

Supervised Learning Models of fMRI Data for Inferring Brain Function and Predicting Behavior

by

Yunzhi Wang

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
May 4, 2014

Keywords: *fMRI, Granger causality, Effective connectivity, Brain networks, Support Vector Machine, Neuroeconomics*

Copyright 2014 by Yunzhi Wang

Approved by

Gopikrishna Deshpande, Chair, Assistant Professor, Electrical & Computer Engineering
Thomas S Denney, Director, Auburn University MRI Research Center
Veena Chattaraman, Associate Professor, Department of Consumer & Design Sciences

Abstract

The development of fMRI has revolutionized cognitive neuroscience. There are two related areas gaining increasing interest: 1) Investigating the directional interactions between different regions. 2) Predicting human behaviors from brain activities. In this thesis, supervised learning models were applied on fMRI data for solving these problems. Firstly, dynamic Granger causality, a regression based supervised learning model, was experimentally demonstrated to be capable of inferring stimulus-evoked sub-100ms timing difference in fMRI responses, providing a reliable data-driven method for effective connectivity analysis of fMRI data. Secondly, Patel's τ – a method which performed best for inferring directional interactions in a previous simulation – was investigated using experimental fMRI data, highlighting the necessity of experimental validation of simulation results. Lastly, recursive cluster elimination based support vector machine, a classification based supervised learning model, was used to predict purchase decisions using spatio-temporal fMRI features, providing a reliable framework for using fMRI data to predict purchase-related decisions.

Acknowledgments

First, I would like to thank my advisor, Dr. Gopikrishna Deshpande. He led me to the interesting world of fMRI, taught me how to be a qualified researcher and spent lots of time reading and revising my papers. I really learned a lot from him. I am deeply indebted to him for his great patience and wonderful advices throughout the work of thesis.

I would like to thank Dr. Thomas Denney and Dr. Veena Chattaraman, who take time to serve on my graduate committee and review my work. I would like to thank my friends in AU MRI center, especially Hao and Karthik, who helped me a lot at the beginning of my graduate study. Without them it would be more difficult to finish the thesis.

I would like to thank Mengdi, who always accompanies and encourages me whenever I am happy or sad.

Finally, I would like to express my deepest gratitude to my parents for their unconditional love and support. They are the power of keeping me going forward.

Table of Contents

Abstract	ii
Acknowledgments.....	iii
List of Tables	vii
List of Figures.....	viii
Chapter 1: Introduction	1
1.1 MRI.....	1
1.2 Functional MRI.....	2
1.3 Functional and Effective connectivity	3
1.4 Supervised Learning	4
1.4.1 Granger causality and Effective connectivity	5
1.4.2 Support Vector Machine and Brain state classification.....	6
1.5 Motivation and Organization	8
Chapter 2: Experimental Validation of Dynamic Granger Causality for Inferring Stimulus- evoked Sub-100ms Timing Differences from fMRI.....	10
2.1 Introduction.....	11
2.2 Methods	12
2.2.1 Data acquisition	12
2.2.2 Dynamic Granger Causality Analysis.....	13
2.2.3 Covariance of Connectivity with Experimental Paradigm	15

2.2.4 Comparison with cross-correlation function.....	16
2.3 Results.....	17
2.4 Discussion.....	20
2.5 Conclusion	23
Chapter 3: Experimental evidence demonstrating the inability of Patel's τ for estimating directionality of brain networks from fMRI	24
3.1 Introduction.....	25
3.2 Methods	27
3.2.1 Animal model selection	27
3.2.2 FMRI/EEG data acquisition and processing.....	28
3.2.3 IEEG experiments and data analysis.....	29
3.2.4 Patel's conditional dependence measures	30
3.3 Results.....	32
3.4 Discussion.....	36
Chapter 4: Predicting Purchase Decisions based on Spatio-temporal Functional MRI Features using Machine Learning	39
4.1 Introduction.....	40
4.2 Method	44
4.2.1 Experiment design and data acquisition	44
4.2.2 ROI selection and feature extraction.....	45
4.2.3 Recursive Cluster Elimination Based Support Vector Machine Classifier	48
4.3 Results.....	51
4.3.1 Prediction accuracy	51
4.3.2 Important spatio-temporal features for classification	52

4.4 Discussion.....	54
4.5 Conclusion	56
Chapter 5: Conclusion.....	57
Bibliography	58

List of Tables

Table 2.1	18
Table 2.2	19
Table 4.1	47
Table 4.2	52

List of Figures

Figure 1.1	1
Figure 1.2	7
Figure 2.1	13
Figure 2.2	17
Figure 2.3	18
Figure 2.4	20
Figure 3.1	33
Figure 3.2	34
Figure 3.3	35
Figure 3.4	36
Figure 4.1	44
Figure 4.2	47
Figure 4.3	50
Figure 4.4	51
Figure 4.5	52
Figure 4.6	53

Chapter 1: Introduction

1.1 MRI

Magnetic Resonance Imaging (MRI) is a noninvasive medical imaging technique using high magnetic fields and wave pulses instead of ionizing radiations or radioactive tracers to image the structures inside the body. When a patient is positioned inside the MRI scanner which forms a strong magnetic field, the randomly spinning nuclei will align with the direction of the magnetic field. Three gradient coils are then used to choose the orientation of the slices in the three directions; the nuclei at different locations will rotate at different speeds because of the spatial variance of the magnetic field. The hydrogen atoms will get excited and emit a radio frequency signal when the RF energy is applied at the appropriate resonant frequency (Larmour frequency). These MR signals detected at the receiver are the mixture of RF signals with different amplitudes, frequencies and phases containing spatial information. Inverse Fourier transform is then applied to recover the spatial information and reconstruct the image of scanned area [1].



Figure 1.1 MRI scanner

Compared with other medical imaging techniques, MRI has three primary advantages [2]: 1) It has the potential of getting very high spatial resolution images for both bone and soft tissue. 2) Ionizing radiation is not required as X-rays or CT scans. 3) It could get images in any plane through the body. Therefore, MRI has become one of the most popular diagnostic imaging techniques over the past two decades.

1.2 Functional MRI

Functional magnetic resonance imaging is a neuroimaging technique using standard MRI scanner to investigate the neuronal changes in brain function over time [2]. The measurement of brain activity is mainly based on the blood oxygenation level dependent (BOLD) contrast. It relies on the fact that Cerebral Blood Flow (CBF) and neuronal activation are normally coupled. Whenever a brain region is activated either simultaneously or driven by some tasks, it will demand more oxygen. The demanded oxygen is carried to neurons by hemoglobin in capillary red blood cells, leading to the increase of blood flow in that region which can be detected by MRI. The change in MR signal from neuronal activation is called hemodynamic response (HDR). HDR always lags 1 to 2 seconds after the neuronal events triggering it, and takes another 5 seconds to rise to a peak.

The spatial resolution of an fMRI image is determined by its voxel dimensions, which are specified by three parameters [2]: field of view, matrix size, and slice thickness. Full-brain experiments will use larger voxel size, while those focusing on the changes in specific regions of interest (ROIs) will use smaller ones. Spatial resolution of fMRI could be as small as the order of millimeter. Temporal resolution of fMRI scan is usually

between 1 second and 2 second. Therefore, fMRI has a poor temporal resolution compared with some other neuroimaging techniques such as electroencephalography (EEG) and magnetoencephalography (MEG), but it has an attractive high spatial resolution and thus has been extensively used in both research and clinical applications.

1.3 Functional and Effective connectivity

One area of rapidly increasing interest in neuroimaging is the mapping of brain network [3]. Different brain regions are assumed to perform different brain functions, while many neuronal processes cannot be localized in a single region and hence are presumed to be encoded by a network formed by several brain regions. Such “mapping” usually starts by defining a set of function nodes [3]. In the context of fMRI, nodes are defined as specific regions of interest (ROIs). Once nodes are identified, various approaches are taken to estimate the edges (connections) between the nodes, using the experimental time courses in ROIs. The most straightforward method may be looking at the correlation between the time courses of the node pair. However, correlation cannot indicate the directionality of the node pair, or whether the connection of this node pair is direct or indirect [3]. Generally, The approaches estimating the interaction between brain regions can be classified into two categories: functional connectivity and effective connectivity. Functional connectivity is defined as “temporal correlations between spatially remote neurophysiological events” while effective connectivity is defined as “the influence one neuronal system exerts over another” [4]. Typically the estimation of directionality of influence is harder than just estimating whether a connection exists or not, but always of greater interest as it provides a mechanistic characterization of the underlying neuronal

processes in terms of information flow. Both functional and effective connectivity can be estimated by data-driven or model-driven approaches.

1.4 Supervised Learning

A supervised learning model is to reason from the external instances given with known output to build a general hypothesis, which is then used to make predictions about future instances [5]. Supervised learning is the process of learning the inherent rules from the training data, creating a classifier or regressor that can be applied to generalize from future instances for prediction of their outputs. There are several steps of supervised learning process. The first step is to collect training data and select features that may be informative for prediction. Feature subset selection is usually the second step to remove the irrelative features and reduce the dimensionality of data. Next part is algorithm selection. There are many approaches proposed for supervised learning. Most commonly used algorithms include artificial neural networks (ANN) [6], support vector machine (SVM) [7], regression models, etc. A particular algorithm will be chosen and performed on the training data. After that, cross-validation is often used in order to estimate the performance of the predictor, by dividing the training data into two exclusive subsets, one for training and the other for testing. Supervised learning models have a broad application in many areas. In the context of fMRI, they can be used for inferring brain function and predicting behavior.

1.4.1 Granger causality and Effective connectivity

Granger causality [8] is an autoregressive (AR) model firstly proposed for assessing the ‘causality’ of different time series in the context of economics. Given two time series X_t and Y_t , if the past values of one time series can help predict the current and future values of the other, then we can say that the former Granger-cause the latter. AR models are used for the estimation of Granger causality. Individually,

$$X_t = \sum_{j=1}^p a_{1j} X_{t-j} + e_{1t}$$
$$Y_t = \sum_{j=1}^p d_{1j} Y_{t-j} + e_{2t}$$

Then bivariate AR models are used for consideration of cross correlation.

$$X_t = \sum_{j=1}^p (a_{2j} X_{t-j} + b_{2j} Y_{t-j}) + e_{3t}$$
$$Y_t = \sum_{j=1}^p (c_{2j} X_{t-j} + d_{2j} Y_{t-j}) + e_{4t}$$

Where a and d represent autocorrelation of each time series, b and c represent cross-correlation between time series. E represent estimation errors (or noises). P is the order of the model, which may be determined by Bayesian Information Criterion (BIC) [9]. Then Granger causality can be calculated from the estimation errors.

$$F(X \rightarrow Y) = \log \frac{\text{var}(e_2)}{\text{var}(e_4)}, F(Y \rightarrow X) = \log \frac{\text{var}(e_3)}{\text{var}(e_2)}$$

The variance ratio cannot be less than 1, because the introduction of additional parameters in the model cannot lead to an increase of estimation errors. Therefore, Granger causality exists on the interval $(0, \infty)$, representing the degree to which one time series can help predict the other.

Previous studies have demonstrated that when applied to electrophysiological data, Granger causality is capable of getting interpretable results in terms of both the directionality and the magnitude of synaptic transmissions [10, 11]. However, in the context of fMRI, the application of Granger causality for estimating effective connectivity is still debated. There are three factors that could have the potential to confound the results of Granger causality: (i) Hemodynamic variability across different brain regions. (ii) Low temporal resolution. (iii) Low Signal-to-noise ratio SNR [12].

However, despite these concerns, highly interpretable results applying Granger causality in fMRI appeared both in simulation works [13, 14] and experimental works [15, 16]. Therefore, the issue of to what extent Granger causality can be applied to fMRI still needs for further considering.

1.4.2 Support Vector Machine and Brain state classification

Support vector machine (SVM) is a supervised learning algorithm developed by Vapnik [7] to solve the classification problems. The goal in a classification problem is to separate different classes by a function which learns the implicit rules by the training data, such that it will be able to assign a novel unlabeled input into a correct class. The fundamental assumption is that samples of the same class will have similar values in feature space, and

thus be near to each other. Therefore, a decision hyperplane can be used to separate the different classes in feature space, as fig.2.2 shows.

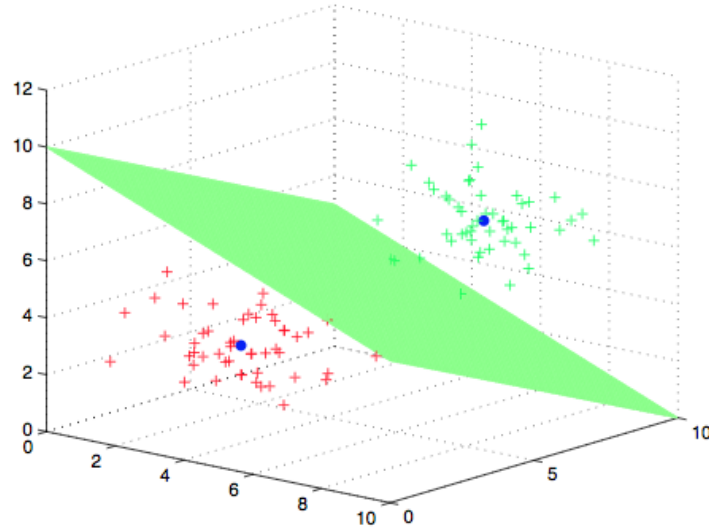


Figure.1.2 An illustration of a decision plane in a three-dimension feature space. This figure is adapted from [22]

The goal of a linear SVM classifier is to find the optimal linear hyperplane in the feature space with the largest margin, since larger margin always means better generalization of the classifier. Given a linear separable training data set (\mathbf{x}_i, y_i) , where \mathbf{x}_i is the input features of the i th sample and y_i is a binary value (either 0 or 1) indicating the class label of the i th sample, then a pair (\mathbf{w}, b) exists such that

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \text{ for all } (\mathbf{x}_i, y_i)$$

Then the decision hyperplane could be given by $\mathbf{w}^T \mathbf{x}_i + \mathbf{b}$, where \mathbf{w} is the weight vector and b is the bias. Then an optimal hyperplane maximizing the distance between different classes can be found by solving a convex quadratic programming problem [5]:

$$\begin{aligned} \underset{\mathbf{w}, b}{\text{Minimize}} \quad & \Phi(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{subject to} \quad & y_i(\mathbf{w}^T \mathbf{x}_i + \mathbf{b}) \geq 1, i = 1, \dots, l \end{aligned}$$

Among the machine learning and data mining methods, SVM has been an active technique and has been applied in a broad range of areas. At the same time, the classification problems in the literature of fMRI have also been gaining more and more attention. The implication of these problems is that brain state can be predicted using fMRI data, which can enhance the understanding of the cognitive process and brain system [17]. SVMs have been extensively applied to solve the classification problems in fMRI, because they have some unique properties appropriate for the context of fMRI. Among these, one fact is that SVM is capable of dealing with small sample sizes and high dimensional features, which matches the situation of fMRI data [17]. Previous studies have demonstrated the feasibility and potential for the application of SVM in Fmri [17, 18], providing a reliable framework for brain state classification.

