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

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

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

Optimal Fractal Feature and Neural Network: EEG Based BCI Applications

Montri Phothisonothai and Katsumi Watanabe

Additional information is available at the end of the chapter

1. Introduction

This chapter provides the theoretical principle of fractal dimension analysis applied for a brain-computer interface (BCI). Fractal geometry is a mathematical tool for dealing with complex systems. A method of estimating its dimension has been widely used to describe objects in space, since it has been found useful for the analysis of biological data. Moreover, fractal dimension (FD) is one of most popular fractal features. The term waveform applies to the shape of a wave, usually drawn as instantaneous values of a periodic quantity versus time. Any waveform is an infinite series of points. Aside of classical methods such as moment statistics and regression analysis, properties such as the Kolmogorov-Sinai entropy [1], the apparent entropy [2] and the FD [3] have been proposed to deal with the problem of pattern analysis of waveforms. The FD may convey information on spatial extent and self similarity and self affinity [4]. Unfortunately, although precise methods to determine the FD have already been proposed [5,6], their usefulness is severely limited since they are computer intensive and their evaluation is time consuming. Recently, the FD is relatively intensive to data scaling and shows a strong correlation with the human movement of EEG data [7-9]. The time series with fractal nature are to be describable by the functions of fractional Brownian motion (fBm), for which the FD can easily be set. Waveform FD values indicate the complexity of a pattern or the quantity of information embodied in a waveform pattern in terms of morphology, spectra, and variance. This chapter, we investigate most widely popular six algorithms to be feature patterns for a BCI in which algorithms are in time and frequency domain approaches. However, many methods in evaluating FDs have been developed. As the existing algorithms, they present different characteristics therefore the optimum condition respects to the requirements of BCI application should be seriously considered.

2. Experimental paradigm[1]

In this study, we tested 7-male and 3-female. Their ages were 21-32 years (mean age, 25.3 years; SD ±3.6), with 180 trials per subject (30 trails per task). At the beginning of each EEG recording session, the subjects were instructed to relax with their eyes open for 30 s. The EEG data from this period was used as the baseline for the tasks. The duration of one trial was 6 s throughout the experiment. Each trail in the experiment was divided into two periods, a relaxing period of 0-3 s and an imaging period of 3-6 s. When the subjects had their eyes open on a blank screen, they would be in a relaxed state of mind (relaxing period). In the imaging period, the subjects imagined with eyes open, as represented by arrows. This study focuses on the motor move-ment functions in the brain. Therefore, imagination of motors movements are used as the tasks throughout this experiment. To assist the subject imagine the task without difficulty, we provide the acting stimuli in accordance to the given tasks. Explicit indicators are also shown on a monitor for informing subject during the experiment, all selected tasks are listed in Table 1.

Task	Imagine	Indicator
Left-hand movement (LH-MI)	Opening and grasping left hand	⬅
Right-hand movement (RH-MI)	Opening and grasping right hand	➡
Feet movement (FT-MI)	Up-down lifting	⬇
Tongue movement (TG-MI)	Up-down movement	▬

Table 1. Selected tasks in this study

3. Fractal analysis and methods

In this chapter, most widely popular six algorithms in time and frequency domain analysis have been addressed; the box-counting method (BCM), Higuchi's method (HM), variance fractal method (VFD), detrended fluctuation analysis (DFA), power spectral density analysis (PSDA), and critical exponent method (CEM). To measure the fractality in short time intervals of time-sequential data from one end of the waveform to the other sequentially, the dynamical changes in the FDs with respect the time series based on the time-dependent fractal dimensions (TDFD) were observed. For the classification process, two new feature parameters on the basis of fractal dimension; Kullback–Leibler (K-L) divergence and the different expected values were presented. In experimental results, DFA was selected to evaluate fraction dimensions on the basis of TDFDs [11, 36]. The reason is that, we found that the DFA provides fast computation

---

1 By NLAB, under supervision of Prof.Masahiro Nakagawa, Department of Electrical Engieering, Nagaoka University of Technology, 1603-1 Kamitomioka, Nagaoka-shi 940-2188 Japan. E-mail: masanaka@vos.nagaokaut.ac.jp Http:// pelican.nagaokaut.ac.jp/

time and also presents reasonable values in terms of accuracy and variability when comparison with each other.

3.1. Time series-based analysis

3.1.1. Box-Counting Method (BCM)

One of the most common methods for calculating the FD of a self-similar fractal is the box-counting method (BCM). The definition of BCM is a bounded set in the Euclidian n -space that it composes self-similar property [25], $N_r \propto (1/r)^D$ by covering a structure with boxes of radii, r, the FD can be determined by

$$D_{BCM} = \lim_{r \to 0} \frac{\log_2 (N_r)}{\log_2 (1/r)} \tag{1}$$

We then repeat this process with several different radii. To implement this algorithm, number of contained box, n_r, is computed from the difference between the maximum and minimum amplitudes of the data divide by the changed radius, as

$$n_r(i) = \left\lfloor \frac{\max (x_r) - \min (x_r)}{r(i)} \right\rfloor, \quad \text{for } \{ r \in 2^k \mid k = 1,2, \ldots, \log_2 (L) - 1 \} \tag{2}$$

$$N_r = \sum_i n_r(i) \tag{3}$$

where N_r is a total number of contained boxes, x_r represents the EEG time series with length L, $r(i)$ is a radius by changing a step of k within the i-th subdivision window, and integer-part function denoted by $\lfloor \bullet \rfloor$. To obtain the FD, least-square linear fitted line corresponds to the slope of the plot $\log_2 (N_r)$ versus $\log_2 (1/r)$ is applied.

3.1.2. Variance Fractal Dimension (VFD)

This method is determined by the Hurst exponent, H, whose calculation was divided from the properties of the fBm data. The basic idea of calculation is based on the power law relationship between the variance of the amplitude increments of the input time series, which was produced by a dynamic process over time. The main advantage of VFD was its support of the real-time computation [17]. We also selected the VFD for estimating the FD of EEG data. The amplitude increments of a datum over a time interval Δt adhere to the following power law relationship $\text{Var}[x(t_2-t_1)] \propto |t_2-t_1|^{2H}$, the Hurst exponent can be calculated by using a log-log plot then given by

$$H = \lim_{\Delta t \to 0} \left(\frac{1}{2} \frac{\log_2 \left(\text{Var}[(\Delta x)_{\Delta t}] \right)}{\log_2 (\Delta t)} \right) \tag{4}$$

The variance in each window per stage k can be calculated as follows

$$\text{Var}[\Delta x_{\Delta t}] = \frac{1}{(N_k - 1)} \left[\sum_{j=1}^{N_k} (\Delta x)_{jk}^2 - \frac{1}{N_k} \left(\sum_{j=1}^{N_k} (\Delta x)_{jk} \right)^2 \right] \qquad (5)$$

$$(\Delta x)_{jk} = x(jn_k) - x((j-1)n_k), \quad \text{for } j = 1, 2, \ldots, N_k \qquad (6)$$

The least-square linear fitted line corresponds to the slope of the plot $\log_2(n_k)$ and $\log_2(\text{Var}[\Delta x]_k)$, the Hurst exponent is computed as $H = 0.5s$ where s is the obtained slope then the FD can be estimated as

$$D_{VFD} = 2 - H \qquad (7)$$

The process of calculating the FD essentially involves segmenting the entire input time series data into numerous subsequence (or window). The values k represents the integer range chosen such that each window of size N_T contains a number $n_k = 2^k$ of smaller windows of size $N_k = \lfloor N_T / n_k \rfloor$.

