AR PROCESSES AND SOURCES CAN BE RECONSTRUCTED FROM DEGENERATE MIXTURES

Radu Balan, Alexander Jourjine, Justinian Rosca

Siemens Corporation Research
755 College Road East
Princeton , NJ 08540
{radu,jourjine,rosca}@scr.siemens.com

ABSTRACT

When mixing of sources is degenerate the known blind source separation methods fail, since in general the degenerate BSS is an ill-posed problem. Here we report that if signal transmission is modeled by AR(p) processes one can reconstruct the processes and estimate the sources from their degenerate mixture using only second order statistics. We also prove that the approach fails for a general ARMA(p,q) model. The theoretical results are verified in the case of degenerate mixing of two voices and on synthetic data.[1]

1. INTRODUCTION AND STATEMENT OF THE PROBLEM

Current Blind Source Separation (BSS) literature addresses the case when the number of sources is equal to the number of microphones [JH91, Com94, BS95, Ama96, Tor96, Car97, PP96]. Little work has been done to address the degenerate case when this constraint is not satisfied. Particularly hard is the case of interest for many BSS applications when there are more sources than the number of microphones.

This report demonstrates that separtion in such a degenerate case is feasible. We propose a source separation architecture where sources are modeled as AR processes. We solve a special case of the singular multivariate AR identification problem, namely when the measurement is scalar but the noise term is a 2-dimensional vector. Our current approach is based on the second order statistics only. Methods based on second order statistics have for regular mutivariate AR identification and signal separation (see for instance [S.M88, S.N96, WFO93]). In contrast to these studies, our work concerns the singular case for both the mutivariate AR identification and signal estimation in the BSS problem.

We apply ths approach to the degenerate case of the BSS problem, specifically when a scalar mixture of independent source signals is recorded with one microphone. The theory for singular multivariate AR process identification that is developed here can be extended to higher dimensions (i.e. more sources than two voices).

Let us consider two independent univariate $AR(p)$ processes of order p and the measurement given by the sum of the two outputs (see Figure 1). The time-domain evolution equations are the following:

$$\begin{array}{c} s_1(n) = -\sum_{k=1}^{p} a_k s_1(n-k) + G_1\nu_1(n) \\ s_2(n) = -\sum_{k=1}^{p} b_k s_2(n-k) + G_2\nu_2(n) \\ x(n) = s_1(n) + s_2(n) \end{array} \quad (1)$$

where ν_1 and ν_2 are two independent unit variance white-noises, $a_1, \ldots, a_p$ and $b_1, \ldots, b_p$ are the parameters of the first and second AR process respectively and G_1 and G_2 are real constants.

The problem is to identify the $2p+2$ real parameters $a_1, \ldots, a_p, b_1, \ldots, b_p$, G_1 and G_2 based on the measurement $\{x(n)\}_{n=1,\ldots,N_0}$ of a realisation of (1). Our solution is based on the second order statistics of the measurements practically given by the sampled autocovariance coefficients

$$\hat{r}(l) = \frac{1}{N}\sum_{k=l}^{N} x(k)x(k-l).$$

The organization of the report is the following: section 2 presents the main theoretical results. First we show how the spectral density of x can be decomposed; second we derive a modified ARMA estimator by a polynomial system that involves second order statistics of the measurements. Section 3 presents a gradient algorithm to solve these equations together with some other algorithm to address the estimation problem. Section 4 contains numerical experiments showing

[1] In Proceedings of the 1st ICA-BSS Conference, Aussois, France, January 1999

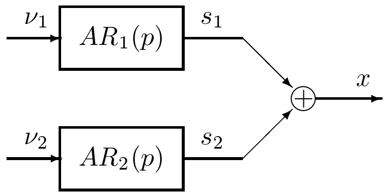

Figure 1: The Singular Multivariate AR Model

a successful application of the theory and is followed by conclusions.

