INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF NEURAL ENGINEERING

J. Neural Eng. **1** (2004) 142–150 PII: S1741-2560(04)82457-1

Neural network classification of autoregressive features from electroencephalogram signals for brain–computer interface design

Nai-Jen Huan[1] and Ramaswamy Palaniappan[2]

[1] Faculty of Information Science and Technology, Multimedia University, Melaka, Malaysia
[2] Biomedical Engineering Research Centre, Nanyang Technological University, Singapore

Received 14 June 2004
Accepted for publication 3 August 2004
Published 31 August 2004
Online at stacks.iop.org/JNE/1/142
doi:10.1088/1741-2560/1/3/003

Abstract

In this paper, we have designed a two-state brain–computer interface (BCI) using neural network (NN) classification of autoregressive (AR) features from electroencephalogram (EEG) signals extracted during mental tasks. The main purpose of the study is to use Keirn and Aunon's data to investigate the performance of different mental task combinations and different AR features for BCI design for individual subjects. In the experimental study, EEG signals from five mental tasks were recorded from four subjects. Different combinations of two mental tasks were studied for each subject. Six different feature extraction methods were used to extract the features from the EEG signals: AR coefficients computed with Burg's algorithm, AR coefficients computed with a least-squares (LS) algorithm and adaptive autoregressive (AAR) coefficients computed with a least-mean-square (LMS) algorithm. All the methods used order six applied to 125 data points and these three methods were repeated with the same data but with segmentation into five segments in increments of 25 data points. The multilayer perceptron NN trained by the back-propagation algorithm (MLP-BP) and linear discriminant analysis (LDA) were used to classify the computed features into different categories that represent the mental tasks. We compared the classification performances among the six different feature extraction methods. The results showed that sixth-order AR coefficients with the LS algorithm without segmentation gave the best performance (93.10%) using MLP-BP and (97.00%) using LDA. The results also showed that the segmentation and AAR methods are not suitable for this set of EEG signals. We conclude that, for different subjects, the best mental task combinations are different and proper selection of mental tasks and feature extraction methods are essential for the BCI design.

1. Introduction

Brain–computer interface (BCI) designs might give new lives to those individuals who are completely paralyzed. With this motive, for the last 10 years, the volume and pace of BCI research have grown tremendously [1, 2]. In 1995, there were no more than six active BCI research groups, and in the year 2000 there were more than 20 [2]. BCI designs are very useful for completely paralyzed individuals[3] to communicate with their external surroundings using their brain thoughts. These individuals could have become completely paralyzed after being involved in an accident (which may cause cerebral palsy, spinal cord injuries) or due to some disease such as amyotrophic lateral sclerosis, brainstem stroke and muscular

[3] Individuals who have lost all forms of control over their peripheral nerves and muscles.

dystrophies. BCI design is also suitable for use in simple hands off menu selection on the screen.

There are a few non-invasive methods for obtaining the brain signals to be utilized in a BCI design. Electroencephalogram (EEG) signals recorded at the scalp during some mental tasks have been used by some of the research groups [3–5]. Some others utilize single-trial visual evoked potential (VEP) signals where the subjects gaze at a screen of letters or menus [6]. Synchronization and desynchronization of μ-rhythm extracted during sensory motor tasks is another method for BCI design [7]. Reviews of some of these technologies and developments in this area are given by Vaughan *et al* [1] and Wolpaw *et al* [2].

In this paper, a two-state BCI was designed using six different methods to extract features from EEG signals that were recorded during five different mental tasks from four different healthy subjects. These mental tasks were: geometrical figure rotation, mathematical multiplication, mental letter composing, visual counting and baseline-resting task. The proposed BCI design is individual BCI, which is suitable for use by a particular individual. Our results showed that we cannot expect to use a particular set of mental tasks for all the subjects, most likely due to the different thought patterns exhibited by different subjects.

The six different feature extraction methods that we used were:

- Autoregressive (AR) coefficients computed with Burg's algorithm using 125 data points.
- Adaptive AR (AAR) coefficients computed with a least-mean-square (LMS) algorithm using 125 data points.
- AR coefficients computed with a least-squares (LS) algorithm using 125 data points.
- All the three methods above were repeated using the same 125 data points but with segmentation into five segments in increments of 25 data points each[4].

All the methods used order six. After these features were computed, multilayer perceptron NN trained by the back-propagation algorithm (MLP-BP) was used to classify the ten different possible combinations of two mental tasks. The output of the two-state BCI design could be used to control the movement of a cursor to select a target on a computer screen, to move a wheel chair or to be used with some translation schemes such as Morse code [5], which would provide a channel for paralyzed individuals to communicate with their external surroundings. Such an application would involve on-line classification of the EEG signal of interest and using the ability to perform the mental tasks to execute commands. As an example, a Morse code scheme could possibly be used to translate the outputs of BCI design into letters/words such as 'water', 'tv', etc. For instance, the letter 'A' in Morse code could be represented by a dot followed by a dash, so the user needs to think of different mental tasks to 'spell' out the letter 'A'. For example, assuming that mental task 'letter' represents dot, and 'maths' represents dash, the user should perform a letter mental task, followed by a maths mental task.

[4] These are coefficients computed using data points 1–25, 1–50, 1–75, 1–100, 1–125.

Also, it was shown by Osaka [8] that arithmetic tasks exhibit a higher power spectrum in the right hemisphere whereas visual tasks do so in the left hemisphere. So one would expect a combination of arithmetic and visual tasks to be better for use with the Morse code scheme.

