

## Research Article

# Corticomuscular Activity Modeling by Combining Partial Least Squares and Canonical Correlation Analysis

Xun Chen,<sup>1,2</sup> Aiping Liu,<sup>1</sup> Z. Jane Wang,<sup>1</sup> and Hu Peng<sup>2</sup>

<sup>1</sup> Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

<sup>2</sup> Department of Biomedical Engineering, School of Medical Engineering, Hefei University of Technology, Hefei, Anhui 230009, China

Correspondence should be addressed to Hu Peng; hpeng@hfut.edu.cn

Received 8 March 2013; Accepted 9 May 2013

Academic Editor: Chang-Hwan Im

Copyright © 2013 Xun Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Corticomuscular activity modeling based on multiple data sets such as electroencephalography (EEG) and electromyography (EMG) signals provides a useful tool for understanding human motor control systems. In this paper, we propose modeling corticomuscular activity by combining partial least squares (PLS) and canonical correlation analysis (CCA). The proposed method takes advantage of both PLS and CCA to ensure that the extracted components are maximally correlated across two data sets and meanwhile can well explain the information within each data set. This complementary combination generalizes the statistical assumptions beyond both PLS and CCA methods. Simulations were performed to illustrate the performance of the proposed method. We also applied the proposed method to concurrent EEG and EMG data collected in a Parkinson's disease (PD) study. The results reveal several highly correlated temporal patterns between EEG and EMG signals and indicate meaningful corresponding spatial activation patterns. In PD subjects, enhanced connections between occipital region and other regions are noted, which is consistent with previous medical knowledge. The proposed framework is a promising technique for performing multisubject and bimodal data analysis.

## 1. Introduction

Corticomuscular activity modeling is important for assessing functional interactions in the motor control system, that is, studying simultaneous cortical and muscular activities during a sustained isometric muscle contraction. Traditionally, the most common method to assess the interactions between motor-related brain areas and the muscles is magnitude-squared coherence (MSC), which is a normalized measure of correlation between two waveforms or signals in the frequency domain. For instance, in monkeys, coherent oscillations in the 20–30 Hz band could be detected between cortical local field potentials and the rectified electromyography (EMG) from contralateral hand muscles that were modulated during different phases of a precision grip task [1]. In humans, similar findings were discovered in the beta-band corticomuscular coherence between magnetoencephalography (MEG) [2] and electroencephalography (EEG) [3] from the primary motor cortex during isometric contractions.

Although MSC has been popular in studying corticomuscular coupling, it suffers from several limitations. First, addressing the intersubject 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 [4]. In fact, recent work has suggested that *interactions between* brain regions correspond more closely to ongoing EMG than activity at discrete sites [5–7]. 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 [8]. 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, partial least squares (PLS), and canonical correlation analysis (CCA) [9]. Among these methods, the latent-variable- (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 to them [10]. An advantage of PLS is that PLS 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 functional magnetic resonance imaging (fMRI) data associated with behavioural measures [11] and the common temporal components between EEG and fMRI signals [12]. 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 [5].

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 [13]. 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 [14], Clercq et al. successfully removed muscle artifacts from a real ictal EEG recording without altering the recorded underlying ictal activity. In [15], 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 [16].

Although PLS and CCA have been investigated in many medical applications, 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, both PLS and CCA have their advantages and disadvantages, but perhaps more importantly, these two methods can be considered complementary. In this paper, 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 with 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.

We will evaluate the performance of the proposed method on both synthetic data and real-world data. We first illustrate its performance using simulations, and we then apply the method to concurrent EEG and EMG data collected from patients with Parkinson's disease (PD) and age-matched normal subjects when they perform a dynamic, visually guided tracking task. We note highly correlated temporal patterns between EEG and EMG signals and meaningful spatial activation patterns. While the proposed method is intentionally proposed for corticomuscular coupling analysis, it can also be applied to analyze other types of concurrent signals, including, but not limited to, fMRI, photoplethysmograph (PPG), electrocardiography (ECG), and kinematic data.

## 2. Materials and Methods

*2.1. Methods.* In this section, 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-mean data sets are stored in two matrices, the predictor matrix  $\mathbf{X}(N \times p)$  and the response matrix  $\mathbf{Y}(N \times q)$ , where  $N$  means the number of observations and  $p$  and  $q$  indicate the numbers of variables in corresponding matrices.

*2.1.1. 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 to them [13]. In other words, the covariance between the extracted LVs should be maximized as

$$\begin{aligned} \max_{\mathbf{w}_1, \mathbf{w}_2} \quad & (\mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2)^2, \\ \text{s.t.} \quad & \mathbf{w}_i^T \mathbf{w}_i = 1, \quad i = 1, 2, \end{aligned} \quad (1)$$

where  $\mathbf{w}_i$ 's ( $i = 1, 2$ ) are the weight vectors. A typical PLS can be implemented by the classical NIPALS algorithm [10]. Also, an alternative calculation way is to perform eigenvalue-eigenvector decomposition [17]. Therefore, the maximum of (1) is achieved by having  $\mathbf{w}_1$  and  $\mathbf{w}_2$  as the largest eigenvectors of the matrices  $\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X}$  and  $\mathbf{Y}^T \mathbf{X} \mathbf{X}^T \mathbf{Y}$ , respectively. To obtain subsequent weights, the algorithm is repeated with deflated  $\mathbf{X}$  and  $\mathbf{Y}$  matrices. The detailed calculation procedure can be found in the appendix.

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  $\mathbf{X}$ , not all of them are generally used. The main reasons for this are the following: the measured data are never noise-free and some small components only describe noise, and it is common to ignore small components because of the

problems of collinearity. 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 [10].

