



## AUTOREGRESSIVE INTEGRATED ADAPTIVE NEURAL NETWORKS CLASSIFIER FOR EEG-P300 SIGNAL CLASSIFICATION

**Demi Soetraprawata\*, Arjon Turnip**

Technical Implementation Unit for Instrumentation Development Division – LIPI  
Kompleks LIPI Gd. 30, Jalan Sangkuriang Bandung, 40135, Indonesia

Received 09 October 2012; received in revised form 22 February 2013; accepted 27 February 2013  
Published online 30 July 2013

### Abstract

Brain Computer Interface has a potency to be applied in mechatronics apparatus and vehicles in the future. Compared to the other techniques, EEG is the most preferred for BCI designs. In this paper, a new adaptive neural network classifier of different mental activities from EEG-based P300 signals is proposed. To overcome the over-training that is caused by noisy and non-stationary data, the EEG signals are filtered and extracted using autoregressive models before passed to the adaptive neural networks classifier. To test the improvement in the EEG classification performance with the proposed method, comparative experiments were conducted using Bayesian Linear Discriminant Analysis. The experiment results show that the all subjects achieve a classification accuracy of 100%.

Keywords: brain computer interface, feature extraction, classification accuracy, autoregressive, adaptive neural networks, EEG-based P300, transfer rate.

### I. INTRODUCTION

A Brain Computer Interface (BCI) is a device that allows users to communicate with the world without utilizing voluntary muscle activity. BCI systems utilize what is known about brain signals to detect the message that a user chose to communicate. These systems rely on the finding that the brain reacts differently to different stimuli, based on the level of attention that is given to the stimulus and the specific processing that is triggered by the stimulus. Therefore brain activity must be monitored with various techniques. Among these techniques, EEG is the most preferred for BCI designs, because of its non-invasive, cost effectiveness, easy implementation, and best temporal resolution [1, 2].

EEGs are usually analyzed in two ways: (i) as free running EEG; (ii) as events related potentials (ERPs) (i.e., P300, slow cortical potentials (SCPs), readiness potentials (RPs), and steady state visual evoked potentials (SSVEPs)) [3]. Around 1964, Chapman and Bragdon as well as Sutton et al., are independently discovered a wave peaking at approximately 300 ms after task-

relevant stimuli [4]. This component is known as the P300. While the P300 is evoked by many types of paradigms, the most common factors that influence it are stimulus frequency and task relevance [5]. The presence, magnitude, topography and time of the response signals are often used as metrics of cognitive function in decision making processes. The P300 has been shown to be fairly stable in locked-in patients, re-appearing even after severe brain's stem injuries. Farwell and Donchin (1988) first showed that this signal may be successfully used in a BCI [6]. Using a broad cognitive signal like the P300 has the benefit of enabling control through a variety of modalities, as the P300 enables discrete control in response to both auditory and visual stimuli. As it is a cognitive component, the P300 has been known to change in response to subject's fatigue [5].

One of the most important task in designing a BCI is in extracting relevant features from the EEG signals, which is naturally noisy and stochastic. In order to avoid the averaging processes and to remove the artifacts, which are computational complexity, poor generalization, and needs a large number of trainings to achieve a desired accuracy and a communication rate,

\* Corresponding Author. Tel: +62-22-2503053  
E-mail: {demi001, arjo001}@lipi.go.id

therefore a new adaptive neural network classifier (ANNC) of different mental activities are proposed. To overcome the over-training caused by noisy and non-stationary data, the features of the EEG-based P300 signals are extracted using autoregressive (AR) method before passed to the proposed classifier algorithm. In order to examine the performance improvements of the proposed classification method, comparative experiments were conducted using Bayesian Linear Discriminant Analysis (BLDA).

The contributions of this paper are as follow:

- a. Enhancements and strengthens the EEG signal according to the small amplitude of the EEG-based P300 which is naturally noisy and stochastic.
- b. Driving the tracking error converges to a small value around zero while the closed-loop stability is guaranteed.
- c. The introduction of the AR method and the application of the proposed classifier improve the classification accuracy and the transfer rate of a BCI even when the subjects are in a fatigue condition.

The structure of the paper is as follows. In Section 2, the EEG data set and pre-processing are described. Feature extraction and classification using AR method and adaptive neural networks, respectively, are explained in Section 3. Results and discussions are presented in Section 4. Conclusions are drawn in Section 5.

