LETTER *Special Section on Adaptive Signal Processing and Its Applications*

A New Speech Enhancement Algorithm for Car Environment Noise Cancellation with MBD and Kalman Filtering

Seungkwon BEACK[†a)], **Seung H. NAM**[††], ***and*** **Minsoo HAHN**[†], ***Nonmembers***

SUMMARY We present a new speech enhancement algorithm in a car environment with two microphones. The car audio signals and other background noises are the target noises to be suppressed. Our algorithm is composed of two main parts, i.e., the spatial and the temporal processes. The multi-channel blind deconvolution (MBD) is applied to the spatial process while the Kalman filter with a second-order high pass filter, for the temporal one. For the fast convergence, the MBD is newly expressed in frequency-domain with a normalization matrix. The final performance evaluated with the severely car noise corrupted speech shows that our algorithm produces noticeably enhanced speech.

key words: *speech enhancement, multichannel blind deconvolution, Kalman filter*

1. Introduction

Recently, speech recognition algorithms have been applied to the car environments, but inevitably their performances are severely degraded due to car noises. Car noises can generally be categorized into spatial and temporal ones. Spatial noises come from single or multiple point sources such as client's speech and car audio signals while temporal noises are the ones generated by the engine vibration, the air turbulence, and the tire friction.

The spatial car noises are one of the hardest obstacles in speech enhancement. Recently, as the MBD has been introduced for the spatial noise suppression, the speech quality becomes dramatically improved. However, the simultaneous occurrence case of the spatial and the temporal noises is still a challenging task. In [3], several types of spatial or temporal noises were independently tested with the MBD combined with the sub-band processing and produced rather successful results. The simultaneous cases in car environment were tested with the spatio-temporal enhancement technique [2]. It produced fairly enhanced speech with rather complicated processes.

In this paper, we present a new speech enhancement algorithm basically based on the spatio-temporal process as in [2]. However, the components of our process are newly proposed and the overall structure is relatively simple. Our goal is to design a successful speech enhancement algorithm in a car environment where various car audio and background noises exist simultaneously. We exclude other clients' interfering speech noises partly because the drivers are usually more often disturbed by the car audio than the client's speech in a running car.

For the robustness of our algorithm, the frequency-domain block-based MBD (FB-MBD) with a normal-ization matrix is proposed as a part of the spatial process. The matrix is designed to overcome the intrinsic problems of the time-domain MBD (TD-MBD) such as the whitening effect and the slow convergence [8]. It is derived from the information-maximization-based TD-MBD with natural gradients [1]. The temporal process based on the high pass and the Kalman filtering to suppress the background noises is subsequently applied [7].

We use the following notation throughout the paper. Bold uppercases and lowercases are used for matrices and vectors, respectively. Normal lowercases are for scalars and vector elements. "n" and "b" are for the time and the block indices, respectively. The superscript f indicates the frequency-domain quantity.

2. Problem Description

The corrupted signal, $x_j(n)$ from the jth-microphone can be represented as the convolution sum,

$$x_j(n)=\sum_{p=0}^{Q} a_{jj,p}s_j(n-p)+\sum_{\substack{i=1\\ i\neq j}}^{l}\sum_{p=0}^{Q} a_{ji,p}s_i(n-p)+v(n),\quad j=1,2,\ldots,m. \tag{1}$$

In this equation, $a_{ji,p}$ is the (j,i) element of the mixing system of $\boldsymbol{A}(z)=\sum_{p=0}^{Q}\boldsymbol{A}_p z^{-p}$ with the finite impulse response (FIR) Q, and $s_i(n)$ is the ith-point source. The second term on the right-hand side indicates the group of spatial noises propagated from the several interfering point sources, where $s_j(k)$ is the target speech signal. $v(n)$ is the microphone-independent temporal noise. Similarly, the jth-unmixing signal can be described as

$$u_j(n)=\sum_{p=0}^{Q}\left(w_{jj,p}(n)x_j(n-p)+\sum_{\substack{i=1\\ i\neq j}}^{l} w_{ji,p}(n)x_i(n-p)+\sum_{i=1}^{l} w_{ji,p}(n)v(n-p)\right) \tag{2}$$

Manuscript received June 24, 2004.
Manuscript revised October 2, 2004.
Final manuscript received November 11, 2004.