2. THE MAIN RESULTS

Since the two signals s_1 and s_2 are independent, the process (1) has the spectral power density given by the following formula:

$$R_x(z) = \frac{G_1^2}{P_1(z)P_1(\frac{1}{z})} + \frac{G_2^2}{P_2(z)P_2(\frac{1}{z})} \tag{2}$$

Now it is easy to prove the following decomposition (factorization) result:

THEOREM 1 *Suppose we are given the sum x of two independent and stable $AR(p)$ process outputs s_1 and s_2. Furthermore suppose the processes have no common poles. Then the second order statistics is generically enough to uniquely identify the two AR processes.*

Remarks 1. By *generical* we mean that the set of "bad" AR processes form an algebraic manifold of positive codimension in the $2p+2$ dimensional space of parameters. Actually we can say a lot more about this algebraic manifold. These results will apear shortly in a full-length report.

2. We point out that the uniqueness of the decomposition (2) holds only for AR processes. If we replace them by $ARMA$ processes, the result no longer holds true, as can be easily seen.

3. Equation (2) shows that $x(n)$ is second-order statistics equivalent with an ARMA(2p,p) process whose transfer function $Q(z)/P(z)$ is related to our AR(p) processes by:

$$\begin{array}{c} P(z) = P_1(z)P_2(z) \\ Q(z)Q(\frac{1}{z}) = G_1^2 P_2(z)P_2(\frac{1}{z}) + G_2^2 P_1(z)P_1(\frac{1}{z}) \end{array}$$

This would suggest the following identification algorithm:

ALGORITHM 1 (ARMA(2p,p) Identification)
1. Identify the process $\{x(n)\}$ as an $ARMA(2p,p)$ process $Q(z)/P(z)$;

2. For each partition of the $2p$ roots of P into two subsets of p zeros each, construct the polynomials P_1 and P_2 that have these roots and compute G_1 and G_2 that best fit the second equation above (we explain what we mean by best fit in the next section after the Algorithm 2);

3. Choose the partition that gives the smallest error, and that will be an estimate of $G_1, P_1,\ G_2, P_2$.

We tried this algorithm but it does not give acceptable estimates, particularly for large p. A second approach to this problem is to look for a *Modified ARMA estimator (MARMA estimator)*, adapted to our special form. To do this we need to obtain the time-domain evolution equation of the measurement. In the z transform domain we have:

$$P_1(\frac{1}{z})P_2(\frac{1}{z})x(z) = G_1 P_2(\frac{1}{z})\nu_1(z) + G_2 P_1(\frac{1}{z})\nu_2(z)$$

which turns into the following equation:

$$\begin{array}{l} x(n) + \sum_{k=1}^{2p}(a*b)_k x(n-k) = G_1\nu_1(n) + G_2\nu_2(n) + \\ \qquad + \sum_{k=1}^{p}(G_1 b_k \nu_1(n-k) + G_2 a_k \nu_2(n-k)) \end{array}$$

where $(a*b)_k = \sum_{l=0}^{k} a_l b_{k-l}$ with the convention $a_0 = b_0 = 1$. To obtain the second order statistics evolution, we correlate $x(n)$ with $x(n-l)$ and $s_1(n)$ with $\nu_1(n-l)$, respectively $s_2(n)$ with $\nu_2(n-l)$ in (1). Let us denote as follows: $r(l) = E[x(n)x(n-l)]$, $\Psi_1(l) = E[s_1(n)\nu_1(n-l)]$, $\Psi_2(l) = E[s_2(n)\nu_2(n-l)]$, where $E[X]$ is the expected value of the random variable X. Then we obtain the following system of polynomial equations:

$$\begin{array}{l} r(l) + \sum_{k=1}^{2p}(a*b)_k r(l-k) = G_1\Psi_1(-l) + G_2\Psi_2(-l) + \\ \qquad + \sum_{k=1}^{p} G_1 b_k \Psi(k-l) + \sum_{k=1}^{p} G_2 a_k \Psi(k-l) \\ \Psi_1(l) = -\sum_{k=1}^{p} a_k \Psi_1(l-k) + G_1\delta_{l,0} \\ \Psi_2(l) = -\sum_{k=1}^{p} b_k \Psi_2(l-k) + G_2\delta_{l,0} \end{array} \tag{3}$$

where δ is the Dirac impulse.

