BLIND SEPARATION AND RESTORATION OF SIGNALS MIXED IN CONVOLUTIVE ENVIRONMENT

Jiangtao Xi and James P. Reilly

The Communications Research Laboratory
McMaster University
Hamilton, Ontario, Canada L8S 4K1
jim@reilly.crl.mcmaster.ca

ABSTRACT

This paper proposes new neural network approaches for separating and restoring signals mixed through FIR channels. Firstly, a set of maximal entropy based train rules are developed. Secondly, a new scheme for restoring the original signals is proposed for the 2 × 2 case. Computer simulation results for speech signals are presented to verify the proposed approaches.

1. INTRODUCTION

Blind signal separation is a fundamental and challenging problem in signal processing which has received a great deal of attention in recent years [1-10]. Most of the work on blind signal separation done so far considers only the case of instantaneous combination [7]. However, in many practical situations, the measurements are convolutive mixures of the original signals. Compared to the instantanous combination, the problem of separating signals mixed in convolutive environment is much more difficult. One of the major issues is to develop effective separation algorithms [3, 4, 5, 9] for this case. Of various approaches, the Neural Network technique using maximal output entropy algorithm [10] is especially attractive due to its simplicity in implementaion. However, for the case of convolutive mixing, there are some difficulties in deriving the training rules because the relations between the probility density functions (pdf) of the measurements and the outputs are too complicated to be expressed in analytical forms. Recently, as an effort to separate convolutive mixed speech signals, a set of algorithms for training the separation filter coefficients have been proposed [9].

The other problem associated with the blind separation of signals mixed in convolutive environment is the restoration of the original signals. It has been shown that even though the signals are separated, the resulting separated outputs are still unknown convolved versions of the signal sources rather than the original signals [3, 5]. If the original signals are of interest, a blind deconvolution module is still necessary in addition to the blind separation module.

The purpose of this paper is twofold. Firstly, it presents a set of new training rules on the basis of maximal entropy [10] for the separation network consisting of FIR filters. Compared to the previous work [9], we not only consider the effect of instantaneous measurements, but also the effect of the delayed measurements, because the delayed measurements have equal contribution to the outputs as the non-delayed measurements do. Secondly, we propose an approach which can restore the original signals from the separated outputs.

2. THE NEW TRAINING RULES

Considering the $N \times N$ separation network depicted in Figure 1. Assuming that the separating network consists of causal FIR filters, the outputs are as follows:

$$u_i = \sum_{k=0}^{L} w_{i1,k}x_{1,k} + \sum_{k=0}^{L} w_{i2,k}x_{2,k} + \ldots + \sum_{k=0}^{L} w_{iN,k}x_{N,k} \qquad i = 1, 2, \ldots, N \quad (1)$$

where $w_{ij,k}$ denotes the coefficient associated with delay k of the FIR filter from jth measurement to ith output, and L is the length of the separation filters, which are assumed to be the same for all the filters. The purpose is to work out the training algorithm to maximize the joint entropy of the auxiliary outputs, y_i shown in Figure 1, given as follows:

$$H(Y_N) = \int f_Y(Y_N) log f_Y(Y_N) dY_N \quad (2)$$

where $Y_N = (y_1, y_2, \ldots, y_N)$, while $f_Y(Y_N)$ is the joint probility desity function (pdf). It is obvious that the outputs not only depend on the instantaneous measurements, but also on delayed version of them. In order to have the complete consideration of all the measurements (including the delayed ones), we have to evaluate the relationship between the joint pdf of x's and the joint pdf of y's. Given that

$$f_Y(Y_N) = \frac{f_U(U_N)}{C} \quad (3)$$

where $U_N = (u_1, u_2, \ldots, u_N)$, and C is the Jacobian determinant from U_N to Y_N. Hence we need to work out the relation between $f_U(U_N)$ and the pdf of x's. This is a difficult task because the number of x's, $N(L+1)$, is different from the number of u's, N. Hence we can not directly use the Jacobian. One way to solve this problem is to construct another NL u's, u_i ($i = N+1, N+2, \ldots, NL+N$), in a convenient way so that the number of u's is the same as the number of x's. In this case the joint pdf of u_i ($i = 1, 2, \ldots, NL+N$) can be obtained on the basis of the Jacobian and the pdf of x's. Then we can get the joint pdf of $u_1, u_2, \ldots, u_N$ by computing the marginal pdf of $u_1, u_2, \ldots, u_{NL+N}$. One possible choice for constructing the NL u's is to simply to copy NL measurements except those associated with delay k, that is,

$$u_{mN+i} = \begin{cases} x_{i,m}, & m = 0, 1, \ldots, L (m \neq k); \ \ i = 1, 2, \ldots, N \\ u_i, & m = k; \ \ i = 1, 2, \ldots, N \end{cases} \quad (4)$$

