2

Design of fMRI Experiments

In addition to the usual questions of experimental design that are relevant for any psychological experiment, fMRI studies present a unique set of design parameters that need to be determined. These parameters refer to the details of the data acquisition from the scanner. In this chapter we explore issues relating to the design of fMRI experiments from imaging and statistical perspectives.

2.1 Imaging Design Issues

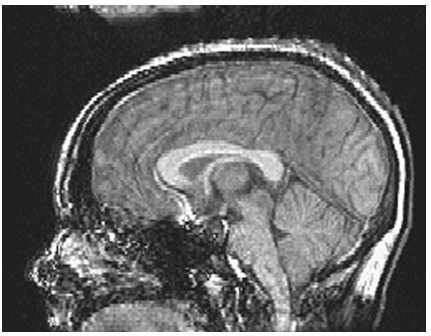
The imaging parameters are fixed at the start of any study. By changing the values of the imaging parameters, the technologist who operates the scanner controls the spatial and temporal resolution of the data, and hence the overall quality of the resultant images. It is standard to report the values of these parameters in the discussion of fMRI studies.

2.1.1 Description of Parameters

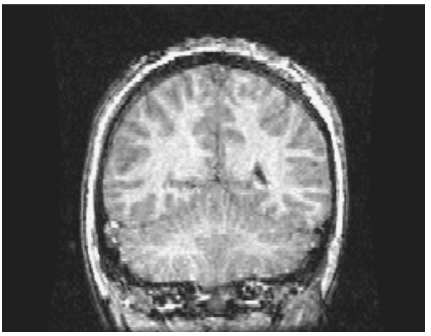
Images can be acquired in one of four *imaging planes*: axial, coronal, sagittal, and oblique. The first three are more common in practice. Axial slices are perpendicular to the longitudinal axis of the body – a series of axial slices will go from the top of the brain, down. Coronal, or frontal, slices are parallel to the front of the body – a series of coronal slices will proceed from the front to the back of the brain. Sagittal slices are obtained in parallel to the midline of the body – a series of sagittal slices will proceed from one side of the brain (say, the left side) to the other. Finally, an oblique slice is taken by tilting any of the three standard views. Examples of the three primary perspectives are shown in Figure 2.1.

The *field of view* (FOV) is the physical size of the image, measured in mm^2, and specifies the region from which the image was sampled. For instance, if the FOV is 20 cm in each direction, this means that the slices encompassing

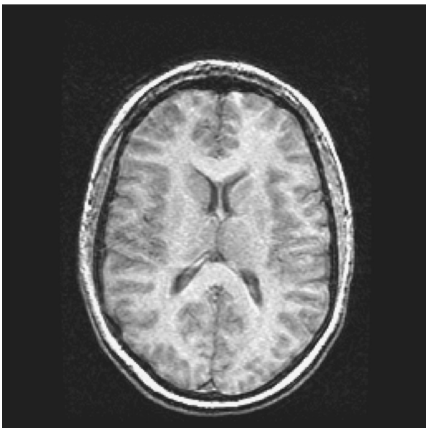
N.A. Lazar, *The Statistical Analysis of Functional MRI Data*,
DOI: 10.1007/978-0-387-78191-4_2,



(a) Sagittal slice



(b) Coronal slice



(c) Axial slice

Fig. 2.1. Examples of the three primary slice orientations. *Figures courtesy of Rebecca McNamee, University of Pittsburgh.*

the regions of interest are contained within a 20 cm x 20 cm region in the plane.

The *acquisition matrix size* is the size of the grid into which the plane of the FOV is divided for each slice. The acquisition matrix is usually square, often 64 x 64 or 128 x 128. The FOV and the size of the acquisition matrix determine the two dimensions of a voxel in the plane of a slice. Thus, if the field of view is 20 cm in a particular direction, and the matrix size is 64 in that same direction, voxels will be of length $200/64 = 3.125$ mm on a side. These are typical values.

Slice thickness is the thickness, measured in mm, of an individual slice. If the x and y dimensions of the three-dimensional object are given by the FOV

and the acquisition matrix, the z dimension is given by the slice thickness. The *slice gap*, also in mm, is the space between consecutive slices. The reason for leaving a gap between slices, historically, is that the radiofrequency pulses were imperfect, and there could be "crosstalk" between slices if there was no gap. This results in contamination of the signal, since the slices aren't perfectly separated from each other (Hashemi et al., 2004). A typical study might use slices that are 5 mm thick with a gap of 1 mm. Together with the FOV and matrix size values described in the previous two paragraphs, this results in voxels that are 3.125 mm x 3.125 mm x 5 mm, which are often seen in practice. With improvements in imaging techniques, many studies now do away with the slice gap altogether.

The order in which slices are acquired is fixed by the *excitation sequence*. While it might be intuitively appealing to image slices in a physically sequential fashion, this leads to crosstalk. Hence slices are usually acquired in an interleaved manner, for instance, all even slices, then all odd slices. Arbitrary sequencing is also possible with some software.

All of the parameters described above relate to the details of the data collection, such as the size of a voxel. Brown and Semelka (1995) call these *extrinsic parameters*. There are also *intrinsic parameters*, which affect the voxel's signal.

Repetition time (TR) is the time, in milliseconds, between successive applications of the radiofrequency pulses to a particular volume of tissue. Suppose then that we have applied a single 90^{o} pulse; the time until we apply the next one is the TR. What happens between the two pulses?

