Robust ordering of independent components in functional magnetic resonance imaging time series data using canonical correlation analysis

Tamer Youssef[a], Abou-Bakr M. Youssef[a], Stephen M. LaConte[b], Xiaoping P. Hu[b], and Yasser M. Kadah[a,b1]

[a]Biomedical Engineering Department, Cairo University, Egypt
[b]Emory University/Georgia Tech Biomedical Engineering, Atlanta GA 30322

ABSTRACT

The application of independent components analysis (ICA) to functional magnetic resonance imaging data has been proven useful to decompose the signal in terms of its basic sources. The main advantage is that ICA requires no prior assumption about the neuronal activity or the noise structure, which are usually unknown in fMRI. This enables the detection of true activation components free of random and physiological noise. Hence, this technique is superior to other techniques such as subspace modeling or canonical correlation analysis, which have underlying assumptions about the signal components. Nevertheless, this technique suffers from a fundamental limitation of not providing a consistent ordering of the signal components as a result of the whitening step involved in ICA. This mandates human intervention to pick out the relevant activation components from the outcome of ICA, which poses a significant obstacle to the practicality of this technique. In this work, a simple yet robust technique is proposed for ranking the resultant independent components. This technique adds a second step to ICA based on canonical correlation analysis and the prior information about the activation paradigm. This enables the proposed technique to provide a consistent and reproducible ordering of independent components. The proposed technique was applied to real event-related functional magnetic resonance imaging data and the results confirm the practicality and robustness of the proposed method.

Keywords: Functional magnetic resonance imaging, independent component analysis, canonical correlation analysis

1. INTRODUCTION

The advent of functional magnetic resonance imaging (fMRI) has resulted in many exciting studies that have exploited its unique capability. It provides a valuable noninvasive tool for investigating brain function. The different magnetic properties of oxy- and deoxyhemoglobin are used to visualize localized changes in blood flow, blood volume and blood oxygenation in the brain[1,2,3]. These in turn become indicators for local changes in neural activity. To observe these homodynamic changes, the subject is exposed to controlled stimuli, which are carefully designed to affect only certain brain functions then rapid acquisition of a series of brain images is performed. The sequence of images is analyzed to detect such changes and the result is expressed in the form of a map of the activated regions, which represents sensory, motor, and cognitive functions in the brain[4]. fMRI analysis approaches range from model-based to exploratory, although model-based approaches are by far the most utilized[5]. Model-based approaches are extremely useful when the time course of the hemodynamic response can be inferred apriori, however the hemodynamics of the brain are still being studied and good *a priori* assumptions are sometimes not available[6]. Besides, the data are in general very noisy, and much statistical research has been devoted to studying how the weak activation signals may be extracted with optimal sensitivity. Thus, the use of data-driven analysis in fMRI is inherently attractive because it does not rely on imposed assumptions about experimental conditions. The concept of data driven decompositions became familiar since principal component analysis (PCA) and independent component analysis (ICA) have been introduced to fMRI data analysis[1,7,8]. Both ICA and PCA are closely related to projection pursuit methods[9,10]. The underlying idea is to find interesting directions or components within the multi-dimensional data set. Both ICA and PCA use linear transformation to get the

[1] E-mail: ymk@ieee.org

components of the observed signals. The key difference however, is in the type of components obtained. The goal of PCA is to obtain principal components, which are defined as the uncorrelated direction of highest variance. In ICA however, the stronger constraint of statistical independence is imposed. It seeks to obtain statistically independent components. Hence, PCA algorithms use only second order statistical information, while ICA algorithms utilize higher order statistical information for separating the signals. Moreover, PCA gives projections of the data in the direction of the maximum variance. The principal components (PCs) are ordered in terms of their variances: the first PC defines the direction that captures the maximum variance possible; the second PC defines (in the remaining orthogonal subspace) the direction of maximum variance, and so forth. This provides a natural ranking for the resulting components (even though this might not provide information about the usefulness of the components since the variance of a signal is not necessarily related to the importance of the variable). On the other hand, ICA involves an intermediate whitening step that does not make the ranking based on projection magnitude possible. Therefore, there is no consensus of how this ordering can be performed[15]. Several attempts were reported including the use of the value of higher order statistics[17]. These methods were not proven useful in many applications including fMRI and electrophysiology data[17]. The lack of consistent ordering of components results in different arrangement of independent components each time the analysis is performed even on the same data set, which mandates the intervention by the user to select the "interesting" components out of the analysis result. This is usually a cumbersome task that takes fairly long time and makes the technique subject-dependent. Therefore, an analysis technique that would allow the robust ordering of independent components without the intervention of the user would be rather useful.

