A JOINT DIAGONALIZATION METHOD FOR CONVOLUTIVE BLIND SEPARATION OF NONSTATIONARY SOURCES IN THE FREQUENCY DOMAIN

Wenwu Wang, Jonathon A. Chambers, and Saeid Sanei

Centre for Digital Signal Processing Research, King's College London, London WC2R 2LS, U.K.
E-mail: [wenwu.wang, jonathon.chambers, saeid.sanei]@kcl.ac.uk

ABSTRACT

A joint diagonalization algorithm for convolutive blind source separation by explicitly exploiting the nonstationarity and second order statistics of signals is proposed. The algorithm incorporates a non-unitary penalty term within the cross-power spectrum based cost function in the frequency domain. This leads to a modification of the search direction of the gradient-based descent algorithm and thereby yields more robust convergence performance. Simulation results show that the algorithm leads to faster speed of convergence, together with a better performance for the separation of the convolved speech signals, in particular in terms of shape preservation and amplitude ambiguity reduction, as compared to Parra's nonstationary algorithm for convolutive mixtures.

1. INTRODUCTION

Blind Source Separation (BSS) is an approach to estimate and recover *independent* source signals using only the information within the mixtures observed at each input channel. Many algorithms have been proposed to solve the standard blind source separation problem in which the mixtures are assumed to be instantaneous and the source signals are assumed to be statistically stationary [1]. In practical situations such as in the radiocommunications, telemetry, radar, sonar and speech context, however, the sources are often nonstationary or (quasi)-cyclostationary and the observed signals are usually convolutive mixtures [2] [3]. Unfortunately, conventional BSS methods have limitations in these situations. Increasing interest has therefore been focused on solving the problem of BSS of convolutive mixtures and the main existing strategies can be approximately classified into: (1) carrying out blind separation in the time domain by extending the existing BSS algorithms for the instantaneous case to the convolutive case [4]; (2) transforming totally or partially the convolutive BSS problem into multiple short-term instantaneous problems in the frequency domain and separating the instantaneous mixtures in every frequency bin [3]; (3) decomposing the problem rather than to learn the possibly huge filters all at once, i.e. the decorrelation approach, subspace method, and subband decomposition [5]; (4) exploiting the statistical properties or special structure contained within the source signals to formulate various separation criteria [3].

Performing convolutive BSS in the frequency domain has shown to be an effective way in some existing literature. The time-frequency domain methods were proposed in [6], which transform part of the calculations into the frequency domain but the other parts remain in the time domain. Methods working wholly in the frequency domain were proposed in [7] and [8]. In [9], the temporal structure of the signals is used to realize the separation for convolutive mixtures. Multistage methods were proposed in [10]. The solution using second order statistics provided by nonstationary signals can be found in [3] and [8]. In this work, we propose a new approach to convolutive blind separation of nonstationary sources in the frequency domain.

This work was supported by the Engineering and Physical Sciences Research Council of the U.K.

2. MODEL FORMULATION IN THE FREQUENCY DOMAIN

Assume that N source signals are recorded by M microphones, where $M \geq N$. The output of the j-th microphone is modeled as a weighted sum of convolutions of the source signals corrupted by additive noise, that is

$$x_j(n) = \sum_{i=1}^{N} \sum_{p=0}^{P-1} h_{jip} s_i(n-p) + v_j(n), \qquad (1)$$

where h_{jip} is the P-point impulse response from source i to microphone j $(j = 1, \cdots, M)$, s_i is the source signal from a source i, x_j is the received signal by a microphone j, v_j is the additional sensor noise, and n is the discrete time index. The problem is to reconstruct the unknown sources from the sensor data by assuming statistically independent sources without any other prior knowledge about the sources and the mixing process. The frequency domain approach to blind source separation of convolutive mixtures

