A MIXED EFFECT MODEL WITH FEATURE EXTRACTION FOR FUNCTIONAL MAGNETIC RESONANCE IMAGING (FMRI) DATA

by

Sooyoung Kim

(Under the Direction of Cheolwoo Park)

ABSTRACT

The Primary goal of this research is an application of feature extraction and a mixed model to functional Magnetic Resonance Imaging (fMRI) data. The goal of study is a comparison of multiple groups of subjects when they conduct a cognitive task. Since fMRI data of interest are the information of stimulus-response reactions from human brain activity over time, they show repeated patterns in the signals. Therefore, we use the feature extraction method that collects the characteristics or patterns of data. Then, we apply a mixed model that includes both fixed and random effects to find any group difference. Through a simulation study we find a mixed model with feature extraction approach effective for detecting a difference between groups. Finally, we have applied the approach on the11 regions of interest in human brain from the cognitive task fMRI data, and found that the region called Striatum shows significant difference.

INDEX WORDS: Functional magnetic resonance imaging data, Feature extraction, Mixed model, Region of Interest.

A MIXED EFFECT MODEL WITH FEATURE EXTRACTION FOR FUNCTIONAL MAGNETIC RESONANCE IMAGING (FMRI) DATA

by

Sooyoung Kim

B.S., Iowa State University, 2012

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2017

A MIXED EFFECT MODEL WITH FEATURE EXTRACTION FOR FUNCTIONAL MAGNETIC RESONANCE IMAGING (FMRI) DATA

by

Sooyoung Kim

Major Professor:	Cheolwoo Park
Committee:	Paul Schliekelman
	Ray Bai

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2017

ACKNOWLEDGEMENTS

I would like to thank my entire committee Dr. Cheolwoo Park, Dr. Paul Schliekelman, and Dr. Ray Bai for their time and patience in shepherding me through the process. I would particularly like to thank Dr. Park, my main advisor, for his consistent and unwavering encouragement that helped me focus on the light at the end of the tunnel. In the processing of this work, I met a lot of questions and doubts. Although Dr. Park was very busy in his teaching, researching and coordinator working, he patiently explained many complex concepts to me every time through his strong background and profound knowledge. In addition, Dr. Park shared many previous experiences to me and he always guided me in the right direction whenever he thought I needed it. All of those encouraged me to accomplish this thesis. On a more personal level, there are several people that deserve acknowledgement. I am grateful for the support of my parents and sister which has enabled me to pursue and achieve my goals.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS iv

LIST OF TABLES vi

LIST OF FIGURES vii

CHAPTER

1 Introduction 1

2 Functional Magnetic Resonance Imaging (fMRI) Data 3

2.1 Functional Magnetic Resonance Imaging (fMRI) 3

2.2 fMRI Data 5

3 Methods 9

3.1 Mixed Model 9

3.2 Feature Extraction 11

4 Data Simulation 15

4.1 Data Simulation 15

4.2 Result 18

5 Real Data Analysis 21

5.1 Feature Extraction 21

5.2 Mixed Model Analysis 26

6 Summary 28

BIBLIOGRAPHY 30

LIST OF TABLES

Page

Table 1: Regions of Interest. *6*

Table 2: Data Example by General Group of ROI1. ..... *7*

Table 3: Data Example by Specific Group of ROI1. ..... *7*

Table 4: Display Example of Simulated Data: S(1, 2, 3) in SNR(1). *16*

Table 5: Feature Extraction Data Example: S(1, 2, 3) in SNR(1). *18*

Table 6: Mixed Model Analysis: S(1, 1, 1) in SNR(0.5, 1, 1.5). *19*

Table 7: Mixed Model Analysis: S(1, 2, 3) in SNR(0.5, 1, 1.5). *19*

Table 8: Mixed Model Analysis: S(1, 1.5, 2) in SNR(0.5, 1, 1.5). *20*

Table 9: Feature Extraction Data Example: ROI1_SEF. *26*

Table 10: Mixed Model Results. *27*

LIST OF FIGURES

Page

Figure 1: Shape of the Hemodynamic Response for a single stimulus.....*4*

Figure 2: Displays an example of BOLD signals in ROI1 .....*8*

Figure 3: Feature Extraction example: General group of ROI1 .....*13*

Figure 4: Data Simulation Plots (Design 1, Design 2, Design 3) *17*

Figure 5: fMRI Time Series Plot for ROI1_ SEF. *22*

Figure 6: fMRI Time Series Plot for ROI2_ FEE. *22*

Figure 7: fMRI Time Series Plot for ROI3_ SPL. *22*

Figure 8: fMRI Time Series Plot for ROI4_ Cuneus. *23*

Figure 9: fMRI Time Series Plot for ROI5_ Thalamus. *23*

Figure 10: fMRI Time Series Plot for ROI6_ IPL. *23*

Figure 11: fMRI Time Series Plot for ROI7_ PFC-L. *24*

Figure 12: fMRI Time Series Plot for ROI8_ PFC-R. *24*

Figure 13: fMRI Time Series Plot for ROI9_ Striatum. *24*

Figure 14: fMRI Time Series Plot for ROI10_ MOG.....*25*

Figure 15: fMRI Time Series Plot for ROI11_ IFC.....*25*

Chapter 1

Introduction

The functional magnetic resonance imaging (fMRI) system that is advanced from magnetic resonance imaging system, known as MRI, is focused on detection of neuronal activities in human brain. Regulation of cognitive control such stimulates neuronal activity on specific region of the brain is detected as rates of blood-oxygenation-level dependent (BOLD) contrast. So thus, fMRI data collect BOLD contrast rates over time, and collected data have unique patterns that follow a specific task conducted in a scanner. Often, a group comparison is of interest in an fMRI study where each group of subjects conducts a different type of task. In such case, an appropriate statistical approach should be applied to capture these differences. This is often challenging because fMRI data are known to have a low signal-to-noise ratio (SNR), and thus these differences are hard to be detected (LeBihan et al., 1993). Driven by the need to detect increasingly subtle differences between groups, a statistical method must be sensitive enough to find these differences.

Feature extraction is a process of extracting main characteristics or patterns in a signal. Because typical fMRI data related to a certain task show peaks (stimulus on) and valleys (stimulus off) by design, a feature extraction can be used to summarize the data using a statistic such as mean or median from peaks and valleys. This is useful especially when fMRI data exhibit low SNR because a direct group comparison based on raw BOLD signals can be difficult. By feature extraction, characterized by extraction of differences between averages or medians of the peak of

a cycle in our analysis, we can convert a function comparison problem into a group comparison under a well-known repeated measurement setting, which can be analyzed by a mixed model. A mixed model approach contains both fixed and random effects, which allow us to detect a group difference (fixed) while taking subject variation (random) into account.

To examine the proposed approach, we create simulated data that mimic fMRI data. In simulation, we compare 3 different groups with 10 subjects each under 3 SNRs. We also try different 3 different curve settings. In real fMRI data where subjects conduct a cognitive task, we have two different task groups with 16 subjects each, and attempt to find which areas show a group difference in 11 regions of interest in the brain. We rearrange the data by feature extraction method, and analyze the extracted data using a mixed model for each ROI. We discover that there is the one region of interest of brain shows statistical significance between two groups according to our approach.

