Fall 1-27-2008

Comparison of linear parametric models for predicting fMRI response

Parina Gandhi
New Jersey Institute of Technology

Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

COMPARISION OF LINEAR PARAMETRIC MODELS FOR PREDICTING FMRI RESPONSE

by

Parina Gandhi

In this study, five different linear parametric models including Autoregressive model (ARX), Autoregressive Moving Average Model (ARMAX), Box-Jenkins Model (BJ), Instrument Variable Model (IV) and Prediction Error Model (PEM) were used to predict the fMRI response and their performances compared. Transfer functions were computed for every voxel time series for every subject using all the parametric models. Cross-correlation was subsequently performed between the predicted response and the actual fMRI data to compare the performance of the five models. The consistency of the models and the transfer function was checked by doing a statistical analysis. Among the five models tested, PEM resulted in the highest correlation coefficient of 0.76 with the measured response, while ARX, which was the simplest of all, gave the least correlation coefficient of 0.23 with the measured response. The PEM model was consistent in predicting the response between the subjects compared to all other models. A significant difference between the PEM model versus the other models was observed for all the subjects.

Keywords: Parametric model, hemodynamic response, Autoregressive, transfer function cross-correlation.

COMPARISION OF LINEAR PARAMETRIC MODELS FOR PREDICTING FMRI RESPONSE

by
Parina Gandhi

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Masters in Biomedical Engineering

Department of Biomedical Engineering

January 2008

APPROVAL PAGE

COMPARISION OF LINEAR PARAMETRIC MODELS FOR PREDICTING FMRI RESPONSE

Parina Gandhi

Dr. Tara L Alvarez, Co-Advisor — Date
Associate Professor of Biomedical Engineering, NJIT

Dr. Bharat B Biswal, Co-Advisor — Date
Associate Professor of Department of Radiology, UMDNJ

Dr. Max Roman, Committee Member — Date
Assistant Research Professor of Biomedical Engineering, NJIT

BIOGRAPHICAL SKETCH

Author: Parina Gandhi

Degree: Master of Science

Date: January 2008

Undergraduate and Graduate Education:

- Master of Science in Biomedical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2008

- Bachelor of Science in Electrical Engineering,
  Sarvajanik College of Engineering and Technology, Surat, India, 2006

Major: Biomedical Engineering

To my beloved family

ACKNOWLDEGEMENT

I would like to express my deepest gratitude to Dr. Bharat Biswal, who was my research supervisor, and provided valuable and countless resources, insight, and intuition, but also constantly gave me support, encouragement, and reassurance. Special thanks are given to Dr. Tara Alvarez for being my advisor at NJIT and providing necessary guidance and Dr. Max Roman for actively participating in my committee.

I would like to show my appreciation to my fellow mates at UMDNJ Dr Kannurpati, Nirvish, Samata, Sheela, and Priyanka for their support, help and motivation during my study at the UMDNJ Research Laboratory.

Last, but not the least I appreciate the support and love of my family that they had given throughout my studies.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES
(continued)

CHAPTER 1

INTRODUCTION

1.1 Preface

Functional Magnetic Resonance Imaging (fMRI) is used to map the changes in the brain hemodynamic that measures neural activity corresponding to a stimulus or task [1-5]. Recent development in hardware and software in MRI has made possible the acquisition of sequential images at unprecedented spatial and temporal resolution. This has made it possible to better understand the neurovasculature of the brain and help explain the underlying neuronal activity associated with eloquent region or regions of the brain. The ability to observe both the structures and the underlying functional aspects associated with the structure has helped propel fMRI to become the method of choice for studying human brain function. fMRI provides high spatial and temporal resolution in addition to being non-invasive, and detects changes in neural activity by measuring the changes in the Blood Oxygen Level Dependent (BOLD) and blood volume in the brain [6].

The ability to observe brain function has opened an array of new opportunities to advance our understanding of brain organization in both patient populations and healthy subjects. This for example has created a potential new standard for assessing neurological status and neurosurgical risk. Clinically fMRI can be used on patients for mapping critical areas before they undergo brain surgery. It is also currently being used in association with other clinically relevant information for early identification of psychiatric and central nervous system (CNS) disorders and for the measuring the effects of therapies on neurodegenerative and neurodevelopment disorders. Thus, in a short time

since its inception in 1990's, fMRI has emerged as one of the most promising techniques in the imaging field.

1.2 Objective

Like any scientific experiment, the objective in most fMRI experiments is to understand the activated response (output) following the stimulus or task presentation (input), and the underlying mechanism associated with it. This is typically done by varying the input stimulus and predicting the change in the output response. Transfer functions which mathematically relate the input stimulus with the output response, are typically used to accomplish this. In this study, five different parametric models including Autoregressive Model (ARX), Autoregressive Moving Average Model (ARMAX), Box-Jenkins (BJ) model, Instrument Variable Model (IV) and Prediction Error Model (PEM) [7, 8] were used to build the transfer function and consequently used for prediction. Thus, the overall purpose of this study was to compare different parametric models and in particular their ability to predict the fMRI response. These models were defined using linear techniques.

The transfer function for each voxel depends upon the complex interactions among neurons, underlying hemodynamic function and the task. However, the linear transform model may vary on a voxel by voxel basis. The noise also plays an important role in predicting the stimulated response. This will further help in understanding the underlying neuronal activity in the human brain. Different parametric models were used so that validation can be carried out to find the method that would give the optimum result. The transfer function associated with healthy subjects and patient populations (with altered neurovasculature) would be different. Hence, the difference between the

two transfer function can be used to understand, quantify, and predict responses from the different subjects.

1.3 Background Information

Investigators have tried to predict the fMRI response using linear-time invariant (LTI) techniques [9-11]. In a LTI system [12], any output can be determined by convolving the input with the transfer function. The properties of a LTI system include linearity, time-invariance, scaling and superposition. According to a theory of LTI system, if the fMRI stimulus used was shifted in time (say by a few seconds), than its response will also shift by the same amount in time. Similarly, if the amplitude of the stimulus was increased (or decreased) then the fMRI response would also change in the same fashion. For the superposition property, for example, if we have two different stimuli for fMRI then the response will be the addition of two individual responses from the two different stimuli.

Boynton and his colleagues [9] characterized LTI relationship between fMRI and neural activity by predicting its response. They tested three specific hypotheses for the study. First, they tested if the fMRI response was time-invariant. Second, they tested if the superposition holds true, i.e., if responses from short stimulus could be used to predict responses from longer stimulus duration. Third, they tested if the noise was independent of the stimulus. They found that the first and third hypothesis was true and hence, fMRI response was time-invariant and independent of noise. However, for the second test, their data suggested that the predicted response estimate obtained by concatenating several short stimuli was not very accurate when compared to the original longer stimulus duration.

In a similar study, Glover [10] measured the temporal characteristics of the BOLD response in sensorimotor and auditory cortices while stimuli were presented that varied between durations of 167 msec to 16 seconds. In this study too, the superposition principle did not hold true in fMRI (i.e., impulse response estimated using a 1 second stimulus did not predict well the output response for long duration stimulus). Glover predicted the response to a long stimulus of finger tapping from the impulse response having different temporal separation at the rate of 1 to 4 Hz. Since there was an overlap between the responses, he used Weiner deconvolution to separate the response for each individual stimulus. He concluded that the separation using Wiener deconvolution was possible only if the response had temporal spacing of $\geq$ 4 sec and hence, this method could be used for removing the temporal blurring during fMRI response.

In another study, Cohen [11] used a gamma-variate model to fit the impulse response function with the behavioral conditions to obtain a better prediction of the fMRI response. In this study, the authors used a very brief visual stimulus (for one second) which was presented several times and averaged the fMRI response to obtain the impulse response function. Because, an impulse response given as a stimulus results in the output being the impulse response function, the author used this to estimate the impulse response function rather than an input-output model to characterize the impulse response function. The mean output, which was the impulse response function, was subsequently modeled using a gamma-variate function [13]. The fitted model of the impulse response function was then applied to estimate the fMRI signal response to complex stimulus of irregular timings and variable intensities.

Recently, ARX method has been applied by Baraldi and colleagues [14] and they

have tried to estimate and predict the fMRI response by using one of the parametric models called ARX which is used in this study also besides, other models like ARMAX, BJ, IV and PEM. They predicted various parameters including time to onset, peak time and amplitude of the BOLD response. They found that ARX model was quite effective in predicting the response.

1.4 Outline

This thesis has been organized as follows: The coming chapter explains the concepts of MRI and fMRI. It explains the generation of MRI signal, acquisition of image, basic theory of fMRI and the understanding the fMRI as an imaging technique. The chapter that follows explains the theory behind the parametric models and understanding the peculiarities of every model and trying to get an idea of how it can be used for the analysis of fMRI data. The chapter after above includes the experimental setup for this study followed by the data processing. The statistical analysis carried out on the data set used for this study is explained in this chapter to have a better understanding of the results. The above chapter is followed by the Results. The result from every subject has been shown. The last chapter is the discussion for the results observed and some of the assumptions made during study. Lastly the references used for this study has been mentioned.

CHAPTER 2

FUNDAMENTALS OF MRI AND FMRI

2.1 Basic Principle Behind MRI

The Magnetic Resonance Imaging (MRI) [15, 16] Scanner has three main components, consisting of the main magnet (generating the magnetic field), the gradient coil (used for spatial encoding of images), and the radiofrequency coil (used to excite the protons and receive signals from the excited protons). All these along with the other components including shimming coil and reconstruction system are essential in generating an MRI image.

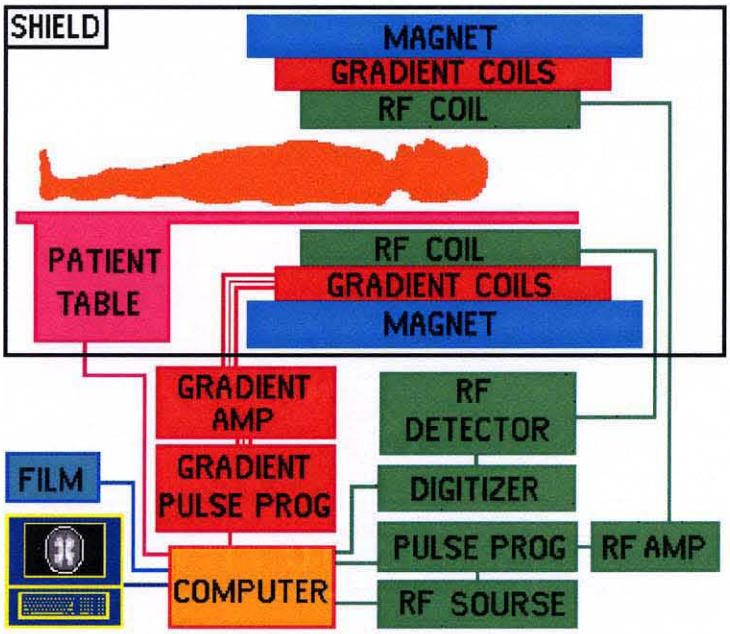

Figure 2.1 Schematic of a MRI system.
Source: http://neurocog.psy.tufts.edu/images/mri-scanner1.gif

The static magnetic field generates constant magnetic field. The magnetic field for MRI should be homogeneous and of suitable strength. If the magnetic field used is not homogeneous than the signal measured from the particular part of the body would depend

on where the body is located in the magnetic field. Permanent magnets have initially been used to generate the main magnetic field. However, in permanent magnets it is difficult to achieve homogeneity over a large volume required for imaging. Further, high field permanent magnets are extremely difficult in addition to being prohibitively expensive; as a result, permanent magnets are typically confined till 0.2 Tesla. Thus, to create large magnetic field, superconducting electromagnets are used instead of permanent magnets. Superconducting magnets, as the name implies, are created by generating magnetic field strength from wires that are cooled by liquid helium to reduce their temperature to near absolute zero, thereby creating superconductivity. For humans, the constant magnetic field strength usually varies between 1T to 7T and for animals it can go as high as 11 T.

The radio-frequency (RF) coil having transmitter and receiver generate the signal at the resonance frequency of the atomic nuclei in the static magnetic field. This coil is called radiofrequency coil, as it has a resonant frequency corresponding the protons to be excited at the magnetic field in the range of radio waves in the spectrum of electromagnetic waves. These coils must also have property of uniformity and sensitivity. When the body is placed in the magnet, the net magnetization of atomic nuclei is aligned with the magnetic field. This electromagnetic coils than generate waves at a resonant frequency depending on the strength of the magnetic field. This is called excitation and nuclei absorb the energy of the radiofrequency pulse. When the pulse ends, the energy that is released by the nuclei is absorbed by the radiofrequency coils in a process called reception. The signal generated and collected is the raw MR signal, and reconstructed to generate the MR image. In MRI, the radiofrequency coils are placed around the head using a surface coil or volume coil. Since, the surface coils are placed on the imaged

sample they usually have high sensitivity on the surface but poor volume (global) coverage while on the other hand volume coil provides uniform spatial coverage for a large volume. Thus, the decision to use the surface coil versus the volume coils should be made depending on the object or area to be imaged. For example, if whole brain coverage is required, volume coil is essential, while if only a superficial region (say the sensorimotor cortex or the spine) is required to scan than surface coil should be used.

