A Practical Guide to an fMRI Experiment

Nasser H. Kashou

Wright State University - Main Campus, nasser.kashou@wright.edu

Follow this and additional works at: http://corescholar.libraries.wright.edu/bie

Part of the Biomedical Engineering and Bioengineering Commons, and the Industrial Engineering Commons

Repository Citation
http://corescholar.libraries.wright.edu/bie/107
1. Introduction

Functional Magnetic Resonance Imaging (fMRI) has been around for two decades and research in this field has been exponentially rising. Much of this research has been dominated by basic science. Recent trends have brought the clinical realm into play in which valuable contributions can still be made. Helping the clinician understand the basic concepts behind an fMRI experiment is crucial to further developing and evaluating functional paradigms and research. Critical to designing an fMRI experiment is understanding the related physics and how fine tuning scanning parameters affects the image quality, which in turn affect the findings of an fMRI study. In addition, understanding the physiology behind the acquired signal and anatomy of the brain is also important. To appreciate the complexity of the fMRI process see (Amaro & Barker, 2006; Savoy, 2005).

In this chapter we present a practical guide to the novice on the important aspects needed to perform an efficient fMRI experiment from idea formulation to understanding the possible limitations of the results. The basic concepts of fMRI, beginning with image resolution and physics, will be discussed along with advice on possible “pearls” and “pitfalls” of this process. Points covered will include: paradigm design, scanning protocol, and limitations.

2. Basic physics

How is an image acquired in MRI? In this section a brief overview of the physics and steps needed to generate an image is introduced. The main components in acquiring an image in MRI are a magnet, three gradients and a radio frequency (RF) coil. The magnet strength can range anywhere from 1.5 to 7 (Besle et al., 2013; Duchin et al., 2012; Yacoub, Harel & Ugurbil, 2008) and 8 Tesla (Novak et al., 2001, 2003) (and even higher for animal systems (Yacoub, Uludag, Ugurbil & Harel, 2008)). Currently, 1.5 and 3 Tesla are the standard strengths used in the clinical environment MRI magnets. To get a grasp on the strength of the magnet, consider that the Earth’s magnetic field is equal to 0.5 Gauss and 10,000 Gauss is equal to 1 Tesla. This means when working with a 1.5 or 3 Tesla system the magnet is 30,000 and 60,000 times stronger than the Earth’s magnetic field. Because of the intensity of the magnetic field, it is critical that ferrous material never be brought in or near the MRI scanner room. This is the most important thing to know when working with an MRI scanner. The MRI magnet itself
cannot provide images without the two other components: the RF coil and three gradients \((G_x, G_y, \text{ and } G_z)\). The RF coil is used to send a pulse at the same precessing frequency of the hydrogen nuclei (see below). In some systems the RF coil both transmits a pulse and receives a signal called the free induction decay (FID). The gradients are turned on and off to cause a slight gradient increase in the magnetic field. All together these components initiate the process of acquiring viewable images. They also form the physical basis of scanning sequences.

What is being measured?

The fMRI process measures the interaction of protons, specifically the hydrogen nuclei from water molecules in the magnetic field. The interaction of protons is called nuclear magnetic resonance (NMR). Protons spin in a manner analogous to tops in that they have an orientation and a frequency and are precessing at an angular frequency, \(\gamma B_0\) in a magnetic field \(B_0\) where \(\gamma = 4.256 \text{MHz/}T\) is a proportionality constant called the gyromagnetic ratio. These protons are randomly oriented in our bodies, Fig. 1.

Figure 1. The human body is mainly made up of water. Each water molecule contains two hydrogen nuclei, which are exploited to extract an image in MRI. These nuclei (precess at 42.58 MHz/Tesla) have a spin with orientation and frequency and a magnetic moment. Because the direction or orientation of the hydrogen nuclei are random (bottom), no net magnetization \((M)\) is seen when the body is not in an external magnetic field. When the body is in a magnetic field, the \(M\) vector (top center) has a \(z\) and \(xy\) component (top right) known as longitudinal \((M_z)\) and transverse \((M_{xy})\) magnetization, respectively. Inside the magnetic field \(B_0\) an equilibrium magnetization \((M_0)\) is aligned with the field in the \(z\)-direction.

The precession frequency of the nuclei can be determined by the Larmor equation.

\[
\omega_o = \gamma B_0
\]

where \(\omega_o\) is the angular frequency of precession of protons in an external magnetic field, and \(B_0\) is the strength of the external magnetic field.

When a subject is put into an MRI scanner, a net magnetization \((M)\) in the direction of the \(B_0\) field is induced by the magnet. An RF pulse at the precession frequency of the hydrogen
nuclei can be used to excite the protons, causing the net \( M \) to tip. The magnetization vector can be divided into two components, longitudinal \( (M_z) \) and transverse \( (M_{xy}) \).

\[
M_z(t) = M_o (1 - \exp^{-t/T1})
\]

\[
M_{xy}(t) = M_o \exp^{-t/T2}
\]

Where \( M_o \) is the equilibrium magnetization value before any excitation is applied and T1 and T2 are time constants. The graphs of Equations 2 and 3, commonly known as T1 and T2 relaxation are illustrated in Fig. 2. The formation of image contrast is dependent on them.

The following steps illustrated in Figs. 2 and 3 are needed for the entire image formation process. First the participant, after safety screening of any ferrous objects, is placed in the scanner. At this point there will be a net magnetization in the direction of the magnetic field \( (B_o) \) as a result of some protons aligning with the MRI’s magnetic field. In a 1.5 Tesla scanner, an excess of 1 nucleus in 100,000 aligns itself with \( B_o \) (Savoy, 2001). Next, calibration and fine tuning, called shimming is performed. Then, an RF pulse (typically 90°, known as flip angle, \( \phi \)) is sent at the appropriate frequency based on the Larmor equation. This knocks the particular protons of interest over by 90° and, as these absorb the energy, they will try to realign with the field. They will emit energy during this process. Then, after turning off the RF pulse the protons return to their original orientations and emit energy in the form of radio waves. This process of the nuclei returning to their original state is called relaxation. Relaxation is divided into two components: longitudinal (T1) and transverse (T2) also known as spin-lattice and spin-spin respectively. Anytime during this process of realignment with \( B_o \) the RF coil can measure the radio waves being emitted. The choice of when to measure makes up the basis of contrasts as will be seen below.