3.1.3. Higuchi's Method (HM)

FD is another measure of data complexity, generally evaluated in phase space by means of correlation dimension. Higuchi proposed an algorithm for the estimation of FD directly in time domain without reconstructing the strange attractor [18]. The HM also gives reasonable estimate of the FD in the case of short time segment. The HM algorithm based on the given finite time series $y = \{y(1), y(2), \ldots, y(N)\}$, a new time series, y_m^k, are constructed by the following equation

$$y_m^k = \left\{ y(m), \ y(m+k), \ y(m+2k), \ \ldots, \ y\left(m + \lfloor \frac{N-m}{k} \rfloor \right) \bullet k \right\} \text{ for } m = 1, 2, \ldots, k \qquad (8)$$

where both m and k are integers and they indicate the initial time and the time interval, respectively. The length, $L_m(k)$, of each curves is computed as

$$L_m(k) = \frac{1}{k} \left\{ \left(\left[\sum_{i=1}^{\lfloor \frac{N-m}{k} \rfloor} | y(m+ik) - y(m+(i-1)k) | \right) \bullet \frac{N-1}{\lfloor \frac{N-m}{k} \rfloor \bullet k} \right\} \qquad (9)$$

The length of curve for time interval k, $L_m(k)$ is computed as the average of the m curves. A relationship of this algorithm is $L_m(k) \propto k^{-D_{HM}}$ therefore we apply the least-squares fitting line of $\log_2(L(k))$ versus $\log_2(k)$, the negative slope of the obtained line is calculated giving the estimate of the FD, D_{HM}.

3.1.4. Detrended Fluctuation Analysis (DFA)

The idea of DFA was invented originally to investigate the long-range dependence in coding and noncoding DNA nucleotide sequence [19]. Due to simplicity in implementation, the DFA is now becoming the most important method in the field of fractal analysis [20]. Therefore the DFA method was also applied to FD estimation in this study.

$$X(k) = \sum_{i=1}^{k} \left[x(i) - \bar{x} \right]$$ (10)

This integrated series is divided into non-overlapping intervals of length n. In each interval, a least squares line is fit to the data (representing the trend in the interval). The series $X(k)$ is then locally detrended by subtracting the theoretical values $X_n(k)$ given by the regression. For a given interval length n, the characteristic size of fluctuation for this integrated and detrended series is calculated by

$$F(n) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left[X(k) - X_n(k) \right]^2}$$ (11)

This computation is repeated over all possible interval lengths (in practice, the minimum length is around 10, and the maximum is a half-length of input data, giving two adjacent intervals). In this experiment, we use the power of 2 based length for input EEG data therefore in this case we can range $n = 2^k$ for $k = 4, 5, \ldots, \log_2(L) - 1$. A relationship is expected, as $F(n) \propto n^\alpha$ where α is expressed as the slope of a double logarithmic plot $\log_2(F(n))$ versus $\log_2(n)$. As PSD, Then α can be converted into the Hurst exponent $H = \alpha - 1$ and the estimated FD according as $D_{DFA} = 3 - \alpha$.

3.2. Frequency-based analysis

3.2.1. Power Spectral Density Analysis (PSDA)

This method is widely used for assessing the fractal properties of time series, the method based on frequency domain analysis, and also works on the basis of the 1/f-like power scaling which can be obtained by the fast Fourier transform (FFT) algorithm. It is called a power-law relationship of Mandelbrot and van Ness [25] can be expressed as follows

$$f(s) \propto \frac{1}{s^{-\beta}}$$ (12)

where s is the frequency, $f(s)$ is the correspondence power at a given frequency, and β is the scaling exponent which is estimated by calculating the negative slope of the linear fit relating $\log(f(s))$ to $\log(s)$. The PSDA was applied to raw EEG data to classify the EEG characteristics. This study used the enhanced version of PSDA that initially proposed by Fougère [26] and

modified by Eke [27]. To improve the stability of spectral estimates, raw EEG data of each individual trial were first detrending by subtracted the mean of the data series from each value, and then each value is multiplied by a parabolic window is given as below

$$\psi(i) = 1 - \left(\frac{2i}{M+1} - 1\right)^2, \qquad \text{for } i = 1, 2, \ldots, M \tag{13}$$

Finally the scaling exponent was estimated by the least-square fitting of log-log domain for the frequencies lower than 1/8 of the maximum sampling rate, i.e., less than 64 Hz in this study. The Hurst exponent can be determined as $H = (\beta - 1)/2$ where $1 < \beta < 3$ and the estimated FD can be computed by the following equation

$$D_{PSDA} = 2 + \frac{(1 - \beta)}{2} \tag{14}$$

3.2.2. Critical Exponent Method (CEM)

This method was initially proposed by Nakagawa [21]. The CEM has been applied in the physiological data analysis and featuring [22-24]. The PSD of observed fBm data, $P_H(v)$, in the frequency domain is determined as

$$P_H(v) \propto v^{2H+1} = v^{-\beta} \tag{15}$$

The moment, I_α, and the moment exponent, α, of the PSD is determined as

$$I_\alpha = \int_1^\Omega P_H(v) v^\alpha dv, \qquad \left(-\infty < \alpha < +\infty\right) \tag{16}$$

We will consider the frequency bands as finite value and substitute Eq. (15) to Eq. (16) thus the equation was given as

$$I_\alpha \propto \int_1^\Omega v^{\alpha-\beta} dv = \int_1^\Omega v^{A-1} dv = \frac{1}{A}(\Omega^A - 1) = \frac{2}{A} \exp\left(\frac{uA}{2}\right) \sinh\left(\frac{uA}{2}\right) \tag{17}$$

where $A = \alpha - \beta + 1$, $u = \log(\Omega)$, and Ω is the upper frequency variable which was normalized to the lower bound of the integration region as 1. Hence taking the logarithm of moment and the third order derivative can be written as

$$\frac{d^3 \log I_\alpha}{d\alpha^3} = \frac{2}{8} u^3 \operatorname{cosech}^3\left(\frac{uA}{2}\right) \cosh\left(\frac{uA}{2}\right) - \frac{1}{A^3} \tag{18}$$

We then determine the critical value $\alpha = \alpha_c$ which satisfies for Eq. (18) is equal to zero. Finally, the FD can be estimated by

$$D_{CEM} = 2 - \frac{\alpha_c}{2} \tag{19}$$

The PSDA and CEM used the FFT algorithm for transforming the EEG time series data into frequency components.

4. Performance assessment

Since high computation time and low variability are the requirement of the BCI system, the optimal algorithm for evaluating fractal dimension should be examined carefully. In this experiment, we selected two signals that fractal dimension value can be easily set for assessing. The usefulness of output results can be helped us know which algorithm is suited for applying it as the feature extractor.

4.1. Assessing with artificially generated signals

1. Fractional Brownian motion (fBm) signal[2]. General model of the fBm signal can be written as the following fractional differential equation [32]

$$\frac{\mathrm{d}^{H+1/2} B_H(t)}{\mathrm{d}t^{H+1/2}} \propto w(t) \tag{20}$$

The mathematical model for solving the fBm signal in Eq.(20) was proposed by Mandelbrot and Van Ness [25], which is defined by

$$B_H(t) = \frac{1}{\Gamma(H+1/2)} \left(\int_{-\infty}^{0} [(t-s)^{H-1/2} - (-s)^{H-1/2}] dB(s) + \int_{0}^{t} [(t-s)^{H-1/2}] dB(s) \right) \tag{21}$$

where $\Gamma(\bullet)$ is the Gamma function $\Gamma(a) = \int_{0}^{+\infty} x^{a-1} e^{-x} dx$ and H is the Hurst exponent where $0 < H < 1$. B is a stochastic process. To simulate the fBm signal, we can simply implement the above definition to discrete time series at n-th index by Riemann's type sums as follows:

$$B_H(n) = \frac{1}{\Gamma(h+1/2)} \left(\sum_{k=-b}^{0} \left[(n-k)^{H-\frac{1}{2}} - (-k)^{H-\frac{1}{2}} \right] G_1(k) + \sum_{k=0}^{n} \left[(n-k)^{H-\frac{1}{2}} \right] G_2(k) \right) \tag{22}$$

In case $H = 1/2$, Eq.(4.22) can be reduced, we then obtain

$$B_{1/2}(n) = \sum_{k=0}^{n} G(k) \tag{23}$$

---

2 fBm data were simulated by NLAB, Department of Electrical Engieering, Nagaoka University of Technology, 1603-1 Kamitomioka, Nagaoka-shi 940-2188 Japan.

