BLIND SIGNAL SEPARATION FOR CONVOLUTIVE MIXING ENVIRONMENTS USING SPATIAL-TEMPORAL PROCESSING

James P. Reilly, Lino Coria Mendoza

Communications Research Laboratory
McMaster University
1280 Main St. W.,
Hamilton, Ontario, Canada
L8S 4K1
reillyj@mcmaster.ca
lcoria@reverb.crl.mcmaster.ca

ABSTRACT

In this paper we extend the *infomax* technique [1] for blind signal separation from the instantaneous mixing case to the convolutive mixing case. Separation in the convolutive case requires an unmixing system which uses present and past values of the observation vector, when the mixing system is causal. Thus, in developing an infomax process, both temporal and spatial dependence of the observations must be considered. We propose a stochastic gradient based structure which accomplishes this task. Performance of the proposed method is verified by subjective listening tests and quantitative measurements.

1. INTRODUCTION

Blind signal separation (BSS) is now becoming a mature topic. Much work [1][3][4][5][6] has been done on BSS for the case of *instantaneous* mixing; i.e., when the transfer functions from the sources to the sensors involve only scaling operations on the inputs. However, less effort has been directed towards the more difficult *convolutive* mixing case, where a much broader class of transfer functions can exist. The convolutive case occurs more often in practice; e.g., acoustic mixing in live reverberative environments. The ability of BSS algorithms to handle the convolutive mixing case greatly expands the potential range of applications where these algorithms can be put to use.

Previous work in BSS for the convolutive mixing case includes [2][7][8][9][10]. However, these methods all suffer from limitations, such as the ability to deal with only two input sources, restrictions on the mixing system, or excessive computational complexity. In this paper, we present a relatively simple *infomax* technique for blind signal separation in the convolutive mixing case that exploits the temporal and spatial properties of the output signals in a straightforward manner.

Our system model is depicted in Fig. 1. M samples from N statistically independent, zero-mean sources are mixed through an $N \times N$ multidimensional dynamic channel to produce an $N \times M$ matrix $\mathbf{X}$ of observations. We denote the nth, $n = 1, \ldots, N$ row $\mathbf{x}_n$ of this matrix to represent the variation in time of the nth signal over all available M samples; likewise, we denote the mth, $m = 1, \ldots, M$ column (snapshot) as $\mathbf{x}(m)$, which represents the spatial variation (i.e., across sensors) at the mth sample instant. An equivalent notation is used for the signals s, v, u and y shown in Fig. 1. Our objective is to produce outputs $\mathbf{v}_n$, which are the desired output signals corresponding to the separated sources. This objective is realized by determining an unmixing system $\mathbf{W} \in \Re^{N \times N}$, and a temporal processing system $\mathbf{A} \in \Re^{N \times N}$ (both whose elements are FIR filters) so that the joint entropy of the outputs y is maximized.

In this paper we assume the elements of the $N \times N$ mixing system $\mathbf{F}$ are FIR filters of known length $K + 1$. It is straightforward to show [11] that separation of the sources can be achieved using an $N \times N$ unmixing system $\mathbf{W}$ whose elements consist of FIR filters of length $L_W + 1 = (N - 1)K$. We can define $\mathbf{W}_\ell \in \Re^{N \times N}, \ell = 0, \ldots, L_W$ as the matrix of FIR filter weights at delay ℓ. The quantity $\mathbf{A}_\ell, \ell = 0, \ldots, L_A$ is defined in a corresponding way from $\mathbf{A}$. In this case however, because there is no cross–coupling between the channels, the $\mathbf{A}_\ell$ are diagonal.

