Available online at www.sciencedirect.com

ScienceDirect

Signal Processing 87 (2007) 1819–1832

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Source separation using single channel ICA

M.E. Davies[a,*], C.J. James[b]

[a]IDCOM and the Joint Research Institute for Signal and Image Processing, The University of Edinburgh, Scotland, UK
[b]Signal Processing and Control Group, ISVR, University of Southampton, UK

Received 1 August 2006; received in revised form 2 January 2007; accepted 9 January 2007
Available online 25 January 2007

Abstract

Many researchers have recently used independent component analysis (ICA) to generate codebooks or features for a single channel of data. We examine the nature of these codebooks and identify when such features can be used to extract independent components from a stationary scalar time series. This question is motivated by empirical work that suggests that single channel ICA can sometimes be used to separate out important components from a time series. Here we show that as long as the sources are reasonably spectrally disjoint then we can identify and approximately separate out individual sources. However, the linear nature of the separation equations means that when the sources have substantially overlapping spectra both identification using standard ICA and linear separation are no longer possible.

Keywords: Independent component analysis; Blind source separation; Sparse coding; Single channel ICA

1. Introduction

Independent component analysis (ICA) was originally proposed for the blind separation of vector-valued observations into independent sources, e.g. [1]. However, it has also been noted that the infomax algorithm of Bell and Sejnowski [2] is equivalent to the model of Olshausen and Field [3] for learning sparse factorial codebooks. Since then there has been a steady interest in using ICA algorithms to learn 'features' or codebooks that provide a more efficient representation of a signal than a fixed (non-adapted) representation (e.g. [4–6]). However, in this context it is not clear what the features actually represent and to what extent they are truly independent. On the other hand, other empirical work [7–13] suggests that this representation can sometimes be used to separate out independent sources from a single time series. Here we will call this type of analysis "single channel ICA" (SCICA).

SCICA must be contrasted with the extensive research effort into underdetermined (overcomplete) ICA where solutions generally require strong additional assumptions such as sparsity and where separation is often only achieved through computationally intensive procedures (e.g. [14–17]). Since SCICA can be regarded as an extreme case of underdetermined ICA (one sensor!) it would be very interesting to know when and how applying a standard ICA solution can solve this problem.

*Corresponding author. Tel.: +44 01316505795; fax: +44 01316506554.
E-mail addresses: mike.davies@ed.ac.uk (M.E. Davies), C.James@soton.ac.uk (C.J. James).

0165-1684/$ - see front matter
doi:10.1016/j.sigpro.2007.01.011

The SCICA approach, as noted in [7,8], is an instance of multi-dimensional independent component analysis (MICA) [18] applied to vectors of delayed samples. This implies that multiple ICA components may be associated with a single independent source and that some form of post-processing is necessary to group the relevant ICA components together. In [7] grouping was done based on cross entropy estimates between components while in [8] synchronization of the analysis blocks with the primary source was used.

The main contribution of this paper is to set out a mathematical framework that explains, under certain assumptions, when and how standard ICA can perform source separation from a single sensor. We show that the special structure induced by mapping a time series problem into a multi-dimensional model means that the sources can only be successfully identified and separated using ICA when they have disjoint spectral support. We also show that dimensions of the subspaces is asymptotically proportional to the bandwidth of the individual sources. Finally, when the individual sources are composed of a filtered sequence of independent identically distributed (i.i.d.) random variables, we argue that the components can be grouped based upon spectral similarity.

The rest of the paper is set out as follows. We begin by reviewing the application of ICA to single channel data and show that this is equivalent to learning a bank of zero phase perfect reconstruction filters. In Section 3 we introduce a simple mathematical model for a mixture of independent sources and then show that SCICA can only separate out sources whose power spectra have disjoint support. The next section discusses two algorithmic solutions to this problem. We complete the paper with some examples, showing both cases where the technique succeeds and where it fails.

2. Single channel ICA

A simple but popular model of a signal, $\mathbf{x} \in \mathbb{R}^N$ is to represent it as a linear superposition of vectors $\mathbf{a}_i$:

$$\mathbf{x} = \sum_i s_i \mathbf{a}_i, \tag{1}$$

where s_i are the weights or coefficients. This can be written more succinctly in vector–matrix form: $\mathbf{x} = \mathbf{A}\mathbf{s}$, where $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_\mathbf{N}]$ and $\mathbf{s}$ is the vector of coefficients s_i. Usually the vectors $\mathbf{a}_\mathbf{i}$ are chosen to form a basis in the signal space so that an inverse (analysis) equation is uniquely defined: $\mathbf{s} = \mathbf{W}\mathbf{x}$, where $\mathbf{W} = \mathbf{A}^{-1}$. This has been an extremely popular framework for signal analysis over the years and encompasses the discrete Fourier transform (DFT), the discrete cosine transform (DCT), orthogonal and bi-orthogonal wavelets and the Karhunen–Loeve basis.

If the coefficients s_i are treated as *independent* random variables then we have a generative linear statistical model. Furthermore if we assume that A is square and invertible we have the classic ICA model.

We can apply such a representation to a scalar time series $x(t)$ by breaking it up into a sequence of contiguous blocks of length N and treat these as a sequence of vector observations:

$$\mathbf{x}(n) = [x(nN), \ldots, x(n(N-1)+1)]^\mathrm{T}, \tag{2}$$

where n is the block index. A standard ICA algorithm then applied to $\mathbf{x}(n)$ can learn a basis that is in some sense 'well matched' to the time series (see for example [19, Chapter 21]). This is the idea behind learning features using ICA and is closely linked with learning sparse factorial codes [3].

2.1. Separation and reconstruction

Since the data in SCICA come from a scalar time series, some thought is required as to how to perform separation. Furthermore we are often interested in seeing the separated signals back in the original observation domain. In standard ICA, for each source this is achieved through the application of an unmixing and mixing pair of operations:

$$\mathbf{x}_s^{(i)} = \mathbf{A}_{(:,i)} \mathbf{W}_{(i,:)} \mathbf{x}, \tag{3}$$

where $\mathbf{x}_s^{(i)}$ is the ith source in the observation domain. One nice property of mapping the sources back to the observation domain is that we have a perfect reconstruction decomposition:

$$\mathbf{x} = \sum_i \mathbf{x}_s^{(i)}. \tag{4}$$

Applying Eq. (3) to the nth block of the scalar time series, as defined in Eq. (2), gives

$$x_s^{(i)}(nN - k + 1) = A_{(k,i)} \sum_{j=1}^{N} W_{(i,j)} x(nN - j + 1), \tag{5}$$

where j and k index over the N samples.

Unfortunately the resulting estimate of the reconstructed sources will be highly dependent on the block alignment with the data. Although this will be partially overcome when we group dependent components together (see Section 3), to obtain complete shift-invariance we need to implement 'cycle-spinning' [20,21].

The solution of cycle-spinning was originally proposed as technique in wavelet analysis to reduce blocking artifacts [20]. Essentially a different estimate is made for each possible block alignment with the data (generating N different estimates due to the block periodicity). The 'cycle-spinning' estimate is then the average of all of these fixed basis estimates.

