Basic concepts and overview

Karl J. Friston

Contents

I. Introduction

This chapter provides an overview and background for subsequent chapters that describe, in detail, some of the different approaches to characterising and analyzing brain activation studies. This chapter is meant to underscore the following point, although we may take for granted that the results of functional neuroimaging studies can facilitate our understanding of how the brain works, the converse is also true. Our understanding of the brain's functional organization constrains and informs the way in which functional neuroimaging data are analyzed and, consequently, affects the ensuing results. In what

follows we review some fundamental ideas about brain organization and how these ideas relate to different approaches to data analysis. The aim of this chapter is to introduce various techniques that have been developed, or adapted, for analyzing functional imaging data and to place them in some broad context. The theoretical foundations and procedural nature of the techniques are then described in subsequent chapters.

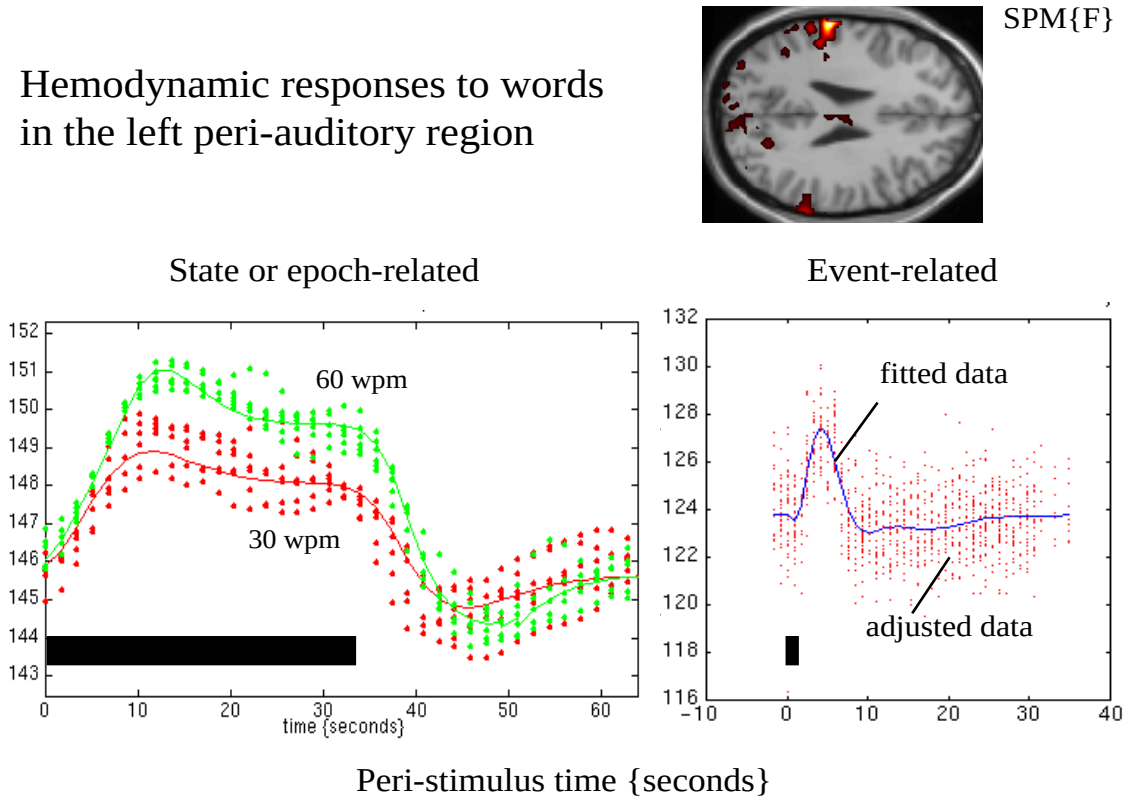

Figure 1

From the early 80s positron emission tomography (PET) dominated the field of functional neuroanatomy. In the past five years functional magnetic resonance imaging (fMRI) has developed into an alternative and powerful technique. PET measures blood flow on a spatio-temporal scale of about 6mm and 30 seconds; fMRI is sensitive to the oxygenation of blood and has a spatio-temporal scale of about 1-3mm and one or more seconds. The lower limits on the *effective resolution* of fMRI are physiological and imposed by the spatio-temporal organization of evoked hemodynamic responses (2-5mm and 5-8 seconds). Local increases in neural activity cause both a relative deoxygenation of blood and an increase in perfusion, that quickly reverses the deoxygenation, leading to an increase in oxygenation that endures for several seconds. To a first approximation one can think of the observed hemodynamic response as a smoothed version of the underlying neural activity (see lower right panel above). PET measures changes in blood flow or perfusion directly in terms of the amount of radiolabelled water that accumulates locally but this takes several tens of second to a minute or so.

Although there is a great overlap between PET and fMRI, there are some fundamental constraints on a experimental design imposed by the different techniques: Consider the distinction between *state-related responses* and *event-related responses*. By virtue of the relatively long half-life of the tracers used, PET can only measure responses summed over fairly long periods of time. Consequently PET can only be used to measure differences between brain states. In contrast, fMRI can be used in two ways. Epochs of repeated

stimuli or continuous task performance can be presented and the ensuing signal interpreted as a brain-state dependent measure (left panel). Alternatively, event-related responses to a single stimuli can be measured by analogy to evoked potentials in electrophysiology (right panel).
The data above were acquired from a single subject using echo planar imaging (EPI) fMRI at a rate of one volume image every 1.7s. The subject listened to words in epochs of 34 seconds at a variety of different frequencies. The fitted peri-auditory responses (lines) and adjusted data (dots) are shown for two rates (30 and 60 words per minute) on the left. The solid bar denotes the presentation of words. By removing confounds and specifying the appropriate design matrix (see figure 2) one can show that fMRI is exquisitely sensitive to single events. The data shown on the right were acquired from the same subject whilst simply listening to single words presented every 34 seconds. Event-related responses were modelled using a small set of temporal basis functions of the peri-stimulus time. The SPM{F} reflecting the significance of these event-related responses (see figure 2) shown has been thresholded at $p = 0.001$ (uncorrected) and displayed on a T_1 weighted structural MRI.

Our interpretation of functional imaging studies relies upon an ensemble of models. Some are very explicit (e.g. the general linear model used in statistical analysis) and some are implicit (e.g. functional specialization when inferring the nature of activations or pure insertion in cognitive subtraction). These models embody the assumptions that are required to make sense of the data. Models relating to what is being measured (Figure 1) and how these measurements are analyzed (Figure 2) are biophysical, anatomical, mathematical, statistical and, usually, explicit. In this chapter we consider the conceptual and mathematical models that are commonly employed in analysing functional imaging data.