The steps described above give the general procedure for acquiring a signal but without specific information on its location. In order to fully understand image acquisition we need to include the roles of the gradient coils. The use of these gradients are needed to encode an array of points in space, Fig. 3.

First, one gradient will phase encode while the second will frequency encode and the final one will slice encode. How this is done is dependent on the imaging parameters and pulse sequences (will not be discussed in this chapter) used. The use of different values of echo time (TE) and repetition time (TR) will determine the contrast or type of image acquired. TR is simply the time between transmitted RF pulses. TE is the time to listen to the signal. Changing these two parameters determines the contrast in an anatomical MRI. Standard image contrasts are divided up into T1-weighted, T2-weighted or proton density (PD). In a T1-weighted image fat has a high (bright) signal and cerebral spinal fluid (CSF) has a low (dark) signal. In a T2-weighted image fat has low signal and CSF has high signal. In a PD image the contrast is between that of T1 and T2.

T1 and T2 weighted images are so named because they are based on the T1 and T2 relaxation times. T1 relaxation is a measure of how quickly protons realign with \( B_o \) (returning to equilibrium) and T2 relaxation is a measure of how quickly protons interact with each other (dephase).

Radio waves have to be a specific frequency to excite the protons. This frequency is proportional to the strength of the magnetic field. Turning on the gradients will slightly
increase the magnetic field allowing for manipulation of the frequencies that will affect protons in different parts of space. The slice encode process determines the slice location and thickness based on the pulse bandwidth per user specification. This is repeated until the desired dataset is acquired.

![Illustration of the interaction between protons, the magnetic field, and the RF pulse.](image)

The final step is to convert the acquired frequencies to image space/domain, forming an image. This is done by using the inverse fast fourier transform (IFFT), Fig. 4. This is a quick overview on generation of an image, which is necessary to understand in order to properly design an efficient fMRI experiment.

3. Image quality

In recent years, MRI image quality has improved. MRI image quality is dependent on many factors, some of which are TE, TR, the number of signal averages (NA) (also known as number of excitations NEX), and resolution. Note, that there are tradeoffs among all these factors. Increasing TE, for instance, allows for more T2 weighting, but decreases the signal to noise ratio (SNR). A long TR, on the other hand, allows for more slices to be acquired, while decreases T1 contrast and thus increases scanning time. Thus, to decrease the scanning time, TR can be decreased, which allows for more T1 contrast, at the expense of a lower SNR and acquisition of few slices. SNR increases as NA is increased, but only by a factor of $\sqrt{N_A}$. Scan time, in turn, is directly proportional to NA. Resolution is a function of the number of phase encode (PE) steps, the number of frequency encode (FE) steps (matrix size), field of view (FOV), and slice thickness (ST). Thinner slices reduce partial volume effects (PVE) and increase resolution, but SNR and anatomic coverage are reduced proportionally. In-plane resolution is increased as FE steps and/or PE steps are increased, but SNR decreases. Scan time increases with the number of PE steps. However, changing the FOV or increasing the
Using the gradient coils adds the ability to spatially encode the signals from the process described in Fig. 2 by slightly modifying the magnetic field with the gradients (causes the precession frequency to change spatially), thus a 2D/3D dataset in k-space (see Fig. 4) can be mapped. A lower magnetic field targets lower frequencies and a higher field generates higher frequencies. The frequency of $\omega_0$ determines which location to collect the signals from based on the gradients $G_y$ (phase encode), $G_x$ (frequency encode) and $G_z$ (slice select). Note, this is for an axial scan. The roles of the gradients switch for coronal and sagittal scans. The bandwidth of $\omega_0$ determines the slice thickness.

Using the gradients $G_x$, $G_y$ and $G_z$ the signals can be detected from specific regions of space. In conventional MRI scanning that is done in a raster scan fashion one slice at a time. In functional MRI an entire slice is acquired. The signals are populated in the $k_x$ and $k_y$ space via incrementing the gradients $G_y$ and $G_x$ slightly. The addition of $G_z$ allows for 3D imaging. This k-space data generated and collected by the gradients and RF coil is then converted to the image domain via the inverse fast fourier transform (IFFT) as the final process to generate an MRI/fMRI image.

Nominal spatial resolution is the smallest tissue volume size that can be represented on an image. It is defined as the prescribed FOV in the frequency- and phase-encoding directions divided by the number of frequency- and phase-encoding points ($|\Delta x = \text{FOV}_x / N_x|$ and $|\Delta y = \text{FOV}_y / N_y|$), respectively (Slavin, 2005). For instance, on older scanner platforms this arises when a matrix of 96x96 is set. As a result the scanner automatically upsamples by zero padding the raw data to 128x128, and thus the displayed resolution is $\text{FOV}_x / 128$ and
FOVy/128 instead of the nominal (acquired) resolution of FOVx/96 and FOVy/96. This is because in the past these scanners interpolated anything that was not a factor of two from a matrix size of 64x64, e.g. 96x96 so the displayed resolution would not be the same as the nominal. Displayed resolution is sometimes mistaken with nominal resolution.