## II. DATA SET AND EEG PRE-PROCESSING

In order to examine the performance improvement of the proposed EEG classification method, the EEG based P300 data used in this paper was obtained previously by Hoffmann et al. (2008) who used the following procedure [5]. The data have been recorded according to the 10-20 standards from the 32 electrode configurations. In this study, however, only the signals from the eight electrodes configuration were used. Each recorded signal has a length of 820 samples with a sampling rate of 2,048 Hz. A six-choice signal paradigm was tested using a population of five disable and four able bodied subjects. The subjects were asked to count silently the number of times a prescribed image flashed on a screen. Four seconds after a warning tone, six different images (a television, a telephone, a lamp, a door, a window, and a radio) were flashed in a random order. Each flash of an image lasted for 100 ms, and for the following 300 ms no image was flashed (i.e., the inter-stimulus interval was 400 ms). Each subject completed four recording

sessions. Each of the sessions consisted of six runs with one run for each of the six images. The duration of one run was approximately one minute and the duration of one session, including setup of electrodes and short breaks between runs, was approximately 30 minutes. One session comprised on average 810 trials, and the entire data for one subject was taken from an average of 3,240 trials.

Our goal is to discriminate all possible combinations of the pairs of mental activities from each other using the corresponding EEG signals. The EEG signals are processed in segments (EEG-trials) in which the BCI attempts to recognize the mental activities.

Before the classification and validation are performed, several pre-processing operations including filtering, and down sampling were applied to the data. A 6th order forward-backward Butterworth band pass filter with cut off frequencies of one Hz and 12 Hz was used to filter the data. The EEG was down sampled from 2,048 Hz to 32 Hz by selecting each 64th sample from the band pass-filtered data.

## III. FEATURE EXTRACTION AND CLASSIFICATION

### A. Feature Extraction

In this section, the feature extraction which is focused on the estimation of statistical measurements from the perturbation free EEG-trials delivered by the pre-processing module, is explored. The features computed on a given EEG-trial are grouped into a vector called feature vector that is sent to the pattern recognition module which evaluates the likelihoods that the EEG-trial was produced during the execution of the mental activities.

The autoregressive (AR) method is built on the hypotheses of stationarity, ergodicity, absence of coupling between the univariate components, and existence of a linear prediction model [7]. Let  $\mathbf{Y}$  be an  $N_e$  dimensional multivariable stochastic EEG signal of length  $N_s$ , composed of randomvectors:  $\{\mathbf{Y}(k) = (y_1(k) \dots y_{N_e}(k))^T | k = 0, \dots, N_s - 1\}$  where  $y_1, \dots, y_{N_e}$  are the univariate components of  $\mathbf{Y}$ . The AR model can be generated by a linear prediction model of the form [7]:

$$\mathbf{Y}(k) = -\sum_{i=1}^Q \mathbf{A}(k, i)\mathbf{Y}(k - i) + e_1(k), \quad (1)$$

where  $Q$  is the model order, the  $\mathbf{A}(k, i)$  are  $N_e \times N_e$  matrices ( $N_e N_e$  denoting the number of electrodes and  $N_s$  denotes the number of temporal samples per EEG channel), and  $e_1(k)$  is

the prediction error with a zero mean random vector. Since the coupling between the channels is ignored, equation (1) can be split into linear prediction models corresponding to each univariate component. Thus, the  $n$ -th univariate component of  $Y$  can be written in the form:

$$y_n(k) = -\sum_{i=1}^{Q_n} a_n(k, i)y_n(k-i) + e_n(k), \quad (2)$$

where the  $a_n(k, i)$  are the AR coefficients and  $Q_n$  is the AR order corresponding to  $y_n$ , and  $e_n$  is the  $n$ -th prediction error process. The indexes  $n$  and  $k$  are used to reference the electrode and time index, respectively. Furthermore, as stationarity and ergodicity are assumed, then the AR model for the  $n$ -th channel becomes:

$$y_n(k) = -\sum_{i=1}^{Q_n} a_n(i)y_n(k-i) + e_n(k). \quad (3)$$

The coefficients  $a_n(1), \dots, a_n(Q_n)$  can be determined by minimizing the averaged squared prediction error:

$$\begin{aligned} \varepsilon(Q_n) &= \frac{1}{N_s} \sum_{k=0}^{N_s-1} e_n^2(k) \\ &= \frac{1}{N_s} \sum_{k=0}^{N_s-1} (y_n(k) + \sum_{i=1}^{Q_n} a_n(i)y_n(k-i))^2 \end{aligned} \quad (4)$$

In this relation, the samples prior to  $y_n(0)$  are assumed to be zero.

