

Functional connectivity analysis of fMRI data based on regularized multiset canonical correlation analysis

Filip Deleus*, Marc M. Van Hulle

Katholieke Universiteit Leuven, Medical School - Labo Neuro- en Psychofysiologie, Campus Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium

ARTICLE INFO

Article history:

Received 22 September 2010

Received in revised form 4 November 2010

Accepted 5 November 2010

Keywords:

fMRI

Functional connectivity

Canonical correlation analysis

ABSTRACT

In this paper we describe a method for functional connectivity analysis of fMRI data between given brain regions-of-interest (ROIs). The method relies on nonnegativity constrained- and spatially regularized multiset canonical correlation analysis (CCA), and assigns weights to the fMRI signals of the ROIs so that their representative signals become simultaneously maximally correlated. The different pairwise correlations between the representative signals of the ROIs are combined using the *maxvar* approach for multiset CCA, which has been shown to be equivalent to the generalized eigenvector formulation of CCA. The eigenvector in the *maxvar* approach gives an indication of the relative importance of each ROI in obtaining a maximal overall correlation, and hence, can be interpreted as a functional connectivity pattern of the ROIs. The successive canonical correlations define subsequent functional connectivity patterns, in decreasing order of importance. We apply our method on synthetic data and real fMRI data and show its advantages compared to unconstrained CCA and to PCA. Furthermore, since the representative signals for the ROIs are optimized for maximal correlation they are also ideally suited for further effective connectivity analyses, to assess the information flows between the ROIs in the brain.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The fMRI signal is generally regarded as a proxy for the underlying local neuronal activity (Logothetis et al., 2001). However, brain function is not only the result of activity within different specialized and isolated brain areas, but is also due to interactions between brain areas (Ramnani et al., 2004). Methods for the analysis of these interactions, or connectivities, are based on the premise that functionally interacting regions will show correlated patterns of activity (Young et al., 1994). Hence, methods for connectivity analysis with fMRI data are always based on some measure related to the correlation between two or more fMRI time signals: pairwise correlations in seed based approaches; measures based on all the pairwise correlations together like in Principal Component Analysis (PCA) (Friston et al., 1993), Generalised Eigenimage Analysis (Friston, 1997), or Independent Component Analysis (ICA) (McKeown et al., 1998); correlations at different time lags leading to the notion of Granger causality (Roebroeck et al., 2005); multiple regression with or without time lags as in Multivariate Autoregressive Modelling (MAR) (Harrison et al., 2003); predicted versus observed correlations in Structural Equation Models (SEMs) (Bullmore et al., 2000;

McIntosh and Gonzalez-Lima, 1994); conditional correlations for structure learning with Bayesian or Dynamic Bayesian networks (Zheng and Rajapakse, 2006; Rajapakse and Zhou, 2007); partial correlations in the concentration matrix (Marrelec et al., 2006, 2007). This ubiquitous modelling of within brain correlations in connectivity studies is in contrast with what is done in activation studies, where the relationship of each individual fMRI signal to the experimental paradigm is modelled.

Friston (1994) makes a distinction between functional and effective connectivity analysis. Functional connectivity is mainly model-free and data-driven, based on seed-based correlations or PCA/ICA-based methods. By contrast, effective connectivity is more directly concerned with the way how the information flows through the brain and starts therefore from a given model of underlying structural connectivity (based on anatomical and/or neurophysiological knowledge).

Functional connectivity analysis can be performed on the level of the individual voxels or on the level of regions-of-interest (ROIs). The advantage of the ROIs is that results are reached in terms of brain regions directly, allowing for a more meaningful interpretation. Methods for effective connectivity analysis are mostly performed on the ROI level, mainly because of the requirement of prior model specification and/or the need of the inverse correlation matrix, which becomes rank deficient in the typical case that there are many more voxels than signal samples (Worsley et al., 1998; Friston et al., 1995). The ROIs can be defined by multi-

* Corresponding author. Tel.: +32 16 34 59 61; fax: +32 16 34 59 60.

E-mail addresses: Filip.Deleus@med.kuleuven.be (F. Deleus), Marc.Vanhulle@med.kuleuven.be (M.M. Van Hulle).

ple criteria: cyto- and myeloarchitecture, anatomical connections, the topographic organization, and functional properties (Orban and Vanduffel, 2007), or by data-driven methods such as cluster analysis (Baumgartner et al., 1997; Penny and Friston, 2003; Thirion et al., 2006), and region-growing methods (Lu et al., 2004; Bellec et al., 2006). Despite the advantages of the ROIs, they may, on the other hand, also cause a lack of robustness in the connectivity results: the results could substantially change due to small changes in the positions of the voxels. This lack of robustness has several causes: misalignment, movement, field distortions and noise, and morphological differences in individual brains. Therefore, it is important to apply a method that is robust when defining representative signals for the ROIs.

Three methods are frequently applied for defining representative signals: (1) an individual voxel is identified within each ROI, e.g., the peak voxel of activation (Jennings et al., 1998; Gonçalves et al., 2001); (2) the average signal is taken across all voxels of a ROI (e.g., Greicius et al., 2003); (3) PCA is performed on the signals within a ROI and the projection on the first principal component defines the ROI's representative signal. The first method is biased by the selected contrast, and the second method may be distorted by non-informative signals, such as noise (Bellec et al., 2006; Kohler et al., 1998). It also assumes functionally homogeneous ROIs, since opposite effects, like activation and deactivation, tend to cancel out when averaging over the ROI (Poldrack, 2007). The third method is mostly used in fMRI connectivity studies (Bullmore et al., 2000; Grafton et al., 1994; Büchel and Friston, 1997; Mechelli et al., 2002; Gavrilescu et al., 2008). However, PCA often results in diffused weightings, i.e. non-zero weights assigned to the majority of the voxels (Zass and Shashua, 2006). Furthermore, in non-homogeneous brain areas, with different activation profiles in different subregions, only the subregion leading to the representative signal with the maximum variance will be selected. This problem can be solved by selecting a number of principal components, but then it becomes more complicated to establish how these signals relate to those of other ROIs. Even worse, opposite effects may be combined by weights with opposite signs, in order to increase the variance of the representative signal.

We propose a method for functional connectivity analysis, which starts from a set of given ROIs, and which delivers three types of output: it identifies the functional connectivity pattern of the ROIs that leads to the maximal overall correlation. Therefore, it first detects the correlated voxels within the ROIs, and it uses them to construct representative signals which admit the best possible overall correlation. Furthermore, the spatial patterns of the correlated voxels within the ROIs may be used as a starting point in a method to refine the given ROIs (e.g., Deleus and Van Hulle, 2009), and the representative signals can also be used in further effective connectivity studies.

Our method is based on nonnegativity constrained, spatially regularized, multiset canonical correlation analysis (CCA). CCA (Hotelling, 1936) is a multivariate statistical technique to identify and quantify the correlation between two sets of variables. CCA has already been used in fMRI for activation detection (Friman et al., 2001, 2002, 2003), employing a two-set CCA in each voxel independently, and in a classification task for categorizing fMRI patterns that correspond to different emotions (Hardoon et al., 2007). Recently, CCA has also been used to identify signal integration (Bruguier et al., 2008), and to combine the different principal components from 2 ROIs in a Granger causality study (Sato et al., 2010).

Nonnegativity constraints have been applied in fMRI for the parcellation of cortical areas based on replicator dynamics (Lohmann and Bohn, 2002; Neumann et al., 2006), and for a functional segmentation based on nonnegative PCA (Ng et al., 2009). Nonnegativity constraints in CCA were introduced by Das and Sen (1994),

and applied in fMRI for improving the activation detection (Friman et al., 2003; Ragnehed et al., 2009). CCA in its conventional form has weak spatial specificity, and an assignment scheme to improve the spatial specificity in fMRI activation detection with CCA was proposed in Nandy and Cordes (2004). We will use a regularization term for spatial improvement, which is similar to the ridge-like regularization term applied in Hardoon et al. (2004), Shawe-Taylor and Cristianini (2004) and Takane et al. (2008) to stabilize the CCA-solution. It may be emphasized that the interpretation of the coefficients is a chief motivation for regularized CCA. CCA has been generalized to multiple datasets by Kettenring (1971), and multi-set CCA has been applied in neuroimaging to perform a joint blind source separation in multi-subject fMRI data (Li et al., 2009), and for the fusion of simultaneously acquired fMRI and EEG data (Correa et al., 2010).

This paper is organised as follows. We describe our method in Section 2, first for 2 sets without constraints (Section 2.1), then with nonnegativity constraints (Section 2.2) and spatial regularization (Section 2.3), for multiple sets (Section 2.4) and for multiple modes (Section 2.5). The interpretation in terms of functional connectivity patterns is given in Section 2.4 and the statistical evaluation is described in Section 2.6. In Section 3 we apply the method on synthetic data and compare it with unconstrained and unregularized CCA and with PCA. Next we apply the method on monkey fMRI data of an action observation and a motion localizer study and we discuss the functional connectivity patterns that were found.

2. Methods

2.1. Unconstrained CCA on 2 ROIs

By convention, a lowercase bold symbol denotes a vector, an uppercase bold symbol a matrix, and a lowercase non-bold symbol a scalar. Let $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \dots \mathbf{x}_{n_x}]$ be a $(t \times n_x)$ -dimensional data matrix consisting of n_x fMRI signals of length t , where each signal \mathbf{x}_i corresponds to one voxel in ROI x . Similarly, let $\mathbf{Y} = [\mathbf{y}_1 \ \mathbf{y}_2 \dots \mathbf{y}_{n_y}]$ be the fMRI signals in ROI y . Each signal is normalized to zero mean and unit variance. With CCA, we obtain the weight vectors \mathbf{w}_x and \mathbf{w}_y so that the correlation, ρ , between the two canonical variates, \mathbf{x} and \mathbf{y} , is maximal, with

$$\mathbf{x} = w_{x,1}\mathbf{x}_1 + w_{x,2}\mathbf{x}_2 + \dots + w_{x,n_x}\mathbf{x}_{n_x} = \mathbf{X}\mathbf{w}_x \quad (1)$$

$$\mathbf{y} = w_{y,1}\mathbf{y}_1 + w_{y,2}\mathbf{y}_2 + \dots + w_{y,n_y}\mathbf{y}_{n_y} = \mathbf{Y}\mathbf{w}_y \quad (2)$$

The CCA problem can be formalized as the following constrained optimization problem:

$$\max_{\mathbf{w}_x, \mathbf{w}_y} \rho = \mathbf{w}_x' \frac{\mathbf{X}'\mathbf{Y}}{t-1} \mathbf{w}_y \quad (3)$$

$$\text{s.t. } \mathbf{w}_x' \frac{\mathbf{X}\mathbf{X}'}{t-1} \mathbf{w}_x = 1 \quad (4)$$

$$\mathbf{w}_y' \frac{\mathbf{Y}'\mathbf{Y}}{t-1} \mathbf{w}_y = 1 \quad (5)$$

Due to Hotelling (1936) this problem can be reduced to two eigenvalue problems (see Appendix A), and the linear combinations $\mathbf{x} = \mathbf{X}\mathbf{w}_x$ and $\mathbf{y} = \mathbf{Y}\mathbf{w}_y$ can be considered as the representative time signals for ROI x and ROI y , respectively. The derivation of the eigenvector solution, see (25), shows that $\mathbf{w}_x = (1/\rho)(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})$ which is the ordinary-least-squares solution (OLS) in the regression from \mathbf{X} to \mathbf{y} , up to a scaling factor ρ . Thus, the CCA problem can also be solved by an iterative least squares algorithm: $\mathbf{x}'\mathbf{y}$ is maximized by optimizing \mathbf{w}_x for a given \mathbf{w}_y , next, \mathbf{w}_y is optimized for a given \mathbf{w}_x , and so on, and whereby \mathbf{x} and \mathbf{y} are rescaled to unit variance, at each iteration. This iterative formulation is effectively an 'Alternating Least Squares' (ALS) method (Cichocki and Zdunek, 2007),

and enables us to introduce two extensions for the CCA-problem, namely, nonnegative weights and spatial regularization. For a discussion about the online realization of the CCA approach (without constraints or regularization) we refer to (Liu et al., 2007).

2.2. Nonnegativity

The weight vectors \mathbf{w}_x and \mathbf{w}_y in (3) may contain positive as well as negative components. Hence, a representative signal cannot be interpreted as a weighted average of the original time signals due to the possibility of negative weights. Therefore, we impose nonnegativity constraints: $\mathbf{w}_x \geq 0$ and $\mathbf{w}_y \geq 0$, and the voxels with a nonnegative weight can then be considered as the refined ROI within an original ROI mask. Since we solve the CCA problem by an iterative least squares formulation, we can revert to standard nonnegative least squares (NNLS)-algorithms (Lawson and Hanson, 1974; Bro and De Jong, 1997) to find the nonnegative weight vector at each iteration. In the remainder of this paper we refer to $\text{nnls}(\mathbf{X}\mathbf{X}, \mathbf{X}\mathbf{y})$ as the nonnegative least squares solution where the the unconstrained solution would be found as $(\mathbf{X}\mathbf{X})^{-1}(\mathbf{X}\mathbf{y})$.

