BAYESIAN MODELING AND INFERENCE FOR FUNCTIONAL MAGNETIC RESONANCE IMAGING OF THE VISUAL CORTEX

by

Meltem Sevgi

B.S, Physics Engineering, Istanbul Technical University, 2009

Submitted to the Institute of Biomedical Engineering
in partial fulfillment of the requirements
for the degree of
Master of Science
in
Biomedical Engineering

Boğaziçi University
2011

ACKNOWLEDGMENTS

It is a pleasure for me to offer my gratitude to my supervisor, Prof. Cengizhan Öztürk, for his encouragement and guidance. He always listens and finds a solution when things go down. It has been a great experience to be a member of his medical imaging group.

Many thanks go in particular to Prof. Ahmet Ademoğlu. He patiently taught me and tirelessly answered to my questions. I would like to acknowladge him for his important contribution for this thesis, and also for our roundtable discussions on science and philosophy.

I am indebted to my friends Esin Karahan and Adil Deniz Duru for sharing their knowledge and skills on neuroimaging. They are the best engineers that a neuroscience group would need.

This thesis would not have been possible without the collaboration of Prof. Tamer Demiralp and his neuroscience group in Istanbul University. I am grateful for their contribution in this study.

Last but not the least, I would like to thank to my family. I am so lucky for having such a family with patience and support throughout my life.

ABSTRACT

BAYESIAN MODELING AND INFERENCE FOR FUNCTIONAL MAGNETIC RESONANCE IMAGING OF THE VISUAL CORTEX

For the most effective use of functional magnetic resonance imaging (fMRI), mapping the brain signals to a statistically valid map is crucial. The common approach to create a statistic at each voxel is applying the frequentist or the classical statistics. However, there are many challenges raised by the use of classical statistics to test the functional data such as the multiple comparison problem, and the limitation in the interpretation of the parameters. As an alternative, a Bayesian approach can be used to assess the data based on the posterior probability distributions of the parameters. In this study, the power of Bayesian inference was compared against classical inference in random effect analyses: A group data collected from visually stimulated volunteers was assessed following a simulation study. In order to assess the results of the statistical inference for the group level, the variation of the effect sizes with respect to stimulus frequency was used. A comparison was performed between the change in the effect sizes of lateral geniculate nuclei (LGN) and primary visual area (V1) during graded visual stimulation by using the posteror probability maps (PPMs) with an effect size threshold of zero. This comparison became possible with the fact that once we had the posterior probabilities the activity in LGN was able to be visualized by changing the effect size threshold and without decreasing the significancy threshold, which is not possible to achieve with classical inference where the data is tested against the null hypothesis. Despite of the small magnitude of activation in LGN we could show the connectivity between V1 and LGN and the differences in response characteristics during graded visual stimulation.

Keywords: fMRI, Bayesian inference, visual cortex, LGN, PPMs.

ÖZET

GÖRSEL KORTEKSİN FONKSİYONEL MANYETİK REZONANS GÖRÜNTÜLEMESİ İÇİN BAYES MODELLEME VE ÇIKARIMI

Fonksiyonel manyetik görüntülemenin en etkin kullanımı açısından beyin sinyallerinin istatistiki bir haritayla gösterimi büyük önem taşır. Her bir vokselde bir istatistik oluşturmak için kullanılan genel yaklaşım klasik istatistiktir. Ancak, klasik istatistiğin kullanımında bir çok zorluk mevcuttur, örneğin, çoklu karşılaştırma problemi veya parametrelerin yorumlanmasındaki sınırlar gibi. Bir alternatif olarak, parametrelerin şartlı olasılık dağılımlarına dayanan Bayesçi yaklaşım verilerin değerlendirmesinde kullanılabilir. Bu çalışmada Bayes çıkarımının gücü klasik çıkarımla rasgele etki analizi üzerinden kıyaslandı: Bir simülasyon çalışmasının ardından, görsel olarak uyarılmış deneklerden toplanan grup datası değerlendirildi. Grup seviyesindeki istatistiksel çıkarımın sonuçlarını değerlendirmek için etki büyüklüğünün uyaran frekansıyla değişimi kullanıldı. Etki büyüklüğü eşiği sıfır olan şartlı olasılık haritaları kullanılarak lateral genikülat çekirdeklerin (LGÇ) ve birincil görme korteksinin etki büyüklüklerindeki değişim karşılaştırıldı. Bu karşılaştırmanın yapılabilmesi şartlı olasılıklar bilindiğinde anlamlılık seviyesini düşürmeden etki büyüklüğü eşiğinin değiştirilerek LGÇ'deki aktivitenin görüntülenebilmesi ile mümkün hale geldi; bu durumu verinin boş hipoteze karşı test edildiği klasik istatistikle yapabilmek mümkün değildir. LGÇ'deki aktivite değerinin küçük olmasına karşın birincil görme korteksi ile arasındaki 'bağlanırlığı' ve artan görsel uyaranlar sırasında bu iki bölge arasındaki cevap karakteristiğinin farklarını gösterebildik.

Anahtar Sözcükler: fMRG, Bayesçi çıkarım, görsel korteks, LGÇ, PPM.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS

T_2	Transverse relaxation time constant
T_1	Longitudinal relaxation time constant
ϵ	Error term
σ^2	Variance
N	Normal distribution
β	Model parameters
α	Height threshold for t-statistics
n	Number of t-statistics
V	Number of voxels
λ	Model hyperparameters
Q	Covariance components
Φ	Cumulative Normal Distribution
γ	Precision (inverse variance)

LIST OF ABBREVIATIONS

EEG	Electroencephalography
MEG	Magnetoencephalography
NIRS	Near Infrared Spectroscopy
fMRI	Functional Magnetic Resonance Imaging
PET	Positron Emission Tomography
BOLD	Blood Oxygen Level Dependent
RFT	Magnetoencephalography
SPM	Statistical Parametric Mapping
PPM	Posterior Probability Maps
MEG	Cerebral blood Flow
$CMRO_2$	Cerebral Metabolic Rate of Oxygen consumption
CBV	Cerebral Blood Volume
HRF	hemodynamic Response Function
FIR	Finite Impulse Response
EPI	Echo Planar Imaging
TE	Echo Time
OLS	Ordinary Least Square
FWE	Family Wise Error
FWHM	Fool Width Half Maximum
EC	Euler Characteristic
PEB	Parametric Empirical Bayes
EM	Expectation Maximization
ReML	Restricted Maximum Likelihood
TR	Time of Repetition
FA	Magnetoencephalography
MNI	Montreal Neurological Institute
ROI	Region of Interest
LGN	Lateral Geniculate Nuclei

V1	Primary Visual Area
GM	Gray Matter
WM	White Matter

1. Introduction

1.1 Motivation and Objectives

Understanding the human brain function is possible with different functional brain imaging technics. While the methods such as single unit, patch clamp, light microscopy can provide information about neuronal dynamics directly from neurons and dendrites, it is possible to map the brain function in terms of larger functional units. For example, electrical potentials and magnetic fields in the brain can be recorded with electroencephalography (EEG) and magnetoencephalography (MEG), respectively. They have good temporal resolution but poor spatial resolution. There are optical imaging technics such as near infrared spectroscopy (NIRS) having a limited spatial resolution because of the light scattering by the skull. In contrast, images acquired with functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) have higher spatial resolution.

While structural MRI is used to distinguish different types of tissue, the goal of the functional MRI is relating the neuronal activity to a certain type of experimental stimuli via the so called blood oxygen level dependent(BOLD) signal. Today, fMRI is widely used in many clinical applications, e.g. in neurology, psychiatry, pharmacology, study of emotion, and also in many research applications of sensorimotor, auditory, language systems, short-term and long-term memory, etc. The data hierarchy of an fMRI experiment is as follows: Subjects, sessions, runs, single run, volume, slices, voxel. Each run has a collection of data as a time series of a volume.

For the most effective use of fMRI, mapping the brain signals to a statistically valid map is crucial. The common approach to create a statistic at each voxel is the frequentist or the classical statistics. However, there are many challenges raised by the use of classical statistics to test the functional data such as the multiple comparison problem, and the limitation in the interpretation of the parameters. As an alternative,

a Bayesian approach can be used to assess the data based on the posterior probability distributions of the activations.

First of all, we may ask questions about the absolute values of the parameters with Bayesian inference. It is available to infer about the probability that a parameter lies in a specified interval [1]. This is an obvious advantage over classical inference where the data can be tested only under the null hypothesis that the value of the parameter is zero. Therefore, the frequentist approach is limited to the rejection of the null hypothesis making it impossible to ask questions about the values of the parameters.

Secondly, the multiple comparison problem associated with classical inference can be overcome with Bayesian inference. The correction for the P-values with a Bonferroni correction will be too conservative because of the spatial correlation in the functional images where the number of independent observations is not equal to the number of voxels. Therefore, the random field theory is used to solve this problem. However, the underlying assumption of RFT is that the residual errors in the general linear model are continuous multivariate Gaussian distributions, which is not in real life, where the voxels might have different noise properties due to the brain anatomy (refer to section 3.3 for a detailed explanation). Hence, a Bayesian inference will circumvent such corrections by using directly the probability that a voxel is activated.

In this study, we aim to use a Bayesian approach to assess group data collected from 40 visually stimulated subjects, following a simulation study. Then we will compare and discuss the power of Bayesian inference against classical inference in random effect analyses. Statistical parametric mapping (SPM) is based on classical inference, hence it is a limited tool. We will use posterior probability maps (PPMs) to visualize the results from the Bayesian inference based on the posterior probability of activations.

1.2 Outline of the Thesis

The next chapter gives a brief overview about the physiological basis of fMRI, the so called BOLD effect, image acquisition with fMRI and also types of experimental paradigms. Because this thesis aims to compare different statistical methods to assess fMRI data, Chapter 3 gives an introduction about common statistical approaches used in fMRI data analysis and discusses many problems associated with the classical inference. Then, Chapter 4 introduces the Bayesian approach as an alternative statistical tool which is further explained with a simulation study. In Chapter 5, the methods used in the two-level analysis of a real large dataset are described. The results are visualized with posterior probability maps compared to statistical parametric maps, and also quantified with effect sizes for the whole brain as well as some specific regions in Chapter 6. Finally, the results are further discussed and compared with similar studies in Chapter 7.

2. Principles of MRI

2.1 The BOLD Signal

Blood has a high concentration of deoxyhemoglobin, and a low oxyhemoglobin concentration during the resting state. In contrast, brain has a high concentration of oxyhemoglobin, which reduces static magnetic field. Because deoxyhemoglobin is paramagnetic and oxyhemoglobin is diamagnetic, at the intersection of brain and blood is static magnetic field inhomogeneities, shortening T2*, which in turn reduces the signal intensity.

Cerebral blood flow (CBF) is a measure of oxygen delivery to the brain tissue. After a neuronal activation, there is an increase in the oxygen consumption so that increased CBF will increase the oxyhemoglobin concentration and reduce the deoxyhemoglobin concentration. Therefore, local magnetic field differences at the intersection of brain and blood will be less than the case of the resting state. It has been observed that reduced inhomogeneity provides enhanced local image intensity in the regions of neuronal activation [2]. This is the so called BOLD effect, which in real life is more complicated.

In an fMRI experiment the correlation between the stimulus and the stimulus-induced BOLD response is measured. It has been shown that the neural activity can be expressed in terms of a linear transform of fMRI response. Figure 2.1 shows the linear transform model for the Retinal-V1 pathway [3]. The fMRI signal is given with a linear transform of the neural response which is a nonlinear function of the stimulus. The fMRI signal can be modeled with the convolution of the stimulus S(t) and the hemodynamic response function (HRF):

$$BOLD(t) = HRF \cdot S(t) = \int h(t-u)S(u)\,du. \tag{2.1}$$

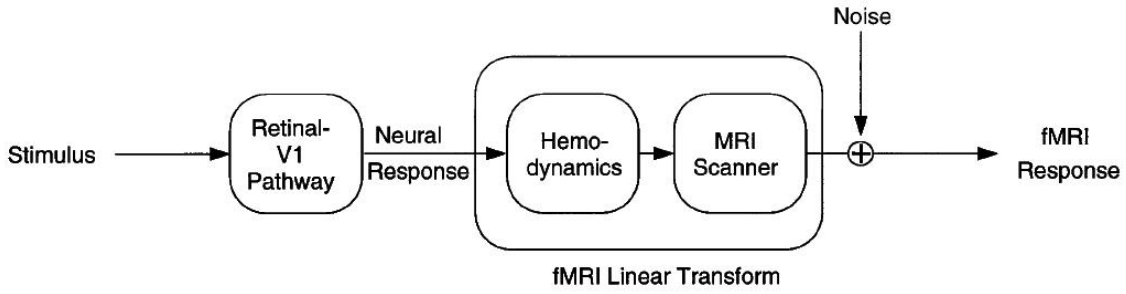

Figure 2.1 fMRI linear transform model for the Retinal-V1 pathway [3].

A typical hemodynamic response begins with a signal decrease, which was observed at high-field strengths [4]. This initial-dip is originated from the higher cerebral metabolic rate of oxygen consumption (CMRO2). Following this signal decrease, both the CBF and the CBV increases (Figure 2.2). CBF transports more oxygen to the activation site so that positive BOLD response is observed during 5-10 s. Afterwards, CBF and CMRO2 turn to their baseline levels faster than CBV which transports more deoxyhemoglobin resulting in the signal undershoot up to 30 s after the onset of the stimulus. There are many techniques used to model the HRF in the literature such as using the temporal basis functions [5], inverse logistic functions [6], smooth finite impulse response (FIR) filter [7], canonical HRF with time and dispersion derivatives [8], and the canonical SPM HRF [9]. Hemodynamic response varies between subjects and in different brain regions [10]. It is important to remember that using a wrong shape will reduce the sensitivity of the model and inference.

