

# Analyzing consistency of independent components: An fMRI illustration

Jarkko Ylipaavalniemi<sup>a,\*</sup> and Ricardo Vigário<sup>a,b,c</sup>

<sup>a</sup>*Adaptive Informatics Research Centre, Helsinki University of Technology, P.O. Box 5400, FI-02015 TKK, Finland*

<sup>b</sup>*Brain Research Unit, Helsinki University of Technology, P.O. Box 2200, FI-02015 TKK, Finland*

<sup>c</sup>*Advanced Magnetic Imaging Centre, Helsinki University of Technology, P.O. Box 3000, FI-02015 TKK, Finland*

Received 28 June 2006; revised 19 June 2007; accepted 15 August 2007

Available online 28 August 2007

**Independent component analysis (ICA) is a powerful data-driven signal processing technique. It has proved to be helpful in, e.g., biomedicine, telecommunication, finance and machine vision. Yet, some problems persist in its wider use. One concern is the reliability of solutions found with ICA algorithms, resulting from the stochastic changes each time the analysis is performed. The consistency of the solutions can be analyzed by clustering solutions from multiple runs of bootstrapped ICA. Related methods have been recently published either for analyzing algorithmic stability or reducing the variability. The presented approach targets the extraction of additional information related to the independent components, by focusing on the nature of the variability. Practical implications are illustrated through a functional magnetic resonance imaging (fMRI) experiment.**

© 2007 Elsevier Inc. All rights reserved.

*Keywords:* Consistency; Variability; Bootstrap; Bagging; Boosting; Clustering; Independent component analysis; Functional magnetic resonance imaging; Brain; Functional areas

## Introduction

Blind source separation (BSS) is defined as finding underlying source signals using only observed signals with unknown mixing and minimal, if any, prior information on the sources (Cardoso, 1990; Jutten and Herault, 1991). Independent component analysis (ICA) (Jutten and Herault, 1991; Comon, 1994; Hyvärinen et al., 2001) is perhaps the most widely used method for performing blind source separation. ICA is implemented in many algorithms, such as, FastICA (Hyvärinen, 1999), Infomax (Bell and Sejnowski, 1995; Amari et al., 1995) and JADE (Cardoso, 1989). ICA works in a clear data-driven manner, based only on the assumption that

the original sources are statistically independent. The solution of the BSS problem has useful applications in many research fields, including biomedicine, telecommunication and finance. ICA has been used for identifying signals of interest, removing artifacts and reducing noise (see, e.g., Makeig et al., 1995; Jahn et al., 1998; Vigário et al., 2000; Tang et al., 2002; Funaro et al., 2003). ICA can also be used in feature extraction, with applications in, for example, natural image processing and modeling of human vision (see, e.g., Olshausen and Field, 1996; Hurri et al., 1996; Hoyer and Hyvärinen, 2000).

However, some methodological problems remain in the wide adoption of ICA. One concern is the tendency of the estimated sources, or independent components, to change slightly each time the analysis is performed. Such variability can be caused by many factors. For example, the theoretical assumption of statistical independence may not hold for the data. The algorithmic implementation of ICA may also be inherently stochastic. Furthermore, additive noise or other features of the data can cause variations in the solutions.

Under expert supervision, such behavior is usually overcome by comparing the estimated components with known sources, or with references obtained by other methods (cf. Calhoun et al., 2001). Another approach is to evaluate the results based on knowledge of the true sources. Such approaches can effectively cancel the true benefits of data-driven analysis. Moreover, they may not even be possible in practice. Recently, bootstrapping methods (Efron and Tibshirani, 1994) have been used successfully to identify consistently reproducible components (see, e.g., Duann et al., 2003). Additionally, some methods (see, e.g., Himberg et al., 2004) allow the analysis of algorithmic stability by grouping similar solutions and measuring the amount of variability. However, the full potential of analyzing the consistency has not yet been exploited.

A novel approach is presented, based on observations made by Ylipaavalniemi and Vigário (2004) and Ylipaavalniemi et al. (2006), offering a better insight to both ICA and the data. Running the FastICA algorithm (Hyvärinen, 1999) multiple times in a bootstrapping manner, followed by a suitable clustering of the

\* Corresponding author.

E-mail addresses: jarkko.ylipaavalniemi@tkk.fi (J. Ylipaavalniemi), ricardo.vigario@tkk.fi (R. Vigário).

URL: <http://www.cis.hut.fi/whyj> (J. Ylipaavalniemi).

Available online on ScienceDirect ([www.sciencedirect.com](http://www.sciencedirect.com)).

estimated solutions, results in an improved set of estimates and allows the nature of the variability to be analyzed. The method easily reveals consistent independent components, but also helps to interpret the less consistent phenomena. The usefulness of the method is further illustrated with functional magnetic resonance imaging (fMRI) data, obtained during speech stimulation.

A brief explanation of ICA is given in the Independent component analysis section, needed to understand the causes of variability considered in the Variability of independent components section. The actual method for analyzing consistency and interpreting the nature of the variability is presented in the Analyzing consistency by exploiting variability section. Related work is discussed in the Related work section, and finally the usefulness of the method is illustrated in the Application to fMRI analysis section.

### Independent component analysis

Before estimating the independent components, the observed data can be whitened, that is, transformed to be uncorrelated and have unit variances. Whitening can be done using a linear transformation and does not constrain the estimation in any way, since independence implies uncorrelation. Additionally, whitening simplifies the following component estimation by restricting the structure of the mixing. If the whitening is done using principal component analysis (PCA) (Jolliffe, 2002), the number of free parameters can be reduced by taking only the  $K$  strongest principal components, and leaving out the weakest. Assuming that the weakest components contain mainly noise, the dimension of the data is reduced in an optimal manner to improve the signal-to-noise ratio.

As mentioned before, ICA is based only on the assumption that the source signals are statistically independent. This seems reasonable in many applications and in fact does not have to hold exactly for ICA to be applicable in practice. The generative model used in ICA is an instantaneous linear mixture of random variables. The original signals can be considered as a source matrix  $\mathbf{S}$ , where each row of the matrix contains one of the  $K$  signals. Respectively, the observed mixed signals are denoted as matrix  $\mathbf{X}$ . Again, each row of the matrix contains one of the  $N$  observed signals. Assuming a noiseless environment, the mixing model can be expressed in matrix form as:

$$\mathbf{X} = \mathbf{A}\mathbf{S}. \quad (1)$$

Each column  $a_k$  of the full rank  $N \times K$  mixing matrix  $\mathbf{A}$  holds the mixing weights corresponding to source  $k$ . The problem of jointly solving both the mixing and the original sources is not only considerably difficult, but also ambiguous. Since both  $\mathbf{A}$  and  $\mathbf{S}$  are unknown, it immediately follows that the signs and scaling of the sources cannot be uniquely defined. One can multiply the mixing weights  $a_k$  and divide the corresponding source respectively with any given coefficient. Additionally, the sources can appear in any order. Fortunately, the ambiguities in the model are not so crucial. For example, the sign and scaling of the independent components can often be fixed with simple normalization.

The remainder of this paper focuses mainly on the use of the FastICA algorithm (FastICA, 1998; Hyvärinen, 1999; Hyvärinen and Oja, 2000), but the considerations should be relatively easy to extrapolate to other algorithmic implementations of ICA. FastICA is very fast and robust, even with big data sets and under somewhat noisy conditions. FastICA uses a fixed-point optimization scheme

based on Newton iteration and an objective function related to negentropy. The basic idea is to search for a solution in the form  $\mathbf{S} = \mathbf{W}\mathbf{X}$ , where  $\mathbf{W} = \mathbf{A}^{-1}$ .

### Variability of independent components

An intuitive explanation of ICA is offered by the central limit theorem. Basically, it states that the distribution of a mixture of independent and identically distributed (i.i.d.) random variables tends to be more Gaussian than the original ones. Therefore, as the sources are made more non-Gaussian they must also become more unmixed. Commonly used measures of non-Gaussianity are skewness and kurtosis, i.e., the third and fourth order cumulants.

#### Theoretical considerations

Statistical independence means that the joint probability density function (pdf) of the sources can be factored in its marginal pdfs. A closed form solution to the ICA problem is not possible since the exact pdfs are unknown. The sources have to be estimated based on concepts such as mutual information and negentropy. It can be shown (see, e.g., Hyvärinen and Oja, 2000) that the resulting objective functions essentially measure non-Gaussianity of the sources.

