Preprocessing and Analysis of Functional MRI Data

Matthijs Vink
2005, version 1.4

Rudolf Magnus Institute of Neuroscience
University Medical Center Utrecht
Department of Psychiatry
Section Functional Neuroimaging
M.Vink@azu.nl
www.matthijs-vink.com

The following people have contributed significantly to this manuscript:

Martijn van den Heuvel, MSc.
Mathijs Raemaekers, PhD.
Rene Mandl, MSc.
Bas Neggers, PhD.
Leila Kushan, BSc.
Nick Ramsey, PhD.

1 Introduction

In this reader, we will discuss all the steps which are typically performed when analysing an fMRI data set. We will begin by briefly explaining how fMRI works and provide some information regarding the fMRI signal itself. In Chapter 2, all the steps of preprocessing of the raw fMRI data needed for statistical analysis are described. In chapter 3, a brief background is given on the characteristics of the fMRI signal (i.e. BOLD signal), and on experimental design. We will discuss two major fMRI designs; blocked and event-related designs. Chapter 4 describes the most common statistical analysis method for individual fMRI data sets; multiple-regression. First, the components which make up the fMRI signal are discussed and explained. Next, the concept of multiple-regression is described. In chapter 5, two main statistical issues are addressed. First, we will explain multicollinearity and how it may affect the final results of your analysis. Second, the issue of thresholding statistical maps is discussed. In chapter 6, the analysis of fMRI data from a group is described. We will discuss both the standard group analysis as well as a so-called Region-of-Interest and Volume-of-Interest approaches.

Throughout the reader, you will come across gray text blocks titled 'Exploration'. These are merely included to provide some background for interested readers. These texts are presented separately, as they are not crucial for the understanding of the issues discussed in the standard text.

Keep reading even when you come across an issue that you do not immediately understand. Most issues discussed in this reader are linked to each other, and are therefore discussed repeatedly.

1.1 Background on fMRI

Functional MRI images are obtained using an MRI scanner, which basically is a large magnet with a field strength of around 1.5 to 4 Tesla. When a subject is placed inside the scanner (i.e. magnetic field), a slight majority of the protons within that subject will tend to align with the field (B_0). The signal measured by an MRI-scanner is based on the emission of electromagnetic radiation from the nuclei of these protons (hydrogen atoms), which are excited by a radio frequency (RF) pulse (see 2.2 IMAGE (RE)CONSTRUCTION). After excitation, the protons will fall back into their normal state. During this relaxation, energy is emitted mainly because of collisions with other protons. The time it takes for hydrogen atoms to lose 67 percent of their energy is called the 'T2' relaxation time. As this relaxation occurs within a very strong magnetic field, some protons will radiate an RF pulse in phase with each other, which can be detected by a receiver coil. This coil is placed as close as possible to the head of the subject, to minimize signal loss. The signal measured by this coil decays

much faster than would be expected solely on the T2 characteristics. This increased decay of the signal is the result of dephasing of the protons caused by microscopic regional differences in the magnetic field. This signal is called T2*, and is the basis of the BOLD-fMRI signal.

1.2 What is measured with BOLD-fMRI ?

Functional MRI does not measure neuronal activation itself, but rather a derivative. This derivative, in essence, is formed by the regional increase in the amount of oxygen. In 1990, pioneering work of Ogawa et al. (Ogawa et al. 1990a;Ogawa et al. 1990b) and Turner et al.(Turner et al. 1991) demonstrated that the MR signal in the vicinity of blood vessels and in perfused brain tissue decreased with a decrease in blood oxygenation. This type of physiological contrast was coined 'blood oxygenation level dependent' (BOLD) contrast by Ogawa et al. (Ogawa et al. 1990a). Presently, most fMRI images are constructed using this BOLD contrast.

When neurons in the brain become active, the amount of blood transported to these neurons is increased. As a consequence, both regional cerebral blood flow (rCBF) as well a regional cerebral blood volume (rCBV) is increased. The increase in blood flow supplies an increase of oxygenated hemoglobin that largely exceeds the regional oxygen consumption. Because oxygenated hemoglobin is diamagnetic (i.e. it exerts a little effect on the regional magnetic field) and deoxygenated hemoglobin in paramagnetic (i.e. it disturbs the regional magnetic field), a relative increase of oxygenated hemoglobin will reduce local instabilities in the magnetic field at the site of the neuronal activation. As a result, the BOLD signal is slightly stronger at sites of activation, which leads to an increase in image intensity.

2 Preprocessing

2.1 Introduction

Raw data obtained from an MRI scanner requires several manipulations to allow statistical analysis on these data. These manipulations include the following steps:

STEP 1 → *Image (re)construction* of the raw data obtained from the scanner
STEP 2 → *realignment* of all functional images to the same orientation and position
STEP 3 → *coregistration* of the T1 weighed image (anatomy) with the functional images

STEP 4 → *normalization* of all images to a standardized size, orientation, and position (MNI space)
STEP 5 → *smoothing* of the functional images by convolution with a Gaussian kernel

Normalization and smoothing are typically only applied when a group analysis is performed.

2.2 Image (re)construction

In this section, we will present a very brief overview of how an MRI image is constructed and how it is transformed into a picture of the brain. This overview is by far not complete and is a very coarse generalization, but hopefully will provide you with some background on (f)MRI.

2.2.1 Constructing an MRI image

An MRI scanner does not make pictures of the brain, but rather collects information of the brain in terms of frequency and phase of spinning particles (i.e. protons) in human tissue. When a subject is placed in the MRI scanner (i.e. a strong magnet), electrically charged parts in human cells called protons align along this main magnetic field (i.e. B_0-field). Due to this alignment all the charged particles point in the same direction, resulting in a net magnetization in the direction of the B_0-field. When one makes an MRI scan a radio frequency (RF) pulse is send through the brain, the net magnetization flips, resulting from the fact that the spinning protons are tipped out of this B_0-alignment.

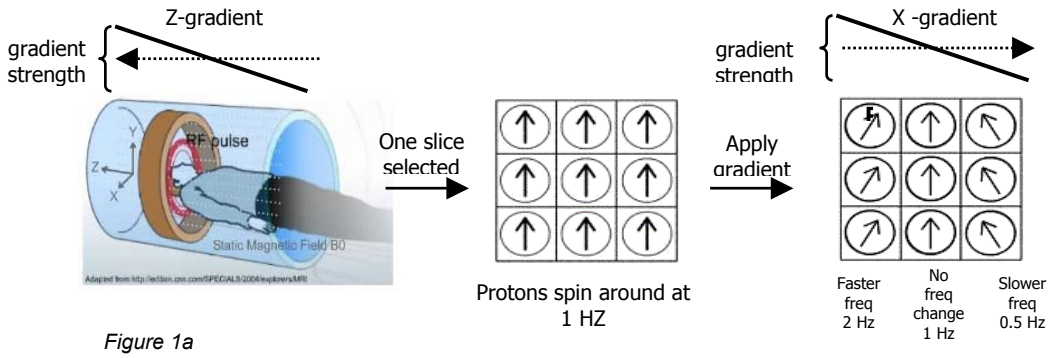


Figure 1a

When the protons return to this B_0-alignment, energy is radiated which is collected by a read-out coil placed around the head of the subject. When such an RF pulse is applied in combination with a linear magnetic field in the length of the scanner (called a gradient), then only those protons that lie at a specific place in the linear gradient will flip, enabling the selection of a specific slice.

Then as two other gradients are applied in the other two dimensions (x and y), the individual voxels of the slice can be determined. This results from the fact that the frequency and phase with which the protons spin around changes due to the two gradients. For example look at figure 1a-A, in which a single slice is depicted. First all the protons in the slice spin at the same frequency. If the strength of a gradient in the x-direction increases in a linear fashion (figure 1a-B), then protons at the end will spin faster (i.e. larger frequency) than protons at the beginning of the gradient. As the linear increase in strength of the gradient is known, the location of the protons can be calculated from the frequency of their spin. In turn, the other gradient in the other dimension (y-direction) will do something similar, but only changes the phase of the spinning protons instead of the frequency. When frequency and phase information of the spinning protons is combined, one can reconstruct the slice in 2D. When all slices are scanned one can construct the 3D brain.

2.2.2 Image reconstruction: from k-space to the image

In k-space, all the information regarding the frequency (i.e. the speed of spin) and the phase (i.e. the difference in angle between the protons at a particular time point) in which the protons spin is stored. For every scan you make, there is also a k-space matrix. This is depicted in figure 1b. Typically, K-space has two dimensions (x and y) in an EPI 2D scan, which represent the image information from one brain slice. Using 3D PRESTO, k-space is made up in 3 dimensions and hence reflects the information from the entire brain. The x and y dimension of k-space are not directly related to the horizontal and vertical dimension of the final brain image. Rather, the axes of k-space correspond to the values measured with different encoding steps (ky-axis) and the values during read out over time (kx-axis). To obtain the MRI image of the brain, the k-space matrix of every slice (2D; or from the entire brain 3D) is converted using a Fourier Transform. This process is called *image reconstruction*.

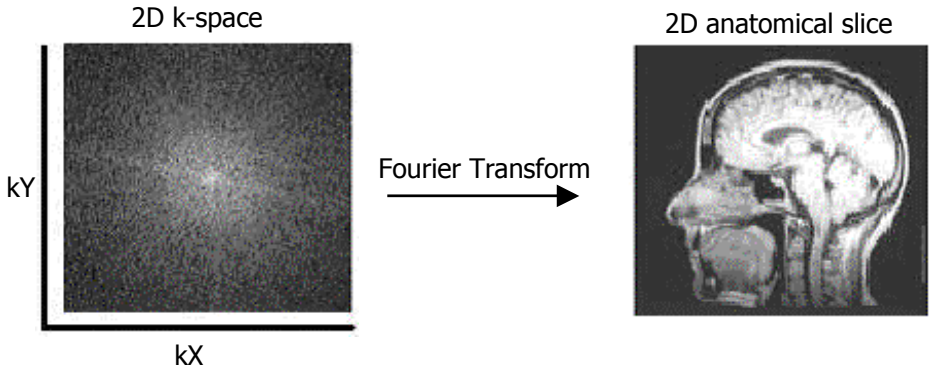


Figure 1b

2.3 Realignment

The goal of realignment is to align all functional images to one specific image. Typically, all functional images are aligned to the first functional image, so that all functional images are in the same orientation and position. Functional images from an fMRI time-series are originally not aligned due to movement of the subject during the fMRI experiment. Therefore, the source of the signal in one voxel can differ between scans (over time), resulting in 'fake activation' (see figure 2), or in a decrease in signal-to-noise ratio (SNR), by an increase in noise within that voxel. A successful realignment procedure of all functional images ensures that the source of the signal in one voxel originates from the same location within each scan. However, some movement artefacts still remain after realignment (see 4.2.4 MOVEMENT RELATED ARTEFACTS for a more detailed discussion).

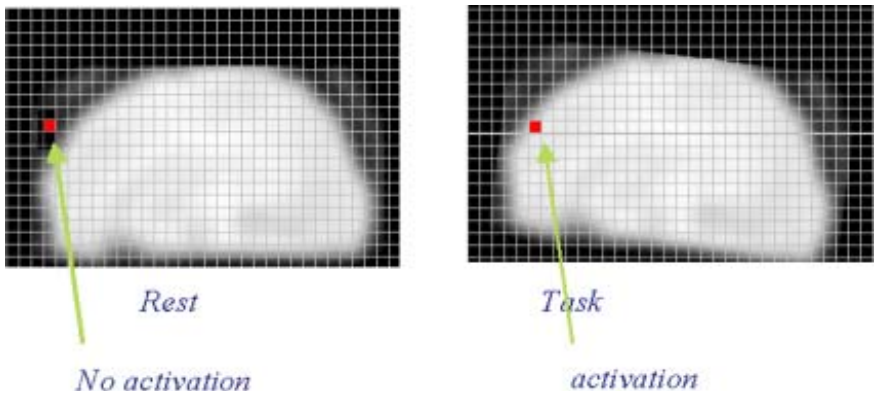


Figure 2 When subjects move their head during task presentation, the signal in the particular voxel will come from brain tissue, while during rest the signal in this voxel originates from outside the brain. Realignment attempts to ensure that the signal in a specific voxel has the same source for the entire time series data.

Realignment is carried out in two steps. First, the parameters needed for the (linear) rigid-body transformation of the images to a user-selected fixed image are determined. This image can be any one of the functional images, but typically is the first functional image. The transformation is a so-called rigid body transform, meaning that 'the size of the brain' is kept constant. There are three rotations (over x, y, and z -axis), and three translations (left-right, up-down, and forward-backward), making a total of 6 parameters which fully describe the movement of the head over time. Second, the parameters are applied to the functional images. To obtain the new voxel values, resampling of the data is required, involving some form of interpolation of the data points. This is depicted in figure 3.

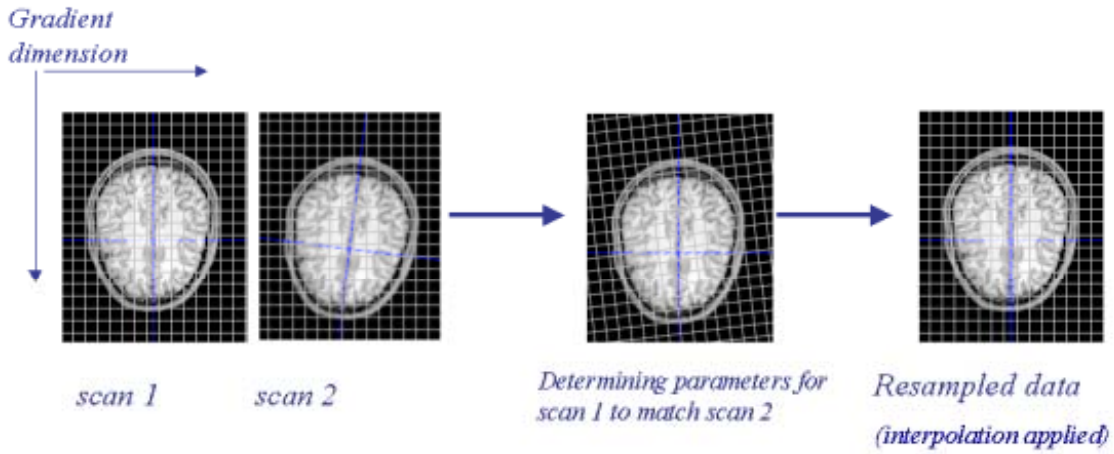


Figure 3 Example of why resampling of the data is needed during image realignment.

2.4 Coregistration

The goal of coregistration is to align the functional images with the anatomical image, so that the activation is superimposed onto the correct anatomical location. There are two methods one can use: (a) segmentation of the images and subsequent matching of the separate segments, or (b) mutual information, which should be used if the images have a different modality. That is, if one image is an anatomical image (T1 weighted) and the other for example a SPECT image or a functional scan (T2* weighted).

2.3 EXPLORATION

The rigid-body parameters are parameterised by:

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & X_{translation} \\ 0 & 1 & 0 & Y_{translation} \\ 0 & 0 & 1 & Z_{translation} \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{Translation}} \times \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\Phi) & \sin(\Phi) & 0 \\ 0 & -\sin(\Phi) & \cos(\Phi) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{Pitch}} \times \underbrace{\begin{pmatrix} \cos(\Theta) & 0 & \sin(\Theta) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\Theta) & 0 & \cos(\Theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{Yaw}} \times \underbrace{\begin{pmatrix} \cos(\Omega) & \sin(\Omega) & 0 & 0 \\ -\sin(\Omega) & \cos(\Omega) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{Roll}}$$

The mutual information method maximizes the mutual information in the 2D histogram in a limited number of iterations. In other words, the difference between the two images is minimized. Because there are a limited number of iterations, it is important in SPM99/2 that prior to coregistration, the images are in approximately the same location.

