Institute of Physics Publishing    Journal of Neural Engineering

J. Neural Eng. **3** (2006) 235–244    doi:10.1088/1741-2560/3/3/006

# Classification of single trial motor imagery EEG recordings with subject adapted non-dyadic arbitrary time–frequency tilings

**Nuri Fırat Ince[1,2], Sami Arica[1] and Ahmed Tewfik[2]**

[1] Department of Electrical and Electronics Engineering, University of Cukurova, Adana 01330, Turkey
[2] Department of Electrical and Computer Engineering, University of Minnesota, MN 55455, USA

Received 2 February 2006
Accepted for publication 27 June 2006
Published 19 July 2006
Online at stacks.iop.org/JNE/3/235

**Abstract**

We describe 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). In an offline step, the EEG data obtained from the $C_3/C_4$ electrode locations of the standard 10/20 system is adaptively segmented in time, over a non-dyadic grid by maximizing the probabilistic distances between expansion coefficients corresponding to left and right hand movement imagery. 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. This reduced feature set is finally fed to a linear discriminant for classification. The online step simply computes the reduced dimensionality features determined by the offline step and feeds them to the linear discriminant. We provide experimental data to show that the method can adapt to physio-anatomical differences, subject-specific and hemisphere-specific motor imagery patterns. 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 error rate of 76.3% on the same subjects, and higher error rates than the proposed approach on each individual subject.

(Some figures in this article are in colour only in the electronic version)

## 1. Introduction

A brain computer interface (BCI) [1] is a system that can translate the electrical activity of the brain for use in communication and control. Such a translation can be achieved invasively, e.g., by measuring single neuron activities [2], or non-invasively, e.g., by recording electroencephalogram (EEG) [3, 4]. Both methods have advantages and disadvantages. The invasive methods provide higher spatial resolution and signal-to-noise ratio (SNR). However, they require implantation of an electrode grid into the cortex, an operation that involves several risks. The long-term recording of single neural activity is also difficult. The EEG, on the other hand, has poor spatial resolution and low SNR. Further, high frequency components such as gamma rhythms are difficult to obtain with EEG. However, the EEG is advantageous because it is non-invasive, inexpensive and easily applied to human beings. In addition, it provides long-term robust signals.

In an EEG-based BCI, the subject is usually asked to execute a mental task in synchronization with a cue. Motor imagery (MI) is used, in particular, in synchronous BCI systems. It can induce short lasting amplitude decrease and enhancement in corresponding cortical areas in the rhythmic components of EEG. The resulting signals are called event related desynchronization (ERD) and event related synchronization (ERS), respectively [5, 6]. The underlying basis of using MI as a BCI strategy is that unilateral hand movement imagery produces contralateral hemisphere dominant activities.

In this paper, we describe a novel procedure for the adaptive time–frequency analysis and classification of EEG signals recorded in MI-based BCI systems. In section 2, we provide a brief summary of previous MI-related EEG signal classification techniques. In section 3 we describe the motor imagery dataset. Section 4 details our proposed approach. We provide experimental results and comparisons against the adaptive autoregressive method in section 5.

## 2. Background

Several methods have been used in the past to extract and classify the ERD/ERS structures for a BCI task. In the earliest research, band powers (BP) obtained from fixed windows were used in combination with learning vector quantization [7]. The BP method requires prior knowledge of the reactive bands. In [8] distinction sensitive learning vector quantization (DSLVQ) was proposed to select the reactive bands in an automated manner in fixed windows. However, the ERD/ERS events do not necessarily occur in fixed time segments. Besides, there is strong evidence in the literature showing that the time courses of ERD and ERS in the alpha and beta bands are dissimilar. It has been reported that the alpha band takes several seconds to attenuate and recover whereas the beta band shows burst activity [5, 6]. The adaptive autoregressive modeling (AAR) method of [9] overcomes differences in time adaptation by defining the spectral information dynamically using autoregressive (AR) parameters. The AAR is used to find the best time point for classification. It has been reported that the best time point is subject dependent. Both DSLVQ and AR model based research indicates that time–frequency domain features [8–10] can be used in BCI systems. These features differ from person to person and may be caused by physio-anatomical and motor imagery strategy differences. As mentioned above, previous research, e.g., [8–10], indicates that a mental state classification algorithm should simultaneously consider multiple heterogeneous time and frequency features. However, none of the methods listed above uses multiple heterogeneous time and frequency features simultaneously for classification. Further, it is clearly desirable to adapt to the time-ERD/ERS patterns in an automated manner to account for subject-dependent variability.

Recently, the authors established the advantages of using an adaptive time–frequency analysis approach with dyadic time segmentation to classify single trial EEG related to real movements [11]. They also analyzed and visualized the resulting most discriminant patterns [12]. Here, we significantly enhance the classification performance of these prior approaches by using non-dyadic time segmentations of the underlying EEG signals and adaptively selecting the most discriminant time–frequency features in each time segment. We show below that the proposed approach is capable of adapting to physio-anatomical differences and subject- and hemisphere-dependent motor imagery patterns, resulting in a much better classifier.

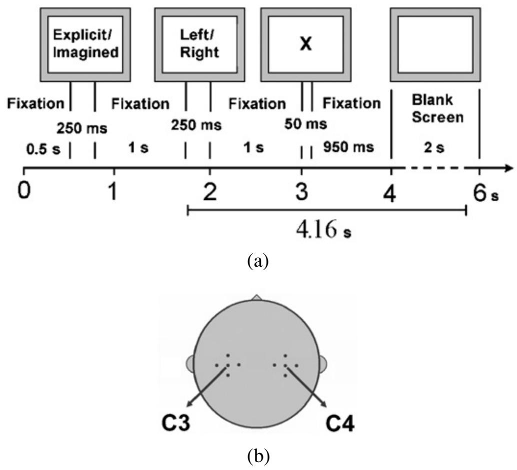


**Figure 1.** The time course of the experimental paradigm. (a) The analysis window is indicated at the bottom and includes the preparation stage. (b) The $C_3/C_4$ electrode locations used in this study.