2.2 Echo-Planar Imaging

MR scanners use a series of radio frequency pulses, gradient waveforms, and data acquisition to manipulate the magnetization in order to produce the desired signal. The series of these events is called pulse sequence [11]. The common pulse sequence used for fMRI is echo-planar imaging (EPI). Basic EPI pulse sequences are gradient-echo EPI, spin-echo EPI, inversion-recovery EPI, single-shot EPI, and multishot EPI. A

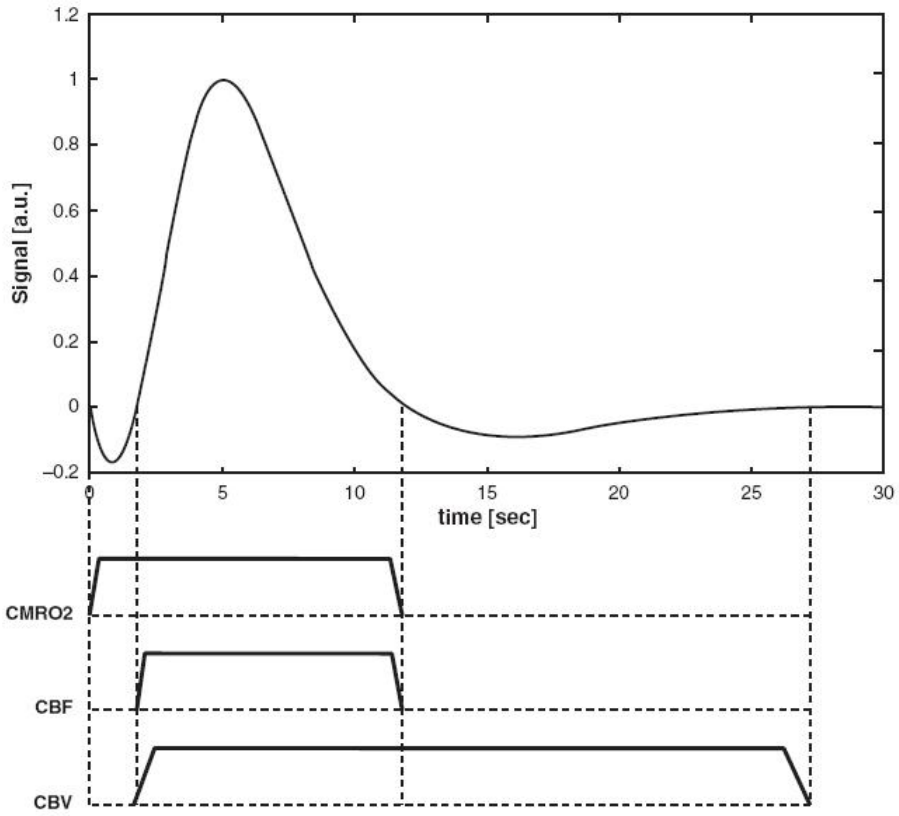

Figure 2.2 Haemodynamic response function has an negative initial dip, a positive BOLD response, and a negative undershoot [2].

representation of gradient-echo EPI sequence, which was used in this thesis as well, is shown in Figure 2.3. A selective excitation pulse produces an FID signal. The flip angle of the excitation pulse is set to 90°. Because each k-space line is acquired at different TE, the amplitude of the gradient echo signal is a function of TE in that echo train, n:

$$S(n) = S_0 \exp(-TE(n)/T2^*). \tag{2.2}$$

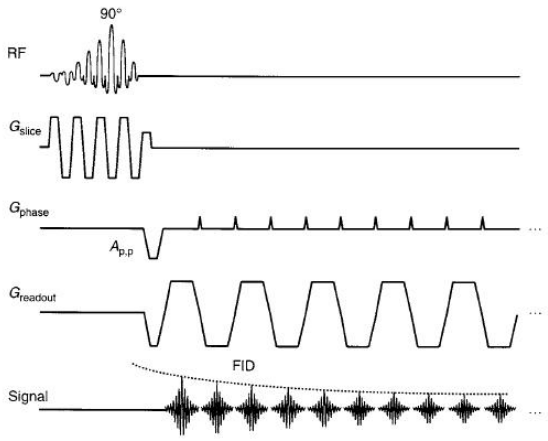

Figure 2.3 An example of a gradient-echo EPI pulse sequence [11].

2.3 Experimental Design

In order to test a research question an experiment must be set up properly. In an experimental design, there are independent and dependent variables. In an fMRI experiment BOLD signal change is used as the dependent variable. The hypothesis is tested by manipulating the independent variables and checking whether the BOLD signal is changed as well. This hypothesis testing enables the researcher to ask questions about heamodynamic activity, neuronal activity, or the psychological processes. First type of hypothesis is about heamodynamic activity which can be measured directly from the BOLD change such as the nonlinearity of the hemodynamic response [12, 13]. Secondly, neuronal activity might be the research question in an fMRI experiment. Biophysical modeling studies for integration of EEG and fMRI [14], or spontaneous fluctuations in neuronal activities can be questioned by transforming the BOLD signal. The third type of hypothesis address the psychological processes such as attention [15], memory [16], perception [17].

2.3.1 Blocked Designs

In a blocked design experiment an experimental condition is compared to a control condition so that the dependent measure is compared in each block condition.

These conditions might consist of two different tasks, or a series of on periods of activation followed by off periods, which are known as null-task blocks. The former design is called as alternating design. Simply, a block is a presentation of similar stimuli together. Duration of the blocks is very important and depends on the research question. For example, in a sustained attention experiment, a task may be composed of 90-sec blocks of continuous stimuli presentation [18]. Further, if the block length is very short (less than about 10 s) then the hemodynamic response cannot return to its baseline during the off period, so the BOLD signal amplitude will decrease [19].

2.3.2 Event-Related (Trial-Based) Designs

A block design would not be a proper design in a recognition experiment where the presentation of the familiar stimuli together would be meaningless. With an event-related design, it is possible to randomize the presentation of familiar and novel stimuli. It is assumed that the neural activity occurs for short and discrete intervals in an event-related designed experiment. No assumptions are made about the shape of the hemodynamic response. Trials are separated with an interstimulus interval.

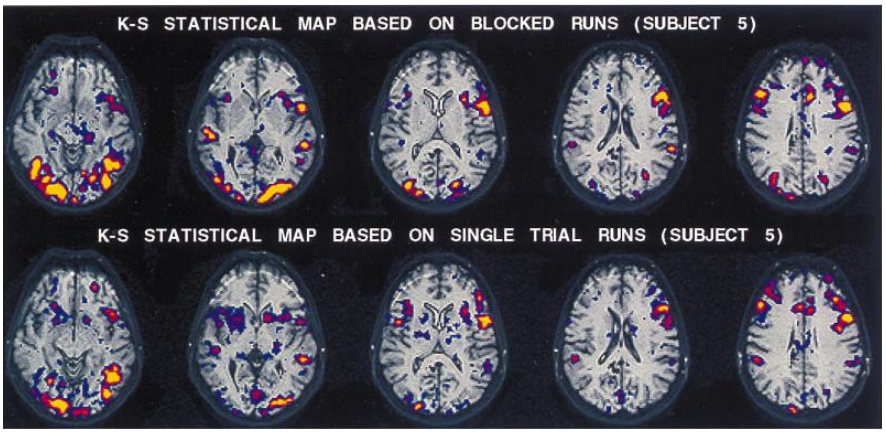

Figure 2.4 A study on block-design and event-related design shows that there are slightly more active voxels observed in the block-design [20].

2.3.3 Mixed Designs

In event-related designs, each trial produces a neural activity as a transient time course. The neural activity is sustained throughout the performance of a task, so event-related designs ignore the âsustainedâ activity. On the other hand, blocked designs would confound the sustained activity and trial-related activity [21]. Embedding the trials in a task block and using the control block alternately form the mixed design (Figure 2.5) allows examining the sustained processes in the brain by separating transient activity from sustained activity.

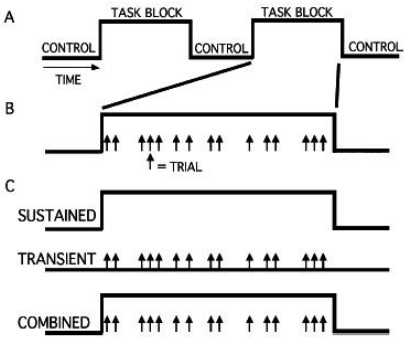

Figure 2.5 (A) Mixed design. (B) Trials with different intervals are embedded in a task block. (C) In the combined condition, sustained activity and transient activity are superimposed [21].

3. fMRI Data Analysis

The purpose of fMRI data analysis is to identify the voxels which show higher intensity during the period where stimulation was applied. An active voxel shows an intensity level significantly higher than the noise level. Preprocessing steps reduce unwanted noise and prepare the data for further statistics. Finally, the significantly active voxels are displayed on statistical maps with a color scale indicating the degree of significance.

3.1 Preprocessing

Realignment: During an fMRI session, subjects move their heads causing variations in the time series of a voxel to refer to another point in the brain over time. A realignment routine is applied using a least squares approach and a rigid body spatial transformation with 6 parameters (three parameters for translation, three parameters for rotation). Commonly the first image is selected as the reference scan, and the rest is realigned accordingly.

Registration: Functional data have low-resolution and little anatomical contrast but anatomical image has a higher resolution with many details. To answer how the functional data correspond to the underlying anatomy, functional-structural coregistration is applied by minimizing the cost function with a rigid body transformation [22]. Echo planar images might have distortions along a certain axis. Then a nine-parameter rigid-body transformation is used. If further distortions exist, some warping algorithms can be applied.

Segmentation: MR images of the brain can be segmented into three tissue types: white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF). These tissues are classified by using the probability of each voxel belonging to WM, GM,

or CSF based on the image intensity ranges [23]. These probability images, which are obtained from 250 young normal subjects, are provided by Montreal Neurological Institute [24].

Normalization: In group studies there are many differences across subjects in shape, size, orientation, and gyral anatomy. In order to compensate the variabilites of individuals, brain of each subject must be transformed into a common space for multisubject comparisons. For this reason, data must be represented in a common coordinate space, the stereotaxic space. There are two spaces used in neuroimaging: Talairach space which was derived from the brain of an elderly woman, and the second template was created by Montreal Neurological Institute (MNI) consisting of an average of 152 T1-weighted brain images.

Smoothing: Noise is an unavoidable random variable in image intensity. In an fMRI experiment, the change in intensity due to stimulation is between 0.5 per cent and 5 per cent of the average, whereas the noise level is between 0.5 per cent and 1 per cent [25]. Time series of each voxel should be filtered to reduce the low and high frequency noise. This is achieved by adjusting the overall intensity level by convolving each volume with a Gaussian profile filter. Typically, a width of between 3 and 10 mm full-width-half-maximum (FWHM) is used. Spatial smoothing is also necessary for further statistical analysis [26] unless a variational Bayes approach is applied.

Temporal Filtering: High-pass filtering is applied to time series of each voxel in order to get rid of low frequency components caused by scanner drift or physiological effects such as breathing or heartbeat.

3.2 The General Linear Model

Suppose the response variable in a voxel is expressed with a random variable, y_j, where $j = 1, \ldots, J$ is the number of observation. Also, for each observation there is a set of L explanatory variable, x_{jl}, where $l = 1, \ldots, L$. The general linear model

explains the response variable y_j in terms of a linear combination of the explanatory variables plus an error term [23].

$$y_j = \sum_{l=1}^{L} x_{jl}\beta_l + \epsilon_j \tag{3.1}$$

Here the β_i are parameters to be estimated, and ϵ_j are independent and identically distributed normal random variables with zero mean and variance σ^2, $\epsilon_j \sim N(0, \sigma^2)$. It is very common and useful to express the general linear model with matrix formulation. Writing the above equation for each observation gives a set of equations which is given with a matrix form:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_j \\ \vdots \\ y_J \end{bmatrix} = \begin{pmatrix} x_{1,1} & \cdots & x_{1,l} & \cdots & x_{1,L} \\ \vdots & & & & \\ x_{j,1} & \cdots & x_{j,l} & \cdots & x_{j,L} \\ \vdots & & & & \\ x_{J,1} & \cdots & x_{J,l} & \cdots & x_{J,L} \end{pmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_l \\ \vdots \\ \beta_L \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_j \\ \vdots \\ \epsilon_J \end{bmatrix} \tag{3.2}$$

and is written in matrix notation as:

$$Y = X\beta + \epsilon \tag{3.3}$$

While modeling fMRI data, Y is the time series of each voxel, X is the design matrix with a size of J x L, and β is the column vector of parameters. Design matrix X includes every assumption made in modeling the data, and the rest is encoded in the error term for each observation. Therefore, all the factors that effect the measurements should be included in the model. Statistical Parametric Mapping is a mass-univariate approach, so that the same model is applied at each voxel. Ordinary least square (OLS) is a common method to solve the general linear model. It finds the parameter estimates so that sum of squared errors becomes minimum:

$$||\epsilon||^2 = ||Y - X\hat{\beta}||^2 \tag{3.4}$$

In this case the parameters β are estimated by using:

$$\hat{\beta} = (X^T X)^- X^T Y \tag{3.5}$$

where X^- denotes the pseudo inverse of X.

3.3 Classical Inference

Once the parameter estimation is made, and the errors are calculated, specific questions can be asked about the effects by using contrasts. Statistical inference is performed through t-contrasts or F-contrasts. The form of a t-contrast is a linear combination of parameter estimates. For example, if we have three parameter estimates, $\beta_1, \beta_2, \beta_3$ and we want to test the first parameter β_1, then the contrast vector should be $c^T = [1, 0, 0]$. Also, a negative relation could be tested with $c^T = [-1, 0, 0]$. After the contrasts are specified t-statistics are performed given by [5, 27].

$$t = \lambda^T \hat{\beta} / SD(\lambda^T \hat{\beta}) \tag{3.6}$$

Each and every voxel is analysed with t-test and the resulting parameters are combined in a statistical parametric map (SPM). The values in SPMs may be distributed according to a known probability density function, usually the Studentâs t or F-distributions. These are also named as t- or F-maps.

In the statistical inference we search for the activation in the whole brain (or a part of it) although we ask questions about regionally specific effects. Because we are looking for the activation at unknown locations the voxels with the larger test statistic Z should be chosen. Therefore, the image is thresholded at the height z excluding the false positives with a probability of say 0.95, which means the chance that we take when we reject the null hypothesis or the probability of falsely assigning a voxel as activated is five per cent. Then we can raise the threshold to make the probability of finding activation in these parts 0.05. This is problematic because of testing the null hypothesis at a very large number of voxels and related with the multiple comparison problem in classical statistics.

