Fetal electrocardiogram extraction by sequential source separation in the wavelet domain

Fetal Electrocardiogram Extraction by Sequential Source Separation in the Wavelet Domain

Maria G. Jafari*, *Member, IEEE,* and Jonathon A. Chambers, *Senior Member, IEEE*

***Abstract*—This paper addresses the problem of fetal electrocardiogram extraction using blind source separation (BSS) in the wavelet domain. A new approach is proposed, which is particularly advantageous when the mixing environment is noisy and time-varying, and that is shown, analytically and in simulation, to improve the convergence rate of the natural gradient algorithm. The distribution of the wavelet coefficients of the source signals is then modeled by a generalized Gaussian probability density, whereby in the time-scale domain the problem of selecting appropriate nonlinearities when separating mixtures of both sub- and super-Gaussian signals is mitigated, as shown by experimental results.**

***Index Terms*—Blind source separation, fetal electrocardiogram extraction, independent component analysis, wavelet transform.**

I. Introduction

THE fetal electrocardiogram (FECG) provides information about fetal maturity, position of the fetus, multiple pregnancies, as well as being a diagnostic tool which can monitor conditions such as arrhythmia, and assess the fetal acidosis, and may be of vital importance to both mother and child when risk factors are present during pregnancy [1], [2]. FECGs can be between 5 and 1000 times smaller in intensity than in the adult [3], because of the layers of tissue between the electrodes and the fetal heart and, therefore, it is generally obscured by noise [4], [5]. In addition, the amplitude of FECG signals changes during pregnancy: it increases during the first 25 weeks, experiences a marked minimum toward the 32nd week, and increases again afterwards [6]. The best and most accurate FECG is obtained when the electrodes are attached directly to the fetal scalp, but this is only achievable during delivery, and clearly cannot be used to monitor the state of the fetus throughout pregnancy, or for an early diagnosis [7]. Thus, the attractiveness of noninvasive techniques for the extraction of FECG signals is obvious. The fetal electrocardiogram extraction problem involves the elimination of maternal ECG (MECG) components, and other interfering signals, from ECG measurements obtained during pregnancy. The application of blind source separation (BSS) methods to FECG extraction is justified in [7]–[10], where the validity of the BSS model for biomedical applications in general is also discussed. The block-based JADE algorithm, which performs independent component analysis (ICA), is applied to the FECG extraction problem in [3], and its performance is compared with that of the least mean square (LMS) and recursive least squares (RLS) algorithms, which are used to implement an adaptive MECG canceller. The authors show that the JADE algorithm performs better than LMS and RLS when the mixing channel satisfies the linear instantaneous mixture assumption, while its performance breaks down when the characteristics of the mixing channel deviate from the ideal. The time-frequency information present in ECG signals is exploited in [11], where the batch algorithm is based on the joint diagonalization of whitened spatial time-frequency distribution matrices, while the performance of higher-order ICA methods, and principal component analysis is compared in [7], [8], [10], and [12]. Results indicate that sources recovered using ICA approaches are clearer and less noisy than those obtained with PCA. Also, the performance of a block based, two-stage, BSS algorithm, consisting of a prewhitening step followed by higher-order processing, has been shown to be superior to Widrow's multireference adaptive noise cancelling method, although the computational complexity increases [9], [13].

Nonetheless, the noisy and nonstationary nature of this type of signal presents a challenge to conventional instantaneous BSS algorithms, which are often derived on the assumption that statistically time-invariant sources are mixed in the absence of noise [14]. In practice, source nonstationarity is a characteristic of speech and audio signals, as well as biomedical measurements where, for instance, the heart rate varies over time, depending on factors such as the comfort of the patient, and additive noise, Gaussian or otherwise due, for example, to the recording equipment. The separation of nonstationary sources is often addressed by methods based only on second-order statistics, which make the crucial assumption of nonstationarity of the source signals, i.e., the variances of the sources are assumed to change with time [15]–[18]. The presence of additive noise within the mixing model complicates significantly the estimation of the separating matrix. When additive noise is not assumed to be zero, it is often regarded as an additional set of sources, leading to an even more difficult problem, which is under-determined even when the number of sensors equals the number of sources. In some applications, it may be desirable to reduce the noise level by applying some form of preprocessing, while in others it may be possible to measure the environmental noise, for instance by appropriately placing additional electrodes during the recording of ECG signals [19]. An adaptive approach to noise cancellation and

Manuscript received April 22, 2003; revised June 20, 2004. This work was supported by the EPSRC of the U.K. *Asterisk indicates corresponding author.*

*M. G. Jafari is with the Centre for Digital Music, Department of Electronic Engineering, Queen Mary University of London, Mile End Road, London, E1 4NS, U.K. (e-mail: maria.jafari@elec.qmul.ac.uk).

J. A. Chambers is with the Centre of Digital Signal Processing, Cardiff School of Engineering, Cardiff University, Queen's Buildings, Cardiff, CF24 OYF, U.K.

Digital Object Identifier 10.1109/TBME.2004.842958

source separation is developed in [20], where it is assumed that the environmental noise is known, and that noise is additive and convolutional. Then, in order to recover the sources, the mixtures are separated and the additive noise is estimated and subtracted. When the noise is Gaussian, it is possible to exploit the fact that Gaussian random variables have zero higher-order cumulants. This is the approach taken in [21], where the explicit use of fourth-order cross cumulants ensures that theoretically the algorithm is robust with respect to Gaussian noise. Alternatively, the noise level can be reduced by means of wavelet de-noising. An experimental study of its use as a preprocessing stage for BSS is presented in [22], where it is concluded that wavelet de-noising is a very efficient preprocessing technique, which improves the performance of certain BSS methods. In [23], the wavelet transform is used to decompose an observed signal into subband signals, and the natural gradient algorithm (NGA) is only applied to those bands having the strongest power. In this paper, we propose a wavelet domain approach, which addresses the BSS problem in general, and the FECG extraction problem in particular, when the sources are nonstationary and contaminated by additive noise. The use of frequency and time-frequency methods for BSS has also been motivated by the observation that when mapping certain signals from the time domain to the frequency domain, the statistics of the sources become less Gaussian [24]. To show that this property is also true in the wavelet domain, we make use of an image processing result[1] to obtain a model for the sample distribution of the wavelet coefficients of the sources, thus extending it to one-dimensional (1-D) signals. The technique improves the speed of convergence of the NGA, and can overcome the problem of having to select the nonlinearities required to separate mixed sub- and super-Gaussian signals. Addressing the fetal ECG extraction problem within a sequential learning framework is particularly advantageous when motion occurs during the data acquisition process. This may be due, for instance, to changes in position by the expectant mother or the fetus. In such cases, it is useful for the separating algorithm to have the property of quickly reconverging.

The BSS problem is described in Section II, while it is shown in Section III that FECG extraction can indeed be formulated as a BSS problem, and the wavelet transform is introduced in Section IV. The time-scale domain approach is explained in Section V, and the performance of the proposed method is shown by simulation in Section VI, while conclusions are drawn in Section VII.

II. Problem Statement

When n source signals are instantaneously mixed by a stationary channel, and additive noise is present, the m observed signals are given by [25], [26]

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) + \mathbf{n}(k) \tag{1}$$

where $\mathbf{x}(k) \in \mathbb{R}^m$ is the vector of observed signals, $\mathbf{s}(k) \in \mathbb{R}^n$ is the vector of source signals, assumed to be zero mean and mutually statistically independent, $\mathbf{n}(k) \in \mathbb{R}^m$ is the vector of noise signals, whose elements are assumed to have zero mean, and be mutually statistically independent, and independent of the source signals [27], and k denotes the discrete time index. $\mathbf{A}$ is the matrix of the source-steering vectors, containing information about the sensors and direction of arrival of the source signals. In general, $\mathbf{A}$, typically referred to as the mixing matrix, is an unknown full column rank matrix and $\mathbf{A} \in \mathbb{R}^{m\times n}$. The aim of BSS is that of recovering the unobserved original source signals from the available measurements. The sources are recovered using the following linear separating system:

$$\mathbf{y}(k) = \mathbf{W}(k)\mathbf{x}(k) \tag{2}$$

where $\mathbf{y}(k) \in \mathbb{R}^n$ is an estimate of $\mathbf{s}(k)$, to within the ambiguities of BSS as explained later, and $\mathbf{W}(k) \in \mathbb{R}^{n\times m}$ is the separating matrix (or unmixing) matrix, whose pseudoinverse is an estimate of the mixing matrix. The NGA algorithm is a BSS technique which updates the separating matrix according to

