Functional Connectivity:

Eigenimages and multivariate analyses

Karl J Friston and Christian Büchel

The Wellcome Dept. of Cognitive Neurology,
University College London
Queen Square, London, UK WC1N 3BG
Tel (44) 020 7833 7456
Fax (44) 020 7813 1445
email k.friston@fil.ion.ucl.ac.uk

Contents

I. INTRODUCTION

This chapter is concerned with the characterisation of imaging data from a multivariate perspective. This means that the observations at each voxel are considered jointly with explicit reference to the interactions among brain regions. The concept of functional connectivity is reviewed and provides the basis for understanding what eigenimages represent and how they can be interpreted. Having considered the nature of eigenimages and variations on their applications, we then turn to a related approach that, unlike eigenimage

analysis, is predicated on a statistical model. This approach is called multivariate analysis of variance (ManCova) and uses canonical variate analysis to create canonical images. The integrated and distributed nature of neurophysiological responses to sensorimotor or cognitive challenge makes a multivariate perspective particularly appropriate, if not necessary for functional integration.

A Functional integration and connectivity

A landmark meeting that took place on the morning of August 4th 1881 highlighted the difficulties of attributing function to a cortical area, given the dependence of cerebral activity on underlying connections (Phillips *et al* 1984). Goltz, although accepting the results of electrical stimulation in dog and monkey cortex, considered the excitation method inconclusive, in that the movements elicited might have originated in related pathways, or current could have spread to distant centres. Despite advances over the past century, the question remains; are the physiological changes elicited by sensorimotor or cognitive challenges explained by functional segregation, or by integrated and distributed changes mediated by neuronal connections? The question itself calls for a framework within which to address these issues. *Functional and effective connectivity* are concepts critical to this framework.

1 Origins and definitions

In the analysis of neuroimaging time-series functional connectivity is defined as the *correlations between spatially remote neurophysiological events*. This definition provides a simple characterisation of functional interactions. The alternative is effective connectivity (*i.e. the influence one neuronal system exerts over another*). These concepts originated in the analysis of separable spike trains obtained from multiunit electrode recordings (Gerstein and Perkel 1969). Functional connectivity is simply a statement about the observed correlations; it does not comment on how these correlations are mediated. For example, at the level of multiunit micro-electrode recordings, correlations can result from *stimulus-locked transients,* evoked by a common afferent input, or reflect *stimulus-induced oscillations;* phasic coupling of neural assemblies, mediated by synaptic connections. Effective connectivity is closer to the notion of a connection and can be defined as *the influence one neural system exerts over another*, either at a synaptic (c.f. synaptic efficacy) or cortical level. Although functional and effective connectivity can be invoked at a conceptual level in both neuroimaging and electrophysiology they differ fundamentally at a

practical level. This is because the time-scales and nature of neurophysiological measurements are very different (seconds *vs.* milliseconds and hemodynamic *vs.* spike trains). In electrophysiology it is often necessary to remove the confounding effects of stimulus-locked transients (that introduce correlations *not* causally mediated by direct neural interactions) in order to reveal an underlying connectivity. The confounding effect of stimulus-evoked transients is less problematic in neuroimaging because promulgation of dynamics from primary sensory areas onwards *is* mediated by neuronal connections (usually reciprocal and interconnecting). However it should be remembered that functional connectivity is not necessarily due to effective connectivity (*e.g.* common neuromodulatory input from ascending aminergic neurotransmitter systems or thalamo-cortical afferents) and, where it is, effective influences may be indirect (*e.g.* polysynaptic relays through multiple areas).

II. EIGENIMAGES, MULTIDIMENSIONAL SCALING AND OTHER DEVICES

In what follows we introduce a number of techniques (eigenimage analysis, multidimensional scaling, partial least squares and generalised eigenimage analysis) using functional connectivity as a reference. Emphasis is placed on the relationships between these techniques. For example, eigenimage analysis is equivalent to principal component analysis and the variant of multidimensional scaling considered here is equivalent to principal coordinates analysis. Principal components and coordinates analyses are predicated on exactly the same eigenvector solution and from a mathematical perspective are essentially the same thing.

A Measuring a pattern of correlated activity

Here we introduce a simple way of measuring the amount a pattern of activity (representing a connected brain system) contributes to the functional connectivity or variance-covariances observed in the imaging data. Functional connectivity is defined in terms of statistical dependencies among neurophysiological measurement. If we assume these measurements conform to Gaussian assumptions then we need only characterise their correlations or

covariance (correlations are normalised covariances)[1]. The point to point functional connectivity between one voxel and another is not usually of great interest. The important aspect of a covariance structure is the pattern of correlated activity subtended by (an enormous number of) pairwise covariances. In measuring such patterns it is useful to introduce the concept of a *norm*. Vector and matrix norms serve the same purpose as absolute values for scalar quantities. In other words, they furnish a measure of distance. One frequently used norm is the 2-norm, which is the length of a vector. The vector 2-norm can be used to measure the degree to which a particular pattern of brain activity contributes to a covariance structure. If a pattern is described by a column vector (p), with an element for each voxel, then the contribution of that pattern to the covariance structure can be measured by the 2-norm of $Mp = \|Mp\|$. M is a (mean-corrected) matrix of data with one row for each successive scan and one column for each voxel:

$$\|Mp\|^2 = p^T M^T Mp \qquad 1$$

(^{T} denotes transposition). Put simply the 2-norm is a number that reflects the amount of variance-covariance or functional connectivity that can be accounted for by a particular distributed pattern. It should be noted that the 2-norm only measures the pattern of interest. There may be many other important patterns of functional connectivity. This fact begs the question "what are the most prevalent patterns of coherent activity?" To answer this question one turns to eigenimages or spatial modes.

B Eigenimages and spatial modes

In this section the concept of eigenimages or spatial modes is introduced in terms of patterns of activity defined above. We show that spatial modes are simply those patterns that account for the most variance-covariance (*i.e.* have the largest 2-norm).

Eigenimages or spatial modes are most commonly obtained using singular value decomposition (*SVD*). *SVD* is an operation that decomposes an original time-series (M) into two sets of orthogonal vectors (patterns in space and patterns in time) V and U where:

[1] Clearly neuronal processes are not necessarily Gaussian. However, we can still characterise the second order dependencies with the correlations. Higher-order dependencies would involve computing cumulants as described in final chapter of this section.

$$[U,S,V] = SVD(M)$$
$$M = USV^T \qquad 2$$

U and V are unitary orthogonal matrices $U^TU = 1$, $V^TV = 1$ and $V^TU = 0$ (the sum of squares of each column is unity and all the columns are uncorrelated) and S is a diagonal matrix (only the leading diagonal has non-zero values) of decreasing singular values. The singular value of each eigenimage is simply its 2-norm. Because *SVD* maximises the first singular value, the first eigenimage is the pattern that accounts for the greatest amount of the variance-covariance structure. In summary, *SVD* and equivalent devices are powerful ways of decomposing an imaging time-series into a series of orthogonal patterns than embody, in a step-down fashion, the greatest amounts of functional connectivity. Each eigenvector (column of V) defines a distributed brain system that can be displayed as an image. The distributed systems that ensue are called *eigenimages* or *spatial modes* and have been used to characterise the spatiotemporal dynamics of neurophysiological time-series from several modalities. Including, multiunit electrode recordings (Mayer-Kress *et al* 1991), EEG (Friedrich *et al* 1991), MEG (Fuchs *et al* 1992), PET (Friston *et al* 1993a) and functional MRI (Friston *et al* 1993b). Interestingly in fMRI the application of eigenimage that has attracted the most interest is in characterising functional connections while the brain is at 'rest'. See Biswal *et al* (1995).

Many readers will notice that the eigenimages associated with the functional connectivity or covariance matrix are simply principal components of the time-series. In the EEG literature one sometimes comes across the Karhunen-Loeve expansion which is employed to identify spatial modes. If this expansion is in terms of eigenvectors of covariances (and it usually is), then the analysis is formally identical to the one presented above.

One might ask what the column vectors of U in Eq(2) correspond to. These vectors are the time-dependent profiles associated with each eigenimage known as *eigenvariates*. They reflect the extent to which an eigenimage is expressed in each experimental condition or over time. See Figure 1 for a simple schematic illustrating the decomposition of a time-series into orthogonal modes. This is sometimes called spectral decomposition. Eigenvariates play an important role in the functional attribution of distributed systems defined by eigenimages. This point and others will be illustrated in the next.

C Mapping function into anatomical space - Eigenimage analysis

To illustrate the approach, we will use the PET word generation study used in previous chapters. The data were obtained from five subjects scanned 12 times whilst performing one of two verbal tasks in alternation. One task involved repeating a letter presented aurally at one per two seconds (*word shadowing*). The other was a paced verbal fluency task, where subjects responded with a word that began with the heard letter (*word generation*). To facilitate inter-subject pooling, the data were realigned and spatially normalised and smoothed with an isotropic Gaussian kernel (FWHM of 16mm). The data were then subject to an AnCova (with 12 conditions, subject effects and global activity as a confound). Voxels were selected using a conventional SPM{F} to identify those significant at $p<0.05$ (uncorrected). The time-series of condition-specific effects, from each of these voxels, were entered into a mean corrected data matrix *M* with 12 rows (one for each condition) and one column for each voxel.

