Learning EEG-based Spectral-Spatial Patterns for Attention Level Measurement

Brahim Hamadicharef, Haihong Zhang, Cuntai Guan, Chuanchu Wang, Kok Soon Phua, Keng Peng Tee, Kai Keng Ang

Learning EEG-based Spectral-Spatial Patterns for Attention Level Measurement

Brahim Hamadicharef, Haihong Zhang, Cuntai Guan, Chuanchu Wang
Kok Soon Phua, Keng Peng Tee, Kai Keng Ang

Institute for Infocomm Research, 1 Fusionopolis Way, #21-01 Connexis, Singapore 138632
Email: {bhamadi, hhzhang, ctguan, ccwang, ksphua, kptee, kkang}@i2r.a-star.edu.sg

Abstract— **In our every day life, our brain is constantly processing information and paying attention, reacting accordingly, to all sorts of sensory inputs (auditory, visual, etc.). In some cases, there is a need to accurately measure a person's level of attention to monitor a sportsman performance, to detect Attention Deficit Hyperactivity Disorder (ADHD) in children, to evaluate the effectiveness of neuro-feedback treatment, etc.**

In this paper we propose a novel approach to extract, select and learn spectral-spatial patterns from electroencephalogram (EEG) recordings. Our approach improves over prior-art methods that was, typically, only concerned with power of specific EEG rhythms from few individual channels. In this new approach, spectral-spatial features from multichannel EEG are extracted by a two filtering stages: a filter-bank (FB) and common spatial patterns (CSP) filters. The most important features are selected by a Mutual Information (MI) based feature selection procedure and then classified using Fisher linear discriminant (FLD). The outcome is a measure of the attention level.

An experimental study was conducted with 5 healthy young male subjects with their EEG recorded in various attention and non-attention conditions (opened eyes, closed eyes, reading, counting, relaxing, etc.). EEGs were used to train and evaluate the model using 4x4fold cross-validation procedure. Results indicate that the new proposed approach outperforms the prior-art methods and can achieve up to 89.4% classification accuracy rate (with an average improvement of up to 16%). We demonstrate its application with a two-players attention-based racing car computer game.

I. Introduction

In our every day life, our brain is constantly processing information and paying attention, reacting accordingly, to all sorts of sensory inputs (auditory, visual, etc.). There is a need to accurately measure someone's level of attention for detection, performance, monitoring, in medicine, sports, gaming, driving, etc.

A large number of studies have reported clear association between attention or relevant mental conditions and spectral features of the EEG. For example, in [1] the authors applied principal component analysis to full EEG log spectrum and used sub-space features to estimate a local error rate in a sustained attention task. Many other studies focused on specific EEG rhythms. In [2], alpha band activities and coherence time courses were examined in an attention study. In [3], experimental results also indicated correlates between alpha power dynamics and errors by subjects in sustained attention tasks. In [4], neuro-cognitive activity during a self-paced visuospatial task were compared from EEG profiles of marksmen and novice shooters. Furthermore, many studies suggest that self-regulation of theta/beta reduces Attention Deficit Hyperactivity Disorder (ADHD) symptoms [5] [6] and the basis of neuro-feedback therapy [7] [8] [9].

Rather than limiting potential useful features to only specific EEG rhythms, we propose to identify and exploit joint spectral-spatial patterns. We recognized the importance and usefulness of spatial filters from other EEG research studies, for e.g. in Brain-Computer Interface (BCI) to improve spatial resolution or for discriminating motor intentions [10].

In this paper we propose a novel approach that combines advanced signal processing and machine learning (ML) to overcome the limitations of the current prior-art methods. Specifically, we use filter banks to cover a broad range of EEG rhythms, together with common spatial pattern (CSP) filtering, that makes it possible to extract (and tailor) patterns in the EEG for discriminating different EEG conditions (i.e. attention and non-attention). Because of the large number of possible features, we use a MI-based feature selection procedure to reduce them. Notice that the Filter Bank Common Spatial Pattern (FBCSP) was first proposed in [11], but to the best of our knowledge, this is the first time CSP is used in attention detection.

An experimental study was conducted with 5 young healthy male subjects in various attention and non-attention conditions. Results using the novel approach indicate significant improvement in accuracy and outperforming the prior-art methods based on alpha, beta, and theta EEG rhythms [6] [12].

One important advantage of using machine learning in our approach is that it can be learn and be tuned for subject-specific performance. The final model can be trained from a relatively small amount of EEG data, stored, reused at later stage, retrained, etc. Such flexibility is important when, for e.g., continuously monitoring an athlete's performance over the course of training program. Models can also be trained on subject-specific EEG recording but also from a pool of EEG recording of a group of subjects.