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta(k)[\mathbf{I} - \mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k)]\mathbf{W}(k) \tag{3}$$

where $\mathbf{f}(\mathbf{y}(k))$ is an odd nonlinear function of the output $\mathbf{y}(k)$, called the activation function, $\eta(k)$ is a positive learning parameter, and $(\cdot)^T$ denotes vector transpose. Nonetheless, it is only possible to recover the sources up to a multiplicative constant, and their order cannot be predetermined. These ambiguities, inherent to the BSS problem, imply that the exact inverse of the mixing matrix cannot be obtained, so that perfect separation is achieved when the global mixing-separating matrix, defined as

$$\mathbf{P}(k) = \mathbf{W}(k)\mathbf{A} \tag{4}$$

tends toward a matrix with only one nonzero term in each row and column [25]. Conventional BSS assumes that at most one source has Gaussian distribution because, for Gaussian random variables, uncorrelatedness corresponds to independence. A measure of non-Gaussianity of a zero mean, unit variance signal z is given by its kurtosis $\kappa(z)$, or normalized fourth-order cumulant, which is defined as [28]

$$\kappa(z) = \langle z^4 \rangle - 3 \tag{5}$$

where $\langle \cdot \rangle$ denotes the statistical expectation. For a Gaussian signal, kurtosis is zero. Signals with negative kurtosis are called sub-Gaussian, while those with positive kurtosis are referred to as super-Gaussian. Additionally, the BSS problem is typically simplified further by assuming that there are at least as many sensors as sources, $m \geq n$ and, by convention, the source signals have unit variance. The performance of BSS methods can be assessed by plotting the following performance index (PI) [29]:

$$\mathbf{PI}(k) = \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{j=1}^{n}\frac{|p_{ij}|^2}{\max_q |p_{iq}|^2} - 1\right\} + \frac{1}{n}\sum_{j=1}^{n}\left\{\sum_{i=1}^{n}\frac{|p_{ij}|^2}{\max_q |p_{qj}|^2} - 1\right\} \tag{6}$$

where $[\mathbf{P}(k)]_{ij} = p_{ij}$, and n is the number of source signals. It is a nonnegative measure of the closeness between $\mathbf{W}(k)$ and the pseudoinverse of the mixing matrix, taking into account the scaling and ordering ambiguities, and generally, a lower PI indicates better performance.

[1]The authors wish to express their gratitude to Dr. A. Evans, at the University of Bath, for bringing this result to their attention.

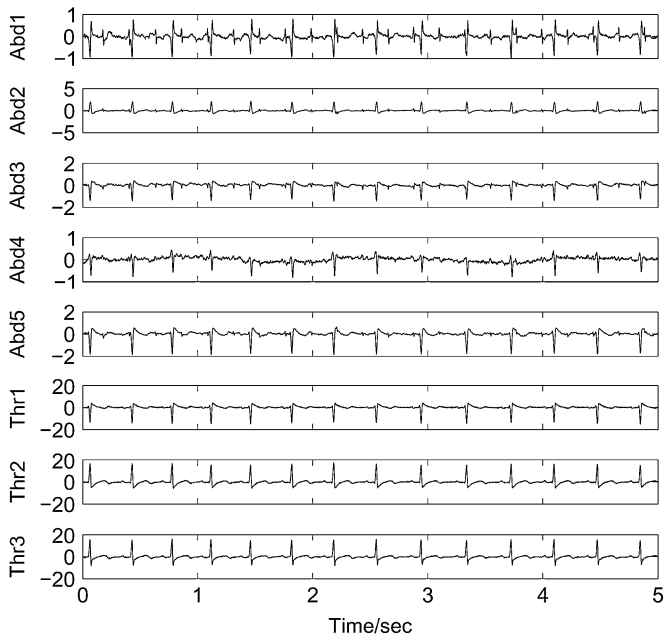

Fig. 1. An 8-channel cutaneous potential recording from a pregnant woman. The signals denoted Abd1–5 were recorded from the abdominal area, while the lower-most recordings Thr1–3 were obtained from the thoracic area.

III. Formulation of FECG Extraction as a BSS Problem

Fig. 1 shows the 8-channel cutaneous potential recording of a pregnant woman, obtained when eight electrodes are placed at different positions on the body of the expectant mother. In particular, five electrodes are located in the abdominal area, and three in the thoracic region. The electrocardiogram measurements were recorded over 5 s, and sampled at 500 Hz,[2] and represent mixtures of FECG and MECG contributions, as well as noise. The fetal heartbeat component, noise and respiratory motion artefacts are noticeable in the abdominal recordings, denoted in Fig. 1 as Abd1—Abd5, whereas FECG contributions are not visible in the thoracic measurements, which are dominated by the maternal heartbeat because of the distance between the fetus and the chest leads. The specific locations of the electrodes are justified by the fact that the maternal ECG components represent the largest signals interfering with the fetal electrocardiogram [13].

The separation of MECG and FECG components can be formulated as a BSS problem due to the following factors [7]–[10].

- MECG and FECG components are generated from physically independent bioelectric sources and, therefore, the sources may be considered statistically independent [9].
- The relationship between the bioelectric current sources and the surface of the body can be considered linear, and propagation delays can be considered negligible, so that the measurements may be assumed to be instantaneous linear mixtures [8], [10], [30].
- The mixing matrix is determined by the geometry of the body, positions of the sources and electrodes, and conductivity of the tissues of the body [10].

[2]Contributed to ICA Central http://sig.enst.fr/~cardoso/icacentral/index.html by L. De Lathauwer, K. U. Leuven, Belgium.

In addition, it is reported in [8]–[10] that the bioelectric activity of the maternal heart can be represented by a three-dimensional (3-D) dipole current. Then, MECG signals may be expressed as the linear combination of three statistically independent vectors. These form the MECG-subspace. The FECG-subspace, on the other hand, is not necessarily 3-D, but changes during pregnancy [8], [10]. Therefore, when ECG sources are separated with BSS algorithms, the resulting independent components should contain the MECG and FECG subspaces.

IV. The Wavelet Transform and BSS

The wavelet transform maps a signal from the time domain to the time-scale domain. A basic wavelet function is defined, called the mother wavelet, which is translated and dilated, resulting in a set of orthonormal wavelet basis functions [31]. Then, the wavelet transform of a signal is given by the inner product of the signal with each of the basis functions, so that the transformed signal is a function of the translation and scale parameters, where the term translation refers to the location of the window, while scale, defined as the inverse of frequency, refers to its width. Discrete wavelets are defined as

$$\psi_{j,q}(k) = 2^{-j/2}\psi(2^{-j}k - q) \tag{7}$$

where $j, q \in \mathbb{Z}$, and $\mathbb{Z}$ denotes the field of integers, and the wavelet transform of a signal $x(k)$ is given by

$$c_{j,q} = (x(k), \psi_{j,q}(k)) \tag{8}$$

where $c_{j,q}$ denotes the transform coefficients, and $(\cdot,\cdot)$ represents the inner product. In general, source separation in the wavelet domain introduces the permutation problem, which however is a well-known limitation of transform domain BSS, particularly for convolutive mixtures, for which the separation matrix will be different in each subband, while in the case of instantaneous mixtures the mixing is in effect identical in each subband. In our implementation, the outputs from the wavelet transform are concatenated, so as to form a single vector for source separation. The greatest disadvantage of performing source separation on the concatenated coefficients is the introduction of discontinuities, which result in fluctuations and absence of excitation in places, hence causing the adaptive process to slow down or even terminate at times.

Nonetheless, the wavelet transform is particularly useful for the analysis of nonstationary signals since it provides the time localization of frequency components, and it has the property of rendering many noiseless signals sparse when transformed in the wavelet domain. This means that the coefficients resulting from the signal are relatively large compared to those arising from noise, and the low-amplitude noise coefficients can be removed by setting them to zero. The process of truncation of the wavelet transform is performed usually by applying either a hard- or a soft-thresholding method [32]. Hard-thresholding involves setting the coefficients whose values are below a certain threshold to zero, and leaving the others unchanged. Soft-thresholding entails modifying the coefficients above the threshold as well. The thresholding method has a major drawback when

adaptive filtering is implemented in the wavelet domain: it may result in most of the coefficients being set to zero, thus causing a persistence of excitation problem. Although, it was found that generally de-noising of real measurements does not lead to a large number of zero coefficients, certain regions could be identified in which the wavelet coefficients are predominantly zero or very small, and they were found to correspond to poor performance of the adaptive filter.

While the wavelet decomposition provides an exact representation of a given signal, de-noising in wavelet bases leads to an approximation which uses a finite number of transform coefficients. In general, when a function f is approximated using only the first N terms of its projection in an orthonormal basis, the approximation is said to be linear, whereas it is nonlinear when the largest N terms are kept [33]. Since there exists many possible N-term approximations of a function, which may result in different rates of approximation, the crucial issue is how fast they converge. In [33], Vetterli compares the squared error (defined as the difference between the function f and the approximation $\hat{f}_N$) for linear and nonlinear approximation, when the Fourier and wavelet series coefficients of a piecewise constant function with a discontinuity are truncated, and shows that the nonlinear wavelet approximation has an exponential rate of convergence, while its linear counterpart, and linear and nonlinear Fourier approximations have convergence speed which is inversely proportional to N. This effectively implies that the nonlinear wavelet approximation requires a much smaller number of coefficients to represent a function accurately than its linear counterpart, and linear and nonlinear Fourier approximation. Thus, since the wavelet de-noising scheme is a nonlinear method, it is particularly advantageous because of the outstanding properties of nonlinear wavelet approximation.

