Neural network based BCI by using orthogonal components of multi-channel brain waves and generalization

著者	Nakayama Kenji, Horita Hiroki, Hirano Akihiro
journal or publication title	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)
volume	5164 LNCS, Issue PART 2
number	2008
page range	879-888
year	2008-09-01
URL	http://hdl.handle.net/2297/18078

doi: 10.1007/978-3-540-87559-8_91

Neural Network Based BCI by Using Orthogonal Components of Multi-channel Brain Waves and Generalization

Kenji Nakayama, Hiroki Horita, and Akihiro Hirano

Graduate School of Natural Science and Technology, Kanazawa University
Kanazawa, 920-1192, Japan
nakayama@t.kanazawa-u.ac.jp
http://leo.ec.t.kanazawa-u.ac.jp/~nakayama

Abstract. FFT and Multilayer neural networks (MLNN) have been applied to 'Brain Computer Interface' (BCI). In this paper, in order to extract features of mental tasks, individual feature of brain waves of each channel is emphasized. Since the brain wave in some interval can be regarded as a vector, Gram-Schmidt orthogonalization is applied for this purpose. There exists degree of freedom in the channel order to be orthogonalized. Effect of the channel order on classification accuracy is investigated. Next, two channel orders are used for generating the MLNN input data. Two kinds of methods using a single NN and double NNs are examined. Furthermore, a generalization method, adding small random numbers to the MLNN input data, is applied. Simulations are carried out by using the brain waves, available from the Colorado State University website. By using the orthogonal components, a correct classification rate P_c can be improved from 70% to 78%, an incorrect classification rate P_e can be suppressed from 10% to 8%. As a result, a rate $R_c = P_c/(P_c+P_e)$ can be improved from 0.875 to 0.907. When two different channel orders are used, P_e can be drastically suppressed from 10% to 2%, and R_c can be improved up to 0.973. The generalization method is useful especially for using a sigle channel order. P_c can be increased up to 84 $\sim$ 88% and P_e can be suppressed down to 2 $\sim$ 4%, resulting in $R_c = 0.957 \sim 0.977$.

Keywords: BCI, Brain waves, Neural network, Mental task, Orthogonal components, Gram-Schmidt, Generalization.

1 Introduction

Among the interfaces developed for the handicapped persons, Brain Computer Interface (BCI) has been attractive recently [1], [2]. Approaches to the BCI technology includes nonlinear classification by using spectrum power, adaptive auto-regressive model and linear classification, space patterns and linear classification, hidden Markov models, and so on [3],[4]. Furthermore, application of neural networks have been also discussed [5],[6],[7], [8], [9], [10]. In our works, FFT of the brain waves and a multilayer neural network (MLNN) have been

V. Kurkova, R. Neruda, and J. Koutnik (Eds.): ICANN 2008, Part II, LNCS 5164, pp. 879–888, 2008.

applied to the BCI. Efficient pre-processing techniques have been also employed in order to achieve a high score of correct classification of the mental tasks [16]. Furthermore, the generalization methods have been applied to the neural network based BCI. The method of adding small random noise to the MLNN inputs can improve classification performance [17].

The multi-channel brain waves have some common features, which make features vague, and make mental task classification difficult. We will try to remove the common and vague features and emphasize individual features of each mental task. For this purpose, essential features of the multi-channel brain waves are extracted. Conventional methods have employed Independent Component Analysis (ICA), Blind Source Separation (BSS) and so on [11],[12]. However, these methods have an essential problem, that is 'Permutation'. Order of the components, which are extracted, is not fixed. It can be changed depends on data sets. Thus, these methods are difficult to be combined with the MLNN.

In this paper, the BCI using the FFT amplitude of the brain waves and the MLNN is employed. The brain wave in some time interval, that is a frame, can be regarded as vectors. Letting M be the number of channels, M vectors are obtained for one mental task and one measuring trial. Let M vectors be $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M\}$. This vector set is transferred to the orthogonal vector set $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_M\}$. This vector set is further pre-processed and is used for the MLNN input data.

2 Brain Waves and Mental Tasks

2.1 Mental Tasks

In this paper, the brain waves, which are available from the web site of Colorado State University [13], are used. The following five kinds of mental tasks are used.

