TB, RMK, UK, PM/341796, 14/05/2010

IOP Publishing PHYSIOLOGICAL MEASUREMENT

Physiol. Meas. **31** (2010) 1–18 UNCORRECTED PROOF

Classification of biological signals using linear and nonlinear features

T Balli and R Palaniappan

School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

E-mail: tballi@essex.ac.uk and rpalan@essex.ac.uk

Received 23 December 2009, accepted for publication 29 April 2010
Published DD MMM 2010
Online at stacks.iop.org/PM/31/1

Abstract

This paper investigates the characterization ability of linear and nonlinear features and proposes combining such features in order to improve classification of biological signals, in particular single-trial electroencephalogram (EEG) and electrocardiogram (ECG) data. For this purpose, three data sets composed of ECG, epileptic EEG and finger-movement EEG were utilized. The characterization ability of seven nonlinear features namely the approximate entropy, largest Lyapunov exponents, correlation dimension, nonlinear prediction error, Hurst exponent, higher order autocovariance and asymmetry due to time reversal are compared with two linear features namely the autoregressive (AR) reflection coefficients and AR model coefficients. The features were tested by their ability to differentiate between different classes of data using a linear discriminant analysis (LDA) method with tenfold cross-validation. The class separability of combined linear and nonlinear features was assessed using sequential floating forward search with linear discriminant analysis method (SFFS-LDA). The results demonstrated that linear and nonlinear features on their own provided comparable results for the ECG data set and the finger-movement EEG data set whilst the linear features provided a better class separability compared to nonlinear features for the epileptic EEG data set. Combining linear and nonlinear features demonstrated a significant improvement in the class separability for all of the data sets where an average improvement of 20.56% was obtained with the ECG data set, 7.45% with finger-movement data set and 6.62% with the epileptic EEG data set. Overall results suggest that the use of combined linear and nonlinear feature sets would be a better approach for the characterization and classification of biological signals such as EEG and ECG. Q1

Keywords: classification, ECG, EEG, linear features, nonlinear features

1. Introduction

Conventional feature extraction methods for the analysis of biological signals utilize time- and frequency-based methods (Anderson *et al* 1998, Chua *et al* 2008, Geng *et al* 2008, Phothisonothai and Nakagawa 2008, Pires *et al* 2007, Tamil *et al* 2008 and Zhang *et al* 2008). The theory behind these methods assumes that these signals are linear. Although such methods provide reasonable results, the underlying nonlinear properties of these signals are ignored. Therefore, the requirements for further characterization and a better understanding of biological signals has led to an increasing interest in nonlinear methods adopted from higher order statistics, nonlinear dynamics and chaos theory domains.

In recent years, there have been many research studies on nonlinear features for the analysis of biological signals. The majority of research studies have been devoted to analysis of electroencephalogram (EEG) signals recorded from subjects with different pathological conditions such as epilepsy (Lehnertz 2008, Greene *et al* 2008, Casdagli *et al* 1996, Gautama *et al* 2003, Andrzejak *et al* 2001, Akay 2001), Alzheimer's (Czigler *et al* 2008), schizophrenia (Sabeti *et al* 2009, Lee *et al* 2001) and sleep disorders (Olbrich *et al* 2003, Berryman *et al* 2005, Shen *et al* 2003) as well as the analysis of heart rate variability (HRV) (Benitez *et al* 2009, Akay 2001) and electrocardiogram (ECG) (Owis *et al* 2002, Small *et al* 2002, Akay 2001) signals recorded from healthy subjects and patients with heart rhythm disorders. In general these research studies have focused specifically on nonlinear dynamic measures, with little to no literature on direct comparisons between such measures and their linear counterparts. The focus on single type of measures also means that few studies have been performed which investigate the combination of linear and nonlinear measures for the characterization and classification of biological signals.

In this study, we set out to investigate the characterization ability of linear autoregressive (AR) features and five nonlinear features adopted from nonlinear dynamics and chaos theory domain, namely the approximate entropy (APEN), largest Lyapunov exponents, correlation dimension (CD), nonlinear prediction error (NLPE) and Hurst exponent. We also test two nonlinear features adopted from higher order statistics domain, namely the higher order autocovariance and asymmetry due to time reversal.

Three biological signal data sets composed of ECG, epileptic EEG and finger-movement EEG are used in this study. The features were assessed by their ability to differentiate between different classes of data. Class separability was investigated through direct comparison between linear and nonlinear features and also by comparing combined linear and nonlinear feature sets with different parameter settings.

2. Data sets

In this study, we have utilized three data sets which are composed of ECG signals recorded from healthy subjects and patients with premature contractions[1], EEG signals recorded from healthy subjects during an idle (resting) state and during flexion/extension of the left index finger and EEG signals recorded from healthy subjects and epilepsy patients.

2.1. ECG data set

The ECG data from Massachusetts General Hospital/Marquette Foundation (MGH/MF) database (see MIT-BIH Arrhythmia Database) were used in this study. For analysis, three

[1] Premature contractions are the heart beats that occur earlier than expected and cause irregularity in the usual rhythm of the heart. The occurrences of premature heart beats are not life threatening but are indication of heart problems that can be predisposition to life-threatening arrhytmias (Beasley 2003).

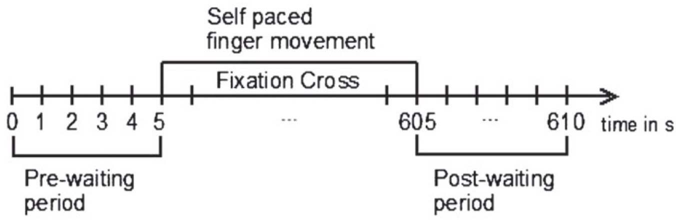

Figure 1. The timing scheme of the experimental paradigm.

classes of ECG beats were considered: the premature supraventricular contraction (PSC) beats, premature ventricular contraction (PVC) beats and normal (N) beats. Data from ten subjects for each class were chosen from tapes mgh001–mgh050 after visual inspection for errors. Here we have utilized 30 randomly chosen recordings from each subject and extracted three consecutive beats (QRS complexes) from each recording using the Pan and Tomkins algorithm (Pan and Tompkins 1985). There were a total of eight signals from each recording which included the data from three ECG leads, arterial pressure, pulmonary arterial pressure, central venous pressure, respiratory impedance and airway CO_2 waveforms. Here we have only utilized the signals from three ECG leads. Thus a total of 8100 beats comprising 2700 N, 2700 PSC and 2700 PVC beats (10 subject $\times$ 30 recordings $\times$ 3 beats $\times$ 3 leads) were extracted for this study.

