



ELSEVIER

Neurocomputing 22 (1998) 173–186

---

---

NEUROCOMPUTING

---

---

# Removing artifacts from electrocardiographic signals using independent components analysis

Allan Kardec Barros\*, Ali Mansour, Noboru Ohnishi

*RIKEN BMC Research Center, 2271-130, Ana Gahora, Shimoshi-dami, Aichi 463-0003, Japan*

Accepted 3 July 1998

---

## Abstract

In this work, we deal with the elimination of artifacts (electrodes, muscle, respiration, etc.) from the electrocardiographic (ECG) signal. We use a new tool called independent component analysis (ICA) that blindly separates mixed statistically independent signals. ICA can separate the signal from the interference, even if both overlap in frequency. In order to estimate the mixing parameters in real time, we propose a self-adaptive step-size, derived from the study of the averaged behavior of those parameters, and a two-layers neural network. Simulations were carried out to show the performance of the algorithm using a standard ECG database. © 1998 Elsevier Science B.V. All rights reserved.

*Keywords:* Independent component analysis; Blind separation; Adaptive filtering; Cardiac artifacts; ECG analysis

---

## 1. Introduction

Many attempts were carried out to eliminate corrupting artifacts from the actual cardiac one when measuring the electrocardiographic (ECG) signal.

Cardiac signals show the well-known repeating and almost periodic pattern. This characteristic of physiological signals was already explored in some works (e.g. [5,17,22]) by synchronizing the parameters of the filter with the period of the

---

\* Corresponding author. E-mail: allan@bmc.niken.go.jp

signal. However, those filters fail to remove the interference when it has the same frequency of the cardiac signal.

On the other hand, many works were carried out in the field of blind source separation (BSS), using a new tool called *independent component analysis* (ICA). This large number of works may be because the ICA algorithms are in general elegant, simple and may deal with problems when second-order statistics (SOS) methods<sup>1</sup> in general do not work. This is because SOS algorithms usually search for a solution that *decorrelates* the input signals whereas ICA looks for statistically *independent* signals.

ICA is based on the following principle. Assuming that the original (or source) signals have been mixed linearly, and that these mixed signals are available, ICA finds in a *blind* manner a linear combination of the mixed signals which recovers the original source signals, possibly re-scaled. This may be carried out by using the principle of entropy maximization of non-linearly transformed signals.

Our study goes toward speed of convergence and quality of the output signal. The justification for a faster algorithm is that biomedical signals [5,17,18,21] are non-stationary and their environment is changing constantly. However, we are rather concerned about finding an algorithm that quickly tracks those changes than about the computational time.

In this work, we propose a self-adaptive step-size for ICA algorithms which accelerates the speed of convergence. Instead of dealing with the non-linear cost function of ICA algorithms which would be optimum, we carry out our analysis in a mean-squared framework. For this approach, we can solve the problem of bounds to the step-size and derive the optimum one for one step convergence. In this field, there is the work of Douglas and Cichocki [15], with focus on decorrelation networks. Cichocki and his colleagues [11] also proposed a self-adaptive step-size. However, our attempt here is to find a step-size which is directly based on the evolution of the algorithm.

Moreover, we propose a neural network consisting of two layers of ICA algorithms. Some works [6,9,16] suggested to carry out whitening before the ICA algorithm in order to orthogonalize the inputs, which yields a faster convergence. The basis of our two-layer network is the same. However, we argue that using a cascade of two ICA algorithms is a stronger principle, because both are searching for independent solutions. Belouchrani et al. [8] also proposed a multi-layer network, but they were not interested in comparing the multi-layer results with the pre-whitening. This is carried out here, by simulations, for different initial conditions.

## 2. Independent component analysis (ICA)

The principle of ICA may be understood as follows. Consider  $n$  source signals  $\mathbf{s} = [s_1, s_2, \dots, s_n]^T$  arriving at  $m$  receivers. Each receiver gets a linear combination  $\mathbf{x}$  of the source signals, so that we have

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}, \quad (1)$$

---

<sup>1</sup> Such as the one proposed in [5,17,22].

where  $\mathbf{A}$  is an  $m \times n$  matrix, and  $\mathbf{n}$  is the noise, which is omitted because it is usually impossible to distinguish noise from the source signals, therefore, we omit it from now on. The purpose of ICA is to find a matrix  $\mathbf{B}$ , that multiplied by  $\mathbf{A}$ , will cancel the mixing effect. For simplicity, we assume that matrix  $\mathbf{A}$  is a  $n \times n$  invertible square matrix. Ideally,  $\mathbf{BA} = \mathbf{I}$ , where  $\mathbf{I}$  is the identity.

The system output is then given by

$$\mathbf{z} = \mathbf{Bx} = \mathbf{BA}s = \mathbf{Cs}, \quad (2)$$

where the elements of vector  $\mathbf{s}$  must be *mutually independent*. In mathematical terms, it means that the joint probability density of the source signals must be the product of the marginal densities of the individual sources [19]

$$p(\mathbf{s}) = \prod_{i=1}^M p(s_i). \quad (3)$$

Thus, instead of searching for a solution that uncorrelates the signals, ICA looks for the most independent signals which is a much stronger principle.

