IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

Automatic screening of obstructive sleep apnea from the ECG based on empirical mode decomposition and wavelet analysis

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2010 Physiol. Meas. 31 273

(http://iopscience.iop.org/0967-3334/31/3/001)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 134.58.253.57
The article was downloaded on 23/03/2011 at 13:03

Please note that terms and conditions apply.

IOP PUBLISHING PHYSIOLOGICAL MEASUREMENT

Physiol. Meas. **31** (2010) 273–289 doi:10.1088/0967-3334/31/3/001

Automatic screening of obstructive sleep apnea from the ECG based on empirical mode decomposition and wavelet analysis

M O Mendez[1,2], J Corthout[3], S Van Huffel[3], M Matteucci[4], T Penzel[5], S Cerutti[1] and A M Bianchi[1]

[1] Department of Biomedical Engineering, Politecnico di Milano, Pzza. Leonardo da Vinci 32, 20131 Milano, Italy
[2] Facultad de Ciencias, Diagonal Sur S/N, Zona Universitaria, C.P. 78290, San Luis Potosi, S.L.P., Mexico
[3] Department of Electrical Engineering—ESAT, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
[4] Department of Electronics and Information, Politecnico di Milano, Pzza. Leonardo da Vinci 32, 20131 Milano, Italy
[5] Das Schlafmedizische Zentrum der Charité, Luisenstrasse 13, 10117 Berlin, Germany

E-mail: martin.mendez@biomed.polimi.it

Received 30 August 2009, accepted for publication 21 December 2009
Published 20 January 2010
Online at stacks.iop.org/PM/31/273

Abstract

This study analyses two different methods to detect obstructive sleep apnea (OSA) during sleep time based only on the ECG signal. OSA is a common sleep disorder caused by repetitive occlusions of the upper airways, which produces a characteristic pattern on the ECG. ECG features, such as the heart rate variability (HRV) and the QRS peak area, contain information suitable for making a fast, non-invasive and simple screening of sleep apnea. Fifty recordings freely available on Physionet have been included in this analysis, subdivided in a training and in a testing set. We investigated the possibility of using the recently proposed method of empirical mode decomposition (EMD) for this application, comparing the results with the ones obtained through the well-established wavelet analysis (WA). By these decomposition techniques, several features have been extracted from the ECG signal and complemented with a series of standard HRV time domain measures. The best performing feature subset, selected through a sequential feature selection (SFS) method, was used as the input of linear and quadratic discriminant classifiers. In this way we were able to classify the signals on a minute-by-minute basis as apneic or nonapneic with different best-subset sizes, obtaining an accuracy up to 89% with WA and 85% with EMD. Furthermore, 100% correct discrimination of apneic patients from normal subjects was achieved independently of the feature extractor. Finally, the same procedure was repeated by pooling features from standard HRV time domain, EMD and WA together in order to investigate if

 © 2010 Institute of Physics and Engineering in Medicine

the two decomposition techniques could provide complementary features. The obtained accuracy was 89%, similarly to the one achieved using only Wavelet analysis as the feature extractor; however, some complementary features in EMD and WA are evident.

Keywords: time-frequency analysis, obstructive sleep apnea, heart rate variability, automatic classification, pattern recognition

1. Introduction

Sleep-related breathing disorders have a high prevalence in the adult population. Epidemiological studies indicate a high prevalence of 4% in males and 2% in females in the general population (Young *et al* 1993). Obstructive sleep apnea (OSA), the most common of the different types of sleep-related breathing disorders with about 84% of the total cases (Malhotra and White 2002), is characterized by repetitive cessations of respiratory flow during sleep.

The cessation of respiratory air flow occurs as a consequence of a collapse of the upper airway at the level of the oropharynx. Upper airways obstruction produces a negative intrathoracic pressure during inspiration which does not result in an effective airflow. This continues until the lack of oxygen and the increase of CO_2 causes a central nervous system activation, called arousal. This activation recovers respiration without reaching the level of conscious wakefulness, thus not being perceived by the subject. This same mechanism could repeat up to 600 times in a single night in patients with severe sleep apnea (Penzel *et al* 2000). This respiratory phenomenon, connected to central and autonomic reflexes, creates a pattern in the heart rate fluctuations that is repeated at each apnea episode, which is characterized by a decrease in the heart rate during the apnea phase and a heart rate increase during the recovery (generally accompanied by an arousal episode).

Patients suffering from sleep apnea report excessive daytime drowsiness even when the sleep period seems long enough. Besides significant social and emotional problems caused by poor mental performances (Martin *et al* 1996), patients can also suffer from a series of physiological problems. Frequent findings are hypertension and cardiac arrhythmia, which in turn are often precursors of heart failure (Himmelmann 1999). This is why a sleep apnea patient has a significantly higher probability of developing cardiovascular diseases (Ancoli-Israel *et al* 2003).

Clinical sleep studies are expensive, because they require overnight evaluation in sleep laboratories with dedicated systems and specialized attending personnel. Due to the cost and relative scarcity of sleep centres, sleep apnea is widely underdiagnosed (Young *et al* 1997). Hence, techniques which could provide a simple screening of sleep apnea without the need for a specialized sleep centre are surely be of benefit. To this purpose, an interesting signal to be used is the electrocardiogram (ECG) since this is highly influenced during the apnea event and can be easily measured in a non-invasive way and with high signal-to-noise ratio even in a nonclinical environment. In the recent years, the international literature presented numerous papers on this subject, and some of them described processing methods able to reach very good performance in apnea detection (Penzel *et al* 2002). In particular, Mendez *et al* (2009) demonstrated the importance of using time-variant or time-frequency approaches for correctly managing the nonstationarities in the signals, typical of apnea episodes. In this paper, we

propose the use of two different methodologies: the empirical mode decomposition (EMD) that decomposes the signals in intrinsic mode functions and the wavelet analysis (WA) that enable a time-frequency (or time scale) view of the signal. Both the methods could put into evidence-specific dynamics for a better description of the signal.

2. Methodological considerations

2.1. Empirical mode decomposition (EMD)

The essence of the EMD is to decompose the signal in different components called intrinsic mode functions (IMFs) that satisfy the necessary conditions for a meaningful Hilbert transform (HT) calculation. An IMF is calculated by subtracting from the original signal $x(t)$ the mean m of the upper and lower envelopes, obtained by a cubic spline interpolation through the local maxima and minima of the signal $x(t)$, respectively:

$$C = x(t) - m \tag{1}$$

This process is further repeated on C until eventually an IMF is obtained according to the following characteristics:

(i) the number of extrema and the number of zero crossings must be either equal or differ at most by one;
(ii) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

This IMF is then subtracted from $x(t)$, leaving a residue R:

$$R = x(t) - C. \tag{2}$$

The described procedure can be applied again to R to obtain a set of IMFs and thus to decompose the signal until the remaining R is no more than a monotonic signal. For the detailed description of the EMD procedure, see Huang *et al* (1998).

