Proceedings of the 3rd International
IEEE EMBS Conference on Neural Engineering
Kohala Coast, Hawaii, USA, May 2-5, 2007

Schizophrenia Classification using Working Memory MEG ERD/ERS Patterns

Nuri F. Ince, Massoud Stephane, Ahmed H. Tewfik, Giuseppe Pellizzer, Kate McClannahan

Abstract—**In this paper we investigate the use of event related desynchronization (ERD) and synchronization (ERS) patterns extracted from magnetoencephalogram (MEG) in a working memory task to discriminate between controls and patients with schizophrenia. In the experimental paradigm, sequential letters appearing on a screen are memorized by subjects. In one of two conditions the letters constituted a word. The ERD and ERS patterns are extracted in the theta, alpha, beta and gamma bands from 248 electrode locations covering the whole head. We noticed that most of the ERD patterns are localized on the left frontotemporal area in both word and nonword conditions in the late memorization stage. The beta band showed the most significant difference in this cortical area between controls and schizophrenia patients. By using a decision tree, 94.7% and 87.5% classification accuracy was obtained for controls and patients individually in both word and nonword conditions. Furthermore, we report that on the left frontotemporal lobe, the discrimination within the beta band between patients and controls in the word condition was higher than in the nonword condition. The higher discrimination within the word condition can be linked to the abnormalities in language processing in schizophrenia patients. Our results show that the ERD/ERS patterns extracted from MEG can be successfully used in patient-control discrimination with appropriate adjustment of spatial, spectral, temporal and functional process parameters.**

I. Introduction

SCHIZOPHRENIA is a chronic, severe, and disabling mental disorder. According to the National Mental Institute of Health (NIMH), this disorder affects approximately 1% of people all over the world [1]. People with schizophrenia generally suffer from cognitive impairments such as difficulties in attention, memory, planning, organization and language. Furthermore, schizophrenia patients experience auditory hallucinations and paranoia. The diagnosis of schizophrenia is based on the self-reported experiences of the patient, in combination with secondary signs observed by a psychiatrist or other clinician. As indicated in [2], the nature of an illness in medicine is often determined by examining behavioral and physiological markers. In the case of mental illnesses, the diagnosis is difficult since the markers mostly appear as behavioral patterns. To date, no medical test for schizophrenia exists. Therefore, finding biological markers for schizophrenia carries significant importance for both diagnosis and therapy.

Several methods have been applied in the past to discriminate between mental illnesses, including schizophrenia. Features extracted from different modalities such as electroencephalogram (EEG) [3, 4, 5], functional magnetic resonance imaging (fMRI) [2, 6] and magnetoencephalogram (MEG) are commonly used to delineate between illnesses. The resting state activity of the brain is used by many research groups [2, 5]. Evoked potentials recorded by auditory stimulation is preferred as another strategy for feature extraction in EEG based studies as it conveys information about abnormal brain functions [3, 4, 6]. Many of these systems used multichannel EEG or MEG data and fed the features to complicated classifiers such as Neural Networks (NN) [2-6] and Fuzzy ARTMAP [4] for final discrimination. Neural networks are frequently used in patient and control classification with the belief that the system can represent complex discriminant information in the hidden layers and nonlinear activation functions of neurons. Although high accuracies are reported, such systems make it difficult to see a functional relationship between multichannel brain activity and the results reported. Also, the resting state of the brain can make the classification task difficult since the impaired functions are not stimulated.

In this paper we use another strategy to achieve patient control discrimination. Specifically we use working memory MEG patterns. Our proposed approach relies on the extraction of ERD and ERS patterns from multichannel MEG recordings and uses these features from activated cortical areas for discrimination. The paper is organized as follows. In the next section we describe the experimental paradigm. Then in section III we describe the features extracted from multichannel MEG recordings and finally provide experimental results from patients and controls.

II. Experimental Paradigm and MEG Recordings

In order to asses the dynamic working memory performance of the subjects, an experimental paradigm shown in Fig. 1 was designed. Five letters were shown to the subjects on a computer screen. Each letter appeared for 150ms followed by a 50 ms blank screen. The letters are

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

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

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

Kate McClannahan, was with VA Medical Center, MN USA. (e-mail: kmccl@u.washington.edu)

1-4244-0792-3/07/$20.00©2007 IEEE.