(FD-CM-BSS) is to transform the problem into multiple short-term instantaneous BSS problems in the frequency domain and the independent component analysis (ICA) is applied to the instantaneous mixtures in every frequency bin. Using a T-point windowed discrete Fourier transformation (DFT) [11], time-domain signals x_j can be converted into frequency-domain time-series signals $X_j(\omega,\ t)$

$$X_j(\omega, t) = \sum_{\tau=0}^{T-1} x_j(t+\tau)w(\tau)e^{-J2\pi\omega\tau} \tag{2}$$

where $w(\tau)$ denotes a window function, such as Hamming, Hanning or Kaiser windows, $J = \sqrt{-1}$, ω is a frequency index, $\omega = 0, \frac{1}{T}, \cdots, \frac{T-1}{T}$, and T is the frame size of the DFT. We apply corresponding expressions to $H_{ji}(\omega)$, $S_i(\omega,\ t)$ and $V_i(\omega,\ t)$, in which $H_{ji}(\omega)$ does not depend on time index t due to the assumption that the mixing system is time invariant and the same assumption will be applied to the separation system as below. As shown in [3], a linear convolution can be approximated by a circular convolution if $P \ll T$, that is

$$\mathbf{X}(\omega, t) = \mathbf{H}(\omega)\mathbf{S}(\omega, t) + \mathbf{V}(\omega, t) \tag{3}$$

where $\mathbf{S}(\omega, t) = [S_1(\omega, t), \cdots, S_N(\omega, t)]^T$ and $\mathbf{X}(\omega, t) = [X_1(\omega, t), \cdots, X_M(\omega, t)]^T$ are the time-frequency representation of the source signals and the observed signals respectively.

Similar to some time domain methods for instantaneous mixtures, an alternative method for convolutive blind separation in the frequency domain is to estimate a backward model in every frequency bin ω,

$$\mathbf{Y}(\omega, t) = \mathbf{W}(\omega)\mathbf{X}(\omega, t) \tag{4}$$

where $\mathbf{Y}(\omega, t) = [Y_1(\omega, t), \cdots, Y_N(\omega, t)]^T$ is the time-frequency representations of the estimated source signals. The parameters in $\mathbf{W}(\omega)$ are determined so that the elements $Y_1(\omega, m), \cdots, Y_N(\omega, m)$ become mutually independent, where m is the discrete time index. The above calculations are carried out in each frequency bin independently. Following the above discussion, the frequency domain methods for blind separation of convolutive mixtures can be formulated as in the following steps: (i) transform the observed fullband signals into narrowband signals by the discrete Fourier transformation; (ii) optimize the inverse of the mixing matrix in the frequency domain; (iii) reconstruct the time domain fullband separated signal from the frequency domain narrowband separated signals. One can see that the advantages of solving the problem in the frequency domain are clear. For instance, we can simplify the convolutive mixture down into simultaneous mixtures by working in the frequency transform, and many mature unsupervised learning algorithms proposed for separating instantaneous mixtures can be readily used to recover the sources at each frequency bin. It is easy to ensure convergence of the separation filter in iterative ICA learning with high reliability. Moreover, natural signals are more non-Gaussian in the frequency domain. However, as will be discussed in the following sections, permutation indeterminacy of each source will be crucial for reconstructing the sources at each frequency bin, and the separation performance can become saturated before reaching a sufficient performance because the independence assumption is weakened in each narrowband.

3. A JOINT DIAGONALIZATION METHOD FOR FD-CM-BSS USING NONSTATIONARY SECOND-ORDER STATISTICS

It has been shown in [3] that there is a set of second-order conditions that can be specified to perform blind separation for nonstationary signals. Although second-order statistics is not sufficient to identify and invert the mixing coefficients for stationary signals [12] and higher order statistics have to be considered either explicitly [13] or implicitly [14], the second-order criterion is sufficient for non-stationary signals under some weak conditions. For example, in [3] [10], the multiple covariance matrices estimated at different times are simultaneously diagonalized for the convolutive mixtures. Non-stationarity of speech signals can stand in various ways, e.g. vocal tract and glottal signal variations, varying powers, changing correlation of neighboring samples, or even non-stationary higher-order moments. In the following discussion, the source signals are assumed to be nonstationary and the cross-power spectrum at multiple times is considered.