[†]The authors are with Multimedia Group, Information and Commun. University, 119 Munjiro, Yuseong-gu, Daejeon, 305-732, Korea.
[††]The author is with Dept. of Engineering, Paichai University, 14 YanJa Gil, Daejeon, 302-735, Korea.
a) E-mail: skbeack@icu.ac.kr

DOI: 10.1093/ietfec/e88–a.3.685

where $w_{ji,p}$ is the (j, i) element at the lag p of the unmixing system at time n.

The fundamental problem of the MBD comes from the fact that it estimates the unmixing system using the mixed signal without any information of the mixing system under the assumption that the number of microphones is equal to that of sources. The car audio signal propagated from multiple sources violates this assumption because only two microphones are adopted in our approach. If the spatial noises are successfully separated, the desired speech signal is obtained as the arbitrarily delayed and scaled version of the original speech with added convolution sum as in Eq. (3). This $u_j(n)$ is generally accepted as the MBD solution.

$$u_j(n) = d'_j s_j(n-\tau_j) + \sum_{p=0}^{Q} w_{ji,p}(n) v(n-p). \tag{3}$$

The channel permutation (CP) has also been accepted as a solution in the MBD [1], [6]. This CP problem occurs when the microphone in the direction to the jth-point source fails to keep the jth-point source power dominantly. The temporal processing requires the unmixed signal identification. Without the channel identification, the added computational burden to identify $u_j(n)$ is required. Several MBD approaches in a car have adopted closely coupled two microphones for other clients' speech suppression. In our algorithm, the spatial approach is mainly for the car audio signal cancellation. In this case, the CP can easily occur. Therefore, it has to be admitted that our algorithm would not always be free from the CP problem.

The MBD process does not influence the inherently reverberant temporal noises since it only maximizes spatial independency when their originations are deterministic. So the temporal noise, the second term in Eq. (3), still exists as a filtered version after the MBD. Hence, the temporal noise reduction needs another process as mentioned in [2].

3. Speech Enhancement Scheme

The scheme of our proposed algorithm is shown in Fig. 1. The microphone positions are selected in consideration of the CP problem. Our algorithm firstly reduces the car audio signal from the mixed x_1 with the MBD. As a result, u_1 becomes the unmixed output corresponding to x_1. Then the temporal noises in u_1 are suppressed by the temporal process.

3.1 Frequency Domain Block-Based MBD (FB-MBD)

Gradient algorithms show fast convergence properties when their estimations are based on block processes due to the more accurate gradient vector estimation [10]. The computation power can also be reduced by overlap-save or overlap-add methods [12]. In this section, we formulate the normalized FB-MBD algorithm (NFB-MBD). It is confirmed that our algorithm is almost free from the speech whitening effect while converging rather fast.

3.1.1 Time Domain Formulation of MBD

As in [1], the gradient equation of the MBD can be represented as

$$\Delta \boldsymbol{W}_p(n) = \left(\boldsymbol{W}_p(n) - \boldsymbol{y}(n-L) \sum_{q}^{L-1} \boldsymbol{W}_{L-q}^T \boldsymbol{u}(n-q) \right) \tag{4}$$

where $\boldsymbol{y}(n)$ is $f(\boldsymbol{u}(n))$ for some monotonic nonlinear function $f(\cdot)$, $\boldsymbol{u}(k)$ is $[u_1(k) \cdots u_m(k)]^T$, and the filter length L is the truncated version of Q. The variation of $\mathrm{E}\left\{\Delta \boldsymbol{W}_p(k)\right\}$ depends mainly on the eigenvalue spread of the cross-correlation matrix between $\boldsymbol{y}(k-L)$ and $\boldsymbol{u}(k)$. The variation can be smoothed through the cross-correlation matrix normalization based on the diagonal terms. The anti-causal filtering, the second term of Eq. (4), however, makes it difficult to derive the normalized form directly, because there is L sample delay assumption to reduce the computation [1]. We ignore the anti-causal part of $\boldsymbol{W}_p(k)$, because the causal FIR can be successfully used as an unmixing system in a small reverberant space such as in a car. Then, Eq. (4) can be represented as the correlation matrix whose lag is a positive integer.

