

# Modeling Differences in the Time-Frequency Representation of EEG Signals Through HMM's for Classification of Imaginary Motor Tasks.

Jaime F Delgado Saa<sup>\*†</sup>, Mujdat Cetin<sup>\*</sup>

<sup>\*</sup>Signal Processing and Information Systems Laboratory, Sabanci University, Istanbul-Turkey.

<sup>†</sup>Department of Electrical and Electronics Engineering, Universidad del Norte, Colombia.

**Abstract**—Brain Computer interfaces are systems that allow the control of external devices using the information extracted from the brain signals. Such systems find applications in rehabilitation, as an alternative communication channel and in multimedia applications for entertainment and gaming. In this work, a new approach based on the Time-Frequency (TF) distribution of the signal power, obtained by autoregressive methods and the use Hidden Markov models (HMM) is developed. This approach take into account the changes of power on different frequency bands with time. For that purpose HMM's are used to modeling the changes in the power during the execution of two different motor tasks. The use of TF methods involves a problem related to the selection of the frequency bands that can lead to over fitting (due to the course of dimensionality) as well as problems related to the selection of the model parameters. These problems are solved in this work by combining two methods for feature selection: Fisher Score and Sequential Floating Forward Selection. The results are compared to the three top results of the BCI competition IV. It is shown here that the proposed method over perform those other methods in four subjects and the average over all the subjects equals the one obtained by the winner algorithm of the competition.

**Keywords:** Brain Computer Interfaces, HMM, AR-Models, Feature Selection.

## I. INTRODUCTION

A Brain Computer Interface (BCI) is a system that provides an alternative communication way for people who suffered a disease or an accident that compromises the ability to perform motor tasks. Also, applications for healthy subjects in areas of multimedia and gaming started to incorporate this technologies in the last years[1]. BCI's make use of the brain signals to control external devices that help the subject to communicate and interact with the environment. Most of the current approaches to BCI are based on the comparison of the values of power of the EEG during the execution of imaginary motor tasks. However, the well known phenomena of Event Related Synchronization and Event Related Desynchronization [2],[3] provides more information that can be employed to improve the performance of the BCI's. This information is related not only to the difference of power of the signals but the time change of them on different frequency bands. For this, algorithms that take into account the change of the signal on time as Hidden Markov Models (HMM) [4],[5]

This work was supported by the Scientific and Technological Research Council of Turkey under Grant 107E135, by a Turkish Academy of Sciences (TUBA) Distinguished Young Scientist Award

have been used in combination with features that describe the temporal behavior of the EEG signals [6],[7]. Although the Time-Frequency analysis of the EEG signals have shown good results in previous works [8],[9],[10],[11],[12] in BCI applications, a combination of the time-frequency power distribution of the signals and algorithms that take into account the changes in the distribution have not been reported. One possible reason for this is that the selection of the parameters of the models (states, Gaussian mixtures, etc. in HMM) along with the selection of the frequency bands becomes problematic. In this work we make use of the Time-Frequency distribution of the power of the signals, using Autoregressive Models for calculation of the Power Spectral Density (PSD) and HMM for classification of two different motor tasks. The problem of selection of parameters is solved by combining two methods for selection of features. The first method is used to select the most representative features and fix the Model Order Parameters (states, Gaussian mixtures, type of transition matrix). The second method is used to incorporate new features (power in selected frequency bands) that increase the performance by using cross validation on the training data. The final results presented, are based on the system's performance using two unseen test sessions.

In this document, the work above presented is described. First a short description of the methods employed is shown. Next, an explanation of how those methods were used, the results obtained and the comparisons with others methods is given. Finally, the conclusion and future work are presented.

