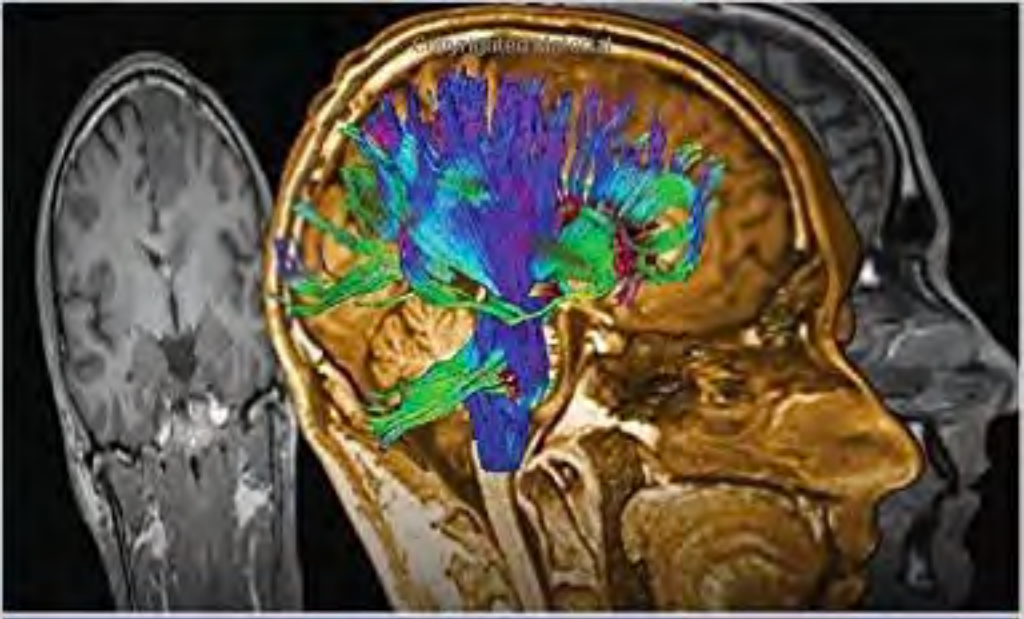


CLINICAL BRAIN MAPPING

DANIEL YOSHOR AND ELI M. MIZRAHI

CHAPTER 3

Functional MRI for Cerebral Localization: Principles and Methodology

Michael S. Beauchamp

Department of Neurobiology and Anatomy, University of Texas Health Science Center, Houston, Texas

▶ INTRODUCTION

The widespread availability of magnetic resonance imaging (MRI) scanners has allowed clinicians a clearer picture of brain structure and function. In addition to the static anatomy imaged with traditional MRI, functional MRI (fMRI) allows for the measurement of brain function. The most widely used fMRI technique is known as blood oxygen level dependent (BOLD) fMRI. BOLD fMRI has become a powerful tool for studying brain function, and has led to important advances in our understanding of the basic neurobiology of the brain. However, there are significant obstacles to translating these advances into clinically useful information. In this chapter, we give a brief overview of the techniques used to create BOLD images and the data analysis necessary to identify active brain areas. Finally, we examine the difficulties in drawing clinical inferences from fMRI studies.

▶ A BRIEF HISTORY OF fMRI

Paul Lauterbur and Peter Mansfield received the 2003 Nobel Prize for developing techniques to transform NMR (nuclear magnetic resonance), which gives information about a uniform sample of material, into a technique that provides a two-dimensional (2D) image of an object.[1,2] This method is now known as MRI, a considerably more patient-friendly term than the original name, which was "NMR Zeugmatography" (zeugma, Greek for yoke, referring to joining a magnetic field gradient to a local region to form an image). Because of its medical importance, it is likely that a future Nobel Prize will honor those involved in the development of fMRI. A probable honoree is Seiji Ogawa, who discovered that hemoglobin could serve as a natural contrast agent to measure blood oxygenation.[3–5] Following Ogawa's discovery, a number of groups realized that this BOLD MRI could be used to measure changes in blood oxygen in the human brain linked to brain activity.[6–8] This new method was termed functional MRI (fMRI) because it measured brain function, and to distinguish it from the more traditional anatomical or structural MRI.

▶ A THUMBNAIL SKETCH OF fMRI

As neuronal activity in the brain fluctuates, the local blood oxygen content varies. A particular type of MR image, a T2*-weighted image (effective T2), is sensitive to the blood oxygen content. What distinguishes fMRI from MRI is that instead of collecting a single image (to inspect brain anatomy), a whole series of brain images is collected to find brain areas with fluctuating blood oxygen content. When this type of image is acquired, active brain regions become brighter. Figure 3–1 shows a thumbnail sketch of a simple fMRI experiment. When the patient views a blank screen, there is little activity in visual cortex in the occipital lobe. When the patient views a picture, neurons in visual cortex become active and the blood oxygenation level increases, resulting in an increase in MR signal intensity. When the picture disappears, neurons in visual cortex cease firing and the MR signal returns to baseline intensity.

In order to perform an fMRI experiment, specialized equipment over and above that required for typical clinical MRI is required. Most importantly, sensory stimuli must be delivered to the patient, and these stimuli must by synchronized with the MR image acquisition (Fig. 3–2).

A thriving cottage industry exists to provide this equipment; often the scanner manufacturers will combine and resell the equipment from a combination of vendors as part of an "fMRI package." However, third party vendors often offer capabilities unavailable from the scanner manufacturer. Headphones allow the delivery of auditory stimuli (such as single words) or

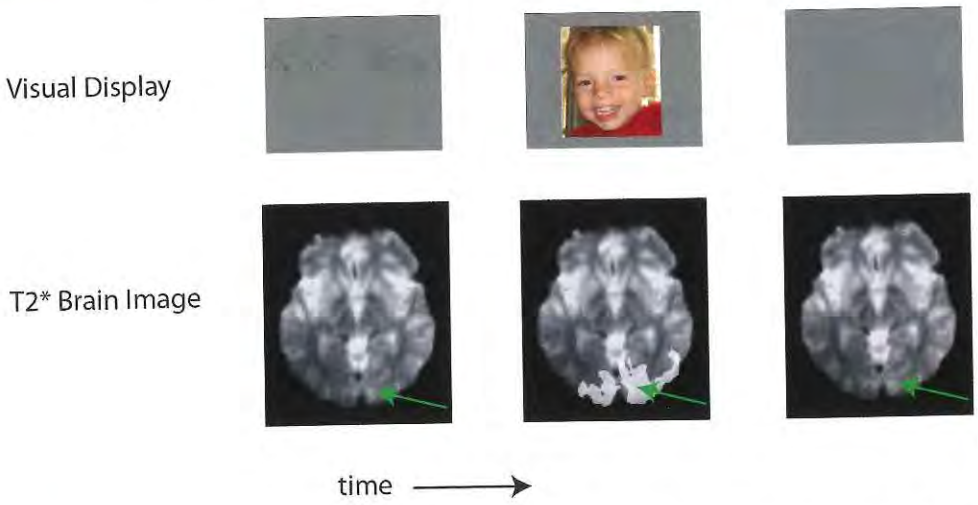


Figure 3–1. Top: visual display viewed by a subject. The subject views a blank screen, followed by an image of a face, then a blank screen. Bottom: T2* MR images from the subject's brain collected while viewing the display. An axial slice showing visual cortex in the occipital lobe (*green arrow*). When neuronal activity in visual cortex increases, there is a concomitant brightening of the voxels in visual cortex. When the stimulus is turned off, the brightness returns to baseline.

instructions (tap the fingers in your right hand). The stock headphones included with MR scanners are pneumatic (as in old-fashioned airline headsets): sound is generated in a speaker and transmitted through plastic tubes to the headset. This is adequate for patient communication, but is not sufficient for an auditory fMRI experiment in which the brain responses to carefully calibrated stimuli are measured. Therefore, at least three vendors produce high-fidelity MR compatible headphones. Visual stimuli may be delivered using goggles (containing small

Figure 3–2. Apparatus for delivering sensory stimuli and recording behavioral responses during a functional MRI experiment. (A) Picture of the MR scanner showing bed and head coil. (B) Close-up view with patient holding button response device, wearing earphones, and viewing visual stimulus in mirror attached to head coil. (C) View behind the scanner, showing screen with visual display viewed by subject via mirror (*white arrow*) and video-based infrared eye tracker (ASL Inc., *green arrow*). (D) Close-up view of button response device (Current Designs, Inc.). (E) Close-up view of somatosensory stimulation device consisting of a piezoelectric bender attached to subject's hand with an elastic bandage.

