Extraction of Unique Independent Components for Nonlinear Mixture of Sources

P. Gao
Newcaslte University, Newcastle upon Tyne, UK
Email: pei.gao@ncl.ac.uk

L.C. Khor, W.L. Woo and S.S. Dlay
Newcaslte University, Newcastle upon Tyne, UK
Email: {l.c.khor, w.l.woo, s.s.dlay}@ncl.ac.uk

***Abstract*—In this paper, a neural network solution to extract independent components from nonlinearly mixed signals is proposed. Firstly, a structurally constrained mixing model is introduced to extend the recently proposed mono-nonlinearity mixing model, allowing that different nonlinear distortion are applied to each source signal. Based on this nonlinear mixing model, a novel demixing system characterized by polynomial neural network is then proposed for recovering the original sources. The parameter learning algorithm is derived mathematically based on the minimum mutual information principle. It is shown that unique extraction of independent components can be achieved by optimizing the mutual information cost function under both model structure and signal constraints. In this framework, the theory of series reversion is developed with the aim to perform dual optimization on the polynomials of the proposed demixing system. Finally, simulation results are presented to verify the efficacy of the proposed approach.**

***Index Terms*—Nonlinear independent component analysis, Nonlinear blind source separation, polynomial neural network, unsupervised learning**

I. INTRODUCTION

Generally, the problem of the blind separation of independent sources involves a set of observations $\mathbf{x} = \left[x_1(t) \quad x_2(t) \quad \cdots \quad x_p(t)\right]^T$ which are generated from a set of unknown independent components $\mathbf{s} = \left[s_1(t) \quad s_2(t) \quad \cdots \quad s_q(t)\right]^T$ according to

$$x_i = f_i\left(s_1, s_2, \ldots, s_q\right) \tag{1}$$

where f_i is an unknown differentiable bijective mapping, $i = 1, 2, \ldots, p$ and t is the time or sample index [1, 15, 18]. A technique known as Independent Component Analysis (ICA) is exploited to estimate both the mixing mappings f_i 's and the original sources $s_i(t),\ i = 1, 2, \ldots, q$. A common assumption by most ICA algorithm is that the mixing mapping takes the form of the linear combination, i.e. $f_i\left(s_1, s_2, \ldots, s_q\right) = m_{i1}s_1 + m_{i2}s_2 + \cdots + m_{iq}s_q$. However, this linear assumption is almost always violated due to the existence of the nonlinear distortion in practice, thus resulting in the failure in extracting the original source signals under nonlinear mixtures [2, 9, 11]. Hence, the search for a reliable nonlinear solution becomes paramount at both theoretical and practical levels.

The contribution of this paper is as follows: Firstly, a multi-nonlinearly constrained system is proposed as the mixing and demixing model. The model generalizes the original mono-nonlinearity model previously presented in [2]. The proposed model is a more general description than the post-nonlinear systems [5] and provides a better representation of a nonlinear mixture. Secondly, a new polynomial-based neural network demixer is proposed and developed as the separation system to estimate the unknown source signals. Finally, the theory of Series Reversion is incorporated into the derivation of the parameter learning algorithm to account for the special structure of the demixing network.

II. NONLINEAR ICA MODEL

A mono-nonlinearity mixing model derived from the theory of functional analysis was proposed in [2] to provide a general description of the mixing system in the following form:

$$\mathbf{x} = f\left(\mathbf{M} f^{-1}(\mathbf{s})\right) \tag{2}$$

where $\mathbf{M} = \left[\mathbf{m}_1 \quad \mathbf{m}_2 \quad \cdots \quad \mathbf{m}_p\right]^T$ with dimension $p \times q$ and $\mathbf{m}_i = \left[m_{i1} \quad m_{i2} \quad \cdots \quad m_{iq}\right]^T$. In this paper, we assume that the number of sources is equal to that of observations, i.e. $p = q = N$. This model is structured in the form of one linear mixing matrix sandwiched between two layers of nonlinearities, one of which is the inverse function of the other. The term 'mono-nonlinearity' represents the

Based on "Blind Source Separation of Nonlinearly Constrained Mixed Sources Using Polynomial Series Reversion", by P. Gao, L.C. Khor, W.L. Woo and S.S. Dlay which appeared in the Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006, Toulouse, France, May 2006. © 2006 IEEE.

condition where an identical nonlinear distortion is applied to each source signal. However, there is no guarantee that this condition is always fulfilled in practice. In fact, the channels between observations and sources are arbitrarily distorted due to the uncertainty of the environment. Hence, to preserve the special relationship between the two layers in (2), we represent the 'multi-nonlinearity' constrained mixing system by the following model:

$$\mathbf{x} = \mathbf{D_f} * (\mathbf{M}\mathbf{D}_{\mathbf{f}^{-1}} * \mathbf{s}) \tag{3}$$

where

$$\begin{aligned}
&\mathbf{D_f} = \mathrm{diag}\begin{bmatrix} f_1 & f_2 & \cdots & f_N \end{bmatrix} \\
&\mathbf{D}_{\mathbf{f}^{-1}} = \mathrm{diag}\begin{bmatrix} f_1^{-1} & f_1^{-1} & \cdots & f_N^{-1} \end{bmatrix} \\
&\mathbf{D_g} * \mathbf{u} = \mathrm{diag}\begin{bmatrix} g_1 * u_1 & g_2 * u_2 & \cdots & g_N * u_N \end{bmatrix} \\
&g_i * u_i = g_i(u_i)
\end{aligned}$$

This model will reduce to the mono-nonlinearity mixing model when $f_1 = f_2 = \cdots = f_N = f$ and can further represent a linear mixture as a special case if $\{f_i\}_{i=1}^{N}$ is linear.

A demixing system for (3) can be described by the inverse of the mixing system where the original sources are estimated as follows:

$$\hat{\mathbf{s}} = \mathbf{D}_{\mathbf{f}^{-1}}^{-1} * (\mathbf{M}^{-1}\mathbf{D}_{\mathbf{f}}^{-1} * \mathbf{x}) \tag{4}$$

Using the identity $\mathbf{D}_{\mathbf{g}}^{-1} = \mathbf{D}_{\mathbf{g}^{-1}}$, (4) can be rewritten as

$$\hat{\mathbf{s}} = \mathbf{D_f} * (\mathbf{W}\mathbf{D}_{\mathbf{f}}^{-1} * \mathbf{x}) \tag{5}$$

where $\mathbf{W}$ is the demixing matrix. Given the observed signals, the aim is to estimate $\{f_i\}_{i=1}^{N}$ and $\mathbf{W}$ such that the resulting transformed signals are mutually as independent as possible and statistically as close as possible to the source signals.