### 3. ICA as a density shaper

With ICA, one wants to find a way to estimate the true distribution  $p(\mathbf{s}, \theta)$  of a random variable, given the samples  $\mathbf{z}_1, \dots, \mathbf{z}_N$ . In other words, ICA is a probability density estimator, or density shaper.

Given the modifiable parameters  $\hat{\theta}$ , we should find a density estimator  $\hat{p}(\mathbf{z}, \hat{\theta})$  of the true density  $p(\mathbf{s}, \theta)$ . This may be performed by entropy maximization, mutual information minimization, maximum likelihood or Kullback–Liebler (K–L) divergence. We take, for instance, the K–L divergence, given by

$$I\{p(\mathbf{s}, \theta), \hat{p}(\mathbf{z}, \hat{\theta})\} = \int p(\mathbf{s}, \theta) \log \frac{p(\mathbf{s}, \theta)}{\hat{p}(\mathbf{z}, \hat{\theta})} d\mathbf{z}. \quad (4)$$

A small value of the K–L divergence  $I\{p(\mathbf{s}, \theta), \hat{p}(\mathbf{z}, \hat{\theta})\}$  indicates that  $\hat{p}(\mathbf{z}, \hat{\theta})$  is close to the true density  $p(\mathbf{s}, \theta)$ . Hence, we should minimize  $I\{p(\mathbf{s}, \theta), \hat{p}(\mathbf{z}, \hat{\theta})\}$ , and this can be carried out by using a gradient method. However, instead of the conventional Euclidean gradient method [6], which reads

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \mu_k \frac{\partial}{\partial \hat{\theta}_k} I\{p(\mathbf{s}, \theta), \hat{p}(\mathbf{z}, \hat{\theta})\}, \quad (5)$$

we rather use the following gradient:

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \mu_k \mathfrak{I} \frac{\partial}{\partial \hat{\theta}_k} I\{p(\mathbf{s}, \theta), \hat{p}(\mathbf{z}, \hat{\theta})\}, \quad (6)$$

where  $\mathfrak{I}$  is a positive-definite matrix.

This is called the relative [9,10], natural, or Riemannian gradient [3,4]. This algorithm works better in general because the parameter space of neural networks is Riemannian [1].

To obtain a better estimation, Pearlmutter and Parra [20] derived an algorithm that extracts many parameters related to the signal, and therefore their parameter space  $S = \{\hat{\theta}\}$  was built with many variables. However, in most of the works, the gradient method shown above was derived using only the weight matrix  $\mathbf{B}$  as the parameter to be estimated. For this case, we have  $\mathfrak{J} = \mathbf{B}^T \mathbf{B}$  and the weights of  $\mathbf{B}$  are updated by [9]

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \mu_k [\mathbf{I} - N(\mathbf{z}_k)] \mathbf{B}_k, \quad (7)$$

where  $N(\cdot)$  is a non-linear function. In this work, we use the following function, as suggested by Bell and Sejnowski [6]:

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \mu_k (\mathbf{I} - \mathbf{y}_k \mathbf{z}_k^T) \mathbf{B}_k, \quad (8)$$

with  $\mathbf{z} = \mathbf{B}\mathbf{x}$  and  $\mathbf{y} = \tanh(\mathbf{z})$ .

And this is the *secret* of ICA: this non-linearity *tries* to shape the sources distribution. In other words, if one expands this non-linearity in a series of Taylor, higher order moments appear. For example, if we use one sigmoidal function, which is used frequently in neural networks, one can see that

$$\tanh(u) = u - \frac{u^3}{3} + \frac{2u^5}{15} + \dots \quad (9)$$

It should be added, however, that even though these methods are said to *blindly* estimate the sources, some *prior* knowledge is necessary in order to choose this non-linearity. As we have seen, ICA is also known to be a density estimator. In other words, the non-linearity  $g$  should be chosen so that

$$y_i = g(u) \approx \int_{-\infty}^u f_s(v) dv, \quad (10)$$

where  $f_s$  is the density of  $s$ .

In practice, it is not very necessary for this equation to hold. For signals with a super-Gaussian distribution (kurtosis  $> 0$ ), it did not pose as a problem to separate them using Eq. (8). In the case of sub-Gaussian signals, Cichocki and his colleagues [12] suggested the following equation:

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \mu_k (\mathbf{I} - \mathbf{z}_k \mathbf{y}_k^T) \mathbf{B}_k. \quad (11)$$

An interesting discussion about this topic was carried out by Amari [2].

