Validating the independent components of neuroimaging time-series via clustering and visualization

Johan Himberg[1], Aapo Hyvärinen[2*] and Fabrizio Esposito[3]

1) Neural Networks Research Centre
Helsinki Univ. of Technology, Finland,
`johan.himberg@hut.fi`
2) Helsinki Institute for Information Technology / BRU
Dept. of Computer Science, Univ. of Helsinki, Finland
`aapo.hyvarinen@helsinki.fi`
3) Second Division of Neurology, Second University of Naples
and Dept of Neurological Sciences,
University of Naples "Federico II", Italy
`faesposi@unina.it`

14th November 2003

Abstract

Recently, independent component analysis (ICA) has been widely used in the analysis of brain imaging data. An important problem with most ICA algorithms is, however, that they are stochastic, i.e. their results may be somewhat different in different runs of the algorithm. Thus, the outputs of a single run of an ICA algorithm cannot be trusted, and some analysis of the algorithmic reliability of the components is needed. Moreover, as with any statistical method, the results are affected by the random sampling of the data, and some analysis of the statistical significance or reliability should be done as well. Here, we present a method for assessing both the algorithmic and statistical reliability of estimated independent components. The method is based on running the algorithm many times with slightly different conditions, and visualizing the clustering structure of the obtained components in the signal space. In experiments with MEG and fMRI data, the method was able to show that expected components are reliable; furthermore, it pointed out components whose interpretation was not obvious but whose reliability should incite the the experimenter to investigate the underlying technical or physical phenomena. The method is implemented in a sofware package called Icasso.

*Corresponding author. Dept of Computer Science, P.O.Box 26, 00014 University of Helsinki, Finland. Email: `aapo.hyvarinen@helsinki.fi`, fax: +358-9-191-44441, tel: +358-9-191-44270

1 Introduction

Independent component analysis (ICA) (Hyvärinen et al., 2001b) is a general-purpose statistical model that has been widely used in the analysis of brain imaging data, see e.g. (Makeig et al., 1997; McKeown et al., 1998; Vigário et al., 2000). A major problem in application of ICA is that the reliability of the estimated independent components is not known. An ICA algorithm gives a specified number of components, but it is not known which ones are to be taken seriously.

As with any statistical method, it is necessary to analyze the statistical reliability (significance) of the obtained components. The finite sample size induces statistical errors in the estimation—this is the case where classical analysis of statistical reliability should be used (Meinecke et al., 2002). Such an analysis can be accomplished using bootstrapping, a well-known computational method for computing the statistical reliability in the case where a simple mathematical formula cannot be found (Efron and Tibshirani, 1993). Bootstrapping is a resampling method, i.e. the data sample is randomly changed by simulating the sampling process, and the algorithm is then run many times with the bootstrapped samples that are somewhat different from each other. The reliability of the original estimate can then be analyzed by looking at the spread of the obtained estimates.

A further problem typical of ICA is that most algorithms have random (stochastic) elements, i.e. the algorithms give somewhat different results at every run of the algorithm. Most ICA algorithms such a FastICA (Hyvärinen, 1999) or the natural gradient (infomax) algorithm (Bell and Sejnowski, 1995; Amari et al., 1996) are based on minimization or maximization of an objective function, such as the likelihood, mutual information, or negentropy, see (Hyvärinen and Oja, 2000; Hyvärinen et al., 2001b) for a discussion. The randomness of the algorithm stems from the fact that it is not possible to find the point that globally minimizes the objective function (for simplicity of terminology, we shall talk about minimization only since the principle is the same in the case of maximization).

The problem is that most ICA algorithms are based on methods related to gradient descent. The basic principle is to start in some initial point, and then make steps in a direction that decreases the objective function, until one finds a point in which the objective function is locally minimized. Depending on the point where the search was started (the "initial point"), the algorithm will find different local minima. In the case of a very high-dimensional signal space, the probability of finding the global minimum may be very small. Yet, in the case of a high-dimensional signal space, the situation is even more complicated, since at each local minimum, a subset of the components may be accurately estimated. This is plausible because the estimation of the individual components is partly decoupled from each other: in fact, the components can be estimated one at a time by using the criterion of negentropy or nongaussianity (Hyvärinen, 1999). Thus, to assess the algorithmic reliability of the estimated components, it is reasonable to run the estimation algorithm many times, using different initial values, and assessing which of the components are found in almost every run.

Algorithmic and statistical reliabilities in the setting of stochastic algorithms and local minima are illustrated in one dimension in Figure 1.