2. INFOMAX CRITERION FOR BSS

Let us define the vector $\mathbf{x} \in \Re^{MN}$ as vec$(\mathbf{X})$[1]. A similar definition holds for the quantities $\mathbf{s}$, $\mathbf{v}$, $\mathbf{u}$ and $\mathbf{y}$. Then the output $\mathbf{v}$ can be expressed in terms of the observations as

$$\mathbf{v} = \tilde{\mathbf{W}}\mathbf{x} \tag{1}$$

where $\tilde{\mathbf{W}}$ is given as

$$\tilde{\mathbf{W}} = \begin{bmatrix} \mathbf{W}_0 & & & & \\ \mathbf{W}_1 & \mathbf{W}_0 & & & \\ \vdots & & \ddots & & \\ \mathbf{W}_{L_W} & \cdots & & \mathbf{W}_0 & \\ & \mathbf{W}_{L_W} & \cdots & & \mathbf{W}_0 \end{bmatrix}. \tag{2}$$

In a similar way, we can define the variable $\mathbf{u}$ as

$$\mathbf{u} = \tilde{\mathbf{A}}\mathbf{v} \tag{3}$$

This work was supported by grants from Consejo Nacional de Ciencia y Tecnología (CONACYT, Mexico), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Telecommunications Research Institute of Ontario (TRIO).

[1]The vec$(\cdot)$ operator concatenates the columns of its matrix argument into one long vector.

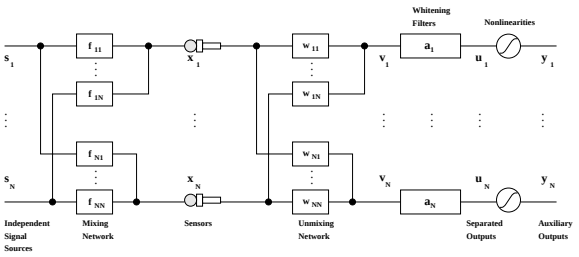

Figure 1: A blind signal separation structure for the convolutive mixing case.

where $\tilde{\mathbf{A}}$ is defined using the $\mathbf{A}_\ell$ in a corresponding way to $\tilde{\mathbf{W}}$ in (2).

The auxiliary output vector $\mathbf{y}$ is defined as

$$y_j = g(u_j), \qquad j = 1, \ldots, MN, \tag{4}$$

where $g(\cdot)$ is a suitable nonlinearity [6][5].

We intend to achieve spatial separation of the outputs $\mathbf{v}$ by maximizing the mutual information between the sources $\mathbf{s}$ and the outputs $\mathbf{y}$ with respect to $\mathbf{A}$ and $\mathbf{W}$, in a manner analogous to the method proposed by [1]. This is equivalent to maximizing the joint entropy $H(\mathbf{y})$ of the outputs $\mathbf{y}$. Maximizing $H(\mathbf{y})$ (under suitable constraints) has the effect of driving the elements of $\mathbf{y}$ towards statistical independence. When the input sources are independent, this criterion is sufficient for blind signal separation. In the convolutive case however, the observations $\mathbf{x}$ are generally dependent in both space and time; thus, $H(\mathbf{y})$ is maximized by forcing both the temporal and spatial dimensions of $\mathbf{x}$ towards independence. In this proposed configuration, the function of $\mathbf{W}$ is to provide spatial independence of the outputs, whereas the function of $\mathbf{A}$ is to provide temporal independence.

To achieve separation, an expression for $H(\mathbf{y})$ in terms of $\mathbf{W}$ and $\mathbf{A}$ is required. In this vein, we can express the joint pdf $f_y(\mathbf{y})$ of $\mathbf{y}$ as

$$f_y(\mathbf{y}) = \frac{f_x(\mathbf{x})}{|J|} \tag{5}$$

where J is the Jacobian of the transformation of $\mathbf{x}$ into $\mathbf{y}$, or

$$J = \det \begin{bmatrix} \frac{\partial \mathbf{y}_1}{\partial \mathbf{x}_1} & \cdots & \frac{\partial \mathbf{y}_1}{\partial \mathbf{x}_M} \\ \vdots & & \vdots \\ \frac{\partial \mathbf{y}_M}{\partial \mathbf{x}_1} & \cdots & \frac{\partial \mathbf{y}_M}{\partial \mathbf{x}_M} \end{bmatrix} \tag{6}$$