### 3.1. Indeterminacy of the solution

Because the system works in a *blind* manner,  $\mathbf{B}$  does not necessarily converge to the inverse of  $\mathbf{A}$ . We can only affirm that  $\mathbf{C} = \mathbf{D}\mathbf{P}$ , where  $\mathbf{D}$  is a diagonal, and  $\mathbf{P}$  is a permutation matrix [13]. Without any a priori information, which is the case of blind source separation, nothing can be done concerning to the permutation, but we can still normalize the weight matrix to avoid the problem of random scaling. An

interesting solution is to preserve the entropy of the input signal, normalizing the weights by<sup>2</sup>

$$\mathbf{B} \leftarrow |\det(\mathbf{B})|^{1/n} \mathbf{B}. \quad (12)$$

### 3.2. Equivariance property

An *equivariant* estimator  $A(\cdot)$  for an invertible  $n \times n$  matrix  $M$  is defined as [9]

$$A(M\mathbf{z}_k) = MA(\mathbf{z}_k). \quad (13)$$

This property can be applied to relative gradient algorithms. Multiplying both sides of Eq. (7) by  $A$  yields

$$\mathbf{C}_{k+1} = \mathbf{C}_k - \mu_k [\mathbf{I} - N(\mathbf{C}_k \mathbf{s}_k)] \mathbf{C}_k. \quad (14)$$

Therefore, the trajectory of the global system  $\mathbf{C} = \mathbf{BA}$  is *independent* of  $A$ . In other words, if one changes the initial weight, the whole learning trajectory will be changed.

### 3.3. Filtering

Sometimes a filtering operation can be very useful in some ill-conditioned mixing problems. Here we discuss a filtering operation that *preserves* the mixing matrix. With this, one can alter the power spectral response of the signals in order to help in the quality of the separation. If the matrix is preserved and its inverse (or a scaled version of it) can be estimated by ICA, then one can easily recover the source signals.

A causal filter for the mixed vector  $\mathbf{x}$  with impulse response  $\mathbf{H}(t, \tau)$ , can be described by

$$\mathbf{y}(t) = \int_{-\infty}^t \mathbf{H}(t, \tau) \mathbf{x}(\tau) d\tau = \int_{-\infty}^t \mathbf{H}(t, \tau) \mathbf{A} \mathbf{s}(\tau) d\tau. \quad (15)$$

We assume that this first-order linear system may be *time-variant*. Moreover, we should find a  $\mathbf{H}(t, \tau)$  so that the following holds:

$$\mathbf{y}(t) = \mathbf{A} \int_{-\infty}^t \mathbf{H}(t, \tau) \mathbf{s}(\tau) d\tau. \quad (16)$$

In other words, the filtering operation should not alter the structure of matrix  $\mathbf{A}$ . For this to happen, the impulse response  $\mathbf{H}(t, \tau)$  should be a diagonal matrix with the same elements, i.e.,  $\mathbf{H}(t, \tau) = h(t, \tau) \mathbf{I}$ , which implies that the elements of vector  $\mathbf{x}$  should be passed through the same filter.

<sup>2</sup> With a unitary determinant, the entropy of the mixed and of the output signal will be the same [14].

#### 4. Time-varying step-size for ICA algorithms

From Eq. (7), the output correlation matrix will be given by

$$\mathbf{T}_k = E[\mathbf{z}_k \mathbf{z}_k^T] = E[\mathbf{B}_k \mathbf{x}_k \mathbf{x}_k^T \mathbf{B}_k^T]. \quad (17)$$

In this analysis, we make use of the independent assumption. This effectively implies that  $\mathbf{x}_k$  is independent of former values and that the elements of  $\mathbf{B}_k$  are mutually independent. This is very common in the field of adaptive filtering to use this assumption, even though it is rarely true in practice.

If the fluctuations in the elements of  $\mathbf{B}_k$  are small, we can thus rewrite Eq. (17) as<sup>3</sup>

$$\mathbf{T}_k = E[\mathbf{B}_k] \mathbf{R} E[\mathbf{B}_k^T], \quad (18)$$

where  $\mathbf{x}$  is assumed to be stationary constant; in other words, the input covariance  $\mathbf{R} = E[\mathbf{x}_k \mathbf{x}_k^T]$  is constant. However, the same cannot be said about  $\mathbf{T}_k = E[\mathbf{z}_k \mathbf{z}_k^T]$ , thus it is assumed to be non-stationary.

Using Eqs. (8) and (18), we can write

$$\begin{aligned} \mathbf{T}_{k+1} &= E[(1 + \mu_k) \mathbf{I} - \mu_k \mathbf{y}_k \mathbf{z}_k^T] \mathbf{T}_k E[(1 + \mu_k) \mathbf{I} - \mu_k \mathbf{z}_k \mathbf{y}_k^T] \\ &= (1 + \mu_k)^2 \mathbf{T}_k - \mu_k (1 + \mu_k) \mathbf{T}_k \mathbf{P}_k - \mu_k (1 + \mu_k) \mathbf{P}_k^T \mathbf{T}_k + \mu_k^2 \mathbf{P}_k^T \mathbf{T}_k \mathbf{P}_k, \end{aligned} \quad (19)$$

where  $\mathbf{T}_k = E[\mathbf{z}_k \mathbf{y}_k^T]$ .