## B. Adaptive Neural Networks Classifier

Artificial neural networks have been proposed in the fields of information and neural sciences following research into the mechanisms and structures of the brain. This has led to the development of new computational models for solving complex problems such as pattern recognition, rapid information processing, learning and adaptation, classification, identification and modelling, speech, vision and control systems [8-14].

In this paper, we are only concerned with the adaptive classifier problem of EEG-based P300 represented by nonlinear discrete-time systems which can be transformed in state space description [15] as follow:

$$\begin{aligned} x_1(k+1) &= x_2(k), \\ x_2(k+1) &= x_3(k), \\ &\vdots \\ x_n(k+1) &= f(x(k)) + g(x(k))u(k), \\ y_k &= x_1(k), \end{aligned} \quad (5)$$

where  $x(k) = [x_1(k), x_2(k), \dots, x_n(k)]^T \in R^n$ ,  $u(k) \in R$ ,  $y(k) \in R$  are the state variables, system input and output, respectively;

$f(x(k))$  and  $g(x(k))$  are unknown which may not be linearly parameterized. The classifier system attempts to make the plant output  $y_d(k)$  match the target output asymptotically, so that  $\lim_{t \rightarrow \infty} \|y_d(k) - y_k\| \leq \varepsilon$  for some specified constant  $\varepsilon \geq 0$ . If  $f(x(k))$  and  $g(x(k))$  are known, the following classifier can be used, and the system would exactly track the target output  $y_d(k)$ .

$$u(k) = g^{-1}(x(k))(y_d(k) - f(x(k))) \quad (6)$$

Since  $f(x(k))$  and  $g(x(k))$  are unknown, neural networks can be used to learn to approximate these functions and generate suitable classifiers. Although the function  $g(x(k))$  is not known, it can be assumed that its sign is known along system trajectories and that there exist two constants  $g_0, g_1 > 0$  such that  $g_0 \leq |g(x(k))| \leq g_1$ ,  $\forall x \in \Omega \in R^n$  with compact subset  $\Omega$  containing the origin. This assumption implies that the function  $g(x(k))$  is strictly either positive or negative. From this point forward therefore, without losing generality, we shall assume  $g(x(k)) > 0$ .

Neural networks are general modelling tools that can approximate any continuous or discrete nonlinear function to any desired accuracy over a compact set [9-11, 16, 17]. In this work, a new adaptive neural network classifier is developed for nonlinear system (5) using high order neural networks. Therefore the mental activities according to the given stimuli could be extracted and classified with high accuracy. It should be noted that although the new states  $x_2, x_3, \dots, x_n$  are not available in practice, we can predict them as will be detailed in the following discussion. Let  $x_d = [y_d(k), y_d(k+1), \dots, y_d(k+n-1)]^T$  the target system states. Define error  $e(k) = x(k) - x_d(k)$ . The equation of  $e(k)$  can be written as:

$$\begin{aligned} e_1(k+1) &= e_2(k), \\ e_2(k+1) &= e_3(k), \\ &\vdots \\ e_n(k+1) &= f(x(k)) + g(x(k))u(k) \\ &\quad - y_d(k+n). \end{aligned} \quad (7)$$

In order to develop the output feedback classifier clearly, define the following new variables  $\bar{y}(k) = [y_{k-n+1}, \dots, y_{k-1}, y_k]^T$ ,  $\bar{u}_{k-1}(k) = [u_{k-1}, \dots, u_{k-n+1}]^T$ , and  $\bar{z}(k) = [\bar{y}^T(k), \bar{u}_{k-1}^T(k)]^T \in \Omega_{\bar{z}} \subset R^{2n-1}$ . According to the definition of the new states,  $\bar{y}(k) =$

