# ANALYSING ICA COMPONENTS BY INJECTING NOISE

*Stefan Harmeling*[1], *Frank Meinecke*[1], *and Klaus-Robert Müller*[1,2]

[1]Fraunhofer FIRST.IDA, Kekuléstrasse 7, 12489 Berlin, Germany
[2]University of Potsdam, Department of Computer Science,
August-Bebel-Strasse 89, 14482 Potsdam, Germany
{harmeli,meinecke,klaus}@first.fhg.de

**ABSTRACT**

Usually, noise is considered to be destructive. We present a new method that constructively injects noise to assess the reliability and the group structure of empirical ICA components. Simulations show that the true root-mean squared angle distances between the real sources and some source estimates can be approximated by our method. In a toy experiment, we see that we are also able to reveal the underlying group structure of extracted ICA components. Furthermore, an experiment with fetal ECG data demonstrates that our approach is useful for exploratory data analysis of real-world data.

## 1. INTRODUCTION

In order to apply unsupervised learning algorithms to real-world problems it is of fundamental importance to determine how trustworthy their results are.

Recently, Meinecke et al. [11, 10] proposed a bootstrap resampling method that estimates the reliability and grouping of independent components found by algorithms for independent component analysis (abr. ICA, see [4, 1, 7]). This method profits from the well-developed theory of bootstrap (see [5]). However, it is not straightforward for all existing ICA algorithms how to define a resampling strategy that preserves the statistical structure relevant to the considered ICA algorithm.

Our approach refers to the inherent ideas of ICA algorithms: according to Cardoso's three easy routes (see [2]) the statistical structure relevant for ICA algorithms are non-Gaussianity, non-whiteness and non-stationarity. Our method partially destroys this structure by corrupting the data with stationary white Gaussian noise. The motivation for this is, that we expect reliable components to be extracted even if they have lost some of their structure.

ICA models multivariate time-series

$$x(t) = [x_1(t), \ldots, x_n(t)]^\top$$

as a linear combination,

$$x(t) = As(t),$$

of statistically independent source signals

$$s(t) = [s_1(t), \ldots, s_n(t)]^\top.$$

An algorithm for ICA estimates a mixing matrix $A$ only from the observed signal $x(t)$. Therefore, the true sources—or equivalently the columns of the mixing matrix $A$—can be estimated at best up to permutation and scaling.

In this paper, reliability of an estimated source is measured as the stability of its direction with respect to noise, i.e. to a fading-out of the marginal non-properties: non-Gaussianity, non-whiteness and non-stationarity. Two or more unreliable components might span a reliable subspace which can be stable and thus reliably separated from other components. E.g. a rotational invariant distribution might form such a subspace, since it could be well-defined but there is no preferred ICA basis inside (see our toy example or [11]).

## 2. ALGORITHM

Real-world signals are usually given as a multivariate time-series $x(t)$ comprising of $n$ components each of length $T$, which we represent as an $n \times T$ matrix,

$$X = [x(1) \cdots x(T)].$$

We assume that all signals have mean zero. The ICA algorithm tries to estimate from this matrix $X$ the mixing matrix $A$ and therewith the demixing matrix $W = A^{-1}$ such that the demixed signals, i.e. the rows of the matrix

$$Y = WX,$$

are as independent as possible. Bearing in mind the usual indeterminacies of ICA solutions, i.e. arbitrary scaling and permutation, we can assume without loss of generality that the mixing matrix $A$, the inverse of $W$, has unit-length columns[1], i.e.

$$A_{:j}^{\top} A_{:j} = 1 \text{ for } j \in \{1, \ldots, n\}.$$

This ensures that the energy of each component $Y_{j:}$ is equal to the sum of the energies of the proportions of $Y_{j:}$ in the components of $X$, which can be written as $A_{:j} Y_{j:}$, mathematically speaking

$$\begin{aligned} \frac{1}{T}\mathrm{tr}(A_{:j} Y_{j:} Y_{j:}^{\top} A_{:j}^{\top}) &= \frac{1}{T}\mathrm{tr}(A_{:j}^{\top} A_{:j} Y_{j:} Y_{j:}^{\top}) \\ &= \frac{1}{T}\mathrm{tr}(Y_{j:} Y_{j:}^{\top}) \\ &= \frac{1}{T} Y_{j:} Y_{j:}^{\top}, \end{aligned}$$

using for the first equality $\mathrm{tr}(CD) = \mathrm{tr}(DC)$, for the second the fact that the columns of $A$ have unit length and for the third that $Y_{j:} Y_{j:}^{\top}$ is a scalar.