If we assume that the variation of  $\mathbf{z}_k$  is bounded to the interval  $[-1, 1]$ , we can then say that in this limit  $\mathbf{y}_k \approx \mathbf{z}_k$  and  $\mathbf{P}_k \approx \mathbf{T}_k$ . Then, Eq. (19) can be written as

$$\mathbf{T}_{k+1} = (1 + \mu_k)^2 \mathbf{T}_k - 2\mu_k (1 + \mu_k) \mathbf{T}_k^2 + \mu_k^2 \mathbf{T}_k^3. \quad (20)$$

There is a unitary matrix  $\mathbf{Q}$  that diagonalizes  $\mathbf{T}_k$  so that  $A_i = \mathbf{Q}^T \mathbf{T}_i \mathbf{Q}$  and  $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$ . Thus, we can rewrite Eq. (20) as

$$A_{k+1} = (1 + \mu_k)^2 A_k - 2\mu_k (1 + \mu_k) A_k^2 + \mu_k^2 A_k^3. \quad (21)$$

Notice that when deriving Eq. (21) from Eq. (20) the orthogonal property of  $\mathbf{Q}$  was used.<sup>4</sup>

From Eq. (21), the eigenvalues of  $\mathbf{T}_k$  are the elements of  $A_k$ , and are given by

$$\lambda_{k+1,i} = (1 + \mu_k)^2 \lambda_{k,i} - 2\mu_k (1 + \mu_k) \lambda_{k,i}^2 + \mu_k^2 \lambda_{k,i}^3. \quad (22)$$

For uniform convergence in a mean-squared sense, it is required that  $\lambda_{k+1,i} < \lambda_{k,i}$ , which yields the following bounds for the step-size:<sup>5</sup>

$$0 < \mu_k < \frac{2}{\lambda_{k,i} - 1}. \quad (23)$$

<sup>3</sup> The following steps are similar to the one carried out by Douglas and Cichocki [15].

<sup>4</sup> For example,  $\mathbf{Q}^T \mathbf{T} \mathbf{V}_k \mathbf{Q} = \mathbf{Q}^T \mathbf{T} \mathbf{Q} \mathbf{Q}^T \mathbf{V}_k \mathbf{Q}$ .

<sup>5</sup> This derivation is carried out simply by substituting  $\lambda_{k+1,i} < \lambda_{k,i}$  in Eq. (22).

From Eq. (23), the optimum step-size which will give one-step convergence is

$$\mu_{\text{opt}} = \frac{2}{\lambda_{k,i} - 1}. \quad (24)$$

Using Eq. (23) and the assumption that  $\mathbf{y}_k \approx \mathbf{z}_k$ , we propose then to use the following step-size to update the weight matrix:

$$\mu_k = \frac{2}{\mathbf{y}_k^T \mathbf{z}_k + 1}. \quad (25)$$

When proposing the step-size above, we had in mind the following:

$$\frac{2}{\sum_i \lambda_{k,i} + 1} < \frac{2}{\lambda_{k,i} + 1} < \frac{2}{\lambda_{k,i} - 1}. \quad (26)$$

## 5. A network for fast blind separation

In this work, we are mainly interested in using ICA to filter noises which are possibly overlapping in frequency the cardiac signal. However, we do not want that this use of ICA implies in a longer time of convergence. Usually, ICA has a slower convergence compared to the classical LMS, because ICA is also estimating higher-order moments. Thus, we propose an architecture to deal with this matter.

It is very common among researchers to use a whitening filter before the ICA algorithm itself. The reason is that the whitening carries out a decorrelation between the input signals. Then, the ICA work is reduced to estimate the moments higher than two, with this, one gains in speed.

Here, we propose a different reasoning. Instead of using only second-order statistics, we suggest the use of a network that substitutes the whitening by an ICA algorithm itself. With this, we are not only estimating the second, but also higher-order moments. We will see that this simple substitution implies in a much *faster* convergence.

The architecture includes other points to improve the speed of convergence as shown in Fig. 1. In resume, they are given below.

- Pre-process the mixed signals by a high-pass filter operation that obeys Eq. (16). Later we will discuss why this is important.
- Use a time-varying step-size for faster convergence as in Eq. (25).
- Use a two-layer network. The two layers are cascaded in series and the first layer is only used for convergence. The second layer is updated in a batch mode, and the first, at every iteration (to avoid instability of convergence) and it is *turned off* after a given number of iterations.

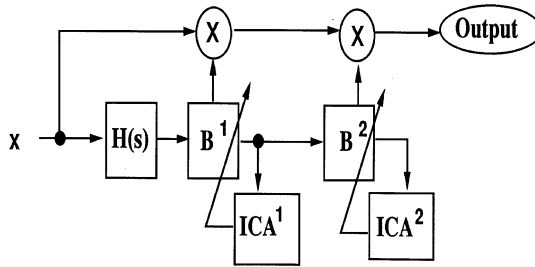


Fig. 1. Block diagram of the proposed method. The signal is inputted into a high-pass filter, then to the first layer (ICA<sup>1</sup>), which is updated only up to a number of iterations. The other layer (ICA<sup>2</sup>) is always “turned on”.

