Signal Separation of Nonlinear Time-Delayed Mixture: Time Domain Approach

W.L. Woo, S.S. Dlay and J.E. Hudson
School of Electrical, Electronic and Computer Engineering
University of Newcastle
Newcastle upon Tyne, NE1 7RU
UNITED KINDOM
Email: w.l.woo@ncl.ac.uk, s.s.dlay@ncl.ac.uk, john-hudson@lycosmax.co.uk

Abstract — **In this paper, a novel algorithm is proposed to solve blind signal separation of nonlinear time-delayed mixtures of statistically independent sources. Both mixing and nonlinear distortion are included in the proposed model. Maximum Likelihood (ML) approach is developed to estimate the parameters in the model and this is formulated within the framework of the generalized Expectation-Maximization (EM) algorithm. Adaptive polynomial basis expansion is used to estimate the nonlinearity of the mixing model. In the E-step, the sufficient statistics associated with the source signals are estimated while in the M-step, the parameters are optimized by using these statistics. Generally, the nonlinear distortion renders the statistics intractable and difficult to be formulated in a closed form. However, in this paper it is proved that with the use of Extended Kalman Smoother (EKS) around a linearized point, the M-step is made tractable and can be solved by linear equations.**

I. INTRODUCTION

Linear blind signal separation has been well learnt so far and a plethora of methods has been proposed. Most methods are based on higher order statistics which require non-Gaussian source signals [1-3]. In addition, these methods yield good performances only if the underlying assumption of the mixture is linear. However, in the practical applications e.g. speech and imaging processing, one of the fundamental issues is to separate the received signals through noisy environment and because the receiving elements such as the microphones [4] or antennae are subject to nonlinear distortion, a more accurate representation of the received signals must be developed to account for the existence of the nonlinearity. The need of accurate representation of the nonlinear distorted signals makes nonlinear blind source separation (NBSS) well researched recently. However, majority of the previous works focus on the instantaneous mixing [5-9]. These techniques are based on statistical neural network approaches [5-7] and functional analysis theory [8-9]. Although powerful, these techniques rely on the condition that the time is aligned during the mixing. To the knowledge of the authors, there is no method that addresses the nonlinear time-delayed mixture. The closest research to this work is [10] which treat a rather general convolutive mixing but the computational complexity is relatively high. Hence, most existing methods will fail to yield satisfactory performance in practical applications.

In this paper, the noisy nonlinear time-delayed mixing problem is considered for the first time and the objective is to provide optimal estimation of the source signals, the mixing process parameters and the parameters of the additive noise. The contribution of this paper is to provide a state space representation of the nonlinear time delayed mixtures of quasi-stationary signals and to formalize a generalized EM framework for estimating the source signals, the mixing process parameters and the additive noise. The generalized EM algorithm is derived where the post-nonlinearity is approximated by a set of polynomials whose coefficients are updated as part of the mixing process parameters. In the proposed algorithm, the sufficient statistics of the source signals are inferred in the E-step and the model parameters are updated in the M-step.

II. PROPOSED MODEL

The nonlinear time-delayed mixture of sources can be expressed as:

$$x_i(t) = f_i\left(\sum_{j=1}^{N_s} h_{ij} s_j(t-\tau_{ij})\right) + n_i(t) \quad (1)$$

where $i = 1, 2, \ldots, M$, $f_i(\cdot)$ is a one-to-one nonlinear function and τ_{ij} is the relative time delay of the mixing between the j^{th} source signal $s_j(t)$ and the i^{th} receiver $x_i(t)$. In vector form, (1) can be formulated as

$$\mathbf{x}_t = \mathbf{f}\left(\sum_{l=0}^{L} \mathbf{H}_l \mathbf{s}_{t-l}\right) + \mathbf{n}_t \quad (2)$$

with $\mathbf{x}_t = \left[x_1(t)\ x_2(t)\ \cdots\ x_M(t)\right]^T$, $\mathbf{H}_l$ is the time-delayed mixing matrix, $\mathbf{s}_t = \left[s_1(t)\ s_2(t)\ \cdots\ s_{N_s}(t)\right]^T$ is the vector of the source signals at time t , $\mathbf{f}(.) = \left[f_1(\cdot)\quad f_2(\cdot)\quad \cdots\quad f_M(\cdot)\right]^T$ and $L = \max_{ij}\{\tau_{ij}\}$. The additive noise $\mathbf{n}_t$ is assumed to be Gaussian. The nonlinear time-delayed mixing model in (2) can be extended to include temporally correlated sources as follows:

$$\mathbf{s}_t = \mathbf{G}_t \mathbf{s}_{t-1} + \mathbf{w}_t \quad (3)$$

978-0-7695-3594-4/09 $25.00 © 2009 IEEE
DOI 10.1109/ICSAP.2009.11

Authorized licensed use limited to: Newcastle University. Downloaded on September 1, 2009 at 09:29 from IEEE Xplore. Restrictions apply.

