Gaussian Mixture Model Based on Genetic Algorithm for Brain-Computer Interface

Boyu Wang, Chi Man Wong, Feng Wan, Peng Un Mak, Pui In Mak, and Mang I Vai

Department of Electrical and Electronics Engineering
University of Macau
Macau SAR, China

Abstract—**Gaussian mixture model (GMM) has been considered to model the EEG data for the classification task in brain-computer interface (BCI) system. In the practical BCI application, however, the performance of the classical GMM optimized by standard expectation-maximization (EM) algorithm may be degraded due to the noise and outliers, which often exist in realistic BCI systems. The motivation of this paper is to introduce the GMM based on the combination between the genetic algorithm (GA) and EM method to give a probabilistic output for further analysis and, more important, to achieve the reliable estimation by pruning the potential outliers and noisy samples in the EEG data, so the performance of BCI system can be improved. Experiments on two BCI datasets demonstrate the improvement in comparison with the classical mixture model.**

Keywords-electroencephalogram; genetic algorithm; Gaussian mixture model; brain-computer interface

I. INTRODUCTION

A brain-computer interface (BCI) is a system that forms a possible output channel for the individuals with severe motor disabilities to have effective control over external devices without using the traditional pathways as peripheral muscle or nerves [1–3]. The brain activities are often recorded noninvasively by electroencephalogram (EEG), which has excellent temporal resolution and usability, and the EEG signal is therefore a popular choice for BCI research.

In order to control an EEG-based BCI, the user must produce different brain activity patterns, which are recorded by electrodes on the scalp, and then features are extracted from the EEG signals and translated into the control commands. In most existing BCIs, this translation relies on a classification algorithm [4], [5]. For a more detailed review on the classification technologies used in BCIs see [3–8].

Gaussian mixture model (GMM) [17] has been applied to model the features extracted from the EEG signal in BCI. The GMM for unsupervised classification was also reported in [10]. In [21] and [22], the mixture of Gaussian was introduced as the online classifier and the parameters were updated in a simulated online scenario. In [9] a GMM-based classifier was used to separate the signal into different classes of mental task, where adaptation is concerned by using a supervised method. Each class was represented by a number of Gaussian prototypes, typically less than four. Similarly, [23] and [24]

also proposed an online GMM classifier via the decorrelated least mean square (DLMS) algorithm. In [11] and [12] the GMM based on sequential expectation-maximization (SEM) algorithm for the unsupervised adaptation was proposed. The failure detection index (FDI) was introduced to detect the failure of the adaptation, and decide when to stop the adaptation, or re-train the system. An improved GMM with parameter initialization was also proposed to enhance the adaptability of the classifier for the shifts between sessions in [13], where the initial parameters were treated as variational, and determined based on historical estimation. In [14] the GMM was used to model the resting brain activity so that changes in EEG signal can be detected rather than classified. Therefore, the proposed method did not require extensive initial signal identification procedures since it only required prior data samples from only one class (i.e., rest), and not form the multiple classes (e.g., signals associated with a variety of different actual or imagined movements). Furthermore, this approach did not require prior determination of the specific brain activity features that best suited for communication, which would benefit the clinical application of BCI technology. On the other hand, in [15] the mental state was modeled by Gaussian distribution, and then the prior distribution of rest state was inferred and subsequently adapted during the ongoing use. This distance-based approach made "anything other than active state" detection possible and allowed the subject to find the way most suitable for him/her. A comparative study of different methods of features and classifiers, including the GMM classifier, for designing of subject-specific and subject-independent BCIs was reported in [16].

BCI can benefit from the unsupervised learning methods, such as GMM, since in the practical application, the subject's intention is usually unknown to the system, and the GMM can be trained online without any feedback and cue as the label of the data. In general, the parameter set of a GMM is optimized by expectation-maximization (EM) algorithm in a way to maximize the likelihood of the data given the model. However, the performance of the classical model optimized by standard EM method will be degraded due to the noise and outliers, which often exist in a realistic BCI system due to the factors such as low signal-to-noise ratio (SNR), measurement inaccuracies, physiological variations in background EEG, muscle and eyes blink artifacts. Therefore, a robust method is needed to reject the noisy data so that the reliable estimation can be achieved.