### 2.1. Partially destroying statistical structure

The statistical structure that most algorithms for ICA exploit is non-Gaussianity, non-stationarity or non-whiteness—Cardoso's three easy routes to ICA (cf. [2]). In order to analyze the reliability and grouping of the extracted components $Y$, we will now partially destroy their statistical structure by adding noise, which is stationary Gaussian distributed and independent in time, and examine how the unmixing results change. More precisely, after adding the noise, the signals are more Gaussian, more stationary and spectrally more flat. Note, that the noise level has to be adjusted for each component separately: otherwise some components—the weak ones—might loose all their statistical structure while others—the strong ones—are not affected at all, which would be undesirable since such a procedure would favor strong components over weak components[2].

Let $E$ be the matrix that contains the square roots of the energies of the extracted components on the diagonal—or equivalently speaking their standard deviations,

$$E = \begin{bmatrix} \sqrt{\frac{1}{T} Y_{1:} Y_{1:}^{\top}} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\frac{1}{T} Y_{n:} Y_{n:}^{\top}} \end{bmatrix}.$$

Being aware of the fact that this matrix contains not exactly energies, we call it nonetheless for simplicity *energy matrix*.

[1] We use the colon notation of Golub (see [6]), i.e. $A_{i:}$ is the $i$-th row of $A$ and $A_{:j}$ is its $j$-th column.

[2] Equivalently, we could normalize the extracted signals to variance one and add noise with the same noise level for each of the signals. We pursue the slightly more complicated way that preserves the true variances in order to avoid any alterations of the statistical structure.

Adding $R$ instances of Gaussian white noise, written as $n \times T$ matrices $N^{(1)}, \ldots, N^{(R)}$, that have been adjusted by the energy matrix $E$, to the extracted components $Y$ provides us with $R$ versions of $Y$, the statistical structure of which has been partially destroyed:

$$\begin{aligned} \tilde{Y}^{(1)} &= \cos(\sigma) Y + \sin(\sigma) E N^{(1)} \\ &\vdots \\ \tilde{Y}^{(R)} &= \cos(\sigma) Y + \sin(\sigma) E N^{(R)}. \end{aligned}$$

$0 \leq \sigma \leq \pi/2$ is a parameter that can be visualized as a turning knob: $\sigma = 0$ adds no noise, i.e. all statistical structure is preserved, $\sigma = \pi/2$ produces only noise, i.e. all statistical structure is destroyed and in between some is kept, some is destroyed. Since $\cos^2(\sigma) + \sin^2(\sigma) = 1$ it is guaranteed that the noisy versions have in each component the same energy as their colleagues in $Y$. Later in our experiments, we fix $\sigma = \pi/8$ using empirical evidence (see Sec. 3.1).

### 2.2. Remixing and demixing again

The versions with the damaged statistical structure are mixed by randomly generated mixing matrices $B^{(1)}, \ldots, B^{(R)}$, which have unit-length columns. Hereby, we obtain mixtures in which the initially extracted components keep their energy. The additional mixing is important for algorithms, that depend very much on the starting conditions, and it does not cause any problems in the later analysis.

By applying the chosen ICA algorithm to the remixed versions $B^{(1)}\tilde{Y}^{(1)}, \ldots, B^{(R)}\tilde{Y}^{(R)}$ we obtain demixing matrices $V^{(1)}, \ldots, V^{(R)}$ and hereby demixed signals,

$$\begin{aligned} Z^{(1)} &= V^{(1)} B^{(1)} \tilde{Y}^{(1)} \\ &\vdots \\ Z^{(R)} &= V^{(R)} B^{(R)} \tilde{Y}^{(R)}. \end{aligned}$$

For these signals we are going to measure the angle to the initially extracted components.

### 2.3. Constructing the relevant transformation

Transforming the remixed noisy versions $B^{(r)}\tilde{Y}^{(r)}$ to $Z^{(r)}$, for $r \in \{1, \ldots, R\}$ can be seen as demixing the remixed, initially extracted components $B^{(r)} Y$ but using only part of the statistical structure of the signals.