2.3. Spatial regularization

The nonnegative weight vectors define subregions in the pre-specified ROI masks, however, without considering the spatial information: the fact that two voxels are neighbours is ignored when optimizing their weights. Two observations motivate a spatial regularization of the weight vectors. First, the time signals of neighbouring voxels in fMRI data are correlated with each other, hence, they are not independent in space (Biswal et al., 1995; Zarahn et al., 1997). Second, in fMRI data obtained during a cognitive task, an area of real activation is likely to be characterized by more than one single voxel. It is assumed that neural activity tends to stimulate signal changes over contiguous groups of voxels (Forman et al., 1995), or, in other words, fMRI activation is more likely to occur in clusters of several contiguous voxels rather than in isolated voxels (Tononi et al., 1998; Katanoda et al., 2002). Despite of the presence of these spatial correlations in the fMRI signals, the weight vectors \mathbf{w}_x and \mathbf{w}_y do not reflect this spatial organisation, instead they tend to define scattered weight patterns with high variances on each individual weight value. This phenomenon is known as the multicollinearity problem in regression analysis: the matrices $(\mathbf{X}\mathbf{X})$ and $(\mathbf{Y}\mathbf{Y})$ become more singular with higher spatial correlations, and the inverse operations on them, $(\mathbf{X}\mathbf{X})^{-1}$ and $(\mathbf{Y}\mathbf{Y})^{-1}$, lead to increased variances. In order to obtain weight vectors which are more compliant with the spatial structure of the data, we add a penalty term for spatial regularization to the goal function of the optimization: for each pair of adjacent voxels, i and j , we add the penalty term $\gamma(\mathbf{w}_{x,i} - \mathbf{w}_{x,j})^2$ when calculating the regression coefficients \mathbf{w}_x . Thus, for example for ROI $_x$, we add $\gamma \sum_{\{i,j\} \in N_x} (\mathbf{w}_{x,i} - \mathbf{w}_{x,j})^2$, where N_x is the set of all pairs of adjacent voxels in ROI $_x$. The weight vector is then found as $\mathbf{w}_x = \text{nnls}(\mathbf{X}\mathbf{X} + \gamma \cdot \mathbf{G}_x, \mathbf{X}\mathbf{y})$, up to a scaling factor. The spatial regularization matrix \mathbf{G}_x is formed by starting from the $[n_x \times n_x]$ dimensional null matrix and adding 1 to $\mathbf{G}_x(i, i)$ and $\mathbf{G}_x(j, j)$ and subtracting 1 from $\mathbf{G}_x(i, j)$ and $\mathbf{G}_x(j, i)$ whenever i and j are adjacent voxels in ROI $_x$. The factor γ controls the degree of spatial regularization. In Section 2.6.3, we describe a bootstrapping procedure to find a suitable value for γ .

2.4. Nonnegative multiset CCA

We now turn to the case that $m > 2$ brain regions are involved. In this case, all the pairwise correlations between the representative signals have to be optimized simultaneously. This problem

is known as multiset CCA (Kettenring, 1971; Horst, 1961; Van de Geer, 1984), and the methods described in the literature differ in the way they combine all the pairwise correlations into one measure of overall correlation. Following the terminology of Kettenring (1971), the 5 most important objective functions to maximize are: (1) *sum-corr*: the sum of all the elements in the correlation matrix; (2) *ssqcor*: the sum of the squared elements in the correlation matrix; (3) *maxvar*: the largest eigenvalue of the correlation matrix; (4) *minvar*: the smallest eigenvalue of the correlation matrix; (5) *genvar*: the determinant of the correlation matrix. We choose the *maxvar* approach since it can be considered as a natural extension of CCA to multiple sets (Borga, 1998; Via et al., 2007).

To see this, notice that for the two data sets case the two constraints (4)–(5) are equivalent to one constraint: $(1/2)(\mathbf{w}'_x(\mathbf{X}\mathbf{X}/(t-1))\mathbf{w}_x + \mathbf{w}'_y(\mathbf{Y}\mathbf{Y}/(t-1))\mathbf{w}_y) = 1$ (since $\rho_x = \rho_y$ in (24)). With m brain regions and corresponding data sets $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$, the optimization problem in (3)–(5) can then be extended to:

$$\max_{\mathbf{h}_1, \dots, \mathbf{h}_m} \rho = \sum_{r \neq k}^m \mathbf{h}'_r \frac{\mathbf{X}'_r \mathbf{X}_k}{t-1} \mathbf{h}_k \quad (6)$$

$$\text{s.t.} \quad \frac{1}{m} \sum_r^m \mathbf{h}'_r \frac{\mathbf{X}'_r \mathbf{X}_r}{t-1} \mathbf{h}_r = 1 \quad (7)$$

which, using the method of Lagrange multipliers, can be solved by the following generalized eigenvalue (GEV) problem:

$$(\mathbf{A} - \mathbf{B})\mathbf{h} = \rho \mathbf{B}\mathbf{h} \quad (8)$$

with

$$\mathbf{A} = \begin{bmatrix} \mathbf{X}'_1 \mathbf{X}_1 & \dots & \mathbf{X}'_1 \mathbf{X}_m \\ \vdots & \ddots & \vdots \\ \mathbf{X}'_m \mathbf{X}_1 & \dots & \mathbf{X}'_m \mathbf{X}_m \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{X}'_1 \mathbf{X}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{X}'_m \mathbf{X}_m \end{bmatrix}, \quad \mathbf{h} = \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_m \end{bmatrix}$$

Let $\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_m]$ be the m normalized representative time signals of length t , with $\mathbf{z}_r = \mathbf{X}_r \mathbf{w}_r$ and $(\mathbf{z}'_r \mathbf{z}_r)/(t-1) = 1$, ($r = 1 \dots m$), and let $\mathbf{s} = \mathbf{Z}\mathbf{v}$ be a linear combination of these representative signals. The *maxvar* approach seeks the vector \mathbf{v} which maximizes the variance of \mathbf{s} , or in other words, which maximizes $\lambda = \mathbf{v}'(\mathbf{Z}\mathbf{Z}/(t-1))\mathbf{v}$ under the constraint that $\mathbf{v}'\mathbf{v} = 1$. The vector \mathbf{v} is then found as the first eigenvector of $\mathbf{Z}'\mathbf{Z}$. After some algebraic manipulations, λ can be rewritten as:

$$\lambda = \sum_{r \neq k}^m v_r \mathbf{w}'_r \frac{\mathbf{X}'_r \mathbf{X}_k}{t-1} \mathbf{w}_k v_k + \sum_r^m v_r \mathbf{w}'_r \frac{\mathbf{X}'_r \mathbf{X}_r}{t-1} \mathbf{w}_r v_r \quad (9)$$

With $\mathbf{h}_r = v_r \mathbf{w}_r$, the first term in (9) is equivalent to (6), and, since $\mathbf{w}'_r(\mathbf{X}'_r \mathbf{X}_r)/(t-1)\mathbf{w}_r = 1$ and $\mathbf{v}'\mathbf{v} = 1$, the second term is equivalent to (7). Due to the equivalence of the GEV formulation and the *maxvar* approach, the eigenvector \mathbf{v} can be interpreted as a functional connectivity pattern between the m brain regions. To see this, note that each \mathbf{w}_r is equal to \mathbf{h}_r , but after a rescaling which gives unit variance to \mathbf{z}_r . This scaling factor is defined by v_r . Thus, the pairwise correlations (i.e. between normalized signals) are weighted by the product of their corresponding elements in \mathbf{v} in order to get a maximum overall correlation. This overall correlation can also be defined as the variance of the total signal $\mathbf{s} = \mathbf{Z}\mathbf{v}$, and each individual element v_r provides an indication of the relative importance of each region in getting a maximum overall correlation, or a maximum overall functional connectivity, between the m brain regions.

We incorporate the nonnegativity constraints and the spatial regularization into the maxvar approach. The optimization problem then becomes:

$$\max_{\mathbf{v}, \mathbf{w}_1 \dots \mathbf{w}_m} \mathbf{v}' \mathbf{Q} \mathbf{v} - \gamma \sum_{r=1}^m \sum_{(i,j) \in N_r} (\mathbf{w}_{r,i} - \mathbf{w}_{r,j})^2 \quad (10)$$

$$\text{s.t. } \mathbf{w}_r' \frac{\mathbf{X}_r' \mathbf{X}_r}{t-1} \mathbf{w}_r = 1, \quad r = 1 \dots m \quad (11)$$

$$\mathbf{w}_r \geq 0, \quad r = 1 \dots m \quad (12)$$

$$\mathbf{v}' \mathbf{v} = 1 \quad (13)$$

$$\mathbf{v} \geq 0 \quad (14)$$

with:

$$\mathbf{Q} = \frac{\mathbf{Z}' \mathbf{Z}}{t-1} \quad (15)$$

and it can be solved by iteratively seeking \mathbf{v} given $\mathbf{w}_1, \mathbf{w}_2 \dots \mathbf{w}_m$, and by seeking \mathbf{w}_r given \mathbf{v} and the other $(m-1)$ weight vectors. We now further focus on these two subproblems.

2.4.1. Optimal vector \mathbf{v}

For given weight vectors \mathbf{w}_r , the matrix \mathbf{Q} is constant and the optimization problem becomes:

$$\max_{\mathbf{v}} \lambda = \mathbf{v}' \mathbf{Q} \mathbf{v} \quad (16)$$

$$\text{s.t. } \mathbf{v}' \mathbf{v} = 1 \quad (17)$$

$$\mathbf{v} \geq 0 \quad (18)$$

Without the nonnegativity constraints on \mathbf{v} , λ and \mathbf{v} correspond to the largest eigenvalue and the corresponding eigenvector of \mathbf{Q} , respectively. Notice that, if all the elements in this first eigenvector would be nonnegative, the constraints in (18) are non-binding, and they could be safely removed from the formulation. Notice also that, if we set an element in \mathbf{v} equal to 0, say $v_r = 0$, the object value $\mathbf{v}' \mathbf{Q} \mathbf{v}$ can be calculated as $\mathbf{v}_J' \mathbf{Q}_{(J,J)} \mathbf{v}_J$ with J the set of all indices $[1 \dots m]$ except r , which is again a (smaller) eigenvalue problem. These considerations prompt for an active set algorithm (Nocedal and Wright, 2006) to solve (16)–(18).

2.4.2. Optimal weight vector \mathbf{w}_r

Let J be again the set of all indices $[1 \dots m]$ except r , and let \mathbf{Z}_J consist of the J columns in \mathbf{Z} and \mathbf{v}_J consist of the corresponding elements in \mathbf{v} . The goal function of the optimization problem with respect to \mathbf{w}_r , but for the moment without the spatial regularization term, is then:

$$\begin{aligned} \max_{\mathbf{w}_r} \lambda &= \mathbf{v}' \mathbf{Q} \mathbf{v} \\ &= [\mathbf{v}_r \mathbf{v}_J'] \begin{bmatrix} \mathbf{w}_r' \mathbf{X}_r' \mathbf{X}_r \mathbf{w}_r & \mathbf{w}_r' \mathbf{X}_r' \mathbf{Z}_J \\ \mathbf{Z}_J' \mathbf{X}_r \mathbf{w}_r & \mathbf{Z}_J' \mathbf{Z}_J \end{bmatrix} \begin{bmatrix} \mathbf{v}_r \\ \mathbf{v}_J \end{bmatrix} \\ &= v_r^2 (\mathbf{w}_r' \mathbf{X}_r' \mathbf{X}_r \mathbf{w}_r) + 2v_r \mathbf{v}_J' \mathbf{Z}_J' \mathbf{X}_r \mathbf{w}_r + \dots \\ &\quad + \mathbf{v}_J' \mathbf{Z}_J' \mathbf{Z}_J \mathbf{v}_J \end{aligned} \quad (19)$$

The last term is a constant, and the first term is constant under the constraint that $\mathbf{w}_r' \mathbf{X}_r' \mathbf{X}_r \mathbf{w}_r = t-1$. With $\mathbf{s}_r = \mathbf{Z}_J \mathbf{v}_J$, the second term represents the correlation between $\mathbf{X}_r \mathbf{w}_r$ and \mathbf{s}_r . Due to the correspondence between correlation and regression, \mathbf{w}_r boils down to the solution of the regression from \mathbf{X}_r to \mathbf{s}_r , up to a scaling factor. The nonnegativity constraints and the spatial regularization penalty term can then be easily added as described above.

The optimal solution is found by iteratively updating one weight vector \mathbf{w}_r , then the eigenvector \mathbf{v} , then another weight vector, and so on, until convergence. Corresponding to constraint (11), the

representative signals in \mathbf{Z} are rescaled to unit variance for each iteration.

