14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

MOTOR IMAGERY BASED BRAIN COMPUTER INTERFACE WITH SUBJECT ADAPTED TIME-FREQUENCY TILING

Nuri F. Ince [1, 2], Ahmed H. Tewfik [2], Sami Arıca [1]

Department of Electrical and Electronics Engineering, University of Cukurova, Turkey [1]
Department of Electrical and Computer Engineering, University of Minnesota, USA [2]
e-mail: firat@ece.umn.edu, tewfik@ece.umn.edu, arica@cu.edu.tr

ABSTRACT

We introduce a new technique for the classification of motor imagery electroencephalogram (EEG) recordings in a Brain Computer Interface (BCI) task. The technique is based on an adaptive time-frequency analysis of EEG signals computed using Local Discriminant Bases (LDB) derived from Local Cosine Packets (LCP). Unlike prior work on adaptive time-frequency analysis of EEG signals, this paper uses arbitrary non-dyadic time segments and adaptively selects the size of the frequency bands used for feature extractions. In an offline step, the EEG data obtained from the C3/C4 electrode locations of the standard 10/20 system is adaptively segmented in time, over a non-dyadic grid. This is followed by a frequency domain clustering procedure in each adapted time segment to maximize the discrimination power of the resulting time-frequency features. Then, the most discriminant features from the resulting arbitrarily segmented time-frequency plane are sorted. A Principal Component Analysis (PCA) step is applied to reduce the dimensionality of the feature space. The online step simply computes the reduced dimensionality features determined by the offline step and feeds them to the linear discriminant. The algorithm was applied to all nine subjects of the BCI Competition 2002. The classification performance of the proposed algorithm varied between 70% and 92.6% across subjects using just two electrodes. The average classification accuracy was 80.6%. For comparison, we also implemented an Adaptive Autoregressive model based classification procedure that achieved an average classification rate of 76.3% on the same subjects, and higher error rates than the proposed approach on each individual subject.

1. INTRODUCTION

The use of the electrical activity of the brain for communication and control has gained significant interest in recent years. The so-called BCI is constructed by recording, processing and classifying the brain signals which are induced by mental tasks [1]. Since these tasks are executed without the need of any muscular act, the related activity has great potential as a new mechanism for handicapped people to interact with their environment.

In general the electroencephalogram (EEG) is used in BCI systems due to its ease of recording and low cost. Subjects are asked to execute a mental task such as imagining a hand movement or solving a mathematical problem. Related rhythmic patterns in the EEG signals from the activated electrodes are then processed and classified to give feedback to the subject. Such movement imagery (MI) is used widely as a BCI strategy. In particular, event-related desynchronization (ERD) and synchronization (ERS) structures caused by MI are analyzed [1, 2]. Frequency bands, that show significant change in a predefined time window, are selected for the filtering process. It is assumed that the selected interval is an active segment. Previously, distinction sensitive learning vector quantization (DSLVQ) had been used as an automated approach to select relevant bands in fix segments [3]. However, ERS and ERD do not necessarily occur in a specific time interval. Therefore the selection of the active segment can be a problem if fixed windows are used. Auto regressive (AR) modeling and adaptive autoregressive (AAR) modeling were then used to deal with the non-stationary nature of the EEG signals [4, 5]. Both DSLVQ and AR model based studies indicate that subject specific time and frequency domain features of EEG signal do exist. However, neither method considers features from multiple time and frequency indexes. They ignore the possibility that subjects may have physio-anatomical differences and/or different imagery strategies to induce ERD and ERS patterns. In addition, there is a rich literature which indicates that the time and frequency characteristics of the alpha (7-13Hz) and beta (14-30Hz) band components can vary widely with beta band showing burst activity, whereas alpha band changes taking seconds to attenuate and recover [1, 2]. Recently adaptive time-frequency methods were used to visualize and segment movement EEG [6, 7, 8]. In these studies, the aim was to extract subject ERD and ERS structures.