In order to calculate the angle between a new component $Z_{i:}^{(r)}$ and an initially extracted component $Y_{j:}$ we have to ensure two things:

1. We need to consider the transformation with respect to the normalized signals

$$Y^{normalized} = E^{-1} Y$$

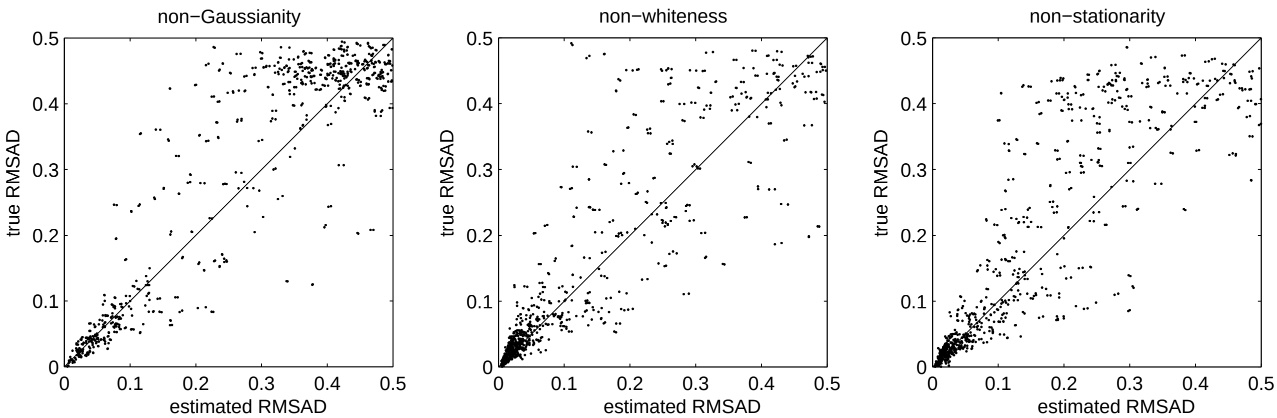


**Fig. 1**. The estimated RMSAD correlates in all three cases with the true RMSAD. Note, that these scatter plots depend on the choice of the noise parameter $\sigma$. All experiments have been carried out with $\sigma = \pi/8$.

each having variance one. Due to

$$\begin{aligned} Z &= V^{(r)}B^{(r)}Y \\ &= V^{(r)}B^{(r)}EY^{normalized} \end{aligned}$$

the transformation to proceed with is

$$V^{(r)}B^{(r)}E.$$

2. The transformed signals must be normalized as well, i.e. we have to left-multiply $V^{(r)}B^{(r)}E$ by a diagonal matrix $D^{(r)}$ such that the rows of

$$U^{(r)} = D^{(r)}V^{(r)}B^{(r)}E$$

have unit-norm. We refer to this matrix as $U^{(r)}$.

The latter matrix describes the relevant transformation: the angle between $Z_{i:}^{(r)}$ and $Y_{j:}$ is the arcus cosine of the absolute value of the $ij$-th entry of that matrix,

$$\alpha_{ij}^{(r)} = \arccos(|U^{(r)}|_{ij}),$$

which is some number between $0$ and $\pi/2$. Note, that we have to take the absolute value of each matrix entry, since orientation does not matter for the calculation of the angle.

### 2.4. Estimating reliability and grouping structure

Using these angles, we compute statistics regarding the initially extracted components $Y$: to begin with we calculate for each component the root mean-squared angle distance to $Y_{j:}$ (abr. RMSAD):

$$v_j = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\min_i(\alpha_{ij}^{(r)})^2} \qquad (1)$$

These values estimate the uncertainty of the extracted components $Y$, as we will see in the experiments section. A large RMSAD means unreliable, a small RMSAD means the corresponding component is reliable.

Furthermore, we define a matrix that displays the grouping structure of the extracted signals, which we call *grouping matrix*:

$$\begin{aligned} S_{jk} &= \frac{1}{R}\sum_{r=1}^{R}\cos(\alpha_{:j}^{(r)})^\top\cos(\alpha_{:k}^{(r)}) \\ &= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{n}\cos(\alpha_{ij}^{(r)})\cos(\alpha_{ik}^{(r)}). \end{aligned}$$

Instead of taking the cosine of the angles we can directly process the values of the transformation $U^{(r)}$,

$$S = \frac{1}{R}\sum_{r=1}^{R}|U^{(r)}|^\top|U^{(r)}|.$$

Intuitively speaking, the $jk$-th entry of $|U^{(r)}|^\top|U^{(r)}|$ is large if there is at least one component $Z_{i:}^{(r)}$ in which both signals $Y_{j:}$ and $Y_{k:}$ concur with large proportions. If that is the case, $Y_{j:}$ and $Y_{k:}$ contribute to the same subspace and are grouped together.