$$\Delta \boldsymbol{W}_p(n) = \sum_{q=0}^{p} \left(\boldsymbol{I}\delta_{p-q} - \boldsymbol{y}(n)\boldsymbol{u}^T(n-p+q) \right) \boldsymbol{W}_q(n), \quad 0 \le q \le p \le L-1. \tag{5}$$

Equation (5) is identical to the causal MBD algorithm derived from the manifold using the isometry of the Riemannian metric [5]. If $\boldsymbol{u}(n)$ is adjusted to have the unit variance, the equilibrium point of Eq. (5) is

$$\mathrm{E}\left\{ \boldsymbol{I}\delta_{p-q} - \boldsymbol{y}(n)\boldsymbol{u}^T(n-p+q) \right\} = 0 \tag{6}$$

Equation (6) can be modified by the normalized matrix $\boldsymbol{\Lambda}^{-1}$ on the constraint of the same equilibrium point,

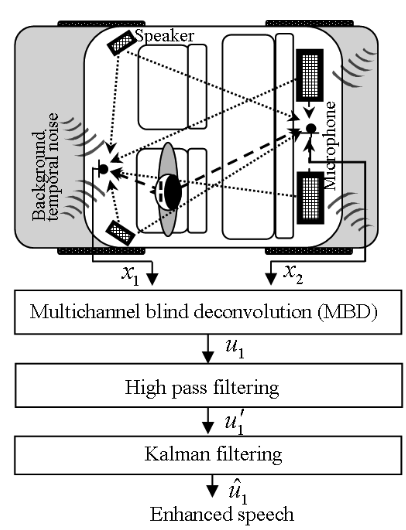

Fig. 1 Proposed speech enhancement process.

$$\mathrm{E}\left\{\boldsymbol{I}-\boldsymbol{\Lambda}^{-1}\boldsymbol{y}(n)\boldsymbol{u}^T(n-p+q)\right\}=0. \tag{7}$$

If $\boldsymbol{\Lambda}^{-1}$ is designed as a diagonal matrix whose elements are the diagonal elements of $\mathrm{E}\left\{\boldsymbol{u}(n)\boldsymbol{u}^T(n-p+q)\right\}$, Eq. (7) has the spatial self-orthogonalization property. And according to the Bussgang property, $\mathrm{E}\left\{\boldsymbol{y}(n)\boldsymbol{u}^T(n)\right\}$ becomes $\mathrm{E}\left\{\boldsymbol{u}(n)\boldsymbol{u}^T(n)\right\}$ [12]. Our normalization technique is realized in preference of the temporal orthognalization rather than the spatial one because the MBD's temporal i.i.d.(identically independent distribution) assumption mainly contributes to the performance degradation of the TD-MBD in case of the convolved acoustic mixture. The NFB-MBD is derived by modifying Eqs. (5) and (7).

3.1.2 Realization of NFB-MBD

For the block-based implementation in frequency domain, all the following fundamental vectors in time domain are defined with L, in the form of 2's power, elements for the convenient FFT implementation. The fundamental vectors are defined as,

$$\boldsymbol{x}_i(b)\triangleq[x_i(bL)\cdots x_i(bL+L-1)]^T \tag{8}$$

$$\boldsymbol{u}_j(b)\triangleq\left[u_j(bL)\cdots u_j(bL+L-1)\right]^T \tag{9}$$

where $i,j=1,2,\cdots,m$. These vectors are packed to be represented by the frame as follows

$$\tilde{\boldsymbol{x}}_i(b)\triangleq[\boldsymbol{x}_i(b-k+1)\cdots\boldsymbol{x}_i(b-1)\boldsymbol{x}_i(b)]^T \tag{10}$$

$$\tilde{\boldsymbol{u}}_j(b)\triangleq\left[\underline{\boldsymbol{u}}_j(b-k+1)\cdots\boldsymbol{u}_j(b-1)\boldsymbol{u}_j(b)\right]^T \tag{11}$$

$$\tilde{\boldsymbol{y}}_j(b)\triangleq f(\tilde{\boldsymbol{u}}_j(b)). \tag{12}$$

