COMPUTATIONAL METHODS FOR PREDICTING BEHAVIOR FROM NEUROIMAGING DATA

BY LI ZHU

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Electrical and Computer Engineering

Written under the direction of

Prof. Laleh Najafizadeh

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2018

ABSTRACT OF THE DISSERTATION

Computational Methods for Predicting Behavior from NeuroImaging Data

by Li Zhu

Dissertation Director: Prof. Laleh Najafizadeh

One of the major goals in neuroscience is to understand the relationship between the brain function and the behavior. Inferring about the behavior, intent, or the engagement of a particular cognitive process from neuroimaging data finds applications in several domains including brain machine interfaces. To date, although a variety of imaging techniques have been developed and various computational techniques have been suggested, the estimation power has been limited to distinguishing very distinct classes of motor activities or cognitive processes. To improve the estimation power, there exist technical challenges that need to be addressed at the three stages of data acquisition (recording brain activities), data processing (processing brain recordings), and data analytics (inferring behavior from brain recordings).

The objective of the dissertation is to address technical challenges at the data processing and the data analytics stages, by leveraging tools from network science, machine learning and signal processing.

The first part of the dissertation focuses on data processing. In brain imaging experiments, typically, to reduce noise and to empower the signal strength associated with task-induced activities, recorded signals (e.g., in response to repeated stimuli or from a group of individuals) are averaged through a point-by-point conventional averaging technique. However, due to the existence of variable latencies in recorded activities, the use of the conventional averaging

technique can lead to inaccuracies and loss of information in the averaged signal, which may result in inaccurate conclusions about the functionality of the brain. To improve the averaging accuracy in the presence of variable latencies, we present new averaging framework that employs dynamic time warping (DTW) algorithm to account for the temporal variation in the alignment of functional Near-Infrared Spectroscopy (fNIRS). As a proof of concept, we focus on the problem of localizing task-induced active brain regions. The proposed framework is extensively tested on experimental data (obtained from both block design and event-related design experiments) as well as on simulated data. The proposed framework is shown to improve the accuracy of the averaging operation compared to conventional averaging techniques and is expected to introduce significant impact in various fNIRS-based neuroscience and clinical research studies.

The second part of the dissertation focuses on data analytics. We first address the problem of inferring behavior from neuroimaging data, by extracting new features based on the temporal characteristics of brain recordings. We hypothesize that there exist characteristics in the time course of cortical activities that are specific to the corresponding behavior. We introduce a method based on visibility graph (VG) to reliably identify such discriminatory characteristics in cortical recordings. An extensive study considering different choice of features and machine learning algorithms is conducted based on recordings obtained via widefield transcranial calcium imaging under spontaneous whisking condition, and recordings obtained via fNIRS under resting state and task execution conditions. It is shown, for the first time, that the characteristics of calcium recordings and fNIRS signals identified by the proposed method, carry discriminatory information that are powerful enough to decode behavior. The proposed method will have applications in advancing the accuracy of brain machine interfaces, and can open up new opportunities to study various aspects of brain function and its relationship to behavior.

Next, we propose to use multilayer perceptron (MLP) to perform classification based on the graph measures of the VGs. We also build a predictive framework using convolutional neural networks (CNN) to perform classification from the constructed multi-channel VGs directly. Multi-channel VGs allows the CNN to learn the discriminatory features utilizing the full temporal information encoded in the VGs, and can, hence, strengthen the inferring power. We evaluate the performance of both approaches using the widefield transcranial calcium imaging

data, and demonstrate improvement compared to classical machine learning methods.

Lastly, we investigate the temporal and spectral characteristics of brain function under spontaneous activity. By combining the wavelet transform coherence and the multivariate permutation test, we present a new method for investigating changes in functional connectivity under spontaneous activity. The proposed method is applied on the widefield transcranial calcium imaging data. Results on how function connectivity corresponding to two forms of spontaneous activity differ across frequency and space are presented.

Acknowledgements

The days and nights spent here at Rutgers University towards the Ph.D. degree have constituted one of the most memorable periods of my life. It has been a fruitful and rewarding experience for me. I would like to express my sincere gratitude to my advisor, Prof. Laleh Najafizadeh, for her generous support and guidance throughout the past years. I deeply appreciate Prof. Najafizadeh for introducing me to the multi-disciplinary field of functional brain imaging research, for inspiring me toward the right direction of the projects, for helping me patiently in scientific writings and presentations, and for offering me many excellent opportunities in attending world-class academic conferences. Her close and insightful guidance has provided me with numerous research opportunities, and her enthusiasm, leadership, and support have constantly encouraged me throughout my Ph.D. journey.

I would like to thank my Ph.D. dissertation committee, Ph.D. proposal exam committee and Ph.D. qualifying exam committee members, Prof. Yingying Chen, Prof. Kristin Dana, Prof. Janne Lindqvist, Prof. Yanyong Zhang, Prof. Saman Zonouz for their precious time in reviewing my dissertation and reports, for their insightful comments, and their generous availability. I would express my special gratitude to Prof. David Margolis, from the Department of Cell Biology and Neuroscience, who also served on my Ph.D. dissertation committee and Ph.D. proposal exam committee, for his insightful guidance and suggestions during the collaborative research. I would like to thank my co-author, Dr. Christian Lee, without whom the work on animal data would not have been completed.

Many thanks to the faculty members of the Department of Electrical and Computer Engineering, who helped me during my studies: Prof. Sasan Haghani, Prof. Shantenu Jha, Prof. Narayan Mandayam, Prof. Athina Petropulu, and Prof. Dipankar Raychaudhuri. Many appreciate to Prof. Han Xiao, from the Department of Statistics and Biostatistics, who instructed me

to learn the analysis of time series. I am also deeply indebted to the current and formal members at Rutgers Integrated Systems and Neuroimaging Laboratory, Dr. Yi Huang, Dr. Foroogh Shamsi, Maria Peifer, Fanpeng Kong, Tianjiao Zeng, Yunqi Wang, Siddharth Ravichandran, Fatemeh Koochaki, Shiva Salsabilian, and Anthorny Yang for all their help and friendship. Furthermore, I am very grateful to the Graduate Director Prof. Zoran Gajic, as well as current and former ECE staff, Mrs. Arletta Hoscilowicz, Mr. John Scafidi, Mrs. Christy Lafferty, Mr. John P. McCarthy, Mrs. Noraida Martinez, and Mrs. Tea Akins, who have given me excellent help.

A major portion of this research was supported by Siemens Healthineers. I am specially grateful to Dr. Dorin Comaniciu, Dr. Francisco Pereira, and Dr. Bin Lou. In the Summer of 2016, I got the opportunity to work at Siemens Healthineers in Princeton, NJ as a research intern. I am thankful to Dr. Francisco Pereira, and Dr. Bin Lou for the generous sharing of their time and expertise. This research was also funded in part by the National Science Foundation (NSF 1605646), the National Institutes of Health, and the New Jersey Commission on Brain Injury Research (CBIR16IRG032).

Last, but by no means least, I am indebted to all of my family members. My greatest appreciation goes to my parents, Xingbing and Zeping, for their unconditional love, support, and encouragement throughout my life. They taught me to think big and to never give up. Every step of my life could not have been made without their constant support and dedication. The appreciation and love also go to my grandparents, Bangyi, the late Yanxiang, the late Yiming, and the late Chen, who have given me all they have but never asked any returns. My gratitude and love go to my wife, Haifang, for her understanding, patience and sacrifice. Many thanks to my in-laws, Guoqin and Changyu, for their selfless support. The appreciation is also given to my aunt, Zeqing, and my late uncle Wenxuan. They have always been examples to follow as honest and hardworking scientists. Heartfelt thanks to my daughters, Caroline and Catalina, whose arrivals have brought so much joy to the whole family.

Dedication

To my beloved family

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

The human brain is known as one of the most complex systems in nature. It is a network containing around 100 billion neurons [9], and is known as one of the most complex systems in nature. To date, as humans, we still know little about how the neurons interact with each other in different scales to form networks, and how the brain activities are related to behavior.

The research on the brain helps to decrease the rate of brain-related disorders and to improve the quality of life of patients, their families, and the whole society. The burden of brain-related disorders is enormous. For example, in the U.S., patients suffering from Alzheimer's and Parkinson's diseases cost about $200 billion annually (see Fig. 1.1) [2]. Therefore, recently, the research and development in the fields of neuroscience, neurotechnology, and the related infrastructure has been attracting increased attention. The number of grants and filed patents in neurotechnology is experiencing an explosive growth at unprecedented rates (see Fig. 1.2) [3].

To understand the structure and the functionality of the brain we rely on various neuroimaging techniques. The goal of the proposed work is to develop computational methods to investigate the dynamics of brain activation patterns from functional brain imaging data, and use it to infer the behavior or brain states. In this chapter, we provide a brief overview of the research problems targeted in this work, and outline the contributions of the dissertation.

The first part of the dissertation focuses on data processing. In typical brain imaging experiments, to reduce noise and to empower the signal strength associated with task-induced activities, recorded signals (e.g., in response to repeated stimuli or from a group of individuals) are averaged through a point-by-point conventional averaging technique. However, due to

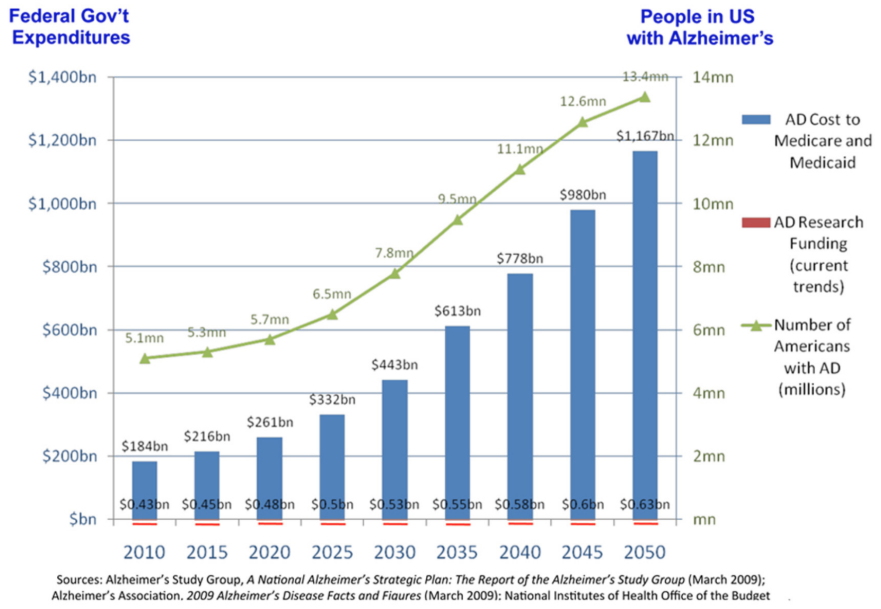


Figure 1.1: U.S. Alzheimer's disease: incidence, cost to Medicare/Medicaid and federal research funding [2].

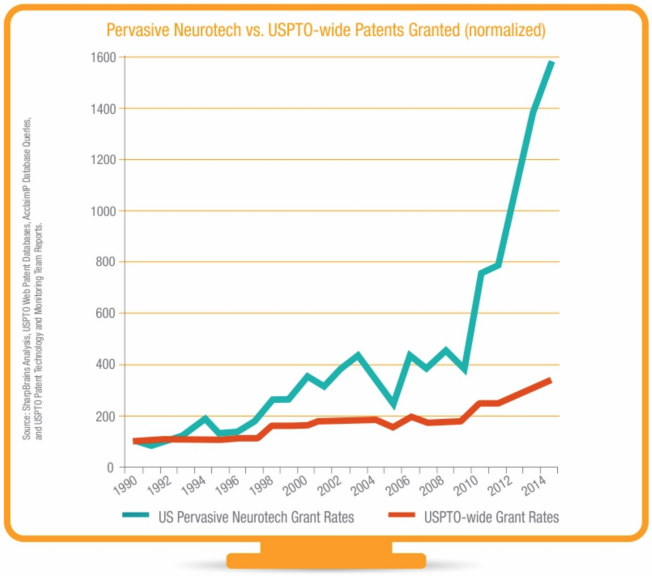


Figure 1.2: There is an explosion in neurotechnology related patent fillings and grants [3].

the existence of variable latencies in recorded activities, the use of the conventional averaging technique can lead to inaccuracies and loss of information in the averaged signal, which may result in inaccurate conclusions about the functionality of the brain. To improve the averaging accuracy in the presence of variable latencies, we present an averaging framework that employs dynamic time warping (DTW) to account for the temporal variation in the alignment of

functional near-infrared spectroscopy (fNIRS) recordings. As a proof of concept, we focus on the problem of localizing task-induced active brain regions. The proposed framework is extensively tested on experimental data (obtained from both block design and event-related design experiments) and simulated data.

The second part of the dissertation focuses on data analytics. To date, to infer the cognitive state of the subject from neuroimaging data, methods based on functional specificity and power spectral density have been suggested. However, the estimation power of such methods has been limited to distinguishing very distinct classes of motor activities or cognitive processes. As such the community has been searching for alternative methods to improve the power of inference. In this part, we present novel frameworks to infer behavior and the brain states by capturing characteristics based on the temporal dynamics of the recordings as well as based on functional connectivity. Various classical machine learning and neural networks models are utilized to perform the classification. These approaches are validated using brain imaging data acquired from animals and humans via widefield calcium imaging and fNIRS.

1.2 Outline and Contribution

Chapter 2 reviews the brain imaging techniques, with a special focus on three imaging modalities that have been used for this work (fNIRS, widefield calcium imaging, and the electroencephalography (EEG)). The multi-modal brain imaging and the experimental design techniques are then presented [10].

Chapter 3 focuses on a major challenge in data preprocessing of fNIRS recordings. The problem of non-linear misalignment of the repeated recordings during fNIRS experiments is first reviewed. A novel DTW-based averaging algorithm is then presented. Finally, the proposed method is validated using both experimental and synthetic datasets [11, 12].

Chapter 4, Chapter 5, and Chapter 6 focus on the data analytics based on visibility graph for data obtained from animal subjects and human subjects, respectively. In Chapter 4, we demonstrate that the temporal characteristics of calcium dynamics obtained through widefield imaging can be utilized to infer the corresponding behavior. In Chapter 5, we propose two new analysis frameworks, which combine visibility graph and neural networks, to further strengthen

the inferring power. In Chapter 6, the fractality of human fNIRS recordings is studied via both conventional methods and the VG-based method. We demonstrate that the VG-based measures can be used to distinguish human brain states (resting-state vs task-execution) [13, 1, 14].

Chapter 7 proposes a framework to study the functional brain connectivity based on the wavelet transform coherence and the multivariate permutation test. The proposed method is applied on data obtained from widefield transcranial calcium imaging on mice. Results on how functional connectivity corresponding to two forms of spontaneous activity differ across frequency and space are presented [15, 16].

Finally, a summary of the dissertation is given in Chapter 8. Future work are discussed.

Chapter 2

Review of Functional Brain Imaging Techniques

This chapter reviews brain imaging techniques, and the experimental design procedures conventionally used in functional neuroimaging studies.

2.1 Overview of Neuroimaging Tools

Neuroimaging techniques can be categorized into different classes: structural vs functional, invasive vs non-invasive, and direct vs indirect.

Structural brain imaging is used to study the anatomical aspects of the brain (providing a static view of the brain) and is targeted to detect abnormalities in brain structure (e.g. tissue). Examples of structural brain imaging techniques include magnetic resonance imaging (MRI), and computed tomography (CT). Functional brain imaging focuses on the investigation of the functionality of the brain (providing a dynamic view of the brain function). Examples of functional brain imaging techniques include fMRI, Magnetoencephalography (MEG) and fNIRS [17, 18]. Outcomes of functional brain imaging tools can be used for example to identify regions of the brain that are active during task execution.

Brain imaging techniques can be also categorized into invasive and noninvasive classes. Invasive brain imaging techniques, such as Electrocorticography (ECoG), require an open-skull surgery that places electrodes directly on the surface of the brain to record neuronal activity from the cortex. In contrast, non-invasive brain imaging techniques, including EEG, fMRI, and fNIRS, do not require surgery, and are therefore more suitable for human studies and have been widely used in the neuroscience research and clinical practices [18].

Brain imaging techniques, depending on the physiological variable they measure, can be categorized as direct or indirect. The electrical activity of neurons can be measured directly by brain imaging tools which are sensitive to electrical potentials or magnetic fields, including

EEG and MEG. On the other hand, brain activities elicit cerebrovascular changes for supplying oxygen and nutrients to active neurons. The brain image tools, such as fMRI and fNIRS, which are sensitive to changes in the oxygenation in the blood, can therefore be used to indirectly measure the underlying neuronal activities.

Fig. 2.1 compares the various functional brain imaging techniques for humans, with respect to the invasiveness and spatio-temporal resolution. It reveals that for non-invasive brain imaging techniques (shown in blue), none of them achieve high resolution in both temporal and spatial aspects [4].

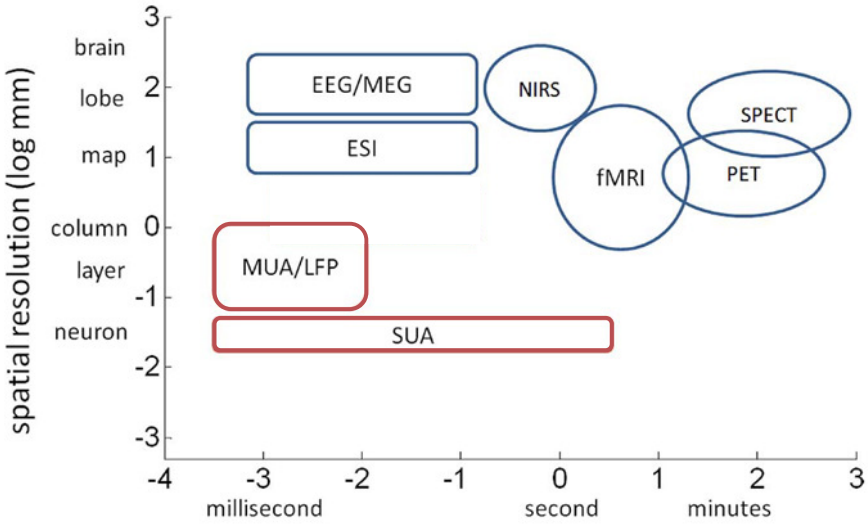


Figure 2.1: Comparison of various noninvasive (illustrated in blue) and invasive (illustrated in red) functional brain imaging techniques with respect to the spatio-temporal resolution [4] (ESI: electromagnetic source imaging, PET: positron emission tomography, SUA: single unit activity, SPECT: single-photon emission computed tomography).

2.2 Multi-Modal Brain Imaging

Multi-Modal brain imaging is the technique to conduct neuroimaging experiments combining two or more imaging modalities. While each of existing neuroimaging techniques has the limitation on the resolution from either spatial or temporal aspects, multi-modal brain imaging provides a possibility of providing the increased spatial-temporal resolution that not any individual modalities can achieve [4]. Multi-Modal brain imaging, when combining modalities that record signals related to both neuronal activity and hemodynamic response, allows for investigating the same brain events from different perspectives and enables us to understand how

brain functions at different functional levels.

The combination of EEG and fMRI has been used as a multi-modal imaging tool in recent studies [19]. However, EEG signals acquired through EEG-fMRI experiments suffer from artifacts induced by the magnetic field during fMRI image acquisition [20]. An alternative approach is to use the combination of EEG and fNIRS modalities. [21, 22]. Due to minimal interference between the optical and electrical measurements, the use of fNIRS in combination with EEG does not impact SNR.

Multi-Modal imaging has been recently used for 1) the investigation of the spatio-temporal relationship between the vascular/hemodynamic response and the neuronal activity (known as neurovascular coupling (NVC)) [23], 2) the enhancement of the performance of brain-computer interfaces (BCIs) [24], 3) the improvement in the accuracy of EEG source localization, by utilizing statistical maps of paradigm-related hemodynamic signals as a prior [25, 26, 27], and 4) helping in better understanding of brain-related diseases [28].

2.3 Optical Brain Imaging

After the fast growing in the past 30 years, optical brain imaging has become a rich and diverse field, and found a series of applications in both scientific research and clinical diagnoses [29]. Light can not only be used to imaging function *in vivo* in exposed animal brain, it can also be used as non-invasive imaging on human brain. Light provides superior sensitivity to functional changes through the intrinsic changes in absorption, scatter, fluorescence, or via the use of extrinsic contrast [29]. This section provides an overview of two optical brain imaging techniques that have been used in the proposed work, fNIRS and widefield calcium imaging.

2.3.1 functional Near-Infrared Spectroscopy

2.3.1.1 Physiological Basis

Brain function is regularized by series of complex mechanisms, including neurometabolic and neurovascular coupling mechanisms, which control the close interaction among neurons, glia cells, and cerebral vessels [30]. The consumption in energy substrates (i.e., oxygen and glucose), caused by neuronal activity, increases cerebral metabolic rate of oxygen ($CMRO_2$) on

one hand, which lowers the blood oxygen saturation (i.e. the fraction of oxygen-saturated hemoglobin relative to total hemoglobin), and increases the cerebral blood flow (CBF) and the cerebral blood volume (CBV) on the other hand, which raise the blood oxygen saturation [30]. This competition ends up with the increase in the local concentration of HbO_2 and the decrease in that of HbR in active brain regions [18, 31, 17]. Fig. 2.2 is an overview of the effect of neuronal activity on the local changes in $[HbO_2]$ and $[HbR]$.

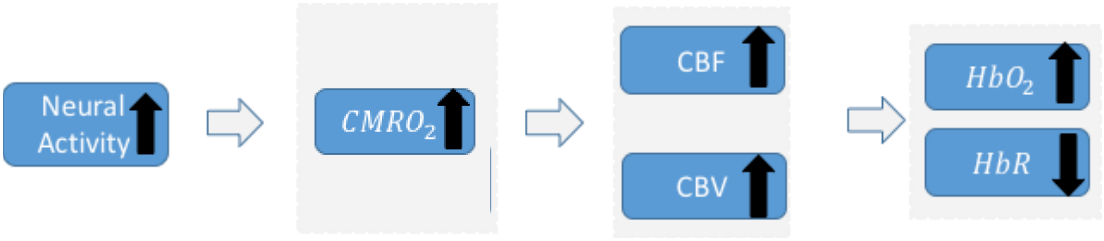


Figure 2.2: Overview of the effect of neuronal activity on the changes in hemodynamic signals.

fNIRS is a brain imaging technique used to indirectly measure the neuronal activities by measuring localized $\Delta[HbO_2]$ and $\Delta[HbR]$. The biological tissue is relatively transparent to the light in the NIR range ($690 - 1000\ nm$), where HbO_2 and HbR are the predominant absorbers.

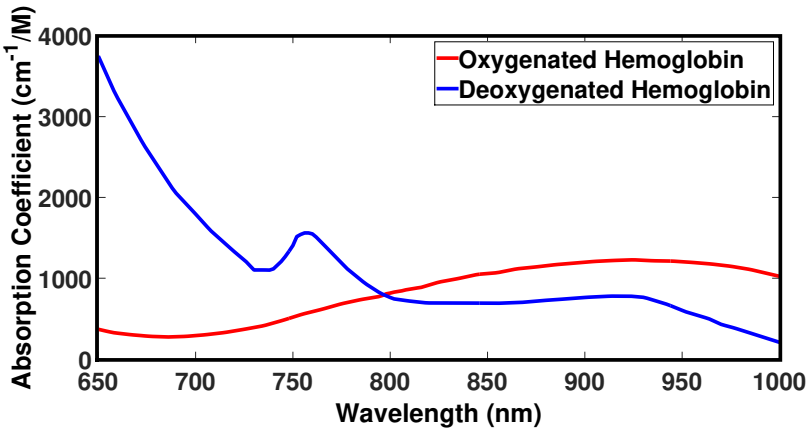


Figure 2.3: Absorption spectra for oxygenated hemoglobin and deoxygenated hemoglobin.

In a typical fNIRS experiment, an array of light emitters and detectors are placed on the surface of the head. The photons that enter the head at the emitter position are either absorbed or scattered. The detector, located centimeters away from the emitter position, samples a diffuse, banana-shaped volume along the path that is dependent on the source-detector distance. The

relative changes in the concentration of both chromophores (HbO_2 and HbR), as a function of time, are estimated by using the modified Beer Lambert law (MBLL) [11] from intensities measured through two or more wavelengths per emitter-detector pair. Due to the low optical absorption of the biological tissue at NIR range, the NIR light can penetrate deep enough to sample the outer $1.5-2$ cm of the head through the skin and skull and reach approximately $5-10$ mm of the brain tissue. Fig. 2.4-**a** conceptual illustrates the components involved in the fNIRS measurement [5].

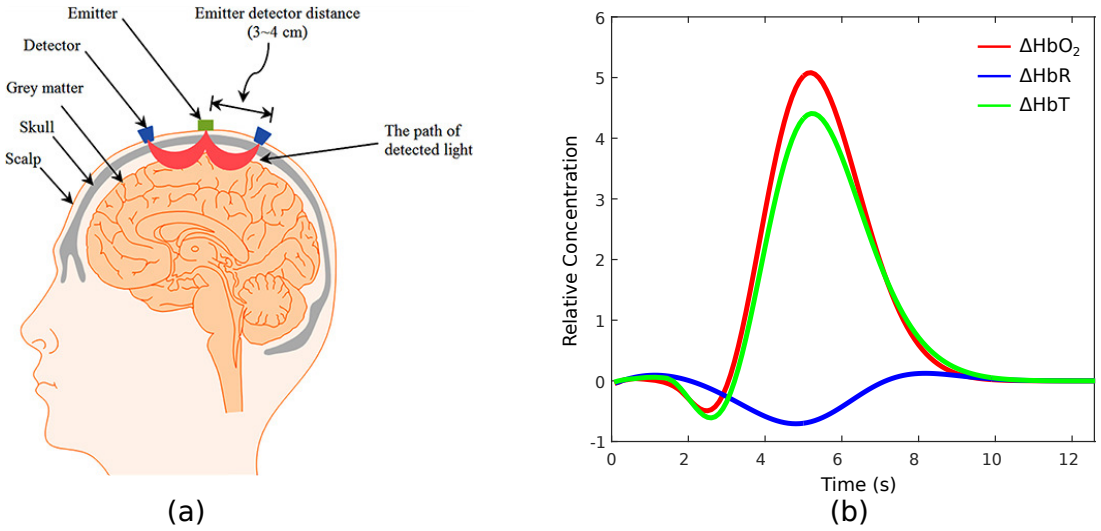


Figure 2.4: (**a**) Conceptual illustration of fNIRS emitter-detector pairs and the banana-shaped light path [5]. (**b**) Typical hemodynamic signals extracted from fNIRS measurement.

2.3.1.2 Modified Beer-Lambert Law

The Modified Beer-Lambert Law is used to convert the measured optical density to the localized $\Delta[HbO_2]$ and $\Delta[HbR]$, assuming the optical absorption remains homogeneous across the illuminated area and the scattering loss is constant over time [32, 33]. For a given source-detector pair (denoted here as channel m), the change in the optical density (ΔOD) can be described as [34]

$$|\Delta OD_m(\lambda, n)| = \ln \frac{I_{m,1}(\lambda, n)}{I_{m,0}(\lambda)} = -\Delta\mu_m(\lambda, n) \cdot DPF(\lambda) \cdot L_m, \tag{2.1}$$

where n is the time index, $I_{m,1}(\lambda, n)$ and $I_{m,0}(\lambda)$ are the acquired optical signals at the detector position during the task period and pre-stimuli-baseline period, respectively, $\Delta\mu_m(\lambda, n)$ is

the wavelength dependent absorption coefficient variation, $DPF(\lambda)$ is the unitless differential pathlength factor, and L_m is the source-detector separation for channel m [35, 34].

Since in the NIR wavelength range, HbO_2 and HbR are the major chromophores in the brain tissue, the change in the absorption coefficient between the task and the baseline conditions can be described as [34]

$$\Delta\mu_m(\lambda, n) = \epsilon_{HbO_2,\lambda}\Delta C_{HbO_2,m}(n) + \epsilon_{HbR,\lambda}\Delta C_{HbR,m}(n), \tag{2.2}$$

where $\epsilon_{HbO_2,\lambda}$ and $\epsilon_{HbR,\lambda}$ are the extinction coefficients for HbO and HbR, respectively, at wavelength λ. The ΔC_{HbO_2} and ΔC_{HbR} can be estimated by acquiring the optical attenuation at two wavelengths λ_1 and λ_2 simultaneously [34]

$$\begin{bmatrix} \Delta C_{HbO,m}(n) \\ \Delta C_{HbR,m}(n) \end{bmatrix} = \frac{1}{L_m} \cdot \begin{bmatrix} \epsilon_{HbO,\lambda_1} & \epsilon_{HbR,\lambda_1} \\ \epsilon_{HbO,\lambda_2} & \epsilon_{HbR,\lambda_2} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \Delta OD_m(\lambda_1, n)/DPF(\lambda_1) \\ \Delta OD_m(\lambda_2, n)/DPF(\lambda_2) \end{bmatrix}. \tag{2.3}$$

Fig. 2.4-**b** illustrates typical hemodynamic signals in response to an external stimulus. The stimulus-related neuronal activation causes an increase in $\Delta[\mathrm{HbT}]$ and $\Delta[\mathrm{HbO_2}]$ and a decrease in $\Delta[\mathrm{HbR}]$, to supply extra oxygen consumption requested by the neuronal activation. As the need of the neuronal activation are met, the hemodynamic signals return to the normal level. An initial dip is visible for $\Delta[\mathrm{HbT}]$ and $\Delta[\mathrm{HbO_2}]$ signals, which reveals the oxygenation changes prior to any subsequent blood flow increases [36].

2.3.2 Widefield Calcium Imaging

Widefield fluorescence imaging of genetically encoded calcium indicators (GECIs) is the imaging technique capable of recording large numbers of spatially distributed neurons, with high temporal resolution [37]. Newly developed GECIs such as GCaMP6 have improved sensitivity and brightness [38, 39], that, when expressed in transgenic reporter mice, enable imaging of neuronal activity of genetically defined neuronal populations over large portions of the cerebral cortex [40]-[43]. Compared to non-linear optical methods such as two-photon laser-scanning microscopy [44, 45], the use of epifluorescence optical imaging allows for easier implementation, higher temporal resolution, and much larger fields of view [46, 47]. Two-photon calcium

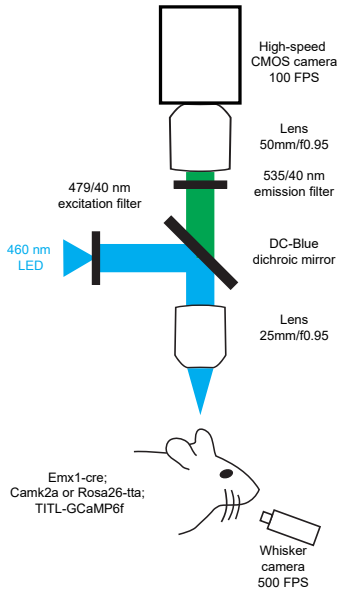


Figure 2.5: Illustration of the experimental setup used for widefield imaging of cortical activity of mice expressing GCaMP6f, and simultaneous recording of whisker movement.

imaging can be used to track individual neurons over time as animals learn [48, 49], but it is difficult to study neurons in spatially segregated cortical areas. Furthermore, long-term widefield imaging can be performed through either cranial windows or a minimally invasive intact skull preparation in living subjects over multiple weeks [50, 51]. These developments in widefield imaging have opened new possibilities for studying large-scale dynamics of brain activity in relation to behavior [51, 52], for example, during locomotion and active whisker movements in mice [40],[53]-[57].

In our experiments, the imaging of GCaMP6f is carried out in head-fixed mice with the transparent skull covered with glycerol and a glass coverslip. A schematic of the imaging system is shown in Fig. 2.5. A custom macroscope [50] allowed for simultaneous visualization of large area of the brain cortex (as seen in Fig. 2.5). The cortex is illuminated with 460 nm LED (Aculed VHL) powered by a Prizmatix current controller (BLCC-2). Excitation light is filtered (479/40; Semrock FF01-479/40-25) and reflected by a dichroic mirror (Linos DC-Blue G38 1323 036) through the objective lens (Navitar 25 mm / f0.95 lens, inverted). GCaMP6f fluorescence is filtered (535/40; Chroma D535/40m emission filter) and acquired using a MiCam Ultima CMOS camera (Brain vision) fitted with a 50 mm / f0.95 lens (Navitar). Images are captured on a 100×100 pixel sensor. The sampling rate is 100 frames per second.

2.4 Electroencephalography

2.4.1 Physiological Basis

EEG measures the brain scalp field potential (SP) that is generated by millions of neurons firing synchronously [58]. Since its first introduction, EEG has been widely used in neuroscience research and clinical studies [58, 18]. In a typical EEG experiment, electrodes are placed on the subject's head surface spreading across a cap. The potential difference of each electrode is measured with respect to a common reference.