In the present work, we present a simple yet effective method for functionally ranking the resulting components of the ICA algorithm by using a priori information about the functional paradigm. This technique applies one of the prominent model-based fMRI tools, namely canonical correlation analysis (CCA), to the outcome of ICA. This combination enables the elimination of the limitations of both techniques and provides a robust model-free mechanism for analyzing fMRI data. The proposed methodology is applied to real event-related fMRI data and the results confirm the consistency of ICA ordering using this method.

2. BACKGROUND

2.1. Independent component analysis

Independent Component Analysis (ICA) is a promising tool in statistical analysis. It was developed as a solution to blind source separation problems. These problems involve multiple sources and sensors for situations in which multiple measurements of linearly mixed independent signals are available. In its most common form, ICA is represented by:

$$X = A \cdot S \ , \quad \text{or equivalently,} \quad S = W \cdot X \ , \tag{1}$$

where X is our data matrix where rows correspond to time and columns to spatial locations (a single column is a time course at a given voxel), S is the matrix of the independent components, A is the mixing matrix and W is unmixing matrix and they are the inverse of each other. A, W, S are all estimated from X. Assuming that the sources S are mutually independent and by constructing a suitable measure of independence among the components of S, we may estimate S by optimizing this measure.

Figure 1. Basic concept of ICA.

The use of ICA for analyzing fMRI data was first proposed by McKeown [1]. Here the measured MRI signal is a mixture of the signals arising from various biological processes. It is desired to recover activations maps and time courses corresponding to the functional task in question. The main advantage is that ICA does not require any a prior assumption about neither the signal response to the neuronal activation nor the noise structure surrounding the signal, which are both unknown in fMRI. However, the main drawback of ICA is that it provides no natural ranking for the resultant components. This is due to the nature of its concept. From Eq. (1), ICA assumes no information about both matrix A and S and try to extract both of them. Whatever the algorithm it uses to find them, if we multiply any row of

matrix *A* with a factor *m* and divide the corresponding column in *S* with a factor *1/m*, nothing will change from the view of the resultant output. Most current ICA algorithms use this factor to maintain a unit variance for all the out components or to maintain any other required criteria. This shows that the variance, entropy and all signal properties (at least those that depends on the second order moment) do not contain any useful information because simply the algorithm has assumed it during solving the problem. And so, we cannot depend on any feature of the generated components to make any relations between them (ordering them for example).

2.2. Canonical correlation analysis

Canonical correlation analysis (CCA) is a well-known tool in statistical analysis, developed by Hotelling[11]. CCA is a way of measuring the linear relationship between two multidimensional variables. It finds two bases, one for each variable, that are optimal with respect to correlations and, at the same time, it finds the corresponding correlations[12,13].

Consider two multidimensional random vectors x and y which are m, n-dimensional respectively. We try to find linear combinations,

$$X = a_1x_1 + a_2x_2 + a_3x_3 + \ldots + a_mx_m = w_x{}^T x \quad , \tag{2}$$

and,

$$Y = b_1y_1 + b_2y_2 + b_3y_3 + \ldots + b_ny_n = w_y{}^T y \quad , \tag{3}$$

which are chosen so that X and Y correlate the most. The linear combination coefficients are gathered in the two vectors w_x and w_y. The projections onto w_x and w_y, i.e. X and Y are called canonical variates. Thus, we attempt to solve the following optimization problem,

$$\max_{W_x,W_y}\{r\} = \frac{C(X,Y)}{\sqrt{V(X)V(y)}} \, . \tag{4}$$

Since both variance V and covariance C are invariant with respect to an addition of a constant, we can without loss of generality assume a zero mean. The expression for the correlation r in Eq. (3) can then be written as,

$$r = \frac{E(XY)}{\sqrt{E(X^2)E(y^2)}} \quad , \tag{5}$$

which can be expressed as,

$$r = \frac{E(w_x{}^T xy^T w_y)}{\sqrt{E(w_x{}^T xx^T w_x)E(w_y{}^T yy^T w_y)}} \quad . \tag{6}$$

