We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Mental Task Recognition by EEG Signals: A Novel Approach with ROC Analysis

Takashi Kuremoto, Masanao Obayashi,
Shingo Mabu and Kunikazu Kobayashi

Additional information is available at the end of the chapter

Abstract

Electroencephalogram or electroencephalography (EEG) has been widely used in medical fields and recently in cognitive science and brain-computer interface (BCI) research. To distinguish metal tasks such as reading, calculation, motor imagery, etc., it is generally to extract features of EEG signals by dimensionality reduction methods such as principle component analysis (PCA), linear determinant analysis (LDA), common spatial pattern (CSP), and so on for classifiers, for example, k-nearest neighbor method (kNN), kernel support vector machine (SVM), and artificial neural networks (ANN). In this chapter, a novel approach of feature extraction of EEG signals with receiver operating characteristic (ROC) analysis is introduced.

Keywords: brain-computer interface (BCI), electroencephalogram or electroencephalography (EEG), artificial neural networks (ANN), support vector machine (SVM), receiver operating characteristic (ROC), Fourier transformation (FT)

1. Introduction

The electrical activity of the brain can be measured by electrodes placed on the scalp and the observed signal is called electroencephalogram or electroencephalography (EEG). EEG is also called "brain wave" and it has been widely used in clinical diagnose of brain disease since the early time of last century [1].

Different mental tasks yield EEG signals in different patterns in the different observation values. For example, in the case of human brain, the resting state (relax state), the most prominent power spectra are 8–15 Hz EEG signals (so-called "alpha-wave") observed in posterior sites,

meanwhile, 16–31 Hz signals (beta-wave) appears in the mental tasks such as active thinking, high alert, anxious, etc. Gamma-wave, EEG with higher than 32 Hz, displays during cross-modal sensory processing such as combining the stimuli of visual and auditory. On the other hand, the location of electrodes on scalp records different EEG signals spatially, and they are called EEG signals in different "channels". The allocation of electrodes is usually with the international 10–20 system. The name of 10–20 system comes from those adjacent electrodes that are allocated in distances of 10 or 20% of the total front-back or right-left of skull. More channels, more spatial features, may result in higher recognition rate of mental tasks. On the other hand, few channels give lower computational cost in the EEG classification systems.

In last decades, EEG has been utilized in the field of the brain-computer interface (BCI) for its ability of the mental task recognition [2–6]. Mental tasks indicate the state of activity of the brain with some specific tasks. For example, imagining writing a letter, counting, calculating, or raising a hand, a leg, etc. There are many classifiers for EEG recognition that have been proposed such as linear discriminant analysis (LDA), support vector machine (SVM), artificial neural networks (ANN), fuzzy inference systems, Bayesian graphical network (BGN), and so on. However, for the reasons of the complex nature of EEG signals, for example, noise and outliers, nonstationarity, high dimensionality, individual difference, etc., the pattern recognition (classification) problem of EEG signals is still a high hurdle for BCI realization.

To normalize the raw EEG signals, Nakayama and Inagaki proposed to reduce the number of the time series data of power spectrum of frequency given by fast Fourier transformation (FFT) with average values and normalize the FFT by a nonlinear normalization function [4]. To extract discriminant features of EEG signals for mental task recognition, Li and Zhang proposed a regularized tensor discriminative feature space, which includes multichannels, power spectrum of frequency, and those data in time series: channel × frequency × time [5]. Obayashi et al. applied Nakayama and Inagaki's pre-processing method to their practical EEG recognition system with single channel information in [6]. In [7], Jrad and Congedo used spatially weighted SVM (swSVM) to build a spatial filter for each temple feature. In the previous works of authors [8], discriminant temporal frequency data were utilized to reduce the flattening of different EEG patterns adopting the pre-processing method of [4], temporal spatial frequency concept, and average moving processing of [7] were adopted to obtain higher rate of mental task recognition.

Recently, we proposed to find the discriminant feature of temporal frequency by receiver operating characteristic (ROC) analysis in [9]. The discriminant feature of temporal frequency indicates the power spectra of FFT in an interval of time series of EEG data, which are higher relative to a mental task comparing with other intervals (windows). ROC analysis has been widely utilized in medical & diagnostic science [10, 11], microarray classification [12], and recently in EEG classification [13]. It is a stochastic criterion to classify two kinds of probability distributions and the details will be described in the next section.

In this chapter, discriminative feature extraction methods of EEG signals, which play an important role for classifiers, are discussed. Specially, an advanced temporal–spatial spectrum feature extraction method is introduced [9].

2. Discriminant feature extraction using ROC analysis

2.1. ROC analysis

Receiver operating characteristic (ROC) analysis was first used in radar signal detection in 1940s. The classification results of data in two kinds of distributions can be divided into four categories: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). A curve is plotted by the rate of TP against the rate of FP and it can be a measure of classification accuracy.

Now, let the TP of class A be in the shadow area α, and FP in area $1-\beta$, where β is the TP of class B (See **Figure 1**). When the dividing line between A and B is slid along x axis, a ROC curve is plotted indexing the divisibility of the two probability density functions (See **Figure 2**). If two distributions completely overlapped, $\alpha = 1-\beta$.