The remainder of the paper is organized as follows. In Section II, the methods are introduced. In Section III, we present an the EEG data from an experimental study carried out to evaluate the novel approach. Results are presented in Section IV. Finally, we conclude the paper in Section V.

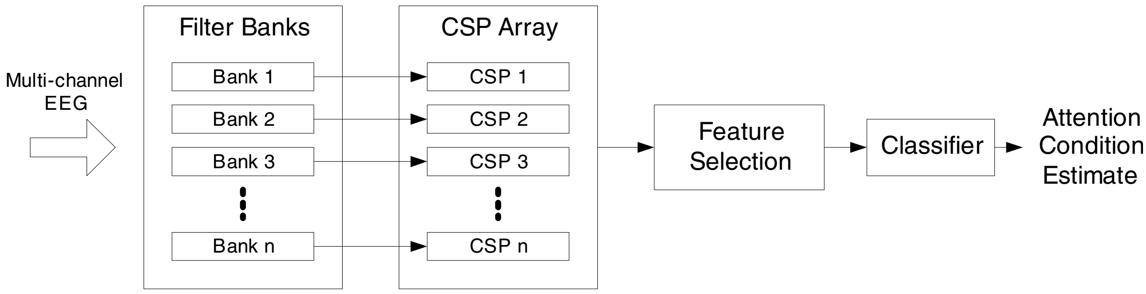

Fig. 1. Schematic flowchart of attention monitoring

II. Methods

The overall flowchart of the novel approach is shown in Figure 1. It consists of four stages: a multiple bandpass filtering stage with Filter-Bank (FB), a spatial filtering stage with common spatial pattern (CSP) algorithm, a feature selection procedure and finally a classifier. The classifier outcome is a measure of attention level.

A. Filter-Bank (FB)

The filter-bank (FB), front-end on the system, consists of 9 bandpass IIR filters (Chebyshev Type II) with equally spaced central frequency (4Hz, 8Hz, 12Hz, 16Hz, 20Hz, 24Hz, 28Hz, 32Hz, 36Hz), each with an 8Hz bandwidth, a stop-band attenuation of 40db, and a pass-band ripple of 3db.

B. Common Spatial Pattern (CSP)

In the present context, the CSP algorithm [10] explores spatial patterns of brain rhythm modulations and will discriminate between two classes of EEG patterns, of class O_1 and O_2, to obtain an optimal projection vector $\mathbf{w}$ onto EEG signal $\mathbf{x}$ that satisfies

$$\begin{aligned} \underset{\mathbf{w}}{\text{argmax}} \quad & E[(\mathbf{w}^T\mathbf{x}_i - E[\mathbf{w}^T\mathbf{x}])^2 | \mathbf{x} \in O_1] \\ \text{s.t.} \quad & E[(\mathbf{w}^T\mathbf{x}_i - E[\mathbf{w}^T\mathbf{x}])^2] = 1 \end{aligned} \tag{1}$$

This can be reformulated as

$$\begin{aligned} \underset{\mathbf{w}}{\text{argmax}} \quad & \mathbf{w}^T \Sigma_{O_1} \mathbf{w} \\ s.t. \quad & \mathbf{w}^T(\Sigma_{O_1} + \Sigma_{O_2})\mathbf{w} = 1 \end{aligned} \tag{2}$$

where Σ_{O_i} denotes the covariance matrix for class O_i. This optimization problem can be solved using joint diagonalization, which consists of computing a projection matrix W and a diagonal matrix D:

$$W\Sigma^{(1)}W^T = D \tag{3}$$

$$W\Sigma^{(2)}W^T = I - D \tag{4}$$

Matrices W and D can be obtained using the following procedure [10]. First we calculate the matrix P which whitens the matrix $\Sigma_{O_1} + \Sigma_{O_2}$: $P(\Sigma_{O_1} + \Sigma_{O_2})P^T = I$. This is typically carried out using singular value decomposition (SVD) and normalization. We then calculate the whitened matrix $\Sigma'_{O_1} = P\Sigma_{O_1}P^T$, and calculate the eigenvector matrix Q for $\Sigma'_{O_1} = QDQ^T$. Finally, we calculate the projection matrix as $W = P^TQ$.

