

# Methods and Approaches for Characterizing Learning Related Changes Observed in functional MRI Data — A Review

Raju S. Bapi <sup>\*</sup>, V. S. Chandrasekhar Pammi, K. P. Miyapuram

Department of Computer and Information Sciences  
University of Hyderabad  
Gachibowli, Hyderabad – 500 046, India  
Email: <sup>\*</sup> bapics@uohyd.ernet.in

## ABSTRACT

Brain imaging data have so far revealed a wealth of information about neuronal circuits involved in higher mental functions like memory, attention, emotion, language etc. Our efforts are toward understanding the learning related effects in brain activity during the acquisition of visuo-motor sequential skills. The aim of this paper is to survey various methods and approaches of analysis that allow the characterization of learning related changes in fMRI data. Traditional imaging analysis using the Statistical Parametric Map (SPM) approach averages out temporal changes and presents overall differences between different stages of learning. We outline other potential approaches for revealing learning effects such as statistical time series analysis, modelling of haemodynamic response function and independent component analysis. We present example case studies from our visuo-motor sequence learning experiments to describe application of SPM and statistical time series analyses. Our review highlights that the problem of characterizing learning induced changes in fMRI data remains an interesting and challenging open research problem.

**KEYWORDS:** General Linear Model, Learning, SPM, modelling HRF, time series analysis, ICA

## 1. INTRODUCTION

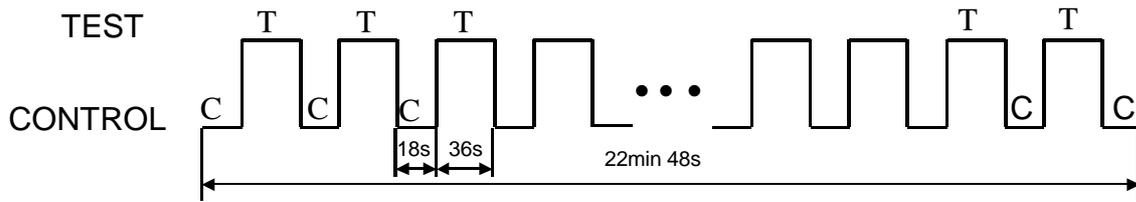
The functional Magnetic Resonance Imaging (fMRI) is a powerful imaging tool that can be used to perform brain activation studies non-invasively *in vivo* while subjects are engaged in meaningful behavioural tasks. The resulting activation of the brain indirectly depends on blood-oxygen-level-dependent (BOLD) signal (Ogawa et al., 1990). fMRI is well suited to measure the dynamic changes in brain activity induced by tasks that involve

learning as it provides a reasonable spatial and temporal resolution (~1–1.5 mm; 3–5 seconds, respectively) compared to other neuroimaging techniques such as positron emission tomography (Volkow et al., 1997). After ten years of fMRI research, there is still much to learn about how neuronal activity, haemodynamics and fMRI signals are interrelated (Heeger and Ress, 2002). A recent review of Ugurbil et al. (2003) suggested the possibility of obtaining spatially accurate and quantitative data on brain function from magnetic resonance technologies.

It has long been established that the brain undergoes learning related changes – both short and long term (for example, Karni et al., 1995; Kaas, 1995). Although "learning" is an important cognitive function that the brain is constantly engaged in, most of the current fMRI studies do not explicitly investigate this phenomenon. A possible reason for not undertaking tasks involving learning could be that the image analysis methodology for unraveling time dependent changes is still rudimentary in the fMRI modality. We illustrate possible approaches for characterizing learning related fMRI activity by taking case studies from our experiments in motor skill learning where subjects learned finger movement sequences. The aim of these experiments is to unravel the neural bases for learning visuomotor sequences (Bapi et al., 2000b; Pammi et al., 2003). The goal of the current paper is to review possible methods and approaches for characterizing learning related behaviour from fMRI data.

## **2. MATERIALS AND METHODS**

We utilized an on-off (boxcar) design in which the subject performed alternating control and test conditions inside the fMRI scanner. Visual stimuli consisting of illuminated squares on a 3x3 grid are projected in front of the subject and the subject learned to press the corresponding keys on a 3x3 keypad placed near the right hand. Subjects performed two tasks — 2x6 and 1x12 tasks — adopted from a modified version of 2x5 visuo-motor sequence learning paradigm (Hikosaka et al., 1995; Bapi et al., 2000a). In 2x6 task, subjects learned, by trial and error, to press two keys in response to lighted squares on the display. In 1x12 task, explicit guidance was given by visual cues for executing one key-press at a time. Eight right-handed normal human subjects (6 male and 2 females) with average age of 24.7 years (range 23 to 28 years) participated in this study. Ethics committee of the Brain Activity Imaging Center (ATR, Kyoto, Japan) approved the experimental protocol.