*2.1.2. Canonical Correlation Analysis.* Different from PLS, CCA is to find linear combinations of both  $\mathbf{X}$  and  $\mathbf{Y}$  variables which have maximum correlation coefficient with each other. This leads to the same objective function but different constraints compared with (1):

$$\begin{aligned} \max_{\mathbf{v}_1, \mathbf{v}_2} & \quad (\mathbf{v}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{v}_2)^2 \\ \text{s.t.} & \quad \mathbf{v}_1^T \mathbf{X}^T \mathbf{X} \mathbf{v}_1 = 1, \quad \mathbf{v}_2^T \mathbf{Y}^T \mathbf{Y} \mathbf{v}_2 = 1, \end{aligned} \quad (2)$$

where  $\mathbf{v}_i$ 's ( $i = 1, 2$ ) are the weight vectors.

The solutions to this problem are the largest eigenvectors of the matrices  $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} (\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{X}$  and  $(\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ , respectively. The subsequent weights are the eigenvectors of the same matrix in the order of decreasing eigenvalues. The predictor LVs  $\mathbf{U}_X$  can be calculated directly from the original  $\mathbf{X}$  matrix as  $\mathbf{U}_X = \mathbf{X} \mathbf{V}_1$ , the columns of which are uncorrelated with each other. The detailed derivation is shown in the appendix. However, the solution depends heavily on whether or not the covariance matrix  $\mathbf{X}^T \mathbf{X}$  is invertible. In practice, it is possible to have  $\text{rank}(\mathbf{X}^T \mathbf{X}) < p$  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.

*2.1.3. 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 with 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 relationships between EEG and EMG signals are 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 cannot 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 cannot ensure that such LVs are nontrivial 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 the appendix, and the specific implementation procedure is shown in Algorithm 1.

*2.2. Synthetic Data.* In this simulation, we applied the proposed method to synthetic data and also reported the results of the PLS and CCA approaches for comparison. As an illustrative example, without loss of generality, four sources were generated and analyzed for each data set.

The following four source signals were considered for the data set  $\mathbf{X}$ :

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

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

**Input:** two data sets  $\mathbf{X}$  (with size  $N \times p$ ) and  $\mathbf{Y}$  (with size  $N \times q$ )  
**Output:** corresponding LVs matrices  $\mathbf{U}_X$ , and  $\mathbf{U}_Y$

**The First Step:**

- (1) Solve the eigen decomposition problems:  
 $(\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X}) \mathbf{w}_1 = \lambda_1 \mathbf{w}_1$  and  $(\mathbf{Y}^T \mathbf{X} \mathbf{X}^T \mathbf{Y}) \mathbf{w}_2 = \lambda_2 \mathbf{w}_2$ .
- (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, that is,  $\mathbf{T}_X = []$  and  $\mathbf{T}_Y = []$ .
- (6) **while**  $count > 0$  **do**
- (7) Set  $\mathbf{w}_1$  and  $\mathbf{w}_2$  to be the largest eigenvectors of the matrices  $\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X}$  and  $\mathbf{Y}^T \mathbf{X} \mathbf{X}^T \mathbf{Y}$ , respectively.
- (8) Calculate the LVs as  $\mathbf{t}_X = \mathbf{X} \mathbf{w}_1$  and  $\mathbf{t}_Y = \mathbf{Y} \mathbf{w}_2$ .
- (9) Set  $\mathbf{T}_X = [\mathbf{T}_X \ \mathbf{t}_X]$  and  $\mathbf{T}_Y = [\mathbf{T}_Y \ \mathbf{t}_Y]$ .
- (10) Deflate  $\mathbf{X}$  by subtracting the effects of the LV  $\mathbf{t}_X$  from the data space:  
 $\mathbf{X} = \mathbf{X} - \mathbf{t}_X (\mathbf{t}_X^T \mathbf{t}_X)^{-1} \mathbf{t}_X^T \mathbf{X}$ .
- (11) Deflate  $\mathbf{Y}$  by subtracting the effects of the LV  $\mathbf{t}_Y$  from the data space:  
 $\mathbf{Y} = \mathbf{Y} - \mathbf{t}_Y (\mathbf{t}_Y^T \mathbf{t}_Y)^{-1} \mathbf{t}_Y^T \mathbf{Y}$ .
- (12) Let  $count = count - 1$ .
- (13) **end while**

**The Second Step:**

- (14) Solve the following eigen decomposition problems:  
 $[(\mathbf{T}_X^T \mathbf{T}_X)^{-1} \mathbf{T}_X^T \mathbf{T}_Y (\mathbf{T}_Y^T \mathbf{T}_Y)^{-1} \mathbf{T}_Y^T \mathbf{T}_X] \mathbf{v}_1 = \eta_1 \mathbf{v}_1$  and  
 $[(\mathbf{T}_Y^T \mathbf{T}_Y)^{-1} \mathbf{T}_Y^T \mathbf{T}_X (\mathbf{T}_X^T \mathbf{T}_X)^{-1} \mathbf{T}_X^T \mathbf{T}_Y] \mathbf{v}_2 = \eta_2 \mathbf{v}_2$ .
- (15) Set  $\mathbf{V}_1$  and  $\mathbf{V}_2$  to be the  $R$  associated eigenvectors, respectively.
- (16) The recovered LVs  $\mathbf{U}_X$  and  $\mathbf{U}_Y$  can be calculated by  
 $\mathbf{U}_X = \mathbf{T}_X \mathbf{V}_1$  and  $\mathbf{U}_Y = \mathbf{T}_Y \mathbf{V}_2$ .

ALGORITHM 1: The combined PLS + CCA method.

TABLE 1: The correlation coefficients between the corresponding source pairs of  $\mathbf{X}$  and  $\mathbf{Y}$ .