III. POLYNOMIAL-BASED NEURAL NETWORK FOR NONLINEAR ICA

In current literature, popular nonlinear network demixers such as SOM [16], GTM [17], RBF [4] and MLP with sigmoidal nonlinearity [7] are inherently nonlinear because of the fixed nonlinearities in the hidden neurons. However, the fixed rigidness of the nonlinearity will lead to the oversized and overfitted network and inevitably increase computational complexity [2]. Instead of using a fixed form of nonlinearity in the hidden neurons, we propose to design a demixer whereby its intrinsic nonlinearity can be flexibly controlled.

A. Polynomial-based Network as the Nonlinear ICA Demixing System

The Weierstrass Approximation Theorem states that for every continuous function $\phi : [\alpha, \beta] \rightarrow \mathbf{R}$, there always exists a polynomial series $p(u) = \sum_{m=0}^{M} \lambda_m u^m$, parameterized by $\boldsymbol{\theta} = \left\{ M,\ \{\lambda_m\}_{m=0}^{M} \right\}$, which can uniformly approximate ϕ with arbitrary accuracy. Therefore, a feedforward polynomial-based network shown in Figure 1 is proposed to reflect the model in (5). The hidden layer neurons in the proposed network perform the polynomial series to approximate the mixing mapping functions $\{f_i\}_{i=1}^{N}$ and $\{f_i^{-1}\}_{i=1}^{N}$. The outputs of the demixing system assume the following form

$$\begin{aligned}
&\mathbf{y}_{[3]} = \mathbf{D_f} * \mathbf{y}_{[2]} = \sum_{m=0}^{M_1} \left(\mathbf{a}_m \circ \mathbf{y}_{[2]}^{m} \right) \\
&\mathbf{y}_{[2]} = \mathbf{W}\mathbf{y}_{[1]} \\
&\mathbf{y}_{[1]} = \mathbf{D}_{\mathbf{f}^{-1}} * \mathbf{x} = \sum_{n=1}^{M_2} \left(\mathbf{b}_n \circ (\mathbf{x} - \mathbf{a}_0)^n \right)
\end{aligned} \tag{6}$$

where $\mathbf{y}_{[i]} = \begin{bmatrix} y_{[i,1]} & \cdots & y_{[i,N]} \end{bmatrix}^T$, $\mathbf{a}_m = \begin{bmatrix} a_{[m,1]} & \cdots & a_{[m,N]} \end{bmatrix}^T$, $\mathbf{b}_n = \begin{bmatrix} b_{[n,1]} & \cdots & b_{[n,N]} \end{bmatrix}^T$, $y_{[j,i]}$ denotes the i^{th} output of the j^{th} layer in the demixer, $\{a_{[m,i]}\}_{m=0;i=1}^{m=M_1;i=N}$ and $\{b_{[n,i]}\}_{n=1;i=1}^{n=M_2;i=N}$ are the coefficients, M_1 and M_2 represents the order of the series expansion and '$\circ$' denotes the Hadamard product.

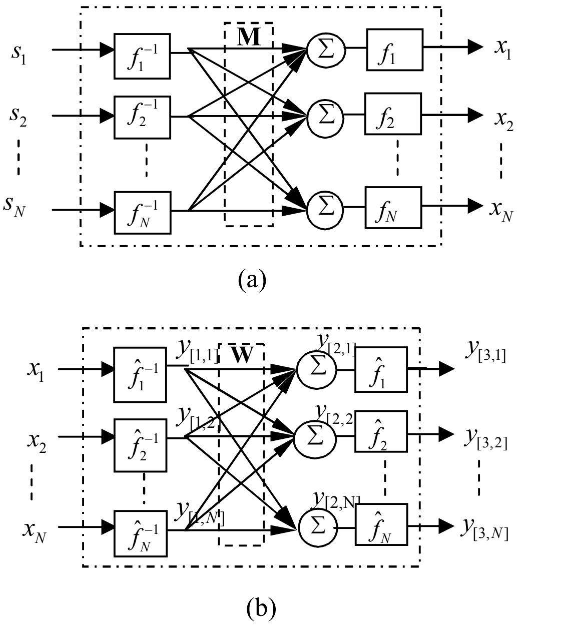


Figure 1: (a). Multi-nonlinearity Constrained Mixing Model. (b) Polynomial-based Nonlinear ICA Demixer.

B. Series Reversion

As shown in Figure 1, the implementation of the proposed demixer requires the inverse function of the polynomial series. It is possible to express the inverse function of a polynomial in a closed form when the order of the forward function is 4 or less. However, computing the inverse function becomes difficult and intractable as the order increases. The theory of the Series Reversion provides an alternative solution and further establishes the foundation for computing the inverse function of a general polynomial expansion. In this paper, instead of presenting the theorem formally, we provide a paraphrase of the main theorem in [6, 13] with further derivation to our proposed demixing system.

Theorem 1: If the function $g(.)$ has a polynomial expression as $g(u)=\sum_{m=0}^{M_1}\lambda_m u^m$, then its inverse function can be given by the similar form of $g^{-1}(u)=\sum_{n=1}^{+\infty}\gamma_n(u-\lambda_0)^n$ and the coefficients computed from

$$\gamma_n = \sum_{k_2,k_3,\ldots}\left[(-1)^{\sum_{i=2}^{M_1}k_i}\frac{(n-1+\sum_{i=2}^{M_1}k_i)!}{n!\prod_{i=2}^{M_1}(k_i!)}\left(\prod_{i=1}^{M_1}\lambda_i^{k_i}\right)\right] \quad (7)$$

where $k_2+2k_3+3k_4+\cdots=n-1,\ k_i\geq 0,\ i=2,3,4,\ldots$ and $k_1=-\left(n+\sum_{i=2}^{M_1}k_i\right)$. In addition, the differential of γ_n with respect to λ_m 's takes the form of

$$d\gamma_n=\sum_{m=1}^{M_1}\left(\sum_{k_2,k_3,\ldots}(-1)^{\sum_{i=2}^{M_1}k_i}\frac{(-k_1-1)!}{n!\prod_{i=2}^{M_1}(k_i!)}\left(\prod_{\substack{i=1\\ i\neq m}}^{M_1}\lambda_i^{k_i}\right)k_m\lambda_m^{k_m-1}\right)d\lambda_m \quad (8)$$

Hence, the derivative of the reverse series with respect to the coefficients in the forward polynomial $\frac{\partial g^{-1}}{\partial\lambda_m}$ can be easily derived from $\frac{\partial g^{-1}}{\partial\lambda_m}=\sum_{n=1}^{M_2}\left(\frac{\partial g^{-1}}{\partial\gamma_n}\frac{\partial\gamma_n}{\partial\lambda_m}\right)$.