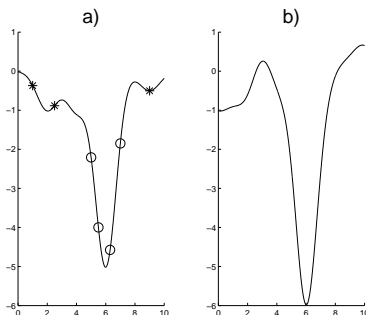In this paper, we present a tool for investigating the reliability of the independent

Figure 1: Illustration of the problem of local minima. a) An objective function that is to be minimized. There is a global minimum at 6, and small local minima at 2 and 9. If a gradient descent algorithm starts at the points marked by circles, it will probably find the global minimum. However, if it starts at the points marked by stars, it will probably converge the the nearest local minimum, at 2 or 9. Running the algorithm many times, it will converge to 6 most of the time. b) When the data is resampled in the bootstrapping method, the optimized function changes a bit. The smaller local minima at 2 and 9 disappear, and a new local minimum appears at 0. The stable minimum at 6 is still a local (and global) minimum.

components. The method is based on estimating a large number of candidate independent components by running an ICA algorithm many times, and visualizing their clustering in the signal space. Each estimated independent component is one point in the signal space. If an independent component is reliable, (almost) every run of the algorithm should produce one point in the signal space that is very close to the "real" component. Thus, reliable independent components correspond to clusters that are small and well separated from the rest of the estimates. In contrast, unreliable components correspond to points which do not belong to any cluster. We investigate both algorithmic and statistical reliability by running the ICA algorithm either with different initial values, or with different bootstrapped data sets, respectively.

Our focus is on constructing a comprehensive set of methods supported by explorative data analysis and visualization. We have developed a software package called *Icasso*[1] to implement these operations and visualize the results. Preliminary results were reported in (Himberg and Hyvärinen, 2003).

2 The *Icasso* software package

2.1 Overview of *Icasso*

First, we provide a concise overview of our method for ICA reliability analysis, *Icasso*. It consists of the following steps:

---

[1] The MATLAB package is available at http://www.cis.hut.fi /jhimberg/icasso/

1. Parameters for the ICA estimation algorithm are selected. We use FastICA so the parameters include the orthogonalization approach (symmetric or deflationary), the nonlinearity, etc.

2. The ICA algorithm is run a number of times using the selected training parameters. Each time the data is bootstrapped and/or the initial point of the optimization is changed.

3. The estimates (estimated components) are clustered according to their mutual similarities. In principle, the clustering method can be freely selected. We apply agglomerative clustering with average-linkage criterion.

4. The clustering is visualized as a 2-D plot. The user investigates how the data is concentrated in the clusters. The clustering of the estimated components is expected to yield information on the reliability (robustness) of estimation. A compact, tight cluster emerges when a similar component repeatedly comes up despite of the randomization.

5. The user can retrieve the estimates belonging to certain cluster(s) for further analysis and visualization.

To complete steps 1–3, the the user simply sets the FastICA parameters and launches a resampling and clustering application. In step 4, the user explores the clustering by launching an interactive visualization application. The user can examine the quality of the clusters and rank them accordingly. Subsequently, *Icasso* visualizes the similarity matrix between all the estimates and their partition into clusters in a single graph. Thus, the user can examine relationships between estimates and clusters in detail. In step 5, the user can retrieve any set of estimates that belong to certain cluster(s).

Our criteria for selecting the specific estimation and visualization methods for *Icasso* were that i) methods for completing each subtask are well-known, ii) they support visualization and explorative data-analysis, and iii) in order to avoid redundant work, existing, publicly available building blocks should be used.[2] In the rest of this section we provide a detailed desciption of the method.

2.2 Generating the estimates and comparing them

We consider the standard linear, noise-free ICA model $\mathbf{x} = \mathbf{As}$ of independent sources $\mathbf{s}$ and a mixing matrix $\mathbf{A}$. However, what is usually estimated in practice, is the demixing matrix $\mathbf{W}$ for $\mathbf{s} = \mathbf{Wx}$, where $\mathbf{W}$ is a (pseudo)inverse of $\mathbf{A}$ (Hyvärinen et al., 2001b).

The FastICA algorithm is run M times on the data matrix $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ldots \mathbf{x}_N]$ consisting of N samples of k-dimensional vectors. The estimates of demixing matrices $\hat{\mathbf{W}}_i$ from each run $i = 1, 2, \ldots, M$ are collected into a single matrix $\hat{\mathbf{W}} = [\hat{\mathbf{W}}_1^T \ \hat{\mathbf{W}}_2^T \cdots \hat{\mathbf{W}}_M^T]^T$. If n_i independent components are estimated on each round, we get $K = \sum_i n_i$ estimates, and the size of $\hat{\mathbf{W}}$ will be $K \times k$.

---

[2]We use the FastICA Toolbox 2.1 and the SOM Toolbox 2.0 (Vesanto et al., 2000) for MATLAB, both freely available from http://www.cis.hut.fi /research/software.shtml

We can recompute independent component estimates by a) *Randomizing the initial condition*: FastICA is run M times for the same data $\mathbf{X}$, so that for each run the algorithm starts from a new random initial condition; b) *Bootstrapping*: FastICA is run M times. The initial condition is kept the same in every run, but the data is resampled by bootstrapping every time; and c) *Bootstrapping with randomized initial condition* as a combination of a) and b).