## II. METHODS

### A. Dataset description

In this work the Dataset Iib of the BCI competition IV [13], which consist in bipolar EEG recordings over C3, Cz and C4 in 9 subjects, was used. The cue based BCI paradigm consisted on two classes, represented by the imagination of the movement of left hand and right hand. The scheme of the sessions is depicted in Figure 1. At the beginning of each trial a fixation cross appear and remain in the screen until 3th second, then the cue is presented and the subject is requested to perform the imaginary movement of the corresponding hand. The dataset contains five sessions, two of them without feedback. In this work only the session with feedback was employed for training because the methods presented here

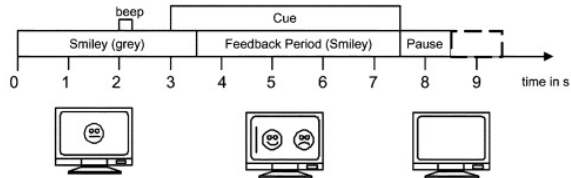


Fig. 1. Time scheme for the experimental setup

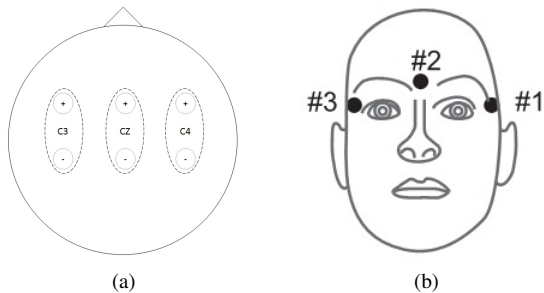


Fig. 2. a.) Montage used to extract the signal on C3, C4 and Cz, b.) EOG Channels

make use of the temporal behavior of the EEG signals and this behavior is modified due to the feedback influence [14]. For testing, the provided data involve only sessions with feedback. The available EEG signals (Figure 2a) were filtered between 0.5 and 100Hz and a notch filter at 50Hz was used. EOG recordings are also available and the electrode montage is shown in Figure 2b. The EOG recordings can be used to reduce the interference caused by the ocular movements during the time period that the session last. For more information about the dataset specifications see [15].

### B. Preprocessing:

1) *EOG artifact reduction:* In order to reduce the EOG interference in the EEG signal, linear regression was employed. In this approach for EOG artifact reduction, the recording signal is modeled as the summation of the actual EEG signal and the noise, represented by a linear combination of the EOG signals [16], as shown in equation (1).

$$Y_{TXM} = S_{TXM} + U_{TXN}b_{NXM} \quad (1)$$

where  $T$  represent the time points,  $M$  the number of EEG channels and  $N$  the number of EOG ( $U$ ) channels. Then the original EEG signal can be found to be  $S = Y - U.b$ . This problem is solved by finding the coefficients that explain how the EOG signal is propagated (volume conduction) to each of the points in the scalp where the measurement is done. Assuming that the EOG and the actual EEG signals are independent, the coefficients can be found by equation (2):

$$b = Cov(U^T U)^{-1} Cov(U^T Y) \quad (2)$$

2) *Spectral Filtering:* The data was sampled a 250Hz and originally filtered between 0.5Hz and 100Hz. A notch filter was also employed in order to eliminate the 50Hz interference.

In addition to this, the bandwidth of the signals was reduced to the interest components (6Hz to 35Hz) using a 6th order Chebyshev Type II filter. The lower frequency was selected in order to help to minimize the effect of the EOG activity which contains components at low frequencies.

### C. Feature Extraction

1) *Power Spectral Density Estimation using AR parameters:* The well known phenomena of Event Related Synchronization/Desynchronization provide useful information for identification of motor tasks. The PSD of the signal is calculated by parametric methods which involve the calculation of autoregressive models of the signal. In this work the Burg method was used because in contrast to the Yuller - Walker method, the former provides always a stable model, minimizing the error on backward and forward direction [17]. The PSD is estimated as the frequency response of the AR model which is estimated according to (5):

$$x(n) = \sum_{k=1}^p a_k x(n-k) + u(n). \quad (3)$$

Where,  $n$  represents the discrete time index,  $p$  is the model order,  $a_k$  is the  $k$ th coefficient of the model and  $u(n)$  is the system input or noise function. Then calculating Z-Transform we obtain:

$$H(z) = \frac{X(z)}{U(z)} = \left( 1 - \sum_{k=1}^p a_k z^{-k} \right)^{-1} \quad (4)$$

Given this, the power spectrum can be obtained by evaluating  $H(z)$  on the unit circle where  $z = \exp(jw)$ , [18].

