Multimodal Biomedical Signal Processing for Corticomuscular Coupling Analysis

by

Xun Chen

B.Sc., University of Science and Technology of China, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Biomedical Engineering)

The University Of British Columbia
(Vancouver)

January 2014

Abstract

Corticomuscular coupling analysis using multiple data sets such as electroencephalogram (EEG) and electromyogram (EMG) signals provides a useful tool for understanding human motor control systems. A popular conventional method to assess corticomuscular coupling is the pair-wise magnitude-squared coherence (MSC). However, there are certain limitations associated with MSC, including the difficulty in robustly assessing group inference, only dealing with two types of data sets simultaneously and the biologically implausible assumption of pair-wise interactions. In this thesis, we propose several novel signal processing techniques to overcome the disadvantages of current coupling analysis methods. We propose combining partial least squares (PLS) and canonical correlation analysis (CCA) to take advantage of both techniques to ensure that the extracted components are maximally correlated across two data sets and meanwhile can well explain the information within each data set. Furthermore, we propose jointly incorporating response-relevance and statistical independence into a multi-objective optimization function, meaningfully combining the goals of independent component analysis (ICA) and PLS under the same mathematical umbrella. In addition, we extend the coupling analysis to multiple data sets by proposing a joint multimodal group analysis framework. Finally, to acquire independent components but not just uncorrelated ones, we improve the multimodal framework by exploiting the complementary property of multiset canonical correlation analysis (M-CCA) and joint ICA. Simulations show that our proposed methods can achieve superior performances than conventional approaches. We also apply the proposed methods to concurrent EEG, EMG and behavior data collected in a Parkinson's disease (PD) study. The results reveal highly correlated temporal patterns among the multimodal

signals and corresponding spatial activation patterns. In addition to the expected motor areas, the corresponding spatial activation patterns demonstrate enhanced occipital connectivity in PD subjects, consistent with previous medical findings.

Preface

This thesis is written based on a collection of manuscripts, resulting from the collaboration of several researchers. The majority of the research, including literature survey, algorithm development and implementation, numerical simulation, real data analysis and result report, are conducted by the author, with suggestions from Prof. Z. Jane Wang and Prof. Martin J. McKeown. The manuscripts are primarily drafted by the author, with helpful revisions and comments from Prof. Z. Jane Wang (papers in Chapters 2–5), Prof. Martin J. McKeown (papers in Chapters 3–5) and Prof. Rabab K. Ward (papers in Chapter 4).

Chapter 2 is based on the following manuscript:

- X. Chen, A. Liu, Z. J. Wang, H. Peng, "Modeling Corticomuscular Activity by Combining Partial Least Squares and Canonical Correlation Analysis," *Journal of Applied Mathematics*, vol. 2013, Article ID 401976, 11 pages, 2013.

Chapter 3 is based on the following manuscript:

- X. Chen, C. He, Z. J. Wang, M. J. McKeown, "An IC-PLS Framework for Corticomuscular Coupling Analysis," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 7, pp. 2022-2033, 2013.

Chapter 4 is based on the following manuscripts:

- X. Chen, X. Chen, R. K. Ward, Z. J. Wang, "A Joint Multimodal Group Analysis Framework for Modeling Corticomuscular Activity," *IEEE Transactions on Multimedia*, Special Issue on Multimodal Biomedical Imaging, vol. 15, no. 5, pp. 1049-1059, 2013.

- X. Chen, Z. J. Wang, M. J. McKeown, "A Tridirectional Statistical Model for Group Corticomuscular Activity Analysis," *Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on*, pp. 309-312, Banff, Canada, Sep. 2012.

Chapter 5 is based on the following manuscript:

- X. Chen, Z. J. Wang, M. J. McKeown, "A Three-step Method for Corticomuscular Activity Modeling with Application to Parkinson's Disease," *IEEE Transactions on Information Technology in Biomedicine*, in press, 2013, doi:10.1109/JBHI.2013.2284480.

Appendix A is based on the following manuscripts:

- X. Chen, Z. J. Wang, "Design and Implementation of a Wearable, Wireless EEG Recording System," *The 5th International Conference on Bioinformatics and Biomedical Engineering*, pp. 1-4, May 2011.
- X. Chen, Z. J. Wang, "Pattern Recognition of Number Gestures Based on A Wireless Surface EMG System," *Biomedical Signal Processing and Control*, vol. 8, no. 2, pp. 184-192, 2013.

All experimental procedures in this work were reviewed and approved by the University of British Columbia Clinical Research Ethics Board under Certificate Numbers H04-70291 and H05-70555.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ACCC autocorrelation and cross-correlation coefficients

ADC analog-to-digital converter

ANN artificial neural network

BEH behavior data

BCI brain-computer interface

BSS blind source separation

CMRR common-mode rejection ratio

CCA canonical correlation analysis

DC direct current

DRL driven right leg

ECG electrocardiogram

EEG electroencephalogram

EMG electromyogram

EOG electrooculogram

PC personal computer

fMRI functional magnetic resonance imaging

FFT fast Fourier transform

HCI human-computer interface

HOS high order statistics

ICA independent component analysis

ICR Independent component regression

ICs independent components

IVA independent vector analysis

JMSF joint multimodal statistical framework

jICA joint independent component analysis

***k*-NN** *k*-nearest neighbor

LDA linear discriminant analysis

LVs latent variables

LV latent variable

M1 primary motor cortex

MAV mean absolute value

MAVR mean absolute value ratio

MAVS mean absolute value slope

M-CCA multiset canonical correlation analysis

MCU microcontroller

MEG magnetoencephalogram

MSC magnitude-squared coherence

MKL multiple kernel learning

MKL-SVM multiple kernel learning support vector machine

MVC maximum voluntary contraction

op-amp operational amplifier

OSC orthogonal signal correction

PCA principal component analysis

PCs principal components

PCB printed circuit board

PD Parkinson's disease

PLS partial least squares

PPG photoplethysmograph

QDA quadratic discriminant analysis

RBF radial basis function

sEMG surface EMG

SPM spectral power magnitudes

SSC slope sign changes

STFT short-time Fourier transform

subLVs sub-latent variables

supLV super latent variable

SVM support vector machine

TD Hudgins' time-domain

UPDRS Unified Parkinson's Disease Rating Scale

USB Universal Serial Bus

WL waveform length

WT wavelet transform

ZC zero crossings

Acknowledgments

First and foremost, I would like to thank my PhD supervisors, Prof. Z. Jane Wang and Prof. Martin J. McKeown, for being great mentors, both professionally and personally. This thesis would never have been written without their patient guidance and insightful suggestions. I am in particular indebted to them for supporting me over four years and giving me sufficient freedom to explore interesting research topics, so that I could finally grow up as an independent researcher.

I would like to thank my thesis examination committee members for their time and efforts. I would also like to thank Prof. Rabab K. Ward and Prof. Victor C.M. Leung for their care and support for my research.

I would like to thank my labmates, friends and colleagues at UBC, with whom I have had a memorable campus life over the years. I would also like to express special thanks again to my senior labmates, including Junning Li, Xiaohui Chen, Xudong Lv, Chen He and Joyce Chiang. Without their professional suggestions and personal encouragement, I would never have grown up that fast.

Finally, I am deeply indebted to my family for their endless love and support. A special thank goes to my wife, Aiping Liu, for sharing all the happy as well as frustrated moments with me all the time.

Chapter 1

Introduction

1.1 Background

During the last two decades, due to advances in molecular biology, electronics, computational techniques and biosignal acquisition technologies, there have been increasingly active research activities in studying the nervous system. Recent such advances allow scientists to investigate fundamental questions in neuroscience research – how the brain is able to perceive, process and act upon the tremendous amount of information flowing across in a seemingly effortless manner. While the anatomical structure of the brain is relatively well studied, the underlying mechanism that coordinates various components to act together synergistically is yet to be elucidated. Of particular interest is how various components of the motor system interact together to produce coordinated movements. From movement planning to actual execution, this process involves a series of coordinated actions in functionally specialized brain regions and muscle groups. Understanding movement control has important implications for the development of brain-computer interface (BCI), prosthetic control, deep brain stimulation therapies and treatment for movement disorders [1, 2].

The emergence of powerful new measurement techniques such as neuroimaging and electrophysiology has provided researchers new opportunities to explore unknown aspects of brain functioning from different angles. Among these techniques, electroencephalogram (EEG) and surface EMG (sEMG) have been the most popular candidates in many practical biomedical applications, principally due to their high temporal resolution, noninvasibility, low cost and suitability for long-term monitoring [11]. For instance, wireless body sensor networks, integrating electrophysiological signals and positioning sensors, have been attracting increasing attention for health-monitoring applications, e.g., greatly facilitating home health care in a long-distance manner.

With the availability of multichannel neural signals, earlier works investigating human motor control systems mainly focus on localization of functionally specialized brain regions during specific motor tasks [1]. However, such approaches only provide a limited view of the motor control mechanism. Alternative methods have been developed to investigate the integration of functionally related neuronal groups, termed as brain connectivity [3]. The study of brain connectivity has provided not only a system-level view of brain functioning in the normal state but also an explanatory framework for pathological conditions such as Parkinson's disease (PD) [4, 5]. Nevertheless, during motor tasks, it is not convincing to only analyze brain activity measured by EEG but neglect muscle activity measured by sEMG. One main reason is that EEG contains a lot of non-task related background activities. It is desirable to investigate the coupling between EEG and sEMG. Such coupling analysis allows the identification of underlying components from EEG signals whose temporal patterns are maximally correlated with those of sEMG (i.e., highly modulated by a motor task), while meanwhile discards any non-task related background components. Therefore, jointly analyzing sEMG together with EEG could highly benefit task-related motor control studies.

1.2 Related Works

1.2.1 Wireless EEG and sEMG systems

EEG Systems

In the late 1800s, Richard Caton (1842-1926) first reported the presence of bipotentials on the surface of the human skull [6]. Later, in 1924, Hans Berger measured these electrical signals in the human brain for the first time and provided the first systematic description for EEG [7]. EEG monitors the electrical activity caused by the firing of cortical neurons across the human scalp. EEG activity always reflects the summation of the synchronous activity of thousands or millions of neurons and shows oscillations at a variety of frequencies. The human EEG activity is mainly categorized into five bands by frequency: Delta (1–4 Hz), Theta (4–8 Hz), Alpha (8–12 Hz), Beta (12–30 Hz) and Gamma (30–100 Hz) [8]. An example is shown in Fig. 1.1.

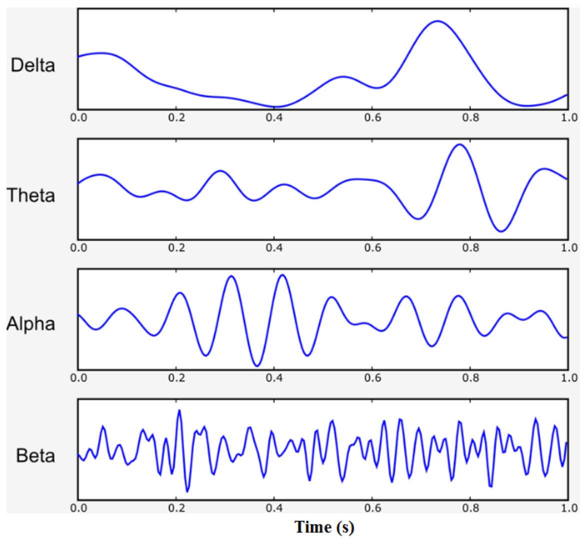


Figure 1.1: EEG frequency bands [9].

EEG has been widely used for studying various neurological conditions such as epilepsy and coma [10]. Diagnostic applications generally focus on spectral content of EEG, reflecting the type of neural oscillations observable in EEG signals. As EEG recording remains the most widespread, noninvasive and inexpensive technology with sufficient temporal resolution and it is suitable for continuous monitoring, the EEG recording systems play an important role in brain studies, especially in diagnosis of brain diseases such as epilepsy, sleeping disorder and abnormal behavior [11].

In current EEG recording systems, the EEG recorders based on personal computer (PC) generally communicate with the medical instruments through the computer I/O interface. Such systems usually adopt a wired, serial port interface, such as RS-232C standard, to transmit the measured EEG data. It can be inconvenient and uncomfortable for daily-life usage because of the wired transmission lines. Recent advances in electronic, communications and information technologies have stimulated great interest world-wide in the development of portable, battery operated biomedical instruments [12]. This has been particularly true in the case of electrocardiogram (ECG) systems, which have been increasingly portable. For continuous monitoring of EEG signals, it is also desirable to have a portable EEG measurement system for reliably measuring brain activity. Size, power consumption, and wireless communication capability are important factors to be considered in designing such EEG systems [13].

sEMG Systems

The first actual recording of electromyogram (EMG) was made by Marey in 1890. Later in 1922, Gasser and Erlanger used an oscilloscope to show the electrical signals from muscles. Then, researchers began to use improved electrodes more widely for the study of muscles. The clinical use of EMG is mainly for diagnosis of neurological and neuromuscular problems. A typical EMG signal and its amplitude envelope are illustrated in Fig. 1.2.

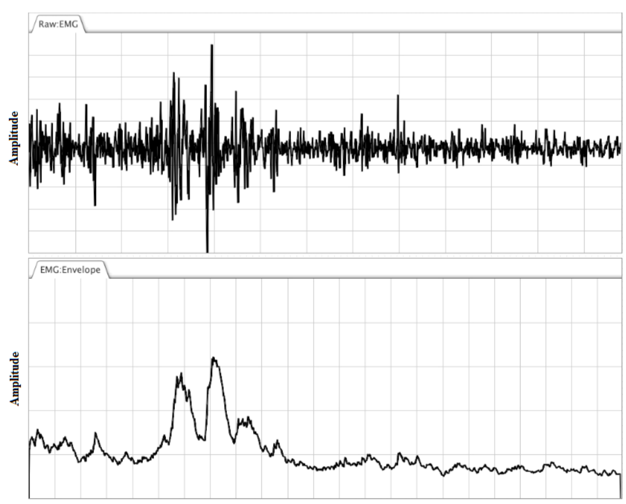


Figure 1.2: An example of an EMG signal [9].

Significant insights have been acquired regarding the underlying neural mechanisms controlling movements, via the simultaneous measurement of sEMG and body kinematics during tasks such as walking, swimming, and scratching [14]. Due to the presence of electrical wires between the electrodes and the sEMG device, as well as the wire between the sEMG device and the computer, sEMG measurements were subject to an inherent limitation in the past. These sEMG devices could also be affected by noise, resulting from the movements of the wires. In recent years, although some commercial wireless sEMG devices, such as Trigno Wireless System produced by Delsys, have appeared, many research groups are still developing their own wireless sEMG systems since it is more convenient for researchers to design and modify their systems according to their specific research needs. We also have designed a sEMG system by appropriately modifying an original wireless EEG system.

Among sEMG applications, hand gesture recognition based on forearm sEMG has been an active research area due to its broad applications in myoelectric control. With using wired or wireless sEMG sensors, human motions can be captured non-invasively by sEMG signals and such sEMG signals can be intelligently recognized as control commands in many myoelectric systems such as multifunction prosthesis, wheelchairs, virtual keyboards, gesture-based interfaces for virtual reality games etc. [15, 16]. Depending on the involved movements, current sEMG-based recognition and classification research can be divided into three main categories: gross hand, wrist and arm movement recognition; individual finger activation and movement detection; and multiple finger gesture classification.

The majority of previous research has been focused on gross hand, wrist and arm movement recognition and quite high recognition accuracy can be achieved. For instance, with two, four, five or eight sEMG channels, a high classification accuracy which is above 95% was reported for classifying four or six movements [17–19]. However, it is worth noting that these movements, such as palm extension and closure, wrist flexion and extension, wrist pronation and supination, represent relatively easy classification problems since they generate distinguishing sEMG activation patterns.

For the category of individual finger activation and movements, such as finger typing, generally involving flexion and extension of the individual thumb, index, middle, ring and little fingers, several recent works have been reported along with this research direction [20–22]. Researchers in [21] used five channels to collect forearm sEMG signals for a piano-tapping task, in which the subjects tapped a keyboard with each of the five fingers, and a 85% recognition accuracy was achieved by using artificial neural network classifiers. In [22], though a 98% high accuracy could be achieved for classifying 12 flexion and extension movements of individual fingers, 32 sEMG channels were used which is highly impractical for real-life online applications.

Regarding the research direction of multiple finger gesture classification, to our knowledge, relatively few research papers were published and they didn't investigate benchmark multi-finger movement tasks.

1.2.2 Corticomuscular Coupling Analysis

Corticomuscular coupling analysis, i.e. studying simultaneous cortical and muscular activities typically during sustained isometric muscle contraction, is a key technique to assess functional interactions in the motor control system. The most popular method is magnitude-squared coherence (MSC), a normalized measure of correlation between two waveforms or signals in the frequency domain, which is calculated by

$$Coh(f) = |S_{e^*e^\circ}(f)|^2 / (S_{e^*e^*}(f) S_{e^\circ e^\circ}(f)), \tag{1.1}$$

where $|S_{e^*e^\circ}|$ is the cross-spectrum between the signals e^* and e°; $S_{e^*e^*}$ is the autospectrum of the signal e^*; $S_{e^\circ e^\circ}$ is the autospectrum of the signal e°. MSC has been successfully used to assess the interactions between motor-related brain areas and the muscles. For example, in monkeys, MSC in the 20-30 Hz band can be detected between cortical local field potentials and the rectified EMG from contralateral hand muscles that are modulated during different phases of a precision grip task [23]. In humans, similar findings of beta-band corticomuscular coherences have also been detected during isometric contractions utilizing both magnetoencephalogram (MEG) [24] and electroencephalography EEG [25] recordings over the primary motor cortex. MSC has proved useful in assessing diseases of the human motor system, particularly PD. An inverse relation exists between beta-band corticomuscular MSC and the clinical sign of bradykinesia in PD [26]. Cerebro-muscular MSC appears unaffected in early PD, yet beta band oscillations of bilateral primary sensorimotor cortices are already increased at the earliest stages of PD [27]. Later on, cerebro-muscular MSC is reduced in non-medicated PD, and medication [28] or deep brain stimulation [26] normalizes this deficiency. Finally primary motor cortex (M1)–muscular MSC is strongly reduced for both alpha and beta bands during repetitive movement compared to static contraction, but this is unaffected by administration of levodopa [29].

Recently, several data-driven multivariate methods, such as partial least squares (PLS), have been developed for analyzing biological data that may be appropriate for assessing corticomuscular coupling as they establish a dependency relationship between two types of data sets. PLS relies on latent variables (LVs), which may enhance aid in the biological interpretation of the results. PLS is widely used in practical applications, probably because PLS can handle high dimensional and collinear data - frequently the case in real-world biological applications. PLS exploits the covariation between predictor variables and response variables and extracts a new set of latent components that maximally relate them [32]. In other words, the covariance between the extracted LVs should be maximized as

$$\begin{aligned} &\max_{w_1,w_2} \quad \left(w_1{}^T X^T Y w_2\right)^2, \\ &\text{s.t.} \quad w_i{}^T w_i = 1, \quad i = 1,2 \end{aligned} \tag{1.2}$$

where w_i's $(i = 1,2)$ are the weight vectors, $X(N \times p)$ is the predictor matrix and $Y(N \times q)$ is the response matrix with N representing the number of observations. PLS and its variants have been investigated in many medical applications, such as determining the spatial patterns of brain activity in functional magnetic resonance imaging (fMRI) data associated with behavioral measures [33], and the common temporal components between EEG and fMRI signals [34]. In a recent PD study, PLS has been extended to perform group level analysis and even accommodate multiway (e.g., time, PLS channel, and frequency) data [30].

1.3 Challenges and Motivation

Although MSC has been popular in studying corticomuscular coupling, it suffers from several limitations. First, incorporating inter-subject variability in order to make a robust group inference is not straightforward with MSC because the exact frequency of maximum coupling may be inconsistent across subjects, necessitating *post hoc* pooled coherence analyses in an attempt to combine the original coherence estimates into a single representative estimate [35]. Second, the pair-wise MSC concentrates on assessing the role of an individual locus in the brain in driving the motor system while motor activity is known to be more distributed [36]. In fact, recent work has suggested that *interactions between* brain regions correspond more closely to ongoing EMG than activity at discrete sites [30]. Moreover, when the brain activity is measured by EEG, applying MSC directly to raw EEG and EMG signals normally yields a very low coherence value, because only a small fraction of ongoing EEG activity is related to motor control [31]. Therefore extensive statistical testing, with all the accompanying assumptions, is required to determine whether the EEG/EMG coherence is, in fact, significant.

Regarding PLS, there exist several concerns with it and its extended methods, which hinder their applications to multimodal corticomuscular activity analysis. First, in the regular PLS, multiway PLS [33] and multiblock PLS [37] frameworks, only two types of data sets can be processed at the same time, e.g. fMRI and behavior measurements [33] or EEG and EMG signals [30], while in many cases more than two types of data sets are available and a better understanding could be achieved from analyzing multimodal data jointly. Another concern lies in the goal of PLS, which is to maximize covariance but not correlation. This may make the extracted underlying components across the data sets highly correlated, but not maximally correlated. In addition, PLS can only extract uncorrelated LVs and their interpretations may be difficult in real applications [38] since it might be insufficient to consider only up to the second-order statistics (e.g. correlation and covariance) for obtaining a unique LV model [39] if the data are not strictly multivariate Gaussian.

According to the aforementioned limitations, the main technical challenges could be summarized as poor signal-to-noise ratio, inter-subject variability, hard biological interpretation and multimodality. The goal of this research work is to develop novel multimodal signal processing techniques to address these challenges arising when modeling corticomuscular activity from EEG, EMG and behavioral recordings. To make this happen, the corresponding objectives are to correlate EEG with the task, perform group analysis, explore higher order statistics and establish multimodal cost functions. Specifically, the main technical contributions are:

1. Investigate the complementary relationship between PLS and canonical correlation analysis (CCA), and combine them to model the cortical and muscular activity.
2. Develop a novel coupling analysis method, termed as IC-PLS, for modeling EEG-EMG activity which incorporates the concept of independence to circumvent the statistical insufficiency of the second-order statistics.
3. Propose a joint multimodal group analysis framework (JMSF) for investigating the relationships between cortical, muscular and behavioral measurements under the uncorrelatedness assumption.
4. Design a three-step method based on the above approach under the independence assumption.

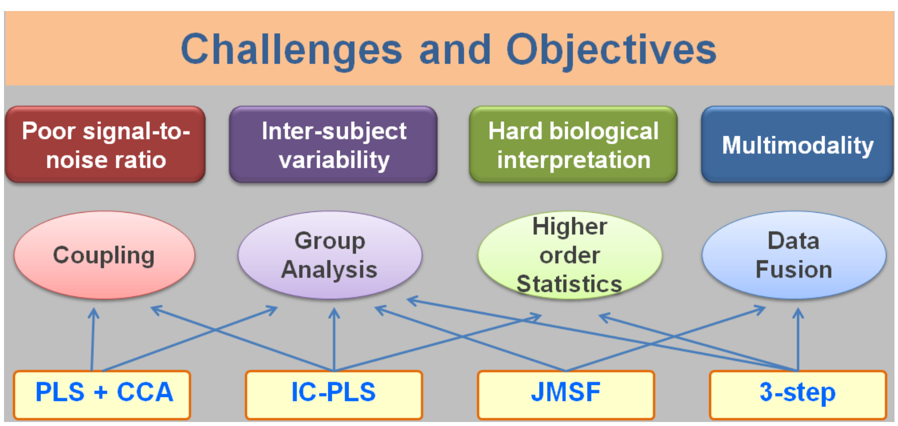


Figure 1.3: The overview of challenges and objectives of the thesis work.

Fig. 1.3 illustrates all the challenges, objectives, contributions and their relationships. The developed techniques are applied to EEG, EMG and behavioral data collected from both healthy subjects and PD patients when they perform a dynamic, visually guided tracking task.

1.4 Thesis Outline

The thesis outline is summarized as follows:

In Chapter 2, we first overview existing coupling analysis methods and summarize their advantages and disadvantages. We then describe the data sets collected from a visually guided tracking task and explain the data preprocessing procedures. After that, we propose combining PLS and CCA to improve the performance of the joint latent variable (LV) extraction. The proposed PLS+CCA has a two-step modeling strategy. PLS first extracts LVs which are able to most explain individual data set and meanwhile are well correlated to the LVs in the other data set. Then CCA is utilized to extract the LVs by maximizing the correlation coefficients. The extracted components are guaranteed to be maximally correlated across two data sets and meanwhile well explain the information within individual data sets.

In Chapter 3, to overcome the insufficiency of the uncorrelatedness assumption, we propose an IC-PLS framework to simultaneously incorporate response-relevance and independence into the regression procedure, keeping the LVs maximally independent and uniquely sorting the LVs in order of relevance. When applied to corticomuscular coupling analysis, the proposed IC-PLS is able to extract the most significant LV pairs from concurrent EEG and EMG data in an orderly manner.

In Chapter 4, to accommodate multimodal data sets, we propose a joint multimodal statistical framework to simultaneously model multiple data spaces, keeping the LVs uncorrelated within each data set and meanwhile highly correlated across multiple data sets.

In Chapter 5, to facilitate interpretations in real medical applications, we incorporate independence into the joint multimodal framework and propose a three-step method by combining multiset canonical correlation analysis (M-CCA) and joint independent component analysis (ICA). This method is able to explore the relationships between multimodal data sets and meanwhile extract independent LVs within each data set. We applied these proposed methods to concurrent EEG, EMG and BEH data collected from normal subjects and patients with PD when performing a dynamic force tracking task. We utilized the proposed methods to extract highly correlated temporal patterns among the three types of signals (features) and reported meaningful connectivity patterns.

Finally, the conclusions of the thesis and suggestions for future research are summarized in Chapter 6.

In Appendix A, we present the entire design process of a wearable and wireless EEG recording system and report a few testing examples. Then, we introduce a real-time sEMG hand gesture recognition system, including the hardware design, data collection, feature extraction and classification components. For feature extraction and classification, we investigate the most popular methods and compare their performances systematically. Further, we propose employing multiple kernel learning support vector machine (MKL-SVM) for hand gesture recognition and demonstrate its superior performance. The reason we put this part in Appendix is that we did not use our developed system to collect data for corticomuscular coupling analysis although our original purpose was to utilize an integrated wireless EEG and EMG system for patients' convenience. We were required to use commercialized medical devices in scientific publications.

Chapter 2

Modeling Corticomuscular Activity by Combining PLS and CCA

2.1 Introduction

Corticomuscular activity modeling is important for assessing functional interactions in the motor control system, i.e. studying simultaneous cortical and muscular activities during a sustained isometric muscle contraction. As mentioned in Section 1.2.2, the most common method to assess the interactions between motor-related brain areas and the muscles is MSC, which is a normalized measure of correlation between two waveforms or signals in the frequency domain. Although MSC has been popular in studying corticomuscular coupling, it suffers from several limitations. First, addressing the inter-subject variability challenge to make a robust group inference is not straightforward with MSC because the exact frequency of maximum coupling may be inconsistent across subjects. Second, MSC emphasizes the role of individual locus in the brain in driving the motor system while motor activity is known to be more distributed [36]. In fact, recent work has suggested that *interactions between* brain regions correspond more closely to ongoing EMG than activity at discrete sites [30, 71–74]. Moreover, when the brain activity is measured by EEG, applying MSC directly to raw EEG and EMG signals normally yields a very low coherence value, because only a small fraction of ongoing EEG activity is related to the motor control [31]. This implies that extensive statistical testing is required to determine whether the EEG/EMG coherence is statistically significant.

Recently, several data-driven multivariate methods have been developed for analyzing biological data, and they seem to be appropriate for modeling corticomuscular activity because these methods explore dependency relationships between data sets. These methods include multiple linear regression, principal component regression, PLS and CCA [75]. Among these methods, the LV based approaches, such as PLS and CCA, play a dominating role, probably due to the fact that the extracted LVs could help the biological interpretations of the results.

PLS, first developed for process monitoring in chemical industry, exploits the covariation between predictor variables and response variables and finds a new set of latent components that maximally relate them [32]. An advantage of PLS is that it can handle high dimensional and collinear data, which is often the case in real-world biological applications. PLS and its variants have been investigated in many medical applications, such as assessing the spatial patterns of brain activity in fMRI data associated with behavioral measures [33], and the common temporal components between EEG and fMRI signals [34]. In addition to the ability of handling high dimensional and collinear data, PLS is sufficiently flexible that it can be extended to perform group level analysis and to accommodate multiway data [30].

CCA is commonly used to seek a pair of linear transformations between two sets of variables, such that the data are maximally correlated in the transformed space. Generally CCA is not as popular as PLS in practical applications [76]. This is probably because real-world data are usually high dimensional and collinear and thus applying CCA directly to the raw data can be ill-conditioned. However, with some appropriate preprocessing strategies, CCA has been shown to be quite useful in many medical applications. For instance, in [77] Clercq et al. successfully removed muscle artifacts from a real ictal EEG recording without altering the recorded underlying ictal activity. In [79], Gumus et al. found that there were significant correlations at expected places, indicating a palindromic behavior surrounding the viral integration site. CCA can be extended to accommodate multiple data sets simultaneously [78].

However, PLS and CCA can only extract uncorrelated LVs and their interpretations may be difficult in real applications [38]. ICA is based on the notion that it is insufficient to consider only up to the second-order statistics (e.g. correlation and covariance) for obtaining a unique LV model [39] if the data are not strictly multivariate Gaussian. ICA assumes that the multivariate data are composed of a linear superposition of mutually statistically independent signal sources. In statistics, independence is a much stronger condition than uncorrelatedness, and thus ICA algorithms typically employ criteria related to information theory and/or non-Gaussianity. ICA has found many potential biomedical applications [80–82]. For example, ICA can be used to reliably separate fMRI data into meaningful constituent components, including consistently and transiently task-related physiological changes, non-task-related physiological phenomena and machine or movement artifacts [80].

Independent component regression (ICR) has been developed as an alternative to traditional LV-based methods (e.g. PLS) [83]. In ICR, independent components (ICs) are first extracted from the measurements and then linear regression is performed to relate the ICs with the response. However, as a two-stage method, ICR inherits the following drawbacks of ICA. First, ICA can only decompose one data set at a time and thus ignores the effects of the response variables. Even though the extracted ICs are maximally independent with each other, they may, and are typically not, directly informative about the response. Second, unlike principal component analysis (PCA) which yields unique and ranked principal components (PCs) by best explaining the variance in the data, classical ICA with stochastic learning rules can result in ICs that vary by permutation and/or dilation even when repeatedly performed on the same data set. Such inherent indeterminacy of classical ICA is due to the fact that the assumptions are mathematically insufficient to extract the independent sources exactly in their original form. To overcome these drawbacks, joint independent component analysis (jICA), was developed to maximize the independence of joint sources of multiple data sets [84]. However, in the jICA framework, all modalities are assumed to share the same mixing matrix, which may not be true in practice [84].

To address the aforementioned technical challenges arising when modeling corticomuscular activity, we aim to develop novel signal processing techniques in the following sections and chapters. We evaluate the performance of the proposed methods both numerically and practically. We first apply the proposed methods to synthetic data to illustrate their performance where the underlying truth is well understood. We then apply the proposed group analysis methods to concurrent EEG, EMG and behavior data (BEH) collected from normal subjects and patients with PD when performing a force tracking task.

2.2 Experimental Data

The study was approved by the University of British Columbia Ethics Board, and all subjects gave written, informed consent prior to participating. Nine PD patients (mean age: 66 yrs) were recruited from the Pacific Parkinson's Research Centre at the University of British Columbia (Vancouver, Canada). They all displayed mild to moderate levels of PD severity (stage 1-2 on the Hoehn and Yahr scale) and were being treated with L-dopa medication (mean daily dose of 720mg). All PD subjects were assessed after a minimum of 12-hour withdrawal of L-dopa medication, and their motor symptoms were assessed using the Unified Parkinson's Disease Rating Scale (UPDRS), resulting in a mean score of 23. In addition, eight age-matched healthy subjects were recruited as controls. During the experiment, subjects seated 2 m away from a large computer screen. The visual target was displayed on the screen as a vertical yellow bar oscillating in height at 0.4 Hz. Subjects were asked to squeeze a pressure-responsive bulb with their right hand. The visual feedback representing the force output of the subject was displayed as a vertical green bar superimposed on the target bar as shown in Fig. 2.1. Applying greater pressure to the bulb increased the height of the green bar, and releasing pressure from the bulb decreased the height of the green bar. Subjects were instructed to make the height of the green bar match the height of target bar as closely as possible. Each squeezing period lasted for 15 seconds and was followed by a 15-second rest period. The squeezing task was performed twice. The force required was up to 10% of each subject's maximum voluntary contraction (MVC), which was measured at the beginning of each recording session.

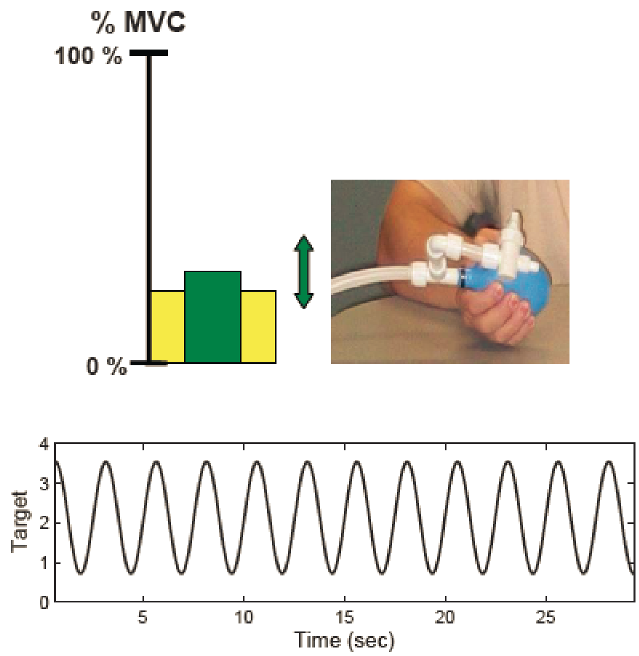


Figure 2.1: The squeezing task: The subject was instructed to follow the target bar (yellow) as close as possible. The force exerted by the subject was shown by the green bar.

The EEG data were collected using an EEG cap (Quick-Cap, Compumedics, Texas, USA) with 19 electrodes based on the International 10-20 system, referenced to linked mastoids. The EEG and EMG data were sampled at 1kHz using SynAmps2 amplifiers (NeuroScan, Compumedics, Texas, USA). A surface electrode on the tip of the nose was used as ground. Ocular movement artifacts were measured using surface electrodes placed above and below the eyes (Xltek, Ontario, Canada). Data were later processed by a band-pass filter (1 to 70Hz) off-line and down-sampled to 250 Hz. Artifacts associated with eye blinks and muscular activities were removed using the Automated Artifact Removal in the EEGLAB Matlab Toolbox (Gomez-Herrero, 2007). To simplify analysis, the EEG signals were divided into five regions as shown in Fig. 2.2. The five regions were: *Fronto-Central (FCentral)* - FP1, FP2, F7, F3, Fz, F4 and F8; *Left Sensorimotor (LSM)* - T7, C3, P7 and P3; *Right Sensorimotor (RSM)* - T8, C4, P8 and P4; *Central* - Cz and Pz; and *Occipital* - O1 and O2. While the proposed framework is capable of handling data at an individual electrode level, we found in our prior work [100] that averaging recordings within a region has several potential benefits: 1) it reduces the effects of possible spurious activity measured at individual electrodes, 2) it reduces inter-subject variability due to slight variations in the placement of the electrode cap, and 3) regions can be chosen as they are biologically related to the motor task we are studying based on prior medical knowledge, making neuroscience interpretations easier. Thus, the raw time courses of the electrodes within each region were averaged as the overall activity of the region, and the averaged time course were then zero-meaned and normalized to unit variance. For subsequent analysis, data collected during the squeezing periods were concatenated in time into a single matrix for each individual subject. Data from the rest periods were excluded from the analysis.

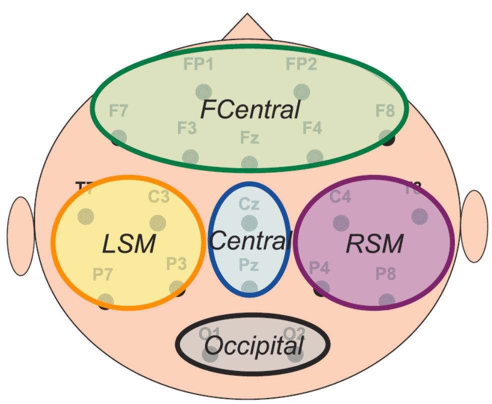


Figure 2.2: Five brain regions: *Fronto-Central (FCentral)* - FP1, FP2, F7, F3, Fz, F4 and F8, *Left Sensorimotor (LSM)* - T7, C3, P7 and P3, *Right Sensorimotor (RSM)* - T8, C4, P8 and P4, *Central* - Cz and Pz and *Occipital* - O1 and O2.

