A BCI motor imagery experiment based on parametric feature extraction and Fisher Criterion

David D'Croz-Baron[1], Juan Manuel Ramirez[1], Mary Baker[2], Vicente Alarcon-Aquino[3], Obed Carrera[1]
[1]Department of Electronics, INAOE, Tonantzintla, Puebla, Mexico
[2]Department of Electrical Engineering, Texas Tech University, Lubbock, Texas, USA
[3]Department of Computing, Electronics and Mechatronics, UDLAP, Puebla, Mexico
d.dcroz@gmail.com; jmram@inaoep.mx; mary.baker@ttu.edu; vicente.alarcon@udlap.mx, obed.carrera@inaoep.mx

Abstract

An EEG-based classification method in the time domain is proposed to identify left and right hand motor imagery as part of a brain-computer interface (BCI) experiment. The feature vector is formed by sixth order autoregressive coefficients (AR) or sixth order adaptive autoregressive coefficients (AAR) representing EEG signals obtained from C3 and C4 channels, according to the EEG 10-20 standard. The signal is analyzed considering 1 second windows with a 50% overlapping. A feature selection process based on the Fisher Criterion (FC) removes irrelevant or noisy information. A Linear Discriminant Analysis (LDA) is applied to both cases: feature vectors formed with the total number of coefficients, and feature vectors formed with coefficients corresponding to larger Fisher Ratio. Classification results obtained using two AR methods, Burg and Levinson-Durbin, and one AAR LMS are presented.

Keywords

Autoregressive coefficients (AR), Adaptive Autoregressive Coefficients (AAR), Brain Computer Interfaces (BCI), EEG, Fisher Criterion (FC).

I. INTRODUCTION

A brain-computer interface (BCI) is a term broadly used to describe a system which translates the electrical signals generated from cognitive processes into control signals for a variety of applications, such as computer controls, speech synthesizers, or mechanical prostheses. The ongoing electroencephalographic signals (EEG) contain information associated to movements, mental tasks or mental responses related to some stimuli. These signals are analyzed and processed through several mathematical techniques to extract useful information represented in the form of feature vectors, which are then translated into meaningful control commands. An important purpose of a direct BCI is to allow individuals with motor disabilities such as locked-in syndrome, which can be caused by amyotrophic lateral sclerosis, high-level spinal cord injury, or some other severe health conditions, to have some control over external devices [1, 2]. In those cases BCIs would evidently lead to an improvement in the quality of life of affected people. The electrophysiological activity addressed in this work is sensorimotor activity, specifically motor imagery, which corresponds to the situation when a person imagine some movements, for example, right leg, left arm, tongue, etc. The feature extraction process can be carried out in time or frequency domains. In the first case, parametric methods such autoregressive (AR), moving average (MA) or adaptive autoregressive (AAR) models [3,4], can be used. In the second case, representative analysis techniques in the frequency domain include Fourier and Wavelet analysis, Wigner-Ville distribution, or Empirical Mode Decomposition [5-7]. Some feature extraction techniques using information from both domains applied simultaneously have also been reported [8]. In [2], Cho et al. used an AR feature vector of each channel (C3 and C4) selecting two reactive bands per channel with a 77% of correct classification. In [3], Huan and Palaniappan compare different segmentation techniques using AR methods and a neural network as a classifier. Wang et al. [8] presented a feature selection analysis showing that noisy features can be removed leading to better classification rates. In this paper we present results obtained on the classification of left and right hand motor imagery using two different methods: 1) a composed feature vector formed with the coefficients of a sixth order autoregressive model (AR) for each channel C3 and C4 according to the 10-20 standard, and 2) a composed feature vector formed with the coefficients of a sixth order adaptive autoregressive model (AAR) for each channel C3 and C4. A comparison of classification performance based on linear discrimination analysis (LDA) after feature selection of coefficients using Fisher Criterion (FC) is presented.

II. METHODS

Figure 1 shows a flowchart of the experimental process involved in the BCI two-class experiment presented in this work.