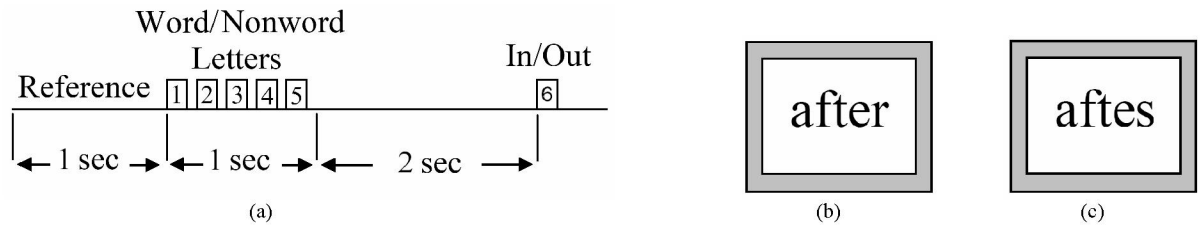


Fig.1. The timing diagram of working memory paradigm (a). 5 sequential letters on the screen resulted in either a word (b) or a nonword (c). Here five letters are visualized for only demonstration purposes. In the experimental paradigm each sequential letter appeared only for 150 ms on the screen.

organized in such a way that the appearance of the last letter results in either a word or a nonword when all five letters are combined. The two second region, following the letters, is referred to as the memorization segment. Two seconds after the last letter a probe letter appeared on the screen. The subject is asked to press a button if this letter is one of the previously shown letters. Word and nonword conditions are randomized with equal probability. Sixty trials were recorded for each condition. Subjects performing correct responses with less than 70% accuracy in the experiment were excluded from the database. During this experimental paradigm, 248 channel MEG data was recorded from 19 controls and 8 schizophrenia patients with a 1 KHz sampling frequency using whole-head neuromagnometer, Magnes 3600 (4-D Neuroimaging, San Diego CA). In parallel, electrocardiogram (EKG) and electrooculogram recordings were also obtained for artifact rejection. Principal component analysis was used to remove EKG and eye movement artifact from the MEG data. Then, the successful trials were bandpass filtered between 2-64Hz and down sampled to 256Hz. All artifact rejection and data preprocessing steps were accomplished in BESA™. All trials were exported into ASCII files for further analysis in Matlab™.

III. Working Memory ERD/ERS Patterns

The oscillatory components of the electrical activity of the brain can be modulated due to an event in a time locked manner. First, Berger reported such changes as the decrease of alpha (α) activity upon opening eyes. When the event causes short lasting amplitude decrease in the rhythmic activity, it is called "Event Related Desynchronization" (ERD) [7]. Conversely, due to an event, an amplitude increase in the rhythmic activity is called "Event Related Synchronization" (ERS) [7]. It is assumed that the ERD and ERS reflect the activation of the underlying neural circuit in the measurement space. When the neural circuit is activated the synchrony between neurons is decreased and this is reflected as ERD. In the opposite case, when the neural circuit is deactivated, the neurons start to have a coherent activity, which in turn induces ERS [7]. In this study we used the ERD and ERS waveforms of several frequency bands of MEG to quantify the activated cortical areas. Let $x(i, j)$ be the power of the jth sample of the ith trial which is obtained by filtering the MEG in a frequency band of interest and then squaring the resulting values. Let *ERD(p)* denote the percentage ERD with respect to a reference. Here *p* stands for percentage. The reference *R* is calculated as the average power of an interval before the trigger signal. The interval should be at least a few seconds in order to satisfy the statistical stability. We calculate *ERD(p)* as follows. Compute

$$R = \frac{1}{k}\sum_{j=n_0}^{n_0+k} A(j) \qquad (1)$$

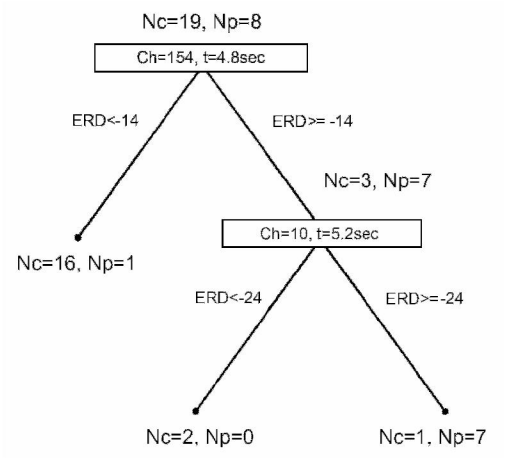
where R is the average power, n_0 is the start index of the reference interval, k is the length of the reference interval and $A(j)$ is the mean power at the jth sample which is obtained from $x(i, j)$ averaging over trials. The ERD for each sample point is then calculated as

$$ERD(j) = \frac{A(j)-R}{R} \times 100\%. \qquad (2)$$

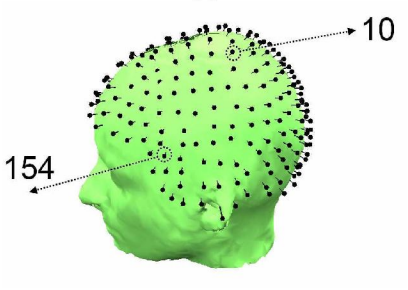
Here a negative percentage value indicates ERD and a positive one ERS. We used a second order Butterworth filter to estimate the *ERD(p)* values in the theta (4-8 Hz), alpha (8-12 Hz), beta (16-24 Hz) and gamma (32-48Hz) bands. Prior to obtaining the percentage values, we filtered the band power data with a 32 point long Gaussian kernel for smoothing. Then we estimated the mean power in 250ms windows with 125ms overlaps to reduce the number of time points for the *ERD(p)* calculation.

