AUTOMATIC SLEEP SCORING BASED ONLY ON ELECTROCARDIOGRAM RECORDS

Tarik Al-ani[1,2], Radek Kazbunda[3], Daniel Novák[3]

[1]Université Paris-Est, Groupe ESIEE-Paris, Laboratoire A2SI
Cité Descartes, 93162 Noisy-Le-Grand, FRANCE
[2] Université de Versailles Saint Quentin, Laboratoire LISV
10-12 Avenue de lEurope, 78140 Velizy, FRANCE
[3] Czech Technical University in Prague, Department of Cybernetics, Czech Republic

t.alani@esiee.fr (Tarik AL-ANI)

Abstract

The technique used in the most sleep laboratories for the diagnosis of the sleep disorders is the polysomnography. This technique is of great discomfort for the patient (hospitalisation, sleeping in a non familiar environment, connected to many sensors and cables). The fundamental signals for assessing the quality of sleep can be recorded. Evaluating these signals in 30 seconds interval is time consuming even for experience physician. Because that these signals are recorded in real time and in digital form, and because that the diagnosis is made directly from these records, they can thus be used for automatic processing. One of the most important problems in ECG analysis is the extraction of appropriate features, and this can be tackled in various ways. The aim of this work is to automatically classify sleep stage using only the electrocardiogram (ECG) records and using the conventional R&K classification criteria. The feature extraction stage of the work described in this paper was performed using methods of Detrended Fluctuation analysis and Heart Rate Variability analysis. All these methods are based on analysis of a Tachogram (record of RR intervals). Feature-spaces formed using these two methods were used as input to a Artificial Neural Network (ANN). Our approach has been tested on a real ECG records from different patients demonstrating the feasibility of the proposed method. The capability to differentiate sleep stages in predefined categories (wake, light sleep, deep sleep, REM) was successful in 65%. The Classification performed on data set containing only deep sleep and REM categories had 83.4% reliability.

Keywords: Electrocardiogram, Sleep scoring, Heart rate variability, feature extraction, Classification, Neural networks.

Presenting Author's Biography

Dr. Tarik AL-ANI received his Bachelor of Science (B.Sc.) in Physics (1973), Electrical Engineering Diploma (1975) from the University of Baghdad then PhD from the University Paris XI in 1984. He is currently Associate Professor in the high Engineering school ESIEE-ASI Laboratory and Associate researcher in the University of Versailles-LISV Laboratory in France. His research area of interests include: Stochastic modeling approaches, Artificial Intelligence, Identification, Control and Biomedical Engineering.

ISBN 978-3-901608-32-2 *Copyright © 2007 EUROSIM / SLOSIM*

1 Introduction

Wake/sleep complaints are second only to complaints of pain as a cause to seek medical attention [1]. Undiagnosed and untreated wake/sleep complaints exact an enormous toll at the personal level in terms of misery and at the societal level in socioeconomic consequences. More than 25% of all European citizens suffer from sleep disorders, like sleep apnea and insomnia.

For the diagnosis of the most frequent sleep disorders the polysomnography in a sleep lab is of great discomfort for the patient (hospitalization, sleeping in a non familiar environment, connected to many sensors and cables, etc.). The fundamental signals for assessing the quality of sleep can be recorded in a non invasive and comfortable way at ambulatory care or at home, and analyzed and scored by a sophisticated software under the supervision of a sleep specialist. The final analysis could help in the choice of the proper treatment and in the identification of a few specific cases that really need a further investigation through a complete polysomnography.

What is urgently needed is to reduce overwhelming number of sleep disorders candidates by means of very simple-to-use, comfortable and cheap methodology. Apart from quite a big number of existing commercial ambulatory systems and racket development of telemedicine which is foreseen as a key element leading for reduction of cost [2] in medical sector, the current situation is still alarming in many countries. Furthermore, with unavoidable scenario of huge increase of elderly population in near future [3], there is a big demand for new approaches based on well-measured vital signals

Many research has been undertaken in automatic sleep scoring in order to reduce analysis time and increase the reliability in the diagnosis results. Most of these research are based on electroencephalogram (EEG) analysis using neural networks, genetic algorithm or stochastic modeling approaches [4, 5, 6, 7, 8, 9, 10]. However, in the case of the EEG, the effects of scalp, fluid and bone on the tiny electrical currents generated in the cortex may be modeled only poorly and the large size of scalp electrodes and the effects of muscle and instrument noise all contribute to the complexity of EEG analysis. The aim of this work is to automatically classify sleep stage using only the electrocardiogram (ECG) records and using the R&K classification criteria. One of the most important problems in ECG analysis is the extraction of appropriate features, and this can be tackled in various ways. The feature extraction stage of the work described in this paper was performed using methods of Detrended Fluctuation analysis and Heart Rate Variability analysis [11, 12, 13, 14]. All these methods are based on analysis of a Tachogram (record of RR intervals). Feature-spaces formed using these two methods were used as input to a Artificial Neural Network (ANN) [15].

This paper is structured as follows. Section 2 presents a brief introduction to the conventional R&K sleep scoring. Section 3 describe our approaches for automatic sleep stage classification from ECG. The performance of the proposed approaches is demonstrated in section 4.

2 Sleep and different sleep stages

Sleep is a state of natural rest and is necessary for physical and psychical recovery. Sleep is often regarded as test situation for autonomic nervous system. Length of sleep necessary for recovery can differ greatly depending on age, etc. According to several studies the cognitive and physical performances are reduced with fewer than eight hours of sleep [16]. Early sleep researchers noted that the electroencephalogram (EEG) waveforms changed as a subject passed from wakefulness to sleep. Not long after this, sleep researchers Rechtschaffen and Kales (R&K) [17] developed the specific rules that are used today in the scoring of sleep. Sleep stage scoring depends on the recognition of specific characteristic EEG waveforms recorded on a device known as a polygraph. In general, two types of sleep stages are considered: REM (Rapid Eye Movement) and Non-REM which include four different sleep stages as show in Fig. 1.