Prior to analysis, the ECG signals were pre-processed to remove high frequency content, baseline noise and 60 Hz powerline interference by using a bandpass filter from 1 to 35 Hz.

2.2. *EEG data set I*

The EEG data set was recorded from nine right-handed subjects (all subjects were male), with ages ranged from 23 to 46. Subject 8 had experience in using a BCI system based on self-paced movement, subjects 3 and 5 had experience in offline BCI experiments and the remaining subjects were naive to BCI use. Signals were acquired using the Guger Technologies g.Bsamp device. EEG signals were recorded over the motor cortex from five bipolar channels located at C3, C1, Cz, C2 and C4, referenced to the right mastoid. Electromyogram (EMG) signals were recorded from the flexors of the left forearm for labeling of movement- and non-movement-related EEG. All data were sampled at 256 Hz.

Within each run, the subjects were asked to perform self-paced flexion/extension of the left index finger whilst a fixation cross was visible on the screen. They were instructed to perform each movement for 5–10 s and to rest for a minimum of 10 s between movements. As the data were un-cued the number of trials within each run was variable. Each subject performed three runs in a single session. Each run lasted for 610 s where the subjects had 5 s of pre-waiting and post-waiting periods before and after the fixation cross appeared on the screen for 600 s. The timing scheme of a run is illustrated in figure 1. Instructions were given to concentrate on the fixation cross as much as possible during each run. After each run the EMG recordings were assessed to ensure that the subjects understood requirements of the experiment.

2.3. *EEG data set II*

The EEG signals in this data set were obtained from Bonn University EEG database (see Epileptic EEG Database). There were a total of five classes (A–E) in this data set, each containing 100 single channel EEG segments. The EEG signals were recorded using a 128 channel system and common average referencing was applied. Signals were digitized with a 12 bit analogue-to-digital conversion and recorded at a sampling rate of 173.61 Hz. For each set, EEG segments of length 23.6 s were selected and cut from multichannel EEG recordings after visual inspection for artifacts such as muscle activity or eye movements.

The sets A and B consist of EEG segments recorded from five healthy subjects (using the international 10–20 electrode placement system) during the resting state with eyes open (set A) and eyes closed (set B). The sets C, D and E consist of EEG segments recorded from five epilepsy patients. All of the patients had achieved complete seizure control after resection of one of the hippocampal formations, indicating correct diagnoses of the epileptogenic zone. Segments in set C were recorded from the hippocampal formation of the opposite hemisphere of the brain whilst segments in set D were recorded from the epileptogenic zone. Segments in sets C and D were recorded during seizure-free intervals and segments in set E during the seizure interval.

3. Methods

3.1. Nonlinear features

In order to characterize the nonlinear properties in EEG and ECG segments, we have utilized two features from higher order statistics domain, namely the higher order autocovariance and asymmetry due to time reversal, and five features from nonlinear dynamics and chaos theory domains namely the APEN, largest Lyapunov exponents, CD, NLPE and Hurst exponent.

3.1.1. Higher order statistics features. The higher order autocovariance and asymmetry due to time reversal features were adopted from the nonlinearity analysis literature where they were used along with the surrogate data method to investigate the indications of nonlinear structures in the time series data (Gautama *et al* 2003, 2004).

Higher order autocovariance. In the nonlinearity analysis literature, this feature is basically defined as the third-order autocovariance (C3) method. This feature measures the dependence of time series on two time shifted versions of itself,

$$C3(\tau) = (N - 2\tau)^{-1} \cdot \Sigma_{n=2\tau+1}^{N} (x(n) \cdot x(n-\tau) \cdot x(n-2\tau)), \tag{1}$$

where $x(n)$ is the time series of the length N and τ is the time lag.

In this study we have extended this feature to estimate the dependence of time series for higher orders such that

$$cX(\tau) = (N - m\tau)^{-1} \cdot \Sigma_{n=m\tau+1}^{N} (x(n) \cdot x(n-\tau) \cdot \ldots \cdot x(n-m\tau)), \tag{2}$$

where τ is the time lag and m is the order. The selection of order, m, and time lag, τ, parameters were treated as embedding parameter selection problem which will be explained in section 3.1.2.

Asymmetry due to time reversal. The asymmetry due to time reversal (REV) measures the irreversibility of time series which is a property of nonlinear time series (Gautama *et al* 2003, 2004). This measure is given by

$$\mathrm{REV}(\tau) = (N - 1 - \tau)^{-1} \cdot \Sigma_{n=1+\tau}^{N} (x(n) - x(n-\tau))^{3}, \tag{3}$$

where $x(n)$ is the time series of the length N and τ is the time lag. Here, the REV feature is extracted with the time lag, τ, ranging from 1 to 10. The time lag that yielded the highest feature separability was chosen as optimal.

3.1.2. Nonlinear dynamic features. The first step in nonlinear dynamic measure estimation is state space reconstruction where the univariate data are transformed to its trajectory in multidimensional state space. Supposing that a single scalar time series $\{x(t), t = 1, \ldots, N\}$ is measured from the m-dimensional system using an observation function $g(\cdot)$ such that

$$x(t) = g(s(t)), \qquad g : M \to R, \qquad s(n) \in M \subseteq R^m, \tag{4}$$

where $s(t)$ stands for the state of system at time t and M is the representation of m-dimensional state space. The observation function $g(\cdot)$ cannot provide the complete representation of the underlying properties of the dynamical system. According to the Takens theorem (Takens 1981), this can be achieved by representing single scalar time series as time lagged versions of itself such that

$$f : R \to R^m, \qquad y_t = f(x(t)) = [x(t), x(t-\tau), \ldots, x(t-(m-1)\tau)], \tag{5}$$

where τ is the time lag, m is the embedding dimension and y_t is the state vector at time t.

The selection of the embedding dimension, m, and time lag, τ, parameters is important to achieve a good reconstruction of the time series in state space. In this study we have used three approaches for the selection of embedding parameters. In the first approach, we have utilized the conventionally used false nearest neighbors method along with first local minimum of mutual information function for estimation of m and τ, respectively. In the second approach, we have used the false nearest neighbors method along with first zero crossing of autocorrelation function for estimation of m and τ. In the third approach, we have selected the embedding dimension, m, and time lag, τ, pairs by minimization of the NLPE on EEG and ECG segments. The parameter pairs estimated by this method lead to reconstruction that exploits the nonlinear deterministic nature of the time series in state space; further information about this approach can be found in our previous work (Balli and Palaniappan 2009). The three approaches were used to investigate the effect of parameter selection methods on the separability of features for different classes of data. Additionally, we have performed an exhaustive search on the parameter space, with m values ranging from 2 through 10 and τ values ranging from 1 through 10, in order to explore the parameter pairs that maximize the class separability of different features.