Immediately after the first pulse is applied, the magnetization flips from being aligned in the z direction to being in the (x, y)-plane, as was described in Chapter 1. Call the net magnetization in the (x, y)-plane $\mathbf{M}_0$. As soon as the pulse is turned off the magnetization in the transverse plane starts to decay, with a concomitant recovery of magnetization in the z direction, according to the formula

$$\mathbf{M}_z(t) = \mathbf{M}_0(1 - e^{-t/T_1}),$$

as we saw previously. At time $t =$TR, i.e., when we apply the *next* pulse (and hence flip the magnetization back into the (x, y)-plane),

$$\mathbf{M}_z(\text{TR}) = \mathbf{M}_0(1 - e^{-\text{TR}/T_1}),$$

which is less than $\mathbf{M}_0$, the initial magnetization in the z direction before the first pulse was applied. Once this second pulse is turned off, again there is recovery toward the z-axis, but as a new pulse will be applied at time $t = 2\text{TR}$, again the recovery will not be complete. That is, for each successive pulse, the system starts with less than total magnetization $\mathbf{M}_0$.

As the TR increases, there is clearly more time for the radiofrequency energy to dissipate, through the relaxation process described in the previous chapter, and we get closer to the initial magnetization $\mathbf{M}_0$ at the start of each successive pulse.

In terms of the received signal, at time $t = 0$ the signal is a strong FID, also as described in Chapter 1. At successive TRs the signal is still an FID, but with dampened amplitude. Ideally, one would want to measure the signal immediately after the RF pulse, with no delay. While this isn't feasible in practice, if we could do so, the FID signal would be proportional to $1 - e^{-\mathrm{TR}/T_1}$. Similar statements hold for each of the successive FIDs, that is, they are maximal if they can be measured right after the application of the RF pulse.

However, limitations on the hardware make it impossible to measure the signal immediately upon the application of the RF pulse. Instead, there is a brief waiting time between the originating pulse of the image, and the peak of the echo, that is, the acquisition of data from the center of k-space (the maximum of the signal). This short time is called *echo time* or TE, and is also measured in milliseconds. Now, the FID in the transverse plane decays at the rapid rate T_2^* according to the function e^{-t/T_2^*}. If we could take the measurement immediately, before any signal decay, the measured signal would be the initial magnetization, $\mathbf{M}_0$, flipped into the (x, y)-plane. After time TE the measured signal is slightly smaller, namely $\mathbf{M}_0 e^{-\mathrm{TE}/T_2^*}$.

Note that both processes of T_1 and T_2 relaxation are occurring simultaneously, so in fact the measured signal is proportional to

$$\mathbf{M}_0(1 - e^{-\mathrm{TR}/T_1})e^{-\mathrm{TE}/T_2^*}.$$

Thus, it is possible to manipulate the MR signal by changing TR and TE, which are under the control of the experimenter. Roughly speaking, TR is related to T_1 effects in the image, and TE is related to T_2^* (or T_2) effects. We will see this in more detail in the next section.

The *excitation* or *flip angle* sets the amount of rotation away from the equilibrium axis following the radiofrequency excitation pulse. The default value for the flip angle is 90^o in most scanners, since this gives the maximum magnetization in the transverse plane. Excitation angle is the other parameter, along with TR, that determines the amount of T_1 weighting in the image.

2.1.2 How Are Resolution and Image Quality Affected by Changes in the Parameters?

Many of the acquisition parameters, such as the flip angle, the size of the acquisition matrix, and the field of view (and hence the voxel size), are fixed by convention, although not by necessity. Others, such as TR, are given more to the control of the experimenter. Yet, in any case it is important to be aware of the tradeoffs in terms of resolution and image quality (for instance, as measured by signal to noise ratio), as well as in terms of total scan time, that result from changes in the values of these parameters. A useful summary of the effects of manipulating the different values on resolution, signal to noise ratio, and scan time can be found in Brown and Semelka (1995).

We consider first the direct effect of manipulating TR and TE. This is seen via the contrast between two different tissue types, call them A and B, with different values of T_1 and T_2^*, which we will denote T_{1A}, T_{1B}, T_{2A}^*, and T_{2B}^*, respectively. The contrast between the two tissue types is given by

$$c_{AB} = \mathbf{M}_{0A}(1 - e^{-\mathrm{TR}/T_{1A}})e^{-\mathrm{TE}/T_{2A}^*} - \mathbf{M}_{0B}(1 - e^{-\mathrm{TR}/T_{1B}})e^{-\mathrm{TE}/T_{2B}^*}.$$

Now, suppose that $T_{1A} > T_{1B}$, in other words, tissue type A has a longer T_1 recovery time than does tissue type B. Put another way, it takes tissue A longer to reach equilibrium than it takes tissue B, so that at any given point in time t the recovery curve for A will be below that for B. Furthermore, at different points in time, the distance between the two recovery curves will be different, as demonstrated in Figure 2.2 – near $t = 0$, the curves will be close together (no recovery in either case), there is no signal from either tissue, and $c_{AB} = 0$. As t goes to infinity, the curves will be close together again (full recovery in both cases), the effect of T_1 is reduced (in the limit vanishing altogether). So, at very short or very long TRs, the contrast between the tissues is not large. Between those two extremes there is a difference between the tissues. Since we have assumed that T_{1B} is the smaller of the two T_1 times, tissue B recovers more quickly and hence has the stronger signal at intermediate TRs. Indeed, for any two tissues that differ in their value of T_1, we have therefore shown that there is a time point that maximizes the distance between the two recovery curves, giving optimal contrast, and in general, shorter TRs will give greater T_1 contrast.

A similar analysis shows that shorter TE values diminish the effect of T_2^* (or T_2), whereas longer TEs enhance the contrast from T_2^*. Often, in the case of fMRI, the TE is set at the gray matter T_2^* time, since this enhances the BOLD effect, although recent studies have employed variable TEs (see, for example, Chen et al. 2003a).

