THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA VOL.25, NO.4E 2006. 12 pp. 148~153

A New Formulation of Multichannel Blind Deconvolution: Its Properties and Modifications for Speech Separation

Seung-Hyon Nam*, Innho Jee**
*Department of Electronic Engineering, Paichai University
**Department of Computer and Information Communication, Hongik University
(Received October 24 2006; Revised November 23 2006; accepted November 27 2006)

Abstract

A new normalized MBD algorithm is presented for nonstationary convolutive mixtures and its properties/modifications are discussed in details. The proposed algorithm normalizes the signal spectrum in the frequency domain to provide faster stable convergence and improved separation without whitening effect. Modifications such as nonholonomic constraints and off-diagonal learning to the proposed algorithm are also discussed. Simulation results using a real-world recording confirm superior performanceof the proposed algorithm and its usefulness in real world applications.

Keywords: *Blind source separation, Multichannel blind deconvolution, Natural gradient, Nonholonomic, Whitening, Speech enhancement.*

I. Introduction

Blind source separation (BSS) is a technique to separate original signals from a set of mixtures without any information on original signals or a mixing system except that original signals are statistically independent each other. In case of speech separation in a noisy room, the received signal at a microphone would be a reverberant and mixed version of sound sources and the mixing is convolutive.

Multichannel blind deconvolution (MBD) is one of practical methods for convolutive BSS. In [1,2], MBD algorithms that employ natural gradient with double-sided unmixing filters has been proposed. These MBD algorithms, however, suffer from whitening effect of output speech, slow convergence speed, and poor separation.

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

Recently, it has been shown by the author of this paper that these shortcomings of the existing MBD algorithm can be overcome by exploiting spectrum normalization and employing right-sided unmixing filters [3]. Further, its application to speech enhancement in a car has been demonstrated [4].

However, the proposed normalized MBD algorithm does not remove the whitening problem completely. In this paper, this deficiency is perfectly resolved by introducing nonholonomic constraints on the MBD algorithm. Furthermore, the computational burden is greatly reduced by introducing off-diagonal learning which does not use diagonal filters. In this case, diagonal filters are absorbed into off-diagonal terms. With these modifications, the resulting MBD algorithm will provide improved performance in terms of quality while reducing the computational burden into half (for the case of two sources and two sensors).

In this paper, the detail structure, properties, and modifications of the newly proposed algorithm are investigated and compared with existing MBD algorithms. Simulation results using real-world recording are presented to confirm the theoretical expectations.

II. Blind Signal Separation

2.1. Blind Signal Separation Problem.

In convolutive mixing with P sources and Q sensors, the mixed signal at the sensor j is given by

$$x_j(k)=\sum_{i=1}^{P}\sum_{p=-\infty}^{\infty}A_{ji,p}\,s_i(k-p),\ j=1,\dots,Q \tag{1}$$

where $s_i(k), i=1,\cdots,P$ are source signals and $A_{ji,p}$ is the p^{th} coefficient of the $(i,j)^{th}$ component (from the source i to the sensor j) of the mixing system. Assume that the unmixing system is given by $\mathbf{W}(z,k)=\sum_{p=-\infty}^{\infty}\mathbf{W}_p(k)z^{-p}$. Then the i^{th} separated signal is given by

$$u_i(k)=\sum_{j=1}^{Q}\sum_{p=-\infty}^{\infty}W_{ij,p}(k)x_j(k-p) \tag{2}$$

where $W_{ij,p}(k)$ is the $(i,j)^{th}$ component of $\mathbf{W}_p(k)$. The number of sensors Q is assumed to be equal to or greater than the number of sources P for successful separation.