M was subject to *SVD* as described above. The distribution of eigenvalues (Figure 2, lower left) suggests only two eigenimages are required to account for most of the observed variance-covariance structure. The first mode accounted for 64% and the second for 16% of the variance. The first eigenimage V_1 is shown in Figure 2 (top) along with the corresponding eigenvariate U_1 (lower right). The first eigenimage has positive loadings in the anterior cingulate, the left DLPFC, Broca's area, the thalamic nuclei and in the cerebellum. Negative loadings were seen bitemporally and in the posterior cingulate. According to U_1 this eigenimage is prevalent in the verbal fluency tasks with negative scores in word shadowing. The second spatial mode (not shown) had its highest positive loadings in the anterior cingulate and bitemporal regions (notably Wernicke's area on the left). This mode appears to correspond to a highly non-linear, monotonic time effect with greatest prominence in earlier conditions.

The *post hoc* functional attribution of these eigenimages is usually based on their eigenvariates (*U*). The first mode may represent an *intentional* system critical for the intrinsic generation of words in the sense that the key cognitive difference between verbal fluency and word shadowing is the intrinsic generation as opposed to extrinsic specification of word representations and implicit mnemonic processing. The second system, that includes the anterior cingulate, seems to be involved in habituation, possibly of attentional or perceptual set.

The is nothing 'biologically' important about the particular spatial modes obtained in this fashion, in the sense that one could 'rotate' the eigenvectors such that they were still

orthogonal and yet gave different eigenimages. The uniqueness of the particular solution given by *SVD* is that the first eigenimage accounts for the largest amount of variance-covariance and the second for the greatest amount that remains and so on. The reason that the eigenimages in the example above lend themselves to such a simple interpretation is that the variance introduced by experimental design (intentional) was substantially greater than that due to time (attentional) and both these sources were greater than any other effect. Other factors that ensure a parsimonious characterisation of a time-series, with small numbers of well-defined modes include (i) smoothness in the data and (ii) using only voxels that showed a non-trivial amount of change during the scanning session.

D. Mapping anatomy into functional space - multidimensional scaling

In the previous section the functional connectivity matrix was used to define associated eigenimages or spatial modes. In this section functional connectivity is used in a different way, namely, to constrain the proximity of two cortical areas in some functional space (Friston *et al* 1996a). The objective here is to transform anatomical space so that the distance between cortical areas is directly related to their functional connectivity. This transformation defines a new space whose topography is purely functional in nature. This space is constructed using multidimensional scaling or principal coordinates analysis (Gower 1966).

Multidimensional scaling (MDS) is a descriptive method for representing the structure of a system. Based on pairwise measures of similarity or confusability (Torgerson 1958; Shepard 1980). The resulting multidimensional spatial configuration of a system's elements embody, in their proximity relationships, comparative similarities. The technique was developed primarily for the analysis of perceptual spaces. The proposal that stimuli be modelled by points in space, so that perceived similarity is represented by spatial distances, goes back to the days of Isaac Newton (1794).

Imagine k measures from n voxels plotted as n points in a k-dimensional space (k-space). If they have been normalised to zero mean and unit sum of squares, these points will fall on an k-1 dimensional sphere. The closer any two points are to each other, the greater their correlation or functional connectivity (in fact the correlation is a cosine of the angle subtended at the origin). The distribution of these points embodies the functional topography. A view of this distribution, that reveals the greatest structure, is simply obtained by rotating the points to maximise their apparent dispersion (variance). In other words one looks at the subspace with the largest 'volume' spanned by the principal axes of the n points

in *k*-space. These principal axes are given by the eigenvectors of MM^T. *i.e.* the column vectors of U_1. From Eq(2):

$$MM^T = U\lambda U^T$$
$$\lambda = SS^T \qquad 3$$

Let Q be the matrix of desired coordinates derived by simply projecting the original data onto axes defined by U: where $Q = M^T U$. Voxels that have a correlation of unity will occupy the same point in MDS space. Voxels that have uncorrelated dynamics will be $\pi/2$ apart. Voxels that are negatively but totally correlated (correlation = -1) will be maximally separated on the opposite sides of the MDS hyperspace. Profound negative correlations denote a functional association that is modelled in MDS functional space as diametrically opposed locations on the hyper-sphere. In other words, two regions with profound negative correlations will form two 'poles' in functional space.

Following normalisation to unit sum of squares over each column M (the adjusted data matrix from the word generation study above) the data were subjected to singular value decomposition according to Eq(2) and the coordinates Q of the voxels in MDS functional space were computed. Recall that only two eigenvalues exceed unity (Figure 2; right), suggesting a functional space that is essentially two dimensional. The locations of voxels in this two-dimensional subspace are shown in Figure 3 (lower row) by rendering voxels from different regions in different colours. The anatomical regions corresponding to the different colours are shown in the upper row. Anatomical regions were selected to include those parts of the brain that showed the greatest variance during the 12 conditions. Anterior regions (Figure 3; right) included the mediodorsal thalamus (blue), the dorsolateral prefrontal cortex (DLPFC), Broca's area (red) and the anterior cingulate (green). Posterior regions (Figure 3; left) included the superior temporal regions (red), the posterior superior temporal regions (blue) and the posterior cingulate (green). The corresponding functional spaces (Figure 3; lower rows) reveal a number of things about the functional topography elicited by this set of activation tasks. First, each anatomical region maps into a relatively localised portion of functional space. This preservation of local contiguity reflects the high correlations within anatomical regions, due in part to smoothness of the original data and to high degrees of intra-regional functional connectivity. Secondly, the anterior regions are almost in juxtaposition as are posterior regions. However, the confluence of anterior and posterior regions forms two diametrically opposing poles (or one axis). This configuration suggests an

anterior-posterior axis with prefronto-temporal and cingulo-cingulate components. One might have predicted this configuration by noting that the anterior regions had high positive loadings on the first eigenimage (see Figure 2) while the posterior regions had high negative loadings. Thirdly, within the anterior and posterior sets of regions certain generic features are evident. The most striking is the particular ordering of functional interactions. For example, the functional connectivity between posterior cingulate (green) and superior temporal regions (red) is high and similarly for the superior temporal (red) and posterior temporal regions (blue). Yet the posterior cingulate and posterior temporal regions show very little functional connectivity (they are $\pi/2$ apart or, equivalently, subtend 90 degrees at the origin).

These results are consistent with known anatomical connections. For example DLPFC - anterior cingulate connections, DLPFC - temporal connections, bitemporal commissural connections and mediodorsal thalamic - DLPFC projections have all been demonstrated in non-human primates (Goldman-Rakic 1988). The mediodorsal thalamic region and DLPFC are so correlated that one is embedded within the other (purple area). This is pleasing given the known thalamo-cortical projections to DLPFC.

E. Functional connectivity between systems - Partial least squares

Hitherto, we have been dealing with functional connectivity between two voxels. The same notion can be extended to functional connectivity between two systems by noting that there is no fundamental difference between the dynamics of one voxel and the dynamics of a distributed system or pattern. The functional connectivity between two systems is simply the correlation or covariance between their time-dependent activity. The time-dependent activity of a system or pattern p_i is given by:

$$\begin{aligned} v_i &= Mp_i \\ C_{ij} &= v_i^T v_j = p_i^T M^T M p_j \end{aligned} \qquad 4$$

where C_{ij} is the functional connectivity between the systems described by vectors p_i and p_j. Consider functional connectivity between two systems in separate parts of the brain, for example the right and left hemispheres. Here the data matrices (M_i and M_j) derive from different sets of voxels and Eq(4) becomes:

$$C_{ij} = v_i^T v_j = p_i^T M_i^T M_j p_j \qquad 5$$

If one wanted to identify the intra-hemispheric systems that showed the greatest inter-hemispheric functional connectivity (*i.e*. covariance) one would need to identify the pair of vectors p_i and p_j that maximise C_{ij} in Eq(5). *SVD* finds another powerful application in doing just this where:

$$\begin{aligned} [U,S,V] &= SVD(M_i^T M_j) \\ M_i^T M_j &= USV^T \\ U^T M_i^T M_j V &= S \end{aligned} \qquad 6$$

The first columns of U and V represent the singular images that correspond to the two systems with the greatest amount of functional connectivity (the singular values in the diagonal matrix S). In other words *SVD* of the (generally asymmetric) cross-covariance matrix, based on time-series from two anatomically separate parts of the brain, yields a series of paired vectors (paired columns of U and V) that, in a step-down fashion, define pairs of brain systems that show the greatest functional connectivity. This particular application of *SVD* is also know as *partial least squares* and has been proposed for analysis of designed activation experiments where the two data matrices comprise (i) an imaging time-series and (ii) a set of behavioural or task parameters (Macintosh *et al* 1996). In this application the paired singular vectors correspond to (i) a singular image and (ii) a set of weights that give the linear combination of task parameters that show the maximal covariance with the corresponding singular image.

F. Differences in functional connectivity - Generalised eigenimages

In this section we introduce an extension of eigenimage analysis using the solution to the generalised eigenvalue problem. This problem involves finding the eigenvector solution that involves two functional connectivity or covariance matrices and can be used to find the eigenimage that is maximally expressed in one time-series relative to another. In other words it can find a pattern of distributed activity that is most prevalent in one data set and least expressed in another. The example used to illustrate this idea is fronto-temporal functional disconnection in schizophrenia (see Friston *et al* 1996b).