C. Gradient Based Parameter Learning Algorithm

The primary goal of the demixing system is to obtain a set of signals as independent as possible. The cost function based on the Kullback-Leibler Divergence (KLD) in [1] is commonly used in blind signal separation problems. However, in nonlinear ICA, the preservation of independence is not strong enough to ensure signal separability and this inadvertently results in non-uniqueness of solutions. Therefore, to reduce the indeterminacy of non-unique solutions, the cost function is modified by incorporating a set of signal constraints into the original KLD cost function as follows:

$$J=-\log\left|\det\frac{d\mathbf{y}_{[3]}}{d\mathbf{x}^T}\right|-\sum_{i=1}^{N}\log\left(p_i(y_{[3,i]})\right)+\underbrace{\sum_{i=1}^{N}\beta_i g_i^{(c)}(y_{[3,i]},s_i)}_{\text{constraints}}$$
$$g_i^{(c)}(y_{[3,i]},s_i)=\sum_{j=1}^{D}\left[cum\left(y_{[3,i]},j\right)-cum\left(s_i,j\right)\right]^2 \quad (9)$$

where β_i's denotes a set of constants that control the weight of the additional constraints; $cum(u,j)$ represents the j^{th} order cumulant of u and D is the maximum order of the cumulant. In fact, these constraints imply the use of *a priori* information about the source distributions which is intended to match the outputs of the demixer to the original source signals in terms of cumulants.

The mutual information of the output will only be zero if the obtained outputs are independent with each other. The use of signal constraints primarily aims to force the demixer outputs to have identical statistics with the source signals up to the 2nd order statistics. It is noted that the use of signal constraints alone is not sufficient but needs to be used in conjunction with the structural constraint in order to effectively ameliorate the indeterminacy problem. When the mutual information is minimised and the signal constraints are satisfied, the outputs of the demixer will correspond to the original sources. According to the structure of the network, we have $\frac{d\mathbf{y}_{[3]}}{d\mathbf{x}^T}=\frac{d\mathbf{D_f}}{d\mathbf{y}_{[2]}}\cdot\mathbf{W}\cdot\frac{d\mathbf{D_{f^{-1}}}}{d\mathbf{x}}$. Since the forward and the inverse functions take the forms of the Weierstrass series:

$$\frac{df_i\left(\mathbf{y}_{[2]}\right)}{d\mathbf{y}_{[2]}}=\sum_{m=1}^{M_1}ma_{[m,i]}\mathbf{y}_{[2]}^{m-1}$$
$$\frac{df_i^{-1}(\mathbf{x})}{d\mathbf{x}}=\sum_{n=1}^{M_2}nb_{[n,i]}\left(\mathbf{x}-a_{[0,i]}\right)^{n-1} \quad (10)$$

where M_1 and M_2 denote the order of the forward and the reverse Weierstrass series respectively. Therefore,

$$\frac{d\mathbf{y}_{[3]}}{d\mathbf{x}^T}=\text{diag}\left[\sum_{m=1}^{M_1}m\left(\mathbf{a}_m\circ\mathbf{y}_{[2]}^{m-1}\right)\right]\mathbf{W}\text{diag}\left[\sum_{n=1}^{M_2}n\left(\mathbf{b}_n\circ(\mathbf{x}-\mathbf{a}_0)^{n-1}\right)\right] \quad (11)$$

Accordingly, the cost function for the demixer assumes the following form:

$$J=-\log|\det\mathbf{W}|-\log\left|\det\left\{\text{diag}\left(\sum_{m=1}^{M_1}m\left(\mathbf{a}_m\circ\mathbf{y}_{[2]}^{m-1}\right)\right)\right\}\right|$$
$$-\log\left|\det\left\{\text{diag}\left(\sum_{n=1}^{M_2}n\mathbf{b}_n(\mathbf{x}-\mathbf{a}_0)^{n-1}\right)\right\}\right| \quad (12)$$
$$-\sum_{i=1}^{N}\left[\log\left(p\left(y_{[3,i]}\right)\right)-\beta_i g_i^{(c)}\left(y_{[3,i]},s_i\right)\right]$$

Since the theory of series reversion is used, by using Theorem 1, for $i=1,2,\ldots,N$, we can obtain the total derivatives of $\left\{\gamma_{[n,i]}\right\}_{n=1}^{n=M_2}$ with respect to $\left\{\lambda_{[m,i]}\right\}_{m=1}^{m=M_1}$ as:

$$d\gamma_{[n,i]}=\sum_{m=1}^{M_1}\xi_{[n,m,i]}d\lambda_{[m,i]} \quad (13)$$

where

$$\xi_{[n,m,i]}=\sum_{k_2,k_3,\ldots}\left[(-1)^{\sum_{q=2}^{M_1}k_q}\frac{\left(n-1+\sum_{q=2}^{M_1}k_q\right)!}{n!\prod_{q=2}^{M_1}(k_q!)}\left(\prod_{\substack{j=1\\ j\neq m}}^{M_1}\lambda_{[j,i]}^{k_j}\right)k_m\lambda_{[m,i]}^{k_m-1}\right].$$ For simplicity, we define the following functions:

$$\boldsymbol{\eta}=\sum_{m=1}^{M_1}m(m-1)\text{diag}\left(\mathbf{a}_m\circ\mathbf{y}_{[2]}^{m-2}\right)$$

$$\boldsymbol{\kappa}_m=\text{diag}\left(\sum_{n=1}^{M_2}\boldsymbol{\xi}_{[n,m]}\circ(\mathbf{x}-\mathbf{a}_0)^n\right)$$

$$\boldsymbol{\varphi}_1=\left[-1\Big/\left(\sum_{m=1}^{M_1}ma_{[m,1]}y_{[2,1]}^{m-1}\right)\ \cdots\ -1\Big/\left(\sum_{m=1}^{M_1}ma_{[m,N]}y_{[2,N]}^{m-1}\right)\right]^T$$