where $\mathbf{G}$ is the state evolution matrix which contains the autoregressive (AR) coefficients. The prior probability distribution over initial states of the source signals $\mathbf{s}_1$ is taken to be Gaussian with mean $\boldsymbol{\mu}_0$ and covariance $\boldsymbol{\Lambda}_0$. The Gaussian noise vector $\mathbf{w}$ and $\mathbf{n}$ have zero mean and covariance matrices $\mathbf{Q}$ and $\mathbf{R}$, respectively. In (2) and (3), the nonlinear time-delayed mixture is constructed in two parts. First, the temporal correlation of the hidden source signal is represented by using higher order AR process model. The AR(*J*) process is adopted to implement the temporal correlation of the non-stationary signal by stacking the state variables to form the state vector, which is expressed as:

$$\mathbf{s}_{i,t} = \mathbf{G}_{i,t}\mathbf{s}_{i,t-1} + \mathbf{w}_t \tag{4}$$

where

$$\mathbf{s}_{i,t} = [s_{i,t} \quad s_{i,t-1} \quad \cdots \quad s_{i,t-J+1}]^T, \mathbf{w}_t = \begin{bmatrix} w_t & \mathbf{0} \end{bmatrix}^{\mathrm{T}}$$

$$\mathbf{G}_{i,t} = \begin{bmatrix} \mathbf{g}_{i,t} \\ \mathbf{I} \quad \mathbf{0} \end{bmatrix}, \mathbf{g}_{i,t} = [g_{i,t,1} \quad g_{i,t,2} \quad \cdots \quad g_{i,t,L}] \tag{5}$$

The vector for all sources is now expressed as:

$$\mathbf{s}_t^{\mathrm{T}} = [\mathbf{s}_{1,t}^{\mathrm{T}} \quad \mathbf{s}_{2,t}^{\mathrm{T}} \quad \cdots \quad \mathbf{s}_{N_s,t}^{\mathrm{T}}] \tag{6}$$

The parameters of the mixing process are defined in the following to satisfy the statistical independence between the source signals:

$$\mathbf{G}_t = diag[\mathbf{G}_{1,t} \quad \mathbf{G}_{2,t} \quad \cdots \quad \mathbf{G}_{N_s,t}]$$

$$\mathbf{Q} = diag[\mathbf{Q}_1 \quad \mathbf{Q}_2 \quad \cdots \quad \mathbf{Q}_{N_s}], (\mathbf{Q}_i)_{j_1 j_2} = \begin{cases} q_i & j_1 = j_2 = 1 \\ 0 & otherwise \end{cases} \tag{7}$$

$\boldsymbol{\Lambda}_0$ and $\boldsymbol{\mu}_0$ are defined in a similar fashion.

Second, the nonlinear mixture and nonlinearity distortion are introduced into the model. To represent the time-delayed mixture in the proposed model, the observation matrix is extended to the full matrix of filters, which can be expressed in the following:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_{11} & \cdots & \mathbf{h}_{1N_s} \\ \vdots & \ddots & \vdots \\ \mathbf{h}_{M1} & \cdots & \mathbf{h}_{MN_s} \end{bmatrix}, \mathbf{h}_{ij} = \begin{bmatrix} h_{ij,1} & h_{ij,2} & \cdots & h_{ij,L} \end{bmatrix} \tag{8}$$

In above, $\mathbf{h}_{ij}$ represents all the paths between sensor *i* and source *j* (*L*=*J*). For the case where the signals are quasi-stationary, they can be segmented into windows in which the source signals can be assumed to be stationary [11]. Hence, the nonlinear time-delayed model can be expressed as:

$$\begin{aligned} \mathbf{s}_t^n &= \mathbf{G}_t^n \mathbf{s}_{t-1}^n + \mathbf{w}_t^n \\ \mathbf{x}_t^n &= \mathbf{f}(\mathbf{H}\mathbf{s}_t^n) + \mathbf{n}_t^n, \quad n = 1, 2, \ldots, N \end{aligned} \tag{9}$$

Hence, a total number of *N* segments are observed. In the following section, the EM algorithm is generalized to estimate the parameters in the proposed model and the Kalman recursion is used to estimate the relevant statistics while the polynomials are used to estimate the nonlinear function.