A natural measure of similarity between the estimated independent components is the absolute value of their mutual correlation coefficients r_{ij}, $i, j = 1, \ldots, K$. Straightforward calculations show that they can be efficiently computed by simple matrix multiplication. In fact, each correlation is given by one element of the matrix $\mathbf{R} = \hat{\mathbf{W}}\mathbf{C}\hat{\mathbf{W}}^T$ where $\mathbf{C}$ is the covariance matrix of the original data $\mathbf{X}$ (this is exactly true in the case of different initial values; for bootstrapping, some extra normalization is necessary). The final similarity matrix then has the elements σ_{ij} defined by

$$\sigma_{ij} = |r_{ij}|. \tag{1}$$

Later, we use clustering methods and validity indices that use dissimilarities (distances). Therefore, we need to transform the similarity matrix into a dissimilarity matrix with elements d_{ij}. A classic way to make this transformation is obviously given by (Everitt, 1993):

$$d_{ij} = 1 - \sigma_{ij}. \tag{2}$$

We will also use a simple modification of this formula as will be explained below.

2.3   Clustering the estimates

We can partition the set of all estimates (estimated independent components) into a number of disjoint clusters using some basic clustering algorithm and the dissimilarity measure in Eq. (2). Agglomerative hierarchical clustering is a well-known method for a modest number of objects (Everitt, 1993; Gordon, 1987)). The tree-like hierarchy (dendrogram) produced by agglomeration is intuitively appealing in the sense that all clusters implied by lower levels of the tree are always subsets of clusters at the higher levels. As a result, the user is able to explore and compare the different level(s) of clustering that are readily computed. The obvious way to obtain a partition of L clusters from a dendrogram is to cut it at level where L clusters are present. A dendrogram is illustrated in Figure 2.

There are numerous reviews and studies on the many agglomeration strategies and cluster validity indices, see, e.g., (Bandyopadhay and Maulik, 2001; Bezdek and Pal, 1998; Everitt, 1993; Gordon, 1987; Maulik and Bandyopadhyay, 2002). Unfortunately, there is no easy way of selecting the optimal agglomeration strategy for a specific set of data, and the selection must be based on problem specific considerations. The same applies also to selecting a clustering validity index for determining a "natural" number of clusters (Bezdek and Pal, 1998).

Three basic agglomeration strategies that operate directly on the similarity matrix are single-link (SL), complete-link (CL), and group average-link (AL). *Icasso* uses AL as the default choice of agglomeration strategy. This is because, firstly, SL is in general reported to be more sensitive to noise than AL and CL. Secondly, our experiments (not
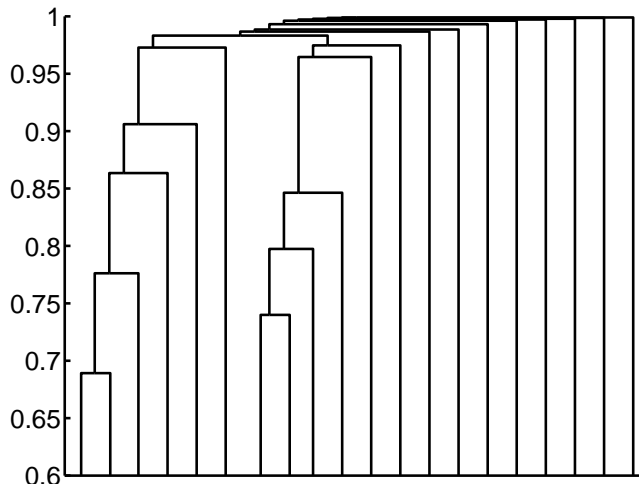
Figure 2: An illustration of a dendrogram and hierarchical agglomerative clustering. The points are successively joined into clusters when moving upwards in the dendrogram. The vertical axis gives the dissimilarity for which the clusters are merged. Clustering can be performed at any level.

shown) revealed that when the number of clusters is smaller than the data dimension ($L < k$), CL starts to join clusters inconsistently.

To direct the attention of the user to those clusters that seem to be the most compact and interesting, we introduce a (conservative) cluster quality index I_q that reflects the compactness and isolation of a cluster. It is computed as the difference between the average intra-cluster similarities and average extra-cluster similarities. Let us denote by C the set of indices of all the estimated components, by C_m the set of indices that belong to the m-th cluster, and by $|C_m|$ the size of the m-th cluster. Then we define the cluster quality index as

$$I_q(C_m) = \frac{1}{|C_m|^2} \sum_{i,j \in C_m} \sigma_{ij} - \frac{1}{|C_m||C_{-m}|} \sum_{i \in C_m} \sum_{j \in C_{-m}} \sigma_{ij}, \tag{3}$$

where C_{-m} is the set of indices that do not belong to the m-th cluster. Eventually, $I_q(C_m)$ is equal to one for an ideal cluster when Eq. (1) is used to compute the similarities σ_{ij}, and decreases when C_m becomes less compact and isolated.

An important parameter that needs to be selected in any clustering method is the number of clusters L. We prefer leaving the final selection of the number of clusters to the user who can interactively explore the results produced by different levels of dendrogram. It is reasonable to start studying the clustering from the number of clusters L equal to the data dimension k and investigate the values of cluster quality index in rank order, as will be done in Fig. 4 below.