**Figure 1.** fMRI box-car design.

A series of whole brain fMRI scans (functional images) are acquired continuously while the subject performed the task inside the scanner. Functional images were acquired in a 1.5 tesla whole-body scanner. The experiment consisted of four sessions and a session comprised 13 epochs of alternating control (C) and test (T) conditions, each lasting for 18 and 36 seconds, respectively (see Fig. 1). During the experiment, a total of 228 scans were obtained with an inter-scan interval of 6 seconds. Each scan covered the entire brain and consisted of 50 axial continuous slices with 3mm thickness. In addition, a high-resolution anatomical brain image of the subject with slice-thickness of 1 mm was collected.

### **3. APPROACHES FOR CHARACTERIZING LEARNING INDUCED CHANGES**

#### **3.1. Statistical Parametric Map (SPM)**

A popular method for fMRI data analysis is called 'Statistical Parametric Mapping' (SPM). SPM is a method of data reduction and condensing of information (in a statistically meaningful way) from a number of individual brain scans into a single image volume that can be more easily viewed and interpreted. A package by the same name is developed by the Wellcome Department of Cognitive Neurology, London. Most fMRI experiments attempt to infer some general property of brain activation from a particular sample of individuals. Conventionally, this requires a pooled analysis corresponding to a group of subjects. Thus, the images need to be in a standardized space prior to statistical analysis. The preprocessing of fMRI images usually proceeds by setting a common origin for all images and correcting for any head movements during the scanning period. fMRI images are then normalized (Friston et al., 1995a) using non-linear warping techniques. A high-resolution anatomical image for each subject is additionally acquired in order to ensure correct anatomical localization of brain activation, which is used to determine the parameters for normalizing the functional images. The matching

of the brains in the normalization step is only possible on a coarse scale, since there is not necessarily a one-to-one mapping of the cortical structures between different brains. Hence, the normalized functional images are spatially smoothed with a gaussian filter. The smoothing process also validates the underlying gaussian assumption for the BOLD activity that is in turn used in statistical inference. Statistical analysis is performed on the preprocessed image data using the general linear model (Friston et al., 1995b). The general linear model (GLM) is used to specify the covariates of interest, such as the experimental design and the nature of hypothesis testing to be implemented in the form of a design matrix. Brain activity specific to task is obtained by specifying linear contrasts. The end result is a statistical parametric map (SPM) that can be used to display significant activations overlaid on a high-resolution anatomical image.

The main idea behind the General Linear Model (GLM) is that the observed signal ( $Y$ ) is a linear combination of explanatory variables ( $x$ ) and gaussian noise ( $\epsilon$ ), thereby equivalent to linear regression.

$$Y_j = x_{j1}\beta_1 + \dots + x_{jk}\beta_k + \dots + x_{jL}\beta_L + \epsilon_j$$

The  $\beta_k$  are unknown parameters to be estimated, corresponding to each of the  $L$  explanatory variables ( $x_{jk}$ ) for the  $j^{\text{th}}$  observation of  $\mathbf{Y}$ . This can be expressed in the matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

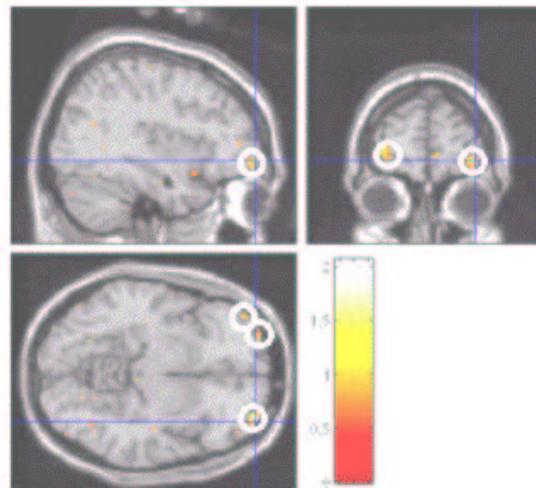
Here,  $\mathbf{X}$  is called the design matrix that contains the explanatory variables and  $\boldsymbol{\beta}$  is the parameter estimate. Each column in the design matrix corresponds to some effect that one has built into the experiment or effects that may confound the results. The rows of the matrix correspond to the number of observations. Commonly used parametric models, such as linear regression, t-tests and analysis of variance (ANOVA) are special cases of the general linear model.