In our prior work, [7] and [8], we demonstrated the advantages of using adapted time-frequency bases for analysis and classification EEG records accompanying real movements. Here, we significantly enhance these approaches by using non-dyadic time segmentations of the underlying EEG signals and adaptively selecting the most discriminant frequency band features in each time segment. In contrast with our earlier works which used fixed bands with the largest energy mean difference, here, we select the band sizes adaptively by estimating the distance between the probability distributions of expansion coefficients which are less sensitive to outliers. Also this latest method overcomes the limitation to dyadic grid of the original local discriminant bases (LDB) procedure. We show that the proposed approach is also capable of adapting physio-anatomical differences and subject dependent motor imagery patterns and resulting in a much better classifier.

2. MATERIALS

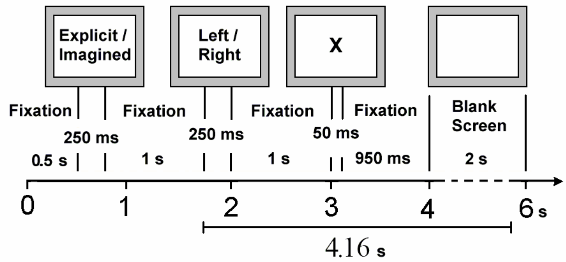
The dataset of BCI competition 2002, which was provided by Dr. Allen Osman from University of Pennsylvania, is used in this investigation [9]. The imagery EEG data was collected from 9 subjects. These subjects were asked to execute an imagined left and right index finger movement in an experimental paradigm given in figure 1. First the subjects are told whether the action will be explicit or imagined. Then a L/R cue appears on the screen indicating whether the movement is left or right. One second after the L/R cue, the letter X appears on the screen indicating it is time to take the required action. EEG was recorded with 100Hz sampling frequency from 59 electrodes placed on site corresponding to the International 10/20 system and referenced to the left mastoid. In this study the EEG data from only the C3 and C4 electrodes are analyzed. These channels are converted to Hjort derivation in order to enhance the local activity [10]. The Hjort derivation C_i^H is calculated as

$$C_i^H = s_{C_i} - \frac{1}{4}\sum_{j \in S_i} s_{C_j} \tag{1}$$

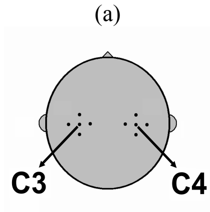
where s_{Ci} is the reading of the center electrode C_i, with $i=3$ or 4 and S_i is the index of 4 electrodes surrounding electrode C_i (c.f., Figure 1). Then the EEG data is bandpass filtered between 2-40Hz. For classification, we use all 90 trials available for each task.

3. METHODS

The proposed EEG signal classification approach consists of five steps (figure 2): The first four steps consist of off-line data preprocessing and adaptive time-frequency segmentation of the EEG signals. The last step is the online classification procedure. The offline time-frequency adaptation step is preceded by a spin cycle procedure to deal with the shift variance of the local cosine packets. It begins with an application of the merge/divide algorithm with local cosine packets to adaptively segment the EEG along the time axis. This is then followed by the frequency domain clustering procedure. Finally, principal component analysis is performed to reduce the dimensionality of the feature space. The on-line step simply computes the reduced



(a)



(b)

Figure 1 - (a) The time course of the experimental paradigm. The analysis window is indicated at the bottom and includes the preparation stage as well. (b) The C3/C4 and surrounding pair of 4 electrode locations used in this study.

dimensionality features selected by the off-line step and feeds them to a linear discriminant. In the remainder of this section, we briefly describe the various steps of the approach emphasizing the key off-line novel steps.