## 3. Materials

The dataset of BCI competition 2002 is used in this investigation [13]. This dataset was provided by Dr Allen Osman from University of Pennsylvania. The imagery EEG data were collected from nine subjects. These subjects were asked to execute an imagined left (L) or right (R) index finger movement according to an experimental paradigm shown 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 the time to take the required action. The EEG is recorded with a sampling frequency of 100 Hz from 59 electrodes at sites corresponding to the international 10/20 system and referenced to the left mastoid. In this study the EEG data from the $C_3$ and $C_4$ electrodes only are analyzed. These channels are converted to Hjort derivation in order to get the local activity [14]. Specifically, let $C_i^H$ be the Hjort derivation corresponding to electrode $C_i$, $i = 3$ or 4. The Hjort derivation $C_i^H$ is calculated as

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

where $s_{C_i}$ is the reading of the center electrode $C_i$, with $i = 3$ or 4 and $S_i$ is the set of indices corresponding to the four

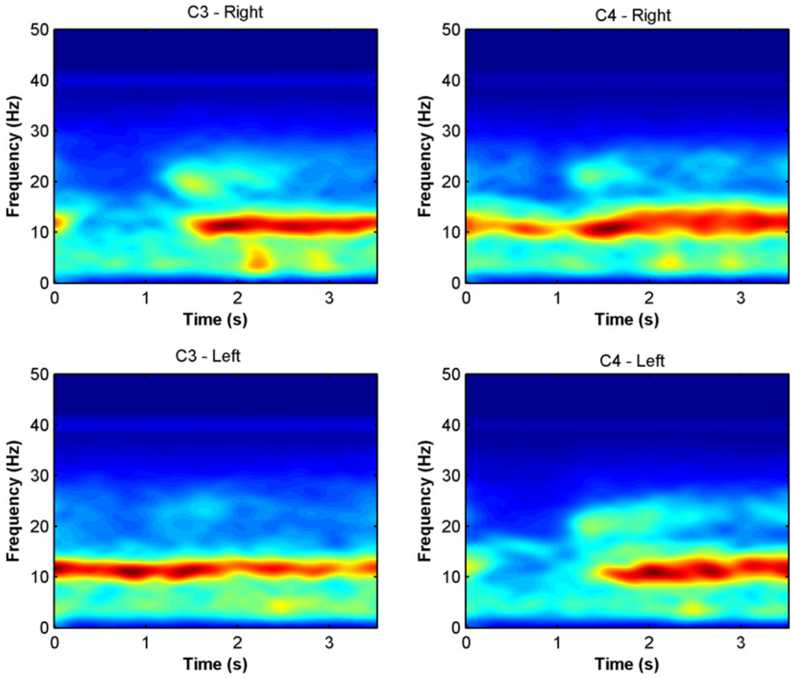


**Figure 2.** Averaged time–frequency maps of right (top) and left (bottom) hand movement imagination of subject S2. Note the energy difference in the interval 0–1.5 s on both channels according to the direction of the imagination. On both channels the ERD in this segment is followed by a short lasting beta ERS around 20 Hz.

electrodes surrounding electrode $C_i$ (cf figure 1(b)). The EEG data are finally bandpass filtered and only information between 2 and 40 Hz was retained. We used all 90 trials available for each task to test the classification approaches discussed below. Figure 2 shows the short time Fourier transforms (STFTs) of the $C3$ and $C4$ electrodes of one subject corresponding to left and right hand movement imageries in the selected time window. The STFTs shown are averages over several sweeps. The maps shown are constructed using overlapping windows and contain redundant information. Note that these time–frequency maps support the findings of previous researchers described in section 2. We will show below how to construct non-redundant overlapping time–frequency tilings that are optimized for discrimination between left and right imagery movements.

## 4. Methods

The proposed EEG signal classification approach consists of five steps (figure 3). 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 dimensionality features selected by the off-line step and feeds them to a linear discriminant.

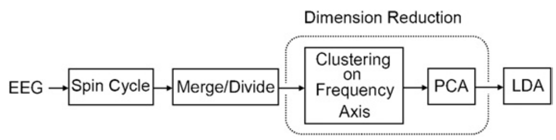


**Figure 3.** Block diagram of the proposed approach.

### 4.1. Data pre-processing and adaptive time–frequency segmentation

*4.1.1. Spin cycle procedure.* Since ERD and ERS are time locked event related potentials, we selected cosine packets (CP) representations to derive an adaptive partition of the time axis [15, 16]. This should enable us to locate discriminant ERD/ERS patterns that have different temporal behaviors. We provide more details about the local cosine packet analysis below. We note here however that CP do not satisfy the shift invariance property. This means that a shift in the original signal can cause major changes in the expansion coefficients. This makes CP hard to use for pattern recognition [17]. To overcome this problem we used the 'spin cycle' procedure of [18, 19]. The procedure expands the training set by generating its shifted versions in both directions in a circular manner. If the desired number of shifts is $\tau$ then the training set is expanded to include its shifts by $(-\tau, -\tau + 1, \ldots, \tau)$. All shifted versions are then analyzed simultaneously as described below.

*4.1.2. Local cosine packets.* The short time Fourier transform (STFT) is traditionally used to analyze the local frequency content of a signal. This type of block transform generates side-lobe artifacts when disjoint rectangular windows are used. When smooth windows are used, orthogonality is lost. It is possible to construct orthogonal transforms with smooth and overlapping windows via trigonometric bases. The construction can be obtained from sine or cosine bases which are referred to as local sine packets and local cosine packets (LCP), respectively. Such transforms partition the time axis rather than the frequency axis by using smooth bells [20]. The bells are constructed using cutoff functions $r(t)$ that satisfy

$$|r(t)^2| + |r(-t)^2| = 1 \quad \text{for all } t \in R;$$
$$r(t) = \begin{cases} 0, & \text{if } t \leqslant -1 \\ 1, & \text{if } t \geqslant 1, \end{cases} \tag{2}$$

in order to preserve the orthogonality. An example of such a function $r(t)$ is

$$r(t) = \begin{cases} 0, & \text{if } t \leqslant -1, \\ \sin\left[\frac{\pi}{4}\left(1 + \sin\left(\frac{\pi t}{2}\right)\right)\right], & \text{if } -1 < t < 1, \\ 1, & \text{if } t \geqslant 1. \end{cases} \tag{3}$$