2.2. Hilbert transform

The HT performed afterward on the IMFs makes the data analytical and calculates its instantaneous frequency in a unique way. For an arbitrary time series, $X(t)$, we can always have its HT, as

$$Y(t) = \frac{1}{\pi} p.v. \int_{-\infty}^{\infty} \frac{X(t')}{t - t'} \, \mathrm{d}t', \tag{3}$$

where $p.v.$ indicates the Cauchy principal value. With this definition, $X(t)$ and $Y(t)$ form a complex conjugate pair, so we can have an analytic signal, $Z(t)$, as

$$Z(t) = X(t) + \mathrm{i}Y(t) = a(t)\, \mathrm{e}^{\mathrm{i}\theta(t)}, \tag{4}$$

in which

$$a(t) = [X^2(t) + Y^2(t)]^{1/2}, \qquad \theta(t) = \arctan\left(\frac{Y(t)}{X(t)}\right). \tag{5}$$

In equation (4), the polar coordinate expression clarifies the local nature of this analytic representation: it is the best local fit of trigonometric function with time-varying modulus and phase to $X(t)$.

2.3. Wavelet analysis

To study the EMD behaviour on HRV with an established decomposition method, an analogous procedure was implemented using WA to subdivide the signal in components. Basically, wavelet analysis decomposes a signal by using different versions of a 'mother' wavelet which is scaled and translated in time (Torrence and Campo 1998).

The general formula for the continous wavelet transform (CWT) is

$$\mathrm{CWT}_x(\tau, a) = \frac{1}{\sqrt{a}} \int x(t) g^* \left(\frac{t-\tau}{a} \right) \mathrm{d}t, \tag{6}$$

where $g(t)$ is the mother or basic wavelet, * denotes a complex conjugate, a is the scale factor and τ is a time shift (Thakor *et al* 2000). Typically, $g(t)$ is a bandpass function centred around a central frequency f_0. Scale a produces the compression or expansion of $g(t)$ in time, scaling the corresponding central frequency and bandwidth $1/a$. WA ensures that the wavelet function at each scale is normalized to have unit energy such that the energy at the different scales can be directly comparable. In the discrete implementation of the wavelet transform (DWT), the scale factor a varies as power of 2. The discrete Daubechies4 wavelet basis has been chosen as suggested in (Pichot *et al* 1999).

3. Material and methods

3.1. Database description

The analysed data come from the Physionet database (www.physionet.org). The ECG recordings come from Polysomnographies recorded at the sleep laboratory of the Philipps Universität in Marburg, Germany. They are continuous recordings and contain all the events that occur during a night which includes apneas, arousals, movements and also some wakefulness episodes. The inclusion criteria of the subjects are reported in the Physionet website and Penzel *et al* (2000). The subjects were normal sleep apnea patients; they had some degree of arterial hypertension but no other cardiac or other medical disorder. They had no medication with effects on blood pressure or heart rate. The patients had some arrhythmias as usual in patients with sleep apnea and in patients with hypertension. Heart failure patients were excluded.

Apnea scoring was carried out on the basis of standard criteria by an expert sleep clinician (Penzel *et al* 2000).

An apnea/hypoapnea event is defined as a transient reduction in, or complete cessation of, breathing. The episode must fulfill the following criteria (AASM task force 1999).

(i) A clear decrease (>50%) from the baseline in the amplitude of a valid measure of breathing during sleep. Baseline is defined as the mean amplitude of stable breathing and oxygenation in the 2 min preceding onset of the event or the mean amplitude of the three largest breaths in the 2 min preceding onset of the event.

(ii) Clear amplitude reduction of a validate measure of breathing during sleep that does not reach the above criterion but is associated with either an oxygen saturation of >3% or an arousal.

(iii) The event lasts 10 s or longer.

Apnea detection was performed through the visual scoring of the respiratory activity that was obtained as oronasal airflow measured using nasal thermistors, together with chest and abdominal respiratory effort measured using inductive plethysmography, each digitized at

25 Hz. In addition also an oxygen saturation signal digitized at 1 Hz was analysed. The apneic–nonapneic annotation was provided on a minute-by-minute basis: a minute was labelled as apneic if it contained at least one apneic or hypo-apneic episode, and if not it was defined as nonapneic (Penzel *et al* 2000). The subject age ranged in the data set between 27 and 63 years (48 ± 10.8 years) and subject weight ranged between 53 and 135 kg (86.3 ± 22.2 kg).

Standard ECG recordings were acquired at a sampling frequency of 100 Hz, with a 16 bits resolution. In the website the recordings were subdivided in three groups: apneic patients (class A, with more than 100 min in apnea), borderline patients (class B, with total apnea duration more than 5 and less than 99 min) and control or normal patients (class C, with less than 5 min in apnea). From the database, 25 recordings were used as a training set for the classification algorithms. A second group with 25 recordings was used as a test set to measure the performance of the algorithms. Recordings of the database with a large number of ectopic beats (more than 10% of the beats in the recording) were not included in the present research. From the selected 50 recordings (from a total of 70 recordings), the heart rate variability (HRV) signal was automatically obtained through a derivative and threshold procedure. The eventual misdetected beats were manually corrected (see section 3.3 for details). The training set consists of 4950 apneic minutes and 7127 nonapneic minutes. The test set contains 4428 apneic and 7927 nonapneic minutes.

3.2. RR signal correction and ECG-derived respiratory signal (AR signal)

In the provided database for each ECG signal a series of automatically detected QRS complexes is included. From the QRS times and ECG signals the R peaks were detected and the RR series were obtained. After that, the RR series were plotted together with the ECG to allow manual correction of the misdetected beats. After manual correction, some errors were still present such as outliers and automatic correction was done based on the deviation with respect to a smooth version of the original time series (Mendez *et al* 2009, de Chazal *et al* 2003). The AR signal was calculated by subtracting the baseline from the original ECG, detecting the minimum value within 100 ms before and after the maximum value of the R peak and integrating the area under the R peak in the interval between these two points. The baseline itself was calculated using a median filter of 200 ms width.