In **Figure 2**, the area below the ROC curve is called "area under the curve" (AUC). This value takes from 0.0 to 1.0, and it is an indicator of the divisibility of the two distributions. If the value of AUC becomes 0.5, two distributions are completely overlapped. Conversely, when the value of AUC reaches 1.0 (or 0.0), it means that the two distributions are completely separated.

In the practice procedure of ROC analysis, the area of α, that is, the rate of TP, and $1-\beta$, the rate of TN, can be calculated by the number of training samples, which are labeled data belonging to different classes.

2.2. Discriminant feature extraction of EEG signals

In [8], power spectrums of an interval of frequencies given by EEG signals FFT, which has a distinguish value to neighbors were used as discriminant features as the input vectors of classifiers. The flow chart of this method is depicted in **Figure 3**. **Algorithm I** shows the method in detail.

Algorithm I.

Step 1. Dividing (windowing) the original EEG signals into several intervals;

Step 2. Executing discrete Fourier transformation (DFT) in different intervals and normalizing the transformation results;

Step 3. Calculating the average power spectrum of banded (limited) frequencies in each phases;

Step 4. Finding a special (feature) interval, in which average power spectrum is the most different one from its neighborhoods;

Step 5. The power spectrum of FFT in the windowed frequencies and their average values are used as the feature data for classifiers.

A sample of the first processing (Step 1) is shown in **Figure 4**. In **Figure 4**, an EEG signal, which is a time series data (the potential of an electrode) of one channel, is divided into five intervals. DFT is executed in each interval at Step 2, and as a sample, the result of the second intervals (at time 30–60) is shown in **Figure 5**.
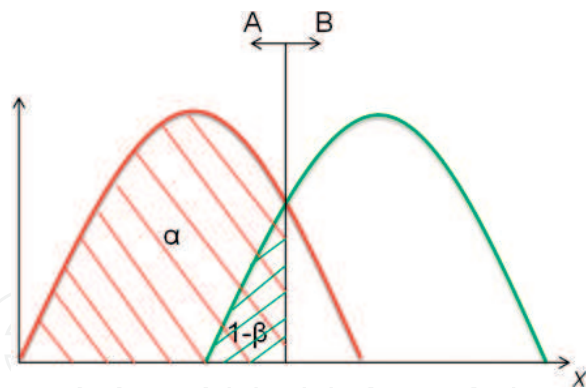
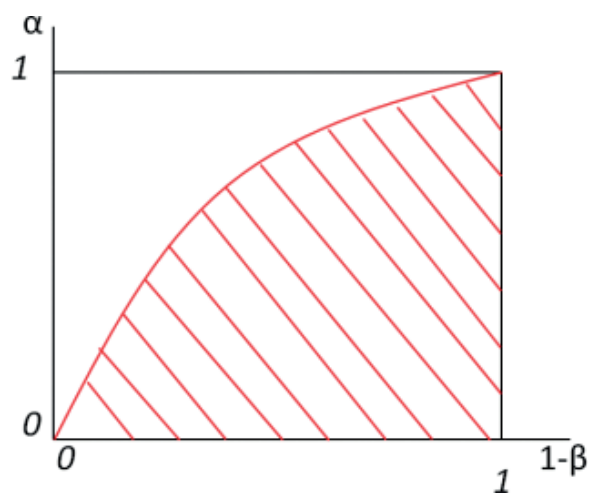**Figure 1.** Overlapping of the probabilities of two classes of data.

Figure 2. AUC of ROC curve.

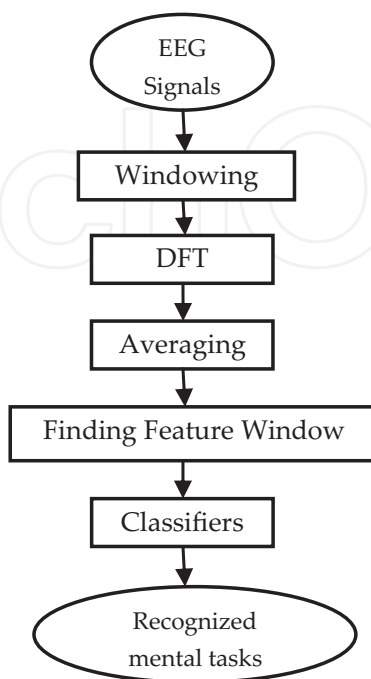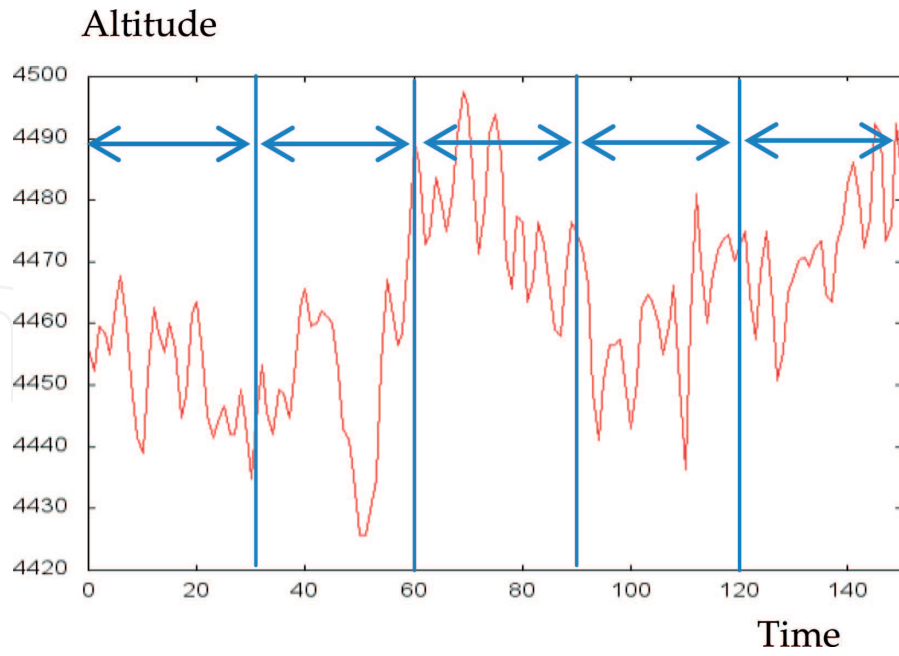**Figure 3.** Flow chart of EEG signal recognition in [8].