$[x_1(k-n+1), \dots, x_1(k-1), x_1(k)]^T$  and from Eq. (5), the following equation is obtained.

$$\begin{aligned} y_{k+1} &= x_2(k) = x_3(k-1) = \dots \\ &= x_n(k-n+2) \\ &= f(\bar{y}(k)) + g(\bar{y}(k))u_{k-n+1} \\ &= \phi_2(\bar{z}(k)), \end{aligned} \quad (8)$$

which  $x_2(k)$  is a function of  $y(k)$  and  $u_{k-n+1}$ . From (5), similarly the following equation is obtained.

$$\begin{aligned} y_{k+2} &= f(\bar{y}(k+1)) + g(\bar{y}(k+1))u_{k-n+2} \\ &= \phi_3(\bar{z}(k)), \\ &\vdots \\ y_{k+n-1} &= f(\bar{y}(k+n-2)) + g(\bar{y}(k+n-2))u_{k-1} \\ &= \phi_n(\bar{z}(k)) \end{aligned} \quad (9)$$

It proves that  $x_n(k)$  is a function of  $\bar{z}(k)$ . Substituting the predicted states into the last equation in (5), we obtain:

$$\begin{aligned} y_{k+n} &= x_n(k+1) \\ &= f(\bar{z}(k)) + g(\bar{z}(k))u(k) \end{aligned} \quad (10)$$

where

$$f(\bar{z}(k)) = f([x_1(k), \phi_2(\bar{z}(k)), \dots, \phi_n(\bar{z}(k))]^T), \quad (11)$$

$$g(\bar{z}(k)) = g([x_1(k), \phi_2(\bar{z}(k)), \dots, \phi_n(\bar{z}(k))]^T). \quad (12)$$

Define a tracking error as  $e_y(k) = y_k - y_d(k)$ . The tracking error dynamics are given by:

$$\begin{aligned} e_k(k+n) &= -y_d(k+n) + f(\bar{z}(k)) \\ &\quad + g(\bar{z}(k))u(k). \end{aligned} \quad (13)$$

Supposing that the nonlinear functions  $f(\bar{z}(k))$  and  $g(\bar{z}(k))$  are known exactly, then a desired classifier, such that the output  $y_k$  follows the target trajectory  $y_d(k)$ , is written as follow:

$$u^*(k) = -\frac{1}{g(\bar{z}(k))} (f(\bar{z}(k))) - y_d(k+n). \quad (14)$$

Substituting the desired classifier equation (14) into error dynamics equation (13), i.e.,  $u(k) = u^*(k)$ , then the error dynamics goes to zero is obtained. This means that after  $n$  steps, we have  $e_y(k) = 0$ . Therefore,  $u^*(k)$  is a  $n$ -step dead-beat classifier. Moreover, the desired classifier  $u^*(k)$  can be expressed as:

$$u^*(k) = \bar{u}^*(\bar{z}(k)), \quad (15)$$

where  $\bar{z}(k) = [\bar{z}^T(k), y_d(k+n)]^T \in \Omega_{\bar{z}} \subset R^{2n}$  with component set  $\Omega_{\bar{z}}$  is defined as

$$\Omega_{\bar{z}} = \{(\bar{y}(k), u_{k-1}(k), y_d) | \bar{u}_{k-1}(k) \in \Omega_{u,y,k,y_d} \in \Omega_y\}. \quad (16)$$

When the nonlinear functions  $f(\bar{z}(k))$  and  $g(\bar{z}(k))$  are unknown, the nonlinearity  $u^*(k)$  is not available. In the following, high order neural networks (HONNs) is introduced to construct the unknown nonlinear functions  $f(\bar{z}(k))$  and  $g(\bar{z}(k))$  for approximating the desired feedback signal  $u^*(k)$ . Under certain conditions, it has been proven that neural networks has function approximation abilities and has been frequently used as function approximators, which include linearly and nonlinearly parameterized networks. Consider the following HONNs [16, 18]:

$$\varphi(W, z) = W^T S(z), W \text{ and } S(z) \in R^l, \quad (17)$$

$$S(z) = [s_1(z), s_2(z), \dots, s_l(z)]^T, \quad (18)$$

$$s_i(z) = \prod_{j \in l_i} [s(z_j)]^{d_j(i)}, i = 1, 2, \dots, l, \quad (19)$$

where  $z = [z_1, z_2, \dots, z_m]^T \in \Omega_z \subset R^m$ ; the positive integer  $l$  indicates the node number of neural network;  $d_j(i)$  indicates the non-negative integers;  $W$  is an adjustable synoptic weight vectors; and  $s(z_j)$  is chosen as a hyperbolic tangent function.

