34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Selection of Abnormal Neural Oscillation Patterns Associated with Sentence-level Language Disorder in Schizophrenia*

Tingting Xu, Massoud Stephane and Keshab K. Parhi, *Fellow, IEEE*

***Abstract*—Language disorder is one of the core symptoms in schizophrenia. We propose a new framework based on machine intelligence techniques to investigate abnormal neural oscillations related to this impairment. Schizophrenia patients and healthy control subjects were instructed to discriminate semantically and syntactically correct sentences from syntactically correct but semantically incorrect sentences presented visually, and 248-channel MEG signals were recorded with a whole head machine during the task performance. Oscillation patterns were extracted from the MEG recordings in 8 frequency sub-bands throughout sentence processing, which form a large feature set. A two-step feature selection algorithm combining F-score filtering and Support Vector Machine recursive feature elimination (SVM-RFE) was designed to pick out a small subset of features which could discriminate patients and controls with high accuracy. We achieved a 90.48% prediction accuracy based on the selected top features, following the leave-one-out cross validation procedure. These top features provide interpretable spectral, spatial, and temporal information about the electrophysiological basis of sentence processing abnormality in schizophrenia which may help understand the underlying mechanism of this disease.**

I. Introduction

Schizophrenia is a chronic, severe, and disabling mental disorder that has affected people throughout history. Language impairment is recognized as one of the core symptoms associated with this disease. Behavioral research has demonstrated disturbances at multiple levels of language processing (sub-lexical, lexical, sentence and discourse) [1]. The importance of language in schizophrenia is such that this illness may be evolutionarily related to the development of language in Homo sapiens [2]. Consequently, understanding the physiological basis of language disorder in schizophrenia would shed light on the underlying mechanisms of the illness itself.

Earlier research efforts that tried to understand the mechanisms of schizophrenia have focused on relating specific cortical regions to the psychotic symptoms, such as language disorder. However, more recent theories suggest that cognitive dysfunctions are not simply due to a spatially circumscribed deficit, but rather represent a distributed impairment involving many cortical areas and their connectivity [3]. Accordingly, neural oscillations have now become a crucial target for schizophrenia research, due to their role in realizing flexible communication within and between cortical areas.

Studies have shown that cognitive functions modulate neural oscillations at multiple frequencies simultaneously and differentially in terms of their frequency, location, and time of occurrence [4, 5]. Consequently, to get a better understanding of oscillatory activity, the full scale of the spectral, spatial, and temporal dimensions of brain oscillations need to be evaluated. Such multi-dimensional evaluation usually results in a dauntingly large scale data set (combination of number of frequency sub-bands, number of magnetoencephalogram (MEG) / electroencephalography (EEG) channels and number of time points). Machine intelligence techniques are well-suited to perform such comprehensive data analysis. For example, artificial neural network (ANN) [6] and linear discriminant analysis (LDA) [7] have been applied to discriminate schizophrenia patients and healthy control subjects, and achieved high classification accuracy. However, due to the complexity of the classifier, it is still difficult to understand the functional basis of the discrimination.

Motivated by the above considerations, we designed an experiment to investigate abnormal neural oscillations associated with language processing impairment in schizophrenia. More specifically, we aimed at building a classifier which can achieve a high degree of discrimination between patients and healthy controls with a small subset of oscillation patterns that could be understood in terms of brain function. These discriminating features reflect the frequency, brain region and time of oscillations that are abnormal during language processing in schizophrenia, and may help us understand the underlying mechanism of this impairment. The detailed explanation of the experimental paradigm as well as the classification and feature selection strategy are described in Section II. The experimental results and conclusion are presented in Section III and IV, respectively.

II. Method

A. Outline and System Framework

To investigate abnormal brain oscillation associated with language impairment, schizophrenia patients and healthy control subjects were instructed to perform a sentence processing task while multi-channel MEG signals were recorded to detect brain oscillation during task performance. Full scale spectral-spatial-temporal oscillation patterns were extracted from the MEG recordings to construct a rich feature set. A two-step feature selection algorithm based on F-score filtering [8] and SVM-RFE (support vector machine –

* Research is partially supported by the IGERT Neuroengineering program under grant "NSF DGE-1069104".

