TÜRK NÖROLOJİ DERGİSİ TURKISH JOURNAL OF NEUROLOGY

Nörolojide Yeni Teknolojiler / *New Technologies in Neurology*

Türk Nöroloji Dergisi 2005; Cilt:11 Sayı:1 Sayfa:8-26

Principles of Functional Magnetic Resonance Imaging and its Applications in Cognitive Neuroscience

Şule Tınaz[1, 2], Chantal E. Stern [1, 2, 3, 4]

[1]*Center for Memory and Brain,* [2]*Program in Neuroscience,* [3]*Dept. of Psychology Boston University Boston, MA 02215*

[4]*Athinoula A. Martinos Biomedical Imaging Center Mass General Hospital Charlestown, MA 02129*

ÖZET

Kognitif Nörobilimde Fonksiyonel Manyetik Rezonans Görüntüleme Prensipleri ve Uygulamaları

Fonksiyonel Manyetik Rezonans Görüntüleme (fMRG), milimetrik düzeydeki lokalizasyon gücü sayesinde beyin aktivasyon haritaları elde etmede kullanılan bir nörogörüntüleme tekniğidir. BOLD (blood oxygenation level dependent), diğer bir deyişle kanın oksijenasyon düzeyine bağlı fMRG metodu, kognitif çalışmalarda en sık kullanılan tekniktir. BOLD fMRG nöral aktiviteye karşı gelişen hemodinamik yanıtı ölçer. Kandaki oksijen taşıyıcısı hemoglobin molekülünün, oksijen bağlamış ve serbest halleri manyetik ortamda farklı davranışlar sergiler. İşte BOLD sinyali, hemoglobin molekülünün oksijenasyon düzeyine göre değişen manyetik özelliklerine bağlıdır.

FMRG invazif değildir ve örneğin pozitron emisyon tomografisinden (PET) farklı olarak bireyler radyasyona maruz kalmaz. Bu özellik aynı bireyden uzun süren ve birden çok deney seanslarında veri toplamayı olanaklı kılar. FMRG'nin PET'e kıyasla zamansal çözünürlüğünün daha iyi oluşu da kognitif deney tasarımlarında esneklik sağlaması açısından önemlidir. Bu yazıda MRG fiziğinin temel ilkelerini ve BOLD sinyalini oluşturan metabolik, hemodinamik ve elektrofizyolojik mekanizmaları gözden geçirip, fMRG'de sıklıkla kullanılan deneysel tasarım paradigmalarını örneklerle açıklamaya çalıştık. Ardından fMRG verilerinin analizinde kullanılan temel yöntemlerin teorik altyapısını ve uygulamasını özetleyip fMRG verilerine dayanılarak geliştirilen nöral ağ modellerini tanıttık. Son olarak, deney tasarım tiplerinin ve analiz yöntemlerinin uygulanışını somut bir örnek üzerinde göstermek amacıyla laboratuarımızda Parkinson hastaları ile yapılan ve halen sürmekte olan bir fMRG çalışmasından söz ettik.

Anahtar Kelimeler: fonksiyonel manyetik rezonans görüntüleme

Yazışma Adresi: Şule Tınaz
Center for Memory & Brain Boston University 2 Cummington St, room 109
Boston, MA 02215 Tel: (617) 353 1433 atinaz@bu.edu

ABSTRACT

Functional magnetic resonance imaging (fMRI) is a neuroimaging technique that provides brain activation maps with a spatial resolution of a few millimeters. The BOLD (blood oxygenation level dependent) fMRI method is the most commonly used technique. It measures the hemodynamic response to neural activity. The BOLD fMRI signal is based on the magnetic properties of the oxygenated / deoxygenated hemoglobin which is the oxygen carrier in blood.

FMRI is noninvasive, and unlike in positron emission tomography (PET) individuals are not exposed to radiation. This allows data collection from the same individual over multiple sessions. The relatively high temporal resolution of fMRI compared to PET provides flexibility in experimental designs of cognitive tasks.

In this paper we review the key principles of MRI physics, and the underlying metabolic, hemodynamic, and electrophysiological mechanisms of BOLD signal. We introduce frequently used experimental design paradigms and present examples. Next, we give an overview of theoretical considerations and applications of analysis methods in fMRI time series. Neural network modeling based on fMRI data is also discussed. Finally, we present an ongoing study in our laboratory to demonstrate the application of design types and analysis methods.

Keywords: *functional magnetic resonance imaging*

Dergiye Ulaşma Tarihi/Received: 17.12.2004
Kesin Kabul Tarihi/Accepted: 17.12.2004

INTRODUCTION

Cognitive neuroscience has received increasing attention from scientists and general public in the last decade. The focus of cognitive neuroscience is to address the longstanding brain-mind questions and to bridge the gap between brain function and mental activities using experimental paradigms adopted mostly from cognitive psychology and new experimental techniques. The advent of modern neuroimaging techniques including magnetoencephalography (MEG), positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) has revolutionized this research field. This paper focuses on fMRI which has shown enormous growth in the last decade.

The relationship between brain function and blood flow changes within the brain has been known for over a century. John Fulton's report in 1928 during his neurosurgery residency training under Harvey Cushing is a remarkable example in this regard. Fulton was presented a patient with an arteriovenous malformation in the occipital cortex. There was a bone defect over the primary visual cortex due to an unsuccessful attempt to surgically remove the lesion. When the patient was asked to use his eyes intentionally (e.g., reading a newspaper), the bruit in the back of his head increased in intensity. However, it did not increase when the patient used his other senses (e.g., smelling vanilla, trying to hear a quiet conversation), not even when he passively viewed a light source without mental effort[1].

The nature of this observation and of several others in the past can now be addressed and investigated using modern functional neuroimaging techniques, specifically fMRI, in which changes in blood flow and blood oxygenation can be measured during cognitive events.

The first fMRI study that investigated the cognitive task related brain activity changes was performed in 1991 at Massachusetts General Hospital. This study examined cerebral blood volume (CBV) changes in the human visual cortex during a visual stimulation task using gadolinium diethylene-triamine-penta-acetic acid (Gd-DTPA)[2]. Gd-DTPA is a paramagnetic contrast agent that initially causes signal loss following injection. As it passes out of the neurovasculature, the MR signal returns back to normal. The MR signal intensity time course can be converted to a contrast agent concentration time course, and the area under the concentration time curve is proportional to the regional CBV[3]. The study showed significant increase in CBV in the primary visual cortex during visual stimulation compared to a baseline state of resting in darkness.

The first fMRI studies using the BOLD (Blood Oxygenation Level Dependent) technique were published in early 1990s[4,5]. Kwong et al. (1992) investigated changes in blood oxygenation (using a gradient echo sequence which is sensitive to magnetic susceptibility) and in cerebral blood flow (CBF) (using a spin echo inversion recovery sequence which is sensitive to flow) during visual stimulation and a manual motor task. Visual stimulation and motor tasks induced increased signal intensity in the primary visual and primary motor cortices, respectively. Increased signal intensity corresponded to increased CBF in the flow sensitive images, and to increased venous blood oxygenation in susceptibility sensitive images[5].

We begin with reviewing the basic principles of MRI physics for a better understanding of fMRI and the nature of the BOLD signal.

Basic Principles of MRI Physics

In MRI, magnetic properties of hydrogen (H) atoms, in other words protons, are utilized because they are abundant in tissues and have the simplest atomic configuration containing only one proton in the nucleus and an orbiting electron. Protons and neutrons possess an intrinsic angular momentum called "spin" which is a purely quantum mechanical phenomenon. Its speed cannot be changed, the only feature that can change is the axis of spin (the direction of the angular momentum). Protons and neutrons have oppositely oriented spins. When they combine to form nuclei, the result is that nuclei with an even number of protons and an even number of neutrons have no net spin because opposite spins cancel out, whereas nuclei with an odd number do have a net spin. Protons also have a magnetic dipole momentum. They behave like a tiny magnet with the north/south axis. When placed in a magnetic field (B_0) the proton rotates into alignment with the field because of its magnetic dipole momentum. Because the proton also has angular momentum, the spin axis of the proton precesses around the field axis and creates its own magnetic field (M_0) that is parallel to and smaller than the main magnetic field B_0 (Figure1)[6].

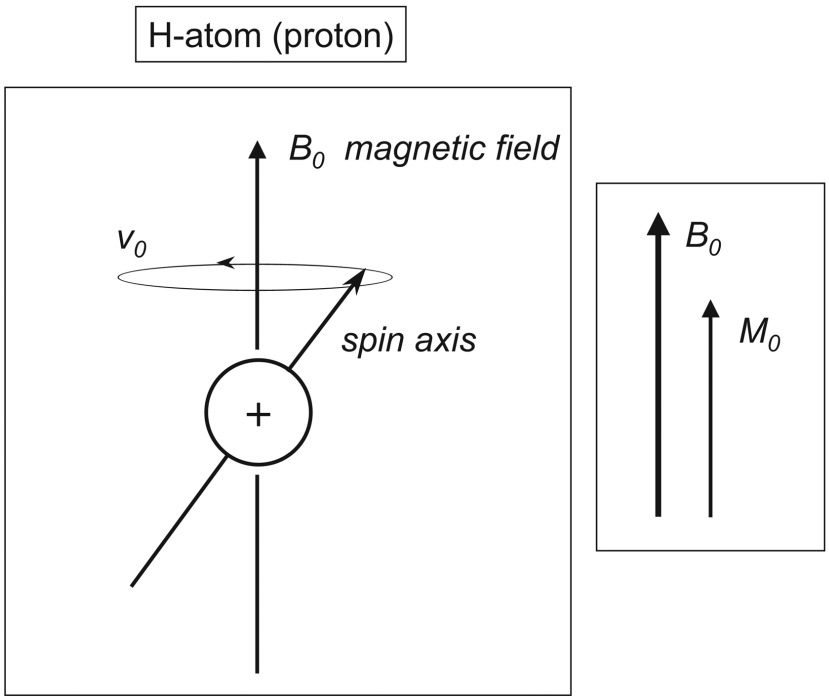


Figure 1. H-atom precessing in a magnetic field (B_0) (adapted from Buxton RB. Nuclear magnetic resonance; Introduction to functional magnetic resonance imaging: Principles and techniques. New York, Cambridge University Press 2002:64-85).