2.5. Successive canonical correlations

Up to this point, the described procedure finds only the first canonical correlation, or also called the first mode. In the GEV-formulation in (8) the successive eigenvalues and eigenvectors define the successive modes. We use the superscript $[\ell]$ to denote the different modes, e.g., $\mathbf{Z}^{[\ell]}$ are the representative signals for mode ℓ . When the successive eigenvectors are collected in the matrix \mathbf{H} , the orthogonality between the successive modes in (8) is defined by $\mathbf{H}' \mathbf{B} \mathbf{H} = \mathbf{I}$. With two data sets, this is equivalent to $\mathbf{z}_k^{[f]'} \mathbf{z}_r^{[\ell]} = 0$, with $k, r = 1 \dots 2$ and $f \neq \ell$. Thus, the representative signals for different modes (within or between ROIs) are orthogonal to each other. For the multiset case, the orthogonality constraints in the GEV formulation are defined by the summations, like in (7).

In order to determine the successive modes in our iterative approach, we define a deflation procedure, which also accounts for the nonnegativity constraints. As described above for the first mode, when calculating weight vector $\mathbf{w}_r^{[1]}$, we take the regression from \mathbf{X}_r to $\mathbf{s}_r^{[1]}$. For the successive modes $\ell > 1$, we first define the reduced signal $\tilde{\mathbf{s}}_r^{[\ell]}$ as the subtraction of $\mathbf{s}_r^{[\ell]}$ minus the representative signals obtained in the previous modes, but using nonnegative regression. Next, we take the nonnegative and spatially regularized regression from \mathbf{X}_r to this reduced signal. More formally, for the calculation of $\mathbf{w}_r^{[\ell]}$, let $J = [1 \dots m] \setminus r$ and $\mathbf{s}_r^{[\ell]} = \mathbf{Z}_J^{[\ell]} \mathbf{v}_J^{[\ell]}$, and let $\mathbf{Z}^{[L]}$ with $L = 1 \dots \ell - 1$ be the representative signals from all the ROIs in all the previous modes. We calculate then $\mathbf{b}_r^{[\ell]} = \text{nnls}(\mathbf{Z}^{[L]'} \mathbf{Z}^{[L]}, \mathbf{Z}^{[L]'} \mathbf{s}_r^{[\ell]})$, and we define the reduced signal $\tilde{\mathbf{s}}_r^{[\ell]} = \mathbf{s}_r^{[\ell]} - \mathbf{Z}^{[L]} \mathbf{b}_r^{[\ell]}$. Due to the nonnegativity constraints, the orthogonality constraints cannot be fulfilled, instead, nonnegativity enforces the successive representative signals not to be positively correlated, i.e., $\mathbf{z}_k^{[f]'} \mathbf{z}_r^{[\ell]} \leq 0$ for $k, r = 1 \dots m$ and $f \neq \ell$.

2.6. Performance measures and statistical evaluation

Three measures are evaluated: (1) the total correlation between the m brain regions, for each mode; (2) the importance of the individual ROIs in the different modes; (3) the weights of the individual voxels. An important issue is the statistical evaluation of these measures. In unconstrained CCA and with independent observations, the asymptotic distribution of the canonical correlations is chi square (Anderson, 1984). However, due to the nonnegativity constraints, spatial regularization, multiple sets, and autocorrelated noise structure in the fMRI data, an analytical solution for testing the significance of the different measures, based on their asymptotic properties, is not likely to be possible. Thus, we have to resort to bootstrap methods such as the one proposed in Das and Sen (1996) and Efron and Tibshirani (1993), for obtaining confidence intervals, and for hypothesis testing.

Bootstrap methods require random samples of data. Thereby, it is crucial to preserve the temporal and spatial dependence structure in the data. We use the stationary bootstrap method (Politis and Romano, 1994), in which blocks of random length (following a geometric distribution) are resampled with replacement. The stationary bootstrap is applied to the indices of the time signals, leading then to a reordering of them, to obtain random samples of the same size as the original data. The same reordering is always applied on all the signals within one ROI, so as to preserve the spatial dependence structure within the ROIs, but one or different orderings are applied on different ROIs (keeping or destroying the correlations between the ROIs) for obtaining confidence intervals or for null-hypothesis testing, respectively. In order to enforce

that one mode captures the same information in the different randomisations, we use the parameters obtained on the original data as starting solutions for the iterative algorithm on the bootstrap samples. The three measures, and their statistical evaluation, are further described in the next three paragraphs.

2.6.1. Canonical correlations: ρ_{tot}

For the evaluation of the canonical correlations, the penalty term in (10) is not taken as part of the correlation measure. The penalty term served as a means to obtain smoother weight patterns, which are then used to calculate the signals \mathbf{Z} and their overall correlation. Incorporating the penalty term in the correlation measure would obstruct a fair comparison of our results with those of other methods. According to (11) we rescale the signals \mathbf{Z} to unit variance before calculating their overall correlation. For each mode ℓ , we obtain then $\lambda^{[\ell]} = \mathbf{v}(\mathbf{Z}'\mathbf{Z}/(t-1))\mathbf{v}$, which ranges between 1 and m , due to (11) and (13). We define the total correlation for mode ℓ as $\rho_{tot}^{[\ell]} = (\lambda^{[\ell]} - 1)/(m - 1)$, which lies between 0 and 1.

For testing the hypothesis $H_0 : \rho_{tot}^{[\ell]} = 0$, we generate a number of bootstrap samples by applying different reorderings on all the ROIs, and we apply CCA and calculate $\rho_{tot}^{[\ell]*}$ for each bootstrap sample. The p -value for rejecting H_0 is given by the fraction of times that $\rho_{tot}^{[\ell]*} \geq \rho_{tot}^{[\ell]}$.

2.6.2. Importance of the different ROIs: v_r and ρ_r

As stated above, the eigenvector \mathbf{v} gives an indication of the relative importance of each ROI in obtaining a maximum overall correlation between the ROIs. However, as the squared elements of \mathbf{v} sum up to 1, each v_r depends on the number of ROIs, m . A more intuitive measure is obtained as $\rho_r = \text{corr}(\mathbf{z}_r, \mathbf{s}_r)$, with $\mathbf{s}_r = \mathbf{Z}_r\mathbf{v}_r$, and $J = [1 \dots m] \setminus r$. This ρ_r represents the correlation between the r th ROI's representative signal and the total signal formed by the other ROIs.

For clarifying the interpretation of these different measures, consider 3 signals \mathbf{z}_r ($r = 1 \dots 3$), in two extreme cases, first, they are totally uncorrelated, and second, they are identical. In both cases, $v_r = 1/\sqrt{3}$ as the 3 signals contribute equally to the variance of \mathbf{s} , no matter its value. This variance is equal to 1 in the first case, and equal to 3 in the second case. The individual correlations $\text{corr}(\mathbf{z}_r, \mathbf{s})$ with $\mathbf{s} = \mathbf{Z}\mathbf{v}$, thus, without excluding component r , would be equal to $1/\sqrt{3}$ and to 1, respectively. When using \mathbf{s}_r , instead of \mathbf{s} , the individual correlations $\rho_r = \text{corr}(\mathbf{z}_r, \mathbf{s}_r)$ are equal to 0 in the first case and equal to 1 in the second case, which are the extreme values for ρ_r .

For testing the hypothesis $H_0 : \rho_r = 0$, we generate a number of bootstrap samples \mathbf{X}_r^* for ROI r by reordering its signals. The other ROIs are unchanged, and \mathbf{Z}_j and \mathbf{v}_j which were obtained from the CCA on the original data, are used to calculate \mathbf{s}_r . Next, we perform a nonnegativity and spatially regularized regression from \mathbf{X}_r^* to \mathbf{s}_r due to which we obtain \mathbf{w}_r^* and the corresponding \mathbf{z}_r^* . Finally, we have $\rho_r^* = \text{corr}(\mathbf{z}_r^*, \mathbf{s}_r)$ and the p -value for rejecting H_0 is given by the fraction of times that $\rho_r^* \geq \rho_r$. This procedure is repeated for the different ROIs, and for the different significant modes.

2.6.3. Voxel weights: \mathbf{w}

The strictly positive voxel weights determine the correlated parts in the ROIs. The confidence intervals for each $\mathbf{w}_{x,i}$ follow directly from the bootstrapping procedure as described above, where we use the same reordering in all the ROIs, and whereby the weight vectors are always rescaled to get normalized signals \mathbf{Z} . For the quantification of the variability of \mathbf{w}_x , we determine $\text{MSE}_{\mathbf{w}_x}$ for ROI x as follows. For each voxel i , we calculate the standard deviation, $\sigma_{x,i}$, of $\mathbf{w}_{x,i}$ in the different bootstrap samples, and we define $\text{MSE}_{\mathbf{w}_x} = \sum_i \sigma_{x,i}^2/n_x$.

The variability of the weight vectors is related to the degree of spatial regularization, which is controlled by the parameter γ . Due to the multicollinearity problem, caused by the spatial dependency structure between neighbouring voxels, the weight vectors are very unsmooth and scattered, and, furthermore, very different weight patterns will be obtained for different bootstrap samples. A higher value for γ increases the role of the penalty term, or, equivalently, imposes stronger constraints to the model. This leads to smoother weight patterns and makes them also less susceptible to small variations in the data. With increasing γ -values more and more the same voxels in different bootstrap samples will be involved with a strictly positive weight, and $\text{MSE}_{\mathbf{w}}$ will continue to decrease. However, stronger constraints also cause a decrease in the overall correlation, ρ_{tot} , as the model has less degrees of freedom. Thus, there is a tradeoff between variability of the weights and the overall correlation. In order to find an optimal γ -value, we define the minimum weight vector, \mathbf{w}_x^{\min} , for a given γ , as the vector in which each element i contains the minimum $w_{x,i}$ found in the different bootstrap samples. These \mathbf{w}_x^{\min} vectors are determined for all the ROIs, and applied to the original data to calculate the overall correlation ρ_{tot}^{\min} .

With a too small γ -value, quite different \mathbf{w}_x vectors are obtained in each bootstrap sample. Taking the minimum over these quite different weight vectors leads to a quasi null-vector, and also to a low overall correlation. With a higher, optimal, γ -value rather the same \mathbf{w}_x vectors are obtained in each bootstrap sample, and \mathbf{w}_x^{\min} will be quasi equal to \mathbf{w}_x . With a still higher γ -value, the weight vectors, \mathbf{w}_x as well as \mathbf{w}_x^{\min} are enforced to be too smooth, causing also bad signals to gain too much influence, and the overall correlation will decrease. Thus, we determine \mathbf{w}_x^{\min} for different γ -values and select this γ that leads to the maximum ρ_{tot}^{\min} . We apply this procedure only for the first canonical correlation, and use the same γ for the successive canonical correlations.

3. Results

We use the terms CCA for classical multiset CCA solved with the GEV approach, CCA-pos for CCA with nonnegativity constraints, CCA-reg for CCA with spatial regularization, and CCA--posreg for CCA with nonnegativity constraints and spatial regularization. Confidence intervals are obtained by generating 1000 random samples and applying one of the four CCA approaches on each sample. We show always the 5–95% percentiles of the estimates to indicate how reliably the parameters are estimated. For hypothesis testing, we generate 1000 null-data sets, as described above, and show the 0.05 threshold level.

3.1. Synthetic data

Simulations with synthetic data are carried out in order to evaluate the effects of the nonnegativity constraints and the spatial regularization. We simulate data from 6 ROIs that are correlated with each other by one or more common sources. We use 3 source signals, $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$, simulating data from 40 runs, each consisting of 30 time steps. The source signals for one run are shown in Fig. 1. The signal $\mathbf{x}_{r,i}$ from voxel i in ROI r is generated using the following data generating process:

$$\mathbf{x}_{r,i} = \alpha_{1,r,i}\mathbf{q}_1 + \alpha_{2,r,i}\mathbf{q}_2 + \alpha_{3,r,i}\mathbf{q}_3 + \alpha_{0,r,i}\mathbf{u}_{r,i} \quad (20)$$

where the source signals \mathbf{q}_j ($j = 1 \dots 3$) were normalized to unit variance, and where the autocorrelated noise signals $\mathbf{u}_{r,i}$ were generated from an AR(1) process and normalised to unit variance. In all simulations, we set $\alpha_{0,r,i} = 2$.

We use (20) in 3 types of experiments. In the first experiment (EXP1), each ROI consists of only a small number of voxels (10 voxels), and there is no spatial dependency structure in the noise. The

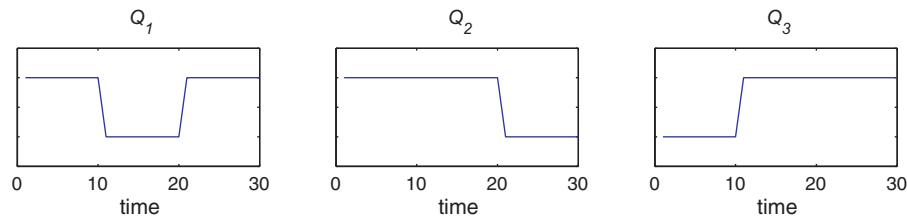


Fig. 1. The 3 underlying common sources for the synthetic data generation.

coefficients α_j in (20) are shown in the top row in Fig. 3 for the 10 voxels in the 6 ROIs. Since spatial regularization is not needed in this experiment, we can focus on the effect of the nonnegativity constraints.

