30th Annual International IEEE EMBS Conference
Vancouver, British Columbia, Canada, August 20-24, 2008

Selection of Spectro-Temporal Patterns in Multichannel MEG with Support Vector Machines for Schizophrenia Classification

Nuri F. Ince, Fikri Goksu, Giuseppe Pellizzer, Ahmed Tewfik, Massoud Stephane

Abstract—**We present a new framework for the diagnosis of schizophrenia based on the spectro-temporal patterns selected by a support vector machine from multichannel magnetoencephalogram (MEG) recordings in a verbal working memory task. In the experimental paradigm, five letters appearing sequentially on a screen were memorized by subjects. The letters constituted a word in one condition and a pronounceable nonword in the other. Power changes were extracted as features in frequency subbands of 248 channel MEG data to form a rich feature dictionary. A support vector machine has been used to select a small subset of features with recursive feature elimination technique (SVM-RFE) and the reduced subset was used for classification. We note that the discrimination between patients and controls in the word condition was higher than in the non-word condition (91.8% vs 83.8%). Furthermore, in the word condition, the most discriminant patterns were extracted in delta (1-4 Hz), theta (4-8Hz) and alpha (12-16 Hz) frequency bands. We note that these features were located around the left frontal, left temporal and occipital areas, respectively. Our results indicate that the proposed approach can quantify discriminative neural patterns associated to a functional task in spatial, spectral and temporal domain. Moreover these features provide interpretable information to the medical expert about physiological basis of the illness and can be effectively used as a biometric marker to recognize schizophrenia in clinical practice.**

I. Introduction

SCHIZOPHRENIA is a mental disorder that affects approximately 1% of people over 18 years of age all over the world [1]. This disorder causes multiple impairments in cognitive domains such as attention, memory, executive functions and language. In addition schizophrenia is also associated with symptoms such as auditory hallucinations, delusions and emotional dysregulation. Combined, the cognitive impairments and the symptoms of the disorder, impact the life of patients significantly. To date, the diagnosis of schizophrenia is based only on clinical observations and the patient self-reported experiences. The diagnosis of schizophrenia remains uncertain since there are no biological markers to validate the clinical diagnosis.

Therefore in the last several years there has been an increasing interest in finding biological markers that may enable early diagnosis of schizophrenia as well as other mental illnesses. In particular several research groups are exploring the brain functions by using a variety of noninvasive measures of brain activity such as functional magneto resonance imaging (fMRI) [2, 3], electroencephalography (EEG) [4, 5 and 6], magnetoencephalography (MEG) [7]. Generally the brain activity is monitored under two conditions: (i) Resting state and (ii) Functional state. During the resting state studies, the subjects are asked to fixate their eyes on a cross during the experiments. In contrast during the functional state studies, the subjects are given a warning stimulus or they are engaged with a specific task and related activity patterns are used for further analysis.

Studies that used features reflecting functional brain activity report higher classification accuracy (≈90%) than studies that used features from resting state brain activity (60-80%). Consequently, it is reasonable to expect that studies of brain activity associated with functional impairments of schizophrenia patients will provide more information than resting state studies. However due to the complexity of used classifiers, it is difficult to understand the physiological basis associated with discrimination power.

In this paper we also use the brain activity related to a functional task to discriminate between schizophrenia patients and healthy controls. In addition we quantify extracted patterns in spatial, spectral and temporal domain to provide interpretable information to the medical expert about physiological basis of the illness. This may provide significant contribution in understanding the pathology in brain function. For this particular purpose we obtained the neural activity with MEG during a verbal memory task. Abnormal working memory neural correlates have been frequently reported in this illness [8]. In the following step we extracted spectro-temporo-spatial features from multichannel MEG data. This constituted a quite large predictor space where the sample size is low. Therefore it is difficult to identify relevant patterns that discriminates schizophrenia patients from controls. To handle this problem we use a maximum margin, support vector machine,

Nuri F. Ince is with the departments of Electrical and Computer Engineering and Neuroscience, University of Minnesota MN USA (phone: 612-625-5006; e-mail: firat@umn.edu).