Given the above choice of u's, it can be shown that the Jacobian determinant from x's to u's, denoted as D_k, is simply the Jacobian determinant from $x_{i,k}$ ($i = 1, 2, \ldots, N$) to u_i

($i = 1, 2, ..., N$). The joint pdf of the u's is thus obtained as follows:

$$f_U(u_1, u_2, ..., u_N) = \frac{\int f_X[G^{-1}(U)]du_{N+1}du_{N+2}...du_{NL+N}}{D_k} \quad (5)$$

where $G()$ represents the mapping between x's and u's. It can be shown that Equation (5) can be written in the form:

$$f_U(u_1, u_2, ..., u_N)|_{U=X_k} = \frac{f_X(x_{1,k}, x_{2,k}, ..., x_{N,k})}{D_k} \quad (6)$$

Now let us go back to the joint entropy of $y_1, y_2, ..., y_N$. By using Equations (3), (5) and (6) it can be shown that

$$H(y_1, y_2, ..., y_N) = H(X_k) - E[lnC] - E[lnD_k] \quad (7)$$

where $X_k = [x_{1,k}, x_{2,k}, ..., x_{N,k}]$. The training rules for a filter coefficients can be obtained by calculating the gradient of the joint entropy with respect to that coefficient. It is obvious that the first term is the entropy of the measurements with delay k which are not related to the separating network. Hence the training rule should be determined by the second and third term. Assuming $y_i = \frac{1}{1+e^{-u_i}}$, the results can be obtained as follows:

$$\triangle w_{ij,k} \propto (1 - 2y_i)x_{j,k} + \frac{D_{ij,k}}{D_k} \quad (8)$$

and

$$\triangle w_{ij,l} \propto (1 - 2y_i)x_{j,l}, \qquad \text{if } l \neq k \quad (9)$$

where $D_{ij,k}$ is the cofactor of D_k associated with excluding the (i, j)th element. When $k = 0$, the above results are the same as the training rules proposed in [10]. It is interesting to note that the training rules for the coefficients with delay k, given in Equation (8) are different from those for other coefficients in Equation (9). This is unfair when we see that all the coefficients have the same contribution to u's. In other words, it is more reasonable that all the coefficients have the same form of training rules.

A training rule which equally considers each of the measurements can be obtained as follows. We only use Equation (8) to update $w_{ij,k}$, while disregarding the results in Equation (9). In other words, for deriving the training rules for $w_{ij,k}$, we simply employ the Jacobian from their associated measurements $x_{j,k}$ to u's. This Jacobian is not used for deriving the training rules for other coefficients. This is reasonable because the effect of $x_{j,k}$ on u's is dependent only on their associated weights $w_{ij,k}$, while the other coefficients are irrelevant. The resulting training rules for all the coefficients are the same as Equation (8), which is rewritten as follows:

$$\triangle w_{ij,k} \propto (1 - 2y_i)x_{j,k} + \frac{D_{ij,k}}{D_k}, k = 1, 2, ..., L \quad (10)$$

Computer simulations are performed on the above training rules for the 2×2 separation system illustrated in Figure 2, where the two measurements are combinations of two speech signals through various mixing network parameters. Some of the results is illustrated in Figures 3 and 4. We have simulated the following situations:

(a). Both the mixing filters and the separation filters are of fourth order, where the coefficients of the mixing filters are given as follows:

H_{11}=[1 0.6 0.45 0.34 0.3] and H_{12}=[0.75 0.5 0.31 0.25 0.2]
H_{21}=[0.78 0.63 0.45 0.25 0.1] and H_{22}=[1 0.65 0.35 0.23 0.2] (11)

The results are shown in Figure 3. Clearly good separation is obtained.

(b). The mixing filters and the separation filters are of tenth order, where the coefficients of the mixing filters are as follows:

H_{11}=[1 0.8 0.7 0.5 0.3 0.3 0.2 0.2 0.2 0.18 0.15]
H_{12}=[0.7 0.65 0.5 0.43 0.37 0.3 0.25 0.2 0.17 0.15 0.13]
H_{21}=[0.78 0.63 0.55 0.45 0.31 0.29 0.24 0.21 0.2 0.11 0.1]
H_{22}=[1 0.85 0.77 0.55 0.43 0.31 0.27 0.25 0.21 0.19 0.14] (12)

The simulation results for this situation are given in Figure 4. It is seen that the separation is still considerable.