Due to the volume conduction effect, the SP signal measured at one electrode location is the summation of contributions from a wide range of distinct brain regions, which could be characterized by the EEG linear forward model [59]. At each time point, modeling the finite number (K) of focal neuronal activity on the cortex as current dipoles $\mathbf{j} = [j_1, j_2, \cdots, j_K]^T \in \mathbb{R}^K$, the SPs picked up by M electrodes placed on the scalp, $\mathbf{y} = [y_1, y_2, \cdots, y_M]^T \in \mathbb{R}^M$, could be expressed as

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{g}_k \cdot j_k + \mathbf{e} = \mathbf{G} \cdot \mathbf{j} + \mathbf{e}, \tag{2.4}$$

where $\mathbf{g}_k = [g_{1k}, g_{2k}, \cdots, g_{Mk}]^T \in \mathbb{R}^M (k = 1, 2, \cdots, K)$ is the column vector representing the weight with which the source activity at dipole k (j_k) contributes to SPs acquired at each electrode (y_m), $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_K] \in \mathbb{R}^{M \times K}$ is the so-called lead-field matrix (or gain matrix) with dimensions equal to the number of EEG electrodes (M) by the number of cortical nodes (K), and $\mathbf{e}(t)$ represents the noise. The lead-field matrix $\mathbf{G}$ describes the relationship between the brain activity in the source space and the measurements obtained in the sensor space, and is dependent on the physical and geometrical properties of the head [60].

2.5 Experimental Design

In a task-based functional neuroimaging experiment, the participant is engaged in a task. The task is designed to evoke specific brain function such as motor movement, working memory, and visual/auditory identification. The most common types are "block-design" and "event-related design."

2.5.1 Block-Design Experiments

Studies which target localizing the brain regions related to a specific activity (i.e. “detection" studies), generally involve block-design experiments (illustrated in Fig. 2.6-**a**). In such experiments, to increase the detection power, the general idea is to closely present multiple trials, with the same type of stimuli, to the subject within each block, and repeat the experiment across multiple blocks. Blocks of different experimental conditions are often interleaved to reduce subject’s prediction of the coming stimuli. Recorded signals across blocks of experimental conditions of the same nature are then averaged, using conventional averaging techniques, and the related active areas are identified [11].

By conducting block-design experiments, adequate SNR can be attained from the recorded signals. But block-design paradigms have some disadvantages. Brain activities in response to individual trials cannot be distinguished.

2.5.2 Event-Related Experiments

In contrast to presenting multiple trials of the same type sequentially within each block, “event-related” experiments separate each trial in time by a randomized, but with sufficient length in time, inter-trial interval (ITI). Trials with different conditions can be interleaved. Depending on the research question of interest, the number of trials across conditions could be different (illustrated in Fig. 2.6-**b**).

While event-related experiments deliver decreased SNR, as compared to the block design experiments, they provide the possibility of estimating the time-course of the hemodynamic response [61].

2.6 Summary

In this chapter, brain imaging techniques and the experimental design procedures in functional neuroimaging studies were reviewed. Multi-modal brain imaging, details of fNIRS, widefield calcium imaging, and EEG were described. Conventional procedures for designing paradigms were discussed.

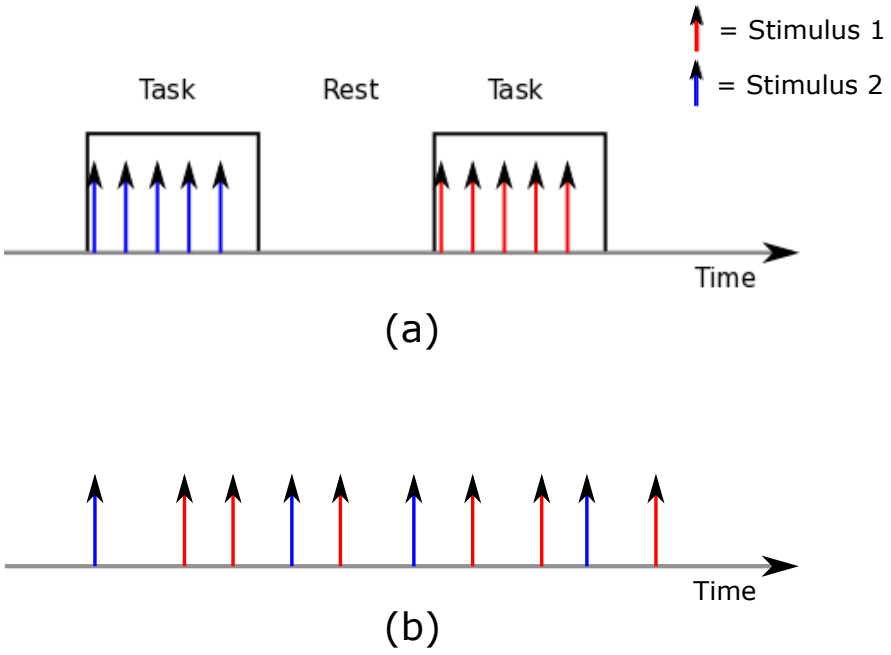


Figure 2.6: (**a**) Illustration of a block-design experimental paradigm. (**b**) Illustration of the an event-related experimental paradigm.

Chapter 3

Data Processing: Dynamic Time Warping-Based Averaging Framework

In this chapter, we first describe the problem with conventional averaging approaches used in fNIRS experiments. We then introduce a novel averaging method based on the DTW algorithm which helps to improve the accuracy of the averaged signal across multiple blocks, hence increasing the detection power. The technique is validated using both real experimental fNIRS data and synthetic data.

3.1 Introduction

In fNIRS brain imaging studies, the averaging operation is performed at different stages of analysis (e.g. across trials, blocks, subjects, channels), with the objective of enhancing the signal strength associated with task-induced brain activities and reducing noise and randomness. For example, in task-based fNIRS studies, the averaging operation is used at the early stages of analysis. Task-based experimental paradigms are categorized as "block design" and "event-related design." In block design experiments, to increase the detection power for estimating the location of task-induced active regions, the general idea is to present multiple trials of the same type for the subject within each block, and repeat the experiment across multiple blocks. Blocks of different experimental conditions are often interleaved. Recorded fNIRS signals across blocks of similar conditions are averaged through the conventional point-by-point averaging technique [62, 34, 63, 64, 65, 66, 67]. In event-related design studies, brain activities associated with individual trials are recorded, allowing for estimating the brain's hemodyamic response related to the stimulus. This hemodyamic response can be obtained through averaging recorded activities in response to several discrete events of the same type. In both categories, to

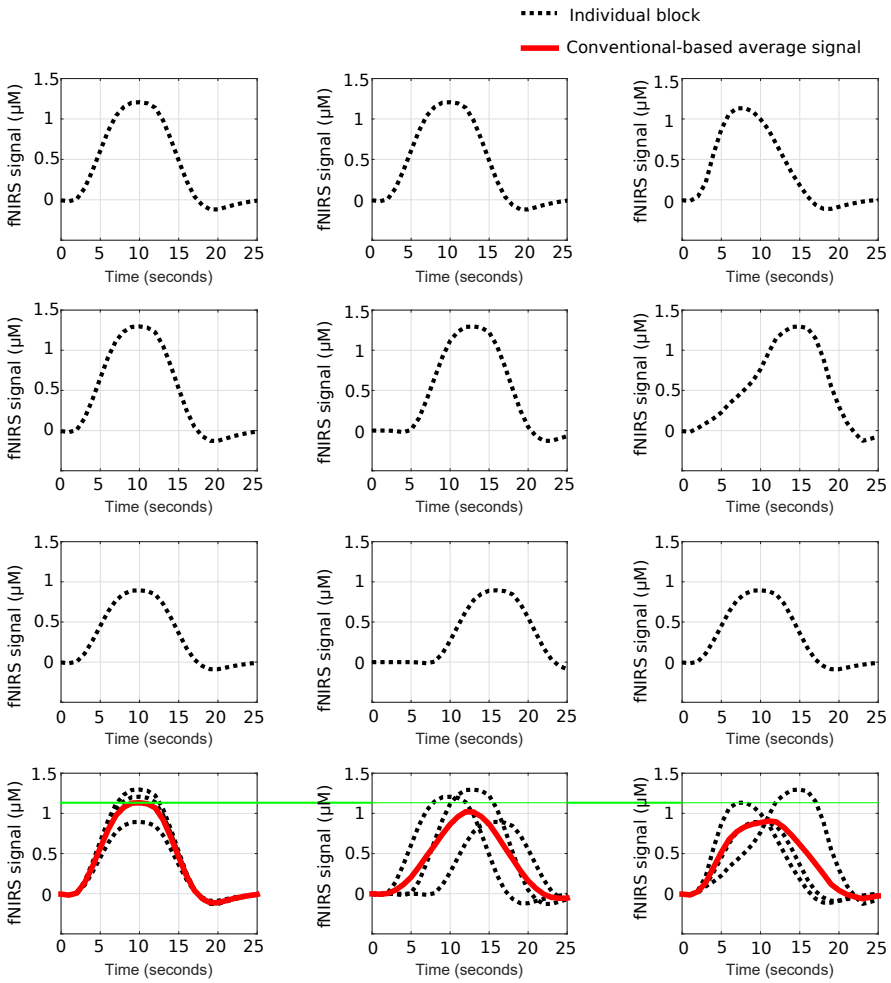


Figure 3.1: Visual illustration of using point-by-point conventional averaging technique for three scenarios: a) when three signals are temporally aligned, b) when there exist *linear* temporal variation in the alignment of three signals, and c) when there exist *nonlinear* temporal variation in the alignment of three signals.

estimate the location of brain regions associated with the task, the stimulus or the event of interest, various statistical tests (e.g. student's *t*-test) are used to evaluate the statistical significance of features (e.g. amplitude) of the averaged signals [68].

As the averaging operation is conducted at the early stages of the analysis [62, 69, 70, 71, 63, 64], inaccuracies in the averaged signal could lead to type I (incorrectly detecting a region as active) or type II (incorrectly detecting a region as inactive) errors in the statistical analysis, resulting in inaccurate conclusions about the functionality of the brain. As stated before, to perform the averaging operation, typically conventional point-by-point averaging technique is used in fNIRS studies. Previous work however, have shown that there exist variability (e.g.

latency differences) in the brain response to trials of the same type (e.g. trial-to-trial variability) due to for example differences or delays in neural responses, or individual's performance [72, 73, 74, 75]. Furthermore, in patient populations (e.g. patients with Autistic Spectrum Disorders (ASD)) several studies have reported variable latencies in their responses to stimuli [76, 77]. While preserving information related to variability would be important in identifying parameters related to behavioral variability (e.g. understanding neural mechanisms related to variability in response time) [72, 75, 78], in several neuroimaging studies (e.g. those interested in functional specificity) conclusions are made based on an "aggregate-then-analyze" approach. These types of studies, that use conventional point-by-point averaging approach and does not take temporal variation of signals into account, can result in inaccuracies (e.g. blurring the peaks and valleys) in the averaged signal. Examples of three scenarios are conceptually visualized in Fig. 3.1. In the scenario shown in Fig. 3.1-a (left column), no temporal variation exists in the alignment of the three signals to be averaged. In the scenario shown in Fig. 3.1-b (middle column), a *linear* temporal variation exists among the three signals. In the scenario shown in Fig. 3.1-c (right column), a *nonlinear* temporal variation exists among the three signals. The averaged signals obtained through point-by-point averaging technique for each scenario are shown in the last row. It can be seen that, compared to Fig. 3.1-a, the presence of linear temporal variation for signals in Fig. 3.1-b has resulted in attenuation of the amplitude of the averaged signal. The presence of nonlinear temporal alignment for signals in Fig. 3.1-c has caused nonlinear distortion in the averaged signal obtained through conventional point-by-point technique. In reality, such inaccuracies in the averaged signal can lead to misunderstandings about the brain function. We had also briefly previously discussed this problem in [11].

To address the problem of temporal variation across fNIRS signals, in this chapter, a dynamic time warping (DTW)-based averaging technique is presented for fNIRS-recorded time series. DTW algorithm was originally introduced in speech processing [79], and has been widely used for measuring the similarity of two time series in various fields of research, such as biometric, data mining, gene expression analysis, human motion recognition, and EEG signal analysis [80, 81, 82, 83, 73, 84]. Various forms of DTW-based averaging approaches have also been proposed for applications such as EEG [73] and satellite image time series [85].

To extensively examine the performance of the proposed DTW-based averaging technique

in fNIRS-based detection studies, experiments and simulation are performed. For the experimental part, both block design and event-related design experiments are considered. For the simulation study, a dataset is generated based on the standard model of hemodynamic response function, and operating characteristic curves (ROC) are used to compare the performance of conventional- and DTW-based averaging techniques. Both experimental and simulation studies show that DTW-based averaging technique outperforms the conventional averaging technique in terms of accurately estimating the location of task-induced brain regions. While for the proof of concept, in this chapter, we focus on the problem of localizing task-induced active brain regions, the presented averaging technique can be used in other averaging stages of fNIRS neuroimaging studies (e.g. computing grand averages).

The remainder of this chapter is organized as follows. In Section 3.2 the DTW-based averaging technique is described. In Section 3.3 the experimental and simulation studies are presented and results are discussed. Finally, the chapter is summarized in Section 3.4.

3.2 DTW-Based Averaging

Let $\mathbf{b_k} = [b_k(1), \cdots, b_k(N)]$ represents the k^{th} time series of a group of K time series, each with N time points. In fNIRS neuroimaging experiments, each of these time-series corresponds to the signal from a channel associated with a given block (in a block design experiment, K being total number of blocks) or associated with a given trial (in an event-related design experiment, K being total number of events), recorded from a channel. The aim is to obtain the averaged representation of these K time-series using DTW [85].

To obtain the DTW-based averaged representation of K signals, first, the best alignment between each signal and a "reference" signal $\mathbf{c} = [c(1), \cdots, c(N)]$ is found. The reference signal could be for example the conventional averaged signal of all K time-series. To optimally align signal $\mathbf{b_k}$ and the reference signal $\mathbf{c}$, a cost matrix $\mathbf{D_k}$ needs to be determined. $\mathbf{D_k}$ is a $N \times N$ matrix in which its elements are obtained through a cost function representing the discrepancy between the i^{th} and j^{th} samples ($i = 1, \cdots N$, $j = 1, \cdots N$) of signals $\mathbf{c}$ and $\mathbf{b_k}$. Measures of the Euclidean distance, or the square of difference between normalized samples, can be used as the cost function [73].

Next, from the cost matrix, an optimal alignment path, $\mathbf{W_k}^{\mathbf{opt}} = [w_1, \cdots, w_L]^T$, $N \leq L \leq 2N-1$, where $w_l = (i(l), j(l)), 1 \leq l \leq L, 1 \leq i(l), j(l) \leq N$, must be determined so that the overall similarity between the two signals is maximized. The optimal alignment path $\mathbf{W_k}^{\mathbf{opt}}$ shows how the mapping between the indices of the two time series $\mathbf{c}$ and $\mathbf{b}_k$ must be made to achieve the best alignment. For example, if $\mathbf{W_k}^{\mathbf{opt}} = [(1,1),(2,3),(2,4),...,(N,N)]^T$, then the sample $c(1)$ is aligned with $b_k(1)$, $c(2)$ is aligned with $b_k(3)$ and $b_k(4)$, etc. To obtain $\mathbf{W_k}^{\mathbf{opt}}$ the solution to the following optimization problem [73]

$$\min \sum_{l=1}^{L} D_k(i(l), j(l)), \tag{3.1}$$

subject to the following constraints should be found [73]:

- *Monotonicity Alignment*: The search for the alignment path must be monotonic, so that the natural time ordering in the sequence is preserved, i.e.,

$$i(l) \geq i(l-1), \text{and } j(l) \geq j(l-1), \tag{3.2}$$

- *Continuity*: The alignment function does not skip any samples in two sequences, i.e.,

$$i(l) - i(l-1) \leq 1, \text{and } j(l) - j(l-1) \leq 1, \tag{3.3}$$

- *End-point Alignment*: The first and the last points of the sequences must be aligned, i.e.,

$$i(1) = j(1) = 1, \text{and } i(L) = j(L) = N.$$

Once $\mathbf{W_k}^{\mathbf{opt}}$ is obtained, a new N-points time series, $\mathbf{b}_{\mathbf{k(aligned)}} = [b_{k(aligned)}(1), \cdots, b_{k(aligned)}(N)]$, is formed as follows

- if the index represented by $i(l)$ in $\mathbf{W_k}^{\mathbf{opt}}$ is unique, $b_{k(aligned)}(m) = b_k(j(l))$, $(m = 1, \cdots, N)$,

- if the index represented by $i(l)$ in $\mathbf{W_k}^{\mathbf{opt}}$ is not unique, $b_{k(aligned)}(m) =$ average of all $b_k(j(l))$'s corresponding to $i(l)$.

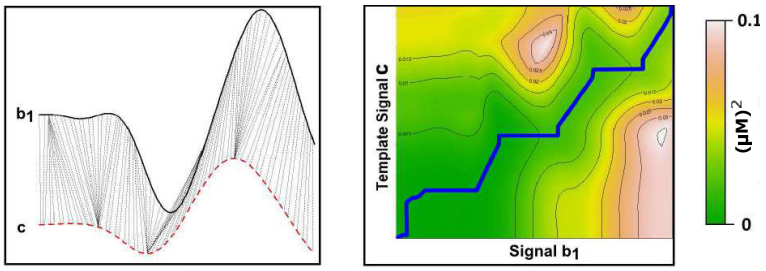


Figure 3.2: Illustration of the alignment process for an exemplary signal $\mathbf{b_1}$ and a reference signal $\mathbf{c}$ through DTW, a) time series $\mathbf{b_1}$ and $\mathbf{c}$ where their aligned points are connected, and b) color-coded cost matrix, representing the distance between the two time series. The warping path is shown in blue. Plots are created using R programming package [6].

As an example, if

$$\mathbf{W_k}^{\mathbf{opt}} = [(1,1),(2,2),(2,3),(2,4),(3,5),\cdots,(N-1,N-1),(N,N)]^T, \tag{3.4}$$

then, $\mathbf{b_{k(aligned)}}$ is obtained as

$$\mathbf{b_{k(aligned)}} = [b_k(1), \frac{b_k(2)+b_k(3)+b_k(4)}{3}, \cdots, b_k(N)]. \tag{3.5}$$

This procedure will be performed for all K signals. Once all signals are aligned with the reference signal $\mathbf{c}$ and their aligned representations are determined, the DTW-based averaged signal is obtained as

$$\mathbf{b_{DTWaveraged}} = \frac{\sum_{k=1}^{K} \mathbf{b_{k(aligned)}}}{K}. \tag{3.6}$$

Fig. 3.2 illustrates an example of aligning signal $\mathbf{b}_1$ with a reference signal $\mathbf{c}$. The color-coded cost matrix and the obtained optimal warping paths (in blue) are shown in Fig. 3.2-b. It can be observed that when a sample in $\mathbf{b}_1$ is aligned with several samples in $\mathbf{c}$ the warping path has a vertical direction, while when a sample in $\mathbf{c}$ is aligned with several points in $\mathbf{b}_1$, the warping path follows a horizontal direction. Note that the optimal warping path is mostly along the anti-diagonal elements of the cost matrices, illustrating that the two signals experience temporally variable latencies. Table 3.1 summarizes the steps involved in the DTW-based averaging procedure using a reference signal.

The DTW-based averaging technique can be realized in various forms [85]. For example, in [73], instead of using a reference signal, the average is obtained sequentially in a pair-wise

Table 3.1: Summary of steps involved in DTW-based averaging procedure using a reference signal.

Step	Operation
1	Define a reference signal $\mathbf{c}$.
2	For each signal $\mathbf{b}_k$, $k = 1, 2, \cdots, K$, generate the cost matrix representing the discrepancy between $\mathbf{b}_k$ and $\mathbf{c}$.
3	Based on the cost matrix, for each signal, find the optimal alignment path such that the overall similarity between the corresponding signal and the reference signal is maximized.
4	Find the average of aligned signals.

manner. As such, errors at early stages of computation could propagate throughout the averaging process, resulting in loss of information [85]. To address the problems associated with pair-wise averaging, a global averaging strategy was introduced in [85], in which the averaged signal is obtained considering all signals and is updated through an iterative process. The method presented here also computes the average considering all signals and therefore, will not be sensitive to ordering effects.

3.3 Evaluation

The performance of the proposed averaging framework for analyzing fNIRS-recorded time series was evaluated through both experimental and simulated data. On the experimental end, two experiments, one block design and one event-related design, are considered. The recorded time series are analyzed through both conventional and DTW-based averaging techniques. As it will be shown, compared to conventional point-by-point averaging, when DTW-averaging is used, increase in the detection power for the block-design task and decrease in false positive rate for the event-related task are observed. For the simulation study, data using a widely-used equation of the hemodynamic response is generated, and ROC curves for the conventional and the DTW-based averaging are obtained and compared. Furthermore, the impact of choice of reference signal on the performance of DTW-based averaging technique is investigated.

3.3.1 Experimental Studies

Two experiments, one block design (Experiment I), and one event-related design (Experiment II) are performed. We first describe the experimental setup for each case, and then present and

discuss the results.

3.3.1.1 Experimental Setup

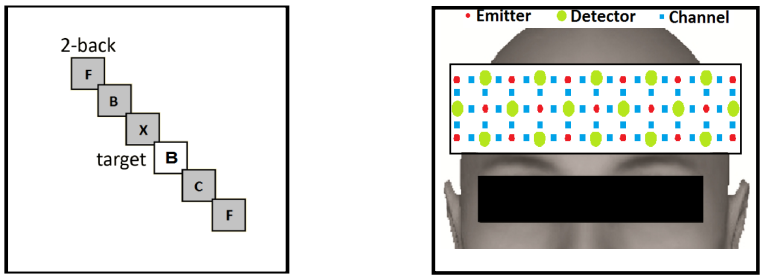
- **Experiment I: Block Design Paradigm**

  Five right-handed healthy volunteers (one female) participated in Experiment I. All volunteers gave their informed consent approved by the Rutgers University Institutional Review Board (IRB) prior to the experiment [11]. The paradigm for this experiment was the 2-back task (see Fig. 3.3-a), which has been widely used in determining brain activities related to working memory [86, 66, 67]. Three blocks were presented to participants. Each block lasted for 32 s, and consisted of 16 letters (computerized in a pseudorandom order with four being target stimuli) with inter-trial interval (ITI) of 2 s. The order of presentation of target stimulus was different across blocks. Participants were asked to respond, by pressing the right button on the mouse, when the presented letter matched the letter shown two stimuli back.

  It is worth to note that the block design experiments are performed with the goal of continuously engaging the brain in the cognitive process of interest over the duration of the block, in order to increase the detection power for localizing the related activities in the brain. Here, the cognitive process of interest is "working memory". The process of memorizing and recovering letters is done on a continuous basis over the duration of the block for each letter (target or non-target). For letters that satisfy the 2-back condition, other additional processes and functions, such as pressing the button (activating the somatosensory region), will become involved. Through the averaging process across blocks, the SNR related to these additional functions would be small, because of the variations in the timing of the target stimulus from one block to other. But information related to "working memory", which is the main objective of the n-back task, should be preserved over the duration of the block.

  fNIRS data were collected using a 52-channel (Hitachi ETG-4000, 17 sources (lasers, 695 nm and 830 nm) and 16 detectors) at a sampling rate of 10 Hz. The source-detector

separation was 3 cm. The headband optode holder was placed on the forehead of volunteers to cover the prefrontal cortex, as shown in Fig. 3.3-b.



(a) (b)

Figure 3.3: a) Experimental paradigm, and b) Optode setup (Red: light emitter, Green: detector, Blue: fNIRS channel), for the 2-back task (Experiment I).

- **Experiment II: Event-Related Paradigm**

  Six right-handed healthy volunteers (all males) participated after giving their informed consent. An event-related modified visual oddball task [84, 87] consisting of three graphical stimuli (plus, square, circle), presented in random order, was used as the paradigm (see Fig. 3.4-a). A total of 220 stimuli (30 target stimuli ("plus"), and 190 non-target stimuli) were presented. Each stimulus was presented for 50 ms, with ITI of $10 - 12$ s. To minimize the periodic systemic effects [88], the ITIs were randomized to prevent the subjects from predicting the presentation time. Participants were asked to press the left button of the mouse once the target stimulus was shown on the screen.

  fNIRS data were measured by a NIRx System (NIRx NIRScout, 16 sources (LEDs, 760 nm and 830 nm) and 16 detectors) at the sampling rate of 10.42 Hz. Measured signals from source-detector pairs with a separation distance of 3 cm were considered, resulting in a total of 38 channels. Optodes were placed over the prefrontal and visual regions of the cortex, as shown in Fig. 3.4-b.

3.3.1.2 Preprocessing

Recorded signals were visually inspected. Bad channels (for example due to loose contact to skin) were excluded from further analysis. In Experiment II, the response to the first trial for majority of participants included traces of subject's movement, and so it was also excluded

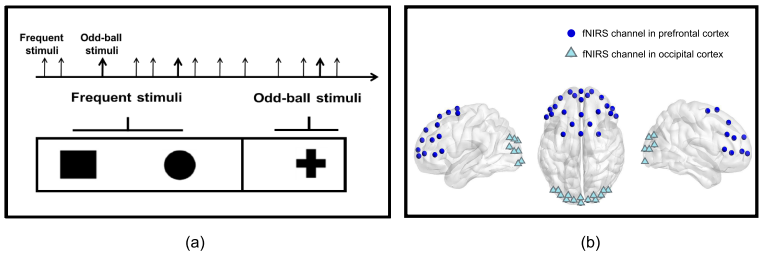


Figure 3.4: a) Experimental paradigm, and b) Optode setup (Dark blue: fNIRS channels in prefrontal cortex, Light blue: fNIRS channels in visual cortex), for the modified visual oddball task (Experiment II). Locations of optodes are visualized using MATLAB, BrainNet Viewer [7].

from the analysis, for all subjects. Signals were then segmented. For Experiment I, signals were segmented by blocks. For each block, the segmentation time window began at 5 s prior to the onset of the first stimulus in the block and ended 15 s after the end of the block. For Experiment II, signals associated with “target" trials and “non-target” trials were segmented separately. Trials associated with missed response (i.e. target trials without subject’s response) were excluded from further analysis. The segmentation window began at 1.9 s prior to the onset of the stimulus and had a duration of 13.9 s.

Next, signals related to $\Delta[\mathrm{HbO}_2]$ and $\Delta[\mathrm{HbR}]$ with respect to the baseline were extracted using the modified Beer Lambert law (MBLL) [89].

Band-pass filters ($0.01 - 0.2$ Hz) were then implemented to remove artifacts and low-frequency drift. Furthermore, following [90] a procedure for the automatic detection of fast changes in signals was implemented and applied to signals from Experiment II, through which signals that showed sudden fast changes where excluded from further analysis.

3.3.1.3 Results-Experiment I

For each subject and each channel, averaged signals, using both the conventional-based and DTW-based averaging techniques, were first obtained. Next, for each averaged signal, the time point at which the signal reaches its maximum value following the onset of stimuli was identified. A temporal window with a duration of 2.1 s [91] around this time point was considered, and the mean of the averaged signal within this window was calculated. This number, to which we refer to as Activation Index (AI), was used in subsequent statistical tests to determine if the region associated with the channel was active in response to the external stimuli.

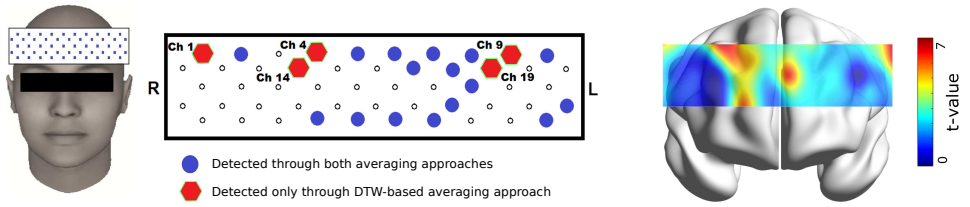


Figure 3.5: a) Statistical activation map for 2-back task based on the averaged signal obtained from the conventional-based averaging technique and the DTW-based averaging technique. Blue-colored circles indicate channels with significant increase in $\Delta[\mathrm{HbO}_2]$ identified based on averaged signal obtained thorough both averaging techniques. Red-colored circles indicate channels with significant increase in $\Delta[\mathrm{HbO}_2]$ that were only detected from the averaged signal obtained through the DTW-based averaging. The statistical significance level is $p < 0.05$. b) Statistical map illustrating the difference between the DTW-based and conventional-based averaged signals.

Table 3.2: Mean and standard deviation (SD) of AIs across subjects (units in μM) as well as results of t-test for channels showing statistically more significant activation when DTW-based averaging technique is used as compared to when conventional point-by-point averaging technique is used (d.f.$= 4$).

Channel	**Conventional averaging technique**			**DTW-based averaging technique**		
	Mean (S.D.)	p**-value**	t**-value**	**Mean (S.D.)**	p**-value**	t**-value**
1	0.65 (0.59)	0.068	2.48	0.84 (0.67)	0.048	2.82
4	0.45 (0.40)	0.068	2.49	0.61 (0.41)	0.035	3.14
9	0.50 (0.48)	0.081	2.32	0.73 (0.49)	0.029	3.34
14	0.44 (0.42)	0.082	2.31	0.61 (0.49)	0.048	2.83
19	0.98 (0.95)	0.082	2.31	1.30 (0.97)	0.040	3.00

To determine active regions, one-sample t-test (with $p < 0.05$) was performed on AIs (obtained through each averaging technique), with the null hypothesis being the region is not active [91]. Fig. 3.5 illustrates the result of the statistical test. As can be seen five channels (channels 1, 4, 9, 14, and 19) were identified as being significantly active in response to the task when DTW-based averaging technique is used to obtain the averaged signal. Table 3.2 summarizes the result of the statistical test for these channels and each averaging technique. These channels were not detected as active channels when conventional-based averaging technique is used. The activation pattern obtained through DTW-based averaging technique appears to be more consistent with the results reported in previous fMRI studies where activation in bilateral prefrontal cortex was observed in response to working memory tasks [92, 93].

While Fig. 3.5-a illustrates the statistical activation map by employing a fixed threshold

($p < 0.05$), it would be informative to also investigate the statistical significance of the difference of the outcomes of the two analysis approaches. To achieve this, for each channel and each subject, we subtracted the conventional-based averaged signals from their DTW-based averaged counterparts and conducted statistical test on the difference signals, with the null-hypothesis that there are no significant differences. Fig. 3.5-b presents the t-map obtained from the one-sample right-tailed students' t-test. For almost all regions, the obtained t-values are significantly larger than zero, indicating that there are statistically significant differences between the averaged signals obtained from the two techniques. Next, we used the metric of contrast-to-noise-ratio (CNR) [94, 95] to quantify the signal to noise ratio (SNR) for the averaged signals obtained from each averaging technique. We considered both $\Delta[\mathrm{HbR}]$ and $\Delta[\mathrm{HbO_2}]$ signals. Only channels that were identified as active through both averaging techniques (shown in blue in Fig. 3.5) were considered in this analysis. Denoting *mean*(**DUR**) and *var*(**DUR**) as the mean and variance of the signal amplitude during $5-15$ s after the onset of the first stimulus of the block (considering the delay in the hemodynamic signals), and *mean*(**ITI**) and *var*(**ITI**) as the mean and variance of the signal amplitude corresponding to $10-15$ s after the presentation of the last stimulus of the block, the CNR is computed as

$$\mathrm{CNR} = \frac{|mean(\mathbf{DUR}) - mean(\mathbf{ITI})|}{\sqrt{var(\mathbf{DUR}) + var(\mathbf{ITI})}}. \tag{3.7}$$

Results are shown in Fig. 3.6. It is shown that for $\Delta[\mathrm{HbO_2}]$, mean CNR across subjects is significantly higher when DTW-based averaging technique is used, as compared to when conventional averaging technique is used ($p < 0.05$, paired t-test, d.f. = 15). Since $\Delta[\mathrm{HbR}]$ signals are generally weaker than $\Delta[\mathrm{HbO_2}]$ signals, we observe less significant difference between the CNR in $\Delta[\mathrm{HbR}]$ signals obtained from the two approaches, though the mean CNR obtained through the DTW-based averaging technique is still higher than that obtained through the conventional-based approach. These results show that the DTW-based averaging technique has increased the SNR in the averaged signals, and can enhance the detection power in studies that aim to identify active brain regions associated to external stimuli.

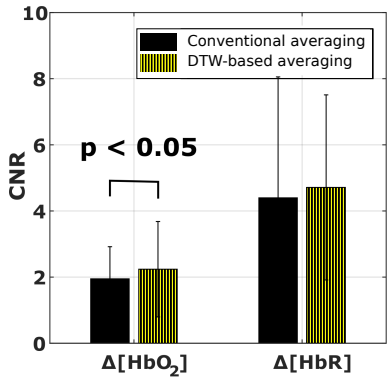


Figure 3.6: Comparison of mean CNR values in averaged $\Delta[\mathrm{HbO}_2]$ and $\Delta[\mathrm{HbR}]$ signals, for Experiment I, obtained through conventional- and DTW-based averaging techniques, respectively.