Now, consider the partitioning of the time axis that produces a set of intervals $[a_j, a_{j+1}]$. Interval $[a_j, a_{j+1}]$ describes a partition of length $l_j$, where $l_j = a_{j+1} - a_j$. Let $\gamma \leqslant l_i/2$ be the length of the overlap between adjacent intervals. As we explain below, LCP are used to calculate a local discriminant basis (LDB) and adaptively segment the EEG signals along the time dimension. To this end, we set $\gamma$ to half the length of the desired shortest time segment, i.e., the segment that corresponds to the deepest node in the tree representation of the LDB to preserve orthogonality. The smooth window function is then defined as

$$w(t) = \begin{cases} r\left(\frac{t-a_j}{\gamma}\right), & t \in [a_j - \gamma, a_j + \gamma], \\ 1, & t \in [a_j + \gamma, a_{j+1} - \gamma], \\ r\left(\frac{a_{j+1}-t}{\gamma}\right), & t \in [a_{j+1} - \gamma, a_{j+1} + \gamma]. \end{cases} \tag{4}$$

Each window overlaps its neighboring windows and satisfies

$$w_{j-1}^2(t) + w_j^2(t) + w_{j+1}^2(t) = 1 \tag{5}$$

for $t \in [a_j - \gamma, a_{j+1} + \gamma]$, as shown in figure 4.

It can be shown that the set of functions

$$\psi_k^j(t) = w_j(t)\frac{\sqrt{2}}{\sqrt{l_j}} \cos\frac{\pi}{l_j}\left(k + \frac{1}{2}\right)(t - a_j) \quad j \in Z,\ k \in N \tag{6}$$

are an orthonormal basis for $L^2(R)$. A signal $s(t) \in L^2(R)$ can therefore be written in terms of the functions $\psi_k^j(t)$ as

$$s(t) = \sum_{j \in Z, k \in N} c_k^j \psi_k^j(t), \tag{7}$$

where $c_k^j = \langle s(t) \cdot \psi_k^j(t)\rangle$. The coefficients $c_k^j$ can be computed with the fast discrete cosine transform (type IV) after a preliminary folding step [20].

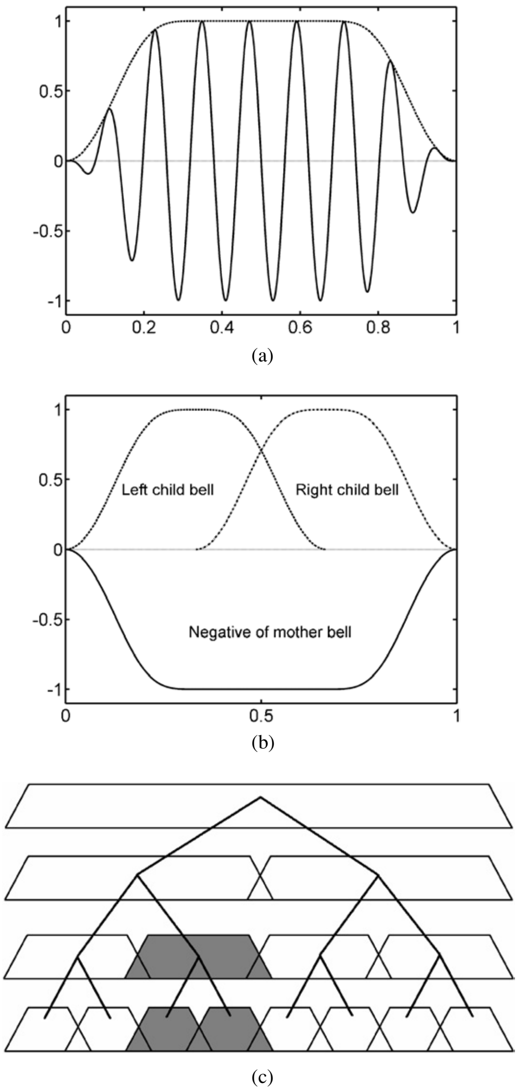


**Figure 4.** (a) The smooth local cosine. (b) The orthonormal smooth windows which represent each subspace on the tree structure are also marked in (c). (c) The dyadic tree that is used for discriminant base search.

*4.1.3. Local discriminant bases.*

*Dyadic local discriminant bases.* Since the MI-related ERD/ERS patterns appear as time locked transient phenomena, it is crucial to focus on local properties of the EEG signal. The best base (BB) algorithm was developed to extract such local information [21]. The BB algorithm expands the signal into orthonormal bases by using wavelet packets (WP) or local trigonometric bases over a dyadic tree (see figure 4(c)). This full tree is pruned to minimize a cost function, such as entropy, by a divide and conquer algorithm. Traditionally, the BB method has been used for signal representation. Since the selected cost function is entropy, such an algorithm ends up selecting the basis that provides good signal compression. For classification, however, the discrimination power of the nodes is what matters. Therefore, we replace the entropy criterion by a cost function that can measure the distance between matching nodes in the trees derived from signals corresponding to left or right imagery movements. By pruning the dyadic tree to maximize the selected distance criterion, an LDB is extracted [22]. In [19] several types of LDB were proposed over a dyadic tree by using wavelets or local trigonometric bases. The LDB-I procedure used the mean energy of the expansion coefficients to capture the distance between classes. However, sometimes the mean may not capture real discriminant information. The LDB-II uses the distance between probability densities (pdf) or cumulative density functions (cdf) of the expansion coefficients corresponding to signals in different classes. It is less affected by energy variations between the sub-bands of EEG signals than the LDB-I method. The LDB-II algorithm can be summarized as follows.

*Step 1.* Expand each training signal into a wavelet or cosine packets representation over a dyadic tree.