The first half of this chapter focusses on principles of brain organization and their relation to the analytic techniques available. We will make a distinction between *functional integration* and *functional segregation* at a theoretical level and a corresponding distinction between *multivariate* and *univariate* analyses at a procedural level (univariate means pertaining to one variable). The multivariate analyses considered here include eigenimage analysis, multidimensional scaling, multivariate analysis of covariance (ManCova) and canonical variates analysis (CVA). The univariate approaches are collectively brought under the heading of statistical parametric mapping. The second half of this chapter considers stages of data analysis that are required to implement the above. These stages include realignment, spatial normalization, smoothing, statistical analysis and, when appropriate, statistical inference. Many of the techniques described here start off with the general linear model and a short introduction to the general linear model is included at the end of this introductory chapter.

II Principles of functional organization and implications for imaging

II.A Specialization or integration?

The brain appears to adhere to two fundamental principles of functional organization; *functional integration* and *functional segregation*, with the integration within and among functionally segregated areas mediated by functional or effective connectivity. The characterization of functional segregation, integration and connectivity is an important endeavour in many areas of neuroscience, not least in functional imaging.

The distinction between functional segregation and integration relates to the distinction between *localisationism* and *[dis]connectionism* that dominated thinking about cortical function in the nineteenth century. Since the early anatomic theories of Gall, the identification of a particular brain region with a specific function has become a central theme in neuroscience. However functional localization *per se* was not easy to demonstrate in a compelling way: For example, a meeting that

took place on the morning of August 4th 1881 addressed the difficulties of attributing function to a cortical area, given the dependence of cerebral activity on underlying connections (1). This meeting was entitled 'Localisation of function in the cortex cerebri'. Goltz (2) although accepting the results of electrical stimulation in dog and monkey cortex, considered the excitation method inconclusive, in that movements elicited might have originated in related pathways, or current could have spread to distant centres. In short the excitation method could not be used to infer functional localisation because localisationism discounted interactions, or functional integration among different brain areas. It was proposed that lesion studies could supplement excitation experiments; ironically, it was observations on patients with brain lesions (3) some years later that led to the concept of *disconnection syndromes* and the refutation of localisationism as a complete or sufficient explanation of cortical organization. Functional localisation and functional segregation are used here to refer to different concepts. Functional localisation implies that a function can be localised in a cortical area, whereas segregation suggests that a cortical area is specialised for some aspects of perceptual or motor processing, and that this specialisation is anatomically segregated in the cortex. The cortical infrastructure supporting a single function may then involve many specialised areas whose union is mediated by the functional integration among them. In this view functional segregation is only meaningful in the context of functional integration and *vice versa*..

The cerebral excitation paradigm of the mid nineteenth century is, in a way, still employed today. With modern methods excitation is usually elicited by asking subjects to perform different tasks and the neuronal responses are measured with functional imaging. Despite advances over the past century, the central issue remains unresolved. Are the physiological changes elicited by sensorimotor or cognitive challenges best characterized by the activation of functionally segregated areas, or are they better described in terms of the behaviour of anatomically distributed, but functionally integrated systems. The position that we have adopted is to simply acknowledge that both perspectives are valid and that the description of, and inferences about, brain activations can proceed at both levels.

What then are the implications for analyzing and interpreting brain activation experiments? There are two basic approaches that are predicated on the distinction between functional integration and segregation presented above. The first set of approaches is based on correlated physiological dynamics in different parts of the brain and deals with distributed brain systems. These approaches are inherently multivariate in nature. The second depends on detecting focal differences using a functional segregation or specialisation view of the brain.

The first set of approaches to analysis includes *eigenimage analysis* and *multidimensional scaling*. These are useful ways of characterizing distributed neurophysiological changes in a parsimonious and sometimes revealing fashion. They are closely related to each other (mathematically speaking) but have different and interesting interpretations in relation to functional imaging data. Neither eigenimage analysis nor multidimensional scaling can be used to make statistical inferences about what is observed (i.e. estimate the significance of results obtained in terms of a p value). The second approach, considered under the heading of functional integration, is multivariate analysis of [co-]variance (*ManCova*) with canonical variates analysis (CVA). Like eigenimage analysis CVA characterizes brain responses in terms of anatomically distributed profiles or *spatial modes*. Unlike any of the techniques mentioned so far ManCova provides for statistical inference about the responses, where this inference is about the entire brain (i.e. it has no regional specificity). To address the significance of regionally specific effects one turns, finally, to *statistical parametric mapping.* Statistical parametric mapping is an approach that is predicated on functional segregation or specialisation and is used to characterize physiology in terms of regionally specific responses. It does this characterization by treating each voxel separately (i.e. it is a univariate approach) and by performing voxel-wise statistical analyses in parallel creates an image of a statistic or ‘significance’. The next two sections review the concepts of functional integration and segregation, as they relate to functional imaging, and describe more fully the multivariate and univariate techniques mentioned above.

II.B Functional integration and connectivity

Functional integration is mediated by the interactions between functionally segregated areas. One way of characterizing these interactions is in terms of functional or effective connectivity. In the analysis of neuroimaging time-series functional connectivity has been defined as the *temporal correlations between spatially remote neurophysiological events* (4). The alternative is to refer explicitly to effective connectivity (i.e.*the influence one neuronal system exerts over another*) (5). These concepts were developed in the analysis of separable spike trains obtained from multiunit electrode recordings (6, 7, 8, 9). Functional connectivity is simply a statement about the observed correlations; it does not provide any direct insight into 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 (7). To examine the integration within a distributed system, defined by functional connectivity, one turns to *effective connectivity.* Effective connectivity is closer to the intuitive notion of a connection. In electrophysiology there is a close relationship between effective connectivity and synaptic efficacy: "It is useful to describe the effective connectivity with a connectivity matrix of effective synaptic weights" (7). It has also been proposed that "the [electrophysiological] notion of effective connectivity should be understood as the experiment and time dependent, simplest possible circuit diagram that would replicate the observed timing relationships between the recorded neurons" (9).

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 the 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 the underlying connectivity. The confounding effect of stimulus-evoked transients is less problematic in neuroimaging because the 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 and, where it is, effective influences may be indirect (e.g polysynaptic relays through multiple areas).

