Special Issues in Functional Magnetic Resonance Imaging

Alistair M. Howseman, Oliver Josephs, Geraint Rees and Karl J. Friston

The Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London WC1N 3BG, UK

Contents

I. Introduction

Functional magnetic resonance neuroimaging is a technique which can be used to perform similar brain activation studies to PET, although non-invasively, but it can also provide additional information to that available from PET. The study of brain activity using functional magnetic resonance imaging (fMRI) poses an additional set of problems, both from an acquisition and processing standpoint, to those that have been comprehensively addressed for PET in previous chapters. The statistical and image processing framework for fMRI has been laid out in some of those chapters. In this chapter we will discuss the particular issues, in both imaging physics and physiology, which pertain to the design of fMRI experiments; and how we go about carrying out the data analysis in the context of the hypothesis which we are testing. Much of this work is ongoing in this laboratory and many others. Nevertheless we will review the substantial progress made in these areas and highlight some outstanding issues which need to be resolved in order for fMRI to become a routine tool.We can group this work into three broad categories.

1) *MRI Physics:* The fMRI signal depends on the effects that changes in cerebral blood flow and metabolism, consequent on changes in neural activity, have on the MRI physics. Optimal data acquisition is achieved through the consideration and refinement of many aspects of the MRI imaging techniques. MRI returns a more complex signal about the distribution and environment of brain water than does PET. Therefore potential sources of activation artefact need to be carefully considered.

2) *General issues in fMRI experimental design*: The design of fMRI experiments reflects the temporal resolution of fMRI and the large number of measurements that will typically be made. In particular the high temporal resolution facilitates event related fMRI experiments, in addition to condition contrasted experiments analogous to those performed in PET. Sources of noise in fMRI time series need to be carefully considered. The data analysis needs to take account of the fact that unlike PET, fMRI is not measuring absolute blood flow changes but a relative change in the oxygenation of venous blood. This undergoes a dynamic evolution in response to neuronal firing which is modelled by SPM. The statistical methods and pre processing steps used to find significant activations need to address the particular characteristics of fMRI data.

3) *Scanner hardware and environment:* The high magnetic field, restricted space and noise of the scanner make

special demands on the methods of stimulus presentation and subject accommodation within the system.

II. MRI physics

Functional MRI is a powerful imaging tool to study function in the human brain. The full potential of fMRI can only be realised when several aspects of this imaging technique are properly understood and have been optimised for experimental use. We can address the key issues by asking the following questions. What are the sources of fMRI signal and how may the experiment be tailored so that the activated areas accurately reflect areas of neuronal firing? What are the limits of spatial and temporal resolution in fMRI? How does the choice of imaging technique (pulse sequence) affect the data acquired and which imaging parameters are important? What types of image distortion and artefact exist and how can they be minimised?

Sources of Contrast

The mechanisms that produce Blood Oxygenation Level Dependent (BOLD) image contrast have been described in detail in the previous chapter. As a function of magnetic field strength BOLD contrast activation changes are approximately linear and so operation at high magnetic fields is desirable. Currently activation experiments are performed at field strengths between 1.5 and 4.0 Tesla. For a comparison of these field strengths see Jezzard and Turner (1994). Even higher field strength systems are being developed but there are very demanding technical difficulties to be overcome before it can be confirmed that these very high fields are practical for fMRI. The T_2^* BOLD effect is the dominant source of activation in fMRI images and has been highly characterised in terms of optimisation of imaging parameters (see later section) but it is not the only source of contrast. Signal increases during neural activation can also occur due to an apparent T_1 effect in the larger blood vessels. The presence of this effect and how it relates to neuronal activity can be paraphrased by the question “brain or vein - oxygenation or flow?” (Frahm et al, 1994). Flowing blood is not subjected to the same T_1 relaxation effects as stationary tissue. As fresh, unsaturated blood moves into the imaging slice it yields higher signal intensity (hence apparent reduction in T_1) than blood which has been repeatedly excited by the previous slice selection radiofrequency pulses. Wrongful interpretation of these signals as BOLD effects can be avoided by anatomical inspection and minimised in the data by the choice of MRI technique. Single slice imaging methods are very susceptible to these in-flow effects because the TR time is short compared to blood T_1. We have demonstrated that in multiple slice echo-planar imaging (EPI) these T_1 based in-flow effects are negligible, suggesting that BOLD contrast dominates.

The magnitude of the BOLD signal changes are a subject of much debate. Some of the very large signal changes reported in the early fMRI experiments have since been shown to be due to T_1 flow effects (Duyn et al, 1994, Frahm et al, 1994). At field strengths of 1.5-2.0 T signal changes of 1-2% from voxels containing pure grey matter are often quoted as being consistent with simulations of the biophysical mechanisms of BOLD contrast (Ogawa et al, 1993, Boxerman et al, 1995). At 4.0 T the equivalent signal changes are 5-6 %. However it has been shown that in the absence of in-flow effects, larger signal changes can still occur on T_2^* weighted images in the vicinity of large vessels (Haacke et al, 1994). Haacke et al proposed that the effect of a changed magnetic susceptibility in large vessels during stimulation may lead to an alternative effect on T_2^* gradient echo images. They suggested that to explain the effect of a large vessel within a voxel, the voxel may need to be compartmentalised and this can lead to bulk magnetisation phase shifts within the voxel. Vector addition of the magnetisation in the different compartments yields a signal which can exhibit very large changes during activation. We frequently measure maximum signal change of up to 20% in visual cortex with a multislice technique with no in-flow contamination (Porter et al , 1997, Howseman et al, 1997). It has also been shown that by increasing the spatial resolution of an imaging method the % signal change can be increased due to a reduction in partial volume effects (Frahm et al, 1993).

Spatial and Temporal Resolution

The spatial and temporal resolution of fMRI may ultimately be limited by the characteristics of the physiological changes accompanying neuronal activation as discussed in chapter 8, rather than by the resolution of the imaging techniques themselves. Consider that BOLD signals can originate within and surrounding the smallest capillaries and the largest venous vessels. Optical imaging techniques (Malonek and Grinvald, 1996) have shown that blood vessels during stimulation become more oxygenated over an area of a few mm in diameter around the site of neuronal activity. This may impose an intrinsic resolution limit for fMRI. An additional constraint is that oxygenation changes can be detected several mm downstream in the venous system from the site of neuronal activity. It has been suggested that mapping of the macrovasculature may be required to remove signals in large vessels to optimize the spatial accuracy of fMRI (Menon et al, 1993). At present there remains some doubt about the exact physiological location of the spins which are the source of the BOLD effect. Theoretical modelling of the magnetic susceptibility gradients and Monte Carlo simulations have been performed to investigate the physical mechanisms of BOLD contrast. Due to the very low volume of blood within brain tissue it has been proposed that extravascular spins contribute heavily to BOLD contrast (Ogawa et al, 1993) because local magnetic field gradients exist around vessels. However other work suggests that the intravascular spins dominate the mechanism of signal change (Boxerman et al, 1995). These models also address the magnitude of signal change associated with different vessel sizes and show that by using spin echo imaging techniques to detect changes in T_2 during activation, the greatest sensitivty to oxygenation changes are limited to small vessels (on the order of 10 μm in diameter) (Weiskoff et al, 1993). This suggests a possible approach to the removal of larger vessels from activation images. However activation imaging using T_2 methods yields much smaller signal changes than those on T_2* images. This is discussed further in the next section.

The ultimate temporal resolution of fMRI is likely to be determined more by physiological dynamics than the rate at which data can be acquired. Single slice EPI images can be acquired for example at up to 40 images per second. Typically whole brain multislice EPI data sets are acquired in 5 seconds. The BOLD signal response to neuronal activity which operates on a time scale of 1ms (synaptic transmission) to 100s of ms (information transport) is heavily lagged and damped by the hemodynamic response. The peak BOLD signals in activated primary cortex have a measured latency of 5 - 8 seconds from stimulus onset and a similar time is required for the signal to return to baseline upon stimulus cessation (Kwong et al, 1992, Blamire et al, 1992). This corresponds to a hemodynamic response function with a FWHM of approx 6 seconds (Friston et al, 1994) which is convolved with the neuronal firing at the data processing stage. Bandettini et al (1993) have looked at the effect of varying the inter stimulus interval on the time course of BOLD contrast and reported that if individual stimuli are presented more rapidly than the width of the hemodynamic response that contrast is reduced because the signal does not have sufficient time to return to the resting level. For a finger tapping task they also reported that activation could be detected for stimulus periods as short as 0.5 sec. Assessing the measured response to a single stimulus or task will be addressed in the section on data analysis of event related fMRI.