The aspect of interference between different thoughts should be considered for different applications. For example, it was reported that two verbal tasks or two visuospatial tasks which are performed simultaneously lead to interference, whereas verbal and visuospatial tasks performed simultaneously do not show this interference [9]. Thus, as an example, one would expect a combination of the letter and the counting task (both involving verbal processing) to be better for wheelchair movement (visuospatial task) than for verbal communication.

2. Methods

2.1. Data

The EEG data used in this study were collected by Keirn and Aunon [4]. The data that we used in this paper were actually the same set as Keirn and Aunon used in their paper [4]. The data are available online at http://www.cs.colostate.edu/~anderson. A brief description of the dataset follows. The recordings of mental tasks were conducted for several trials and the number of times that the tasks were repeated varied from one subject to another. There were all together seven subjects and the number of trials for each subject is tabulated below.

Subject	Number of trials
1	10
2	5
3	10
4	10
5	15
6	10
7	5

Each channel from each trial produced 2500 sample points for the 10 s recording because the amplified EEG traces were sampled and stored at 250 sample points per second. In our study, since half-second windows were used, we extracted the features using 125 data points. As a result, for each trial, we obtained 20 segments with each segment consisting of 125 data points. In this study, we planned to use the data points from 10 trials to give 200 patterns for each mental task for training and testing purposes. As a result, we discarded the data from subjects 2 and 7 since they completed only 5 trials and as for the subject 5 with 15 trials, we used only the first 10 trials in order to have a fair comparison with the other subjects. Even though subject 4 had completed 10 trials, the trials were not selected due to some missing data. As such, the selected subjects were 1, 3, 5 and 6 and for convenience, we denote them as subjects 1, 2, 3 and 4 in this paper.

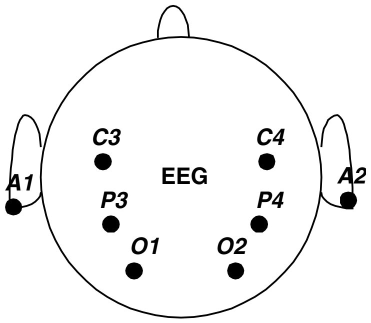

Figure 1. Electrode placement.

The subjects were seated in an industrial acoustics company sound controlled booth with dim lighting and a noiseless fan (for ventilation). An electro-cap elastic electrode cap was used to record EEG signals from positions C3, C4, P3, P4, O1 and O2 (shown in figure 1), defined by the 10–20 system [10] of electrode placement. The impedances of all electrodes were kept below 5 kΩ. Measurements were made with reference to electrically linked mastoids, A1 and A2. The electrodes were connected through a bank of amplifiers (Grass7P511), whose band-pass analogue filters were set at 0.1 to 100 Hz. The data were sampled at 250 Hz with a Lab Master 12-bit A/D converter mounted on a computer. Before each recording session, the system was calibrated with a known voltage. Signals were recorded for 10 s during each task and as mentioned earlier, data from 10 trials were used. Next, the EEG signal for each mental task was segmented into 20 segments with length 0.5 s. As the sampling rate was 250 Hz, each half-second EEG segment was 125 data points (samples) in length.

In this paper, EEG signals from four subjects performing five different mental tasks have been used. These mental tasks are:

(a) *Baseline task*. The subjects were asked to relax and think of nothing in particular. This task was used as a control and as a baseline measure of the EEG signals.
(b) *Math task*. The subjects were given nontrivial multiplication problems, such as 42 times 18, and were asked to solve them without vocalizing or making any other physical movements. The tasks were non-repeating and designed so that an immediate answer was not apparent. The subjects verified at the end of the task whether or not he/she arrived at the solution and no subject completed the task before the end of the 10 s recording session.
(c) *Geometric figure rotation task*. The subjects were given 30 s to study a particular three-dimensional block object, after which the drawing was removed and the subjects were asked to visualize the object being rotated about an axis. The EEG signals were recorded during the mental rotation period.
(d) *Mental letter-composing task*. The subjects were asked to mentally compose a letter to a friend or a relative without vocalizing. Since the task was repeated for several times the subjects were told to continue with the letter from where they left off.
(e) *Visual counting task*. The subjects were asked to imagine a blackboard and to visualize numbers being written on the board sequentially, with the previous number being erased before the next number was written. The subjects were instructed not to verbalize the numbers but to visualize them. They were also told to resume counting from the previous task rather than starting over each time.

Keirn and Aunon [4] specifically chose these tasks since they involve hemispheric brainwave asymmetry (except for the baseline task). As such, Keirn and Aunon and later Anderson *et al* [3] proposed that these tasks are suitable for brain–computer interfacing.

Keirn and Aunon [4] in their paper used 2 s segments that comprised 512 data points (artefact-free EEG signals) from a total of 2500 sample points recorded in 10 s for one session. Furthermore, the first quarter second (64 data points) of each 2 s segment was also used for analysis. However, in this paper, we used all the 2500 sample points recorded for 10 s without any form of filtering. This saves computational time and design cost. Keirn and Aunon used three feature extraction methods: AR coefficients computed with Burg's algorithm, power spectral density values computed using Wiener Khintchine theorem and the AR model, where they studied the classification performance across all subjects rather than individual subjects. Their paper showed the average per cent of classification accuracy ranging from ten task combinations from five subjects for the three different feature extraction methods, but in this paper, our focus is on individual BCI design. This is to accommodate the fact that each individual would have their own style and pattern of thinking, so we studied and analysed the results according to individual performance. Because the methods of using the data set varied from this study to theirs, a comparison of classification results would not be appropriate.