- Baseline - Relaxed situation - (B)
- Multiplication (M)
- Letter-composing (L)
- Rotation of a 3-D object (R)
- Counting numbers (C)

2.2 Brain Wave Measurement

Location of the electrodes to measure brain waves is shown in Fig.1. Seven channels including C3, C4, P3, P4, O1, O2, EOG, are used. EOG, which does not appear in this figure, is used for measuring movement of the eyeballs. In this paper, channel numbers Ch1 through Ch7 are assigned to C3, C4, P3, P4, O1, O2 and EOG, respectively, for convenience.

The brain waves are measured for a 10sec interval and sampled by 250Hz for each mental task. Therefore, 10sec $\times$ 250Hz $= 2,500$ samples are obtained for one channel and one mental task. One data set includes 2,500 samples for each channel and each mental task. Five mental tasks and seven channels are included in one data set.

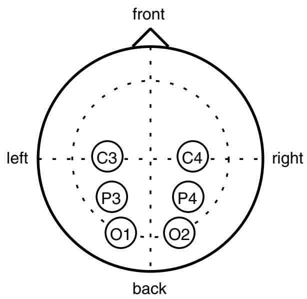

Fig. 1. Location of electrodes measuring brain waves

2.3 Mental Task Classification by Using Multilayer Neural Network

An MLNN having a single hidden layer is used. Activation functions used in the hidden layer and the output layer are a hyperbolic tangent and a sigmoid function, respectively. The number of input nodes is 10 samples$\times$7 channels=70. Five output neurons are used for five mental tasks. The target for the output has only one non-zero element, such as $(1, 0, 0, 0, 0)$. In the testing phase, the maximum output becomes the winner and the corresponding mental task is assigned. However, when the winner have small value, estimation becomes incorrect. Therefore, the answer of the neural network is rejected, that is any mental task cannot be estimated. The error back-propagation algorithm is employed for adjusting the connection weights.

3 Pre-processing of Wave Forms

Several techniques for pre-processing proposed in [16] are also employed in this paper, and are briefly described here.

3.1 Amplitude of FFT of Brain Waves

In order to avoid effects of brain wave shifting along the time axis, which is not essential, the brain wave is first Fourier transformed and its amplitude is used.

3.2 Reduction of Samples by Averaging

In order to make the neural network size to be compact and to reduce effects of the noises added to the brain waves, the FFT samples in some interval are averaged. By this averaging, the number of samples is reduced from 2,500 to 20. Since the brain waves are real values and their FFT amplitude are symmetrical, a half of the 20 samples can represents all information. Finally, 10 samples are used.

3.3 Nonlinear Normalization

The amplitude of the FFT is widely distributed. Small samples also contain important information for classifying the mental tasks. However, in the neural networks, large inputs play an important role. If large samples do not include important information, correct classification will be difficult. For this reason, the nonlinear normalization as shown in Eq.(1) has been introduced [17]. x is the FFT amplitude before normalization and $f(x)$ is the normalized amplitude. In Eq.(1), x_{min} and x_{max} mean the minimum and the maximum values of x in all channels. The small samples are expanded and the large samples are compressed. In this paper, usefulness this nonlinear normalization method for the orthogonal components of the brain waves will be also investigated.

$$f(x) = \frac{\log(x - x_{min})}{\log(x_{max} - x_{min})} \tag{1}$$

The linear normalization given by $f_{linear}(x) = (x - x_{min})/(x_{max} - x_{min})$ will be examined for comparison.

4 Generalization by Adding Small Random Numbers

The brain waves are very sensitive, which easily change depending on health conditions of the subjects and the measuring environment. The data sets measured for the same subject, have different features. Therefore, generalization is very important for the BCIs. In our previous work, two kinds of generalization techniques, which are adding small random numbers to the MLNN input data [14] and a weight decay technique [15], have been applied. The former method can provide good classification performance [17].

In this paper, the method of adding small and different random numbers to the MLNN input data at each epoch of the learning process is applied, and its usefulness for the orthogonal components of the multi-channel brain waves will be investigated.

5 Orthogonal Components of Multi-channel Brain Waves

5.1 Orthogonal Component Analysis

There are several kinds of methods for analyzing orthogonal components, including blind source separation (BSS), independent component analysis (ICA), principal component analysis (PCA) and so on. They have an essential problem, that is 'Permutation'. In these methods, it is not guaranteed that the same component is analyzed in the same order. This point is described in detail here.