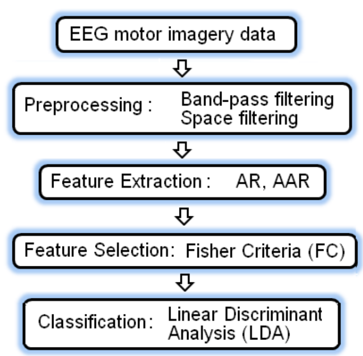

Figure 1. Block diagram of the experimental process.

A. Feature extraction: AR parameters.

Autoregressive analysis aims to identification, estimation, and forecasting of a time series, with spectral content represented inherently in the model parameters or coefficients of a transfer function, assuming some constraints. A typical approach for estimating the time-varying coefficients of an AR model is performing time segmentation of the signal; the AR coefficients are obtained from each time segment. The result is a time-course of the AR coefficients which inherently contains information about the time-varying characteristics of the process. In this work, the EEG time series obtained from the BCI experiment are fitted with an AR model in order to generate the feature vectors for classification purposes. A real-valued, zero-mean, stationary, autoregressive process can be expressed as:

$$y[n] = \sum_{i=1}^{P} a_i y[n-i] + x[n], \tag{1}$$

$$\boldsymbol{x}[\boldsymbol{n}] = N\{0, \sigma_x^2[n]\},$$

where p is the order model, $y[n]$ is the time series at sampled point n, a_i are the real-valued AR coefficients and $x[n]$ represents the error term independent of past samples. The error term is assumed to be a zero-mean noise with finite variance. The optimum model order is a trade-off between model complexity and the fitness. It depends on segment length, frequency components, sampling frequency and applications. Akaike's Final Prediction Error FPE is used in order to obtain an optimum model order [4] and it is given by equation 2, where N is the segment length, p is the model order and E_p is the Final Error.

$$FPE(p) = \left(\frac{N+p+1}{N-p-1}\right) E_p \tag{2}$$

AR modeling may yield poor estimates if the segment length is too short. Since EEG signals are non-stationary, too high segment lengths may not be completely described by AR parameters. For that reason in each trial the EEG signals were divided into 1 second window segments with 500ms overlap in order to improve time resolution, as shown in figure 1. It is important to note that database IIIA has different sampling frequency Fs, and AR parameters can be compared only if they are extracted from a signal or time series with the same Fs.

In this work, two autoregressive methods were used for comparison purposes: Burg's method and Levinson Durbin's recursive method. A linear discriminant analysis was used in both cases, with the goal of comparing classification performance. The feature vector is represented by equation 3, where C3 and C4 indicate the used channel according to the 10-20 standard denomination for EEG signals acquisition.

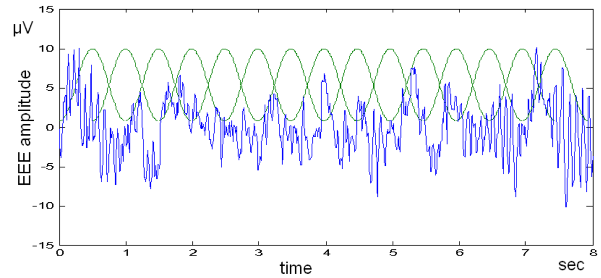

Fig.2 Typical EEG signal with a 500ms overlap windowing

The sixth order AR coefficients are represented by a_i in equation 3:

$$\boldsymbol{d_i} = \{a_1^{C3}, a_2^{C3}, \ldots, a_6^{C3}, a_1^{C4}, a_2^{C4}, \ldots, a_6^{C4}\} \tag{3}$$

B. Feature extraction: AAR parameters

In order to consider the non-stationarity of EEG signals, the AR parameters may change with time recursively. This yields to the use of Adaptive Autoregressive parameters shown in equation 4. The basic difference between equations 1 and 4 is that a_i coefficients are time-dependent, i.e., each time k a new set of coefficients arises.

$$y[n] = \sum_{i=1}^{P} a_i[k] y[n-i] + x[n], \tag{4}$$

$$\boldsymbol{x}[\boldsymbol{n}] = N\{0, \sigma_x^2[n]\},$$

In equation 4, p is the order model, $y[n]$ is the time series at sampled point n, $a_i[k]$ are the real-valued AR coefficients and $x[n]$ represents the error term independent of past samples. The error term is assumed to be a zero-mean noise with finite variance. For comparison reasons, the order p will be the same as the AR coefficients.