*Step 2.* For each expansion coefficient (node in the tree), calculate the distance between classes as the distance between the pdf or cdf of the coefficient obtained from signals corresponding to the two different classes. Accumulate the distances of the expansion coefficients in each subspace.

*Step 3.* Prune the tree from bottom to top by maximizing the distance between the two classes.

*Step 4.* Order the expansion coefficients from the pruned tree by using a class seperability criterion (CS). Select the top $k \ll n$ coefficients for classification, where $n$ is the number of samples in each training signal.

Let $p$ and $q$ represent the probability distributions of each expansion coefficient estimated via histogram. We have used the Euclidean distance given by

$$D(p, q) = \|p_i - q_i\|^2 = \sum_{i=1}^{n}(p_i - q_i)^2 \tag{8}$$

for pruning the tree. We considered the Fisher class separability criterion

$$F = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \tag{9}$$

for ordering the features. In (2), $\mu$ and $\sigma$ are the mean and the standard deviation of the feature they correspond to.

*Non-dyadic LDB—merge/divide.* The original LDB uses a dyadic tree structure as explained above. However, subject-specific EEG patterns will not necessarily fall along the dyadic segments produced by the original LDB method. We therefore modify the time segmentation procedure by adopting a merge/divide strategy that is used in geophysical waveform compression [23]. We illustrate the algorithm with the schematic diagram given in figure 5. Assume that the signals shown at the top of the figure are recorded from the same electrode location during task A and B. Instead of building a tree structure as described above, the signal is analyzed from left to right with three windows. The windows have the traditional children and mother structure. That is, the union of children windows covers the same time interval

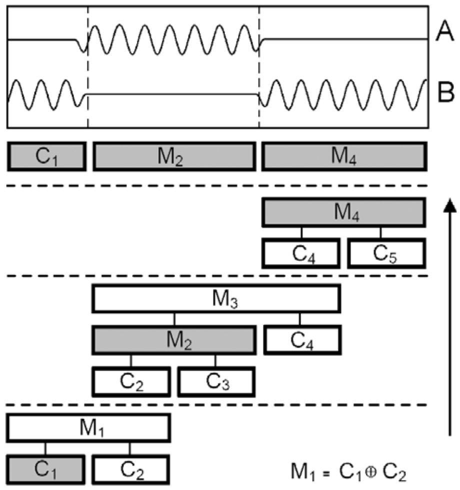


**Figure 5.** A schematic illustration of the merge/divide algorithm to locate the segments where classes A and B are well separated. Note that the rectangular segments correspond to the smooth windows given in (4).

covered by the mother window. The children windows that match the nodes at the bottom of the tree correspond to a LCP representation. The distance between the expansion coefficients in the mother and children subspaces ($M_1$, $C_1$ and $C_2$) are then compared. Whenever the distance between the coefficients corresponding to the left and right imagery signals in the children subspaces is greater than the distance between the coefficients corresponding to the left and right imagery signals in the mother subspace, we retain the partitioning of the time axis corresponding to the children windows, i.e., the signal is divided at that point. Otherwise, we retain the time interval corresponding to the mother window. If we retain the mother window, the mother segment is used as the left child in the next iteration ($M_2$ in figure 5).

Note that, unlike the dyadic case, the children windows are *not* necessarily half the length of the mother window. The right child is *always* the time segment of smallest size used by the procedure and will have a fixed length. The length of the left child can grow to be multiples of that of the basic smallest segment. Hence, unlike the dyadic case, there is no limitation on the length of the resulting time segments.

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 time window (cell) and construct a children mother decomposition structure.

*Step 2.* In each sub-space, expand the signal into cosine packets. For each expansion coefficient, calculate the distance between the two classes and accumulate the distances of the 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.* Order the expansion coefficients of the segmented signal by using a class separability criterion. Select the top $k \ll n$ coefficients for classification, where $n$ is the number of samples in each training signal.

*4.1.4. Frequency band adaptation.* The LDB algorithm produces a feature space with high dimensionality equal to the dimension of the original space. A high dimension feature space can decrease the generalization capability of the classifier when a limited number of training samples exist—this is the curse of dimensionality [24]. In addition, during our studies we observed that the center frequency of the oscillations differ from sweep to sweep. This increases uncertainty along the frequency axis.

To deal with these issues, in our prior work, we grouped the expansion coefficients in 4 Hz frequency bins between 0 and 40 Hz in each segment. Such an approach has direct connections with the Mel-scale approach used in acoustic signal processing and classification [25] and resulted in significant classification performance improvement over the case where individual expansion coefficients are used [11]. However, we have found experimentally that this approach does not necessarily maximize the discrimination capability of the retained time–frequency features. To deal with this limitation, we elected in this study to use adaptive partitions along the frequency axis. Specifically, within each time segment, 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. Similar approaches were utilized to select relevant bands in near infrared spectroscopy [26].

The procedure can be summarized as follows.

*Step 1.* Construct adaptive time segmentation by using the merge/divide procedure.

*Step 2.* In each time segment, calculate the distance between the pdf of frequency coefficients corresponding to the left and right MI when the coefficients are considered individually and again after they are merged.

*Step 3.* Merge coefficients when the merger leads to a larger distance between the classes.

*Step 4.* Choose coefficients from the constructed arbitrary time–frequency segmentation and order them according to a given class separability criterion. Select the top $k \ll n$ coefficients for classification, where $n$ is the number of samples in each training signal.

The resulting tilings obtained by combining the LDB and clustering approaches result in an arbitrary segmented time–frequency plane (cf figure 6(c)). The segmentation is reminiscent of that of [27]. In [27] a double tree approach was used. Unlike our procedure, the tree in [27] was also limited to a dyadic grid.