For $n = 1, 2, \ldots, N$ and $b = \sqrt{N^3}$. G is a normal distribution space.

2. Synthetic (Syn) signal. This signal was produced using the deterministic Weierstrauss cosine function [28]. To simulate the Syn signal, we can simply implement the above definition to discrete time series at n-th index by [33]

$$W_h(n) = \sum_{i=0}^{M} \gamma^{-iH} cos\left(2\pi\gamma^i n\right) \tag{24}$$

where $\gamma > 1$, and we set $M = 26$ and $\gamma = 5$. The fractal dimension of this signal is given by $D = 2 - H$.

4.2. Conditions for assessing

A summary of the selected parameter values in this experiment is shown in Table 2. The selection of signal lengths that are powers of two was motivated by the requirements of frequency-based algorithms (PSDA and CEM). Output performances of each algorithm were shown in Fig. 1.

4.3. Results of performance assessing

In order to test the effect of signal lengths on H estimation and each algorithm was applied on the entire series. The estimated fractal dimension values are then averaged at the change of H. Before utilizing those of algorithms, we should regard as the following things:

1. *Computation time*. Short computation time is the requirement of the BCI applications. The fBm and Syn signals are performed to complete the entire series on a laptop 1.6 GHz Pentium with memory of 512 MB. Table 3 shows the results of the average computation time.

2. *Accuracy*. To compare with the theoretical FD value (true FD value), we used a mean-squared error (MSE) value, which can be defined by

$$E_D = \frac{1}{L}\sum_{i=1}^{L}\left(D_i - \hat{D}_i\right)^2 \tag{25}$$

where D is a theoretical FD value (true value), $\hat{D}$ is an estimated FD value at ith step of H, and L is a signal length. Results are shown in Tables 4 and 5.

1. *Variability (or robustness)*. The standard deviation (SD) value is computed to indicate that that variability of algorithms. Results are shown in Tables 6 and 7.

Algorithm	Parameter	Value
BCM	Step size of radius	$k = 1, 2, \ldots, \log_2(p) - 1$
VFD	Step size of window	$k = 2, 3, \ldots, \log_2(p) - 1$

Algorithm	Parameter	Value
HM	Maximum interval length	$k_{max} = p - 3$
DFA	Step size of interval length	$k = 4, 5, \ldots, \log_2(p) - 1$
PSDA	FFT point; Maximum frequency	$N_{FFT} = p; f_{max} = 64\,\text{Hz}$
CEM	FFT point; Step size of moment exponent	$N_{FFT} = p; a_\Delta = 0.001$

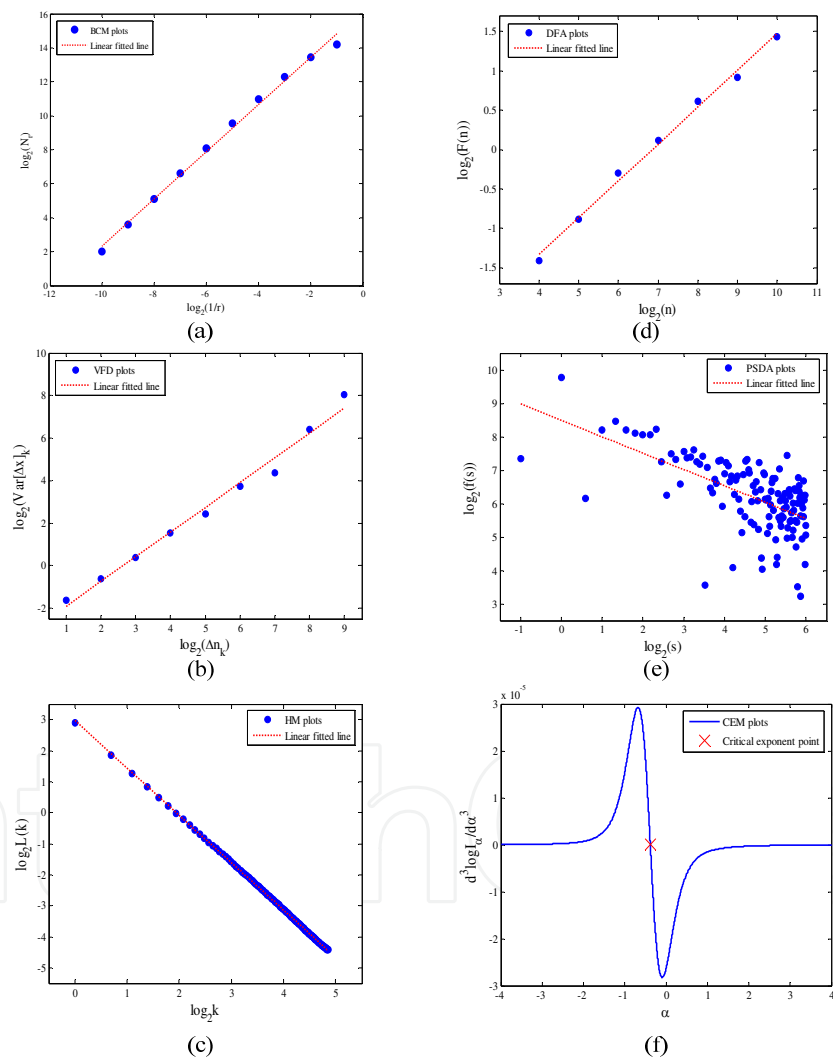**Table 2.** Selected parameters in the experiment.

Figure 1. The results for fractal analysis, performed on a series of test EEG data. (a) BCM algorithm: a log-log plot of a total number of contained boxes versus radius. (b) VFD algorithm: a log-log plot of variance versus window length. (c) HM algorithm: a log-log plot of a curve's length versus interval length. (d) DFA algorithm: a log-log plot of a total length F(n) versus interval length. (e) PSDA algorithm: a log-log plot of power spectra versus frequency. (f) CEM algorithm: a log-log plot of moment function versus moment exponent value [16].

Signal length	t_{avr} [ms]					
BCM	VFD	HM	DFA	PSDA	CEM	
$L = 2^7$	2.60	**0.77**	14.90	10.05	2.07	140.40
$L = 2^8$	3.50	**1.10**	148.60	20.20	2.90	227.50
$L = 2^9$	6.80	**2.80**	385.20	40.20	4.90	1,767.80
$L = 2^{10}$	12.60	**4.30**	697.90	82.50	10.80	3,351.20
$L = 2^{11}$	26.00	**7.10**	1,400.70	168.80	20.30	6,538.50
$L = 2^{12}$	54.00	**12.40**	2,490.50	329.70	46.40	12,964.80
$L = 2^{13}$	114.90	**61.10**	4,160.70	895.10	137.80	25,318.10
$L = 2^{14}$	265.50	**164.70**	14,630.80	2,971.50	478.70	50,157.60
$L = 2^{15}$	676.10	**607.90**	47,082.50	11,210.20	1,882.30	101,809.70
Average	129.11	**95.79**	7,890.20	1,747.53	287.35	22,475.06

Table 3. Average computation time in millisecond (best value is marked in bold).