Now note two things; First we do not know the theoretical autocovariance coefficients, so we have to replace $r(l)$ by the sampled values $\hat{r}(l)$; Second note the causality relations between s_1, s_2 and the noise inputs. This causality implies $\Psi_1(l) = \Psi_2(l) = 0$ for every $l < 0$. Therefore the system (3) becomes:

$$\begin{array}{l} \hat{r}(l) + \sum_{k=1}^{2p}(a*b)_k \hat{r}(l-k) - (G_1^2 + G_2^2)\delta_{l,0} - \\ \quad - \sum_{k=l}^{p}(G_1 b_k \Psi_1(k-l) + G_2 a_k \Psi_2(k-l)) = 0 \\ \Psi_1(l) = -\sum_{k=1}^{min(l,p)} a_k \Psi_1(l-k)\ ,\quad \Psi_1(0) = G_1 \\ \Psi_2(l) = -\sum_{k=1}^{min(l,p)} a_k \Psi_2(l-k)\ ,\quad \Psi_2(0) = G_2 \end{array} \tag{4}$$

We solve this nonlinear system in $G_1, G_2, \underline{a}, \underline{b}$ by looking for the least square solution that minimizes a quadratic criterion of the form:

$$J = \sum_{l=0}^{L} \alpha_l |\hat{r}(l) + \sum_{k=1}^{2p}(a*b)_k \hat{r}(l-k)(G_1^2+G_2^2)\delta_{l,0} - \\ - \sum_{k=l}^{p}(G_1 b_k \Psi_1(k-l) + G_2 a_k \Psi_2(k-l))|^2 \quad (5)$$

where $L \geq 2p+1$ and $(\alpha_l)_{l\geq 0}$ are some positive weights. Thus the *Least Square estimator (LS estimator)* is given by solving the following optimization problem:

$$(\hat{G_1}, \hat{G_2}, \underline{\hat{a}}, \underline{\hat{b}}) = argmin\ J(\hat{r}) \quad (6)$$

3. IDENTIFICATION AND SEPARATION ALGORITHMS

In this section we present an algorithm to solve the identification issue and then we discuss the degenerate case of the BSS problem. Here we report only one algorithm we tried so far. A longer discussion will follow in an extended version of this report.

3.1. The Least Square Estimator

The Least Square estimator presented before is based on a gradient descent scheme. One issue related to this algorithm is how to choose an initial point $(G_1, G_2, \underline{a}, \underline{b})$. We present here an algorithm for obtaining this intialization.

The idea is the following: we identify first the time series $\{x(n)\},\ n = 1, \ldots, N_0$ as a "long" AR process, say $AR(2L_0)$, and then we approximate its spectral power density by a decomposition of the type (2).

ALGORITHM 2 (Initialization of $G_1, G_2, \underline{a}, \underline{b}$) *1. Choose $L_0 > p$ and find an $AR(2L_0)$ estimator of the time series $\{x(n)\}_{n=1\ldots N_0}$, say $\tilde{G}/\tilde{P}(z)$.*

2. For each partition of the $2L_0$ roots of $\tilde{P}$ into two groups of p zeros, construct $P_1(z)$ and $P_2(z)$ the polynomials corresponding to these zeros. Let $S(z)$ be the remainder polynomial in $\tilde{P}$, $\tilde{P} = P_1 P_2 S$. Find G_1 and G_2 that best approximate the equation:

$$\tilde{G}^2 = G_1^2 P_2(z)S(z)P_2(\frac{1}{z})S(\frac{1}{z}) + G_2^2 P_1(z)S(z)P_1(\frac{1}{z})S(\frac{1}{z}) \quad (7)$$

(we indicate below how to obtain G_1 and G_2)

3. Choose the best partition with respect to the approximation error and obtain the corresponding estimates for G_1, G_2, $\underline{a}$, $\underline{b}$.

To choose G_1 and G_2 in (7) we have tried both a Padé approximation [K.K87] and a least 2-norm solution. Both seem to work equaly fine. We describe here the 2-norm approximation.

