

EEG Source Extraction by Autoregressive Source Separation Reveals Abnormal Synchronization in Parkinson's Disease

Joyce Chiang, Z. Jane Wang, and Martin J. McKeown

Abstract—Recent research efforts in studying brain connectivity has provided new perspectives to understanding of neurophysiology of brain function. Connectivity measures are typically computed from electroencephalogram (EEG) signals, yet the presence of volume conduction makes interpretation of results difficult. One possible alternative is to model the connectivity in the source space. In this study, we proposed a novel source separation technique in which EEG signals are represented as a state-space framework. The framework jointly models the underlying brain sources and the connectivity between them in the form of a generalized autoregressive (AR) process. The proposed technique was applied to real EEG data collected from normal and Parkinson's patients during a motor task. The extracted sources revealed the abnormal beta activity in Parkinson's subjects and showed similar biological networks as previous studies.

I. INTRODUCTION

In neurobiology, there has been increasing interest in identifying functional connectivity between brain regions, as connectivity is believed to provide an integrating framework for a variety of complex brain functions. Several mathematical methods have been explored to provide a quantitative measure of brain connectivity using electroencephalography (EEG) data, including correlation, coherence and Granger causality [1], [2]. In most studies, these connectivity measures are directly computed from scalp EEG signals. However, owing to volume conduction, the signal measured from an EEG electrode does not exclusively represent the activity of one local neural source, but rather the superposition of active sources throughout the brain [3]. This can give rise to spurious correlations between scalp EEG signals and potentially lead to misinterpretation of the connectivity results.

To address the problem of volume conduction in EEG, several studies have proposed studying the brain connectivity in the "source space" which involves estimating the underlying neural sources from scalp EEG signals. Determining neural sources from EEG signals is a mathematically ill-conditioned inverse problem, and it has no unique solution without prior knowledge or strong statistical assumptions [4]. One common approach to solving the source identification problem is the dipole modeling technique, which aims to find current dipoles that best describe the observed EEG signals. However, the successful estimation of sources depends on many factors, including models of the generators of the

electrical activity, knowledge of the number of active dipoles, and the shape and conductivity of the head model [5].

Another possible approach to finding brain sources is to use blind source separation (BSS) algorithms. BSS aims to extract a set of underlying sources from observed signals by making certain assumptions about the statistical properties of the source signals. One of the most widely used BSS algorithms in neuroscience studies is independent component analysis (ICA). Fundamentally, ICA assumes the sources are mutually independent. Most of the commonly used ICA algorithms further assume that observed signals are instantaneous linear mixtures of the underlying sources and their temporal ordering is irrelevant [6]. However, the underlying assumption of statistical independence between the activations of different neural assemblies still remains to be validated experimentally [4]. Furthermore, as we try to determine the causal relationships between EEG sources, it is implicitly assumed that the sources are not only temporally correlated within themselves but also cross-correlated between each other. This contradicts the fundamental assumption of ICA.

To address these concerns, we propose a state-space framework which allows the underlying brain sources and the connectivity between them to be modeled jointly. In this framework, the scalp EEG signals are represented as linear mixtures of unobserved brain sources resulted from volume conduction. The sources are modeled by a multivariate autoregressive (mAR) process, where the residuals are assumed to be mutually independent and follow generalized Gaussian distributions. This framework is similar to the one proposed by Gómez-Herrero *et al.* [7], where the source separation is done by a heuristic, multi-stage approach which involves performing mAR estimation and ICA sequentially. However, as noted by the authors that the successful separation of sources depends strongly on the accuracy of the mAR estimation step, including the fitting of the mAR model to EEG data and parameter estimation. Here, we propose solving this framework analytically by modeling the residual process of the sources explicitly using generalized Gaussian distribution. This allows the simultaneous estimation of the sources and mAR parameters using a maximum likelihood approach.