T. Xu is with the Department of Electrical and Computer Engineering, University of Minnesota, MN 55455 USA (e-mail: xuxxx591@umn.edu).

M. Stephane is with the VA Medical Center, Minneapolis, MN 55417 USA (e-mail: mstephan@ umn.edu).

K. K. Parhi is with the Department of Electrical and Computer Engineering, University of Minnesota, MN 55455 USA (e-mail: parhi@ umn.edu).

recursive feature elimination) backward elimination [9] was then employed to select the most informative features which could discriminate patients and controls with high accuracy. The frequency, space and time information contained in these discriminating features were examined for analyzing the language processing abnormality in schizophrenia. The system framework is shown in Figure 1.

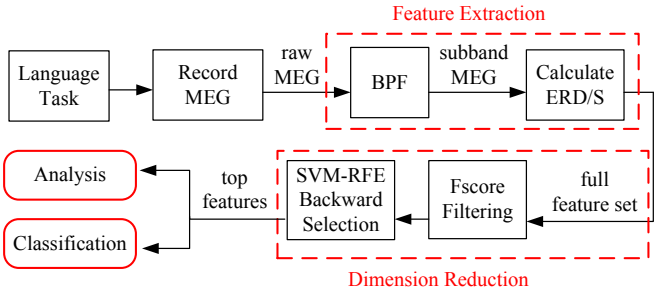

Figure 1. System framework.

B. Subjects and Language Task

Ten schizophrenia patients (10 male) meeting the DSM-IV diagnostic criteria and eleven healthy control subjects (9 male, 1 female) participated in this study. All the subjects were native English speakers and were right-handed. The patient group did not differ significantly ($p<0.05$) from the control group with respect to age, personal or parental level of education, and premorbid overall and verbal intelligence. The experimental protocol was approved by the Minneapolis VA Medical Center and the University of Minnesota Institutional Review Boards.

Subjects were asked to distinguish between correct and incorrect sentence stimuli. Each stimulus is a set of five English words. A correct stimulus is a semantically and syntactically correct sentence (e.g., The boy ate the bagel.) while an incorrect stimulus is a syntactically correct but semantically incorrect sentence (e.g., The bagel ate the boy) . The elements in each stimulus were presented one at a time in the center of a monitor placed in front of the subjects, at a rate of one per second with 10 seconds inter-stimulus interval. The subjects were instructed to read the stimuli silently and press a button with their right index finger for incorrect stimuli.

C. MEG Recordings

During task performance, MEG signals were recorded with a sampling rate of 1KHz, using a 248-detector whole-head neuromagnometer equipped with first order axial gradiometer Magnes 3600 (4-D Neuroimaging, San Diego CA), in a 2-layer mu-metal magnetically shielded room (IMEDCO, Hagendorf, Switzerland). In parallel, Electrocardiogram (ECG) and electrooculogram (EOG) were recorded to identify and correct epochs contaminated by heartbeats and eye movements.

As we were interested in detecting between-group oscillation differences during normal language processing, here we only analyze epochs with correct stimuli. Each epoch was defined to include a baseline period (3 seconds immediately before the first word) and an active period (8.5 seconds after the onset of the first word). The active period includes a 5 seconds "encoding phase" (sentence presentation) and a 3.5 seconds "post-stimuli phase" (after sentence presentation). The timing diagram of the task is shown in Figure 2. After removing heartbeat and eye movement, 40 artifact-free trials for each subject were band pass filtered between 1 to 64Hz and down sampled to 256Hz for further analysis.