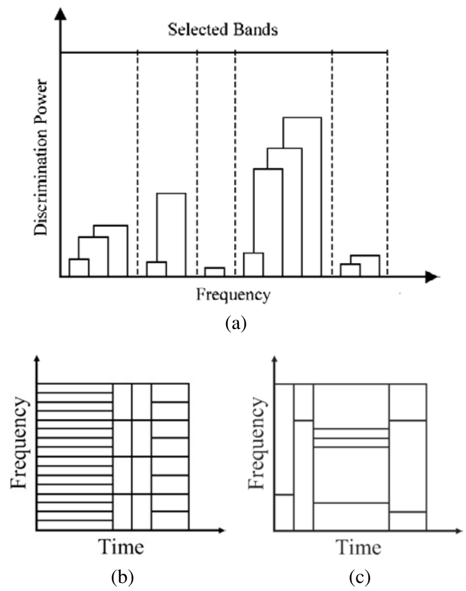


**Figure 6.** (a) Illustration of the frequency axis clustering that is obtained by merging consecutive expansion coefficients in each segment. (b) Example of a time frequency tiling constructed with a dyadic tree. (c) Example of a non-dyadic/flexible segmentation followed by frequency axis clustering. Note that the segmentation in (c) cannot be obtained with the original LDB.

### 4.2. Dimension reduction and classification

The adaptive time–frequency coefficients selected by the merge/divide and frequency domain clustering procedures may be correlated. A correlated feature set may decrease the generalization capability of the resulting classifier. Principal component analysis (PCA) was used in [28] on wavelet packet coefficients for dimension reduction in the classification of electromyogram signals. Here we use PCA for dimensionality and correlation reduction. We project the highest ranked set of ordered features onto the eigenvectors computed using a PCA analysis of the set. Specifically, let $x$ be a feature vector, $\lambda_k$ and $W_k$ the $k$th eigenvector of the covariance matrix of the feature set and $u_k$ the projection of $x$ onto $W_k$. We compute $u_k$ as

$$u_k = W_k^T x. \tag{10}$$

We then retain the projected features $u_k$ that correspond to the largest $N$ eigenvalues $\lambda_k$. In our work, we typically selected $N$ in the range of 32 to 48 because we observed that most of the discrimination power is concentrated in this range. Finally, the resulting reduced feature set is used for classification as an input to a Fisher linear discriminant analysis (LDA). The LDA is one of the most commonly used statistical classifiers due to its ease of implementation and trainability. It can be viewed as a method for identifying the best discriminating hyper plane in an $n$-dimensional space. A detailed description can be found in [29]. The weight vector used in the LDA is calculated as

$$v = \left(\sum_1 + \sum_2\right)^{-1} (m_1 - m_2), \tag{11}$$

**Table 1.** The classification accuracy (%) of the different approaches discussed in section 5. The bottom row represent the average classification accuracy.

| Subjects | Dyadic LDB 4 Hz Bin | Merge/divide 4 Hz Bin | Merge/divide clustered bands | AAR ($p = 6$) LMS, mu = 0.010 | AAR ($p = 6$) RLS, mu = 0.012 |
|---|---|---|---|---|---|
| S1 | 82.4 | 81.8 | 83.6 | 81.5 | 77.6 |
| S2 | 91.2 | 92.7 | 92.6 | 88.5 | 87.8 |
| S3 | 65.4 | 70.2 | 70 | 67.5 | 70.2 |
| S4 | 66.2 | 68.2 | 70.7 | 71 | 71.8 |
| S5 | 77.1 | 77 | 77.8 | 71 | 70.2 |
| S6 | 81.9 | 85.5 | 87.2 | 78.5 | 75.2 |
| S7 | 87.5 | 88.3 | 89.7 | 80.5 | 80.8 |
| S8 | 62 | 66.1 | 70 | 67 | 70.6 |
| S9 | 80.3 | 84.4 | 83.7 | 81.4 | 78.4 |
| Average | 77.1 | 79.4 | 80.6 | 76.3 | 75.8 |

where $\Sigma_i$ and $m_i$ are the covariance matrix and mean of the feature vector corresponding to class $i$ respectively.

The distance of a feature vector to the discriminating hyper-plane is calculated as

$$d = v^T x - v_0, \qquad (12)$$

where $x$ is the feature vector, $v$ the weight vector in (12), $v_0$ the threshold and $d$ the calculated distance.

We close this section by noting that further reduction of the dimensionality of the feature space may be possible. Preliminary results indicate that some PCA reduced classification features may be further combined in nearly all the subjects we investigated while preserving or improving classification performance. Promising avenues for investigating further reduction of the dimensionality of the feature space include applying classifiers that select subsets of features, e.g., stepwise discriminant analysis or logistic regression and studying interactions between different cortical locations and time–frequency patterns.

## 5. Results

In order to assess the efficiency of the developed algorithm, we compared its performance against the adaptive autoregressive (AAR) model based approach for feature extraction. We chose the AAR approach because it is considered to be the golden standard by many researchers in the field. An AAR model is described by

$$y[n] = a_{i,n} y[n-1] + \cdots + a_{p,n} y[n-p] + w[n] \qquad (13)$$

where $y[n]$ is the output sequence and $p$ the model order. The coefficients $a_{i,n}$ are time changing model parameters and $w[n]$ is a white noise process with zero mean and variance $\sigma_x^2$. To capture three spectral peaks as is customarily done in the literature, we selected $p = 6$. We calculated the model parameters for every sample using both the least mean square (LMS) and recursive least square (RLS) approaches for the $C_3$ and $C_4$ electrodes. A learning rate (mu) of 0.010 for LMS and 0.012 for RLS algorithms was used. The model parameters obtained from the $C_3/C_4$ electrode locations resulted in a feature vector of dimension 12. We noticed that, for the datasets we used, the classification rate of the AAR approach

**Table 2.** The classification accuracy obtained by class seperability sorting only (CS $= F$) and PCA applied on top sorted CS features. NoF stands for the number of features used to achieve minimal classification.

| Subjects | CS Accuracy (%) | CS NoF | PCA Accuracy (%) | PCA NoF |
|---|---|---|---|---|
| S1 | 83.9 | 13 | 83.6 | 12 |
| S2 | 92.3 | 20 | 92.6 | 19 |
| S3 | 72.4 | 8 | 70 | 12 |
| S4 | 69.7 | 51 | 70.7 | 25 |
| S5 | 77.4 | 46 | 77.8 | 43 |
| S6 | 86.4 | 30 | 87.2 | 9 |
| S7 | 88.5 | 34 | 89.7 | 10 |
| S8 | 69.4 | 47 | 70 | 43 |
| S9 | 84 | 9 | 83.7 | 7 |
| Average | 80.4 | 29 | 80.6 | 20 |

was slightly better when the model parameters were computed using the LMS algorithm.