$$s(z_j) = \frac{e^{z_j} - e^{-z_j}}{e^{z_j} + e^{-z_j}} \quad (20)$$

According to Girosi and Poggio (1989) [19], there exist ideal weight  $W^*$  such that the function  $\varphi(z)$  can be approximated by an ideal neural network on a compact set  $\Omega_z \subset R^m$ :

$$\varphi(z) = W^{*T} S(z) + \varepsilon_z, \quad (21)$$

where  $\varepsilon_z$  is called the neural network approximation error. It is representing the minimum possible deviation of the ideal approximator  $W^{*T} S(z)$  from the unknown function  $\varphi(z)$ . In general, the ideal neural network weight  $W^*$  is not known and needs to be estimated. In this paper, there exist an integer  $l^*$  and an ideal constant weight vector  $W^*$ , such that for all  $l \geq l^*$ ,

$$\bar{u}^*(\bar{z}(k)) = W^{*T} S(\bar{z}(k)) + \varepsilon_{\bar{z}}, \forall \bar{z} \in \Omega_{\bar{z}}, \quad (22)$$

where  $\varepsilon_{\bar{z}}$  is the neural network estimation error satisfying  $|\varepsilon_{\bar{z}}| < \varepsilon_0$ . Based on Lyapunov technique, it has been proven in Ge et al., 2003 [17] that the adaptive classifier law and the updating law can be chosen as:

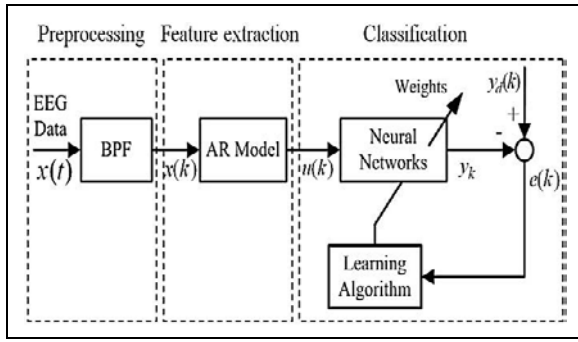


Figure 1. Structure of feature extraction and classification algorithms

$$u(k) = \widehat{W}^T S(\bar{z}(k)), \quad (23)$$

$$\widehat{W}(k+1) = \widehat{W}(k_1) + \Gamma[S(\bar{z}(k_1))(y_{k+1} - y_d(k+1)) + \rho \widehat{W}(k)], \quad (24)$$

where  $k_1 = k - n + 1$ , diagonal gain matrix  $\Gamma > 0$ , and  $\rho > 0$ . Therefore, the tracking error converges to a small neighborhood of zero by increasing the approximation accuracy of the neural networks and the closed-loop stability is guaranteed.

Figure 1 shows the structure of the pre-processing, feature extraction, and the ANNC algorithm. In Figure 1,  $x(t)$  indicates a non-pre processed (raw) EEG signal;  $x(k)$  indicates a filtered signal;  $y(k)$  indicates an extracted signals in which the artifact was removed;  $y_d(k)$  and  $y_k$  indicate a target and classified signal, respectively.

#### IV. RESULT AND DISCUSSION

In this study, a new method using adaptive neural networks for the classification of the EEG-based P300 signals is proposed. This method is supported by the AR model to extract the features