W clearly satisfies Eq. 4 with the columns of W being the optimal CSP projection vectors. Typically, only the vectors with the m highest eigenvalues (for the positive class, i.e. O_1) and lowest m eigenvalues (for the negative class i.e. O_2) are used. Given one $\mathbf{w}$ from the selected $2m$ projection vectors, the CSP feature, f_p, is computed as the normalized log power (variance) of the filtered signal.

$$f_p = \log\left(\frac{\text{var}\left(\mathbf{w}^T\mathbf{x}\right)}{\sum_{i=1}^{2m}\text{var}\left(\mathbf{w}_i^T\mathbf{x}\right)}\right) \tag{5}$$

C. Feature Selection

Feature selection is an important aspect in machine learning and can be defined as *selecting a subset of size k, from a given a set of d features, that leads to the smallest classification errors*. One commonly used method, based on statistics from the data distribution, is the Principal Component Analysis (PCA) which employs covariance, the second-order statistics of the empirical data distribution [13]. In this work we choose Mutual Information (MI), a method from higher order statistics [14], also considered as a good indicator, robust to noise and data transformation, to find relevance between variables. Let us consider a variable X and its corresponding class label variable Y, then the MI between X and Y is defined as:

$$I(X;Y) = H(Y) - H(Y|X) \tag{6}$$

where $H(X)$ denotes the entropy of the feature variable X and $H(Y|X)$ represents the conditional entropy of class label variable Y given feature variable X.

$$H(X) = -\int_{\mathbf{x}\in X} p(\mathbf{x})\log_2 p(\mathbf{x})d\mathbf{x} \tag{7}$$

$$H(Y|X) = -\int_{\mathbf{x}\in X} \sum_{y\in Y} P(y|\mathbf{x})\log_2 P(y|\mathbf{x})d\mathbf{x} \tag{8}$$

We opted for the Naive Bayesian Parzen Window (NBPW) approach described in [15]. The procedure consists of the following steps:

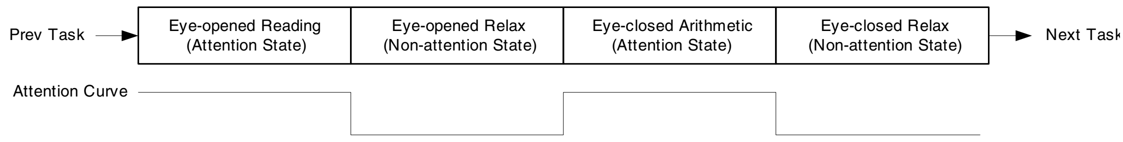

Fig. 2. Sequence of attention / non-attention tasks during one run

1) Initialization: Initial candidate set of d features $F = \{f_1, f_2, \ldots, f_d\}$, and the initial select feature set as a null set $F_{opt} = \emptyset$, $I_0 = 0$;
2) For each feature f_k in the candidate set:
   a) Form a tentative feature vector $F_k = F_{opt} \bigcup \{f_k\}$;
   b) Use F_t and NBPW to predict the class label Y_k;
   c) Compute the MI of the predicted class label and the true label $I(Y_k; Y)$;
3) Select the feature, say f_k, which maximizes $I(Y_k; Y)$;
4) If $F_{opt} \neq \emptyset$ and mutual information gains less than a preset threshold δ: $I(Y_k; Y) - I_0 < \delta$, exit;
5) $I_0 = I(Y_k)$;
6) Update the candidate set by $F \to F \setminus \{f_k\}$ and update the select feature set by $F_{opt} \to F_{opt} \bigcup \{f_k\}$;
7) If the candidate set is empty, exit;
   otherwise go to Step 2;

III. Experimental Study

An experiment was conducted to record EEG for the evaluation of the novel approach to discriminate the two conditions: attention and non-attention. The experiment consists of runs with 4 different tasks as shown in Figure 2. Task 1 was a *concentration task* with eyes opened, in which the subject was asked to continuously read an recent IEEE Transactions paper. Task 2 was a *non-concentration task* with eyes opened, in which the subject stay in a relax state and while looking at surroundings without paying particular attention to any objects. Task 3 was a *concentration task* but this time with eyes closed, during which the subject is asked to perform a mental arithmetic calculation (400 minus 7, minus 7, minus 7, etc.). Finally, Task 4 was a *non-concentration task* with with eyes closed, where a subject is in relax state and not focusing on any particular thought. The ground truth used by the classifier is shown as the attention curve in Figure 2. Five healthy young male subjects participated in the experiment. Two subjects completed 3 sessions of data collection, while the remaining 3 subjects completed only 2 sessions. Each session consists of 5 runs and each run include all 4 tasks. The user was permitted a short pause of few minutes between runs to avoid fatigue. EEG data was check for apparent EMG artefacts. Subjects were seated in a comfortable armchair, and reminded to not to have large body movements in order to avoid too much EMG artefacts. During Task 3 and Task 4 of the experiment, subjects were monitored to avoid dozing off episodes.