One practical method to estimate the unmixing system $\mathbf{W}(z,k)$ is multichannel blind deconvolution (MBD) which minimizes iteratively mutual information between separated signals. In [1], the cost function,

$$\phi(W(z,k))=-\frac{1}{2\pi j}\int \log|\det W(z,k)|z^{-1}dz-\sum_{i=1}^{P}\log p_i(k) \tag{3}$$

where $p_i(u_i(k))$ is a probability density function of $u_i(k)$, is minimized with respect to $\mathbf{W}(k)$. Natural gradient (or relative gradient [5]) is known to be more efficient for this information geometry than standard gradient [6]. The update rule using natural gradient is then given by

$$\Delta\mathbf{W}_p(k)=\sum_{q=-\infty}^{\infty}\left\{\delta_{pq}\mathbf{I}-\mathbf{y}(k)\mathbf{u}^T(k-p+q)\mathbf{W}_q(k)\right\} \tag{4}$$

Here, $\mathbf{y}(k)=\mathbf{f}(\mathbf{u}(k))$ where $f_i(u_i)=\frac{d}{du_i}\log p_i(u_i)$. Since the unmixing filter $W_{ij}(z,k)$ is assumed to be doubly infinite, it should be approximated to a causal double-sided filter for practical implementation.

In [1], the unmixing filters are approximated with double-sided filters which are truncated and shifted so that the separated signal is given by

$$\mathbf{u}(k)=\sum_{p=0}^{L}\mathbf{W}_p(k)\mathbf{x}(k-p) \tag{5}$$

The MBD algorithm with the natural gradient (NGMBD) is then presented as

$$\Delta\mathbf{W}_p(k)=\mathbf{W}_p(k)-\mathbf{y}(k-L)\mathbf{v}^T(k-p) \tag{6a}$$

$$\mathbf{v}(k)=\sum_{q=0}^{L}\mathbf{W}_{L-q}^T(k)\mathbf{u}(k-q) \tag{6b}$$

The unmixing system is initialized with $\mathbf{W}(z,0)=\mathbf{I}z^{-p}$ for some $0\le p\le L$. If $0\le p\le L$, it converges to a set of $(P\times Q)$ double-sided finite impulse response (FIR) filters.

The NGMBD algorithm (6a) has equilibrium points

$$E\left\{y_i(k)u_j(k-l)\right\}=\delta_{ij}\delta_l \tag{7}$$

so that the unmixed output signals are whitened. Consequently quality of the unmixed speech signal is generally poor although it is still intelligible. Furthermore, nonstationarity of speech signals yields slow convergence rate and poor separation since convergence rate differs at each spectral region due to spectral tilt. In [3], it is demonstrated that natural gradient still suffers from slow convergence fornonstationary signals such as speech. In addition, delay in (6b) causes performance degradation.

2.2. A Frequency-Domain Normalized MBD (FNMBD) Algorithm.

If the right-sided filters are used, the natural gradient updating rule (4) can be written as

$$\Delta\mathbf{W}_p(k)=\sum_{q=0}^{L-1}\left\{\delta_{pq}\mathbf{I}-\mathbf{y}(k)\mathbf{u}^T(k-p+q)\right\}\mathbf{W}_q(k) \tag{8}$$

without any delay and approximations. The filter is initialized with $W(z,0)=I$. The update rule (8) is the same as derived in [7]. A major advantage of this alternative formulation is accessibility of $\Delta X=(I-yu^T)$. The gradient term $\Delta\mathbf{X}$ can be modified for better convergence and separation performance.

Notice that MBD algorithms (6) and (8) adopt sample-by-sample processing. The MBD algorithm (8) can be implemented in the frequency domain using the advantage of FFT in an overlap-save manner. It can be expressed in the frequency domain, in simplified matrix form omitting time-domain constraints, as

$$\Delta\mathbf{W}^f(b)=\left\{\bar{\mathbf{I}}-\mathbf{y}^f(b)(\mathbf{u}^f(b))^H\right\}\mathbf{W}^f(b) \tag{9}$$

where b denotes block index, the superscript f the quantity in the frequency domain, H the hermitian transposition. In addition, $\bar{\mathbf{I}}=diag(\bar{1},\cdots,\bar{1})$ where $\bar{1}=(1,\cdots,1)$. Notice that (9) is actually a time-domain algorithm so that forward/inverse Fourier transforms and proper time-domain constraints are required to compute linear convolution and correlation via circular convolution and correlation, respectively. Notice that the algorithm (9) with double-sided filters is presented in [2]. However, as we will discuss later, algorithms with double-sided filters are not as robust as those with right-sided filters in real-world applications.

From (9), we can now express a new normalized MBD algorithm by normalizing the signal spectrum $y^f(b)$ and $(u^f(b))^H$ in the frequency domain. In this paper, only the algorithm with right-sided filters is considered. (Frame packing and corresponding time-domain constraints for the double-sided filters can be found in [8].)