Signal length	E_D(fBm signal)					
BCM	VFD	HM	DFA	PSDA	CEM	
$L = 2^7$	0.091	0.018	**0.003**	0.004	0.027	0.039
$L = 2^8$	0.069	0.013	0.031	**0.002**	0.031	0.047
$L = 2^9$	0.054	0.008	0.033	**0.001**	0.096	0.040
$L = 2^{10}$	0.044	0.007	0.001	**0.001**	0.099	0.045
$L = 2^{11}$	0.038	0.005	0.019	**0.002**	0.102	0.042
$L = 2^{12}$	0.032	0.003	0.015	**0.001**	0.102	0.041
$L = 2^{13}$	0.023	**0.001**	**0.001**	0.004	0.101	0.046
$L = 2^{14}$	0.022	0.004	**0.001**	**0.001**	0.105	0.046
$L = 2^{15}$	0.019	0.007	0.002	**0.001**	0.103	0.053
Average	0.044	0.007	0.012	**0.002**	0.085	0.044

Table 4. MSE value performed with the fBm signal (best value is marked in bold).

Signal length	SD(fBm signal)					
BCM	VFD	HM	DFA	PSDA	CEM	
$L = 2^7$	0.083	0.009	0.054	**0.003**	0.059	0.038
$L = 2^8$	0.067	0.023	0.004	**0.002**	0.048	0.048
$L = 2^9$	0.044	**0.002**	0.003	0.009	0.173	0.026

Signal length	SD(fBm signal)					
$L = 2^{10}$	0.030	0.003	0.002	**< 0.001**	0.107	0.033
$L = 2^{11}$	0.020	0.057	0.003	**< 0.001**	0.098	0.041
$L = 2^{12}$	0.018	0.083	**0.002**	0.005	0.089	0.036
$L = 2^{13}$	0.026	0.105	**0.004**	0.022	0.027	0.044
$L = 2^{14}$	0.037	0.182	**0.004**	0.042	0.051	0.035
$L = 2^{15}$	0.049	0.183	**0.003**	0.065	0.029	0.029
Average	0.034	0.072	**0.009**	0.016	0.076	0.037

Table 5. MSE value performed with the Syn signal (best value is marked in bold).

Signal length	SD(fBm signal)					
BCM	VFD	HM	DFA	PSDA	CEM	
$L = 2^7$	**0.056**	0.201	0.109	0.112	0.248	0.069
$L = 2^8$	**0.047**	0.150	0.089	0.095	0.154	0.056
$L = 2^9$	0.061	0.116	**0.040**	0.069	0.090	0.047
$L = 2^{10}$	0.034	0.088	**0.032**	0.049	0.056	0.042
$L = 2^{11}$	0.032	0.078	**0.019**	0.039	0.036	0.037
$L = 2^{12}$	0.023	0.079	**0.016**	0.036	0.028	0.031
$L = 2^{13}$	0.023	0.062	**0.014**	0.028	0.019	0.027
$L = 2^{14}$	0.018	0.054	**0.013**	0.024	0.014	0.026
$L = 2^{15}$	0.015	0.021	**0.011**	0.022	0.013	0.014
Average	**0.034**	0.094	0.038	0.053	0.073	0.039

Table 6. Deviation performed with the fBm signal (best value is marked in bold).

Signal length	SD(Syn signal)					
BCM	VFD	HM	DFA	PSDA	CEM	
$L = 2^7$	**0.039**	0.151	0.056	0.085	0.278	0.036
$L = 2^8$	0.031	0.116	**0.026**	0.044	0.236	0.025
$L = 2^9$	0.023	0.074	**0.022**	0.043	0.141	0.046
$L = 2^{10}$	0.017	0.068	0.021	**0.016**	0.148	0.038
$L = 2^{11}$	0.018	0.070	0.019	**0.011**	0.165	0.030
$L = 2^{12}$	0.019	0.091	0.018	**0.010**	0.175	0.022
$L = 2^{13}$	0.022	0.063	0.017	**0.009**	0.127	0.020

Signal length	SD(Syn signal)					
L = 2¹⁴	0.019	0.049	0.015	**0.014**	0.121	0.017
L = 2¹⁵	**0.008**	0.011	0.009	0.012	0.119	0.012
Average	**0.022**	0.077	**0.022**	0.027	0.168	0.027

Table 7. Deviation performed with the Syn signal (best value is marked in bold).

4.4. Time-Dependent Fractal Dimensions (TDFD)

One of common measure of irregularity in fluctuations as time-sequential data may be the fractal dimension. Fractal dimensions estimate the degree of freedom of fluctuations in a signal. To measure the fractality in short time intervals of time-sequential data from one end of the waveform to the other sequentially, we may observe the dynamical changes in the fractal dimensions with respect the time series. These fractal dimensions, namely, the time-dependent fractal dimensions (TDFD) [22].

In this study, EEG signal was sampled at 512 Hz (or L =3,072 points) whose duration is 6 seconds. We set a window function is a rectangular type, window size = 512 points (1 s), moving window with intervals = 10 points (19.5 ms) because the temporal resolution of EEG is millisecond or even better [29]. The number of obtained points can compute from $N_{FD} = \lfloor (L - L_w)/\Delta t \rfloor + 1$ where L is a signal length, L_w is a window length ($L_w \leq L$), and Δt is an interval. Thus, we then obtained 257 points of fractal dimension. The time series-based algorithms of fractal dimension present more effective than frequency-based algorithms [16].
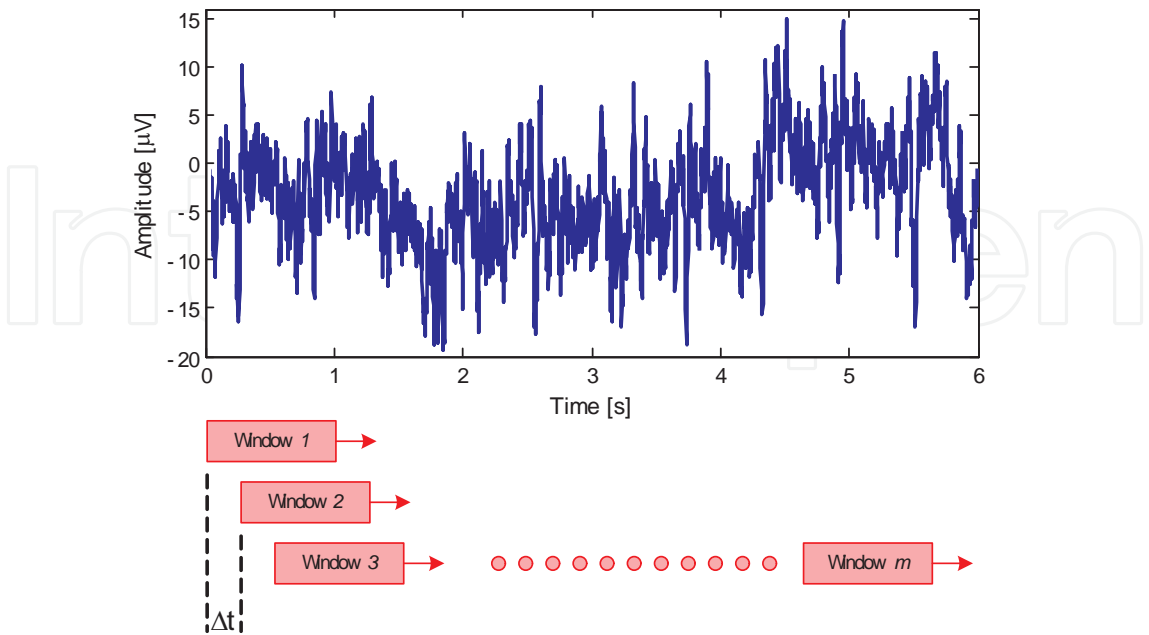
Figure 2. Computation process of TDFDs.