where each element of the above is given by

$$\frac{\partial \mathbf{y}_p}{\partial \mathbf{x}_q} = \begin{bmatrix} \frac{\partial y_{p1}}{\partial x_{q1}} & \cdots & \frac{\partial y_{p1}}{\partial x_{qN}} \\ \vdots & & \vdots \\ \frac{\partial y_{pN}}{\partial x_{q1}} & \cdots & \frac{\partial y_{pN}}{\partial x_{qN}} \end{bmatrix}, \qquad p, q = 1, \ldots, M. \tag{7}$$

From (6) we have

$$\begin{aligned} J &= \det \begin{bmatrix} \frac{\partial \mathbf{v}_1}{\partial \mathbf{x}_1} & \cdots & \frac{\partial \mathbf{v}_1}{\partial \mathbf{x}_M} \\ \vdots & & \vdots \\ \frac{\partial \mathbf{v}_M}{\partial \mathbf{x}_1} & \cdots & \frac{\partial \mathbf{v}_M}{\partial \mathbf{x}_M} \end{bmatrix} \det \begin{bmatrix} \frac{\partial \mathbf{u}_1}{\partial \mathbf{v}_1} & \cdots & \frac{\partial \mathbf{u}_1}{\partial \mathbf{v}_M} \\ \vdots & & \vdots \\ \frac{\partial \mathbf{u}_M}{\partial \mathbf{v}_1} & \cdots & \frac{\partial \mathbf{u}_M}{\partial \mathbf{v}_M} \end{bmatrix} \\ &\quad \cdot \det \begin{bmatrix} \frac{\partial \mathbf{y}_1}{\partial \mathbf{u}_1} & \cdots & \frac{\partial \mathbf{y}_1}{\partial \mathbf{u}_M} \\ \vdots & & \vdots \\ \frac{\partial \mathbf{y}_M}{\partial \mathbf{u}_1} & \cdots & \frac{\partial \mathbf{y}_M}{\partial \mathbf{u}_M} \end{bmatrix} \\ &= (\det \tilde{\mathbf{W}})(\det \tilde{\mathbf{A}}) \prod_{j=1}^{MN} y_j'. \end{aligned} \tag{8}$$

In this paper we use $g(u) = \tanh(u)$. Then, y_j' is given as

$$y_j' = 1 - y_j^2. \tag{9}$$

Using (8) and (5) we have

$$\begin{aligned} H(\mathbf{y}) &= \mathrm{E}[\ln f_y(\mathbf{y})] \\ &= \mathrm{E}[\ln |J|] - \mathrm{E}[\ln f_x(\mathbf{x})] \\ &= \mathrm{E}\left[\ln |\det \tilde{\mathbf{W}}| + \ln |\det \tilde{\mathbf{A}}| + \ln \prod_{j=1}^{MN} |y_j'|\right] \\ &\quad + H(\mathbf{x}). \end{aligned} \tag{10}$$

We now have an expression for $H(\mathbf{y})$ in terms of the parameters of interest. An off– line algorithm for separation given a block of M samples could be achieved by directly minimizing (10) with respect to $\mathbf{W}$ and $\mathbf{A}$. However, an on–line algorithm is more desirable. In this respect, we now propose a stochastic gradient ascent algorithm for blind signal separation, which takes into account the previous M samples of data, by maximizing $H(\mathbf{y})$ with respect to $\mathbf{W}$ and $\mathbf{A}$.