First, define the signal with normalized power spectrum as follows:

$$\bar{\mathbf{y}}_i^f(b)=\mathbf{y}_i^f(b)\varnothing\sqrt{P_{y_i}(b)} \tag{10a}$$

$$\bar{\mathbf{u}}_i^f(b)=\mathbf{u}_i^f(b)\varnothing\sqrt{P_{u_i}(b)} \tag{10b}$$

where $\varnothing$ denotes the component-wise division and

$$\mathbf{P}_{y_i}(b)=(1-\gamma)\mathbf{P}_{y_i}(b-1)+\gamma\left|\mathbf{y}_i^f(b)\right|^2 \tag{11a}$$

$$\mathbf{P}_{u_i}(b)=(1-\gamma)\mathbf{P}_{u_i}(b-1)+\gamma\left|\mathbf{u}_i^f(b)\right|^2 \tag{11b}$$

where $0<\gamma<1$. Let $\mathbf{F}$ and $\mathbf{F}^{-1}$be the Fourier transform and the inverse Fourier transform matrices, respectively. Further, let $\mathbf{P}_{L,0}$ be the matrix that discards all the samples except the first L samples. Then the right hand side of (9) can be converted into normalized form with proper time-domain constraints as follows:

$$\Delta X_{ij}^f(b)=\bar{1}\delta_{ij}-\mathbf{F}\mathbf{P}_{L,0}\mathbf{F}^{-1}\left\{\bar{\mathrm{y}}\odot(\overline{\mathrm{u}_j^f})^*\right\}. \tag{12a}$$

$$\Delta W_{ij}^f(b)=\mathbf{F}\mathbf{P}_{L,0}\mathbf{F}^{-1}\left\{\sum_{l=1}^{P}\Delta X_{ij}^f(b)\odot W_{ij}^f(b)\right\} \tag{12b}$$

where $\odot$ denotes the component-wise multiplication. The time-domain constraint $\mathbf{P}_{L,0}$ in (12a) is to preserve only the first L cross-correlation lags that are computed using circular correlation. To obtain the unbiased L cross-correlation lags between $\mathbf{y}_i(b)$ and $\mathbf{u}_j(b)$, the support of $\mathbf{y}_i(b)$ should be shorter than the support of $\mathbf{u}_i(b)$ by L samples so that the first $2L$ samples of $\mathbf{y}_i(b)$ are zero. That is,

$$\mathbf{y}_i(b)=\mathbf{P}_{0,N-2L}\mathbf{f}(\mathbf{u}_i(b)) \tag{13}$$

For this reason, the frame length N should satisfy $N\geq 4L$. On the other hand, the time-domain constraint $\mathbf{P}_{L,0}$ in (12b) is to limit the filter length to L. Notice that (12b) is a convolution of two sequences having supports L and aliasing does not occur if $N\geq 4L$. Owing to time-domain constraints in (12a) and (12b), the algorithm does not suffer from the frequency permutation problem.

Using the same notational convention as in (9), the resulting frequency-domain normalized MBD (FNMBD) algorithm can be expressed in a simplified form as

$$\Delta\mathbf{W}^f(b)=\left\{\bar{\mathbf{I}}-\Lambda_y^{-1/2}(b)\mathbf{y}^f(b)(\mathbf{u}^f(b))^H\Lambda_u^{-1/2}(b)\right\}\mathbf{W}^f(b) \tag{14}$$

where $\Lambda_y(b)$ and $\Lambda_u(b)$ are diagonal matrices with diagonal elements $\mathbf{P}_{y_i}(b)$ and $\mathbf{P}_{u_i}(b)$, respectively. Due to spectral normalization employed in (14), the spectral tilt is effectively compensated so that the stability of the algorithm is greatly enhanced.

Notice that the off-diagonal terms of $\mathbf{y}^f(b)(\mathbf{u}^f(b))^H$ in (14) is normalized by the power of two sources. This is quite different from the method presented in [12,13] where the off-diagonal terms are normalized with respect to just single source.

