula to decorrelate far-end and microphone signals for NLMS tap-weight update are:

$$x_d(n) = x(n) - ax(n-1) \tag{7}$$

$$y_d(n) = y(n) - ay(n-1) \tag{8}$$

where $0 < a < 1$. We set a=0.85.

Two decorrelation filters are necessary to de-correlate two signals that are fed into the NLMS adaptive filter. In addition, an inverse decorrelation filter is necessary to recover true error signal as $e(n) = e_d(n) + ae_d(n-1)$. These blocks are shown in Fig. 2.

3.2. Voice activity detection

A short-term magnitude estimation of each block of input signal is computed as [8]

$$\overline{|x(n)|} = (1-\gamma(n))|x(n)| + \gamma(n)\overline{|x(n-1)|} \tag{9}$$

where

$$\gamma(n) = \begin{cases} \gamma_r, & \text{if } |x(n)| > \overline{|x(n-1)|} \\ \gamma_f, & \text{ortherwise.} \end{cases}$$

For 8 kHz sampling rate, we choose γ_r=0.95 and $\gamma_f = 0.97$. Voice activity of a far-end signal is

Table 1. Implementation of the Block-online BSS algorithm [5].

On-line Part:	
1.	Collect off-line audio samples Let N denote the frame size, $M(=N\times R)$ the block size in a frame. The collected signal is given by, $\mathbf{x}_q(m')=\left[\mathbf{x}_q(m'KM),\cdots,\mathbf{x}_q(m'KM+KM-1)\right]^T$ where R is the overlap ratio between frames, $0<R\le 1$.
2.	Form K off-line blocks. The off-line block index is $m=m'K,\cdots,m'K+K-1$.
Off-line Part:	
	$j=1,\cdots,j_{\max}$
	$m=m'K,\cdots,m'K+K-1$
3.	Perform batch BSS $\mathbf{u}_p(m)=\tilde{\mathbf{w}}_{pq}^{j}(m)*\mathbf{x}_q(m)$
4.	Compute the average $\Delta\tilde{\mathbf{w}}_{pq}$. $\Delta\tilde{\mathbf{w}}_{pq}^{j-1}(m)=\frac{1}{K}\sum_{n=0}^{L-1}\Delta\tilde{\mathbf{w}}_{pq}^{j-1}(m)$ $\Delta\mathbf{w}_{pq}^{j}(m')=\sum_{n=0}^{L-1}\Delta\tilde{\mathbf{w}}_{pq}^{j-1}(m)$
5.	Perform off-line update. $\tilde{\mathbf{w}}_{pq}^{j}(m)=\tilde{\mathbf{w}}_{pq}^{j-1}(m-1)+\mu_{off}\Delta\mathbf{w}_{pq}^{j-1}(m)$
On-line Part:	
6.	Perform online update. $\mathbf{w}_{pq}(m')=\lambda\mathbf{w}_{pq}(m'-1)+(1-\lambda)\Delta\tilde{\mathbf{w}}_{pq}^{j_{\max}}(m')$
7.	Use online filter as an initial off-line filter and repeat step 2-6. $\tilde{\mathbf{w}}_{pq}^{0}(m'+1)=\mathbf{w}_{pq}(m')$

detected if the short-term magnitude exceeds a predefined noise threshold of 35 dB.

3.3. Double talk detection

Double talk detection is a typical problem that is deeply concerned in AEC implementation to detect the situation in which both sides talk simultaneously. During the double talk period, the residual error increases due to local speech so that the AEC stability bound decreases and the algorithm may start to diverge. This situation must be prevented and the adaptive filter coefficients must be frozen during the double talk. Two typical methods to detect the double talk are correlation-based and energy-based methods. The first one uses degree of similarity between microphone signal $y(n)$ and the output signal $\hat{d}(n)$ of adaptive filter or the loudspeaker signal $x(n)$ in term of correlation [6,8] whereas the second method compares the microphone signal level with the speaker signal level.

3.4. Frequency-domain block NLMS adaptive filter

The frequency domain block NLMS filter provides a possible solution to the computational structure and complexity problem. The algorithm consists of the following main steps [9].

The most recently acquired block of $N/2$ input samples and the previous block make up the new input block $\mathbf{X}_n$ and this new block was transformed into frequency domain by the discrete Fourier transform: $\mathbf{X}_n=\mathbf{F}\{\mathbf{x}_n\}$. The most recent M frequency domain input blocks are retained as $\mathbf{X}_n=[\mathbf{X}_n\mathbf{X}_{n-1}\cdots\mathbf{X}_{n-M+1}]$.