The notion that schizophrenia represents a disintegration or fractionation of the psyche is as old as its name, introduced by Bleuler (1911) to convey a 'splitting' of mental faculties. Many of Bleuler's primary processes, such as 'loosening of associations' emphasise a fragmentation and loss of coherent integration. In what follows we assume that this mentalistic 'splitting' has a physiological basis, and furthermore that both the mentalistic and physiological disintegration have precise and specific characteristics that can be understood in terms of functional connectivity

The idea is that although localised pathophysiology in cortical areas may be a sufficient explanation for some signs of schizophrenia it does not suffice as a rich or compelling explanation for the symptoms of schizophrenia. The conjecture is that symptoms such as hallucinations and delusions are better understood in terms of abnormal interactions or impaired integration between different cortical areas. This dysfunctional integration, expressed at a physiological level as abnormal functional connectivity, is measurable with neuroimaging and observable at a cognitive level as a failure to integrate perception and action that manifests as clinical symptoms. The distinction between a regionally specific pathology and a pathology of interaction can be seen in terms of a first order effect (*e.g.* hypofrontality) and a second order effect that only exists in the relationship between activity in the prefrontal cortex and some other (*e.g.* temporal) region. In a similar way psychological abnormalities can be regarded as first order (e.g. a poverty of intrinsically cued behaviour in psychomotor poverty) or second order (*e.g.* a failure to integrate intrinsically cued behaviour and perception in reality distortion).

1 The generalised eigenvalue solution

Suppose that we want to find a pattern embodying the greatest amount of functional connectivity in control subjects, relative to schizophrenic subjects (*e.g.* fronto-temporal covariance). To achieve this we identify an eigenimage that reflects the most functional connectivity in control subjects relative to a schizophrenic group (*d*). This eigenimage is obtained by using a generalised eigenvector solution:

$$\begin{aligned} C_i^{-1} C_j d &= d\lambda \\ C_j d &= C_i d\lambda \end{aligned} \qquad 7$$

where C_i and C_j are the two functional connectivity matrices. The generalised eigenimage d is essentially a single pattern that maximises the ratio of the 2-norm measure [Eq(1)] when

applied to C_i and C_j . Generally speaking, these matrices could represent data from two [groups of] subjects or from the same subject[s] scanned under different conditions. In the present example we use connectivity matrices from control subjects and people with schizophrenia showing pronounced psychomotor poverty.

The data were acquired from two groups of six subjects. Each subject was scanned six times during the performance of three word generation tasks (A B C C B A). Task A was a verbal fluency task, requiring subjects to respond with a word that began with a heard letter. Task B was a semantic categorisation task in which subjects responded "man-made" or "natural", depending on a heard noun. Task C was a word-shadowing task in which subjects simply repeated what was heard. In the current context, the detailed nature of the tasks is not very important. They were used to introduce variance and covariance in activity that could support an analysis of functional connectivity.

The groups comprised six control subjects and six schizophrenic patients. The schizophrenic subjects produced less than 24 words on a standard (one minute) FAS verbal fluency task (generating words beginning with the letters 'F', 'A' and 'S'). The results of a generalised eigenimage analysis are presented in Figure 4. As expected the pattern that best captures differences between the two groups involves prefrontal and temporal cortices. Negative correlations between left DLPFC and bilateral superior temporal regions are found (Figure 4; upper panels). The amount to which this pattern was expressed in each individual group is shown in the lower panel using the appropriate 2-norm $\left\| d^T C_i d \right\|$. It is seen that this eigenimage, whilst prevalent in control subjects, is uniformly reduced in schizophrenic subjects.

G. Summary

In the preceding sections we have seen how eigenimages can be framed in terms of functional connectivity and the relationships among eigenimage analysis, multidimensional scaling, partial least squares and generalised eigenimage analysis. In the next section we use the generative models perspective, described in the previous chapter, to take component analysis into the nonlinear domain.

III NONLINEAR PCA AND ICA

A Generative models

Recall from the previous chapter how generative models of data could be framed in terms of a *prior* distribution over causes $p(v;\theta)$ and a *generative* distribution or likelihood of the inputs given the causes $p(u \mid v;\theta)$. For example, factor analysis corresponded to the generative model

$$\begin{aligned} p(v;\theta) &= N(v:0,1) \\ p(u|v;\theta) &= N(u:\theta v,\Sigma) \end{aligned} \qquad 8$$

Namely, the underlying causes of inputs are independent normal variates that are mixed linearly and added to Gaussian noise to form inputs. In the limiting case of $\Sigma \to 0$ the model become deterministic and conforms to PCA. By simply assuming non-Gaussian priors one can specify generative models for sparse coding

$$\begin{aligned} p(v;\theta) &= \prod p(v_i;\theta) \\ p(u|v;\theta) &= N(u:\theta v,\Sigma) \end{aligned} \qquad 9$$

where $p(v_i;\theta)$ are chosen to be suitably sparse (*i.e.* heavy-tailed) with a cumulative density function that corresponds to the squashing function below. The deterministic equivalent of sparse coding is ICA that obtains when $\Sigma \to 0$. These formulations allow us to consider simple extensions of PCA by looking at nonlinear versions of the underlying generative model.

B Nonlinear PCA

Despite its exploratory power, eigenimage analysis is fundamentally limited because the particular modes obtained are uniquely determined by constraints that are biologically implausible. This represents an inherent limitation on the interpretability and usefulness of eigenimage analysis. The two main limitations of conventional eigenimage analysis are that the decomposition of any observed time-series is in terms of linearly separable components. Secondly, the spatial modes are somewhat arbitrarily constrained to be orthogonal and account, successively, for the largest amount of variance. From a biological perspective, the linearity constraint is

a severe one because it precludes interactions among brain systems. This is an unnatural restriction on brain activity, where one expects to see substantial interactions that render the expression of one mode sensitive to the expression of others. Nonlinear PCA attempts to circumvent these sorts of limitations.

The generative model implied by Eq(8), when $\Sigma \to 0$, is linear and deterministic

$$\begin{aligned} p(v;\theta) &= N(v:0,1) \\ u &= \theta v \end{aligned} \qquad 10$$

Here the causes v correspond to the eigenvariates and the model parameters to scaled eigenvectors $\theta = VS$. u is the observed data or image that comprised each row of M above. This linear generative model $G(v,\theta) = \theta v$ can now be generalised to any static nonlinear model by taking a second order approximation

$$\begin{aligned} p(v;\theta) &= N(v:0,1) \\ u &= G(v,\theta) \\ &= \sum_i V_i v_i + \tfrac{1}{2}\sum_{ij} V_{ij} v_i v_j + \ldots \\ V_i &= \frac{\partial G}{\partial v} \\ V_{ij} &= \frac{\partial^2 G}{\partial v_i \partial v_j} \end{aligned} \qquad 11$$

This nonlinear model has two sorts of modes. First-order modes V_i that mediate the effect of any orthogonal cause on the response (*i.e.* maps the causes onto voxels directly) and second-order modes V_{ij} which map interactions among causes onto the measured response. These second-order modes could represent the distributed systems implicated in the interaction between various experimentally manipulated causes. See the example below.

The identification of the first- and second-order modes proceeds using expectation maximisation (EM) as described in the previous chapter. In this instance the algorithm can be implemented as a simple neural net with forward connections from the data to the causes and backward connections from the causes to the predicted data. The **E**-step corresponds to *recognition* of the causes by the forward connections using the current estimate of the first-order modes and the **M**-Step adjusts these connections to minimise the prediction error of the

generative model in Eq(11), using the recognised causes. These schemes (*e.g.* Kramer 1991, Karhunen and Joutsensalo 1994, Friston 2000) typically employ a 'bottleneck' architecture that forces the inputs through a small number of nodes (see the insert in Figure 5). The output from these nodes then diverges to produce the predicted inputs. After learning, the activity of the bottleneck nodes can be treated as estimates of the causes. These representations obtain by projection of the input onto a low-dimensional curvilinear manifold that is defined by the activity of the bottleneck. Before looking at an empirical example we will briefly discuss ICA.

C. Independent Component Analysis

ICA represents another way of generalising the linear model used by PCA. This is achieved, not through nonlinearities, but by assuming non-Gaussian priors. The non-Gaussian form can be specified by a nonlinear transformation of the causes $\tilde{v} = \sigma(v)$ that renders then normally distributed, such that when $\Sigma \to 0$, in Eq(9) we get

$$\begin{aligned} p(\tilde{v};\theta) &= N(\tilde{v};0,1) \\ u &= \theta\sigma^{-1}(\tilde{v}) \\ \tilde{v} &= \sigma(\theta^{-1}u) \end{aligned} \qquad 12$$

This is not the conventional to present ICA but is used here to connect the models for PCA and ICA. The form of the nonlinear squashing function $\tilde{v} = \sigma(v)$ embodies our prior assumptions about the marginal distribution of the causes. These are usually supra-Gaussian. There exist simple algorithms that implicitly minimise the objective function F (see previous chapter) using the covariances of the data. In neuroimaging, this enforces an ICA of independent spatial modes, because there are more voxels than scans (McKeown *et al* 1998). In EEG there are more time-bins than channels and the independent components are temporal in nature. The distinction between *spatial* and *temporal* ICA depends on whether one regard the Eq(12) as generating data over space or time. See Friston (1998) for a discussion of their relative merits. The important thing about ICA, relative to PCA, is that the prior densities model independent causes not just uncorrelated causes. This difference is expressed in terms of statistical dependencies beyond second order. See Stone (2002) for an introduction to these issues. .