It is important to note that a cycle-spinning estimate is *always* superior to or at least as good as block processing when we have a stationary random process.[1] This is because cycle-spinning is an example of Rao–Blackwellization [22] which reduces the variance of the estimator while not effecting the bias.

In our case the cycle-spinning separation is given by

$$x_s^{(i)}(t) = \frac{1}{N}\sum_{k=1}^{N} A_{(k,i)} \sum_{j=1}^{N} W_{(i,j)} x(t-j+k) = \frac{1}{N} a_i(-t) * w_i(t) * x(t), \tag{6}$$

where $a_i(t)$ is an finite impulse response (FIR) filter associated with the column vector $A_{(:,i)}$ and $w_i(t)$ is an FIR filter associated with row vector $W_{(i,:)}$. In other words the estimate is a filter. Note the same arguments can be used to define a component time series, $s_i(t) = w_i(t) * x(t)$, however, here we will concentrate on the components mapped back to the observation domain (i.e. $x_s^{(i)}(t)$).

To simplify things further let us assume that the signal has already been pre-whitened. Note that in SCICA spatial pre-whitening is equivalent to temporal pre-whitening which can be achieved with a whitening filter. Whitened data imply that the separating matrix W is orthogonal and therefore $a_i(t) = w_i(t)$. That is: the filter is symmetric around $t = 0$ and hence has zero phase.

Thus to separate a component and then reconstruct it in the observation domain requires three steps: pre-whitening; separation using Eq. (6) and inverse whitening. However, since filters commute this is equivalent to just the single application of the separation filter. Thus while phase plays an important role in the identification of the independent components it plays no role in the separation.

In summary: using ICA with pre-whitening and cycle-spinning we can learn a set of zero phase, perfect reconstruction separation filters:

$$f_i(t) = \frac{1}{N} w_i(-t) * w_i(t). \tag{7}$$

[1] Implicitly, by having no preference of the position of the signal we are introducing a stationary model: $p(x_t) = p(x_{t-\tau}), \forall\tau$.

2.2. Independent or merely interesting?

What is not clear at this stage is when we can really expect the components $x_s^{(i)}(t)$ to be independent. Indeed it is not generally possible to linearly decompose a single time series into independent components and one might reasonably question whether the resulting coefficients are truly independent.

If we merely wish to learn 'interesting' features then this may not be of primary concern. In this case it is also possible to relax the notion of component-wise independence by, for example, grouping the components on a topographical map so that the dependence between components is represented by distance [24].

In contrast, if we are interested in separating out independent components present in the single time series [7–11] then the notion of independence is very important. However, since the learnt ICA bases generally contain shifted versions of the same 'feature' [19] we will find that MICA [18] provides us with the appropriate model.

MICA assumes that an observed data vector $\mathbf{x} \in \mathbb{R}^N$ can be decomposed into the sum of C independent vectors:

$$\mathbf{x} = \sum_{p=1}^{C} \mathbf{x}_p, \tag{8}$$

where each $\mathbf{x}_p \in \mathbb{R}^N$ is the pth multi-dimensional component that is assumed to lie in an n_p-dimensional subspace E_p and the set $E_1, \ldots, E_C$ are assumed to be linearly independent. For clarity we recall the definition of MICA given in [18].

Definition 1. The canonical MICA decomposition (if it exists) of a vector $\mathbf{x}$ is the unique decomposition of $\mathbf{x}$ into independent multi-dimensional components $\mathbf{x} = \sum_{p=1}^{C} \mathbf{x_p}$ such that:

(1) there is at most one Gaussian component and
(2) no non-Gaussian component can be further decomposed into other independent components (maximal decomposition)

Cardoso [18] also noted that a standard ICA algorithm can be used to learn the MICA decomposition by first estimating one-dimensional components. These components can then be grouped based on dependency and the ICA mixing matrix can be organized into a set of C submatrices $A = [\tilde{A}_1, \ldots, \tilde{A}_C]$, where the columns of $\tilde{A}_p$ span the subspace E_p.

We next need to determine under what conditions SCICA can be treated as a legitimate MICA problem and how best to identify which components (and how many) are associated with each source. We address these problems in the next two sections.

3. Independent mixtures of random processes

Suppose the observed signal, $x(t)$, admits a decomposition into the sum of mutually independent random processes, $x_p(t)$:

$$x(t) = \sum_p x_p(t). \tag{9}$$

Note that this is an additive rather than multiplicative model similar to MICA [18].

It is instructive to first consider an interesting subset of this model that fits closely to the ICA/sparse coding formalism. Suppose that each stochastic process $x_p(t)$ is a filtered i.i.d. random process:

$$x_p(t) = h_p(t) * s_p(t), \tag{10}$$

where $h_p(t)$ is an FIR filter of length M. While this model is more restrictive than that in Eq. (9) it is essentially the model that is implicitly assumed when applying ICA to a time delay representation.

As before, let $\mathbf{x}(t) = [x(t), x(t-1), \ldots, x(t-N+1)]^{\mathrm{T}}$ represent our N-dimensional delay vector. Then we can write Eq. (9) in a matrix form as

$$\mathbf{x}(t) = \sum_p \mathbf{H}_p \mathbf{s}_p(t), \tag{11}$$

where $\mathbf{H}_p$ is the following Toeplitz matrix associated with the filter $\mathbf{h}_p = [h_p(0), \ldots, h_p(M-1)]^{\mathrm{T}}$:

$$\mathbf{H}_p = \begin{pmatrix} \mathbf{h}_p & & & & \\ & \mathbf{h}_p & & & \\ & & \mathbf{h}_p & & \\ & & & \ddots & \\ & & & & \mathbf{h}_p \end{pmatrix} \tag{12}$$

and $\mathbf{s}_p(t) = [s_p(t), s_p(t-1), \ldots, s_p(t-N-M+2)]^{\mathrm{T}}$.

For a single source ($C = 1$), solving for $s_1(t)$ is precisely the blind deconvolution problem [25]. The relationship between blind source separation and blind deconvolution is discussed in detail in [26]. Indeed it can be shown that the single-unit FastICA, when applied to a delay vector, is equivalent to the Shalvi–Weinstein blind deconvolution algorithm [19]. That is it learns a single separating vector that is equivalent to the Shalvi–Weinstein deconvolution filter. Applying full ICA (i.e. learning the full unmixing matrix) to the blind deconvolution problem learns an unmixing matrix W whose rows consist of N approximate shifted versions of the deconvolution filter [19].

At this point it does not look promising that we will be able to obtain anything useful in the case of multiple sources without considering overcomplete representations. However, the blind deconvolution problem is only well defined when the convolution operator $\mathbf{h}_1$ is invertible. In that case, in the limit $N \to \infty$, the ICA blind deconvolution matrix $\mathbf{H}_1$ has full rank and the shifted versions of $\mathbf{h}_1$ span the full space. However, when $\mathbf{h}_p$ is not invertible the matrix $\mathbf{H}_p$ may be rank deficient. Although this immediately precludes the possibility of extracting $s_p(t)$, it does not necessarily preclude the estimation of $x_p(t)$.