1.5 Motivation and Organization

fMRI is a non-invasive technique, and it allows researchers to obtain indirect estimates of neural activity at a spatial resolution of millimeters within a matter of seconds. Consequently, mining fMRI data provides a powerful tool for understanding human

cognitive processes. Among all the applications of fMRI in neuroscience, there are two areas gaining particular increasing interest: 1) Mapping the functional network of brain and investigating the directional interactions of different regions. 2) Decoding and predicting human behaviors from brain activities.

The goal of this thesis is to apply supervised learning models to infer brain function and predict behavior from fMRI data. There are two categories of supervised learning models: regression models and classification models. Both of them were used for the analysis of fMRI data in different chapters. In chapter 2, dynamic Granger causality [19], a regression based supervised learning model, is experimentally demonstrated to be capable of inferring stimulus-evoked sub-100 ms timing difference from fMRI, providing a reliable data-driven method for effective connectivity analysis of fMRI data. Chapter 3 is a continuance of chapter 2, where Patel's τ [20] – a higher order statistics method which performed best for inferring directional interactions from fMRI in a previous simulation study [3] – was verified using experimental fMRI data, highlighting the necessity of experimental validation of simulation results. In chapter 4, recursive cluster elimination based support vector machine [21], a classification based supervised learning model, was used to predict purchase decisions using spatio-temporal fMRI features. This provides a reliable framework for using fMRI data to predict purchase-related decision-making as well as infer its neural correlates. Chapter 5 presents a conclusion of the whole work in the thesis.

Chapter 2: Experimental Validation of Dynamic Granger Causality for Inferring Stimulus-evoked Sub-100ms Timing Differences from fMRI

Abstract

Decoding the sequential flow of events in the human brain non-invasively is critical for gaining a mechanistic understanding of brain function. In this study, we propose a method based on dynamic Granger causality analysis to measure timing differences in brain responses from fMRI. We experimentally validate this method by detecting sub-100ms timing differences in fMRI responses obtained from bilateral visual cortex using fast sampling, ultra-high field and an event-related visual hemifield paradigm with known timing difference between the hemifields. Classical Granger causality was previously shown to be able to detect sub-100 ms timing differences in the visual cortex. Since classical Granger causality does not differentiate between spontaneous and stimulus-evoked responses, dynamic Granger causality has been proposed as an alternative, thereby necessitating its experimental validation. In addition to detecting timing differences as low as 28 ms during dynamic Granger causality, the significance of the inference from our method increased with increasing delay. Therefore, it provides a methodology for understanding mental chronometry from fMRI in a data-driven way.

2.1 Introduction

Correct measurements of small temporal differences in brain activities play a critical role in fully understanding the neural connectivities underlying brain processes. Functional MRI (fMRI) is an indirect measure of neuronal activity based on the blood oxygenation level-dependent (BOLD) hemodynamic response. Typically the hemodynamic response takes 5-8 seconds to reach its peak and 15-30 seconds to return to baseline. On the other hand, neural latencies are typically of the order of tens to hundreds of milliseconds. Therefore, accurate detection of the timing difference of neuronal activities using fMRI is challenging. However, using innovative experimental designs, previous studies have shown that fMRI is sensitive to latency differences of the order of hundreds of milliseconds in the human brain, notwithstanding its poor temporal resolution and hemodynamic smoothing [23,24]. Recently, a study performed by Katwal et al. [16], suggested that recent advances in ultrahigh field image acquisition, fast temporal sampling, and techniques for increasing the available signal-to-noise ratio (SNR) may improve the ability to detect shorter timing differences. Using these strategies, Katwal et al. attempted to detect small timing differences in BOLD signals by introducing known timing differences between left and right visual cortices. They showed that Granger Causality (GC) works well for detecting small temporal precedence in BOLD responses in the visual cortex [16]. GC is a widely-applied method for mapping effective connectivity over the brain, which is based on a statistical measure of how one time series predicts the future values of another [8,14,25,26]. However, conventional GC is sensitive to both spontaneous and stimulus evoked responses [27]. Therefore, previous

studies have proposed that by estimating time-varying coefficients using a dynamic Granger causality model (dGC), temporal precedence due to stimulus-evoked BOLD responses can be separated from that due to spontaneous activity using both EEG [27,28] and BOLD fMRI data [30-32]. In this study, we validate this method by demonstrating the ability of the dGC model to detect sub-100 milliseconds of timing differences between BOLD fMRI time series from left and right visual cortices. Note that dGC was used to detect temporal precedence and not to infer causal influences in this study.

2.2 Methods

2.2.1 Data acquisition

Gradient-echo EPI data (TR=250 ms, TE=25 ms, flip angle=30°, FOV=128 mm×128 mm and voxel size=1 mm×1 mm×2 mm) were acquired from a 7T Philips Achieva scanner from 5 healthy subjects in two coronal slices (with no slice gap) around the calcarine fissure. An event-related visual hemifield paradigm with known timing difference between the hemifields was used. Each visual stimulus comprised a 2-s flashing of checkerboard followed by a 16-s fixation cross for total trial duration of 18 s. Each run included 17 trials and the total run time was 306 seconds. For each subject, five runs were executed by introducing known delays (including 0, 28, 56, 84, 112 ms) between right and left hemifield stimulus. Fig.2.1 shows the stimulus paradigm. FMRI data which consisted of average time series from two activated visual cortical regions (denoted as X and Y for right and left hemisphere, respectively) were obtained using voxels selected by a novel graph-based visualization of self-organizing maps [16,33]. These fMRI data were used for the current analyses.

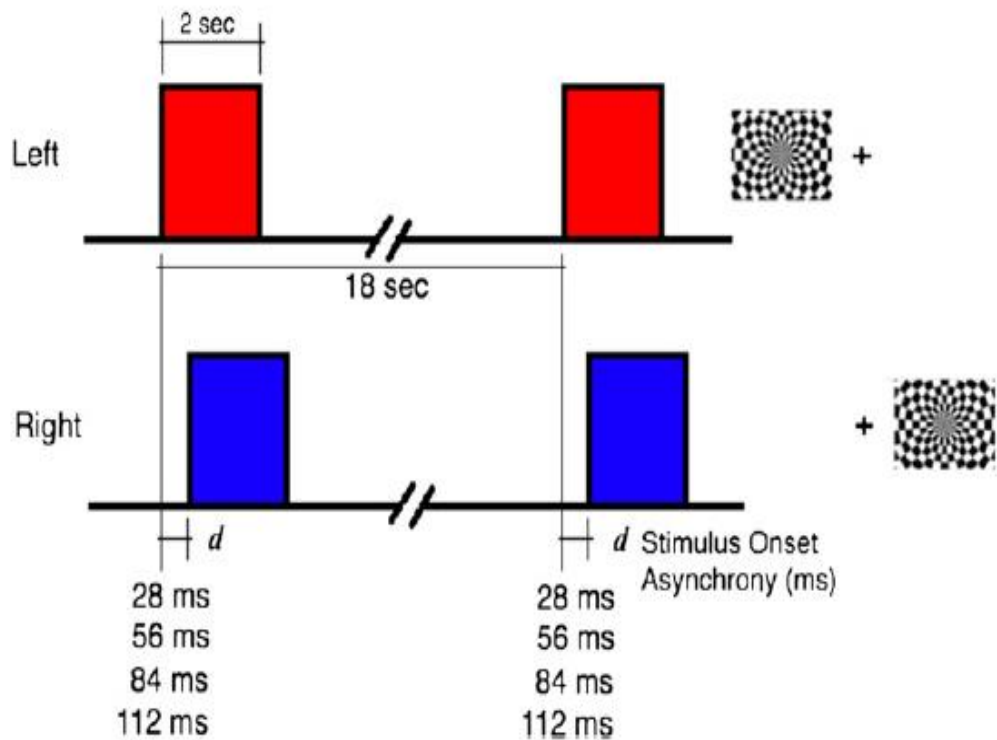


Figure 2.1 The stimulus paradigm. Adapted from Katwal et al 2012 with permission

2.2.2 Dynamic Granger Causality Analysis

As mentioned above, conventional GC is incapable of separating temporal precedence due to spontaneous and stimulus-evoked brain activity. One possible approach for inferring temporal precedence only from stimulus-evoked brain activity is to show that Granger causal estimates covary with the experimental paradigm. Such modulation can confirm temporal precedence due to stimulus-evoked activity and rule out temporal precedence from spontaneous activity. However, conventional GC can only provide one connectivity measure for the entire experiment, because it assumes that the model coefficients are stationary and invariant across time as shown below. Let k fMRI time

series be represented as $X(t) = [x_1(t) \ x_2(t) \ \dots \ x_k(t)]$. The fMRI time series can be input into a multivariate autoregressive (MVAR) [4] as follows:

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ \cdot \\ \cdot \\ x_k(t) \end{bmatrix} = \begin{bmatrix} 0 & a_{12}(0) & \dots & a_{1k}(0) \\ a_{21}(0) & 0 & & a_{2k}(0) \\ \cdot & \cdot & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{k1}(0) & a_{k2}(0) & \dots & 0 \end{bmatrix} \times \begin{bmatrix} x_1(t) \\ x_2(t) \\ \cdot \\ \cdot \\ x_k(t) \end{bmatrix} + \sum_{n=1}^p \begin{bmatrix} a_{11}(n) & a_{12}(n) & \dots & a_{1k}(n) \\ a_{21}(n) & a_{22}(n) & & a_{2k}(n) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{k1}(n) & a_{k2}(n) & \dots & a_{kk}(n) \end{bmatrix} \times \begin{bmatrix} x_1(t-n) \\ x_2(t-n) \\ \cdot \\ \cdot \\ x_k(t-n) \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_2(t) \\ \cdot \\ \cdot \\ e_k(t) \end{bmatrix}$$

Where p is the order of the model determined by the Akaike or Bayesian information criterion [25,34], a are the model coefficients and e is the model error. Note that $a(0)$ represent the instantaneous influences between time series while $a(n)$, $n=1 \dots k$ represent the causal influences between time series. The effect of instantaneous correlation on causality can be minimized by modeling both instantaneous and causal terms in a single model as shown before [35]. The model coefficients were allowed to vary as a function of time in order to make the MVAR model dynamic as given below.

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ \cdot \\ \cdot \\ x_k(t) \end{bmatrix} = \begin{bmatrix} 0 & a_{12}(0,t) & \dots & a_{1k}(0,t) \\ a_{21}(0,t) & 0 & & a_{2k}(0,t) \\ \cdot & \cdot & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{k1}(0,t) & a_{k2}(0,t) & \dots & 0 \end{bmatrix} \times \begin{bmatrix} x_1(t) \\ x_2(t) \\ \cdot \\ \cdot \\ x_k(t) \end{bmatrix} + \sum_{n=1}^p \begin{bmatrix} a_{11}(n,t) & a_{12}(n,t) & \dots & a_{1k}(n,t) \\ a_{21}(n,t) & a_{22}(n,t) & & a_{2k}(n,t) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{k1}(n,t) & a_{k2}(n,t) & \dots & a_{kk}(n,t) \end{bmatrix} \times \begin{bmatrix} x_1(t-n) \\ x_2(t-n) \\ \cdot \\ \cdot \\ x_k(t-n) \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_2(t) \\ \cdot \\ \cdot \\ e_k(t) \end{bmatrix}$$

Considering the model coefficients $a_{ij}(n,t)$ as a state vector of a Kalman filter, they were adaptively estimated using the algorithm proposed by Arnold *et al.* [36]. Dynamic Granger causality (dGC) was then obtained as follows:

$$dGC_{ij}(t) = \sum_{n=1}^p |a_{ij}(n,t)| \quad \forall i = 1 \dots k, j = 1 \dots k$$

A similar metric has been previously used in the static case [37]. Since we had only one time series each from right and left visual cortices for every subject, we used a bivariate model with $k=2$ in this study. We denote the fMRI time series obtained from right and left visual cortices as X and Y , respectively. The dGC model was used to get $X \rightarrow Y$ and $Y \rightarrow X$ connectivity time series for all subjects and delays using a model order of one as determined by the Bayesian Information Criterion [9,14]. The forgetting factor for the Kalman filter was estimated based on minimization of relative error variance [32]. Subsequently, we calculated dynamic Granger causality difference (dGCD) time series, i.e. $X \rightarrow Y - Y \rightarrow X$ to infer the difference in timing between X and Y . If dGCD is larger than zero, it means that the precedence is from X to Y , and vice versa if dGCD is negative.

2.2.3 Covariance of Connectivity with Experimental Paradigm

A time series representing the experimental paradigm was generated by the convolution of the stimulus boxcar function with Statistical Parametric Mapping (SPM)'s canonical HRF. Fig.2.2 shows the stimulus function and the time series representing the experimental paradigm. In order to evaluate how dGCD covaried with the experimental paradigm, a general linear model (GLM) was used, considering the dGCD time series as the response variable and the experimental paradigm as the predictor variable [30,31]. The t -value obtained from the GLM represents the strength of co-variance between dGCD time series and the experimental paradigm. For each delay, we obtained 5 t -values corresponding to the five subjects. Subsequently, a one-side z -test was performed to examine whether the sample represented by the 5 t -values had a mean significantly larger

than zero. The null hypothesis of z -test was that the sample belonged to a normal distribution with zero mean, and standard deviation of σ . For all z -tests, we set σ equal to the standard deviation of the t -values obtained by 0 ms delay.

2.2.4 Comparison with cross-correlation function

The conventionally used metric for estimating delays between time series is the cross-correlation function which computes Pearson's correlation coefficient between two time series at various delays and infer the delay corresponding to the highest correlation coefficient as the time delay between the time series. We compared the efficacy of dGCD with that of the conventionally used cross-correlation function for inferring neural latencies. The minimum latency that can be inferred using the cross-correlation method is equal to the sampling period. The TR of the fMRI time series we used was 250 ms. Since we were interested in inferring sub-100 ms delays, we upsampled the data 25 times such that the resampled data had a sampling period of 10 ms. For each delay and subject, we obtained the cross-correlation function between the upsampled data from bilateral visual cortices. The delay corresponding to the maximum cross-correlation value was found in each case. A one-side z -test was performed to test whether the timing differences obtained from the cross-correlation function were significantly greater than zero, similar to the procedure adopted in the case of dGCD.