V. Time-Scale Approach to Blind Source Separation

A. Modeling the Sample pdf in the Wavelet Domain

The statistics of the wavelet coefficients of natural images can be highly non-Gaussian, and can be modeled using a generalized Gaussian probability density function (pdf) of the form [34], [35]

$$f_{s,p}(c) = \frac{e^{-|c/s|^p}}{N(s,p)} \tag{9}$$

where $N(s,p) = 2s\Gamma(1/p)/p$, and $\Gamma(l) = \int_0^\infty t^{l-1}e^{-t}\,dt$ is the Gamma function. Expressions for the variance σ^2, and kurtosis κ of the distribution are defined in [34] and [35], and an analytical derivation of (9) can be found in [34]. To demonstrate the validity of the model in the 1-D case, we use the least squares curve fitting method from MATLAB (lsqcurvefit.m) to fit the generalized Gaussian distribution in (9) to the sample distribution of the wavelet domain representation of the signals depicted in Fig. 2, where the maternal and fetal ECG recordings, $s_1(k)$ and $s_2(k)$ are selected from the components extracted with the JADE algorithm, and de-noised with the wavelet transform. Their length is 2500 samples, while the length of the speech signals $s_3(k)$ and $s_4(k)$, is 10 000 samples.

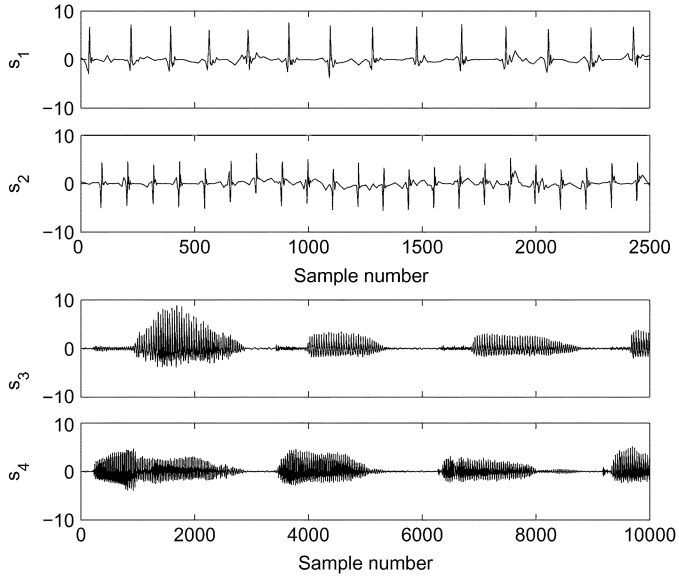

Fig. 2. Four original source signals: maternal ECG (s_1), fetal ECG (s_2), and two speech signals (s_3 and s_4).

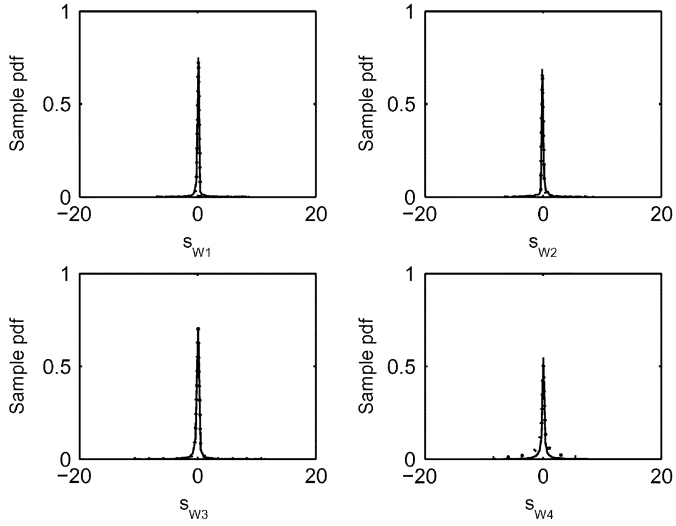

Fig. 3. Sample distributions (solid lines) and fitted Generalized Gaussian sample pdfs (dotted lines) for the wavelet coefficients of the sources in Fig. 2.

The sample pdfs and fitted sample distributions of the wavelet coefficients of the sources are illustrated in Fig. 3. The plots show that, for the selected sources, the generalized Gaussian distribution models the statistics of the wavelet coefficients very closely. It is not surprising that the poorest results are obtained when the sample pdf of the wavelet coefficient of $s_4(k)$ is modeled, since the two speech signals used here as sources, are in fact the output of the combined time-delayed decorrelation and ICA algorithm, and are estimated from two sensors recorded in a normal office room [36]. Consequently, each source is likely to be contaminated by noise, reverberation in the room and, to a certain extent, by the other source. Thus, the Gaussian pdf is fitted to the sample distributions of the wavelet representation of sources $s_3(k)$ and $s_4(k)$, as shown in Fig. 4. A comparison with the plots in Fig. 3, indicates that the generalized Gaussian distribution is a better fit than the Gaussian pdf. Figs. 5 and 6 show the sample distributions of the sources in the time domain and

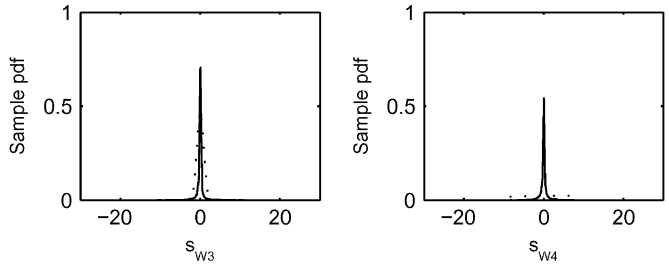

Fig. 4. Sample distributions (solid lines) and fitted Gaussian sample pdfs (dotted lines) for the wavelet coefficients of the sources $s_3(k)$ and $s_4(k)$.

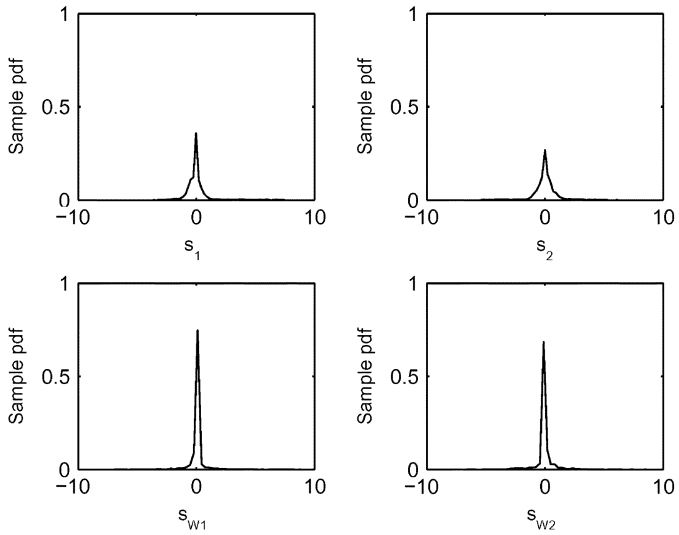

Fig. 5. Sample pdfs of the sources s_1 and s_2 in Fig. 2, and of their corresponding wavelet coefficients.

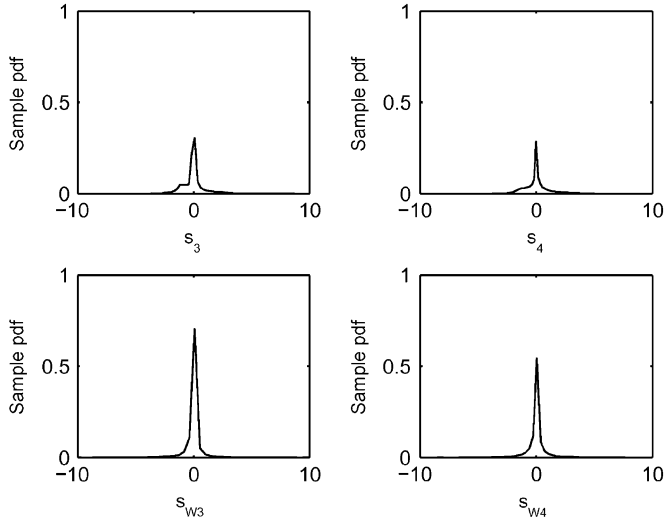

Fig. 6. Sample pdfs of the sources s_3 and s_4 in Fig. 2, and of their wavelet coefficients.

in the wavelet domain, while the kurtoses of the signals in both domains, evaluated according to (5), are recorded in Table I. The kurtoses of the sources increase quite dramatically in some cases. Kurtosis is a measure of non-Gaussianity of a random variable: the further away the pdf of a random variable is from the Gaussian distribution, the further away is its kurtosis from zero. In this case, it is clear from Figs. 5 and 6 that the sample

TABLE I
KURTOSES OF THE SOURCES IN FIG. 2, $\kappa(s_i)$, AND OF THEIR WAVELET COEFFICIENTS, $\kappa(s_{Wi})$

Source	$\kappa(s_i)$	$\kappa(s_{\mathcal{W}i})$
$s_1(k)$	19.96	26.00
$s_2(k)$	9.22	18.41
$s_3(k)$	9.85	28.72
$s_4(k)$	2.48	12.63

pdfs of the sources in the time domain are closer to the Gaussian distribution than their respective sample density functions in the time-scale domain. Thus, the statistics of the source signals are less Gaussian in the wavelet domain than they are in the time domain. These results will be useful in examining the convergence behavior of the time-scale approach.