Suppose now that we have multiple components and that each $\mathbf{h}_p$ is non-invertible and each corresponding matrix $\mathbf{H}_p$ is rank deficient: i.e. its columns only span a subspace E_p in $\mathbb{R}^N$. In this case the vectors $\mathbf{x}_p(t) = \mathbf{H}_p \mathbf{s}_p(t)$ lie in this subspace. Furthermore, if all the subspaces, $\{E_p\}$, are linearly independent then we have a valid MICA model.

Any such process should then be decomposable by ICA into basis vectors $\mathbf{a}_i$ that can be grouped into C subsets γ_p, such that if $i \in \gamma_p$ then $\mathbf{a}_i \subset E_p$ [18]. Furthermore since these basis vectors tend to be shifted approximations of the individual filter $\mathbf{h}_p$ all the basis vectors associated with the subset γ_p will have very similar spectral support. This will prove very useful when grouping the components (see Section 4).

Finally, to estimate the individual independent processes we must sum over the contributions from the separate ICA components. Recall from Section 2, each component contribution can be estimated using the filter defined in Eq. (6). We therefore have the following estimate for $x_p(t)$:

$$x_p(t) = \sum_{i \in \gamma_p} f_i(t) * x(t). \tag{13}$$

3.1. SSA, MICA and SCICA

More generally we can consider the case of when the stochastic processes $x_p(t)$ are singular, as in singular systems analysis (SSA) [27], but not necessarily satisfying the $x_p(t) = h_p(t) * s_p(t)$ restriction. SSA assumes that a process lies approximately on a low dimension subspace E_p within the N-dimensional delay vector space. For such a delay vector $\mathbf{x}_p$ an orthogonal projector, $\mathbf{\Pi}_p$, onto this subspace can be defined by the significant eigenvectors of the N-dimensional covariance matrix $\mathbf{C}_p$. Again, if the subspaces $E_1, \ldots, E_C$ are linearly independent then we have a valid MICA problem.

Note that unlike general MICA the observation dimension is not predefined. Indeed practical issues aside, the dimension merely constrains the length of the potential separating filters. We can therefore consider the case where $N \to \infty$. Since, for our problem, $\mathbf{C}_p$ has Toeplitz structure we know that the distribution of eigenvalues tends to the power spectrum of $x_p(t)$ [28] and the dimension of the component subspace E_p is proportional the size of the frequency support (bandwidth) of the stochastic process. Finally, for the subspaces, $\{E_p\}$, to be linearly independent, the components must have disjoint spectral support (also a necessary condition for separating sources using time-invariant linear filters).

We can now give a completely general definition of SCICA:

Definition 2. The canonical SCICA decomposition (if it exists) of a stationary random process $x(t)$ is the unique decomposition into independent stationary random processes $x(t) = \sum_{p=1}^{C} x_p(t)$ such that:

(1) there is at most one Gaussian process;
(2) all the independent random processes must be bandlimited with disjoint spectral support; and
(3) no non-Gaussian process can be further decomposed into multiple independent, spectrally disjoint processes (maximal decomposition).

As above, the general SCICA process is decomposable by ICA (at least in the limit $N \to \infty$) into basis vectors $\mathbf{a}_i$ that when grouped into subsets γ_p span the independent subspaces E_p. However, in general, we cannot guarantee that the basis vectors, $\mathbf{a}_i$, will be shifted approximations of some generating vector, or even that they will have common spectral support. In this case grouping based on additional (higher level) dependencies may be required, e.g. as in [7].

In the remainder of this paper we will assume that this problem does not occur.

4. Solving SCICA with ICA

We now consider practical methods to solve the SCICA problem. While any ICA algorithm can be used to estimate the one-dimensional ICA components large computational savings can be had by using an algorithm that makes use of pre-whitening, such as FastICA [19]. This is because spatial whitening of the delay vector is equivalent to temporal whitening and can be efficiently implemented using the DFT or the DCT. Dimension reduction, i.e. principal component analysis (PCA), can also be included by projecting to zero the values of the frequency bins with low energy content.

To cluster the independent components we will utilize the fact that the basis functions spanning the subspace E_p are shifted approximations for the generating filter $\mathbf{h}_p$. Depending on the bandwidth of $\mathbf{h}_p$ there will be a different number of shifted versions of the same filter. Furthermore following the SCICA model the other components should have disjoint spectral support. We therefore propose to group the components based on the separating filter's transfer function.

This can be done by first calculating the magnitude transfer functions for the learnt basis: $|A_i(\omega)|^2$. The individual components can then be clustered into groups γ_p based on the similarity of $|A_i(\omega)|^2$ by, for example, using a standard k-means clustering algorithm. Re-organizing the separating matrix into the groups of submatrices $A = [\tilde{A}_1, \ldots, \tilde{A}_C]$ then generates the MICA decomposition.

Finally the separating/reconstruction filter for source p is defined as

$$\tilde{f}_p(t) = \frac{1}{N} \sum_{i \in \gamma_p} a_i(-t) * w_i(t). \tag{14}$$

We can summarize the SCICA algorithm steps as follows:

(1) Temporally whiten the signal (with possible dim. reduction) using PCA, DFT or DCT.
(2) Apply an ICA algorithm to learn the mixing matrix A.

(3) Calculate the magnitude transfer functions (TFs) of the basis vectors $a_i(t)$ and cluster into groups γ_p, $p = 1, \ldots, C$ using k-means (or another clustering algorithm).
(4) Calculate the shift-invariant separation and reconstruction filters, $\tilde{f}_p(t)$, for each source.

4.1. A quick and dirty solution

We now describe a quick and dirty solution that can be substantially faster than the ICA method described above and avoids explicit clustering. It is based upon a deflationary scheme and is particularly attractive when there are only a small number of independent processes to be extracted. This is the case, for example, when using the constrained ICA model [11].

The idea is to modify the previous algorithm to remove the clustering step. Suppose that we begin with a single-unit deflationary ICA algorithm (such as FastICA [19]). The first extracted component should be an approximate shifted version of the generating filter for one of the sources. Note that if each of the independent components that span E_p are shifted versions of the generating filter then the individual separation and reconstruction filters:

$$f_i(t) = a_i(-t) * w_i(t), \quad \text{for any } i \in \gamma_p \tag{15}$$

will be identical. The shift is cancelled out by mapping back to the observation domain. Thus much of the work done in Eq. (14) is largely redundant.

We therefore propose replacing equation (14), which requires prior clustering of the components, with a single component estimate:

$$\tilde{f}_p(t) \approx \alpha_p f_i(t) = \frac{\alpha_p}{N} a_i(-t) * w_i(t) \quad \text{for any } i \in \gamma_p. \tag{16}$$

The scaling parameter α_p is required since we do not know the size of γ_p (i.e. the dimension of the independent subspace E_p). However, as this is a simple rescaling, we can estimate α in closed form by minimizing the norm of the residual $r(t) = x(t) - \alpha_p f_i(t) * x(t)$ giving

$$\alpha_p = \frac{\langle x(t), f_i(t) * x(t)\rangle}{\langle x(t), x(t)\rangle}, \tag{17}$$

where $\langle \cdot, \cdot \rangle$ is the usual vector inner product.

We can therefore apply a single-unit ICA algorithm to $x(t)$ and extract the component with the separating filter in Eq. (16). The remaining signal, $r(t)$ should then only contain the other sources and we can repeat the process by successive application of the single-unit ICA algorithm to the residual $r(t)$.