The ICA algorithms are usually derived under strict assumptions, such as, statistical independence, stationarity and noiseless data. With real data the assumptions often fail and the behavior of the estimation cannot be guaranteed, even if the estimation would seem robust. The measures commonly used in the optimization, like kurtosis, give relatively high scores for sparse sources, whether or not they are truly independent. Therefore, independence and sparseness are strongly connected (see, e.g., Li et al., 2004), which can result in a natural tendency of any ICA algorithm to bias towards sparse, rather than strictly independent, solutions. Indeed, when the degrees of freedom in ICA are too high, the model is likely to overfit the data, which is known to result in strong sparse contamination of the estimated components (Säreälä and Vigário, 2003). Additionally, the overfitting contamination can change freely on each application of the algorithm. On the other hand, a too low model order can result in poor separation altogether. Whitening, and suitable dimension reduction, is important in controlling the model order, or the size of the decomposition.

The algorithm itself may also be a source of variability. The selected optimization scheme and implemented iteration are usually tuned for fast and robust convergence, which can affect the estimation accuracy of the algorithm. Also, since the ICA model is usually noiseless, the performance of the algorithm in the presence of noise depends on the implementation. Particularly, FastICA can search for the independent components one at a time, in *deflation* mode, or all at once, in *symmetric* mode. The effect is, respectively, either accumulating all the inaccuracies to the last components or spreading them evenly among all components. The sensitivity to outliers in the data can also be tuned somewhat by choosing the nonlinearity associated with the contrast function. Additionally, the initial conditions of the algorithm are usually selected at random.

#### Practical considerations

Arguably, the good performance of ICA in practice, even under ill conditions, can be in part attributed to the fact that the sparse

components often form a very natural and easily interpreted decomposition of the data (cf. McKeown and Sejnowski, 1998; Vigário et al., 2000). In some cases, the best goal may even be to search for sparse, rather than independent components. The problem is that the natural and stable biasing towards sparse components can easily be confused with the sparseness produced by overfitting. Thus, it is crucial to analyze the presence and nature of the variability.

All the optimization is performed in an optimization landscape that is partly defined by the data itself. Therefore, the distribution and structure of the data will also affect, and possibly further bias, the convergence of the algorithm. Perhaps the simplest example of such effects is the increase in the number of local optima when the signal-to-noise ratio of the data is low. In addition to solving the reliability problems when using ICA, the analysis of consistency and nature of the potential variability may actually reveal more about the properties of the underlying data. For example, components that are clearly separated from other components but contain a high degree of variability could characterize a subspace phenomenon. On the other hand, consistent components with very weak signal-to-noise ratios could be reliably discriminated from overfitting.

### Analyzing consistency by exploiting variability

The variability can be considered on a surface defined by the objective function and the given data. An illustrative example of a 3-dimensional optimization landscape is shown in Fig. 1, where in addition to the global minimum, the surface presents several local minima and some noise. Different starting points and directions, and the tendency to bias towards a certain solution, cause the optimization to converge along different paths, shown as thick curves. Ideally, a robust algorithm should always reach the same true optimum, but in practice it can get stuck on a local minimum and even the optimal solutions reached along different paths can vary slightly, depending on the implementation of the algorithm. The estimates should still form consistent groups with high

similarity. Thus, one may expect that the true solution can be found as the mean of a consistent group and that the spread of the group can be used to analyze the variability of that solution. The analysis is based on applying FastICA multiple times in an efficient and controlled manner.

### Randomizing initial conditions and resampling data

Randomizing the initial conditions changes the starting point and direction of optimization. This allows converging to the optimum solution by approaching it from different directions. Actually, some of the directions may allow a more optimal convergence than others and optimal points surrounded by local minima, defined by properties of the data or noise, may only be reachable from certain directions. Starting with different initial conditions is important for complete analysis since it helps detecting even the weakest local minima. This can be crucial with poor signal-to-noise ratio, or when the data are such that it causes strong bias or overfitting tendency.

Resampling the data with reposition (Meinecke et al., 2002), i.e., picking a set of samples from the original data in random order, where the size of the resulting set can be different from the original, and each sample can potentially appear many times, reveals statistical properties of the estimation. The process slightly changes the optimization landscape, which is partly defined by the resampled data. The strong and stable optima define the true decomposition of the data, while the changing shape of local minima and noise leads to their discarding. Resampling is also required to fully characterize the variability since it allows distinguishing between biasing and overfitting. Both can produce the tendency towards, e.g., sparse rather than independent components, but only the biased estimates are consistent and natural components of the data.

The idea is further illustrated in Fig. 2, using a 2-dimensional optimization surface, shown as thin curves. The resampling causes the shape of the surface, thick curves, to change slightly and the algorithm converges to different individual solutions, marked as “o”. Some of the solutions can fall on local minima, but the mean

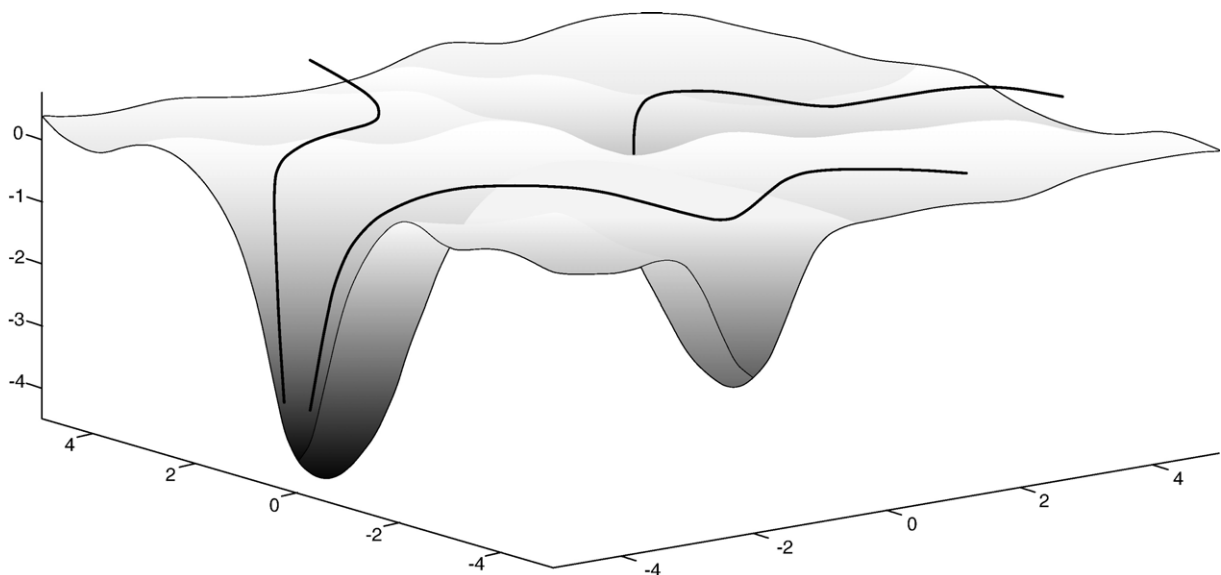


Fig. 1. Estimation landscape. Different initial conditions and data properties cause the algorithm to converge along different paths (shown as curves) on the 3-dimensional optimization surface. Thus, the algorithm may get stuck on a local minimum and even the optimal estimates may be slightly different.