It is found in our simulations that the performance is highly dependent on the complexity of the mixing network. When the order of the mixing filter is small (for example, less than 10), good separation result can be attained. The performance becomes worse when increasing the number of order of the mixing filters. It is also found that the performance of the algorithm is sensitive to signal magnitudes, the learning rate and the initial values of coefficients. In our simulations, the speech signals are scaled within the range [-1 1], the learning rate is 0.00001, initial vaules of coefficients are $w_{ij,k} = 1$ for $i = j$ and $w_{ij,k} = 0$ for $i \neq j$.

3. ORIGNAL SIGNAL RESTORATION BASED ON IDENTIFICATION

The approaches proposed above are able to separate the signals which are mixed in convolutive environments. However, in the general case the separated signals are unknown convolved versions of the original signals. In this section, we will develop a relationship between the separating newtork and the mixing network. The relationship can be used to identify the unknown mixing network. A restoration scheme is established on the basis of the identification results.

Consider the system depicted in Figure 2. The output signals in frequency domain can be written as follows:

$$U_1(\omega) = [H_{11}(\omega)W_{11}(\omega) + H_{21}(\omega)W_{12}(\omega)]S_1(\omega) + [H_{12}(\omega)W_{11}(\omega) + H_{22}(\omega)W_{12}(\omega)]S_2(\omega) \quad (13)$$

and

$$U_2(\omega) = [H_{11}(\omega)W_{21}(\omega) + H_{21}(\omega)W_{22}(\omega)]S_1(\omega) + [H_{12}(\omega)W_{21}(\omega) + H_{22}(\omega)W_{22}(\omega)]S_2(\omega) \quad (14)$$

where $H_{ij}(\omega)(i, j = 1, 2)$ are the frequency responses of the unknown mixing systems, and $W_{ij}(\omega)(i, j = 1, 2)$ are the frequency responses of the unknown mixing systems, which are all assumed to be FIR filters. As indicated, there have been a number of blind signal separation techniques. Therefore it is reasonable for us to assume that the signals has already been separated. In this case there are two possible results described as follows: (1) $U_1(\omega)$ contains $S_1(\omega)$ only and $U_2(\omega)$ contains $S_2(\omega)$ only, and (2) $U_1(\omega)$ contains $S_1(\omega)$ only and $Y_1(\omega)$ contains $S_1(\omega)$ only.

Let us consider the first case. From Equation (13) and (14) we have:

$$H_{11}(\omega)W_{21}(\omega) + H_{21}(\omega)W_{22}(\omega) = 0 \quad (15)$$

and

$$H_{12}(\omega)W_{11}(\omega) + H_{22}(\omega)W_{12}(\omega) = 0 \quad (16)$$

here we have assumed that $S_1(\omega)$ and $S_2(\omega)$ are not zero. Equation (15) and (16) can be interpreted as the cases in Figure 5 and Figure 6 respectively. These situations are exactly the same as that in [6], in which similar conditions as indicated in Equation (15) and (16) can be used to identify the unknown mixing filters, as long as the following conditions are satisfied:

a. The polynominals $H_{ij}(\omega)$, the Fourier transform of $h_{ij}(t)$are coprime or do not share any common roots;

b. The input $s_i(t)$ has $2L+1$ or more modes, or it can be expressed as a linear sum of $2L+1$ exponentials, where L is the maximun order among $h_{ij}(t)$ for $i=1,2$.

The unknown mixing filters can be identified by the following equations [6]:

$$H_{12}(\omega) = -W_{12}(\omega), H_{22}(\omega) = W_{11}(\omega) \quad (17)$$

and

$$H_{21}(\omega) = -W_{21}(\omega), H_{11}(\omega) = W_{22}(\omega) \quad (18)$$

Substituting Equations (15), (16), (17) and (18) into (13) and (14) gives:

$$\begin{aligned} U_1(\omega) &= C(\omega)S_1(\omega) \\ U_2(\omega) &= C(\omega)S_2(\omega) \end{aligned} \quad (19)$$

where $C(\omega) = W_{11}(\omega)W_{22}(\omega) - W_{21}(\omega)W_{12}(\omega)$. Assuming $C(\omega)$ is invertible, the original signals can be restored by passing the outputs through a filter whose frequency response is the reciprocal of $C(\omega)$. Note that both channels require the same filter to restore the original signals.