On an MRI console there are options for the user to set the resolution. This, ultimately controls the image quality. TE and TR affect image contrast (increasing TE, increases T2-weighted, decreases SNR), (increasing TR, decreases T1-weighted, increases SNR). The user has the option to set TE and TR. The resolution is directly determined by the FOV, the number of frequency encode steps, the number of phase encode steps, and slice thickness for 2D imaging. The FOV in the x-direction divided by the number frequency encode steps gives the size in the x-direction. The FOV in the y-direction divided by the number phase encode steps gives the size in the y-direction. The slice thickness determines the size in the z-direction. Multiplying the x, y and z sizes results in the total voxel size, Fig. 5. In regards to 3D imaging, the FOV in the z-direction divided by the number phase encode steps gives the size in the z-direction instead of using slice thickness. The total size of a 3D volume is the (number of frequency encode steps) by (number of phase encode steps in y-direction) by (number of phase encode steps in z-direction). In this example, we chose x-direction to be the frequency encode as is the case for a standard axial/transverse scan; however, this can be switched with y- or z- directions if coronal or sagittal scans are desired. Smaller voxels result in a decrease in SNR. All these parameters are displayed on the MRI console, and can be changed within certain ranges. The choice of settings depends on the power of the gradient; the stronger the gradient, the smaller FOV’s and thinner the slices that can be obtained. There are other factors that the user can not directly control but still affect image quality. For instance, good shimming results in better $B_0$ homogeneity, which improves SNR and minimizes artifacts and geometric distortion. The slew rates on most of scanners are typically 150mT/meter/msec. This determines the maximum number of slices we can choose, as well as set the minimum bound of TE, FOV and slice thickness.

It is immediately apparent that from these acquisition parameters many things can either make for an efficient or inefficient dataset before beginning the fMRI task paradigm. In MRI, higher image resolution, i.e. smaller pixel/voxel sizes, is directly proportional to the magnet strength. Thus, going from 1.5 to 3 Tesla, the resolution can be doubled while using the same imaging parameters with the added advantage of shorter acquisition time. But with increase in resolution or signal there is also an increase in the noise. Note, for some applications (i.e. infant imaging) it may be advantageous to image at 1.5 Tesla.

In fMRI it is standard to achieve isotropic in plane resolution. For instance to attain a 3x3 pixel size then a matrix size of 64x64 with a field of view (FOV) of 192mm can be set in the scanner. The voxel is determined by the additional parameter of slice thickness which is dependent on the third (slice encode) gradient as was previously alluded to, Fig. 5.

3.0.1. SNR

The signal to noise ratio (SNR) is directly proportional to voxel volume (linear), which of course is inversely proportional to resolution. SNR is proportional to $\frac{1}{\sqrt{2}}$ of the number of phase encodes. SNR is proportional to $\sqrt{2}$ of the number of excitations. There are many different coils out there, but in general, a phased array coil (SNR proportional to $\sqrt{2}$ the
Figure 5. Image resolution is determined by the number of pixels in the x,y plane know as matrix size. This is dependent on how the gradients are phase and frequency encoded. In this case x=40, y=50. Assuming a field of view of 120 mm yields an in plane resolution of (120/40, 120/50) or (3, 2.4). Typically fMRI resolution at 3 Tesla is 3x3x5 with a FOV of 192 mm and matrix size of 64. In the typical axial slice acquisition, the slice encoding is performed in the z plane using gradient $G_z$ to set the slice thickness, the phase encoding is in the y plane using the gradient $G_y$ and the frequency encoding in the x plane using the gradient $G_x$.

number of coil elements) is best, followed by quadrature, and lastly linear. Any coil will give a better signal when closer to the region of interest (by inverse square). Also, with everything being equal, spin echo gives better SNR than gradient echo, because the 180° refocusing pulse corrects for field homogeneity (signal decays as a function of T2, not T2*).

SNR can be adjusted and optimized by choosing the proper imaging parameters. Note, there will be a tradeoff between resolution and SNR. For example if an increase from the 3x3x5 voxel resolution is desired then simply adjusting the matrix size to 128x128 while keeping all other parameters the same would yield a voxel of 1.5x1.5x5. Here the in plane resolution is doubled however the signal also suffers more noise. Similarly if a smaller slice thickness is chosen then SNR decreases and chances of partial volume effects increases. Next an explanation of the differences between MRI and fMRI is presented.

3.1. MRI vs fMRI

For both MRI and fMRI the process explained above applies. The difference is in the acquisition parameters and pulse sequences used. The most obvious difference is in the resolution, Fig. 6. MRI is denoted as the anatomical high resolution (< 1 mm in-plane) image. In general this is one anatomical (T1 weighted) dataset with three spatial dimensions. Whereas fMRI is a set of low resolution ( 3 mm in plane) datasets with the addition of a time dimension, 4D. The difference in resolution is based on the imaging sequence used to acquire the data. In MRI about 5 minutes are used to scan the entire brain which allows for a very fine grid whereas in fMRI more than 100 volumes are acquired in the same amount
of time. So, rather than 5 minutes to acquire one brain (MRI), only 2 seconds are allotted to acquire a brain (fMRI). The spatial resolution again is just a function of voxel size, but the temporal resolution is a function of gradient strength/slew rate, which determines how fast we can acquire images. When scanning for an anatomical image the participant lies in the scanner without engaging in any task. But for an fMRI, there is a specific task that needs to be repeated by the participant over the span of the 5 minutes. In doing so, the functional signal can be derived and analyzed. A description on preparing this task is given in Sections 5 and 6.

Figure 6. Illustrated is a visual comparison of the difference between MRI and fMRI. MRI is a set of high resolution slices that make up one 3D dataset while fMRI is a series of low resolution slices that make up many volumes, a 4D dataset (volumes + time). Note, images not to scale.

3.1.1. Artifacts

The main artifact in fMRI is susceptibility due to structures such as the ear and nasal canals because of the air tissue interface. These artifacts cause a signal loss in the auditory and frontal regions, respectively. For example, if we are interested in frontal lobe function then the imaging parameters need to be optimized to minimize this susceptibility artifact. For a 3 Tesla system, being conservative and using 64x64 matrix size helps alleviate this. Other common artifacts are caused by retainers and braces which result in signal loss. This should be kept in mind when recruiting participants for a study. There are other numerous types of artifacts that can occur in MRI such as chemical shift (fat vs water protons do not resonate at the same frequency), and ghosting. Chemical shift artifact can be minimized by increasing the receive bandwidth, increasing the FOV, and/or increasing the number of frequency encodes, but all these will also decrease SNR (Parizel et al., 1994; Zumowski & Simon, 1994).