Fikri Goksu is with the department of Electrical and Computer Engineering, University of Minnesota MN USA (e-mail: goks0002@umn.edu).

Giuseppe Pellizzer, is with VA Medical Center, MN USA. (e-mail: pelli001@umn.edu).

Ahmed H. Tewfik is with the department of Electrical and Computer Engineering, University of Minnesota MN USA (phone: 612- 625-6024; fax: (612) 625-4583; e-mail: tewfik@umn.edu).

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

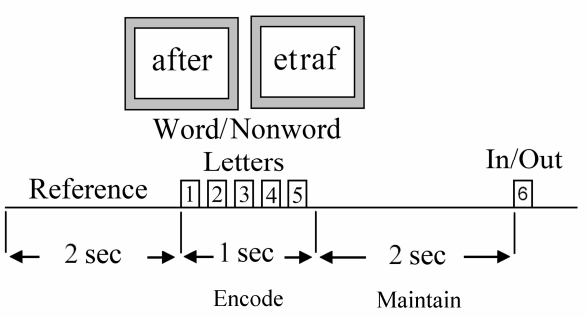

Fig. 1. The timing diagram of working memory paradigm. Five sequential letters appearing on the screen resulted in either a word or a pronounceable nonword. Each sequential letter appeared only for 150 ms on the screen. The neural activity related to "Encode" and "Maintain" stages are used in further steps in classification.

classifier with a recursive feature selection procedure that has been previously used in the area of Genomics. Finally we study the spectral, temporal and spatial content of selected features for discrimination.

This paper is organized as follows. In the next section we describe the verbal memory experimental paradigm and provide a detailed explanation of feature extraction and classification strategy. Finally we provide experimental results and discussion.

II. Materials and Methods

A. Experimental Paradigm and MEG recordings

A modified Sternberg paradigm was used to assess verbal working memory [9]. (See Fig. 1) Five letters were presented sequentially on a computer screen in 1 sec. Let us call this time segment the "encode stage" in the rest of the paper. Each letter appeared for 150 ms followed by a 50 ms blank screen. The letters constituted either a word (50% of the trials) or a pronounceable non-word. Two seconds after the last letter a probe letter appeared on the screen. We will use the term "maintain stage" for this two second time period following the letter presentation. There was a 50% probability that the probe letter was one of the previously shown letters. The subjects were asked to decide whether the probe letter was one of the previously shown letters or not (in/out). The response was provided by a button press (left button for yes and right button for no). Word and non-word, and in/out trials were randomized with equal probability. Sixty trials were recorded for each condition. All subjects, except one control, had correct responses in about 70% of the trials. Those subjects with less than 70% accuracy were rejected from the analyses. The correct trials were analyzed in further steps. For the controls, the numbers of available trials in word and non-word conditions were 52.4±6.2 and 50.6±6.1, respectively. For the patients, the numbers of available trials were 43.3±11.7 and 41.8±11.5, respectively. There was a significant difference between the correct responses in word and non-word condition for controls ($p<0.03$, paired t-test) but not for patients ($p=0.17$, paired t-test). The performance of the controls was significantly better than patients in both word and non-word conditions ($p<0.0039$ and $p<0.0053$) respectively. MEG recordings were obtained from 15 patients (12 males and 3 females) meeting the DSM IV diagnostic criteria for schizophrenia and 23 healthy controls (14 males and 9 females) that were all right handed subjects. MEG data were recorded from 248 axial gradiometers with a 1 KHz sampling frequency during this experimental paradigm (Magnes 3600WH (4D Neuroimaging, San Diego CA). In parallel, electrocardiogram (ECG) and electrooculogram recordings were also obtained for artifact correction. Principal component analysis was used to remove ECG and eye movement artifact from the MEG data [10]. Epochs with residual artifact activity were visually inspected. Then, the successful trials were bandpass filtered between 1-64Hz and down sampled to 250Hz. All artifact rejection and data preprocessing steps were accomplished using BESA™. All trials were exported into ASCII files for further analysis using Matlab™.

