CONVOLUTIVE BLIND SOURCE SEPARATION BASED ON GDFT FILTERBANKS AND PRE-DETERMINED SUBBAND WHITENING

Ebrahim Ghanavati[1], Hamid Sheikhzadeh[2,3], Kamraan Raahemifar[3], and Amin Kheradmand[4]

[1,2,4]Department of Electrical Engineering, Amirkabir University of Technology
P.O.Box 15875-4413, 424 Hafez Ave, Tehran, Iran
[3]Department of Electrical and Computer Engineering, Ryerson University
P.O.Box M5B 2K3, 245 Church St., Toronto, Ontario, Canada
[1]e.ghanavati@aut.ac.ir, [2]hsheikh@aut.ac.ir, [3]kraahemi@ee.ryerson.ca, [4]a_kheradmand@aut.ac.ir

ABSTRACT

This paper focuses on the convolutive blind source separation. Confronted by drawbacks of time-domain and frequency-domain approaches, we propose a novel approach for source separation in subband-domain based on the pre-emphasis processing of subband signals. A time-domain algorithm based on the entropy maximization principle, using the natural gradient algorithm for adaptation task, is employed for subband signal separation. Instead of signal whitening based on frame-by-frame linear prediction analysis, we propose a fixed, pre-determined signal whitening scheme in the subbands to improve the separation performance while decreasing artifacts. With less computational complexity and side-effects, the proposed method is experimentally evaluated and shown to be superior to several other subband-based approaches.

1. INTRODUCTION

Blind source separation (BSS) is a statistical signal processing method which aims to extract source signals from their observed mixtures, assuming almost no *a priori* information about the characteristics of the sources or the mixing environment. The only assumption about the source signals is that they are statistically mutually independent.

This paper focuses on the BSS of convolutive mixtures of speech signals. In literature, different methods have been proposed to tackle this problem. Most of the methods are based on the concept of independent component analysis (ICA) [5].

Early BSS algorithms tried to solve the problem exclusively in the time-domain. In realistic environments, one must adapt fairly long separating filters to adequately separate the observed mixtures. Because of this, time-domain approaches converge very slowly, especially when dealing with colored signals. Moreover, many of these algorithms were originally developed to separate i.i.d signals. When applied to colored signals, these methods could not distinguish between time correlations and spatial correlations of the observed signals. As a result, recovered signals have flattened spectra compared to the source signals. Besides, these algorithms have extreme computational complexities.

Encountered with these obstacles, Smaragdis [14] proposed to take the problem into the frequency-domain. Solving an instantaneous mixing problem in each frequency bin, the convergence rate increases considerably. Moreover, using the benefits of fast Fourier transform (FFT), computational complexity reduces greatly. Still, permutation and scaling problems exceedingly degrade the overall separation performance of these algorithms [14]. Besides, there are fundamental limitations on the separation ability of the frequency-domain algorithms [4].

To alleviate the problems of the time-domain and the frequency-domain algorithms, some researches proposed to tackle the BSS problem in the subband-domain [3, 7, 9, 13]. By use of a reasonable number of subbands and separating filters of appropriate length in each subband, the effects of long reverberations can be properly covered. Using shorter separating filters in subbands and whiter signals to adapt them, the convergence rate is increased considerably. Moreover, applying a time-domain BSS algorithm in each subband, the permutation problem is avoided within subbands. Although the permutation problem might occur between subbands, due to the existence of more information in each subband compared to the frequency-domain algorithms, it is much easier to mitigate the permutation problem [3]. Since the subbands often heavily overlap in the frequency-domain, likelihood of permutation problem between subbands greatly decreases. Our extended experiments with subband-based BSS algorithms also confirm this. Moreover, whiter signals are used in subband-based methods to adapt separating filters, and the whitening artifact of the time-domain algorithms arises independently in each subband. Thus, the whitening distortion yields less overall artifacts in the recovered signals [3].