Ghosting/motion artifacts (depending on the source) can be minimized with saturation (SAT) bands, increasing excitations, flow compensation (aka gradient moment nulling),
respiratory compensation, respiratory gating, cardiac gating or breath-hold (Morelli et al., 2011). Susceptibility artifact can be minimized by choosing spin echo over gradient echo scanning sequences (Stradiotti et al., 2009). If gradient echo must be done, minimizing TE and increasing the bandwidth will reduce susceptibility. With a basic understanding of MR imaging, we will follow with an explanation of the origins of the functional signal.

4. BOLD

Different regions are specialized to perform different sensory, motor and cognitive functions. fMRI has been developed as a technique for mapping brain activation over the last two decades and has found widespread interest in basic and clinical research aimed at better understanding brain function. The fMRI technique is based on the detection of local perturbations of the deoxyhemoglobin concentration in the vicinity of neuronal activity. Neuronal activity at the synaptic level results in both an increase in oxygen consumption by the active cortex and an even greater increase of blood flow to the site. Because oxygen delivery exceeds oxygen utilization, the net effect is a local decrease in deoxyhemoglobin concentration near the activation site. The decrease in deoxyhemoglobin concentration at the site of neuronal activity causes a local increase in the magnetic resonance signal, Fig. 7. This effect has been termed the Blood Oxygenation Level Dependent (BOLD) contrast mechanism (Ogawa et al., 1990). This is possible because of the magnetic properties of hemoglobin which contains four Fe2+ ions. Specifically, deoxygenated blood is paramagnetic, meaning it has a small additive intrinsic magnetic field and oxygenated blood is diamagnetic meaning it tends to oppose external magnetic field. The ratio of deoxygenated to oxygenated blood changes when a particular task is performed as a result of the neurons firing which cause an increase in both blood flow and oxygen consumption level in that particular region of the brain. However, the blood flow increase is larger than proportional oxygen consumption. The result of this brief stimulus and in turn neural activation is the hemodynamic response function (HRF). The HRF has a characteristic shape with an initial dip immediately following the stimulus then an increase and finally an undershoot although this shape can vary amongst participants (Aguirre et al., 1998). Understanding the behavior of the HRF can help in designing an efficient fMRI paradigm. In summary, BOLD fMRI capitalizes on the difference between two conditions: i.e., an active condition during which a specific stimulus that reflects a specific neural activity is generated and a passive condition during which the stimulus-related neural activity is absent or kept to a minimum and is generated by applying a low threshold, Fig. 8.

![Figure 7](image-url) A simplified flow of events that lead to the BOLD fMRI signal. A specific stimulus/task causes an increase in neural activity which triggers a complex chain of changes to cause an increase in cerebral blood oxygenation. This complex series of events includes an increase in cerebral blood flow (CBF), an increase in cerebral metabolic rate for oxygen which in turn causes the cerebral blood volume (CBV) to increase. These events cause a decrease in the local deoxyhemoglobin (HHb) content. This then allows for the detection of the signal which after post processing can be overlaid on an anatomical MRI.
5. Paradigms

There are three types of fMRI design paradigms: block, event-related (widely spaced and rapid) and mixed (block and event related), Fig. 8. The development of the event related studies (Buckner et al., 1996, 1998; Burow et al., 1998; Clark et al., 1998; Dale & Buckner, 1997; D’Esposito et al., 1999; Friston et al., 1998; Josephs et al., 1997; Rosen et al., 1998; Wagner et al., 1998; Wiener et al., 1996; Zarahn et al., 1997) came several years after the advent of BOLD fMRI. Choosing the proper TR, TE, FOV and matrix size values are all important and are dependent on the problem or question that is being investigated as will be discussed in section 6, but of equal importance is the type of paradigm used. This in fact is intertwined with the imaging parameters. The importance of choosing a suitable TR to and interstimulus interval (ISI) ratio has been known early on, namely because BOLD is not necessarily steady-state rather transient signal (Price et al., 1999). For block designs, this is fairly straightforward. If using an event related paradigm then caution should be taken in choosing the proper TR. Software such as optseq (http://surfer.nmr.mgh.harvard.edu/optseq/) allows for a good randomization of stimuli for rapid event related designs based on specific imaging parameter. This ensures a robust randomized design.

![Figure 8](image-url)

**Figure 8.** The typical paradigm designs include: (1) block, (2a) rapid event with fixed ISI, (2b) rapid event with random ISI and (3) mixed. In general they should consist of at least two states, A (no task) and B (task). The hemodynamic response is expected to rise as a result of the tasks. The ratio of blood oxygenation is measured to determine the pixel with the fMRI BOLD signal change.

