# EEG-BASED COMMUNICATION: A TIME SERIES PREDICTION APPROACH

*Damien Coyle, Girijesh Prasad and Martin McGinnity*

Intelligent Systems Engineering Laboratory (ISEL), School of Computing and Intelligent Systems, Faculty of Engineering, Magee Campus, University of Ulster, Northland Road, Derry, Northern Ireland, BT48 7JL, UK.
dh.coyle@ulster.ac.uk

## ABSTRACT

Recently, a new technology known as the brain-computer interface (BCI) has received a substantial amount of interest among various research groups worldwide. The human brain can be represented by self-organising and complex biochemical states. Due to continuous neuronal activity in the brain, chaotic electric potential waves are observed in Electroencephalogram (EEG) recordings of the brain. A BCI involves extracting information from the highly complex EEG. This is achieved by obtaining the dominant discriminating features from different EEG signals recorded during specific thought processes. A class of features is usually obtained from each thought process and subsequently a classifier is trained to learn which feature belongs to which class. This ultimately leads to a system that can determine which thoughts belong to which set of EEG signals. This work outlines a novel method which utilises cybernetic intelligence in the form of Neural Networks (NN). Three NNs are coalesced to perform simplified simulations of a number of the characteristic and complex processes that are sub-consciously performed in the human brain. These include prediction, feature extraction and classification. These processes are combined in this system to produce a pattern recognition system which distinguishes between similar complex patterns from a noisy environment with classification accuracy which compares satisfactorily to current reported results. The classification accuracy is achieved by increasing the separability between the features extracted from two EEG signals recorded from subjects during imagination of left and right arm movement.

## 1. INTRODUCTION

Nearly two million people in the United States [1] are affected from neuromuscular disorders. A conservative estimate of the overall prevalence is that 1 in 3500 of the worlds population may be expected to have a disabling inherited neuromuscular disorder presenting in childhood or later life (1991) [2]. In many cases those affected may have no control over any of their muscles. Brain-Computer Interface (BCI) technology may help improve the standard of living for these people by offering an alternative communication channel which does not depend on the peripheral nerves or muscles [3]. A BCI replaces nerves and muscles and the movements they produce with electrophysiological signals in conjunction with the hardware and software that translate those signals into actions [1].

This paper proposes a novel approach for human computer interaction utilising computational models which are based on the architecture of the human brain and inspired by the way neuronal information processing is carried out. A system is developed that can learn to recognise features of complex signals generated by the human brain. Researchers utilizing systems which exhibit cybernetic intelligence hope to obtain a better understanding of the dynamics of the EEG signal in relation to brain function. The EEG signal can provide information about the state of complex neuronal networks often considered as nonlinear dynamical systems. Greater understanding of the brain has helped to develop new technologies that reproduce brain-like operations; an example being artificial neural networks, as well as progressing towards the development of more biologically inspired intelligent systems in the form of spiking NNs.

The underlying generator of EEG is neuronal activity of large numbers of neurons communicating and interacting through low voltage electrical signaling. The highly non-stationary EEG signal is produced by the temporal and spatial summation of electrical currents that arise from pre and postsynaptic potentials [4] of parallel and synchronously active neurons. The normal geometry of synaptic distributions over the pyramidal cells makes it impossible to know whether an EEG event at the scalp (i.e. frequency or amplitude change) is due to an inhibitory or excitatory postsynaptic potential and which specific neuron is firing. In general the EEG can be considered as an information-carrying neural signal that is not used internally as a neural code [5].

A BCI normally involves a feature extraction procedure (FEP), a translation algorithm (TA), and a feedback mechanism. The FEP extracts the dominant discriminating features from the EEG activity during two or more specific thought processes. The TA translates these features into specific control signals which can be used for cursor control, menu selection, letter typing or for control of prosthetic devices. The optional feedback mechanism helps the user become more efficient in controlling his/her EEG for specific tasks [6][7][8].

