社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
CAS2007-19, VLD2007-35, SIP2007-49(2007-6)

FFTと階層形ニューラルネットワークによるブレイン・コンピュータ・インタフェイス

特徴抽出と汎化能力の向上について

金田 泰明† 中山 謙二† 平野 晃宏†

† 金沢大学 大学院自然科学研究科 〒920-1192 石川県金沢市角間町
E-mail: †kaneda@leo.ec.t.kanazawa-u.ac.jp, ††{nakayama,hirano}@t.kanazawa-u.ac.jp

あらまし 脳波のFFTと階層形ニューラルネットワークを用いるブレイン・コンピュータ・インタフェイス（BCI）に関して，以前に前処理の方法をいくつか提案し，メンタルタスクの分類性能を向上した．本稿では，まず，階層形ニューラルネットワークでメンタルタスクを分類するために用いられる特徴の解析を行った．特徴は結合荷重の分布に基づいて解析した．隠れ層から出力層への結合荷重はメンタルタスクに対して独立になる傾向があった．従って，入力層から各メンタルタスクに対応する隠れユニットへの結合荷重分布がメンタルタスク毎の特徴を表している．次に，汎化能力を向上する２通りの学習法について検討を行った．一つは，ニューラルネットワークの入力データに乱数を加える方法であり，もう一つは，結合荷重を圧縮する方法する方法である．シミュレーションの結果，いずれの方法もテストデータに対する分類性能を向上することが出来たが，乱数を加える方法が有効であることが分かった．

キーワード ブレイン・コンピュータ・インタフェイス，メンタルタスク，階層形ニューラルネットワーク，特徴抽出，汎化能力

A Brain Computer Interface Based on FFT and Multilayer Neural Network

Feature Extraction and Generalization

Yasuaki KANEDA†, Kenji NAKAYAMA†, and Akihiro HIRANO†

† Graduate School of Natural Science and Technology, Kanazawa Univ. Kakuma-machi, Kanazawa, Ishikawa, 920-1192 Japan
E-mail: †kaneda@leo.ec.t.kanazawa-u.ac.jp, ††{nakayama,hirano}@t.kanazawa-u.ac.jp

Abstract In this paper, a multilayer neural network is applied to 'Brain Computer Interface' (BCI), which is one of hopeful interface technologies between humans and machines. Amplitude of the FFT of the brain waves are used for the input data. Several techniques have been introduced for pre-processing the brain waves. They include segmentation along the time axis for fast response, nonlinear normalization to emphasize important information, averaging samples of the brain waves to suppress noise effects, reduction in the number of the samples to realize a small size network, and so on. In this paper, two kinds of generalization techniques, including adding small random noises to the input data and decaying connection weight magnitude, are applied. Their usefulness are analyzed and compared based on correct and error classifications. Simulation is carried out by using the brain waves, which are available from the web site of Colorado State University. The number of mental tasks is five. Some data sets are used for training the multilayer neural network, and the remaining data sets are used for testing. In our previous work, classification accuracy of 64%∼74% for the test data have been achieved. In this paper, by applying the generalization techniques, the accuracy can be improved up to 80%∼88%.

Key words Brain computer interface, Mental task, Multilayer neural network, Feature extraction, Generalization

1. Introduction

Nowadays, several kinds of interfaces between humans and computers or machines have been proposed and developed. For persons being in a healthy condition, keyboards and mouses are useful and practical interfaces. On the other hand, for handicapped persons, several interface techniques, which use available organs and functions, have been studied and developed.

Among the interfaces developed for the handicapped persons, Brain Computer Interface (BCI) has been attractive recently. A subject imagines some mental tasks, and the brain waves are measured. The brain waves are analyzed and the mental tasks are estimated. Furthermore, based on the estimated mental task, computers and machines are controlled [1].

It can be expected that severely handicapped persons, who cannot control any parts of their own body, can control a wheelchair, computers and other machines through the BCI [2]. Furthermore, in the virtual reality (VR) technology, it may be possible to control a person in the VR world, and to have many kinds of experiences in the VR world. For instance, training how to move in danger situations may be possible by using the BCI technology.

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 have been applied to the BCI. Efficient pre-processing techniques have been also employed in order to achieve a high probability for correct classification of the mental tasks [15], [16], [17].