For the dyadic and flexible LDB algorithms we select an analysis window of 416 samples and a tree depth of 4. The cell size is taken to be equal to the time length of the deepest segment, which is 260 ms. We used ten times ten-fold cross validation to estimate the classification accuracy.

In order to study the performance of each unit of the proposed algorithm, we compared the classification results obtained with the following procedures:

- dyadic LDB with fixed band features,
- merge/divide with fixed band features,
- merge/divide with frequency axis clustering,
- classification without PCA.

Tables 1 and 2 show the classification accuracy for all nine subjects.

The flexible LDB that combines the merge/divide and frequency domain clustering algorithms outperformed all other methods. Performance improves as we move from the dyadic to the merge/divide approach combined with frequency domain clustering. This demonstrates the importance of adaptation in time and frequency in a flexible manner. Also the superior performance of the proposed approach over the dyadic LDB and AAR techniques is further evidenced by the results of the paired $t$-test ($p = 0.0011$, $p = 0.0021$,

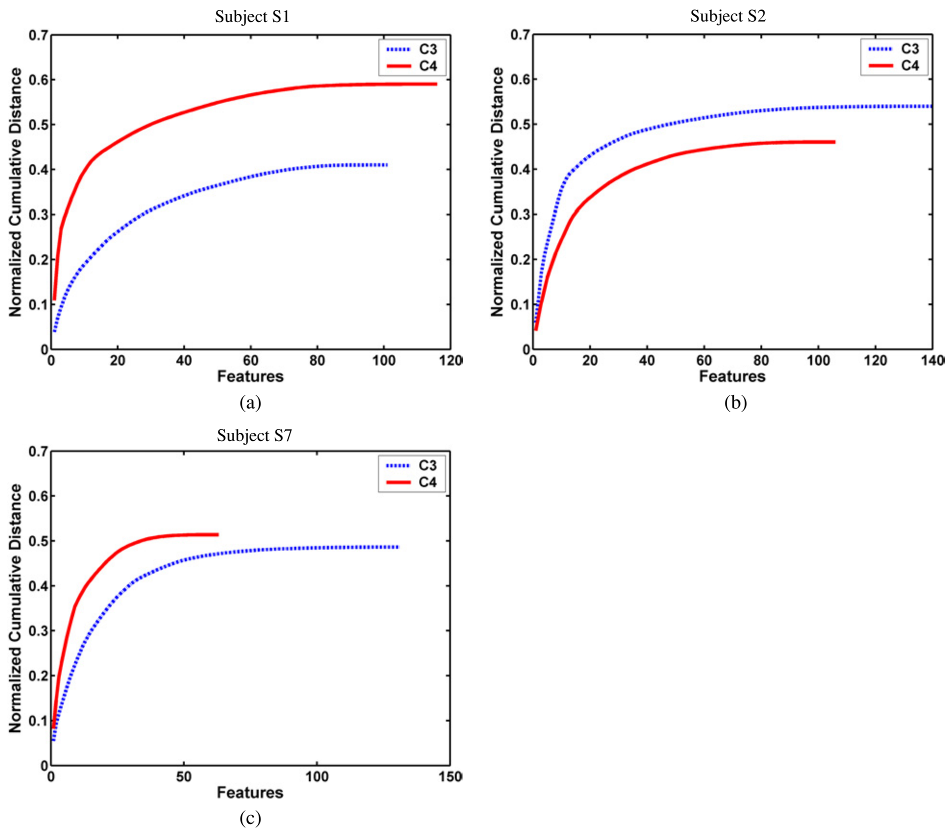


**Figure 7.** The discrimination power of the $C_3/C_4$ electrode locations of subjects S1 (a), S2 (b) and S7 (c). Note the differences between the total distances at each location. This supports a hemispheric asymmetry.

respectively). Finally, observe that the performance of our proposed approach is slightly better than that of Wang and He [31] on the dataset used in this paper, despite the fact that we use only 2 electrodes and Wang and He [31] use 20 electrodes.

We observed that our proposed algorithm constructs different time–frequency segmentations for each subject. Furthermore, it segments the EEG signals corresponding to activities in different hemispheres of the same subject differently. Figures 7 and 8 illustrate the discrimination power of each electrode and the time–frequency features selected for three representative subjects. It is well known that left and right motor cerebral cortices control contralateral finger movements. As mentioned earlier, this behavior is indeed the reason that MI-related ERD/ERS changes are used in some BCI systems. In the past, many articles assumed that the ERD/ERS patterns in the two hemispheres corresponding to left hand imagery are mirror images of those corresponding to right hand imagery. However, our current findings, as illustrated by figures 7 and 8, do not support this assumption. Since an equal number of features from both channels are included in the conventional AR and band power methods, these techniques are unable to model hemispheric asymmetry.

Since we only used two electrodes in this study, we were unable to expand our findings and investigate the effect of different cortical areas.

Note also that figure 8 indicates that most of the time wide bands are selected. For instance, although subject S7 has a very strong alpha activity in the C3 electrode, the beta band is selected as the most discriminant feature. It appears that all oscillatory components do not have the same discrimination power. Since AR based methods only model the peaks of the spectrum, all energetic components are directly included in the feature set.

Careful analysis of the ERD/ERS activities of the various subjects and the time–frequency segmentations produced by our procedure indicates that the arbitrary time segmentation created by the merge/divide approach is responsible for the significant improvement in classification accuracy on subjects S3, S4, S6 and S8.