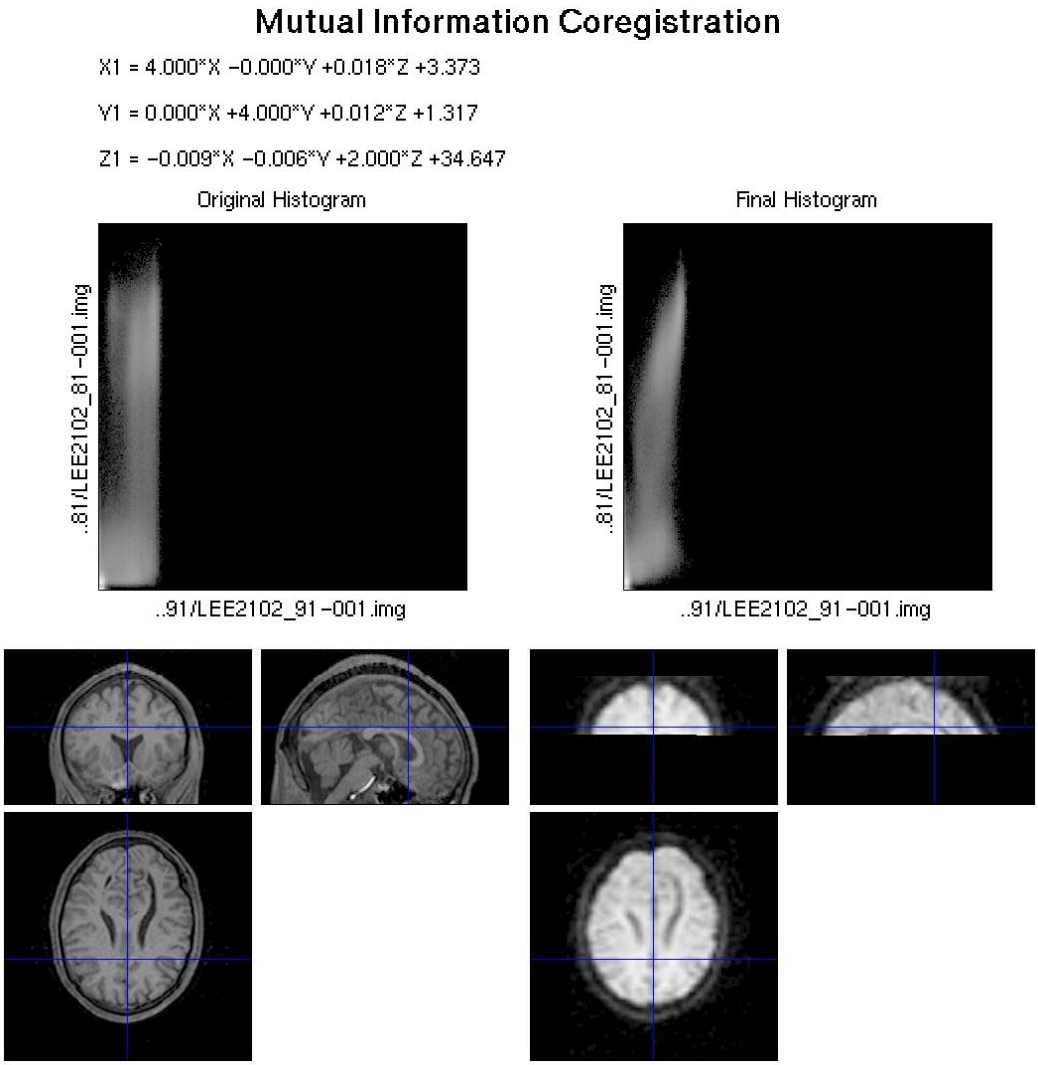


Figure 4 Example of SPM99 output for mutual-information coregistratoin.

2.5 Normalization

The brain of every individual is different. For an individual analysis, in which one is only interested in the regions that are active due to a task in a particular subject, this is no problem. However, when performing a group analysis it is essential that all brains are of the same size and orientation. During normalization, the images are warped so that functionally homologous regions from different subjects are as close together as possible. However, there is no exact match between function and structure (cf. the use of Brodmann maps), and in addition the structure itself differs between subjects (for example, the cingulate cortex in some subjects consists of two but in other subjects of three gyri). To somehow correct for these factors, the image is blurred (see Smoothing). Here, we will discuss how normalization is performed.

In the first step, a template brain is selected. This can be either a template as shown in figure 5 or any linear combination of templates. A template is typically in MNI space (Montreal Neurological Institute). This template is used worldwide, so once your images are in MNI space, you can compare your results per coordinate with results from all other institutes. The second step in normalization involves the minimalization of the sums of squared differences between the template brain and the original brain, and also the squared number of standard deviations away from the expected parameter values. In contrast to realignment, which is a rigid-body transformation (i.e. the size of the brain is kept constant), normalization involves also changing the size of the brain using a linear 12 parameter registration (called an affine transformation) to match size and position of the template (see figure 6). By masking the original image non-brain voxels are deleted and hence cannot affect this affine registration. Such a brain mask can also be applied when normalizing lesioned brains, by masking out the lesion so that it is not used in the normalization process (the MNI template has no lesions). A possible third step involves a global non-linear warping of the original brain to match the template (SPM). This non-linear warping is based on a Bayesian framework to simultaneously maximize the smoothness of the warps

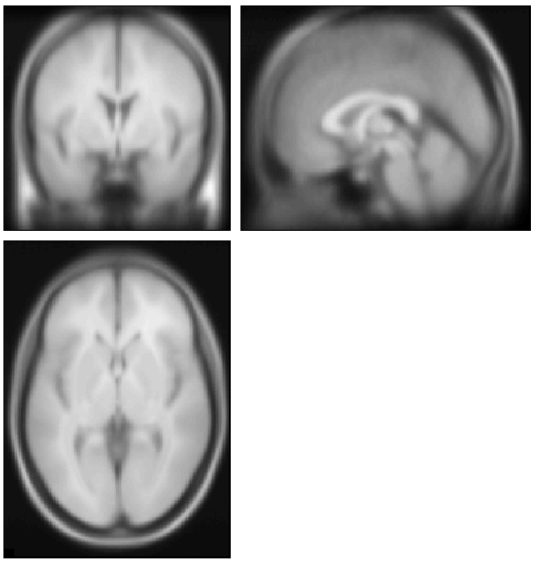


Figure 5. The montreal standard brain

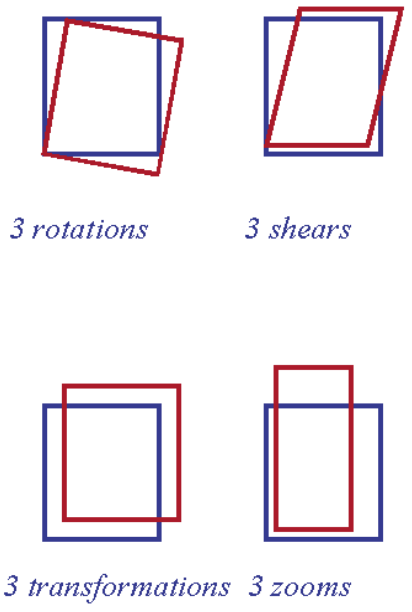


Figure 6. Schematic representation of the transformations applied during normalization. All translations are possible for X, Y, and Z direction

In figure 6b, some examples of normalised brains are presented. As you can see, the brains are not similar even after normalisation. Another step in preprocessing is required to increase the overlap between subjects.

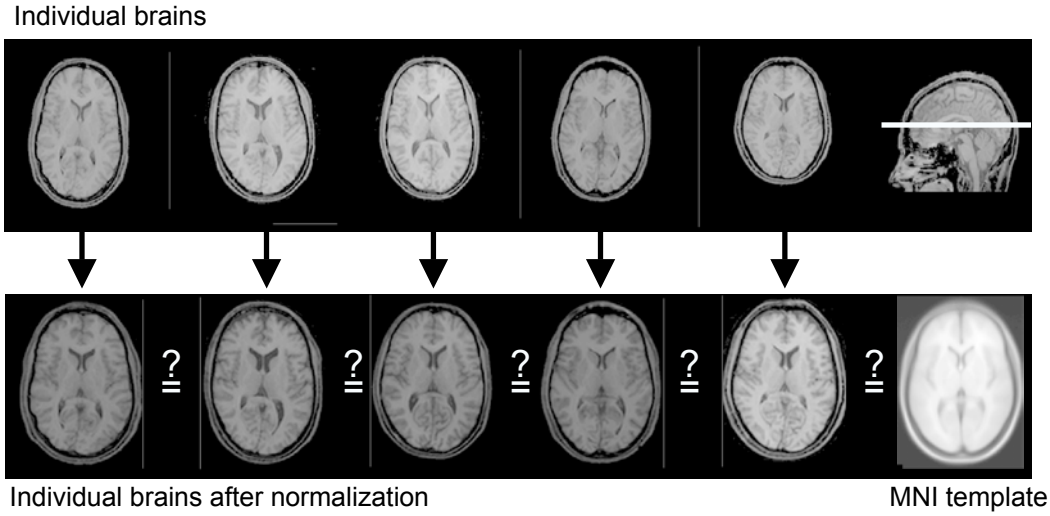


Figure 6b

2.6 Smoothing

Smoothing involves blurring *functional* MRI images using (typically) a Gaussian filter (i.e. data is convolved with a Gaussian kernel; see figure 8). Smoothing, in practice, is only applied when a group analysis is performed. By smoothing the image, the overlap of activation between subjects is increased.

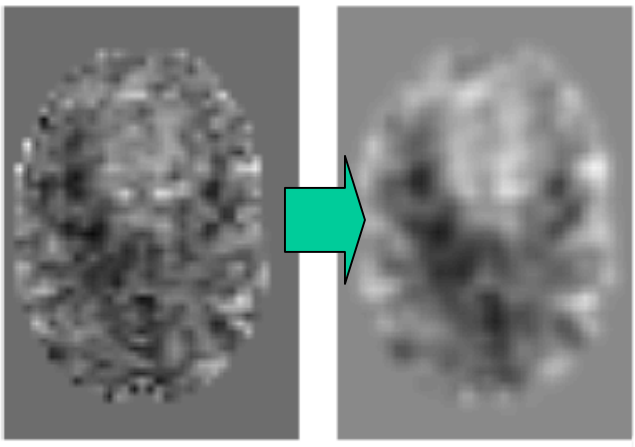


Figure 7. Blurring of a t-map with a Gaussian kernel (FWHM = 8 mm)

When one blurs (smooths) an image, each voxel effectively becomes the result of applying a weighted region of interest (ROI; the voxels under the kernel). The size of this kernel (see figure 7) is determined by the full width at half maximum (FWHM). The FWHM is an indication of the distribution of the kernel values, meaning that when the FWHM is 8 mm, the kernel is 8 mm wide at 50 percent of its peak value. The voxel falling within the range defined by the FWHM receive the highest weights, while the voxel falling outside this range receive lower weights. In the ideal case, a FWHM kernel size is chosen so that it matches the size of the expected activation. For example, if the activation is an area of 10 mm^3, then

the FWHM should be 10. If a FWHM of 20 is chosen, then the activation in that activation area is averaged with the activation in the surrounding voxels. If these are not significantly active, then the overall activation in that region (which is now $20mm^3$) is reduced compared to the ROI of $10mm^3$. Typically, a FWHM size of twice the voxel size is chosen.

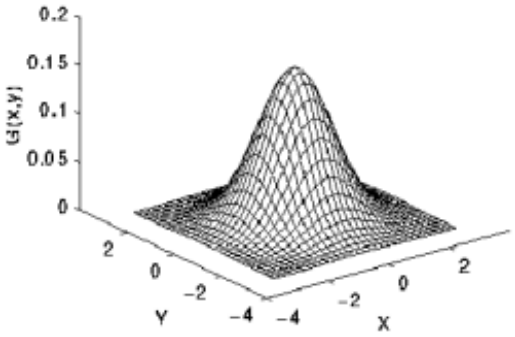


Figure 8. 3D representation of a Gaussian kernel

How is the data convolved with the Gaussian kernel? Figure 8 depicts a 3D representation of the kernel. This can be rewritten in a 2D form (figure 9). The data (middle matrix in figure 9) is then convolved with this 2D Gaussian kernel. As an example, of how the smoothed values are obtained, we will describe this process for the center voxel (which now has value 5) . In figure 9, on the left, the 2D Gaussian kernel is depicted. This kernel is applied to the data (e.g. from a t-map or functional scan) presented next to the kernel in figure 9. To calculate the new value for the center voxel in the data, multiply each voxel value with the corresponding value of the Gaussian kernel and divide by 273 (which is the total numeric value below the Gaussian curve). This division is done so that the data are only scaled, but not altered in any other way. The value of

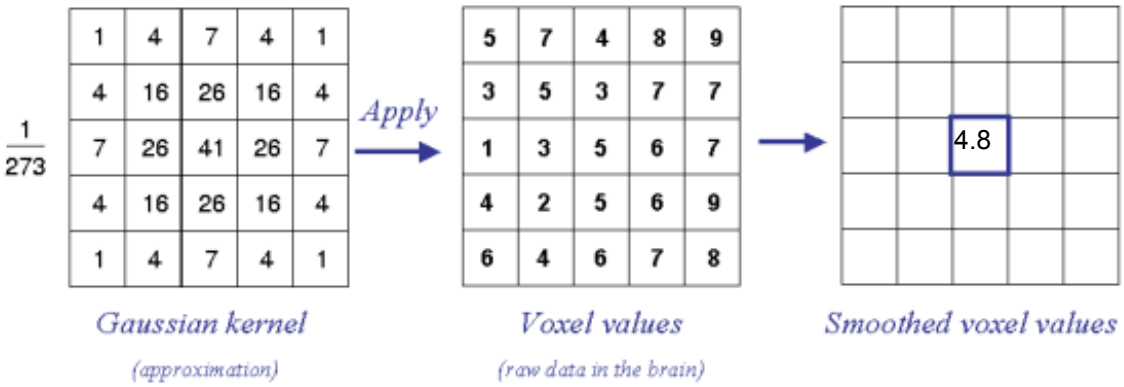


*Figure 9. 2D representation of a Gaussian kernel. The value of the smoothed voxel in the middle is calculated by: Value middle voxel: ((1 * 5)/273) + ((4 * 7)/273) + ((4 * 7)/273) + ... + ((1 * 8)/273) = 4.8242*

the center voxel in the smoothed image effectively depends for $\frac{41}{273}$ part upon it's own old value, and for $\frac{273-41}{273}$ part upon the values of it's neighbouring voxels. After the new value of the center voxel is calculated, the kernel is

moved one voxel to the left (or up), and the procedure is subsequently repeated for all voxels in the three dimensions.

After smoothing the image, the number of voxels and the size of the voxels (i.e. the sampling rate) remains the same. However, the resolution of the image becomes less (see figure 7). Whereas in the not-smoothed image, the resolution was the same as the voxel size (for example 4 mm), the resolution of the blurred image is specified in terms of *resels* (i.e. *res*olution *el*ement*s*). A resel consists of a number of voxels that fall within the FWHM. For example, if the FWHM is 8 mm (3D) and the voxels are $4 \times 4 \times 4$ mm, then a resel consists of $2 \times 2 \times 2$ = 8 voxels. The number of resels is then the number of voxel divided by 8. Compared to voxels, these resels are a better estimation of the number of independent observations (see 5.3 THRESHOLDING for the implications).

There are three reasons for smoothing: *(1) Increase the signal-to-noise ratio (SNR)*. Because the signal in a voxel in a smoothed images originates not only from the voxel itself but also from it's neighbouring voxels, the effect of random (uncorrelated) noise is reduced. *(2) Increase inter-subject overlap*. After normalization, the brains are in MNI space, but there may be slight differences between subjects in the relationship between function and structure. By averaging the signal over a larger area, the overlap between activation spots between subjects (possibly) increases. *(3) Increase validity of the analysis*. Neighbouring voxel values in an image are spatially correlated. Recall that the goal of pre-processing is to allow statistical analysis on the data. The analysis we will discuss in chapter 4 is voxel-based, which means that per voxel a test of significance is performed. To assess which voxels are significantly active during for example a task, a threshold needs to be set. In order to determine the height of that threshold, the number of independent tests that are performed is needed. This number is not the same as the number of voxels, since these voxels are correlated (the results of these voxel-based tests are also correlated). An estimation of the number of independent tests that are performed is the number of *resels* in an image. However, these resels are also not completely independent, but are a better estimation of the true number of independent observations than the number of voxels (see also 5.3 THRESHOLDING).