This thesis is organized as follows. In Chapter 2, we provide a brief review of fMRI data and describe the data we analyze. In Chapter 3, we review our approach, mixed model and feature extraction. In Chapter 4, we illustrate how we generate the simulated data and report the simulation results under various settings. In Chapter 5, we analyze the fMRI data described in Section 2.2, and discuss the obtained results from our feature extraction and mixed model approach. We finally conclude in Chapter 6.

Chapter 2

Functional Magnetic Resonance Imaging (fMRI) Data

For clear understanding of our research project, we introduce general information of fMRI, and describe the data we analyze.

2.1 Functional Magnetic Resonance Imaging (fMRI)

Functional magnetic resonance imaging (fMRI) is a one of the MRI scan that is specialized for brain imaging. A MRI scanner has built with a strong static magnet that creates the magnetic field, and it detects atomic particles such as proton, electron, and neurons. Since human bodies have abundance of tissues that consist of these small particles, the MR scanner system is useful to study them.

As a principle of MR scanner, fMRI is focused on the detection of neuronal activity. By responses to a specific stimulus, our neuronal system promotes the high tension on blood pressure, so that it increases blood flow to the part of brain. In human blood, there is called hemoglobin that carries the diamagnetic property in a magnetic field. In 1936, the Nobel laureate Pauling and his student Coryell discovered that hemoglobin weakly repels a magnetic field when it is attached to oxygen (Hb). On the other hand, hemoglobin without oxygenation (dHb) causes it to become magnetic. Thus, deoxygenated hemoglobin has a paramagnetic property in a magnetic field that we use to identify the brain regions of fMRI studies.

The intensity of magnetization in brain, also called blood-oxygenation-level dependent (BOLD) contrast, is measured based on the rate between oxygenated hemoglobin (Hb) and deoxygenated

hemoglobin (dHb). Once the neuronal activity increases in specific part of the brain, a glucose level increases highly in the area of stimuli with blood flow. The level of glucose also indicates the level of oxygen supply to the area, which indicates much more oxygen consumption by neurons. Thus, the Hb level with oxygen has a regular rate in blood vessel in general, so we can measure contrast between Hb and dHb (Huettel et al., 2009).

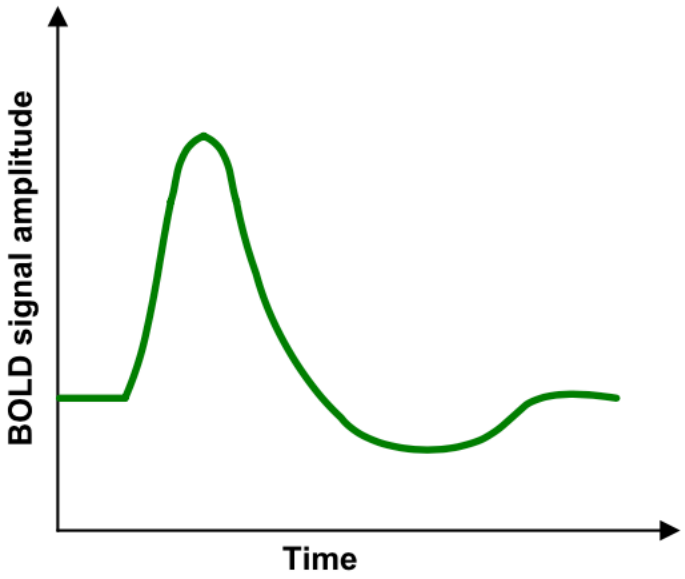

Figure 1 Shape of the hemodynamic response for a single stimulus.

The fMRI data are hemodynamic responses that are the changes in the ratio of oxygenated to deoxygenated blood. The shape of the hemodynamic response depends on the properties of stimuli and the related brain functions. Figure 1 summarizes a hemodynamic response function (HRF), which is a time lag of changes in the hemodynamic response after the stimulus is given.

Further, the MRI technology has finally advanced to obtain a functional image of the brain by measuring the BOLD signal relative to a particular stimulus. Also, comparing with other imaging technologies such as X-ray and CT, fMRI has advantages of low invasiveness (non-surgery),

lack of radiation exposure, and relatively wide availability. Therefore, fMRI, a neuronal imaging technology, is now dominating at clinics and hospital.

2.2. FMRI Data

Typically, BOLD fMRI signals are reduced in cognitive control and attentional networks over time (Chein & Schneider, 2005). With cognitive control, we can build a map that indicates the changes of stimulus-response in our task groups. Between trials of prosaccade (look toward a cue) and antisaccade (look away from a cue), low and high level cognitive controls are advanced to a result of the direction of fMRI BOLD signal changes over time. The study combines these records in practice of saccade tasks with a probability trial that examines how antisaccade and prosaccade trials in different tasks are affected by either specific or general saccade practice. (Pierce & Mcdowell, 2017)

We have 32 participants who completed an initial fMRI session with blocks (cycles) of antisaccade to prosaccade trials and then were divided into two task groups for 4 days of saccade practice in the laboratory before an identical posttest fMRI session. During the 4 days, half of the participants are practiced the mixed probability (0%, 25%, 50%, and 100%) blocks each day (specific group). Here, for example, 25% means 25% of events consist of antisaccade tasks. The other half of the participants practiced only blocks of 100% antisaccade trials (general group). Our main interest is to test whether these two groups show different brain response before and after practice in some regions of human brain.

Each subject is measured in 160 time points of fMRI session, and we particularly interested in comparing two task groups at 11 bilateral neural regions of interest (ROI) that are established by an anatomical criterion as in the Table 1. According to Dyckman etl al. (2007), these areas are

the main components of the saccade-related neural circuitry based on previous fMRI studies of saccadic performance.

Table 1 Regions of Interest

ROIs'	Regions of Interest	Abbreviation
ROI1	Supplementary eye field	SEF
ROI2	Frontal eye field	FEF
ROI3	Superior parietal lobe	SPL
ROI4	Cuneus	Cuneus
ROI5	Thalamus	Thalamus
ROI6	Inferior parietal lobule	IPL
ROI7	Left PFC	PFC-L
ROI8	Right PFC	PFC-R
ROI9	Striatum (Basal Ganglia)	Striatum
ROI10	Middle occipital gyrus	MOG
ROI11	Right inferior frontal cortex	IFC

The collected raw fMRI data are properly pre-processed (see Pierce & McDowell (2017) for more details) and de-trended to remove trends typically generated from the scanner machine. Also, we take the average over voxels in each ROI (typically there are 30-70 voxels for each ROI), which yields 16 (subjects) by 160 (time points) observations for both specific and general groups. Exemplary observations are presented in Tables 2 and 3.

Figure 2 displays an example of BOLD signals for both general and specific groups in ROI1. We subtract two curves before and after practice, and take the average over subjects to create this plot. The goal of this study is to determine whether there is a difference between two task groups at each ROI.