This algorithm can be summarized as:

(1) temporally whiten signal (with possible dimension reduction);
(2) extract one component (deflationary ICA) and form separation filter $\tilde{f}_p(t) = (\alpha_p/N)w_p(-t) * w_p(t)$;
(3) calculate the residual $r_p(t) = r_{p-1}(t) - \tilde{f}_p(t) * r_{p-1}(t)$, where we initialize $r_0(t) = x(t)$;
(4) to extract further sources go back to step 2.

Note that since we only use one basis vector $\mathbf{a}_i \in E_p$ to estimate the filter $\tilde{f}_p(t)$ there is no clustering step. Furthermore we only have to determine C components (as opposed to N previously) in a deflationary manner. Thus if, as is often the case, $C \ll N$ we have considerable computational savings.

We have, however, found that this approach tends to give inferior results to the full ICA solution. This is presumably because the filters $f_i(t), i \in \gamma_p$ are not completely identical and as such each individual filter only provides a crude approximation of the generating filter. This can also result in some of the source being left in the residual (see the example in the next section).

5. Experiments

We now demonstrate the results of SCICA on three examples. The first data set is a toy example, specifically designed to show that approximately spectrally distinct independent processes can be blindly extracted from a single time series. The second and third examples deal with real data showing both a successful application and a failing SC ICA method. For brevity we only examine the first algorithm in Section 4.

5.1. Toy example

We begin by presenting a toy example with a single channel observation of two highly sparse i.i.d. sources each passed through separate bandlimited filters with approximately disjoint support. The filters and their transfer functions are shown in Fig. 1. A third source of a single sinusoid was also added to the mixture.

The power spectrum of the mixture of sources is shown in Fig. 2. Note that just from the power spectrum one might be fooled into believing that there were four distinct sources present.

Applying FastICA to the mixture of the sources (delay dimension = 100, reduced to 36 using PCA) resulted in the basis functions (columns of A) shown in Fig. 3. Fig. 3 (b) also shows the magnitude of the transfer functions for each column of A when interpreted as a filter. It is clear that a number of the basis functions are various shifted and distorted approximations of one of the two generating filters in Fig. 1. The transfer functions were then grouped into three clusters using k-means. The cluster labels are also included in the figure. Grouping into more than three clusters resulted in separation filters with significant spectral overlap and so could be discounted.

We therefore obtained three sets of filters that approximately spanned the same frequency ranges and three independent subspaces with dimensions 20, 14 and 2, respectively. The dimension is a

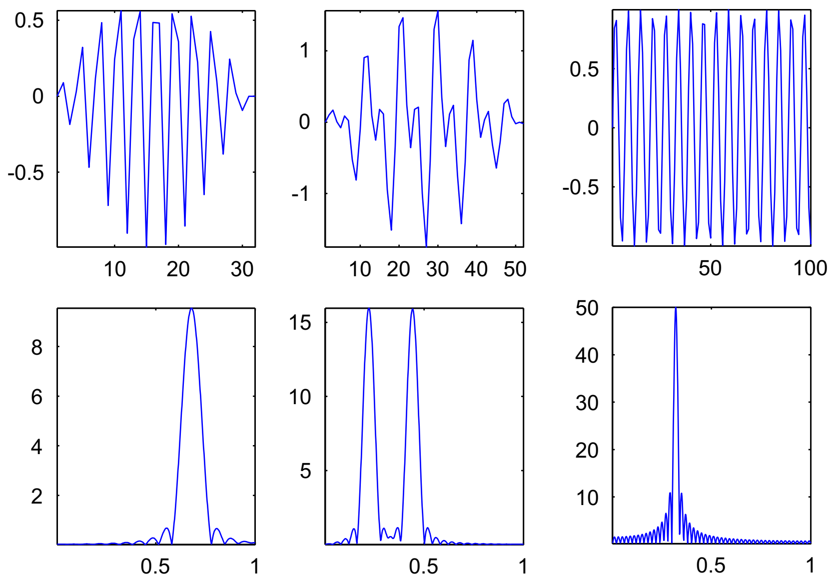

Fig. 1. The impulse responses and transfer functions of the two mixing filters (left and center) and a section from the single sinusoid (right).

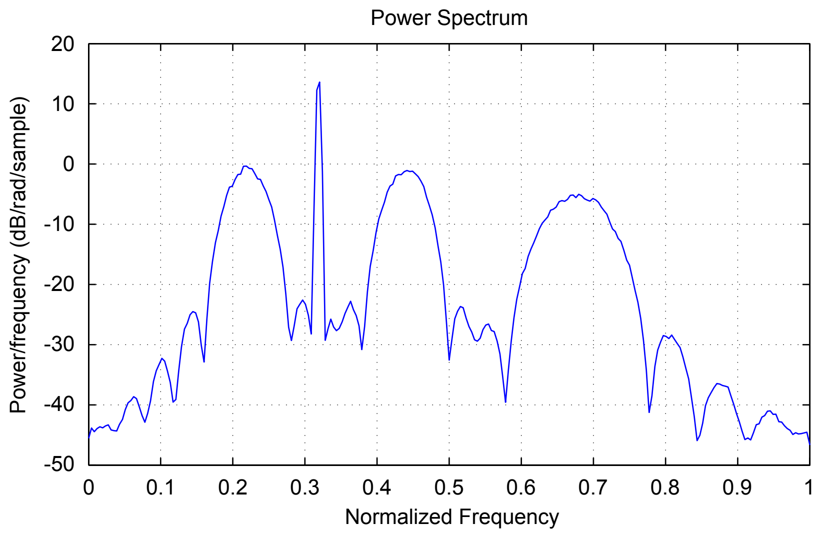

Fig. 2. The power spectrum of the mixture of sources.

