

# Discriminative Methods for Classification of Asynchronous Imaginary Motor Tasks from EEG Data

Jaime F. Delgado Saa, *Student member, IEEE*, Mijdat Çetin, *Member, IEEE*,

**Abstract**—In this work, two methods based on statistical models that address temporal changes in the Electroencephalographic (EEG) signal are proposed for asynchronous brain computer interfaces (BCI) based on imaginary motor tasks. Unlike current approaches to asynchronous BCI systems that use windowed versions of EEG data combined with static classifiers, the methods proposed here are based on discriminative models that allow sequential labeling of data. In particular, the two methods we propose for asynchronous BCI are based on Conditional Random Fields (CRF) and Latent Dynamic CRFs (LDCRF), respectively. We describe how the asynchronous BCI problem can be posed as a classification problem based on CRFs or LDCRFs, by defining appropriate random variables and their relationships. CRF allows modeling the extrinsic dynamics of data, enabling modeling the transitions between classes, which in this context correspond to distinct tasks in an asynchronous BCI system. On the other hand, LDCRF surpasses this approach by incorporating latent variables that permit modeling the intrinsic structure for each class and at the same time allows modeling extrinsic dynamics. We apply our proposed methods on the publicly available BCI competition III dataset V. Results are compared to the top algorithm in the BCI competition as well as to methods based on Hierarchical Hidden Markov Models (HHMM), Hierarchical Hidden CRF (HHCRF), neural networks based on particle swarm optimization (IPSONN), and to a recently proposed approach based on neural networks and fuzzy theory, the S-dFasArt. Our experimental analysis demonstrates the improvements provided by our proposed methods in terms of classification accuracy.

**Index Terms**—Brain computer interface, sequential labeling, brain states, sensorimotor rhythms, imaginary motor tasks, discriminative models, conditional random fields.

## I. INTRODUCTION

**B**RAIN computer interfaces (BCI) are systems that provide an alternative non-muscular communication path for people who suffer severe muscular disabilities resulting from disease or accident [1]. Moreover, BCIs have found application for healthy subjects in multimedia and gaming in recent years [2]. BCIs use brain signals generated by various physiological mechanisms such as slow cortical potentials, sensorimotor rhythms, P300 potentials and steady-state visually evoked potentials [3], to provide control over the environment through a computer. In the case of non-invasive BCI systems based on electroencephalographic (EEG) signals, two types of BCI systems are used: synchronous and asynchronous. In a synchronous BCI approach, the subject receives cues that indicate when the mental task should be executed. Although this approach can be appropriate for laboratory research, it is not useful for most real life

applications in which the subject will need to control the interface continuously without cues or temporal constraints for execution of the mental task. A BCI system operating in this manner is called asynchronous. Most existing pieces of work on asynchronous systems use windowed EEG signals (or features of the EEG signals) and static classifiers (e.g., LDA, Gaussian classifiers, neural networks) [4], [5], [6], [7], [8], [9], [10], [11]. In those approaches, the difference of power in the EEG signals in different frequency bands is used to determine the subject's intention. Other research involves the detection of transitions between tasks by identification of abrupt changes in the estimated power densities of the EEG signals [12], [10]. This so-called mental task transition detector offers increased performance in the classification accuracy of EEG signals [12], [10]. However, the temporal structure of the EEG signal which has been shown to increase the performance of the synchronous BCI systems [13], [14], [15], [16], [17], [18] has not been not exploited.

In an asynchronous scenario, the subjects execute different mental tasks without cues, so when the subjects start the execution of a specific task is unknown. In this case, the problem is labeling sequential data. Statistical models such as hidden Markov models (HMM) and conditional random fields (CRF) have been used with success in other fields such as gesture recognition and natural language processing [19], [20], [21], [22]. Given that CRFs can in principle be used to model the dynamics of sequential data, they are attractive for asynchronous BCI applications. However, although CRFs can model the extrinsic dynamics of the data (or features), which in asynchronous BCI corresponds to dynamics across different tasks, CRFs lack the ability to model intrinsic dynamics, i.e., the temporal evolution in the course of execution of a particular task. Physiological theory indicates that different states in the human brain emerge during the execution of mental tasks and those states are observed in the EEG signal through the well known phenomena of event related synchronization and de-synchronization (ERS/ERD) [23]. Several studies have attempted to capture that structure through various random process models. Of particular interest is a method capable of modeling the intrinsic structure, proposed by Sugiura et al. [24]. This method is based on hierarchical hidden CRFs (HHCRFs), which generalize the hidden conditional random field (HCRF) model of [22]. Sugiura et al. apply HHCRF to EEG signal segmentation in an asynchronous BCI application and demonstrate the

performance improvements it provides over the generative counterpart, the hierarchical HMM [25], [26]. Sugiura et al.'s work shares certain aspects of our research. In particular, similar to our work, Sugiura et al.'s work also involves a discriminative model for asynchronous BCI. However, their model focuses on building the hierarchy of various state variables and leads to a rather complicated structure requiring an extra level involving indicator variables. We propose that nature of the asynchronous BCI problem can be effectively captured by a simpler discriminative model, as presented in our work. We experimentally demonstrate the advantages offered by our model over that proposed by Sugiura et al. in Section V. Another algorithm used for classification of temporal patterns is presented in a recent work by Cano et al. [27]. This algorithm is based on neural networks and fuzzy theory, the S-dFasArt. Cano et al. show that the S-dFasArt algorithm provides an improvement in the classification rate of spontaneous mental activity by using the dataset V of the BCI competition III.