Figure 4. A sample of Step 1 processing: dividing EEG signals into several intervals.
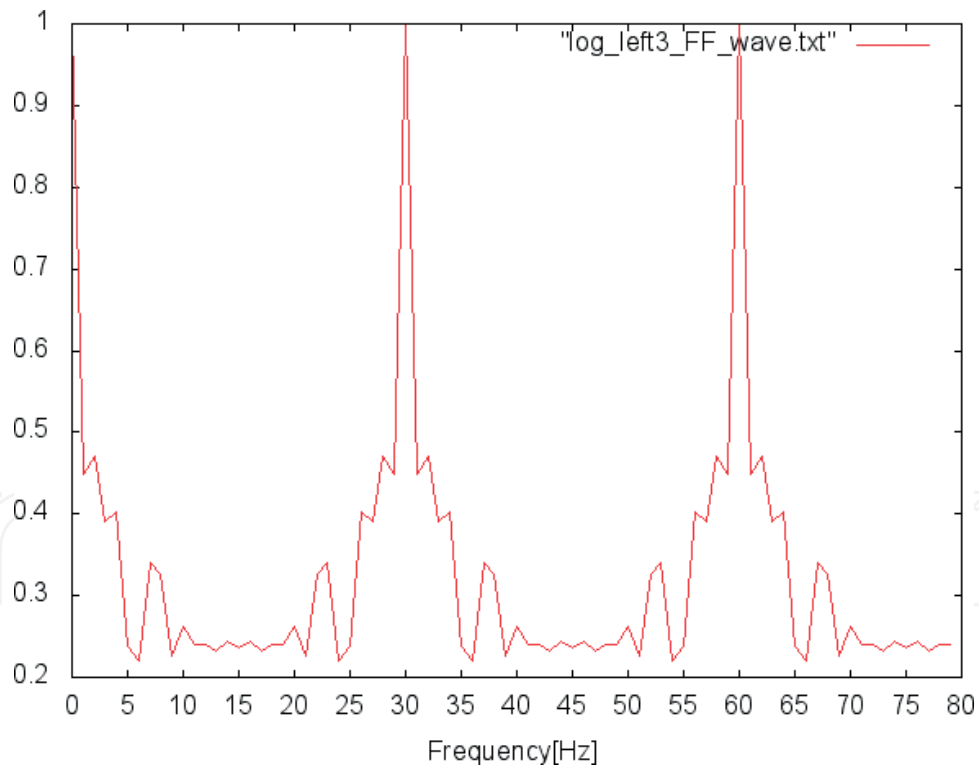

Figure 5. A sample of Step 2 processing: DFT and normalizing results of an interval of EEG time series data.

The normalization of DFT results is given by a nonlinear function [4].

$$x(n) = \frac{\log(x(n) - \max(x(n)) + 1)}{\log(\max(x(n)) - \min(x(n)) + 1)} \tag{1}$$

where $x(n)$ is the original DFT power spectrum of frequency n.

This nonlinear normalization reduces the vibration of time series of DFT results, avoiding the overfitting when classifiers are designed.

A frequency interval, which has distinguished power spectra for a certain mental task is chosen by Eq. (2).

$$\underset{p}{\arg\max}\, L(p) = \sum_{h=h_{low}}^{h=h_{up}} \mid F_{(p+1)h} - F_{ph} \mid \tag{2}$$

where p = 1, 2, …, P is the number of intervals, F_{ph} is the power spectrum on the frequency, $h = h_{low,}\ h_{low+1},\ …,\ h_{up}$ is the frequency, h_{low} and h_{up} are bands of feature frequencies of mental tasks and they were 4 and 45 Hz, respectively in our experiments.

For ROC analysis, it gives a measure of the difference between two probability distributions, it is validly used to find the discriminant features for EEG signal classification. In [13], Nguyen et al. utilized the AUC of ROC curve to select the elite wavelet coefficients, and in [9], we adopted an algorithm that using high AUC values to select metal task-related frequencies of EEG signals in different channels, respectively, and using the power spectrum of these frequencies as discriminant features for various classifiers such as SVM, ANN [including multi-layer perceptron (MLP), and deep neural networks (DNN)], k-nearest neighbor, decision tree (DT), and so on. The discriminant feature extraction method using ROC analysis is given by **Algorithm II**.