D. An example

This example comes from Friston *et al* (2000)[1] and is based on an fMRI study of visual processing that was designed to address the interaction between colour and motion systems. We had expected to demonstrate that a 'colour' mode and 'motion' mode would interact to produce a second order mode reflecting. (i) Reciprocal interactions between extrastriate areas functionally specialised for colour and motion, (ii) interactions in lower visual areas mediated by convergent backwards efferents or (iii) interactions in the pulvinar mediated by cortico-thalamic loops).

1 Data acquisition and experimental design

A young subject was scanned under four different conditions, in 6 scan epochs, intercalated with a low-level (visual fixation) baseline condition. The four conditions were repeated 8 times in a pseudo-random order giving 384 scans in total or 32 stimulation/baseline epoch pairs. The four experimental conditions comprised the presentation of (i) radially moving dots and (ii) stationary dots, using (i) luminance contrast and (ii) chromatic contrast in a two by two-factorial design. Luminance contrast was established using isochromatic stimuli (red dots on a red background or green dots on a green background). Hue contrast was obtained by using red (or green) dots on a green (or red) background and establishing isoluminance with flicker photometry. In the two movement conditions the dots moved radially from the centre of the screen, at 8 degrees per second to the periphery, where they vanished. This creates the impression of optical flow. By using these stimuli we hoped to excite activity in a visual motion system and one specialised for colour processing. Any interaction between these systems would be expressed in terms of motion-sensitive responses that depended on the hue or luminance contrast subtending that motion.

2 Nonlinear PCA

The data were reduced to an eight-dimensional subspace using SVD and entered into a nonlinear PCA using two causes. The functional attribution of the resulting sources was established by looking at the expression of the corresponding first-order modes over the four conditions (right lower panels in Figure 5). This expression is simply the score on the first

[1] Although an example of nonlinear PCA, the generative model actually used finessed Eq(10) with a nonlinear function of the second order terms.

$$u = G(v) = \sum_i V_i v_i + \tfrac{1}{2}\sum_{ij} V_{ij}\sigma(v_i v_j)$$

This endows the causes with a unique scaling.

principal component over all 32 epoch-related responses for each cause. The first mode is clearly a motion-sensitive mode but one that embodies some colour preference in the sense that the motion-dependent responses of this system are accentuated in the presence of colour cues. This was not quite what we had anticipated; the first-order effect contains what would functionally be called an interaction between motion and colour processing. The second source appears to be concerned exclusively with colour processing. The corresponding anatomical profiles are shown in Figure 5 (left panels). The first-order mode, that shows both motion and colour-related responses shows high loadings in bilateral motion sensitive complex V5 (Brodmann Areas 19 and 37 at the occipto-temporal junction) and areas traditionally associated with colour processing (V4 - the lingual gyrus). The second first order mode is most prominent in the hippocampus, parahippocampal and related lingual cortices on both sides. In summary the two first-order modes comprise: (i) an extrastriate cortical system including V5 and V4 that is responds to motion, and preferentially so when motion is supported by colour cues. (ii) A [para]hippocampal/lingual system that is concerned exclusively with colour processing, above and beyond that accounted for by the first system. The critical question is where do these modes interact?

The interaction between the extrastriate and [para]hippocampal/lingual systems conforms to the second order mode in the lower panels. This mode is highlights the pulvinar of the thalamus and V5 bilaterally. This is a pleasing result in that it clearly implicates the thalamus in the integration of extrastriate and [para]hippocampal systems. This integration being mediated by recurrent [sub]cortico-thalamic connections. It is also a result that would not have obtained from a conventional SPM analysis. Indeed we looked for an interaction between motion and colour processing and did not see any such effect in the pulvinar.

E. Summary

We have reviewed eigenimage analysis and generalisations based on nonlinear and non-Gaussian generative models. All the techniques above are essentially descriptive, in that they do not allow one to make any statistical inferences about the characterisations that obtain. In the second half of this chapter we turn to multivariate techniques that do embody statistical inference and explicit hypothesis testing. We will introduce *canonical images* that can be thought of as statistically informed eigenimages pertaining to a particular effect introduced by experimental design. We have seen that patterns can be identified using the generalised eigenvalue solution that are maximally expressed in one covariance structure relative to another. Consider now using this approach where the first covariance matrix

reflected the effects we were interested in, and the second embodied covariances due to error. This corresponds to canonical image analysis, and is considered in the following sections.

IV. MANCOVA AND CANONICAL IMAGE ANALYSIS

A Introduction

In the following sections we review multivariate approaches to the analysis of functional imaging studies. The exemplar analysis described uses standard multivariate techniques to make statistical inferences about activation effects and to describe their important features. Specifically, we introduce multivariate analysis of covariance (ManCova) and canonical variates analysis (CVA) to characterise activation effects. This approach characterises the brain's response in terms of functionally connected and distributed systems in a similar fashion to eigenimage analysis. Eigenimages figure in the current analysis in the following way. A problematic issue in multivariate analysis of functional imaging data is that the number of samples (*i.e.* scans) is usually very small in relation to the number of components (*i.e.* voxels) of the observations. This issue is resolved by analysing the data, not in terms of voxels, but in terms of eigenimages, because the number of eigenimages is much smaller than the number of voxels. The importance of the multivariate analysis that ensues can be summarised as follows:. (i) Unlike eigenimage analysis, it provides for statistical inferences (based on classical p-values) about the significance of the brain's response in terms of some hypothesis. (ii) The approach implicitly takes account of spatial correlations in the data without making any assumptions. (iii) The canonical variate analysis produces generalised eigenimages (canonical images) that capture the activation effects, while suppressing the effects of noise or error. (iv) The theoretical basis is well established and can be found in most introductory texts on multivariate analysis (see also Friston *et al* 1996c).

Although useful, in a descriptive sense, eigenimage analysis and related approaches are not generally considered as 'statistical' methods that can be used to make statistical inferences; they are mathematical devices that simply identify prominent patterns of correlations or functional connectivity. It must be said, however, that large sample, asymptotic, multivariate normal theory could be used to make some inferences about the relative contributions of each eigenimage (*e.g.* tests for non-sphericity) if a sufficient number of scans were available. In what follows we observe that multivariate analysis of covariance (ManCova) with canonical variate analysis combines some features of statistical parametric mapping and eigenimage

analysis. Unlike statistical parametric mapping, ManCova is multivariate. In other words, it considers as one observation all voxels in a single scan. The importance of this multivariate approach is that effects, due to activations, confounding effects and error effects, are assessed both in terms of effects at each voxel *and interactions among voxels*. This means one does not have to assume anything about spatial correlations (*c.f.* stationariness with Gaussian field models) to assess the significance of an activation effect. Unlike statistical parametric mapping these correlations are explicitly included in the analysis. The price one pays for adopting a multivariate approach is that inferences cannot be made about regionally specific changes (*c.f.* statistical parametric mapping). This is because the inference pertains to all the components (voxels) of a multivariate variable (not a particular voxel or set of voxels). Furthermore, because the spatial non-sphericity has to be estimated, without knowing the observations came from continuous spatially extended processes, the estimates are less efficient and inferences are less powerful.

In general, multivariate analyses are implemented in two steps. First, the significance of a hypothesised effect is assessed in terms of a p-value and secondly, if justified, the exact nature of the effect is determined. The analysis here conforms to this two-stage procedure. When the brain's response is assessed to be significant using ManCova, the nature of this response remains to be characterised. Canonical variate analysis (CVA) is an appropriate way to do this. The canonical images obtained with CVA are similar to eigenimages but are based on both the activation and error. CVA is closely related to de-noising techniques in EEG and MEG time-series analyses that use a generalised eigenvalue solution. Another way of looking at canonical images is to think of them as eigenimages that reflect functional connectivity due to activations, when spurious correlations due to error are explicitly discounted.

B. Dimension reduction and eigenimages

The first step in multivariate analysis is to ensure that the dimensionality (number of components or voxels) of the data is smaller than the number of observations. Clearly for images this is not the case, because there are more voxels than scans; therefore the data have to be transformed. The dimension reduction proposed here is straightforward and uses the scan-dependent expression Y of eigenimages as a reduced set of components for each multivariate observation (scan). Where:

$$[U,S,V] = SVD(M)$$
$$Y = US \qquad 13$$

As above M is a large matrix of adjusted voxel values with one column for each voxel and one row for each scan. Here 'adjusted' implies mean correction and removal of any confounds using linear regression. The eigenimages constitute the columns of U, another unitary orthonormal matrix, and their expression over scans corresponds to the columns of the matrix Y. Y has one column for each eigenimage and one row for each scan. In our work we use only the j columns of Y and U associated with eigenvalues greater than unity (after normalising each eigenvalue by the average eigenvalue).

