

Eigenimages and multivariate analyses

Karl J Friston

Contents

- I. Introduction**
- II. Functional integration and connectivity**
 - II.A. Origins and definitions
- III. Eigenimages, multidimensional scaling and other devices**
 - III.A. Measuring a pattern of correlated activity
 - III.B. Eigenimages and spatial modes
 - III.C. Mapping function into anatomical space - eigenimage analysis
 - III.D. Mapping anatomy into functional space - multidimensional scaling
 - III.E. Functional connectivity between systems - partial least squares
 - III.F. Differences in functional connectivity - generalised eigenimages
 - III.F.1. The generalised eigenvalue solution
 - III.G. Summary
- IV. ManCova and canonical image analysis**
 - IV.A. Introduction
 - IV.B. Dimension reduction and eigenimages
 - IV.C. The general linear model revised
 - IV.D. Statistical inference
 - IV.E. Characterising the effect
 - IV.F. Relationship to eigenimage analysis
 - IV.G. An illustrative application
 - IV.H. Canonical Variates Analysis
 - IV.I. Summary

I. Introduction

This chapter is concerned with the characterization of imaging data from a multivariate perspective. This means that the observations at each voxel are considered conjointly with explicit reference to the interactions among brain regions. The concept of functional connectivity is introduced 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 variates analysis to create canonical images. In contradistinction to previous chapters this, and the next chapter are less concerned with functional segregation but more with functional integration. The integrated and distributed nature of neurophysiological responses to sensorimotor or cognitive challenge makes a multivariate perspective particularly appropriate and provides a complementary characterization of activation studies.

II. 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 [1]. 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.

II.A. Origins and definitions

In the analysis of neuroimaging time-series functional connectivity is defined as the *temporal correlations between spatially remote neurophysiological events* [2]. This definition provides a simple characterization of functional interactions. The alternative is effective connectivity (i.e. *the influence one neuronal system exerts over another*) [3]. These concepts originated in the analysis of separable spike trains obtained from multiunit electrode recordings [4, 5]. 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 [6]. 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).

III. Eigenimages, multidimensional scaling and other devices

In what follows we introduce a number of techniques (eigenimage analysis, multidimensional scaling, partial least squares and generalized 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 analysis are predicated on exactly the same eigenvector solution and from a mathematical perspective are essentially the same thing.

III.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 correlations or covariance (correlations are normalized covariances). 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 (\mathbf{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 $\mathbf{M}\mathbf{p} = \|\mathbf{M}\mathbf{p}\|_2$. \mathbf{M} is a (mean corrected) matrix of data with one row for each successive scan and one column for each voxel (T denotes transposition):

$$\|\mathbf{M}\mathbf{p}\|_2^2 = \mathbf{p}^T \mathbf{M}^T \mathbf{M} \mathbf{p} \quad 1$$

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: If e time-dependent changes occur predominantly in regions described by the pattern (\mathbf{p}) then the correlation between the pattern of activity and \mathbf{p} *over space* will vary substantially *over time*. The 2-norm measures this temporal variance in the spatial correlation. The pattern \mathbf{p} can be used to define the functional connectivity of interest. For example, if one were interested in the functional connectivity between left dorsolateral prefrontal cortex (DLPFC) and left superior temporal region one could test for this interaction using the 2-norm in Eq. (1) where \mathbf{p} had large values in the frontal and temporal regions. This approach has been used to demonstrate abnormal prefronto-temporal integration in schizophrenia [7]; an example we shall return to below.

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.

III.B Eigenimages and spatial modes

In this section the concept of eigenimages or spatial modes is introduced in terms of patterns of activity (\mathbf{p}) defined in the previous section. 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 (\mathbf{M}) into two sets of orthogonal vectors (patterns in space and patterns in time) \mathbf{V} and \mathbf{U} where:

$$[\mathbf{U} \ \mathbf{S} \ \mathbf{V}] = \text{SVD}\{\mathbf{M}\}$$

such that: $\mathbf{M} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T \quad 2$

\mathbf{U} and \mathbf{V} are unitary orthogonal matrices (the sum of squares of each column is unity and all the column are uncorrelated) and \mathbf{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 maximizes the largest 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 stepdown fashion, the greatest amounts of functional connectivity. Each eigenvector (column of \mathbf{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 characterize the spatiotemporal dynamics of neurophysiological time-series from several modalities; including multiunit electrode recordings [8], EEG [9], MEG [10], PET [2] and functional MRI [11].

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 \mathbf{U} in Equ(2) correspond to. These vectors are the time-dependent profiles associated with each eigenimage. They reflect the extent to which an eigenimage is expressed in each experimental condition or over time. These vectors play an important role in the functional attribution of distributed systems defined by eigenimages. This point and others will be illustrated in the next section.