Table 2 General group (100%) Data Example (ROI1)

Obs	X0	X0.1	X0.2	…..	X0.166	X0.167	X0.168
1	1.2135	-0.4726	-0.8796	…..	0.0798	-0.1810	-0.8961
2	0.3231	-0.4239	-0.5881	…..	0.1169	-0.5143	-1.0483
3	0.3252	4.3563	4.2135	…..	-2.0467	-1.9811	-1.3034
:				:			:
14	0.6071	1.0074	-0.4537	…..	0.8834	0.8904	-0.2745
15	-0.0282	-0.5860	-0.6728	…..	0.6415	0.0996	-0.4948
16	0.0792	0.1731	-0.1428	…..	-1.5260	-1.8521	-1.7350

Table 3 Specific Group (100%) Data Example (ROI1)

Obs	X0	X0.1	X0.2	…..	X0.166	X0.167	X0.168
1	-0.0572	0.3141	0.5405	…..	-1.2103	-0.9022	-0.4693
2	-0.0434	-0.1938	0.2628	…..	-0.2163	-0.1437	-0.0808
3	-0.2981	-0.7969	-0.6260	…..	0.5427	0.3754	0.2970
:				:			:
14	0.6071	1.0074	-0.4537	…..	-0.1159	0.5594	1.2574
15	-0.0282	-0.5860	-0.6728	…..	0.2014	0.3602	0.1248
16	0.0792	0.1731	-0.1428	…..	-0.6365	-0.9360	-0.8177

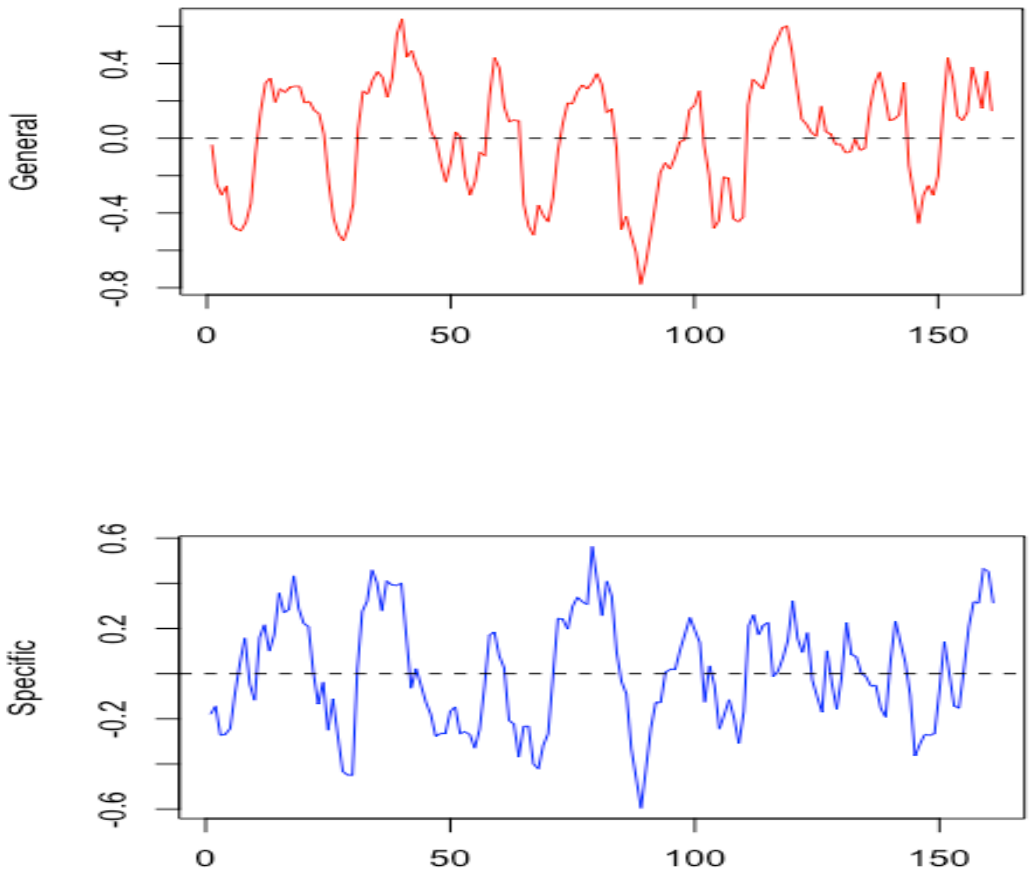

Figure 2 Displays an example of ROI1 BOLD signals (Top-General, Bottom-Specific).

Chapter 3

Methods

In this chapter, we introduce two methods, mixed models and feature extraction.

3.1 Mixed Model

A mixed model, or mixed effect model, is the statistical model which includes both fixed and random effects, such we known a “fixed” as the parameters of explanatory variables that we are interested in specific, and a “random” as the level of factors that is the variables of randomly picked from the population of research interested.

For instance, in our fMRI data, there are two task groups, each compared pre- & post-tests, and the trials are collected as antisaccade and prosaccade from 16 different examinees at each group. The 16 selected examinees are representatives for our experiment because any one among all examinees can be chosen. Hence, the factor of task group is a fixed effect and the factor of subject is a random effect. Since the method of an analysis is affected by whether a factor is treated as fixed or random, inappropriate classification of factors may lead to an inaccurate result.

In a mixed model, the observations in different blocks (or cycles) are assumed to have different variance, and the observations in the same group have equal variance. Hence, the two sources of variation are considered; the variation between the groups and the other variation within the groups.

Mixed models are typically in the analysis of longitudinal data, which is useful in settings of repeated measurements on the experimental units. The longitudinal data, where each time series constitutes an individual curve (Demidenko, 2004). The repeated measures are defined as data obtained repeatedly on each experiment unit. From difference between longitudinal data and repeated measures, the repeated measures are not necessarily obtained over time. By Demidenko (2004), "A mixed model is well suited for biological and medical data, which display notorious heterogeneity of responses to stimuli and treatment. An advantage of the mixed model is the ability to genuinely combine the data by introducing multilevel random effects".

The standard form of the mixed model is

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

where,

- y is an n × 1 response vector, and n is the number of observations.
- X is an n × p fixed effect design matrix with p fixed effect coefficients.
- β is a p × 1 vector of fixed effect vector.
- Z is an n × q random effect design matrix with q random effect coefficients.
- b is a q × 1 random effect vector.
- ε is an n × 1 error vector.

We assume that random effect vector, b, and error vector, ε, follows;

$$b \sim N\big(0, \sigma^2 D(\theta)\big),$$

$$\varepsilon \sim N(0, \sigma^2 I),$$

respectively, where D is asymmetric matrix parameterized by a variance component vector θ, I is the n × n identity matrix, and σ^2 is the error variance. Furthermore, we assume that the random effect vector b and the error vector ε are independent each other. And it makes correct comparisons between task groups with variations, which is the variations between groups instead of the overall measurement errors.

In addition, fMRI data are highly correlated temporally since stimuli are given continuously or periodically over time, and thus the response to the stimulus at the current scan time point will be influenced by the stimulus at the previous scan time point, and the response to it (Lazar, 2008). A typical approach to address this issue of analysis is to assume a temporal model for characterizing the dependence structure in fMRI time series (Worsley, 2003). This temporal correlation has been evaluated using different types of time series models ranging from autoregressive of low-order to long-range dependence (Bullmore et al., 2003; Park et al., 2010). In our analysis, we use autoregressive structure for mixed models, which gives homogeneous variances and correlations that decline exponentially with time period.