3.1. Flexible Local Discriminant Bases

As emphasized in the previous section the alpha and beta band ERD/ERS patterns have transient behavior. Therefore it is crucial to focus on the local properties of the EEG. The LDB algorithm [11] has been developed to extract such local information. The original LDB algorithm expands the signals of given classes into orthonormal bases by using wavelets or local trigonometric bases over a dyadic tree structure. It then finds the nodes of the tree where the classes are well separated by using a distance function maximization strategy [11]. Therefore the LDB algorithm is a very powerful method for extracting discriminant time-frequency features. Since the MI related ERD/ERS patterns appear as time locked transient phenomena, we use Local Cosine Packets to describe the signal. However, subject specific EEG patterns will not necessarily fall in dyadic segments of the original LDB. Also to capture the discriminant information in the EEG with individual expansion coefficients is difficult. Therefore we developed the Flexible-LDB algorithm which

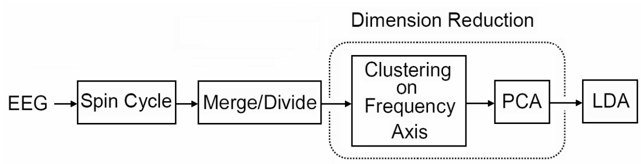


Figure 2 - The block diagram of the proposed approach.

enhances the time segmentation procedure by adopting a merge/divide strategy that is used in geophysical waveform compression [12]. Then we extract most discriminant band features in each time segment with a clustering procedure. Let us first illustrate the time adaptation algorithm with a schematic diagram given in Figure 3. Here the signal is analyzed with three smooth windows which have children and mother structure. In order to preserve orthogonality, these windows are constructed using cutoff functions *r(t)* that satisfy

$$\left|r(t)^2\right| + \left|r(-t)^2\right| = 1 \quad \text{for all } t \in R; \qquad (2)$$
$$r(t) = \begin{cases} 0, & \text{if } t \le -1, \\ 1, & \text{if } t \ge 1. \end{cases}$$

An example of such a function *r(t)* is

$$r(t) = \begin{cases} 0, & \text{if } t \le -1, \\ \sin\left[\dfrac{\pi}{4}\left(1+\sin\left(\dfrac{\pi t}{2}\right)\right)\right], & \text{if } -1 < t < 1, \\ 1, & \text{if } t \ge 1. \end{cases} \qquad (3)$$

Unlike the dyadic case, the children windows are not necessarily half the length of the mother window. In each segment the signal is expanded using Local Cosine Packets which provides local spectral representation. Then the distance between the expansion coefficients are compared in the mother and children subspaces. Whenever the total distance between classes in the children subspaces is greater than the mother subspace, the signal is divided at that point. Otherwise, the children segments are discarded. In the next iteration, the mother segment is used as the left child (M_2 in Figure 3). Note that the right child is the basic smallest size time segment used by the procedure and will have a fixed length. The left child can grow to be multiples of the basic smallest segment. This algorithm is iterated from left to right along the time axis by implementing the above procedure to achieve the desired time adaptation.

The algorithm can be summarized as follows.

Step 1. Select a basic time window (Cell) size and construct a children mother structure.

Step 2. In each space expand the signal into Cosine Packets. For each expansion coefficient calculate the distance between each class and accumulate the distances of expansion coefficients in each subspace.

Step 3. Merge the children subspaces if their discrimination power is less than that of the mother subspace else divide the signal at that point.

Step 4. Iterate the previous step from left to right until the desired time adaptation is obtained.