IV. Discrimination Between Controls and Schizophrenia Patients

In order to discriminate patients from controls, we use a decision tree (DT) as a classifier [8]. The decision tree uses a 'divide and conquer' strategy for classification. In each level it uses a cost function to evaluate the discrimination power for a given feature. In each node, a single feature is used to split the dataset into child nodes such that in the child nodes the overlap between classes is reduced. One of the reasons for using the DT is that the selected features and classification strategy can be easily understood and may uncover the relationships within the data. As previously indicated, the cost function to evaluate the effectives of a split is critical. Here we selected the Gini Impurity criterion for generating the tree and evaluating the features.



(a)



(b)

Fig. 2. (a) The structure of the decision tree showing the patterns selected in each node. Nc and Np stands for the number of patients and controls in each Node. (b) The electrode locations selected by DT. Note that the electrode location 154 was selected from the left frontotemporal area as a primary feature.

The Gini Impurity (GI) function is defined as:

$$GI(N) = \sum_{l \neq k} P(w_l) * P(w_k) \qquad (3)$$

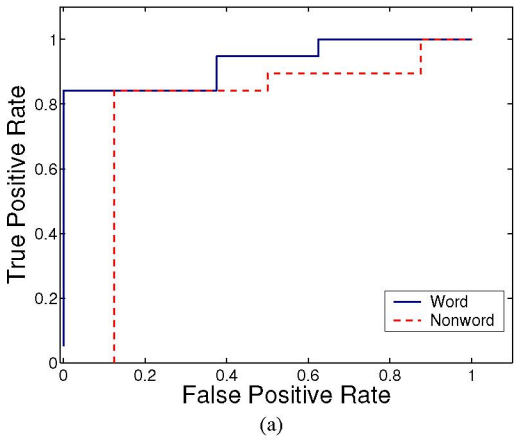
where $P(w_j)$, $P(w_k)$ are the fraction of a class j,k at node N. The efficiency of a split is estimated by evaluating the impurity gain dG,

$$dG = GI(N) - [GI(N_L) + GI(N_R)] \qquad (4)$$

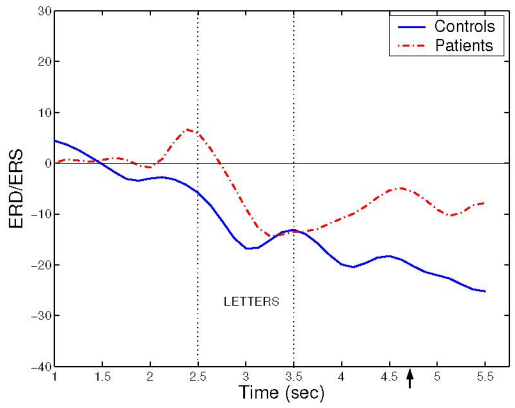
where $GI(N_L), GI(N_R)$ are the impurities in the left and right child branches of node N respectively.

V. Results

We used the one subject leave out method to estimate classification performance. In each step we use one subject for testing and the remaining subjects for training the DT classifier. Separately, for the word and nonword conditions, we used the features individually from the theta, alpha and beta frequency bands to see the effect of the functional activity and the physiological bands in classification performance. We observed that the algorithm achieved the highest classification performance in the beta frequency band in both word and nonword conditions. Specifically 94.7% of the controls were correctly classified. One subject (5.3% of the controls) was recognized as schizophrenic by



(a)



(b)

Fig. 3. (a) The ROC curves of most discriminant time point of channel 154 on the left temporal area in Word and Nonword condition. Note that the discrimination between patients and controls are higher for Word condition. At bottom (b) the ERD/ERS curves of the same electrode from controls and patients in Word condition. The maximum discrimination point is marked with arrow on the time axis.

the DT. For the patients, the true positive and false negative accuracy were 87.5% and 12.5% respectively. In both word and nonword condition the same classification accuracies were obtained. The DT robustly selected the same channels and time location while achieving minimal error. The structure of the classifier, selected features and thresholds are given in Fig. 2 (a). In Fig.2 (b) we visualize the electrode locations selected by DT. We noticed that DT always selected channel 154 in the top level as a primary feature. In the second level electrode 10 is selected. The electrode 154 is located in the left frontotemporal lobe around **Brocca's area**. This area is known as the specialized cortical region responsible for language production. The receiver operating characteristics (ROC) curve for this particular electrode location within word and nonword conditions and their related ERD curves for patients and controls are given in Fig. 3. The ROC curve shows that a simple threshold can help discriminate between controls and patients with around 85% accuracy in channel 154. Although the area under the ROC curve (AUC) is higher for the word condition, in testing the DT, both conditions resulted in the same accuracy percentages. This can be due to the lack of enough training data for better generalization. The location of the