B. Extracting Working Memory ERD/ERS Patterns

The brain electrical activity related to a working memory task is well studied in literature by using event related potentials. Since this activity is phase locked to the event, related raw data is averaged over trials. In this paper we use patterns embedded in oscillatory components of brain activity and which are not phase locked to the event. In particular we computed event related desynchronization (ERD) and synchronization (ERS) which are described as short lasting power changes related to a stimulus [11]. In literature ERD and ERS patterns has been used to quantify activated cortical areas. It has been shown that the ERD and ERS patterns have frequency specific behavior. Therefore in this study we used ERD and ERS patterns to quantify activated cortical areas. We compute the ERD and ERS waveforms of several frequency bands of MEG. Let $x(i, j)$ be the power of the jth sample of the ith trial which was obtained first by filtering the MEG in a frequency band of interest and then squaring the subband activity. The ERDS is denoted as the percentage of power change with respect to a reference interval. The reference R was calculated as the average power of a two second interval before the onset of letters in our experimental paradigm. The ERD/ERS for each sample point was then calculated as

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

where and $A(j)$ is the mean power at the j^{th} sample. Here a negative percentage value indicates ERD and a positive one ERS. Rather than focusing on a fixed frequency interval, the MEG activity was decomposed into 6 different frequency subbands by using a second order Butterworth filter. The bandwidth of each subband was set to 4Hz in 1-16Hz range and 8Hz in 16-32Hz. Prior to obtaining the percentage values in each band, the power data were smoothed with a

250ms long Gaussian kernel. Then the mean power was estimated in 250ms windows with 125ms overlaps to reduce the number of time points for the *ERDS* calculation.

C. Recursive Feature Elimination and Classification with Support Vector Machines

The ERDS waveforms are computed for all 248 MEG channels and six frequency subbands. Each waveform contained 24 time points starting from the onset of the letter presentation and extending to the end of memorization stage. This constituted a predictor space with a dimension N=35712 per condition. Obviously this is a very high dimensional space and it is difficult to use all features for classification due to overlearning problem in a small sample size. Similar problems do also exists in the area of Genomics where the expression profiles of a large number of genes measured on a limited sample space where one is interested into those genes that discriminates between two groups [12]. Recently this problem has been successfully tackled by a recursive feature elimination strategy by using a support vector machine classifier (SVM-RFE) for relevant gene selection in cancer classification [12]. The support vector machine classifier searches for the maximum margin that is estimated from to the training patterns, support vectors, which are most difficult to classify. The resulting classifier has higher generalization capacity in higher dimensions which is the result of its large margin. In linear SVM the maximum margin is defined as $\frac{2}{\|v\|^2}$ where the margin is maximized by determining the weight vector according to

$$\min \frac{1}{2}\|v\|^2 + C\sum_j \zeta_j^2 \text{ subject to}$$
$$y_j(v^T x_j + b) \geq 1 - \zeta_j \quad \text{and}$$
$$\zeta_j \geq 0 .$$

Here $x \in \Re^k$ is the feature vector of a sample with label $y \in \{-1,1\}$, v is the weight vector and b is the bias term ζ_j is the slack variable allowing margin errors, and $\boldsymbol{C}$ is a parameter set by the user that controls the trade off between the size of the margin and the number of support vectors inside the slab. In the optimization problem above the objective function is a quadratic function of the variables v and ζ and the constraints are linear function of the variables.

As defined above the maximum margin is defined by as $\frac{2}{\|v\|^2}$. Obviously the margin is defined from weight magnitude. In SVM-RFE framework the maximum margin is initially estimated by training the classifier with all features in the training set. This corresponds to the maximum margin level that can be achieved by the classifier. In the following steps the weight values are squared and those ones that are smaller than a threshold are recursively eliminated. This step corresponds discarding feature indices that have minimal contribution on the margin definition. This procedure is iterated to select a subset of features. In the meanwhile the size of the margin is reduced with each elimination step. This algorithm has been successfully used in detection of biomarkers for cancer classification by inspecting the expression profiles of thousand of genes.