2.6 EXPLORATION

Calculation of the smoothed value

Multiply the value of each voxel with the corresponding kernel value and divide this by the total value of the kernel (= 273). The sum of these values is the smoothed value of the middle voxel:

$$((1 \times 5)/273) + ((4 \times 7)/273) + ((7 \times 4)/273) + \ldots + ((1 \times 8)/273) = 4.8242$$

Convoluting the image with a Gaussian kernel poses a problem on the edges of the image (see figure 10, (Maisog and Chmielowska 1998)). The value of the smoothed voxel in the middle still depends for $\frac{41}{273}$ part on it's old value and $\frac{232}{273}$ part on it's neighbouring voxels. However, around the edges some of these neighbouring voxels are outside the brain and have value 0. A similar problem occurs at borders between white and grey matter. The values of the voxels outside the brain as well as of the voxels in white matter are not correlated with the value of the voxel in the middle, but these values are included in the calculation of the smoothed value for this middle voxel. This results in an underestimation of the smoothed voxel value. To correct for this border-effect, one can either (a) mirror the region, so that the values of the voxels outside the brain (or in white matter) are replaced by the values of the voxels in the grey matter, (b) apply some form of edge truncation so that the voxels which fall outside the brain (defined by a brain mask) do not affect the weighing of the smoothed voxel values, or (c) not smooth these voxels. These choices are typically not made by the user, but are implemented in the various software packages for fMRI preprocessing (such as SPM99/2).

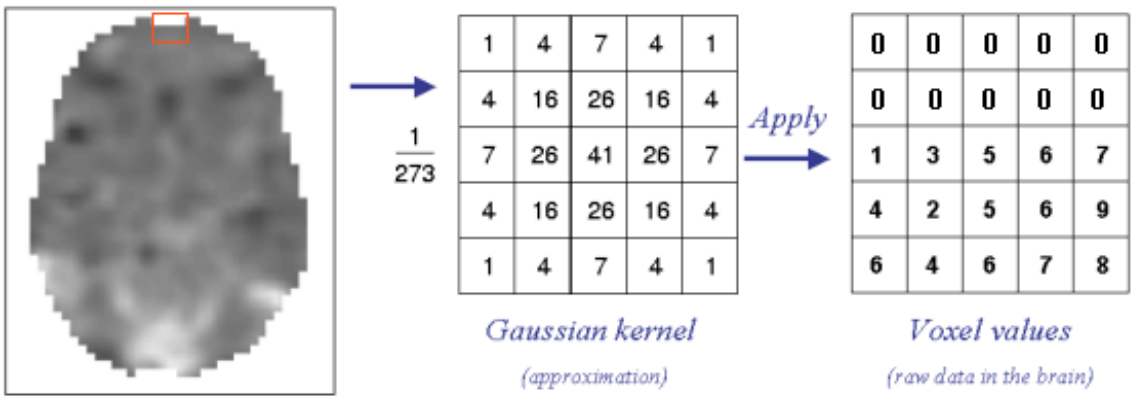


Figure 10. Smoothing around the edges poses a problem

3 Experimental design

3.1 Introduction

In this chapter, we will discuss two main design types used in fMRI research; blocked and event-related (ER) designs. A blocked design refers to a design in which the task is presented in so-called blocks of for example 30 seconds, alternated with periods of rest (rest blocks). The signal that is measured effectively arises from a combined brain response to a number of trials (events). In contrast, the idea behind an event-related design is to obtain the brain response to a single event. In discussing these designs, we will focus on power and efficiency. The *power* of a design refers to how well two conditions can be separated in terms of brain activation (for example condition 1 versus 2, but also condition 1 versus rest). *Efficiency* refers to the amount of data that can be obtained within a fixed time period. First, let's look at the characteristics of the signal (BOLD dependent) fMRI is based on.

3.2 BOLD and HRF characteristics

The signal that is commonly measured using functional MRI is called the BOLD (Blood Oxygenation-Level Dependent) signal (see figure 11). As suggested by the name, this signal depends upon the oxygen-level in the blood rather than on direct neuronal activation. There is, however, a relationship between neural activation and the BOLD signal. Neuronal firing will commence immediately in response to stimulus presentation. This process requires oxygen, which is supplied through the blood. Because immediately after stimulus onset no extra oxygen is supplied, a slight decrease in the signal can be observed. This initial dip quickly transforms into a positive signal when the oxygen supply starts to build up. After approximately six seconds post-stimulus, the signal that is detected with (BOLD dependent) fMRI reaches its maximum (peak). After 20 seconds, the signal from the original stimulus has returned to zero. This BOLD signal is modelled by a haemodynamic response function (hrf; (Friston et al. 1995)), which is a generalized approximation of the actual BOLD curve (figure 14, green line). However, the specific characteristics of the BOLD curve can differ between *(a) brain regions within one subject*, and *(b) subjects*. So, for example, in some regions (or subjects) the actual BOLD response may peak after 5 instead of 6 seconds, or the BOLD response is much wider in some brain regions (subjects). Due to these possible variations, the signal we attempt to describe using a fixed hrf may not be present at all in the data. Then the conclusion may be that there is no significant brain response to a specific stimulus, while in fact there is a response but is differs from the function we use to define a brain response. To prevent such 'false conclusions', a correction is needed to account for inter-region and inter-

subject variability in BOLD responses. SPM99/2 can allow for a correction for both onset time variation and dispersion (i.e. wider or smaller curve) of the BOLD curve, by including a time and a dispersion derivative in the hrf (see figure 12, blue and red line). By including (one of) these derivatives, more of the variation in the fMRI signal is explained (see CHAPTER 4). Alternatively, the derivatives can be used to visualize deviations from the hrf. Timing differences (i.e. time when the peak of activation is reached) between conditions in a particular brain area can be visualized by looking at the time derivative during these conditions.

Brain activation can also be modelled using different functions such as (a) a set of Fourier functions, (b) Gamma functions, and (c) a Finite Impulse function. However, the problem with using input functions which are not based on (at least an estimation of) the true physiological signal lies in the interpretation of the results. Namely, when activation in a specific region is highly correlated with a specific Gamma function, it is difficult to understand the source of this relation in terms of the known BOLD response. Using an input function that is based on the true BOLD signal will yield areas in which the signal behaves in a way described by this function. Since this input function is physiologically valid, brain activation can be explained in terms of the mechanisms underlying the BOLD signal.

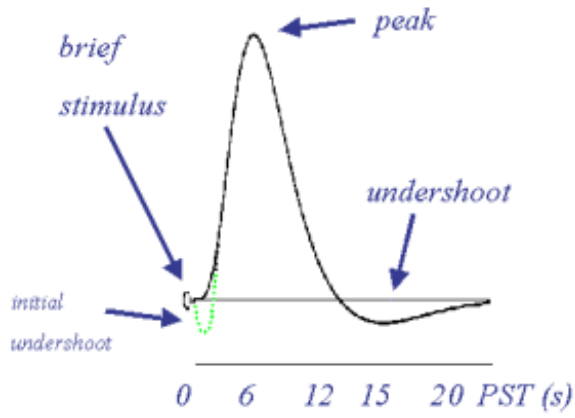


Figure 11. Representation of the BOLD curve (adapted from(Friston et al. 1995)) without permission)
PST = post stimulus time in seconds

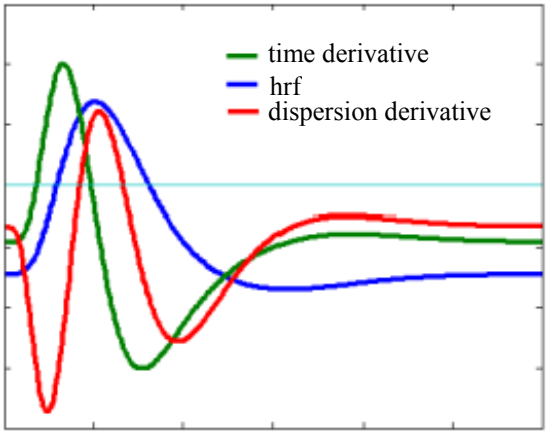


Figure 12. The blue line depicts the HRF which is thought to best characterize the BOLD signal. The other two factors model the differences in onset time of the BOLD signal (green line), and differences in width of the BOLD curve (red line)

3.3 Blocked designs

A blocked design refers to a design in which the task is presented in so-called blocks of for example 30 seconds, alternated with periods of rest (rest blocks). A blocked design potentially has a lot of power (i.e. the ability to differentiate between two conditions), since the brain is either repeatedly doing the task, or doing 'nothing' during rest blocks. The signal associated with performing the task is maximized during task blocks and low during resting periods, thereby maximizing the difference in brain activation during these two conditions. A blocked design is also efficient in that a lot of brain activation is recorded within a fixed time period, because the separate task trials are spaced closely together.

One should consider (at least) two points when designing a blocked paradigm:

(a) Ideally, the number of scans and events should be equal in all conditions, so that the variance in all factors is the same. This means also that the power for all conditions is equal. Consider for example, a task with two conditions and rest periods. Every condition should then consist of 1/3 of the scans.

(b) The length of a block should be between 14 to 20 seconds, yielding a task frequency of 0.036 to 0.025 Hz. However, when there is no accurate description of the BOLD response, then the block should be longer (around 30 seconds; 0.016 Hz, (Aguirre et al. 1998)). Using slightly longer blocks places less importance on how well the rise and subsequent fall at the beginning and the end of the block is described (figure 13 A and B, and see text below). When this description is very accurate, for instance because the actual BOLD response is known, then the blocks can be short. The longer the blocks are, the more chance there is for a correlation with low-frequency scanner-related artefacts (see also 4.2.2 LOW-FREQUENCY NOISE).

A limitation when using a blocked design is that randomization of trials from different conditions is not possible. The idea of a blocked design is to maximize the signal related to a specific task by presenting a number of trials from the same condition in close succession. So although there may be different blocks consisting of different conditions, each block typically consists of only one condition. This poses a problem for some cognitive tasks, for which it is not possible to divide the different conditions into separate blocks. Consider for example a classic paradigm called Go-NoGo. In this task, subjects should press a button when a 'A' appears (i.e. Go trials), but withhold that response when an 'X' appears (i.e. NoGo trial). Activation during the NoGo trials then represents brain activation involved in inhibiting the response. A blocked version of this task might entail blocks of Go trials and blocks of NoGo

trials. However, during blocks of NoGo trials, there will be no inhibition as subjects do not respond at all. NoGo blocks are in fact rest blocks (see also 3.5 REST PERIODS) and will not, in any case, represent inhibition related brain activation.

Another drawback of using a blocked design is that the strength of the brain signal can decrease over time. That is, neurons which fired strongly at the beginning of the block may reduce their firing rate over time due to neuronal mechanisms like habituation. The idea of a blocked design is that presenting multiple trials in close succession maximizes the signal. This BOLD response is modelled by a so-called box-car function (see text below). When the neuronal activation drops (i.e. the BOLD response decreases) due to habituation, then this box-car function is not a good estimation of the neuronal response anymore. Rather, the decrease of neuronal activation should be taken into account. If no such correction is applied, then the model that is used to describe this particular BOLD response is inadequate, leading to bad statistical values (see chapter 4 STATISTICS).

A commonly used function to describe the signal variance in a blocked design is a *box-car* (see figure 13, A), which encodes a '0' for no-task (i.e. rest or another task) and a '1' for task. However, brain regions do not suddenly activate and stop activating in the way modelled by the box-car function, but rather the BOLD response peaks after (approximately) 6 seconds of stimulus onset, and has an undershoot at the end of the block. By convolving the box-car input function with a function which best describes how the BOLD signal rises (e.g. a hrf or haemodynamic response function), a better estimation of the true block-related BOLD signal is obtained (figure 13, B).

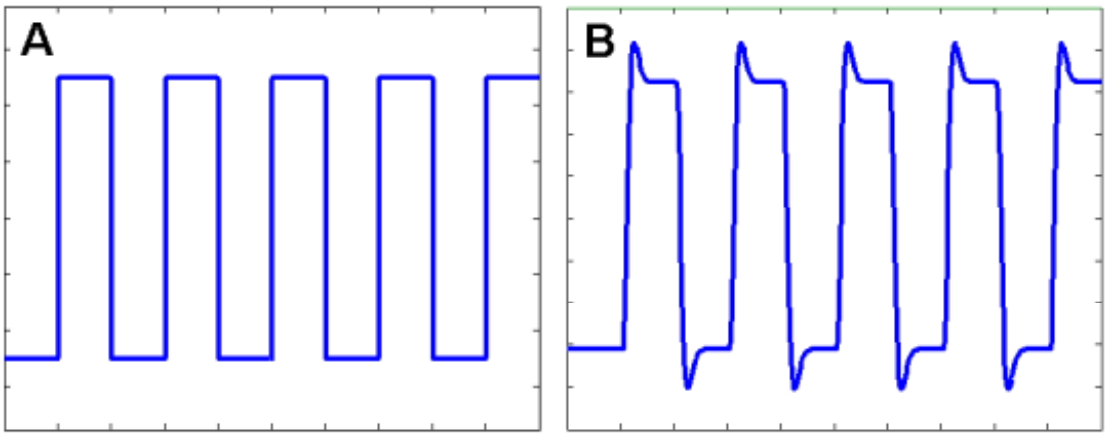


Figure 13. A Box-car input function for a block factor, B same box-car function convolved with a HRF

3.4 Event-related designs

An event-related design is used when you want to obtain the brain response (i.e. BOLD-curve) to a single event. When two stimuli are presented in close succession (for example 1 second apart), the corresponding BOLD curves will overlap, presumably in a linear fashion (meaning that the response is an addition of both curves). If the form of the BOLD curve is known, then these two overlapping curves can be disentangled (alternatively, one can use a hrf to estimate the two BOLD curves). However, the way BOLD curves add up depends upon (a) the distance in time between two stimuli, and (b) the number of consecutive stimuli. When a large number of stimuli (e.g. 20) are presented in close succession (i.e. 1 second apart), the individual BOLD curves cannot be disentangled anymore. The response will reach a plateau with little variance, so that individual BOLD curves originating from individual events cannot be calculated anymore. In an event-related design, in which different trials are randomly intermixed, one would like to determine these event-related BOLD curves. What would be an optimal inter stimulus interval (ISI) for an event-related design? On the one hand, one would like to present as many events as possible in a short time period (i.e. high task frequency). However, then the power (i.e. the ability to differentiate between two conditions) of each task factor (representing the various task conditions) is low. To increase the power, stimuli should be spaced more widely over time (i.e. low task frequency). However, at lower frequencies, noise occurs (such as scanner-drifts, see 4.2.2 and 4.2.3), thereby disturbing the task related signal. Combining these factors (high task frequency with low power, and low task frequency which is affected by scanner noise) suggests an optimum frequency for an event-related design with a fixed inter stimulus interval, namely around 0.05 - 0.08 Hz (a/o. (Bandettini and Cox 2000). This effectively comes down to presenting an event every 12 to 20 seconds.