III. PARAMETER ESTIMATION

For linear mixing model, the EM algorithm is the standard algorithm used to estimate the parameters where the Kalman smoother is used to infer the hidden states. The algorithm of Kalman smoother contains two parts: a forward recursion which used the observation from $\mathbf{x}_1$ to $\mathbf{x}_t$, named Kalman filter, and a backward recursion which uses the observation from $\mathbf{x}_T$ to $\mathbf{x}_{t+1}$ where T is the length of the data. However, with a model defined by (9) the conditional densities are in general non-Gaussian and can lead to intractable solution. To deal with this problem, the generalized EM algorithm is derived with the use of extended Kalman smoother and polynomial nonlinearity estimator. Firstly, the likelihood function can be defined as:

$$L(\theta) = \log p(\mathbf{x} \mid \theta) = \log \int d\mathbf{s} p(\mathbf{x}, \mathbf{s} \mid \theta) \tag{10}$$

where θ denotes all the parameters in the proposed model (9). Following Jensen's inequality:

$$\begin{aligned} L(\theta) &= \log p(\mathbf{x} \mid \theta) = \log \int d\mathbf{s} p(\mathbf{x}, \mathbf{s} \mid \theta) \\ &= \log \int d\mathbf{s} \frac{\hat{p}(\mathbf{s})}{\hat{p}(\mathbf{s})} p(\mathbf{x}, \mathbf{s} \mid \theta) \geq \int d\mathbf{s} \hat{p}(\mathbf{s}) \log \frac{p(\mathbf{x}, \mathbf{s} \mid \theta)}{\hat{p}(\mathbf{s})} \\ &= \varepsilon_1(\theta, \hat{p}) - \varepsilon_2(\hat{p}) = \mathcal{F}(\theta, \hat{p}) \end{aligned} \tag{11}$$

where $\varepsilon_1(\theta, \hat{p}) \equiv \int d\mathbf{s} \hat{p}(\mathbf{s}) \log p(\mathbf{x}, \mathbf{s} \mid \varepsilon)$ and $\varepsilon_2(\hat{p}) \equiv \int d\mathbf{s} \hat{p}(\mathbf{s}) \log \hat{p}(\mathbf{s})$. It is well known that in the E-step $\mathcal{F}(\theta, \hat{p})$ is maximised when $\hat{p}(\mathbf{s})$ is chosen to be exactly the conditional distribution of $\mathbf{s}$ with the parameters estimated in the previous iteration, $\hat{p}(\mathbf{s}) = p(\mathbf{s} \mid \mathbf{x}, \theta^{[k]})$ at which point the bound becomes an equality. Then in the M-step, $\varepsilon_1(\theta, \hat{p})$ is maximized with respect to θ. Each iteration cannot decrease $\mathcal{F}(\theta, \hat{p})$.

(i) E-step

The relevant statistics of the posterior distribution of the source signals $p(\mathbf{s}_t \mid \mathbf{x}_{1:T}, \theta^{[k]})$ are inferred in the E-step and represented in the form of the parameters obtained in the previous iteration. For nonlinear model in (9), this is achieved by using the Extended Kalman Smoother (EKS). The EKS applies the basic Kalman Smoother to output the relevant first and second order statistics based on a linearized point of the nonlinear system. Here, the nonlinearity is linearized at the mean of the current filtered (not smoothed) state $\hat{\mathbf{s}}_t$. Hence, after the linearization, the model (9) becomes

Authorized licensed use limited to: Newcastle University. Downloaded on September 1, 2009 at 09:29 from IEEE Xplore. Restrictions apply.

$$\mathbf{s}_t^n = \mathbf{G}_t^n \mathbf{s}_{t-1}^n + \mathbf{w}_t^n$$
$$\mathbf{x}_t^n = \mathbf{f}(\mathbf{H}\hat{\mathbf{s}}_{t|t-1}^n) + \mathbf{C}_{\hat{\mathbf{s}}_{t|t-1}^n}(\mathbf{s}_t^n - \hat{\mathbf{s}}_{t|t-1}^n) + \mathbf{n}_t^n \quad (12)$$

where the matrix $\mathbf{C}$ is defined by the derivative of the vector-valued function $\mathbf{f}$, at point $\hat{\mathbf{s}}_{t|t-1}^n$

$$\mathbf{C}_{\hat{\mathbf{s}}_{t|t-1}^n} \equiv \left.\frac{\partial \mathbf{f}}{\partial \mathbf{s}_t^n}\right|_{\mathbf{s}_t^n = \hat{\mathbf{s}}_{t|t-1}^n} \quad (13)$$