Instead of referring to the whole brain we can restrict the analysis to a family

of voxel statistics. In this case the error rate will be the family wise error. One of the ways for performing a family-wise null hypothesis test is height thresholding which accounts for the number of tests. The Bonferroni correction is a simple way for adjusting the height threshold. Given a family-wise error rate P^{FWE} and a brain volume of n t-statistics the height threshold will be

$$\alpha = P^{FWE}/n. \tag{3.7}$$

Data from a voxel is not independent from its neighboring voxels in most functional images. This spatial correlation is based on three factors: Firstly, during the acquisition and reconstruction the signal from a voxel will also produce some signal change in nearby voxels because of the noise sources and the BOLD activity that spans large regions. Secondly, the preprocessing steps such as normalization and realignment performed prior to statistical analysis contain resampling that require averaging over neighboring voxels. The final factor causing the correlation is spatial smoothing. Although the noise is uncorrelated throughout the image, the signal is not independent. Smoothing also requires averaging of neighboring voxels to increase the signal to noise ratio according to the matched filter theorem [28], so it will increase the spatial correlation.

The Bonferroni correction defined in equation 3.7 relies on the assumption that the family-wise probability of each voxel is independent, with n is the number of independent voxels. In fact, the number of independent observations will be much smaller than the number of voxels due to the spatial correlation in the functional images as explained above. Hence, the Bonferroni correction will overestimate the number of independent observations. Simply putting the number of independent observations, n_i instead of n will hold for the right correction. However, after the spatial smoothing that blurs the image, there is not a simple way of calculating the number of independent values. And the Bonferroni correction will be too conservative and cannot be used anymore. This problem can be solved with random field theory (RFT).

Before an introduction to the random field theory, the concept of resolution

element (resel) should be introduced. A resel can be defined as the volume of voxels that has the same dimensions as the FWHM of the smoothed image [29]. For example, if a 2D image has a 8 pixels by 8 pixels of smoothness, then the resel is a block of 64 pixels. Having 128 x 128 pixels in the image, the number of resels is 128 x 128 /64 = 256. For a 3D dataset consisting x x y x z voxels and a smoothness with FWHM of V voxels, we can formulate the number of resells as

$$R = (x \text{ x } y \text{ x } z) \ / \ V^3 \tag{3.8}$$

According to the RFT, the appropriate height threshold can be found in a few steps. First, the spatial correlation of the image is estimated. Then this smoothness value is substituted in the RFT equation to compute the corresponding expected Euler characteristic at different thresholds. Euler characteristic (EC) can be thought of as the number of active clusters after thresholding, so the expected EC, E[EC], will be the probability of finding a cluster that is above threshold in the image. In this case, the E[EC] can be treated as our family wise error rate, $P^{FWE} \approx E[EC]$.

Once we know the number of resels in a 2D image, the E[EC] can be estimated with the equation given by Worsley et al. [29]:

$$E[EC] = R(4log_e2)(2\pi)^{-3/2} Z_t \exp\left({Z_t}^2/2\right) \tag{3.9}$$

See Figure 3.1 for a comparison of uncorrected and a family wise error (FWE) corrected statistical maps. One can conclude that a voxel statistic of the magnitude 4.61 has a 5 per cent chance of arising anywhere in the whole brain drawn from the null distribution.

A comparison for RFT and Bonferroni correction can be made for a dataset conta ining 100 resels, there is a 0.05 probability of seeing a cluster of activation at threshold Z of 3.8, whereas in the Bonferroni correction the threshold will be a Z-threshold of 3.3, giving a less conservative test of significance (0.05/100).

The underlying assumption of RFT is that error fields are continuous multivariate Gaussian distributions. This is provided only in case of sufficiently smoothing the

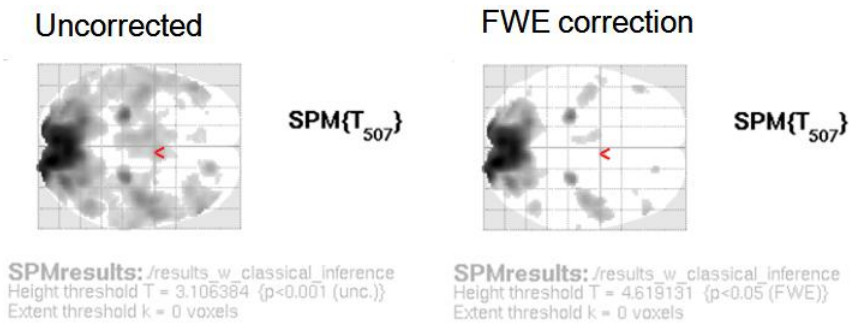

Figure 3.1 The p-value is corrected using the random field theory.

data and correctly specifying the general linear model so that the residual errors are Gaussian. In real life, not all the voxels would have similar noise properties. One voxel may contain a major blood vessel or it can be on the edge of the brain.

The proposed method is the Bayesian inference where instead of referencing to a null hypothesis, inferences are based on the posterior distributions. Chapter 4 will describes the Bayesian approach.

3.4 Intersubject Analyses

A single result for a group of subjects can be created with statistical methods including 'fixed-effects' and 'random-effects' analyses.

3.4.1 Fixed-effects analyses

This method includes comparing the data points for all subjects. It is assumed that experimental effect is constant across subjects and the stimulus has the same effect upon BOLD signal in every subject. Because this method involves averaging the conditions across the subjects, the inference will have a high sensitivity to the extreme results form individual subjects. This allows making inference about the subjects who were run in that particular study.

3.4.2 Random-effects analyses

Unlike fixed-effects analyses, random-effects analyses take into account between-session errors, and the inference tells us about the population from which the group of subjects was drawn. Subjects might have BOLD responses with large differences in amplitude, so at first level activations for each subject are calculated independently. At the second level within a classical approach a t-test is used to evaluate whether first-level parameters of each subject are drawn from a distribution with zero mean. If the test is significant then experimental manipulation has an effect which is valid for the population.

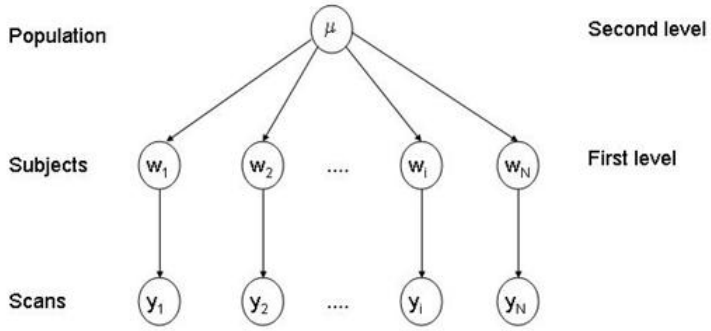

Figure 3.2 Generative model of random effects analysis [23].

4. Bayesian Advances in Neuroimaging

4.1 Bayes' Theorem

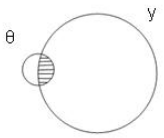

Figure 4.1 The posterior probability can be derived from the joint probability of two events.

Given probabilities $p(\theta), p(y)$, and the joint probability $p(\theta, y)$, we can write the conditional probabilities

$$\begin{aligned} p(y|\theta) &= \frac{p(\theta, y)}{p(\theta)} \\ p(\theta|y) &= \frac{p(\theta, y)}{p(y)} \end{aligned} \tag{4.1}$$

Eliminating $p(\theta, y)$ gives Bayes rule

$$p(y|\theta) = \frac{p(\theta|y)p(y)}{p(\theta)} \tag{4.2}$$

Given some observed data y, and the probability distribution of the model parameter $p(\theta)$, the posterior or conditional probability distribution of the parameter $p(\theta|y)$ can be calculated with equation 4.2. $p(\theta)$ is called as prior and it expresses our initial belief about the parameter. $p(\theta|y)$ is called as the posterior and it expresses our belief about the parameter in the presence of data. $p(y|\theta)$ is called the likelihood function of θ for given data y. $p(y)$ is constant with respect to parameter θ [30].

For any voxel, under Gaussian assumptions, one can calculate the posterior density using Bayes rule, with the specification of the appropriate priors for the parameters

and the likelihood function (Figure 4.2)

$$p(\theta|y) \approx p(y|\theta)p(\theta) \tag{4.3}$$

These distributions can be specified in terms of their means and variances in a fully Bayesian approach. In this scheme, the mean and the precision of the posterior distribution m and d, can be computed by the mean and precision of the prior (m_o and p_o) and the likelihood (m_D and p_D) with the formulae below. Precision is defined as the inverse of the variance.

$$\begin{aligned} p &= p_o + p_D \\ m &= \frac{p_o}{p} m_o + \frac{p_D}{p} m_D \end{aligned} \tag{4.4}$$

Instead of using a fully Bayesian approach we will use the empirical Bayes which was

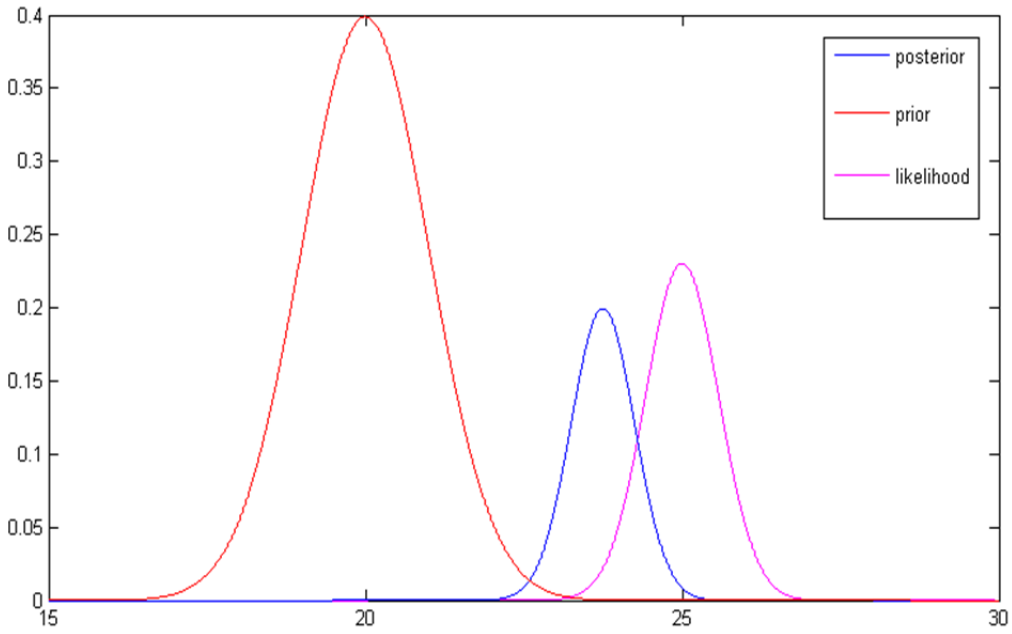

Figure 4.2 According to Bayes rule, the posterior distribution is closer to the likelihood because the likelihood has lower variance.

defined by Friston et al., [31, 32]. In this approach variances of the prior distributions are estimated from the data. This is why it is called empirical Bayes.

4.2 Parametric Empirical Bayes (PEB)

Almost all the models used in neuroimaging have a hierarchical structure. Empirical Bayes is the approach for the treatment of hierarchical models in a Bayesian framework. In hierarchical models, equation 4.3 can be written in terms of a factorization of intermediate terms, $p(\theta|y_1, y_2)$. These terms are called empirical priors and play an important role between the likelihood and priors. There are many examples of this type of models where the parameters at one level control the distribution of parameters at lower levels, eg. Markov models of hidden states or mixture models.

Estimating the variance components in fMRI time-series induced by multisubject studies is performed through the Expectation-Maximization (EM) algorithm. The covariance component estimation is also the connection between classical inference and empirical Bayesian inference [32]. PEB is a special case of an EM algorithm for linear Gaussian models (Figure 4.3). The goal of PEB is to estimate the intermediate level parameters.

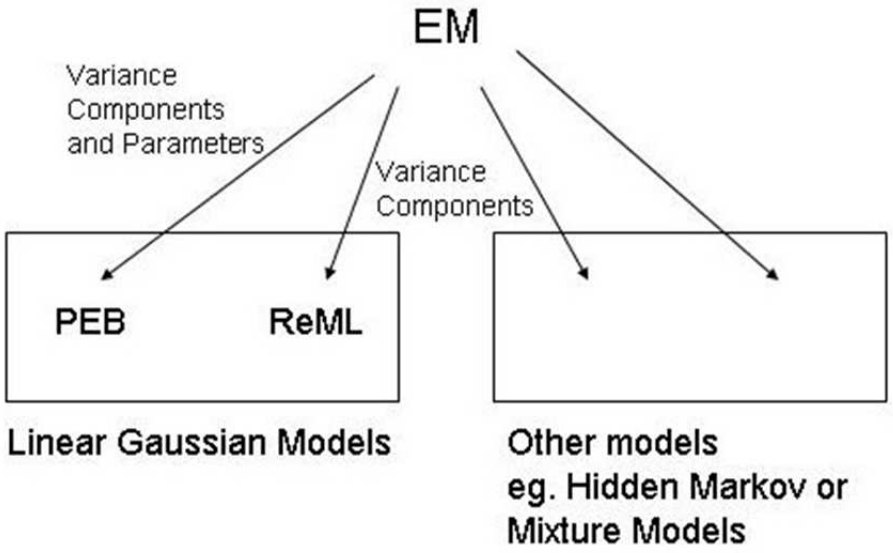

Figure 4.3 Parametric Empiric Bayes is a special case of an Expectation-Maximizastion Algorithm to estimate posterior distirubutions for parameters in E-step and to update hyperparameters in M-step [33].

The PEB algorithm used in the next chapter returns the moments of the posterior probability distribution functions of a hierarchical linear observation model under

Gaussian assumptions.

$$\begin{aligned} y &= X_1\theta_1 + \epsilon_1 \\ \theta_1 &= X_2\theta_2 + \epsilon_2 \\ &\vdots \\ \theta_{n-1} &= X_n\theta_n + \epsilon_n \end{aligned} \tag{4.5}$$

At the final level θ_n is set to 1 and the known prior covariance is used. This covariance comes from the variance in parameters over voxels: Because we are looking for the same effect at each voxel there is a natural hierarchy in neuroimaging, where the first level is the effect at any voxel, and the second is the effects over voxels [34]. Hence, the voxels are used as a second level providing the variances for any one voxel in order to be used as the prior covariance. Then these priors are used to estimate the error covariance hyperparameters at the higher levels.