III.C Mapping function into anatomical space - Eigenimage analysis

To illustrate the approach, we will use a standard word generation study. 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 letter presented (*intrinsic word generation*). To facilitate intersubject pooling, the data were realigned and spatially normalized 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 the omnibus F ratio to identify those significant at $p < 0.05$ (uncorrected). The adjusted time-series from each of these voxels formed a mean corrected data matrix \mathbf{M} with 12 rows (one for each condition) and one column for each voxel.

The images data matrix \mathbf{M} was subject to SVD as described in the previous section. The distribution of eigenvalues (Figure 1, 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 (the first column of \mathbf{V}) is shown in Figure 1 (top) along with the corresponding vector in time (the first column of \mathbf{U} - 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 \mathbf{U} 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.

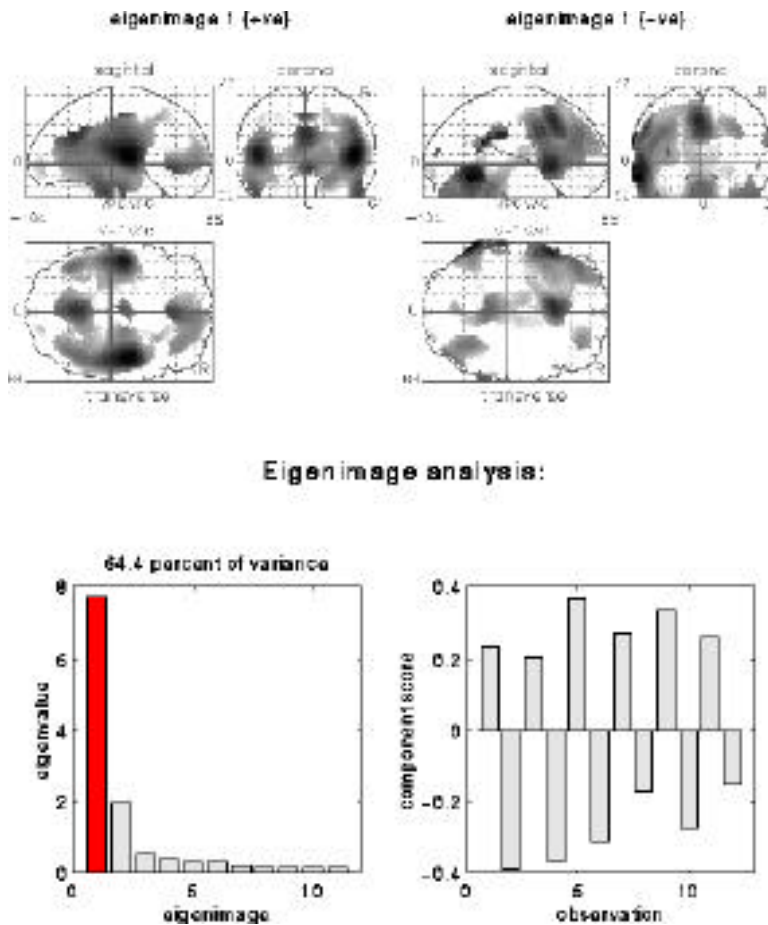


Figure 1

Eigenimage analysis 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 in the stereotactic space of Talairach and Tournoux (1988) (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 characterizing this time-series can be considered two-dimensional. Lower right: The temporal eigenvector reflecting the expression of this eigenimage over the 12 conditions (i.e. the first column of U).

The *post hoc* functional attribution of these eigenimages is usually based on their time-dependent profiles (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.

There 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 characterization 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.

III.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. 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[12].

Multidimensional scaling (MDS) is a descriptive method for representing the structure of a system. Based on pairwise measures of similarity or confusability [13, 14]. 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 modeled by points in space, so that perceived similarity is represented by spatial distances, goes back to the days of Isaac Newton[15]. The implementation of this idea is however relatively new [14].

Imagine K measures from n voxels plotted as n points in a K -dimensional space (K -space). If they have been normalized 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 maximize 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 $\mathbf{M}\cdot\mathbf{M}^T$. i.e. the column vectors of \mathbf{U} . From Equ(2):

$$\mathbf{M}\cdot\mathbf{M}^T = \mathbf{U}\cdot\mathbf{S}^2\cdot\mathbf{U}^T$$

Let \mathbf{Q} be the matrix of desired coordinates derived by simply projecting the original data (\mathbf{M}^T) onto axes defined by \mathbf{U} :

$$\mathbf{Q} = \mathbf{M}^T\cdot\mathbf{U} \quad 3$$

Voxels that have a correlation of unity will occupy the same point in MDS space. Voxels that have independent dynamics (correlation = 0) will be 2 apart. Voxels that are negatively but totally correlated (correlation = -1) will be maximally separated (by a distance of 2). Profound negative correlations denote a functional association that is modeled in MDS functional space as diametrically opposed locations on the hypersphere. In other words two regions with profound negative correlations will form two 'poles' in functional space.