The EMG signals were recorded using self-adhesive, silver, silver-chloride pellet surface electrodes with 7 mm diameter. A bipolar montage was used with a fixed inter-electrode distance of 30 mm. The surface EMG signals were simultaneously collected together with the EEG signals and were amplified and sampled at 1000 Hz. To be consistent with the EEG preprocessing, the EMG signals were down-sampled off-line to 250 Hz and only the squeezing periods were used for subsequent analysis. The BEH signals were recorded from the pressure-responsive bulb, representing the force preformed by the subjects. Similarly, the BEH signal from each subject was resampled to ensure its proper alignment with other signals. It is worth noting that it took a graduate student in Neuroscience about six months to collect these data sets, including recruiting subjects and designing experimental protocol.

2.3 A Combined PLS and CCA Method

2.3.1 Motivation and Objectives

PLS and CCA have been investigated in many medical applications. However, to the best of our knowledge, no report has profoundly explored their underlying differences, compared their characteristic performances, and combined their advantages to overcome their drawbacks. For corticomuscular activity modeling, as we will elaborate more in Section 2.3.2, both PLS and CCA have their advantages and disadvantages, but perhaps more importantly, these two methods can be considered complementary. In this chapter, we propose combining PLS and CCA to improve the performance of the joint LV extraction and the proposed method is denoted as PLS+CCA. More specifically, the proposed PLS+CCA has a two-step modeling strategy: we first adopt PLS to obtain LVs across two data sets and then perform CCA on the extracted LVs. In the first step, PLS is performed for preliminary LV preparation. The aim of this step is to extract LVs which can most explain its own data set and meanwhile are well correlated to the LVs in the other data set. Besides, this step can also prevent the ill-conditioned problem when applying CCA directly to the raw data. In the second step, CCA is applied to the extracted LVs by PLS to construct the LVs by maximizing the correlation coefficients. With these two steps, it is ensured that the extracted components are maximally correlated across two data sets and meanwhile can well explain the information within individual data sets.

2.3.2 Methods

In this subsection, we first analyze the properties of PLS and CCA and demonstrate their complementarity. Based on this observation, we then propose combining the two approaches to have the PLS+CCA method. The two zero-meaned data sets are stored in two matrices, the predictor matrix $X(N \times p)$ and the response matrix $Y(N \times q)$, where N means the number of observations and p and q indicate the numbers of variables in corresponding matrices.

Partial Least Squares

PLS exploits the covariation between predictor variables and response variables and tries to find a new set of LVs that maximally relate them [76]. In other words, the covariance between the extracted LVs should be maximized as

$$\begin{aligned} &\max_{w_1,w_2} \quad \left(w_1{}^T X^T Y w_2\right)^2, \\ &\text{s.t.} \quad w_i{}^T w_i = 1, \quad i = 1,2 \end{aligned} \tag{2.1}$$

where w_i's $(i = 1,2)$ are the weight vectors. A typical PLS can be implemented by the classical NIPALS algorithm [32]. Also, an alternative calculation way is to perform eigenvalue-eigenvector decomposition [85]. Therefore, the maximum of Equation (2.1) is achieved by having w_1 and w_2 as the largest eigenvectors of the matrices $X^T YY^T X$ and $Y^T XX^T Y$ respectively. To obtain subsequent weights, the algorithm is repeated with deflated X and Y matrices. The detailed calculation procedure can be found in Appendix B.1.1.

The number of components to be extracted is a very important parameter of a PLS model. Although it is possible to extract as many PLS components as the rank of the data matrix X, not all of them are generally used. The main reasons for this are the following: the measured data are never noise-free and some trivial components only describe noise. Therefore appropriate measures are needed to determine when to stop. Typically, the number of components needed to describe the data matrices is determined based on the amount of variation remained in the residual data [32].

Canonical Correlation Analysis

Different from PLS, CCA is to find linear combinations of both X and Y variables which have maximum correlation coefficient with each other. This leads to the same objective function but different constraints compared with Equ. (2.1):

$$\begin{aligned} &\max_{v_1,v_2} \quad \left({v_1}^T X^T Y v_2\right)^2 \\ &\text{s.t.} \quad {v_1}^T X^T X v_1 = 1, \quad {v_2}^T Y^T Y v_2 = 1 \end{aligned} \tag{2.2}$$

where v_i's $(i = 1,2)$ are the weight vectors.

The solutions to this problem are the largest eigenvectors of the matrices – $(X^TX)^{-1}X^TY(Y^TY)^{-1}Y^TX$ and $(Y^TY)^{-1}Y^TX(X^TX)^{-1}X^TY$ – respectively. The subsequent weights are the eigenvectors of the same matrix in the order of decreasing eigenvalues. The predictor LVs U_X can be calculated directly from the original X matrix as $U_X = XV_1$, the columns of which are uncorrelated with each other. The detailed derivation is shown in Appendix B.1.2. However, the solution depends heavily on whether or not the covariance matrix X^TX is invertible. In practice, it is possible to have $\text{rank}(X^TX) < p$ so that the invertibility cannot be satisfied and directly applying eigenvalue decomposition in the raw data space may lead to the ill-conditioned problem. Therefore, some appropriate preprocessing strategies are needed in practice before applying CCA.

The Combined PLS+CCA Method

Based on the discussion above, we can see that the fundamental difference between PLS and CCA is that PLS maximizes the covariance while CCA maximizes the correlation. The objective of PLS is to construct LVs which could most explain their own data set and meanwhile are well correlated to the corresponding LVs in the other set. In other words, the first priority of PLS is to find the LVs which can explain significant proportion of variance in each data set and the second priority is to find the LVs with relatively high correlation coefficients between the two data sets. In contrast, the only objective of CCA in the construction of LVs is to maximize their correlation coefficients with the LVs in another data set. From this point of view, the LVs extracted by PLS are able to represent major information for individual data sets while the ones extracted by CCA may be trivial (e.g. noises with similar patterns) even if their correlation coefficient is maximum. This is an advantage of PLS over CCA. Besides, PLS can handle high dimensional and collinear data, which is often the case in real-world biological applications, while applying CCA directly to the raw data may be ill-conditioned. However, we should note that our goal is to find the relationships between two data sets, not just to explore the information within individual data sets. It is possible that a higher covariance merely results from the larger variance of LVs, which may not necessarily imply strong correlations. To overcome this, CCA is a powerful tool to ensure that the extracted LVs have similar patterns across the data sets.

For corticomuscular activity modeling, the coupling relationship between EEG and EMG signals is what to be explored. In practice, EEG and EMG signals can be contaminated by other types of signals and are never noise-free. In addition, the signals from adjacent channels generally are similar, which leads to collinear data. By employing PLS, we can deal with the collinear EEG/EMG data sets and extract significant LVs, but it can not guarantee that the corresponding LVs are highly correlated with each other. With using CCA, we can extract highly correlated LVs from EEG and EMG signals, but it can not ensure that such LVs are non-trivial and we may face the ill-conditioned problem.

For corticomuscular coupling analysis, both PLS and CCA have their advantages and disadvantages, but perhaps most importantly, these two methods can be considered complementary. It is natural for us to think of combining PLS and CCA to form a two-step modeling strategy. In the first step, PLS is performed for preliminary LV preparation. The aim of this step is to extract LVs which can most explain its own data set and meanwhile are well correlated to the LVs in another data set. In this case, the trivial and irrelevant information across data sets could be removed. Besides, this step can also prevent the ill-conditioned problem when applying CCA directly to the raw data. In the second step, CCA is applied to the prepared LVs by PLS to construct the LVs by maximizing the correlation coefficients. After these two steps, it is ensured that the extracted components are maximally correlated across data sets and meanwhile can well explain the information within each individual data set. The details of the proposed PLS+CCA method are given in Appendix B.1 and the specific implementation procedure is

Algorithm 1 The Combined PLS+CCA Method

Input: two data sets X (with size $N \times p$) and Y (with size $N \times q$)
Output: corresponding LVs matrices U_X, and U_Y

The First Step:

1: Solve the eigen decomposition problems:
$\left(X^T Y Y^T X\right) w_1 = \lambda_1 w_1$ and $\left(Y^T X X^T Y\right) w_2 = \lambda_2 w_2$.
2: Determine R_1 and R_2, the numbers of LVs extracted, corresponding to the above two problems by the ratio of explained variance.
3: Determine the final number of LVs: $R = \min(R_1, R_2)$.
4: Set $count = R$.
5: Initialize both LVs matrices to be empty, i.e., $T_X =[\ ]$ and $T_Y =[\ ]$.
6: **while** $count > 0$ **do**
7:     Set w_1 and w_2 to be the largest eigenvectors of the matrices $X^T Y Y^T X$ and $Y^T X X^T Y$ respectively.
8:     Calculate the LVs as $t_X = X w_1$ and $t_Y = Y w_2$.
9:     Set $T_X =[T_X\ t_X]$ and $T_Y =[T_Y\ t_Y]$.
10:     Deflate X by subtracting the effects of the LV t_X from the data space: $X = X - t_X ({t_X}^T t_X)^{-1} {t_X}^T X$.
11:     Deflate Y by subtracting the effects of the LV t_Y from the data space: $Y = Y - t_Y ({t_Y}^T t_Y)^{-1} {t_Y}^T Y$.
12:     Let $count = count - 1$.
13: **end while**

The Second Step:

14: Solve the following eigen decomposition problems:

$$\left[({T_X}^T T_X)^{-1} {T_X}^T T_Y ({T_Y}^T T_Y)^{-1} {T_Y}^T T_X\right] v_1 = \eta_1 v_1 \text{ and}$$

$$\left[({T_Y}^T T_Y)^{-1} {T_Y}^T T_X ({T_X}^T T_X)^{-1} {T_X}^T T_Y\right] v_2 = \eta_2 v_2.$$

15: Set V_1 and V_2 to be the R associated eigenvectors respectively.
16: The recovered LVs U_X and U_X can be calculated by
$U_X = T_X V_1$ and $U_Y = T_Y V_2$.

shown in Algorithm 1.

2.3.3 Data Processing and Results

Simulation

In this simulation, we apply the proposed method to synthetic data and also report the results of the PLS and CCA approaches for comparison. As an illustrative example, without loss of generality, four sources are generated and analyzed for each data set.

Synthetic Data The following four source signals are considered for the data set X:

$$\begin{aligned}
s_{11} &= 1.5\sin(0.025(t+63))\sin(0.2t),\\
s_{12} &= 1.5\sin(0.025t),\\
s_{13} &= \text{sign}(\sin(0.3t)+3\cos(0.1t)),\\
s_{14} &= \text{uniformly distributed noise in the range } [-1.5, 1.5],
\end{aligned} \tag{2.3}$$

where t denotes the time index vector, valued from 1 to 1000, and s_{1i}'s $(i = 1,2,3,4)$ represent four simulated sources, as shown in Fig. 2.3a. Note that here s_{1i}'s are column vectors.

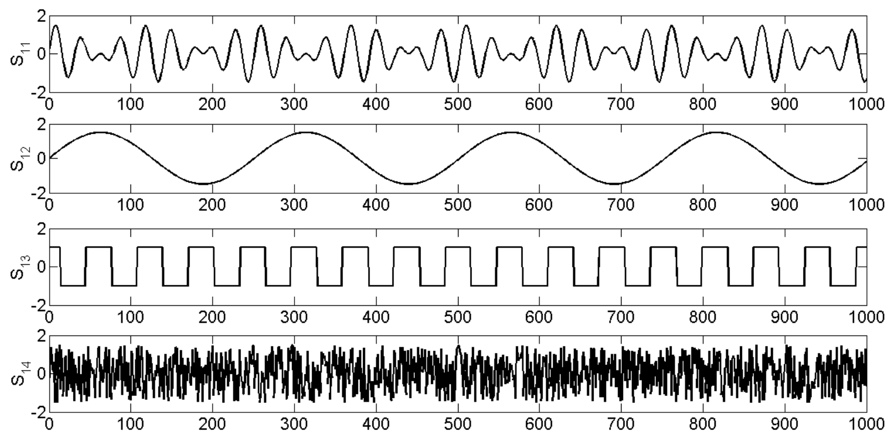
Also, four source signals are considered for the data set Y:

$$\begin{aligned}
s_{21} &= 1.5\sin(0.025(t+69))\sin(0.2(t+6))\\
s_{22} &= 1.5\sin(0.025(t+20))\\
s_{23} &= \text{sign}(\sin(0.3(t+7))+3\cos(0.1(t+7)))\\
s_{24} &= \text{uniformly distributed noise (the same as } s_{14})
\end{aligned} \tag{2.4}$$

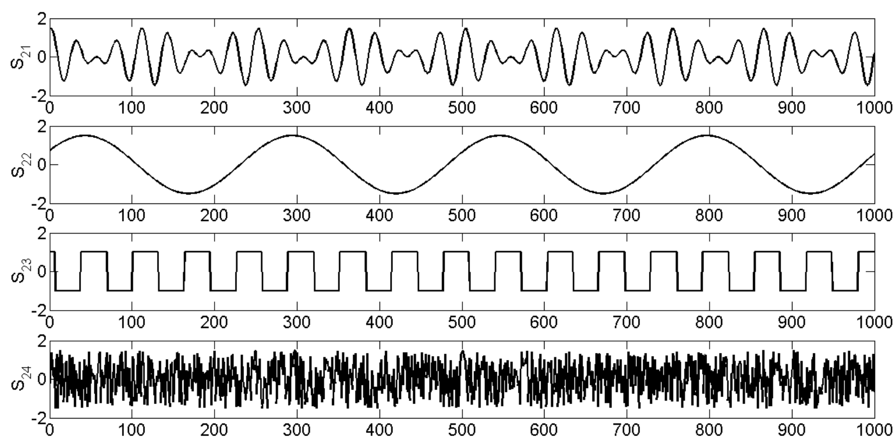
where the notations are similarly defined. The four simulated sources are shown in Fig. 2.3b.

Two mixed data sets X and Y are generated as follows with each row denoting one observation in their respective data space:

$$X = S_1 \cdot A, \quad Y = S_2 \cdot B, \tag{2.5}$$



(a)



(b)

Figure 2.3: The four simulated source signals: (a) for X; (b) for Y.

where $S_1 = [s_{11}\, s_{12}\, s_{13}\, s_{14}]$ and $S_2 = [s_{21}\, s_{22}\, s_{23}\, s_{24}]$ with

$$A = \begin{bmatrix} 0.76 & -0.65 & 0.77 & 0.83 & 0.82 \\ 0.49 & 0.25 & 0.12 & 0.22 & -0.17 \\ 0.28 & -0.21 & 0.11 & 0.19 & -0.11 \\ 0.07 & 0.06 & -0.08 & 0.07 & -0.04 \end{bmatrix}, \tag{2.6}$$

Table 2.1: The correlation coefficients between the corresponding source pairs of X and Y.

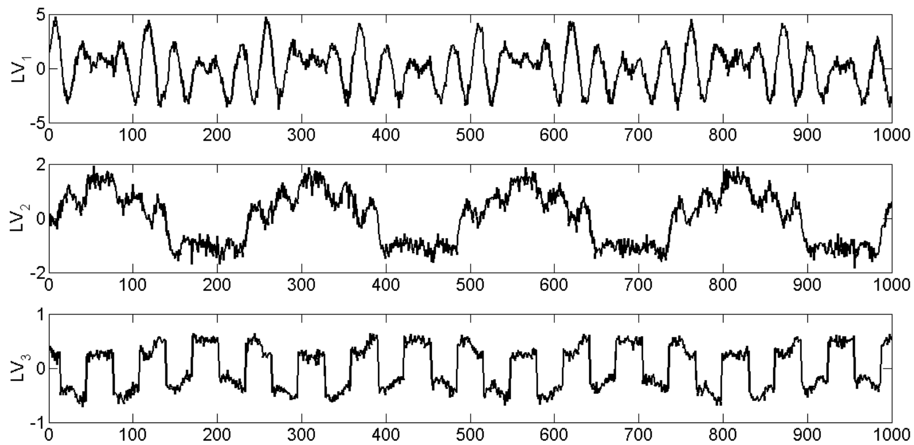
	s_{11} and s_{21}	s_{12} and s_{22}	s_{13} and s_{23}	s_{14} and s_{24}
CC*	0.3655	0.8787	0.5520	1.00

* Here CC stands for correlation coefficient between two source signals.

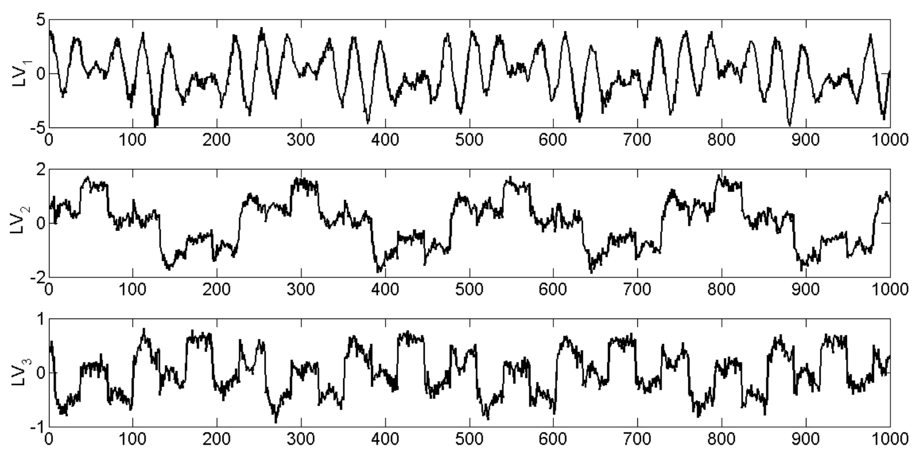
$$B = \begin{bmatrix} 0.73 & -0.82 & 0.91 & -0.79 & 0.88 \\ 0.42 & -0.27 & 0.17 & -0.20 & -0.30 \\ 0.27 & 0.26 & -0.18 & 0.17 & -0.24 \\ 0.08 & -0.01 & 0.01 & 0.09 & -0.01 \end{bmatrix}. \tag{2.7}$$

The patterns of the corresponding sources are similar across the two data sets, representing common information. However, from Equations (2.3) and (2.4), we can see that there are some time-shifts between corresponding source pairs and their correlation coefficients are given in Table 2.1. The first pair of sources have the lowest CC, but in the mixed data sets we intentionally assign the highest weights to this pair of sources, as shown in the mixing matrices A and B. This pair can represent the major information within individual data sets, but can not reflect too much the coupling relationships between the two sets. The second and third pairs have relatively high CCs and moderate weights in the mixed data sets. These two pairs generally not only contain the major information within individual data sets but also represent the coupling relationships across data sets. The fourth pair sources have the highest CC, but we assign the smallest weights. Although this pair sources have the highest CC, they do not represent significant information due to the small weights. Generally, they could be regarded as trivial information. Moreover, different white Gaussian noise with 10% power was added to each source in each data space.

Results The extracted components using PLS, CCA and the proposed PLS+CCA methods are shown in Figs. 2.4, 2.5 and 2.6 respectively. The LVs extracted by PLS are automatically ordered in terms of their significance. To some extent, the LVs successfully reflect the corresponding relationships of the underlying sources



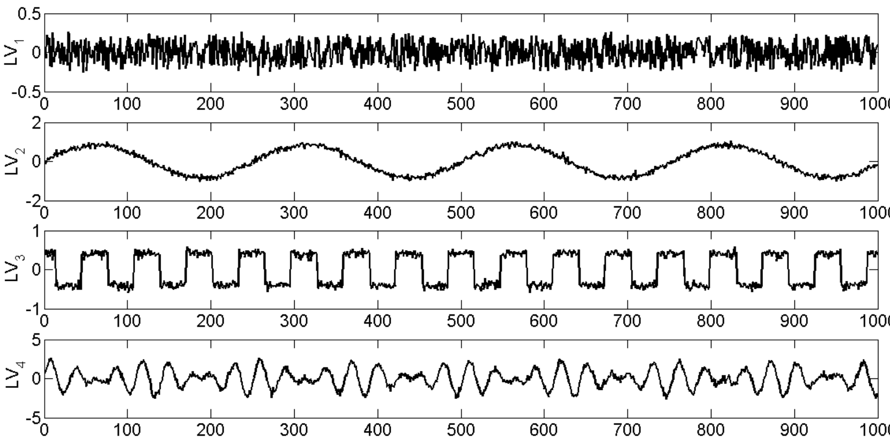
(a)



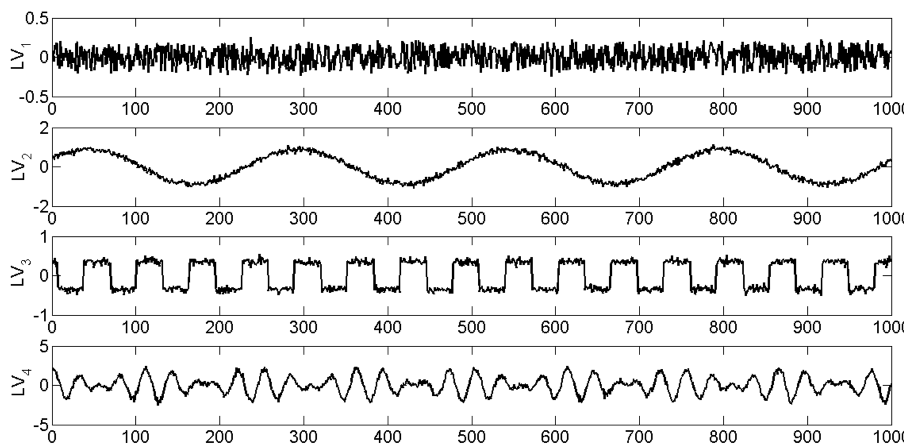
(b)

Figure 2.4: (a) The LVs estimated in X using PLS. (b) The LVs estimated in Y using PLS.

between X and Y. However, compared with the original sources, the extracted LVs are distorted, suggesting that a higher covariance may merely result from the larger variance of LVs, which may not necessarily imply strong correlations. We can see that CCA can recover the original sources accurately in both data spaces and the LVs are ordered strictly according to their correlation coefficients, but it completely ignores the influence of the variance and thus the extracted LVs may only reflect



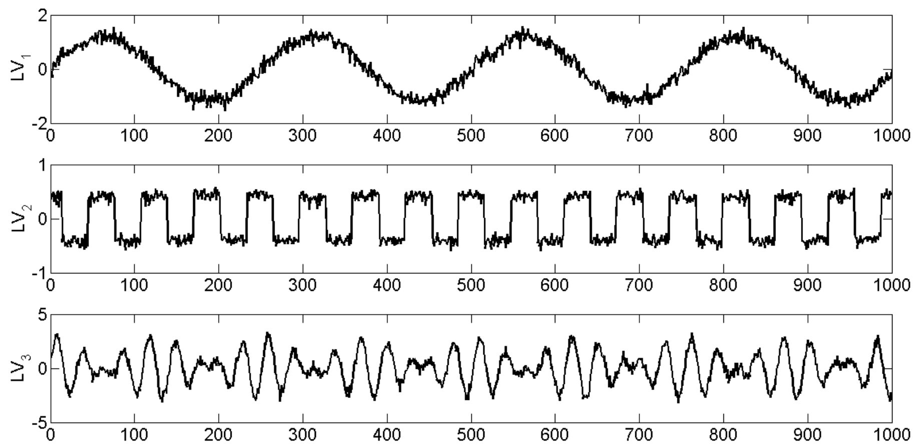
(a)



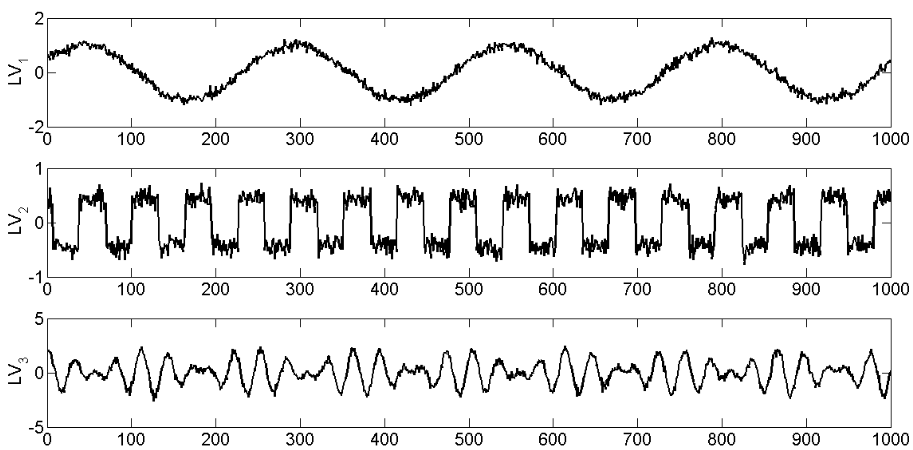
(b)

Figure 2.5: (a) The LVs estimated in X using CCA. (b) The LVs estimated in Y using CCA.

trivial information of the data sets (e.g., the 1st LV). For instance, although the first pair of LVs have the highest correlation coefficient, they do not contain major information of the data spaces. In practice, such LVs generally represent the noises with similar patterns simultaneously coupled into the two data modalities. When jointly modeling the data sets, they should be removed. We also note that PLS only extracts three LVs since they are sufficient to describe the data sets. These LVs



(a)



(b)

Figure 2.6: (a) The LVs estimated in X using the proposed PLS+CCA. (b) The LVs estimated in Y using the proposed PLS+CCA.

do not include the first pair recovered by CCA due to their triviality. The above observations motivates us to employ the proposed PLS+CCA method.

When the proposed method is employed, the dominant sources which make significant contributions to both data spaces are first identified and ordered in terms of covariance. At the same time, trivial information is removed. Then, within the extracted major information, sources that are highly correlated are accurately recovered with the focus on correlation. In this case, it is ensured that the extracted LVs are maximally correlated across two data sets and meanwhile can well explain the information within each individual data set.

Real Data

In many medical applications, the results of analyzing one subject's data can not be generalized to the population level because of the inter-subject variability concern. Therefore, it is necessary to recruit a proper number of subjects to perform a group analysis. For modeling the corticomuscular activity, we apply the proposed method to concurrent EEG and EMG signals collected from normal subjects and patients with PD during a motor task, described in Section 2.2.

Feature Extraction In most existing studies, the analysis for corticomuscular coupling is performed directly on the raw EEG and EMG data. This typically yields quite small correlation values. Nonetheless, with appropriate preprocessing steps, highly correlated EEG and EMG feature(s) can be extracted from the raw signals. In this chapter, we examine the coupling relationships between time-varying EEG features and amplitudes of the EMG signals, constituting X_b and Y_b respectively for each subject b (for $b = 1,2,...,B$). We have a total of B subjects ($B = 17$ in this study). To achieve a group analysis, all subjects' data sets are concatenated together as:

$$\begin{aligned} X &= [X_1, X_2, ..., X_B], \quad \forall b = 1,2,...,B \\ Y &= [Y_1, Y_2, ..., Y_B], \quad \forall b = 1,2,...,B \end{aligned} \tag{2.8}$$

with the assumption that all subjects share common group patterns in the temporal dimension [82].

EEG Features: pair-wise Pearson's correlations [86] are considered in this study. Pearson's correlation measures the dependency between a pair of EEG signals $e^* = (e^*_1, e^*_2, ..., e^*_n)$ and $e^\circ = (e^\circ_1, e^\circ_2, ..., e^\circ_n)$ in the time domain as follows:

$$\gamma_{e^* e^\circ} = \frac{\sum_{i=1}^n (e^*_i - \bar{e^*})(e^\circ_i - \bar{e^\circ})}{\sqrt{\sum_{i=1}^n (e^*_i - \bar{e^*})^2 \sum_{i=1}^n (e^\circ_i - \bar{e^\circ})^2}}, \tag{2.9}$$

where $\bar{e^*}$ and $\bar{e^\circ}$ are the sample means of e^* and e°. In this chapter, we calculate the time-varying pair-wise correlations between EEG channels, using a Hamming window with length 300 and with a 95% overlap. Therefore, the raw EEG information can be represented by a two-dimensional matrix with size $N \times M$, where the rows correspond to the samples at different time points and the columns correspond to the features, i.e. pair-wise correlations between the EEG channels.

EMG Features: an individual EMG channel signal can be considered as a zero-mean, band-limited and wide-sense stationary stochastic process modulated by the EMG amplitude, which represents the overall muscle activity of individual underlying muscle fibers [87]. While different techniques have been proposed for accurate amplitude estimation, in this study, we employ the root-mean-square approach to calculate the EMG amplitude of short-duration EMG signals $e = (e_1, e_2, ..., e_n)$:

$$e_{rms} = \sqrt{\frac{1}{n}({e_1}^2 + {e_2}^2 + \cdots + {e_n}^2)}. \tag{2.10}$$

A moving window with length $n = 300$ and a 95% overlap is applied here, the same as in the EEG feature calculation, to ensure that the obtained EEG and EMG features are temporally aligned and matched.

In the above setting, for each subject b (for $b = 1, 2, ..., B$), X_b and Y_b represent the time-varying feature matrices of EEG and EMG respectively. The length of the time sequences here is 480 associated with the 300-length moving window and a 95% overlap. For the EEG correlation feature, since we have 19 EEG channels based on the International 10-20 system and thus there are a total of $C_2^{19} = 171$ correlation connections. Therefore, X_b is with size 480×171. For the EMG amplitude feature, since there are three surface EMG channels, Y_b is with size 480×3.

Significance Assessment To determine the statistical significance levels of the extracted LVs, we employ a non-parametric permutation test [88] in which the temporal order of EEG features X_b is uniformly permuted for all subjects while keeping the EMG features Y_b intact. Two hundred random permutations are generated. The proposed PLS+CCA method is applied to each of these permutations. The correlation coefficients among the extracted temporal patterns from permuted EEG features and unchanged EMG features are then calculated to form an empirical null distribution. The p-value of the original EEG/EMG correlation coefficient is then computed from the null distribution as the proportion of sampled permutations whose correlation coefficients are greater than or equal to the original correlation coefficient. The components with p-value being less than 0.05 are considered to be statistically significant, denoted as LV_{EEG} and LV_{EMG}, both with size $(N \times K)$, where K means the number of significant components.

Spatial Pattern Extraction Our goal is to investigate the differences in spatial patterns of EEG channels between the normal and PD patient groups when the subjects perform a motor task. After the identification of significant temporal patterns, we can regress the EEG-related components LV_{EEG} back onto the EEG features X_b (for $b = 1, 2, ..., B$) for each subject as follows:

$$p_{bk} = \sqrt{\frac{1}{lv_k^T X_b X_b{}^T lv_k}} X_b{}^T lv_k, \quad k = 1, 2, ..., K. \tag{2.11}$$

where lv_k is the k-th column of LV_{EEG} and p_{bk} is the spatial pattern of the k-th component for subject b. In addition, we also want to determine which EEG features in the spatial patterns have significant contributions to the corresponding temporal patterns. This is done by identifying EEG features that have weights statistically different from zero. To determine the group-level spatial pattern, for each significant component, we apply a two-tailed t-test to each element of the spatial patterns of all subjects with each group.

Results In this real data study, we apply the proposed method for corticomuscular activity modeling to the EEG and EMG features generated using the procedure described in Section 2.3.3 from 8 normal and 9 PD subjects simultaneously. The joint modeling of normal and PD data allows the identification of common temporal patterns across the groups. Meanwhile, the spatial patterns may be different across subjects, from which we could identify specific correlation connections that are differently recruited by PD subjects during the motor task.

Using the permutation test, two components were deemed significant ($P \leq 0.05$) (Fig. 2.7). Note that in the figure only connections whose weights are statistically different from zero are shown. The results based on real data from PD

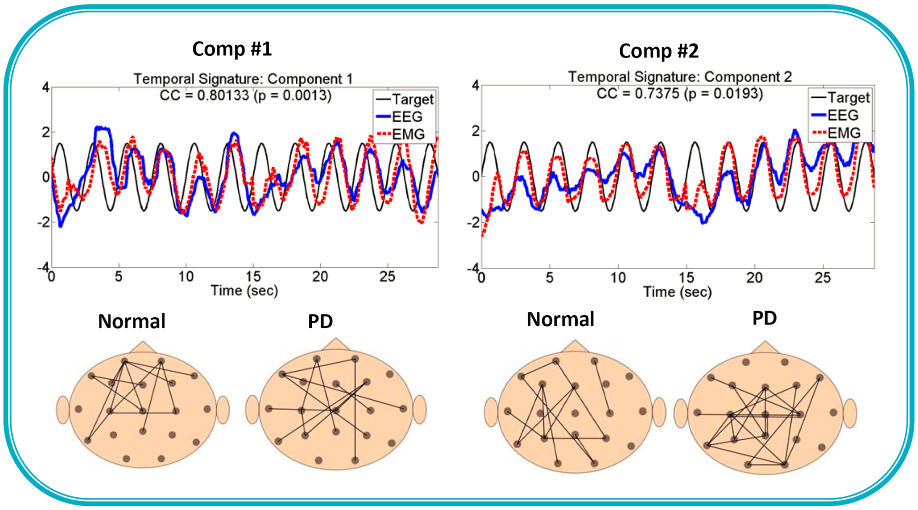


Figure 2.7: The two components from the proposed PLS+CCA method when using the EEG correlation features and the EMG amplitude features as data sets. Top panel: Temporal patterns of the EEG (blue, solid) and the EMG (red, dashed). The oscillation of the target bar is also shown (black, solid). Bottom panel: EEG spatial patterns of normal subjects (left) and PD subjects (right). The connections reflect the respective spatial patterns in the two groups. CC means correlation coefficient.

and normal subjects performing a dynamic motor task are promising. In the past, most EEG/EMG coupling studies have compared EEG activity at a specific locus (e.g. sensorimotor cortex contralateral to the hand performing the task) with the EMG during sustained contractions. However, we found that in normal subjects, correlations between the contralateral sensorimotor cortex and other regions are closely associated with ongoing EMG features during dynamic motor tasks (Fig. 2.7). It is likely that the dynamic nature of the task might require the recruitment of additional regions such as frontal regions for motor selection [89], contralateral (i.e. ipsilateral to hand movement) sensorimotor cortex for fine modulatory control [90] and occipital regions for post-error adaptations [91].

2.3.4 Discussion

From Fig. 2.7, we note similar connections between the PD and control groups, especially when comparing connections associated with each component, but we also note significant differences when comparing the PD and control groups. It is noted that PD subjects have increased connectivity between the frontal regions and central and sensorimotor cortices, compared with control subjects. This may reflect the enhanced effort required by PD subjects for motor task switching [92], a problem common in this PD population [93]. In addition, PD subjects have a significant connection between the left sensorimotor and occipital regions that is not present in the control group. We note that the connections with occipital regions are prominent in PD subjects. Compared to normal subjects, the PD subjects heavily rely on visual cues for the initiation [94] and ongoing control of movement [95]. Moreover, the increased intra and inter-hemispheric connections observed in the PD subjects are consistent with the findings in previous MEG studies [5].

Chapter 3

An IC-PLS Framework for Corticomuscular Coupling Analysis

3.1 Motivation and Objectives

In the previous chapter, we have presented the combined PLS+CCA method for corticomuscular coupling analysis. However, both PLS and CCA can only extract uncorrelated LVs and their interpretations may be difficult in real applications [38]. ICA is based on the notion that it is insufficient to consider only up to the second-order statistics (e.g. correlation and covariance) for obtaining a unique LV model [39] if the data are not strictly multivariate Gaussian. ICA assumes that the multivariate data are composed of a linear superposition of mutually statistically independent signal sources. In statistics, independence is a much stronger condition than uncorrelatedness. As mentioned in Section 2.1, both PLS and ICA have their advantages and disadvantages, but most importantly, these two methods can be considered complementary.

In this chapter, we propose combining their advantages and minimizing their drawbacks by formulating a multi-objective optimization problem to simultaneously incorporate response-relevance and independence into the regression procedure, to produce a so-called IC-PLS model. The proposed IC-PLS extracts LVs from both the measured data X and the response Y, keeping the LVs maximally independent and uniquely sorting the LVs in order of relevance. When applied to corticomuscular coupling analysis, the proposed IC-PLS can extract the most significant LV pairs from concurrent EEG and EMG data in an orderly manner. Furthermore, to infer consistent activation patterns across subjects within a group in the face of inter-subject variability, we also develop a group analysis framework based on the proposed IC-PLS model to accommodate the multi-subject case and achieve robust group inference.

3.2 Methods

In this section, we first formulate the corticomuscular coupling analysis problem as a multi-objective optimization problem. The basic idea is to extract highly correlated, but still maximally independent components in both the EEG and EMG so that the covariance between each of the extracted EEG and EMG components is maximized. Since the components are a result of an optimization based on a weighted combination of statistical independence and response-similarity goals, we then design some strategies to adjust each sub-objective's parameters simultaneously and make the corresponding sub-objectives change in parallel. Because of the importance of the initial solution, we also describe the initialization setting procedure.

3.2.1 A Multi-objective Optimization

To combine the advantages of PLS and ICA and design an overall optimization objective, the following conditions should be satisfied simultaneously:

First, PLS exploits the covariation between predictor variables and response variables and tries to find a new set of LVs that maximally relate them [76]. In other words, the covariance between the extracted LVs should be maximized as

$$\begin{aligned} &\max_{w_1,w_2} \quad \left(\mathrm{E}\left((w_1{}^T x)(w_2{}^T y)\right)\right)^2 \\ &\text{s.t.} \quad w_i{}^T w_i = 1, \quad i = 1,2. \end{aligned} \tag{3.1}$$

where w_i's $(i = 1, 2)$ are the weight vectors, x is the predictor vector with size $p \times 1$ and y is the response vector with size $q \times 1$. Here it is assumed that x and y are preprocessed (the specific procedure will be presented in Section 3.2.3).