Thus, after the linearization, with the noise distribution and the prior distribution of the initial of the hidden state is Gaussian, the conditional distribution of the hidden states given the output at every time will also be Gaussian. Hence, the Kalman Smoother can be applied on the model (12) where the generalized EM algorithm is used to infer the conditional distribution. The inferred first order statistics is the source conditional mean $\hat{\mathbf{s}}_t^n$ for segment n, which is expressed as:

$$\hat{\mathbf{s}}_t^n = \langle \mathbf{s}_t^n \rangle \quad (14)$$

where the $\langle . \rangle$ denotes for the integral over the source posterior $p(\mathbf{s}_t \mid \mathbf{x}_{1:T}, \theta^{[k]})$. The second order statistics inferred by EKS is the autocorrelation matrix along with the time lag-one autocorrelation matrix of source i for segment n, which is expressed as $\mathbf{A}_{ii,tt}^n$ and $\mathbf{A}_{ii,t(t-1)}^n$, respectively:

$$\mathbf{A}_{ii,tt}^n \equiv \langle \mathbf{s}_{i,t}^n (\mathbf{s}_{i,t}^n)^{\mathrm{T}} \rangle \equiv \begin{bmatrix} \mathbf{a}_{ii,1,tt}^n & \mathbf{a}_{ii,2,tt}^n & \cdots & \mathbf{a}_{ii,L,tt}^n \end{bmatrix}^{\mathrm{T}} \quad (12)$$

$$\mathbf{A}_{ii,t(t-1)}^n \equiv \langle \mathbf{s}_{i,t}^n (\mathbf{s}_{i,t-1}^n)^T \rangle = \begin{bmatrix} \mathbf{a}_{ii,1,t(t-1)}^n & \mathbf{a}_{ii,2,t(t-1)}^n & \cdots & \mathbf{a}_{ii,L,t(t-1)}^n \end{bmatrix}^T \quad (13)$$

Here, the first element in $\mathbf{a}_{ii,1,tt}^n$ is defined as $a_{ii,1,tt}^n$. With the relevant statistics are inferred in current E-step, the model parameters can be updated to maximize the likelihood in (11) in the following M-step.

(ii) M-step

In the M-step, $\varepsilon_1(\theta)$ in (11) is maximized with respect to all the parameters in the proposed model. This involves using the statistics obtained from the previous E-step:

$$\varepsilon_1(\theta,\bar{p}) = -\frac{1}{2}\sum_{n=1}^{N}\Bigg[\sum_{i=1}^{N_S}\log\det\mathbf{\Lambda}_{0,i}^n + (T-1)\sum_{i=1}^{N_S}\log q_i^n + T\log\det\mathbf{R}$$
$$+\sum_{i=1}^{N_S}\langle(\mathbf{s}_{i,1}^n-\boldsymbol{\mu}_{0,i}^n)^{\mathrm{T}}(\mathbf{\Lambda}_{0,i}^n)^{-1}(\mathbf{s}_{i,1}^n-\boldsymbol{\mu}_{0,i}^n)\rangle + \sum_{t=2}^{T}\sum_{i=1}^{N_S}\langle\frac{1}{q_i^n}(s_{i,t}^n-(\mathbf{g}_{i,t}^n)\mathbf{s}_{i,t-1}^n)^2\rangle$$
$$+\sum_{t=1}^{T}\langle(\mathbf{x}_t^n - \mathbf{f}(\mathbf{H}\hat{\mathbf{s}}_{t|t-1}^n) - \mathbf{C}_{\hat{\mathbf{s}}_{t|t-1}^n}(\mathbf{s}_t^n - \hat{\mathbf{s}}_{t|t-1}^n))^{\mathrm{T}}\mathbf{R}^{-1}$$
$$(\mathbf{x}_t^n - \mathbf{f}(\mathbf{H}\hat{\mathbf{s}}_{t|t-1}^n) - \mathbf{C}_{\hat{\mathbf{s}}_{t|t-1}^n}(\mathbf{s}_t^n - \hat{\mathbf{s}}_{t|t-1}^n))\rangle\Bigg] \quad (14)$$

