Convolutive Blind Source Separation Based on Disjointness Maximization of Subband Signals

Tiemin Mei and Alfred Mertins, *Senior Member, IEEE*

***Abstract*—The concept of disjoint component analysis (DCA) is based on the fact that different speech or audio signals are typically more disjoint than mixtures of them. This letter studies the problem of blind separation of convolutive mixtures through the subband-wise maximization of the disjointness of time-frequency representations of the signals. In our approach, we first define a frequency-dependent measure representing the closeness to disjointness of a group of subband signals. Then, this frequency-dependent measure is integrated to form an objective function that only depends on the time-domain parameters of the separation system. Lastly, an efficient natural-gradient-based learning rule is developed for the update of the separation-system coefficients.**

***Index Terms*—Convolutive blind source separation, disjointness maximization.**

I. Introduction

BLIND source separation (BSS) has been an active research topic during the past decade due to its potential applications in many areas. As a special case, separation of instantaneous mixtures is very successful so far and many approaches have been proposed [1]–[5]. However, a more challenging situation is the separation of convolutive mixtures with long mixing channels [6]–[10].

A general way for solving the convolutive BSS problem is to extend the approaches for instantaneous mixtures to the case of convolutive mixtures, which can be done in either the time or frequency domain. An advantage of the time-domain approaches is that they usually do not suffer from the so-called unknown permutation problem [6]. Frequency-domain approaches are considered as promising techniques for BSS in cases of very long mixing channels. It is known that convolutive mixtures in the time domain can be considered as instantaneous mixtures in the frequency domain, so approaches for instantaneous mixture separation can be applied to frequency-domain representations of convolutive mixtures. However, the permutation ambiguity, which is inherited from instantaneous BSS, makes convolutive BSS very difficult [7], [8].

Extensive work has been done to remedy the permutation problem. One way is to identify the permutation based on signal and/or BSS-system properties [7], [8]. Other approaches cope with the problem by trying to avoid permutations rather than identifying them. This is, for example, done by building objective functions in the frequency domain that keep the advantages of the frequency-domain approaches while capturing the optimizing parameters in the time domain [13], [14]. Because the frequency-domain objective function is integrated in order to yield the time-domain filter coefficients, the method is referred to as the frequency-domain integrated objective approach. One of the advantages of this technique is the high computational efficiency due to the use of the fast Fourier transform. Another one is the ability to overcome the local permutation problem. This is achieved by constraining the separation filters to a limited length with a window during the implementation, which is equivalent to imposing smoothness constraints on the separation filters in the frequency domain.

In [11], the author proposed a BSS method for instantaneous mixtures based on the maximization of the disjointness of separated signals. In this setup, disjointness of signals means that the signals have periods of activity and periods of inactivity and that the active periods of different signals are distributed in a disjoint manner along the time axis. For time-domain speech signals of competing speakers, at least a partial disjointness is present, as there are usually periods of silence in a speech signal. A higher degree of disjointness is often achieved in the time-frequency domain, in which speech signals may even be disjoint orthogonal, a property which is also known as W-disjoint orthogonality [16], [17]. The reason for the disjointness in the time-frequency domain is that speech has a very sparse representation in this domain [18].

In this letter, we combine the DCA principle from [11] with the concept of an integrated frequency-domain objective function from [13]. The DCA is carried out in subbands in the time-frequency domain. The result is a novel approach for blind source separation of convolutive mixtures with state-of-the-art separation performance.

II. Problem Statement

In this letter, we only consider the N-by-N cases. The mixing channels are assumed to be FIR of length L, and the separation filters are also FIR with length $M \geq (N-1)(L-1)+1$ [9]. We assume that the sources are real and at least partly disjoint to each other in the time-frequency domain. Furthermore, we assume that the mixing system is linear and time invariant. We use $\mathbf{s}(n)$, $\mathbf{x}(n)$ and $\mathbf{y}(n)$ to denote the sources, the mixtures, and the separated outputs, respectively.