3.3. Analysis description

On the basis of previous knowledge regarding the physiological effects of OSA on the ECG signal, we extracted the RR interval series and the AR signal (QRS complex area), according to the QRS annotations available on the website. The RR intervals show a characteristic oscillation (brady-tachycardia) during periods with apneic events, producing a frequency component around 0.02 Hz. AR is a signal that is highly affected by respiration; thus it can be used as a rough estimation of respiratory activity and shows the same 0.02 Hz oscillation during the sleep periods with apneas. During periods without apneic events, an oscillation at the respiration frequency is observed in both the signals around 0.25 Hz. Each beat-by-beat signal was resampled at 1 Hz, and then the time series were decomposed into their respective modes by EMD. Thereafter, the Hilbert transform (HT) was applied to each mode, from which features were extracted and complemented with a series of standard heart rate variability (HRV) measures in time domain and through nonlinear analysis. All these parameters have been obtained with 1 min time resolution. From the obtained feature set, the best performing feature subset was selected as the input of linear and quadratic discriminant classifiers. This whole process has been repeated using WA and features were extracted from wavelets details.

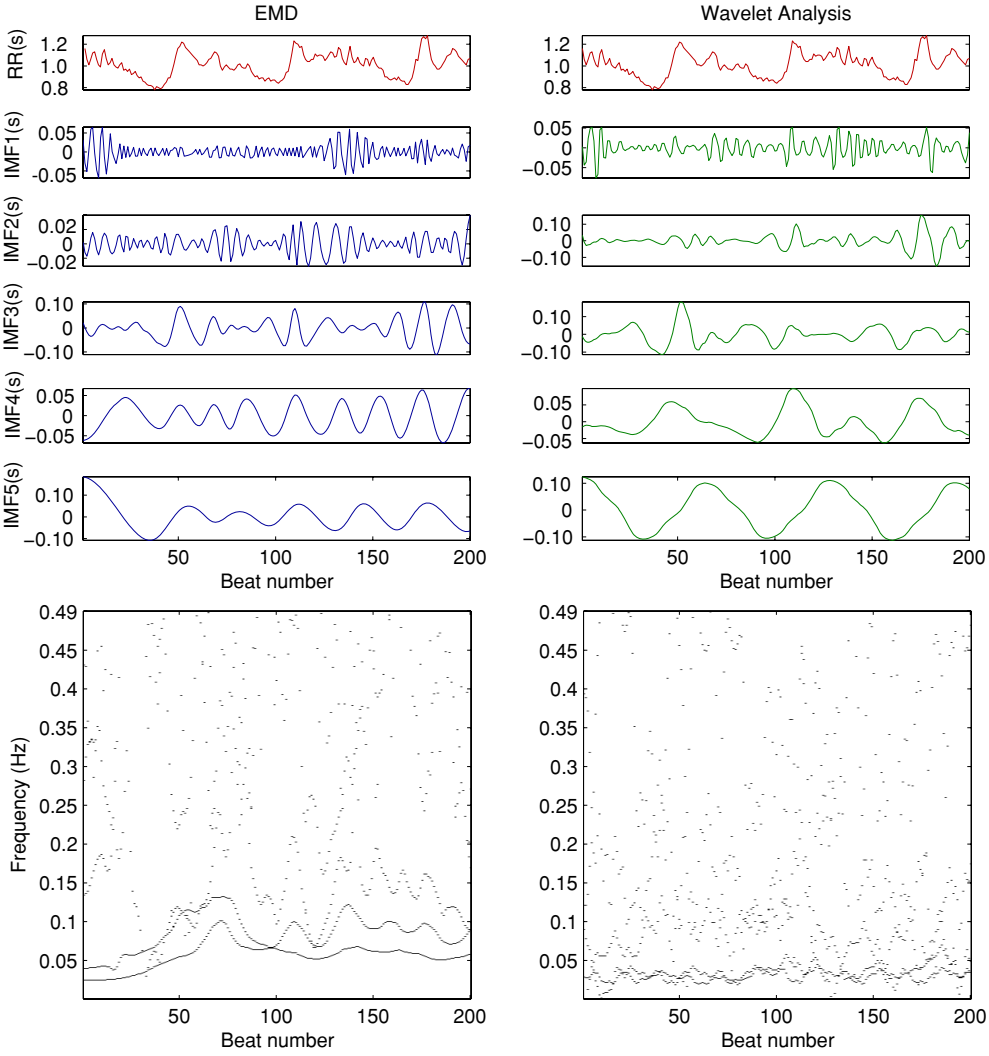


Figure 1. Empirical mode decomposition (EMD) and wavelet analysis (WA) of an apneic RR signal segment and their respective time-frequency spaces, obtained after calculation of the Hilbert transform on the components. Displayed are the instantaneous amplitudes of the respective components (above) and the modulus of their Hilbert transform (below).

(This figure is in colour only in the electronic version)

An illustration of both decompositions applied to an apneic segment of a representative RR signal is shown in figure 1. Finally the features extracted by means of EMD and WA were pooled in one set and the best features that separate apnea and nonapnea sleep time were selected.

3.4. Feature set

From the EMD the first five IMFs were retained as they contained the relevant information for this application (last IMFs contained information related to the trend). Analogous selection was made for the WA: also here only the first five components were used (see figure 1),

as they cover all relevant frequencies in the HRV signal. The Hilbert transforms of these components were then calculated. Instantaneous frequencies were not normalized, while for the instantaneous modulus a whole variety of normalization factors were explored: for each sample the instantaneous amplitude of every component was divided (1) by the mean value of the instantaneous amplitudes of all other components and (2) by the sum of the amplitudes of all the modes. Using only the first five components this gave rise to 15 different ratios. Means and standard deviations of these ratios are calculated on a minute-by-minute basis. All these results were then added to the feature collection. Furthermore, corresponding RR and AR signal components were also multiplied with each other to provide a sort of normalized correlation measure between the contributions of the considered band in the RR and AR signal. Thereafter, the ratios between instantaneous modulus were logarithmically transformed to obtain a normal distribution. Then all the features, that were not normalized with respect to the corresponding total power, were normalized to zero mean and unit variance.

These features were then complemented with five additional feature types. Five features from the standard time domain measures such as RMSSD, SDSD, HRV triangular index and TINN were calculated (Task Force 1996) as well as the sample entropy on a minute-by-minute basis.

3.5. Feature subset and classifier selection

Feature selection is a fundamental step inside the classification process. The selection of a small subset, with high discriminatory power, aids to reduce the complexity of the classification procedure and avoids the course of dimensionality in estimating the *a posteriori* distribution during classification. Different approaches have been proposed to select the best set of features for a specific classification task, among them we can find statistical analysis of features, wrap methods, principal component analysis and factor analysis (Duda *et al* 2001). Wrap methods search for a feature subset that maximizes the accuracy of the learning algorithm. The accuracy is evaluated by cross-validation on the training set following an induction rule based on adding or removing a feature from the feature set. Different wrap methods are usually used in the literature, the most common ones are forward, backward and bidirectional (Kohavi and John 1997, John *et al* 1994). In this study, sequential feature selection (SFS) was used to select the best feature subset that separates sleep apnea from normal sleep periods. SFS starts with the accuracy evaluation of the learning algorithm for each single feature in the feature set. In other words, SFS measures the leave-one-out cross validation (LOOCV) for each single feature in the feature set (one subject is used for testing the algorithm performance at each iteration). Afterwards, SFS selects the individual feature that maximized accuracy. Then, SFS computes LOOCV for two features, the one that maximized accuracy in the first iteration and each other feature from the remaining feature subset. SFS selects the couple with the highest accuracy. After that, SFS evaluates feature subsets with three, four and more features. Finally, SFS gives a *tuple* consisting of the combination of features that was the winner out at each iteration. Generally, the SFS procedure terminates when the accuracy estimated at a certain iteration is less than or equal to that evaluated at the previous step or when the feature subset achieves a predefined number of features. In this study, SFS was terminated when the best feature subset reaches 20 features. This process was performed for both linear and quadratic discriminant classifiers (LD and QD).