Applying the expectation operators leads to the following formula for the objective function,

$$r = \frac{w_x{}^T C_{xy} w_y}{\sqrt{w_x{}^T C_{xx} w_x w_y{}^T C_{yy} w_y}} \quad . \tag{7}$$

Here, C_{xx} and C_{yy} are the auto-correlation matrices between x, y and C_{xy} is the cross-correlation matrix between x, y. By taking the partial derivatives of Eq. (4) with respect to w_x and w_y we arrive in the following system of equations,

$$C_{xy}\hat{w}_y = r\lambda_x C_{xx}\hat{w}_x \, , \tag{8}$$

and

$$C_{yx}\hat{w}_x = r\lambda_y C_{yy}\hat{w}_y \, , \tag{9}$$

where λ_x and λ_y are expressed as,

$$\lambda_x = \lambda_y{}^{-1} = \sqrt{\frac{\hat{w}_y{}^T C_{yy}\hat{w}_y}{\hat{w}_x{}^T C_{xx}\hat{w}_x}} \quad . \tag{10}$$

A simple scaling of the linear combination coefficients in Eqs. (1) and (2) would not affect the correlation coefficient. The norms of w_x and w_y are for this reason free of choice. A natural choice is to make $\|w_x\|^2 = \|w_y\|^2 = 1$, which explains the normalized vectors w_x and w_y in Eq. (5). A solution to the equation system in Eq. (5) can be obtained as,

$$C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} \hat{w}_x = r^2 \hat{w}_x \quad , \tag{11}$$

and,

$$C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy} \hat{w}_y = r^2 \hat{w}_y \quad . \tag{12}$$