A method that provides the combined advantages of CRF with the use of hidden states has been proposed by Morency et al. for gesture recognition [28]. The so called latent dynamic CRF (LDCRF) allows modeling extrinsic dynamics of the sequential data as well as the intrinsic dynamics within each class by means of hidden states. This approach permits modeling different states during the execution of a specific mental task and at the same time modeling transitions between different mental tasks. Given these features, LDCRF can be applied directly to sequential data avoiding the need for windowing the signal. In this work two methods for asynchronous BCI, one based on CRF and another on LDCRF, are presented. For CRF the nodes in the model represent the mental task executed by the user. For LDCRF, hidden variables are incorporated and represent different states that occur during the execution of a specific task. Nodes in a second layer of the graph represent different mental tasks. We use surface Laplacian filters to obtain the signals over centro-parietal electrode positions and power spectral densities of the signals in specific frequency bands are used as features. Feature selection is performed by sequential floating forward selection (SFFS), producing an optimal set of features used as input to the CRF-based and LDCRF-based classifiers. The Dataset V of the BCI competition III has been used. We compare the performance of our proposed methods with the BCI competition winner algorithm as well as methods recently proposed by Sugiura et al. [20], Cano et al. [3], and Lin et al. [11]. The superiority of our proposed methods is evidenced by the higher levels of classification accuracy they provide.

## II. CONDITIONAL RANDOM FIELDS

CRFs are discriminative graphical models. Lafferty et al. [29] define the probability of a particular label sequence  $\mathbf{y} = \{y_1, y_2, \dots, y_m\}$  given an observation sequence  $\mathbf{x} = \{x_1, x_2, \dots, x_m\}$  with  $x_j \in \mathcal{R}^d$  to be of the form:

$$P_{\theta}(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp\left\{ \sum_{l \in L_1} \sum_{j=1}^m f_{1,l}(y_{j-1}, y_j, \mathbf{x}, j) \theta_{1,l} + \sum_{l \in L_2} \sum_{j=1}^m f_{2,l}(y_j, \mathbf{x}, j) \theta_{2,l} \right\} \quad (1)$$

where  $f_{1,l}$  and  $f_{2,l}$  are feature functions related to the edges and nodes of the graph, respectively; both functions are given and fixed.  $L_1$  and  $L_2$  are the set of indices for the feature functions related to the edges and nodes respectively (see Figure 1). The feature functions are real-valued and express sufficient statistics describing their arguments and relationships.

The conditional probability expressed in (1) can be simplified by writing:

$$P_{\theta}(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp\left\{ \sum_{l \in L} \sum_{j=1}^m f_l(y_{j-1}, y_j, \mathbf{x}, j) \theta_l \right\} \quad (2)$$

where  $L$  is a set of indices for the feature functions, each  $f_l(y_{j-1}, y_j, \mathbf{x}, j)$  is either a state (node) function of a transition (edge) function and  $Z(\mathbf{x})$  is a normalization factor.

In an asynchronous BCI scenario with reference to Figure 1(a), the observation sequence  $\mathbf{x}$  corresponds to EEG features and each element  $y_j$  of the label sequence  $\mathbf{y}$  corresponds to the imagined mental/motor task (relax, right finger movement, left finger movement, mathematical mental operation, etc.) at time point  $j$ . Then the feature functions provide sufficient statistics for classification of motor tasks.

Parameter estimation in CRFs for a linear chain (considered here for BCI signals) can be performed through a maximum likelihood approach [29], as we describe next. Given independent identically distributed (i.i.d.) training data  $D = \{\mathbf{x}^{(i)}, \mathbf{y}^{(i)}\}_{i=1}^N$ , where  $\mathbf{x}^{(i)} = \{x_1^{(i)}, x_2^{(i)}, \dots, x_m^{(i)}\}$  is a sequence of inputs and each  $\mathbf{y}^{(i)} = \{y_1^{(i)}, y_2^{(i)}, \dots, y_m^{(i)}\}$  is a sequence of mental/motor task labels, the conditional log-likelihood of the training data can be expressed as follows:

$$l(\theta) = \sum_{i=1}^N \log P(\mathbf{y}^{(i)}|\mathbf{x}^{(i)}) - \frac{\theta^2}{2\sigma^2} \quad (3)$$

where the regularization term  $\frac{\theta^2}{2\sigma^2}$  is the log of a Gaussian prior with variance  $\sigma^2$ , that is  $P(\theta) = \exp(-\frac{1}{2\sigma^2} \|\theta\|^2)$ . By substituting (2) into (3) and including a regularization term as a measure to avoid over fitting [29] the following expression is obtained:

$$\begin{aligned} \tilde{l}(\theta) &= \sum_{i=1}^N \sum_{j=1}^m \sum_{l \in L} f_l(y_{j-1}^{(i)}, y_j^{(i)}, \mathbf{x}^{(i)}, j) \theta_l \\ &\quad - \sum_{i=1}^N \log Z(\mathbf{x}^{(i)}) - \sum_{l \in L} \frac{\theta_l^2}{2\sigma^2}. \end{aligned} \quad (4)$$

The parameters  $\theta_l$  which maximize the regularized conditional log-likelihood above can be found by iterative optimization methods. In our work, we use a quasi-Newton algorithm using Hessian updates based on the Broyden-Fletcher-Goldfarb-Shanno (BFGS) formula.