	$\mathbf{s}_{11}$ and $\mathbf{s}_{21}$	$\mathbf{s}_{12}$ and $\mathbf{s}_{22}$	$\mathbf{s}_{13}$ and $\mathbf{s}_{23}$	$\mathbf{s}_{14}$ and $\mathbf{s}_{24}$
CC*	0.3655	0.8787	0.5520	1.00

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

Also, four source signals were considered for the data set  $\mathbf{Y}$ :

$$\begin{aligned}
 \mathbf{s}_{21} &= 1.5 \sin(0.025(t + 69)) \sin(0.2(t + 6)), \\
 \mathbf{s}_{22} &= 1.5 \sin(0.025(t + 20)), \\
 \mathbf{s}_{23} &= \text{sign}(\sin(0.3(t + 7))) + 3 \cos(0.1(t + 7)), \\
 \mathbf{s}_{24} &= \text{uniformly distributed noise (the same as } \mathbf{s}_{14}),
 \end{aligned} \tag{4}$$

where the notations are similarly defined. The four simulated sources are shown in Figure 1(b).

Two mixed data sets  $\mathbf{X}$  and  $\mathbf{Y}$  were generated as follows with each row denoting one observation in their respective data space:

$$\mathbf{X} = \mathbf{S}_1 \cdot \mathbf{A}, \quad \mathbf{Y} = \mathbf{S}_2 \cdot \mathbf{B}, \tag{5}$$

where  $\mathbf{S}_1 = [\mathbf{s}_{11} \ \mathbf{s}_{12} \ \mathbf{s}_{13} \ \mathbf{s}_{14}]$  and  $\mathbf{S}_2 = [\mathbf{s}_{21} \ \mathbf{s}_{22} \ \mathbf{s}_{23} \ \mathbf{s}_{24}]$  with

$$\begin{aligned}
 \mathbf{A} &= \begin{bmatrix} 0.76 & -0.65 & 0.77 & 0.83 & 0.82 \\ 0.49 & 0.25 & 0.12 & 0.22 & -0.17 \\ 0.28 & -0.21 & 0.11 & 0.19 & -0.11 \\ 0.07 & 0.06 & -0.08 & 0.07 & -0.04 \end{bmatrix}, \\
 \mathbf{B} &= \begin{bmatrix} 0.73 & -0.82 & 0.91 & -0.79 & 0.88 \\ 0.42 & -0.27 & 0.17 & -0.20 & -0.30 \\ 0.27 & 0.26 & -0.18 & 0.17 & -0.24 \\ 0.08 & -0.01 & 0.01 & 0.09 & -0.01 \end{bmatrix}.
 \end{aligned} \tag{6}$$

The patterns of the corresponding sources are similar across the two data sets, representing common information. However, from (3) and (4), we can see that there are some time-shifts between corresponding source pairs, and their correlation coefficients are given in Table 1. The first pair 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  $\mathbf{A}$  and  $\mathbf{B}$ . This pair can represent the major information within individual data sets but cannot 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

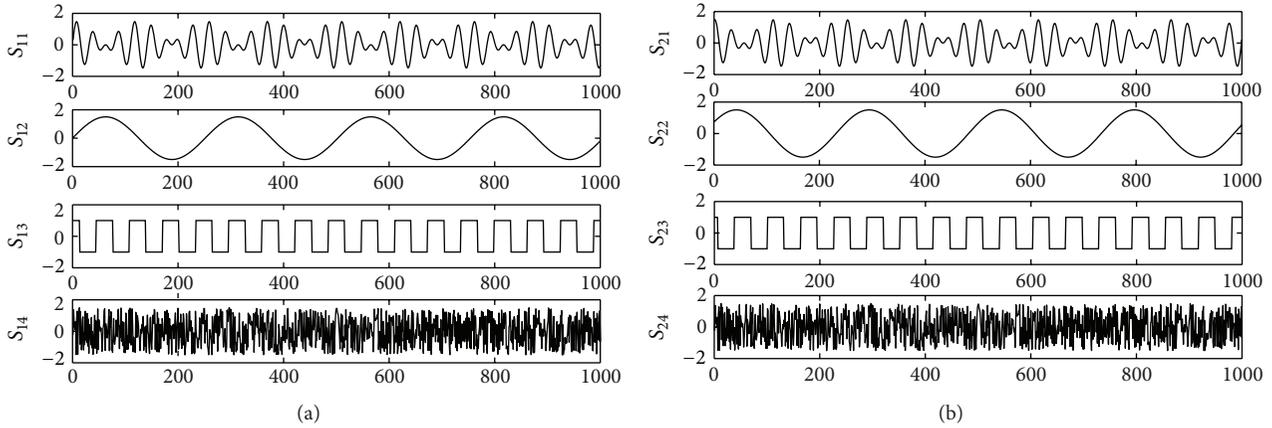


FIGURE 1: The four simulated source signals: (a) for  $X$ ; (b) for  $Y$ .

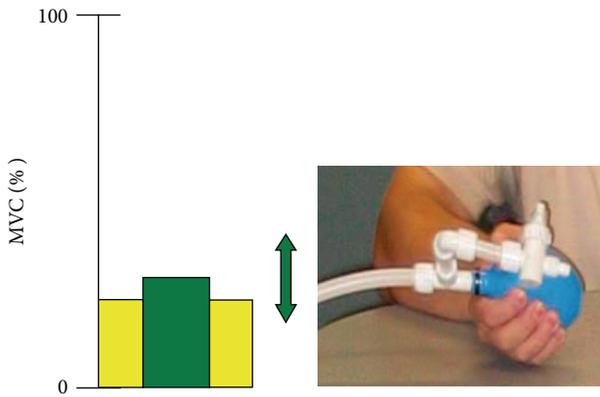


FIGURE 2: The squeezing task: the subject was instructed to follow the target bar (yellow) as closely as possible, and the green bar shows the force exerted by the subject.