The noise-free convolutive mixing model is given as

$$\mathbf{x}(n) = \boldsymbol{A}(n) * \mathbf{s}(n) = \sum_{l=0}^{L-1} \boldsymbol{A}(l)\mathbf{s}(n-l) \tag{1}$$

where $\boldsymbol{A}(n) = [a_{ij}(n)]_{N\times N}$ is the FIR-filter mixing matrix. We assume that the transfer function matrix of the mixing system, $\mathbf{A}(z) = \sum_{n=0}^{L-1} \boldsymbol{A}(n) z^{-n}$, is nonsingular on the unit circle of the

Manuscript received December 19, 2007; revised April 22, 2008. This work was supported by the German Research Foundation under Grant No. ME1170/1. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Richard J. Kozick.

T. Mei is with the Institute for Signal Processing, University of Lübeck, Lübeck 23538, Germany, on leave from the School of Information science and Engineering, Shenyang Ligong University of Technology, Shenyang 110168, China (e-mail: mei@isip.uni-luebeck.de; meitiemin@163.com).

A. Mertins is with the Institute for Signal Processing, University of Lübeck, Lübeck 23538, Germany (e-mail: mertins@isip.uni-luebeck.de).

Digital Object Identifier 10.1109/LSP.2008.2001114

complex plane, which guarantees that the sources are separable. The separation-system output is given as

$$\mathbf{y}(n)=\boldsymbol{H}(n)*\mathbf{x}(n)=\sum_{l=0}^{M-1}\boldsymbol{H}(l)\mathbf{x}(n-l) \tag{2}$$

where $\boldsymbol{H}(n)=[h_{ij}(n)]_{N\times N}$ is the separation system. From (1) and (2), we have $\mathbf{y}(n)=\boldsymbol{G}(n)*\mathbf{s}(n)$ with $\boldsymbol{G}(n)=\boldsymbol{H}(n)*\boldsymbol{A}(n)$ being the global system. This can be rewritten in the z-domain as $\mathbf{Y}(z)=\mathbf{G}(z)\mathbf{S}(z)$ with $\mathbf{G}(z)=\mathbf{H}(z)\mathbf{A}(z)$. BSS is considered to be successful if the components of the output vector $\mathbf{y}(n)$ are permuted and filtered versions of the signal sources in $\mathbf{s}(n)$, which means that $\mathbf{G}(z)=\mathbf{PD}(z)$, where $\mathbf{P}$ is a permutation matrix and $\mathbf{D}(z)$ is a diagonal transfer function matrix. The transfer functions in $\mathbf{D}(z)$ lead to a coloration of sources. However, such effects are not considered in this letter. They can, for example, be reduced through postfiltering based on the Minimal Distortion Principle [21].

III. BSS Based on the Maximization of Disjointness of Subband Signals

In the time-frequency domain, the observed signals are decomposed into a set of narrowband components via the short-time Fourier transform (STFT), and the separation criterion is defined for each frequency bin, that is, for each subband of the time-frequency representation. The separation process can be described by the equation

$$\mathbf{Y}(l,e^{j\omega})=\mathbf{H}(e^{j\omega})\mathbf{X}(l,e^{j\omega}) \tag{3}$$

where l is the time index and

$$\mathbf{Y}(l,e^{j\omega})=\left[y_1(l,e^{j\omega}),y_2(l,e^{j\omega}),\ldots,y_N(l,e^{j\omega})\right]^{\mathrm{T}} \tag{4}$$

$$\mathbf{X}(l,e^{j\omega})=\left[x_1(l,e^{j\omega}),x_2(l,e^{j\omega}),\ldots,x_N(l,e^{j\omega})\right]^{\mathrm{T}}. \tag{5}$$

For a mixture of disjoint signals, the maximization of the disjointness of the outputs of the separation system will lead to the separation of the sources. In a time-domain setup, this is equivalent to the minimization of the "overlap" between the outputs of the separation system [11]. The "overlap" is defined as

$$O_{ij}=\mathrm{E}\left[|y_i||y_j|\right] \tag{6}$$

where E[.] denotes the expectation and $i,j=1,2,\ldots,N$. O_{ij} is essentially the cross-correlation between the signals $|y_i|$ and $|y_j|$. For audio signals containing pauses, we have $|y_i|\geq 0$ during their active periods and $|y_i|=0$ during their silent periods. So, if two signals y_i and y_j are exactly disjoint to each other, the overlap O_{ij} will be zero. Of course, in practice, there will typically be some overlap.