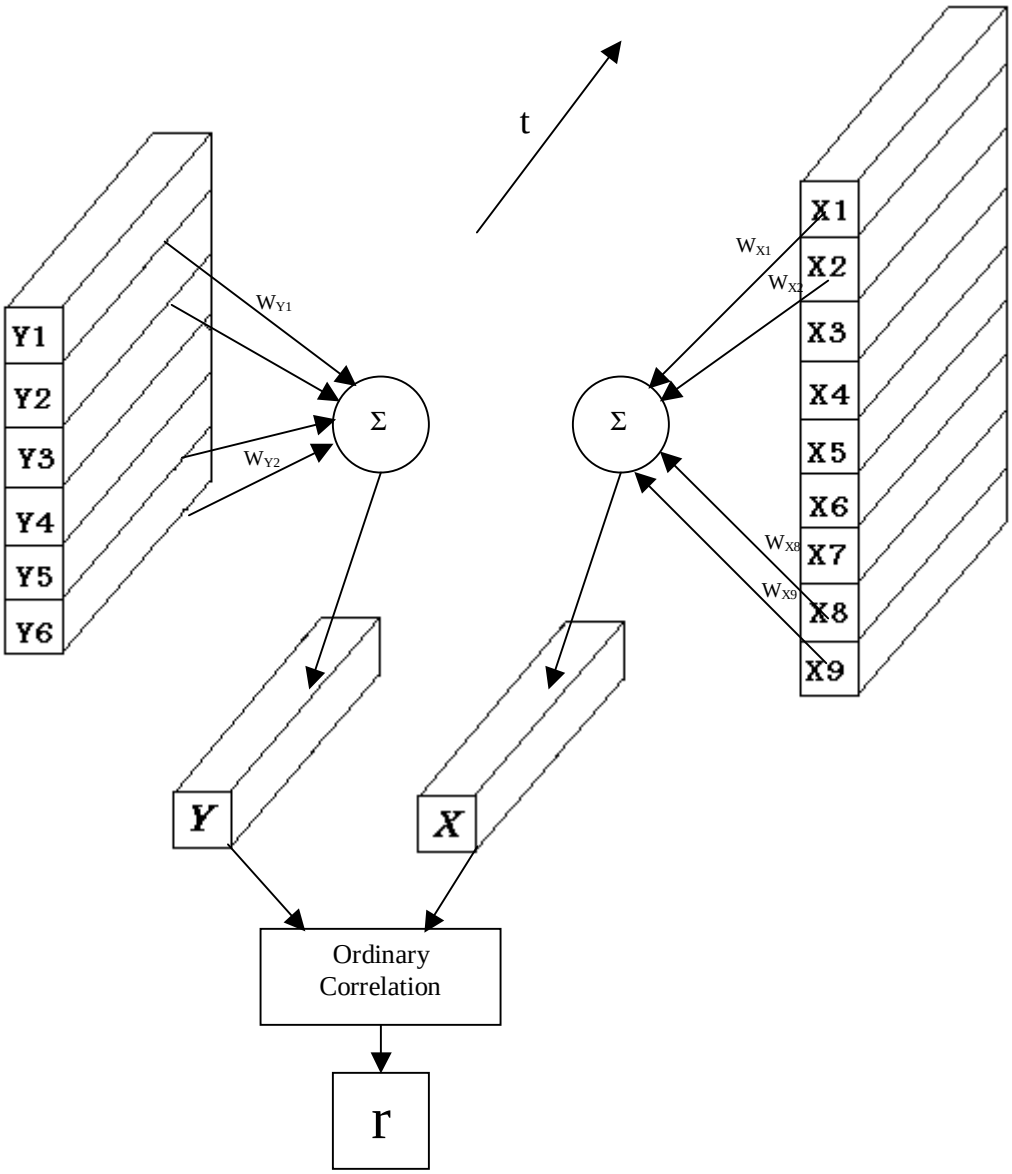

Figure 2. CCA between two-dimensional vectors X and Y.

Thus w_x and w_y are eigenvectors to the matrices $C_{xx}^{-1}C_{xy}C_{yy}^{-1}C_{yx}$ and $C_{yy}^{-1}C_{yx}C_{xx}^{-1}C_{xy}$, and the eigenvalues r_2 are the squared canonical correlations. Note that we obtain several solutions, as there are several eigenvectors to the above

matrices. In fact, we obtain *min(m, n)* possible solutions, but the eigenvectors $w_{x,1}$ and $w_{y,1}$ corresponding to the largest eigenvalue, i.e., the largest canonical correlation.

3. METHODS

3.1. Algorithm

Here, we describe the methodology by which the combination of ICA and CCA will be performed. In event-related fMRI, the activation signal seems to be our helper. It is not used in the calculation of the ICs. Beside, we can assume without any restriction that the neuronal response is periodic and have in some way the same periodicity of the activation signal, i.e., it contains some harmonics that looks like the harmonics in the activation signal. And from here we can start.

In an fMRI experiment, a number of image slices *N* are acquired at *N* subsequent time points. With each pixel in each image slice we obtain a time series of length *N*. The independent components of this time series can then be computed using ICA. Some of these components are relevant to the neuronal activation; others are relevant to the physiological noise while others show the random noise components. We are searching for ICs whose time series has a component that follows the paradigm. Due to the low Signal-to-Noise ratio, most of these components are heavily contaminated with noise. So, we cannot use a normal correlation between them and the activation signal to know which of them is more relevant to the neuronal activity. So, we use the canonical correlation (CCA) between all the ICs as vector *x* in Eq. (2) and a model for the activation signal as vector *y* in Eq. (3). This model is chosen depending on the symmetric square wave paradigm of the activation signal. It is then likely to assume that the response will have the same fundamental frequency as the paradigm. It is probably also a good choice to include components with frequencies equal to a following few harmonics in the Fourier series expansion of the symmetric square-wave. In our work, we have chosen the following simple basis[14],

$$y(t)=\begin{pmatrix}\sin(\omega t)\\ \cos(\omega t)\\ \sin(3\omega t)\\ \cos(3\omega t)\\ \sin(5\omega t)\\ \cos(5\omega t)\end{pmatrix}, \qquad \omega=\frac{2\pi}{T}, \qquad t=1,2,\ldots,N. \tag{13}$$

That is, the chosen basis consists of the truncated Fourier-series expansion of a square-wave, which contains odd harmonics only) where T is the period in the paradigm and t is a discrete time index. This is similar to the decomposition in[16] where also the assumption of little power in the frequencies beyond the fifth harmonic is asserted.