Here, k determines the hop size in the correlation matrix and $\underline{\boldsymbol{u}}_j(b-k+1)$ is a surplus circulant part produced by the circular convolution. We use $\mathcal{F}$ to denote the $kL\times kL$ FFT matrix [10] , and the frequency domain input vector is

$$\boldsymbol{x}_j^f(b)=\mathcal{F}\tilde{\boldsymbol{x}}_j(b) \tag{13}$$

In order to represent the point-wise multiplication procedure mathematically, the diagonal matrix $\boldsymbol{X}_j^f(b)$ can be defined as $\mathrm{diag}\left[\boldsymbol{x}_j^f(b)\right]$.

When the frequency domain unmixing vector of length L is defined as follows,

$$\boldsymbol{w}_{ij}^f(b)\triangleq\mathcal{F}\left[w_{ij,0}(b)\cdots w_{ij,L-1}(b),\underbrace{0\cdots0}_{(k-1)L}\right]^T \tag{14}$$

the output vector in (10) can be evaluated as

$$\underline{\boldsymbol{u}}_j^f(b)=\boldsymbol{X}_j^f(b)\boldsymbol{w}_{ij}^f(b). \tag{15}$$

For the linear convolution, the circulant part of $\tilde{\boldsymbol{u}}_j(b)$ should be removed by

$$\boldsymbol{u}_j^f(b)=\mathcal{P}_{O,kL-L}\underline{\boldsymbol{u}}_j^f(b) \tag{16}$$

where $\mathcal{P}_{O,kL-L}$ is the following window matrix.

$$\mathcal{P}_{O,kL-L}\triangleq\mathcal{F}\begin{bmatrix}\boldsymbol{0}&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{I}_{kL-L}\end{bmatrix}\mathcal{F}^{-1}. \tag{17}$$

The cross correlation matrix between the jth-unmixing vector $\tilde{\boldsymbol{u}}_j(b)$ and the ith-nonlinear vector $\tilde{\boldsymbol{y}}_i(b)$ can be obtained as $\boldsymbol{\varphi}_{ij}^f(b)$ in frequency domain,

$$\boldsymbol{\varphi}_{ij}^f(b)=\mathcal{P}_{L,O}\underline{\boldsymbol{\varphi}}_{ij}^f(b) \tag{18}$$

$$\underline{\boldsymbol{\varphi}}_{ij}^f(b)=\mathrm{diag}\left[\boldsymbol{y}_i^f(b)\left(\boldsymbol{u}_j^f(b)\right)^H\right] \tag{19}$$

where $\boldsymbol{y}_i^f(b)$ is $\mathcal{F}\tilde{\boldsymbol{y}}_i(b)$ and $\mathcal{P}_{L,O}$ is another window matrix.

$$\mathcal{P}_{L,O}\triangleq\mathcal{F}\begin{bmatrix}\boldsymbol{I}_L&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{0}\end{bmatrix}\mathcal{F}^{-1}. \tag{20}$$

Therefore, we can formulate the (i,j)th-element of the causal FB-MBD algorithm,

$$\tilde{\boldsymbol{w}}_{ij}^f(b)=\sum_{r=1}^{m}\mathcal{P}_{L,O}\left(\boldsymbol{I}_N\delta_{ijr}-\boldsymbol{\varphi}_{ir}^f(b)\right)\boldsymbol{w}_{rj}^f(b). \tag{21}$$

where $\delta_{ijr}=1$ when $i=j=r$, otherwise zero. The hop size of correlation is proportional to k within the frame length kL. It can be expected that the large hop size with the large k estimates more accurate cross correlation quantities despite of the relatively short filter length L. The increased hop size, however, does not always guarantee the accuracy improvement mainly because speech and acoustic signals are considered to be stable only for the short-time duration. In other words, the extended filter length is not essential for the separation performance improvement and the details about this can be found in [11]. Hence, the value of k is experimentally determined in our study.