Optimizing (14) with respect to the model parameters, the new estimator for segment n of source i is given by the following closed form equations:

$$\boldsymbol{\mu}_{0,i}^n = \hat{\mathbf{s}}_{i,1}^n$$
$$\mathbf{\Lambda}_{0,i}^n = \mathbf{A}_{ii,11}^n - \boldsymbol{\mu}_{0,i}^n(\boldsymbol{\mu}_{0,i}^n)^{\mathrm{T}}$$
$$\mathbf{g}_{i,t}^n = \left(\mathbf{a}_{ii,1,t(t-1)}^n\right)\left[\mathbf{A}_{ii,(t-1)(t-1)}^n\right]^{-1}$$
$$q_i^n = \frac{1}{T-1}\left[\sum_{t=2}^{T}\left(a_{ii,1,tt}^n - \mathbf{g}_i^n(\mathbf{a}_{ii,1,t(t-1)}^n)^{\mathrm{T}}\right)\right] \quad (15)$$

The reconstructions of $\boldsymbol{\mu}_0^n, \mathbf{\Lambda}_0^n, \mathbf{G}^n, \mathbf{Q}^n$ can be achieved following the definitions in Section II. However, the estimators for $\mathbf{H}$ and $\mathbf{R}$ include the statistics of all observed segments and they are much different from the ones for linear deconvolution and more complex. Because the new estimator for $\mathbf{H}$ cannot be expressed in a closed form, the gradient ascent algorithm is used to update $\mathbf{H}$, the new estimator for $\mathbf{h}_i$ is given by:

$$\mathbf{h}_{i,t+1} = \mathbf{h}_{i,t} + \alpha\frac{\partial\varepsilon_1(\theta,\bar{p})}{\partial\mathbf{h}_i}\Big|_{\mathbf{h}_i=\mathbf{h}_{i,t}} \quad (16)$$

where

$$\frac{\partial\varepsilon_1(\theta,\bar{p})}{\partial\mathbf{h}_i} \approx \sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{M} r_{ik}^{-1}\left(-x_{k,t}^n + f_k + f_k'\,\mathbf{h}_k\,\mathbf{e}_t^n\right)f_i'(\hat{\mathbf{s}}_{t|t-1}^n)^{\mathrm{T}}$$

In above, $\mathbf{e}_t^n = (\hat{\mathbf{s}}_t^n - \hat{\mathbf{s}}_{t|t-1}^n)$ and $\mathbf{h}_i = \begin{bmatrix} h_{i1} & \cdots & h_{iN_S}\end{bmatrix}$, α is the learning rate, r_{ij}^{-1} is ij^{th} element of the inverse matrix of $\mathbf{R}$, f_i is $f_i\left(\sum\mathbf{h}_{ij}\hat{\mathbf{s}}_{j,t|t-1}^n\right)$ for short, and f' is the first-order derivative with respect to the argument. The new estimator for $\mathbf{R}$ is

$$\mathbf{R} = \frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\Bigg[\mathbf{x}_t^n(\mathbf{x}_t^n)^{\mathrm{T}} - \mathbf{x}_t^n\mathbf{f}^{\mathrm{T}} - \mathbf{x}_t^n(\mathbf{e}_t^n)^{\mathrm{T}}\mathbf{C}_{\hat{\mathbf{s}}_{t|t-1}^n}^{\mathrm{T}} - \mathbf{f}(\mathbf{x}_t^n)^{\mathrm{T}} + \mathbf{f}\mathbf{f}^{\mathrm{T}}$$
$$+\mathbf{C}_{\hat{\mathbf{s}}_{t|t-1}^n}\left(\mathbf{A}_{tt}^n - \hat{\mathbf{s}}_t^n(\hat{\mathbf{s}}_{t|t-1}^n)^{\mathrm{T}} - \hat{\mathbf{s}}_{t|t-1}^n(\hat{\mathbf{s}}_t^n)^{\mathrm{T}} + \hat{\mathbf{s}}_{t|t-1}^n(\hat{\mathbf{s}}_{t|t-1}^n)^{\mathrm{T}}\right)\mathbf{C}_{\hat{\mathbf{s}}_{t|t-1}^n}^{\mathrm{T}}\Bigg]$$
$$+\mathbf{f}(\mathbf{e}_t^n)^{\mathrm{T}}\mathbf{C}_{\hat{\mathbf{s}}_{t|t-1}^n}^{\mathrm{T}} - \mathbf{C}_{\hat{\mathbf{s}}_{t|t-1}^n}(\mathbf{e}_t^n)(\mathbf{x}_t^n)^{\mathrm{T}} + \mathbf{C}_{\hat{\mathbf{s}}_{t|t-1}^n}(\mathbf{e}_t^n)\mathbf{f}^{\mathrm{T}} \quad (17)$$

where $\mathbf{A}_{tt}^n$ is diagonal autocorrelation matrix for $\mathbf{s}_t^n$ and $\mathbf{f}$ is $\mathbf{f}(\mathbf{H}\hat{\mathbf{s}}_{t|t-1}^n)$ for short. Since the nonlinearity $\mathbf{f}$ cannot be accessed directly, a self-adaptive algorithm for estimating $\mathbf{f}$ is essential. Here, a set of polynomial is used to estimate the