LCD panels) or via a mirror viewed by the subject. This allows the patient to see visual stimuli projected onto a screen behind or in front of the scanner. Piezoelectric vibrotactile devices allow delivery of somatosensory stimuli. Because subjects may close their eyes or fall asleep, a useful tool for fMRI is an eye tracker, which allows experimenters to view a video image of the subject's eyes and monitor their behavioral state. It is also important to monitor responses to be sure subjects are able to perform the task. For instance, a subject might press one button when seeing a male face and another button when seeing a female face. This helps the subject maintain alertness and makes sure that the subject understands the instructions and is performing the task.

▶ MRI AND MR SCANNERS

MRI scanners can be viewed as complex computers that can use different programs (known as pulse sequences) to acquire many different kinds of images that depend on different physiological properties. Because no ionizing radiation is used during MRI scanning, a typical MR examination includes many different pulse sequences, with the primary limit being the time that the subject can remain in the scanner.

By far, the most common method for performing fMRI of the brain is BOLD fMRI. To perform BOLD fMRI, a pulse sequence must be sensitive to the amount of blood oxygenation in the tissue, and it must acquire images quickly. A typical spoiled-GRASS pulse sequence used for anatomical imaging of the brain takes minutes to acquire a single brain image. As generally applied, BOLD fMRI uses a pulse sequence with two key features: first, it generates images with T2* contrast, a contrast that is sensitive to the amount of oxygenated blood in the tissue, and second, it generates images of the entire brain in a second or two (rather than minutes) using echo-planar imaging. These images are referred to as T2*, echo-planar imaging (EPI), or functional images (as opposed to purely anatomical images). Figure 3–3 shows the basic physiological mechanism of BOLD as it is currently understood: increased neuronal activity causes spatially localized increase in blood flow without as large an increase in oxygen consumption, resulting in a net increase in blood oxygenation and an increased MR signal.

▶ COLLECTING MRI DATA

Because the rapidly acquired T2* images used to measure BOLD activity have poor contrast and are of low resolution, fMRI experiments nearly always acquire a high-resolution T1-weighted anatomical brain volume to serve as an anatomical underlay for the functional brain map (Fig. 3–4A). To acquire this single anatomical volume takes about 5 minutes. Images for BOLD activity are acquired much more quickly, usually limited by the speed of the magnetic field gradients used for image acquisition. The typical goal in an fMRI study is to collect data from the entire brain. If a single slice of approximately 3 mm thickness is acquired in 50 ms, we can collect 40 such slices within a 2-second repetition time or TR (the TR is the time required to collect the entire volume of interest). This allows us to collect data from a 12-cm-thick slab of tissue, or about the inferior to superior extent of the average human brain (Fig. 3–4B).

BOLD fMRI studies generate huge amounts of data. A 5-minute anatomical MR scan series generates a single brain volume. The single volume is viewed as a series of slices on a workstation or a light box, and can be quickly read by a radiologist who examines each slice for abnormalities. In the same 5-minute period, a typical fMRI scan series would collect brain volumes every 2 seconds, producing 150 brain volumes (Fig. 3–4C). Even if these brain volumes were each manually examined, it would be difficult to detect the slight changes in image intensity that signal the presence of brain activity. Therefore, we can consider BOLD fMRI data analysis as a problem of compression, or reduction, in which the hundreds of brain volumes acquired are reduced to one or a few brain volumes that can be easily examined by the scientist or clinician.

▶ ANALYZING fMRI DATA

fMRI studies acquire data from thousands of brain locations, leading to the analysis problem of deciding which brain locations are involved in the experimental task. To address this problem, techniques developed to analyze other types of time varying data have been adapted to fMRI.

Figure 3–5 shows some sample raw data from voxels in the ventricle (where there is no neural response to the visual stimulus) and from voxels in the visual cortex (where there is a great deal of response to the visual stimulus).

The challenge faced by fMRI analysis is to accurately distinguish these two populations, differentiating active voxels from inactive voxels. An arbitrary statistical threshold is used, which inevitably classifies some truly active voxels as inactive, or vice versa. There is no universally agreed upon threshold for use in creating activation maps, or even universally agreed upon way of calculating significance; that is, a map that is labeled as "active voxels, $p < 0.05$" in one study may use a completely different statistical test than that used in a different study that uses the same nominal threshold.

Before fMRI, the most popular method for brain imaging was positron emission tomography (PET). In a

