THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA VOL.25, NO.4E 2006. 12 pp. 142~147

Speech Enhancement Using Blind Signal Separation Combined With Null Beamforming

Seung-Hyon Nam*, Rodrigo C. Muñoz, Jr.**
*Department of Electronic Engineering, Paichai University.
**Institute of Engineering and Architecture, Bataan Polytechnic State College, Philippines
(Received February 22 2006; Revised November 23 2006; accepted November 25 2006)

Abstract

Blind signal separation is known as a powerful tool for enhancing noisy speech in many real world environments. In this paper, it is demonstrated that the performance of blind signal separation can be further improved by combining with a null beamformer (NBF). Cascading the blind source separation with null beamforming is equivalent to the decomposition of the received signals into the direct parts and reverberant parts. Investigation of beam patterns of the null beamformer and blind signal separation reveals that directional null of NBF reduces mainly direct parts of the unwanted signals whereas blind signal separation reduces reverberant parts. Further, it is shown that the decomposition of received signals can be exploited to solve the local stability problem. Therefore, faster and improved separation can be obtained by removing the direct parts first by null beamforming. Simulation results using real office recordings confirm the expectation.

Keywords: *Blind source separation, Multichannel blind deconvolution, Null beamforming, Speech enhancement.*

I. Introduction

Enhancement of noisy speech in real world environments has been studied for a long time. It can be applied to most speech communication systems including automatic speech recognition, speech compression, and acoustic echo cancellation.

Speech enhancement using multiple microphones can be categorized into beamforming and blind signal separation (BSS). In beamforming, signals are separated using the directional information of the signals with respect to the microphones. On the other hand, BSS does not use any information on the signals and assumes that source signals are statistically independent each other. In general, BSS provides better performance over the beamforming techniques in terms of the number of microphones, speech quality (distortion), and interference reduction.

In case of blind speech separation in a noisy room, the received signal at a microphone would be a reverberant and mixed version of sound sources so that it is modeled as convolutive mixing. Restoration of the sound sources then can be accomplished by estimating the unmixing (separating) filters that deconvolve the room impulse responses.

Multichannel blind deconvoltion (MBD) is one practical method to estimate the unmixing filters [1]. Existing MBD algorithms, however, suffer from whitening effect of output speech, slow convergence speed, and poor separation. Recently, the frequency-domain normalized MBD (FNMBD) algorithm has been proposed to overcome the shortcomings of existing MBD algorithms [2]. It employs right-sided unmixing filters instead of double-sided filters and operates on the signals with normalized spectrum. It also has been successfully applied to speech enhancement in real world noisy environments [3]. Further, its properties and modifications are presented in [4].

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

Performance of BSS algorithms can be improved further by utilizing directional information of the source signals through beamforming. Several approaches have been reported [5-8]. In [5], the BSS algorithm is embedded into the adaptive beamformer to compensate the poor performance of the blocking matrix. In [6, 7], null beamforming (NBF) is utilized to resolve the frequency permutation problem inherited in the frequency-domain BSS algorithms. In [8], beamformer filters are utilized as initial values for the unmixing filters of BSS.

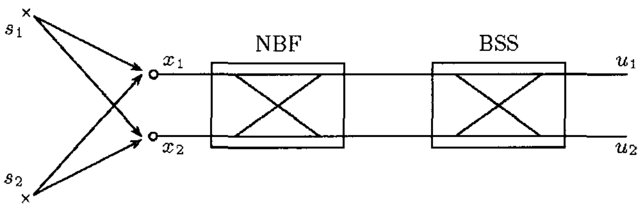

Fig. 1. BSS with a NBF front-end.

In this paper, however, beamforming is utilized as a front-end cascaded with BSS to remove direct parts of interfering signals as shown in Fig. 1. It is assumed that directional information is available from the various existing methods. In this way, the received signals are effectively decomposed into direct and reverberant parts using beamforming and BSS. This decomposition provides significant performance improvement as well as solves the stability problem discussed in [9].

For this purpose, NBF is chosen among various beamforming techniques for its capability of making sharp directional nulls toward interfering signals with only two microphones.

This paper is organized as follows: first, a newly proposed MBD (called FNMBD [4]) is described briefly. Next, geometry and filters for NBF is presented with its implications. Then beam pattern of NBF, FNMBD, and FNMBD when combined with NBF are investigated and discussed. Finally, simulation results in real world environment are presented and discussed.