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.

**2.3. Real Data.** In many medical applications, the results of analyzing one subject's data cannot be generalized to the population level because of the intersubject variability concern. Therefore, it is necessary to recruit a proper number of subjects to perform a group analysis in many medical applications. 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.

**2.3.1. Data Collection.** 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 (stages 1-2 on the Hoehn and Yahr scale) and were being treated with L-dopa medication (mean daily dose of 720 mg). 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 Figure 2. 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.

The EEG data were collected using an EEG cap (Quick-Cap, Compumedics, TX, USA) with 19 electrodes based on the International 10–20 system. The EEG data were sampled at 1000 Hz using SynAmps2 amplifiers (NeuroScan, Compumedics, TX, 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, ON, Canada). Data were later processed by a band-pass filter (1 to 70 Hz) offline and downsampled to 250 Hz. Artifacts associated with eye blinks and muscular activities were removed using the Automated Artifact Removal in the EEGLAB Matlab Toolbox [18]. The raw time sequences of the electrodes were then normalized to have zero-mean and 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.

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 downsampled offline to 250 Hz, and only the squeezing periods were used for subsequent analysis.

**2.3.2. 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 work, we examine the coupling relationships between time-varying EEG features and amplitudes of the EMG signals, constituting  $\mathbf{X}_b$  and  $\mathbf{Y}_b$ , respectively, for each subject  $b$  (for  $b = 1, 2, \dots, B$ ). 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

$$\begin{aligned} \mathbf{X} &= [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_B], \quad \forall b = 1, 2, \dots, B, \\ \mathbf{Y} &= [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_B], \quad \forall b = 1, 2, \dots, B \end{aligned} \quad (7)$$

with the assumption that all subjects share common group patterns in the temporal dimension [19].

**EEG Features.** Pairwise Pearson's correlations [20] are considered in this study. Pearson's correlation measures the dependency between a pair of EEG signals  $\mathbf{e}^* = (e_1^*, e_2^*, \dots, e_n^*)$  and  $\mathbf{e}^\circ = (e_1^\circ, e_2^\circ, \dots, e_n^\circ)$  in the time domain as follows:

$$\gamma_{e^*e^\circ} = \frac{\sum_{i=1}^n (e_i^* - \bar{e}^*)(e_i^\circ - \bar{e}^\circ)}{\sqrt{\sum_{i=1}^n (e_i^* - \bar{e}^*)^2 \sum_{i=1}^n (e_i^\circ - \bar{e}^\circ)^2}}, \quad (8)$$

where  $\bar{e}^*$  and  $\bar{e}^\circ$  are the sample means of  $\mathbf{e}^*$  and  $\mathbf{e}^\circ$ . In this work, we calculate the time-varying pairwise 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, that is, pairwise 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 [21]. While different techniques have been proposed for accurate amplitude estimation, in this study, we employ the root-mean-square (RMS) approach to

calculate the EMG amplitude of short-duration EMG signals  $\mathbf{e} = (e_1, e_2, \dots, e_n)$ :

$$e_{\text{rms}} = \sqrt{\frac{1}{n} (e_1^2 + e_2^2 + \dots + e_n^2)}. \quad (9)$$

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  $b = 1, 2, \dots, B$ ),  $\mathbf{X}_b$  and  $\mathbf{Y}_b$  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, thus there are a total of  $C_2^{19} = 171$  correlation connections. Therefore,  $\mathbf{X}_b$  is of size  $480 \times 171$ . For the EMG amplitude feature, since there are three surface EMG channels,  $\mathbf{Y}_b$  is of size  $480 \times 3$ .

**2.3.3. Significance Assessment.** To determine the statistical significance levels of the extracted LVs, we employ a nonparametric permutation test [22] in which the temporal order of EEG features  $\mathbf{X}_b$  is uniformly permuted for all subjects while keeping the EMG features  $\mathbf{Y}_b$  intact. Two hundred random permutations are generated. The proposed PLS + CCA method described in Section 2.1.3 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  $\mathbf{LV}_{\text{EEG}}$  and  $\mathbf{LV}_{\text{EMG}}$ , both with size  $(N \times K)$ , where  $K$  means the number of significant components.

**2.3.4. 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  $\mathbf{LV}_{\text{EEG}}$  back onto the EEG features  $\mathbf{X}_b$  (for  $b = 1, 2, \dots, B$ ) for each subject as follows:

$$\mathbf{p}_{bk} = \sqrt{\frac{1}{\mathbf{lv}_k^T \mathbf{X}_b \mathbf{X}_b^T \mathbf{lv}_k}} \mathbf{X}_b^T \mathbf{lv}_k, \quad k = 1, 2, \dots, K, \quad (10)$$

where  $\mathbf{lv}_k$  is the  $k$ th column of  $\mathbf{LV}_{\text{EEG}}$  and  $\mathbf{p}_{bk}$  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.