II.C Functional connectivity at different spatio-temporal scales

Clearly there is an enormous difference between correlations in regional cerebral blood flow (rCBF) measured over seconds and the fast phasic dynamics that mediate neural interactions on a time-scale of milliseconds. The purpose of this section is to suggest that there is a relationship between functional connectivity at a large-scale physiological level and at a neural level. Functional connectivity between two areas (in terms of functional imaging) implies that their pool activity goes up and down together. Is it reasonable to suppose that two regions with high pool activity will share a significant number of neurons whose dynamic interactions occur within a time frame of milliseconds? We suggest it is. There are two lines of evidence in support of a relationship between fast dynamic interactions and slower variations in pool activity: (i) Aertsen & Preissl (9) have investigated the behaviour of artificial networks, analytically and using simulations. They concluded that *short term effective connectivity* varies strongly with, or is modulated by, *pool activity.* Pool activity was defined as the product of the number of neurons and their mean firing rate. The mechanism is simple; the efficacy of sub-threshold EPSPs (excitatory post-synaptic potentials) for establishing dynamic interactions is a function of postsynaptic depolarization, which in turn depends on tonic background activity. (ii) The second line of evidence is experimental and demonstrates that the presence of fast interactions is associated with intermediate or long term correlations between distant neurons. Nelson et al (10) have characterized effective connections between neurons, or small groups of neurons, in BA 17 and BA 18 of cat

extrastriate cortex. By cross-correlating activity they demonstrated that the most likely temporal relationship between spikes was a synchronous one. Furthermore, the cross-correlograms segregated into three non-overlapping groups with modal widths of 3ms, 30ms and 400ms. The short term correlation structures (3 and 30ms) were almost always associated with the intermediate (400ms) correlations. These observations suggest an interaction between short term (<100 ms) and intermediate (100 - 1000ms) correlations. In summary, the idea is that co-activated regions will have increased (correlated) perfusion and neuronal pool activity. Higher background discharge rates augment post-synaptic depolarization and susceptibility to fast dynamic interactions at a neural level, both within and between the regions co-activated. This susceptibility may facilitate functional and effective connectivity at a neural level.

II.D Eigenimage analysis and functional connectivity

A powerful use of functional connectivity is in the characterization of distributed brain systems. Usually pair-wise functional connectivity between any two points in the brain is not as interesting as patterns of correlations that obtain on considering all correlations together. Important patterns can be identified as follows. The functional connectivity (covariance) matrix, obtained from a time-series of images, is subject to Principal Component Analysis (PCA) or Singular Value Decomposition (SVD). The resulting eigenimages (principal components or spatial modes) each correspond to a spatial pattern or distributed system, comprising regions that are jointly implicated by virtue of their functional interactions. This analysis of neuroimaging time-series is predicated on established techniques in electrophysiology (both EEG and multiunit recordings). For example, in an analysis of multichannel EEG data the underlying spatial modes that best characterize the observed spatiotemporal dynamics are identified with a Karhunen Loeve expansion (11). Commonly, this expansion is in terms of eigenvectors of a covariance matrix associated with a time-series. The spatial modes are then identical to the principal components identified with a PCA. SVD is a related technique (12) that has been used with the Karhunen Loeve expansion, to identify spatial modes in multiunit electrode recordings (13). Having identified important spatial modes their functional role can be inferred on the basis of their expression over time or over different brain states. Note that this sort of analysis is entirely data-led and does not refer to any model or hypothesis.

II.E Multidimensional scaling and functional connectivity

In neuroimaging, functional mapping usually implies *mapping function into an anatomical space*, for example, using statistical parametric mapping to identify activation foci, or characterization of distributed changes with spatial modes or eigenimages as described in the previous section. This section introduces a complementary approach namely, *mapping anatomy into a functional space*. The basic idea here is to construct a space where functional connectivity is used to define the proximity of two anatomical areas. In a subsequent chapter a simple variant of multidimensional scaling (principal coordinates analysis) is described that uses functional connectivity as a metric (or measure of distance). This scaling transformation maps *anatomy* into a functional space. The topography, or proximity relationships, in this space embody the functional connectivity among brain regions. The higher the functional connectivity the closer the regions. Like eigenimage analysis the technique represents a descriptive characterization of anatomically distributed changes in the brain that reveals the structure of cortico-cortical interactions in terms of functional connectivity and, in fact, uses the same mathematics to do so. This form of multidimensional scaling can be regarded as equivalent to eigenimage analysis but looked at from a different point of view.

II.F ManCova-CVA and eigenimage analysis

Neither eigenimage analysis nor multidimensional scaling are statistical techniques in the sense that they do not have an explicit statistical model that allows for statistical inference. ManCova, on the other hand, uses the general linear model (an important linear equation that is described more fully below and in subsequent chapters) to make inferences about distributed activation effects. The nature of these effects can be characterised using an eigenimage analysis or a CVA ManCova is a multivariate technique. In other words it considers as one observation all the voxels in a single scan. The importance of this multivariate approach is that the effects due to activations, confounding effects and error effects are assessed statistically, both in terms of effects at each voxel, also and interactions among voxels. The price paid for adopting a multivariate approach is that one cannot make statistical inferences about regionally specific changes. This is because the inference pertains to all the components (voxels) of the multivariate variable (not a particular voxel or set of voxels). In functional imaging there is a caveat to the application of ManCova. Generally speaking ManCova requires the number of observations (i.e. scans) to be greater than the number of components of the multivariate observation (i.e. voxels). Clearly this is not the case for most functional imaging studies. The number of components is therefore reduced by using eigenimages as opposed to voxels. This works because the number of eigenimages is always less than or equal to the number of scans.

In general multivariate analyses are implemented in two steps. First the significance of the hypothesised effect is assessed in an omnibus sense in terms of a p-value and then the nature of this response is characterised. Canonical Variate Analysis (CVA) is one way of doing this. The canonical images obtained with CVA are similar to eigenimages but use both the activation and the error effects. Intuitively these approaches can be understood as finding the eigenimages that capture the effects of interest and simultaneously avoid patterns due to error.

This concludes a brief review of multivariate techniques that take explicit account of interactions among brain regions or voxels. In general the effects elicited by experimental design are described in terms of spatial modes (i.e. eigenimages or canonical images) that are single, anatomically distributed patterns. We now turn to functional segregation and univariate approaches that describe neurophysiological effects in terms of a number of regionally specific changes, whose significance can be assessed independently of changes elsewhere in the brain.