Algorithm II.

Let the input signals be $x_{kc,m,n}$ (c = 1 or 2, k = 1, 2, …, K), where k indicates the kth EEG signal of a set of EEG data, and c indicates the class of mental task, m indicates the channel number, and n is the time of signal.

   Step 1. Perform FFT to all the EEG signals $x_{kc,m,n}$ and let the result be power spectrum $E_{m,p}$ (p = 1, 2, …, P) corresponding to frequency $F_{kc,m,p,}$ where p indicates the order number of frequencies.

   Step 2. Obtain $P_{k1,m,p}$ and $P_{k2,m,p'}$ which are two probability density functions of $F_{kc,m,p}$ at p frequency, where class c = 1 and 2 of K signals of channel m.

   Step 3. Calculate the ROC curve and its AUC $A_{m,p}$ of $P_{k1,m,p}$ and $P_{k2,m,p}$.

   Step 4. Repeat Step 2 and Step 3 on all channels, a set $AUC_{m,n}$ of frequency p in channel m is obtained.

   Step 5. Find P points of frequencies, in which $A_{m,p}$ is high.

   Step 6. Power spectrum $E_{m,p}$(p = 1, 2, …, P) of the unknown EEG signal are used as input feature vector of a classifier.

The main difference between **Algorithm I** and **II** is that the power spectra in a special interval of frequencies, which is mostly related to an event of brain activity, are chosen in the former, meanwhile the power spectra of special frequencies chosen by high AUC of ROC are chosen as discriminant features in the later algorithm. The flow chart of EEG signal classification using ROC analysis is depicted in **Figure 6**.
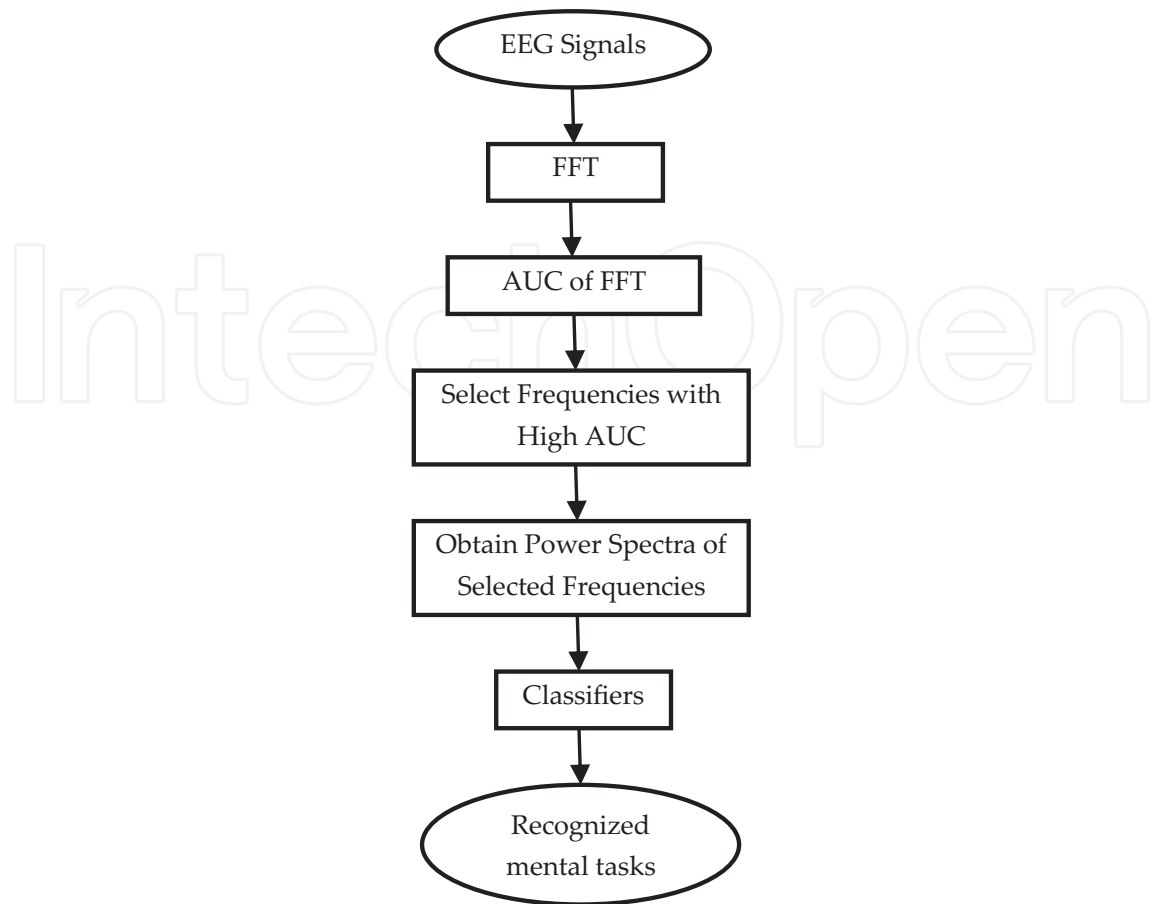
Figure 6. Flow chart of the EEG signal classification using ROC analysis.