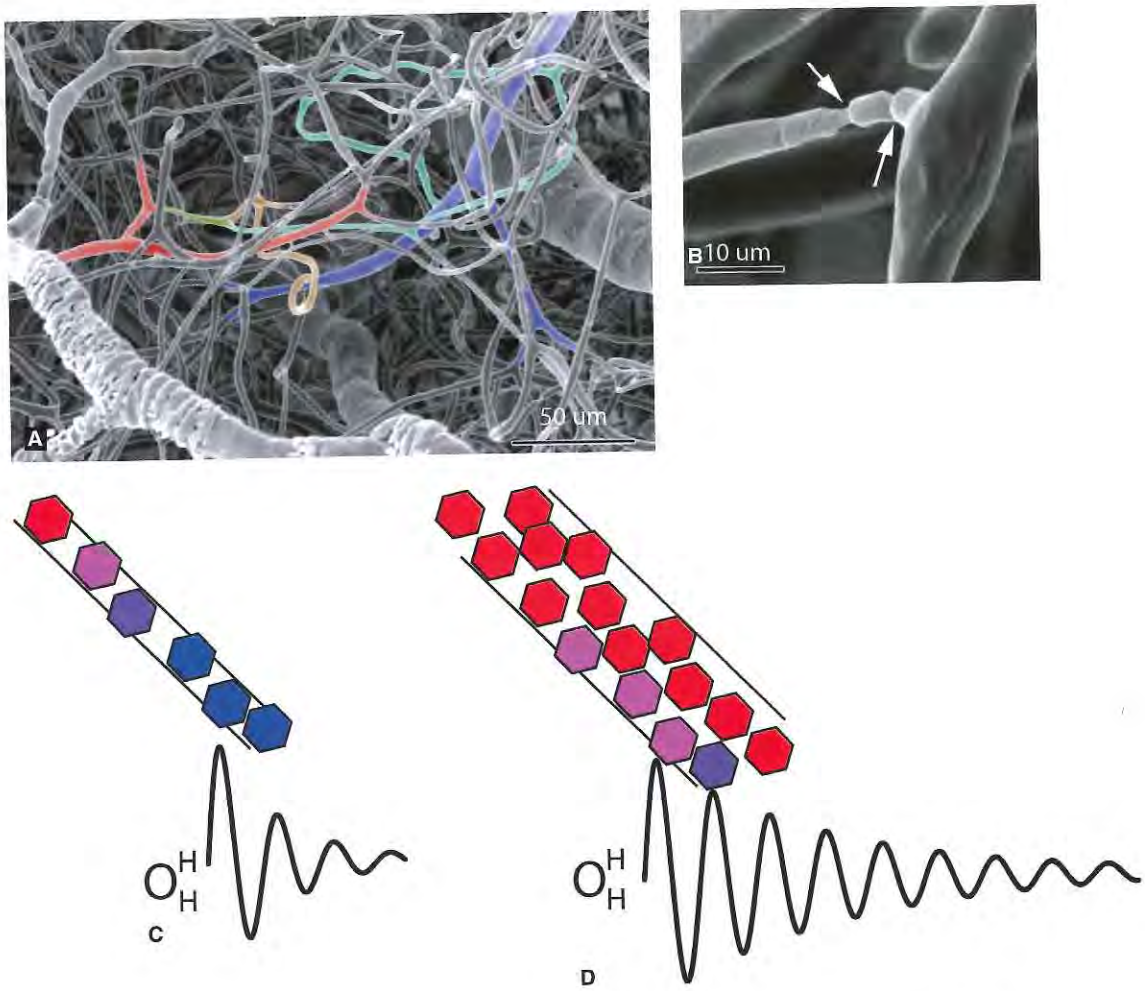


Figure 3–3. Mechanism of blood oxygen level dependent functional MRI. (A) The cerebral vasculature in a small region of auditory cortex. A precapillary arteriole (*red*) gives rise to a collateral capillary (*green*) which divided into two terminal capillaries (*orange and cyan*) before joining with other capillary loops and entering a postcapillary venule (*blue*). (Adapted from Harrison RV, Harel N, Panesar J, Mount RJ. Blood capillary distribution correlates with hemodynamic-based functional imaging in cerebral cortex. Cereb Cortex 2002;12(3):225–233.) (B) Enlargement showing perivascular control structures near capillary branching points that may control blood flood. (Adapted from Harrison RV, Harel N, Panesar J, Mount RJ. Blood capillary distribution correlates with hemodynamic-based functional imaging in cerebral cortex. Cereb Cortex 2002;12(3):225–233.) (C) At rest, there is relatively little blood flow into an area of cortex. Black lines show a capillary, hexagonal shapes represent red blood cells, red color represents oxyhemoglobin, and blue color represents deoxyhemoglobin. Most of the hemoglobin is deoxygenated, weakening the echo from nearby protons in water molecules (*black sinusoidal curve*). (D) During activation, more blood flow enters the capillary. Only some of the extra oxygen is extracted, resulting in less deoxygenated hemoglobin and greater signal from protons in nearby water molecules.

typical PET experiment, only a single brain image would be collected for each condition for each subject. Then, a subtraction was performed to show areas with differential blood flow between conditions.[9]

In the early days of fMRI, people would treat the fMRI data like PET data, simply performing a *t*-test between task epochs (e.g., subject looking at a picture) and rest epochs (e.g., subject looking at a blank screen). The flaws in this approach are immediately apparent when examining Fig. 3–5C. A *t*-test assumes that values collect in task and baseline periods are similar to each other. However, the MR time series in active voxels in visual

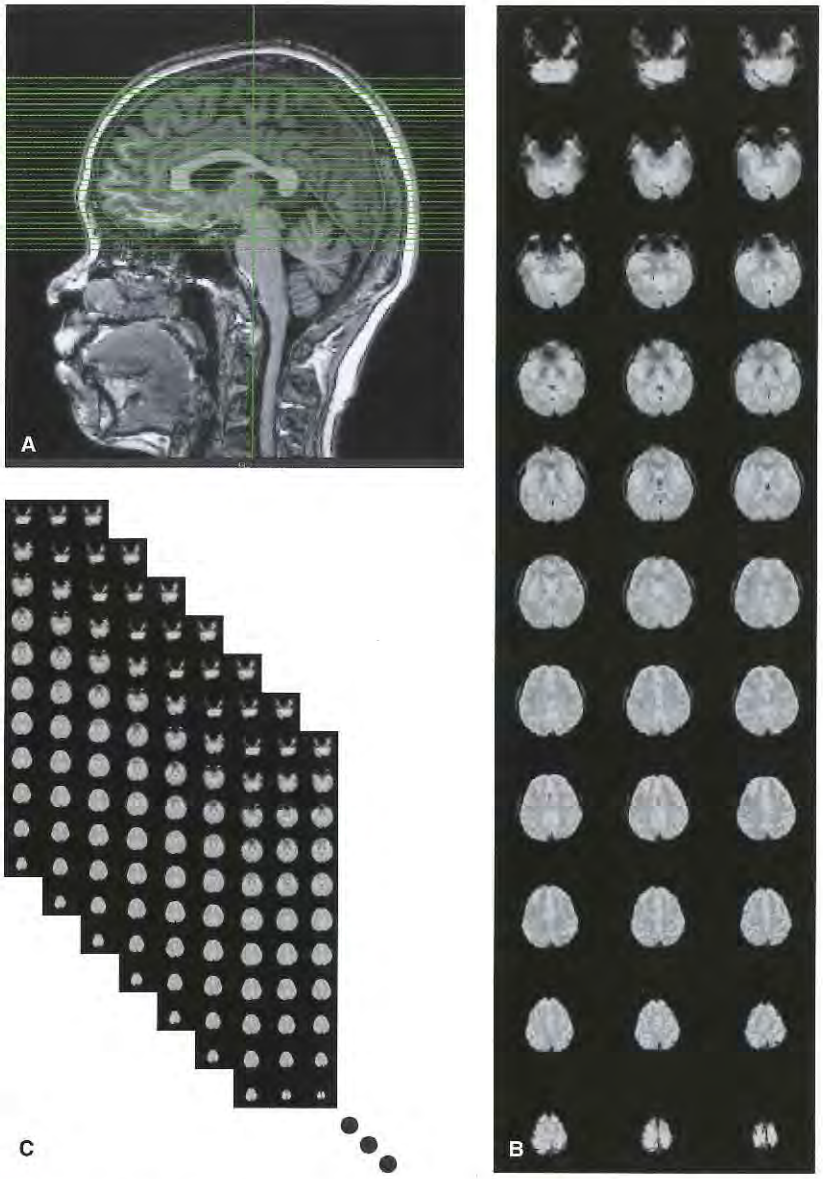
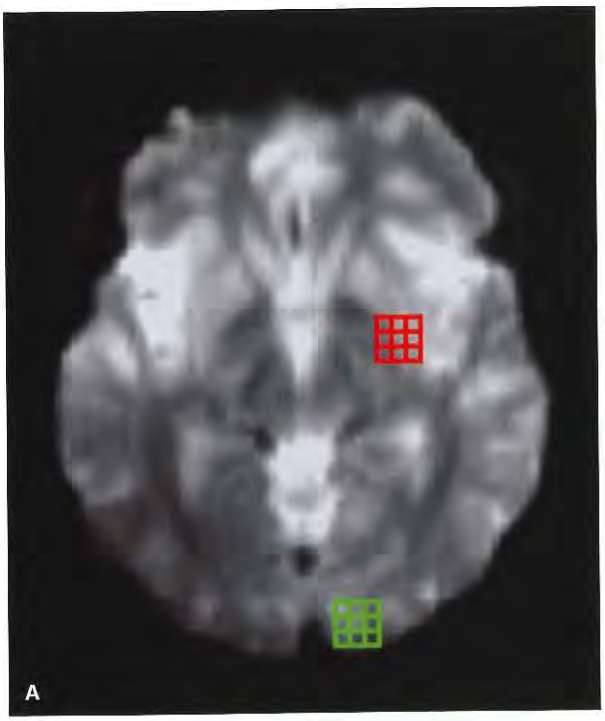
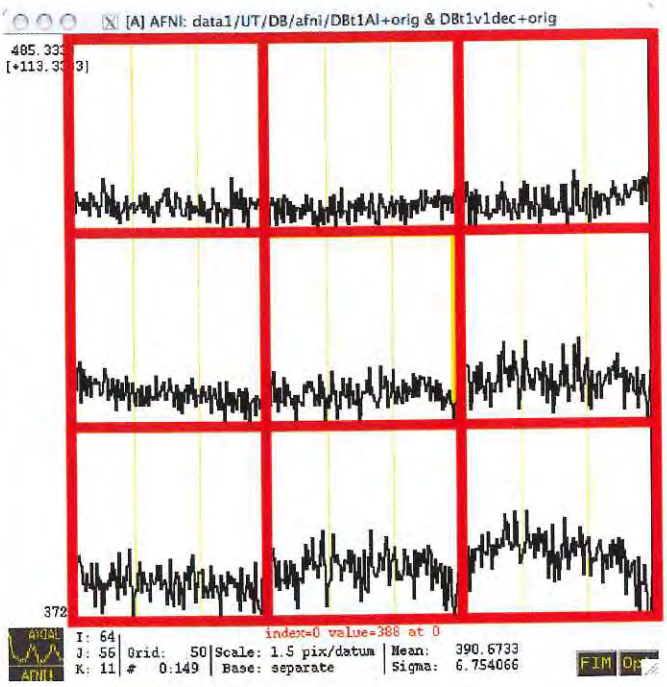