Approximate entropy method. The APEN is a measure that quantifies the irregularity of a time series proposed by Pincus (1991). This measure can be estimated as follows:

$$C_i^m = \sum_{j=1}^{Nv} \frac{\Theta(r - \|y_i - y_j\|)}{Nv}, \tag{6}$$

where Nv is the number of vectors in state space, r is the tolerance of comparison value, y_i and y_j are vectors reconstructed in m-dimensional state space, $\|\cdot\|$ represents the Euclidean distance between vectors and $\Theta(x)$ is the Heaviside function such that $\Theta(x) = 1$ if $x > 0$ and $\Theta(x) = 0$ if $x \leqslant 0$. APEN(m, r) is obtained by

$$ApEn(m, r) = \Phi^m(r) - \Phi^{m+1}(r), \tag{7a}$$

$$\Phi^m(r) = (N-(m-1))^{-1} \cdot \sum_{i=1}^{N-(m-1)} \ln\left[C_i^m(r)\right], \tag{7b}$$

where N is the length of time series and m is the embedding dimension.

The estimation of APEN measure involves selection of tolerance of the comparison value, r, in addition to embedding parameters, m and τ. In previous studies, it has been suggested that the r value of 0.2 times standard deviation of time series provides good statistical validity (Akay 2001). However in our prior studies, we found out that feature separability was improved using r values higher than 0.2 times standard deviation of time series (Dyson *et al* 2008). Therefore, this feature was estimated with r parameter values ranging from 0.2 to 5 times standard deviation of time series with increments of 0.2 in this study. The best r parameter was selected according to classification results. In our prior studies, we have also investigated that the APEN features estimated with different r values were complementary to each other, providing an improved characterization of EEG and ECG signals. Therefore, in this study the APEN features with r values ranging from 0.2 to 5 times standard deviation of time series were included in the feature space for selection of complimentary features.

Largest Lyapunov exponent. The largest Lyapunov exponent (LLE) quantifies the average exponential divergence of nearby trajectories in state space where the sensitive dependence on initial conditions is obtained. In the literature several algorithms have been proposed for the calculation of LLE (Akay 2001, Kantz and Schreiber 1997, Sprott 2003). In this study we have used Rosenstein's algorithm (Rosenstein *et al* 1993) where the LLE measure was estimated as follows: for each state space vector y_j the distance to the nearest neighbor y_i was calculated (equation (8*a*)), and then two neighboring points are evolved in state space by time t to calculate the new separation distance (equation (8*b*)):

$$d_j(0) = \|y_j - y_i\|, \tag{8a}$$

$$d_j(t) = \|y_{j+t} - y_{i+t}\|. \tag{8b}$$

The largest Lyapunov can be calculated using a least-squares fit to the average line defined by $L(t) = \langle \ln d_j(t) \rangle$.

Correlation dimension. CD is a measure of the dimensionality of the space occupied by state vectors, which is also referred to as fractal dimension[2] (Akay 2001, Kantz and Schreiber 1997, Sprott 2003). There are several algorithms for the estimation of CD, and in this study we have utilized the Grassberg–Procaccia algorithm (Akay 2001, Kantz and Schreiber 1997, Sprott 2003). Using this algorithm, the CD was estimated by first calculating correlation integral, $C(r)$ which is defined in equation (6), over a range of r values. Then, the plot of $\log C(r)$ versus $\log r$ should have a linear scaling region whose slope, in the limit of small r and large N, was the CD,

$$\mathrm{CD} = \lim_{r\to 0} \lim_{N\to\infty} \frac{d \ln C(r)}{d \ln r}. \tag{9}$$

Nonlinear prediction error. The NLPE is a simple algorithm which exploits the deterministic structure in the time series (Kantz and Schreiber 1997). This algorithm works by constructing local linear models on a given state space vector.

[2] A fractal dimension is any dimension measurement that allows noninteger values.

First of all, the state vectors, y_t, reconstructed from univariate time series, $x(t)$, were divided into train, Y_{train}, and test sets, Y_{test}, in which every state vector y_t in the train and test sets has a target data point, $x(t+T)$, for T step ahead prediction. Therefore for every state vector y_i, with corresponding target $x(i+T)$, in the test set, k nearest neighbors from the train set $\{y_j;\ j=1,\dots,k\}$, with corresponding targets $\{x(j+T);\ j=1,\dots,k\}$, were grouped together. In order to do the prediction, a linear model defined by

$$x(j+T) = a_0 + \sum_{k=1}^{m} a_k y_j(k), \tag{10}$$

was fitted to k state vectors and their target values. The model parameters were estimated using the recursive least-squares algorithm. Following this the estimated model was used to predict $x(i+T)$ using the y_i vector as input and the prediction error was calculated as $e = \langle |x(i+T) - \widehat{x}(i+T)| \rangle$.

In this study we have set T to 1 and k to $1/d$ $\{d;\ d=1,\dots,20\}$ of total number of state vectors in the train set. The k value leading to the highest feature separability was selected while testing the individual features. Furthermore, preliminary studies have shown that NLPE features with different k values were complimentary to each other with some of the embedding parameter pairs leading to improved characterization of EEG and ECG signals. Therefore, the NLPE features with k values $1/d$ $\{d;\ d=1,\dots,20\}$ were included in the feature space for selection of complimentary features.

Hurst exponent. The Hurst exponent (HURST), which is also called rescaled range statistics (R/S), was used for quantifying the self-similarity of time series, providing information on the recurrence rate of similar patterns in time series at different scales (Akay 2001, Sprott 2003). Unlike the rest of nonlinear dynamic features, this feature does not involve state space reconstruction.

The HURST exponent estimates range between 0 and 1. HURST $= 0.5$ indicates that there is no correlation in the time series; $0 <$ HURST < 0.5 indicates that time series has long-range anti-correlations and $0.5 <$ HURST < 1 indicates that there is long-range correlations in the time series.

In order to estimate the Hurst exponent from time series, $x(t)$, first, the accumulated deviation from the mean of time series over time T was calculated:

$$X_T(t) = \sum_{i=1}^{t} x(i) - \overline{x}, \tag{11}$$

with $t = 1,\dots,T$.