Following normalization to unit sum of squares over each column \mathbf{M} (the adjusted data matrix from the word generation study above) the data were subjected to singular value decomposition according to Equ(2) and the coordinates \mathbf{Q} of the voxels in MDS functional space were computed as in Equ(3). Recall that only two eigenvalues exceed unity (Figure 1 right) suggesting a functional space that is essentially two dimensional. The locations of voxels in this two-dimensional subspace are shown in Figure 2 (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 2, right) included the mediodorsal thalamus (blue), the dorsolateral prefrontal cortex (DLPFC), Broca's area (red) and the anterior cingulate (green). Posterior regions (Figure 2, left) included the superior temporal regions (red), the posterior superior temporal regions (blue) and the posterior cingulate (green). The corresponding functional spaces (Figure 2, 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 localized 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 1) 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 π apart or equivalently subtend 90 degrees at the origin).

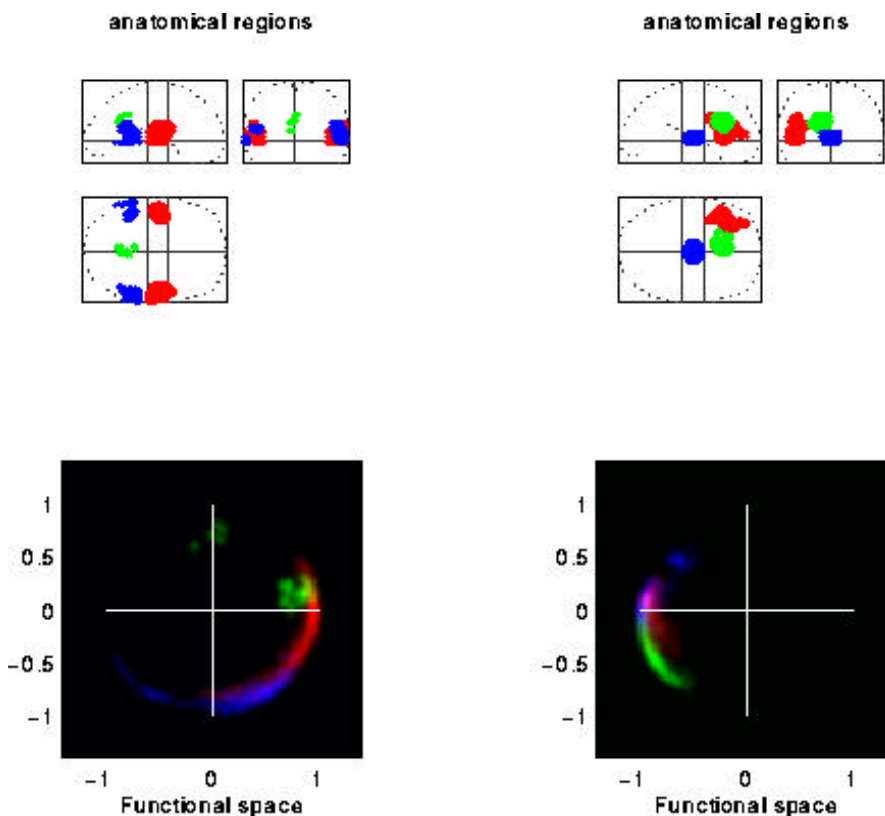


Figure 2

Classical or metric scaling analysis of the functional topography of intrinsic word generation in normal subjects. Top: Anatomical regions categorized 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 between 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.

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 [16]. 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.

III.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 \mathbf{p} is given by:

$$\mathbf{m}_p = \mathbf{M} \cdot \mathbf{p}$$

$$\text{therefore: } \rho_{pq} = \mathbf{m}_q^T \cdot \mathbf{m}_p = \mathbf{q}^T \cdot \mathbf{M}^T \cdot \mathbf{M} \cdot \mathbf{p} \quad 4$$

where ρ_{pq} is the functional connectivity between the systems described by vectors \mathbf{p} and \mathbf{q} . Consider next functional connectivity between two systems in separate parts of the brain, for example the right and left hemispheres. Here the data matrices (\mathbf{M}_p and \mathbf{M}_q) derive from different sets of voxels and Equ(4) becomes:

$$\rho_{pq} = \mathbf{m}_q^T \cdot \mathbf{m}_p = \mathbf{q}^T \cdot \mathbf{M}_q^T \cdot \mathbf{M}_p \cdot \mathbf{p} \quad 5$$