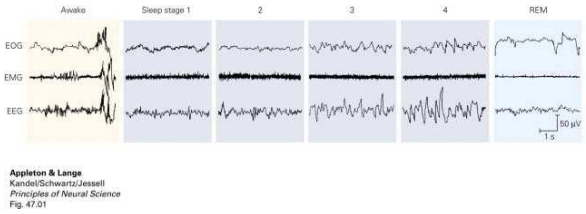


Fig. 1 Manual R&K Sleep Scoring System.

The classification is performed for each 30 seconds epoch. The reason is historic: At a paper speed of 10 mm/s, 1 page equates to 30 seconds and is defined as 1 epoch. Computerized polysomnography usually displays one video screen as one 30-second epoch.

In this work we investigate an automatic sleep scoring which is based only on the ECG records to replicate the conventional R&K sleep scoring.

3 Methods

Our aim is to develop an automatic sleep stage classification based on ECG signal analysis. To simplify this task, some sleep stages were merged in one category (see section 3.4). The classifier should be able to classify each 30 seconds of record in one of the four sleep categories. To accomplish this task, three steps must be performed: choose proper methods for feature extraction from ECG, then prepare feature sets and compare performance of features on large dataset and finally choose proper classifiers and compare their performance

3.1 Automatic sleep stage classification

Sleep stage influences autonomic nervous system. This influence can be investigated using features based on

 Copyright © 2007 EUROSIM / SLOSIM

the analysis of ECG signal. These features use usually a Tachogram. Several approaches have been tested and proposed for this analysis: Detrended Fluctuation Analysis (DFA), Heart Rate Variability (HRV) [13], Progressive Detrended Fluctuation analysis (PDFA) [18].

Tachogram can be obtained from ECC signal, but as well it can be obtained from pulse record. Using pulse may be deceptive when some heart beats do not have much cardiac output.

Classifying each interval at the basis of these features is a pattern recognition problem. To perform classification of ECG record intervals, the feed-forward artificial neural networks (FFN) and Elman artificial neural networks (ELN) approach was adopted.

3.2 Electrocardiogram (ECG)

An electrocardiogram (ECG) is a recording of the electric potential, generated by the electric activity of the heart, on the surface of body, Fig. 2.

Fig. 2 Electrocardiogram

Electrocardiogram is a basic tool for Cardiac electrophysiology, for studying the cardiac mechanisms and for studying the performance of the electrical activities of specific regions of the heart.

ECG signal is governed by autonomous nervous system, this common source is the cause of correlations with breathing [19] and can be source of the correlation with the different sleep stages [14, 20].

3.3 ECG based methods

There are a number of methods for the processing of ECG signal, but they have to be analyzed in order to decide whether they are useful for automatic estimation of sleep stage using ECG signal.

There are few difficulties in processing of the ECG signal which are imposed by its biological origin: Heart rate has many individual components and is driven by competitive forces (sympathetic and parasympathetic) and more over there are more of regulation mechanisms. This causes the creation of complex fluctuations. These fluctuations are not simply the result of responses on external factors, but they are persistent during physical load, rest and sleep. This non-stationarity is common for stochastic processes and therefore imposes that similar methods of processing may be used such as Heart Rate Variability (HRV), Detrended Fluctuation Analysis (DFA), Progressive Detrended Fluctuation Analysis (PDFA), Heart Rate Morphology, Multiscale Entropy Analysis and Information-Based Similarity.

3.3.1 Heart rate variability

Heart rate variability is a term used for the interpretation of oscillation in the interval between consecutive heart beats and oscillations in consecutive instantaneous heart rates. Heart rate variability covers large number of methods but not all of them are suitable for analyzing short intervals of records in which we try to detect sleep stages. As a referential specification and interpretation guide, see [12]. All the measurements have to be performed using a Tachogram.

HRV can be divided in two major groups of statistics: Time domain methods and Frequency domain methods.

Time domain methods

These methods employ Statistical or Geometrical methods for gathering features from RR record. Generally Geometrical methods require large number of samples (more than 20 minute interval) and therefore are unusable for sleep stage detection. Similar problem appears for some of statistical features, see the commentary of each of them below. These were selected as usable for sleep stage detection.

Selected variables

- *RMSSD*: square root of mean squared differences of successive NN intervals. This short term measurement estimates high frequency variations in heart rate.
- *SDNN*: standard deviation of RR interval reflects all the cyclic components responsible for variability in the period of recording. Because total variance of HRV increases with the length of the analyzed recording SDNN is dependent on the analyzed cycle length. Using this feature, this will make that classifier to work properly only on intervals of the same length as that used for training.
- *RRmean*: mean of RR interval length
- *NN50*: number of intervals longer than 50 ms, this feature can be substituted by RMSSD (there is power law relation between them), whose usage is encouraged because of its statistical properties. The interchangeability should be verified and therefore it was included among used features.

Unsuitable variables

- *SDANN* - requires long intervals
- *pNN50* - can be substituted by RMSSD.

Frequency domain methods

Variety of spectral methods are applied to Tachogram. Power spectral analysis (PSD) gives us an estimate of power distribution as a function of frequency. All of the calculated values can be used for sleep stage classification (see Tab. 1 for detailed description). The description nomenclature given in Table 1 are defined as follows. Var: Variance of NN over the temporal segment,

ISBN 978-3-901608-32-2 Copyright © 2007 EUROSIM / SLOSIM

PL: Power in low frequency spectrum, PH: Power in high frequency spectrum, PVL: Power in very low frequency spectrum, LFN: LF power in normalized units given by

$$\frac{LF}{Total\,power - VLF} \times 100,$$

HFN: HF power in normalized units given by

$$\frac{HF}{Total\,power - VLF} \times 100,$$

R: Ration given by

$$\frac{LF}{HF}.$$

m, s and n.u. are the abbreviations of meters, seconds and normalized units respectively.

Tab. 1 Table of frequency variables for short time analysis. V: Variable, U: Units, D: Description, F: Frequency range [12].

V	U	D	F
5 min. total power	ms^2	Var	$\leq 0.4\ Hz$
VLF	ms^2	PVL	$\leq 0.04\ Hz$
LF	ms^2	PL	0.04-0.15 Hz
LF norm	n.u.	LFN	
HF	ms^2	PH	0.15-0.4 Hz
HF norm	n.u.	HFN	
LF/HF	-	Ratio	

The technical specification of power spectral components is summarized in Tab. 2.