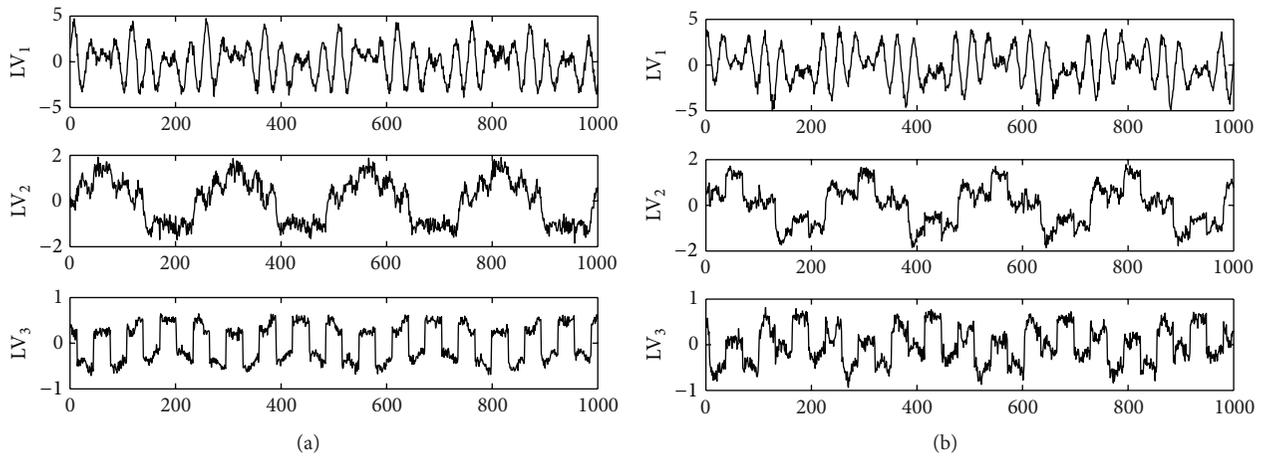


FIGURE 3: (a) The LVs estimated in X using PLS. (b) The LVs estimated in Y using PLS.

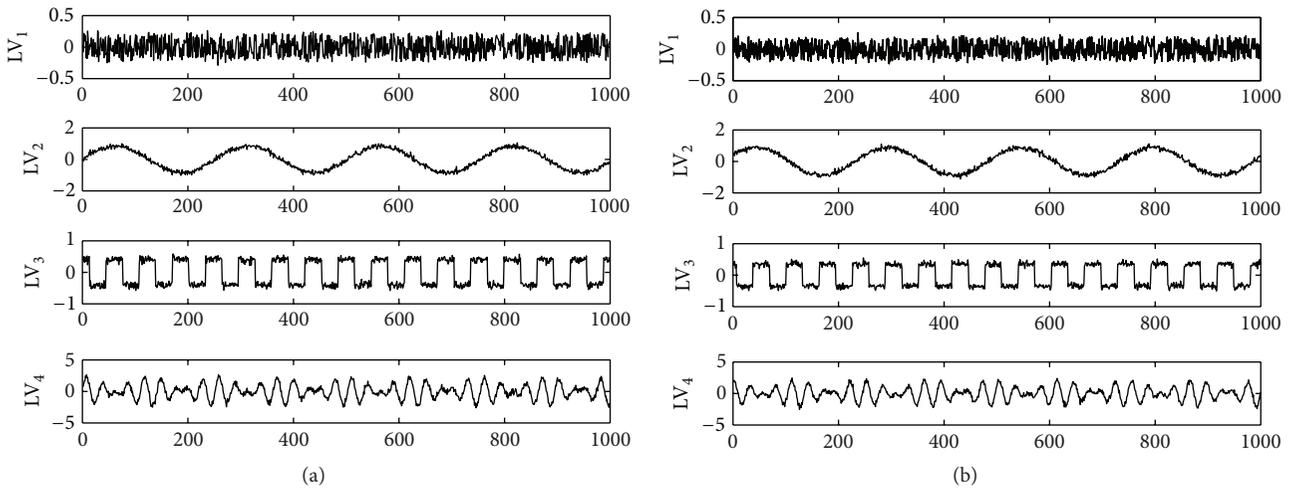


FIGURE 4: (a) The LVs estimated in X using CCA. (b) The LVs estimated in Y using CCA.

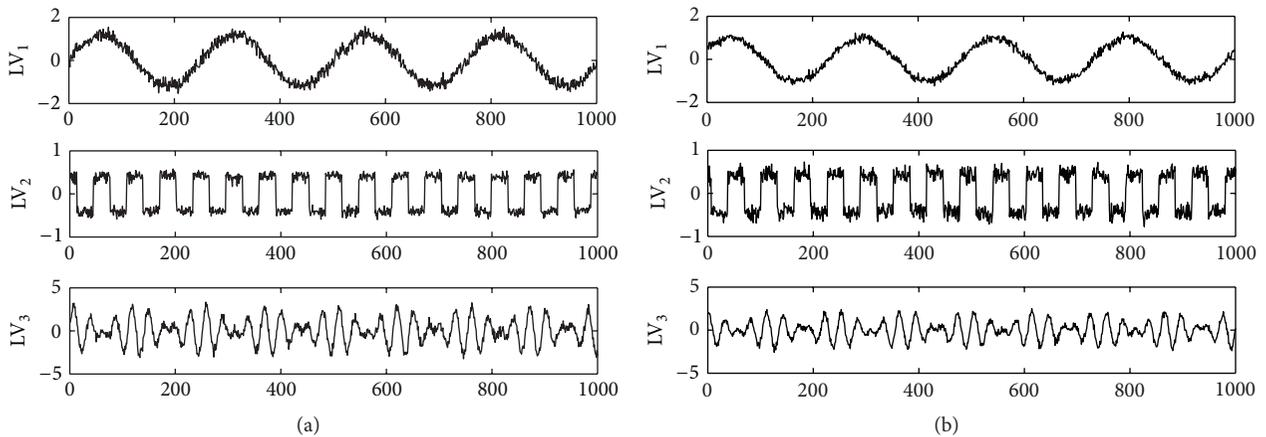


FIGURE 5: (a) The LVs estimated in X using the proposed PLS + CCA. (b) The LVs estimated in Y using the proposed PLS + CCA.