Imaging Techniques

Echo-planar imaging (EPI), an ultra high speed imaging method, and its variants have become the methods of choice for the majority of functional MRI studies. The ability of EPI to monitor changes in the oxygenation state of the brain on a subsecond basis was first demonstrated by Turner et al (1991). EPI was proposed by Mansfield (1977) but has only recently been installed on clinical MRI systems. One of the reasons for the slow emergence of EPI as an MRI imaging method of widespread availability is that it places stringent demands on the performance of the MRI scanner hardware, in particular the gradient subsystem. Even today there are perhaps only 100 - 200 scanners in the world with sufficient gradient performance. EPI allows functional imaging experiments to be performed with similar or slightly improved spatial resolution to PET whilst introducing a temporal dimension to activation experiments. Although not close to the ms resolution of EEG the spatiotemporal features of EPI are unique in functional neuroimaging.

Higher spatial resolution (on the order of 1mm in-plane) can be achieved in fMRI by using rapid gradient echo techniques such as FLASH which can acquire a single slice of data in 1-10 seconds. FLASH may be the method of choice if one is attempting to very accurately localise activation to sulcal anatomy. A limitation of FLASH

is that if it is to be used for multislicing the acquisition times become rather long. Consequently it should be thought of as a high spatial resolution method with limited brain coverage. Used as a single slice method with a short TR, FLASH is very sensitive to in-flow effects and to minimise these a tip angle of 10^0 is usually used and possibly saturation bands are placed either side of the imaging slice (Frahm et al, 1994).

One of the great advantages of EPI is that it can be used to map the whole brain with excellent temporal resolution when used as a multislicing technique. For example a 64x64x64 image matrix with isotropic 3mm resolution can be acquired in approximately 5 seconds. The TR for each slice is then 5 seconds and with tissue and blood T1s on the order of 1 second at fMRI field strengths, saturation effects are minimal. We have tested this assumption by acquiring data for different slice orders (Howseman et al, 1997). Visual stimulation was used and in a single imaging session identical fMRI experiments were performed for ascending, descending and interleaved slices. The differences between the measurements were assessed by contrasting the activations using a condition by experiment interaction in SPM. The data showed that even at a lenient threshold of p=0.001 (uncorrected) there were no significant differences between the data sets.

A variant on multislice 2D EPI is to apply phase encoding in the 3rd dimension (Guilfoyle et al, 1989). We have implemented this for fMRI of the whole human brain (Porter et al, 1997). A direct comparison between 2D and 3D can be made because the same matrix size and resolution can be acquired for each in the same imaging time. A factor of 2 improvement in S/N ratio with the 3D technique has been achieved (using a tip angle of 30^0) but to optimise image contrast a tip angle of 15^0 has been chosen which yields a S/N ratio improvement of 1.7. Very similar patterns of activation are seen with the two techniques. The purpose of developing this method was to get more significant activations by virtue of the improved S/N. This is based on the assumption that instrumental noise rather than physiological noise is the dominant source of noise in fMRI time series. It is unclear at present whether consistently higher Z-scores have been achieved. Comparing the data sets on a voxel by voxel basis is difficult because of the different effects of magnetic susceptibility. Characterisation of the distribution of Z-scores in the two data sets requires further work.

The majority of fMRI studies use gradient echo imaging techniques to obtain T_2^* contrast. Confirmation that the signals being measured are T_2^* effects was established by an experiment of Bandettini (1994) in which the echo time dependence of BOLD contrast was studied. Optimal contrast to noise ratio (not magnitude of signal change) is obtained by setting the gradient echo time approximately equal to the T_2^* of grey matter. Neuronal activation can also be detected on spin-echo images due to changes in tissue T_2 relaxation times during stimulation. The mechanism for T_2 contrast also results from changing deoxyhemoglobin concentrations but is due to diffusion of water in the microscopic field gradients set up around the capillaries and venules (Weisskoff et al, 1993). However the sensitivity of T_2 methods is much less than T_2^* imaging because the static dephasing components are refocussed. One of the attractions of spin echo methods is that in regions of gross field inhomogeneity which occur at interfaces between tissue and air or bone, signal dephasing in long echo time images can be reduced (see next section). It is precisely this feature which improves image quality which also limits the sensitivity to changes in deoxyhemoglobin concentration. An alternative approach known as assymetric spin echo (ASE) imaging is a compromise between the two techniques and is well suited to EPI in which the receiver is kept open for a relatively long time (approx. 50ms). In this case the spin echo is timed to coincide with a position in k-space which is not central. i.e. spin rephasing occurs earlier than the acquisition of the low spatial frequencies (the bulk of the signal). From a BOLD point of view the contrast echo time is effectively reduced to the time between the spin and gradient echoes.

Image Distortions and Artefacts

Functional MRI techniques are deliberately sensitised to differences in T_2^* so that the magnetic properties of deoxy-haemoglobin can manifest themselves as image contrast. An unfortunate consequence of this is that we are also sensitised to other sources of magnetic field inhomogeneity, most significantly macroscopic effects which are detrimental to image quality. The magnetic susceptibility differences between air, bone and different tissue types cause large scale inhomogeneity, particularly at the high field strengths typically used for fMRI. Components of inhomogeneity within the imaging plane cause geometric distortion of the images and those co-

linear with the slice selection direction cause a loss of image intensity due to spin dephasing (Frahm et al, 1988). This signal loss effect is particularly severe at the anterior frontal pole near to the sinuses and in the temporal lobe above the petrous bone. Geometric distortions are a particularly serious problem with EPI because of the very low frequency per point in the phase encoding direction. This distortion needs to be minimised by selection of the imaging parameters otherwise serious mis-registration in relation to anatomical images can result. Typically for echo-planar fMRI, magnetic field inhomogeneity sufficient to cause dephasing of 2π through the slice corresponds to a 1 voxel distortion in the plane. 'Unwarping' schemes have been proposed to correct for the distortion effects in post-processing (Jezzard and Balaban, 1995). To improve the image homogeneity such that the most difficult areas of the brain may be investigated without sacrificing too much sensitivity in BOLD contrast is a delicate tradeoff. As mentioned above, work on techniques using a combination of spin and gradient echoes may tackle this problem. An alternative approach is to locally shim the region of brain which is of interest (at the expense of the rest of the brain) (see Blamire et al, 1992).

Another common problem with echo-planar images is the presence of Nyquist ghosts. These are low intensity (approx 1%) secondary images which appear a half FOV away from the real image. Their presence is due to timing or phase differences between the odd and even echoes in the echo train. They are minimised by calibration using a pre scan under a bipolar gradient prior to data acquisition. Nyquist ghosts may represent a problem in fMRI because they can be unstable over time due to shim and gradient amplifier instabilities. The total power of the real and ghost image remains constant but the intensity can leak between the two. Provided that their intensity is not correlated with the experimental task there should not be a serious problem in analysis although if their intensity becomes too great image realignment may be affected.

III. The nature of fMRI signals and noise

In this section we will discuss the nature of the signals present in an fMRI experiment and the potential sources of noise. Figure 9.1 represents a single voxel fMRI measurement. A pre-requisite to designing a successful fMRI paradigm is to ensure that the evoked neural activity generated by the stimulus or task constitutes a signal which can passes through all the transformations shown, to the stage of statistical analysis, without being completely masked by noise and distorting processes.