In the following, we exploit the DCA idea to define an optimality criterion for each subband of the time-frequency representation. For this purpose, we define a vector of the absolute values of subband signals as

$$\widetilde{\mathbf{Y}}(l,e^{j\omega})=\left[\left|y_1(l,e^{j\omega})\right|,\left|y_2(l,e^{j\omega})\right|,\ldots,\left|y_N(l,e^{j\omega})\right|\right]^{\mathrm{T}} \tag{7}$$

and the correlation matrix of vector $\widetilde{\mathbf{Y}}(l,e^{j\omega})$ as

$$\mathbf{P}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}}(l,\omega)=\mathrm{E}\left[\widetilde{\mathbf{Y}}(l,e^{j\omega})\widetilde{\mathbf{Y}}^{\mathrm{T}}(l,e^{j\omega})\right]. \tag{8}$$

The off-diagonal entries of $\mathbf{P}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}}(l,\omega)$ contain the overlaps $O_{ij}=\mathrm{E}[|y_i(l,e^{j\omega})||y_j(l,e^{j\omega})|]$, which means that the diagonalization of $\mathbf{P}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}}(l,\omega)$ is equivalent to the minimization of the overlaps between all pairs of subband signals. Therefore, according to [11], the diagonalization of $\mathbf{P}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}}(l,\omega)$ will result in the separation of sources.

In accordance with the frequency-domain integrated objective function in [13] and similar second-order statistics-based approaches found in [14] and [15], we define the following frequency-domain integrated objective function that will be optimized according to the time-domain parameters of the separation system:

$$f\left(l,\boldsymbol{H}(n)|_{n=0,1,\ldots,M-1}\right)=\frac{1}{2}\int_{-\pi}^{\pi}\frac{\det\left[\mathbf{D}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}}(l,\omega)\right]}{\det\left[\mathbf{P}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}}(l,\omega)\right]}\mathrm{d}\omega \tag{9}$$

where $\mathbf{D}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}}(l,\omega)=\mathrm{diag}[\mathbf{P}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}}(l,\omega)]$. The operator diag[.] sets the off-diagonal elements of a square matrix to zero, and it turns a vector into a diagonal matrix.

The derivation according to [13] and [19] of the natural-gradient based learning rule for this optimization problem yields

$$\boldsymbol{H}^{l+1}(n)=\boldsymbol{H}^{l}(n)-\mu\int_{-\pi}^{\pi}\mathrm{E}\left[\mathbf{B}\left(\mathbf{D}^{-1}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}}(l,\omega)-\mathbf{P}^{-1}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}}(l,\omega)\right)\widetilde{\mathbf{Y}}\mathbf{Y}^{\mathrm{H}}\right]\mathbf{H}^{l}(e^{j\omega})e^{j\omega n}d\omega \tag{10}$$

where

$$\mathbf{B}=\mathrm{diag}([(y_1(l,e^{j\omega})/|y_1(l,e^{j\omega})|),$$

$(y_2(l,e^{j\omega})/|y_2(l,e^{j\omega})|),\ldots,(y_N(l,e^{j\omega})/|y_N(l,e^{j\omega})|)]^{\mathrm{T}})$ is a diagonal matrix which is made up of the phases of the subband signals.