In this paper, features of the brain waves, which are extracted by the multilayer neural network, are analyzed. Furthermore, two kinds of generalization techniques are applied to increasing accuracy of classification. Simulations are carried out by using the brain waves, which are available from the web site of Colorado State University [11]. Estimation results of the proposed method are compared to the conventional methods.

2. Mental Tasks and Brain Wave Measurement

2.1 Mental Tasks

In this paper, the brain waves, which are available from the web site of Colorado State University [11], are used. The following five kinds of mental tasks are used as imaging.

- Baseline (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.

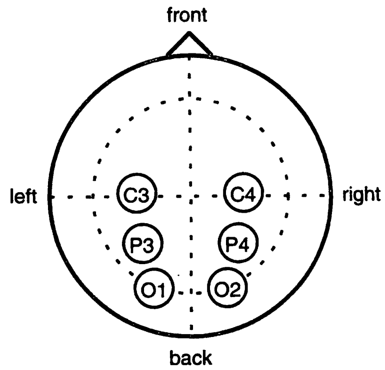

図 1 Location of electrodes measuring brain waves.

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. Therefore, 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. One example of the brain wave is shown in Fig.2.

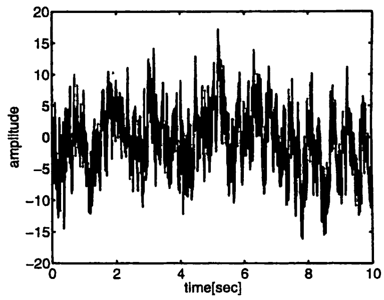

図 2 Example of brain waves in time domain.

3. Pre-Processing

3.1 Segmentation along Time Axis

In order to make the BCI response fast, the brain wave measured during 10sec is divided into the segments of a 0.5sec length as shown in Fig.3. The segmentation is shifted by 0.25sec. This means the segment of a 0.5sec length can be obtained every 0.25sec. The brain wave segment in a 0.5sec length is used to classify the mental tasks.

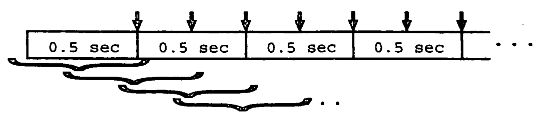

図 3 Segmentation of brain wave along time axis.

3.2 Amplitude of FFT of Brain Waves

What kinds of features of the brain waves should be used to classify the mental tasks is very important. 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. The segment of the brain wave of a 0.5sec length includes $2,500 \times 0.5/10 = 125$ samples. An example of the FFT amplitude of the segment is shown in Fig.4(Left).

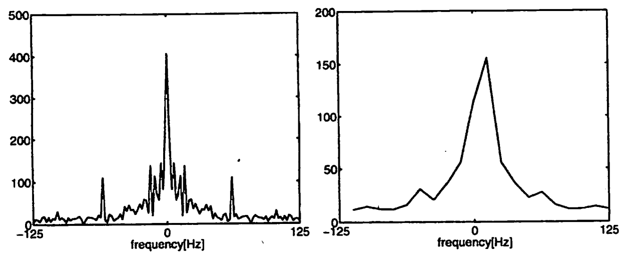

図 4 Amplitude of FFT of brain wave segment with 0.5sec length. (Left) 125 samples. (Right) 20 samples obtained by averaging.

3.3 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. The average value is used to express the representative sample in this interval, and is used for the neural network input. By this averaging, the number of samples is reduced from 125 to 20. The amplitude of the FFT for the reduced samples is also shown in Fig.4(Right).

3.4 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) and Fig.5 is introduced in this paper. x is the 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. The small samples are expanded and the large samples are compressed.

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

3.5 Input of Neural Network

Since the amplitude response is symmetrical in the frequency range from 0 to f_s, which is a sampling frequency, only the right hand side is used. Furthermore, the amplitude response of the seven channels are simultaneously applied to the neural network. An example of the neural network input is shown in Fig.6.

4. Mental Task Classification by Using Multilayer Neural Network

A multilayer neural network having a single hidden layer is used.