Tab. 2 Approximate correspondence of time domain and frequency domain variables. TD: Time domain variable, AFDC: Approximate frequency domain correlate.

TD	AFDC
SDNN	Total power
RMSSD	HF
NN50	HF

3.3.2 Detrended Fluctuation Analysis

DFA is method for quantifying the correlation property in non-stationary time series based on the computation of time dependent fluctuation function $F(N)$. The ability to detect long-range correlations in non-stationary time series is the biggest advantage over conventional methods for RR analysis. It also permits to avoid spurious detection of non-stationarity and noise artifacts.

The correlation property of $F(N)$ function is usually expressed by scaling exponents α (see below the details) defined for different ranges of analyzed heartbeats [21].

Conventional DFA-1 computation

To compute fluctuation function $F(n)$ from time-series $x(i)$ $[i = 1, ..., N]$, the time series is first integrated:

$$y(k) = \sum_{i=1}^{k} [x(i) - M]$$

where M is the average value of the series $x(i)$, and k ranges between 1 and N.

Next, the integrated series $y(k)$ is divided into boxes of equal length n and the local trend $y_n(k)$ fitting the data in each box is calculated. The integrated time series is detrended by subtracting polynomial local trend $y_n(k)$, see Fig. 3, then the root-mean square fluctuation of the detrended series is computed and finally the fluctuation function :

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N} [y(k) - y_n(k)]^2}$$

$F(n)$ is computed for all time-scales n. Typically, $F(n)$ increases with n, the "box-size". If $logF(n)$ increases linearly with $log(n)$, then the slope of the line relating $F(n)$ and n in a log-log scale gives the scaling exponent α, see example in Fig. 4

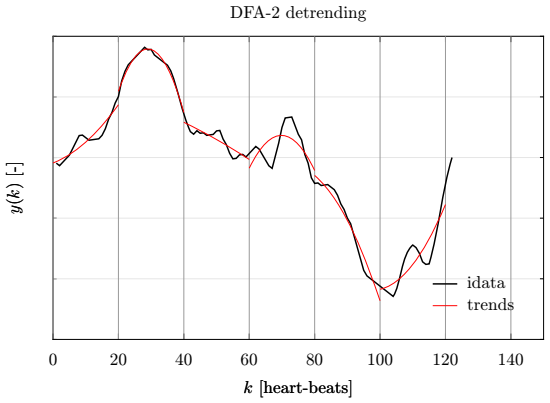


Fig. 3 Local detrending in the DFA algorithm.

According to random walk theory for uncorrelated data is the scaling exponent $\alpha = 0.5$. If $\alpha = 1.0$, the correlation of the time-series is the same as 1/f noise. If $\alpha = 1.5$, $x(i)$ behaves like Brown noise. If the scaling exponent is in the interval $\alpha = (0.5, 1\rangle$, then there is persistent long-range power-law correlation, such that large RR interval is likely to be followed by large interval. In contrast scaling exponent in interval $\alpha = (0, 0.5)$ indicates power-law correlation when the large values of time series are followed by small ones, see example in Fig. 4

DFA of higher orders

DFA of higher orders can be used to eliminate the noise and to refine the results obtained from DFA-1. The DFA of higher orders has ability to eliminate effects of trends of lower orders [22]. This means that the fluctuation function of DFA-l will not be influenced by polynomial

 Copyright © 2007 EUROSIM / SLOSIM

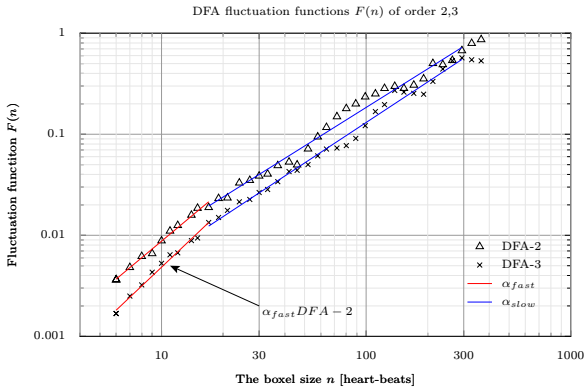


Fig. 4 DFA scaling exponents fitted in RR interval series. Fitted lines for scaling exponents $\alpha_1 = 1$; $\alpha_2 = 1.476$.

trend of lower order than l. DFA enables us to quantify correlations in noisy signals embedded in polynomial trends, moreover if scaling and crossover features of $F(n)$ are used we can also determine the order of trends and also to detect sinusoid trend.

The order of DFA is specified by order of local trend polynomial function used for fitting the integrated data series in detrending step.

3.4 HRV and DFA sleep stage classification

The classification will be performed using the features prepared by DFA and HRV. The classification must be based on non-invasive measurements and for each 30 seconds of record has to be classified in one of the following groups:

- Deep sleep: group includes sleep stage 3 and sleep stage 4 where stage 3 is considered as transitional to stage 4.
- Light sleep: group includes sleep stage 1 and sleep stage 2. Both sleep stages lack presence of B waves. Patient is easy to wake up.
- REM sleep: includes only REM sleep stage.
- Wake state: in this period none of sleep stage can be classified.

This work does not assume any manual modification of signal. Usually ECG record contains some artifacts which cannot be avoided, e.g. movements. If these cannot be handled by artifact filtering and QRS complex detection algorithm, the interval is invalidated. According to references, the movements occur often in sleep stage change, this may be possibly used in future as a feature for classification [14, 17].

3.4.1 Classification schema

Following the capabilities requirements of the classifier, the classification in our framework can be divided into several steps: signal preprocessing, selection of diagnostic intervals and data feature extraction and sleep stage classification in one of four specified groups: Deep sleep, Light sleep, REM sleep and Wake state. Fig. 5 gives the classifier's schema based on the classifier's specification.