There are also quantitative indices for suggesting the number of clusters that best

fits to the data. We considered some of the cluster validity indices that can be computed knowing only the dissimilarity matrix. These included four of the Dunn-like indices in (Bezdek and Pal, 1998) and the R-index defined in (Levine and Domany, 2001). Empirical studies on such indices, e.g., (Bandyopadhay and Maulik, 2001; Bezdek and Pal, 1998; Maulik and Bandyopadhay, 2002) yield often different results depending on the character of the data used without no clear indication of general superiority. Our own experiments on these indices did not suggest any definitive winner either. Currently, *Icasso* shows the R-index (I_R) in the user interface. The index is defined as

$$I_R = \frac{1}{L} \sum_{m=1}^{L} \frac{S_m^{in}}{S_m^{ex}},$$ (4)

where

$$S_m^{in} = \frac{1}{|C_m|^2} \sum_{i,j \in C_m} d_{ij}, \quad \text{and} \quad S_m^{ex} = \min_{m' \neq m} \frac{1}{|C_m||C_{m'}|} \sum_{i \in C_m} \sum_{j \in C_{m'}} d_{ij}.$$

The index is a variant of the Davies-Bouldin index, see (Bezdek and Pal, 1998). It looks for compact and well-separated clusters and the minimum of I_R suggests the best partition. However, we note that such an index should be used only by side of the explorative investigation.

As a spin-off of the reliability analysis, *Icasso* is also able to improve the estimates of the components. In fact, it is possible to integrate information over the many runs performed by computing a representative point for each tight cluster. *Icasso* provides the "centrotype" of the cluster as such a representative point. The centrotype is the point in the cluster that has the maximum sum of similarities (as measured by correlation coefficients) to other points in the cluster.

2.4 Visualization by nonlinear 2-D projection

In addition to assigning the estimated components to clusters, *Icasso* provides a tool for getting a detailed look into the clustering results and relations between the clusters and individual estimates. The result of the hierarchical clustering is typically presented as a dendrogram, but also other types of visualization exist. In *Icasso*, each estimate is plotted as a point on the display, and a convex hull bounds the estimates belonging to the same cluster (Gordon, 1987). This presentation allows visualizing similarities σ_{ij} rather explicitly: the points are connected with lines whose thickness/color represent the similarities between them. This will be shown in Fig. 3 below.

We apply projection methods related to multidimensional scaling (MDS) as suggested in (Gordon, 1987) to approximate the original dissimilarities between estimates by Euclidean distances in two dimensions. This should result in a projection, where the smaller a convex hull is, the more compact the corresponding cluster is. An ideal cluster should contract into a single point.

For this purpose, we compared the linear metric MDS (MMDS) (Torgerson, 1952), and two non-linear methods: Sammon's projection (Sammon, 1969), and Curvilinear Component Analysis (CCA) (Demartines and Hérault, 1997). In addition to visual comparison, we used a trustworthiness index proposed in (Venna and Kaski, 2001).

Spatial proximity is one of the strongest visual indicators of grouping (Ware, 2000). In order to be trustworthy, a projection should be such that one can trust the visual proximity as an indicator of similarity. The trustworthiness index in (Venna and Kaski, 2001) is a function of the visual neighborhood size, and it must be evaluated for the neighborhood sizes of interest: according to (Venna and Kaski, 2001) it is especially important that the trustworthiness is retained for small neighborhoods.

According to our experiments (results not shown), CCA produces more trustworthy projections than MMDS and Sammon's method for dissimilarity measure in Eq. (2). We considered also visualization based on the Self-Organizing Map (SOM) since it is reported to be more trustworthy than many MDS-like methods (Nikkilä et al., 2002; Venna and Kaski, 2001). However, we abandoned this method since its regular grid visualization forces the lines of the similarity graph to shadow each other more than they do on a non-uniform projection.

The projection can be further controlled by modifying the definition of dissimilarity in Eq. (2) suitably, e.g., as

$$d_{ij}^* = \sqrt{1 - \sigma_{ij}}. \tag{5}$$

This spreads the distribution of the distances so that differences in size among the most compact clusters can be seen better. For this reason, *Icasso* uses transformation in Eq. (5) instead that of Eq. (2) for making the visualization.

3 Illustrative experiments

Here we illustrate the utility of *Icasso* by reanalyzing two data sets whose properties are well-known to us. First, we experimented with a magnetoencephalographic (MEG) data set described in more detail in (Vigário et al., 1998). Next we applied *Icasso* on functional magnetic resonance imaging (fMRI) data from a finger-tapping experiment, described in more detail in (Esposito et al, in press).

3.1 Magnetoencephalographic data

Methods   The data consist of preprocessed signals originating from 122-channel whole-scalp MEG measurements from the brain (Vigário et al., 1998). The original signals are band-pass filtered between $0.5\ldots45$ Hz, and the data dimension (k) is reduced from 122 to 20 using principal component analysis in order to reduce noise and overlearning. The recording lasts about 2 minutes and contains 17730 samples. The measurements from the brain are disturbed by signals originating from various sources: heart beat, eye blinks and saccade, and other muscular activity—and a digital watch.