The above results can be used to weight images differentially, according to T_1, T_2, or T_2^*. T_1 weighted images will result from an experiment with an intermediate TR and a small TE. Tissues that have long T_1, such as cerebrospinal fluid, will be downweighted and hence will not be as visible in the image, as tissues with shorter T_1 times, such as white matter. T_2 weighted images are often used as structural reference points in functional studies. These are generated from a long TR and a medium TE; the resultant images highlight regions of cerebrospinal fluid and downweight the white matter. Similar to T_2 contrast, T_2^* contrast is obtained by long TR and intermediate TE. Images weighted in this way are sensitive to the amount of deoxygenated blood in the tissue, which is itself a function of the changing metabolism of the active neurons. In general, it is possible to take advantage of the T_1 and T_2 times of different types of tissues, to tailor images that highlight the areas of interest (see Hashemi et al. 2004, for more detail).

Another important aspect of an fMRI experiment is to control the *signal to noise ratio* (SNR). In fMRI, SNR is proportional to the product of volume of a voxel and $\sqrt{N_y \mathrm{NEX}/\mathrm{BW}}$, where N_y is the number of phase encoding steps

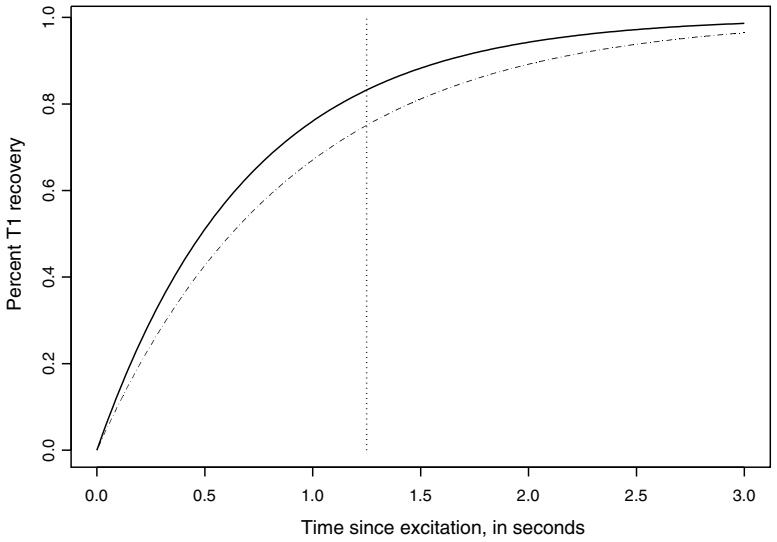


Fig. 2.2. The differences in relaxation times for different tissues can be exploited to weight images, in such a way that different aspects of the brain are emphasized.

(that is, in the y direction), *NEX* is the number of excitations (the number of times the scan is repeated) and *BW* is the bandwidth (the range of frequencies in the slice selection). From this formula a number of things are clear. First, as we increase the voxel volume, the SNR increases linearly. This is reasonable, since increasing the size of a voxel also increases the number of protons in a voxel, and the measured signal is related to the number of protons. Also, as the number of excitations increases, so too does the SNR, however, the increase is not linear. Rather, it follows a square root law: increasing NEX by a factor of 2, for instance, increases SNR by a factor of $\sqrt{2}$, since the signal doubles but the noise goes up only by a factor of $\sqrt{2}$ (typical variance behavior). In a similar fashion, doubling the number of phase encoding steps, N_y, increases the SNR by $\sqrt{2}$. The last component in this formula for SNR, the bandwidth BW, enters in an inverse relationship. Moving to a larger bandwidth increases the noise, and diminishes SNR. Again, the change follows a square root law.

Different ways of looking at SNR provide insight into the various tradeoffs among the parameters. SNR depends on two elements: the volume of an individual voxel and the total time to sample all signals. We can rewrite the expression for SNR analyzed in the previous paragraph, as follows. The size of a voxel along the y-axis, call it Δy, is equal to the field of view (FOV) in the y direction divided by the number of phase encoding steps. Similarly, Δx is the FOV in the x direction divided by the number of frequency encoding steps,

N_x. Hence the volume of a voxel can be rewritten as $\mathrm{FOV}_x\mathrm{FOV}_y\Delta z/N_xN_y$. Combining this with the previous expression for SNR, we have that SNR is proportional to

$$(\mathrm{FOV}_x/N_x)(\mathrm{FOV}_y)\Delta z\sqrt{\mathrm{NEX}/N_y\mathrm{BW}}.$$

With this formulation, we can explore the tradeoffs between resolution of the image and signal to noise ratio. Namely, for a fixed FOV, as we increase the number of phase encoding steps, SNR decreases. Conversely, if the FOV increases along with N_y (thus keeping the voxel size, and the spatial resolution, constant), SNR increases. However, there is a tradeoff in that the acquisition time will also increase. Increasing the slice thickness increases signal to noise ratio, but decreases resolution.

The scan, or acquisition, time is given by $\mathrm{TR} \times N_y \times \mathrm{NEX}$. Looking now at the effect of modifying the TR, it is clear that increasing TR increases the scan time. It also increases coverage – with a longer TR, it is possible to acquire more slices of the brain in a single scan. And, increasing TR leads to better SNR. On the other hand, increasing TE has no effect on the scan time, but does cause a worsening of SNR.

In summary, there are tradeoffs between the various desiderata from an imaging perspective – good signal to noise ratio, short scan times (to prevent discomfort of subjects in the scanner, for instance), and adequate spatial resolution. It is not always possible to achieve all of these goals simultaneously, and it is the work of MR technologists to figure out what parameter settings will balance the differing needs of a given experiment. Configuring the scanner parameters appropriately is a key component of fMRI experimental design.