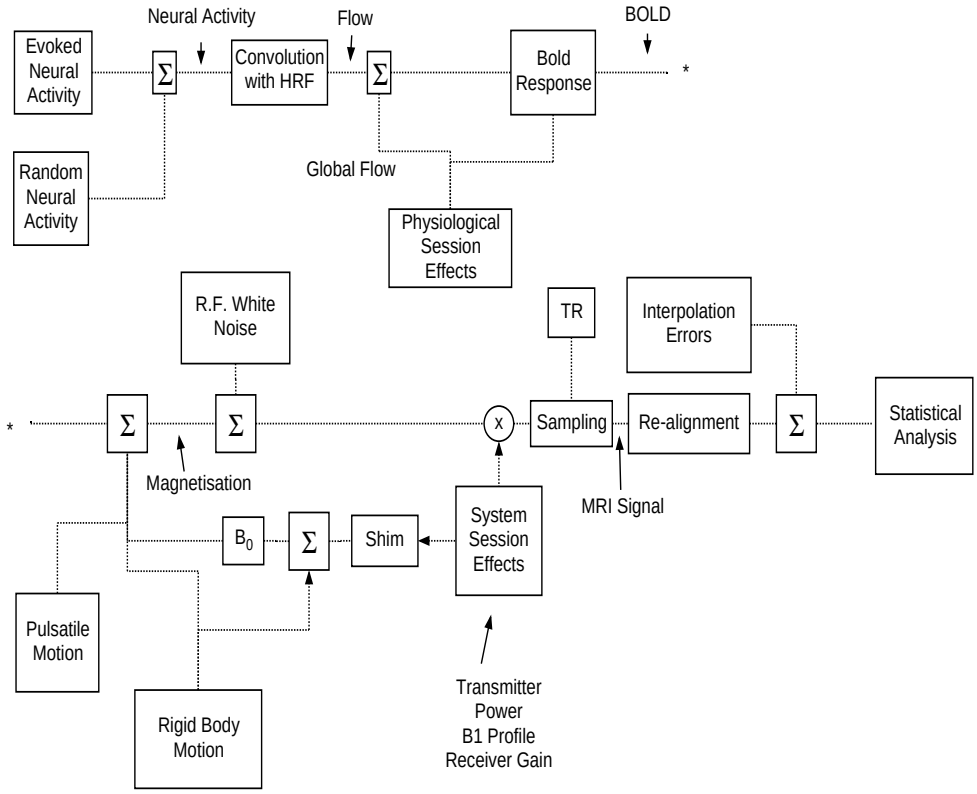

Figure 9.1

Signal Path

As discussed in the previous chapter, neural activity requires increased local blood flow to supply additional oxygen. Neural activity can change quickly. Whole neural populations can activate in under 10 ms. The haemodynamic response is slower (usually greater than 1 s passes before flow increases). This response is usually modelled as affecting the signal by a process of convolution. The locally enhanced flow (and increased blood volume) increases the effective T_2^* and gives rise to the so-called BOLD (Blood Oxygenation Level Dependent) contrast effect. The BOLD contrast magnetisztion signal is sampled at discrete time points (once every TR [repetition time]) to yield the MRI signal. This is digitised and then further processing, which may include spatial re-alignment, normalisation and smoothing, is performed.

Sources of Noise

In addition to the activity evoked by the experiment, endogenous neural activity will cause fluctuations in flow, as may global physiological effects. In addition there may be other influences of the physiological state on the BOLD response.

Various forms of real and apparent motion may constitute a fluctuating noise source to the measurement. Some causes are rigid body motion of the subject's head, pulsatile motion of the brain caused by the cardiac and respiratory cycles and local modulation of the static magnetic field (B_0) by movement (e.g. respiration and speech). These effects are typically low frequency (e.g. slow nodding of an incompletely restrained head during the course of an experiment) or wide-band (e.g. aliased cardiac-locked pulsatile motion).

R.F. (Johnson) noise contributes further measurement noise. This noise which is completely wide-band in character and which originates on the r.f. coil and within the subject, determines the SNR of the MRI images.

Several instrumentation effects (e.g. transmitter power calibration, B_1 coil profile and receiver gain) will scale the signal before the MRI sampling process.

At the stage of spatial re-alignment (which aims to correct for rigid body motion), processing time constraints prohibit the use of ideal interpolation schemes. Interpolation errors will be be introduced. The errors are functions of the predominantly low frequency motion and thus constitute a low frequency noise source.

Frequency Analysis

Data from an fMRI experiment forms a time-series embodying many hundreds of measurements at each voxel. If the data are acquired sufficiently fast (e.g. every 6 seconds or less) this time-series can evidence temporal autocorrelations or temporal smoothness due to the hemodynamic response function and other sources of physiological noise. The effect of the hemodynamic response function can be thought of as convolving an underlying neural time-series with the response function to give the observed hemodynamic time-series (Friston *et al* 1993). In this context, time-dependent changes in the BOLD signal may be represented, in frequency space, as the sum of several frequency components (where the spectral density is the Fourier transform of the autocovariance function). Each frequency component has a different source; heuristically these sources can be thought of as either (i) signal, related to functionally activated brain areas (i.e. neuronal activity convolved with the hemodynamic response function) or (ii) related to noise such as arises with aliasing of physiological bio-rhythms or slow movement artefacts. The successful experimental design will therefore be one in which signal and noise in the fMRI time series are not confounded. Noise is usually a low frequency component of the fMRI time-series and is largely due to cardiac and respiratory aliasing. For example, if scans were acquired every six seconds and the subject's respiratory cycle was 6.1 seconds, then each scan would be 0.1 seconds out of phase. Over a prolonged period of scanning, this asynchrony would create an artefactual signal of low frequency (0.1 Hz). Other low frequency components relate to Meyer's waves and related phenomena that are secondary to slow periodic hemodynamic changes (probably due to intrinsic autoregulatory feedback of cerebrovascular tone). In what follows we present a characterization of noise, dealing first with periodic noise and then with non-periodic noise components.

Periodic Noise Sources

Two important sources of periodic physiological noise in brain fMRI are the cardiac and respiratory cycles. The heart cycle causes both periodic blood flow and pulsatile bulk motion due to induced pressure variation. The blood flow effects are confined mainly to vessels but the pulsatile motion can extend throughout the entire brain. Respiration causes both a generalised variation in blood oxygenation and bulk displacement of the head in reaction to breathing. Both oxygenation changes and small displacements of tissue can be causes of image voxel intensity changes. The blood oxygenation level contributes directly through the BOLD contrast mechanism. Motion, which is small compared to the image voxel size, can be considered to operate on a given voxel at a certain time according to a Taylor expansion in image spatial derivatives about the voxel (see Friston *et al* 1996).

The expected time course of measured physiological noise depends not only on the sources but also on the imaging TR (repetition time). If the TR is short compared to both cardiac and respiratory cycles (i.e. TR << 1 second) then both will be seen as straightforward periodic functions. In a frequency spectrum (neglecting harmonics) each will appear as a peak with a central "mean" frequency and with widths representing departure from strict periodicity. A schematic spectrum is shown in Figure 9.2.

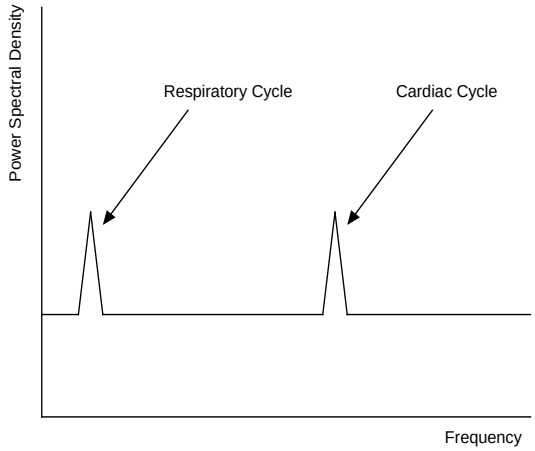

Figure 9.2

A very short TR is achievable with single slice EPI imaging (TR ~ 100 ms). For such short TR measurements it may be worthwhile to apply simple notch filters centred on the mean cardiac and respiratory frequencies. Clearly any activation frequency component within the notch filter stop bands will be discarded along with the noise. If whole brain multislice EPI, or other slow single slice techniques, are used the TR is typically several seconds. This is above the Nyquist limit for both the cardiac and respiratory noise and so *aliasing* will occur (Aliasing is the strobing effect apparent in discrete sampled time series whereby whole periods between sample points are not registered. For example a 1.1 Hz signal sampled at 1 Hz is indistinguishable from a 0.1 Hz signal). In general, for a sampling frequency F_R and a noise of frequency F_N the measured frequency F_M is as shown in Figure 9.3.