3.2 Feature Extraction

The feature extraction is the recognition and extraction of the pattern of complex data that involves reducing the amount of resources required to describe a large data set. In general, analysis with a large data set causes over-fit with classification algorithm, and it leads to poor generalization to new data. With feature extraction, final combinations of variables from the pattern recognition and reduction get rid of these problems where the data lead to sufficient accuracy.

The most important step of feature extraction is the imaging processing that is used to detect and isolate the portion or shapes of the data we are interested in. For the imaging processing, we can use any types of information that we can recognize as the characteristics such as waves, skewedness, and any type of patterns. There are many different techniques of imaging process, but, we focus at structural pattern recognition, which we also use in our data analysis. The goal of this structural recognition is to simplify the representation of the pattern into more meaningful and easier features. In our fMRI data, we can see that peaks and valleys are regularly present in signals, and these peaks and valleys give cycles of the signals that represents feature extracted values.

Structural pattern recognition is originally based on its morphological interconnections present within the data. This method is effective for the data with identifiable patterns and shape such as image data and time series data. However, the method is limited as a consequence of complications associated with the implementation of the description task because there is no general solution to extract the structural features from data. As a result, this extraction is focused on both simplest and most generic as possible, and finally, existing structural pattern recognition must be connected to morphological features which are correctly established as being effective for the original data under analysis.

Although we extract features that characterize our data, there is overall exemption from complete analysis with the extracted data from the original data set. This is because of the redundancy in features may cause the transform from original set of the data. Therefore, the determination of a subset of the features must be careful and be tested with relevant simulation instead of the direct analysis of complete initial data set.

In our mixed model analysis, a feature extraction is approached with detection of peaks and valleys of curve that determines the difference across defined populations or factors of interest (Morris, 2012; Morris et al., 2005). As illustrated in Figure 2, we use the mean or median values from peak and valley of each cycle as extracted features representing the stimulus-response periods. The representative of the cycle can be condensed into one value by using the difference between those two averages or medians.

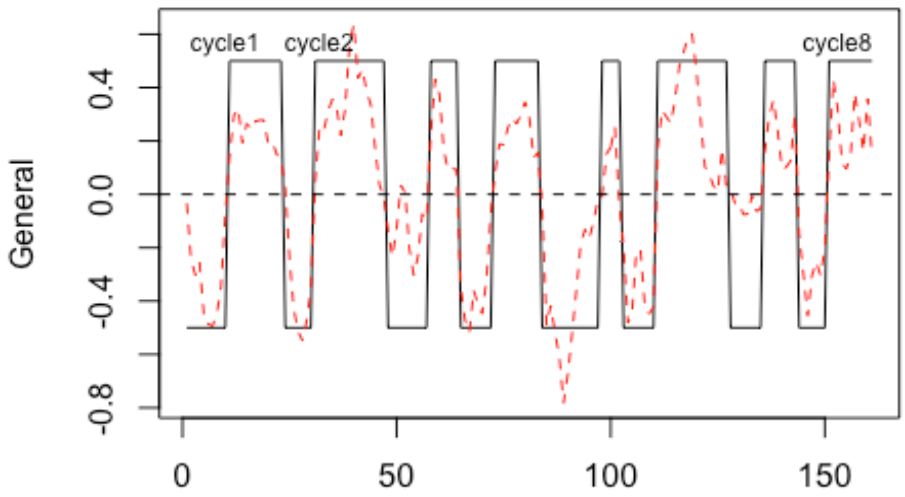

Figure 3 Feature extraction example: General group of ROI1

Thus, for a given subject in a practice group, each cycle, which has mostly 12 responses as can be seen from Figure 2, is represented by one summary value, that is the difference between means or medians, extracted from the original twelve values. With the feature extraction approach, the data structure, finally, becomes that each subject has eight extracted values (repeated measures), i.e., one value for each cycle.

Denote the response variable (extracted features) for the j-th subject at the k-th cycle in the i-th task group by Y_{ijk}, then the mixed model can be written as,

$$Y_{ijk} = \mu + \alpha_i + a_{j(i)} + e_{ijk}$$

where, μ is the grand mean, α_i is the fixed effect of the i-th task group, $a_{j(i)}$ is the random effect associated with the j-th subject in the i-th task group, e_{ijk} is the random error associated with the j-th subject at the k-th cycle in the i-th task group:

$i = 1,2$, $j = 1,...,n_i$, $n_1 = 16$ (general), $n_2 = 16$ (specific), and $k = 1, \ldots, 8$.

We assume $a_{j(i)} \sim i.i.d.\ N(0, \sigma_i^2)$ and e_{ijk} is normally distributed with $E(e_{ijk}) = 0$, $V(e_{ijk}) = \sigma^2$, and, $Cov(e_{ijk}, e_{ijk'}) = \sigma^2\rho$ for $k' \neq k$ for all i and j. Since ρ is defined by $\sigma_i^2/[\sigma^2 + \sigma_i^2]$, which corresponds to autocorrelation of autoregressive model with order 1, AR(1). Note that,

$$E(y_{ijk}) = \mu + \alpha_i,$$

$$Var(y_{ijk}) = \sigma_i^2 + \sigma^2$$

$$Cov(e_{ijk}, e_{ijk'}) = \sigma^2\ Var(e_{ijk}).$$

With the results of application of the final model based on the feature extraction approaches, we can analyze whether the differences between task groups are significant for each ROI. Finally, the test results show which ROIs have significant differences between task groups for both mean and median values.

Chapter 4

Simulation

In this chapter, we introduce our data simulation designs, and analysis results by feature extraction and mixed models.

4.1 Data Simulation

From fMRI data, we find there are certain features/patterns that represent by terms of fixed and random effects, which we base these to our design of simulation data. To signify our simulation, we set the 3 different design models corresponding to different signals with variation on their noises, i.e. Signal-to-Noise Ratio (SNR).

The SNR is defined as the ratio of the power of a signal and the power of background noise (Sohn, 1999):

$$SNR = \sigma^2_{signal} / \sigma^2_{noise}$$

where $\sigma^2$is a variance of signal or noise.

We conduct 3 sets of our model simulation based on the following model:

$$X_{it} = S_{it} + Z_{it}, \qquad i = 1, \ldots, 10\ (subjects), t = 1, \ldots, 100(time\ points)$$

I. Design I: Signals of (1 to -1) with *SNR*(0.5, 1, 1.5)

- S_i: each S_i with 5 cycles of the same signal varying between -1 and 1.

- $Z_i \sim N(0, \frac{Var(S_i)}{j})$, *i = 1,2,3,* $j_{(SNR)}$ *= 0.5, 1, 1.5.*

II. Design II: Signals of (1 to -1), (2 to -2), and (3 to -3) with *SNR*(0.5, 1, 1.5)

- S_i: each S_i with 5 cycles of (1 to -1), (2 to -2), and (3 to -3). *i = 1,2,3.*

- $Z_i \sim N(0, \frac{Var(S_i)}{j})$, *i = 1,2,3,* $j_{(SNR)}$ *= 0.5, 1, 1.5.*

III. Design III: Signals of (1 to -1), (1.5 to -1.5), and (2 to -2) with *SNR*(0.5, 1, 1.5)

- S_i: each S_i with 5 cycles of (1 to -1), (1.5 to -1.5), and (2 to -2), *i = 1,2,3.*

- $Z_i \sim N(0, \frac{Var(S_i)}{j})$, *i = 1,2,3,* $j_{(SNR)}$ *= 0.5, 1, 1.5.*

Each design has three 3 task groups indexed by *i* and is variated with 3 sets of *SNR* in a same repeated-measurement of signals, and different repeated measurement of signals with 10 subjects for each. As a result, the data consist of (30 x 100) observations.