In a second experiment (EXP2), we use only one source signal, \mathbf{q}_1 , and 2 ROIs, but the ROIs consist now of many (i.e., 300) voxels. Furthermore, the autocorrelated noise was convolved with a 4×4 mm FWHM Gaussian kernel in order to introduce a spatial dependency structure in the noise. The voxels that were correlated by \mathbf{q}_1 are depicted by the yellow blobs shown in the top panels of Fig. 4. This experiment is designed to directly compare our method with the classical unconstrained two-set CCA.

In the third experiment (EXP3), the 3 common sources \mathbf{q}_j are again used for generating data in 6 ROIs, as in EXP1, but now with many (i.e., 300) voxels in each ROI and using the same Gaussian kernel, as in EXP2, to introduce spatial dependency structures in these 6 ROIs. The ground truth is shown in the top panels of Fig. 5.

The results from EXP1 are shown in Figs. 2 and 3. The left panels in Fig. 2 show the overall correlations in the 7 first modes with CCA-pos (top) and unconstrained CCA (bottom). CCA-pos indicates 3 statistically significant modes while only two significant modes are found in the unconstrained case. The right panels show the importance of each ROI in the different modes, $\rho_r^{[e]}$, which is related to the eigenvector $\mathbf{v}^{[e]}$. The importance of the individual voxels is shown in Fig. 3, by means of their weight vectors \mathbf{w} for CCA-pos (middle) and CCA (bottom). When comparing these weight vectors with the underlying ground truth it becomes clear that mode 1 in CCA-pos captures voxels that are correlated by \mathbf{q}_2 , mode 2 corresponds to \mathbf{q}_1 , and mode 3 to \mathbf{q}_3 . This clear distinction is not possible with unconstrained CCA as the 3 underlying sources are merged into only 2 significant modes.

The results from EXP2 are shown in Fig. 4. The middle row of plots shows, as a function of the modes, the total correlations with the 4 CCA approaches. We see that the classical CCA finds 16 significant modes while the voxels in the 2 ROIs are correlated by only 1 common source. The voxels to which CCA assigns a non-zero

weight are mostly taken in the correlated parts, as can be seen in the weight vectors for the different modes (bottom panels), but these voxels are combined with positive as well as with negative weights. This freedom is taken away with CCA-pos, and, as a result, only one significant mode is found. Also spatial regularization, even without nonnegativity constraints, solves the problem of too many significant modes in this case. The reason is that highly positive and negative weights in nearby voxels are penalized by the spatial regularization term, and hence, will not occur.

In EXP3 we compare spatially regularized CCA with and without nonnegativity constraints, thus, CCA-posreg and CCA-reg, respectively. The total correlations, ρ_{tot} (not shown), are comparable to the ones in EXP1, i.e. 3 significant modes are found with CCA-posreg and only 2 with CCA-reg. For the first 7 modes in CCA-posreg, we have $\rho_{tot} = [0.4049; 0.4043; 0.4008; 0.0813; 0.0793; 0.0767; 0.0615]$, while for CCA-reg they are $\rho_{tot} = [0.6947; 0.4051; 0.1670; 0.1666; 0.1598; 0.1580; 0.1542]$. The weight vectors are shown in Fig. 5 from which we see that CCA-posreg delineates the spatially contiguous blobs that are correlated with each other by the different sources \mathbf{q}_j , while CCA-reg without constraints suffers from the fact that anticorrelated parts could be merged and could compensate for each other.

3.2. Real fMRI data

We used contrast agent-enhanced fMRI data from 3 experiments. In the experiments, awake monkeys were scanned in a horizontal bore, 1.5T MR scanner (Sonata; Siemens, Erlangen, Germany) equipped with echoplanar imaging. Each time series consisted of gradient-echo echoplanar whole brain images: repetition time (TR), 2.4 s; echo time (TE), 27 ms; 64×64 matrix; $2 \text{ mm} \times 2 \text{ mm} \times 2 \text{ mm}$ voxels (32 sagittal slices). The monkey functional volumes were realigned and non-rigidly co-registered with the anatomy of a template. The brain of the template has been scanned at a spatial resolution of 1 mm^3 . The monkey functional volumes were subsampled to 1 mm^3 , but not smoothed.

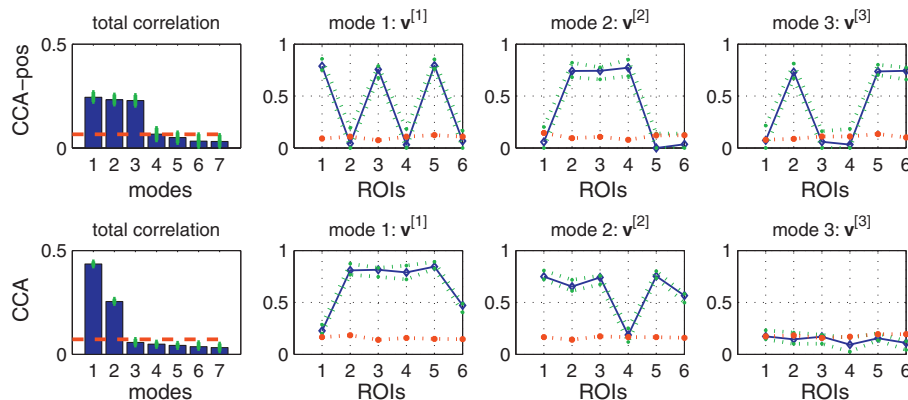


Fig. 2. Synthetic data-EXP1: total correlations, ρ_{tot} (left), and the importance of each ROI in the first 3 modes, $\rho_r^{[e]}$ (right), when applying CCA-pos (top) and CCA (bottom).

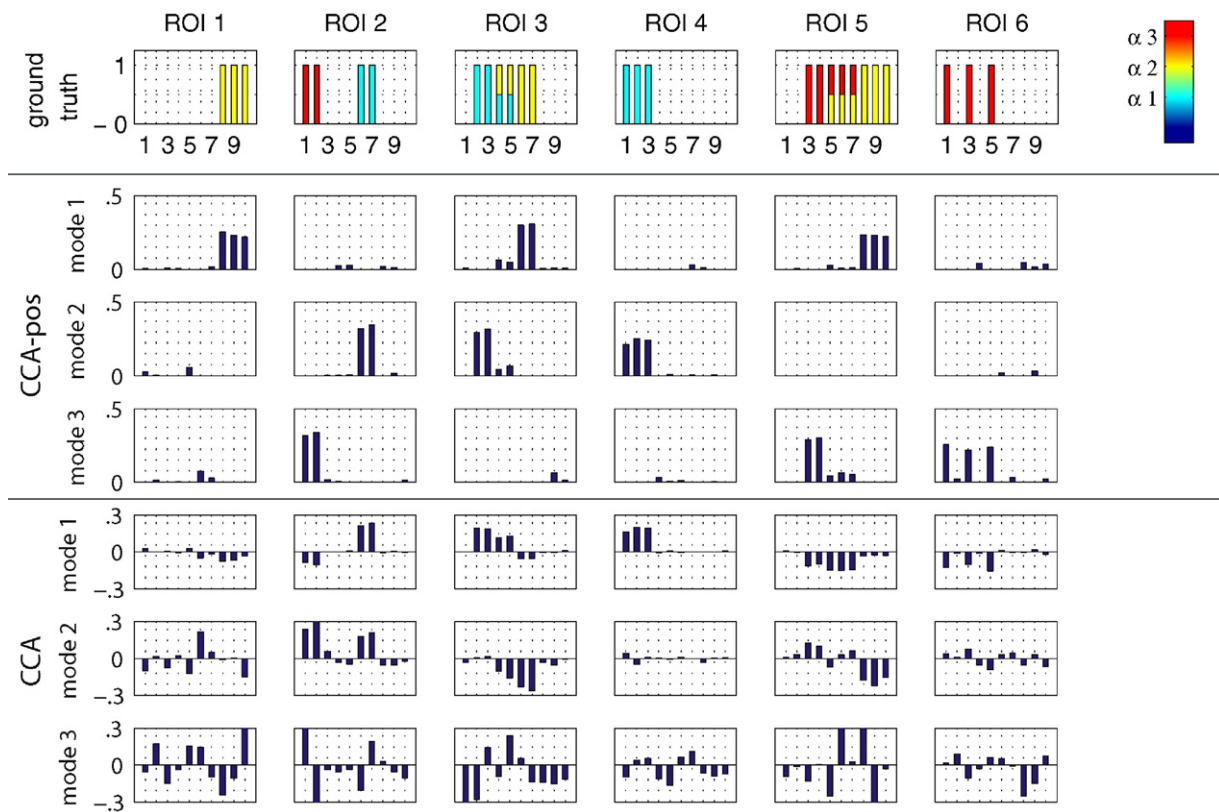


Fig. 3. Synthetic data-EXP1: ground truth (top) and the weight vectors obtained with CCA--pos (middle) and CCA (bottom).

The first experiment comes from an action observation study by Nelissen et al. (2005): monkeys were scanned when viewing video-clips in which a person was grasping objects. Two types of control stimuli were used: static single frames of the action videos, and scrambled videos produced by phase scrambling each frame of the video sequences. A block design was used with 4 conditions per run (the action video, the scrambled sequence, the static frame, and a fixation condition). Within the same run, the presentation order of the images was repeated twice, thus, 8 blocks per run. Between the runs, the order of the conditions was randomized. In each block, 15 scans were obtained. We selected data from 14 runs (only runs in which the monkey fixated for more than 85% of the time were retained). In the further analysis, we removed the volumes obtained during the fixation condition, thus, our time signals are from 1260 ($-14 \times 3 \times 2 \times 15$) volumes.

The second data set comes from the same action observation study, but with video-clips only showing the hand (and forearm) grasping the objects, instead of the full actor. The same type of control stimuli and block design are used.

The third data set comes from a motion localizer study by Nelissen et al. (2006). In this experiment, random textured patterns (50% white 0.075° dots, 50% black 0.075° dots) were presented in a circular aperture (14° diameter) either stationary or translating at five different speeds (1, 2, 4, 8 and 16°/s). A block design was used, with 7 conditions per run (5 speeds, 1 stationary and 1 fixation condition). Within the same run, the presentation order of conditions was repeated two times. Between the runs, the order of the conditions was randomized. Again, the fixation condition was removed and 10 runs were selected (where the monkey fixated for more than 85% of the time). This leads to time signals with 1200 ($=10 \times 6 \times 2 \times 10$) volumes.

3.2.1. Analysis with 2 ROIs

We delineated two broad ROIs (400 voxels each), one in the frontal lobe and one in the intraparietal sulcus (IPS), and applied the different CCA-approaches on these two ROIs with the full-actor action data set. The correlations for successive modes with the different CCA approaches (not shown) are comparable with those that we obtained on the synthetic data in EXP2, i.e., a large number of significant modes with classical unconstrained and unregularized CCA. We focus further on the effect of the regularization parameter, γ , in the first (and only significant) mode with CCA--posreg. In the left plot of Fig. 7, ρ_{tot} is shown for different γ -values and it decreases when γ increases. The optimal degree of spatial regularization is selected as $\gamma = 0.9$ since ρ_{tot} obtained on the minimum weight vectors (see Section 2.6.3) is then maximal. The variability of the weight vectors over different bootstrap samples is shown in the right plot of Fig. 7. For the sake of comparison, we apply PCA within the ROIs separately, using the same bootstrap samples. The voxel weights from the PCA are also rescaled so that the corresponding representative signals have unit variance, in order to be comparable with the CCA results. The green lines in Fig. 7 show the results when using these PCA-weight vectors, where we see that they have less spatial variability over different bootstrap samples, but also that the correlation between the two representative signals is lower than with CCA--posreg. The weight vectors projected on the brain slices are shown in Fig. 6 for CCA--pos (this is $\gamma = 0$), for CCA--posreg with $\gamma = 0.9$, and for PCA. We see that CCA--pos selects non-contiguous blobs and PCA gives very (and maybe too) smooth weight patterns. On the other hand, CCA--posreg seems to delineate only the important subpart. This is located very anterior in the postarcuate bank, which is in agreement with known anatomical connectivity between the two selected ROIs (Borra et al., 2008; Belmalih et al., 2009).