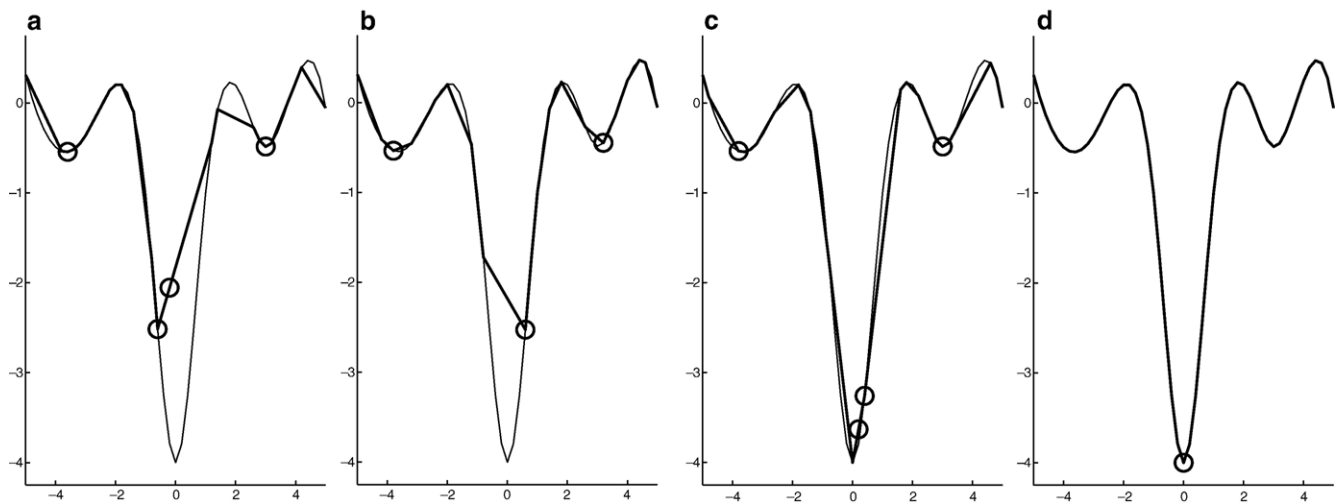


Fig. 2. The idea of resampling. In panels a–c, the data are resampled (shown as thick curves) to slightly change the shape of the true 2-dimensional optimization surface (shown as thin curves). This allows the algorithm to converge to different solutions (marked as “O”), which may fall on local minima. The best solution is given by the mean of all solutions, in panel d.

of all solutions gives a very good estimate of the true optimum, as shown in Fig. 2d.

#### Controlling instead of constraining

Performing ICA many times under resampling is actually rather difficult in practice, e.g., due to the sign, scaling and permutation ambiguities. Some earlier methods (cf. Meinecke et al., 2002) use a simple trick to make it faster and easier. They perform an initial single run and the consecutive runs are applied to resampled versions of the estimated independent components from the first run. Ideally, the consecutive applications of ICA should result in an identity mixing matrix. The estimation takes place in the local vicinity of the initial solution and should converge quickly with minimal effects on, e.g., the sign and permutation of the components. However, with real data including noise and other artifacts such approaches can be very constrained, since the estimates become only slightly altered versions of the initial components. The biggest drawback is that the initial run essentially defines the set of independent components to be analyzed, and a poor solution on the initial run will render the consecutive runs practically useless. Also, the variability is heavily restricted, so there is a danger that the optimization landscape may not be covered sufficiently and the original assumptions on consistency could be violated.

In the presented method, a separate initial run is not used, allowing for any permutation, and indeed, any combination of independent components to be found on each run. The benefits are that ICA has more freedom to find the optimal solution on each run and the solutions can include local minima that are very difficult to find on all runs. Clearly, this leads to a more difficult combination problem. The FastICA algorithm is well suited for such use, but there is a theoretically justified way of making it even faster. The amount of data used on each run can be reduced by resampling the original data into smaller subsets, i.e., picking less than the total number of samples available. In such approach, a too strong reduction can eventually cause an increase in the variance of the estimates, but it should not affect the bias. Although the behavior is

easy to confirm in practice, the amount of reduction should be controlled carefully since suitable values most likely depend on the data.

As mentioned before, other considerations include the definition of the degree of dimension reduction typically performed during the whitening stage, as well as the number of independent components estimated from that reduced space. Whitening dimension reduction has the practical effect of removing redundant and noisy information from the input space, while helping preventing the ICA algorithm from overfitting (see, e.g., Särelä and Vigário, 2003). It is common practice to estimate as many independent components as the number of retained whitened signals. In the presented method, the number of components is smaller. The rationale behind the decision can be explained from a deflation versus symmetric algorithmic reasoning. In the latter, any deviation from the assumed data model is distributed throughout all the estimated components, often resulting in less overfitting and faster convergence. On the other hand, in the deflation approach, the estimated components are affected by the accumulated errors of all previously estimated components. Resulting in higher overfitting tendency of the first components and larger deviation from the assumed data model among the last ones. Therefore, the choice of estimating a smaller set of independent components than the total number of whitened signals allows for a pool of dimensions to better account for data modeling errors. In the multiple runs of the algorithm, different sets of independent components can be estimated, resulting in a total number of consistent components that may well exceed that of the single runs.

#### Clustering estimated independent components

The following presents an efficient and practical clustering approach, which naturally takes into account the ambiguities and convergence behavior of ICA discussed above. The method only requires two parameters and is not very sensitive to their values. The estimated components from the multiple runs are assumed to be represented as a concatenation of the respective mixing matrices. The information needed for clustering could also be

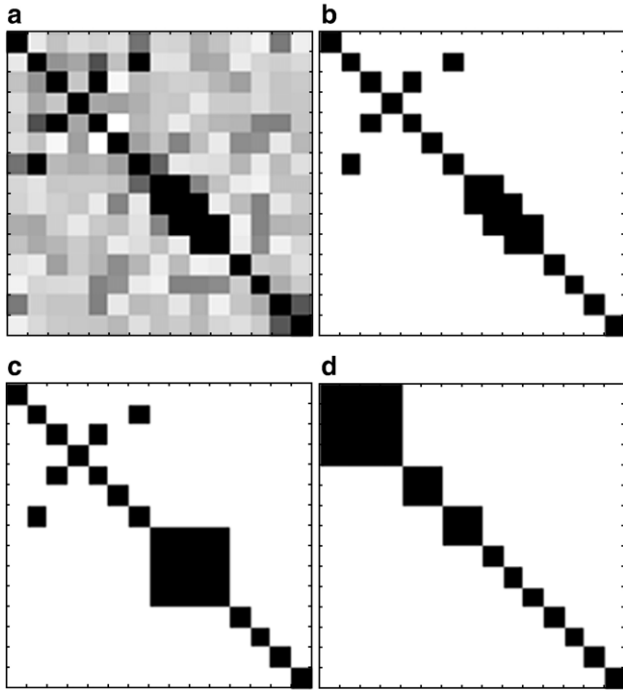


Fig. 3. Phases of correlation calculations. Grayscale matrix images with white being 0 and black 1 of the succession from absolute correlation values  $|C|$ , in panel a, through thresholded binary correlations  $\tilde{C}$ , panel b, into relations  $R$ , panel c, after raising to a power. In panel d, the final relations reordered to show the separate groups more clearly.

gathered from the actual sources, but the amount of data would be far greater. Denoting the total number of runs by  $B$ , and by  $A^b$  the mixing matrix from run  $b$ , as defined in Eq. (1), the concatenated matrix is defined as:

$$\bar{A} = [A^1 | A^2 | \dots | A^B]. \quad (2)$$

Additionally, it is assumed that the columns  $\bar{a}$  of matrix  $\bar{A}$  are normalized, e.g., by making their mean zero and variance one, which still leaves the sign and permutation ambiguities. An efficient way to measure the similarity of the normalized estimates, that is easy to understand in both the mixing and signal spaces, while being relatively fast to compute, is through correlation:

$$C \propto \bar{A}^T \bar{A}. \quad (3)$$

Considering that a set of estimates presenting a correlation over a given threshold value  $\epsilon$  correspond to a common underlying source leads to a binary correlation matrix  $\tilde{C}$ , whose elements  $\tilde{c}_{ij}$  mark whether the two corresponding estimates  $\bar{a}_i$  and  $\bar{a}_j$  are related. Denoting the elements of matrix  $C$  by  $c_{ij}$ , the binary values of the elements of  $\tilde{C}$  are defined as:

$$\tilde{c}_{ij} = \begin{cases} 1, & \text{if } |c_{ij}| > \epsilon \\ 0, & \text{if } |c_{ij}| \leq \epsilon \end{cases}. \quad (4)$$

The absolute values are needed because of the sign ambiguity. Furthermore, if the matrix  $\tilde{C}$  shows that estimate  $\bar{a}_i$  is related to  $\bar{a}_j$ , and that  $\bar{a}_j$  is to  $\bar{a}_k$ , this is a strong suggestion that  $\bar{a}_i$  and  $\bar{a}_k$  are also related through a longer path. These somewhat weaker and discontinuous paths can also be considered as estimates of a common underlying source, even if  $\bar{a}_i$  and  $\bar{a}_k$  would directly

exhibit a lower correlation than the threshold value  $\epsilon$ . To archive this, the matrix  $\tilde{C}$  can be raised to a suitable power, whose value  $p$  prelates to the length of the longest acceptable path:

$$R = \tilde{C}^p. \quad (5)$$

The relations matrix  $R$ , whose elements  $r_{ij}$  mark whether the two corresponding estimates  $\bar{a}_i$  and  $\bar{a}_j$  are linked, can be considered as a binary matrix with all non-zero elements defining a link. This would allow the rather heavy computation involved in the matrix-power calculation to be implemented very efficiently. The different phases of the correlation calculations are illustrated in Fig. 3. It shows the succession from absolute correlation values  $|C|$  through thresholded binary correlations  $\tilde{C}$  into relations  $R$  after raising to a power. Finally, a reordered version of the relations matrix is shown, where groups are made more evident. The element values are shown as grayscale images with white being 0 and black 1. In particular, note how raising the matrix  $\tilde{C}$  to a power makes the groups in the data more visible.