The $R(T)$ was calculated as the difference between the maximum and minimum value of $X_T(t)$ and $S(T)$ was calculated as the standard deviation of time series over time T:

$$\frac{R(T)}{S(T)} = \frac{\max(X_T(t)) - \min(X_T(t))}{\left(\frac{1}{T}\sum_{t=1}^{T}(x(t)-\overline{x})^2\right)^{1/2}}. \tag{12}$$

The Hurst exponent was obtained by calculating the average rescaled range over different scaling regions and the slope of line produced by $\ln(R(n)/S(n))$ versus $\ln(n)$ resulted in an estimate of the Hurst exponent.

3.2. Linear features

We have employed the linear AR model for characterizing the amplitude-dependent properties of EEG and ECG signals.

3.2.1. Linear autoregressive model. An AR model of order p is given by

$$x(t) = -\sum_{i=1}^{p} a_k x(t-k) + e_t, \tag{13}$$

where $x(t)$ are the data at time t, a_k are the AR model coefficients and e_t is the Gaussian white noise with mean zero.

In this study, we have used AR model coefficients and AR reflection coefficients, estimated by Burg's method (Burg 2005), as features from EEG and ECG segments. The optimal model order was selected by extracting the AR coefficients with order ranging from 1 to 10 and the model order that maximized the separability of features for different classes of data was chosen.

3.3. Feature extraction and classification

The feature sets from the ECG data set were created by extracting the linear and nonlinear features from three consecutive beats taken from each recording belonging to each subject. For each class (N, PSC, PVC) a total of 900 features (10 subjects $\times$ 30 recordings $\times$ 3 leads) were extracted for testing the separability of linear and nonlinear features. The separability of individual features was assessed using the linear discriminant analysis (LDA) method with tenfold cross-validation. Training and testing sets for classification were created using feature sets from different subjects (i.e. with tenfold cross-validation, the features from nine subjects were used for training and the features from remaining one subject were used for testing); therefore, the system was not subject dependent.

The features from EEG data set I were extracted using a window of 256 samples with an overlap of 64 samples. During the recording session, the subjects were asked to perform self-paced movement; therefore, the length of movement and non-movement segments was variable for different runs recorded from subjects. To ensure equal sample sizes for each class, the minimum number of samples, N, for each class within a run was taken and N sample points were used from each class. The separability of individual features was assessed using the LDA method with tenfold cross-validation. The training and testing sets were created by shuffling movement and following non-movement trials from all sessions where 9/10 of trials were used for training and 1/10 of trials were used for testing in each fold. The separability of individual features for each subject was tested based on channels.

The signals in EEG data set II were divided into eight non-overlapping segments with 512 data points for feature extraction. We have adopted the weak stationarity criterion from Andrzejak *et al* (2001) for choice of the segment length. A total of 800 features (i.e. 100 single channel EEG divided into eight segments) were extracted from each class (A, B, C, D and E) for testing the feature separability. The separability of individual features was assessed using the LDA method with tenfold cross-validation.

The class separability using combined linear and nonlinear features for three data sets was further assessed using sequential floating forward search with linear discriminant analysis method (SFFS-LDA) (Pudil *et al* 1994) using tenfold cross-validation. A maximum of ten features were selected for the ECG data set and EEG data set II and a maximum of five features were selected for EEG data set I. The number of features used was based upon our preliminary studies which investigated the convergence of classification accuracy.

It should be noted that our primary aim in this study was to make a direct comparison between the separability of linear and nonlinear features using individual and combined feature sets. Therefore, we have utilized the LDA classifier as it is more robust compared to its nonlinear counterparts such that it has a low computational complexity and is less prone to

Table 1. LDA classification results for nonlinear dynamic features—ECG.

Feature name	FNN&MMI		FNN&AC		Minimum NLPE		Exhaustive search			
	Accuracy	Lead	Accuracy	Lead	Accuracy	Lead	Accuracy	m	τ	Lead
cX	47.33	2	39.89	1	44.89	3	53.22	2	6	2
CD	43.89	3	44.44	3	41.78	1	47.44	4	8	1
LLE	62.89	3	63.56	3	48.44	3	66.56	5	10	3
APEN	58.78	3	57.78	3	61.89	3	67.44	2	10	3
NLPE	48.67	3	51.78	3	48.89	3	57.56	10	6	3

overfitting (Muller *et al* 2003). It is likely that the use of different classifiers (i.e. nonlinear) might bring further improvement to the classification accuracy. The use of different classifiers for testing separability of these feature sets will be addressed in our future works.

4. Results and discussion

4.1. ECG data set

The separability of nonlinear features that involved state space reconstruction (cX, CD, LLE, APEN and NLPE) was investigated using three different embedding parameter selection methods. The first method was false nearest neighbors and minimum mutual information (FNN&MMI) which yielded an embedding dimension of $m = 5$ and time lag of $\tau = 8$. The second method was false nearest neighbors and first zero crossing of autocorrelation function (FNN&AC), yielding an embedding dimension of $m = 7$ and time lag of $\tau = 9$. The third method was minimum NLPE that estimates the embedding parameter pairs by minimization of NLPE; this method gave an embedding dimension of $m = 8$ and time lag of $\tau = 1$. The best accuracy of each feature among three recording leads is presented with corresponding parameter selection methods in table 1. Additionally, an exhaustive search was performed on the class separability of nonlinear features that were estimated with embedding dimension, m, of 2 through 10 and time lag, τ, of 1 through 10. The embedding parameters and lead giving the best classification accuracy for each feature is also presented in table 1.

Estimation of the HURST feature does not require any parameter selection and the REV feature only involves selection of time lag. An accuracy of 46.55% was obtained from lead 2 with the HURST exponent and an accuracy of 43.89% was obtained from lead 1 with the REV feature. The optimal time lag leading to highest accuracy was $\tau = 1$ for the REV feature.

The separability of AR features was investigated with model order 1 through 10. Using both AR reflection coefficients and model coefficients, an accuracy of 51.22% was obtained from lead 2 with an order of 1.