C. The general linear model revisited

Recall the general linear model from previous chapters:

$$Y = X\beta + \varepsilon \qquad 14$$

where the errors are assumed to be independent and identically normally distributed. The design matrix X has one column for every effect (factor or covariate) in the model. The design matrix can contain both covariates and indicator variables reflecting an experimental design. β is the parameter matrix with one column vector of parameters for each mode. Each column of X has an associated unknown parameter. Some of these parameters will be of interest, the remaining parameters will not. We will partition the model accordingly.:

$$Y = X_1\beta_1 + X_0\beta_0 + \varepsilon \qquad 15$$

where X_1 represents a matrix of 0s or 1s depending on the level or presence of some interesting condition or treatment effect (*e.g.* the presence of a particular cognitive component) or the columns of X_1 might contain covariates of interest that could explain the observed variance in Y (*e.g.* dose of apomorphine or 'time on target'). X_0 corresponds to a matrix of indicator variables denoting effects that are not of any interest (*e.g.* of being a particular subject or block effect) or covariates of no interest (i.e. 'nuisance variables' such as global activity or confounding time effects).

D. Statistical inference

Significance is assessed by testing the null hypothesis that the effects of interest do not significantly reduce the error variance when compared to the remaining effects alone (or alternatively the null hypothesis that β_1 is zero). The null hypothesis is tested in the following way. The sum of squares and products matrix (SSPM) due to error is obtained from the difference between actual and estimated values of the response:

$$S_R = (Y - X\hat{\beta})^T (Y - X\hat{\beta}) \qquad 16$$

where the sums of squares and products due to effects of interest is given by

$$S_T = (X_1\hat{\beta}_1)^T (X_1\hat{\beta}_1) \qquad 17$$

The error sum of squares and products under the null hypothesis *i.e.* after discounting the effects of interest are given by:

$$S_0 = (Y - X_0\hat{\beta}_0)^T (Y - X_0\hat{\beta}_0) \qquad 18$$

The significance can now be tested with:

$$\lambda = \frac{S_R}{S_0} \qquad 19$$

This is Wilk's statistic (known as Wilk's Lambda). A special case of this test is Hotelling's T^2 test and applies when one simply compares one condition with another, *i.e.* X_1 has only one column (Chatfield and Collins 1980). Under the null hypothesis, after transformation, λ has Chi-squared distribution with degrees of freedom jh. The transformation is given by:

$$-(v - ((j - h + 1)/2)\ln\lambda \sim \chi^2_{jh} \qquad 20$$

where v are the degrees of freedom associated with error terms, equal to the number of scans (n) minus the number of effects modelled = n - rank(X). j is the number of eigenimages in

the *j*-variate response variable and *h* are the degrees of freedom associated with effects of interest = rank(X_1).

E. Characterising the effect

Having established that the effects of interest are significant (*e.g.* differences among two or more activation conditions) the final step is to characterise these effects in terms of their spatial topography. This characterisation uses canonical variates analysis or CVA. The objective is to find a linear combination (compound or contrast) of the components of *Y*, in this case the eigenimages, that best express the activation effects when compared to error effects. More exactly we want to find $\mathbf{c}_1$ such that the variance ratio:

$$\frac{c_c^T S_T c_1}{c_c^T S_R c_1} \qquad 21$$

is maximised. Let $z_1 = Yc_1$ where z_1 is the first canonical variate and c_1 is a canonical image (defined in the space of the spatial modes) that maximises this ratio. c_2 is the second canonical image that maximises the ratio subject to the constraints $c_i^T c_j = 0$ (and so on). The matrix of canonical images $c = [c_1, \ldots c_h]$ is given by solution of the generalised eigenvalue problem:

$$S_T c = S_R c \lambda \qquad 22$$

where λ is a diagonal matrix of eigenvalues. Voxel-space canonical images are obtained by rotating the canonical image in the columns of *c* back into voxel-space with the original eigenimages $C = Vc$. The columns of *C* now contain the voxel values of the canonical images. The *k*th column of *C* (the *k*th canonical image) has an associated canonical value equal to the *k*th leading diagonal element of λ times r/h. Note that the 'activation' effect is a multivariate one, with *j* components or canonical images. Normally only a few of these components have large canonical values and only these need to be reported. There are procedures based on distributional approximations of λ that allow inferences about the dimensionality of a response (number of canonical images). We refer the interested reader to Chatfield and Collins (1980) for further details.

1 Relationship to Eigenimage Analysis

When applied to adjusted data eigenimages correspond to the eigenvectors of S_T. These have an interesting relationship to the canonical images: On rearranging Eq(22), we note that the canonical images are eigenvectors of $S_R^{-1}S_T$. In other words, an eigenimage analysis of an activation study returns the eigenvectors that express the most variance due to the effects of interest. A canonical image, on the other hand, expresses the greatest amount of variance due to the effects of interest *relative to error*. In this sense, a CVA can be considered an eigenimage analysis that is 'informed' by the estimates of error and their correlations over voxels.

F An illustrative application

In this section we consider an application of the above theory to the word generation study in normal subjects, used in previous sections. We assessed the significance of condition-dependent effects by treating each of the 12 scans as a different condition. Note that we do not consider the word generation (or word shadowing) conditions as replications of the same condition. In other words, the first time one performs a word generation task is a different condition from the second time and so on. The (alternative) hypothesis adopted here states that there is a significant difference among the 12 conditions, but does not constrain the nature of this difference to a particular form. The most important differences will emerge from the CVA. Clearly one might hope that these differences will be due to word generation, but they might not be. This hypothesis should be compared with a more constrained hypothesis that considers the conditions as six replications of word shadowing and word generation. This latter hypothesis is more directed and explicitly compares word shadowing with word generation. This comparison could be tested in a single subject. The point is that the generality afforded by the current framework allows one to test very constrained (*i.e.* specific) hypotheses or rather general hypotheses about some unspecified activation effect[1]. We choose the latter case here because it places more emphasis on canonical images as descriptions of what has actually occurred during the experiment.

The design matrix partition for effects of interest X_1 had 12 columns representing the 12 different conditions. We designated subject effects, time and global activity as uninteresting confounds X_0. The adjusted data were reduced to 60 eigenvectors as described above. The first 14 eigenvectors had (normalised) eigenvalues greater than unity and were used in the

subsequent analysis. The resulting matrix data *Y*, with 60 rows (one for each scan) and 14 columns (one for each eigenimage) was subject to ManCova. The significance of the condition effects was assessed with Wilk's Lambda. The threshold for condition or activation effects was set at p = 0.02. In other words the probability of there being no differences among the 12 conditions was 2%.

1 Canonical Variates Analysis

The first canonical image and its canonical variate are shown in Figure 6. The upper panels show this system to include anterior cingulate and Broca's area, with more moderate expression in the left posterior infero-temporal regions (right). The positive components of this canonical image (left) implicate ventro-medial prefrontal cortex and bitemporal regions (right greater than left). One important aspect of these canonical images is their highly distributed yet structured nature, reflecting the distributed integration of many brain areas. The canonical variate expressed in terms of mean condition effects is seen in the lower panel of Figure 6. It is pleasing to note that the first canonical variate corresponds to the difference between word shadowing and verbal fluency.

Recall that the eigenimage in Figure 2 reflects the main pattern of correlations evoked by the mean condition effects and should be compared with the first canonical image in Figure 6. The differences between these characterisations of activation effects are informative: The eigenimage is totally insensitive to the reliability or error attributable to differential activation from subject to subject whereas the canonical image does reflect these variations. For example, the absence of the posterior cingulate in the canonical image and its relative prominence in the eigenimage suggests that this region is implicated in some subjects but not in others. The subjects that engage the posterior cingulate must do so to some considerable degree because the average effects (represented by the eigenimage) are quite substantial. Conversely, the medial prefrontal cortical deactivations are a more pronounced feature of activation effects than would have been inferred on the basis of the eigenimage analysis. These observations beg the question 'which is the best characterisation of functional anatomy'? Obviously there is no simple answer but the question speaks to an important point. A canonical image characterises a response *relative to error,* by partitioning the observed variance into effects of interest and a residual variation about these effects. Experimental design, a hypothesis, and the inferences that are sought determine this

---

[1] This is in analogy to the use of the SPM{F}, relative to more constrained hypotheses tested with SPM{T}, in conventional mass-univariate approaches.

partitioning. An eigenimage does not embody any concept of error and is not constrained by any hypothesis.