$$\boldsymbol{\varphi}_2=\left[-1\Big/\left(\sum_{n=1}^{M_2}nb_{[n,1]}\left(x_1-a_{[0,1]}\right)^{n-1}\right)\ \cdots\ -1\Big/\left(\sum_{n=1}^{M_2}nb_{[n,N]}\left(x_N-a_{[0,N]}\right)^{n-1}\right)\right]^T$$

$$\boldsymbol{\xi}_{[n,m]}=\left[\xi_{[n,m,1]}\ \cdots\ \xi_{[n,m,N]}\right]^T$$

$$\boldsymbol{\psi}=\tilde{\boldsymbol{\psi}}+\boldsymbol{\beta}\circ\mathbf{g}^{(\mathbf{c})}$$

$$\boldsymbol{\beta}=\left[\beta_1\ \cdots\ \beta_N\right]^T$$

$$\tilde{\boldsymbol{\psi}} = -\left[\frac{d\left[\log\left(p_1(y_{[3,1]})\right)\right]}{dy_{[3,1]}} \cdots \frac{d\left[\log\left(p_N(y_{[3,N]})\right)\right]}{dy_{[3,N]}}\right]^T$$

$$\mathbf{g}^{(c)} = \left[\frac{d\left[f_1^{(c)}\left(y_{[3,1]}, s_1\right)\right]}{dy_{[3,1]}} \cdots \frac{d\left[f_N^{(c)}\left(y_{[3,N]}, s_N\right)\right]}{dy_{[3,N]}}\right]^T$$

The total differential of the cost function is then derived as

$$\begin{aligned} dJ &= -\mathrm{tr}\left[d\mathbf{W}\mathbf{W}^{-1}\right] + \boldsymbol{\psi}^T d\mathbf{y}_{[3]} \\ &+ \boldsymbol{\varphi}_1^T\left[\sum_{m=1}^{M_1} m\left(\mathbf{y}_{[2]}^{m-1} \circ d\mathbf{a}_m\right) + \sum_{m=1}^{M_1} m(m-1)\mathrm{diag}\left(\mathbf{y}_{[2]}^{m-2}\right)\left(\mathbf{a}_m \circ d\mathbf{y}_{[2]}\right)\right] \\ &+ \boldsymbol{\varphi}_2^T \sum_{m=1}^{M_1}\left(\left(\sum_{n=1}^{M_2} n\boldsymbol{\xi}_{[n,m]} \circ \left(\mathbf{x} - \boldsymbol{\lambda}_0\right)^{n-1}\right) \circ d\mathbf{a}_m\right) \\ &+ \boldsymbol{\varphi}_2^T \sum_{n=1}^{M_2} n(n-1)\mathrm{diag}\left(\left(\mathbf{x} - \mathbf{a}_0\right)^{n-2}\right)\left(\mathbf{b}_n \circ d\left(\mathbf{x} - \mathbf{a}_0\right)\right) \end{aligned} \quad (14)$$

From the structure of the network, the following equations can be found:

$$d\mathbf{y}_{[1]} = \sum_{m=1}^{M_1} \boldsymbol{\kappa}_m \circ d\mathbf{a}_m + \mathrm{diag}^{-1}\left(\boldsymbol{\varphi}_2\right) d\mathbf{x} - \mathrm{diag}^{-1}\left(\boldsymbol{\varphi}_2\right) \circ d\mathbf{a}_0 \quad (15)$$

$$\begin{aligned} d\mathbf{y}_{[2]} &= d\mathbf{W}\mathbf{y}_{[1]} + \mathbf{W}\sum_{m=1}^{M_1}\left(\boldsymbol{\kappa}_m \circ d\mathbf{a}_m\right) \\ &+ \mathbf{W}\left[\mathrm{diag}^{-1}\left(\boldsymbol{\varphi}_2\right) d\mathbf{x} - \mathrm{diag}^{-1}\left(\boldsymbol{\varphi}_2\right) d\mathbf{a}_0\right] \end{aligned} \quad (16)$$

$$\begin{aligned} d\mathbf{y}_{[3]} &= \left[\mathrm{I} - \mathrm{diag}^{-1}\left(\boldsymbol{\varphi}_1\right)\mathbf{W}\mathrm{diag}^{-1}\left(\boldsymbol{\varphi}_2\right)\right] d\mathbf{a}_0 \\ &+ \sum_{m=1}^{M_1}\left[\left(\mathrm{diag}^{-1}\left(\boldsymbol{\varphi}_1\right)\mathbf{W}\boldsymbol{\kappa}_m + \mathbf{y}_{[2]}^m\right) \circ d\mathbf{a}_m\right] \\ &+ \mathrm{diag}^{-1}\left(\boldsymbol{\varphi}_1\right)\mathbf{W}\mathrm{diag}^{-1}\left(\boldsymbol{\varphi}_2\right) d\mathbf{x} + \mathrm{diag}^{-1}\left(\boldsymbol{\varphi}_1\right) d\mathbf{y}_{[1]} \end{aligned} \quad (17)$$

The derivative of the cost function with respect to the parameters can therefore be derived as (18)-(20).

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{W}}\mathbf{W}^T &= \mathrm{I} - \left(\sum_{m=1}^{M_1} m(m-1)\mathrm{diag}\left(\mathbf{a}_m \circ \mathbf{y}_{[2]}^{m-2}\right)\right)\boldsymbol{\varphi}_1\mathbf{y}_{[2]}^T \\ &- \left(\sum_{m=1}^{M_1} m\,\mathrm{diag}\left(\mathbf{a}_m \circ \mathbf{y}_{[2]}^{m-1}\right)\right)\boldsymbol{\psi}\mathbf{y}_{[2]}^T \end{aligned} \quad (18)$$

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{a}_0} &= \left[\mathrm{I} - \mathrm{diag}^{-1}\left(\boldsymbol{\varphi}_1\right)\mathbf{W}^T\mathrm{diag}^{-1}\left(\boldsymbol{\varphi}_2\right)\right]\boldsymbol{\psi} + \mathrm{diag}^{-1}\left(\boldsymbol{\varphi}_2\right)\mathbf{W}^T\boldsymbol{\eta}\boldsymbol{\varphi}_1 \\ &- \left(\sum_{n=1}^{M_2} n(n-1)\mathrm{diag}\left(\mathbf{b}_n \circ \left(\mathbf{x} - \mathbf{a}_0\right)^{n-2}\right)\right)\boldsymbol{\varphi}_2 \end{aligned} \quad (19)$$

$$\begin{aligned} \frac{\partial J}{\partial \mathrm{diag}\left(\mathbf{a}_m\right)} &= \boldsymbol{\varphi}_1\left(m\mathbf{y}_{[2]}^{m-1}\right)^T + \mathrm{diag}\left[\boldsymbol{\psi}\left(\mathbf{y}_{[2]}^m\right)^T\right] \\ &+ \boldsymbol{\varphi}_2\left(\sum_{n=1}^{M_2} n\boldsymbol{\xi}_{[n,m]} \circ \left(\mathbf{x} - \mathbf{a}_0\right)^{n-1}\right)^T \\ &+ \boldsymbol{\kappa}_m\mathbf{W}^T\boldsymbol{\eta}\mathrm{diag}\left(\boldsymbol{\varphi}_1\right) - \mathrm{diag}^{-1}\left(\boldsymbol{\varphi}_1\right)\mathbf{W}^T\boldsymbol{\kappa}_m\mathrm{diag}\left(\boldsymbol{\psi}\right) \end{aligned} \quad (20)$$