Note, the possible block-structure of $S$ can be automatically obtained using the second eigenvector using ideas from spectral clustering.

## 3. EXPERIMENTS

In order to validate our approach, we show empirically that our method approximates true angle deviations, and we apply it to a toy data set, where we know the true signals and finally to some real-world data. For all experiments, we use

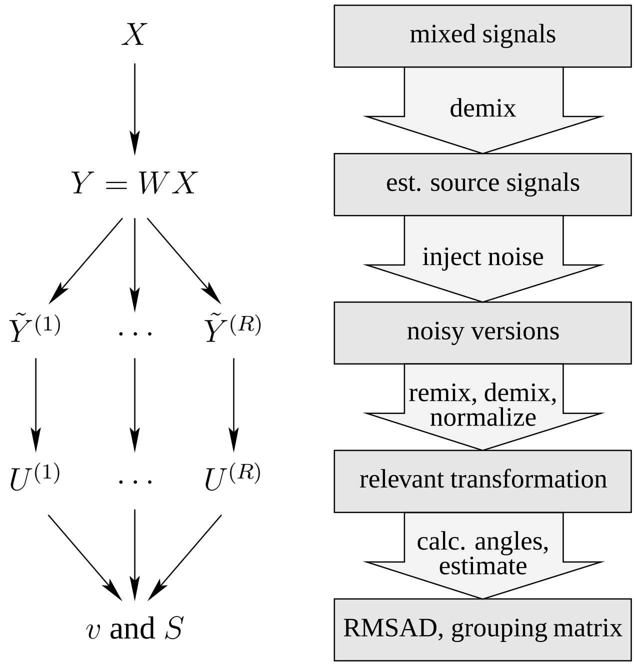


**Fig. 2**. This schematic view shows our method at a glance.

the same three algorithms, each going exclusively one of the easy routes to ICA: as an example of an algorithm that exploits only non-Gaussianity we use JADE (see [3]), for non-whiteness TDSEP (see [13]) and for non-stationarity a simplistic variant of SEPAGAUS (see [12]), which we forced to ignore all statistical information but non-stationarity[3].

### 3.1. True versus estimated RMSAD

Comparing the reliability of signals obtained by different algorithms requires to check whether the RMSAD estimated by our method is correlated to the true one. The latter is defined as the RMSAD of the projection directions (rows of the demixing matrix) estimated from different instances of a certain process. Our estimated RMSAD is given by Eq. (1), which is an estimate from noisy versions of one specific instance of the same process.

The coordinates of each point in Fig. 1 show the true and estimated RMSAD for one particular process. In all three plots we observe that these values are correlated. This means that our estimate of the RMSAD for one particular instance is with high probability close to the true one. Therefore it makes sense to compare reliability, expressed as our estimated RMSAD, between different algorithms.

Note, that this finding depends very much on the choice of $\sigma$ which controls the signal to noise ratio. In all experiments reported here $\sigma$ has been fixed to $\pi/8$.

[3]The input parameters of SEPAGAUS allow us to specify that only one positive frequency channel will be used.

### 3.2. Toy data

Running our method in a completely controlled environment enables us to easily evaluate our results. We consider seven signals that show different combinations of statistical structure:

| 7 signals | non-Gaussianity | non-whiteness | non-stationarity |
|---|---|---|---|
| Speech | + | + | + |
| Music | + | + | + |
| Cosine | + | + | - |
| Sine | + | + | - |
| Uniform noise | + | - | - |
| Gaussian noise | - | - | - |
| Gaussian noise | - | - | - |