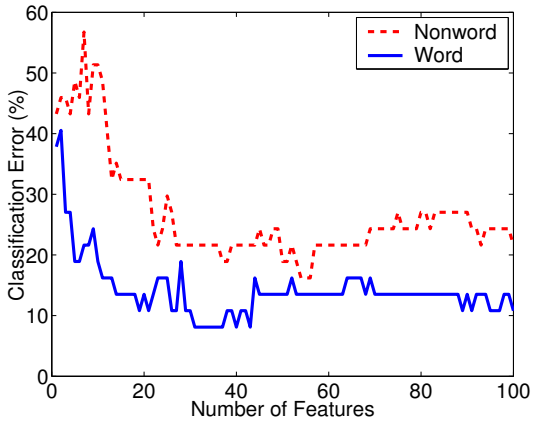

Fig. 2. The classification error rate versus the number of features in Word and Nonword conditions. Note that the system provided lower error rates in Word condition by using smaller number of features than Nonword case ($\boldsymbol{C}$=1).

In our study we used SVM-RFE method to select a subset of spectro-temporo-spatial patterns for discrimination between controls and schizophrenia patients. Due to high dimensionality and to speed the feature selection procedure we eliminated more than one feature in each step as suggested in [12]. In particular we removed 1000 features initially and slowly decreased this number to one feature at a time.

III. Results

We conducted several experiments to asses the effect of functional activity in classification accuracy. A leave one subject out method was used to asses the classification performance of the proposed system. Interestingly we observed that the best classification result (91.9%) was obtained in Word condition. The classification accuracy in Nonword condition was 83.8%. The classification error rate versus the number of selected features is given in Fig. 2. In Nonword condition the best classification accuracy was obtained with 53 features. We note that the SVM classifier provided the best classification rate with 31 features in Word condition. In particular the recognition accuracy was 93.3% for schizophrenia patients and 90.1% for controls. In order to understand better the underlying physiological characteristic of the selected features, we extracted their spatial and spectral content and visualized in Fig. 3 by using

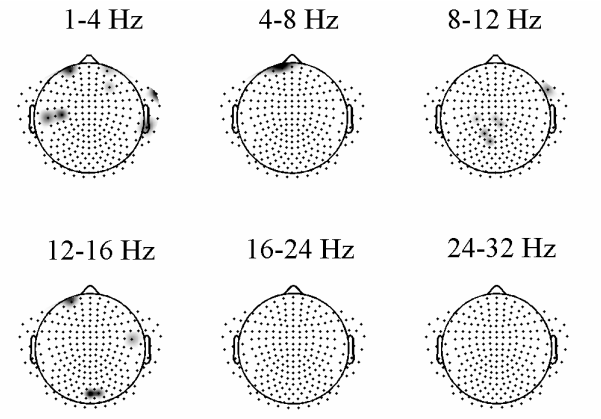

Fig. 3. Subband specific two dimensional head topographies of selected features by SVM-RFE in Word condition. The darker regions indicate the more features was selected from these positions. Note that most of the features were localized in left frontal, occipital and left temporal areas.

EEGLab [13]. We note that the classifier selected most of the features from left frontal, occipital and left temporal areas. The features selected from left frontal area were located in theta (4-8 Hz) and delta (1-4 Hz) bands. In contrast the features in occipital regions were located in alpha band (12-16Hz). We note distinct cortical areas and subbands were selected by the algorithm. In our previous study which was on a smaller subject database, we identified the most discriminant patterns in the left temporal and parietal are in beta frequency band by using a decision tree classifier [14]. We note that the SVM-RFE algorithm has selected not only different frequency bands but also different channel locations.

In SVM-RFE method the ***C*** parameter has a critical role in defining the margin. We conducted several experiments to study the effect of ***C*** parameter in classification accuracy. The Table-1 shows obtained results with ***C***= {0.1, 1, 10, 100}. We note that better results were obtained with smaller values of the ***C*** parameter which results in a larger margin construction.