Second, to incorporate the advantages of ICA, the extracted LVs should be as independent with each other as possible. According to Hyvärinen and Oja [39], this can be achieved by solving the following problems:

$$\begin{aligned} \max_{w_1} \quad & \left(\mathrm{E}(\mathrm{G}({w_1}^T x)) - \mathrm{E}(\mathrm{G}(u_1))\right)^2 \\ \text{s.t.} \quad & {w_1}^T w_1 = 1, \end{aligned} \tag{3.2}$$

and

$$\begin{aligned} \max_{w_2} \quad & \left(\mathrm{E}(\mathrm{G}({w_2}^T y)) - \mathrm{E}(\mathrm{G}(u_2))\right)^2 \\ \text{s.t.} \quad & {w_2}^T w_2 = 1, \end{aligned} \tag{3.3}$$

where u_1 and u_2 are standard Gaussian variables and $\mathrm{G}(\cdot)$ is a non-quadratic function. The following choices of $\mathrm{G}(\cdot)$ have been previously suggested:

$$\mathrm{G}_1(u) = \frac{1}{a_1} \log\cosh(a_1 u) \quad \text{and} \quad \mathrm{G}_2(u) = -\exp(-\frac{u^2}{2}) \tag{3.4}$$

where the constant a_1 is generally $1 \leq a_1 \leq 2$. In this chapter, we adopt $\mathrm{G}_1(\cdot)$ for $\mathrm{G}(\cdot)$ in the following optimization procedure.

To encapsulate the above three maximization objectives, an intuitive approach is to combine the three objectives into a single aggregate objective function and achieve a good trade-off between their respective goals. We therefore employ the well-known weighted linear sum of the sub-objectives, and propose the following multi-objective optimization problem:

$$\begin{aligned} \max_{w_1, w_2} \quad & \alpha \left[\mathrm{E}(({w_1}^T x)({w_2}^T y))\right]^2 \\ & + \beta \left[\mathrm{E}(\mathrm{G}({w_1}^T x)) - \mathrm{E}(\mathrm{G}(u_1))\right]^2 \\ & + \theta \left[\mathrm{E}(\mathrm{G}({w_2}^T y)) - \mathrm{E}(\mathrm{G}(u_2))\right]^2 \\ \text{s.t.} \quad & {w_i}^T w_i = 1, \quad i = 1, 2. \end{aligned} \tag{3.5}$$

where α, β and θ are the weights for the corresponding sub-objectives respectively, and satisfy $\alpha+\beta+\theta=1$. In Section 3.2.2, we will further discuss how to adjust these parameters in detail.

Based on the above optimization formulation, a solution can be calculated using an approximate Newton iteration approach with the detailed derivation being shown in Appendix B.2. At each iteration, w_i's are updated as:

$$\begin{aligned} & w_i \leftarrow w_i + d_i, \\ & \text{i.e.,} \quad w_i \leftarrow w_i - (\mathrm{J}\Phi(w_i))^{-1} \nabla \mathrm{F}_{w_i}, \quad i = 1,2. \end{aligned} \tag{3.6}$$

where $\mathrm{J}\Phi(w_i)$ is the Jacobian matrix and $\nabla \mathrm{F}_{w_i}$ is the first-order derivative of $\mathrm{F}(w_1, w_2, \lambda_1, \lambda_2)$ defined in Equation (B.20) with respect to w_i.

3.2.2 Determining the Optimization Weights

The primary goal during the multi-objective optimization is to simultaneously optimize two or more separate objectives subject to certain constraints. Due to the complicated and possibly conflicting nature of the sub-objectives, the overall optimal solution obtained depends on the relative values of the weights specified (i.e., α, β and θ here). Therefore, to achieve a good trade-off between different sub-objectives, it is important to determine the weights appropriately. Here two key issues should be taken into consideration: the different scales and the different convergence speeds of the sub-objectives [96].

The following strategy was employed to avoid the undesirable situation in which one sub-objective overwhelms the others: the aforementioned weights α, β and θ are decomposed into three factors respectively: $\alpha_{sig} \cdot \alpha_{scale} \cdot \alpha_{adj}$, $\beta_{sig} \cdot \beta_{scale} \cdot \beta_{adj}$ and $\theta_{sig} \cdot \theta_{scale} \cdot \theta_{adj}$. Here α_{sig}, β_{sig} and θ_{sig}, called significance factors, denote the relative significance attached to each sub-objective. They can be set subjectively according to the specific application. α_{scale}, β_{scale} and θ_{scale}, named scale factors, are adopted to unify the scales of the three sub-objectives. Just as their names imply, the scale factors are defined as follows:

$$\alpha_{scale} = \frac{1}{|\mathrm{F}_{1,w_1 w_2}|},\ \beta_{scale} = \frac{1}{|\mathrm{F}_{2,w_1}|},\ \theta_{scale} = \frac{1}{|\mathrm{F}_{3,w_2}|}, \tag{3.7}$$

where the denominators are the absolute values of the corresponding sub-objective functions defined as

$$
\begin{aligned}
\mathrm{F}_{1,w_1w_2} &= \left(\mathrm{E}((w_1{}^T x)(w_2{}^T y))\right)^2 + \kappa_{11}(w_1{}^T w_1 - 1) \\
&\quad + \kappa_{12}(w_2{}^T w_2 - 1), \\
\mathrm{F}_{2,w_1} &= \left(\mathrm{E}(\mathrm{G}(w_1{}^T x)) - \mathrm{E}(\mathrm{G}(u_1))\right)^2 + \kappa_2(w_1{}^T w_1 - 1), \\
\mathrm{F}_{3,w_2} &= \left(\mathrm{E}(\mathrm{G}(w_2{}^T y)) - \mathrm{E}(\mathrm{G}(u_2))\right)^2 + \kappa_3(w_2{}^T w_2 - 1),
\end{aligned}
\tag{3.8}
$$

where κ_{11}, κ_{12}, κ_2 and κ_3 are Lagrange multipliers. α_{adj}, β_{adj} and θ_{adj}, called adjustable factors, are used to balance their different convergence speeds and can be updated in real-time during the iterative searching procedure. In general, the gradient of a function can be employed to estimate speed of convergence. Therefore, the adjustable factors can be defined as below:

$$
\begin{aligned}
\alpha_{adj} &= \frac{1}{(\|\nabla \mathrm{F}_{1,w_1}\| + \|\nabla \mathrm{F}_{1,w_2}\|)/2} \\
\beta_{adj} &= \frac{1}{\|\nabla \mathrm{F}_{2,w_1}\|},\ \theta_{adj} = \frac{1}{\|\nabla \mathrm{F}_{3,w_2}\|}
\end{aligned}
\tag{3.9}
$$

where the denominators are the L^2 norm of the corresponding first-order derivatives, which are calculated in a similar way to the calculation of $\nabla \mathrm{F}_{w_i}$ $(i = 1,2)$ in Appendix:

$$
\begin{aligned}
\nabla \mathrm{F}_{1,w_1} &= 2\mathrm{E}((w_1{}^T x)(w_2{}^T y))\mathrm{E}(x(w_2{}^T y)) \\
&\quad - 2w_1{}^T \mathrm{E}((w_1{}^T x)(w_2{}^T y))\mathrm{E}(x(w_2{}^T y))w_1 \\
\nabla \mathrm{F}_{1,w_2} &= 2\mathrm{E}((w_1{}^T x)(w_2{}^T y))\mathrm{E}(y(w_1{}^T x)) \\
&\quad - 2w_2{}^T \mathrm{E}((w_1{}^T x)(w_2{}^T y))\mathrm{E}(y(w_1{}^T x))w_2 \\
\nabla \mathrm{F}_{2,w_1} &= 2(\mathrm{E}(\mathrm{G}(w_1{}^T x)) - \mathrm{E}(\mathrm{G}(u_1)))\mathrm{E}(x\mathrm{g}(w_1{}^T x)) \\
&\quad - 2w_1{}^T (\mathrm{E}(\mathrm{G}(w_1{}^T x)) - \mathrm{E}(\mathrm{G}(u_1)))\mathrm{E}(x\mathrm{g}(w_1{}^T x))w_1 \\
\nabla \mathrm{F}_{3,w_2} &= 2(\mathrm{E}(\mathrm{G}(w_2{}^T y)) - \mathrm{E}(\mathrm{G}(u_2)))\mathrm{E}(y\mathrm{g}(w_2{}^T y)) \\
&\quad - 2w_2{}^T (\mathrm{E}(\mathrm{G}(w_2{}^T y)) - \mathrm{E}(\mathrm{G}(u_2)))\mathrm{E}(y\mathrm{g}(w_2{}^T y))w_2.
\end{aligned}
\tag{3.10}
$$

where $\mathrm{g}(\cdot)$ represents the first-order derivative of $\mathrm{G}(\cdot)$. From the above definitions, we can see that α_{adj}, β_{adj} and θ_{adj} will be adjusted online during the optimization procedure. If one sub-objective changes faster, i.e., its gradient norm is larger, the corresponding adjustable factor will be smaller and vice versa. This makes all sub-objectives change in parallel. However, when the iteration results are close to the optima, the gradients will approach zero. Therefore we set the three weights as below:

$$\begin{cases} \alpha = \alpha_{sig} \cdot \alpha_{scale} \cdot \alpha_{adj}, & \alpha_{adj} \leq TH \\ \beta = \beta_{sig} \cdot \beta_{scale} \cdot \beta_{adj}, & \beta_{adj} \leq TH \\ \theta = \theta_{sig} \cdot \theta_{scale} \cdot \theta_{adj}, & \theta_{adj} \leq TH \\ \alpha = \alpha_{sig} \cdot \alpha_{scale}, & \alpha_{adj} > TH \\ \beta = \beta_{sig} \cdot \beta_{scale}, & \beta_{adj} > TH \\ \theta = \theta_{sig} \cdot \theta_{scale}, & \theta_{adj} > TH \end{cases} \tag{3.11}$$

where TH is a predefined small threshold that envelops a narrow range around the optimum. Whenever the searching route enters the range, the corresponding gradient approximates zero so that the weights should be determined only by the first two factors. In short, the final weights take into account the effects of all the three factors.

3.2.3 Initialization Setting

When solving optimization problems, the final solutions not only depend upon the search algorithm employed, but also upon the initial setting, which determines the searching path and starting point respectively. In the proposed IC-PLS model, to prevent the optimization being stuck in an uninformative local minimum, we introduce a joint statistical analysis method [70, 71] and use the extracted LVs as initial conditions for the subsequent iterative search.

Suppose the original predictor and response data are denoted by Z_1 (with size $N \times P$) and Z_2 (with size $N \times Q$) respectively and define the sub-latent variables (subLVs) in each data space to be linear combinations of the original variables, i.e., $Z_1 v_1, Z_2 v_2$. One super latent variable (supLV) t_g is designed to relate the subLVs by solving the following optimization problem:

$$\begin{aligned} \max \quad & ({t_g}^T Z_1 v_1)^2 + ({t_g}^T Z_2 v_2)^2 \\ \text{s.t.} \quad & {t_g}^T t_g = 1, \quad {v_i}^T v_i = 1, \quad \forall i = 1,2 \end{aligned} \tag{3.12}$$

with all data assumed to be zero-mean and normalized to unit variance in advance. The subLV $Z_i v_i$ in each data space carries its associated variation information and $({t_g}^T Z_i v_i)^2$ models the covariance information between the subLV $Z_i v_i$ and the supLV t_g. The supLV t_g relates all subLVs simultaneously and actually plays a role as a link. It is expected that, by solving the above optimization problem, the extracted subLVs will carry as much variation as possible in each data space and at the same time be correlated as closely as possible.

By the method of Lagrange multipliers, the supLV and subLVs can be readily derived as:

$$(Z_1 {Z_1}^T + Z_2 {Z_2}^T) t_g = \rho_g t_g, \tag{3.13}$$

$$t_i = Z_i v_i = \sqrt{\frac{1}{\rho_i}} Z_i {Z_i}^T t_g, \tag{3.14}$$

where $\rho_i = {t_g}^T Z_i {Z_i}^T t_g$. Thus with Lagrange multipliers the optimization problem is now characterized as a standard algebra problem.

Several supLVs, represented by T_g and ranked by descending ρ_g's can be extracted and a same number of subLVs, represented by T_i, can be readily calculated by using Eq. (3.14):

$$T_i = Z_i {Z_i}^T T_g D_i, \tag{3.15}$$

where D_i is a diagonal matrix with corresponding $\sqrt{\frac{1}{\rho_i}}$'s as its diagonal elements. In practice, we need to specify the number of supLVs to be extracted, which can be done by extracting a sufficient number to explain an adequate fraction of the variance (e.g. 90%).

From the above derivation, it is apparent that the supLVs T_g are actually the principal components (PCs) of the concatenated data shown in Equation (3.13) and the subLVs T_i are extracted by regressing each data space on the supLVs as shown in Equation (3.15). Hence, the supLVs can represent the general systematic variations for both data spaces and the subLVs can capture the local systematic variations within each data space. However, collinearity may exist in the subLVs calculated through the above procedure since each data set is used repetitively as shown in Equation (3.14). Furthermore, the subLVs are not sorted according to the correlation of the corresponding between-set subLV pairs. To efficiently implement the subsequent Newton iteration, we need to address the above concerns and therefore modify the original subLVs as follows:

First, find the maximal correlation and the corresponding supLV t_g^* as well as subLVs t_1^*, t_2^*;

Second, exclude t_g^* from T_g and update each data space using t_1^*, t_2^* in a deflation manner;

Third, update the original subLVs based on the new supLVs and the updated data spaces by using Equation (3.15).

After repeating these steps, it is ensured that the modified subLVs (MsubLVs) are orthogonal to each other and automatically ordered. The extracted MsubLVs are then used as the initial solutions for the proposed multi-objective optimization problem.

3.2.4 A Group Analysis Framework

In many medical applications, the desire to make group inferences requires recruitment of an adequate number of subjects and the application of a suitable group analysis which accommodates inter-subject variability. Here we propose a group analysis framework for modeling corticomuscular coupling activity, and apply the proposed group analysis framework to the EEG/EMG data collected from normal subjects and subjects with PD during a motor task.

As illustrated in Fig. 3.1, there are two stages in this framework. Suppose we have B subjects in total (e.g. $B = 17$ in this study) and two types of data sets for each subject. In the first stage, the subLVs T_{bj}'s ($b = 1, 2, ..., B$; $j = 1, 2$) for each subject are calculated by using Equation (3.13) and Equation (3.15). Note that these subLVs include most of the useful information for further analysis, and there is no need to use the modified algorithm here because the collinearity and ordering problems will be solved together in the initialization step of the group level IC-PLS model.

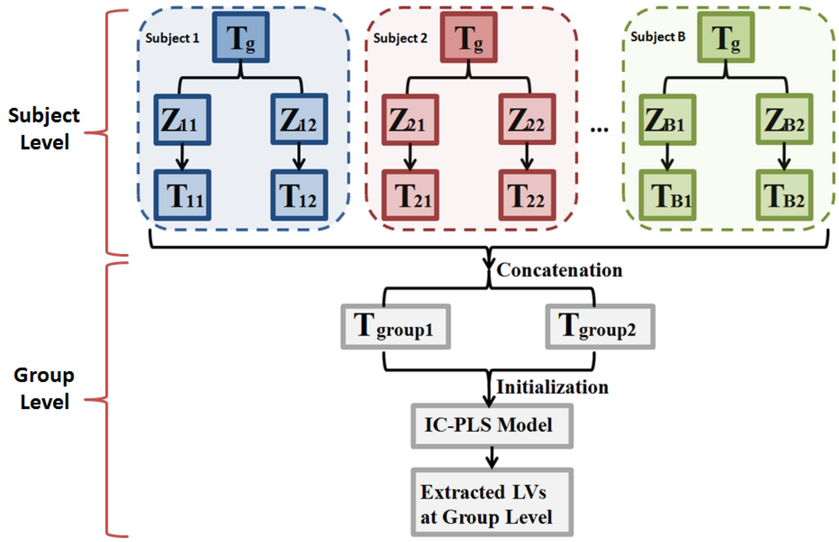


Figure 3.1: The diagram of the group analysis framework based on the proposed IC-PLS model.

In the second stage, all subjects' subLVs are correspondingly concatenated as:

$$T_{group\,j} = [T_{1j}, T_{2j}, ..., T_{Bj}], \quad \forall j = 1, 2, \tag{3.16}$$

with the assumption that all subjects share common temporal group patterns [82]. Then the whole initialization procedure described in Section 3.2.3 is applied to the concatenated data $T_{group\,j}$. Finally, the proposed IC-PLS model is employed to process the initialized data sets and the expected LVs could be extracted to represent group patterns.

3.3 Data Processing and Results

3.3.1 Simulation

Synthetic Data

In this simulation, we apply the proposed IC-PLS model to synthetic data and also implement PLS and ICA for comparison. As an illustrative example, without loss of generality, four sources are generated and analyzed, similar to that used by Hakyin [98].

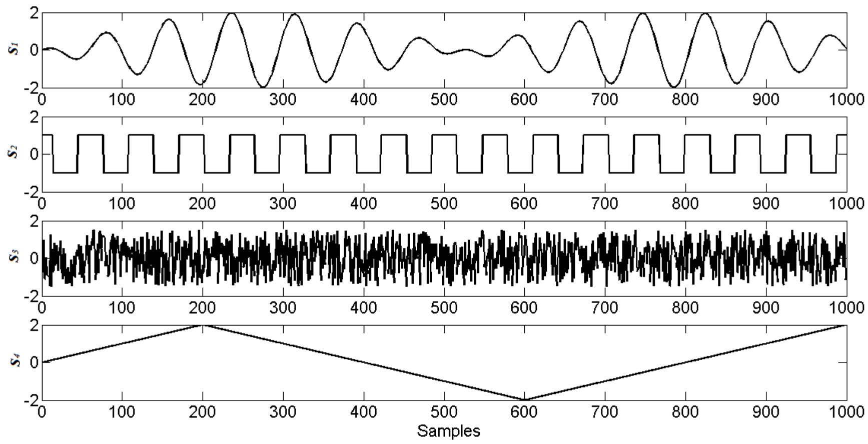


Figure 3.2: The four source signals.

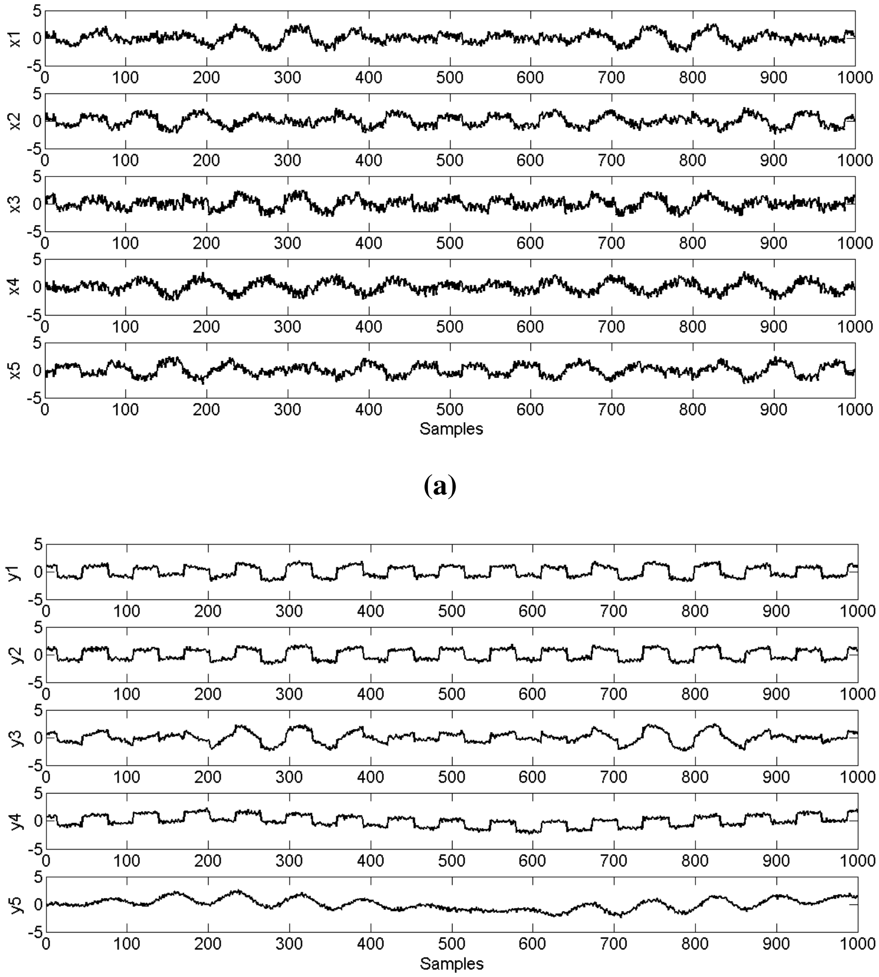
The following four source signals are considered:

$$
\begin{aligned}
s_1 &= 2\cos(0.08t)\sin(0.006t) \\
s_2 &= \text{sign}(\sin(0.3t) + 3\cos(0.1t)) \\
s_3 &= \text{uniformly distributed noise in the range } [-1.5, 1.5] \\
s_4 &= \begin{cases} 0.01t & 1 \leq t \leq 200 \\ -0.01t + 4 & 201 \leq t \leq 600 \\ 0.01t + 8 & 601 \leq t \leq 1000 \end{cases}
\end{aligned}
\tag{3.17}
$$

where t denotes the time point index, valued from 1 to 1000, and s_i's $(i = 1, 2, 3, 4)$ represented four simulated sources, as shown in Fig. 3.2.

Two mixed data sets X and Y, shown in Fig. 3.3, were generated as follows:

$$x = As_x \quad \text{and} \quad y = Bs_y \tag{3.18}$$



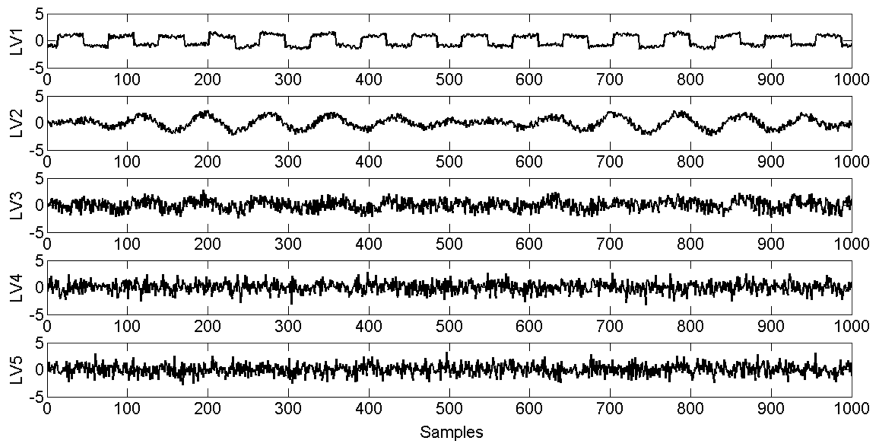
(a)

(b)

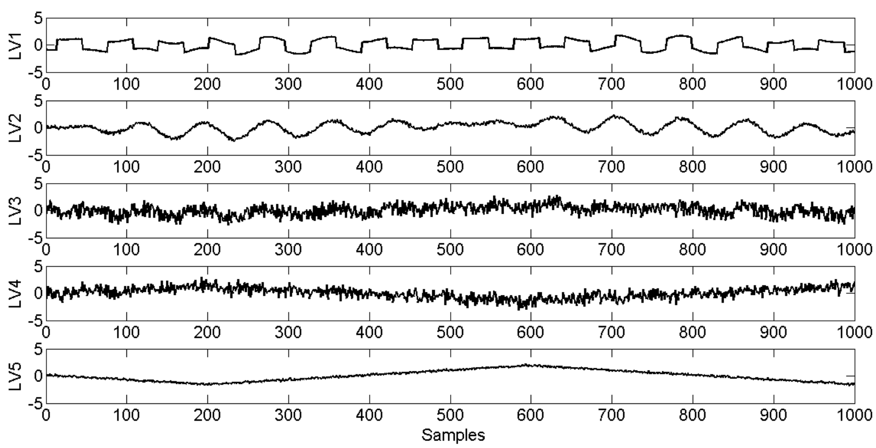
Figure 3.3: (a) The mixed data X. (b) The mixed data Y.

where $s_x = [s_1\, s_2\, s_3]^T$ and $s_y = [s_1\, s_2\, s_4]^T$ with

$$A = \begin{bmatrix} 0.86 & 0.79 & 0.67 \\ -0.55 & 0.65 & 0.46 \\ 0.17 & 0.32 & -0.28 \\ -0.33 & 0.12 & 0.27 \\ 0.89 & -0.97 & -0.74 \end{bmatrix} \tag{3.19}$$



(a)

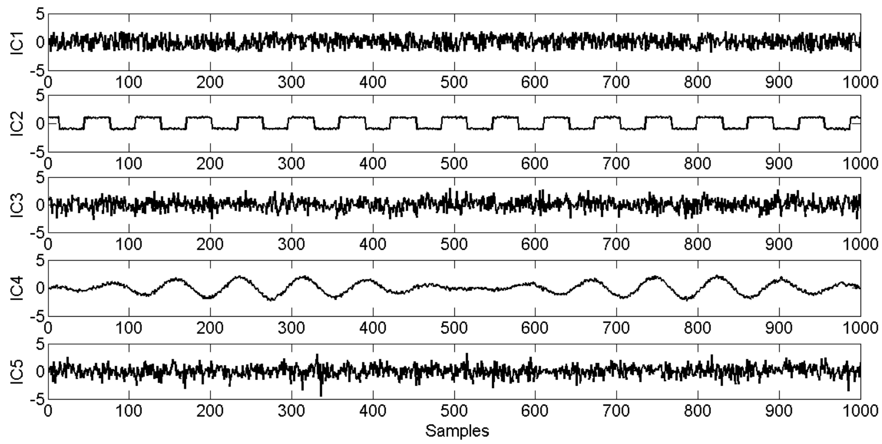


(b)

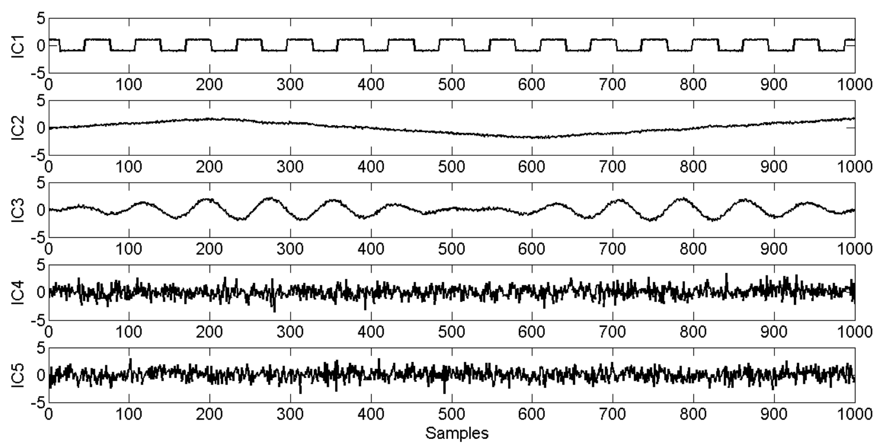
Figure 3.4: (a) The LVs estimated in X using PLS. (b) The LVs estimated in Y using PLS.

$$B = \begin{bmatrix} 0.21 & 0.69 & 0.05 \\ 0.22 & 0.98 & 0.03 \\ 0.68 & 0.74 & 0.10 \\ 0.05 & 0.63 & 0.42 \\ 0.83 & 0.10 & 0.79 \end{bmatrix} \tag{3.20}$$

Here x, y, s_x and s_y are all column vectors, denoting one observation in their respective data space.



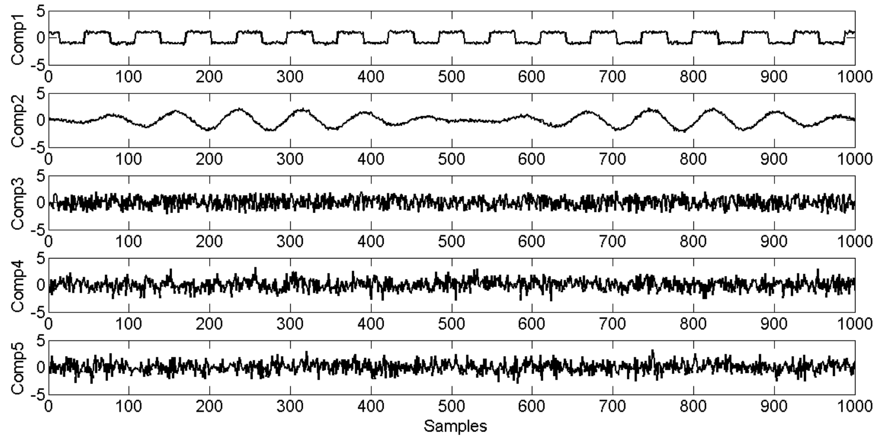
(a)



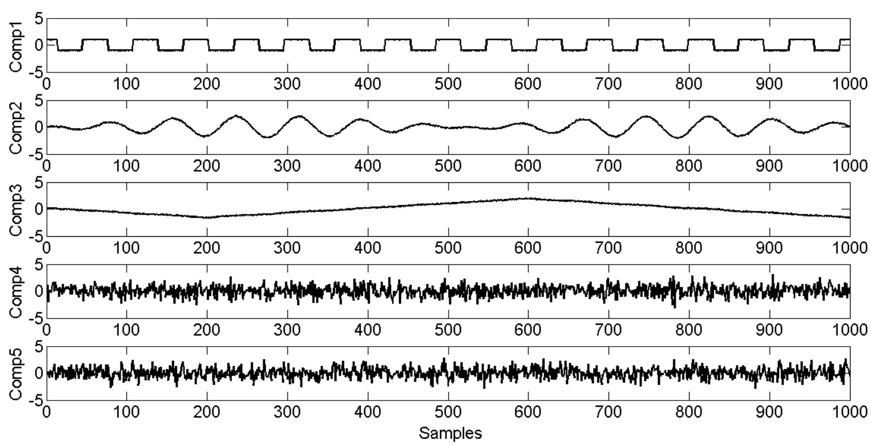
(b)

Figure 3.5: (a) The LVs estimated in X using ICA. (b) The LVs estimated in Y using ICA.

The sources s_1 and s_2 exist in both data spaces X and Y, representing common information. The source s_3 only contributes to X and s_4 only to Y, representing their own unique information. In Y, we intentionally assigned a relatively high weight to the source s_2. Moreover, different white Gaussian noise with 5% power was added to each source in each data space.



(a)



(b)

Figure 3.6: (a) The LVs estimated in X using IC-PLS. (b) The LVs estimated in Y using IC-PLS.

Results

The extracted components using PLS, ICA and the proposed IC-PLS model are shown in Figs. 3.4, 3.5 and 3.6 separately. The LVs extracted by PLS are automatically ordered in terms of their significance and successfully reflected the corresponding relationships of the underlying sources between X and Y. However, compared with the original sources, the extracted LVs are distorted, suggesting that the fact that the PLS sources are uncorrelated are insufficient to accurately recover the underlying sources. In contrast, ICA recovers the original sources accurately in both data spaces, but it does not relate the two data spaces and the LVs are unordered. When the proposed IC-PLS model is employed, the dominant sources which make significant contributions to both data spaces are accurately identified and ordered, with the focus entirely on sources that are common across the predictor and the response.

3.3.2 Real Data

The experimental data sets have been described in Section 2.2. The concurrently collected EEG and EMG from controls and PD subjects are utilized here.

Feature Extraction

The basic idea is similar to that in Section 2.3.3. We examined the coupling relationships between time-varying EEG features and amplitudes of the EMG signals, constituting Z_{b1} and Z_{b2} respectively for Subject b.

EEG Features Two types of EEG features were considered: pair-wise Pearson's correlation [86] and band-limited, pair-wise EEG coherences [97]. Pearson's correlation has been defined in Section 2.3.3 and can be referred to Equation (2.9). EEG coherence can also be used to examine the relation between two EEG signals, which includes both power and frequency information. The calculation can be referred to Equation (1.1). In this chapter, we calculated both the time-varying pair-wise correlations and coherences between EEG channels, using a Hamming window with length 300 and with a 95% overlap. To be consistent with common EEG practice, we further divide the coherence into four frequency bands, i.e., theta band (4-8 Hz), alpha band (8-13 Hz), beta band (13-30 Hz) and gamma band (30-70 Hz). The time-varying energy signals in each band were then used as the EEG features. Therefore, the raw EEG information can be represented by a two-dimensional matrix with size $N \times M$, where the rows correspond to the samples at different time points and the columns correspond to the features, i.e. pair-wise correlations or band-limited coherences between the EEG channels.

EMG Features In this study, we still employ the root-mean-square approach described in Equation (2.10) to calculate the EMG amplitude of short-duration EMG signals by a moving window with length $n = 300$ and a 95% overlap, the same as in the EEG feature calculation, to ensure that the obtained EEG and EMG features are temporally aligned and matched.

Regarding significance assessment and spatial pattern extraction, we could refer to Section 2.3.3.

Results

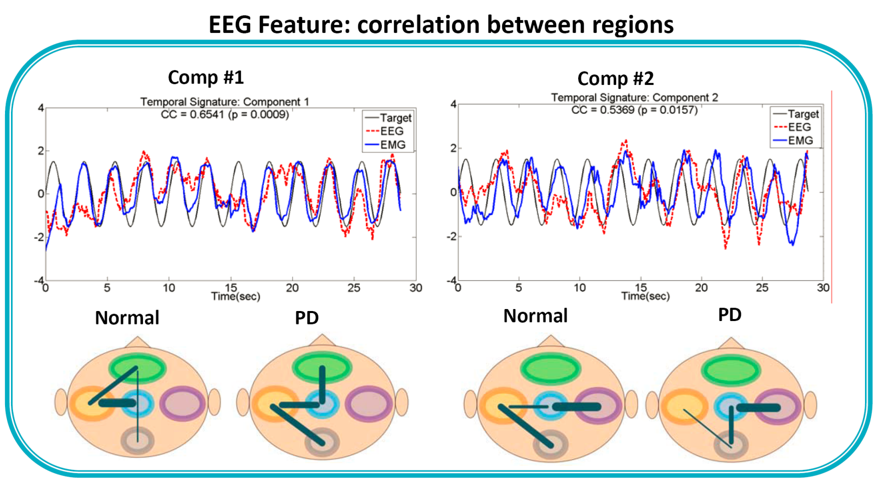


Figure 3.7: The two components of the group analysis framework when using the EEG correlations and the EMG amplitude data sets. Top panel: Temporal patterns of the EEG (red, dashed) and the EMG (blue, solid). The oscillation of the target bar is also shown (black, solid). Bottom panel: EEG spatial patterns of normal subjects (left) and PD subjects (right). The connections reflect the respective spatial patterns in the two groups and the width is made proportional to the weighting of the respective connections in the group EEG spatial pattern. CC means correlation coefficient.

When the proposed group analysis framework is applied to EEG and EMG

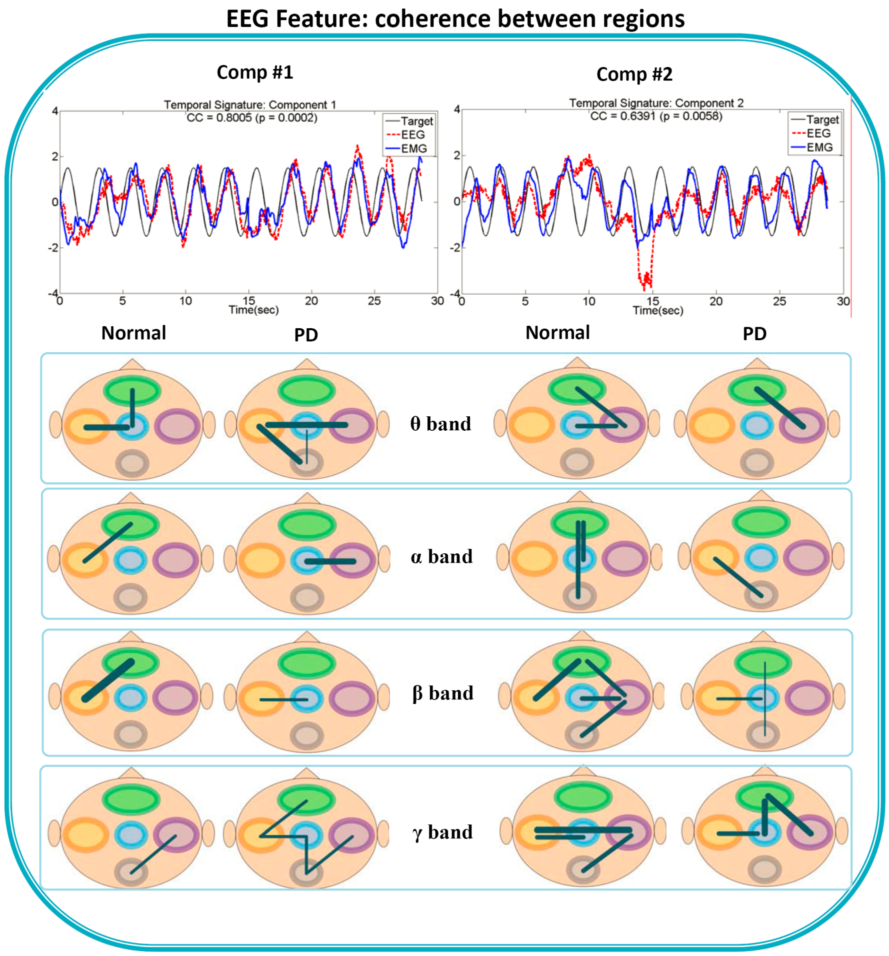


Figure 3.8: The two components of the group analysis framework when using the EEG coherence and the EMG amplitude data sets.

features generated using the procedure described in Section 3.3.2 from all 8 normal and 9 PD subjects simultaneously, the identification of common temporal patterns, but different spatial patterns could be identified. Using the permutation test, two components are deemed significant ($P \leq 0.05$) (Figs. 3.7 and 3.8). Note that in the figures only connections whose weights are statistically significantly different from zero are shown.