The correlations  $C$ , from Eq. (3), and their binary relations  $R$ , from Eq. (5), make it possible to cluster the estimated components  $\bar{A}$ , from Eq. (2), into consistent groups. The clustering is done with a simple and efficient algorithm, requiring no additional parameters. It is based on the fact that each element  $c_{ij}$  and  $r_{ij}$ , of matrices  $C$  and  $R$ , respectively, link two estimated components  $\bar{a}_i$  and  $\bar{a}_j$ . It consists of two phases:

- (1) All elements that are related, that is, have a value  $r_{ij}$  of 1, are sorted into a list in descending order of the corresponding absolute correlation value  $|c_{ij}|$ . The resulting list will define the links between all component pairs in order from the most similar to the least similar pair.
- (2) Then, the clusters are formed by going through this list in order until all components have been clustered by making the following decisions for each link:
  - If neither of the linked components  $\bar{a}_i$  and  $\bar{a}_j$  belongs to an existing cluster, create a new cluster containing both components.
  - If only one of the linked components  $\bar{a}_i$  and  $\bar{a}_j$  belongs to an existing cluster, add the other component to that cluster as well.
  - If both linked components  $\bar{a}_i$  and  $\bar{a}_j$  already belong to different clusters, do nothing.

The clustering resembles the complete linkage approach in hierarchical clustering, with the addition that the clusters are growing from the most dense parts of the space towards the more empty parts. Additionally, the maximum size of a cluster is not restricted, so a cluster can have several members even from the same run. This improves the idea of complete linkage in the case of a subspace that would be spanned by more than one component, by clustering all components of the subspace together.

The number of clusters is not restricted either and together with the last decision, of doing nothing when both components in a pair have already been clustered to different clusters, potentially results in a number of left-out components in the end. These can be safely discarded from further analysis, since they are outliers resulting from overfitting. The last decision could, in theory, be used to gather a hierarchical tree of the clusters. One way of achieving this would be to count the number, or accumulated correlation, of these weakest links between the clusters and then form the tree based on

those values. However, such extensions to the method have not been studied.

#### *Nature of variability*

Consistent components are easy to identify, e.g., by the number of occurrences. In first approximation, the mean, or centroid, of each cluster can be considered as the best estimate of the true decomposition given by the clusters. The mean representatives can also depart somewhat from strict independence. In cases where independence is crucial, the estimates from a single run closest to the mean of the respective cluster can be picked instead. However, as discussed in Variability of independent components section, and shown by Ylipaavalniemi and Vigário (2004), the nature of the variability can be very different for each cluster and using the mean should produce a more robust decomposition of the data.

To characterize the differences among the components, measures of compactness and discrimination can be calculated. The clustering method does not restrict the number of estimates in each cluster to be equal, so even the estimate count can reveal whether the component is easy or difficult for ICA to separate. Since the order of components is not fixed in ICA, the properties can be combined into a measure used for ranking the components. Thus, significant components can be shown first, and similar phenomena can be shown close to each other. In some cases, it is possible to describe features of the component signals instead of the clustering. The variability can be characterized by analyzing it in all of the similarity, mixing and signal spaces. The additional information can be used to identify the underlying phenomena, e.g., to detect artifacts.

Fig. 4 shows a way to visualize such measures. The discrimination rings show how consistent a component is, and how separated it is from all the other components. Intuitively, the intra-cluster ring (left) should be a compact dot and the inter-cluster ring (right) a narrow band with a hole big enough to fit the intra-cluster ring. In some cases, few outliers can increase the radius of the intra-cluster ring, but the mean value should remain relatively small. The spread of the mixing vectors shows whether the variability is distributed evenly in all dimensions, or if certain subspaces exhibit more variation.

#### **Related work**

There has been an exhaustive amount of research in resampling and multiple-run-based approaches. Several view points have been considered, but most often the goal has been to improve the performance of an existing algorithm by forming ensembles, weighted combinations, or simply by picking the best, out of many individual instantiations. There exist many other uses of resampling approaches in machine learning, but the following are the most relevant ones.

#### *Boosting, bagging and bootstrap*

The approach of resampling observed data into many individual training sets is called either boosting or bagging, depending on the context, like neural networks or decision trees (see, e.g., Freund and Schapire, 1996; Breiman, 1996). Originally, the idea of resampling the input data was proposed in supervised learning, e.g., for classification, to allow combining many weakly performing units to an ensemble that outperforms even the best performing single units. However, the theories are very general and many

algorithms have been developed. The theoretical connections and explanations for the increased performance have been studied recently (see, e.g., Breiman, 2001; Andonova et al., 2002), and the process is still ongoing.

In unsupervised learning, resampling with replacement is usually called bootstrapping. The main idea is to cover the whole distribution of an estimator with the independent samples (Efron and Tibshirani, 1994). Most commonly the goal has been to analyze the stability of an algorithm or to select the best performing model (see, e.g., Rao and Tibshirani, 1997), but despite the somewhat different viewpoint the theory is basically the same as in boosting and bagging (see, e.g., Bauer and Kohavi, 1999). Resampling the available data to discover the whole distribution of an estimator, while retaining the structure within the data (see, e.g., Breakspear et al., 2004), also has strong connections to Bayesian theory (see, e.g., Rubin, 1981; Clyde and Lee, 2001). In other sections of the paper, the term bootstrapping is used in a broader sense, generally referring to any of the boosting, bagging or bootstrap concepts.

#### *Consistency and clustering approaches*

The majority of previous work on resampling in machine learning has been done in connection with supervised algorithms, where the performance is easy to verify (Strother et al., 2002). The behavior of unsupervised algorithms, like ICA, under re-sampling is more difficult to quantify (Baumgartner et al., 2000). The effects of increasing the level of noise in the data on the behavior of ICA and consistency of the independent components have also been studied (Harmeling et al., 2003, 2004). However, essentially lowering the signal-to-noise ratio is based on very different assumptions and has very different effects on the variability than, e.g., resampling the

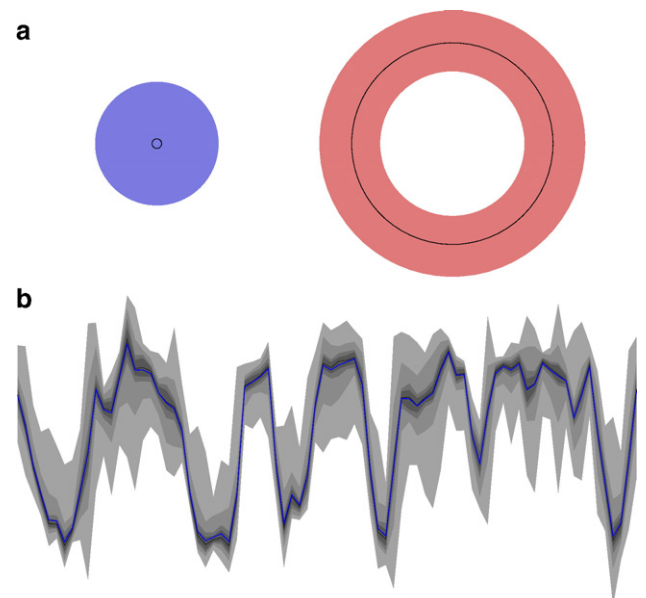


Fig. 4. Illustration of consistency measures. In panel a, rings visualizing cluster discrimination. The spread of Euclidean distances between members of a cluster (left blue ring) and between all the other clusters (right red ring). The black circles mark the mean distances, respectively. In panel b, spread of the variability of mixing vectors. The mean mixing vector (solid blue line) surrounded by bands of different gray shades representing quantiles of the mixing values. The darkest shade corresponding to the 0.95 quantile and the lightest to 0.05. See the online version of this article for the color figure.

data. In terms of boosting, bagging and bootstrap literatures, increasing the noise should affect the variance of the underlying algorithm, whereas suitable resampling has more to do with its bias.