IV. Conclusion

In this paper we investigated the use of SVM-RFE method on to find biomarkers for the classification of schizophrenia. In particular we extracted spectro-temporal patterns from multichannel MEG recordings in a verbal memory task and selected a small subset of features with SVM-FRE algorithm for classification. During the experiment the subjects memorized five sequentially presented letters which constitute a word in one condition and a pronounceable nonword in the other. We note that the classifier provided the highest classification accuracy 91.8% in word condition versus 82% in nonword condition. We observed that the selected features are located in left prefrontal, occipital and left temporal areas. Our results indicate that the SVM-RFE algorithm can successfully select features from a large predictor space associated to neural activity in a functional task and that these features can be used effectively in recognizing patients in schizophrenia.

TABLE I
Classification Accuracies for Different Values of ***C*** Parameter

C	Word	Nonword
0.1	91.9	75.5
1	**91.9**	**83.8**
10	91.9	81.1
100	86.5	75.7

Acknowledgment

This work was supported by grants from the VA Medical Center (MS, GP), the mind Institute (MS, AL) and the William and Martha Muska Foundation (MS).

References

[1] http://www.nimh.nih.gov/
[2] M.J. Jafri, V.D. Calhoun, "Functional Classification of Schizophrenia Using Feed Forward Neural Networks," Proceedings of 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS '06, pp. 6631 – 6634, 2006.
[3] V. Calhoun, G. Pearlson and K. Kiehl, " Automated Classification of bipolar, schizophrenic and healthy individuals via multiple spatial ICA functional brain 'modes'," Proceedings of the 16th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing, pp. 371 – 376, Sept. 2006.
[4] M.R. Saatchi, B.W. Jervis, "PC-based integrated system developed to diagnose specific brain disorders", Computing & Control Engineering Journal, vol. 2, pp. 61 – 68, March 1991.
[5] B.W. Jervis, T. Roberts, et. al "Detection of brain conditions from evoked responses using artificial neural networks," IEE Colloquium on Artificial Intelligence Methods for Biomedical Data Processing, pp. 9/1 - 9/8, 26 Apr 1996.
[6] Y. Li, F. Fan, "Classification of Schizophrenia and Depression by EEG with ANN," Proceedings of 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS '06, pp.2679-2682, September 2006.
[7] A. P. Georgopoulos et al., Synchronous neural interactions assessed by magnetoencephalography: a functional biomarker for brain disorders, J. Neural Eng. 4 (2007) 349–355
[8] J.D. Cohen, W.M. Perlstein, T.S. Braver, L.E. Nystrom, D.C. Noll, J. Jonides et al. "Temporal dynamics of brain activation during a working memory task". Nature 1997; 386(6625):604-608.
[9] S. Sternberg. "High-speed processing in human memory". Science. 153, 652-654, 1966.
[10] N. Ille, P. Berg, M. Scherg "Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies". Journal of Clinical Neurophysiology. 2002;19:113-124.
[11] C. Neuper, G. Pfurtscheller, "Event-related dynamics of cortical rhytms: frequency-specific features and functional correlates," Inter. Jour. of Psychophys., 43, pp. 41-58, 2001.
[12] I. Guyon, J. Weston, S.M.D. Barnhill, V. Vapnik, "Gene Selection for Cancer Classification using Support Vector Machines". Machine Learning. 2002;46:389–422.
[13] A. Delorme, and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial dynamics including indenpendent component analysis," Journal of Neuroscience Methods, vol. 134, pp. 9-21, 2004. Available at: http://www.sccn.ucsd.edu/eeglab/
[14] N. F. Ince, M. Stephane, A. H. Tewfik, G. Pellizzer, K. McClannahan, "Schizophrenia Classification using Working Memory MEG ERD/ERS Patterns", 3rd Int. IEEE EMBS Conference on Neural Engineering, 2-5 May 2007, Kohala Coast, Hawaii, USA.