Spatial encoding required to create an image are generated using the gradient coil. The gradient coil makes the MR signal spatially dependent and hence different location in space has a unique frequency centered around the Larmor frequency and contribute differently to the measured signal over time. The gradient coils generate the magnetic field strength linearly in one spatial direction. Like MR scanner and the radiofrequency coils, the gradient coils should also be uniform and sensitive. The strength of the gradient coil depends both on the current density and the physical size of the coil. It is directly related to the current density (There will be an increase in gradient field by increasing the current density) while it is indirectly related to the size of the coil (Reducing the size of the gradient increases the strength of the gradient field). The gradient field strength determines the spatial resolution of the image. In practice, the main magnet cannot be perfectly uniform especially in the presence of an object (or sample) and cannot be linear either. Thus, there is a need to correct for these inhomogeneties in the MRI scanners. In the MRI scanners, the magnetic field strength may be too strong in some location and too weak in other location. To overcome this problem, shimming coils are used which generates high magnetic field and compensate for the inhomogeneity in the magnetic field strength. For every subject, shimming coils must be adjusted differently unlike

magnetic fields as every person's head will distort the magnetic field in a different way. They are adjusted so to have an optimum signal and they are turned on for the entire imaging session unlike radiofrequency and gradient coils.

2.2 MR Signal Generation

A proton has a magnetic moment as well as angular momentum. When proton spins, its moving magnetic field generates magnetic moment and its moving mass produces angular momentum. When the nuclei have both the above properties, it is referred as nuclear magnetic resonance property (NMR). These types of nuclei are generally referred as spins. When the strong magnetic field B_0 is applied, the spins tries to align themselves either parallel to the magnetic field or in lower energy state while some might align in the direction anti-parallel to the main magnetic field i.e., in the higher energy state. This has been explained in the Figure 2.2 given below. The number of protons in the lower energy state is slightly higher than the higher energy state and is determined by the thermal energy of the region being imagined.

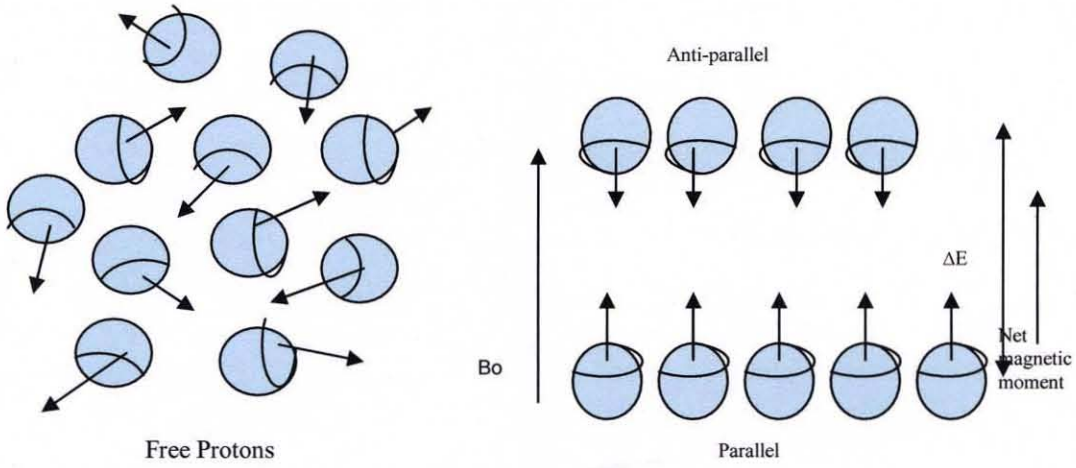

Figure 2.2 Alignment of the spin in the parallel and the anti-parallel direction on the application of the strong magnetic field.

Protons when placed in the magnetic field will try to change their orientation and undergo the gyroscopic motion known as precession. MR signal is the measure of the net magnetization of all the spins in a given volume. The net magnetization is the result of the three components. The longitudinal component that is parallel or anti-parallel to the magnetic field is defined by M_z. The longitudinal component reaches maximum amplitude at equilibrium and is denoted by M_o.The transverse component is perpendicular to the magnetic field. This component is shown in the Figure 2.3 as M_{xy}. Since, there is large number of spins within a given volume, the transverse component cancels out each other. Hence, the net magnetization is due to the longitudinal component only. This is the basis of MR signal generation.

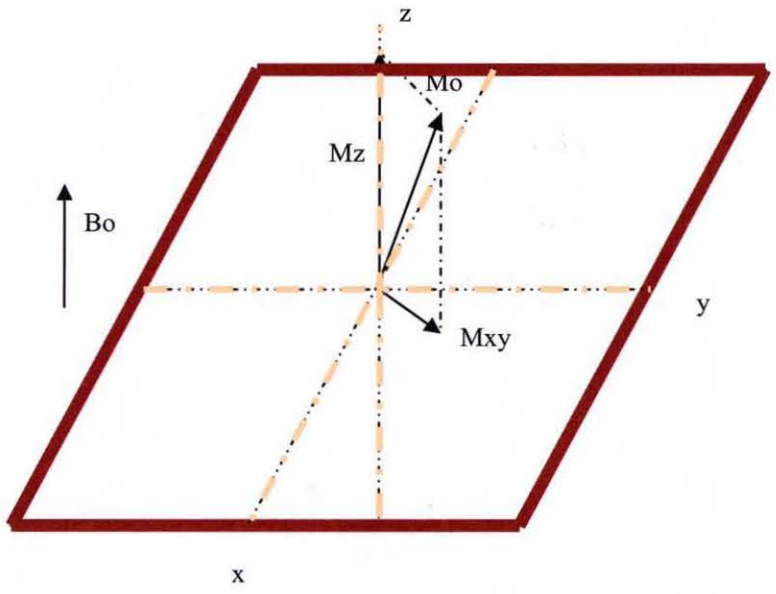

Figure 2.3 Generation of the MR signal due to the longitudinal component M_z, transverse component M_{xy} and the equilibrium component M_o.

A spin in the low energy state absorbs photons with energy equal to the difference between the low and high transition state and can jump to high energy state. This process is called excitation. Similarly, when spin makes the transition between the high and low

state it emits the energy that is the difference of the two transition state. Usually, when the spin gets excited, it disturbs the thermal equilibrium and hence, the excess spins tries to return back to the lower energy state. This process is called reception and during this time, the electromagnetic energy that is emitted is detected by the radiofrequency coils. The frequency that is required to change the spin from one state to another is called the Larmor frequency and is given by the following equation.

$$v = (\gamma / 2\pi) Bo \tag{2.1}$$

where γ is the known as the gyromagnetic ratio and Bo is the magnetic field strength. The magnetic moment is larger than the angular moment by a factor γ.

2.3 Spin Relaxation

Excitation pulse creates the MR signal, but it won't last long and will decay within few seconds. This is called spin relaxation. This decay of MR signal is due to the two components called the longitudinal relaxation and the transverse relaxation. The spin system gradually loses energy to the environment or lattice of nuclei once the excitation pulse is removed. This process is called as longitudinal relaxation or spin lattice relaxation. Since most of the nuclei return to their lower energy, the net magnetization becomes parallel to the magnetic field. Form the point of classical mechanics, the transverse magnetization goes to the longitudinal direction as it was before the excitation pulse. This results in smaller transverse leading to a smaller MR signal. The time associated with this longitudinal relaxation is called T_1 and the process is called the T_1 recovery.

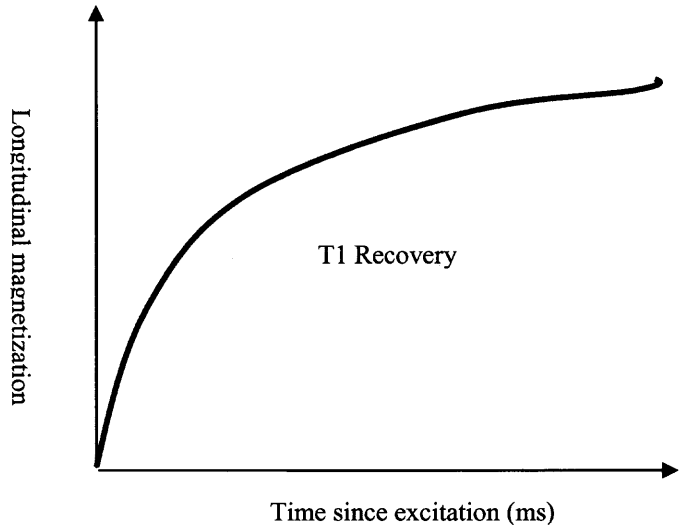

Figure 2.4: Changes in the longitudinal component of the magnetization when the magnetic field is applied.

The longitudinal magnetization M_z at a given time t is given by the following equation. M_0 is the original magnetization.

$$M_z = M_0(1 - e^{-t/T1}) \tag{2.2}$$

T_1 Relaxation is dependent on the absorbed energy dissipation which occurs in the surrounding molecular lattice. Different tissues structure and pathologies have varying relaxation time. Usually the large and slowly moving molecules have low frequency of vibration, the moderate sized molecules and viscous fluids have intermediate vibrational frequency while the small have the high frequency. Hence, the T_1 relaxation is dependent on the structures of tissues. T_1 relaxation is also directly related to the magnetic field strength.

The excitation pulse causes the spins in the field to transverse around the main field vector at about the same phase. Hence, they have their precession in the transverse plane at the same starting point. With the passage of time, in about 100 msec, this coherence is gradually lost and the spins will be out of phase. This process is called as

transverse relaxation. This phenomenon occurs due to two reasons intrinsic property of the spin. The intrinsic property is due to the spin-spin interaction between the different spins when the excitation pulse is applied. This loss of signal by the spin-spin interaction is called T_2 decay and is given by the following equation.

$$M_{xy} = M_0 e^{-t/T_2} \tag{2.3}$$

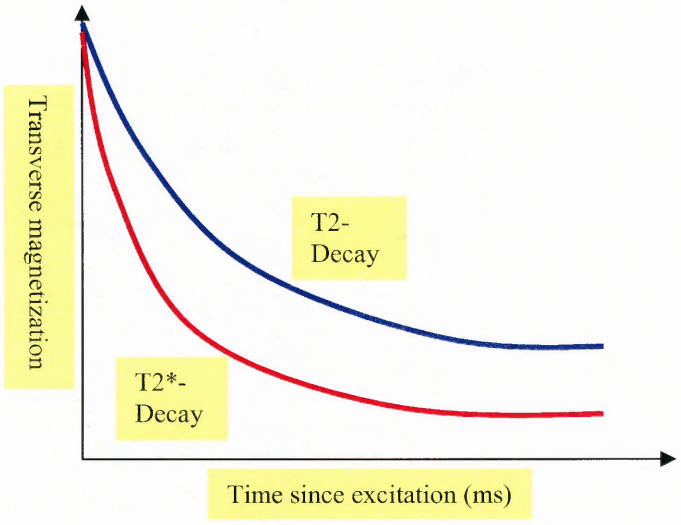

Figure 2.5 Changes in the transverse component of the magnetization when the magnetic field is applied.

The extrinsic property is due to the inhomogeneity in the magnetic field. Since the frequency of precession of spin is dependent on the magnetic field strength at that location, the variation in field strength causes the spin to precess at different frequencies. This lead to the loss of signal called T_2^* decay. This T_2^* is much faster than the T_2 as it includes the loss of signal due both the spin-spin interaction as well as due to the inhomogeneity of magnetic field. The molecular structure of the sample determines the T_2. Molecules in the cerebral spinal fluid (CSF) have mobile molecules which give a long T_2 as the fast and rapid molecular motion reduces the magnetic inhomogeneities.

Hence the non moving structures having stationary magnetic inhomogeneities have a very short T_2.

T_2 is usually shorter than the T_1. Consider a soft tissue in the body where T_1 is 500ms while the T_2 will be 5 to 10 times shorter than the T_1.The factors affecting the values of T_1 and T_2 are motion of the molecule in the given region and the size of the molecules. The size of molecule can be small, medium and large and accordingly the corresponding molecules will have fast, medium and slow frequency of vibration. Usually if the size of the molecules is small than it would have long T_1 and T_2 while the intermediate molecules have short. Similarly if the molecules are large and moving slowly will have long T_1 and short T_2. How the molecular size affects the length of T_1 and T_2 is shown in Figure 2.6.