By inserting (18)-(20) into (21)-(22), the gradient descent based learning algorithm can be obtained.

$$\mathbf{W}(t+1) = \mathbf{W}(t) - \mu_{\mathbf{W}}\frac{\partial J}{\partial \mathbf{W}}\mathbf{W}^T\mathbf{W}(t) \quad (21)$$

$$\mathbf{a}_m(t+1) = \mathbf{a}_m(t) - \mu_{\mathbf{a}_m}\frac{dJ}{d\mathbf{a}_m} \quad ; \quad m = 0, 1, \ldots, M_1 \quad (22)$$

where $\mu_{\mathbf{W}}$ and $\mu_{\mathbf{a}_m}$ represent the step size of the demixing matrix $\mathbf{W}$ and $\mathbf{a}_m$'s respectively.

IV. RESULTS

In this section, three experiments will be conducted to verify the efficacy of the proposed approach. In the first experiment, the robustness of the linear algorithm in nonlinear mixture is investigated to show the importance of the solution specifically tailored for the blind signal separation under nonlinear environment. The second experiment is carried out for the blind source separation based on synthetic data while the third experiment on real life speech signals. Analysis based on the order of the Weierstrass series is presented. Before showing the results, a performance index which is able to evaluate the performance is introduced. It is well known that Mean Square Error (MSE) is a great performance index to measure the similarity between two signals. However, it is sensitive to the variability of both scale and phase of the signals. In the context of blind source separation, the estimated signals can be subject to scale and phase reversal ambiguities and therefore, the MSE criterion is not suitable for direct performance comparison between original sources and estimated sources. As an alternative, the following performance index is proposed in [2] as

$$\mathsf{P} = 2\left(1 - \frac{1}{N}\sum_{i=1}^{N} |\,\theta_i\,|\right) \quad (23)$$

where

$$\theta_i = \frac{E\left[\overline{(s_i - E[s_i])}(y_{[3,i]} - E[y_{[3,i]}])\right]}{\sqrt{E\left[|s_i - E[s_i]|^2\right]E\left[|y_{[3,i]} - E[y_{[3,i]}]|^2\right]}} \quad (24)$$

θ_i is denoted as the normalised cross-correlation between the original and the estimates source signals. In above, ‘ $\overline{u}$ ’ and ‘ $|u|$ ’ denote the complex conjugate and absolute operation of u, respectively. In short, this performance index P is essentially a variant of the MSE criterion that implicitly takes into account the scale and phase reversal ambiguities.

A. Experiment 1

Two sinusoid signals are generated synthetically as the original source signals. The source signals are passed through the nonlinear mixing system defined as $\mathbf{x} = f(\mathbf{Ms})$ where $\mathbf{M}$ is selected to be an orthogonal matrix parameterised by the angle of α with the form of

$$\mathbf{M} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \quad (25)$$

and $f(\cdot)$ is the memoryless nonlinear function which is given by $f(\mathbf{Ms}) = \mathbf{Ms} + \beta(\mathbf{Ms})^3$ where β represents the degree of the nonlinearity. It can be implied that the higher value β has, the more severe the intrinsic nonlinearity of $f(\cdot)$ is. Along with the MSE, the following performance index is evaluated to formulate explicitly the normalised cross correlation between original and the estimates source signals.

$$\delta=\frac{1}{2N}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\left(\frac{E\left[\left(y_{[3,i]}-E\left(y_{[3,i]}\right)\right)\left(y_{[3,j]}-E\left(y_{[3,j]}\right)\right)\right]}{\sqrt{E\left[\left|y_{[3,i]}-E\left(y_{[3,i]}\right)\right|^{2}\right]E\left[\left|y_{[3,j]}-E\left(y_{[3,j]}\right)\right|^{2}\right]}}\right) \quad (26)$$

The MSE performance index shows how accurate the estimate signals resemble the original sources while the normalised cross correlation implies the statistical independence among the estimated signals. In this experiment, the parameter α from the equation (26) which characterise the degree of mixing is within [0°,45°] and β∈[0, 2]. The achieved MSE and δ with respect to α and β using JADE algorithm [12] is shown in Figure 2.

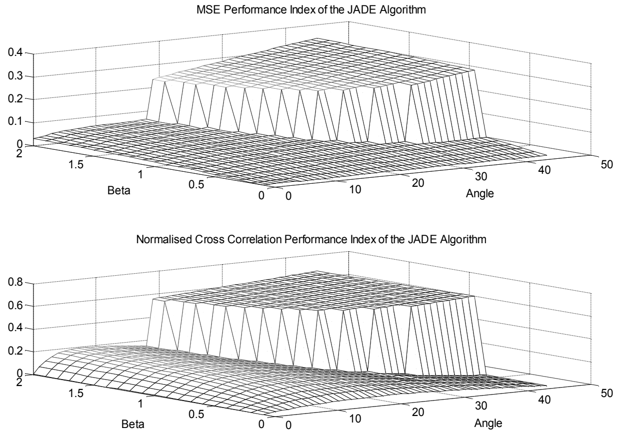
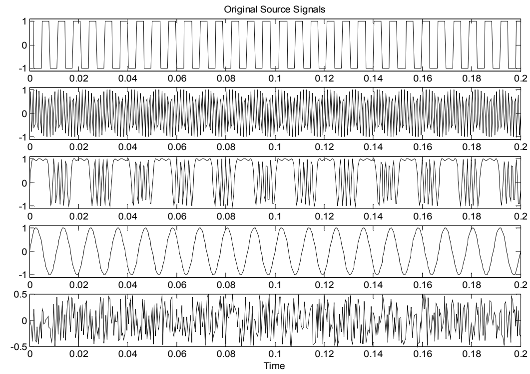


Figure 2: The performance index with respect to α and β of the JADE algorithm.