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 set of vectors \mathbf{p} and \mathbf{q} that maximize ρ_{pq} in Equ(5). SVD finds yet another powerful application in doing just this where:

$$[\mathbf{U} \ \mathbf{S} \ \mathbf{V}] = \text{SVD}\{\mathbf{M}_q^T \cdot \mathbf{M}_p\}$$

$$\text{such that: } \mathbf{M}_q^T \cdot \mathbf{M}_p = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T$$

$$\text{and } \mathbf{U}^T \cdot \mathbf{M}_q^T \cdot \mathbf{M}_p \cdot \mathbf{V} = \mathbf{S} \quad 6$$

The first columns of \mathbf{U} and \mathbf{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 \mathbf{S}). In other words SVD of the (generally asymmetric) covariance matrix, based on time-series from two anatomically separate parts of the brain, yields a series of paired vectors (paired columns of \mathbf{U} and \mathbf{V}) that, in a stepdown fashion, define pairs of brain systems that show the greatest functional connectivity. See Friston [7] for further details. This particular application of SVD is also known 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 [17]. 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.

III.F. Differences in functional connectivity - Generalised eigenimages

In this section we introduce an extension of eigenimage analysis using the solution to the generalized 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.

The notion that schizophrenia represents a disintegration or fractionation of the psyche is as old as its name, introduced by Bleuler [18] to convey a 'splitting' of mental faculties. Many of Bleuler's primary processes, such as 'loosening of associations' emphasize 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

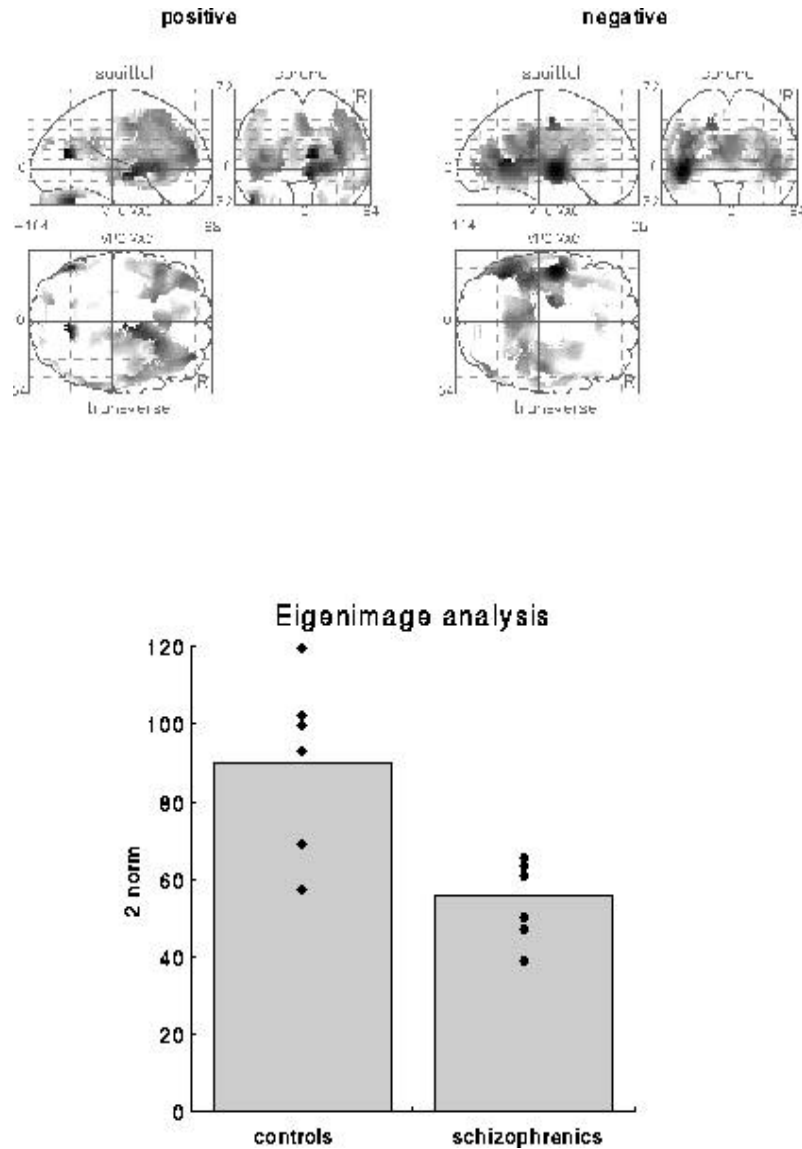


Figure 3

Generalized eigenimage analysis of the schizophrenic and control subjects. Top left and right: Positive and negative loadings of the first eigenimage that is maximally expressed in the normal group and minimally expressed in the schizophrenic group. This analysis used ^{15}O PET activation studies of word generations 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).