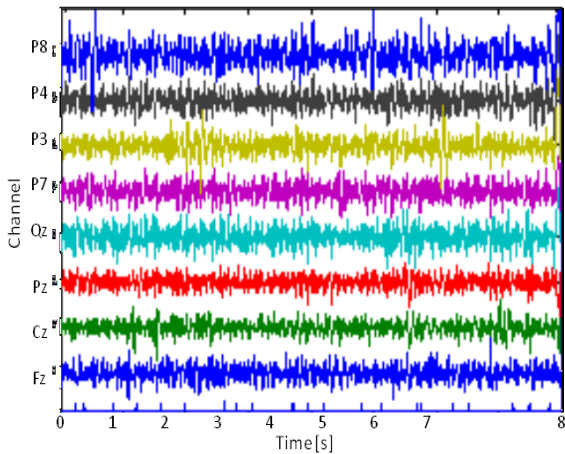


Figure 2. Pre-processed EEG signals using the Butterworth band pass filter

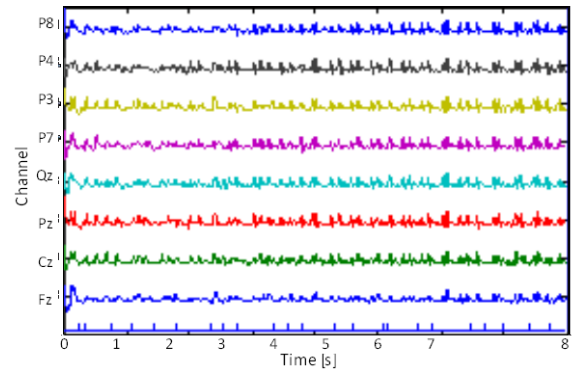


Figure 3. Extracted EEG signals using the AR model

and reduce the artifact that is contained within the EEG signals. The methods mentioned above were applied to the training of eight subjects who participated in four training sessions with six runs for each session. Figure 2 and 3 are the pre-processed EEG-based P300 signals using Butterworth band pass filter and after applying the AR method as feature extractor and artifacts remover, respectively. Although we can notice some improvement, at Figure 3, it is still difficult to classify the signals with respect to the P300 component. Thus, a new adaptive neural networks classifier is proposed.

The tracking error graph with and without applying the AR model approach is shown in Figure 4. The curves show that a level of accuracy is attained after about 250 iterations by applying the AR model approach. On the other hand, the same level of accuracy is attained after 1,800 iterations if the proposed feature extraction method is not applied. It means the introduction of the AR method is relevant to accelerate the training processes.

The tracking error converges to a small value around zero while the closed-loop stability is guaranteed. Furthermore, the tracking error with the AR model was converging in faster. According to [17], the data set for subject five were not included in the simulation since the subject misunderstood the instructions given before the experiment. Comparative plots of the