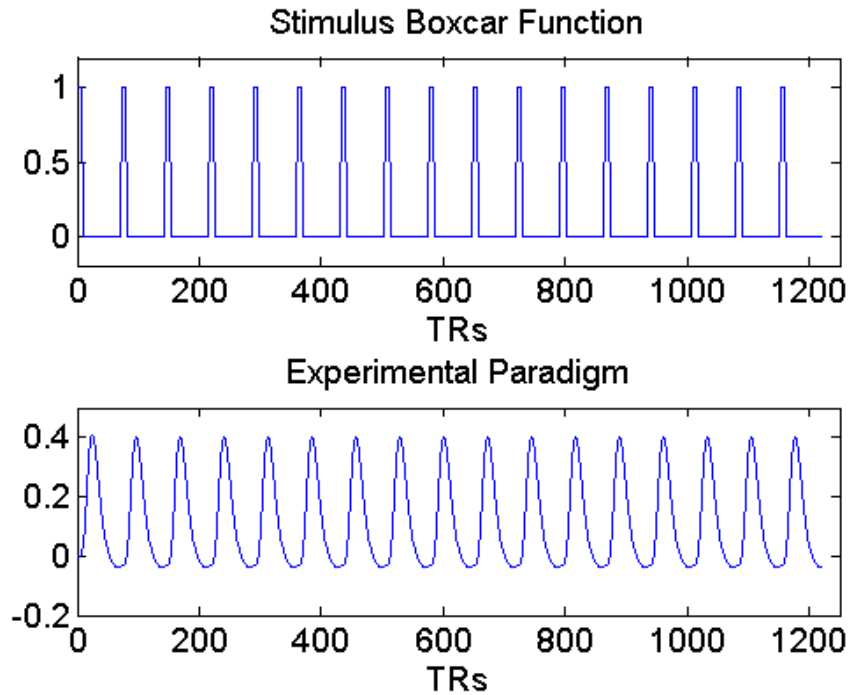


Figure 2.2 Stimulus boxcar function and experimental paradigm

2.3 Results

Fig.2.3 shows the t -values of the GLM fit between the experimental paradigm and dGCD time series. Table 2.1 shows the p -values of a one-sided z -test used to test whether the t -value sample was significantly greater than zero. It is notable that no causality was detected for a delay of zero, while dGCD significantly covaried with the experimental paradigm for all other delays. Also, the significance of causality generally increased with increasing delay. This indicates that dGCD derived from fMRI data was sensitive to even 28 ms latency and that the sensitivity increased with increasing delay time.

Table 2.1 *p*-values of a *z*-test with the null hypothesis that the distribution of *t*-values obtained from all subjects has zero mean

Delay (ms)	<i>p</i> -value
0	0.1120
28	0.0034
56	0.0141
84	5.18×10^{-07}
112	3.81×10^{-12}

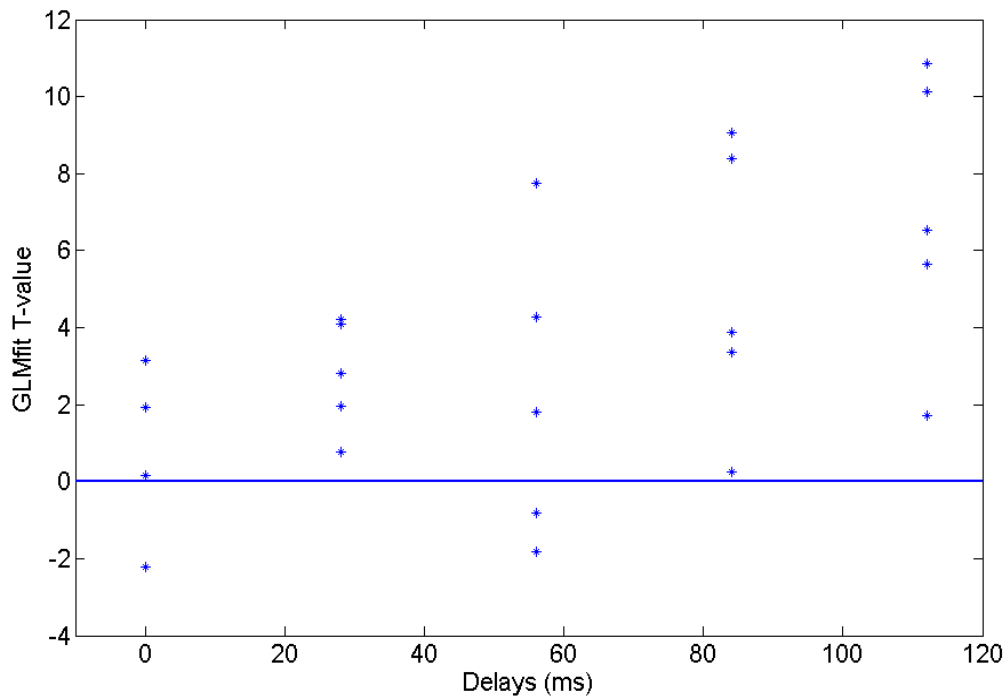


Figure 2.3 *t*-values for the GLM fit between *dGCD* and experimental paradigm versus delay times

Fig.2.4 shows the delays inferred from the cross-correlation function on the y-axis and the true delays on the x-axis. The p -values of the one-sided z-test used to test whether the delays inferred from the cross-correlation function were significantly different from zero are shown in Table 2.2. It is apparent that the cross-correlation function infers a delay when there is no true delay and does not infer a delay when there is one.

Table 2.2 *p-values of a z-test with the null hypothesis that the distribution of delays inferred from the cross-correlation function has zero mean*

Delay (ms)	p-value
0	0.0079
28	0.9939
56	0.6852
84	1
112	0.8326

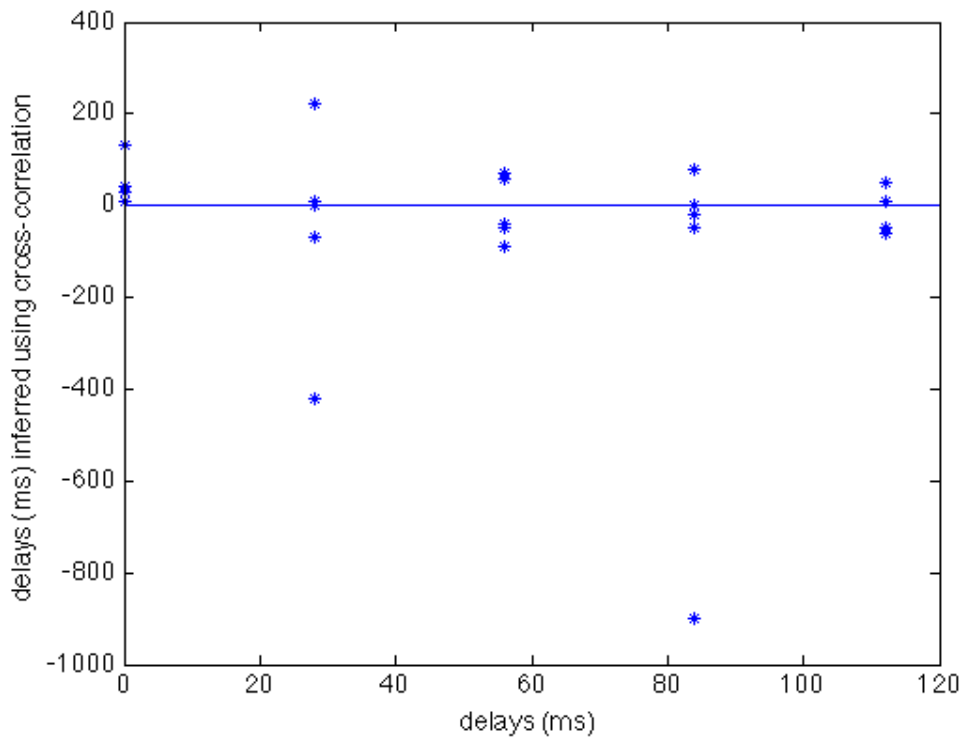


Figure 2.4 Delays inferred from the cross-correlation function on the y-axis and the true delays on the x-axis

2.4 Discussion

There has been intense debate in the past 2 years regarding methods which are suitable to infer directional connectivity information from fMRI [38-42]. Simulations by Smith et al [3] showed that Patel's τ [20] was more suitable than lag-based methods such as Granger causality for detecting directional connectivity. However, studies conducted by different groups have shown that under certain conditions, such as fast sampling and hemodynamic variability being within a range typically observed in healthy individuals, Granger causality can faithfully capture directionality information from fMRI based on

neuronal latencies [12,15,43,44]. The most recent and compelling experimental evidence in favor of Granger causality corresponds to the study by Katwal et al. which showed that using Granger causality for relative timing measurement and self-organizing maps for voxel selection, timing differences as low as 28 ms can be inferred from fMRI time series in bilateral visual cortices, which had experimentally controlled timing differences induced by time-lagged hemi-field stimulation [16]. This makes the data obtained from the Katwal et al.'s study ideal for testing and validating potential approaches for inferring latencies from fMRI.

One outstanding issue with conventional GC is that it is sensitive to both spontaneous and stimulus evoked responses [27]. Previous studies using both EEG [28,29] and BOLD fMRI data [30-32] have proposed that by estimating time-varying coefficients using a dynamic Granger causality model (dGC), temporal precedence due to stimulus-evoked BOLD responses can be separated from that due to spontaneous activity. Therefore, in this study, we have reused the data from the study conducted by Katwal et al [16] to demonstrate and validate the use of dynamic Granger causality to infer tens of milliseconds of stimulus-evoked timing differences from BOLD fMRI. In order to be consistent with the study by Katwal et al., we used dynamic Granger causality difference between bilateral visual cortices as our metric.

We tested three primary hypotheses. First, whether the covariance of dynamic Granger causality difference with the experimental paradigm was non-significant for a delay of 0 ms. This was indeed the case as shown in the results of Table.2.1 wherein the null result would indicate that there was no underlying timing difference. Second, the amount of covariance of dynamic Granger causality difference with the experimental paradigm must

increase with increasing latency. The increasing t -value of the GLM in Fig.2.3 (and decreasing p -value in Table.2.1) supports this hypothesis. Consequently, the t -value can be interpreted in terms of the amount of latency between the time series. Third, we hypothesized that even for a 28 ms delay, we would find significant ($p<0.05$) covariance between dynamic Granger causality difference and the experimental paradigm. This was proven right as shown from the results in Table.2.1. Results obtained from the conventionally used cross-correlation function demonstrated its inability to infer neuronal latencies from fMRI data.

Finally, we provide a few cautionary notes for interpreting the results presented in this report. First, given the confounding effect of the variability of the hemodynamic response [45,46] on Granger causal estimates obtained from BOLD fMRI [15,47], it is noteworthy that hemodynamic variability was probably not a factor influencing the results of both the Katwal et al.'s study [16] as well as the current study since left and right visual cortices are likely to have the same hemodynamics as they are fed by a common hemodynamic source. However, if the proposed dGC technique is applied to other situations where this may not be the case, we recommend that the dGC model be applied on deconvolved fMRI data [48,49]. Second, the performance of the dGC model was aided by the high SNR obtained from the 7T magnet as well as high temporal resolution provided by a TR of 250 ms. More studies are required to ascertain the applicability of these results at lower field strengths and longer TRs. Third, our results should be strictly interpreted within the framework of detecting neuronal delays and not directional connectivity in general. Neuronal delays are an established electrophysiological signature of directional connectivity; however the activity of region A may directionally influence (or predict) the

activity of region B regardless of an explicit delay between the activities obtained from both the regions.

2.5 Conclusion

In this study, dynamic Granger causality analysis was performed to detect sub-100ms timing differences in BOLD responses from the visual cortex. While Katwal et al. [16] demonstrated this possibility using conventional Granger causality, our proposed dynamic Granger causality metric relies on experimental modulation of causality with time. Consequently, the proposed model was able to infer only stimulus-evoked (and not spontaneous) neural timing differences. In summary, our experimental validation of dynamic Granger causality to detect sub-100ms (as small as 28 ms) timing differences provides a reliable data-driven method for effective connectivity analysis of fMRI data.

Chapter 3: Experimental evidence demonstrating the inability of Patel's τ for estimating directionality of brain networks from fMRI

Abstract

Investigating the directional interactions between brain regions plays a critical role in fully understanding brain function. Consequently, multiple methods have been developed for non-invasively inferring directional connectivity from the human brain using functional magnetic resonance imaging (fMRI). Recent simulations by Smith et al showed that Patel's τ , a method based on higher order statistics, was the best approach for inferring directional interactions from fMRI. Since simulations make restrictive assumptions about reality, we set out to verify this finding using experimental fMRI data obtained from a three-region network in a rat modal with electrophysiological validation. Our hypothesis was that Patel's τ obtained from fMRI data should correctly estimate the directionality of neuronal influences obtained from intra-cerebral EEG in this network. However, our results indicate that the accuracy of network directionality estimated using Patel's τ was not better than chance. First, our results highlight the necessity of experimental validation of simulation results. Second, it is unclear which brain mechanism is modeled by a directionality inferred from Patel's τ . Third, this study shows that a directional connection ascertained by different methods may mean different things and more experimental studies are needed for investigating the neuronal mechanisms underlying the direction of a connection in the brain ascertained by fMRI using different methods.

3.1 Introduction

Functional magnetic resonance imaging (fMRI) has primarily been used to explore the spatial localization of brain function [89] where in different brain regions are assumed to perform different brain functions. However, many neuronal processes cannot be localized in a single region and hence are presumed to be encoded by a network formed by several brain regions. Therefore, it is increasingly being recognized that investigating the interactions between brain regions plays a critical role in fully understanding brain function. There are many different methods that have been proposed to characterize the interactions between brain regions. These can be broadly classified into two categories: functional connectivity and effective connectivity. Functional connectivity is defined as “temporal correlations between spatially remote neurophysiological events” while effective connectivity is defined as “the influence one neuronal system exerts over another” [4]. The estimation of directionality of influence is often of great interest as it provides a mechanistic characterization of the underlying neuronal processes in terms of information flow. However, it is much harder to estimate the direction of influence than just estimate whether a connection exists or not. There are three general classes of methods to accomplish this. The first one is “lag-based” methods such as Granger causality [8]. The assumption of these methods is that if one time course is similar to a time-shifted version of the other, then the one with temporal precedence may cause the other. The second class utilizes the concept of conditional independence such as Bayes net methods [50]. The last class is based on higher order statistics such as Patel’s τ [20] wherein asymmetries in the probability of activation of brain region A given the activation of another brain region B versus the probability of activation of region B given

the activation of region A are used to infer the directionality of the influences between regions A and B.