The idea is that although localized 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).

III.F.1 The generalized eigenvalue solution

Suppose that we want to find a pattern embodying the greatest amount of functional connectivity in normal subjects that, relatively speaking, was not evident in schizophrenia (eg. fronto-temporal covariance). To achieve this result we identify an eigenimage that reflects the most functional connectivity in normal subjects relative to a schizophrenic group (\mathbf{d}). This eigenimage is obtained by using a generalized eigenvector solution:

$$\begin{aligned} & \mathbf{C}_1^{-1} \cdot \mathbf{C}_2 \cdot \mathbf{d}_1 &= & \mathbf{d}. \\ \text{or} & \mathbf{C}_2 \cdot \mathbf{d}_1 &= & \mathbf{C}_1 \mathbf{d}. \end{aligned}$$

where \mathbf{C}_1 and \mathbf{C}_2 are the two functional connectivity matrices. The generalized eigenimage (\mathbf{d}) is essentially a single pattern that maximizes the ratio of the 2-norm measure [Equ(1)] when applied to \mathbf{C}_1 and \mathbf{C}_2 . 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 normal 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 normal 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 generalized eigenimage analysis are presented in Figure 3. 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 3, upper panels). The amount to which this pattern was expressed in each individual group is shown in the lower panel of Figure 5 using the appropriate 2-norm $\|\mathbf{d} \cdot \mathbf{C} \cdot \mathbf{d}\|$. It is seen that this eigenimage, whilst prevalent in normal subjects, is uniformly reduced in schizophrenic subjects.

III.G. Summary

In the preceding sections we have seen how eigenimages can be framed in relation to functional connectivity and the relationship between eigenimage analysis, multidimensional scaling, partial least squares and generalized eigenimage analysis. All these techniques are essentially descriptive, in that

they do not allow one to make any statistical inferences about the characterizations 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 the advantage of using this approach where the first covariance matrix reflected the effects we were interested in, and the second embodied covariances due to error. This application generates canonical images, and is considered in the following sections.

IV. ManCova and canonical image analysis

IV.A Introduction

In the following sections we review a general multivariate approach to the analysis of functional imaging studies. This analysis uses standard multivariate techniques to make statistical inferences about activation effects and to describe important features of these effects. Specifically we introduce multivariate analysis of covariance (ManCova) and canonical variates analysis (CVA) to characterize activation effects and address special issues that ensue. 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 analyzing 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 multivariate analysis that ensues can be summarised as follows: (i) Unlike eigenimage analysis it provides for statistical inferences (based on a p value) 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 variates analysis produces generalized 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.

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 variates analysis, combines many of the attractive 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 feature 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 an 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).

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. Then the brain's response is assessed to be significant using ManCova, the nature of this response remains to be characterised. We propose that 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 effects. 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 discounting.

IV.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 \mathbf{X} of eigenimages as a reduced set of components for each multivariate observation (scan). Where:

$$[\mathbf{U} \mathbf{S} \mathbf{V}] = \text{SVD}\{\mathbf{M}\}$$

$$\text{and} \quad \mathbf{X} = \mathbf{U} \cdot \mathbf{S} \quad 7$$

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

IV.C The general linear model revisited

Recall the general linear model from previous chapters:

$$\mathbf{X} = \mathbf{G}\mathbf{b} + \mathbf{e} \quad 8$$

where the errors \mathbf{e} are assumed to be independent and identically normally distributed. The design matrix \mathbf{G} 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. \mathbf{b} is the parameter matrix with one column vector of parameters for each mode. Each column of \mathbf{G} has an associated unknown parameter. Some of these parameters will be of interest, the remaining parameters will not. As before we will split \mathbf{G} (and \mathbf{b}) into two partitions $\mathbf{G} = [\mathbf{H} \mathbf{D}]$ and similarly $\mathbf{b} = [\mathbf{a}^T \mathbf{g}^T]^T$ with estimators $\mathbf{b} = [\mathbf{a}^T \mathbf{g}^T]^T$. Here effects of interest are denoted by \mathbf{H} and confounding effects of no interest by \mathbf{D} . Equ(8) can be expanded:

$$\mathbf{X} = \mathbf{H} \cdot \mathbf{a} + \mathbf{D} \cdot \mathbf{g} + \mathbf{e} \quad 9$$

where \mathbf{H} 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 \mathbf{H} might contain covariates of interest that could explain the observed variance in \mathbf{X} (e.g. dose of apomorphine or 'time on target'). \mathbf{D} 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).