2.1.3 Filling in k-Space

Another aspect of fMRI design relates to how the k-space data are themselves acquired – the planning of *pulse sequences*. Here, too, there are various tradeoffs, each trajectory through k-space having advantages and disadvantages. The planning of optimal sequences is still a topic of research among MR physicists. This section describes two of the more common types of data acquisition for fMRI: echo-planar imaging (EPI) and spiral imaging.

With *echo-planar imaging*, the gradients move through k-space in a *boustrophedonic* pattern, i.e., using alternate left to right and right to left lines (literally, as the oxen turns while plowing). EPI, originally proposed by Mansfield (1977), is the fastest MRI imaging technique currently available (Hashemi et al., 2004). However, it requires specific hardware to perform, namely, special gradients that can be turned on and off rapidly. *Single-shot EPI* allows for all of k-space to be filled in following a single RF pulse, whereas other fast imaging techniques require multiple pulses. In the most widely used version, "blip EPI," a large phase encoding gradient places the first echo at the edge of k-space. The readout, or frequency encoding, gradient forms a trail

of echoes, that is, we acquire a line of k-space data in the x-coordinate. The direction of the trail alternates back and forth, as subsequent phase encoding "blips" move the gradient along to the opposite side, thus covering all of k-space. In this version of EPI, the phase encoding gradient is turned on only when the readout gradient is 0 (i.e., at one or the other end of the k_x-axis in k-space). An advantage of the blipped EPI, compared to earlier versions of EPI in which the phase encode gradient was on continuously, is that the resultant trajectory through k-space is truly rectilinear. This makes it easier to Fourier transform the data to obtain an image. Figure 2.3 shows the k-space trajectory for blipped EPI.

Fig. 2.3. The k-space trajectory for standard, single-shot blipped echo-planar imaging (EPI).

It is worth emphasizing a technical point here, namely that since all of k-space is filled in following a single RF pulse, the data must be acquired quickly, before there is significant T_2 or T_2^* decay. On the other hand, the experimenter also wants to sample a large part of k-space in order to have adequate spatial resolution of the images, and this takes time. That is, we see again the types of tradeoffs that appear so frequently in fMRI experimental design. The way the question manifests itself here is that there is a need to compromise on the number of lines of k-space that are acquired over the space in question, a sacrifice of spatial resolution to the demands of time. Typical EPI images will be 64 x 64, or at most 128 x 128 voxels in a slice, whereas other pulse sequences can produce images of as much as 512 x 512 pixel resolution. The latter take longer to acquire, since they use more than a single pulse to

acquire all of k-space. A single image using EPI can be acquired in 30 to 50 milliseconds, with a whole volume scan (multiple slices) in only 2 to 4 seconds. By contrast, a particular sequence called FLASH, with resolution of 256 x 256 pixels, acquires a single image in 2.5 to 10 seconds, and a multislice volume can take up to 4 minutes (Jezzard and Clare, 2001). The limiting factor on EPI is the ability to rapidly alternate the gradients back and forth, which seriously taxes the hardware.

For single-shot EPI, TR refers to the time between successive images. The speed of single-shot EPI is offset by a number of drawbacks, aside from the relatively low spatial resolution and the stress the method places on the gradient hardware. First, the data that are collected during the transitions from one line of k-space to the next are not used in the creation of the images, so the technique entails an intrinsic loss of data. Second, EPI data are prone to various artifacts, in particular, *susceptibility artifacts*, which manifest themselves as distortions in the interfaces between air sinuses and brain tissue. Third, the long readout time of single-shot EPI may result in *geometric distortions*, which are seen as stretching or shearing when there is a distortion in the (x, y)-plane, or a dampening of signal when the distortion is in the z direction.

Multi-shot EPI is similar to single-shot EPI, with the main difference that the readout is divided into multiple segments. The segments are acquired in an interleaved manner, with one shot applied to each, as demonstrated in Figure 2.4. This variant of EPI puts much less stress on the gradient hardware, but images take longer to acquire.

Spiral imaging, like EPI, is a fast image acquisition technique. It differs from EPI in how the gradients are used to traverse k-space. Whereas EPI takes a rectilinear trajectory, based on rapidly switching gradients, spiral imaging uses sinusoidal gradients to induce a spiraling path through k-space. The spirals may be from the edge of k-space inward, or from the center of k-space out (see Figure 2.5). It is also possible to combine the two trajectories in a "spiral in/out" pattern, which increases SNR and decreases susceptibility artifacts (Glover and Law, 2001); the advantages of the spiral in/out method over conventional spiral methods appear to be greater at higher field strengths (1.5T versus 3T) for a variety of tasks that activate a range of brain regions that are particularly susceptible to artifacts (Preston et al., 2004). Spiral imaging induces less stress on the gradient hardware, since it samples k-space continuously, rather than switching back and forth as in EPI. The continuous sampling also means that all the k-space data are used in creating the image. Finally, spiral imaging is faster than EPI. A disadvantage of spiral acquisition is that the data no longer fall on a Cartesian grid, hence, in order to apply the Fourier transform, the data must be *resampled* prior to image reconstruction. This involves interpolating the data in order to map from a spiral to a rectangle. Like EPI, spiral imaging is prone to certain artifacts arising from inhomogeneities in the magnetic field, and spatial distortions, although these can be mitigated by using the spiral in/out trajectory. The patterns of the