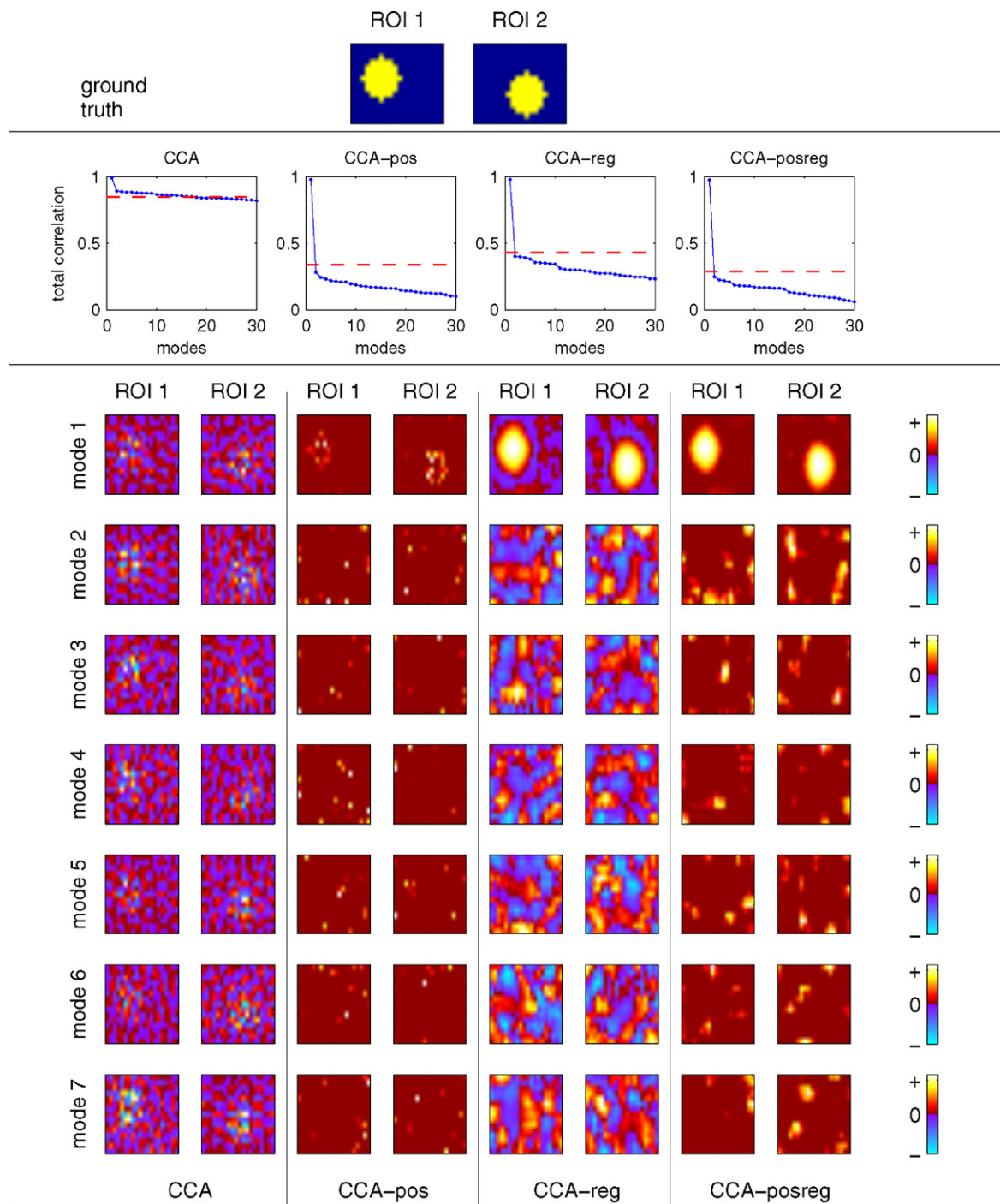


Fig. 4. Synthetic data-EXP2: ground truth (top), total correlations in the 4 different CCA-approaches (middle) and their corresponding weight vectors (bottom).

3.2.2. Analysis with 2×48 ROIs

We select 48 ROIs from an in-house ROI atlas in both hemispheres. The atlas defines the early visual areas in the occipital lobe, regions in the superior temporal sulcus (STS), regions in the parietal cortex, and premotor areas. The selected ROIs are:

- The early visual areas:** areas V1d, V1v, V2d, V2v, V3A, V3d, V3v, V4d, V4v, which are defined in Fize et al. (2003), based on their retinotopic organization; and V6 and V6A defined in Galletti et al. (1999).
- Regions in the intraparietal sulcus (IPS):** areas PIP, MIP, VIPp, pIPS (posterior, medial, ventral, and posterior at the fundus);

next, subdivisions of the lateral intraparietal area (LIP): LIPdi, LIPvi, LIPda, LIPva (dorsal, ventral and intermediate, anterior); finally, the anterior part of the IPS: AIPp and AIPa (posterior and anterior). These regions are described in Durand et al. (2007).

- Regions in the intraparietal lobule (IPL):** here we have 4 distinct areas forming the cortical convexity of the IPL, in agreement with Pandya and Seltzer (1982) and Rozzi et al. (2006), referred to as PF, PFG, PG and Opt.
- Regions in the superior temporal sulcus (STS):** area MT/V5 and its satellites, namely dorsal and ventral MST (MSTd, MSTv) and FST at the fundus of the STS; next, LST, first described in Nelissen et al. (2006) and located anterior to FST. Further, STPm (middle

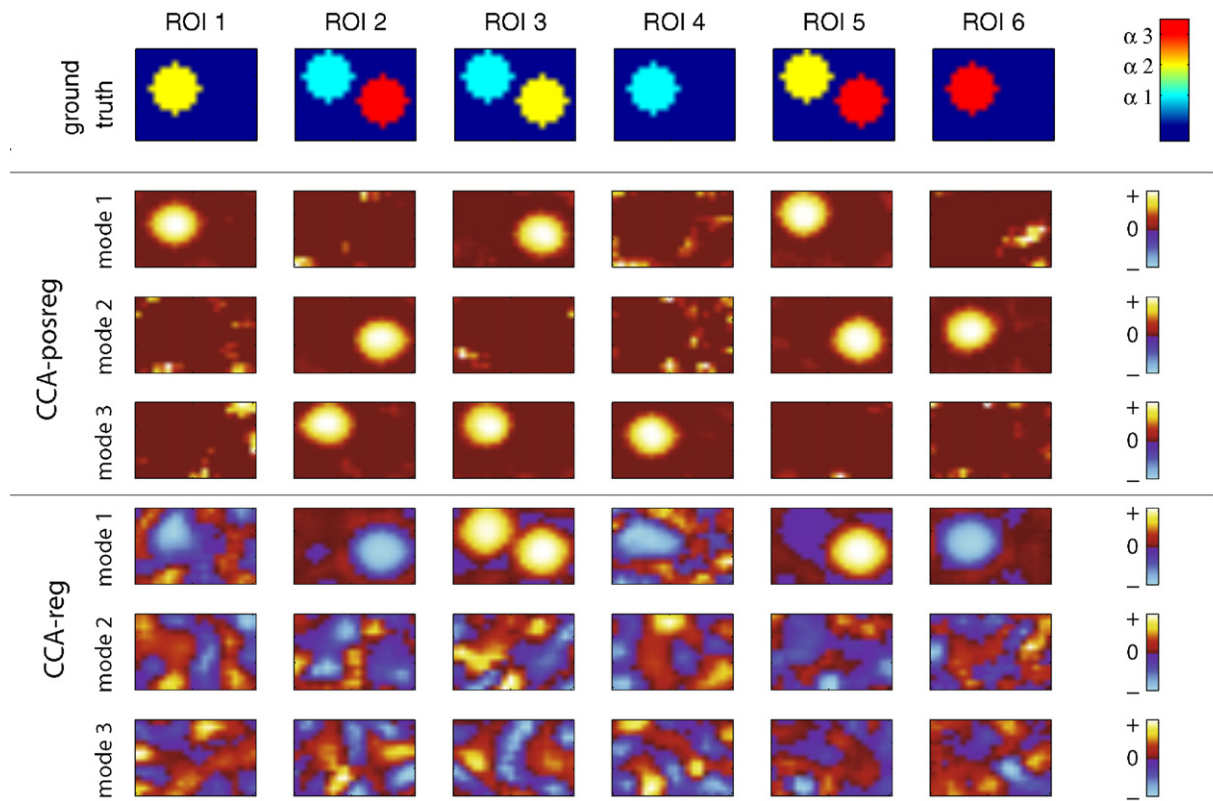


Fig. 5. Synthetic data-EXP3: ground truth (top) and weight vectors obtained with CCA--posreg (middle) and CCA--reg (bottom).

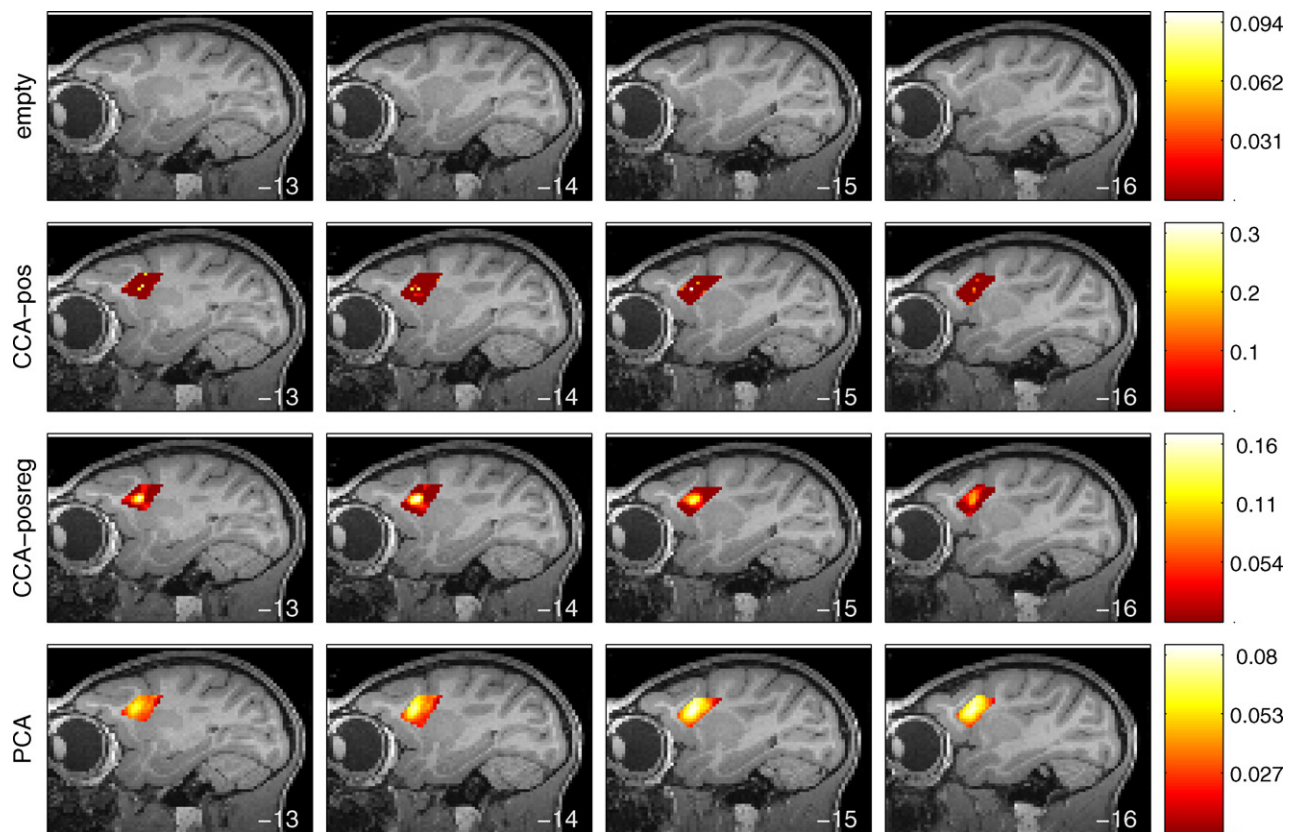


Fig. 6. Weight vectors for CCA with one ROI in the frontal lobe and one ROI in the intraparietal sulcus (only the frontal ROI is shown). The results are obtained with CCA--pos (top), CCA--posreg (middle) and PCA (bottom). CCA--pos gives scattered and unsmooth results whereas PCA gives diffused weightings. CCA--posreg delineates only the really important part of the ROI.