Motivated by the approach in [9], in this paper, we propose a new method for BSS in the subband-domain. In the approach of [9], linear prediction residuals of the subband signals are used to adapt the separating filters of each subband. After convergence, the adapted filters of each subband are applied to the original subband signals for separation. Inspired by this research, we propose a novel approach based on the pre-emphasis and de-emphasis processing of the subband signals. In contrast to the method of [9], in which block-by-block linear prediction analysis is

used to estimate the whitening filters of each subband signal, we use a single pre-determined pre-emphasis filter to remove the spectral tilt of the subband signals to make them whiter. Then, the whitened mixtures in each subband adapt the separating filters of that subband. After convergence, the adapted filters are applied to the whitened mixtures to extract the pre-emphasised separated outputs. Finally, using a single pre-determined de-emphasis filter, the colors of separated signals are recovered. Additionally, in contrast to the most of the earlier approaches for subband BSS, in which the single sideband (SSB) modulated filterbanks decompose the observed mixtures, we use generalized discrete Fourier transform (GDFT) filterbanks. Our experiments verify that using the proposed approach, superior performance, in terms of separation quality and convergence rate can be achieved.

2. BSS OF CONVOLUTIVE MIXTURES

In a realistic scenario, N_s unobserved source signals (with discrete time index of t), $\mathbf{s}(t)=[s_1(t),\ldots,s_{Ns}(t)]^T$, are travelled through an unknown environment and mixtures of them are received to N_m sensors. The observed mixtures can be described as

$$\mathbf{x}(t)=\sum_{k=0}^{M-1}\mathbf{H}(k)\,\mathbf{s}(t-k),\ t=0,1,\ldots \tag{1}$$

where $\mathbf{x}(t)=[x_1(t),\ldots,x_{Nm}(t)]^T$ is the vector of observed mixtures and $\mathbf{H}$ is the matrix of mixing filters of length M.

BSS aims to adapt a system, $\mathbf{W}$, consisting of separating filters of length T, so that its outputs be as independent as possible. The separated outputs can be modelled as

$$\mathbf{y}(t)=\sum_{k=0}^{T-1}\mathbf{W}(k)\,\mathbf{x}(t-k),\ t=0,1,\ldots \tag{2}$$

where $\mathbf{y}(t)=[y_1(t),\ldots,y_{Ns}(t)]^T$ is the vector of recovered signals. The task of BSS is to adapt the separating system $\mathbf{W}$, such that the global system, $\mathbf{G}$, be of the form

$$\mathbf{G}(z)=\mathbf{H}(z)\,\mathbf{W}(z)=\mathbf{P}\mathbf{D}(z) \tag{3}$$

in which $\mathbf{P}$ is a permutation matrix and $\mathbf{D}(z)$ is a diagonal matrix of arbitrary filters.

3. SUBBAND BSS

Subband-domain BSS consists of three distinct stages, ordered as: subband analysis, subband separation, and subband synthesis. The following subsections describe these stages, respectively.

3.1 Subband analysis stage

In the first stage, all observed mixtures must be decomposed over multiple subbands [6]. This may be done using GDFT filterbanks [13] or SSB filterbanks [3, 7, 9]. In most of the previous subband approaches, the SSB filterbanks are used. In this way, the subband signals are real-valued and one can employ any of the previously proposed time-domain BSS algorithms in each subband, without extending them to deal with complex valued signals. In this paper, we use a GDFT filterbank. The benefit is that the phases of the subband signals are not dropped, which leads to more accurate filter adaptation as shown in our experiments. In the analysis stage, each observed mixture, $x_j(t)$, is decomposed into N subband signals, $X_j^{GDFT}(\kappa,m)$, where $\kappa = 0,\ldots N-1$ is the subband index and m is the time index of the subband signals, related to the full-band time index as $m=Rt$, where R is the decimation rate of the filterbank. One can choose any value for R such that $R\le N$. The prototype low-pass analysis window is of the form