Figures 7–9 showed a sample of the processing. In **Figure 7**, a raw EEG signal and its FFT result are shown. Note that the number of horizontal axis indicates the order of frequencies, and the value of vertical axis is the power spectrum. In **Figure 8**, the distribution of the power
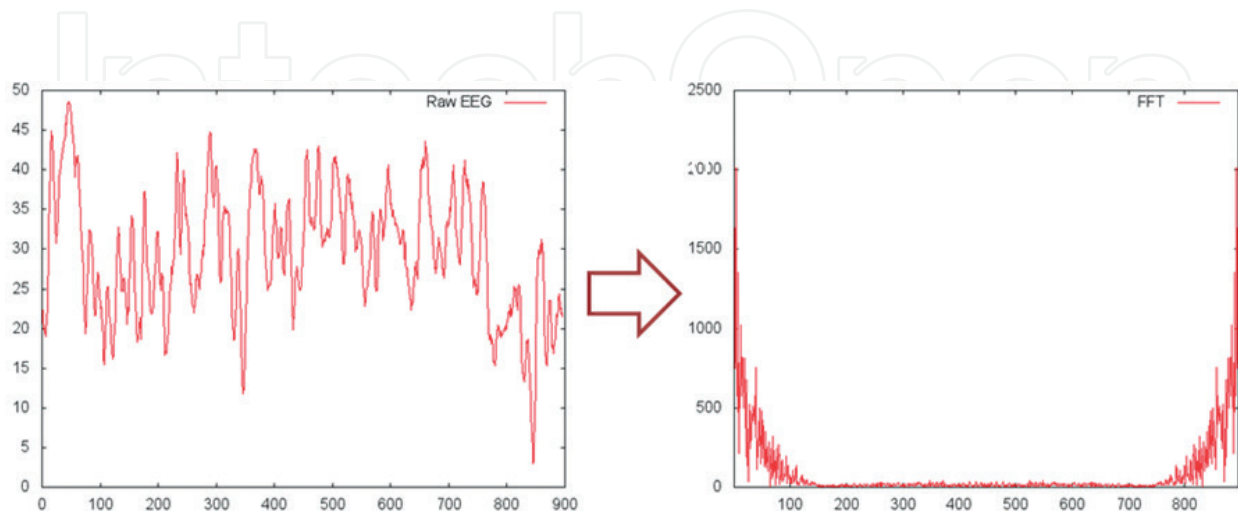

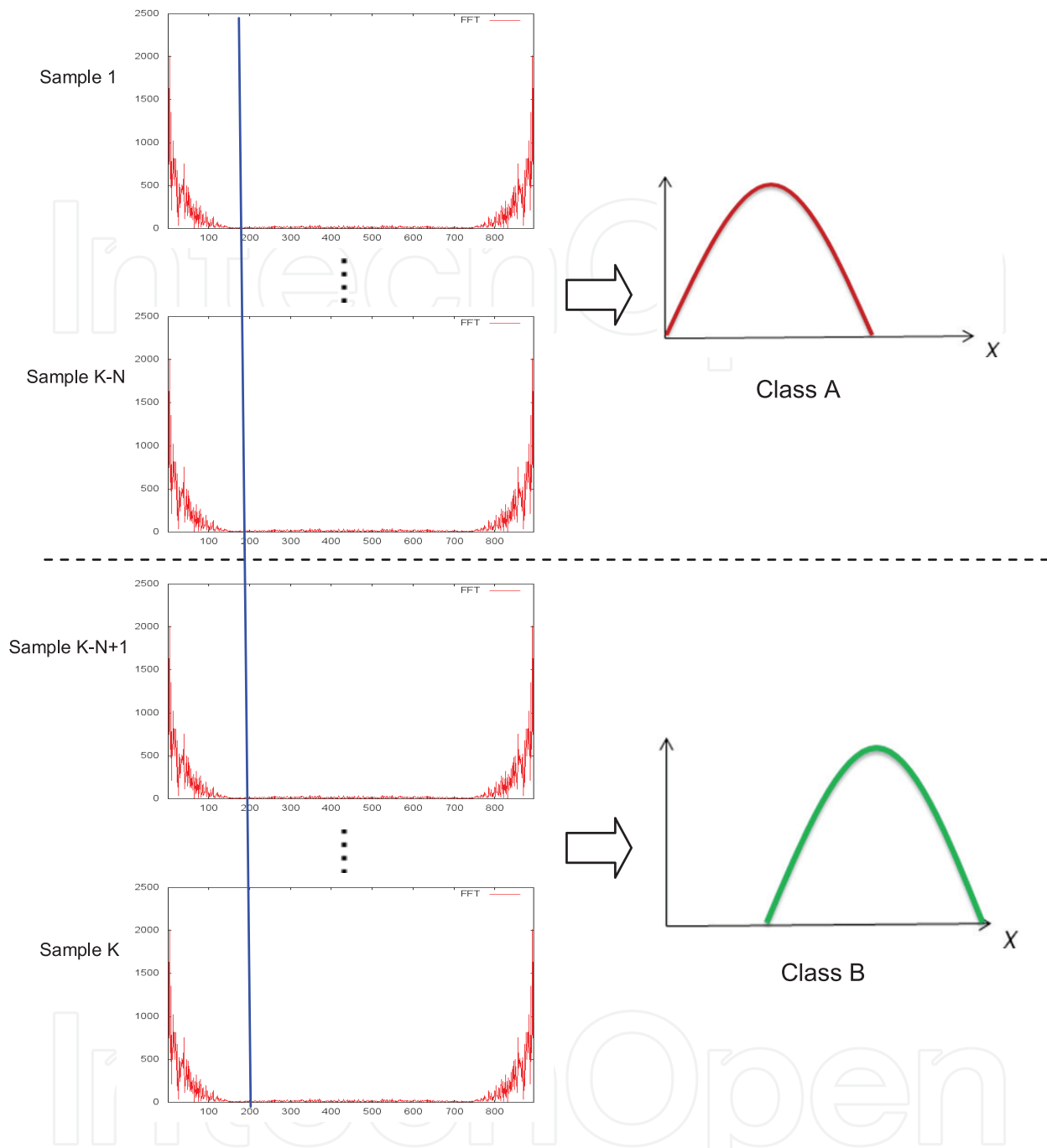**Figure 7.** A raw EEG signal (left) and its FFT results (right).