Various approaches for the FEP and TA have been reported. These include using classical techniques such as Auto-Regression (AR) and Linear Discriminant Analysis [9], Neural Networks [10] and Fuzzy ARTMAP [11]. Many statistical methods and artificial intelligent processing techniques have been experimented with, each with varying degrees of accuracy and complexity. Depending on the subject utilizing the system, classification accuracy between 70-95% for two and three class (mental tasks) BCIs with information transfer rates between 10-25 bits/min have been reported. Features from the time domain, frequency domain or both can be utilized. NNs, regarded as the universal approximator, provide a well established framework for pattern recognition problems [10] and are very adaptable to solving nonlinearly separable problems. Because the EEG signals result from the summation of the action potentials of millions of neurons the EEG signals have deterministic features that are intertwined with noise to produce a wide variety of behaviours, including chaotic behaviour. Therefore, the choice of translation algorithm and feature extraction procedure is crucial and must be robust enough to distinguish specific EEG patterns from a multitude of patterns as well as being adaptable to different users. The BCI operation depends on the interaction of two adaptive controllers: the user's brain, which produces the signals measured by the BCI, and the BCI itself, which translates these signals into specific commands [1].

The proposed BCI approach using NN-based time series prediction allows a highly adaptable system which does not require subject specific frequency analysis and focuses on features extracted from the time domain only. The authors believe that a time series prediction approach to extract features using NNs has not been reported in BCI research yet. EEG is recorded from two electrodes (channels) attached to the scalp. The EEG data is uniquely configured so that data from both channels can be processed by a single network. Two NNs are used to predict a value of EEG signals at point *t,* in the future, using past values. The system is configured in three stages. The first stage involves training two NNs separately to perform one-step-ahead prediction using each successive four previous values of each time series. These NNs are labelled 'left' and 'right' corresponding to the type of EEG data on which they are trained (i.e. either left or right arm movement EEG data). The next stage involves inputting each type of training data into both NNs. Each 'prediction'-NN provides a one-step-ahead prediction for the data that is input. The MSE of the prediction measured over a segment of the prediction provides a feature for the signal that is input to the NN. Because each NN is trained to predict the values of two channels of EEG data there can be two or more features extracted from each NN. Also, each type of data is fed into both prediction-NNs therefore from the FEP there can be four or more features extracted for each type of data (i.e. two from each channel prediction of the left NN and two from the right NN). Measuring the MSE of the prediction over a number of segments from each channel provides the option of extracting more than one feature for each channel. Combining the features obtained from left data input to both NNs forms a left feature vector and similarly, for right data, a right feature vector is obtained. Thus, performing this procedure for a number of training signals a class of features is obtained for left EEG data and class of features is obtained for right EEG data. The third stage is classification. This involves training a linear or nonlinear classifier on the features obtained. When applying the system the unknown data is fed into both prediction-NNs and a feature vector is extracted. Subsequently these features are fed into the classifier and a class prediction is made. The following section provides a rationale for the development of this particular process.

Section 3 describes the data acquisition and section 4 outlines the data configuration. Section 5 explains the choice of architecture and training parameters for the prediction-NNs. Sections 6 and 7 detail the FEP and classification procedure. Section 8 is a discussion of the results and section 9 concludes the paper.

## 2. SIGNAL ANALYSIS (ERS/ERD)

A frequency domain analysis of EEG recorded from electrodes C3 and C4 can provide insight as to why there are discriminative properties between the EEG of the two thought processes. The frequency components are obtained using a Fast Fourier Transform (FFT) and a temporal resolution is employed by sliding a window along the data sequence with a certain overlap on each window. Usually during a left arm movement imagery there is a desynchronization of the $\mu$ (8-12Hz) rhythm observed from the spectral plot of the C4 signal and synchronization on the C3 spectral plot. These changes are referred to as Event Related Synchronisation/ Desynchronization (ERS/ERD). Usually an ERD is observed over the contralateral (opposite) hemisphere and an ERS is observed on the ipsilateral (same side) hemisphere. Many BCIs use the power of the frequency spectrum in subject specific frequency bands as features for maximal discrimination [10][11].