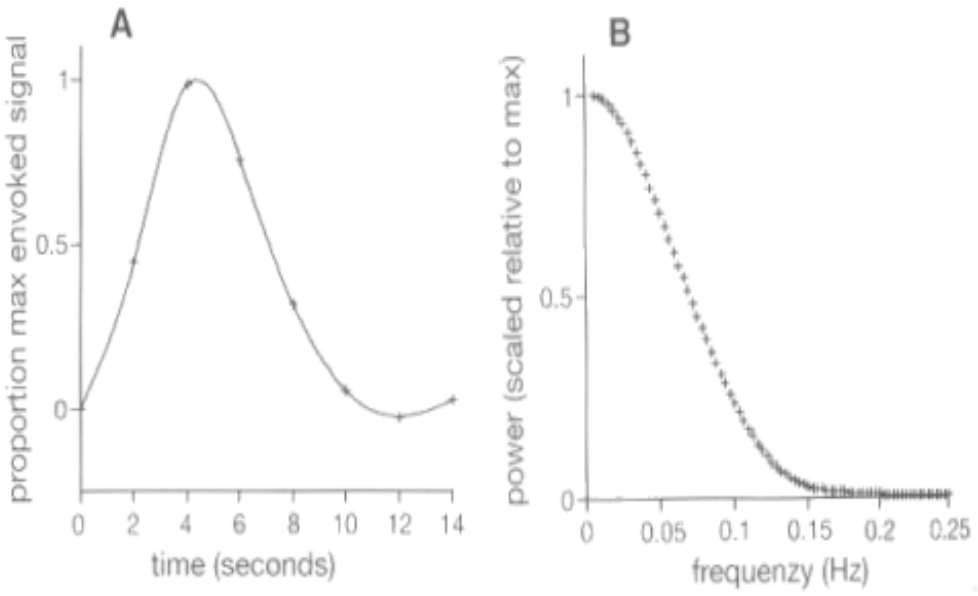


Figure 14. A gives the BOLD curve. B represents the power spectrum of this HRF. Note that at low frequencies power is high, whereas at high frequencies power is low. However, using a low frequency interacts with scanner artefacts (see 4.2.2 Low-frequency noise). The optimal frequency is about 0.05 Hz.

There are alternative event-related designs that allow a shorter inter stimulus interval. Basically there are four types of event-related designs:

1. fixed long inter stimulus interval (ISI) [a/o.(Bandettini and Cox 2000)]
2. variable ISI [a/o.(Dale 1999)]
   variable ISI with occasional small blocks of the same condition (Vink et al. 2005b)
3. fixed short ISI (task 1) randomly interspersed with other trials (task 2) (Buckner et al. 1996;Vink et al. 2005a)

3.4.1 Fixed long ISI
3.4.2 Variable ISI
3.4.3 Blocked variable ISI
3.4.4 Fixed short ISI

3.5 Rest periods

Typically, rest periods should occur throughout an fMRI experiment. These periods serve two main purposes. First, they allow the subject to pause and recuperate from performing the task. Second, the signal in the rest periods serve as a baseline for the fMRI signal. During the rest period, the signal from the scanner is not zero, but rather reflects the signal from the scanner and the subject in rest. All signal variations due to performing a task are defined relative to this baseline. Consider for example a blocked design in which subjects have to press a button for periods of 30 seconds, while during rest periods they have to refrain from responding. The signal from a voxel in the motor cortex will increase during these activation blocks, compared to rest. If the rest periods would be left out, the there will be no signal increase in that motor voxel, as there is no baseline from which activity can differ. Consider another example with two types of blocks. In one block, subjects have to press a button with their left thumb, in the other block they have to use their right thumb. If no rest periods are included, the only signal you can obtain from pressing the left thumb is only relative to pressing the right thumb, as there is no baseline during which no responses are made. So, you will never obtain the true signal related to pressing with the left thumb, but only the signal from pressing with the left thumb compared to pressing with the right thumb. Including rest periods allows you to validate your design, by looking at activation of left thumb presses versus baseline. This is particularly important for studies with patients, in which case you want to be able to check whether these subjects do perform the task correctly by looking a the activation during a particular condition. If rest periods are left out, the you can only observe the contrast between two conditions. This difference may be greater in patients compared to controls, not because of the patients being patients, but because they might not be doing one of the two tasks.

In event-related designs, rest periods are also crucial. Again, subject have to respond with their left thumb. Stimuli (which require a button press) are presented every 8 seconds. Eight seconds after responding, the BOLD-curve (see figure 11) has not returned to baseline yet. By responding every 8 seconds, the signal will never return to baseline. So, if we then want to look at motor activation during responding with the left thumb, we can only obtain the difference between actual responding (peak of the BOLD curve) and the residual activation during that responding. This difference will not be so big as the difference between the peak of the BOLD curve versus a true baseline, when there is no residual activation. Consequently, including a rest period in your design allows you to obtain a true baseline, which is used to estimate the strength of the BOLD signal during activation compared to doing nothing.

4 STATISTICS

4.1 Introduction

The goal of performing an fMRI experiment is to determine which brain activation is significantly associated with a cognitive/physiological process/mechanism. Essential for localising the activation are the realignment and co-registration steps of pre-processing. Spatial normalization and smoothing are only applied when a group analysis is performed (see CHAPTER 6). To determine which activation is significant, a statistical analysis is required. The analysis described here is called multiple-regression. We will begin this section by describing the strategy for such an analysis.

When looking at the signal in a particular voxel over time, one immediately sees that this is not a straight line, but rather goes up and down in a seemingly random fashion (figure 15). In other words, the signal values vary over time. There are several factors, or rather *effects*, that are responsible for this variation in signal values over time (i.e. signal variance). The goal of the statistical analysis is to describe (i.e. to model) these effects and subsequently assign weights to them.

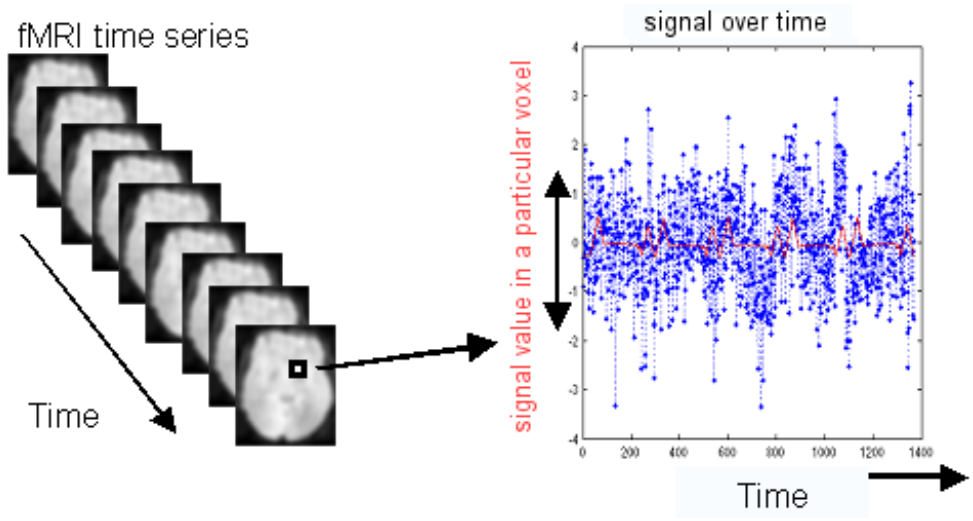


Figure 15 Time series data from a particular voxel

Let's start with an example. A subject is performing a motor task, during which the subject has to press a button for blocks of 30 seconds, while no response is required during rest periods. The signal in voxels which are located in brain regions involved in performing the task (e.g. motor cortex), will increase during the task, while during rest the signal will return to baseline state. In this case, (part of) the variance in the fMRI signal is caused by, and can therefore be described by, a factor which specifies when the task is

performed and when there is a rest period. The variance in the fMRI signal is also affected by various other effects. We can make a distinction between effects of interest (variance due to performing the task), effects of no interest (variance in the signal due to 'known' effects like scanner artefacts), and error.

[3.1] fMRI signal = effects of no interest + effects of interest + error
(variance due to - global effects - low frequency noise - high frequency noise - movement related noise)
(variance due to task)

The idea is to best describe the fMRI signal in terms of effects of interest and effects of no interest. The better you can describe the variance in the signal (i.e. explain the variance), the less random noise you have. As I will describe below, this is important as the main statistical outcome value, representing the significance (i.e. the t-value) of the activation during a particular task or contrast of tasks is in fact a weighted ratio between explained variance versus unexplained variance. The goal during model building therefore is to make a model which explains a lot of variance (i.e. signal variation) so that there remains little variance which is unexplained. More specific for fMRI, you want to have task factors that each explain as much variance as possible, as the significance of the activation is usually calculated for each task factor separately (or a contrast between them). That is, you want to see the activation during a particular task (condition). A raw version of the formula for the t-value is given below and will be discussed in more detail in 4.5 MAKING A T-MAP FOR A SINGLE FACTOR. The t-value is given by:

$$([3.14])\ t = \text{regression coefficient} * \sqrt{\frac{\text{explained variance}}{\text{unexplained variance}}}$$

The regression coefficient reflects how strong the signal modelled by a particular factor is in the data. The explained variance refers to the variance that is *uniquely explained by a particular factor*. The *unexplained variance depends upon the model as a whole*. The more variance the model explains, the less unexplained variance there is. We therefore want to model both the effects of interest (i.e. task related activation) and the effects of no interest.

The following section describes what kinds of effects of no interest are present in the data, and how can we correct for them.

4.2 Effects of no interest

Effects of no interest can be generated by events which you cannot prevent from occurring (e.g. breathing, heartbeat, (head)-motion) or which cannot be controlled (e.g. scanner drifts, scanner instabilities). The goal of including factors that model effects of no interest is to remove the variance these effects generate in the signal. By including these factors, the amount of unexplained variance is reduced. This is important for the t-ratio (see above). However, the t-ratio may be negatively affected when the factors correcting for these effects of no interest are correlated with the task factors. In that case, the amount of uniquely explained variance by the task factors is reduced (see 5.2 MULTICOLLINEARITY).

Basically, the effects of no interest are a source of error, but in contrast to *random* error, these effects can be described by some factor(s). By removing these sources of error by correcting for them, only random error will remain. Randomness of the error is in fact an essential assumption which has to be met when doing a statistical analysis like multiple-regression. In this section, we will discuss four different effects of no interest.

4.2.1 Global effects

The global scaling factor corrects the differences in the entire (i.e. global) image intensity over time within an fMRI time-series. These differences in image intensity are mainly caused by scanner drifts. For example, the scanner warms up over the course of the experiment, causing a overall reduction in signal intensity. The global scaling correction is done by dividing intensity values for each scan by the mean value for all voxels for each scan.

It is implicitly assumed that global intensity (i.e. global volume mean) is not correlated with the task, but is only caused by the scanner instabilities. This might be the case if a very large number of voxels is highly active during the task, and not active at all during rest. In that case, including a factor for global scaling will cause task activation to be scaled down as well (see figure 16).

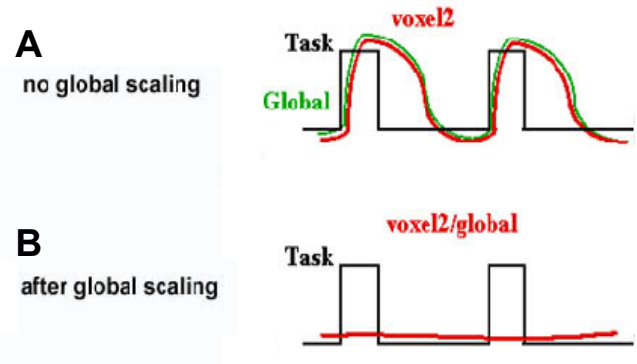


Figure 16. (A) No global scaling: Task-related activation in voxel2 is highly correlated with the global volume intensity. Hence, after correction for the global volume intensity (B), activation in voxel2 is strongly diminished.

In a typical fMRI volume, there are 20,000 voxels (which cover the entire brain). For activation to have a great effect on the global volume intensity, a very large number of voxels needs to be very active during the task. In most cases, it is not likely that activation and global volume mean are highly correlated. Rather, other factors such as scanner drifts are more likely to be responsible for changes in global volume mean. Therefore, including a global scaling factor will generally be beneficial.

4.2.2 Low frequency noise

An MRI scanner, although very expensive, is not stable over time. That means that there might be drifts in the signal, due to this instability. So, in the fMRI time series data, some low frequency signals are likely to be present, which are not due to the task. Removing these signals will decrease the amount of unexplained variance. To remove these signals, SPM99/2 offers the possibility to apply a high-pass filter (i.e. signals with a high frequency may pass) over the data. This high-pass filter consists of a user-specific number of low frequency cosine functions, ranging from a half cosine which describes a linear trend in the fMRI time-series up to the cosine function with the highest frequency. The high-pass filter is constructed using the formula:

[3.3] $$f_r(t) = \cos\left(r\pi \frac{t}{t_N - t_1} \right)$$

The function $f_r(t)$ is applied to an fMRI time series of N scans, acquired at times $t_1,\ldots,t_N$, whereby ***r*** ranges from 1 (which will give a half cosine function describing a linear trend in the data) to a specified number reflecting the highest frequency (see figure 17). It is important that the cosine function with the highest frequency is not correlated with the task factor (see 5.2 MULTICOLLINEARITY). To ensure that the cosine function with the highest frequency does not correlate with the task factor, a cut-off period is calculated to calculate the highest frequency allowed for the high-pass filter. The cut-off period is given by:

[3.4] cut-off period = $(rest + task) \times 2$

(This number should be multiplied by the TR (time for one scan) for SPM99/2)

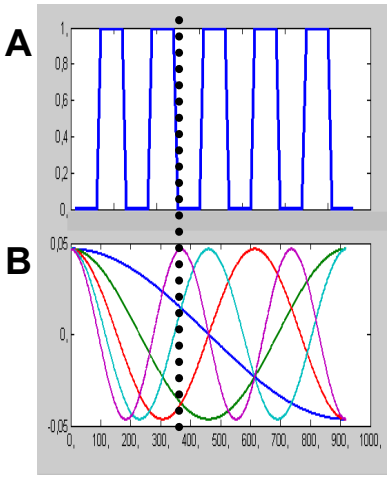


Figure 17. Example of the high-pass filter. (A) represents the task.(B) represents the high-pass filter, with the blue bar as the lowest (i.e. linear trend) and the purple as the highest frequency cosine function. The dotted line depicts the cut-off period.

The maximum value for ***r*** (and hence the highest frequency) is calculated by:

[3.4a] maximum $\boldsymbol{r} = \frac{N}{(\text{cut-off period})}$

where N is the total number of scans and (rest + task) refers to one experimental cycle consisting of one rest period and one task period.

To calculate the most optimal high-pass filter, you can use an SPM extension called Design Magic (http://www.fil.ion.ucl.ac.uk/spm/ext/#D_magic).

4.2.3 High frequency noise

Because fMRI data is in fact data from a time-series, the value in a particular (brain) voxel at a given time t_n is correlated with the value in that same voxel at time t_{n-1} or t_{n+1} (where 'n' is in scans). This fact is important for the way the analysis is performed. Originally, you treat the data as if you have an *independent observation* of the brain signal each time you make a scan. However, because temporally neighbouring scans are correlated with each other over time, these observations are not independent. So, the number of true independent observations is less that the number of scans. Therefore, when evaluating individual data a correction is needed. There are two possibilities: (a) Convolve the data with a specific function (e.g. a Gaussian curve). This way, the correlation between the neighbouring scans is forced in a specific pattern (in the same way as spatial smoothing, see 2.6 SMOOTHING). A disadvantage of this method is that the temporal specificity decreases. (b) The other option is to calculate the true temporal correlation between the residual error of neighbouring scans, using an AR(1) model (i.e. first-order auto-regression model):

[3.5] $$y_t = \alpha_t y_{t-1} + e_t$$

with y_t is the residual error from scan obtained at time t. Basically, this model calculates how well the residual error of $scan_{t-1}$ can be described by the residual error in $scan_t$ (expressed by α). This is calculated for each voxel separately. In the analysis, the number of observations will be adjusted for the factor α . This adjustment consists of a correction of the degrees of freedom by the factor α (i.e. the amount of temporal correlation).

When you are only interested in group results, including a correction for temporal auto-correlation is not needed. In fact, including such a correction does not affect your group results. This is because the correction involves a proportional decrease in the number of degrees of freedom, which will hence

lead to proportional lower t-value, but also to a proportional lower standard deviation (see CHAPTER 6 GROUP ANALYSIS).