### D. Feature Selection

To select the band of frequencies which provide more information about the oscillatory brain activity related to the two mental tasks realized by the Subjects, the training dataset was analyzed calculating the degree of separability of the frequency components during the realization of imaginary motor activity. For this purpose, two methods, Fisher Score and Sequential Floating Forward Selection, were employed.

1) *Fisher Ratio:* The Fisher criterion, which is a measure of the class separability for two classes [19] is introduced. The separability between class  $i$  and class  $j$  is determined by:

$$Fisher_{(f,t,i,j)} = \frac{(m_{(i,f,t)} - m_{(j,f,t)})^2}{\sigma_{(i,f,t)}^2 + \sigma_{(j,f,t)}^2}. \quad (5)$$

Where  $m_{(i,f,t)}$  and  $m_{(j,f,t)}$  correspond to the inter trial average of the PSD at time  $t$  and frequency  $f$  for each class where  $i \neq j$ ,  $\sigma_{(i,f,t)}^2$  and  $\sigma_{(j,f,t)}^2$  are, as well, the inter trial variances of the PSD for each class. For a two-class problem, this produces two matrices (one for each class) from where the Fisher ratio (as indicator of the separability between classes) can be selected for a specific frequency component as well as for a specific time instant.

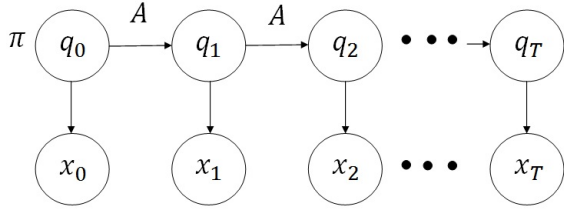


Fig. 3. Hidden Markov Model

2) *Sequential Floating Forward Selection*: SFFS (see [20]) is a method for feature selection in which new features are added (to a selected set of features) according to a cost function. From the unselected features, the one which maximizes the value of the cost function is selected and included in the set. This method is differentiated from Sequential Forward Selection (SFS) in the sense that the meaningless feature can be removed in each iteration. The algorithm will add a new feature until the number of features currently selected equals the number of features requested by the user.

### E. Classification: Hidden Markov Models

An example of a Hidden Markov Model (HMM) is depicted in Figure 3. A HMM is a finite automaton which contain a discrete number of states  $Q$  emitting a feature vector  $X$  at each time point, the distribution of the output at each time point depends on the current state. Given that this kind of models are generative, it is necessary to determine the joint probability over observation and labels, which requires all possible observation sequences to be enumerated. In order to make the inference problem tractable, conditional independence is assumed, meaning that the future states are independent from the past states given that the current state is known.

The problem is then to find the parameters of the model that maximize the log-likelihood of the observation given those parameters. The parameters are the transition matrix  $A$ , for which each entry  $a_{i,j}$  represent the probability to pass from state  $i$  to state  $j$ , the vector of initial probabilities  $\Pi$  which represent the probability of  $q_i$  to be the initial state  $q_0$ . Finally, distribution of the data on each stated is modeled using Gaussian mixtures in this work.

For classification, a HMM is trained for each of the possibles classes, where each new sequence of data (EEG features) is evaluated and the output class is defined by the model with higher probability to generate the observed sequence of data. For more information on HMM's the reader is referred to [21]

## III. RESULTS

The EOG interference was minimized by using Linear Regression. The coefficients were calculated using the signals at the beginning of each session which involves the execution of eye movement and intervals with close eyes and open eyes, as described in [16]. This provides information that permit to establish the propagation of the EEG signals on the scalp. Also, given that the EOG signals are large in magnitude, the

interference of EEG in the EOG recordings can be neglected [16]. Next, each of the signals obtained from the electrode montage in Figure 2a, was filtered between 6 and 35Hz and the PSD was calculated using a sliding window of 1 second length with overlapping of 80%. The frequency resolution was set to 0.25Hz by selecting a FFT of 1024 points in the calculation of the frequency response of the AR-Model and the order of the models was set to 10 referring to previous works [22],[18]. By doing this, a representation on time and frequency is obtained making possible to observe in which frequencies the power of the signal is concentrated for a specific time interval.