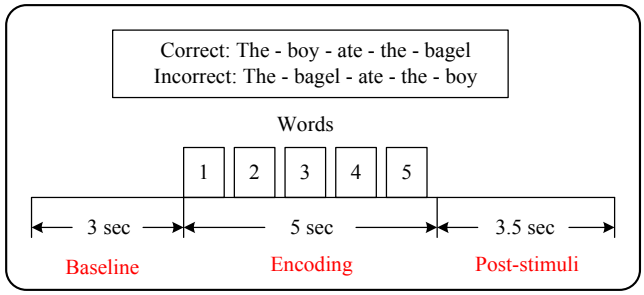

Figure 2. Timing diagram of the semantic processing task

D. Extraction of Oscillation Patterns

Certain events like cognitive tasks can cause frequency specific changes of the ongoing oscillation activity and may lead to either decrease or increase of power in given frequency bands. The former case is called event-related desynchronization (ERD) and the latter event-related synchronization (ERS) [4]. To investigate neural oscillations, we computed ERD/ERS on multiple dimensions including frequency, time and space. As ERD/ERS contain frequency specific behavior, MEG signals were first filtered between 1-48Hz and then decomposed into eight frequency sub-bands using a second order Butterworth filter. The bandwidth of each sub-band was 4Hz for the 1-16Hz range and 8Hz for the 16-48Hz range. For each sub-band, power data were first averaged across all trials and then smoothed using a 250ms window with 125ms overlap. For each 11.5 seconds trial, the total time point was reduced to 92 per channel, including 24 points in the baseline period (3 seconds) and 68 points (8.5 seconds) in the active period. Finally, the ERD/ERS value was calculated as the percentage power change of each smoothed mean power data relative to the mean power within baseline period:

$$ERDS(i) = \frac{A(i) - R}{R} \times 100\% \quad (1)$$

where $A(i)$ is the ith smoothed mean power sample and R is the average power of baseline period, which is calculated as $R = \frac{1}{24}\sum_{i=1}^{24} A(i)$. The reason for using relative power change is to remove the additive effects like medication, coffee and tobacco consumption, as those effects affect both the baseline and the active period.

E. Feature Selection and Classification

Multidimensional evaluation of oscillations result in a very large ERD/ERS feature set: 8 frequency sub-bands*248 MEG channels*92 time points per channel, leading to a total of 182,528 features per subject. Compared to the small sample size (21 subjects) and small number of groups to discriminate (patient group vs. control group), the features are in an

extremely high dimensional space. To avoid spurious group differences, we employed a classification based two-step feature selection algorithm to select a small sub-set of features that have highest discriminating power in patient and control classification. These top discriminating features reflect the frequency, brain region and time of oscillations that are abnormal during semantic processing in schizophrenia, and can help us understand the underlying mechanism of this impairment.

In the two-step feature selection stage, we first used F-score filtering to eliminate large number of "garbage" features. F-score is a simple and generally effective technique to measure the discrimination of two sets of real numbers [8]. Consider n training samples: x_k, $k = 1,\ldots,n$ and let the number of positive and negative samples be n_+ and n_- respectively. Each sample is a vector with m features. The F-score of the ith feature is defined as:

$$F(i) = \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}(x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}(x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2} \quad (2)$$

where $\bar{x}_i$, $\bar{x}_i^{(+)}$, $\bar{x}_i^{(-)}$ represent the averages of the ith feature of the whole, positive, and negative data samples, respectively; $x_{k,i}^{(+)}$ and $x_{k,i}^{(-)}$ are the ith feature of the kth positive and negative sample, respectively. In short, the numerator indicates the discrimination between the positive and negative sets, and the denominator represents the discrimination within each of the two sets. The higher the F-score is, the more likely this feature is discriminative.

After F-score filtering, large number of irrelevant features were eliminated except the top 150 features with the highest F-score were kept for next step feature ranking by SVM-RFE, a classification based feature selection algorithm [9]. Basically, it is a backward selection strategy using the weights of SVM model [10] to produce a feature ranking. Due to the nature of our dataset (small sample size vs. large feature number) as well as computational consideration, we employed linear kernel SVM, in which the weight vector $\mathbf{w}$ is obtained by solving the following quadratic optimization problem:

$$\min_{\boldsymbol{\omega},\xi,b} \quad \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + C\sum_{k=1}^{n}\xi_k, \quad (3)$$

$$\text{s.t.} \quad y_k(\mathbf{w}^T\mathbf{x}_k + b) \geq 1 - \xi_k, \quad \xi_k \geq 0, \quad k = 1,\ldots,n$$

where $\mathbf{x_k}$ is the training vector and $y_k \in \{1, -1\}$ is the corresponding class label; ξ_k is the so called slack variable allowing margin errors and b is a bias term. C is a penalty parameter set by the users to control the tradeoff between margin size (generalization ability of the classifier) and the number of samples inside the slab (training error). To get the optimal C value, the training samples were subdivided into learning set and validation set. We used the learning set to build SVM models with different C values ($\log_2 C \in \{-1, 0, 1, \ldots, 10\}$) and used these models to classify the validation set. The C value associated with the smallest validation error was used to build the final SVM classifier using all the training samples. Then the weight values in the classifier were squared and the feature with smallest weight was removed from the ranking list, based on the idea that the smaller the weight is, the less relevant the feature is. This procedure was repeated after all the features were removed from the list. According to the backward elimination characteristic of SVM-RFE algorithm, the later a feature is removed from the list, the higher its ranking is.