Both discriminants assume that the class densities have a constant covariance matrix and the class prediction is based on the maximum likelihood rule. LD and QD give estimation of the *a posteriori* probability for each class. The largest estimated *a posteriori* probability defines

the class of the examined sample. To explain how discriminants work, let $\mathbf{x} = [x_1, x_2, \ldots, x_d]$ be a row vector with d feature values and assume that we want to assign $\mathbf{x}$ to one of the k possible classes i, then the discriminant value f_i for each class i for LD and QD can be computed, respectively, as

$$f_{i_{LD}} = \mu_i \mathbf{C}^{-1} \mathbf{x}^T - \tfrac{1}{2} \mu_i \mathbf{C}^{-1} \mu_i^T + \log(\pi_i) \tag{7}$$

$$f_{i_{QD}} = -(\mathbf{x} - \mu_i)^T \mathbf{C}_i^{-1} (\mathbf{x} - \mu_i) - \left|\mathbf{C}_i^{-1}\right| + 2\log(\pi_i) \tag{8}$$

where T means transpose and μ_i is the row mean vector evaluated from n training vectors belonging to the class i and π_i represents the *a priori* probability that $\mathbf{x}$ belongs to a class i. To evaluate μ_i let N be the total number of $\mathbf{x}$ in the training set and N_i be the number of $\mathbf{x}$ of the class i, then μ_i is obtained as follows:

$$\mu_i = \frac{1}{N_i} \sum_{n=1}^{N_i} \mathbf{x}_{ni}. \tag{9}$$

For a QD classifier $\mathbf{C}_i$ represents the pooled covariance of each class and it is evaluated as

$$\mathbf{C}_i = \frac{1}{N_i - 1} \sum_{n=1}^{N_i} (\mathbf{x}_{ni} - \mu_i)^T (\mathbf{x}_{ni} - \mu_i) \tag{10}$$

while for LD the pooled covariance is obtained for all classes as

$$\mathbf{C} = \frac{1}{N - k} \sum_{i=1}^{k} \sum_{n=1}^{N_i} (\mathbf{x}_{ni} - \mu_i)^T (\mathbf{x}_{ni} - \mu_i). \tag{11}$$

Finally, a practical way to evaluate π_i is

$$\pi_i = \frac{N_i}{N}. \tag{12}$$

However, since we are assuming the same *a priori* probability for both apneic and nonapneic classes, π_i could be eliminated in the above equations.

Figure 2(a) shows the accuracy performance, for both learning algorithms LD and QD, during SFS feature selection using features obtained by EMD and time domain parameters calculated from the HRV signal from the training set. Performance of the learning algorithms, LD and QD, during SFS with features extracted by WA and HRV time measures from the training set is shown in figure 2(c). Finally, since EMD and WA decompose a time series in a number of nonlinear and linear scales, features extracted from EMD and WA and HRV time domain measures were pooled together in order to observe if different ways of decomposition could be complementary, and a higher classification could be obtained due to the complementarity or the higher classification performance could be obtained with a reduced number of features. The results for both learning algorithms, LD and QD with SFS and features obtained through EMD-WA along with HRV time measures from the training set, are shown in figure 2(e).

In an attempt to boost the classification performance, each feature in the best 20-feature subset (coming from SFS feature selection for WA, EMD and WA-EMD independently) was smoothed by a moving average filter with different time windows (in minutes). LOOCV was used to evaluate the learning algorithms performance which were fed with the best 20-feature subset (obtained by SFS) plus its single features but smoothed with a 2 min window. Then, LOOCV was used again to evaluate the learning algorithms performance which were fed

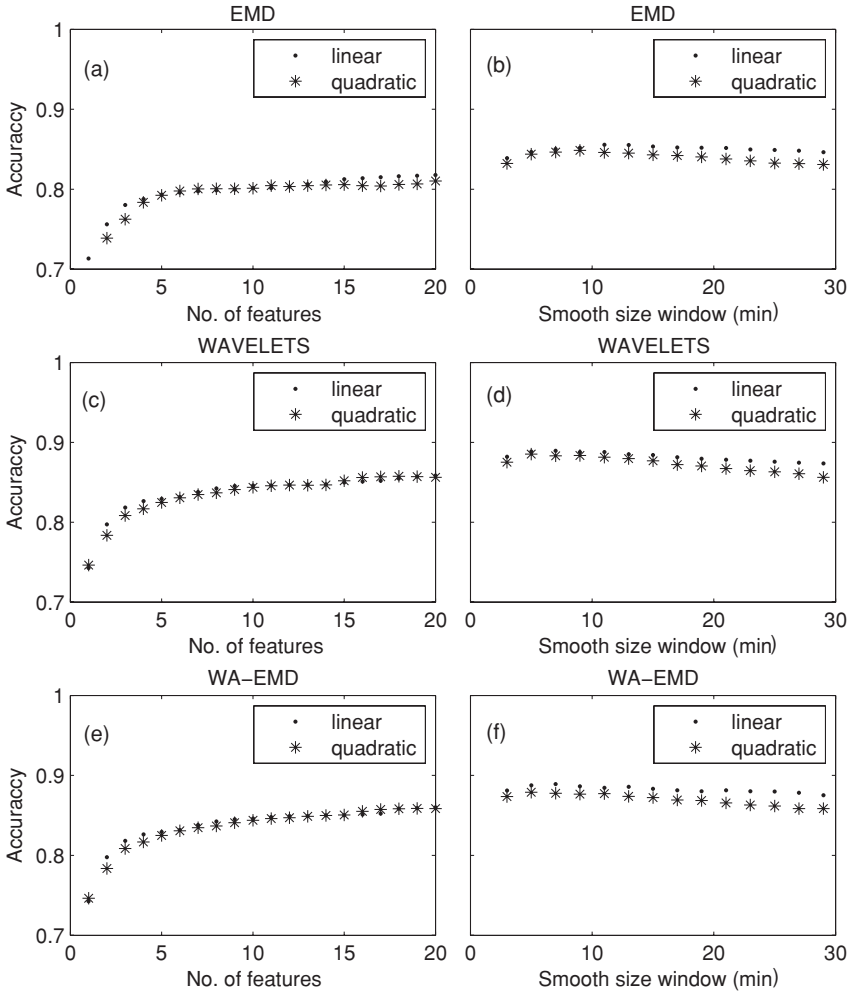


Figure 2. Classification performance measures calculated by sequential feature selection (SFS) on the training set. Subplots a, c and e (left column) show performance of both linear and quadratic discriminant classifiers for EMD and WA and combined WA-EMD features as the best set of features increase. Subplots b, d and f (right column) show cross-validation results for both discriminant classifiers using best 20 feature subset, obtained by SFS from EMD, WA and WA-EMD features, respectively, and complemented with their smoothed versions filtered at different windows sizes by a moving average filter.