G Multivariate models

CVA rests upon *i.i.d.* assumptions about the errors over time. Violation of these assumptions has motivated the study of multivariate linear models (MLMs) for neuroimaging that allow for temporal non-sphericity (see Worsley *et al* 1997). Although MLMs are important this book has chosen to focus more on univariate models. There is a reason for this: Any MLM can be reformulated as a univariate model by simply vectorising the multivariate response. For example the MLM;

$$\begin{aligned} Y &= X\beta + \varepsilon \\ [y_1, \ldots y_j] &= X[\beta_1, \ldots \beta_j] + [\varepsilon_1, \ldots \varepsilon_j] \end{aligned} \qquad 23$$

can be rearranged to give a univariate model

$$\begin{aligned} vec(Y) &= (1 \otimes X)vec(\beta) + vec(\varepsilon) \\ \begin{bmatrix} y_1 \\ \vdots \\ y_j \end{bmatrix} &= \begin{bmatrix} X & & \\ & \ddots & \\ & & X \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_j \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_j \end{bmatrix} \end{aligned} \qquad 24$$

where $\otimes$ denotes the Kronecker Tensor product. Here $\mathrm{cov}(vec(\varepsilon)) = \Sigma \otimes V$, where Σ are the covariances among components and V encodes the temporal correlations. In MLMs Σ is unconstrained and requires full estimation (in terms of S_R). Therefore, any MLM and its univariate version are exactly equivalent, if we place constraints on the non-sphericity of the errors that ensure it has the form $\Sigma \otimes V$. This speaks to an important point; any multivariate analysis can proceed in a univariate setting with appropriate constraints on the non-sphericity. In fact MLMs are special cases that assume the covariances factorise into $\Sigma \otimes V$ and Σ is unconstrained. In neuroimaging there are obvious constraints on the form of Σ because this embodies the spatial covariances. Random field theory harnesses these constraints. MLMs do not and are therefore less sensitive.

V. SUMMARY

This chapter has described multivariate approaches to the analysis of functional imaging studies. These use standard multivariate techniques to describe or make statistical inferences about distributed activation effects and characterise important features of functional connectivity. The multivariate approach differs fundamentally from statistical parametric mapping, because the concept of a separate voxel or region of interest ceases to have meaning. In this sense inference is about the whole image volume not any component of it. This feature precludes statistical inferences about regional effects made without reference to changes elsewhere in the brain. This fundamental difference ensures that mass-univariate and multivariate approaches are likely to be regarded as distinct and complementary approaches to functional imaging data (see Kherif *et al* 2002).

In this chapter we have used correlations among brain measurements to identify systems that respond in a coherent fashion. This identification proceeds without reference to the mechanisms that may mediate distributed and integrated responses. In the next chapter we turn to models of effective connectivity that ground the nature of these interactions.

REFERENCES

B Biswal FZ Yetkin VM Haughton and JS Hyde. (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Mag. Res. Med.* **34**:537-541

E. Bleuler. Dementia Praecox or the group of schizophrenias: Translated into English 1987 in "*The clinical roots of the schizophrenia concept*." (J. Cutting and M. Shepherd, eds.), Cambridge University Press, UK, 1913.

C. Chatfield and A.J. Collins. (1980) Introduction *to multivariate analysis*. Chapman and Hall, London. pp189-210

R. Friedrich, A. Fuchs, H. Haken. (1991) Modelling of spatio-temporal EEG patterns in "*Mathematical approaches to brain functioning diagnostics*" (I. Dvorak and AV Holden, ed.), Manchester University Press, New York, .

A. Fuchs, J.A.S. Kelso, H. Haken. (1992) Phase transitions in the human brain: spatial mode dynamics. Int. J. of Bifurcation and Chaos **2**:917-939

KJ Friston, C.D. Frith, P.F. Liddle, R.S.J. Frackowiak. (1993a). Functional connectivity: the principal component analysis of large (PET) data sets.. *J. Cereb. Blood Flow Metab.* **13**:5-KJ Friston, P. Jezzard, R.S.J. Frackowiak, R. Turner. (1993b) Characterising focal and distributed physiological changes with MRI and PET. In *Functional MRI of the Brain*, Society of Magnetic Resonance in Medicine, Berkeley CA. pp207-216

Friston KJ Frith CD Fletcher P Liddle PF Frackowiak RSJ (1996a) Functional topography: multidimensional scaling and functional connectivity in the brain *Cerebral Cortex* **6**:156-164

Friston KJ Herold S Fletcher P Silbersweig D Cahill C Dolan RJ Liddle PF Frackowiak RSJ and Frith CD.(1996b) Abnormal fronto-temporal interactions in schizophrenia In*: Biology of Schizophrenia and Affective Disease* Ed Watson SJ ARNMD Series Vol. **73**;421-429

Friston KJ Poline J-B Holmes AP Frith CD Frackowiak RSJ (1996c) A multivariate analysis of PET activation studies. *Human Brain Mapping* **4**:140-151

Friston KJ (1998) Modes or models: a critique on independent component analysis for fMRI. *Trends in Cognitive Sciences* 2:373-374

Friston KJ, Phillips J, Chawla D & Bűchel C. (2000). Nonlinear PCA: Characterising interactions between modes of brain activity. *Phil Trans R Soc. Lond. B*; **355**:135-146

GL Gerstein, DH Perkel. (1969). Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science* **164**:828-830

P.S. Goldman-Rakic.(1988) Topography of cognition: Parallel distributed networks in primate association cortex. *Ann. Rev. Neurosci*. **11**:137-156

J.C. Gower. (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**:325-328

Kherif F, Poline JB, Flandin G, Benali H, Simon O, Dehaene S, Worsley KJ (2002) .Multivariate model specification for fMRI data. *NeuroImage*. **16**:1068-83.

J Karhunen and Joutsensalo J (1994) Representation and separation of signals using nonlinear PCA type learning. *Neural Networks* **7**:113-127

MA Kramer (1991) Nonlinear principal component analysis using auto-associative neural networks. *AIChE Journal*, **37**:233-243

G. Mayer-Kress, C. Barczys, W. Freeman. (1991). Attractor reconstruction from event-related multi-electrode EEG data in "*Mathematical approaches to brain functioning diagnostics"* (I. Dvorak and AV Holden ed.), Manchester University Press, New York.

A.R. McIntosh, F.L. Bookstein, J.V. Haxby, CL Grady. (1996) Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage* **00**:00-00

McKeown MJ, Makeig S, Brown GG, Jung TP, Kindermann SS, Bell AJ, Sejnowski TJ. (1998) Analysis of fMRI data by blind separation into independent spatial components. *Hum Brain Mapp*. **6**:160-88.

I. Newton. (1794) *Opticks*. Book 1, part 2, prop. 6 London: Smith and Walford, .

C.G. Phillips, S. Zeki , H.B. Barlow. Localisation of function in the cerebral cortex. Past, present and future. *Brain* **107**:327-361(1984).

JV Stone (2002) Independent component analysis: an introduction, *Trends in Cognitive Sciences*, **6**; 59-64

P. Talairach and J. Tournoux. (1988) *A Stereotactic coplanar atlas of the human brain*. Stuttgart: Thieme.

W.S. Torgerson. (1958). *Theory and methods of scaling* New York: Wiley

R.N Shepard. (1980) Multidimensional scaling, tree-fitting and clustering. *Science* **210**:390-398

Worsley KJ, Poline JB, Friston KJ, Evans AC (1997) .Characterizing the response of PET and fMRI data using multivariate linear models. *NeuroImage*. **6**;305-19

Legends for Figures

Figure 1

Schematic illustrating a simple spectral decomposition or singular-decomposition of a multivariate time-series. The original time series is shown in the upper panel with time running along the *x* axis. The first three eigenvariates and eigenvectors are shown in the middle panels together with the spectrum [hence spectral decomposition] of singular values. The eigenvalues are the square of the singular values $\lambda = SS^T$. The lower panel shows the data reconstructed using only three principal components. Because they capture most of the variance the reconstructed sequence is very similar to the original time-series.

Figure 2

Eigenimage analysis of the PET activation study of word generation Top: Positive and negative components of the first eigenimage (*i.e.* first column of *V*). The maximum intensity projection display format is standard and provides three views of the brain (from the back, from the right and from the top). Lower Left: Eigenvalues (singular values squared) of the functional connectivity matrix reflecting the relative amounts of variance accounted for by the 11 eigenimages associated with this data. Only two eigenvalues are greater than unity and to all intents and purposes the changes characterising this time-series can be considered two-dimensional. Lower right: The temporal eigenvariate reflecting the expression of this eigenimage over the 12 conditions (*i.e.* the first column of *U*).

Figure 3

Classical or metric scaling analysis of the functional topography of intrinsic word generation in normal subjects. Top: Anatomical regions categorised according to their colour. The designation was by reference to the atlas of Talairach and Tournoux (1988). Bottom: Regions plotted in a functional space following the scaling transformation. In this space the proximity relationships reflect the functional connectivity among regions. The colour of each voxel corresponds to the anatomical region it belongs to. The brightness reflects the local density of points corresponding to voxels in anatomical space. This density was estimated by binning the number of voxels in 0.02 'boxes' and smoothing with a Gaussian kernel of full width at half maximum of 3 boxes. Each colour was scaled to its maximum brightness.

Figure 4

Generalised eigenimage analysis of schizophrenic and control subjects. Top left and right: Positive and negative loadings of the first eigenimage that is maximally expressed in the control group and minimally expressed in the schizophrenic group. This analysis used PET activation studies of word generation with six scans per subject and six subjects per group. The activation study involved three word generation conditions (word shadowing, semantic categorisation and verbal fluency) each of which was presented twice. The grey scale is arbitrary and each image has been normalised to the image maximum. The display format is standard and represents a maximum intensity projection. This eigenimage is relatively less expressed in the schizophrenic data. This point is made by expressing the amount of functional connectivity attributable to the eigenimage in (each subject in) both groups, using the appropriate 2-norm (lower panel).