Letting M be the number of channels, a set of M vectors is measured for one mental task and one measuring trial. Let a whole vector be $\boldsymbol{X} = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \cdots, \boldsymbol{x}_M^T]^T$. $\boldsymbol{x}_i$ corresponds to the brain wave measured at the ith channel. Let the corresponding orthogonalized vector be $\boldsymbol{V} = [\boldsymbol{v}_1^T, \boldsymbol{v}_2^T, \cdots, \boldsymbol{v}_M^T]^T$. The whole vector $\boldsymbol{V}$ is used as the MLNN input data after the pre-processing. Therefore, the order of $\boldsymbol{v}_i$ in $\boldsymbol{V}$ is very important. Let consider two sets of the vectors $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ for the same mental task and for different measuring trials. Let $\boldsymbol{v}_{1i}$ and $\boldsymbol{v}_{2j}$ be the ith and jth vectors in $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$, respectively. If $\boldsymbol{v}_{1i}$ is most similar to $\boldsymbol{v}_{2j}, i \neq j$, for instanse $\boldsymbol{v}_{1i}^T \boldsymbol{v}_{2j} / \parallel \boldsymbol{v}_{1i} \parallel \parallel \boldsymbol{v}_{2j} \parallel$ is most close to 1, then $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ cannot express the same feature even though they belong to the same mental task. BSS, ICA and PCA cannot guarantee the same order of the orthogonal components in the different measuring trials due to 'Permutation' problem [11],[12]. This is an essential problem to use the orthogonal components as the MLNN input data.

For this reason, in this paper, Gram-Schmidt orthogonalization is applied to the vector set $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M\}$. The order of the orthogonal components can be controlled by selecting the channels to be orthogonalized in the specified order. 'Permutation' of the orthogonal components does not occur.

5.2 Gram-Schmidt Orthogonalization

The vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M\}$, which express the brain waves at M-channels, are usually linearly independent. This set can be transferred into the orthogonal vector set $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_M\}$ by Gram-Schmidt orthogonalization [18]. $\boldsymbol{x}_1$ is used for $\boldsymbol{v}_1$. $\boldsymbol{v}_2$ is a part of $\boldsymbol{x}_2$, which is orthogonal to $\boldsymbol{v}_1$. In the same way, $\boldsymbol{v}_k$ is a component of $\boldsymbol{x}_k$, which is orthogonal to all the previous orthogonal vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{k-1}$.

$\{\boldsymbol{v}_i\}$ are Fourier transformed and their amplitude are pre-processed as described in Sec.3, and are used as the MLNN input data.

5.3 Order of Orthogonalization

There exists degree of freedom of selecting the channel order, in which the brain waves are orthogonalized by the Gram-Schmidt method. The first channel can hold a whole information, and the following channels provide only a part of the vector, which is orthogonal to the previous orthogonal vectors. Therefore, the order of the channels will affect accuracy of classifying the mental tasks. We will investigate effects of the channel order through simulation.

5.4 Input Data Sets by Using Two Channel Orders

As described in the previous sections, we can use a plural number of the channel orders for generating the MLNN input data. Let the input data sets, which are generated from the same brain waves by using two kinds of the channel orders, be I_1 and I_2. Thus, the input data are equivalently doubled for each mental task.

Single MLNN Method. A single MLNN, denoted NN, is used for classifying the mental tasks. In the training phase, both I_1 and I_2, generated from the training brain waves, are used to train NN. In the testing phase, I_1, generated from the test brain waves, is applied to NN, and the output O_1 is obtained. Separately, I_2, generated from the same brain waves, is applied to NN, resulting in the output O_2. The final output O_t is given by $O_t = (O_1 + O_2)/2$. The mental task is estimated based on the maximum value of O_t.

Double MLNN Method. Two independent MLNNs are used, denoted NN_a and NN_b. I_1 of the training brain waves is used to train NN_a, and I_2 of the same brain waves is used to train NN_b, respectively. In the testing phase, I_1 of the test brain waves is applied to NN_a and the output O_{a1} is obtained. In the same way, I_2 of the same brain waves is applied to NN_b and O_{b2} is obtained. The final output O_t is evaluated by $O_t = (O_{a1} + O_{b2})/2$. The mental task is classified based on the maximum value of O_t.