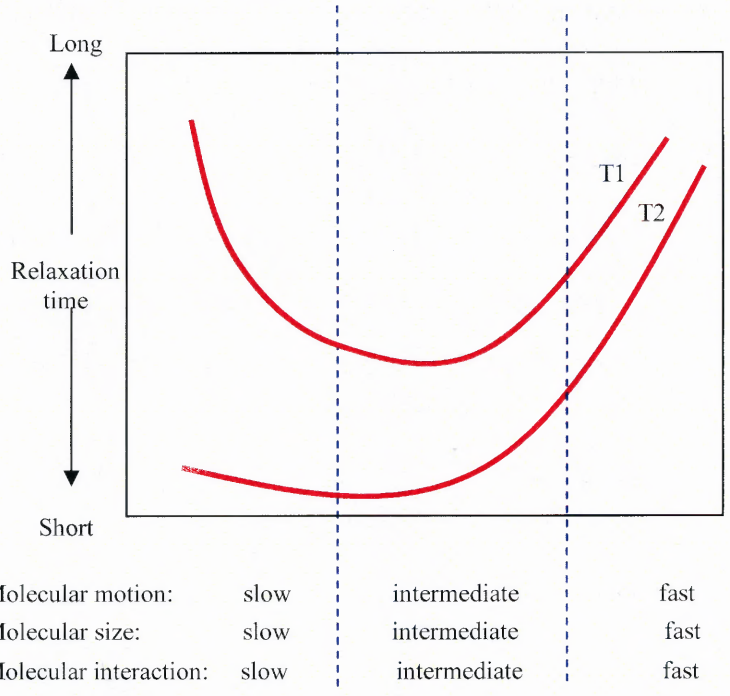

Figure 2.6 Molecular size and motion affecting the length of T_1 and T_2.

T_1 Relaxation gets influenced by the strength of the magnetic field while it does not affect the T_2 decay. This is due to the Larmor frequency and the overlap of the

molecules due to their vibration. If the magnetic field strength applied is increased the higher Larmor frequency is obtained while the reduction in the overlap of spectral is observed, which produces a longer T_1. T_1 and T_2 are the most basic properties of the tissues and for various tissues their values are shown in the Table 2.1.

Table 2.1 Values of the T_1 and T_2 for Different Tissues in the Brain

TISSUE	T_1 **for 0.5T scanner (msec)**	T_1 **for 1.5T scanner (msec)**	T_2 **(msec)**
Fat	210	260	80
Liver	350	500	40
Muscle	550	870	45
White matter	500	780	90
Gray Matter	650	900	100
Cerebrospinal fluid	1,800	2,400	160

2.4 Localization of the MR Signal

The magnetic field of one or more coils is superimposed to obtain the magnetic field gradients. This magnetic field gradient is proportional to the field strength over a given field of view. This will alter the precessional frequency of the proton corresponding to their position along the gradient. The steepness of the gradient field is determined by its peak amplitude. This peak amplitude is decided by the slew rate. Usually, a shorter slew rate is preferred. Typical values of the slew rate are 5mT/m/msec to 250mT/m/msec. MR signal is localized using slice select gradients, frequency encode gradient and the phase encode gradient.

The slice select gradient spatially directs the RF energy and hence determines the tissue in the body to be imagined. To acquire an axial MR image, slice select gradient is applied along the cranial-caudal axis of the body. The bandwidth of the RF pulse and the

gradient strength across the field of view determines the slice thickness. If the gradient field strength is held constant and a narrow bandwidth RF pulse is applied a narrow slice of tissue is excited, while the broader bandwidth will excite thicker slice. The waveform called the 'sinc' pulse is used as a RF pulse for the excitation of rectangular slice of protons. The other gradient used is the frequency encode gradient (FEG). This gradient is applied in the direction perpendicular to the slice select gradient. To acquire an axial image FEG should be applied along the X-axis. If the frequency of the precession is higher, than it will occur at the positive pole while lower precessional frequency will occur at negative pole. Phase encode gradient determines the spin position in the third spatial direction. This gradient is applied before the frequency encode gradient and after the slice encode gradient. The linear variation in the precessional frequency is seen on the application of this gradient, across the tissue. There is an advance of phase for protons in the positive direction and retardation for protons in the negative direction. A change of phase across the field of view is introduced due to strength of phase encode gradient for each TR interval. The strength of the phase encode gradient causes definite changes in the field of view. If the phase encode gradient is incremented sequentially from positive to negative while image acquisition introduces phase shift at each position along the direction of the phase encode gradient.

2.5 Image Formation in MRI

The image acquisition is MRI is a map that represents the spatial distribution of the property of the spin within a given volume of a sample. This property includes the density of the spin, also the mobility related to them and T_1 or T_2 relaxation times for

the tissues where the spins are located. A single antenna is used as a receiver which covers the entire brain region. The signal measured by the antenna is the sum of the transverse magnetization of all the voxels.

$$S(t) = \int_x \int_y \int_z M_{xy}(x, y, z, t) dx dy dz \tag{2.4}$$

S(t) is the spatial summation of the MR signal form every voxel. Since the above equation is in 3D it is difficult to analyze it. So instead it is done in two steps. First a slice selection is done as described in the above section and than the 2D encoding scheme is used within the slice to resolve the spatial distribution of the spin magnetizations.

The data acquisition in MRI is typically done using the k-space technique. K-space is a matrix which has positive and negative values of spatial frequencies encoded as a complex numbers. The k-space matrix is a four quadrant matrix having origin at the center representing frequency at that point as 0.

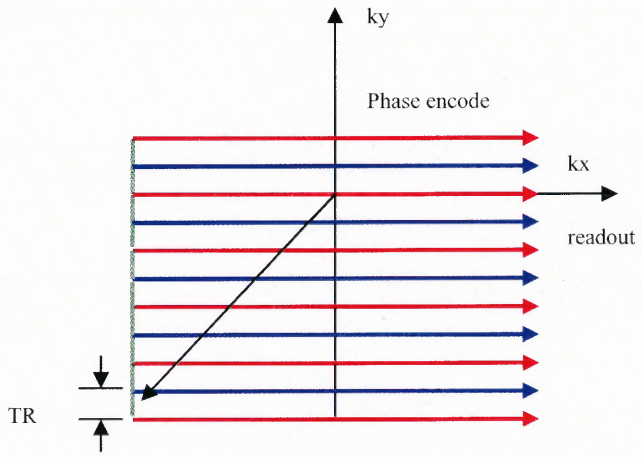

Figure 2.7 k-space matrix-Data acquisition in the MRI.

The kx direction has a data encoded by the frequency encode gradient while the ky has the data encoded by the phase encode gradient. The k-space matrix can be filled using spiral method also. The image reconstruction is done by using inverse Fourier

transform on the k-space data. Similarly, taking Fourier transform of the image gives the data of k-space. MRI image is an ensemble of spatial frequency components. The images obtained through k-space are brightest in the center and darkest near the edges. The signal to noise ratio of the MRI signal is determined by the low spatial frequency data from near the center of k-space. Conversely, high-spatial frequency data collected in the k-space helps to increase the spatial resolution of the image.

2.6 Echo Planar Imaging

Initially MR images are obtained using voxel by voxel method. In this method k-space was filled in line by line fashion which needed high number of excitation even for a image of low resolution. Hence, the entire procedure almost takes 4 hours. Instead of this method, a new approach is used called echo planar imaging (EPI) to acquire an image. In this approach, in a single excitation entire k-space is filled using rapid gradient switching. In other words, before the T_2^* and the T_2 decay occurs data must be acquired in a single excitation.

The tomographic image formation in MRI requires that the spatial encoding should be done in three dimensions in which one dimension is determined by the slice select excitation. The phenomenon of magnetic resonance depends on the match between the radio frequency excitation, pulse frequency and proton spin frequency which indirectly depends on the local magnetic field and hence will excite the MR signal over a small range of locations. The differences between EPI and conventional imaging occur in the remaining "in-plane" spatial encoding. The spatial resolution can be defined as the product of the imaging gradients amplitudes and the ON time of the stimulus. The spatial

resolution determines the pixel size during the acquisition of a data. The pixel size is equal to 1/gGt, where g is the Larmor constant (4258 Hz/gauss), G is the gradient amplitude, usually expressed in gauss/cm, and t is the gradient on time. If having a gradient of 0.5 gauss/cm and the on time of 10msec will give the spatial resolution of 0.47 mm. But this is only the spatial encoding along one in-plane dimension only. If a two-dimensional Fourier transform imaging is considered than the encoding for the second in-plane dimension will occur by having a small gradient pulse before each readout line. It is easier to understand this in terms of k-space where the k-space represents the MRI raw data before using Fourier transform to obtain and image. The signal location in k-space is the integral of the gradient amplitude and on-time

$$k_i(t) = \gamma \int G_i(t)dt \tag{2.5}$$

where ki is the location in k-space along the i axis, Gi(t) is the gradient amplitude along the i axis as a function of time, g is the Larmor constant, and t is the gradient on-time.

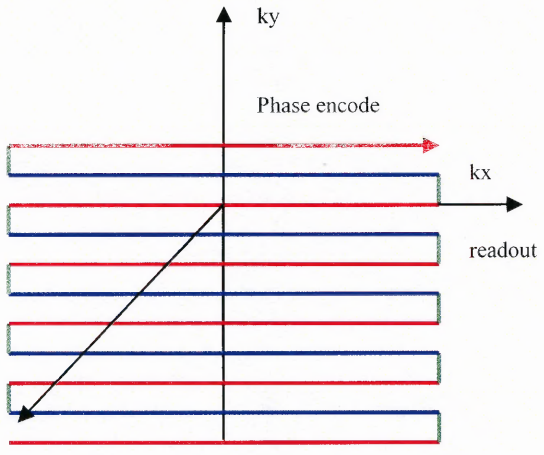

Figure 2.8 k-space-Data acquisition technique for the EPI sequence.

It can be seen that as the gradient time product increases signal can be encoded to higher k values which in turn will increase the resolution. Hence, in order to have a better resolution in EPI larger k-space must be sampled which usually takes longer time and hence k-space must be filled rapidly.

This needs stronger gradient of about 2.5 Gauss/cm. As can be seen from the figure 2.7 k-space in the conventional MRI is filled line by line after very RF excitation. Here data is collected in a single line i.e., in kx direction. Since separate excitation is required for every line the total imaging time depends on the time between excitation and the total number of data lines collected. While in EPI since all data are acquired during single excitation, here the imaging time depends on the resolution and the field of view in the final images.

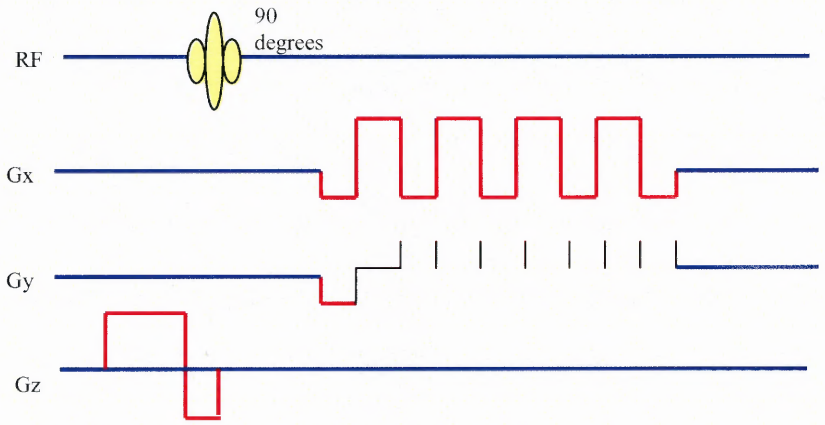

Figure 2.9 Schematic of the pulse sequence used for the EPI image.

The schematic of the k-space and the EPI for the same is shown in Figure 2.8 and Figure 2.9 respectively. Since all the data are collected in a single shot for EPI this has considerable advantage over a conventional method. The time required for the imaging is decreased considerably leading to a very high efficiency. The speed of the EPI is due to use of very high amplitude field gradients which in turn require very high sampling rate.

But in contrast to this EPI has a poor bandwidth. This in turn drops the SNR of the signal. The reason behind the low bandwidth along the phase encoding axis is due to the continuous encoding scheme used in EPI. This in turn causes a shape distortion artifact. Another most common artifact in EPI is ghosting from eddy currents. Due to back and forth trajectory in k-space used in EPI, this artifact occurs. This can be overcome by using a robust gradient coils such that induction of current in them is minimized. At present, shielded or screened gradient coils are used in which different set of gradients coils is wound around outside the primary coil set to nullify the external magnetic fields Imperfection in the magnetic fields causes artifacts in the image acquired by EPI. Besides this, inhomogeneties in static field also cause the loss of signal and geometric distortions.