There are clearly discriminative components between the two thought images, in the frequency domain although the synchronization and desynchronization of the signals does have implications in the time domain. During testing of the prediction accuracy of the right network it was observed that in most cases the MSE of the prediction on the C4 output of the network was less than that of the C3 output prediction. It is postulated that this is due to the fact that the right network is trained to predict a synchronized signal for the C4 output (ipsilateral) and a desynchronized signal on the C3 output (contralateral). The opposite occurs for the left network. Considering these observations, it appears that both NNs should be different and can produce significantly different prediction accuracy when input with the opposite data. This also led to the hypothesis that training each network to predict the next value of both channels simultaneously would make each network unique and be more specific to each type of movement imagination data. Therefore, each network would predict the data type which it was trained on more accurately than the opposite data.

An observation of the time evolution of the frequency components (ERD and ERS) from the beginning of each signal led to the hypothesis that an increase in the classification accuracy may be possible if the error was measured using a segment of the prediction error instead of using the total MSE of the entire prediction. This deduced that if only the MSE of the prediction accuracy of a small segment of data or the MSE of a number of small segments of data was used then BCI classification would not only be more accurate but a faster BCI could be established. Classification accuracy and speed are crucial requirements of the BCI.

## 3. DATA AQUISITION

The EEG data used to demonstrate this approach was recorded by a research group at the Institute for Biomedical Engineering, University of Technology Graz, Austria [9][10][11]. The Graz research group has developed a BCI which uses μ (8-12Hz) and central β (18-26Hz) EEG rhythms recorded over the sensorimotor cortex. Several factors have suggested that μ and/or β rhythms could be good signal features for EEG based communication. These signals are associated with those cortical areas most directly connected to the brain's normal motor output channels [1].

The data is recorded from two different subjects in an explicit experimental paradigm where each subject is instructed to think about left and right arm movement in accordance to a cue stimulus displayed on a computer monitor. In each recording session a number of EEG patterns relating to the imagined right or left arm movement thought process are produced by a subject over a number of trials. All signals are amplified and digitized for storage and manipulation on computer. The EEG in this experiment is recorded using electrodes C3 and C4 which are positioned over the left and right hand sides of the motor cortex (movement related area of the brain), respectively, according to the American Electroencephalographic society standard (10/20 system) electrode positioning nomenclature [4]. A detailed description of the experimental setup for recording these EEG signals is available in [9].

## 4. DATA CONFIGURATION

In this study the recorded EEG data is structured so that the values of every successive four time points in the time-series from each channel are used to predict the value of the next time point in the series from that channel. Each training data input sequence contains four values from the data recorded from C3 and four from C4. This forms an 8 element input vector for each training sample. The training data output contains every subsequent value from each of the training input data vectors of C3 and C4. Thus each training output vector corresponding to each training input vector is a 2 element vector.

The data is sampled at 128 Hz and each trial consists of approximately 5 seconds of relevant data. The data was recorded from two subjects (S1 and S2) during two different sessions. For the subject S1 a total of 280 trials were executed, 140 of which were of right arm movement imagery and 140 left arm movement imagery. For the subject S2 there were 320 trials, 160 of which were of right arm movement imagery and 160 left. One session provides a substantial amount of data for training and testing, and a validation set if required. Each trial consists of 640 samples ($5/128^{-1} = 640$) therefore, the training input/output data for each trial consists of 636 samples (data points 636 → 639 are used to predict 640).

## 5. PREDICTION - NNs ARCHITECTURE AND TRAINING PROCEDURE

Two feed-forward multilayered perceptron NNs are used for prediction. One NN is trained on the left EEG data and the other on the right EEG data. For training purposes 50% of the trials are concatenated to form the training data set. By using separate NNs for each type of data, each trained NN has certain uniqueness, in that it is more apposite to each type of time-series data. Combining the data from both electrode channels (C3 and C4) also enhances each NN's expediency to the type of data on which it is trained.