Authorized licensed use limited to: Newcastle University. Downloaded on September 1, 2009 at 09:29 from IEEE Xplore. Restrictions apply.

nonlinearity distortion functions $\mathbf{f}$. The estimation only requires the statistics obtained from the E-step. Thus, for every scalar function of $\mathbf{f}$, $f(.)$ can be approximated by polynomials as

$$f_i(\mathbf{h}_i\hat{\mathbf{s}}^n_{t|t-1}) = \sum_{z=0}^{Z} b_{i,z}(\mathbf{h}_i\hat{\mathbf{s}}^n_{t|t-1})^z = \mathbf{b}_i\mathbf{v}_i^n \tag{18}$$

where Z represents the order and $b_{i,z}$ being the coefficients of the polynomial, $\mathbf{b}_i = \left[b_{i,0} \ \cdots \ b_{i,Z}\right]$, $\mathbf{v}_i = \left[1 \ \ \mathbf{h}_i\hat{\mathbf{s}}^n_{t|t-1} \ \ \cdots \ \ (\mathbf{h}_i\hat{\mathbf{s}}^n_{t|t-1})^Z\right]^{\mathrm{T}}$. The polynomial coefficient $b_{i,z}$ can be updated as one of the model parameters by maximizing $\varepsilon_1(\theta,\hat{p})$. With the learning rate η, the gradient ascent algorithm for $b_{i,z}$ is expressed as

$$b_{i,z,t+1} = b_{i,z,t} + \eta\,\frac{\partial\varepsilon_1(\theta,\hat{p})}{\partial b_{i,z}}\Big|_{b_{i,z}=b_{i,z,t}} \tag{19}$$

where for $b_{i,0}$ we have

$$\frac{\partial\varepsilon_1(\theta,\hat{p})}{\partial b_{i,0}} \approx -\sum_{n=1}^{N}\sum_{t=1}^{T}\left[-\left(\sum_{k_1=1}^{M} x^n_{k_1,t} r^{-1}_{k_1 i}\right)+\left(\sum_{k_2=1}^{M} (\mathbf{b}_{k_2}\mathbf{v}^n_{k_2}) r^{-1}_{ik_2}\right)\right] \tag{20}$$

As for $b_{i,z}, 1 \le z \le Z$,

$$\frac{\partial\varepsilon_1(\theta,\hat{p})}{\partial b_{i,z}} \approx -\sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{M} r^{-1}_{ki}\left(\mathbf{b}_k\mathbf{v}^n_k - x^n_{k,t}\right)(\mathbf{h}_i\hat{\mathbf{s}}^n_{t|t-1})^z \tag{21}$$

Thus, all new estimators for parameters are obtained and represented with the statistics obtained in the E-step. The EM algorithm alternates between the E-step and M-step until it converges to a steady solution, which normally requires only a few iterations.

IV. RESULTS

To investigate the efficacy of the proposed approach, the algorithm is tested on a post-nonlinear convolutive mixture of two independent speech signals with additive Gaussian noise under different settings of the signal-to-noise ratio (SNR) environment. The parameters J and L are set to be 3. The mixing is randomly selected as:

$$\begin{aligned} x_1(t) &= \tanh\left(s_1(t-\tau_{11}) + 0.7 s_2(t-\tau_{12})\right) + n_1(t) \\ x_2(t) &= \left(-0.8 s_1(t-\tau_{21}) + s_2(t-\tau_{22})\right)^3 + n_2(t) \end{aligned} \tag{22}$$

where $\tau_{11} = \tau_{22} = 1$, $\tau_{12} = 3$, $\tau_{21} = 5$ and additive Gaussian noise is added to the mixture to obtain the required SNR. The function $f_1(\cdot) = \tanh(\cdot)$ and $f_2(\cdot) = (\cdot)^3$ in (22) (from (1)) are selected as the post-nonlinear distortions since $f_1(\cdot)$ is bounded whereas $f_2(\cdot)$ unbounded and this selection is taken merely to study the performance of the proposed algorithm under two different forms of nonlinearity.