2.7 Functional MRI

Functional magnetic resonance imaging (fMRI) [16] measures the hemodynamic response related to neuronal activity in the brain in response to a task or stimulus [1-5]. It is currently hypothesized that when neuronal activity increases, there is an increased demand for oxygen and consequently the local response is an increase in blood flow to eloquent regions of the brain.

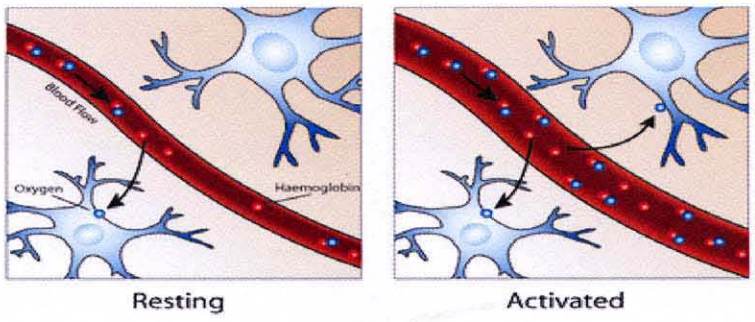

Figure 2.10 Changes in the blood flow during the resting and the activation state.
Source http://www.fmrib.ox.ac.uk/education/fmri/introduction-to-mri/images/BOLDeffect.png/image_preview

The difference in magnetic properties of hemoglobin leads to small differences in the MR signal of blood, depending on the degree of oxygenation. The Figure 2.10 explains the difference in the blood flow during the resting and when there is activation. Since blood oxygenation varies according to the levels of neural activity these differences can be used to detect brain activity. This contrast mechanism which is currently the most popular method in fMRI is known as blood oxygenation level dependent (BOLD) contrast imaging [6].

The change in the BOLD MR signal caused by the neuronal activity is known as hemodynamic response function (HRF) [16]. The amplitude of the HRF is proportional to the rate of neuronal firing and the width of HRF is proportional to the duration of neuronal activity. Suppose a sensory stimulus is given than the cortical neuronal responses will occur within tens of milliseconds of stimulus while the first HRF is initiated 1-2 seconds later. Hence HRF lags the neuronal events by some amount. It has been shown by researchers that the initial dip of 1 or 2s duration has been due to a transient increase in the amount of deoxygenated hemoglobin. After a short span of time, there is an increase in the metabolic demand of neuronal activity over a baseline levels due to increased flow of oxygenated blood. Now the signal reaches the peak of the hemodynamic response. After a few seconds the signal decreases in amplitude and goes below the baseline level and remain there for a small span of time. This effect is called as post-stimulus undershoot. This is shown in the Figure 2.11 which is the hemodynamic response to single shot duration. The changes in the hemodynamic response due to a block of multiple consecutive events are shown in the Figure 2.12.

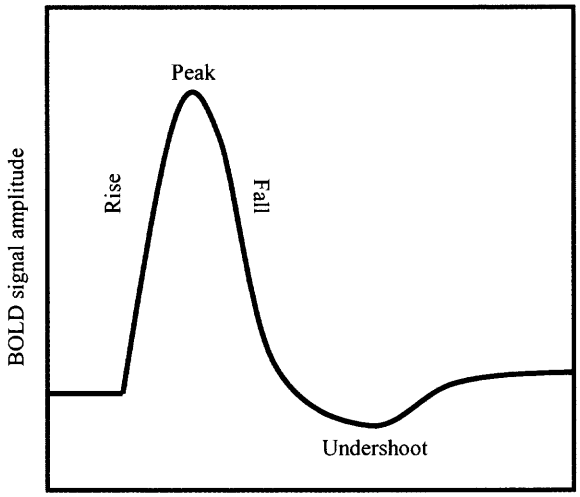

Figure 2.11 Schematic representation of the fMRI hemodynamic response to a single short duration.

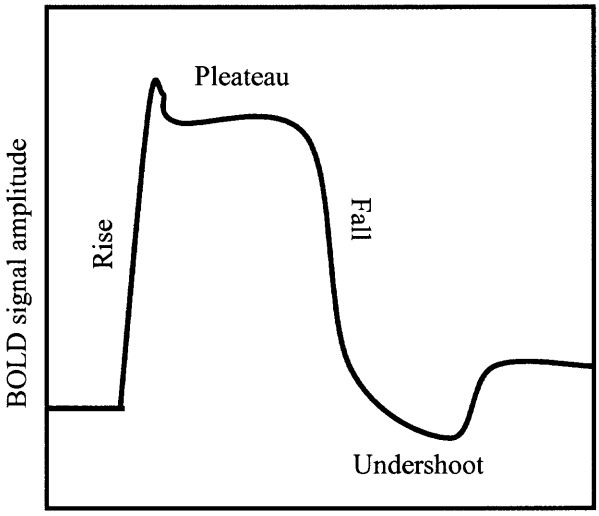

Figure 2.12 Schematic representation of the fMRI hemodynamic response to a block of multiple consecutive events.

This phenomenon can be understood by considering the changes in the blood flow and blood volume separately. When the neuronal activity ceases, there is a rapid decrease in blood flow compared to the blood volume. Hence, if the blood volume remained above baseline levels and the blood flow at baseline then a greater amount of deoxygenated hemoglobin would be present. This would lead in the reduction of the entire fMRI signal below the baseline levels and no sooner the blood volume returns to normal level the

signal rise to the baseline. The BOLD signal in a given voxel is proportional to the total amount of deoxygenated hemoglobin and the noise present.

The neuronal activity associated with the cognitive process has both the spatial and temporal properties. The ability to distinguish changes in maps in a spatial location is called the spatial resolution or in other words it is the ability to separate adjacent regions with different functional properties. The important parameter that determines the spatial resolution is the voxel size which is typically 3 to 5 mm for entire brain. The temporal resolution distinguishes the changes in the activity in a single location over time. fMRI is considered to have an intermediate level of temporal resolution. The determining factor for the temporal resolution is the repetition time (TR) which is about 1 to 3 sec.

MRI and fMRI differ in number of ways. The purpose of structural MRI is to differentiate between the different types of tissue. Therefore, each image acquired in the structural MRI is a snapshot of the underlying tissue so acquiring one image is sufficient for mapping the brain structure if the contrast to noise is high. fMRI tries to relate the changes in neurophysiology over time corresponding to the experimental task/stimulus. In fMRI, therefore, one image is not sufficient to get the desired information of the brain activity. Hence, a sequential series of images are obtained over a span of time. For this, fMRI data is collected as a time series in a temporal order at a specified rate. The corresponding time series data is analyzed for the changes corresponding to the stimulus given.

The purpose of doing fMRI experiment is to first detect the activated voxels, i.e., voxels that change in a predictive way over time, and consequently estimating the time course of an active voxel. These two factors determine the spatial and the temporal

resolution in the fMRI experiment. Therefore, at times the designs used might not be able to be optimal for both detection and estimation.

The fMRI studies can be categorized in two types. First is block design and second one is the event related designs. In the block design fMRI, one or more conditions is presented in the alternating fashion. The advantage of using block design is that they are very good at detecting significant fMRI activity.

Figure 2.13 Schematic of the block design used for the fMRI response.

For a good block design, the difference between the conditions used in it, should be large enough so that when there is a response during a task block, the signal can return to the baseline during the non-task blocks. This gives maximal variability between the blocks. When the short block design is used, the hemodynamic response generated due to the task does not have sufficient time to return to the baseline during the non block design which will in turn cause the reduction in the BOLD amplitude. This will decrease the total variability of the data which in turn reduces the experimental power. Besides this, to minimize the noise present in the data the condition should have as many transitions as possible. Sometimes while using a long and control blocks it is difficult to identify if the change in the signal from one block to another occurred due to the experimental manipulation or due to drifts in low-frequency noise. Hence, a tradeoff should be set by keeping both of the above condition. Usually block length equal to the duration of the hemodynamic response is used. The block designs are very insensitive to the shape of the hemodynamic response.

In the event related design, stimuli is presented as individual events or trials. In this design it is assumed that the neural activity would occur only for short and discrete intervals. There is separation of time between different events by an interstimulus interval (ISI) which usually ranges from 2s to 20s depending on the requirement of the experiment. Unlike block design, here the stimuli are presented in the random fashion rather than an alternating pattern.

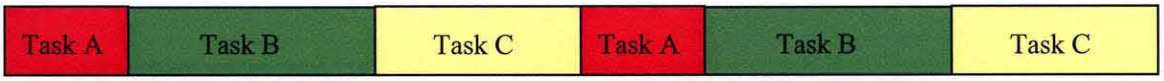

Figure 2.14 Schematic of the event related design used for the fMRI response.

The event related designs have very good estimation power for detecting the shape of the hemodynamic response. This helps researchers to infer about the relative timing of the neuronal activity, about the feedback processes and the sustained activity within a region. The flexibility in using the event related design is that one can distinguish different brain processes associated with specific parts of a task based on their relative timing.

Sometimes combination of the block and event related designs are used, called mixed designs. In this case, response is obtained to a stimulus which is presented in discrete and regular blocks but within each block there are multiple types of events. The advantage of using this design is that we have best combination of detection and estimation and help in dissociating transient and sustained component of the activity. The disadvantage of using this design is that analysis becomes complicated and the assumption of linearity may not be valid.

Functional signal to noise ratio (functional SNR) is the difference between the two states of the brain caused by the experiment setup while the noise is the variability of

that state over time. The functional SNR is depended on the intensity difference within a voxel, group of voxels or region over time. The amplitude of the BOLD response determines the functional SNR. Besides this the overall noise in the data also determines the functional SNR. The magnitude of the noise is usually larger than the magnitude of the BOLD signal. Part of the noise in fMRI is due to the thermal noise but majority of the noise signal is due to the artifacts such as head motion, physiological variation related to heart rate and respiration, unrelated neural activity, and cognitive or behavioral variability when the subject performs an experimental task. The most common method for improving SNR in fMRI study is averaging the signal. This will improve both estimation and detection power but might mask some other effects.

Once the fMRI data are collected, a number of different data analyses can be carried out on the data. In the hypothesis driven analyses, test is carried out to see how well the time course obtained from the analysis matches the idealized waveform. The analysis that needs to be carried out depends on the hypothesis to be tested. If there is need to find whether a voxel has a different mean signal intensity levels in two different experimental conditions than the common t-test can be used. Similarly, if there was a need to test the hypothesis for specific prediction about the activity change when using a model for the expected hemodynamic response then correlation test can be carried out. T-test is used to find the differences between the mean of any two distribution. Thus, while t-test conceptually easy to understand and implement, it has several drawbacks. For example, t-test can't find the differences in variability or shape differences between two different tasks. Further, for complex stimulus, the differences may not be in the mean but in more subtle parameters including time to peak, onset time, etc. For such cases, non-

parametric methods including Kolmogorov-Smirnov (K-S test) or Kruskal-Wallis test may be used.

For certain experiments, the signal obtained from the block design fMRI task has a periodic nature. This type of signal can be analyzed using Fourier transform. The analysis on the fMRI data can also be carried out by representing it as a general linear model. In this test an assumption is made that the experimental data are composed of the linear combination of the different model factors, along with uncorrelated noise.

The other different types of analysis that can be carried out on the fMRI data are data driven analyses. In the data driven analyses, a complementary approach is used for testing each of the voxel's time course against a hypothesis. The data driven techniques that are usually used is the independent component analysis (ICA) and clustering. ICA tries to identify the voxels whose activity varies similarly together over time. In clustering, voxels that have similar time course are clustered together and this process is repeated till certain criteria are reached i.e., the number of desired clusters or distance between clusters.

fMRI has a wide range of application ranging from the field of neuroscience, psychology, neurobiology, psychiatry, neuroradiology and many others. The response of the brain to a complex cognitive task can be monitored using fMRI. An example of this can be given as the activation of the frontal area during the retrieval of an episodic memory. fMRI can be used in the clinical application in number of ways. The patients can undergo an fMRI before getting operated for the brain tumor. This will help in a more precise knowledge of the tumor's proximity of the surgical field to what is known as the eloquent cortex i.e., the areas of the brain that, if injured, will result in a major

impairment such as paralysis, blindness, or loss of language function. This would in turn help planning the surgery. The task related fMRI can be used in the diagnosis and treatment of disorders of the CNS such as such as attention-deficit/hyperactivity disorder, Alzheimer's disease (AD), and Parkinson's disease. In Alzheimer's disease it can be used to study the effects of the drug that might influence the rate of cognitive decline. fMRI can also be used to find the changes in brain of the people who suffers from the genetically predisposed to neurological disorders such as Huntingdon's Disease or (AD). Especially for the former it is possible to see the changes on fMRI images of the brain long before it can be seen in structural imaging (MRI).

fMRI could be useful in providing necessary information during brain surgery for the person suffering from epilepsy where drug treatment has not being able to control the seizures. Wada test is used as a pre-surgical test for patients suffering from the epilepsy in which the catheter is inserted into an artery at the groin and than threaded into other arteries leading to the brain. This procedure is invasive and painful. fMRI has been shown to be able to correlate significantly with the Wada test.