Figure 3-4. Data collected in a typical functional MRI experiment. (A) High-resolution anatomical image (T1-weighted) showing fine detail of anatomical structures. Green lines show location of T2*-weighted images pictured in (B). (B) T2*-weighted echo-planar images collected in a single 2-second repetition time (TR). 33 slices are pictured in a 3 × 11 montage, from inferior (*top left*) to superior (*bottom right*). Each slice is 3-mm thick, with 2.75 mm × 2.75 mm in plane resolution. (C) A T2*-weighted brain volume is collected every 2 seconds, leading to 150 volumes in a typical 5-minute scan series.

cortex does not resemble a square wave, but more closely resembles a sine wave. Treating all points on the top half of the sine wave (for task epochs) and bottom half of the sine wave (for baseline epochs) as equivalent is inaccurate. Therefore, the next stage in the development of fMRI analysis techniques was to use cross-correlation,[10] which allows knowledge about the expected shape of the MR time series to be incorporated into the analysis. The next (and consensus) method for analyzing fMRI data was to use multiple regression.[11] In this most common analysis method, also referred to as the multivariate generalized linear hypothesis or general linear model, regressors are created that correspond to the temporal sequence of different events experienced by the subject during scanning.[12–14] Brain regions in which the MR time series is time-locked to these events are classified as "active." In this section, we discuss the methods and procedures for deciding whether any





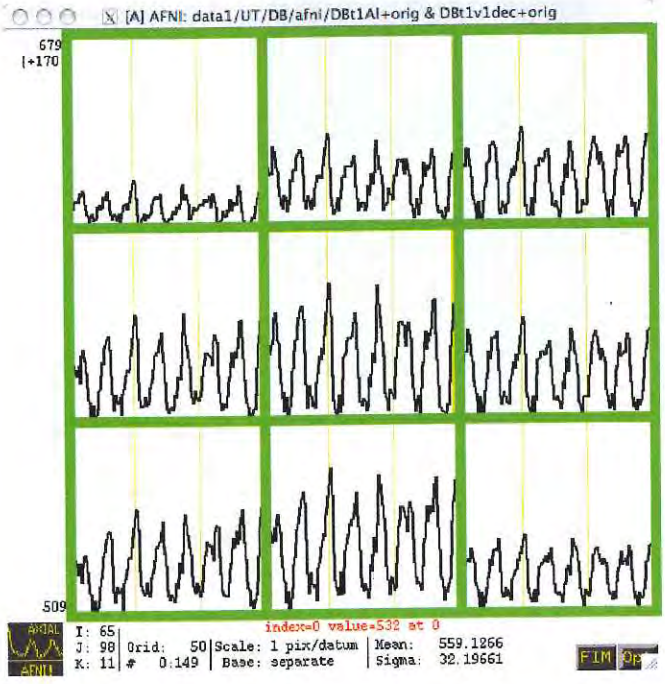


Figure 3–5. (A) Axial slice through the brain. Grids show voxel locations (enlarged) of voxels shown in (B) and (C). (B) A plot of the intensity in nine voxels in the red grid in (A), located in a ventricle. Each black trace shows the intensity of that voxel in each of 150 images, collected every 2 seconds over a 5-minute scan series. No regular variation in the intensity is apparent. (C) A plot of the intensity of the nine voxels in the green grid shown in (A) located in visual cortex. The signal becomes brighter and dimmer as the subject views visual stimuli.

brain area is active in a given stimulus or behavioral task.

To understand these statistical analyses, we use as an example a simple visual stimulation experiment. Figure 3–5 shows a single axial slice through a brain volume, along with the time series. In a single 5-minute scan series, we collected 150 such images. We may plot the intensity of each voxel in a simple x–y plot, with time along the horizontal axis and image intensity along the vertical axis. Plotting the time series of a few neighboring voxels gives even more information. In voxels in visual cortex, we see a periodic variation in the image intensity, with seven peaks over the course of the 5-minute scan. These peaks correspond to the times at which the subject was looking at visual stimuli (in this case, a movie entitled "Winged Migration"). The troughs correspond to times at which the subject was viewing a blank fixation screen.

A number of important points can be made by looking at these time series. First, notice that the time series in neighboring voxels is not identical. Even though they are only separated by 3 mm, the voxel in the top left has a relatively small signal change and the voxel in the middle has a very large signal change (on the order of 10%). This demonstrates that the BOLD fMRI signal is not "all or nothing." The amount of BOLD signal change in a voxel is thought to be proportional to the summed neuronal activity of all of the neurons in that voxel (on the order of a million neurons if the voxel contains entirely gray matter). The voxel in the upper left may have fewer neurons, or it may have neurons that do not respond as strongly to the visual stimulus as the voxel in the center of the display. Next, note that (even for the center voxel) the height of each peak varies. Because each peak represents presentation of a different segment of the movie, the different peaks are a measure of how strongly the neurons in the voxel responded to that particular segment of the movie. Note that the third peak is the highest in the center voxel, and in the other voxels as well, suggesting that all of the neurons in this neighborhood preferred the third movie segment.

Next, notice that the troughs of the signal are not identical, even though the subject viewed the exact same visual stimulus (a blank screen) in each trough. This suggests that the variation in the signal intensity is not determined solely by the neuronal activity in the voxel.

This can be seen more clearly by examining the time series from voxels in the ventricle and adjacent white matter. No periodic variation of the signal intensity related to visual stimulation is observed. However, the signal does have peaks and dips. These variations in signal intensity have a number of sources, including respiration, which introduces changes in tissue oxygenation; the cardiac cycle, which can introduce changes in tissue oxygenation; gross movement of the brain because of pulsations or subject head movements; and scanner or thermal noise. Effectively, this noise in the MR time series limits the ability to detect active brain regions. If the fluctuations in the MR signal introduced by the stimulus are much smaller than the noise in the MR signal, active regions cannot be detected.

Visual examination of the MR time series is a critical step in performing quality control in fMRI experiments. Hardware problems, such as hardware-generated "spikes" in the MR signal, can be easily detected. However, visual inspection of the time series in each voxel to determine whether it is active or not is impractical. Therefore, quantitative methods are used to generate a measure of the degree of activity in each voxel.

We next show how the generalized linear hypothesis can be used to analyze fMRI time series (Fig. 3–6). Each voxel is fit separately, termed univariate analysis. The model requires the analyst to generate a series of predictions about what might be happening in the time series in each voxel. These are known as "predictors" or "regressors" and each must have the same length as the time series in each voxel.

The regressors are placed into two groups: one is the "baseline" group, which can be considered as the "null hypothesis": what should the time series look like in a voxel in which there is no neural activity related to the stimulus. In Fig. 3–6A, a single baseline regressor is shown, consisting of a flat line. The null hypothesis in this case is that the brightness of the voxel should not change over time, because there is change in voxel intensity due to blood oxygenation changes. In a real fMRI experiment, a number of other baseline regressors would also be used. For instance, there are often slow fluctuations in intensity due to respiration. This can be modeled as slow changes in voxel intensity. Or, the subject's head movement can be estimated. This means that an inactive voxel might be predicted to have some intensity fluctuations related to head motion.

The second set of regressors represent the "experimental hypothesis." These are the predictions about what should happen in voxels in which there is neural activity related to the stimulus or task. In Fig. 3–6B, a single experimental regressor is shown, representing a sine wave of the same frequency of the visual stimulus (on and off 8 times over the course of 5-minute run). Voxels containing neurons responding to the visual stimulus would be predicted to get brighter and darker as the visual stimulus turns on and off. In a real fMRI experiment, we might have multiple experimental regressors. If two different types of visual stimuli were shown (such as faces and houses), we would have two separate regressors, each representing the amplitude of the response to each type of stimulus.