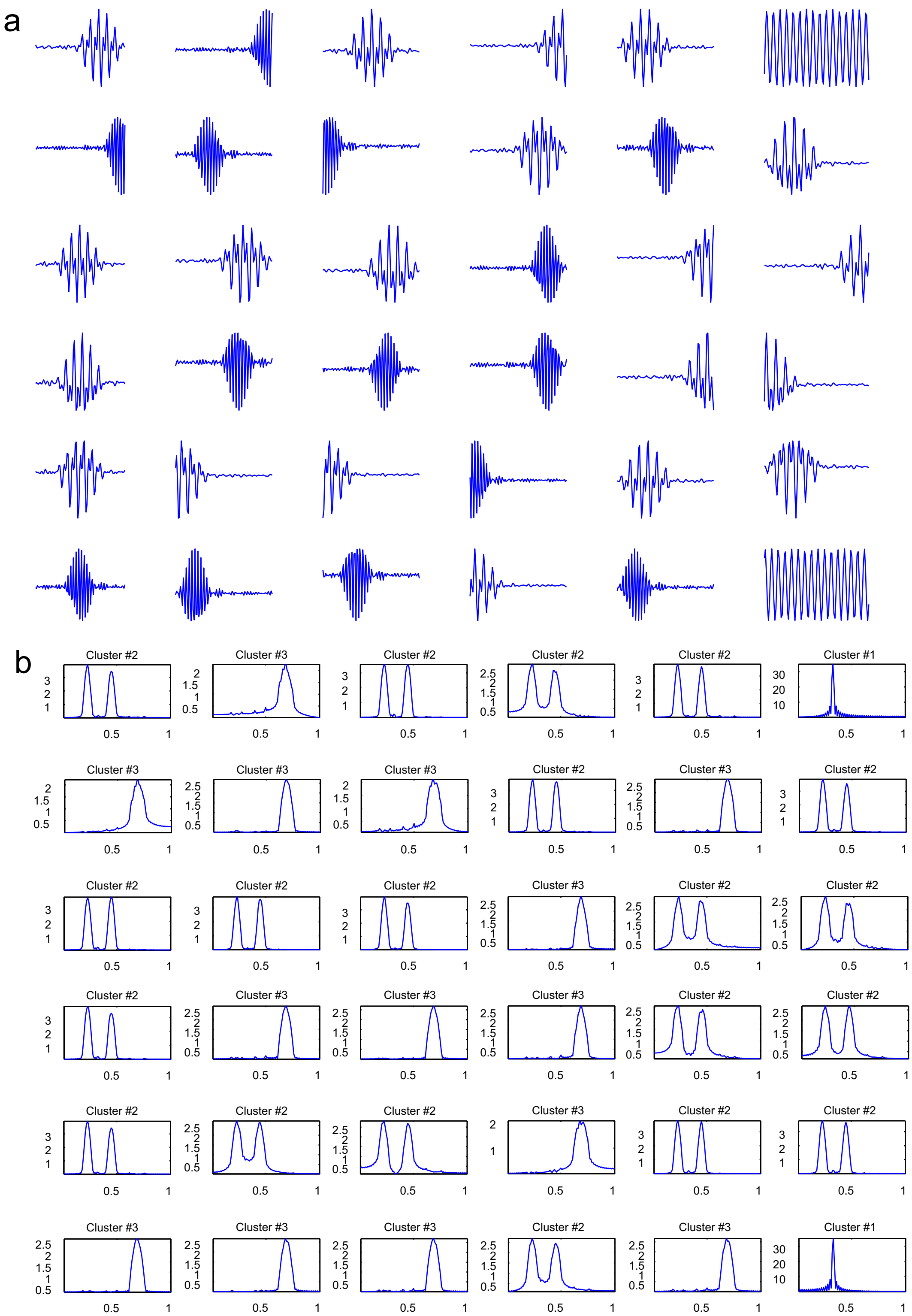

Fig. 3. (a) Basis functions from the columns of the mixing matrix, A, learnt using FastICA. (b) The magnitude frequency responses for the same functions.

function of the bandwidth of the separated component, as predicted from the theory.

To extract the sources we construct the filters, given by Eq. (14), and filter the original time series. The filters are plotted in Fig. 4, along with their magnitude frequency responses. It can be seen that they appear as symmetric adaptively selective bandpass filters. The signal-to-interference ratios (SIR) of the extracted sources are given in Table 1 and separation has clearly been achieved.

We repeated the experiment using the deflationary algorithm given in Section 4.1. The first independent component found was associated with source 2. Once this source has been filtered out, using Eq. (16), the single-unit ICA algorithm was applied again and found a component associated with the source 1. Filtering out source 1 therefore leaves an estimate for source 3. The SIRs are also included in Table 1. While separation has been achieved the SIR values are not as high as for the previous algorithm. On the other hand, the process only required two passes of the single-unit ICA algorithm and there was not need for clustering.

Finally, to highlight a potential weakness of the greedy deflationary algorithm we also applied it a third time. However, rather than extracting a component associated with source 3 it found a component associated with the reminents of source 2 that had not been fully extracted at the first attempt.

5.2. Experiment 2: epileptic EEG data

In our second experiment we applied SCICA to a single channel from a multi-channel EEG recording of the onset of an epileptic seizure. The data consisted of 25 sensors spaced evenly around the head and digitally recorded at a rate of 200 samples/second. As this is the real data there is no ground truth. Instead we make a comparison between what can be extracted using SCICA and using traditional multi-channel (spatial) ICA. Note that these methods are in fact complementary since multi-channel performs blind *spatial* filtering while SCICA performs blind *temporal* filtering.

We first analyzed the full complement of 25 sensors spatially using the FastICA algorithm [19]. The independent components were then hand labelled based upon the component waveform morphology and the localization of the component on the head (obtainable from the estimated mixing matrix). From this it was deduced that ICs #17 and #19 predominantly captured the rhythmic seizure activity. These are shown in Fig. 5. The seizure

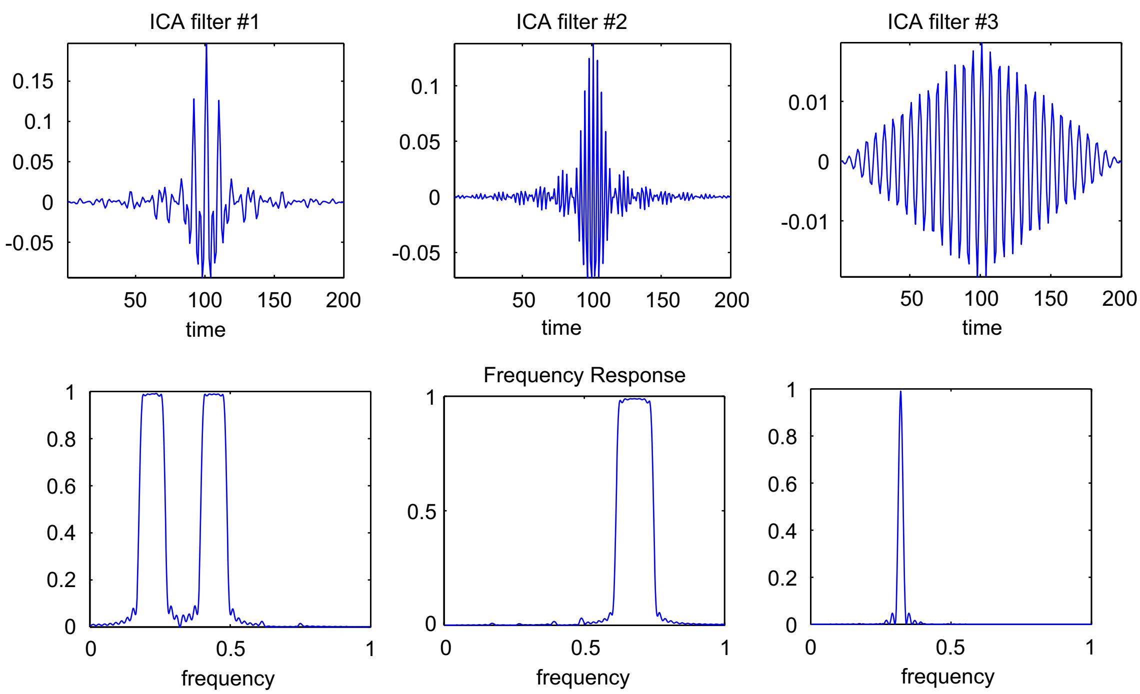

Fig. 4. The zero-phase reconstruction filters for the three different sources.

onset commences at about a third of the way through. Note that even after ICA the data are still extremely noisy.