The frequency at which the protons precess is called the Larmor frequency (ω):

$$\omega=\gamma B/2\prod$$

B: Magnetic field strength

γ: gyromagnetic ratio (an intrinsic magnetic property specific to each nucleus)

The Larmor frequency for H-atoms in 1.5 Tesla (T) is 64 Mhz. This is also called the resonance frequency of the H-atoms in 1.5 T (Just to give a sense of how strong a 1.5 T magnetic field is: 1 T=10000 Gauss; earth's magnetic field=0.5 Gauss. 1.5 T-magnet is 30000 times stronger than earth's magnetic field).

How is the MRI signal generated and detected?

Even though the events underlying the MRI signal take place in the quantum world, the basic mechanisms can be described in terms of classical mechanics for simplicity, namely Faraday's Law of electromagnetic induction. Faraday's Law states that charged particles in motion create magnetic fields, and changing magnetic fields create electric fields. This is the basic physical process in MRI that generates a measurable signal in a detector coil. Electric fields can be generated by varying the magnitude of the magnetic field B, or the area A that is enclosed by the loops of the coil, or the angle between B and A with time. In MRI, a radiofrequency (RF) pulse is applied to the protons, which forces them to rotate around 90° in the main magnetic field B_0. This creates a dipole magnetic field that changes with time as the protons rotate. The changing magnetic field will induce current in the receiver coil, and the MR signal can be detected as a voltage change generated by the rotating protons.

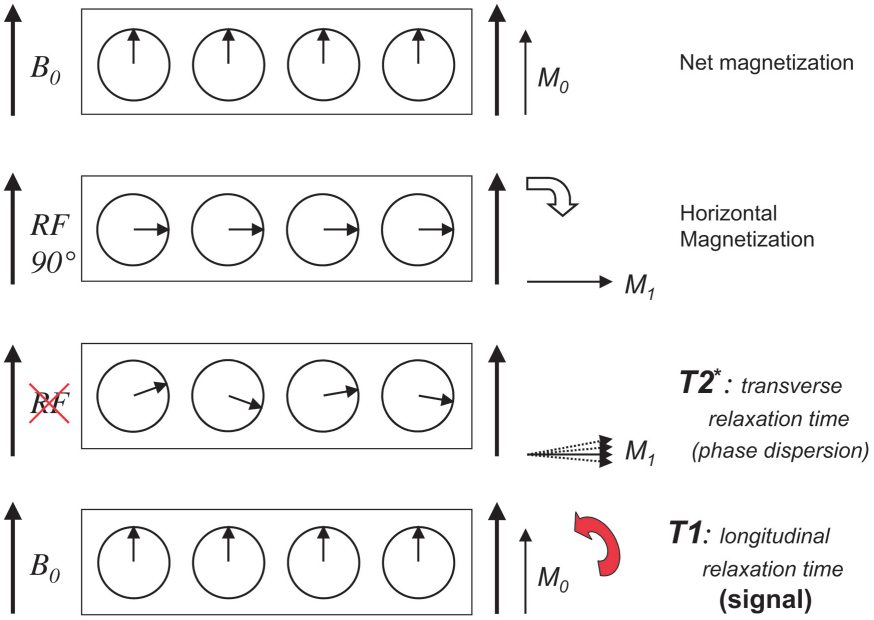


Figure 2. Generation of MRI signal *(see text)*.

Important is that the H-atoms are targeted selectively by the RF pulse. To do that, an RF pulse that contains the Larmor frequency of the H-atoms must be used (64 MHz in 1.5 Tesla). In this case, Larmor frequency corresponds to the resonance frequency of H-atoms. Resonance is the natural frequency of a system's oscillation, and a system absorbs energy best at this frequency. When a radiofrequency (RF) pulse at the Larmor frequency of H-atoms is emitted, they absorb the energy, and their spin axis changes its orientation. This phenomenon is called "resonance absorption"[6].

Figure 2 demonstrates the generation of the MRI signal. Protons aligned with the main magnetic field B_0 are bombarded with a 90° RF pulse. Their spin axis rotates,

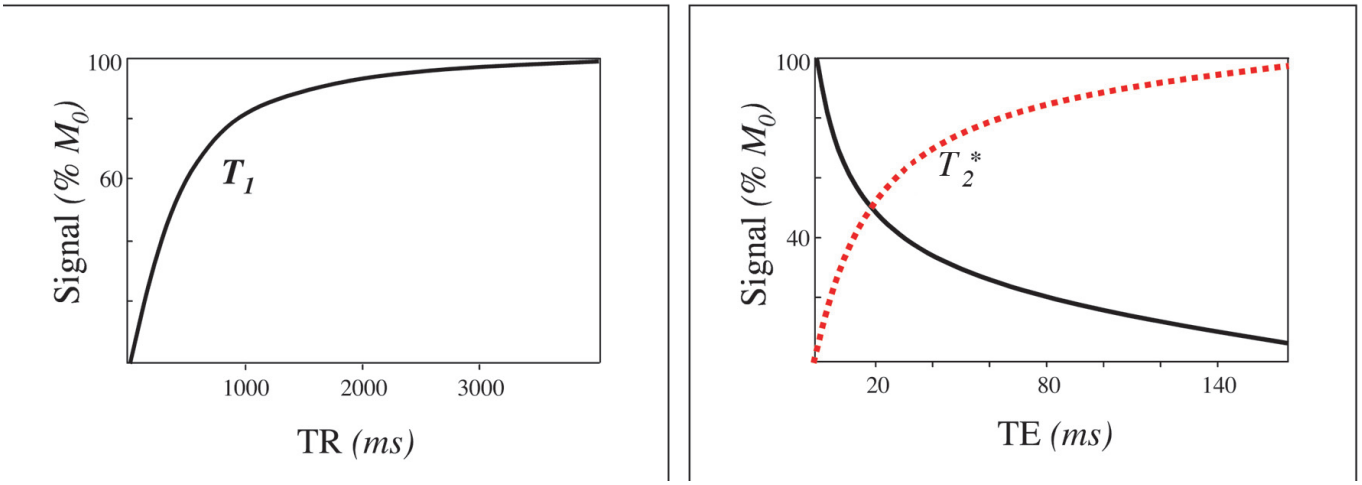


Figure 3. T_1 and T_2 relaxation curves (adapted from Buxton RB. Relaxation and contrast in MRI; Introduction to functional magnetic resonance imaging: Principles and techniques. New York, Cambridge University Press 2002:155-183).

and the net magnetization (M_1) is only in the horizontal plane now. When the RF pulse is stopped, the spins will realign with B_0 again because B_0 is the only dominant force in the absence of the RF pulse. In other words, protons will emit the energy they absorbed from the RF pulse which is detected as a voltage change. The MRI signal is based on this voltage change. When the RF pulse stops, the spins of H-atoms enter a brief fuzzy period on the horizontal plane in which they lose phase coherence and start precessing at different rates. This phenomenon is called "phase dispersion", and it is caused by interference of microenvironmental, technical and physiological factors within the process. Each H-atom is exposed to different intra - and intermolecular interactions depending upon its microenvironment. For instance, a H-atom within myelin would behave differently than the H-atom in the cerebrospinal fluid. Technical factors include noise caused by magnetic field inhomogeneity due to the imperfections of the magnetom. There is also physiological noise including cardiac pulsation, breathing and head movement[6].

In summary, the MRI signal is generated by the RF energy at the resonance frequency of H-atoms that is first absorbed and then released by these atoms when they relax. There are two relaxation times, known as T_1 and T_2. T_1 is the time required for M_0 to return to 63% of its original value following an excitation RF pulse. It is also called the "longitudinal" or "spin-lattice" relaxation time[7]. When the spins return to their original position after excitation they emit their energy to their surroundings ("lattice"). T_2 is the time required for M_1 to decay to 37% of its initial value on the horizontal plane. It is also called the "transverse" or "spin-spin" relaxation time. After the RF pulse, spins start loosing their phase coherence while at the same time they reorient themselves along B_0. The spin energy will dissipate in the form of energy exchange among nearby spins with incoherent phases[7]. Obviously, T_2 will always be less than or equal to T_1. In gradient echo pulse sequences T_2 is referred to as T_2*, the asterisk indicating the contribution of extrinsic noise to T_2.

Gradient Echo Sequences

As can be seen in the diagram in Figure 3, longitudinal relaxation corresponds to signal gain, whereas transverse relaxation corresponds to signal decay (T_2*, black line). Given that the process during T_2 is actually a decay of signal, then how is it possible to acquire T_2–weighted images?

The main idea in gradient echo sequences is to exert control on phase dispersion phenomenon in order to reverse this inevitable process and benefit from it. To do that, a small magnetic gradient (perturbation) is applied to the main magnetic field B_0 with a magnitude of approximately 1% of B_0, immediately after the RF pulse is stopped. Since the Larmor frequency of protons is proportional to the magnetic field strength, protons will start precessing at different rates depending upon their position in the gradient. The Larmor formula can be rewritten as follows: $\omega i = (B_0 + Gri)$ (r: position of the proton, Gri: magnetic gradient vector). Protons in the lower end of the gradient will slow down, whereas protons in the higher end will speed up (Figure 4, left panel). In other words, phase dispersion will be created experimentally. When another gradient with the same magnitude and duration as the first one is applied immediately in the

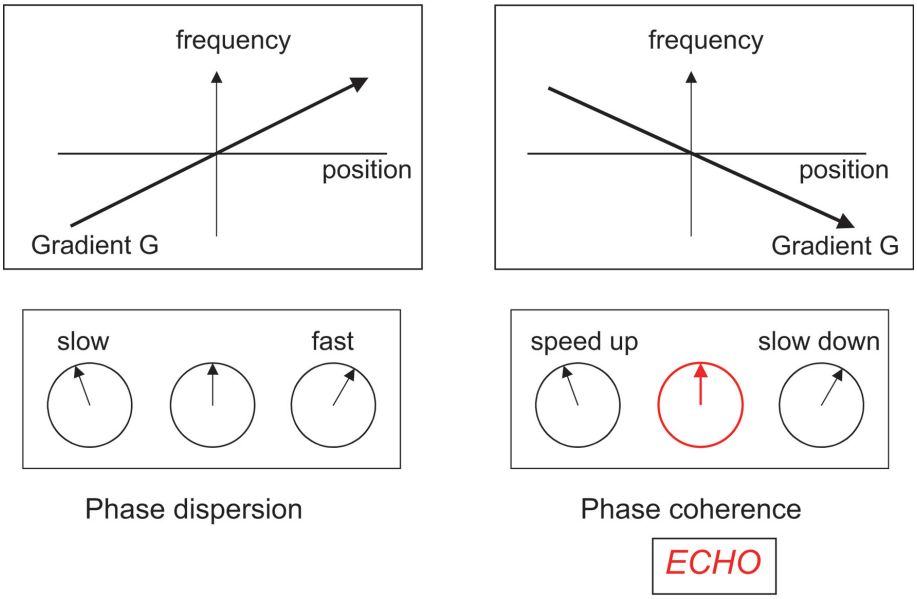


Figure 4. Generation of "echo" signal (adapted from Buxton RB. Magnetic resonance imaging; Introduction to functional magnetic resonance imaging: Principles and techniques. New York, Cambridge University Press 2002: 86-103).

opposite direction, protons that have slowed down will speed up, and those that have speeded up will slow down, and all protons will start precessing at the same rate again (Figure 4, right panel). Consequently, the phases that have dispersed due to the first gradient will reach coherence after the second gradient, and form the MRI signal that is called "echo" (reversal of decay) which corresponds to signal gain as illustrated in Figure 3 (red dashed line)[8]. Application of these small gradients in opposite directions in rapid succession is called "echo planar imaging" (EPI) which is the most widely used technique in functional MRI experiments[3].