In practice, the statistical expectation is replaced with time-averaging. For implementation reasons, we do not want the time-averaged term $\mathbf{D}^{-1}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}}(l,\omega)-\mathbf{P}^{-1}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}}(l,\omega)$ in (10) to be involved in a second time-averaging, so we rewrite (10) in the following form:

$$\boldsymbol{H}^{l+1}(n)=\boldsymbol{H}^{l}(n)-\mu\int_{-\pi}^{\pi}\left(\sum_{k=1}^{N}\widetilde{\mathbf{a}}_k\mathrm{E}\left[\mathbf{B}\mathrm{L}(N)\mathrm{diag}[\mathbf{Y}^*]\times\left|y_k(l,e^{j\omega})\right|\right]\right)\mathbf{H}^{l}(e^{j\omega})e^{j\omega n}d\omega \tag{11}$$

where $\widetilde{\mathbf{a}}_k$ $(k=1,2,\ldots,N)$ is a diagonal matrix whose diagonal entries are the kth column of the matrix $\mathbf{D}^{-1}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}}(l,\omega)-\mathbf{P}^{-1}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}}(l,\omega)$, $\mathrm{L}(N)$ is an N-by-N matrix whose entries are ones, and the superscript * is the conjugation operator.

For comparison purposes, we also define the zero-mean amplitude-modulation signals [12]

$$\widehat{y}_i(l,e^{j\omega})=\left|y_i(l,e^{j\omega})\right|-\mathrm{E}\left[\left|y_i(l,e^{j\omega})\right|\right] \tag{12}$$

and derive a corresponding learning rule based on the matrices

$$\mathbf{P}_{\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}}(l,\omega)=\mathrm{E}[\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}^{\mathrm{T}}] \tag{13}$$

and

$$\mathbf{D}_{\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}}(l,\omega)=\mathrm{diag}\left[\mathbf{P}_{\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}}(l,\omega)\right] \tag{14}$$

where $\widehat{\mathbf{Y}} = [\widehat{y}_1(l, e^{j\omega}), \widehat{y}_2(l, e^{j\omega}), \ldots, \widehat{y}_N(l, e^{j\omega})]^{\mathrm{T}}$. The learning rules corresponding to (15) and its implementation version (16) are

$$\boldsymbol{H}^{l+1}(n) = \boldsymbol{H}^{l}(n) - \mu \int_{-\pi}^{\pi} \mathrm{E}\left[\mathbf{B}\left(\mathbf{D}^{-1}_{\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}}(l,\omega) - \mathbf{P}^{-1}_{\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}}(l,\omega)\right)\widehat{\mathbf{Y}}\mathbf{Y}^{\mathrm{H}}\right]\mathbf{H}^{l}(e^{j\omega})e^{j\omega n}d\omega \quad (15)$$

$$\boldsymbol{H}^{l+1}(n) = \boldsymbol{H}^{l}(n) - \mu \int_{-\pi}^{\pi} \left(\sum_{k=1}^{N} \widehat{\mathbf{a}}_k \mathrm{E}\left[\mathbf{B}\mathrm{L}(N)\mathrm{diag}[\mathbf{Y}^{*}] \times \widehat{y}_k(l, e^{j\omega})\right]\right)\mathbf{H}^{l}(e^{j\omega})e^{j\omega n}d\omega \quad (16)$$

where $\mathbf{B}$ is the same as that defined in (10), but $\widehat{\mathbf{a}}_k$ $(k = 1, 2, \ldots, N)$ is a diagonal matrix whose diagonal entries are the kth column of the matrix $\mathbf{D}^{-1}_{\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}}(l,\omega) - \mathbf{P}^{-1}_{\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}}(l,\omega)$.

Both algorithms (10) and (15) are essentially decorrelation approaches. The only difference between them is rooted in the choice of signals that are to be decorrelated. While the overlap $O_{ij} = E[|y_i(l, e^{j\omega})|\ |y_j(l, e^{j\omega})|]$ will directly measure the disjointness of two signals $y_i(l, e^{j\omega})$ and $y_j(l, e^{j\omega})$, the correlation term $\mathrm{E}[\widehat{y}_i(l, e^{j\omega})\widehat{y}_j(l, e^{j\omega})]$ cannot do this. As we will see in simulations, the algorithm (15) is not as good as the disjointness maximization algorithm (10).