### 3.1.1. Fixed Effects Model and Conjunction analysis

The GLM approach to neuroimaging has been used to great success during the past 10 years. For majority of behavioral studies, the "box-car" design is adopted where the activity of brain areas correlated with test epochs is obtained by comparing the averaged activation in test epochs to that of the control epochs (see Fig. 1). This approach is called the "cognitive subtraction" approach (Friston, 1997). The subtractive approach assumes that brain activity scales in a linear fashion. Higher order effects can be specified by adding *non-linear* terms to the GLM. However, this requires that the nature of hypothesized relationship between the brain activation and an

experimental variable is already known. Further, the statistical assumption in such *fixed effects models* is that the test epochs are replicated and hence can be averaged and compared to the control epochs. Fixed effects model can also be effectively used to find the typical activation of a group of subjects or compare between subject groups. For example, if among eight subjects, three of them show significant activation then the average activation will also show significant effect.

Average result from group analysis considers only the within-subject / session variability. One can perform *Conjunction analysis* at the second level to fixed effects model to account for variability of activation across the subjects. Fixed effects models along with conjunction analysis are powerful tools to characterize typical and generic aspects of the functional architecture of the human brain (Friston et al., 1999). For the case study that we present in this paper, we performed a fixed effects analysis on a pool of five subjects' data. Corresponding to the trial and error task when compared to the guided task, a typical activation was found in the brain area called Orbitofrontal Cortex (OFC). We further verified this activation using conjunction analysis of subject specific contrasts. As the results shown in Figure 2 indicate, OFC activation remains intact in the conjunction analysis also. In summary, one serious limitation of the fixed effects analysis is that the variability across subjects / sessions is ignored and hence the inference cannot be generalized to population.



**Figure 2.** Orbitofrontal Cortex activity using  $2 \times 6 > 1 \times 12$  contrast.

### 3.1.2. Partitioning the Model into Learning Stages

Strictly speaking, fixed effects method is not suitable for learning studies as it ignores the fact that the epochs are not replicated and that learning causes differential changes in the early as compared to late epochs. A simple

way to incorporate early versus late differences is to partition the overall average into two parts – one corresponding to the average of early epochs and the other corresponding to the late epochs. The comparison of these two averages gives us the typical features of the early stages of learning and that of the late stages. In the current case study, it is interesting that the activation in OFC is observed in late stage but not in the early stage.

### **3.1.3. Time effects within a session (Parametric Modulation)**

In the approach outlined thus far, we ignored any session specific improvements. These can be accounted by using the "parametric modulation" approach in SPM. Parametric modulation is a special case of the approach of using "user specified regressors" in SPM. Regressors are used to study session specific adaptation in brain activity, general effects of time on brain activity and Brain-Behaviour correlations effectively.

### **3.1.4. Random Effects Analysis (RFX)**

All the above methods use fixed effects model where inter-session variance is ignored. Subjects continually adapt to a particular task and both neuronal and cognitive adaptations take place. It can thus be argued that no observation is truly a replicate of a previous one (Vazquez and Noll, 1998). An approach called random effects analysis (RFX) takes inter-session variance also into account and provides generalizability of results to populations and also to make quantitative inferences (Friston et al., 1999). The implementation of RFX in SPM package involves a two-level analysis — in the first level a fixed effects analysis is done and then followed by a second level analysis. In the second level analysis, the contrast images from the first level are analyzed using student's t-test or ANOVA-like methods. Problems with RFX are that it requires large sample and is more complex to implement.

In this section we summarized the current approach to fMRI analysis using general linear models (GLM). Statistical Parametric Mapping (SPM) is one popular and comprehensive implementation of GLM. We outlined several ways of characterizing learning effects within the SPM framework. These involve either partitioning the overall model into learning stages, or inclusion of regressors or parameters that represent time effects, or using a random effects analysis. All these methods offer different ways of adapting the basic SPM framework for

studying learning related effects. As such these methods address the problem only partially. In the next section, we point out alternative approaches.