In order to implement the NFB-MBD with the temporal self-orthogonalization property, we introduce a diagonal matrix $\boldsymbol{\Lambda}_{ij}^f(b)$,

$$\boldsymbol{\Lambda}_{ij}^f(b)\triangleq\mathrm{diag}\left[\boldsymbol{u}_i^f(b)\left(\boldsymbol{u}_j^f(b)\right)^H\right] \tag{22}$$

and the matrix is updated with the forgetting factor γ as follows

$$\boldsymbol{\Lambda}_{ij}^f(b)=(1-\gamma)\boldsymbol{\Lambda}_{ij}^f(b-1)+\gamma\boldsymbol{\Lambda}_{ij}^f(b)+\alpha\boldsymbol{I}_N. \tag{23}$$

Here, α is an arbitrary small constant to keep the normalization matrix nonsingular. The cross correlation matrix in frequency domain is now redefined as the normalized form of $\boldsymbol{\varphi}_{ij}^{f,N}(b)$.

$$\boldsymbol{\varphi}_{ij}^{f,N}(b)=\mathcal{P}_{L,O}\left(\boldsymbol{\Lambda}_{ij}^f(b)\right)^{-1}\underline{\boldsymbol{\varphi}}_{ij}^f(b). \tag{24}$$

Finally, we can formulate the NFB-MBD by substituting Eq. (24) for Eq. (21):

$$\tilde{\boldsymbol{w}}_{ij}^f(b)=\sum_{r=1}^{m}\mathcal{P}_{L,O}\left(\boldsymbol{I}_N\delta_{ijr}-\boldsymbol{\varphi}_{ir}^{f,N}(b)\right)\boldsymbol{w}_{rj}^f(b). \tag{25}$$

We tested two matrices $\mathbf{\Lambda}_{ij}^{f}(b)$ and $\left\|\mathrm{diag}\left[\underline{\boldsymbol{\varphi}}_{ij}^{f}(b)\right]\right\|^{2}$ for the use as the normalization matrix. Theoretically, they are supposed to produce the same results at the Bussgang equilibrium point but experimentally, i.e., in the vicinity of the Bussgang equilibrium point, $\mathbf{\Lambda}_{ij}^{f}(b)$ is slightly superior to $\left\|\mathrm{diag}\left[\underline{\boldsymbol{\varphi}}_{ij}^{f}(b)\right]\right\|^{2}$ and that is why we selected $\mathbf{\Lambda}_{ij}^{f}(b)$.

3.2 Temporal Processing

The temporal noises such as engine and air friction noises usually have colored spectra and the spectra also change themselves in accordance with the car speed. Fortunately, the characteristics of the temporal noise spectra are rather slowly changing compared to the spatial noise and speech.

Firstly, the almost band limited noises can be removed with a simple second-order HPF, proposed for the GSM-EFR, with 80 Hz cutoff frequency [9].

$$H(z) = \frac{0.9273 - 1.8545z^{-1} - 0.9273z^{-2}}{1 - 1.9059z^{-1} + 0.9114z^{-2}} \tag{26}$$

With this filter, most components of the engine vibration and the air friction noises can be successfully reduced but still there are noises having above 80 Hz components.

The Kalman filter with a whitening process is adopted to cancel the remaining temporal noises. Whitening is achieved by using the autoregressive inverse model of $\boldsymbol{\rho}^{(n)}(z)$ estimated for the initial 100 msec of $u_i'(n)$, the HPFed version of $u_i(n)$. After Kalman filtering, the distortion of $\hat{u}_i(n)$ caused by the whitening process is compensated by $1/\boldsymbol{\rho}^{(n)}(z)$ and then the final enhanced speech $s(n)$ is obtained. The details of our temporal noise cancellation procedure can be found in [7].