Block designs commonly consist of two states, Fig. 8, A (rest) and B (task). However in some situations the factors of time or budget are an issue thus a third or even a fourth state is added. For example, doing two separate two state runs can take 12 minutes total however if they are combined into one run consisting of three states the total scan time can be reduced by several minutes. Also some institutions charge by the hour while others charge by the run, so knowing this can help in designing a paradigm that is optimal for either situation in order to stay within the budget allocated. Having four states complicates the design and strategies should be taken to design efficiently. A minimum block of 8 seconds of rest and 8 seconds of task has been achieved in the motor cortex without degrading the fMRI signal amplitude (Moonen et al., 2000). In two states the conditions would alternate a suggested minimum of three times, e.g. ABABABA. It is good practice to begin and end on a rest state in order to have a baseline measurement. For three conditions there are several combinations...
for presentation, e.g. CACBACBC; CABABABC; CABCABCABC where C can be the rest condition and A and B are task 1 and task 2 respectively. The disadvantage of this paradigm is that the participant may start to predict or anticipate the task. In contrast, event-related can be more easily randomized because of the small stimulus duration. But how long should the stimulus last? Several groups have noted different durations that can still be detected by fMRI such as 2 seconds (Bandettini & Cox, 2000; Blamire et al., 1992), 1 second (Dale & Buckner, 1997), 0.5 seconds (Bandettini et al., 1993) and 34 milliseconds (Savoy et al., 1995). Specifically, Dale (1999) illustrated that presenting a stimulus every 1 second is possible if the ISI is varied. Burock et al. (1998) showed that it is possible to have a mean ISI of 500 milliseconds if the stimuli presentation order is randomized. This temporal resolution is a clear advantage in event-related over block designs, however there is a tradeoff of SNR in the fMRI signal. It has been shown that going from a block to a variably spaced event-related design decreases the SNR by approximately 33% (Bandettini & Cox, 2000) and about 17% (Miezin et al., 2000) from a widely spaced to a rapid event-related paradigm. In general, block designs generate an increased magnitude in the BOLD signal intensity under the Buxton model (Buxton et al., 1998; Glover, 1999) and better statistical power (Friston et al., 1999). Thus, a standard MRI sequence protocol requires 5 minutes to acquire on brain (MRI scan), while 2 seconds are allotted to acquire the total volume of a brain during the acquisition of an fMRI sequence (fMRI). An optimal ISI for a fixed stimulus duration of less than 2 seconds is about 12 seconds (Bandettini & Cox, 2000). Additionally by randomizing the ISI the statistical power increases and allows for reducing the ISI (Bandettini & Cox, 2000; Dale, 1999). A third possibility is using a mixed design which is a combination of event-related and block design. Fig. 8. Note, for clinical use the majority of fMRI scans will follow the block paradigm. However, advances in experimental design and as clinicians become more informed, the use of event and mixed designs may start becoming more commonly used. Similar to choosing the imaging parameters, determining which paradigm to use will depend on the goal of the experiment.

6. Preparing an experiment

With the basics of MR physics and fMRI paradigms presented a more informed decision can be made on the experiment. In beginning the journey into an fMRI experiment some basic questions need to be asked.

- Why are we doing this experiment? This is generally our hypothesis.
- What are we looking for? For example, it could be a specific behavior or a physiological measurement that we are interested in.
- Where? This involves knowing the neuroanatomy.
- How? This is the type of fMRI design.

The best way to explain this is to walk through an example by asking these questions.

- Why? We hypothesize that eye movements will cause activation.
- What? Moving the eyes.
- Where? In the cortex.
- How? Tell participant to move eyes.
Initially this may look like a good set of answers. But if we investigate further we find that it is not. What is wrong with these answers? They are too general and leave room for error, confounds, and reproducibility will be difficult as nothing is mentioned about the paradigm or scanning protocol. As a consequence, it will be difficult for different investigators to reproduce the results (i.e. activation maps). The addition of scanning parameters would alleviate some of the problem. For instance, full brain coverage axial scan with TR=3 seconds, TE=35ms, \( \phi = 90^\circ \), matrix size=128x128, FOV=256mm and ST=8mm. This would yield a voxel resolution of 2x2x8mm, SNR=75%, and allow for about 45 slices to be acquired on a 3 Tesla system. If full brain coverage is necessary then these parameters would be good, however from the vague answers given above this cannot be determined. Also, note scanning time is still missing and there is no mention of the number of volumes to be acquired, leaving even the scanning parameters lacking some detail for reproducibility.

Let us now try to reword the answers to come up with a better starting point.

- **Why?** We hypothesize that eye movements will activate the visual cortex.
- **What?** Specifically moving the eyes from left to right in a saccadic fashion.
- **Where?** Want to see activation differences between V1 and V2 in visual cortex.
- **How?** We will have a block paradigm of 30 seconds fixation followed by 30 seconds of visually guided saccadic eye movements. A backprojection system will be used to display the visual stimulus that will cue the participant to fixate on a white dot in the center of the screen.

The dot will then automatically proceed to alternate back and forth between the center and the left side of the screen. Note since the participant being scanned is looking through a mirror, the direction the dot moves may be opposite of what it is on the stimulus computer screen depending on the setup. Forgetting this fact is a common mistake made by novice users which could lead to incorrect interpretation of results especially when studying oculomotor function where directionality is important. In regards to any stimulus that involves the visual cortex, it is good practice to either turn the lights off in the scanner room or keep on for all participants. This may seem like a trivial issue however it will introduce extra confounds and variables which could have been avoided. It is always recommended to test the paradigm in the scanner before recruiting participants in order to debug and pick out issues like that of having the mirror flip the direction of the visual stimulus. For statistical power we want to repeat this on-off cycle six times. Using the same imaging protocol just mentioned allows for 120 volumes to be acquired in 6 minutes and 21 seconds. The 21 seconds are used for dummy scans to allow for the participant to become acclimated in the scanner and more importantly for steady-state to be reached for the imaging sequence. For a particular participant whole brain coverage may only require 20 slices thus total slices acquired would be 120*20=2400. Hence in this latter experimental design example it is seen that by providing a more informed set of answers and paradigm the study can be changed from an inefficient fishing expedition to an efficient specific study. Is this the most optimized design? Probably not but it is a better starting point. Since the interest is in the known regions of the visual cortex then whole brain coverage may not be necessary. Reducing the coverage can result in enhanced scanning time, SNR, statistical power and/or image resolution.
Answering the questions differently will affect the way we want to scan the brain. For example let us ask the following questions and see how they will affect the scanning options from Fig. 9.

- What brain areas are active during a visual stimulus X? This is a very general question and may require an axial full brain coverage scan.
- Is the primary visual cortex (area 17) active during visual stimulus X? This is more specific and allows for a more localized scan in the visual cortex.
- What areas of the brain are active during working memory task Y? Again, this is very general.
- Is the prefrontal cortex active during working memory task Y? Here the frontal cortex is targeted hence no need to scan the visual cortex.

The conventional scanning plane has been the axial however there is no rule against scanning in other planes or at an oblique angle. In fact, there are situations where going away from convention is advantageous specifically when the anatomy of interest does not lie along one of the three orthogonal imaging planes. Depending on the answers given to the list of questions a very specific orientation and number of slices desired can be determined.