## **4. ALTERNATIVE APPROACHES FOR CHARACTERIZING LEARNING INDUCED CHANGES**

### **4.1. Statistical Time Series Approaches**

As described in the previous section, the SPM framework has special limitations when used for studies that investigate learning and similarly for investigating pharmacological effects. Hence there is an immediate need to explore alternative methods that explicitly characterize the effects of time in fMRI data. One alternative is to augment the current methodology with a second-level analysis where statistical time series techniques are applied on regions selected from the first-level ANOVA based analysis. Explicit modelling of time series can potentially reveal and enable the characterization of learning induced changes in the brain activity. A series of 228 BOLD intensity values are obtained from the activated brain areas. These intensity values are extracted from the local maxima (of voxels) in the volume of interest as revealed by the SPM fixed effects analysis described in section 3.1.1. We used the standard Auto Regressive Integrated Moving Average (ARIMA) statistical time series model (Box and Jenkins, 1976), as it is well known for uncovering hidden patterns in time series data.

Brahmaiah and Rao (2002) conducted preliminary investigations on our fMRI time series data. The coefficients of ARIMA model from different brain regions revealed a pattern of periodicities in the control condition but dissimilar behaviour in the test condition. The time series data of the sequence learning (test condition) fitted optimally in an auto regressive (AR) model reflecting dependencies on previous history. The dependency could possibly be related to the progressive sequence learning accomplished by the subject. The data from the control condition fitted optimally in a moving average (MA) model revealing that there are no history-related dependencies. This is in accordance with the experimental design as there is no learning involved in the control condition. Further investigation needs to be carried out to tease out various components in the AR model related to sequence learning (test condition).

#### **4.2. Modelling haemodynamic response function (HRF)**

The other possible approach to characterize learning effects is to explicitly model and estimate the haemodynamic response function (for example, Svensen et al., 2000; Marrelec et al., 2002). HRF is the theoretical signal that BOLD fMRI would measure in response to a single, very short stimulus of unit intensity. The local change of BOLD is not immediate but usually has a delay of 2-6 seconds from the onset of stimulus. This change is observed to increase slowly and attains maximum and returns to the baseline. The signal obtained from the fMRI experiments generally consists of noise and BOLD response components. The noise is due to physiological sources such as breathing, heartbeat and system sources like scanner and noise drift.

Estimation of the HRF is of great interest when analyzing fMRI data, since it can give a deep insight into the underlying dynamics of brain activation and relationship between activated areas. There are broadly two categories of techniques for HRF estimation — parametric approaches and non-parametric approaches. In parametric approach HRF is modelled by functions such as Gaussian, Gamma, Poisson etc. These functions give a parsimonious representation of the underlying HRF. With these models the problem simplifies to estimating the parameters of the above functions. In the non-parametric approach the HRF is modelled by an FIR filter. The coefficients of FIR filter are then estimated from the fMRI time series data.

#### **4.3. Independent Component Analysis (ICA)**

Another powerful alternative approach is to use model-free methods such as the Independent Component Analysis. ICA is a data-driven method and can potentially reveal the complex spatiotemporal dynamics of a task (Bell and Sejnowski, 1995). The ICA algorithm is similar to principal component analysis (PCA) in that it decomposes a data set into discrete components. PCA orients the first component in the direction of maximal variance in the data set with subsequent components oriented orthogonally. The ICA algorithm decomposes the data set using the principle of minimizing the mutual information between components. The resultant spatial maps in ICA are independent, but the corresponding time courses are not constrained to be independent. ICA may be better at identifying task-related signals in the brain, wherein the contribution of cognitive effects to the overall variance in functional imaging data is relatively small (Berns et al., 1999). It is particularly useful in paradigms in which the time course of the brain response is unknown. This is a powerful approach, because it

allows one to design experiments in the absence of fixed effects, which are necessary for conventional ANOVA-type models.

## **5. SUMMARY AND CONCLUSIONS**

In this review, we have presented four different approaches for studying the temporal dynamics of brain activity, especially for learning tasks. The first approach is the traditional analysis using linear methods such as Statistical Parametric Mapping. Fixed effects model together with conjunction analysis in SPM can potentially help in characterizing typical activation in subjects. Random Effects analysis can be used to generalize results to the population. SPM framework is useful but has serious limitations for characterizing learning effects. The second approach is to use statistical time series methods at the second level on the results obtained from the first level fixed effects analysis. We presented example case studies to illustrate these two approaches.