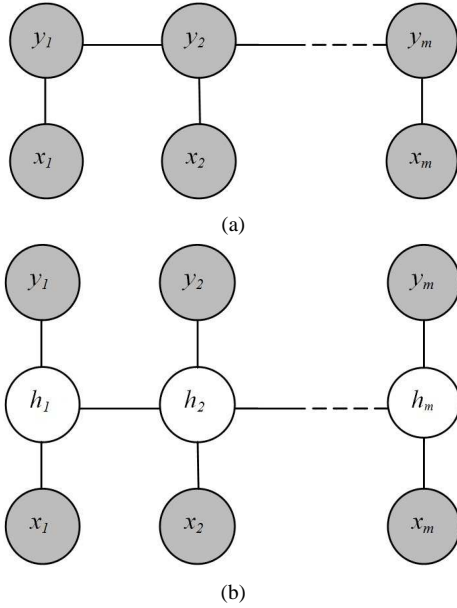


Fig. 1. (a) CRF model (b) LDCRF model. Shaded nodes represent observed variables in the training set. Although only one link between  $x_j$  and hidden nodes  $h$  is shown in the graph for simplicity, long range dependencies are also possible in these models.

### III. LATENT DYNAMICS CONDITIONAL RANDOM FIELDS

CRFs allow modeling transitions between classes, capturing the extrinsic dynamics of the EEG features, but cannot represent internal states for each class, an ability which increase the differentiability between classes. A model that incorporates the ability to capture both extrinsic and intrinsic dynamics is the Latent Dynamics CRF (LDCRF) proposed by Morency et al. [28]. By combining the strengths of CRF and Hidden conditional random fields (HCRF) [20]; LDCRF offers several advantages. As in CRF, LDCRF models the transitions between classes; and as in HCRF, includes hidden states allowing to model within class dynamics. These characteristics allow the LDCRF model to be directly applied for labeling unsegmented sequences.

In the application of LDCRF models to BCI, the task is to learn a mapping between a sequence of EEG features  $\mathbf{x} = \{x_1, x_2, \dots, x_m\}$  obtained during the subject's imagination of motor activity and a sequence of labels  $\mathbf{y} = \{y_1, y_2, \dots, y_m\}$  for the imaginary task executed; where each  $y_j$  is a class label for the  $j^{th}$  element of the sequence  $\mathbf{x}$  and is a member of the set  $\mathcal{Y}$  of possible class labels. LDCRFs also contain a vector of substructures  $\mathbf{h} = \{h_1, h_2, \dots, h_m\}$  which form a set of hidden variables in the model, because they are not observed in the training examples, and represent different mental states in the brain during the execution of each of the imaginary tasks.

Morency et al. define the latent conditional model:

$$P(\mathbf{y}|\mathbf{x}, \theta) = \sum_{\mathbf{h}} P(\mathbf{y}|\mathbf{h}, \mathbf{x}, \theta) P(\mathbf{h}|\mathbf{x}, \theta). \quad (5)$$

where  $\theta$  are the parameters of the model. In order to keep training and inference tractable, Morency et al. restrict the model to have disjoint sets of hidden states associated with each class. Then, the set of all possible states  $\mathcal{H}$  is the union of

all  $\mathcal{H}_y$  sets, where  $\mathcal{H}_y$  refers to the class-specific set of hidden states for class  $y$ . Under this assumption, the conditional probability in (5) can be written as:

$$P(\mathbf{y}|\mathbf{x}, \theta) = \sum_{\mathbf{h}: \forall h_j \in \mathcal{H}_y} P(\mathbf{h}|\mathbf{x}, \theta). \quad (6)$$

The equality on Equation 6 follows from the assumption of disjoint sets of hidden states, which produces  $P(\mathbf{y}|\mathbf{h}, \mathbf{x}, \theta) = 0$  for  $h_j \notin \mathcal{H}_y$  and  $P(\mathbf{y}|\mathbf{h}, \mathbf{x}, \theta) = 1$  for  $h_j \in \mathcal{H}_y$ . Using the usual conditional random field formulation:

$$P(\mathbf{h}|\mathbf{x}, \theta) = \frac{1}{Z(\mathbf{x}, \theta)} \exp\left\{\sum_l \mathbf{F}_l(\mathbf{h}, \mathbf{x}) \theta_l\right\}. \quad (7)$$

with  $\mathbf{F}_l$  defined as:

$$\mathbf{F}_l(\mathbf{h}, \mathbf{x}) = \sum_{j=1}^m f_l(h_{j-1}, h_j, \mathbf{x}, j) \quad (8)$$

Each feature function  $f_l(h_{j-1}, h_j, \mathbf{x}, j)$  as in the case of CRF is either a transition function or a state function.

The parameters of the LDCRF model can be learnt as is done for those in CRF by finding the optimal parameters  $\theta^*$  that maximize the objective function in Equation 3.

The feature functions in the LDCRF model correspond to transition and state feature functions. Note that transitions can be among hidden states within the same class (hence intrinsic) or among hidden states of different classes (hence extrinsic). Accordingly, weights associated with the hidden states in the same subset  $\mathcal{H}_y$  model the intrinsic dynamics while weights associated with hidden variables from different sets model the extrinsic dynamics. The number of transition functions in the model is given by the square of the cardinality of the set  $\mathcal{H}$ .

The number of state feature functions equals to the dimension of  $\mathbf{x}$  times the number of possible hidden states  $|\mathcal{H}|$ . Figure 1(b) shows a diagram for the LDCRF model where the input sequence  $\mathbf{x}$  corresponds to EEG features and the labels  $y_j$  represent the mental task executed. Given that  $\mathbf{x}$  and  $\mathbf{y}$  are observed in the training set, they are represented by shaded nodes in the graph of Figure 1(b).