4.2.4 Movement related noise

Despite rigid fixation of the head, subjects do move while performing an fMRI experiment. This movement is calculated during the realignment and the images are transformed accordingly (see 2.3 REALIGNMENT). However, even after this realignment process, there is still an effect of motion. Movement induces a disturbance in the signal in voxels throughout the entire scan. For example, the signal to noise ratio (SNR) changes in all voxels. Basically, signal variance is increased due to movement. By including the movement parameters in the design matrix, this additional variance can potentially be explained for a part. This is when one assumes that, for example, each time the subjects moves to the left, the signal decreases, while it increases when moving to the right. This might not be the case. In other words, the relationship between the movement parameters obtained during realignment and the effect of movement on the signal in the voxels is not linear or stable over time. By including the realignment parameters to model out the effects of motion, indeed some additional variance is explained, but not all signal variance due to movement is effectively removed.

It is possible that the movement (realignment parameters) is correlated with the task (see first figure in 2.3 REALIGNMENT). In that case, either a task factor or the movement factor can explain the variance in the signal. If you still decide to include the realignment parameters in your model, then the amount of uniquely explained variance by these task factors is reduced. Consequently, the t-ratio is affected negatively. However, the residual variance you subsequently explain with the task factor is reliably related to that task factor, because movement related variance is deleted.

When you find a high correlation between movement (realignment parameters) and the task factors, you might want to consider a more rigid fixation of the head of the subjects. Or just delete the data-set, if the correlation occurs only in a few subjects.

4.3 Effects of interest

The effects of interest refer to the variations in the signal which are induced by the task. For example the signal may increase when the task is performed, and decrease during a rest period. At the beginning of this chapter, the term 'model fit' was introduced. This term refers to how well the model explains the variance in the signal. It is important that the model as a whole explains as much variance as possible, because the amount of unexplained variance (i.e. *residual error*) should be as small as possible. In addition, specifically in the case of fMRI, we also look at the variance explained by each factor separately, as to make an estimation of how strong this particular effect is present in the signal. If a particular factor describes task related activation, then we are very interested in how much variance is explained by that factor in particular, rather than what the entire model explains. Consider for example a factor describing a blocked design (see CHAPTER 3), encoding a '0' for rest and a '1' for activation. The factor would then look like this:

# scan	1	2	3	4	5	6	7	8	9	…	30	31
event	rest	rest	Task	Task	Task	Task	Task	Task	Task	…	rest	rest
			1	1	1	2	2	2	2			
Factor 1	0	0	1	1	1	0	0	0	0		0	0
Factor 2	0	0	0	0	0	1	1	1	1	…	0	0

Figure 18a gives a graphical representation of the same factor:

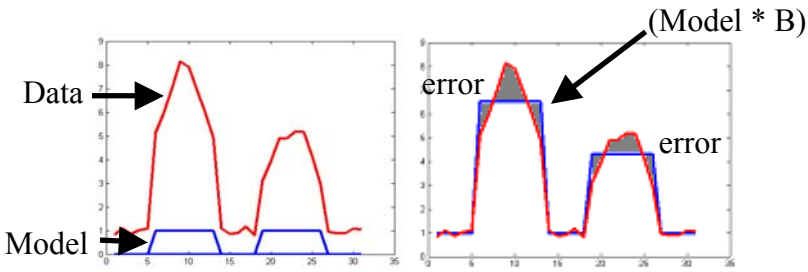


Figure 18a

So we expect the activation in some voxels to follow the pattern of the factor. If there is activation that behaves according to this pattern, then this would be interpreted as being activation (variance) due to this factor. We want to make an estimation of how strong this activation (variance) is present in the data. This is roughly measured in terms of the number of times the factor is present in the data. In figure 17b, the letter 'B' depicts this multiplication scalar. This letter B represents a matrix containing b-values (the multiplication scalars) for all factors separately. We expect there is a linear relation between the factor and the fMRI signal (hence the name of the analysis, General Linear Model, GLM). If the variance related to the factor is strongly present, then the b-value for that factor is high (and vice versa). Remember that this signal processing is done for each voxel separately. So one factor can result in high b-values in some voxels, but low b-values in others.

Taken together, the factors modelling the effects of no interest and effects of interest form the model that as a whole describes as much of the variance in the fMRI signal as possible. The remaining part of the variance in the fMRI signal is presumably random noise (randomness of the noise is one of the assumptions which have to be met before statistical analysis can continue).

The next step after building the model is to 'fit the model to the data'. In other words, we are going to estimate how strong each factor is present in the fMRI signal.

4.4 Fitting the model: b-values

We will now describe the statistical analysis, called multiple-regression. It is called 'multiple' as there is more than one regressor (i.e. factors) used to describe the data. These factors together form the model (represented by 'X'). This model describes the signal variation in the fMRI data time-series (data represented by Y). Using a statistical technique called multiple-regression, weights are given to each of the factors in the model, reflecting how strong the signal variation described by these factors is present in the data. These weights are called regression-coefficients, or b-values. The analysis is voxel-based, meaning that a separate regression is performed for every voxel in the brain.

For simplicity, matrix annotation is used. This annotation differs from normal algebraic annotation, but matrix annotation offers a very intuitive way of presenting the data and formulas. For a brief overview of matrix algebra, see 4.4 EXPLORATION.

4.4 EXPLORATION

Brief overview of matrix algebra

In matrix algebra, a single captital letter describes a matrix (for example a design matrix).

For example, A is a matrix consisting of 4 elements and is called a 4 by 2 matrix (4 rows, 2 columns)

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 6 \\ 4 & 8 \end{pmatrix}$$

You can multiply this matrix with any scalar , for example:

$$A*10 = \begin{pmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \\ 4 & 8 \end{pmatrix} *10 = \begin{pmatrix} 10 & 20 \\ 20 & 40 \\ 30 & 60 \\ 40 & 80 \end{pmatrix}$$

When multiplying matrices with other matrices, the number of columns in the first matrix must equal the number of rows in the second matrix. A multiplication is performed by adding up the products of the first row of the first matrix and the first column of the second matrix. In matrix algebra, you can only multiply matrices with the same inner dimensions. So you can multiply a 4 by 2 matrix with a 2 by 6 matrix, but not with a 6 by 2 matrix. In the example, we will multiply the 4 by 2 matrix A with itself, to obtain the sums of squares and cross-products matrix. For this multiplication, we need to transpose the matrix A, because A * A (4 by 2 * 4 by 2) cannot be calculated. We make A' which is the same matrix, but the rows are now columns. So the multiplication A' * A (2 by 4 * 4 by 2) is possible. This will generate a 2 by 2 matrix with the sums of squares (diagonal elements) and crossproducts (non-diagnonal elements). Of course, we can also calculate A * A'. This will generate a 4 by 4 matrix. As the multiplication symbol '*' is implied in matrix algebra, it is often omitted. So A'A is the same as A'*A.

$$A'A = \begin{pmatrix} 10 & 20 & 30 & 40 \\ 20 & 40 & 60 & 80 \end{pmatrix} * \begin{pmatrix} 10 & 20 \\ 20 & 40 \\ 30 & 60 \\ 40 & 80 \end{pmatrix} = \begin{pmatrix} 3000 & 6000 \\ 6000 & 12000 \end{pmatrix}$$

The first element of A'A, is in the first row and the first column of A'A and is by the sum of the multiplications of the elements of the first row of A' with the elements of the first column of A:

$$3000 = (10*10) + (20*20) + (30*30) + (40*40)$$

The basic formula for multiple-regression is given by:

[3.6] $Y = XB + error$

which states that the signal (Y) is equal to our model (X) plus some additional signal variation we cannot model as it is random noise (error). Recall that we want to minimize the amount of error. The amount of error, or unexplained variance, depends on how well the model describes the data Y (i.e. whether the factors describe the signal variations accurately, and whether a lot of signal variation is accounted for by adding factors for effects of no interest). The idea now is to find out whether, and if so, how strong each factor of the model is present in the data. This is done by multiplying the model X by a factor B (i.e. actually a different multiplication factor for each regressor of the model).

Assume we have the following situation ***(which is different from the example data set!)***:

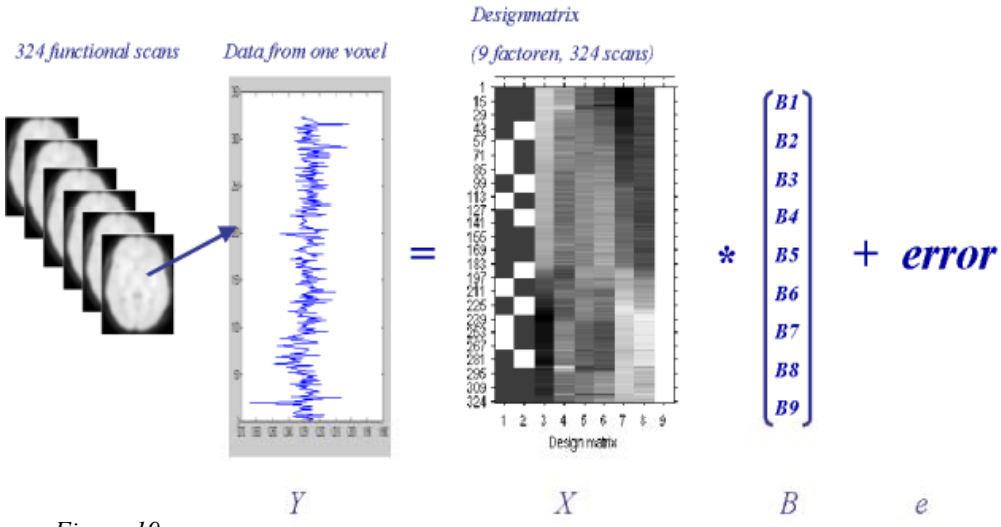


Figure 19

Y : represents the data, in this case the data from one voxel is displayed. The analysis is voxel-based, meaning that the entire analysis is performed for each separate (brain) voxel (voxels outside the brain are discarded using a mask).

X : represents the design matrix or model, which contains 9 factors:

Factor 1. factor for task 1
Factor 2. factor for task 2
Factor 3. linear trend (see 4.2 EFFECTS OF NO INTEREST)
Factors 4-8. movement related parameters (see 4.2 EFFECTS OF NO INTEREST)
Factor 9. intercept

The *intercept* actually is a vector (column in the matrix) consisting of elements with the value 1. This factor is included to model the basic activation level in the signal. During rest periods, the signal is not zero as there is still a signal coming out of the scanner. This is the signal that the scanner generates, and is always present. Therefore, this signal modelled by the intercept is called the *baseline* signal. During task periods, the signal effectively is the sum of the baseline activation level PLUS the increase (or decrease) due to the task. So all variations in the signal are relative to the baseline (see also 3.4 REST PERIODS).

B : represents the vector of multiplication scalars

E : represents the error (i.e. the variance in the signal that is not explained by the model)

In mathematical terms, this is stated by:

[3.6] $Y = XB + error$ (i.e. normal linear regression)

This is represented in figure 20, for the example data set (see also figure 18).

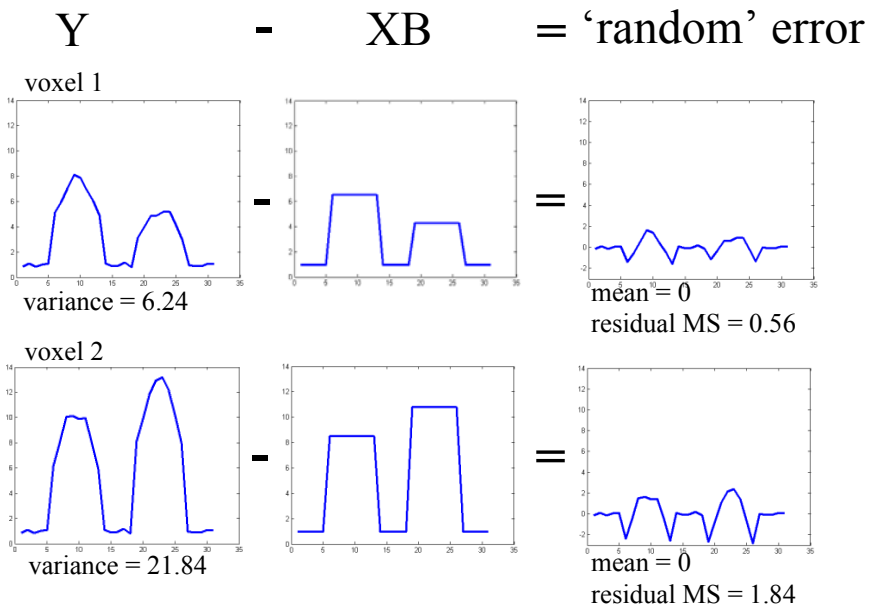


Figure 20

At this point, we want to find the multiplication scalars (i.e. b-values) for this particular model which best describe the data. This 'best description' is defined

in terms of yielding the smallest residual error. The residual error refers to that part of the data that is not described in the model, and hence is given by:

[3.7] $error = Y - XB$

A more formal approach of obtaining b-values that yield the smallest residual error is called the 'least squares criterion'. In mathematical annotation, this becomes minimizing:

[3.8] $\sum(error^2) = (Y - XB)'(Y - XB) = e'e = SS_e$ = sums of squares of the error

The solution for the minimalization of the sums of squares of the error is given by:

[3.9] $B = (X'X)^{-1}X'Y$

provided that $(X'X)^{-1}$ exists (i.e. if the matrix is not singular, see 5.2 MULTICOLLINEARITY). The B matrix contains estimations of the true B values.

The residual error, meaning the part in the signal which is not explained by the model, is called the *residual mean square*, and is given by:

[3.11] $MS_e = \frac{e'e}{N-H}$ with df = N-H (in contrast, the error variance is given by: $\frac{e'e}{N-1}$)

where N is the number of scans (or number of independent observations) and H is the number of factors of the model. Now we can make a b-map for every factor which represents the how strong this factor is present in the data, specified for each individual voxel.

4.4 EXPLORATION

Formula 3.9 is obtained using:

[3.10] $Y = X^{\perp} + XB$

which states that the best estimation of Y is a vector X straight below Y plus a correction for the height of Y.

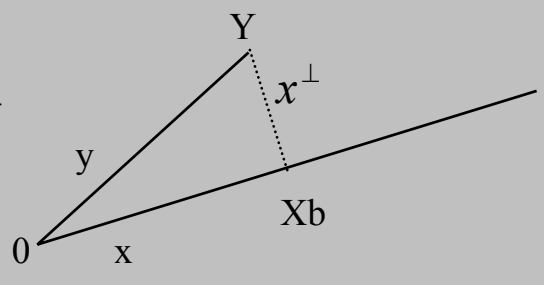
Multiplying [3.10] with X' gives:

[3.10] $X'Y = X'X^{\perp} + X'XB = 0 + X'XB = X'XB$

Then multiply [3.10] with $(X'X)^{-1}$ gives:

([3.9]) $B = (X'X)^{-1}X'Y$

visualization



4.4.1 An example

Assume that a specific subject has performed an fMRI experiment with two conditions. We will look at the activation time series of two voxels in a brain. Figure 21 depicts both the fMRI signal for two voxels as well as the model describing this signal (note that it is a limited model with only factors modelling effects of interest). The model encodes a '0' for rest and a '1' for activation.

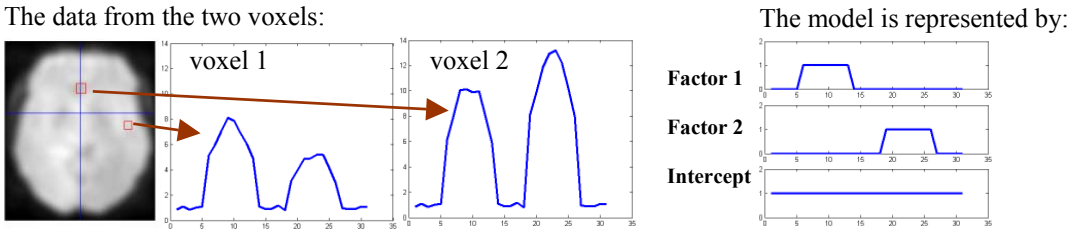


Figure 21

Now we can determine which linear combination of the model best describes the data for each voxel separately.