To ensure that $Y_1(\omega, m), \cdots, Y_N(\omega, m)$ are mutually uncorrelated, we can resort to the cross-power spectra matrices $\mathbf{R}_Y(\omega, k)$

$$\mathbf{R}_Y(\omega, k) = \mathbf{W}(\omega)[\mathbf{R}_X(\omega, k) - \mathbf{R}_V(\omega, k)]\mathbf{W}^H(\omega) \tag{5}$$

where $\mathbf{R}_X(\omega, k)$ is the covariance matrix of $\mathbf{X}(\omega, k)$, and $\mathbf{R}_V(\omega, k)$ is the covariance matrix of $\mathbf{V}(\omega, k)$. The objective is to find a $\mathbf{W}(\omega)$ that diagonalizes these matrices simultaneously for all time blocks $k, k = 1, \ldots, K$, that is

$$\mathbf{R}_Y(\omega, k) \rightarrow \mathbf{\Lambda}_C(\omega, k) \tag{6}$$

where $\mathbf{\Lambda}_C(\omega, k)$ is an arbitrary diagonal matrix, which can be derived by the following equation based on the independent assumptions of the source signals s_i $(i = 1, \cdots, N)$ and the sensor noise

$$\mathbf{R}_X(\omega, k) - \mathbf{R}_V(\omega, k) = \mathbf{H}(\omega)\mathbf{\Lambda}_S(\omega, k)\mathbf{H}^H(\omega) - \mathbf{\Lambda}_V(\omega, k) \tag{7}$$

where $\mathbf{\Lambda}_S(\omega, k)$ and $\mathbf{\Lambda}_V(\omega, k)$ are respectively the different diagonal covariance matrices of the source signals and noise signals for each k, and $(\cdot)^H$ denotes Hermitian transpose operator.

As in [3], we would like to estimate directly a stable multi-path backward FIR model in (4). Recalling the separation criteria in [3], we have

$$J(\mathbf{W}) = \arg\min_{\mathbf{W}} \sum_{\omega=1}^{T} \sum_{k=1}^{K} \|E(\omega, k)\|_F^2 \qquad (8)$$

where $||\cdot||_F^2$ is the squared Frobenius norm, $E(\omega, k)$ is the error between $\mathbf{\Lambda}_S(\omega, k)$ and $\mathbf{R}_Y(\omega, k)$. Alternatively, this separation criteria can be replaced by

$$J(\mathbf{W}) = \arg\min_{\mathbf{W}} \sum_{\omega=1}^{T} \sum_{k=1}^{K} J_M(\mathbf{W})(\omega, k) \qquad (9)$$

where $J_M(\mathbf{W})(\omega, k)$ is defined as

$$J_M(\mathbf{W})(\omega, k) = \|\mathbf{R}_Y(\omega, k) - diag[\mathbf{R}_Y(\omega, k)]\|_F^2 \qquad (10)$$

where $diag(\cdot)$ is an operator which zeros off-diagonal elements of a matrix. However, this cost function is not sufficient for joint diagonalization because $\mathbf{W}(\omega) = \mathbf{0}$ leads to a trivial solution to the minimization of $J_M(\mathbf{W})(\omega, k)$. Alternatively, we impose a constraint on $\mathbf{W}(\omega)$ to prevent this degenerate solution at every iteration. Effectively, we incorporate a penalty term to the original cost function. Using a non-unitary matrix constraint with the form $J_C(\mathbf{W})(\omega, k) = \|diag[\mathbf{W}(\omega) - \mathbf{I}]\|_F^2$, we have

$$J(\mathbf{W}) = \arg\min_{\mathbf{W}} \sum_{\omega=1}^{T} \sum_{k=1}^{K} \{J_M(\mathbf{W})(\omega, k) + \lambda J_C(\mathbf{W})(\omega, k)\} \qquad (11)$$

where λ is a weighting factor. The problem in (11) is actually a least squares (LS) estimation problem. In terms of the derivation of any real valued function of a complex valued variable, the gradients in (11) with respect to their parameters can be readily derived as