Figure 5

Upper panel: Schematic of the neural net architecture used to estimate causes and modes. Feed-forward connections from the input layer to the hidden layer provide an estimate of the causes using some recognition model (the **E**-Step). This estimate minimises prediction error under the constraints imposed by prior assumption about the causes. The modes or parameters are updated in an **M**-Step. The architecture is quite ubiquitous and when 'unwrapped' discloses the hidden layer as a 'bottleneck' (see insert). These 'bottleneck' architectures are characteristic of manifold learning algorithms like nonlinear PCA.

Lower panel (left): Condition-specific expression of the two first orders modes ensuing from the visual processing fMRI study. These data represent the degree to which the first principal component of epoch-related responses over the 32 photic stimulation/baseline pairs was expressed. These condition-specific responses are plotted in terms of the four conditions for the two modes. **Motion** - motion present. **Stat.** - stationary dots. **Colour** - isoluminant, chromatic contrast stimuli. **Isochr.** - isochromatic, luminance contrast stimuli.

Lower panels (right): The axial slices have been selected to include the maxima of the corresponding spatial modes. In this display format the modes have been thresholded at 1.64 of each mode's standard deviation over all voxels. The resulting excursion set has been superimposed onto a structural T1 weighted MRI image.

Figure 6

Top: The first canonical image displayed as maximum intensity projections of the positive and negative components. The display format is standard and provides three views of the brain from the front, the back and the right hand side. The grey scale is arbitrary and the space conforms to that described in the atlas of Talairach and Tournoux (1988). Bottom: The expression of the first canonical image (*i.e.* the canonical variate) averaged over conditions. The odd conditions correspond to word shadowing and the even conditions correspond to word generation. This canonical variate is clearly sensitive to the differences evoked by these two tasks.

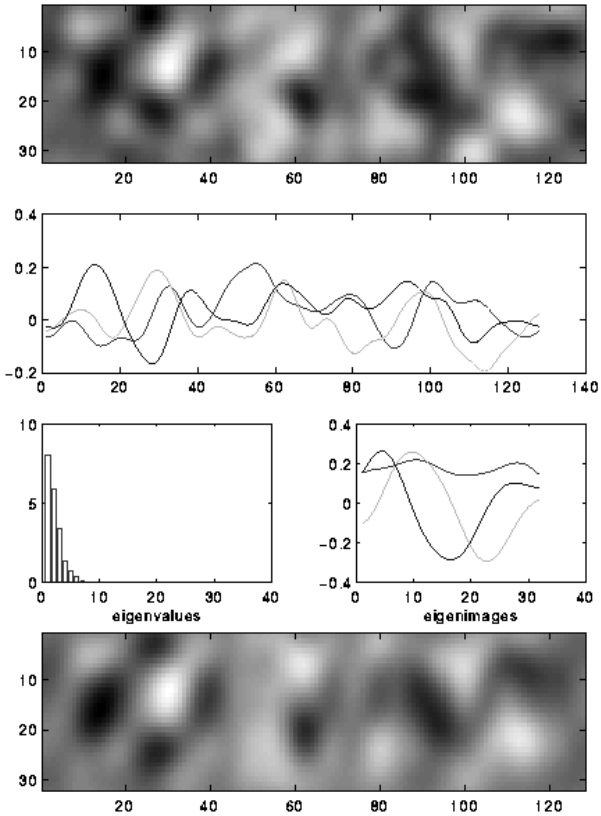
Eigenimages

Time-series (M)
128 scans of 40 "voxels"

Eigenvariates (U)

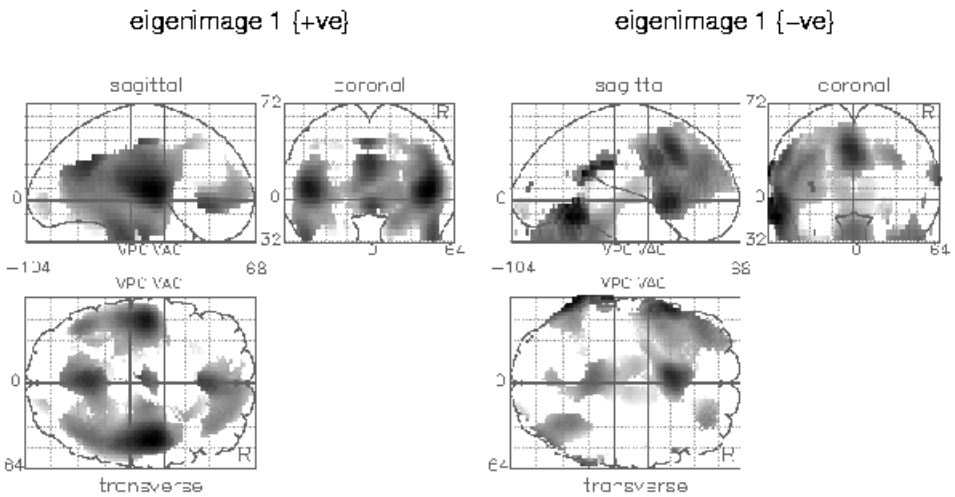
Singular values (S)
and spatial "modes" or
eigenimages (V)

'Reconstituted'
time-series
$\widetilde{M}$

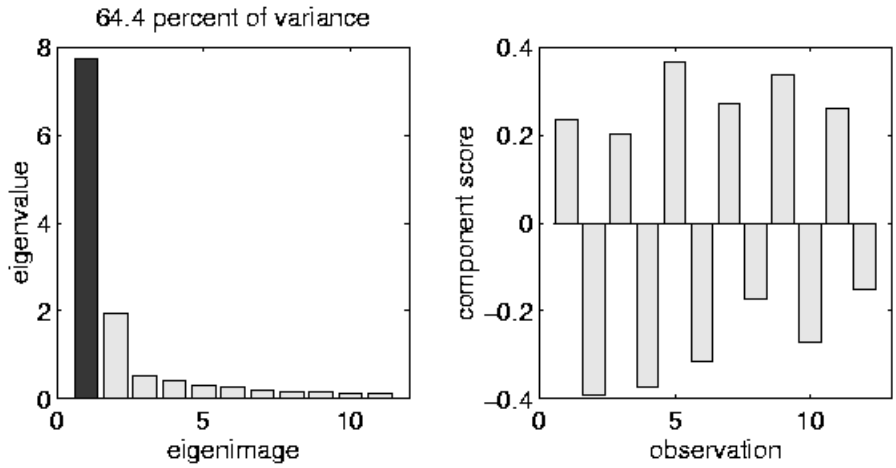


$$M = USV^T = s_1 U_1 V_1^T + s_2 U_2 V_2^T + \ldots$$

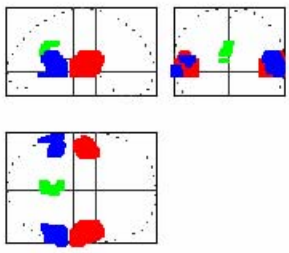
$$\widetilde{M} = s_1 U_1 V_1^T + s_2 U_2 V_2^T + s_3 U_3 V_3^T$$

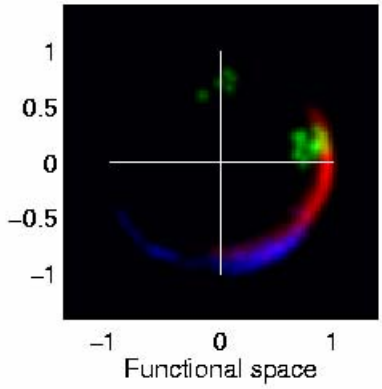


Eigenimage analysis:

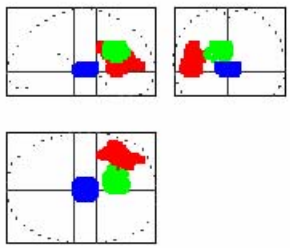


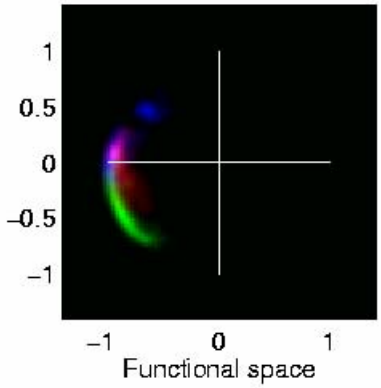
anatomical regions

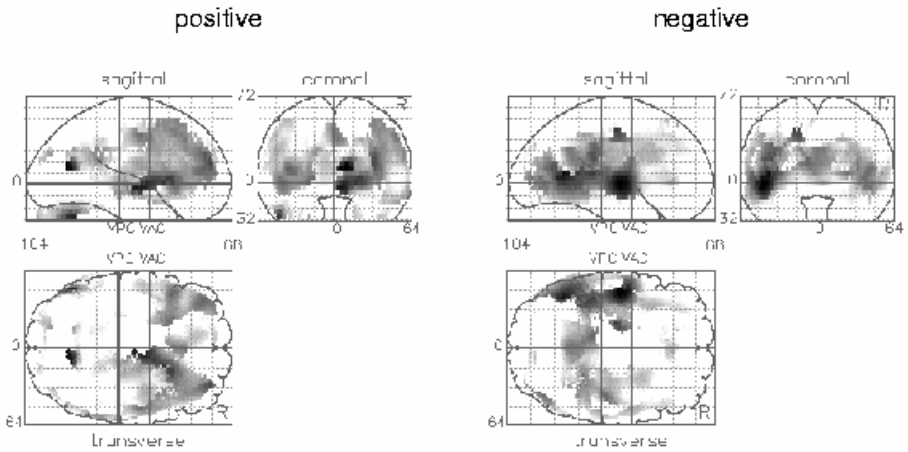


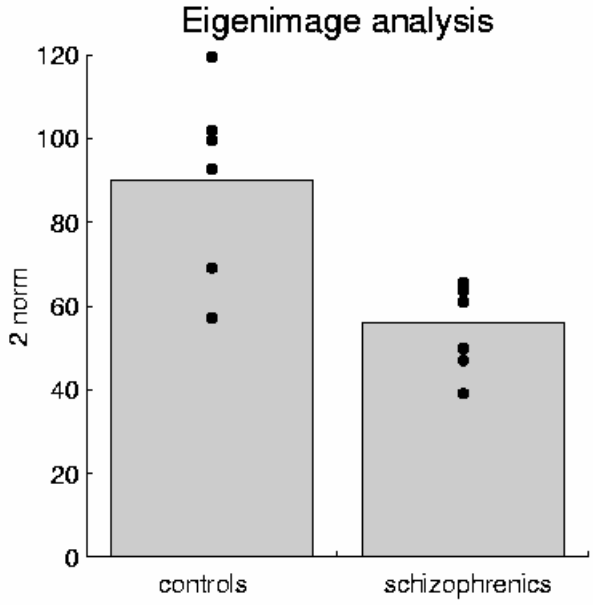


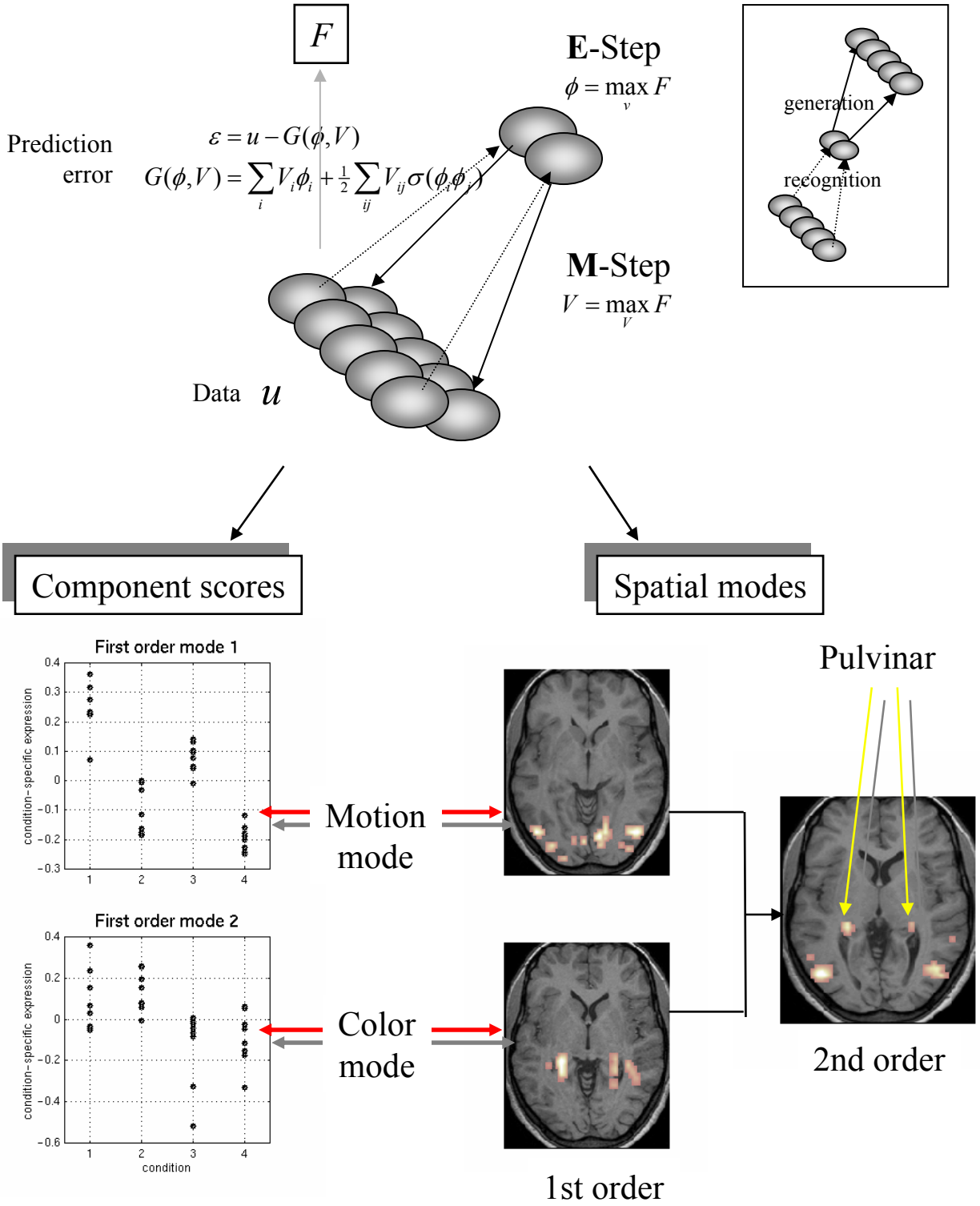
anatomical regions

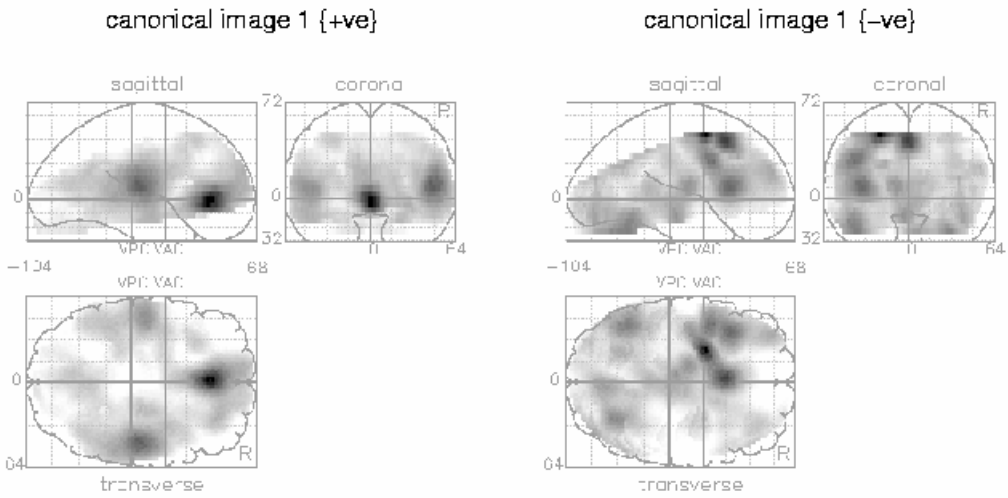


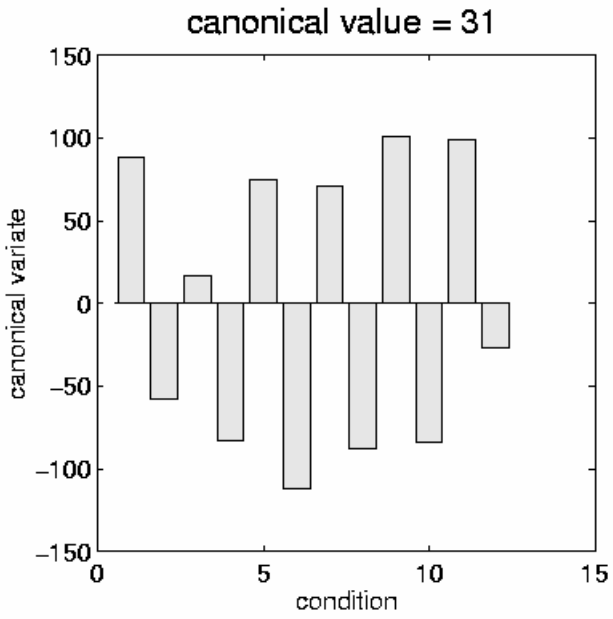



positive
negative
sagittal
coronal
transverse


Eigenimage analysis
2 norm
120
100
80
60
40
20
0
controls
schizophrenics


F
E-Step
$\phi = \max_v F$
Prediction error
$\varepsilon = u - G(\phi, V)$
$G(\phi, V) = \sum_i V_i \phi_i + \frac{1}{2} \sum_{ij} V_{ij} \sigma(\phi_i \phi_j)$
M-Step
$V = \max_V F$
Data u
generation
recognition
Component scores
Spatial modes
First order mode 1
condition-specific expression
First order mode 2
condition
Motion mode
Color mode
Pulvinar
2nd order
1st order


canonical image 1 {+ve}
sagittal
corona
transverse
canonical image 1 {−ve}
sagittal
coronal
transverse


canonical value = 31
canonical variate
condition
150
100
50
0
−50
−100
−150
0
5
10
15