The voxels which are going to be anaylsed with Bayesian inference, are identified by maximum likelihood estimation at first. Conditional paremeter estimation and ReML hyperparameter estimation are performed through Bayesian scheme. A hyperparameter is related to the probabilistic behaviour of a parameter. It is assumed that on average over all voxel, the net experimental effect is zero but each voxel has a variability determined by their prior variance. The prior variances of each voxel is computed with empirical Bayes and used in the estimation of ReML hyperparameters. To consider the ReML hyperparameter estimation as a variance component estimation, covariance components which are induced by paremeter variations (Q), and which are induced by error non-sphericity (V) are collapsed in a big covariance component structure.The estimated errors are given by

$$C = \lambda_1 Q_1 + \lambda_2 Q_2 \tag{4.6}$$

where λ is the variance component or the hyperparameter. This estimation uses the covariance over voxels given by YY^T/n, with n is the number of voxels. The confounds are treated as fixed with infinite prior variance. The estimated errors which are going

to be used for the estimation of conditional parameters can be written in the matrix form:

$$C_\theta = \begin{bmatrix} \sum_{i=1} \lambda_i Q_i & \cdots & 0 \\ \vdots & \infty & \\ \vdots & \ddots & \\ \vdots & \ddots & \\ 0 & & \infty \end{bmatrix} \tag{4.7}$$

This prior covariance is input for the next ReML which estimates the voxel-specific error hyperparametersλ_ϵ which in turn defines the error covariance C_ϵ with the appropriate correlation or nonsphericity of the errors V.

Once we estimate the prior covariance C_θ and the error covariance C_ϵ, the posterior moments of the parameters can be computed for each voxel. The posterior covariance $C_{\theta|y}$ and the posterior mean $\mu_{\theta|y}$ are given by [32].

$$\begin{aligned} C_{\theta|y} &= (X^T C_\epsilon^{-1} X + C_\theta^{-1})^{-1} \\ \mu_{\theta|y} &= C_{\theta|y}(X^T C_\epsilon^{-1} y + C_\theta^{-1} \mu_\theta) \end{aligned} \tag{4.8}$$

Finally, in order to form a posterior probabilty map, the posterior moments are used in the equation below for each voxel

$$p = 1 - \Phi(\frac{\gamma - c^T \mu_{\theta|y}}{\sqrt{c^T C_{\theta|y} c}}) \tag{4.9}$$

where the Φ is the cumulative normal distribution and the c is the contrast vector which is used to specify the corresponding effect (see section 3.3). This whole procedure to construct a PPM is shown in Figure 4.4.

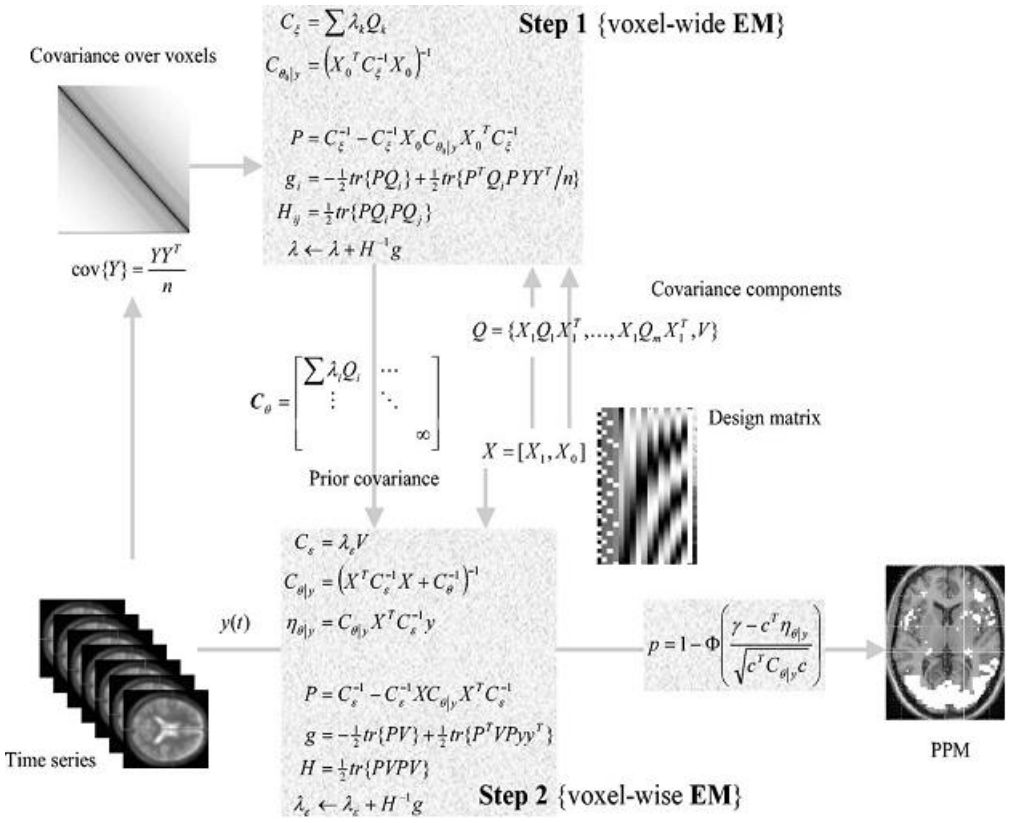

Figure 4.4 Derivation of posterior probability maps with ReML algorithm for a hierarchical model [34].

4.3 Simulation Study

Assume we have a dataset which was collected from 40 subjects. And the search volume includes 100 voxels. We want to make a random-effect analysis and find the maximum likelihood estimates and compare them with PEB estimates. This simulation was adapted from [33] to mimic the real dataset used in the next section. The two-level linear model is given by

$$\begin{aligned} y &= X\theta + \epsilon_1 \\ \theta &= \mu + \epsilon_2 \end{aligned} \tag{4.10}$$

We will use global shrinkage priors so that our prior belief is that across the whole brain the average effect size is zero, $\mu = 0$. Variability of the voxels will be determined by their prior precision which is estimated by their prior precision. This prior precision can be estimated from the data using Empirical Bayes. In this simulation we will simply

take the variability of the voxels as normally distributed with the precision $\alpha = 1$. Also, the effect size across the subjects follow a normal distribution around the true effect size with the precision λ_i, $i = 1, .., 40$ (Figure 4.5). In this case the covariance components for the first and second level are given by

$$\begin{aligned} C_1 &= \sum_{i=1} \lambda_i Q_i \\ C_2 &= \frac{1}{\alpha} I_{100} \end{aligned} \tag{4.11}$$

The update equations for the hyperparameters λ_i and α are as follows

$$\begin{aligned} \beta(i) &= \frac{\gamma_i}{N} \sum_{n=1}^{N} y_{in} \\ \frac{1}{v_i} &= \frac{1}{N - \gamma_i} \sum_{n=1}^{N} (y_{in} - \beta(i))^2 \\ \gamma_i &= \frac{N v_i}{N v_i + \alpha} \\ \frac{1}{\alpha} &= \frac{1}{\sum_i \gamma_i} \sum_{i=1}^{V} \beta(i)^2 \end{aligned} \tag{4.12}$$

As one can see in Figure 4.9, for the gamma values which are defined as the ratio of the data precision to the posterior precision for each voxel, PEB estimates are quite accurate on average although it fails at a minority of the voxels with a high variability between subjects. For example the gamma is extremely small for the voxel 29 which has a large variance. For such voxels, Bayesian estimate of the population effect size might be wrong. However, Figures 4.6-4.8 show that PEB estimates are more precise than ML estimates for the majority.

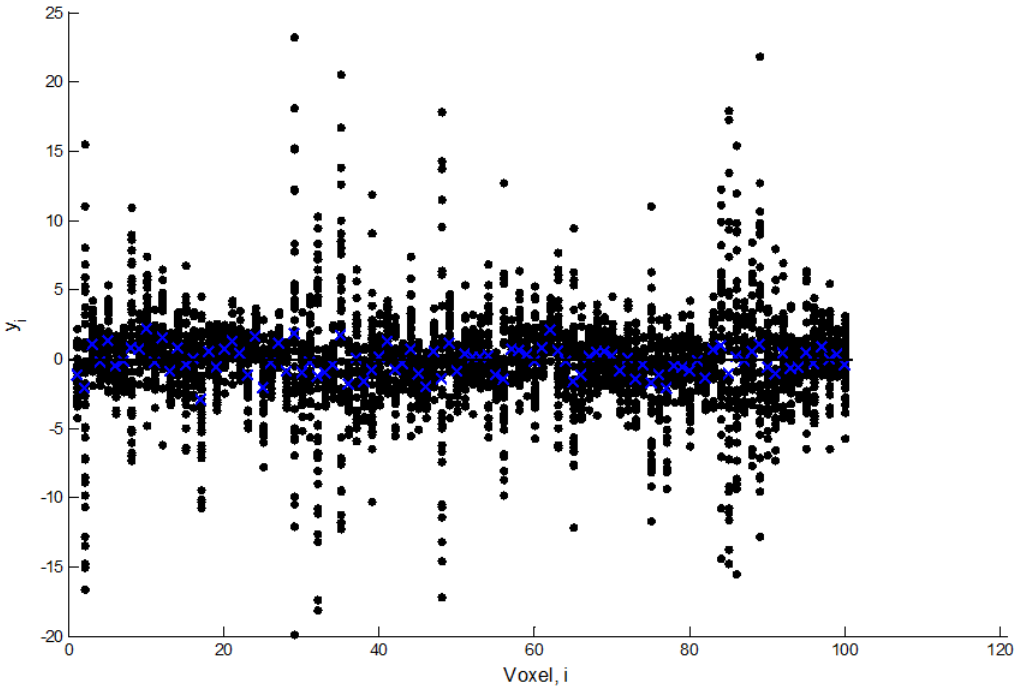

Figure 4.5 Data from 40 subjects (black dots) are normally distributed around the true effect size with the precision λ_i. Blue crosses are the sample means.

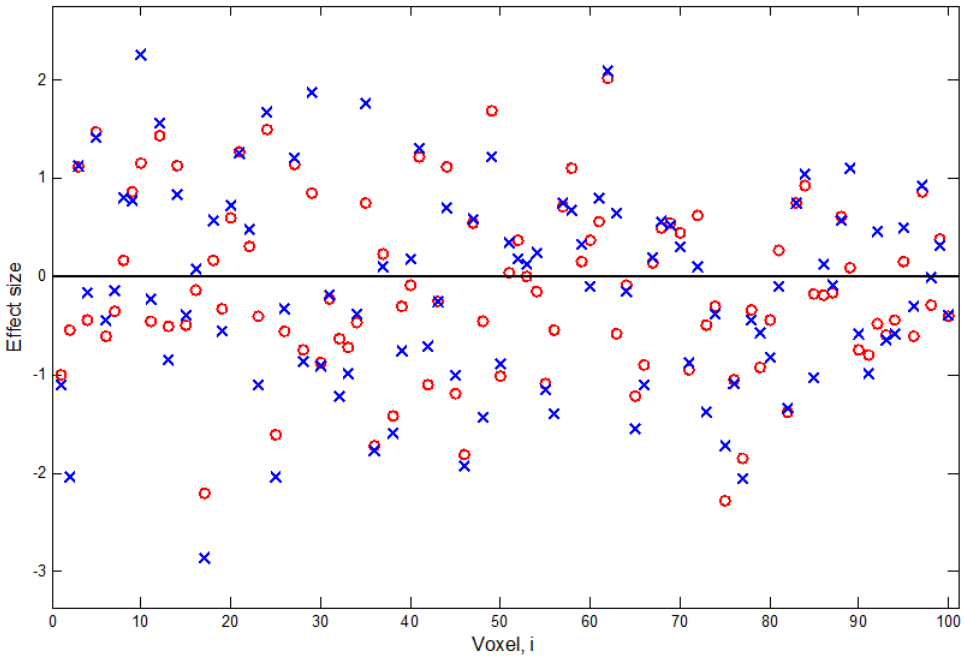

Figure 4.6 Maximum Likelihood estimates (blue crosses) with an estimation error of 1.98. Red circles are the real effect sizes for each voxel.

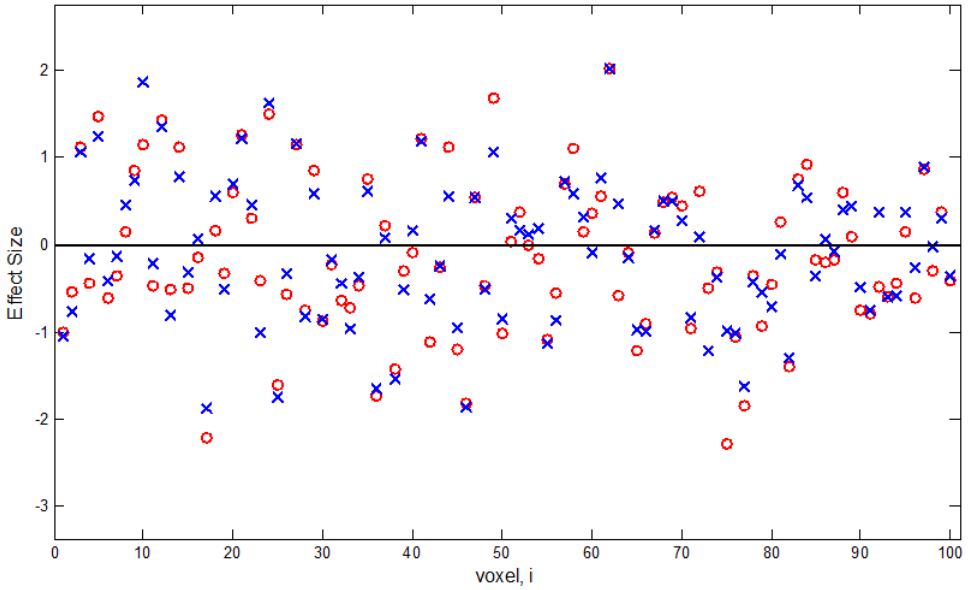

Figure 4.7 After 5 PEB iteration.