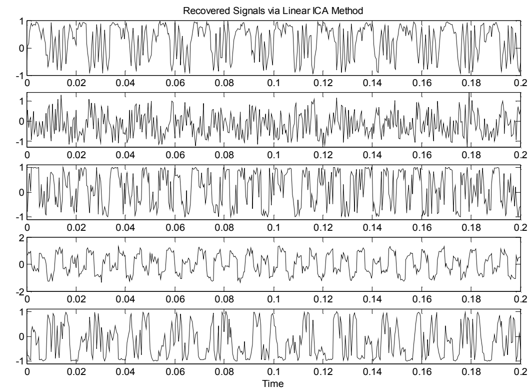
Within the region of $\left\{\alpha\in\left[15^{\circ},45^{\circ}\right],\beta\in\left[0.5,2\right]\right\}$, it can be seen that both the MSE and cross correlation performance of the linear algorithm under normal nonlinear mixtures are high and we conclude that the linear algorithm fails to extract the original source signals within this region. Nonlinear BSS solution is required under typical nonlinear environment. It can be concluded that in linear mixtures performances of linear algorithms are unaffected by the degree of mixing whereas in nonlinear mixtures performance of linear algorithms rely on both the degree of nonlinearity and the degree of mixing.

B. Experiment 2

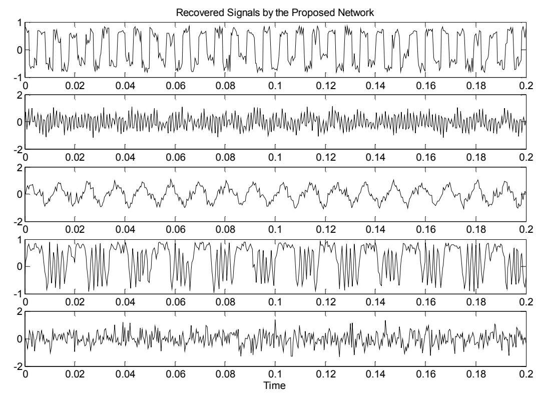
Five subgaussian signals are generated synthetically as the original sources and expressed as $\mathbf{s}(t)=$ [Binary signal; $\sin(1600\pi t)$; $\sin\left(600\pi t+6\cos(120\pi t)\right)$; $\sin(180\pi t)$; Uniform-distributed signal$]^{T}$. The source signals are then mixed according to (3) where $\mathbf{M}$ is a 5×5 random mixing matrix and $\mathbf{D_f}=\mathrm{diag}\left[\tanh\quad \sinh^{-1}\quad \tanh\quad \sinh^{-1}\quad \tanh\right]$. The learning rates for the weights and the coefficients $\mathbf{a}_m$ are set to $\mu_{\mathbf{W}}=0.001$ and $\mu_{\mathbf{a}_m}=0.00003$, respectively. In order to assess the performance of the proposed algorithms, we compare the proposed method with existing algorithms (Linear ICA [1], RBF [4] and FMLP Network [7]) based on the MSE performance index.



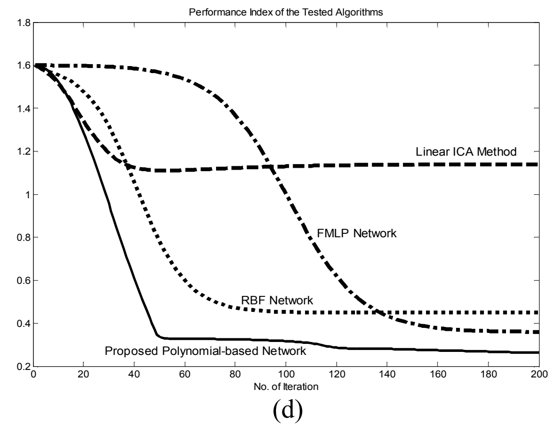
(a)



(b)



(c)



(d)

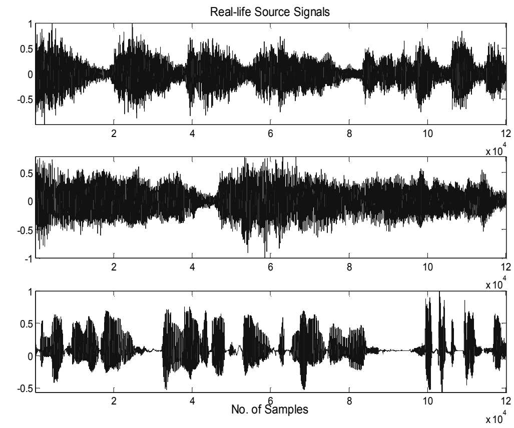
Figure 3: (a) Original sources.
(b) Recovered signals via Linear ICA method.
(c) Recovered signals via the proposed network.
(d) MSE Performance index of the tested algorithms.

The source signals, signals recovered by Linear ICA method and the proposed network, the performance index of the tested algorithms are shown in Figure 3. We have also simulated the RBF and FMLP demixers with different number of hidden neurons respectively but no substantial improvement of results has been obtained. A Monte-Carlo experiment of 100 trials has been conducted for the RBF and FMLP demixer and in each simulation, the convergence of the RBF and the MLP demixers have been monitored to ensure that both demixers do not converge to local minima. In Figure 2(d), the proposed approach has demonstrated its efficacy in separating signals under the nonlinear mixture. The success is consecutively followed by the MLP and RBF but the separation results achieved by the linear method falls far from optimal and this further indicates the crucial need for nonlinear separation techniques.