While some neuroimaging techniques such as electroencephalography (EEG) and magnetoencephalography (MEG) have attractive temporal resolution that may favor the application of the above approaches, they are limited by their poor spatial resolution. In contrast, functional magnetic resonance imaging (fMRI) can provide excellent spatial resolution of millimeters and thus has become a popular choice for network estimation. However, fMRI is an indirect measure of neural activities suffering from hemodynamic smoothing and poor temporal resolution [23]. Some approaches whose application to EEG has been established, remain debated when applied to fMRI data (e.g. Granger Causality) [12]. Therefore, careful validation is necessary for the application of network estimation methods, especially for the estimation of directionality of influence. Recently, Smith et al. performed extensive analyses of simulated fMRI data to evaluate the validity of various network estimation methods [3], observing that Patel's τ performed best in the estimation of connection directionality compared to other methods such as Granger causality. However, any simulation is limited by the underlying assumptions, and Smith et al. used a generative biophysical model without an explicit delay, which may have favored Patel's τ over lag-based methods because the former is not based on a delay assumption while the latter is [26,40]. Meanwhile Roebroek et al. [14] and Luo et al. [13] reported excellent results with Granger causality while explicitly including delays in their simulated data. It is notable that we can never make a final conclusion from simulations since they often make restrictive assumptions about reality which might not hold true. Therefore experimental validation of simulations is required. Two recent

studies for experimental validation of effective connectivity methods for fMRI are noteworthy. In the first one, Katwal et al. showed that Granger Causality was capable of inferring sub-100 ms timing differences between right and left visual hemifield stimuli [16]. In the second work, David et al. performed simultaneous EEG and fMRI measurements followed by intra-cerebral EEG (iEEG) recordings in rats [15]. Effective connectivity obtained from both from raw and deconvolved fMRI data [8] using Granger Causality and that obtained from Dynamic Causal Modeling (DCM) [51] were compared with directed functional coupling estimated from iEEG recording for validation. The results showed that Granger causality applied to deconvolved fMRI data as well as DCM were able to estimate network directionality which was consistent with that obtained from iEEG.

In this study, we aimed to use experimental data from the study by David et al to verify the validation of simulation results obtained by Smith et al., specifically with reference to the superiority of Patel's τ for obtaining the directionality of brain connectivity. Patel's τ was performed on these data to estimate the directionality of the three-voxel network; the results were compared to the network estimated by iEEG for validation. Our results indicate that Patel's τ cannot correctly estimate the directionality of brain network.

3.2 Methods

3.2.1 Animal model selection

Genetic Absence Epilepsy Rats from Strasbourg (GAERS) [52] were used in this experiment. GAERS results from genetic selection of more than 80 generations. The rats show spontaneous spike and wave discharges (SWDs), lasting 20 seconds on average and

repeating every minute when they are at rest. Previous studies using genetic model of absence epilepsy have shown that SWDs originate from the perioral regions of the first somatosensory cortex [52,53], thus providing a reference for validating directionality estimation results using fMRI.

3.2.2 fMRI/EEG data acquisition and processing

Six male adult GAERS were included in the fMRI/EEG study. Spontaneous SWDs during MR experiments were measured using EEG. Three carbon electrodes were used, locating on the skull near the midline (frontal, parietal and occipital). Two additional carbon electrodes were introduced for measurement of cardiac activity (electrocardiography [ECG]). MR experiments were performed in a horizontal-bore 2.35 T magnet. fMRI data were acquired using gradient-echo echo-planar imaging (EPI) sequence with the following parameters: two shots, data matrix = 48×48 , FOV = 35×35 mm², 15 contiguous 1.5-mm-thick slices covering the whole brain, alpha = 90°, TE = 20 ms, TR = 3 s. T₁ weighted anatomical scans were also obtained using a 3D-MDEFT sequence [54] with the following parameters: voxel size = $0.33 \times 0.33 \times 0.33$ mm³, TI = 605 ms, quot = 0.45, alpha = 22°, TR/TE = 15/5 ms, and BW = 20 kHz.

SPM 5 (<http://www.fil.ion.ucl.ac.uk/spm/software/spm5/>) was used for data processing and analysis [55]. Standard spatial preprocessing was performed including realigning, normalizing and smoothing. A SWD regressor was then obtained by convolving the down-sampled EEG signal with a canonical HRF. This regressor was then used to obtain SWD-related t-statistic maps and identify ROIs. Several significant activated and deactivated regions were found at the group level. Three of them were identified as ROIs:

Primary somatosensory cortex, barrel field (S1BF), Thalamus and Striatum (caudate-putamen; CPu). Activations were found in S1BF and Thalamus while deactivations were found in CPu. There were several reasons for selecting these 3 regions: (1) they were most consistently activated over different sessions and rats. (2) they exhibit different hemodynamics, which can provide a rigorous validation since HRF variability is a vital concern for many approaches. (3) our current understanding of SWDs can easily integrate them. The time courses from these three regions were used in the following analysis. Please refer to David et al [15] for complete details of data acquisition and activation analysis.

3.2.3 IEEG experiments and data analysis

Five adult GAERS (two males, three females) were used in the iEEG experiment. GAERS were implanted with intra-cerebral electrodes locating in the three ROIs (S1BF, ventrobasal thalamus and striatum). Another two electrodes were fixed in the nasal and occipital bones for reference. EEG data were obtained in awake and freely moving rats. Please refer to David et al [15] for details of the iEEG experiments. IEEG connectivity was obtained by spike averaging and generalized synchronization. Significant directionality were found from S1BF to the striatum ($p < 10^{-9}$) and from S1BF to the thalamus ($p < 0.02$), while the connectivity from striatum to thalamus was not significant ($p > 0.3$). These results were also in accordance with some previous studies [52,53], thus were used as ground truth to validate the results obtained from fMRI connectivity analysis.

3.2.4 Patel’s conditional dependence measures

A data-driven and hypothesis-unconstrained Bayesian approach was proposed by Patel et al. to examine both functional connectivity and effective connectivity [20]. It assesses the connectivity between two voxel or ROI time series by comparing joint and marginal probabilities of evoked activity of voxel/ROI pairs using a Bayesian method. Note that the inferences are made using time series, which could either be extracted from a single voxel or be obtained by averaging time series from multiple voxels within an ROI. In this study, we employed mean time series extracted from the ROIs S1BF, Thalamus and Striatum. Three steps were taken for the measurement of directionality: determining voxel activation, Patel’s kappa for measurement of functional connectivity, and Patel’s tau for measurement of ascendancy (or directionality).

Determining voxel activation. Patel’s method derives inferences based on a binary time series indicating whether a ROI is evoked or not at each time point (or TR). Here, we use the word “evoked” in order to refer to both activations and deactivations. In order to determine the evocation of ROIs at each time point, we first normalized the time series to the range [0,1], limiting the largest 10 percentile data to 1 and lowest 10 percentile data to 0, and linearly mapping other data to 0:1. A pre-chosen threshold T ($0 < T < 1$) was used for binarization. Since S1BF and thalamus were shown to be activated by David et al [14], normalized values larger than T were binarized to 1 and others were binarized to 0. However, David et al showed that the striatum (CPu) was deactivated [14], and hence normalized values smaller than $1 - T$ were binarized to 1 and others were binarized to 0 [8]. Given two ROIs a and b , joint evocation probabilities were calculated from the binary time series extracted from corresponding mean ROI time series. $\theta_1, \theta_2, \theta_3, \theta_4$ were

denoted as the probability of ROI a evoked and ROI b evoked, the probability of ROI a evoked and ROI b rest, the probability of ROI a rest and ROI b evoked, and the probability of ROI a rest and ROI b rest, respectively.

Patel's κ is a measure of functional connectivity based on the posterior distributions. κ is defined as follows [20]:

$$\kappa_{ab} = \frac{\theta_1 - E}{D(\max(\theta_1) - E) + (1 - D)(E - \min(\theta_1))}$$

Where

$$E = (\theta_1 + \theta_2)(\theta_1 + \theta_3)$$

$$\max(\theta_1) = \min(\theta_1 + \theta_2, \theta_1 + \theta_3)$$

$$\min(0, 2\theta_1 + \theta_2 + \theta_3 - 1)$$

$$D = \begin{cases} \frac{\theta_1 - E}{2(\max(\theta_1) - E)}, & \text{if } \theta_1 - E \geq 0 \\ 0.5 - \frac{\theta_1 - E}{2(E - \min(\theta_1))}, & \text{otherwise} \end{cases}$$

Here the numerator of κ is the difference between the joint probability of both ROIs being evoked and the expected joint probability under the case of independence, while the denominator is just a constant restricting κ from -1 to 1. Specifically, if ROI a and ROI b are statistically independent, κ will be 0. When ROI a and ROI b are evoked simultaneously ($\theta_2 = \theta_3 = 0$), κ will be 1. Large value of κ indicates a stronger dependence relationship, or say, functional connectivity between the two ROIs.

Patel's τ is a measure of ascendancy (or directional influence) between any given ROIs a and b . The assumption is that if ROI b is ascendant to ROI a , the time period when ROI

a is evoked should be in the subset of the period when ROI b is evoked, if ROIs a and b are functional connected. τ is define as follows [20]:

$$\tau_{ab} = \begin{cases} 1 - \frac{\theta_1 + \theta_3}{\theta_1 + \theta_2}, & \text{if } \theta_2 \geq \theta_3 \\ \frac{\theta_1 + \theta_2}{\theta_1 + \theta_3} - 1, & \text{otherwise.} \end{cases}$$

τ_{ab} ranges from -1 to +1, with a positive value indicating that the influence is from ROIs a to b , and a negative value indicating directional influence from ROI b to a , given $\kappa \neq 0$.

3.3 Results

Functional connectivity between S1BF, Thalamus and Striatum was determined using Patel's κ with a threshold 0.75. The choice of this value for the threshold is guided by the recommendations of Smith et al [3]. Fig.1 shows Patel's κ values obtained from each of the five rats between the three ROIs. It is notable that a negative κ indicates that the ROI under consideration tends to be evoked at the rest period of the other. In our study, since the three ROIs have been previously demonstrated to be involved in SWDs and were identified using an fMRI regressor, negative κ values were regarded as errors in the estimation of κ . Fig.3.1 shows that most of the κ values were positive, indicating that the three ROIs were functionally connected.

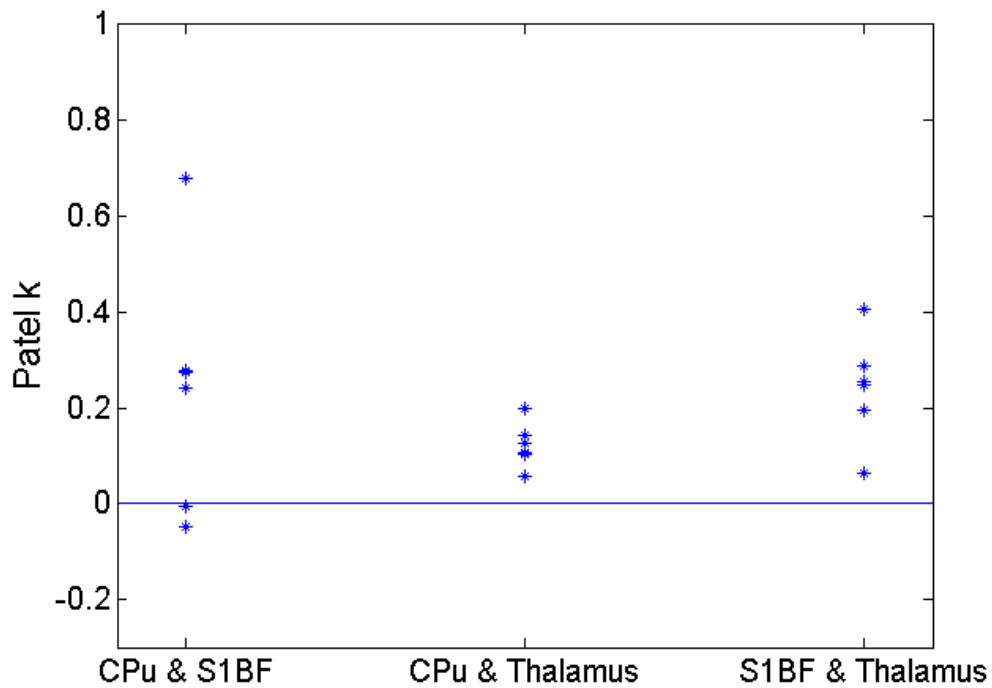


Figure 3.1 Patel's κ for all ROI pairs for each rat showing that the ROIs are functionally connected

Subsequently, the directionality of connectivity in the 3-ROI network was estimated by computing Patel's τ with threshold being 0.75. The results were compared with the ground truth network obtained from iEEG data for validation (please refer to David et al [15] for this result). Fig.3.2 shows the network estimated using Patel's τ for each rat.

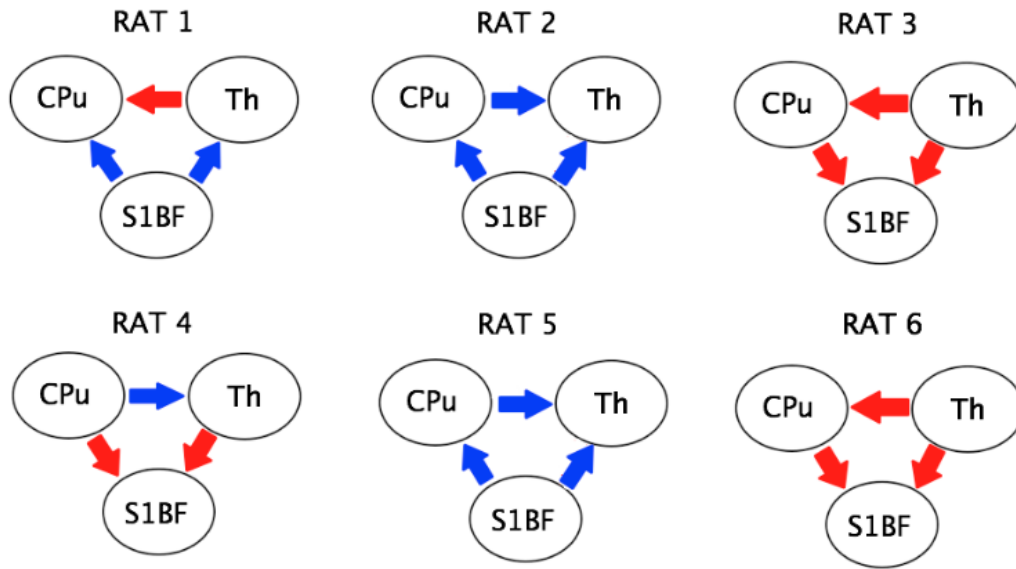


Figure 3.2 Estimated directionality of the network using Patel's τ . Blue arrows indicate estimated directions which agree with that obtained from iEEG while red arrows indicate estimated directions which disagree with that obtained from iEEG [14]

At the group level, the correct estimation rate for each ROI pair was obtained over the six rats. In order to get the statistical significance of the correct estimation rate over chance level, a binomial null distribution $B(\eta, \rho)$ was formed, where η is the number of rats (i.e. 6), ρ is the success probability at chance level (i.e. 0.5). The correct estimation rates based on Patel's τ were compared with the null distribution to get p-values, as shown in Fig.3.3. All the p-values were greater than 0.05, indicating that Patel's τ cannot correctly estimate the directionality of network.