III.Functional segregation and statistical parametric mapping

III.A Functional segregation and specialisation

The functional role played by any component (e.g. A neuron) of a connected system (e.g. the brain) is largely defined by its connections. Certain patterns of cortical projections are so common that they could amount to rules of cortical connectivity. 'These rules revolve around one, apparently, overriding strategy that the cerebral cortex uses - that of functional segregation' (14). Functional segregation demands that cells with common functional properties be grouped together. This architectural constraint in turn necessitates both convergence and divergence of cortical connections. Extrinsic connections between cortical regions are not continuous but occur in patches or clusters. This patchiness has, in some instances, a clear relationship to functional segregation. For example; V2 has a distinctive cytochrome oxidase architecture, consisting of thick stripes, thin stripes and inter-stripes. When recordings are made in V2, directionally selective (but not wavelength or colour selective) cells are found exclusively in the thick stripes. Retrograde (i.e. backwards) labelling of cells in V5 is limited to these thick stripes. All the available physiological evidence suggests that V5 is a functionally homogeneous area that is specialized for visual motion. Evidence of this nature supports the notion that patchy connectivity is the anatomical infrastructure that mediates functional segregation and specialisation (see Zeki (14) for a full discussion). If it is the case that neurons in a given cortical area share a common responsiveness (by virtue of their extrinsic connectivity) to some sensorimotor or cognitive attribute, then this functional segregation is also an anatomical one. Challenging a subject with the appropriate sensorimotor attribute or cognitive process should lead to activity changes in, and

only in, the area of interest. This is the model upon which a search for regionally specific effects is based

III.B Statistical parametric mapping and regionally specific effects

Functional mapping studies are usually analyzed with some form of statistical parametric mapping. Statistical parametric mapping refers to the construction of spatially extended statistical processes to test hypotheses about regionally specific effects. Statistical parametric maps (SPMs) are image processes with voxel values that are, under the null hypothesis, distributed according to a known probability density function (usually Gaussian).

The success of statistical parametric mapping is largely due to the simplicity of the idea. Namely, one analyzes each and every voxel using any standard (univariate) statistical test. The resulting statistical parameters are assembled into an image - the SPM. SPMs are interpreted as a spatially extended statistical process by referring to the probabilistic behaviour of stationary Gaussian fields (15). Stationary fields model both the univariate probabilistic characteristics of an SPM and any stationary spatial covariance structure (stationary means not a function of position). 'Unlikely' excursions of the SPM are interpreted as regionally specific effects, attributable to a sensorimotor or cognitive process that has been manipulated experimentally. This characterization of physiological responses appeals to functional segregation as the underlying model of brain function. One could regard all applications of statistical parametric mapping as testing some variant of the functional segregation hypothesis.

The ideas behind statistical parametric mapping are, of course, not new. Statistical parametric mapping represents the convergence of two earlier ideas; change distribution analysis and significance probability mapping. Change distribution analysis was a voxel-based assessment of neurophysiological changes developed by the St. Louis group for PET activation studies (16). The essential difference between change score maps and SPMs is that the former simply reflect the size of the activation whereas SPMs are images of the statistical reliability or significance of the effect. Significance probability mapping was developed in the analysis of multichannel EEG data and involves the construction of interpolated pseudomaps of a statistical parameter. Unlike pseudomaps in electrophysiology SPMs are well behaved in the sense that they are stationary (due to the fact that neuroimaging samples uniformly and that the point spread function is stationary). This well behaved aspect of SPMs (under the null hypothesis) meant that theoretical advances could be made, to a point where this analytical area is now a rapidly growing and exciting part of applied spatial statistics. This development has occurred in the context of the theory of Gaussian Fields (17, 18, 19, 20). In particular; the theory of level-crossings (17) and differential topology (18).

III.C Statistical parametric mapping - implementation and applications

Statistical parametric mapping is a framework, subsuming the general linear model and the theory of Gaussian fields, that allows for a diverse interrogation of functional imaging data. The same ideas and implementation can accommodate approaches that range from simple unpaired t tests, to multifactorial AnCova. The experimental design and the model used to test for specific neurophysiological responses are embodied in a mathematical structure called the *design matrix*. The nature and role of the design matrix is described more fully below and in subsequent chapters. The approach adopted in our version of statistical parametric mapping partitions the design matrix according to whether the effect is interesting (e.g. an activation) or not (e.g. a nuisance effect like global activity). These effects can reflect factor levels (using indicator-type variables to denote the presence of a particular cognitive component) or can be continuous functions (a covariate, like reaction time).

The contribution of each effect (i.e. each column of the design matrix modelling the various response components) to the observed physiological responses is estimated using the general linear model and standard least squares. These estimated contributions are known as parameter estimates and can be as

simple as the mean activity associated with a particular condition or as complicated as an interaction term in a multifactorial experiment. Regionally specific effects are framed in terms of differences among these parameter estimates (e.g. an activation effect) and are specified using linear compounds or contrasts. The significance of each contrast is assessed with a statistic whose distribution has Student's t distribution under the null hypothesis. For each contrast or differences in parameter estimates, a t statistic is computed (in parallel) for each and every voxel to form a SPM{t}. For convenience the SPM{t} is transformed to a Gaussian Field or SPM{Z}. Statistical inferences are then made about local excursions of the SPM{Z} above a specified threshold, using distributional approximations from the theory of Gaussian fields. These distributions pertain to simple characterizations of the 'blobs' of activation; namely, their maximal value and their spatial extent (or both). The resulting p-values can be considered corrected for the volume of brain tested. By specifying different contrasts one can test for a variety of effects. These effects generally fall into one of three broad categories:

III.C.1 Cognitive and sensorimotor subtraction and conjunctions

The tenet of this approach is that the difference between two tasks can be formulated as a separable cognitive or sensorimotor component and that the regionally specific differences in brain activity identify the corresponding functionally specialized area. Early applications of subtraction range from the functional anatomy of word processing (21) to functional specialization in extrastriate cortex (22). The latter studies involved presenting visual stimuli with and without some specific sensory attribute (e.g. colour, motion etc). The areas highlighted by subtraction were identified with homologous areas in monkeys that showed selective electrophysiological responses to equivalent visual stimuli.

Cognitive conjunctions can be thought of as an extension of the subtraction technique, in the sense that they combine a series of subtractions. In subtraction we test a hypothesis pertaining to the activation in one task relative to another. In cognitive conjunctions several hypotheses are tested, asking whether all the activations, in a series of task pairs, are jointly significant. Consider the problem of identifying regionally specific activations due to a particular cognitive component (e.g. object recognition). If one can identify a series of task pairs whose differences have only that component in common, then the region which activates in all the corresponding subtractions can be uniquely associated with the component in question.