We ran *Icasso* five times using three different settings. Setting I: random initial conditions, third power as nonlinearity; II: as I, but hyperbolic tangent (tanh) as the nonlinearity; and III: as I, but using both bootstrapping and random initial conditions. Each time number of randomizations (M) was 15, and the symmetrical approach was used in FastICA.

Results and discussion   In the following, we present results for a particular test run from setting I. First, we select the number of clusters L to equal 20, which is the dimension of the data (after PCA). The clusters and their interrelations relations are visualized in Fig. 3. Fig. 4 shows the quality index I_q for each cluster. Note how the diameter of the convex hulls representing the clusters grows when the value of quality index I_q decreases.

We notice a "knee" in the graph presenting the ranked I_q when moving from cluster #10 to #12 in Fig. 4. Also, the clustering validity index I_R has a local minimum for $L = 13$ (not shown). Thus, it is interesting to consider the case $L = 13$. Convex hulls marked A and B show how clusters are merged if $L = 13$ is selected instead. The estimated source signals for centrotypes associated to the most robust clusters #1–11 (being outside of convex hulls A and B), are presented in quality rank order in Fig. 5. From the previous studies, we know that source estimates #1 and #2 correspond to eye movements, #3 to heart and #7 to the digital watch. Sources #5 and #6 are related to muscular activities due to biting. As a result, known, strong artifacts are all ranked to the top which is quite reasonable. In repeated runs of *Icasso* with different settings, all top 4 estimates were always ranked 1–4 with the first and the second only occasionally changing places. The next seven estimates remained usually in top 11, except that estimates #5–6 related to biting became less reliably estimated, especially in setting II.

Icasso is thus able to point out some components that are worth investigating further. Above all, source #4 is interesting since it is clearly well estimated—even in repeated experiments and in other settings—but the physiological explanation, if any, is not yet known. The same applies, to a lesser extent, to the components #8–11. These components may not look very interesting at first sight, which is possibly why they were not investigated in detail by (Vigário et al., 1998).

3.2   Functional magnetic resonance images

Methods   A healthy volunteer participated in a dominant-hand finger-tapping fMRI experiment (Esposito et al, in press). The experimental paradigm consisted of five intervals of five time-points during which a self-paced finger-tapping task (sequential opposition of all fingers of the right hand against the thumb) at a specified frequency of 2 Hz, was carried out alternated with five intervals of five time-points of resting. The alternation between task and rest conditions was verbally triggered and the frequency and quality of the task were controlled by visual inspection. An additional "rest" block was acquired before the first "task" block of which the first two time-points were discarded to avoid the global effects induced by incomplete magnetization equilibrium.

Each image of the slice time-series was smoothed in space by applying an isotropic gaussian kernel to low-pass filter each slice at each time-point. Then, the dimensionality of the data was reduced to 30 by PCA. In FastICA, we used a symmetric approach, the nonlinearity being tanh. We tested the algorithmic reliability by running the ICA algorithm 15 times with random initial conditions.

Results and discussion   The similarity graph is shown in Fig. 6, and the five first centrotype activation maps are shown in Fig. 7. Based on previous studies we interpret
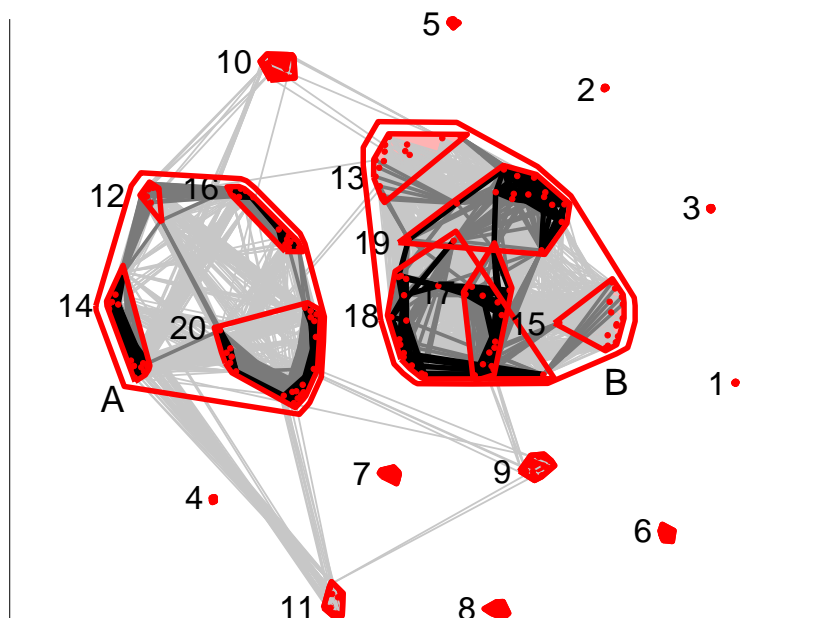
Figure 3: The similarity graph of the estimates for MEG data. Clusters are indicated by red convex hulls. Grey and lines connect estimates whose similarity is larger than a threshold, the darker the line the stronger the similarity. Labels 1–20 correspond to Figure 4. Convex hulls A and B show how clusters agglomerate further if the number of clusters is set to $L = 13$ instead of $L = 20$.
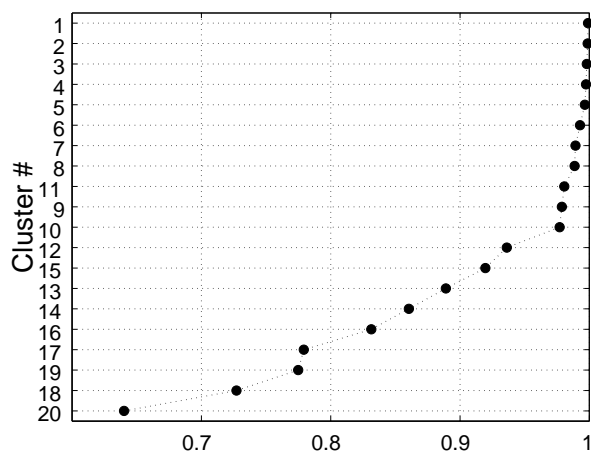