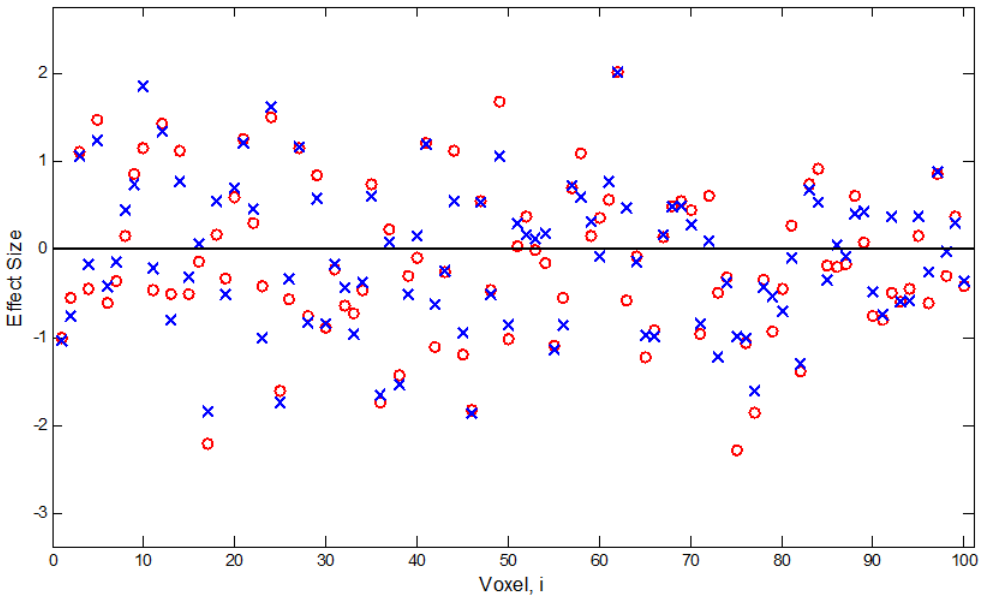

Figure 4.8 After 10 PEB iteration. Estimation error is 0.70. True $\alpha = 1.00$, Estimated $\alpha = 1.24$.

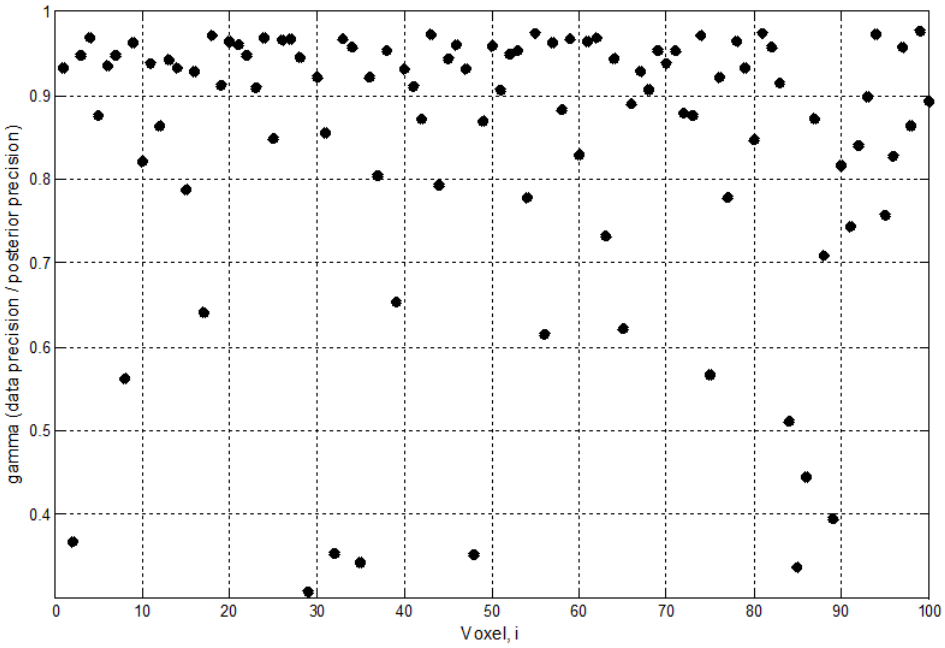

Figure 4.9 Ratio of the data precision to the posterior precision.

5. Bayesian Estimation and Inference for the fMRI Data with Steady State Visual Evoked Stimuli

5.1 Data Acquisition

fMRI was recorded from 40 volunteers (20 male, 20 female, mean age 25.8) without any neurological disorder. The procedure was approved by the local ethics committee of Istanbul University, Istanbul Faculty of Medicine.

Flashing light stimuli were presented at 6, 8, 10, 12, 14, 18, 22, 26, 30, 34, 38, 42 and 46 Hz. Each session (Figure 5.1) started with 29.8 s (10 TR) baseline recording followed by three blocks each consisting of 44.7 s (15 TR) of stimulation and 44.7 s baseline recording yielding a total period of 298 s, and 70 minutes for all sequences. A 1.5 T MR system (Achieva, Philips Healthcare, Best, The Netherlands) was used to acquire $T2^*$-weighted images (TR/TE/FA = 2981 ms/50 ms/90°, matrix size = 64 x 64, 32 axial slices, voxel size = 3.59 x 3.59 x 4) with a gradient echo EPI sequence. More detailed information about the acquisition protocol for the experiment can be found in [35].

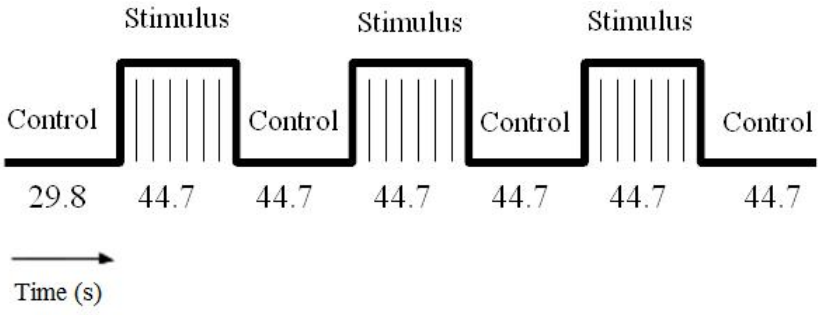

Figure 5.1 Block design for the stimuli used in the acquisition.

5.2 Data Analyses

5.2.1 Preprocessing

Preporocessing steps included realignment of functional images and creating the mean image, coregistration of structural image to mean functional image, segmentation of coregistered structural image, normalization of bias corrected structural image and realigned functionals into a standard space (MNI152, 2mm) template. The code was modified from Rick Henson's script source available at [36]. Finally, spatial smoothing with a Gaussian kernel of 8 mm FWHM and a high-pass temporal filtering were applied.

5.2.2 First Level Analyses

For the general linear model being used at first level, the design matrix comprised stimulus pattern convolved with the hemodynamic response function and its dispersion and time derivatives as well as the motion parameters (Figure 5.2). The last column of ones models the constant or mean signal. Baseline is used as a reference because even without the stimuli the signal is never zero.

Scans from each subject were subjected to a similar analysis with a design matrix which is specific to that subject. Parameters, which model the contribution of each component in the design matrix, were estimated through classical esimation. The reason for using the classical estimation procedure is because the Bayesian estimation is computationally expensive. A variational Bayesian scheme was also performed but it takes too many hours for a single subject. After parameter estimation, the contrast vectors were defined for each column of design matrix except motion parameters. Then, each contrast vector was estimated accordingly.

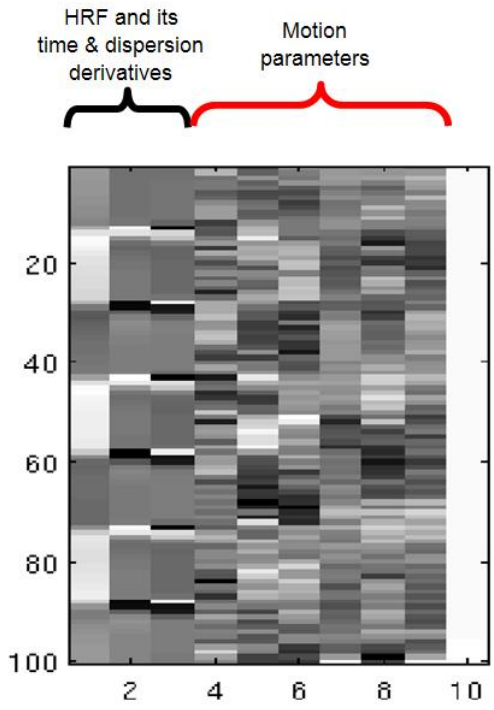

Figure 5.2 Design matrix for the first level analysis.

5.2.3 Second Level Analysis

Parameters modeling the first column of the design matrix and being estimated with corresponding contrast vector for each subject were used at the second level analysis to estimate group level parameters. The design matrix simply involves âonesâ to estimate population effect for each stimulus frequency (Figure 5.3). Under Gaussian assumptions, a parametric empirical Bayes approach, which is defined in Section 4.2, was used to estimate population-wise parameters. The contrast vectors were defined and estimated for each parameter. Whole inference was based on the posterior distribution of the parameters so that we used posterior probability of each voxel to construct posterior probability maps. Effect size thresholds are based on the estimate of the prior variance. In order to compare PPMs with classical inference results, we also specified effect size thresholds as zero.

After constructing frequency wise PPMs, we grouped them according to alpha, beta and gamma bands so as to make the random effect analysis and the Bayesian

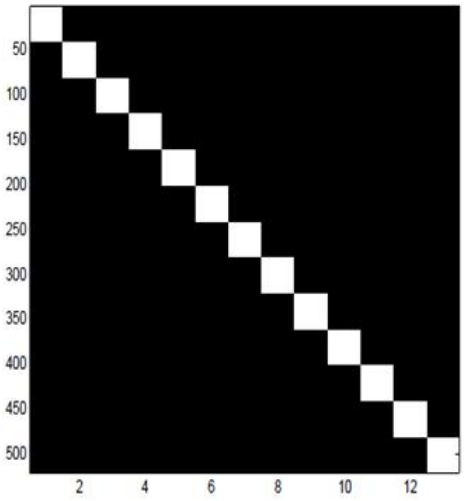

Figure 5.3 The design matrix for the second level analysis.

inference from a band-wise perspective: Contrasts were estimated through 6-8-10-12 Hz, 14-18-22-26 Hz, 30-34-38-42-46 Hz for alpha, beta and gamma bands, respectively.

Finally, we used the differences in effect sizes with respect to stimulus frequency to assess results of Bayesian inference further. Effect sizes were plotted for the whole brain as well as some specific ROIs: The LGN sends most of its axons to primary visual area (V1), so the information coming from the LGN is processed in V1, then passes to the other visual areas V2, V3, V4, etc. (Figures 5.4, 5.5). Therefore, the activation of LGN must be present during this information flow.

Detecting the LGN activity is a challenge for most neuroimaging applications because of its small size. fMRI is capable of detecting LGN activity [38]–[39]. It has been shown that there is a linear correlation of activation sizes between LGN and V1 [40], where the activation size was defined in terms of the number of voxels. In this study, we compared the change in the effect sizes of LGN and V1 during graded visual stimulation by using PPMs with an effect size threshold of zero. In order to get effect sizes in these specific regions, we masked the contrast images derived from group level analysis with the probabilistic ROI images from Juelich Histological Atlas [41].

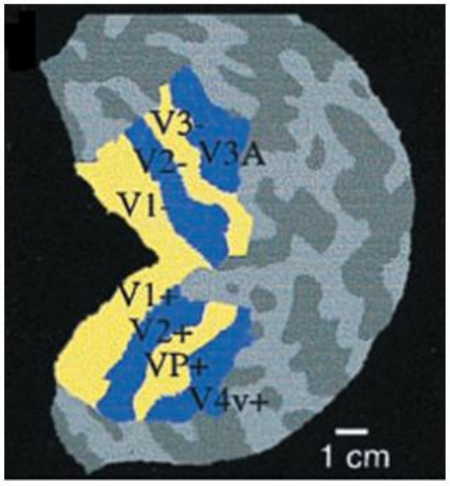

Figure 5.4 Retinotopic mapping of the visual cortical areas V1-, V2-, V3-, and V3A, superior to the calcarine sulcus, and areas V1+, V2+, VP+, and V4v+, inferior to the calcarine. 1 and 2 indicate upper and lower visual field representations, respectively.adapted from [37].

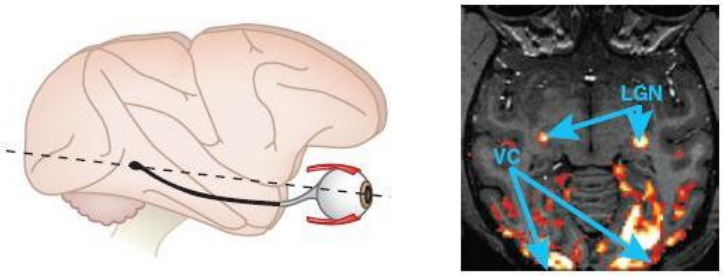

Figure 5.5 LGN is the gateway for sensory information that reaches to visual cortex [42].

6. RESULTS

Below are the second level group analysis results over 40 subjects. Maps were constructed in two ways: (i) frequency and (ii) band-wise. In Figure 6.1, PPM for each frequency is shown on the axial template slice where the global maximum of the effect size is observed. Height threshold for significancy is 0.95.

To assess the results of the statistical inference for the group level we used the variation of the effect sizes with respect to stimulus frequency. As shown in Figure 6.2, we found the global maximum response at 10 Hz. There is a general decline in effect size as the stimulus frequency increases. Local maxima can also be observed at 10, 14 and 30 Hz for each band respectively.

Figure 6.3 shows the band wise PPMs for alpha, beta, and gamma bands (from top to bottom) in left and right hemispheres on the inflated template volume. Here the significancy threshold was specified as 0.99. The extent of activation is largest in alpha and smallest in gamma with a steady decrease as the frequency increases.

In Figure 6.1 the effect size thresholds are based on the estimate of the prior variance. This type of thresholding is more meaningful because we are not interested in effects of any size. For example, the evoked sensory responses are in the range of 1-2 % of global mean value, and some cognitive effects are about 0.5 %. On the other hand, we specified the thresholds as zero in the following Figures 6.4-6.16, so that the effects of any size are shown with a colormap, if significant. Hence, we could compare the PPMs with corresponding SPMs.

We could detect the activation in LGN by decreasing the effect size threshold without decreasing the significancy. PPMs with zero effect size threshold show that the effect sizes in LGN are less than that in V1 during graded visual stimulation. The threshold for significancy is 0.95 (Figures 6.4-6.16, left). No activation was detected in

right LGN at 26 Hz. ROI analyses yield that the normalized mean effect sizes in LGN have a global maximum at 12 Hz with local maxima at 18 and 42 Hz, whereas V1 has a maximum at 10 Hz (Figure 6.18).