EEG was recorded using a 40-channel EEG amplifier (Compumedics Inc., USA). The EEG cap is standard 32 channels (See Figure 3) but we limited our recordings only 15 channels

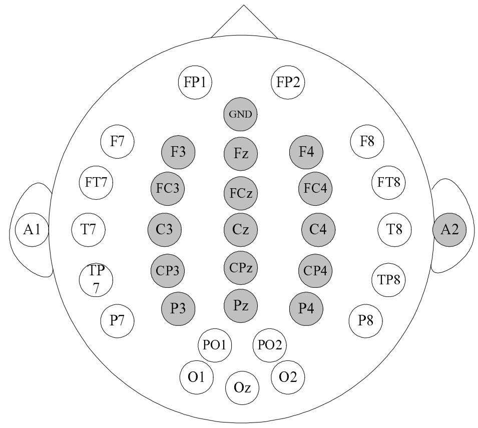

Fig. 3. EEG cap with highlighted in grey the electrodes used in this study

(F3, Fz, F4, FC3, FCz, FC4, C3, Cz, C4, CP3, CPz, CP4, P3, Pz, P4) using the right earlobe (A2) as reference. The sampling rate was 250Hz, and a hardware 50Hz notch filter was used to reduce the main interference. All electrodes had less than 5kΩ impedance.

IV. Results

The objective of this study was to examine the capability of the novel approach to differentiate attention from non-attention conditions. Thus, we considered Task 1 and Task 3 as representing attention class, and Task 1 plus Task 4 as the non-attention class. Continuous data were segmented into EEG epochs, and our method was tested to classify each epoch into attention or non-attention class. It was of interest to see the system performance over difference lengths of EEG epochs. Thus, we examined EEG epochs of 2s, 4s, 8s, or 16s long.

To avoid any bias in the evaluation results, we used a typical 4x4fold cross-validation. Briefly, each subject's EEG data was divided into n parts, then where in each fold a particular part is used for testing while the other $n-1$ parts were used for training. In other words the test part is not used in the training. This process is repeated n times and accuracy is averaged.

The CSP was applied to find the most discriminative spatial patterns between the attention and non-attention classes from channels filtered by the filter-bank. After the feature extraction stage, we obtain $9 \times 2 = 18$ candidate feature set. An MI-based feature selection procedure is then applied for further reduction (typically 2 to 4 features) in combination with the FLD classification. We used a threshold δ of 0.1 for all experiments.

The baseline for comparison is the prior-art method spectral-based method [6] [12] with alpha (8-12Hz), beta (12-30Hz), and theta (4-8Hz) band joint to form a feature input vector to a FLD classifier.

We wanted to investigate how robust was our approach when changing the window length as typically, near real-time situation (online system with continuous EEG) requires short window length, while longer windows can be used for offline situation.

Results for classification accuracy are presented in Table I. Our method outperforms the prior-art method with higher accuracies from 60.1% to 89.4% (mean from 65.25% to 85.58%) while prior-art only achieving 50.60% to 71.9% (mean 55.50% to 69.55%). With subject #2, for e.g., the improvement is significant with more than 16% (mean from 69.55% to 85.58%). Classification accuracies are also shown to be constant across all different window lengths. We also notice that performance do not drop for short window (e.g. 2s), which is an important issue when aiming at near real-time operation.

TABLE I

ACCURACY RESULTS FOR PRIOR-ART AND NOVEL APPROACH

	Window length				
	2s	4s	8s	16s	Mean
Prior-art method					
Subject #1	58.0%	55.8%	56.3%	56.3%	56.60%
Subject #2	71.9%	69.4%	69.4%	67.5%	69.55%
Subject #3	66.1%	64.3%	59.7%	60.6%	62.67%
Subject #4	69.1%	67.6%	68.1%	65.0%	67.45%
Subject #5	58.6%	56.2%	56.6%	50.6%	55.50%
Novel approach					
Subject #1	64.3%	65.0%	65.8%	65.9%	65.25%
Subject #2	82.3%	86.8%	89.4%	83.8%	85.58%
Subject #3	74.3%	72.8%	72.8%	69.4%	72.33%
Subject #4	60.1%	72.2%	65.0%	70.0%	66.83%
Subject #5	65.1%	64.0%	70.3%	75.0%	68.60%
Mean	69.2%	72.2%	72.7%	72.8%	71.72%

V. CONCLUSIONS

In this paper we proposed a novel approach to concentration taking advantages of signal processing and machine learning paradigms. Spectral-Spatial features are extract using filter bank and CSP filtering from EEG data. Best features are selected using a MI-based procedure and classified using FLD. Results from an 5 subjects experiment showed that the novel approach provides significant improvement over prior-art method which uses specific EEG band powers. The resulting EEG-based attention level measurement system is robust, run in near real-time, and requires only few electrodes. The system can be implemented as a low-cost portable EEG system. The novel approach can be demonstrated with a two-players attention-based racing car computer game.