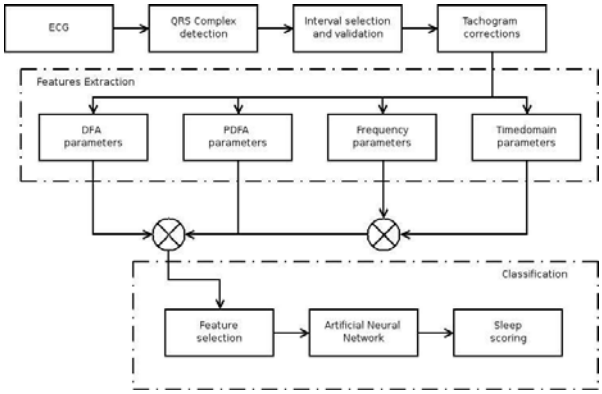


Fig. 5 Sleep stage classifier schema.

Signal pre-processing

Because of the requirement to classify each 30 seconds, it is very important to select proper diagnosis interval length. Many methods used for feature extraction (e.g. Heart Rate Variability (HRV), Detrended Fluctuation Analysis (DFA)) require relatively long record in order to give reliable values [12, 14]. There are several possible solutions: use only features which are disposable at short intervals or try some averaging from longer intervals. We have selected only features accurate enough on short intervals. The basic solution uses only 30 seconds interval length to compute features, improper methods of feature extraction were omitted.

The other problem is to do an accurate QRS complex detection. Using short intervals, this problem gain importance: Missing QRS complexes degrade quality of extracted features and can cause failure of certain features extraction algorithms (these algorithms demand minimal number of samples) [23]. Therefore a simple condition for invalidation of intervals was implemented: Algorithm expects QRS complex detection to be quite successful: from first 15 minutes counts average number of heartbeats for 30 second interval and all intervals, which don't overcome limit of 66% of this value are invalidated.

Feature extraction

The first step for successful classification is to perform a good feature extraction. This step is essential for further classification. A good feature extraction can greatly increases reliability of the final classification. The goal of this step is to create large set of features which can be used for classification in step 2. This set is distributed in smaller sets that can be used for training, cross-validation and test as described in ANN section. The training will be performed several times using different features combinations. The performance of several features combinations is then compared.

 Copyright © 2007 EUROSIM / SLOSIM

Classification

In our approach, we use only the ECG records for sleep stage classification. The classification will be performed by Artificial neural networks (ANN), see next section. There are numerous HRV based approaches for sleep stage and sleep apnea classification [13, 18].

3.4.2 Feature extraction

To follow the classifier requirements as specified in the beginning of section 3, it is necessary to choose variables which can be used for very short record intervals. Ideal value would be 30 seconds or even shorter, but this requires reliable QRS complex detection, because each missed QRS complex (even corrected) can greatly degrade the statistics. After pre-processing of ECG signal (see section 3.4.1) the features, as specified bellow, are extracted and saved in database.

Spectral components

Every sleep stage has its own characteristic spectrum components. The problem is how to define this specificity. We inspired us by a table of characteristic spectra given in the paper [13]. To compute frequency values we used parametrical method "pburg" of order 9, the propriety of selection of this order can be verified by "arfit toolbox" [24, 25]. This selection is a compromise, because the proper order may vary. This is caused by short interval length and by the natural diversity in heart rates and cannot be avoided. At basis of spectrum presented for each sleep stage we selected the following features (see section 3.3.1): *Low Frequency/Hight Frequency (LF/HF)* - ratio of this features, *Total power (TP)* - total power for analyzed interval, *Low Frequency (LF)*, *High Frequency (HF)*, *Very Low Frequency (VLF)*, *normalized LF* and *normalized HF*.

Time domain measurements

Generally Time domain measurements are hard to interpret, especially for short term recordings. Only methods for short time intervals were chosen [12].

Selected features

The selected features are: *RMSSD* for all lengths of intervals, *SDNN* for all lengths of intervals, *RRmean* - mean of RR interval length over analyzed interval and *NN50* - this feature is deprecated and probably it can be easily substituted by RMSSD.

A Genetic algorithm [26] was implemented in our work to select the best feature combination (see section 4).

Detrended fluctuation analysis

Generally DFA needs a longer analyzed intervals due to the logarithmic dependence of scaling exponent α. Normally the scaling exponent is computed for several boxel size ranges n, but the size of the analyzed interval of 30 seconds does not permits fitting of the second independent scaling exponent, therefore 2 overlapping exponents were chosen. The basic boxel size was set to 6 to remove degradation of scaling coefficient caused by detrending at small number of samples. This effect is stronger for higher order of DFA. The default DFA order for analysis was set to 2 and for comparison we added a scaling coefficients of DFA-3, in order to investigate the impact of steeper slop of $\alpha_1$coefficient.

- $DFA-2\alpha_{fast}$ is computed from the basic boxel size $n = 6$ to boxel size $n = 16$, DFA order is 2, the used computation step is 1.2
- $DFA-2\alpha_{total}$ scaling coefficient computed from basic boxel size $n = 6$ to boxel size of the total number of samples in the analyzed interval, DFA order is 2, the used computation step is 1.2
- $DFA-3\alpha_{fast}$ is computed from basic boxel size $n = 6$ to boxel size $n = 16$, DFA order is 3, the used computation step is 1.2
- $DFA-3\alpha_{total}$ scaling coefficient computed from basic boxel size $n = 6$ to boxel size of the total number of samples in the analyzed interval, the used for computation step is 1.2.

Data

The raw data records used in our work are real data issued from the polysomnograph of Medatec company [27] which is used in some hospitals in France, particularly in the Hospital Raymond Poincar-Garches [28]. In this preliminary study, four ECG records corresponding to four patients [27, 28] were used. These all night recordings sampled at 200 Hz were scored according to the R&K rules. The authors actually extend the results to include more records corresponding to different patients.

4 Results

The purpose of this section is to assess whether the chosen ANN inference approaches together with our feature extraction approach for getting reliable classifications is useful in the context of sleep staging.

ANN training

As mentioned above, this classification task is rather a pattern recognition problem and this influences possible training mechanisms. In this study, feed-forward neural networks (FFN) and Elman neural networks (ELN) were used as a classifiers. The speed of convergence at larger networks should be also considered as a criterion for algorithm selection, slow training algorithm would prevent comparison of a different training data sets. To satisfy these criteria, the following Matlab ANN toolbox algorithms [29] were used to construct our ANN: "trainrp", "trainscg", "trainoss" and "traingdx".