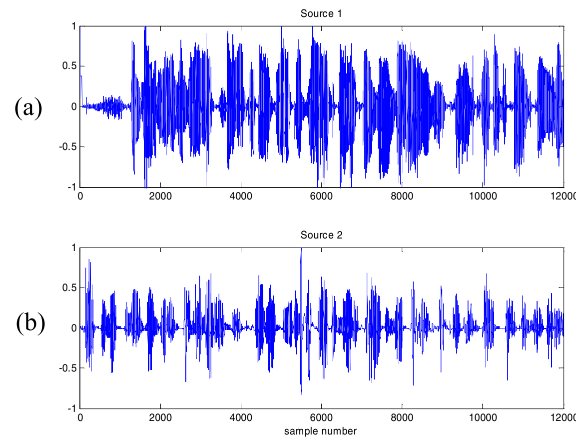

Figure 1: Original sources. (a) $s_1(t)$. (b) $s_2(t)$.

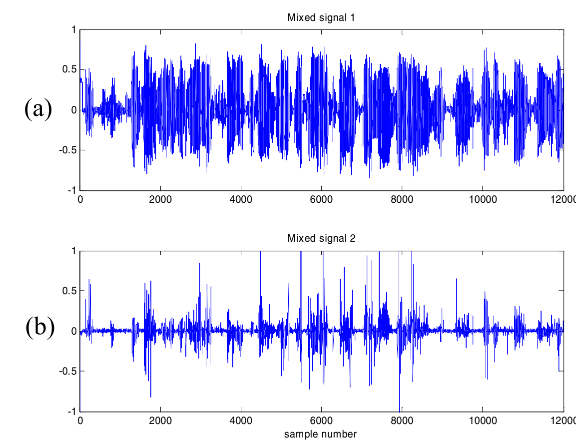

Figure 2: Mixed signals. (a) $x_1(t)$. (b) $x_2(t)$.

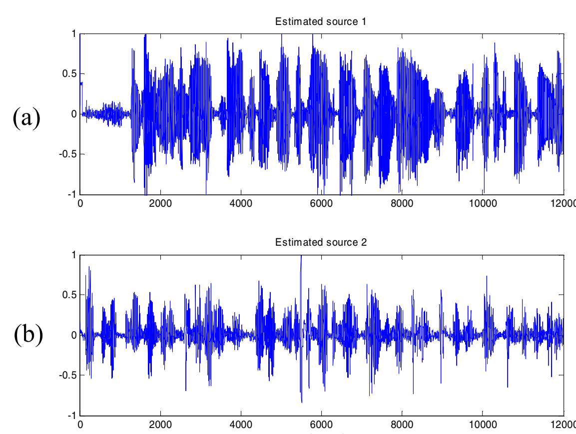

Figure 3: Estimated sources. (a) $\hat{s}_1(t)$. (b) $\hat{s}_2(t)$.

Authorized licensed use limited to: Newcastle University. Downloaded on September 1, 2009 at 09:29 from IEEE Xplore. Restrictions apply.

Figures 1(a)-(b) show the original speech sources which have been captured and sampled at 16kHz. The resulting signals are segmented into windows of time length $T = 60$ samples. This is to allow both signals to be stationary within the window but its statistics being variable from window to window; thus satisfying the characteristic of quasi-stationarity in the proposed algorithm. Figures 2(a)-(b) show the result of the mixing according to (24). Finally, the separation results are displayed on Figures 3(a)-(b) where we can visually inspect that the estimated signals resembles very closely to the original speech signals in Figure 1. In addition, we have tested the proposed algorithm under different signal-to-noise ratio (SNR). All model parameters are estimated by the proposed algorithm. To compare the performance of the signal separation, we have compared the proposed algorithm with two other algorithms: DUET algorithm [12] (which assumes the time-delayed mixing to be linear) and Variational Bayesian nonlinear separation algorithm in [13] (which treats the mixing to be time aligned i.e. all time delays τ_{ij} in the mixing to be identical). In addition, to quantify the separation performance a signal to interference ratio (SIR) is used:

$$\mathrm{SIR} = 10\log_{10}\left(\frac{P_{11}+P_{22}}{P_{12}+P_{21}}\right)\ \mathrm{dB} \tag{23}$$

where P_{ij} is the power of the signal which contributes the i^{th} estimated source signal to the j^{th} original source signal where the normalized cross-autocorrelation is used. The obtained results are shown in the Table 1 where the improvements of the proposed algorithm over DUET algorithm [12] and Variational Bayesian algorithm [13] are dramatically high under low SNR. The results show that linear algorithm fails miserably and thus points out the importance of incorporating a nonlinear model in the case where the observed mixture has been distorted. This also shows that the proposed algorithm is robust under high level of noise in a post nonlinearly mixed sources environment.