To test the robustness and the generalization ability of the selected discriminating features, a leave-one-out double cross validation procedure [11] was performed. Each time, 20 subjects were used for feature selection and training a SVM classifier based on the selected top features while the other one subject was used for testing the classification result. The testing sample was completely left out before testing and the procedures were repeated until all subjects were classified.

III. Results

In Figure 3, we compare the discriminating power of the top features selected by our two-step feature selection algorithm with the features ranked by F-score only and using SVM-RFE only. The lowest prediction error based on the two-step feature selection algorithm is obtained with 53 top features. As presented in Table 1, a 90.48% overall prediction accuracy can be achieved for all the subjects. The classification accuracies for control group and patient group are 100% (true negative rate) and 80% (true positive rate), respectively.

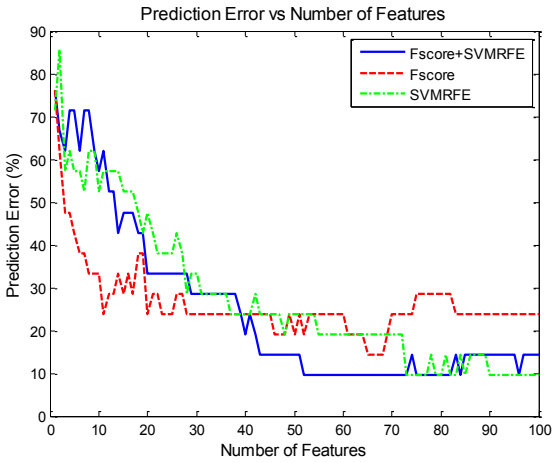

Figure 3. Prediction error rate versus feature number using different algorithms

TABLE I. Prediction Results Using Top 53 Features (F-score+SVMRFE)

Control			Patient			Average	
error	*TN*[a]	*FP*[b]	*error*	*TP*[c]	*FN*[d]	*error*	*accuracy*
0/11	100%	0	2/10	80%	20%	2/21	90.48%

a. True Negative Rate; b. False Positive Rate; c. True Positive Rate; d. False Negative Rate

In comparison, the lowest prediction error based on the top features ranked only by F-score is 14.28% which is higher than the two-step algorithm. This is because F-score does not

reveal mutual information among features, i.e., the power of the combination of the features, which affects its generalization ability. Top features selected by using SVM-RFE directly without pre-selection can also achieve 90.48% overall accuracy but needs more than 70 features. For analysis consideration, a smaller feature number may help better locate the dysfunctional brain region and frequency. In addition, from the view of computation, applying SVM-RFE directly without pre-selection is very time consuming in our case, due to its backward elimination characteristic and the large feature size fed into the algorithm. Thus, we use F-score filtering as a pre-selection step to remove large number of irrelevant features before applying SVM-RFE as a main feature selection step. The combination of the two algorithms not only help achieve a better discrimination with a smaller number of features but also reduce the time and complexity in computation.

To better understand the underlying physiological characteristics of the discriminating features, we next present the time courses of 12 most often selected features and their spatial locations in Fig. 4. Other top features were similarly located, i.e., at adjacent time points, channels and frequency bands. We note that the spatial locations of the top features were not restricted to one specific cortical area, but rather involve several different brain regions, which support many recent theories that emphasize the role of disturbed coordination in the pathophysiology of schizophrenia [3].