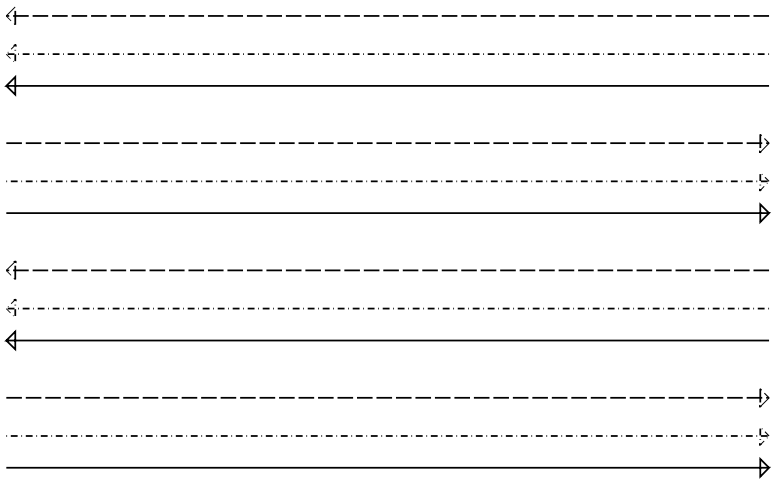


Fig. 2.4. The k-space trajectory of multi-shot EPI.

artifacts are somewhat different for the two types of trajectories, however (Jones et al., 2001).

2.2 Statistical Design Issues

The other aspect of experimental design for fMRI is statistical, namely, how should the study itself be carried out? While this question is common to all scientific studies, fMRI presents some interesting challenges.

2.2.1 Common Experimental Designs

There are two main approaches to the design of fMRI experiments, from the perspective of stimulus presentation. The first, *block design*, will be familiar to statisticians as a traditional way of designing an experiment. The second, usually called *event-related design*, arises from functional neurology studies, in particular, *event-related potentials*, or ERP. More recently, hybrid, or mixed, designs, which combine aspects of block and event-related, have been used.

Traditionally, due to limitations in resolution, fMRI experiments utilized simple block designs, in which periods of rest (or fixation) alternated with periods of task, or periods of different tasks were alternated. This was necessary in order to accumulate enough data to make a statistical analysis feasible.

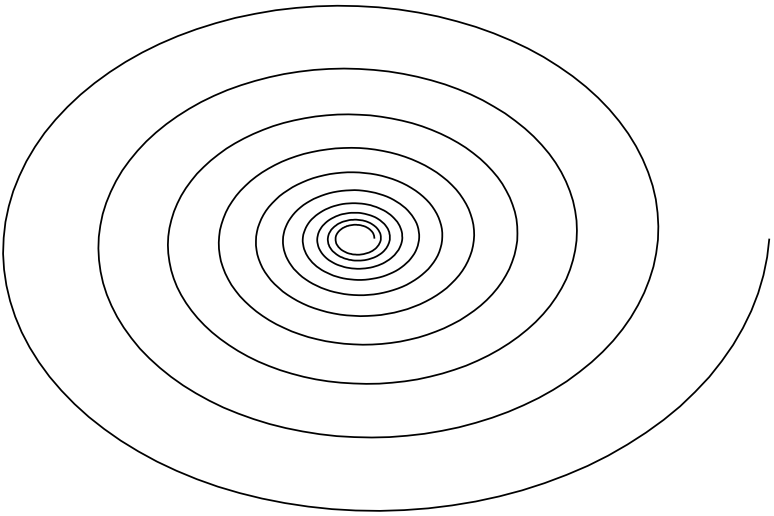


Fig. 2.5. The k-space trajectory for spiral imaging.

These "boxcar designs" (so-called because of their "on-off" nature, which can be depicted graphically as "up-down"; see Figure 2.6) lend themselves to a variety of statistical approaches. Block design experiments are easy to carry out, with presentation of the stimulus taking place in blocks of a fixed length, say 30 seconds. In each block one stimulus type is presented. A simple example of this type of experiment is the presentation of a flashing checkerboard during the task blocks, alternating with blocks of fixation, in which no visual stimulus is presented. More complex designs to accommodate more than one type of task are, of course, possible.

Block designs are powerful for locating voxels in which the level of activity is significantly different in the task versus the control conditions. To understand this recall the hemodynamic response described in Chapter 1. Following presentation of a stimulus there is a gradual rise in the signal until a peak is reached, after which the system returns to baseline levels in the absence of further stimulation. In a block design there is constant stimulation for the duration of the task blocks, meaning that the hemodynamic response does not return to baseline during this time. Instead, as the stimulus is repeatedly presented, the hemodynamic response in the active voxels accumulates, rising to a plateau instead of a short-lived peak. Decay back to baseline occurs only when the stimulus presentation is turned off, that is, during the control blocks. Voxels that are not active do not exhibit the characteristic response, and so

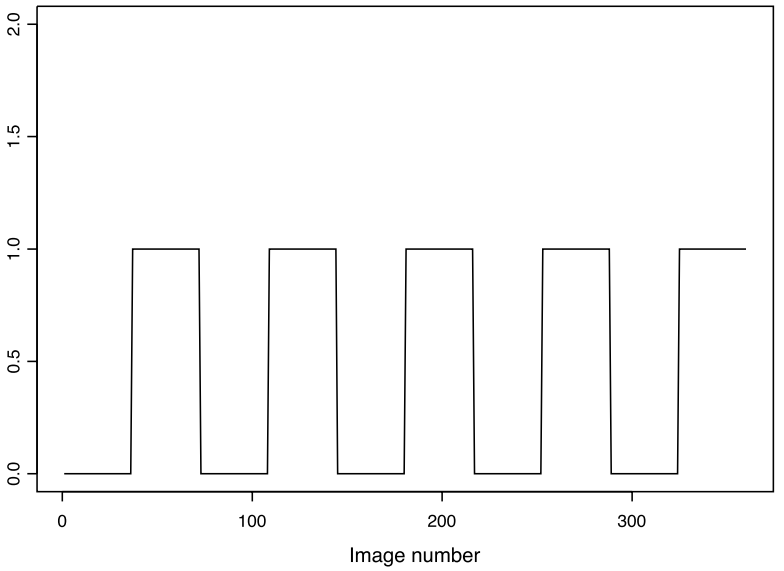


Fig. 2.6. Stimulus path of a simple block design. Blocks of the experimental task alternate with blocks of rest, or control. Of interest is the comparison between levels of activation during the task and during rest.

will not show increased levels of signal during the task blocks compared to the control blocks.