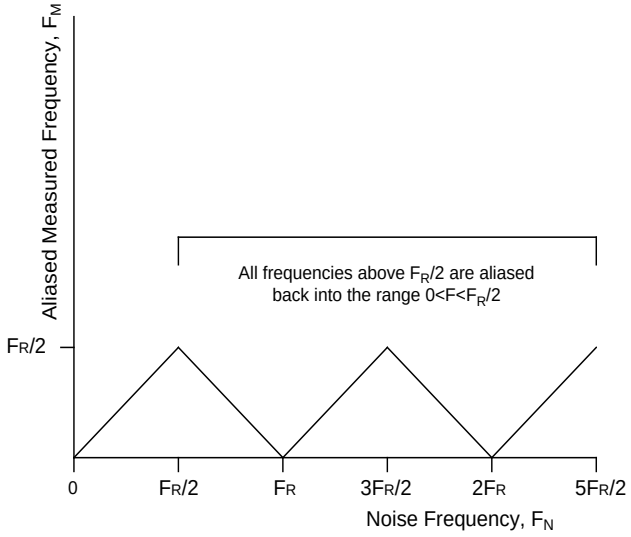

Figure 9.3

The graph demonstrates that the aliased spectral density function depends primarily on the spectral **width** of the noise. If the noise is centred at an arbitrary frequency but extends over a width of $\Delta F_N > F_R$ then its aliased contribution will extend across the entire measurement frequency spectrum. In this situation (i.e. non-monochromatic noise and TR long compared with noise period) a simple notch filter would not be appropriate.

Non-Periodic Noise Sources

In addition to respiratory and cardiac related noise, low frequency noise components - typically manifesting as

"drifts and shifts"- are evident in fMRI time courses. These can arise because of long term physiological shifts, movement related noise remaining after re-alignment and adjustment, or from instrumental instability. The spectrum of this noise typically has a "1/f" characteristic, so-called because the spectral power density falls linearly with increasing frequency. In addition both the subject and scanner will contribute thermally generated white noise. These two generic types of noise combine to generate the form of spectrum seen in Figure 9.4. In summary the noise spectrum of the fMRI signal comprises a low-frequency and wideband component. Periodic physiological noise sources can appear as focal peaks in the spectrum (short TR) or increased wideband noise (long TR).

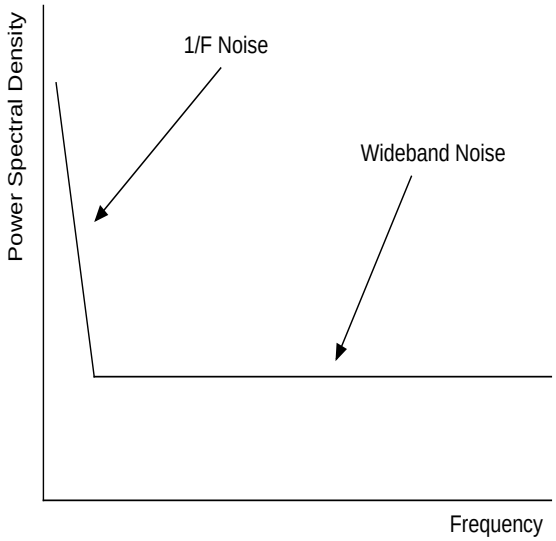

Figure 9.4

IV. Experimental design and data analysis

An fMRI experiment to test a given biological hypothesis must be designed within the constraints of the temporal characteristics of the fMRI BOLD signal and of the various confounding effects to which fMRI is susceptible. Broadly these limit the range of frequencies for which an activation signal is likely to be unambiguously detectable to a range extending from approximately one cycle per few minutes to one cycle per few seconds as explained above. Within these frequency constraints many different forms of experimental design can be considered. It is convenient to characterise the stimulus or task intensity and timing as functions of time. These functions are typically of two general forms, representing either blocks of stimulation, or describing discrete, non-zero events (i.e.. the occurence of brief stimuli). Experiments akin to those performed in PET acquire data in blocks (e.g. boxcar functions of alternating activation and rest) in which the steady conditions are maintained for say 30 to 60 sec. duration. Additionally data may be recorded to monitor the BOLD response following a single stimulus or task. This is known as event related fMRI.

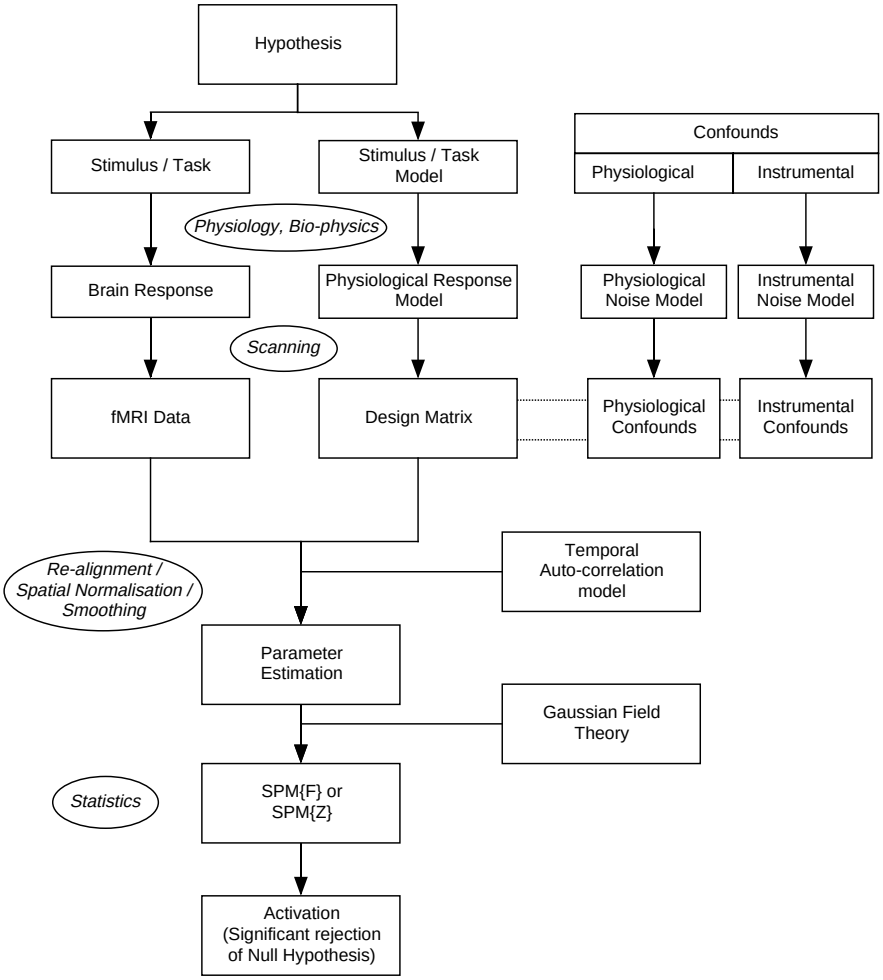

Figure 9.5

Figure 9.5 is a schematic diagram showing how SPM is used to process a generic fMRI experiment. Upon the specification of a hypothesis and after designing an experiment and acquiring data, the hypothesis is framed as a design matrix model. This takes account of the physiological (BOLD) response to neuronal activity and the presence of confounding effects in the data. In other words, in order to test the significance of any activation the biological effect of the stimulus and the response of the scanner must be modelled and confounding effects must be accounted for. The neural responses elicited by a given task give rise to haemodynamic effects which are sampled at discrete intervals by the scanner. The data are re-aligned to a reference image, spatially normalised to a reference template and spatially smoothed as described in earlier chapters. The application of appropriate physiological and bio-physical models transform the stimulus / task model to a physiological response model. In the case of state-related fMRI the physiological model is the stimulus / task model convolved by a postulated haemodynamic (BOLD) response function. In primary sensory cortex a peak in the fMRI BOLD signal is reached between 5-8 seconds after the onset of stimulus (Bandettini, 1993). The hemodynamic response function for such data has been fitted to a Poisson distribution by Friston et al (1994). If this is not used, and depending on the timing of the experiment, correlations between brain activity and sensory input can either be missed entirely or a loss of statistical power can result. Alternatively, and, for event-related fMRI necessarily, the precise form of the haemodynamic response function may not be assumed *a priori* in which case the physiological response model may comprise several terms, each corresponding to one basis function of a potential haemodynamic response convolved by the stimulus / task function. The physiological response model is sampled to obtain the design matrix.