The next step is to fit all of the baseline regressors to the voxel time series. For a sample ventricle voxel time series, this is shown in Figs. 3–6C and D. The flat baseline regressor is fit to the time series, and the

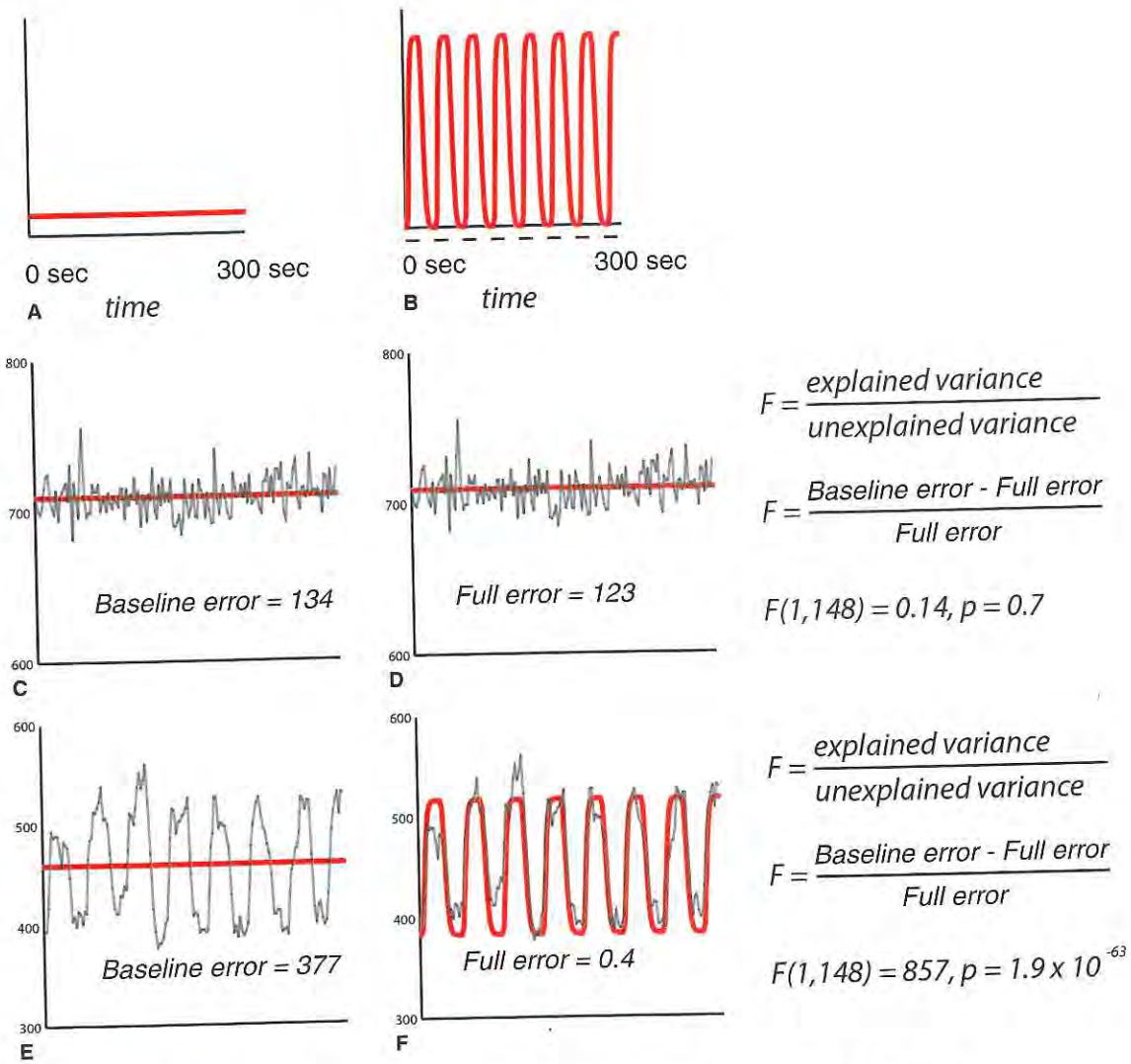


Figure 3–6. (A) Baseline (null-hypothesis) regressor consisting of a flat line, equivalent to constant intensity in the voxel over time. The *y*-axis is arbitrary, since the waveform fits separately to each voxel. (B) Experimental regressor consisting of a sinusoidal waveform. Black lines under *x*-axis show the time when each visual stimulus is on screen. The voxel intensity is expected to increase when the stimulus is on screen and decrease when it is not. (C) The black line shows the actual MR time series from a voxel in a ventricle over the course of a 5-minute MR scan series. The red line shows the best-fit baseline regressor from (A). The baseline error is the difference between the black and red lines. (D) The red line shows the best-fit combination of the baseline regressor in (A) and the experimental regressor in (B) (full statistical model). The full error is the difference between the black and red lines. (E) The black line shows the actual MR time series from a voxel in visual cortex over the course of a 5-minute MR scan series. The red line shows the best-fit baseline regressor from (A). The baseline error is the difference between the black and red lines (large error). (F) The red line shows the best-fit combination of the baseline regressor in (A) and the experimental regressor in (B) (full statistical model). There is a good match between the lines, resulting in low error.

error (difference between best fit and actual data) is calculated. Next, the combination of baseline regressors and experimental regressors are fit to the voxel time series and error is again calculated. This error will always be less, because fitting additional curves always fits better. To determine if the experimental hypothesis is correct, the key question is how much extra variance is accounted for by the experimental regressors. To do this, we calculate an *F*-ratio, the ratio of the explained variance to unexplained variance. In this case, the explained variance is the extra variance accounted for by the experimental regressors (full error – baseline

error) divided by the total error. In Figs. 3–6C and D, this extra variance is very small, leading to a small F-ratio (0.14). To calculate statistical significance, we can examine a table that requires two values, the numerator and denominator degrees of freedom. The numerator degrees of freedom specifies how many experimental regressors are used (as more regressors are used, more variance is accounted for by chance) and the denominator degrees of freedom is approximately the number of points in the MR time series (as this number increases, power increases as well, meaning that it is important to always collect as much fMRI data as possible).

The same process for an active voxel is shown in Figs. 3–6E and F. The baseline error is large because the stimulus-evoked fluctuations in the MR signal are not well fit by the baseline straight-line function. In contrast, the full model error is low, because the sinusoidal experimental regressor fits the MR intensity very well. This results in a very large F-statistic.

This process of model fitting is repeated for every voxel in the brain.

▶ CREATING ACTIVATION MAPS

To construct an activation map, the range of F-values is mapped to a color scale. In the Fig. 3–7, low F-values are assigned a green color, and higher F-values are assigned yellow to red colors. Then, each voxel is colored in according the F-value calculated for it. Figure 3–7B shows the result of this process. Most of the brain is green (meaning low F-rations) except for voxels in occipital lobe, which have yellow to red colors. In order to make the binary judgement of which voxels are "active" or not, a statistical threshold must be chosen. In this example, an F-statistic of ten corresponds to a chance probability of $p < 10^{-7}$. If there are 10^5 voxels in the brain volume, this means the odds of a single voxel being active by chance is $p < 0.01$ using the conservative Bonferroni correction for multiple comparisons. If we display only voxels that have an F-value greater than 15, we see mainly voxels in the occipital lobe.

In order to be more sensitive to weak activations, other processing steps are often performed before the F-ratio is calculated for every voxel. An important preprocessing step is motion correction. Gross head motion can occur as the head settles into the foam padding of the MR head coil, or the body gradually relaxes, over the course of the scan session. By comparing each volume to the initial volume (or the mean of all volumes), these slow motions can be corrected fairly well. Quick motions, such as those that occur during a sneeze are difficult to correct. Therefore, it is important to explain to the patient the necessity of holding the head still and attempt to restrict head motion with padding or commercial products (such as foam beads) while still maintaining patient comfort.

Another popular preprocessing step is spatial smoothing. If each voxel contains noise from physiological or scanner sources, averaging data across several voxels can reduce the noise. The averaging is usually performed only within images acquired at the same time (spatial smoothing) as opposed to images acquired across time (temporal smoothing).

Both motion correction and spatial smoothing reduce the error in the full model fit, increasing the resulting F-statistic and meaning that more voxels pass the significance threshold. There is no single statistical threshold that is correct for all patients and experiments. It is important to examine fMRI data using a range of statistical thresholds, or completely unthresholded, as in the Fig. 3–7, in order to get a clear understanding of the data.

fMRI activations can be considered as a virtual "mountain range." At a very high threshold, only the very tops of the highest mountain peaks will be visible. As the threshold is lowered, the peaks will enlarge as the flanks of the tall mountains emerge above the threshold, and the tops of smaller activation mountains will also be visible.