As should be evident even from this brief discussion, block design experiments are amenable to a range of statistical analyses, most of which involve comparison of (time-averaged) activity across experimental conditions. Since analysis is driven in large part by comparison, much of the focus in devising a block design experiment is on finding suitable controls. Ideally, the scientist wants the conditions to differ only in the variable of interest, that is, the control task should be identical, insofar as possible, to the experimental task. With complicated, and even not so complicated, cognitive processes such as are measured in a typical fMRI experiment, it isn't always clear what an appropriate control should be.

A study by Newman et al. (2001) demonstrates the potential impact of the baseline, or control, condition on the results of an fMRI experiment. They compared three baselines – rest, passive, and task-related – against the same task. The *rest* baseline, in which the subject lies in the scanner with no stimulus being presented, is the simplest, and is often used as the default control condition. In the *passive* baseline, stimuli are presented just as in the task condition, but subjects are instructed not to process them. As pointed out by Newman et al., it is impossible to know whether or not subjects truly

follow the instruction not to process the stimuli, but the rationale behind this control is clear. Finally, the *task-related* baseline presents the subject with a well-defined task. The experimental task in the study at hand was one of "phoneme discrimination" – subjects were presented with pairs of three letter nonsense words (consonant-vowel-consonant) and had to decide whether or not the two nonwords in each pair ended with the same sound. The three baseline (control) conditions were rest (relax and don't think of anything), passive listening (listen to the nonsense words but don't do anything in response), and monitoring tones, in which subjects were presented with triplets of high and low pitched tones, and had to judge if the final tone was high pitch. Note that the third baseline condition is similar to the experimental task, but with tones replacing the nonsense words. The authors found that there were clear differences in the activation patterns, depending on the control that was used. In general, more activation was detected in the resting baseline condition and less in the passive listening baseline. In related work, Marx et al. (2004) found that the choice of rest baseline – eyes open versus eyes shut – was critical for a simple visual task.

The event-related design, known also as *single-trial fMRI*, moves away from blocking the experimental conditions. In these studies, trials, or stimuli, are presented individually, separated by an *interstimulus interval* (ISI), which can be fixed within an experiment, or may vary from trial to trial. For example, instead of showing a flashing checkerboard continuously for 30 seconds, as we would do in a blocked design, in an event-related design, the checkerboard is flashed only once, for a short period of time. Another short checkerboard burst may follow some time later, or perhaps a completely different stimulus may be presented (Figure 2.7). This paradigm greatly expands the flexibility of the fMRI experiment, since researchers are no longer bound by the constraints of a formal block design. For instance, it is possible to let the stimulus presentation depend on the response of the subject: A correct response to a question could lead to a more difficult question on the next trial. Studies can be "self-paced" in the sense that the subject himself controls when stimuli are presented (e.g., Maccotta et al. 2001). Stimuli can also be presented randomly, again in contrast to the block design, wherein the stimulus within a block is fixed, and stimuli across blocks alternate. Furthermore, if the ISI is long enough, the neural activation following a stimulus will return to baseline, hence the event-related paradigm allows researchers to learn about the hemodynamic or BOLD response (the time course of activity) at a single voxel. This is not feasible with a block design, as it averages over hemodynamic responses, thereby blurring the individual features. Therefore, whereas block designs are good for *detection* of activated voxels, event-related designs are more effective at *estimation* of the hemodynamic response function. Another feature of event-related analysis is that it allows for separation depending on the response to the task (for instance, trials in which the subject responded correctly versus trials in which the subject responded incorrectly). As a result

of the greater flexibility afforded by event-related studies, their statistical analysis is often more challenging.

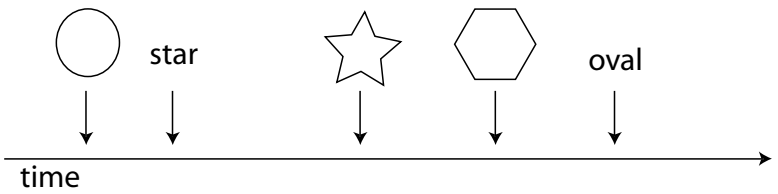


Fig. 2.7. Schematic of an event-related fMRI design. Instead of a "boxcar" describing the stimulus series, as in the block design, the event-related design can be described as a sequence of (possibly) irregularly spaced spikes. Stimuli of different types are presented in random order and at different time lags. In this example, a researcher might be interested in the differences between visual and verbal presentations of shapes.

A major design issue in event-related studies is the interstimulus interval, specifically the optimal length of the ISI, and whether it should be fixed or random. Early studies (for instance, Buckner et al. 1996) used constant ISI and allowed the hemodynamic response to fully evolve; thus, for a stimulus of duration 2 seconds, the optimal span of time between the end of a stimulus and the beginning of the next was 12 seconds (Bandettini and Cox, 2000). The resultant design – 2 seconds of stimulus followed by 12 seconds of ISI in which the subject would fixate, repeated multiple times – has the flavor of a "quasi-block" design and doesn't fully exploit the flexibility of the event-related paradigm. In addition, such studies are time consuming since the stimulus is presented very infrequently, they lack statistical power, and subjects become distracted during the long ISI. On the other hand, as noted by Bandettini and Cox (2000),with shortened constant ISI, there is an associated decrease in information (functional contrast) compared to the traditional block design.