Figure 8. Calculation of the power spectra distribution (histogram) of each frequency of two classes of EEG signals. Frequency 200 (series number) is illustrated as a sample here.

spectrum of each frequency is calculated using the labeled samples. For example, there are K samples including N samples of class A and $K-N$ samples of class B as shown in **Figure 8**. AUC of the power spectra on each frequency is shown in **Figure 9**. Additionally, frequencies with high AUC extracted by a threshold line are used as criteria of discriminant feature selection. For example, in the case of three input dimensions for a classifier, the input vector is the power spectra with high AUC of frequencies as shown in **Table 1**.
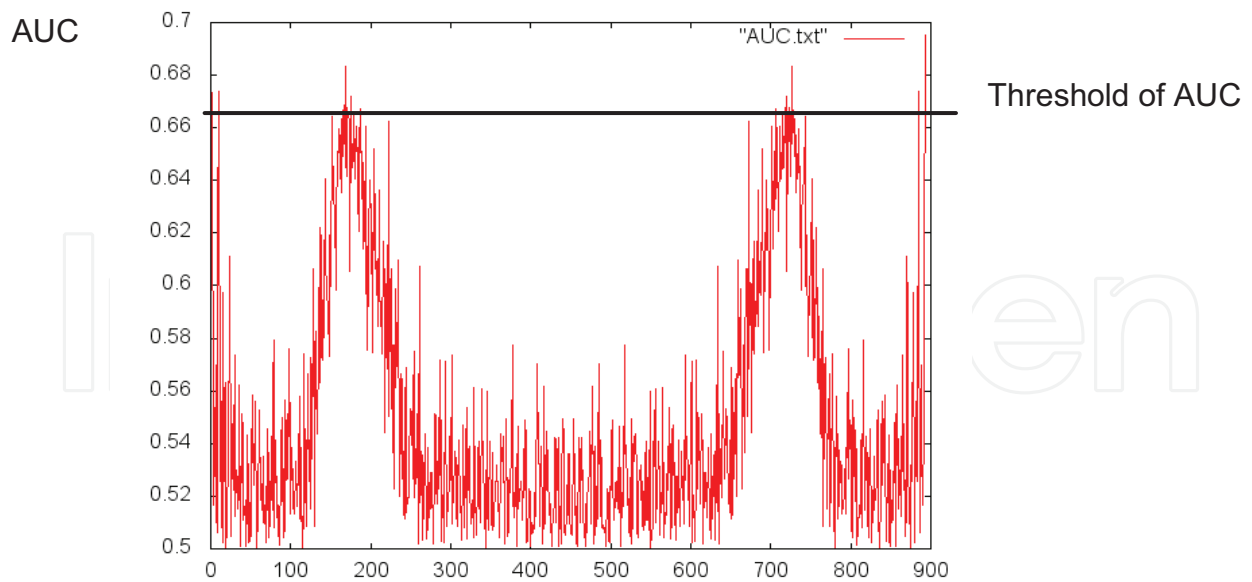
Figure 9. AUC of the power spectra on each frequency of two classes of EEG signals.

Ordered AUC	Number of freq.	Power spectrum (input of classifiers)
0.695	896	1545.186
0.691	158	10.093
0.688	726	9.535

Table 1. A sample of discriminant features extracted by ROC analysis.

3. Experiments

To compare the performance of different feature extract methods for EEG signal classification, experiments with two kinds of EEG data were performed [9]. One was a benchmark data set given by Brain-Computer Interfaces Laboratory, Colorado State University [14, 15], and another was from BCI competition II [16]. Classifiers used in the comparison experiments for different feature extraction methods were kernel SVM, MLP, kNN, deep neural network (DNN), and DT, in which source coded are in a software package R [17] as shown in **Table 2**.

The evaluation of the performance of different feature extraction methods uses the accuracy of classification, which is given by Eq. (3).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Name	Function
ROCR	ROC analysis/AUC calculation
Kernlab	Support vector machine (kernel SVM)
nnet	Neural network (MLP)
class	k-nearest neighbor (kNN)
h2o(+JavaVM)	deep neural network (DNN)
rpart	decision tree (DT)

Table 2. Software R [17] and its function used in the experiment.