3. TRAINING RULES FOR W AND A

We now consider the case for $\mathbf{W}$. The stochastic gradient update for the ℓth weight matrix $\mathbf{W}_\ell$ is given from (10) by

$$\begin{aligned} \Delta \mathbf{W}_\ell &\propto \frac{\partial H(\mathbf{y})}{\partial \mathbf{W}_\ell} \\ &= \frac{\partial}{\partial \mathbf{W}_\ell} \ln |\det \tilde{\mathbf{W}}| + \frac{\partial}{\partial \mathbf{W}_\ell} \ln \prod_{j=1}^{MN} |y_j'|. \end{aligned} \tag{11}$$

The first term evaluates as

$$\begin{aligned} \frac{\partial}{\partial \mathbf{W}_\ell} \ln |\det \tilde{\mathbf{W}}| &= M \frac{\partial}{\partial \mathbf{W}_\ell} \ln |\det \mathbf{W}_0| \\ &= \begin{cases} M \left[\mathbf{W}_0^T\right]^{-1} & \text{if } \ell = 0 \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{12}$$

We now consider the differentiation of the second term in (11) with respect to a particular element $w_{pq\ell}, p, q = 1, \ldots, N$ of $\mathbf{W}_\ell$. This term separates into a sum of log–terms, where only one term depends on $w_{pq\ell}$:

$$\begin{aligned} \frac{d}{dw_{pq\ell}} \sum_{j=1}^{MN} \ln \left|\frac{dy_j}{du_j}\right| &= \frac{d}{dw_{pq\ell}} \sum_{m=1}^{M} \sum_{n=1}^{N} \ln \left|\frac{dy_{mn}}{du_{mn}}\right| \\ &= \frac{d}{dw_{pq\ell}} \sum_{m=1}^{M} \sum_{n=1}^{N} \ln(1 - y_{mn}^2) \\ &= \sum_{m=1}^{M} \sum_{n=1}^{N} \frac{d}{du_{mn}} \ln(1 - y_{mn}^2) \frac{du_{mn}}{dw_{pq\ell}} \end{aligned}$$

$$= \sum_{m=1}^{M}\sum_{n=1}^{N}(-2y_{mn})\frac{du_{mn}}{dw_{pq\ell}} \quad (13)$$

Now, we must concentrate on the term $du_{mn}/dw_{pq\ell}$. Considering only $\mathbf{W}_\ell$, the mth block $\mathbf{u}(m) \in \Re^N$ of $\mathbf{u}$ is a convolution:

$$\mathbf{u}(m) = \sum_{k=0}^{L_A} \mathbf{A}_{L_A-k}\left[\mathbf{W}_\ell \mathbf{x}(k-\ell+m)\right], \qquad m \geq L_A. \quad (14)$$

The differentiation of the above with respect to the pqth element of $\mathbf{W}_\ell$ involves only the pth column of $\mathbf{A}_{L_A-k}$ and the qth element of $\mathbf{x}(k-\ell+m)$. Further, because $\mathbf{A}_\ell$ is diagonal, the only non-zero element in the pth column of $\mathbf{A}_{L_A-k}$ is the scalar $a_p(L_A-k)$. So we get

$$\frac{du_{mn}}{dw_{pq\ell}} = \begin{cases} \sum_{k=0}^{L_A} a_p(L_A-k)x_q(k-\ell+m) & \text{if } p=n \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

Substituing Eq. (15) into Eq. (13) we have:

$$\frac{d}{dw_{pq\ell}}\sum_{m=1}^{M}\sum_{n=1}^{N}\ln(1-y_{mn}^2) = \sum_{m=1}^{M}(-2y_{mp})\sum_{k=0}^{L_A}a_p(L_A-k)x_q(k-\ell+m) \quad (16)$$

Let

$$z_{pq}(m-\ell) = \sum_{k=0}^{L_A} a_p(L_A-k)x_q(k-\ell+m) \quad (17)$$

Then Eq. (16) becomes

$$\frac{d}{dw_{pq\ell}}\sum_{m=1}^{M}\sum_{n=1}^{N}\ln(1-y_{mn}^2) = \sum_{m=1}^{M}(-2y_{mp})z_{pq}(m-\ell). \quad (18)$$