The AAR parameters are very attractive because they offer a high time resolution and they are calculated recursively by different methods. Recursive methods have in common the calculation of the prediction error given by equation 5. In this work, a least mean square (LMS) algorithm was used. LMS is briefly described as follows:

$$e[n] = y[n] - \boldsymbol{\hat{a}}[\boldsymbol{n-1}]\boldsymbol{y}[\boldsymbol{n-1}] \tag{5}$$

$$\boldsymbol{\hat{a}}[\boldsymbol{n}] = \boldsymbol{\hat{a}}[\boldsymbol{n-1}] + \left(\frac{UC}{MSY}\right) e[n]\boldsymbol{y}[\boldsymbol{n-1}] \tag{6}$$

Equations 5 and 6, $\boldsymbol{\hat{a}}[\boldsymbol{n}]$ represents the vector of AR coefficients at time sample n, $\boldsymbol{\hat{a}}[\boldsymbol{n-1}]$ is the set of coefficients in the previous sampled point, UC is the update coefficient, MSY is the variance of the signal $\boldsymbol{y}[\boldsymbol{n}]$, $e[n]$ is the prediction error and $\boldsymbol{y}[\boldsymbol{n-1}]$ is the vector composed by the p previous samples of the signal $\boldsymbol{y}[\boldsymbol{n}]$ at the sampled point n.

C. Feature Selection based on Fisher Criteria.

Feature selection algorithms are used to find the most informative features which increase the percent of classification. A feature selection process is particularly relevant in problems with high dimensional input data, and it can reduce the complexity of the classification problem [1]. In this work, Fisher Criterion (FC) is used to remove features that are noisy or irrelevant; a detailed description of the algorithm can be found in [9]. It basically considers C class labels, with n_i training vectors on each one, so the a priori probability for a class i is given by $P_i = {n_i}/{\sum_{i=1}^{C} n_i}$.

The mean for each class is estimated as $\widehat{\mu_\iota} = \frac{1}{n_i}\sum_{j=1}^{n_i} \boldsymbol{x_j^i}$, and the gross mean is given by $\hat{\mu} = \sum_{i=1}^{C} P_i \widehat{\mu_\iota}$.

The covariance matrix for each class is estimated by equation 7, the within-class scatter matrix is obtained through equation 8, and the between-class scatter matrix is given by equation 9. It is well known that the scatter matrices obtained through equations 5 and 6 contain information about class separability in a classification problem.

$$\widehat{S_\iota} = \frac{1}{n_i}\sum_{j=1}^{n_i}\left(\boldsymbol{x_j^i} - \widehat{\boldsymbol{\mu_\iota}}\right)\left(\boldsymbol{x_j^i} - \widehat{\boldsymbol{\mu_\iota}}\right)^T \quad (7)$$

$$S_w = \sum_{i=1}^{C} P_i \widehat{S_\iota} \quad (8)$$

$$S_b = \sum_{\boldsymbol{i=1}}^{\boldsymbol{C}} P_i(\widehat{\boldsymbol{\mu_\iota}} - \widehat{\boldsymbol{\mu}})(\widehat{\boldsymbol{\mu_\iota}} - \widehat{\boldsymbol{\mu}})^T \quad (9)$$

Finally, the Fisher Criterion for the *k*th feature is given by the ratio between the *k*th elements in the diagonal of matrices 8 and 9, as expressed in equation 10. Features with greater Fisher Ratio are more important, while features with lower Fisher Ratio are considered noisy features.

$$Fisher(k) = \frac{S_b^{(k)}}{S_w^{(k)}} \quad (10)$$

D. Linear Discriminant Analysis (LDA).

LDA is a linear classification method, fast and computationally attractive, which is used in this work as a basic method of exploring data separability on the described experiments. Basically, LDA is a method for identifying the best discriminating hyperplane in an n-dimensional feature space. Further experiments using additional classification methods, such as Support Vector Machine (SVM) and neural networks are currently in progress. LDA classification is given by equation 11:

$$D_i = \boldsymbol{w^T} * \boldsymbol{d_i} + w_0 \quad (11)$$

The weight vector $\boldsymbol{w}$ and threshold w_0 are obtaining using equations 12 and 13, respectively, where $\boldsymbol{S_w^{-1}}$ is the same as in equation 8; $\widehat{\boldsymbol{\mu}}$ is the gross mean, and $\widehat{\boldsymbol{\mu}}_{\boldsymbol{1}}, \widehat{\boldsymbol{\mu}}_{\boldsymbol{2}}$ are the estimated mean for each class.

