Independent component analysis based on symmetrised scatter matrices

S. Taskinen [a,*], S. Sirkiä [a], H. Oja [b]

[a]*Department of Mathematics and Statistics, University of Jyväskylä*

[b]*Tampere School of Public Health, University of Tampere*

Abstract

A new method for separating the mixtures of independent sources has been proposed recently in [8]. This method is based on two scatter matrices with the so called independence property. The corresponding method is now further examined. Simple simulation studies are used to compare the performance of so called symmetrised scatter matrices in solving the independence component analysis problem. The results are also compared with the classical FastICA method. Finally, the theory is illustrated by some examples.

Key words: affine equivariance, FastICA, independence property, source separation, symmetrised scatter matrix

1 Introduction

Independent component analysis (ICA) is a useful technique in various signal processing as well as data analysis applications. Briefly, the idea in ICA is to find a transformation that transforms the data to independent components. See e.g. [4] for a review.

The classical ICA model is in its simplest form

$$\boldsymbol{x} = A\boldsymbol{s},$$

where the random k-vector $\boldsymbol{x}$ is the observed mixture of independent components, A is an unknown $k \times k$ mixing matrix of full-rank and $\boldsymbol{s} = (s_1, \ldots, s_k)^T$

---

* Corresponding author: Sara Taskinen, Department of Mathematics and Statistics, P.O. Box 35, FIN-40014 University of Jyväskylä, Finland; tel: +358-14-260 2991, email: slahola@maths.jyu.fi

has independent components. At most one of the independent components is allowed to be gaussian. Now using only observations $\boldsymbol{x}$, the goal in ICA is to find a matrix B such that $B\boldsymbol{x}$ has independent components. Matrix B is then called as unmixing matrix. Note that if D is a $k \times k$ diagonal matrix and P is a $k \times k$ permutation matrix, then

$$\boldsymbol{x} = (AP^{-1}D^{-1})(DP\boldsymbol{s}),$$

where $DP\boldsymbol{s}$ has independent components as well. Therefore, $\boldsymbol{s}$ may be defined only up to multiplying constants and a permutation. This ambiguity is, however, insignificant in most applications.

In this paper, we will show how the ICA problem may be solved using two different scatter matrices. This approach to ICA is introduced in [8]. In Section 2, we will recall the definition of scatter matrix and show how scatter matrices can be used to build up the so called symmetrised scatter matrices. Some examples of symmetrised scatter matrices are also given. In Section 3, we will show how scatter matrices are used to solve the ICA problem. Different methods are compared through simulation studies in Section 4 and some examples are given in Section 5.

2 Symmetrised scatter matrices

In the next section, we will show how two different scatter matrices may be used to find the unmixing matrix that transforms the observations back to independent components, but recall first the definition of a scatter matrix.

Definition 1 *Let $\boldsymbol{x}$ be a random k-vector with cdf F. A symmetric $k \times k$ matrix valued functional $S(\boldsymbol{x})$ is a scatter matrix if it is positive definite and affine equivariant in the sense that*

$$S(A\boldsymbol{x} + \boldsymbol{b}) = AS(\boldsymbol{x})A^{T},$$

where A is any nonsingular $k \times k$ matrix and $\boldsymbol{b}$ is a k-vector.

As will be seen in the next section, scatter matrices cannot be used in independent component analysis unless they have the following property.

Definition 2 *If $S(\boldsymbol{x})$ is a diagonal matrix for all $\boldsymbol{x}$ with independent components, then $S(\boldsymbol{x})$ is said to have the independence property.*

An example of the scatter matrix with independence property is the regular covariance matrix, but for example regular M-functionals and S-functionals do not have this property. In [8] it is, however, shown that by computing

any scatter matrix using pairwise differences one obtains a matrix with the independence property. In other words, if $S(\boldsymbol{x})$ is a scatter matrix, then

$$S(\boldsymbol{x}_1 - \boldsymbol{x}_2),$$

where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent copies of $\boldsymbol{x}$, is a scatter matrix with the independence property. Note that if pairwise differences are used to compute the scatter matrix, no explicit location vector is needed in the definition. In the following the scatter matrices based on pairwise differences will be called as symmetrised scatter matrices.