3.3.1.4 Results-Experiment II

Experiment II is an event-related experiment. The histogram of response time for the "target" stimuli for each subject is shown in Fig. 3.7. Variability in response time across trials is observed for all individuals. This variability can also be an indication of the presence of temporal variation in the timing of the occurrence of task-induced events in the recorded fNIRS signals, and if not addressed, would affect the accuracy of the averaged signal.

For each subject, $\Delta[\mathrm{HbO}_2]$ signals from each channel were averaged for each condition (target and non-target), using the conventional and the DTW-based averaging techniques, separately. Fig. 3.8 shows examples of $\Delta[\mathrm{HbO}_2]$ signals and their averaged signals obtained through both averaging techniques, under the "target" condition (shown in left) and "non-target" condition (shown in right) for channel 1, for a representative subject. It is observed that in both cases, the peak amplitude of the conventional-based averaged signal is lower than that of the DTW-based averaged signal. Furthermore, comparing the two conditions, the peak amplitude of the conventional-based averaged signal for the non-target condition is lower than the peak amplitude of the conventional-based averaged signal for the target condition, which may indicate that the brain region under channel 1, has been active for the target response (and hence sensitive to the oddball effect). However, caution should be taken when making inference from this observation to avoid false alarm. Indeed, when the averaging process is conducted using DTW-based averaging technique, the difference in the peak amplitude of the two conditions is minimized, and such conclusion can not be made.

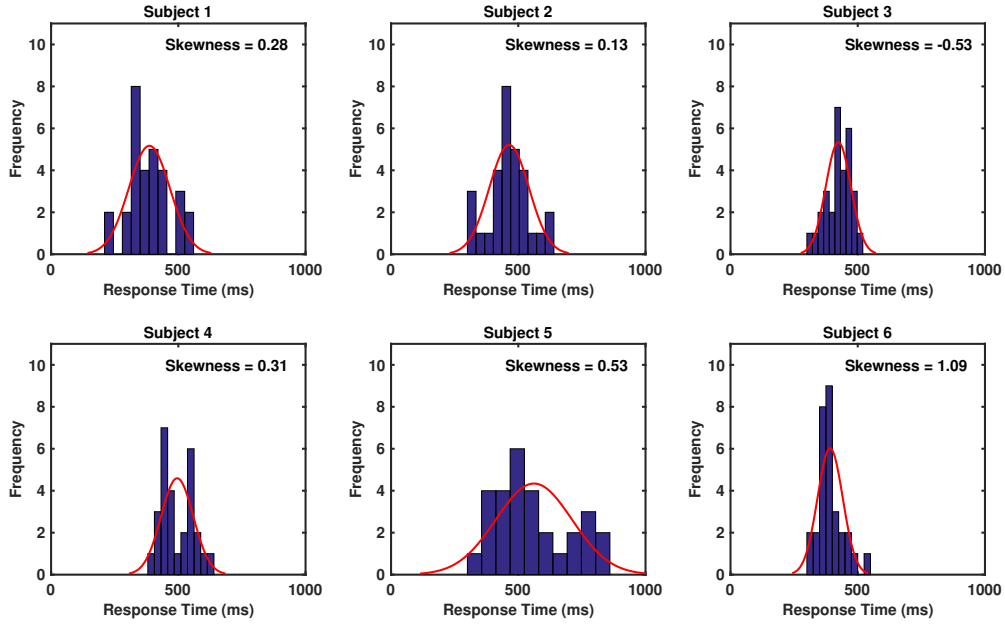


Figure 3.7: Histogram of the response time for "target" stimuli for individual subjects in Experiment II.

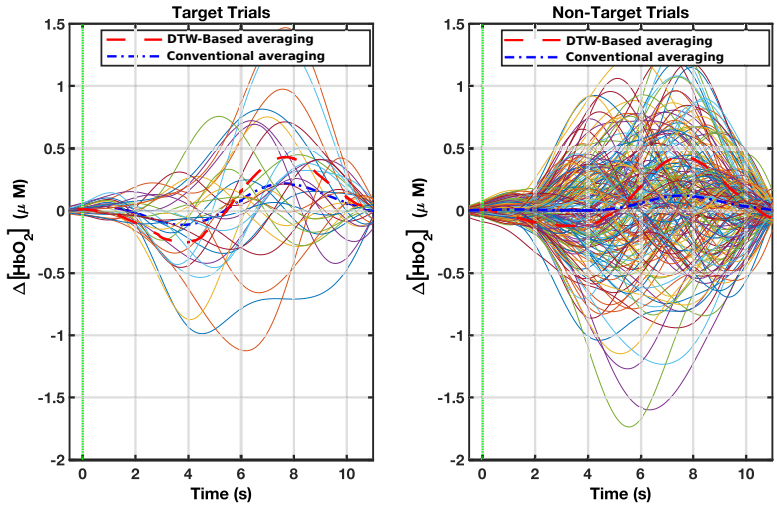


Figure 3.8: Exemplary recorded $\Delta[\mathrm{HbO}_2]$ signals from one channel (Channel 1) under "target" condition (shown in left) and "non-target" condition (shown in right) in Experiment II. Each trace represents a signal associated with a trial. The averaged signals obtained through conventional (shown in blue) and DTW-based (shown in red) averaging techniques are also shown. The vertical bar represents the timing of the onset of the experimental stimulus.

Next, for each subject and each channel, the AIs under the target and non-target conditions, denoted as AI_{ta} and AI_{nt}, respectively, were estimated using the window length of 21 samples

(2.02 s). The difference in AIs across two conditions, defined as $\mathrm{AI}_{diff} = \mathrm{AI}_{ta} - \mathrm{AI}_{nt}$, was then computed. For each channel, AI_{diff} obtained from all subjects, were pooled and tested using one-sample right-tail student's t-test, to determine whether the channel is active in response to the oddball effect, with the null-hypothesis of $mean(\mathrm{AI}_{diff}) = 0$. With a threshold of $p < 0.05$, the identified active channels, using the conventional-based averaging technique and the DTW-based averaging technique are shown in Figs. 3.9-a and 3.9-b, respectively. One can observe that compared to the patterns obtained through the conventional averaging, the active regions identified through the DTW-based averaging are located mostly in the prefrontal and occipital cortices, which appears to be consistent with previous fMRI studies [96].

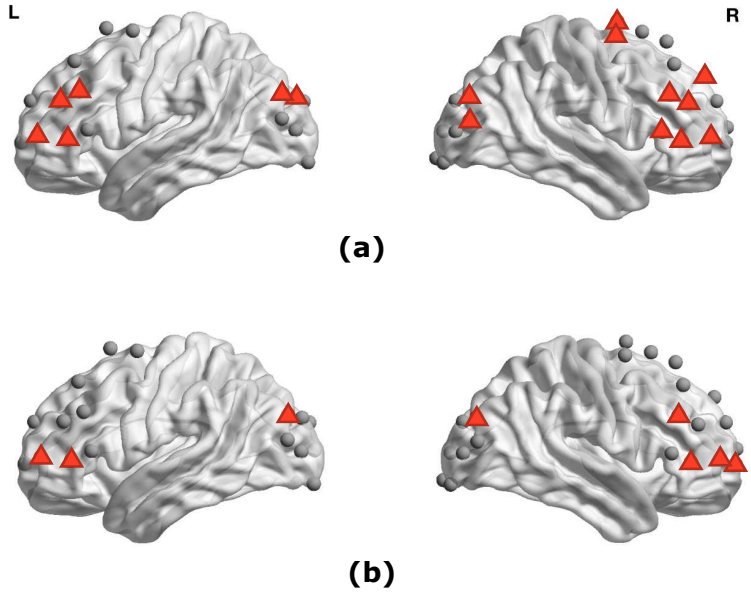


Figure 3.9: Location of channels (shown in red) that are specifically sensitive to the target condition compared to the non-target condition (sensitive to the oddball effect) in Experiment II, using the conventional-based averaging (shown in (a)) and the DTW-based averaging (shown in (b)), respectively.

3.3.2 Simulation Studies

Given that in simulations, the "ground truth" in terms of the location of brain activation is known, simulations are performed to compare the performance of the conventional averaging and the DTW-based averaging techniques. Furthermore, we also investigate the impact of choice of reference signal on the performance of DTW-based averaging procedure.

3.3.2.1 Simulation Platform

We considered a scenario similar to Experiment II where there are 50 fNIRS channels, and the task is an oddball task, with 20 target trials, and 150 non-target trials. The designed "ground truth" was that 10 channels are sensitive to the oddball effect (target>non-target).

Simulated signals were generated based on the widely-used equation that is typically used to model the hemodynamic response function (HRF) [97]

$$\text{HRF}(t) = A_1(\frac{t-d}{\tau_1})^{\delta_1} e^{(-(\delta_1/\tau_1)(t-\tau_1))} - A_2(\frac{t-d}{\tau_2})^{\delta_2} e^{(-(\delta_2/\tau_2)(t-\tau_2))}, \tag{3.8}$$

where parameters A_1, A_2, τ_1 and τ_2 determine the amplitude of the peak and the undershoot, d represents the time delay, and δ_1 and δ_2 form the general shape of the peak and the undershoot. Through fitting this model to a typical measured oxygenation response from fNIRS recording, we used $A_1 = 1$, $A_2 = 0.4$, $\delta_1 = 10$, and $\delta_2 = 20$ for signals under target and non-target conditions. The amplitude of the signals for the target condition was set to be 3% larger than that of the non-target condition.

To simulate the nonlinear variability in the latency, parameters τ_1, τ_2, and d were designated as normally distributed random variables. Their variance was set as 20, 25, and 10, respectively. Using this model, 20 target trials, and 150 non-target trials were created 40 times (corresponding to 40 participants). Additive white Gaussian noise was added to each of the simulated signals so that the SNR of the obtained signals equals 10 dB.

3.3.3 Performance Comparison

The analysis procedure for the simulated dataset followed the one we used for Experiment II. First, the signals associated with each channel, each condition, and each subject were averaged separately using the conventional and the DTW-based averaging techniques. Next, AI_{diff} values were computed for each channel and each subject. One-sample right-tail student's t-tests were conducted across subjects for each channel, with the null-hypothesis being $mean(\text{AI}_{diff}) = 0$. As the "ground truth" is known, we expected that for the 10 "true" active channels, the estimated AI_{diff} be significantly larger than zero.

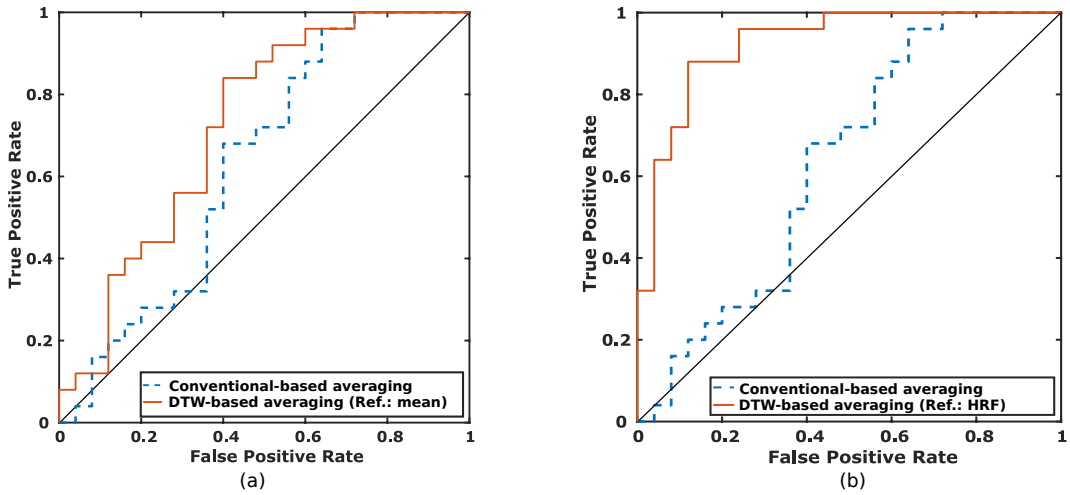


Figure 3.10: ROC curves for the Conventional and the DTW-based averaging techniques. For the right figure, a standard HRF is used as the reference signal for the DTW-based averaging technique.

We computed the receiver operating characteristic (ROC) curves for each of the averaging techniques to evaluate their performance. Here, the ROC curve illustrates the fraction of detected active channels and the associated false positive rate, when the threshold (given as t-values) varies. Fig. 3.10-a shows the resultant ROC curves for the conventional (shown in blue) and DTW-based (shown in red) averaging techniques using the simulation dataset. It is clearly shown that the DTW-based averaging technique outperforms the conventional one.

Furthermore, to examine the effects of the choice of the reference signal on the outcome of the DTW-based averaging technique, we repeated the simulation study where instead of the point-by-point averaged signal, we used the standard hemodynamic response as the reference signal. The result is shown in Fig. 3.10-b. It illustrates that the performance of the DTW-based averaging, with the standard HRF as reference, is still better than that of the conventional-based averaging. This result shows that regardless of the choice of the reference signal, the DTW-based averaging technique outperforms the conventional-based averaging technique in fNIRS-based detection studies.

3.4 Discussion

The averaging operation is performed at different stages of a wide range of fNIRS brain imaging studies (e.g. across trials, blocks, subjects), with the objective of enhancing the signal strength

associated with task-induced activities, and reducing noise and randomness. The commonly used averaging approach used for fNIRS signals is the point-by-point averaging technique. As shown in this chapter, due to the existence of variable latencies across fNIRS signals, the use of conventional point-by-point averaging technique can lead to inaccuracies in the averaged signal, and consequently, incorrect conclusions about the functionality of the brain. To address this problem, we presented a DTW-based averaging technique for fNIRS signals that takes into account that their alignment properties may vary in time. The technique optimally aligns each fNIRS signal with a reference signal such that their similarity is maximized. Once all fNIRS signals are optimally aligned, the averaged signal is computed.

To compare the performance of the DTW-based and conventional point-by-point averaging techniques for fNIRS signals, we focused on the problem of localizing task-induced active regions in the brain. Results from both block design and event-related design experiments showed that the location of task-induced active regions estimated based on the DTW-based averaged signals are better aligned with the results reported from prior fMRI studies. Furthermore, since in the presented algorithm, fNIRS signals are individually aligned with a reference signal, we also investigated the question of whether choosing a different signal as the reference would impact the performance of the algorithm. With a reference signal modeled based on a standard HRF, the DTW-based averaging technique continued to show improved performance compared to the conventional point-by-point averaging technique, in localizing active brain regions. The results also show that the performance of the DTW-based averaging can be further improved with a proper choice of reference signal. In our simulation setup, all generated fNIRS signals were created based on the HRF model (with randomized parameters and added noise), hence the choice of the standard HRF as the reference signal resulted in an improved performance, compare to the scenario where point-by-point averaged signal was chosen as the reference signal.

Improvement in the accuracy of the average of fNIRS signals is expected to introduce significant impact in various fNIRS-based neuroscience and clinical research studies. For example, in multi-modal EEG-fNIRS experiments [98, 99, 64], the location of active regions estimated from fNIRS signals, has been used as constraint for the EEG source localization problem

[64]. Errors in localizing active brain regions due to inaccuracies in the average signal can negatively influence the outcome of the EEG source localization problem, and therefore, the method presented here can be employed to avoid such impact. While for the proof of concept, in this chapter, we focused on the problem of localizing task-induced active brain regions, DTW-based averaging framework can be employed in other steps of the analysis of fNIRS signals, to avoid loss of information. For example, several studies report grand averages of the hemodynamic response across subjects [100, 101, 102, 103, 104, 105]. We suggest DTW-based averaging be used in these averaging steps, instead of the commonly used point-by-point technique, since variations in latencies for signals recorded from different individuals are inevitable. This issue is of particular importance when the study focuses on patient population, as several investigations have confirmed the existence of variable latencies in responses of patients (e.g. those with ASD) to stimuli [76, 77].

3.5 Summary

In this chapter, we presented a novel averaging framework based on DTW algorithm. Our results suggest that the conventional point-by-point averaging technique, commonly used in fNIRS brain imaging studies, can result in inaccurate conclusions about the brain function, and therefore, use of advanced averaging techniques, such as DTW-based averaging, that take into account the temporal variations in the alignment of recorded fNIRS signals, is highly recommended. Simulation results showed that the performance of the DTW-based averaging technique can be further improved based on the choice of reference signal. Future work involve investigating this issue in depth and developing quantitative statistical measures and iterative procedures to obtain an optimum choice for a reference signal, that maximizes the accuracy of the averaged signal.

Chapter 4

Data Analytics: Decoding Cortical Brain States from Widefield Calcium Imaging Data Using Visibility Graph

In this chapter, we focus on the problem of inferring behavior from widefield calcium imaging data. We introduce a method based on visibility graph to reliably identify temporal characteristics of brain recordings which carry discriminative information across behaviors. The presented method is validated using widefield calcium imaging under spontaneous whisking condition.

4.1 Introduction

It had been shown that the brain states is highly related to the functions of sensory perception, sensorimotor coordination and learning [106]. Decoding the cortical brain states and inferring about the behavior, intent, or the engagement of a particular cognitive process, from neuroimaging data, finds applications in several domains including brain machine interfaces (BMIs) [107, 108, 109]. Depending on the type of physiological activity that is monitored, various computational techniques have been suggested to infer or decode the intent or the cognitive state of the subject from recorded brain activities. Methods based on functional specificity [110, 111], brain connectivity patterns [112, 113, 114], and power spectral density [115], to name a few, have been suggested. However, the estimation power of such methods has been limited to distinguishing very distinct classes of motor activities or cognitive processes [116]. As such the community has been searching for alternative methods to improve the power of inference.

Given the time-varying nature of the brain function, in this work, we focus on the time domain information. We hypothesize that there exist “characteristics" in the time course of cortical activities that are specific to the corresponding behavior. The key challenge is to develop methods that can reliably identify such discriminatory characteristics in cortical recordings. To

test the hypothesis, we use transgenic calcium reporter mice expressing GCaMP6f specifically in neocortical pyramidal neurons to image neural activity in nearly the entire left hemisphere and medial portions of the right hemisphere in head-fixed mice, including sensory and motor areas of the neocortex. For behavior, we focus on active whisking (AW) and no whisking (NW). In many rodent cortical neurophysiological studies, whisking behavior has been associated with brain state changes on different levels of cell and systems function [55]. Quiet wakefulness, in the absence of locomotion or whisking, is associated with low frequency synchronized cortical activity, while locomotion and whisking are associated with higher frequency desynchronized activity in primary sensory areas of the cortex [53], [106], [117], [118], [119]. Recent studies indicate that active, arousal-related behaviors such as locomotion and whisking are associated with widespread modulation of cortical activation [40], [57]. Therefore, AW and NW behavioral states could be a good proxy for distinct brain states.

To identify features in calcium imaging data that would be unique to behavior (here AW or NW), we propose to use visibility graph (VG) [120]. As will be discussed, VG provides a means to "quantify" various properties of a given time series, enabling a path to extract temporal-based features that are unique to the characteristics of the time series. We construct the VG representation of the recordings for each region of interest (ROI), extract the graph measures, and build features based on the graph measures for all ROIs. We conduct an extensive study to identify the best model capable of inferring AW and NW for each subject, from cortical recordings. Fig. 4.1 provides a summary of the procedure.

The novelty of our work is the introduction of the visibility graph for extracting features that are related to the temporal characteristics of recorded calcium time series. It is shown that the temporal features of calcium recordings extracted through VG, carry discriminatory information for inferring the corresponding behavior. While in this study, we consider cortical signals from the entire left hemisphere and medial part of the right hemisphere, and focus on whisking condition, given the data-driven nature of the proposed approach, we expect that it would be also applicable to recorded activity from other areas of the brain, such as the thalamus and deep layers of motor cortex, for inferring other forms of behavior or cognitive states.

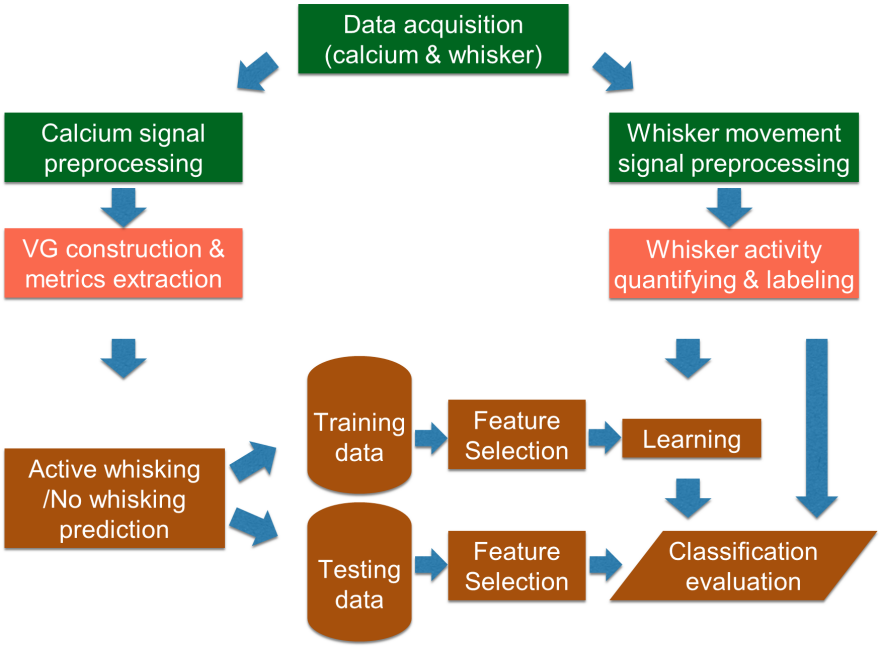


Figure 4.1: Summary of the proposed analysis procedure.

4.2 Materials and Methods

Before discussing details of data collection and the analysis procedure, we provide clarification about some terminologies used throughout the paper. Note that in this study we use the term "decode" and "infer" interchangeably.

The *imaged area* here refers to the optically accessible cortical area. The imaged area in this study covers the entire left hemisphere, and medial part of the right hemisphere of the cortex.

Behavior in this study is related to whisking condition. Two classes of behavior, *active whisking (AW)* and *no whisking (NW)*, are considered here. We use the term "brain state" and "behavior" interchangeably.

Features are measures extracted from cortical recordings. To examine how well the proposed features from recorded calcium transients can discriminate the two classes of AW and NW, classification experiments are performed. In these experiments, a *classifer* refers to the algorithm that is used to perform classification.

A *predictive model* refers to a trained classifier. The ability of the model to correctly infer (or predict) the whisking condition (AW or NW) from features extracted from cortical recordings, is tested using k-fold cross validation.

We now discuss the widefield imaging experiments, and the methods used in the analysis.

4.2.1 Animals and Surgery

Six mice expressing GCaMP6f in cortical excitatory neurons were used for widefield transcranial imaging [40, 13]. All procedures were carried out with the approval of the Rutgers University Institutional Animal Care and Use Committee. Triple transgenic mice expressed Cre recombinase in Emx1-positive excitatory pyramidal neurons (The Jackson Laboratory; 005628), tTA under the control of the Camk2a promoter (The Jackson Laboratory; 007004) or ZtTA ($3/6$ mice) under the control of the CAG promoter into the ROSA26 locus (The Jackson Laboratory; 012266) and TITL-GCaMP6f (The Jackson Laborotory; Ai93; 024103). At 7 to 11 weeks of age, mice were outfitted with a transparent skull and an attached fixation post using methods similar to those described previously [51, 40, 121]. Mice were anesthetized with isoflurane (3% induction and 1.5% maintenance) in 100% oxygen, and placed in a stereotaxic frame (Stoelting) with temperature maintained at 36 °C with a thermostatically controlled heating blanket (FHC). The scalp was sterilized with betadine scrub and infiltrated with bupivacaine (0.25%) prior to incision. The skull was lightly scraped to detach muscle and periosteum and irrigated with sterile 0.9% saline. The skull was made transparent using a light-curable bonding agent (iBond Total Etch, Heraeus Kulzer International) followed by a transparent dental composite (Tetric Evoflow, Ivoclar Vivadent). A custom aluminum headpost was affixed to the right side of the skull and the transparent window was surrounded by a raised border constructed using another dental composite (Charisma, Heraeus Kulzer International). Carprofen (5 mg/kg) was administered postoperatively. Following a recovery period of one to two weeks, mice were acclimated to handling and head fixation for an additional week prior to imaging. Mice were housed on a reversed light cycle and all handling and imaging took place during the dark phase of the cycle.

4.2.2 Widefield Calcium Imaging and Whisker Movement Recording

Imaging of GCaMP6f was carried out in head-fixed mice with the transparent skull covered with glycerol and a glass coverslip. A schematic of the imaging system is shown in Fig. 4.2. A

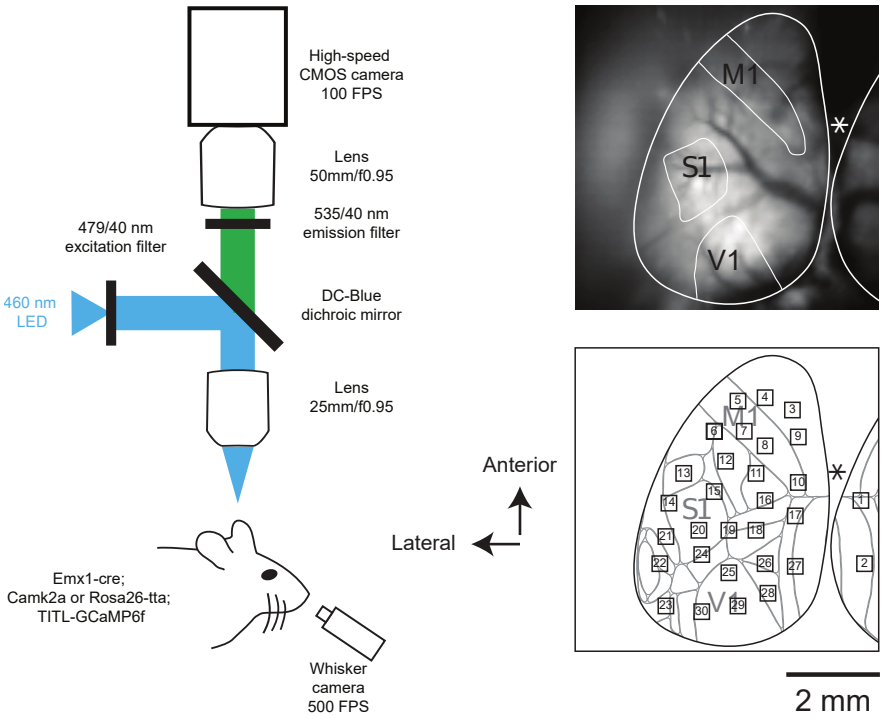


Figure 4.2: Left: Illustration of the experimental setup used for widefield imaging of cortical activity of mice expressing GCaMP6f and simultaneous recording of whisker movement. Right, top: raw image of neocortical surface through transparent skull preparation. M1, S1, and V1 are schematically labeled. Asterisk indicates position of Bregma. Right, bottom: ROIs are superimposed on a map based on the Allen Institute common coordinate framework v3 of mouse cortex (brain-map.org; adapted from [8]). ROI: 1, Retrosplenial area, lateral agranular part (RSPagl); 2, Retrosplenial area, dorsal (RSPd); 3, 4, 9, Secondary motor area (MOs); 5, 7, 8, 10, Primary motor area (MOp); 6, Primary somatosensory area, mouth (SSp-m) / upper limb (SSp-ul); 11, 16, Primary somatosensory area, lower limb (SSp-ll); 12, SS-ul; 13, Primary somatosensory area, nose (SSp-n); 14, 20, Primary somatosensory area, barrel field (SSp-bfd); 15, SSp-bfd / Primary somatosensory area, unassigned (SSp-un); 17, Retrosplenial area, lateral agranular part (RSPagl); 18, Anterior visual area (VISa) / Primary somatosensory area, trunk (SSp-tr); 19, VISa / SSp-tr / SSp-bfd; 21, Supplementary somatosensory area (SSs); 22, Auditory area (AUD); 23, Temporal association areas (TEa); 24, SSp-bfd / Rostrolateral visual area (VISrl); 25, 29, 30, Primary visual area (VISp); 26, Anteromedial visual area (VISam); 27, RSPagl / RSPd; 28, Posteromedial visual area (VISpm).

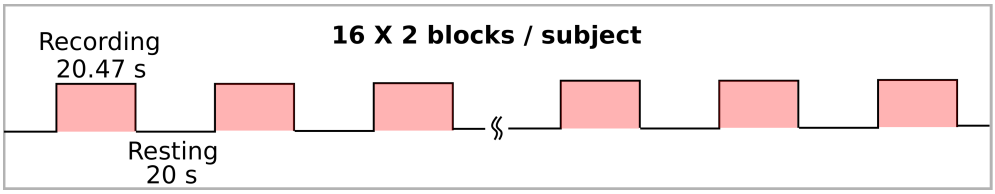


Figure 4.3: Experimental protocol that was followed for each subject. Each subject participated in two sessions per day. In each session, spontaneous activity was acquired for sixteen 20.47 *s* blocks, with 20 *s* of rest between blocks.

custom macroscope [50] allowed for simultaneous visualization of nearly the entire left hemisphere and medial portions of the right hemisphere (as seen in Fig. 4.2. The cortex was illuminated with 460 nm LED (Aculed VHL) powered by a Prizmatix current controller (BLCC-2). Excitation light was filtered (479/40; Semrock FF01-479/40-25) and reflected by a dichroic mirror (Linos DC-Blue G38 1323 036) through the objective lens (Navitar 25 mm / f0.95 lens, inverted). GCaMP6f fluorescence was filtered (535/40; Chroma D535/40m emission filter) and acquired using a MiCam Ultima CMOS camera (Brain vision) fitted with a 50 mm / f0.95 lens (Navitar). Images were captured on a 100×100 pixel sensor. Spontaneous cortical activity was acquired in 20.47 *s* blocks at 100 frames per second with 20 *s* between blocks (Fig. 4.3). Sixteen blocks were acquired in each session and mice were imaged in two sessions in a day.

In addition, all whiskers contralateral to the imaged cortical hemisphere were monitored with high-speed video at 500 frames/s using a Photonfocus DR1 camera triggered by a Master-9 pulse generator (AMPI) and Streampix (Norpix) software. Whiskers were illuminated from below with 850 nm infrared light. The mean whisker position was tracked and measured as the changes in angle (in degree) using a well-established, automated whisker-tracking algorithm, freely available in MATLAB [122], that computes the frame-by-frame center of mass of all whiskers in the camera's field of view. The angle of the center of mass of all whiskers is similar to the average angle of all whiskers tracked individually, because the whiskers do not move independently.

4.2.2.1 Preprocessing of Calcium Signals

Changes in GCaMP6f relative fluorescence ($\Delta F/F_0$) for each frame within a recording were calculated by subtracting and then dividing by the baseline. The baseline was defined as the average intensity of the first 49 frames. Two blocks (one from subject #2 and one from subject #3) were excluded from further analysis due to loss of whisker movement data. The length of blocks were shortened to 20 s from 20.47 s for the remaining parts of analysis.

Thirty 5×5 pixel regions of interest (ROIs) distributed over the cortex (see Fig. 4.2) in each frame were defined based on location relative to the bregma point on the skull. In $5/6$ mice, whisker stimulation by piezo bending element was used to map the location of S1 barrel cortex. The 30 ROIs were positioned to cover and fill space between areas including somatosensory, visual and motor areas of the cortex (S1, V1, M1) (see Fig. 4.2). Each pixel is 65 μm side length, and 5×5 pixel ROI is $325x325$ μm. This size ROI is the approximate dimension of a cortical column in sensory cortex, and is consistent with the standard practices in the field [41, 123]. These studies, which examined sensory mapping, spontaneous activity, and task-related activation, have shown that widefield calcium signals do not display signals with resolution better than these dimensions, and therefore, smaller ROIs are not beneficial. The choice of ROI size is therefore, suitable and standard for comparison across different existing datasets. ROI locations were kept the same across subjects. Time series associated with each ROI were obtained by finding the average of pixel intensities within the corresponding ROI.

4.2.2.2 Labeling Data Related to Active Whisking and No Whisking Conditions

In order to investigate the relationship between behavior and the cortical activity, it is necessary to identify the duration in the recordings that are related to "active whisking" (AW) and "no whisking" (NW) conditions. Here we developed a method to automatically label the duration related to each condition, according to the whisker movement recordings.

The whisker movement time series was segmented using a sliding window. For a given segment i, the standard deviation (SD) of the signal (σ_{w_i}) is computed as $\sigma_{w_i} = \sum_{j=1}^{N}(x_j - \mu_i)^2$, where μ_i represents the mean and N denotes the number of samples within the segment. This procedure generates a new time series of σ_{w_i}s, representing the extent to which the whisker

is in motion over the course of observation. A threshold was then set to identify whether the recordings correspond to active whisking (above the threshold) or no whisking (below threshold) conditions. After testing different threshold values and visually inspecting the raw whisker movement signals, a threshold value of 10 was used.