The paper is organized as follows. In Section II, we introduce the generalized multivariate AR (GmAR) source separation algorithm and the clustering method used for finding group-representative source estimates. The proposed source separation algorithm is applied to EEG data collected from 6 healthy subjects and 5 Parkinson's disease (PD) patients during a right-handed bulb-squeezing task. The

J. Chiang and Z. J. Wang are with Department of Electrical and Computer Engineering, University of British Columbia, Canada joycehc@interchange.ubc.ca, zjanew@ece.ubc.ca

M. J. McKeown is with Pacific Parkinson's Research Centre, University of British Columbia, Canada mmckeown@interchange.ubc.ca

clustering results are discussed in Section III.

II. METHODS

A. AR Source Separation

1) *EEG Signal Model*: Let $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T$ denote an M -dimensional EEG data vector at time t , where the superscript T denotes the transpose. We assume that $\mathbf{x}(t)$ are generated from an instantaneous mixing of the sources, which is written as

$$\mathbf{x}(t) = \mathbf{C}\mathbf{s}(t), \quad 1 \leq t \leq N \quad (1)$$

where $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_M(t)]^T$ denotes the underlying EEG sources, and \mathbf{C} is an $M \times M$ mixing matrix where the i -th column represents the projection from the i -th source to the scalp electrodes, also commonly referred to as ‘‘scalp maps’’ in ICA literature.

In the proposed signal model, we assume that the sources $\mathbf{s}(t)$ are modeled by a generalized multivariate AR (GmAR) process of order P , which is written as

$$\mathbf{s}(t) = \sum_{p=1}^P \mathbf{A}_p \mathbf{s}(t-p) + \mathbf{v}(t) \quad (2)$$

where $\mathbf{A}_p \in \mathbb{R}^{M \times M}$ is the mAR coefficient matrix at time lag p and $\mathbf{v}(t) = [v_1(t), v_2(t), \dots, v_M(t)]^T$ denotes the residual process whose elements are assumed to be mutually independent. Furthermore, we assume that each element of \mathbf{v} follows a zero-mean, univariate generalized Gaussian distribution with the density

$$p(v_i(t)|R_i, w_i) = \frac{R_i}{2w_i \Gamma(1/R_i)} e^{-|\frac{v_i(t)}{w_i}|^{R_i}}, \quad (3)$$

where $\Gamma(\cdot)$ is the gamma function and the noise parameters R_i and w_i determine the shape and the width of the distribution. Generalized Gaussian distribution is a more general parametric form of Gaussian which can model super-Gaussian when $R_i < 2$, Gaussian when $R_i = 2$, and sub-Gaussian when $R_i > 2$. This noise model has been used in an ICA model recently developed by Penny *et al.* for modeling EEG signals [8].

In summary, the signal model can be specified by the following parameter set

$$\boldsymbol{\theta} = \{\mathbf{C}, \mathbf{A}_p, R_i, w_i\}, \quad \begin{aligned} 1 \leq p \leq P \\ 1 \leq i \leq M. \end{aligned} \quad (4)$$

Our objective is to estimate the model parameter set $\boldsymbol{\theta}$, together with the time courses of $\mathbf{s}(t)$, using the maximum likelihood approach.

2) *Maximum Likelihood Estimation*: In this paper, the model parameters are estimated using an iterative maximum-likelihood approach. Specifically, at each estimation, the unobserved sources $\mathbf{s}(t)$ are first estimated using \mathbf{W} (defined below) estimated from previous iteration, and the sources are used, as if they were observed, to update the log-likelihood function. The model parameters are then re-estimated based on the updated log-likelihood function.

Let us first define the data likelihood (likelihood of all samples of observations) with respect to $\boldsymbol{\theta}$ as follows

$$L \triangleq p_{\mathbf{x}}(\mathbf{x}(1), \mathbf{x}(1), \dots, \mathbf{x}(N)|\boldsymbol{\theta}) \quad (5)$$