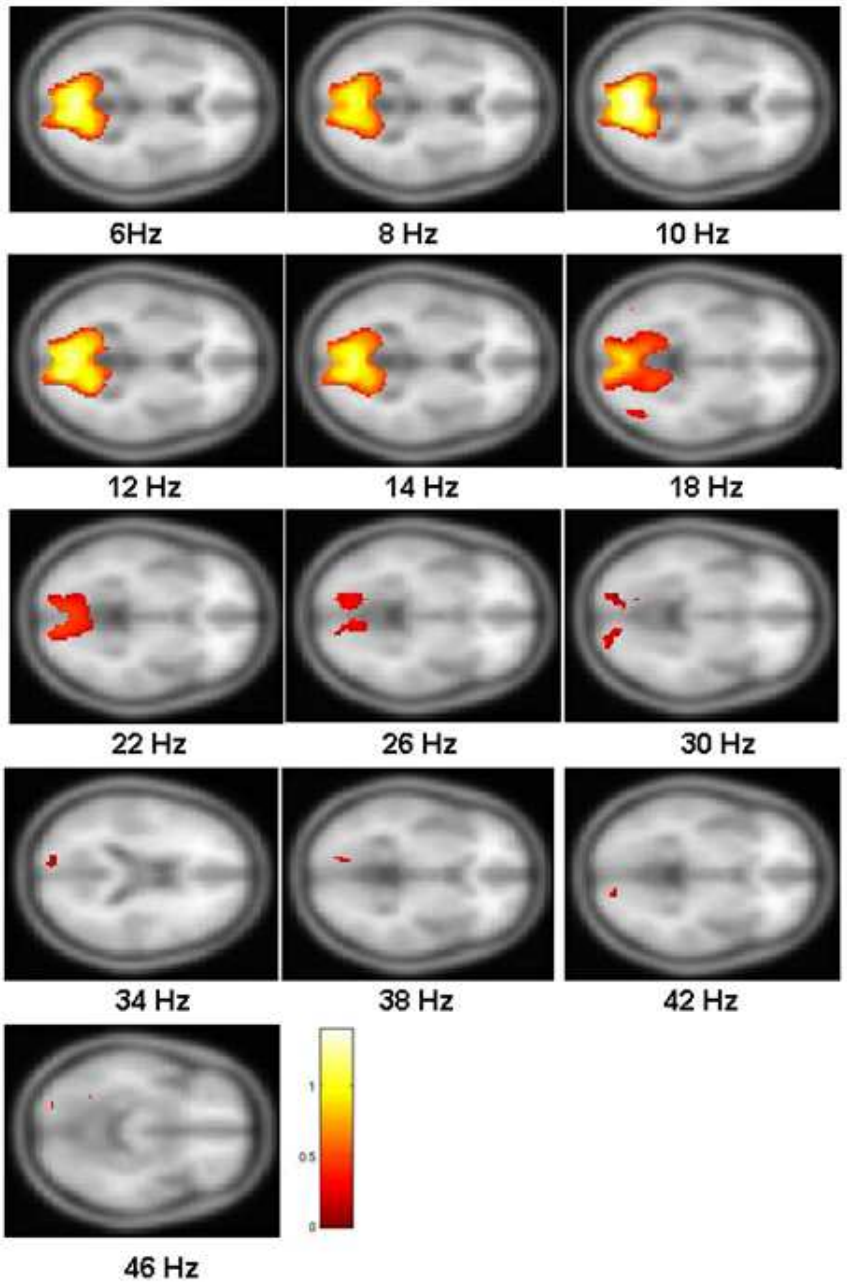

Figure 6.1 PPM for each frequency is shown on the axial template slice where the global maximum is observed (Height threshold, $P = 0.95$).

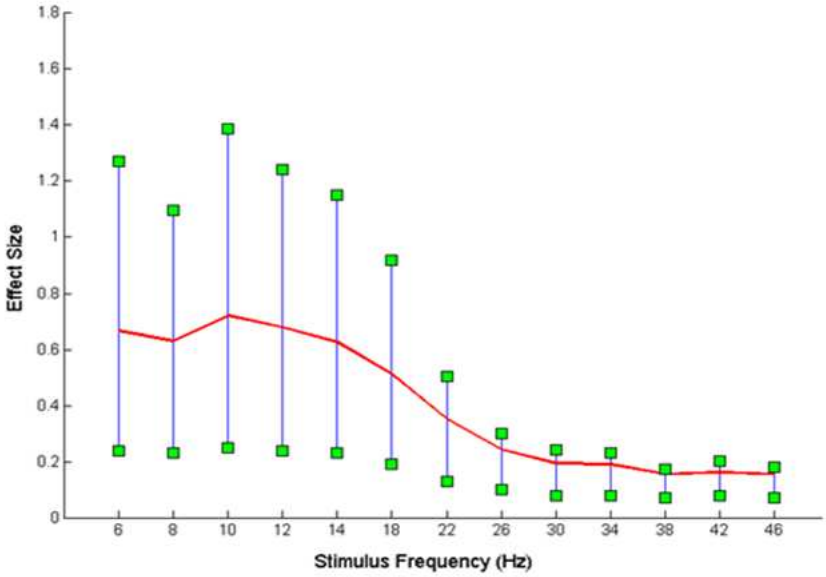

Figure 6.2 Mean effect sizes (red) of the frequency based PPMs. The maximum effect sizes and the thresholds are also shown with green.

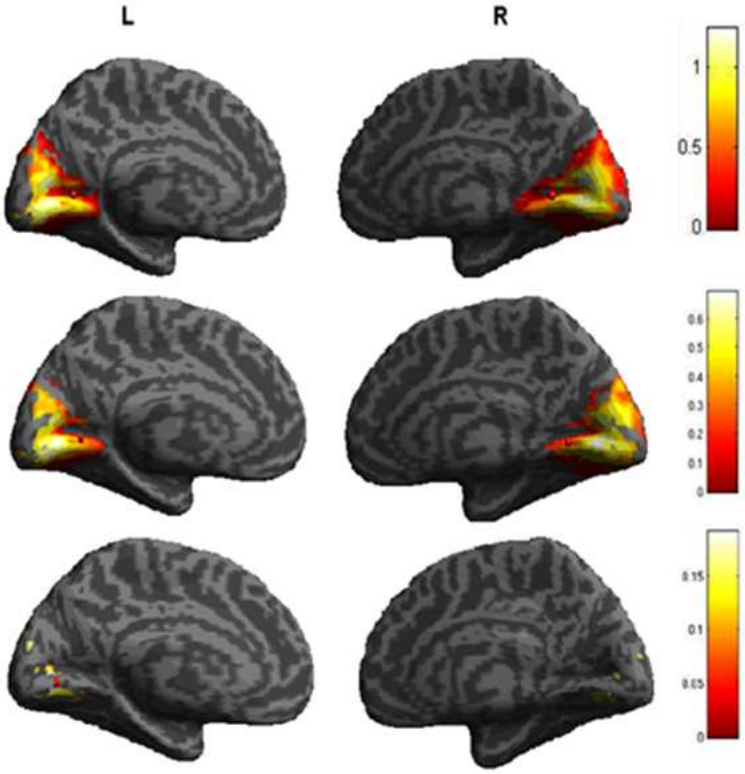

Figure 6.3 Band wise PPMs (P = 0.99) for alpha, beta, and gamma bands (from top to bottom) in left and right hemispheres on the inflated template volume.

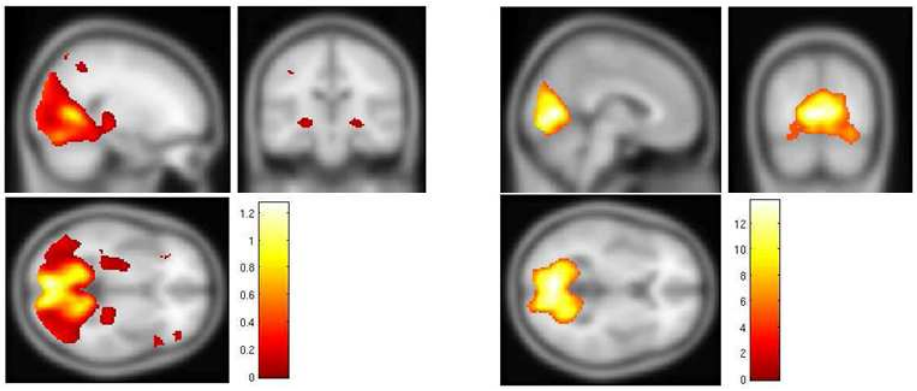

Figure 6.4 6 Hz.

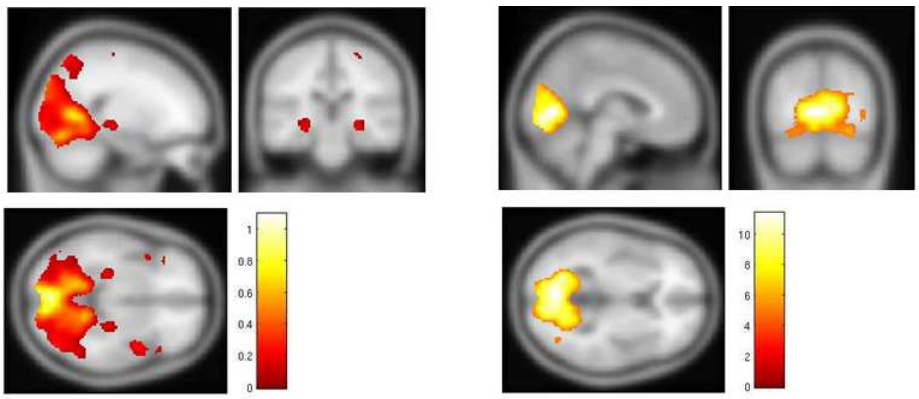

Figure 6.5 8 Hz.

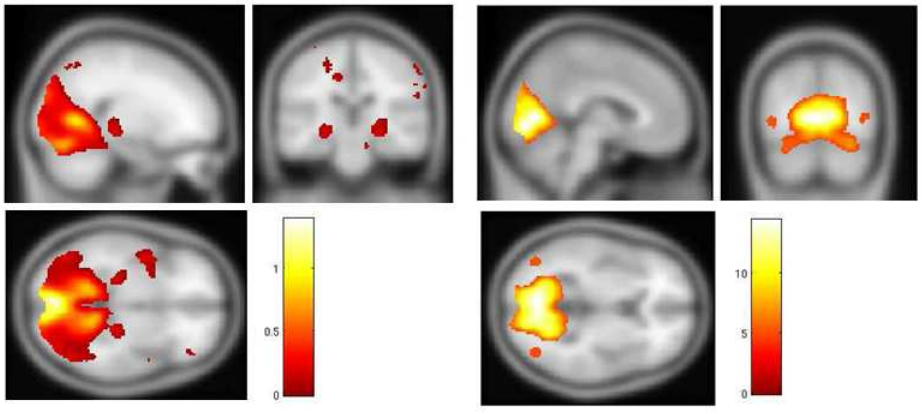

Figure 6.6 10 Hz.

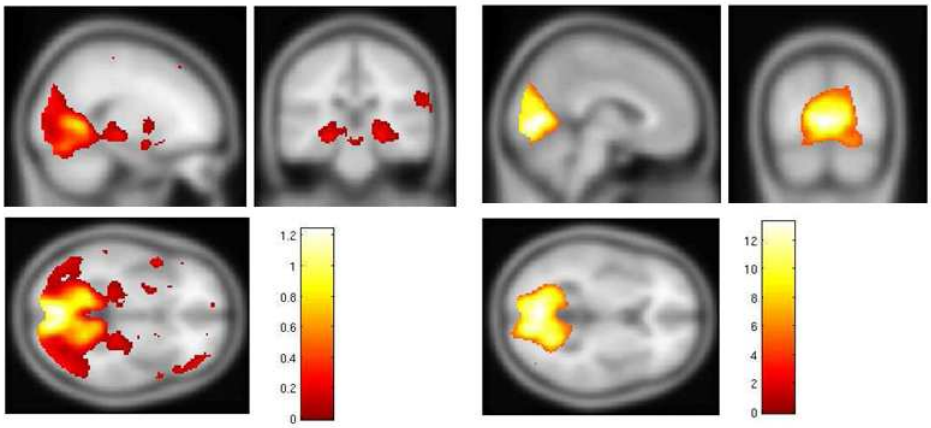

Figure 6.7 12 Hz.

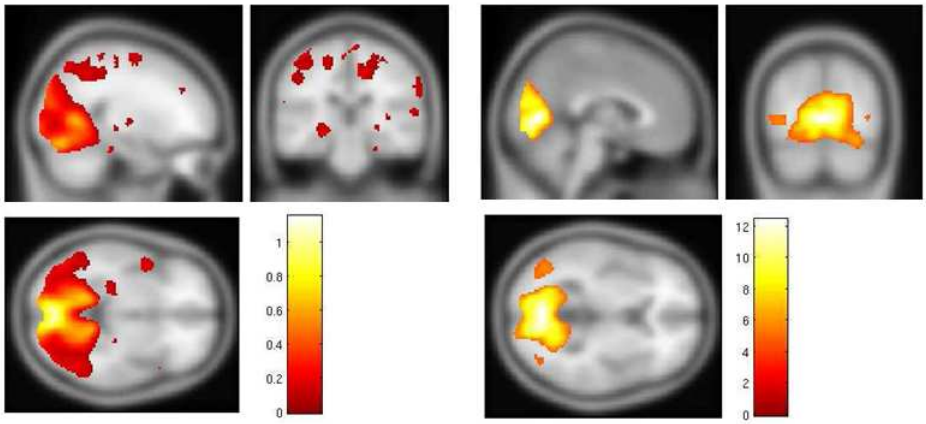

Figure 6.8 14 Hz.

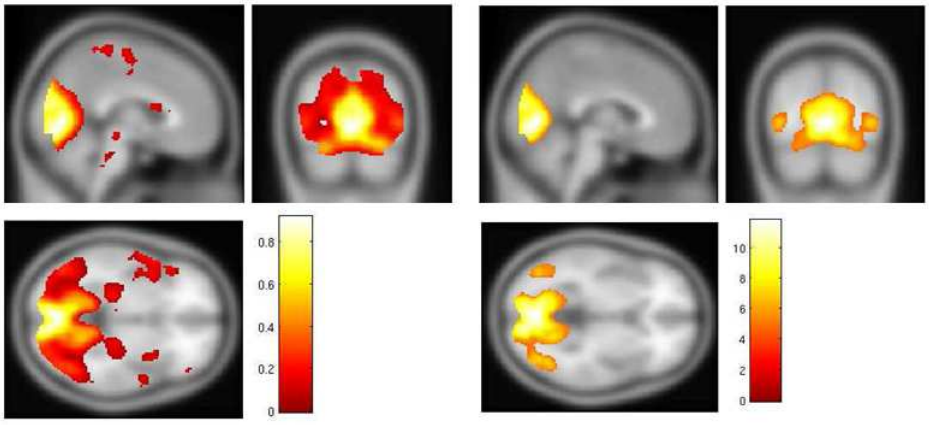

Figure 6.9 18 Hz.