In this paper, we have used six different feature extraction methods to extract the feature from the EEG signals. In the first method, AR coefficients were computed using Burg's method [10–12] using 125 data points (one segment of 0.5 s) of the EEG signals. Model order six was used for this AR process based on the suggestions in [3, 4]. The second method was the same as the first method except that the AR coefficients were computed for the same 125 data points but segmented into five segments in increments of 25 data points. In other words, AR coefficients were computed for data points 1–25, 1–50, 1–75, 1–100 and 1–125. In the third method, AAR coefficients were computed using the LMS algorithm using 125 data points. The fourth method was the same as the third, but instead of using 125 data points, AAR coefficients were computed for data points 1–25, 1–50, 1–75, 1–100 and 1–125. For the fifth method, AR coefficients were estimated by the LS algorithm using 125 data points and the last method was the same as the fifth except that AR coefficients were computed for data points 1–25, 1–50, 1–75, 1–100 and 1–125.

The following discussion details in brief the different feature extraction processes.

2.2. Feature extraction

2.2.1. Sixth-order AR coefficients estimated using Burg's algorithm. A real-valued, zero-mean, stationary, autoregressive process of order p is given by

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + e(n), \qquad (1)$$

where p is the model order, $x(n)$ is the signal at the sampled point n, a_k are the real-valued AR coefficients and $e(n)$ represents the error term independent of past samples. The term autoregressive implies that the process $x(n)$ is seen to be regressed upon previous samples of itself. The error term is assumed to be a zero-mean noise with finite variance. In applications, the values of a_k have to be estimated from finite samples of data $x(1), x(2), x(3), \ldots, x(N)$.

In this paper, we used Burg's method [11–13] to estimate the AR coefficients. The method is more accurate as compared to other methods such as Levinson–Durbin as it uses the data point directly. Furthermore, the Burg algorithm uses more data points by minimizing both forward error and backward error. The Burg algorithm is given in the appendix.

In computing the AR coefficients, order six was used because other researchers [3, 4] have suggested the use of order six for AR process for mental task classification. Therefore, we had six AR coefficients for each channel, giving a total of 36 features for each EEG segment for a mental task. For the first method, AR coefficients were calculated using 125 data points and a total of 36 features were obtained for each EEG segment for a mental task. For the second method, AR coefficients were computed every 25 data points. As a result, a total of 180 features were obtained for each EEG segment for a mental task.

2.2.2. Sixth-order AAR. An AR model with time-varying coefficients of order p is defined by

$$Y_t = a_{1,t}Y_{t-1} + a_{2,t}Y_{t-2} + \cdots + a_{p,t}Y_{t-p} + E_t, \qquad (2)$$

where E_t is a white noise process [14, 15]. In this paper, in order to estimate the time-varying AR coefficients, we used the LMS algorithm [16] as follows:

$$E(t) = Y(t) - a_1(t)Y(t-1) - \cdots - a_p(t)Y(t-p)$$
$$a_k(t+1) = a_k(t) + cE(t)Y(t-k) \qquad (3)$$

where $c = f/\mathrm{var}(Y)$.

Two of the commonly used adaptive algorithms are the recursive-least-squares algorithm (RLS) and LMS [17], but the LMS algorithm was used here due to its computational simplicity [18, 19]. The difference of AAR to the AR model is that the parameters $a_{1,t}$, $a_{2,t}, \ldots, a_{p,t}$ can vary with time, however it is assumed that the parameters change only 'slowly' [15, 16]. Since there were six AAR coefficients estimated at any time point t, for the third method, we chose the six coefficients from the 125th data point as features to represent the EEG and overall, we had 36 features for each EEG segment from the six channels. The number of features is the same as the first method, which is to ensure that a fair comparison could be conducted later. For the fourth method, we computed six AAR coefficients from the 25th, 50th, 75th, 100th and 125th data points to give 30 features and overall, we had 180 features from six channels. Again, this ensured the same number of features for fair comparison with the second method.

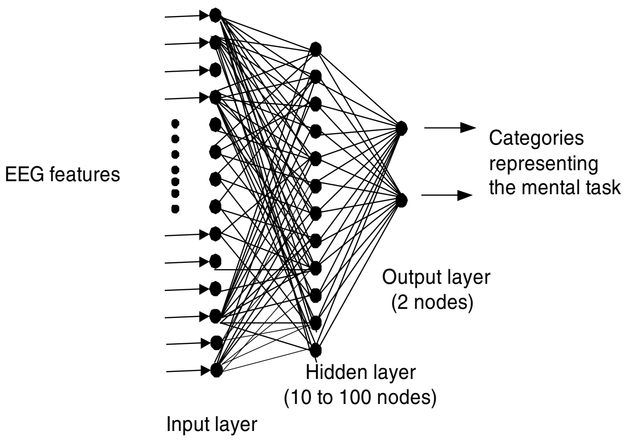

Figure 2. MLP-BP NN architecture.

2.2.3. Sixth-order AR coefficients with the LS algorithm. In this paper, we used the LS algorithm proposed in [20] in addition to Burg's method to estimate the AR coefficients and we computed the features for 125 data points. For this fifth method, we had 36 features for each EEG segment from six channels. This ensured a fair comparison with the first and third methods. As for the sixth method, AR coefficients were computed for data points 1–25, 1–50, 1–75, 1–100, 1–125 to obtain 180 features from six channels, which is similar to the second and fourth methods.