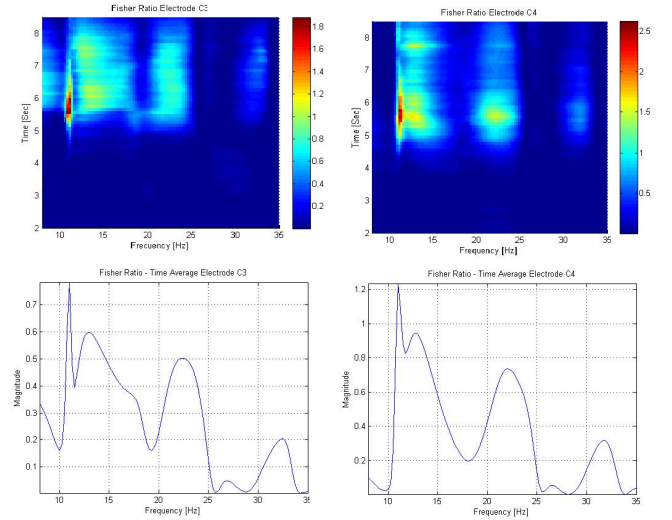


Fig. 4. Fisher Ratio (Time vs Frequency) and Average across time for each frequency component

Once the PSD of the signal is obtained it is necessary select the features which provide more information for discrimination between the four classes. Initially the Fisher Ratio (FR) is calculated according to the Equation (8). The resultant matrices for electrodes C3 and C4 are shown in Figure 4 for subject 8. The time average of the FR is used to select the frequencies for which the separability of the signal is the highest (See Figure 4). The spectrum of the signal was divided in non-overlapping frequency bands form 9 - 35Hz and then those frequency bands coinciding with the frequency values with higher separability (according to the FR), are preselected. This initial set of features is used to select the Model Order parameters for the HMM. The Model order parameters (MOP) make reference to the determination of the number of mixtures used to describe the distribution of the data on each state, the number of states and the type of model (ergodic or right). The last option establish restrictions over the transition matrix. In the case of the right model, transitions to previous estates are forbidden while in a ergodic model there is no such restrictions. The MOP were selected using 3X3 Folds Cross validation on the preselected features, then 2/3 of the data was used each time as training data and 1/3 as test data. This is repeated 3 times changing the number of states, the number of Gaussian Mixtures and the type of model. The Model with

Subject	Zheng Yang Chin et al.	Huang Gan et al.	Damien Coyle et al.	Proposed Method.
1	0.40	0.42	0.19	<b>0.46</b>
2	0.21	0.21	0.12	<b>0.27</b>
3	<b>0.22</b>	0.14	0.12	0.19
4	<b>0.95</b>	0.94	0.77	0.93
5	0.86	0.71	0.57	<b>0.88</b>
6	0.61	0.62	0.49	<b>0.64</b>
7	0.56	<b>0.61</b>	0.38	0.54
8	<b>0.85</b>	0.84	<b>0.85</b>	0.71
9	0.74	<b>0.78</b>	0.61	0.75
Average	<b>0.60</b>	0.58	0.46	<b>0.60</b>

TABLE I  
COMPARISON WITH THE RESULTS OF THE BCI COMPETITION IV (KAPPA VALUES)

the set of MOP that provides the higher average accuracy was selected. Figure 5 shows the final frequency bands selected for each subject for the signals measured over C3 and C4.

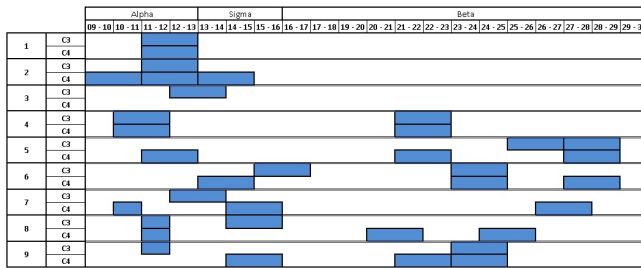


Fig. 5. Selected frequency bands for each subject

After this process, the MOP are fixed and the remaining features obtained by taking the average power in the frequency band previously defined are processed using SFFS. The size of the final set of features was limited to six, to avoid over fitting due to the course of dimensionality. This process is done for two HMM one corresponding to the modeling of signals for each motor task.