Choosing ANN structure

To ensure good recognition it is preferred to construct the first layer of ANN such that the dimensions of this layer and the input vector are the same. Similarly the dimensions of the output layer is the same as that of the output states vector (4 states, see output definition in section 3.4). The output function is proposed to be log

 Copyright © 2007 EUROSIM / SLOSIM

of the sigmoid function (log-sig) to ensure that the output is in range of (0,1). The output state vector is composed of the four states: wake,light sleep,deep sleep and REM.

In our work, a simple method to create ANN with different topology of inner layers has been developed. The method generates ANN by adding hidden layer with different number of neurons to the current ANN scheme used for training. These networks were trained several times and the medium performance function on training data, medium performance on test data and the best reached performance on test data were captured and plotted in graphs.

ANN types comparison

Two types of ANN were proposed: feed-forward (FFN) and Elman neural ANN. ELN is based on FFN, but due to recurrent connection between hidden layer and input layer it can learn temporal patterns. This capability may prove as useful.

4.1 Sleep stages classification

In order to choose the best ANN topology for our study, a simple test was performed. During this test ANNs with different topology of the inner layers are created. The results of this test has been used to select the best performing networks. These networks are Elman network with one hidden layer and FFN with one hidden layer. To get the best performance, during training process, a cross-validation data set was specified. This data set is not used for training but, the performance of trained function is measured on it and if the performance on cross-validation data get worse for defined number of consecutive training iterations the training is stopped. This procedure is called early stopping. The limit was set to 20 iterations in which the performance degrades.

4.1.1 Features property comparison

In the subsection 3.4.2, suitable features were selected for a robust data representation at the basis of theoretical analysis (see section 3.3). This does not necessary means that they are proper to perform ANN learning. Some combinations of these features perform better than others. Some combinations are nearly equivalents, because they express the same variation property.

Tab. 3 and 4 show some combinations that we used to initialize the GA, where the features combination vector is defined, from left-to-right successively, as a binary vector which specifies presence or absence of the following features at the input of ANN (see section 3.4.2): RMSSD, NN50,HF,HFn, $2\alpha_{fast}$, $2\alpha_{total}$, $3\alpha_{fast}$, $3\alpha_{total}$, TP, SDNN, VLF, LF, LFn, RRmean, LF/HF.

The performance for the estimated feature combinations was estimated on 3 different data sets. Each data set contains training data, cross-validation data and test data.

DATA SET-1 is randomly selected data set, which containing only deep sleep and REM episodes. This data set correspond to the simplest task because they may be clearly classified and therefore it is suitable to evaluate the fitness of features for sleep stage estimation. The results for the created Elman ANNs are given as an example in Tab. 5.

Tab. 3 Basic feature combinations.

ID	Features Combination vector
1	$ch(1,:) = [1,0,0,1,1,0,1,0,0,1,0,0,1,1,1]$
2	$ch(2,:) = [1,0,0,1,1,0,1,0,1,1,0,0,1,1,1]$
3	$ch(3,:) = [1,0,0,1,0,1,0,1,0,1,0,0,1,1,1]$
4	$ch(4,:) = [1,0,0,1,0,1,0,1,1,1,0,0,1,1,1]$
5	$ch(5,:) = [1,0,0,0,0,0,0,0,0,1,0,0,0,1,0]$
6	$ch(6,:) = [0,0,0,1,0,0,0,0,1,0,0,0,1,0,1]$
7	$ch(7,:) = [0,0,0,0,1,0,1,0,0,0,0,0,0,0,0]$
8	$ch(8,:) = [0,0,0,0,0,1,0,1,0,0,0,0,0,0,0]$
9	$ch(9,:) = [0,0,0,0,1,1,1,1,0,0,0,0,0,0,0]$
10	$ch(10,:) = [1,0,0,0,1,0,1,0,0,1,0,0,0,1,0]$
11	$ch(11,:) = [0,0,0,1,1,0,1,0,1,0,0,0,1,0,1]$
12	$ch(12,:) = [1,0,0,1,0,0,0,0,1,1,0,0,1,1,1]$
13	$ch(13,:) = [1,1,1,1,1,1,1,1,1,1,1,1,1,1,1]$
14	$ch(14,:) = [1,1,0,0,0,0,0,0,0,1,0,0,0,1,0]$
15	$ch(15,:) = [0,0,1,0,0,0,0,0,1,0,0,1,0,0,1]$
16	$ch(16,:) = [0,0,1,1,0,0,0,0,1,0,1,1,1,0,1]$

Tab. 4 Description of the feature combinations in Tab. 3.

ID	Description
1	Best by statistical properties (Default)
2	Default + TP
3	Default, DFA order changed to 3
4	Default, TP, DFA order changed to 3
5	Time domain
6	Frequency domain (only normalized)
7	DFA-2
8	DFA-3
9	DFA-2, DFA-3
10	Time domain, DFA-2
11	Frequency domain, DFA-2
12	Time domain, Frequency domain (only normalized)
13	Complete features vector
14	Time domain
15	Frequency domain
16	Complete Frequency domain

Data SET-2 is randomly selected data set, which contains only REM and wake stage episodes. This set presents more complicated task than *DATA SET-1*.

Data SET-3 is randomly selected data set, which contains all the sleep categories (see section 3.4). This presents the most complicated classification task and goal of our measurements.

All the following tables contain values mean of the ANN performance function (Mean Square Error ("MSE")) on training data, mean of ANN performance on test data ("test MSE") and best performance value on test data (min test MSE).

 Copyright © 2007 EUROSIM / SLOSIM

Tab. 5 Performance on *DATA SET-1* for Elman network.