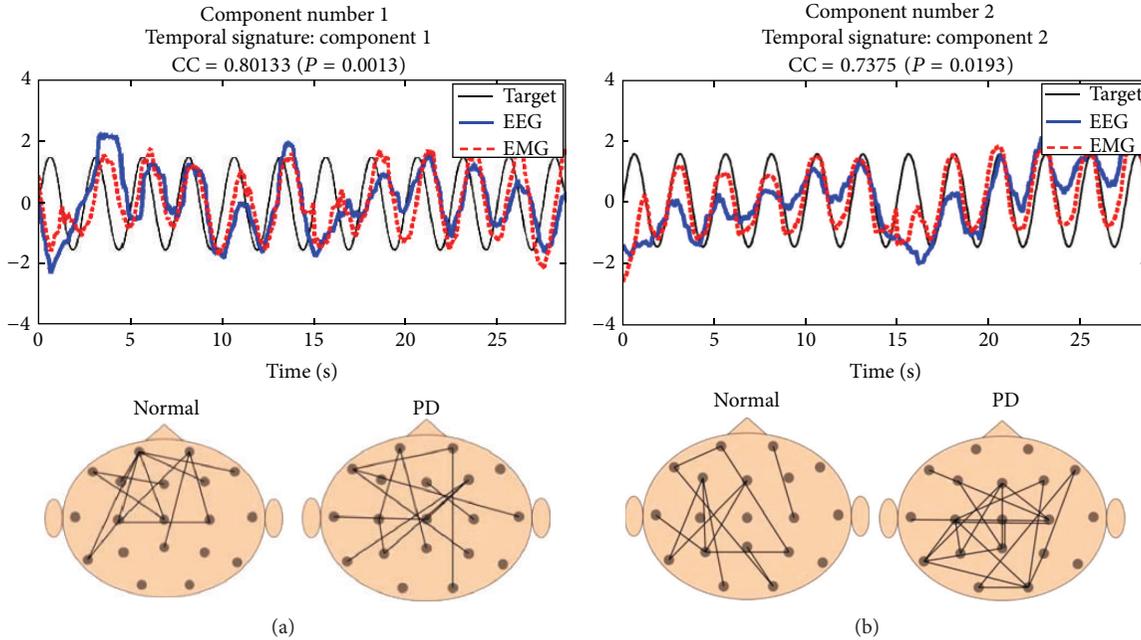


FIGURE 6: The two components from the proposed PLS + CCA method when using the EEG correlation features and the EMG amplitude features as data sets. Top panel: temporal patterns of the EEG (blue, solid) and the EMG (red, dashed). The oscillation of the target bar is also shown (black, solid). Bottom 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.

### 3. Results and Discussion

**3.1. The Synthetic Data Case.** The extracted components using PLS, CCA, and the proposed PLS + CCA methods are shown in Figures 3, 4, and 5, 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 between  $\mathbf{X}$  and  $\mathbf{Y}$ . 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 trivial information of the data sets (e.g., the 1st LV). For instance, although the first pair of LVs has 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 do not include the first pair recovered by CCA due to their triviality. The above observations motivate 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.

**3.2. The Real Data Case.** In this case study, we applied the proposed method for corticomuscular activity modeling to the EEG and EMG features generated using the procedure described in Section 2.3.2 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 ( $P \leq 0.05$ ) (Figure 6). Note that in the figure only connections whose weights are statistically different from zero are shown. The results based on real data from PD and normal subjects performing a dynamic motor task are promising. In the past, most EEG/EMG coupling studies have compared EEG/EMG 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 (Figure 6). It is likely that the dynamic nature of the task might require the recruitment of additional regions such as

frontal regions for motor selection [23], contralateral (i.e., ipsilateral to hand movement) sensorimotor cortex for fine modulatory control [24], and occipital regions for posterror adaptations [25].

From Figure 6, 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 [26], a problem common in this PD population [27]. 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 [28] and ongoing control of movement [29]. Moreover, the increased intra- and interhemispheric connections observed in the PD subjects are consistent with the findings in previous MEG studies [30].

#### 4. Conclusions

In this paper, we combine the advantages of PLS and CCA and propose a PLS + CCA method to improve the performance of the joint LV extraction. We illustrate the performances of the proposed approach using both synthetic data and real-life data. For corticomuscular activity modeling, we note highly correlated temporal patterns between EEG and EMG signals and meaningful spatial activation patterns. The proposed method is a promising analysis technique for multisubject and bimodal data sets, including, but not limited to, fMRI, PPG, ECG, and kinematic data.

### Appendix

#### A. The Derivation of the Algorithm

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

*A.1. The First Step: PLS.* The cost function of PLS is as follows (the same as (1) in Section 2.1.1):

$$\begin{aligned} \max_{\mathbf{w}_1, \mathbf{w}_2} \quad & (\mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2)^2 \\ \text{s.t.} \quad & \mathbf{w}_i^T \mathbf{w}_i = 1, \quad i = 1, 2, \end{aligned} \quad (\text{A.1})$$

where  $\mathbf{w}_i$ 's ( $i = 1, 2$ ) are the weight vectors.

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

$$L(\mathbf{w}_i, \lambda_i) = (\mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2)^2 - \sum_{i=1}^2 \lambda_i (\mathbf{w}_i^T \mathbf{w}_i - 1), \quad (\text{A.2})$$

where  $\lambda_i$ 's are Lagrange multipliers.