This effect is shown in Fig. 3–7. At very high thresholds, only voxels in early visual cortex (V1/V2) are active; as the threshold is lower, voxels in other regions of extrastriate cortex become active. In order to determine whether activations are "real," or represent false positives, it is important to examine the time series from the voxels in question.

In this example, we focussed only on a simple two-condition experiment, in which the subject alternately views a moving visual stimulus and a blank fixation screen. However, the exact same analysis procedures can be used with much more complex designs. For instance, we could alternate between viewing a visual stimulus, hearing an auditory stimulus, and viewing a blank screen (and hearing nothing). In this experiment, the full model would contain two different predictor functions, one for the visual stimulus and one for the auditory stimulus. Activation maps could then be created for both functions, either functions, or the contrast between auditory and visual functions. This process can be repeated indefinitely, for instance, ten different types of visual stimuli. A predictor function is created for each predictor functions, and the full model fit to each voxel's time series. Activation maps can be created to show voxels responding to any stimulus (full-F) or any combination of individual stimuli.

▶ VISUALIZATION OF ACTIVATION ON THE CORTICAL SURFACE

Because the T2* EPI images used to create fMRI activation maps are relatively low resolution, they are not

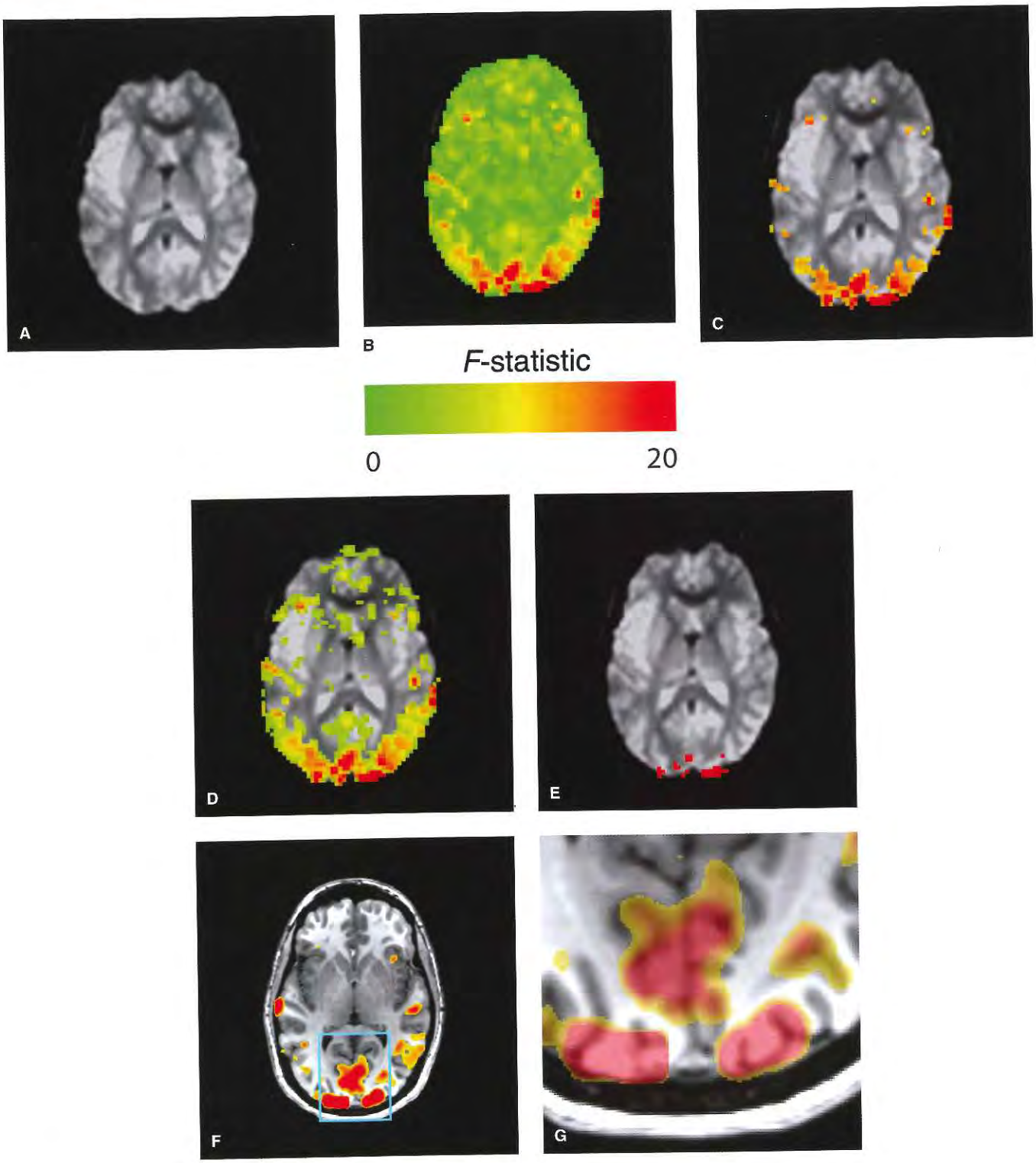


Figure 3–7. Visualization of active brain areas. (A) Original T2* brain image. (B) Brain image with each voxel colored by the significance of the activation (calculated as in Fig. 3–6). Green is low significance and red is high significance (*F*-values shown by color scale). (C) Combination of (A) and (B): active voxels above a certain threshold ($F > 10$) are colored; all other values are from original brain image. (D) Map created with a more liberal threshold ($F > 5$). (E) Map created with a stricter threshold ($F > 15$). (F) Thresholded voxels ($F > 10$) interpolated and overlaid on a high-resolution T1 image from the same subject. (G) Enlargement of visual cortex region in (F). Colored voxels are made transparent to allow visualization of anatomical structure and functional activation. Notice correspondence between gray matter in T1 image and significant (*colored*) voxels.

ideal for clinical decision-making. The typical strategy is to collect a high-resolution T1 anatomy in the same patient during the same scanning session. Then, the active voxels calculated from the EPI images are overlaid on this high-resolution T1 anatomy (see Fig. 3–8).

There are two problems inherent in this approach. First, the subject may move between collection of the T1 anatomy and the functional EPI images. By aligning (motion correcting) the EPIs relative to the T1, this can be compensated for. Second, EPI images are distorted relative to the true brain anatomy pictured in T1 images because of magnetic susceptibility and other artifacts. This means that even without any patient motion, there can be shifts of many millimeters between the EPI and the T1.

Therefore, even if activations maps are viewed overlaid on a T1, it is critical to also visualize them overlaid on the original EPI to ascertain the true location of activation.

Additional processing steps are often performed to make the T1 anatomy more visualization friendly. These include skull stripping (removing the skull) and normalizing the brain so that it is in a standard space. This means that regardless of the patient's position in the scanner, the brain will have a consistent alignment with a template. If the template has labels attached, for instance the location of Brodmann areas or anatomical structures, these labels can be applied to the individual subject brain to make identification of different anatomical structures easier.

Another useful tool for visualizing MRI data are three-dimensional (3D) models of the cortical surface. As shown in the Fig. 3–8, computerized tools are available to automatically segment the MR image into contours tracing the boundary between gray and white matter and the boundary between gray matter and CSF (pial surface of the brain). After creation of these contours, a surface tessellation is created consisting of a mesh of thousands of triangles (middle panel and enlargement). This tessellation is physically equivalent to the cortical surface, which also consists of a single sheet of tissue, folded to fit within the confines of the skull.

When viewed in 2D slices, it can be difficult to identify specific sulci or gyri. When viewed as a whole, it is easy to identify a specific sulcus, such as the superior temporal sulcus, at any location along its anterior to posterior extent. Most of the cortex is buried in the sulcal folds. With a cortical surface model, the links between nodes can be relaxed, allowing the surface to inflate like a balloon. This reveals tissue previously hidden in the sulci. fMRI activations can be visualized on the cortical surface (with the same caveat as overlaying them on the T1). This has the enormous advantage of allowing viewing of all active regions in the brain in a single glance, instead of sorting through stacks of images.