2.3. MLP-BP NN

In this paper, MLP NN with a single hidden layer trained by the BP algorithm [21] was used to classify different combinations of two mental tasks represented by the different EEG features. Figure 2 shows the architecture of the MLP-BP NN used in this study. The output nodes were set at two so that the NN could be classified into one of the two categories representing the mental task. The hidden layer nodes were varied from 20 to 100 in steps of 20.

A total of 200 EEG patterns (20 segments for EEG each signal × 10 trials) were used for each subject for each mental task in this experimental study. Therefore, for each simulation, there were 400 EEG patterns from two mental tasks, where half of the patterns were used in training and the remaining half in testing. The selection of the patterns for training and testing were chosen randomly. Training was conducted until the average error fell below 0.01 or reached a maximum iteration limit of 10 000. The average error denotes the error limit to stop NN training. The average error is the average of NN target output subtracted by the desired target output from all the training patterns. The desired target output was set to 1.0 for the particular category representing the mental task of the EEG pattern being trained, while for the other category, it was set to 0. Figure 3 shows the flow of the methodology.

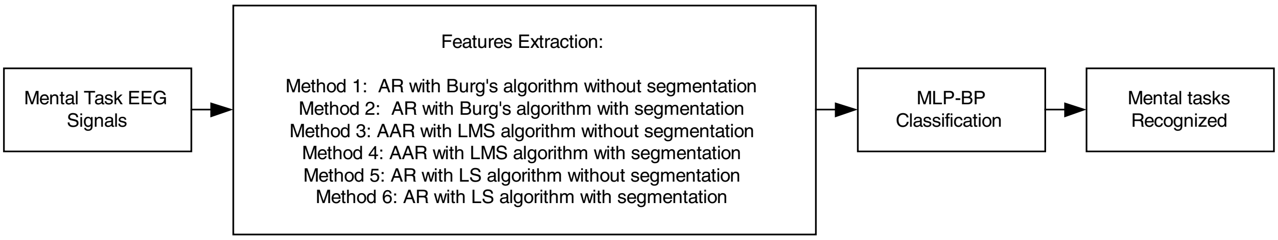

Figure 3. Flow of methodology.

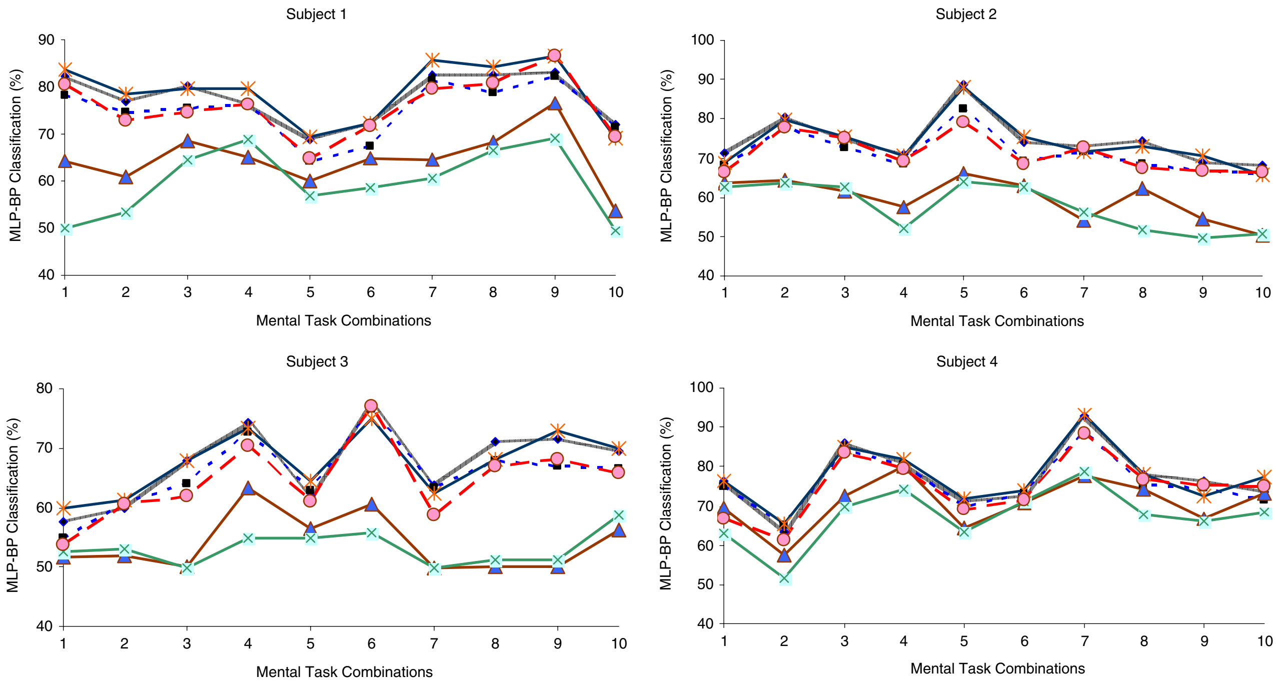

Figure 4. MLP-BP classification performance for all subjects.

2.4. Linear discriminant analysis (LDA)

LDA is one of the linear classification methods that require fewer examples in order to obtain a reliable classifier output [22]. LDA was used in addition to MLP-BP NN to compare the classification performances.

For the LDA method, assume that each data element s_i has m features and the number of examples is n where each example is assigned to one of the two classes $C = \{0, 1\}$. Then S is a matrix of size $n \times m$, and C is a vector of size n. N_0 and N_1 are the number of elements for classes 0 and 1, respectively.