Then, CCA begins to work to find a linear combination between the ICs time series $X=w_x^T x$ and the basis functions $Y=w_y^T y$ so that the correlation between *X(t)* and *Y(t)* is the largest achievable as shown in Fig. (3). This procedure results in a six sample canonical correlation, r_i, i=1,2…6 with a corresponding w_x and w_y. It is clear that the higher the value of r_i, the higher the relation with the activation paradigm and the corresponding w_x, moreover, the higher a certain coefficient in w_x, the higher the relation between the activation paradigm and the IC that corresponds to this coefficient. Based on this note, a certain grade can be assigned for each IC based on its relation with the activation paradigm. This can be given by:

$$M=\sum_{i=1}^{n} w_x^T \cdot r_i \qquad . \tag{14}$$

So each factor of the w_x will be weighted by how strong it will correlate with the activation paradigm. By ordering the vector *M*, we can get the order of the ICs according to its relation to the activation paradigm.

3.2. Noise considerations

Unfortunately, the sine/cosine basis choice may lead to problems with noise especially in fMRI environment. That is mainly because noise can just by pure chance make a strong correlation with some signal in *Y*. A linear combination of our chosen basis (which forms *Y(t)*) can be:

$$y(t) =, w_{y1}\sin(\omega t) + w_{y2}\cos(\omega t) + w_{y3}\sin(3\omega t) + w_{y4}\cos(3\omega t) + w_{y5}\sin(5\omega t) + w_{y6}\cos(5\omega t). \quad (15)$$

Applying trigonometric identities to the above equation yields,

$$y(t) =, k_1\sin(\omega t + \theta_1) + k_{23}\sin(3\omega t + \theta_2) + k_3\sin(5\omega t + \theta_3) \quad , \quad (16)$$

where *k1, k2,* and *k3* represent the magnitudes of the different frequency components, which determine the shape of Y(t). And θ1, θ2, θ3 represent the phase delay between the frequency components. To illustrates the importance of these different variables, different samples of Y(t) are generated with different sets of these variables as shown in Fig. 3.

