Testing different ICA algorithms and connectivity analyses on MS patients*

Muthuraman M, Anjum T, Droby A, Fleischer V, Reitz S, Mideksa KG, Schmidt G, Zipp F, Groppa S

Abstract— Multiple sclerosis (MS) is a progressive neurological disorder that affects the central nervous system. Functional magnetic resonance imaging (fMRI) has been employed to track the course and disease progression in patients with MS. The two main aims of this study were to apply in a data-driven approach the independent component analysis (ICA) in the spatial domain to depict the active sources and to look at the effective connectivity between the identified spatial sources. Several ICA algorithms have been proposed for fMRI data analysis. In this study, we aimed to test two well characterized algorithms, namely, the fast ICA and the complex infomax algorithms, followed by two effective connectivity algorithms, namely, Granger causality (GC) and generalized partial directed coherence (GPDC), to illustrate the connections between the spatial sources in patients with MS. The results obtained from the ICA analyses showed the involvement of the default mode network sources. The connectivity analyses depicted significant changes between the two applied algorithms. The significance of this study was to demonstrate the robustness of the analyzed algorithms in patients with MS and to validate them before applying them on larger datasets of patients with MS.

I. Introduction

Multiple sclerosis (MS) is characterized by demyelination, axonal injury, and axonal loss in the central nervous system (CNS). These processes disrupt the communication between cortical and subcortical regions [1]. Conventional MRI only weakly correlates with the clinical course and evolving functional impairment. This can be partially improved by quantitative MR-based techniques such as diffusion weighted (DW) imaging [2] or functional MRI (fMRI) [4]. Commonly, reconstruction techniques in MRI acquire the data in complex frequency space, followed by an inverse discrete Fourier transformation into complex image space. However, most fMRI studies are based on magnitude images even though the data acquired is complex data [5]. Analysis of the complex fMRI data allows for improved sensitivity over that based on magnitude fMRI data. Independent component analysis (ICA) is a data-driven and not hypothesis-driven method that utilizes complex images to depict independent spatial sources in the analyzed data. We selected two ICA algorithms, namely, the fast ICA [6] and the complex infomax algorithm [7] to reconstruct the spatial sources in a data set of MS Patients.

We were interested in describing the effective connectivity between sources, so performed a connectivity analysis to obtain the exact hierarchy of the network of sources. We tested two different algorithms, namely, the Granger causality (GC) [8] and the generalized partial directed coherence (GPDC) [9]. In general, GC operates in the time domain whereas GPDC operates in the frequency domain. Here, both algorithms are applied in the frequency domain to allow comparison. Both GC and GPDC algorithms are based on multivariate autoregressive models and provide the direction of information flow between two or more signals.

In this study, we aimed to optimize distinguish the performance of two ICA algorithms followed by two connectivity algorithms, for use in future studies with larger datasets. Important features of fMRI signal course that differentiate MS patients from healthy subjects should be described in large cohorts of subjects in future studies. For the quantification of our results we estimate quantitative parameters, namely the number of activated voxels from the spatial sources, the strength of the peak voxel, and the euclidean distance between the identified peak voxels for the ICA algorithms, and the number of connections between the sources and the strength of the connection for the connectivity analysis.

II. Methods

A. Data acquisition and preprocessing

The study was performed in nine patients with multiple sclerosis (MS) (six female and three male; mean age 42±12.06 years). The study was acknowledged by the local ethics committee, University of Mainz, Germany and the patients gave their informed consent. fMRI data were obtained using a GE-EPI sequence (TR = 3000 ms, TE = 30