III.C.2 Parametric designs

The premise is that regional physiology will vary systematically with the degree of cognitive or sensorimotor processing. Examples of this approach include the experiments of Grafton *et al* (23) who demonstrated significant correlations between rCBF and the performance of a visually guided motor tracking task. On the sensory side Price *et al* (24) demonstrated a remarkable linear relationship between perfusion in periauditory regions and frequency of aural word presentation. Significantly this correlation was not observed in Wernicke's area, where perfusion was appeared to correlate, not with the discriminative attributes of the stimulus, but with the presence or absence of semantic content.

III.C.3 Factorial designs

At its simplest an interaction represents a change in a change. Interactions are associated with factorial designs where two or more factors are combined in the same experiment. The effect of one factor, on the effect of the other, is assessed by the interaction term. Factorial designs have a wide range of applications. An early application in neuroimaging examined physiological adaptation and plasticity (25) during motor performance by assessing time by condition interactions (26). Psychopharmacological activation studies are examples of factorial design (27). In these studies subjects perform a series of baseline-activation pairs before and after being given a drug. The interaction term reflects the modulatory drug effect on the task-dependent activation (28). We will

expand on the role of factorial designs in the context of cognitive science and additive factors logic in a later chapter.

In summary the applications of statistical parametric mapping have been introduced using a framework of activation studies that distinguishes between subtractive (categorical), parametric (dimensional) and factorial (interaction) designs. Subtractive designs are well established and powerful devices in functional mapping but are predicated on possibly untenable assumptions about the relationship between brain dynamics and the functional processes that ensue. Parametric approaches avoid many of the shortcomings of 'cognitive subtraction' by testing for non-linear but systematic relationships between neurophysiology and sensorimotor, psychophysical, pharmacologic or cognitive parameters. The fundamental difference between subtractive and parametric approaches lies in treating a cognitive process not as a categorical invariant but as a dimension or attribute that can be expressed to a greater or lesser extent. It is anticipated that parametric designs will find an increasing role in psychological and psychophysical activation experiments. Finally, factorial experiments provide a rich way of assessing the affect of one manipulation on the effects of another and incidently allow one to test some of the assumptions implicit in cognitive subtraction.

Having established some of the theoretical precedents, and placed the approaches considered here in a general context, we will now review a series of data transformations that are required to implement data analysis at a practical or operational level.

IV. The operational components of functional imaging data analysis

The following represents a summary of the stages required for the analysis of PET and fMRI activation studies. With regard to fMRI we are concerned principally with the analysis of fast techniques (e.g. echo-planar imaging [EPI]) that provide volumetric data (by multislice acquisition). Slower fMRI techniques, with an interscan interval that is greater than the hemodynamic response time (about six to eight seconds), can be treated in the same way as PET data because temporal autocorrelations due to the hemodynamic response function or introduced by temporal smoothing, do not have to be considered. The stages of analysis include (i) *realignment*, (ii) *spatial normalisation*, (iii) *spatial smoothing*, (iv) *voxel-wise statistical analysis* using the general linear model (for fast fMRI we have to use an extended model that accommodates temporal smoothing and autocorrelations over time) and (v) *statistical inference* based on the spatial extent and maxima of thresholded activation foci that ensue (using the theory of Gaussian fields). See Figure 2 for further details.

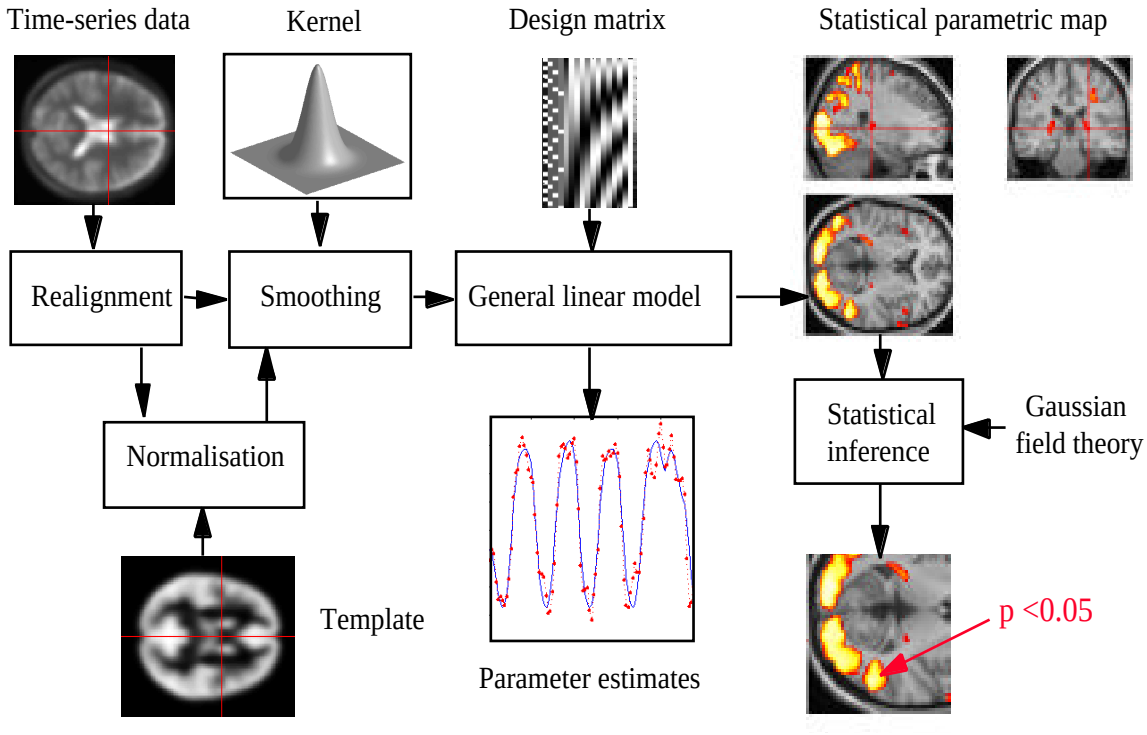

Figure 2

Overview

This schematic depicts the transformations that start with the imaging time-series and end with a statistical parametric map (SPM). SPMs that can be thought of as "X-rays" of the significance of an effect. Voxel-based analyses require the data to be in the same anatomical space: This is effected by realigning the data [in fMRI it is usually necessary to remove signal components, that are correlated with displacements, that remain after realignment]. After realignment the images are subject to non-linear warping so that they match a template that already conforms to a standard space. After smoothing, the general linear model is employed to perform the appropriate univariate test at each and every voxel. The test statistics that ensue (usually t or F statistics) constitute the SPM. The final stage is to make statistical inferences on the basis of the SPM and characterize the responses observed using the fitted responses or parameter estimates. If one knew where to look beforehand then this inference can be based on the value of the statistic without correction. If however one did not predict an anatomical site *a priori*, then a correction for the multiple dependent comparisons performed has to be made. These corrections are usually made using distributional approximations from the theory of Gaussian fields.