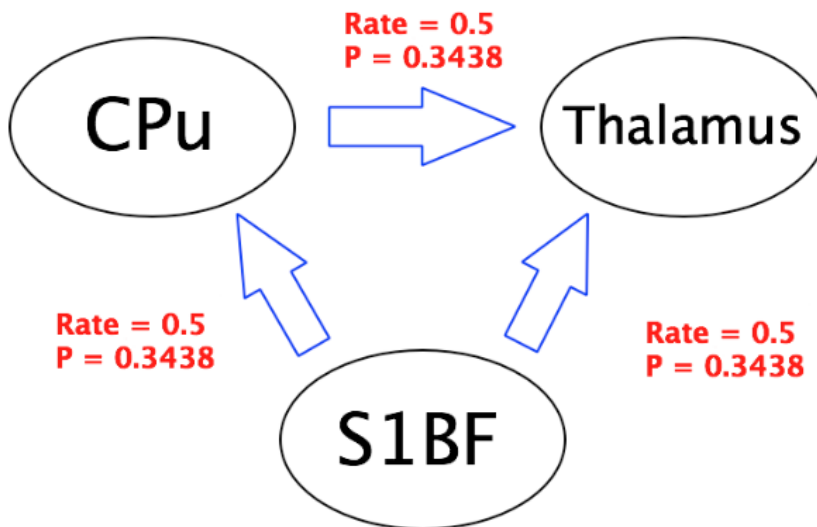


Figure 3.3 Correct group level estimation rates and corresponding p-values for each ROI pair.

In order to consider the effect of different thresholds, Patel's τ with two other thresholds (0.5 and 0.25) as well as the case of no binarization was performed to get estimates of the 3-ROI network. Fig.3.4 shows the correct estimation rates and corresponding p-values in these three cases.

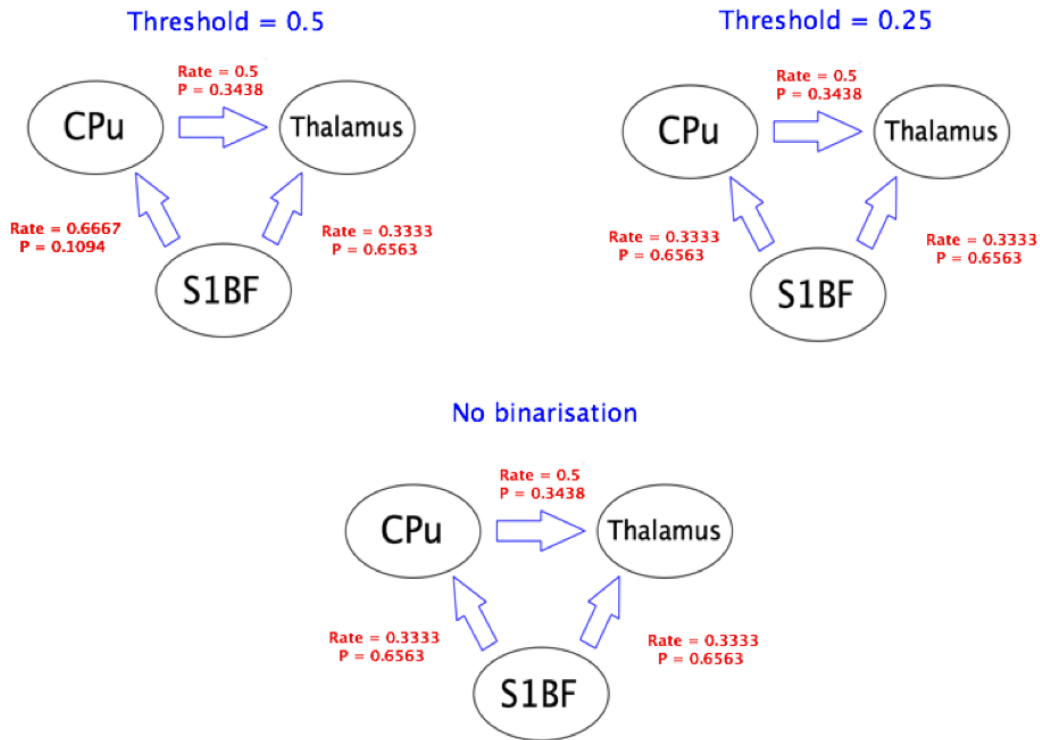


Figure 3.4 Correct group level estimation rates and corresponding p -values for each ROI pairs for different thresholds.

3.4 Discussion

In this study, we experimentally demonstrate that Patel's τ is unable to correctly estimate the directionality of neuronal influences from fMRI data. This finding does not support previous simulations that suggested Patel's τ as an effective measure of directional brain connectivity [3]. Our experimental conditions were fairly similar to those used in the simulations by Smith et al [3]: both used a TR of 3 s, we had data lengths up to 3 times more than those used by Smith et al (note that different rats had different data lengths [15]) which should favor all methods including Patel's τ , and HRF variability was simulated by Smith et al [3] and was reported in the dataset used in the current work by David et al [15]. Further, Smith et al showed that the accuracy of the

estimation of network directionality using Patel's τ showed a strong positive dependence on the length of the time series and strength of connections. As mentioned earlier, the fact that we used a longer time series than that used by Smith et al should favor Patel's τ . However, Rat-2 had the shortest time series of all rats and yet we were able to correctly estimate the directionality of all the three paths in this rat. Also, David et al showed based on directional connectivity estimated from iEEG data that S1BF had very strong directional influences on Thalamus and the Striatum. This should also favor Patel's τ . Taken together, our results demonstrate that even though we had reasonable similarity of experimental conditions as compared to the simulations by Smith et al [3], with certain parameters which were favorable to Patel's τ , we were not able to correctly estimate the directionality of connections in the rat brain. This opens up the possibility that certain assumptions made Smith et al [3] may not hold true in reality. However, we are unable to speculate on specific factors which led to this negative result, and this is an aspect that may be probed in future studies. It is noteworthy that simulations are required to model experimental conditions and not vice versa. Hence, our results could be viewed independently from the simulations of Smith et al as well. Generally speaking, our results demonstrate the need for experimental validation of simulations, since the latter often make restrictive assumptions about reality which might not hold true.

Another issue that should be highlighted is the lack of clarity in the neuroimaging community regarding the neuronal mechanisms underlying the direction of a connection in the brain ascertained by fMRI using various methods. On the one hand, lag-based methods have a clear neuroscientific connotation as it is linked to the concept of mental chronometry. Also, electrophysiological experiments such as the ones employing

syntactic event-related potentials have observed latencies in the primary visual cortex 100 ms post stimulus and in the parietal cortex 500-600 ms post stimulus [57]. These “causal” chains of events in the brain make the interpretation of directionality in lag-based methods fairly straightforward. On the other hand, other methods which assign directionality to connections, specifically Patel’s τ derived from higher order statistics, may not capture “directional influence” in a temporal sense, the way it is intuitively construed by many people. Rather, they rely on other concepts, such as the asymmetries in the probability of activation of a region A given the activation of another region B, versus the probability of activation of a region B given the activation of region A, as in Patel’s τ . Our results indicate that such concepts may not be capable of correctly estimating the directionality of neuronal influence. Therefore, more research is needed to ascertain the neuronal mechanistic underpinnings of methods claiming to ascertain directionality of neuronal influence from fMRI data.

Chapter 4: Predicting Purchase Decisions based on Spatio-temporal Functional MRI Features using Machine Learning

Abstract

Machine learning algorithms allow us to directly predict brain states based on functional magnetic resonance imaging (fMRI) data. In this study, we demonstrate the application of this framework to neuromarketing by predicting purchase decisions from spatio-temporal fMRI data. A sample of 24 subjects were shown product images and asked to make decisions of whether to buy them or not while undergoing fMRI scanning. Eight brain regions which were significantly activated during decision-making were identified using a general linear model. Time series were extracted from these regions and input into a recursive cluster elimination based support vector machine (RCE-SVM) for predicting purchase decisions. This method iteratively eliminates features which are unimportant until only the most discriminative features giving maximum accuracy are obtained. We were able to predict purchase decisions with 71% accuracy, which is higher than previously reported. In addition, we found that the most discriminative features were in signals from medial and superior frontal cortices, both before and after the decision point. Therefore, this approach provides a reliable framework for using fMRI data to predict purchase-related decision-making as well as infer its neural correlates.

4.1 Introduction

The development of functional MRI (fMRI) has greatly promoted our understanding of human brain function. One fundamental and valued problem in neuroimaging is its potential to predict human behaviors from their brain activations. fMRI is a non-invasive technique, and it allows researchers to obtain indirect estimates of neural activity at a spatial resolution of millimeters within a matter of seconds [58]. Consequently, mining fMRI data provides a powerful tool for understanding human cognitive processes.

Recently, multivariate pattern recognition (MPR) methods have been extensively applied to analyze fMRI data for decoding behaviors and cognitive processes [18,59,60,61]. In these approaches, fMRI data are used to detect the differences in activation patterns of cognitive state (state 1 vs. state 2) and discriminate one from the other. Many earlier studies have shown that MPR methods can successfully predict behaviors from brain activations. For example, Haxby et al. distinguished the category of perceived visual stimuli [62], Kamitani et al. decoded the direction of movement [63], Mitchell et al. predicted whether the subjects were looking at a picture or a sentence [64], etc.

Methodologically, the framework of MPR methods applied in neuroimaging usually consists of three parts: feature extraction, feature selection and a particular pattern recognition algorithm [65]. Feature extraction is to obtain some specific characteristics from fMRI data, with the hope that they may have the power to discriminate different classes. The most commonly used features are voxel intensities from specific brain regions of interest (ROIs) [60]. While traditional fMRI studies focused on the spatial features to identify the relevant brain regions, Mourão-Miranda et al. proposed a spatio-

temporal classifier by considering both spatial and temporal features [66]. This approach is valuable for finding out not only ‘where’ but also ‘when’ the brain activation predicts behavior. After feature extraction, feature selection is used to select subsets of features that possess the most discriminatory power. Feature selection is essential for fMRI studies, not only because it can improve the performance of classifiers in terms of prediction accuracy, but also because it can identify the relevant regions and time points that are most useful for classifying different cognitive states. There are two main categories of existing feature selection approaches: “filter methods” such as t-test [64] and “wrapper methods” such as recursive feature elimination (RFE) [67]. Generally, the wrapper methods perform better than filter methods for fMRI studies [59]. In the last part, the selected features are input to a specific pattern recognition algorithm (example: support vector machine) to separate the different classes and correctly predict the class of a novel pattern. Strategies using any combination of each individual part (i.e. feature extraction, feature selection and pattern recognition algorithm) can be used for brain state classification.

While most of the previous studies related to brain state classification focused on the prediction of sensory stimulus perceived by human beings, the possibility of using fMRI data to directly predict human decisions is greatly attractive in many applications. One such example is neuromarketing – the application of neuroimaging techniques to objectively characterize the effect of product marketing on human brains – which has gained increasing popularity [68]. A mechanistic insight into the neurocognitive processes underlying an individual’s decision on whether to buy a product or not is the most fundamental quest in marketing analysis. There are a couple of previous studies

using fMRI data and MPR methods to predict purchase decisions. Knutson et al. extracted fMRI data from the following brain regions – bilateral nucleus accumbens (NAcc), bilateral medial prefrontal cortex (MPFC) and right insula – and used a simple logistic regression model to predict purchase decisions with prediction accuracy at 60% [69]. In order to increase prediction accuracy and yield interpretable coefficients for gaining insight into the neural mechanisms underlying the decision process, Grosenick et al. reused the data from Knutson et al. and applied six different classification models to predict purchase behavior [70]. They showed that PDA-ENET (penalized discriminant analysis – elastic net) classifier [71, 72] performs best with across-subjects classification rate at 66% and within-subjects classification rate at 67.05% and 63.15% for the two presentation datasets respectively [70]. Although Grosenick et al. significantly increased the prediction accuracy as well as temporal and spatial interpretability compared to Knutson et al, we highlight the following outstanding issues. First, three periods were included in the experiment designed by Knutson et al.: product period, price period and choice period [69]. Although prices are an essential element for marketing analysis, the induction of price period may make it difficult to disentangle the effects of product design features and price on the purchase decision. Therefore, in this study, we asked participants to make purchasing decisions solely based on product design features without price considerations, i.e. prices were not displayed. Second, the authors employed three different regression models (i.e. Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net (ENET) and Univariate Soft Thresholding (UST)) for Penalized Discriminant Analysis (PDA) classifiers. All of the three models have the mechanism of automatic variable selection, i.e. removing less discriminative features

from the model. However, the linear Support Vector Machine (SVM) used by the authors does not actually have such ‘feature selection’ part embedded into it [70]. Thus, the fact that PDA classifiers gave a better prediction accuracy and spatial-temporal interpretability than linear SVM are less convincing. Therefore we employed a method based on RFE for selecting features, which were then input into the SVM classifier so that the efficacy of SVMs for brain state classification in the context of predicting purchase decisions using fMRI data can be established.

Specifically, we used spatio-temporal fMRI features with Recursive Cluster Elimination based Support Vector Machine (RCE-SVM) [21] for predicting purchase decisions. The signal features were extracted from time series obtained from 8 different ROIs activated during the task. The reason for adopting RCE as feature selection method is that the wrapper methods have been shown to be advantageous over filter methods for feature selection [59], and RCE considers feature clusters rather than individual features (assuming that features are usually correlated with each other) which makes it faster than the RFE method [21]. Also, RCE-SVM has been reported to be a very reliable classification method in some earlier fMRI studies [65, 73]. Using this method, we were able to achieve an average classification accuracy of 71%, which is better than that obtained by previous purchase decision prediction studies. Also we ranked the features based on their discriminability and in order to infer where and when brain activation can best predict purchase decisions.