$$\frac{\partial J}{\partial \mathbf{W}^*(\omega)} = 2 \sum_{k=1}^{K} \{2[\mathbf{R}_Y(\omega, k) - diag(\mathbf{R}_Y(\omega, k))]\mathbf{W}(\omega) [\mathbf{R}_X(\omega, k) - \mathbf{R}_V(\omega, k)] + \lambda diag[\mathbf{W}(\omega) - \mathbf{I}]\} \qquad (12)$$

$$\frac{\partial J}{\partial \mathbf{R}_V^*(\omega, k)} = -4 \sum_{k=1}^{K} diag\{\mathbf{W}^H(\omega)[\mathbf{R}_Y(\omega, k) - diag(\mathbf{R}_Y(\omega, k))]\mathbf{W}(\omega)\} \qquad (13)$$

Similar ideas have been exploited in the literature for blind source separation of instantaneous mixtures in the time domain. A unitary constraint within the Stiefel manifold has been used in [15] [16] [17], which can be analogously realized by applying a projection operation to the gradient of $J_M(\mathbf{W})(\omega, k)$, or incorporating a penalty term of the form $\left\|\mathbf{W}(\omega)\mathbf{W}^H(\boldsymbol{\omega}) - \mathbf{I}]\right\|_F^2$ with $J_M(\mathbf{W})(\omega, k)$, or even a post SVD decomposition operation $\mathbf{W}(\omega) = \mathbf{U}(\omega)\mathbf{\Sigma}(\omega)\mathbf{V}^H(\omega)$. Further discussion can be found in [18]. The non-unitary matrix constraint for joint-diagonalization can be found in [19] [20]. The LS solution to (11) can be obtained using the well-known stochastic gradient algorithm to find the unknown parameters

$$\mathbf{W}^{(l+1)}(\omega) = \mathbf{W}^{(l)}(\omega) - \mu(\omega) \cdot \frac{\partial J}{\partial \mathbf{W}^{(l)*}(\omega)} \qquad (14)$$

In practical implementation, we propose to use a modified gradient of the penalty term, that is, $2diag[\mathbf{W}(\omega) - \mathbf{I}]\mathbf{W}(\omega)$ rather than $2diag[\mathbf{W}(\omega) - \mathbf{I}]$ in (12), which adapts in a more uniform way, and therefore leads to a better separation performance with a little cost of convergence speed as revealed in Section 5. A tradoff between the separation performance and the convergence speed can be obtained using a Lagrange multiplier operator η and the penalty term becomes $diag[\mathbf{W}(\omega) - \mathbf{I}][\eta\mathbf{I} + (1-\eta)\mathbf{W}(\omega)]$. Assuming that the sources are known as a desired response $\mathbf{S}_0(\omega, k)$, η can be chosen in a normalized form as

$$\eta = -\frac{4\,\|diag[\mathbf{W}_0(\omega) - \mathbf{I}]\mathbf{W}_0(\omega)\|_F}{\mathbf{S}_0^H(\omega, k)\mathbf{R}_X^{-1}(\omega, k)\mathbf{S}_0(\omega, k)} \qquad (15)$$

Because the cross-correlation $\mathbf{R}_X(t, t+\tau) = E\{\mathbf{x}(t)\mathbf{x}(t+\tau)^T\}$ is time dependent for nonstationary signals, it is difficult to estimate the cross-power-spectrum with a relatively short stationary time resolution. However, $\mathbf{R}_X(\omega, k)$ can be estimated by the following block form in practice

$$\hat{\mathbf{R}}_X(\omega, k) = \frac{1}{D} \sum_{m=0}^{D-1} \mathbf{X}(\omega, Dk+m)\mathbf{X}^H(\omega, Dk+m) \qquad (16)$$

where D is the number of intervals used to estimate each cross power matrix. In implementation, the normalized step sizes for adaptation of $J_M(\mathbf{W})(\omega, k)$ and $J_C(\mathbf{W})(\omega, k)$ for each frequency bin take the form

$$\mu_{J_M}(\omega) = \alpha\left(\textstyle\sum_{k=1}^{K} ||\mathbf{R}_X(\omega, k)||_F^2\right)^{-1} \qquad (17)$$

$$\mu_{J_C}(\omega) = 0.2/\left(0.04 + \textstyle\sum_{k=1}^{K} \left\|\frac{\partial J_C(\mathbf{W})(\omega, k)}{\partial \mathbf{W}^*(\omega)}\right\|_F\right) \qquad (18)$$

respectively, where α is a scalar adjustable for adaptation.