As an example, sample images and time series corresponding to two ROIs (6 and 27) along with whisking movement signal, recorded in block #1 from subject #1, are shown in Fig. 4.4. The top row illustrates a series of baseline-corrected images. Shown also are the averaged image for the duration of $(6.01 - 6.20)$ s (labeled in red in Fig. 4.4(c)), where no clear calcium transients are present, and the averaged image for the duration of $(13.21 - 13.40)$ s (labeled in blue in Fig. 4.4(c)), where calcium transients are present.

The measured angle corresponding to whisker movement recordings of the same block is shown in Fig. 4.4(d), and in Fig. 4.4(e) the time series obtained based on the standard deviation calculation of sliding window approach discussed in Section 4.2.2.2 is plotted. The threshold level for determining AW and NW conditions over time, is visualized by a red horizontal line.

4.2.3 Visibility Graph

Here, we first describe the procedure used to construct the visibility graph for a given time series and extracting graph measures.

4.2.3.1 VG Construction

Visibility graph is an effective tool that can be employed to reveal the temporal structure of the time series at different time scales [120, 124, 125, 126]. Recently, VG has been receiving increased attention in various studies related to human brain function such as those involving sleep [127], epilepsy [128, 129], Alzheimer's disease [130], and differentiating resting-state and task-execution states [14]. In these studies, VG has been applied to time series obtained from various imaging modalities such as electroencephalography (EEG) [127, 128, 129, 130, 131], functional near-infrared spectroscopy (fNIRS) [14], and functional magnetic resonance imaging (fMRI) [132].

VG maps a time series to a graph, thereby, providing a tool to "visually" investigate different properties of the time series [120, 126]. The VG associated with a given time series $\mathbf{x} =$

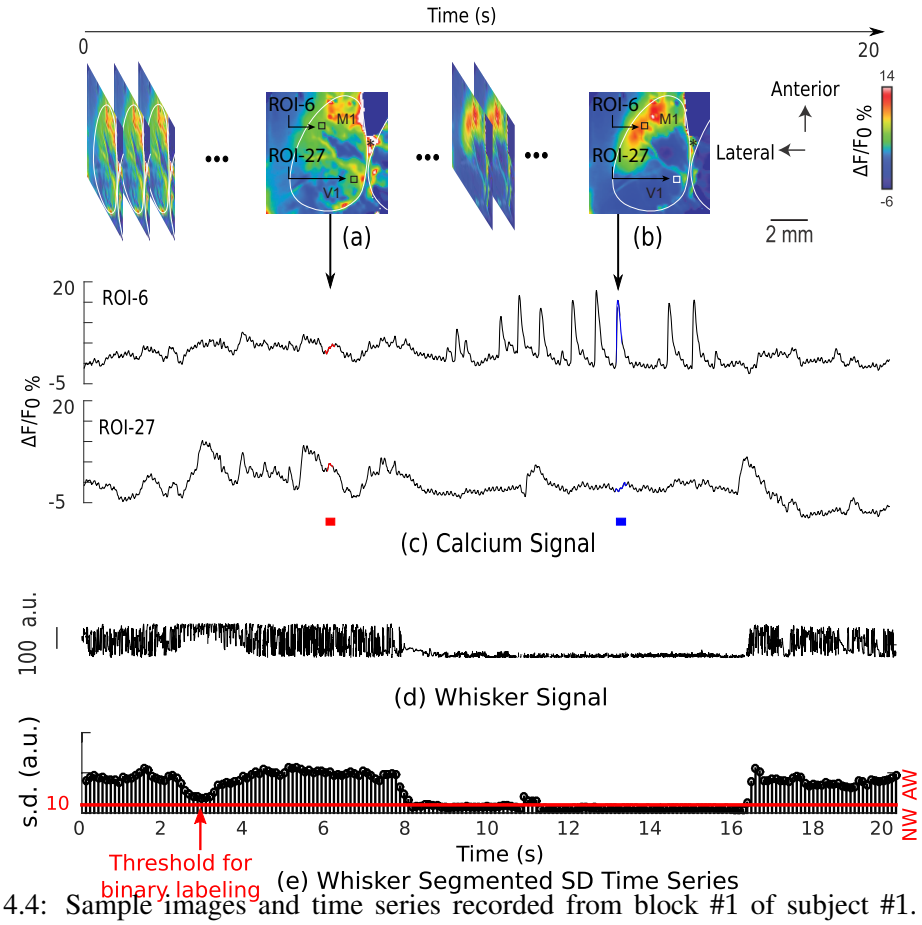


Figure 4.4: Sample images and time series recorded from block #1 of subject #1. (a)-(b) baseline-corrected images, (c) time series corresponding to ROI-6 and ROI-27, (d) measured angle corresponding to whisker movement signal recorded from the same block, and (e) standard deviation-based time series of the signal, (d) where the threshold level used for labeling AW and NW conditions is shown as a red line.

$[x(1), \cdots, x(N)]$ of N points is constructed as follows. Each point in $\mathbf{x}$ is considered as a node in the graph (i.e. for an N-point time series, the graph will have N nodes). The link between node pairs is formed only if the nodes are considered to be *naturally visible*. That is, in the graph, there will be an undirected and unweighted link between nodes i and j, if and only if, for any point p ($i < p < j$) in the time series, the following condition holds

$$x(p) < x(j) + [x(i) - x(j)][\frac{t(j) - t(p)}{t(j) - t(i)}], \tag{4.1}$$

where $t(j)$, $t(p)$ and $t(i)$ are the time corresponding to points j, p, and i [120]. That is, two

nodes i and j are connected, if the straight line connecting two data points $(t(i), x(i))$ and $(t(j), x(j))$, does not intersect the height of any data point $(t(p), x(p))$ that exists between them. Accordingly, in the adjacency matrix $A_{\mathbf{x}} = \{a_{i,j}\}$ $(i, j = 1, \cdots, N)$, the element $a_{i,j}$ will be set to 1 if the nodes i and j are connected given the definition above, and 0 if otherwise.

4.2.3.2 Graph Measures as Features

Once the time series $\mathbf{x}$ of N points is mapped to a graph with adjacency matrix $A_{\mathbf{x}} = \{a_{i,j}\}$ $(i, j = 1, \cdots, N)$ via VG, the topological measures of the graph can be utilized to investigate different properties of the time series. Here, we consider three of such measures: Edge Density (D), Averaged Clustering Coefficient (C), and Characteristic Pathlength (L), as defined below.

- *Edge Density* (D) measures the fraction of existing edges in the graph with respect to the maximum possible number of edges [133]. The edge density is obtained as

$$D = \frac{1}{N(N-1)} \sum_{i,j} a_{i,j}. \tag{4.2}$$

  It can be shown that for a globally convex time series, the value of D would be 1, and for a time series with large number of fluctuations, the value of D would be small. Therefore, the edge density can be considered as a measure of irregularity of fluctuations in the time series [134].

- *Averaged Clustering Coefficient* (C) is obtained as the average of local clustering coefficients of all nodes in the graph. The local clustering coefficient of the node i (C_i) is defined as the fraction of its connected neighboring nodes to the maximum number of possible connections among the neighboring nodes [133]. The averaged clustering coefficient is computed as

$$C = \frac{1}{N} \sum_{i=1}^{N} C_i = \frac{1}{N} \sum_{i,j,l} \frac{a_{ij} a_{il} a_{jl}}{K_i (K_i - 1)}, \tag{4.3}$$

  where K_i represents the degree of node i (the number of edges connected to node i). A large value of C indicates dominant convexity of the time series [134].

- *Characteristic Pathlength* (L) is found as the average of the shortest pathlength between all node pairs in the graph. The characteristic pathlength is obtained as

$$L = \frac{1}{N(N-1)} \sum_{i,j} l_{ij}, \tag{4.4}$$

where l_{ij} denotes the shortest pathlengh between nodes i and j.

4.2.4 Classification

To learn models of inferring behavior (as measured by AW and NW) from recordings obtained via widefield calcium imaging of cortical activity, classification experiments are performed. Specifically, we wish to learn classifiers in the following form:

$$f: \text{ VG Measures } (t_0,\ t_0 + w) \rightarrow \{AW,\ NW\}, \tag{4.5}$$

where VG Measures $(t_0,\ t_0 + w)$ represents graph measures that are extracted from VGs associated with calcium signals within the segment $[t_0, t_0 + w]$, and w denotes the window length used for segmentation 4.3.1.

Here, we briefly describe the feature extraction process, the classifiers, and the measures used to evaluate the classification performance. Classification experiments were executed using GraphLab [135].

4.2.4.1 Feature Extraction

Three graph measures were extracted from the VG associated with each segment (identified by the sliding window) of recordings obtained from individual ROIs. To extensively investigate which measures will result in a better model, seven types of feature vectors were formed. These were D, C, L, $D + C$, $D + L$, $C + L$, and $D + C + L$. In all cases, feature vectors were constructed using measures from all the ROIs. For example, when considering D as features, for each segment, a feature vector of 30×1 is constructed (where 30 represents the number of ROIs).

Table 4.1: Number of blocks and number of AW/NW segments for each subject, when the window length of 2 s with a step size of 0.5 s is used.

Subject ID	**1**	**2**	**3**	**4**	**5**	**6**	**Total**	**% Total Segments**
# Blocks	32	31	31	32	32	32	190	
# AW Segments	238	360	240	416	227	153	1634	23.24
# NW Segments	946	787	907	768	957	1031	5396	76.76

Five different sliding window duration (1, 1.5, 2, 2.5, and 3 s) were considered for segmentation. As such, the number of segments per recording block varies based on the sliding window duration (39 for 1 s window, 38 for 1.5 s window, 37 for 2 s window, 36 for 2.5 s window, and 35 for 3 s window). There are 32 blocks for subject #1, 4, 5, 6, and 31 blocks for subject #2 and 3. Table 4.1 summarizes the number of blocks, and the number of AW/NW segments for each subject, when the window duration of 2 s, and window step of 0.5 s are used.

4.2.4.2 Classifiers and Evaluation Measures

Three commonly-used classifiers were used to perform classification: 1) k-nearest neighbor (kNN), 2) regularized logistic regression (LR), and 3) random forest (RF). These classifiers have been shown to offer good performance with neuroimaging data in several studies [136, 137, 138, 139, 140, 141, 142, 143, 144]. Here, for kNN, k in the range of 1 and 10 is used, for LR, ℓ_2-norm regularization is used, and the weight of the regularization was set between 10^{-2} and $10^{1.5}$, and for RF, the subsampling ratio is selected to be 40%, 70% or 100%.

To evaluate the classification performance, three measures, accuracy (AC), sensitivity (SE), and specificity (SP), were used [145].

First, separate classifiers were trained for each subject. A ten-fold cross-validation was used to test the performance of the models. For each subject, the data were randomly partitioned into ten subsamples. Classification experiments were repeated ten times, where during each, one subsample was assigned as the testing dataset, and the remaining subsamples were assigned as training dataset. For every subject, the classification performance was evaluated using the measures described above, and then results were averaged across the ten repetitions.

4.3 Results

4.3.1 VG Construction from Calcium Signals

The preprocessed calcium signals were segmented using sliding windows with the fixed step of 50 time points (0.5 s). Five different window lengths were used: 100, 150, 200, 250, and 300 time points (corresponding to 1, 1.5, 2, 2.5, and 3 s, respectively). The VG was constructed for each segment of the time series obtained from each ROI. For each VG, three graph measures, D, C, and L were extracted. As a result, for a given sliding window length, recordings from each ROI of each recording block, result in three time series for D, C and L. Our objective is to use these information and develop models to predict the behavior of active whisking and no whisking, from recorded calcium signals.

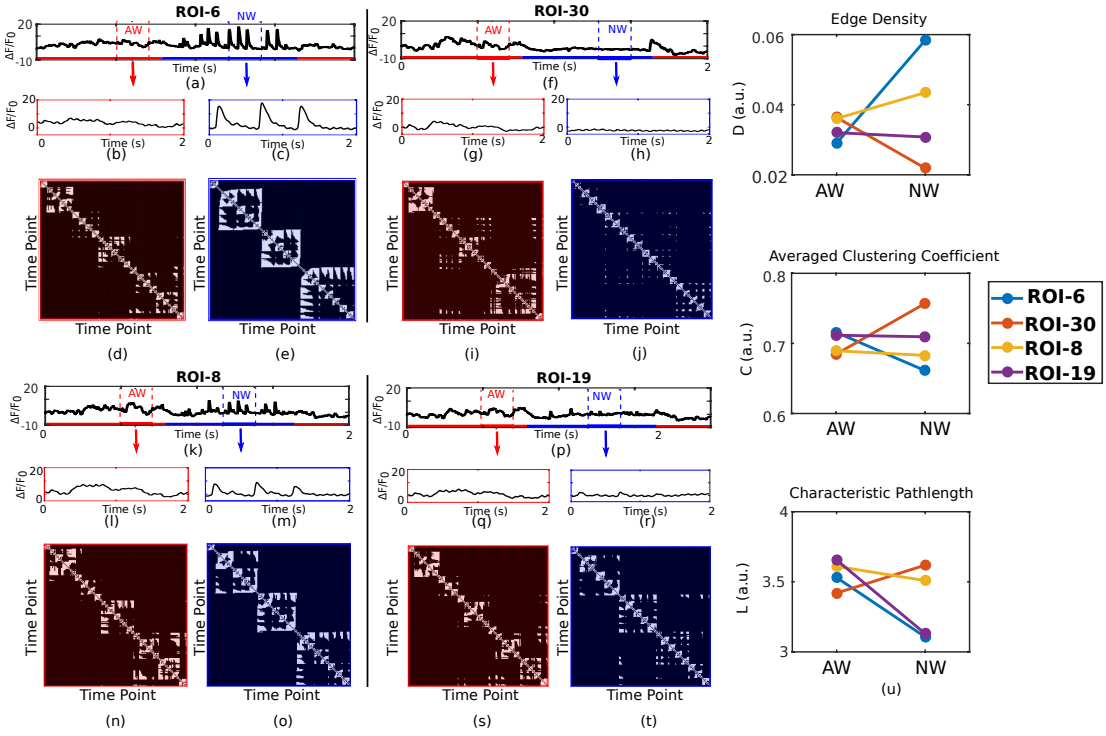


Figure 4.5: Preprocessed calcium signals of recording block #1 from subject #1 from ROI-6 (a), ROI-8 (f), ROI-30 (k) and ROI-19 (p). For each case, 2 s segments of signals corresponding to AW (shown in red in (b), (g), (l) and (q)) and NW (shown in blue in (c), (h), (m) and (r)) conditions as determined from whisker movement recordings. For each ROI, the adjacency matrices for 2 s AW are shown in (d), (i), (n), and (s), and for 2 s NW are shown in (e), (j), (o), and (t). Measures extracted from VG of 2 s duration of AW time series (shown in red) and from VG of 2 s NW time series (shown in blue) are also shown in (u) for each ROI.

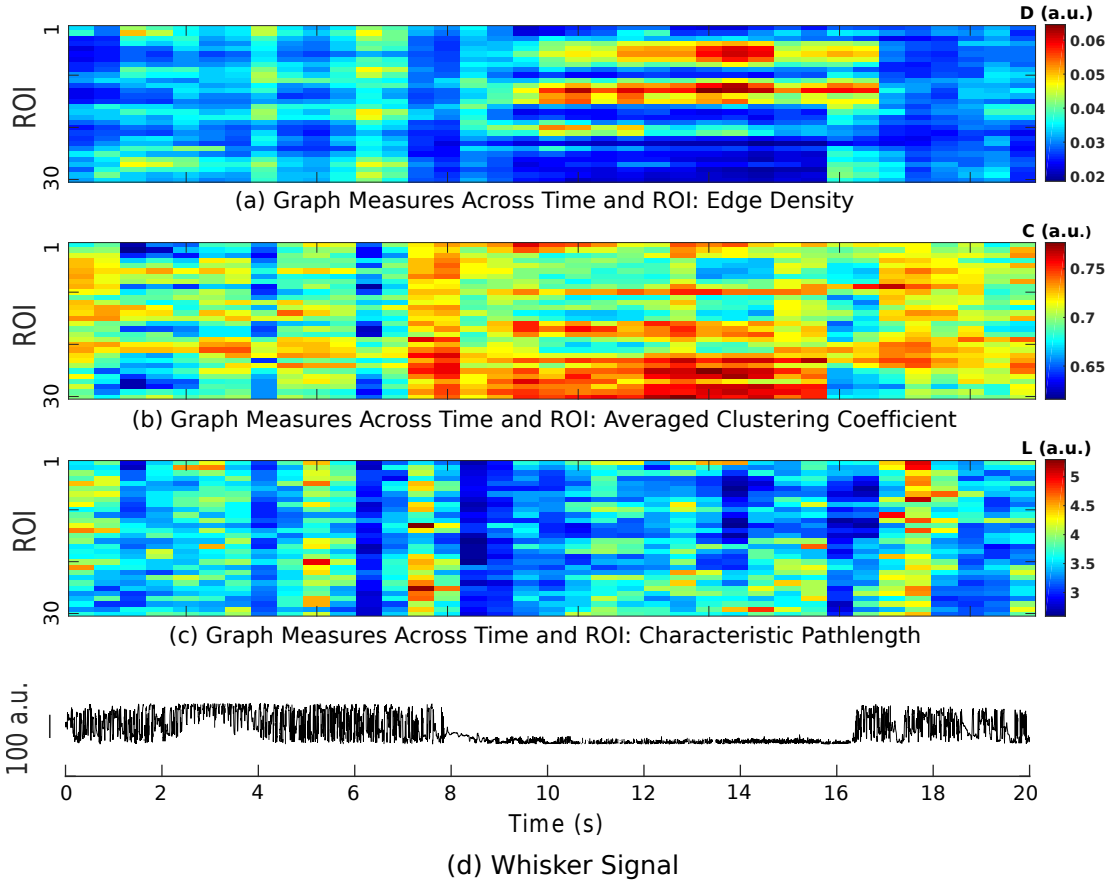


Figure 4.6: Color-coded graph measures for all ROIs as a function of time during a recording block. (a) Edge density (D), (b) Averaged clustering coefficient (C), and (c) Characteristic pathlength (L). (d) Whisker movement recording obtained simultaneously in the same block.

Representative preprocessed calcium signals from four ROIs (6, 8, 19 and 30) of the recording block #1 from subject #1 are shown in Fig. 4.5. For signals from each ROI, two segments of 2 s duration, corresponding to AW and NW, are also shown. For each of these segments, the VG is constructed and their corresponding adjacency matrices are presented. As segments have the same duration (2 s or 200 time points), the number of nodes in all graphs will be the same. In these matrices, the dark color represents no connection, and the light color represents the existence of an edge. For each ROI, the distinctions between the patterns of the matrices related to AW and NW can be revealed via the three graph measures D, C and L. The values for these measures are compared for AW and NW and each ROI in Fig. 4.5(u).

As can be seen, distinct patterns (e.g. in terms of amplitude and width of calcium transients), for the same whisking condition (AW or NW) are observed in signals obtained from

different ROIs distributed over the cortex, suggesting that different cortical regions have potentially different relationships with behavior. For example, for ROIs in or close to M1 (ROI-6 and ROI-8) the measure D is larger during NW compared to AW, suggesting that there are more number of edges in the VG representation of recordings from this region for NW as compared to AW. For ROIs close to S1 (ROI-19) the measure L appears to be smaller during NW compared to AW, suggesting that there are less connections in the VG representation of recordings from this region for NW as compared to AW. In V1 (e.g. ROI-30), the measure C is larger during NW as compared to AW, indicating the presence of smaller clusters in the VG representation of recordings from this region during AW as compared to NW. These results suggest that different regions of the brain follow different temporal dynamics during behavior, and such differences can be revealed and quantitatively described via VG measures D, C and L.

The graph measures shown in Fig. 4.5(u) correspond to two segments of the time series for each of the four ROIs. Using the sliding window of length 2 s, VGs can be constructed for each segment of the time series, and from each VG, the three mentioned graph measures can be extracted. Figs. 4.6(a) to (c) show the results of such analysis for all ROIs, illustrating the temporal evolution of D, C and L, respectively. The simultaneously obtained whisker movement recording is also shown in Fig. 4.6(d). It can be clearly seen that different patterns are observed for VG measures for duration corresponding to AW and NW across all ROIs.

4.3.2 Classification Results

For each subject, we performed comprehensive investigation on how the selection of various parameters (e.g. various window sizes for extracting VG measures, and performing classification based on different selection of feature types) will impact the classification results. For each choice of window size, features were constructed based on individual or a combination of measures from the corresponding VG. Figs. 4.7, 4.8 and 4.9 illustrate the evaluation measures obtained for each subject when kNN, LR, and RF were used as the classifier, respectively.

It was found out that while the performance is subject dependent (due to individual variability as well as variability in whisking behavior across subjects (see Table 1)), with a proper choice for features and window length, all classifiers result in high levels of accuracy and specificity for all subjects. The sensitivity remains to be relatively modest, however, considering the

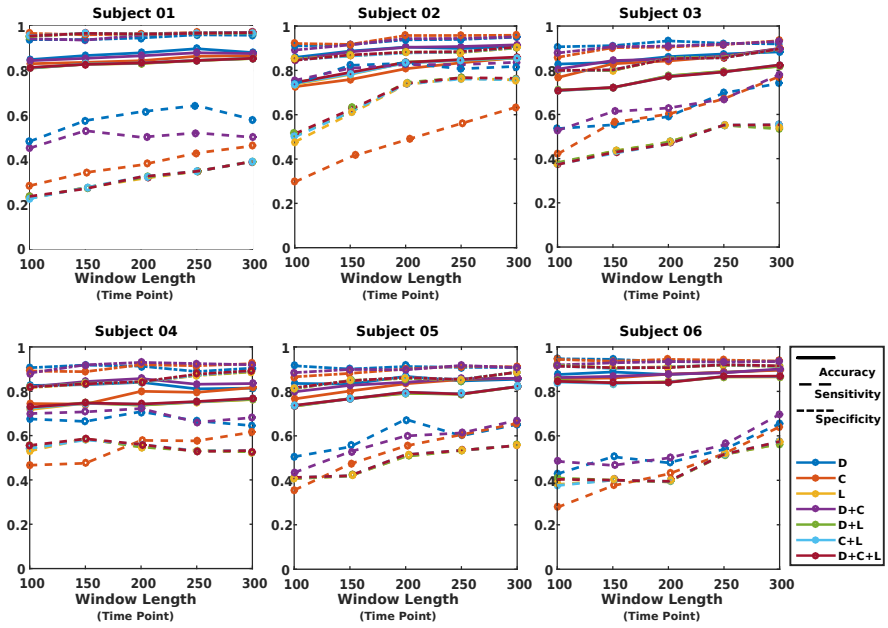


Figure 4.7: Classification results when using kNN as classifier.

imbalanced dataset between AW and NW (e.g. only 23% of the samples belonged to the AW condition for 2 s window duration), the obtained significantly better accuracy than naive classifier (in which all the testing samples are assigned the label associated to the majority class in the training set), demonstrating the effectiveness of the VG measures in providing features that carry discriminatory information for AW and NW. In majority of scenarios, classification based on either C or L did not result in good performance, while classification based on feature $D + C$ or D led to the best sensitivity results for majority of the subjects.

For each classifier, information about the choice of window length (w), features, and parameters that have resulted in the best sensitivity among all the explored options, are summarized in Table 4.2. Consistent with the observation made from Figs. 4.7, 4.8, and 4.9, it can be seen that, in all cases, the graph measure D, either individually or jointly with others, has been identified as the optimum feature. For classifiers kNN and LR, the feature $D + C$ across most subjects has resulted in the best sensitivity results, while for the RF classifier, the measure D by itself has worked as the optimum feature. In terms of duration of segments for constructing VGs, window duration of equal or larger than 2 s has resulted in the optimum performance. In addition, for most cases, the sensitivity measure dropped as the window size for extracting features goes below 150 points.

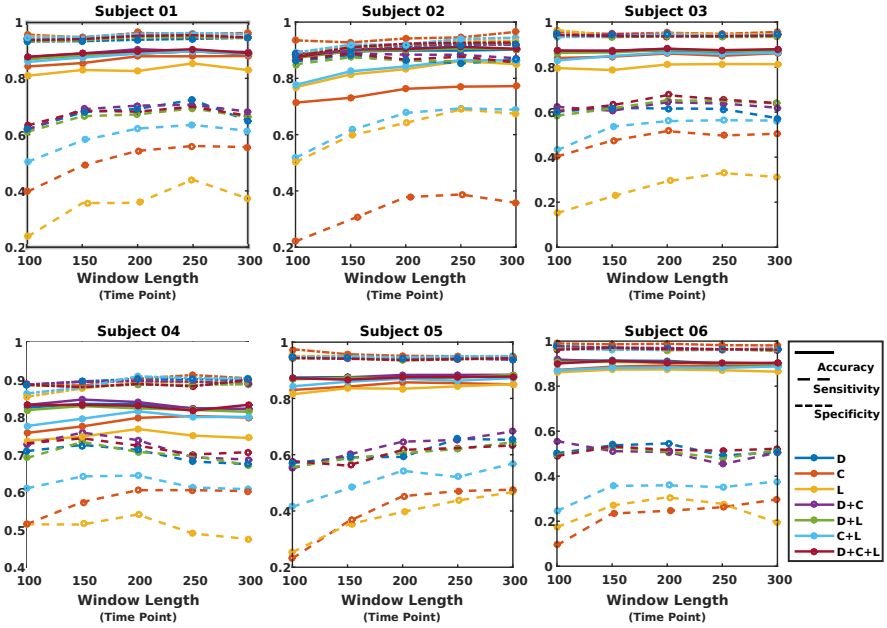


Figure 4.8: Classification results when using regularized logistic regression (LR) as classifier.

Overall, *k*NN and LR deliver almost always slightly better performance than RF, but using the right features, all classifiers are able to successfully differentiate the whisking conditions, demonstrating that features based on visibility graph carry discriminatory information. To summarize the performance of the classifiers, we repeated the classification across subjects by using unified parameters that led to the best classification performance in majority of the subjects in Table 4.2. We used $D + C$ as the feature, and 2 s as the window length. The results are presented in Table 4.3, where as can be seen, on average, an accuracy larger than 86% is achieved across all subjects.

4.4 Discussion

Measuring brain states over wide areas of cortex is of central importance for understanding sensory processing and sensorimotor integration. Changes in brain states influence the processing of incoming sensory information. For example, data from several sensory modalities including somatosensation, vision, and audition, indicate that the cortical representations of stimuli vary depending on the neocortical state when the stimulus arrives [146, 147, 148, 149]. In mice, natural spontaneous behaviors such as locomotion and self-generated whisker movements influence brain states through increased behavioral arousal and activation of ascending

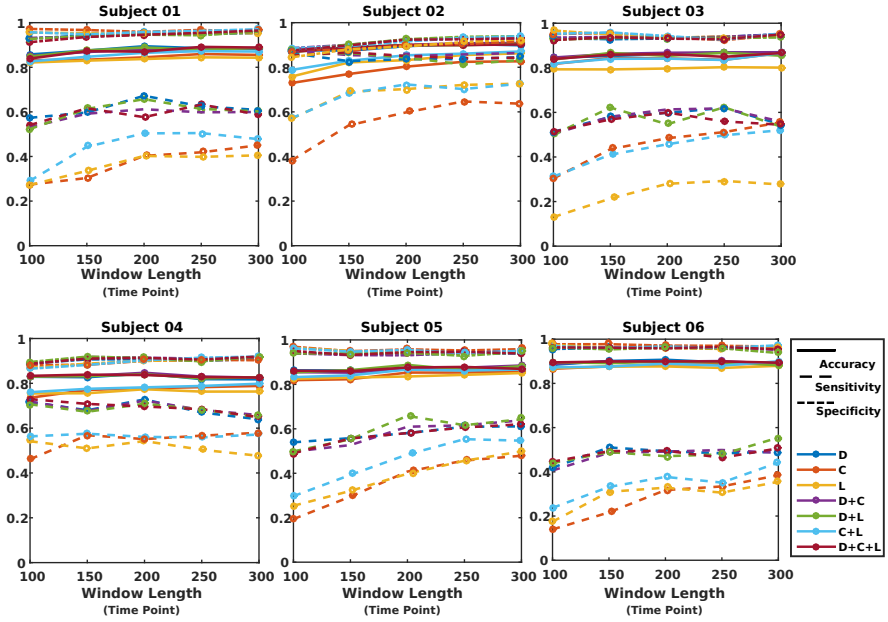


Figure 4.9: Classification results when using random forest (RF) as classifier.

neuromodulator systems [118, 150]. Studies in mice using widefield imaging of voltage and calcium sensors during whisking or locomotion have provided important information on the spatiotemporal modulations of brain states [40, 57], and relating these dynamic optical signals to behavior is an area of great interest. This line of research will be advanced by the development of several new transgenic calcium reporter mice [151, 152] and cranial window methods [42].

Studies from several sensory modalities including somatosensation, vision, and audition have reported changes in the cortical representation of stimuli that vary depending on the neocortical state when the stimulus arrives.

The VGs constructed here corresponded to segments of recordings as identified by the moving window of length w. We performed a comprehensive study (five different window lengths, seven types of features per choice of window length, and three classifiers) to find the model that can be used to infer the behavior (AW or NW) from calcium imaging data. All classifiers delivered high accuracy and specificity and moderate sensitivity, with kNN and LR offering better performances than RF. Considering the imbalanced dataset between AW and NW (e.g. only 23% of the samples belonged to the AW condition for 2 s window length), the obtained significantly better-than-naive-classifier demonstrates the effectiveness of the VG measures in

Table 4.2: Classification results for best sensitivity obtained for each subject when using kNN, regularized logistic regression (LR), and random forest (RF) as classifier. Features, window lengths (w), and related parameters from which the optimum results have been obtained are also listed (SS is short for subsample). Note that "$+$" in the "Feature" rows represent using multiparametric approach for performing the classification.

Classifier	Performance Measure	1	2	3	4	5	6
k**NN**	**AC (%)**	89.84	91.52	89.78	85.72	86.74	90.09
	SE (%)	64.33	83.64	78.01	72.18	67.42	69.66
	SP (%)	96.08	95.15	93.03	93.01	91.20	93.47
	Feature	D	$D+C$	$D+C$	$D+C$	D	$D+C$
	w **(points)**	250	300	300	200	200	300
	k	7	3	1	5	1	1
LR	**AC (%)**	89.66	88.75	88.14	84.70	88.39	91.83
	SE (%)	72.36	89.54	67.53	75.89	68.30	55.58
	SP (%)	94.02	88.37	93.64	89.43	93.89	97.49
	Feature	D	$D+C$	$D+C+L$	$D+C$	$D+C$	$D+C$
	w **(points)**	250	100	200	150	300	100
	ℓ_2	0.01	0.5623	0.01	0.0178	0.010	0.0316
RF	**AC (%)**	88.93	91.94	87.00	84.37	88.77	88.21
	SE (%)	67.57	87.80	62.03	72.82	65.81	55.48
	SP (%)	94.15	93.73	93.91	90.68	94.20	93.74
	Feature	D	D	D	D	$D+L$	$D+L$
	w **(points)**	200	250	250	200	200	300
	Row SS Ratio	0.7	1.0	1.0	0.7	0.7	0.7
	Col. SS Ratio	1.0	0.4	0.7	1.0	0.4	1.0

providing features that carry discriminatory information for AW and NW. Other techniques for learning from imbalanced data, such as [153], can also be incorporated to achieve an even better performance. Regardless, as it was shown, the obtained performance was comparable to the scenario in which the number of spikes, inferred from calcium signals, are used as features.

Additionally, among the three considered visibility graph measures (D, C and L), it was observed that the measure D, was identified as the feature providing the best sensitivity results, for all subjects and all choice of classifiers, either individually or jointly with other measures (e.g. $D+C$)). This observation indicates that the measure D carries the strongest discriminatory information among the three considered VG measures. Given that D is related to the number of edges in the graph that are associated with the fluctuations in the time series, this result shows that variations in the patterns, and in the relative timing of the fluctuations with respect to one another, play key roles in differentiating the two states. Furthermore, it was demonstrated that the proposed method is capable of providing features common across subjects, which result in successful classification performance.

It is worth noting that, the three different classifiers were implemented independently, to

Table 4.3: Classification performance using unified parameters across subjects and classifiers. $D + C$ is used as the feature, and $w = 200$ points is used as the window length for extracting features in all cases.

Classifier	Performance Measure	Mean	SD
*k*NN	**AC (%)**	**86.54**	2.86
	SE (%)	**66.42**	11.68
	SP (%)	**91.67**	2.82
	AUC	**0.774**	0.057
LR	**AC (%)**	**88.76**	2.67
	SE (%)	**68.65**	11.93
	SP (%)	**93.81**	2.34
	AUC	**0.927**	0.025
RF	**AC (%)**	**87.48**	2.16
	SE (%)	**64.37**	13.07
	SP (%)	**93.31**	1.78
	AUC	**0.912**	0.026

demonstrate the robustness of the VG measures as features. The logistic regression classifier is robust to noise and can avoid overfitting by using regularization. The random forest classifier can handle nonlinear and very high dimensional features. The *k*NN classifier is considered computationally expensive but it is simple to implement and supports incremental learning in data stream. As presented, all classifiers were able to successfully differentiate the whisking conditions demonstrating the robustness of the VG metrics in capturing the temporal characteristic of optical imaging data.