$$h_a(t)=N\,\mathrm{sinc}\left(\frac{t}{N}\right)\mathrm{win}(t),\ t=0,\ldots,4N-1 \tag{4}$$

in which win(t) is a L=4N taps Hamming window. The oversampling (OS) ratio of the GDFT filterbanks is defined as

$$OS=\frac{N}{R}. \tag{5}$$

Using oversampled filterbanks (OS>1), aliasing distortion can be avoided [6] and BSS can be performed in different subbands independently [16]. Besides, as each subband signal has an approximate bandwidth of π/OS, it is much whiter than the full-band signals. This reduces the whitening artifact of the time-domain BSS algorithm which is used in the separation stage.

3.2 Time-domain BSS in subbands stage

Generally, the decimation rate of filterbanks is much smaller than the length of mixing and demixing filters, $R<<\{M, T\}$. So, unlike the frequency-domain approaches, the subband mixtures ought to be considered as convolutive mixtures. Due to decimation by R, it is sufficient to adapt filters of length T/R in each subband.

Based on the entropy maximization principle, a natural gradient algorithm was derived by Amari *et al* [2] for separation of instantaneous mixtures. Employing the isomorphism between scalar and FIR polynomial matrices [11], the algorithm could be generalized for separation of convolutive mixtures [12]. This generalized time-domain algorithm adapts separating filters of κ-th subband, $\mathbf{W}^{(\kappa)}$, based on

$$\tilde{\mathbf{W}}_{l+1}^{(\kappa)}=\tilde{\mathbf{W}}_{l}^{(\kappa)}+\mu(\kappa)\,\Delta\tilde{\mathbf{W}}_{l}^{(\kappa)}, \tag{6}$$

$$\Delta\tilde{\mathbf{W}}_{l}^{(\kappa)}=\left[\mathbf{I}-\mathrm{FFT}\left\{\boldsymbol{\varphi}\left(\mathbf{u}_{l}^{(\kappa)}\right)\right\}\left(\tilde{\mathbf{u}}_{l}^{(\kappa)}\right)^{H}\right]\tilde{\mathbf{W}}_{l}^{(\kappa)} \tag{7}$$

$$\tilde{\mathbf{u}}_{l}^{(\kappa)}=\tilde{\mathbf{W}}_{l}^{(\kappa)}\,\tilde{\mathbf{x}}^{(\kappa)}, \tag{8}$$

in which l is the iteration index, $\mu(\kappa)$ is the step size in κ-th subband, $(.)^H$ is the Hermitian operator, and $\tilde{(.)}$ represents a variable in the frequency-domain. $\mathbf{x}^{(\kappa)}$ and $\mathbf{u}_l^{(\kappa)}$ are the vectors of observed mixtures and separated outputs, respectively. $\mathbf{I}$ is the unit FIR polynomial matrix which its

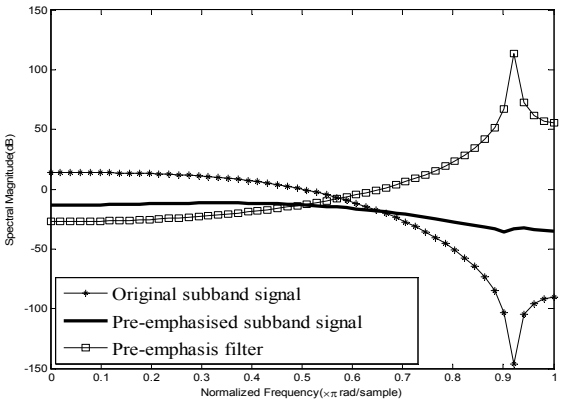

Figure 1 – Spectral magnitudes of a subband signal, pre-emphasis filter and pre-emphasised subband signal