4.5. Evaluation of features by Kullback-Leibler divergence

In probability theory and information theory, the Kullback–Leibler divergence [34] (or called K-L divergence) is a measure of the difference between two probability distributions: from an actual probability distribution P to an arbitrary probability distribution Q.

$$D_{KL}\ (P\parallel Q)=\int_{-\infty}^{+\infty}p(x)\log\frac{p(x)}{q(x)}dx \tag{26}$$

For probability distributions P and Q of a discrete random variable is defined to be

$$D_{KL}\ (P\parallel Q)=\sum_{i}P(i)\log\frac{P(i)}{Q(i)} \tag{27}$$

Important properties of the K-L divergence are

1. The K–L divergence is always positive $D_{KL}\ (P\parallel Q)\geq0$,

2. $D_{KL}\ (P\parallel Q)=0$ if and only if $P=Q$,

3. The K-L divergence is not symmetric $D_{KL}\ (P\parallel Q)\neq D_{KL}\ (Q\parallel P)$.
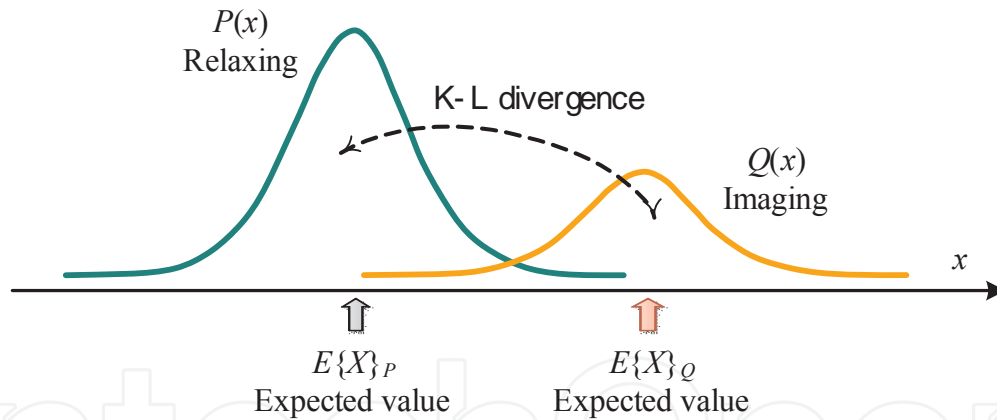
Figure 3. Feature extraction concepts on the basis of the K-L divergence.

The different expected values between imaging and relaxing periods are also proposed to use as the featuring parameter together with the K-L divergence. In general, if X is random variable defined on a probability space, then the different expected value of two random variables can be defined by

$$\Delta E\{X\}_{PQ}=\sum_{i}p_{i}x_{i}-\sum_{j}q_{i}x_{j} \tag{28}$$

In this study, we applied the K-L divergence and different expected values for finding the features between relaxing and imaging periods. According to Eq.(28), $P(i)$ is regarded as relaxing period,

and $Q(i)$ is regarded as imaging period. The probability distributions can be approximated by normalizing histogram of the data. The output patterns of K-L divergences versus the different expected value are also shown in Fig. 4. These obtained results can be classified by the neural network. More explanations of this classification process are in chapter 5.
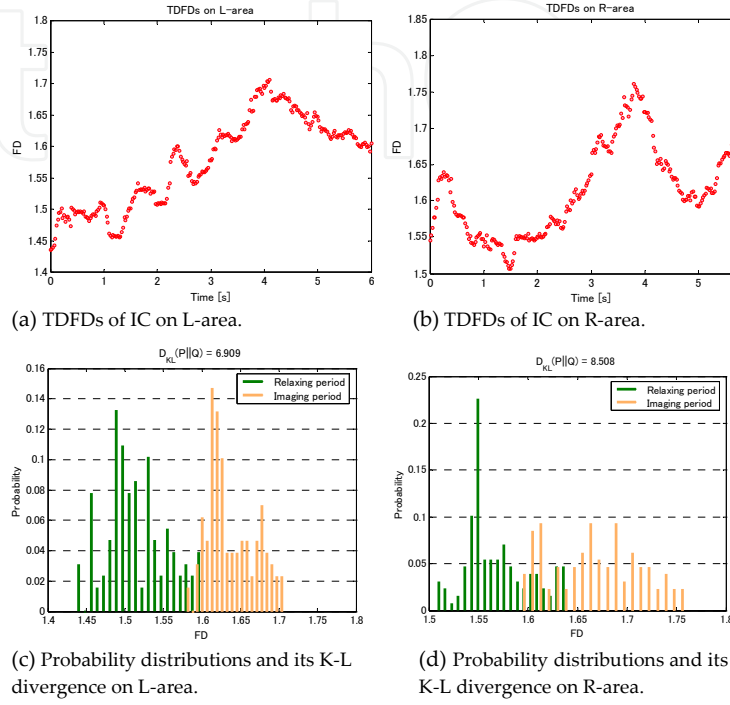

(a) TDFDs of IC on L-area.

(b) TDFDs of IC on R-area.

(c) Probability distributions and its K-L divergence on L-area.

(d) Probability distributions and its K-L divergence on R-area.

Figure 4. Left-hand movement imagination (LH-MI).

5. Neural network classifier

This study is concerned with the most common class of neural networks (NNs) called back-propagation algorithm. Backpropagation is an abbreviation for the backwards propagation of error. The learning process is processed based on the Levenberg–Marquardt (LM) backpropagation method [30] the Hessian matrix can be approximated as the following equation

$$H \approx 2J^{T}J \tag{29}$$

and the gradient can be determined as

$$\nabla J = 2J^{T}\mathbf{e} \tag{30}$$

where J is the Jacobian matrix that contains first derivatives of the network error with respect to weights and biases, and $\mathbf{e}$ is a vector of network errors.

$$J\left(w\right) = \begin{pmatrix} \dfrac{\partial e_1(w)}{\partial w_1} & \cdots & \dfrac{\partial e_1(w)}{\partial w_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial e_N(w)}{\partial w_1} & \cdots & \dfrac{\partial e_N(w)}{\partial w_n} \end{pmatrix} \tag{31}$$

The LM algorithm can use for updating the vector of weights, **w**, by

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \left(J^T J + \mu I\right)^{-1} J^T e \tag{32}$$

$$\Delta \mathbf{w} = - \left(J^T J + \mu I\right)^{-1} J^T e \tag{33}$$

where I is an identity matrix, μ is a learning rate, and n is an iteration step. This algorithm is particularly well suited for training NN with MSE. Moreover, the LM algorithm has been shown to be much faster than backpropagation in a variety of applications [31].
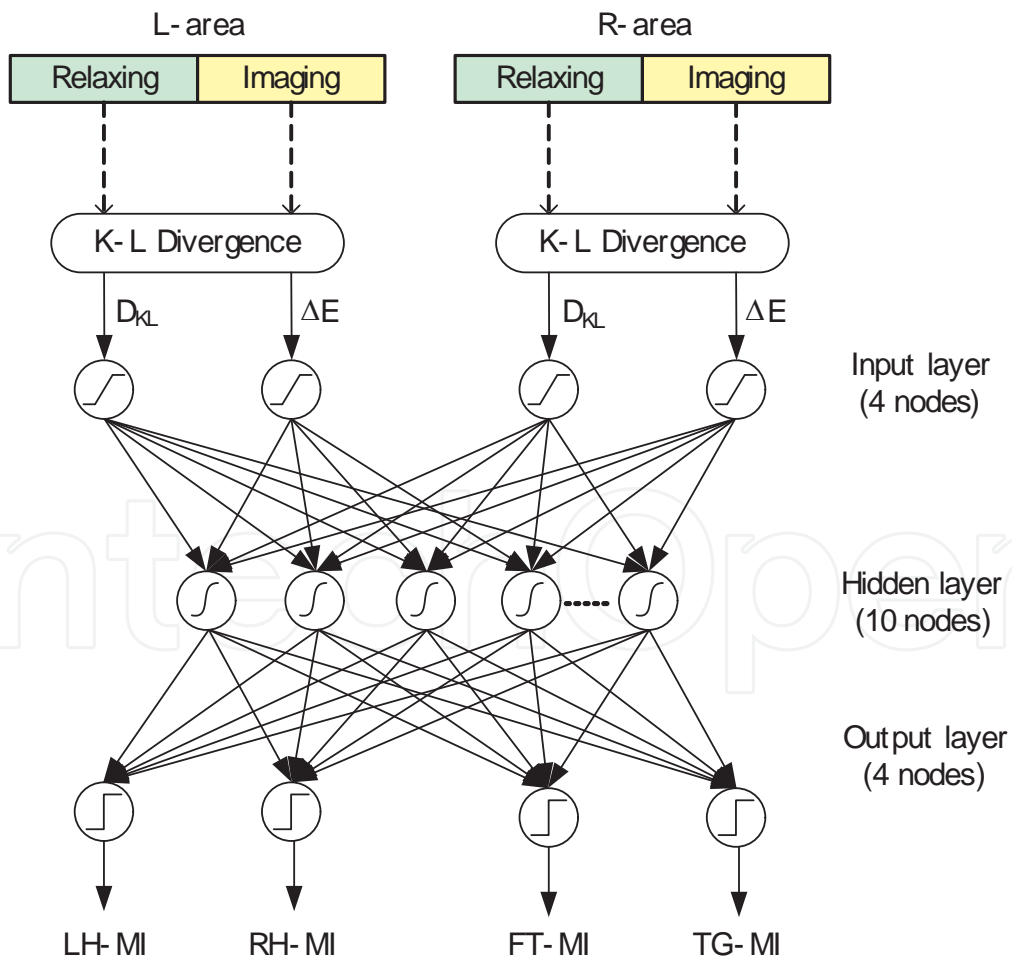
Figure 5. Classification diagram of the proposed method.