In short, fMRI has revolutionized the current understanding of brain function in both patients' populations, and healthy controls. A number of different studies including language, episodic or autobiographical memory, conceptual categorization and mathematical calculations have been performed using fMRI. Also studies related to mental state such as goal setting, decision making, consciousness and emotion have been performed.

CHAPTER 3

THEORY OF PARAMETRIC MODELS.

3.1 Introduction

Researchers have used linear-time invariant (LTI) to predict the fMRI response. Boynton and colleagues [9] have shown that the neural activity is a non-linear function of the contrast of a visual stimulus but the fMRI response was a linear transform of the neural activity in V1 when the signal was averaged over time. But the nonlinear effects were seen for stimuli of a shorter duration in the V1 region of the brain. In another study, by Cohen [11], a measured response from a visual experiment was used as the basis to quantitate, the activation in the sensorimotor cortex with the finger tapping at different rates. However, departure from the linearity in the V1 especially for a shorter stimulus was observed at similar times as same as Boynton and colleagues.

The parametric models used for this study assume that the fMRI response is linear. While several studies have shown that the fMRI response is non-linear [17, 18], linearity was assumed for two reasons. First, for several systems including the visual and motor system, the response can be assumed as linear [9] for a longer stimulus. Further, in studies where a non-linear effect has been demonstrated in V1, it has been found that only a small initial part of the signal represents the non-linearity while the rest (majority of the signal) can be assumed as linear [9]. In some sense fMRI response is still linear so it is possible to predict the fMRI response using linear techniques. Assumption of the linearity for the fMRI simplifies the analysis and provides better understanding for the underlying neuronal activity. Therefore, for this study five different linear parametric

models were used for the analysis. The linear parametric models [7] that have been used included the Autoregressive (ARX) Model [7], Autoregressive Moving Average (ARMAX) Model [7], Box-Jenkins (BJ) model [7], Prediction Error Model (PEM) [7] and Instrumental Variable (IV) Model [7].

3.2 Linear Parametric Models

A parametric model [7, 8] can be defined as a system which has an input, an output and some noise. The Figure 3.1 shows the general parametric linear model for fMRI. In the Figure, u(t) is the stimulus for the fMRI, HRF is the hemodynamic response function of the fMRI, y(t) is the fMRI response and e(t) is the noise source present in the fMRI signal intensity.

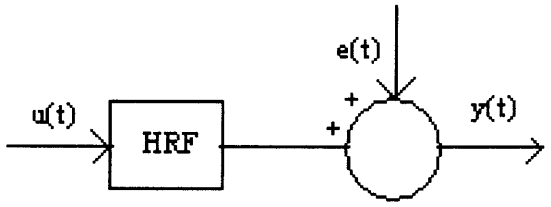

Figure 3.1 Schematic diagram of the general parametric models for fMRI is shown in the figure.

The general description of linear model for fMRI is given by:

$$Y = Gu + He \tag{3.1}$$

Here y and u is the vector representation of the output y(t) and input u(t) respectively. G is defined as the transfer function of the fMRI system, and described a linear relationship between the input and the output. Also, H is the property of the noise present in the fMRI system and is called the noise model. In this study, for all the models, both the transfer function and the noise model were considered.

The general linear parametric model for a voxel time series having n time points is given by the following equation [14]:

$$y(t)+a_1y(t-1)+.....+a_{na}y(t-n_a)= \frac{b_1}{f_1}u(t-nk)+.....+\frac{b_{nb}}{f_{nf}}u(t-nk-nb+1)+\frac{e(t)+c_1e(t-1)+...+c_{nc}e(t-n_c)}{1+d_1q^{-1}+...+d_1q^{-nd}} \quad (3.2)$$

Where a, b, c, d and f are the coefficients of the general linear model of the fMRI. This can be written in short as follows.

$$A(q)y(t)=\sum_{i=1}^{nu}\frac{B_i(q)}{F_i(q)}u_i(t-nk_i)+\frac{C(q)}{D(q)}e(t) \quad (3.3)$$

Where

$$A(q)=1+a_1q^{-1}+.....+a_{na}q^{-na} \quad (3.4)$$

$$B(q)=b_1+b_2q^{-1}+.....+b_{nb}q^{-nb+1} \quad (3.5)$$

$$C(q)=1+c_1q^{-1}+.....+c_{nc}q^{-nc} \quad (3.6)$$

$$D(q)=1+d_1q^{-1}+.....+d_{nd}q^{-nd} \quad (3.7)$$

$$F(q)=1+f_1q^{-1}+.....+f_{nf}q^{-nf} \quad (3.8)$$

For the fMRI data, the system was assumed as a single input multiple output system, where every voxel response was treated as different output. Thus, u, y, and n represent the input stimulus, the output fMRI signal, and the number of time points in every voxel respectively. The noise parameter present in the voxel time series was represented as e(t). The A(q) polynomial was the auto regressive part of the output which means how the previous instances of the output affects the input and B(q) shapes the input. The C(q) parameter makes the model a moving average auto regressive model, D(q) parameter makes the model more general to noise properties and F(q) makes the

model error free. Here, q was the model order required to adequately represent the input and output regressors.

3.3 Autoregressive Model (ARX)

The first parametric model used for our analysis was the Autoregressive Model (ARX) [7, 8]. It is defined as the following;

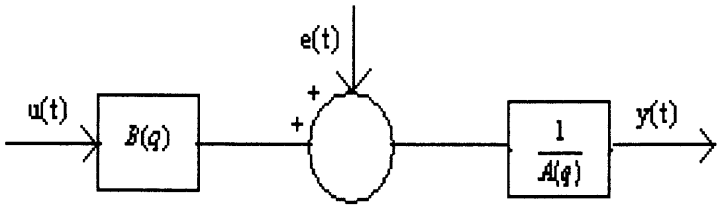

Figure 3.2 Schematic of the ARX model.

$$y(t)+a_1y(t-1)+.....+a_{na}y(t-na)=b_1u(t-nk)+.....+b_{nb}u(t-nk-nb+1)+e(t) \quad (3.9)$$

It can also be written as shown below:

$$A(q)y(t)=\sum_{i=1}^{nu}B_i(q)u_i(t-nk_i)+e(t) \quad (3.10)$$

The schematic diagram of the ARX model is shown in Figure 3.2. If we compare it with the general linear model of fMRI described earlier we can say that the transfer function and the noise model can be defined as below:

$$G=\frac{B(q)}{A(q)} \quad ; \qquad H=\frac{1}{A(q)} \quad (3.11)$$

For the ARX model assumption of a white noise was made. Also ARX model assumes that the disturbance is the part of the system dynamics and hence, treats it

together. If the system has high order and has good signal to noise ratio, ARX gives good estimation. Thus, ARX model is quite a simple model.

3.4 Autoregressive Moving Average Model (ARMAX)

This model is the extended model of the ARX model. The schematic diagram of the ARMAX [7, 8] model is shown in the Figure 3.3. It is the Autoregressive Moving Average Model with exogenous input. The coefficient of polynomial C(q) makes it more flexible while treating noise. It is generally used when the noise in the system is significant and the analysis by ARX model is not accurate.

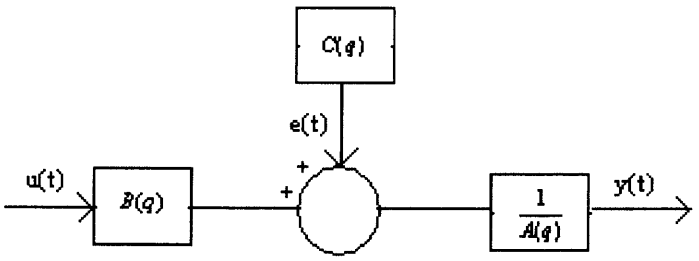

Figure 3.3 Schematic of the ARMAX model.

The ARMAX model is given by the following equation:

$$y(t)+a_1y(t-1)+.....+a_{na}y(t-n_a)= \\ b_1u(t-nk)+.....+b_{nb}u(t-nk-nb+1)+e(t)+c_1e(t-1)+...+c_{nc}e(t-n_c) \tag{3.12}$$

This can be summed up as follows:

$$A(q)y(t)=\sum_{i=1}^{nu}B_i(q)u_i(t-nk_i)+C(q)e(t) \tag{3.13}$$

3.5 Box-Jenkins Model (BJ)

The schematic diagram of the Box-Jenkins (BJ) [7, 8] model is given in the figure 1(d). Disturbance properties are modeled separately from the system dynamics using BJ model. It is more flexible than ARX and ARMAX model while modeling a noise. The coefficient of the polynomial D(q) and F(q) makes the BJ model more efficient in treating noise compared to ARX and ARMAX model.

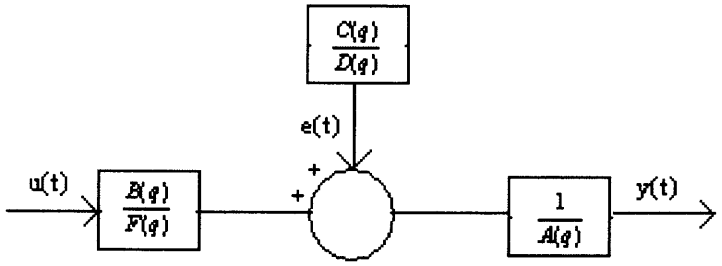

Figure 3.4 Schematic of the BJ model.

It is defined by the following equation:

$$y(t)+a_1y(t-1)+.....+a_{na}y(t-n_a)= \frac{b_1}{f_1}u(t-nk)+.....+\frac{b_{nb}}{f_{nf}}u(t-nk-nb+1)+\frac{e(t)+c_1e(t-1)+...+c_{nc}e(t-n_c)}{1+d_1q^{-1}+...+d_1q^{-nd}} \quad (3.14)$$

This can be summed up as follows:

$$A(q)y(t)=\sum_{i=1}^{nu}\frac{B_i(q)}{F_i(q)}u_i(t-nk_i)+\frac{C(q)}{D(q)}e(t) \quad (3.15)$$

3.6 Prediction Error Model (PEM)

Prediction Error Model (PEM) [7, 8] is the state space model. This model gives maximum likelihood estimate for a Gaussian distribution that minimizes prediction

errors. Using PEM model, space variables are defined by first order differential equation. From the input-output data y, u, the error e(t) can be computed as:

$$e(t) = H^{-1}(q)[y(t) - G(t)u(t)] \tag{3.16}$$

These errors are, for given data y and u, functions of G and H.

$$V_n(G,H) = \sum_{t=1}^{N} e^2(t) \tag{3.17}$$

That is,

$$[G_N, H_N] = \text{argmin} \sum_{t=1}^{N} e^2(t) \tag{3.18}$$

3.7 Instrument Variable Model (IV)

The Instrument Variable model [7, 8] is also an extended version of the ARX model. The only difference is that this method is independent of white noise. Thus, when assumed that the system have white noise at that time this method would give better approximation than the ARX model. By using filtered versions of the input we get,

$$N(q)s(t) = M(q)u(t) \tag{3.19}$$

Hence, for this study as explained above five different Parametric Models were used. All five models follow the property of linearity, superposition, scaling and are time invariant. (Stimulus and the response are linearly related. Also, the response from the two different stimuli is the addition of the response from each stimulus. If the stimulus is shifted in time by some amount response would also have the same effect. Similarly, if the amplitude of the stimulus is changed than there will be corresponding change in the amplitude of the response). Therefore, studying five models helps in studying fMRI response in a better way. Also, it can be used as validation method.

CHAPTER 4

EXPERIMENTAL SETUP

4.1 Subjects and Stimulus

Data from all the subjects were collected using a 3T Siemens Allegra system, equipped with a transmit/receive birdcage radio-frequency head coil. Volunteers were positioned supine on the gantry with the head in a midline location in the coil. Foam padding was placed between the forehead and the coil to reduce motion artifacts.

Six healthy volunteers (2 females, 4 males between 22 and 38 years of age) with no history of head trauma, neurological disease or hearing disability were scanned. All protocols in this study were approved by the local Institutional Review Board. Written consent was obtained from all subjects after explanation of the nature and possible consequences of the study. All subjects were paid a nominal fee for their participation.

For each subject, echo-planar images were obtained in the axial plane in six slices covering the motor cortex using the following parameters: 64×64 matrix, TR/TE=1000/27 ms, FOV=22 cm, slice thickness=5 mm, bandwidth=125 kHz. The imaging procedure for each volunteer was as follows: sagittal localizer images were first obtained with a conventional gradient echo sequence. The midsagittal image was used to select the axial slice over the motor cortex for functional imaging. In each subject, 1024 echo-planar images were obtained for a total scanning time of about 17 min. Verbal instructions to finger-tap matching the ON/OFF paradigm of the each task was delivered through a microphone and a speaker headphone system worn by the subject.