main diagonal elements are sequences of ones, all of length T, and its other elements are sequences of zeros. Moreover, $\boldsymbol{\varphi}(\mathbf{u})=[\varphi_1(u_1),\ldots,\varphi_{Ns}(u_{Ns})]^T$ is the vector of nonlinear activation function, which acts on the time-domain signals. The optimal form of φ (.) is defined as [17]

$$\varphi_i(u_i) = -\frac{\partial}{\partial u_i}\log\left(p_{u_i}(u_i)\right), \tag{9}$$

where $p_u(u)$ is the pdf of the sequence u. The pdf of speech signals can be properly approximated by the generalized Gaussian distributions (GGD). So, the activation function is defined as [17]

$$\varphi_u(u) = \text{sign}(u)|u|^{\alpha-1} \tag{10}$$

in which α is the Gaussian exponent of GGD distributions. Note that although the adaptation algorithm (6)-(8) operates in the frequency-domain to benefit from the advantages of FFT, as φ(.) exclusively acts on the time-domain sequences, the permutation problem is avoided within subbands.

Similar to most of the time-domain BSS approaches, the algorithm (6)-(8) causes whitening distortion in the recovered signals. Several approaches were proposed to tackle this problem, e.g. [10]. In [10], a so called LP-NGA approach is proposed, in which, linear prediction analysis is used to extract the residuals of the observed mixtures. These residuals have flat spectra and are used to adapt separating filters. Reaching to a convergence point, the adapted filters are applied to the original mixtures to produce the recovered signals. Note that in the subband analysis stage, a narrow sector, approximately of bandwidth π/N, of the spectrum of each observed mixture is decimated and extended to an approximate bandwidth of π/OS. So, in the subband BSS algorithms, especially for small values of OS, the adaptation signals are nearly white. Consequently, with respect to the full-band algorithms, the whitening distortion has less adverse effect on the recovered subband signals. However, in order to perform filter adaptation in different subbands, independently, and to prevent the aliasing distortion, the filterbank should be sufficiently oversampled [16]. As a result, the coloration of the subband signals increases and the whitening effect may

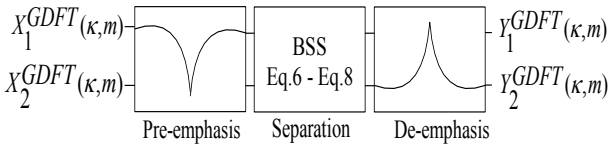

Figure 2 – Configuration of the proposed BSS method

become considerable. To tackle this dilemma, the LP-NGA algorithm was incorporated into the subband BSS framework [9]. Our proposed solution to this problem is much simpler and more efficient. The method is based on the pre/de-emphasis processing of the subband signals. Our motivation comes from the success of the pre-emphasis method for increasing the convergence rate of the subband adaptive filtering [1].

Note that the subband signals are colored, but the color of them is quite known and is determined by the prototype analysis filter. Specifically, as in subband analysis stage, the observed mixtures are decomposed to subband signals using $h_a(t)$ and its modulated versions, the spectral tilt of all subband signals is fairly analogous to the spectrum of $h_a(t)$, as depicted in Figure 1. Accordingly, we propose to remove the spectral tilt of all subband signals and whiten them, using a single, pre-determined pre-emphasis filter, h_{pre}. The spectrum of a three-tap IIR pre-emphasis filter and the spectral shape of a pre-emphasised subband signal are depicted in Figure 1, as well. The pre-emphasised signals in each subband κ, are used as the input to the algorithm (6)-(8). After convergence, employing the adapted system, $W^{(\kappa)}$, on the pre-emphasised signals of the κ-th subband, the pre-emphasised separated signals of the subband are estimated. As denoted in Eq.8, these outputs are in the frequency-domain. We convert these outputs back to the time-domain using the overlap-save method. Finally, using the inverse of h_{pre} as de-emphasis filter in each subband, the color of separated outputs is recovered. The block diagram of the proposed BSS method is depicted in Figure 2. Note that in contrast to the method of [9], in which for each block of each subband, linear prediction coefficients should be estimated and used as the whitening filter of that block, in the proposed method only a single pre-determined filter is employed as the whitening filter for all subband signals. In this way, in addition to the reduced computational complexity, the problems associated with linear prediction analysis in speech processing applications are prevented. In [7], we proposed to incorporate a similar BSS algorithm into the SSB filterbanks. In that work, however, the subband signals were real-valued and the spectra of the adapting signals were centered around π/2. Also, an activation function different from the one in Eq.10 was employed. Our experiments demonstrate the advantage of the proposed method in this paper over the method of [7].