The two other well-known parameters in MRI are repetition time (TR) and echo time (TE). TR is the time between successive RF excitation pulses. It determines the amount of T_1 weighting. TE is the time between the excitation pulse and the echo. It determines the amount of T_2 weighting[8]. TR and TE are parameters that can be manipulated by the experimenter in order to acquire the desired weighting in scans, whereas T_1 and T_2 are parameters that are specific to tissue properties and provide the contrast between different tissue types. The contrast information is only useful if it is considered in the three dimensional (3D) anatomical image of the brain. To obtain the 3D image of the brain, three small magnetic field gradients need to be applied perpendicular to each other. These three gradients are called slice selection, readout or frequency encoding, and phase encoding [8]. The basic principle is based on the fact that each proton experiences a different magnetic field strength in the gradient and precesses at a different rate compared to the other protons according to its position in the gradient. The differences in precessing frequencies of protons along three dimensions provide the necessary information to localize the protons in 3D space.

The raw signal is the sum of frequencies in time with different amplitudes. It is analog (continuous). The mathematical tool Fourier transformation (FT) allows us to describe any function of position as a function of spatial frequencies. By applying FT to decompose the signal into its multi-component frequencies we obtain the k-space. K-space is the distribution in space in terms of amplitudes of different spatial frequencies k (k is the inverse wavelength). Low spatial frequencies usually have the largest amplitude and so contribute most to the image intensity, but the high spatial frequencies provide spatial resolution in the image[9]. Application of an inverse FT to the k-space representation yields the 3D image. The 3D image unit is called a "voxel" (analogous to pixel in 2D).

BOLD (Blood Oxygenation Level Dependent) Signal

Neuronal activity causes increase in blood flow and energy requirement in terms of glucose and O_2 consumption in the active region[10]. When a neuron fires an action potential the intracellular concentration of Na^+ increases, and the Na^+-K^+-ATPase pump works harder to reestablish the homeostasis which is an energetically demanding process. The extracellular concentration of some metabolites and ions increases as well which decreases the extracellular pH resulting in vasodilatation and increased blood flow. The reuptake and repackaging of glutamate in the presynapyic terminal also requires energy[22] (Figure 5).

In a PET study using somatosensory stimulation, Fox and Raichle observed 30% increase in local CBF, but only 5% increase in local O_2 metabolism compared to the resting state[10]. This discrepancy (uncoupling) between blood flow and O_2 consumption during neuronal activation underlies the BOLD signal. With increased blood flow to

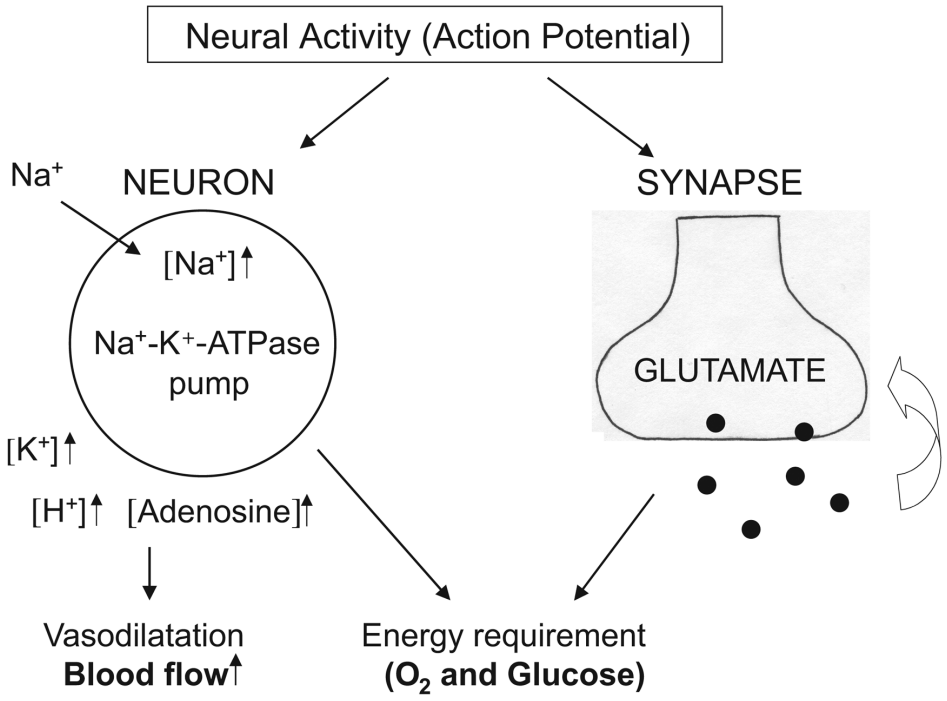


Figure 5. Neuronal activity increases blood flow and energy requirement.

the active neural tissue, the oxygenation level of the venous blood increases relative to the oxygenation level under resting condition, because increased blood flow is not accompanied by proportionately increased O_2 utilization by the active neural tissue. (For instance, in their experiments, Kwong et al. (1992) found that, assuming an initial venous oxygenation of 60% with constant O_2 consumption, the 2% MR signal increase in the active state corresponded to a final venous oxygenation of 75% [5]). Hemoglobin molecule (Hb) is the oxygen carrier in blood and composed of four globin chains, each containing a heme group. In the center of each heme group is an iron (Fe^{++}) atom that binds the elemental oxygen molecule. Hb in its deoxygenated form is paramagnetic. When Fe is not covered by oxygen molecules it interferes with and distorts the magnetic field (increases magnetic susceptibility), causing MR signal decay [11]. When the venous oxygenation increases due to neural activation, in other words, the oxy-Hb / deoxy-Hb ratio increases in the venous end, Fe is covered by oxygen molecules and cannot interfere with and distort the magnetic field anymore. Consequently, MR signal will increase relatively in the active state of neurons [12] (Figure 6). The changes in CBF and oxy-Hb / deoxy-Hb ratio in response to neural activation are local phenomena. This is the rationale behind the fact that BOLD scans are collected T_2-weighted because T_2 is sensitive to local dynamic changes in the microenvironment. The oxygenation level of Hb determines the signal strength during T_2[4].

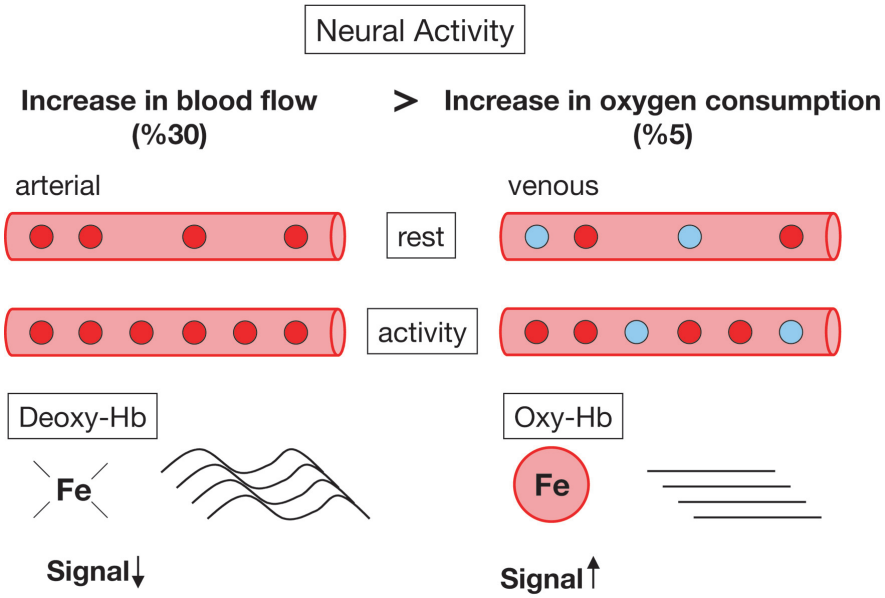


Figure 6. BOLD signal is related to the oxygenation state of Hb. Red balls: Oxy-Hb; blue balls: Deoxy-Hb.

Figure 7 shows the canonical hemodynamic response to a brief stimulus. The canonical hemodynamic response function (hrf) is delayed in onset and temporally extended. It begins to increase within 2s of the stimulus onset, reaches the peak value within 5-7s, and returns to baseline in 20-30s[13, 14].

BOLD signal depends upon combined changes in CBF, cerebral metabolic rate of O_2 ($CMRO_2$) and CBV. The complex relationship between these variables is still an area of debate and intensive investigation.