This paper handles the noise and outliers detection task by introducing a robust method which combines the genetic algorithm (GA) with the EM to maximize trimmed likelihood function [18]. The performance of this method is tested on two datasets, which are dataset III and IV from BCI competition 2003.

II. METHODS

A. Standard EM algorithm for GMM

A mixture model with K Gaussian densities can be defined as

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k N(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{1}$$

which is called Gaussian mixture model (GMM). The parameters $\pi_1, ..., \pi_k$ are the mixture coefficients, and each Gaussian density $N(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ with mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{\Sigma}_k$. Being probabilities, the parameters π_k must satisfy

$$0 \leq \pi_k \leq 1 \text{ and } \sum_{k=1}^{K} \pi_k = 1 \tag{2}$$

Define $\Theta \equiv \{\pi_1, ..., \pi_k, \boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_k\}$ as the complete set of parameters needed to specify the GMM.

Suppose there is a data set of N observations $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$, and we model the data set with a GMM. Hence the corresponding log-likelihood function is

$$\log p(\mathbf{X} \mid \Theta) = \sum_{n=1}^{N} \log p(\mathbf{x}_n \mid \Theta) \tag{3}$$

One approach to estimate the parameter set is maximum likelihood (ML)

$$\hat{\Theta}_{ML} = \arg \max_{\Theta} \{\log p(\mathbf{X} \mid \Theta)\} \tag{4}$$

It is well known that the closed-form solution for (4) cannot be found analytically. One powerful method for finding the ML solution is the EM algorithm [19], which comprises two alternate steps: the E-step which computes the posterior probabilities by using current estimated parameters, and the M-step which updates the parameters using the current posterior probabilities. For more details on EM for GMM, as well as other mixture models, see [17], [19].

B. Maintaining the Integrity of the Specifications

In a practical BCI system, the noise and outliers due to the factors such as low signal-to-noise ratio (SNR), measurement inaccuracies, physiological variations in background EEG, muscle and eyes blink artifacts often exist, which will definitely degrade the performance of standard EM for GMM. In general, if outliers exist in the data, they may pull a cluster's mean estimated towards their location, while away from the cluster's true center. On the other hand, they can also enlarge the cluster's covariance in their direction. Therefore, the noise and the outliers can have two negative effects: the "masking" where the presence of an outlier masks the appearance of another one, and "swamping" which means certain non-outlying sample maybe wrongly observed as atypical [18].

To handle this problem, we introduce the genetic-based EM, which is proposed in [18] for GMM. In this method, the goal is to maximize the trimmed likelihood function, which is also used as the fitness function for GA [20]. The definition of the trimmed likelihood function is

$$\log p_{TLE}(\mathbf{X} \mid \Theta) = \sum_{n=1}^{N} \omega_n \log p(x \mid \Theta) \tag{5}$$

where $\omega_n \in \{0,1\}$, and $\sum_{n=1}^{N} \omega_n = M$. When $\omega_n = 0$, it indicates that $\mathbf{x}_n$ is trimmed off the data set, otherwise, $\omega_n = 1$. The chromosomes in the genetic population are binary vectors of length N, each bit of which indicated whether the sample $\mathbf{x}_n$ is trimmed or not. Only the samples survived (the corresponding bit is 1) can contribute to the Gaussian mixture. Thus, the output of this method is a subset of "normal data" of size M and the corresponding estimated parameter set for GMM.