The architecture of each set of NNs was adjusted each time the NNs were trained. It was observed that the size of the NN (i.e. no. of hidden layers and no. of neurons in each layer) did not have a huge effect on the overall prediction error. For this reason the prediction was not used as criterion for the choice of NN architecture but instead the overall classification accuracy of the complete system was used as the criterion for selecting the best NN architecture. It was found that NNs with different numbers of hidden layers and the different numbers of neurons in each layer resulted in significant changes in the classification accuracy. The NN weights were updated using Levenberg-Marquardt method. This method allowed fast convergence to a global minimum error but has the disadvantage of being computationally expensive.

To determine which sets of NNs provide features which allow the best classification accuracy a comparative analysis was performed for the NNs trained on data recorded from subject S1. Many types of NN architectures were experimented with, of which three are presented in this work. Each set of NNs were trained a number of times, each time each set of NNs were trained for 50 epochs. Usually the NNs would converge and reach a global minimum error in less than 20 epochs. It was thought that if each NN was allowed to reach a global minimum error and then continue training for a certain number of epochs then each NN would become over-generalized to the type of data on which it was trained and that this would be advantageous to this method (i.e. that each NN would be more specific to the data on which it was trained). Over generalization in most applications is not regarded as advantageous because the networks don't generalize well to new unseen data. There is much similarity in the magnitude of change in the signal from point to point in certain data sequences for all channels and for both types of data. Therefore, a substantial amount of training data can be representative of all the data as a whole. It was postulated that if each NN was over trained on the data on which it was trained then each NN would still generalize well to the testing data and any unseen data. Three sets of NNs were trained, one with one hidden layer containing 10 neurons, one with two hidden layers with 6 and 8 neurons in the first and second hidden layers, respectively, and another with 3 hidden layers with 8, 10 and 6 neurons in the first, second and third hidden layers, respectively.

Three sets of NNs with the same architectures as those just previously described were also trained, this time using a validation data set to stop the training early. After each training epoch the NN is tested using the validation data. If the NN does not reduce the validation data prediction error further for a specified number of epochs during training, training is stopped. This prevents the NNs becoming over-generalized. This meant that if there were significant differences in the test data which were similar to those in the validation data, the NNs trained using validation would generalize well to the test data, thus produce a smaller prediction error and perhaps provide features which are more specific to the signals being predicted.

## 6. FEATURE EXTRACTION PROCEDURE

After each NN has been trained the training data is configured so that it can be input to both the NNs again. This time the data is input to each NN trial by trial. After the data for a trial is fed into both NNs each predicted output for C3 is subtracted from the actual C3 data. The MSE of the difference between the actual and the predicted output, for a segment of the trial, is calculated. This is a measure of the prediction accuracy. This procedure is repeated for the C4 prediction also. The data from each training trial is input to both the left