Task	Desired vector
LH-MI	$[1 \quad 0 \quad 0 \quad 0]^T$
RH-MI	$[0 \quad 1 \quad 0 \quad 0]^T$
FT-MI	$[0 \quad 0 \quad 1 \quad 0]^T$
TG-MI	$[0 \quad 0 \quad 0 \quad 1]^T$

Table 8. Desired vectors from the proposed network

5.1. Cross-validation procedure

Then, the featured parameters typically perform in terms of fractal properties. The NN was selected as classifier and it includes adaptive self-learning that was learnt to produce the correct classifications based on a set of training examples via cross validation process. A tenfold cross validation method was used to increase the reliability of the results [13]. For each experiment, we divided all the data into three sets, namely, the training set (50% of the entire data), cross validation set (20% of the entire data), and testing set (30% of the entire data). According to the tasks, a set of desired values for network training are shown in Table 8. The MSE rate is succeeded versus the iteration step is shown in Fig. 6. In Table 9, we compared the proposed method with the conventional feature, namely, band power estimation (BPE) [14]. The training process makes the network to define its decision boundaries. Figures 7 to 8 show a tight decision boundary after training the network.

6. Discussions

This chapter discusses the major contributions of this study, 12 electrode channels were used to measure EEG signal over the sensorimotor area of the cortex. The accuracy rates of motor imagery tasks are satisfactory. The error in this classification is due to the ambiguous data between the relaxing and imaging periods, particularly when the subjects have imagined the movements. However, in this experiment, throughout the period of recoding raw EEG signals without training and also without rejecting a bad trial. On contrary, they can be different for different individuals. This makes the EEG signal depend on the measured subject. Even if different persons perform the same tasks and the measurement was done under identical conditions the resulting signals could differ [35].

The proposed method provides 2880 features for each subject, whereas BPE method provides 21600 features. The current features obtained from the proposed method present an easy practicable and reliable method for training the classification algorithms. Nevertheless, the number of features from BPE method can be reduced by changing the number of electrodes. We used the NN to classify the features processed by BPE method. After learning the NN, learning curves and decision boundaries of BPE method is shown in Fig. 9. As the results,
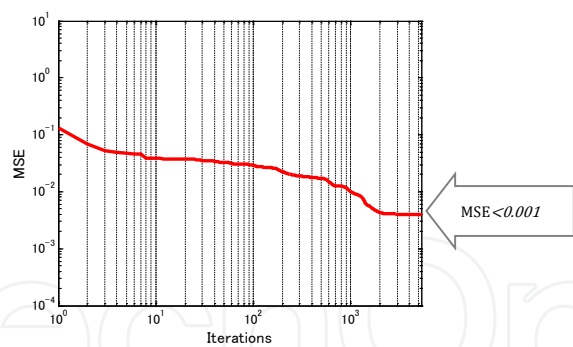
Figure 6. Log-log plot of learning curve.

Method	Accuracy [%]			
	LH-MI	RH-MI	FT-MI	TG-MI
BPE	77.6	79.5	76.4	77.3
Proposed	82.5	84.3	83.6	82.6

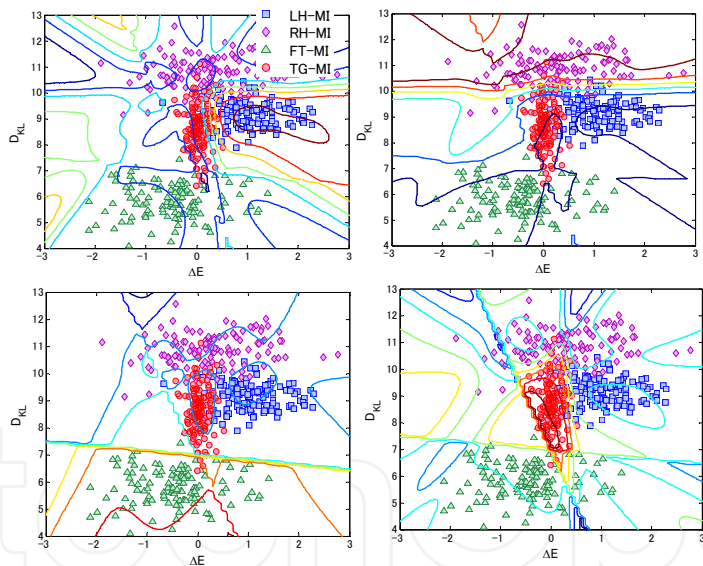**Table 9.** A comparison result of average accuracy rates.

Figure 7. Decision boundary of the trained network.

output decision functions of classifying features based on BPE method does not perform well and cannot made clearly separation boundaries.

The six algorithms present specific properties in terms of capability and variability. Some algorithms appeared inapplicable for the EEG time series data in this study, such as PSDA and CEM algorithms, for example. The main reason was that two algorithms based on frequency analysis since estimating the Hurst exponent can be utilized by means of the FFT method where
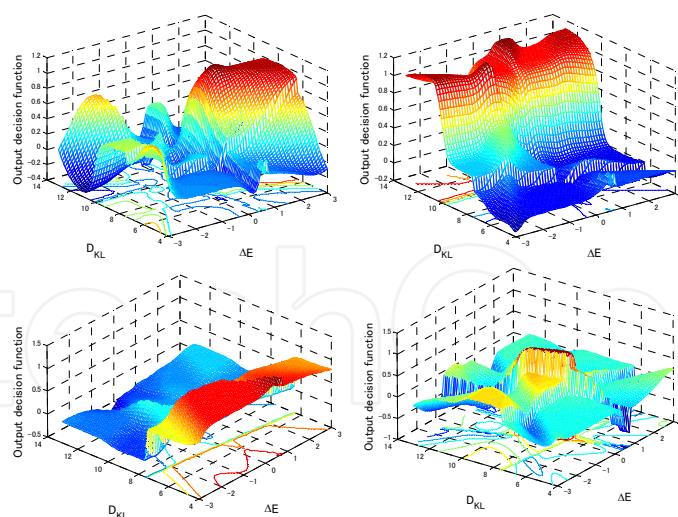
Figure 8. Corresponding 3-D surface of decision boundaries for each task.
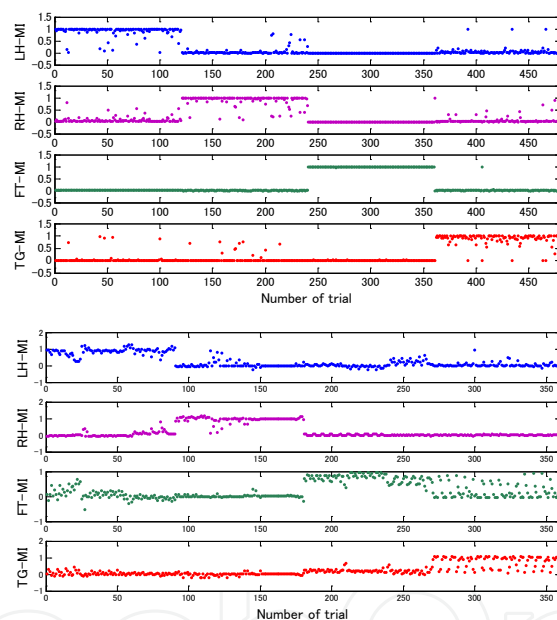
Figure 9. Updated weight values of the NN after learning; (Top panel) Proposed method, (Bottom panel) BPE Method.