The threshold of rejection, that is 'No estimation', is also employed in all methods. If all the outputs are less than the threshold, then the MLNN answers 'any mental task cannot be estimated'.

6 Simulations and Discussions

6.1 Simulation Setup

Training and Testing Brain Waves
The brain waves with a 10 sec length for five mental tasks were measured 10 times. Therefore, 10 data sets are available. Among them, 9 data sets are used for training and the remaining one data set is used for testing. Five different combinations of 9 data sets are used for the training. As a result, five different data sets are used for testing. Thus, five independent trials are carried out. Classification accuracy is evaluated based on the average over five trials [3].

Score of Correct and Error Classifications
Estimation of the mental tasks is evaluated based on a correct classification rate (P_c) and an error classification rate (P_e), and a rate of correct and error classification (R_c) as follows:

$$P_c = \frac{N_c}{N_t} \times 100\%, \quad P_e = \frac{N_e}{N_t} \times 100\% \tag{2}$$

$$R_c = \frac{N_c}{N_c + N_e}, \quad N_t = N_c + N_e + N_r \tag{3}$$

N_c, N_e and N_r are the numbers of correct and incorrect classifications and rejections, respectively. N_t is the total number of the testing data. R_c is used to evaluate a correct classification rate except for 'Rejection'.

Parameters in Neural Network Learning
A hyperbolic tangent function and a sigmoid function are used fin the hidden layer and the output layer, respectively. The number of hidden units is 20. The threshold for rejection is 0.7. A learning rate is 0.02.

6.2 Brain Waves Before and After Orthogonalization

Figure 2 shows the brain waves before (gray) and after (black) orthogonalization. The horizontal axis shows the sample number in the time domain. The channel

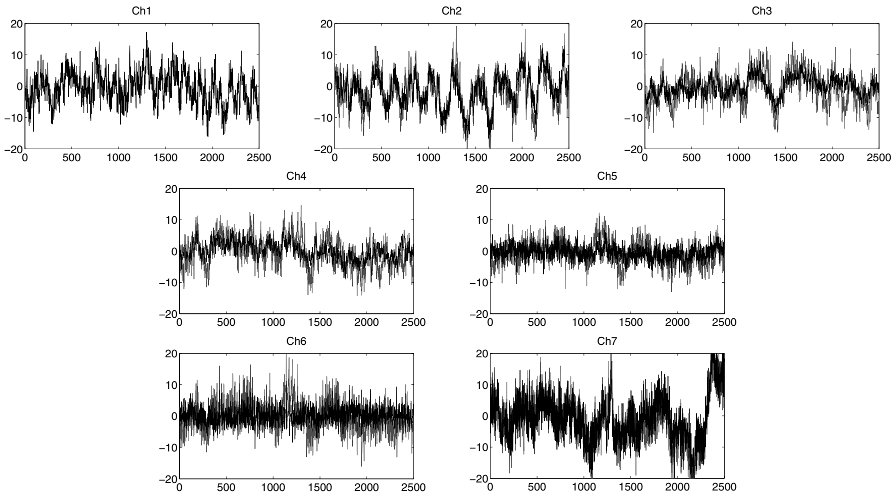

Fig. 2. Brain waves of 7 channels before (gray) and after (black) orthogonalization

order of orthogonalization is Ch1, 2, 3, 4, 5, 6, 7. The orthogonalized brain waves are gradually decreased along the channel order. Since Ch7 is used for detecting blinking, which is not the mental task, its brain wave is not correlated with that of the other channels, resulting in a large orthogonal component.

Figure 3 shows the MLNN input data before (dashed line) and after (solid line) orthogonalization. They are normalized by using the nonlinear function Eq.(1) [16]. The FFT amplitude responses are arranged from Ch1 through Ch7 along the horizontal axis from the left side to the right side. One channel includes 10 samples.

6.3 Classification by Using Orthogonal Components

Table 1 shows classification rates by using the orthogonal components of the brain waves. 7 kinds of channel orders are used, which are selected by circular shifting. Although they do not include all permutations, effects of the channel order can be investigated. 'Conventional' means our method, which employs the original brain waves and the pre-processing techniques [16]. By using the orthogonal components, P_c can be improved from 70% to 78%, and P_e can be suppressed from 10% to 8%, and R_c is increased from 0.875 to 0.907.