with the best feature subset plus its single features but smoothed with a 3 min window. This procedure was done for smoothing windows of 4, 5, 6 until 30 min.

The results of the learning algorithms, for the best 20-feature subsets and complemented with their respective smoothed versions at different time smooth windows for subsets coming from (1) EMD and temporal HRV measures, (2) WA and temporal HRV measures (3) WA-EMD and HRV time measures, are shown in figures 2(b), (d), and (f), respectively. Note in figures 2(b), (d) and (f) how the accuracy performance is higher for small smooth size windows and lower for large smooth size windows. This process tries to include information from large scales and reduces possible outliers that can be found in the raw features.

Table 1. Classification performance of linear and quadratic discriminant classifier on the test set using empirical mode decomposition (EMD) and wavelet analysis (WA) as feature extractors. Classification was done using the best set feature with 10 and 20 features complemented with their respective smoothed versions obtained by a moving average filter with a size window of 7 min.

Method	Performance of wavelets			
No of features	10 features		20 features	
Discriminant	Linear	Quadratic	Linear	Quadratic
Accuracy	0.8755	0.8907	0.8664	0.8630
Sensitivity	0.8591	0.9037	0.857	0.9054
Specificity	0.9050	0.8673	0.8834	0.8200
Method	Performance of EMD			
No of features	10 features		20 features	
Discriminant	Linear	Quadratic	Linear	Quadratic
Accuracy	0.8071	0.8433	0.8581	0.8401
Sensitivity	0.7777	0.8855	0.8282	0.8714
Specificity	0.832	0.8119	0.88832	0.7555
Method	Performance of WA+EMD			
No of features	10 Features		20 Features	
Discriminant	Linear	Quadratic	Linear	Quadratic
Accuracy	0.8711	0.8907	0.8566	0.8673
Sensitivity	0.8569	0.9037	0.8442	0.90589
Specificity	0.8967	0.8673	0.8797	0.7920

4. Results

The best set of features with their best smoothed versions were used to classify HRV periods of 1 min in the test set into apnea and no apnea events. The trained classifiers were tested on 25 recordings coming from 13 apnea, 4 borderline and 8 control subjects never seen during the training phase. Table 1 shows the performance of all the combinations of classifiers and features extractors during the test process by using best feature sets composed with 10 and 20 features and complemented with their respective smoothed versions (obtained with a smoothing window of 7 min) that maximized the learning algorithms performance.

One can observe from table 1 that the performance of the classifier based on WA-EMD does not outperform to the classifier based only in WA. Figures 3 and 4 show minutes of apnea for each subject: the circle represent apneic subjects, the cross borderline subjects and the star normal subjects. In figure 3 the results were obtained automatically by LD-EMD with a best feature set of 20 features plus their smoothed versions, while figure 4 presents the results obtained by QD-WA with a best feature set of ten features plus their smoothed versions. Note that complete separation between normal and pathologic subjects can be achieved using a threshold of 50 apnea minutes per night (dashed line) for the LD-EMD classifier while for QD-WA a threshold of 15 apnea minutes per night (dashed line) separates completely both subject classes. Furthermore, one can observe that most of the borderline subjects are located between 15 and 160 apnea minutes per night for both classifiers.

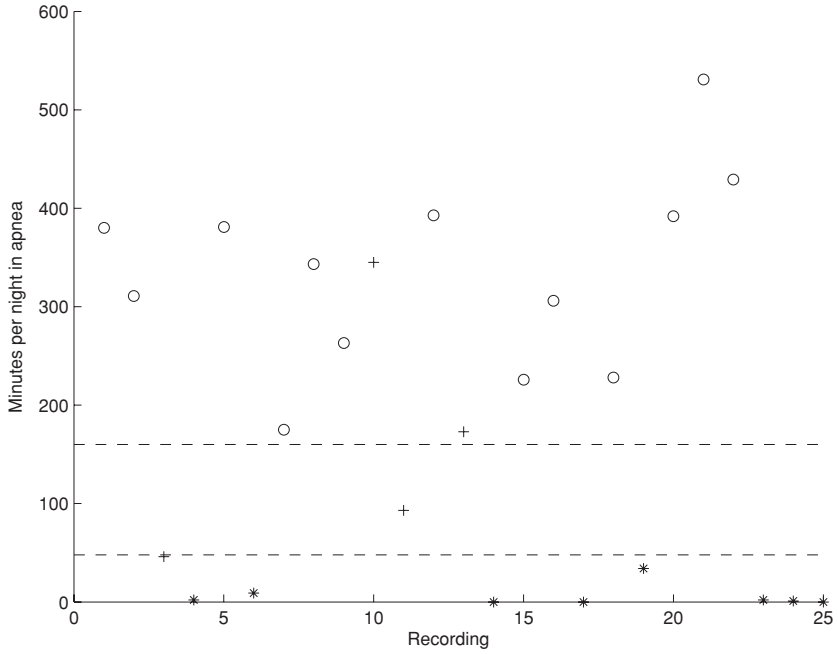


Figure 3. Class separation based on minutes per night calculated by the LD-EMD classifier processing 20 features for 25 recordings of the test set. Note that applying a threshold of 55 min per night, apnea and normal classes are separated. To compare with real classification, ∘ represents subjects with apnea, + is borderline and ∗ is for normal subjects.

Figure 5 shows how well the apnea-hypopnea index (AHI) is correlated with the minutes in apnea defined by an expert, as well as the correlation of the AHI with the minutes in apnea detected automatically when the EMD or WA was used as feature extractors. From the figure, it is possible to note that the AHI correlates very well with the minutes defined by an expert, this is around 0.95%. Regardless from the fact that the automatic classification does not present so high correlation values, the values are very good (i.e. 0.84 with EMD and 0.88 with WA). However, high correlation values, up to 0.97, were obtained among the minutes in apnea defined by an expert and those computed automatically. It is worth noting from the results, that WA seems to perform better.