Specifically, the genetic-based EM algorithm produces a sequence of estimates by applying the EM and GA alternately. In each EM step, the parameters of GMM are evaluated based on the selected "normal samples", and in the GA step, the chromosomes are updated according to the likelihood contribution for the current estimated mixture model. Each pair of the chromosome and the corresponding mixture model will be a possible solution. It should be emphasized that to avoid the random manner in mutation process, another operator called guided mutation is carried to trim off the samples with small likelihood. Since the trimmed likelihood function is proven that it preserves the monotonic property of standard EM, and the best individuals are always unaltered to the next generation due to the guided mutation, the convergence of this method can be guaranteed. Fig. 1 and Fig. 2 show the flowchart of the guided mutation and genetic based EM algorithm for GMM. For more detailed and comprehensive accounts on the genetic-based EM algorithm, see [18].
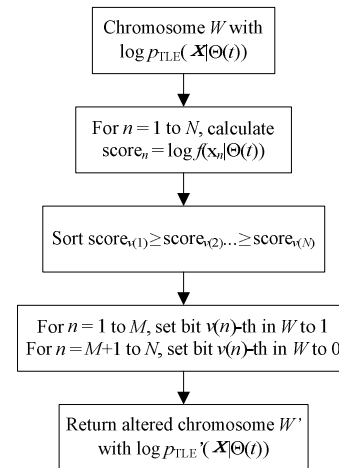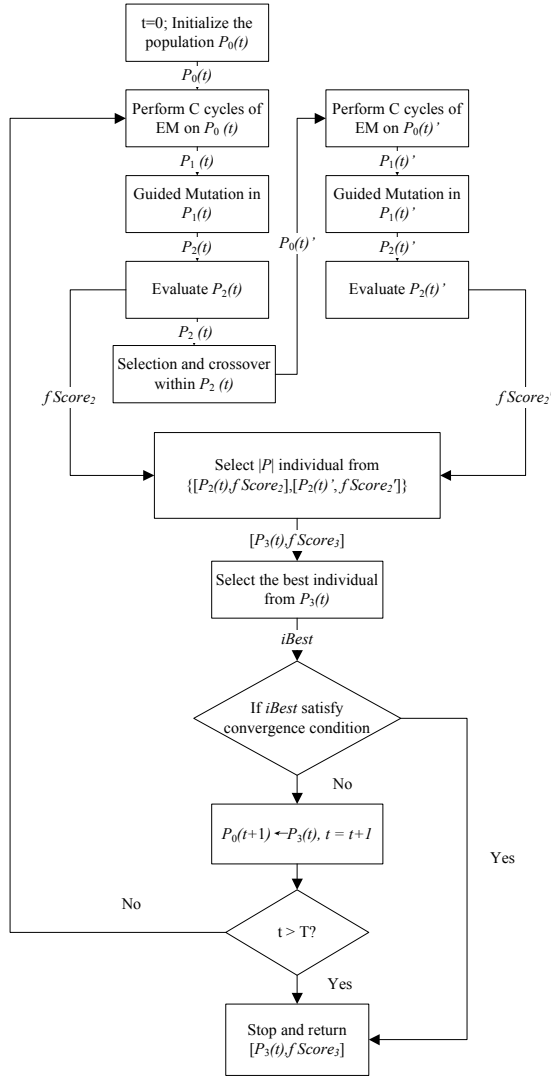

Figure. 1. Flowchart of guided mutation

Figure. 2. Flowchart of genetic-based EM for GMM

III. EXPERIMENTS

A. Data Set

The performance of GMM method based on genetic-based EM is evaluated on two datasets, which are the dataset III and dataset IV from BCI competition 2003 [32].

Dataset III consists of 420 motor imagery EEG samples (210 left trials and 210 right trials) of 9s length, which are recorded from a 25 years old female during a feedback session. Three pairs of bipolar EEG channels (anterior '+', posterior '-') were positioned over C3, Cz and C4. It was sampled at 128 Hz and bandpass filtered between 0.5 and 30Hz.

Dataset IV contains 416 finger movement trials of 500 ms length which are recorded from a normal subject. The task is to press with the index and little fingers the corresponding keys in a self-chosen order and timing 'self-paced key typing'. The EEG was collected by 28 electrodes at the positions of the international 10/20-system, and the sample rate was 100Hz.

B. Feature Extraction Methods

Various feature extraction methods have been proposed for EEG analysis [25]. For the first data set, since there are only three channels, the power spectral density (PSD) and band power (BP) are used. On the other hand, common spatial pattern (CSP), as a technique to analyze multichannel data, is adopted to extract the features on the second data set.

The PSD describe how the power (or variance) of a signal distributed with different frequencies, and BP is extracted in a given frequency band by filtering it, squaring it and average it. The CSP is designed based on a decomposition of the raw EEG signals into spatial patterns so that the variances of the resulting signals carry the most discriminative information [26], [27].