In the next, consider some examples of symmetrised scatter matrices. The symmetrised M-functional $S(\boldsymbol{x})$ is based on the regular M-functional [7,5] and it solves

$$E[w_1(r)(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T - w_2(r)S(\boldsymbol{x})] = 0,$$

where $r = [(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T S(\boldsymbol{x})^{-1}(\boldsymbol{x}_1 - \boldsymbol{x}_2)]^{1/2}$ and w_1 and w_2 are some non-negative weight functions. This scatter functional is proposed and studied in [10]. The examples of symmetrised M-functionals include the following:

(i) The weight functions for symmetrised Huber's M-estimator are given by

$$w_1(r) = \begin{cases} 1/\sigma^2, & r^2 \leq c^2 \\ c^2/r^2\sigma^2, & r^2 > c^2 \end{cases} \quad \text{and} \quad w_2(r) = 1,$$

where c is a tuning constant defined so that $q = Pr(\chi_k^2 \leq c^2/2)$ for a chosen q. The scaling factor σ is such that $E[w_1(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)] = k$.

(ii) Dümbgen's estimator [2] uses $w_1(r) = k/r^2$ and $w_2(r) = 1$. This estimator is the symmetrised version of Tyler's M-estimator [11]. Note that Dümbgen's estimator is defined only up to a constant. Therefore it is affine equivariant in the sense that $S(A\boldsymbol{x} + \boldsymbol{b}) \propto AS(\boldsymbol{x})A^T$.

Another symmetrised covariance estimator used in this paper is the spatial Kendall's tau matrix

$$S(\boldsymbol{x}) = E\left[\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2}\right].$$

This matrix is introduced and studied in [12] and it is a symmetrised version of the so called spatial sign covariance matrix (see e.g. [6,12]). Note that the spatial Kendall's tau matrix is not a real scatter matrix, since it is only orthogonal equivariant, that is, $S(U\boldsymbol{x} + \boldsymbol{b}) \propto US(\boldsymbol{x})U^T$ for any orthogonal $k \times k$ matrix U. As shown in [8], it can, however, be used in the independent component analysis.

The above estimators are highly robust, therefore we will include one nonrobust estimator to our comparisons, that is, the matrix of fourth moments of

3

differences
$$S(\boldsymbol{x}) = E\left[||\boldsymbol{x}_1 - \boldsymbol{x}_2||^2(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T\right].$$
Like the spatial Kendall's tau matrix, this matrix is only orthogonal equivariant.

3   Independent component analysis based on scatter matrices

In this section, we will show how two different scatter matrices may be used to solve the ICA problem. Let now S_1 and S_2 be two scatter matrices with the independence property and assume that

$$\boldsymbol{x} = A\boldsymbol{z},$$

where $\boldsymbol{z}$ has independent components, and scatter matrices are such that

$$S_1(\boldsymbol{z}) = I_k \quad \text{and} \quad S_2(\boldsymbol{z}) = D(\boldsymbol{z}),$$

where $D(\boldsymbol{z})$ is a diagonal matrix with diagonal elements $d_1 > \cdots > d_k \geq 0$. The assumption on distinct eigenvalues guarantees that the transformation matrix A is uniquely defined. Therefore, in this approch, the independent components cannot have the same marginal distributions.

In [8], it is shown that if the unmixing matrix is computed using

$$B(\boldsymbol{x}) = \left[U_2\left([S_1(\boldsymbol{x})]^{-1/2}\boldsymbol{x}\right)\right]^T\left[S_1(\boldsymbol{x})\right]^{-1/2}, \tag{1}$$

where $U_2(\boldsymbol{x})$ is the matrix of eigenvectors of $S_2(\boldsymbol{x})$, then

$$B(\boldsymbol{x})\boldsymbol{x} = J\boldsymbol{z}$$

for some diagonal matrix J with diagonal elements ± 1. Thus the transformation in (1) gives the independent components up to a sign. In this approach, the first scatter matrix is used to whiten the data. By writing $A = ULV^T$, where U and V are orthogonal matrices and L is a diagonal matrix (singular value decomposition), it is easy to see that $[S_1(\boldsymbol{x})]^{-1/2}\boldsymbol{x} = UV^T\boldsymbol{z}$. Then, since $U_2\left([S_1(\boldsymbol{x})]^{-1/2}\boldsymbol{x}\right) = UV^T$, the second scatter matrix may be used to find the rotation matrix that yields to independent components. Note also that in this approach, the components become ordered according to their marginal kurtosis.