4.2 Method

4.2.1 Experiment design and data acquisition

Twenty-four healthy subjects (17 female and 7 male; mean age = 23.6; age range – 19 – 59 years) who were recruited from Auburn University participated in this study. The study protocol was approved by the Institutional Review Board at the university and informed consent was obtained from each subject prior to their participation.

While being scanned, the subjects participated in an event-related task. There were 64 actual product images, with equal number of complex and simple product designs (32 each). There were also an equal number of products reflecting both hedonic and utilitarian product categories (32 each). Please refer to our other publication for details regarding how the product images were chosen based on behavioral testing [74]. For each trial, the subjects were shown one of the 64 product images for 5 seconds, and allowed another 5 seconds to make a purchase decision (buy or not buy?). The 64 stimuli were shown in pseudo-random order, using the E-prime software (<http://www.pstnet.com/software.cfm?ID=101>). Inter-trial intervals were also randomly chosen using optseq software (<http://surfer.nmr.mgh.harvard.edu/optseq/>). The schematic of this event-related design is shown in fig.4.1.

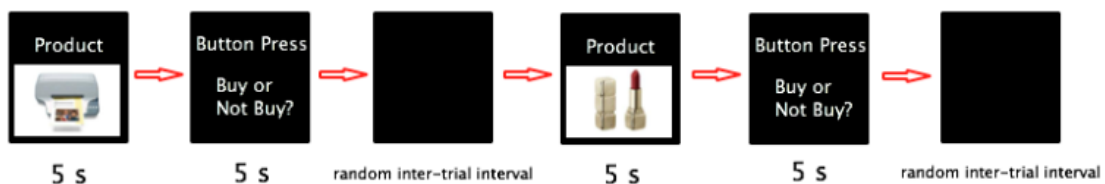


Figure 4.1 The schematic of the event-related experimental design

Functional MRI images were acquired with a 3 Tesla Siemens Verio scanner. 64 visual stimuli were presented to the subjects using an MR-compatible projection system while they lay in the scanner. An MR-compatible button box was used to record the subjects' response to each stimulus. Whenever the subjects had a choice they would press different buttons for buy/not buy decisions. fMRI data were obtained using Echo-planar imaging (EPI) sequence [75] with a 32-channel head coil and the following parameters: TR = 1000 ms, TE = 30 ms, FOV = 24 cm, matrix = 64×64 , 3×3 mm² in-plane resolution and contiguous slices of 5 mm thickness with whole brain coverage. High-resolution anatomical scans were also obtained for an anatomical reference using the 3D magnetization-prepared rapid gradient echo (MPRAGE) [76] sequence (TE/TR = 5/35 ms, matrix = $256 \times 208 \times 196$, FOV = $256 \times 208 \times 192$ mm², and a 1 mm isotropic resolution). fMRI data were subjected to standard pre-processing using statistical parametric mapping (SPM) software (www.fil.ion.ucl.ac.uk/spm/).

4.2.2 ROI selection and feature extraction

Using a general linear model [77], brain regions activated more when a product was bought compared to when it was not and vice versa were identified. We employed a stringent threshold of $p < 0.01$ FWE corrected for multiple comparisons so that only the most discriminative activations were used for further analysis.

One fundamental assumption in MPR methods is that all the training trials in the same class (e.g. state 1) will have the same properties and thus can be exchanged with each other. If only spatial features are extracted for prediction, the stationarity and exchangeability assumption cannot hold [66]. Simple temporal embedding can solve this

problem (i.e. MPR methods would then use both spatial and temporal features for prediction). The feature selection part can produce a discriminating score for both voxel and time point, providing insight into dynamic changes in discriminating power of voxels/ROIs.

Eight activated ROIs were selected; their names and coordinates are shown in Table 4.1. Ten time points (5 from the viewing window and 5 from the decision window) extracted for each subject and each trial was aligned with respect to the exact time point when the subjects made the purchase decision (indicated via pressing the button). Most of the subjects made the decision between time point 6 and 8. So the length of aligned time series was 8 time points, with the decision point being 6th, as shown in Fig.4.2. All the aligned time series were arranged wherein the input space covered both voxels and time points [66, 70]. Specifically, the data was arranged as a three dimensional ($N \times F \times S$) matrix \mathbf{X} , with N corresponding to the number of trials per subject (64), F corresponding to the input features of the classifier (64), and S corresponding to the number of subjects. For each trial, the extracted features were the 8-timepoint aligned time series in the 8 ROIs, so F (input features) was $8 \times 8 = 64$. In this study, we focused on the classification within individual subjects. For each subject, the 64 input features were used in a classifier to obtain the prediction accuracy.

Table 4.1 ROI names and peak MNI coordinates

Name	Peak MNI coordinate
Inferior Temporal Gyrus (ITG)	-52, -36, -24
Medial Frontal Gyrus (MFG)	1, 3, 46
Angular Gyrus (AG)	-50, -74, 16
Superior Frontal Gyrus (SFG)	-14, 32, 62
Middle Frontal Gyrus (MiFG)	52, 24, 38
Left Middle Temporal Gyrus (L MTG)	-46, -40, -6
Right Middle Temporal Gyrus (R MTG)	56, -44, -12
Mid Orbitofrontal Cortex (MOFC)	32, 58, -14

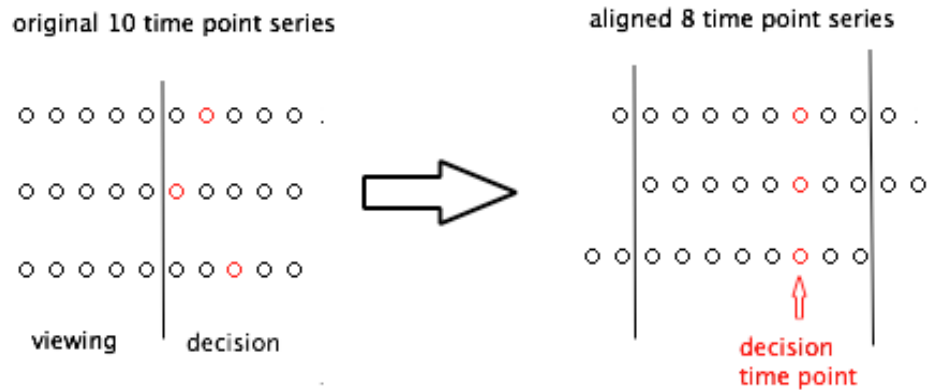


Figure 4.2 An illustration of the process of alignment of ROI time series with respect to the decision time point (red point)

4.2.3 Recursive Cluster Elimination Based Support Vector Machine Classifier

Support Vector Machine (SVM), which was initially proposed by Vapnik [7], is a widely used machine learning method for classification in many different fields of research [78]. Earlier studies have shown that using discriminatory input features will enhance the performance of SVM classifier [67]. Filtering methods and wrapper methods are two commonly used approaches for feature selection [79,80]. In filtering methods, statistical tests such as t-test are performed to select the features that are statistically different between classes [81]. The limitation of this method is that the features are selected independent of the classification process, and the measures are univariate without considering the relationship between features [82]. Wrapper methods can successfully solve these problems by embedding the feature selection into the classification process. In this method, features are iteratively eliminated to minimize prediction error [59, 79, 83]. RCE-SVM is a wrapper methods based SVM. It was firstly proposed for gene classification to enhance both classification accuracy and computational efficiency [21], and then successfully applied in some previous fMRI studies [65, 73]. In this study, we propose a method that takes advantage of both filter and wrapper methods. Selection of spatio-temporal features using GLM analysis represents a “filtering” of the input space for dimensionality reduction using mass univariate models. By using these selected features in an RCE-SVM wrapper model, our approach represents a fusion of both filter and wrapper methods.

There are three main steps in RCE-SVM algorithm: cluster step, SVM scoring step and RCE step, as shown in Fig.4.3 [65, 21]. Firstly, the input features on the 64 trials for each subject were equally divided into two sets, one for training and the other for testing. In

the cluster step, an unsupervised learning method *K-means algorithm* [84] was performed to identify features correlated in the training set which were clustered into n clusters, where n was the initial number of clusters and set to a pre-chosen number (i.e. 35 in this study). In the next step, SVM score of each cluster was defined as its ability to discriminate the two classes. The scores were obtained through a cross-validation by a linear SVM. The training data were randomly and equally partitioned into 5 non-overlapping folds; linear SVM using the features in one particular cluster was trained over 4 folds and performance was calculated from the remaining fold. The procedure was repeated for 50 times in order to take into account different partitions to ensure the reliability of performance. The mean classification accuracy over all the folds and repetitions was assigned as the SVM score of each feature. In the RCE step, the 20% of features with the lowest SVM scores were removed. The remaining features were merged and n was set to $n - 0.2*n$. All the three steps were repeated until n is equal to 2. Testing set was used to evaluate the prediction performance at the end of each iteration. There is no bias in the performance accuracy using this procedure because of total separation of training and testing data [85]. The accuracy at each RCE-SVM loop was obtained from the average accuracy of all 50 repetitions using the feature clusters of testing data at the corresponding loop.

A within-subject prediction performance was calculated as the mean value of the individual subject RCE-SVM classifier prediction accuracies. In order to calculate the statistical significance of prediction accuracy over chance level, a binomial null distribution $B(\eta, \rho)$ [61] was formed, where η is the number of trials (i.e. 64), ρ is the success probability of chance level (i.e. 0.5). The prediction accuracies were compared

with the null distribution to get p-values and only accuracies whose p-values were less than 0.05 were considered as significantly higher than guess level.

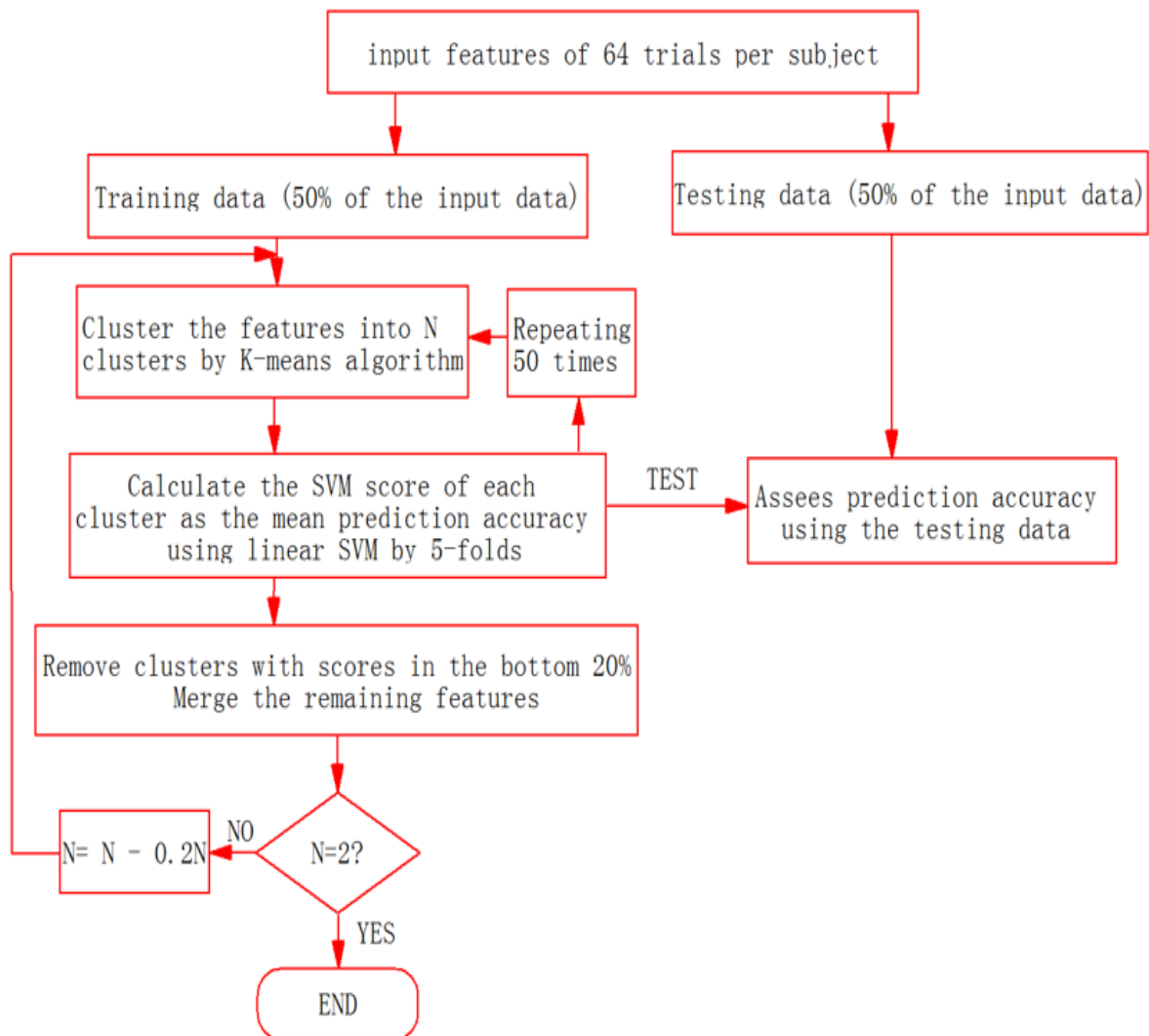


Figure 4.3 Flowchart of RCE-SVM algorithm

4.3 Results

4.3.1 Prediction accuracy

The average prediction accuracies at each RCE-SVM step (i.e. using particular number of feature clusters) are shown in Fig.4.4. The figure illustrates that the average performance of prediction increased with the removal of non-discriminative features and reached a highest accuracy of 70.73% using 2 clusters and 4 features. The p-values corresponding to accuracies at each step are shown in Table 4.2. For each individual subject, the prediction accuracy was defined as the maximum accuracy in the accuracy curve. When prediction accuracy was calculated separately for male and female subjects, no significant differences in accuracy were observed ($p>0.05$). Fig.4.5 shows the histogram and statistics of the individual prediction accuracies for all the subjects. The mean and median accuracy is 70.98% and 70.56% respectively. 11 subjects had rates $> 70\%$, with a maximum individual accuracy of 83.35%.