### IV. DATA PROCESSING METHODOLOGY AND EXPERIMENTS

#### A. Problem and Dataset Description

This work uses the Dataset V of the BCI competition III. The dataset contains data from three normal subjects during four non-feedback sessions. The subject is requested to execute one out of three mental tasks: 1) Imagination of repetitive left hand movements, 2) Imagination of repetitive right hand movements, and 3) Generation of words beginning with the same random letter. The subject executes a mental task during fifteen seconds and then switches randomly to another task at the operator's request. For each subject, four sessions of four minutes length are available. The first three sessions are used for training; the fourth session, for testing. The data provide pre-computed features, obtained as follows. EEG signals are spatially filtered using a surface Laplacian filter and the power spectral density of these signals is calculated every 62.5 ms

using the last second of data. The power spectral density was calculated between 8Hz - 30Hz with a resolution of 2Hz over centro-parietal electrodes C3, Cz, C4, CP1, CP2, P3, Pz, and P4. As a result, the pre-computed feature vector for each temporal window is a 96-dimensional vector (8 channels  $\times$  12 frequency components).

### B. Feature Extraction and Selection

1) *Feature Extraction.*: Using the vector of pre-computed features, the average power across frequency in Alpha (8Hz - 12Hz), Sigma (12Hz - 16Hz), and Beta (18Hz - 26Hz) bands were computed for each of the eight electrodes. Figure 2 shows the topographic power distribution in the selected bands, for each subject. The topographic distribution shows, for each class and frequency band, the logarithm of the average power during the execution of each mental task (class), using all data available for each class in the training set. Differences in the amplitude of the signal provide information about the type of CRF-features and LDCRF-features that could be used, as will be discussed later. The frequency bands alpha, sigma, and beta, were selected because these rhythms are related to the well-known phenomena of ERS/ERD observed during the execution of mental tasks. This frequency band choice provides a new feature vector with 24 features, based on which we perform automatic feature selection for maximizing classification performance.

2) *Feature Selection.*: Feature selection is performed using the sequential floating forward selection algorithm (SFFS, see [30], [31], [32]). Given a set of features  $\mathcal{F} = \{f_1, f_2, \dots, f_D\}$ , we are interested in finding a new set  $\overline{\mathcal{F}}_k = \{\overline{f}_1, \overline{f}_2, \dots, \overline{f}_k\}$  such that  $k \leq D$ . Ideally the new set of features  $\overline{\mathcal{F}}$  increases the performance of the system or produces the same performance with a reduced number of features, and hence reduces the computational cost. The selection of the feature subset  $\overline{\mathcal{F}}_i$  from set  $\mathcal{F}$  is performed according to an objective function  $J(\overline{\mathcal{F}}_i)$ , where if  $J(\overline{\mathcal{F}}_i) > J(\overline{\mathcal{F}}_j)$  the subset  $\overline{\mathcal{F}}_i$  performs better than subset  $\overline{\mathcal{F}}_j$  does. SFFS adds sequentially a new feature from the original set to the output set according to the objective function. On each iteration the effect of removing each one of the previously selected features is evaluated. If one feature is found to reduce accuracy, that feature is removed to avoid the monotonic growth of the feature vector size, as encountered in sequential forward selection (SFS). In SFFS, we use the classification accuracy as the cost function based on three-fold cross-validation in the training data.

### C. Model Selection and Classification

1) *CRF model.*: For the case of linear-chain CRF, given a new input sequence  $\mathbf{x}$ , the most likely labeling  $\mathbf{y}^* = \arg \max_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})$  can be efficiently and exactly calculated by variants of dynamic programming algorithms for HMM, as described in [29]. The particular form we use for the conditional probability of the labels given the data is as