Table 4 Displays examples of simulated data S(1,2,3) with SNR(1); 30 x 100 obs.

	x1	**x2**	**x3**	**…..**	**x97**	**x98**	**x99**	**x100**
S1obs1	-0.455	1.237	1.019	…..	-2.567	-1.936	-2.411	-0.934
S1obs2	1.122	1.861	1.101	…..	-1.700	-1.324	-2.395	-1.637
:				:				:
S2obs1	1.356	0.623	-3.356	…..	1.729	-0.945	-3.809	-0.617
:				:				:
S3obs9	-3.002	3.146	-1.684	…..	1.460	-5.458	-4.717	-4.070
S3obs10	0.590	6.412	1.193	…..	-11.564	-3.924	-3.823	0.231

Next, we plot the data from each simulation design that shows a clear view of signals and cycles. The plot is conducted by the mean values of each point of S_i.

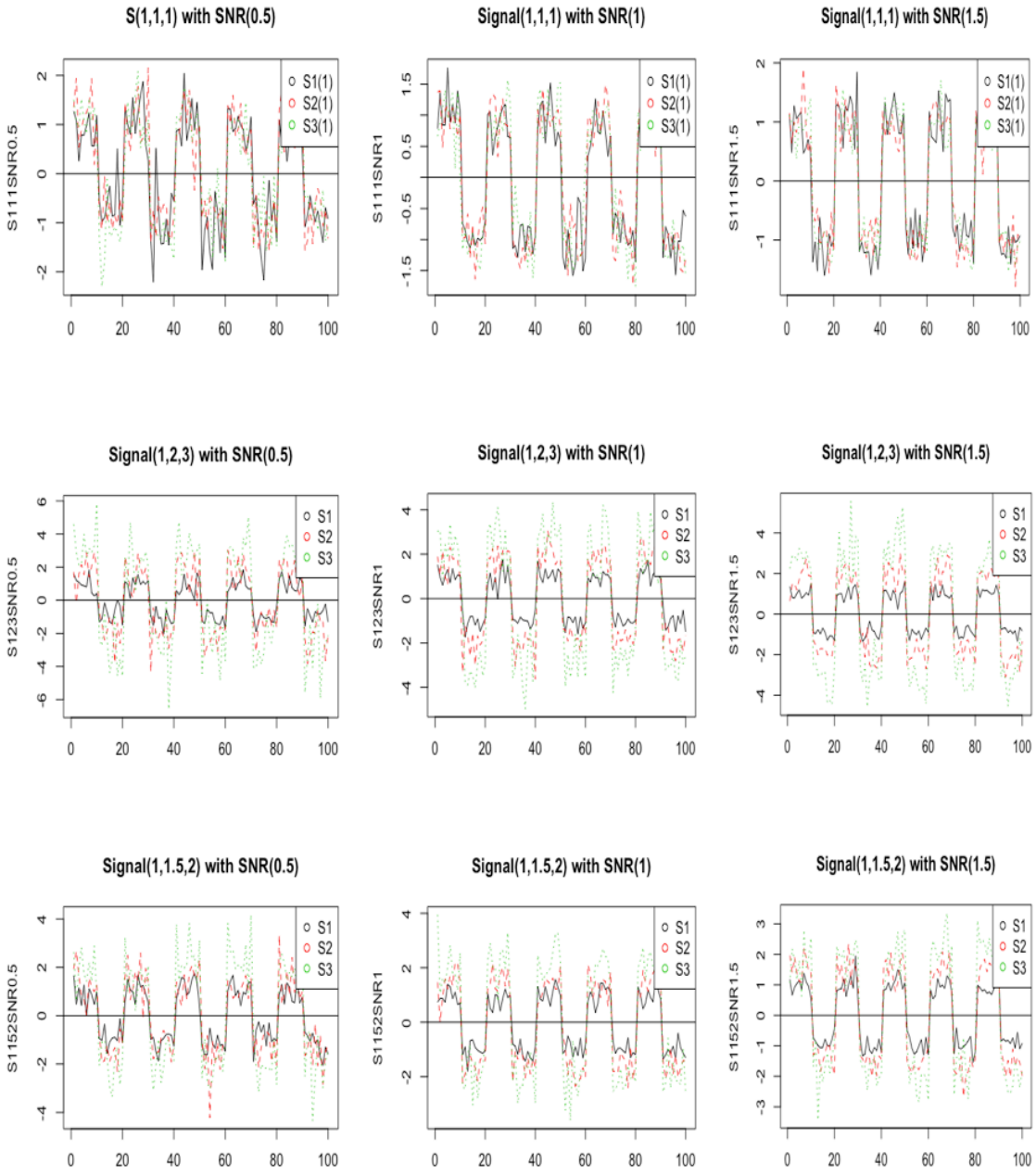

Figure 4 Data Simulation Plots (From top to bottom, Design 1, Design 2, Design 3)

Using the feature extraction procedure, we measure the mean or median at each peak and valley and take a difference between these two in each cycle. Table 5 shows examples of the extracted data.

Table 5 Data Example: S(1,2,3) with SNR(1)

person	task	cycle	AvgDiff	MedDiff
1	1	1	1.837696	1.555361
1	1	2	2.619058	2.307668
1	1	3	1.795086	2.278572
1	1	4	2.113585	1.893811
1	1	5	1.670764	1.957463
2	1	1	2.103943	1.625988
2	1	2	2.628392	2.666016
2	1	3	1.995191	1.991915
:	:	:	:	:
:	:	:	:	:

The observed signal *Y*'s are rearranged in cycles that consist of extracted values from peaks and valleys of signals. As a result, we have 5 cycles for each person. Hence, these differences are responses, task group with 3 levels is fixed effect, and subject (10) is random effect, and 5 cycles are repeated measurements in our mixed model. As mentioned in Section 3.1, we use the AR structure for the covariance of the random effect.

4.2 Results

From the mixed model analysis, we have 3 results of both fixed and random effects by variation of SNR. Here, AvegDiff, and MedDiff indicate Average Difference, and Median Difference. Thus, we compare 6 analysis results from each simulation design.

Table 6 Mixed Model Analysis: S(1, 1, 1) with SNR(0.5, 1, 1.5)

Obs	S(1,1,1), SNR(0.5)	Effect	NumDF	DenDF	FValue	ProbF
1	AvgDiff	task	2	32.9	0.60	0.5559
2	MedDiff	task	2	22.4	0.75	0.4852
Obs	**S(1,1,1), SNR(1)**	**Effect**	**NumDF**	**DenDF**	**FValue**	**ProbF**
1	AvgDiff	task	2	34.9	0.36	0.7005
2	MedDiff	task	2	38	0.02	0.9792
Obs	**S(1,1,1), SNR(1.5)**	**Effect**	**NumDF**	**DenDF**	**FValue**	**ProbF**
1	AvgDiff	task	2	21	0.79	0.4674
2	MedDiff	task	2	34.2	1.62	0.2137

From the first simulation design, $S(1, 1,1)$ with $SNR(0.5, 1, 1.5)$, all of the analysis results show non-significance of the task group effect as expected. Since there is no difference between S_i's in our first design, there should be no task effect among the three groups. In addition, these results are consistent for all levels of *SNR*.