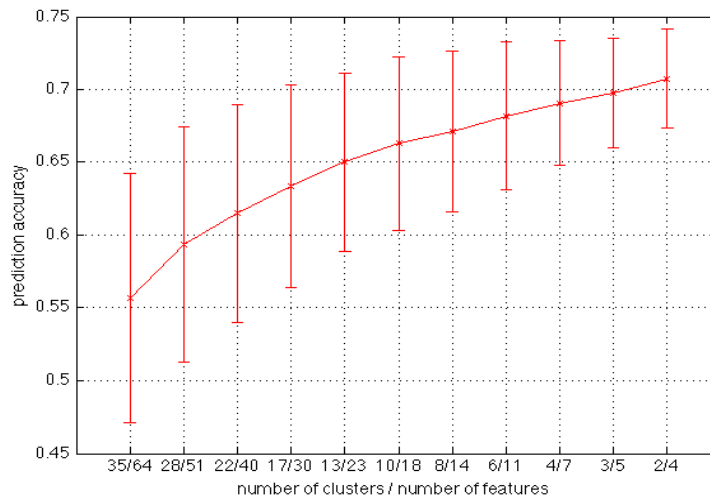


Figure 4.4 The evolving prediction accuracy of RCE-SVM with decreasing number of features

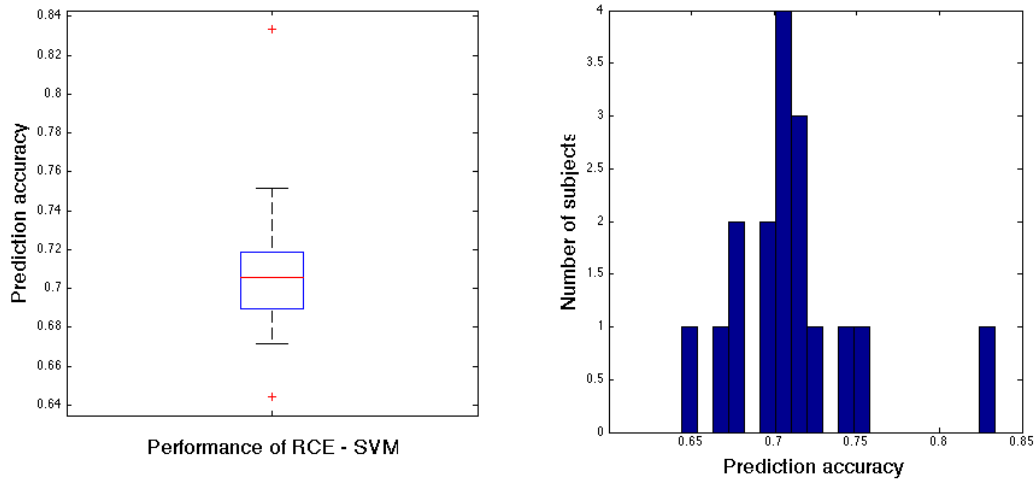


Figure 4.5 Statistics and histogram of RCE-SVM within-subject prediction accuracies

Table 4.2 Prediction accuracies and corresponding p-values at each step of RCE-SVM classifier

Number of features	Prediction accuracy (%)	P-value
64	55.70	0.1919
51	59.39	0.0517
40	61.50	0.0300
31	63.35	0.0164
24	65.01	0.0084
19	66.28	0.0041
15	67.13	0.0041
11	68.19	0.0018
8	69.07	0.0008
6	69.76	0.0008
4	70.73	0.0003

4.3.2 Important spatio-temporal features for classification

The SVM scores indicating discriminatory power of individual features were averaged over all the RCE-SVM classifiers and ranked in descending order, i.e. the feature with

rank-1 was the most discriminative. Fig.4.6 shows the dynamic changes of discriminatory power in the 8 ROIs across the entire trial. The top 4 ranked are indicated in red since they gave the best prediction accuracy.

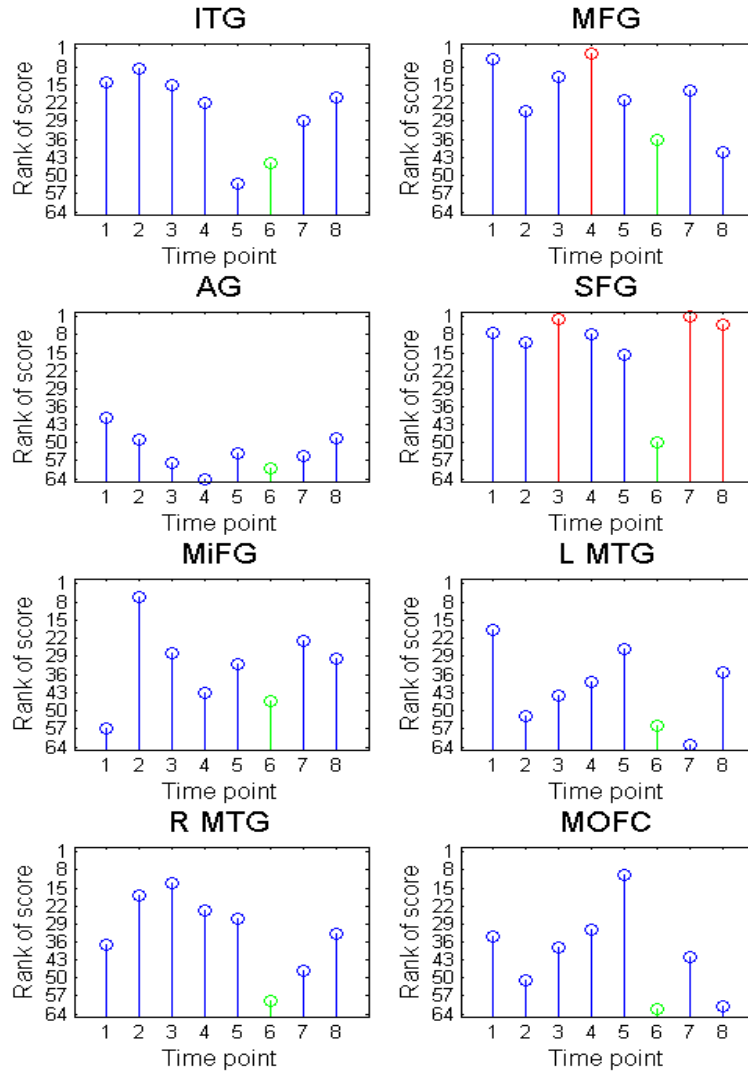


Figure 4.6 Dynamic changes of discriminatory power in 8 ROIs (Red-labeled points are top-4-ranked features; green-labeled points are decision points)

4.4 Discussion

The goals of this study were: (1) to predict purchase decisions by utilizing machine learning methods (2) to find the spatio-temporal features which were most important for prediction, and (3) to interpret them. We employed a recursive cluster elimination based support vector machine for prediction of purchase decisions based on spatial-temporal fMRI features and obtained those features which have the highest predictive power. We obtained >70% which was significantly higher than chance level. Further, the most discriminative features were in medial and superior frontal cortices, both before and after the decision point. We elaborate on these themes below.

Recently, SVMs have been extensively applied in fMRI data analysis [59, 65, 66]. There are two main motivations for utilizing machine learning methods for fMRI analysis: (1) Classifiers can be seen as a pattern recognition method used to predict cognitive behaviors from brain activity. The optimization of prediction accuracy is crucial for this case. (2) The classification procedure can also provide an insight into the neuronal mechanisms underlying the cognitive process [66]. The two motivations also correspond to the two goals in our study.

The average prediction accuracy curve in Fig.4.4 shows that without the feature selection part (i.e. RCE part), the average accuracy was 55.70% with a p-value < 0.05. In contrast, the average individual accuracy obtained from RCE-SVM after eliminating uninformative features was much higher. The results demonstrate that the feature selection part is of great importance for advancing the utility of machine learning algorithms for brain state classification. Therefore, the results obtained by Grosenick *et al*

which claimed that the LDA classifier performs better than linear support vector machine for purchase prediction needs to be viewed in the context that they did not employ feature selection before using the SVM [70].

The dynamic changes of SVM scores in ROIs shown in Fig.4.6 provide the spatio-temporal information about ‘when’ and ‘where’ the human brain activations are most important for purchase decision prediction. It shows that the most important features for purchase prediction included signal amplitudes in SFG and MFG before the decision point, and in SFG after the decision point. Therefore, understanding the role of SFG and MFG in decision-making process is essential for insight into the underlying neural mechanism. SFG has been proved to be less important when a single action is selected, but necessary when the decision rules change dynamically [86,87]. In our experiment, products with different design features (i.e. hedonic or utilitarian products with either simple or complex design) were presented to the participants in a random order. It is obvious that the rules for whether to buy a hedonic product will be different from the rules for whether to buy a utilitarian product. SFG was activated and generated discriminatory power soon after participants viewed the products, but 3 s before they made the purchase decision. Medial frontal gyrus (MFG, also referred to as medial prefrontal cortex) plays an important role in integrating gains and losses [74]. Previous studies have shown that MFG is activated in economic decisions [69, 74]. After viewing the products and before making a decision whether to buy it or not, considering the gains and losses of a decision activated MFG. This probably explains the predictive power of the amplitude of MFG signal 2 s before the decision point. After the decision point, participants will hold a short term-memory for what decision they made. SFG has been

implicated in working memory (WM) [88], indicating a probable reason for its discriminatory power after participants made a purchase decision.

4.5 Conclusion

In this study, we adopted recursive cluster elimination based support vector machine to predict purchase decisions, (i.e. whether an individual decides to buy a product or not), using spatio-temporal fMRI features with more than 70% accuracy. We combined filter methods (i.e. GLM) with wrapping methods (i.e. RCE) for feature selection. This enabled us to identify the signal values in medial and superior frontal gyrus, both before and after the decision point, as spatio-temporal features possessing the most discriminatory power for predicting purchase decisions. Our approach provides a reliable multivariate pattern recognition framework for brain state classification using neuroimaging data, in terms of both improving prediction accuracy and generating interpretable spatio-temporal information.

Chapter 5: Conclusion

In this thesis supervised learning models were applied for estimating effective connectivity and predicting purchase decisions from fMRI data. Our proposed dynamic Granger causality relies on experimental modulation of causality with time, and therefore was able to infer only stimulus-evoked (and not spontaneous) neural timing differences. Subsequently, we experimentally demonstrated that Patel's τ was unable to correctly estimate the directionality of neuronal influence of spontaneous spike and wave discharges (SWDs) in Genetic Absence Epilepsy Rats from Strasbourg (GAERS). These findings do not support previous simulations that suggested Patel's τ as the most effective measure of directional connectivity, and demonstrate the need for experimental validation of simulations since the latter often make restrictive assumptions about reality which might not hold true. Last but not least, we adopted recursive cluster elimination based support vector machine to predict purchase decisions, using spatio-temporal fMRI features with more than 70% accuracy. The combination of filter methods with wrapping methods for feature selection enabled us to identify the signal values in medial and superior frontal gyrus, both before and after the decision point, as spatio-temporal features possessing the most discriminatory power for predicting purchase decisions. In conclusion, this thesis provides some reliable validation and methodology for the application of supervised learning models in the context of fMRI.

Bibliography

- [1] S. C. Bushong, Magnetic resonance imaging, Elsevier Health Sciences, 2003.
- [2] S. A. Huettel, A. W. Song and G. McCarthy, Functional magnetic resonance imaging, Sunderland, MA: Sinauer Associates, 2004.
- [3] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey and M. W. Woolrich, "Network modelling methods for fMRI," *Neuroimage*, vol. 54, no. 2, pp. 875-891, 2011.
- [4] K. J. Friston, "Functional and effective connectivity in neuroimaging: a synthesis," *Human brain mapping*, vol. 2, no. 1-2, pp. 56-78, 1994.
- [5] S. B. Kotsiantis, "Supervised machine learning: a review of classification techniques," *Informatica (03505596)*, vol. 31, no. 3, 2007.
- [6] B. Yegnanarayana, Artificial neural networks, PHI Learning Pvt. Ltd., 2009.
- [7] V. Vapnik, The nature of statistical learning theory, springer, 2000.
- [8] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: Journal of the Econometric Society*, pp. 424-438, 1969.
- [9] G. Schwarz, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, no. 2, pp. 461-464, 1978.
- [10] A. Bollimunta, Y. Chen, C. E. Schroeder and M. Ding, "Neuronal mechanisms of cortical alpha oscillations in awake-behaving macaques," *The Journal of neuroscience*, vol. 28, no. 40, pp. 9976-9988, 2008.
- [11] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura and S. L. Bressler, "Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 26, pp. 9849-9854, 2004.
- [12] X. Wen, G. Rangarajan and M. Ding, "Is Granger causality a viable technique for analyzing fMRI data?," *PloS one*, vol. 8, no. 7, p. e67428, 2013.
- [13] Q. Luo, W. Lu, W. Cheng, P. A. Valdes-Sosa, X. Wen, M. Ding and J. Feng, "Spatio-temporal Granger causality: A new framework," *NeuroImage*, vol. 79, pp. 241-263, 2013.
- [14] A. Roebroeck, E. Formisano and R. Goebel, "Mapping directed influence over the brain using Granger causality and fMRI," *Neuroimage*, vol. 25, no. 1, pp. 230-242, 2005.
- [15] O. David, I. Guillemain, S. Saitet, S. Reyt, C. Deransart, C. Segebarth and A. Depaulis, "Identifying neural drivers with functional MRI: an electrophysiological validation," *PLoS biology*, vol. 6, no. 2, p. e315, 2008.
- [16] S. B. Katwal, J. C. Gore, J. C. Gatenby and B. P. Rogers, "Measuring relative timings of brain activities using fMRI," *NeuroImage*, vol. 66, pp. 436-448, 2013.
- [17] S. LaConte, S. Strother, V. Cherkassky, J. Anderson and X. Hu, "Support vector machines for temporal classification of block design fMRI data," *NeuroImage*, vol. 26, no. 2, pp. 317-329, 2005.
- [18] D. D. Cox and R. L. Savoy, "Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex," *Neuroimage*, vol. 19, no. 2, pp. 261-270, 2003.