III. Properties and Modifications of the FNMBD Algorithm

3.1. Properties.

In [7], stability analysis of the algorithm (14) has been accomplished by assuming that signals are not only mutually independent but also i.i.d. temporally. Likewise stability of the FNMBD algorithm can be treated indirectly by considering stability of the corresponding time-domain algorithm. In the corresponding time-domain algorithm, $\mathbf{y}(k)$ and $\mathbf{u}(k)$ are replaced by $\bar{\mathbf{y}}(k)$ and $\bar{\mathbf{u}}(k)$, respectively, having normalized spectrum. In fact, flattened spectrum of $\bar{u}_i(k)$ reinforces the i.i.d. assumption employed in the stability analysis so that stability of the FNMBD algorithm (14) is justified.

The FNMBD algorithm has equilibrium points

$$E\left\{\bar{y}_i^f(b)\bar{u}_j^f(b)\right\} = \delta_{ij} \tag{15}$$

Unlike (7), equilibrium points (15) do not impose any compulsory constraints on the spectra unmixed signals since $\bar{y}_i^f(b)$ and $\bar{u}_j^f(b)$ are already whitened.

Another good property of the FNMBD algorithm is equivariant property [5]. Assume that mixing and unmixing systems have the same supports. Post-multiplying the mixing filter $\mathbf{A}^f(b)$ to both sides of (14) yields

$$\Delta\mathbf{H}^f(b) = \left\{\mathbf{I} - \Lambda_y^{-1/2}(b)\mathbf{y}^f(b)(\mathbf{u}^f(b))^H\Lambda_u^{-1/2}(b)\right\}\mathbf{H}^f(b) \tag{16}$$

where $\mathbf{H}^f(b) = \mathbf{W}^f(b)\mathbf{A}^f(b)$. Dependency on the mixing matrix is absorbed as an initial condition so that separation is independent of the mixing system. Furthermore, dependency on the input signal power is removed through normalization so that the FNMBD algorithm provides uniform convergence regardless of the input signal as well as the mixing system. Therefore, convergence of the FNMBD algorithm depends only on the step size and the filter length.

3.2. Modifications.

As pointed out earlier, the frame length N should be larger than $4L$ to compute L unbiased cross-correlation lags in (12a). For the double-sided filters, however, $2L$ cross-correlations of lag from $-(L+1)$ to L are required. If we omit this constraint, the cross-correlations are biased but more samples of $\mathbf{u}(b)$ are utilized.

The nonholonomic constraint is that diagonal components of $\Delta\mathbf{X} = \mathbf{I} - \mathbf{y}\mathbf{u}^T$ being zero [11]. The nonholonomic constraint prevents the gradient terms from being affected by time-varying signal powers. The nonholonomic constraint is therefore proposed as a solution to the whitening effect. Nevertheless, the exact nonholonomic constraint is not possible to implement in the NGMBD algorithm since $\Delta\mathbf{X}$ is not accessible due to backward filtering. On the other hand, the FNMBD algorithm is approximatelynonholonomic, and exact nonholonomicity can be achieved in the FNMBD algorithm by setting $\Delta\mathbf{X}_{ii}^f(b) = 0$.

In general, all (i,j) components of the filter are learned and utilized in separating the signals. In off-diagonal learning, however, only diagonal components of the filter are fixed to unit impulses and only off-diagonal components are learned. Diagonal components are in factabsorbed in the off-diagonal components since there exists inherit indeterminacy to arbitrary scaling and filtering. However, off-diagonal learning provides computational saving. This savings can be significant if it is combined with nonholonomic constraints for $P = Q = 2$. Consider the off-diagonal learning of the gradient $\Delta\mathbf{W}^f(b)$ in (12a) and (12b) with the nonholonomic constraints. For the 2x2 case, off-diagonal components of the filter are rewritten as

$$\Delta\mathbf{W}_{12}^f(b) = -\Delta\mathbf{X}_{12}^f(b) = \mathbf{F}\mathbf{P}_{L,0}\mathbf{F}^{-1}\left\{\bar{\mathbf{y}}_1^f(b)\,\mathrm{e}\left(\bar{\mathbf{u}}_2^f(b)\right)^*\right\} \tag{17a}$$

$$\Delta \mathbf{W}_{21}^{f}(b) = -\Delta \mathbf{X}_{21}^{f}(b) = \mathbf{F}\mathbf{P}_{L,0}\mathbf{F}^{-1}\left\{\bar{\mathbf{y}}_{2}^{f}(b)\, \mathrm{e}\, \left(\bar{\mathbf{u}}_{1}^{f}(b)\right)^{*}\right\} \quad \textbf{(17b)}$$

Notice that computation (12b) is not required and the computational burden is reduced to nearly half. However, the modified update rule (17) does not have natural gradient.