C. Parameter Setting

Some parameters for the feature extraction and classification algorithms are needed to be selected carefully. Specifically, for the first data set, the EEG signals are filtered by a bandpass filter (10-12 Hz), and then we calculate the PSD from C3 and C4 channels using periodogram within 10-12Hz. The length of time window for BP is 15 ($\approx$ 117.2 ms) and the corresponding starting point is t = 560 (= 4.375s). For the second data set, the EEG signals are filtered by a zero-phase filter (0-7Hz). Then we apply CSP on the segment with sample points of 43-47 (200ms-160ms before the keypress), and only the first pair of the spatial filter is used.

In this paper, since the GMM is trained in an unsupervised manner, the number of the Gaussian components is two, which equals to the number of the classes (left and right hand imagery/movement), otherwise the labels of the samples cannot be assigned and classified. For the genetic-based EM algorithm, there are also a few parameters need to be defined to specify the learning process. The maximum number of each EM cycles C is set to 10. The trimming level ϵ, which is defined as $M = (1-\epsilon) \times N$, is set to be 0.1, and the size of the population $|P|$ is 8.

IV. RESULTS

The performances of the GMM optimized by using genetic-based EM and standard EM algorithm, together with that of linear discriminant analysis (LDA), are demonstrated in Table I. From the table it can be found that LDA, as a supervised learning method, has higher accuracies than the other two classifiers, which is not surprising since it utilizes the information of the class labels.

On the other hand, the performance of GMM optimized by genetic-based EM is superior to the classical GMM. The main reason of it, as mentioned above, may be that the outliers and noisy samples are trimmed off the original data by the introduced method, which will improve the clustering quality.

TABLE I
CLASSIFICATION ACCURACY (%) OF EACH CLASSIFIER

Methods	First data set		Second data set
	PSD	BP	
LDA	82.3 ± 2.3	83.7 ± 3.3	83.1± 2.6
Classical GMM	79.3 ± 1.9	80.3 ± 4.5	78.2 ± 3.2
EM-GA based GMM	78.9 ± 2.9	81.4 ± 3.7	81.7 ± 2.8

This can be demonstrated in Fig. 3 and Fig. 4. In Fig. 3, the training samples of the second dataset (preprocessed by bandpass filter and CSP) are clustered by the classical GMM (dash lines) and the robust method (solid lines), of which the centers are marked with cross and circle respectively. The outliers are indicated by the green panes. It can be observed that the covariances of the both clusters are enlarged due to the noisy data and outliers. In addition, the directions of the dispersions and the center of the cluster 2 are also drawn toward the noisy data. All of these factors may cause the negative effects on the model estimation and the further analysis. The similar situations can also be observed in Fig. 4, in which the features of training samples of the first dataset are extracted by BP and PSD respectively. Here the covariances of clusters are also be enlarged, but the centers and the directions of the dispersions are not changed obviously, which may explain to some extent why these two methods have the similar performances for the first data set.

The parameters of the genetic-based EM algorithm are determined by experience. In particular, we found that the convergence of the algorithm can be guaranteed if we set C to be 10. Large C cannot further improve the performance but increase the execution time, and it is same to the size of the population $|P|$. On the other hand, the choice of the trimming level ϵ can affect the classification result. Unfortunately, this prior knowledge is not usually available for many practical applications. If ϵ is set too high, the remaining data may not be able to reflect the true distribution in the feature space, and too few data will also cause the overftting problem. Conversely, if ϵ is too small, some noisy data may still degrade the classification performance. In our experiment, we have found that $\epsilon \in [0.8, 0.9]$ usually provides desirable result. Since the noise is ubiquitous in EEG signals, we suggest that one should trim some training data before construct the classifier.

V.   DISCUSSION AND CONCLUSION

A robust clustering algorithm by combining GA with EM for GMM was introduced in this paper. The motivation of this work is to reduce the negative effects caused by the noisy data. The experimental results demonstrate that the performance of the classical GMM can be improved by the introduced method.

LDA outperforms the mixture model. However, in an unsupervised manner, the mixture model can reduce the training time since the mapping from the EEG pattern changes to the intention of the user can be learned online without any feedback [10]. If the GMM is trained in a supervised manner, the number of the Gaussian components should also be optimized by some information criterion (e.g., the Akaike information criterion (AIC) [28], Bayesian inference criterion (BIC) [29], minimum description length (MDL) [30]). In [31], an algorithm was proposed to select the number of the components and avoid the sensitivity to initialization, which can be used for number of Gaussian components selection.