As described by the theory of the general linear model (Friston et al, 1995c) the design matrix is formed by

designating a set of columns which correspond to experimental conditions of interest (the hypothesis under test) and a set of columns which model effects of no interest. These confounds may consist of sampled effects such as cardiac and respiratory noise and instrumental confounds (e,g, receiver gain effects) and other unspecified low frequency baseline drifts which may be modelled using a high pass filter (see Chapter 3).

As described in previous chapters the purpose of the parameter estimation stage is to discover, for each voxel, the linear combination of design matrix columns which best fits the data (in a least-squares sense). The estimated parameters are then converted to a Fisher F statistic for every voxel in the brain volume and displayed as an SPM{F}. Additionally an SPM{Z} may be produced which represents Student's t-statistic. Using the theory of Gaussian Random Fields the spatial distribution of the F (or Z) statistic in the SPM under the Null hypothesis may be determined and thus significances, corrected for multiple comparisons may be assigned to voxels, clusters and sets of voxels.

State-related fMRI

Two techniques can be used (often, in combination) to counter noise. Firstly, noise processes can be modelled and their estimated contribution removed from a measured signal. This technique is used, for example, to remove movement-related signal variation mediated by differential spin excitation histories (Friston *et al* 1996). Secondly, an experiment can be designed to take account of the characteristic features of the measurement system, for example the noise and sensitivity frequency spectra. An appropriate example in the present context is the method of "chopping" illustrated in figure 9.6. Chopping involves alternating between stimulus or task conditions to generate an time-dependent activation with a fundamental frequency sufficiently high to be clear of the 1/f region in the noise spectrum without being so high as to be attenuated by the sluggish hemodynamic response. Chopping utilises the high temporal resolution of fMRI to compensate for its poorer low frequency noise performance. This leads to a choice of chopping frequencies between ~1 cycle / few minutes (a typical low frequency noise cut-off point) and 1 cycle / 10 seconds (above which the signal would be attenuated by the hemodynamic response function).

Most fMRI studies to date have been performed using a boxcar design with alternating epochs of rest and a single activation task of about thirty seconds. In the context of the theoretical analysis presented above, this ensures that activations of interest and physiological noise are confounded as little as possible. The recurring rest periods provide a constant baseline that can be used to assess low-frequency baseline drift secondary to physiological noise. These other low frequency effects can largely be attributed to residual motion effects and may be assessed explicitly by comparing rest periods at the beginning and end of the experiment. In practice these effects are removed by the columns of the design matrix which constitute the high pass filter (next section). A very simple example of a state related fMRI experiment is shown in figure 9.7. A visual stimulus consisting of a black and white alternating checkerboard is presented in 30 sec. blocks interspersed with rest. The SPM MIPs show the activated areas in visual cortex during presentation of the stimulus (N.B. MIPs are not normalised to glass brains). The time course of a highly significant voxel is also shown. Note that the magnitude of signal change between rest and activation is appproximately 20 %.

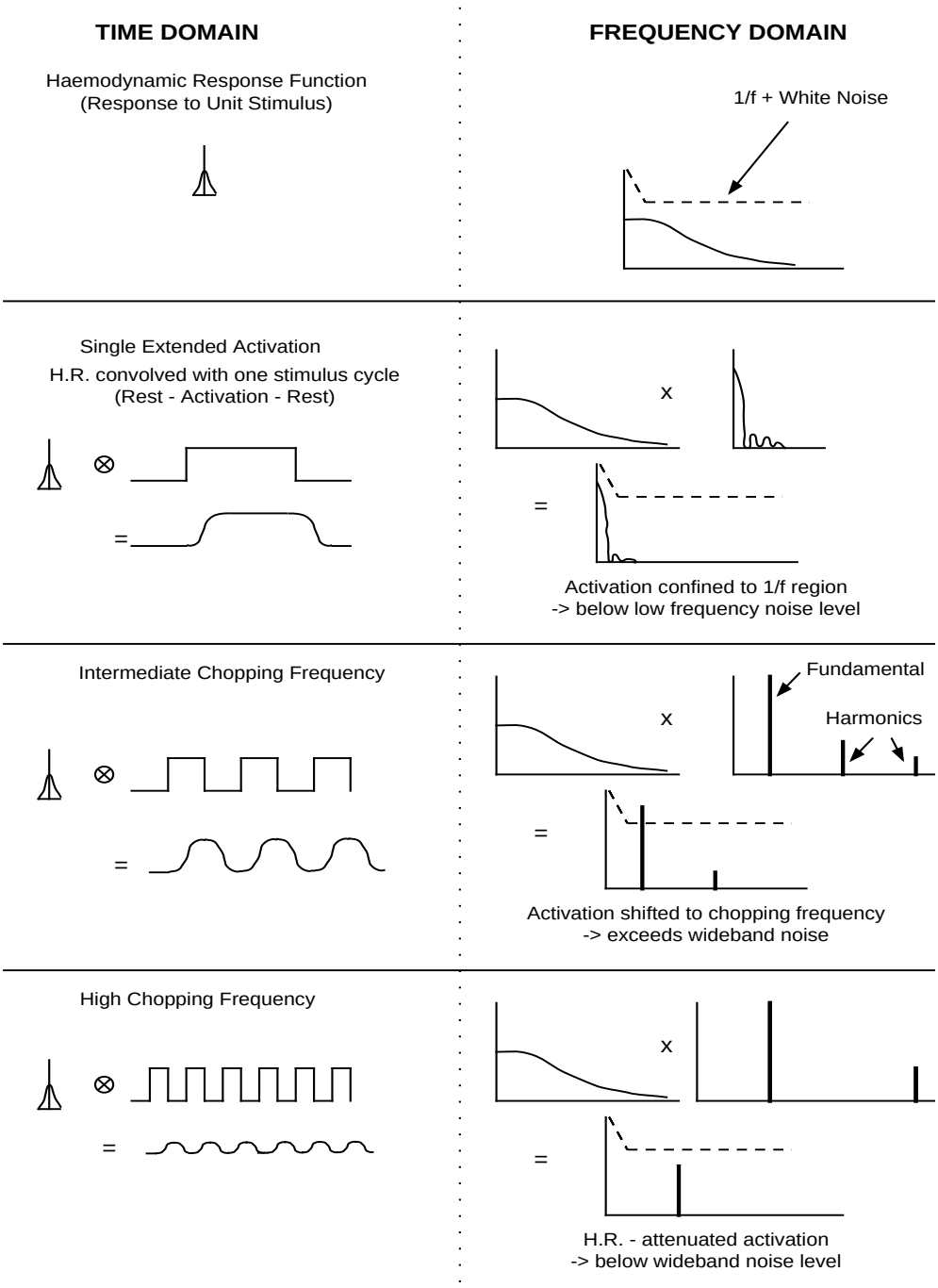

Figure 9.6

It should be noted that depending on the paradigm used it is still not clear how long the alternating periods of rest and activation should be. State related fMRI studies usually assume that sustained activation occurs during each cycle. For primary sensory cortex the time course of activation shows that this is generally the case. However it may not be safe to assume that this applies to all areas of the brain, for different tasks, or that activation is sustained at a constant level for the duration of an experiment. It should also be considered that the hemodynamic response function may be time dependent and may vary across brain regions.

As an example of a non sustained response during a cyclic fMRI experiment consider that in a cognitive motor task Friston et al (1995a) demonstrated *biphasic* temporal responses in the anterior cingulate over a thirty second period. This is an important observation because these significant responses could not be classified as 'activations' when averaged over the entire epoch. The form of the hemodynamic transients evoked during this task have been explored using a multivariate analysis by Friston et al (1995b). Having identified a significant evoked response using MANCOVA, its form was then characterised with canonical variates analysis (see the chapter on eigenimages and multivariate analyses). This approach has the advantage that it makes no prior assumptions about the form of the evoked hemodynamic response. The study investigated two motor sequencing conditions, differing only in that in one the subject made a predetermined movement (cued visually) and in the other a random direction was specified by a visual cue. The evoked responses for the two conditions were significantly different. The important message to take from this work is that the evoked hemodynamic response to different types of task may differ, not only in the degree to which they are expressed but also in their form and shape.