There are several hypotheses addressing the discrepancy between blood flow and O_2 utilization. One of them is the oxygen

Hemodynamic Response Function (hrf)

- 4-6 s peak
- 20-30 s baseline
- Signal ratio: 0-5%

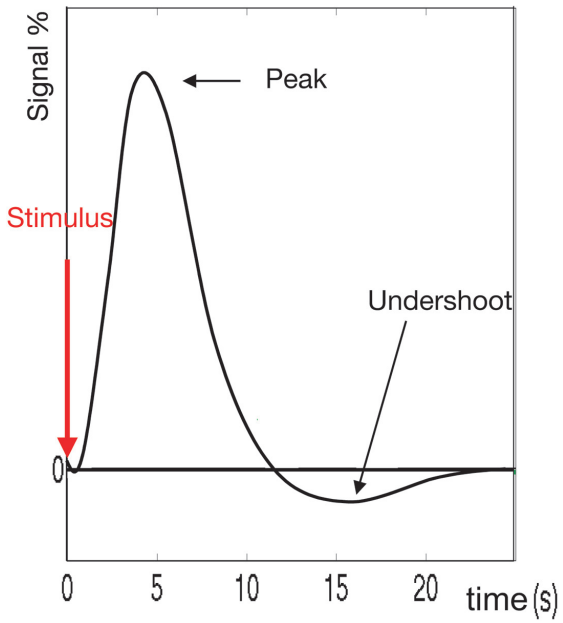


Figure 7. Hemodynamic response function (adapted from Buechel C, Fsiston K, Henson R, Josephs O. (Epoch &) Event-related fMRI; SPM course 2002 slide presentation. Wellcome Dept. Of Cognitive Neurology& Institute of Cognitive Neuroscience, University College London, UK)

limitation model that states that O_2 transport is difficult due to its low water solubility[15]. In addition, local blood velocity in arterioles, capillaries and venules increases which in turn decreases the capillary transit time of O_2. As a consequence CBF increases substantially to support a small increase in the $CMRO_2$.

CBF change and CBV change have competitive effects on the BOLD signal change. Increase in CBF generates positive BOLD signal. But when CBV increases, the amount of blood increases as well, and that in turn increases the deoxy-Hb ratio, which causes a negative BOLD signal. There is evidence that CBV changes lag behind CBF changes provided by the "Balloon" and "Windkessel delayed compliance" models [16, 17]. Both models take into account the biomechanical properties of the vessels such as viscoelastic resistance to rapid volume changes. When there is a steady relationship between CBF and CBV the BOLD signal has a simple hrf shape. But when the CBV changes rapidly the initial viscoelastic resistance and then the slow relaxation of vessels complicate the dynamics, causing nonlinearities in

Glucose Metabolism

Both blood flow and glucose metabolism increase substantially in activated areas of the brain. However, this parallel increase does not necessarily indicate a link between the two. There is evidence suggesting that increase in cerebral metabolic rate of glucose (CMRGlc) is independent of CBF changes. A PET study demonstrated that under different levels of induced hypoglycemia the CBF in response to vibrotactile stimulation did not change[18]. There is also a discrepancy between glucose and oxygen metabolism during activation. Neural activity increases glucose metabolism much more than oxygen consumption[19]. This is due to nonoxidative mechanisms of glucose consumption. Most of the

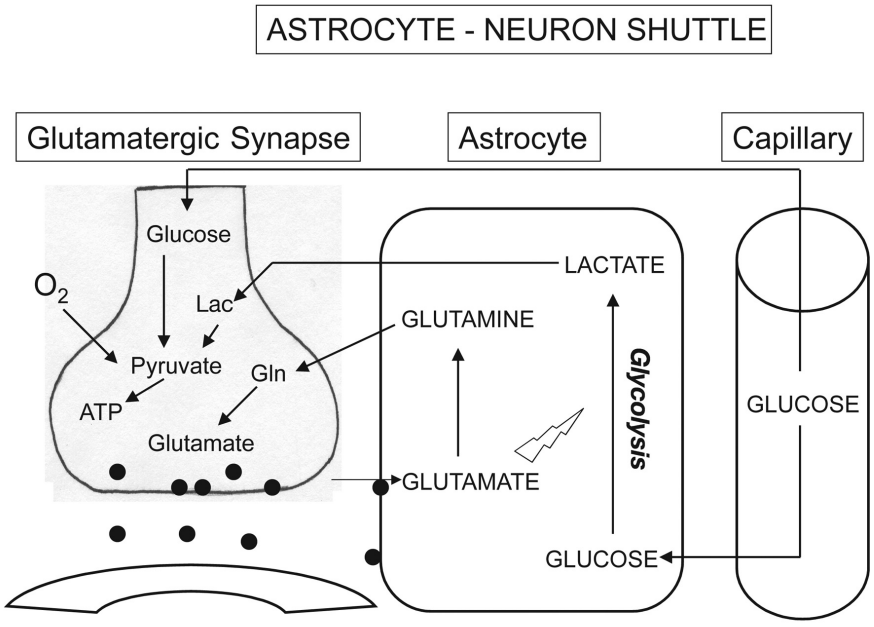


Figure 8. Astrocyte-Neuron shuttle (adapted from Magistretti PJ, Pellerin L. Cellular bases of brain energy metabolism and their relevance to functional brain imaging: evidence for a primary role of astrocytes. Cerebral Cortex 1996; 6:50-61).

increase in CMRGlc is related to glycolysis, and not to full oxidative metabolism[19, 20, 21]. This transient, nonoxidative glucose consumption together with increased blood flow may provide an increase in the oxy-/deoxy-Hb ratio sufficient to be detectable by fMRI.

Glucose metabolism is regulated by the astrocyte-neuron lactate shuttle. For each synaptically released glutamate molecule taken up by an astrocyte, one glucose molecule enters the astrocyte. Glutamate triggers glycolysis. Glycolysis produces two ATP molecules and two lactate molecules that are released and consumed by the neuron to yield 18 ATPs through oxidative phosphorylation. Glutamate is cotransported with Na^+, resulting in an increase in the intracellular Na^+ concentration, leading to the activation of the Na^+/K^+ ATPase pump. One glucose consumed through glycolysis produces two ATPs, one is used by the Na^+/K^+ ATPase pump, the other ATP is used for the synthesis of glutamine from glutamate by the enzyme glutamine synthase[22] (Figure 8).

Synaptic activity is tightly coupled to glucose uptake. The largest portion of energy consumption is attributed to the post-synaptic effects of glutamate. In an attempt to examine the correlation between hemodynamic response and electrophysiological activity, Logothetis et al (2001) recorded multi-unit activity (MUA) and local field potentials (LFP) from the monkey primary visual cortex in a 4.7 Tesla scanner under visual stimulation with a rotating checkerboard pattern[23]. MUA reflects the spiking output of a neural population, whereas LFP is the weighted average of input signals of a neuronal population. It reflects not only the spike activity but also the subthreshold local integrative processes. They found that the BOLD signal significantly correlated with LFPs, but not with MUA. It seems that the BOLD signal correlates with the energetically expensive synaptic activity rather than the spiking activity[23].

Experimental Designs

There are several experimental design types in fMRI including categorical, parametric and factorial designs. In parametric designs, changes in brain activation in response to a certain parametric modulation can be investigated (e.g., learning across time, manipulation of working memory load). For example, Maguire *et al.* (2003) examined the effect of remoteness on the neural basis of autobiographical memory retrieval using a parametric design. They found a gradual decrease in right hippocampus activity the more remote the autobiographical memories[24]. In factorial designs, the effect of two ore more variables each with different levels can be measured. In addition, the main effects of each variable and the interaction between them can be calculated. Henson *et al.* (2000) investigated the role of stimulus familiarity in repetition priming using a two-by-two factorial design. The events of interest were first and second presentation of familiar (F1 and F2) and unfamiliar (U1 and U2) stimuli. They performed two orthogonal comparisons in order to identify regions showing greater response to familiar than to unfamiliar stimuli, (F1+F2) – (U1+U2), and regions showing an interaction between familiarity and repetition, (F1 – F2) – (U1 – U2). A right fusiform region demonstrated reduced response to repetition of familiar stimuli and increased response to repetition of unfamiliar stimuli[25]. A similar design was used to examine the effects of pharmacological modulation on repetition priming in a face recognition paradigm where subjects were given placebo, scopolamine (cholinergic challenge) or lorazepam (GABAergic challenge) prior to the study phase. A right fusiform region showed decreased response to repetition of famous faces and enhanced response to repetition of unfamous faces in the placebo group. The same region showed decreased activation to repetition of famous faces in the lorazepam group (i.e. a main effect of repetition). No significant repetition effects were found in the scopolamine group suggesting that cholinergic challenge impairs repetition priming by possibly interfering with the acquisition phase of the stimuli[26].

Categorical designs, specifically the "cognitive subtraction" approach, are widely used in fMRI studies to examine the underlying neural substrates of a particular cognitive function. In cognitive subtraction designs, an activation task that includes the cognitive component of interest is contrasted with a control task that has all components of the activation task in common, except the cognitive component of interest. The assumption is that by subtracting the activation observed in the control task from that in the activation task one obtains a subtraction map of the brain regions that are associated with the cognitive component of interest. Subtraction designs have some shortcomings. One of them is the difficulty in finding the optimal control task that controls for all but the component of interest. Also, the addition of a cognitive component does not necessarily correspond to the activation of extra brain regions in a linear fashion. More likely, the newly added cognitive component would affect the neural implementation of the previous ones that are common to both activation and control tasks. Consequently, the subtraction map would include not only the differences due to the added component but also the interaction between the added and shared components[27, 28].

Block and Event-Related Design Paradigms

Block design paradigms were introduced to the field of

functional brain imaging with the advent of positron emission tomography (PET) studies in late 1980s. Block designs with relatively long durations are a necessity in PET studies because it takes about 20 s for the radiopharmaceutical (radioactive H_2O) to arrive in the brain after intravenous injection, and blood flow changes can be measured in about one minute [1] . Whereas in fMRI, the higher temporal resolution of this technique compared to PET provides the capability of detecting activity within seconds of stimulus onset and therefore does not require the use of block designs. Instead, fast event-related designs can be used in fMRI experiments[29].

The fact that the hemodynamic response to individual trials is temporally extended causes a potential problem to mixed trial event-related experiments. The hemodynamic response to temporally adjacent trials may overlap making it impossible to separate out the contribution of each trial to the signal. One way of alleviating the overlap is to space the trials sufficiently far apart. However, this sets limits to the number of trials in a certain period of time which in turn reduces the power of the signal. A more favorable approach is to insert a variable length interval between the trials and to counterbalance the presentation of different trial types. Counterbalancing means that each type of trial is followed and preceded by the other equally often. Then the hemodynamic response to each trial type can be selectively averaged. The overlap between adjacent responses cancels out by simple substraction when the two trial types are contrasted with each other[14, 30].