For example in Fig. 9 it can be seen that it may not be necessary or the most efficient use of time and resolution to choose the axial or coronal planes even for visual cortex activation. The oblique plane scan at the visual cortex may be the best choice. Also if the other scans were chosen it would not have been necessary to acquire slices from the entire brain rather enough to cover the visual cortex. If only a few slices are required, this could reduce the scanning time which could be invested in acquiring more volumes and/or increasing the resolution. Note, by acquiring more volumes we can increase the SNR; i.e., the SNR is proportional to the square root of the number of scans (rows) and the voxel volume. Overall, a priori knowledge of the particular problem that is anticipated to be studied can go a long way in optimizing the fMRI experiment. Indeed it is advantageous to do full brain coverage if that is what the problem entails however it is not a good idea to go on a fishing expedition for activation sites when one already knows they are interested in one region, e.g. primary visual cortex. A high resolution scan also helps in neurosurgical cases, such as localization of a brain function that could be affected by tumor resection. Second, and most importantly the brain mapping field has shown that it is not a matter of specifying exact regions when a hypothesis is formulated but rather a network of regions. The studies/hypotheses that are executed need to report brain networks; to do this one has to scan the entire brain. These are some aspects of several questions to ask before preparing an experiment. If thought out thoroughly they would minimize errors and confounding variables and in turn optimize the fMRI design and experience.

7. Collecting data

Once the questions to the problem are answered, the parameters such as time required, number of slices, number of volumes, matrix size, resolution and FOV are entered into the console of the MR scanner. It is recommended that the participant is made familiar with the task outside the scanner to minimize confusion and error inside the scanner. After entering the scanner the anatomical high resolution MR images are acquired followed by the fMRI
Figure 9. Illustration of four possible orientations overlaid on top fMRI results from an oculomotor paradigm coregistered with an anatomical MRI. In this case, all orientations except the third from top (oblique orientation at frontal cortex) would have captured the visual cortex activation.
scan. The latter is where the participant is instructed to perform a task. This process can take an average of an hour. In our experience, it is not recommended to go over this time too much mainly because of the attention and fatigue factor on the participant’s end. If prepared and thought out properly the hour should be sufficient to collect the MRI and two or three fMRI sessions. At the end of the scanning session the images can be copied onto a DVD in DICOM (http://medical.nema.org/) format and taken to a local workstation for analysis.

In the case of strictly clinical fMRI, for example, presurgical planning support, the scanners have a built in analysis tool that can output the results of activation on the console itself and these results can be pushed onto a picture archiving and communication system (PACS). PACS is a standard in clinical imaging departments but usually not available in research imaging centers. In this case the paradigm timing and design need to be entered into the console before the scanning begins. The use of the MR scanner’s software application is not recommended for research purposes, as it can only handle the statistical analysis (statistical, see Section 8.1) of one participant at a time. To date scanner vendors have not added the option of performing group analysis. Section 8.2 lists possible software packages that can be used instead of the built-in software on the MRI console. For clinicians, it is also suggested that they use these packages for a more robust analysis especially for case studies and for confirming the results in situations of more critical care, such as surgical resections.

8. Preparing data

After fMRI data is acquired, motion correction and filtering may be required. Each slice has to be aligned with the next within the volume followed by image registration between the volumes using the first as the fixed (stationary) volume. From the saccadic eye movement example in Section 6, if each volume consists of 20 slices, this means 19 slices need to be aligned to the first. The 2D slice alignment is repeated for all 120 volumes independently. Afterwards the slice aligned volumes are registered (translated and rotated) to each other in 3D. Accordingly, 119 volumes are registered to this first fMRI volume (fixed). This is repeated for each participant dataset. For group comparisons, the processed 4D datasets then have to be registered to a common space. Note, after statistical processing (described in next section) is performed, the slice-aligned volumes, which are generated by the echo planar image (EPI; functional) runs are registered with the 3D images, which are high resolution anatomical, T1-weighted runs.

8.1. Statistics

Overall, there are two statistical approaches in fMRI, hypothesis and data driven. What will be described is the former. Once all the data sets are aligned, time courses can be plotted for each voxel, Fig. 10. The signal pattern is predicted to follow the block, event or mixed design and a general linear model (GLM) can be used to fit the data with a particular p-value. The fMRI signal from one voxel over time can be defined as $y(t)$:

$$y(t) = \beta x(t) + \epsilon(t)$$  \hspace{1cm} (4)

Where $\beta$ is the parameter estimate (PE) for $x(t)$ and $\epsilon(t)$ is the error term. The boxcar stimuli (from Fig. 8) can be denoted by 0s and 1s for rest and task respectively and defined as $x(t)$. Even though it is not explicitly shown here $x(t)$ is convolved with the HRF function $h(t)$. 
This convolution of the stimuli patterns with HRF is used to model the predictors in the GLM. Thus, $x(t)$ is more correctly defined as $x(t) = x(t) \ast h(t)$. An example predictor is seen in Fig. 11. In the case of three stimuli being presented, the equation becomes as follows:

$$y(t) = \beta_1 x_1(t) + \beta_2 x_2(t) + \beta_3 x_3(t) + \epsilon(t)$$

(5)

In this case, three waveforms are estimated with the $\beta$ terms. The higher the value for a particular $\beta$, the closer the waveform fits the corresponding $x$ model. For instance, a high $\beta_2$ value signifies a good fit with the model $x_2$. The different models are commonly called explanatory variables (EVs). The linear model from Equation 4 can be written in matrix form.

$$
\begin{pmatrix}
Y_1 \\
Y_2 \\
Y_3 \\
\vdots \\
Y_n
\end{pmatrix}
= 
\begin{pmatrix}
X_{11} & X_{12} & X_{13} & \ldots & X_{1p} \\
X_{21} & X_{22} & X_{23} & \ldots & X_{2p} \\
X_{31} & X_{32} & X_{33} & \ldots & X_{3p} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
X_{n1} & X_{n2} & X_{n3} & \ldots & X_{np}
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\vdots \\
\beta_n
\end{pmatrix}
+ 
\begin{pmatrix}
\epsilon_1 \\
\epsilon_2 \\
\epsilon_3 \\
\vdots \\
\epsilon_n
\end{pmatrix}
$$

(6)

This equation can be rewritten as:

$$Y = X\beta + \epsilon$$

(7)

After solving for $\beta$, the t statistic can be calculated by:

$$T = \frac{\beta}{\sigma(\beta)}$$

(8)
where $\sigma$ is the standard error. The t statistic signifies how well the data fit the predictor models $X$. To determine whether $\beta_1$ better fits EV1 than $\beta_2$, a simple subtraction is performed and a new t statistic is computed.