II. THE FNMBD ALGORITHM FOR BSS

In convolutive mixing with P sources and Q sensors, the mixed signal at time k 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 $\mathbf{W}(z,k)=\sum_{p=0}^{L-1}\mathbf{W}_p(k)z^{-p}$ where $\mathbf{W}_p(k)$ is a $P\times Q$ matrix. Then the i^{th} separated signal is given by

$$u_i(k)=\sum_{j=1}^{Q}\sum_{p=0}^{L-1} W_{ij,p}(k)x_j(k-p) \tag{2}$$

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.

Now, assume that the unmixing process is performed in block mode and define $\mathbf{W}^f(b)$ be a $P\times Q$ matrix at frequency bin f obtained by applying the Fourier transform to the unmixing filter $\mathbf{W}(z,b)$ at block time b. The FNMBD algorithm adopts normalization of the separated signals in the frequency domain [4] and can be expressed in a simplified vector form as

$$\Delta\mathbf{W}^f(b)=\left\{\bar{\mathbf{I}}-\Lambda_y^{-1/2}(b)\mathbf{y}^f(b)(\mathbf{u}^f(b))^H\Lambda_u^{-1/2}(b)\right\}\mathbf{W}^f(b). \tag{3}$$

Here, the superscript f the quantity in the frequency domain, H the hermitian transposition. In addition, $\bar{\mathrm{I}}=diag(\bar{1},\cdots,\bar{1})$ where $\bar{1}=(1,\cdots,1)$. The separated signal vector $\mathbf{u}^f(b)$ is transformed into $\mathbf{y}^f(b)$ in the time domain using $\mathbf{y}(k)=\mathbf{f}(\mathbf{u}(k))$ wher e $f_i(u_i)=\frac{d}{du_i}\log p_i(u_i)$. Both $\mathbf{y}^f(b)$ and $\mathbf{u}^f(b)$ are then normalized using the diagonal matrices $\Lambda_y(b)$ and $\Lambda_u(b)$ whose diagonal elements $\mathbf{P}_{y_i}(b)$ and $\mathbf{P}_{u_i}(b)$, respectively, are updated at each block time by

$$\mathbf{P}_{y_i}(b)=(1-\gamma)\mathbf{P}_{y_i}(b-1)+\gamma\left|\mathbf{y}_i^f(b)\right|^2 \tag{4a}$$

$$\mathbf{P}_{u_i}(b)=(1-\gamma)\mathbf{P}_{u_i}(b-1)+\gamma\left|\mathbf{u}_i^f(b)\right|^2 \tag{4b}$$

where $0<\gamma<1$. The detail description of the FNMBD algorithm can be found in [4]. Notice that equilibrium points of (3) are given by, at each frequency bin,

$$E\left\{\frac{y_i^f(b)\left(u_j^f(b)\right)^*}{\sqrt{\left|y_i^f(b)\right|^2\left|u_j^f(b)\right|^2}}\right\}=1$$

which do not impose any constraint on the spectrum so that the whitening problem is clearly avoided. In addition, due to spectral normalization employed in (4), the spectral tilt is effectively compensated and the stability and separation performance of the algorithm is greatly enhanced.

III. FNMBD WITH NBF

There are several types of beamformers varying from adaptive and non-adaptive form. In this paper, the conventional null-beamformer (NBF) is presented. A significant characteristic of the NBF is a steep null toward the interfering signal. It greatly supplements the performance of BSS when connected in cascade.

3.1. Geometry of Beamforming

The geometry of a beamformer is shown in Fig. 2. The locations of the sources are known a priori. Angles are measured from the normal line with respect to the microphone array and the clockwise rotation is denoted as positive. If we consider only the direct part, time lag of source s_l at the sensor at d_k with respect to the phase center is given by

$$\tau_{kl}=\frac{d_k\sin\theta_l}{c} \tag{5}$$

where c is the sound velocity.