The combination strategy with ICA has to also take into account the sign, scale and order ambiguities in the model. Even the optimal number of components is usually not known. Some simple combination strategies, or validation of manual combinations, have been studied in connection with the producibility of ICA decompositions (see, e.g., Duann et al., 2003; Ito et al., 2003; Stögbauer et al., 2004). The combination approach explained in the Clustering estimated independent components section is rather simple and allows easy interpretation of the results, but it has strong connections to recently developed clustering methods (see, e.g., Meilá and Shi, 2000, 2001) based on the eigenstructure of a similarity matrix. There has even been research on the stability of clustering solutions on themselves (see, e.g., Lange et al., 2004).

Similar approaches as the one presented in this paper have been recently developed, partly in collaboration with the authors of this paper. However, those methods have very different goals, and therefore the selected strategies are also somewhat different.

The approach proposed by Meinecke et al. (2002) relies on a single initial decomposition, as any single run ICA approach does, and aims at validating the consistency of those independent components. Thus, allowing the analysis of the robustness of the ICA method, or the removal of severely overfitted components. On the other hand, Himberg et al. (2004) have proposed a visualization method (Icasso, 2003), based on clustering components from multiple runs. However, the method is computationally very heavy and leaves the user with a non-linear two-dimensional cluster representation, which itself is unstable and changes every time the method is applied. Thus, the visualization can be difficult or even misleading to interpret.

In addition to avoiding some of the previous problems, the approach in this paper provides additional intuition and motivation for the use of any similar resampling method in connection with ICA.

**Application to fMRI analysis**

The usefulness of the method was tested with data from a functional magnetic resonance imaging (fMRI) study. The data were

collected from 10 healthy adults while they were listening to spoken safety instructions in 30 s intervals, interleaved with 30 s resting periods. All the data were acquired at the Advanced Magnetic Imaging Centre of Helsinki University of Technology, using a 3.0 Tesla MRI scanner (Signa EXCITE 3.0 T; GE Healthcare, Chalfont St. Giles, UK) with a quadrature birdcage head coil, and using Gradient Echo (GRE) Echo Planar Imaging (EPI) (TR 3 s, TE 32 ms, 96 × 96 matrix, FOV 20 cm, slice thickness 3 mm, 37 axial slices, 80 time points (excl. 4 first ones), flip angle 90°).

*Spatial ICA of fMRI data*

Spatial ICA was first applied to fMRI data by McKeown et al. (1998) and McKeown and Sejnowski (1998) and has since been used in many studies (for recent reviews, see, e.g., Calhoun et al. (2003); McKeown et al. (2003)). To be able to use the fMRI signal in the standard ICA model presented in Eq. (1), each scanned volume must be transformed into vectorized form, in a bidirectional manner. Since ICA is only concerned with the statistics of the observations and not the order of samples in them, the voxels in the volumes can be freely reordered into vectors (cf. Calhoun et al., 2003). Naturally, all volumes must be transformed using the same reordering. The ICA model used in fMRI analysis is illustrated in Fig. 5.

The fMRI signal is represented by a  $t \times v$  data matrix  $\mathbf{X}$ , where  $t$  is the number of time points and  $v$  is the number of voxels in the volumes. Thus, each row of  $\mathbf{S}$  contains one of the  $k$  independent spatial patterns and the corresponding column of  $\mathbf{A}$  holds its activation time course. The statistical constraint applies to the spatial domain. Contrary to the commonly used fMRI analysis methods (cf. Worsley and Friston, 1995; SPM, 1999), based on general linear model, no assumptions are made on the shape of the activation time courses. They are formed in the demixing process, in a purely data-driven way.

*Consistency analysis*

Before ICA analysis, the recorded volumes were preprocessed with the SPM software (SPM, 1999). The processing included slice time correction, realigning for motion compensation and spatial smoothing (Gaussian 8 mm FWHM). The analysis was limited to voxels inside the skull, using anatomical images. The observation

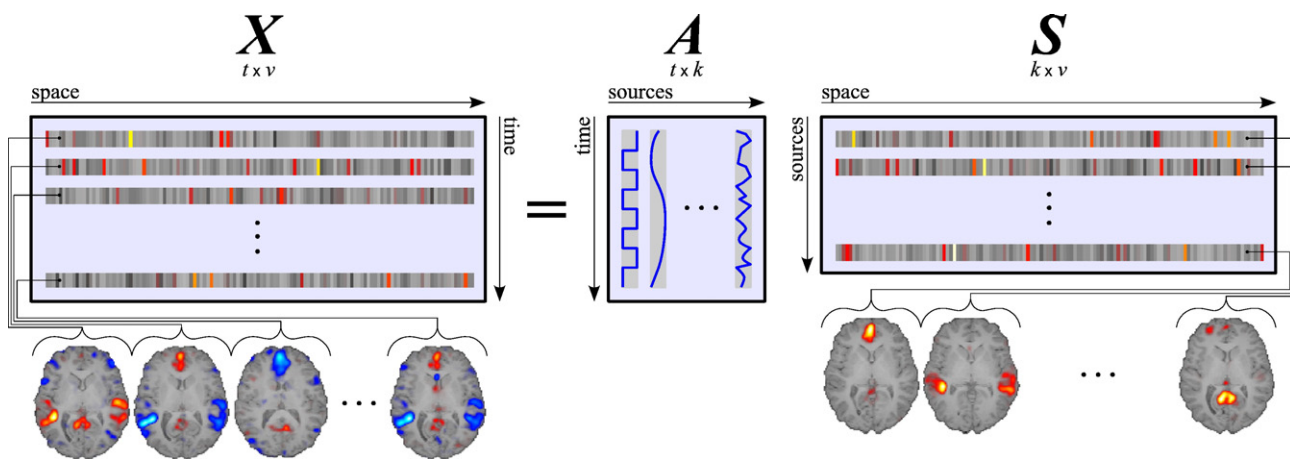


Fig. 5. Spatial ICA of fMRI data. The rows of the data matrix  $\mathbf{X}$  and source matrix  $\mathbf{S}$  are vectorized volumes. The corresponding columns of the mixing matrix  $\mathbf{A}$  are the time courses. Note, that the statistical independence applies to the volumes.

matrix was whitened using PCA, and the dimension reduced to the 30 strongest principal components, resulting in a  $30 \times 254484$  data matrix.

The assumptions on the magnitude of the variability and the effects of the data resampling were tested first. Fig. 6 shows the learning curves depicting the magnitude of the variability with the fraction of samples ranging from 1% to 100%, in steps of 1%. For each percentage value, ICA was run 50 times and a set of 5 mean estimates was created using the presented clustering procedure. The sets were compared with estimates produced without resampling. The distance of the estimates was defined as the Frobenius norm of the difference of the mixing matrices. Whether due to variance or bias, the overall variability only increases at very low sampling values. Therefore, the fraction of samples used in practice can be quite small, e.g., between 10% and 20%. The results also suggest that the fMRI data contains a lot of redundancy for estimating independent components.

Finally, the data of each subject were analyzed separately by running FastICA 200 times using a resampling of 10% of the samples, and estimating 15 independent components in each run (symmetric mode with *tanh* nonlinearity). The resulting 3000 component estimates were then clustered, with a correlation threshold of 0.8 and raising to power 8. All parameter values were chosen heuristically, based on earlier testing. Increasing the model complexity produces only more overfits, which are left out in the clustering.

### Experiment results

Fig. 7 shows the results for a single subject. Separate high-resolution color figures with the complete results for all 10 subjects are available online (see Appendix A). The following highlights the different types of components found.

Component 1 is extremely consistent, because the intra-cluster ring is reduced to a dot and the time course shows minimal variability. Also, the estimate count is 200. The almost linear time course suggests that the component represents a scanner-related drifting artifact. Likewise, components 5 and 6 are very consistent. Component 5 corresponds to activation of auditory cortex on the superior temporal lobe, which is strongly related to stimulation. However, component 6 is not related to stimulation, but represents

activity along the anterior and parietal cingulate gyrus. Components of such types are also consistent with earlier studies (see, e.g., McKeown and Sejnowski, 1998; Calhoun et al., 2001, 2003; van de Ven et al., 2004), showing that ICA can find independent components representing weakly, transiently or strongly stimulus-related brain activity.

On the other hand, component 12 shows stimulus-related activity on the lateral parietal lobe, but both the intra-cluster ring and the spread of the time course reveal considerable variability. The mean intra-cluster distance is still very small, so the spread is due to only a few outliers. However, the inter-cluster distances are relatively small and the estimate count 91 is lower than 200. The component estimation appears consistent, but the component is very difficult to separate from other components, causing the lower probability of detecting it with ICA.