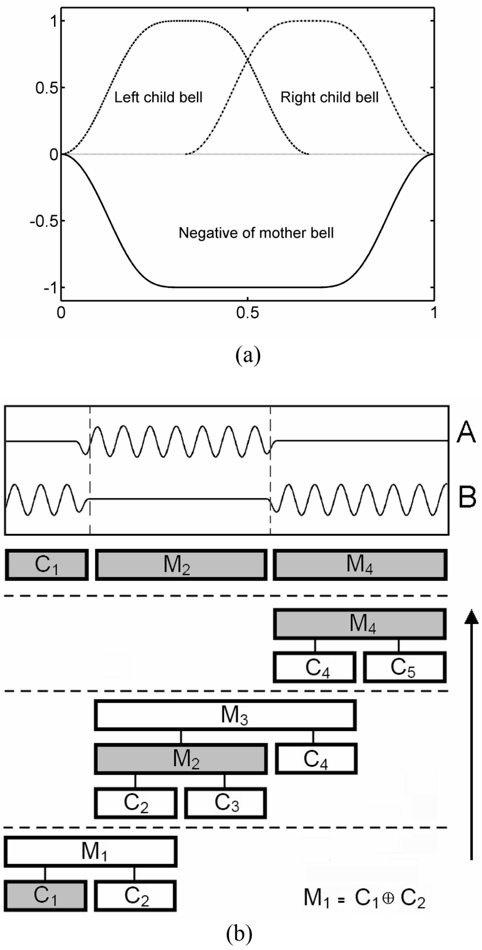


Figure 3 - (a) The smooth mother and children windows. (b) A schematic illustration of merge/divide algorithm to locate the segments where the signals belonging to the classes **A** and **B** are well separated. Note that the rectangular segments correspond to smooth windows given in (a). The C_1, M_2 and M_4 segments are obtained by iterating the merge/divide algorithm. Please follow from bottom to top to see how time adaptation is obtained.

Step 5. Order the expansion coefficients from the segmented signal by using a class separability criterion and pick up the top $k << n$ coefficients for classification.

There are various choices for distance measures. Assume p, q represent the probability distribution of each expansion coefficient estimated via a histogram. We have used the Euclidean distance

$$D(p,q) = \|p_i - q_i\|^2 = \sum_{i=1}^{n} (p_i - q_i)^2 \qquad (4)$$

for pruning the tree. Further, we implemented the Fisher class separability criterion

$$F = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \qquad (5)$$

for ordering the features, where μ and σ are the mean and standard deviation of the feature they belong to.

3.2. Feature Extraction, Dimension Reduction and Classification

The original LDB algorithm produces a feature space with high dimensionality. A feature space of high dimension can decrease the generalization capability of the classifier when there is a limited number of training samples available [13]. Also during our studies we observed that the center frequency of the oscillations differ from sweep to sweep. This increases uncertainty along the frequency axis. Therefore, in our previous studies we merged the expansion coefficients in 4Hz frequency bins. Here we enhance the frequency adaptation within each time segment. Specifically, we merge coefficients that are adjacent in the frequency domain only if their union has larger discrimination power than the individual coefficients. Note that this is basically a coefficient clustering approach obtained via cost function maximization and results in an adaptive frequency band adaptation for discrimination.

The Flexible-LDB which is obtained by combining of adaptive time segmentation and frequency axis clustering approach results in an arbitrary segmented t-f plane, which has interesting connections to [14]. However, in [14] the authors used a double tree approach which was also limited to a dyadic grid. Here we have overcome this limitation.

Note also that Cosine Packet representations are not shift invariant. To get around this problem we implement the "spin cycle" procedure of [11]. The procedure expands the training set by generating its time shifted versions in both directions in a circular manner. If the desired number of shifts is τ then the training set is expanded to $2\tau+1$ including the original signal and its shifts by $-\tau, -\tau+1, ..., \tau$.

We use linear discriminant analysis (LDA) as a classifier. However the F criterion does not take into account the correlations between ordered features. Therefore prior to feeding the features to the LDA procedure, we apply principal component analysis (PCA) on the top sorted features. The PCA procedure removes the correlation between features and reduces the dimensionality of the feature space. Then we sort the projected feature set according to corresponding eigenvalues in the descending order. Finally we supply this PCA re-ordered feature set to LDA.