The adaptive filter uses low-latency structure where the filter response is evenly divided into non-overlapping segments and filter output can be calculated in time-domain as

$$\hat{y}(n)=\sum_{j=0}^{M*N/2-1}h(j)x(n-j)=\sum_{i=0}^{M-1}\sum_{k=0}^{N/2-1}h_i(k)x_j(n-k) \tag{10}$$

which can be computed in the frequency domain using the overlap-save or overlap-add method as

$$\hat{\mathbf{Y}}_n(k)=\sum_{i=0}^{M-1}\mathbf{X}_{n-i}(k)\mathbf{H}_i(k),\quad k=0,\ldots,N/2 \tag{11}$$

where H_i is a segments of NLMS filter response in frequency domain.

IV. Simulations

4.1. Block-online BSS algorithm

We investigated the performance of the block-online BSS algorithm in a room environment with

reverberation time T_{60}=500 ms. Two speakers are 1 m apart from two microphones and microphone spacing is set to 15 cm. In order to investigate the performance of the block-online BSS algorithm, we first use two fixed speakers at 70° and 110°. Signals are sampled at 16 kHz and collected for 10 sec. Parameters used for the batch part are frame size 4096, block size 2048, frame rate 50%, μ=0.08, γ=0.5. For the block-online BSS, we set K=4, μ_{off}=0.16, λ=0.2, j_{max}=5. Table 2 is the performance of block-online for the fixed speakers. Notice that the block-online BSS provides slightly better SIR than the batch version. This is due to the fact that the block-online version adapts the time-varying nature of speech signals better than the batch version.

To investigate the separation performance of the block-online BSS algorithm, the speaker 2 at 110° was moved to 150° after 10sec. Figure 3 show the SIR performance of the block-online BSS algorithm for the 20 sec.

After moving of speaker 2 at 10 sec, the SIR of the fixed speaker was dropped while that of moved speaker increased. This is because mixing filters for the fixed speaker are estimated easily whereas those for the moved speaker are not. It takes about 2 sec to be recovered completely from the change. Compare to the SIR result in Table 2, we can see that the average SIR of the block-online BSS algorithm is degraded only by 1 dB for the changing environment.

Table 2. Separation performance for the fixed speakers.

	ch1 (dB)	ch2 (dB)	Avg (dB)
Batch	11.19	9.94	10.57
Block-online	12.59	10.18	11.39

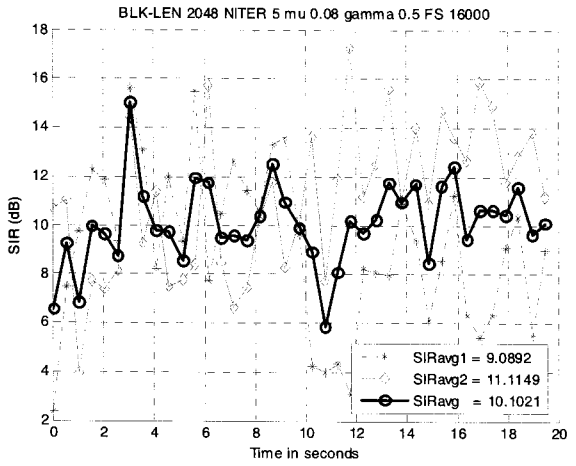

Fig. 3. SIR of the block-online BSS algorithm for a changing environment.

4.2. AEC combined with block-online BSS

Simulation Setup:

Microphone placement in a car and the processing steps for the AEC+BSS are shown in Fig. 4. This simulation setup is very tough because the far-end voice is played back by the six speakers of the car. Thus, this arrangement is very helpful to prevent channel permutation of the separated output [2]. It should be noticed that this microphone arrangement may not be sufficient for the frequency-domain BSS algorithm to avoid the channel permutation problem. In a car, the reverberation time is approximately 50 msec and sampling rate is 8 kHz. Since background noise level in the car (including engine noise and wind noise) is noticeably high, both BSS and AEC performances may degrade and Wiener filtering is necessary to eliminate the noise. In our simulation, AEC is setup with the NLMS adaptive filter of length

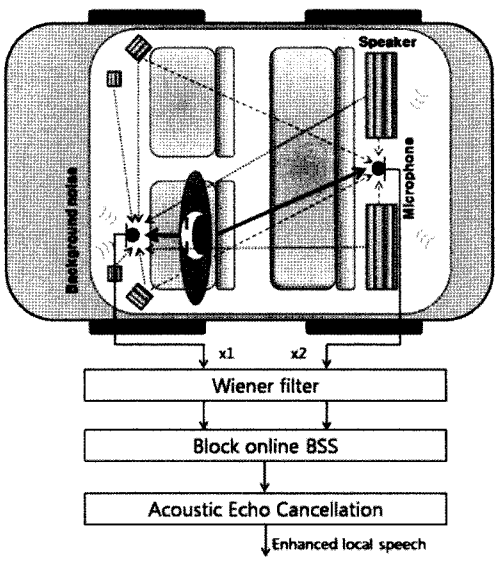

Fig. 4. Simulation setup and speech enhancement process.