$$ll3\_e = \frac{1}{M}\sum_{t=1}^{M}(c3\_l(t) - c3\_ll(t))^2 \qquad (1)$$

$$ll4\_e = \frac{1}{M}\sum_{t=1}^{M}(c4\_l(t) - c4\_ll(t))^2 \qquad (2)$$

$$lr3\_e = \frac{1}{M}\sum_{t=1}^{M}(c3\_l(t) - c3\_lr(t))^2 \qquad (3)$$

$$lr4\_e = \frac{1}{M}\sum_{t=1}^{M}(c4\_l(t) - c4\_lr(t))^2 \qquad (4)$$

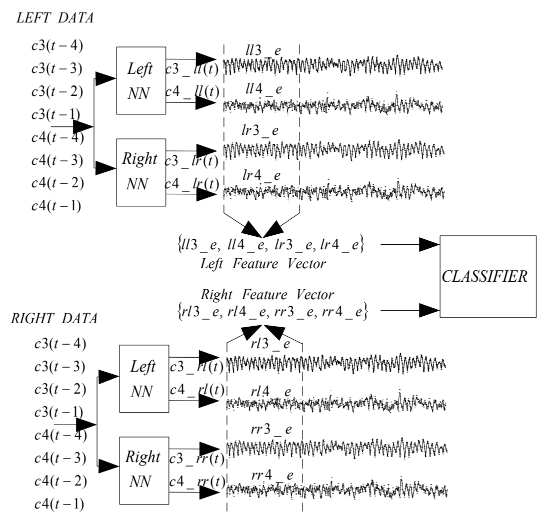


Figure 1:*Feature Extraction Procedure*

and right NNs. Each NN provides a prediction result for both channels. When left data is fed into both NNs then there are four features obtained. Equations (1) to (4) show the equations for obtaining the four left features where the index *ll*=left data→left NN and *lr*=left data→right NN. The right features are obtained by inputting the right data to both NNs. In this case the variable indexes are replaced in equations (1) to (4) with *rl*=right data→left NN and *rr*=right data→right NN. *M* is the number of prediction samples the MSE is calculated over. Therefore for each trial of left data a four element feature vector of errors is obtained and similarly for the right data. Figure 1 illustrates the FEP.

## 7. CLASSIFICATION

In addition to feature extraction, the most crucial step in the process of pattern recognition is classification. The operation of the classification step can be simplified as being that of a transform of quantitative input data to qualitative output information [13]. For initial classification it was assumed that the features from both classes were linearly separable. Fisher's linear discriminant analysis (LDA) was used to obtain a linear function which can separate the two classes. Even though LDA is not considered a universal approximator or classifier it has the ability to find the line of maximum separability between the two classes very quickly and efficiently. For classes which have very similar features LDA may not be as successful as a NN. There are few disadvantages of NNs with comparison to LDA. If the application involves a lot of parameter tuning or feature selection then any changes made to the training features or data requires that the NN has to be retrained and in some cases this can be time consuming especially if a lot of parameter tuning is involved in the FEP. An LDA classifier can be retrained very quickly.

It is suggested that for applications which do involve a lot of adjustment to the parameters of the FEP, thus a lot of classification accuracy testing, then LDA is an efficient approach. When the maximum classification accuracy is obtained using LDA (i.e. the features from all classes have maximum separability linearly), if there is possibility for improvement a NN may be used to increase the accuracy. For this approach the LDA proved to have the ability to classify the features quickly and efficiently and at the same time achieve a good accuracy. A NN classifier was also trained and comparison of both classifiers is provided.

When each classifier has been trained on the features extracted from the prediction errors the unseen test data is used to validate the effectiveness of the system i.e. the one step ahead prediction, feature extraction and subsequently the classification accuracy. The test data is fed into each NN trial by trial. The test data is configured the same as the training data.

## 8. RESULTS AND DISCUSSION

Figure 2 shows the prediction accuracy from one left trial input to the left NN. The classification technique described was tested on unseen data (140 trials) for subject S1 and 160 trials for subject S2. The best results obtained for subject S2 was 75% classification accuracy using an NN classifier. Table 1 shows various classification results obtained utilizing an LDA classifier as well as the results obtained utilizing the NN classifier for subject S1. The first and second columns specify the number of hidden layers and the number neurons in each layer, respectively. The third column specifies the training duration. The fourth column specifies the number of points the MSE of the prediction was calculated over and the number of features extracted for each channel. Using a 200 point MSE calculation the maximum number of features that can be used is 3 because each trial contains 640 data points. A 200 point MSE calculation was chosen after a number of tests because this appeared to provide the best classification accuracy. There are number of ways the features can be extracted and the results may be subject specific. For example, using a smaller number of points for the MSE calculation and a greater number