When using the correlation between brain regions as the EEG feature, many connections between are similar between PD and normals, especially when connections seen in either components are considered. The only difference is a connection between the L sensorimotor and frontal regions that is present in controls but missing in PD (Fig. 3.7).

When the pairwise EEG coherence features are used to couple with the EMG features, there are also two components deemed significant via the permutation test (Fig. 3.8). Again, when considering across components, there are similarities in the connections between PD and controls, but also significant differences. In the beta band, normal subjects have connections between the R sensorimotor cortex and occipital, central and frontal regions that are completely missing in PD subjects. In the gamma band, PD subjects have increased connectivity between the frontal regions and central and sensorimotor cortices that are missing in controls. Finally, PD subjects have a significant connection between the L sensorimotor and occipital regions in the theta and alpha bands that is not present in controls.

3.4 Discussion

In this chapter, we have proposed a new method to assess EEG/EMG coupling based on combining the favourable properties of ICA and PLS. The simulations support utilization of proposed method, and results based on real data from Parkinson's and normal control subjects performing a dynamic motor task are promising.

The results we observed in the PD and control subjects are consistent with prior studies describing the changes in EEG observed in PD. When using the correlation between brain regions as EEG features, we found widespread connectivity between central regions and the ipsilateral (right) sensorimotor regions and ongoing EMG activity. A body of literature suggests that widespread oscillatory disruption in PD is detectable in the EEG during visually guided tasks (e.g. [100]). This likely represents compensatory expansion of cortical regions to overcome deficiencies in the basal ganglia to perform motor tasks, as has been observed in fMRI studies [101]. The lack of fronto-central alpha activity seen in PD subjects compared to controls may reflect impaired spatial attention [102], as impairment of maintenance of attention has been well described in PD subjects (e.g. [115]). The increased frontal connectivity we observed in the gamma region in PD may reflect the enhanced effort PD subjects required for motor task switching [92], a problem common in this population [93].The enhanced occipital connectivity seen in PD in the theta and beta bands likely relates to the fact that PD subjects heavily rely on visual cues for the initiation [94] and ongoing control of movement [95].

We are able to find common temporal patterns of corticomuscular coupling in both PD and controls despite having different spatial origins of this common signal. We note that our results are consistent with a recent study that suggests the magnitude of traditional MSC is not significantly different early in PD [106]. However, we have shown that the distributed EEG connectivity patterns required to generate these common temporal patterns differ between PD and controls. One could envision situations where the spatial extent is similar, but the temporal profiles are significantly different (e.g. focal stroke affecting the sensorimotor cortex contralateral to the hand performing the task) which may not be well-served by the proposed approach. However, even in this scenario, compensatory mechanisms would still presumably modify the connectivity patterns giving rise to corticomuscular coherence, and this would still be captured by the current approach.

A limitation of our approach is that it treats corticomuscular interactions as bidirectional; however, in addition to the cortex driving the muscle, there may be proprioceptive information from the muscle to the cortex. This reflects the inherent limitation of the PLS step, however recent work on directed partial least squares (e.g. [107]), may provide future directions to extend the current approach. Another possible refining is that the cross-validation and bootstrap method may be used to test the robustness and infer the population level information.

Chapter 4

A Joint Multimodal Group Analysis Framework for Modeling Corticomuscular Activity

4.1 Motivation and Objectives

In the previous two chapters, we have proposed two novel coupling analysis methods. However, in both methods, only two types of data sets can be processed at the same time, while in many cases more than two types of data sets can be available and a better understanding could be achieved from analyzing multimodal data jointly. To address the above concern, in this chapter, we design a joint multimodal statistical framework (JMSF) for corticomuscular coupling analysis to relate multiple data sets.

Unlike previous approaches where only two data sets are considered and the interest is to interpret one data set by the other in a unidirectional fashion, the proposed framework models multiple data spaces simultaneously in a multidirectional fashion. Here, we extend the "bidirection" concept in [108] to be "multidirection" to accommodate multimodal cases. The framework has a two-step modeling strategy. In the first step, a multidirectional LV extraction solution (denoted as multi-LV extraction) is established for preliminary LV preparation. It is formulated by solving a simple optimization problem. In the second step, a joint postprocessing is performed on the extracted LVs to acquire common and specific information in each data space. Furthermore, to address the challenging issue of inter-subject variability in biomedical applications where the problem is to infer certain activation patterns consistently shared within a population, we develop a group analysis architecture based on the proposed JMSF to accommodate the multi-subject case and to summarize the information from each individual subject's data to achieve group inference.

4.2 Methods

4.2.1 The Joint Multimodal Statistical Framework

The main contributions of the proposed method lie in two folds: (1) it changes the traditional unidirectional regression fashion to a multidirectional fashion where data spaces are explored under the supervision of each other and (2) it models multiple data sets/signals simultaneously overcoming the traditional constraint of having only two types of data sets. In the following subsections, the two-step modeling strategy of the proposed JMSF, which is illustrated in Fig. 4.1, is described in details.

Multi-LV Extraction

Suppose we have m data sets X_1, which is with size $N \times P_1$, X_2 with size $N \times P_2$, ..., and X_m with size $N \times P_m$, and we define the sub-latent variables (denoted as subLVs) in each data space to be linear combinations of the original variables, i.e., $X_1 w_1, X_2 w_2, \ldots, X_m w_m$. One super latent variable (denoted as supLV) t_g is designed to relate the subLVs, and it can be obtained by solving the following optimization problem:

$$\begin{aligned} \max_{t_g, w_i} \quad & \sum_{i=1}^{m} ({t_g}^T X_i w_i)^2, \\ \text{s.t.} \quad & {t_g}^T t_g = 1, \quad {w_i}^T w_i = 1, \quad \forall i = 1, 2, \ldots, m. \end{aligned} \tag{4.1}$$

All columns of the m data matrices are assumed to be zero-mean and normalized to unit variance in advance. The subLV $X_i w_i$ for the ith data space carries its associated variation information and $({t_g}^T X_i w_i)^2$ models the covariance information between each subLV $X_i w_i$ and the supLV t_g. The supLV t_g relates all subLVs simultaneously and actually plays a role as a link bridge. By solving the above optimization problem, it is expected that the extracted subLVs should carry as many variations as possible in each data space and at the same time be correlated to each other as closely as possible.

By employing the method of Lagrange multipliers, we rewrite the initial cost function as:

$$L_{t_g,w_i} = \sum_{i=1}^{m} ({t_g}^T X_i w_i)^2 - \lambda_g({t_g}^T t_g - 1) - \sum_{i=1}^{m} \lambda_i({w_i}^T w_i - 1), \tag{4.2}$$

where λ_g and λ_i's are Lagrange multipliers.

Taking the derivatives of L_{t_g,w_i} with respect to t_g and w_i's and setting them to be zero, we have:

$$\frac{\partial L_{t_g,w_i}}{\partial t_g} = 2\sum_{i=1}^{m} ({t_g}^T X_i w_i) X_i w_i - 2\lambda_g t_g = 0, \tag{4.3}$$

$$\frac{\partial L_{t_g,w_i}}{\partial w_i} = 2({t_g}^T X_i w_i) {X_i}^T t_g - 2\lambda_i w_i = 0. \tag{4.4}$$

By left multiplying Equation (4.3) with ${t_g}^T$ and Equation (4.4) with ${w_i}^T$, we can easily derive the following equations:

$$\sum_{i=1}^{m} ({t_g}^T X_i w_i)^2 = \lambda_g, \tag{4.5}$$

$$({t_g}^T X_i w_i)^2 = \lambda_i, \quad \forall i = 1, 2, ..., m. \tag{4.6}$$

The above equations indicate that λ_g is the underlying optimization objective and λ_i's stand for the suboptimal objective parameters, which is actually equivalent to the following relation:

$$\sum_{i=1}^{m} \lambda_i = \lambda_g. \tag{4.7}$$

According to Equations (4.5) and (4.6), we can modify Equations (4.3) and (4.4) as follows:

$$\sum_{i=1}^{m} \sqrt{\lambda_i} X_i w_i = \lambda_g t_g, \tag{4.8}$$

$$\frac{1}{\sqrt{\lambda_i}} {X_i}^T t_g = w_i, \quad \forall i = 1,2,...,m. \tag{4.9}$$

Then, the above equations can be combined as

$$(\sum_{i=1}^{m} X_i {X_i}^T) t_g = \lambda_g t_g, \tag{4.10}$$

which means that the optimization problem is now transferred to be a standard algebra problem. The solution can be easily found through the eigenvalue decomposition of the matrix $(\sum_{i=1}^{m} X_i {X_i}^T)$.

After t_g is calculated, the suboptimal objective parameters λ_i's can be obtained by combining Equations (4.6) and (4.9):

$$ {t_g}^T X_i {X_i}^T t_g = \lambda_i, \quad \forall i = 1,2,...,m. \tag{4.11}$$

Finally, in terms of Equations (4.9) and (4.11), the subLVs can be expressed as

$$t_i = X_i w_i = \sqrt{\frac{1}{{t_g}^T X_i {X_i}^T t_g}} X_i {X_i}^T t_g, \quad \forall i = 1,2,...,m. \tag{4.12}$$

Usually, using Equation (4.10), several supLVs t_g's, denoted by the matrix T_g, can be derived according to the descending λ_g's. For each data set X_i, the same number of subLVs, denoted by T_i, can be readily calculated by Equation (4.12), rewritten as

$$T_i = X_i {X_i}^T T_g D_i, \quad \forall i = 1,2,...,m, \tag{4.13}$$

where D_i is a diagonal matrix with corresponding $\sqrt{\frac{1}{\lambda_i}}$'s being its diagonal elements. A practical issue is to determine the number of supLVs. In our study, we determine the number by setting a threshold that corresponds to the ratio of explained variance (e.g. 90%).

From the entire derivation process, we can see that the supLVs T_g are actually the principle components (PCs) in the concatenation data space, as shown in Equation (4.10), and the subLVs T_i's are extracted by regressing each data set onto the supLVs, as shown in Equation (4.13). Hence, the supLVs can represent the general systematic variations for all data spaces and the subLVs T_i's can capture the local systematic variations in the ith data space. However, the collinearity problem may exist in the subLVs calculated through the above procedure since each data set is used repetitively as shown in Equation (4.12). The extracted subLVs are not necessarily orthogonal to each other. Furthermore, in practice, we are interested in identifying the most highly correlated components across multiple data sets. However, the subLVs derived above are not sorted according to the average correlation coefficient (acc) between corresponding subLV pairs, i.e. $acc(l) = \sum_{\substack{i,j=1 \\ i \neq j}}^{m} cc(t_{i,l}, t_{j,l})/(m(m-1)/2)$. To effectively implement the joint post-processing step which is to be described shortly, we need to address these orthogonality and sorting issues and thus we design a modified algorithm as described in Algorithm 2. In Algorithm 2, it can be ensured that the modified subLVs (MsubLVs) for each data space are orthogonal to each other and are automatically ordered according to the descending average correlation.

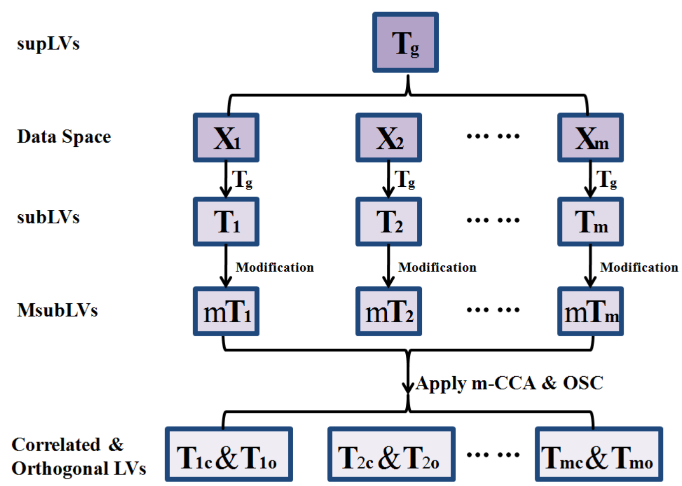


Figure 4.1: The diagram of the joint multimodal statistical framework.

Algorithm 2 Multi-LV Extraction

Input: multiple data sets X_1 with size $N \times P_1$, X_2 with size $N \times P_2$, ..., and X_m with size $N \times P_m$
Output: corresponding MsubLVs matrices mT_1, mT_2, ..., mT_m

Set $count = L$, and extract L supLVs $T_g = [t_{g,1}, t_{g,2}, ..., t_{g,L}]$ from the concatenated data space $[X_1, X_2, ..., X_m]$ by using Equation (4.10).
Calculate the original subLVs $T_i = [t_{i,1}, t_{i,2}, ..., t_{i,L}], \quad \forall i = 1, 2, ..., m$ within each data space by Equation (4.13).
Initialize all MsubLVs matrices to be empty, i.e., $mT_i =$[ ], $i = 1, 2, ..., m$.
while $count > 0$ **do**
  For $l = 1, 2, ...L$, calculate the correlation coefficients (cc) between each pair of subLVs (e.g. $cc(t_{i,l}, t_{j,l}), i \neq j$, l denotes the l-th subLV).
  For each l, compute the average correlation coefficient (acc), i.e., $acc(l) = \sum_{\substack{i,j=1 \\ i \neq j}}^{m} cc(t_{i,l}, t_{j,l})/A$, where $A = m(m-1)/2$ is the total number of possible unique pairs.
  Find the maximum acc and the corresponding supLV t_g^* as well as the subLVs $t_1^*, t_2^*, ..., t_m^*$.
  Set $mT_i =[mT_i \ t_i^*]$, $i = 1, 2, ..., m$
  Deflate X_i's by substracting the effects of the corresponding subLV from each data space as follows:
  for $i = 1$ to m **do**
    $p_i{}^T = (t_i^{*T} t_i^*)^{-1} t_i^{*T} X_i$
    $E_i = X_i - t_i^* p_i{}^T$
  end for
  Exclude t_g^* from T_g to obtain an updated supLVs matrix T_g^{Δ}.
  Calculate the updated subLVs matrices by using Equation (4.13) based on T_g^{Δ} and E_i
  Let $count = count - 1$.
end while

Joint Postprocessing

Although the corresponding subLVs for the data spaces may show similar patterns under the link of supLVs, the correlation among the subLVs can not be necessarily ensured to be maximized. Therefore, a joint postprocessing step is needed to further decompose the underlying information of the results from the first-step. Recently, joint blind source separation (BSS) techniques on multiple data sets have been successfully applied to a wide range of applications such as the estimation of brain activations [82]. Some popular methods include group ICA [82], independent vector analysis (IVA) [109] and M-CCA [110]. Through numerical and real data study, it has been shown that M-CCA can achieve better performance than the others, especially when the data sets are large [110]. Therefore, in this chapter, we adopt M-CCA as the joint postprocessing method.

M-CCA extends the theory of CCA to more than two random vectors to identify canonical variates that summarize the correlation structure among multiple random vectors by linear transformations. Unlike CCA where correlation between two canonical variates is maximized, M-CCA aims to optimize an objective function of the correlation matrix of the canonical variates from multiple random vectors in order to make the canonical variates achieve the maximum overall correlation. Details about the implementation procedure of M-CCA can be found in [110].

In this study, the inputs to M-CCA are the extracted MsubLVs from the first-step, i.e. $mT_1, mT_2, ..., mT_m$, and the outputs are the extracted canonical components with maximum correlation, revealing their common information, i.e., T_{1c}, T_{2c}, ..., T_{mc}. The associated subspace decomposition can be expressed as below:

$$\begin{aligned} &{P_{ic}}^T = ({T_{ic}}^T T_{ic})^{-1} {T_{ic}}^T X_i \\ &X_i = X_{ic} + X_{ie} = T_{ic} {P_{ic}}^T + X_{ie}, \\ &\forall i = 1, 2, ..., m \end{aligned} \tag{4.14}$$

where P_{ic} is the loading matrix and X_{ie} is the residual information within the ith data space. Regarding the number of the canonical components, we can first keep it the same as the number of subLVs and then apply the permutation test (see Section 4.3.2) to identify significant components as the final common information.

Since the residual X_{ie} may contain some specific information within the data space X_i besides noises, it is desirable to explore this kind of underlying information, reflecting the unique characteristics of each data space. Here, we adopt the method of orthogonal signal correction (OSC) [111] to draw the orthogonal components from the M-CCA residuals, i.e., T_{1o}, T_{2o}, ..., T_{mo}. The associated decomposition can be described as

$$\begin{aligned}
&{P_{io}}^T = ({T_{io}}^T T_{io})^{-1} {T_{io}}^T X_i \\
&X_{ie} = X_{io} + X_{ir} = T_{io}{P_{io}}^T + X_{ir}, \\
&\forall i = 1, 2, ..., m
\end{aligned} \tag{4.15}$$

where P_{io} is the loading matrix, and X_{ir} is the final residual within each data space.

In summary, each original data space X_i is decomposed into three subspaces:

$$X_i = X_{ic} + X_{io} + X_{ir} = T_{ic}{P_{ic}}^T + T_{io}{P_{io}}^T + X_{ir}, \tag{4.16}$$

where the first subspace stands for the similarity among multiple data sets and the second reveals the unique information in each data space.

4.2.2 A Group Analysis Architecture

To accommodate the multi-subject case, the proposed method in Section 4.2.1 is extended to the group level and we propose a group analysis method. For modeling the corticomuscular activity, we apply the proposed group analysis method to concurrent EEG, EMG and BEH signals collected from normal subjects and patients with PD during a motor task.

Suppose we have a total of B subjects ($B = 17$ in this study) and m types of data sets for each subject ($m = 3$ in this study). As illustrated in Fig. 4.2, there are mainly three steps in this architecture. In the first step, the subLVs T_{bj}'s ($b = 1, 2, ..., B$; $j = 1, 2, ...m$) for each subject are calculated using Equations (4.10) and (4.13). These subLVs include most of the useful information for further analysis. It is not necessary to use the modification algorithm here because the collinearity and ordering problems will be solved together later. In the second step, all subjects'

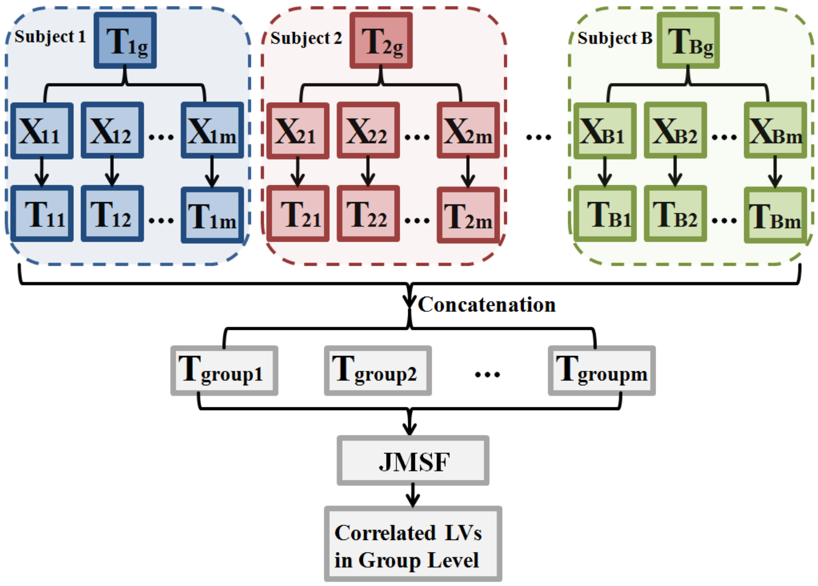


Figure 4.2: The diagram of the group analysis architecture based on the proposed JMSF in Section 4.2.1.

subLVs are correspondingly concatenated as:

$$T_{groupi} = [T_{1i}, T_{2i}, ..., T_{Bi}], \quad \forall i = 1, 2, ..., m, \tag{4.17}$$

with the assumption that all subjects share common group patterns in the temporal dimension. In the third step, the proposed JMSF is applied to the m grouped data sets and correlated components can be extracted.

4.3 Data Processing and Results

4.3.1 Simulation

Synthetic Data

In this simulation, without loss of generality, three data sets are generated and analyzed as an illustrative example. The following three data space with five variables are considered, where each data space contains a common source and a unique source:

$$
\begin{aligned}
X_1 &= [s_1, s_1, s_2, s_2, s_2], \quad with \quad size \quad 100 \times 5 \\
X_2 &= [s_2, s_2, s_3, s_3, s_3], \quad with \quad size \quad 100 \times 5 \\
X_3 &= [s_4, s_4, s_2, s_2, s_4], \quad with \quad size \quad 100 \times 5 \\
s_1 &= cos(0.01t)sin(t) \\
s_2 &= sin(0.015t) + 2cos(0.005t) \\
s_3 &= 2sin(0.025t) \\
s_4 &= 2cos(0.08t)sin(0.006t)
\end{aligned}
\tag{4.18}
$$

where t denotes the time point index, valued from 1 to 100, and s_i's $(i = 1, 2, 3, 4)$ represent four major simulated sources, which are uncorrelated with each other. In each data space, the common source s_2 represents the common information shared with other spaces and the unique source contributes to its different behavior. White Gaussian noises with 5% power level are added to each source in each data set. Moreover, to further demonstrate the robustness, we add another two higher noise levels and summarize the results in Table 4.1.

Results

We apply the proposed JMSF to synthetic data to demonstrate its performance. We illustrate the entire process of modeling the underlying variations in each data space under the supervision of each other.

All variables in three data sets are normalized to have zero-means and unit variances. According to the accumulative explained variance, four supLVs, denoted by T_g, are extracted to explain the general systematic information in the concatenated data space $[X_1, X_2, X_3]$, accounting for more than 90% of the overall variation. The four supLVs are shown in Fig. (4.3), compared with the original sources s_i's $(i = 1, 2, 3, 4)$. The extracted supLVs are sorted in a descending order of their associated variances. We can see that the supLVs recover the major sources in some sense. As expected from our method, the first supLV resembles the common source s_2 in the three data spaces. The second and third supLVs are actually the combinations of s_3 and s_4. The fourth is quite similar to s_1.

Furthermore, the corresponding subLVs and MsubLVs are also extracted from

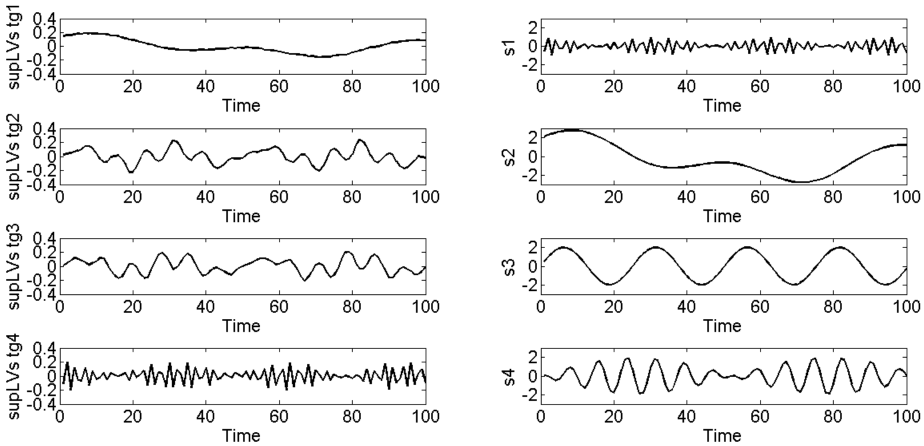


Figure 4.3: The extracted four supLVs (left hand side) and the original four sources s_i's (right hand side).

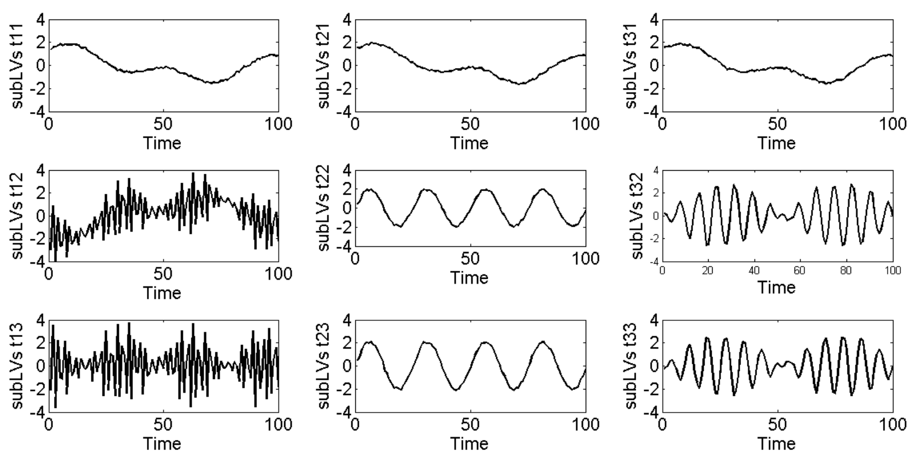


Figure 4.4: From left to right: the original subLVs of the data spaces X_1, X_2 and X_3 respectively.

each data space, and the results are shown in Fig. (4.4) and Fig. (4.5) respectively. Obviously, the subLVs from each data space are not orthogonal to each other, which indicates that the information in each data space is repetitively used. By using the modified algorithm presented in Section 4.2.1, orthogonal MsubLVs can be extracted and sorted automatically according to their correlation relationships with the supLVs. Compared with the true underlying signal model, we can

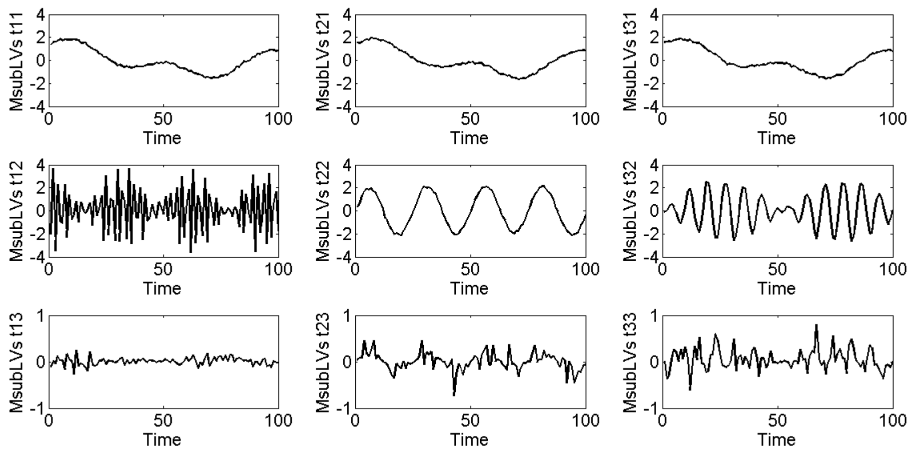


Figure 4.5: From left to right: the modified subLVs for the data spaces X_1, X_2 and X_3 respectively.

see that the first two MsubLVs can summarize the major variation in each data space and the third one is mainly due to noise.

Finally, the postprocessing step is performed to separate the highly correlated information from the irrelevant parts. As shown in Fig. (4.6), the recovered CCA components and OSC components are comparatively present for the three data spaces. The CCA components are similar to the common source s_2 and meanwhile the OSC components are similar to their corresponding unique source.

This simulation study illustrates the main steps of the proposed framework and shows that the underlying information in each data space can be effectively extracted. To further demonstrate its robustness, we consider another two higher noise levels (i.e., 10% and 15%) and summarize the results in Table 4.1. From the table, we can see that although the noise power significantly increases, cc and acc still maintain at quite high values, indicating the good robustness of the proposed method against additional noise.

4.3.2 Real Data

The detailed description for the experiment can be referred to Section 2.2. The concurrently collected EEG, EMG and BEH signals from controls and PD subjects are utilized here.

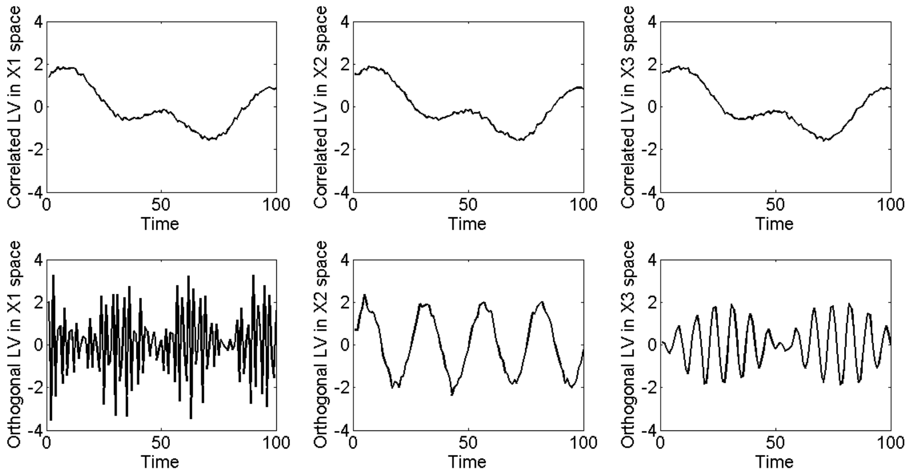


Figure 4.6: The postprocessing results for Multi-LVs.

Table 4.1: The estimation performances of the proposed method at different noise levels.

Noise Level	cc1	cc2	cc3	acc
5%	0.9985	0.9982	0.9978	0.9969
10%	0.9943	0.9920	0.9923	0.9871
15%	0.9842	0.9831	0.9880	0.9752

[a] Here noise level indicates the noise power in terms of the percentage of the source signal power; cc1 means the correlation coefficient between the original source s_2 and the correlated LV extracted from $X1$; cc2 and cc3 are similarly defined; acc is defined in Algorithm 2.

Feature Extraction

In this chapter, we examine the coupling relationships among the EEG signals, the amplitudes of the EMG signals and the behavioral performance measurements, constituting X_1, X_2, and X_3 respectively for each subject.

EEG Features Two types of EEG features are considered: Pearson's correlation [86] and EEG spectrum [112]. Pearson's correlation has been defined in Section 2.3.3 and can be referred to Equation (2.9). Spectrum represents the energy distribution of an EEG signal $x = (x_1, x_2, ..., x_n)$ in the frequency domain, which can be generated via a discrete Fourier transform:

$$X_k = \sum_{i=1}^{n} x_i e^{-j2\pi k \frac{i}{n}}, \qquad k = 0, 1, ..., n-1, \tag{4.19}$$

where X_k's represent the frequency domain information. In this chapter, we calculate the time-varying correlations between each two EEG channels and the time-varying spectra of each EEG channel, using a Hamming window with length 300 and with a 95% overlap. To be consistent with the EEG practice, we further divide the spectrum into three frequency bands, i.e., theta band (4-8 Hz), alpha band (8-13 Hz) and beta band (13-30 Hz). The time-varying energy signals in each band are used as the EEG features. Therefore, now the raw EEG signals can be transferred into a two-dimensional matrix with size $N \times M$, where the rows correspond to the samples at different time points and the columns correspond to the features, i.e. pair-wise correlations between the EEG channels or band-limited energies of individual EEG channel signals.

EMG and BEH Features The same EMG feature described in Equation (2.10) is employed here. Note that the window size used here is the same as in the EEG correlation calculation. This is to ensure that the obtained EEG and EMG features are temporally aligned and matched. Similarly, the BEH measurement sequence from each subject is resampled to ensure the same temporal length as that of the EEG and EMG features.

Significance Assessment

To determine the statistical significance levels of the extracted components, we exploit the non-parametric permutation test [88] in which temporal correlations among EEG, EMG and BEH are removed by permuting the temporal order of EEG features X_1 and EMG features X_2 uniformly for all subjects while keeping the BEH features X_3 unchanged. In this study, 200 random permutations are generated and the proposed group analysis architecture is applied to each of these permutations. The average correlation coefficients among the extracted temporal patterns from permuted EEG features, permuted EMG features, and unchanged BEH features are then calculated to form an empirical null distribution. The p-value of the original EEG/EMG/BEH average correlation can be computed from the null distribution as the probability of observing a value at least as extreme as the original correlation in the null distribution. The components with p-value being less than 0.05 are considered to be statistically significant, denoted as $T_{group1c}$, $T_{group2c}$ and $T_{group3c}$, all with size $(N \times K)$, where K is the number of significant components.

Spatial Pattern Extraction

Spatial patterns of EEG channels represent functional connections between brain regions. Especially in this study, we want to investigate the differences between the normal group and the PD patient group when performing a motor task. After the identification of significant temporal patterns, we regress the EEG-related components $T_{group1c}$ back onto the EEG features X_{b1} (for $b = 1, 2, ..., B$) for all subjects as follows:

$$p_{bk} = \sqrt{\frac{1}{{t_k}^T X_{b1} {X_{b1}}^T t_k}} {X_{b1}}^T t_k, \quad k = 1, 2, ..., K. \tag{4.20}$$

where t_k is the k-th column of $T_{group1c}$ and p_{bk} is the spatial pattern of the k-th component for subject b. For subject b, p_{bk} can reflect which correlation connections make important contributions to the corresponding significant temporal pattern. To allow for easier visualization and comparison of spatial patterns between the two groups (normal vs. PD), for each significant component, we concatenate the spatial patterns of the subjects within a group horizontally, and perform PCA on the concatenated spatial patterns. The group level spatial pattern is then represented by the first principal component [30].

Results

In this real data study, we apply the proposed group analysis method for corticomuscular activity analysis to concurrent EEG, EMG and BEH signals collected from normal subjects and patient subjects with PD when they perform a force tracking task. We aim to extract common temporal patterns among the three types of signals and explore the differences in their spatial patterns between the control and

PD groups.

The proposed group analysis method is applied to the EEG, EMG and BEH features generated using the procedure described in Section 4.3.2 from all 8 normal and 9 PD subjects simultaneously. The joint modeling of normal and PD data allows the identification of common temporal patterns across the groups. Meanwhile, the spatial patterns may be different across subjects, from which we can identify the specific correlation connections or brain regions that are differentially recruited by PD subjects during the motor task.

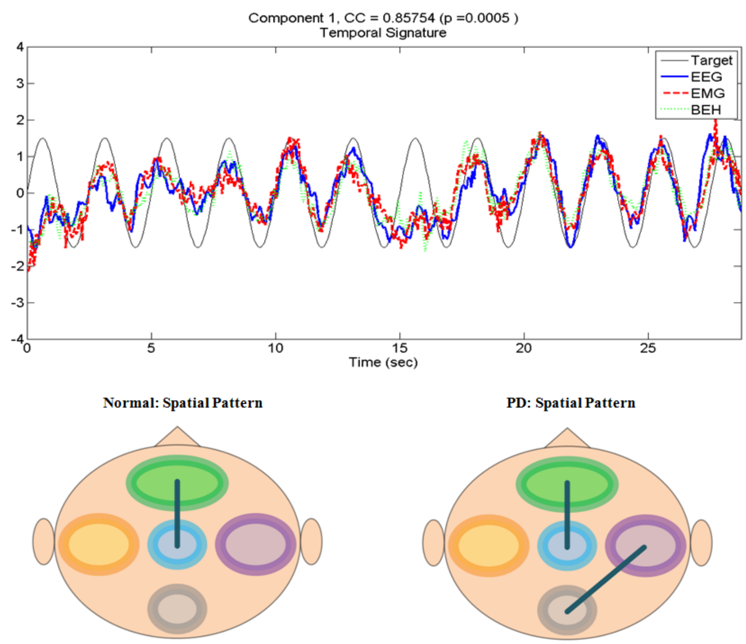


Figure 4.7: Component 1 of the group JMSF when using the EEG correlations, the EMG amplitude and the force output data sets. Top panel: Temporal patterns of the EEG (solid), the EMG (dashed) and the BEH signals (dotted). The oscillation of the target bar is also shown (black, solid). Bottom panel: EEG spatial patterns of normal subjects (left) and PD subjects (right). The connections reflect the respective spatial patterns in the two groups.