These signals are mixed by a randomly chosen matrix and analyzed with the three ICA algorithms mentioned above. The results are visualized in Fig. 3: the RMSAD—depicted in the right-most column—reveals which components have been recovered reliably. However, the group structure can not be directly infered from the RMSAD: the reliability barplot using non-whiteness resembles the plot based on non-stationarity, but the underlying group structure is very different, as can be seen in the grouping matrices. Furthermore, the statistical structure shown in the table above is matched by those matrices: in the matrix for non-Gaussianity we see five blocks that correspond to five subspaces. Both two-dimensional subspaces—sine vs. cosine and Gaussian noise vs. Gaussian noise—are examples where the corresponding distributions are rotation-invariant. Besides, the algorithm exploiting only non-whiteness is not able to differentiate between the i.i.d. signals, as can be seen by the three-dimensional block. Using only non-stationarity even less structure can be identified. Note, that this finding does not rank the three non-properties: it reflects only the statistical properties of the investigated signals, which is exactly what our method is intended to do.

### 3.3. Fetal ECG

As an example of a real-world data set, we present our results on the fetal ECG data (from [9]) which contains 2500 data points sampled at 500Hz with 8 electrodes located at abdomen and thorax of a pregnant woman. The goal of applying ICA algorithms to this data is to separate the fetal heartbeat from the heartbeat of the mother. The uncertainty, shown as the RMSAD in the right-most column of Fig. 4, reveals that the non-Gaussianity based algorithm is the method of choice for this dataset, very much in agreement with Meinecke et al. [11]. The grouping matrices underline this, because they clearly show that six independent