IV. Simulations

A. Simulation Results for the New Algorithm

Two 4-s long speech signals were selected from the TIMIT database [20] and used in this simulation. The selected files are named TEST\DR1\FAKS0\SA1.wav and TEST\DR1\FELC0\SI1386.wav. The sampling frequency was 16 kHz. For illustration purposes, the signals are depicted in Fig. 1(a). As one can see from the plots, the time-domain waveforms have a partial disjointness property. The mixing channel impulse responses have been measured in a regular laboratory room of size $6.9\ \mathrm{m} \times 5.0\ \mathrm{m} \times 3.0\ \mathrm{m}$. Two loudspeakers were placed at a distance of 1.25 m from each other, and two microphones were placed at a distance of 1 m from the axis connecting the loudspeakers. The distance between the microphones was 35 cm. The mixing channel impulse responses were identified to a length of 4096 taps at the sampling frequency of 16 kHz. They are depicted in Fig. 2.

In the following, results will be presented with respect to signal-to-interference ratios (SIRs), which were calculated with the following formulas for both the mixtures and the separated channels: $\mathrm{SIR}_1 = 10\log_{10}(p_{s_1}/p_{s_2})$ if the channel is considered to contain source 1, and $\mathrm{SIR}_2 = 10\log_{10}(p_{s_2}/p_{s_1})$ when the channel is considered to contain source 2, where p_{s_1} and p_{s_2} are the powers of sources 1 and 2 contained within the channel, respectively.

The unmixing filters have been initialized with zeros except for the zero-time delayed matrix $\boldsymbol{H}(0)$, which was chosen as the identity matrix: $\boldsymbol{H}(0) = \mathbf{I}$. For algorithms (10) and (15), the time-varying learning rate was set to $\mu = 0.001 - (0.001 - 0.0001)t/t_{\max}$ ($t_{\max} = 3345$) and $\mu = 0.01 - (0.01 - 0.001)t/t_{\max}$ ($t_{\max} = 1300$), respectively, where t is the iteration index and $t_{\max}$ is the maximum number of iterations. The expectations of stochastic variables were replaced with time-moving averages, where the time-moving averaging factor was set to $\alpha = 0.6$. The separation filters were of length 2048. The FFT block size was set to 8192. The overlaps between FFT blocks were 7936 and 4096 for (10) and (15), respectively.

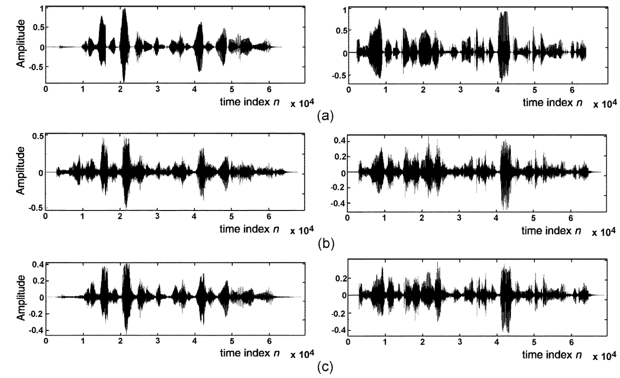

Fig. 1. Signals from the experiment. (a) Two sources from TIMIT database, sampled at 16 kHz. (b) Mixtures with SIRs of 5.86 dB and 6.36 dB, respectively. (c) Separated signals with SIRs of 21.56 dB and 15.14 dB, respectively.

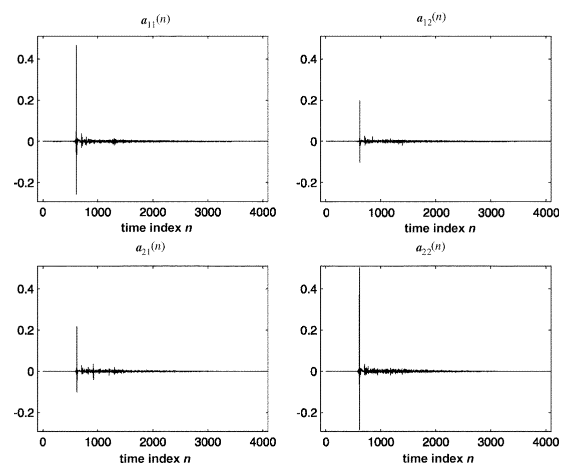

Fig. 2. Mixing system $\boldsymbol{A}(n)$, consisting of four 4096-tap FIR filters.