3.3 Placement of Microphones

The CP problem is critical in temporal processes. It is insisted that the constraint $w_{ii,p} = \delta_{ij}\delta_p$ can solve this problem [2], but it is not hard to find the failures even under the constraint in real applications especially when the closely coupled microphones are used. When each sound is recorded on each microphone with the comparatively large energy, the problem can be avoided under the above constraint. In our experiment, we cope with this CP problem by simply locating two microphones separately, i.e., one in the front and the other in the rear section as shown in Fig. 1.

4. Experiments

The performance of our NFB-MBD is firstly checked and then the overall performance of our proposed algorithm is evaluated for the test materials containing both the temporal and the spatial noises.

4.1 Performance Evaluation of NFB-MBD

To demonstrate the usefulness of our proposed speech enhancement algorithm, the performance of our NFB-MBD is compared with those of the causal FB-MBD (CFB-MBD) and the nonstationary statistics-based MBD (NSB-MBD) in [5], [6]. The CFB-MBD is realized based on Eq. (23). The mixture is recorded with two omni-directional condenser microphones (Audio-Technica AT9500II). The distance between two microphones is 215 cm and the recorded signals are digitized at 16 kHz with 16 bit resolution. Our present goal is to evaluate only the NFB-MBD performance; the data are collected in a stand-still car.

The learning parameters of the NFB-MBD and the CFB-MBD are equal; the filter length is 128, the frame size for the FFT is 512 with 75% overlapping. Both algorithms are on-line operation ones and their step size μ_0 is 0.005. The step size adaptation for the robustness of the CFB-MBD can be achieved as follows [1],

$$\mu_i(b) = \frac{\mu_0}{\left(0.01 + \sum_{p} u_i(n-p) f\left(u_i(n-p)\right)\right)} \tag{27}$$

The NSB-MBD is an off-line multi-iteration algorithm with the same filter length but the frame size is 1024.

To verify the usefulness of our NFB-MBD, the performance is evaluated for the music-Korean speech mixture and the results are summarized in Fig. 2. As easily can be seen from the figure, channel 1 contains stronger speech than channel 2, and in case of the music signal, the situation is reversed. Listening tests confirm that our NFB-MBD algorithm separates the speech and the music with almost no quality degradation and outperforms the others. The causal MBD is also successful in signal separation but the results still suffer from the whitening effect. Surprisingly, the NSB-MBD fails to separate the signals.

4.2 Overall Performance Evaluation

To show the efficiency of our algorithm, the computation

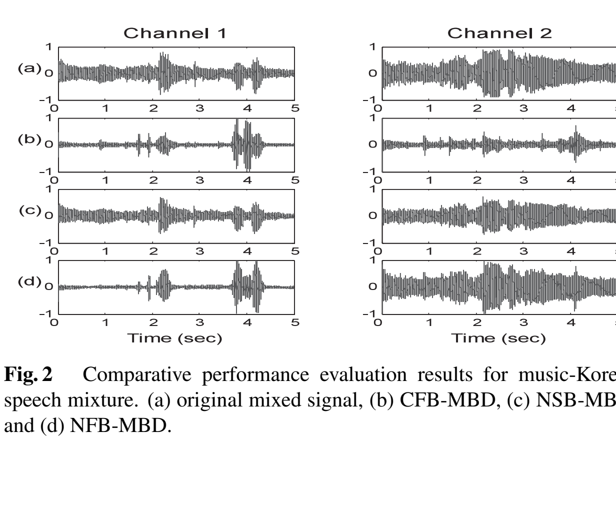

Fig. 2 Comparative performance evaluation results for music-Korean speech mixture. (a) original mixed signal, (b) CFB-MBD, (c) NSB-MBD, and (d) NFB-MBD.

Table 1 Comparison of computation loads between proposed and conventional method.

Operation	Proposed (times)	Conventional (times)
FFT/IFFT	24	28
Multiplication	131,790	425,984
Addition	75,815	292,966
Division	4,112	8,192

Table 2 Comparison of output SNR between proposed and conventional method.