The mean μ_c of each class c is the mean over all s_i with i being all elements in class c. The total mean μ of the data is

$$\mu = \frac{N_0\mu_0 + N_1\mu_1}{N_0 + N_1}. \tag{4}$$

The covariance matrix C of the data is the expectation value for

$$C = E\langle(s-\mu)^T(s-\mu)\rangle. \tag{5}$$

Then, the weight vector w and the offset w_0 are

$$w = C^{-1}(\mu_1 - \mu_0)^T \qquad w_0 = -\mu w. \tag{6}$$

The weight vector w determines a separating hyperplane in the m-dimensional feature space. The normal distance $D(x)$ of any element x is

$$D(x) = xw + w_0 = (x-\mu)w = (x-\mu)C^{-1}(\mu_1 - \mu_0)^T. \tag{7}$$

If $D(x)$ is bigger than 0, x is assigned to class 1 while if $D(x)$ is smaller than 0, x is assigned to class 0. However, if $D(x) = 0$, it means all elements of x are part of the separating hyperplane.

3. Results

Figures 4 and 5 show the classification performances of MLP-BP NN and LDA, respectively, for different combinations of two mental tasks using sixth-order AR coefficients computed using Burg's algorithm (AR-BG) with segmentation (AR-BG seg) and without segmentation (AR-BG no seg); sixth-order AR coefficients computed using the LS algorithm (AR-LS) with segmentation (AR-LS seg) and without segmentation

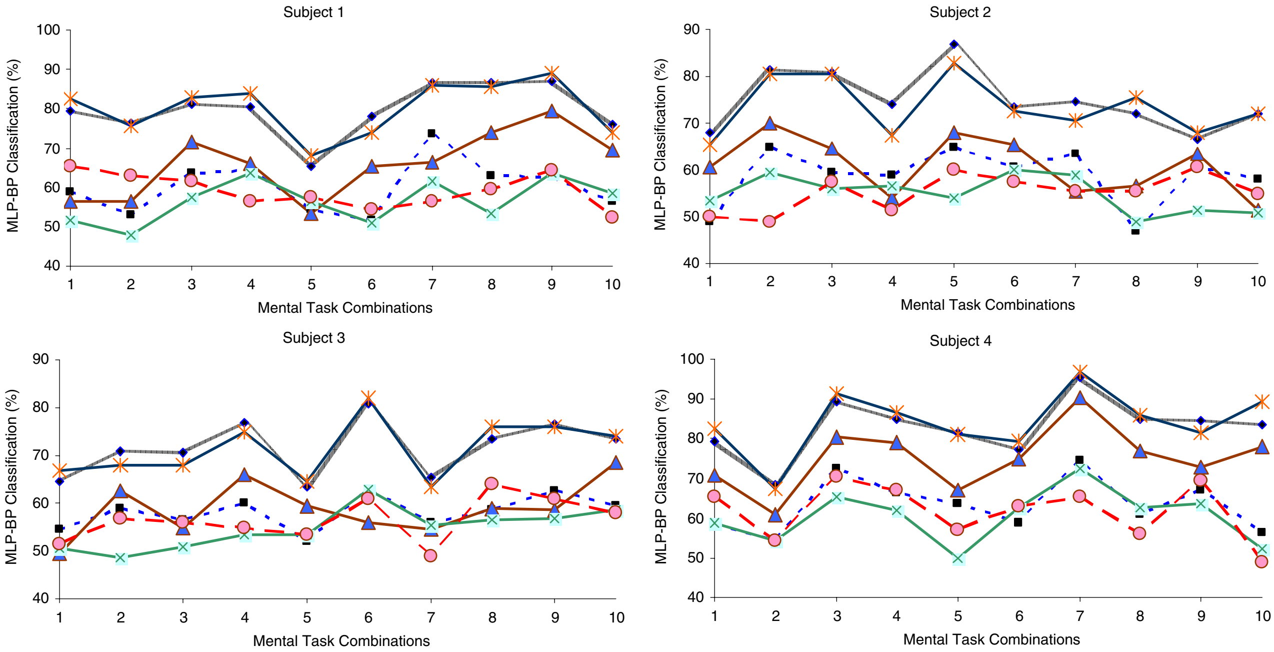

Figure 5. LDA classification performance for all subjects.

Table 1. Legends for figures 4 and 5.

No.	Mental task combination	Methods
1	Baseline, count	
2	Baseline, letter	AR-BG no seg
3	Baseline, maths	AR-BG seg
4	Baseline, rotation	AAR no seg
5	Letter, count	AAR seg
6	Letter, rotation	AR-LS no seg
7	Maths, count	AR-LS seg
8	Maths, letter	
9	Maths, rotation	
10	Rotation, count	

(AR-LS no seg); sixth-order AAR coefficients computed using the LMS algorithm (AAR) with segmentation (AAR seg) and without segmentation (AAR no seg) for all subjects. Table 1 gives the legend for figures 4 and 5. Tables 2 and 3 show the summary of results where the average performance, coefficient variation (standard deviation/mean) and maximum performance of each mental task combination for each method are given. The last column lists the method that gave the best performances for a particular mental task combination and its occurrences.

4. Discussion

From figures 4 and 5 and tables 2 and 3, it can be seen that for all the subjects, using EEG data with segmentation resulted in poorer classification performances as compared to EEG data without segmentation. The latter option not only performed better but was also simpler in terms of complexity and computational time. Therefore, it is suggested that for future work, one should consider classification without data segmentation (especially short segments) for AR representations of EEG signals. Also, none of the best mental task combinations for any of the subjects involved AAR method. The best mental task combination for all the subjects was either from AR-BG or AR-LS methods. These facts are true for both MLP-BP NN and LDA classifications.