Let us denote

$$P_1(z)S(z)P_1(\frac{1}{z})S(\frac{1}{z}) = \sum_{l=-2L_0+p}^{2L_0-p} f_l^1 z^l, \quad \text{and}$$

$$P_2(z)S(z)P_2(\frac{1}{z})S(\frac{1}{z}) = \sum_{l=-2L_0+p}^{2L_0-p} f_l^2 z^l$$

Then the 2-norm error computed on the unit circle in the complex plane is given by:

$$Error = \sum_{l=-2L_0+p}^{2L_0-p} |G_1^2 f_l^2 + G_2^2 f_l^1 - \tilde{G}^2 \delta_{l,0}|^2$$

Then we easily obtain a linear system in G_1^2 and G_2^2 by setting to zero the derivatives of the *Error* with respect to G_1^2, respectively G_2^2.

3.2. The Estimation Problem

Recall the problem is the following: we have two voice signals recorded by the same microphone and we want, based on this mixed signal, to estimate the original two signals. The solution we propose is represented in Figure 2 and consists of two stages: an identification part and a linear estimation part. For identification, we assume the two voices are approximated respectively by AR(p) processes and our task is to identify the proccesses parameters. For the linear estimation we tried both the Wiener filtering [Poo94] as well the causal part of the Wiener filter. It seems the causal part gives better results is terms of the sound quality. The Wiener filter formulae are given by:

$$F_1(z) = \frac{G_1^2 P_2(z)P_2(\frac{1}{z})}{G_1^2 P_2(z)P_2(\frac{1}{z}) + G_2^2 P_1(z)P_1(\frac{1}{z})}$$
$$F_2(z) = \frac{G_2^2 P_1(z)P_1(\frac{1}{z})}{G_1^2 P_2(z)P_2(\frac{1}{z}) + G_2^2 P_1(z)P_1(\frac{1}{z})} \quad (8)$$

and the causal parts are then:

$$F_1^c(z^{-1}) = \frac{G_1 P_2(z^{-1})}{T(z^{-1})} \qquad F_2^c(z^{-1}) = \frac{G_2 P_1(z^{-1})}{T(z^{-1})} \quad (9)$$

where $T(z)$ is the spectral factor in the factorization $T(z)T(\frac{1}{z}) = G_1^2 P_2(z)P_2(\frac{1}{z}) + G_2^2 P_1(z)P_1(\frac{1}{z})$.

The adaptation algorithm is the following:

ALGORITHM 3 (On-line Adaptation) *1. Initialize the parameter estimation on the first N_0 samples using the previous algorithm.*

2. Apply a couple of gradient descent steps to "polish" the approximation.

3. At each new sample, update the sampled autocovariance coefficient by using a rectangular window (or

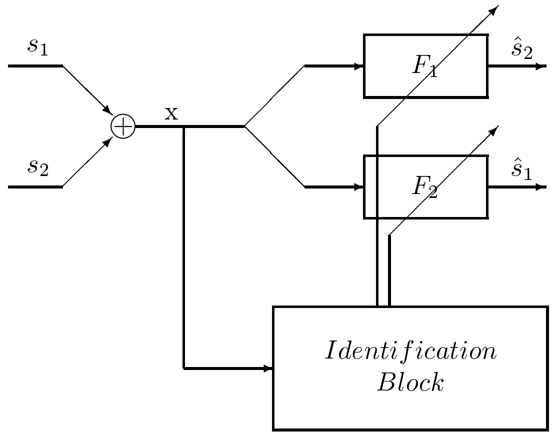

Figure 2: The Adaptive Estimation Diagram.

an exponential window) and apply a gradient step to adapt the estimation of $\hat{G}_1, \hat{G}_2, \underline{\hat{a}}, \underline{\hat{b}}$

4. Estimate $\hat{s_1}, \hat{s_2}$ *using the updated (causal) Wiener filters.*

Experimentally, the gradient correction at Step 3 seems not to track well the actual values of the parameters (obtained by an AR(p) estimator on the actual signals). Nonetheless, the more computationally expensive algorithm that simply applies the estimation on a sliding nonoverlapping window gives better results. Future work is needed to obtain a better on-line algorithm.

4. NUMERICAL EXPERIMENTS

We report experiments on both synthetic and voice data. First we describe AR identification experiments of singular mutivariate AR processes on synthetic data. Second we describe an application of the theory to the estimation of two voices from one scalar mixture. All experiments were performed in Matlab.