Elman	MSE	test MSE	min test MSE
1	0.1227	0.1235	0.0855
2	0.1185	0.1178	0.0790
3	0.1224	0.1244	0.0930
4	0.1173	0.1203	0.0903
5	0.1246	0.1247	0.1032
6	0.0942	0.1003	0.0810
7	0.1157	0.1186	01125
8	0.1099	0.1190	0.1160
9	0.1065	0.1185	0.1099
10	0.1241	0.1241	0.0817
11	0.0898	0.0938	0.0708
12	0.1162	0.1186	0.0901
13	0.1105	0.1143	0.0966
14	0.1238	0.1246	0.1096
15	0.0907	0.0986	0.0737
16	0.0904	0.0928	0.0712

Choosing features for ANN training

It is presumed that features obtained by different methods express autonomic regulation in different ways and their proper combination can improve the performance of selected ANN. Our capability to evaluate the property of features and to create the best training set is limited, therefore simple Genetic Algorithm (GA) [26] was implemented to reduce this inability and to help us for selecting the best features combinations necessary for training. This algorithm tries to evaluate the suitability of different combinations of features. This genetic algorithm utilizes only the simple methods: pre-initialization, breeding and mutation. For this purposes chromosomes was defined as a binary vector which specifies presence or absence of feature at the input of ANN. The chromosome was defined according to the variation property of features: features considered as equivalent or at least of common origin were put as neighbors. If a gen is set to zero the feature is really absent and thus the dimension of ANN input is also reduced.

Even now it can be claimed that some features are redundant (DFA-2 vs. DFA-3). To select the best combination of features it is better to chose features from different groups: Time domain, Frequency domain and DFA. As a guide, Tab. 2 may be used. At the basis of theoretical analysis (see 3.3) the feature sets were created, these were also used for GA pre-initialization (see above). These sets are described in detail in the table 3. The following features (see 3.4.2) give the best statistical properties and therefore they were used to form pre-initialization of GA: RMSSD, HFn, $2\alpha_{fast}$, $2\alpha_{total}$, SDNN, LFn, RRmean and LF/HF.

Genetic algorithm results

The genetic algorithm got stable after 40 generations: the population contained the same individuals. This was specified as stopping condition.

The best feature combination was selected as

$$ch = [1, 0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1]$$

From the rundown of genetic algorithm the other individuals whose chromosomes could contain interesting information were then chosen. These individuals were selected because their quality rating was near to the performance of the best individual. These chromosomes are different from the best individual's chromosome. These individuals are summarized in Tab. 6 and Tab. 7.

Tab. 6 The best feature combinations selected by GA.

ID	Chromosome
17	$ch(17,:) = [0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1]$
18	$ch(18,:) = [0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1]$
19	$ch(19,:) = [1, 0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1]$
20	$ch(20,:) = [1, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1]$
21	$ch(21,:) = [1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1]$
22	$ch(22,:) = [1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 1]$

Tab. 7 Description of the feature combinations in Tab. 6

ID	Description
17	NN50,$2\alpha_{fast}$,$3\alpha_{fast}$,SDNN,LF/HF
18	NN50,$2\alpha_{fast}$,$3\alpha_{total}$,SDNN,LF/HF
19	RMSSD,HFn,$2\alpha_{fast}$,$3\alpha_{fast}$,SDNN,LF/HF
20	RMSSD,HFn,$2\alpha_{fast}$,$2\alpha_{total}$,$3\alpha_{fast}$,SDNN,LF/HF
21	RMSSD,HFn,$3\alpha_{total}$,SDNN,LFn,LF/HF
22	RMSSD,HF,$3\alpha_{fast}$,$3\alpha_{total}$,SDNN,LFn,LF/HF

To be able to perform comparison, the performance of these features was evaluated at the same 3 data sets (*DATA SET-1*, *DATA SET-2* and *DATA SET-3*) as preselected feature combinations. As an example, the results for feature combinations generated by GA, evaluated on *DATA SET-1* and using Elman network, are summarized in Tab. 8.

Tab. 8 GA generated features evaluated on *DATA SET-1* using Elman network.

Elman	MSE	test MSE	min test MSE
17	0.0718	0.0928	0.0852
18	0.0694	0.0821	0.0708
19	0.0747	0.0751	0.0639
20	0.0625	0.0809	0.0659
21	0.0767	0.0833	0.0716
22	0.0723	0.0782	0.0621

4.1.2 ANN structure comparison

To select proper topology of ANN statistic tests were run on 3 different ANN topologies (see section 4). The number of hidden layers and neurons inside for tested topologies are summarized in Tab. 9 and the other parameters for input and output layer are as specified in default FFN topology.

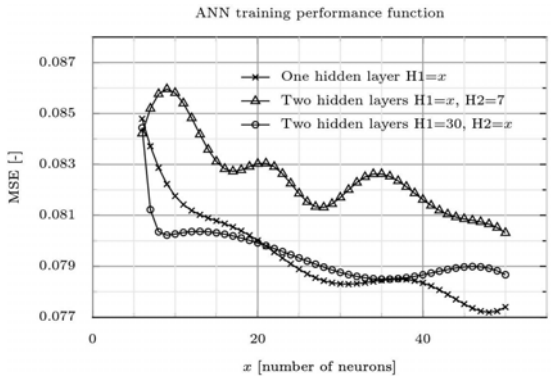
Tab. 9 ANN patterns. ANN-P: ANN property, H-L: Hidden layers

ANN-P	ANN 1	ANN 2	ANN 3
H-L	H1=x	H1=x,H2=7	H1=30,H2=x

For all the ANNs, the used features vector identifier (ID) =1, the Matlab training function is "trainp", the transfer function is "log-sig", the early stopping has 20 iterations and the input states are the sleep stage 4 and REM, performance function is MSE. x is the variable number of neurons in the given hidden layer. The fixed numbers (7 and 30) of neurons in certain tested topologies was estimated at the basis of the first test. The test was performed as follows:

1. Random selection of training, cross-validation and test data from the database of features.
2. Creation of ANN by using the selected topology.
3. Re-training ANN for 500 times on the same data to get some usable statistics.
4. Getting mean of the performance ANN function on the training data, mean of the performance ANN function on the test data and the best performance value on the test data.
5. The obtained means were fitted with Gaussian curves to make the results readable (see Fig. 6 as example). For fitting, the Curve fitting toolbox was used. [30]

Fig. 6 ANN topology comparison (MSE vs. hidden units number).



All the topologies achieved nearly the same best results on the performances but these results showed that ANN with one hidden layer performs better in all the manners. Increasing the number of neurons and hidden layers can augment the performance function, but network with only one hidden layer generalized better. Thus it was chosen to make all the comparisons with FFN using only one hidden layer. In order to determine the best number of neurones in this hidden layer, the test was re-run several times for this ANN pattern and it showed the performance is significantly better for ANNs with 25-35 neurons in the hidden layer.