From a cognitive point of view, ER designs have several advantages. Cognitive performance includes two main components. One is stimulus-driven, bottom-up processing. The other one is rule-based, top-down processing in which subjects develop and apply cognitive strategies based on the rules and demands of the task at hand. The presentation of two different trial types A and B in blocks may easily enhance skill acquisition on those trials because subjects are exposed to the same trial type repetitively. In addition, subjects may use different cognitive strategies in blocks A and B. These two confounding factors render the interpretation of the performance differences between A and B difficult in blocked designs. Using mixed event-related designs alleviates these problems because subjects have no way to predict which trial type will be presented next and no basis to bias their responses by using differential strategies[31].

ER designs allow post hoc classification of trials based on subject performance. For instance, an ER design can be used in encoding tasks in which different categories of stimuli (e.g. abstract-concrete words, indoor-outdoor pictures) are presented in a series of intermixed trials[32,33]. After the scan session subjects can be tested on their subsequent memory of the stimuli that were studied during the scan session. The fMRI scans can then be classified based on whether the stimulus was subsequently remembered ("R") or forgotten ("F") on the postscan memory test. Analyzing the R and F scans separately would reveal brain regions whose activity during the encoding period predicts subsequent remembering.

ER fMRI designs can easily adopt the "oddball tasks" (target stimulus detection in the presence of interleaved distractor stimuli) widely used in event-related potential (ERP). A better understanding of the neuroanatomical substrates of this task can be achieved by fMRI due to the high spatial resolution of this technique compared to ERP. ER fMRI designs are the method of choice in this regard because the hemodynamic response evoked by oddball and distractor stimuli can be tested separately and the underlying neural networks can be identified [34].

Some events can only be indicated by subject. Kleinschmidt et al. (1998) developed an object perception task using ambiguous figures (same figure can bee seen as a vase or as the profiles of two people, or as a young woman with a hat or an elderly woman with a scarf) to dissociate perceptual from stimulus-driven mechanisms. An observer can only be aware of one the two incompatible percepts at any given moment, but over time experiences spontaneous reversals ("flips") between the two percepts. Subjects were instructed to report their conscious experience of the visual scene by key-presses, defining the occurrence of perceptual reversals and the presence of stable percepts[35].

For some cognitive tasks block designs are more favorable, such as tasks that evaluate learning across time, or when the task cannot be accomplished in only few seconds. For instance, in an implicit motor sequence learning study, blocks of repeating sequences (task of interest) and random sequences (control task) were used so that subjects could have enough exposure to the repeating sequence throughout the block[36]. Block designs also provide powerful signal. This feature becomes especially important when one has to work with a small group of subjects (e.g. a patient population).

Many cognitive functions involve a combination of state-

related ("mental mode") and trial-related processes. Mixed blocked and event-related designs provide a powerful tool to dissociate transient trial-related activity from sustained state-related activity. In mixed designs, experimental blocks are intermixed with rest blocks (visual fixation), and each experimental block consists of different trial types that are presented at varying intervals ("jittered")[37]. The reason to insert variable gaps between trials is to ensure that the trial-related transient activity declines during these gaps, so that overlap between different trials can be avoided. Trials and blocks can be modeled and analyzed separately. This way, sustained brain activation observed throughout the blocks and transient brain activation in trials can be attributed to state-related and trial-related processes, respectively[38].

Donaldson *et al*. (2001) used a mixed design paradigm in a recognition memory task with old and new words in order to identify brain regions that are transiently active and are associated with "retrieval success" of the items, those that show sustained activation and are related to the "retrieval mode", and finally active areas reflecting a combination of both state- and trial-related processing. They found transient, item-related activity in parietal areas and anterior left frontal cortex; sustained, state-related activity including decreased activity in bilateral parahippocampal cortex; sustained and transient activity in left middle frontal gyrus, bilateral frontal operculum, and medial frontal gyrus[38].

Preprocessing of Images

Prior to the statistical analysis, imaging data need to be preprocessed. Basic preprocessing steps include head motion correction, spatial normalization to an anatomical template, and smoothing. Data processing and analysis can be carried out using several different available analysis streams (e.g. Statistical Parametric Mapping-SPM, Wellcome Dept. Of Cognitive Neurology, University College London; Analysis of Functional Neuroimages-AFNI, National Institute of Mental Health; FSFAST, Mass General Hospital). This section provides a brief overview of the preprocesing and analysis steps included in SPM software. For more detailed information and thorough descriptions the SPM website is recommended (see references).

Motion correction (realignment)

Despite proper stabilization of the head and cooperative efforts of subjects, head movement is inevitable during scanning (e.g. due to breathing, brain pulsation). Even the slightest movement in millimeter range can be a serious confound introducing artifacts in the voxel values. Therefore all images (scans) need to be realigned with respect to a reference image which is usually the first image in the time series. Realignment involves estimation and application of motion correction parameters. Each image is matched to the reference image using spatial transformations, namely affine "rigid body" transformation. Affine refers to a geometric transformation that maps the coordinates of each point in a scan onto the coordinates of another space (reference scan). In three dimensions, an affine rigid body transformation requires six parameters, three translations (head shifts along x, y, and z axes) and three rotations (head rotations about x, y, and z axes) in order to take head motion in all directions in to account. Parameter estimation is an iterative process. At each iteration, the fit between the data and the reference image is evaluated and modified if necessary. Once the difference between the data and the reference image is minimized to a certain convergence criterion, iterations stop. The estimated transformation parameters are then applied to every image. This process is called "resampling", and refers to determining for each voxel in the transformed images the corresponding signal intensity in the reference image using interpolation methods (trilinear, sinc, or spline interpolation)[39].

After realignment, there is still some residual variance remaining in the data due to nonlinear effects that are not accounted for by linear affine transformation. On of them is the movement-by-field distortion interaction. An object moving in the magnetic field distorts the field introducing magnetic field inhomogeneities. Linear affine rigid body transformation does not take this dynamic process into account. "Unwarping" is a process that estimates the field inhomogeneities with respect to object position based on the data and realignment parameters, and improves the sensitivity by reducing the residual variance [40].

Spatial Normalization

The majority of fMRI studies involve multiple subjects in order to make population level inferences about the cognitive function of interest. The signal collected from each subject is averaged, and the resulting image is a generic representation of the functional neuroanatomy underlying that cognitive function over individuals. Because of the differences in each individual's brain anatomy it is important to bring their images into the same standard anatomical space in order to generalize the interpretation

of the results [41]. The frequently used anatomical templates are Talaraich & Tournoux (1988)[42], MNI305 (Montreal Neurological Institute, average of 305 normal MRI scans obtained from all right handed 239 males and 66 females, mean age 23.4 +/- 4.1) or ICBM152 (International Consortium for Brain Mapping, average of 152 normal MRI scans that have been matched to the MNI305 template). Images are warped onto the template using 12-parameter affine transformation and resampled as described above.

Smoothing

Data are convolved with a smoothing kernel (filter), such as a Gaussian shaped kernel that has a width of several millimeters (FWHM-full width at half maximum, meaning the width of the kernel at half maximum of the height of the Gaussian curve). Smoothing is a process by which each data point in the image is replaced with a weighted average of itself and its neighbors[43]. Smoothing provides several advantages for the analysis of imaging data. First, it increases the signal to noise ratio. The hemodynamic response to an experimental manipulation extends over several millimeters in space, whereas noise has usually higher spatial frequencies. Smoothing recovers the signal from the noisy data. Secondly, for statistical inference about the data the theory of Gaussian Random Fields (GRF) is used (see below). Imposing a Gaussian shape to the data makes it to conform to the GRF model [43]. Finally, no two individuals have the same anatomy at a microscopic scale. The spatial scale of several millimeters provided by smoothing takes the anatomical variations among individuals into account and makes intersubject averaging meaningful.

Statistical Analysis

As mentioned, voxels are the three-dimensional image resolution elements in fMRI. Statistical analyses are performed at every single voxel using standard univariate tests. The resulting statistics are assembled into an image which is called a "statistical parametric map" (SPM). SPMs are spatially extended statistical image processes with voxel values that have approximately a Gaussian distribution. Hypotheses about regionally specific effects are tested based on SPMs[44].

Statistical analysis of imaging data involves two main components:

1. Modeling the data in order to break it down into components of interest, confounds, and error,
2. Making statistical inferences about effects using t or F tests[45].

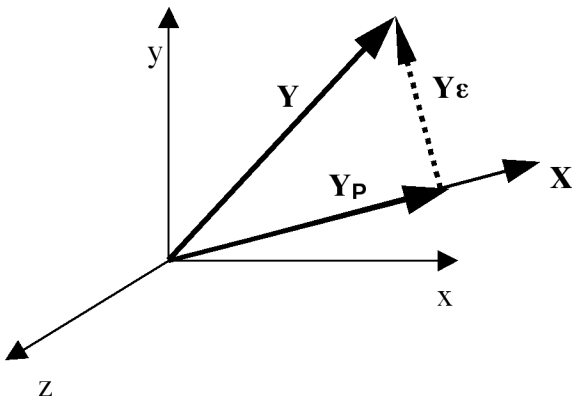
Data modeling employs the general linear model (GLM). GLM is also known as "analysis of (co)variance" or "multiple regression" in statistics[44] . The GLM equation is the following:

$$y = X\beta + \varepsilon$$

- y is the measured hemodynamic response (BOLD signal)
- X is the design matrix that is specified by the experimenter based on the effects built into the experiment. Each column in the design matrix contains some effect referred to as explanatory variable, covariate or regressor.
- β corresponds to parameter estimates. The relative contribution of each column in the design matrix to the measured response is determined by these parameter estimates. In other words, parameter estimates are measures of how well the model fits the actual response. Parameters are estimated using standard least squares.
- ε represents the error term. Ideally, it is an independently and identically distributed variable.

GLM can be represented geometrically using vectors in space:

The Y vector represents the measured data, and X is the expected response defined by the design matrix. The idea is to find the closest fit between the measured and expected



data. Geometrically this refers to the minimal distance between Y and X which is the orthogonal projection of Y onto X, namely the Y_P vector. The "least square" method estimates the parameters (ß) for this projection. Y corresponds to the residual error between Y and Y_P[44].