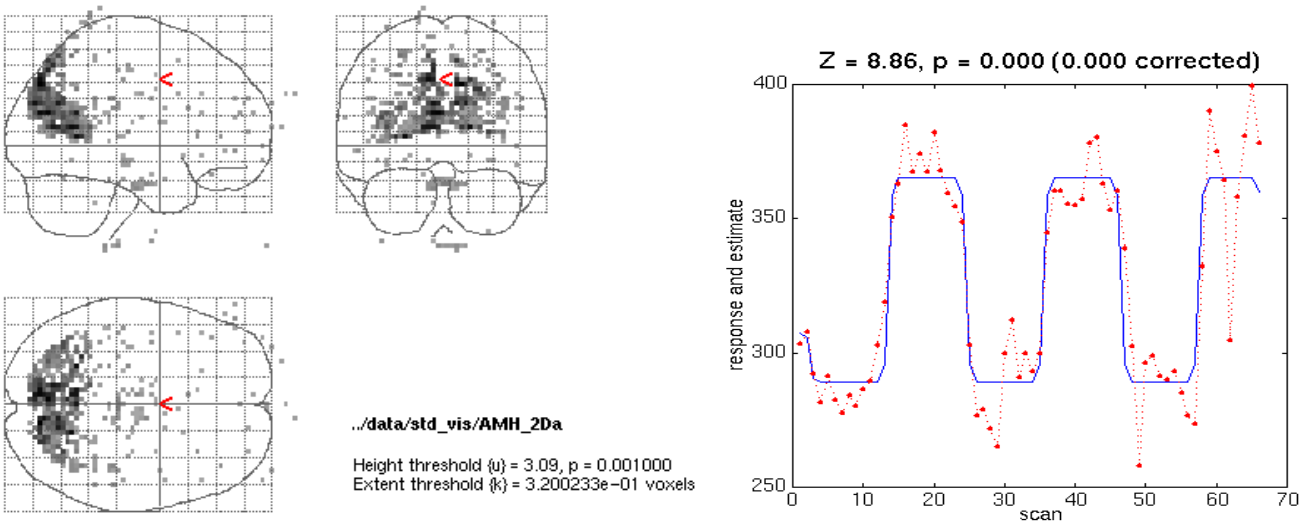

Figure 9.7

High pass filter

The application of a high pass filter (Holmes et al, 1997, see chapter 3) to the data isolates the high frequency effects of interest and removes most of the (low frequency) physiological noise. It is important to set the filter periodicity appropriately with regard to the experimental design in order not to remove signals of interest. For example, a signal of interest of periodicity 60s can be achieved by alternating periods of thirty seconds activation (five whole brain acquisitions with a TR of 6 seconds) with thirty seconds rest. Subsequently applying a high-pass filter of periodicity 120s will remove noise but leave the signal of interest intact. This is illustrated in figure 9.8 where a high-pass filter is applied to a data set which involved alternating cycles of rest and finger opposition. The first two columns in the design matrix tell us that there is activation for ten scans (60 seconds) then rest for ten scans (60 seconds), i.e. the rest-activation cycle is 120 seconds. In this example we have set the high-pass filter at 240 seconds to remove noise but leave the signal intact. The next 12 columns of the design matrix are the high-pass filter. Each column represents a sinusoid of gradually increasing frequency. It can be seen from the third from right column that the highest frequency column of the design matrix only has three complete cycles over 120 scans (720 seconds), i.e. 240 seconds. Thus the signal we are interested in, which has a periodicity of 120 seconds, is left intact.

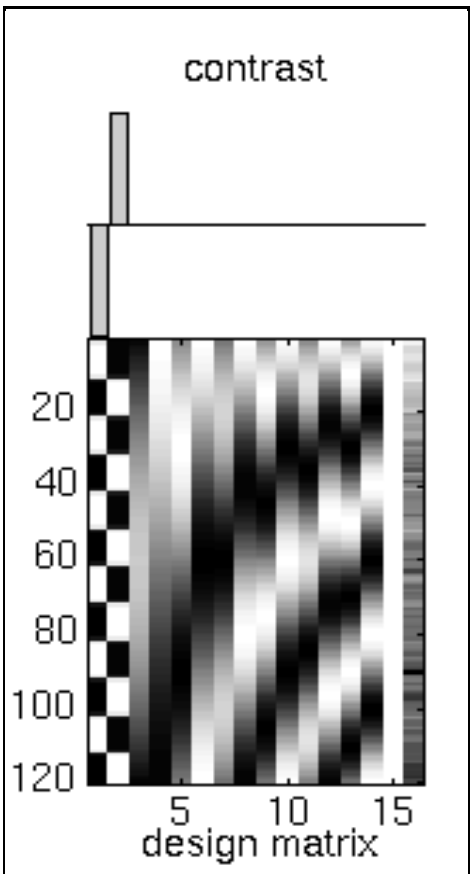

Figure 5. Implementation of a high pass filter within the design matrix

Figure 9.8

Event related fMRI

fMRI paradigms can be performed *without* placing the subject in a number of contrasting cognitive "states". Instead, "event-related" haemodynamic responses may be measured directly (Josephs et al, 1997). Voxels are considered activated if they show a consistent repeated response to each event, irrespective of the precise time course of the response. One key advantage of this approach is that the experiment need not be constrained to successive, rapidly repeated trials of the same type as are current blocked designs. For example, the response to odd-ball stimuli can be assessed. In addition the increased lattitude of design allows for the analysis of uncontrollable events, e.g. hallucinations.

The statistical inference provided by this technique is tailored specifically to time-locked activation and is formulated in terms of the General Linear Model as implemented within SPM. The model for the response is base on Fourier series (or other) smooth basis functions. The SPM{F} is applied to make inferences about the significance of all components of the model. The inference may concern simply the significance of activation, or demonstrate the significance of differences between responses to different event types (analogous to N300/P400 mismatch measurements in electrophysiology).

Measurements can be made over the entire brain using multisliced EPI. Although this usually entails a repeat time of approx. 5 s, the finer temporal structure in the event-related response is characterised by varying the phase of the stimulus relative to the acquisition. This gives a temporal resolution which is greater than the scanning repeat time. An example event related fMRI paradigm is to present a stimulus repeatedly, approximately every 20 seconds with the TR of the acquisition set to 5.2 sec. The timing of the stimulus and scanning is therefore arranged so that each presentation of the stimulus is at a slightly different phase of the scanning cycle. An illustrative auditory evoked response SPM{F} and the associated time course from the marked voxel is shown in figure 9.9 (TR=1.7 sec., ISI=33.15 sec.).

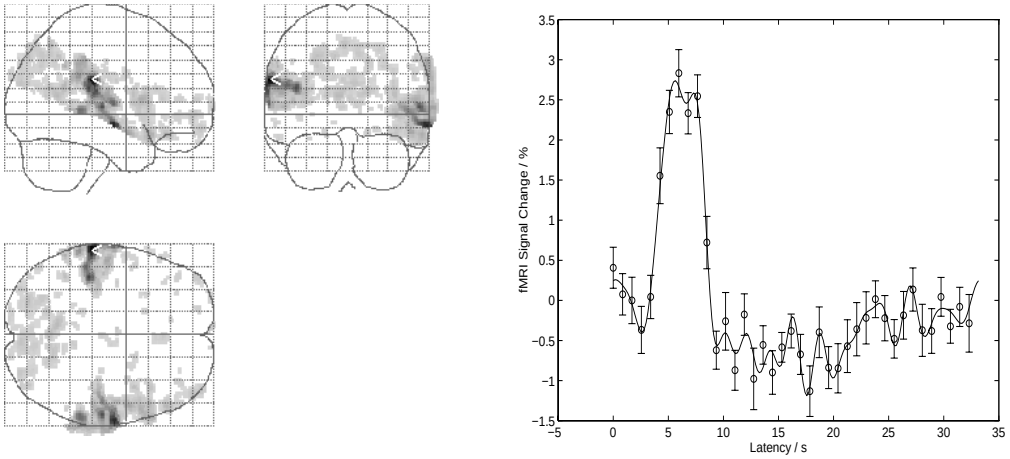

Figure 9.9

An event-related analysis may be viewed as the least-squares deconvolution of the measured data with respect to the stimulus function to yield the underlying haemodynamic response function. This is in contrast to deconvolving the data with respect to an estimated haemodynamic response function to obtain an estimate of the underlying neural activity.

Events may be extended in time. The event-related haemodynamic response is then defined as the response at the voxel to the extended event. Such an event might, for example, be the “block-followed-by-rest” of a state-related design. Thus a state-related design may be analysed using an event-related model. From this perspective the event-related model is a generalisation of earlier “temporal basis function” models.