Since the above ICA transformation requires scatter matrices with the independence property, the symmetrised scatter matrices are very useful here. In [8], it is also shown that it is enough to assume that S_2 is only orthogonally equivariant.

4 Comparison of the methods

In this section, we will compare the performance of our scatter matrix based methods with the popular FastICA method [3] through a simple simulation study. Most of the ICA algorithms are based on the idea that due to the central limit theorem, the distribution of mixed sources tends to gaussian distribution. Therefore, the algorithms search for the unmixing matrix that maximizes the nongaussianity of the sources. In FastICA, the nongaussianity is measured with approximation to negentropy, that is, the negentropy of y is approximated with

$$J(y) = \{E[G(y)] - E[G(\nu)]\}^2,$$

where ν is a random variable from a standard normal distribution. According to [3], some useful choices for function G are

$$G(u) = \alpha^{-1} \log \cosh(\alpha u) \quad \text{and} \quad G(u) = -\exp(-u^2/2),$$

where $1 \le \alpha \le 2$.

In the following we will compare the accuracies of estimated unmixing matrices. The accuracies are measured using the so called Amari-error [1], which is a standard performance index in the field of signal separation. If the estimated unmixing matrix is denoted by $\widehat{B}$ and true mixing matrix by A, then the Amari-error is defined as

$$E = \frac{1}{2k(k-1)} \left\{ \sum_{i=1}^{k} \left(\sum_{j=1}^{k} \frac{|p_{ij}|}{max_h |p_{ih}|} - 1 \right) + \sum_{j=1}^{k} \left(\sum_{i=1}^{k} \frac{|p_{ij}|}{max_h |p_{hj}|} - 1 \right) \right\},$$

where $P = (p_{ij}) = \widehat{B}A$. Since the unmixing matrix may only be estimated up to sign changes and permutation of rows, the perfect separation implies that P is a permutation matrix consisting of values 1 and/or -1. In that case $E = 0$. Moreover, when the estimation procedure fails to separate the sources, the value of E increases. The inaccuracy measure is scaled so that the maximum value is 1.

In the first experiment, four independent sources were considered. Sources were generated from a standard normal distribution, t_5-distribution, uniform distribution and cauchy-distribution and then mixed with a random 4×4 matrix. The unmixing matrices were then estimated using different methods and the Amari-errors were computed in each case. The sample size was 1000 and the process was replicated 200 times with the same mixing matrix. The boxplots of resulting simulation errors are illustrated in Figure 1. Algorithm (1) corresponds to FastICA with $G(u) = \log \cosh(u)$ and (2) to FastICA with $G(u) = -\exp(-u^2/2)$. In all our scatter matrix based methods, we use the

sample covariance matrix as S_1. The other scatter matrix S_2 is chosen to be the symmetrised Huber's M-estimator in (3), the Dümbgen's estimator in (4), the spatial Kendall's tau matrix in (5) and the matrix of fourth moments of the differences in (6). For the symmetrised Huber's M-estimator, we chose $q = 0.9$ as the tuning constant. The corresponding estimator is then highly efficient, but still robust against some very extreme outliers. Also smaller values of q were applied here, but the resulting Amari-errors were then higher than in the case of $q = 0.9$. In the second case, two skewed distributions were included in the simulation study, that is, lognormal distribution and Rayleigh distribution. These two distributions are widely used in modeling several signal processing phenomena. The simulation setup was similar as before. The results are also given in Figure 1.
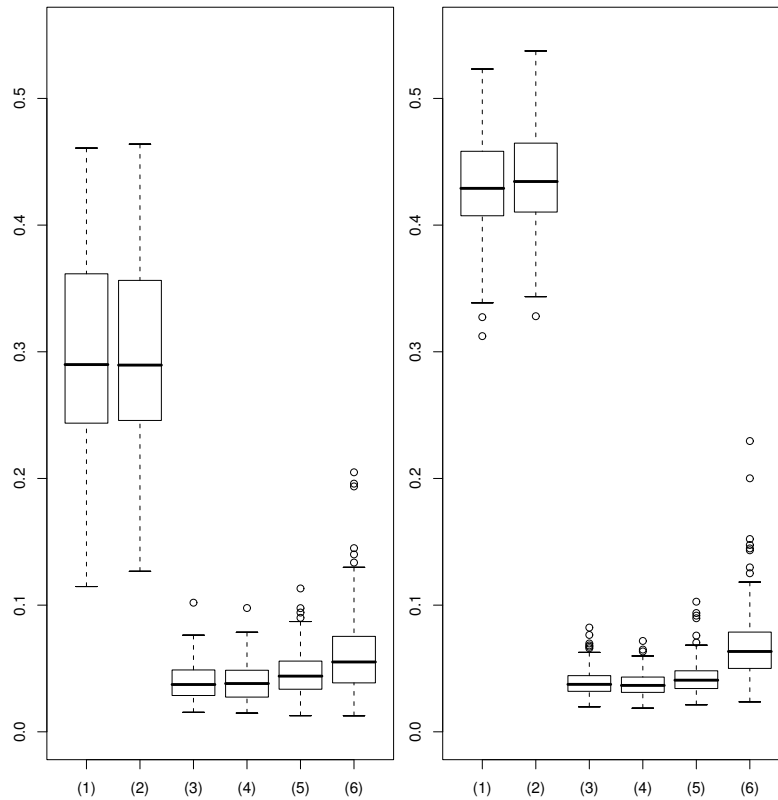