IV.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 \mathbf{a} is zero). The null hypothesis can be tested in the following way. The sum of squares and products matrix (SSPM) due to error $\mathbf{R}(\mathbf{e})$ is obtained from the difference between actual and estimated values of \mathbf{X} :

$$\mathbf{R}(\mathbf{e}) = \mathbf{R}(\mathbf{X} - \mathbf{G} \cdot \mathbf{b}) = (\mathbf{X} - \mathbf{G} \cdot \mathbf{b})^T (\mathbf{X} - \mathbf{G} \cdot \mathbf{b}) \quad 10$$

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

$$\mathbf{T} = (\mathbf{H}\mathbf{a})^T(\mathbf{H}\mathbf{a}) \quad 11$$

The error sum of squares and products under the null hypothesis $\mathbf{R}(0)$ i.e. after discounting the effects of interest (\mathbf{H}) are given by:

$$\mathbf{R}(0) = (\mathbf{X} - \mathbf{D}\mathbf{g})^T(\mathbf{X} - \mathbf{D}\mathbf{g}) \quad 12$$

Clearly if \mathbf{D} does not exist this simply reduces to the sum of squares and products of the response variable $(\mathbf{X}^T\mathbf{X})$. The significance can now be tested with:

$$= \frac{|\mathbf{R}(0)|}{|\mathbf{R}(0)|} \quad 13$$

where λ is Wilk's statistic (known as Wilk's Lambda). A special case of this test is Hotelling's T^2 test and applies when \mathbf{H} simply compares one condition with another [19]. Under the null hypothesis, after transformation, λ has a χ^2 distribution with degrees of freedom $J.h$. The transformation is given by:

$$-(r - ((J - h + 1)/2)).\log(\lambda) \sim \chi^2(J.h)$$

where r are the degrees of freedom associated with error terms, equal to the number of scans (I) minus the number of effects modelled = $I - \text{rank}(\mathbf{G})$. J is the number of eigenimages in the J -variate response variable \mathbf{X} and h are the degrees of freedom associated with effects of interest = $\text{rank}(\mathbf{H})$.

IV.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 characterize these effects in terms of their spatial topography. This characterization uses canonical variates analysis or CVA. The objective is to find a linear combination (compound or contrast) of the components of \mathbf{X} , 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:

$$(\mathbf{c}_1^T\mathbf{H}\mathbf{c}_1) / (\mathbf{c}_1^T\mathbf{R}\mathbf{c}_1)$$

is maximised [19]. Let $\mathbf{Z}_1 = \mathbf{X}\mathbf{c}_1$ where \mathbf{Z}_1 is the first canonical variate and \mathbf{c}_1 is a canonical image (defined in the space of the spatial modes) that maximises this ratio. \mathbf{c}_2 is the second canonical image that maximises the ratio subject to the constraints of $\text{Cov}\{\mathbf{c}_1, \mathbf{c}_2\} = 0$ (and so on). The matrix of canonical images $\mathbf{c} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_j]$ is given by solution of the generalised eigenvalue problem:

$$\mathbf{T}\mathbf{c} = \mathbf{R}\mathbf{c} \quad 14$$

where \mathbf{T} is a diagonal matrix of eigenvalues. Voxel-space canonical images \mathbf{C} are obtained by rotating the canonical image in the columns of \mathbf{c} back into voxel-space with the original eigenimages \mathbf{V} :

$$\mathbf{C} = \mathbf{V}\mathbf{c} \quad 15$$

The columns of \mathbf{C} now contain the voxel values of the canonical images. The k th column of \mathbf{C} (the k th canonical image) has an associated canonical value equal to the k th leading diagonal element of \mathbf{T} 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 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 [19] for further details.

IV.F. Relationship to Eigenimage Analysis

When applied to adjusted data eigenimages [2] correspond to the eigenvectors of \mathbf{T} . These have an interesting relationship to the canonical images: On rearranging Equ(14):

$$\mathbf{R}^{-1} \cdot \mathbf{T} \cdot \mathbf{c} = \mathbf{c}.$$

we note that the canonical images are eigenvectors of $\mathbf{R}^{-1} \cdot \mathbf{T}$. In other words an eigenimage analysis of an activation study returns the eigenvectors that express the most variance due to the effects of interest - $\text{eig}(\mathbf{T})$. A canonical image, on the other hand, expresses the greatest amount of variance due to the effects of interest *relative to error* - $\text{eig}(\mathbf{R}^{-1} \cdot \mathbf{T})$. In this sense a CVA can be considered an eigenimage analysis that is 'informed' by the estimates of error effects.

IV.G. 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 6 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. 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 \mathbf{H} has 12 columns representing the 12 different conditions. We designated subject effects, time and global activity as uninteresting confounds \mathbf{D} . The corrected data were reduced to 60 eigenvectors as described above. The first 14 eigenvectors had (normalized) eigenvalues greater than unity and were used in the subsequent analysis. The resulting matrix \mathbf{X} , 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%.