Let \mathbf{W} denote the unmixing matrix which is defined as $\mathbf{W} = \mathbf{C}^{-1}$ such that $\mathbf{s}(t) = \mathbf{W}\mathbf{x}(t)$. Using the multivariate transformation, the likelihood of observations defined in (5) can be calculated from the likelihood of the source densities as follows

$$\begin{aligned} L &= \left(\frac{1}{|\det(\mathbf{W}^{-1})|} \right)^N p_{\mathbf{s}}(\mathbf{W}\mathbf{x}(1), \mathbf{W}\mathbf{x}(2), \dots, \mathbf{W}\mathbf{x}(N)) \\ &= (|\det(\mathbf{W})|)^N p_{\mathbf{s}}(\mathbf{s}(1), \mathbf{s}(2), \dots, \mathbf{s}(N)) \end{aligned} \quad (6)$$

where $|\cdot|$ denotes the absolute value. Since the sources follow a P -th order multivariate AR process, the conditional probability of each time point depends only upon its previous P time points, i.e.,

$$\begin{aligned} p_{\mathbf{s}}(\mathbf{s}(t)|\mathbf{s}(t-1), \dots, \mathbf{s}(1)) \\ &= p_{\mathbf{s}}(\mathbf{s}(t)|\mathbf{s}(t-1), \dots, \mathbf{s}(t-P)) \\ &= p_{\mathbf{v}}(\mathbf{s}(t) - \sum_{p=1}^P \mathbf{A}_p \mathbf{s}(t-p)) \end{aligned} \quad (7)$$

Under the assumption of mutual independence of the residuals, the right-hand side of (7) can be simplified as

$$\prod_{i=1}^M p_{v_i}(\mathbf{e}_i^T(\mathbf{s}(t) - \sum_{p=1}^P \mathbf{A}_p \mathbf{s}(t-p))|w_i, R_i), \quad (8)$$

where \mathbf{e}_i is the i -th column of an $M \times M$ identity matrix.

Combining (6), (7) and (8), we can rewrite the data likelihood as follows

$$\begin{aligned} L &= (|\det(\mathbf{W})|)^N p_{\mathbf{s}}(\mathbf{s}(1), \dots, \mathbf{s}(P)) \\ &\quad \prod_{t=P+1}^N \prod_{i=1}^M p_{v_i}(\mathbf{e}_i^T(\mathbf{s}(t) - \sum_{p=1}^P \mathbf{A}_p \mathbf{s}(t-p))|w_i, R_i) \end{aligned} \quad (9)$$

By taking the logarithm of (9), we obtain the log likelihood function of the observations

$$\begin{aligned} LL &\triangleq \log p_{\mathbf{x}}(\mathbf{x}(1), \mathbf{x}(1), \dots, \mathbf{x}(N)) \\ &= N \log(|\det(\mathbf{W})|) + \log p_{\mathbf{s}}(\mathbf{s}(1), \dots, \mathbf{s}(P)) \\ &\quad + \sum_{t=P+1}^N \sum_{i=1}^M \log \frac{R_i}{2w_i \Gamma(1/R_i)} - \left| \frac{z_i(t)}{w_i} \right|^{R_i} \end{aligned} \quad (10)$$

where

$$z_i(t) = \mathbf{e}_i^T(\mathbf{s}(t) - \sum_{p=1}^P \mathbf{A}_p \mathbf{s}(t-p)). \quad (11)$$

ML estimates of model parameters are found by setting the respective partial derivative of the log-likelihood (i.e.,

$\frac{\partial LL}{\partial \mathbf{W}}$, $\frac{\partial LL}{\partial \mathbf{A}_p}$, $\frac{\partial LL}{\partial R_i}$ and $\frac{\partial LL}{\partial w_i}$ to zero. This gives

$$w_i = \left(\frac{R_i}{N} \sum_{t=P+1}^N |z_i|^{R_i} \right)^{\frac{1}{R_i}} \quad (12)$$

$$\frac{\partial LL}{\partial R_i} = \frac{N}{R_i} + \frac{N}{R_i^2} \Gamma' \left(\frac{1}{R_i} \right) - \sum_{t=P+1}^N \left(\frac{|z_i|}{w_i} \right)^{R_i} \log \frac{|z_i|}{w_i} = 0, \quad (13)$$

$$\frac{\partial LL}{\partial A_p(i, j)} = \quad (14)$$

$$\frac{R_i}{w_i^{R_i}} \sum_{t=P+1}^N |z_i(t)|^{R_i-1} \text{sign}(z_i(t)) s_j(t-p) = 0$$

$$\frac{\partial LL}{\partial W_{i,j}} = \quad (15)$$

$$N W_{i,j} - \frac{R_i}{w_i^{R_i}} |z_i(t)|^{R_i-1} \text{sign}(z_i(t)) \mathbf{e}_j^T \mathbf{W}^T \mathbf{s}(t) = 0.$$