Fig. 1. Separation performance of four symmetrically distributed sources (left figure) and four symmetric and two skewed sources (right figure) of length 1000. The boxplots (1) and (2) correspond to FastICA with two different G functions. Other boxplots correspond to scatter matrix based methods, where S_1 is the sample covariance matrix and S_2 is the symmetrised Huber's M-estimator (3), the Dümbgen's estimator (4), the spatial Kendall's tau matrix (5) and the matrix of fourth moments of the differences (6).

Consider first the simulation results in case of symmetrical sources. The use of symmetrised M-estimators or spatial Kendall's tau matrix in separating the sources yields to very low inaccuracies. The matrix of fourth moments of the differences works also well, even though some higher inaccuracies are encountered. Both FastICA methods fail to separate the sources more often than the scatter matrix based methods. When two sources from the skewed distributions were included in the study, the variances of inaccuracies given by FastICA methods decrease and the separation performances are always poorer than those of the scatter matrix based methods. This result was, however, expected, since it is well known that the FastICA method performs poorly in case of asymmetric distributions.

Table 1 lists the times needed to compute one ICA transformation using the algorithms described before. The transformations were computed with C-functions called from R. As seen from the table, the methods based on symmetrised M-estimators are quite slow to compute. However, those algorithms that use the spatial Kendall's tau matrix or the matrix of fourth moments of the differences are very fast even in the case $k = 4$ and $n = 2000$. The FastICA algorithms are naturally the fastest ones, but as seen from the simulation study, these algorithms often yield to poor estimates of unmixing matrices.

Table 1
Computation times (in seconds) of different ICA algorithms on an AMD Athlon(tm) XP 2200+ (1.8 GHz) and 515 MB of RAM.

		Algorithm					
	n	(1)	(2)	(3)	(4)	(5)	(6)
$k = 2$	500	0.00	0.01	1.31	1.40	0.10	0.08
	1000	0.01	0.01	5.16	5.13	0.39	0.32
	2000	0.01	0.01	20.97	20.37	1.57	1.27
$k = 3$	500	0.01	0.00	1.75	1.93	0.18	0.13
	1000	0.01	0.01	6.88	7.61	0.67	0.54
	2000	0.01	0.01	27.41	29.44	2.66	2.13
$k = 4$	500	0.01	0.01	2.54	4.23	0.27	0.21
	1000	0.01	0.01	33.87	19.07	1.06	0.83
	2000	0.01	0.03	40.23	71.97	4.23	3.37

5 Examples

5.1 Signal separation

To illustrate the performance of our method, the example with mixed signals is first considered. Three independent signals were generated using the Matlab function *demosig* and are illustrated in Figure 2. The sample size was 500 and as a mixing matrix A we used a random 3×3 matrix. The true unmixing matrix is now

$$B = \begin{pmatrix} 1.829 & -1.254 & -0.203 \\ -0.358 & -0.140 & 1.256 \\ -1.743 & 3.189 & -0.009 \end{pmatrix}$$

and the mixed signals are given in Figure 3.