IV.H. Canonical Variates Analysis

The first canonical image and its expression in each condition is shown in Figure 4. The upper panels show this system to include anterior cingulate and Broca's area, with more moderate expression in the left posterior inferotemporal 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 4. This variate is simply $\mathbf{a} \cdot \mathbf{c}_1$. It is pleasing to note that the first canonical variate corresponds to the difference between word shadowing and verbal fluency.

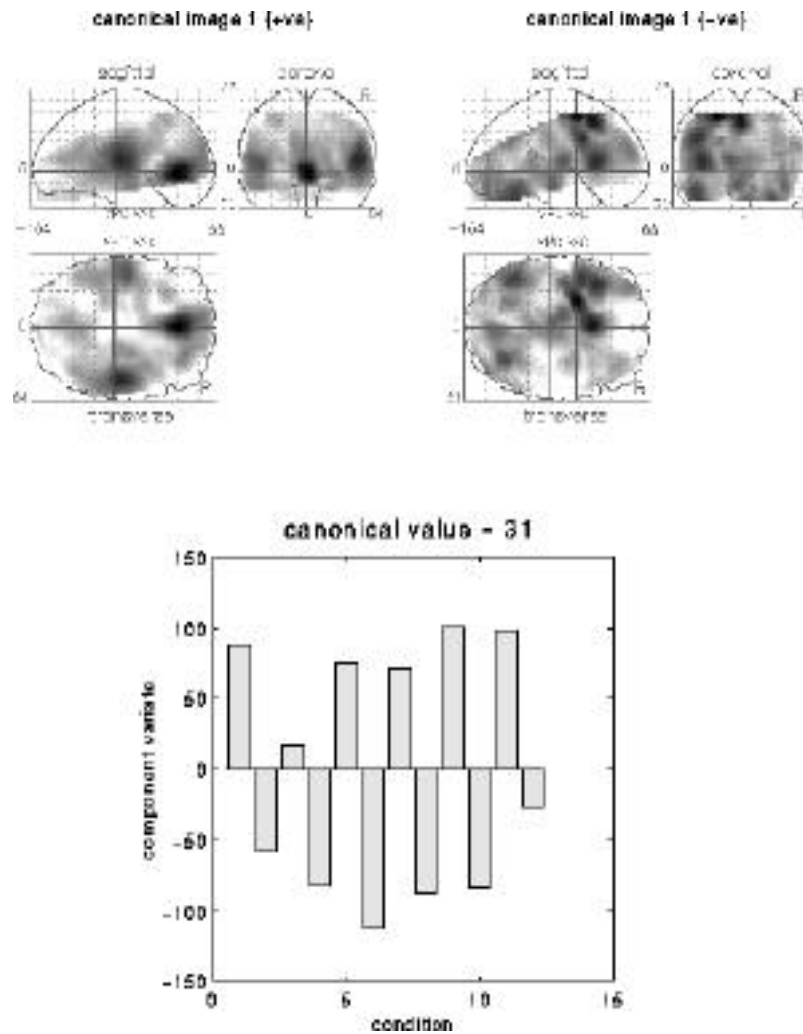


Figure 4

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.

Recall that the eigenimage in Figure 1 reflects the main pattern of correlations evoked by the mean condition effects and should be compared with the first canonical image in Figure 4. 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 much more generic 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 characterization 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 (in the J larger spatial modes) into effects of interest and a residual variation about these effects (error). This partitioning is

determined by experimental design, a hypothesis, and the inferences that are sought. An eigenimage does not embody any concept of error and is not constrained by any hypothesis.

IV.I. Summary

These sections have described a multivariate approach to the analysis of functional imaging studies. This analysis uses standard multivariate techniques to make statistical inferences about activation effects and to describe the important features of these effects. More specifically the proposed analysis uses multivariate analysis of covariance ManCova with Wilk's Lambda to test for specific effects of interest (e.g. differences among activation conditions) and canonical variates analysis (CVA) to characterize these distributed responses. The generality of this approach is assured by the generality of the linear model used. The design and inferences sought are embodied in the design matrix and can, in principle, accommodate most parametric statistical analyses. This multivariate approach differs fundamentally from statistical parametric mapping, because the concept of a separate voxel or region of interest ceases to have meaning. One scan represents one observation (not 10^5 voxels). In this sense the statistical 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 SPM and multivariate approaches are likely to be regarded as distinct and complementary approaches to functional imaging data.

At the time of writing the extension of the ManCova/CVA approach described above to serially correlated fMRI time-series remains problematic. Recent work by Keith Worsley and Jean-Baptiste Poline (personal communication) has focussed on the sum of F ratios in an $SPM\{F\}$ as an omnibus test and a characterization, of multivariate effects, based on an extensions of orthogonalized partial least squares that accommodates the serial or temporal correlations extant in fMRI data.