For the other possible situation where $U_1(\omega)$ contains $S_2(\omega)$ only and $U_2(\omega)$ contains $S_1(\omega)$ only. It can be shown that,

$$\begin{aligned} U_1(\omega) &= C(\omega)S_2(\omega) \\ U_2(\omega) &= C(\omega)S_1(\omega) \end{aligned} \quad (20)$$

It is interesting to note that the same restoration filter is used not only for both channels, but also for the two cases considered. This result is very useful in that we do not need to discriminate between the two above cases.

Now let us examine the performance of the proposed restoration technique by looking at the impulse responses from each of the two inputs to the two outputs. Firstly, computer simulations have been performed on the above restoration scheme using the examples given in Figure 4. We have evaluated the impulse responses from the original signal ports to the separation outputs and to the restoration outputs. The results are given in Figure 7. T_{ij} refers to the impulse response from s_i to u_j, while $V_{i,i}$ are the impulse response from u_i to the restored outputs. It is seen that T_{12} and T_{21} are much smaller than T_{11} and T_{22}, which means that the two original signals are separated at the outputs in that u_1 contains s_1 only and u_2 contains s_2 only, although there is some leakage as reflected by non-zero T_{12} and T_{21}. It is also clear that u_1 and u_2 are convolved versions of the original signals, because both T_{11} and T_{22} are far from delta function. However, it is seen that V_{ii} are much closer to delta function. Hence the proposed the proposed technique can effectively deconvolve the signals.

In order to verify the proposed technique further, consider the example which was used in [5]. The two original signals are a female voice singing without music and musical track respectively. 5-tap mixing filters were used to produce $x_1(t)$ and $x_2(t)$ as artifical mixture of $s_1(t)$ and $s_2(t)$. 5-tap FIR separating filters were estimated using the constant diagonal algorithm proposed in [5]. It is seen from Figure 8 that the proposed restoration scheme can also deconvolve the signals.

4. CONCLUSIONS

We have derived a set of new training rules for blind separation systems consisting of FIR filters. Computer simulations with speech signals have been performed to verify the performance of the proposed training rules. It was found that good separation can be attained, although the signal magnitude, learning rates and the initial values should be carefully selected. We also proposed an approach to deconvolve the separated signals in order to restore the original signals. It is seen by computer simulation the proposed restoration scheme can effectively deconvolve the signals.

5. ACKNOWLEDGMENT

We wish to appreciate the support from TRIO (Tecommunication Research Institute, Ontario) on the work of this paper. In addition, we want to thank Mr. D. C. B. Chan, Dr. P. J. W. Rayner and Dr. S. J. Godsill, Dr. A. J. Bell and Professor T. J. Sejnowski as well as Dr. Kari Torrkola for their kind help in providing the details of their work. We also appreciate the kind assistance of Zhengbin Wang.

REFERENCES

[1] 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.

[2] J. L. Lacoume and P. Ruiz, "Separation of independent sources from correlated inputs,"*IEEE Trans. Signal Processing*, Vol. 40, no.12, Dec. 1992, pp. 3074–3078.

[3] D. Yellin and E. Weinstein, "Criteria for multichannel signal separation,"*IEEE Trans. Signal Processing*, Vol. 42, no.8, Aug. 1994, pp. 2158–2168.

[4] S. Li and T. J. Sejnowski, "Adaptive separation of mixed broad-band sound sources with delays by a beamforming Herault-Jutten Network,"*IEEE Journal of Oceanic Engineering*, Vol. 20, no.1, Jan. 1995, pp. 73–78.

[5] D. C. B. Chan, P. J. W. Rayner and S. J. Godsill, "Multi-channel blind signal separation by decorrelation,"*IEEE Workshop on Speech Processing*, 1995.