B. Performance Index

Before introducing the new BSS approach in the wavelet domain, some measure is needed to assess its performance, and to compare it to time domain methods. As mentioned in Section IV, the wavelet coefficients $c_{j,q}$ of a signal $x(k)$ are given by the inner product of the signal with each wavelet

$$c_{j,q} = (x(k), \psi_{j,q}(k)). \tag{10}$$

This expression, together with the linearity property of the inner product

$$((x(k) + z(k)), y(k)) = (x(k), y(k)) + (z(k), y(k)) \tag{11}$$

$$(\alpha x(k), y(k)) = \alpha(x(k), y(k)) \tag{12}$$

leads to the following result, useful for performance evaluation purposes, provided the mixing matrix is known. When the mixing matrix $\mathbf{A}$ is real and time-invariant, the wavelet transform of the ith observed signal $x_i(k)$ in (1) is given by

$$\mathcal{W}\{x_i(k)\} = \mathcal{W}\left\{\sum_{l=1}^{n} a_{il} s_l(k)\right\}$$

$$(x_i(k), \psi_{j,q}(k)) = \sum_{l=1}^{n} a_{il}(s_i(k), \psi_{j,q}(k)). \tag{13}$$

Taking the wavelet transform of the vector of observed signals, we obtain, in matrix form

$$\mathbf{c}_x = \mathbf{A}\mathbf{c}_s \tag{14}$$

where $\mathbf{c}_x = [(x_1(k), \psi_{j,q}(k)), \ldots, (x_n(k), \psi_{j,q}(k))]$ and $\mathbf{c}_s = [(s_1(k), \psi_{j,q}(k)), \ldots, (s_m(k), \psi_{j,q}(k))]$ are, respectively, vectors of wavelet transformed sensors and sources. It follows then that the sources estimated in the wavelet domain $\mathbf{y}_{\mathcal{W}}(k)$ are given by

$$\mathbf{y}_{\mathcal{W}}(k) = \mathbf{W}(k)\mathbf{c}_x = \mathbf{W}(k)\mathbf{A}\mathbf{c}_s. \tag{15}$$

Therefore, the PI remains a meaningful performance measure for BSS algorithms operating in the time-scale domain.

C. Time-Scale Approach

To perform separation in the time-scale domain, the mixed signals are divided into blocks of length N, so that the independent components can be extracted as each new block of data becomes available. When the length of the measurements is M samples, the first block comprises samples $\mathbf{x}(1),\ldots,\mathbf{x}(N)$, the second block is made up by $\mathbf{x}(N+1),\ldots,\mathbf{x}(2N)$, and so on, and in general the ith block includes samples $\mathbf{x}(iN+1),\ldots,\mathbf{x}((i+1)N)$, where $i=0,\ldots,(M/N)-1$. The wavelet transform of each block of data is then evaluated, and hard-thresholding is applied. Reducing the noise level is expected to improve the performance of NGA because, invariably, true measurements are noisy, while the algorithm is derived on the assumption that the sources are mixed in the absence of noise. Although not satisfying this hypothesis fully, the mixtures obtained after de-noising are better suited for processing by NGA than prior to noise removal. Finally, NGA is used to separate sequentially the transformed signals according to

$$\begin{aligned}\mathbf{y}_{\mathcal{W}}(k) &= \mathbf{W}(k)\mathbf{x}_{\mathcal{W}_N}(k)\\ \mathbf{W}(k+1) &= \mathbf{W}(k)+\eta(k)\left(\mathbf{I}-\mathbf{f}_{\mathcal{W}}(k)\mathbf{y}_{\mathcal{W}}^{T}(k)\right)\mathbf{W}(k).\end{aligned}\quad(16)$$

Since the algorithm operates on a single block of data, this gives a vector of estimated signals, whose inverse wavelet transform represents the corresponding block of recovered sources. Thus, the NGA in the time-scale domain approach can be summarized as shown in Fig. 7. It should be noted that the block-by-block application of NGA has the potential drawback of introducing the permutation problem, due to the concatenation of the outputs, prior to the evaluation of the inverse wavelet transform. Nonetheless, as mentioned in Section IV, this problem is typically not experienced in the case of instantaneous mixtures, and it was not encountered by the authors.

To ease the problem of absent filter excitation associated with discontinuities described in Section IV, the approach above can be modified such that it includes a scheme to switch to conventional NGA once convergence is achieved, and switch back to the time-scale method if reconvergence is necessary.

NGA requires *a priori* knowledge about the statistics of the sources, because different nonlinearities $\mathbf{f}(\mathbf{y}(k))$ in (3) are selected for the separation of sub- and super-Gaussian sources. Moreover, the algorithm may fail to separate the sources when mixtures of both sub- and super-Gaussian signals are observed. To address these difficulties, Douglas *et al.* propose in [37] to employ time-varying nonlinearities, appropriately selected for each channel according to the statistics of the corresponding estimated source. In the wavelet domain, however, the problem of switching between activation functions can be mitigated, because the wavelet coefficients of all the sources can generally be modeled by a generalized Gaussian density and, therefore, the activation function need not change when the sources are sub-Gaussian.

1) Convergence of Algorithm: It has been established previously that the statistics of the wavelet representation of certain signals are less Gaussian than the statistics of the signal itself, and more precisely, they are more super-Gaussian. In particular, the kurtosis was found to increase in the time-scale domain. Mathematically, let κ_i and $\kappa_i^{\mathcal{W}}$ be, respectively, the kurtosis of the ith signal and of its wavelet coefficients; it was observed from simulations that in general

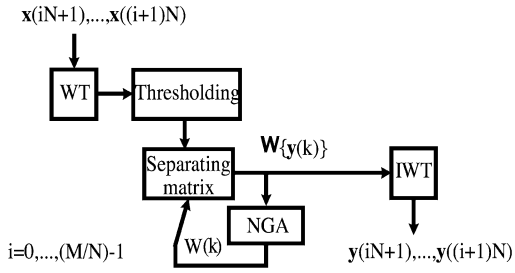

Fig. 7. Diagram of the NGA in the wavelet domain approach. The mixtures are divided into blocks of length N, so that the ith block is formed by samples $\mathbf{x}(iN+1),\ldots,\mathbf{x}((i+1)N)$, where $i=0,\ldots,(M/N)-1$. The wavelet transform of each block of data is then evaluated, and hard-thresholding is applied to the concatenated wavelet coefficients. NGA separates the transformed signals sequentially, giving a vector of estimated signals, whose inverse wavelet transform represents the corresponding block of recovered sources.

$$\kappa_i^{\mathcal{W}} > \kappa_i \quad(17)$$

An expression for the global mixing-separating system in terms of the kurtoses of the sources can be derived. In the time domain, the global system is given by

$$\mathbf{P}(k+1)=\mathbf{P}(k)+\eta(k)[\mathbf{I}-\mathbf{f}(\mathbf{y}(k))\mathbf{y}^{T}(k)]\mathbf{P}(k). \quad(18)$$

Characterization of the transient behavior of the above system is typically a very challenging task, due to the cross-coupling of the elements of the global mixing-separating matrix [38]. The convergence speed of the algorithm depends on the second term on the right-hand side of the above expression, i.e., the product $\eta(k)[\mathbf{I}-\mathbf{f}(\mathbf{y}(k))\mathbf{y}^{T}(k)]\mathbf{P}(k)$. Let $\mathbf{F}_y(k)=\mathbf{f}(\mathbf{y}(k))\mathbf{y}^{T}(k)$. To a first approximation, an increase in $(\mathbf{I}-\mathbf{F}_y(k))$ results effectively in the algorithm taking a larger step in the descent direction, which is desirable during initial convergence, when the filter parameters are away from their optimal values. Convergence of the mean of the algorithm, on the other hand, is ensured when

$$\lim_{k\longrightarrow\infty}\langle\mathbf{F}_y(k)\rangle=\mathbf{I}. \quad(19)$$

Thus, some growth in the diagonal elements of $-\mathbf{F}_y(k)$ will generally increase the convergence speed of the algorithm, as will a more rapid decay of the off-diagonal elements. It will be shown later that these reasons represent a first step toward justifying the improved performance of NGA operating in the wavelet domain.