Adaptability of an online BCI system should also be considered to cope with the non-stationarity in the EEG signal [11–13], and how to design the adaptation strategy for GMM is also one step in future work. Moreover, the idea of GA can also be adopted to choose the subset to design a subject-independent
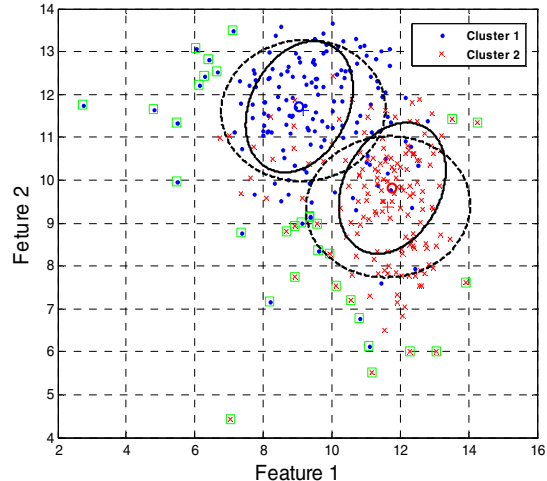

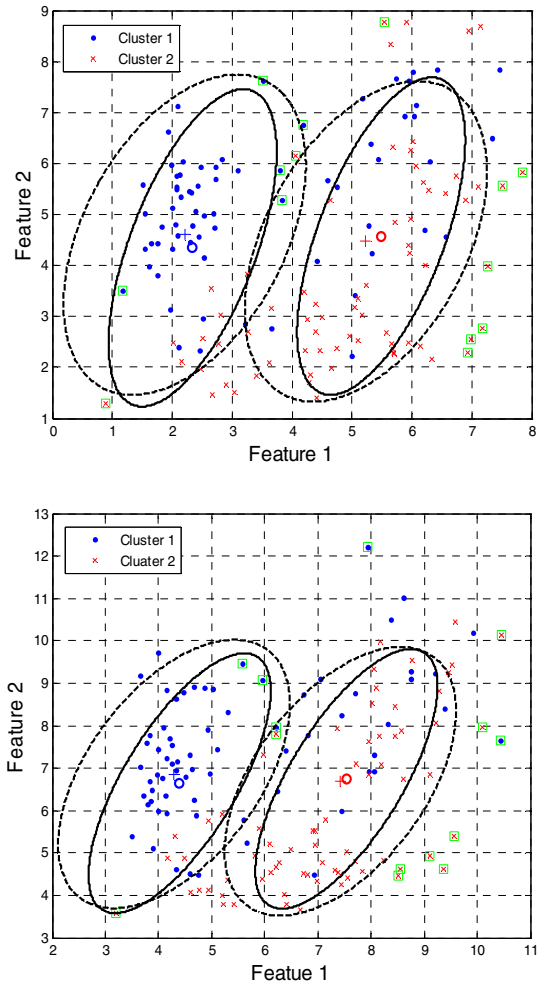Figure. 3.   The clusters estimated by classical GMM and robust method

Figure. 4.   The clusters estimated by classical GMM and robust method for features extracted by BP (top) and PSD (bottom)

BCI, which is robust to the variance between different BCI users, so that the users could start real-time BCI use without any prior calibration.

REFERENCES

[1] J. R. Wolpaw *et al.*, "Brain–computer interface technology: a review of the first international meeting," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 2, pp. 164–173, June 2000.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interface for communication and control," *Clin. Neurophsiol.,* vol. 133, pp. 767-791, June 2002.

[3] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals," *J. Neural Eng.*, vol. 4, pp. R32–R57, 2007.

[4] D. J. McFarland, C.W. Anderson, K.-R.Müller, A. Schlogl, and D. J. Krusienski, "BCI meeting 2005-workshop on BCI signal processing: feature extraction and translation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 135-138, June 2006.

[5] F. Lotte, M. Congedo, A. Lecuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 4, pp. R1–R13, 2007.

[6] K.-R.Müller, C.W. Anderson, and G. E. Birch, "Linear and nonlinear methods for brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, pp. 165–169, June 2003.