Sustained BOLD contrast

Although it is desirable to alternate epochs of rest and activation during an fMRI experiment, it is apparent that certain phenomena such as motor learning or sustained attention do not lend themselves to this approach. In these cases the psychological phenomenon of interest is only elicited during a sustained period of activity. These types of experiment are therefore extremely vulnerable to confounding with the low frequency effects of physiological noise. It may not be possible to unambiguously attribute any changes in BOLD signal to the psychological process of interest rather than to unrelated physiological states. It is therefore helpful to compare activation states with adjacent reference baseline states wherever possible. In short this means that the assessment of enduring cognitive or sensorimotor states must be framed in terms of condition x time interactions. In relation to this it should be noted that the approach outlined here of alternating epochs of a reference condition and an activated condition does not require that the reference condition is rest. The constraint is merely that reference conditions are repeated through the course of the experiment at a frequency that is high relative to low-frequency sources of noise. We reemphasise here the need for alternating task conditions. All BOLD contrast fMRI experimental designs should embody the concept of a contrast between two experimentally

specified states; there is probably no such thing as ‘steady state’ BOLD contrast fMRI.

As an example of the difficulty of steady state fMRI consider the controversy surrounding the time course of activation in primary visual cortex throughout an extended period of stimulation (Kruger et al, 1995, Bandettini et al, 1995, Howseman et al, 1996). In the case of the Gottingen group very rapid decay of signal is seen, suggesting that the BOLD signal has returned to baseline within 2-3 minutes of stimulus onset. Our results and those of others show no evidence of a ‘vanishing’ BOLD signal.

Temporal autocorrelations and statistical modelling

Analysis of an fMRI time series involves fitting a model to the observed data; the model is composed of several different components specified in the design matrix (see the chapter on statistical models and experimental design). However the general linear model used for PET data analysis assumed that each scan represents an independent observation. The presence of temporal autocorrelation in fMRI time-series is a violation of this assumption. The approach we have adopted is to extend the general linear model to accommodate these temporal autocorrelations. Details will be found in the chapter on statistical models and experimental design. In brief the variance estimators are recalculated under the assumption that the data represent an uncorrelated set of observations that have been convolved with some smoothing kernel. The resulting statistics are then distributed with the t distribution whose *effective degrees of freedom* are a function of the smoothing kernel. The effective degrees of freedom are less than the number of scans and correspond roughly to the number of scans divided by the width of the smoothing kernel. This approach assumes the temporal autocorrelation function is known, or can be estimated. In many instances it can be deduced from the temporal smoothing kernel applied to the time-series to increase signal to noise. This temporal smoothing is considered in the next section

Smoothing in space and time

Smoothing in space enhances the signal to noise ratio of the data and allows intersubject averaging by blurring differences in gyral anatomy between subjects. The validity of statistical inferences also depends on the data being sufficiently smooth to allow use of the Theory of Gaussian Fields (see the chapter of statistical inference and the theory of Gaussian fields). On the other hand, fMRI has high anatomical resolution and so there is a trade off between the degree of smoothing and the spatial resolution to be considered. Our approach has been to use the least degree of smoothing that is compatible with the smoothness being large enough to ensure the validity of the statistical inferences (i.e. a smoothness of at least 2 voxels).

As noted in previous sections fMRI time-series contain a number of signals; these include uncorrelated noise (e.g. Quantum noise and thermal noise), correlated noise (e.g. physiological noise from cardiac and respiratory cycles) and correlated signal that conforms approximately to changes in neural activity convolved with a hemodynamic response function (Friston et al 1994). Correlated noise can arise directly from cardiac and respiratory cycles, their physiological modulation (e.g. the heart rate variability signal) or aliasing of these effects due to an interaction with the repeat time.

Smoothing in time or convolution of the fMRI time-series with the hemodynamic response function will, in principle, enhance signal relative to noise, particularly thermal and other noise with high frequency components. This de-noising device is predicated on the 'matched filter theorem' and is based on the conjecture that 'interesting' hemodynamics are the result of convolving an underlying neuronal process with a hemodynamic response function. As such the signal is always 'smoother' than (or as smooth as) the response function. Clearly the optimum smoothing kernel is related to the hemodynamic response function. 'Optimum' refers to the kernel that maximizes variance in the signal frequencies relative to other frequencies. It is generally accepted that the hemodynamic response function has an associated delay and dispersion of about 6-8 seconds. This corresponds to a Gaussian kernel of width $\sqrt{6} = 2.45$ to $\sqrt{8} = 2.83$ seconds and it is this sort of kernel that is usually applied to time-series with short (< 6 seconds) TR.

Intersubject Averaging

In many fMRI experiments it is likely that several subjects will be studied, and this raises the question of how best to proceed with data analysis. As there are no limits on the volume of data that can be acquired in a single subject (unlike PET), it is possible to perform an experiment as a series of single case studies. The resultant statistical parametric maps can be displayed directly as an anatomical rendering of that subject's brain. Localisation of function is therefore undertaken by referring directly to the surface anatomy. Such a solution characterises the data most fully in terms of both the similarities and differences between the functional anatomy of each subject. Reporting of such data poses problems; a qualitative description of areas activated in common across subjects is necessary rather than a quantitative comparison between subjects. The comparison of results between different groups becomes more difficult as results are not reported in a standard anatomical space.

The ability to normalise fMRI data into a common stereotactic space, such as that of Talairach and Tournoux allows the reporting of significant activations in a common reference space. However the deformations introduced by single subject normalisation may compromise anatomical resolution. Spatial normalisation also allows intersubject averaging of data. However, the intersubject variability of gyral anatomy is significantly greater than the anatomical resolution of fMRI and so it is likely that a significant degree of spatial smoothing will be required to detect areas activated in common across subjects. This inevitably compromises the spatial resolution of fMRI, one of its most attractive features.

It is likely that the use of these related approaches will depend on the nature of the question being studied. For example, studies requiring high anatomical resolution may rely more on single subject analyses.

Movement-related effects

Imaging the human brain with functional MRI relies on detecting small signal changes in the presence of very large baseline signals. However changes in signal intensity can also arise from head motion and this represents one of the most serious confounds in fMRI studies. Despite restraints on head movement willing and cooperative subjects still show displacements of up to a millimetre or so. With very young, old, ill or disturbed subjects head restraints may not be appropriate. In such circumstances head movements of several millimetres are not uncommon. Movement-related components in fMRI time-series can arise from:

1) *Differences in the position of the object in the scanner* Spatial variation in sensitivity will render the signal a function of the object's position at the time of scanning. This spatial variability can include large scale field inhomogeneity or can be expressed at a much finer scale. An important example of the latter is found in slice-selective irradiation, used for example in multi-slice acquisition. The degree to which spins are excited in any small volume of the object will depend on an interaction between the local magnetic field and the Fourier transform of the slice-selective pulse. For example the excitation of spins in a small region on the edge of a slice will be exquisitely sensitive to small displacements in and out of that slice. In other words signal intensity will be a strong function of position relative to the volume excited or the scanner's frame of reference.

2) *Differences due to the history of the position of the object*. If the number of excited spins is a function of position in the scanner it follows that the number of excited spins (and implicitly the signal) will also be a function of position in previous scans. This dependency is due to changes in saturation of spin magnetization, that is a function of the number of spins excited in the previous scan. This excitation will in turn be a function of position and so, by induction, a function of all previous positions. In summary the current signal is a function of current position and the spin excitation history. The spin excitation history is in turn determined by the history of past movements. This effect will manifest if the recovery of magnetization in the z direction is incomplete by the time the next slice-selective pulse arrives (i.e. if TR is comparable to T_1 which is certainly the case for many fMRI studies). In general movement within the plane of the slice will not change the set of spins excited and should not contribute to this effect.

In summary, movement-related effects can be divided into those that are some function of position of the object in the frame of reference of the scanner, and a component that is due to *movement in previous scans.* This

second component depends on the history of excitation experienced by spins in a small volume and consequent differences in local saturation. The spin excitation history will itself be a function of previous positions. This suggests an autoregression-moving average (ARMA) model for the effects of previous displacements on the current signal (Friston *et al* 1996). This spin excitation history is important because it suggests that realignment alone is not sufficient to deal with all movement-related effects. Friston *et al* (1996) have described an approach which entails; (i) estimating the movement parameters by comparing each scan in the time series to a reference scan, as described in the chapter on spatial transformation, (ii) realigning the time-series using these parameter estimates and (iii) adjusting the values of each voxel by removing any component that is correlated with a function of movement estimates, obtained at the time of the current scan and the previous scan. This function is based on first order ARMA model of movement-related effects.