256, block size $B=N/2=64$, step size $\mu=0.05$. Block-online BSS is set up with filter size $L=1024$, frame size 4L, and 10 learning iterations.

Contribution of the block-online BSS to double-talk detection

Since echo signal is suppressed by BSS and its average power is reasonably smaller than the power of local speech. So we can simply use energy-based double talk detection to detect double-talk period. For this purpose, we selected the Geigel method which compares the echo signal level to the reference signal level (from speaker) to detect double talk.

The average energy of last K blocks of the reference signal $|\bar{x}_i|$ is compared with the average energy of echo in current block $|\bar{d}|$, double talk is declared if

$$|\bar{d}| \geq c * \max\left(|\bar{x}_0|, |\bar{x}_1|, \ldots, |\bar{x}_{K-1}|\right) \tag{12}$$

where c is a threshold to reflect environment attenuation.

Fig. 5 shows the double-talk detection by the proposed method with the attenuation parameter $c_1=0.85$ for the AEC-only case and $c_2=0.35$ for the AEC+BSS case, parameters are optimized for the best results in each case. As we can see, the false double talk detection is reduced remarkably in the proposed AEC+BSS case due to echo suppression by BSS.

Contribution of block-online BSS to echo suppression

Fig. 6 shows the recorded signal by two microphones and the resulting echo remaining after BSS and AEC. After BSS, echo is remarkably suppressed while retaining the local speech. The residual echo after BSS is suppressed further in AEC so that only local speech is transmitted to the far-end side.

To evaluate the performance of AEC and AEC combines with BSS, the error return loss enhancement (ERLE) is calculated at time index n as follows

$$ERLE = \frac{\overline{x_{echo}^2}}{\overline{e_{echo}^2}}$$

where $\overline{x_{echo}^2}$ and $\overline{e_{echo}^2}$ are the power of the echo signal at microphone and the remain echo at the output

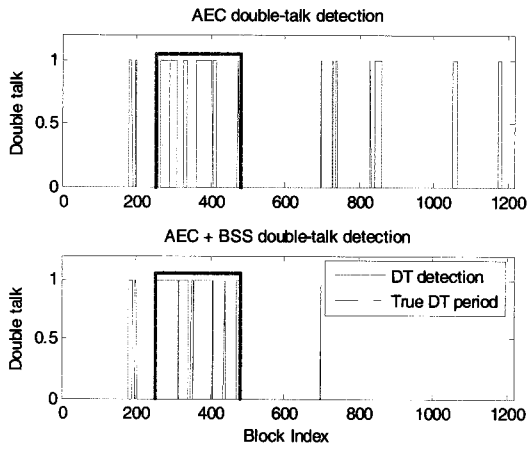

Fig. 5. Geigel double-talk detection for the AEC-only (above) and BSS+AEC (below) cases.

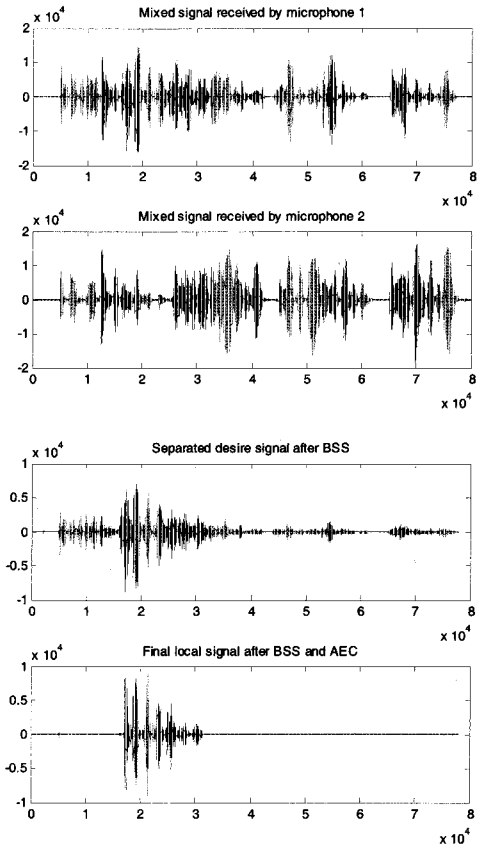

Fig. 6. Microphone speech signals and processed signals after BSS and AEC.