4.1. Experiments on Synthetic Data

We constructed two stable and independent AR(p) processes. Then we estimated the processes from their sum by filtering the observed signal with Wiener filters defined by parameters estimated by applying gradient descent with the initialization given by the Algorithm 2. Here we report only one set of results, those corresponding to $p = 4$.

We considered $N_0 = 500$ samples to estimate the autocovariance coefficients at various lags $\hat{r}(l)$; this corresponded to a 30ms window of a speech signal sampled at 16000Hz, on which the signal may be considered stationary (see [LJ93]). For the criterion (5) we took $L = 3p$ to avoid the contribution of the fluctuations in $\hat{r}(l)$ for large l. For diferent values of L_0 we obtained different initializations. Surprisingly, the best initialization has been given by the lowest value $L_0 = p = 4$.

In Figure 3 we plot the initial spectral power obtained with the Algorithm 2 using the 2-norm minimization, for several values of L_0. The first plot gives the Yule-Walker estimation [GH94] of the spectral powers of the two AR(4) processes. The theoretical spectral power is depicted in Figure 4, top plots, using the actual values of the parameters. In Figure 4 we also present the spectral power densities where the gradient algorithm converged after 1000 steps. Different initializations implied different limiting densities each of them corresponding to a different local minimum of the criterion J. For $L_0 = 4$ we show in Figure 5 the convergence of the parameters. Figure 6 plots the decimal logarithm of the criterion. Note how fast J decays during the first 40 steps. The limiting spectral power densities obtained in Figure 4, second row, approximates very well the original spectral densities. The gradient descent steps decreased the criterion to about 40 times less the initial value. The parameters of the two AR(4) proccesses used were the following: $G_1 = G_2 = 1$, $a_1 = 0.1, a_2 = -0.13, a_3 = -0.001, a_4 = 0.0012$ and $b_1 = -0.4, b_2 = 0.63, b_3 = -0.43, b_4 = 0.365$. The identification algorithm gave a better estimate for the second process which was the most powerful process.

From these experiments we conclude that the algorithm we proposed gives a fairly good estimate of the spectral power densities of the two sources. The parameters of the most powerful process are estimated better.

4.2. Experiments on Voice Data

We performed experiments with voices from the TIMIT database. The two voice signals (called A1 and A2 here) consist of 57448 samples at 16kHz sampling frequency (about 3 seconds of data). We tested how feasible the estimation problem is. We identified the two voices as AR(4) processes, directly on the actual signals, and then we estimated the two voices from their sum using the filters from equations (8) and (9). In Figure 7 we show the time-series of the original voices (upper graphs), of their sum (the middle plot), and the estimated signals (lower plots). We used $p = 4$ and $N_0 = 500$. The quality of the outputs is good for this rather low dimensional AR models we are approximat-

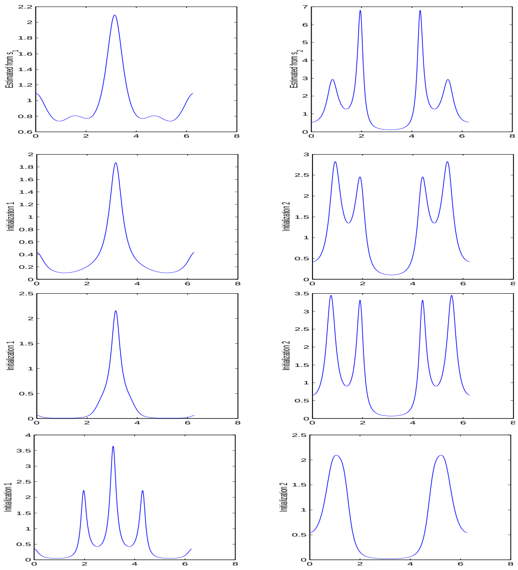

Figure 3: Spectral power densities for the Yule-Walker estimations (first row), initial spectral powers for $L_0 = 4$ (second row), $L_0 = 5$ (third row) and $L_0 = 6$ (fourth row)