## 6. Simulations

We have carried out simulations to test the validity of the proposed method. The simulation consisted in mixing actual ECG and electrode motion artifact (usually the result of intermittent mechanical forces acting on the electrodes) signals. We used signals from the MIT-BIH noise stress test database, which are standard for testing ECG analyzers. Their power spectrum is shown in Fig. 2. Notice that the fundamental frequency of the ECG signal (around 1 Hz) is overlapped in frequency by the electrode artifact one. The mixing was carried out using different random matrices.<sup>6</sup> The first layer (ICA<sup>1</sup>) in Fig. 1 was turned off after 1000 iterations. For all cases, we initialized the weight matrix by the identity matrix. The filter cutoff frequency was 2 Hz, and the weights were updated every block size of 50 iterations.

## 7. Results

We used as figure of merit to measure the quality of separation at the  $k$ th iteration the following equation:

$$\varphi_k = 200 \sum_{j=1,2} (\tilde{c}(k)_j - 0.5),$$

$$\tilde{c}(k)_j = \max \left\{ \frac{|c_{i,j}|}{\sum_i |c_{i,j}|} \right\} \quad \text{for } j = 1, 2. \quad (27)$$

With this index, we are measuring how far the matrix  $C$  is from the solution  $DP$  at each iteration. When  $C = DP$ , only one element at each column/line is different from zero. The index  $\varphi$  will be 100 for the best case, and will be null for the worst.

<sup>6</sup> By the equivariance property, we can conclude that this would be equivalent to keeping the mixing matrix constant and changing the weight initial value.



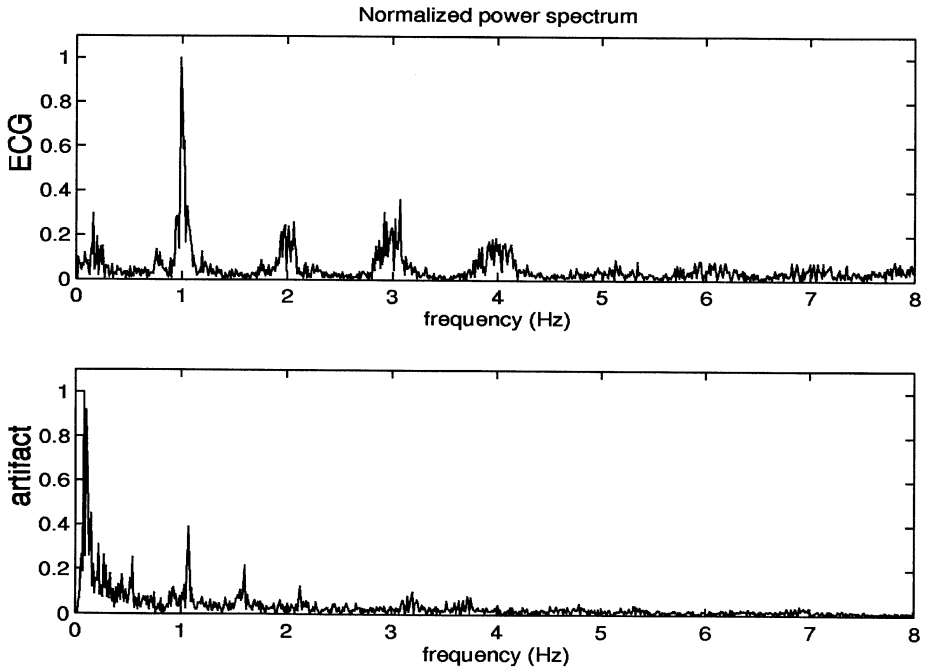


Fig. 2. Power spectrum of the two source signals: ECG and electrode motion artifact. Notice that the first harmonic of the ECG signal (around 1 Hz) is overlapped in frequency by the respiratory one.

We have extensively run the proposed algorithm for different randomly mixed vectors. Fig. 3 shows a summary of the results. The simulations were carried out using:

- Filtering and *no* filtering. For these two cases, we carried out the simulations using:
  - The proposed method as in Fig. 1, with two ICA algorithms and a time-varying step-size as in Eq. (11).
  - The same configuration of the item above, but instead of an ICA algorithm, we used a *whitening* one in the first layer. This was conducted by substituting the non-linear function in Eq. (8) by  $y = \text{erf}(z)$ .<sup>7</sup>
  - One ICA algorithm as in Eq. (11).

The index as in Eq. (27) was calculated for each simulation, and they are shown in Fig. 3. The signals recovered by the proposed network are shown in Fig. 4. The “recovered signals” were obtained after normalizing the weights as in Eq. (12).

<sup>7</sup>This algorithm was named by Bell and Sejnowski [6] as “Gaussian component analysis” (GCA), to differ from PCA and ICA. This function searches “for the decorrelated solution which gives the most Gaussianly distributed outputs”.