Figure 4: The quality index I_q in rank order for the clusters in Fig. 3.
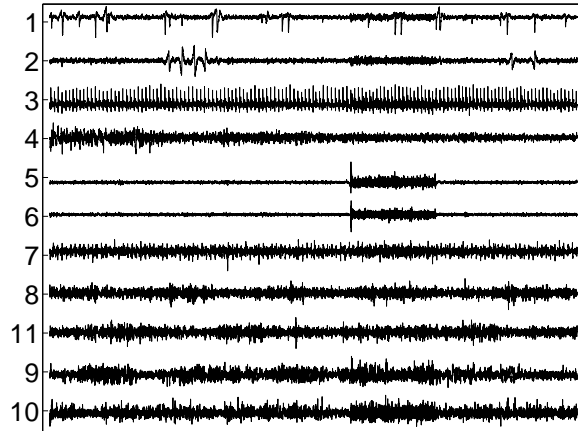
Figure 5: Estimated sources corresponding to centrotypes of clusters #1–11 in Fig. 3.

the first five clusters as follows:

1. Task-related source. The highest values of the component clearly appear to cluster anatomically in contra-latarel rolandic region of the primary motor cortex and the supplementary motor area. The primary somatosensory area (SI) is active as well.

2. Vascular source, showing activation foci that cluster in the regions of large blood vessels. The time-course (not shown) was poorly related to task execution.

3. This source is difficult to interpret and worthy of further study. The nice bilateral structure and the shape of the time-course suggest the classification of this source as an "auditory source" caused by the auditory cue (GO-STOP) but it is not possible to confirm this hypothesis in the absence of more inferior slices that covered primary and secondary cortices.

4. Motion-related source, active at the boundaries of brain (typical ring-like activation patterns). These phenomena are representative of head motion effects during scanning. The associated time-course (not shown) exhibited a long-term effect on the data, starting at the onset of the first task block.

5. Another vascular source, showing activation foci that cluster in the regions of large blood vessels. This source exhibits an evident correlation with the task and includs the motor cortex activity. Possibly, it corresponds to modulation of task-related activity.