The estimation of component 15 is less consistent than in the previous examples, shown by the large mean intra-cluster distance. The hole in the middle of the ring means that even the two closest estimates are quite far apart. However, the inter-cluster distances are still consistently larger, showing a clear separation from other components. Additionally, the estimate count 280 means that the component is a combination of more than one estimate. The cluster characteristics reveal that ICA can separate an independent subspace, but cannot consistently identify a single component within the subspace.

Fig. 8 shows one component consistently found on all 10 subjects. The matching component of each subject was selected manually but was always among the most consistent, e.g., component 5 in Fig. 7. The normalized activation time courses are strongly correlated and stimulus related. The spatial locations on the superior temporal lobes are also reproduced, with some differences due to individual brain anatomy.

Similarly, Fig. 9 shows a component that appears consistently for most of the 10 subjects. The matching component of each subject was selected manually. For a single subject, a matching component was not found, and for two subjects the different spatial extent of the activation cannot be explained by differences in brain anatomy. The spatial activation pattern along the anterior and parietal cingulate gyrus is consistently reproduced for the remaining 7 subjects. The activation is not related to stimulation and the time courses are different for each subject (not shown).

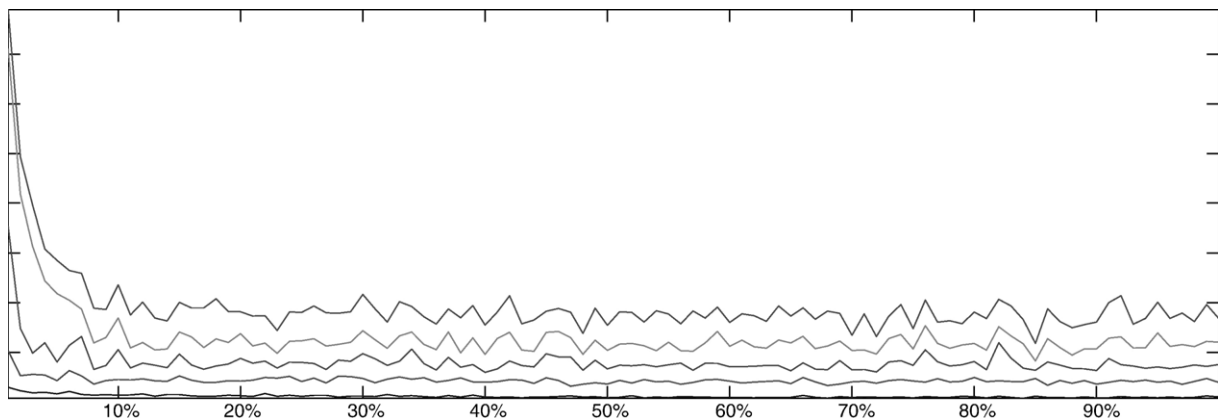


Fig. 6. Controlling variability by resampling. The cumulative mean magnitude of variability, for 5 independent components measured between estimates using a given percentage of samples and estimates using all of the samples. The bottom curve is the magnitude of the first component. The second curve is the cumulative magnitude of the first two components, and so on.

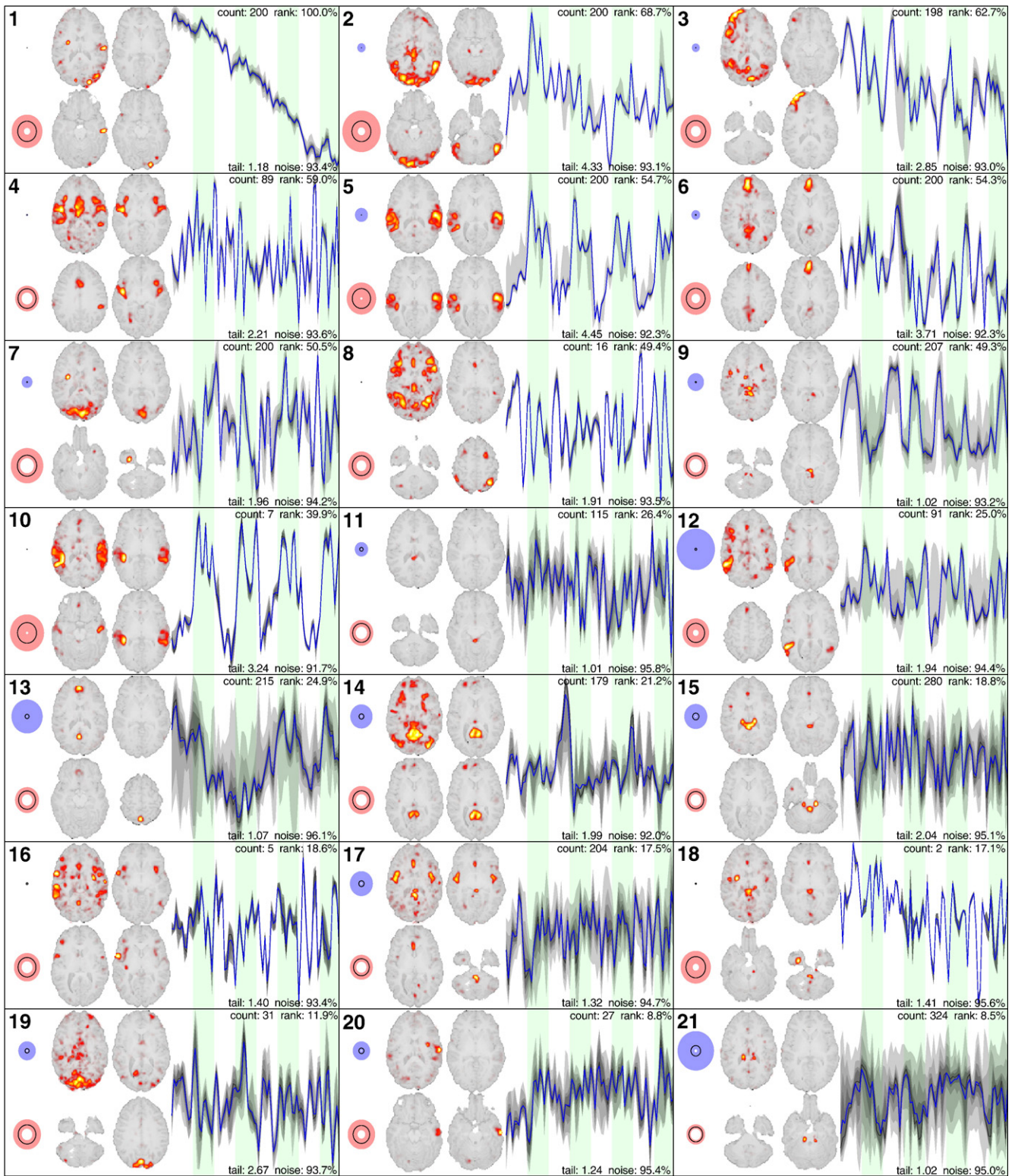


Fig. 7. Highlights of clustered independent components. Each component is shown with consistency information. The upper and lower rings on the left show the spread of intra- and inter-cluster distances, respectively. The black circles depict mean distances. Four axial slices of the brain are shown, with the overlaid activation patterns representing, clockwise from top left: the sum through all slices; the slice with maximum activation power; the slice containing the highest valued voxel; and the slice containing the lowest valued voxel. Finally, the activation time course is shown on top of the stimulation blocks. The gray scale bands surrounding the mean time course depict quantiles of the spread ranging from 0.05 to 0.95. Textual values related to consistency and non-Gaussianity are also shown: the count is the number of consistent estimates; the rank is a normalized measure of overall consistency, used for sorting; the tail measures the skewness of the component distribution as a ratio between the positive and negative tails; and the noise measures the energy in the symmetric main lobe.