In terms of BSS, Fig. 2(a) can be treated with only right-sided causal filters. The FNMBD algorithm with right-sided filters is very efficient for this setup. On the other hand, Fig. 2(b) requires double-sided noncausal filters as well as right-sided causal filters due to time lags (5) as pointed out in [9]. was pointed out in [9] that the algorithms with right-sided filters are more robust than those with double-sided filters. This robustness is

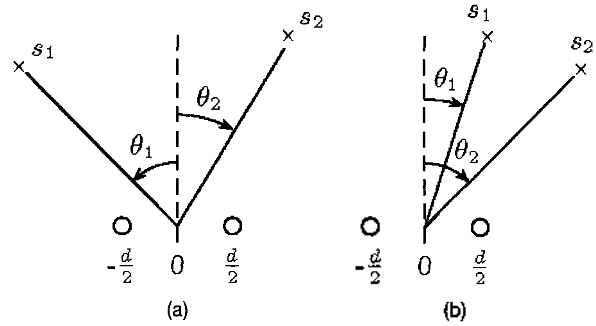

Fig. 2. Geometries of a linear beamformer.

explained further in the next section in terms of local stability.

3.2. Local Stability of BSS and NBF Filters

This requirement prevents the use of efficient normalized MBD algorithm with right-sided filters presented in [4]. However, these time lags can be effectively equalized by NBF, where the coefficients are computed in the frequency domain as

$$W_{kl}^{NBF}=\left(-1\right)^{k+1}\exp\left(-j2\pi f\tau_{kl}\right). \tag{6}$$

These NBF coefficients are then converted into the time domain and delayed to get double-sided noncausal filters $w_{kl}^{NBF}(t)$, which are sinc functions since time lags are not-integers in general. After processing the sensor signals using NBF filters, the direct parts of interfering signals are effectively removed and Signal-to- Interference Ratio (SIR) becomes approximately 6-8dB. What remain after NBF are reverberant parts which do not cause any causality problem since time lags are equalized by NBF. The NBF output can be further separated using the FNMBD algorithm with casual filters.

It is noted that only few tenth of filter coefficients are required to compensate the time lags (5) for most speech applications. For example, NBF filters of length 32-64 are enough to cover the time lags for speech sampled at 8kHz rate with a two-microphone array ($d=4$cm). If we use a BSS algorithm with double-sided filters, on the other hand, center coefficients of diagonal filters are initialized with 1 or NBF filters may be used as initial values [8]. Since unmixing filters should be long enough for enough separation, anticausal (left-sided) parts are wasted. Moreover, these anticausal parts can produce pre echo that deteriorates speech quality [10].

As pointed out in [9], it can be shown that, after some manipulation, right-sided filters are sufficient for the geometry in Fig. 2(a) whereas one double-sided filter is necessary for Fig. 2(b). In [9], the necessity of the double-sided filter has been treated as "the causality problem" and the BSS algorithm with double-sided filters has been proposed in [9] for geometry in Fig. 2(b). However, the requirement on the double-sided filter is deduced from the assumption that the mixed sensor signals contain only the direct part of the original source signals which is clearly not true. The FNMBD algorithm with double-sided filters would be very inefficient to separate the reverberant parts.

These inefficiency can be explained further using the local stability conditions derived in [11], which is

$$1-A_{12,0}A_{21,0}>0 \tag{7}$$

for causal filters and

$$1-\sum_p\left|A_{12,p}\right|-\sum_p\left|A_{21,p}\right|>0 \tag{8}$$

for noncausal filters (in a case of 2 sensors 2 sources). From condition (7), we see that the local stability of the BSS algorithm depends on the direct parts of the received signal. If microphones are placed closely enough so that the delay between the microphones is small enough, then condition (7) would not be easily satisfied and the BSS algorithm would suffer from the local instability. On the other hand, we can see from condition (8) that the BSS algorithm with noncausal filters would suffer from the local instability regardless microphone spacing since (8) is hardly satisfied. This explains the low separation performance of the BSS algorithm with noncausal filters.

This problem can be effectively solved by cascading the NBF and FNMBD with right sided filters together. By processing the mixed signals using NBF filters, the direct parts of interfering signals are effectively removed. Stability conditions (7) and (8) are then easily satisfied. What remain after NBF are reverberant parts which do not cause any causality problem since time lags are already equalized by NBF. The NBF output can be separated further using right-sided filters.

IV. SIMULATIONS

4.1. Simulation Setup

Speech sounds of a male voice located at $-30°$ and a female voice located at $40°$, recorded in a normal office, are used as the two mixing sources. The mixing signals, approximately 120cm from the array, are applied to two. microphones separated at 4cm. The mixed signals are recorded at 8kHz sampling rate. The recording is arranged to have 3 segments of data with 10 seconds length per segment. The first segment is used for learning and the remaining segments are used for SIR calculations. The last two segments consist of two 10 seconds segment where only one source is active at each segment. Signal-to-Interference ratio (SIR) is used as a performance measure. The SIR is defined as the ratio of the signal power of the target signal from that of the interfering signal.