The General Linear Model

The general linear model is an equation which expresses the observed response variable in terms of a linear combination of explanatory variables plus a well behaved error term. The general linear model is variously known as 'analysis of covariance' or 'multiple regression analysis' and subsumes simpler variants, like the 't test' for a difference in means. The matrix that contains the explanatory variables (e.g. designed effects or confounds) is called the design matrix. Each column of the design matrix corresponds to some effect one has built into the experiment or that may confound the results. The example above relates to an fMRI study of visual stimulation under four conditions. The effects on the response variable are modelled in terms of functions of the presence of these conditions [smoothed with some appropriate hemodynamic response

function] and constitute the first four columns of the design matrix. There then follows a series of terms that are designed to remove or model low frequency variations in signal due to artifacts such as aliased biorhythms. The final column is whole brain activity. The relative contribution of each of these columns is assessed using standard least squares and inferences about these contributions are made using t statistics or F ratios, depending upon whether one is looking at a particular linear combination of these contributions, or all of them together. In fMRI the data are sometimes smoothed in time and one has to include the resulting serial correlations in the general linear model.

Statistical Inference and the theory of Gaussian fields
Correcting the p values for the volume analyzed is complicated by the fact that the data are not independent, by virtue of smoothness in the original data. This is accounted for by treating the SPMs as realizations of Gaussian random fields. The inferences that obtain are based on the probability of obtaining *c* or more clusters of volume *k* or more voxels above a threshold of *u.* in the volume analyzed. This formulation has a number special cases that give corrected p values pertaining to the number of clusters, the size of clusters or the heights of voxels within each cluster.

IV.A Realignment

Movement-related variance components in PET and fMRI time-series represent one of the most serious confounds of analysis. For PET this stage involves estimating movement relative to the first scan (using a least squares analysis that obtains after linearising the problem with a first order Taylor series), and realigning the scans *post hoc* using these estimates. Removing movement-related effects from fMRI time-series is not as simple as it may seem at first glance. This is because movement in earlier scans can affect the signal in subsequent scans (due to differential spin-excitation histories). The procedure we have adopted is to estimate head movements as for PET data. These estimates are then used to (i) realign the images and (ii) perform a mathematical adjustment to remove movement-related components that persist after simple realignment. The adjustment procedure is based on a moving average-auto-regression model of spin-excitation history effects. The latter components can be prominent and are specific to the way the fMRI signal is acquired.

IV.B Spatial normalisation

Clearly, to implement voxel-based analysis of imaging data, data from different subjects must derive from homologous parts of the brain. Spatial transformations are therefore applied that move and 'warp' the images such that they all conform (approximately) to some idealized or standard brain. This normalisation facilitates intersubject averaging and the reporting of results in a conventional way even if intersubject averaging is not used. The transformation of an image into a standard anatomical space [usually that described in the atlas of Talairach & Tournoux (29) as proposed by Fox *et al* (30) corresponds to spatial normalisation. For PET the normalizing transformations can be computed on the basis of the PET data themselves, or on the basis of co-registered high resolution anatomical MRI scans. If the T_2*-weighted fMRI data are of sufficiently good quality it is possible to normalise these directly.

IV.C Spatial Smoothing

Smoothing (convolving the data with a smoothing kernel) has several important objectives. Firstly, it generally increases signal relative to noise. The neurophysiological effects of interest are produced by hemodynamic changes that are expressed over spatial scales of several millimetres, whereas noise usually has higher spatial frequencies. For PET the spatial frequency structure of noise is determined by the reconstruction process used to create the images. In fMRI the noise can (to a first approximation) be regarded as independent for each voxel and has therefore very high spatial frequency

components. Secondly, convolving with a Gaussian (or other) kernel conditions the data in the sense that the data conform more closely to a Gaussian field model. This is important if one wants to use the theory of Gaussian fields to make statistical inferences about (i.e. assign p-values to) the ensuing regionally specific effects. The requirement that the data be a good lattice representation of a Gaussian field includes (i) that the autocorrelation function be twice differentiable and (ii) that the spatial correlations be stationary. Both these requirements are assured (approximately) after smoothing. The final reason for smoothing is specific to intersubject averaging. It ensures that hemodynamic changes from subject to subject are assessed on a spatial scale at which homologies in functional anatomy are typically expressed. For example, at a scale of a hundred microns to a millimetre it is highly unlikely that (even with perfect spatial normalization) the functional anatomy and organization of the middle frontal gyri in two subjects will show meaningful homologies. However, at an anatomical scale of 8mm they do as evidenced by the success of multi-subject PET activation studies. In other words a mapping of function onto anatomy may only be meaningful at a resolution that is not confounded by microscopic and mesoscopic organizational details that are unique to a given individual.

IV.D Statistical Analysis

This stage corresponds to modelling the data in order to partition observed neurophysiological responses into components of interest, confounds or components of no interest and an error term. This partitioning is effected using the general linear model to estimate the components in terms of parameters associated with the design matrix (see below).

The analysis of regionally specific effects uses the general linear model to assess the contribution or differences among parameter estimates (specified by a contrast) in a univariate sense, by referring to the error variance. This assessment is in terms of a F or t value for each and every voxel (i.e. a SPM{F} or SPM{t}). The SPM{t} is transformed to the unit normal distribution to give a Gaussian field or SPM{Z}. If one wishes to make inferences about several effects at the same time (e.g. using polynomial expansions or basis functions of some parameter of interest, then the SPM{F} is usually employed. For fast fMRI data there is a caveat. In order to maximize signal components one can smooth in time. The autocovariances in the error terms, that result from this smoothing, require analysis by an extended general linear model that takes account of such autocorrelations or serial correlations.