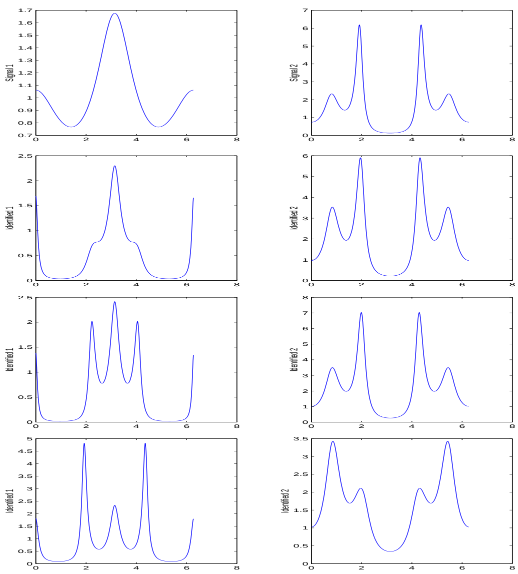

Figure 4: The theoretical spectral power densities (first row), limiting spectral powers for $L_0 = 4$ (second row), $L_0 = 5$ (third row) and $L_0 = 6$ (fourth row)

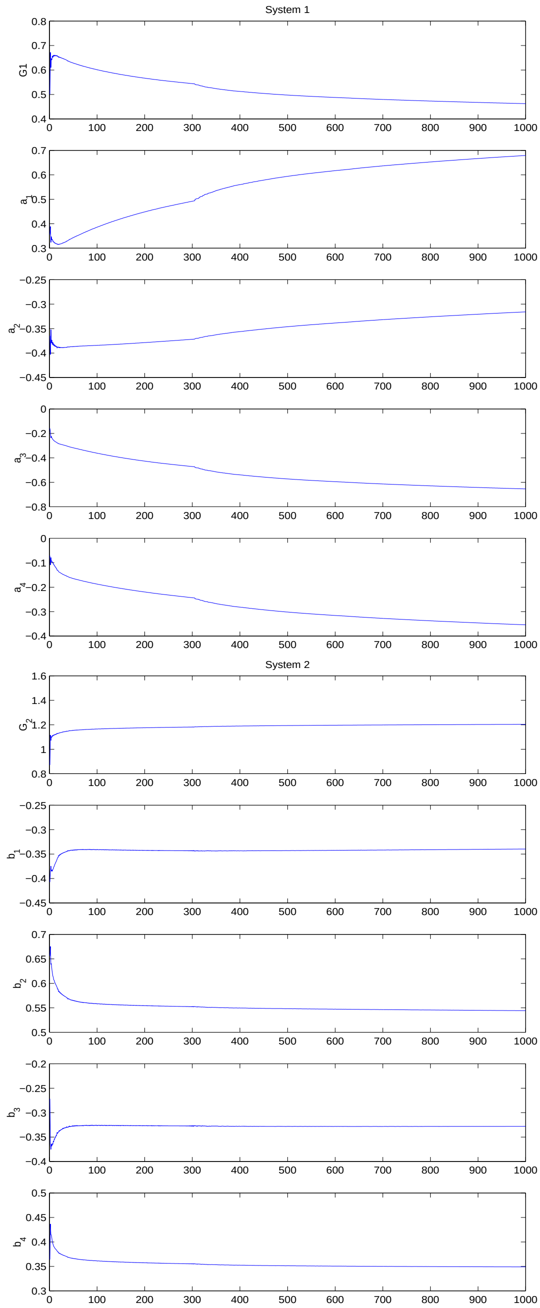

Figure 5: The parameters of the two processes.