Table 1: *A Comparison of Classification Results for Different NN Architectures, Training Durations and Feature Extraction Parameters*

| Hidden Layers | Neurons | Dur. | M | LDA % | NN % |
|---|---|---|---|---|---|
| 3 | 8-10-6 | 50 | **200x1** | **84.3** | **82** |
| | | | 200x2 | 75.7 | 80 |
| | | | 200x3 | 71.4 | 81.4 |
| 2 | 6-8 | 50 | **200x1** | **84.3** | **83.6** |
| | | | 200x2 | 75.7 | 82 |
| | | | 200x3 | 75 | 82.9 |
| 1 | 10 | 50 | **200x1** | 76.4 | 82.1 |
| | | | 200x2 | **75.7** | **83.6** |
| | | | 200x3 | 80.7 | 83.6 |
| 3 | 8-10-6 | Val. Stop | **200x1** | **79.3** | **82.9** |
| | | | 200x2 | 79.3 | 81.4 |
| | | | 200x3 | 75.7 | 80.7 |
| 2 | 6-8 | Val. Stop | **200x1** | 80 | 80.7 |
| | | | 200x2 | **78.6** | **85** |
| | | | 200x3 | 77.1 | 81.4 |
| 1 | 10 | Val. Stop | **200x1** | 76.4 | 83 |
| | | | 200x2 | **77.8** | **85** |
| | | | 200x3 | 80 | 85 |

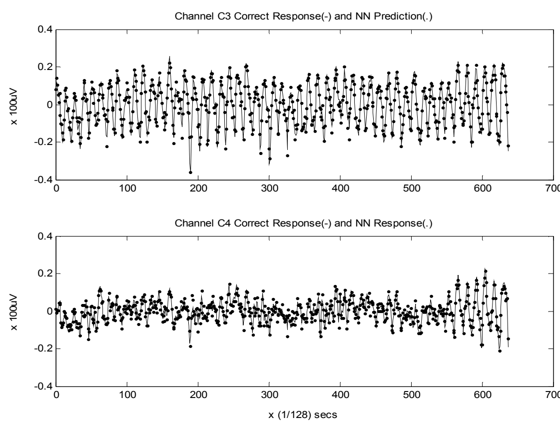


Figure 2: *Prediction Accuracy (C3 & C4) Left Data → Left Network*

of features from each channel may improve the classification accuracy. As can be seen from Table 1, in a number of experiments using 2 x 200 point MSE calculations for each channel helped improve the classification accuracy although this improvement is obtained at the expense of classification time. For example, if only one feature from each channel is used then the duration of the classification would be approximately 1.56 secs. (200*(1/128)) plus the time it takes to amplify and digitize the data, do the feature extraction and the classification. If two features from each channel are used then the classification would be approximately 3.1 secs. plus the time required for data processing. With a classification rate of approximately 1.56 seconds and 85% accuracy, information transfer rates approach 15 bits/min. Information transfer rate is a standard method of measuring effectiveness of a BCI and the calculation is derived in [8]. BCI systems must have the ability to classify signals rapidly (ideally realtime) and accurately therefore a trade-off must be made. For example, the first and second prediction NNs of Table 1 can provide features which can be classified with 84% accuracy using LDA with only one feature from each channel, thus requiring shorter classification time. For the fifth and sixth NNs of Table 1 the classification accuracy is increased by using 2 features from each channel.

The two smaller prediction NN architectures with validation-stop provided the best accuracy overall. This classification was obtained using a NN classifier. The fastest classification accuracy was obtained using the two largest prediction NNs with overgeneralization using an LDA classifier. The speed for this classification is twice as fast as that of the other two at the expense of losing a percentage in classification accuracy therefore, either of first two prediction NNs of Table 1 for prediction and feature extraction along with an LDA classifier would be the best choice as an outcome from this comparative analysis. Due to the random weight initialization of the NNs, there are always discrepancies in the results from the NNs with the same architectures so it is advisable to train the networks a number of times and compare results.