Friston *et al* (1996) showed that in an extreme case as much as 90% of the variance in an fMRI time-series can be accounted for by the effects of movement after realignment and that this component can be successfully be removed. A danger with this approach is in the event of stimulus correlated motion, where true signal correlations are removed from the time series.

V. FMRI hardware and experimental environment

The design of commercially available fMRI scanners can constrain experimental design. The Siemens Vision system can acquire 128 measurements during a single run before pausing for data processing. A measurement can be any number of axial slices (64 slices representing whole brain acquisition) at variable repeat times; nevertheless data acquisition is likely to be discontinuous within an experiment. This may be significant for paradigms studying sustained physiological or psychological phenomena.

At short TR (repetition time), T_1 relaxation effects become prominent in the early measurements of a time series. If TR is chosen such that significant saturation occurs during the time series (likely for any TRs less than 10 seconds) then a number of dummy scans (typically 2-8) should be acquired prior to those in the experiment. This will establish an equilibrium magnetisation by the time the experiment begins and will improve the re-alignment and statistical analysis. We therefore usually discard the first 2-8 scans in each scanning block prior to realignment and data analysis.

The scanning environment

In contrast to PET, the fMRI environment embodies physical constraints which may restrict the stimulus presentation and subject's response. In addition to the limitations discussed below, it should be noted that access of the experimenter to the subject during the experiment is limited, as the subject needs to be placed completely inside the bore of the magnet.

The bore of the magnet is wide enough to just enclose the subject. The subject's head is enclosed within the head coil so no head movement is possible. A mirror mounted above the subject allows viewing of projected visual stimuli. Although the bore is restricted, it is possible to produce a wide field of view by placing the projection screen close to the subject (an apparent distance of 40cm). This allows a field of view of greater than fifty degrees. Arm movement at the shoulder is possible but restricted and may compromise head immobilisation.

Customised apparatus must be made for the recording of responses within a magnetic environment. Recording of EEG and EMG presents special technical problems.

The gradient switching required by echo-planar imaging can create high levels of auditory noise. By means of careful acoustic engineering, this noise can be very substantially reduced, but this is not normally found in commercial MRI equipment. Despite the use of shielded gradient coils, the ambient noise levels within the bore of the magnet during imaging may exceed 90dB. Nevertheless, stimulus delivery through appropriately isolated

earpieces make presentation of auditory stimuli such as words feasible. It has proved possible to detect activation in both primary auditory cortex and auditory association areas using these techniques. However noise may still present a limitation to certain types of experiment. The noisy environment also makes communication with the subject difficult whilst imaging is in progress.

References

Bandettini, P.A. (1993) MRI studies of brain activation: temporal characteristics. In: Functional MRI of the brain. Berkeley: SMRM, California 143-151

Bandettini, P.A., Wong, E.C., Jesmanowicz, A., Hinks, R.S. and Hyde, J.S. (1994) Spin echo and gradient echo EPI of human brain activation using BOLD contrast: A comparative study at 1.5 T. NMR in Biomedicine**, 7,** 12-20

Bandettini, P.A., Davis, T.L. et al (1995) Proc. **3rd** SMR, 453

Blamire , A.M., Ogawa, S., Ugurbil, K., Rothman, D.L. et al (1992) PNAS **89**, 11069

Boxerman, J.L., Bandettini, P.A., Kwong, K.K., Baker, J.R., Davis, T.L., Rosen, B.R. and Weisskoff, R.M. (1995) The intravascular contribution to fMRI signal change: Monte Carlo modeling and diffusion weighted studies *in vivo*. Mag. Res. Med. **34**, 4-10

Frahm, J., Merboldt, K.-D. and Hanicke, W. (1988) Direct FLASH MR imaging of magnetic field inhomogeneities by gradient compensation. Mag. Res. Med. **6**, 474-480

Frahm, J., Merboldt, K.-D. and Hanicke, W. (1993) Functional MRI of human brain activation at high spatial resolution. Mag. Res. Med. **29** 139-144

Frahm J, Merboldt K-D, Hanicke W, Kleinschmidt A and Boecker H (1994) Brain or vein - oxygenation or flow? On signal physiology in functional MRI of human brain. NMR in Biomed. **7**, 45-53

Kruger, G., Frahm, J. et al (1995) Proc **3rd meeting** SMR, 454

Kwong, K.K., Belliveau, J.W., Chesler, D.A., Goldberg, I.E., Weisskoff, R.M., Poncelet, B.P., Kennedy, D.N., Hoppel, B.E., Cohen, M.S., Turner, R., Cheng, H.-M., Brady, T.J. and Rosen, B.R. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc. Natl. Acad. Sci., USA **89**, 5675-5679 (1992)

Friston, K., Jezzard, P. and Turner R (1994) Analysis of functional MRI time series. Human Brain Mapping **1**, 153-171

Friston, K., Frith, C.D., Turner, R. and Frackowiak, R.S.J. (1995a) Characterising evoked hemodynamics with fMRI. NeuroImage **2**, 157-165

Friston, K, Frith, C.D., Turner, R., and Frackowiak, R.S.J. (1995b) Characterising dynamic brain responses with fMRI: A multivariate approach. NeuroImage **2**, 166-172

Friston, K., Holmes, A.P., Worsley, K., Poline, J-B, Frith, C. and Frackowiak, R.S.J. (1995c) Statistical parametric maps in functional imaging: A general linear approach. Human Brain Mapping **2,** 189-210

Friston K, Williams S, Howard R, Frackowiak R S J and Turner R (1996) Movement related effects in fMRI time series. Mag. Res. Med. **35**, 346-355

Guilfoyle, D.N. et al (1989) Magn. Res. Med. **10**, 282-287

Haacke, E.M., Hopkins, A., Lai, S., Buckley, P., Friedman, L., Meltzer, H., Hedera, P., Friedland, R., Klein, S., Thompson, L., Detterman, D., Tkach, J. and Lewin J.S. 2D and 3D high resolution gradient echo functional imaging of the brain: venous contributions to signal in motor cortex studies. NMR in Biomed. **7**, 54-62 (1994)

Howseman, A.M., Josephs, O., Porter, D., Muller, E., Frackowiak, R.S.J. and Turner, R. (1996) Sustained activation in visual cortex using EPI and FLASH fMRI. **2nd meeting** Functional Mapping of the Human Brain, Boston

Jezzard, P. and Balaban, R.S. (1995) Correction for geometric distortion in echo-planar images from B_0 field variations. Mag. Res. Med. **34**, 65-73

Josephs, O., Turner, R. and Friston, K. (1997) Event related fMRI. Human Brain Mapping (submitted)

Malonek, D. and Grinvald, A. (1996) Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: Implications for functional brain mapping. Science **272**, 551-554

Mansfield, P. (1977) Multi-planar image formation using NMR spin echoes J. Phys. C **10**, L55-58

Menon R S, Ogawa S, Tank, D W and Ugurbil K (1993) 4 Tesla gradient recalled echo characteristics of photic stimulation induced signal changes in the human primary visual cortex. Mag. Res. Med. **30**, 380-386

Menon, R.S., Hu, X., Adriany, G., Andersen, P., Ogawa, S. and Ugurbil, K. (1994) **2nd meeting** SMR, 622

Ogawa, S., Menon, R.S., Tank, D.W., Kim, S.-G., Merkle, H., Ellerman and Ugurbil, K. (1993) Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging: A comparison of signal characteristics with a biophysical model. Biophysical Journal **64**, 803-812

Turner R., Le Bihan, D., Moonen, C.T., Despres, D. and Frank, J. (1991) Echo-planar time course MRI of cat brain oxygenation changes. Mag. Res. Med. **22**, 159-166

Weisskoff R.M., Boxerman, J.L., Zuo, C.S. and Rosen, B.R. (1993) Endogenous susceptibility contrast:Principles of relationship between blood oxygenation and MR signal change. In: Functional MRI of the brain. Berkeley: SMRM, California 143-151