- [19] M. Arnold, X. H. R. Milner, H. Witte, R. Bauer and C. Braun, "Adaptive AR modeling of nonstationary time series by means of Kalman filtering," *Biomedical Engineering, IEEE Transactions on*, vol. 45, no. 5, pp. 553-562, 1998.
- [20] R. S. Patel, F. D. Bowman and J. K. Rilling, "A Bayesian approach to determining connectivity of the human brain," *Human brain mapping*, vol. 27, no. 3, pp. 267-276, 2006.
- [21] M. Yousef, S. Jung, L. C. Showe and M. K. Showe, "Recursive Cluster Elimination (RCE) for classification and feature selection from gene expression data," *BMC bioinformatics*, vol. 8, no. 1, p. 144, 2007.
- [22] T. C. Clancy, A. Khawar and T. R. Newman, "Robust signal classification using unsupervised learning," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 4, pp. 1289-1299, 2011.
- [23] R. Menon, & S. G. Kim, "Spatial and temporal limits in cognitive neuroimaging with fMRI," *Trends Cogn. Sci.*, vol. 3, pp. 207–216, Jun. 1999.
- [24] R. Menon, D. Luknowsky, & J. Gati, "Mental chronometry using latency-resolved functional MR," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, pp. 10902-10907, Sept. 1998.
- [25] G. Deshpande, S. LaConte, G. James, S. Peltier, & X. Hu, "Multivariate Granger causality analysis of brain networks," *Human Brain Mapping*, vol. 30, no. 4, pp. 1361-1373, Apr. 2009
- [26] M. Dhamala, G. Rangarajan, & M. Ding, "Estimating Granger causality from fourier and wavelet transforms of time series data," *Physical Review Letters*, vol. 100, no. 1, p. 018701, Jan. 2008
- [27] S. Ryali, K. Supekar, T. Chen, & V. Menon, "Multivariate dynamical systems models for estimating causal interactions in fMRI," *NeuroImage*, vol. 54, no. 2, pp. 807-23, Jan 2011
- [28] W. Hesse, E. Moller, M. Arnold, & B. Schack, "The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies," *Journal of Neuroscience Methods*, vol. 124, no. 1, pp. 27-44, Jan. 2003
- [29] L. Astolfi, F. Cincotti, D. Mattia, F. De Vico Fallani, A. Tocci, A. Colosimo, . . . F. Babiloni, "Tracking the time-varying cortical connectivity patterns by adaptive multivariate estimators," *IEEE Transactions on Biomedical Engineering*, vol. 55, pp. 902-913, Mar. 2008
- [30] S. Lacey, H. Hagtvedt, V. Patrick, A. Anderson, R. Stilla, G. Deshpande, . . . & K. Sathian, "Art for reward's sake: Visual art recruits the ventral striatum," *NeuroImage* vol. 55, pp. 420–433, Mar. 2011
- [31] D. Kapogiannis, G. Deshpande, F. Krueger, M. Thornburg, & J. Grafman, "Brain Networks Shaping Religious Belief," *Brain Connectivity*, 2013 (in press)
- [32] M. Havlicek, J. Jan, M. Brazdil, & V. Calhoun, "Dynamic Granger causality based on Kalman filter for evaluation of functional network connectivity in fMRI data," *NeuroImage*, vol. 53, pp. 65-77, Oct. 2011
- [33] S. Katwal, J. Gore, & B. Rogers, "Unsupervised spatiotemporal analysis of fMRI data using graph-based visualizations of self-organizing maps," *IEEE Transactions in Biomedical Engineering*, vol. 60, pp. 2472-2483, Sept. 2013
- [34] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 9, pp. 716-723, Dec. 1974

- [35] G. Deshpande, K. Sathian, & X. Hu, "Assessing and Compensating for Zero-lag Correlation Effects in Time-lagged Granger Causality Analysis of fMRI," *IEEE Transactions on Biomedical Engineering*, vol. 57, pp. 1446-1456, Jun. 2010
- [36] M. Arnold, W. Miltner, H. Witte, R. Bauer, & C. Braun, "Adaptive AR Modeling of Nonstationary Time Series by Means of Kalman Filtering," *IEEE Transactions on Biomedical Engineering*, vol. 45, pp. 553-562, May 1998
- [37] M. Kaminski, M. Ding, W. Truccolo, & S. Bressler, "Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance," *Biological Cybernetics*, vol. 85, pp. 145-157, Aug. 2001
- [38] K. Friston, "Causal modelling and brain connectivity in functional magnetic resonance imaging," *PLoS Biology*, vol. 7, p. e33, Feb. 2009
- [39] G. Deshpande, K. Sathian, X. Hu, & K. Buckhalt, "A rigorous approach for testing the constructionist hypotheses of brain function," *Behavioral and Brain Sciences*, vol. 35, pp. 148-149, Jun. 2012
- [40] G. Deshpande, & X. Hu, "Investigating effective brain connectivity from FMRI data: past findings and current issues with reference to granger causality analysis," *Brain Connectivity*, vol. 2, pp. 235-245, Oct. 2012
- [41] A. Roebroeck, E. Formisano, & R. Goebel, "The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution," *NeuroImage*, vol. 58, pp. 296-302, Sept. 2011
- [42] S. Bressler, & A. Seth, "Wiener-Granger Causality: A well established methodology," *NeuroImage*, vol. 58, pp. 323-329, Sept. 2011
- [43] A. Seth, P. Chorley, & L. Barnett, "Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling," *NeuroImage*, vol. 65, pp. 540-555, Jan. 2013
- [44] M. Schippers, R. Renken, & C. Keysers, "The effect of intra- and inter-subject variability of hemodynamic responses on group level Granger causality analyses," *NeuroImage*, vol. 57, pp. 22-36, Jul. 2011
- [45] G. Aguirre, E. Zarahn, & M. D'Esposito, "The variability of human, BOLD hemodynamic responses," *NeuroImage* vol. 8, pp. 360-369, Nov. 1998
- [46] D. Handwerker, J. Ollinger, & M. D'Esposito, "Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses," *NeuroImage* vol. 21, pp. 1639-1651, Apr. 2004
- [47] G. Deshpande, K. Sathian, & X. Hu, "Effect of hemodynamic variability on Granger causality analysis of fMRI," *NeuroImage* vol. 52, pp. 884-896, Sept. 2010
- [48] M. Havlicek, K. Friston, J. Jan, M. Brazdil, & V. Calhoun, "Dynamic modeling of neuronal responses in fMRI using cubature Kalman filtering," *NeuroImage*, vol. 56, pp. 2109-2128, Jun. 2011
- [49] G. Wu, W. Liao, S. Stramaglia, J. Ding, H. Chen, & D. Marinazzo, "A blind deconvolution approach to recover effective connectivity brain networks from resting state fMRI data," *Med Image Anal.*, vol. 17, pp. 365-374, Apr. 2013
- [50] J. D. Ramsey, S. J. Hanson, C. Hanson, Y. O. Halchenko, R. A. Poldrack and C. Glymour, "Six problems for causal inference from fMRI," *Neuroimage*, vol. 49, no. 2, pp. 1545-1558, 2010.

- [51] K. J. Friston, L. Harrison and W. Penny, "Dynamic causal modelling," *Neuroimage*, vol. 19, no. 4, pp. 1273-1302, 2003.
- [52] L. Danover, C. Deransart, A. Depaulis, M. Vergnes and C. Marescaux, "Pathophysiological mechanisms of genetic absence epilepsy in the rat," *Progress in neurobiology*, vol. 55, no. 1, pp. 27-57, 1998.
- [53] H. Meeren, G. van Luijtelaar, F. L. da Silva and A. Coenen, "Evolving concepts on the pathophysiology of absence seizures: the cortical focus theory," *Archives of neurology*, vol. 62, no. 3, pp. 371-376, 2005.
- [54] P. O. Polack, I. Guillemain, E. Hu, C. Deransart, A. Depaulis and S. Charpier, "Deep layer somatosensory cortical neurons initiate spike-and-wave discharges in a genetic model of absence seizures," *The Journal of neuroscience*, vol. 27, no. 24, pp. 6590-6599, 2007.
- [55] R. Deichmann, C. Schwarzbauer and R. Turner, "Optimisation of the 3D MDEFT sequence for anatomical brain imaging: technical implications at 1.5 and 3 T," *Neuroimage*, vol. 21, no. 2, pp. 757-767, 2004.
- [56] P. Schweinhardt, P. Fransson, L. Olson, C. Spenger and J. L. Andersson, "A template for spatial normalisation of MR images of the rat brain," *Journal of neuroscience methods*, vol. 129, no. 2, pp. 105-113, 2003.
- [57] A. C. Gouvea, C. Phillips, N. Kazanina and D. Poeppel, "The linguistic processes underlying the P600," *Language and Cognitive Processes*, vol. 25, no. 2, pp. 149-188, 2010.
- [58] K. A. Norman, S. M. Polyn, G. J. Detre and J. V. Haxby, "Beyond mind-reading: multi-voxel pattern analysis of fMRI data," *Trends in cognitive sciences*, vol. 10, no. 9, pp. 424-430., 2006.
- [59] F. De Martino, G. Valente, N. Staeren, J. Ashburner, R. Goebel and E. Formisano, "Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns," *Neuroimage*, vol. 43, p. 44-58, 2008.
- [60] J. D. Haynes, K. Sakai, G. Rees, S. Gilbert, C. Frith and R. E. Passingham, "Reading hidden intentions in the human brain," *Curr. Biol.*, vol. 17, p. 323-328, 2007.
- [61] F. Pereira, T. Mitchell and M. Botvinick, "Machine learning classifiers and fMRI: a tutorial overview," *Neuroimage*, vol. 45, p. S199-209, 2009.
- [62] J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Scgityeb and P. Pietrini, "Distributed and overlapping representations of faces and objects in ventral temporal cortex," *Science*, vol. 293, p. 2425-2429, 2001.
- [63] Y. Kamitani and F. Tong, "Decoding seen and attended motion directions from activity in the human visual cortex," *Current Biology*, vol. 16, no. 11, pp. 1096-1102, 2006.
- [64] T. M. Mitchell, R. Hutchinson, R. S. Niculescu, F. Pereira, X. Wang, M. Just and S. Newman, "Learning to decode cognitive states from brain images," *Machine Learning*, vol. 57, no. 1-2, pp. 145-175, 2004.
- [65] G. Deshpande, Z. Li, P. Santhanam, C. D. Coles, M. E. Lynch, S. Hamann and X. Hu, "Recursive cluster elimination based support vector machine for disease state prediction using resting state functional and effective brain connectivity," *PloS one*, vol. 5, no. 12, p. e14277, 2010.
- [66] J. Mourao-Miranda, K. J. Friston and M. Brammer, "Dynamic discrimination analysis: a spatial-temporal SVM," *Neuroimage*, vol. 36, no. 1, pp. 88-99, 2007.

- [67] R. C. Craddock, P. E. Holtzheimer, X. Hu and H. S. Mayberg, "Disease state prediction from resting state functional connectivity," *Magnetic resonance in Medicine*, vol. 62, no. 6, pp. 1619-1628, 2009.
- [68] D. Ariely and G. S. Berns, "Neuromarketing: the hope and hype of neuroimaging in business," *Nature Reviews Neuroscience*, vol. 11, no. 4, pp. 284-292, 2010.
- [69] B. Knutson, S. Rick, G. E. Wimmer, D. Prelec and G. Loewenstein, "Neural predictors of purchases," *Neuron*, vol. 53, no. 1, pp. 147-156, 2007.
- [70] L. Grosenick, S. Greer and B. Knutson, "Interpretable classifiers for fMRI improve prediction of purchases," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 16, no. 6, pp. 539-548, 2008.
- [71] T. Hastie, A. Buja and R. Tibshirani, "Penalized discriminant analysis," *The Annals of Statistics*, vol. 23, no. 1, pp. 73-102, 1995.
- [72] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301-320, 2005.
- [73] G. Deshpande, L. E. Libero, K. R. Sreenivasan, H. D. Deshpande and R. K. Kana, "Identification of neural connectivity signatures of autism using machine learning," *Frontiers in human neuroscience*, p. 7, 2013.
- [74] V. Chattaraman, G. Deshpande, H. Kim and K. Sreenivasan, "Form 'Defines' Function: Converging Evidence from Functional MRI and Behavioral Studies on the Predictive Influence of Product Beauty on Purchase," *Journal of Marketing Research*, 2014, under review.
- [75] M. Poustchi-Amin, S. A. Mirowitz, J. J. Brown, R. C. McKinstry and T. Li, "Principles and Applications of Echo-planar Imaging: A Review for the General Radiologist," *Radiographics*, vol. 21, no. 3, pp. 767-779, 2001.
- [76] J. P. Mugler and J. R. Brookeman, "Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE)," *Magnetic Resonance in Medicine*, vol. 15, no. 1, pp. 152-157, 1990.
- [77] K. J. Friston, A. P. Holmes, K. J. Worsley, J. P. Poline, C. D. Frith and R. S. Frackowiak, "Statistical parametric maps in functional imaging: a general linear approach," *Human brain mapping*, vol. 2, no. 4, pp. 189-210, 1994.
- [78] L. Wang, *Support Vector Machines: Theory and Applications*, New York: Springer, 2005.
- [79] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *The Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2003.
- [80] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial intelligence*, vol. 97, no. 1, pp. 273-324, 1997.
- [81] J. Mourao-Miranda, E. Reynaud, F. McGlone, G. Calvert and M. Brammer, "The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data," *Neuroimage*, vol. 33, no. 4, pp. 1055-1065, 2006.
- [82] S. Ryali, K. Supekar, D. A. Abrams and V. Menon, "Sparse logistic regression for whole-brain classification of fMRI data," *NeuroImage*, vol. 51, no. 2, pp. 752-764, 2010.

- [83] I. Guyon, J. Weston, S. Barnhill and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine learning*, vol. 46, no. 1-3, pp. 389-422, 2002.
- [84] X. Yang, D. Lin, Z. Hao, Y. Liang, G. Liu and X. Han, "A fast SVM training algorithm based on the set segmentation and k-means clustering," *Progress in Natural Science*, vol. 13, no. 10, pp. 750-755, 2003.
- [85] N. Kriegeskorte, W. Simmons, P. Bellgowan and C. Baker, "Circular analysis in systems neuroscience: the dangers of double dipping," *Nature Neuroscience*, vol. 12, no. 5, pp. 535-540, 2009.
- [86] M. S. Rushworth, K. A. Hadland, T. Paus and P. K. Sipila, " Role of the human medial frontal cortex in task switching: a combined fMRI and TMS study," *Journal of Neurophysiology*, vol. 87, no. 5, pp. 2577-2592, 2002.
- [87] M. S. Rushworth, M. E. Walton, S. W. Kennerley and D. M. Bannerman, "Action sets and decisions in the medial frontal cortex," *Trends in cognitive sciences*, vol. 8, no. 9, pp. 410-417, 2004.
- [88] F. du Boisgueheneuc, R. Levy, E. Volle, M. Seassau, H. Duffau, S. Kinkingnehun, Y. Samson, S. Zhang and B. Dubois, "Functions of the left superior frontal gyrus in humans: a lesion study," *Brain*, vol. 129, no. 12, pp. 3315-3328, 2006.
- [89] K. Friston, P. Jezzard and R. Turner , "Analysis of functional MRI time series," *Hum Brain Mapping 2*, p. 69-78, 1994.