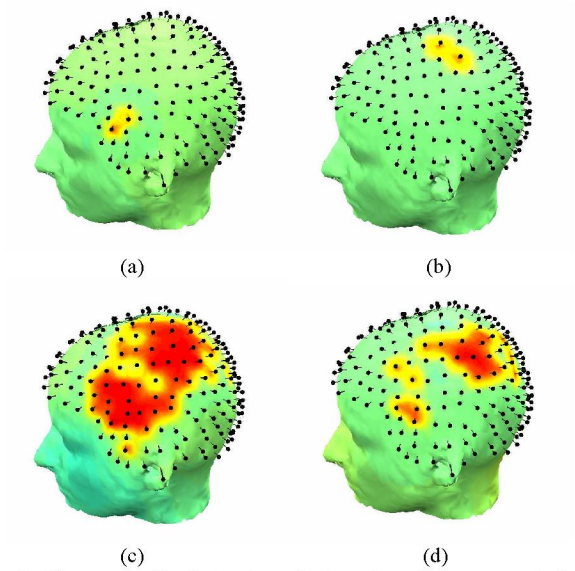


Fig. 4. The most discriminant cortical regions between control and schizophrenia patients in theta (a), alpha (b), beta (c) bands for word condition. Note that the most discriminant cortical region is located on left temporal area in beta band. The most discriminative cortical regions in beta band for nonword condition are visualized in (d). Note that the left temporal area is not as discriminative as in word condition.

most discriminative time point is shown with an arrow in Fig.3.b. The controls showed significant lower ERD than patients at this location.

In order to understand the underlying reason why DT has selected these electrode locations, we investigated the ERD and ERS values on all electrode locations by using the student t-test. The significance tests were implemented within 3 second intervals following the onset of the first letter. False Discovery Rate was used to eliminate multiple comparison errors at the $p<0.05$ level. We noticed that for controls, the significant ERD/ERS patterns were located in the beta frequency band in the late memorization stage in the left frontal and parietal temporal lobe. No significant activation was observed across patients in this area.

Furthermore, in order to asses the performance of each MEG channel in discrimination, we calculated the AUC between controls and patients. For each channel and frequency band, we implemented a search along the time axis to find the maximum AUC value. This enabled us to quantify the maximum discrimination power of each physiological band for a given cortical location in the whole experiment. Fig. 4 shows the AUC values on 3D head maps for several frequency bands for the word and nonword conditions. We used with EEGLAB toolbox to create 3D-headmaps [9]. We observed that the most discriminative cortical areas were located within the beta band. Within this band we noticed that there exist two clusters for channels located in the left frontal and parietal temporal lobes. We note that in the classification, the DT selected two representative electrodes from these two clusters to discriminate between controls and patients. Furthermore, we noticed that the ERD patterns in the left frontotemporal area provided the largest discrimination in the word condition. The AUC values in the nonword condition in the left frontotemporal cluster was significantly lower than the patients ($p<0.00000053$, for 17 electrodes). Although the DT achieved the same accuracy in both conditions, discriminant 3D maps show that the word condition generated larger discrimination on this cortical region. Our results strongly suggest that functional activity carried out by a subject affects the discriminant cortical areas.

VI. Conclusion

In this paper we investigated the use of ERD ERS patterns related to a working memory task for schizophrenia control discrimination. We noticed that most of the ERD patterns are localized on the left frontal and parietal temporal areas in both word and nonword conditions in the memorization stage. The beta band showed the most significant difference in these cortical areas between controls and schizophrenia patients. By using a decision tree, 94.7% and 87.5% classification accuracies were obtained for controls and patients in both word and nonword condition. Furthermore, by using the AUC values, we noticed that the discrimination within the beta band between patients and controls in the word condition was significantly higher than in the nonword condition in the left frontotemporal area. The higher accuracy within the word condition can be linked to the abnormalities in language processing in schizophrenia patients. Our results show that the working memory ERD/ERS patterns extracted from MEG can be successfully used in patient-control discrimination with appropriate adjustment of spatial, spectral, temporal and functional process parameters.

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] 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.

[4] B.W. Jervis, T. Roberts, etal; "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.

[5] 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.

[6] 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.

[7] 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.

[8] L. Breiman, J. H. Friedman, etal,"Classification and Regression Trees," Wadsworth International Group, Belmont, California, (1984).

[9] Delorme, A. and Makeig, S. "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/