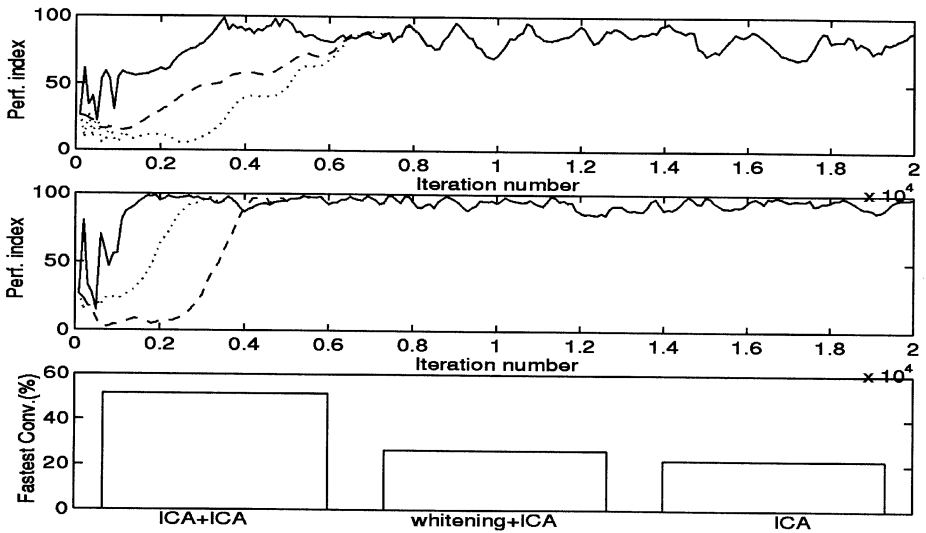


Fig. 3. *top*: This plot shows the performance index as in Eq. (27) for different algorithms, but *without filtering*. The *solid* line is for the case of the proposed ICA + ICA architecture; the *dashed*, for *whitening* + ICA one; and the *dotted*, for ICA only. *middle*: same as the top, but *with filtering*. Notice the lower variance compared to the plot in the top. *bottom*: this plot shows which algorithm was the fastest in an ensemble of 70 runnings for different initial matrices  $C$ , not mattering the time of convergence.

## 8. Discussion

By looking at Fig. 3 we arrive at the following conclusions:

- Comparing the plots in the top and in the middle, we can see that the filtering was important in order to have less variance after convergence. This is because the lower-frequency signal (trend) was removed. The trend is usually accounted in the literature as non-stationary mean. Besides, the high-pass filter is also removing the overlapping that occurs at 1 Hz, and we believe that this fact also helped in order to have a better quality in the output.
- The two-layer ICA network performed better in general than the others, as one can see in the plot in the bottom. We should enforce, however, the fact that in the referred plot, we did not care about the speed of convergence, but rather, about which one reached first an acceptable level of separation ( $\varphi$  in Eq. (27) around 80).

Another point that should be emphasized is that of the adaptive step size. When we started using ICA, the greatest problem in our point of view was that of step-size. Since we wanted a fast convergence, we had to fix the step-size to some upper value, otherwise the algorithm would not converge. Therefore, for these ECG and electrode noise data, we found *heuristically* an upper bound of  $2 \times 10^{-4}$  for the learning rate. We compared then this learning rate with the adaptive one, derived here. Fig. 5 shows

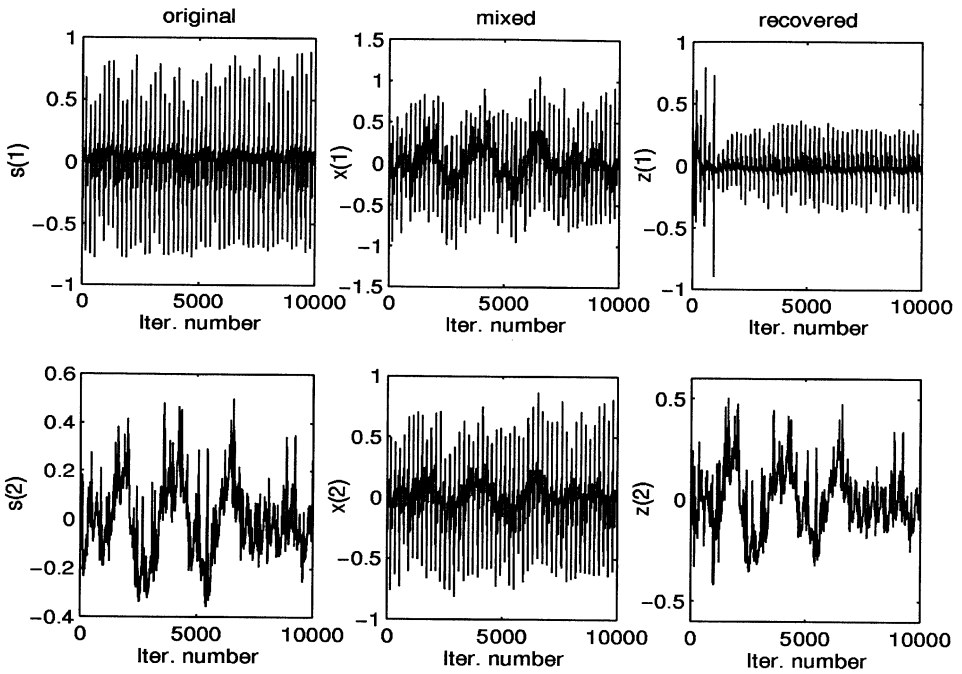


Fig. 4. An example of the original, mixed and recovered signal by the proposed network. In this case,  $A = [1 \ 1; 0.9 \ 1]$ .