The results obtained are compared to the three top results of the BCI Competition IV dataset 2b. For this, according to the methodology employed in the competition the Kappa values [23] were calculated. The results are shown in Table I, where the best results appear in bold. It is possible to see that our proposed method over perform all the other works in 4 subjects and the average performance equals the highest obtained in the competition

Also the accuracy across time was calculated for each subject in each of the two testing sessions. Figure 6 shows that the higher accuracy for the majority of the subjects is obtained at the end of the trial. This was expected, because for training the models the data from second 4 up to the end of the trial were used. Those results are summarized in table II, where the highest accuracy across time and the time at which is achieved are shown.

#### IV. CONCLUSION

In this work the changes in the power of the EEG signals were employed for classification of two motor tasks. The main idea is that the well known phenomena of ERD/ERS can be

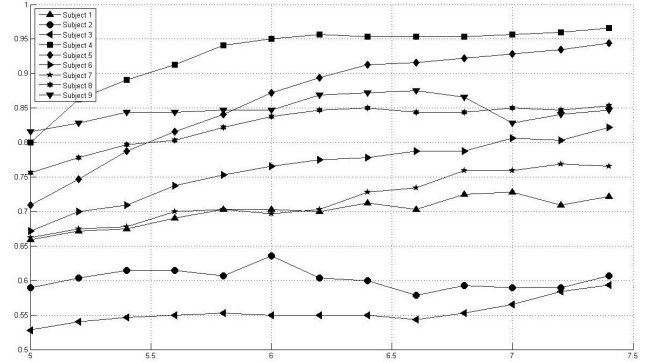


Fig. 6. Accuracy Vs Time for each subject

Subject	Max Acc	Time [sec]
1	73%	7.0
2	64%	6.0
3	59%	7.4
4	97%	7.4
5	94%	7.4
6	82%	7.4
7	77%	7.2
8	85%	7.4
9	88%	6.6

TABLE II  
TIME POINTS OF MAXIMUM ACCURACY FOR EACH SUBJECT

modeled by algorithms that take into account the change of the input signal. For that a HMM's were employed looking for modeling those changes. The results showed that this method is suitable for this task and is also comparable with the methods currently used. Although HMM has been used before, in this work we show that the use of the TF AR-Power estimation and the proper selection of the changes of these estimations on specific frequency bands (by combining two feature selections methods) can provide good results. One important point is that time points where (in the trial time line) the maximum accuracy is obtained. As shown in Table II in 7 out of the 9 subjects the best accuracy is obtained at the end of the trial, which in cases where not continuous output is required represent an advantage over others methods based on static classifiers because the time for best performance is "a priori" known. This work leaves the door open to a discussion

related to the selection of the MOP. Future work will be concentrated to determinate the number of states according to Neurophysiologic information which should lead to a better understanding of the definition of state in terms of the behavior of the EEG signals and in agreement with the medical theory. In respect to the selection of the general model, (a HMM for modeling the EEG signals for each of the two tasks) in this work the brain is modeled as a unified system in the way that the MOP for each of the two models (left and right) include signals from different regions. Our future work will also include revision of more specific models where each region can be modeled independently and relationships found in those specific models could also be exploited with more general type of graphical models.