$$\boldsymbol{w^T} = \boldsymbol{S_w^{-1}} * (\widehat{\boldsymbol{\mu}}_{\boldsymbol{1}} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{2}}) \quad (12)$$

$$w_0 = \boldsymbol{w^T} * \widehat{\boldsymbol{\mu}} \quad (13)$$

III. EXPERIMENTS

A. Experimental paradigm and databases.

The EEG data used in this work was obtained from two sources: 1) the public repository of the BCI Competition III (IIIA and IIIB), available to the international community for academic and research purposes, and 2) an own database generated according to the experimental paradigm of the mentioned competitions, which was obtained in the Autumn's Dawn NICE (Neuro- Imaging Cognition and Engineering) Laboratory at Texas Tech University.

The data base IIIA corresponds to experiments with imagined left and right hand, a foot and tongue movements. A detailed description about the signal acquisition and experimental paradigm can be found in [10]. Each trial started with an empty black screen; at time point t = 2 s a short beep tone was presented and a cross '+' appeared on the screen to raise the subject's attention. Then at second 3 (t = 3 s) an arrow appearing for 1.25 s pointed either to the left, right, upwards or downwards. Each position of this arrow instructed the subject to imagine the movements or left hand, right hand, tongue or foot [10]. Only left and right hand movements are addressed in this work.

The data base IIIB was obtained from the same public repository and corresponds to BCI experiments with imagined left and right hand movements. A detailed description about the signal acquisition and experimental paradigm can be found in [11]. The EEG was obtained using a sampling frequency of 125 Hz, and passband filtered between 0.5 and 30Hz. Each trial started with the presentation of a fixed cross at the center of the monitor, followed by a short warning tone at 2s. At 3s, the fixed cross was overlaid with an arrow at the center of the monitor for 1.25 s, pointing either to the left or to the right as experimental cue. Depending on the direction of the arrow, the subject was instructed to imagine a movement of the left or the right hand.

The database TTU was obtained at the Autumn's Dawn NICE (Neuro- Imaging Cognition and Engineering) Laboratory, Texas Tech University, and corresponds to EEG experiments with imagined left and right movements for two subjects. The EEG was obtained using a sample frequency of 125Hz, and passband filtered between 0.3 and 30Hz. The trial duration is 7 seconds, and in each trial the subject was instructed to imagine only one movement depending on the direction of the arrow at time t= 3s.

Characteristics of all the data bases are summarized in table I. Database IIIA and TTU were obtained using a reference electrode, and database IIIB using bipolar electrodes.

TABLE I
Characteristics of databases obtained from BCI competition.

Data Base	Subjects	Fs	Number Electrodes	Trial Duration
IIIA	S1,S2,S3	250 Hz	60	7 sec
IIIB	S4,S5,S6	125 Hz	3	8 sec
TTU	S7, S8	125 Hz	64	7 sec

IV. RESULTS

Figure 3 depicts the Final Prediction Error (FPE) obtained from the experiment corresponding to the subject S3 using a 3.0-4.0 seconds epoch, and several AR model orders. A sixth order AR model was selected to be used in the described experiments. A series of experiments using the described methods for extracting AR coefficients, Burg's method and Levinson Durbin's recursion, are presented in Table II. The table summarizes the results obtained from the 8 subjects. From these results it is evident that Burg's algorithm yields to a better classification than Levinson-Durbin's Recursion. The results were obtained using LDA as the classification method. The percent of recognition was calculated as the number of correct recognitions over the total number of trials of each class using the data test, which is the half part of the data for each subject. Table III summarizes the results obtained from the 8 subjects using LMS algorithm to calculate the AAR coefficients with a coefficient UC=0.004.