IV.E Statistical inference

Statistical inference can proceed at a number of levels. If one has employed a ManCova-CVA approach then a single p value obtains (based on Wilk's Lambda or a related statistic) that compares the variance-covariance accounted for by effects of interest relative to error. This single p-value tells one nothing about regionally specific effects. Statistical inference about specific regional changes requires statistical parametric mapping. With an anatomically constrained hypothesis about effects in a particular brain region *a priori* the Z value in that region in the SPM{Z} can be used to test the hypothesis. With an anatomically 'open' hypothesis was anatomically 'open' (i.e. a null hypothesis that there is no effect anywhere in the brain) a correction for multiple non-independent comparisons is required. It is at this point that the theory of Gaussian fields provides a way of computing a corrected p-value.

The SPM{Z} of the previous section is subject to standard procedures developed for statistical parametric mapping using the theory of Gaussian fields. These procedures give p-values that pertain to different levels of inference; these levels of inference can be in terms of (i) the number of activated regions (i.e. number of clusters above some height and volume thresholds), (ii) the number of activated

voxels (i.e. volume) comprising a particular region and (iii) the p-value for each voxel within that region. These p-values are corrected for the multiple non-independent comparisons implicit in an analysis. These results [the SPM{Z} and associated p-values] are the endpoint of analysis.

V. The general linear model

Many of the above techniques depend, at some level, on the general linear model. This section introduces the general linear model and terms that will be adopted in subsequent chapters. The general linear model is simply an equation that relates what one observes, to what one expected to see, by expressing the observations (*response variable*) as a linear combination of expected components (or *explanatory variables*) and some residual error. The general linear model comes in a number of guises, for example multiple linear regression, analysis of covariance or a simple t test. The general linear model for a response variable x_{ij} such as regional CBF at voxel j = 1,...,J is:

$$x_{ij} = g_{i1}ß_{1j} + g_{i2}ß_{2j} + \ldots\ldots g_{iK}ß_{Kj} + e_{ij} \qquad 1$$

where i = 1,....,I indexes the observation (e.g. scan). The general linear model assumes the errors (e_{ij}) are independent and identically distributed normally [$N(0, \sigma^2_j)$]. For example, in activation studies this means assuming an equal error variance (σ^2_j) across conditions and subjects (but not from one voxel or brain structure to the next). $ß_{kj}$ are K unknown parameters for each voxel j that represent the relative contribution of each of the explanatory variables.

The coefficients g_{ik} are explanatory variables relating to the conditions under which the observation (scan) *i* was made. These coefficients can be of two sorts: (i) a *covariate* (e.g. global CBF, time, plasma prolactin level, ... etc) in which case Equ(1) is a familiar multivariate regression model or (ii) indicator-type or dummy variables, taking integer values to indicate the level of a factor (e.g condition, subject, drug ... etc) under which the response variable (e.g. perfusion) is measured. Mathematically speaking there is no distinction between these two sort of variables but we make a distinction here for didactic reasons. Equ(1) can be written in matrix form as a multivariate general linear model;

$$\mathbf{X} = \mathbf{Gß} + \mathbf{e} \qquad 2$$

$$\mathbf{G}^T\mathbf{G}\,\mathbf{b} = \mathbf{G}^T\mathbf{X}$$

if **G** is of full rank then $\mathbf{G}^T\mathbf{G}$ is invertible and the least squares estimates are given uniquely by

$$\mathbf{b} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{X}$$

where $E\{\mathbf{b}_j\} = ß_j$ and $\mathrm{Var}\{\mathbf{b}_j\} = \sigma^2_j (\mathbf{G}^T\mathbf{G})^{-1}$ 3

If the errors are normally distributed then the least squares estimates are also the maximum likelihood estimates and are themselves normally distributed (32). $E\{\mathbf{b}_j\}$ is simply the expectation or average of the parameter estimates. $\mathrm{Var}\{\mathbf{b}_j\}$ is the variance-covariance matrix for the parameter estimates corresponding to the *j*th voxel. These equations can be used to implement a vast range of statistical analyses. The issue is therefore not so much the mathematics but the formulation of a design matrix (**G**) appropriate to the study design and inferences that are sought.

The design matrix can contain both covariates and indicator variables reflecting the experimental design. Each column of **G** has an associated unknown parameter in the vectors $ß_j$. Some of these parameters will be of interest [e.g. the effect of particular sensorimotor or cognitive condition or the regression coefficient of rCBF (the response variable) on reaction time (covariate)]. The remaining parameters will be of no interest and pertain to confounding effects (e.g the effect of being a particular subject or the regression slope of voxel activity on global activity). Confounding here is used to

denote an uninteresting effect that could confound the estimation of interesting effects (e.g. the confounding effect of global changes on regional activations).

The general linear model, and associated parameter estimation, have a central role in nearly everything that follows and forms the starting point for most of the statistical procedures considered in this book.

VI. Summary

This chapter has reviewed the basic distinction between approaches inspired by functional integration (eigenimage analysis, multidimensional scaling and ManCova-CVA) and those predicated on functional segregation (statistical parametric mapping). Functional integration is mediated by anatomical, functional and effective connections. In neuroimaging these concepts form the basis for characterizing patterns of correlations using eigenimages, singular images or canonical images (i.e. spatial modes). These spatial modes describe distributed systems. Functional segregation and specialisation speaks to regionally specific aspects of functional organization and requires a different approach to characterizing brain responses. Statistical parametric mapping provides for many ways in which to make inferences about functional segregation. Three important classes of experimental design, used in this endeavour, are subtractive, parametric and factorial. Each of these is a particular instance of the general linear model and can be framed in terms of an SPM{Z}. Statistical inferences about the resulting SPM{Z} are made using distributional approximations from the theory of Gaussian fields.

The two basic approaches to data analysis can be contrasted in a number of ways. SPMs rely on functional segregation as a conceptual model, Eigenimage analyses and related approaches (e.g. CVA) address the integration of distributed changes in terms of systems. The characterization of changes with a SPM is in terms of (multiple) activation foci that are identified following thresholding. The equivalent characterization with eigenimage analyses is in terms of spatially extended anatomical profiles (spatial modes). SPM is a statistical procedure that permits statistical inferences about regionally specific findings (e.g. the probability of finding an activation focus by chance). Approaches based on eigenvalue solutions are mathematical devices that simply orthogonalise the variance-covariance, and make no comment on the significance of the resulting spatial modes (although a set of canonical images can be said to characterize a significant effect if ManCova is used to make an inferences about these effects beforehand).