The LDB algorithm of course is not without shortcomings. Both the dyadic and flexible methods suffer from the effects of small expansion coefficients and noise mentioned in [19]. The same information is often repeated in the side lobes or in high frequency coefficients. The clustering procedure partially groups the correlated features. Unfortunately the effect of side lobe artifacts is still present. This is the main reason for using PCA prior to LDA. Principal components analysis can remove correlations between features that might be repeated in side lobes. Therefore, classification remains constant or improves when using a smaller number of PCA features (see table 2). On the other hand, PCA is very sensitive to outliers. In particular, the lower components show significant energy changes from

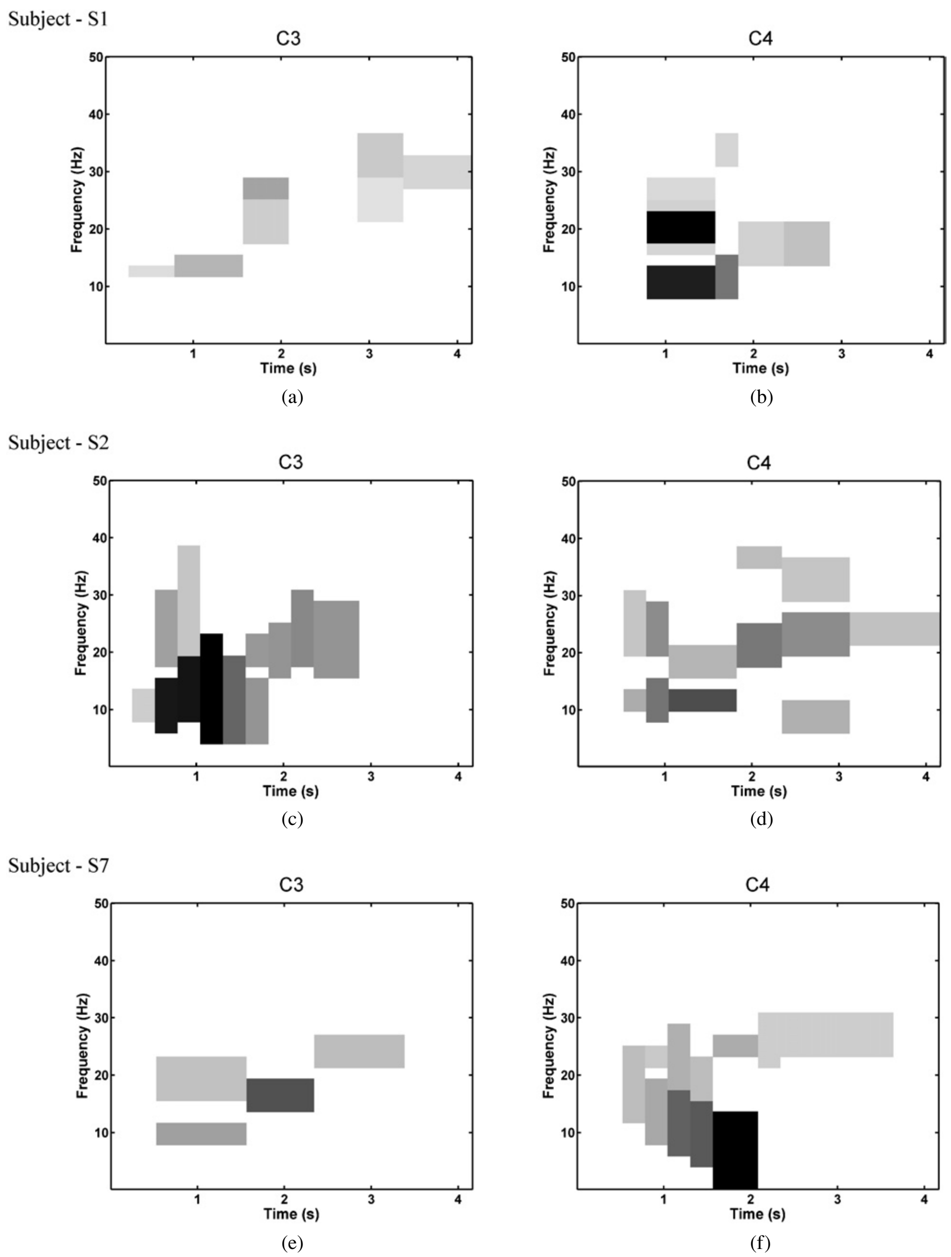


**Figure 8.** The time–frequency features of three representative subjects obtained from the $C_3$ (a), (c), (e) and $C_4$ (b), (d), (f) electrode locations. The darker locations have more discrimination power. Note the differences between the tilings and feature characteristics and their discrimination power in each hemisphere.

sweep to sweep. This may eliminate the advantages of using PCA.

Note also that one could apply a denoising procedure to the EEG signals prior to classification. However, the imbalance between the energy in the EEG bands makes it hard to select a suitable denoising procedure. An inappropriate threshold can remove a component which carries significant information.

The occurrence of ERD and ERS events on the same electrode in different bands was previously reported in [4]. Here we observe similar structures. Our algorithm successfully adjusted the adaptive time–frequency segmentation maps to the multi-component nature of the EEG. (cf figure 8). We also visually observed that the temporal variability of beta activity is less than that of the alpha activity. We noted poor classification for subjects with only alpha ERD/ERS. This may be due to our assumption that all components exhibit time locked behavior. Algorithms which are less sensitive to temporal variations, such as hidden Markov models, may give better results with these subjects.

## 6. Conclusion

We introduced an adaptive arbitrary time–frequency segmentation and feature selection approach for the classification of single trial MI-related EEG recordings. The segmentations and feature characteristics calculated by our procedure vary from subject to subject and depend on the hemisphere in which activity occurs for each subject. Our results therefore support a hemispheric asymmetric behavior. This also agrees with the study in [30] where a hidden Markov model and a genetic algorithm were combined to assess hemispheric asymmetry in an MI task.

The online computational complexity of our procedure is very low. The segmentation is implemented offline during the learning stage. During the online phase, the algorithm simply calculates and merges the smooth DCT-IV on the selected intervals. The expansion coefficients are then classified with a linear discriminant. Experimental results demonstrate the ability of our procedure to adapt to inter-subject variability and physio-anatomical differences, establish its low computational cost and clearly indicate that heterogeneous time–frequency features can improve classification results. Note however that our procedure uses 2 to 3 seconds worth of data after the instant at which movement is imagined to make its classification decision. This latency may be excessive in certain BCI tasks and future research should focus on reducing this delay.