3.3 Subband synthesis stage

In the last stage, the separated signals in different subbands are combined together via a GDFT synthesis filterbank [6] to form the separated signals in the time-domain. To maintain the linear phase property in the subband processing, a time reversed version of $h_a(t)$ can be used as the prototype synthesis window, $h_s(t)$,[15]

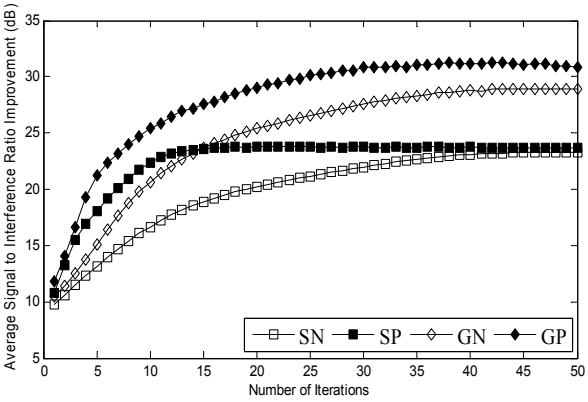

Figure 3 – SIRI values of the four methods in the first experiment

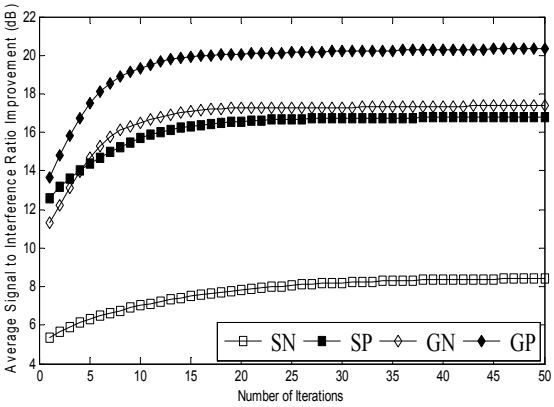

Figure 4 – SIRI values of the four methods in the second experiment

$$h_s(t) = h_a(L-t-1),\ t = 0,\ldots,4N-1 \tag{11}$$

4. EXPERIMENTAL RESULTS

In this section, the performance of the proposed method in separation of convolutive mixtures of speech is examined. The standard case of $N_s=N_m=2$ is considered. The separation ability of BSS algorithms is measured via signal to interference ratio improvement (SIRI). This measure is defined as [3]

$$SIRI_i = SIR_{Oi} - SIR_{Ii},\ i = 1,\ldots,N_s \tag{12}$$

$$SIR_{Oi} = 10\log\left\{\frac{\Sigma_t y_{i,si}^2(t)}{\Sigma_t\left(\Sigma_{j\neq i} y_{i,s_j}(t)\right)^2}\right\}\text{[dB]}, \tag{13}$$

$$SIR_{Ii} = 10\log\left\{\frac{\Sigma_t x_{i,s_i}^2(t)}{\Sigma_t\left(\Sigma_{j\neq i} x_{i,s_j}(t)\right)^2}\right\}\text{[dB]}, \tag{14}$$

where $x_{i,sj}$ and $y_{i,sj}$ specify the contributions of s_j into x_i and y_i, respectively. SIRI determines the ability of BSS algorithm in removing interference and recovering target sources. In order to calculate the SIRI values, the contributions of each source to each of the sensors and outputs should be determined. The contributions of each source can be calculated by activating that source while the other sources are de-activated.