4. PERMUTATION PROBLEM

As demonstrated in [21], although there are a couple of indeterminacies in the model (1), such as *sign*, *amplitude*, *spectral shape* and *permutation*, all but the permutation ambiguity can be ignored. In section 2, when we try to combine the results from the individual frequency bins in the

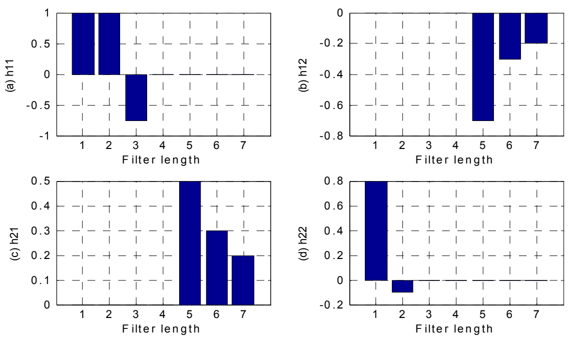

Fig. 1. Coefficients of the mixing filters contained in H(z)

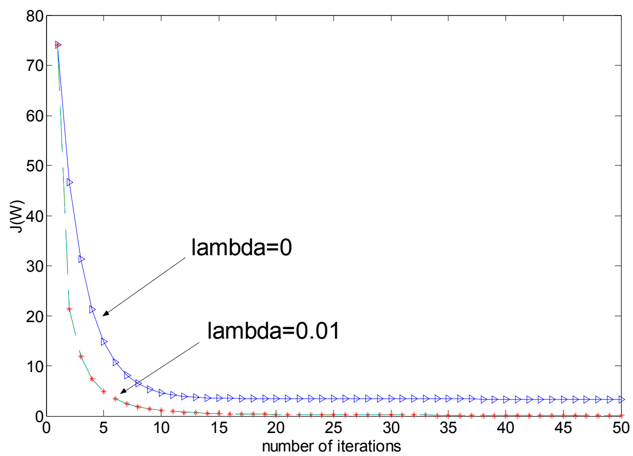

Fig. 2. Convergence performance of the proposed algorithm (λ=0.01) as compared with Parra's algorithm (λ=0)

time domain, the *permutation problem* occurs because of the inherent permutation ambiguity in the rows of $\mathbf{W}(\omega)$. Existing methods try to solve the problem in the following ways: (1) constraints on the filter models in the frequency domain [3] [14]; (2) exploiting the continuity of the spectra of the recovered signals [22]; (3) co-modulation of different frequency bins [23]; (4) using a time-frequency source model [21] [6]; (5) using beamforming view to align solutions [24]. Here, we still use the constraint method as in [3] to allow comparison.

5. NUMERICAL EXPERIMENTS

In this section, we examine the performance of the proposed algorithm by simulation, when applied to artificially mixed signals and real room recordings.

ε^2	$\eta = 0$	$\eta = 1$	$\eta = 0.5$
$\lambda = 0.001$	4.80e-5	4.03e-5	4.61e-5
$\lambda = 0.01$	3.38e-5	1.32e-5	2.86e-5
$\lambda = 0.1$	2.99e-6	2.97e-6	3.17e-6
$\lambda = 1$	9.07e-6	3.78e-6	5.89e-5
$\lambda = 10$	3.60e-4	4.62e-5	5.66e-4
$\lambda = 0$	5.37e-5		

Table 1. Estimation error of the proposed algorithm ($\lambda \neq 0$) as compared with Parra's algorithm ($\lambda = 0$)