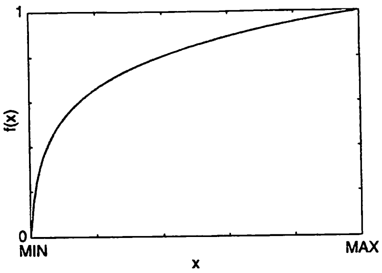

図 5 Nonlinear normalization for amplitude of FFT.

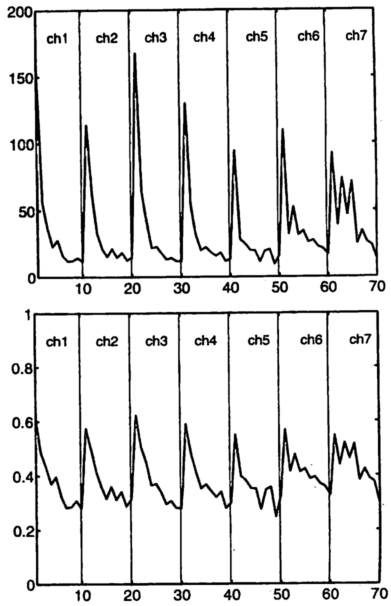

図 6 Input of neural network including 7-channels for one mental task. (Upper) Before normalization. (Lower) After normalization.

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.

5. Genaralization

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. The generalization is equivalent to make a boundary at the middle point between the different mental task regions in the input space, taking a probability of the input data into account.

In this section, two kinds of the generalization techniques , which are adding random noise to the neural network input [12] and a weight decay technique [13], are applied to learning the neural network.

5.1 Adding Random Noise to NN Input Data

By adding small and different random noises to the neural network input data at each epoch of the learning process, the region, where the input data of a mental task are distributed, can be expanded. The boundary between the different mental tasks in the input space can be set at the middle point of their regions.

5.2 Weight Decay Method

When the magnitude of the connection weights of the neural network are large, slopes of the hyperbolic tangent and the sigmoid function become steep. When the slope is steep, the boundary can locate at any point between the different mental task regions in the input space. By suppressing the magnitude of the connection weights in a learning process, the slope can be controlled to be gentle. As a result, the boundary can be set at the middle point between the different mental task regions [13], [14].

6. Simulations and Discussions

6.1 Simulation Setup

6.1.1 Training and Testing Brain Waves

The brain waves with 10sec 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].

6.1.2 Probability of Correct and Error Classifications

Estimation of the mental tasks is evaluated based on probabilities of correct classification (P_c) and error classification (P_e), and a correct classification rate (R_c).

$$P_c = \frac{N_c}{N_t} \times 100\%, P_e = \frac{N_e}{N_t} \times 100\%, R_c = \frac{N_c}{N_c + N_e} \quad (2)$$

$$N_t = N_c + N_e + N_r \quad (3)$$

N_c, N_e and N_r are the numbers of correct and error classifications and rejections. N_t is the total number of the training data or the testing data. P_c can express probability of correct classifications, and P_e expresses mis-classifications for all data. R_c is used to evaluate a correct classification rate except for 'Rejection'.

6.1.3 Parameters in Neural Network Learning

- Activation functions:
  Hidden layer: Hyperbolic tangent
  Output layer: Sigmoid function
- The number of hidden neurons: 20
- A learning rate: 0.2
- Initial weights: Random numbers in -0.1∼+0.1
- The threshold for rejection: 0.8

6.2 Probabilities of Classification

6.2.1 Learning Curves

Figure 7 shows learning curves for 'Training data' and 'Testing data'. The vertical axis indicates P_c and the horizontal axis is the iteration number. P_c of the training data denoted 'Learning' converges almost 100%, and P_c of the testing data denoted 'Test' approaches to around 80%.

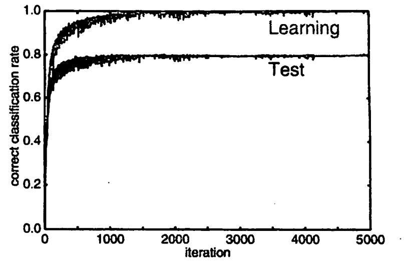

図 7 Learning curves for training data (Learning) and testing data (Test).