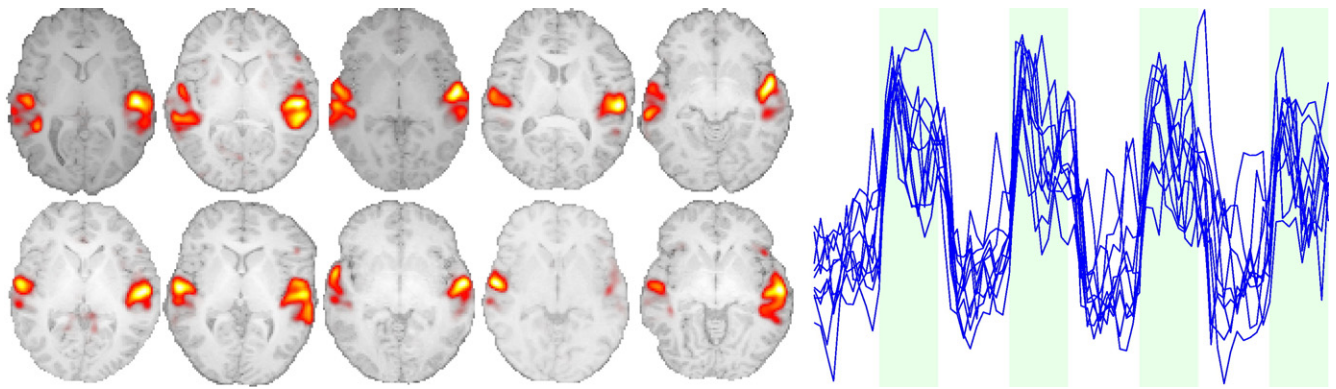


Fig. 8. Consistency across all 10 subjects. Axial slices containing the highest valued voxel (left), from the matching auditory activation-related component of each subject. Superposition of the corresponding normalized activation time courses (right).

In addition to the examples illustrated in Figs. 8 and 9, several other components appear consistently among the 10 subjects. Many of the components are not reproduced in all 10 subjects, or the spatial extents of the activation vary individually, making the comparison more complicated.

Previously, the nature of the variability was mainly considered in the time courses. The analysis can also be carried out in the spatial domain, and Fig. 10 shows how two consistent components can be linked by spatial variability. Component 5 in Fig. 7 is very consistent, but the spatial standard deviation, shown in Fig. 10b, reveals a significant spatial variability. Since the highest variability does not coincide with the highest mean values, the focal points of activity are consistent. However, the spatial variability strongly correlates with component 12. The consistency information and estimate count 91 of component 12 reveal that it could be a weaker local minimum, or relatively close

to component 5, causing ICA to more probably converge to the solution given by component 5.

A similar linking exists between components 5 and 10. The components are very much alike. However, component 5 is more focal and has a lower correlation with the stimulation. The estimate count 7 of component 10 is also very low. Component 10 could represent more accurately the truly independent activity, but due to structure of the data, ICA tends to bias towards the more sparse solution of component 5.

## Discussion

The paper presents a method for analyzing the consistency of independent components. Its usefulness is illustrated by analyzing fMRI data. The method works by exploiting the variability of ICA

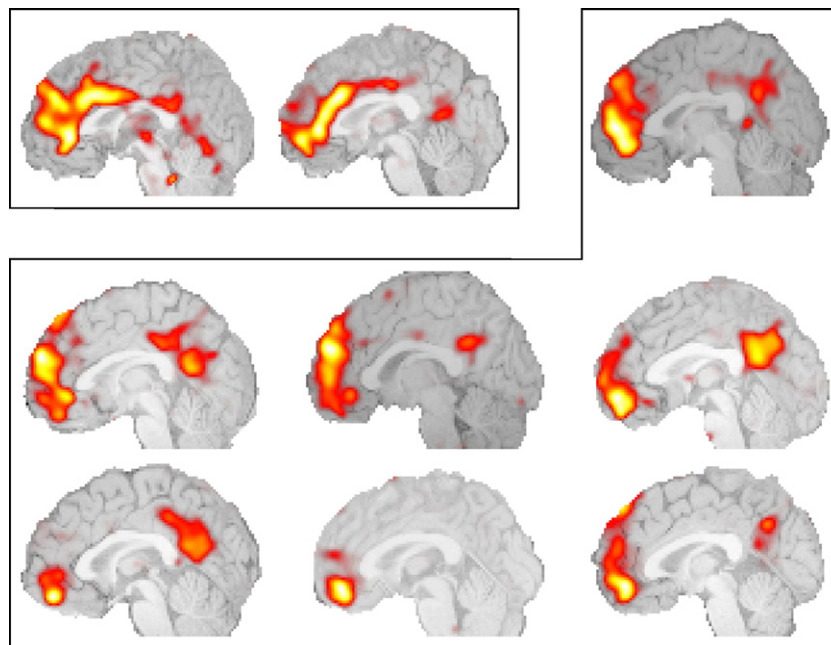


Fig. 9. Consistency across some of the subjects. Sagittal slices containing the highest valued voxel, from the matching component of each subject. The activity is located along the anterior and parietal cingulate gyrus. Two of the subjects are shown separated due to the somewhat different spatial extent of activation.

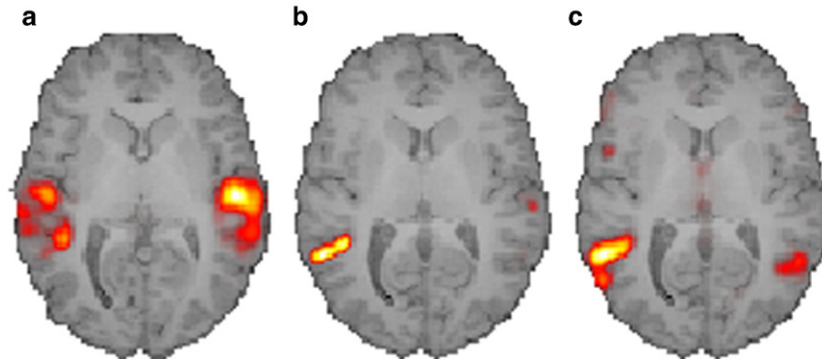


Fig. 10. Spatial variability linking consistent components. Axial slices of component 5 showing the mean, in panel a, and the standard deviation, panel b, of the estimates. In panel c, slice from the identical location of component 12, showing the mean of the estimates.

algorithms in a resampling and clustering based approach. Although the method was applied using the FastICA algorithm, it should be relatively straightforward to implement using other ICA and BSS algorithms. Unlike previous methods, the method is computationally efficient and aims at providing additional information for better interpretation of the decomposition.

The method can identify consistent components and neglect overfits. Several consistent components appear to be reproducible also across many subjects. Additionally, the method can reliably identify components that contain some variability, or are difficult to separate based on statistical independence alone, regardless of whether the component represents stimulus-related activation or not. Such abilities are crucial in situations where hypothesis-driven methods cannot be used.

Activity subspaces are also identified, when ICA cannot consistently converge to a single component estimate. Closely related estimates are combined as one component and several components can be linked through their variability. The underlying subspace should then be further characterized by using other criteria than independence, possibly applied to the internal variability.

The proposed approach has been successfully used in a recent study (Ylipaavalniemi and Vigário, 2007) to provide additional insight into an fMRI data set, not otherwise attainable.

### Acknowledgments

We thank Riitta Hari, Tuukka Raij and Mika Seppä from the Advanced Magnetic Imaging Centre (AMI Centre) and the Brain Research Unit in the Low Temperature Laboratory of Helsinki University of Technology for the fMRI data and invaluable discussions on the subject of this paper.

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.neuroimage.2007.08.027](https://doi.org/10.1016/j.neuroimage.2007.08.027).

### References

Amari, S.-I., Cichocki, A., Yang, H.H., 1995. A new learning algorithm for blind signal separation. *Proceedings of the 1995 Conference on Advances in Neural Information Processing Systems (NIPS 1995)*. Vol. 8. Denver, CO, pp. 757–763. November.