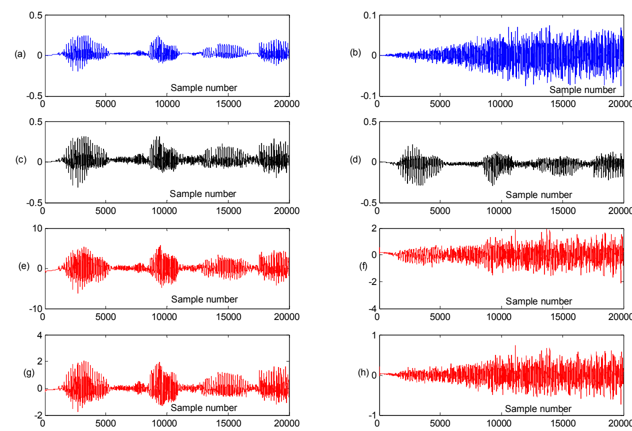

Fig. 3. Separation results of a reading sound and a street background sound mixed by a non-minimum phase system with the proposed algorithm as compared to Parra's algorithm: (a)(b) original sources; (c)(d) mixtures; (e)(f) separated sources by Parra's algorithm; (g)(h) separated sources by the proposed algorithm

5.1. Artificially Convolved Mixtures

A system with two inputs and two outputs is considered, that is, $N = M = 2$. To measure the resemblance between the original and the reconstructed source waveforms, we resort to their mean squared difference. By assuming the signals are zero-mean and unit-variance, we have

$$\varepsilon^2 = E\left[\|\mathbf{y} - \mathbf{s}\|^2\right] = \frac{1}{N}\sum_{i=1}^{N} E\left[|y_i(k) - s_i(k)|^2\right]$$

The percentage of waveform similarity on a logarithmic scale can be defined as $\varepsilon^2(dB) = 10\log_{10}(100(1-\varepsilon^2/2))$. Firstly, we compare the proposed algorithm with Parra's algorithm with respect to the convergence performance and waveform similarity. The source signals are downloaded from the website http://sound.media.mit.edu/ica-bench/. Both of the signals are sampled at 22.05kHz with a duration of 9 seconds. The samples are 16-bit 2's complement in little endian format. One source signal is a recording of the reading sound

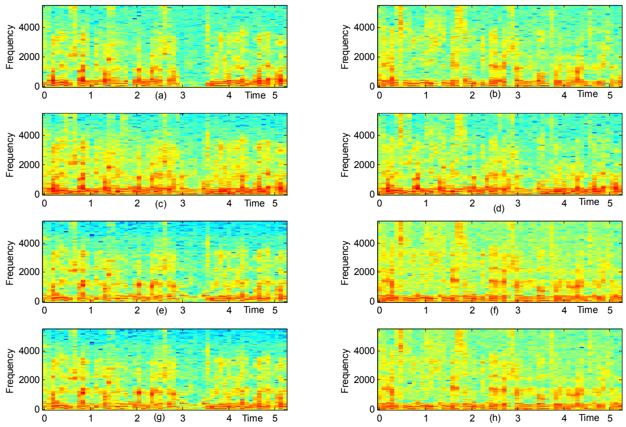

Fig. 4. Separation results of two speech signals mixed by a non-minimum phase system with the proposed algorithm as compared to Parra's algorithm by spectrogram: (a)(b) original sources; (c)(d) mixtures; (e)(f) separated sources by Parra's algorithm; (g)(h) separated sources by the proposed algorithm