These values are found using formula 3.9. We will present the values here without presenting the data (Y, which is a 30 by 1 vector) and the model (X, which is a 30 by 3 matrix). For **voxel 1**, the b-values are:

([3.9]) $$B = (X'X)^{-1}X'Y = \begin{pmatrix} 0.192 & 0.07 & -0.00 \\ 0.07 & 0.192 & -0.00 \\ -0.00 & -0.00 & 0.032 \end{pmatrix} \times \begin{pmatrix} 26.026 \\ 8.256 \\ 101.00 \end{pmatrix} = \begin{pmatrix} 5.54 \\ 3.32 \\ 3.26 \end{pmatrix}$$

For both voxels, the b-values become:

	Voxel 1 b-value	**Voxel 2** b-value
Task 1	5.5386	7.5386
Task 2	3.3173	9.8173
Intercept	3.2581	5.4516

Note that the b-values for both task factors are higher in voxel 2 than in voxel 1. As can be seen in figure 18, this is caused by the higher signal in voxel 2. Signal strength is represented by the variance of the signal (variance(voxel1) = 6.24, variance(voxel2) = 21.84).

4.5 Making a t-map for a single factor

In this section, we will describe the transformation from b-value to t-value in order to determine the significance of the b-value. So far, we have obtained a b-value for each voxel in the brain, resulting in a b-map. There is a b-map for each factor of the model. The next step is to determine per b-map in which voxels this factor is significantly present. This is not directly reflected by the size of the b-value, but rather in every voxel a so-called t-test is performed to assess if the factor adds to the variance which is explained by the entire model. After performing the t-test, a t-map can be made for each factor, reflecting in which voxels this factor adds to the explained variance of the entire model. In other words, whereas the b-value reflects the weight of the factor within the model, the t-value reflects the contribution of this factor, corrected for the other factors, compared to the amount of error of the entire model. Whether this contribution is significant depends upon whether the t-value exceeds a certain threshold (see 5.3 THRESHOLDING). In essence, the t-value depends upon the ratio between the amount of variance this factor uniquely contributes (see 5.2 MULTICOLLINEARITY) and the amount of unexplained variance. It is therefore essential to reduce the amount of unexplained variance (i.e. error) by modelling as many sources of signal variance as possible. As discussed before, it is therefore important to model the effects of no-interest as well as those which are of interest (i.e. task related).

How can we obtain the t-value associated with the b-value? This transformation is actually the result of a test to determine if the b-value significantly differs from zero. In other words, a test is performed to determine whether the factor, which is multiplied by a certain b-value, adds something to the model or if this factor explains no variance and should be deleted from the model. Factors that do not add variance to the model only ensure lower t-values for all factors of the model (because residual mean square increases). Basically, we must decide whether the increase in the amount of explained variance (so-called regression sums of squares) is sufficient to warrant using the additional factor in the model. The hypotheses for testing the significance of any individual b-value (regression coefficient) are:

$H_0 : B_j = 0$ (null-hypothesis: no correlation between the data Y and the model X; b is zero)

$H_1 : B_j \neq 0$ alternative hypothesis: the correlation between Y and X is not zero; b is not zero

If H_0 (i.e. the null-hypothesis) is not rejected, then this indicates that the factor *j* can be deleted from the model, because the factor does not uniquely explain some part of the variance. The test statistic for this hypothesis is:

[3.14] $t = \frac{c'B}{se(B_j)}$ with $df = N - H - 1$

This test is actually a test of the contribution of the factor *j* given the other factors in the model. This is easy to see because the standard error of the mean, $se(B_j)$, depends upon both the residual error of the model and the unique contribution of the factor given the other factors.

Recall the informal definition of t (in 4.1 INTRODUCTION):

([3.14]) t = regression coefficient * $\sqrt{\frac{\text{explained variance}}{\text{unexplained variance}}}$

How is this definition obtained and how can it be calculated? Let's rewrite this informal definition into a mathematical expression.

The formal definition of the t-ratio (or t-value) is given by:

([3.14]) $$t = \frac{c'B}{se(B_j)} = \frac{c'B}{\sqrt{MS_e c'(X'X)^{-1}c}} \text{ with } df = N - H - 1$$

This can be rewritten is a more standard algebraic annotation as:

([3.14]) $$t = \frac{B_j}{\sqrt{\frac{1 - R^2_{Y.H}}{(1 - R^2_{j.G}) * (N - K - 1)}}}$$ (see text below for the legend)

where H contains all the factors of the model, and G all those factors but X_j.

In text, this would become:

([3.14]) t = $\frac{\text{regression coefficient}}{\sqrt{\frac{\text{unexplained variance}}{\text{explained variance}}}}$

So that in essence:

([3.14]) $$\text{t = regression coefficient} * \sqrt{\frac{\text{explained variance}}{\text{unexplained variance}}}$$

The regression coefficient reflects how strong the signal modelled by a particular factor is present in the data. The explained variance refers to the variance that is uniquely explained by a particular factor. The unexplained variance depends upon the model as a whole. The more variance the model explains, the less unexplained variance there is.

From this notation, it becomes clear that the t-ratio depends upon not only the size of the regression coefficient (i.e. b-value), but also on the ratio between explained versus unexplained variance. It is therefore important to (1) model the effects of no-interest as well as effects of interest, and (2) to calculate the uniquely explained variance for a particular task factor (see 5.2 MULTICOLLINEARITY).

Another way of looking at the t-value is to say that the t-value reflects the reliability of the b-value. The more reliable the b-value is, the smaller the standard error of the mean becomes, and hence the higher the t-value. This becomes more clear as we look at the complete formula for the t-value:

([3.14]) $$t = \frac{c'B}{se(B_j)} = \frac{c'B}{\sqrt{MS_e c'(X'X)^{-1}c}} \text{ with } df = N - H - 1$$

Let's look at the different components of this formula:

$\boldsymbol{c'}$ refers to the contrast vector. This is a matrix used to select a specific factor. For example, if you want to work only with the first factor of a design matrix with 9 factors (from figure 17; the first two factors model the task), the contrast matrix will become:

$$c' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$\boldsymbol{B}$ is the matrix containing the b-values for each factor of the design matrix:

$$B' = (b_1 \quad b_2 \quad b_3 \quad b_4 \quad b_5 \quad b_6 \quad b_7 \quad b_8 \quad b_9)$$

MS_e is the mean of the squared residuals (residual mean square) and is a weighted representation of the error variance $1-R^2_{y.H}$ (see [3.11]). If a particular factor is included in the model, but explains no unique variance, then the residual sums of squares remains the same, but the residual mean square increases (see [3.11]), because of the increase in the number of factors H.

$c'(X'X)^{-1}c$ is a weighted representation of $\frac{1}{1-R^2_{j.G}}$ and refers to that part of the variance uniquely explained by a particular factor *j*. The more unique variance this factor explains, the smaller this element becomes, and hence the higher the t-value becomes.

4.5 EXPLORATION

Model fit; the F-test

The residual sums of squares, the part of the data that is NOT explained, is:

$$SS_E = S_{yy} - SS_R$$

The regression sums of squares, the part of the data that IS explained, is:

$$SS_R = B'X'y - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

and the total sums of squares, the total data that can be explained, is:

$$S_{yy} = y'y - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

so that the residual sums of squares effectively becomes:

$$SS_E = y'y - B'X'y$$

The model fit can be tested using the F statistic:

$$F = \frac{SS_R / H}{SS_E / N - H} = \frac{MS_R}{MS_E} \text{ with } df = N, N - H$$

where H is the number of factors and N is the number of observations (scans)

4.5 EXPLORATION

Exploring the meaning of $(X'X)^{-1}$

$(X'X)^{-1}$ is the weighted inverse covariance matrix of X. The covariance matrix is given by:

[3.15] $Cov(X) = X_{dev}'X_{dev}$

where [3.16] $X_{dev} = X - \bar{X}$

where $\bar{X}$ has columns containing factor averages. In SPM99/2, all factors are weighted so that each factor has mean 0. Therefore, in this particular case, we get:

([3.16]) $X_{dev} = X - \bar{X} = X - 0 = X$

so that ([3.15]) $Cov(X) = X_{dev}'X_{dev} = X'X$

A covariance matrix has diagonal elements which represent factor variance and non-diagonal elements which represent covariance. Factor variance is given by:

[3.17] $$S^2_{factor} = \frac{\sum_N (x - \bar{x})^2}{N - 1}$$

and covariance between factors by:

[3.18] $S_{12} = \bar{x}_1\bar{x}_2 - \bar{x}_1 \times \bar{x}_2$

Element (1,1) of $X'X$ represent the factor variance for factor 1, although not corrected for the number of observations (see formula 3.17). This element is also called the sums of squares for factor 1 (SS). The matrix $X'X$ is also referred to as sums of squares and cross products (SSCP), since the non-diagonal elements are cross products of pairs of factors.

The next step is to invert $X'X$:

[3.19] $$(X'X)^{-1} = \frac{1}{\det}\begin{pmatrix} m(X'X)_{(1,1)} & . & . & m(X'X)_{(1,H)} \\ . & . & . & . \\ . & . & . & . \\ m(X'X)_{(H,1)} & . & . & m(X'X)_{(H,H)} \end{pmatrix}$$

where H is the number of factors, and $m(X'X)_{(h,h)}$ refers to the *minor* of the h[th] element $X'X$ which is defined by the determinant of the matrix formed by deleting the h[th] row and h[th] column from $X'X$. This element reflects the combined variance, corrected for the covariances, of the factors other than h. The next step is calculate the determinant of the entire matrix $X'X$. This determinant reflects the generalized variance of the entire model. Covariance between factors would decrease this variance, since in that case more than one factor explains a segment of the total variance. For an example, see 4.5.1 EXPLORATION.

4.5 EXPLORATION - continued

The first element of the matrix $(X'X)^{-1}$ then becomes:

[3.20] $$inv(X'X)_{(1,1)} = \frac{m(X'X)_{(1,1)}}{\det(X'X)}$$

and basically reflects the ratio between generalized variance of the other factors and the generalized variance of the entire model. In more practical terms, this element reflects the variance explained by the first factor (increasing the factor variance four times will result in four times decrease in value of this element).

Note that the factors means do not have to be zero for $(X'X)^{-1}$ to reflect the correct ratio between the variances! (see EXPLORATION 4.5.1)

4.5.1 An example - continued

In this example, we will calculate the reliability of the b-values by rewriting them as t-values for both task factors in both voxels. Recall that the formula for t-values is given by:

([3.14]) $$t = \frac{c'B}{\sqrt{MS_e c'(X'X)^{-1}c}} \text{ with } df = N - H - 1$$

Let's start with B. B refers to an entire matrix containing b-values for all factors for each voxel. These have been calculated in the previous part of the example:

B for voxel 1 : $B_{voxel1} = \begin{pmatrix} 5.54 \\ 3.32 \\ 3.26 \end{pmatrix}$ and for voxel 2: $B_{voxel2} = \begin{pmatrix} 7.54 \\ 9.81 \\ 5.45 \end{pmatrix}$

So the scalar by which to multiply the factor, the b-value, is known for each factor. To calculate the t-value for the first factor in voxel 1, we need the first element of the B matrix for voxel 1. This is accomplished by multiplying the B matrix with a contrast vector. To select the first element, this contrast vector becomes:

$$c' = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$$

and to select the second element (b-value for the second factor), this vector becomes:

$$c' = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}$$

So that $c'B$ for the first factor becomes:

$$\begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} 5.54 \\ 3.32 \\ 3.26 \end{pmatrix} = 5.54$$

Next, we calculate the residual mean square:

([3.11]) $$MS_e = \frac{e'e}{N - H}$$

whereby e(rror) is defined as:

([3.7]) $error = Y - XB$

This is depicted in figure 19 for both voxels.

Recall that Y (the data; figure 19) for voxel 1 and for voxel 2 is a 31×1 matrix (31 scans, 1 column) and that X (the design matrix) is a 31×3 matrix (31 scans, 2 factors + intercept). So calculating the error for every data point (i.e. scan) results in a 31×1 matrix:

([3.7]) $e = Y - XB = (31 \times 1)$ matrix

containing the difference per scan between Y (real data) and XB (estimation of that data).

Consequently, the residual sums of squares becomes one number, because of the matrix multiplication. The dimensions are given by:

[3.7b] $e'e = (1\times 31)\times(31\times 1) = (1\times 1)$ matrix (see also 4.4 EXPLORATION)

Now the residual mean square of the variance can be calculated using:

([3.8]) $MS_e = \frac{e'e}{N-H} = \frac{15.68}{31-3} = 0.56$

Finally, we calculate $c'(X'X)^{-1}c$, which reflects a division by the variance of the first factor, corrected for the correlation with the other factors. Calculation of this element is done in three steps:

First, we calculate the sums of squares and cross products matrix of X (with column means ≠ 0):

[3.17] $$(X'X) = \begin{pmatrix} 8 & 0 & 8 \\ 0 & 8 & 8 \\ 8 & 8 & 31 \end{pmatrix}$$

which has the sums of squares for each factor on the diagonal and the cross products between the various factors as non-diagonal elements. These are weighted variances and covariances (divide the elements by N-1 to obtain the variances and covariances).

Second, this square matrix is then inverted:

[3.19] $$(X'X)^{-1} = \frac{1}{\det(X'X)}\begin{pmatrix} 184 & 64 & -0 \\ 64 & 184 & -0 \\ -0 & -0 & 31 \end{pmatrix} = \begin{pmatrix} 0.192 & 0.067 & -0.000 \\ 0.067 & 0.192 & -0.000 \\ -0.000 & -0.000 & 0.032 \end{pmatrix}$$

Third, apply the contrast vector to this inverted matrix, so that the first element of this matrix is selected:

$$c'(X'X)^{-1}c = \begin{pmatrix}1 & 0 & 0\end{pmatrix} \times \begin{pmatrix}0.192 & 0.067 & -0.000 \\ 0.067 & 0.192 & -0.000 \\ -0.000 & -0.000 & 0.032\end{pmatrix} \times \begin{pmatrix}1 \\ 0 \\ 0\end{pmatrix} = 0.192$$

The t-value for the first factor in the voxel 1 is subsequently calculated by:

([3.14]) $$t = \frac{c'B}{\sqrt{MS_e c'(X'X)^{-1}c}} = \frac{5.5386}{\sqrt{0.56 \times 0.1917}} = 16.91 \text{ with } df = N - H - 1$$

The other t-values are computed in the same way:

	Voxel 1		**Voxel 2**	
	b-value	t-value	b-value	t-value
Task 1	5.5386	16.91	7.5386	12.70
Task 2	3.3173	10.13	9.8173	16.54
Intercept	3.26	-	5.45	-

Note that the highest b-value (voxel 2, task 2) is NOT associated with the highest t-value. This is due to the higher error variance (residual mean square, MS_e) in voxel 2. Let's work out the numbers for the second factor for voxel 2:

([3.14]) $$t = \frac{c'B}{\sqrt{MS_e c'(X'X)^{-1}c}} = \frac{9.8173}{\sqrt{1.84 \times 0.1917}} = 16.54 \text{ with } df = N - H - 1$$

4.5.1 EXPLORATION

To understand what the elements of this matrix mean, a little exploration:

Element(1,1) of the matrix $(X'X)^{-1}$ (0.192) depicts the ratio between variance explained by the factors 2 and 3, and the generalized variance of all the factors (i.e. the model). This value is obtained in two steps.

Step 1:
Obtain the first minor of the matrix $(X'X)$, which is obtained by deleting the first row and the first column of the matrix $(X'X)$. We get:

4.5.1 EXPLORATION - continued

$$\begin{pmatrix} 8 & 8 \\ 8 & 31 \end{pmatrix}$$

The first minor is then calculated by:

$\text{minor}(1,1) = (8 \times 31) - (8 \times 8) = 184$

This value reflects the generalized variance of the factors 2 and 3 of the model (NOT factor 1).