Comparison with spike rate inference-based feature extraction approach

The proposed approach was applied directly to the recorded calcium signals, without using methods such as template matching [154, 155], deconvolution [156, 157], Bayseian inference [158, 159], supervised learning [160], or independent component analysis [161]. Here, we compare the classification performance of the proposed approach with the scenario in which the number of spikes are used as features for each condition.

To infer the spiking events from calcium recordings, we used the FluoroSNNAP [155] toolbox in MATLAB, which utilizes a commonly-used template-matching algorithm. The same window sizes that were considered in VG-based analysis, were also considered for spike-based analysis. For each segment, feature vectors were constructed by concatenating the number of detected spikes from all ROIs. The regularized logistic regression was used as the classifier, with the same ℓ_2 penalty weights as was set before. Similar to the VG-based feature extraction

Table 4.4: Performance comparison of classification experiments based on *i)* VG-based feature extraction from all ROIs, *ii)* Spike-based feature extraction from all ROIs, *iii)* Variance-based feature extraction from all ROIs, *iv)* VG-based feature extraction only from ROI-20, and *v)* VG-based feature extraction from ROIs 25-30.

Window Size (*s*)	1	1.5	2	2.5	3
AUC (All ROIs VG-based)	0.916	0.923	0.927	0.923	0.923
AUC (Spike-based)	0.849	0.882	0.868	0.894	0.896
AUC (Variance-based)	0.914	0.919	0.920	0.915	0.916
AUC (ROI-20, VG-based)	0.841	0.856	0.860	0.857	0.857
AUC (ROIs 25-30, VG-based)	0.825	0.846	0.853	0.857	0.854

technique, the performance was evaluated using the same cross-validation procedure described earlier.

Results for the area-under-the-ROC-curve (AUC) are presented in Table 4.4 for each window size. It is shown that the VG-based approach provides a better performance. This result further confirms the capabilities of VG-based measures in identifying discriminatory features related to different behavior from calcium recordings.

Comparison with signal variance-based feature extraction approach

We carried out another analysis to compare the classification performance of the proposed approach with the scenario in which the variance of the signal is used as features for all candidate window sizes. For each segment, feature vectors were constructed by concatenating the variance from all ROIs. The same classifier and regularization optimization process similar to VG-based approach was used. The AUC values based on *10*-fold cross validation was used to compare the classification performance. The results are summarized in Table 4.4 for each window size, which shows the VG-based method provides a better performance regardless of the selection of window sizes.

Comparison with VG-based features from the somatosensory cortex

We carried out an additional analysis to examine whether the classification results will be different if only signals recorded from the ROIs located in the primary somatosensory cortex are considered, since layer 4 “barrels” in primary somatosensory cortex receive sensory input from the whiskers. Among the ROI locations, the ROI-20 was in close proximity of the primary somatosensory cortex, according to the location of bregma and functional mapping experiments

in a subset of mice. Using the same parameter settings used earlier, classification was performed based on VG measures extracted from ROI-20 signals. Results for AUC are shown in Table 4.4. It can be seen that when features from all ROIs (covering large area of the cortex) are used, the classification performance is significantly better. This result is consistent with previous work [40, 56, 57, 162], which suggest that brain state modulation is widespread across many cortical regions.

In a related analysis, we further used VG-based features extracted from ROIs 25-30, which did not show the epileptiform-like events during NW (as seen in signals obtained from ROI 6). Results are summarized in Table 4.4, suggesting that VG is capable of decoding behavior from ROIs with various dynamic properties. It should be noted that VG analysis in this paper, uses a relatively fast time scale (2 s) compared to the blood-flow related signals that can reduce fluorescent calcium signals. Contamination is particularly strong for sensory-evoked signals [37, 152], but less of a concern here for signals related to spontaneous behavioral state transitions.

4.5 Summary

In this chapter, we presented a novel method for predicting behavior based on VG.To the best of our knowledge, this work is the first study demonstrating that it is possible to infer behavior from the temporal characteristics of calcium recordings, extracted through visibility graph. As such the proposed method could have applications in BMIs involving human [107], or in rodents and primates [108, 109], where from brain recordings subject's intention should be inferred. Due to differences in the nature of recorded signals or experimental conditions, a direct and fair comparison with these studies and the results shown here cannot be made, but the classification results for accuracy presented here are comparable to the results that have been reported in [107, 163, 164]. Additionally, the proposed methodology in combination with widefield optical imaging of ensembles of neurons in awake behaving animals, can open up several new opportunities to study various aspects of brain function and its relationship to behavior. It could also be employed to develop quantitative biomarkers based on VG measures. While here we considered three VG measures (D, L and C), a wide range of other graph measures [165] could also be used to possibly improve the classification performance. It can

be concluded that VG is very effective in providing “quantitative” measures that can reveal differences in recorded calcium time series.

Chapter 5

Data Analytics: Inferring Behavior from Brain Activity Using Visibility Graph and Neural Networks

In this chapter, we extend the work of Chapter 4 by presenting two new analysis frameworks that combine visibility graph and neural networks (NN). We evaluate the frameworks using widefield calcium imaging data and present their strengthened inferring power as compared to using conventional machine learning method.

5.1 Introduction

As discussed earlier, understanding the relation between brain states and behavior is one of the major goals in cognitive neuroscience [56]. Widefield calcium imaging allows detecting excitatory neurons over large area of cortex with relatively high time resolution. It has been shown in rodent cortical neurophysiological studies that, whisking behavior is associated with brain state changes on different levels of cell and systems function [55]. Hence, using widefield calcium imaging to record large-scale dynamics of brain activity, with "active whisking" (AW) and "no whisking" (NW) behavioral states as the proxy of ground truth, opens new windows to investigate the relationship between brain functions and behavior [1, 13].

In Chapter 4, we introduced a method based on VG [166]. The spatial properties of the VG inherit the temporal dynamics of the underlying time series [120]. After the construction of VG, a series of graph measures can be extracted to encode the different aspects of the temporal dynamics of the calcium transient [134]. In [1], classical machine learning algorithms were employed to predict the whisking behavior based on the selected VG measures.

The recent advancement in neural networks has resulted in revolutionary development in many research domains [167]. The main stream of the NN family can be categorized into several architectures, among which the multilayer perceptron (MLP) is the most traditional

one. It can be viewed as an expanded version of the classical logistical regression (LR). MLP has the capability to learn representative characteristics of the data with different levels of abstraction and allows for modeling of complex nonlinear problems [168]. On the other hand, MLP lacks the capability of scaling to various computer vision problems, partially because, as the complexity of the input data becomes higher, the number of parameters and the hardness of the training increase exponentially. The convolutional neural networks (CNN) is another NN architecture which solves this problem by using a detector, known as *kernel* or *filter bank*, to convolve the input signal, and thus largely decreases the number of parameters. Meanwhile, this strategy allows the networks to be sensitive to certain features that could be located everywhere on the input image, which is specially favorable in object detection problems. What's more, while the classical machine learning approaches require considerable domain-specific expertise to design good feature extractors, the CNN performs automatic feature extraction [167].

In this chapter, we present two new analysis frameworks that combine VG and NN.

- With the hypothesis that compared to the classical LR, the MLP will be providing better discrimination power, we build a predictive model using MLP, along with the VG measures that are the same as the ones used in Chapter 4 and [1]. We refer to this approach as VGm-MLP. We expect that by using VGm-MLP, the nonlinear properties of the dataset could be better identified and thus the classification accuracy will be improved.

- With the aim of eliminating the information loss during the VG metric extraction process, we build a CNN model to learn the topological properties from the VGs directly. We refer to this approach as VGi-CNN. Based on previous studies [1, 166], demonstrating that the VG encodes the temporal dynamics of the underlying time series, we expect that the CNN could be successfully capture additional information over the ones that being extracted by the graph measures, and thus will deliver better inferring power, and improved accuracy.

5.2 Methods

We use the same dataset used in Chapter 4 to evaluate the new frameworks. In this section, we first briefly recap the data acquisition and preprocessing procedures. Next, the two frameworks, the VGm-MLP and the VGi-CNN, are introduced. For details about data acquisition, preprocessing, visibility graph, and graph measures, please refer to Section 4.2 in Chapter 4.

5.2.1 VG Metrics-Based Multilayer Perceptron (VGm-MLP)

The Multiple Perceptron is a computational model which is built by cascading a series of perceptron layers, each of which is formed by a set of perceptron units known nodes. The classical LR can be considered as one perceptron unit. A perceptron assigns one weight to each input feature, and calculates the weighted sum for the feature vector. This weighted sum is fed into an activation function (sigmoid function in this study) and is converted to a number between 0 and 1. With a threshold, the perceptron generates a binarized output. In order to fit more complex nonlinear boundaries, a series of perceptron units are organized and cascaded to form a MLP. In MLP, the final layer of perceptrons that creates the predicted labels, is called the "output layer", and those perceptrons are called "output nodes". The layers of perceptrons between the input features and the output layer, are called "hidden layers", and the nodes are called "hidden nodes" accordingly. In the proposed VGm-MLP framework, the stacked VG metrics D, C, L from 30 ROIs are used as the 1-D input feature vector. Two hidden layers are used for the MLP model.

5.2.2 VG Image-Based Convolutional Neural Networks (VGi-CNN)

The convolutional neural networks are designed to process data with multiple arrays, for instance, colored pictures which are composed of three 2-D arrays, and videos which are composed of three 3-D arrays. While there are many variants of CNN, the principle of the CNN is based on four operation stages: convolution, nonlinear activation, pooling, and classification based on full-connected layer. Typically, the convolution, non-linearity and pooling are stacked together to form basic building blocks. Several blocks are cascaded and followed by fully-connected layers to generate the single or multiple binary outputs. In the proposed VGi-CNN

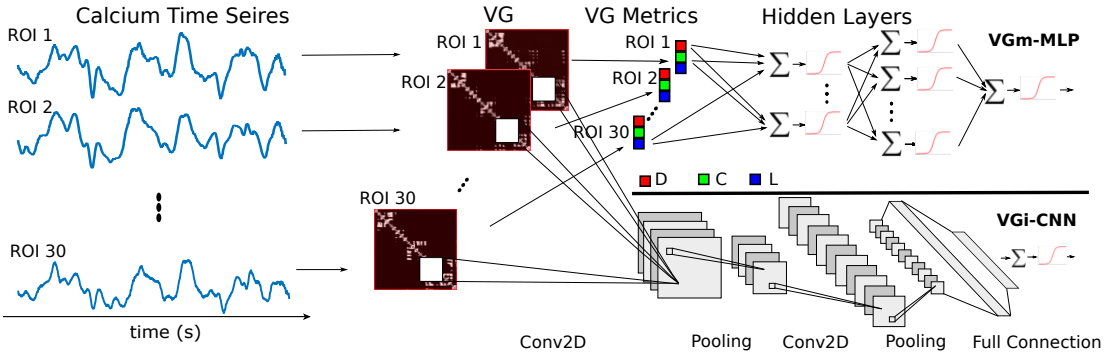


Figure 5.1: Visual illustration for the proposed VG metrics-based multilayer perceptron (VGm-MLP) and the VG image-based convolutional neural networks (VGi-CNN) frameworks, for behavior classification from calcium recordings. The calcium time series are first converted into VGs. To perform the classification using VGm-MLP, several graph metrics are extracted from the VGs (D: Edge Density, C: Averaged Clustering Coefficient, L: Characteristic Pathlength). The stacked graph metrics across ROIs are used as the input vector, to a two-hidden-layer MLP. To perform the classification using VGi-CNN, the VGs are used directly as 30-channel 2-D images to a CNN model.

framework, our task is to use CNN to capture the topological patterns of the VGs directly, which are known as associated with the temporal characteristics of the underlying calcium recordings. To do so, the VGs constructed across 30 ROIs are stacked together to form the 30-channel input images (in contrast to 3-channel (R, G, B) colored images as normally used in computer vision problems). To accommodate the computation complexity, the VG images are re-sampled to 40×40. Two convolutional blocks are used in the CNN model. Fig. 5.1 visually describes the VGm-MLP and VGi-CNN frameworks.

5.3 Results

The preprocessed calcium time series recorded from the six mice were used to perform the behavior decoding process based on VGm-MLP and VGi-CNN frameworks. To improve the generalization capability and to prevent overfitting, the dropout technique [169] was used for both frameworks. We employed the backpropagation and Adam algorithm [170] to train the models. The optimal hyperparameters (learning rate, batch size, and dropout ratio) were identified using grid search. To evaluate the classification performance, a five-fold cross validation is employed. We computed the accuracy, sensitivity, and specificity as measures to evaluate the performance. Using the classical logistical regression as a baseline [1], we evaluated the

Table 5.1: Classification results for the proposed VGm-MLP and VGi-CNN frameworks for each subject, along with their mean (standard deviation) across subjects. Their improvement in the performance are evaluated with respect to the classical logistical regression (LR) model reported in [1], using paired-sample one-tail student's t-tests (df=5). A positive t-value means that the proposed framework delivers higher values for the respective evaluation metric compared to LR model. * and *** represent $p < 0.05$ and $p < 0.005$, respectively.

Classifier	Performance	1	2	3	4	5	6	mean (std)	Comp. to LR
VGm-MLP	**AC (%)**	89.41	92.23	88.56	85.42	89.15	90.93	89.28 (2.32)	$t = 1.15$
	SE (%)	73.14	87.28	70.23	73.06	71.02	68.72	73.91 (6.77)	$t = 1.01$
	SP (%)	93.41	94.57	93.22	92.03	93.33	94.36	93.49 (0.91)	$t = 1.15$
VGi-CNN	**AC (%)**	91.27	91.44	90.04	85.08	89.15	91.02	89.67 (2.41)	$t = 2.15, *$
	SE (%)	84.62	93.73	73.35	79.01	75.46	69.64	79.30 (8.72)	$t = 4.29, ***$
	SP (%)	93.00	90.65	94.50	88.36	92.42	94.27	92.20 (2.34)	$t = -0.77$

improvement in three classification performance metris when using VGm-MLP and VGi-CNN, using paired-sample one-tail student's t-test. The results are presented in Table 5.1. A positive t-value means that the proposed framework improves the performance as compared to the LR model. It is shown that compared to LR, the VGm-MLP framework provides a better classification performance, and the VGi-CNN model provides additional significant improvement, especially for accuracy ($p < 0.05$) and sensitivity ($p < 0.005$).

5.4 Summary

In this chapter, we proposed two analysis frameworks for predicting behavior from calcium imaging data. Both frameworks are based on the temporal dynamics of the calcium recordings, obtained via the VG technique. These frameworks were evaluated by the same dataset described in Chapter 4. Compared to classical machine learning techniques, the proposed frameworks delivered improved performance.

The contribution of these two frameworks is significant. Through the VGm-MLP framework we presented that by using the VG technique, the nonlinear complexity in the temporal dynamics of calcium recordings could be characterized such that a neural networks with relatively simple architecture can discriminate the brain states. A simple model is favorable in the neuroscience field, because the size of experimental data is usually small. By cascading multiple layers of perceptrons, MLP is extremely flexible to fit complex nonlinear data. It is worth to note that while on the presented dataset, VGm-MLP only raises the performance to a small amount, we expect that, compared to LR, it could be more successful when dealing with

dataset with more nonlinear complexity.

The the VGi-CNN framework by taking the advantage of the CNN's capability of automatically discovering the representations needed for classification, the required domain specific knowledge could be minimized. Moreover, instead of using limited number of graph measures to characterize the topological properties, VGi-CNN is able to learn the discriminant features utilizing the full information encoded in the VGs. Another advantage of VGi-CNN comes from the translation invariant property of CNN. As described earlier, VG encodes the temporal dynamics of brain recordings. It is well known that the brain activity exhibits variability in time domain, due to for example differences or delays in neural responses, individuals performance, or brain-related diseases [12]. This temporal variation on the time series would lead to the spatial shift in motif patterns in VG. While this phenomenon may negatively affect the performance of a traditional machine learning method, it wouldn't affect the performance of VGi-CNN due to its translation invariance property. Results showed that the VGi-CNN framework has improved classification accuracy compared to using traditional LR. To the best of our knowledge, this work is the first utilizing CNN and VG to solve the classification problem on the time series data.

While in this study the proposed frameworks are evaluated using the calcium imaging data, we expect that they can find their application outside of the neuroscience domain.

Chapter 6

Data Analytics: On the Fractality of fNIRS Time Series Using Visibility Graph

In this chapter, we present a novel approach, based on visibility graph, for revealing nonlinear features in signals recorded via fNIRS corresponding to brain activities. The technique is demonstrated to differentiating brain states using experimental fNIRS data.

6.1 Introduction

Most existing functional brain imaging studies can be classified into two general categories: *localization studies*, which aim to identify brain's activation patterns in response to specific stimuli [91, 171, 172, 173, 174, 175, 176], and *connectivity studies* (functional or effective) which focus on investigating the functional interactions among brain regions, either when the brain is at rest or engaged in performing a particular task [71, 90, 177, 178, 179, 180, 181]. In exception of a few [107, 182], the majority of these studies investigate the brain function in the spatial domain and typically ignore changes that occur in the temporal domain [14, 180, 183, 184, 185]. However, it is now well known that the brain is highly dynamic [84, 107, 182, 186, 187, 188], and therefore, to gain a more comprehensive picture about its function, methods capable of extracting information about its non-stationarity function are required. In this study, using visibility graph (VG), we present a new approach for revealing the non-linear physical characteristics of fNIRS time series, which has been rarely examined in the literature [189, 190].

Fractal analysis of time series provides an interesting opportunity to study their temporal structure in terms of self similarity, power law scaling relationship, and scale-invariance [191]. Since its introduction, fractal analysis has been used in several fields such as physics, geography, biology, and psychology [192] to reveal such properties in time series, that will

not be visible through conventional analysis. Recently, fractal analysis has also been applied to recordings obtained through functional brain imaging studies [193, 194, 195, 196] to better characterize the observed temporal fluctuations of the signals, or to differentiate a diseased group from a healthy population with an improved accuracy. For instance, using fractal analysis of fMRI data, Sokunbi et al. showed that the hemodynamic response measured from patients with Schizophrenia presented larger complexity, compared to that measured from healthy controls [197]. He, B. showed that the fractal properties of the fMRI signal altered with the changes of brain functional state [198]. Using fNIRS, Khoa and Nakagawa showed that the complex characteristics of the signals recorded during physical motion and imaginary motion of the right arm were different, which can be potentially used for BCI [189]. Lei et al. showed that the power spectra of both EEG and fMRI signals follow the power-law distribution, and scale free brain activity can be characterized by robust temporal structures [199].

There exist several methods to estimate the fractality of a time series [191]. VG is a recently introduced method, which maps a time series into a graph (called a visibility graph), such that its topological properties are related to the fractality and complexity of the time series [196, 120]. VG is computationally less complex compared to conventional fractal analysis approaches [196], and has been successfully used in different studies [200, 201].

For example, using Electrocardiogram (ECG), Jiang et al. showed that employing VG analysis can reveal the dynamic changes caused by mediation training, manifested as regular heartbeat, which is closely related to the adjustment of the autonomous neural system [202]. Zhu et al. applied a VG based approach for alcoholism identification, showing that this approach is promising in separating alcoholic subjects from controlled drinkers [131]. A VG-based approach was used successfully in distinguishing children with Autism Spectrum Disorder from non-autistic children [203].

It is important to note here the difference between VG and the graph theoretical-based methods used commonly in functional connectivity studies [204, 165]. In functional connectivity studies, graphs are constructed in the spatial domain, i.e. nodes in the graph correspond to the location of channels or voxels, and links between two given nodes are formed based on the statistical similarity of the time series associated with the two nodes, quantified by measures such as correlation. In VG, on the other hand, which is formed for a given time series (see Fig.

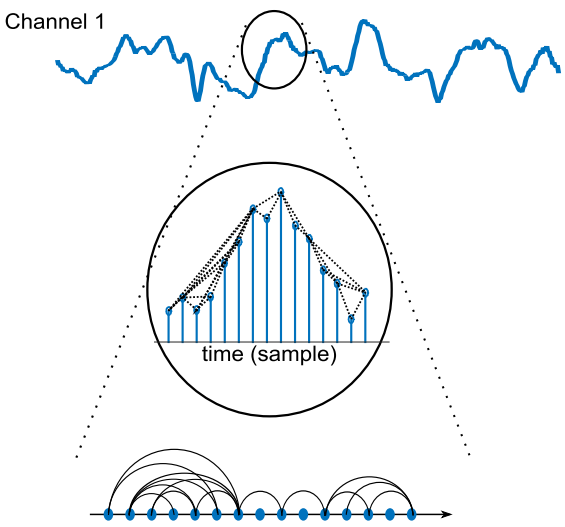


Figure 6.1: Conceptual illustration of constructing visibility graph from a time series. Nodes in the graph correspond to the time points in the time series, and links are formed based on the visibility condition.

6.1), the nodes correspond to the time points of the time series, and the links are formed based on natural visibility between the time points (discussed in Chapter 4). Once the graph is formed for each time series, graph metrics can be extracted to represent different properties of the time series.

This chapter presents a VG-based framework for analyzing the fractality of fNIRS time-series, and demonstrates its potential in detecting whether the brain is at rest or is engaged in executing tasks. fNIRS time series are recorded from nine healthy male subjects at two resting-state conditions and two task conditions. Artifact removal procedure is performed to reduce the interference related to motion [205]. VGs are constructed for each channel and each condition, from which the power of scale-freeness of visibility graph (PSVG) is extracted and compared across conditions. The PSVG values estimated during resting-states and task sessions are then used to form features to feed into different classifiers, including k-Nearest Neighbors (kNN), linear and non-linear Support Vector Machine (SVM), and Linear (LDA) and Quadratic Discriminant Analysis (QDA), to verify the efficacy of using the VG-based metric to infer brain states. To the best of our knowledge, this is the first study to employ VG to reveal temporal features of fNIRS-recorded time series and demonstrate its feasibility in identifying features in fNIRS recordings that can be used to distinguish different brain states.

The rest of this chapter is organized as follows. Section 6.2 describes the methods used for

analysis in this study. The details of the experimental setup are given in Section 6.3. Section 6.4 presents the experimental results using VG, and finally the summary is provided in Section 6.5.

6.2 Methods

In this section, theoretical basis of the methods used in the analysis are described. The fractal properties of fNIRS time series will be evaluated using two approaches, Scaled Windowed Variance (SWV) analysis and VG. These methods are described here. Furthermore, modalities used for classification are briefly reviewed. In the following, we denote an N-point time series (e.g. recorded by a given fNIRS channel) with $\mathbf{x} = \{x_i\}_{i=1}^{N}$.

6.2.1 Scaled Windowed Variance Analysis

We first evaluate the fractal dimension of fNIRS time series using conventional methods. Fractal dimension can be estimated in time domain, frequency domain, or time-frequency domain, using various methods [191]. The choice of the proper method has to be made based on the properties of the time series, and whether it is stationary or non-stationary. As it is well known that physiological signals and brain activities are non-stationary [191, 206], Scaled Windowed Variance Analysis method [191, 207] is used here to estimate the fractal dimension (defined below) of fNIRS-recorded time series.

To estimate the fractal dimension of a time series using SWV, the time series is partitioned into non-overlapping segments of size n. If $\hat{\mu}$ represents the mean of the segment, the standard deviation for each segment, $\hat{\sigma}_n$, is computed as [191]

$$\hat{\sigma}_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2}. \tag{6.1}$$

This measure is computed for all segments, and then is averaged to obtain $\bar{\sigma}_n$. The procedure is repeated for different window sizes.

For non-stationary fractal time series, the windowed mean standard deviation, $\bar{\sigma}_n$, and its

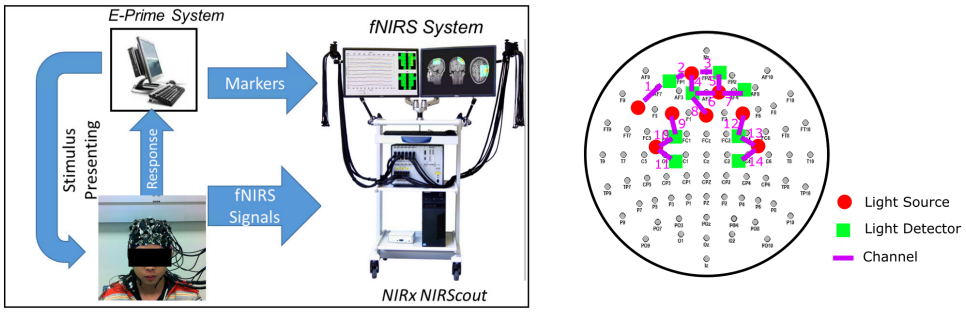


Figure 6.2: a) Illustration of the experimental setup, and b) map of fNIRS optode locations.

corresponding window size n, follow a power law relation given by [191]

$$\bar{\sigma}_n \overset{\mathrm{d}}{=} p \cdot n^H, \tag{6.2}$$

where $\overset{\mathrm{d}}{=}$ represents equal in distribution, H is the Hurst coefficient that can be obtained by calculating the slope of the least-squares linear regression line of $\log(\bar{\sigma}_n)$ versus $\log(n)$, and p is a proper prefactor [191]. The fractal dimension, D, is linearly related to H as $D = 2 - H$ [208].

6.2.2 Power of Scale-freeness from Visibility Graph

The construction of visibility graph is described in Chapter 4. The constructed VG reveals the dynamic properties of the time series in unique ways. For example, periodic signals result in regular graphs, and fractal time series result in scale-free networks [196, 120, 200, 134]. Scale-free corresponds to the property of the graph that, independent of the number of nodes, its degree distribution $P(k)$, has a power-law tail where the tail exponent obeys the power law, i.e.,

$$P(k) \sim k^{-\gamma}. \tag{6.3}$$

In (6.3), k represents the degree of the node (i.e., the number of links connected to a node), $P(k)$ denotes the degree distribution (i.e., the fraction of nodes with degree k), and γ denotes the power of scale-freeness (PSVG). It has been proved that the PSVG is indicative of the fractality of the time series [166].

6.2.3 Classification

The feasibility of VG-based metrics in distinguishing different brain states will be evaluated via classification. kNN, linear and non-linear SVM, and LDA and QDA classification techniques are employed. A brief description of each techniques follows.

kNN is a widely used classification technique in brain studies [141, 142, 143, 209, 210], in which the label of a testing data is determined by the labels of the k neighbors in the training dataset that are nearest to the testing data, according to some distance measures (in this chapter, the Euclidean distance is used) [144]. In kNN, the best value of k can be determined by cross-validation [143], and in this study the values of k in the range of 1 to 6 are used.

SVM is another supervised classification algorithm, which tries to maximize the distance between the separating hyperplane and the nearest training points [5, 211]. Although regular SVM is a linear classifier, it can make non-linear decision boundaries by using kernel functions such as radial basis functions (RBF) or polynomial kernels. Non-linear SVM might provide increased classification accuracy, with the expense of extensive computation load and the risk of overfitting. SVM have been shown to offer good performance with hemodynamic activity data in recent studies [5, 32, 212, 213, 214]. In this study both linear and non-linear SVMs are used.

Discriminant Analysis has been frequently used in fNIRS studies [215, 216, 217, 218, 219]. Discriminant Analysis looks for a boundary by which the samples are well separated from each other. The boundary can be linear for LDA, or quadratic for QDA, both of which maximize the between-class variance and minimize within-class variance [216, 215]. In this study we implement both LDA and QDA.

The classification performance is evaluated by a leave-two-out-cross-validation procedure [220, 221], which is an exhaustive cross-validation method and when compared to the traditional leave-one-out method, helps to improve the evaluation confidence for small sample-sized data. The classification results are measured using the following metrics,

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}, \quad \text{Sensitivity} = \frac{\text{TP}}{\text{TP}+\text{FN}}, \quad \text{Specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}}. \tag{6.4}$$

In (6.4), TP, FP, TN and FN refer to true positive, false positive, true negative, and false negative, respectively.

Accuracy represents how successful the classifier is in correctly classifying the samples. Sensitivity indicates how well the classifier detects a test data from group 1 (here resting-state), while specificity suggests how good the classifier is at identifying a test data from group 2 (here task execution).

6.3 Experimental Procedure

In this section, we describe the experimental setup and the preprocessing steps used to remove artifacts from fNIRS data. Fig. 6.3 shows the pipeline of the data analysis.

6.3.1 Experimental Paradigm and Data Acquisition Procedure

Nine healthy male subjects (age mean: 25, age SD: 4.8) participated in our experiments, after providing their written informed consents, approved by the Rutgers Institutional Review Board (IRB). The experiments included two resting-state sessions (eyes-closed (EC) and eyes-opened (EO)), and two block-designed (number of blocks $= 3$) tasks. Each resting-state session lasted for 10 minutes. The tasks were the Response Time (RT) task, and the modified Go No-Go (GNG) task. Each block in RT task consisted of 50 left and 50 right arrows presented to participants in random order with inter-trial interval (ITI) of $[800 - 1200]$ ms. Participants were asked to press right/left mouse button depending on the direction of presented arrow. Each block in the modified GNG task consisted of 60 Go and 30 No-Go symbols with ITI of $[800 - 1200]$ ms [222]. Participants were asked to click only when the Go symbol is presented. Recordings for the two resting-state and the two tasks sessions were all completed in one setting.

Changes in the optical signals were recorded at two wavelengths (685 nm and 830 nm) at the sampling rate of 12.5 Hz, using NIRScout system (NIRx Medical Technologies, LLC). Note that although the maximum sampling rate of NIRScout system is 62.5 Hz, in this experiment, 12.5 Hz was used as the practical sampling rate due to the time-division multiplexing on the illuminated light sources to avoid crosstalk. A total of 14 channels were used (8 light sources and 8 detectors arrangement). Optodes were placed to cover the prefrontal and motor cortices,

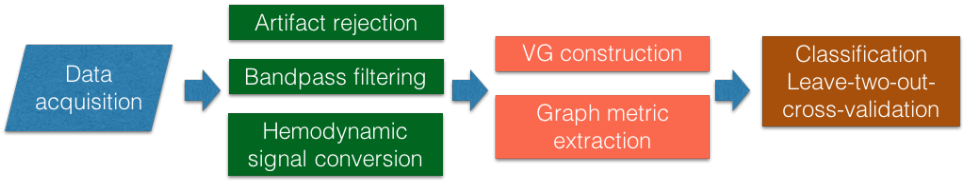


Figure 6.3: The data analysis pipeline.

where brain activities are expected according to the nature of tasks. Fig. 6.2 shows the location of optodes.

6.3.2 Data Preprocessing

In this section, we describe the artifact rejection, bandpass filtering and hemodynamic signal conversion process.

Motion-related artifacts inevitably exist in fNIRS-recorded time series, and if not removed, can negatively impact the outcome of VG analysis. Therefore, artifact rejection procedure was first performed [205]. Two types of artifacts, "spikes" and "discontinuities" (jumps), were considered. For each channel, short intervals containing spikes were identified by visual inspection, and then replaced by the average of the signals of the same interval from two neighboring channels. Discontinuities were detected by examining the difference in signal values at each two successive time point. A time point pair with discontinuity was identified when this differences becomes larger than four times of the standard deviation of the time series, and was eliminated by subtracting the difference from the values for points after the jump [223].

Next, optical signals were band-pass filtered in the range of 0.01 Hz and 0.2 Hz to exclude the low-frequency drift, and the interference caused by physiological sources such as eye-blinking, heart beaten, and respiration [224, 225]. The filtered signals were converted into $\Delta[\mathrm{HbO}_2]$ and $\Delta[\mathrm{HbR}]$ according to the modified Beer-Lambert law [226]. The artifact rejection, bandpass filtering, and data conversion procedures were performed using the nirsLAB toolbox [223, 32]. For the RT and GNG tasks, time series were segmented according to the border of the blocks (170 s for each block). The same duration was used to segment the signals from the resting-state recordings, into three non-overlapping blocks. This procedure for all subjects resulted in a total of 108 14-channel time series, each with 2125 data points.

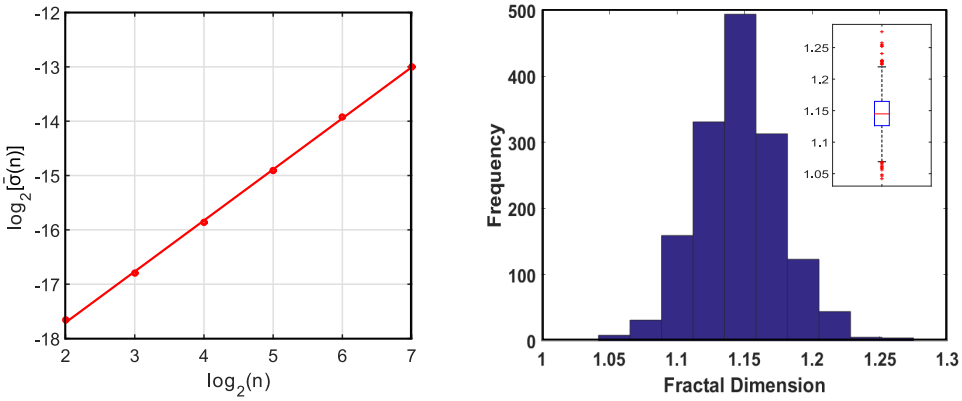


Figure 6.4: Results from SWV analysis. a) Calculated windowed mean standard deviation as a function of the window size in a double logarithmic plot for a given time series. The slope of the fitted least square trend is related to the fractal dimension of the time series, and b) distribution of the estimated fractal dimension for all fNIRS time series.