## Acknowledgments

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

## References

[1] Wolpaw J R, Birbaumer N, William J H, Dennis J M, Hunter P, Gerwin S, Donchin E, Luis A Q, Charles J R and Vaughan T M 2000 Brain-computer interface technology: a review of the first international meeting *IEEE Trans. Rehabil. Eng.* **8** 164–73

[2] Wessberg J, Stambaugh C, Kralik J, Beck P, Laubach M, Chapin J, Kim J, Biggs S, Srinivasan M and Nicolelis M 2002 Real-time prediction of hand trajectory by ensembles of cortical neurons in primates *Nature* **408** 361–5

[3] Wolpaw J R and McFarland D J 1994 Multichannel EEG-based brain-computer communication *Electroencephalogr. Clin. Neurophysiol.* **90** 444–9

[4] Pfurtscheller G and Neuper C 2001 Motor imagery and direct brain-computer interface *Proc. IEEE* **89** 1123–34

[5] Lopes da Silva F H and Pfurtscheller G 1999 Basic concepts on EEG synchronization and desynchronization *Event Related Desynchronization. Handbook of Electroencephalography and Clinical Neurophysiology* ed G Pfurtscheller and F H Lopes da Silva (Amsterdam: Elsevier)

[6] Neuper C and Pfurtscheller G 2001 Event-related dynamics of cortical rhythms: frequency-specific features and functional correlates *Int. J. Psychophysiol.* **43** 41–58

[7] Kalcher J, Flotzinger D and Pfurtscheller G 1993 Brain-computer interface—a new communication device for handicapped persons *J. Microcomput. Appl.* **16** 293–9

[8] Prezenger M and Pfurtscheller G 1999 Frequency component selection for an EEG-based brain computer interface *IEEE Trans. Rehabil. Eng.* **7** 413–9

[9] Schlögl A, Flotzinger D and Pfurtscheller G 1997 Adaptive autoregressive modeling used for single trial EEG classification *Biomed. Tech.* **42** 162–7

[10] Pfurtscheller G, Neuper C, Schlogl A and Lugger K 1998 Separability of left and right hand movement imageries with adaptive autoregressive modeling *IEEE Trans Rehabil. Eng.* **6** 316–25

[11] Ince N F, Tewfik A and Arica S 2005 Classification of movement EEG with local discriminant bases *ICASSP: 30th Int. Conf. on Acoustics, Speech and Signal Processing* (Philadelphia, PA: IEEE)

[12] Ince N F, Tewfik A, Arica S and Yagcioglu S 2005 Analysis and visualization of movement related EEG activities using local discriminant bases *2nd Int. IEEE EMBS Conf. on Neural Eng.* (Washington, DC)

[13] Osman A and Robert A 2001 Time-course of cortical activation during overt and imagined movements *Proc. Cognitive Neuroscience Ann. Meet* (New York)

[14] Hjorth B 1975 An online transformation of EEG scalp potentials into orthogonal source derivations *Electroencephalogr. Clin. Neurophysiol.* **39** 526–30

[15] Wesfreid E and Wicherhauser M 1993 Adapted trigonometric transform and speech processing *IEEE Trans. Acoust. Speech Process.* **41** 3596–600

[16] Saito N 1996 Classification of geophysical acoustic waveforms using time–frequency atoms *Am. Stat. Assoc. Statistical Computing Proc.*

[17] Mallat S 2000 *A Wavelet Tour of Signal Processing* 2nd edn (New York: Academic)

[18] Coifman R R and Donoho D 1995 Translation invariant de-noising *Wavelets and Statistics (Lecture Notes in Statistics)* ed A Antoniadis and G Oppenheim (Berlin: Springer) pp 125–50

[19] Saito N, Coifman R R, Geshwind F B and Warner F 2002 Discriminant feature extraction using empirical probability density and a local basis library *Pattern Recognit.* **35** 1842–52

[20] Wickerhauser M V 1994 *Adapted Wavelet Analysis from Theory to Software* (Wellesley, MA: A K Peters)

[21] Wickerhauser M V and Coifman R R 1992 Entropy based algorithms for best basis selection *IEEE Trans. Inform. Theory* **32** 712–18

[22] Saito N 1994 Local feature extraction using a library of bases *Dissertation* Yale University

[23] Wu R S and Wang Y 1999 New flexible segmentation technique in seismic data compression using local cosine transform *Proc. SPIE* **3813** 784–94

[24] Bishop C M 1995 *Neural Networks for Pattern Recognition* (Oxford: Clarendon)

[25] Cetin A E, Pearson T C and Tewfik A H 2004 Classification of closed- and open-shell pistachio nuts using impact acoustical analysis *ICASSP: Int. Conf. on Acoustics, Speech and Signal Processing* (Philadelphia, PA: IEEE)

[26] Akgül C B, Sankur B and Akın A 2005 Spectral analysis of event-related hemodynamic responses in functional near infrared spectroscopy *J. Comput. Neurosci.* **18** 67–83

[27] Herley C, Kovacevic J, Ramchandran K and Vetterli M 1993 Tilings of the time–frequency plane: construction arbitrary orthogonal bases and fast tiling algorithms *IEEE Trans. Signal Process.* **41** 3341–59

[28] Englehart K, Hudgins B, Parker P A and Stevenson M 1999 Classification of the myoelectric signal using time–frequency based representations *Med. Eng. Phys.* **21** 431–8

[29] Duda R, Hart P and Stork D 2000 *Pattern classification* 2nd edn (New York: Wiley)

[30] Obermaier B, Munteanu C and Pfurtscheller G 2001 Asymmetric hemisphere modeling in an offline brain computer interface *IEEE Trans. Syst. Man Cybern. C* **4** 536–40

[31] Wang T and He B 2004 Classifying EEG-based motor imagery tasks by means of time–frequency synthesized spatial patterns *Clin. Neurophysiol.* **115** 2744–53