4.2. Performance of FNMBD Combined with NBF

Beam pattern plays an important role in the experimentation and analysis and is expressed as

$$F_l(f,\theta)=\sum_{k=1}^{Q}W_{lk}(f)\exp\left(j2\pi f d_k \sin\theta/c\right) \tag{9}$$

for given unmixing filters $W_{lk}(f)$.

In Fig. 3, beam patterns of NBF, FNMBD, and FNMBD with the NBF front-end are shown. The beam pattern of NMBD with the NBF front-end is obtained for the system in Fig. 1. It is clear that NBF provides a deep null at the specified DOA of the interfering signal. This reveals the fact that NBF removes the direct parts of the interfering signals effectively. On the other hand, FNMBD provides null over the wide range of DOA which implies the fact that BSS removes reverberant parts as well as direct parts of the interfering signals. By directly cascading NBF into FNMBD, the resulting beam pattern of FNMBD shows clear null toward interfering signals as well as reverberant parts.

Figure 4(a) shows the results for a casual case where the speech sources are at $-30°$ and $40°$ to the normal direction. When double-sided filters are initialized with NBF filters of length 1024 as in [8], it starts from SIR of

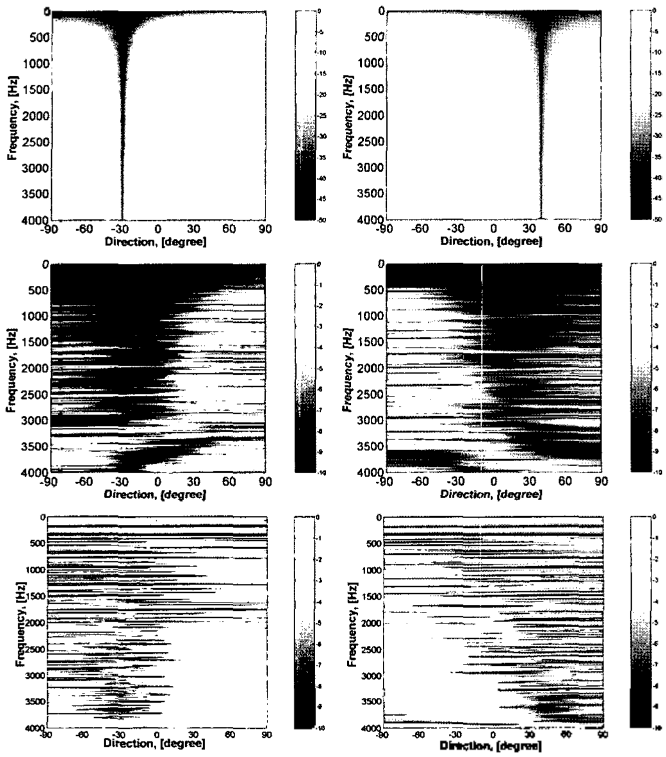

Fig. 3. Beam patterns of NBF, FNMBD, and FNBMD after NBF.

7.5dB and settles down after 2dB improvement. On the other hand, the algorithm with right-sided filters gives SIR of 9.5dB. The algorithm with right-sided filters provides additional 4.5dB when it is cascaded with NBF of length 64. For the noncausal case where sources are at 40° and 0°, the algorithm with right-sided filters performs poorly as shown in Fig. 4(b). The NBF provides SIR of 6dB for this case. The algorithm with double-sided filters initialized with NBF coefficients again improves SIR about 2dB. However, the algorithm with right-sided filters provides additional 5dB when it is cascaded with NBF. These results demonstrate that NBF, when it is combined with a BSS algorithm with right-sided filters, not only improves the separation performance but also solves the causality problem.