6.4 Results

In this section, we present the results of fractal analysis and classification experiments.

6.4.1 Results from Scaled Windowed Variance Analysis

Fractality of fNIRS recordings, for each channel, subject, condition, and block, was first examined using SWV analysis, which is an appropriate method for non-stationary fractal time series [191]. Fig. 6.4-a shows the obtained mean standard deviation ($\bar{\sigma}_x(n)$) as a function of the window size (n) in a double logarithmic plot, for a given time series. As can be seen, for the window size ranging from 2^2 to 2^7, the data follows a linear trend. By definition, the slope of this trend equals the Hurst exponent, H, from which the fractal dimension D can be estimated from $D = 2 - H$ [191]. As a metric related to the self-similarity of the time series, the value of D ranges between 1.0 and 2.0. D for a random time series is 1.5. For a time series with high correlation or memory over time, the extracted D value is near 1 [208].

The distribution of estimated D values for all preprocessed time series (resting-state and task) is shown in Fig. 6.4-b. As it can be seen, the estimated D value for all time series is less than 1.3. This result confirms the existence of high degree of fractality (positive correlation) in fNIRS time series.

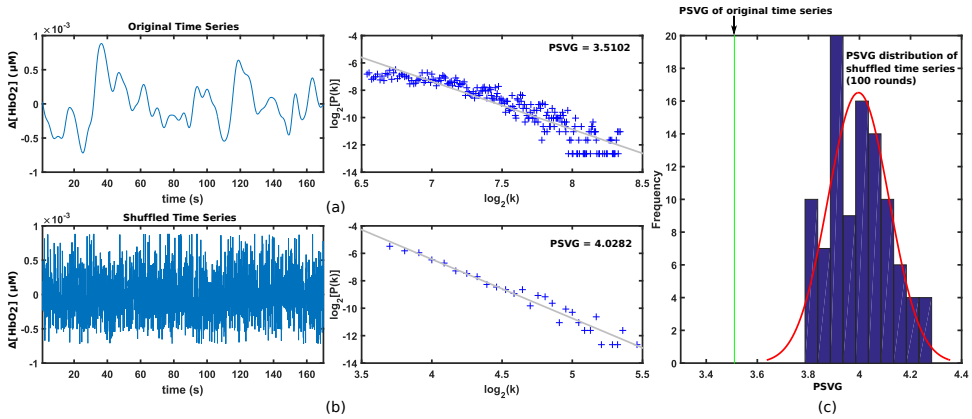


Figure 6.5: Comparison of the PSVG, estimated for a representative resting-state block of fNIRS time series, and its randomly shuffled version. a) The time course of a representative fNIRS recording and its power-law tail of the degree distribution, b) the time course of randomly shuffled version of the same time series, and its power-law tail of the degree distribution, and c) comparison between the PSVG of the original time series, and the distribution of PSVGs for 100 randomly shuffled versions of the original time series.

6.4.2 Results from Visibility Graph Analysis

To evaluate the effectiveness of VG in capturing the temporal structure of time series, first, the VG was constructed for two cases: a representative preprocessed fNIRS time series, and its randomly shuffled version (i.e. the order of appearance of data points in time was randomly shuffled). The PSVG was then estimated for each case. The time series and their associated power-law tail of the degree distribution, for a representative original time series (a $\Delta[\mathrm{HbO}_2]$ signal associated with one EC block) and its shuffled version, are shown in Figs. 6.5-a and 6.5-b, respectively. It is shown that the degree distribution in the two cases decay at different rates.

The estimated PSVG for the original time series, $PSVG_{orig}$, is 3.512. For the shuffled version, random shuffling was performed for 100 rounds, and for each round the power of scale-freeness, $PSVG_{shuf}$, was estimated. The averaged $PSVG_{shuf}$ is obtained as 3.996 (standard deviation is 0.120, range $[3.787 - 4.283]$, refer to Fig. 6.5-c). It is observed that although both the original time series and its randomly shuffled version have the same distribution in terms of amplitude of data points, they have different PSVG values. This result implies that the temporal structure of the time series, rather than the distribution of amplitude of data points, can be characterized from the PSVG values.

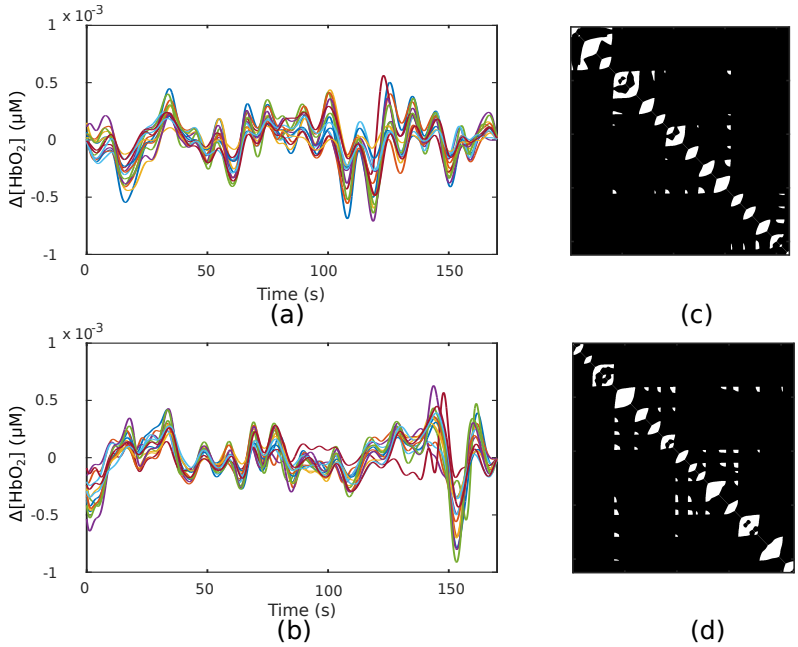


Figure 6.6: The post-preprocessed $\Delta[\mathrm{HbO}_2]$ time series recorded from one representative subject under EC condition (shown in a)) and GNG condition (shown in b)). The waveforms with different colors represent the time series recorded from different channels. c) and d) represent adjacent matrices of VGs associated with channel 1 under the two conditions shown in a) and b), respectively. The dark color represents that there is no connection (no visibility), and the light color represents the existence of an link (visibility).

The visibility graphs were then constructed for the time series associated with each channel, subject, condition, and block, separately. As an example, Figs. 6.6-a and 6.6-b show two representative $\Delta[\mathrm{HbO}_2]$ signals recorded from one subject under EC and GNG blocks, respectively. The waveforms with different colors represent the $\Delta[\mathrm{HbO}_2]$ time series recorded from different channels. Figs. 6.6-c and 6.6-d illustrate two adjacent matrices of VGs associated with channel 1 under the two conditions (EC and GNG) as shown in Figs. 6.6-a and b, respectively. In the following analysis steps, we extract the graph metrics to infer whether the subject is under resting-state or is executing a task.

Figs. 6.7-a and 6.7-b present the degree distribution patterns averaged across subjects for each of the 14 channels, and their zoomed in power-law tail, respectively. The colors correspond to different conditions (EC, EO, RT, and GNG). A least-square regression line was fitted to the power-law tail and from it, the PSVG (equal to the negative of the slope) was calculated ($\gamma = -\frac{\log[P(k)]}{\log(k)}$). Fig. 6.7-b shows that for most channels, the PSVG are different across conditions. The mean and standard deviation of the estimated PSVG values, associated with each

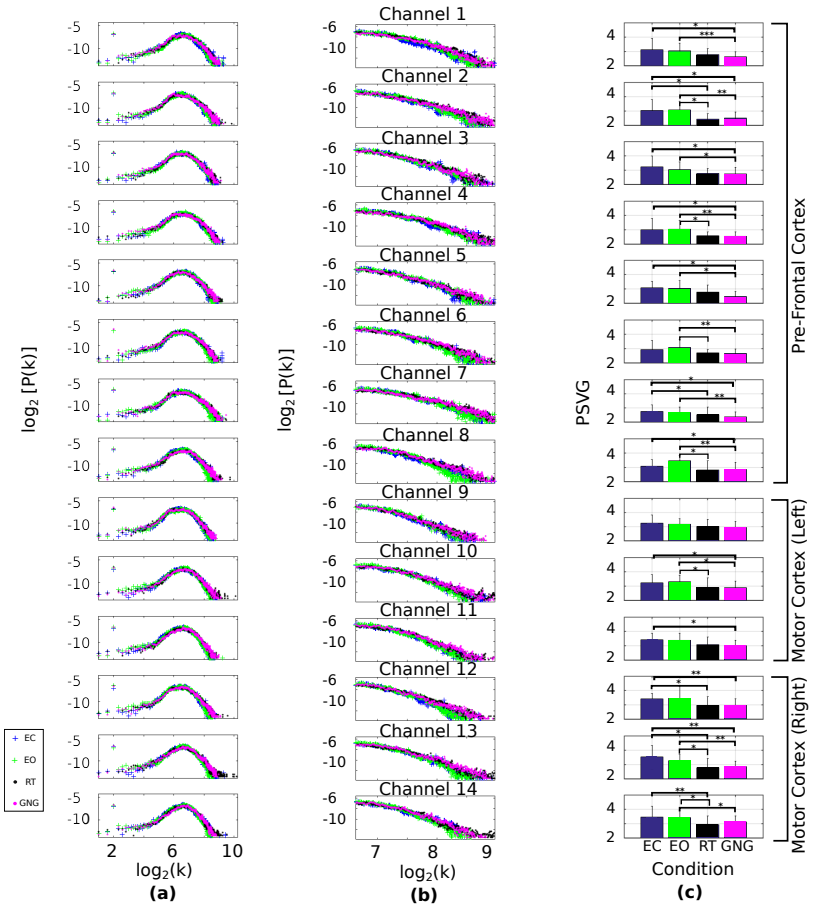


Figure 6.7: a) Degree distribution results averaged across subjects for 14 channels. b) The linear range of the averaged degree distribution for 14 channels. c) The mean and standard deviation of estimated PSVG for each condition across subjects. The pairs of conditions presenting statistical significant difference are labeled using $*$ notations. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

condition, and each channel, are presented as bar plots in Fig. 6.7-c. Paired-sample student's t-tests were performed for each channel across each pair of conditions (EO - EC, EO - RT, EO - GNG, EC - RT, EC - GNG, RT - GNG). For each channel, the pairs of conditions presenting statistical significant difference were labeled using "star" notations. It is shown clearly that for all subjects, the PSVG values associated with resting-states (EC and EO) are larger than those associated with task-execution (RT and GNG) conditions. Using this result, we hypothesize that the PSVG values of VGs can be used to distinguish brain states. In the following, we perform classification to test this hypothesis.

6.4.3 Classification Results

Binary classification was performed to evaluate the feasibility of using PSVG in distinguishing brain states. The individual multi-channel PSVG values associated with EC and EO conditions were pooled to form the resting-state sample data, and those PSVG values associated with RT and GNG conditions were pooled to form the task-execution sample data. Before pooling, the individual PSVG values were normalized across conditions, so their ℓ_2-norm were equal to 1 for each channel. This procedure left 18 data for each binary condition.

Two main components are necessary for performing classification: feature selector and classifier [227]. The objective of the feature selector is to identify channels that are most relevant for the subsequent classification task. Considering the small number of samples in the study, reducing features from all channels to a small group of, but highly relevant channels would help to prevent the *curse of dimensionality* problem [228, 229]. To achieve this, a two-sample t-test was performed across resting-state and task-execution data for each channel. Fig. 6.8 shows the t-values of all channels in descending order where the t-values associated with significant levels of $p = 0.01$ and $p = 0.001$ are also shown ($t = 2.750$ and 3.646 corresponding to $p = 0.01$ and 0.001, respectively). While Fig. 6.8 shows that the PSVG values for most channels are significantly different (for $p = 0.01$), it is expected that the amount of information related to differentiating between resting-state and task-execution conditions is mostly contained in the top-ranked channels (Channels 2, 13 and 4 corresponding to a $p = 0.001$). To confirm this, classification experiments were performed where an increasing number of top-ranked channels were selected to form the feature vector (starting with the top-ranked channel, i.e., Channel 2,

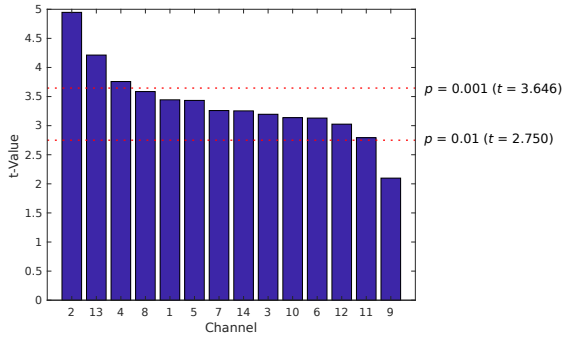


Figure 6.8: The results of two-sample t-tests between two brain-states (resting-states vs. task-execution) across subjects for each channel. The channels are ranked in descending order according to their t-values. The red horizontal lines represent t-values associated with $p = 0.001$ and $p = 0.01$, respectively.

and sequentially adding more channels with the largest t-values). For the classifiers, we chose kNN, linear and non-linear SVM, LDA, and QDA. We note that previous studies have used artificial neural networks (ANN) for classification of fNIRS data [216], but considering the small sample size in this study, we did not implement ANN.

Leave-two-out-cross-validation procedure was used, in which the classification was repeated $\binom{36}{2} = 630$ times, where each time two samples were assigned to the testing dataset and the remaining samples were assigned to the training dataset. The classification results were evaluated using the measures of accuracy, sensitivity, and specificity. Next the results were averaged across all repetitions for a given selected number of channels. For kNN, k was selected in the range of 1 to 6. For SVM, different kernels (linear, RBF, quadratic, 3^{rd} and 4^{th} polynomial) were implemented. For Discriminant Analysis, LDA and QDA were used.

The accuracy of classification results as a function of the number of top-ranked channels are illustrated in Figs. 6.9-a, 6.9-b, and 6.9-c. It can be seen that all classifiers achieved significantly better accuracy than that of naive classifier (50%) and better results are obtained when only top-ranked channels are used. The performance of kNN generally improved with increasing values of k. For example, when using the two top-ranked channels and when $k = 5$, kNN achieved the best result. Similarly, LDA and SVM achieved their best results when the top two or three ranked channels were used, respectively. The linear and non-linear SVM (using RBF kernel) worked equally well and slightly better than that of kNN and Discriminant Analysis. For Discriminant Analysis, the performance of LDA is better than QDA. Overall,

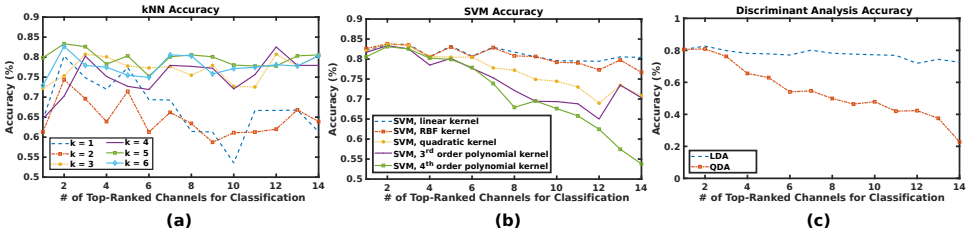


Figure 6.9: The classification accuracy using kNN with different values of k (shown in (a)), SVM with different kernels (shown in (b)), and Discriminant Analysis with different discriminant types (shown in (c)).

Table 6.1: Classification performance using different classifiers.

	Classification Performance			
Classification Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)	Parameters
kNN	83.33	88.89	77.78	$k = 5$, Channels 2 and 13
SVM	83.73	87.46	77.78	kernel: linear, Channels 2, 4, and 13
Discriminant Analysis	82.62	87.46	77.78	model: LDA, Channels 2 and 13

it is observed that the non-linear classifiers (QDA and non-linear SVM) did not improve the accuracy of classification due to the relatively small sample size.

Table 6.1 is a summary for the best classification performance for kNN, SVM, and Discriminant Analysis evaluated by accuracy, sensitivity, and specificity. Channels 2, 4, and 13 contributed most to the accuracy of classification. The results revealed that the PSVGs of hemodynamic signals measured from both prefrontal (Channels 2 and 4) and motor (Channel 13) cortices had contributed to the success in differentiating resting-state and task-execution conditions. It is worth to note that in our previous study using EEG-fNIRS multimodal imaging, the location of channel 2 was detected as the most sensitive channel to GNG tasks, which is consistent with the present study [10].

6.5 Summary

The human brain is a highly complex system with non-linear dynamical behavior [230]. As such, methods such as fractal analysis, that can reveal non-linearities in the recordings associated with brain activities, can serve as an important complimentary approach to commonly used methods (e.g. used for functional localization or functional connectivity), to gain a more comprehensive knowledge about how the brain functions.

In this study, we investigated the fractality of fNIRS recorded time series obtained during

task execution or when the brain is at rest using visibility graph. Fractal dimension for all conditions and recordings were first estimated through SWV analysis. It was shown that fNIRS recordings, regardless of the experimental conditions, exhibit high fractality. This result is in line with previous studies that used EEG or fMRI to monitor brain function [196, 231, 232].

Next, VGs were constructed for all fNIRS time series recordings, and the corresponding PSVG were calculated. Results showed that for most channels, the difference in PSVG values for cases when the brain is at rest and when the brain is engaged in executing tasks is statistically significant. To the best of our knowledge, this is the first study that explores the possibility of employing VG in differentiating brain states.

The capability of VG-based metrics in differentiating brain states was further examined by performing classification using the PSVG values estimated during rest or task execution as features to the classifier. A wide range of classifiers were used, and although the number of training and testing samples in this study was small, a reasonably good accuracy is obtained using each classification technique.

Our results indicate that for various classification modalities, the top-ranked channels corresponding to the prefrontal and motor cortices are the most relevant in identifying the brain state. This reveals that both cortical regions have committed to the execution of the experimental task. Note that this result does not imply that other channels are not relevant to the task-execution. Rather, the result suggests that the top-ranked channels have brought adequate information for distinguishing experimental conditions, and the information brought by lower-ranked channels might be overlapped by those brought by top-ranked channels. A larger sample size will be needed to further examine the capability of VG-based metrics in distinguishing different brain states.

One application of VG-based metrics being used as features would be to use them as biomarkers for diagnosing brain-related disorders. Due to the instrumental advantages of fNIRS technology (e.g. portability and low-cost), fNIRS has great potential for clinical settings. As such, finding features in fNIRS recordings that can serve as biomarkers for diagnosis is of great interest to clinicians. Based on the initial results presented in this study and given that VG-based metrics can reveal non-linear properties of the brain recordings, VG-based metrics could be potentially used as biomarkers for brain-related diseases for cases where there are

disease-related information hidden in the non-linear properties of the recordings.

Chapter 7

Data Analytics: Probing the Dynamics of Spontaneous Cortical Activities

In this chapter, we present a novel framework, based on wavelet transform coherence and multivariate permutation test, for evaluating spatial and spectral changes in functional connectivity as related to spontaneous activity. This method is validated using data obtained from widefield calcium imaging of GCaMP6 transgenic mice.

7.1 Introduction

Over the past decade an increased number of studies have investigated functional connectivity under various conditions using imaging modalities such as fMRI [233, 234], fNIRS [78, 185, 15, 235, 236], and EEG [84, 237, 238, 239]. Many of these studies have suggested that functional connectivity varies in time, even without the presence of external stimulus [240, 233, 84, 241, 242, 243, 244, 245, 246, 247]. Accordingly, different methods, such as short time Fourier transform, and wavelet coherence [244, 245, 246, 247], have been utilized to capture the dynamic nature of functional connectivity across time, space and frequency. Frequency-based methods, however, mostly require a prior knowledge on which frequency bands are involved in the activity.

In this study, the cortical brain activity are recorded using widefield calcium imaging, and spontaneous activity are differentiated by monitoring whisking activity. Two conditions are considered: no whisking (NW) and active whisking (AW). The proposed method is used to identify differences in functional connectivity (across space and frequency) when the mouse is involved in AW and NW.

The rest of this chapter is organized as follows: the proposed methods are described in Section 7.2, results are presented in Section 7.3, and discussed in Section 7.4. For the details

of experiments and preprocessing steps please refer to Chapter 4.

7.2 Methods

7.2.1 Wavelet Transform Coherence

Wavelet transform coherence (WTC) enables finding the coherence between two time series across time and frequency [233, 248, 249, 250, 15, 251]. When applied to time series obtained from two ROIs, WTC can be used as a measure of functional connectivity of the two regions.

For a time series x_n of length N, its continuous wavelet transform, $W_{x_n}(n,s)$, sampled at equal time steps of size Δt is obtained as [233]

$$W_{x_n}(n,s) = \sqrt{\frac{\Delta t}{s}} \sum_{m=1}^{N} x_n \psi_0^*[(m-n)\frac{\Delta t}{s}], \tag{7.1}$$

where the parameter n denotes the time index, s represents the wavelet scale, and ψ_0 is the wavelet function (here we used Morlet). The WTC ($R^2_{x_n,y_n}(n,s)$) of two time series x_n and y_n is find as [233]

$$R^2_{x_n,y_n}(n,s) = \frac{| < s^{-1} W_{x_n,y_n}(n,s) > |^2}{| < s^{-1}|W_{x_n}(n,s)| > |^2 \cdot < |s^{-1}|W_{y_n}(n,s)| > |^2}, \tag{7.2}$$

where $< . >$ represents smoothing in both time and scale.

7.2.2 Multivariate Permutation Test

Permutation test was developed based on the reasoning that if two datasets were drawn from the same population, the labels of the data from the datasets are exchangeable [252, 15]. The conventional univariate permutation test starts by computing a statistic, t-value for instance, for the observed data across two conditions. Then the data observed from the two conditions are randomly permuted, and a new t-value is computed. This permutation procedure is repeated for many times (repeated exhaustively for all possible arrangements or by randomly re-sampling a large number of data). For each case, a new t-value is computed. These set of t-values form a distribution under the null hypothesis that the data from two conditions can be swapped. The permutation test does not rely on assumptions about the distribution of the data [253, 254].

The framework of multivariate permutation test was generalized based on the procedure described above, in which multiple measures under different conditions can be considered simultaneously [252, 255]. Multivariate permutation test had been applied previously in assessing the statistical significance of waveform difference across conditions in EEG studies [255]. In this study, we use multivariate permutation test to evaluate statistical changes in functional connectivity, as measured by WTC. For each permutation, *a set of* observed WTC values is permuted across sessions. This results in a set of statistics (here t-values) for each frequency bins. The most extreme statistic is reserved as an element for the null hypothesis distribution. The statistical significance (p-value) of WTC values associated with individual frequency bins are then obtained by being referenced on that null hypothesis distribution.

This procedure has been automatically adjusted for multiple comparisons because of the outwardly extended null hypothesis distribution which is constructed by the extreme statistics across variables. As more multiple comparisons (more frequency bins are involved) are performed, more conservative p-values would be obtained [252].

7.3 Results

There were a total of 11 recording sessions for six mice, with each session consisting of 15 or 16 blocks. Fig. 7.1 illustrates an example for the post-preprocessed signals from all ROIs, as well as the corresponding spontaneous whisker movement recordings, obtained from subject 1 in block 5 of the recording session 1.

For every possible ROI-pair, the WTC scalogram was first computed for signals obtained in a session. Segments corresponding to AW and NW conditions (as identified based on the whisker movement recordings) were then identified on the scalogram. Fig. 7.2 shows a representative WTC scalogram obtained for recordings from ROI-3 and ROI-23 in block 5, and recording session 1. The WTC values (columns of the scalogram) associated with similar condition were pooled and averaged across recording blocks, respectively. This procedure was done for recordings from each session, resulting in 11 109×1 vectors for each ROI-pair per experimental condition, where 109 refers to the number of frequency bins, ranging from $0 - 50$ Hz.

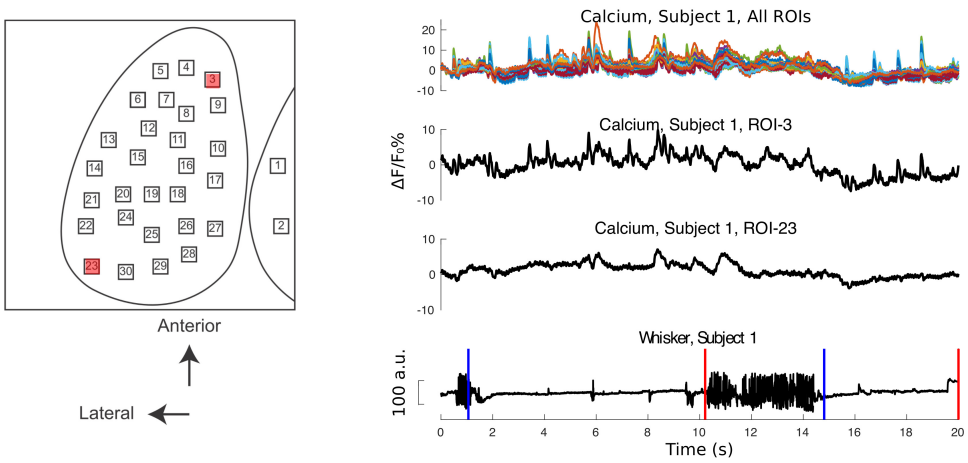


Figure 7.1: Example of post-preprocessed calcium signals and the corresponding spontaneous whisker movement recordings obtained from subject 1 in block 5 of recording session 1. Right from top to bottom: signals from all ROIs; signal from ROI-3; signal from ROI-23; spontaneous whisker movement recording. The blue vertical bars represent the starting points of NW periods, and the red vertical bars represent the ending points of NW periods. Left: Locations of ROI-3 and ROI-23 have been highlighted.

Multivariate permutation tests were then performed on data from two conditions obtained for each ROI-pair. The procedure is summarized below for a given ROI-pair.

- **forming the sample sets:** The sample sets for each condition (AW and NW) consists of the corresponding mean WTC values obtained from each session. That is, for a given ROI, each sample set consists of 11 109×1 vectors.

- **constructing the null hypothesis distribution:** The basis of permutation test is that if the null hypothesis (no difference across conditions) is true, exchanging the WTC values across conditions would have no effect on the statistical measures (here the t-value). For each permutation run, the vectors are shuffled between samples, and the t-value for each frequency bin is computed. The maximum absolute t-value, t_{max}, across all frequency bins of the permutation run as an element of the null hypothesis distribution is identified. The null-hypothesis distribution is constructed by pooling t_{max} from all permutations. A total of 2^{11} permutation runs were performed.

- **evaluating statistical significance:** For each frequency bin, the t-value based on the original observed WTC values is computed. This t-value is referenced to the null hypothesis distribution to evaluate the p-value.

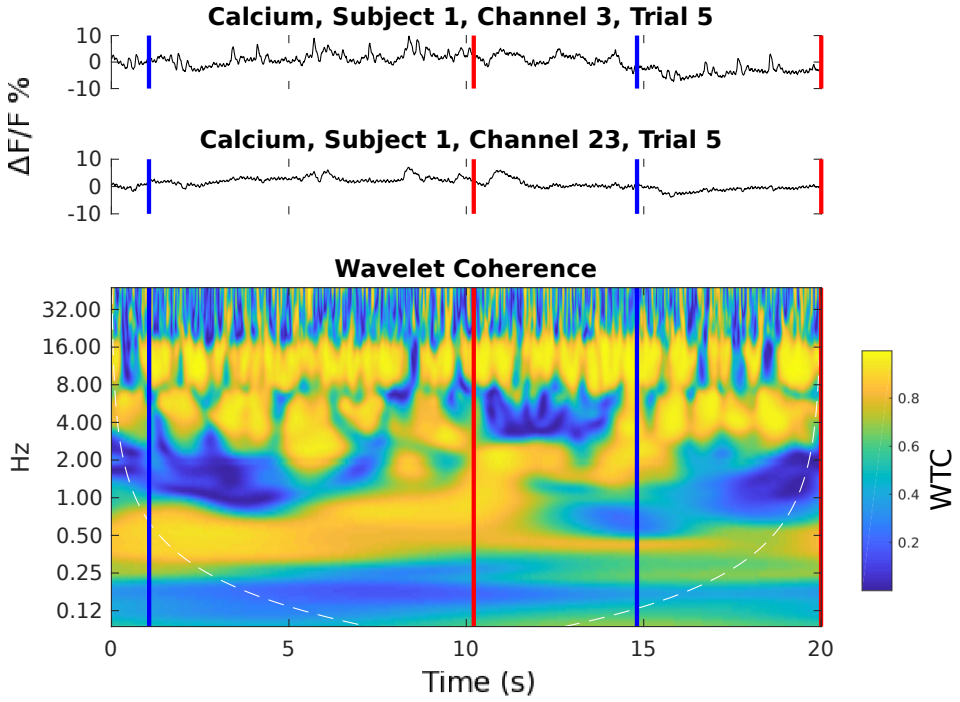


Figure 7.2: Representative WTC scalogram for recordings obtained from ROI-3 and ROI-23 in block 5 of recording session 1 for subject 1. The colored vertical bars carry the same information as those in Fig. 7.1.

The computed *t*-values for all frequency bins and ROI-pairs are presented in Fig. 7.3. It can be seen that large *t*-values are obtained for many ROI-pairs. They can be clustered into two groups: one for those with positive *t*-values, and the other for those with negative *t*-values.

By referencing these *t*-values on the constructed null hypothesis distribution, along with the location of the ROIs on the cortex, we obtained a topological map for ROI-pairs revealing significant changes across spontaneous whisking conditions for each frequency bin, with an α level of 0.001, corrected for multiple comparison. Fig. 7.4 presents the ROI-pairs that reveal significant differences in functional connectivity, as measured by WTC, across spontaneous whisking conditions. The ROI-pairs showing larger WTC values in NW condition are presented in the left panel. The ROI-pairs showing larger WTC values in AW condition are presented in the right panel. Results demonstrate that frequencies between $4 - 8$ Hz correspond to NW condition (shown in the left graph) and the frequencies around 15 Hz correspond to AW condition (shown in the right graph). This result is consistent with previous study [106] which showed that passive brain states (NW) are associated with slower, larger amplitude oscillatory activity,

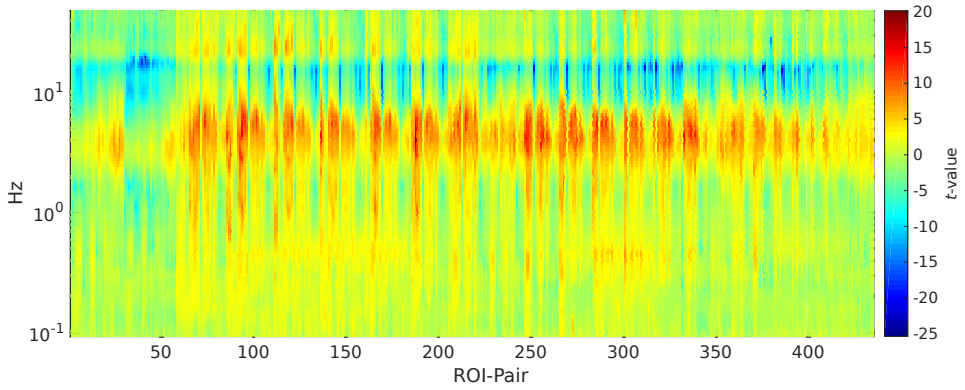


Figure 7.3: Raster diagram illustrating t-values for all frequency bins and ROI-pairs computed from the WTC values.

where the neighboring neurons show synchronized activity. The slow oscillations are reduced during active brain states (AW), and neighboring neurons become more desynchronized.