In this study, we repeat the analysis to two types of EEG features. For each case, X_1, X_2 and X_3 represent the time-varying feature matrices of EEG, EMG and

BEH respectively. The length of the time sequences here is 480, associated with the 300-length moving window and a 95% overlap. For the EEG correlation case, since the EEG signals are divided into five regions as mentioned in Section 2.2, there are a total of $C_2^5 = 10$ correlation connections. Thus, X_1 is with size 480×10. For the EEG band-energy case, there are five brain regions and three frequency bands, indicating 15 energy features can be generated. Thus, X_1 is with size 480×15. For both cases, X_2 and X_3 are with size 480×1.

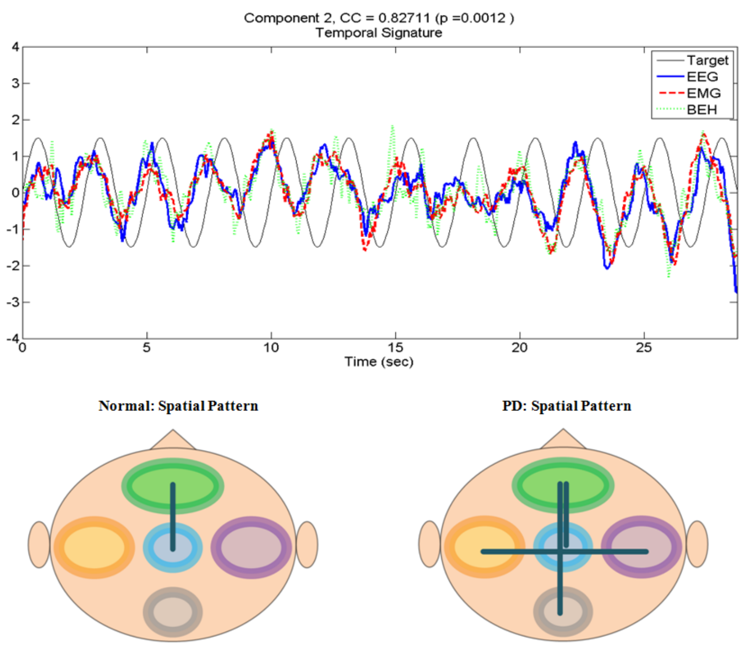


Figure 4.8: Component 2 of the group JMSF when using the EEG correlations, EMG amplitude and force output signals as the three data sets.

EEG Correlation Components Corresponding to EMG and BEH Data Using the permutation test, three significant components ($P \leq 0.05$) are identified. The temporal and the corresponding spatial patterns for Component 1 are shown in Fig. 4.7. The average correlation coefficient among the EEG, EMG and BEH scores is 0.86 (with the p-value P = 0.0005). The group-level EEG spatial activation patterns of normal and PD subjects given by PCA are shown in the bottom panel of

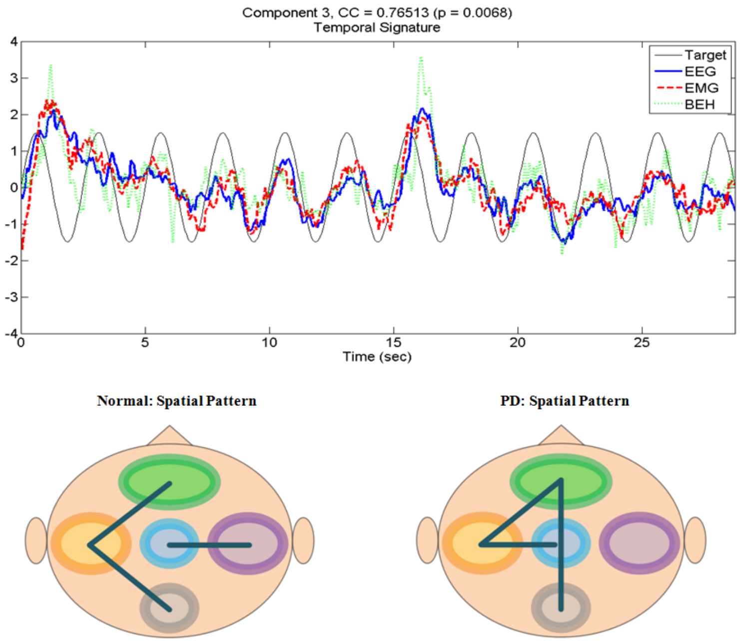


Figure 4.9: Component 3 of the group JMSF when using the EEG correlations, EMG amplitude and force output as the three data sets.

Fig. 4.7. In this EEG feature setting, there are a total of 10 correlation connections. For easier interpretation, only connections whose weights are greater than 80% of the maximum value are shown. From Fig. 4.7, we can see that, for both normal and PD groups, this component is dominated by the connection *Central – FCentral*, while in PD subjects, there is an additional connection *Occipital – RSM*. The temporal and spatial patterns for Component 2 are shown in Fig. 4.8. The average correlation coefficient among the EEG, EMG and BEH scores is 0.83 (with P = 0.0012). From Fig. 4.8, we can see that, for both normal and PD groups, there exists the connection *Central – FCentral*, whereas for the PD group, there are two additional connections *LSM – RSM* and *Occipital – FCentral*. The temporal and spatial patterns for another significant component are shown in Fig. 4.9. The average correlation coefficient among the EEG, EMG and BEH scores is 0.77 (with P = 0.0068). In normal subjects, this component is based largely on the connections *LSM – FCentral*, *Occipital – LSM* and *Central – RSM*, while in PD subjects, it is based on the connections *LSM – FCentral*, *Occipital – FCentral* and *Central –*

LSM.

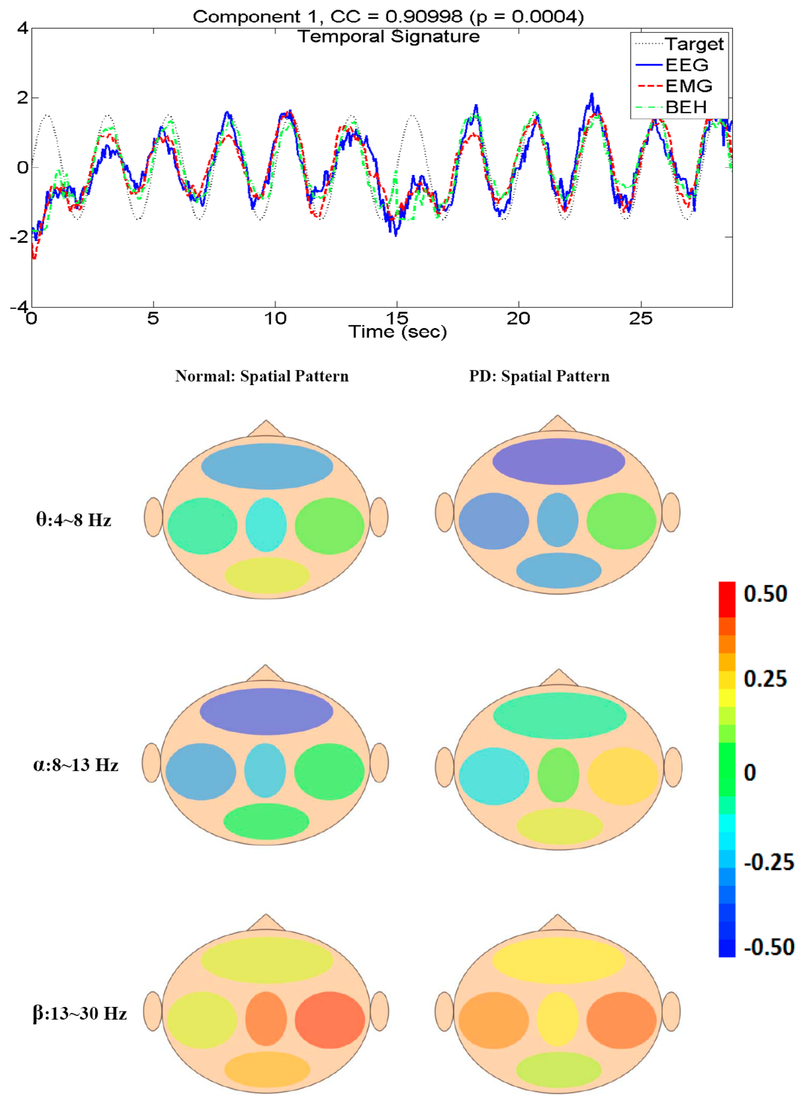


Figure 4.10: Component 1 of the group JMSF when using EEG band-limited energies, EMG amplitude and force output signals as data sets. Top panel: Temporal patterns of EEG (solid), EMG (dashed) and BEH (dotted). The oscillation of the target bar is also shown (black, solid). Bottom panel: EEG spatial patterns of normal subjects (left) and PD subjects (right). The colors reflect the relative weighting of the respective band-limited energies in five brain regions.

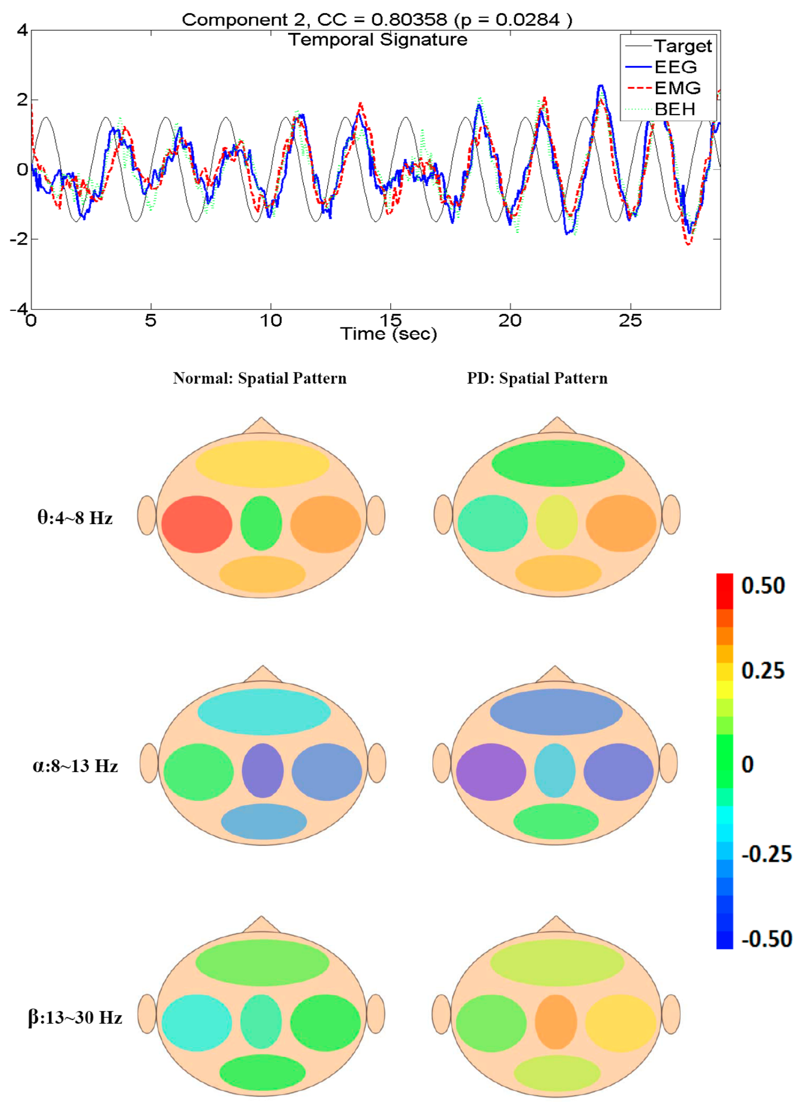


Figure 4.11: Component 2 of the group JMSF when using the EEG band-limited energies, EMG amplitude and force output signals as data sets.

EEG Band-Energy Components Corresponding to EMG and BEH Data Using the permutation test, two significant components ($P \leq 0.05$) are identified. The temporal and spatial patterns for Component 1 are shown in Fig. 4.10. The average correlation coefficient among the EEG, EMG and BEH scores is 0.91 (with P = 0.0004). The group-level EEG spatial activation patterns of normal and PD subjects given by PCA are shown in the bottom panel of Fig. 4.10. In this EEG feature setting, there are five brain regions and three frequency bands, i.e. 15 energy features. For the theta band, in both normal and PD groups, *FCentral* demonstrates higher weights, while in PD subjects *LSM*, *Central* and *Occipital* also have relatively higher weights. For the alpha band, in normal subjects *FCentral* and *LSM* play important roles, whereas in PD subjects there are no obvious differences among the five regions. For the beta band, in the normal group this component is largely associated with activities in *Central*, *RSM* and *Occipital* regions, while in PD the activity is seen more in the *RSM* and *LSM* regions. The temporal and spatial patterns for the other component are shown in Fig. 4.11. The average correlation coefficient among the EEG, EMG and BEH scores is 0.80 (with P = 0.0284). For the theta band, in both normal and PD groups *RSM* and *Occipital* demonstrates higher weights, while in normal subjects *LSM* shows a particularly high weight. For the alpha band, in normal subjects the right brain area plays an important role, whereas in PD subjects it is seen more in the frontocentral area. For the beta band, all regions display low values in normal subjects, while in PD this component is associated with activity in the *Central* region.

4.4 Discussion

When the EEG correlation feature is used, we note that the connections with occipital regions are prominent in PD subjects. Compared to normal subjects, the PD subjects heavily rely on visual cues for the initiation [94] and ongoing control of movement [95]. In addition, the increased intra and inter-hemispheric connections observed in the PD subjects are in accord with the findings in previous MEG studies during the resting state [5].

For the case when the EEG Band-Energy feature is used, we note that an increased number of brain regions in the theta band are recruited in PD subjects, which is in line with the observed association between synchronization in the theta band and motor symptoms, in particular tremor [5]. In addition, the PD subjects show increased interhemispheric synchronization in the alpha range. This finding may be positively related to cognitive perseveration in Parkinson's disease [5]. Besides, compared to normal subjects, the PD subjects heavily rely on the *Central* region in the beta band during the motor task, which is quite consistent with our previous study in the same experiment setting by mutual information network [112].

Chapter 5

A Three-step Method for Corticomuscular Activity Modeling

5.1 Motivation and Objectives

In the last chapter, we have proposed a framework to handle multimodal data sets under the uncorrelatedness assumption. However, LVs that are uncorrelated can sometimes complicate interpretations in real medical applications [38]. We will demonstrate this point in the simulation part. In this chapter, we aim to develop a framework that can explore the relationship between multimodal data sets and meanwhile can extract independent LVs within each data set. For multimodal corticomuscular activity analysis, both M-CCA and jICA have their individual advantages and disadvantages and fortunately, these two methods can be considered complementary.

In this chapter, we propose combining M-CCA and jICA to improve the overall performance of the joint source extraction. More specifically, we first adopt M-CCA to obtain CVs across multiple data sets and then perform jICA on the extracted LVs. In contrast to prior approaches [113, 114], we generate common components from *sources* rather than the *mixing matrices* and concentrate on multimodal/multiset data fusion. In fact, we employ a three-step modeling strategy. The first two steps are similar to JMSF and the last one is to help improve the limitation of M-CCA. Because M-CCA requires a stringent assumption that correlation coefficients of the corresponding sources between multiple data sets should be sufficiently distinct [110], which will be specified in Section 5.2.1 by Equation (5.2). This assumption may not be satisfied in practice.

5.2 Methods

The main contributions of the proposed method lies in its ability to: (1) model multiple data sets/signals simultaneously, overcoming the traditional constraint of analyzing only two types of data sets, and (2) extract maximally correlated components across multiple data sets while simultaneously keeping components within each individual data set maximally statistically independent. In the following subsections, the three-step modeling strategy, illustrated in Fig. 5.1, is described in detail. However, we will begin from the second step since the first step can be referred to Section 4.2.1.

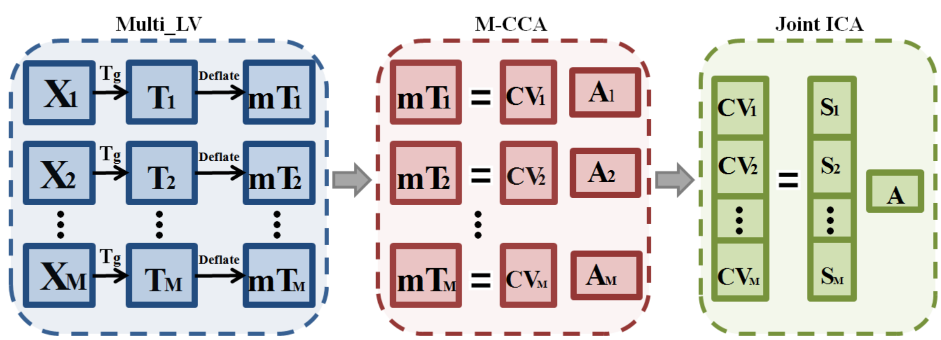


Figure 5.1: The diagram of the proposed joint multimodal analysis method.

5.2.1 Multiset Canonical Correlation Analysis

Although the corresponding subLVs for multiple data sets could show similar patterns because of the common supLVs, the correlation coefficients among the subLVs can not be necessarily ensured to be maximized. Therefore, a second step is needed to further decompose the underlying information of the results from the first-step. M-CCA extends the theory of CCA to more than two random vectors to identify CVs that summarize the correlation structure among multiple random vectors by linear transformations. Unlike CCA where correlation between two canonical variates is maximized, M-CCA aims to optimize an objective function of the correlation matrix of the canonical variates from multiple random vectors in order to make the canonical variates achieve the maximum overall correlation [110]. It has been shown that M-CCA can achieve excellent performances, especially when the data sets are large. Details about the implementation procedure of M-CCA can be found in [110].

In this study, the inputs to M-CCA are the extracted MsubLVs from the first-step, i.e. mT_1, mT_2, ..., mT_M, and the outputs are the extracted canonical variates with maximum correlation, revealing their common information, i.e., CV_1, CV_2, ..., CV_M. The associated decomposition can be expressed as below:

$$mT_i = CV_i A_i, \quad \forall i = 1,2,...,M \tag{5.1}$$

where A_i is the mixing matrix within the ith data set. Regarding the number of the canonical variates in each data set, we can keep it the same as the number of MsubLVs.

As mentioned before, M-CCA may fail to separate sources whose correlation coefficients are equal or very close, which could often occur in real biomedical applications. In other words, if K sources can be extracted from each of M data sets correspondingly, the following requirement must be met to successfully recover the sources by M-CCA:

$$|r_{m,n}^{(\alpha)}| \neq |r_{m,n}^{(\beta)}|, \quad (1 \leq \alpha < \beta \leq K, \forall m,n \in 1,2,...,M) \tag{5.2}$$

where $|r_{m,n}^{(\beta)}|$ represents the correlation coefficient between the β-th source from the m-th data set and the β-th source from the n-th data set. Therefore, CV_i should be regarded as incompletely decomposed sources, i.e. mixtures of the real independent components.

5.2.2 Joint Independent Component Analysis

Due to the possible incompleteness of the source separation, in Step-3, we perform jICA on the concatenated CVs to maximize the independence among joint components by reducing their higher order statistical dependencies as below [84]:

$$[CV_1; CV_2; \dots; CV_M] = [S_1; S_2; \dots; S_M]A \tag{5.3}$$

where A represents the common mixing matrix and S_i is the extracted ICs from the i-th data set. For the implementation of ICA, many algorithms have been developed based on different cost functions. Here we employ the classical one, FastICA [39]. We then correlate the corresponding columns of S_i and calculate the average correlation coefficients, in terms of which the corresponding columns are sorted from high to low. As to the number of the joint ICs in each data set, we can first keep it the same as the number of CVs and then apply the permutation test to identify significant components.

We should note that jICA assumes that all modalities share the same mixing matrix A. This constraint is not always easily satisfied in practice. In the proposed method, M-CCA first links multiple data sets via correlation and prepares more relevant CVs correspondingly across the data sets. Therefore we can perform jICA on the extracted CVs with more confidence, meaning that M-CCA helps jICA relax the required constraint. In turn, jICA further decomposes the remained mixtures in CVs and also helps M-CCA relax the constraint of the distinctiveness for correlation coefficients. In summary, M-CCA and jICA in the proposed method are complementary to each another and combining them appropriately can somehow overcome their limitations.

5.3 Data Processing and Results

5.3.1 Simulation

In this simulation, we apply the proposed method to synthetic data and also implement separate ICA, M-CCA and jICA respectively for comparison.

Synthetic Data

As an illustrative example, without loss of generality, six sources are generated and analyzed.

The following six source signals are considered:

$$\begin{aligned} s_1 &= \sin(0.015t) + \cos(0.005t) \\ s_2 &= 2\cos(0.08t)\sin(0.006t) \\ s_3 &= ECG, \quad s_4 = EMG \\ s_5 &= 1.5\cos(0.01t)\sin(0.5t) \\ s_6 &= 1.5\sin(0.025(t+63))\sin(0.2t) \end{aligned} \tag{5.4}$$

where t denotes the time index vector, valued from 1 to 1000, and s_i's $(i = 1, 2, ..., 6)$ represent six simulated sources, as shown in Fig. 5.2. Note that here s_i's are column vectors.

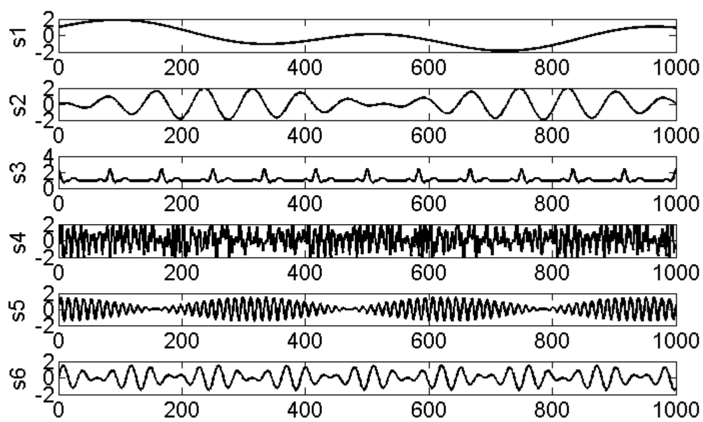


Figure 5.2: The six simulated source signals.

Three mixed data sets X_1, X_2 and X_3 are generated as follows, with each row denoting one observation in their respective data space:

$$X_i = S_i \cdot A_i, \quad i = 1, 2, 3 \tag{5.5}$$

where $S_1 = [s_1\, s_2\, s_3\, s_4]$, $S_2 = [s_1\, s_2\, s_3\, s_5]$ and $S_3 = [s_1\, s_2\, s_3\, s_6]$ with

$$A_1 = \begin{bmatrix} 0.76 & -0.55 & 0.17 & -0.33 & 0.82 \\ 0.79 & 0.65 & 0.32 & 0.12 & -0.97 \\ 0.87 & 0.46 & -0.58 & 0.27 & -0.74 \\ 0.59 & 0.45 & 0.37 & 0.22 & 0.11 \end{bmatrix}, \tag{5.6}$$

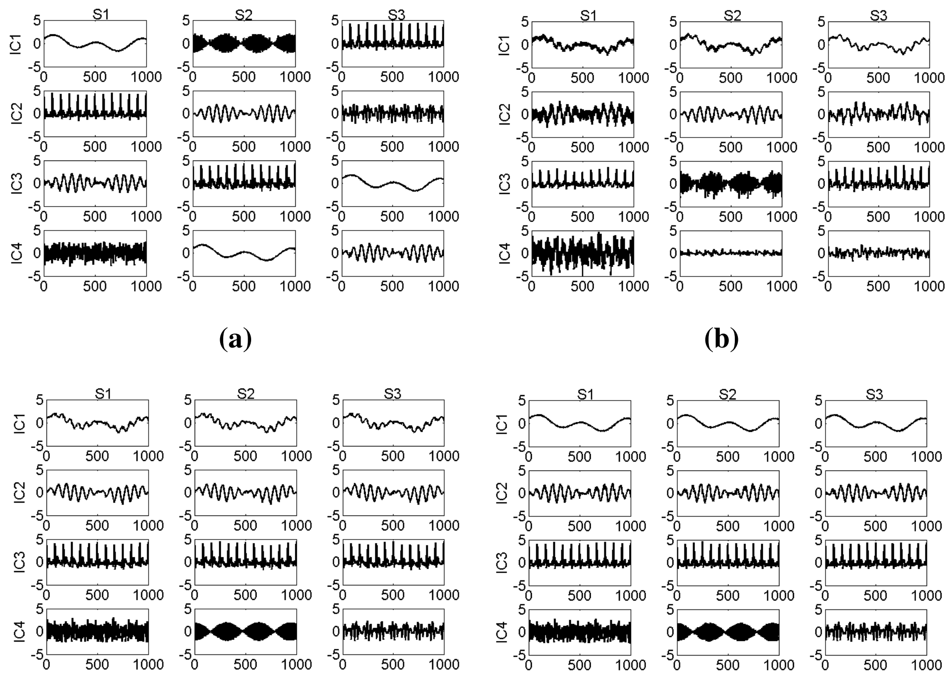
$$A_2 = \begin{bmatrix} 0.73 & -0.52 & 0.21 & -0.29 & 0.78 \\ 0.82 & -0.67 & 0.6 & -0.20 & -0.90 \\ 0.78 & -0.1 & 0.71 & 0.29 & -0.51 \\ 0.52 & 0.39 & 0.30 & 0.27 & 0.15 \end{bmatrix}, \tag{5.7}$$

$$A_3 = \begin{bmatrix} 0.69 & -0.55 & 0.22 & -0.34 & 0.77 \\ 0.76 & 0.69 & 0.22 & 0.25 & -0.60 \\ 0.71 & -0.49 & -0.35 & 0.4 & -0.9 \\ 0.60 & 0.41 & 0.30 & 0.16 & 0.20 \end{bmatrix}. \tag{5.8}$$

The sources s_1, s_2 and s_3 exist in all data sets, representing common information. The source s_4 only contribute to X_1, s_5 only to X_2 and s_6 only to X_3, representing their own unique information. Random Gaussian noise with 5% power were added to S_i's before generating the mixed data sets.

Results

The extracted underlying components using separate ICA, joint ICA, M-CCA and the proposed method are shown in Fig. 5.3a, 5.3b, 5.3c and 5.3d respectively. We also made a quantitative comparison between these methods by calculating average correlation coefficients for each method (Fig. 5.3e.) The separate ICA approach recovered the original sources accurately in all data sets as expected, but was unable to meaningfully relate the three data sets or rank the ICs appropriately. Although a subsequent source match via cross-correlation could be used to oder the sources this may introduce ambiguity especially when the estimated component number is high. By using joint ICA, the first jointly extracted sources, although perhaps less accurate, are at least highly relevant to each other. But the remaining extracted sources seem to be uninformative due to the stringent assumption that all modalities



(a) (b) (c) (d)

The Performance Comparison Table by Average Correlation Coefficient

Dataset \ Method	Separate ICA	Joint ICA	M-CCA	The proposed method
X1	0.5573	0.8623	0.9573	0.9901
X2	0.4928	0.7004	0.9539	0.9905
X3	0.0091	0.8185	0.9494	0.9850

(e)

Figure 5.3: The extracted underlying components by the following four methods: (a) separate ICA; (b) joint ICA; (c) M-CCA; (d) the proposed method; (e) performance comparison of the different methods.

share the same mixing matrix. The LVs extracted by the M-CCA were at least automatically and neatly ordered in terms of their average correlation coefficient values among the data sets. However, compared with the original sources, the extracted LVs were distorted, suggesting 1) performance of M-CCA may suffer when the condition indicated by Equation (5.2) is not satisfied, and 2) uncorrelatedness may not be a sufficiently rigorous criterion to accurately recover the underlying sources. When the proposed method was employed, the dominant sources which made significant contributions to all data sets were accurately identified and ordered, with the focus entirely on sources common across multiple data sets. In summary, the proposed method mitigates the deficiencies of both M-CCA and joint ICA and can separate sources accurately and link them correctly in a less-constrained condition.

5.3.2 Real Data

The experiment description can be referred to Section 2.2. In this study, we utilize concurrently EEG, EMG and BEH signals collected from controls and PD subjects.

Feature Extraction

In this chapter, we examine the coupling relationships among time-varying EEG features, the amplitude of the EMG signals, and behavioral performance measurements, constituting X_{1b}, X_{2b}, and X_{3b} respectively for each subject b (for $b = 1,2,...,B$). Suppose we have a total of B subjects ($B = 17$ in this study) and m types of data sets for each subject ($m = 3$ in this study). To achieve a group analysis, all subjects' data sets are correspondingly concatenated as:

$$\begin{aligned} X_1 &= [X_{11}, X_{12}, ..., X_{1B}], \\ X_2 &= [X_{21}, X_{22}, ..., X_{2B}], \\ X_3 &= [X_{31}, X_{32}, ..., X_{3B}], \end{aligned} \tag{5.9}$$

with the assumption that all subjects share common group patterns in the temporal dimension [82].

Band-limited, pair-wise EEG coherences [97] are considered in this chapter since this type of time-varying features include temporal, spatial and spectrum information simultaneously. The calculation can be referred to Equation (1.1) and Section 3.3.2. EMG and BEH Features are exactly the same as in Section 4.3.2.

Significance Assessment

The non-parametric permutation test performed here is similar to Section 4.3.2. The p-value of the original EEG/EMG/BEH average correlation was then estimated from the null distribution as the probability of observing a value at least as extreme as the original correlation in the null distribution. The components with p-value being less than 0.05 were considered statistically significant, denoted as U_1, U_2 and U_3, all with size $(N \times K)$, where K is the number of significant components.

Spatial Pattern Extraction

In order to aid in the biological interpretation of the results, we also derived connectivity patterns between EEG channels representing functional connections between brain regions. After the identification of significant temporal patterns, we regressed the EEG-related components U_1 back onto the EEG features X_{1b} (for $b = 1, 2, ..., B$) for all subjects as follows:

$$p_{bk} = \sqrt{\frac{1}{{u_k}^T X_{1b} {X_{1b}}^T u_k}} {X_{1b}}^T u_k, \quad k = 1, 2, ..., K. \tag{5.10}$$

where u_k is the k-th column of U_1 and p_{bk} is the connectivity pattern of the k-th component for subject b. In addition, we also want to determine which EEG features in the connectivity patterns have significant contributions to the corresponding temporal patterns. This is done by identifying EEG features that have weights statistically significant different from zero. To determine the group-level connectivity pattern, for each significant component, we applied a two-tailed t-test to each connectivity feature across all subjects in each subject group.

Results

The proposed method was applied to the EEG, EMG and BEH features generated using the procedure described in Section 5.3.2 from 8 normal and 9 PD subjects simultaneously. The joint modeling of normal and PD data allows for the identification of common temporal patterns across the groups, while the connectivity patterns may be different across subjects. This allows for identification of specific correlation connections that are differentially recruited by PD subjects during the motor task.

Here X_{1b}, X_{2b} and X_{3b} $(b = 1,2,...,B)$ represent the time-varying feature matrices of EEG, EMG and BEH for the b-th subject respectively. The length of the original raw EEG and simultaneously recorded EMG data for each subject was 7485 associated with the sampling frequency 250 Hz. As mentioned before, after using the 300-length moving window and a 95% overlap, the length of time points (the matrix row) here was 480. For the EEG coherences, there are five brain regions and four frequency bands, indicating there are a total of $C_2^5 \times 4 = 40$ variables. Thus, X_{1b} is of size 480×40. For the EMG amplitude feature, since there are two surface EMG channels, X_{2b} was 480×2, and the BEH data set X_{3b} was size 480×1. The group analysis data sets X_1, X_2 and X_3 were generated by Equation (5.9).

Using the permutation test, one significant component (p-value ≤ 0.05) was identified. The temporal and spatial patterns for this component are shown in Fig. 3.8. The *acc* among the EEG, EMG and BEH scores was 0.7946 (with p-value = 0.0015). Note that in the figure only connections whose weights were statistically significantly different from zero are shown.

Although there were some similarities in the brain regions and connections recruited by PD and control groups, there still existed significant differences Fig. 5.4. In theta and alpha bands, normal subjects demonstrated frontoparietal and fronto-occipital connections which were completely missing in PD subjects. In the beta band, normal subjects had connections between the L sensorimotor cortex and the frontal and R sensorimotor regions that were absent or reduced in PD subjects. In the gamma band, PD subjects had increased connectivity between the frontal regions and central and sensorimotor cortices that were missing in controls. Finally, PD subjects had a significant connection between the L sensorimotor and occipital regions in the theta, alpha and gamma bands that was not present in controls.

5.4 Discussion

In this chapter, we have proposed a three-step multimodel data analysis method to assess corticomuscular coupling based on combining the advantages of M-CCA and jICA. The simulation results support the usefulness of the proposed method.

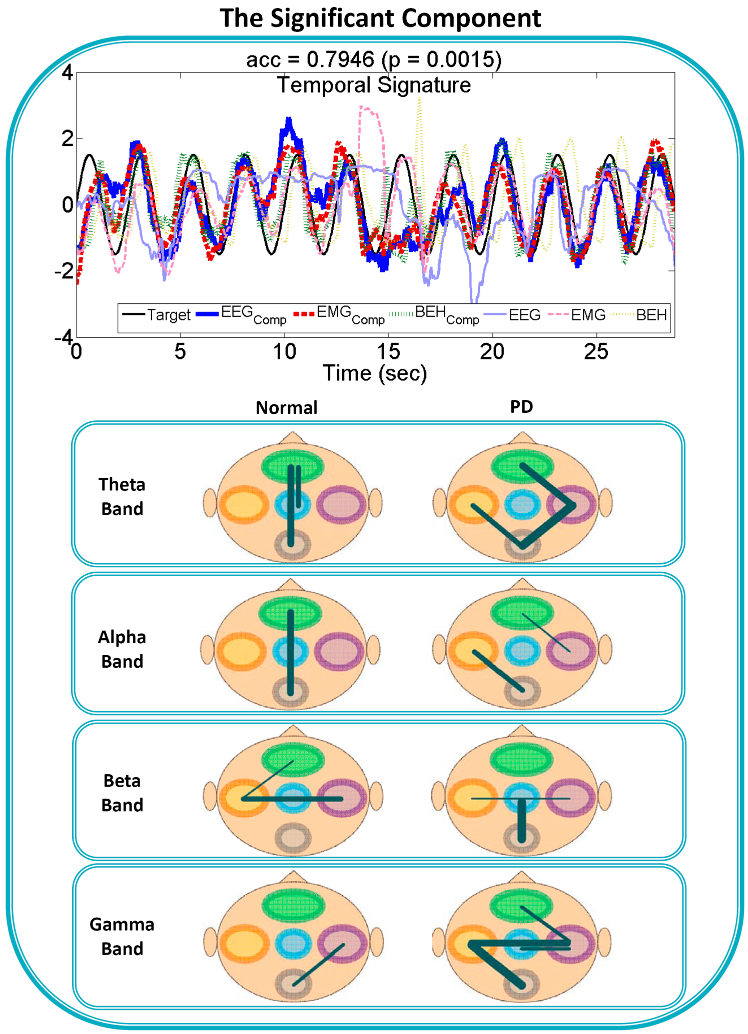


Figure 5.4: The significant component jointly extracted from the EEG coherence, EMG amplitude and BEH data sets. Top panel: Temporal patterns of the EEG (solid), EMG (dashed) and BEH signals (dotted). The oscillation of the target bar is also shown (gray, solid). Bottom panel: EEG connectivity patterns of normal subjects (left) and PD subjects (right). The connections reflect the respective connectivity patterns in the two groups with the width of each connection proportional to the weighting in the group EEG spatial pattern. Here *acc* refers to average correlation coefficient.

Furthermore, when applied to real data promising results based on detecting significant differences between Parkinson's disease subjects and age-matched controls were detected.

The observed results in PD and control groups are in accord with previous medical findings. We found widespread connectivity between the frontal, central and ipsilateral (right) sensorimotor regions and ongoing EMG/BEH activity from Fig. 5.4. It is suggested that during visually guided tasks PD subjects may generate detectable widespread oscillatory disruption in the EEG (e.g. [100]). This likely represents a compensatory mechanism of cortical regions to overcome deficiencies in the basal ganglia to perform motor tasks, as has been observed in fMRI studies [101]. The lack of fronto-central and fronto-occipital activities in the theta and alpha bands seen in PD subjects compared to controls may reflect impaired spatial attention [102], as impairment of maintenance of attention has been well described in PD subjects (e.g. [115]). The weaker beta band connection between the L and R sensorimotor regions in PD may reflect the decreased ability for fine modulatory control [90]. The increased gamma connectivity in PD likely represents the extra effort PD subjects require for motor task switching [73, 92, 116], a problem common in PD [93]. We also found that connections with occipital regions are prominent in PD subjects. The enhanced occipital connectivity in the theta, beta and gamma bands may relate to the fact that PD subjects heavily rely on visual cues for the initiation [94] and ongoing control of movement [95]. In addition, the increased intra and inter-hemispheric connections we observed in PD subjects are consistent with the findings in previous MEG studies during the resting state [5].

It may seem paradoxical that PD subjects have reduced beta-band connectivity in the sensorimotor cortex compared with controls, especially since PD subjects may have enhanced beta-band EEG activity associated with the clinical signs of bradykinesia and rigidity [104, 116]. We believe that this is because the current method specifically looks at the common sources among the EEG, EMG and behavior signals. Because a key aspect in PD is the inability to modulate the beta-band oscillations [105], it is very likely that the fraction of beta-band activity that varies with phasic EMG/BEH activity is reduced in PD.