For MLP-BP NN classification for subject 1, it can be seen that for all the methods, the best mental task combination was maths–rotation. For this mental task combination, the information transfer rate (bit rate) [23] was 0.2930. AR-LS without segmentation gave the maximum classification percentage of 86.70%. For subject 2, it can be seen that letter–count mental task combination gave the best classification performance for all the feature extraction methods with the maximum classification percentage of 88.80% given by the AR-BG method without segmentation with information transfer rate of 0.2407. As for subject 3, it can be seen that for both AR-BG and AR-LS methods (both without segmentation), the mental task combination that gave the best performance was letter—rotation with maximum classification percentage of 77.30% with the information transfer rate of 0.2132; whereas for the AAR method with and without segmentation gave baseline–rotation and rotation–count, respectively as the best mental task combinations. For subject 4, the AR-BG and AR-LS methods (both without segmentation) gave maths–count as the best mental task combination with maximum classification percentage of 93.10% with the information transfer rate of 0.4787 whereas with the AAR method, the best mental task combination was baseline–rotation.

From these discussions, it can be seen that for each subject, the best mental task combination was different: for

Table 2. Summary of MLP-BP NN classification results for all subjects.

		Classification performance (%)						
		AR-BG		AAR		AR-LS		
	Method	No seg	Seg	No seg	Seg	No seg	Seg	Best methods (occurrences)
Subject 1	Average	77.68	75.06	64.70	59.84	78.91	75.74	AR-LS no seg (eight times)
	Coefficient variation	0.07	0.08	0.09	0.12	0.08	0.08	
	Maximum	83.10	82.30	76.70	69.10	86.70	86.50	AR-BG no seg (two times)
	Best mental task combination	Maths, rotation	Maths, rotation	Maths, rotation	Maths, rotation	Maths, rotation	Maths, rotation	
Subject 2	Average	74.41	71.15	59.64	57.60	73.87	70.92	AR-LS no seg (three times)
	Coefficient variation	0.08	0.07	0.09	0.11	0.09	0.07	
	Maximum	88.80	82.50	65.90	64.10	88.00	79.00	AR-BG no seg (seven times)
	Best mental task combination	Letter, count	Letter, count	Letter, count	Letter, count	Letter, count	Letter, count	
Subject 3	Average	67.46	65.68	54.00	53.22	67.55	64.42	AR-LS no seg (six times)
	Coefficient variation	0.10	0.09	0.09	0.05	0.08	0.10	
	Maximum	77.30	76.65	63.40	58.80	74.90	77.10	AR-BG no seg (four times)
	Best mental task combination	Letter, rotation	Letter, rotation	Baseline, rotation	Rotation, count	Letter, rotation	Letter, rotation	
Subject 4	Average	77.14	75.83	70.67	67.45	77.58	74.68	AR-LS no seg (eight times)
	Coefficient variation	0.11	0.09	0.09	0.11	0.10	0.11	
	Maximum	92.70	88.70	80.10	78.60	93.10	88.30	AR-BG no seg (two times)
	Best mental task combination	Maths, count	Maths, count	Baseline, rotation	Maths, count	Maths, count	Maths, count	

AR-BG = AR with Burg's algorithm, AR-LS = AR with least-square algorithm, seg = segmentation, no seg = without segmentation.

Table 3. Summary of LDA classification results for all subjects.

		Classification performance (%)						
		AR-BG		AAR		AR-LS		
	Method	No seg	Seg	No seg	Seg	No seg	Seg	Best methods (occurrences)
Subject 1	Average	79.70	60.15	65.90	56.50	80.15	59.15	AR-LS no seg (five times)
	Coefficient variation	0.08	0.11	0.13	0.10	0.08	0.07	
	Maximum	87.00	73.50	79.50	63.50	89.00	65.50	AR-BG no seg (five times)
	Best mental task combination	Maths, rotation	Maths, count	Maths, rotation	Maths, rotation	Maths, rotation	Baseline, count	
Subject 2	Average	75.00	58.70	60.95	55.00	73.55	55.20	AR-LS no seg (two times)
	Coefficient variation	0.08	0.10	0.10	0.07	0.08	0.07	
	Maximum	87.00	65.00	70.00	60.00	83.00	60.50	AR-BG no seg (eight times)
	Best mental task combination	Letter, count	Letter, count	Baseline, letter	Letter, rotation	Letter, count	Maths, rotation	
Subject 3	Average	71.65	58.15	58.90	54.75	71.40	56.60	AR-LS no seg (five times)
	Coefficient variation	0.08	0.06	0.10	0.08	0.08	0.08	
	Maximum	81.00	63.00	68.50	63.00	82.00	64.00	AR-BG no seg (five times)
	Best mental task combination	Letter, rotation	Letter, rotation	Rotation, count	Letter, rotation	Letter, rotation	Maths, letter	
Subject 4	Average	83.00	63.35	75.20	60.45	84.25	61.75	AR-LS no seg (seven times)
	Coefficient variation	0.09	0.11	0.11	0.11	0.09	0.12	
	Maximum	95.50	74.50	90.50	72.50	97.00	70.50	AR-BG no seg (three times)
	Best mental task combination	Maths, count	Maths, count	Maths, count	Maths, count	Maths, count	Baseline, maths	

AR-BG = AR with Burg's algorithm, AR-LS = AR with least-squares algorithm, seg = segmentation, no seg = without segmentation.

subject 1, it was maths–rotation, for subject 2, it was letter–count, for subject 3, it was letter–rotation and for subject 4, it was maths–count. It can also be noted that none of the best mental task combinations involved the baseline task. Overall, for three subjects, the best method was AR-LS without segmentation whereas for the other subject, it was AR-BG without segmentation.