Expressing the above in matrix form, we have the on-line learning rule for $\mathbf{W}_\ell$

$$\Delta\mathbf{W}_\ell \propto \begin{cases} [\mathbf{W}_0^T]^{-1} - 2\mathbf{y}_m \odot \mathbf{z}_m & \text{if } \ell = 0 \\ -2\mathbf{y}_m \odot \mathbf{z}_{m-\ell} & \ell = 1,\ldots,L_W. \end{cases} \quad (19)$$

where

$$\boldsymbol{z}_m = \begin{bmatrix} z_{11}(m) & \cdots & z_{1N}(m) \\ z_{21}(m) & \cdots & z_{2N}(m) \\ \vdots & & \vdots \\ z_{N1}(m) & \cdots & z_{NN}(m) \end{bmatrix} \quad (20)$$

and the $\odot$ operator, which maps an $N \times 1$ vector $\mathbf{y}_m$ and an $N \times N$ matrix $\boldsymbol{z}_m$ into an $N \times N$ matrix, is defined according to the following rule:

$$\boldsymbol{y}_m \odot \boldsymbol{z}_m = \begin{bmatrix} y_{m1}z_{11}(m) & \cdots & y_{m1}z_{1N}(m) \\ y_{m2}z_{21}(m) & \cdots & y_{m2}z_{2N}(m) \\ \vdots & & \vdots \\ y_{mN}z_{N1}(m) & \cdots & y_{mN}z_{NN}(m) \end{bmatrix} \quad (21)$$

We can now turn our attention to training $\mathbf{A}$. In this case, we realize that ideally, $\mathbf{A}$ must produce outputs $\mathbf{u}$ which are *temporally independent*. This may not be possible using only the FIR filter structure proposed in Fig. 1. However, the structure can produce uncorrelated outputs, which in many cases closely approximates independence. The training rules we develop generate an $\mathbf{A}$ for which the resulting temporal dependence is minimized.

The development of the training rule for $\mathbf{A}$ is similar to that for $\mathbf{W}$. Using (10) the adjustment for a stochastic gradient ascent rule for a particular $\mathbf{A}_\ell$ satisfies

$$\Delta\mathbf{A}_\ell \propto \frac{\partial H(\mathbf{y})}{\partial \mathbf{A}_\ell} = \frac{\partial}{\partial \mathbf{A}_\ell}\ln|\det\tilde{\mathbf{A}}| + \frac{\partial}{\partial \mathbf{A}_\ell}\ln\prod_{j=1}^{MN}|y_j'|. \quad (22)$$

As before, the first term evaluates as

$$\frac{\partial}{\partial \mathbf{A}_\ell}\ln|\det\tilde{\mathbf{A}}| = \begin{cases} M\left[\mathbf{A}_0^T\right]^{-1} & \text{if } \ell = 0 \\ 0 & \text{otherwise.} \end{cases} \quad (23)$$

By analogy to (13), we differentiate the second term of (22) with respect to a particular element $a_{pq\ell}$ to get

$$\frac{d}{da_{pq\ell}}\sum_{j=1}^{MN}\ln\left|\frac{dy_j}{du_j}\right| = \sum_{m=1}^{M}\sum_{n=1}^{N}(-2y_{mn})\frac{du_{mn}}{da_{pq\ell}}. \quad (24)$$

The derivative $\frac{du_{mn}}{da_{pq\ell}}$ evaluates to

$$\frac{du_{mn}}{da_{pq\ell}} = \begin{cases} v_p(m-l) & \text{if } p=q=n \\ 0 & \text{otherwise} \end{cases} \quad (25)$$

Substituting (23) (24) and (25) into (22) and combining into matrix form, we have

$$\boldsymbol{\Delta}\mathbf{A}_\ell \propto \begin{cases} [\mathbf{A}_0]^{-1} - \text{diag}[2\mathbf{y}_m \otimes \mathbf{v}_m] & \text{if } \ell = 0 \\ -\text{diag}[2\mathbf{y}_m \otimes \mathbf{v}_m] & \ell = 1,\ldots,L_A. \end{cases} \quad (26)$$

where $\otimes$ means element–by–element multiplication.