The signals were then separated using the transformation in (1). In this example S_1 was chosen to be the sample covariance matrix and as S_2 we used the spatial Kendall's tau matrix. Other symmetrised scatter estimators, however, yielded to results of the same type. The estimated unmixing matrix is now

$$\widehat{B} = \begin{pmatrix} 1.876 & -1.275 & -0.315 \\ -0.201 & -0.249 & 1.234 \\ 1.775 & -3.206 & -0.015 \end{pmatrix},$$

which is, in spite of some sign changes, very close to the true one. The separation results are shown in Figure 4. Note that except the flipped axis of third signal, the sources are very well separated.

In the second example, four independent speech signals of length 1000 were considered. This data is downloaded from www2.ele.tue.nl/ica99/. The signals were mixed using a random 4×4 matrix and then separated using our scatter matrix based method with the sample covariance matrix as S_1 and the Dübgen's estimator as S_2. The original speech signals are given in Figure 5 and the separated signals in Figure 6. As seen from these Figures, the resulting signals are permutated, but very well separated. The permutation results from the fact that in the second scatter matrix computed on original independent components, the diagonal values were not ordered. Again, if the other symmetrised scatter matrices are used here, the results are more or less the same.

In this example, we use our estimation procedure to separate artificial mixtures of three natural images. The images were downloaded from the website www.cis.hut.fi/projects/ica/data/images/. The original 130×130 pixel images were vectorized by stacking the rows next to each other. The random 3×3 matrix was then used as a mixing matrix A. The original images are given in the first row of Figure 7 and mixed images in the second row. The true unmixing matrix is now

$$B = \begin{pmatrix} 15.439 & 3.815 & -21.871 \\ -15.552 & -1.718 & 21.376 \\ 32.892 & 3.104 & -42.021 \end{pmatrix}.$$

And using the transformation (1) with sample covariance matrix as S_1 and spatial Kendall's tau matrix as S_2 we obtain

$$\widehat{B} = \begin{pmatrix} 35.547 & 3.745 & -45.776 \\ -13.534 & -3.616 & 19.489 \\ 15.687 & 1.496 & -21.338 \end{pmatrix},$$

which is, except the permutation, very close to the true unmixing matrix. For the separated images, see the third row of Figure 7. Note that in this example, our data set consisted of 16900 3-dimensional observations, therefore the use of symmetrised M-estimators here was a bit too time-consuming. When the matrix of fourth moments of the differences was used as S_2, the separation was as good as before.

Finally, note that the both FastICA methods were also applied to the above data sets. These methods, however, failed to separate the sources. The corresponding Amari-errors were approximately 0.5, while the Amari-errors for the methods based on scatter matrices were below 0.1.

References

[1] Amari, S., Cichocki, A. and Yang, H.H. (1996). A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems 8*, MIT Press.

[2] Dümbgen L. (1998). On Tyler's M-functional of scatter in high-dimension. *Ann. Inst. Statist. Math.*, **50**, 471–491.

[3] Hyvärinen, A. and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation*, **9(7)**, 1483–1497.

[4] Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis*, Wiley, New York.

[5] Huber P. J. (1981). *Robust Statistics*, Wiley, New York.

[6] Marden (1999). Some robust estimates of Principal components. *Statistics & Probability Letters*, **43**, 349–359.

[7] Maronna R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.*, **4**, 51–67.

[8] Oja, H., Sirkiä, S. and Eriksson, J. (2006). Scatter matrices and independent component analysis. *Austrian Journal of Statistics*. To appear.

[9] Ollila, E., Hettmansperger, T.P. and Oja, H. (2003). Affine equivariant multivariate sign methods. Under revision.

[10] Sirkiä, S., Taskinen, S. and Oja, H. (2006). Symmetrised M-estimators of scatter. Under revision.

[11] Tyler D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.*, **15**, 234–251.

