Functional MRI: Genesis, State of the art and the Sequel

Rose Dawn Bharath

Department of Neuroimaging and Interventional Radiology and Faculty In charge Advanced Brain Imaging Facility, Cognitive Neuroscience Centre, National Institute of Mental Health And Neurosciences, Bangalore, Karnataka, India

Correspondence: Dr. Rose Dawn Bharath, Department of Neuroimaging and Interventional Radiology, Ashwini Block, National Institute of Mental Health and Neurosciences, Hosur Road, Bangalore - 29, Karnataka, India.
E-mail: drrosedawn@yahoo.com

Introduction

The last 25 years have seen functional magnetic resonance imaging (fMRI) grow from an interesting experimental imaging technique in the hands of some to a primary investigation of choice in the localization and lateralization of brain function prior to surgery. Developments in the field of computational neurosciences have transformed fMRI analysis from classical subtractive type analysis to dynamic casual modeling, and now to graph theory analysis. This has widened the scope of fMRI, and is therefore finding applications in understanding neural correlates of diseases like autism and Alzheimer’s disease,[1,2] prognostication of diseases like traumatic brain injury,[3] and has the potential to direct therapy.[4,5] It is unfortunately true that this widened ambit has not received the clinical attention it deserves, probably because fMRI is susceptible to artifacts from skull base and blood products and has reduced sensitivity in patients with vascular malformations, or because a change in medical practice usually lags behind the technological and scientific developments that make it possible. This review focuses on the developmental chronology of fMRI image analysis in the last 25 years with highlights on major milestones like developments in the field of paradigms, analysis methods, resting state fMRI, and functional connectivity. To make the statistical images of brain at work more colorful, the article starts with genesis of fMRI and ends with the hope of a promising bright future. Many inputs for this article are obtained from a series of 103 review articles edited by Bandettini et al.,[6] compiling personal experiences of pioneers in this field. Interested readers are encouraged to refer to these for a more complete overview.

Genesis of BOLD: Stories

Segi Ogawa coined the term blood oxygen level dependent (BOLD) imaging for functional imaging though three groups were in close competition for the title: The MGH group with Kenneth Kwong as the lead author,[7] the University of Minnesota group with Segi Ogawa as the lead author,[8] and the University of Wisconsin group with Peter Bandettini as the lead author.[9]

BOLD imaging takes advantage of the magnetic differences between oxyhemoglobin and deoxyhemoglobin. Deoxyhemoglobin is highly paramagnetic and causes loss of signal on MRI and, thus, creates a natural contrast between highly oxygenated areas of the brain and less oxygenated areas. During task induced brain activation, there is localized increase in blood flow rich in oxyhemoglobin which increases the MR signal. It is thought that this localized increase in blood flow reflects neuronal activity since both are found to be temporally correlated. The change in the MR signal after the onset of a stimulus is called hemodynamic response function (HRF). HRF begins 1-2 s after a stimulus, rises to a peak within 3 s, and falls immediately (within 2 s) after the stimulus to a level below the baseline called the undershoot and recovers to the baseline overtime. Even with widespread use of this technology, the physiological mechanisms underlying the genesis of BOLD are not well understood. In the well-meaning attempt to understand the physiological basis of BOLD, a controversy hatched called the “coupling controversy” which is primarily a debate between the principles of fMRI and positron emission tomography (PET) models. The fMRI models measure the % change and are heavily dependent upon assumed parameters, including the coupling ratio. The PET models are too slow and complex to be physiologically reliable,
Despite the paucity of assumed parameters, in 1986, Fox et al., using PET radiotracers, found that after a stimulus, the oxygen extraction fraction (OEF) response “uncoupled” and fell by a highly significant 19%, whereas cerebral blood flow (CBF), cerebral blood volume, and cerebral oxygen metabolic rate (CMRO2) were coupled with the stimulus and showed significant increase. A 90-year-old, nearly universally accepted hypothesis of Roy and Sherrington that focal, stimulus-induced increases in brain blood flow are driven by local metabolic demand was contradicted by this finding, giving rise to the coupling controversy. Subsequent studies by Fox et al. in 1988 on cerebral glucose metabolic rate (CMR gluc) also revealed an “uncoupled” increase 10-fold greater than the CMRO2 increase. Based on these, they speculated that the disproportionate increase in CMR gluc (over CMRO2) might indicate a shift in metabolic pathway from glucose oxidation to glycolysis, which was supported by “enzymatic limitation” hypothesis. In an attempt to detangle the “uncoupling controversy,” in 1996, Malonek et al. employed intrinsic signal optical imaging and proposed that despite the evident vascular hyperoxia during task, brain tissue might experience an oxygen debt, at least transiently, due to oxygen diffusion limitation from the blood vessel to the active neurons. To make issues more complex, in 2010, Lin et al. explained ABOLD changes using astrocyte neuron shuttling model and proposed that “neurotransmitters, particularly glutamate, rather than energy use are probably the principal agents generating activity-induced blood flow...” Ultimately, the coupling controversy has greatly stimulated growth and has left us richer in knowledge to presume that task-induced increases in blood flow are multifactorial and complex and could be the end result of dynamic interactions induced by energy demand, regulated by neurotransmitters, and limited by enzymatic and diffusion functions.