4.1.3 ANN types comparison

Two types of ANN were selected for comparison: FFN and ELN. Using the results from previous section we made comparison of the best FFN (with one hidden layer) with the ELN defined by several topologies. The test was performed as in the section 4.1.2. The used topologies of network are summarized in Tab. 10. For all the ANNs, the used features vector identifier (ID) =1, the Matlab train function is "trainp ", the transfer function is "log-sig", the early stopping is 20 iterations and the input state is the sleep stage 4 and REM. Because it was shown in the previous section that adding more neurons did not improved results, the training of FFN was terminated at the limit of 35 neurons in the hidden layer.

Tab. 10 Networks for ANN types performance comparison. ANN-P: ANN property, H-L: Hidden layers, FFN: feed-forward network, ELN: Elman network

ANN-P	ANN 1	ANN 2	ANN 3
ANN type	FFN	ELN	ELN
H-L	H1=x	H1=x,H2=7	H1=30,H2=x

On one hand, from the results plotted in Fig. 7 and Fig. 8 it is obvious that Elman network performs better with respect to the capability of generalization and convergence.

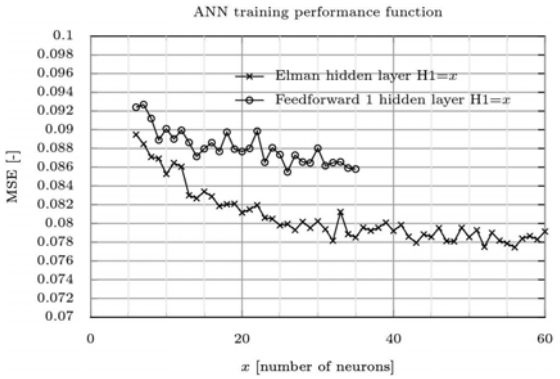


Fig. 7 Performance comparison of Elman and feed-forward ANNs using train data

On the other hand, the best reached performance using the test data set is nearly the same. This is the fact in nearly all the tests.

4.1.4 Topology results summary

At the basis of tests, two artificial neural networks were constructed for classification of sleep stages. Their properties are summarized in Tab. 11. For all the ANNs, the used Matlab train function is "trainrp", the transfer function is "log-sig", the early stopping is 20 iterations and the performance function is MSE.

4.2 Sleep stage classification

The capability to classify was so far tested on four patients, with total record length of 1980 minutes. The ca-

 Copyright © 2007 EUROSIM / SLOSIM

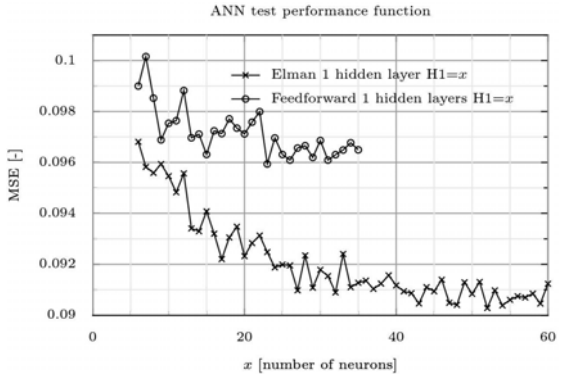


Fig. 8 Performance comparison of Elman and feed-forward ANNs using test data

Tab. 11 Final ANN properties. ANN-P: ANN property, H-L: Hidden layers, FFN: feed-forward network, ELN: Elman network

ANN-P	ANN 1	ANN 2
ANN type	ELN	FFN
H-L	H1=27	H1=17

pability to differentiate sleep stages in predefined categories (see section 3.4) is successful in 65%. The Classification performend at data set containing only deep sleep and REM categories had 83.4% reliability.

By analyzing the results given in the Tables of performances on *DATA SET-1*, *DATA SET-2*, *DATA SET-2* for Elman (see Tab. 5 as an example) and feed-forward ANN, and by analyzing the Tables of generated features using *DATA SET-1*, *DATA SET-2*, *DATA SET-2* (see Tab. 8 as an example) with respect to the quality of training, we found that Elman network performs worse than feed-forward ANN. Although both networks reached similar mean values of performance on test data feed-forward ANN was able to reach much better absolute results on test data.

Surprisingly Elman neural network performed better when analyzing quality of classification. Both ANNs were evaluated on all the available data. Elman reached the reliability of 83.4% with "ID = 19" and feed-forward ANN reached 82.4% with "ID= 2". The best performance had feature combination selected by GA with "ID = 19" and was followed by the feature combination with "ID = 2" and "ID = 3".

Comparing all the tables of the generated features on the three data *DATA SET-1*, *DATA SET-2*, *DATA SET-2* and all the tables of performances of Elman and feed-forward ANNs, we found out that the best mean test performance the GA generated feature sets, but the feature combination which was estimated to have the best statistical properties with "ID = 1" reached the best test value. This can be explained as that the feature combination with "ID = 1" contains most of the relevant informations, but it is more complicated to overcome local minima during the training. On the other side using the feature combination selected by GA with "ID = 19" the ANN can be easily trained.

5 Conclusion

In this preliminary work, we studied the feasibility of automatic sleep scoring using only the electrocardiogram (ECG) records. This feasibility has been demonstrated on real ECG records issued from 4 patients. The capability to differentiate sleep stages in predefined categories is successful in 65%. The Classification performend at data set containing only deep sleep and REM categories had 83.4% reliability. In order to obtain more reliable results, the authors extend actually this work to include more number of patients.

The above results show that the methods used to evaluate the property of the features and the ANN topology was not sufficiently satisfactory. The performance rating on test data and the ability to classify properly has to be clearly linked. Thus, in the future work, the performance function for the ANN and the evaluation function for the Genetic algorithm will be revisited.

The main benefit will be sleep quality assessment using only vital signals as electrocardiogram or pulse transition time which can be easily measured. Using Holter analysis, that is a wide spread and cheap methodology in cardiology area, the sleep quality monitoring could be applied straightforward by implementing project results.

Furthermore, others objective well be the reduction of the costs for the sleep evaluation, better prevention of chronic illness or disabilities related to sleep disorders, though the earlier and simpler examination of potential patients and early diagnosis. A positive effect on the quality of medical care can be expected, while lowering the overall economic burden on the health care systems.