When the sources are super-Gaussian, the activation function can be chosen to be $f_i(y_i(k))=\tanh(y_i(k))$. In [39], Amari *et al.* approximate $\tanh(y_i(k))$ with Maclaurin's series up to degree 5. In this paper, since we seek to express the system in (18) in terms of the kurtoses of the sources, we truncate Maclaurin's approximation to degree 3. Thus, ignoring the time index for convenience, this gives

$$\tanh(y_i)\simeq y_i-\frac{1}{3}y_i^3 \quad(20)$$

and the elements of $\mathbf{F}_y(k)$ in (18) can be written as

$$\tanh(y_i)y_j\simeq\left(y_i-\frac{1}{3}y_i^3\right)y_j,\quad\forall i,j. \quad(21)$$

Substituting the following expression into (21)

$$\mathbf{y} = \mathbf{AWs} = \mathbf{Ps} \Longrightarrow$$
$$y_i = \sum_{j=1}^{m} p_{ij} s_j \tag{22}$$

and applying the statistical expectation operator, the entries in $\mathbf{F}_y(k)$ can be written in terms of the elements of the global mixing-separating system $\mathbf{P}(k)$, giving

$$\langle \tanh(y_i) y_j \rangle \simeq \left\langle \left(1 - \sum_{l=1}^{m} p_{il}^2\right) \sum_{l=1}^{m} p_{il} p_{jl} \right\rangle - \frac{1}{3} \sum_{l=1}^{m} \langle p_{il}^3 p_{jl} \rangle \kappa_l \tag{23}$$

where the expectation is with respect to the elements of the separating matrix $\mathbf{W}(k)$, and the source signals. Moreover, as in [38] and [40], it has been assumed that the elements of the separating matrix are independent of the sources, implying that the elements of the global matrix $\mathbf{P}(k)$ are also independent of the sources; the sources are zero mean and unit variance, and the definition of kurtosis $\kappa_i = \langle s_i^4 \rangle - 3$ has been used. Considering only the diagonal terms of $\langle \mathbf{F}_y(k) \rangle$, namely $\langle \tanh(y_i) y_i \rangle, i = 1, \ldots m$, and reintroducing the time index, at time $k = 0$ we have

$$\langle \tanh(y_i(0)) y_i(0) \rangle \simeq \left\langle \left(1 - \sum_{l=1}^{m} p_{il}^2(0)\right) \sum_{l=1}^{m} p_{il}^2(0) \right\rangle - \frac{1}{3} \sum_{l=1}^{m} \langle p_{il}^4(0) \rangle \kappa_l. \tag{24}$$

When the sources are separated in the time-scale domain, (18) becomes

$$\mathbf{P}^{\mathcal{W}}(k+1) = \mathbf{P}^{\mathcal{W}}(k) + \mu(k) \left[\mathbf{I} - \mathbf{f}(\mathbf{y}_{\mathcal{W}}(k)) \mathbf{y}_{\mathcal{W}}^T(k)\right] \mathbf{P}^{\mathcal{W}}(k). \tag{25}$$

Equation (25) indicates that generally, at any time k, the global matrix $\mathbf{P}(k) \neq \mathbf{P}^{\mathcal{W}}(k)$, because typically the algorithms (18) and (25) will have different dynamical characteristics. At time $k = 0$, however, and assuming that the initialization of the separating matrix $\mathbf{W}(k)$ in the wavelet domain is the same as in the time domain, (24) becomes

$$\langle \tanh(y_{\mathcal{W}i}(0)) y_{\mathcal{W}i}(0) \rangle \simeq \left\langle \left(1 - \sum_{l=1}^{m} p_{il}^2(0)\right) \sum_{l=1}^{m} p_{il}^2(0) \right\rangle - \frac{1}{3} \sum_{l=1}^{m} \langle p_{il}^4(0) \rangle \kappa_l^{\mathcal{W}}. \tag{26}$$

Since $\langle p_{il}^4(0) \rangle > 0, \forall i, l$ and from (17), the inequality $\langle p_{il}^4(0) \rangle \kappa_l^{\mathcal{W}} > \langle p_{il}^4(0) \rangle \kappa_l$ holds. It follows that

$$\sum_{l=1}^{m} \langle p_{il}^4(0) \rangle \kappa_l^{\mathcal{W}} > \sum_{l=1}^{m} \langle p_{il}^4(0) \rangle \kappa_l. \tag{27}$$

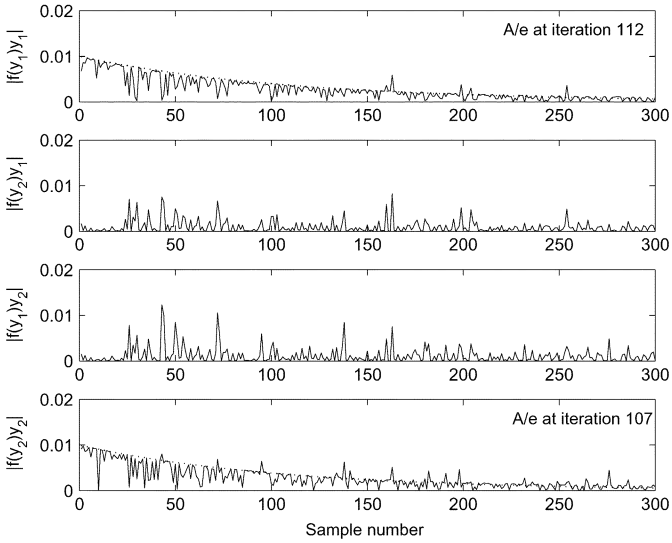

Fig. 8. Magnitude plots of the evolution of the elements of $\eta(k)(\mathbf{I} - \mathbf{F}_y(k))$. The diagonal elements (solid lines) are fitted with exponential envelopes (dotted lines), and the number of iterations required for the amplitude of the exponential fits to decay to $1/e$ of its initial value is evaluated.

Multiplying (27) by $-((1/3))$ and adding $\langle (1 - \sum_{l=1}^{m} p_{il}^2(0)) \sum_{l=1}^{m} p_{il}^2(0) \rangle$ to both sides gives

$$\left\langle \left(1 - \sum_{l=1}^{m} p_{il}^2(0)\right) \sum_{l=1}^{m} p_{il}^2(0) \right\rangle - \frac{1}{3} \sum_{l=1}^{m} \langle p_{il}^4(0) \rangle \kappa_l^{\mathcal{W}} < \left\langle \left(1 - \sum_{l=1}^{m} p_{il}^2(0)\right) \sum_{l=1}^{m} p_{il}^2(0) \right\rangle - \frac{1}{3} \sum_{l=1}^{m} \langle p_{il}^4(0) \rangle \kappa_l. \tag{28}$$

Finally, considering $\langle -\mathbf{F}_y(0) \rangle$, and $\langle -\mathbf{F}_y^{\mathcal{W}}(0) \rangle$, (28) becomes

$$\left\langle -\left(1 - \sum_{l=1}^{m} p_{il}^2(0)\right) \sum_{l=1}^{m} p_{il}^2(0) \right\rangle + \frac{1}{3} \sum_{l=1}^{m} \langle p_{il}^4(0) \rangle \kappa_l^{\mathcal{W}} > \left\langle -\left(1 - \sum_{l=1}^{m} p_{il}^2(0)\right) \sum_{l=1}^{m} p_{il}^2(0) \right\rangle + \frac{1}{3} \sum_{l=1}^{m} \langle p_{il}^4(0) \rangle \kappa_l. \tag{29}$$

Thus, in an element-by-element sense

$$\operatorname{diag}\left(\mathbf{I} - \mathbf{f}_y^{\mathcal{W}}(0)\right) > \operatorname{diag}(\mathbf{I} - \mathbf{f}_y(0)). \tag{30}$$

In general, assuming that at time k, the global matrix $\mathbf{P}(k)$ has the same value in both the time and wavelet domains, (30) becomes

$$\operatorname{diag}\left(\mathbf{I} - \mathbf{f}_y^{\mathcal{W}}(k)\right) > \operatorname{diag}(\mathbf{I} - \mathbf{f}_y(k)). \tag{31}$$

In the described analysis, the off-diagonal terms of $(\mathbf{I} - \mathbf{f}_y(k))$ and $(\mathbf{I} - \mathbf{f}_y^{\mathcal{W}}(k))$ have been ignored because during initial convergence the diagonal elements are large and dominate the performance of the algorithm. This is illustrated in Figs. 8 and 9 which show the evolution of the elements of the matrices $\eta(k)(\mathbf{I} - \mathbf{f}_y(k))$ and $\mu(k)(\mathbf{I} - \mathbf{f}_y^{\mathcal{W}}(k))$, for the case of two randomly generated Laplacian sources, with zero mean and unit variance, and $n = m$, averaged over 30 independent trials, where the contributions of the step-size parameters have been taken into account because, due to their self-adaptive nature, they play a role in the behavior of the algorithm. Evidently, in the

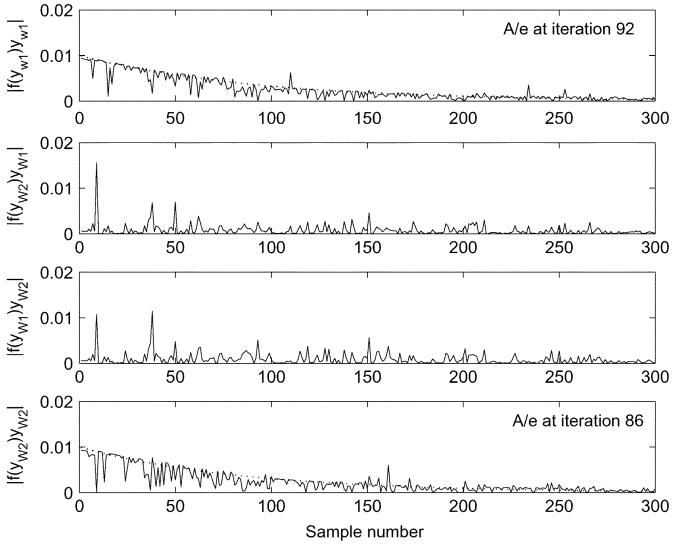

Fig. 9. Magnitude plots of the evolution of the elements of $\mu(k)(\mathbf{I}-\mathbf{f}_y^{\mathcal{W}}(k))$. The diagonal elements (solid lines) are fitted with exponential envelopes (dotted lines), and the number of iterations required for the amplitude of the exponential fits to decay to $1/e$ of its initial value is evaluated.