M. Muthuraman.,the Department of Neurology, Christian Albrechts University Kiel, 24105 Germany m.muthuraman@neurologie.uni-kiel.de,
A. Droby, V. Fleischer, F. Zipp, S. Groppa are with the Department of Neurology, Johannes Gutenberg University Mainz, 55131 Germany
S. Reitz is with the department of Neurology at the Goethe University Frankfurt am Main, 60325 Germany
T. Anjum, KG. Mideksa, G. Schmidt are with the Institute for Digital Signal Processing and System Theory, Faculty of Engineering, Christian Albrechts University Kiel, 24143 Germany

ms, flip angle = 90°, a matrix size of 64 × 64, spatial in-plane resolution of 3 mm, 45 slices with slice thickness of 2 mm (inter-slice gap of 1 mm). A series of 200 volumes were acquired, resulting in a scan duration of 10 min. Preprocessing was performed using statistical parametric mapping toolbox (SPM8). The preprocessing steps were realignment, normalization, and smoothing. The realignment was done by taking the initial image of each patient as the reference image to eradicate motion artifacts in the fMRI images due to head movement. The default normalization was done using the standard MNI template brain to remove the effect of different brain sizes in the group of patients to provide a standard platform for comparing the results from different groups. Smoothing was performed using a Gaussian kernel of width 8 mm to reduce noise and compensate for inaccuracies caused by the normalization.

B. Independent component analysis

ICA was conducted using the GIFT toolbox (http://mialab.mrn.org/software/gift). ICA is a signal processing tool used to transform an observed multidimensional random vector into components that are mutually independent. In simplest terms the ICA can be defined as:

$$y = As \tag{1}$$

where $y = (y_1, \ldots, y_m)$ is the vector of observed random variables, $s = (s_1, \ldots, s_n)$ is the vector of statistically independent latent variables referred to as the independent components, and X is an unknown constant mixing matrix. The fast ICA algorithm is based on the fixed point algorithm for complex signals [6]. The independent components s in are found by searching for a matrix A such that $s = A^H y$ up to some uncertainty. The algorithm searches for the maximum of $E\left\{T\left(\left|A^H y\right|^2\right)\right\}$ in which T is a smooth even function, A is an $n-$dimensional complex weighting vector and $E\left\{\left|A^H y\right|^2\right\} = 1$. The algorithm requires whitening of the data in which the observed variable y_{old} is linearly transformed to a zero-mean variable $y = Py_{old}$, $y = (y_{1q} + iy_{1i}, \ldots, y_{nq} + iy_{ni})$ such that $E\{yy^H\} = I$. The fixed point algorithm for one unit is defined as:

$$a^+ = E\left\{y\left(a^H y\right) * t\left(\left|a^H y\right|^2\right)\right\}$$
$$- E\left\{t\left(\left|a^H y\right|^2\right) + t\left|a^H y\right|^2 t'\left(\left|a^H y\right|^2\right)\right\} a \tag{2}$$

$$a_{new} = \frac{a^+}{\left\|a^+\right\|}$$

The one-unit algorithm can be extended to the estimation of the whole ICA transformation $s = A^H y$. The de-correlation is done to prevent different neurons from converging to the same maximum in the outputs $A_1^H y, \ldots, A_n^H y$. In order to estimate all the independent components simultaneously, it is preferable to use a symmetric de-correlation matrix. This can be written as follows:

$$A = A\left(A^H A\right)^{-1/2} \tag{3}$$

where $A = (a_1 \cdots a_n)$ is the matrix of the vectors.

The complex infomax algorithm starts by maximizing the entropy of the output, b_i , of a single layer neural network which can be written as [7]:

$$b_i = r \text{ if } v_i = Zy_i \tag{4}$$

where r is the weighting factor, v_i are the estimated sources and Z is the inverse of the mixing matrix. The entropy of a complex number is defined as the joint differential entropy of its real and imaginary parts [10]:

$$D_E(b) = D_E\left(b_R + jb_I\right) \hat{=} D_E(b_R, b_I) \tag{5}$$

which can be written in terms of the output probability density function, $f_y(y)$, which depends on the weight function z and the Jacobian matrix:

$$D_E(b) = E\left[\left|\frac{\partial b}{\partial v}\right|\right] - E\left[\ln f_y(y)\right] - \ln z \tag{6}$$

The complex nonlinearity can be defined as follows:

$$\Delta Z = \eta\left\{I + \left[I - \left(C_1(v_1), \ldots, C_N(v)\right)\right] v^H\right\} z \tag{7}$$

If we consider the data to be real then equation (7) can be reduced to the form:

$$\Delta Z = \eta\left\{I + (1 - 2b)v^{"}\right\} z \tag{8}$$

Equation (8) can be solved analytically using a non-bounded function. However, care should be taken for the domain usage due to the singularity and periodicity of the function [11].