**The state of the art**

**Paradigm design: Block, event, and mixed design**

A stimulus forms the basic component of a paradigm and are otherwise called trials. A stimulus can be of various types based on the function a researcher wishes to elicit. It can be of auditory, sensory, motor, visual, or cognitive type.

Block designs were the earliest designs to be used in fMRI and they mirrored the design of PET experiments. Block design experiments utilize blocks of either identical stimulus types to establish a task-specific condition, e.g. finger tapping task to visualize the motor cortex, or a mixture of trial types to establish a mixed task condition, e.g. picture naming task to locate the Broca’s area. Salient feature of this design is its efficiency in obtaining adequate signal-to-noise ratio in a single subject in the shortest possible time, making it the most popular design in preoperative functional mapping. Since this design collapses multiple stimuli, we are only able to assess the function of brain once in 15-30s by clubbing together multiple HRFs and making us relatively blind to what is happening in the brain at the beginning of events, across the event, and at the end of events. By adopting a subtraction comparison strategy, we are only seeing functioning areas which are higher than the baseline, leading to much criticism related to the neuropsychological drawbacks. Finally, in tasks where both positive and negative responses occur in a single block, the block design averages the two responses resulting in a canceling effect that does not represent the complexity and magnitude of brain function.

With the insight that the hemodynamic response function could be a marker of the underlying neuronal activity, there were attempts in increasing temporal resolution of fMRI analysis which resulted in event-related designs allowing us to see brain function once in 2-4 s. This initial approach to event-related designs was inherited from event related potential ERP research (trial averaging) in electrophysiology. An “event-related fMRI” involves separating the conditions of an experiment into discrete points in time, so that the associated brain responses with each element can be analyzed independently. The core idea of an event-related design is the separation of cognitive processes into discrete points in time (i.e., “events”) allowing differentiation of their associated fMRI signals. By modeling brain function as a series of transient changes, rather than as an ongoing state, event-related fMRI allowed researchers to create much more complex paradigms and more dynamic analysis methods. As the complexity of experimental designs increased, fMRI analyses became increasingly abstracted from the original data. The downside of the event-related design included a decrease of signal-to-noise leading to less power than block designs of similar timing. Further, while the event-related design created a more fine-grained characterization of the BOLD activity, this methodology still ignored certain signals, including transients at the block transitions and sustained activity that begins and ends with the performance of the task.

Mixed block/event-related designs allow for the simultaneous modeling of the transient, trial-related activity and the sustained, task-related BOLD activity. Transient BOLD activity which is seen prior to after a HRF is thought to reflect the neural response to processing of stimuli, computation of responses to stimuli, and many intermediate processes which are to be identified. The sustained signal is most often believed to represent a putative task maintenance signalsignal. Perhaps most probably related to a task set. For example, in a learning memory task, regions showing sustained effects during various processes of encoding, retrieval, and object naming are active in encoding-, retrieval-, and object naming-mode, respectively all three modes. This is not synonymous as with the regions and or neural activity...
associated with encoding, retrieving, or naming of a specific stimulus\(^{[33]}\) because of multitasking of brain regions and interlinking of neural processes. It is the sustained signal that we commonly see in the routine fMRI analysis whereas modelling of the transient BOLD events is probably better as this can reflect the neuronal processes associated with these specific tasks.