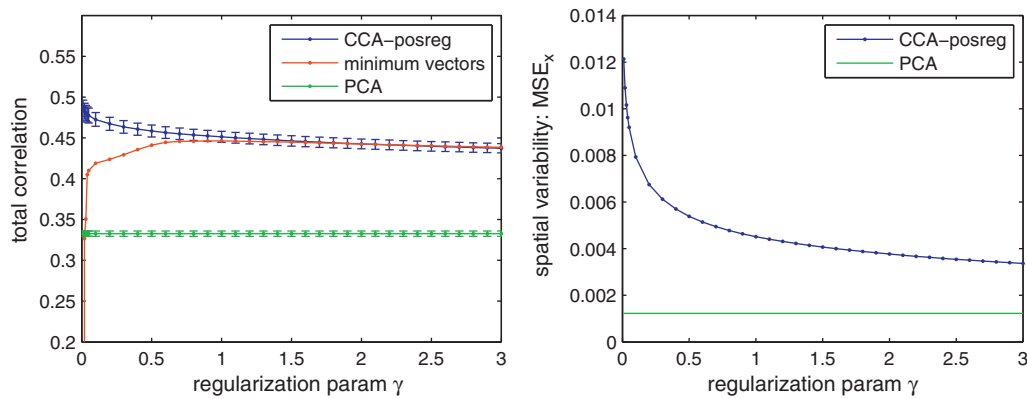


Fig. 7. Effect of the degree of spatial regularization (γ) on the total correlation (left) and on the spatial variability of the weights in the frontal ROI (right). The minimum weight vector, \mathbf{w}_x^{\min} , for a given γ -value is computed as the minimum vector over the weight vectors that are obtained with CCA-posreg for the different bootstrap samples. The parameter γ is not involved when applying PCA within the ROIs.

part of the superior temporal polysensory region) in the upper bank of the STS. Finally, LB1, LB2, and UB1, UB2 which are two anterior regions in the lower and upper bank, respectively, of the rostral portion of the STS. These regions are described in Nelissen et al. (2006).

5. Regions in the frontal cortex: areas F1, F2, F3 (=SMA-proper) and F4; next, three subparts of the premotor field F5: F5c, F5p, F5a which are at the convexity, and at the posterior and anterior sector of the posterior bank, respectively. Further areas 45B and FEF in the prearcuate cortex and the arcuate sulcus, respectively, and areas F6 (=pre-SMA) and F7. The parcellation of F5 and neighbouring cortex was based on anatomical criteria described in Nelissen et al. (2005) and Belmalih et al. (2007, 2009); the definition of the other premotor ROIs was based on earlier studies reviewed in Rizzolatti and Luppino (2001), and the procedure to define the coordinates of these regions on the template brain has been described in Joly et al. (2009).

We apply CCA-posreg on the 3 data sets and we find 2 significant modes for the action observation data where the monkey sees the full actor (see left panel in Fig. 8), 2 significant modes for hand action observation, and one significant mode for the moving dots data (the two latter results are not shown). The signals $\mathbf{s}^{[\ell]} = \mathbf{Z}^{[\ell]} \mathbf{v}^{[\ell]}$, for $\ell = 1, 2$ for the full-person action observation are plotted in the two panels on the right in Fig. 8, after averaging these signals over

the different runs. From these signals, we can conclude that the correlation in the first mode is mainly created by ROIs which are activated during the action observation and to a lesser degree during the observation of the static frames, while the correlation in the second mode is mainly created by ROIs that are activated during the observation of meaningless motion (the scrambled videos). The functional connectivity patterns, indicating the importance of each of the 96 ROIs in these two modes, are shown in Fig. 9. The functional connectivity patterns for the hand-action are plotted in Fig. 10.

The first-mode connectivity patterns for action observation are mostly determined by regions in the frontal lobe, regions in the lower bank of the STS, and regions in the IPS. The connectivity patterns for the left and the right hemisphere are different for the early visual regions, due to the asymmetric placement of the stimulus, but seem to become similar starting from the IPL (regions PF/PFG), for both the full-actor and the hand-data. When we compare the first-mode patterns for full-actor versus hand, we see that the IPL regions, mostly PFG, are present for the full-actor connectivity pattern, but not for the hand. In the STS, the regions of the lower bank are more present in the connectivity patterns than those in the upper bank, where mostly the middle portion (STPm) is involved. In the IPS, mainly the anterior parts of LIP and AIP are involved. In the frontal lobe, the most important regions are F5a, 45B, and FEF. Region F5c is involved in the functional connectivity pattern for the

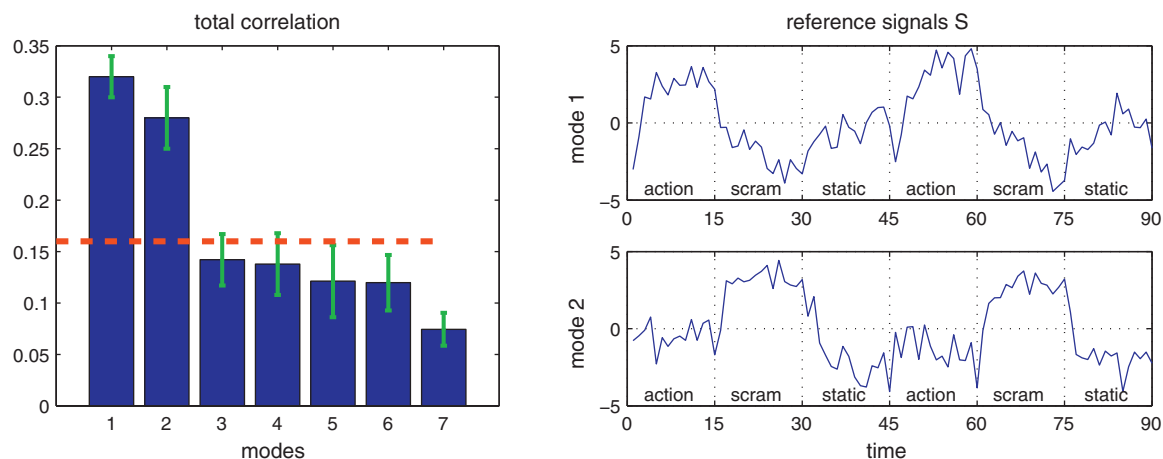


Fig. 8. Results for CCA-posreg on the full-actor action observation data set. Left: total correlation in the successive CCA modes. The red dotted line represents the 0.05-threshold for $H_0 : \rho_{tot} = 0$ (1000 null-samples). Right: reference signals for the two significant modes after averaging over the runs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

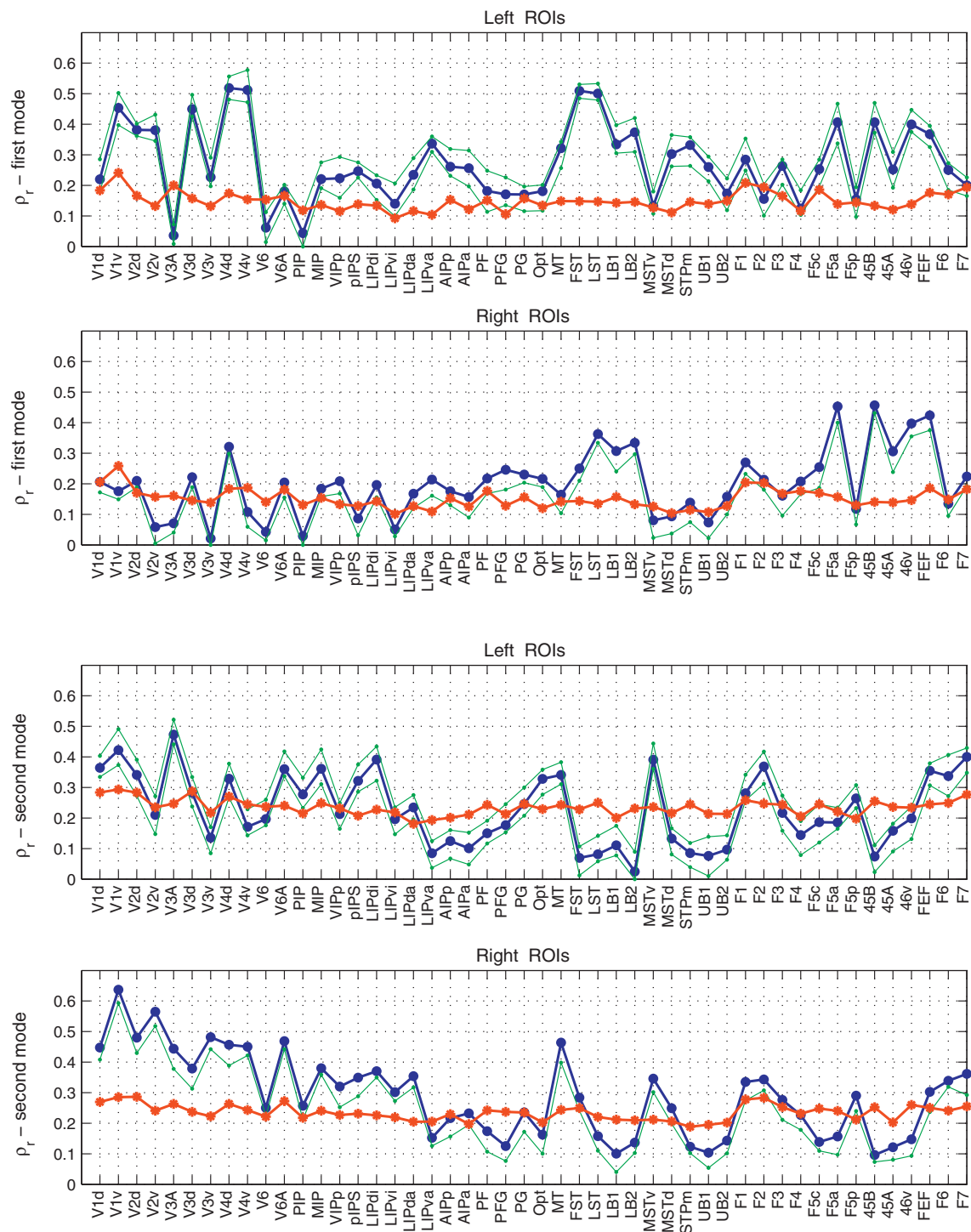


Fig. 9. Action–observation, full person: functional connectivity patterns corresponding to the first and second mode in CCA–posreg with the 96 ROIs. The blue line denotes the importance of each ROI, measured as ρ_r , the green lines represent the confidence intervals and the red line denotes the 0.05-threshold for $H_0: \rho_r = 0$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

full-actor but not for the hand. This is in agreement with Nelissen et al. (2005) who states that F5c is context-dependent (thus, the full actor is needed), whereas F5a is more abstract and less context-related (the observed hand-action is enough). This different these involvements of F5c in the functional connectivity patterns for full person or hand action observations is similar to what we observed in PFG. This fits well with known anatomical connections between PFG and F5c (Nelissen, 2006; Luppino et al., 1999; Matelli et al., 1986).

The second-mode functional connectivity patterns for action observation seem mainly involved in (meaningless) motion, as could be seen from the averaged time signals in Fig. 8, and the regions that are involved in this mode are MT, MST, FEF, F5p. These regions are also part of the functional connectivity pattern for the moving dots, shown in Fig. 11. In this figure, we see clear similarities between both hemispheres. The most important regions here are the early visual regions, regions in the STS, but not so deep as for the action obser-

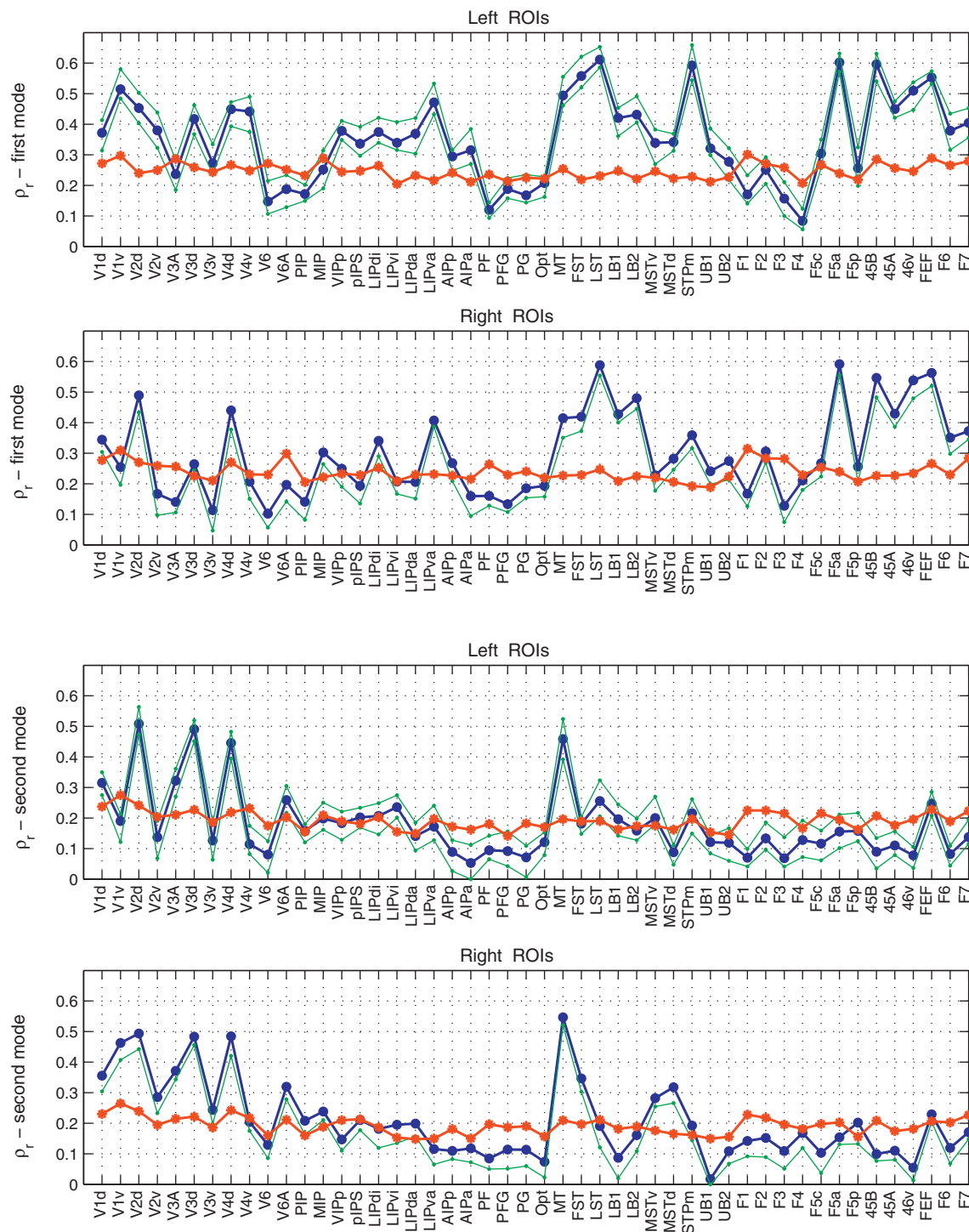


Fig. 10. Action–observation, hand: functional connectivity patterns corresponding to the first and second mode in CCA--posreg with the 96 ROIs. The blue line denotes the importance of each ROI, measured as ρ_r , the green lines represent the confidence intervals and the red line denotes the 0.05-threshold for $H_0: \rho_r = 0$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

vation (no LB1, UB1, etc.), the anterior part of LIP, F5 and FEF.