4. Benchmark data and experiment results

Open access free website of BCI laboratory of Colorado State University [14, 15] provides Benchmark EEG data with five kinds of mental tasks as shown in **Table 3**. The data were measured by six channels with EEG sensors (See **Figure 10**) and one channel data of an EOG sensor (to measure the movement of an eye). The sampling rate is 250 Hz, and EEG data are recorded in 10 seconds, that is, 2500 time series data obtained by one trial. EEG signals of each mental task are recorded in 10 trials of five subjects. For the ROC analysis classifies two classes data, "Baseline" (relaxing state) and "Multiplication" (Multiplication calculation mentally) data, were used in our experiment. Additionally, training samples and testing samples used EEG data of the same subject, which were chosen randomly with a ratio of 15:5.

The classification accuracies of **Algorithm I** [8], and **Algorithm II** [9] by different classifiers are shown in **Table 4**. In **Table 4**, it is also shown that different dimensionalities of the input vector influenced the classification accuracy. Feature extraction method using **Algorithm II**. (FFT and ROC analysis) had a prior performance especially in the case of 140-dimension input vector. The highest classification accuracy 97.5% was given by kernel SVM classifier, and DNN stood the second position with 95.37% using **Algorithm II** feature extraction method, respectively.

4.1. BCI competition II data and experiment results

BCI competition II data [16] were also used in the performance comparison of different feature extraction methods. There are two-class data named "Ia" and "Ib," which are EEG data obtained

Mental task	Contents
Baseline	Relaxing as much as possible
Multiplication	Calculating multiplication mentally.
Letter-composing	Considering the contents of a letter
Rotation	Imagining rotation of a 3-D object
Counting	Imagining writing a number in order

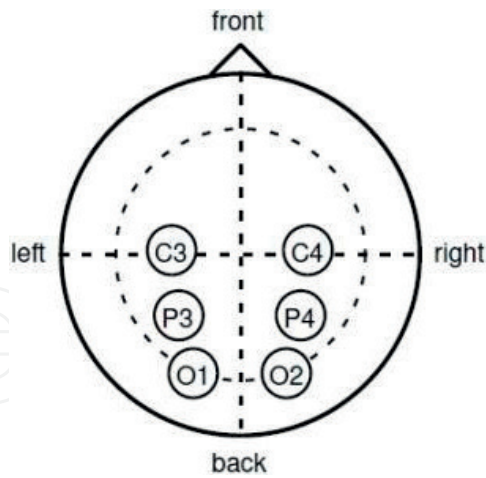**Table 3.** Mental tasks in a benchmark database [14, 15].

Figure 10. Positions of EEG sensors with six channels [8].

Classifier	Feature extraction method			
	Algorithm I (Temporal FFT)		Algorithm II (FFT and ROC analysis)	
	140-D	1120-D	140-D	1120-D
Kernel SVM	59.58	70	**97.5**	75
MLP	49.58	38.33	**55.0**	52.92
k-Nearest neighbor	55.92	66.67	**73.33**	66.03
Deep neural network	61.67	71.67	**95.37**	94.58
Decision tree	34.5	35.5	**50.0**	**50.0**

Unit: %.
The bold values indicate the best recognition result between different feature extraction algorithms for one classifier in the case of benchmark data.

Table 4. Classification results of benchmark data [14, 15].

by a healthy subject and an amyotrophic lateral sclerosis (ALS) patient. In each data set, two kinds of mental tasks were required, respectively. One was to move a cursor up (class A) and another was to move the cursor down (class B). Details of these EEG data descriptions are shown in **Table 5**. Additionally, training samples and testing samples were chosen randomly with a ratio of 240:28 for Ia and 180:20 for Ib.

The accuracies of classification of Ia and Ib by different feature extraction methods and classifiers are shown in **Tables 6** and **7**, respectively. **Algorithm II** (FFT and ROC analysis) showed the highest classifications for all classifiers. The highest accuracy for data Ia was 91.23%, given by

Data set	Mental tasks	Trials	Channels	Samples/Ch.	Sampling freq.
Ia	2	135/133	6	896	256
Ib	2	100/100	7	1152	256

Table 5. Description of EEG data of BCI competition II [16].

Classifier	Feature extraction method			
	Algorithm I (Temporal FFT)		Algorithm II (FFT and ROC analysis)	
	140-D	1120-D	140-D	1120-D
Kernel SVM	61.10	58.98	87.04	**91.23**
MLP	49.09	49.95	68.06	**70.86**
k-Nearest neighbor	50.55	55.46	**79.17**	55.76
Deep neural network	57.72	62.08	83.48	**86.10**
Decision tree	41.79	43.15	67.5	**73.22**

Unit: %.

The bold values indicate the best recognition result between different feature extraction algorithms for one classifier in the case of data Ia.

Table 6. Classification results of BCI competition II data Ia [16].