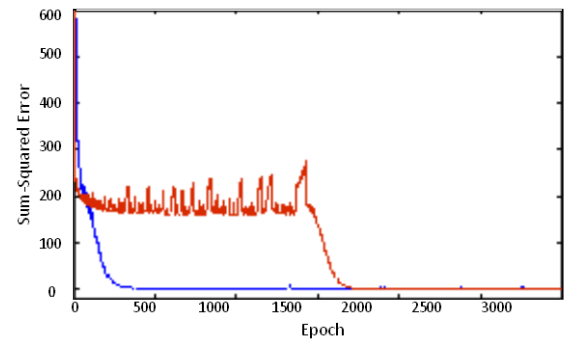


Figure 4. Network's performance according to mean squares errors

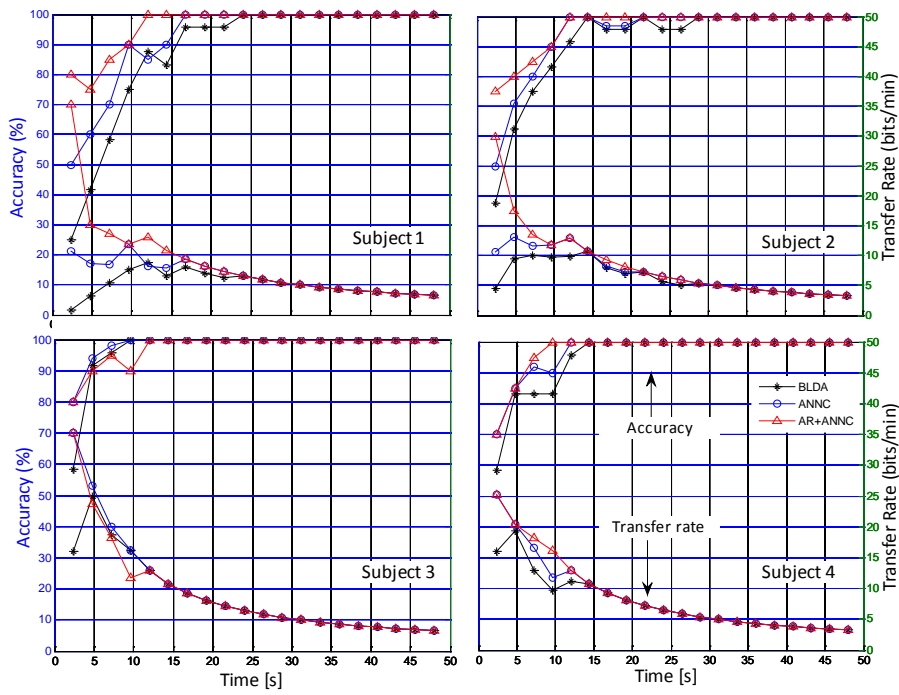


Figure 5. Comparison of classification accuracy and transfer rate plots obtained with BLDA, ANNC, and the combination of the AR model and ANNC for disable-bodied subjects

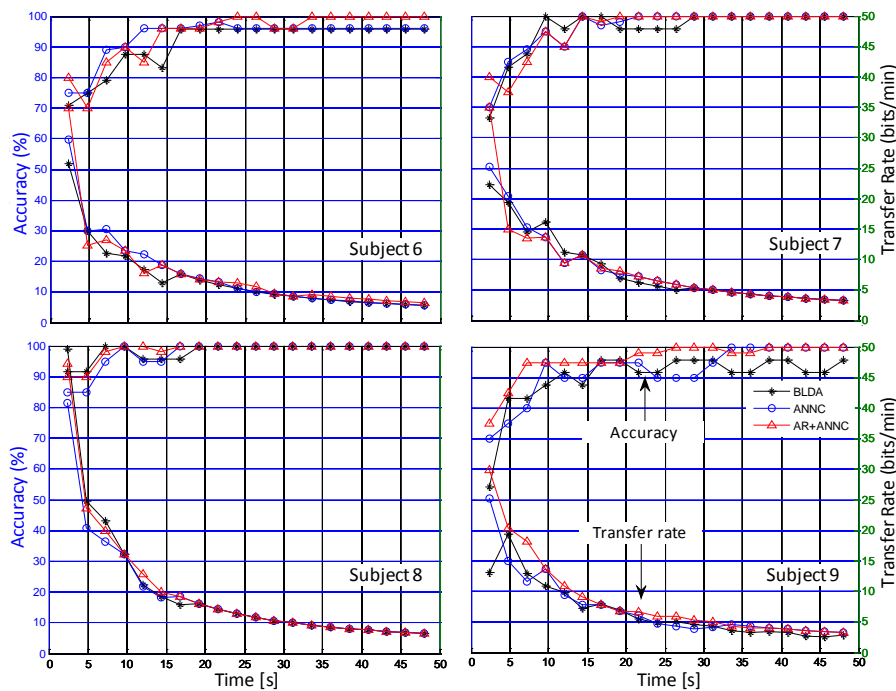


Figure 6. Comparison of classification accuracy and transfer rate plots obtained with BLDA, ANNC, and the combination of the AR model and ANNC for able-bodied subjects

classification accuracies and transfer rates (obtained with BLDA, ANNC, and the combination of the AR model and ANNC methods and averaged over four sessions) for the disable-bodied subjects (subjects 1 – 4) and able-bodied subjects (subjects 6 – 9) are shown in Figure 5 and 6, respectively.

All of the subjects (with the combination of the AR model and ANNC methods), except for subjects 6 and 9, achieved an average classification accuracy of 100% after five blocks of stimulus presentations were averaged (i.e., 14 second). However, subjects 6 and 9, compared with BLDA, still achieved an average classification accuracy of 100% after nine and ten

blocks of stimulus presentations were averaged, respectively. This results give a significant improvement compared with the results presented in [17] in which subject 6 and subject 9 were not achieved an average classification accuracy of 100%. It means the introduction of the AR method and the application of the proposed classifier enables the BCI to extract and classify the information in terms of the classification accuracy from a fatigue subject. Thus, fatigue as one of the reasons for the poorer performance of subject 9, as mentioned in Hoffmann *et al.* [5], can be solved using the proposed method.