The surface model can also be rotated to get different views of how it would look like in various clinical situations, for instance when exposed during a craniotomy.

Because of the many complex steps required to analyze and visualize fMRI data, analysis is typically conducted with specialized software packages. Many of these packages, including AFNI[15] and SPM,[9] are freely available and include sample datasets, providing an easily accessible way to become familiar with fMRI data.

▶ DRAWING CLINICAL INFERENCES FROM fMRI STUDIES

In order to draw clinical inferences from fMRI studies, we must make a number of inferences from our fMRI activation maps. The first inference is that active voxels reflect activity in neurons located inside the voxel. How justified is this assumption? There is good evidence from experiments in animal models that changes in blood oxygenation are tightly coupled to local neuronal activity.[16,17] This coupling is made possible because of the dense network of capillaries in cortex (see Fig. 3–3).[18] Individual capillaries appear to open and close with metabolic demand, producing a spatially localized fMRI BOLD signal. This is seen in fMRI studies, which image cortical ocular dominance columns (on the order of ~1 mm in size) in human primary visual cortex. These maps are reproducible within the same scanning session and across different scanning sessions on different days.[19] In diseases in which the vasculature is impaired, such as stroke or tumor, the spatial localization of the BOLD signal may also be impaired. In general, it is important to obtain converging evidence for neural organization observed with fMRI. An example of converging evidence is shown in Fig. 3–9.

With high-resolution fMRI, it was shown that regions of human superior temporal sulcus contains segregated, patchy regions that respond to auditory stimulation, such as speech, and visual stimulation, such as face movements, or both. However, this conclusion is tempered by the knowledge that fMRI is an indirect measure of neural activity. Therefore, critical additional evidence is provided by two additional methods, anatomical information from tracer injections in macaque monkeys;[20] and direct recordings of neural responses using penetrating microelectrodes in monkeys.[21] Both of these very different techniques showed similar evidence of patchy organization, suggesting that the organization observed with fMRI is real, and not an artifact of differential vascularization, experimental design, stimulus set, or analysis method.

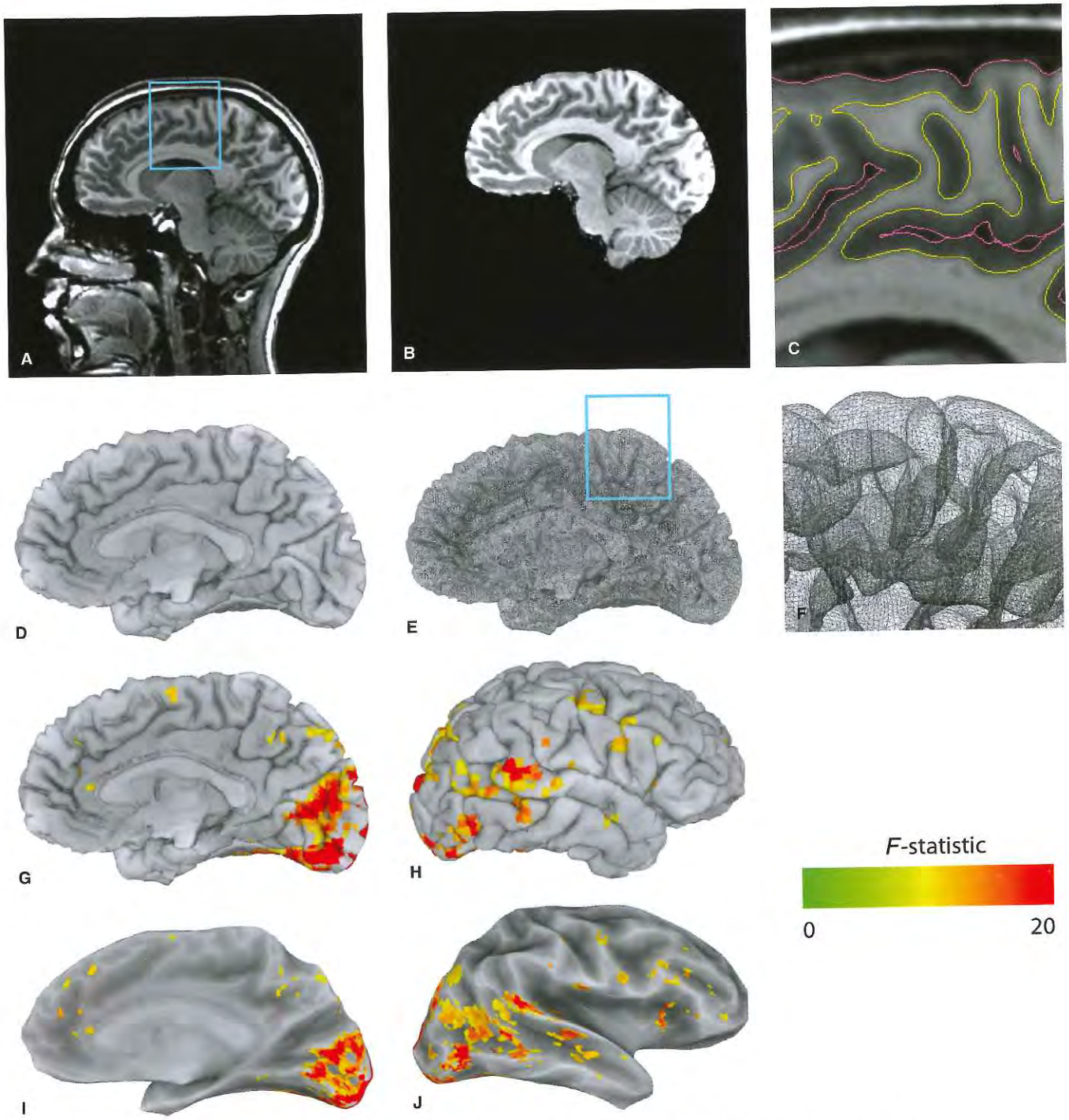


Figure 3–8. Structural MRI analysis stream for functional MRI. (A) Midsagittal section of T1-weighted anatomical scan. (B) Skull-stripped and intensity normalized brain. (C) Automated tracing of cortical gray matter and CSF boundary (*pink line*) and cortical gray matter and white matter boundary (*yellow line*). Enlargement of blue area shown in (A). (D) Cortical surface model, rendered as surface. (E) Cortical surface model rendered to show individual mesh elements. (F) Enlargement showing individual triangular elements (face sets) of cortical surface mesh. (G) Functional activation (threshold $F > 15$) mapped to cortical surface as colored regions. (H) Lateral view of cortical surface with activation. (I) Inflated cortical surface model (medial view). (J) Inflated cortical surface model (lateral view).