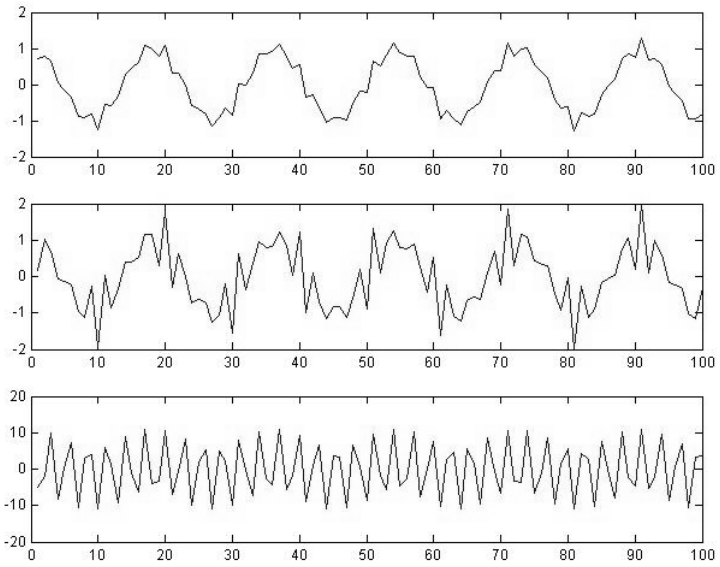

Figure 3. Different generated time series form the chosen basis functions

It should be noted that the top generated time series is very close to the activation paradigm itself (actually, it is the projection of the square wave activation paradigm on our basis space). The second can be accepted as an activation response to this paradigm, while the third time series seems to be just a noise. Note that all of them are generated using the same basis in our work. So, some signal that looks like the third one can deceive our system and give a very high correlation with a noisy component while we will convert this high correlation factor to high grade for this component as an activation response. To solve this problem, we return back to the activation paradigm. We can imagine that the activation response component should be more similar to this paradigm than any other component especially the noise components that cause the problem we are facing. So, the projection of the activation response on our chosen basis should give similar coefficients to those comes from the projection of the paradigm itself on these basis. By applying any measurement technique to measure this similarity between the two coefficient vectors (mean square error for example), we can get a new rank of how close the vector Y(t) corresponds to the activation response not to a noisy component. By combining these two ranks, we can get a general rank for each generated IC by:

$$M = \sum_{i=1}^{n} \frac{w_x^T \cdot r_i}{MSE\left(w_{yi} - w_{y0}\right)} \quad . \quad (17)$$

Here M is the resultant rank vector, r_i is the correlation coefficient result from the CCA, w_{yi} is the i_{th} row of w_y corresponds to the i^{th} IC in the CCA (i.e. the projection of the i_{th} linear combination of the ICs on the basis functions) and w_{y0} is the projection vector of the activation paradigm on our basis. The higher the term $MSE\ (w_{yi}\text{-}w_{y0})$, the wider the

distance between this linear combination of the ICs and the activation paradigm and most probably it is correlated to a noise component and so decrease the grade of this combination of the ICs.

4. RESULTS

The present technique was applied to ICA results for event-related fMRI data. The data were obtained from an activation study performed on a volunteer using a Siemens 1.5T clinical scanner. In this study, an oblique slice through the motor and the visual cortices was imaged using a T2*-weighted EPI sequence (TE/TR= 60/300 ms, Flip angle=55°, FOV=22cm×22cm, slice thickness=5 mm). The subject performed rapid finger movement cued by flashing LED goggles. The study consisted of 31 epochs, with 64 images per epoch[3]. Temporal ICA was applied to process groups of pixels within a user-specified region of interest of variable size between 4×4 to 16×16. The proposed ranking method based on CCA was used to order the outcome of ICA, which came in different order each run. The results of ICA before and after applying the ordering technique are shown in Figs. 4-7. As can be seen, the components are consistently ordered in these examples. Furthermore, in each of these examples, even though the order of components came different from ICA from the data before the application of the proposed technique, the results after its application were exactly the same each time. This suggests that the addition of the proposed method as a postprocessing step after ICA makes the technique more practical for use in clinical settings.

5. CONCLUSIONS

A new technique for ordering the outcome of temporal ICA is proposed. With the aid of CCA, our method attempts to distribute noise in fMRI signals to many projections so its effect on the overall decision of ranking is minimal. The results obtained in our experiments suggest this technique to be highly robust, which makes it suitable as a postprocessing step after ICA to make its results easier to evaluate and the technique more practical to use.

ACKNOWLEDGEMENTS

The authors would like to acknowledge support by International Electronics, Egypt, and NIH (grants RO1MH55346 and RO1EB00321), Georgia Research Alliance, and The Whitaker Foundation.

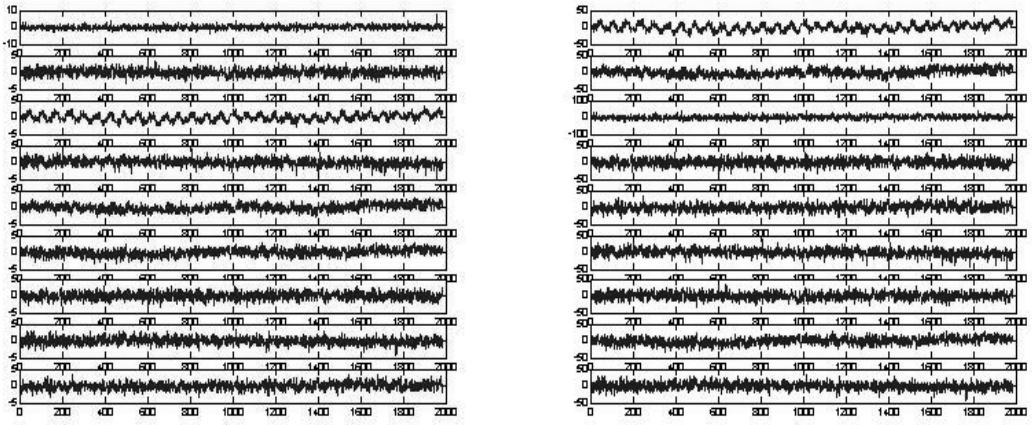

Figure 4. The generated ICs (left) without ordering. and (right) after ordering – Example 1.