6.2.2 Effects of Segmentation

Aim of the segmentation is to make the BCI response fast. On the other hand, the length is limited to 0.5sec. In this section, effects of this short length is investigated. As described in the previous section, The 10 sec length of the brain waves is divided into 0.5 sec, which has 125 samples.

The probability of correct and error classifications and their rate, that is P_c, P_e and R_c, are shown in Tabel 1. The probabilities were averaged over the iterations from 4,001 to 5,000. From these results, accuracy of the segmentation is almost the same as that of using all data.

表 1 Probability of classifications for segmentation.

	Training data			Testing data		
	P_c	P_e	R_c	P_c	P_e	R_c
No Segmentation	100.0	0.0	1.00	78.0	0.0	1.00
Segmentation	99.7	0.1	0.99	79.7	10.5	0.88

6.2.3 Effects of Nonlinear Normalization

Effects of the nonlinear normalization given by Eq.(1) on the mental task classification accuracy is investigated. For reference, linear normalization, by which the sample values are linearly normalized from 0 to 1, is also used. The segmentation is used, and 125 samples are reduced to 20 samples by averaging.

The learning curves are shown in Fig.8. Furthermore, the probability of correct and error classifications and their rate are shown in Table 2. From these results, the nonlinear normalization can make convergence of the learning fast, and the probability can be also improved.

6.2.4 Threshold for Rejection

When the winner output has small value, the estimation by the neural network is not accurate. Therefore, the result is rejected. In this section, effects of the threshold for rejection is investigated. The probability of correct and error classifications and their rate are

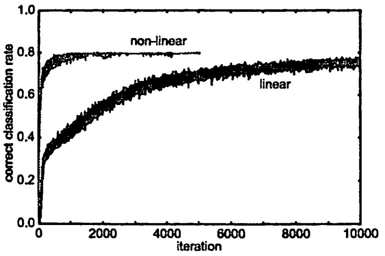

図 8 Learning curves for linear and nonlinear normalization.

表 2 Probability of classifications for linear and nonlinear normalization.

	Training data			Testing data		
Normalization	P_c	P_e	R_c	P_c	P_e	R_c
Linear	81.8	1.9	0.98	68.4	9.8	0.88
Nonlinear	99.7	0.1	0.99	79.7	10.5	0.88

listed in Table 3. By setting the threshold to be low, P_c can be improved, however, P_e is also increased at the same time. In some applications, P_e should be minimized while the rejection can permitted to some extent. In these cases, high threshold is preferred. Like this, the optimum threshold is highly dependent on applications, to which the BCI system is applied.

表 3 Probability of classifications different thresholds.

Threshold	P_c	P_e	R_c
0.8	80.3	10.0	0.89
0.6	81.7	11.9	0.87
0.4	82.9	13.2	0.86
0.2	83.7	14.2	0.86
0.0	84.5	15.5	0.84

6.3 Feature Extraction

The connection weights of the trained multilayer neural network are investigated in order to analyze feature extracted by the neural network in the learning process. Magnitude of the connection weights from the hidden layer to the output layer are shown in Fig.9. The horizontal axis indicates the hidden unit number and the vertical axis is the output unit number. 20 hidden units and 5 output units, which correspond to the 5 mental tasks, are used. Dark color means large magnitude and light color is small magnitude. For example, the 4th and the 9th hidden units have large connection weights for the 1st mental task. Furthermore, the 1st, 7th and 17th hidden units have large connection weights for the 3rd mental task. The connection weights from these hidden units to the other output units do not have large magnitude. Thus, they play an important role for the 1st and the 3rd mental tasks. Therefore, it can be recognized that these hidden units extract the features for these mental tasks. The feature of the FFT of the brain waves for these mental tasks can be expressed by using the connection weights from the input layer to these hidden units, as shown in Fig.10. In this figure, the horizontal axis indicates the input unit number and the vertical axis means the hidden unit number.

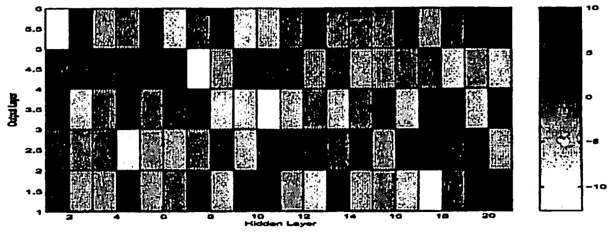

図 9 Connection weights from hidden layer to output layer.