4. RESULTS

To assess the efficiency of the proposed algorithm we compared its performance with that of the standard reference AAR model based approach for feature extraction. The AAR approach is widely used in BCI research due to its efficiency and ease of applicability. An AAR model is described as follows

$$y[n] = a_{1,n} y[n-1] + .. + a_{p,n} y[n-p] + x \qquad (6)$$

where y is the output sequence, p is the model order, $a_{m,n}$ are the time varying model parameters and x is white noise with a zero mean and variance σ^2. We selected $p = 6$ and calculated the model parameters for every sample using the least mean square approach for C3 and C4 electrodes. Note that with $p=6$, the model represents 3 peaks in the signal spectrum. Hence, our parameter choices resulted in a feature vector of dimension 12 which is commonly preferred in BCI studies.

To compare AAR, Dyadic and Flexible LDB algorithms, we selected an analysis window of 416 samples and a tree depth of 4. The cell size for the merge/divide approach is chosen to be equal to the deepest segment, which is 260ms. For the PCA procedure, we typically select k in the range of 32 to 48 because most of the discrimination power is concentrated in these coefficients. We have used 10 times 10 fold cross validation to estimate the classification accuracy. Table 1 shows the classification accuracy for all 9 subjects.

The Flexible-LDB algorithm outperformed other methods. One way paired t-test indicates that the classification accuracy of the introduced method is significantly better than Dyadic-LDB and AAR methods ($p<0.0012$, $p<0.0022$). We observed that the algorithm constructed different time-frequency tiling for different subjects. We visualize selected time-frequency features for a representative subject in figure 4. Further, for a given subject, the two hemispheres are represented by distinct segmentations and features. Also the total discrimination power of the features obtained from C3/C4 electrodes was different. Therefore our results support a hemispheric asymmetry behavior. We noticed that the minimal classification error is obtained after combining several features. This indicates that the interactions between

TABLE 1
The classification accuracy (%) of Dyadic, Flexible LDB and AAR methods

Subjects	Dyadic LDB	Flexible - LDB	AAR (p=6)
S1	82.4	83.6	81.5
S2	91.2	92.6	88.5
S3	65.4	70	67.5
S4	66.2	70.7	71
S5	77.1	77.8	71
S6	81.9	87.2	78.5
S7	87.5	89.7	80.5
S8	62	70	67
S9	80.3	83.7	81.4
Avg	77.1	**80.6**	76.3

different cortical regions and time frequency locations are important for discrimination. Currently we are investigating an approach which can construct the t-f tiling by accounting the interactions between features. A recently introduced t-f approach in [15] has achieved 80% by using a simultaneous analysis of 20 electrodes on the same data set. Our algorithm is capable of achieving similar (or slightly better) performance just by using 2 electrodes. Its efficiency of constructing arbitrary tiling for each subject and weighting space/electrode locations can be a reason for this. Also being able to capture same error rates with minimal number of electrodes makes the algorithm computational efficient. Obtained classification accuracy and capability of adapting to inter-subject variability and physio-anatomical differences make the proposed algorithm a promising candidate for future BCI systems.

Acknowledgements

This work is supported by National Scientific Research Council of Turkey (TUBITAK).

4. REFERENCES

[1] G. Pfurtscheller, C. Neuper, "Motor Imagery and Direct Brain-Computer Interface," *Proceeding of IEEE*, vol.89, pp. 1123-1134, 2001.

[2] 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.

[3] M. Prezenger, G. Pfurtscheller, "Frequency component selection for an EEG-based brain computer interface," *IEEE Trans. on Rehabil. Eng.* 7, pp. 413-419, 1999.

[4] J. R. Wolpaw and D.J. McFarland, "Multichannel EEG-based brain-computer communication," *Electroenceph. Clin. Neurophy.*, vol. 90, pp. 444-449, 1994.

[5] A. Schlögl, D. Flotzinger, G. Pfurtscheller, "Adaptive autoregressive modeling used for single trial EEG classification," *Biomed. Technik,*42, pp. 162-167, 1997.

[6] J. Ginter Jr., K. J. Blinowska, M. Kaminski, P. J. Durka, "Phase and Amplitude analysis in time- frequency space-application to voluntary finger movement," *Journal of Neuroscience Methods* 110, pp. 113-24, 2001.