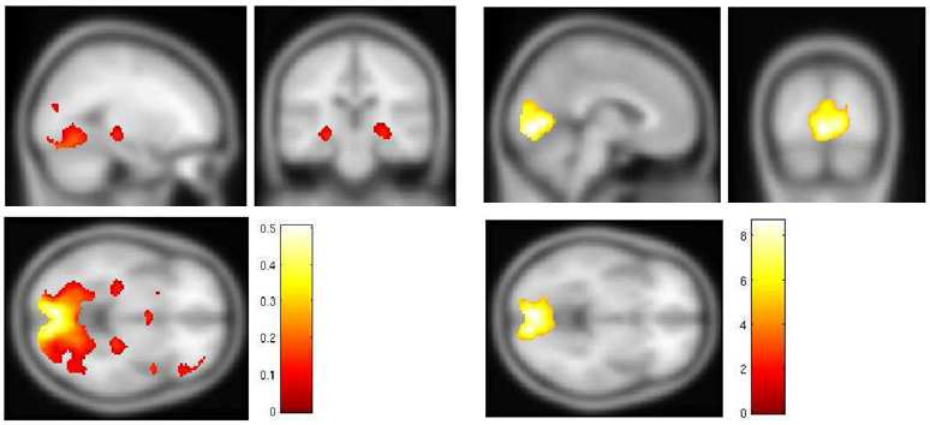

Figure 6.10 22 Hz.

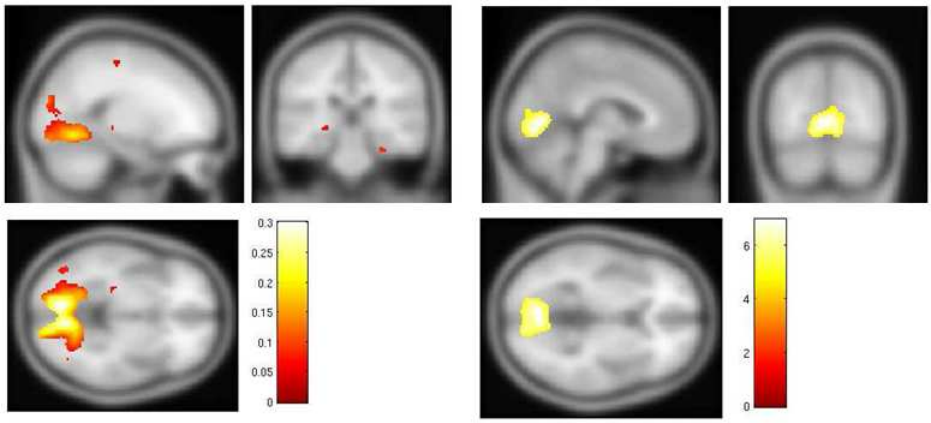

Figure 6.11 26 Hz.

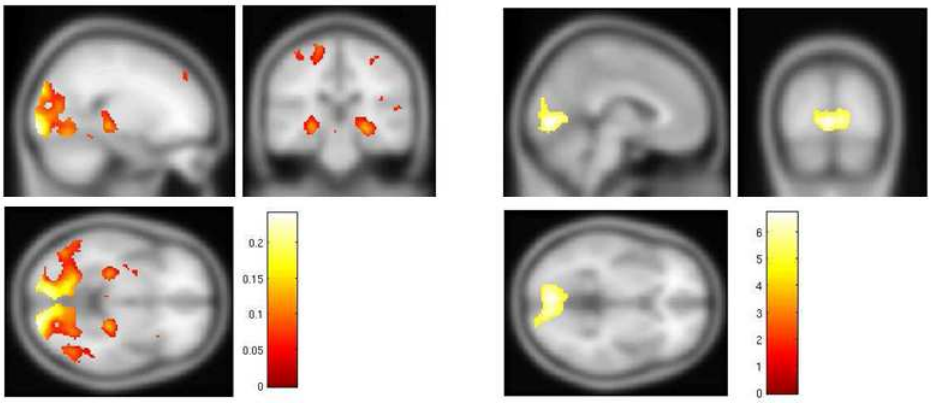

Figure 6.12 30 Hz.

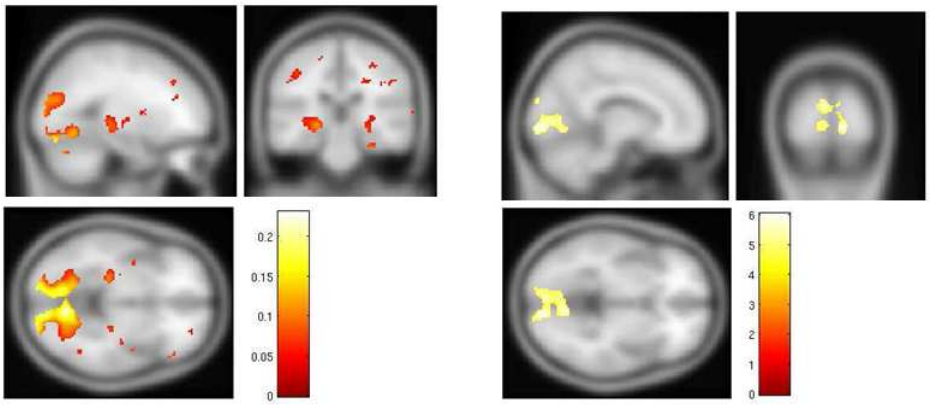

Figure 6.13 34 Hz.

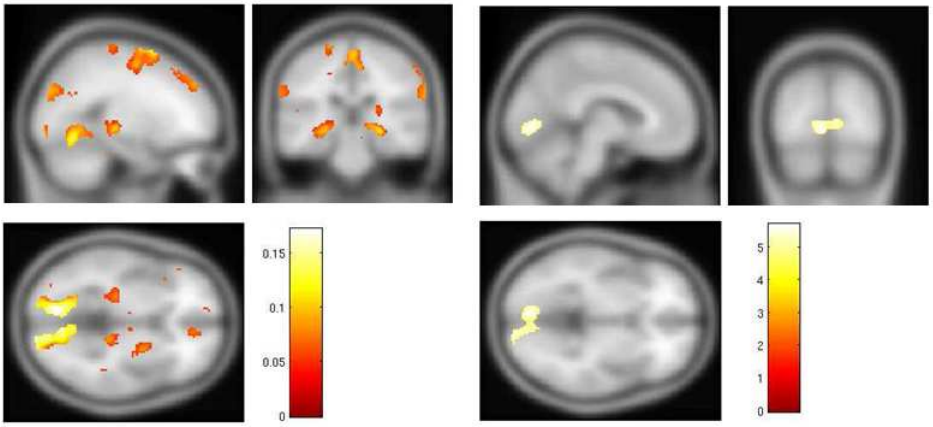

Figure 6.14 38 Hz.

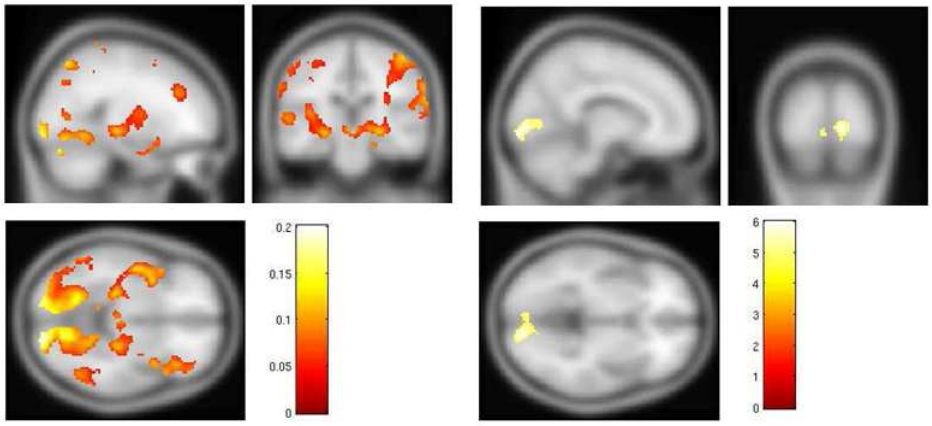

Figure 6.15 42 Hz.

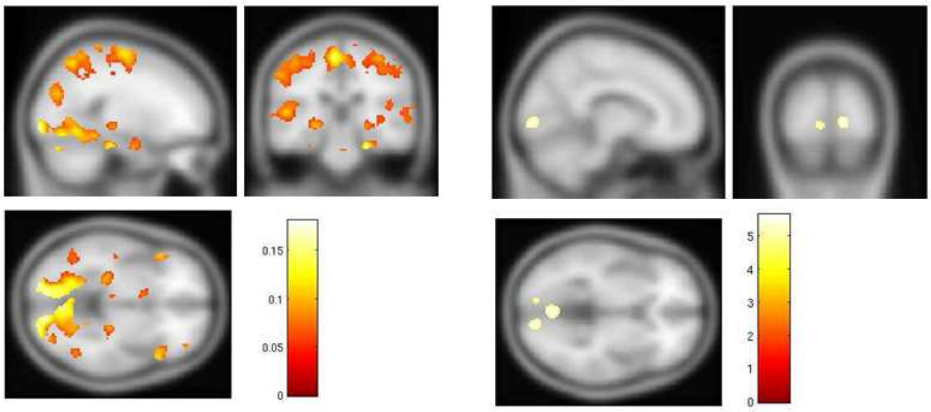

Figure 6.16 46 Hz.

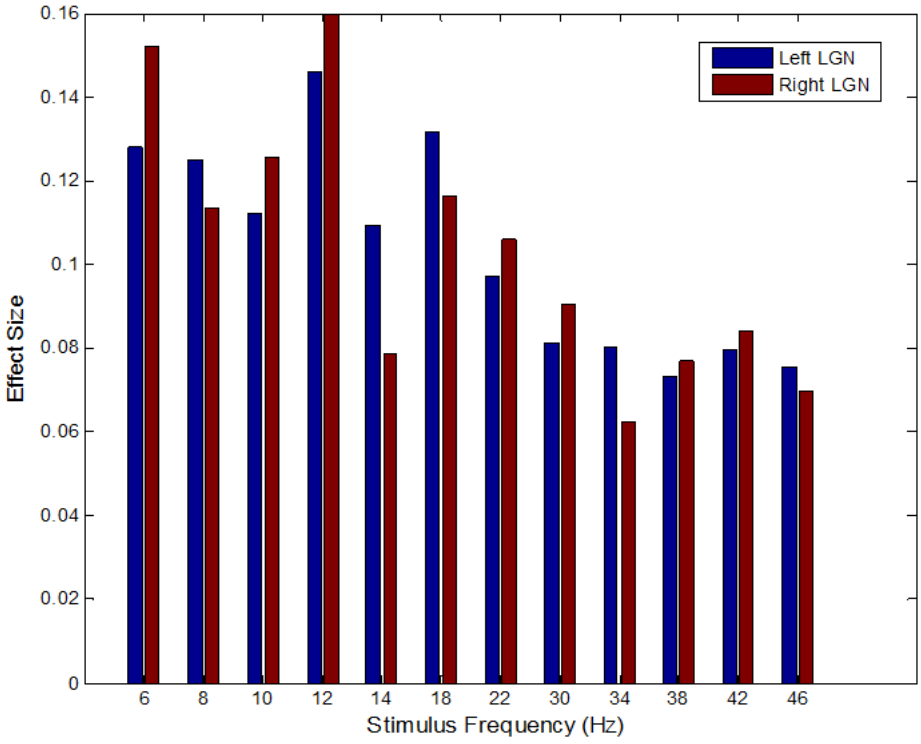

Figure 6.17 Mean effect sizes of the right (red) and left (blue) LGN.

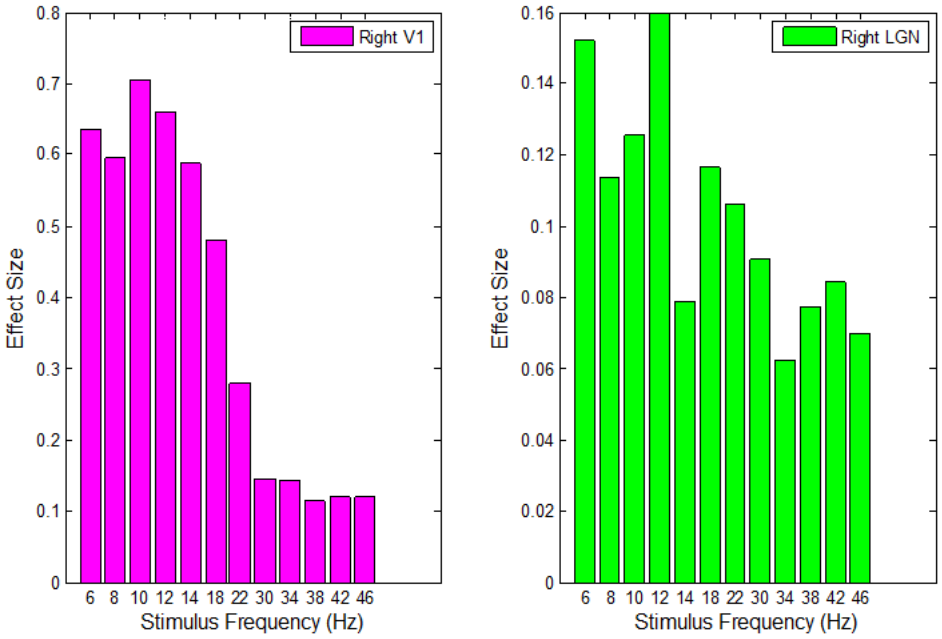

Figure 6.18 Comparison of mean effect sizes wrt stimulus frequency of right V1 (magenta) and right LGN (green).

7. DISCUSSION

It was suggested by Friston et al. [32], having enough scans and subjects, the potential of false positives is more likely with classical inference where a voxel with a trivially small activation can be declared significant due to the high degrees of freedom. Although a small effect size could be interesting in any other statistical test, in neuroimaging we are not interested in trivially small activations. Since we have a large number of subjects, we propose that using a random effects analysis over a group of 40 subjects with Bayesian inference enhanced the reliability of our results.