Two scans were performed on all subjects. During the first scan, subjects performed bilateral finger tapping for 20 sec followed by 20 sec of rest for three cycles

resulting in 90 images. During the second scan, subjects performed bilateral finger tapping alternating with periods of rest for 1024 seconds. Both the "ON" and "OFF" periods were randomized during the second scan. The minimum and maximum "ON" and "OFF" periods were kept between 8 and 32 seconds. For the finger-tapping task, subjects were instructed to successively touch each finger with the thumb of the fingers' respective hand in a self-paced manner.

4.2 Pre-Processing

The pre-processing is required to carry out on the fMRI data after the image reconstruction and prior to statistical analysis. There are two aims of pre-processing. The first one is to remove the uninteresting variability from the data. This in turn improves the signal to noise ratio (SNR) and the contrast to noise (CNR). Another aim of pre-processing data is to prepare the data for the statistical analysis. The basic framework of the pre-processing is slice timing correction, head motion correction, distortion correction, coregistration or normalization and the spatial and temporal smoothing.

In fMRI most of the data are acquired using two-dimensional pulse sequence in which one slice is acquired at a time due to the spatial gradient which in turn restricts the effects of an excitation pulse to a single slice within the brain. Sometimes interleaved slice acquisition is used in which scanner first collects the data of all the odd slices and than the even slices to avoid cross-slice excitation. Since each slices are acquired at different time, this creates a problem for interleaving sequences where the spatial consecutive slices are not acquired successively. As a result, the time series measured will have particular delay between the observed data and the experimental data. The

event related design have a more dire consequences of this compared to the blocked design since event related designs depends on a modeling of the timing of the experiments. Slice timing errors are corrected by using a technique called temporal interpolation during preprocessing. In this technique amplitude of MR at the onset of TR signal is estimated using the information of the nearby points. Either linear, spline or sinc function is used for the interpolation. It is not possible to get perfectly the missing information using interpolation. The accuracy of getting the missing data back depends on the variability of the data present in the experiment and the rate of sampling. This can be explained as the interpolation will be able to recover the data back if the data changes very rapidly over time compared to the sampling rate but if the variability of data is slow compared to the rate of sampling than this interpolation technique would be quite effective.

The distortion in data due to head motion is the major cause of worry in fMRI. In other word, in fMRI analyses it is assumed that every voxel represents a definite part of the brain and if the head moves than each voxel's time course is derived from more than one brain location. Head motion also affects the activation time. When a long TR imaging sequence is used for imaging head motion could lead to an error in estimating the activation. This can be seen in the Figure 4.1. Besides this the head motion can also lead to the changes in the pattern of spin excitation in the brain. Lastly the head motion may cause the loss of data on the edge of the imaging volume. The head motion can be prevented in the number of ways. Head restraints like bite bars, masks, padding or tapping can be used. The head motion correction can also be done by aligning successive image volumes in the time series which is called coregistration.

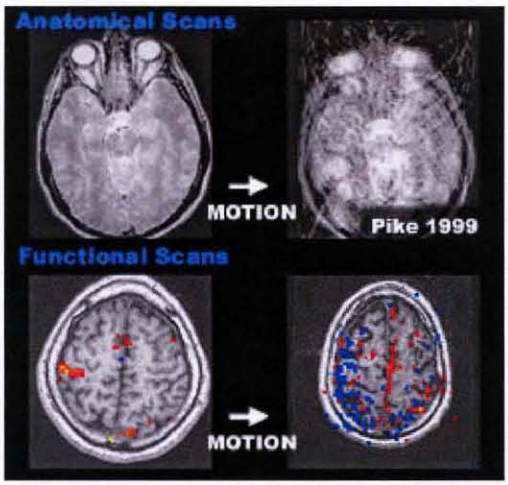

Figure 4.1 Head motion affecting the activation map during fMRI response.
Source: http://www.sunnybrook.ca/files/RESH_FMRI_headmotion.jpg

Since the brain is the same in every image of the time series, a transformation called the rigid body transformation can be used. The assumption behind using this transformation is that the shape and size of the two objects that are being co-registered are identical and the one can be superimposed on the other by using the three translations and three rotations. The other challenge while correcting the head motion is that the head of the subject might move in any direction and hence to overcome these realignment algorithms uses the iterative approach for motion correction. After this it is necessary to resample the original data without head motion. This is accomplished using spatial interpolation, where signal intensity points are considered in two or three dimension unlike temporal interpolation where only point in one dimension is considered.

The fMRI data was also pre-processed for the distortion correction in which the functional images suffer from the geometric or intensity distortion. This distortion occurs due to the inhomogeneities in the static field or excitation field. The former causes the geometric distortion of the image and may also sometime cause a loss of signal while the latter is responsible for the intensity distortion. The non homogeneity can be corrected by

using a shimming coil or alternatively approach called magnetic field mapping should be used. In this approach, by obtaining two images of signal phase having different echo time is used to create a field map of the main magnetic field. Hence the difference between the two images will be related to the field strength at a given location and if the field is uniform than the phase difference induced by the echo times will be same in all the voxels giving the resulting image as uniform gray. The goal of using the temporal and spatial correction was to make sure that each voxels contains data from a single brain region sampled at regular intervals throughout the time-series.

All data for this study were pre-processed using **A**nalysis of **F**unctional **N**euro **I**mages (AFNI) [19]. This tool was used for processing, analyzing and displaying the fMRI data. The fMRI data were pre-processed for slice timing errors and head motion using AFNI where the reconstructed images were corrected for motion using a rigid-body volume registration. The fMRI images from each of the finger-tapping scans were registered using 3D registration in AFNI [19].

4.3 Statistical Analysis

The two statistical analyses were carried out on the data set for this study. The first one was the cross-correlation method [15]. The correlation coefficient indicates how well the two variables (i.e., input and output) are related. It shows the degree of relation between them. It is given by the following equation. Here r_{xy} is called the correlation coefficient. x is the mean of the sample and y is the mean of another sample. While s_x and s_y are the standard deviation of the two samples that are being considered.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - x)(y_i - y)}{(n-1)S_x S_y} \tag{4.1}$$

The cross correlation was also carried out on the data. It can also be referred as covariance. Generally when there is a need to know the unknown signal from the known signal and see how the unknown signal is related to the known signal cross relation is performed.

Voxel-wise correlation of signal intensity time course with a boxcar reference function was used to determine activation using a threshold of 0.5 for the correlation coefficient ($P<.001$; after Bonferroni correction for multiple comparisons). Voxels that passed the threshold, during the finger-tapping task were considered as belonging to the sensorimotor cortex and stored for further data analysis on a subject-wise basis.

System Identification methods were used to build mathematical models to explain the dynamical nature of the fMRI response. Parameters from each of the models were adjusted to maximize the goodness of fit with the fMRI response. Parametric Identification Methods were used to estimate parameters in given model structures. Values of the Parameters that gave the best agreement between the model's (stimulated or predicted) output and the measured one were noted and stored for further analysis. By modeling the output fMRI response with a number of different system identification models, comparison between the various models was also preformed. The linear parametric models [12] that have been used included the Autoregressive (ARX) Model [12], Autoregressive Moving Average (ARMAX) Model [12], Box-Jenkins (BJ) model [12], Prediction Error Model (PEM) [12] and Instrumental Variable (IV) Model [12].

All voxels from the whole brain including the sensorimotor cortex were used for comparing and building input-output models using each of the five parametric models described earlier. We used the system identification tool box of MATLAB (Mathworks, MA) for data analysis.

After the model parameters for each of the voxel time-series was obtained using each of the models, an estimated response was generated using the corresponding parameters. Cross correlation was then calculated between every voxel time-series, measured response and the corresponding predicted response obtained using the model parameter. This was done for every subject using each of the five different models. Pair-wise correlation values were then obtained between the five models for each of the subjects. The computed cross correlation between the actual voxel and the predicted response was used to determine differences between models.

In fMRI data usually the temporal drifts are seen and these drifts might cause pronounced autocorrelation in the residual noise process. This would further complicate the signal detection and analysis that need to be carried out on the data set. Hence, it is necessary to remove these drifts from the data set before doing any kind of analysis. The method used for this purpose is Detrending. There are many different types of Detrending methods like non linear running lines and spline Detrending. Here in this study simple linear Detrending was done where the mean from the every voxel time series is removed or in other words it can be said that linear trend is removed from the data.

The other statistical test that was used was the t-test [15]. In this test null hypothesis that the mean of the two samples are true is tested. T-test is characterized on the basis of the mean, standard deviation and the number of the time points in a given two

data sets. Hence using this test it is possible to find if the mean are distinct under the assumption that the distribution of both the sample is normal. To perform this test researchers use the following equation. Here X_1 and X_2 are the mean of the given samples and S_1 and S_2 the standard deviation of the given two samples. The resulting t value is converted to the probability based on the degree of freedom (df). This term is defined as the number of time points minus 1.

$$t = \frac{X1 - X2}{S_{x1-x2}} where S_{x1-x2} = \sqrt{(s_1^2 + s_2^2)/n} \tag{4.2}$$

T-test was carried out for this study to check for the significant differences between the models. To check the consistency of the transfer function of the models, the transfer function of the high correlated voxels was used to predict the response of other high correlated voxels. This was carried out for every subject.

To test the reliability of the prediction of all the methods, a statistical test was done whereby 100 voxel time series data sets were randomly chosen. Each of the voxel time courses obtained from one of the models was fitted with each of the remaining four models and the correlation of the fit was obtained for all the methods. This was performed for all the 100 voxels time series. A t-test was then used for the analysis between the five methods. A significance value $P<0.01$ was used to identify differences between the two distributions.

4.4 Akaike's Theory

It is necessary to determine the order of the model to be used to have optimal results for the study. It can be determined using many different theories like Akaike's information criterion (AIC) [20, 21], Schwarz information criterion (SIC) Hannan-Quinn criterion

(HQC) , final prediction error (FPE) , and Bayesian information criterion (BIC). The AIC was used in this study for determining the order of the model.

$$AIC_p = -2T[l_n(\sigma)^2] + 2p$$

The order and delay were chosen automatically for each voxel time series using the basis of the Akaike's information theoretic criterion [12] for all the models. All five parametric models used the same order and different delays. For all the five parametric methods, second order model was used while the delay was allowed to vary to optimize the correlation values corresponding to the model. The consistency of all the five models was analyzed between the subjects as well as within subjects.

Thus, the data obtained were pre-processed for motion correction, slice timing correction and than the five different models were used for building the model for fMRI response. The various statistical analyses were carried out to have a better understanding for the results obtained.

CHAPTER 5

RESULTS

Motion correction was performed for all subjects and for each of the scans. No significant motion was found across any of the runs. The mean and standard deviation was found to be about 0.0106±0.0715. Table 5.1 shows the mean and the standard deviation of motion correction for each subject for the six motion parameters.

Table 5.1 Mean and the Standard Deviation of the Motion Parameters for each Subjects

Subjects	X	Y	Z	XY	YZ	ZX
Subject1	0.0413 ±0.0281	0.0392 ±0.0173	-0.0735 ±0.0632	0.0729 ±0.0421	-0.1201 ±0.0629	0.1831 ±0.0282
Subject2	-0.0676 ±0.0738	-0.1789 ±0.0938	0.2008 ±0.0582	0.0753 ±0.0582	0.1932 ±0.0797	0.3884 ±0.2105
Subject3	-0.1193 ±0.0577	0.0362 ±0.0246	-0.0015 ±0.0141	-0.0167 ±0.0156	-0.1978 ±0.111	-0.1138 ±0.0751
Subject4	-0.0831 ±0.0621	-0.0282 ±0.0411	0.0371 ±0.0172	0.0269 ±0.0183	0.0729 ±0.0821	0.1731 ±0.0328
Subject5	-0.0929 ±0.0428	0.0636 ±0.0314	0.0613 ±0.0244	-0.0632 ±0.0479	-0.1327 ±0.1053	-0.1883 ±0.0849
Subject6	-0.1084 ±0.0601	-0.2793 ±0.1482	0.0547 ±0.0304	0.5801 ±0.2034	-0.0020 ±0.0899	0.0652 ±0.2268

An idealized reference waveform representing the "ON/OFF" cycle was correlated with every voxel in the brain. All voxels in the brain that had a correlation coefficient of 0.5 or more were considered active. A correlation coefficient of 0.5 represented a significance level of $P<0.0001$. These activated voxels were mostly clustered together in the left sensorimotor, right sensorimotor, and supplementary motor regions. A representative image with sensorimotor cortex and its associated cortex, activated is shown in figure 5.1.