In this chapter, we were able to find common temporal patterns in both PD and controls assuming different connectivity origins of the common signal. We have shown that the distributed EEG connectivity patterns required to generate the common temporal pattern differ between PD and controls. One could envision situations where the connectivity patterns are similar, but the temporal profiles are significantly different. Nevertheless, usually behavioral tasks are specifically chosen so that PD and control subjects can still perform the task with roughly equal accuracy. Otherwise any differences in brain activation may be dismissed as simply that the PD and control subjects were performing different tasks.

Chapter 6

Conclusions and Future Works

6.1 Conclusions

In this thesis, we proposed four novel multimodal signal processing methods for corticomuscular coupling analysis and investigated different brain connectivity patterns between normal and PD subjects during a dynamic visually guided tracking task. The goal of all these methods is to maximize the estimated source independence within an individual data set and to maximize the source dependence across datasets. As illustrated in Fig. 6.1, the relationships between these methods can be summarized as follows: both PLS+CCA and IC-PLS are bimodal approaches for handling only two types of data; both JMSF and 3-step method can handle multimodal data; both PLS+CCA and JMSF can only extract uncorrelated components within each data set, while either of IC-PLS and 3-step method is able to obtain independent components using higher-order statistics.

PLS+CCA is a bimodal method which only explores second-order statistics with a two-stage procedure. It is a simple, yet efficient method to extract uncorrelated components within each data set and meanwhile keep corresponding components across data sets highly correlated. It is especially suitable for a situation where the uncorrelatedness assumption is competent to decompose the signals and computational speed is the major concern (e.g. online detection). IC-PLS is also a bimodal method which incorporates both second-order and high-order statistics into a multi-objective optimization problem. It ensures that the components within

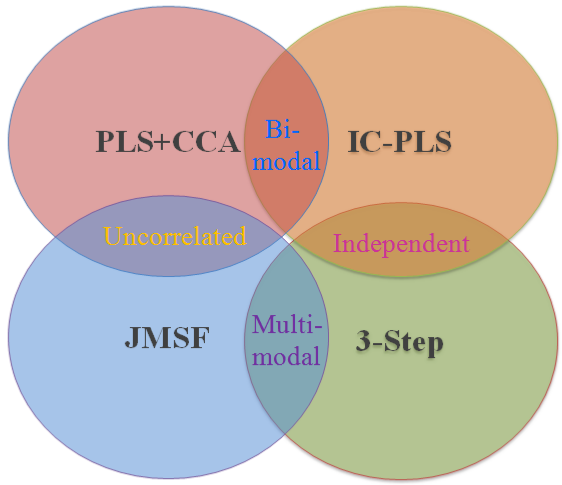


Figure 6.1: The relationships between the four proposed methods.

each data set are as independent as possible and their orders across data sets are the same. When two types of data are available and the uncorrelatedness assumption is insufficient to decompose the signals, IC-PLS is a good option. However, as a multi-objective optimization problem, IC-PLS represents a compromise between the conventional PLS and ICA.

JMSF is preferred when more than two types of data are available. It is able to extract similar temporal patterns from multimodal data sets. It particularly accommodates the needs in recent years, e.g., performing data fusion when concurrent data sets are available (e.g. fMRI, EEG, EMG, etc). However, the estimated components within each data set are only uncorrelated, which may not be enough in certain applications where independent components are preferred. To address this concern, the proposed 3-step method aims to extract independent components by assuming all data sets share the same mixing matrix. For the purpose of group analysis, we could employ both JMSF and the proposed 3-step methods.

One should note that in this thesis we have analyzed the brain spatial patterns by several data-driven coupling analysis methods based on multimodal data, while the conventional EEG-EMG coherence only explores individual locus and traditional EEG-based connectivity analysis ignores EMG-related information. In the past, most EEG/EMG coupling analysis studies have compared EEG activity at a specific locus (e.g. sensorimotor cortex contralateral to the hand performing the task) with the EMG during sustained contraction. However, we found that spectrum/correlation/coherence features between the contralateral sensorimotor cortex and other regions were closely associated with ongoing EMG/BEH features during dynamic motor performance. It is likely that the dynamic nature of the task might require the recruitment of additional regions such as frontal regions for motor selection [89], contralateral (i.e. ipsilateral to hand movement) sensorimotor cortex for fine modulatory control [90] and occipital regions for post-error adaptations [91].

While the proposed methods are intentionally developed for corticomuscular coupling analysis, we would like to emphasize that they can also be applied to analyze other forms of concurrent signals including but not limited to fMRI, photoplethysmograph (PPG), ECG and kinematic data. Therefore, they are promising tools for multi-subject and multi-modal data analysis.

In Appendix A, we designed and implemented a wireless wearable EEG/sEMG recording system. Based on the developed system, we investigated a novel application – Chinese number gesture recognition. Number gestures are the most frequently used hand gestures in real-life especially when there are language communication difficulties. One is able to develop similar applications for their own languages based on our reported work. We believe that such applications could be very useful in multifunction prosthesis, remote control, human-computer interface (HCI), etc. We also conducted a complete performance comparison by investigating several most popular feature extraction and recognition methods. To further improve the recognition accuracy, we proposed using MKL-SVM as a more effective classifier by fitting the features with multiple kernel functions. Among these combinations, we noted that 3F + MKL-SVM provided the best performance, but the efficiency of MKL-SVM is much lower than that of the others. To implement an online recognition system, we suggest that Hudgins' time-domain (TD) + linear discriminant analysis (LDA) can be the best trade-off between efficiency and performance.

6.2 Future Research Directions

To better model corticomuscular activity by data-driven coupling methods, we still need to address several challenges, including relaxing the statistical assumption (e.g. sharing the same mixing matrix) for independence, dealing with the underdetermined situation when there exist more sources than the measurement channels, and developing novel approaches for dynamic/time-varying coupling analysis.

Decomposing signals into independent components has been proven very useful in many biomedical applications [80–82]. An interesting problem is to jointly model multiple data sets to decompose each into independent components and relate corresponding components across the data sets. In this thesis, we have finally proposed a three-step framework to realize this goal. However, a stringent statistical assumption has been made in the framework that the mixing matrices are the same. This assumption could be too strong in practice. We have to relax this practically-unfavored assumption to explore underlying information in a more realistic setting. A possible solution might lie in a recently popular joint BSS method – the IVA [109]. IVA, as an extension of ICA from one to multiple datasets, has attracted increasing research attention during the past few years. IVA was originally designed to address the permutation problem in the frequency domain for the separation of acoustic sources [117]. Recently, it has been used to do group inference for fMRI data [118]. In [119], IVA has been investigated under the multivariate Gaussian model. Nevertheless, the distribution of sources in practice can not be easily modeled as Gaussian. For the corticomuscular coupling analysis problem, we will explore the specific distributions of EEG and EMG signals and design specific coupling analysis methods under the IVA framework.

Another possible future direction is to develop underdetermined coupling analysis methods. In many biomedical applications, only a limit number of sensors are available (e.g. one or two EEG/EMG channels), while more underlying sources actually exist. It is of particular importance for the situations where a large number of sensors should be avoided such as in the application of health care at home. Possible solutions may include decomposing signals as a preprocessing step (e.g. empirical mode decomposition) [120], separating signals into multiple segments [121] and performing sparse representation [122].

It is also desirable to develop dynamic models for time-varying coupling analysis. All previous methods are based on the assumption that the overall signals are stationary and the corresponding patterns are the same. However, it may not be the case in practice. A more natural assumption should be that the brain patterns are dynamically changed in terms of their states (e.g. different force levels during motor tasks). We could combine hidden Markov models with existing coupling methods and dynamically analyze the learned brain patterns. This may allow researchers obtaining insights into the changing process of brain patterns during some motor tasks and provide more information about specific diseases.

Bibliography

[1] T. Mima and M. Hallett, "Corticomuscular Coherence: A Review," *Journal of Clinical Neurophysiology*, vol. 16, no. 6, p. 501, 1999.

[2] P. Grosse, M. Cassidy, and P. Brown, "EEGEMG, MEGEMG and EMGEMG frequency analysis: physiological principles and clinical applications," *Clinical Neurophysiology*, vol. 113, no. 10, pp. 1523-1531, 2002.

[3] K. Friston, "Functional and effective connectivity in neuroimaging: a synthesis," *Human Brain Mapping*, vol. 2, no. 1-2, pp. 56-78, 1994.

[4] S. Farmer, Neural rhythms in Parkinsons disease, Brain, vol. 125, no. 6, pp. 1175-1176, 2002.

[5] D. Stoffers, J. Bosboom, etc., "Increased cortico-cortical functional connectivity in early-stage Parkinson's disease: An MEG study," *NeuroImage*, vol. 41, no. 2, pp. 212-222, 2008.

[6] L. F. Haas, "Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography," *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 74, no. 1, pp. 9-9, 2003.

[7] H. Berger, "On the Electroencephalogram of Man," *Electroencephalogr Clin Neurophysiol*, Suppl 28, pp. 37-73, 1969.

[8] E. Niedermeyer, F. H. L. da Silva, "Electroencephalography: Basic Principles, Clinical Applications, and Related Fields," *Lippincot Williams and Wilkins*, 2004.

[9] http://cec.sonus.ca/econtact/14_2/ortiz_biofeedback.html

[10] J. J. Carr and J. M. Brown, "Introduction to Biomedical Equipment Technology," *Prentice Hall*, 4th Edition, 2001.

[11] Webster JG, “Medical Instrumentation: Application and Design,” *John Wiley and Sons Inc*, 3th Edition, 1998.

[12] M.J. Burke and D.T. Gleeson, “A micropower dry-electrode ECG preamplifier,” *IEEE Trans. Biomed. Eng.*, vol. 47, pp. 155-162, 2000.

[13] M. Engine, T. Dalbasti, M. Gulduren, E. Davash, E. Z. Engin, “A prototype portable system for EEG measurements,” *Measurement*, vol. 40, pp. 936-942, 2007.

[14] J. L. Smith, S. H. Chug, R. F. Zernicke, “Gait-related motor patterns and hindlimb kinetic for the cat trot and gallop,” *Exp Brain Res*, vol. 94, pp. 308-322, 1993.

[15] K. R. Wheeler and C. C. Jorgensen, “Gestures as input: Neuroelectric joysticks and keyboards,” *Pervasive Comput.*, vol. 2, no. 2, pp. 56-61, Apr.-Jun 2003.

[16] M. A. Oskoei and H. Hu, “Myoelectric control systems-a survey,” *Biomed. Signal Process. Control*, vol. 2, no. 4, pp. 275-294, Oct. 2007.

[17] B. Karlik, M. O. Tokhi and M. Alci, “A fuzzy clustering neural network architecture for multifunction upper-limb prosthesis,” *IEEE Trans. Biomed. Eng.*, vol. 50, no. 11, pp. 1255-1261, 2003.

[18] J. U. Chu, I. Moon and M. S. Mun, “A Real-Time EMG Pattern Recognition System Based on Linear-Nonlinear Feature Projection for a Multifunction Myoelectric Hand,” *IEEE Trans. Biomed. Eng.*, vol. 53, no. 11, pp. 2232-2239, 2006.

[19] K. S. Kim, H. H. Choi, C. S. Moon, C. W. Mun, “Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions,” *Current Applied Physics*, vol. 11, no. 3, pp.740-745, 2011.

[20] G. Tsenov, A. H. Zeghbib, F. Palis, N. Shoylev, V. Mladenov, “Neural networks for online classification of hand and finger movements using surface EMG signals,” *NEUREL 2006*, pp. 167-171, Sep. 2006.

[21] G. Kondo, R. Kato, H. Yokoi, T. Arai, “Classification of individual finger motions hybridizing electromyogram in transient and converged states,” *in Proc. IEEE Eng. Med. Biol. Soc. Conf.*, pp. 2909-2915, 2010.

[22] F. Tenore, A. Ramos, A. Fahmy, S. Acharya, R. Etienne-Cummings, and N. V. Thakor, "Towards the control of individual fingers of a prosthetic hand using surface EMG signals," *in Proc. IEEE Eng. Med. Biol. Soc. Conf.*, pp. 6145-6148, 2007.

[23] S. N. Baker, E. Olivier, R. N. Lemon, "Coherent oscillations in monkey motor cortex and hand muscle EMG show task-dependent modulation," *J Physiol*, vol. 501, no. 1, pp. 225-241, 1997.

[24] B. A. Conway, D. M. Halliday, S. F. Farmer, U. Shahani, P. Maas, A. I. Weir and J. R. Rosenberg, "Synchronization between motor cortex and spinal motoneuronal pool during the performance of a maintained motor task in man," *J Physiol*, vol. 489, pp. 917-924, 1995.

[25] D. Halliday, B. Conway, S. Farmer, J. Rosenberg, "Using electroencephalography to study functional coupling between cortical activity and electromyograms during voluntary contractions in humans," *Neuroscience Letters*, vol. 241, no. 1, pp. 5-8, 1998.

[26] P. Brown, J. Marsden, etc., "Intermuscular coherence in Parkinson's disease: relationship to bradykinesia," *Neuroreport*, vol. 12, no. 11, pp. 2577-2581, 2001.

[27] B. Pollok, V. Krause, W. Martsch, C. Wach, A. Schnitzler, M. Sudmeyer, "Motor-cortical oscillations in early stages of Parkinson's disease," *The Journal of Physiology*, vol. 590, no. 13, pp. 3203-3212, 2012.

[28] S. Salenius, S. Avikainen, S. Kaakola, R. Hari and P. Brown, "Defective cortical drive to muscle in Parkinson's disease and its improvement with levodopa," *Brain*, vol. 125, pp. 491-500, 2002.

[29] J. Hirschmann, T. E. zkurt, Markus Butz, M. Homburger, S. Elben, C. J. Hartmann, J. Vesper, L. Wojtecki, and A. Schnitzler, "Differential modulation of STN-cortical and cortico-muscular coherence by movement and levodopa in Parkinson's disease," *NeuroImage*, vol. 68, pp. 203-213, March 2013.

[30] J. Chiang, Z. J. Wang, M. J. McKeown, "A multiblock PLS model of cortico-cortical interactions in Parkinson's disease during corticomuscular activity," *NeuroImage*, vol. 63, no. 3, pp. 1498-1509, 2012.

[31] R. Bortel, P. Sovka, "EEG-EMG coherence enhancement," *Signal Processing*, vol. 86, no. 7, pp. 1737-1751, 2006.

[32] P. Geladi and B. Kowalski, "Partial least-squares regression: a tutorial," *Analytica Chimica Acta*, vol. 185, pp. 1-17, 1986.

[33] A. McIntosh, F. Bookstein, J. Haxby, C. Grady, "Spatial pattern analysis of functional brain images using partial least squares," *NeuroImage*, vol.3, no. 3, pp. 143-157, 1996.

[34] E. Martinez-Montes, P. A. Valdes-Sosa, F. Miwakeichi, R. I. Goldman, M. S. Cohen, "Concurrent EEG/fMRI analysis by multiway partial least squares," *NeuroImage*, vol. 22, no. 3, pp. 1023-1034, 2004.

[35] A.M. Amjad, D.M. Halliday, J.R. Rosenberg, B.A. Conway, "An extended difference of coherence test for comparing and combining several independent coherence estimates: theory and application to the study of motor units and physiological tremor," *J Neurosci Methods*, vol. 73, no. 1, pp. 69-79, 1997.

[36] V. Murthy, E. Fetz, "Coherent 25-to 35-Hz oscillations in the sensorimotor cortex of awake behaving monkeys," *Proceedings of the National Academy of Sciences*, vol. 89, no. 12, pp. 5670-5674, 1992.

[37] S. Wold, N. Kettaneh, K. Tjessem, "Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection," *Journal of Chemometrics*, vol. 10, no. 5-6, pp. 463-482, 1998.

[38] M. G. Gustafsson, "Independent Component Analysis Yields Chemically Interpretable Latent Variables in Multivariate Regression," *J. Chem. Inf. Model.*, vol. 45, no. 5, pp. 1244-1255, 2005.

[39] A. Hyvarinen, E. Oja, "Fast fixed-point algorithm for independent component analysis," *Neural Comput.*, vol. 9, no. 7, pp. 1483-1492, 1997.

[40] G. R. Naik, D. K. Kumar, H. Weghorn, M. Palaniswami, "Subtle hand gesture identification for hci using temporal decorrelation source separation bss of surface EMG," *In: Digital image computing techniques and applications, 9th Biennial conference of the Australian pattern recognition society*, Adelaide, Australia, pp. 30-37, 2007.

[41] G. Lanckriet, T. De Bie, N. Cristianini, M. I. Jordan, and W. Stafford Noble. "A statistical framework for genomic data fusion," *Bioinfomatics*, vol. 20, no. 16, pp. 2626-2635, 2004.

[42] R. Agarwal, J. Gotman, "Adaptive Segmentation of Electroencephalographic Data Using a Nonlinear Energy Operator," *Proc. IEEE ISCAS 1999*, pp. 199-202, 1999.

[43] B. B. Winter, J. G. Webster, “Driven-Right-Leg Circuit Design,” *IEEE Trans. Biomed. Eng.*, vol. 30, pp. 62-66, 1983.

[44] H. L. Lew, S. J. TSAI, Pictorial guide to muscles and surface anatomy, *Johnson’s practical electromyography*, Lippincott Williams and Wilkins; pp. 145-212, 2007.

[45] G. Li, Y. Li, Z. Zhang, Y. Geng, R. Zhou, “Selection of sampling rate for EMG pattern recognition based prosthesis control,” *in Proc. IEEE Eng. Med. Biol. Soc. Conf.*, pp. 5058-5061, 2010.

[46] X. Zhang, X. Chen, Z. Zhao, Y. Tu, J. Yang, V. Lantz, K. Wang, “Research on gesture definition and electrode placement in pattern recognition of hand gesture action SEMG,” In *Medical Biometrics*, pp. 33-40, Springer Berlin Heidelberg, 2007.

[47] K. Englehart, B. Hudgins, P. A. Parker, M. Stevenson, “Classification of the myoelectric signal using time-frequency based representations,” *Medical Engineering and Physics*, vol. 21, no. 6-7, pp. 431-438, 1999.

[48] B. Hudgins, P. A. Parker, R. Scott, “A new strategy for multifunction myoelectric control,” *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82-94, 1993.

[49] S. Leowinata, B. Hudgins, P.A. Parker, “A multi function myoelectric control strategy using an array of electrodes,” *16th Annual Congress of the International Society Electrophysiology and Kinesiology*, Montreal, PQ, Canada, 1998.

[50] N. Uchida, A. Hiraiwa, N. Sonehara, K. Shimohara, “EMG pattern recognition by neural networks for multi fingers control,” *Proc. Ann. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 3, pp. 1016-1018, 1992.

[51] K. A. Farry, I. D. Walker, R.G. Baraniuk, “Myoelectric teleoperation of a complex robotic hand,” *IEEE Transactions on Robotics and Automation*, vol. 12, no. 5, pp. 775-788, 1996.

[52] K. Englehart, B. Hudgins, P. A. Parker, “A wavelet-based continuous classification scheme for multifunction myoelectric control,” *IEEE Trans. Biomed. Eng.*, vol. 48, pp. 302-311, 2001.

[53] K. Nazarpour, A. R. Sharafat, S. M. P. Firoozabadi, “Application of higher order statistics to surface electromyogram signal classification,” *IEEE Trans. Biomed. Eng.*, vol. 54, no. 10, pp. 1762-1769, Oct. 2007.

[54] F. Chan, Y.-S. Yang, F. K. Lam, Y.-T. Zhang, P. A. Parker, "Fuzzy EMG classification for prosthesis control," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 3, pp. 305-311, 2000.

[55] K. Englehart, B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 7, pp. 848-854, 2003.

[56] Y. Huang, K. Englehart, B. Hudgins, A. D.C. Chan, "A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 11, pp. 1801-1811, 2005.

[57] M. R. Ahsan, M. I. Ibrahimy, O. O. Khalifa, "EMG signal classification for human computer interaction: a review," *Eur J Sci Res*, vol. 33, no. 3, pp. 480-501, 2009.

[58] J. Kim, S. Mastnik, E. Andre, "EMG-based hand gesture recognition for realtime biosignal interfacing," *in Proc. 13th Int. Conf. Intell. User Interfaces*, Gran Canaria, Spain, pp. 30-39, 2008.

[59] T. Lorrain, N. Jiang, D. Farina, "Influence of the training set on the accuracy of surface EMG classification in dynamic contractions for the control of multifunction prostheses," *Journal of NeuroEngineering and Rehabilitation*, 8:25, 2011.

[60] L. Hargrove, E. Scheme, K. Englehart and B. Hudgins, "Principal components analysis tuning for improved myoelectric control," *Proceeding of the 30th Annual Conference of Canadian Medical and Biological Engineering Society*, 2007, Toronto.

[61] M.F. Lucas, A. Gaufriau, S. Pascual, C. Doncarli, D. Farina, "Multi-channel surface EMG classification using support vector machines and signal-based wavelet optimization," *Biomed. Signal Process. Control*, vol. 3, no. 2, 169-174, 2008.

[62] M. A. Oskoei, H. Hu, "Support vector machine-based classification scheme for myoelectric control applied to upper limb," *IEEE Trans. Biomed. Eng.*, vol. 55, pp.1956-1965, 2008.

[63] C. W. Hsu, C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Networks*, vol. 13, pp.415-425, 2002.

[64] Y. H. Liu, H. P. Huang, C. H. Weng, "Recognition of electromyographic signals using cascaded kernel learning machine," *IEEE/ASME Trans. Mechatron.*, vol. 12, no. 3, pp. 253-264, 2007.

[65] F. R. Bach, G. R. G. Lanckriet, M. I. Jordan, "Multiple kernel learning, conic duality, and the SMO algorithm," *Twenty-first international conference on Machine learning*, 2004.

[66] J. Friedman, T. Hastie, R. Tibshirani, *The Elements of Statistical Learning - Data Mining, Inference and Prediction*, 2nd ed., Springer, New York, 2001.

[67] S. Sonnenburg, G. Raetsch, C. Schaefer, and B. Schoelkopf. "Large scale multiple kernel learning," *J. Mach. Learn. Res.*, 7:1531-1565, 2006.

[68] A. Rakotomamonjy, F.R. Bach, S. Canu and Y. Grandvalet, "SimpleMKL," *J. Mach. Learn. Res.*, 9:2491-2521, 2008.

[69] G. Tropini, J. Chiang, Z. J. Wang, E. Ty, M. J. McKeown, "Altered Directional Connectivity in Parkinson's Disease During Performance of a Visually Guided Task," *Neuroimage*, vol. 56, no. 4, pp. 2144-2156, 2011.

[70] X. Chen, Z. J. Wang, M. J. McKeown, "A tridirectional method for corticomuscular coupling analysis in Parkinson's disease," *Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on*, pp. 309-312, 17-19 Sept. 2012.

[71] X. Chen, X. Chen, R. K. Ward, Z. J. Wang, "A Joint Multimodal Group Analysis Framework for Modeling Corticomuscular Activity," *IEEE Trans. on Multimedia*, vol. 15, no. 5, pp. 1049-1059, 2013.

[72] X. Chen, A. Liu, Z. J. Wang, H. Peng, "Modeling Corticomuscular Activity by Combining Partial Least Squares and Canonical Correlation Analysis," *Journal of Applied Mathematics*, vol. 2013, Article ID 401976, 11 pages, 2013.

[73] X. Chen, C. He, Z. J. Wang, M. J. McKeown, "An IC-PLS Framework for Group Corticomuscular Coupling Analysis," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 7, pp. 2022-2033, 2013.

[74] X. Chen, Z. J. Wang, M. J. Mckeown, "A Three-step Method for Corticomuscular Activity Modeling with Application to Parkinson's Disease," *IEEE Transactions on Information Technology in Biomedicine*, in press, 2013, doi:10.1109/JBHI.2013.2284480.

[75] B. S. Everitt, G. Dunn, *Applied Multivariate Data Analysis*, 2nd Ed., John Wiley and Sons, 2010.

[76] H. L. Yu, J. F. MacGregor, "Post processing methods (PLS-CCA): simple alternatives to preprocessing methods (OSC-PLS)," *Chemom. Intell. Lab. Syst.*, vol. 73, pp. 199-205, 2004.

[77] W. D. Clercq, A. Vergult, B. Vanrumste, W. V. Paesschen, S. V. Huffel, "Canonical Correlation Analysis Applied to Remove Muscle Artifacts From the Electroencephalogram," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2583-2587, 2006.

[78] N. M. Correa, T. Adali, Y. Li and V. D. Calhoun , "Canonical Correlation Analysis for Data Fusion and Group Inferences," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 39-50, 2010.

[79] E. Gumus, O. Kursun, A. Sertbas and D. Ustek, "Application of canonical correlation analysis for identifying viral integration preferences," *Bioinformatics*, vol. 28, no. 5, pp. 651-655, 2012.

[80] M. J. McKeown, S. Makeig, G. G. Brown, T. P. Jung, S. S. Kindermann, A. J. Bell, T. J. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," *Hum. Brain Mapp.*, vol. 6, no. 3, pp. 160-188, 1998.

[81] R. Vigario, J. Sarela, V. Jousmaki, M. Hamalainen, and E. Oja, "Independent Component Approach to the Analysis of EEG and MEG Recordings," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 5, pp. 589-593, 2000.

[82] V. D. Calhoun, J. Liu, T. Adali, "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data," *NeuroImage*, vol. 45, no. 1, pp. S163-S172, 2009.

[83] H. Kaneko, M. Arakawa, K. Funatsu, "Development of a new regression analysis method using independent component analysis," *J. Chem. Inf. Model*, vol. 48, pp. 534-541, 2008.

[84] V. D. Calhoun, T. Adali, K. A. Kiehl, R. Astur, J. J. Pekar, G. D. Pearlson, "A method for multitask fMRI data fusion applied to schizophrenia," *Human brain mapping*, vol. 27, no. 7, pp. 598-610, 2006.

[85] F. Lindgren, P. Geladi, S. Wold, "The kernel algorithm for PLS," *Journal of Chemometrics*, vol. 7, no. 1, pp. 45-59, 1993.

[86] S. Weiss, P. Rappelsberger, "Long-range EEG synchronisation during word encoding correlates with successful memory performance," *Cognitive Brain Research*, vol. 9, pp. 299-312, 2000.

[87] E. Clancy, E. Morin, and R. Merletti, "Sampling, noise-reduction and amplitude estimation issues in surface electromyography," *Journal of Electromyography and Kinesiology*, vol. 12, no. 1, pp. 1-16, 2002.

[88] P. Good, *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*, 2nd Ed., Springer Series in Statistics, Springer-Verlag, Heidelberg, 2000.

[89] P. Praamstra, L. Boutsen, G. W. Humphreys, "Frontoparietal control of spatial attention and motor intention in human EEG," *Journal of Neurophysiology*, vol. 94, no. 1, pp. 764-774, 2005.

[90] J. Liepert, C. Dettmers, C. Terborg, C. Weiller, "Inhibition of ipsilateral motor cortex during phasic generation of low force," *Clinical Neurophysiology*, vol. 112, no. 1, pp. 114-121, 2001.

[91] C. Danielmeier, T. Eichele, B. U. Forstmann, M. Tittgemeyer, M. Ullsperger, "Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas," *The Journal of Neuroscience*, vol. 31, no. 5, pp. 1780-1789, 2011.

[92] P. Sauseng, W. Klimesch, R. Freunberger, T. Pecherstorfer, S. Hanslmayr, M. Doppelmayr, "Relevance of EEG alpha and theta oscillations during task switching," *Experimental Brain Research*, vol. 170, no. 3, pp. 295-301, 2006.

[93] I. G. M. Cameron, M. Watanabe, G. Pari, D. P. Munoz, "Executive impairment in Parkinson's disease: response automaticity and task switching," *Neuropsychologia*, vol. 48, no. 7, pp. 1948-1957, 2010.

[94] P. Praamstra, D. F. Stegeman, A. R. Cools, M. W. Horstink, "Reliance on external cues for movement initiation in Parkinson's disease. Evidence from movement-related potentials," *Brain*, vol. 121, no. 1, pp. 167-177, 1998.

[95] J. K. R. Stevenson, M. M. K. Oishi, S. Farajian, E. Cretu, E. Ty, M. J. McKeown, "Response to sensory uncertainty in Parkinson's disease: a marker of cerebellar dysfunction?" *European Journal of Neuroscience*, vol. 33, no. 2, pp. 298-305, 2011.

[96] X. Lu, *Optimization Applications and Basis* , Shanghai: Tongji University Press; 2003.

[97] M. A. Guevara, M. Corsi-Cabrera, "EEG coherence or EEG correlation?" *Int. J. Psychophysiol.*, vol. 23, no. 3, pp. 145-153, 1996.

[98] S. Haykin, *Neural Networks*, Upper Saddle River, NJ: Prentice Hall International; 1999.

[99] D. G. Luenberger, *Optimization by Vector Space Methods*, New York: Wiley, 1969.

[100] G. Tropini, J. Chiang, Z. J. Wang, E. Ty, M. J. McKeown, "Altered Directional Connectivity in Parkinson's Disease During Performance of a Visually Guided Task," *Neuroimage*, vol. 56, no. 4, pp. 2144-2156, 2011.

[101] B. Ng, S. Palmer, R. Abugharbieh, M. J. McKeown, "Focusing effects of L-dopa in Parkinson's disease," *Hum. Brain Mapp.*, vol. 31, no. 1, pp. 88-97, Jan 2010.

[102] P. Capotosto, C. Babiloni, G. L. Romani, M. Corbetta, "Frontoparietal cortex controls spatial attention through modulation of anticipatory alpha rhythms," *J Neurosci*, vol. 29, pp. 5863-5872, 2009.

[103] M.J. Wright, R.J. Burns, G.M. Geffen, L.B. Geffen, "Covert orientation of visual attention in Parkinson's disease: An impairment in the maintenance of attention," *Neuropsychologia*, vol. 28, no. 2, pp. 151-159, 1990.

[104] P. Brown, "Abnormal oscillatory synchronisation in the motor system leads to impaired movement," *Current Opinion in Neurobiology*, vol. 17, no. 6, pp. 656-664, 2007.

[105] N. Jenkinson, P. Brown, "New insights into the relationship between dopamine, beta oscillations and motor function," *Trends in Neurosciences*, vol. 34, no. 12, pp. 611-618, 2011.

[106] B. Pollok, V. Krause, W. Martsch, C. Wach, A. Schnitzler, M. Sudmeyer, "Motor-cortical oscillations in early stages of Parkinson's disease," *J Physiol*, DOI: 10.1113/jphysiol.2012.231316, 2012.

[107] H. Gu, Z. Pan, B. Xi, V. Asiago, B. Musselman, D. Raftery, "Principal component directed partial least squares analysis for combining nuclear magnetic resonance and mass spectrometry data in metabolomics: application to the detection of breast cancer," *Anal Chim Acta*, vol. 686, no. 1-2, pp. 57-63, 2011.

[108] C. Zhao, F. Gao, "A bidirectional between-set statistical analysis method and its applications," *JAIChE Journal*, 57(5):1233-1249, 2011.

[109] H. Attias, S. Lee, T. Lee, "Blind Source Separation Exploiting Higher-Order Frequency Dependencies," *IEEE Transactions on Audio, Speech, and Language Processing*, 15(1):70-79, 2007.

[110] Y. Li, T. Adali, W. Wang, V. Calhoun, "Joint Blind Source Separation by Multiset Canonical Correlation Analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918-3929, 2009.

[111] T. Fearn, "On orthogonal signal correction," *Chemom Intell Lab Syst.*, vol. 50, pp. 47-52, 2000.

[112] Z. J. Wang, P. W. Lee, and M. J. McKeown, "A Novel Segmentation, Mutual Information Network Framework for EEGAnalysis of Motor Tasks," *BioMedical Engineering Online*, vol. 8, no. 9, 2009.

[113] J. Sui, T. Adali, G. Pearlson, H. Yang, S. Sponheim, T. White, and V. D. Calhoun, "A CCA+ ICA based model for multi-task brain imaging data fusion and its application to schizophrenia," *Neuroimage*, vol. 51, no. 1, pp. 123-134, 2010.

[114] J. Sui, G. Pearlson, A. Caprihan, T. Adali, K. A. Kiehl, J. Liu, J. Yamamoto, and V. D. Calhoun, "Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA+ joint ICA model," *Neuroimage*, vol. 57, no. 3, pp. 839-855, 2011.

[115] M. J. Wright, R.J. Burns, G.M. Geffen, L.B. Geffen, "Covert orientation of visual attention in Parkinson's disease: An impairment in the maintenance of attention," *Neuropsychologia*, vol. 28, no. 2, pp. 151-159, 1990.

[116] E. Florin, R. Erasmi, C. Reck, M. Maarouf, A. Schnitzler, G. R. Fink, L. Timmermann, "Does increased gamma activity in patients suffering from Parkinson's disease counteract the movement inhibiting beta activity?," *Neuroscience*, vol. 237, pp. 42-50, 2013.

[117] T. Kim, T. Eltoft, TW Lee, "Independent vector analysis: An extension of ICA to multivariate components," *Independent Component Analysis and Blind Signal Separation*, pp. 165-172, Springer Berlin Heidelberg, 2006.

[118] J. H. Lee, T. W. Lee, F. A. Jolesz, S. S. Yoo, "Independent vector analysis (IVA): multivariate approach for fMRI group study," *Neuroimage*, vol. 40, no. 1, pp. 86-109, 2008.

[119] M. Anderson, T. Adali, XL Li, "Joint blind source separation with multivariate Gaussian model: algorithms and performance analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1672-1683, 2012.

[120] M. J. McKeown, R. Saab, R. Abu-Gharbieh, "A combined independent component analysis (ICA)/empirical mode decomposition (EMD) method to infer corticomuscular coupling," *IEEE EMBS 2nd International Conference on Neural Engineering*, pp. 679-682, 2005.

[121] M. E. Davies, C. J. James, "Source separation using single channel ICA," *Signal Processing*, vol. 87, no. 8, pp. 1819-1832, 2007.

[122] P. Bofill, M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal processing*, vol. 81, no. 11, pp. 2353-2362, 2001.

Appendix A

Prototype Development

The reason we put this part in Appendix is that we did not use our developed system to collect data for corticomuscular coupling analysis although our original purpose was to utilize an integrated wireless EEG and EMG system for patients' convenience. We were required to use commercialized medical devices in scientific publications. However, the prototype is still very useful to develop novel applications for daily-life usage.

A.1 Introduction

Biosensor monitoring systems have been attracting increasing interest, stimulated by recent advances in electronic, communications and information technologies. For instance, EEG is the most popular non-invasive recording technology studied in BCI, a hot topic aiming at creating a direct link between the brain and a computer or a control device. However, most BCI systems use bulky and wired EEG measurements, which is uncomfortable and inconvenient for users to perform daily-life tasks. To address this concern, in Section A.2 we design a wearable, wireless EEG acquisition and recording system, including a data acquisition unit and a data transmission/receiving unit. Our testing results show that the proposed design is effective with high common-mode rejection ratio (CMRR). The developed EEG system is suitable for biomedical applications such as patient monitoring and BCI.

Based on the developed EEG system, in Section A.3 we modify it and build a sEMG hand gesture recognition system. As mentioned in Section 1.2.1, regarding the research direction of multiple finger gesture classification, to our knowledge, there are relatively few research works in literature and there has been no multiple finger gesture benchmark proposed yet. In [40], the authors used BSS and artificial neural network (ANN) techniques to off-line classify four subtle hand gestures by four wired sEMG channels and achieved 97% average accuracy for seven subjects, however, only three of the four gestures can be categorized as multiple finger gesture tasks. In this chapter, we plan to define a set of multi-finger movement tasks which could be commonly used for benchmark testing. Based on our extensive preliminary investigation, we choose number gestures as an appropriate set of multi-finger movements because of the following reasons: they satisfy the basic definition of multi-finger movements; and people use number gestures frequently in real-life especially when there are language communication difficulties, which is indeed a major reason why originally number gestures were invented.

In this chapter, we plan to study Chinese number gestures representing the numbers zero to nine, as illustrated in Fig. A.1. We propose to recognize these ten classes of subtle hand gestures based on a 4-channel wireless sEMG system. The recognition procedure is implemented in two phases. In Phase-1, we develop off-line recognition algorithms to study their recognition performances and check the feasibility of implementing a real-time recognition system. We first investigate the most popular feature extraction and classification algorithms; then based on observed performances we further propose to combine all three features together and employ multiple kernels to adapt the combined feature set structure by multi-

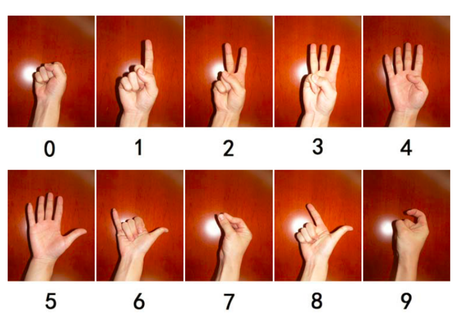


Figure A.1: Illustration of the Chinese number gestures.

ple kernel learning (MKL). In support vector machine (SVM), a kernel function implicitly maps samples to a feature space given by a feature map. It is often unclear what the most suitable kernel for the task at hand is, and hence the user may wish to combine several possible kernels. One problem with simply adding kernels is that using uniform weights is possibly not optimal. For instance, if one kernel is not correlated with the labels at all, then giving it a positive weight just adds noise [41]. MKL is an efficient way of optimizing kernel weights. In Phase-2, we implement a real-time sEMG recognition system for Chinese number gestures and demonstrate its online performance as a preliminary study for possible practical applications. For the two phases, we achieved 97.93% and 95% high average recognition accuracies respectively.