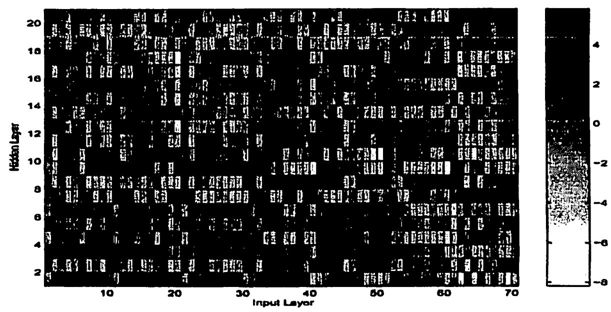

図 10 Connection weights from input layer to hidden layer.

6.4 Effects of Generalization Techniques

6.4.1 Adding Random Noise

The input data of the multilayer neural network are distributed from 0 to 1 by the nonlinear normalization. One example is shown in Fig.6, in which the input data are distributed from 0.25 to 0.6. In this simulation, random numbers, which are uniformly distributed in the $-0.1 \sim 0.1$ interval, are added to the input data. This range is determined by experience. Actually, it should be optimized for each problem.

The probability of correct and error classifications and their rate for the testing data are listed in Table 4. In Subject 1, P_c is increased from 74% to 88% and P_e is decreased from 6% to 2%. Thus, R_c is increased from 0.925 to 0.978. In Subject 2, P_c is increased from 64.0% to 80.0%. P_e is decreased from 8.0% to 4%. Therefore, R_c is also well improved from 0.889 to 0.952. As a result, efficiency of generalization of adding random noises to the input data can be recognized.

表 4 Improvement of classifications by generalization.

	Subject 1			Subject 2		
Methods	P_c	P_e	R_c	P_c	P_e	R_c
No generalization	74.0	6.0	0.925	64.0	8.0	0.889
Adding random noises	88.0	2.0	0.978	80.0	4.0	0.952
Weight decay	82.0	4.0	0.954	60.0	4.0	0.938
Weight decay & Scaling	88.0	12.0	0.88	84.0	16.0	0.84

6.4.2 Generalization by Weight Decay

The connection weights are compressed at each epoch by multiplying the following variable factor $g(n)$.

$$g(n) = g_0 + (1 - g_0)\frac{1 - e^{2\pi an}}{1 + e^{2\pi an}}, \quad n \geq 0 \tag{4}$$

$$\hat{w}(n) = g(n)w(n) \tag{5}$$

$g_0 = 0.99 \sim 0.994$ and $a = 0.5 \sim 1$ are used, which are determined by experience. $\hat{w}(n)$ will be used in the next epoch. The probabilities of the classification are also listed in Tabel 4, denoted 'Weight decay'. In the case of Subject 1, P_c, P_e and R_c are all improved from thoese of 'No generalization'. However, the improvements are inferior to 'Adding random noises'. In the case of Subject 2, even though P_c is decreased from 'No generalization', P_e and R_c can be improved. In this case, 'Adding random noises' still can provide good performances.

In the weight decay method, the boundary can be controlled at the middle point between different classes. However, the outputs of the multilayer neural network gradually change from one class to the other class. In order to emphasize the outputs, that is to make the outputs more clear, all connection weights are scaled up by multiplying $1.5 \sim 2$. Probabilities of classification are also listed in Table 4 denoted 'Weight decay & Scaling'. In this case, P_c can be improved from 82% to 88% and from 60% to 84% for Subject 1 and 2, respectively. However, P_e is also increased from 4% to 12% and from 4% to 16%, resulting in lower R_c.

6.5 Comparison to Conventional Methods

The same brain waves were used in [5]. The coefficients of a 6-th order auto-regression model are used for the neural network input. The data length of 10 sec is also divided into 0.5 sec segments. The probability of correct classifications is $30\% \sim 60\%$ for four subjects. Therefore, the proposed method, which employs the amplitude response of the FFT, the nonlinear normalization, the averaging, the rejection and the generalization technique can provide higher probability of correct mental classification.