**General linear model and fMRI: Statistical parametric mapping analysis**

The statistical parametric mapping (SPM) software which incorporates the general linear model (GLM) has been at the heart of fMRI analyses for the past 25 years. The main reason for this is that SPM is (1) conceptually simple, (2) readily available in standard packages, (3) is an incredibly flexible tool, (4) implements the standard statistical testing framework, and (5) does not require heavy computation and can be implemented in only a few lines of MATLAB or Python code. SPM software does not really refer to a single piece of software, as many changes are made between each release to make life easier for users, to fix bugs, and to ensure compatibility with other software. Karl Friston first wrote SPM in 1991 which has been referred to as SPM 91 or classical SPM; the subsequent releases have been named SPM 94, 95, ..., 5, 8, etc., to coincide with the year of release. Standard processing steps include several preprocessing steps like image import, image registration, realignment, normalizing to a template image, smoothing, etc., and statistical inference steps like design, analysis, estimate, and results. The end result of an experiment is typically a set of statistic values (e.g., T or F values) that comprise an image. The last display results section in SPM deals with various rendering modes which make the statistics colorful and visually appealing.

The GLM is written as follows

\[ y = Xh + n, \]

where \( y \) is the observed fMRI time series, \( X \) is the design matrix that contains the stimulus timing information, \( h \) is the hemodynamic response function, and \( n \) represents additive noise.

An illustrative in-house example of SPM analysis in 53 healthy controls during encoding of working memory is demonstrated in Figure 1. This is a simple subtractive type of statistical analysis and represents only areas which had more blood flow during the task, in comparison to the rest. This is otherwise called activation type of analysis.

However, it is not sure whether the association of fMRI with GLM will last forever since GLM relies on assumptions that are difficult to check. First, the matrix \( X \) should contain the appropriate regressors. Too few or too many regressors will lead to either loss in sensitivity or specificity. Second, the normality assumption about the linearity of hemodynamic response may not be true, especially when it comes to cognition. Third, the assumptions on the variance–covariance structure of the noise are difficult to model.

One emerging area that looks to overcome some of the restrictions of SPM is the use of non-parametric, permutation testing approaches. Another emerging area in fMRI that could potentially do this is Bayesian statistics.\(^{[35]}\)

**Masterly inactivity: Advantages of resting state fMRI**

In the early 1990s, Dr. Raichle recognized a very important fact, i.e. even at rest, the brain uses about 20% of the energy of the entire body. The change in energy utilization concomitant to specific task performance is miniscule compared to the energy consumption of the “resting” brain.\(^{[36]}\) In 1995, Dr. Bharat Biswal demonstrated for the first time coherent brain spontaneous fluctuations using BOLD fMRI.\(^{[37]}\) Though the origin of these spontaneous fluctuations is still not completely known, multiple neuroimaging modalities have converged on the fact of its existence. Its practicality and ease of use has resulted in exponential increase in publications using this technique. The practical advantage of the resting state paradigm is that it does not require patients to perform cognitively challenging or experimentally controlled tasks. Brain
function can be measured with fMRI simply by asking patients to lie quietly in the scanner for about 10-20 min, or indeed by scanning patients who are unconscious. Hence, this technique is like any other sequence in structural imaging study adding to its popularity. Another interesting feature is that we are able to visualize the function of whole brain and not only one hemisphere. Many functions of the brain, especially cognitive functions like language, are interlinked and are not localized to one specific area like Broca’s or Wernicke’s. It is increasingly reported that human brain function might be having a small-world network topology,[58] and hence the importance of visualizing the whole brain. Resting state network analysis relies on finding a set of voxels that exhibit spontaneous BOLD fluctuations using a variety of exploratory and multivariate analysis algorithms, the most popular being spatial independent component analysis. The most commonly visualized networks are sensorimotor, default mode, visual, auditory, and language, and frontoparietal networks, and these are demonstrated in the Figure 2. The default mode network is particularly interesting because of its dynamic character and has been extensively studied in assessing brain maturation,[39,40] aging,[41,42] epilepsy,[43] and brain tumors.[44] Default mode network is dynamic in that it is positively correlated at rest and demonstrates task-induced negative correlations.

Cluster to networks: Advantages of functional connectivity analysis

The field of fMRI gradually made a fundamental shift from the activation clusters to the connectivity analysis around 2003, motivated by the idea that connectivity gets us closer to the actual mechanisms of brain function. This was as a result of changes in conceptual focus and methodological procedures with a shift from PET type of analysis to electroencephalography (EEG) and time series type of analysis. Whereas the activation paradigm emphasized the univariate (single-voxel or regional) response in amplitude to an exogenous stimulus, the connectivity paradigm emphasized the bivariate or multivariate (systems or network) covariance.[45] Application of already existing computer science knowledge of networks to fMRI using graph theory analysis paved the way for connectivity analysis. The connectivity could be either task-based functional connectivity or resting functional connectivity. Connectivity is popularly described as graphs which are data units that have nodes (C) and edges between nodes (L). The clustering coefficient (C) is a measure of local network connectivity. A network with a high average clustering coefficient is characterized by densely connected local clusters and has a high C value. The characteristic path length (L) is a measure of network edges. A network with a low characteristic path length is characterized by short distances between any two nodes. Small-world network is characterized by a high clustering coefficient and a low characteristic path length. Connectivity analysis is increasingly finding its place in understanding network correlates of diseases like epilepsy,[48,49]