Table 7 Mixed Model Analysis: S(1, 2, 3) with SNR(0.5, 1, 1.5)

Obs	S(1,2,3), SNR(0.5)	Effect	NumDF	DenDF	FValue	ProbF
1	AvgDiff	task	2	28.3	131.65	<.0001
2	MedDiff	task	2	27.3	90.54	<.0001
Obs	**S(1,2,3), SNR(1)**	**Effect**	**NumDF**	**DenDF**	**FValue**	**ProbF**
1	AvgDiff	task	2	27.6	214.19	<.0001
2	MedDiff	task	2	28.4	146.07	<.0001
Obs	**S(1,2,3), SNR(1.5)**	**Effect**	**NumDF**	**DenDF**	**FValue**	**ProbF**
1	AvgDiff	task	2	23.7	290.10	<.0001
2	MedDiff	task	2	24.2	328.02	<.0001

Table 8 Mixed Model Analysis: S(1, 1.5, 2) with SNR(0.5, 1, 1.5)

Obs	S(1,1.5,2), SNR(0.5)	Effect	NumDF	DenDF	FValue	ProbF
1	AvgDiff	task	2	40.9	63.00	<.0001
2	MedDiff	task	2	35.5	47.15	<.0001
Obs	**S(1,1.5,2), SNR(1)**	**Effect**	**NumDF**	**DenDF**	**FValue**	**ProbF**
1	AvgDiff	task	2	31.6	98.67	<.0001
2	MedDiff	task	2	34.8	69.77	<.0001
Obs	**S(1,1.5,2), SNR(1.5)**	**Effect**	**NumDF**	**DenDF**	**FValue**	**ProbF**
1	AvgDiff	task	2	23.5	171.89	<.0001
2	MedDiff	task	2	32.7	111.73	<.0001

For the results of last two designs, $S(1, 2, 3)$ and $S(1, 1.5, 2)$ with $SNR(0.5, 1, 1.5)$, since the design built on variation of S_i's, all task effects are statistically significant on all possible combinations of simulation designs. We repeat this simulation numerous times and have consistent results.

In summary, our data simulation shows that the mixed model analysis with feature extraction is significantly appropriate to use in various simulation settings with low and high *SNR*s. Thus, next, we apply mixed model analysis to our real fMRI data.

Chapter 5

Real Data Analysis

In this chapter, we present our data analysis results for the real data described in Section 2.2. Our process is conducted as follows.

(i) At each ROI, we start from the de-trended and averaged over voxels. Hence, each subjects has two curves, i.e. before and after practice.

(ii) We subtract these two time series and obtain one curve for each subject.

(iii) We extract the features from the subtracted time series for each subject. We use mean and median values to extract the features representing the stimulus-response periods in fMRI data.

(iv) We analyze the extracted features using a mixed model with repeated measurements.

5.1 Feature Extraction

From Tables 2 and 3 in Section 2.2, there are two task groups, specific and general groups for each region of interest (ROI). For the analysis using mixed models, we need to rearrange the fMRI time series by feature extraction. We create cycles based on peaks of signals. In this process, cycles are determined based on the observations near by a peak or a valley for each group. As a result, number of observations per cycle is varying for different ROIs and groups.

From ROI1 to ROI11, most signals are conducted with 8 cycles for both groups. However, in ROI5 (Thalamus), the signals have no specific patterns after the 4th cycle. Hence, the features are extracted with 1 to 4th cycles only at ROI5 groups.

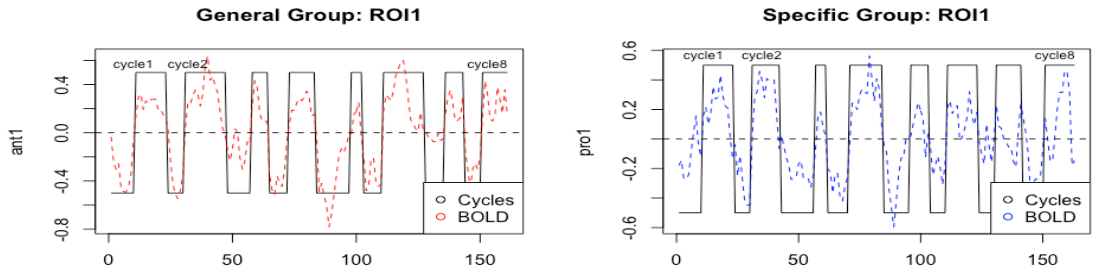

Figure 5 ROI1_ SEF: 8 cycles.

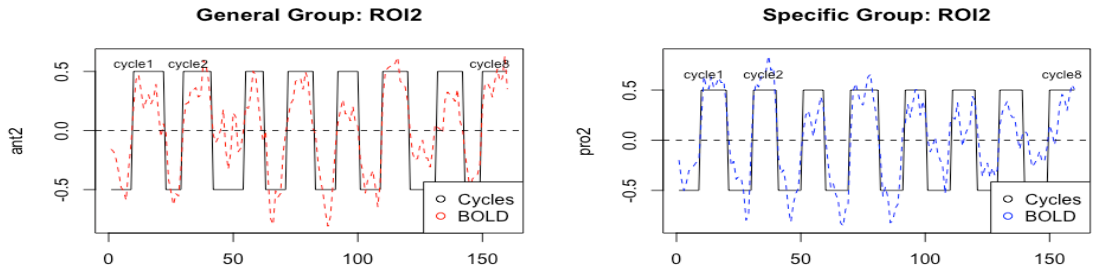

Figure 6 ROI2_FEF: 8 cycles.

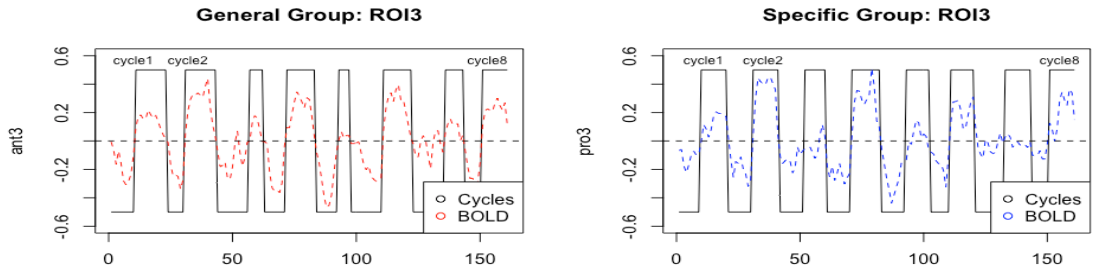

Figure 7 ROI3_SPL: 8 cycles.