[12] Visuri, S., Koivunen, V. and Oja, H. (2000). Sign and rank covariance matrices. *Journal of Stat. Plan. Inf.*, **91**, 557–575.
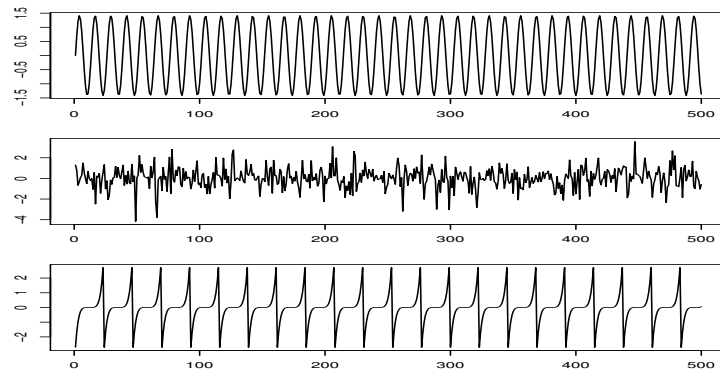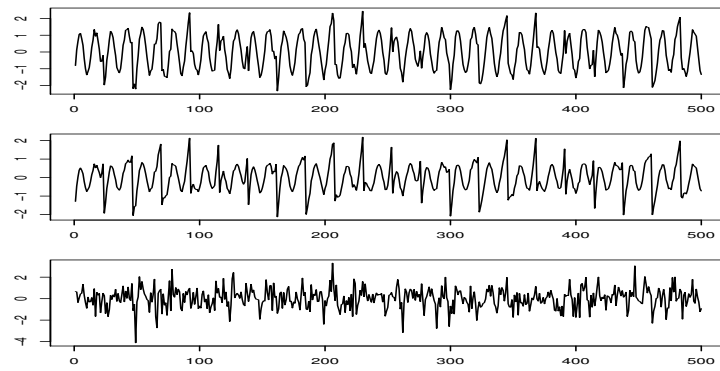
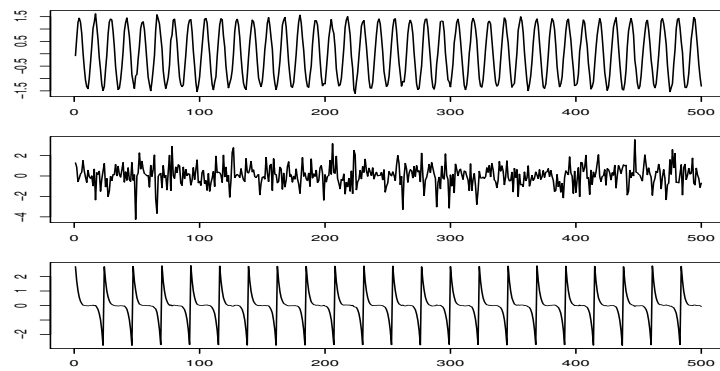Fig. 2. Original signals.

Fig. 3. Mixed signals.
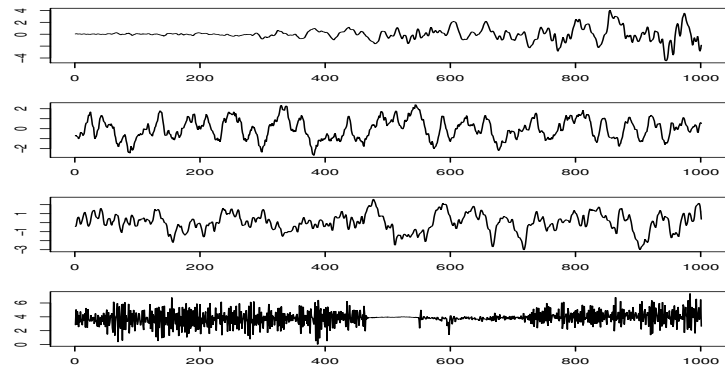

Fig. 4. Separated signals.
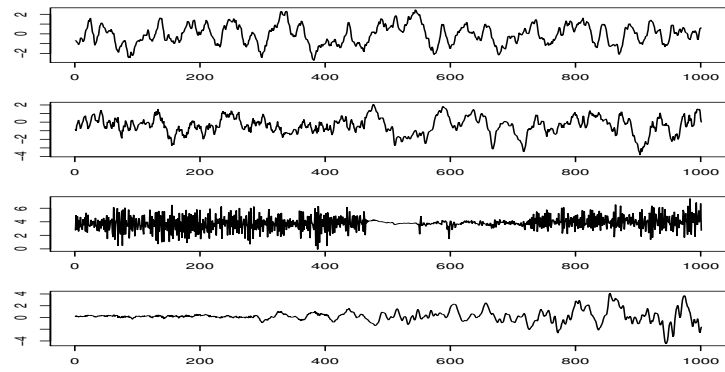
11

Fig. 5. Original speech signals.

Fig. 6. Separated speech signals.

Fig. 7. Original images (first row), mixed images (second row) and separated images (third row).