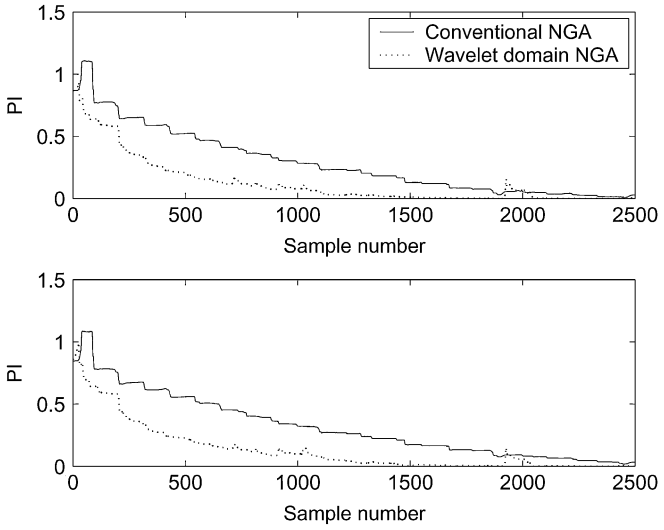

Fig. 10. Performance index obtained with NGA and NGA and wavelet transform, for the fetal and maternal ECG sources in Fig. 2, when zero mean Gaussian noise is present. The upper plots depict the PI when the SNR is 10 dB, while the lower plots correspond to a SNR of 5 dB.

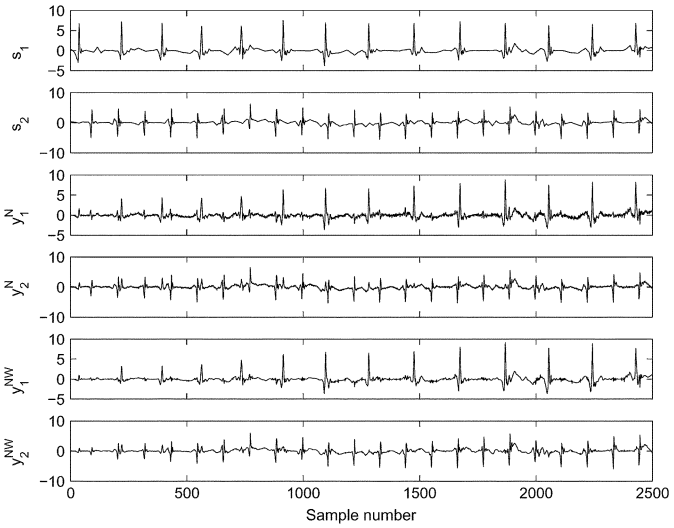

Fig. 11. Sources recovered with conventional NGA and the NGA in the wavelet domain method, when the SNR is 10 dB. The waveforms s_1 and s_2 correspond to the original maternal and fetal ECGs; y_1^N and y_2^N represent the components extracted with NGA, and y_1^{NW} and y_2^{NW} those obtained with the time scale approach.

time as well as in the wavelet domain, the diagonal elements are initially greater in magnitude than the off-diagonal elements. In addition, exponential envelopes are fitted to the diagonal terms, so that the rate of convergence of the two algorithms can be compared by evaluating the number of iterations required for the amplitude of each exponential fit to decay to $1/e$ of its initial value (see Figs. 8 and 9). The results indicate that the diagonal elements of $\mu(k)(\mathbf{I}-\mathbf{f}_y^{\mathcal{W}}(k))$ decay more rapidly than those of $\eta(k)(\mathbf{I}-\mathbf{f}_y(k))$. Thus, the convergence speed of NGA becomes faster when the algorithm separates the signals in the wavelet domain.

VI. Experimental Results

In order to study the behavior of the wavelet domain approach, the signals in Fig. 2 are taken in pairs and mixed by a stationary instantaneous mixing matrix $\mathbf{A}$, randomly generated, and given by

$$\mathbf{A}=\begin{bmatrix}-1.493 & 0.729\\ 1.496 & 2.367\end{bmatrix} \tag{32}$$

Also, zero mean Gaussian noise at 10 dB and 5-dB signal-to-noise ratio (SNR) is added. Since the sources are super-Gaussian, the activation function in (3) is chosen as $f_i(y_i)=\tanh(2y_i)$ [25]; also the separating matrix $\mathbf{W}(k)$ is initialized to $0.2\mathbf{I}$. The mixtures are separated, in 30 independent trials, with the NGA, and with the NGA in the wavelet domain method, using the wavelet transform; the Daubechies filters of length 8 are selected. The performance indices for the two methods, and for the two sets of original source signals, are compared in Fig. 10, when the noise level is 10-dB and 5-dB SNR. Plots of PIs for ordinary NGA are depicted in the figure as solid lines, while the dotted lines represent the performance indices obtained when NGA operates in the time-scale domain. Separation with the NGA in the wavelet domain results generally in faster convergence speed than when it separates in the time domain. Figs. 11 and 12 show the original maternal and fetal ECGs (s_1 and s_2), and the sources recovered with conventional (y_1^N and y_2^N) and wavelet domain NGA (y_1^{NW} and y_2^{NW}), when the noise level is 10 dB and 5 dB SNR, respectively. Both figures illustrate that each estimated signal is contaminated by noise and by the other source to a higher degree when separation is performed with time domain NGA rather than the wavelet domain method. In particular, the FECG component is more prominent in the output y_1^N, representing an estimate of the maternal ECG recording by conventional NGA, than in y_1^{NW}. Similarly, the contribution of the MECG source to y_2^{NW} is significantly reduced after about 800 samples, while it has generally larger amplitude, and is still visible in y_2^N up to about 1400 iterations. Improved algorithm performance, observed when the sources are estimated in the

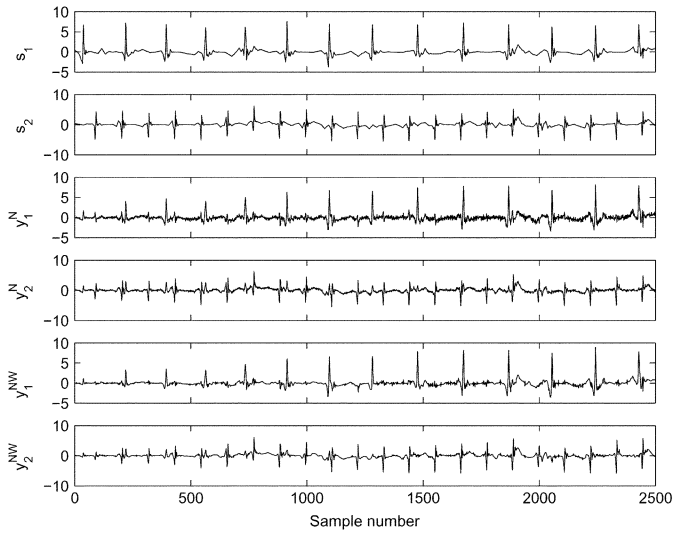

Fig. 12. Original MECG and FECG sources (s_1 and s_2), and signals recovered by time domain NGA (y_1^N and y_2^N), and the wavelet domain method (y_1^{NW} and y_2^{NW}), when the SNR is 5 dB.

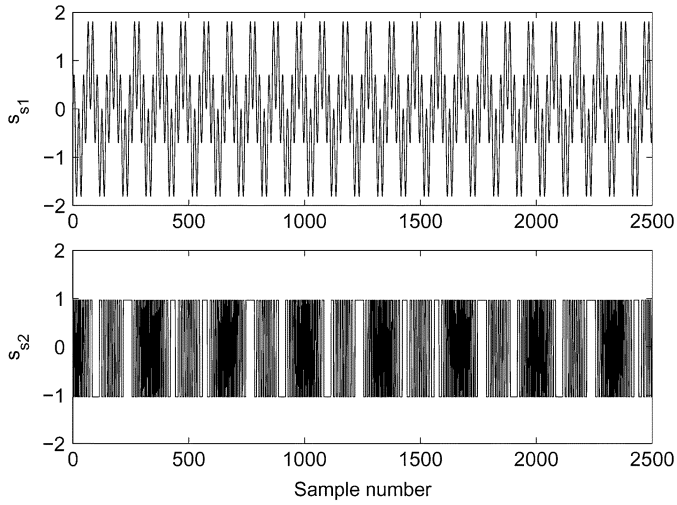

Fig. 13. Two periodic sub-Gaussian source signals.