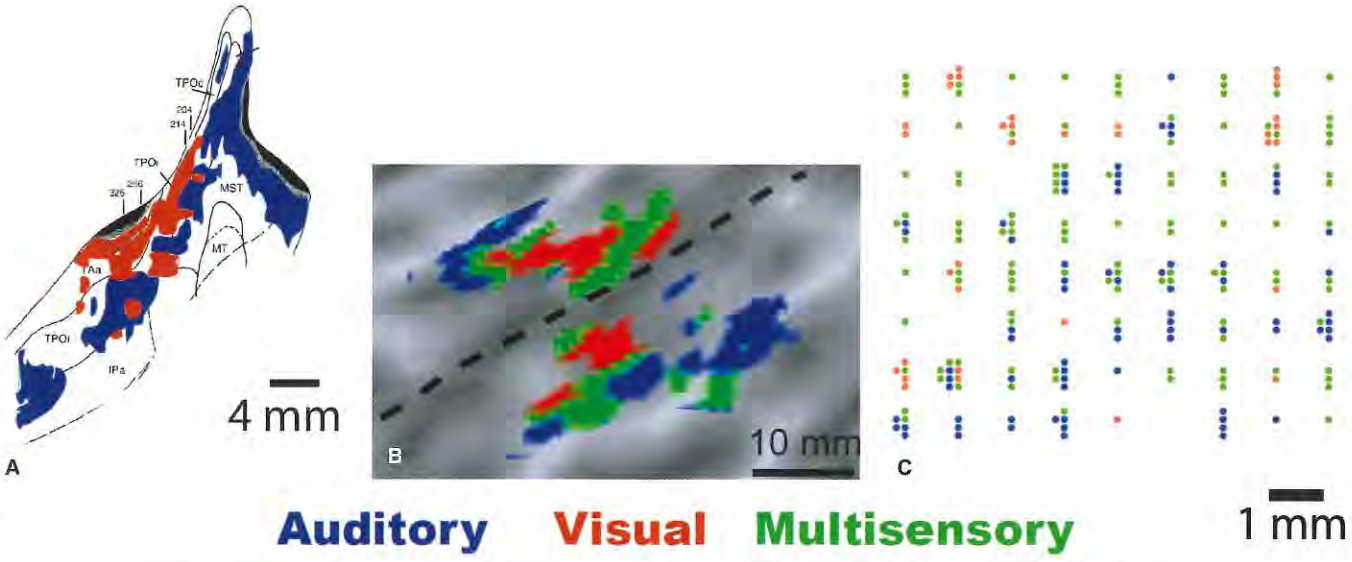


Figure 3–9. Converging evidence on neural organization from functional MRI (fMRI) and other methods. (A) Anatomical data set showing the superior temporal sulcus (STS) of the macaque monkey. (Adapted from Seltzer B, Cola MG, Gutierrez C, Massee M, Weldon C, Cusick CG. Overlapping and nonoverlapping cortical projections to cortex of the superior temporal sulcus in the rhesus monkey: double anterograde tracer studies. J Comp Neurol 1996;370(2):173–190.) Tracer injection of anterograde tracer into macaque monkey auditory cortex and visual cortex. Blue regions of STS receive input from auditory cortex. Orange regions receive input form visual cortex. Note the patchy organization, with some regions receiving auditory input, some regions receiving visual input. (B) fMRI data demonstrating patchy organization of human STS, with interleaved regions responding to auditory stimulation (*blue*), visual stimulation (*orange*), or both (*green*). (Adapted from Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci 2004;7(11):1190–1192.) (C) Single-unit recording from macaque STS showing uneven patchy distribution of neurons responding to auditory stimulation (*blue*), visual stimulation (*orange*), or both (*green*). (Adapted from Dahl CD, Logothetis NK, Kayser C. Spatial organization of multisensory responses in temporal association cortex. J Neurosci 2009;29(38):11924–11932.)

A second inference is that the region of activity is necessary for the cognitive function being studied. This problem is particularly acute in studies of language, memory, and other complex cognitive functions. In a typical cognitive study (such as picture naming) in which subjects are presented with stimuli, make cognitive decisions about them, and then produce a motor response, activations would be expected in unisensory regions responding to the sensory stimulus, multisensory regions that integrate across modalities, cognitive regions that are important for decision-making, and response selection and motor regions that produce the behavioral output. Even without an explicit behavioral task, (such as viewing, but not naming, pictures) subjects may perform language and memory operations when presented with a stimulus (and without a task there is no measure of subjects' alertness or attention, or the amount of processing performed on each stimulus). With or without a task, fMRI experiments typically find activity in many brain regions. Statistical criteria or *a priori* criterion are applied to the fMRI data in order to classify a subset of these active regions as being specifically involved in particular elements of the cognitive task. For instance, if the subjects' task is to press a button in response to an auditory tone, activations on the superior temporal gyrus would be classified as due to the auditory tone, whereas activations along the central sulcus would be attributed to the motor response and associated somatosensory feedback.[22] However, even in this simple task, other brain activations are not so clear cut: secondary somatosensory cortex in the parietal operculum, which might be active in response to somatosensory feedback from the button press, is adjacent to auditory association areas, which might also be active depending on the nature of the auditory tone, with no clear boundary between them. A further complicating factor is brain reorganization or anatomical distortion due to injury or tumor.[23]

Finally, BOLD fMRI depends on patient compliance. For instance, normal volunteers will reliably tap their fingers and keep their head still when instructed to, resulting in excellent maps of motor cortex. Patients who do not understand instructions, or are unwilling to comply, will produce poor quality or uninterpretable activation maps. This is one reason that it is important to collect behavioral data from subjects while they are in the scanner. For instance, if there is little activation in motor cortex during a finger-tapping task, is it because motor cortex is impaired? Or, because the subject was asleep in the scanner and was not performing the task? If an objective measure, such as the number of finger taps per minute, is available, this can answer the question.

In summary, fMRI provides the most powerful technique currently available to measure human brain function. Rapid advances continue more than 20 years after the basic principles were first demonstrated, suggesting a bright future for the technique.

REFERENCES

1. Lauterbur PC. Image formation by induced local interactions: examples employing nuclear magnetic resonance. Nature 1973;242:190-191.
2. Lauterbur PC. Progress in n.m.r. zeugmatography imaging. Philos Trans R Soc Lond B Biol Sci 1980;289(1037):483-487.
3. Ogawa S, Lee TM. Magnetic resonance imaging of blood vessels at high fields: in vivo and in vitro measurements and image simulation. Magn Reson Med 1990;16(1):9-18.
4. Ogawa S, Lee TM, Kay AR, Tank DW. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci USA 1990;87(24):9868-9872.
5. Ogawa S, Lee TM, Nayak AS, Glynn P. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magn Reson Med 1990;14(1):68-78.
6. Bandettini PA, Wong EC, Hinks RS, Tikofsky RS, Hyde JS. Time course EPI of human brain function during task activation. Magn Reson Med 1992;25(2):390-397.
7. Kwong KK, Belliveau JW, Chesler DA, et al. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci USA 1992;89(12):5675-5679.
8. Ogawa S, Tank DW, Menon R, et al. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc Natl Acad Sci USA 1992;89(13):5951-5955.
9. Friston KJ, Frith CD, Liddle PF, Frackowiak RS. Comparing functional (PET) images: the assessment of significant change. J Cereb Blood Flow Metab 1991;11(4):690-699.
10. Bandettini PA, Jesmanowicz A, Wong EC, Hyde JS. Processing strategies for time-course data sets in functional MRI of the human brain. Magn Reson Med 1993;30(2):161-173.
11. Worsley KJ, Friston KJ. Analysis of fMRI time-series revisited–again. Neuroimage 1995;2(3):173-181.
12. Bowerman BL, O'Connell RT. Linear Statistical Models: An Applied Approach. Boston: PWS-Kent Pub. Co., 1990, pp. xvi-1024.
13. Neter J, Wasserman W, Kutner MH. Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs. Homewood, IL: Irwin, 1990, pp. xvi-1181.
14. Rencher AC. Methods of Multivariate Analysis. New York: Wiley, 1995.
15. Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 1996;29:162-173.
16. Logothetis NK. The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. Philos Trans R Soc Lond B Biol Sci 2002;357(1424):1003-1037.
17. Lee JH, Durand R, Gradinaru V, et al. Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. Nature 2010;465(7299):788-792.
18. Harrison RV, Harel N, Panesar J, Mount RJ. Blood capillary distribution correlates with hemodynamic-based functional imaging in cerebral cortex. Cereb Cortex 2002;12(3):225-233.
19. Cheng K, Waggoner RA, Tanaka K. Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. Neuron 2001;32(2):359-374.
20. Seltzer B, Cola MG, Gutierrez C, Massee M, Weldon C, Cusick CG. Overlapping and nonoverlapping cortical projections to cortex of the superior temporal sulcus in the rhesus monkey: double anterograde tracer studies. J Comp Neurol 1996;370(2):173-190.
21. Dahl CD, Logothetis NK, Kayser C. Spatial organization of multisensory responses in temporal association cortex. J Neurosci 2009;29(38):11924-11932.
22. Beauchamp MS, Lee KE, Argall BD, Martin A. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 2004;41(5):809-823.
23. Ulmer JL, Hacein-Bey L, Mathews VP, et al. Lesion-induced pseudo-dominance at functional magnetic resonance imaging: implications for preoperative assessments. Neurosurgery 2004;55(3):569-579; discussion 580-581.