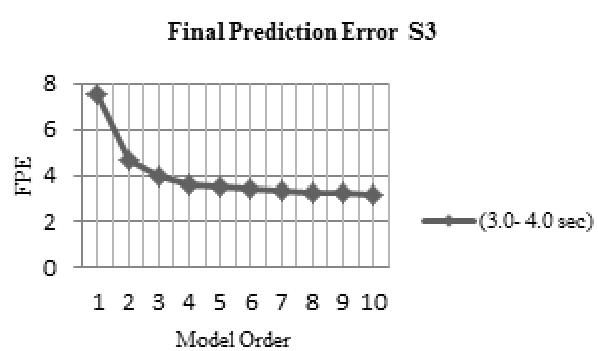

Fig 3. Final Prediction Error for subject 3 with different model orders.

TABLE II.
COMPARISON BETWEEN BURG'S METHOD AND LEVINSON-DURBIN RECURSION FOR 6^{TH} ORDER AR

	Burg's Algorithm			Levinson- Durbin's recursion		
Subject	**Left(%)**	**Rigth (%)**	**Total**	**Left(%)**	**Rigth (%)**	**Total**
S1	86.7	80	83.4	80	73.3	76.7
S2	81.8	86.4	84.1	72.7	90.9	81.8
S3	71.4	71.4	71.4	64.3	78.6	71.4
S4	76.7	66.7	71.7	66.7	68.3	67.5
S5	81.7	80	80.8	86.7	76.7	81.7
S6	66.7	75	70.9	68.3	68.3	68.3
S7	63.8	78.7	71.3	70.2	76.6	73.4
S8	64.5	74.2	69.4	67.7	64.5	66.1
Total	**74.2**	**76.6**	**75.4**	**72.1**	**74.7**	**73.4**

TABLE IIII.
PERCENT OF CLASSIFICATION FOR 6^{TH} ORDER AAR CALCULATED BY LMS ALGORITHM AND UC00.004.

	LMS Algorithm		
Subject	**Left(%)**	**Rigth (%)**	**Total**
S1	83.4	93.3	88.4
S2	81.8	95.5	88.7
S3	71.4	85.7	78.6
S4	70	83.3	76.7
S5	73.3	66.7	70
S6	85	68.3	77
S7	57.4	87.2	71.3
S8	74.2	67.7	71
Total	**74.6**	**81.8**	**77.8**

According to the method described in Figure 1, a feature selection process based on Fisher criterion was incorporated previous to the classification process. Since Burg's method offered better results than Levinson-Durbin's Recursion, it was the method used to compare the classification ratio with Fisher Criterion. In each trial, the EEG signals were divided into 1 second window segments using a rectangular window with 500ms overlap. In every case, the use of Fisher Criterion to select the best features yielded to a better percent of classification. The total percent of classification for each subject is presented in Table IV using Burg's method for AR algorithm and LMS method for AAR algorithm to extract features and Fisher Criterion as the feature selection method. The selected features show the minimum features needed to obtain the higher percent of classification.

TABLE IV.
COMPARISON BETWEEN THE PERCENT OF CLASSIFICATION BETWEEN AR (BURG'S METHOD) AND AAR (LMS) ALGORITHMS

	AR Algorithm		AAR Algorithm	
	Total Percent of Classification	Selected Features	Total Percent of Classification	Selected Features
S1	86.7	2	93.4	11
S2	95.5	3	88.7	3
S3	78.6	8	78.6	8
S4	73.8	11	80	10
S5	82.5	3	75	10
S6	73.8	4	79.2	10
S7	80.9	9	81.9	5
S8	74.2	2	74.2	10
Total	80.8		81.4	

V. Conclusion

In this paper we presented a BCI motor imagery experiment based on parametric feature extraction and Fisher criterion as a feature selection process. The aiming was discrimination between left and right hand motor imagery based on a sixth order AR or AAR parameters and a selection of those features using Fisher Criterion. The results show that the composed feature vector using AR or AAR for feature extraction and Fisher Criterion to select some coefficients is an adequate and effective technique in motor imagery recognition. The feature extraction method performed by using autoregressive AR analysis is described. A comparison between two methods to estimate AR parameters, indicated that Burg's method provide a better performance in classification, minimizing both forward and backward prediction error. In every case, the feature selection process based on Fisher criterion improved the classification performance. When the Fisher ratio of the elements in the feature vector assumes similar values, a feature reduction does not lead to better classification rates. In those cases it is possible that the feature selection is not contributing to eliminate noisy characteristics, but rather, important features are being removed, with a consequent reduction in classification performance. Additional experiments for comparison purposes using neural networks and Support Vector Machine in the classification stage, are currently in progress.