Statistical inferences about the parameters are made using

t or F statistics. T tests are used when one is interested in planned comparisons between particular parameters, whereas F tests can be thought of as a collection of T tests when one is looking at all comparisons jointly.[46]

The parameter estimates in each column of the design matrix are multiplied by contrast weight vectors (e.g. 1, 0, or –1, depending on which effect is of interest or which effects are being compared) that give a weighted sum of parameter estimates referred to as a "contrast"[47]. To obtain the t value, the contrast of the parameter estimates is divided by the standard error of that contrast (standard error is the square root of variance estimates of the parameters).

The general rules of hypothesis testing in statistics also apply to the interpretation of fMRI data. The null hypothesis, H_0, states that the treatment (independent variable) has no effect on the dependent variable for the population. The alternative hypothesis, H_1, predicts the independent variable will have an effect on the dependent variable for the population. It is possible to reject H_0 when in reality the treatment has no effect. This is called a Type I error or false positive. One needs to set criteria to reject H_0 and accept H_1 vice versa. The level of significance, also called alpha (or p value), defines the probability of false positives. Alpha is often 0.05 (5%). This is straightforward process for a single t test. However, statistical analysis of imaging data entails many t tests performed on many voxels. If one does not have an *a priori* hypothesis as to where exactly in the brain the effect will occur, then the whole volume of observed data has to be considered, and some correction for false positives associated with multiple tests has to be performed across the whole volume. Bonferroni correction is one standard way which assumes that the observations and tests performed on them are independent and discrete. In this case, the p value threshold can be adjusted by dividing alpha by the number of tests. However, fMRI data are not independent. They are continuous and spatially correlated (due to smoothing)[43]. This suggests that the Bonferroni correction for imaging data would be too strict and less powerful in detecting effects, because the actual number of independent tests is smaller than assumed in Bonferroni correction (the denominator would be smaller). The theory of Gaussian Random Fields (GRF) is a method to determine the number of independent observations and to adjust the p value in spatially correlated, continuous data sets[43]. GRF theory deals with the behavior of random, probabilistic processes defined over a space of any dimensions (in fMRI, 3D). The application of this theory requires the data to be sufficiently smoothed (with a kernel that has a FWHM 2-3 times the voxel size) so that the error fields are "a reasonable lattice approximation to the underlying GRF"[43], and that one has a well-specified statistical model with enough subjects so that the error distribution is normal (Gaussian).

Modeling

The analysis of functional neuroimaging data produces activation maps that show the brain areas engaged in the cognitive task at hand. Though exciting, this is not the end of the story. Mental processes emerge from the interaction among widely distributed neural systems ("neural networks"). It is valuable to examine the functional networks underlying mental processes and to produce "interaction maps" when interpreting the imaging data. Different mental operations may engage very similar brain regions. However, the interaction between these regions may be quite different. In other words, the "neural context" may differ across operations [48, 49]. Modeling of imaging data is a developing field that addresses the behavior of neural networks. Neural network models examine the "effective connectivity" between brain regions. They combine mathematical models with neuroanatomy. Anatomical connectivity can be derived from primate neuroanatomy or from developing diffusion tensor imaging (DTI) techniques that trace the white matter tracts in humans. [50, 51] Here we introduce two types of modeling that have different computational features.

Structural Equation Modeling (SEM)

SEM examines the correlations (covariances) in the activity between a set of brain regions involved in a task with respect to the anatomical constraints. The activity correlations are estimated that best predict the observed variance-covariance structure of the experimental data and used to determine the connection strengths (path coefficients) among these pre-specified regions. SEM minimizes the difference between the observed variance-covariance structure of the data and the one that is implied by the model by adjusting the path coefficients[52].

SEM has been widely used in neuroimaging. In an fMRI study, Buechel and Friston (1997) showed that attention to visual motion increased the effective connectivity between the primary visual cortex (V1) and motion processing area V5, and also between V5 and posterior parietal cortex (part of the attention network)[53]. In a PET study

investigating the role of medial temporal lobe (MTL) in a spatial and nonspatial object retrieval task Kohler *et al.* (1998) found that MTL showed differential interaction with dorsal (parietal) and ventral (occipito-temporal) visual streams. MTL interaction with the dorsal regions was positive in the spatial retrieval task, but negative in the object retrival task. MTL interaction with the ventral regions showed the reverse pattern.[54] Rowe *et al* (2002) investigated attention to action in Parkinson's disease (PD) using fMRI. Subjects executed an overlearned motor sequence task with and without attentional demands. In control subjects, attention to action increased the effective connectivity between prefrontal cortex and both the lateral premotor cortex and the supplementary motor area. This pattern was not observed in PD subjects suggesting a context-specific functional disconnection between those regions[55].

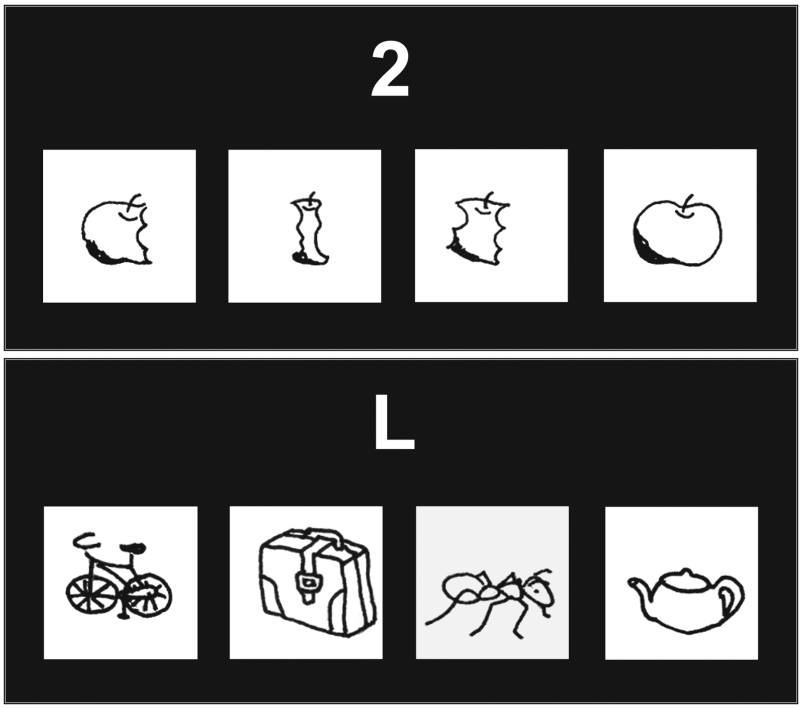


Figure 9. Picture sequencing (PS) and object discrimination control (CON) tasks.

Dynamic Causal Modeling (DCM)

The main difference between SEM and DCM is that SEM assumes the observed data emerge from intrinsic qualities of the neural network, without making a specific reference to the experimental variables, whereas DCM assumes the data are driven by the experimental variables (hence the term "causal") and examines how coupling between brain regions is affected by changes in experimental context. In addition, DCM treats the interactions among brain regions as dynamic processes that evolve over time, whereas SEM treats them as instantaneous[56]. DCMs are "dynamic input-state-output models". Inputs correspond to the designed changes in the experimental context (i.e., the specified design matrix is the input of the model). Outputs are the observed hemodynamic responses. States correspond to the neuronal activity in different brain regions. State variables are composed of several neurophysiological and biophysical variables (e.g. vasodilatory signal, normalized flow, normalized venous volume, normalized deoxy-Hb content, rate of signal decay, resting O2 extraction fraction etc). The objective of the model is to figure out how different brain states are effectively connected to each other within a given experimental context based on known deterministic input and measured output[57]. As opposed to SEM, DCM models coupling between brain regions at the neuronal rather than the hemodynamic level. This approach gives a more realistic picture of the neural network because neuronal interactions may not always be reflected in the hemodynamic response.[56]

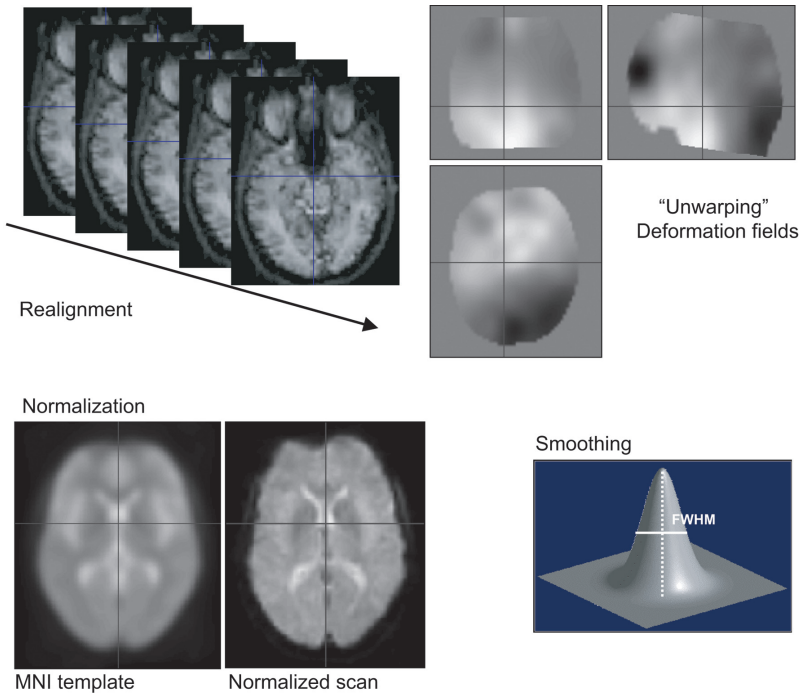


Figure 10. Preprocessing of fMRI time series.

The application of DCM to neuroimaging data has been

fruitful. In an attempt to investigate the changes in effective connectivity in response to attentional modulation, Friston *et al* (2003) used DCM on the fMRI data of an "attention to motion" study. Based on their previous modeling work and the fMRI data they entered the areas V1, V5, superior parietal cortex (SPC), and inferior frontal gyrus (IFG) as nodes of the network into the DCM. The inputs were entered in a hierarchical fashion. The sensory input was simply the visual stimulation entered at V1 only. The first contextual input was visual motion entered as a reciprocal modulation of the V1-V5 connectivity. The second contextual input was attentional set, modeled as a top-down modulation of connections between higher-lower cortical areas. The model demonstrated that visual motion modulates the reciprocal connections between V1 and V5, and that attention to visual motion modulates the reciprocal connections between IFG-SPC and SPC-V5[57].