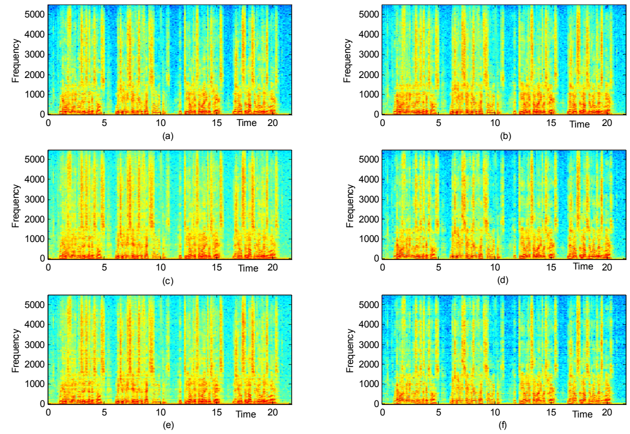

Fig. 5. Separation results of a real room recordings of two speech signals with the proposed algorithm as compared to Parra's algorithm by spectrogram: (a)(b) mixtures; (c)(d) separated sources by Parra's algorithm; (e)(f) separated sources by the proposed algorithm

of a man. The other is a street acoustic background. We artificially mix the two sources by a non-minimum phase system as follows (see [6] and Figure 1).

$$
\begin{aligned}
H_{11}(z) &= 1 + 1.9z^{-1} - 0.75z^{-2} \\
H_{21}(z) &= -0.7z^{-5} - 0.3z^{-6} - 0.2z^{-7} \\
H_{12}(z) &= 0.5z^{-5} + 0.3z^{-6} + 0.2z^{-7} \\
H_{22}(z) &= 0.8 - 0.1z^{-1}
\end{aligned}
$$

The parameters are set to be $\mu = 1$, $K = 5$, $T = 512$, $\alpha = 1$, $D = 1$, $\mathbf{W}_0(\omega) = \mathbf{I}$. The convergence performance and separation results of the proposed algorithm with respect to its parameters are shown in Table 1 (the wave similarity is calculated when it converges) and Figure 2 as compared with Parra's algorithm, from which we see that the proposed algorithm lead to a quicker convergence speed and an improved separation performance. The improved separation performance can also be directly perceived through Figure 3, where $\eta = 0.1$, λ are set to be 0 (corresponding to Parra's case) and 0.01 respectively, and the waveform similarity index is $\varepsilon^2(dB) = 17.168dB$ and $\varepsilon^2(dB) = 18.533dB$ (calculated through all samples) respectively. From Figure 3, we see that the proposed algorithm reduces the amplitude ambiguity to a lower extent at the same time. We further examine the performance of the proposed algorithm by another simulation where the two short speech intervals are downloaded from the website http://medi.uni-oldenburg.de, which are mixed with the same mixing matrix and the parameters are all the same as the first one. The separation results are now plotted by spectrogram, a clearer way to examine the time-frequency representation of speech signals, in Figure 4. Again, it demonstrates an improved separation performance, which is also true with respect to the performance index respectively $\varepsilon^2(dB) = 17.036dB$ and $\varepsilon^2(dB) = 18.264dB$.

5.2. Real Room Recordings

We now apply the algorithm to real room recordings, which is a more complicated case for BSS. The microphone sounds are downloaded from the website http://www.esp.ele.tue.nl/. The room which was used for the recordings was 3.4 x 3.8 x 5.2 m (height x width x depth). Two persons read 4 sentences aloud. The resulting sound was recorded by two microphones which were spaced 58cm apart. The recordings are 16 bit, 24kHz. The separation results are shown in Figure 5. Although giving an objective evaluation to the separation results is not possible, we find a better effect of the proposed algorithm by listening to the separation results which will be demonstrated at the conference.

6. CONCLUSIONS

A joint diagonalization algorithm for separating convolutive mixtures of nonstationary source signals in the frequency domain has been presented. Using the cross-power spectrum and nonstationarity of speech signals, the algorithm incorporates a non-unitary penalty term to the conventional cost function in the frequency domain, which leads to a different search direction to find the minimum of the cost function. Simulation results have shown that the algorithm leads to faster convergence, together with a better performance for the separation of the convolved speech signals, in particular in terms of shape preservation and amplitude ambiguity reduction, as compared to Parra's nonstationary algorithm for convolutive mixtures.

7. REFERENCES

[1] S. Haykin, "Unsupervised adaptive filtering, vol. 1: Blind source separation," *Wiley, New York*, 2000.