C. Connectivity analysis

Both the connectivity analyses were done using the FNC toolbox (http://mialab.mrn.org/software/fnc/). Clive Granger did the practical implementation of the causality by using the concept of an autoregressive process [12]. We can think of two scenarios: First, the prediction of the time series x_1 is based on the present state information of x_1 and x_2. Second, the prediction of the future state of time series x_1 is based only on the present state of x_1. If the prediction using the first scenario is better than that of the second scenario, then time series x_2 "Granger causes" time series x_1. Mathematically, this can be described by the following sets of equations:

$$x_1(t+1) = a_{r1}x_1(t) + \varepsilon_1(t) \tag{9}$$
$$x_1(t+1) = a_{r2}x_1(t) + b_{r2}x_2(t) + \varepsilon_2(t) \tag{10}$$

In equations (9) and (10), $\varepsilon_{1,2}(t)$ is the residual error. The variance of $\varepsilon_2(t)$ needs to be smaller than the variance of $\varepsilon_1(t)$ in order for $x_2(t)$ to "Granger cause" $x_1(t)$. Furthermore, the extent of causality can be estimated as:

$$\lambda_{x_2 \to x_1} = \ln \frac{\mathrm{var}(\varepsilon_1(t))}{\mathrm{var}(\varepsilon_2(t))}. \qquad (11)$$

Here $\lambda_{x_2 \to x_1}$ is the magnitude of Granger causality from time series $x_2(t)$ to $x_1(t)$. In the case of a generalized partial directed coherence method, the basic formulation is taken from the partial directed coherence (PDC) as described in [13]. From equation (10) the multivariate coefficients are a_{rij}, if we take the Fourier transform we have the equation (12)

$$A_{rij}(\omega) = -\sum_{o=1}^{n} a_{rij}(r) e^{-i\omega o}. \qquad (12)$$

Finally, the performance of PDC is severely affected by the difference in magnitude of noise covariances. This shortcoming is overcome by the generalized partial directed coherence [9]. The denominator and numerator of the PDC formulation were both scaled by the individual variance of the time series, as shown below:

$$gPDC_{rij}(\omega) = \frac{\frac{1}{\sigma_i} A_{rij}(\omega)}{\sqrt{\sum_{k=1}^{N} \frac{1}{\sigma_k^2 A_{rkj}(\omega) A_{rkj}^*(\omega)}}} \qquad (13)$$

Here σ_i^2 is the variance of the innovation noise in time series x_i and k represents the overall mean of all the connections.

III. Results

Spatial and connectivity differences

Figure 1 shows axial slices from both algorithms separated by white dashed line boxes. In the case of the infomax algorithm, the left and right visual cortical areas were not found. In order to quantitatively compare the results from both algorithms, the three parameters were estimated. Firstly, the number of voxels activated in each of the 7 networks were compared with a non-parametric Wilcoxon Mann-Whitney test, which showed significant differences between the two algorithms. The number of voxels activated in the fast ICA algorithm was significantly higher ($p = 0.009$) than that in the infomax. The second parameter, location of the peak voxel,estimated by the euclidean distance between the identified mean voxels in the two algorithms, also showed differences between the two algorithms, as shown in Table 1. The ICA component signal strength from the fast ICA algorithm was found to be significantly higher than that from the infomax ($p = 0.006$).

Figure 2 shows the connectivity plots for both ICA algorithms and the two different connectivity algorithms. First, the total number of connections was estimated by counting the uni-directional connections as one and the bi-directional connections as two. In total, for the GC analysis of the fast ICA spatial sources, we found 10 connections, and the same analysis of the infomax sources also showed 10 connections. For the GPDC analysis, the total number of connections was 13 for both ICA algorithms, even though one spatial source was additionally identified with the fast ICA algorithm.