Figure 11. Design matrix, regressors and fitted hemodynamic response.

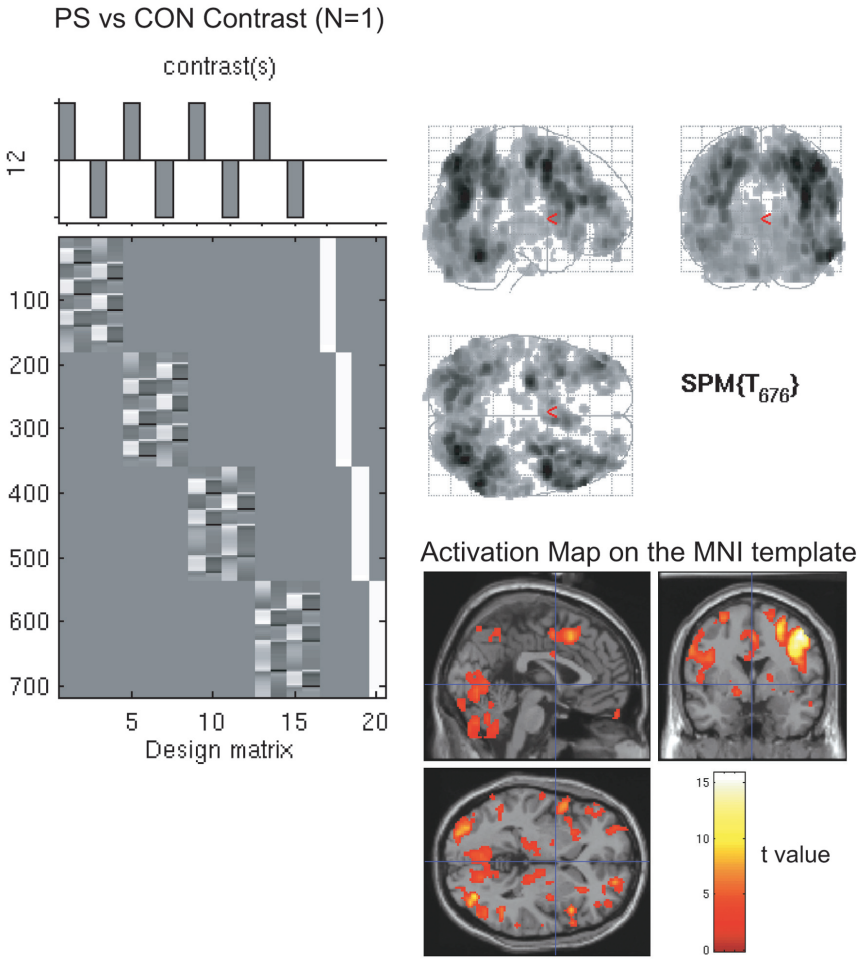


Figure 12. Contrast specification and statistical parametric map rendered on the cortical surface of MNI template.

Mechelli *et al* (2003) investigated how area V3 and parietal cortex may influence the category-responsive occipito-temporal regions. Objects are represented in a distributed occipito-temporal network. It has been shown that within this network there are subregions that respond preferentially to certain categories of objects (e.g. living-nonliving, tools, faces etc.). Using DCM, they showed that the presentation of houses, faces, and chairs affected the effective connectivity between V3 and occipito-temporal regions, but not between the occipito-tempral regions and the parietal areas. This finding suggests that the category effects are mediated by bottom-up processes, namely by inputs from the early visual cortex to the occipito-temporal areas[58].

Example: An fMRI Study of Parkinson's Disease

To demonstrate the application of some of the analysis methods discussed above we present an overview of the methods

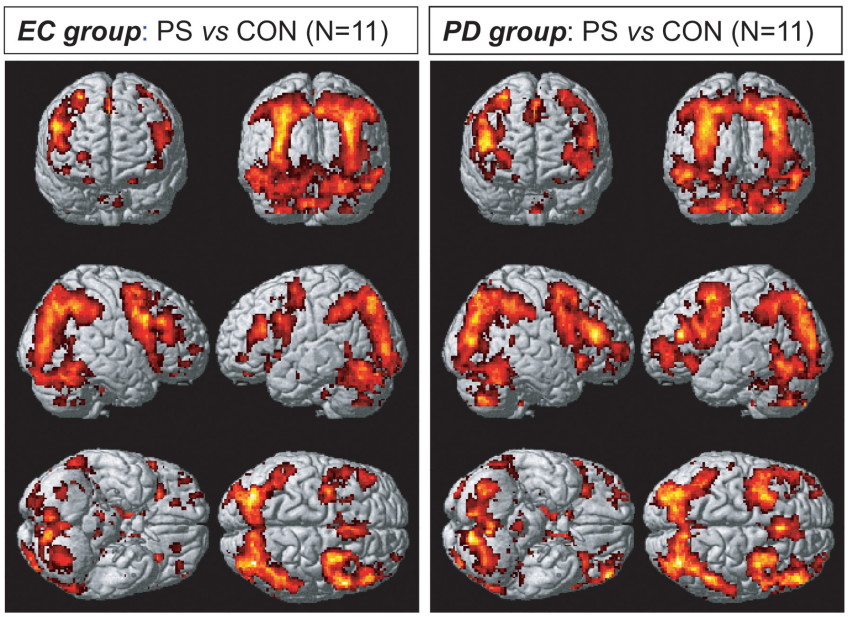


Figure 13. Group averaged statistical parametric maps in patients with Parkinson's disease and in healthy elderly control subjects.

and preliminary data of a recent study in our laboratory that investigates semantic event sequencing in Parkinson's disease[59]. We developed a picture sequencing task (PS) for use in the scanner. Subjects ordered a series of four pictures that were presented in a scrambled order. An object discrimination control task (CON) required subjects to find the living item among a set of four objects (Figure 9). The CON task served as a control for the visuospatial, semantic, and motor components of the PS task, and did not include a semantic sequencing component. 11 subjects with Parkinson's disease (PD) (with an average Hoehn & Yahr score of 2, mean age 57, mean education 16 years), and 11 age- and education matched healthy control subjects (HC) participated in the study. PD subjects were scanned when they were on dopaminergic medication ("ON" state).

Statistical analysis was performed using the SPM2 software package (Wellcome Dept. of Neurology). All images were realigned and normalized to the MNI template, smoothed using a 4 mm3 Gaussian kernel (Figure 10). Statistical analyses employed the general linear model. Design matrices were modeled in scans convolved with a canonical hemodynamic response function with time derivative (Figure 11). Sequencing related activation was assessed in linear contrasts of PS relative to CON blocks (Figure 12). Group averaged, statistical parametric maps (SPMs) were corrected across the whole brain for multiple voxel-wise comparisons. Behavioral measures including reaction time and accuracy did not differ significantly between the PD and HC group.

Figure 13 shows the group activation maps of the PS versus CON contrast rendered on the cortical surface of the MNI canonical brain. Bilateral occipito-temporal, parietal and frontal areas showed significant activation in both groups. However, a two sample t test comparing the two groups revealed that activation was more significant and widespread in the PD group.

In a recent study with young healthy subjects we showed that the dorsolateral prefrontal cortex (DLPFC) and globus pallidus interna (GPi) (an output nucleus of basal ganglia) were specifically involved in semantic event sequencing. We proposed that DLPFC may be important in manipulating the visual stimuli according to the temporal relationship among them, and GPi may be important in selectively updating the neural representations provided by the posterior cortical areas and enhancing the hierarchical selection of those representations by DLPFC to guide semantic event sequencing[60].

We did not observe GPi activation in the PS versus CON contrast in PDs, whereas GPi was active bilaterally in the same contrast in HCs. A simple correlation analysis demonstrated that DLPFC and GPi were functionally connected in the HC group during the PS task, but not in the PD group. In conclusion, there seems to be a functional disconnection between GPi and DLPFC in PDs in the PS task. Regarding the unimpaired behavioral performance in the PS task, PD patients seem to have compensatory changes in the neuronal and/or cognitive architecture.

Conclusion

FMRI can be used in investigating the functional neuroanatomical correlates of a vast array of cognitive functions. Activation maps with high spatial resolution combined with neural network models provide valuable tools for examining the interaction of different brain regions in cognition.

However, the temporal resolution of fMRI is limited to few seconds. On the other hand, imaging techniques including electroencephalography (EEG) and magnetoencephalography (MEG) that measure the neural activity directly have a high

temporal resolution to the order of milliseconds, but poor spatial resolution. Combining of fMRI and MEG data at the analysis stage will provide dynamic statistical parametric maps with a high spatiotemporal resolution. The spread of cortical activation in a cognitive task can be visualized as movies of brain activation[61].

An interdisciplinary approach to fMRI including developments in fMRI techniques, combination of fMRI with different imaging modalities and evaluation of imaging data in a neural network context using modeling will lead us to a deeper understanding of brain dynamics in cognition and behavior.