4. Discussion and conclusions

We introduced a method for detecting correlated subparts within given ROIs, based on multiset nonnegativity constrained and spatially regularized CCA. The usefulness of the CCA--posreg

approach was shown on synthetic data, where opposite effects were merged when using classical unconstrained CCA. Thus, we showed that the use of constrained optimization enhances separability greatly, as compared with standard CCA approaches. The results on real fMRI data are in agreement with known neurophysiological data.

In the original introduction of nonnegativity constraints for two-set CCA (Das and Sen, 1994), it was shown that the constrained solution equals the unconstrained solution to a modified CCA prob-

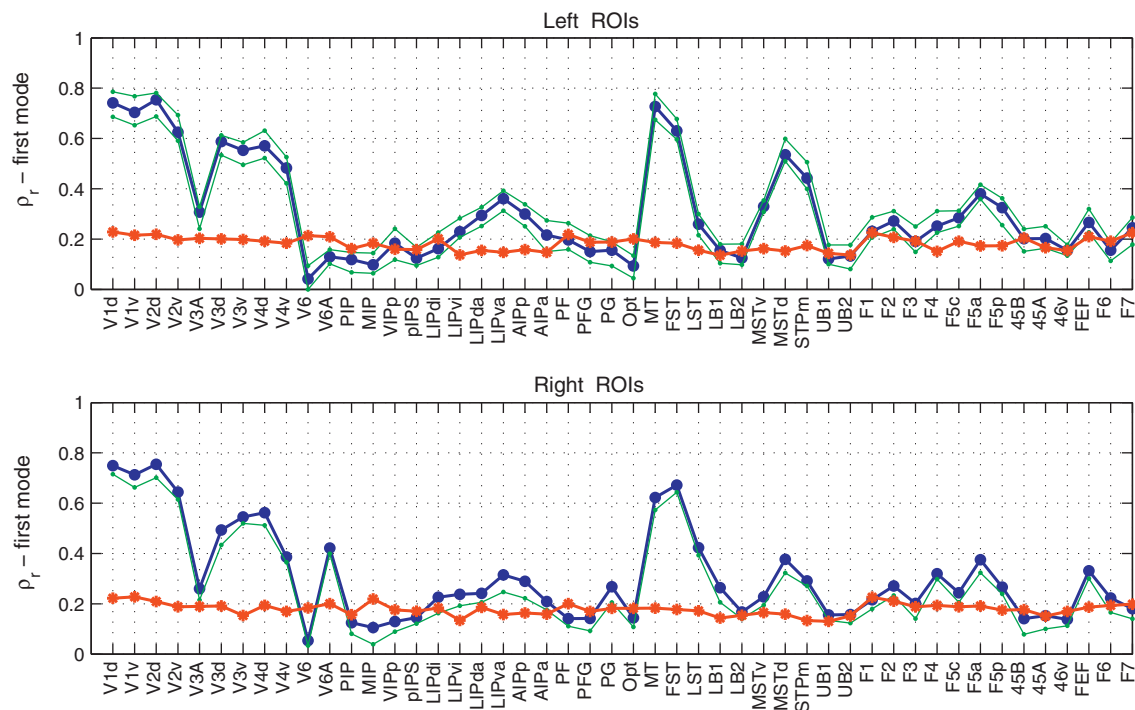


Fig. 11. Moving dots: functional connectivity pattern corresponding to the first mode in CCA--posreg with the 96 ROIs. The blue line denotes the importance of each ROI, measured as ρ_r , the green lines represent the confidence intervals and the red line denotes the 0.05-threshold for $H_0: \rho_r = 0$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

lem, i.e. where one or several variables have been excluded, or unconstrained CCA to 'proper' submatrices based on the Kuhn-Tucker conditions. However, for two sets with p and q variables respectively, there are $(2^p - 1)(2^q - 1)$ proper submatrices, and no procedure was given to find the optimal submatrices. All possible deletions must be tested to find the solution, which gives the constrained CCA the unpleasant property of growing exponentially in the number of dimensions (here voxels). By contrast, by applying an iterative regression approach for the CCA problem, we can revert to standard *npls*-algorithms to find the nonnegative weight vector at each iteration. As the applied *npls*-algorithm of Bro and De Jong (1997) is based on active set theory, the passive constraints define the 'proper' submatrices in the terminology of Das and Sen (1994). Furthermore, since we impose nonnegativity constraints in multiset-CCA, these constraints are also to be considered when combining the different pairwise correlations between the multiple sets. This is done by using the nonnegative eigenvector in the *maxvar* approach, which is equivalent to the GEV formulation for multiset-CCA. Hence, our method could be considered as nonnegative PCA on the ROI level, but with ROIs that are optimized (by means of their weight vectors) in the sense that they admit the best possible PCA representation.

In order to enforce smoother weight patterns, we added a spatial regularization term in a ridge-like fashion. However, we penalize the differences between adjacent voxel weights, instead of their absolute values, as is done in ridge regression using the identity matrix. Ridge-like regularization has also been applied in (multiset) CCA by adding the identity matrix to the within-set correlation matrices in the (generalized) eigenvalue solutions (Hardoon et al., 2004; Takane et al., 2008). Note, however, that the regularization also influences the orthogonality constraints as they then become $\mathbf{H}(\mathbf{B} + \mathbf{G})\mathbf{H} = \mathbf{I}$, with \mathbf{G} the regularization matrix (possibly $\mathbf{G} = \mathbf{I}$). Spatial regularization was necessary because of the multicollinearity problem in fMRI. PCA doesn't suffer from multicollinearity (since it does not use the inverse of the correlation matrices). Therefore, we compared CCA--posreg with PCA

within the ROIs, and we found very smooth weight patterns with PCA. However, these weight vectors lead to lower correlations between the representative signals. Since CCA--posreg delineates the really important parts, and not much more, our CCA-method may be better suited to refine the given ROIs, or possibly, as a starting point in a method to indicate that some ROIs should be splitted and/or joined with other neighbouring ROIs. Furthermore, applying PCA within the ROIs for defining their representative signals may become complicated when different underlying modes are present in the data, since it is not guaranteed then that the principal components in the different ROIs for one mode correspond to each other. With CCA, the signals from different ROIs are always maximally correlated to each other within each mode. Since the method detects spatially contiguous blobs of voxels, based on their correlation, it can be considered as a method for functional connectivity modelling. The constructed representative signals can also directly be used in further analyses of effective connectivity.

Our iterative algorithm is an 'Alternating Least Squares' method, but with nonnegativity and regularization constraints at each iteration. It is important to mention that, theoretically, spatially regularized multiset CCA could be directly solved with the GEV approach (also possibly with nonnegativity constraints). However, it is required then that the total number of voxels (in all the ROIs together) is lower than the length of the time signals, which is not reasonable for real fMRI data. On the other hand, in our iterative method, only the number of voxels, in each of the ROIs separately, must be lower than the length of the time signals, which is more likely within the limits for practical use.

Acknowledgments

We would like to thank Prof. Dr. Wim Vanduffel and Dr. Koen Nelissen for providing us the fMRI data.

FD is supported by research grants received from the Flemish Regional Ministry of Education (Belgium) (GOA 10/019). MMVH

is supported by research grants received from the Excellence Financing program (EF 2005) and the CREA Financing program (CREA/07/027) of the K.U.Leuven, the Belgian Fund for Scientific Research Flanders (G.0588.09), the Interuniversity Attraction Poles Programme – Belgian Science Policy (IUAP P6/054), the Flemish Regional Ministry of Education (Belgium) (GOA 10/019), and the European Commission (IST-2007-217077), and by the SWIFT prize of the King Baudouin Foundation of Belgium.

Appendix A.

The Lagrange-formulation for (3)–(5) is:

$$L = \mathbf{w}_x' \mathbf{X}' \mathbf{Y} \mathbf{w}_y - \frac{\rho_x}{2} (\mathbf{w}_x' \mathbf{X}' \mathbf{X} \mathbf{w}_x - 1) - \frac{\rho_y}{2} (\mathbf{w}_y' \mathbf{Y}' \mathbf{Y} \mathbf{w}_y - 1) \quad (21)$$

and is solved by:

$$\frac{\partial L}{\partial \mathbf{w}_x} = \mathbf{X}' \mathbf{Y} \mathbf{w}_y - \rho_x \mathbf{X}' \mathbf{X} \mathbf{w}_x = 0 \quad (22)$$

$$\frac{\partial L}{\partial \mathbf{w}_y} = \mathbf{Y}' \mathbf{X} \mathbf{w}_x - \rho_y \mathbf{Y}' \mathbf{Y} \mathbf{w}_y = 0 \quad (23)$$

Subtracting \mathbf{w}_x' times (22) from \mathbf{w}_y' times (23), we have

$$\mathbf{w}_y' \mathbf{Y}' \mathbf{X} \mathbf{w}_x - \rho_y \mathbf{w}_y' \mathbf{Y}' \mathbf{Y} \mathbf{w}_y - \mathbf{w}_x' \mathbf{X}' \mathbf{Y} \mathbf{w}_y + \rho_x \mathbf{w}_x' \mathbf{X}' \mathbf{X} \mathbf{w}_x = 0 \quad (24)$$

and, with $\mathbf{x}'\mathbf{x} = 1$ and $\mathbf{y}'\mathbf{y} = 1$, we must conclude that $\rho_x = \rho_y = \rho$. We then get from Eq. (22):

$$\mathbf{w}_x = \frac{(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})\mathbf{w}_y}{\rho} \quad (25)$$

Substitution of (25) in (23) gives:

$$((\mathbf{Y}'\mathbf{Y})^{-1}(\mathbf{Y}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}))\mathbf{w}_y = \rho^2 \mathbf{w}_y \quad (26)$$

which is the eigenvector solution for \mathbf{w}_y . The solution for \mathbf{w}_x can be derived in a similar way.