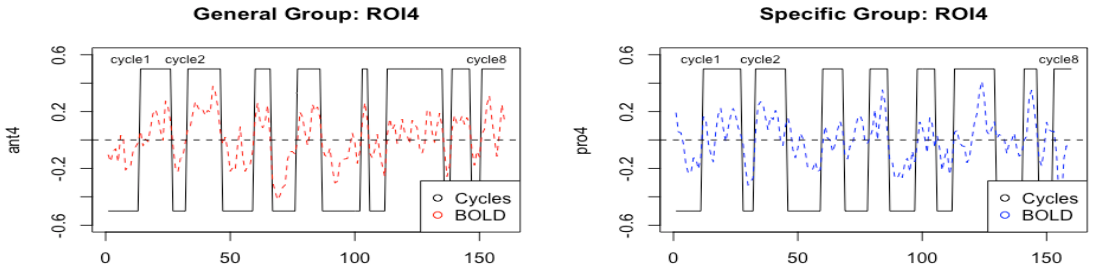

Figure 8 ROI4_Cuneus: 8 cycles.

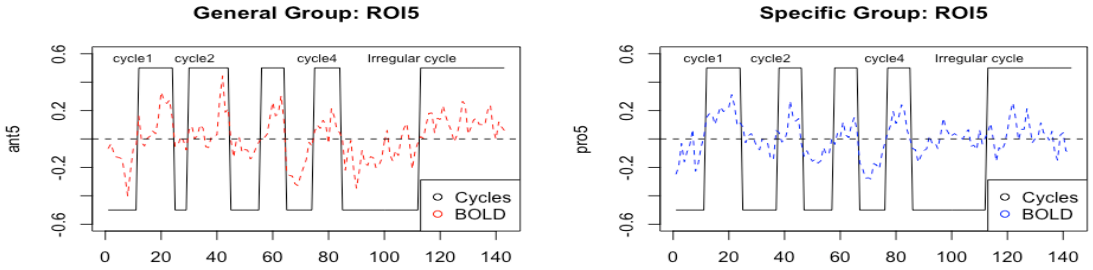

Figure 9 ROI5_Thalamus: 4 cycles.

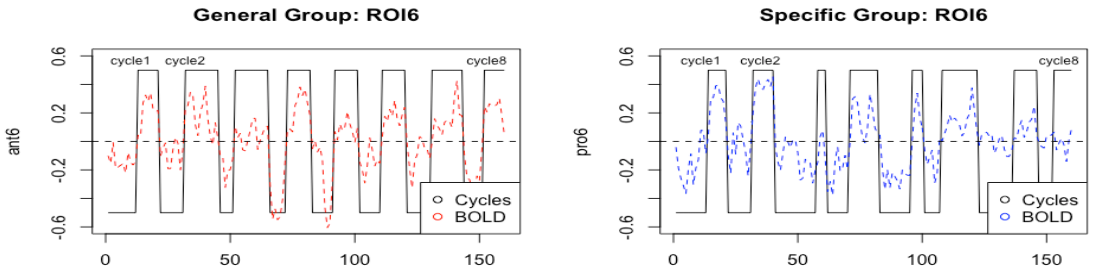

Figure 10 ROI6_IPL: 8 cycles

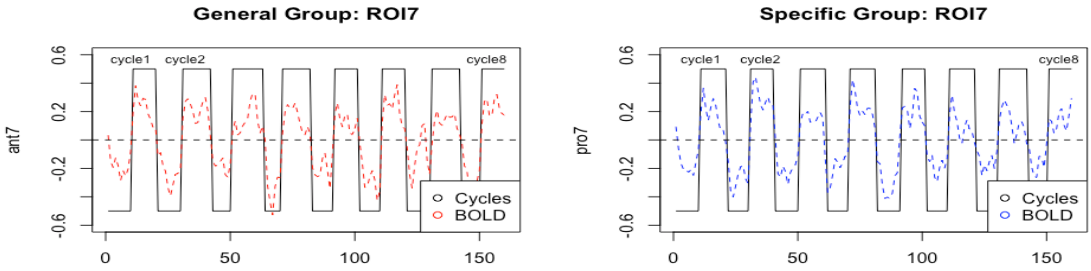

Figure 11 ROI7_PFC-L: 8 cycles.

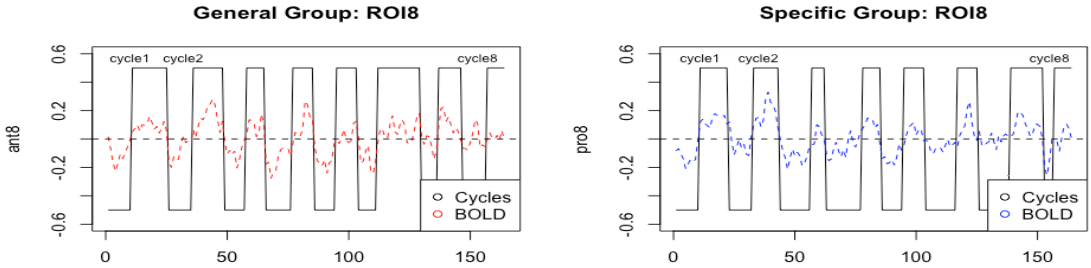

Figure 12 ROI8_PFC-R: 8 cycles

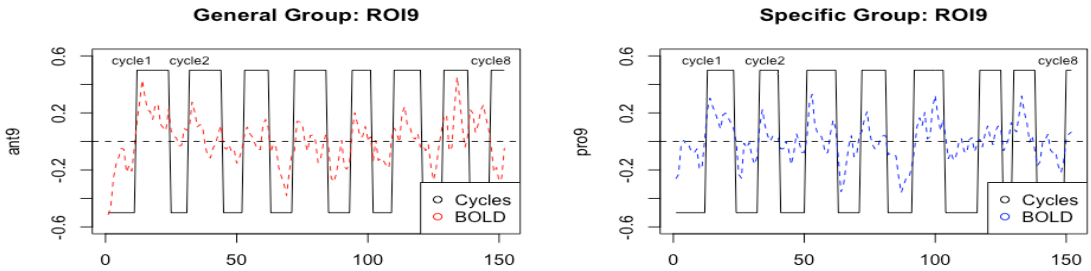

Figure 13 ROI9_Striatum: 8 cycles.

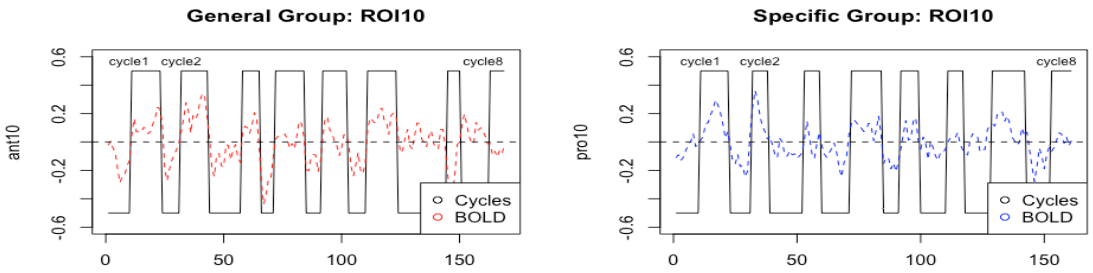

Figure 14 ROI10_MOG: 8 cycles.