REFERENCES

1. Raichle ME. A brief history of human functional brain mapping; Brain Mapping: Systems (A.W. Toga & J.C. Mazziotta, edsAcademic Press: San Diego, California2001: 33-75.
2. Belliveau JW, Kennedy DN, Jr., McKinstry RC, Buchbinder BR, Weisskoff RM, Cohen MS, et al. Functional mapping of the human visual cortex by magnetic resonance imaging. Science 1991;254(5032):716-719.
3. Cohen MS. Rapid MRI and functional applications; Brain Mapping: Methods (A.W. Toga & J.C. Mazziotta, eds) Academic Press: San Diego, California 1996: 223-255.
4. Ogawa S, Lee TM, Kay AR, Tank DW. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci U S A 1990;87(24):9868-9872.
5. Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, et al. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci U S A 1992;89(12):5675-5679.
6. Buxton RB. Nuclear magnetic resonance; Introduction to functional magnetic resonance imaging: Principles and techniques. New York, Cambridge University Press 2002: 64-85.
7. Buxton RB. Relaxation and contrast in MRI; Introduction to functional magnetic resonance imaging: Principles and techniques. New York, Cambridge University Press 2002: 155-183.
8. Buxton RB. Magnetic resonance imaging; Introduction to functional magnetic resonance imaging: Principles and techniques. New York, Cambridge University Press 2002: 86-103.
9. Buxton RB. Mapping the MR signal; Introduction to functional magnetic resonance imaging: Principles and techniques. New York, Cambridge University Press 2002: 218-248.
10. Fox PT, Raichle ME. Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. Proc Natl Acad Sci U S A. 1986 Feb;83(4): 1140-4.
11. Weisskoff RM, Kiihne S. MRI susceptometry: image-based measurement of absolute susceptibility of MR contrast agents and human blood. Magn Reson Med. 1992 Apr;24(2):375-83.
12. Davis TL, Kwong KK, Weisskoff RM, Rosen BR. Calibrated functional MRI: mapping the dynamics of oxidative metabolism. Proc Natl Acad Sci U S A. 1998 Feb 17;95(4):1834-9.
13. Buckner RL, Bandettini PA, O'Craven KM, Savoy RL, Petersen SE, Raichle ME, Rosen BR. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. Proc Natl Acad Sci U S A. 1996 Dec 10;93(25):14878-83.
14. Dale AM, Buckner RL. Selective averaging of rapidly presented individual trials using fMRI. Hum Brain Mapp 1997;5:329-340.
15. Buxton RB, Frank LR. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. J Cereb Blood Flow Metab 1997;17(1):64-72.
16. Buxton RB, Wong EC, Frank LR. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn Reson Med 1998;39(6):855-864.
17. Mandeville JB, Marota JJ, Ayata C, Zaharchuk G, Moskowitz MA, Rosen BR, et al. Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. J Cereb Blood Flow Metab 1999;19(6):679-689.
18. Powers WJ, Hirsch IB, Cryer PE. Effect of stepped hypoglycemia on regional cerebral blood flow response to physiological brain activation. Am J Physiol. 1996
19. Fox PT, Raichle ME, Mintun MA, Dence C. Nonoxidative glucose consumption during focal physiologic neural activity. Feb;270(2 Pt 2):H554-9. Science. 1988 Jul 22;241(4864):462-4.
20. Prichard J, Rothman D, Novotny E, Petroff O, Kuwabara T, Avison M, Howseman A, Hanstock C, Shulman R. Lactate rise detected by 1H NMR in human visual cortex during physiologic stimulation. Proc Natl Acad Sci U S A. 1991 Jul 1;88(13):5829-31.
21. Sappey-Marinier D, Calabrese G, Fein G, Hugg JW, Biggins C, Weiner MW.
    Effect of photic stimulation on human visual cortex lactate and phosphates using 1H and 31P magnetic resonance spectroscopy. J Cereb Blood Flow Metab. 1992 Jul;12(4):584-92.
22. Magistretti PJ, Pellerin L. Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging. Philos Trans R Soc Lond B Biol Sci. 1999 Jul 29;354(1387):1155-63.
23. Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. Neurophysiological investigation of the basis of the fMRI signal. Nature 2001;412(6843):150-157.
24. Maguire EA, Frith CD. Lateral asymmetry in the hippocampal response to the remoteness of autobiographical memories. J Neurosci 2003;23(12):5302-5307.
25. Henson R, Shallice T, Dolan R. Neuroimaging evidence for dissociable forms of repetition priming. Science 2000;287(5456):1269-1272.
26. Thiel CM, Henson RN, Dolan RJ. Scopolamine but not lorazepam modulates face repetition priming: a psychopharmacological fMRI study. Neuropsychopharmacology. 2002;27(2):282-92.
27. Friston KJ, Price CJ, Fletcher P, Moore C, Frackowiak RS, Dolan RJ. The trouble with cognitive subtraction. Neuroimage. 1996;4(2):97-104.
28. Price CJ, Friston KJ. Cognitive conjunction: a new approach to brain activation experiments. Neuroimage. 1997;5(4 Pt 1):261-70.
29. Rosen BR, Buckner RL, Dale AM. Event-related functional MRI: past, present, and future. Proc Natl Acad Sci U S A 1998;95(3):773-780.
30. Dale AM. Optimal experimental design for event-related fMRI. Hum Brain Mapp 1999;8(2-3):109-114.
31. Strayer DL, Kramer AF. Strategies and Automaticity: I. Basic Findings and Conceptual Framework. J Exp Psychol Learn Mem Cogn 1994;20(2):318-341.
32. Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale AM, Rosen BR, Buckner RL. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. Science. 1998;281(5380):1188-91.
33. Schon K, Hasselmo ME, LoPresti ML, Tricarico MD, Stern CE. Persistence of parahippocampal representation in the absence of stimulus input enhances long-term encoding: A functional magnetic

resonance imaging study of subsequent memory after a delayed match-to-sample task. (2004). J Neurosci. 2004:24(50):11088-11097.
34. Clark VP, Fannon S, Lai S, Benson R, Bauer L. Responses to rare visual target and distractor stimuli using event-related fMRI. J Neurophysiol. 2000;83(5):3133-9.
35. Kleinschmidt A, Buchel C, Zeki S, Frackowiak RS. Human brain activity during spontaneously reversing perception of ambiguous figures. Proc R Soc Lond B Biol Sci. 1998;265(1413):2427-33.
36. Schendan HE, Searl MM, Melrose RJ, Stern CE. An FMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. Neuron. 2003;37(6):1013-25.
37. Visscher KM, Miezin FM, Kelly JE, Buckner RL, Donaldson DI, McAvoy MP, Bhalodia VM, Petersen SE. Mixed blocked/event-related designs separate transient and sustained activity in fMRI. Neuroimage. 2003;19(4):1694-708.
38. Donaldson DI, Petersen SE, Ollinger JM, Buckner RL. Dissociating state and item components of recognition memory using fMRI. Neuroimage. 2001;13(1):129-42.
39. Ashburner J, Friston K. Rigid body registration, Human Brain Function 2nd Edition (J Ashburner, K Friston, W Penny, eds) The Wellcome Dept. of Cognitive Neurology,University College London. http://www.fil.ion.ucl.ac.uk/spm
40. Andersson JL, Hutton C, Ashburner J, Turner R, Friston K. Modeling geometric deformations in EPI time series. Neuroimage. 2001;13(5):903-19.
41. Ashburner J, Friston K. Spatial normalization using basis functions; Human Brain Function 2nd Edition (J Ashburner, K Friston, W Penny, eds) The Wellcome Dept. of Cognitive Neurology,University College London.
42. Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain 3-D proportional system: An approach to cerebral imaging. Thieme: Stuttgart, Germany. 1988.
43. Brett M, Penny W, Kiebel S. Introduction to Random Field Theory; Human Brain Function 2nd Edition (J Ashburner, K Friston, W Penny, eds) The Wellcome Dept. of Cognitive Neurology,University College London. http://www.fil.ion.ucl.ac.uk/spm
44. Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: A general linear approach. Hum Brain Mapp. 1995;2:189-210.
45. Friston KJ. Models of brain function in neuroimaging. Annu Rev Psychol. 2005;56:1-31.
46. Friston K, Penny W. Classical and bayesian inference: Human Brain Function 2nd Edition (J Ashburner, K Friston, W Penny, eds) The Wellcome Dept. of Cognitive Neurology,University College London.
47. Poline JB, Kherif F, Penny W. Contrasts and classical inference: Human Brain Function 2nd Edition (J Ashburner, K Friston, W Penny, eds) The Wellcome Dept. of Cognitive Neurology,University College London.
48. McIntosh AR. Understanding neural interactions in learning and memory using functional Neuroimaging. Ann NY Acad Sci. 1998;30(855):556-571.
49. Horwitz B, Tagamets MA, McIntosh AR. Neural modeling, functional brain imaging, and cognition. Trends Cogn Sci. 1999;(3):91-98.
50. Mori S, Barker PB. Diffusion magnetic resonance imaging: its principle and applications. Anat Rec. 1999;257(3):102-9.
51. Kubicki M, Westin CF, Maier SE, Mamata H, Frumin M, Ersner-Hershfield H, Kikinis R, Jolesz FA, McCarley R, Shenton ME. Diffusion tensor imaging and its application to neuropsychiatric disorders. Harv Rev Psychiatry. 2002;10(6):324-36. Review.
52. Buchel C, Friston K. Assessing interactions among neuronal systems using functional neuroimaging. Neural Netw. 2000;13(8-9):871-82.
53. Buchel C, Friston KJ. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb Cortex. 1997;(8):768-78.
54. Kohler S, McIntosh AR, Moscovitch M, Winocur G. Functional interactions between the medial temporal lobes and posterior neocortex related to episodic memory retrieval. Cereb Cortex. 1998;8(5):451-61.
55. Rowe J, Stephan KE, Friston K, Frackowiak R, Lees A, Passingham R. Attention to action in Parkinson's disease: impaired effective connectivity among frontal cortical regions. Brain. 2002;125(Pt 2):276-89.
56. Penny WD, Stephan KE, Mechelli A, Friston KJ. Modelling functional integration: a comparison of structural equation and dynamic causal models. Neuroimage. 2004;23 Suppl 1:S264-74.
57. Friston KJ, Harrison L, Penny W. Dynamic causal modelling. Neuroimage 2003;19(4):1273-1302.
58. Mechelli A, Price CJ, Noppeney U, Friston KJ. A dynamic causal modeling study on category effects: bottom-up or top-down mediation? J Cogn Neurosci. 2003;15(7):925-34.
59. Tinaz S, Stern CE. FMRI study of cognitive event sequencing in Parkinson's disease. Poster presentation at the 34th Annual Meeting of Society for Neuroscience, Oct 2004, San Diego.
60. Tinaz S, Schon K, Stern CE. Neuroanatomical correlates of picture and motor sequencing: An fMRI study. Poster presentation at the 33rd Annual Meeting of Society for Neuroscience, Nov 2003, New Orleans.
61. Dale AM, Liu AK, Fischl BR, Buckner RL, Belliveau JW, Lewine JD, Halgren E. Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron 2000;26:55-67.

Useful links:

1- SPM website:
Online bibliography provides free access to classic theoretical papers and book chapters related to several topics in fMRI reviewed in this paper, including the analysis of fMRI time series.
2- Analysis of Functional Neuroimages (AFNI) developed by Robert W. Cox: http://afni.nimh.nih.gov/afni
3- Freesurfer (FSfast) software developed by scientists at Mass General Hospital, Athinoula A. Martinos Center for Biomedical Imaging. http://surfer.nmr.mgh.harvard.edu
4- International Consortium for Brain Mapping: http://www.loni.ucla.edu/ICBM/