Design with variable ISI is a more complex question. Early studies with variable ISI revealed somewhat contradictory results regarding how often the stimulus or task should be presented and how rapidly the design should alternate between task and control conditions. Dale (1999) found that the optimal design for hemodynamic estimation is one that alternates rapidly between conditions (short average ISI) and varies the time that passes between them (varying ISI). Friston et al. (1999c) explored a range of designs, from the

purely stochastic to the purely deterministic (block design), and found the latter to be the most efficient. The difference in findings, as pointed out by Birn et al. (2002), could be postulated to lie in the different goals that were being assessed in the two studies: estimation of the hemodynamic response function in the first case, and detection of active voxels in the second. Birn et al. (2002) examine the situation in more detail via simulation. They place the block design and the event-related design with short ISIs as two extremes – the former has stimuli presented continuously over a (relatively) long period of time, alternating with long periods of control; the latter switches frequently between task and control. In between these two, with variable ISI, it is possible to have a wide range of designs, and in all cases keep the proportion of time spent in the task condition constant. As expected from the discussion above, and indeed not surprisingly, their simulation studies reveal that there is no one optimal design when using variable ISI; rather, the "best" design depends on the goals of the study (detection versus estimation).

This issue is explored more fully by Liu et al. (2001), who develop a theoretical framework of the relationship between detection power and estimation efficiency. The tradeoff between the two is fundamental, in the sense that increasing one perforce degrades the other. Hence it is not possible to design a study, no matter how the ISI and stimulus presentation are manipulated, that will be able to achieve both goals simultaneously and efficiently. An additional factor included in their model is the "predictability" or perceived randomness of the design. Block designs have a high degree of predictability, since they alternate between two conditions, opening up the possibility (or even certainty) that subjects will anticipate stimuli, with concomitant confounding on the effects of interest. By deriving bounds on the estimation efficiency, Liu et al. demonstrate theoretically that this quantity is maximized for the case of two conditions (stimulus and control) when each condition is equally likely and the stimulus time course (i.e., the string of "stimulus" and "control" labels that describe the state of the experiment at each time point) is obtained as a sequence of Bernoulli trials, with probability $p = 0.5$ of success. Likewise, by deriving theoretical bounds on the detection power, they show that this quantity is maximized in block designs, again with equal probability of being in the task and control conditions. Similar results hold when there is more than one trial type (Liu and Frank, 2004).

In between the two extremes of maximizing detection power at the expense of estimation efficiency and maximizing estimation efficiency at the expense of detection power lie a range of what Liu and colleagues term *semirandom designs*, which have the potential to be useful when both goals are desired in the same experiment. The cost of trying to achieve a balance between the two extremes via a semirandom design is time; the semirandom designs take longer because the researcher is now trying to both estimate and detect. Essentially, in the semirandom design the probability of being in the task condition is allowed to vary over time, as opposed to being fixed, as in the purely random design.

Finally, the relationship between predictability and the other two quantities is examined. Liu et al. (2001) demonstrate that as the ability to estimate the hemodynamic response goes up, average predictability goes down. This makes sense if we recall that random designs are optimal for estimation, and random designs are less predictable, by definition. Another finding is that small increases in predictability may be worthwhile, since they can lead to gains in detection power without seriously impairing the estimation efficiency.

The design of event-related studies is an active field in the fMRI community and the best way to learn about the most recent state of the art ideas is to read the current literature. To wit, we describe one final twist on the topic, studied by Visscher et al. (2003), the *mixed block/event-related design*. In this design, blocks of task are alternated with blocks of control, as in a standard block design. However, within the task blocks, trials are assigned at random (that is, with varying ISIs), as in an event-related study. The advantage of this design, as the authors showed in an extensive simulation study, is that it allows for the separation of transient activity, related to the stimulus as it is presented, from sustained activity, which carries across tasks and stimuli. Many other design variants are possible (Liu, 2004).

Which of these various designs should be used in a given study? It is evident from the discussion above that there is no one "correct" design, and even optimality of a chosen design is contingent on the specific questions of scientific interest. My experience has generally been that the design of the experiment, and particularly the choice between block design and event-related design, depends more on the constraints of the study than on purely statistical considerations, which in any case can accommodate any of the differing paradigms described above. To some extent scientific trends play a part as well – as event-related, mixed, and semirandom designs become more popular and more accepted in the literature, researchers are going to want to exploit their strengths and flexibilities in order to get the most out of the data. As statisticians, it is important that we be aware of the latest designs, their advantages and disadvantages, and steer our neuroimaging colleagues away from designing experiments that will not allow them to answer the questions in which that truly interest them.

2.2.2 Additional Issues

In this section, a number of miscellaneous issues relating to the statistical design of fMRI experiments are explored.

We start with two specific questions on the use of block designs: first, the timing of data acquisition within the block, and second, identifying activation that is the result not of the task, but of the transition from task to control or vice versa. These questions are of potential impact on both the statistical analysis and the interpretation of block design studies.

Veltman et al. (2002) look at the first question in the context of language processing. Conventionally, the stimulus is presented at the onset of a TR,

but Veltman and colleagues argue that this might produce bias. To check this, they varied the length of the fixation condition in a block design study, thereby affecting the relationship between stimulus presentation and the TR; instead of the two always coinciding, the presentation of the stimulus was shifted relative to the onset of the TR. They found that effect sizes were larger in the "no shift" condition than in the conditions where the stimulus presentation was decoupled from the TR. Furthermore, activation was detected in some areas only at certain timing conditions, and not others. The authors conclude that even for block designs, it is important to distribute the sampling of the hemodynamic response, as is done in event-related studies.