## 9. CONCLUSION

This paper has proposed a BCI system that has the ability to learn complex sequences and classify thought processes for a two class problem. This BCI can be easily adapted to different subjects, adaptability being a fundamental requirement for a BCI. This system also has potential to be fully adaptable online.

A frequency analysis of the EEG for subject S2 shows that there is a lot of activity in frequency bands other than the μ band thus suggesting that there is much more noise in this EEG. It is conjectured that this may be the cause of lower classification accuracy and that this method is not as robust for data that is heavily contaminated with noise. Future work will involve improvement of the classification accuracy to increase the information transfer rate. This may be achieved by preprocessing the data for artifact removal and noise reduction using Independent Component Analysis (ICA).

Further development of the algorithm to enable it to be periodically adapted to maintain classification accuracy and learn in parallel with the progressively varying complexity of the human brain, is also intended. Prediction, feature extraction and classification are complex processes performed in the brain. This BCI system depicts these complex tasks in a highly simplified approach utilizing cybernetic intelligence.

## 10. ACKNOWLEDGEMENTS

The authors would like to acknowledge Dr. N. Siddique for his guidance and the Institute for Biomedical Engineering, University of Technology Graz, Austria for providing the EEG data. The first author of this work is funded by a William Flynn scholarship.

## 11. REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, T. M. Vaughan, "Brain-computer interfaces for communication and control", (invited review) *J. Clinical Neurophysiology Elsevier*, vol. 113, pp. 767-791, 2002.

[2] A. E. H. Emery, "Population Frequencies of Inherited Neuromuscular Diseases – A World Survey", *Neuromuscular Disorders*, Vol. 1, No. 1, pp. 19-29, 1991

[3] J.R. Wolpaw, N, Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson and T. M. Vaughan. "Brain-Computer Interface Technology: A Review of the first International Meeting", *IEEE Trans. on Rehab. Eng*. vol. 8, no. 2, 2000.

[4] B. J. Fisch, "Fisch & Spellmann's EEG Primer: Basic principles of digital and analog EEG", *Elsevier* 1999

[5] Lytton, W. W., "Computer to Brain: Foundation of Computational Neuroscience", *Springer*, 2002.

[6] D. Coyle, G. Prasad, M. McGinnity, "The Brain-Computer Interface: A Review and a Novel Proposal", *Postgraduate Research Conference in Electronics (PREP), Photonics, Communications and Software*, 2003.

[7] D. Coyle, G. Prasad, M. McGinnity, O. Nibouche, "The Brain-Computer Interface: A Review and Some New Concepts", *Irish Signals and Systems Conference (ISSC)*, 2003.

[8] J.R. Wolpaw, H. Ramouser, D. J. McFarland and G. Pfurtscheller, "EEG-Based Communication: Improved Accuracy by Response Verification", *IEEE Trans. on Rehab. Eng*. vol. 6. No.3, 2000.

[9] C. Guger, A. Schlogl, C. Neuper, T. Strein, D. Walterspacher, and G. Pfurtscheller, "Rapid Prototyping of an EEG Based BCI", *IEEE Trans. on Neural Sys. and Rehab. Eng*. vol. 9 no.1, 2001.

[10] E. Haselsteiner and G. Pfurtscheller "Using Time-Dependent NNs for EEG Classification", *IEEE Trans. on Rehab. Eng*. vol. 8, no. 4 2000.

[11] R. Palaniappan, R. Paramesran, S. Nishida, N. Saiwaki, "A New Brain_Computer Interface Design Using Fuzzy ARTMAP", *IEEE on Neural Sys. and Rehab. Eng*. vol. 10, no. 3, 2002.

[12] G. Pfurtscheller, C. Neuper, A. Schlogl and K. Lugger, "Separability of EEG signals Recorded During Right and Left Motor Imagery Using Adaptive Autoregressive Parameters", *IEEE Trans. On Rehab. Eng*. Vol.6 No.3, 1998.

[13] C. T. Leondes(editor), "Image Processing and Pattern Recognition", *Academic Press*, 1998.