The results demonstrated that the best overall accuracy of 67.44% was obtained from the APEN feature with exhaustive search. The LLE feature with exhaustive search also provided an accuracy of 66.56% which was comparably good. Among the nonlinear features that employ embedding parameter selection methods, the best class separabilities of 62.89% and 63.56% were obtained from the LLE feature using parameters estimated with FNN&MMI and FNN&AC methods and an accuracy of 61.89% was obtained from the APEN feature using parameters estimated with the minimum NLPE method. The rest of the features provided poor classification accuracies. The findings show that the selection of embedding parameters is an

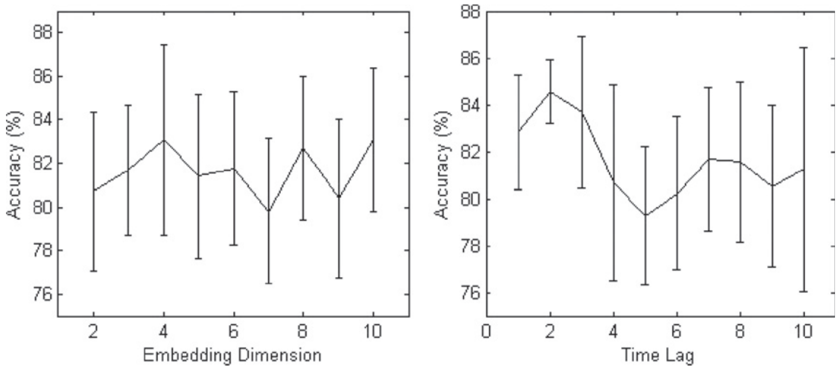

Figure 2. The classification accuracy versus embedding dimension and the classification accuracy versus time lag for three class ECG beats.

Table 2. SFFS-LDA classification results for combined feature sets—ECG.

FNN&MMI	FNN&AC	Minimum NLPE	Exhaustive search		
Accuracy	Accuracy	Accuracy	Accuracy	m	τ
80.22	77.11	84.89	88.00	10	10

important factor in the characterization ability of nonlinear features. However based on these results, none of the parameter selection methods can be reported as superior to others for all features as the parameter pairs that led to best class separability varied among the features.

The separability of three class ECG beats was further assessed using combined linear and nonlinear feature sets where the complimentary features were chosen using the SFFS algorithm with LDA. In order to create the feature spaces, the nonlinear features were extracted using embedding parameter pairs obtained from each method of parameter estimation (FNN&MI, FNN&AC, minimum NLPE methods) and combined with ARC and ARK features (orders 1 through 10) generating three independent feature sets which were searched for complimentary features. An exhaustive search was also performed on the nonlinear feature space. Ninety feature sets were generated where nonlinear features were estimated using embedding parameter pairs over embedding dimensions, m, of 2 through 10, and time lag, τ, values from 1 to 10. Each feature set also included ARC and ARK features (orders 1 through 10). The best classification accuracy among the SFFS-LDA results of 90 feature sets that were searched through is presented.

The results of SFFS-LDA classification with combined linear and nonlinear feature sets are shown in table 2. Combined linear and nonlinear features demonstrate a clear improvement in the class separability with all of the embedding parameter selection methods. Based on these results the minimum NLPE method could be accepted as superior to the conventional methods of embedding parameter selection, FNN&MMI and FNN&AC.

It should be noted that, the exhaustive search is a computationally expensive method of parameter selection; here, we have employed this method in order to investigate the highest and the lowest limit of class separability using nonlinear features estimated with different embedding parameter pairs. Figure 2 shows embedding dimension versus classification accuracy and also time lag versus classification accuracies for combined feature sets. The error bars were obtained by calculating the mean and the standard deviation of all accuracies with time lags of 1 through 10 for the corresponding embedding parameter for embedding dimension versus classification accuracy graph and vice versa. Figure 2 demonstrates that the selection of time lag and embedding dimension parameters is crucial for obtaining higher classification accuracies with combined feature sets as well. While the embedding dimension does not seem to have a significant effect on the classification accuracy, the smaller time lag values seem to be better for obtaining higher classification accuracies for this data set.

4.2. EEG data set I

The results of LDA classification for finger-movement versus idle state EEG segments are shown in table 3. For each subject, the best classification accuracy among linear and nonlinear features along with corresponding feature name, electrode site and parameters is presented.

The embedding dimension and time lag pairs were estimated individually for the subjects. The FNN&MI and FNN&AC methods led to the same embedding parameter pairs for all subjects. The embedding dimension was set to $m = 4$ for subject 10 and it was set to $m = 5$ for the rest of the subjects; time lag was set to $\tau = 2$ for subject 8, $\tau = 3$ for subjects 4 and 7, $\tau = 4$ for subjects 1, 2, 3, 9, and $\tau = 5$ for subjects 5 and 6 using both FNN&MI and FNN&AC methods. The minimum NLPE method gave an embedding dimension of $m = 8$ and time lag of $\tau = 1$ for all subjects.

The class separabilities of the best linear features were found to be significantly better than the best nonlinear features estimated with FNN&MMI, FNN&AC and minimum NLPE methods (paired *t*-test, $df = 11$, $p > 0.05$). Furthermore, the class separabilities obtained through nonlinear features with exhaustive search were found to be significantly better than the separability of nonlinear features estimated with FNN&MMI, FNN&AC and minimum NLPE methods (paired *t*-test, $df = 11$, $p < 0.005$) which was expected as the embedding parameter pair leading to highest class separability was selected using exhaustive search. However, no significant differences were found between nonlinear features with exhaustive search and linear features (paired *t*-test, $df = 11$, $p > 0.05$).

The separability of movement versus idle state EEG segments was further assessed using combined linear and nonlinear feature sets where the complimentary features were selected with the SFFS algorithm using LDA. The feature sets for each subject were created using ARC and ARK features (orders 1 through 10) combined with nonlinear features generated using embedding parameter pairs estimated by one of the parameter selection methods (FNN&MI/FNN&AC, minimum NLPE). This led to two independent feature sets to search through. Also an exhaustive search was performed by creating 90 independent feature sets to be searched using nonlinear features estimated with embedding dimension, m, of 2 through 10 and time lag, τ of 1 through 10, combined with ARC and ARK features (orders 1 through 10).

The class separability with combined feature sets using LDA classification is shown in table 4. In order to test the significant improvements between class separability across subjects for combined and individual features, a paired *t*-test with one tail was performed. The results showed a significant improvement in class separation across subjects for combined feature sets with FNN&AC, FNN&MMI, minimum NLPE and exhaustive search compared to best linear features (paired *t*-test, $df = 11$, $p < 0.001$). Also there were significant differences between the class separability of best nonlinear features estimated with FNN&AC and FNN&MMI and combined linear and nonlinear features with FNN&AC and FNN&MMI (paired *t*-test, $df = 11$, $p < 0.001$); best nonlinear features estimated with minimum NLPE and combined

Table 3. Best LDA classification results among linear and nonlinear features—finger-movement EEG.