The question of activation during transitions between conditions in the block design is taken up by Konishi et al. (2001). They found a set of regions that were transiently activated at the transitions between blocks, consistently for a variety of conditions involving different visual stimuli (verbal and facial) as well as different trial presentation rates. Interestingly, these regions did not always coincide with regions in which task-related activation was detected by the usual statistical analysis. Moreover, the size of the transition effect was similar in all four of the conditions considered by the authors. Perhaps most significant, the transient activation was sometimes detected even when there was no detected activity in the relevant task block, indicating that it is not the activation in the task block alone that is pertinent.

We next turn to efficient and optimal experiment design, in particular for event-related studies, a focus of two recent papers. Block designs can be incorporated into the framework of finding a good design, since they can be seen as one extreme of a continuum of designs. The search for an optimal design involves an exploration of the space of possible experimental setups, which can be accomplished in many different ways. When we discuss "sequences" below, the intention is to reference the string of conditions representing the presentation of stimuli at each time point. For conceptual simplicity, think of each condition in the experiment being assigned an integer value. Then a design is a string, or sequence, of integers.

Buračas and Boynton (2002) consider efficient estimation of the hemodynamic response curve, noting that, by use of a more efficient design it is possible to reduce scan time without loss of signal, compared to a less efficient design. They point out that researchers often pick an event-related sequence of trials in a rather arbitrary fashion, namely, generating many sequences at random, and picking the one that yields the best estimation efficiency. Clearly this is not a satisfactory approach in general, as researchers are left without any guidelines for choosing the design sequence in their next study, and of course there is no guarantee that the best sequence from among a set of randomly generated sequences is optimal in any other sense. Buračas and Boynton propose instead to use *maximum length shift register sequences*, or *m-sequences*, to find good stimulus presentation sequences at little computational cost (unlike the random sequence approach). The emphasis here is on identifying the specific sequence(s) that should be used in an event-related

study. m-sequences are sequences of integers (each integer representing here a condition type, for instance, task or control) which are generated using modulo arithmetic, where the modulus depends on the number of conditions. Sequences are created using the formula

$$s_k = \sum_{i=1}^{r} c_i s_{k-r+i-1},$$

where c_i are set coefficients, s_k is the next value in the sequence (appended to the previous $s_1, \ldots, s_{k-1}$), and r is the *order* of the shift, the number of previous terms included in generating a new element of the sequence. For the simplest fMRI experiments, with only two conditions, both c_i and s_k take on the values 0 and 1, and the generation of new elements uses arithmetic modulo 2.

Buraccas and Boynton show, based on simulations, that generation of event-related designs via m-sequences results in higher estimation efficiency than randomly generated designs, especially for shorter sequence lengths. Even using as many as one million random designs and picking the best among these, estimation efficiency was less than for the m-sequence. Having more than one event type also improves the advantage of m-sequences over random sequences. Indeed, Liu (2004) proves that m-sequences come close to attaining his theoretical upper bound on estimation efficiency and have low predictability, although they have low detection power. The drawback of m-sequences is that there is less flexibility in design, since length and type of allowable sequence are restricted. In practice, the authors claim that this is not any real barrier to using their approach, since the class of acceptable m-sequences is large.

Wager and Nichols (2003) propose the use of a genetic algorithm to accomplish the search over design space. Genetic algorithms start with an initial set of designs and evolve new designs via three steps that mimic the ways changes occur in DNA: *selection*, *crossover*, and *mutation*. Selection of designs is akin to natural selection – the genetic algorithm tests the initial set of designs (which are usually randomly generated) according to some prespecified goodness criterion, selects the best ones, and creates "offspring" from them (new designs). In this way, the best features of the existing designs are passed on. Crossover is similar to the biological exchange of DNA across chromosomes – two designs will exchange "material," or sequence patterns, from a randomly chosen point onwards. And in mutation an element of the design sequence is randomly switched to take another value. These processes are iterated until an appropriate model is identified. Wager and Nichols consider three goodness criteria: effectiveness at detection, estimation efficiency, and counterbalancing of the design. The genetic algorithm is flexible enough to optimize multiple criteria simultaneously (this is accomplished by creating a new, weighted, score, which combines individual criteria), and can also take account of both statistical and psychological requirements (for instance, it might be undesirable

to have the same stimulus repeated too many times in a row, for reasons of anticipation on the part of the subject; the genetic algorithm can incorporate such constraints by rejecting designs that have unwanted properties).

Via a series of simulation experiments, the authors demonstrate the efficacy of the genetic algorithm over random search, although they note that it is hard to achieve all three goodness criteria with one design, in agreement with findings by others. They are also able to give specific recommendations on the length of the ISI in an event-related design, and the length of a block for a block design (subject to the particular model assumptions made in their simulations), suggesting that the use of a genetic algorithm coupled with simulation will be a useful pilot tool in the early stages of planning an experiment. In this scenario, the researcher would run versions of the genetic algorithm with different assumptions and different weighting of the criteria of interest, specific to the research questions at hand, to arrive at an optimal experimental design.

As a final point we mention briefly the problem of using fMRI in clinical populations (that is, groups of subjects with neuropsychological disorders, such as autism, schizophrenia, or Alzheimer's disease) and children, a theme that we will revisit in Chapter 3. These groups present special challenges, among them the need for the researcher to devise experiments that the subjects will be able to perform. For example, it might not be reasonable to expect children or patients with attentional difficulties to carry out tasks that put a heavy load on memory. Choice of task, then, should be suited to the experimental group of interest, as discussed in Jessen et al. (2002)for clinical groups and by Gaillard et al. (2001)(see also references therein) for children.

http://www.springer.com/978-0-387-78190-7

The Statistical Analysis of Functional MRI Data

Lazar, N.

2008, XIV, 299 p., Hardcover
ISBN: 978-0-387-78190-7