LDA classification performance in terms of best mental task combination for subject 1 was maths–rotation where the maximum percentage using AR-LS was 89.00% with the information transfer rate of 0.2731. For subject 2, the best mental task combination was letter–count from AR-BG method (maximum classification of 87%) with information transfer rate of 0.2460. For subject 3, letter–rotation was the best mental task combination given by AR-LS method with maximum classification of 82% with information transfer rate of 0.1480. The performance of subject 4 was the best among the four subjects where it can be noted that the maximum classification percentage was 97.00% for maths–count mental task combination given by AR-LS method with information transfer rate of 0.4158.

As mentioned earlier, the performance of methods without segmentation were better than with segmentation. Thus, for comparing MLP-BP NN and LDA classification performances, we will only take methods without segmentation into consideration. Comparing figures 4 and 5 and tables 2 and 3, it can be seen that in general, LDA performed better that MLP-BP NN.

For subject 1, the classification results for LDA were better than MLP-BP NN not only in terms of average percentages but also maximum percentages. As for the maximum percentage, LDA classification results for AR-BG, AAR and AR-LS without segmentation were 87.00%, 79.50% and 89.00%, respectively compared with MLP-BP NN classification results which were 83.10%, 76.70% and 86.7%, respectively.

For subject 2, it can be seen that the LDA classification results for both AR-BG and AAR without segmentation were better than MLP-BP NN classification results with a difference of 0.59% and 1.31%, respectively. Even though the MLP-BP NN classification results for AR-LS without segmentation were better than LDA classification, the difference in percentage was small, i.e. only 0.32%.

As for subject 3, the performance of LDA was better than MLP-BP NN for all the methods. Performance of AR-BG without segmentation using LDA classification gave 71.65% as compared to 67.46% using MLP-BP NN and in terms of maximum percentage, LDA classification performance was higher than MLP-BP with a difference of 3.70%. As for AAR without segmentation, the average classification percentage using LDA was 58.90% as compared to 56.30% using MLP-BP NN and the maximum percentage using LDA was 5.10% higher than using MLP-BP NN. For AR-LS without segmentation, the average percentage using LDA was 71.40% as compared to 69.90% using MLP-BP NN and the maximum percentage using LDA was 7.10% higher than using MLP-BP NN.

Similarly, for subject 4, the performance of LDA was better than MLP-BP NN where the differences in terms of average percentages were 5.86%, 4.53% and 6.67% for AR-BG, AAR and AR-LS without segmentation, respectively. In terms of maximum percentages, all the methods using LDA gave above 90.00% where the maximum percentages for AR-BG, AAR and AR-LS without segmentation were 95.50%, 90.50% and 97.00%, respectively with a difference of 2.80%, 10.40% and 3.90% as compared to MLP-BP NN classification results.

5. Conclusion

In this paper, we have used MLP-BP NN and LDA to classify mental tasks using features that were extracted from EEG signals. We used six different methods (AR-BG, AR-LS, AAR; with and without segmentation) to extract the features and compared the NN and LDA classification performances for each subject. Our results showed that the AR-LS method without segmentation gave the best classification performances followed by the AR-BG method without segmentation. Therefore, the AAR method (with and without segmentation) is not suitable for this set of mental task EEG signals. Data without segmentation performed better than segmentation. Most likely, this was due to segmentation that results in too few data points for both AR and AAR to model these signals accurately. For future work, we suggest that data segmentation into small segments could be avoided due to its poor performance, increased computational complexity and increased time for training and testing. In addition, the AR method (both with Burg's algorithm and the LS algorithm) performed better and more consistently than AAR. Besides this, the results indicated that different feature extraction methods might be suitable for different subjects such as AR-LS (for subjects 1, 3 and 4) and AR-BG (for subject 2). We draw the conclusion that different subjects have different combination of best mental tasks and proper selection of the best mental tasks is essential for the BCI design. In general, LDA gave better classification performance as compared to MLP-BP NN. LDA is also advantageous because of its lower computational complexity as compared to MLP-BP.

Finally, some important aspects to consider in BCI designs are how the activation patterns associated with specific mental tasks change over time and automaticity. Automaticity in evoking the patterns is necessary to use the method to execute commands, because attentional resources are limited. However, automaticity changes the activation patterns and classification results may be different. However, we were not able to do these analyses for the current study because the data were pre-recorded.

Acknowledgments

The authors would like to acknowledge the assistance of Dr C Anderson of Colorado State University, USA for giving permission to use the EEG data. The assistance of Professor Raveendran of University Malaya, Malaysia in the early stages of the work is greatly appreciated.

Appendix

The steps for implementing Burg's algorithm are as follows:

Step 1. Initial conditions:

$\hat{\pi}_0 = r(0)$.

The forward prediction errors, $\varepsilon_0^f(n) = y(n)$, where $n = 1, 2, 3, \ldots, N-1$.

The backward prediction errors, $\varepsilon_0^b(n) = y(n)$, where $n = 0, 1, 2, \ldots, N-2$.

N is the data size.

Step 2. Reflection coefficients.