	Linear features				Nonlinear features										
					FNN&MMI - FNN&AC			Minimum NLPE			Exhaustive search				
Subject no	Accuracy	Feature	Order	Site	Accuracy	Feature	Site	Accuracy	Feature	Site	Accuracy	Feature	m	τ	Site
1	60.84	ARC	10	C4	57.83	APEN	C4	56.64	APEN	C4	59.80	APEN	3	5	C4
2	69.16	ARC	10	C4	65.38	APEN	C4	64.83	APEN	C4	69.05	APEN	10	5	C4
3	55.37	ARC	2	Cz	55.49	APEN	C1	55.37	APEN	Cz	57.76	NLPE	6	5	C3
4	58.30	ARC	10	C4	57.89	APEN	C3	58.40	APEN	C3	59.94	APEN	2	10	C3
5	58.82	ARC	9	C4	53.92	APEN	C3	53.97	NLPE	Cz	56.21	cX	2	10	C4
6	88.39	ARC	8	C4	80.09	APEN	C4	81.72	APEN	C4	84.78	NLPE	3	1	C4
7	57.86	ARC	9	C4	54.75	NLPE	C2	54.88	APEN	Cz	56.71	APEN	4	5	C2
8	55.61	ARC	10	C3	52.63	APEN	C2	54.33	APEN	C2	55.01	APEN	9	7	C3
9	64.64	ARC	10	C4	55.43	APEN	C2	56.72	APEN	C4	57.94	NLPE	3	1	C2
Mean ± Std	63.22 ± 10.41				59.26 ± 8.63			59.65 ± 8.90			61.91 ± 9.50				

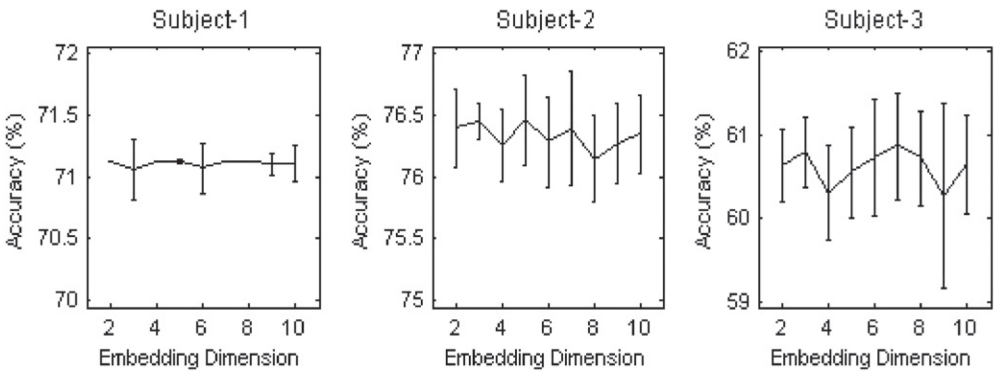

Figure 3. The classification accuracy versus embedding dimension for finger-movement EEG.

Table 4. SFFS-LDA classification results for combined feature sets—finger-movement EEG.

	FNN&MMI - FNN&AC	Minimum NLPE	Exhaustive search		
Subject no	Accuracy	Accuracy	Accuracy	m	τ
1	71.13	71.13	71.25	10	10
2	76.73	76.28	77.21	5	5
3	60.10	61.77	62.37	9	1
4	71.92	72.41	73.04	9	3
5	61.84	62.46	63.02	10	1
6	92.51	91.78	92.67	3	5
7	64.75	64.40	65.43	4	6
8	59.44	59.38	60.60	2	8
9	69.33	69.44	70.35	9	6
Mean $\pm$ Std	69.75 $\pm$ 10.37	69.89 $\pm$ 9.93	70.66 $\pm$ 9.93		

linear and nonlinear features with minimum NLPE (paired t-test, $df = 11$, $p < 0.001$); best nonlinear features with exhaustive search and combined linear and nonlinear features with exhaustive search (paired t-test, $df = 11$, $p < 0.001$).

Furthermore, the results have shown that there were no significant differences between the class separability of minimum NLPE, FNN&MI/FNN&AC and exhaustive search methods (paired t-test, $df = 11$, $p > 0.1$). Figures 3 and 4 demonstrate the embedding dimension versus classification accuracy and time lag versus classification accuracy for subjects 1–3. The graphs show that the amount of deviation in the classification accuracy based on embedding dimension and time lag is approximately 2%. These findings are consistent with subjects 4–9 as well. The results based on finger-movement EEG data set suggest that while the selection of embedding parameters were highly subject and feature dependent in the case of individual features, the selection of these parameters becomes less crucial with the use of combined linear and nonlinear feature sets.

4.3. EEG data set II

The results of nonlinear features that involved state space reconstruction are presented in table 5. Embedding parameter pairs for estimation of nonlinear features were set to $m = 6$,

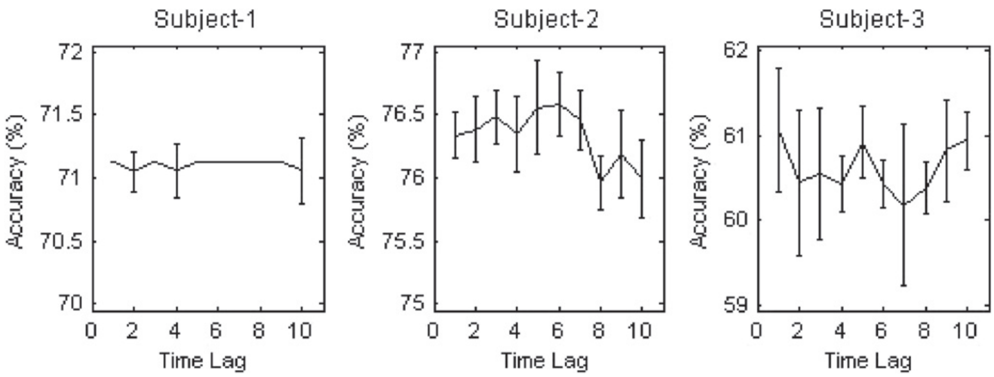

Figure 4. The classification accuracy versus time lag for finger-movement EEG.

Table 5. LDA classification results for nonlinear dynamic features—epileptic EEG.