Future work will focus on EEG recording conditions with a EEG recording protocol based on variants of the stroop test [16]. We will also investigate the use of Relevance Vector Machine (RVM) [17] as classifier.

ACKNOWLEDGMENT

This work is supported by the Agency for Science, Technology and Research (A*STAR), Singapore.

REFERENCES

[1] T.-P. Jung, S. Makeig, M. Stensmo, and T. J. Sejnowski, "Estimating alertness from the EEG power spectrum," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 1, pp. 60–69, January 1997.

[2] S. P. Kelly, P. Docktree, R. B. Reilly, and I. H. Robertson, "EEG Alpha power and coherence time courses in a sustained attention task," *Proceedings of the International Conference on Neural Engineering, March 20-22, 2003*, pp. 83–86.

[3] R.-S. Huang, T.-P. Jung, and S. Makeig, "Multi-scale EEG brain dynamics during sustained attention tasks," *Proceedings of the 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2007), Honolulu, Hawaii, April 15–20, 2007*, vol. 4, pp. 1173–1176.

[4] A. J. Haufler, T. W. Spalding, D. L. Santa Maria, and B. D. Hatfield, "Neuro-cognitive activity during a self-paced visuospatial task: comparative EEG profiles in marksmen and novice shooters," *Biological Psychology*, vol. 53, no. 2–3, pp. 131–160, July 2000.

[5] R. J. Barry, A. R. Clarke, and S. J. Johnstone, "A review of electrophysiology in attention-deficit/hyperactivity disorder: I. qualitative and quantitative electroencephalography," *Clinical Neurophysiology*, vol. 114, no. 2, pp. 171–183, 2003.

[6] V. J. Monastra, S. Lynn, M. Linden, J. F. Lubar, J. Gruzelier, and T. J. LaVaque, "Electroencephalographic Biofeedback in the Treatment of Attention-Deficit/Hyperactivity Disorder," *Applied Psychophysiology and Biofeedback*, vol. 30, no. 2, pp. 95–114, June 2005.

[7] J. F. Lubar, "Discourse on the development of EEG diagnostics and biofeedback for attention-deficit/hyperactivity disorders," *Applied Psychophysiology and Biofeedback*, vol. 16, no. 3, pp. 201–225, September 1991.

[8] M. Linden, T. Habib, and V. Radojevic, "A controlled study of the effects of EEG biofeedback on cognition and behavior of children with attention deficit disorder and learning disabilities," *Applied Psychophysiology and Biofeedback*, vol. 21, no. 1, pp. 35–49, March 1996.

[9] P. N. Friel, "EEG biofeedback in the treatment of attention deficit/hyperactivity disorder," *Alternative Medicine Review*, vol. 12, no. 2, pp. 146–151, 2007.

[10] J. Müller–Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Electroencephalography and Clinical Neurophysiology*, vol. 110, no. 5, pp. 787–798, May 1999.

[11] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter Bank Common Spatial Pattern (FBCSP) in Brain-computer Interface," *Proceedings of the 2008 International Joint Conference on Neural Networks (IJCNN2008), Hong Kong, June 2-6, 2008*, pp. 2391–2398.

[12] U. Leins, G. Goth, T. Hinterberger, C. Klinger, N. Rumpf, and U. Strehl, "Neurofeedback for Children with ADHD: A Comparison of SCP and Theta/Beta Protocols," *Applied Psychophysiology and Biofeedback*, vol. 32, no. 2, pp. 73–88, June 2007.

[13] T. W. S. Chow and D. Huang, "Estimating optimal feature subsets using efficient estimation of high-dimensional mutual information," *IEEE Transactions on Neural Networks*, vol. 16, no. 1, pp. 213–224, January 2005.

[14] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley–Interscience, 2006, IBSN 0471062596.

[15] K. K. Ang and C. Quek, "Rough set-based neuro-fuzzy system," *Proceedings of the International Joint Conference on Neural Networks (IJCNN2006), Vancouver, BC, Canada, July 16–21, 2006*, pp. 742–749.

[16] C. MacLeod, "Half a Century of Research on the Stroop Effect: An Integrative Review," *Psychological Bulletin*, vol. 109, no. 2, pp. 163–203, 1991.

[17] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, June 2001.