4. RESULTS

We demonstrate the performance of the algorithm for an $N = 2, K = 6$ mixing system. The sources are two segments of speech, 4.1 seconds in duration, sampled at 8kHz, normalized so that their maximum amplitude is unity. The speech segment $\mathbf{s}_1$ is a male speaking the phrase "*Marge, it takes two to lie– one to lie, and one to listen*", while $\mathbf{s}_2$ is a female utterance of "*How could you Krusty? I'd never lend my name to an inferior product!*".

The matrices $\mathbf{W}_\ell$ and $\mathbf{A}_\ell$ are all initialized to the identity matrix. The updates at iteration i are made in accordance with (19) and (26) as

$$\mathbf{W}_\ell[i+1] = \begin{cases} \mathbf{W}_0[i] + \eta_W\left(\{\mathbf{W}_0[i]\}^{-T} - 2\mathbf{y}[i] \odot \mathbf{z}[i]\right) & \text{if } \ell = 0 \\ \mathbf{W}_\ell[i] - \eta_W\left(2\mathbf{y}[i] \odot \mathbf{z}[i-\ell]\right) & \ell = 1,\ldots,L_W \end{cases} \quad (27)$$

A corresponding rule is applied for the $\mathbf{A}_\ell$ from (26) , but using the quantity η_A instead. The mixing system $\mathbf{F}$ is specified as

$$\begin{aligned} F_{11}(z) &= 1 + 0.8z^{-1} + 0.7z^{-2} + 0.4z^{-3} + 0.3z^{-4} + 0.2z^{-5} + 0.1z^{-6} \\ F_{12}(z) &= 0.6 + 0.5z^{-1} + 0.5z^{-2} + 0.4z^{-3} + 0.3z^{-4} + 0.2z^{-5} + 0.1z^{-6} \\ F_{21}(z) &= 0.5 + 0.5z^{-1} + 0.4z^{-2} + 0.35z^{-3} + 0.3z^{-4} + 0.2z^{-5} + 0.1z^{-6} \\ F_{22}(z) &= 1 + 0.9z^{-1} + 0.8z^{-2} + 0.6z^{-3} + 0.4z^{-4} + 0.3z^{-5} + 0.1z^{-6} \end{aligned}$$

A *signal to interference ratio*, SIR_n, defined as the desired signal energy to interfering signal energy on the nth channel after convergence is obtained, was calculated. The results for SIR_n vs. η_W are shown in Table 1, for values $L_A = 20$ and $\eta_A = 5 \times 10^{-8}$. The converged speech waveforms corresponding to the bold entry of Table 1 are shown in Fig. 2. The SIR's of the observations $\mathbf{x}$ themselves *before* $\mathbf{W}$ are 3.21 dB and 5.22dB respectively. These quantitative results, in conjunction with subjective listening results which were performed, confirm that a significant level of separation is indeed achieved. Note that it is difficult to assess the level of separation by direct visual comparison of $\mathbf{s}$ and $\mathbf{v}$ in Fig. 2, because the output signals $\mathbf{v}$ have been subjected to a significant filtering operation imposed by the mixing and unmixing networks.

Experimental results have verified that the Frobenius norms $||\mathbf{\Delta W}||_F$ and $||\mathbf{\Delta A}||_F$ approach zero after about 3 seconds of speech, indicating that converged values for $\mathbf{W}$ and $\mathbf{A}$ exist, at least for the case discussed. Qualitative experiments have also indicated the proposed technique is insensitive to initial conditions of $\mathbf{W}$ and $\mathbf{A}$.

η_W	$\text{SIR}_1(dB)$	$\text{SIR}_2(dB)$
0.005	**30.15**	**16.54**
0.05	19.45	16.77
0.5	15.83	6.06

Table 1: SIR's for different learning rates when $K = 6$.