As expected in the previous section, also from Table 1, the classification accuracy depends on the channel order to be orthogonalization. The channel orders Ch2, 3, 4, 5, 6, 7, 1 and Ch3, 4, 5, 6, 7, 1, 2 can provide good classification accuracy. The optimum channel order can be searched for in advance, and can be fixed for an individual subject.

Furthermore, the generalization method of adding small random numbers to the MLNN input data was carried out for the best channel order Ch2, 3, 4, 5, 6,

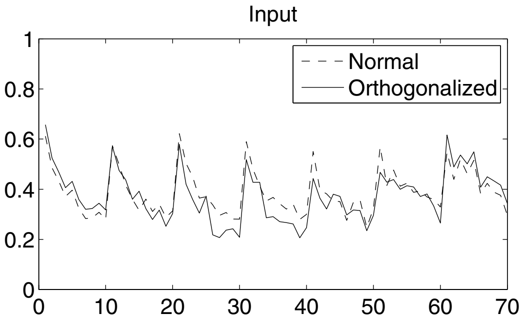

Fig. 3. Input data of MLNN. Dashed line and solid line indicate input data before and after orthogonalization, respectively.

Table 1. Score of classification by using orthogonal components

	P_c	P_e	R_c
Conventional	70	10	0.875
Ch1, 2, 3, 4, 5, 6, 7	70	12	0.854
Ch2, 3, 4, 5, 6, 7, 1	78	8	0.907
Ch3, 4, 5, 6, 7, 1, 2	74	8	0.902
Ch4, 5, 6, 7, 1, 2, 3	70	12	0.854
Ch5, 6, 7, 1, 2, 3, 4	68	12	0.85
Ch6, 7, 1, 2, 3, 4, 5	66	24	0.733
Ch7, 1, 2, 3, 4, 5, 6	70	12	0.854
Generalization ±0.1 Ch2, 3, 4, 5, 6, 7, 1	88	4	0.957
Generalization ±0.05 Ch2, 3, 4, 5, 6, 7, 1	84	2	0.977

7, 1. Random numbers uniformly distributed during ±0.1 and ±0.05 are used. As shown in the same table, P_c is well improved from 78% up to 84 $\sim$ 88%, P_e is well suppressed from 8% to 2 $\sim$ 4%, resulting in $R = 0.957 \sim 0.977$.

The linear normalization $f_{linear}(x)$ described in Sec.3.3 is also examined. The best channel order is Ch5, 6, 7, 1, 2, 3, 4, and $P_c = 62\%$, $P_e = 14\%$ and $R_c = 0.816$, which are not good compared with the nonlinear normalization.

6.4 Classification by Using Two Channel Orders

Two channel orders, Ch2, 3, 4, 5, 6, 7, 1 and Ch3, 4, 5, 6, 7, 1, 2, which provide good classification accuracy in Table 1, are used to generate the MLNN input data I_1 and I_2, respectively.

Table 2 shows simulation results. A method of using a single NN is not good. By using double NN, P_e is well suppressed and R_c can be well increased before the generalization. After the generalization, its performances $P_c = 82\%$, $P_e = 2\%$

Table 2. Score of classification by using two channel orders

	P_c	P_e	R_c
Conventional	70	10	0.875
Single NN	66	8	0.892
Generalization ±0.1	76	6	0.927
±0.05	78	8	0.907
Double NN	72	2	0.973
Generalization ±0.1	82	2	0.976
±0.05	72	4	0.947

and $R_c = 0.976$ are almost same as those of the method using a single channel order Ch2, 3, 4, 5, 6, 7, 1.

When the generalization method, adding random numbers to the MLNN input data, is embedded in the learning process, a single channel order can provide good performance by optimizing the channel order.

6.5 Dependence on Individual Subjects

Since brain waves are dependent on the subjects, the MLNN is needed to be optimized or tuned up for individual subjects. It is not useful to apply the same MLNN to different subjects. From our experiences, the best channel order of orthogonalization also depends on both the subjects and mental tasks. However, it can be searched for by using the training data in advance. The proposed method, using the orthogonal components, has been applied to three subjects, and almost the same improvement on the classification rates have been achieved.