Amongst the other ICs identified were ocular artifacts (#5 and #10) and muscle/electrode artifacts (#6, #8 and #18).

We then analyzed channel #20 with the aim of extracting the rhythmic seizure activity component.

Table 1
The SIRs for the original (mixed) and separated sources

SIR (dB)	s_1	s_2	s_3
Original	−9.5	−2.5	3.16
SCICA (alg. I)	17.3	21.5	27.8
SCICA (alg. II)	10.6	10.4	12.4

Channel #20 was particularly interesting since, while it contained seizure activity, this was masked by other components (particularly the ocular artifact, IC #5).

We used a delay vector of dimension 100, reduced to 28 using PCA. We then clustered the components into three groups (as in the toy example, above, clustering into larger groups resulted in significantly overlapping filter transfer functions and could be discounted). The filter transfer functions are shown in Fig. 6. While we expect that there are many different sources present there appear to be three spectrally distinct components.

Finally, we can compare the separated components with the spatially estimated seizure components mapped onto channel #20. These results are shown in Fig. 7. The first filter (source 1) has

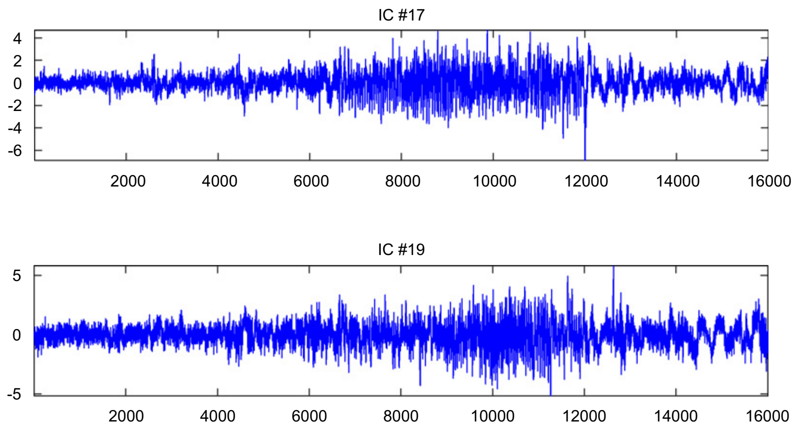

Fig. 5. Independent components #17 and #19 which were determined to contain the majority of rhythmic seizure activity.

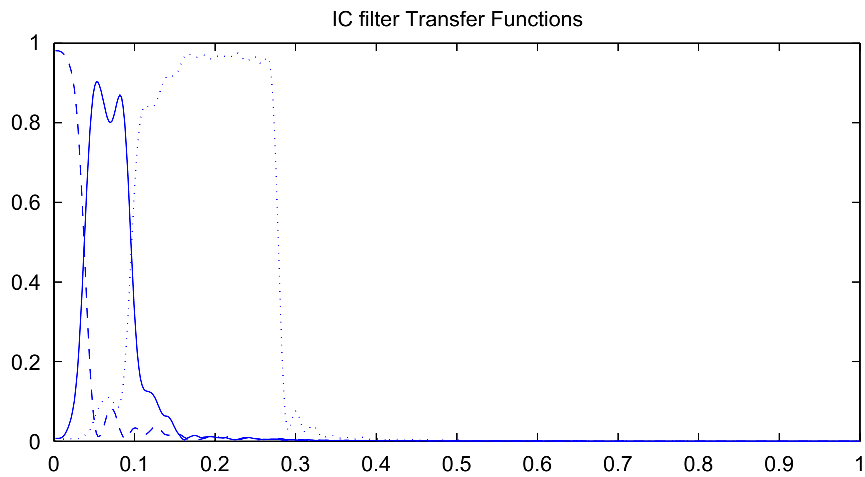

Fig. 6. Transfer functions for the three learnt separating filters. Note they are reasonably disjoint in frequency.

removed a number of spikes and some high frequency noise. The second filter (bandpass between around 5 and 9 Hz) has successfully extracted the seizure component. Furthermore the SCICA component appears more distinct than that learnt by the multi-channel ICA (second row). This is because the multi-channel ICA is not able to perform temporal filtering. Finally, the third filter extracted a low-frequency component. Interestingly this component is quite well correlated to the ocular artifact that was found in the multi-channel spatial ICA at IC #5.

5.3. Experiment 3: a combined maternal/fetal ECG signal

Our final experiment is another a real world problem that has been extensively examined in the ICA community. In prenatal diagnostics the fetal ECG signal is of great value. However, in a non-invasive setting it is impossible to directly observe the fetal heart beat without contamination from the maternal ECG. Like the EEG analysis above, this problem fits the classical ICA assumptions very well [29]. However, we will see here that it does not fit the