7. Conclusion

A multilayer neural network has been applied to the BCI problem. In order to improve accuracy of mental task classification and to achieve a fast response and a small size network, several efficient pre-processing methods for the input data of the neural network have been proposed in out previous work. The probabilities of correct classification of $64\% \sim 74\%$ have been obtained. In this paper, two kinds of generalization techniques are applied. The accuracy is more increased to $80\% \sim 88\%$. Compared to the conventional methods, the higher probability of correct classification can be obtained. Furthermore, features, which are used to classify the mental tasks, are analyzed.

文　献

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

[2] B. Obermaier, G. R. Muller, and G. Pfurtscheller, "Virtual keyboard controlled by spontaneous EEG activity", IEEE Trans. Neural Sys. Rehab. Eng., vol. 11, no. 4, pp.422-426, Dec. 2003.

[3] C. Anderson and Z. Sijercic, "Classification of EEG signals from four subjects during five mental tasks", EANN'96, ed. by Bulsari, A.B., Kallio, S., and Tsaptsinos, D., Systems Engineering Association, PL 34, FIN-20111 Turku 11, Finland, pp. 407-414, 1996.

[4] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication, "Proc. IEEE, vol. 89, no. 7, pp. 1123-1134, July 2001.

[5] J. R. Millan, J. Mourino, F. Babiloni, F. Cincotti, M. Varsta, and J. Heikkonen,"Local neural classifier for EEG-based recognition of metal tasks,"IEEE-INNS-ENNS Int. Joint Conf. Neural Networks, July 2000.

[6] K. R. Muller, C. W. Anderson, and G. E. Birch, " Linear and non-linear methods for brain-computer interfaces,"IEEE Trans. Neural Sys. Rehab. Eng., vol. 11, no. 2, pp. 165-169, 2003.

[7] J. R. Millan, "On the need for on-line learning in brain-computer interfaces", Proc. IJCNN, pp. 2877-2882, 2004.

[8] G. E. Fabiani, D. J. McFarland, J. R. Wolpaw, and G. Pfurtscheller, "Conversion of EEG activity into cursor movement by a brain-computer interface (BCI)", IEEE Trans. Neural Sys. Rehab. Eng., vol. 12, no. 3, pp. 331-338, Sept. 2004.

[9] B. Obermaier, C. Neuper, C. Guger, and G. Pfurtscheller, "Information transfer rate in a five-classes brain-computer interface", IEEE Trans. Neural Sys. Rehab. Eng., vol.9, no.3, pp.283-288, 2001.

[10] C.W. Anderson, S.V. Devulapalli, and E.A. Stolz, "Determining mental state from EEG signals using neural networks", Scientific Programming, Special Issue on Applications Analysis, vol.4, no.3, pp.171-183, Fall, 1995.

[11] Colorado State University: http://www.cs.colostate.edu/eeg/

[12] J.Robert, M.Burton and G.J.Mpitsos, "Event-dependent control of noise enhances learning in neurla networks", Neural Networks, vol.5, no.4, pp.627-637, 1992.

[13] N.K.Treadgold and T.D.Gedeon, "Simulated annealing and weight decay in adaptive learning: The SARPROP algorithm", IEEE Trans. Neural Networks, vol.9, no.4, pp.662-668, July 1998.

[14] M.Tonomura and K.Nakayama, "A hybrid learning algorithm for multilayer perceptrons to improve generalization under sparse training data conditions", Proc. IJCNN2001, Washington DC, pp.967-972, July, 2001.

[15] K.Inagaki and K.Nakayama, "Classification of mental tasks based on brain waves by using neural network (in Japanese)", IEICE, Technical Report, Vol.105 No.174 pp.25-30, SIP2005-54, 2005.07.

[16] K.Inagaki and K.Nakayama, "On brain computer interface by using multilayer neural network (in Japanese)", IEICE, 20th SIP Symposium, Kouchi, 2005.11.

[17] K.Nakayama and K.Inagaki, "A brain computer interface based on neural network with efficient pre-processing", Proc. IEEE, ISPACS2006, Yonago, Japan, pp.673-676, Dec. 2006.