	FNN&MMI	FNN&AC	Minimum NLPE	Exhaustive search		
Feature names	Accuracy	Accuracy	Accuracy	Accuracy	m	τ
cX	27.73	26.75	20.60	40.08	2	8
CD	38.10	34.48	35.00	39.20	4	3
LLE	38.00	20.75	28.30	48.38	7	3
APEN	39.75	38.10	42.00	48.00	2	1
NLPE	50.88	50.92	40.48	53.10	2	1

$\tau = 6$, using the FNN&MMI method; $m = 6$, $\tau = 10$, using the FNN&AC method and $m = 10$, $\tau = 1$, using the minimum NLPE method. The maximum class separability obtained through exhaustive search is also presented with corresponding embedding parameter pairs for each feature.

The Hurst exponent gave an accuracy of 40.72% and the REV feature gave an accuracy of 33.87%. The optimal time lag that led to highest accuracy using the REV feature was $\tau = 1$.

Using linear features, an accuracy of 74.78% and 48.25% was obtained with AR model coefficients (order 10) and AR reflection coefficients (order 3), respectively.

Overall results demonstrate that the highest accuracy of 74.78% was obtained with ARC features whilst the ARK features and nonlinear features provided poor classification accuracies. Class separability of nonlinear features presented in table 5 shows that none of the embedding parameter selection methods could consistently provide the best accuracy for all features. The results of exhaustive search also demonstrate that the parameter pairs that led to highest class separability were variable among features.

The classification accuracies obtained with combined linear and nonlinear features are shown in table 6. The feature sets that were searched through were created using four different settings which was explained in section 4.1. For this data set the use of combined linear and nonlinear features led to an improved class separability of 81.40% through exhaustive search. The separability of combined features that employed embedding parameter selection methods for nonlinear features was also comparable to the exhaustive search. Figure 5 demonstrates the embedding dimension versus classification accuracy and time lag versus classification accuracy graphs obtained through exhaustive search. The graphs illustrate that all of the embedding parameter pairs provide comparable classification accuracies ($\pm 2\%$) for combined

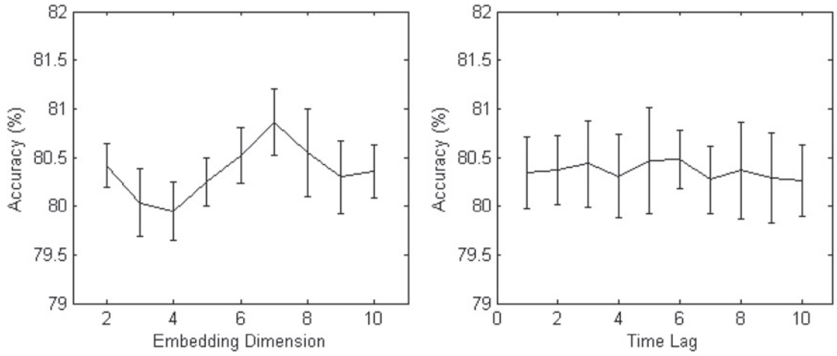

Figure 5. The classification accuracy versus embedding dimension and the classification accuracy versus time lag for five class EEG.

Table 6. SFFS-LDA classification results for combined feature sets—epileptic EEG.

FNN&MMI	FNN&AC	Minimum NLPE	Exhaustive search		
Accuracy	Accuracy	Accuracy	m	τ	Accuracy
80.58	80.25	80.67	7	5	81.40

Table 7. Estimated computation time for each feature.

Feature name	Estimated time (ms)
APEN	41.24
CD	42.51
HURST	12.32
cX	0.34
REV	0.43
NLPE	62.60
LLE	108.55
AR	2.31

linear and nonlinear feature sets. This implies that the selection of the embedding parameter and time lag did not have a significant effect on the separability of combined feature sets for this data set.

4.4. The applicability of linear and nonlinear features for real-time processing

The estimated computation time for extraction of each feature is listed in table 7. These time measures were obtained by running the algorithms on MATLAB (R2009b) using a dual core computer with 2.00 GHz processing speed. Algorithms were run using a sample signal selected from EEG data set I, with a window length of 256 data points. Embedding dimension and time lag parameters for nonlinear features were selected as estimated parameters using the minimum NLPE method (see section 4.2). Note that the time measures do not differ based on different embedding parameter selections. The time measures listed in table 7 illustrate that the estimated computation time of each feature vary based on the complexity of the algorithm. The difference in computational time between linear and nonlinear features was small and negligible considering the improvement in classification accuracy. Overall, the listed estimated computation times are acceptable for real-time processing of EEG/ECG signals. It should be noted that the algorithms for nonlinear features were implemented by the authors and the features were estimated with personal computers. Thus, in the case of real-time design and implementation, the algorithms could be further improved in terms of processing time. Besides use of the dedicated hardware (in a real-time system) for extraction of these features would also bring improvement to the processing times.

5. Conclusion

In this study the characterization ability of linear AR features and nonlinear features was investigated using three data sets composed of ECG, epileptic EEG and finger-movement EEG signals. The linear and nonlinear features were tested by assessing their ability to differentiate between different classes of data.

On their own, the linear and nonlinear features provided comparable results with ECG and finger-movement EEG data sets whilst the linear features provided a better class separability with the epileptic EEG data set. Furthermore, the parameter selection methods and exhaustive search on the parameter space for estimation of nonlinear features demonstrated that the optimal embedding dimension and time lag pairs that provide highest class separability were feature dependent. This implies that none of the parameter selection methods can be proposed as superior to others for all features.

The results based on combined linear and nonlinear features demonstrated significant improvements in the class separability for all of the data sets. Furthermore, no significant differences were found between the class separabilities of combined linear and nonlinear feature sets where the nonlinear features were estimated with embedding parameters generated by FNN&AC, FNN&MI and minimum NLPE methods for finger-movement and epileptic EEG segments. In other words, the class separabilities of combined feature sets in which the embedding parameters for estimation of nonlinear features were investigated through exhaustive search have shown that the selection of embedding parameter pairs does not have significant effect on class separability. This is true for both epileptic EEG and finger-movement EEG data sets where the variability of class separation was approximately 2% with all combinations of embedding parameter pairs. For the ECG data set, the results has shown that the use of the minimum NLPE method provided comparable results to the highest class separability obtained through exhaustive search.