An illustrative example of connectivity analysis can be seen in Figure 3 which demonstrates the connectivity of the posterior cingulate cortex (PCC) and rest of the brain during encoding of working memory in the same data set described in Figure 1. It should be noted that the entire PCC and negative connections were not seen in Figure 1 because of the subtractive type of design only looking at areas which are more active compared to rest of the brain.

**The Sequel**

The future of fMRI lies beyond functional labeling of different regions, encouraging us to devote our resources to characterise underlying neural processes. New analysis methods are already changing the way we analyze imaging data, by improving the sensitivity of fMRI to spatial and temporal patterns. Combining methods is also necessary for studying processes that span spatiotemporal scales. With technological advances, it is now possible to simultaneously combine EEG recording and Transcranial magnetic stimulation (TMS) with fMRI, making integration and modulation of electrical signal with hemodynamic processes a reality. MRI and PET systems have been recently combined in a single scanner, allowing hemodynamic characterization of neurotransmitter function. fMRI at high resolution and high field (7 T and more) promises greater sensitivity to fine-grained patterns[53] which is called pattern imaging fMRI (pi-fMRI).[52] pi-fMRI might have the power to reveal sub-voxel–scale columnar pattern information through small biases in the sample each voxel takes of the underlying neuronal pattern.[53] “Imaging genetics” is a research
approach in which genetic information and fMRI data in the same subject are combined to define neuromechanisms linked to genetic variation. The incorporation of new sources of biological information such as whole genome sequencing, proteomic, lipidomic, and expression profiles, and cellular models derived from induced pluripotent stem cells opens new vistas for imaging genetics in a translational enterprise and is ultimately hoped to improve and create therapeutic options for several diseases. There is increasing interest in implanting stimulating electrodes in subcortical and cortical structures to modulate brain dynamics with therapeutic impact on tremor in Parkinson’s disease and mood disorders. As the familiarity of image-guided minimally invasive procedures continues to advance and find more extensive clinical applications, it is likely that there will be parallel developments using fMRI in early diagnosis of treatment-responsive patients without waiting for the end stage, to plan procedures based on the anatomical coordinates of abnormally connected systems. Another possible growth point is the greater use of fMRI as a biomarker in patients with spinal cord injury, neurodegenerative diseases, etc., to identify patients who would benefit from possibly expensive or prolonged courses of treatment with disease-modifying drugs. Other examples include the potential to use fMRI as a measure of brain age or maturation of brain functional systems that could support educational, judicial, or medical interventions in children with learning or behavioral disorders. One looks forward to the future of functional neuroimaging, which promises to be proactive, increasingly creative, and interdisciplinary in its approach, accommodative of individual differences, quantitative, and more relevant than ever before to the understanding of complex human behaviors.

Conclusion

Applications of the developments in the field of computational sciences to functional image analysis has already transformed the way we look at brain function. This article gives a brief overview of some of the developments in the functional image analysis in the last twenty five years. Knowledge of these methods will enhance our preparedness, as we apply them into our routine clinical practice.

Acknowledgement

I acknowledge the Support from the Department of Science and Technology for the 3TMRI equipment which is a national facility to promote research in cognitive sciences. I am extremely grateful to Director Vice Chancellor NIMHANS, Dr P Satish Chandra and Registrar NIMHANS, Dr V Ravi and Dr A K Gupta Head Department of Neuroimaging and Interventional Radiology whose encouragement makes patient care, teaching and research most satisfying. I acknowledge the hard work done by Prof Shobini L Rao and her vision for this field. I acknowledge all my students especially Mr Rajanikanth Panda, Mrs Lija George, Thamodharan A and the radiographer staff in the department for their untiring work.

References

18. Brockway JP. Two functional magnetic resonance imaging f (MRI) tasks that may replace the gold standard, Wada testing, for language lateralization while giving additional localization information. Brain Cognition 2000;34:57-9.
Bharath: Functional MRI: Genesis, State of the art and the Sequel


Cite this article as: Bharath RD. Functional MRI: Genesis, State of the art and the Sequel. Indian J Radiol Imaging 2014;24:6-12.

Source of Support: The scanner in which the studies were conducted has been funded by the Department of Science and Technology, Conflict of Interest: None declared.