For comparison purposes, simulations with the methods in [7], [8], and [13] were carried out with separation-filter lengths of 2048 taps as well. The method in [7] was used with a time-varying learning rate of $\mu = 0.003 - (0.003 - 0.0003)t/t_{\max}$ with $t_{\max} = 1840$, the overlap between FFT blocks was 7168, and the nonlinear function was $\tanh(5\mathcal{R}(Y)) + j * \tanh(5\mathcal{I}(Y))$, where $\mathcal{R}(.)$ and $\mathcal{I}(.)$ are the real and imaginary part operators, respectively. For the method of [8], five correlation matrices were simultaneously diagonalized, where 8000 iterations at a learning rate on 1.0 were carried out. The method of [13] was used with the same parameters as those for algorithm (10).

The SIRs before and after the separation are listed in Table I. Firstly, we see that the disjointness maximization algorithm (10) is better than the amplitude-modulation decorrelation algorithm given in (15). Secondly, the results in Table I show that the new approach (10) performs better than the methods in [7] and [8], and slightly better than the one in [13]. While the advantages of algorithm (10) over the methods in [7] and [8] were found to be consistent over many experiments with different settings,

TABLE I
SIRs BEFORE AND AFTER THE SEPARATION WITH DIFFERENT ALGORITHMS

	Before sep.	Alg. (10)	Alg. (15)	Parra's Alg. [8]	Smaragdis' Alg. [7]	Mei's Alg. [13]
SIR_1	5.86 dB	21.56 dB	14.92 dB	12.96 dB	13.45 dB	21.85 dB
SIR_2	6.36 dB	15.14 dB	13.20 dB	12.95 dB	14.84 dB	14.37 dB

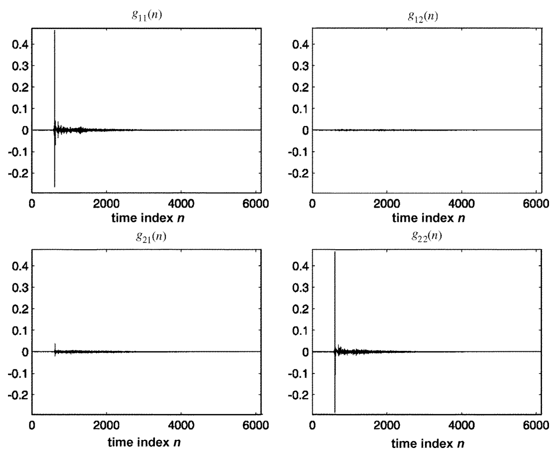

Fig. 3. Impulse responses of the global system $\boldsymbol{G}(n) = \boldsymbol{H}(n) * \boldsymbol{A}(n)$.

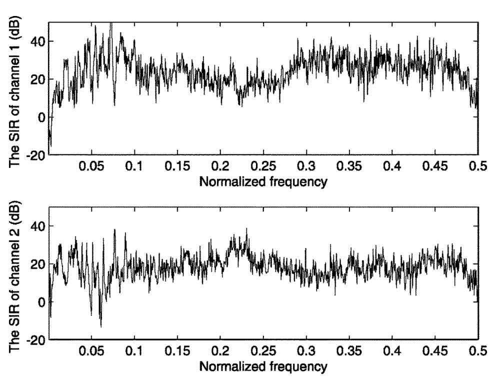

Fig. 4. Subband SIR via frequency of the separated signals.

the comparison with the method from [13] showed that the performance of both algorithms is quite compatible. Depending on the learning rate and iteration time, either the one or the other algorithm performed slightly better.

To further illustrate the results obtained with the proposed algorithm (10), the mixtures and the separated signals are shown in Fig. 1(b) and (c), respectively. The impulse responses of the global system are depicted in Fig. 3.