In a block paradigm, the signal is lower during fixation (condition 1) at a specific voxel than when the participant performs visually guided saccades (condition 2). Regression analysis is then performed on all voxels ($Y_1$ to $Y_n$) of an fMRI dataset. The voxels that follow this trend or fit the boxcar (Fig. 8) can be identified as significantly activated regions of the brain responsible for the associated task or stimulus chosen. If each block lasts 30 seconds with TR=3 seconds this means that 10 volumes are acquired in each block. If the total time is 6 minutes (360 seconds) this yields $360/3=120$ total volumes or time points used. The intensity of one voxel from each volume will result in a time series of 120 points with time as the x-axis and fMRI signal as the y-axis, again with a total time of 6 minutes ($3*120=360$ seconds; Fig. 11). The mean and standard deviations for the duration between the two conditions are then calculated. The voxels that significantly follow this trend are depicted on a color-coded map, which denotes threshold of activation and overlaid onto the original fMRI dataset. Depending on the resolution of the dataset this can be performed on more than 100,000 voxels, $128x128x20 = 327,680$ voxels from the above example. Because of this number many voxels will be activated by chance (for $P<0.01$, $0.01*327,680 = 3,276$) so bonferroni corrections are used; i.e., rather than use $P<0.01$, a corrected $P<0.0000001$ is used $(0.01/100,000 = 0.0000001)$. A subsequent analysis that is commonly used is cluster grouping rather than analyzing individual voxels. If a group of interconnected voxels fit the general linear model from the paradigm then, the chances of false activation due to random noise can be reduced. Thus far, the statistics described have been based on the hypothesis-driven method. However, it is possible to use another technique from statistics known as independent component analysis (ICA). This method is widely used in functional connectivity analysis but will not be discussed in this chapter.

Figure 11. The signal intensities from the ROI in Fig. 10 are plotted out over time (orange curve) and fitted to the model predictor (black curve) based on the paradigm from Fig. 8. Points are sampled from the signal at the two states and statistical difference are determined. Looking at the ROI the raw signal is found to fit the model well for the two conditions. The orange pixels are the significant activation results of the GLM overlaid on top of the fMRIs.
8.2. Software packages

There are several software packages available for fMRI analysis. FSL (http://www.fmrib.ox.ac.uk/fsl/), AFNI (afni.nimh.nih.gov/afni), and SPM are all free with the exception that the latter requires a Matlab (http://www.mathworks.com/) license. FSL and SPM are available for all operating systems. AFNI operates on a UNIX platform but not Windows. More groups seem to be using a combination of these packages. For instance, it is common to use FSL for pre processing and SPM for post-processing. AFNI can be used to evaluate possible susceptibility artifacts, as a result of fMRI scanning parameters, such as TE (Gorno-Tempini et al., 2002) and test the efficacy of the protocol before deciding to collect data across many subjects. This specific use of AFNI is especially helpful when interested in studying frontal lobe activation, which is the area, mostly affected by these artifacts. The packages are all good in their own way and it comes down to user preference.

9. Pitfalls and limitations

Throughout the entire process of an fMRI experiment we need to be aware of our limitations. Are we using a 1.5, 3, or 7 Tesla magnet? The same design paradigm and protocol can be run on all three magnet strengths and give different results. For instance, Krasnow et al. (2003) found a 23% increase in striate and extrastriate activation volume at 3 T compared with that for 1.5 T during a visual perception task. Also a review conducted by Voss et al. (2006) summarized some of the advantages and disadvantages of using a 1.5 or 3 T system for fMRI. The advantages of 3T included: extent and strength of activation, resolution, imaging time. Whereas the advantages for using a 1.5 T were reduced susceptibility and chemical shift artifacts. Note, the strength of activation is directly related to the threshold that is set by the user which can lead to an area being activated at 1.5 and not at 3.0 T. More importantly, at 3 T, activation was detected in several regions, such as the ventral aspects of the inferior frontal gyrus, orbitofrontal gyrus, and lingual gyrus, which did not show significant activation at 1.5 T (Krasnow et al., 2003). The change in signal intensity is also lower for lower magnet strengths. For example, in a visual experiment the change in signal intensity was found to be 15% at 4 T and only 4.7% at 1.5 T (Turner et al., 1993). Thus the strength of the magnet has its own “pearls” and pitfalls and we should be able to identify them. The stimulus can easily be confounded if not designed efficiently. If one studies visual perception then, care needs to be taken to ensure that factors and parameters, such as the lighting in the room, the stimulus size, contrast, luminance, spatial and temporal frequency and distance to the participant stays fixed. If eye movements are of primary concern then investing in an MR compatible eye tracker is necessary in order to confirm that indeed the participants are performing the task. In addition, fMRI is susceptible to motion thus the head has to be held still and padding used. If motion is present then post processing algorithms can be utilized to correct them however only if they are minor. Otherwise the data will not be useful and need to be discarded. It is crucial that the participants understand all the instructions. In an eye movement study, using the same visual stimulus but changing the instruction slightly found that significantly different regions were activated and recorded (Kashou et al., 2010). Participants comfort is important to be ensured to eliminate any confounding factors; if for example, they are in pain then this will be reflected in the data. In designing the task paradigm things to consider are the timings, e.g. 15 vs 30 seconds. The latter may be too taxing on the eyes if a for instance the participant is to focus on a visual stimulus consisting
of counter-phase checkerboards. It may advantageous to use a two-state block design over a three state to alleviate participant confusion and simplifying the statistical analysis process. Reducing the time to 15 seconds and using only two states allow for more cycles to be repeated and give more power for statistical averaging. Alternatively, 15 seconds can be used to add a third state yet keep the total scan time of 6 minutes and introduce a new sequence of events/cycles, however that would increase the complexity. Other limitations include both anatomical and physiological components. For instance, the HRF varies with location in the brain for each participant, as well as across participants (Aguirre et al., 1998). The size and location of the neuroanatomical site can affect the fMRI activation that can be derived, i.e., frontal lobe may not be distinguishable if the susceptibility artifact is too large. Also, delineating smaller regions will only be possible if higher resolution is acquired. Physiologically, the vascular response time may vary from approximately 2 to 6 seconds depending on the region of interest. Through visual cortex stimulation the signal change onsets were shown to vary from 4 to 8 seconds in gray matter and 8 to 14 seconds in sulci (Lee et al., 1995). Another group used a checkerboard as a visual stimulus and found that the hemodynamic response was delayed by 1-2 seconds and reached 90% of its peak after 5 seconds (DeYoe et al., 1994).