A.2 A Wireless Wearable EEG Recording System

A.2.1 System Design

The architecture of the designed EEG recording system is shown in Fig. A.2.

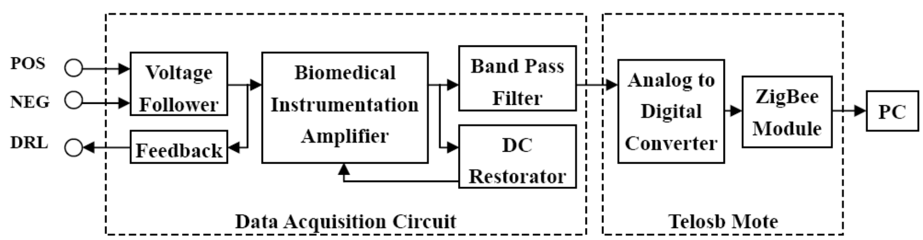


Figure A.2: The architecture of the designed EEG recording system.

EEG Acquisition Circuit

In order to achieve a reliable EEG recording system, the design of the data acquisition circuit is the most important step. The schematic of a single channel EEG acquisition circuit is shown in Fig. A.3. The details of each element in the circuit will be presented as follows.

1. **EEG features:** A typical human adult EEG signal is about 10 to 100 μV in

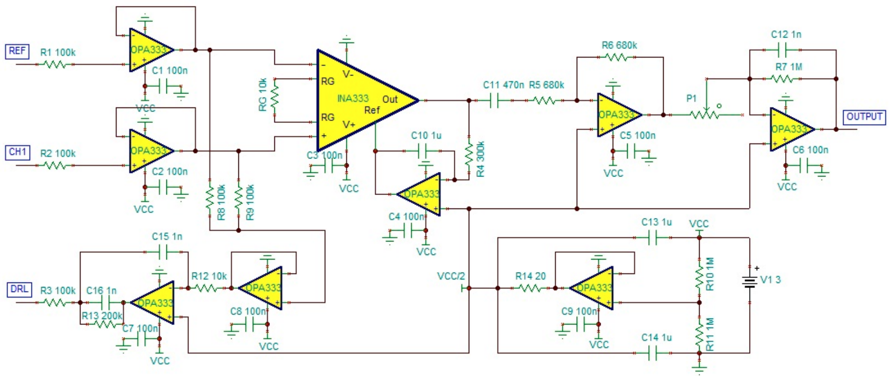


Figure A.3: The schematic of a single channel EEG acquisition circuit.

amplitude when measured from the scalp and it is about 10 to 20 mV when measured from subdural electrodes. Since EEG signals are usually acquired through passive electrodes from head surface, they are easily contaminated by artifacts that are not directly related to brain electrical activity [42]. The artifacts in the recorded EEG can be either technical such as AC power line noise or person-related such as EMG, ECG and electrooculogram (EOG).

2. **Voltage follower:** The voltage follower is constructed by applying a full series negative feedback to an operational amplifier (op-amp) simply by connecting its output to its inverting input, and connecting the signal source to the non-inverting input. This circuit has an output which is identical to the input voltage. The importance of this circuit does not come from any change in voltage, but from the input and output impedances of the op-amp. The input impedance of the op-amp is very high (usually 1 MΩ to 10 TΩ), meaning that the input of the op-amp does not load down the source or draw any current from it. On the other hand, the output impedance of the op-amp is very low and thus it behaves as a perfect voltage source. Both the connections to and from the voltage follower are therefore voltage bridging connections, which can reduce power consumption in the source, distortion from overloading, crosstalk and other electromagnetic interferences. In this design, all op-amps are OPA333 (Texas Instruments, Inc., Dallas, TX). The OPA333 series of CMOS op-amps adopt a proprietary auto-calibration technique to

simultaneously provide very low offset voltage (10 μV max) and near-zero drift over time and temperature. These miniature, high-precision, low quiescent current amplifiers offer high-impedance inputs and low-impedance outputs with micro-power, making them quite suitable for battery-powered, micro-sized portable medical devices.

3. **Instrumentation Amplifier:** The function of instrumentation amplifier is to amplify the voltage difference between the signals from *REF* and *CH1* EEG electrodes, and simultaneously reject common-mode noise, such as power line interference, at both inputs. The INA333 (Texas Instruments, Inc., Dallas, TX) is the instrumentation amplifier chosen for this design since it is the industrial lowest power zero-drift instrumentation amplifier. It provides very low offset voltage, excellent offset voltage drift, micro power consumption, very low quiescent current and high CMRR (110 dB typically), which ensures excellent precision and stability while extending battery life in portable medical devices. In our design, *RG* is set to 10 kΩ and the amplification gain is 11.

4. **DC restorator:** The direct current (DC) restorator is used to eliminate DC offset which would otherwise saturate the op-amps subsequent to the instrumentation amplifier. Here the DC restorator is implemented by using an op-amp in the feedback loop of INA333. The model in Fig. A.4 is used to design the DC restorator.

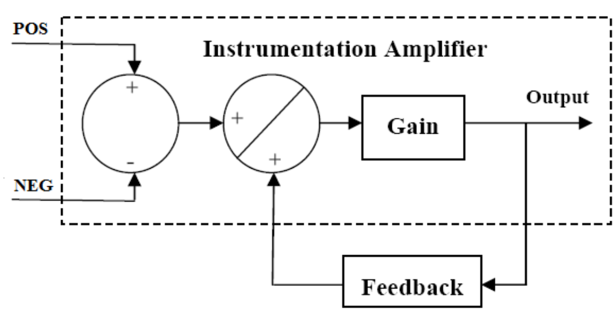


Figure A.4: The DC restorator model.

5. **Drive Right Leg:** To reduce common-mode noise, such as power line interference (50/60Hz) which is unavoidably coupled into the human subject, the driven right leg (DRL) is necessary to be used in the design. The name, DRL, has been preserved from its first use in ECG equipment [43], although the feedback electrode is not connected to the subject's right leg for EEG applications. As shown in Fig. A.3, the DRL is a feedback circuit, feeding the inverse of common-mode voltage back to the human subject, which acts to reduce the common-mode noise present at the electrodes.

6. **Band pass filter:** The band pass filter is composed of a first order active high pass filter and a first order active low pass filter, cutting off the signals with frequency components below 0.5 Hz and above 150 Hz. The high pass filter is designed to remove DC voltage offsets and to reduce the baseline drift, while the low pass filter is used to prevent aliasing in the analog-to-digital converter (ADC) step. The gain of the band pass filter is set to 1000, which can be changed by adjusting the adjustable resistor *P1*. Therefore, the total amplification gain of the EEG acquisition circuit is 11000.

7. **Power supply:** Power for the board is supplied by a button battery, Panasonic CR2032 with 3V (VCC) output voltage. Through a voltage divider and a voltage follower, a stable 1.5V (VCC/2) can be obtained and used as a virtual ground for the entire system. If we use a CR2032 to supply power for both EEG acquisition circuit and the TelosB mote (Crossbow, Inc., Milpitas, CA), the entire system can work normally for at least 8 hours, with most energy being consumed by the TelosB mote for wireless data transmission.

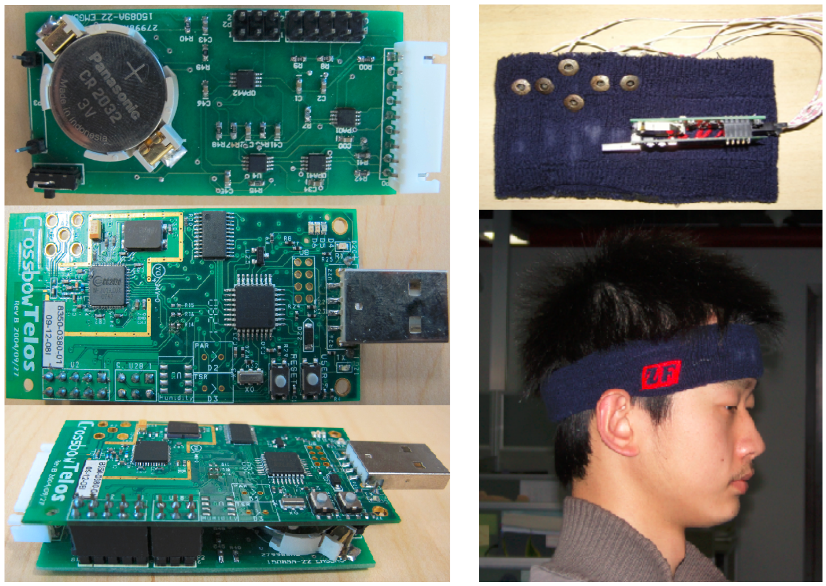
8. **Physical circuit:** A two-layer 3.2 cm $\times$ 6.6 cm printed circuit board (PCB) is designed for the 4-channel EEG acquisition unit. The analog channels are laid out in both sides of the board. A high board density is achieved by using the second smallest available hand-solderable parts (size 0603 for the passive components), amplifiers with MSOP8 footprint (3.0 mm $\times$ 3.0 mm), 8 mil signal trace width and 35 mil vias with 20 mil drill holes. We feel that such design is suitable for a portable EEG recording system.

Wireless Transmission Unit

The wireless transmission unit consists of three parts, which are all based on the TinyOS platform. Two of the three are TelosB motes (one for the transmission side and the other for receiver on the server side), and another is a Java package for the server. TelosB mote is an open source platform designed to enable cutting-edge experimentation for research community. The TelosB mote bundles all the essentials for lab studies into a single platform, which includes: Universal Serial Bus (USB) programming capability, an IEEE 802.15.4 radio with integrated antenna, a low-power microcontroller (MCU) with extended memory, and an optional sensor suite. To transmit data wirelessly, the EEG signals acquired from the analogue circuit go through the ADC equipped with MSP430 on the TelosB mote. It provides 12-bit resolution, and can be configured in different modes. The sampling frequency is selectable and is set to be 333 Hz here, satisfying the Nyquist Theory in terms of the -3dB cutoff frequency of 150 Hz. The data receiver side uses another TelosB mote as the base station which is connected with the PC via USB, and the PC continuously saves the received data into files by a Java-based application. The indoor transmission range can achieve 15 m, which is satisfactory for people's daily usage. The TelosB mote is also a 3.2 cm $\times$ 6.6 cm PCB. With the two PCBs, the height and the weight of the designed system are only 1.2 cm and 56 g (including the button battery) respectively as in Fig. A.5a, which includes data acquisition unit, Telosb mote and their combination from top to bottom. Furthermore, since ZigBees can sleep most of the time, the average power consumption can be very low, resulting in a longer battery life. To make the device portable, we design a head band which can be easily worn as shown in Fig. A.5b.

A.2.2 Experiments and Results

The developed EEG recording system has been tested on one human subject. To evaluate the signal quality of the designed system, the system was first used to measure ECG signals by adjusting the adjustable resistor *P1* to make the gain of band pass filter equal to 100. We connected three electrodes to V4, V5 and right leg of the subject, and asked the subject to remain static in order to avoid movement artifacts. From Fig. A.6a, we can see that the designed system can successfully

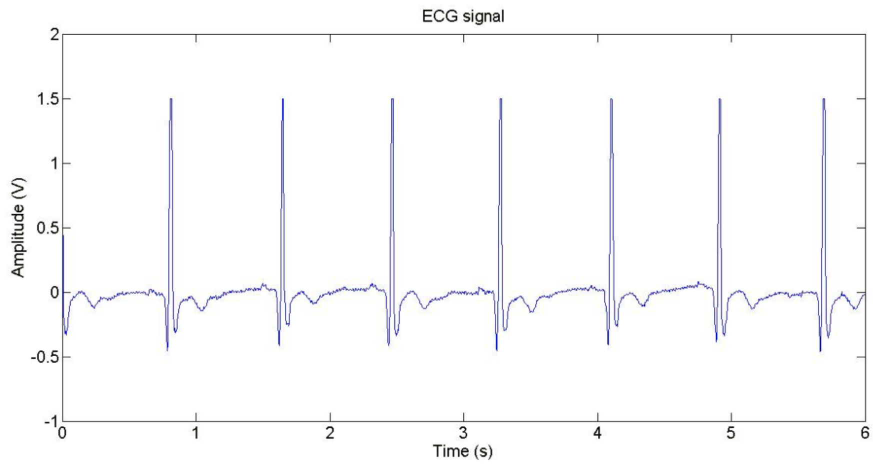


(a) **(b)**

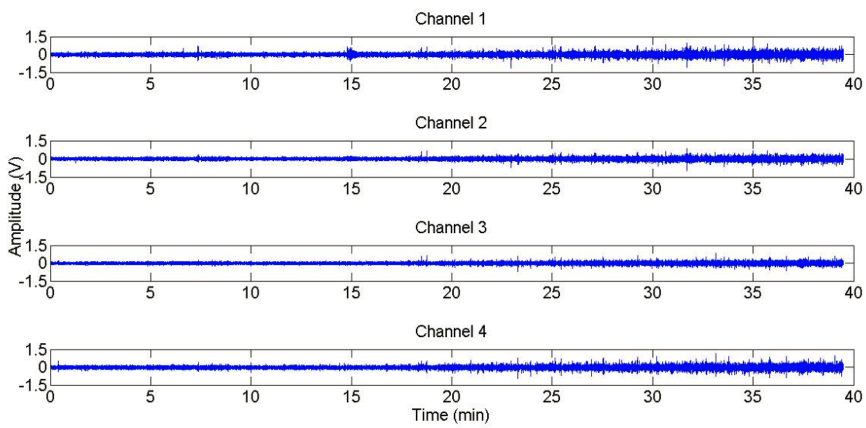
Figure A.5: (a) Data acquisition unit, telosb mote and their combination from top to bottom; (b) The wireless EEG head band.

capture P wave, QRS complex and T wave of the subject's ECG signal. The measured ECG signal seems to have high quality with little noise.

Further, the designed system was used to measure EEG signals by six electrodes placed on the frontal lobe of the head. Four electrodes were placed on FP1, FP2, F7 and F8 (10-20 System of Electrode Placement), and the reference and DRL electrodes were placed on midline position of the frontal head. The subject was lying in bed at night, being mentally prepared for sleep. The measurement lasted for about 40 minutes, and an observation should be noticed that the subject showed light symptoms of waking up temporarily around the 10-minute time point. The EEG signals recorded from the designed 4-channel EEG system were shown in Fig. A.6b. The subject experienced from awake and relaxed state to asleep, as supported by the observations that Delta wave, normally seen in adults in slow-wave sleep, was the highest in amplitude and Alpha wave, emerging with closing of the eyes and with relaxation, was relatively higher in amplitude.



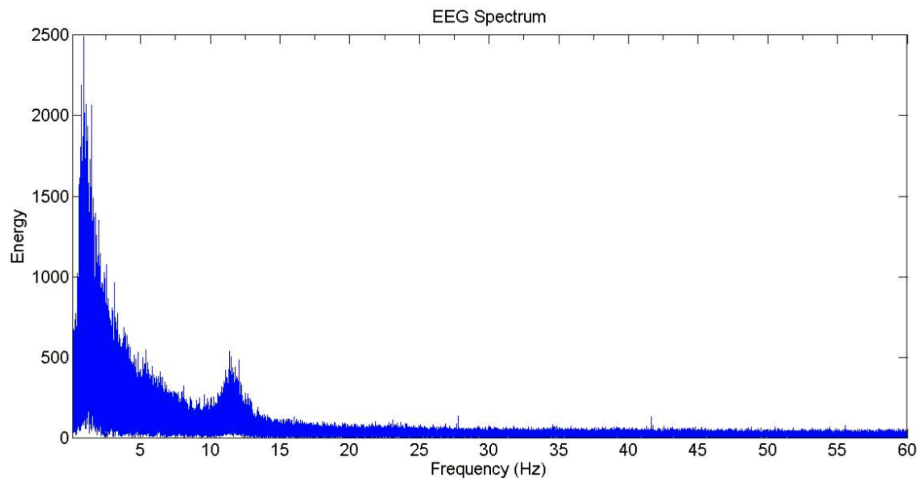
(a)



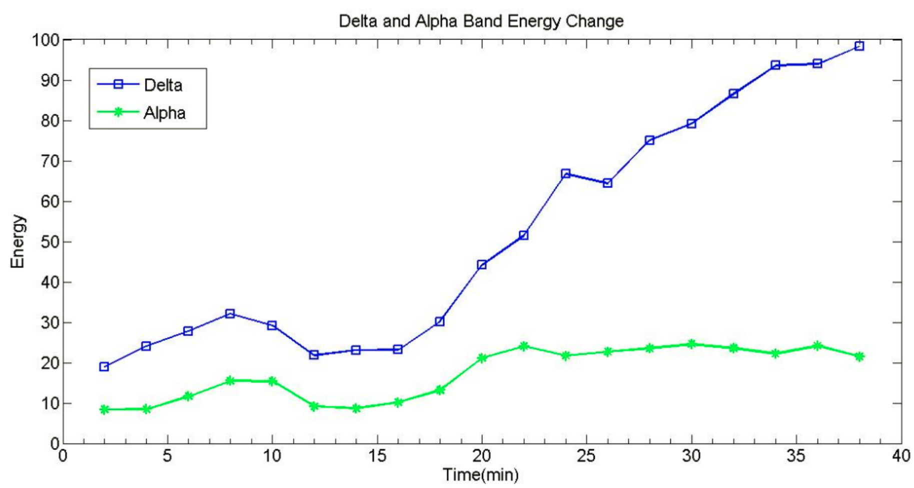
(b)

Figure A.6: (a) ECG reading from the designed EEG recording system; (b) EEG readings from the 4-channel EEG recording system.

After performing fast Fourier transform (FFT) of the EEG signal from Channel 1, we illustrated its frequency properties in Fig. A.7a, where the energies between 1 and 4 Hz (the Delta band) and that between 8 and 12 Hz (the Alpha band) are obviously stronger than that in other frequency bands. Also, from Fig. A.7a, we can see that there is no obvious unwanted interference in the EEG signals.



(a)



(b)

Figure A.7: (a) FFT results of the Channel-1 EEG signal; (b) The temporal energy changes of Delta and Alpha waves;

To have an insight into the measured EEG signals, we used the EEG signal from Channel 3 for further analysis. The EEG signal was uniformly divided into 19 equal segments, with each segment including two-minute EEG signal. For each segment, the average energy between 1 and 4 Hz and that between 8 and 12 Hz were calculated, and thus we could plot the energies of the Delta and Alpha bands as a function of time. As shown in Fig. A.7b, the energy in Delta band was generally increasing and had a trend to further increase with the subject falling asleep deeper and deeper, while the energy in Alpha band seemed keep stable after 20 minutes, which means that the subject had relaxed totally after 20 minutes, but not entered the deepest sleep state yet. We also noted another interesting observation: around the 10-minute time point, both the energies in Delta and Alpha bands tended to increase first and then decreased, which coincided with what we observed about the subject at that time during the experiment.

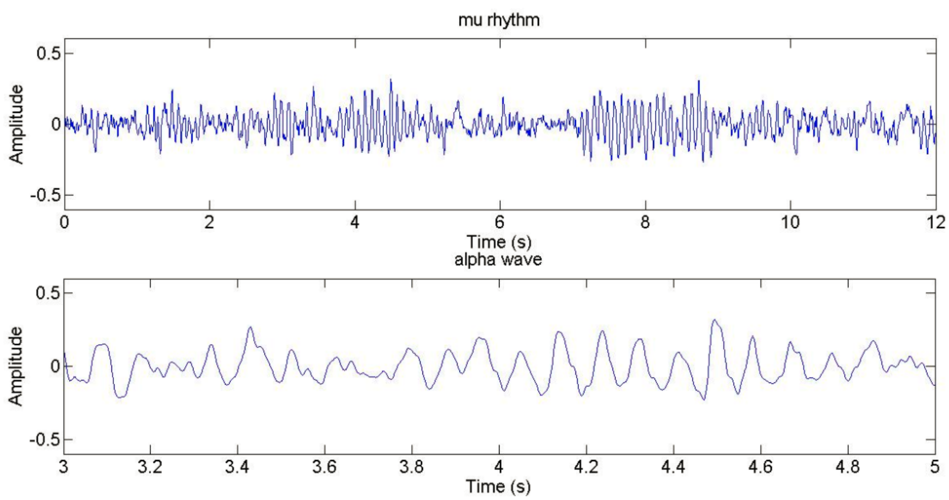


Figure A.8: Alpha waves detected on posterior regions of head.

At last, to demonstrate that the EEG signals were with high quality, we placed two electrodes on Cz and Pz positions to test Alpha waves when the subject closed eyes and relaxed. The testing results were illustrated in Fig. A.8, which presented clear mu rhythm and alpha waves.

A.3 A Wireless Portable sEMG Recognition System

A.3.1 System Design

The basic architecture of our developed 4-channel wireless portable sEMG recording system is the same as that of the EEG system. However, there are some important differences between them. First, each channel has a reference electrode in the sEMG system while all channels in the EEG system share one common reference electrode; second, the bandwidth of the band pass filter in sEMG system is from 20 Hz to 250 Hz since the most useful information reside in the selected band and then the sampling frequency should be changed; third, the overall gain of the sEMG circuit is set to be about 1000 while the gain in EEG system was about 10000 since the amplitude of sEMG measured from human forearms is generally 10 times higher that that of EEG from human scalp; fourth, the feedback (DRL) of sEMG system could be the direct circuit ground while it was not the case for EEG. Besides, a real-time recognition module has been added into the sEMG recording system which would be shown in the next section.

A.3.2 An Application for Chinese Number Guesture Recognition

Electrode Placement

Based on the movement tasks and the muscle anatomy of human upper limb, several forearm muscles can contribute to the designed multi-finger movements, with details of which can be found in [44]. Through a large number of preliminary experiments and our previous study [46], we identify four forearm muscles as suitable candidates for the designed recognition problem of Chinese number gestures. The four selected muscles are Extensor Pollicis Brevis, Extensor Digitorum, Flexor Digitorum Profundus for little finger and Flexor Digitorum Superficialis. The corresponding four sEMG electrode pairs are placed over these muscles as shown in Fig. A.9.

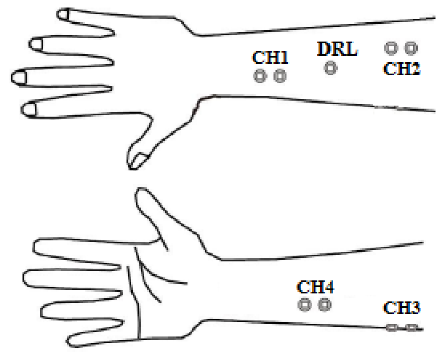


Figure A.9: Illustration of the sEMG electrode pair placement: CH1: over Extensor Pollicis Brevis muscle; CH2: over Extensor Digitorum muscle; CH3: over Flexor Digitorum Profundus muscle; and CH4: over Flexor Digitorum Superficialis muscle.

Experimental Protocol

Six healthy subjects (three females and three males, aged from 23 to 28, all right-handed) volunteered for this study. We used the 4-channel wireless sEMG system to collect sEMG signals from each subject by disposable AgCl electrodes (Junkang Medical Equipment, Inc., Shanghai) placed on the locations as indicated in Fig. A.9. The electrodes have the size 3 cm $\times$ 2 cm with pre-placed electric conduction gel in the middle circle part (1 cm diameter). The sampling frequency was set to be 500 Hz since the most useful energy in sEMG signals is in the range of 0-250 Hz [45]. The experimental procedure consists of two phases: the first one is for off-line analysis and the second one for online analysis. At the beginning of each session in the experiments, we needed to re-place the electrodes. To find corresponding muscles correctly, we asked the subjects to perform certain hand actions suggested by an anatomist. For example, to locate Extensor Pollicis Brevis, the subjects were required to extend their thumbs. Also, before the electrode placement, we used alcohol prep pads to clean the corresponding locations for reducing the electrode-skin impedance.

In the off-line experiment, each subject sat in a comfortable chair and naturally performed the ten Chinese number gesture movements. Each movement lasted for about 0.5 second and the time interval between two adjacent movements was about 1 to 2 seconds. For each of the ten number gestures in the set, we collected 50 trials from each subject to provide enough data samples for training and classification. All sEMG data from one subject were collected within one session. The popular leave-one-out cross validation method is employed on trial bases to calculate the classification accuracy during off-line analysis.

In the online experiment, we collected sEMG data from each subject in nine separate sessions, with 20 trails for each number gesture during each session. Each subject completed the nine separate sessions within ten days. The time interval between two adjacent sessions was more than 8 hours and at the beginning of each session the new electrodes were re-placed as in Fig. A.9. When a subject did the i-th session recording, sEMG data from the 1-st to the $(i-1)$-th sessions were treated as the training set. Therefore, subjects can adjust his/her movements appropriately according to the feedback results of the real-time recognition system.

Recognition Procedure

In this part, we first describe how to detect active segments in multichannel sEMG signals which represent the multi-finger movements. We then describe several popular feature extraction methods and classification algorithms, and further propose a MKL-SVM approach by combining all the features. Finally, we evaluate their classification performances for Chinese number gesture recognition. The basic diagram of the recognition procedure is shown in Fig. A.10.

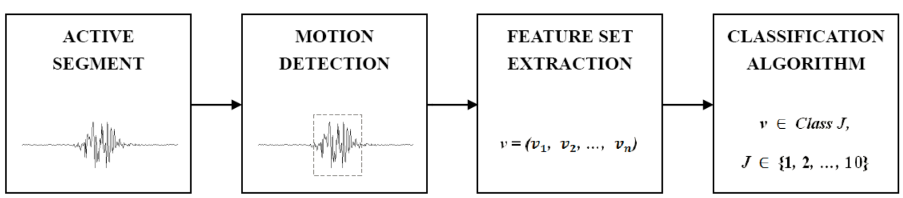


Figure A.10: The basic diagram of the recognition procedure.

sEMG motion detection The simple moving average method is employed here to process the transient energy of sEMG signals for motion detection, which generally consists of three steps.

First, the summation of sEMG signals $s_c(t)$'s is computed and the average $\bar{s}(t)$ is calculated as

$$\bar{s}(t) = \frac{1}{C}\sum_{c=1}^{C} s_c(t), \tag{A.1}$$

where C means the total number of sEMG channels and c means the sEMG channel index. Then, the square of $\bar{s}(t)$ is calculated to get the transient energy $E(t)$ as

$$E(t) = \bar{s}^2(t). \tag{A.2}$$

Second, the width of the moving window is set to be $W = 32$ points (about 64 ms when the sampling frequency is 500 Hz). Using the moving window to calculate the average of the transient energy $E(t)$, we have $E_{MA}(t)$ as

$$E_{MA}(t) = \frac{1}{W}\sum_{i=t-W+1}^{t} E(i). \tag{A.3}$$

Third, a suitable threshold value TH is chosen and the start point and end

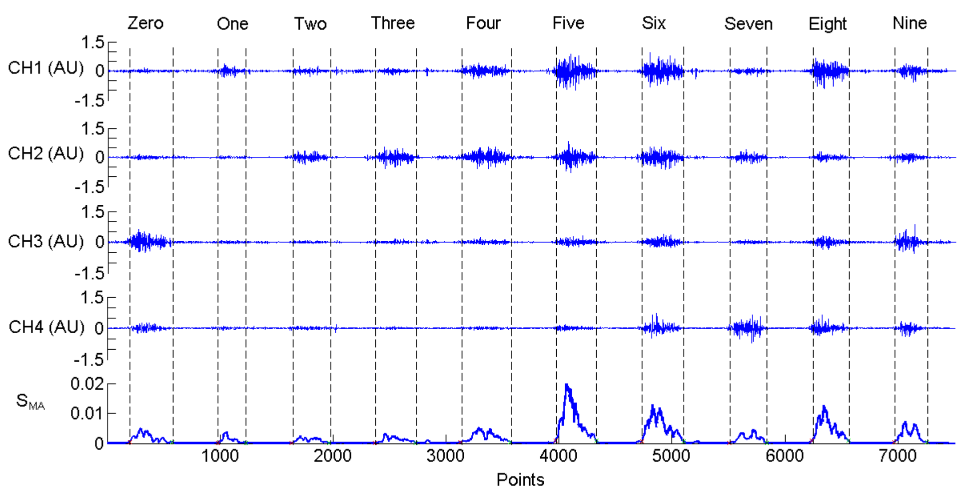


Figure A.11: An example of 4-channel sEMG signals and the corresponding motion detection results of Chinese number gestures.

point of each motion segment can be detected. The start point is defined as the point where the value of $E_{MA}(t)$ is larger than TH and the subsequent W values of $E_{MA}(t)$ are also larger than TH. The end point is defined as the point where the value of $E_{MA}(t)$ is smaller than TH and the subsequent W values of $E_{MA}(t)$ are also smaller than TH. If the time interval between a pair of the start point and the end point is too short, e.g. less than 50 points, we suggest that the corresponding segment is due to noise and should be discarded. Due to our good-quality sEMG signals, we set TH to be 1% of the maximum value of $E_{MA}(t)$'s.

An example of the 4-channel sEMG signals and the motion detection results of the ten Chinese number gestures are illustrated in Fig. A.11.

Feature extraction After the motion detection operation, suitable features should be extracted from each active sEMG segment to compose a feature vector for further classification. Because feature extraction has been shown to have a greater effect on classification accuracy than the type of classifier selected [47], it has been studied thoroughly in previous sEMG-based recognition and classification research. The most popular features are the following six types: TD [48], autocorrelation and cross-correlation coefficients (ACCC) [49], spectral power magnitudes (SPM) [50, 51], short-time Fourier transform (STFT) [47], wavelet transform (WT) [52] and high order statistics (HOS) [53]. However, we feel that STFT, WT and HOS are unsuitable for our case due to several reasons: to use STFT, the original active segments are required to have the same length to acquire transformed feature vectors with the same dimension, however the lengths of active segments for the transient hand gestures are highly unlikely to be the same; Since the time duration of a transient motion is generally about 300 ms, if STFT is used, the frequency resolution would be so low that there will be no meaningful information existing even if it produces high dimensional feature vectors; WT has a more strict requirement that the length of active segments must be a power of two, i.e. 2^n; HOS usually performs well for stationary or weak-sense stationary signals such as sEMG signals with continuous muscle contraction, which is different from our case. Therefore, only TD, ACCC and SPM will be investigated in details in this chapter.

Hudgins' time-domain features: Hudgins' time-domain feature set was introduced in 1993 [48] and has been widely used in the myoelectric control field. It was originally comprised of five different features, including mean absolute value (MAV), mean absolute value slope (MAVS), zero crossings (ZC), slope sign changes (SSC) and waveform length (WL). However, it was reported that the inclusion of SSC in the feature set contributed either a negative effect or no significant effect on classification performances [54]. The MAVS was another feature usually excluded from the TD feature set [55, 56] perhaps due to the similar reason although no clear justification was stated in the literature. Based on our preliminary study, to improve the classification accuracy, we add a new feature called the mean absolute value ratio (MAVR) into the Hudgins' time-domain feature set. MAVR can eliminate the influence of inequable strengthes when the subject performs the same hand gesture at different times. Therefore, we adopt the following four features: MAV, MAVR, ZC and WL.

Autocorrelation and cross-correlation coefficients: Using ACCC as features for myoelectric control was first proposed by Leowinata et al. in 1998 [49], who suggested that useful information might reside in the crosstalk between channels. Therefore, the autocorrelation for each channel and the cross-correlation between channels are adopted as features.

Spectral power magnitudes: A feature set comprised of SPM has been proposed in several studies [50, 51] and was reported to provide good performances. SPM are calculated by taking the average of power spectrum within disjoint bandwidths after performing FFT for the data in active segments. In our case, for each channel we calculate SPM for four equal bandwidths between 75 Hz and 250 Hz.

Classification algorithms The classification algorithm is a key element of the recognition procedure. Its reliability, computational complexity and recognition efficiency have a great effect on the overall performance of the whole system. In recent years, many algorithms for sEMG recognition have been developed to specifically serve certain scenarios [57]. Among them, classical algorithms such as k-nearest neighbor (k-NN), LDA, quadratic discriminant analysis (QDA) and SVM have been most widely employed and reported to perform robustly in many studies [19, 58, 59, 61, 64]. In this chapter, associated with the three sets of features described above, we first evaluate these four popular classical classification approaches and study their effects on the recognition accuracy of Chinese number gestures.

k-nearest neighbor: The k-NN classification algorithm predicts the test sample's class according to its k nearest training samples by classifying the test sample to a class using majority vote among the k neighbors. The Euclidean distance in the feature space is used to determine which k training samples are the nearest neighbors [66]. In this chapter, since there are fifty samples for each hand gesture for each subject, we choose $k = 10$ as a proper setting.

Linear discriminant analysis: The LDA classifier is carried out by calculating linear discriminant functions and selecting the maximum one as the classification rule. LDA assumes that the feature vector x is multivariate normally distributed in each class group and the K classes have a common covariance matrix $\Sigma_k = \Sigma, \forall k$.

Quadratic discriminant analysis: QDA is closely related to, but different from LDA, due to different assumption of the covariance matrix. In QDA, the matrices Σ_k's of the K classes are not assumed to be identical as in LDA. The covariance matrix needs to be estimated separately for each class. In general, QDA and LDA are interchangeable, and which to use depends on the personal preference and the availability of the training data to support the QDA analysis. Both LDA and QDA perform well on a large and diverse set of classification tasks [66].

Support vector machine: SVM aims to find the optimal separating hyperplane between classes by focusing on the training samples that lie at the edge of the class distributions, named the support vectors, and with other training samples being effectively discarded. The basic idea of the SVM classifier is that only the training samples lying on the class boundaries are necessary for discrimination. Detailed discussions of SVM can be found in [66]. SVM was originally designed for binary classification as described above, but can be extended to multiclass classification. Several approaches have been suggested for multiclass classification by SVM [63], and here we adopt the one-against-all approach. In this approach, a set of binary classifiers, each of which is trained to separate one class from the rest, are undertaken and each test sample is allocated to the class for which the largest decision value is determined. For the kernel type, we select radial basis function (RBF) kernel.

We are interested in whether the performance of recognizing the 10 Chinese number gestures can be further improved. Based on our preliminary results shown in Fig. A.13, we can see that all three types of features contain useful, but probably compensating information and could yield good recognition performances, we therefore propose combining the three feature sets together. From Fig. A.13, we also note that SVM provides good performances consistently when using different features. In SVM, a kernel function k implicitly maps samples x to a feature space Φ given by a feature map $k(x_i,x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$. Since there are different kernel types such as RBF kernel, Polynomial kernel, Hyperbolic tangent kernel etc., it is often unclear what the most suitable kernel for the task at hand is, and hence the user may wish to combine several possible kernels. One problem with simply adding kernels is that using uniform weights is possibly not optimal. For instance, if one kernel is not correlated with the labels at all, then giving it positive weight will just add noise and degrade the performance [41]. It is thus desirable to use an optimized kernel function that could fit the data structure very well. Therefore, to efficiently combine multiple features and to optimally combine multiple kernels in SVM, we propose employing the following multiple kernel learning approach for Chinese number gesture recognition.

Multiple kernel learning: MKL is an efficient way of optimizing kernel weights and can design a kernel which is optimal for a given data set. In the MKL approach, now let us consider convex combinations of M kernels

$$k(x_i,x_j) = \sum_{m=1}^{M} w_m k_m(x_i,x_j) \tag{A.4}$$

where $w_m \geq 0$, $\sum_{m=1}^{M} w_m = 1$, and M is the total number of kernels used. In the multiple kernel learning problem, for binary classification, suppose we have a training set of N samples $\{x_i, y_i\}$, $i = 1,...,N, y_i \in \{1,-1\}$, where x_i is translated via M mappings $\Phi_m(x_i) \longmapsto \mathbb{R}^{D_m}$, $m = 1,...,M$, from the input into M feature spaces $(\Phi_1(x_i),...,\Phi_M(x_i))$, where D_m denotes the dimensionality of the m-th feature space. We need to solve the following MKL primal problem [65], which is equivalent to the linear SVM for M = 1:

$$\begin{aligned}
&\min && \frac{1}{2}\left(\sum_{m=1}^{M} w_m \|\beta_m\|_2\right)^2 + C\sum_{i=1}^{N}\xi_i \\
&\text{w.r.t.} && \beta_m \in \mathbb{R}^{D_m}, \xi \in \mathbb{R}^N, \beta_0 \in \mathbb{R} \\
&\text{s.t.} && \xi_i \geq 0 \quad \text{and} \quad y_i\left(\sum_{m=1}^{M} w_m \langle \beta_m, \Phi_m(x_i)\rangle + \beta_0\right) \geq 1 - \xi_i, \forall i = 1, \ldots, N
\end{aligned} \tag{A.5}$$

where β_m is the normal vector to the hyperplane for the feature space Φ_m. The authors in [65] derived the MKL dual for the problem (A.5) as below:

$$\begin{aligned}
&\min && \gamma \\
&\text{w.r.t.} && \gamma \in \mathbb{R}, \alpha \in \mathbb{R}^N \\
&\text{s.t.} && 0 \leq \alpha \leq 1C, \sum_{i=1}^{N} \alpha_i y_i = 0 \\
& && \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j k_m(x_i, x_j) - \sum_{i=1}^{N} \alpha_i \leq \gamma, \forall m = 1, \ldots, M.
\end{aligned} \tag{A.6}$$

where $\alpha = (\alpha_1, \alpha_2, ..., \alpha_N)$, α_i is a Lagrange multiplier and $\beta_m = \sum_{i=1}^{N} \alpha_i y_i \Phi_m(x_i)$.