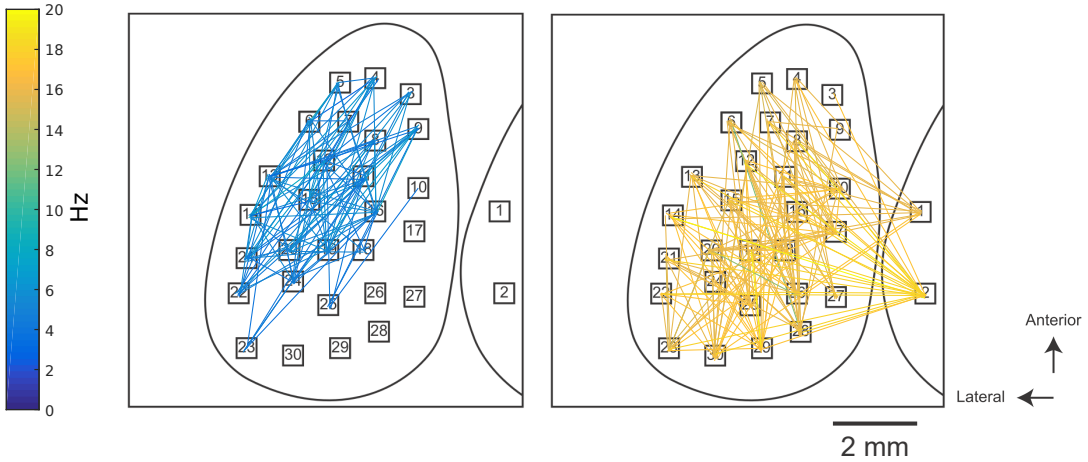


Figure 7.4: The ROI-pairs that reveal significant changes (with an α level of 0.001, corrected for multiple comparison) in the functional connectivity, as measured by wavelet transform coherence, across spontaneous whisking conditions. Left: ROI-pairs that reveal stronger synchronization in NW condition. Right: ROI-pairs that reveal stronger synchronization in AW condition.

7.4 Summary

Characterizing the dynamics of brain function under spontaneous conditions has drawn increased attention in recent years. Widefield transcranial imaging allows measuring calcium transients in neurons across large areas of the cortex with high temporal resolution. Capturing brain activity using widefield imaging in combination with the capability of recording

spontaneous whisker movement activity, provides an interesting opportunity to evaluate how functional connectivity between distributed regions of the brain vary according to spontaneous conditions.

Wavelet coherence is an important computational tool for measuring similarity of time series, and has been used as a measure of functional connectivity in several neuroscience studies [233, 15, 251]. One standard practice in frequency-based approaches is to predefine frequency bands that are known to be specifically sensitive to the experimental paradigm. However, in many scenarios, this prior knowledge is not known. This problem motivated the development of the framework presented in this chapter, in which several frequency bins can be considered. Performing large number of statistical tests on multiple frequency bins raises the necessity for running multiple comparisons. In this study, we adapted the multivariate permutation test, which strongly controls the family-wise error rate, and does not rely on the distribution of the underlying structure of sample data [252, 255]. As demonstrated in Section 7.3, by combining wavelet transform coherence and the multivariate permutation test, NW-specified functional connectivity was detected in lower frequency band, while AW-specified functional connectivity was observed in higher frequency bands.

The advantages of the presented framework include *i)* by considering all frequency bands, there is no need to have a prior knowledge about the frequency bands that are involved in the activity, and *ii)* the proposed approach has no requirement on the sample distribution and its parameters. It should be noted that although we demonstrated the presented approach on spontaneous data obtained through widefield imaging, it can be also extended to studies with task-based experimental paradigms and other imaging modalities, for instance, EEG, fNIRS, and fMRI with applications in brain-computer interfaces, as well as, for identifying the mechanisms underlying human brain-related diseases.

Chapter 8

Conclusions and Future Work

The theme of this dissertation is to present novel computation solutions for challenges on data processing and analytics of functional brain imaging data for predicting behavior/brain state. The first part of the dissertation focused on data processing and proposed an advanced DTW-based averaging technique which considered the temporal non-linear variations in the alignment of recorded fNIRS signals. The second part of the dissertation focused on data analytics in which various computational methods, based on signal processing and machine learning techniques, were presented, for capturing temporal and spectral characteristics of the recordings and building predictive models.

8.1 Contributions

The contributions of this work are summarized as follows:

- A new averaging strategy for improving the detection power of fNIRS experiments based on the dynamic time warping was presented. By taking consideration of the non-linear alignment across the repeated fNIRS recordings, it was shown that the proposed method outperforms the conventional point-by-point averaging technique in terms of the averaging accuracy.

- A data-driven framework for inferring behavior/brain states based on visibility graph was proposed, in which the temporal characteristics of brain recordings were quantified and utilized. It was demonstrated that the temporal dynamics of calcium imaging and fNIRS recordings carry discriminatory information and are powerful enough for inferring behavior and brain state.

- Two novel analysis frameworks combining visibility graph and neural networks were

presented, towards investigating relationship between the brain activity and behavior. These frameworks open new window to utilize the knowledge from both well-developed network science and neural networks techniques, to investigate the temporal dynamics of time series.

- A novel computational framework for studying the changes in the functional connectivity of calcium imaging was proposed by combining wavelet transform coherence and multivariate permutation test. The proposed method does not impose any prior assumption about the frequency bands that are involved in the activity, nor on the distribution of the data.

8.2 Future Work

This research establishes several interesting research directions for future work:

- With the focus on data acquisition stage, one can investigate the optimal sensors configurations for the EEG-fNIRS multimodal experiments in terms of capturing the brain activity originated from the same location from different modalities.

- Explore the spatio-temporal characteristics of the brain imaging data using multilayer visibility graph.

- Develop quantitative statistical measures and iterative procedures for DTW-based averaging to obtain an optimum choice for a reference signal that maximizes the accuracy of the averaged signal.

- Extend the methods combining VG and NN to predict behavior from neuroimaging data by using recurrence neural networks and the transfer learning techniques.

References

[1] L. Zhu, C. R. Lee, D. J. Margolis, and L. Najafizadeh, "Decoding cortical brain states from widefield calcium imaging data using visibility graph," *Biomedical Optics Express*, vol. 9, no. 7, pp. 3017–3036, 2018.

[2] "The cost of dementia," https://ind.ucsf.edu/supporting-our-work/cost-dementia, accessed: 2018-06-23.

[3] "Neruotechnology report," https://sharpbrains.com/blog/2015/05/06/first-ever-pervasive-neurotechnology-report-finds-10000-patent-filings-transforming-medicine-entertainment-and-business/, accessed: 2018-06-23.

[4] B. He and Z. Liu, "Multimodal functional neuroimaging: integrating functional mri and eeg/meg," *Biomedical Engineering, IEEE Reviews in*, vol. 1, pp. 23–40, 2008.

[5] N. Naseer and K.-S. Hong, "fNIRS-based brain-computer interfaces: a review," *Frontiers in Human Neuroscience*, vol. 9, p. 3, 2015.

[6] T. Giorgino, "Computing and visualizing dynamic time warping alignments in R: the dtw package," *Journal of statistical Software*, vol. 31, no. 7, pp. 1–24, 2009.

[7] M. Xia, J. Wang, and Y. He, "BrainNet Viewer: a network visualization tool for human brain connectomics," *PLoS One*, vol. 8, no. 7, p. e68910, 2013.

[8] S. Musall, M. T. Kaufman, S. Gluf, and A. Churchland, "Movement-related activity dominates cortex during sensory-guided decision making," *bioRxiv*, p. 308288, 2018, doi: 10.1101/308288.

[9] S. Herculano-Houzel, "The human brain in numbers: a linearly scaled-up primate brain," *Frontiers in human neuroscience*, vol. 3, 2009.

[10] L. Zhu, A. E. Haddad, T. Zeng, Y. Wang, and L. Najafizadeh, "Assessing Optimal Electrode/Optode Arrangement in EEG-fNIRS Multi-Modal Imaging," in *Clinical and Translational Biophotonics*. Optical Society of America, 2016, pp. JW3A–39.

[11] L. Zhu, M. Peifer, and L. Najafizadeh, "Towards Improving the "Detection" Power of Brain Imaging Experiments Using fNIRS," in *Biomedical Optics Meeting*. Optical Society of America, 2014, pp. BM3A–29.

[12] L. Zhu and L. Najafizadeh, "Dynamic time warping-based averaging framework for functional near-infrared spectroscopy brain imaging studies," *Journal of biomedical optics*, vol. 22, no. 6, p. 066011, 2017.

[13] L. Zhu, C. R. Lee, D. J. Margolis, and L. Najafizadeh, "Predicting Behavior from Cortical Activity Recorded through Widefield Transcranial Imaging," in *CLEO: Applications and Technology*. Optical Society of America, 2017, pp. ATu3B–1.

[14] L. Zhu and L. Najafizadeh, "Temporal Dynamics of fNIRS-Recorded Signals Revealed Via Visibility Graph," in *Cancer Imaging and Therapy*. Optical Society of America, 2016, pp. JW3A–53.

[15] ——, "Does brain functional connectivity alter across similar trials during imaging experiments?" in *Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE, 2014, pp. 1–4.

[16] L. Zhu, C. R. Lee, D. J. Margolis, and L. Najafizadeh, "Probing the dynamics of spontaneous cortical activities via widefield ca+ 2 imaging in gcamp6 transgenic mice," in *Wavelets and Sparsity XVII*, vol. 10394. International Society for Optics and Photonics, 2017, p. 103940C.

[17] M. Peifer, "Decoding brain states using functional brain imaging techniques," Ph.D. dissertation, Rutgers University-Graduate School-New Brunswick, 2015.

[18] W. Ou, "Spatio-temporal analysis in functional brain imaging," Ph.D. dissertation, Citeseer, 2010.

[19] J. Jorge, W. Van Der Zwaag, and P. Figueiredo, "EEG–fMRI integration for the study of human brain function," *NeuroImage*, vol. 102, pp. 24–34, 2014.

[20] P. J. Allen, O. Josephs, and R. Turner, "A method for removing imaging artifact from continuous EEG recorded during functional MRI," *Neuroimage*, vol. 12, no. 2, pp. 230–239, 2000.

[21] M. T. Talukdar, H. R. Frost, and S. G. Diamond, "Modeling Neurovascular Coupling from Clustered Parameter Sets for Multimodal EEG-NIRS," *Computational and Mathematical Methods in Medicine*, 2015.

[22] A. Dutta, A. Jacob, S. R. Chowdhury, A. Das, and M. A. Nitsche, "EEG-NIRS based assessment of neurovascular coupling during anodal transcranial direct current stimulation-a stroke case series," *Journal of medical systems*, vol. 39, no. 4, pp. 1–9, 2015.

[23] H. Girouard and C. Iadecola, "Neurovascular coupling in the normal brain and in hypertension, stroke, and Alzheimer disease," *Journal of Applied Physiology*, vol. 100, no. 1, pp. 328–335, 2006.

[24] S. Fazli, J. Mehnert, J. Steinbrink, G. Curio, A. Villringer, K.-R. Müller, and B. Blankertz, "Enhanced performance by a hybrid NIRS–EEG brain computer interface," *Neuroimage*, vol. 59, no. 1, pp. 519–529, 2012.

[25] W. Ou, A. Nummenmaa, J. Ahveninen, J. W. Belliveau, M. S. Hämäläinen, and P. Golland, "Multimodal functional imaging using fMRI-informed regional EEG/MEG source estimation," *Neuroimage*, vol. 52, no. 1, pp. 97–108, 2010.

[26] L. Meyer, J. Obleser, S. J. Kiebel, and A. D. Friederici, "Spatiotemporal dynamics of argument retrieval and reordering: an fMRI and EEG study on sentence processing," *Frontiers in psychology*, vol. 3, 2012.

[27] M. Brass, M. Ullsperger, T. R. Knoesche, D. Y. Von Cramon, and N. A. Phillips, "Who comes first? The role of the prefrontal and parietal cortex in cognitive control," *Journal of cognitive neuroscience*, vol. 17, no. 9, pp. 1367–1375, 2005.

[28] C. Gerloff, K. Bushara, A. Sailer, E. M. Wassermann, R. Chen, T. Matsuoka, D. Waldvogel, G. F. Wittenberg, K. Ishii, L. G. Cohen *et al.*, "Multimodal imaging of brain reorganization in motor areas of the contralesional hemisphere of well recovered patients after capsular stroke," *Brain*, vol. 129, no. 3, pp. 791–808, 2006.

[29] E. M. Hillman, "Optical brain imaging in vivo: techniques and applications from animal to man," *Journal of biomedical optics*, vol. 12, no. 5, p. 051402, 2007.

[30] P. A. Jackson and D. O. Kennedy, "The application of near infrared spectroscopy in nutritional intervention studies," *Frontiers in human neuroscience*, vol. 7, p. 473, 2013.

[31] F. Scholkmann, S. Kleiser, A. J. Metz, R. Zimmermann, J. M. Pavia, U. Wolf, and M. Wolf, "A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology," *Neuroimage*, vol. 85, pp. 6–27, 2014.

[32] M. Peifer, L. Zhu, and L. Najafizadeh, "Real-time classification of actual vs imagery finger tapping using functional near-infrared spectroscopy," in *Biomedical Optics*. Optical Society of America, 2014, pp. BM3A–34.

[33] L. Kocsis, P. Herman, and A. Eke, "The modified Beer–Lambert law revisited," *Physics in medicine and biology*, vol. 51, no. 5, p. N91, 2006.

[34] S. Tak and J. C. Ye, "Statistical analysis of fnirs data: A comprehensive review," *NeuroImage*, vol. 85, pp. 72–91, 2014.

[35] L. Ji, J. Zhou, R. Zafar, S. Kantorovich, R. Jiang, P. R. Carney, and H. Jiang, "Cortical neurovascular coupling driven by stimulation of channelrhodopsin-2," *PloS one*, vol. 7, no. 9, p. e46607, 2012.

[36] X. Hu and E. Yacoub, "The story of the initial dip in fmri," *Neuroimage*, vol. 62, no. 2, pp. 1103–1108, 2012.

[37] Y. Ma, M. A. Shaik, S. H. Kim, M. G. Kozberg, D. N. Thibodeaux, H. T. Zhao, H. Yu, and E. M. Hillman, "Wide-field optical mapping of neural activity and brain haemodynamics: considerations and novel approaches," *Phil. Trans. R. Soc. B*, vol. 371, no. 1705, p. 20150360, 2016.

[38] L. Madisen, A. R. Garner, D. Shimaoka, A. S. Chuong, N. C. Klapoetke, L. Li, A. van der Bourg, Y. Niino, L. Egolf, C. Monetti, H. Gu, M. Mills, A. Cheng, B. Tasic, T. N. Nguyen, S. M. Sunkin, A. Benucci, A. Nagy, A. Miyawaki, F. Helmchen, R. M. Empson, T. Knopfel, E. S. Boyden, R. C. Reid, M. Carandini, and H. Zeng, "Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance," *Neuron*, vol. 85, no. 5, pp. 942–58, 2015.

[39] T. W. Chen, T. J. Wardill, Y. Sun, S. R. Pulver, S. L. Renninger, A. Baohan, E. R. Schreiter, R. A. Kerr, M. B. Orger, V. Jayaraman, L. L. Looger, K. Svoboda, and D. S. Kim, "Ultrasensitive fluorescent proteins for imaging neuronal activity," *Nature*, vol. 499, no. 7458, pp. 295–300, 2013.

[40] N. A. Steinmetz, C. Buetfering, J. Lecoq, C. R. Lee, A. J. Peters, E. A. Jacobs, P. Coen, D. R. Ollerenshaw, M. T. Valley, S. E. de Vries *et al.*, "Aberrant cortical activity in multiple GCaMP6-expressing transgenic mouse lines," *eNeuro*, vol. 4(5), pp. ENEURO.0207–17, Sep 2017.

[41] M. P. Vanni and T. H. Murphy, "Mesoscale transcranial spontaneous activity mapping in GCaMP3 transgenic mice reveals extensive reciprocal connections between areas of somatomotor cortex," *Journal of Neuroscience*, vol. 34, no. 48, pp. 15 931–15 946, 2014.

[42] T. H. Kim, Y. Zhang, J. Lecoq, J. C. Jung, J. Li, H. Zeng, C. M. Niell, and M. J. Schnitzer, "Long-term optical access to an estimated One million neurons in the live mouse cortex," *Cell Rep*, vol. 17, no. 12, pp. 3385–3394, 2016.

[43] D. Xiao, M. P. Vanni, C. C. Mitelut, A. W. Chan, J. M. LeDue, Y. Xie, A. C. Chen, N. V. Swindale, and T. H. Murphy, "Mapping cortical mesoscopic networks of single spiking cortical or sub-cortical neurons," *Elife*, vol. 6, p. e19976, 2017.

[44] W. Denk and K. Svoboda, "Photon upmanship: why multiphoton imaging is more than a gimmick," *Neuron*, vol. 18, no. 3, pp. 351–357, 1997.

[45] F. Helmchen and W. Denk, "Deep tissue two-photon microscopy," *Nature methods*, vol. 2, no. 12, pp. 932–940, 2005.

[46] A. Grinvald, D. Omer, S. Naaman, and D. Sharon, "Imaging the dynamics of mammalian neocortical population activity in-vivo," in *Membrane Potential Imaging in the Nervous System and Heart*, 2015, pp. 243–271.

[47] B. A. Wilt, L. D. Burns, E. T. Wei Ho, K. K. Ghosh, E. A. Mukamel, and M. J. Schnitzer, "Advances in light microscopy for neuroscience," *Annual review of neuroscience*, vol. 32, pp. 435–506, 2009.

[48] M. L. Andermann, A. M. Kerlin, and R. Reid, "Chronic cellular imaging of mouse visual cortex during operant behavior and passive viewing," *Frontiers in cellular neuroscience*, vol. 4, 2010.

[49] J. L. Chen, D. J. Margolis, A. Stankov, L. T. Sumanovski, B. L. Schneider, and F. Helmchen, "Pathway-specific reorganization of projection neurons in somatosensory cortex during learning," *Nature Neuroscience*, vol. 18, no. 8, pp. 1101–1108, 2015.

[50] M. Minderer, W. R. Liu, L. T. Sumanovski, S. Kugler, F. Helmchen, and D. J. Margolis, "Chronic imaging of cortical sensory map dynamics using a genetically encoded calcium indicator," *Journal of Physiology-London*, vol. 590, no. 1, pp. 99–107, 2012.

[51] G. Silasi, D. Xiao, M. P. Vanni, A. C. Chen, and T. H. Murphy, "Intact skull chronic windows for mesoscopic wide-field imaging in awake mice," *Journal of neuroscience methods*, vol. 267, pp. 141–149, 2016.

[52] T. Murakami, T. Yoshida, T. Matsui, and K. Ohki, "Wide-field Ca2+ imaging reveals visually evoked activity in the retrosplenial area," *Frontiers in molecular neuroscience*, vol. 8, 2015.

[53] M. J. McGinley, M. Vinck, J. Reimer, R. Batista-Brito, E. Zagha, C. R. Cadwell, A. S. Tolias, J. A. Cardin, and D. A. McCormick, "Waking state: rapid variations modulate neural and behavioral responses," *Neuron*, vol. 87, no. 6, pp. 1143–61, 2015.

[54] M. Vinck, R. Batista-Brito, U. Knoblich, and J. A. Cardin, "Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding," *Neuron*, vol. 86, no. 3, pp. 740–754, 2015.

[55] J. Reimer, E. Froudarakis, C. R. Cadwell, D. Yatsenko, G. H. Denfield, and A. S. Tolias, "Pupil fluctuations track fast switching of cortical states during quiet wakefulness," *Neuron*, vol. 84, no. 2, pp. 355–62, 2014.

[56] M. J. McGinley, S. V. David, and D. A. McCormick, "Cortical membrane potential signature of optimal states for sensory signal detection," *Neuron*, vol. 87, pp. 179–92, 2015.

[57] D. Shimaoka, K. D. Harris, and M. Carandini, "Effects of Arousal on Mouse Sensory Cortex Depend on Modality," *Cell reports*, vol. 22, no. 12, pp. 3160–3167, 2018.

[58] H. Hallez, B. Vanrumste, R. Grech, J. Muscat, W. De Clercq, A. Vergult, Y. D'Asseler, K. P. Camilleri, S. G. Fabri, S. Van Huffel *et al.*, "Review on solving the forward problem in eeg source analysis," *Journal of neuroengineering and rehabilitation*, vol. 4, no. 1, p. 46, 2007.

[59] O. Hauk, A. Keil, T. Elbert, and M. M. Müller, "Comparison of data transformation procedures to enhance topographical accuracy in time-series analysis of the human EEG," *Journal of neuroscience methods*, vol. 113, no. 2, pp. 111–122, 2002.

[60] S. Dähne, F. Bießmann, W. Samek, S. Haufe, D. Goltz, C. Gundlach, A. Villringer, S. Fazli, and K.-R. Müller, "Multivariate machine learning methods for fusing multimodal functional neuroimaging data," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1507–1530, 2015.

[61] S. E. Petersen and J. W. Dubis, "The mixed block/event-related design," *Neuroimage*, vol. 62, no. 2, pp. 1177–1184, 2012.

[62] L. Pollonini, C. Olds, H. Abaya, H. Bortfeld, M. S. Beauchamp, and J. S. Oghalai, "Auditory cortex activation to natural speech and simulated cochlear implant speech measured with functional near-infrared spectroscopy," *Hearing research*, vol. 309, pp. 84–93, 2014.

[63] T. Wilcox, H. Bortfeld, R. Woods, E. Wruck, J. Armstrong, and D. Boas, "Hemodynamic changes in the infant cortex during the processing of featural and spatiotemporal information," *Neuropsychologia*, vol. 47, no. 3, pp. 657–662, 2009.

[64] N. Roche-Labarbe, F. Wallois, E. Ponchel, G. Kongolo, and R. Grebe, "Coupled oxygenation oscillation measured by NIRS and intermittent cerebral activation on EEG in premature infants," *Neuroimage*, vol. 36, no. 3, pp. 718–727, 2007.

[65] J. Harrison, K. İzzetoğlu, H. Ayaz, B. Willems, S. Hah, U. Ahlstrom, H. Woo, P. A. Shewokis, S. C. Bunce, and B. Onaral, "Cognitive workload and learning assessment during the implementation of a next-generation air traffic control technology using functional

near-infrared spectroscopy," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 4, pp. 429–440, 2014.

[66] H. Ayaz, B. Willems, B. Bunce, P. A. Shewokis, K. Izzetoglu, S. Hah, A. Deshmukh, and B. Onaral, "Cognitive workload assessment of air traffic controllers using optical brain imaging sensors," *Advances in understanding human performance: Neuroergonomics, human factors design, and special populations*, pp. 21–31, 2010.

[67] R. McKendrick, H. Ayaz, R. Olmstead, and R. Parasuraman, "Enhancing dual-task performance with verbal and spatial working memory training: continuous monitoring of cerebral hemodynamics with NIRS," *NeuroImage*, vol. 85, pp. 1014–1026, 2014.

[68] M. N. Kim, T. Durduran, S. Frangos, B. L. Edlow, E. M. Buckley, H. E. Moss, C. Zhou, G. Yu, R. Choe, E. Maloney-Wilensky *et al.*, "Noninvasive measurement of cerebral blood flow and blood oxygenation using near-infrared and diffuse correlation spectroscopies in critically brain-injured adults," *Neurocritical care*, vol. 12, no. 2, pp. 173–180, 2010.

[69] L. M. Ward, G. Morison, W. A. Simpson, A. J. Simmers, and U. Shahani, "Using Functional Near Infrared Spectroscopy (fNIRS) to Study Dynamic Stereoscopic Depth Perception," *Brain Topography*, pp. 1–9, 2016.

[70] S. Urakawa, K. Takamoto, A. Ishikawa, T. Ono, and H. Nishijo, "Selective medial prefrontal cortex responses during live mutual gaze interactions in human infants: an fNIRS study," *Brain Topography*, vol. 28, no. 5, pp. 691–701, 2015.

[71] J. Zhang, X. Lin, G. Fu, L. Sai, H. Chen, J. Yang, M. Wang, Q. Liu, G. Yang, J. Zhang *et al.*, "Mapping the small-world properties of brain networks in deception with functional near-infrared spectroscopy," *scientific Reports*, vol. 6, 2016.

[72] G. K. Aguirre, E. Zarahn, and M. D'esposito, "The variability of human, bold hemodynamic responses," *NeuroImage*, vol. 8, no. 4, pp. 360–369, 1998.

[73] L. Gupta, D. L. Molfese, R. Tammana, and P. G. Simos, "Nonlinear alignment and averaging for estimating the evoked potential," *Biomedical Engineering, IEEE Transactions on*, vol. 43, no. 4, pp. 348–356, 1996.

[74] R. Henson, C. J. Price, M. Rugg, R. Turner, and K. J. Friston, "Detecting latency differences in event-related bold responses: application to words versus nonwords and initial versus repeated face presentations," *NeuroImage*, vol. 15, no. 1, pp. 83–97, 2002.

[75] L. Holper, N. Kobashi, D. Kiper, F. Scholkmann, M. Wolf, and K. Eng, "Trial-to-trial variability differentiates motor imagery during observation between low versus high responders: A functional near-infrared spectroscopy study," *Behavioural brain research*, vol. 229, no. 1, pp. 29–40, 2012.

[76] J. McPartland, G. Dawson, S. J. Webb, H. Panagiotides, and L. J. Carver, "Event-related brain potentials reveal anomalies in temporal processing of faces in autism spectrum disorder," *Journal of Child Psychology and Psychiatry*, vol. 45, no. 7, pp. 1235–1245, 2004.

[77] L. M. Oberman, P. Winkielman, and V. S. Ramachandran, "Slow echo: facial EMG evidence for the delay of spontaneous, but not voluntary, emotional mimicry in children with autism spectrum disorders," *Developmental Science*, vol. 12, no. 4, pp. 510–520, 2009.

[78] T. Zeng, L. Zhu, Y. Wang, and L. Najafizadeh, "On the Relationship Between Trial-to-Trial Response Time Variability and fNIRS-Based Functional Connectivity," in *Optical Tomography and Spectroscopy*. Optical Society of America, 2016, pp. JW3A–41.

[79] L. R. Bahl, F. Jelinek, and R. Mercer, "A maximum likelihood approach to continuous speech recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 2, pp. 179–190, 1983.

[80] Z. M. Kovács-Vajna, "A fingerprint verification system based on triangular matching and dynamic time warping," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 11, pp. 1266–1276, 2000.

[81] E. J. Keogh and M. J. Pazzani, "Scaling up dynamic time warping for datamining applications," in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2000, pp. 285–289.

[82] J. Aach and G. M. Church, "Aligning gene expression time series with time warping algorithms," *Bioinformatics*, vol. 17, no. 6, pp. 495–508, 2001.

[83] A. Corradini, "Dynamic time warping for off-line recognition of a small gesture vocabulary," in *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 2001. Proceedings. IEEE ICCV Workshop on*. IEEE, 2001, pp. 82–89.

[84] N. Karamzadeh, A. Medvedev, A. Azari, A. Gandjbakhche, and L. Najafizadeh, "Capturing dynamic patterns of task-based functional connectivity with EEG," *NeuroImage*, vol. 66, pp. 311–317, 2013.

[85] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognition*, vol. 44, no. 3, pp. 678–693, 2011.

[86] A. M. Owen, K. M. McMillan, A. R. Laird, and E. Bullmore, "N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies," *human Brain Mapping*, vol. 25, no. 1, pp. 46–59, 2005.

[87] L. Wang, Y. Kuroiwa, T. Kamitani, T. Takahashi, Y. Suzuki, and O. Hasegawa, "Effect of interstimulus interval on visual P300 in ParkinsonâĂŹs disease," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 67, no. 4, pp. 497–503, 1999.

[88] T. Ozawa, T. Aihara, Y. Fujiwara, Y. Otaka, I. Nambu, R. Osu, J. Izawa, and Y. Wada, "Detecting event-related motor activity using functional near-infrared spectroscopy," in *6th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2013, pp. 1529–1532.

[89] A. Bozkurt, A. Rosen, H. Rosen, and B. Onaral, "A portable near infrared spectroscopy system for bedside monitoring of newborn brain," *Biomedical engineering online*, vol. 4, no. 1, p. 29, 2005.

[90] G. Taga, H. Watanabe, and F. Homae, "Spatiotemporal properties of cortical haemodynamic response to auditory stimuli in sleeping infants revealed by multi-channel near-infrared spectroscopy," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 369, no. 1955, pp. 4495–4511, 2011.

[91] H. Watanabe, F. Homae, and G. Taga, "General to specific development of functional activation in the cerebral cortexes of 2-to 3-month-old infants," *NeuroImage*, vol. 50, no. 4, pp. 1536–1544, 2010.

[92] G. Ball, P. R. Stokes, R. A. Rhodes, S. K. Bose, I. Rezek, A.-M. Wink, L.-D. Lord, M. A. Mehta, P. M. Grasby, and F. E. Turkheimer, "Executive functions and prefrontal cortex: a matter of persistence?" *Frontiers in systems neuroscience*, vol. 5, p. 3, 2011.

[93] J. H. Callicott, V. S. Mattay, A. Bertolino, K. Finn, R. Coppola, J. A. Frank, T. E. Goldberg, and D. R. Weinberger, "Physiological characteristics of capacity constraints in working memory as revealed by functional MRI," *Cerebral Cortex*, vol. 9, no. 1, pp. 20–26, 1999.

[94] X. Cui, S. Bray, D. M. Bryant, G. H. Glover, and A. L. Reiss, "A quantitative comparison of NIRS and fMRI across multiple cognitive tasks," *NeuroImage*, vol. 54, no. 4, pp. 2808–2821, 2011.

[95] Y. Zhang, D. H. Brooks, M. A. Franceschini, and D. A. Boas, "Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging," *Journal of biomedical optics*, vol. 10, no. 1, pp. 011 014–01 101 411, 2005.

[96] E. Kirino, A. Belger, P. Goldman-Rakic, and G. McCarthy, "Prefrontal activation evoked by infrequent target and novel stimuli in a visual target detection task: an event-related functional magnetic resonance imaging study," *The Journal of Neuroscience*, vol. 20, no. 17, pp. 6612–6618, 2000.

[97] H. Cho, G. A. Ojemann, D. Corina, and L. G. Shapiro, "Detection of neural activity in event-related fMRI using wavelets and dynamic time warping," in *Optical Science and Technology, SPIE's 48th Annual Meeting*. International Society for Optics and Photonics, 2003, pp. 638–647.

[98] Y. Sun, H. Ayaz, and A. N. Akansu, "Neural correlates of affective context in facial expression analysis: a simultaneous EEG-fNIRS study," in *Signal and Information Processing (GlobalSIP), 2015 IEEE Global Conference on*. IEEE, 2015, pp. 820–824.

[99] A. V. Medvedev, J. Kainerstorfer, S. V. Borisov, R. L. Barbour, and J. VanMeter, "Event-related fast optical signal in a rapid object recognition task: improving detection by the independent component analysis," *Brain research*, vol. 1236, pp. 145–158, 2008.

[100] T. Limongi, G. Di Sante, M. Ferrari, and V. Quaresima, "Detecting mental calculation related frontal cortex oxygenation changes for brain computer interface using multi-channel functional near infrared topography," *International Journal of Bioelectromagnetism*, vol. 11, no. 2, pp. 86–90, 2009.

[101] C. Herff, D. Heger, O. Fortmann, J. Hennrich, F. Putze, and T. Schultz, "Mental workload during n-back taskâĂŤquantified in the prefrontal cortex using fNIRS," *Frontiers in human neuroscience*, vol. 7, p. 935, 2014.

[102] C.-H. Han, H. Song, Y.-G. Kang, B.-M. Kim, and C.-H. Im, "Hemodynamic responses in rat brain during transcranial direct current stimulation: a functional near-infrared spectroscopy study," *Biomedical Optics Express*, vol. 5, no. 6, pp. 1812–1821, 2014.

[103] K. Yamamuro, S. Kimoto, J. Iida, N. Kishimoto, Y. Nakanishi, S. Tanaka, T. Ota, M. Makinodan, and T. Kishimoto, "Reduced prefrontal cortex hemodynamic response in adults with methamphetamine induced psychosis: Relevance for impulsivity," *PloS one*, vol. 11, no. 4, p. e0152373, 2016.

[104] L. P. van de Rijt, A. J. van Opstal, E. A. Mylanus, L. V. Straatman, H. Y. Hu, A. F. Snik, and M. M. van Wanrooij, "Temporal cortex activation to audiovisual speech in normal-hearing and cochlear implant users measured with functional near-infrared spectroscopy," *Frontiers in human neuroscience*, vol. 10, 2016.

[105] M. Carrieri, A. Petracca, S. Lancia, S. B. Moro, S. Brigadoi, M. Spezialetti, M. Ferrari, G. Placidi, and V. Quaresima, "Prefrontal cortex activation upon a demanding virtual hand-controlled task: a new frontier for neuroergonomics," *Frontiers in human neuroscience*, vol. 10, p. 53, 2016.

[106] J. F. Poulet and C. C. Petersen, "Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice," *Nature*, vol. 454, no. 7206, pp. 881–5, 2008.

[107] N. Naseer and K.-S. Hong, "Classification of functional near-infrared spectroscopy signals corresponding to the right-and left-wrist motor imagery for development of a brain–computer interface," *Neuroscience letters*, vol. 553, pp. 84–89, 2013.

[108] S. J. Bensmaia and L. E. Miller, "Restoring sensorimotor function through intracortical interfaces: progress and looming challenges," *Nature Reviews Neuroscience*, vol. 15, no. 5, pp. 313–325, 2014.

[109] D. J. O'shea, E. Trautmann, C. Chandrasekaran, S. Stavisky, J. C. Kao, M. Sahani, S. Ryu, K. Deisseroth, and K. V. Shenoy, "The need for calcium imaging in nonhuman primates: New motor neuroscience and brain-machine interfaces," *Experimental neurology*, vol. 287, pp. 437–451, 2017.