6 Acknowledgments

The authors would like to thanks Daniel Coquelle, manager engineer in MEDATEC France, the Service De Physiologie et d'Exploration Fonctionnelle at Hpital Raymond Poincar, Garches-France in particular Professor F. Lofaso, Dr. MA Duera Salva, Dr. G. MROUE and their team, for providing the data used in this work and their help.

7 References

[1] M. W. Mahowald and C. Schenck. Insights from studying human sleep disorders. *NATURE*, 437(27), 2005.

[2] The world in 2006, unhealthy numbers, 2005.

[3] K. Kinsella, Philips DR., and Global Aging. The challenge of success. *Population Bulletin*, 60(1), 2005.

[4] Zhancheng Li, Minfen Shen, and Patch Beadle. *Classification of EEG Signals Under Different Brain Functional States Using RBF Neural Network*, volume 3174. Lecture

 Copyright © 2007 EUROSIM / SLOSIM

Notes in Computer Science, Springer-Verlag, Berlin/Heidelberg, 2004.

[5] Schaltenbrand N, Lengelle R, Toussaint M, Luthringer R, Carelli G, Jacqmin A, Lainey E, Muzet A, and Macher JP. Sleep stage scoring using the neural network model: comparison between visual and automatic analysis in normal subjects and patients. *Sleep*, 19(1):26 – 35, Jan 1996.

[6] Barnett D.W., Laposky A., Thomas C., and Anch M. Neural network scoring of rat sleep stages. In *Proceedings of the 21st Annual International Conference of the IEEE Eng Med Bio Soc*, volume 1, page 389, Atlanta, GA, USA, October 10, 1999.

[7] Kim B.Y. and Park K.S. Automatic sleep stage scoring system using genetic algorithms andneural network. In *Proceedings of the 21st Annual International Conference of the IEEE Eng Med Bio Soc*, volume 2, pages 849–850, Chicago, IL, USA, July 23, 2000.

[8] Flexer A., Sykacek P., Rezek I., and Dorffner G. Using hidden markov models to build an automatic, continuous andprobabilistic sleep stager. In *IJCNN 2000*, volume 3, pages 627 – 631, Como, Italy, July 23, 2000.

[9] FLEXER Arthur, GRUBER Georg, DORFFNER Georg, and Dorronsoro Jos R. *Continuous Unsupervised Sleep Staging Based on a Single EEG Signal*, volume 2415. Lecture Notes in Computer Science, Springer-Verlag, Berlin/Heidelberg, 2002.

[10] D. Novák, T. Al-Ani, Y. Hamam, and L. Lhotská. Using hidden markov models to build an automatic, continuous andprobabilistic sleep stager. In *5th EUROSIM Congress on Modelling and Simulation*, pages 627 – 631, Noisy-Le-Grand, France, September 06, 2004.

[11] Absil P.-A.and Sepulchre R., Bilge A., and Gerard P. Nonlinear analysis of cardiac rhythm fluctuations using dfa method. *Physica A*, 272(10):235 – 244, October 1999.

[12] Task Force of The European Society of Cardilogy, The North American Society of Pacing, and Electrophysiology. Heart rate variability: Standarts of measurements, physiological interpretation, and clinical use. *European Heart Journal*, 17:354 – 381, March 1996.

[13] Thomas Penzel, Jan W. Kantelhardt, Ludger Grote, Jorg-Hermann Peter, , and Armin Bunde. Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea. *IEEE*, February 2003.

[14] Thomas Penzel, Jan W. Kantelhard, HF Becker, JH Peter, and A Bunde. Detrended fluctuation analysis and spectral analysis of heart rate variability for sleep stage and sleep apnea identification. *Computers in Cardiology*, 30:307 – 310, 2003.

[15] B. Kroose and P. Van der Smagt. An introduction to neural networks, 1993.

[16] Van Dongen HA., Maislin G., Mullington J., and Dinges DF. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *SLEEP*, 26:117–26, 2002.

[17] A. Rechtschaffen and A. Kales. *Techniques and Scoring System for Sleep Stages of Human Subjects*. University of California, 1968.

[18] M. Staudacher, S. Telser, A. Amann, H. Hinterhuber, and M. Ritsch-Marte. A new method for change-point detection developed for on-line analysis of the heart beat variability during sleep. *Physica A*, 349(3-4):582 – 596, April 2005.

[19] Jan W.Kantelhardt, Thomas Penzel, Sven Rostig, Heinrich F Becker, Shlomo Havlin, and Armin Bunde. Breathing during rem and non-rem sleep: correlated versus uncorrelated behaviour. *Physica A*, 319:447 – 457, March 2003.

[20] Armin Bunde, Schlomo Havlin, Jan W. Kantelhard, Thomas Penzel, Jorg-Hermann Peter, and Karlheinz Voigt. Correlated and uncorrelated regions in heart-rate fluctuations during sleep. *Physical Review Letters*, 85(17):3736 – 3739, October 2000.

[21] Peng CK, Havlin S, Stanley HE, and Goldberger AL. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos*, 5(1):82–87, March 1995.

[22] Kun Hu, Plamen Ch. Ivanov, Zhi Chen, Pedro Carpena, and H. Eugene Stanley. Effect of trends on detrended fluctuation analysis. *Physical Review E*, 64(011114):1–19, 2001.

[23] J. Pan and W. L. Tompkins. A real-time qrs detection algorithm. *IEEE Trans. Biomed. Eng.*, 32:230–236, 1985.

[24] Arnold Neumaier and Tapio Schneider. Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Trans. Math. Softw.*, 27(1):27 – 57, 2001.

[25] Tapio Schneider and Arnold Neumaier. Algorithm. arfit — a matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software*, 27:58–65.

[26] Holland J.H. *Adaptation in Neural and Artificial Systems*. MIT Press, 1992.

[27] Medatec france: http://www.medatec.fr/.

[28] Service de physiologie et d'exploration fonctionnelle, hôpital raymond poincaré, garches, france.

[29] The Mathworks Inc MATLAB. Artificial neural network toolbox.

[30] The Mathworks Inc MATLAB. Curve fitting toolbox.

 Copyright © 2007 EUROSIM / SLOSIM