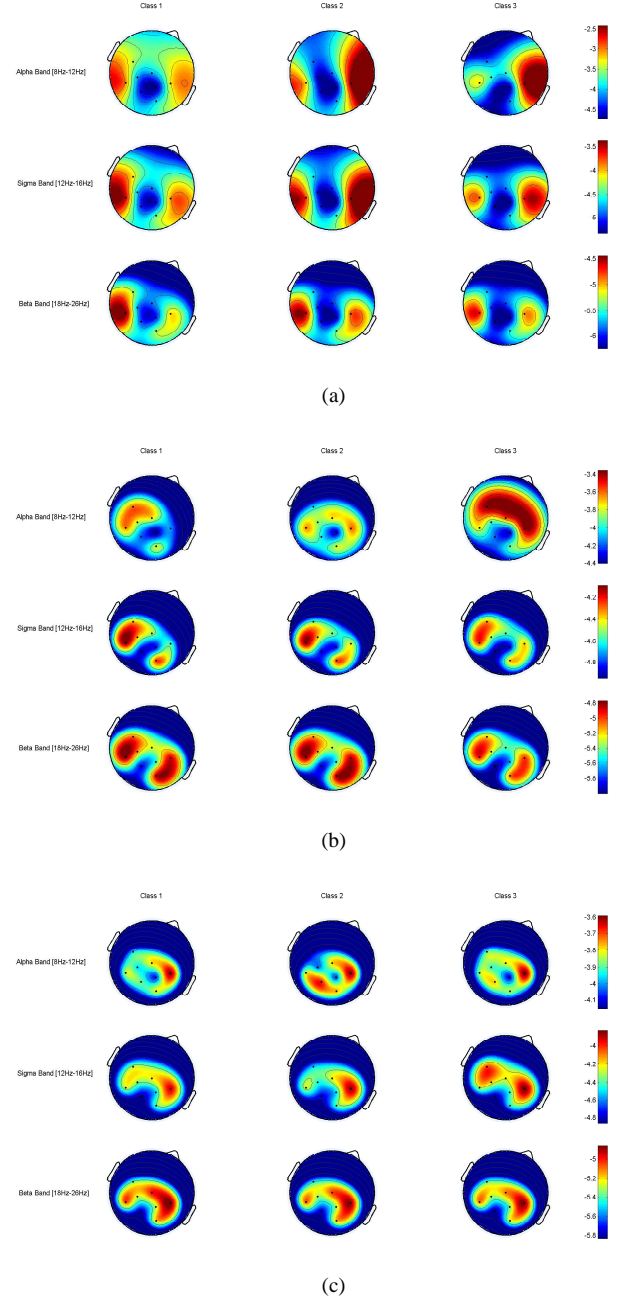


Fig. 2. Topographic distribution of power in different frequency bands (a) Subject 1. (b) Subject 2. (c) Subject 3.

follows:

$$P_{\theta}(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp \left\{ \sum_{j=1}^m f_{1,1}(y_{j-1}, y_j) \cdot \theta_{1,1} + \sum_{j=1}^m f_{2,1}(x_j) \cdot \theta_{2,1}[y] \right\} \quad (9)$$

The dot product  $f_{1,1}(y_{j-1}, y_j) \cdot \theta_{1,1}$  measures the compatibility of a transition from a particular motor task at  $j-1$  to the same or another motor task at  $j$ . Each element of the *edge weight* vector  $\theta_{1,1}$  contains a weight for a particular pair of labels. The feature function  $f_{1,1}(y_{j-1}, y_j)$  is an indicator vector, with a value of 1 for the entry corresponding to the particular set of



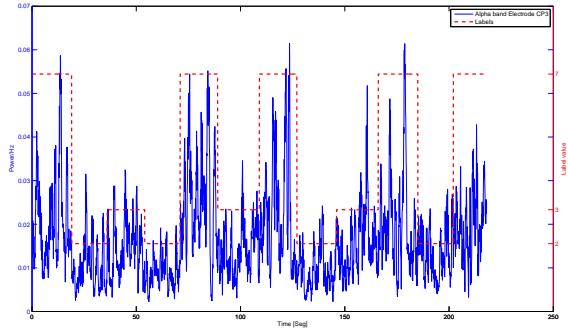


Fig. 3. EEG dynamics example for different classes. Differences between classes and also intra-class differences are observed. The signal corresponds to alpha band in electrode CP3.

values  $(y_{j-1}, y_j)$ , and 0 for all the other entries. The second term, which involves  $f_{2,1}(x_j) \cdot \theta_{2,1}[y]$  with  $f_{2,1}(x_j) = x_j$ , measures the compatibility between the current EEG feature  $x_j$  and the label  $y_j$ .

The class-dependent structure of the features as shown by the topographic distributions in Figure 2 suggest that the node compatibility function chosen in this manner has the potential for use in classification.

2) *LDCRF model.*: In the case of LDCRF, parameter selection is performed according to the description in Equation (4). The topographic power distributions shown in Figure 2 highlight the differences in power distribution when different motor tasks are executed. However, one can also observe temporal variations of power during the execution of a particular task. Figure 3 aims to display both phenomena. For the case of motor tasks, phenomena such as ERD and ERS explain the within class temporal variations. As observed in Figure 3, the magnitude of the signal is class-dependent but variations of the power during execution of the same task are also evident. The LDCRF model has the potential to fit and explain such data well, because LDCRFs are able to model extrinsic and intrinsic dynamics of the signal. Based on this, the feature functions are selected to obtain information about those dynamics. The conditional distribution of the labels given the data can be written as:

$$P(y|\mathbf{x}, \theta) = \sum_{\mathbf{h}: \forall h_j \in H_y} \frac{1}{Z(\mathbf{h}, \mathbf{x})} \exp\left\{ \sum_{j=1}^m f_1(h_{j-1}, h_j) \cdot \theta_1 + \sum_{j=1}^m f_2(x_j) \cdot \theta[h_j] \right\}. \quad (10)$$

were the dot product  $f_1(h_{j-1}, h_j) \cdot \theta_1$  measures the compatibility of the state transitions, where states could correspond to the same or different classes. Each element of the *edge weight* vector  $\theta_1$  contains a weight for a particular pairs of hidden states. The feature function  $f_1(h_{j-1}, h_j)$  is an indicator vector, with a value of 1 for the entry corresponding to the particular set of values  $(h_{j-1}, h_j)$ , and 0 for all the other entries. It is worth noting that this feature function models the

TABLE I  
CROSS VALIDATION RESULTS IN TRAINING DATA FOR THE PROPOSED CRF AND LDCRF BASED METHODS.

Subject	CRF(%)	LDCRF (%)	Hidden states (LDCRF)
B01	89.34	91.55	2
B02	78.08	83.89	2
B03	59.73	59.30	3

intrinsic dynamics by means of the weights associated with pairs of hidden states in the same subset  $\mathcal{H}_y$  and extrinsic dynamics by means of the weights associated with hidden states in different subsets. The second term, which involves the dot product  $f_2(x_j) \cdot \theta[h_j]$  with  $f_2(x_j) = x_j$  measures the compatibility of the current EEG feature  $x_j$  with the hidden state  $h_j$ .