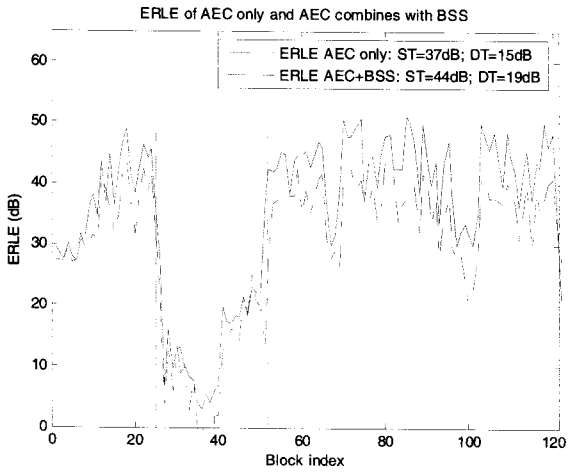

Fig. 7. Comparison of ERLE: AEC-only (dash-dot) and BSS+ AEC (solid). The ERLE values represent running-average over 10 blocks.

after AEC, respectively, averaged over 10 blocks. Note that, in this simulation, we use the same true double -talk period for AEC so that the improvement of ERLE only comes from SIR gain of BSS. The combined structure provides higher ERLE of approximately 7 dB in single talk and 4 dB in double talk period when compare to the AEC-only case as shown in Fig 7. With the help of more accurate double talk detection by energy based method, the AEC+BSS can provides higher improvement of ERLE, especially during double talk period.

V. Conclusions

In this paper, we have proposed a new application of BSS to enhance acoustic echo cancellation in a car environment. The proposed method is to combine BSS with AEC. Basic AEC and block-online BSS implementations have been described. Draft simulation results in a car revealed that putting BSS before AEC can enhance echo cancellation performance during both single talk and double talk periods in a hands -free communication system. Moreover, double talk detection becomes easier and more accurate. This study proves that BSS would be a viable technique for AEC in a real world situation.

Acknowledgement

This work was supported by the Korea Research Foundation Grant funded by Korean Government (MOEHRD) (KRF-2005-041-D00603) and Basic Research Promotion Fund (KRF-2006-211-D00114).

References

1. S.H. Nam and S. Beack, "A frequency-domain normalized multichannel deconvolution algorithm for acoustical signals," Proc. ICA/BSS 2004 (LNCS 3195), Springer, 522-529, 2004.
2. S. Beack, S.H. Nam, and M. Hahn, "A new speech enhancement algorithm for car environment noise cancellation with MBD and Kalman filter", IEICE trans. Fundamental, E88-A(3), March 2005.
3. S.H. Nam, "Frequency-domain normalized multichannel blind deconvolution: modifications and properties," Proc. of 2006 IEEE Signal Pro. Workshop on Machine Learning for Signal Processing, 307-312, Sept., 2006.
4. R. Mukai, H. Sawada, S. Araki, and S. Makino, *Real-time blind source separation for moving speech signals*, (in Speech Enhancement, Eds. J. Benesty, S. Makino, and J. Chen, Springer, 2005.)
5. R. Aichner, H. Buchner, F. Yan, and W. Kellermann, "Real-time convolutive blind source separation based on a broadband approach," Proc. ICA/BSS 2004 (LNCS 3195), Springer, 840-548, 2004.
6. D.Q.K. Ngoc and S.H Nam, "Implementation of a basic echo canceller", Proc. of the 19th Joint Signal Processing Conf., Sept., 2006.
7. D.Q.K. Ngoc, C. Park, and S.H Nam, An acoustic echo canceller combined with blind source separation, Proc. of the Acoustic Society of Korea conf., 24(1), Aug., 2007.
8. E. Hansler and G. Schmidt, *Acoustic Echo and Noise Control-A Practical Approach*, (Wiley, 2004.)
9. Microchip, *Acoustic Echo Cancellation Library*, (DS7013A.)
10. T. Liu, S. Gazor, "De-correlated algorithms for faster adaptation", Proc. of IEEE Int. Conf. on Signal Processing, 2002.

[Profile]

• Duong Q.K. Ngoc

Duong Q.K. Ngoc received the B.S. degree from the Posts and Telecommunications Institute of Technology (PTIT), Vietnam in 2004, and the M.S. degree in electronic engineering from Paichai University, Korea, in 2008, respectively. From 2004 to 2006, he was with Visco JSC as a system engineer for the audio /video conferencing system. He is now with EMERSYS as a research engineer. His research interest is speech enhancement in hands-free systems.

• Chul Park

Chul Park received the B.S. degree in electronic engineering from Paichai University, Korea, in 2007. Currently, he is a graduate student at master level in the Department of Electronic Engineering, Paichai University. His research interests include independent component analysis for blind source separation and speech enhancement.

• Seung-Hyon Nam

1980: BS, Dept. of Electronic Engineering, Sogang University
1987: MS, Dept. of Electrical and Computer Engineering, The University of Alabama, Huntsville
1992: PhD, Dept. of Electrical Engineering, Texas A&M University
1979-1985: Researcher, Agency for Defense Development, Daejeon, Korea
1993 - Present: Professor, Dept. of Electronic Engineering, Paichai University
Interested Areas: Speech and audio processing, data compression, adaptive filters