[6] H. Liu, G. Xu and L. Tong, "A deterministic approach to blind identification of multi-channel FIR systems," *ICASSP'94*, pp.IV-581-584

[7] K. J. Pope and R. E. Bogner, "Blind signal separation, II. Linear, instantaneous combinations," *Digital Signal Processing*, Vol. 6, 1996, pp. 5-16

[8] K. J. Pope and R. E. Bogner, "Blind signal separation, II. Linear, convolution combinations," *Digital Signal Processing*, vol. 6, 1996, pp. 12-28

[9] Kari Torkkola, "Blind separation of convolved sources based on information maximization," *IEEE Workshop on Nenral Networks for Signal Processing*, Kyoto, Janpan, Sept. 4-6, 1996

[10] 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

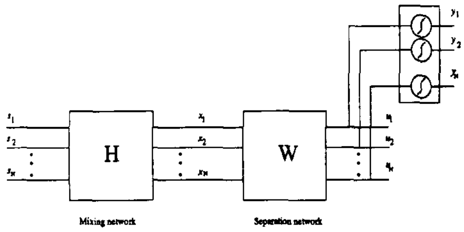

Figure 1. A blind signal separation system

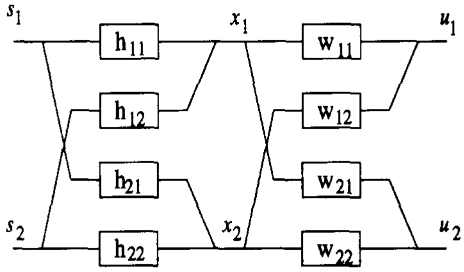

Figure 2. A 2X2 blind signal separation system

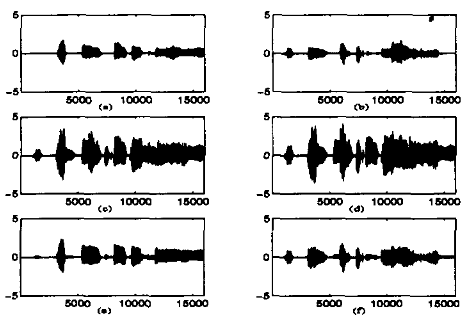

Figure 3. Simulation results: (a) and (b) are the two original signals, (c) and (d) are the measurements, (e) and (f) are the separated results, L=4

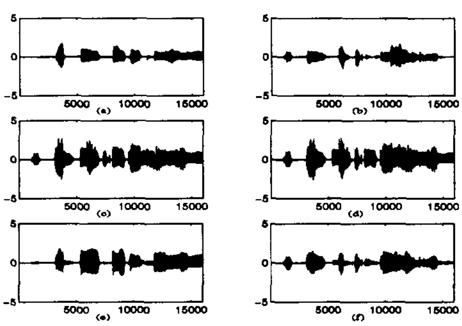

Figure 4. Simulation results: (a) and (b) are the two original signals, (c) and (d) are the measurements, (e) and (f) are the separated results, L=10

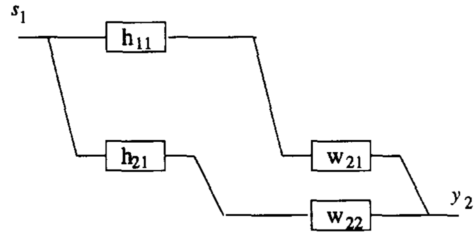

Figure 5. A equivalent network

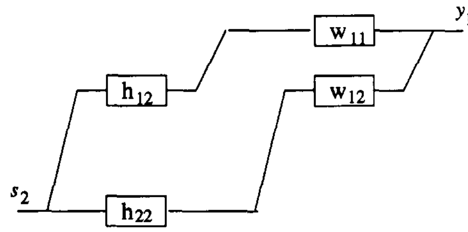

Figure 6. A equivalent network

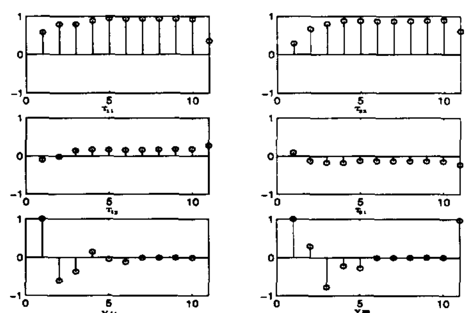

Figure 7. Performance of the proposed restoration scheme

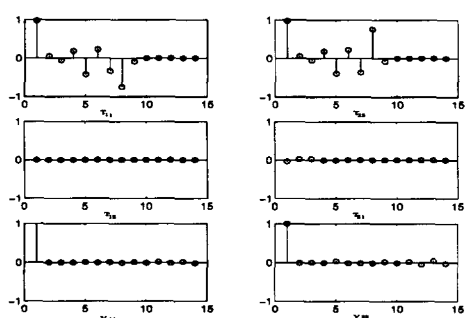

Figure 8. Performance of the proposed restoration scheme