For testing, given a new test sequence  $\mathbf{x}$ , we want to estimate the most probable sequence  $\mathbf{y}^*$  that maximizes the conditional model [28]:

$$\mathbf{y}^* = \arg \max_{\mathbf{y}} \sum_{\mathbf{h}: \forall h_i \in \mathcal{H}_y} P(\mathbf{h}|\mathbf{x}, \theta^*) \quad (11)$$

To estimate the label  $y_j^*$  of the element  $x_j$  of the sequence  $\mathbf{x}$ , the marginal probabilities  $P(h_j = a|\mathbf{x}, \theta^*)$  are evaluated for all possible hidden states  $a \in \mathcal{H}$ . Then the probabilities of hidden states corresponding to each distinct label are summed up, and the label corresponding to the maximum probability hidden state set is chosen. That is, assuming that the states are not shared across classes, the set of states with the higher global probability defines the label to be declared. The marginal probabilities mentioned above can be calculated by belief propagation [28], [33].

Our experiments, use three different models with 2,3, and 4 states per class. For each model SFFS is employed to select the optimal set of features and the accuracies in the three-fold cross - validation process in the training data are compared. The model which provides the best accuracy is selected for use in labeling the test data.

## V. RESULTS

Table I shows the classification accuracies obtained by cross-validation in the training set using CRF and LDCRF, as well as the number of states in the LDCRF model, that provides the best results. Table II shows the selected electrodes and frequency bands using SFFS for each subject for CRF and LDCRF. The input feature vector is formed by concatenation of the power of the signals in each of the selected frequency bands for each electrode in Table II. Experimental results on test data are shown in Table III. The proposed CRF and LDCRF-based methods are compared to the top result in the BCI competition, to the HHMM and HHCRF-based methods presented in [24], to a method proposed by Lin et al. that employs neural networks based on particle swarm optimization [11], and to the recently proposed S-dFasArt method of Cano et al. [27]. Results evidence the superiority of the proposed methods. LDCRF performs better than CRF does, which can be explained by the use of hidden variables that allow modeling, besides extrinsic dynamics, the intrinsic

TABLE II  
FREQUENCY BANDS FOR EACH ELECTRODE SELECTED BY SFFS FOR THE LDCRF AND THE CRF BASED METHODS.

Subj	Chn	LDCRF			CRF		
		Frequency Band			Frequency Band		
		Alpha	Sigma	Beta	Alpha	Sigma	Beta
B01	C3	✓	-	✓	-	✓	✓
	CP1	✓	-	-	✓	✓	-
	P3	-	✓	-	-	-	-
B02	C3	✓	-	-	✓	-	-
	Cz	-	-	-	-	-	✓
	C4	✓	✓	✓	✓	-	✓
	CP1	-	✓	-	-	-	-
	P4	-	-	✓	-	-	✓
B03	C3	-	✓	-	✓	✓	-
	CP2	-	-	✓	-	-	-
	Pz	-	-	-	-	-	✓

TABLE III  
COMPARISON BETWEEN DIFFERENT METHODS.

Subject	B01	B02	B03	Average
Galan [10]	79.60	70.31	56.02	68.64
HHMM [24]	79.05	61.58	34.40	58.34
HHCRF [24]	94.58	70.17	32.11	65.62
IPSONN [11]	78.31	70.27	56.46	68.35
S-dFasArt [27]	87.21	82.26	58.72	76.07
<b>CRF</b>	92.95	89.63	61.81	<b>81.46</b>
<b>LDCRF</b>	95.63	89.75	72.36	<b>85.91</b>

dynamics of the signal during the execution of a particular task.

In the evaluation above, we compared our approach against the winner of the BCI competition, and hence have demonstrated our approach offers better classification performance than all methods considered by the competition organizers. There were a number of other methods submitted to the BCI competition and not considered by the competition organizers as they did not follow the requirements for evaluation. These excluded methods, however, may provide interesting information. In particular from the excluded method with the highest performance, proposed by John Q. Gan et al., includes post processing stages following a linear classifier. The post processing stage smoothes the output of the classifier, that is, previous values of the output were used to define the current output under the assumption that rapid changes are not observed during the execution of the mental tasks. This method obtains an average accuracy of 80.97%. The proposed CRF and LDCRF methods yield better performance in terms of the accuracy. Furthermore, they do not need any post-processing of the output (See Figure 4). The proposed models are able to learn from training data that fast changes in the executed task are unlikely. However if such transitions do appear in the training data, they will be automatically taken into consideration in the learning phase. We believe this is a principled approach to learning and exploiting the dynamics of transitions among tasks in an asynchronous BCI system.

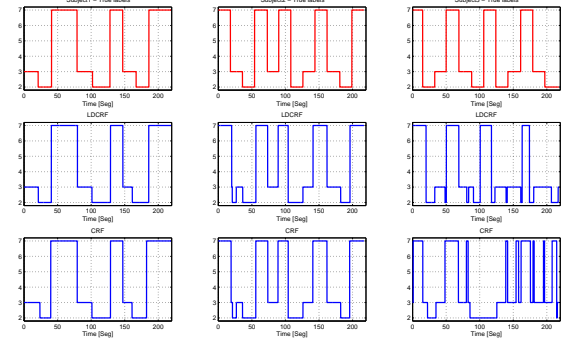


Fig. 4. Classification output for the proposed methods, CRF and LDCRF on the test data. Labels 2,3 and 7 correspond to right hand imaginary, left hand imaginary and word association respectively.