IV. Simulations

4.1. Simulation Setup.

Real world recordings were obtained in a normal office (3.2m x 7.8m x 3m) which has a reverberation time T60=500ms Using two microphones. Speech signals are played back at the speakers located at -30° and 40° with respect to the normal direction of the microphone array. The distance between the microphone and speaker is set to 70 cm and the two microphones are 12 cm apart each other. The mixed signals are recorded at 16 kHz sampling rate. As a performance measure, the signal-to-interference ratio (SIR), defined by the ratio of the signal power of the target signal from that of the jammer signal, is used.

4.2. Performance of the FNMBD Algorithm.

To see the effects of spectrum normalization and right-sided filters, we compared the proposed FNMBD algorithm with existing MBD algorithms as shown in Fig. 1. The filter length is set to 1024 for all algorithms. The step size is chosen differently for each algorithm to get proper stable convergence and the best SIR. The NGMBD algorithm (6) is initialized with $\mathbf{W}(z,0) = \mathbf{I}z^{-L/2}$ and $\mathbf{W}(z,0) = \mathbf{I}$ for double-sided (bidirectional) and right-sided (unidirectional) filters, respectively. The double-sided NGMBD algorithm performs poorer than the right-sided case. This is due to bias presented in the double-sidedfilters. The MBD algorithm (9) with nonholonomic constraints performs better than the right-sidedNGMBD since there is no delay in (9). Finally, the FNMBD algorithm with right-sided filters (14) performs better than the MBD algorithm (9) with nonholonomic constraints. The result clearly confirms the advantage of spectrum normalization and right-sided filters.

We compared the performance of the FNMBD with its modified versions (i) the FNMBD with nonholonomic constraints (ii) the FNMBD with nonholonomic plus off-diagonal learning. From Fig. 2, we can see that the SIR of the FNMBD algorithm is continuously increasing. It implies that the whitening effect is still remaining although it is reduced significantly. As nonholonomic constraints applied, however, the learning stops after reaching its equilibrium state so that there is no whitening occurs any longer. The algorithms becomes slower slightly as off-diagonal learning is used.

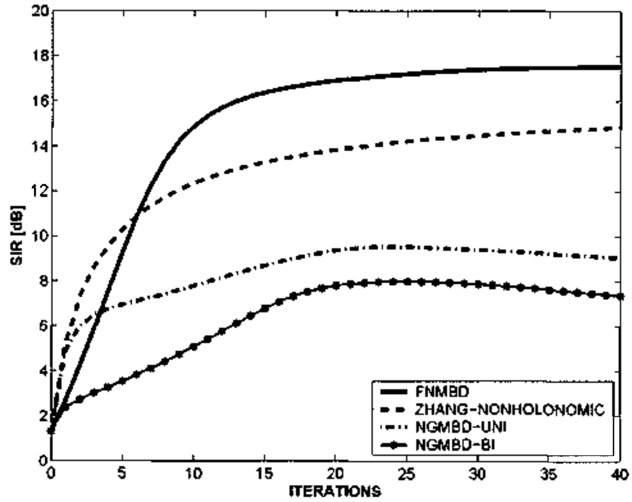

Fig. 1. Performance of the FNMBD algorithm and other existing MBD algorithms.

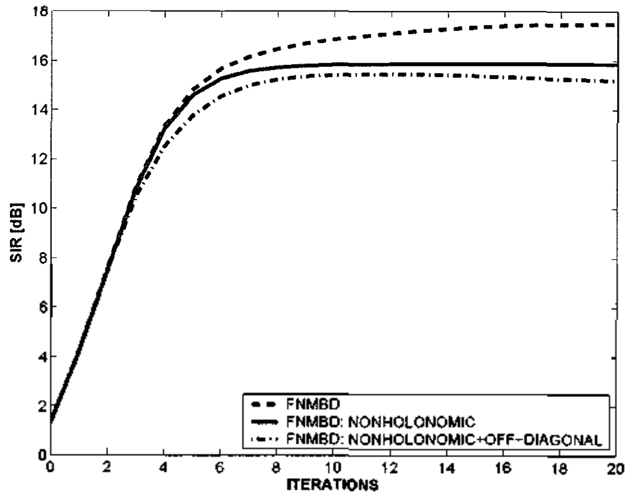

Fig. 2. Performance of the FNMBD algorithm and its modified versions.