[110] N. Kanwisher, "Functional specificity in the human brain: a window into the functional architecture of the mind," *Proceedings of the National Academy of Sciences*, vol. 107, no. 25, pp. 11 163–11 170, 2010.

[111] F. Hutzler, "Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data," *Neuroimage*, vol. 84, pp. 1061–1069, 2014.

[112] J. Richiardi, H. Eryilmaz, S. Schwartz, P. Vuilleumier, and D. Van De Ville, "Decoding brain states from fMRI connectivity graphs," *Neuroimage*, vol. 56, no. 2, pp. 616–626, 2011.

[113] W. Shirer, S. Ryali, E. Rykhlevskaia, V. Menon, and M. Greicius, "Decoding subject-driven cognitive states with whole-brain connectivity patterns," *Cerebral cortex*, vol. 22, no. 1, pp. 158–165, 2012.

[114] S. Salsabilian, C. R. Lee, D. J. Margolis, and L. Najafizadeh, "Using connectivity to infer behavior from cortical activity recorded through widefield transcranial imaging," in *Optics and the Brain*, 2018, pp. BTu2C–4.

[115] B. Blankertz, C. Sannelli, S. Halder, E. M. Hammer, A. Kübler, K.-R. Müller, G. Curio, and T. Dickhaus, "Neurophysiological predictor of SMR-based BCI performance," *Neuroimage*, vol. 51, no. 4, pp. 1303–1309, 2010.

[116] R. A. Poldrack, "Can cognitive processes be inferred from neuroimaging data?" *Trends in cognitive sciences*, vol. 10, no. 2, pp. 59–63, 2006.

[117] E. Eggermann, Y. Kremer, S. Crochet, and C. C. Petersen, "Cholinergic signals in mouse barrel cortex during active whisker sensing," *Cell reports*, vol. 9, no. 5, pp. 1654–1660, 2014.

[118] K. D. Harris and A. Thiele, "Cortical state and attention," *Nature reviews neuroscience*, vol. 12, no. 9, p. 509, 2011.

[119] I. Ferezou, S. Bolea, and C. C. Petersen, "Visualizing the cortical representation of whisker touch: voltage-sensitive dye imaging in freely moving mice," *Neuron*, vol. 50, no. 4, pp. 617–629, 2006.

[120] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuno, "From time series to complex networks: the visibility graph," *Proc. Natl. Acad. Sci.*, vol. 105, no. 13, pp. 4972–5, 2008.

[121] C. R. Lee and D. J. Margolis, "Pupil dynamics reflect behavioral choice and learning in a go/nogo tactile decision-making task in mice," *Frontiers in behavioral neuroscience*, vol. 10, 2016.

[122] P. M. Knutsen, D. Derdikman, and E. Ahissar, "Tracking whisker and head movements in unrestrained behaving rodents," *J. Neurophysiology*, vol. 93, no. 4, pp. 2294–301, 2005.

[123] W. E. Allen, I. V. Kauvar, M. Z. Chen, E. B. Richman, S. J. Yang, K. Chan, V. Gradinaru, B. E. Deverman, L. Luo, and K. Deisseroth, "Global Representations of Goal-Directed Behavior in Distinct Cell Types of Mouse Neocortex," *Neuron*, vol. 94, no. 4, pp. 891–907, 2017.

[124] B. Luque, L. Lacasa, F. J. Ballesteros, and A. Robledo, "Analytical properties of horizontal visibility graphs in the Feigenbaum scenario," *Chaos*, vol. 22, no. 1, p. 013109, 2012.

[125] L. Lacasa, V. Nicosia, and V. Latora, "Network structure of multivariate time series," *Scientific reports*, vol. 5, p. 15508, 2015.

[126] M. Stephen, C. Gu, and H. Yang, "Visibility graph based time series analysis," *PloS one*, vol. 10, no. 11, p. e0143015, 2015.

[127] G. Zhu, Y. Li, and P. Wen, "Analysing epileptic EEGs with a visibility graph algorithm," in *IEEE Int. Conf. on Biomedical Engineering and Informatics (BMEI)*, 2012, pp. 432–436.

[128] C. Hao, Z. Chen, and Z. Zhao, "Analysis and prediction of epilepsy based on visibility graph," in *IEEE Int. Conf. on Information Science and Control Engineering (ICISCE)*, 2016, pp. 1271–1274.

[129] Z. K. Gao, Q. Cai, Y. X. Yang, W. D. Dang, and S. S. Zhang, "Multiscale limited penetrable horizontal visibility graph for analyzing nonlinear time series," *Sci Rep*, vol. 6, p. 35622, 2016.

[130] J. Wang, C. Yang, R. Wang, H. Yu, Y. Cao, and J. Liu, "Functional brain networks in AlzheimerâĂŹs disease: EEG analysis based on limited penetrable visibility graph and phase space method," *Physica A: Statistical Mechanics and its Applications*, vol. 460, pp. 174–187, 2016.

[131] G. Zhu, Y. Li, P. P. Wen, and S. Wang, "Analysis of alcoholic EEG signals based on horizontal visibility graph entropy," *Brain Inform*, vol. 1, no. 1-4, pp. 19–25, 2014.

[132] L. Lacasa, S. Sannino, S. Stramaglia, and D. Marinazzo, "Visibility graphs for fMRI data: multiplex temporal graphs and their modulations across resting state networks," *Network Neuroscience*, 2017.

[133] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–69, 2010.

[134] R. V. Donner and J. F. Donges, "Visibility graph analysis of geophysical time series: Potentials and possible pitfalls," *Acta Geophysica*, vol. 60, no. 3, pp. 589–623, 2012.

[135] Y. Low, J. E. Gonzalez, A. Kyrola, D. Bickson, C. E. Guestrin, and J. Hellerstein, "GraphLab: A new framework for parallel machine learning," *arXiv preprint arXiv:1408.2041*, 2014.

[136] A. Subasi and E. Ercelebi, "Classification of eeg signals using neural network and logistic regression," *Computer Methods and Programs in Biomedicine*, vol. 78, no. 2, pp. 87–99, 2005.

[137] N. D. Schiff, J. T. Giacino, K. Kalmar, J. D. Victor, K. Baker, M. Gerber, B. Fritz, B. Eisenberg, T. Biondi, J. O'Connor, E. J. Kobylarz, S. Farris, A. Machado, C. McCagg, F. Plum, J. J. Fins, and A. R. Rezai, "Behavioural improvements with thalamic stimulation after severe traumatic brain injury," *Nature*, vol. 448, no. 7153, pp. 600–3, 2007.

[138] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, D. Rueckert, and I. Alzheimer's Disease Neuroimaging, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *Neuroimage*, vol. 65, pp. 167–75, 2013.

[139] C. Donos, M. Dümpelmann, and A. Schulze-Bonhage, "Early seizure detection algorithm based on intracranial EEG and random forest classification," *International journal of neural systems*, vol. 25, no. 05, p. 1550023, 2015.

[140] W. Chen, Y. Wang, G. Cao, G. Chen, and Q. Gu, "A random forest model based classification scheme for neonatal amplitude-integrated EEG," *Biomed Eng Online*, no. 2, 2014.

[141] A. Page, C. Sagedy, E. Smith, N. Attaran, T. Oates, and T. Mohsenin, "A flexible multichannel EEG feature extractor and classifier for seizure detection," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 62, no. 2, pp. 109–113, 2015.

[142] H. U. Amin, A. S. Malik, N. Kamel, and M. Hussain, "A novel approach based on data redundancy for feature extraction of EEG signals," *Brain topography*, vol. 29, no. 2, pp. 207–217, 2016.

[143] W. A. Chaovalitwongse and R. C. Sachdeo, "On the time series K-nearest neighbor classification of abnormal brain activity," *IEEE Trans. on Systems Man and Cybernetics Part a-Systems and Humans*, vol. 37, no. 6, pp. 1005–1016, 2007.

[144] F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fMRI: a tutorial overview," *NeuroImage*, vol. 45, pp. S199–209, 2009.

[145] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[146] S. Sachidhanandam, V. Sreenivasan, A. Kyriakatos, Y. Kremer, and C. C. Petersen, "Membrane potential correlates of sensory perception in mouse barrel cortex," *Nature Neuroscience*, vol. 16, no. 11, pp. 1671–7, 2013.

[147] S. J. Kayser, S. W. McNair, and C. Kayser, "Prestimulus influences on auditory perception from sensory representations and decision processes," *Proc. Natl. Acad. Sci.*, vol. 113, no. 17, pp. 4842–7, 2016.

[148] I. Carcea, M. N. Insanally, and R. C. Froemke, "Dynamics of auditory cortical activity during behavioural engagement and auditory perception," *Nature Communications*, vol. 8, p. 14412, 2017.

[149] A. Kyriakatos, V. Sadashivaiah, Y. Zhang, A. Motta, M. Auffret, and C. C. Petersen, "Voltage-sensitive dye imaging of mouse neocortex during a whisker detection task," *Neurophotonics*, vol. 4, no. 3, p. 031204, 2017.

[150] J. Reimer, M. J. McGinley, Y. Liu, C. Rodenkirch, Q. Wang, D. A. McCormick, and A. S. Tolias, "Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex," *Nature communications*, vol. 7, p. 13289, 2016.

[151] H. Dana, T.-W. Chen, A. Hu, B. C. Shields, C. Guo, L. L. Looger, D. S. Kim, and K. Svoboda, "Thy1-GCaMP6 transgenic mice for neuronal population imaging in vivo," *PloS one*, vol. 9, no. 9, p. e108697, 2014.

[152] J. B. Wekselblatt, E. D. Flister, D. M. Piscopo, and C. M. Niell, "Large-scale imaging of cortical dynamics during sensory perception and behavior," *Journal of neurophysiology*, vol. 115, no. 6, pp. 2852–2866, 2016.

[153] H. B. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[154] J. Oñativia, S. R. Schultz, and P. L. Dragotti, "A finite rate of innovation algorithm for fast and accurate spike detection from two-photon calcium imaging," *Journal of neural engineering*, vol. 10, no. 4, p. 046017, 2013.

[155] T. P. Patel, K. Man, B. L. Firestein, and D. F. Meaney, "Automated quantification of neuronal networks and single-cell calcium dynamics using calcium imaging," *Journal of Neuroscience Methods*, vol. 243, pp. 26–38, 2015.

[156] J. Friedrich, P. Zhou, and L. Paninski, "Fast online deconvolution of calcium imaging data," *PLoS Comput Biol*, vol. 13, no. 3, p. e1005423, 2017.

[157] I. J. Park, Y. V. Bobkov, B. W. Ache, and J. C. Principe, "Quantifying bursting neuron activity from calcium signals using blind deconvolution," *Journal of neuroscience methods*, vol. 218, no. 2, pp. 196–205, 2013.

[158] J. T. Vogelstein, A. M. Packer, T. A. Machado, T. Sippy, B. Babadi, R. Yuste, and L. Paninski, "Fast nonnegative deconvolution for spike train inference from population calcium imaging," *Journal of Neurophysiology*, vol. 104, no. 6, pp. 3691–704, 2010.

[159] E. A. Pnevmatikakis, D. Soudry, Y. Gao, T. A. Machado, J. Merel, D. Pfau, T. Reardon, Y. Mu, C. Lacefield, W. Yang *et al.*, "Simultaneous denoising, deconvolution, and demixing of calcium imaging data," *Neuron*, vol. 89, no. 2, pp. 285–299, 2016.

[160] L. Theis, P. Berens, E. Froudarakis, J. Reimer, M. R. Rosón, T. Baden, T. Euler, A. S. Tolias, and M. Bethge, "Benchmarking spike rate inference in population calcium imaging," *Neuron*, vol. 90, no. 3, pp. 471–482, 2016.

[161] E. A. Mukamel, A. Nimmerjahn, and M. J. Schnitzer, "Automated analysis of cellular signals from large-scale calcium imaging data," *Neuron*, vol. 63, no. 6, pp. 747–60, 2009.

[162] V. Sreenivasan, V. Esmaeili, T. Kiritani, K. Galan, S. Crochet, and C. C. Petersen, "Movement initiation signals in mouse whisker motor cortex," *Neuron*, vol. 92, no. 6, pp. 1368–1382, 2016.

[163] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from eeg data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, 2014.

[164] J. Zhang, X. Li, S. T. Foldes, W. Wang, J. L. Collinger, D. J. Weber, and A. Bagić, "Decoding brain states based on magnetoencephalography from prespecified cortical regions," *IEEE Trans. on Biomedical Engineering*, vol. 63, no. 1, pp. 30–42, 2016.

[165] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.

[166] L. Lacasa, B. Luque, J. Luque, and J. C. Nuno, "The visibility graph: A new method for estimating the hurst exponent of fractional brownian motion," *Epl*, vol. 86, no. 3, p. 30001, 2009.

[167] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[168] J. Tang, C. Deng, and G.-B. Huang, "Extreme learning machine for multilayer perceptron," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 4, pp. 809–821, 2016.

[169] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[170] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[171] P. Vannasing, I. Cornaggia, C. Vanasse, J. Tremblay, P. Diadori, S. Perreault, M. Lassonde, and A. Gallagher, "Potential brain language reorganization in a boy with refractory epilepsy; an fNIRS–EEG and fMRI comparison," *Epilepsy & Behavior Case Reports*, vol. 5, pp. 34–37, 2016.

[172] F. Li, H. Zhu, Q. Gao, G. Xu, X. Li, Z. Hu, and S. He, "Using functional near-infrared spectroscopy (fNIRS) to detect the prefrontal cortical responses to deception under different motivations," *Biomedical Optics Express*, vol. 6, no. 9, pp. 3503–3514, 2015.

[173] H. J. Foy, P. Runham, and P. Chapman, "Prefrontal cortex activation and young driver behaviour: a fNIRS study," *PLoS one*, vol. 11, no. 5, p. e0156512, 2016.

[174] I. Schelkanova and V. Toronov, "Independent component analysis of broadband near-infrared spectroscopy data acquired on adult human head," *Biomedical Optics Express*, vol. 3, no. 1, pp. 64–74, 2012.

[175] A. T. Eggebrecht, S. L. Ferradal, A. Robichaux-Viehoever, M. S. Hassanpour, H. Dehghani, A. Z. Snyder, T. Hershey, and J. P. Culver, "Mapping distributed brain function and networks with diffuse optical tomography," *Nature Photonics*, vol. 8, no. 6, pp. 448–454, 2014.

[176] M. Ferrari and V. Quaresima, "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application," *NeuroImage*, vol. 63, no. 2, pp. 921–935, 2012.

[177] C.-M. Lu, Y.-J. Zhang, B. B. Biswal, Y.-F. Zang, D.-L. Peng, and C.-Z. Zhu, "Use of fNIRS to assess resting state functional connectivity," *Journal of Neuroscience Methods*, vol. 186, no. 2, pp. 242–249, 2010.

[178] V. C. Kavuri, Z.-J. Lin, F. Tian, and H. Liu, "Sparsity enhanced spatial resolution and depth localization in diffuse optical tomography," *Biomedical Optics Express*, vol. 3, no. 5, pp. 943–957, 2012.

[179] X.-S. Hu, K.-S. Hong, and S. S. Ge, "Reduction of trial-to-trial variability in functional near-infrared spectroscopy signals by accounting for resting-state functional connectivity," *Journal of Biomedical Optics*, vol. 18, no. 1, pp. 017 003–017 003, 2013.

[180] C.-J. Huang, P.-H. Chou, H.-L. Wei, and C.-W. Sun, "Functional connectivity during phonemic and semantic verbal fluency test: a multichannel near infrared spectroscopy study," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 22, no. 3, pp. 43–48, 2016.

[181] Z. Einalou, K. Maghooli, S. K. Setarehdan, and A. Akin, "Functional near infrared spectroscopy for functional connectivity during stroop test via mutual information." *Advances in Bioresearch*, vol. 6, no. 1, 2015.

[182] N. Naseer and K.-S. Hong, "Determination of temporal window size for classifying the mean value of fnirs signals from motor imagery," in *Robotics, Automation and Mechatronics (RAM), 2013 6th IEEE Conference on*. IEEE, 2013, pp. 237–240.

[183] S. L. Novi, R. B. Rodrigues, and R. C. Mesquita, "Resting state connectivity patterns with near-infrared spectroscopy data of the whole head," *Biomedical Optics Express*, vol. 7, no. 7, pp. 2524–2537, 2016.

[184] E. Bullmore and O. Sporns, "The economy of brain network organization," *Nature Reviews Neuroscience*, vol. 13, no. 5, pp. 336–349, 2012.

[185] M. Imai, H. Watanabe, K. Yasui, Y. Kimura, Y. Shitara, S. Tsuchida, N. Takahashi, and G. Taga, "Functional connectivity of the cortex of term and preterm infants and infants with down's syndrome," *NeuroImage*, vol. 85, pp. 272–278, 2014.

[186] R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun, M. Corbetta, S. Della Penna, J. H. Duyn, G. H. Glover, J. Gonzalez-Castillo *et al.*, "Dynamic functional connectivity: promise, issues, and interpretations," *NeuroImage*, vol. 80, pp. 360–378, 2013.

[187] L. Deng, J. Sun, L. Cheng, and S. Tong, "Characterizing dynamic local functional connectivity in the human brain," *scientific Reports*, vol. 6, 2016.

[188] R. M. Hutchison, T. Womelsdorf, J. S. Gati, S. Everling, and R. S. Menon, "Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques," *human Brain Mapping*, vol. 34, no. 9, pp. 2154–2177, 2013.

[189] T. Q. D. Khoa and M. Nakagawa, "Recognizing brain activities by functional near-infrared spectroscope signal analysis," *Nonlinear Biomedical Physics*, vol. 2, no. 1, p. 3, 2008.

[190] T. Q. D. Khoa, N. Yuichi, and N. Masahiro, "Recognizing brain motor imagery activities by identifying chaos properties of oxy-hemoglobin dynamics time series," *Chaos, Solitons & Fractals*, vol. 42, no. 1, pp. 422–429, 2009.

[191] A. Eke, P. Herman, L. Kocsis, and L. Kozak, "Fractal characterization of complexity in temporal physiological signals," *Physiological measurement*, vol. 23, no. 1, p. R1, 2002.

[192] B. Kaulakys and M. Alaburda, "Modeling scaled processes and 1/fβ noise using nonlinear stochastic differential equations," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 02, p. P02051, 2009.

[193] J. M. Foss, A. V. Apkarian, and D. R. Chialvo, "Dynamics of pain: fractal dimension of temporal variability of spontaneous pain differentiates between pain states," *Journal of Neurophysiology*, vol. 95, no. 2, pp. 730–736, 2006.

[194] K. Linkenkaer-Hansen, V. V. Nikouline, J. M. Palva, and R. J. Ilmoniemi, "Long-range temporal correlations and scaling behavior in human brain oscillations," *Journal of Neuroscience*, vol. 21, no. 4, pp. 1370–1377, 2001.

[195] S. A. Akar, S. Kara, S. Agambayev, and V. Bilgiç, "Nonlinear analysis of eeg in major depression with fractal dimensions," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference*. IEEE, 2015, pp. 7410–7413.

[196] M. Ahmadlou, H. Adeli, and A. Adeli, "New diagnostic EEG markers of the Alzheimer's disease using visibility graph," *Journal of Neural Transmission*, vol. 117, no. 9, pp. 1099–1109, 2010.

[197] M. O. Sokunbi, V. B. Gradin, G. D. Waiter, G. G. Cameron, T. S. Ahearn, A. D. Murray, D. J. Steele, and R. T. Staff, "Nonlinear complexity analysis of brain fMRI signals in Schizophrenia," *PloS One*, vol. 9, no. 5, p. e95146, 2014.

[198] B. J. He, "Scale-free properties of the functional magnetic resonance imaging signal during rest and task," *Journal of Neuroscience*, vol. 31, no. 39, pp. 13 786–13 795, 2011.

[199] X. Lei, Y. Wang, H. Yuan, and A. Chen, "Brain scale-free properties in awake rest and NREM sleep: A simultaneous EEG/fMRI study," *Brain Topography*, vol. 28, no. 2, pp. 292–304, 2015.

[200] L. Telesca, M. Lovallo, A. Ramirez-Rojas, and L. Flores-Marquez, "Investigating the time dynamics of seismicity by using the visibility graph approach: Application to seismicity of Mexican subduction zone," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 24, pp. 6571–6577, 2013.

[201] S. Bhaduri and D. Ghosh, "Electroencephalographic data analysis with visibility graph technique for quantitative assessment of brain dysfunction," *Clinical EEG and Neuroscience*, vol. 46, no. 3, pp. 218–223, 2015.

[202] S. Jiang, C. H. Bian, X. B. Ning, and Q. L. D. Y. Ma, "Visibility graph analysis on heartbeat dynamics of meditation training," *Applied Physics Letters*, vol. 102, no. 25, p. 253702, 2013.

[203] M. Ahmadlou, H. Adeli, and A. Adeli, "Improved visibility graph fractality with application for the diagnosis of Autism Spectrum Disorder," *Physica a-Statistical Mechanics and Its Applications*, vol. 391, no. 20, pp. 4720–4726, 2012.

[204] Y. He and A. Evans, "Graph theoretical modeling of brain connectivity," *Curr Opin Neurol*, vol. 23, no. 4, pp. 341–50, 2010.

[205] I. W. Selesnick, H. L. Graber, D. S. Pfeil, and R. L. Barbour, "Simultaneous low-pass filtering and total variation denoising," *IEEE Transactions on Signal Processing*, vol. 62, no. 5, pp. 1109–1124, 2014.

[206] P. Herman, B. G. Sanganahalli, F. Hyder, and A. Eke, "Fractal analysis of spontaneous fluctuations of the bold signal in rat brain," *NeuroImage*, vol. 58, no. 4, pp. 1060–1069, 2011.

[207] M. J. Cannon, D. B. Percival, D. C. Caccia, G. M. Raymond, and J. B. Bassingthwaighte, "Evaluating scaled windowed variance methods for estimating the hurst coefficient of time series," *Physica A: Statistical Mechanics and its Applications*, vol. 241, no. 3-4, pp. 606–626, 1997.

[208] H. E. Schepers, J. H. Van Beek, and J. B. Bassingthwaighte, "Four methods to estimate the fractal dimension from self-affine signals (medical application)," *IEEE Engineering in Medicine and Biology Magazine*, vol. 11, no. 2, pp. 57–64, 1992.

[209] T. Liu, A. W. Moore, A. G. Gray, and K. Yang, "An investigation of practical approximate nearest neighbor algorithms," in *NIPS*, vol. 12, 2004.

[210] Y. Wang and L. Najafizadeh, "On the invariance of EEG-based signatures of individuality with application in biometric identification," in *IEEE Annual Int. Conf. of Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 4559–4562.

[211] S. Suthaharan, *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*. Springer, 2015, vol. 36.

[212] L. Hausfeld, G. Valente, and E. Formisano, "Multiclass fmri data decoding and visualization using supervised self-organizing maps," *NeuroImage*, vol. 96, pp. 54–66, 2014.

[213] A. Khazaee, A. Ebrahimzadeh, and A. Babajani-Feremi, "Identifying patients with AlzheimerâĂŹs disease using resting-state fMRI and graph theory," *Clinical Neurophysiology*, vol. 126, no. 11, pp. 2132–2141, 2015.

[214] Z. Li, Y. Wang, W. Quan, T. Wu, and B. Lv, "Evaluation of different classification methods for the diagnosis of schizophrenia based on functional near-infrared spectroscopy," *Journal of neuroscience methods*, vol. 241, pp. 101–110, 2015.

[215] N. Naseer, F. M. Noori, N. K. Qureshi, and K.-S. Hong, "Determining optimal feature-combination for lda classification of functional near-infrared spectroscopy signals in brain-computer interface application," *Frontiers in human neuroscience*, vol. 10, 2016.

[216] N. Naseer, N. K. Qureshi, F. M. Noori, and K.-S. Hong, "Analysis of different classification techniques for two-class functional near-infrared spectroscopy-based brain-computer interface," *Computational Intelligence and Neuroscience*, vol. 2016, 2016.

[217] B. W. McMenamin and L. Pessoa, "Discovering networks altered by potential threat ("anxiety") using quadratic discriminant analysis," *NeuroImage*, vol. 116, pp. 1–9, 2015.

[218] M. R. Bhutta, M. J. Hong, Y.-H. Kim, and K.-S. Hong, "Single-trial lie detection using a combined fnirs-polygraph system," *Frontiers in psychology*, vol. 6, p. 709, 2015.

[219] K.-S. Hong, N. Naseer, and Y.-H. Kim, "Classification of prefrontal and motor cortex signals for three-class fNIRS–BCI," *Neuroscience Letters*, vol. 587, pp. 87–92, 2015.

[220] D. J. Greene, J. A. Church, N. U. Dosenbach, A. N. Nielsen, B. Adeyemo, B. Nardos, S. E. Petersen, K. J. Black, and B. L. Schlaggar, "Multivariate pattern classification of pediatric tourette syndrome using functional connectivity mri," *Developmental science*, 2016.

[221] G. Sudre, D. Pomerleau, M. Palatucci, L. Wehbe, A. Fyshe, R. Salmelin, and T. Mitchell, "Tracking neural coding of perceptual and semantic features of concrete nouns," *NeuroImage*, vol. 62, no. 1, pp. 451–463, 2012.

[222] D. J. Simmonds, J. J. Pekar, and S. H. Mostofsky, "Meta-analysis of Go/No-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent," *Neuropsychologia*, vol. 46, no. 1, pp. 224–232, 2008.

[223] Y. Xu, H. L. Graber, R. L. Barbour, I. Blanco, P. Zirak, A. Fortuna, G. Cotta, M. Mayos, A. Mola, T. Durduran *et al.*, "nirsLAB: A Computing Environment for fNIRS Neuroimaging Data Analysis," in *Biomedical Optics*. Optical Society of America, 2014, pp. BM3A–1.

[224] A. P. Buccino, H. O. Keles, and A. Omurtag, "Hybrid EEG-fNIRS Asynchronous Brain-Computer Interface for Multiple Motor Tasks," *PloS one*, vol. 11, no. 1, p. e0146610, 2016.

[225] M. Honma, T. Soshi, Y. Kim, and K. Kuriyama, "Right prefrontal activity reflects the ability to overcome sleepiness during working memory tasks: a functional near-infrared spectroscopy study," *PLoS One*, vol. 5, no. 9, p. e12923, 2010.

[226] L. Zhu and L. Najafizadeh, "Dynamic time warping-based averaging framework for functional near-infrared spectroscopy brain imaging studies," *Journal of Biomedical Optics*, vol. 22, no. 6, p. 066011, 2017.

[227] A. M. Chan, E. Halgren, K. Marinkovic, and S. S. Cash, "Decoding word and category-specific spatiotemporal representations from MEG and EEG," *NeuroImage*, vol. 54, no. 4, pp. 3028–3039, 2011.

[228] V. Pappu and P. M. Pardalos, "High-dimensional data classification," in *Clusters, Orders, and Trees: Methods and Applications*. Springer, 2014, pp. 119–150.

[229] G. V. Trunk, "A problem of dimensionality: A simple example," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 3, pp. 306–307, 1979.

[230] T. McKenna, T. McMullen, and M. Shlesinger, "The brain as a dynamic physical system," *Neuroscience*, vol. 60, no. 3, pp. 587–605, 1994.

[231] W. Lutzenberger, H. Preissl, and F. Pulvermüller, "Fractal dimension of electroencephalographic time series and underlying brain processes," *Biological Cybernetics*, vol. 73, no. 5, pp. 477–482, 1995.

[232] P. Ciuciu, G. Varoquaux, P. Abry, S. Sadaghiani, and A. Kleinschmidt, "Scale-free and multifractal time dynamics of fMRI signals during rest and task," *Frontiers in Physiology*, vol. 3, 2012.

[233] C. Chang and G. H. Glover, "Time–frequency dynamics of resting-state brain connectivity measured with fMRI," *NeuroImage*, vol. 50, no. 1, pp. 81–98, 2010.

[234] C. Luo, W. Song, Q. Chen, Z. Zheng, K. Chen, B. Cao, J. Yang, J. Li, X. Huang, Q. Gong *et al.*, "Reduced functional connectivity in early-stage drug-naive Parkinson's disease: a resting-state fMRI study," *Neurobiology of aging*, vol. 35, no. 2, pp. 431–441, 2014.

[235] E. Sakakibara, F. Homae, S. Kawasaki, Y. Nishimura, R. Takizawa, S. Koike, A. Kinoshita, H. Sakurada, M. Yamagishi, F. Nishimura *et al.*, "Detection of resting state functional connectivity using partial correlation analysis: A study using multi-distance and whole-head probe near-infrared spectroscopy," *NeuroImage*, vol. 142, pp. 590–601, 2016.

[236] B. Keehn, J. B. Wagner, H. Tager-Flusberg, and C. A. Nelson, "Functional connectivity in the first year of life in infants at-risk for Autism: a preliminary near-infrared spectroscopy study," *Frontiers in human neuroscience*, vol. 7, 2013.

[237] J.-L. Chen, T. Ros, and J. H. Gruzelier, "Dynamic changes of ICA-derived EEG functional connectivity in the resting state," *Human brain mapping*, vol. 34, no. 4, pp. 852–868, 2013.

[238] M. Yu, A. A. Gouw, A. Hillebrand, B. M. Tijms, C. J. Stam, E. C. van Straaten, and Y. A. Pijnenburg, "Different functional connectivity and network topology in behavioral variant of frontotemporal dementia and Alzheimer's disease: an EEG study," *Neurobiology of aging*, vol. 42, pp. 150–162, 2016.

[239] A. Haddad and L. Najafizadeh, "Source-informed segmentation: Towards capturing the dynamics of brain functional networks through EEG," in *Proc. of Asilomar Conference on Signals, Systems and Computers*, 2016, pp. 1290–1294.

[240] X. Li, C. Lim, K. Li, L. Guo, and T. Liu, "Detecting brain state changes via fiber-centered functional connectivity analysis," *Neuroinformatics*, pp. 1–18, 2013.

[241] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, "Tracking whole-brain connectivity dynamics in the resting state," *Cerebral cortex*, vol. 24, no. 3, pp. 663–676, 2014.

[242] L. Holper, F. Scholkmann, and E. Seifritz, "Time-frequency dynamics of the sum of intra-and extracerebral hemodynamic functional connectivity during resting-state and respiratory challenges assessed by multimodal functional near-infrared spectroscopy," *NeuroImage*, vol. 125, p. 1174, 2016.

[243] J. Muraskin, S. Dodhia, G. Lieberman, J. O. Garcia, T. Verstynen, J. M. Vettel, J. Sherwin, and P. Sajda, "Brain dynamics of post-task resting state are influenced by expertise: Insights from baseball players," *Human brain mapping*, vol. 37, no. 12, pp. 4454–4471, 2016.

[244] J. Cabral, M. Kringelbach, and G. Deco, "Functional connectivity dynamically evolves on multiple time-scales over a static structural connectome: Models and mechanisms," *NeuroImage*, 2017.

[245] C. Fang, H. Li, and L. Ma, "EEG Brain Functional Connectivity Dynamic Evolution Model: A Study via Wavelet Coherence," in *Advances in Brain Inspired Cognitive Systems: 8th International Conference, BICS 2016, Beijing, China, November 28-30, 2016, Proceedings 8*, 2016, pp. 264–273.

[246] R. F. Rojas, X. Huang, and K. L. Ou, "Bilateral connectivity in the somatosensory region using near-infrared spectroscopy (NIRS) by wavelet coherence," in *SPIE BioPhotonics Australasia*, 2016, pp. 100 131Z–100 131Z.

[247] J. M. Baker, N. Liu, X. Cui, P. Vrticka, M. Saggar, S. H. Hosseini, and A. L. Reiss, "Sex differences in neural and behavioral signatures of cooperation revealed by fNIRS hyperscanning," *Scientific reports*, vol. 6, p. 26492, 2016.

[248] A. Klein, T. Sauer, A. Jedynak, and W. Skrandies, "Conventional and wavelet coherence applied to sensory-evoked electrical brain activity," *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 2, pp. 266–272, 2006.

[249] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.

[250] A. Grinsted, J. C. Moore, and S. Jevrejeva, "Application of the cross wavelet transform and wavelet coherence to geophysical time series," *Nonlinear processes in geophysics*, vol. 11, no. 5/6, pp. 561–566, 2004.

[251] A. Haddad and L. Najafizadeh, "Multi-scale analysis of the dynamics of brain functional connectivity using EEG," in *Proc. of IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2016, pp. 240–243.

[252] D. M. Groppe, T. P. Urbach, and M. Kutas, "Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review," *Psychophysiology*, vol. 48, no. 12, pp. 1711–1725, 2011.

[253] M. X. Cohen, *Analyzing neural time series data: theory and practice*. MIT Press, 2014.

[254] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D: Nonlinear Phenomena*, vol. 58, no. 1-4, pp. 77–94, 1992.

[255] R. C. Blair and W. Karniski, "An alternative method for significance testing of waveform difference potentials," *Psychophysiology*, vol. 30, no. 5, pp. 518–524, 1993.