the FFT can be substantially applicable in the long-time series data, and it easily declines when the EEG data has been analyzed in a short-time series. However, we can enhance the variability and avoid such drawbacks by increasing the sampling rate at the same period of experiment. In terms of variability, the HM algorithm presents the lowest average values of SD. The DFA provides the fast computation time and also presents reasonable values in terms of accuracy and variability when comparison with each other. Although, VDF gives extremely fast computation time but the VFD itself has the drawbacks in terms of accuracy and robustness. The VFD algorithm appropriates for real-time application, since its consumed computation time was less than 10 ms. Moreover, as proposed algorithms we suggest that the most

appropriate algorithm depends on an application purpose; for instance, the HM algorithm was suited for evaluating the FD of EEG data with high precision, whereas HM algorithm extremely consumed the processing time in comparison with the other algorithms. To show the performance of all algorithms, we then compared all estimated FDs with the theoretical value especially computed by short interval, i.e., 2^7, 2^8 and 2^9 points; these results were shown in Figs. 10 to 15.
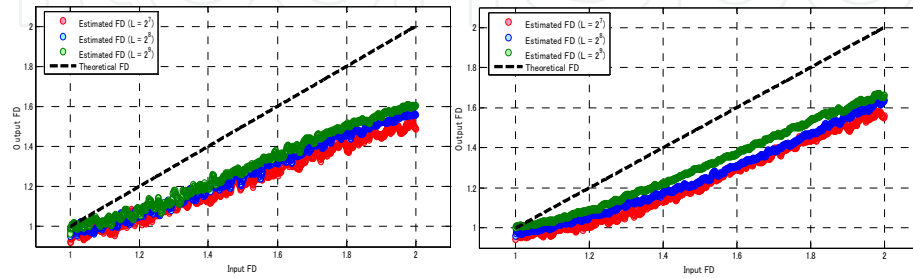

Figure 10. Fractal dimension evaluated by BCM. (Right) fBm signal; (left) Syn signal.
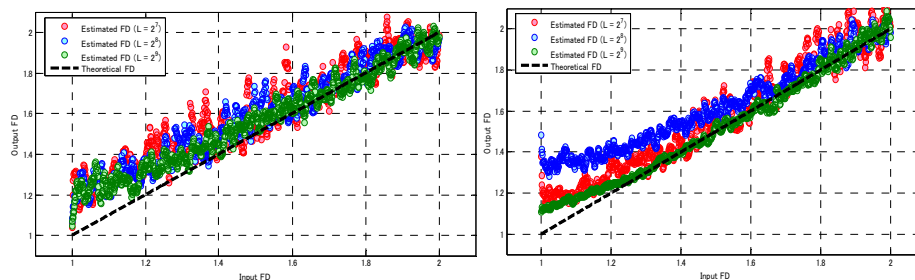
Figure 11. Fractal dimension evaluated by VFD. (Right) fBm signal; (left) Syn signal.
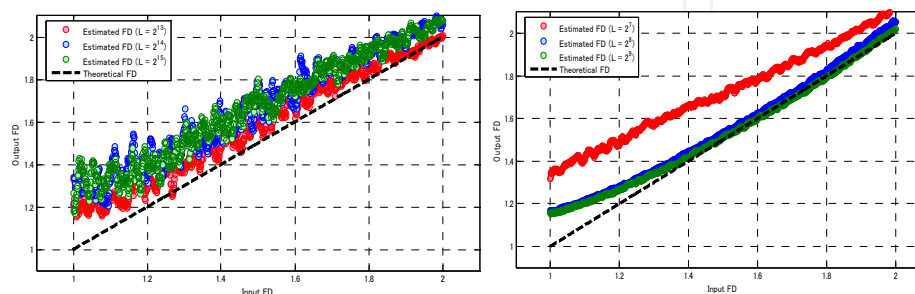

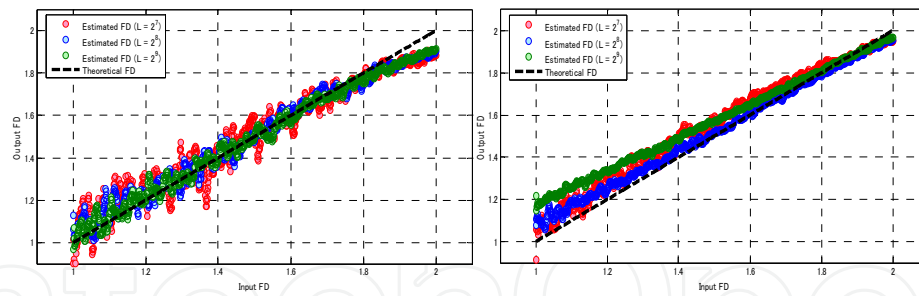**Figure 12.** Fractal dimension evaluated by HM. (Right) fBm signal; (left) Syn signal.

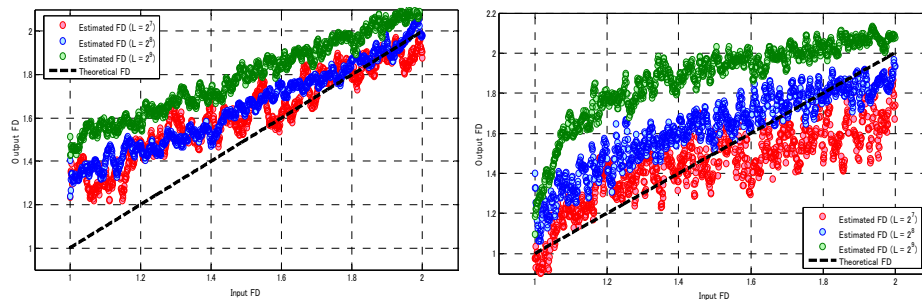**Figure 13.** Fractal dimension evaluated by DFA. (Right) fBm signal; (left) Syn signal.

Figure 14. Fractal dimension evaluated by PSDA. (Right) fBm signal; (left) Syn signal.
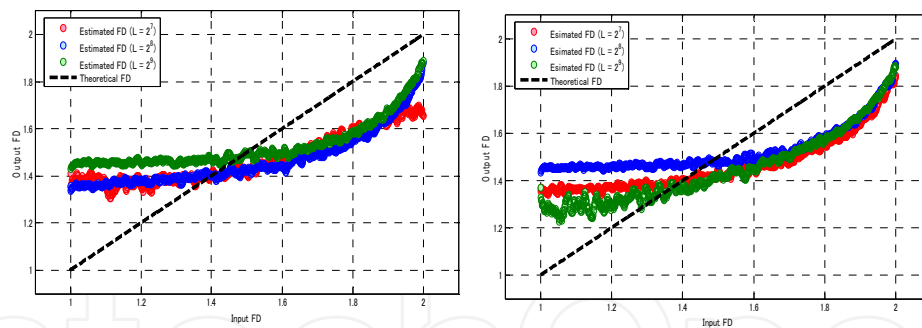

Figure 15. Fractal dimension evaluated by CEM. (Right) fBm signal; (left) Syn signal.