Furthermore, variability in the class separabilities based on different data sets could be related to the nature of the time series. In our previous studies, we have shown that the indications of nonlinearity in the EEG and ECG data sets vary depending on the features utilized to investigate nonlinearity and also depending on different classes of the data sets. In other words, different nonlinear features are able to characterize different properties of the signals on varying rates. This is directly related to the discrimination ability of these features on the corresponding time series. Therefore, it could be concluded that the varying class separabilities based on different data sets might be dependent on the linear or nonlinear nature of time series in the corresponding data set. The relationship between the characterization ability of different nonlinear features and the nature of the signals will be further explored in our future works.

Overall findings suggest that the utilized linear AR features and nonlinear dynamic features are complimentary features that quantify different properties of the corresponding EEG and ECG signals, providing an improved class separability compared to individual features.

References

Q2

Akay M 2001 *Nonlinear Biomedical Signal Processing* vol II (Piscataway, NJ: IEEE)

Anderson C W *et al* 1998 Multivariate autoregressive models for classification of spontaneous electroencephalogram during mental tasks *IEEE Trans. Biomed. Eng.* **45** 277–86

Andrzejak R G *et al* 2001 Indication of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state *Physica* D **64** 0619071–8

Balli T and Palaniappan R 2009 Minimising prediction error for optimal nonlinear modelling of EEG signals using genetic algorithm *Proc. 4th Int. IEEE EMBS Conf. on Neural Engineering (Antalya/Turkey)* pp 364–6

Beasley B M 2003 *Understanding EKGs: A Practical Approach* (Englewood Cliffs, NJ: Prentice-Hall)

Benitez R *et al* 2009 Characterisation of the nonlinear content of the heart rate dynamics during myocardial ischemia *Med. Eng. Phys.* **31** 660–7

Berryman J M *et al* 2005 Nonlinear aspects of the EEG during sleep in children *Proc. SPIE: Fluctuations and Noise in Biological, Biophysical, and Biomedical Systems* pp 40–8

Burg J P 2005 Nonlinear aspects of the EEG during sleep in children *Mod. Spectr. Anal.* 42–8 Q3

Casdagli M C *et al* 1996 Characterising nonlinearity in invasive EEG recordings from temporal lobe epilepsy *Physica* D **99** 381–99

Chua K C *et al* 2008 Automatic identification of epilepsy by HOS and power spectrum parameters using EEG signals: a comparative study *30th Annual Int. IEEE EMBS Conf. (Vancouver/Canada)* pp 3824–7

Czigler B *et al* 2008 Quantitative EEG in early Alzheimer's disease patients—power spectrum and complexity features *Int. J. Psychophysiol.* **68** 75–80

Dyson M *et al* 2008 Approximate entropy for EEG based movement detection *The 4th Int. Workshop on Brain–Computer Interfaces (Graz/Austria)*

Epileptic EEG Database. Bonn University: EEG Database Available at http://www.meb.unibonn.de/epileptologie/science/physik/eegdata.html

Gautama T *et al* 2003 Indications of nonlinear structures in brain electrical activity *Phys. Rev.* E **67(4)** 0462041–5

Gautama T *et al* 2004 The delay vector variance method for detecting determinism and nonlinearity in time series *Physica* D **3–4** 167–76

Geng T *et al* 2008 Novel design of 4-class BCI using two binary classifiers and parallel mental tasks *Comput. Intell. Neurosci.* **437306** Q4

Greene B R *et al* 2008 A comparison of quantitative EEG features for neonatal seizure detection *Clin. Neurophysiol.* **119** 1248–61

Kantz H and Schreiber T 1997 *Nonlinear Time Series Analysis* (Cambridge: Cambridge University Press)

Lee Y *et al* 2001 Detection of non-linearity in the EEG of schizophrenic patients *Clin. Neurophysiol.* **112** 1288–94

Lehnertz K 2008 Epilepsy and nonlinear dynamics *J. Biol. Phys.* **34** 253–66

Mandic D P *et al* 2008 On the characterisation of the deterministic/stochastic and linear/nonlinear nature of time series *Proc. R. Soc.* A **464** 1141–60

MIT-BIH Arrhythmia Database. PhysioBank, signal archives for biomedical research. Available at http://www.physionet.org/physiobank/database/

Muller K R *et al* 2003 Linear and nonlinear methods for brain–computer interfaces *IEEE Trans. Neural Syst. Rehabil. Eng.* **11** 165–9

Olbrich E *et al* 2003 Dynamics of human sleep EEG *Neurocomputing* **52–54** 857–62

Owis M I *et al* 2002 Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification *IEEE Trans. Biomed. Eng.* **49** 733–6

Pan J and Tompkins W J 1985 A real-time QRS detection algorithm *IEEE Trans. Biomed. Eng.* **32** 230–6

Phothisonothai M and Nakagawa M 2008 EEG-based classification of motor imagery tasks using fractal dimension and neural network for brain–computer interface *IEICE Trans. Inform. Syst.* E **91-D** 44–53

Pincus S M 1991 Approximate entropy as a measure of system complexity *Proc. Natl Acad. Sci.* **88** 2297–307

Pires G *et al* 2007 Single-trial EEG classification of movement related potential *Proc. IEEE 10th Int. Conf. on Rehabilitation Robotics (Noordwijk/Netherlands)* pp 569–74

Pudil P *et al* 1994 Floating search methods in feature selection *Pattern Recognit. Lett.* **15** 1119–25

Rosenstein M T *et al* 1993 A practical method for calculating largest lyapunov exponents from small data sets *Physica* D **65** 117–34

Sabeti M *et al* 2009 Entropy and complexity measures for EEG signal classification of schizophrenic and control participants *Artif. Intell. Med.* **47** 263–74

Shen Y *et al* 2003 Dimensional complexity and spectral properties of the human sleep EEG *Clin. Neurophysiol.* **114** 199–209

Small M *et al* 2002 Uncovering non-linear structure in human ECG recordings *Chaos Solitons Fractals* **13** 1755–62

Sprott J C 2003 *Chaos and Time-Series Analysis* (Oxford: Oxford University Press)
Takens F 1981 Detecting strange attractors in turbulence *Lect. Notes Math.* **898** 366–81
Tamil E B M *et al* 2008 A review on feature extraction & classification techniques for biosignal processing: part I. Electrocardiogram) *4th Kuala Lumpur Int. Conf. on Biomedical Engineering 2008* vol 21 pp 107–12
Zhang A *et al* 2008 Feature extraction of EEG signals using power spectral entropy *Proc. 2008 Int. Conf. on BioMedical Engineering and Informatics* vol 02 pp 435–9