- Andonova, S., Elisseeff, A., Evgeniou, T., Pontil, M., 2002. A simple algorithm for learning stable machines. *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI 2002)*. Vol. 1. Lyon, France, pp. 513–517. July.
- Bauer, E., Kohavi, R., 1999. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. *Mach. Learn.* 36 (1–2), 105–139 (July).
- Baumgartner, R., Somorjai, R., Summers, R., Richter, W., Ryner, L., Jarmasz, M., 2000. Resampling as a cluster validation technique in fMRI. *J. Magn. Reson. Imaging* 11 (2), 228–231 (February).
- Bell, A.J., Sejnowski, T.J., 1995. An information–maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7 (6), 1129–1159 (November).
- Breakspear, M., Brammer, M.J., Bullmore, E.T., Das, P., Williams, L.M., 2004. Spatiotemporal wavelet resampling for functional neuroimaging data. *Hum. Brain Mapp.* 23 (1), 1–25 (September).
- Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24 (2), 123–140 (August).
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32 (October).
- Calhoun, V.D., Adali, T., McGinty, V.B., Pekar, J.J., Watson, T.D., Pearlson, G.D., 2001. fMRI activation in a visual-perception task: network of areas detected using the general linear model and independent components analysis. *NeuroImage* 14 (5), 1080–1088 (November).
- Calhoun, V.D., Adali, T., Hansen, L.K., Larsen, J., Pekar, J.J., 2003. ICA of functional MRI data: an overview. *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003)*. Nara, Japan, pp. 281–288. April.
- Cardoso, J.-F., 1989. Blind identification of independent components with higher-order statistics. *Proceedings of the Workshop on Higher-Order Spectral Analysis*. Vail, CO, pp. 157–162. June.
- Cardoso, J.-F., 1990. Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem. *Proceedings of the 1990 International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1990)*. Vol. 5. Albuquerque, NM, pp. 2655–2658. April.
- Clyde, M.A., Lee, H.K.H., 2001. Bagging and the Bayesian bootstrap. *Proceedings of the 8th International Workshop on Artificial Intelligence and Statistics (AISTATS 2001)*. Key West, Florida. January.
- Comon, P., 1994. Independent component analysis, a new concept? *Signal Process.* 36 (3), 287–314 (April).
- Duann, J.-R., Jung, T.-P., Makeig, S., Sejnowski, T.J., 2003. Consistency of infomax ICA decomposition of functional brain imaging data. *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003)*. Nara, Japan, pp. 289–294. April.
- Efron, B., Tibshirani, R.J., 1994. *An Introduction to the Bootstrap*, 1st edition. Routledge/Taylor Francis Group, Oxford, UK. May.
- FastICA, 1998. MATLAB™ Package. URL <http://www.cis.hut.fi/research/ica/fastica>.

- Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm. Proceedings of the 13th International Conference on Machine Learning 1996 (ICML 1996). Bari, Italy, pp. 148–156. July.
- Funaro, M., Oja, E., Valpola, H., 2003. Independent component analysis for artefact separation in astrophysical images. *Neural Netw.* 16 (3–4), 469–478 (April).
- Harmeling, S., Meinecke, F., Müllerr, K.-R., 2003. Analysing ICA components by injecting noise. Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003). Nara, Japan, pp. 149–154. April.
- Harmeling, S., Meinecke, F., Müllerr, K.-R., 2004. Injecting noise for analysing the stability of ICA components. *Signal Process.* 84 (2), 255–266 (February).
- Himberg, J., Hyvärinen, A., Esposito, F., 2004. Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage* 22 (3), 1214–1222 (July).
- Hoyer, P., Hyvärinen, A., 2000. Independent component analysis applied to feature extraction from colour and stereo images. *Netw.: Comput. Neural Syst.* 11 (3), 191–210 (August).
- Hurri, J., Hyvärinen, A., Karhunen, J., Oja, E., 1996. Image feature extraction using independent component analysis. Proceedings of the IEEE 1996 Nordic Signal Processing Symposium (NORSIG 1996). Vol. 2. Espoo, Finland, pp. 475–478. September.
- Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* 10 (3), 626–634 (May).
- Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. *Neural Netw.* 13 (4–5), 411–430 (June).
- Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis, 1st edition. Wiley-Interscience, New York, NY. May.
- Icasso, 2003. MATLAB™ Package. URL <http://www.cis.hut.fi/jhimberg/icasso>.
- Ito, D., Mukai, T., Murata, N., 2003. An approach of grouping decomposed components. Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003). Nara, Japan, pp. 745–750. April.
- Jahn, O., Cichocki, A., Ioannides, A., Amari, S.-I., 1998. Identification and elimination of artifacts from MEG signals using efficient independent components analysis. Proceedings of the 11th International Conference on Biomagnetism (BIOMAG 1998). Vol. 11. Sendai, Japan. August.
- Jolliffe, I.T., 2002. Principal Component Analysis, 2nd edition. Springer-Verlag, New York, NY. October.
- Jutten, C., Herault, J., 1991. Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture. *Signal Process.* 24 (1), 1–10.
- Lange, T., Roth, V., Braun, M.L., Buhmann, J.M., 2004. Stability-based validation of clustering solutions. *Neural Comput.* 16 (6), 1299–1323 (June).
- Li, Y., Cichocki, A., Amari, S.-I., 2004. Analysis of sparse representation and blind source separation. *Neural Comput.* 16 (6), 1193–1204 (June).
- Makeig, S., Bell, A.J., Jung, T.P., Sejnowski, T.J., 1995. Independent component analysis of electroencephalographic data. Proceedings of the 1995 Conference on Advances in Neural Information Processing Systems (NIPS 1995). Vol. 8. Denver, CO, pp. 145–151. November.
- McKeown, M.J., Sejnowski, T.J., 1998. Independent component analysis of fMRI data: examining the assumptions. *Hum. Brain Mapp.* 6 (5–6), 368–372 (December).
- McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J., 1998. Analysis of fMRI data by blind separation into independent spatial components. *Hum. Brain Mapp.* 6 (3), 160–188 (August).
- McKeown, M.J., Hansen, L.K., Sejnowski, T.J., 2003. Independent component analysis of functional MRI: what is signal and what is noise? *Curr. Opin. Neurobiol.* 13 (5), 620–629 (October).
- Meilä, M., Shi, J., 2000. Learning segmentation by random walks. Proceedings of the Neural Information Processing Systems Conference 2000 (NIPS 2000). Denver, CO. December.
- Meilä, M., Shi, J., 2001. A random walks view of spectral segmentation. Proceedings of the 8th International Workshop on Artificial Intelligence and Statistics (AISTATS 2001). Key West, Florida. January.
- Meinecke, F., Ziehe, A., Kawanabe, M., Müller, K.-R., 2002. A resampling approach to estimate the stability of one-dimensional or multidimensional independent components. *IEEE Trans. Biomed. Eng.* 49 (12), 1514–1525 (December).
- Olshausen, B.A., Field, D.J., 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381 (6583), 607–609 (June).
- Rao, J.S., Tibshirani, R., May 1997. The out-of-bootstrap method for model averaging and selection. Technical Report. University of Toronto, Canada, pp. 1–23.
- Rubin, D.B., 1981. The Bayesian bootstrap. *Ann. Stat.* 9 (1), 130–134 (January).
- Särelä, J., Vigário, R., 2003. Overlearning in marginal distribution-based ICA: analysis and solutions. *J. Mach. Learn. Res.* 4, 1447–1469 (December).
- SPM, 1999. MATLAB™ Package. URL <http://www.fil.ion.ucl.ac.uk/spm>.
- Stögbauer, H., Andrzejak, R.G., Kraskov, A., Grassberger, P., 2004. Reliability of ICA estimates with mutual information. Proceedings of the 5th International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2004). Granada, Spain, pp. 209–216. September.
- Strother, S.C., Anderson, J., Hansen, L.K., Kjems, U., Kustra, R., Sidtis, J., Frutiger, S., Muley, S., LaConte, S., Rottenberg, D., 2002. The quantitative evaluation of functional neuroimaging experiments: the NPAIRS data analysis framework. *NeuroImage* 15 (4), 741–771 (April).
- Tang, A., Pearlmuter, B., Malaszenko, N., Phung, D., Reeb, B., 2002. Independent components of magnetoencephalography: localization. *Neural Comput.* 14 (8), 1827–1858 (August).
- van de Ven, V.G., Formisano, E., Prvulovic, D., Roeder, C.H., Linden, D.E.J., 2004. Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. *Hum. Brain Mapp.* 22 (3), 165–178 (July).
- Vigário, R., Särelä, J., Jousmäki, V., Hämäläinen, M., Oja, E., 2000. Independent component approach to the analysis of EEG and MEG recordings. *IEEE Trans. Biomed. Eng.* 47 (5), 589–593 (May).
- Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited—Again. *NeuroImage* 2 (3), 173–235 (September).
- Ylipaavalniemi, J., Vigário, R., 2004. Analysis of auditory fMRI recordings via ICA: a study on consistency. Proceedings of the 2004 International Joint Conference on Neural Networks (IJCNN 2004). Vol. 1. Budapest, Hungary, pp. 249–254. July.
- Ylipaavalniemi, J., Vigário, R., 2007. Subspaces of spatially varying independent components in fMRI. In: Proceedings of the 7th International Conference on Independent Component Analysis and Signal Separation (ICA 2007). London, UK, pp. 665–672. September.
- Ylipaavalniemi, J., Mattila, S., Tarkiainen, A., Vigário, R., 2006. Brains and phantoms: an ICA study of fMRI. Proceedings of the 6th International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2006). Charleston, SC, pp. 503–510. March.