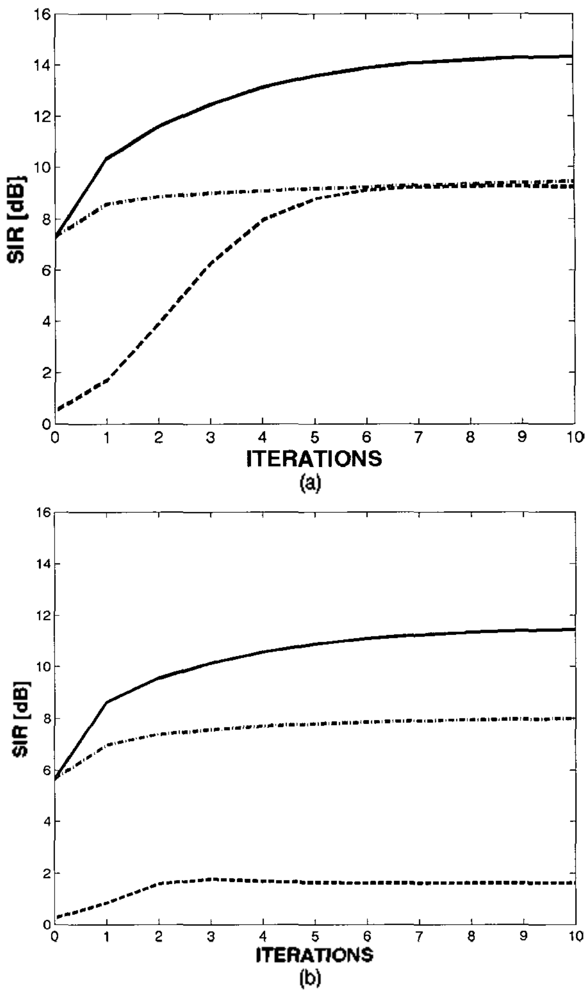

Fig. 4. SIR Improvement of NBF+FNMBD over FNMBD for (a) the causal case (-30° and +40°) and (b) the noncausal case (40° and +0°): right-sided filter (dashed), double-sided filters initialized with NBF (dotted), and right-sided filters with NBF front end (solid).

V. CONCLUSIONS

The performance of FNMBD cascaded with NBF is investigated. It is demonstrated the FNMBD combined with the NBF front-end is equivalent to the decomposition of received signals into direct and reverberant parts. This decomposition not only improves the separation performance of the FNMBD algorithm but also solves the causality problem. Simulation results using a real-world recording confirm the expectation.

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. S. H. Nam and S. Beack, "A frequency-domain nomalized multichannel deconvolutive algorithm for acoustical signals," Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA'04), Granada, Spain, 522-529, 2004.
3. 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.
4. Seung-Hyon Nam, "A new formulation of multichannel blind deconvolution and its properties for speech enhancement," Proc. of Western Pacific Acoustics Conf., Seoul, Korea, June 2006.
5. C. Fancourt and L. Parra, "The generalized sidelobe decorrelator," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 167-170, Oct. 2001.
6. H. Saruwatari, T. Kawamura, K. Sawai, A. Kaminuma, M. Sakata, "Blind source separation based on fast-convergence algorithm using ICA and beamforming for real convolutive mixture," Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 3097-3100, May 2002.
7. H. Saruwatari, S. Kurita, K. Takeda, F. Itakura, and K. Shikano, "Blind source separation based on subband ICA and beamforming," Proc. ICSLP2000, 3 99-97, Oct. 2000.
8. R. Aichner, S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Time-domain blind source separation of nonstationary convolutive signals by utilizing geometry beamforming," Proc. IEEE Workshop on NNSP' 02, 445-454, 2002.
9. R. Aichner, H. Buchner, and W. Kellermann, "On the causality problem in the time-domain blind source separation and deconvolution algorithms," Proc. ICASSP, 2005.
10. R. Mukai, S. Araki, and S. Makino, "Separation and dereverberation performance of frequency domain blind source separation," Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation, 230-235, Dec. 2001.
11. J. Bourgeois, "An LMS viewpoint on the local stability of second order blind source separation," in Proc. IEEE Workshop on Statistical Signal Processing, Bordeaux, France, 1995, 49-53.

[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

• **Rodrigo C. Muñoz, Jr.**

1998 : BS, Dept. of Electrical and Computer Engineering, Saint Louis University, Pilippines
2004 : MS, Dept. of Electrical and Computer Engineering, De La Salle University, Manila, Philippines
2006 : MS, Dept. of Electronic Engineering, Paichai University
2006 : Dean, Institute of Engineering and Architecture, Bataan Polytechnic State College, Pilippines
Interested Areas : Speech signal processing, Speech enhancement, Beamforming.