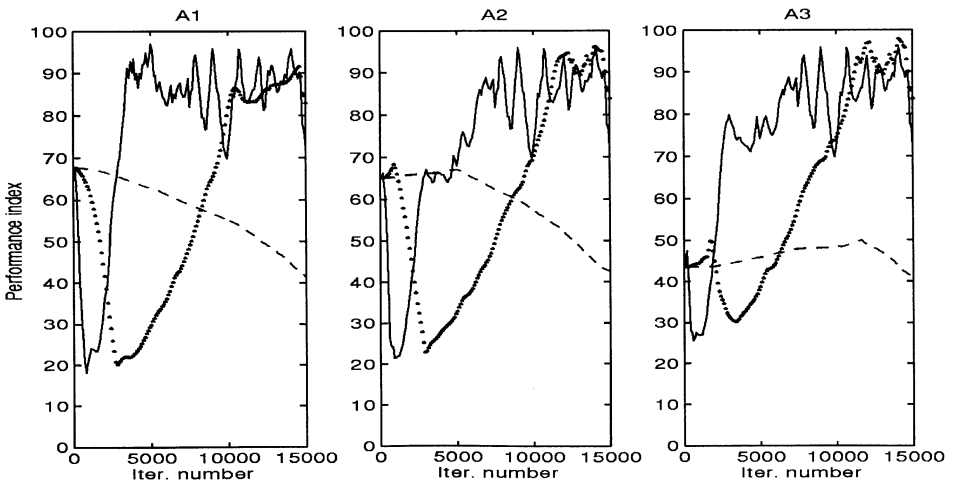


Fig. 5. Values of  $p$  as in Eq. (27) for different initial random weight matrices  $C1$ ,  $C2$  and  $C3$ . The labels are as follows: *solid*: ICA with adaptive learning rate; *dotted*: ICA with a constant learning rate of  $2 \times 10^{-4}$ ; *dashed*: ICA with a constant learning rate of  $2 \times 10^{-5}$ .

this result. We can see that the self-adaptive learning rate allowed a much faster learning, without diverging.

Some words are necessary about the two-layer networks. Some works, e.g., [7,16] have proposed to carry out whitening before the ICA processing in order to orthogonalize the input components. In the same way, we have used the first layer to *force* the algorithm to search for independent components. We argue, however, that cascading two ICA algorithms is a stronger principle because the first layer looks for an *independent* rather than an orthogonal solution. Contrary to whitening, that uses only second-order statistics, ICA makes also use of higher-order moments.

It is important to emphasize that the proposed network, in all the simulations, converged in less than 2000 iteration, while it did not always happen with the others. Again Fig. 3 shows a good example of it. By looking at the plot in the middle, we can see that the ICA algorithm reached convergence at around 3000 iterations, and the ICA + whitening network, in around 4000 iterations. For ECG signals, which are usually sampled at frequencies around 100 Hz, it means a delay of 10–20 s, i.e., the data in this interval should be disregarded.

Probably, the reader is asking why we did not use the two layers in the whole trajectory, but rather, we switched it off after some iterations. We carried it out, but the variance after convergence for such configuration was higher. Therefore, roughly speaking, the first layer works as a *propulsion* to *put the algorithm in the way* to converge to one of the solutions  $C = DP$ .

## 9. Conclusions

In this work, we proposed an architecture to blindly separate linearly mixed signals, based on the independent component analysis principle. The architecture consisted of a high-pass filter, a two-layer network based on ICA algorithm and a self-adaptive step-size. The self-adaptive step-size was theoretically derived from the mean behavior of the output signal.

The proposed network composed of two ICA algorithms converged faster than the one composed of whitening plus an ICA algorithm, where *whitening* stands for an algorithm designed to orthogonalize the input signals. We argued that the two-layer network of ICA algorithm behaves better because the first layer is searching for an independent solution, rather than an orthogonal. This conclusion was confirmed by simulations. The proposed self-adaptive step-size also lead to a fast convergence, though with a greater error.

## References

- [1] S. Amari, Information geometry of the EM and em algorithms for neural networks, *Neural Networks* 8 (9) (1995) 1379–1408.
- [2] S. Amari, T. Chen, A. Cichocki, Stability analysis of adaptive blind source separation, *Neural Networks* 10 (8) (1997) 1345–1351.