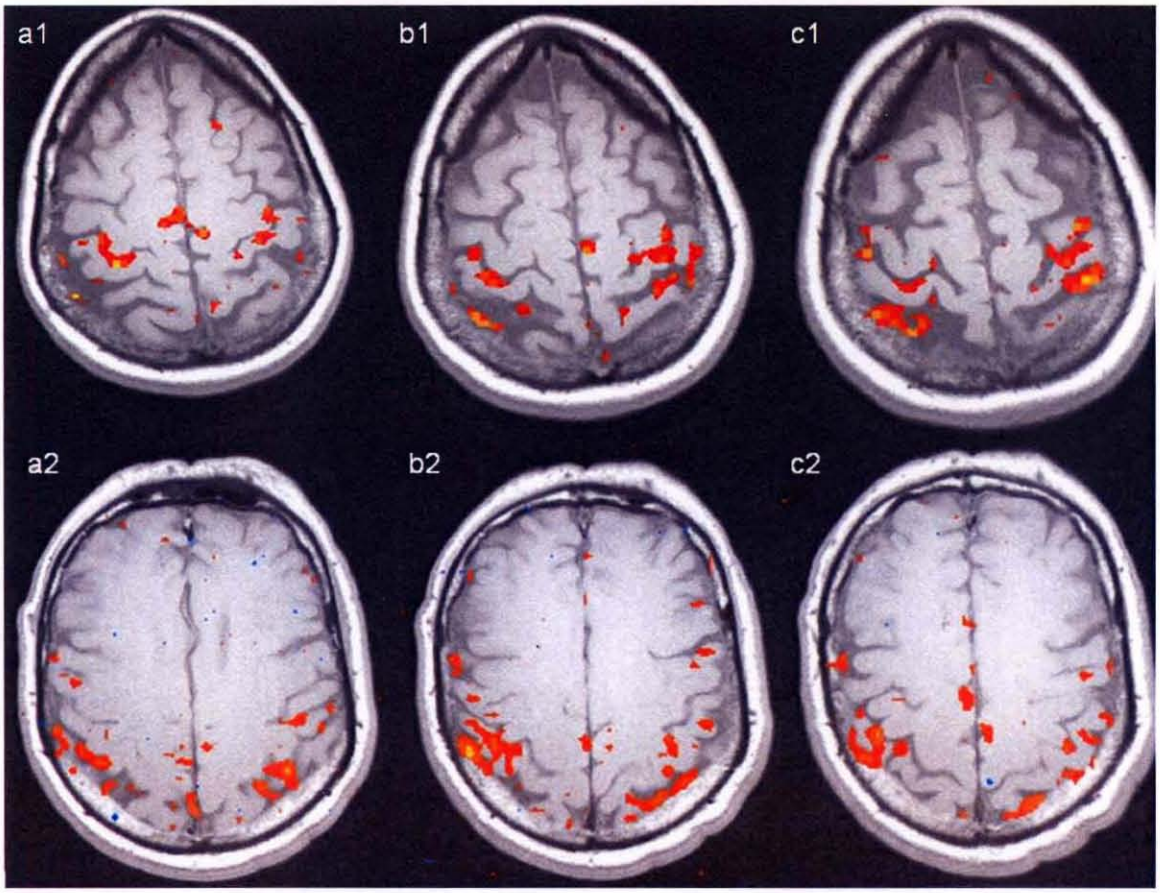

Figure 5.1 The brain image showing the activation for the bilateral finger tapping. a1, b1 and c1 are the images from the subject 1 showing the activation in the sensorimotor cortex of the brain. Similarly a2, b2, and c2 shows the activation for subject2.

A representative voxel time course along with the idealized "ON/OFF" reference waveform is shown in Figure 5.2. The time course of the BOLD signal response had changed by about 1.327% during bilateral finger tapping. All voxels from the sensorimotor cortex were built as models for each of the subjects. The output data sets were then fitted to each of the five models.

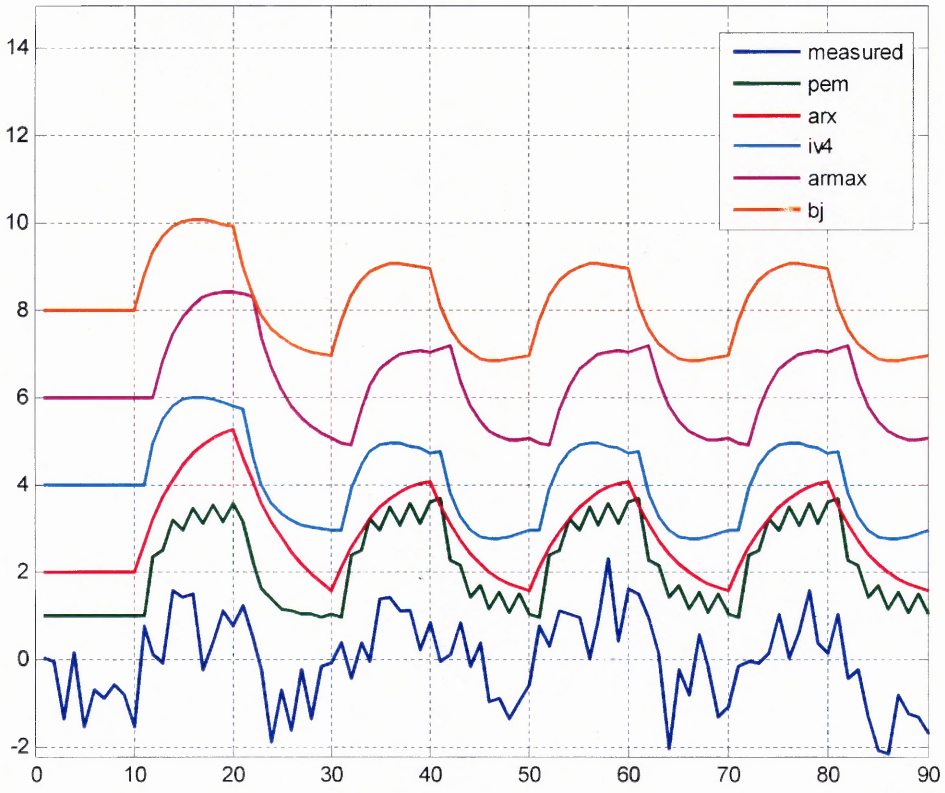

Figure 5.2 The measured response and the modeled fMRI response using all the five model structure for a representative voxel of one of the subject.

A second order model was determined to be optimal for all the data sets using Akaike's information criterion. However, delay values of 1, 0, 1, 2 and 0 were selected for PEM model, ARX model, IV model, ARMAX model and BJ model respectively to get the optimal fit between the measured and the predicted output of the fMRI response. These values of the order and delays were used for every voxel time series in every subject. Figure 5.2 shows the modeled response compared to the actual data using each of the five models. The correlation value for PEM, ARX, IV, ARMAX and BJ models with the measured data from the fMRI response is 0.71, 0.67, 0.66, 0.56 and 0.67 respectively. This high value of the correlation demonstrates that all the models could

predict the output quite reliably. In the present study, the PEM model had the best fit with the measured fMRI data. The estimated fit from other models also gave statistically significant results. The results indicate the PEM model to perform relatively better as it gave best fit compared to the other models.

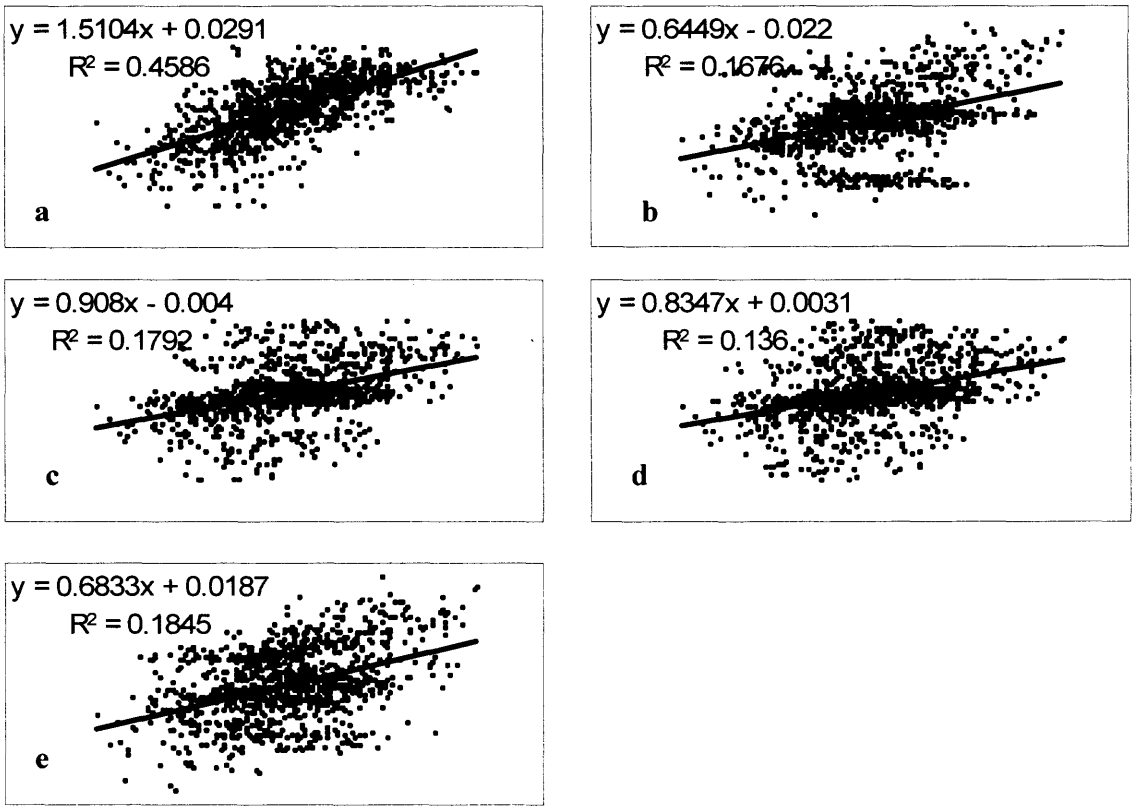

Figure 5.3 The scatter plot of the correlation coefficient of the measured response from the fMRI compared to the modeled response from all the methods for one of the subject is shown. a-PEM, b-ARX, c-IV, d-ARMAX and e-BJ.

Figure 5.3 shows a scatter plot for a representative subject illustrating the fit obtained from the PEM method with the actual voxel time series for all the voxels in the brain. A correlation coefficient of 0.63 was found. A similar analysis was performed for all the other methods for all the subjects. Using all the voxels from the brain, in each of the subject, PEM method was found to have the greatest correlation coefficient.

The reliability of the prediction of PEM model was checked by the method described above where a statistical t-test was done whereby 100 voxel time series data sets were randomly chosen. Each of the voxel time courses were fitted with the output of each of the five models and the correlation of the fit were obtained. The result of t-test showed that the average p-value $P < 0.0001$, which shows that there was a significant difference between PEM and the other models.

To test the predictability of the transfer function, the output response was predicted by using the transfer function of one voxel to get the response for another voxel. Correlation was then obtained between the predicted response and the actual data to quantify the match. A significant correlation of 0.7745 ± 0.1741 was obtained using the PEM method. Using the other four methods like ARX, ARMAX IV4 and BJ, resulted in correlation values of 0.7513 ± 0.1741, 0.7488 ± 0.1821, 0.7587 ± 0.1476 and 0.7689 ± 0.1157 respectively. The results of this test showed that there was no significant difference in the transfer function used to predict an output response for different models.

Qualitative comparison between the five methods were calculated by performing a pair wise correlation between predicted response obtained from each of the method with every other method. Figure 5.4 shows the pair wise correlation matrix obtained for a representative subject. As can be seen from the Figure 5.4, all the five methods had significant correlation with the other methods. The mean correlation value was found to be 0.43 with a minimum and maximum value of 0.13 and 0.76, respectively. Across all the subjects, the mean correlation value was found to be 0.49.