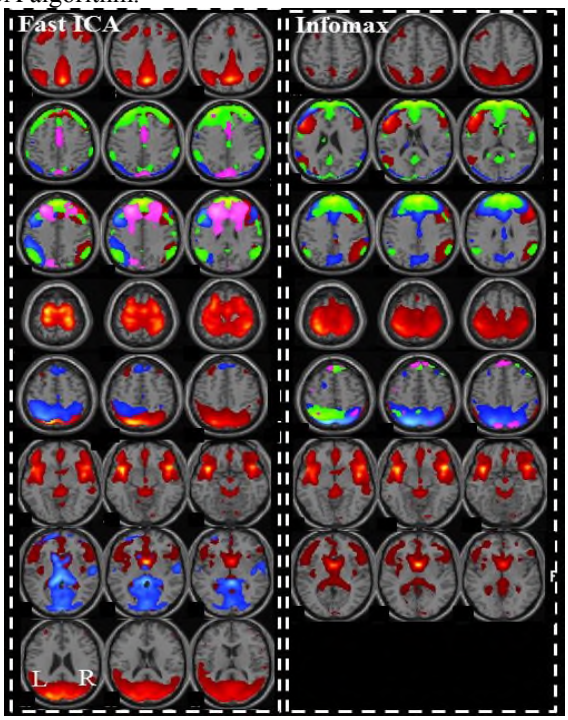

Figure 1. The first three columns show the results from the fast ICA algorithm on three contigous axial slices and the last three colums shows the results from the infomax algorithm. All seven default mode networks were found using fast ICA, whereas, infomax did not find the sources in the visual cortical areas. These results are shown by taking into account all 9 patients as a single group ICA analysis.

ICA	Number of Voxels		Maximum Voxel						Euclidean Distance (mm)
	Fast ICA (a)	Infomax (b)	Fast ICA (a)			Infomax (b)			
1	824	500	0	-55	39	7	-58	37	7.87
2	1840	2015	3	52	32	3	51	28	4.12
3	3247	2987	3	42	39	3	49	31	10.63
4	4712	4517	-21	-7	55	-34	-38	55	33.61
5	3574	3478	-25	-62	56	1	-60	56	26.07
6	3214	3410	-43	14	-9	44	14	-9	87
7	6478	5021	3	13	8	1	2	8	11.18
8	1475	-	-43	-55	23	-	-	-	-

Table 1. The table shows the values for the parameters: the number of voxels, the MNI co-ordinates of the activated peak voxels in the cluster, and the euclidean distance metric between the two algorithms for the peak voxel. a – fast ICA algorithm; b – infomax algorithm.

There were no differences in the number of connections between the two connectivity algorithms. However, the connection strength description was significantly stronger ($p = 0.002$) in the GPDC analysis compared to the GC analysis.

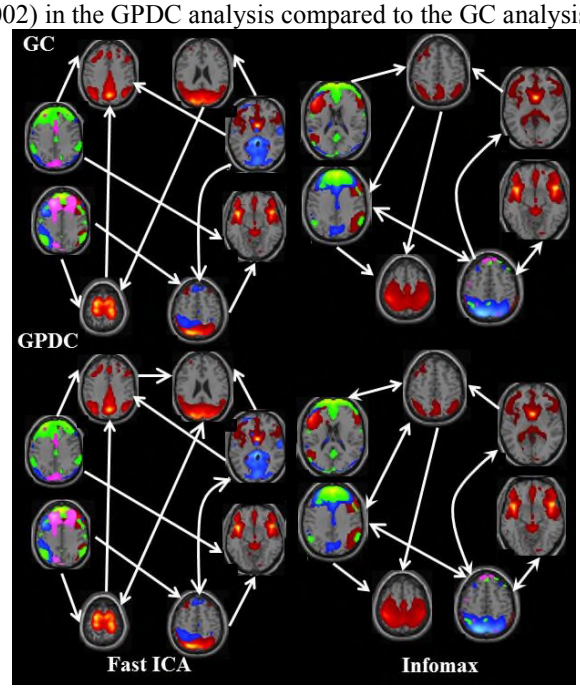

Figure 2. Shows the conectivity between the seven identified networks of spatial sources. The first column shows the results from the fast ICA and the second column shows the reuslts from the infomax algorithm. The first row depicts the results from the Granger causality (GC) analysis, and the second row shows the results from the generalized partial directed coherence (GPDC) analysis. The connections are shown for the group of 9 MS patients.