The second half of this chapter reviewed the necessary steps employed to implement these analyses. This set of data transformations can be broadly classed as spatial transformations (i.e. realignment, spatial normalization and smoothing) and statistical transformations (i.e. eigenimage analysis of the adjusted data or the construction of SPMs). Statistical inference is usually based on SPMs and can either be made about a specific region or about regionally specific effects in a specified brain volume. In the latter case a corrected p value is required that protects against false positives. This correction is achieved using the theory of Gaussian Fields.

V1.A The ensuing chapters on data analysis and characterization

The next chapter deals with spatial transformations and subsequent chapters deal in a more formal way with the techniques introduced above, covering first univariate approaches and then multivariate approaches. These chapter discuss; (i) statistical models in the context of the general linear model, (ii) statistical inference in the context of the theory of Gaussian fields, (iii) Eigenimage analysis and multivariate analyses based on the general linear model. The next chapter addresses functional integration from a different perspective and introduces effective connectivity and related models (e.g. structural equation modelling and nonlinear extensions). It complements the previous chapter about characterizing distributed brain responses, but frames the characterization explicitly in terms of

connections or influences among brain regions. The following chapter returns to the different sorts of experimental design that have been employed in elucidating functional anatomy. It focusses on parametric and factorial designs to illustrate how modern brain activation experiments can (i) address some fundamental issues about functional organization and (ii) call for revisions of some of our basic assumptions. The order in which these chapters are presented reflects the ontology of their development and application. The vast majority of imaging research is predicated on some form of statistical parametric mapping so that a greater emphasis has been placed on these techniques.

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.Goltz F. in "Transactions of the 7th international medical congress" (W MacCormac, ed) Vol. I, pp. 218-228. JW Kolkmann :London, 1881.

3. J.R. Absher and D.F Benson. Disconnection syndrome: an overview of Geschwind's contributions. Neurology 43:862-867 (1993).

4. 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 (1993).

5. 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 (1993).

6. G.L. Gerstein, D.H. Lerkel. Simultaneously recorded trains of action potentials: analysis and functional interpretation. Science 1969;164:828-830.

7. G.L. Gerstein, P. Bedenbaugh, A.M.H.J. Aertsen. Neuronal assemblies IEEE. Trans. on Boimed. Engineering 36:4-14 (1989).

8. P.M. Gochin, E.K. Miller, C.G. Gross, G.L. Gerstein. Functional interactions among neurons in inferior temporal cortex of the awake macaque. Exp. Brain Res. 84:505-516 (1991).

9. 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.

10. J.I. Nelson, P.A. Salin, N.M.J. Munk, M Arzi, J. Bullier. Spatial and temporal coherence in cortico-cortical connections: a cross-correlation study in areas 17 and 18 in the cat. Visual Neurosci. 9:21-37 (1992).

11. 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.

12. G.H. Golub and C.F Van Loan. in "Matrix computations" 2nd ed. pp.241-248. The Johns Hopkins University Press, Baltimore and London, 1991.

13. 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.

14. S.Zeki. The motion pathways of the visual cortex in "Vision: coding and efficiency" (C. Blakemore ed.),pp.321-345. Cambridge University Press, UK, 1990.

15. R.J. Adler. in "The geometry of random fields" Wiley, New York, 1981.

16. P.T. Fox, M.A. Mintun. Non-invasive functional brain mapping by change distribution analysis of averaged PET images of H15O2 tissue activity. J.Nucl. Med. 30:141-149 (1989).

17. K.J. Friston, C.D. Frith, P.F. Liddle, R.S.J. Frackowiak. Comparing functional (PET) images: the assessment of significant change. J. Cereb. Blood Flow Metab. 11:690-699 (1991).

18. K.J. Worsley, A.C. Evans, S. Marrett, P. Neelin.A three-dimensional statistical analysis for rCBF activation studies in human brain. J. Cereb. Blood. Flow Metab. 12:900-918 (1992).

19. K.J. Worsley Local Maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields. Advances Appl. Prob. 26:13-42 (1994)

20. K.J. Friston, K.J.Worsley, R.S.J. Frackowiak, J.C. Mazziotta, A.C. Evans, Assessing the significance of focal activations using their spatial extent. Human Brain Mapping 1:214-220 (1994).

21. S.E. Petersen, P.T. Fox, M.I. Posner, M. Mintun, M.E. Raichle. Positron emission tomographic studies of the processing of single words. J Cog. Neurosci 1:153-170 (1989).
22. C.J. Lueck, S. Zeki, K.J. Friston, N.O Deiber, P. Cope V.J. Cunningham, A.A. Lammertsma, C. Kennard, R.S.J. Frackowiak. The colour centre in the cerebral cortex of man. Nature 340:386-389 (1989).
23. S. Grafton, J. Mazziotta, S. Presty, K.J. Friston, R.S.J. Frackowiak, M. Phelps. Functional anatomy of human procedural learning determined with regional cerebral blood flow and PET. J. Neuroscience 12:2542-2548 (1992).
24. C. Price, R.J.S. Wise, S. Ramsay, K.J. Friston, D. Howard, K. Patterson, R.S.J. Frackowiak. Regional response differences within the human auditory cortex when listening to words. Neurosci. Letters 146:179-182 (1992).
25. P.F.C. Gilbert and W.T. Thach. Purkinje cell activity during motor learning. Brain Res. 128:309-328 (1977).
26. K.J. Friston, C. Frith, R.E. Passingham, P.F. Liddle, R.S.J. Frackowiak. Motor practice and neurophysiological adaptation in the cerebellum: a positron tomography study. Proc. Royal Soc. Lond. Series B.248:223-228 (1992).
27. K.J Friston, P. Grasby, C. Bench, C. Frith, P. Cowen, P. Little, R.S.J. Frackowiak, R. Dolan. Measuring the neuromodulatory effects of drugs in man with positron tomography. Neuroscience Letters 141:106-110 (1992).
28. P. Grasby, K.J. Friston, C. Bench, P. Cowen, C. Frith, P. Liddle, R.S.J. Frackowiak, R. Dolan. Effects of 5-HT1A partial agonist buspirone on regional cerebral blood flow in man. Psychopharmacology 108:380-386 (1992).
29. P. Talairach and J. Tournoux. A Stereotactic coplanar atlas of the human brain. Stuttgart:Thieme, 1988.
30. P.T. Fox, J.S. Perlmutter, M.E. Raichle. A stereotactic method of anatomical localization for positron emission tomography. J Comput Assist Tomogr 9:141-153 (1985)
31. C. Chatfield, A.J. Introduction to multivariate analysis. Chapman and Hall, London. p189-210 (1980)