There are many potential applications of the analysis presented above. One particularly interesting application concerns the ability to test various models in a comprehensive and direct fashion. Hitherto there has been no 'omnibus' test for a particular neurophysiological response or model of this response that did not rely on some assumptions about the multivariate structure of the data (e.g. Gaussian Fields). Wilk's statistic could provide this test. For example the controversy over the appropriate model for removing confounding effects of global activity on regional effects has been dogged by the lack of any compelling comparative assessment of different models. Wilk's statistic could, in principle, be used to resolve this issue by explicitly testing hierarchies of models (a succession of extra effects modelled in the design matrix).

An attractive neuroscience application of the multivariate approach considered here pertains to the significance of interaction terms in a design matrix. Cognitive subtraction is based on the assumption that extra components of a task can be inserted without affecting pre-existing components. In order to verify the assumptions behind cognitive subtraction one needs to demonstrate that these interactions can be ignored when modelling the brain's response. This can be effected simply and rigorously using Wilk's statistic to show that interaction terms in the design matrix are not significant (here one would treat the interaction terms as effects of interest and the remaining effects of no interest). Of course if the interactions were significant this leads to a richer understanding of functional anatomy and provides a basis for more sophisticated experimental designs. We take up this theme again in a subsequent chapter.

References

1. C.G. Phillips, S. Zeki, H.B. Barlow. Localization of function in the cerebral cortex. Past, present and future. *Brain* 107:327-361(1984).
2. K.J. Friston, C.D. Frith, P.F. Liddle, R.S.J. Frackowiak. Functional connectivity: the principal component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.* 13:5-14 (1993a).
3. K.J. Friston, C.D. Frith, R.S.J. Frackowiak. Time-dependent changes in effective connectivity measured with PET. *Human Brain Mapping* 1:69-80 (1993b).
4. G.L. Gerstein, D.H. Perkel. Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science* 164:828-830 (1969).
5. A. Aertsen and H. Preissl. Dynamics of activity and connectivity in physiological neuronal networks in "Non Linear Dynamics and Neuronal Networks" (H.G. Schuster publishers Inc., ed.) pp. 281-302. New York, USA, 1991.

6. G.L. Gerstein, P. Bedenbaugh, A.M.H.J. Aertsen. Neuronal assemblies IEEE. Trans. on Biomed. Engineering 36:4-14 (1989).
7. K.J. Friston, S. Herold, P. Fletcher, D. Silbersweig, C. Cahill, R.J. Dolan, P.F. Liddle, R.S.J. Frackowiak, C.D. Frith. Abnormal fronto-temporal interactions in schizophrenia in "Biology of Schizophrenia and Affective Disease" (S.J. Watson, ed.) ARNMD Series Vol. 73, 1994.
8. G. Mayer-Kress, C. Barczys, W. Freeman. Attractor reconstruction from event-related multi-electrode EEG data in "Mathematical approaches to brain functioning diagnostics" (I. Dvorak and A.V. Holden ed.), Manchester University Press, New York, 1991.
9. R. Friedrich, A. Fuchs, H. Haken. Modelling of spatio-temporal EEG patterns in "Mathematical approaches to brain functioning diagnostics" (I. Dvorak and A.V. Holden, ed.), Manchester University Press, New York, 1991.
10. A. Fuchs, J.A.S. Kelso, H. Haken. Phase transitions in the human brain: spatial mode dynamics. Int. J. of Bifurcation and Chaos 2:917-939 (1992)
11. K.J. Friston, P. Jezzard, R.S.J. Frackowiak, R. Turner. Characterizing focal and distributed physiological changes with MRI and PET. In Functional MRI of the Brain, Society of Magnetic Resonance in Medicine, Berkely CA . 207-216 (1993)
12. J.C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325-328 (1966)
13. W.S. Torgerson. Theory and methods of scaling New York: Wiley (1958).
14. R.N Shepard. Multidimensional scaling, tree-fitting and clustering. Science 210:390-398 (1980)
15. I. Newton. Opticks. Book 1, part 2, prop. 6 London: Smith and Walford, 1794.
16. P.S. Goldman-Rakic. Topography of cognition: Parallel distributed networks in primate association cortex. Ann. Rev. Neurosci. 11:137-156 (1988)
17. A.R. McIntosh, F.L. Bookstein, J.V. Haxby, C.L. Grady. Spatial pattern analysis of functional brain images using partial least squares. NeuroImage 00:00-00 (1996)
18. 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.
19. C. Chatfield and A.J. Collins. Introduction to multivariate analysis. Chapman and Hall, London. P189-210 (1980)
20. P. Talairach and J. Tournoux. A Stereotactic coplanar atlas of the human brain. Stuttgart:Thieme, 1988.