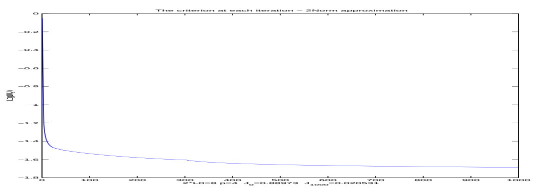

Figure 6: The criterion $\log_{10} J$

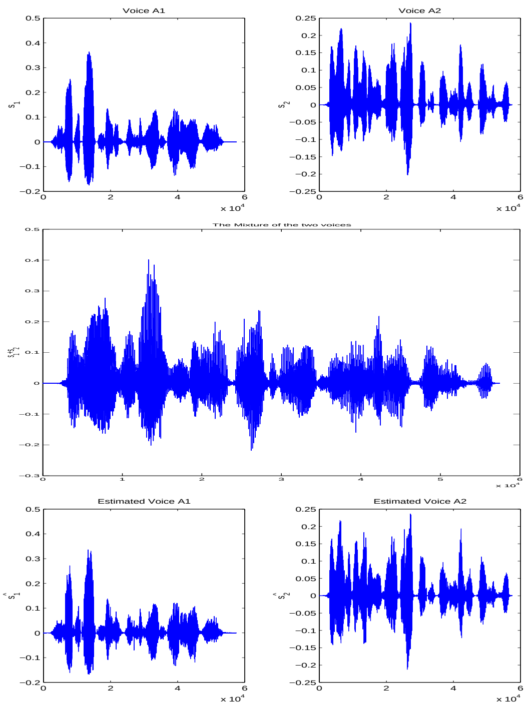

Figure 7: Voice experiments: source voices (upper row), mixture calculated as sum (middle row), and estimated outputs (bottom row).

ing voices with. We experimented with longer AR processes as well, but the quality of the outputs does not improve significantly. These experiments were meant to show that the estimation problem can be solved reasonably well when we identify the two voices as AR processes.

5. CONCLUSIONS

In this report we solved the identification problem of a sum of two independent AR processes. First we proved that this system is identifiable, next we deduced an estimator for the processes parameters and finally we presented a family of algorithms to implement this estimator. As a direct application we considered the degenerate case of the Blind Source Separation problem where a mixture of two voices is given, as recorded with one microphone. From this one measurement (more specifically one sequence of samples) we estimated the original two signals. This application raised the adaptation problem of the singular AR identification algorithm. We showed how to adapt the previous algorithm to an on-line procedure.

The present study shows that the second order statistics is sufficient for both the identification of singular multivariate AR processes of the particular form considered here, as well as estimation of independent signals in a scalar mixture of two voices when these voices can be well approximated by AR processes.

Future work will deal with various issues raised in the on-line implementation, such as faster and more reliable algorithms.

6. REFERENCES

[Ama96] S. Amari. Minimum mutual information blind separation. *Neural Computation*, 1996.

[BS95] A.J. Bell and T.J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[Car97] J.F. Cardoso. Infomax and maximum likelihood for blind source separation. *IEEE Signal Processing Letters*, 4(4):112–114, April 1997.

[Com94] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

[GH94] Arthur A. Giordano and Frank M. Hsu. *Least Square Estimation with Applications to Digital Signal Processing*. Springer-Verlag, 1994.

[JH91] C. Jutten and J. Herault. Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1):1–10, 1991.

[K.K87] K.Kumar. Identification of autoregeressive-moving average (arma) models using padé approximations. *Bull.Inst.Statis.Inst.*, Proc.46th Session, 52(2):377–389, 7 1987.

[LJ93] L.Rabiner and B-H. Juang. *Fundamentals of Speech Recognition*. PTR Prentince Hall, 1993.

[Poo94] H. Vincent Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 1994.

[PP96] B. A. Pearlmutter and L. C. Parra. A context-sensitive generalization of ica. In *International Conference on Neural Information Processing*, Hong Kong, 1996.

[S.M88] S.M.Kay. *Modern Spectral Estimation*. Prentice Hall, 1988.

[S.N96] S.Nakamori. Estimation of multivariate signals by output autocovariance data in linear discrete-time systems. *Math.and Comp.Moddeling*, 24(1):97–114, 7 1996.

[Tor96] K. Torkkola. Blind separation of convolved sources based on information maximization. In *IEEE Workshop on Neural Networks for Signal Processing, Kyoto, Japan*, 1996.

[WFO93] E. Weinstein, M. Feder, and A. Oppenheim. Multi-channel signal separation by decorrelation. *IEEE Trans. on Speech and Audio Processing*, 1(4):405–413, 1993.