We compare the performances of four subband BSS methods: 1) using SSB filterbank without pre/de-emphasising (method SN); 2) using SSB filterbank and pre/de-emphasising (SP) [7]; 3) using GDFT filterbank without pre/de-emphasising (GN); 4) using GDFT filterbank and pre/de-emphasising (GP) . All of these algorithms outperform the purely time-domain algorithm, in terms of separation performance, convergence rate and spectral conservation. For all algorithms, we choose N=128 and OS=2. Note that for SSB filterbanks, unlike Eq.5, OS=N/2R. For algorithms in which SSB filterbank is used, we choose $\alpha=1$ [7]. Otherwise, for GDFT filterbank-based algorithms $\alpha = 0.6$ is chosen. These selections are adjusted for the optimal separation performance through experimentation. The four algorithms are examined in two sets of experiments. In the first, two zero-mean speech signals, one of a male and another of a female speaker, sampled at 8-kHz and normalized to their maximum amplitude, are used as the source signals. These sources are mixed by use of a mixing system related to a dummy head. Assuming that sources are at angles 30^o and -40^o with respect to the perpendicular bisector of the microphone array, the MATLAB code "headmix.mat" [18] is used to simulate the mixing system. The observed mixtures (obtained by filtering the sources via simulated mixing system), have SIR values of 5.5 dB and 2.07 dB. We use separating filters of length 16-taps in each subband. This is equivalent to $16\times128 = 2048$ filter taps in the time-domain. In Figure 3, the SIRI values, averaged over two channels, are depicted versus the number of iterations for the algorithms. In all experiments, the step size, μ, is tuned for optimum separation performance. As can be deduced from this figure, the methods based on the GDFT filterbank, with or without whitening, are more efficient than those using SSB filterbank, in terms of separation performance. As depicted in the figure, methods GN and GP outperform the SSB-based approaches by about 5.5 and 7 dB, respectively. Although the method SP converges faster than the method GN, using the proposed whitening scheme, the convergence rate of the GDFT-based approach (GP) can be increased to reach to the rate of method SP.

In the second experiment, a substantially more challenging separation task is considered, in which, a speech signal of a female is mixed with babble noise in a chamber, while the sources are at the angles 20^o and 60^o with respect to the perpendicular bisector of the microphone array. Since in this configuration the mixing system is non-minimum phase, the separation task is known to be very

	First experiment			Second experiment		
	SIRI1 [dB]	SIRI2 [dB]	No. of iterations	SIRI1 [dB]	SIRI2 [dB]	No. of iterations
TD	14.6	5.4	44	5.5	12.8	43
SN	19	27.5	35	4.5	12.4	35
SP	19.8	28.3	24	22.1	11.5	25
GN	27	30.2	40	10	24.8	21
GP	26.9	34.9	26	28.3	12.5	14

Table 1 – Comparison of simulated methods with respect to SIRI and convergence speed

hard [8]. Signals are sampled at 16 kHz and are normalized to have a magnitude of one. Input SIRs are equal to 1.23 dB and 2.45 dB. We utilize separating filters of length 32 in each subband, equal to 32×128 = 4096 filter taps in the time-domain. Other parameters are set as the first experiment. Average of SIRI values versus the number of iterations are depicted in Figure 4. As can be seen from this figure, the methods SP and GN have almost similar separation performances and both outperform the method SN by about 9 dB. In this case, the GDFT-based approaches converge almost twice faster than the SSB-based approaches. Moreover, the proposed whitening scheme increases the separation performance by about 3.5 dB, with respect to the methods SP and GN. These results confirm the superiority of the GDFT filterbank over the SSB filterbank in subband BSS applications. Moreover, the proposed whitening further improves the performance. Our documented results reveal the ability of the proposed method to preserve the spectral shape of the recovered signals. It is worth pointing out that based on our extensive experiments none of the implemented subband-domain algorithms suffers from the permutation ambiguity, as expected due to the subband frequency-band overlaps. Table 1 summarizes the performances of all the implemented subband and time-domain (TD in Table 1) algorithms in terms of SIRI and the required number of iterations for convergence.