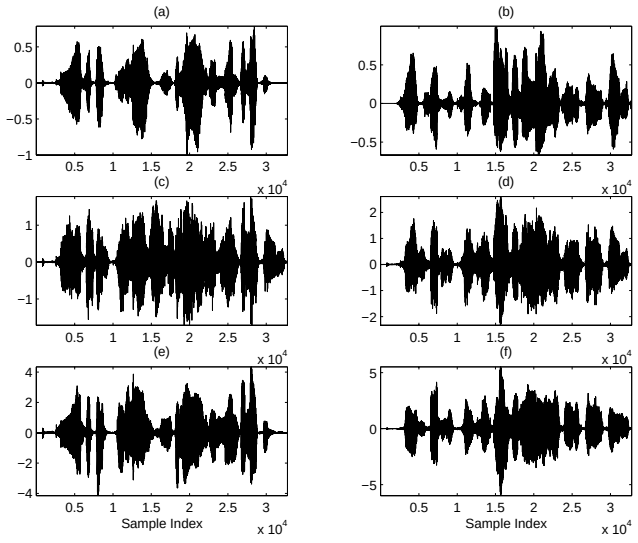

Figure 2: Performance of the algorithm for the $L = 6$ case: sources (a) $\mathbf{s}_1$ and (b) $\mathbf{s}_2$ get mixed through matrix $\mathbf{F}$ to create signals (c) $\mathbf{x}_1$ and (d) $\mathbf{x}_2$. Separation is achieved after convergence, as seen in (e) $\mathbf{v}_1$ and (f) $\mathbf{v}_2$.

5. DISCUSSION AND CONCLUSIONS

We have presented an extension of the infomax technique for blind signal separation to the convolutive mixing case. The extension involves maximization of the joint entropy of $\mathbf{y}$ with respect to $\mathbf{W}$ and $\mathbf{A}$, which are responsible for minimizing statistical dependence amongst the elements of $\mathbf{y}$ in space and time, respectively. Stochastic gradient ascent rules have been derived. The performance of the method has been verified by subjective listening tests and by quantitative measurements.

The natural gradient [4][5], which is well recognized to yield better performance for the instantaneous mixing case, has not yet been derived for this proposed technique. That is a topic for further work.

6. REFERENCES

[1] A. J. Bell and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution, *'Neural Computation*, vol. 7(6), 1995, pp. 1129-1159

[2] K. J. Pope and R. E. Bogner, "Blind signal separation II: Linear, convolutive combinations," *Digital Signal Processing*, vol. 6, 1996, pp. 17-28

[3] P. Comon, "Independent components analysis: A new concept?", *Signal Processing* 36:287-314, 1994.

[4] J-F Cardoso and B.H. Laheld, "Equivariant adaptive source separation", *IEEE Trans. Signal Processing*, 44(12):3017-3030, Dec. 1996.

[5] S. Amari, A. Cichoki and H.H. Yang, "A new learning algorithm for blind signal separation", *Advances in Neural Information Processing Systems*, 8:757-763, 1996.

[6] M. Girolami and C. Fyfe, "Negentropy and kurtosis as projection pursuit indices provide generalized ICA algorithms", *Proc. NIPS*, 1996.

[7] Kari Torkkola, "Blind separation of convolved sources based on information maximization," *IEEE Workshop on Neural Networks for Signal Processing*, Kyoto, Japan, 1996.

[8] E. Weinstein, M. Feder, and A. V. Oppenheim, "Multichannel signal separation by decorrelation," *IEEE Trans. Speech and Audio Processing*, vol. 1, no.4, Oct. 1993, pp. 405–413.

[9] Y. Cao and M. Moody, "Multichannel speech separation by eigendecomposition and its application to co-talker interference removal", *IEEE Trans. Speech and Audio Proc.*, 5(3):209-219, May, 1997.

[10] R. Lambert and A. Bell, "Blind separation of multiple speakers in a multipath environment", *ICASSP98*, Munich, Germany, Apr. 1997.

[11] J. Xi and J. Reilly, "Blind signal separation for convolutive mixing environments", in preparation.