[2] K. A. Meraim, Y. Xiang, J. H. Manton, and Y. Hua, "Blind source separation using second-order cyclostationary statistics," *IEEE Trans. on SP*, vol. 49, pp. 694–701, Apr. 2001.

[3] L. Parra and C. Spence, "Convolutive blind source separation of nonstationary sources," *IEEE Trans. on SAP*, pp. 320–327, May 2000.

[4] S. Amari, S. Douglas, A. Cichocki, and H. Yang, "Novel on-line algorithms for blind deconvolution using natural gradient approach," *Proc. SYSID-97, Kitakyushu, Japan*, pp. 1057–1062, Jul. 1997.

[5] A. Mansour, C. Jutten, and P. Loubaton, "Adaptive subspace algorithm for blind separation of independent sources in convolutive mixtures," *IEEE Trans. on SP*, vol. 48, pp. 583–586, Feb. 2000.

[6] T. W. Lee, A. J. Bell, and R. Lambert, "Blind separation of delayed and convolved sources," *Advances in neural information processing systems 9, MIT Press, Cambridge MA*, pp. 758–764, 1997.

[7] P. Smaragdis, "Information theoretic approaches to source separation," Master's thesis, MIT Media Lab, June 1997.

[8] W. E. Schobben and C. W. Sommen, "A frequency domain blind source separation method based on decorrelation," *IEEE Trans. on SP*, vol. 50, pp. 1855–1865, Aug. 2002.

[9] S. Ikeda and N. Murata, "An approach to blind source separation of speech signals," *Proc. ICANN98, Skovde Sweden,*, pp. 761–766, Sept. 1998.

[10] M. Z. Ikram and D. R. Morgan, "A multiresolution approach to blind separation of speech signals in a reverberant environment," *Proc.ICASSP2001*, May 2001.

[11] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Polar coordinate based nonlinear function for frequency-domain blind source separation," *Proc. ICASSP2002,*, pp. 1001–1004, May 2002.

[12] S. Van-Gerven and D. Van-Compernolle, "Signal separation by symmetric adaptive decorrelation: Stability, convergence and uniqueness," *IEEE Trans. on SP*, vol. 43, no. 7, pp. 1602–1612, 1995.

[13] D. Yellin and E. Weinstein, "Multichannel signal separation: Methods and analysis," *IEEE Trans. on SP*, vol. 44, pp. 106–118, 1996.

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

[15] J. F. Cardoso and A. Souloumiac, "Jacobi angles for simulataneous diagonalization," *SIAM J. Matrix Analysis and Application*, vol. 17, pp. 161–164, Jan. 1996.

[16] S. Choi and A. Cichocki, "Correlation matching approach to source separation in the presence of spatially correlated noise.," *Proc. ISSPA, Singapore, 6-8*, Aug. 2001.

[17] J. H. Manton, "Optimisation algorithms exploiting unitary constraints," *IEEE Trans. SP*, vol. 50, pp. 635–650, Mar. 2002.

[18] M. Joho and H. Mathis, "Joint diagonalization of correlation matrices by using gradient methods with application to blind signal separation," *Proc. SAM 2002, Rosslyn, VA, 4-6*, Aug. 2002.

[19] D. T. Pham and J. F. Cardoso, "Blind separation of instantaneous mixtures of non-stationary sources," *IEEE Trans. SP*, vol. 49, pp. 1837–1848, Sept. 2001.

[20] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. SP*, vol. 50, pp. 1545–1553, Jul. 2002.

[21] M. Davies, "Audio source separation," *Mathematics in Signal Separation*, 2000.

[22] V. Capdevielle, C. Serviere, and J. L. Lacoume, "Blind separation of wide-band sources in the frequency domain," *Proc. ICASSP95*, pp. 2080–2083, 1995.

[23] J. Anemuller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," *Proc. ICA2000, Helsinki, Finland*, pp. 215–220, June 2000.

[24] L. C. Parra and C. V. Alvino, "Geometric source separation:merging convolutive source separation with geometric beamforming," *accepted by IEEE Trans. on SAP*, May 2002.