For $p = 1, 2, 3, \ldots, P$, where P is the required model order,

$$\hat{\pi}_p = \frac{-2\sum_{n=p}^{N-1} \varepsilon_{p-1}^f(n)\varepsilon_{p-1}^b(n-1)}{\sum_{n=p}^{N-1}\left\{\left[\varepsilon_{p-1}^f(n)\right]^2 + \left[\varepsilon_{p-1}^b(n-1)\right]^2\right\}}$$

$$s_p^2 = \left(1 - \left|\hat{\pi}_p\right|^2\right)s_{p-1}^2.$$

For $p = 1$,

$$\hat{a}_1(1) = \hat{\pi}_1.$$

For $p > 1$,

$$a_p(i) = \begin{cases} a_{p-1}(i) + \hat{\pi}_p a_{p-1}(p-i) \\ \quad \text{for} \quad i = 1, 2, 3, \ldots, p-1 \\ \hat{\pi}_p \qquad \text{for} \quad i = p. \end{cases}$$

Step 3. Prediction errors for next orders:

$$\varepsilon_p^f(n) = \varepsilon_{p-1}^f(n) + \hat{\pi}_p \varepsilon_{p-1}^b(n-1)$$
$$\text{for} \quad n = k+1, k+2, \ldots, N-1$$

$$\varepsilon_p^b(n) = \varepsilon_{p-1}^b(n-1) + \hat{\pi}_p \varepsilon_{p-1}^f(n)$$
$$\text{for} \quad n = k+1, k+2, \ldots, N-2.$$

Note. $k =$ the first n in the previous step to determine errors.

Step 4. Repeat steps 2 and 3 (with p incremented by 1) until the selected model order p is reached.

References

[1] Vaughan T M, Wolpaw J R and Donchin E 1996 EEG based communications: prospects and problems *IEEE Trans. Rehabil. Eng.* **4** 425–30

[2] Wolpaw J R, Birbaumer N, Hectderks W J, McFarland D J, Pecleham P H, Schalk G, Donchin E, Quatrano L A, Robinson C J and Vaughan T M 2000 Brain–computer interface technology: a review of the first international meeting *IEEE Trans. Rehabil. Eng.* **8** 164–73

[3] Anderson C W, Stolz E A and Shamsunder S 1998 Multivariate autoregressive models for classification of spontaneous electroencephalogram during mental tasks *IEEE Trans. Biomed. Eng.* **45** 277–86

[4] Keirn Z A and Aunon J I 1990 A new mode of communication between man and his surroundings *IEEE Trans. Biomed. Eng.* **37** 1209–14

[5] Palaniappan R, Paramesran P, Nishida S and Saiwaki N 2002 A new brain–computer interface design using Fuzzy ARTMAP *IEEE Trans. Neural Syst. Rehabil. Eng.* **10** 140–8

[6] Donchin E, Spencer K M and Wijesinghe R 2000 The mental prosthesis: assessing the speed of a P300-based brain–computer interface *IEEE Trans. Rehabil. Eng.* **8** 174–9

[7] Pfurtscheller G, Neuper C, Guger C, Harkam W, Ramoses H, Schlogl A, Obermaier B and Pregenzer M 2000 Current trends in Graz brain–computer interface (BCI) research *IEEE Trans. Rehabil. Eng.* **8** 216–9

[8] Osaka M 1984 Peak alpha frequency of EEG during a mental task: task difficulty and hemispheric differences *Psychophysiology* **21** 101–5

[9] Logie R H, Zucco G M and Baddeley A D 1990 Interference with visual short-term memory *Acta Psychol. (Amst)* **75** 55–74

[10] Jasper H 1958 The ten twenty electrode system of the international federation *Electroencephalographic Clin. Neurophysiol.* **10** 371–5

[11] Burg J P 1968 A new analysis technique for time series data *NATO Adv. Study Inst. Signal Process. Emphasis Underwater Acoust.*

[12] Fante R L 1988 *Signal Analysis and Estimation* (New York: Wiley)

[13] Shiavi R 1991 *Introduction to Applied Statistical Signal Analysis* (MA: Aksen Associates)

[14] Schloegl A, Neuper C and Pfurtscheller G 1997 Subject specific EEG patterns during motor imaginary *Proc. IEEE EMBS Conf.* pp 1530–2

[15] Schloegl A, Lugger K and Pfurtscheller G 1997 Using adaptive autoregressive parameters for a brain–computer-interface experiment *Proc. IEEE EMBS Conf.* pp 1533–5

[16] Schloegl A 1997 Dynamic spectral analysis based on an autoregressive model with time-varying coefficients *Proc. IEEE–EMBC and CMBEC Conf.* pp 881–2

[17] Widrow B and Stearns S D 1984 *Adaptive Signal Processing* (Englewood Cliffs, NJ: Prentice-Hall)

[18] Wu W-R and Chen P-C 1997 Adaptive AR modeling in white Gaussian noise *IEEE Trans. Signal Process.* **45** 1184–92

[19] Haykin S 1996 *Adaptive Filter Theory* 3rd edn (Englewood Cliffs, NJ: Prentice-Hall)

[20] Neumaier A and Schneider T 2001 Estimation of parameters and eigenmodes of multivariate autoregressive models *ACM Trans. Math. Software* **27** 27–57

[21] Rumelhart D E and McCelland J L 1986 *Parallel Distributed Processing: Exploration in the Microstructure of Cognition* vol 1 (Cambridge, MA: MIT Press)

[22] Schlogl A 2001 A new linear classification method for an EEG-based brain–computer interface *Technical Report* MDBC Austria

[23] Obermaier B, Neuper C, Guger C and Pfurtscheller G 2001 Information transfer rate in a five-classes brain–computer interface *IEEE Trans. Neural Syst. Rehabil. Eng.* **9** 283–8