V. Conclusions

A new normalized MBD algorithm has been presented for nonstationary convolutive mixtures and its properties have been discussed in details. The proposed algorithm uses right-sided filters combined with spectrum normalization. As a result, it provides faster stable convergence and improved separation while relieving the whitening effect. Modifications such as exact nonholonomic constraints and off-diagonal learning to the proposed algorithm also have been discussed. These modifications provide better performance while reducing the computational burden greatly. Simulation results using real-world recordings confirm superior performance of the proposed algorithm and its usefulness in real world applications.

Acknowledgment

This work was supported by the Korea Research Foundation Grant funded by Korean Government (MOEHRD) (R05-2004-000-10290-0).

References

1. S.-I. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Novel on-line adaptive learning algorithms for blind deconvolution using the natural gradient approach," Proc. IEEE 11th IFAC Symposium on System Identification, SYSID-97, Kitakyushu, Japan, 1057-1062, 1997.
2. T.-W. Lee, A. J. Bell, and R. Lambert, "Blind separation of delayed and convolved sources," Advances in Neural Information Processing Systems, MIT Press, **9** 758-764, 1997.
3. S. H. Nam and S. Beack, "A frequency-domain normalized multichannel blind deconvolution algorithm for acoustical signals," Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'04), Granada, Spain, 522-529, 2004.
4. S. Beack, S. H. Nam, and M. Hahn, "A new speech enhancement algorithm for car environment noise cancellation with MBD and Kalman filtering," IEICE Trans. Fundamentals, vol. E88-A, 3 685-689, March 2005.
5. J.-F. Cardoso, "Blind signal separation: statistical principles," Proceedings of the IEEE, **9**(10) 2009-2025, 1998.
6. S.-I. Amari, "Natural gradient works efficiently in learning," Neural Computation, **10**(2) 251-276, 1998.
7. L. Zhang, A. Cichocki, and S. Amari, "Geometrical structures of FIR manifold and their application to multichannel blind deconvolution," Proc. IEEE Workshop on Neural Networks for Signal Processing (NNSP'99), Madison, Wisconsin, 303-312, 1999.
8. M. Joho and P. Schniter, "Frequency domain realization of multichannel blind deconvolution algorithm based on the natural gradient," Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'03), 2003.
9. C. Fancourt and L. Parra, "The generalized sidelobe decorrelator," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 167-170, Oct. 2001.
10. H. Buchner, R. Aichner, and W. Kellermann, "A generalization of a class of blind source separation algorithms for convolutional mixtures," Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'03), 2003.
11. S.-I. Amari, T.-P. Chen, and A.Cichocki, "Nonholonomic orthogonal learning Algorithms for blind source separation," Neural Computation, 12(6) 1463-1484, 2000.
12. C. Fancourt and L. Parra, "The coherence function in blind source separation of convolutive mixtures of non-stationary signals," Proc IEEE Workshop on Neural Networks for Signal Processing, 303-312, 2001.
13. H. Buchner, R. Aichner, and W. Kellermann, "Blind source separation for convolutional mixtures: A unified treatment," in Audio Signal Processing: For next-generation multimedia communication systems, Y. Huang and J. Benesty (Eds), Kluwer Academic Pub., 2004.

[Profile]

• **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: Junior 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 filter

• **Innho Jee**

1980 : BS, Dept. of Electronics Engineering, Seoul National University
1983 : MS, Dept. of Electronics Engineering, Seoul National University
1995 : PhD, Dept. of Electrical and Computer Engineering, Polytechnic University, New York City
1982-1988 : Junior Researcher, Agency for Defense Development, Daejeon, Korea
2004-2005 : Research Professor, Dept. of Electrical and Computer Engineering, University of Maryland at College Park
1995-Present : Professor, Dept. of Computer and Information Communications, Hongik University
Interested Areas : Multimedia Signal Processing, Multimedia Security, 3D Image Processing.