SNR / **SIR (dB)**	10dB	20dB	30dB
DUET algorithm [12]	6.5	8.7	9.8
Variational Bayesian algorithm [13]	4.2	5.6	8.2
Proposed algorithm	9.8	11.2	12.7

Table 1: Performance comparisons

V. CONCLUSIONS

A new statistical approach for post nonlinear time-delayed mixing model of quasi-stationary signals has been proposed. The proposed model is based on maximizing the log-likelihood of the observations and the estimation of the model parameters is facilitated by the generalized EM algorithm. In the proposed EM framework, the Extended Kalman Smoother is used to infer the source signals in the E-step while the parameters are updated iteratively in the M-step. Results show that for given nonlinear speech dataset, the proposed algorithm performs significantly better than the linear algorithm and nonlinear algorithm that fail to take into account the time-delays of the mixing.

REFERENCES

[1] S.C. Alvarez, A. Cichocki and L.C. Ribas, "An Iterative Inversion Approach to Blind Source Separation," *IEEE Trans. on Neural Networks*, vol. 11, no.6, pp. 1423 – 1437, 2000.

[2] M. Castella, J.C. Pesquet and A.P. Petropulu, "A Family of Frequency- and Time-Domain Contrasts for Blind Separation of Convolutive Mixtures of temporally Dependent Signals," *IEEE Trans. on Signal Processing*, vol. 53, no.1, 2005.

[3] H. Attias and C.E. Schreiner, "Blind source separation and deconvolution: the dynamic component analysis algorithm," *Neural Computation*, vol. 10 no. 6. pp. 1373-1424, 1998.

[4] T.F. Quatieri, D.A. Reynolds and G.C. O'Leary, "Estimation of handset nonlinearity with application to speaker recognition," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 5, pp. 567-584, Sept. 2000.

[5] W.L. Woo and L.C. Khor: 'Blind Restoration of Nonlinearly Mixed Signals using Multilayer Polynomial Neural Network', *IEE Proc. on Vision, Image and Signal Processing*, vol. 151, no.1, pp. 51-61, 2002.

[6] C. Jutten, M. Babaie-Zadeh and S. Hosseini, "Three easy ways for separating nonlinear mixtures?," *Signal Processing*, vol. 84, no. 2, pp.217-229, 2004.

[7] W.L. Woo and S.S. Dlay: 'Nonlinear Blind Source Separation using a Hybrid RBF-FMLP Network', *IEE Proc. on Vision, Image and Signal Processing*, vol. 152, no. 2, 173-183, 2005

[8] W.L. Woo and S.S. Dlay, "Neural Network Approach to Blind Signal Separation of Mono-nonlinearly Mixed Signals," *IEEE Trans. on Circuits and System* - Part 1, vol. 52, no. 2, pp. 1236-1247, 2005.

[9] P. Gao, W.L. Woo and S.S. Dlay, "Nonlinear Signal Separation for Multi-Nonlinearity Constrained Mixing Model," *IEEE Trans. on Neural Networks*, vol. 17, no. 3, pp. 796-802, 2006.

[10] J. Zhang, W.L. Woo and S.S. Dlay, "Blind Source Separation of Post-Nonlinear Convolutive Mixture," *IEEE Trans. on Audio, Speech and Language Processing,* vol. 15, no. 8, pp. 2311-2330, 2007.

[11] L. Parra and C. Spence, "Convolutive Blind Separation of non-stationary Sources," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no.3, pp. 320-327, 2000.

[12] O. Yilmaz and S. Rickard, "Blind Separation of Speech Mixtures via Time-Frequency Masking," *IEEE Trans on Signal Processing*, vol. 52, no. 7, pp. 1830-1847, 2004.

[13] A. Honkela, H. Valpola, A. Illin and J Karhunen "Blind Separation of Nonlinear Mixtures by Variational Bayesian Learning," *Digital Signal Processing*, vol. 17, no. 5, pp. 914-936, 2007.

Authorized licensed use limited to: Newcastle University. Downloaded on September 1, 2009 at 09:29 from IEEE Xplore. Restrictions apply.