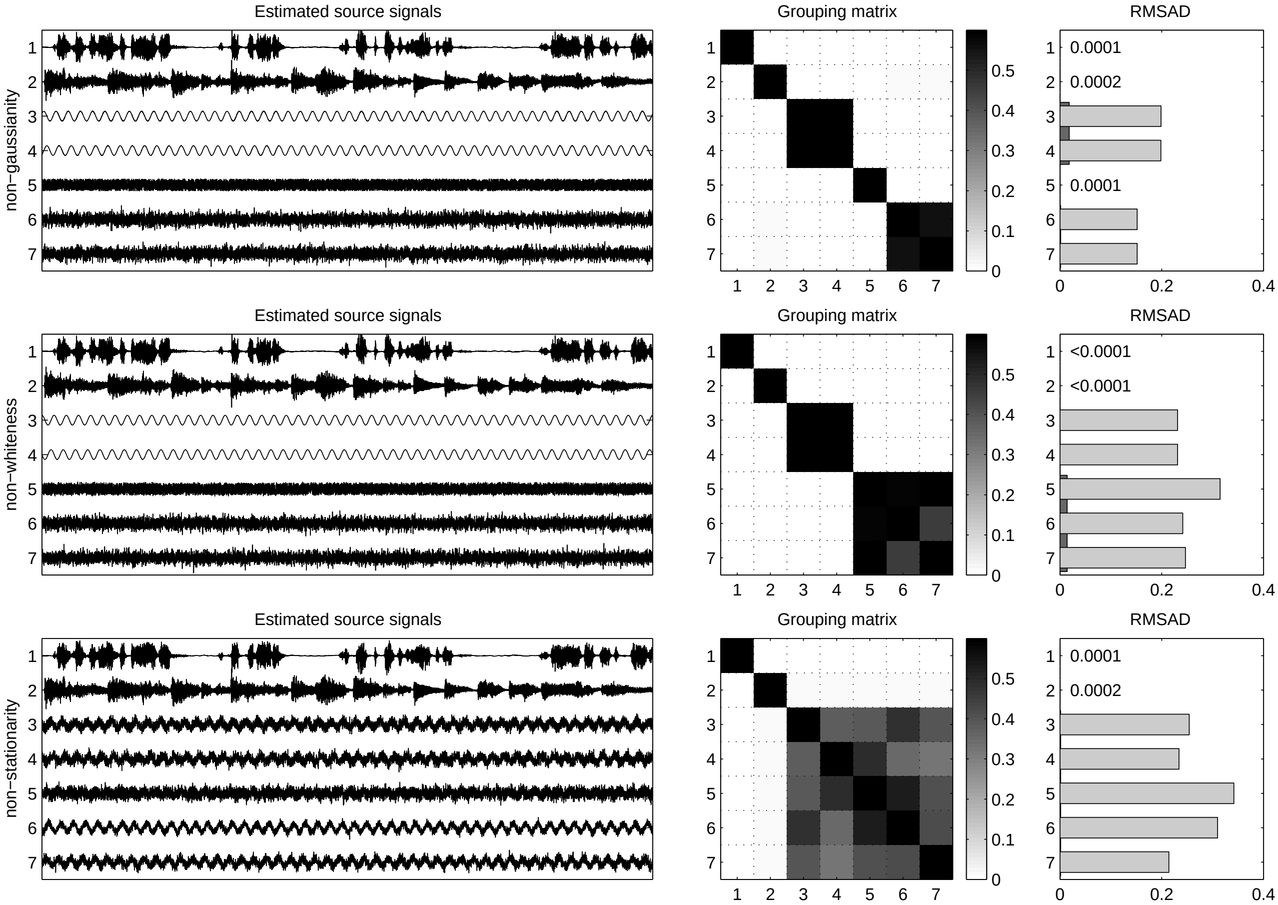


**Fig. 3**. Shown are the results for the toy data. The structure revealed by the grouping matrix and by the RMSADs fits the table in the explaining text. For the first row we used JADE, for the second row we used TDSEP with time-lags $0, \ldots, 20$, for the third we used SEPAGAUS being restricted to one positive frequency channel.

signals have been found. Finally, looking at the estimated waveforms in the first row of the figure, we see that channels 1, 2, 3 and 4 contain the mother's heartbeat and channels 7 and 8 the fetal's heartbeat. The other algorithms did not extract those signals very well.

## 4. CONCLUSION

We presented a new method to assess the reliability of ICA components that can be easily applied to any ICA algorithm. In contrast to our previous work that uses bootstrap (cf. [11]), there is not yet a developed mathematical theory supporting this approach. However, we showed empirically that our estimator is able to approximate the true RMSAD. Controlled toy experiments and experiments with fetal ECG data underlines the usefulness of our approach.

Interesting open questions are: can we improve on our results by choosing the signal to noise ratio more cleverly? How does our method react in an overfitting-scenario (cf. the work in [8])? Future work will also strive for a better understanding of the theoretical properties of our approach.

**Acknowledgements:** The authors wish to thank Andreas Ziehe, Motoaki Kawanabe, Benjamin Blankertz. This work was partly supported by the EU project (IST-1999-14190 – BLISS) and the support from the BMBF under contract FKZ 01IBB02A.

## 5. REFERENCES

[1] J.-F. Cardoso. Blind signal separation: statistical principles. *Proceedings of the IEEE*, 9(10):2009–2025, 1998.

[2] J.-F. Cardoso. The three easy routes to independent component analysis: contrasts and geometry. In T.-W. Lee, editor, *Proc. of the ICA 2001 workshop*, 2001.

[3] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non Gaussian signals. *IEE Proceedings-F*, 140(6):362–370, 1993.

[4] P. Comon. Independent component analysis—a new concept? *Signal Processing*, 36:287–314, 1994.