In [68], the authors tackled the MKL problem through a weighted 2-norm regularization formulation with an additional constraint on the weights that encourages sparse kernel combinations. This algorithm, called SimpleMKL, can converge rapidly with comparable efficiency. In this chapter, we will adopt SimpleMKL as our MKL algorithm.

Results and Discussion

Off-line recognition results and discussion Three general feature types described in Section A.3.2 and four popular classification algorithms described in Section A.3.2, representing a total of twelve combinations, are first applied to our recognition problem of Chinese number gestures. Then based on their performances, we suggest to combine the three feature (3F) together and apply MKL-SVM method.

To make a fair comparison between single kernel and multiple kernels, we also employ the single kernel SVM combined with 3F for classification. In total, 14 approaches were investigated in this study, as shown in Fig. A.13.

There were six subjects in total and each subject performed 50 sEMG experimental trials for each of the 10 designed Chinese number gestures. Therefore, there were totally 300 trials for each movement in the defined number gesture movement set. For each subject, the leave-one-out method, a commonly used method of cross-validation, was used to calculate the classification accuracy and form a subject-level confusion matrix. Adding the six subject-level confusion matrices together produces an overall confusion matrix for a given approach. The resulting classification results from different approaches are summarized by the overall confusion matrices, some of which are presented in Fig. A.12 and the overall classification accuracies are reported in Fig. A.13.

From the confusion matrices, we can see that in general the recognition accuracies for individual number gestures and the overall average accuracies are quite satisfying. By studying the existing literature, we also note that the classification results in our multi-finger number gesture case is comparable to, or even outperform, the ones in gross hand, wrist and arm movement cases when employing similar recognition approaches. For instance, in [19], the difference absolute mean value was used to construct a feature map and the k-NN, LDA and QDA algorithms were used to classify five wrist-motion direction movements, including up, down, right, left, and the rest state. The reported recognition accuracy rates were 84.9% for k-NN, 82.4% for QDA, and 81.1% for LDA, while our resulting accuracy rates were 96.17% for k-NN, 95.90% for LDA, and 95.67% for QDA. Although only two forearm sEMG channels were used in [19] while we used four, it is worth noting that our problem is to recognize twice the number of movements and that our defined subtle number gestures belong to the category of subtle multi-finger movements, which is much more difficult to be classified than the category of gross hand, wrist and arm movements. Using eight sEMG channels, the authors in [61] applied discrete wavelet transform and SVM to classify six gross hand movements such as wrist flexion, wrist extension, hand supination, hand pronation, hand opening and hand closure, and their misclassification rate for six subjects was 4.7% $\pm$ 3.7% while ours was 3.37% $\pm$ 1.19% by using TD and SVM.

Feature Extraction and Classification Methods: ACCC + *k*-NN

	1	2	3	4	5	6	7	8	9	0	Accuracy (%)
1	161	45	11	14	2	10	18	3	33	3	53.67
2	45	184	26	5	3	0	15	4	15	3	61.33
3	4	24	214	33	9	0	9	5	2	0	71.33
4	15	30	53	154	20	4	15	7	1	1	51.33
5	8	7	23	24	228	1	6	3	0	0	76.00
6	3	1	1	3	1	276	3	10	1	1	92.00
7	18	24	19	24	7	5	182	7	4	10	60.67
8	5	5	5	9	18	21	5	224	6	2	74.67
9	45	21	15	14	8	2	22	19	145	9	48.33
0	20	18	4	5	2	10	24	12	32	173	57.67

Overall Accuracy (Mean ± SD): (64.70 ± 13.59) %

(a)

Feature Extraction and Classification Methods: TD + LDA

	1	2	3	4	5	6	7	8	9	0	Accuracy (%)
1	291	1	0	0	0	2	0	0	5	1	97.00
2	2	291	1	2	0	0	4	0	0	0	97.00
3	2	1	287	7	0	0	1	0	2	0	95.67
4	0	1	15	283	0	0	1	0	0	0	94.33
5	0	0	1	0	297	2	0	0	0	0	99.00
6	0	6	0	0	3	288	1	0	1	1	96.00
7	2	5	3	0	0	0	290	0	0	0	96.67
8	1	4	0	0	0	6	0	287	1	1	95.67
9	6	0	0	2	0	0	1	2	285	4	95.00
0	0	1	0	1	0	9	3	0	8	278	92.67

Overall Accuracy (Mean ± SD): (95.90 ± 1.71) %

(b)

Feature Extraction and Classification Methods: SPM + SVM (Signal Kernel)

	1	2	3	4	5	6	7	8	9	0	Accuracy (%)
1	290	0	1	0	0	0	2	1	6	0	96.67
2	0	294	2	3	0	1	0	0	0	0	98.00
3	1	5	280	9	0	1	2	1	1	0	93.33
4	0	3	11	280	3	1	0	0	1	1	93.33
5	0	0	0	4	295	1	0	0	0	0	98.33
6	4	0	1	2	0	286	0	3	1	3	95.33
7	1	1	3	0	0	1	291	0	2	1	97.00
8	0	1	1	0	0	5	0	285	6	2	95.00
9	14	0	0	0	1	0	1	2	274	8	91.33
0	1	1	0	0	3	1	4	2	12	276	92.00

Overall Accuracy (Mean ± SD): (95.03 ± 2.47) %

(c)

Feature Extraction and Classification Methods: 3F + SVM (Multiple Kernels)

	1	2	3	4	5	6	7	8	9	0	Accuracy (%)
1	295	0	0	0	0	0	0	1	2	2	98.33
2	1	299	0	0	0	0	0	0	0	0	99.67
3	1	0	288	8	0	0	1	0	2	0	96.00
4	0	2	4	294	0	0	0	0	0	0	98.00
5	0	0	0	1	299	0	0	0	0	0	99.67
6	1	0	0	0	0	293	0	2	1	3	97.67
7	2	1	0	0	0	1	296	0	0	0	98.67
8	0	0	0	0	1	4	0	295	0	0	98.33
9	1	1	0	1	0	2	1	1	289	4	96.33
0	1	0	0	0	0	3	1	2	3	290	96.67

Overall Accuracy (Mean ± SD): (97.93 ± 1.28) %

(d)

Figure A.12: Some confusion matrices of different combinations. Here the notation "ACCC+*k*-NN" is used to represent the recognition approach that uses the ACCC features and the *k*-NN classifier, and other notations are similarly defined. (a) ACCC+*k*-NN. (b) TD+LDA. (c) SPM+SVM (SK), here SK means single kernel. (d) 3F+SVM (MKL).

However, we also note some limitations in the recognition results reported in Fig. A.12. Several pairs of number gestures are relatively more difficult to be distinguished, such as the gestures representing the numbers 3 and 4, and the gestures representing the numbers 0 and 9. This is probably due to the high movement similarity in such pairs, as seen in Fig. A.1. From Fig. A.12 and Fig. A.13, we note that the recognition approach using ACCC features and the *k*-NN classifier, which is notated as "ACCC + *k*-NN", performs very different from the other recognition approaches and yields the worst recognition results, as presented in Fig. A.12a. It seems that each number gesture can be misclassified into other classes, even if the two gestures (e.g. the numbers 0 and 3) do not seem similar and not been misclassified by any of the rest recognition approaches. We think this observation might reflect *k*-NN's possible disadvantages associated with the Euclidean distance mea-

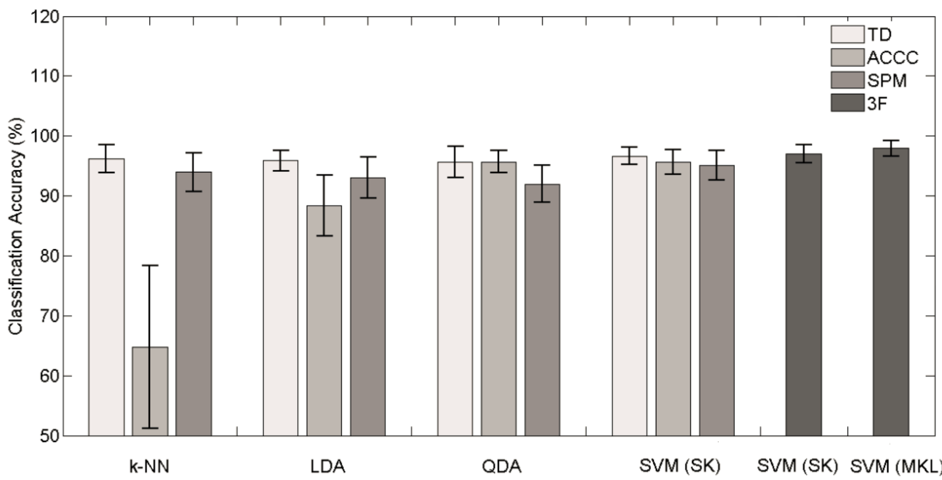


Figure A.13: Comparisons of the overall recognition accuracies when employing different feature sets and different classifiers.

sure. When two class groups are very close to each other, the classifier based on the Euclidean distance could misclassify the points residing around the boundary. It is worse if the two class groups overlap. When one class group has several other class groups surrounded with, the points lying around the boundary in the class can be randomly classified into the surrounding classes by *k*-NN and only the ones in the center can be classified accurately. This can explain our results observed in Fig. A.12a.

From Fig. A.13, we also note that the ACCC + LDA approach yields the second worst accuracy 88.40%, which suggests that ACCC might not be a good feature set for the LDA classifier. We note that the combinations of ACCC and QDA or SVM provide quite high accuracies, both above 95%. In general, it seems that *k*-NN and LDA may not be as robust as QDA and SVM, probably due to the explanations above for *k*-NN and the common covariance assumption for LDA. Compared with the ACCC feature set, SPM and TD can provide more stable performances no matter which classifier is employed, and it is noted that all combinations involving them achieve the accuracy rates above 91%. SPM generally yields a slightly worse recognition performance than TD does. Therefore, the TD features is slightly preferred than other features. Regarding the studied popular classifiers,

Table A.1: Some combinations of different kernels and corresponding results.

Index	Kernel Type	Kernel Number	Kernel Parameter*	Accuracy
1	RBF	10	[0.5 1 2 5 8 10 12 15 17 20]	97.47%
2	RBF	4	[0.5 1 2 5]	97.73%
3	RBF	5	[0.5 1 2 5 8]	97.77%
4	RBF, Poly	7	[0.5 1 2 5 8], [1 2]	97.80%
5	RBF, Poly	8	[0.5 1 2 5 8], [1 2 3]	97.83%
6	RBF, Poly	6	[1 2 5 8], [1 2]	97.87%
7	RBF, Poly	7	[1 2 5 8], [1 2 3]	97.93%

* Here Kernel Parameters include the Gaussian width of RBF kernel and the freedom degree of Polynomial kernel, e.g. "[1 2 5 8], [1 2]" indicates that four RBF kernels with Gaussian width 1, 2, 5, 8 and two Poly kernels with freedom degree 1, 2 are used.

QDA and SVM are both reliable and well-performed. The approaches combining QDA (and SVM) with TD achieved 95.67% and 96.63% average accuracies with small standard errors respectively.

Since all the three types of feature sets contain useful information based on their performance, we further investigate to combine the three features together. We first use single kernel SVM with 3F and achieve the accuracy 97.03% which is higher than that of the above 12 combinations. To fit the combined feature set better and further improve the recognition performance, we apply MKL-SVM method described in Section A.3.2. Regarding the kernel type selection, RBF is generally suggested [62] and we also examine the polynomial kernel function. Regarding the parameter selection, a parameter range is first set for each kernel function: RBF (min:0.5;max:20) and polynomial (min:1;max:5); then the number of kernels and the corresponding parameters (all are integers except 0.5 for RBF) are automatically selected uniformly according to a uniform distribution by MATLAB which is repeated 100 times; finally, the combinations with the best performances are selected. We provide some kernel parameters and report their performances in Table A.1 for reference. From Table A.1, we can see that using polynomial kernel function can help improve the performance and a highest rate 97.93% can be achieved by combination-7. In addition, the results also indicate that though SVM with single kernel provides excellent performance, it does not perform as well as MKL-SVM does. In particular, the combination TD + LDA was reported multiple times to provide the most stable performance for myoelectric control [59, 60]. To validate the significance of the results achieved by 3F + MKL-SVM upon that by TD + LDA, we perform a paired samples t-test to compare the recognition rates by the two combinations. The test indicates that there is a significant difference in the accuracy ($t = 5.303$, $p < 0.01$) and the proposed combination outperforms TD + LDA for classification rate. However, the efficiency of MKL-SVM is much lower than that of LDA since the former involves a slow optimization procedure. Overall, MKL-SVM is a promising method for sEMG-based number gesture pattern recognition.

On-line recognition results and discussion Based on high classification accuracies observed in the offline sEMG recognition analysis in Section A.3.2, we implement a real-time classification system for Chinese number gestures. The purpose of this section is to examine the possibility of practical applications in a controlled

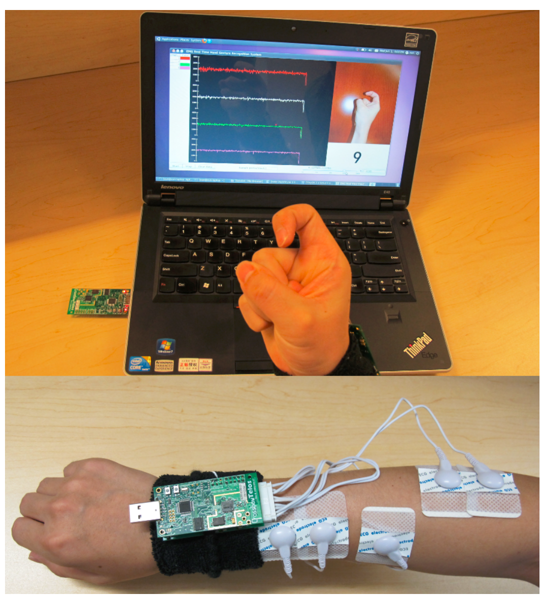


Figure A.14: Pictures of the hardware and software implementation of the proposed real-time sEMG recognition system.

laboratory setting, that is, to check if the accuracy can be improved to a satisfying level when subjects become familiar with the system and if the re-placement of electrodes at the same locations at different time can affect the classification accuracy. From the results in Section A.3.2 and considering the tradeoff between accuracy and algorithm complexity, we select QDA combined with TD as the algorithm for the real-time recognition system of Chinese number gestures. Fig. A.14 presents a demo of the implemented real-time system, in which the recognition algorithm part and graphical user interface are realized by the programming language Java.

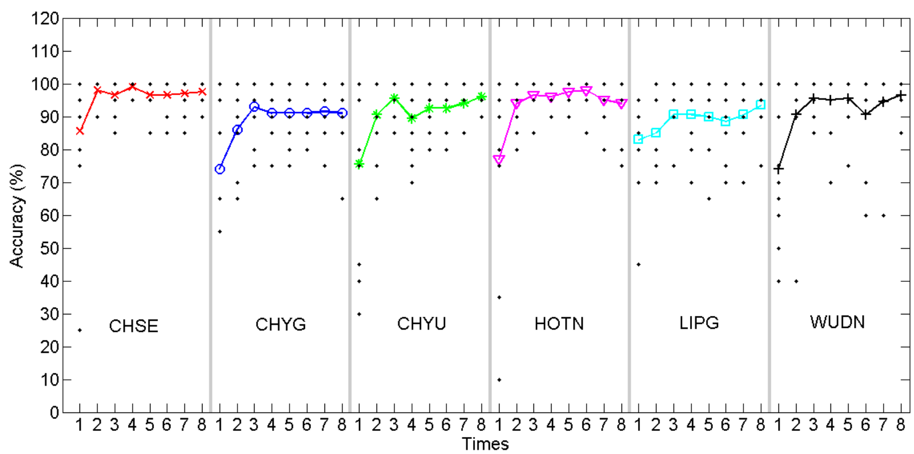


Figure A.15: The average recognition accuracy rates of the six subjects over the eight training sessions. The spread of data points are also provided. Note that some points are overlapped.

The corresponding real-time experiment was described in the Section A.3.2. The classification results for each subject were shown in Fig. A.15 and the results for each of the designed number gesture movements were shown in Fig. A.16. From Fig. A.15, we note that the average accuracy rates for the six subjects generally increase gradually as the number of training sessions increases. After four training sessions, the accuracy rate for each subject remains nearly steady. All six subjects could achieve above 90% accuracy in average and three of them could achieve above 95% accuracy for the ten number gestures. From Fig. A.16, similar observations can be noted for each number gesture movement. The average recog-

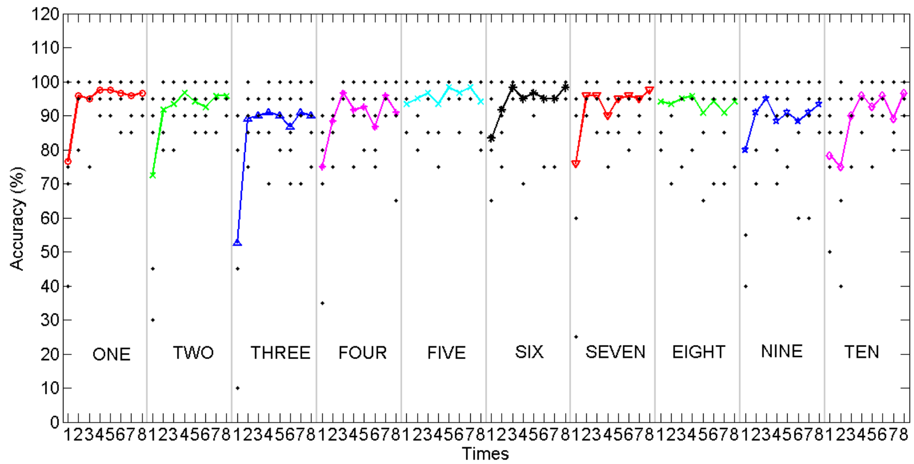


Figure A.16: The average recognition accuracy rates for the ten number gesture movements over the eight training sessions. The spread of data points are also provided. Note that some points are overlapped.

nition accuracy rates for each number gesture movement were above 90% after four training sessions and above 95% accuracy can be achieved for seven of the ten movements.

The reason for our excellent recognition performance should be the following factors: reasonable electrode placement, effective feature extraction methods and advanced classification algorithms. Reasonable electrode placement ensures that meaningful sEMG signals from corresponding muscles are acquired; effective extraction methods ensure that useful information contained in the signals are extracted appropriately; advanced classification algorithms guarantee that different patterns can be distinguished from each other. In this , through a large number of preliminary experiments and our previous study [46], we can design the superior electrode placement for our proposed number gestures using only four sEMG channels, while in most previous studies (e.g. [22]) a large number of channels were employed and arranged orderly in which case only a part of them were actually useful. Also, we extensively investigate the most popular feature extraction and recognition methods so that we can clearly tell the most proper approaches for number gestures and select the most effective features and classification algorithms. We believe that our super recognition performance of recognizing ten multi-finger gestures is jointly contributed by the above factors. With that excellent performance, we believe the number gestures are therefore promising for practical applications such as HCI.

Appendix B

Algorithm Derivation

B.1 The Derivation for The PLS+CCA Method

In this appendix, we show how to mathematically derive the solution of the proposed PLS+CCA method.

B.1.1 The First Step: PLS

The cost function of PLS is as follows (the same as Equation (2.1) in Section 2.3.2):

$$\begin{aligned} &\max_{w_1,w_2} \quad \left({w_1}^T X^T Y w_2\right)^2 \\ &\text{s.t.} \quad {w_i}^T w_i = 1, \quad i = 1,2 \end{aligned} \tag{B.1}$$

where w_i's $(i = 1,2)$ are the weight vectors.

By employing the method of Lagrange multipliers, we rewrite the initial cost function as:

$$\mathrm{L}(w_i,\lambda_i) = \left({w_1}^T X^T Y w_2\right)^2 - \sum_{i=1}^{2} \lambda_i({w_i}^T w_i - 1), \tag{B.2}$$

where λ_i's are Lagrange multipliers.

Now we only present the detailed derivations regarding w_1, since w_2 can be similarly derived. Taking the derivatives of $\mathrm{L}(w_i,\lambda_i)$ with respect to w_1 and λ_1 and setting them to be zero, we have:

$$\nabla \mathrm{L}_{w_1} = 2\left|{w_1}^T X^T Y w_2\right| X^T Y w_2 - 2\lambda_1 w_1 = 0, \tag{B.3}$$

$$\nabla \mathrm{L}_{\lambda_1} = {w_1}^T w_1 - 1 = 0. \tag{B.4}$$

Left multiplying both sides of Eq. (B.3) by ${w_1}^T$, we have:

$$2\left({w_1}^T X^T Y w_2\right)^2 - 2\lambda_1 {w_1}^T w_1 = 0. \tag{B.5}$$

According to Eq. (B.4), λ_1 can be calculated as

$$\lambda_1 = \left({w_1}^T X^T Y w_2\right)^2. \tag{B.6}$$

Through the similar procedure, $\nabla \mathrm{L}_{w_2}$ and λ_2 can be easily derived as

$$\nabla \mathrm{L}_{w_2} = 2\left|{w_1}^T X^T Y w_2\right| Y^T X w_1 - 2\lambda_2 w_2 = 0, \tag{B.7}$$

$$\lambda_2 = \left({w_1}^T X^T Y w_2\right)^2. \tag{B.8}$$

Substituting Eq. (B.8) into Eq. (B.3) and Eq. (B.7) respectively, we have the following two expressions:

$$\sqrt{\lambda_2} X^T Y w_2 = \lambda_1 w_1, \tag{B.9}$$

$$\frac{1}{\sqrt{\lambda_2}} Y^T X w_1 = w_2. \tag{B.10}$$

By substituting Eq. (B.10) into Eq. (B.9), we can formulate an eigenvalue-eigenvector decomposition problem:

$$\left(X^T Y Y^T X\right) w_1 = \lambda_1 w_1. \tag{B.11}$$

Similarly, we can have the formulation for w_2 as:

$$\left(Y^T X X^T Y\right) w_2 = \lambda_2 w_2. \tag{B.12}$$

The above solutions are straightforward. A practical issue is to determine the number of LVs. In our study, we determine the number R by setting a threshold that corresponds to the ratio of explained variance (e.g. 95%). Therefore, the corresponding LVs in X and Y can be calculated by

$$T_X = XW_1, \quad T_Y = YW_2, \tag{B.13}$$

where W_1 is composed of the first R eigenvectors associated with Eq. (B.11) and the columns of T_X represent the R components extracted from X. W_2 and T_Y are similarly defined.

However, the collinearity problem may exist in the LVs calculated through the above procedure, since each data set is used repetitively for each LV's calculation. The extracted LVs are not necessarily uncorrelated to each other. To effectively implement the second step and avoid the ill-conditioned problem, we need to address this uncorrelatedness concern and thus we design a deflation procedure: Before extracting the second common LV in each data space, X and Y are deflated by their corresponding first LVs as follows:

$$X = X - t_X({t_X}^T t_X)^{-1}{t_X}^T X, \quad Y = Y - t_Y({t_Y}^T t_Y)^{-1}{t_Y}^T Y. \tag{B.14}$$

Then the above procedure will be repeated for the further extraction of common LVs. In this way, the following new LVs are uncorrelated to the previous ones.

The purpose of this step is to extract LVs which can most explain the individual data sets and meanwhile are well correlated to the LVs in another data set. With this step, trivial and irrelevant information across data sets could be removed. However, a higher covariance may merely result from the larger variance of LVs, which may not necessarily imply strong correlations. To address this concern, the 2nd step will help further refine the results.

B.1.2 The Second Step: CCA

Based on the extracted LVs in the first step, the objective function of CCA can be constructed as follows:

$$\begin{aligned} \max_{v_1, v_2} \quad & \left({v_1}^T {T_X}^T T_Y v_2 \right)^2 \\ \text{s.t.} \quad & {v_1}^T {T_X}^T T_X v_1 = 1, \quad {v_2}^T {T_Y}^T T_Y v_2 = 1 \end{aligned} \tag{B.15}$$

where v_i's $(i = 1, 2)$ are the weight vectors.

By employing the method of Lagrange multipliers, we rewrite the initial objective function as:

$$\mathrm{L}(v_i, \eta_i) = \left({v_1}^T {T_X}^T T_Y v_2 \right)^2 - \eta_1 ({v_1}^T {T_X}^T T_X v_1 - 1) - \eta_2 ({v_2}^T {T_Y}^T T_Y v_2 - 1), \tag{B.16}$$

where η_i's are Lagrange multipliers. Similar to the derivation in the first step, we can obtain the following eigenvalue-eigenvector decomposition problem:

$$\left[({T_X}^T T_X)^{-1} {T_X}^T T_Y ({T_Y}^T T_Y)^{-1} {T_Y}^T T_X \right] v_1 = \eta_1 v_1. \tag{B.17}$$

Similarly, for v_2, we have:

$$\left[({T_Y}^T T_Y)^{-1} {T_Y}^T T_X ({T_X}^T T_X)^{-1} {T_X}^T T_Y \right] v_2 = \eta_2 v_2. \tag{B.18}$$

The solutions to this problem are the R largest eigenvectors of the corresponding matrices. The recovered LVs U_X and U_Y can be calculated directly from the matrices T_X and T_Y by

$$U_X = T_X V_1, \quad U_Y = T_Y V_2, \tag{B.19}$$

where V_1 is composed of the R eigenvectors associated with Eq. (B.17) and the columns of U_X represent the R components extracted from T_X. V_2 and U_Y are similarly defined.

After these two steps, it is ensured that the extracted components U_X and U_Y are maximally correlated across data sets and meanwhile can well explain the information within each individual data set.

B.2 The Derivation for The IC-PLS Model

In this appendix, we show how to mathematically derive the optimization solution of the proposed IC-PLS model.

By employing the method of Lagrange multipliers, we can rewrite the initial cost function as follows:

$$\begin{aligned}\mathrm{F}(w_1,w_2,\lambda_1,\lambda_2) = \alpha \left(\mathrm{E}((w_1{}^T x)(w_2{}^T y))\right)^2 \\ + \beta \left(\mathrm{E}(\mathrm{G}(w_1{}^T x)) - \mathrm{E}(\mathrm{G}(u_1))\right)^2 \\ + \theta \left(\mathrm{E}(\mathrm{G}(w_2{}^T y)) - \mathrm{E}(\mathrm{G}(u_2))\right)^2 \\ + \lambda_1(w_1{}^T w_1 - 1) + \lambda_2(w_2{}^T w_2 - 1)\end{aligned} \tag{B.20}$$

where λ_1 and λ_2 are Lagrange multipliers.

Here, we only present the detailed derivations regarding w_1. As to those of w_2, it is straightforward from the results obtained on w_1. Taking the derivatives of $\mathrm{F}(w_1,w_2,\lambda_1,\lambda_2)$ with respect to w_1 and λ_1 and setting them to be zero, we have:

$$\begin{aligned}\nabla \mathrm{F}_{w_1} = \frac{\partial \mathrm{F}}{\partial w_1} = 2\alpha \mathrm{E}((w_1{}^T x)(w_2{}^T y))\mathrm{E}(x(w_2{}^T y)) \\ + 2\beta(\mathrm{E}(\mathrm{G}(w_1{}^T x)) - \mathrm{E}(\mathrm{G}(u_1)))\mathrm{E}(x\mathrm{g}(w_1{}^T x)) \\ + 2\lambda_1 w_1 = 0,\end{aligned} \tag{B.21}$$

$$\nabla \mathrm{F}_{\lambda_1} = w_1{}^T w_1 - 1 = 0, \tag{B.22}$$

where $\mathrm{g}(\cdot)$ represents the corresponding first-order derivative of $\mathrm{G}(\cdot)$.

Left multiplying both sides of Eq. (B.3) by $w_1{}^T$, we have:

$$\begin{aligned}2\alpha w_1{}^T \mathrm{E}((w_1{}^T x)(w_2{}^T y))\mathrm{E}(x(w_2{}^T y)) \\ + 2\beta(\mathrm{E}(\mathrm{G}(w_1{}^T x)) - \mathrm{E}(\mathrm{G}(u_1)))w_1{}^T \mathrm{E}(x\mathrm{g}(w_1{}^T x)) \\ + 2\lambda_1 w_1{}^T w_1 = 0.\end{aligned} \tag{B.23}$$

According to Eq. (B.4), λ_1 can be calculated as

$$\begin{aligned}\lambda_1 = -\alpha w_1{}^T \mathrm{E}((w_1{}^T x)(w_2{}^T y))\mathrm{E}(x(w_2{}^T y)) \\ - \beta(\mathrm{E}(\mathrm{G}(w_1{}^T x)) - \mathrm{E}(\mathrm{G}(u_1)))w_1{}^T \mathrm{E}(x\mathrm{g}(w_1{}^T x)).\end{aligned} \tag{B.24}$$

Based on Kuhn-Tucker conditions [99], the optima of the multi-objective function shown in Eq. (3.5) will be at the points shown in Eq. (B.21) under the constraint (B.22). In this work, to improve the convergence speed, Newton's method is employed to solve this problem.

Suppose $\Phi_1 = \nabla \mathrm{F}_{w_1}$ and then its Jacobian matrix can be derived as

$$
\begin{aligned}
\mathrm{J}\Phi_1 = {} & 2\alpha \mathrm{E}(x({w_2}^T y))\mathrm{E}(x^T({w_2}^T y)) \\
& + 2\beta(\mathrm{E}(\mathrm{G}({w_1}^T x)) - \mathrm{E}(\mathrm{G}(u_1)))\mathrm{E}(xx^T \mathrm{g}'({w_1}^T x)) \\
& + 2\beta \mathrm{E}(x\mathrm{g}({w_1}^T x))\mathrm{E}(x^T \mathrm{g}({w_1}^T x)) + 2\lambda_1 \mathrm{I},
\end{aligned}
\tag{B.25}
$$

where $\mathrm{g}'(.)$ means the second-order derivative of G(.).

Substituting Eq. (B.24) into Eq. (B.25), we have the following expression for $\mathrm{J}\Phi_1$:

$$
\begin{aligned}
\mathrm{J}\Phi_1 = {} & 2\alpha \mathrm{E}(x({w_2}^T y))\mathrm{E}(x^T({w_2}^T y)) \\
& + 2\beta(\mathrm{E}(\mathrm{G}({w_1}^T x)) - \mathrm{E}(\mathrm{G}(u_1)))\mathrm{E}(xx^T \mathrm{g}'({w_1}^T x)) \\
& + 2\beta \mathrm{E}(x\mathrm{g}({w_1}^T x))\mathrm{E}(x^T \mathrm{g}({w_1}^T x)) \\
& - (2\alpha {w_1}^T \mathrm{E}(({w_1}^T x)({w_2}^T y))\mathrm{E}(x({w_2}^T y)) \\
& + 2\beta(\mathrm{E}(\mathrm{G}({w_1}^T x)) - \mathrm{E}(\mathrm{G}(u_1))){w_1}^T \mathrm{E}(x\mathrm{g}({w_1}^T x)))\mathrm{I}.
\end{aligned}
\tag{B.26}
$$

Since the data have been initialized, to simplify the inverse of the Jacobian matrix, the second term can be approximated as in the original FastICA algorithm as [39]:

$$
\mathrm{E}(xx^T \mathrm{g}'({w_1}^T x)) \approx \mathrm{E}(xx^T)\mathrm{E}(\mathrm{g}'({w_1}^T x)) = \mathrm{E}(\mathrm{g}'({w_1}^T x))\mathrm{I}. \tag{B.27}
$$

Therefore, the Jacobian matrix can be approximately expressed in a simple form:

$$
\begin{aligned}
\mathrm{J}\Phi_1 = {} & 2\alpha \mathrm{E}(x({w_2}^T y))\mathrm{E}(x^T({w_2}^T y)) \\
& + 2\beta \mathrm{E}(x\mathrm{g}({w_1}^T x))\mathrm{E}(x^T \mathrm{g}({w_1}^T x)) - c_1 \mathrm{I},
\end{aligned}
\tag{B.28}
$$

where c_1 is a constant defined as

$$
\begin{aligned}
c_1 = {} & 2\alpha {w_1}^T \mathrm{E}(({w_1}^T x)({w_2}^T y))\mathrm{E}(x({w_2}^T y)) \\
& + 2\beta(\mathrm{E}(\mathrm{G}({w_1}^T x)) - \mathrm{E}(\mathrm{G}(u_1))){w_1}^T \mathrm{E}(x\mathrm{g}({w_1}^T x)) \\
& - 2\beta(\mathrm{E}(\mathrm{G}({w_1}^T x)) - \mathrm{E}(\mathrm{G}(u_1)))\mathrm{E}(\mathrm{g}'({w_1}^T x)).
\end{aligned}
\tag{B.29}
$$

Substituting Eq. (B.24) into Eq. (B.3), we can express the derivatives of $\mathrm{F}(w_1, w_2, \lambda_1, \lambda_2)$ with respect to w_1 as:

$$
\begin{aligned}
\nabla \mathrm{F}_{w_1} = {} & 2\alpha \mathrm{E}(({w_1}^T x)({w_2}^T y))\mathrm{E}(x({w_2}^T y)) \\
& + 2\beta(\mathrm{E}(\mathrm{G}({w_1}^T x)) - \mathrm{E}(\mathrm{G}(u_1)))\mathrm{E}(x\mathrm{g}({w_1}^T x)) \\
& - 2(\alpha {w_1}^T \mathrm{E}(({w_1}^T x)({w_2}^T y))\mathrm{E}(x({w_2}^T y)) \\
& + \beta(\mathrm{E}(\mathrm{G}({w_1}^T x)) - \mathrm{E}(\mathrm{G}(u_1))){w_1}^T \mathrm{E}(x\mathrm{g}({w_1}^T x)))w_1,
\end{aligned}
\tag{B.30}
$$

Through the similar procedure, $\mathrm{J}\Phi_2$ and $\nabla \mathrm{F}_{w_2}$ can be easily derived as

$$
\begin{aligned}
\mathrm{J}\Phi_2 = {} & 2\alpha \mathrm{E}(y({w_1}^T x))\mathrm{E}(y^T({w_1}^T x)) \\
& + 2\theta \mathrm{E}(y\mathrm{g}({w_2}^T y))\mathrm{E}(y^T \mathrm{g}({w_2}^T y)) - c_2 \mathrm{I},
\end{aligned}
\tag{B.31}
$$

where c_2 is a constant defined as

$$
\begin{aligned}
c_2 = {} & 2\alpha {w_2}^T \mathrm{E}(({w_1}^T x)({w_2}^T y))\mathrm{E}(y({w_1}^T x)) \\
& + 2\theta(\mathrm{E}(\mathrm{G}({w_2}^T y)) - \mathrm{E}(\mathrm{G}(u_2))){w_2}^T \mathrm{E}(y\mathrm{g}({w_2}^T y)) \\
& - 2\theta(\mathrm{E}(\mathrm{G}({w_2}^T y)) - \mathrm{E}(\mathrm{G}(u_2)))\mathrm{E}(\mathrm{g}'({w_2}^T y)).
\end{aligned}
\tag{B.32}
$$

and

$$
\begin{aligned}
\nabla \mathrm{F}_{w_2} = {} & 2\alpha \mathrm{E}(({w_1}^T x)({w_2}^T y))\mathrm{E}(y({w_1}^T x)) \\
& + 2\theta(\mathrm{E}(\mathrm{G}({w_2}^T y)) - \mathrm{E}(\mathrm{G}(u_2)))\mathrm{E}(y\mathrm{g}({w_2}^T y)) \\
& - 2(\alpha {w_2}^T \mathrm{E}(({w_1}^T x)({w_2}^T y))\mathrm{E}(y({w_1}^T x)) \\
& + \theta(\mathrm{E}(\mathrm{G}({w_2}^T y)) - \mathrm{E}(\mathrm{G}(u_2))){w_2}^T \mathrm{E}(y\mathrm{g}({w_2}^T y)))w_2.
\end{aligned}
\tag{B.33}
$$

The Newton iteration direction vectors d_i $(i = 1, 2)$ are derived by solving the following equations:

$$
\mathrm{J}\Phi(w_i) \cdot d_i = -\nabla \mathrm{F}_{w_i}, \quad i = 1, 2. \tag{B.34}
$$

Therefore, finally, w_i $(i = 1, 2)$ can be derived as follows

$$
\begin{aligned}
& w_i \leftarrow w_i + d_i, \\
& \text{i.e.,} \quad w_i \leftarrow w_i - (\mathrm{J}\Phi(w_i))^{-1} \nabla \mathrm{F}_{w_i}, \quad i = 1, 2.
\end{aligned}
\tag{B.35}
$$