- [3] S. Amari, A. Cichocki, H.H. Yang, A new learning algorithm for blind signal separation, *Adv. Neural Inform. Processing Systems*, vol. 8, MIT Press, Cambridge, MA, 1996.
- [4] S. Amari, A. Cichocki, H.H. Yang, Gradient learning in structured parameter spaces: adaptive blind separation of signal sources, *WCNN*, 1996.
- [5] A.K. Barros, N. Ohnishi, MSE behavior of biomedical event-related filters, *IEEE Trans. Biomed. Eng. BME-44* (1997) 848–855.
- [6] A.J. Bell, T.J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, *Neural Comput.* 7 (1995) 1129–1159.
- [7] A.J. Bell, T.J. Sejnowski, Fast blind separation based on information theory, *Proc. Int. Symp. on Nonlinear Theory and Applications (NOLTA)*, vol. 1, Las Vegas, December 1995, pp. 43–47.
- [8] A. Belouchrani, A. Cichocki, K.A. Meraim, A blind identification and separation technique via multi-layer neural networks, *Proc. ICONIP'96*, vol. 2, Springer, Singapore, 1996, pp. 1195–1200.
- [9] J-F. Cardoso, On the performance of source separation algorithms, *Proc. EUSIPCO*, Edinburgh, September 1994, pp. 776–779.
- [10] J-F. Cardoso, B.H. Laheld, Equivariant adaptive source separation, *IEEE Trans. Signal Process. SP-44* (1996) 3017–3030.
- [11] A. Cichocki, S. Amari, M. Adachi, W. Kasprzak, Self-adaptive neural networks for blind separation of sources, *Proc. Int. Symp. Circuits Systems 2* (4) (1996) 157–160.
- [12] A. Cichocki, S. Amari, M. Adachi, W. Kasprzak, Local adaptive learning algorithms for blind separation of natural images, *Neural Network World* 6 (4) (1996) 515–523.
- [13] P. Comon, Independent component analysis, a new concept?, *Signal Process.* 24 (1994) 287–314.
- [14] G. Deco, W. Brauer, Nonlinear higher-order statistical decorrelation by volume-conserving neural architectures, *Neural Networks* 8 (1995) 525–535.
- [15] S. Douglas, A. Cichocki, Neural networks for blind decorrelation of signals, *IEEE Trans. Signal Process.* 45 (11) (1997) 2829–2842.
- [16] J. Karhunen, Neural approaches to independent component analysis and source separation., *Proc. 4th European Symp. on Artificial Neural Networks (ESANN'96)*, Bruges, Belgium, 1996.
- [17] P. Laguna, R. Janè, O. Meste, P. Poon, P. Caminal, H. Rix, N.V. Thakor, Adaptive filter for event-related bioelectric signals using an impulse correlated reference input: Comparison with signal averaging techniques, *IEEE Trans. Biomed. Eng. BME-39* (1992) 1032–1043.
- [18] S. Makeig, A.J. Bell, T-P. Jung, T. Sejnowski, Independent component analysis of electroencephalographic data, *Advances in Neural Information Processing Systems*, vol. 8, MIT Press, Cambridge, MA, 1996.
- [19] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1991.
- [20] B.A. Pearlmutter, L.C. Parra, A Context-Sensitive Generalization of ICA, *Int. Conf. on Neural Information Processing*, Hong Kong, September 1996.
- [21] R. Vigarío, Extraction of ocular artefacts from EEG using independent component analysis, *Electroencephalogr. Clin. Neurophysiol.* 103 (1997) 395–404.
- [22] C. Vaz, X. Kong, N.V. Thakor, An adaptive estimation of periodic signals using a Fourier linear combiner, *IEEE Trans. Signal Process. ASSP-42* (1994) 1–10.



**Allan Kardec Barros** received the B.S. degree in electrical engineering from the Universidade Federal do Maranhão, Brazil, in 1991, the M.S. degree from Toyohashi University of Technology, Japan, in 1995 and the D.Eng. degree from Nagoya University, Japan, in 1998. He is currently a Frontier Researcher at RIKEN in Japan.

His research interests are biomedical signal processing, speech processing and blind signal separation.



**Ali Mansour** was born at Tripoli in Lebanon, in 19 October 1969. He received the M.S. degree in the electronic electric engineering in September 1992 from the Lebanese university (Tripoli - Lebanon), and the PhD degree on Signal Processing, in January 1997 from the Institut National Polythenique de Grenoble INPG (France).

In 1993, he joined the staff of Prof. Jutten. His PhD subject was: The Blind Separation of Sources. In August 1997, he joined the staff of Prof. Ohnishi at the Bio-Mimetic Sensory Systems.



**Noboru Ohnishi** received the B.Eng., M.Eng. and D.Eng. degrees from Nagoya University, Nagoya, Japan, in 1973, 1975 and 1984, respectively. From 1975 to 1986 he was with the Rehabilitation Engineering Center under the Ministry of Labor. From 1986 to 1989 he was an Assistant Professor in the Department of Electrical Engineering, Nagoya University. From 1989 to 1994 he was an Associate Professor. He is a Professor in the Department of Information Engineering, and concurrently a Head of Laboratory for Bio-mimetic Sensory System at the Bio-mimetic Control Research Center of RIKEN. His research interests include computer-vision and -audition, robotics, bio-cybernetics, and rehabilitation engineering. Dr. Ohnishi is a member of IEEE, IEICE, IPSJ, SICE, JNNS, IIITE and RSJ.