## REFERENCES

- [1] A. Nijholt, B. Reuderink, and D. Oude Bos, "Turning shortcomings into challenges: Brain-computer interfaces for games," in *Intelligent Technologies for Interactive Entertainment*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, O. Akan, P. Bellavista, J. Cao, F. Dressler, D. Ferrari, M. Gerla, H. Kobayashi, S. Palazzo, S. Sahni, X. S. Shen, M. Stan, J. Xiaohua, A. Zomaya, G. Coulson, A. Nijholt, D. Reidsma, and H. Hondorp, Eds. Springer Berlin Heidelberg, 2009, vol. 9, pp. 153–168.
- [2] *Basic concepts on EEG synchronization and desynchronization. Event-related desynchronization and related oscillatory phenomena of the brain.* Elsevier, 1999, vol. 6, pp. 3–11.
- [3] *Basic concepts on EEG synchronization and desynchronization.* Elsevier, 1999, vol. 6, pp. 1–14.
- [4] 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.
- [5] H.-I. Suk and S.-W. Lee, *2010 First Workshop on Brain Decoding: Pattern Recognition Challenges in Neuroimaging.*
- [6] B. Hjorth, "Eeg analysis based on time domain properties," *Electroencephalography and Clinical Neurophysiology*, vol. 29, no. 3, pp. 306 – 310, 1970.
- [7] C. Vidaurre, N. Krmer, B. Blankertz, and A. Schlögl, "Time domain parameters as a feature for eeg-based brain-computer interfaces," *Neural Networks*, vol. 22, no. 9, pp. 1313 – 1319, 2009, brain-Machine Interface. [Online]. Available: <http://www.sciencedirect.com/science/article/B6T08-4WTRS8W-2/2/9185d6ddbcb1005544d62c8f1087ce4f>
- [8] R. Magjarevic, A. Yonas, A. S. Prihatmanto, and T. L. Mengko, "Time-frequency features combination to improve single-trial eeg classification," in *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany*, ser. IFMBE Proceedings, O. Dssel and W. C. Schlegel, Eds. Springer Berlin Heidelberg, 2010, vol. 25/4, pp. 805–808.
- [9] R. T. Mina, A. Atiya, M. I. Owis, and Y. M. Kadah, "Brain-Computer Interface Based on Classification of Statistical and Power Spectral Density Features," *Biomedical Engineering*, pp. 2–5, 2006.
- [10] R. Palaniappan, "Brain Computer Interface Design Using Band Powers Extracted During Mental Tasks," *Conference Proceedings. 2nd International IEEE EMBS Conference on Neural Engineering, 2005.*, pp. 321–324, 2005.
- [11] Z. Mu, D. Xiao, and J. Hu, "Classification of Motor Imagery EEG Signals Based on TimeFrequency Analysis," *International Journal of Digital Content Technology and its Applications*, pp. 116–119, 2009. [Online]. Available: <http://www.aicit.org/jdcta/page11.html>
- [12] J. Delgado Saa and M. Sotaquir, "Eeg signal classification using ar-power spectral features and linear discriminant analysis," *Proceedings of Latin American and Caribbean Consortium of Engineering Institutions, Arequipa - Per*, 2010.
- [13] G. M. u.-P. A. S. o. C. Brunner, R. Leeb and G. Pfurtscheller, "Bci competition 2008 - graz data set b." 2008.
- [14] A. Neuper, C. Schlögl and G. Pfurtscheller, "Enhancement of left-right sensorimotor eeg differences during feedback-regulated motor imagery," *Clin. Neurophysiol*, vol. 16, no. 4, pp. 373–382, 1999.
- [15] G. Pfurtscheller *et al.*, "Brain-computer communication: motivation, aim, and impact of exploring a virtual apartment." *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 15, no. 4, pp. 473–82, Dec. 2007.
- [16] a. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings." *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 118, no. 1, pp. 98–104, Jan. 2007.
- [17] P. Stoica and M. R. *Introduction to Spectral Analysis.* Prentice Hall, '1997.
- [18] B. Jansen, J. Bourn, and J. Ward, "Autoregressive stimation of short segament spectra for computerized eeg analysis," *IEE Tmraction on Biomedical Engineering*, vol. BME-28, no. 8, pp. 630–637, 1981.
- [19] X. Pei and C. Zheng, "Classification of left and right hand motor imagery tasks based on eeg frequency component selection," may. 2008, pp. 1888 –1891.
- [20] J. N. c. P. Pudil and J. Kittler, "Floating Search Methods in Feature Selection," *Pattern Recognition Letters*, vol. 15, no. 11, pp. 1119–1125, 1994.
- [21] L. R. Rabiner, "Readings in speech recognition," A. Waibel and K.-F. Lee, Eds. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990, ch. A tutorial on hidden Markov models and selected applications in speech recognition, pp. 267–296.
- [22] A. Schlögl, "The electroencephalogram and the adaptive autoregressive models: Theory and applications," Ph.D. dissertation, vorgelegt an der Technischen Universitt Graz, Graz, April 2000.
- [23] G. Dornhege *et al.*, *Toward Brain-Computer Interfacing.* Cambridge, Massachusetts: MIT Press, 2007, ch. 19.