References

- Anderson T. An introduction to multivariate statistical analysis. New York: Wiley & Sons; 1984.
- Baumgartner R, Scarth G, Teichtmeister C, Somorjai R, Moser E. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex: Part I. Reproducibility. *J Magn Reson Imaging* 1997;7:1094–101.
- Bellec P, Perlberg V, Jbabdi S, Pelegrini-Issac M, Anton JL, Doyon J, et al. Identification of large-scale networks in the brain using fMRI. *Neuroimage* 2006;29(4):1231–43.
- Belmalih A, Borra E, Contini M, Gerbella M, Rozzi S, Luppino G. A multiarchitectonic approach for the definition of functionally distinct areas and domains in the monkey frontal lobe. *J Anat* 2007;211:199–211.
- Belmalih A, Borra E, Contini M, Gerbella M, Rozzi S, Luppino G. Multimodal architectonic subdivision of the rostral part (area F5) of the macaque ventral premotor cortex. *J Comp Neurol* 2009;512:183–217.
- Biswal B, Yetkin F, Haughton V, Hyde J. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn Reson Med* 1995;34:537–41.
- Borga M. Learning multidimensional signal processing. Ph.D. thesis. Sweden: Linköping University; 1998.
- Borra E, Belmalih A, Calzavara R, Gerbella M, Murata A, Rozzi S, et al. Cortical connections of the macaque anterior intraparietal (AIP) area. *Cereb Cortex* 2008;18(5):1094–111.
- Bro R, De Jong S. A fast non-negativity-constrained least squares algorithm. *J Chemometr* 1997;11(5):393–401.
- Bruguier A, Preuschoff K, Quartz S, Bossaerts P. Investigating signal integration with canonical correlation analysis of fMRI brain activation data. *Neuroimage* 2008;41:35–44.
- Büchel C, Friston K. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cereb Cortex* 1997;7:768–78.
- Bullmore E, Horwitz B, Honey G, Brammer M, Williams S, Sharma T. How good is good enough in path analysis of fmri data? *Neuroimage* 2000;11:289–301.
- Cichocki A, Zdunek R. Regularized alternating least squares algorithms for non-negative matrix/tensor factorization. *Lecture Notes Comput Sci* 2007;4493:793–802.
- Correa N, Eichele T, Adali T, Li Y-O, Calhoun V. Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI. *Neuroimage* 2010;50(4):1438–45.
- Das S, Sen P. Restricted canonical correlations. *Linear Alg Appl* 1994;210:29–47.
- Das S, Sen P. Asymptotic distribution of restricted canonical correlations and relevant resampling methods. *J Multivar Anal* 1996;56(1):1–19.
- Deleus F, Van Hulle M. A connectivity-based method for defining regions-of-interest in fMRI data. *IEEE Trans Image Process* 2009;18(8):1760–71.
- Durand J, Nelissen K, Joly O, Wardak C, Todd J, Norman J, et al. Anterior regions of monkey parietal cortex process visual 3D shape. *Neuron* 2007;55(3):493–505.
- Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman & Hall; 1993.
- Fize D, Vanduffel W, Nelissen K, Denys K, Chef d'Hotel C, Faugeras O, et al. The retinotopic organization of primate dorsal V4 and surrounding areas: a functional magnetic resonance imaging study in awake monkeys. *J Neurosci* 2003;23:7395–406.
- Forman S, Cohen J, Fitzgerald M, Eddy W, Mintun M, Noll D. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med* 1995;33:636–47.
- Friman O, Borga M, Lundberg P, Knutsson H. Detection of neural activity in fMRI using maximum correlation modeling. *Neuroimage* 2002;15(2):386–95.
- Friman O, Borga M, Lundberg P, Knutsson H. Adaptive analysis of fMRI data. *Neuroimage* 2003;19(3):837–45.
- Friman O, Cedefamn J, Lundberg P, Borga M, Knutsson H. Detection of neural activity in functional MRI using canonical correlation analysis. *Magn Reson Med* 2001;45:323–30.
- Friston K. Functional and effective connectivity in neuroimaging: a synthesis. *Hum Brain Mapp* 1994;2(1–2):56–78.
- Friston K. Characterizing distributed functional systems. In: Frackowiak R, Friston K, Frith C, Dolan R, Mazziota J, editors. Human brain function. Academic Press; 1997. p. 107–26.
- Friston K, Frith C, Frackowiak R, Turner R. Characterizing dynamic brain responses with fMRI: a multivariate approach. *Neuroimage* 1995;2:166–72.
- Friston K, Frith C, Liddle P, Frackowiak R. Functional connectivity: the principal component analysis of large (PET) data sets. *J Cereb Blood Flow Metab* 1993;13:153–71.
- Galletti C, Fattori P, Gamberini M, Kutz D. The cortical visual area v6: brain location visual topography. *Eur J Neurosci* 1999;11:3922–36.
- Gavrilescu M, Stuart G, Rossell S, Henshall K, McKay C, Sergejew A, et al. Functional connectivity estimation in fMRI data: influence of preprocessing and time course selection. *Hum Brain Mapp* 2008;29(9):1040–52.
- Van de Geer J. Linear relations among k sets of variables. *Psychometrika* 1984;49(1):79–94.
- Goncalves M, Hall D, Johnsrude I, Haggard M. Can meaningful effective connectivities be obtained between auditory cortical regions? *Neuroimage* 2001;14:1353–60.
- Grafton S, Sutton J, Couldwell W, Lew M, Waters C. Network analysis of motor system connectivity in parkinson's disease: modulation of thalamocortical interactions after pallidotomy. *Hum Brain Mapp* 1994;2:45–55.
- Greicius M, Krasnow B, Reiss A, Menon V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc Natl Acad Sci USA* 2003;100(1):253–8.
- Hardoon D, Mourao-Miranda J, Brammer M, Shawe-Taylor J. Unsupervised analysis of fMRI data using kernel canonical correlation. *Neuroimage* 2007;37(4):1250–9.
- Hardoon D, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: an overview with application to learning methods. *Neural Comput* 2004;16(12):2639–64.
- Harrison L, Penny W, Friston K. Multivariate autoregressive modeling of fMRI time series. *Neuroimage* 2003;19(4):1477–91.
- Horst P. Relations among m sets of measures. *Psychometrika* 1961;26:129–49.
- Hottelling H. Relations between two sets of variates. *Biometrika* 1936;28:321–77.
- Jennings J, McIntosh A, Kapur S. Mapping neural interactivity onto regional activity: an analysis of semantic processing and response mode interactions. *Neuroimage* 1998;7:244–54.
- Joly O, Vanduffel W, Orban G. The monkey ventral premotor cortex processes 3D shape from disparity. *Neuroimage* 2009;47:262–72.
- Katanoda K, Matsuda Y, Sugishita M. A spatial-temporal regression model for the analysis of functional MRI data. *Neuroimage* 2002;17:1415–28.
- Kettenring J. Canonical analysis of several sets of variables. *Biometrika* 1971;58(3):433–51.
- Kohler S, McIntosh A, Moscovitch M, Winocur G. Functional interactions between the medial temporal lobes and posterior neocortex related to episodic memory retrieval. *Cereb Cortex* 1998;8:451–61.
- Lawson C, Hanson R. Solving least squares problems. Englewood Cliffs, NJ: Prentice-Hall; 1974.
- Li YO, Adali T, Wang W, Calhoun V. Joint blind source separation by multiset canonical correlation analysis. *IEEE Trans Signal Process* 2009;57(10):3918–29.
- Liu W, Mandic D, Cichocki A. Analysis and online realization of the CCA approach for blind source separation. *IEEE Trans Neural Netw* 2007;18(5):1505–10.
- Logothetis N, Pauls J, Augath M, Trinath T, Oeltermann A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 2001;412(6843):150–7.
- Lohmann G, Bohn S. Using replicator dynamics for analyzing fMRI data of the human brain. *IEEE Trans Med Imaging* 2002;21(5):485–92.
- Lu Y, Jiang T, Zang Y. A split-merge-based region-growing method for fMRI activation detection. *Hum Brain Mapp* 2004;22(4):271–9.
- Luppino G, Murata A, Govoni P, Matelli M. Largely segregated parietofrontal connections linking rostral intraparietal cortex (areas AIP and VIP) and the

- ventral premotor cortex (areas F5 and F4). *Exp Brain Res* 1999;128(1–2): 181–7.
- Marrelec G, Horwitz B, Kim J, Pélégriani-Issac M, Benali H, Doyon J. Using partial correlation to enhance structural equation modeling of functional MRI data. *Magn Reson Imaging* 2007;25(8):1181–9.
- Marrelec G, Krainik A, Duffau H, Pélégriani-Issac M, Lehericy S, Doyon J, et al. Partial correlation for functional brain interactivity investigation in functional MRI. *Neuroimage* 2006;32(1):228–37.
- Matelli M, Camarda R, Glickstein M, Rizzolatti G. Afferent and efferent projections of the inferior area-6 in the macaque monkey. *J Comp Neurol* 1986;251(3): 281–98.
- McIntosh A, Gonzalez-Lima F. Structural equation modeling and its application to network analysis in functional brain imaging. *Hum Brain Mapp* 1994;2(1–2): 2–22.
- McKeown M, Makeig S, Brown G, Jung T, Kindermann S, Bell A, et al. Analysis of fMRI data by blind separation into independent spatial components. *Hum Brain Mapp* 1998;6:160–88.
- Mechelli A, Penny W, Price C, Gitelman D, Friston K. Effective connectivity and intersubject variability: using a multisubject network to test differences and commonalities. *Neuroimage* 2002;17:1459–69.
- Nandy R, Cordes D. Improving the spatial specificity of canonical correlation analysis in fMRI. *Magn Reson Med* 2004;52(4):947–52.
- Nelissen K. Functional anatomy of visual processing in the cerebral cortex of the macaque. Ph.D. thesis. Katholieke Universiteit Leuven; 2006.
- Nelissen K, Luppino G, Vanduffel W, Rizzolatti G, Orban G. Observing others: multiple action representation in the frontal lobe. *Science* 2005;310:332–6.
- Nelissen K, Vanduffel W, Orban G. Charting the lower superior temporal region, a new motion-sensitive region in monkey superior temporal sulcus. *J Neurosci* 2006;26(22):5929–47.
- Neumann J, von Cramon D, Forstmann B, Zysset S, Lohmann G. The parcellation of cortical areas using replicator dynamics in fMRI. *Neuroimage* 2006;32(1): 208–19.
- Ng B, Abugharbieh R, McKeown M. Functional segmentation of fMRI data using adaptive non-negative PCA ANSPCA. In: Yang G, editor. MICCAI 2009. Lecture notes in computer science, vol. 5762. Berlin/Heidelberg: Springer-Verlag; 2009. p. 490–7.
- Nocedal J, Wright S. Numerical optimization. Berlin/New York: Springer-Verlag; 2006.
- Orban G, Vanduffel W. Comment on devlin and poldrack. *Neuroimage* 2007;37(4): 1057–8.
- Pandya D, Seltzer B. Intrinsic connections and architectonics of posterior parietal cortex in the rhesus-monkey. *J Comp Neurol* 1982;204(2):196–210.
- Penny W, Friston K. Mixtures of general linear models for functional neuroimaging. *IEEE Trans Med Imaging* 2003;22(4):504–14.
- Poldrack R. Region of interest analysis for fMRI. *Soc Cogn Affect Neurosci* 2007;2:67–70.
- Politis D, Romano J. The stationary bootstrap. *J Am Stat Assoc* 1994;89:1303–13.
- Ragheh M, Engstrom M, Knutsson H, Soderfeldt B, Lundberg P. Restricted canonical correlation analysis in functional MRI – validation and a novel thresholding technique. *J Magn Reson Imaging* 2009;29(1):146–54.
- Rajapakse J, Zhou J. Learning effective brain connectivity with dynamic Bayesian networks. *Neuroimage* 2007;37:749–60.
- Ramnani N, Behrens T, Penny W, Matthews P. New approaches for exploring anatomical and functional connectivity in the human brain. *Biol Psychiatry* 2004;56(9):613–9.
- Rizzolatti G, Luppino G. The cortical motor system. *Neuron* 2001;31:889–901.
- Roebroeck A, Formisano E, Goebel R. Mapping directed influence over the brain using granger causality and fMRI. *Neuroimage* 2005;25(1):230–42.
- Rozzi S, Calzavara R, Belmalih A, Borra E, Gregoriou G, Matelli M, et al. Cortical connections of the inferior parietal cortical convexity of the macaque monkey. *Cereb Cortex* 2006;16:1389–417.
- Sato J, Fujita A, Cardoso E, Thomaz C, Brammer M, Amaro E. Analyzing the connectivity between regions of interest: an approach based on cluster granger causality for fMRI data analysis. *Neuroimage* 2010;52(4):1444–55.
- Shawe-Taylor J, Cristianini N. Kernel methods for pattern analysis. New York: Cambridge University Press; 2004.
- Takane Y, Hwang H, Abdi H. Regularized multiple-set canonical correlation analysis. *Psychometrika* 2008;73(4):753–75.
- Thirion B, Flandin G, Pinel P, Roche A, Ciuciu P, Poline JB. Dealing with the shortcomings of spatial normalization: multi-subject parcellation of fMRI datasets. *Hum Brain Mapp* 2006;27(8):678–93.
- Tononi G, McIntosh A, Russell D, Edelman G. Functional clustering: identifying strongly interactive brain regions in neuroimaging data. *Neuroimage* 1998;7:133–49.
- Via J, Santamaria I, Perez J. A learning algorithm for adaptive canonical correlation analysis of several data sets. *Neural Netw* 2007;20(1):139–52.
- Worsley K, Poline JB, Friston K, Evans A. Characterizing the response of PET and fMRI data using multivariate linear methods. *Neuroimage* 1998;6:305–19.
- Young M, Scannell J, Burns G, Blakemore C. Analysis of connectivity: neural systems in the cerebral cortex. *Rev Neurosci* 1994;5(3):227–50.
- Zarahn E, Aguirre G, D'Esposito M. Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *Neuroimage* 1997;5:179–97.
- Zass R, Shashua A. Nonnegative sparse PCA. In: Advances in neural information processing systems; 2006. p. 1561–8.
- Zheng X, Rajapakse J. Learning functional structure from fMRI images. *Neuroimage* 2006;31(4):1601–13.