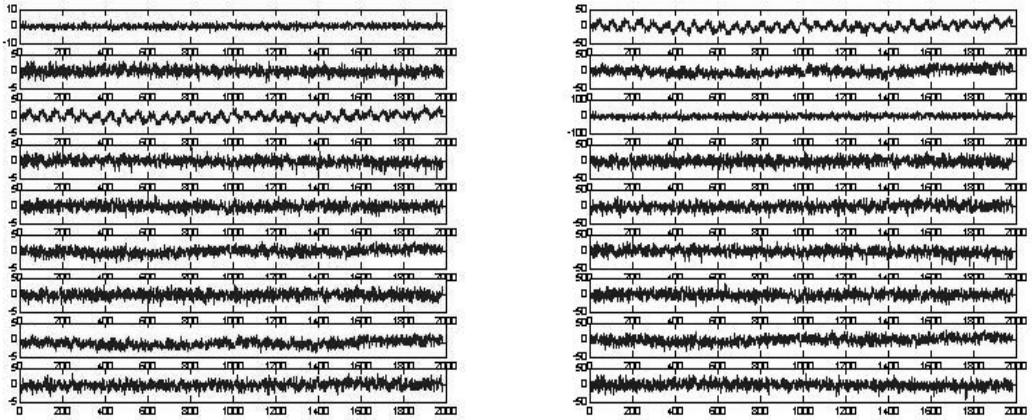

Figure 5. The generated ICs (left) without ordering. and (right) after ordering – Example 2.

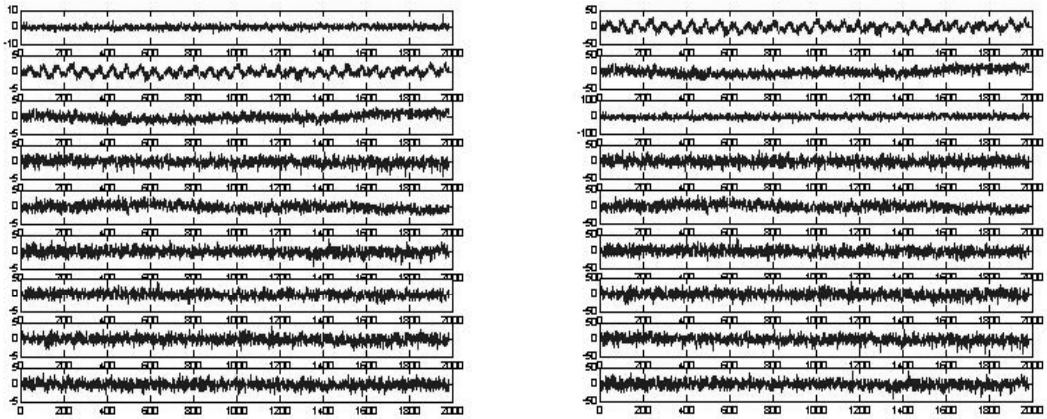

Figure 6. The generated ICs (left) without ordering. and (right) after ordering – Example 3.