Thus, again we see that in addition to showing that some estimated components are reliable (here, in the algorithmic sense), *Icasso* is able to point out a component (#3) whose interpretation is not obvious and which might not capture the attention of the experimenter who just runs ICA once and is confronted with a large number of estimated components.
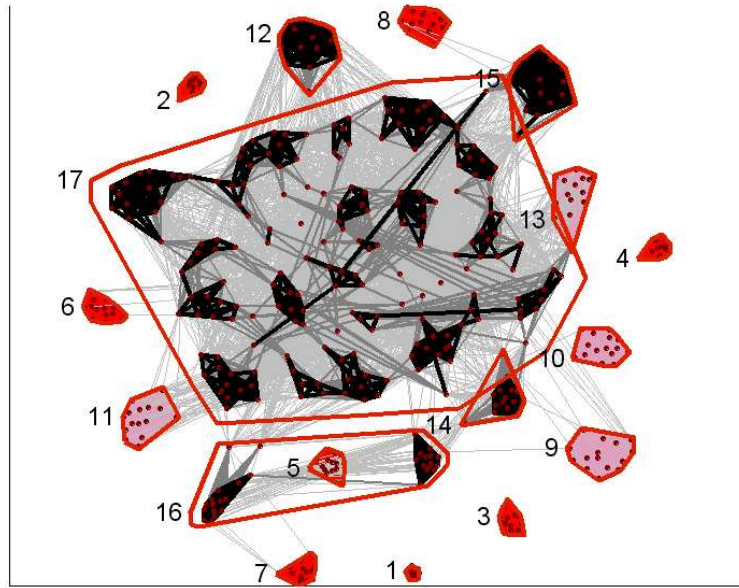
Figure 6: The similarity graph of the estimates for fMRI data. Clusters are indicated by red convex hulls and grey/black lines connect similar estimates.
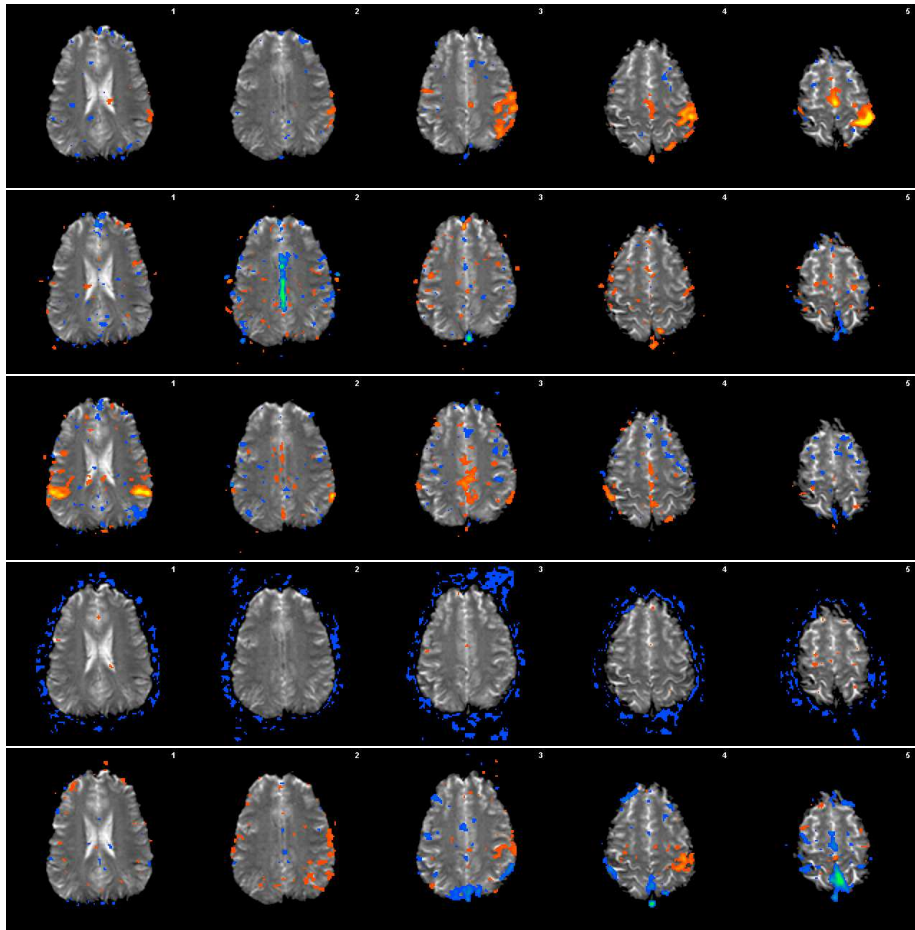
Figure 7: The centrotypes of the five best clusters (top is best) in the fMRI experiment.

4   Discussion

4.1   Related work

Meinecke et al (2002) have previously proposed a method for analyzing the statistical reliability of independent components. However, it seems that their method is only applicable in the case of deterministic algorithms, whereas most analysis of independent components of brain imaging data uses non-deterministic algorithms. The reason is that they assume that small perturbations in the data results in small perturbations in the estimated mixing matrix. While this assumption is true for the algorithms they used, it is not true for the most widely-used ICA algorithms which are stochastic. Meinecke et al (2002) did not consider the algorithmic reliability, either.

An alternative approach to ours would be to measure the independence of the components given by the algorithm. Basically, as most ICA algorithms are based on the maximization of some measure of independence (Hyvärinen et al., 2001b), this boils down to measuring how well the algorithm optimized its objective function. Thus, we could run the ICA algorithm many times and choose the results of the run that gave the most independent components. Such an approach seems to be quite unsatisfactory, however. Some of the components may be well estimated in some runs, other in other runs, and it is not reasonable to assume that the best run gives the best estimates for each component. In *Icasso*, we are able to combine information from several runs of the algorithms. Indeed, we thus obtain a set of components (given by the cluster centers in the signal space) that is better than any of component sets provided by a single run.

Let us also mention some further methods that may be of help in finding the most interesting or meaningful components given by ICA. Formisano et al (2002) developed some indices that are correlated with the interestingness of the components; the components can be ordered using these indices after ICA estimation. Hyvärinen et al (2001a) developed a method that orders components on a one-dimensional line or a two-dimensional grid so that components that belong together, possibly being manifestations of the same multidimensional signal source, are close to each other.

4.2   Future directions

As for the computational load of the algorithm, we have noted that the hierarchical clustering is clearly the bottleneck in the current computational environment. Currently, say, $600 \ldots 1000$ estimated components can be handled on an ordinary PC, which corresponds to, for example, 10 runs of the algorithm when the dimension of the data is $60 \ldots 100$. On the current implementation the computational load in terms of time consumption rises as a cubic function of sample size. Consequently, for bigger amounts of estimates, a more sophisticated implementation may be needed. We are currently investigating alternative methods for clustering.

Finding clusters in the high-dimensional signal space involves fixing the number of clusters to be modeled, as well as the values of other internal parameters. Automatic determination of optimal values for these parameters is a most difficult theoretical problem; ultimately, the optimal values also depend on application-specific and subjective considerations. Therefore, we propose an interactive method based on visualization of

the clustering structure. Yet, it may be possible to find purely automatic methods that offer satisfactory performance. Then the visualization step could be omitted, and numbers describing the reliability of each components could be automatically computed. At this point it seems, however, that the visualization gives the user useful additional information on the structure of the clusters, and may be quite valuable in its own right.