The other two approaches are explicit modelling of HRF and ICA. Modelling HRF is also of interest because HRFs are increasingly suspected to vary from region to region, from task to task, and from subject to subject. ICA appears to be promising, as it does not make any assumptions about the brain response signals other than their statistical independence. These approaches will be investigated on our data in the future.

Learning is an important cognitive function and fMRI studies have not yet started investigating this phenomenon seriously. One possible reason is that there are still many interesting open questions related to the analysis methodology for studying learning induced changes in fMRI data.

## **ACKNOWLEDGMENTS**

We would like to thank Dr. Kenji Doya and Dr. K. Samejima of Human Information Science Laboratories, ATR Labs, Kyoto for their help and intellectual inputs in the design of experiments and the acquisition of fMRI data. We would also like to acknowledge the help of Dr. Chakravarthy Bhagvati, University of Hyderabad for help with the time series analysis and Mr. R. Srikanth, Indian Institute of Science, Bangalore for discussion on modelling and estimation of HRF.

## REFERENCES

- Bapi RS, Doya K, Harner AM (2000a) Evidence for effector independent representations and their differential time course of acquisition during motor sequence learning. *Exp Brain Res* 132:149–162.
- Bapi RS, Graydon FX, Doya K (2000b) Time course of learning of visual and motor sequence representations. Annu meeting of society for Neurosci USA.
- Bell AJ, Sejnowski TJ (1995) An information-maximization approach to blind separation and blind deconvolution. *Neural Computation* 7:1129–1159.
- Berns GS, Song AW, Mao H (1999) Continuous functional magnetic resonance imaging reveals dynamic nonlinearities of "dose-response" curves for finger opposition. *J Neurosci* 19:RC17:1–6.
- Box GEP, Jenkins G (1976) *Time series analysis: Forecasting and control*. 2nd Edition. Holden-Day, San Francisco.
- Brahmaiah T, Rao CCS (2002) *Time Series Modelling of Learning Induced Changes in Brain Areas as Revealed by Functional MRI Data*. Master's thesis, University of Hyderabad, India.
- Friston KJ, Ashburner J, Poline JB, Frith CD, Heather JD, Frackowiak RSJ (1995a) Spatial registration and normalization of images. *Hum Brain Mapp* 2:165–189.
- Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RSJ (1995b) Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp* 2:189–210.
- Friston KJ (1997) Imaging cognitive anatomy. *Trends Cogn Sci* 1:21–27.
- Friston KJ, Holmes AP, Worsley KJ (1999) How many subjects constitute a study? *Neuroimage* 10:1–5.
- Heeger DJ, Ress D (2002) What does fMRI tell us about neural activity? *Nature Reviews* 3:142–150.
- Hikosaka O, Rand MK, Miyachi S, Miyashita K (1995) Learning of sequential movements in the monkey: Process of learning and retention of memory. *J Neurophysiol* 74:1652–1661.
- Kaas JH (1995) The reorganization of sensory and motor maps in adult mammals. In: *The Cognitive Neurosciences* (Gazzaniga M, ed), pp. 51–71. MIT press, Cambridge, MA.
- Karni A, Meyer G, Jezzard P, Adams MM, Turner R, Ungerleider LG (1995) Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature* 377:155–158.
- Marrelec G, Benali H, Ciuciu P, and Poline JB (2002) Bayesian estimation of the haemodynamic response function in Functional MRI. In: *Bayesian Inference and Maximum Entropy Methods* (Fry R, ed), pp. 229–247, MD, USA.

- Ogawa S, Lee TM, Kay AR, Tank DW (1990) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci USA* 87:9868–9872.
- Pammi VSC, Bapi RS, Miyapuram KP, Samejima K, Doya K (2003) The activation of orbitofrontal cortex reflects trial and error processes in a visuomotor sequence learning task. In *Networks and Behavior*, NCBS Neurobiology Symposium, Bangalore, India.
- Svensen M, Kruggel F, von Cramon DY (2000) Probabilistic modeling of single-trial fMRI data. *IEEE Trans on Med Imaging* 19: 25–35.
- Ugurbil K, Toth L, Kim DS (2003) How accurate is magnetic resonance imaging of brain function? *Trends in Neurosci* 26:108–114.
- Vazquez AL, Noll DC (1998) Nonlinear aspects of the BOLD response in functional MRI. *Neuroimage* 7:108–118.
- Volkow ND, Rosen B, Farde L (1997) Imaging the living human brain: Magnetic resonance imaging and positron emission tomography. *Proc Natl Acad Sci USA* 94:2787–2788.