Step 2:

Obtain the determinant of $(X'X)$, which is 960. This determinant reflects the generalized variance of all the factors within the model.

The first element of the matrix $(X'X)^{-1} = 184/960 = 0.1917$. This element depicts the ratio between variance explained by the factors G and the generalized variance of the entire model (all factors H). Hence this value represents the amount of variance explained by the factor j.

4.6 Making a t-map for a contrast between two factors

In the previous section, the b-value for one factor was rewritten as a t-value for that factor, thereby reflecting how strong the b-value deviates from zero. However, in many experiments one would like to answer the question if there are regions that are significantly more active during one condition compared to another. For example: Is activation in the motor cortex higher during the experimental condition than during the control condition?

You might think that since we have t-values for every condition, we can subtract these to test whether the activation is higher in one condition compared to the other. This is NOT the case, as the factors are correlated with each other. Therefore, when answering the question, you should take into account this correlation (i.e. covariance) between the factors. The correct way is therefore to compare the b-values of both conditions (factors) while taking into account the variances and covariances of the factors.

Recall the formula for t:

([3.14]) $$t = \frac{c'B}{\sqrt{MS_e c'(X'X)^{-1}c}} \text{ with } df = N - H - 1$$

The B matrix contains the b-values for each separate factor. We simply subtract the second b-value from the first, which is done by multiplying the B matrix with a contrast matrix similar to the one used to select only the first b-value.

Whereas for one factor (from a model with nine factors, see figure 17) the contrast vector was:

$$c' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

for a two-factor t-test (e.g. factor 1 > factor 2) the contrast matrix for the same model becomes:

$$c' = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Multiplying the latter contrast vector with the B matrix will result in subtracting the second b-value (for the second factor) from the first b-value (from the first factor).

Subsequently, the multiplication $c'(X'X)^{-1}c$

will now yield a combination of the weighted variances of both the first and the second factor plus a correction for the covariance between these two factors.

Finally, we need to determine the size of MS_e. The residual mean square (error, MS_e) is a fixed number for each voxel and is dependent upon the entire model (recall, error = Y - XB). So for all contrasts that can be made between the factors of the model, this component does not change.

4.6 EXPLORATION

When only one factor was tested, $c'(X'X)^{-1}c$ gave a weighted representation of the unique variance of the first factor. In a contrast between two factors, this gives:

$$c'(X'X)^{-1}c = \begin{pmatrix} 1 & -1 & 0 \end{pmatrix} \times \begin{pmatrix} \text{var}(1) & \text{cov}(1,2) & \text{cov}(1,3) \\ \text{cov}(2,1) & \text{var}(2) & \text{cov}(2,3) \\ \text{cov}(3,1) & \text{cov}(3,2) & \text{var}(3) \end{pmatrix} \times \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = (\text{var}(1) - \text{cov}(2,1)) - (\text{cov}(1,2) - \text{var}(2)) = (\text{var}(1) + \text{var}(2)) - 2 \times \text{cov}(1,2)$$

4.6.1 An example - continued again

In this example, we will perform a t-test between the two task factors for the data in voxel 1, to determine if factor 1 explains significantly more variance than factor 2 in this voxel.

The contrast becomes: $c' = \begin{pmatrix} 1 & -1 & 0 \end{pmatrix}$

So $c'B = \begin{pmatrix} 1 & -1 & 0 \end{pmatrix} \times \begin{pmatrix} 5.54 \\ 3.32 \\ 3.26 \end{pmatrix} = 5.54 - 3.32 = 2.22$

And

$$c'(X'X)^{-1}c = \begin{pmatrix} 1 & -1 & 0 \end{pmatrix} \times \begin{pmatrix} 0.192 & 0.067 & -0.000 \\ 0.067 & 0.192 & -0.000 \\ -0.000 & -0.000 & 0.032 \end{pmatrix} \times \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \approx 0.192 + 0.192 - 0.067 - 0.067 \approx 0.25$$

We can now determine the t-value of the difference between the two factors in voxel 1 by applying the general formula:

([3.11]) $$t = \frac{c'B}{\sqrt{MS_e c'(X'X)^{-1}c}} = \frac{2.22}{\sqrt{0.56 \times 0.25}} = 5.93 \text{ with } df = N - H - 2$$

So, now we have a contrast t-value. What can we say about the difference in brain activation during these two task conditions? We need to know if this difference t-value is significant. The significance issue is discussed in 5.3 THRESHOLDING.

5 Statistical issues

5.1 Introduction

In this chapter, we will look at several important statistical issues. In 5.2 we will discuss the implications of multicollinearity. Multicollinearity refers to the correlation between the factors in the design matrix. Before scanning subjects with a particular design, one should calculate the multicollinearity of that design, because high a correlation between factors can results in low t-values. In 5.3 we will discuss thresholding of statistical maps (t-maps) so one can determine 'which voxel is significant'.

5.2 Multicollinearity

Multicollinearity refers to the correlation between the factors in the design matrix. Multicollinearity is a form of *singularity* of the matrix. When a matrix is *singular*, then one factor can be rewritten as a linear combination of (some of) the other factors. See for an example, figure 23a. In that particular case, adding this factor to the model does not add anything in terms of extra information concerning the signal you want to explain, since this information is already present in the other factors. There are degrees of multicollinearity, from none to complete singularity. Consider the following example of a design matrix with two (blocked) task factors:

$$X = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

where X is a (simple) design matrix, with two task factors (column 1 and 2) and an intercept (column 3). The first factor (column 1) can be written as a linear combination of the other two factors:

$$\text{First column} = \begin{pmatrix}1\\1\\1\\0\\0\\0\end{pmatrix} = \begin{pmatrix}0\\0\\0\\1\\1\\1\end{pmatrix} - \begin{pmatrix}1\\1\\1\\1\\1\\1\end{pmatrix} \Rightarrow \begin{pmatrix}-1\\-1\\-1\\0\\0\\0\end{pmatrix} * -1 = \begin{pmatrix}1\\1\\1\\0\\0\\0\end{pmatrix}$$

In this case, the matrix X is called singular, because the first factor is redundant. A simple change in the experimental design can remove this problem completely. The new matrix X then becomes:

$$X = \begin{pmatrix}1 & 0 & 1\\1 & 0 & 1\\0 & 0 & 1\\0 & 0 & 1\\0 & 1 & 1\\0 & 1 & 1\end{pmatrix}$$

In this case, trying to write the first column in terms of the other two is only partially successful:

$$\text{First column} = \begin{pmatrix}1\\1\\0\\0\\0\\0\end{pmatrix} \Rightarrow \begin{pmatrix}0\\0\\0\\0\\1\\1\end{pmatrix} - \begin{pmatrix}1\\1\\1\\1\\1\\1\end{pmatrix} \Rightarrow \begin{pmatrix}-1\\-1\\-1\\-1\\0\\0\end{pmatrix} * -1 = \begin{pmatrix}1\\1\\1\\1\\0\\0\end{pmatrix}$$

Note that the elements in the third and fourth row are not correct. In fact, a third of the factor cannot be written using linear transformations. So, the matrix is not singular, but there still is multicollinearity, as there still is some overlap between the first factor and the linear transformation from the other two factors.

Even small degrees of multicollinearity can pose a problem for the analysis, because of three reasons:

1. limits the amount of variance explained by the model. Because a particular factor (or factors) can be partially rewritten in terms of other factors, these factors in fact all explain the same part of the variance in the signal, instead of explaining additional variance. As a consequence of this overlap in variance explanation, the model fit will be non-optimal and therefore the error variance will remain unnecessarily large. (alteration of the model may fix this)

2. makes determining the importance of a given factor difficult. Because multiple factors explain the same part of the variance in the fMRI signal, it is difficult to decide which of these factors is 'responsible' for this variance. For example, a task factor may be correlated with the linear trend factor. Then, task activation can be attributed to either the task factor or the linear trend factor. Maybe the task factor is not important at all because a linear trend is present in the data.

5.1 EXPLORATION

An example for calculating the multiple correlation between factor j and the other three factors in a design:

[3.XX] $R^2_{j.G} = r^2_{j1} + r^2_{j2.1(s)} + r^2_{j3.12(s)}$

where

[3.XX] $r^2_{j2.1(s)} = \left(\frac{r_{j1} - r_{j1} r_{12}}{\sqrt{1 - r^2_{12}}} \right)^2$

and

[3.XX] $r^2_{j3.12(s)} = \left(\frac{r_{j3.2} - r_{j2.1(s)} r_{23.1}}{\sqrt{1 - r^2_{23.1}}} \right)^2$

The easiest way to calculate the squared multiple correlation is by using SPSS (linear regression).

3. increases the standard error of the b-values and hence decreases t-values. To better comprehend this, inspect the (normal algebraic) formula for the standard error of the b-values:

[3.14a] $se(B_j) = \sqrt{\frac{1 - R^2_{y.H}}{(1 - R^2_{j.G})(N - H - 1)}}$

The term $R^2_{j.G}$ refers to the squared multiple correlation between factor j and the other factors. In other words, this term reflects the collinearity for this factor in the model. The greater this term becomes, the greater the estimated standard error of b (for factor j) becomes. This, in turn, will result in a smaller t-value since the b-value is divided by this estimated standard error of b.

To get an indication of the variance which is explained uniquely by a particular factor, one can correct the factor variance for the squared multiple correlation with the other factors using the variance inflation factor:

[3.16] $\frac{1}{1-R_{j.G}^2}$

where G refers to the factors in the design matrix minus factor j, and hence $R_{j.G}^2$ denotes the explained variance of factor j by the other factors.

To investigate the size of the multicollinearity in the design matrix, one can calculate the multiple correlation between a specific factor and the remaining factors in the design (see 4.1 EXPLORATION).

5.3 Thresholding

After the statistical analysis, we have statistical maps with t-values for each (brain) voxel (t-maps). The question now becomes which t-values represent significant activation in the brain. To answer this question, a threshold is set above which t-values are *significant*. Significant means that the occurrence of such a high (or low) t-value by chance within the normal (student-t) distribution is very unlikely (i.e. below a predetermined chance level). Normally, a threshold is set so that significant values are those that have a likelihood of five percent or less to occur in that particular distribution. Using this threshold, we can determine, for each individual voxel, if the t-value has a chance of 5 percent or less to occur in the known (student-t) distribution.

This threshold (or alpha) reflects the percentage of so-called 'Type 1' errors. Type 1 errors represent values which are in fact part of the distribution, but which are wrongly treated as being not a part of the distribution. The amount of Type 1 errors is normally set at 5 percent, which is comfortably small enough. When one test is performed, we allow a chance of 5 percent for Type 1 errors. This can be written as:

[5.1] $p(\text{one or more Type 1 errors}) = 1-(1-\alpha)^J = 1-(1-0.05) = 1-0.95 = 0.05$

where *p* stands for 'chance of' and J refers to the number of tests being performed.

Similarly, when we perform two tests, we want to maintain this overall 5 percent level. Simply performing more tests should not mean more chance of Type 1 errors over all tests being performed. If we calculate the chance of Type 1 errors when performing two tests, we get:

([5.1])

$$p(\text{one or more Type 1 errors}) = 1-(1-\alpha)^{J} = 1-(1-0.05)^{2} = 1-0.9025 = 0.0975$$

The overall chance of a Type 1 error has increased from 5 percent to almost 10 percent, simply by performing two tests instead of one. This is not something we want.

In a standard fMRI t-map, there are about 16,000 voxels. As a t-test is performed for every voxel, this means that about 16,000 tests are performed. And yet, when we evaluate the activation in the entire brain, we would like to have an overall chance of 5 percent on Type 1 errors. Consequently, some form of correction is needed to keep that chance at 5 percent.

5.3.1 The Bonferroni correction for multiple tests

The Bonferroni method offers a correction for thresholding for multiple tests by dividing the threshold (i.e. alpha) by the number of independent tests that are performed:

[3.17] $\alpha_{overall} = \frac{\alpha}{N}$

for N = number of independent tests. So if $\alpha = 0.05$ and N = 16,000 (=number of voxels) then:

([3.17]) $$\alpha_{overall} = \frac{\alpha}{N} = \frac{0.05}{16{,}000} = 0.000003125$$

So voxel-wise, t-values should be thresholded at $\alpha = 0.000003125$. The corresponding t-value depends upon the degrees of freedom. For an analysis of a single subject, this is number is given by the test that is performed (see [3.14]).

Applying the Bonferroni correction with the number of voxels as the number of independent tests is rather conservative, because the actual number of independent observations IS NOT the number of voxels. Rather, the voxel-values are spatially correlated (see 2.1.5 SMOOTHING). In practice, this means that if one particular voxel-value is not-significant, it is likely that its immediate neighbour also is not significant. The number of resels (see 2.1.5 SMOOTHING) is a better estimation of the true number of independent elements. (If one chooses to use resels for the Bonferroni correction, then the resolution of the t-map should be equal to the resel size).

However, the number of resels also does not represent the true number of independent observations since resels themselves are correlated, although this correlation is known. The true number of independent observations is not easy to work out, so a different approach is needed. One such approach is the Gaussian Random Field theory, which is used in SPM99/2.

5.3.1 EXPLORATION

Why a threshold of t = 4.51 ?

The critical t-value of 4.51, which is commonly applied in the UMC-Utrecht, is based on a voxel-wise alpha of 10 percent. This value reflects a two-tailed alpha of 5 percent, since there is no a priori assumption whether the test value is either in the low 5 percent or the high 5 percent of the distribution. This alpha is divided by 16,000 voxels, resulting in an overall alpha of 0.00000625. This value is converted to a t-value using a t-distribution with infinite degrees of freedom. Of course, the number of degrees of freedom is not infinite, rather the number of scans minus the number of factors (minus 1 or 2, depending on the contrast). However, above 500 degrees of freedom, the exact number does not matter that much anymore.

And how about t = 3.09 ?

Whereas a threshold of t = 4.51 reflects an alpha of 5 percent corrected for multiple tests, a threshold of t = 3.09 refers to an two-tailed alpha of 0.002, NOT corrected for multiple tests. This threshold is often used for exploratory purposes. For example, when one wants to know if there is any trend present in the data.

How can I calculate the correct threshold for my own data ?

You can calculate your very own critical t-value using the t-inverse (T.INV) function in MS Excel. This might result in a different critical t-value if the number of brain voxels of your data (falling within the brain mask) is significantly lower or higher than the 16,000 associated with a t-value of 4.51.

6 Group analysis

6.1 Introduction

Performing a group analysis allows generalization of the results to a population, for example the population of healthy subjects or schizophrenia patients. There are two main strategies for performing a group analysis, namely look at:

1. group effects in the entire brain
2. mean group activation for different conditions in Regions of Interest

The first method involves performing statistics over the entire brain. A disadvantage is the correction of the significance threshold for multiple testing. This might prove to be a very stringent correction if differences between two conditions are very small. As an alternative, one can define Regions of Interest (ROIs) and look at the mean activation differences between conditions within these ROIs, thereby allowing a lower threshold. A slightly different approach is to define Volumes of Interest. Originally, the volume of interest is the entire brain, consisting of approximately 16,000 voxels. However, if one expects differences between conditions to occur only within the visual cortex, one does not want to correct for voxels in which no effect is expected. To bypass the correction for the entire brain, one can select any area of the brain as the volume of interest (VOI). The subsequent correction for multiple testing is hence less stringent.

6.2 Standard group analysis

6.2.1 A single group and a single factor

The Bonferroni method offers a correction for thresholding for multiple tests. Significant values in the group-map *reflect activation that is stable over subjects in a particular group*. The voxel values do not need to be significant on an individual level for the group voxel value to become significant. For example, consider the example in figure 22. The t-value (see also 6.1 EXPLORATION) for a specific voxel is obtained from each individual brain. Note that none of the individual t-values is above the threshold of 't = 4.51' (see 4.5.1 EXPLORATION). The next step is to calculate the group t-value, which represents the stability of the individual t-values. One possibility to obtain the group-value for this voxel is to perform a standard one-sample t-test:

[6.1] $t_{group} = \sqrt{n}\,\frac{\overline{b}}{SD(t)}$ with $df = n-1$, and 'n' is the number of subjects.