Input SNR (dB)	Output SNR (dB)	
	Proposed	Conventional
- 7	2.56	-1.64
- 5	6.52	1.74
0	10.76	3.18
5	12.13	11.76
10	16.02	13.58

loads are compared with that of conventional method in [2] at first. Each computation load is counted till termination of one-output block process with equal size. From Table 1, it is clear that our method is much more efficient.

To evaluate the overall performance, input noisy speech contains both temporal and spatial noises. Temporal noises are recorded in a running car at various speeds with partly opened windows, and spatial ones are constructed by using different speech and a popular music from the one mentioned above. The parameter setting is identical to the NFB-MBD performance evaluation case. Final results in terms of the segment signal-to-noise ratio (SNR) are summarized in Table 2 [7]. It is clear that our method is superior to the conventional one in all cases.

5. Conclusion

A new speech enhancement algorithm using the NFB-MBD and the temporal process is presented for the use in car noise environments. Our algorithm is proved to reduce not only the background audio signal but also the typical temporal car-noises successfully. The NFB-MBD is introduced to deal with the background audio from the multiple sources, and its performance is shown to be better than those of the causal MBD and the nonstationary statistics-based MBD. The HPF and the Kalman filter procedure with the noise whitening also work fairly well as the temporal noise reduction process. In addition, the overall structure and the computation of our algorithm are relatively simple and efficient compared with the previous approaches. Our final results show remarkably enhanced speech waveforms and it is believed that we can recommend the use of our proposed algorithm for the car-environment applications.

Our future works include more extended evaluation of the proposed algorithm in conjunction with a speech recognizer and more detailed convergence property analysis on the NFB-MBD. In addition, the performance analysis with more versatile types of noises would be included for more practical and reliable application of our proposed algorithm.

Acknowledgments

Seung H. Nam was supported by grant No. R05-2004-000-10290-0 from Ministry of Science & Technology.

References

[1] S. Amari, S.C. Douglas, A. Cichocki, and H.H. Yang, "Novel online adaptive learning algorithms for blind deconvolution using the natural gradient approach," Proc. IEEE 11th IFAC Symposium on System Identification, SYSID-97, pp.1057–1062, 1997.
[2] E. Visser, M. Otsuka, and T.W. Lee, "A spatio-temporal speech enhancement scheme for robust speech recognition in noisy environments," Speech Commun., vol.41, pp.393–407, 2003.
[3] H. Saruwatari, K. Sawai, A. Lee, K. Shikano, A. Kaminuma, and M. Sakata, "Speech enhancement and recognition in car environment using blind source separation and subband elimination processing," Proc. Int. Workshop on Independent Component Analysis and Signal Separation, pp.367–372, 2003.
[4] M. Joho and P. Schniter, "Frequency domain realization of a multichannel blind deconvolution algorithm based on the natural gradient," Proc. Int. Workshop on Independent Component Analysis and Signal Separation, pp.543–548, 2003.
[5] 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, pp.303–312, 1999.
[6] L. Parra and C. Spence, "Convolutive blind separation of nonstationary sources," IEEE Trans. Speech Audio Process., vol.8, no.3, pp.320–327, May 2000.
[7] S. Jeong and M. Hahn, "Speech quality and recognition rate improvement in car noise environments," Electron. Lett., vol.37, no.12, pp.800–801, June 2001.
[8] K. Torkkola, "Blind separation of convolved sources based on information maximization," Proc. IEEE Workshop on Neural Networks and Signal Processing, pp.423–432, Sept. 1996.
[9] ETS 300 700, "Digital cellular telecommunications system: EFR speech transcoding," EFR Speech Transcoding (GSM 06.06)', ETSI, p.19, 1997.
[10] B.F. Boroujeny and K.S. Chan, "Analysis of the frequency domain block LMS algorithm," IEEE Trans. Signal Process., vol.48, no.8, pp.2332–2342, Aug. 2000.
[11] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," IEEE Trans. Speech Audio Process., vol.11, no.2, pp.109–116, March 2003.
[12] S. Haykin, Adaptive Filter Theory, 4th ed., Prentice-Hall, 2002.