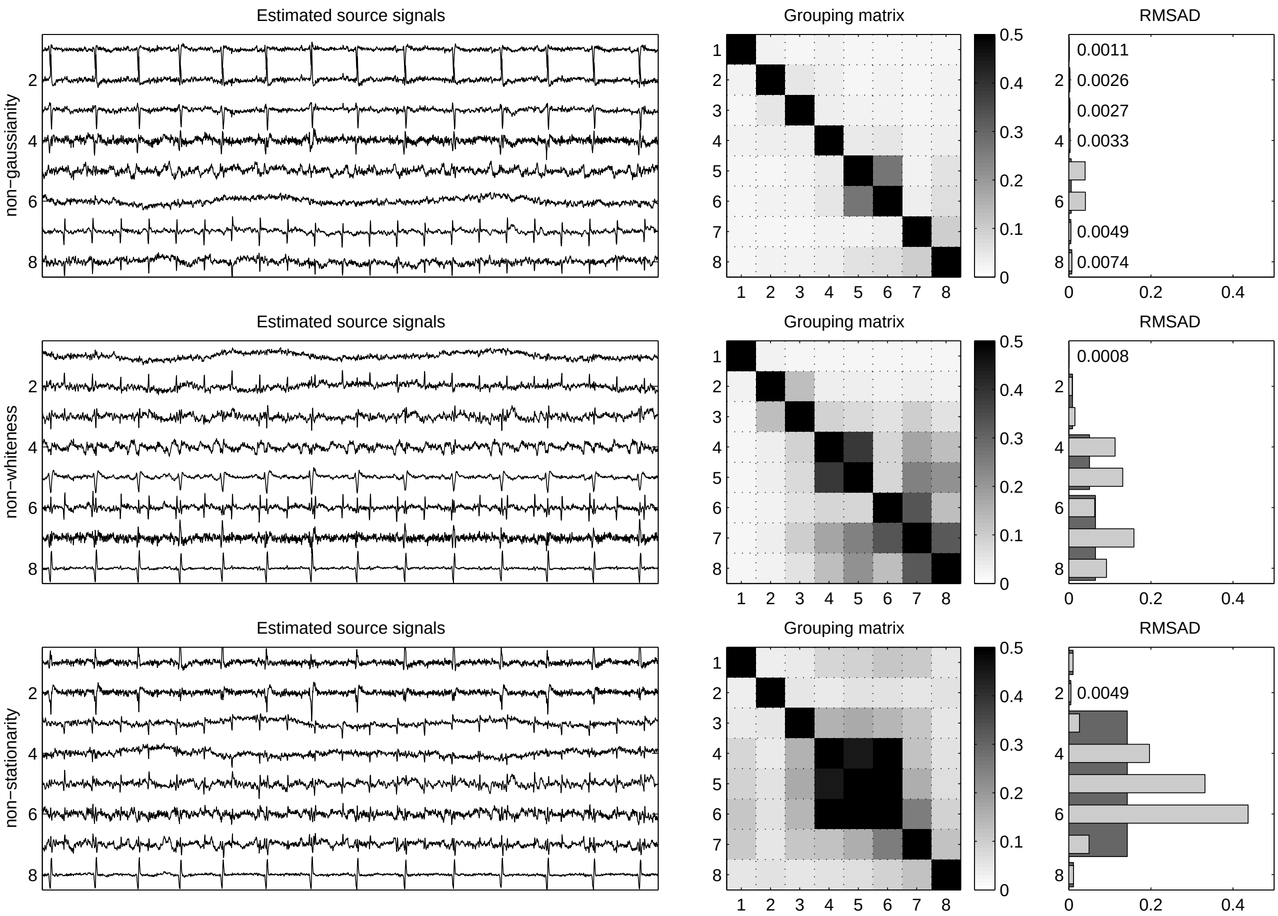


**Fig. 4**. Shown are the results for the fetal ECG. As explained in the text, the extracted components from the ICA algorithm based on non-Gaussianity are the most reliable. For the first row we used JADE, for the second row we used TDSEP with time-lags $0, \ldots, 20$, for the third we used SEPAGAUS being restricted to one positive frequency channel.


[5] B. Efron and R.J. Tibshirani. Improvements on cross-validation: the .632+ bootstrap method. *J. Amer. Statist. Assoc*, 92:548–560, 1997.

[6] G. Golub and C. van Loan. *Matrix Computations*. The Johns Hopkins University Press, 3rd edition, 1996.

[7] A. Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley, 2001.

[8] A. Hyvärinen, J. Särelä, and R. Vigário. Spikes and bumps: Artefacts generated by independent component analysis with insufficient sample size. In *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pages 425–429, Aussois, France, 1999.

[9] L. D. Lathauwer, B. D. Moor, and J. Vandewalle. Fetal electrocardiogram extraction by source subspace separation. In *Proceedings of HOS'95*, Aiguabla, Spain, 1995.

[10] F. Meinecke, A. Ziehe, M. Kawanabe, and K.-R. Müller. Estimating the reliability of ICA projections. In T.G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems*, volume 14. MIT Press, 2002.

[11] F. Meinecke, A. Ziehe, M. Kawanabe, and K.-R. Müller. A resampling approach to estimate the stability of one- or multidimensional independent components. To appear in IEEE Transactions on Biomedical Engineering, 2002.

[12] Dinh-Tuan Pham. Blind separation of instantaneous mixture of sources via the gaussian mutual information criterion. *Signal Processing*, 81:855–870, 2001.

[13] A. Ziehe and K.-R. Müller. TDSEP - an efficient algorithm for blind separation using time structure. In L. Niklasson, M. Bodén, and T. Ziemke, editors, *Proceedings of the 8th International Conference on Artificial Neural Networks*, Perspectives in Neural Computing, pages 675–680, Berlin, 1998. Springer Verlag.