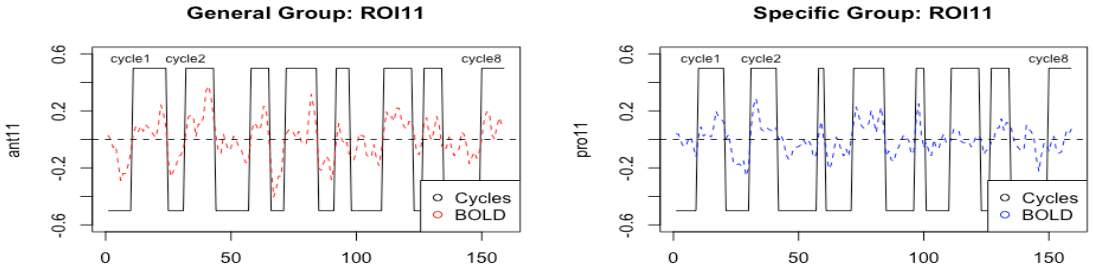

Figure 15 ROI11_IFC: 8 cycles.

The average time series plots for each BOLD signal (dash lines) are drawn for the 2 task groups in Figures 5-15. Black lines represent time points of task blocks for each group.

In summary, we rearrange the data using the feature extraction method. The signals from each group at each ROI are rearranged in cycles that consist of peak and valley, and each cycle contains different number of observations based on the cycles composed. Next, to fit a mixed model, we measure the mean or median of the peaks and valleys, and take a difference between these values for each cycle. Hence, these differences are responses, 2 levels of tasks, which are specific and general groups, are for fixed effect, and 16 subjects for each group are for random effect, and 8 cycles (4 cycles, ROI5) are repeated measurements in our mixed model.

Table 9 shows an example of the extracted data at ROI 1.

Table 9 Feature Extraction Data Example (ROI1)

person	task	cycle	avgdiff	meddiff
1	1	1	-1.2927075	-1.2376377
1	1	2	-0.7310966	-0.6536747
:	:	:	:	:
:	:	:	:	:
16	2	7	0.4117672	0.4347927
16	2	8	-0.3457455	-0.2741749

5.2 Mixed Model Analysis

From feature extraction, we now conduct statistical analysis of mixed model for each region of interest.

The results of application of this mixed model based on the feature extraction approach are summarized in Table 10. F-tests and p-values indicate whether the differences between task groups are significant for each ROI. Large F-values imply that there are significant differences between task groups. The test results show that Stratium (ROI9) has significant differences between 2 task groups at significance level $\alpha = 0.05$ for mean values and at significance level $\alpha = 0.1$ for median values.

Table 10 reports the results from the mixed model analysis for all ROIs.

Table 10. Mixed model results

ROIs	Mean		Median	
	F-value	p-value	F-value	p-value
SEF	0.56	0.4568	0.57	0.4529
FEF	0.02	0.8973	0.24	0.6246
SPL	0.09	0.7634	0.27	0.6047
Cuneus	< 0.01	0.9659	< 0.01	0.9956
Thalamus	0.14	0.7111	0.23	0.6326
IPL	0.46	0.5028	0.14	0.7128
PFC-L	0.11	0.7414	0.22	0.644
PFC-R	< 0.01	0.9724	0.05	0.8231
Striatum	**4.18**	**0.0496**	**3.15**	**0.0852**
MOG	1.96	0.1722	1.31	0.2623
IFC	0.09	0.7590	0.41	0.5212

Chapter 6

Summary

The purpose of this research is an application of a mixed model with feature extraction to functional magnetic resonance imaging (fMRI) data where a comparison of groups is the main interest. Feature extraction captures a prominent pattern from fMRI data, and converts the analysis into a well-known problem for group comparison. Analysis in evaluation of the mixed model consists of fixed and random effects, represented as two tasks as fixed (specific and general groups) and subjects as random. Therefore, we evaluate the group difference using the fixed effect while accounting for the subject variability by the random effect.

In the simulation study, it is evidenced that the mixed model with feature extraction approach correctly detects a group difference if there exists and yields non-significant results when there is no group difference.

For the real fMRI data analysis, the only region of interest, Striatum, shows a significant difference between two task groups at significance level $\alpha = 0.1$ when we use mean and median values for feature extraction. Since the data are the collection of neuronal activities, we can conclude that the brain responses are significantly different before and after practice between specific and general groups in the Striatum region of human brain. We have not found any statistically significant results for the other 11 ROIs using our approach.

For future study, we consult with psychologists whether our findings are consistent with their previous studies. Also, we plan to apply our mixed model with feature extraction approach to other types of fMRI data where a group analysis is of interest.

Bibliography

[1] LeBihan, D., Turner, R., Ze_ro, T., Cuenod, C., Jezzard, P., and Bonnerrot, V. (1993). Activation of human primary visual cortex during visual recall: a magnetic resonance imaging study. *Proceedings of the National Academy of Sciences*, 90:11802-11805.

[2] Huettel, S. A., Song, A. W., and McCarthy, G. (2009). *Functional Magnetic Resonance Imaging*. Sinauer Associates Inc., Sunderland, MA, second edition.

[3] Chein, J. M., & Schneider, W. (2005). Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning. *Brain Research. Cognitive Brain Research*, 25, 607–623.

[4] Pierce J. and McDowell J. (2017). Reduced Cognitive Control Demands after Practice of Saccade Tasks in a Trial Type Probability Manipulation, *Journal of Cognitive Neuroscience*, 2, 368-381.

[5] Dyckman, K. A., Camchong, J., Clementz, B. A., and McDowell, J. E. (2007). An effect of context on saccade-related behavior and brain activity. *NeuroImage*, 36(3):774−784.

[6] Demidenko, E. (2004). *Mixed Models: Theory and Applications.* John Wiley & Sons, Inc., Hoboken, New Jersey.

[7] Lazar, N. A. (2008). *The Statistical Analysis of Functional MRI Data*. Springer, New York.

[8] Worsley, K. (2003). Detecting activation in fMRI data. *Statistical Methods in Medical Research*, 12(5), 401–418.

[9] Bullmore, E., Suckling, J., Zelaya, F., Long, C., Honey, G., Routledge, C., Ng, V., Fletcher, P., and Williams, S., (2003). Wavelets and statistical analysis of functional magnetic resonance images of the human brain. *Statistical Methods in Medical Research*, 12(5), 375– 399.

[10] Sohn, J., (1999). A Statistical Model-Based Voice Activity Detection*, IEEE Signal Processing Letters*, 6,1-3.

[11] Park, C., Lazar, N., Ahn, J., and Sornborger, A. (2010). A Multiscale Analysis of the Temporal Characteristics of Resting-State fMRI Data, *Journal of Neuroscience Methods*, 193, 334-342.

[12] Morris, J. S. (2012). Statistical methods for proteomic biomarker discovery based on feature extraction or functional modeling approaches. *Statistics and Its Interface*, 5(1), 117–136.

[13] Morris, J. S., Coombes, K. R., Koomen, J., Baggerly, K. A., and Kobayashi, R. (2005). Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics*, 21(9), 1764–1775.