Other physiological sources of noise include the respiratory and cardiac cycles but there are algorithms to minimize them, such as the RETROICOR (Glover et al., 2000). Participant habituation to a task will cause a signal intensity decrease in particular areas, such as the amygdala (Breiter et al., 1996; Sakai et al., 1998). Similarly, significantly lower responses were found for auditory stimulation as a result of habituation (Pfeiderer et al., 2002; Rabe et al., 2006). Brain activation can also be modulated by motor fatigue (van Duinen et al., 2007) and mental fatigue (Cook et al., 2007) has been correlated with decreased activation.

Switching coils, or scanners in a middle of a study may confound the data as the equipment may vary and should be kept in mind. For instance switching from 3 T to 1.5 T will cause the spatial resolution to decrease even if the same scanning sequence is performed. Going from a GE 3 T to a Siemens 3 T will not necessarily give the same quality (SNR, resolution), as their sequences and hardware capabilities may vary. Using the same magnet strengths and vendors, e.g. GE 3 T, may also yield differences as the quality and timing of the sequences depend on the maintenance and shimming of the systems. More specifically the timing is dependent on the gradients’ capability to switch on and off, and thus from a physics perspective, variance can be introduced. In addition, the location of one magnet versus another can have an effect on the amount of radio frequency interference or vibration artifacts seen in the images. Some magnets are placed in locations where building vibration is an issue, thus special vibration dampeners have to be installed on various floors in a building. These structural and locational differences will certainly affect the physics aspects of the scanner. The SNR would be directly affected if switching coils even within the same location. Downgrading from a 32-channel to an 8-channel will decrease the SNR and resolution. To properly correct for these variables, baseline/normative data have to be collected from all the potential systems that may be used in a study. This is an essential issue in multicenter trials, especially if different MR system vendors are being used. Again, all these factors can be minimized before the study begins and the data is collected.

After the data is collected, other issues arise. What is the minimum significance level that will be accepted, \( p=0.05 \) vs \( p=0.01 \)? What about gaussian smoothing the data, none, 5mm, 10mm?
This depends on the SNR of the dataset. By smoothing the data, activation clusters will spread and spatial localization limited as well as more regions will be activated (Mikl et al., 2008; Scouten et al., 2006). If fMRI voxel resolution is 2x2x2 mm and a smoothing kernel of 10mm is used then the resolution is decreased by a factor of five (10x10x10 mm). That in way defeats the purpose of sacrificing other parameters to achieve the 2mm$^3$ voxel size. The degrees of freedom (DOF) in the image registration process need to be understood as well. Simply choosing DOF of 3, 6, 7 or 12 can cause alignment errors and erroneous activation maps. These numbers correspond to the standard types of image alignment methods available for 2D and 3D. For example, DOF=6 in 3D includes three translation ($x,y,z$) and three rotation ($x,y,z$) components. DOF=12 includes an additional three scaling factors ($x,y,z$). This becomes a bigger issue if partial brain images are acquired for fMRI and need to be aligned with a whole brain high resolution MRI. Even though the software packages above have built-in tools for these steps, the user has to be cognizant of the potential pitfalls. The registration algorithms vary and the types of interpolation used will introduce more blurring in addition to the degree of the gaussian smoothing chosen/employed. Many users think that because the fMRI activation maps are overlaid on top of a high resolution anatomical MRI, that they have the same accuracy and resolution and do not appreciate the amount of approximations involved in the process. Overall an fMRI study needs much thought. There are many variables and confounding factors involved in setting up an fMRI design and this section touched on the importance of being aware of the limitations and keeping these variables as constant as possible when clinical, longitudinal or group studies are performed.

10. Conclusion

In designing an fMRI experiment there is no real (one) gold standard set of parameters for all participants and stimulus paradigms. The concern thus becomes on what we can control. Here are some concluding tips in answering this question. Before beginning anything we need to fully understand the experimental goals. Inevitably tweaking the fMRI imaging parameters needs to performed, such as the TR, number of slices etc. As a result the paradigm will also be adjusted accordingly, especially in the case of event related designs. Here are some design questions to answer: Design type? Blocked, Event-Related, Mixed? How many subjects? How much data to collect for each subject? How many stimulus conditions? How many repetitions for each condition? When should each stimulus be presented? Getting in the habit of asking many questions before starting a study is key. Overall, the best solution is experience.

Author details

Nasser H Kashou
Department of Biomedical, Industrial and Human Factors Engineering, Wright State University, USA
References


A Practical Guide to an fMRI Experiment
http://dx.doi.org/10.5772/58260

fmri study at 3t., Brain Res 1336: 10–21.
URL: http://dx.doi.org/10.1016/j.brainres.2010.04.017


URL: http://dx.doi.org/10.1006/nimg.2000.0568

URL: http://dx.doi.org/10.1016/j.mri.2007.08.006


URL: http://dx.doi.org/10.1148/rg.313105115