IV. DISCUSSION AND CONCLUSION

In this study, both ICA algorithms were found to be robust in showing similar networks of spatial sources, except for the visual cortical areas omitted by the infomax algorithm. The quantitative measures showed that the fast ICA had a higher number of activated voxels and the signal strength was also higher compared to that of the infomax algorithm. Both algorithms performed robustly on the complex fMRI data and not on the magnitude images as is usually the case in most fMRI studies [5]. Complex-data algorithms have an increased sensitivity as shown previously with simulated data [7], without imposing strong assumptions on the data. In the case of the connectivity analysis, both algorithms showed a similar network pattern, which indicates the robustness of both methods. However, the connection strength was higher in GPDC. The additional inclusion of the variance factor in GPDC could be the improvement of the detection power [9]. In conclusion, the outcome of this study shows that both ICA methods and the connectivity methods are robust. Significant differences in the quantitative measures were seen in the number of voxels activated and signal strength, indicating the superiority of the fast ICA method over the infomax. The limitation of this study is that these analyses were not tested on simulated data in order to obtain a real validation of the results. This could be a future direction before applying it to larger datasets.

ACKNOWLEDGMENT

Support from the German Research Council (Deutsche Forschungsgemeinschaft, DFG, SFB 855, Project D2) is gratefully acknowledged. This work was further supported by a grant from DFG (SFB-128, Project B5).

REFERENCES

[1] C. P. Kamm, B. M. Uitdehaag, and C. H. Polman, "Multiple sclerosis: current knowledge and future outlook," *Eur Neurol,* vol. 72, pp. 132-41, 2014.

[2] D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet, "MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders," *Radiology,* vol. 161, pp. 401-7, Nov 1986.

[3] D. Le Bihan, J. F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, *et al.*, "Diffusion tensor imaging: concepts and applications," *J Magn Reson Imaging,* vol. 13, pp. 534-46, Apr 2001.

[4] M. Filippi and M. A. Rocca, "Functional MR imaging in multiple sclerosis," *Neuroimaging Clin N Am,* vol. 19, pp. 59-70, Feb 2009.

[5] P. A. Bandettini, A. Jesmanowicz, E. C. Wong, and J. S. Hyde, "Processing strategies for time-course data sets in functional MRI of the human brain," *Magn Reson Med,* vol. 30, pp. 161-73, Aug 1993.

[6] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans Neural Netw,* vol. 10, pp. 626-34, 1999.

[7] V. Calhoun, T. Adali, G. Pearlson, and J. Pekar, "On complex infomax applied to functional MRI data," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, 2002, pp. I-1009-I-1012.

[8] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica,* vol. 37, pp. 424-438, 1969.

[9] L. A. Baccala and F. de Medicina, "Generalized Partial Directed Coherence," in *Digital Signal Processing, 2007 15th International Conference on*, 2007, pp. 163-166.

[10] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *Information Theory, IEEE Transactions on,* vol. 39, pp. 1293-1302, 1993.

[11] T. Kim and T. Adali, "Complex backpropagation neural network using elementary transcendental activation functions," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, 2001, pp. 1281-1284 vol.2.

[12] S. L. Bressler and A. K. Seth, "Wiener-Granger causality: a well established methodology," *Neuroimage,* vol. 58, pp. 323-9, Sep 15 2011.

[13] L. A. Baccala and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biol Cybern,* vol. 84, pp. 463-74, Jun 2001.