Since (13), (14) and (15) do not have closed-form solutions, we solve them numerically using a gradient ascent approach such as the Newton-Raphson method. The iteration is repeated until convergence is reached. It is important to note that the algorithm does not guarantee a global maximum and the accuracy of the solution depends on the choice of the initial parameter set. To address this issue, we propose using a multi-start strategy where multiple random initial conditions are used to generate multiple solutions and the one with highest likelihood is chosen as the final estimate.

3) *Group Analysis using Clustering Method*: One challenge in the application of blind source separation to real EEG data is how to perform group analysis, in which we want to be able to extract a set of sources from subjects that are representative of the group as a whole. This is complicated by the fact that each subject may have a different set of active brain sources. To solve this problem, we take a similar approach as in [7], that is, we perform source separation on each subject separately and use the clustering technique from the *Icasso* software package [9] to find source components that have high repeatability across subjects. *Icasso* was originally developed to address the stochastic nature of ICA algorithms by providing methods for assessing repeatability and reliability of ICA-estimates. It works by running the ICA algorithms many times to generate a large number of independent components. These estimated components are clustered according to their mutual similarities (e.g., correlation coefficients) using agglomerative clustering. The clusters containing components that are found consistently across different runs of ICA are selected for subsequent analysis.

For the purpose of EEG group analysis, we made two modifications to the clustering method from *Icasso*. First, the sources are clustered based on the correlation coefficients between their scalp maps (i.e., columns of \mathbf{C}) as opposed to their time courses. This modification is made because

we are more interested in finding brain regions associated with the sources and less interested in their exact activation waveforms. Second, clustering is applied to sources obtained from all available subjects. This allows the detection of sources that are consistent across subjects.

After the similarities between sources are calculated, the clustering results can be visualized using a dendrogram. The number of clusters in the data is determined manually based on the tightness of clusters. The significance of a cluster is determined based on the inter-subject repeatability. In this paper, a cluster is considered significant if it contains sources from at least 50% of the subjects. Each significant cluster is represented by a single source estimate called ‘‘centrotype’’, which is the scalp map in the cluster that has the maximum sum of similarities to other scalp maps in the cluster.

It is important to point out that since the source separation is applied to each subject separately, the number of active sources and their ordering may vary from subject to subject. Thus, we define a new mixing matrix for each subject that combines both group-representative and subject-specific scalp map patterns. First, we define a group-level mixing matrix as follows: $\Phi_{group} = [\phi_1, \dots, \phi_K]$, where $\phi_i \in \mathbb{R}^{19 \times 1}$ denotes the scalp map corresponding to the i -th centrotype. Next, we create a subject-specific mixing matrix Φ^* for each subject by augmenting Φ_{group} with $\hat{\mathbf{C}}$, i.e., $\Phi^* = [\Phi_{group} \quad \hat{\mathbf{C}}]$, where $\hat{\mathbf{C}}$ which is defined as

$$\hat{\mathbf{C}} = \mathbf{C} - (\mathbf{C}^T \Phi \Phi^\dagger)^T, \quad (16)$$

where \dagger denotes the pseudo-inverse operation, and \mathbf{C} is the subject’s own mixing matrix give by the source separation algorithm. The resulting temporal dynamics of the EEG sources of each subject can be obtained by the means of spatial filtering, i.e., $\mathbf{s}^*(t) = (\Phi^*)^\dagger \mathbf{x}^*(t)$.

III. RESULTS

In this section, we evaluate the performance of the proposed framework for modeling real EEG data and discuss the group analysis results when applied to data collected from PD and normal subjects.