7 Conclusion

In this paper, the BCI based on the FFT amplitude and the MLNN is dealt with. Especially, the orthogonal components of the multi-channel brain waves are used to generate the MLNN input data. Gram-Schmidt orthogonalization is applied. The proposed approach can improve P_c from 70% to 78%, P_e from 10% to 8%, and R_c from 0.875 to 0.907. When two channel orders are used, P_e can be well suppressed from 10% to 2%, and R_c can be well improved up to 0.973. The generalization method is also useful, which can improve P_c up to 88% and P_e down to 2%, resulting in $R_c = 0.977$.

References

1. Pfurtscheller, G., Neuper, C., Guger, C., Harkam, W., Ramoser, H., Schlögl, A., Obermaier, B., Pregenzer, M.: Current trends in Graz braincomputer interface (BCI) research. IEEE Trans. Rehab. Eng. 8, 216–219 (2000)

2. Obermaier, B., Muller, G.R., Pfurtscheller, G.: Virtual keyboard controlled by spontaneous EEG activity. IEEE Trans. Neural Sys. Rehab. Eng. 11(4), 422–426 (2003)
3. Anderson, C., Sijercic, Z.: Classification of EEG signals from four subjects during five mental tasks. In: Bulsari, A.B., Kallio, S., Tsaptsinos, D. (eds.) EANN 1996, Systems Engineering Association, PL 34, FIN-20111 Turku 11, Finland, pp. 407–414 (1996)
4. Pfurtscheller, G., Neuper, C.: Motor imagery and direct brain-computer communication. Proc. IEEE 89(7), 1123–1134 (2001)
5. Millan, J.R., Mourino, J., Babiloni, F., Cincotti, F., Varsta, M., Heikkonen, J.: Local neural classifier for EEG-based recognition of metal tasks. In: IEEE-INNS-ENNS Int. Joint Conf. Neural Networks (July 2000)
6. Muller, K.R., Anderson, C.W., Birch, G.E.: Linear and non-linear methods for brain-computer interfaces. IEEE Trans. Neural Sys. Rehab. Eng. 11(2), 165–169 (2003)
7. Millan, J.R.: On the need for on-line learning in brain-computer interfaces. In: Proc. IJCNN, pp. 2877–2882 (2004)
8. Fabiani, G.E., McFarland, D.J., Wolpaw, J.R., Pfurtscheller, G.: Conversion of EEG activity into cursor movement by a brain-computer interface (BCI). IEEE Trans. Neural Sys. Rehab. Eng. 12(3), 331–338 (2004)
9. Obermaier, B., Neuper, C., Guger, C., Pfurtscheller, G.: Information transfer rate in a five-classes brain-computer interface. IEEE Trans. Neural Sys. Rehab. Eng. 9(3), 283–288 (2001)
10. Anderson, C.W., Devulapalli, S.V., Stolz, E.A.: Determining mental state from EEG signals using neural networks. Scientific Programming, Special Issue on Applications Analysis 4(3), 171–183 (1995)
11. Kachenoura, A., Albera, L., Senhadji, L., Comon, P.: ICA: A potential tool for BCI systems. IEEE Signal Processing Magazine, 57–68 (January 2008)
12. Wang, S., James, C.J.: Enhancing evoked responses for BCI through advanced ICA techniques. In: Proc., 3rd Int. Conf. on Advances in Medical, Signal and Information Processing (MEDSIP 2006), Stevenage, UK (2006)
13. Colorado State University, `http://www.cs.colostate.edu/eeg/`
14. Robert, J., Burton, M., Mpitsos, G.J.: Event-dependent control of noise enhances learning in neurla networks. Neural Networks 5(4), 627–637 (1992)
15. Treadgold, N.K., Gedeon, T.D.: Simulated annealing and weight decay in adaptive learning: The SARPROP algorithm. IEEE Trans. Neural Networks 9(4), 662–668 (1998)
16. Nakayama, K., Inagaki, K.: A brain computer interface based on neural network with efficient pre-processing. In: Proc. IEEE, ISPACS 2006, Yonago, Japan, pp. 673–676 (December 2006)
17. Nakayama, K., Kaneda, Y., Hirano, A.: A brain computer interface based on FFT and multilayer neural network-Feature extraction and generalization. In: Proc. IEEE, ISPACS 2007, Xiamen, China, pp. 101–104 (December 2007)
18. Strang, G.: Linear Algebra and its Applications. Academic Press, Inc., New York (1976)