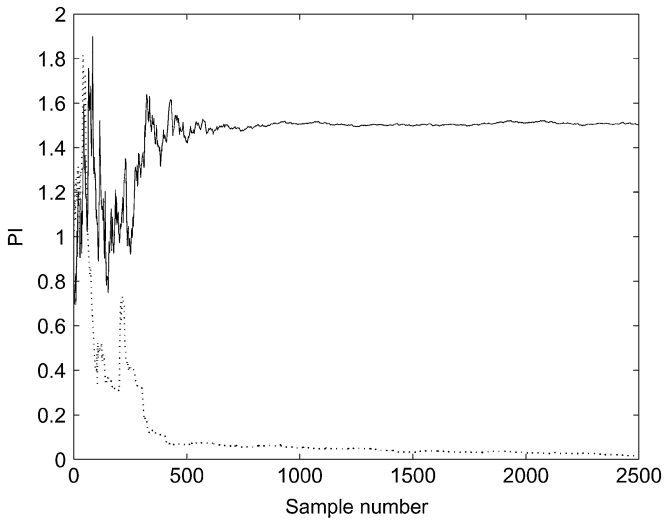

Fig. 14. Performance indices obtained for NGA (solid line), and wavelet domain NGA (dotted line), when the two original sources are sub-Gaussian and the nonlinearity is super-Gaussian.

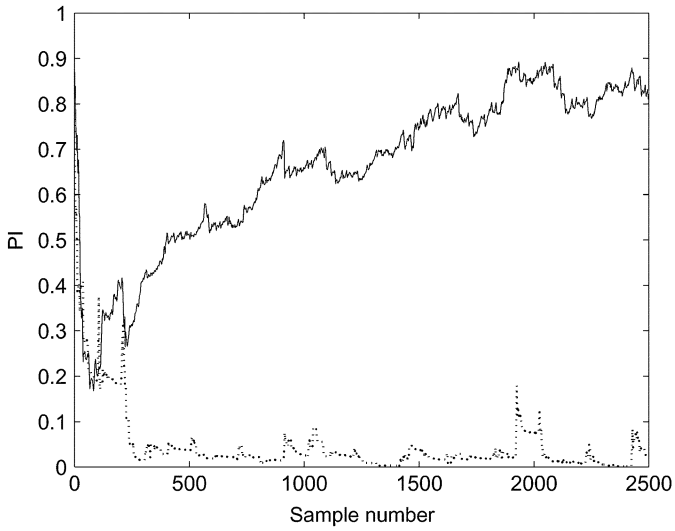

Fig. 15. Performance indices obtained for NGA (solid line), and NGA and WT (dotted line), when two of the original sources are sub-Gaussian and one is super-Gaussian, and the nonlinearity is super-Gaussian.

time-scale domain, is partly explained by the de-noising action of the wavelet transform. However, a more important factor is that the sample pdfs of the original source signals are closer to a Gaussian distribution than the pdfs of the corresponding wavelet coefficients. It should also be noted that the issue of wavelet selection has not been addressed in this study. Some of the difficulties encountered when dealing with natural signals might be alleviated by choosing an alternative wavelet family whose characteristics are more closely matched to the particular signal, but this is outside of the scope of this work.

The most remarkable advantage of the NGA in the time-scale domain approach is that it can allow the separation of sub-Gaussian, and mixtures of both sub- and super-Gaussian signals without the need to switch between distinct nonlinearities. Fig. 14 shows the average performance of the NGA and NGA in the wavelet domain methods, over 30 independent trials, when the two synthetic sub-Gaussian sources in Fig. 13 are mixed by a time-invariant mixing matrix shown in (32), and the separating matrix is initialized to the identity matrix $\mathbf{I}$. We select a super-Gaussian nonlinearity $f_i(y_i(k)) = \tanh(2y_i(k))$ to separate the sources with both techniques. The results clearly show that when operating in the time domain NGA diverges, thus failing to separate the sources, because the activation function does not match the statistics of the sources. When separation is carried out in the time-scale domain, on the other hand, the algorithm converges quite quickly and the PI remains low thereafter: since the wavelet coefficients of the sources have a super-Gaussian distribution, the nonlinearity matches their statistics.

Next, the maternal ECG in Fig. 2 $(s_1(k))$ and the two sub-Gaussian sources in Fig. 13 are mixed as above. In practice, such a situation arises when, for instance, ECG recordings are corrupted by periodic power line noise, which arises due to the power supply. The separation results shown in Fig. 15 indicate that the wavelet approach successfully estimates the sources using a single, super-Gaussian nonlinearity, while

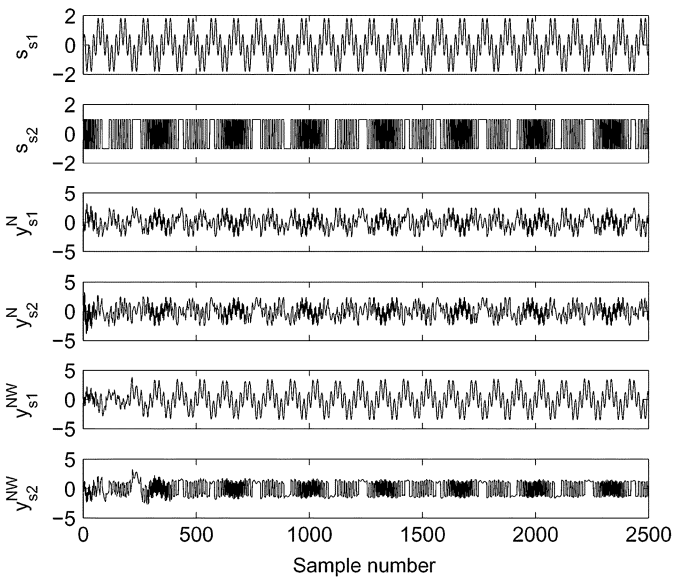

Fig. 16. Original speech signals (s_{s1} and s_{s2}), and sources recovered with conventional NGA (y_{s1}^{N} and y_{s2}^{N}) and wavelet domain NGA (y_{s1}^{NW} and y_{s2}^{NW}).

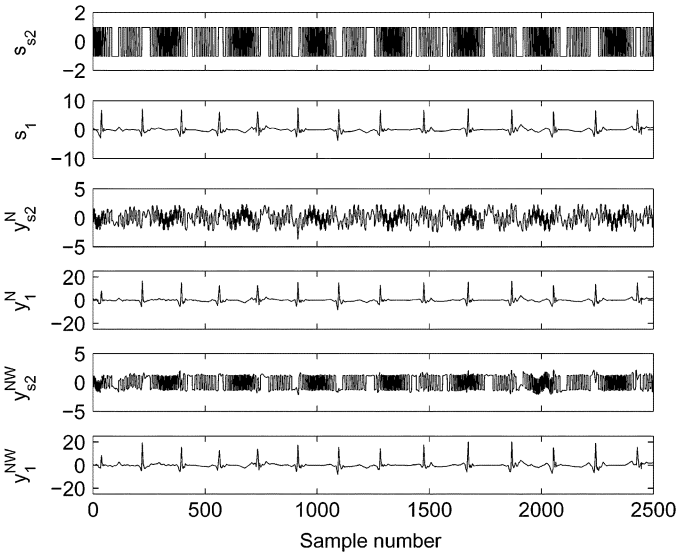

Fig. 17. Original signals (s_{s2} and s_1), and sources recovered with conventional NGA (y_{s2}^{N} and y_1^{N}) and wavelet domain NGA (y_{s2}^{NW} and y_1^{NW}).

ordinary NGA diverges. The selection of appropriate nonlinearities, matching each output signal is a crucial issue in BSS, and becomes even more important as the number of sources increases. In the time domain, separating mixtures of sub- and super-Gaussian sources requires sub- and super-Gaussian activation functions, operating on the relative outputs. In the wavelet domain, on the other hand, the transform coefficients of the sources become super-Gaussian and, therefore, a single super-Gaussian nonlinearity leads to successful source separation.

The sources recovered when only two sub-Gaussian sources are mixed, are depicted in Fig. 16. The third and fourth plots (y_{s1}^{N} and y_{s2}^{N}) show how ordinary NGA fails to converge, while in the time-scale domain the algorithm converges within less than 300 iterations, and the sources are separated successfully (y_{s1}^{NW} and y_{s2}^{NW}). When the original sources are the two sub-Gaussian signals and the maternal ECG, the recovered sources are as illustrated in Fig. 17. The only sub-Gaussian output plotted is s_{s2}, for the sake of clarity, and because, as it can be seen from Fig. 16, it is the hardest to recover. In general, the signals recovered appear to be more accurate when NGA operates in the time-scale domain.