Table 2 shows the ten most important features selected from EMD and WA pools independently and those selected when EMD and WA features were pooled together during the SFS. Note how the second and fifth levels of the RR interval decomposition are the most important features in both classifiers. In addition, most of the features belong to the RR interval series. For the EMD and WA pool, we can observe that most of the features comes from WA.

5. Discussion

This study deals with automatic methods to screen obstructive sleep apnea in a non-invasive and simple way and more precisely the study of EMD and WA behaviour when the HRV signal is analysed. The presented methods produce a minute-by-minute classification with an accuracy similar to the best performing classification algorithms published before (Penzel

Table 2. List of used features for both empirical mode decomposition and wavelet analysis.

Features	Empirical mode decomposition
1	RR: mean ratio modulus IMF2 to total
2	RR: mean ratio modulus IMF5 to total
3	RR, AR: mean modulus IMF2 X IMF2
4	RR: mean ratio modulus IMF2 by IMF3
5	RR: mean ratio modulus IMF1 by IMF3
6	RR: mean ratio modulus IMF1 by IMF2
7	RR: standard deviation frequency IMF2
8	RR: mean ratio modulus IMF2 to total
9	RR: mean ratio modulus IMF3 to total
10	AR: variance ratio modulus IMF3 to total
	Wavelet analysis
1	RR: mean ratio modulus scale5 to total
2	RR: mean ratio modulus scale2 to total
3	RR, AR: mean modulus scale2 X scale2
4	RR: mean ratio modulus scale2 to total
5	RR: mean Triangular
6	RR, AR: ratio mean ratios modulus scale2 to total
7	RR: mean ratio modulus scale1 by scale3
8	RR, AR variance modulus scale5 X scale5
9	AR: mean ratio modulus scale5 to total
10	AR: mean ratio modulus scale3 to total
	EMD and Wavelet Analysis
1	RR-WA: mean ratio modulus scale5 to total
2	RR-WA: mean ratio modulus scale2 to total
3	RR-WA, AR: mean modulus scale2 X scale2
4	RR-WA: mean ratio modulus scale2 to total
5	RR-WA: mean Triangular
6	RR, AR -WA: ratio mean ratios modulus scale2 to total
7	RR-WA: mean ratio modulus scale1 by scale3
8	RR, AR -WA :variance modulus scale5 X scale5
9	AR-WA: mean ratio modulus scale5 to total
10	RR-EMD: mean ratio modulus IMF5 to total

et al 2002), using a generally smaller set of features. Our main observations are that (a) HRV and QRS area yield fundamental information on the physiologic process that can be used to develop simple and fast sleep apnea screening methods, (b) WA and EMD present different ways to decompose a signal (WA in diatomic scales and EMD in nonlinear way); however, the dynamic of the cardio-respiratory system is contained in the decomposition levels that represent the respiratory frequency and the apnea repetition (see figure 1) in both approaches. This leads to the consideration that the most relevant information, for apnea detection, is the respiratory arrhythmia in the heart frequency and the frequency of repetition of the apneas. This finding confirms previous results obtained through different methodologies (Penzel *et al* 2002, Mendez *et al* 2009, de Chazal *et al* 2003).

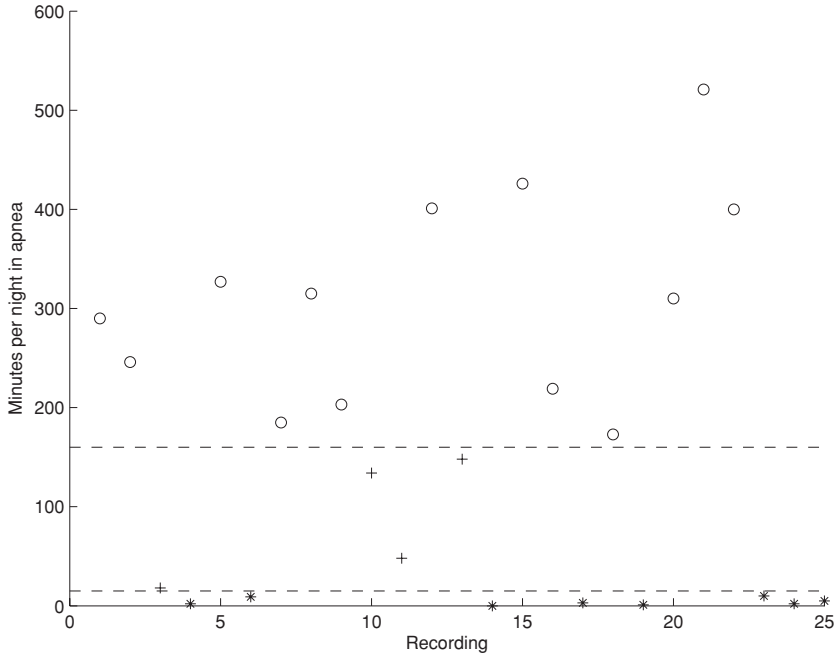


Figure 4. Class separation based on minutes per night calculated by the QD-WA classifier processing 10 features for 25 recordings of the test set. Note that applying a threshold of 15 min per night, apnea and normal classes are separated. To compare with real classification, ○ represents subjects with apnea, + is borderline and ∗ is for normal subjects.

The originality of this study lies mainly in the application of EMD and WA combined with HT. In the closest related earlier work of Mietus (Mietus *et al* 2000) the Hilbert transform is used after applying bandpass filtering to the HRV signal, and the use of the EMD however had never been investigated extensively before. Also this combination of methods had never been investigated in detail before.

It is worth noting that the most important feature for WA is the detail at scale 5 that captures the apnea repetition while for EMD the most important feature is IMF 2 where the respiratory frequency is found (see figure 1). This result suggests that EMD follows better the fast dynamics while WA the slow ones (see table 2). This can be clearly appreciated from figure 1. Another important feature is the relation between AR and RR at the second level for WA and EMD decompositions that suggests a close relation between QRS area and RR intervals variations due to respiration. The instantaneous frequencies from the HT were never included in the feature subset selected by SFS; those features did not seem to provide useful information for classification. One notable point, in the features selected by SFS, is that most of the features belong to the RR measure. This selection could be a right choice, since R peaks are quite robust to noise, while the QRS area could be largely affected by the electrode position or probably by the heart electrical axis of the subject.