Acknowledgments

The first author gratefully acknowledges the financial support from the Mexican National Council for Science and Technology (CONACYT), and the Autumn's Dawn NICE (Neuro- Imaging Cognition and Engineering) Laboratory at Texas Tech University for the research stay.

References

[1] A. Bashashati, M. Fatourechi, R. K. Ward, "A survey of signal processing algorithms in brain- computer interfaces based on electrical brain signals", Journal of Neural Engineering, R32-R57, 2007.

[2] S. Cho, J. Ae Kim, D. Hwang, S. Kee Han, "Single Trial Discrimination between Right and Left Hand Movement-Related EEG Activity", Lecture notes in Computational Science Vol. 3036. pp. 255- 262, 2004.

[3] N. Huan and R. Palaniappan, "Neural network classification of autoregressive features from electroencephalogram signals for brain-computer interfaces design". Journal of Neural Engineering 1, pp. 142-150, 2004.

[4] Lin Liu, Qing Guo Wei, "A Research Combining Spatial Filter with Autoregressive Model for EEG Feature Extraction", Journal of Applied Mechanics and Materials, Vol. 20, pp- 605-611, 2010.

[5] X. Bao-Guo and S. Ai-Guo, "Pattern recognition of motor imagery EEG using Wavelet transform", Journal of Biomedical Science and Engineering 1, pp. 64-67, 2008.

[6] Qin L and He B, "A wavelet-based time-frequency analysis approach for classification of motor imagery for brain–computer interface applications", Journal of Neural Engineering, Vol. 2, pp. 65–72, 2005.

[7] Carlos Guerrero-Mosquera, Armando Malanda Trigueros, Jorge Iriarte Franco, Angel Navia-Vázquez,"New feature extraction approach for epileptic EEG signal detection using time-frequency distributions", Medical and Biological Engineering and Computing, Vol. 48, No. 4, pp.321-330, 2010.

[8] Baoguo Xu, Aiguo Song, Juan Wu, "Algorithm of Imagined Left-Right Hand Movement Classification Based on Wavelet Transform and AR Parameter Model", The 1st International Conference on Bioinformatics and Biomedical Engineering, ICBBE 2007, pp. 539 – 542, 6-8 July 2007.

[9] S.A. Wang, L. Cheng-Lin, L. Zheng, "Feature selection by combining Fisher Criterion and Principal Feature Analysis", Proceeding of the sixth international conference on machine learning and cybernetics Hong Kong, pp. 19-22, 2007.

[10] A. Schlögl, F. Lee, H. Bischof and G. Pfurtscheller, "Characterization of four-class motor imagery EEG data for the BCI-competition 2005", J. Neural Eng. 2 , L1–L9, 2005.

[11] C. Vidaurre, A. Schlögl, R. Cabeza, G. Pfurtscheller, "A fully on-line adaptive brain computer interface", Biomed. Tech. Band, special issue, vol. 49, 2004.

[12] J. Pardley, S.J. Roberts, L. Tarassenko, "A review of parametrical model techniques for EEG analysis", Medical Engineering and Physics, Vol. 18, No. 1, pp. 2-11, 1996.

[13] K. Fukunaga, "Introduction to statistical pattern recognition". 2nd edition. Academic Press, 1990.

[14] R. Palaniappan, "Identifying individuality using mental task based brain-computer interface", Third International Conference on Intelligent Sensing and Information Processing ICISIP 2005, pp. 238- 242, 2005.

[15] Dennis J McFarland and Jonathan R Wolpaw, "Sensorimotor rhythm-based brain–computer interface (BCI): model order selection for autoregressive spectral analysis", Journal of Neural Engineering, Vol. 5, No. 2, pp. 155-162, 2008.

[16] A. Schlögl, S.J. Roberts, G. Pfurtscheller, "A criterion for adaptive autoregressive models". Proceedings of the 22nd annual international conference of the IEEE Engineering in Medicine and Biology Society, Vol. 2, pp. 1581-1582, 2000.