A. EEG Data Collection and Preprocessing

5 PD subjects and 6 age-matched controls (Mean age=66 yrs, mean PD UPDRS score=23) were recruited from the

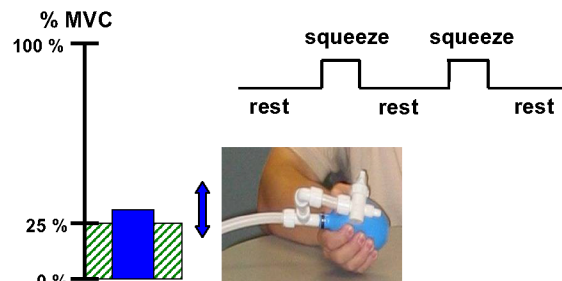


Fig. 1. Bulb-squeezing task

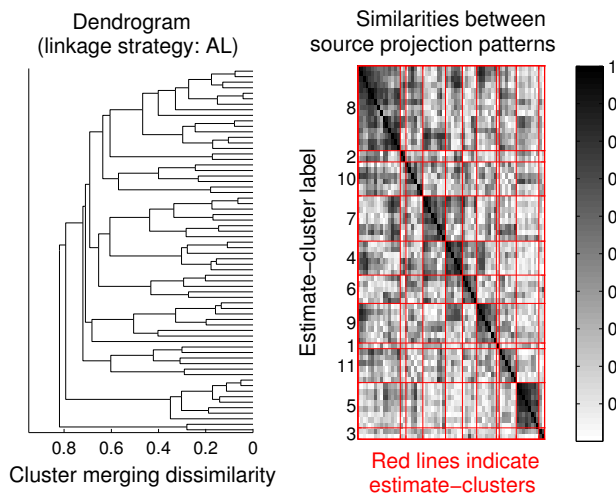


Fig. 2. Dendrogram and similarity matrix

community to perform a right-handed motor task. PD subjects were studied after a minimum of 12 hours withdrawal from L-dopa medication.

During the experiment, subjects were seated 2 m away from a large computer screen. They were asked to squeeze a pressure responsive bulb with their right hand in order to match vertical target bars on the screen that represented 25 % of maximum voluntary contraction (MVC). The task consisted of 8 squeezing trials, where each trial contained 10 seconds of rest period followed by 2 seconds of squeezing (see Fig. 1). Calibration of the squeeze bulb to record MVC for each subject was conducted at the beginning of the experiment.

Brain activity data was collected using an EEG cap (Quick-Cap, Compumedics, Texas, USA) with 19 electrodes using the International 10-20 system, referenced to linked mastoids. EEG data was sampled at 1000Hz using SynAmps2 amplifiers (NeuroScan, Compumedics, Texas, USA). A surface electrode on the tip of the nose was used as ground. Eye movement artifacts were measured using surface electrodes above and below the eyes (Xltek, Ontario, Canada). Data were resampled to 250Hz and band-passed from 1 to 70Hz. Artifacts associated with eye blinks and muscular activity were removed using the Automated Artifact Removal in the EEGLAB Matlab Toolbox [10].

B. Source Separation Results

For each subject, EEG signals recorded during the squeeze periods of 8 trials are concatenated. Furthermore, to reduce computational load and prevent overlearning, data dimension (M) is reduced from 19 to 6 by dividing electrodes into 6 regions and averaging signals within each region. The 6 regions are: Region 1 - (FP1, F7 and F3), Region 2 - (FP2, F4 and F8), Region 3 - (T7, C3, P7, P3), Region 4 - (T8, C4, P8, P4), Region 5 - (Cz, Pz) and Region 6 - (O1, O2). Since averaging is a linear operation, the separability of the underlying source will not be affected. The proposed source separation algorithm is applied to the averaged signals of