## VI. CONCLUSION

In this work two statistical methods are proposed for use in modeling the dynamics of the EEG signal during the execution of mental tasks in an asynchronous BCI scenario. The preprocessing of the signals involve the use of global Laplacian filters and estimation of the spectral density of the segmented EEG signals using the last second of data. SFFS was used for selection of relevant features. A CRF-based model and a LDCRF-based model were employed. The former method is able to model extrinsic dynamics of the EEG features. Those dynamics are related to the transitions from one mental task to the other in an asynchronous BCI system. LDCRF surpasses that approach and models, in addition to the extrinsic dynamics, the internal structure of the signals. We assert that this structure is related to different mental states during the execution of a specific mental task (ERD / ERS for imaginary motor tasks). The superiority of the presented CRF-based and LDCRF-based methods is evidenced in the results presented using a publicly available dataset, and by comparison with recent work. Furthermore it is worth of noting that the proposed methods do not need to use post-processing methods as they automatically learn the dynamics of data. Another advantage of the proposed methods is that there is no need for windowing the EEG features, thanks to the fact that the proposed methods inherently model the temporal structure of the signals and carry temporal information through the state variables. Future work will involve the analysis of the hidden state sequences in different brain regions in order to track the activation of those regions during the execution of mental tasks.

## ACKNOWLEDGMENT

This work was partially supported by the Scientific and Technological Research Council of Turkey under Grant 111E056, Sabanci University Internal Grant IACF-11-00889, and by a Turkish Academy of Sciences Distinguished Young Scientist Award.

## REFERENCES

- [1] N. Birbaumer and L. G. Cohen, "Brain computer interfaces: communication and restoration of movement in paralysis," *The Journal of Physiology*, vol. 579, no. 3, pp. 621–636, 2007.
- [2] A. Nijholt, B. Reuderink, and D. Oude Bos, "Turning shortcomings into challenges: Brain - computer interfaces for games," in *Intelligent Technologies for Interactive Entertainment*, O. Akan, P. Bellavista, and Cao, Eds. Springer Berlin Heidelberg, 2009, vol. 9, pp. 153–168.
- [3] A. Kübler and K. R. Müller, *Toward Brain-Computer Interfacing*. Cambridge, Massachusetts: MIT Press, 2007, ch. 1.
- [4] J. d. R. Millán, P. W. Ferrez, and A. Butfield, "The idiap brain-computer interface: An asynchronous multi-class approach," in *Towards Brain-Computer Interfacing*, G. Dornhege, J. d. R. Millán, T. Hinterberger, D. McFarland, and K. R. Müller, Eds. The MIT Press, 0 2007.
- [5] R. Leeb, D. Friedman, G. R. Müller-Putz, R. Scherer, M. Slater, and G. Pfurtscheller, "Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic," *Intell. Neuroscience*, vol. 2007, pp. 7:1–7:12, April 2007. [Online]. Available: <http://dx.doi.org/10.1155/2007/79642>
- [6] J. Millan and J. Mourino, "Asynchronous BCI and local neural classifiers: an overview of the adaptive brain interface project," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 159–161, June 2003.
- [7] E. Sadeghian and M. Moradi, "Continuous detection of motor imagery in a four-class asynchronous bci," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, Aug. 2007, pp. 3241–3244.
- [8] C. Tsui and J. Gan, "Asynchronous BCI Control of a Robot Simulator with Supervised Online Training," in *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, ser. Lecture Notes in Computer Science, H. Yin, P. Tino, E. Corchado, W. Byrne, and X. Yao, Eds. Springer Berlin / Heidelberg, vol. 4881, pp. 125–134.
- [9] F. Velasco-Álvarez and R. Ron-Angevin, "Asynchronous brain-computer interface to navigate in virtual environments using one motor imagery," in *Proceedings of the 10th International Work-Conference on Artificial Neural Networks: Part I: Bio-Inspired Systems: Computational and Ambient Intelligence*, ser. IWANN '09, 2009, pp. 698–705.
- [10] F. Galán, F. Oliva, and J. Guàrdia, "Using mental tasks transitions detection to improve spontaneous mental activity classification," *Medical and Biological Engineering and Computing*, vol. 45, pp. 603–609, 2007.
- [11] Cheng-Jian Lin and Ming-Hua Hsieh, "Classification of mental task from EEG data using neural networks based on particle swarm optimization," *Neurocomputing*, vol. 72, no. 46, pp. 1121–1130, 2009.
- [12] R. Aler, I. M. Galván, and J. M. Valls, "Transition detection for brain computer interface classification," in *Biomedical Engineering Systems and Technologies*, ser. Communications in Computer and Information Science, A. Fred, J. Filipe, and H. Gamboa, Eds. Springer Berlin Heidelberg, vol. 52, pp. 200–210.
- [13] B. Obermaier, C. Guger, C. Neuper, and G. Pfurtscheller, "Hidden Markov Models for online classification of single trial EEG data," *Pattern Recognition Letters*, vol. 22, no. 12, pp. 1299–1309, Oct. 2001.
- [14] H.-I. Suk and S.-W. Lee, "Two - Layer Hidden Markov Models for Multi-class Motor Imagery Classification," *First Workshop on Brain Decoding: Pattern Recognition Challenges in Neuroimaging*, pp. 5–8, Aug. 2010.
- [15] A. O. Argunsah and M. Çetin, "AR-PCA-HMM Approach for Sensorimotor Task Classification in EEG-based Brain-Computer Interfaces," *20th International Conference on Pattern Recognition*, pp. 113–116, Aug. 2010.
- [16] J. F. Delgado Saa and M. Çetin, "Modeling differences in the time - frequency presentation of EEG signals through HMMs for classification of imaginary motor tasks," *Sabancı University*, 2011.
- [17] J. F. Delgado Saa and M. Çetin, "Hidden conditional random fields for classification of imaginary motor tasks from eeg data," in *Proceedings of 19th European Signal Processing Conference*, ser. EUSIPCO, 2011.
- [18] J. F. Delgado Saa and M. Çetin, "A latent discriminative model-based approach for classification of imaginary motor tasks from eeg data," *Journal of Neural Engineering*, vol. 9, no. 2, p. 026020.
- [19] D. Kelly, J. McDonald, and C. Markham, "Evaluation of threshold model hms and conditional random fields for recognition of spatiotemporal gestures in sign language," in *Computer Vision Workshops (ICCV Workshops)*, 2009 IEEE 12th International Conference on, 27 2009–oct. 4 2009, pp. 490–497.
- [20] S. B. Wang, A. Quattoni, L.-P. Morency, D. Demirdjian, and T. Darrell, "Hidden conditional random fields for gesture recognition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006, pp. 1521–1527.
- [21] H.-S. Yoon, J. Soh, Y. J. Bae, and H. S. Yang, "Hand gesture recognition using combined features of location, angle and velocity," *Pattern Recognition*, vol. 34, no. 7, pp. 1491–1501, 2001. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320300000960>
- [22] A. Gunawardana, M. Mahajan, A. Acero, and J. C. Platt, "Hidden conditional random fields for phone classification," in *Interspeech*, 2005, pp. 1117–1120.
- [23] G. Pfurtscheller and F. Lopes da Silva, "Event - related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999.
- [24] T. Sugiura, N. Goto, and A. Hayashi, "A discriminative model corresponding to hierarchical hms," in *Proceedings of the 8th international conference on Intelligent data engineering and automated learning*, ser. IDEAL'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 375–384. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1777942.1777982>
- [25] K. P. Murphy and M. A. Paskin, "Linear time inference in hierarchical hms," in *IN PROCEEDINGS OF NEURAL INFORMATION PROCESSING SYSTEMS*, 2001.
- [26] S. Fine, Y. Singer, and N. Tishby, "The hierarchical hidden markov model: Analysis and applications," *Machine Learning*, vol. 32, pp. 41–62, 1998, 10.1023/A:1007469218079. [Online]. Available: <http://dx.doi.org/10.1023/A:1007469218079>
- [27] J.-M. Cano-Izquierdo, J. Ibarrola, and M. Almonacid, "Improving Motor Imagery Classification With a New BCI Design Using Neuro-Fuzzy S-dFasArt," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 20, no. 1, pp. 2–7, Jan. 2012.
- [28] L.-P. Morency, A. Quattoni, and T. Darrell, "Latent-Dynamic Discriminative Models for Continuous Gesture Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition. CVPR '07.*, June 2007, pp. 1–8.
- [29] J. D. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of the Eighteenth International Conference on Machine Learning*, ser. ICML '01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 282–289.
- [30] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recogn. Lett.*, vol. 15, pp. 1119–1125, November 1994. [Online]. Available: [http://dx.doi.org/10.1016/0167-8655\(94\)90127-9](http://dx.doi.org/10.1016/0167-8655(94)90127-9)
- [31] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern Recognition*, vol. 33, no. 1, pp. 25–41, 2000. [Online]. Available: <http://www.sciencedirect.com/science/article/B6V14-40CK1NW-3/2/6c05ecc3096175ece4bd4c8cf3c1eed1>
- [32] J. Schenk, M. Kaiser, and G. Rigoll, "Selecting Features in On-Line Handwritten Whiteboard Note Recognition: SFS or SFFS?" in *Document Analysis and Recognition, 2009. ICDAR '09. 10th International Conference on*, July 2009, pp. 1251–1254.
- [33] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988.