[7] N. F. Ince, S. Arica, "Analysis and Visualization of ERD and ERS with Adapted Local Cosine Transform," Proceedings of 26th Annual International Conference of IEEE Engineering in Medicine and Biology Society, EMBC, San Francisco, 2004.

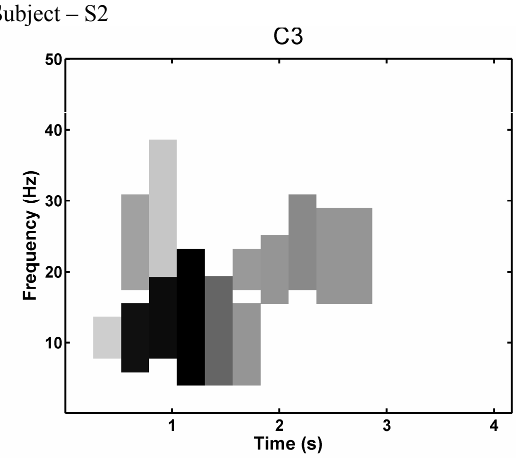
[8] N. F. Ince, A. Tewfik, S. Arica, "Classification of movement EEG with Local Discriminant Bases," 30th International Conference on Acoustics, Speech, and Signal Processing (ICASSP) Society, IEEE, Philadelphia, 2005.

[9] A. Osman, and A. Robert, "Time-course of cortical activation during overt and imagined movements", in Proc. Cognitive Neuroscience Annu. Meet., New York, 2001.

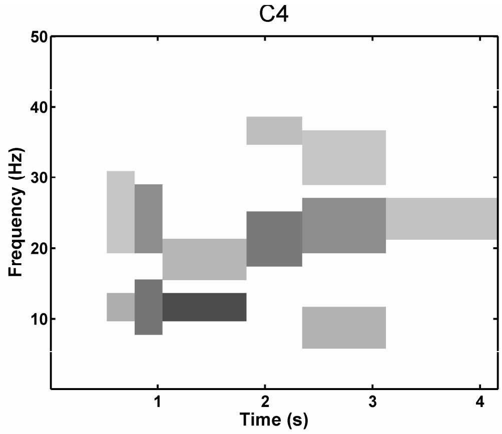
[10] B. Hjorth, "An online transformation of EEG scalp potentials into orthogonal source derivations". Electroencephalography and Clinical Neurophysiology, 39, pp. 526-530, 1975.

[11] N. Saito, R. R. Coifman, F. B. Geshwind, F. Warner, "Discriminant feature extraction using empirical probability density and a local basis library," *Pattern Recognition*, vol.35, pp. 2842-2852, 2002.

Subject – S2



(a)



(b)

Figure 4 - The time-frequency features of a representative subject obtained from (a) C3 and (b) C4 electrode locations. The darker locations have the more discrimination power. Note the differences between tilings, feature characteristics and their discrimination power of each electrode location.

[12] R. S. Wu, and Y. Wang, "New Flexible Segmentation Technique in sesimic data compression using local cosine transform". Wavelet Applications in signal and image processing VII, Proc. SPIE, 3813, pp. 784-794., 1999.

[13] Bishop C. M., *Neural Networks for Pattern Recognition*, Oxford, U.K., Clarendon, 1995.

[14] C. Herley, J. Kovacevic, K. Ramchandran, and M. Vetterli, "Tilings of the Time –Frequency plane: Construction arbitrary orthogonal bases and fast tiling algorithms", IEEE Trans. Signal Proc., special issue on Wavelets and Signal Processing, vol. 41, pp. 3341-3359., 1993.

[15] T. Wang, and B. He, "Classifying EEG-based motor imagery tasks by means of time–frequency synthesized spatial patterns", *Clin. Neuro.* vol.115, pp. 2744–2753, 2004.