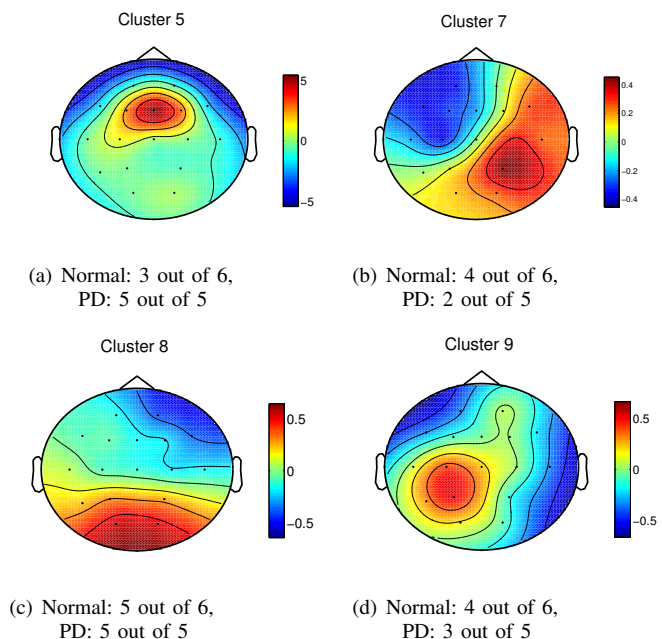


Fig. 3. Source projection patterns corresponding to the centropypes of Cluster 5, 7, 8 and 9. The numbers of normal and PD subjects that contribute to a given cluster are also shown.

these 6 regions, yielding 6 source estimates. To recover the projection patterns from the sources to the 19 electrodes, linear regression is used:

$$\mathbf{x}_i = \mathbf{s}^* \mathbf{c}_i^T + \boldsymbol{\epsilon} \quad \forall i = 1, \dots, 19 \quad (17)$$

where $\mathbf{x}_i \in \mathbb{R}^{N \times 1}$ is a column vector representing the EEG signal from the i -th electrode, $\mathbf{s}^* \in \mathbb{R}^{N \times 6}$ is the time courses of all source estimates and $\mathbf{c}_i \in \mathbb{R}^{1 \times 6}$ is the i -th row of the mixing matrix \mathbf{C} . The resulting \mathbf{C} is of dimension 19×6 .

For each subject, the proposed GmAR source separation algorithm is setup to run 10 times with random initial conditions and the solution with the highest likelihood is used as the final estimates. Clustering is applied to source projection patterns from all normal and PD subjects. We select the number of clusters to be 11. The clustering results including the dendrogram and the similarity matrix are shown in Fig. 2. Cluster 5, 7, 8 and 9 are identified the significant clusters with high inter-subject repeatability. In particular, cluster 8 is found in 10 out of 11 subjects whereas cluster 5, 9 and 7 are found in 8, 7 and 6 subjects, respectively.

The scalp maps corresponding to the centropypes of cluster 5, 7, 8 and 9 are shown in Fig. 3 and are consistent with biological networks previously described. Cluster 8 corresponds to the occipital area which is related to the visual task. Cluster 5 reflects possible error-monitoring processes (often called "Error Related Negativity") in the psychology literature. Cluster 9 is over the left sensorimotor cortex which would be expected to be activated with a right-handed task. Cluster 7 reflects a left frontal/right sensorimotor system previously shown in to be important for rhythmic tasks.

Spatial filters derived using the procedure described in Section II-A.3 are applied to the entire EEG recording (in-