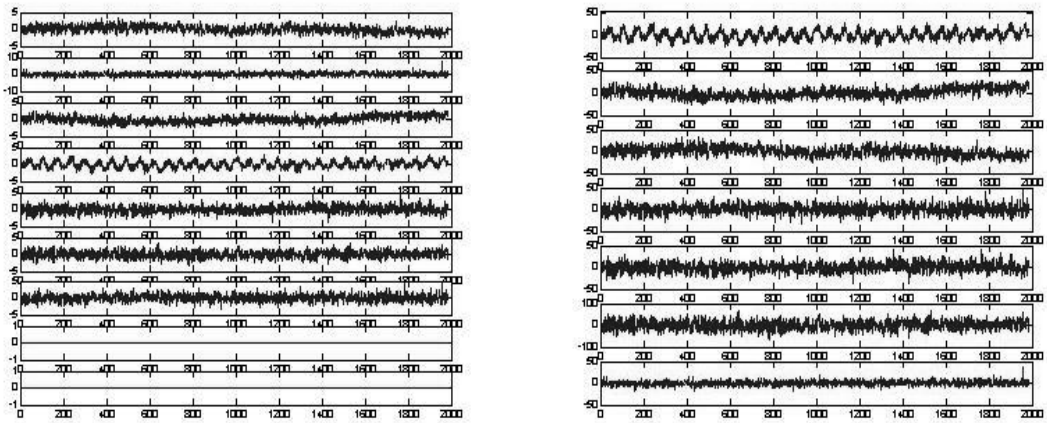

Figure 7. The generated ICs (left) without ordering. and (right) after ordering – Example 4.

REFERENCES

1. S. Ogawa, D.W. Tank, R. Menon, J.M. Ellermann, S.-G. Kim, H. Merkle, and K. Ugurbil, “Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging,” *Proc. Natl. Acad. Sci. USA* **89**, pp. 5951–5955, 1992.
2. K.J. Worsley and K.J.Friston, “Analysis of fMRI time-series revisited-again,” *Neuroimage* **2**, pp. 173-181, 1995.
3. W.W. Orrison, J.D. Lewine, J.A. Sanders, and M.F. Hartshorne, *Functional Brain Imaging*, Mosby-Year Book, St. Louis, 1995.
4. N. Lange, S. C. Strother, J. R. Anderson, F. A. Nielsen, A. P. Holmes, T. Kolenda, R. Savoy, and L. K. Hansen, “Plurality and resemblance in fMRI data analysis,” *Neuroimage* **10**, pp. 282-303, 1999.
5. R.L. Buckner, “Event-related fMRI and the hemodynamic response,” *Hum. Brain Mapp.* **6**, pp. 373-377, 1998.
6. L. K. Hansen, J. Larsen, F. A. Nielsen, S. C. Strother, E. Rostrup, R. Savoy, N. Lange, J. Sidtis, C. Svarer, and O. B. Paulson, "Generalizable patterns in neuroimaging: How many principal components?," *Neuroimage* **9**, pp. 534-544, 1999.
7. J. R. Moeller and S. C. Strother, "A regional covariance approach to the analysis of functional patterns in positron emission tomographic data," *J. Cereb. Blood Flow Metab*. **11**, pp. A121-A135, 1991.
8. V. Cherkassky and F. Mulier, *Learning From Data: Concepts, Theory, and Methods*, John Wiley & Sons, New York, 1998.
9. J. H. Friedman, "Exploratory projection pursuit," *J. of the American Statistical Association* **82**, pp. 249-266, 1987.
10. H. Hotelling, “Relations between two sets of variates,” *Biometrika* **28**, 321–377, 1936.
11. T.W. Anderson, “An introduction to multivariate statistical analysis,” John Wiley & Sons, New York, 1984.
12. M. Borga, *Learning multidimensional signal processing*, Ph.D. thesis, Linkoping University, 1998.
13. O. Friman, J. Cadefamn, P. Lundberg, M. Borga, and H. Knutsson, “Detection of neural activity in functional MRI using canonical correlation analysis”, *Magnetic Resonance in Medicine* **45**, pp. 323-330, 2001.
14. M. J. McKeown, S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," *Hum. Brain Mapp.* **6**, pp. 160-188, 1998.
15. S. LaConte, S.-C. Ngan, and X. Hu, "Enhancing functional paradigm specific idependent components with the AFRICA technique," *Proc. 2001 ISMRM Meeting*, 2001.
16. B.A. Ardekani, J. Kershaw, K. Kashikura, and I. Kanno, “Activation detection in functional MRI using subspace modeling and maximum likelihood estimation,” *IEEE Trans. Med. Imaging* **18**, no. 2, pp. 101-114, 1999.
17. A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, 2001.