5. CONCLUSIONS

Subband BSS was considered in this paper. A brief review of advantages and disadvantages of different BSS approaches was presented. Using a GDFT filterbank and pre-emphasis processing of the subband signals via a single pre-determined filter, a time-domain BSS algorithm was improved. The proposal was motivated by the advantages of pre-emphasis processing in subband adaptive filtering and drawbacks of time-domain and frequency-domain algorithms. Our experiments reveal the superiority of the proposed method in terms of separation performance, convergence rate and spectral preservation.

REFERENCES

[1] H. R. Abutalebi, H. Sheikhzadeh, R. L. Brennan and G. H. Freeman, “Convergence improvement for oversampled subband adaptive noise and echo cancellation”, *Proc. of Eurospeech 2003*, Geneva, Switzerland, 2003

[2] S. Amari, A. Cichocki and H. H. Yang, “A new learning algorithm for blind signal separation,” In *Adv. Neural Informat. Process. Systems*. Cambridge, MA: MIT Press, 1996, Vol. 8, pp. 757–763.

[3] S. Araki, S. Makino, R. Aichner, T. Nishikawa, and H. saruwatari, "Subband-based blind separation for convolutive mixtures of speech", *IEICE Trans. Fundamentals*, vol.E88-A, no.12. 2005.

[4] 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, 2003.

[5] P. Comon, “Independent component analysis: A new concept?", *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[6] R.E.Crochiere, L.R.Rabiner, *Multirate digital signal processing*, Prentice Hall. 1983.

[7] E. Ghanavati, H. Sheikhzadeh, K. Raahemifar and A. Kheradmand, “Subband-based convolutive blind source separation with pre-designed data whitening filters”, accepted to *IEEEIITA 2010*. in press.

[8] A. Kheradmand, H. Sheikhzadeh, K. Raahemifar and E. Ghanavati, "Blind source separation in nonminimum-phase systems based on filter decomposition". unpublished.

[9] K. Kokkinakis, P. Loizou, "Subband-based blind signal processing for source separation in convolutive mixtures of speech", *Proc. ICASSP 2003*, 917-920. 2003.

[10] K. Kokkinakis and A. K. Nandi, “Multichannel blind deconvolution for source separation in convolutive mixtures of speech,” *IEEE Trans. Audio, Speech, Lang. Process.*, Vol. 14, No. 1, pp. 200–212, Jan. 2006.

[11] R. H. Lambert, Multichannel Blind Deconvolution: FIR matrix algebra and separation of multipath mixtures. Ph.D. Thesis, University of Southern California, Los Angeles, May 1996.

[12] T.-W. Lee, A. J. Bell, and R. H. Lambert, “Blind separation of delayed and convolved sources,” in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 1997, vol. 9, pp. 758–764.

[13] H.-M. Park, S.-H Oh and S.-Y Lee, “A uniform oversampled filter bank approach to independent component analysis,” In *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Hong Kong, April 6–10, 2003, Vol. V, pp. 249–252.

[14] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain", *Neurocomp.*, Vol. 22, No. 1–3, pp. 21–34, Nov. 1998.

[15] P.P. Vaidyanathan, *Multirate Systems And Filter Banks*, Prentice Hall PTR 1992

[16] S.Weiss, "On adaptive filtering on oversampled Subbands", Ph.D. thesis, Signal Processing Division, University of Sbathclyde, Glasgow, May 1998.

[17] L. Zhang, A. Cichocki and S. Amari,"Self-adaptive blind source separation based on activation functions adaptation", *IEEE Transactions on neural networks*, vol.15, NO.2. 2004.

[18] ICA '99.[Online]. http://sound.media.mit.edu/ica-bench/code