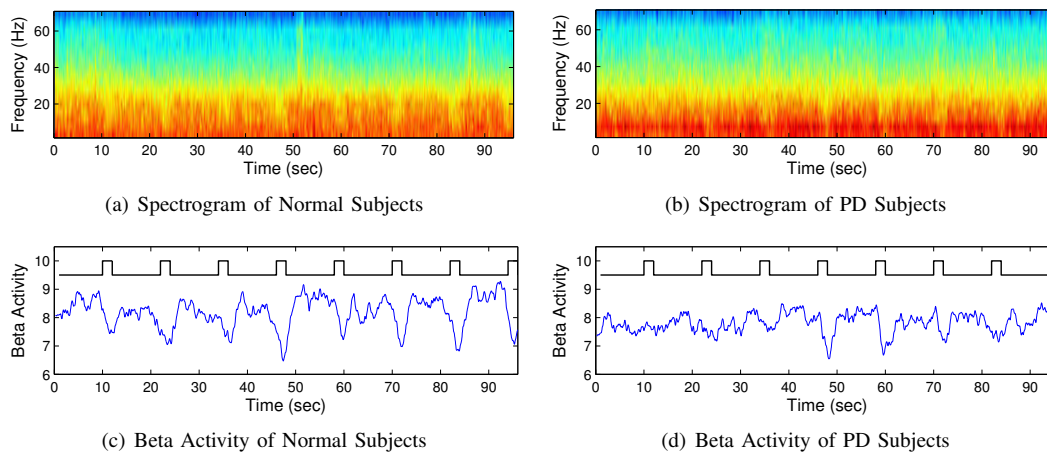


Fig. 4. (a) and (b): Spectrograms of the sources corresponding to Cluster 9. Eight trials are shown. (c) and (d): Beta (15-30Hz) activity. The square wave shows the state of the visual stimuli (low-rest; high-squeeze). Notice the strong attenuation in normal subjects during the squeezing periods.

cluding rest and squeeze periods) of the corresponding subjects. Spectrograms of the estimated sources are computed and averaged across subjects. The averaged spectrograms of the source corresponding to Cluster 9 (the motor area) from normal and PD subjects are shown respectively in Fig. 4(a) and (b). It can be seen that the estimated sources are well modulated by the motor state (rest/squeeze) in normal subjects, whereas this pattern is less prominent in PD subjects. Furthermore, when we average the spectrograms over the beta-band (15-30Hz), we notice a strong attenuation of beta activity in the normal subjects during the squeezing periods, whereas the attenuation in PD subjects is considerably less noticeable. This observation is in line with existing PD studies which reported abnormal beta oscillations during movement in Parkinson's patients [11].

IV. CONCLUSION

We proposed an EEG source separation technique which addressed the issues of volume conduction and the temporal and spatial correlations between underlying EEG sources. Furthermore, clustering technique employed in this study resolved the source ambiguity problem commonly seen in applications of source separation techniques to group analysis. The real EEG results demonstrated that the proposed technique is promising for finding task-related brain sources.

V. ACKNOWLEDGMENTS

This work was partly funded by a Collaborative Health Research Project from NSERC and CIHR awarded to MJM.

REFERENCES

- [1] E. Pereda, R. Quiroga, and J. Bhattacharya, "Nonlinear multivariate analysis of neurophysiological signals," *Progress in Neurobiology*, vol. 77, no. 1-2, pp. 1-37, 2005.
- [2] L. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biological Cybernetics*, vol. 84, no. 6, pp. 463-474, 2001.
- [3] S. Makeig, A. Bell, T. Jung, T. Sejnowski *et al.*, "Independent component analysis of electroencephalographic data," *Advances in neural information processing systems*, pp. 145-151, 1996.
- [4] S. Baillet, J. Mosher, and R. Leahy, "Electromagnetic brain mapping," *IEEE Signal processing magazine*, vol. 18, no. 6, pp. 14-30, 2001.
- [5] Z. Koles, "Trends in EEG source localization," *Electroencephalography and clinical Neurophysiology*, vol. 106, no. 2, pp. 127-137, 1998.
- [6] C. James and C. Hesse, "Independent component analysis for biomedical signals," *Physiol. Meas.*, vol. 26, no. 1, pp. R15-R39, 2005.
- [7] G. Gómez-Herrero, M. Atienza, K. Egiazarian, and J. Cantero, "Measuring directional coupling between EEG sources," *Neuroimage*, vol. 43, no. 3, pp. 497-508, 2008.
- [8] W. Penny, R. Everson, and S. Roberts, "Hidden markov independent component analysis," *Advances in Independent Component Analysis*, pp. 3-22, 2000.
- [9] J. Himberg, A. Hyvärinen, and F. Esposito, "Validating the independent components of neuroimaging time series via clustering and visualization," *Neuroimage*, vol. 22, no. 3, pp. 1214-1222, 2004.
- [10] G. Gomez-Herrero, "Automatic Artifact Removal (AAR) toolbox v1.3 (Release 09.12. 2007) for MATLAB."
- [11] S. Farmer, "Neural rhythms in Parkinson's disease," *Brain*, vol. 125, no. 6, p. 1175, 2002.