## V. CONCLUSIONS

The results presented in this study show that corresponding to the classification accuracy, the data indicates that a P300-based BCI system can communicate at the rate of 31.2 bits/min and 36.7 bits/min for the disabled-bodied and able-bodied subjects, respectively. The classification and transfer rate accuracies obtained based ANNC with the AR models approach are found to be far superior in comparison with the BLDA approach and therefore better suited for BCI applications.

## ACKNOWLEDGEMENT

This work is a part of thematic project research funded by UPT BPI LIPI (DIPA NO. 3425.01.011) budgeting year of 2012. The authors would like to thank the Deputy Chairmant for Scientific Services Dr. Fatimah Zulfah S. Padmadinata for supporting to publish this paper.

## REFERENCE

- [1] B. E. Hillner, *et al.*, "Impact of positron emission tomography/computed tomography and positron emission tomography (PET) alone on expected management of patients with cancer: initial results from the National Oncologic PET Registry," *J Clin Oncol*, vol. 26, pp. 2155-61, May 1 2008.
- [2] F. Jouret, *et al.*, "Single photon emission-computed tomography (SPECT) for functional investigation of the proximal tubule in conscious mice," *Am J Physiol Renal Physiol*, vol. 298, pp. F454-60, Feb 2010.
- [3] E. Niedermeyer and F. L. D. Silva, *Electroencephalography*, 4th ed. Baltimore: Lippincott, Williams & Wilkins, 1999.
- [4] R. M. Chapman and H. R. Bragdon, "Evoked Responses to Numerical and Non-Numerical Visual Stimuli while Problem Solving," *Nature*, vol. 203, pp. 1155-1157, 1964.
- [5] U. Hoffmann, *et al.*, "An efficient P300-based brain-computer interface for disabled subjects," *Journal of Neuroscience Methods*, vol. 167, pp. 115-125, 2008.
- [6] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr Clin Neurophysiol*, vol. 70, pp. 510-23, Dec 1988.
- [7] W. D. Penny, *et al.*, "EEG-based communication: a pattern recognition approach," *IEEE Trans Rehabil Eng*, vol. 8, pp. 214-5, Jun 2000.
- [8] H. Chen and L. Li, "Semisupervised multicategory classification with imperfect model," *IEEE Trans Neural Netw*, vol. 20, pp. 1594-603, Oct 2009.
- [9] S. S. Ge, *et al.*, "Nonlinear adaptive control using neural networks and its application to CSTR systems," *Journal of Process Control*, vol. 9, pp. 313-323, 1999.
- [10] S. S. Ge, *et al.*, "Adaptive MNN control for a class of non-affine NARMAX systems with disturbances," *Systems & Control Letters*, vol. 53, pp. 1-12, 2004.
- [11] S.-J. Liu, *et al.*, "Adaptive output-feedback control for a class of uncertain stochastic non-linear systems with time delays," *International Journal of Control*, vol. 81, pp. 1210-1220, August 1 2008.
- [12] S. Jaiyen, *et al.*, "A very fast neural learning for classification using only new incoming datum," *IEEE Trans Neural Netw*, vol. 21, pp. 381-92, Mar 2010.
- [13] S. Ozawa, *et al.*, "A multitask learning model for online pattern recognition," *IEEE Trans Neural Netw*, vol. 20, pp. 430-45, Mar 2009.
- [14] Y. Washizawa, "Feature extraction using constrained approximation and suppression," *IEEE Trans Neural Netw*, vol. 21, pp. 201-10, Feb 2010.
- [15] A. Isidori, *Nonlinear control systems*, 2nd ed. Berlin: Springer-Verlag, 1989.
- [16] S. S. Ge, *et al.*, *Stable Adaptive Neural Network Control*, 1st ed. Norwell: Kluwer Academic, 2001.
- [17] S. S. Ge, *et al.*, "Adaptive NN control for a class of strict-feedback discrete-time

- nonlinear systems," *Automatica*, vol. 39, pp. 807-819, 2003.
- [18] E. B. Kosmatopoulos, *et al.*, "High-order neural network structures for identification of dynamical systems," *IEEE Trans Neural Netw*, vol. 6, pp. 422-31, March 1995.
- [19] F. Girosi and T. Poggio, "Networks and the best approximation property," *Biological Cybernetics*, vol. 63, pp. 169-176, July 1 1990.