Classifier	Feature extraction method			
	Algorithm I (Temporal FFT)		Algorithm II (FFT and ROC analysis)	
	140-D	1120-D	140-D	1120-D
Kernel SVM	52.99	53.91	76.16	**77.65**
MLP	46.40	53.36	**58.15**	49.09
k-Nearest neighbor	45.90	48.25	**60.04**	57.81
Deep neural network	43.90	49.25	69.93	**75.25**
Decision tree	28.43	47.99	45.44	**55.55**

Unit: %.

The bold values indicate the best recognition result between different feature extraction algorithms for one classifier in the case of data Ib.

Table 7. Classification results of BCI competition II data Ib [16].

kernel SVM using 1120 dimensions of input vector, which were discriminant features extracted by **Algorithm II**, and the same methods yielded the highest classification rate 77.65% for data Ib. These accuracies are higher than best classification rates 90.10 and 56.67%, which are the results of a state-of-the-art method of EEG signal recognition [13]. The future work of the improvement of **Algorithm II** is to find the optimal dimensionality of the discriminant feature space. It is hard to consider higher dimensionality results higher classification accuracy as shown in these experiments. It was better to choose 140-D in the case of benchmark data (**Table 4**), and oppositely, 1120-D was more suitable for BCI competition II data (**Tables 6** and **7**).

5. Conclusion

To recognize the mental tasks by EEG signals, two kinds of temporal-spatial frequency–based feature extraction methods were introduced in this chapter. In **Algorithm I**, event-related

intervals of the raw EEG time series data (temporal information) was extracted at first, and the averaged power spectra of frequencies given by FFT within the interval (frequency information) were used as the discriminant features. In **Algorithm II**, event-related frequencies of EEG's FFT were extracted by ROC analysis with high AUCs. The input space for classifiers was composed by all features extracted by two algorithms from multiple channels, so the spatial information was also included in these feature extraction methods.

Pattern recognition of EEG signals has been studied for decades, and it plays an important role in the field of human robot interaction (HRI). So, we expect that the feature extraction methods introduced in this chapter can be adopted in the real HRI systems in the near future.

Acknowledgements

Author details

Takashi Kuremoto[1]*, Masanao Obayashi[1], Shingo Mabu[1] and Kunikazu Kobayashi[2]

*Address all correspondence to: wu@yamaguchi-u.ac.jp

1 Yamaguchi University, Japan

2 Aichi Prefectural University, Japan

References

[1] Malmivuo J, Plonsey R. Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields. Oxford University Press, Oxford; 1995. http://www.bem.fi/book/

[2] Lotte F, Congedo M, Lecuyer A, Lamarche F, Arnaldi B. A review of classification algorithms for EEG-based brain-computer interfaces. Journal of Neural Engineering. 2007;**4**: 24-48

[3] Cheng SY, Hsu HT. Mental Fatigue Measurement Using EEG, Risk Management Trends. In: Nota G, editor. InTech, Rijeka, Croatia; 2011. pp. 203-228

[4] NakayamaK, Inagaki K. A brain computer interface based on neural network with efficient pre-processing. In: Proceedings of 2006 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2006). 2006. pp. 673-676

[5] Li J, Zhang L. Regularized tensor discriminant analysis for single trial EEG classification in BCI. Pattern Recognition Letters. 2010;**31**:619-628

[6]   Obayashi M, Watanabe K, Kuremoto T, Kobayashi K. Development of a brain computer interface using inexpensive commercial EEG sensor with one-channel. In: Proceedings of the 17th International Symposium on Artificial Life and Robotics (ISAROB). 2012. pp. 714-717

[7]   Jrad N, Conedo M. Identification of spatial and temporal features of EEG. Neurocomputing. 2012;**90**:66-71

[8]   Kuremoto T, Baba Y, Obayashi M, Mabu S, Kobayashi K. To extraction the feature of EEG signals for mental task recognition. In: Proceedings of 54th Annual Conference of the SICE. 2015. pp. 353-358

[9]   Kuremoto T, Baba Y, Obayashi M, Mabu S, Kobayashi K. A method of feature extraction for EEG signals recognition using ROC curve. In: Proceedings of 2017 International Conference on Artificial Life and Robotics. 2017. pp. 654-657

[10] Hanley JA, Mcneil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;**143**:29-36

[11] Mamitsuka H. Selecting features in microarray classification using ROC curves. Pattern Recognition. 2006;**39**(12):2393-2404

[12] Pereira P. Evaluation of rapid diagnostic test performance. In: Saxena SK, editor. Chapter 8. Proof and Concepts in Rapid Diagnostic Tests and Technologies. InTech, Rijeka, Croatia; 2016

[13] Nguyen T, Khosravi A, Creighton D, Nahavandi S. EEG signal classification for BCI applications by wavelets and interval type-2 fuzzy logic systems. Expert Systems with Applications. 2015;**42**:4370-4380

[14] Benchmark EEG Data: Brain-Computer Interfaces Laboratory, Colorado State University: http://www.cs.colostate.edu/eeg/main/data/1989_Keirn_and_Aunon

[15] Anderson CW, Sijercic Z. Classification of EEG signals from four subjects during five mental tasks. In: IEEE Proceeding on Engineering Application in Neural Network. 1997. pp. 407-414

[16] BCI Competition II: http://www.bbci.de/competition/ii/#datasets

[17] The R Project for Statistical Computing: https://www.r-project.org/