4.3   Conclusions

We have developed an interactive visualization method and software package for analyzing the reliability (significance or robustness) of independent components of brain imaging data. The basic principle is to run an ICA algorithm many times, and look at the clustering of the estimated components in the signal space. Basically, each reliable cluster should produce a "tight" cluster of estimated components that are very close to each other, and well separated from the rest. Reliability has two aspects, algorithmic and statistical, which can be probed by running the algorithm with different initial values or bootstrap samples, respectively.

References

Amari, S.-I., Cichocki, A., and Yang, H. (1996). A new learning algorithm for blind source separation. In *Advances in Neural Information Processing Systems 8*, pages 757–763. MIT Press.

Bandyopadhay, S. and Maulik, U. (2001). Nonparametric genetic clustering: Comparison of validity indices. *IEEE Trans. on Systems, Man and Cybernetics C*, 31(1):120–125.

Bell, A. and Sejnowski, T. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159.

Bezdek, J. and Pal, N. (1998). Some new indexes of cluster validity. *IEEE Trans. on Systems, Man and Cybernetics B*, 28:301–315.

Demartines, P. and Hérault, J. (1997). Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets. *IEEE Trans. on Neural Networks*, 8(1):148–154.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.

Esposito, F., Seifritz, E., Formisano, E., Morrone, R., Scarabino, T., Tedeschi, G., Cirillo, S., Goebel, R., and Di Salle, F. Real-time independent component analysis of functional MRI time-series. *NeuroImage*. In press.

Everitt, B. (1993). *Cluster Analysis*. Edward Arnold, London, third edition.

Formisano, E., Esposito, F., Kriegeskorte, N., Tedeschi, G., Salle, F. D., and Goebel, R. (2002). Spatial independent component analysis of functional magnetic resonance imaging time-series: characterization of the cortical components. *Neurocomputing*, 49(1–4):241–254.

Gordon, A. (1987). A review of hierarchical classification. *Journal of the Royal Statistical Society. Series A (General)*, 150(2):119–137.

Himberg, J. and Hyvärinen, A. (2003). Icasso: software for investigating the reliability of ica estimates by clustering and visualization. In *Proc. IEEE Workshop on Neural Networks for Signal Processing (NNSP2003)*, pages 259–268, Toulouse, France.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634.

Hyvärinen, A., Hoyer, P. O., and Inki, M. (2001a). Topographic independent component analysis. *Neural Computation*, 13(7):1527–1558.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001b). *Independent Component Analysis*. Wiley Interscience.

Hyvärinen, A. and Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4-5):411–430.

Levine, E. and Domany, E. (2001). Resampling method for unsupervised estimation of cluster validity. *Neural Computation*, 13(11):2573–2593.

Makeig, S., Jung, T.-P., Bell, A. J., Ghahramani, D., and Sejnowski, T. (1997). Blind separation of auditory event-related brain responses into independent components. *Proc. National Academy of Sciences (USA)*, 94:10979–10984.

Maulik, U. and Bandyopadhay, S. (2002). Performance evaluation of some clustering algorithms and validity indices. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(12):1650–1654.

McKeown, M., Makeig, S., Brown, S., Jung, T.-P., Kindermann, S., Bell, A., Iragui, V., and Sejnowski, T. (1998). Blind separation of functional magnetic resonance imaging (fMRI) data. *Human Brain Mapping*, 6(5-6):368–372.

Meinecke, F., Ziehe, A., Kawanabe, M., and Müller, K.-R. (2002). A resampling approach to estimate the stability of one-dimensional or multidimensional independent components. *IEEE Transactions on Biomedical Engineering*, 49(12):1514–1525.

Nikkilä, J., Törönen, P., Kaski, S., Venna, J., Castrén, E., and Wong, G. (2002). Analysis and visualization of gene expression data using self-organizing maps. *Neural Networks*, 15:953–966.

Sammon, J. W. (1969). A nonlinear mapping for data structure analysis. *IEEE Trans. on Computers C*, 18(5):401–409.

Torgerson, W. (1952). Multidimensional scaling I—Theory and methods. *Psychometrica*, 17:401–419.

Venna, J. and Kaski, S. (2001). Neighborhood preservation in nonlinear projection methods: An experimental study. In *Artificial Neural Networks (ICANN 2001)*, pages 485–491. Springer.

Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J. (2000). SOM toolbox for Matlab 5. Report A57, Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland.

Vigário, R., Jousmäki, V., Hämäläinen, M., Hari, R., and Oja, E. (1998). Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In *Advances in Neural Information Processing Systems*, volume 10, pages 229–235. MIT Press.

Vigário, R., Särelä, J., Jousmäki, V., Hämäläinen, M., and Oja, E. (2000). Independent component approach to the analysis of EEG and MEG recordings. *IEEE Trans. Biomedical Engineering*, 47(5):589–593.

Ware, C. (2000). *Information Visualization: Perception for Design*. Morgan Kaufmann, San Francisco.