[7] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol.11, pp. 141-144, June 2003.

[8] K.-R.Müller, M. Krauledat, G. Dornhege, G. Curio, and B. Blankertz, "Machine learning techniques for brain-computer interfaces," *Biomed. Technol.*, vol. 49, pp.11-22, 2004.

[9] A. Buttfield, and J. del R. Millán, "Online classifier adaptation in brain-computer interfaces," *Tech. Report*, IDIAP–RR 06-16, March 2006.

[10] S. E. Eren, M.Grosse-Wentrup, and M. Buss, "Unsupervised classification for non-invasive brain-computer-interfaces," in *Proc. Automed Workshop*., Düsseldorf, Germany, 2007, pp. 65–66.

[11] B. A. S. Hasan, and Q. J. Gan, "Unsupervised adaptive GMM for BCI," in *Proc. 4th Int. IEEE EMBS Conf. Neural Engineering*, Antalya, Turkey, 2009, pp. 295-298.

[12] B. A. S. Hasan, and Q. J. Gan, "Sequential EM for unsupervised adaptive Gaussian mixture model based classifier," in *Proc. 6th Int. Conf. Machine Learning and Data Mining in Pattern Recognition*, Leipzig, Germany, 2009, pp. 96-106.

[13] G. Liu, G. Huang, J. Meng, D. Zhang, and X. Zhu, "Improved GMM with parameter initialization for unsupervised adaptation of brain-computer interface," *Int. J. Commun. Numer. Meth. Eng.*, to be published.

[14] G. Schalk, P. Brunner, L. A. Gerhardt, H. Bischof, and J. R. Wolpaw, "Brain-computer interfaces (BCIs): Detection instead of classification," *Neurosci. Meth.*, vol. 167, pp. 51-62, Jan 2008.

[15] S. Fazli, M. Danóczy, F. Popescu, B. Blankertz, and K.-R.Müller, "Using rest class and control paradigms for brain computer interfacing," in *Proc. 10th Int. Work-Conf. Artificial Neural Networks*, Salamanca, Spain, 2009, pp. 651-665.

[16] F. Lotte, C. Guan, and K. K. Ang, "Comparison of designs towards a subject-independent brain-computer interface based on motor imagery," in *Proc. 31th Annu. Int. Conf. IEEE EMBS*, Minneapolis, USA, 2009, pp. 4543-4546.

[17] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer Press, New York, 2006.

[18] D.T. Nguyen, L. Chen, and C. K. Chan, "An outlier-aware data clustering algorithm in mixture model," in *Proc. 7th Int. Conf. Information, Communication and Signal Processing*, Macau, China, 2009, pp. 1-5.

[19] G.J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, Wiley, New York, 1997.

[20] D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Professional, 1989.

[21] J. R. Millán, F. Renkens, J. Mouriño, and W. Gerstner, "Brain-actuated interaction," *Artif. Intell.*, vol. 159, pp. 241-259, Nov. 2004.

[22] J. R. Millán, "On the need for on-line learning in brain–computer interfaces," in *Proc. Int. Joint Conf. Neural Networks*, Budapest, Hungary, 2004, pp. 2877–2882.

[23] S. Sun, C, Zhang, and N. Lu, "On the on-line learning algorithms for EEG signal classification in brain computer interfaces," *Lect Notes Comput Sci*, vol. 3614, pp. 638-647, 2005.

[24] S. Sun, and C, Zhang, "Learning on-line classification via decorrelated LMS algorithm: application to brain–computer interfaces," *Lect Notes Comput Sci*, vol. 3735, pp. 215-226, 2005.

[25] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals," *J. Neural Eng.*, vol. 4, pp. R32–R57, 2007.

[26] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 4, pp. 441-446, Dec. 2000.

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

[28] H. Akaike, "A new look at the statistical model identification," *IEEE Trans Aut. Ctrl.*, vol. 19, no.6, pp. 716-723, 1974.

[29] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, pp. 461-464, 1978.

[30] J. Rissanen, "Modeling by shortest data description," *Automatica.*, vol.14, pp. 465-471, 1978.

[31] M. Figueiredo and A. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381–396, Mar. 2002

[32] BCI competition. Available: http://www.bbci.de/competition/ii/