On the other hand, it seems that even if the features that come from the sample entropy and standard time HRV measures contain fundamental information of the process, they do not override those coming from the WA and EMD. Another important point is the feature selection when features from EMD and WA are combined. The most important features belonged to WA, however, EMD features also are found in the best feature subset which indicates the

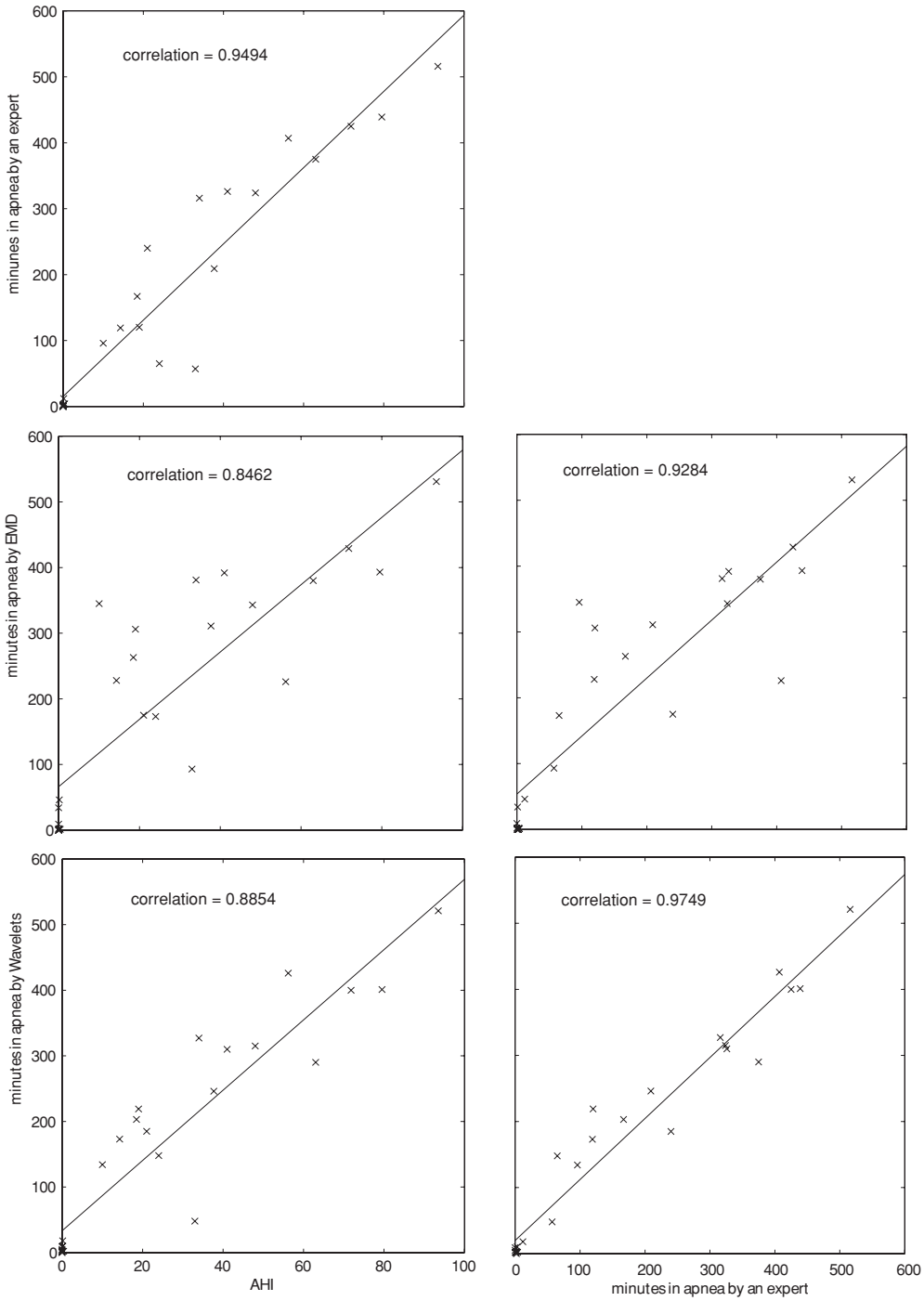


Figure 5. Correlations among AHI, minutes in apnea by an expert and minutes in apnea detected automatically by Wavelets or EMD

complementarity of information between the WA and EMD decompositions. Further to this mark, it is worth mentioning that if one decomposes a time series linearly in scales (WA) or nonlinearly in modes (EMD), the information remains in levels or scales that represent the same frequency. This information provides better insight into the nonlinear and linear intrinsic dynamics of the HRV.

One of the major problems of HT, when it is used for HRV analysis, is the large spread in the instantaneous frequency on IMF1 due to the fast variations between consecutive beats. Some studies have tried to solve this problem in order to better explain or better define this situation. Rilling *et al* (2003) proposed to couple the EMD with the reassigned spectrogram to extract in the clearest way the real information on the heart rate fluctuations. This method tries to correct the intrinsic blurring in a spectrogram caused by the limited size of the analysis window by reassigning the symmetrical energy distributions due to the blurring back to their original positions, which is supposed to be the centre of gravity of the local energy distributions (Auger and Flandrin 1994). The slightly higher sensitivity of this spectrogram method improved the accuracy of this method a little bit compared to the method with EMD and HT. However as the computation time needed for calculating the reassignment of the spectrogram is much longer, one may still prefer the method with HT to guarantee a better usability of the algorithm. In another attempt to overcome the high spread in the IMF1 when HT deals with HRV signals, Huang (1998) used a mean Gaussian window. All methods are good and could be used indifferently since both give a rough measure of the high frequency dynamic of HRV. A thorough analysis of those methods together with the classical ones such as time-variant and time-frequency approaches could be important in order to define the real accuracy of the approaches when instantaneous frequency dynamics have to be analysed.

This study used a variety of features and the best combination was selected based on their classification performance. The feature information which is eventually used corresponds to the information used by experts during manual classification, which confirms the right choice of features by the automatic feature selection mechanism.

Finally, the evaluation of sleep apnea based only on features extracted from ECG seems to have a limit around 90% independently from the method for feature extraction and classifier that can be noted from the current and previous studies (Penzel *et al* 2002). However, this difference between automatic and human scoring remains smaller to the inter-rater variability of visual computer screen-based scoring since scoring between experts presents an agreement level around 85% (Penzel and Conradt 2000). The similar performance, in all the automatic approaches, confirms the hypothesis that automatic screening is robust and reproducible, and then this idea reinforces the possible use of automatic screening for supporting the medical decision and evaluating the sleep apnea in different environments as home. However, more efforts must be done in order to standardize or define the features which carry fundamental information for the apnea classification.