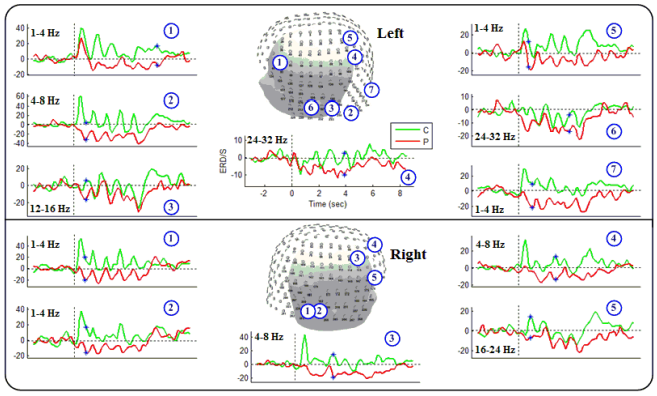

Figure 4. Top discriminant features (* denotes time point that discriminate between groups. Horizontal axis represents time: baseline (-3-0 sec), sentence presentation (0-5sec) and post-stimuli (5-8.5 sec). Vertical axis represents ERD/ERS value.)

During sentence presentation, patients showed reduced delta band (1-4Hz) power at the left parietal-occipital and right temporal areas (ERD as opposed to ERS in controls); reduced theta band (4-8Hz) power at the occipital and right frontal lobes; as well as less alpha (12-16Hz) and beta (16-32 Hz) band ERS power at the left temporal-parietal and right frontal lobes. In addition, patients showed reduced delta band synchrony at the left frontal lobe after sentence presentation. Therefore, schizophrenia patients have reduced synchrony during both sentence presentation and integration periods, which indicate dysfunction during semantic processing as well as failure of neural system to resume idle state.

IV. Conclusion

In this study, we designed an experiment and employed machine intelligence methods to investigate the electrophysiological basis of language processing abnormality in schizophrenia. Specifically, we extracted spectral-spatial-temporal oscillation patterns from 248 MEG channels when subjects performed a sentence processing task. A two-step feature selection algorithm based on F-score filtering and SVM-RFE backward elimination was then performed to select a small subset of features which could discriminate patients and controls with high accuracy. Following the leave-one-out cross validation procedure, we achieved a 90.48% prediction accuracy using 53 top ranked features. These top features provide information about the frequency, location and time course of oscillations that differentiate between schizophrenia patients and healthy controls, which may help understand the underlying mechanism of the language processing abnormalities in schizophrenia. In the future, we will incorporate more feature selection and classification algorithms into the system and test the performance on a larger dataset.

References

[1] M. Stephane, G. Pellizzer, C.R. Fletcher, and K. McClannahan, "Empirical Evaluation of Language Disorder in Schizophrenia", *Rev Psychiatric Neuroscience*, vol.32, pp. 250, 2007.

[2] C.D. Crow, "Is schizophrenia the price Homo sapiens pay for language?", *Schizophr Res*, vol.28, pp.127:141, 1997.

[3] P.J. Uhlhaas and W. Singer, "Abnormal neural oscillations and synchrony in schizophrenia", *Nat Rev Neurosci*, vol.11, pp. 100:113, 2010.

[4] G. Pfurtscheller and F.H. Lopes da Silva, "Functional meaning of event-related desynchronization (ERD) and synchronization (ERS)", *Handbook of Electroencephalography and Clinical Neurophysiology*, Elsevier, Amsterdam, 1999: 51:65.

[5] F.H. Lopes da Silva, "Neural mechanisms underlying brain waves: from the neural membranes to networks", *Electroencephalogr Clin Neurophysiol*, vol.79, pp.81:93, 1991.

[6] Y. Li and F. Fan, "Classification of schizophrenia and depression by EEG with ANN", *IEEE EMBS'06*, pp.2679:2682.

[7] A.P. Georgopoulos, E. Karageorgiou, S.M. Lewis, et al., "Synchronous neural interactions assessed by magnetoencephalography: a functional biomarker for brain disorders", *J Neural Eng*, vol.4, pp. 349:355, 2007.

[8] Y. Chen and C. Lin, "Combining SVMs with various feature selection strategies", *Feature extraction, foundations and applications*, Springer-Verlag, Berlin, 2006.

[9] I. Guyon, J. Weston, S.M.D. Barnhill and V. Vapnik, "Gene Selection for Cancer Classification using Support Vector Machines", *Machine Learning*, vol.46, pp.389:422, 2002.

[10] B.E. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers", *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pp. 144:152, 1992.

[11] V. Cherkassky and F. Mulier, *Learning from data: Concepts, theory, and methods*. 2nd ed. Wiley-IEEE Press, 2007, pp 78-80.