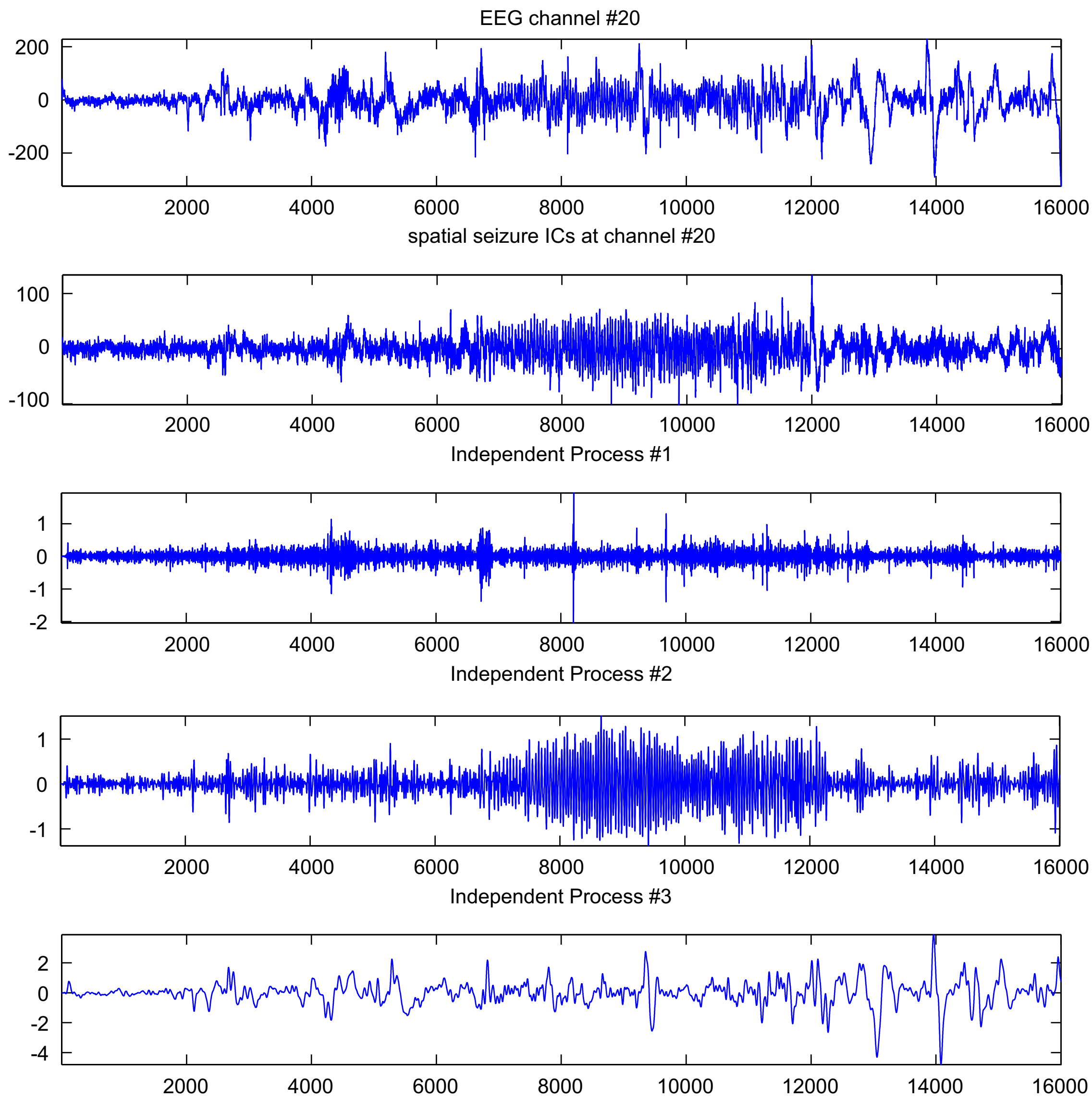

Fig. 7. Plots of the EEG channel #20 (top); the spatial IC seizure components at channel #20; and the three sources separated using SCICA (third, fourth and fifth).

SCICA model. This is because both the maternal and fetal QRST waveforms have very similar broadband spectral profiles. Thus we should expect the proposed methods to fail.

Fig. 8 shows a single channel of mixed maternal/fetal ECG obtained from [30]. Both components are clearly visible and can be to a certain extent 'separated by eye'. Fig. 9 shows the first 7 ICA bases learnt using FastICA (delay vector dimension $=100$ reduced to 49 with PCA). All the basis functions appeared to be approximately the same function shifted by varying degrees.

When trying to learn two clusters of the independent components all the basis functions were clustered together apart from extreme shifts (such as the seventh function in Fig. 9). This did not separate the maternal and fetal ECG signals as desired. Instead the first source contained both ECG signals while the second source identified a low-frequency drift component: see Fig. 10.

In contrast to this failing, Castells et al. [8] have successfully separated out arterial fibrillation (AF) from the ventricular activity (VA) using a single ECG signal. Here the success can be attributed to the synchronization of the analysis blocks to the primary source VA and the localized frequency content of the AF activity.

It is also interesting to note that it is in fact possible to separate the maternal and fetal components, but not using SCICA. In [31,32] fetal and maternal signals were successfully separated from a single channel using a generative sparse coding model and a computationally intensive Monte Carlo inference scheme.

6. Discussion

In this paper we have formalized the concept of SCICA and have shown that it is an interesting special case of Cardoso's MICA. This demonstrates that under certain circumstances single channel blind source separation can be practically addressed using the SCICA framework, but a necessary condition is that the independent source processes have disjoint spectral support. When this is not the case (as in the ECG example) stronger prior information is typically necessary and separation becomes much harder.

We also noted in experiment 2 that the SCICA framework is actually complimentary to the spatial ICA. It is therefore possible to combine these ideas in the form of a space–time ICA codebook. Such ICA codebooks have recently been successfully applied to separate audio sources from stereo microphone recordings [33]. It would be interesting to determine the general conditions required for such space–time codebooks to be able to achieve separation.

In future it would be interesting to explore more general solutions to the SCICA problem, for example when the sources follow a general SSA model. Also, although the ICA codebook is related

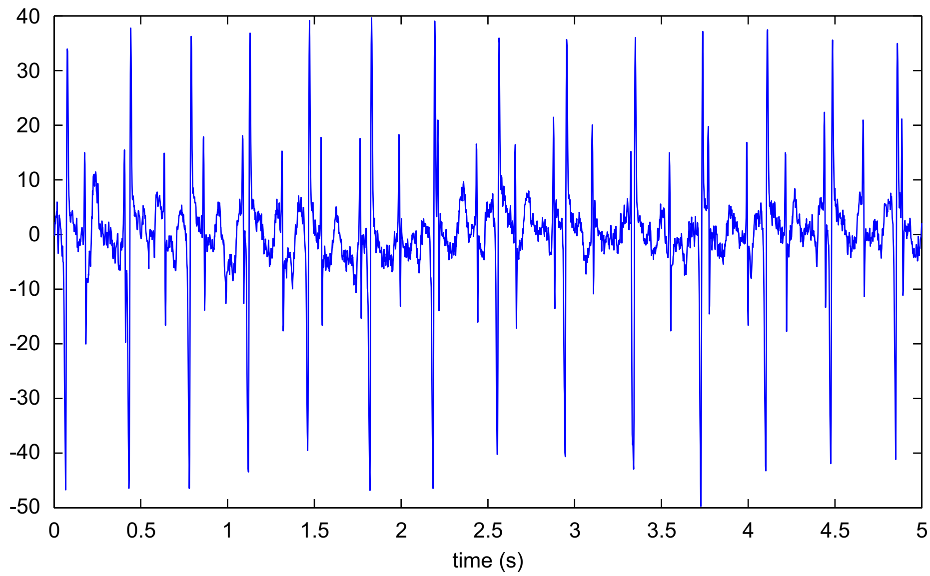

Fig. 8. The single channel ECG signal. Both maternal and fetal components are visible.

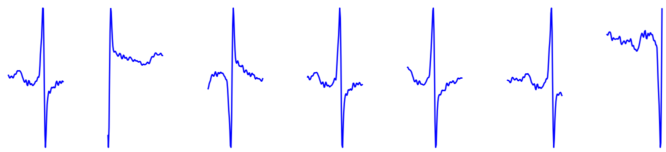

Fig. 9. The first seven basis functions from the columns of the mixing matrix, A, learnt using FastICA for delay vectors of the ECG signal. All the bases in A appear to be the same underlying function shifted by different amounts.

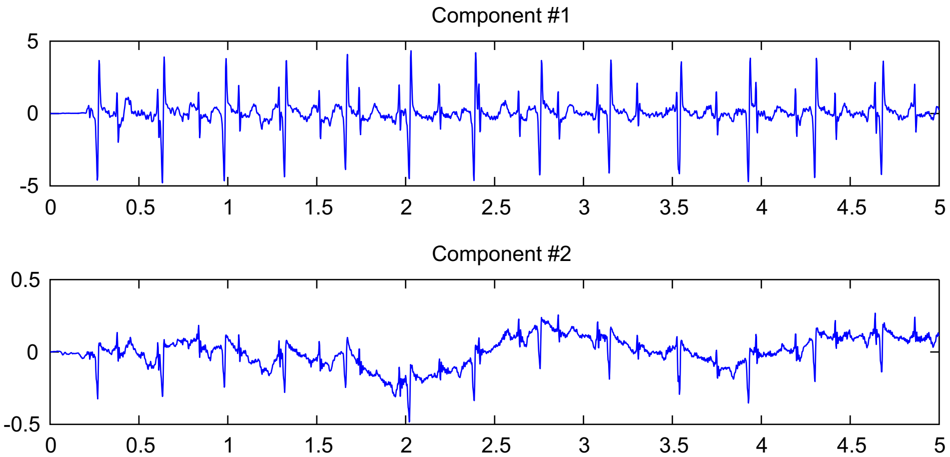

Fig. 10. The separated and reconstructed components of the ECG signal. Note the fetal and maternal components have not been separated.

to sparse factorial codes [3], the ICA model relies on a linear separation procedure, which distinguishes it from overcomplete ICA. However, it might be possible to adapt the fast SCICA approach by incorporating a nonlinear estimation step (such as basis pursuit or matching pursuit) into the estimation step.