Simulation results also showed that PEB approach is a better estimator for the majority of the voxels. The choice of the prior is an important issue in Bayesian inference. For the real data, we used the subject-wise variation of effect sizes about the mean effect size as prior to estimate posterior distribution of each voxel.

The frequentist approach is based on the rejection of the null hypothesis, whereas the Bayesian inference is more flexible to ask question about the values of the parameters from the posterior distiributions without reference to a null distribution. Because SPM uses conventional testing methodology, we used PPMs to map the posterior probability of each voxel. The resulting PPMs showed a global maximum response at 10 Hz, that is in line with the previous studies [43]. They yield activations even at high stimulation frequencies, despite being more limited in extent compared with those at lower frequencies. Band-wise PPMs showed that the largest extent of activation is also in the alpha band.

We specified the effect size thresholds for the PPMs as zero (figures 6.4-6.16, left), so that the effects of any size are shown with a colormap, if significant. This type of thresholding was performed for several reasons: Firstly, because this is what one does in classical statistics where each significant effect is of interest, we could compare the PPMs with SPMs. Secondly, decreasing the effect size threshold enabled us to infer

about the activations in some specific regions.

In a visual motion processing study, it was reported that SPMs missed a critical part of the V5 complex [34]. Similarly in our study, it is clearly seen that SPMs identified a smaller number of voxels than PPMs. This can be explained with the multiple comparison problem that is induced by the classical approach: If one increases the search volume, the threshold will rise that is adjusted by Gaussian field correction. In contrast, we did not need to make a correction when we applied PPM so that the inference in a region does not depend on the inspection of another region.

The visual information from the retina reaches the visual cortex (VC) through the LGN. By decreasing the effect size threshold of the PPMs, we could make inference about the responses in LGN without lowering the significancy. Having an effect size threshold of zero, we could compare the effect sizes between LGN and V1. PPMs showed that the effect sizes in LGN are less than that in V1 and the functional connectivity between LGN and V1 is explicit during graded visual stimulation.

ROI analyses yield that the effect sizes in LGN have a global maximum at 12 Hz, whereas V1 has a maximum at 10 Hz. Although different stimuli have been used, studies in the macaque brain suggest that at low temporal frequencies, there is a little contribution of magnocellular pathway to visual sensitivity [44]. For both hemisphere, LGN has the local maximum at 18 Hz, which can be explained with the contribution of magnocellular units being more sensitive to stimuli modulated at frequencies nearer 20 Hz [45].

Bayesian estimation also provides asking questions about the differences between stimuli. Although F-contrasts can be used for a comparison between different stimuli, in fact it simply says whether there is a significant difference between stimuli without saying which one is different [25]. To assess the monotonicity of parameter estimates, the use of multiple t-tests for all pairs of stimuli would lead to a Bonferroni correction for multiple comparisons. Instead, within a Bayesian context one can compute the monotonicity, $P(\beta_1 < \cdots < \beta_n | y)$ directly [30].

Similar inferences were performed for the first-level analyses. A Variational Bayes algorithm was used to compute the posterior distributions of each voxel where the autoregressive coefficients and the precision of the error enter as the priors [46]–[47]. Hence, the LGN activation can be detected successfully for each subject. However, the update equations of the posteriors make the algorithm computationally expensive so that the parameter estimation took very long hours for a single subject. For this reason, we preferred a classical estimation at the first level.

REFERENCES

1. Woolrich, M. W., E. J. T. Behrens, C. F. Beckmann, M. Jenkinson, and S. M. Smith, "Multilevel linear modeling for fmri group analysis using bayesian inference," *Neuroimage*, Vol. 21, pp. 1732–1747, 2004.

2. Deichmann, R., U. Nöth, and N. Weiskopf, "The basics of functional magnetic resonance imaging," in *EEG-fMRI: Physiological Basis, Technique and Applications* (Mulert, C., and L. Lemieux, eds.), pp. 39–62, Berlin Heidelberg: Springer-Verlag, 2010.

3. Boynton, G. M., S. A. Engel, G. H. Glover, and D. J. Heeger, "Linear systems analysis of funtional magnetic resonance imaging in human v1," *Journal of Neuroscience*, Vol. 16, pp. 4207–4221, 1996.

4. Buxton, R. B., "The elusive initial dip," *Neuroimage*, Vol. 13, pp. 953–958, 2001.

5. Friston, K. J., C. D. Frith, R. Turner, and R. S. J. Frackowiak, "Characterizing evoked hemodynamics with fmri," *Neuroimage*, Vol. 2, pp. 157–165, 1995.

6. Lindquist, M. A., and T. D. Wager, "Validity and power in hemodynamic response modeling: A comparison study and a new approach," *Human Brain Mapping*, Vol. 28, p. 764â784, 2007.

7. Goutte, C., F. A. Nielsen, and L. K. Hansen, "Modeling the haemo-dynamic response in fmri using smooth fir filters.," *IEEE Transactions on Medical Imaging*, Vol. 19, pp. 1188–1201, 2000.

8. Caulhoun, V. D., G. D. P. M C Stevens, and K. A. Kiehl, "fmri analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms.," *Neuroimage*, Vol. 22, pp. 252–257, 2004.

9. Friston, K. J., O. Josephs, G. Rees, and R. Turner, "Nonlinear event-related responses in fmri.," *Magnetic Resonance in Medicine*, Vol. 39, pp. 41–52, 1998.

10. Aguirre, G. K., E. Zarahn, and M. D'Esposito, "The variability of human, bold hemodynamic responses.," *Neuroimage*, Vol. 8, pp. 360–376, 1998.

11. Bernstein, M. A., K. E. King, and X. J. Zhou, *Handbook of MRI Pulse Sequences.*, Amsterdam; Boston: Academic Press, 2004.

12. Soltysik, D., R. Brigs, K. White, K. Peck, and B. Crosson, "Comparison of hemodynamic response nonlinearity across primary cortical areas.," *Neuroimage*, Vol. 22, pp. 1117–1127, 2004.

13. Talavage, T. M., and W. B. Edmister, "Nonlinearity of fmri responses in human auditory cortex.," *Human Brain Mapping*, Vol. 22, no. 3, pp. 216–228, 2004.

14. Sotero, R. C., C. Roberto, and N. J. Trujillo-Barreto, "Biophysical model for integrating neuronal activity, eeg, fmri and metabolism.," *Neuroimage*, Vol. 39, no. 1, pp. 290–309, 2008.

15. Pugh, K., "Auditory selective attention: An fmri investigation.," *Neuroimage*, Vol. 4, no. 3, pp. 159–173, 1996.

16. Gilbert, S. J., P. W. Burgess, M. L. Scholvinck, J. S. Simons, and C. D. Frith, "Differential components of prospective memory? evidence from fmri.," *Neuropsychologia*, Vol. 44, no. 8, pp. 1388–1397, 2006.

17. Ganis, G., "Brain areas underlying visual mental imagery and visual perception: an fmri study.," *Cognitive Brain Research*, Vol. 20, no. 2, p. 226 â 241, 2004.

18. Lawrance, N. S., R. Hoffmann, T. J. Ross, E. A. Stein, and H. Garavan, "Multiple neuronal networks mediate sustained attention.," *Journal of Cognitive Neuroscience*, Vol. 15, pp. 1028–1038, 2003.

19. Huettel, S. A., A. W. Song, and G. McCarthy, *Functional magnetic resonance imaging.*, Sunderland, Mass: Sinauer Associates, Publishers, 2004.

20. Buckner, R. J., P. A. Bandettini, K. M. O'Craven, R. L. Savoy, S. E. Petersen, M. E. Raichle, and B. R. Rosen, "Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging," in *Proc. Natl. Acad. Sci.*, Vol. 93, (USA), pp. 14878–14883, 1996.

21. Visscher, K. M., F. M. Miezina, J. E. Kelly, R. L. Buckner, D. Donaldson, M. P. McAvoy, V. M. Bhalodia, and S. E. Petersen, "Mixed blocked/event-related designs separate transient and sustained activity in fmri.," *Neuroimage*, Vol. 19, p. 1694â1708, 2003.

22. Collignon, A., F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multi-modality image registration based on information theory.," in *Proc. Information Processing in Medical Imaging* (Bizais, Y., C. Barillot, and R. D. Paola, eds.), (Dordrecht, The Netherlands), pp. 263–274, Kluwer Academic Publishers, 1996.

23. Friston, K. J., J. T. Ashburner, S. J. Kiebel, T. E. Nichols, and W. D. Penny, eds., *Statistical Parametric Mapping: The Anlaysis of Functional Brain Images*, Academic Press, 2006.

24. Evans, A. C., D. L. Collins, and B. Milner, "An mri-based stereotactic atlas from 250 young normal subjects.," *Society of Neuroscience Abstracts*, Vol. 18, no. 408, 1992.

25. Jezzard, P., P. M. Matthews, and S. M. Smith, eds., *Functional MRI: An Introduction to Methods.*, Oxford: Oxford University Press, 2002.

26. Friston, K. J., K. J. Worsley, R. S. J. Frackowiak, J. C. Mazziotta, and A. C. Evans, "Assessing the significance of focal activations using their spatial extent.," *Human Brain Mapping*, Vol. 1, pp. 214–220, 1994.

27. Friston, K. J., and K. J. Worsley, "Analysis of fmri time-series revisited - again.," *Neuroimage*, Vol. 2, pp. 173–181, 1995.

28. Turin, G. L., "An introduction to matched filters.," *IRE Transactions on Information Theory*, Vol. 6, no. 3, pp. 311– 329, 1960.

29. Worsley, K. J., A. C. Evans, and S. Marrett, "A three-dimensional statistical analysis for rcbf activation studies in human brain.," *J Cereb Blood Flow Metab*, Vol. 12, pp. 900–18, 1992.

30. Neumann, J., and G. Lohmann, "Bayesian second-level analysis of functional magnetic resonance images.," *Neuroimage*, Vol. 4, p. 58â73, 2003.

31. Friston, K. J., D. E. Glaser, R. N. A. Henson, S. J. Kiebel, C. Phillips, and J. Ashburner, "Classical and bayesian inference in neuroimaging: Applications.," *Neuroimage*, Vol. 16, pp. 484–512, 2002.

32. Friston, K. J., W. D. Penny, C. Phillips, S. J. Kiebel, and G. Hinton, "Classical and bayesian inference in neuroimaging: Theory.," *Neuroimage*, Vol. 16, pp. 465–483, 2002.

33. Penny, W. http://www.fil.ion.ucl.ac.uk/ wpenny/.

34. Friston, K. J., and W. D. Penny, "Posterior probability maps and spms.," *Neuroimage*, Vol. 19, no. 3, pp. 1240–1249, 2003.

35. Bayram, A., Z. Bayraktaroğlu, E. Karahan, B. Erdoğan, B. Bilgiç, M. Özker, I. Kaşıkçı, A. Ademoğlu, C. Öztürk, K. Arıkan, N. Tarhan, and T. Demiralp, "Simultaneous EEG-fMRI analysis of the resonance phenomena in steady-state visual evoked responses.," *Clin EEG Neuroscience*, Vol. 42, no. 2, pp. 98–106, 2011.

36. http://imaging.mrc-cbu.cam.ac.uk/imaging/SpmBatch5.

37. Somers, D. C., A. M. Dale, A. E. Seiffert, and R. B. Tootell, "Functional mri reveals spatially specific attentional modulation in human primary visual cortex.," in *Proceedings of the National Academy of Science*, Vol. 96, (USA), p. 1663â1668, 1999.

38. Buchel, C., R. Turner, and K. J. Friston, "Lateral geniculate activations can be detected using intersubject averaging and fmri.," *Magnetic Resonance in Medicine*, Vol. 38, pp. 691–694, 1997.

39. Chen, W., T. Kato, X.-H. Zhu, J. Strupp, S. Ogawa, and K. Ugurbil, "Mapping of lateral geniculate nucleus activation during visual stimulation in human brain using fmri.," *Magnetic Resonance in Medicine*, Vol. 39, pp. 89–96, 1998.

40. Chen, W., and X.-H. Zhu, "Correlation of activation sizes between lateral geniculate nucleus and primary visual cortex in humans.," *Magnetic Resonance in Medicine*, Vol. 45, pp. 202–205, 2001.

41. Eickhoff, S. B., K. E. Stephan, H. Mohlberg, C. Grefkes, G. R. Fink, K. Amunts, and K. Zilles, "A new spm toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data.," *Neuroimage*, Vol. 25, p. 1325â1335, 2005.

42. Logothetis, N. K., "The ins and outs of fmri signals.," *Nature Neuroscience*, Vol. 10, no. 10, pp. 1230–32, 2007.

43. Wan, X., J. Riera, K. Iwata, M. Takahashi, W. T ., M., and R. Kawashima, "The neural basis of the hemodynamic response nonlinearity in human primary visual cortex: Implications for neurovascular coupling mechanism.," *NeuroImage*, Vol. 32, pp. 616–625, 2006.

44. Merigan, W. H., and H. R. Maunsell, "Macaque vision after magnocellular lateral geniculate lesions.," *NeuroImage*, Vol. 5, no. 4, pp. 347–352, 1990.

45. Derrington, A. M., and P. Lennie, "Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque.," *J Physiol.*, Vol. 357, p. 219â240, 1984.

46. Friston, K. J., R. N. A. Henson, C. Phillips, and J. Mattout, "Bayesian estimation of evoked and induced responses.," *Human Brain Mapping*, Vol. 27, pp. 722–735, 2005.

47. Penny, W. D., S. J. Kiebel, and K. J. Friston, "Variational bayesian inference for fmri time series.," *Neuroimage*, Vol. 19, no. 3, pp. 727–741, 2003.

48. Worsley, K. J., "Local maxima and the expected euler characteristic of excursion sets of x2, f and t fields.," *Adv Appl Prob*, Vol. 26, pp. 13–42, 1994.

49. Worsley, K. J., S. Marrett, and P. Neelin, "A unified statistical approach for determining significant signals in images of cerebral activation.," *Hum Brain Mapp*, Vol. 4, p. 58â73, 1996.

50. Penny, W. D., N. Trujillo-Bareto, and K. J. Friston, "Bayesian fmri time series analysis with spatial priors.," *Neuroimage*, Vol. 24, no. 2, pp. 350–362, 2005.

51. Daunizeau, J., K. J. Friston, and S. J. Kiebel, "Variational bayesian identification and prediction of stochastic nonlinear dynamic causal models.," *Physica D: nonlinear phenomena*, Vol. 238, no. 21, pp. 2089–2118, 2009.