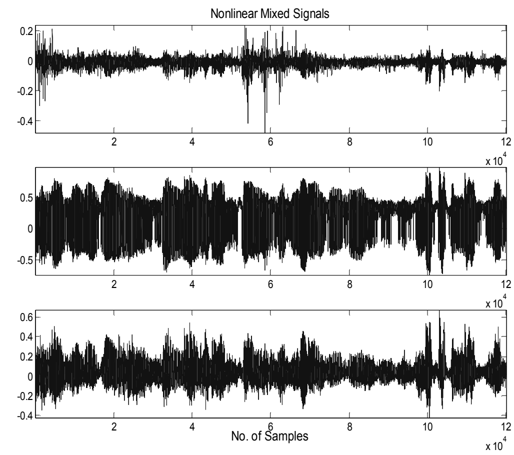
C. Experiment 3

To further investigate the efficacy of the proposed scheme in practical scenario, the third experiment is conducted for separating real data which are obtained from the ICA'99 datasets [8]. Three speech signals are used as the original source signals. Similar mixing model is applied to the sources where $\mathbf{M}$ is changed to an unknown 3×3 random mixing matrix and $\mathbf{D}_{\mathbf{f}^{-1}} * \mathbf{s} = \mathrm{diag}\left\{\tanh^{-1}(s_1) \quad s_2^3 \quad s_3 + 0.8 s_3^3\right\}$. The mixing system is expected to represent the combined recording amplifier [10] and the nonlinearity due to the carbon-button microphones [14] working in the saturation region whose characteristics can be approximated by the hyperbolic function. The parameter settings used here remains identical to the first experiment except the learning rates are now changed to $\mu_{\mathbf{W}} = 0.0005$ and $\mu_{\lambda_m} = 1\times 10^{-5}$. Furthermore, different orders of the Weierstrass series are applied to this experiment to investigate the influence induced by the truncation. Both the forward and the reverse series will be truncated to the 5th, 7th and 9th orders, respectively. The original sources, the nonlinearly mixed signals and the restored signals via the proposed 7th order Weierstrass Network are displayed in Figure 4(a)-(c). Figure 4(d) shows the performance evaluated in term of convergence and accuracy of the proposed algorithm with different orders of polynomials.

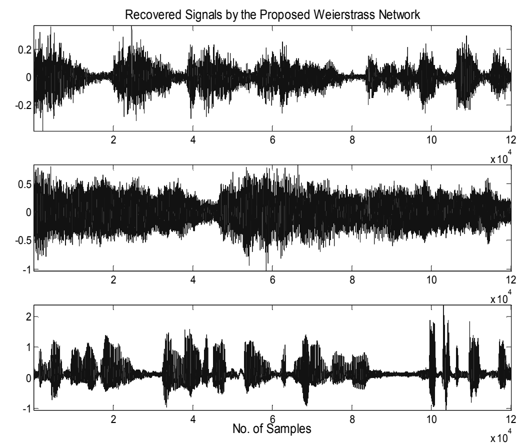
Similar to the previous experiment, the analysis shows that the proposed method has successfully recovered the real-life recorded signals. Since the nonlinear layers are updated from linear functions and $\mu_{\mathbf{W}} > \mu_{\lambda_m}$, the initial convergence (0-25 iteration) is mainly dominated by the linear demixing matrix, which leads to the similarity of all three curves. The adaptation of the nonlinear functions contributes primarily after 30 iterations. It is also observed that the low-order coefficients in the Weierstrass series are firstly updated since they determine the shape and tendency of the nonlinear function. The effect of the high-order items in the Weierstrass series is shown clearly when the system tends to be stable. It is seen that the 9th-order Weierstrass Network outperforms others in term of accuracy. Concurrently, the convergences of the performance index for the 5th, 7th and 9th order Weierstrass Network are almost identical but only differ in terms of steady-state values. The improvement achieved from the 5th order to the 7th is significant whereas from the 7th to the 9th order, the decrement is only 0.012. This result shows that the low-order items in the Weierstrass series dominate the performance of the approximation.



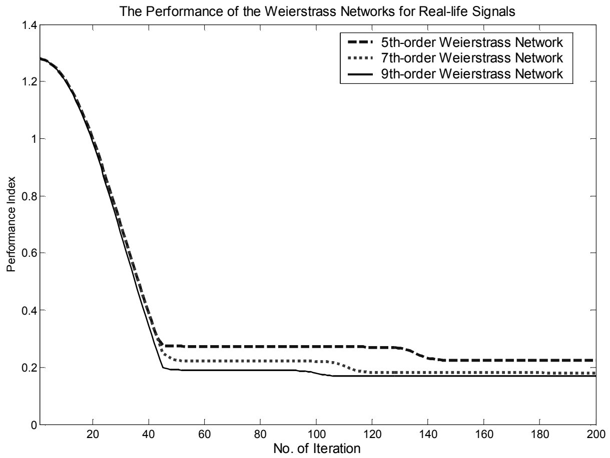
(a)



(b)



(c)



(d)

Figure 4: (a) Real-life speech source signals.
(b) Nonlinear mixed signals.
(c) Recovered signals by the 7th-order Weierstrass Network.
(d) MSE performance index of the Weierstrass Network with different orders.

IV. CONCLUSION

In summary, we proposed a new algorithm for the polynomial neural network nonlinear ICA demixer based on the extension of the mono-nonlinearity mixing model. Unlike the MLP-based network, a set of finite order polynomials are used as the activation function of the hidden neurons to approximate the true nonlinear distortions in the channel. This is primarily to reduce the size of the network by providing flexible and adjustable degree of nonlinearity. According to the structural requirement of the demixing system, the inverse function is calculated as a truncated polynomial according to the theory of reversion which further provides the relationship between the coefficients of the forward and the reverse series. The parameter learning algorithm is therefore derived and takes this relationship into consideration to provide optimal performance. Finally, based on the simulation results carried out, we show that the proposed method outperformed other linear and nonlinear algorithms in terms of accuracy and dynamic convergence speed.