Acknowledgments

Both authors would like to acknowledge the EPSRC funded UK ICA Research Network (http://icarn.org). M.E.D. also acknowledges support for his position from the Scottish Funding Council and for their support of the Joint Research Institute with the Heriot-Watt University as part of the Edinburgh Research Partnership.

References

[1] P. Comon, Independent component analysis—a new concept? Signal Process. 36 (1994) 287–314.

[2] A.J. Bell, T.J. Sejnowski, An information maximization approach to blind separation and blind deconvolution, Neural Comput. 7 (6) (1995) 1129–1159.

[3] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive-filed properties by learning a sparse code of natural images, Nature 381 (1996) 607–609.

[4] S. Abdallah, M.D. Plumbley, If edges are the independent components of natural images, what are the independent components of natural sounds? in: Proceedings of International Conference on ICA 2001, 2001, pp. 534–539.

[5] D.-C. Balcan, J. Rosca, Independent component analysis for speech enhancement with missing TF content, in: Proceedings of the Sixth International Conference on ICA 2006, Lecture Notes in Computer Science, vol. 3889, 2006, pp. 552–560.

[6] A.K. Barros, N. Ohnishi, Single channel speech enhancement by efficient coding, Signal Process. 85 (2005) 1805–1812.

[7] M. Casey, A. Westner, Separation of mixed audio sources by independent subspace analysis, in: Proceedings of International Computer Music Conference, Berlin, 2000.

[8] F. Castells, C. Mora, J. Millet, J.J. Rieta, C. Sanchez, Juan Sanchis, Multidimensional ICA for the separation of atrial and ventricular activities from single lead ECGs in paroxysmal atrial fibrillation epsiodes, in: Proceedings of International Conference on ICA 2004, Lecture Notes in Computer Science, vol. 3195, 2004, pp. 1229–1234.

[9] C.J. James, D. Lowe, Extracting information from multichannel versus single channel EEG data in epilepsy analysis, in: A.K. Hyder, E. Shahbazian, E. Waltz (Eds.), Multisensor Fusion, Kluwer Academic Publishers, Dordrecht, 2002, pp. 889–895.

[10] C.J. James, D. Lowe, Extracting multisource brain activity from a single electromagnetic channel, Artif. Intell. Med. 28 (1) (2003) 89–104.
[11] C.J. James, O. Gibson, M.E. Davies, On the analysis of single versus multiple channels of electromagnetic brain signals, Artif. Intell. Med. 37 (2) (2006) 131–143.
[12] E.S. Warner, I.K. Proudler, Single-channel blind signal separation of filtered MPSK signals, IEE Proc. Radar Sonar Navig. 150 (6) (2003) 396–402.
[13] W.L. Woon, D. Lowe, Can we learn anything from single-channel unaveraged MEG data?, Neural Comput. Appl. 13 (2004) 360–368.
[14] M.E. Davies, N. Mitianoudis, A simple mixture model for sparse overcomplete ICA, IEE Proc. VISP 151 (1) (2004) 35–43.
[15] M. Girolami, A variational method for learning sparse and overcomplete representations, Neural Comput. 13 (11) (2002) 2517–2532.
[16] M.S. Lewicki, T.J. Sejnowski, Learning overcomplete representations, Neural Comput. 12 (2000) 337–365.
[17] E. Moulines, J.-F. Cardoso, E. Gassiat, Maximum likelihood for blind signal separation and deconvolution of noisy signals using mixture models, in: Proceedings of ICASSP-97, vol. 5, 1997, pp. 3617–3620.
[18] J.-F. Cardoso, Multidimensional independent component analysis, in: Proceedings of ICASSP'98, Seattle, WA, 1998, pp. 1941–1944.
[19] A. Hyvarinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, New York, 2001.
[20] R.R. Coifman, D.L. Donoho, Translation-invariant denoising, in: A. Antoniadis, G. Oppenheim (Eds.), Wavelets and Statistics, Springer Lecture Notes in Statistics, vol. 103, Springer, New York, 1995, pp. 126–150.
[21] A. Hyvärinen, P. Hoyer, E. Oja, Image Denoising by Sparse Code Shrinkage, in: S. Haykin, B. Kosko (Eds.), Intelligent Signal Processing, IEEE Press, New York, 2001.
[22] S.M. Kay, Fundamentals of Statistical Signal Processing, Vol. 1: Estimation Theory. Signal Processing Series, Prentice-Hall, Englewood cliffs, NJ, 1993.
[24] A. Hyvärinen, P.O. Hoyer, M. Inki, Topographic independent component analysis, Neural Comput. 13 (7) (2001) 1527–1558.
[25] O. Shalvi, E. Weinstein, New criteria for blind deconvolution of nonminimum phase systems (channels), IEEE Trans. Inform Theory 36 (2) (1990) 312–320.
[26] S.C. Douglas, S. Haykin, Relationships between blind deconvolution and blind source separation, in: S. Haykin (Ed.), Unsupervised Adaptive Filtering, vol. 2, Wiley, New York, 2002, pp. 113–145.
[27] D. Broomhead, G. King, Extracting qualitative dynamics from experimental data, Physica D 20 (1986) 217–236.
[28] R.M. Gray, On the asymptotic eigenvalue distribution of Toeplitz matrices, IEEE Trans. Inform Theory IT-18 (6) (1972) 725–730.
[29] V. Zarzoso, A.K. Nandi, Noninvasive fetal electrocardiogram extraction: Blind separation versus adaptive noise cancellation, IEEE Trans. Biomed. Eng. 32 (3) (2001) 230–236.
[30] DaISy: Database for the Identification of Systems. ESAT/SISTA, K.U. Leuven, Belgium. Available at: ⟨http://www.esat.kuleuven.ac.be/sista/daisy⟩.
[31] T. Blumensath, M.E. Davies, Shift-invariant sparse coding for single channel blind source separation, in: Proceedings of SPARS'05, 2005, pp. 75–78.
[32] T. Blumensath, M.E. Davies, Blind separation of maternal and fetal ECGs using any number of channels, Preprint, 2006.
[33] M.G. Jafari, S. Abdallah, M.D. Plumbley, M.E. Davies, Sparse coding for convolutive blind audio source separation, in: Proceedings of the Sixth International Conference on ICA 2006, Lecture Notes in Computer Science, vol. 3889, 2006, pp. 132–139.