B. Investigation on Local Permutation

Permutation is always a problem when frequency-domain BSS approaches are mentioned. However, there will be no such problem if a frequency-domain defined integrated objective function is optimized with respect to time-domain parameters [13]. To show this, let us suppose that each of the outputs is dominated by one source, the other source is considered as reference. For each subband signal, the SIR is computed. Obviously, if all subbands and all outputs have SIRs of the same sign, it implies that no permutation has happened. For the example in Section IV-A, Fig. 4 shows that the proposed algorithm is not disturbed by the permutation problem.

V. CONCLUSIONS

In this letter, we proposed a frequency-domain integrated objective function for convolutive BSS on the basis of the maximization of subband-wise disjointness of the separated signals. The permutation problem was avoided through the frequency-domain integration and time-domain optimization. Simulation results for convolutive mixtures of speech show that the algorithm, which primarily maximizes disjointness rather than independence, is valid and of high performance for the separation of convolutive mixtures.

REFERENCES

[1] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, pp. 287–314, 1994.

[2] S. Amari and A. Cichocki, "Adaptive blind signal processing-neural network approaches," *Proc. IEEE*, vol. 86, no. 10, pp. 2026–2048, Oct. 1998.

[3] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–443, Feb. 1997.

[4] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Proc. Inst. Elect. Eng. F, Radar Signal Process.*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[5] A. Cichocki and J. Karhunen *et al.*, "Neural networks for blind separation with unknown number of sources," *Neurocomputing*, vol. 24, pp. 55–93, 1999.

[6] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. IEEE Int. Workshop Wireless Communication*, Paris, France, Apr. 1997, pp. 101–104.

[7] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[8] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 320–327, May 2000.

[9] K. Rahbar and J. Reilly, "Blind source separation of convolved sources by joint approximate diagonalization of cross-spectral density matrices," in *Proc. ICASSP' 01*, 2001, vol. 5, pp. 2745–2748.

[10] I. Sabala, A. Cichocki, and S. Amari, "Relationships between Instantaneous blind source separation and multichannel blind deconvolution," *Proc. IEEE Int. Joint Conf. Neural Networks*, vol. 1, pp. 39–44, 1998.

[11] J. Anemüller, "Maximization of component disjointness: A criterion for blind source separation," in *Proc. ICA 2007, LNCS4666*, 2007, pp. 325–332.

[12] J. Anemüller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," in *Proc. 2nd Int. Workshop Independent Component Analysis and Blind Signal Separation*, 2000, pp. 215–220.

[13] T. Mei, J. Xi, F. Yin, A. Mertins, and J. F. Chicharo, "Blind source separation based on time-domain optimization of a frequency-domain independence criterion," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 6, pp. 2075–2085, Nov. 2006.

[14] M. Kawamoto and Y. Inouye, "Blind deconvolution of MIMO-FIR systems with colored inputs using second-order statistics," *IEICE Trans. Fundam.*, vol. E86-A, no. 3, pp. 597–604, 2003.

[15] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutive mixtures based on second order statistics," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 120–134, Jan. 2005.

[16] A. Jourjine, S. Rickard, and O. Yilmaz, "Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP2000)*, 2000, vol. 5, pp. 2985–2988.

[17] S. Rickard and Ö. Yilmaz, "On the approximate W-disjoint orthogonality of speech," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP2002)*, 2002, vol. 1, pp. I-529–I-532.

[18] P. Bofill and M. Zibulevsky, "Blind separation of more sources than mixtures using sparsity of their short-time Fourier transform," in *Proc. Int. Workshop Independent Component Analysis and Blind Signal Separation (ICA2000, Helsinki)*, 2000, pp. 87–92.

[19] S. Amari, "Natural gradient works efficiently in learning," *Neural Comput.*, vol. 10, pp. 251–276, 1998.

[20] J. S. Garofolo *et al., TIMIT Acoustic-Phonetic Continuous Speech Corpus*. Philadelphia, PA: Linguistic Data Consortium, 1993.

[21] K. Matsuoka, "Minimal distortion principle for blind source separation," in *Proc. 41st SICE Annu. Conf.*, Aug. 5–7, 2002, vol. 4, pp. 2138–2143.