VII. Conclusion

The problem of noisy, instantaneous mixtures is addressed in this paper by separating the sources in the time-scale domain, and using wavelet de-noising. It has been shown that the sample pdf of the wavelet coefficients of certain 1-D source signals fits a generalized Gaussian distribution, and that the global mixing-separating system depends on the kurtoses of the sources, which are higher in the wavelet domain. As a consequence, the NGA generally has higher convergence speed when operating in the time-scale domain, than in the time domain. Finally, the time-scale approach has been shown to have the ability to separate sub-Gaussian sources, and mixtures of both sub- and super-Gaussian signals using a single, super-Gaussian nonlinearity. Thus, performing BSS in the wavelet domain can mitigate the problem of switching between different activation functions.

References

[1] L. Y. Shyu, C. F. Huang, Y. S. Wu, W. C. Hu, and K. C. Chao, "The use of thoracic-abdominal transfer function in extracting fetal electrocardiogram," in *Proc. Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society*, vol. 4, 1996, pp. 1644–1645.

[2] *Handbook of Electromedicine: Basic Principles, Applications, Equipment*, H. Kresse, Ed., Siemens, Chichester, U.K., 1985. initiated by J. Patzold.

[3] A. Kam and A. Cohen, "Maternal ECG elimination and fetal ECG detection—Comparision of several algorithms," in *Proc. Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society*, vol. 1, 1998, pp. 174–177.

[4] Y. Datian, C. Yu, and G. Qin, "Wavelet analysis method for processing and recognition of abdominal fetal ECG waveform," in *Proc. IEEE Int. Conf. Electronics, Circuits, and Systems*, vol. 3, 1998, pp. 121–124.

[5] T. Kao, K. L. Wu, B. C. Yu, and J. H. Hung, "Digital signal enhancement of the abdominal fetal ECG," in *Proc. Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society*, vol. 2, 1989, pp. 771–772.

[6] J. A. Crowe, W. Peasgood, and M. S. Woolfson, "Extraction of the abdominal fetal electrocardiogram for use as an indicator of antenatal fetal status," in *Proc. Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society*, vol. 6, 1992, pp. 2499–2500.

[7] V. Zarzoso, A. K. Nandi, and E. Bacharakis, "Maternal and fetal ECG separation using blind source separation methods," *IMA J. Math. Appl. Med. Biol.*, vol. 14, pp. 207–225, 1997.

[8] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "Fetal electrocardiogram extraction by blind source subspace separation," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 5, pp. 567–572, May 2000.

[9] V. Zarzoso and A. K. Nandi, "Comparison between blind separation and adaptive noise cancellation techniques for fetal electrocardiogram extraction," *Inst. Elect. Eng. Colloq. Medical Applications for Signal Processing*, pp. 1/1–1/6, 1999.

[10] A. K. Nandi and V. Zarzoso, "Fetal ECG separation," *Inst. Elect. Eng. Colloq. Use of Model Based Digital Signal Processing Techniques in the Analysis of Biomedical Signals*, pp. 8/1–8/6, 1997.

[11] A. Boushia-Salah, A. Belouchrani, and A. Chichocki, "Application of time-frequency distributions to the independent component analysis of ECG signal," in *Proc. Int. Symp. Signal Processing and Its Applications*, 2001, pp. 238–241.

[12] E. Bacharakis, A. K. Nandi, and V. Zarzoso, "Fetal ECG extraction using blind source separation methods," in *Proc. Eur. Signal Processing Conf.*, 1996, pp. 395–398.

[13] V. Zarzoso and A. N. Nandi, "Noninvasive fetal electrocardiogram extraction: Blind separation versus adaptive noise cancellation," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 1, pp. 12–18, Jan. 2001.
[14] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. New York: Wiley, 2002.
[15] D. L. Jones, "A new method for blind source separation of nonstationary signals," in *Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Processing*, vol. 5, 1999, pp. 2893–2896.
[16] K. Matsuoka, M. Ohya, and M. Kawamoto, "A neural net for blind separation of nonstationary signals," *Neural Netw.*, vol. 8, pp. 411–419, 1995.
[17] S. Choi and O. Y. Lee, "Nonstationary source separation," in *Proc. IEEE Region 10 Conf.*, vol. 1, 1999, pp. 670–673.
[18] D. Pham and J. F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1837–1848, Sep. 2001.
[19] S. Haykin, Ed., *Unsupervised Adaptive Filtering*: Wiley, 2000, vol. I, Blind Source Separation.
[20] A. Cichocki, W. Kasprzak, and S. Amari, "Adaptive approach to blind source separation with cancellation of additive and convolutional noise," in *Proc. IEEE Int. Conf. Signal Processing*, vol. 1, 1996, pp. 412–415.
[21] S. Cruces, A. Cichocki, and L. Castedo, "Blind source extraction in Gaussian noise," in *Proc. Int. Workshop Independent Component Analysis and Blind Signal Separation*, 2000, pp. 63–68.
[22] A. Parashiv-Ionescu, C. Jutten, K. Aminian, B. Nahafi, and P. Robert, "Source separation in strong noisy mixtures: A study of wavelet de-noising preprocessing," in *Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Processing*, 2002, pp. 1681–1684.
[23] Y. Qi, P. Krishnaprasad, and S. Shamma, "The subband-based independent component analysis," in *Proc. Int. Workshop Independent Component Analysis and Blind Signal Separation*, 2000, pp. 199–204.
[24] M. Davies, "Audio source separation," in *Proc. Fifth IMA Int. Conf. Mathematics in Signal Processing*, 2000, pp. 11–14.
[25] S. Amari and A. Cichocki, "Adaptive blind signal processing—Neural network approaches," *Proc. IEEE*, vol. 86, no. 10, pp. 2026–2048, Oct. 1998.
[26] J. Karhunen, "Neural approaches to independent component analysis and source separation," in *Proc. Eur. Symp. Artificial Neural Networks*, 1996, pp. 249–266.
[27] V. Zarzoso, "Closed-form higher-order estimators for blind separation of independent signals in instantaneous linear mixtures," Ph.D. dissertation, Univ. Liverpool, Liverpool, U.K., 1999.
[28] H. Manthis, "On the kurtosis of digitally modulated signals with timing offesets," in *Proc. IEEE Workshop Signal Processing Advances in Wireless Communications*, 2001, pp. 20–23.
[29] A. Cichocki, B. Orsier, A. Back, and S. Amari, "On-line adaptive algorithms in nonstationary environments using a modified conjugate gradient approach," in *Proc. IEEE Workshop Neural Networks for Signal Processing*, 1997, pp. 316–325.
[30] T. P. Jung, S. Makeig, T. W. Lee, M. J. McKeown, G. Brown, A. J. Bell, and T. J. Sejnowski, "Independent component analysis of biomedical signals," in *Proc. Int. Workshop Independent Component Analysis and Blind Signal Separation*, 2000, pp. 633–644.
[31] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
[32] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
[33] M. Vetterli, "Wavelets, approximation, and compression," *IEEE Signal Process. Mag*, vol. 18, pp. 59–73, 2001.
[34] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
[35] R. W. Buccigrossi and E. P. Simoncelli, "Image compresion via joint statistical characterization in the wavelet domain," Univ. Pennsylvania, Philadelphia, Tech. Rep. 414, 1997.
[36] T. W. Lee, A. Ziehe, R. Orglmeister, and T. Sejnowski, "Combining time-delayed decorrelation and ICA: Toward solving the cocktail party problem," in *Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Processing*, vol. 2, 1998, pp. 1249–1252.
[37] S. C. Douglas, A. Cichocki, and S. Amari, "Multichannel blind separation and deconvolution of sources with arbitrary distributions," in *Proc. IEEE Workshop Neural Networks for Signal Processing*, 1997, pp. 436–445.
[38] S. C. Douglas and A. Cichocki, "Convergence analysis of local algorithms for blind decorrelation," in *Adv. Neural Inf. Process. Syst.*, 1996, pp. 2–7.
[39] S. Amari, T. P. Chen, and A. Cichocki, "Stability analysis of adaptive blind source separation," *Neural Netw.*, vol. 10, pp. 1345–1351, 1997.
[40] Y. Deville, "Analysis of the convergence properties of self-normalized source separation neural networks," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1272–1287, May 1999.

Maria G. Jafari (S'01–M'02) received the M.Eng. degree in electrical and electronic engineering from Imperial College London, U.K., in 1999, where the same year she began the Ph.D. degree study in blind source separation which she completed at King's College London in 2003.

Currently, she is a Research Assistant at Queen Mary University of London, U.K., Her research is mainly in audio source separation.

Jonathon Chambers (S'85–M'85–SM'98) was born in Peterborough, U.K., in 1960. He received the Ph.D. degree in 1990 after study at Peterhouse, Cambridge University, and Imperial College London, U.K.

In January 2004, he became a Cardiff Professorial Research Fellow within the Cardiff School of Engineering, Cardiff, U.K. His research contributions have been in adaptive and blind signal processing. He has authored/co-authored approaching 200 conference and journal publications.

Dr. Chambers has served as an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, and is currently serving as an Associate Editor for IEEE SIGNAL PROCESSING LETTERS.