REFERENCES

[1] S.I. Amari and A. Cichocki, "Adaptive blind signal processing - Neural network approaches", *Proceedings of the IEEE*, **86**(10), pp. 2026-2048, Oct. 1998.
[2] W.L. Woo and S.S. Dlay, "Neural Network Approach to Blind Separation of Mono-nonlinearly Mixed Sources", *IEEE Trans. on Circuits and System-1*, **52**(6), pp. 1236-1247, 2005.
[3] A. Hyvarinen and P. Pajunen, "Nonlinear Independent Component Analysis: Existence and Uniqueness Results", *Neural Network*, 12(3), pp. 429-439, 1999.
[4] Y. Tan, J. Wang, and J. Zurada, "Nonlinear Blind Source Separation using a Radial Basis Function Network", *IEEE Trans. on Neural Network*, **12** (1), pp. 124-134, 2001.
[5] A. Taleb and C. Jutten, "Source Separation in Post-Nonlinear Mixtures", *IEEE Transactions on Signal Processing*, 47 (10), pp. 2807-2820, 1999.
[6] P.Gao, W.L.Woo and S.S. Dlay, "Weierstrass approach to nonlinear blind signal separation" in *Proc. of 4th Int. Symp. on Communication Systems, Networks and Digital Signal Processing*, Newcastle, UK, July 2004, pp. 588-591.
[7] W.L. Woo And S. Sali, "General Multilayer Perceptron Demixer Scheme For Nonlinear Blind Signal Separation", *IEE Proc. On Vision, Image And Signal Processing*, 149(5), Pp. 253-262, 2002.
[8] http://sound.media.mit.edu/ica-bench/
[9] 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, pp. 173-183, April 2005.
[10] W. Frank, R. Reger, and U. Appel, "Loudspeaker nonlinearities - Analysis and compensation", *in Proc. Int. Asilomar on Signals, Systems and Computers*, Pacific Groove, California, USA, 1992, pp. 756-760.
[11] 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, February 2004.
[12] J. Cardoso, "High-Order Contrasts for Independent Component Analysis", *Neural Computation* 11, pp. 157–192, 1999.
[13] P.M. Morse, and H. Feshbach, *Methods of Theoretical Physics*, McGraw-Hill Book Company, Inc., New York, 1953.
[14] 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, 2000.
[15] W.L. Woo and S.S. Dlay, "Regularised Nonlinear Blind Signal Separation using Sparsely Connected Network", *IEE Proc. on Vision, Image and Signal Processing*, vol. 152, no. 1, pp. 61-73, February 2005.
[16] P. Pajunen, A. Hyvarinen, and J. Karhunen, "Nonlinear Blind Source Separation by Self-Organising Maps". *Progress in Neural Information Processing 9*, 1996, pp. 1207-1210.
[17] C. Bishop, M. Svensen, and C. William, "GTM: The Generative Topographic Mapping", *Neural Computation*, 1998, **10**, pp. 215-234.
[18] L.C. Khor, W.L. Woo and S.S. Dlay, "Nonlinear Blind Signal Separation with Intelligent Controlled Learning", *IEE Proc. on Vision, Image and Signal Processing*, vol. 152, no. 3, pp. 297-306, June 2005.

P. Gao was born in China. She received the BEng degree (First Class Hons) in Electronic Information and Control Engineering from Beijing University of Technology and the MSc degree (Pass with Distinction) in Communications and Signal Processing from University of Newcastle upon Tyne, UK. She is currently a PhD student with the School of Electrical, Electronics and Computer Engineering at Newcastle University. She was awarded Overseas Research Students Awards (ORS) by Universities UK and International Research Scholarship by Newcastle University for her research work. Her major research interests are in the algorithms for nonlinear signal and image processing. In particular, these are in the areas of blind source separation, independent component analysis, signal/image restoration.

L.C. Khor received her BEng (First Class Hons) and PhD degrees from University of Newcastle upon Tyne, U.K. She has completed her PhD in 2005 and later joined the university as a Researcher in the School of Electrical, Electronic and Computer Engineering. During that time, she has helped to secure and worked on two research projects funded by QinetiQ and Foster Findlay and Associates (ffA). Since joining the university, she has continued to engage in the research of nonlinear signal processing techniques for blind source separation, seismic signal processing and digital watermarking. She has published over 30 papers on these topics in various journals and international conference proceedings. To date, some of her work has been short-listed for the "Set for Britain" Award for Britain's Younger Engineers presented in the House of Commons in 2004. Dr Khor is a member of the IET and IEEE.

W.L. Woo was born in Malaysia. He received the BEng degree (First Class Hons) in Electrical and Electronics Engineering and the PhD degree from University of Newcastle upon Tyne, UK. He was awarded the IEE Prize for academic performance and subsequently, awarded the British Scholarship in 1998 to continue his research work. During this time, he was appointed as a Research Associate to develop robust algorithms based on Higher Order Statistics for adaptive signal estimation, a project funded by the Engineering and Physical Science Research Council (EPSRC). He is currently a Lecturer with the School of Electrical, Electronics and Computer Engineering at same university. His major research interests are in the mathematical theory and algorithms for nonlinear signal and image processing. In particular, these are in the areas of blind source separation, independent component analysis, signal/image deconvolution and restoration. He has published over 150 papers on these topics on various journals and international conference proceedings. Currently, he serves on the Editorial Board of the IET Signal Processing journal (formerly known as the IEE Proceedings in Vision, Image and Signal Processing), Journal of Computers and the World Scientific and Engineering Academy and Society for Transactions on Signal Processing. He is also a Member of the Technical Committee for Signal Processing of the International Association of Science and Technology for Development. He co-chaired and organized the IEEE/IEE 4th International Symposium on Communication Systems, Network and Digital Signal Processing in 2004. He is the co-guest editor of the special issue of the Mediterranean Journal of Computers and Networks. He also actively participate in international conferences and workshops, and serves on their organizing and technical committees. In addition, he acts as a consultant to a number of industrial companies that involve the use of statistical signal and image processing techniques. Dr Woo is a member of the IET and IEEE.

S.S Dlay: received his BSc (Hons.) degree in Electrical and Electronic Engineering and his PhD in VLSI Design from the University of Newcastle upon Tyne in 1979 and 1983 respectively. During this time he held a Scholarship from the Engineering and Physical Science Research Council (EPSRC) and the Charles Hertzmann Award. In 1984 he was appointed as a Post Doctoral Research Associate at Newcastle University and helped to establish an Integrated Circuit Design Centre, funded by the EPSRC. In November 1984 he was appointed as a Lecturer in the Department of Electronic Systems Engineering at the University of Essex. In 1986 he re-joined the University of Newcastle upon Tyne as a Lecturer in the School of Electrical, Electronic and Computer Engineering, then in 2001 he was promoted to Senior Lecturer. In recognition of his major achievements in the area of Signal Processing he has recently been appointed to a Personal Chair in Signal Processing Analysis. He is currently Head of the Signal Processing group. He has published over 200 research papers and his research interests lie in the mathematical advancement and application of modern signal processing theory. More particularly this is in the areas of Blind Source Separation; Biometrics and Security, Biomedical Signal Processing and implementation of Signal Processing Architectures. He is the co-guest editor of the special issue of the IEE Proceeding Circuit, Devices and Systems. He organised and chaired the IEEE/IEE 4th International Symposium on Communication Systems, Network and Digital Signal Processing in 2004, and was also the Editor of the resulting proceedings. He serves on the Editorial Committee of The Molecular and Cellular Pathology of Cancer Progression and Prognosis and is a Member of the Editorial Board of The World Scientific and Engineering Academy and Society for Transactions on Signal Processing. He has played an active role in numerous international conferences in terms of serving on technical and advisory committees as well as organising special sessions on Applications of Signal Processing. Professor Dlay is a College Member of the EPSRC for the term 2006-2009.