Now we only present the detailed derivations regarding  $\mathbf{w}_1$  since  $\mathbf{w}_2$  can be similarly derived. Taking the derivatives of  $L(\mathbf{w}_i, \lambda_i)$  with respect to  $\mathbf{w}_1$  and  $\lambda_1$  and setting them to be zero, we have

$$\nabla L_{\mathbf{w}_1} = 2 \left| \mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2 \right| \mathbf{X}^T \mathbf{Y} \mathbf{w}_2 - 2\lambda_1 \mathbf{w}_1 = 0, \quad (\text{A.3})$$

$$\nabla L_{\lambda_1} = \mathbf{w}_1^T \mathbf{w}_1 - 1 = 0. \quad (\text{A.4})$$

Left multiplying both sides of (A.3) by  $\mathbf{w}_1^T$ , we have

$$2 \left( \mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2 \right)^2 - 2\lambda_1 \mathbf{w}_1^T \mathbf{w}_1 = 0. \quad (\text{A.5})$$

According to (A.4),  $\lambda_1$  can be calculated as

$$\lambda_1 = \left( \mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2 \right)^2. \quad (\text{A.6})$$

Through the similar procedure,  $\nabla L_{\mathbf{w}_2}$  and  $\lambda_2$  can be easily derived as

$$\nabla L_{\mathbf{w}_2} = 2 \left| \mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2 \right| \mathbf{Y}^T \mathbf{X} \mathbf{w}_1 - 2\lambda_2 \mathbf{w}_2 = 0, \quad (\text{A.7})$$

$$\lambda_2 = \left( \mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2 \right)^2. \quad (\text{A.8})$$

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

$$\sqrt{\lambda_2} \mathbf{X}^T \mathbf{Y} \mathbf{w}_2 = \lambda_1 \mathbf{w}_1, \quad (\text{A.9})$$

$$\frac{1}{\sqrt{\lambda_2}} \mathbf{Y}^T \mathbf{X} \mathbf{w}_1 = \mathbf{w}_2. \quad (\text{A.10})$$

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

$$\left( \mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X} \right) \mathbf{w}_1 = \lambda_1 \mathbf{w}_1. \quad (\text{A.11})$$

Similarly, we can have the formulation for  $\mathbf{w}_2$  as

$$\left( \mathbf{Y}^T \mathbf{X} \mathbf{X}^T \mathbf{Y} \right) \mathbf{w}_2 = \lambda_2 \mathbf{w}_2. \quad (\text{A.12})$$

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  $\mathbf{X}$  and  $\mathbf{Y}$  can be calculated by

$$\mathbf{T}_X = \mathbf{X} \mathbf{W}_1, \quad \mathbf{T}_Y = \mathbf{Y} \mathbf{W}_2, \quad (\text{A.13})$$

where  $\mathbf{W}_1$  is composed of the first  $R$  eigenvectors associated with (A.11) and the columns of  $\mathbf{T}_X$  represent the  $R$  components extracted from  $\mathbf{X}$ .  $\mathbf{W}_2$  and  $\mathbf{T}_Y$  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,  $\mathbf{X}$  and  $\mathbf{Y}$  are deflated by their corresponding first LVs as follows:

$$\begin{aligned}\mathbf{X} &= \mathbf{X} - \mathbf{t}_X(\mathbf{t}_X^T \mathbf{t}_X)^{-1} \mathbf{t}_X^T \mathbf{X}, \\ \mathbf{Y} &= \mathbf{Y} - \mathbf{t}_Y(\mathbf{t}_Y^T \mathbf{t}_Y)^{-1} \mathbf{t}_Y^T \mathbf{Y}.\end{aligned}\quad (\text{A.14})$$

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.

*A.2. The Second Step: CCA.* Based on the extracted LVs in the first step, the objective function of CCA can be constructed as follows:

$$\begin{aligned}\max_{\mathbf{v}_1, \mathbf{v}_2} & \quad (\mathbf{v}_1^T \mathbf{T}_X^T \mathbf{T}_Y \mathbf{v}_2)^2 \\ \text{s.t.} & \quad \mathbf{v}_1^T \mathbf{T}_X^T \mathbf{T}_X \mathbf{v}_1 = 1, \quad \mathbf{v}_2^T \mathbf{T}_Y^T \mathbf{T}_Y \mathbf{v}_2 = 1,\end{aligned}\quad (\text{A.15})$$

where  $\mathbf{v}_i$ 's ( $i = 1, 2$ ) are the weight vectors.

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

$$\begin{aligned}L(\mathbf{v}_i, \eta_i) &= (\mathbf{v}_1^T \mathbf{T}_X^T \mathbf{T}_Y \mathbf{v}_2)^2 - \eta_1 (\mathbf{v}_1^T \mathbf{T}_X^T \mathbf{T}_X \mathbf{v}_1 - 1) \\ &\quad - \eta_2 (\mathbf{v}_2^T \mathbf{T}_Y^T \mathbf{T}_Y \mathbf{v}_2 - 1),\end{aligned}\quad (\text{A.16})$$

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

$$\left[ (\mathbf{T}_X^T \mathbf{T}_X)^{-1} \mathbf{T}_X^T \mathbf{T}_Y (\mathbf{T}_Y^T \mathbf{T}_Y)^{-1} \mathbf{T}_Y^T \mathbf{T}_X \right] \mathbf{v}_1 = \eta_1 \mathbf{v}_1. \quad (\text{A.17})$$

Similarly, for  $\mathbf{v}_2$ , we have

$$\left[ (\mathbf{T}_Y^T \mathbf{T}_Y)^{-1} \mathbf{T}_Y^T \mathbf{T}_X (\mathbf{T}_X^T \mathbf{T}_X)^{-1} \mathbf{T}_X^T \mathbf{T}_Y \right] \mathbf{v}_2 = \eta_2 \mathbf{v}_2. \quad (\text{A.18})$$

The solutions to this problem are the  $R$  largest eigenvectors of the corresponding matrices. The recovered LVs  $\mathbf{U}_X$  and  $\mathbf{U}_Y$  can be calculated directly from the matrices  $\mathbf{T}_X$  and  $\mathbf{T}_Y$  by

$$\mathbf{U}_X = \mathbf{T}_X \mathbf{V}_1, \quad \mathbf{U}_Y = \mathbf{T}_Y \mathbf{V}_2, \quad (\text{A.19})$$

where  $\mathbf{V}_1$  is composed of the  $R$  eigenvectors associated with (A.17) and the columns of  $\mathbf{U}_X$  represent the  $R$  components extracted from  $\mathbf{T}_X$ .  $\mathbf{V}_2$  and  $\mathbf{U}_Y$  are similarly defined.