7. Conclusion

By selecting the best performance of fractal algorithm, we can obviously improve the rate of classification and can develop the novel methods in terms of fractal properties. Fractal features and the experimental framework can be applied not only binary-command BCIs, but also multi-command BCIs. The measurement of FD from EEG data can be capable not only in particular works, but also in publicity data. The FD characterizes the self-affine property of EEG data and has a direct relation to the different tasks.

Author details

Montri Phothisonothai and Katsumi Watanabe

Research Center for Advanced Science and Technology, The University of Tokyo, Japan Society for the Promotion of Science, Tokyo, Japan

References

[1] Grassberger, P, & Procaccia, I. Estimation of the Kolmogorov Entropy from a Chaotic Signal. Phys. Rev. A (1983). , 28, 2591-2593.

[2] Pincus, S. M, Gladstone, I. M, & Ehrenkranz, R. A. A Regularity Statistics for Medical Data Analysis. J. Clin. Monit. (1991). , 7, 335-345.

[3] Katz, M. J. Fractals and the Analysis of Waveforms. Comput. Biol. Med. (1998).

[4] Barnsley, M. F. Fractals Everywhere, Ch. II and III, Academic Press Professional; (1993).

[5] Grassberger, P, & Procaccia, I. Measuring Strangeness of Strange Attractors. Phys. Lett. D (1983). , 148, 63-68.

[6] Badii, R, & Politi, A. Statistical Description of Chaotic Attractors: The Dimension Function. J. Stat. Phys. (1985). , 40, 725-750.

[7] Boostani, R, & Moradi, M. H. A New Approach in the BCI Research based on Fractal Dimension as Feature and Adaboost as Classifier. J. Neural Eng. (2004). l., 1, 212-217.

[8] Popivanov, D, Jivkova, S, Stomonyakov, V, & Nicolova, G. Effect of Independence Component Analysis on Multifractality of EEG during Visual-Motor Task. Sig. Pro. (2005). , 85, 2112-23.

[9] Lutzenberger, W, Elbert, T, Birbaumer, N, Ray, W. J, & Schupp, H. The Scalp Distribution of the Fractal Dimension of the EEG and its Variation with Mental Tasks. Brain Topo. (1992). , 5, 27-34.

[10] Prei, l H, Lutzenberger, W, Pulvermüller, F, & Birbaumer, N. Fractal Dimensions of Short EEG Time Series in Humans. Neurosci. Lett. (1997). , 225, 77-80.

[11] Phothisonothai, M, & Nakagawa, M. EEG-based Classification of Motor Imagery Tasks using Fractal Dimension and Neural Network for Brain-Computer Interface. IEICE Trans. Inf. Syst. (2008). ED: 44-53., 91.

[12] Phothisonothai, M, & Nakagawa, M. EEG Signal Classification Method based on Fractal Features and Neural Network. The 30th IEEE EMBC (2008). , 3880-3883.

[13] Setiono, R. Feedforward Neural Network Construction using Cross Validation. Neural Comput. (2005). , 13, 2783-2786.

[14] Palaniappan, R. Utilizing Gamma Band to Improve Mental Task based Brain-Computer Interface Design. IEEE Trans. Neural Net. Rehabil. Eng. (2006). , 14, 299-303.

[15] Mandelbrot, B. B. The Fractal Geometry of Nature. W.H. Freeman, New York; (1983).

[16] Phothisonothai, M, & Nakagawa, M. Fractal-based EEG Data Analysis of Body Parts Movement Imagery Tasks. J Physiol Sci. (2007). , 57(4), 217-226.

[17] Kinsner, W. Batch and Real-time Computation of a Fractal Dimension based on Variance of a Time Series. Univ. Manitoba Canada. Tech. Report (1994). del, 94-6.

[18] Higuchi, T. Approach to an Irregular Time Series on the Basis of the Fractal Theory. Physica D (1988). , 31, 277-83.

[19] Peng, C. K, Mietus, J, Hausdorff, J. M, Havlin, S, Stanley, H. E, & Goldberger, A. L. Long Range Anticorrelations and Non-Gaussian Behavior of the Heartbeat. Phys. Rev. Lett. (1993). , 70, 1343-1346.

[20] Peng, C. K, Havlin, S, Stanley, H. E, & Goldberger, A. L. Quantification of Scaling Exponents and Crossover Phenomena in Nonstationary Heartbeat Time Series. Chaos (1995). , 5, 82-87.

[21] Nakagawa, M. A Critical Exponent Method to Evaluate Fractal Dimension of Self-affine Data. J. Phys. Soc. Jpn. (1993). , 62, 4233-4239.

[22] Sabanal, S, & Nakagawa, M. The Fractal Properties of Vocal Sound and their Application in the Speech Recognition Model. Chaos Soli. Frac. (1996). , 7, 1825-1843.

[23] Petry, A, Augusto, D, & Barone, C. Speaker Identification using Nonlinear Dynamical Features. Chaos Soli Frac. (2002). , 13, 221-231.

[24] Phothisonothai, M, & Nakagawa, M. EEG-based Fractal Analysis of Different Motor Imagery Tasks using Critical Exponent Method. Int. J. Bio. Life Sci. (2006). , 1(3), 175-180.

[25] Mandelbrot, B. B, & Van Ness, J. W. Fractional Brownian Motions, Fractional Noises and Applications. SIAM Rev. (1968). , 10, 422-437.

[26] Fougère, P. F. On the Accuracy of Spectrum Analysis of Red Noise Processes using Maximum Entropy and Periodogram Methods: Simulation Studies and Application to Geographical Data. J. Geograp. Res. (1985). , 95, 4355-4366.

[27] Eke, A, Hermann, P, Kocsis, L, & Kozak, L. R. Fractal Characterization of Complexity in Temporal Physiological Signals. Physio. Meas. (2002). R, 1-38.

[28] Tricot, C. Curves and Fractal Dimension. New York: Springer-Verlag;(1995).

[29] Koenig, T, Studer, D, Hubl, D, Melie, L, & Strik, W. K. Brain connectivity at different time-scales measured with EEG. Philos. Trans R. Soc. Lond. B Biol. Sci. (2005). , 360, 1015-1024.

[30] Marquardt, D. W. An Algorithm for Least-Squares Estimation of Non-linear Parameters. J. Soc. Indust. Appl. Math. (1963). , 11, 431-441.

[31] Principe, J. C, Euliano, N. R, & Lefebvre, W. C. Neural and Adaptive Systems: Fundamentals through Simulations. John Wiley & Sons; (2000).

[32] Nakagawa, M. Chaos and Fractals in Engineering. World Sciencetific; (1999).

[33] Esteller, R, Vachtsevanos, G, Echauz, J, & Litt, B. A Comparison of Waveform Fractal Dimension Algorithms. IEEE Trans. Circuits Syst. (2001). , 48, 177-183.

[34] Kullback, S, & Leibler, R. A. On Information and Sufficiency. Ann. Math. Statist. (1951). , 22, 79-86.

[35] Curran, E, & Stokes, M. Learning to Control Brain Activity: A Review of the Production and Control of EEG Components for Driving Brain-Computer Interface (BCI) Systems. Brain Cogn. (2003). , 51, 326-336.

[36] Phothisonothai, M, & Nakagawa, M. A Classification Method of Different Motor Imagery Tasks Based on Fractal Features for Brain-Machine Interface. J. Intre. Neurosci. (2009)., 8, 95-122.