**Jaime F. Delgado Saa** (S09) received a B.S. degree in electronics engineering and the M.S. degree in electronics engineering from Universidad del Norte, Barranquilla, Colombia, in 2003 and 2009, respectively. From 2004 to 2009 he worked for the Universidad del Norte and the Colombian Navy in teaching and research. He is currently working toward the Ph.D. degree from Sabanci University, Istanbul. His research interests include signal processing, pattern recognition and graphical models.



**Müjdat Çetin** (S98M02) received the Ph.D. degree in electrical engineering from Boston University, Boston, MA, in 2001. From 2001 to 2005, he was with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge. Since September 2005, he has been a faculty member with Sabanci University, Istanbul, Turkey. His research interests include statistical signal and image processing, inverse problems, radar imaging, braincomputer interfaces, machine learning, computer vision, data fusion, wireless sensor

networks, biomedical information processing, and sensor array signal processing. Dr. Çetin has served as the Technical Program Cochair for the 2010 International Conference on Pattern Recognition and for the 2006 IEEE Turkish Conference on Signal Processing, Communications, and their Applications. He is currently an Area Editor for the Journal of Advances in Information Fusion, a Guest Editor for Pattern Recognition Letters, and a EURASIP Liaison Officer for Turkey. He was the recipient of several awards including the 2010 IEEE Signal Processing Society Best Paper Award, the 2010 METU Parlar Foundation Research Incentive Award, the 2008 Turkish Academy of Sciences Distinguished Young Scientist Award, the 2007 Elsevier Signal Processing Journal Best Paper Award, and the 2006 TBİTAK Career Award.