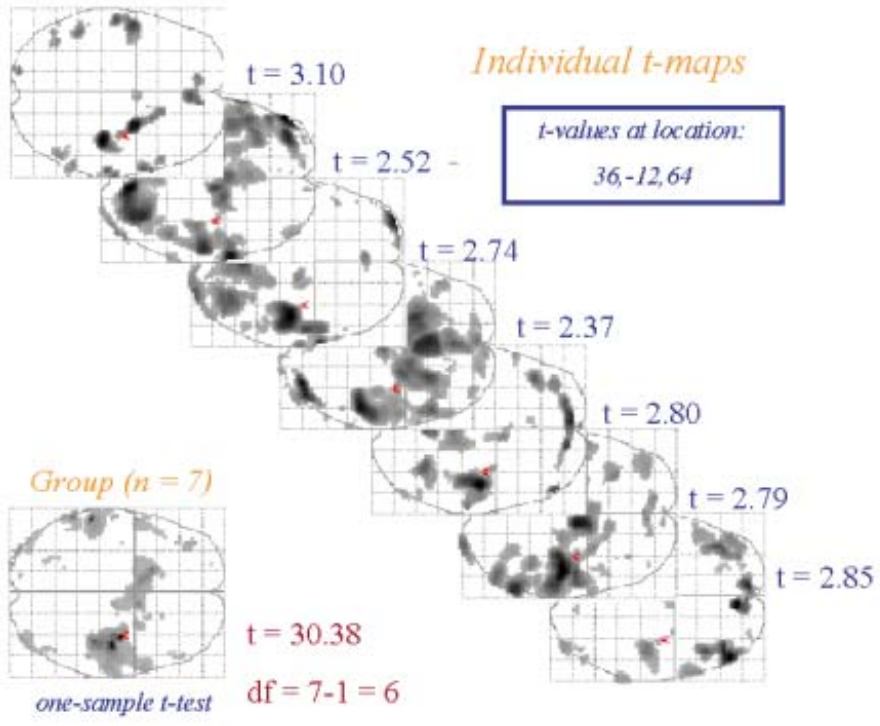


Figure 22. Example of a group analysis over 7 healthy subjects with individual t-values from voxel at location [36,-12,64]

Applying this formula to the data of figure 23 (on the individual t-values) gives:

([6.1]) $$t_{group} = \sqrt{n}\frac{\bar{t}}{SD(t)} = \sqrt{7}\frac{2.74}{0.2356} = 30.38$$

The resulting t-value for the group seems to be rather high (t = 30.38), compared to the low individual t-values. However, because these values do not differ so much between individuals, the standard deviation (SD) is rather low. Subsequently, the group t-value is high, reflecting the fact that the activation in that particular voxel is present in all subjects in a stable manner. Another way of looking at this is to consider all subjects as merely different measurements (observations) of a particular activation. When all the measurements indicate roughly the same value, then this value is very robust. In other words, the likelihood that this voxel (with the high group t-value) is involved in performing the task is very high.

To determine the significance of this value (actually, to assess the chance that such a t-value occurs in a particular distribution), one needs to know from which distribution this t-value is obtained. There are two ways to describe the distribution for this t-value:

1. use the distribution from the group of t-values over which the test in performed (*voxel-based standard deviation; between subject variation*)
2. use the ratio between between and within subject variation (*multi-level, or mixed model approach*)

When you want to compare the activation of several factors in a single group, other tests can be used. For example, an ANOVA can be used to test for the difference between multiple factors.

Comparing two groups involves a two-sample t-test.

6.1 EXPLORATION

Should a group analysis be performed over individual b or t values?

In the example discussed above, t-values were used to calculate the group t-value. The advantage of using t-values is that these are normalized for the error variance in each subject. Subjects who have high b-values, but also high error variance, will have low t-values, reflecting the fact that the high b-values are not so valid. So then the group analysis will be done over the significances of the b-values.

However, consider for example two groups; a patient group and a healthy control group. For the sake of the example, lets assume that the error variance in the patient group is higher (maybe due to movement related noise). It is possible that a patient and a healthy control have the same effect-size (b-value for a specific factor), but differ in the amount of error. While the b-values are equal, the t-values will differ.

6.2.2 Including a covariate in the group analysis

When performing a group analysis, the subjects actually represent different measurements of the same activation. However, subjects can differ quite a lot in, for example, age. Also, when comparing a patient group and a healthy control group, the group may differ in age (or educational level). In sum, there are factors which may affect the outcome of the group analysis, but which fall outside the scope of the experiment (i.e. you want to know the difference between patients and controls, independent of their age or education). In that case, you can add a factor to your group analysis, called a covariate. By including a covariate for age, you effectively remove the influence of age on the test values.

Another reason for adding a factor into the group analysis is that there might be a correlation between brain activation and performance on the experimental task. Subjects with high b-values might have also a good performance on the task, and vice versa.

To illustrate this, we will present sample data from a group analysis which is analysed using multiple-regression. In case of a single group and a single factor, this multiple-regression analysis is the same as a one-sample t-test (see 6.2.1).

Example: A group of healthy controls (n=8) performed a motor task, in which they had to respond to a target as fast as possible. The b-values for a particular voxel in the motor cortex are given by:

$$Y = \begin{pmatrix} 2 \\ 3 \\ 2.5 \\ 2.7 \\ 0.5 \\ 0.2 \\ 1.0 \\ 0.4 \end{pmatrix}$$

Now, we can construct a design matrix which describes the data (similar as to the single-subject analysis). The first factor describes 'performing the task'. As all subjects performed the task, they are all encoded as '1'. Furthermore, we can add a factor representing performance on the task (i.e. reaction time of the motor responses). X becomes:

$$X = \begin{pmatrix} 1 & 390 \\ 1 & 389 \\ 1 & 370 \\ 1 & 375 \\ 1 & 310 \\ 1 & 355 \\ 1 & 360 \\ 1 & 325 \end{pmatrix}$$

in which the first column reflects the task being performed (i.e. in every subject) and the second column reflects task performance (i.e. reaction time of the motor response). Using the same formulas as in CHAPTER 4, we can calculate the b-values and the t-values. The b-value for the first column (performing the task) becomes 0.152, and for the second factor 0.03. The t-values become .393 and 5.01, respectively. With a threshold of t > 4.5 (i.e. alpha = 5 percent, corrected for multiple comparisons), the overall task

activation is not significant in this voxel (i.e. factor 1). However, the activation in this voxel is significantly correlated with the performance on the task (factor 2).

Calculating the correlation between brain activation and task performance can also be done at the single-subject level. This entails including a factor in the single-subject matrix, which represents task performance.

6.3 Standard-deviation in group analysis

The voxel-based standard deviation (SD) is calculated using the individual values for that voxel in the group:

[6.3] $$SD_{voxelbased} = \sqrt{\frac{\sum_{n}(x-\bar{x})^2}{n-1}}$$

where n is the number of subjects and x is either t or b statistic (see 6.1 EXPLORATION).

**
WORSLEY, 2002 → MULTI-LEVEL APPROACH
**

6.4 Region of Interest analysis

Many hypotheses that are addresses using fMRI are stated in terms of the specific functionality of brain regions of interest (ROIs). These regions can be based on their (a) *cytoarchitectonic structure* (e.g. Brodmann areas) or (b) *anatomical landmarks* such as sulci. Other definitions of ROIs can be based on (c) *previous findings in the literature,* (d) *own pilot data,* or (d) *the experimental (functional) data.*

Another, more practical, reason for using an ROI approach is that the Bonferroni correction for multiple tests (i.e. all brain voxels) is too stringent for the expected effects (differences) between task conditions (Nieto-Castanon et al. 2003).

There are two main approaches, being (a) Region of Interest (ROI) approach, and (b) Volume of Interest (VOI) approach. In the ROI approach, a particular measure of the entire region is used. This can be, for example, the average b-value or the maximum b-value. Using the average signal, you assume that the entire ROI is involved in the task condition. These values can then be entered into a standard statistical package for further analysis. Using the VOI approach, the volume of interest is no longer the entire brain (with 16,000 voxels), but rather some predefined region (with significantly fewer voxels). Within this volume, a group analysis can be performed similar to a whole-brain group analysis, but with a lower threshold (as the number of tests/voxels is lower). Using this method, you can test for differences between the conditions voxel by voxel. Consequently, you can detect whether the conditions each activate a different part of the volume, or if there is a difference in activation only in a particular location.

Consider for example a group of 20 healthy controls which performed a blocked visual task in which three different colors are presented, being red, blue, and yellow. Rest periods were also included. The question is whether there is a difference in activation in the primary visual cortex during these conditions. Performing a standard (whole-brain) group analysis might not show any significant voxels above threshold, as the neuronal differences between the conditions are very small. In order to detect the differences, we can define an area of interest, namely the primary visual cortex. This definition can be based on the Brodmann map or can be manually segmented using the anatomical landmarks. The potential problem with these selection methods is that the whole primary visual cortex is selected, while only a small part is activated during the task. For the VOI approach, this means that the Bonferroni correction is still too stringent. The less voxels there are, the lower the correction is. However, the selection of an area based on anatomical landmarks (or Brodmann atlas) is a methodologically sound approach, if you

do not have any prior knowledge of the exact location of the activation within the area.

Using the ROI approach, there are two possibilities: obtain the average b-value of the entire region, or take the maximum b-value of the region. The first method will result in a low average when the region is large and hence contains a lot of voxels which are not activated by the tasks. The second method seems more promising as this does not depend upon the size of the anatomically defined region. However, the location of the maximum b-values may differ for the three conditions. Comparing these values hence implies comparing different locations within the primary visual cortex to each other. The validity of this method is hence questionable.

Selection of areas based on the functional data itself has the advantage that only those voxels which are actually activated during the task are selected. A group analysis can be performed to obtain the areas which show task related activation. A problem may arise, when you compare two groups. Consider the following example. A patient group (n=20) and a healthy control group (n=20) are scanned during a motor task. In a particular voxel, the patients have a low value (b = -2) and the controls have a high value (b = 2). If you define the areas based on the total group activation (i.e. patients and controls together), then this voxel will not be selected, as it has a total group mean of 0. More general, using this group analysis approach there is a bias towards voxels which are activated (or deactivated) in both groups. The chance of finding a difference between the groups in these voxels is low. A possible solution is to lower the threshold, so that those voxels with a low total group average are still included.

References

Reference List

Aguirre,G.K., Zarahn,E., and D'Esposito,M., 1998. The variability of human, BOLD hemodynamic responses. Neuroimage. 8, 360-369.

Bandettini,P.A. and Cox,R.W., 2000. Event-related fMRI contrast when using constant interstimulus interval: theory and experiment. Magn Reson Med. 43, 540-548.

Buckner,R.L., Bandettini,P.A., O'Craven,K.M., Savoy,R.L., Petersen,S.E., Raichle,M.E., and Rosen,B.R., 1996. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging . Proc Natl Acad Sci U S A. 93, 14878-14883.

Dale,A.M., 1999. Optimal experimental design for event-related fMRI. Hum Brain Mapp. 8, 109-114.

Della-Maggiore,V., Chau,W., Peres-Neto,P.R., and McIntosh,A.R., 2002. An empirical comparison of SPM preprocessing parameters to the analysis of fMRI data. Neuroimage. 17, 19-28.

Friston,K.J., Fletcher,P., Josephs,O., Holmes,A., Rugg,M.D., and Turner,R., 1998. Event-related fMRI: characterizing differential responses. Neuroimage. 7, 30-40.

Friston,K.J., Frith,C.D., Turner,R., and Frackowiak,R.S., 1995. Characterizing evoked hemodynamics with fMRI. Neuroimage. 2, 157-165.

Henson,R.N., Price,C.J., Rugg,M.D., Turner,R., and Friston,K.J., 2002. Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. Neuroimage. 15, 83-97.

Hopfinger,J.B., Buchel,C., Holmes,A.P., and Friston,K.J., 2000. A study of analysis parameters that influence the sensitivity of event-related fMRI analyses. Neuroimage. 11, 326-333.

Liao,C.H., Worsley,K.J., Poline,J.B., Aston,J.A., Duncan,G.H., and Evans,A.C., 2002. Estimating the delay of the fMRI response. Neuroimage. 16, 593-606.

Maisog,J.M. and Chmielowska,J., 1998. An efficient method for correcting the edge artifact due to smoothing. Hum Brain Mapp. 6, 128-136.

Mechelli,A., Henson,R.N., Price,C.J., and Friston,K.J., 2003. Comparing event-related and epoch analysis in blocked design fMRI. Neuroimage. 18, 806-810.

Nieto-Castanon,A., Ghosh,S.S., Tourville,J.A., and Guenther,F.H., 2003. Region of interest based analysis of functional imaging data. Neuroimage. 19, 1303-1316.

Ogawa,S., Lee,T.M., Kay,A.R., and Tank,D.W., 1990a. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci U S A. 87, 9868-9872.

Ogawa,S., Lee,T.M., Nayak,A.S., and Glynn,P., 1990b. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magn Reson Med. 14, 68-78.

Pollmann,S., Dove,A., Yves,v.C., and Wiggins,C.J., 2000. Event-related fMRI: comparison of conditions with varying BOLD overlap. Hum Brain Mapp. 9, 26-37.

Ramsey,N.F., Hoogduin,H., and Jansma,J.M., 2002. Functional MRI experiments: acquisition, analysis and interpretation of data. Eur Neuropsychopharmacol. 12, 517-526.

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

Vink, M., Kahn, R. S., Raemaekers, M., Heuvel van den M., Boersma, M., and Ramsey, N. F. Function of striatum beyond inhibition and execution of motor responses. Human Brain Mapping . 2005a.
Ref Type: In Press

Vink, M., Kahn, R. S., Raemaekers, M., and Ramsey, N. F. Perceptual bias following visual target selection. Neuroimage . 2005b.
Ref Type: In Press

Wager,T.D., Vazquez,A., Hernandez,L., and Noll,D.C., 2005. Accounting for nonlinear BOLD effects in fMRI: parameter estimates and a model for prediction in rapid event-related studies. Neuroimage. 25, 206-218.

Worsley,K.J., Liao,C.H., Aston,J., Petre,V., Duncan,G.H., Morales,F., and Evans,A.C., 2002. A general statistical analysis for fMRI data. Neuroimage. 15, 1-15.

APPENDIX – some useful literature

Books on data preprocessing, analysis and (basic) physics:

J. Ashburner, K. Friston, and W. Penny (Eds.), Human Brain Function 2nd edition. http://www.fil.ion.ucl.ac.uk/spm/doc/books/hbf2/

P. Jezzard, P.M. Matthews, and S.M. Smith (Eds.), Functional MRI: An introduction to methods, Oxford University Press, Oxford.

C.T.W. Moonen and P.A. Bandettini (Eds.), Functional MRI, Springer, Berlin.

Book on statistics and matrix algebra:

J.P. Stevens, Applied multivariate statistics for the social sciences, Lawrence Erlbaum Associates, Publishers, London.

Articles:

on fMRI experiments (Ramsey et al. 2002)

on model specification details (Della-Maggiore et al. 2002)

on event-related designs (Hopfinger et al. 2000), (Pollmann et al. 2000)

on mixed-model group analysis (Worsley et al. 2002)

on ROI analysis (Nieto-Castanon et al. 2003)

on non-linearity in the BOLD signal (Wager et al. 2005)

on the difference of modelling blocks as events or blocks in SPM (Mechelli et al. 2003)

on the hrf-derivatives (Liao et al. 2002), (Henson et al. 2002), (Friston et al. 1998)

Websites:

Overview of statistics: http://www.fmrib.ox.ac.uk/fmri_intro/

fMRI for dummies: http://defiant.ssc.uwo.ca/Jody_web/fmri4dummies.htm

Page with links to all aspects: http://www.functionalmri.org/