After these two steps, it is ensured that the extracted components  $\mathbf{U}_X$  and  $\mathbf{U}_Y$  are maximally correlated across data sets and meanwhile can well explain the information within each individual data set.

## References

- [1] S. N. Baker, E. Olivier, and R. N. Lemon, "Coherent oscillations in monkey motor cortex and hand muscle EMG show task-dependent modulation," *Journal of Physiology*, vol. 501, no. 1, pp. 225–241, 1997.
- [2] B. A. Conway, D. M. Halliday, S. F. Farmer et al., "Synchronization between motor cortex and spinal motoneuronal pool during the performance of a maintained motor task in man," *Journal of Physiology*, vol. 489, no. 3, pp. 917–924, 1995.
- [3] D. M. Halliday, B. A. Conway, S. F. Farmer, and J. R. 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.
- [4] V. N. Murthy and E. E. Fetz, "Coherent 25- to 35-Hz oscillations in the sensorimotor cortex of awake behaving monkeys," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 12, pp. 5670–5674, 1992.
- [5] J. Chiang, Z. J. Wang, and 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.
- [6] X. Chen, X. Chen, R. K. Ward, and Z. J. Wang, "A joint multimodal group analysis framework for modeling corticomuscular activity," *IEEE Transactions on Multimedia*, 2013.
- [7] X. Chen, C. He, Z. J. Wang, and M. J. McKeown, "An IC-PLS framework for group corticomuscular coupling analysis," *IEEE Transactions on Biomedical Engineering*, 2013.
- [8] R. Bortel and P. Sovka, "EEG-EMG coherence enhancement," *Signal Processing*, vol. 86, no. 7, pp. 1737–1751, 2006.
- [9] B. S. Everitt and G. Dunn, *Applied Multivariate Data Analysis*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2010.
- [10] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Analytica Chimica Acta C*, vol. 185, pp. 1–17, 1986.
- [11] A. McIntosh, F. Bookstein, J. V. Haxby, and C. L. Grady, "Spatial pattern analysis of functional brain images using partial least squares," *NeuroImage*, vol. 3, no. 3, pp. 143–157, 1996.
- [12] E. Martínez-Montes, P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen, "Concurrent EEG/fMRI analysis by multiway partial least squares," *NeuroImage*, vol. 22, no. 3, pp. 1023–1034, 2004.
- [13] H. Yu and J. F. MacGregor, "Post processing methods (PLS-CCA): simple alternatives to preprocessing methods (OSC-PLS)," *Chemometrics and Intelligent Laboratory Systems*, vol. 73, no. 2, pp. 199–205, 2004.
- [14] W. D. Clercq, A. Vergult, B. Vanrumste, W. V. Paesschen, and S. V. Huel, "Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2583–2587, 2006.
- [15] 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.
- [16] N. M. Correa, T. Adali, Y. O. 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.
- [17] F. Lindgren, P. Geladi, and S. Wold, "The kernel algorithm for PLS," *Journal of Chemometrics*, vol. 7, no. 1, pp. 45–59, 2005.
- [18] G. Gomez-Herrero, "Automatic Artifact Removal (AAR) toolbox v1.3 (Release 09.12.2007) for MATLAB," 2007.

- [19] V. D. Calhoun, J. Liu, and T. Adali, "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data," *NeuroImage*, vol. 45, supplement 1, pp. S163–172, 2009.
- [20] S. Weiss and P. Rappelsberger, "Long-range EEG synchronization during word encoding correlates with successful memory performance," *Cognitive Brain Research*, vol. 9, no. 3, pp. 299–312, 2000.
- [21] E. A. Clancy, E. L. 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.
- [22] P. Good, *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*, Springer Series in Statistics, Springer, Heidelberg, Germany, 2nd edition, 2000.
- [23] P. Praamstra, L. Boutsen, and 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.
- [24] J. Liepert, C. Dettmers, C. Terborg, and C. Weiller, "Inhibition of ipsilateral motor cortex during phasic generation of low force," *Clinical Neurophysiology*, vol. 112, no. 1, pp. 114–121, 2001.
- [25] C. Danielmeier, T. Eichele, B. U. Forstmann, M. Tittgemeyer, and M. Ullsperger, "Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas," *Journal of Neuroscience*, vol. 31, no. 5, pp. 1780–1789, 2011.
- [26] P. Sauseng, W. Klimesch, R. Freunberger, T. Pecherstorfer, S. Hanslmayr, and M. Doppelmayr, "Relevance of EEG alpha and theta oscillations during task switching," *Experimental Brain Research*, vol. 170, no. 3, pp. 295–301, 2006.
- [27] I. G. M. Cameron, M. Watanabe, G. Pari, and D. P. Munoz, "Executive impairment in Parkinson's disease: response automaticity and task switching," *Neuropsychologia*, vol. 48, no. 7, pp. 1948–1957, 2010.
- [28] P. Praamstra, D. F. Stegeman, A. R. Cools, and M. W. I. M. 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.
- [29] J. K. R. Stevenson, M. M. K. Oishi, S. Farajian, E. Cretu, E. Ty, and 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.
- [30] D. Stoffers, J. L. Bosboom, J. B. Deijen, E. C. Wolters, and C. J. Stam, "Increased cortico-cortical functional connectivity in early-stage Parkinson's disease: an MEG study," *NeuroImage*, vol. 41, no. 2, pp. 212–222, 2008.