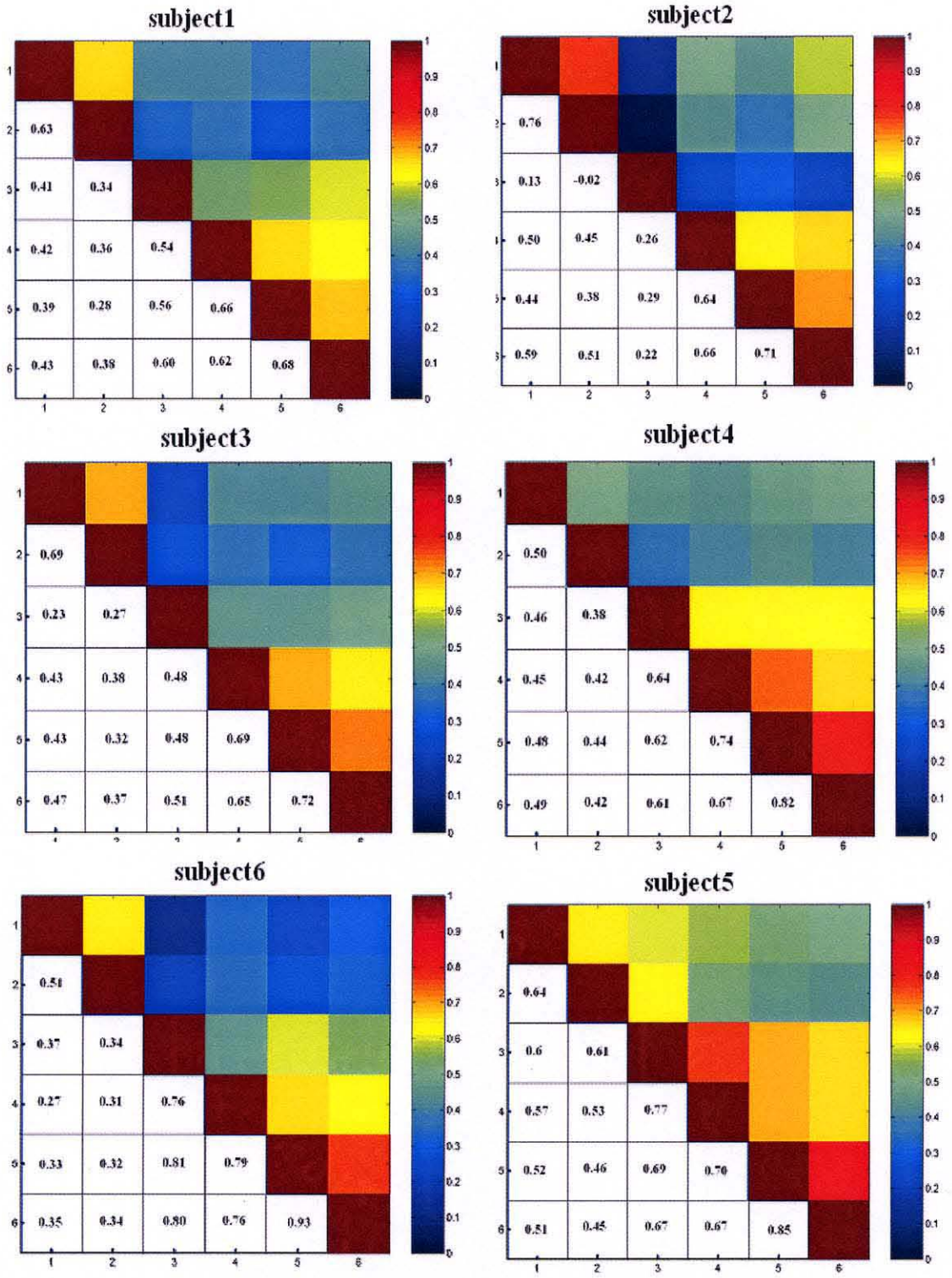

Figure 5.4 Pair-wise cross correlation coefficient for the predicted response from the fMRI response and models as well between the models for all the subjects. In the figure the 1 represents predicted response from the fMRI, 2 represents PEM model response, 3 represents ARX model response, 4 represents IV model response, 5 represents ARMAX model response and 6 represents BJ model response.

CHAPTER 6

DISCUSSION AND CONCLUSION

The exact relationship between the input and the fMRI response is still not exactly understood. While most statistical tests and analysis assume a linear relationship, a few studies have shown the fMRI signal to be non-linear. The fMRI response used in this study using a block design methodology in the region like visual and motor cortex of the brain, can be treated as linear. This was further validated using parametric models that gave accurate results of the estimation of the fMRI response. This kind of analysis simplifies the understanding of the input-output fMRI response. Although, there may well exist a non-linear component of the response, for this study it does not appear to be significant.

For this study, bilateral finger tapping was used for a number of reasons. Bilateral tapping is one of the simplest and most commonly used paradigms currently being used in fMRI. Bilateral finger tapping produces robust results with significant signal changes. Further, number parametric modulations of fMRI signal using tapping rate [22, 23], tapping duration [24, 25], etc has been established by number of groups [26, 27]. Recently, Glover used different duration of bilateral finger tapping to characterize and demonstrate the variability of the impulse response function. As a consequence, in this study stimulus duration of at least 10 seconds "ON" period was used to optimize the estimation of the impulse response function. Although bilateral finger tapping during block design was used, other kinds of stimulus need to be tested and modified as required for the models used. For other stimulus including event related paradigms, those kind of

stimulus may be non-linear methods would be better suited. The bilateral finger tapping was used as it is most simple and robust signal. It gives good signal change and easily reproducible.

Recently, ARX method has been applied by Baraldi and colleagues [13] to estimate and predict the fMRI response. Besides, they predicted various parameters including time to onset, peak time and amplitude of the BOLD response. They found that ARX model was quite effective in predicting the response. The results from this study were similar to the results obtained by Baraldi and colleagues [13]. However, several additional models were tested in this study. Though,, ARX gave quite good approximation to the fMRI response, results of this study indicate that the PEM model performs better compared to the ARX model.

The temporal characteristics of fMRI noise have been typically assumed to be a white Gaussian noise by most statistical methods. The noise has also been assumed as temporally independent. It has been earlier shown that the fMRI noise to be temporally correlated and is not random in nature. Most statistical methods assume the noise to be independent white Gaussian noise. It is thus possible that the various system identification methods here other than PEM have not accounted this property of the noise and have tried to model it along with the signal resulting in a prediction less than that obtained using PEM.

ARX is the simplest model defined in the parametric models. Models like ARMAX, BJ and IV try to minimize the noise and also make the system error free. This may be the reason for ARX model being not consistent between the subjects. For the

ARX model, good signal to noise ratio is needed which plays an important role in predicting the fMRI response of the voxel.

PEM estimates the maximum likelihood of Gaussian distribution while minimizing the prediction error. Using PEM, the estimated transfer function is similar to a Gaussian distribution and the influence of noise on the estimation of impulse response function is minimized. Since fMRI response of a bilateral finger is similar to Gaussian distribution PEM predicts the response more accurately than other models.

A second order model was determined for this study. For this study, young healthy subjects were used using identical image sequence parameters. If the image sequence were changed, that would change for example the flow weighting in the signal, the signal model and the optimal parameters may become different. In certain clinical cases like epilepsy, stroke or tumor where the underlying neurovasculature has been altered, it is possible that the model parameters will be different. Clearly, the order selection is dependent on a number of factors including the underlying neurovasculature, the sampling rate, region of interest, pulse sequence used.

Five different models were used such that the model giving the best fit could be determined. The primary advantage of using multiple approaches is that it can be used as a validation method. Further, limitation of a method or methods can be determined compared to other methods. Also, the effect of SNR and noise parameters can be estimated and its effect on each of the methods can be determined in case some criteria cannot be optimized.

Recently, Ward and Mazahari [28] treated the fMRI experiment as communication channel system and predicted the input stimulus from the response

obtained from the fMRI and the hemodynamic impulse response function (IRF) which gets delayed and blurred due to noise present during the fMRI experiment. The input was predicted by using the Kalman filter. Clearly, the accuracy of the transfer function affects the accuracy of the predicted input response. One immediate application from this study is that, the efficacy of such methods could be tested by using the five transfer functions used in this study. Since both the transfer function and response are known it is possible to calculate the stimulus using these models.

As fMRI becomes more prevalent in the clinical setting, reliable methods for detecting fMRI signal becomes even more important. The knowledge of transfer function will be useful in finding the optimal stimulus that can be presented during an experiment to elicit the maximum response. Also, once the transfer function for each voxel in a particular region of the brain is calculated, it is possible to predict the response from that voxel for other input stimulus. Hence, a comparative analysis can be done on the basis of the transfer function for the two different regions of the brain. In several clinical cases abnormal pattern of detection could be due to a variety of reasons, including subjects' inability to perform the task, or changes to patients' vascular or neuronal effects that would consequently affect the estimated transfer function. Using the proposed method, it is possible that detection of abnormal transfer function can be estimated and differentiated from the subjects being able to perform the task.

REFERENCES

[1] Bandettini PA, Wong EC, Hinks RS, Tikofsky RS, Hyde JS. Time course EPI of human brain function during task activation. Magn Reson Med. 1992 Jun 25, 390-397.

[2] Belliveau JW, Kennedy DN, McKinstry RC, et al. Functional mapping of the human visual cortex by magnetic resonance imaging. Science1991 Nov 1,716 -719

[3] Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weiskoff RM, Poncelet BP, Kennedy DN, Hoppel BE, Cohen MS, Turner R, Rosen B, Brady TJ (1992) Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci, 5675-5679.

[4] Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc Natl Acad Sci U S A. 1992 Jul 1 5951-5955

[5] DeYoe EA, Bandettini P, Neitz J, Miller D, Winans P. Functional magnetic resonance imaging (FMRI) of the human brain. J Neurosci Methods. 1994 Oct 5,171-187.

[6] Ogawa S, Lee TM, Nayak AS, Glynn P. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magn Reson Med. 1990 Apr 14, 68-78.

[7] Lennart L. System Identification-theory for the user. Englewood Cliffs, New Jersey, 1987

[8] Lennart L. System Identification toolbox 7 User's guide. Natick, MA, March 2007

[9] Boynton GM, Engel SA, Glover GH, Heeger DJ. Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci. 1996 Jul 1, 4207-4221.

[10] Glover GH. Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage. 1999 Apr 9, 416-429.

[11] Cohen MS. Parametric analysis of fMRI data using linear systems methods. Neuroimage. 1997 Aug 6, 93-103.

[12] Lyons RG. Understanding Digital Signal Processing. Addison Weseley Longman 1997. (Reading, MA)

[13] Zierler KL. Equations for measuring blood flow by external monitoring of radioisotopes. Circ Res. 1965 Apr 16, 309-321.

[14] Baraldi P, Manginelli AA, Maieron M, Liberati D, Porro CA An ARX modelbased approach to trial by trial identification of fMRI-BOLD responses. Neuroimage. 2007 Aug 1, 189-201

[15] Huettel SA, Song AW, McCarthy G, Functional Magnetic Resonance Imaging. Sunderland, MA, 2004.

[16] Bushberg JT, Seibert JA, Leidholdt EM, Boone JM The Essential Physics of Medical Imaging Philadelphia, PA, 2002.

[17] Vazquez AL, Noll DC Nonlinear aspects of the BOLD response in functional MRI Neuroimage. 1998 Feb 7,108-118

[18] Thomas D, Olivier F. Using nonlinear models in fMRI data analysis: Model selection and activation detection neuroimage.2006. March 03.

[19] Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res. 1996 Jun 29, 162-173.

[20] Akaike, H. Fitting autoregressive models for prediction. Annals of the Institute of Statistical Mathematics, 1969, March 21, 243-247.

[21] Akaike, H. Information theory and an extension of the maximum likelihood principle. 2nd International Symposium on Information Theory, B. N. Petrov and F. Csaki (eds.), Akademiai Kiado, Budapest, 1973, 267-281.

[22] Rao SM, Bandettini PA, Binder JR, Bobholz JA, Hammeke TA, Stein EA, Hyde JS. Relationship between finger movement rate and functional magnetic resonance signal change in human primary motor cortex. J Cereb Blood Flow Metab. 1996 Nov;16(6):1250-4. J Cereb Blood Flow Metab. 1996 Nov16,1250-1254.

[23] Robson MD, Dorosz JL, Gore JC. Measurements of the temporal fMRI response of the human auditory cortex to trains of tones. Neuroimage. 1998 Apr 7,185-198.

[24] Temporal and intensity coding of pain in human cortex. J Neurophysiol. 1998 Dec 8 Porro CA, Cettolo V, Francescato MP, Baraldi P., 3312-3320.

[25] Birn RM, Bandettini PA. The effect of stimulus duty cycle and "off" duration on BOLD response linearity. Neuroimage. 2005 Aug 1, 70-82.

[26] Birn RM, Cox RW, Bandettini PA. Detection versus estimation in event-related fMRI: choosing the optimal stimulus timing. Neuroimage. 2002 Jan15, 252-264.

[27] Bandettini PA, Jesmanowicz A, Wong EC, Hyde JS. Processing strategies for time-course data sets in functional MRI of the human brain. Magn Reson Med. 1993 Aug 30, 161-173.

[28] Douglas W Yousef M. State-space estimation of the input stimulus function using the Kalman filter: A communication system model for fMRI experiments Journal of Neuroscience Methods. 2006 May 24, 271-278.