It is important to stress that such a result is connected to the quality of the signal, and, especially for long-term recordings during the night, some periods may occur in which the signal is really bad, absent or the heart rhythm is destroyed by the ectopic beats. When the ECG presents a low signal-to-noise ratio, any R peak detector could fail and then alternative measures have to be taken in order to obtain reliable results. For instance, if some isolated peaks are misdetected, the procedure described in section correction of RR intervals could be used, as well as the procedure described in de Chazal *et al* (2003) and Mendez *et al* (2009) when consecutive misdetected beats occur (from 3 to 10 misdetected beats). In addition, when more than ten consecutive misdetected beats occur, the whole 1 min epoch could be discarded and remains as unclassified. Thus if these situations are managed correctly, the performance of the current screening procedure is not largely affected. However, for future

clinical applications, it will be necessary to define the minimal amount of the total night recording needed for a reliable diagnosis.

It is worth remembering, however, that the method for the detection of apnea episodes is proposed as a tool for home monitoring or screening and is not intended as a new diagnostic tool that could replace the current standard procedure for a complete and exhaustive diagnosis. It is rather a support tool for screening of people who could be at risk for apneas especially in departments of cardiology where long-term ECG recording is a standard tool. People identified with some degree of apneic events through the present analysis will be object of a further clinical examination and referred to a sleep centre with sleep experts and may be investigated with cardiorespiratory polysomnography. As the extracted features come mainly from the analysis of the HRV signal, any pathology, event or drug that have an effect on the HRV may influence the accuracy in the detection of the apneic episodes. It is beyond the aims of the present work to provide detailed guidelines for the clinical use of the methods. To do this work, a prospective study on a cardiology recruited patient group is needed. The aim of the present work is to demonstrate the feasibility of an apnea screening based on ECG analysis as can be observed from the current results. In order to come to a clinical application, further investigations are still required on a larger number of subjects covering different subgroups of the general population (i.e. different age ranges, presence of particular pathologies, etc).

6. Conclusions

The obtained results, with comparable accuracy up to 89%, showed that apnea screening can be done in a very simple way using only data obtainable from the ECG signal, and in multiple alternative ways. Moreover apneic and normal subjects could be separated completely. These results further confirm the possibility of developing fast home screenings for sleep apnea patients. Though EMD is a promising method, EMD did not seem to outperform WA in this application; the latter performed slightly better. The frequency information found at the lowest level and pointing to the apnea repetition was one of the main features for both methods. The importance of the first IMF1 in EMD, however, could indicate a significant difference with WA. Nevertheless in this first study on the use of EMD for the automatic screening of obstructive sleep apnea, the possible usefulness of the EMD in HRV analysis has been shown and it could be interesting to investigate its use in more different applications. Finally, our results suggest that the apnea minute-by-minute classification could be a useful index for screening proposes, it is because the minute-by-minute classification presents a close relation with the AHI.

Acknowledgments

This work was supported in part by the HeartCycle project ICT FP7 216695 of the European Community and by the Government of Mexico (convenio CONACyT-UASLP No. 290554). S Van Huffel is supported by the Research Council KUL: GOA-AMBioRICS, CoE EF/05/006 Optimization in Engineering (OPTEC), and by the Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, 'Dynamical systems, control and optimization', 2007–2011)

References

Ancoli-Israel S, DuHamel E R, Stepnowsky C, Engler R, Cohen-Zion M and Marler M 2003 The relationship between congestive heart failure, sleep apnea, and mortality in older men *Chest* **124** 1400–5

American Academy of Sleep Medicine Task Force 1999 Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research *Sleep* **22** 667–89

Auger F and Flandrin P 1994 The why and how of time-frequency reassignment *IEEE Int. Symp. on Time-Frequency and Time-Scale Analysis* pp 197–200

de Chazal P, Heneghan C, Sheridan E, Reilly R, Nolan P and O'Malley M 2003 Automated processing of the single lead electrocardiogram for the detection of obstructive sleep apnoea *IEEE Trans. Biomed. Eng.* **50** 686–96

Duda R O, Hart P E and Stork D G 2001 *Pattern Classification* (New York: Wiley)

Himmelmann A 1999 Hypertension: an important precursor of heart failure *Blood Press* **8** 253–60

Huang N E, Shen Z, Long S L, Wu M C, Shih H H, Zheng Q, Yen N, Tung C C and Liu H H 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis *Proc. R. Soc.* **454** 903–95

John G H, Kohavi R and Pfleger K 1994 Irrelevant features and the subset selection problem *Proc. 11th Int. Conf. on Machine Learning* pp 121–9

Kohavi R and John G H 1997 Wrappers for feature subset selection *Artif. Intell.* **97** 273–324

Malhotra A and White D P 2002 Obstructive sleep apnoea *Lancet* **360** 237–45

Martin S E, Engleman H M, Deary I J and Douglas N J 1996 The effect of sleep fragmentation on daytime function *AJRCCM* **153** 1328–32

Mendez M O, Bianchi A M, Matteucci M, Cerutti S and Penzel T 2009 Sleep apnea screening by autoregressive models from a single ECG lead *Trans. Biomed. Eng.* **56** 2828–50

Mietus J E, Peng C K, Ivanov P C and Goldberger A L 2000 Detection of obstructive sleep apnea from cardiac interbeat interval time series *Comput. Cardiology* **27** 753–6

Penzel T, Moody G B, Mark R G, Goldberger A L and Peter J H 2000 The apnea-ECG database *Comput. Cardiol.* **27** 255–8

Penzel T and Conradt R 2000 Computer based sleep recording and analysis *Sleep Med. Rev.* **4** 131–48

Penzel T, McNames J, Murray A, de Chazal P, Moody G and Raymond B 2002 Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings *Med. Biol. Eng. Comput.* **40** 402–7

Pichot V, Gaspoz J, Molliex S, Antoniadis A, Busso T, Roche F, Costes F, Quintin L, Lacour J and Barthelemy J 1999 Wavelet transform to quantify heart rate variability and to assess its instantaneous changes *J. Appl. Physiol.* **86** 1081–91

Rilling G, Flandrin P and Gonçalvès P 2003 On empirical mode decomposition and its algorithms *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03 Grado(I)*

Task Force of the European Society of Cardiology and the North American Society of Pacing Electrophysiology 1996 Heart rate variability: standards of measurement, physiological interpretation, and clinical use *Circulation* **93** 1043–65

Thakor N V, Gramatikov B and Sherman D 2000 Wavelet (time-scale) analysis in biomedical signal processing *The Biomedical Engineering Handbook* 2nd edn (Boca Raton, FL: CRC Press LLC) pp 886–905

Torrence C and Campo G P 1998 A practical guide to wavelet analysis *Bull. Am. Meteorol. Soc.* **79** 61–78

Young T, Palta M, Dempsey J, Skatrud J, Weber S and Badr S 1993 The occurrence of sleep-disordered breathing among middle-aged adults *N. Engl. J. Med.* **3281** 230–35

Young T, Evans L, Finn L and Palta M 1997 Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women *Sleep* **20** 705–6