# Sleep Apnea Screening by Autoregressive Models From a Single ECG Lead

Martin O. Mendez*, Anna Maria Bianchi, *Member, IEEE*, Matteo Matteucci, *Member, IEEE*,
Sergio Cerutti, *Fellow, IEEE*, and Thomas Penzel, *Senior Member, IEEE*

*Abstract*—This paper presents a method for obstructive sleep apnea (OSA) screening based on the electrocardiogram (ECG) recording during sleep. OSA is a common sleep disorder produced by repetitive occlusions in the upper airways and this phenomenon can usually be observed also in other peripheral systems such as the cardiovascular system. Then the extraction of ECG characteristics, such as the RR intervals and the area of the QRS complex, is useful to evaluate the sleep apnea in noninvasive way. In the presented analysis, 50 recordings coming from the apnea Physionet database were used; data were split into two sets, the training and the testing set, each of which was composed of 25 recordings. A bivariate time-varying autoregressive model (TVAM) was used to evaluate beat-by-beat power spectral densities for both the RR intervals and the QRS complex areas. Temporal and spectral features were changed on a minute-by-minute basis since apnea annotations where given with this resolution. The training set consisted of 4950 apneic and 7127 nonapneic minutes while the testing set had 4428 apneic and 7927 nonapneic minutes. The K-nearest neighbor (KNN) and neural networks (NN) supervised learning classifiers were employed to classify apnea and non apnea minutes. A sequential forward selection was used to select the best feature subset in a wrapper setting. With ten features the KNN algorithm reached an accuracy of 88%, sensitivity equal to 85%, and specificity up to 90%, while NN reached accuracy equal to 88%, sensitivity equal to 89% and specificity equal to 86%. In addition to the minute-by-minute classification, the results showed that the two classifiers are able to separate entirely (100%) the normal recordings from the apneic recordings. Finally, an additional database with eight recordings annotated as normal or apneic was used to test again the classifiers. Also in this new dataset, the results showed a complete separation between apneic and normal recordings.

*Index Terms*—Electrocardiogram, heart rate variability, neural network, pattern classification, time-varying autoregressive model.

## I. INTRODUCTION

**S**LEEP apnea is a common sleep disorder characterized by the repetitive cessation of breathing during sleep. In clinics, sleep apnea is divided into three classes: obstructive, central, and mixed. Obstructive sleep apnea (OSA) is the most common form in the general population, with prevalence between 2% and 4%, and is commonly associated with aging and obesity. Obstructive apnea consists of an interruption of the airflow to the lungs caused by a collapse in the upper airways. Regardless of the type of sleep apnea, it is typically accompanied by reductions in blood oxygen saturation and arousal events. During obstructive sleep apnea, respiratory muscles produce mechanical effort in order to overcome the occlusion. If these efforts are not sufficient, the blood oxygen level begins to decrease, and muscle effort increases in response to the hypoxia until an arousal event takes place to reactivate all of the systems and restore the normal respiration [1]. In mild and severe stages, sleep apnea events may occur hundreds of times during a single night with serious health and social consequences. The most common sleep apnea symptoms are sleepiness during the day, irritability, tiredness, impaired concentration, and reduced learning capabilities [2]. These symptoms typically produce even more serious consequences, including social problems in the work place and traffic accidents. In addition, severe OSA generates diurnal hypertension and severe cardiovascular health implications that can lead to death when apnea remains undiagnosed or is not properly treated [3].

In order to diagnose sleep apnea, dedicated personnel, adequate infrastructure, and special acquisition systems are needed, which make this diagnosis an expensive procedure. OSA diagnosis, in fact, requires the analysis of different signals that are obtained from polysomnography recordings taken during the duration of one night in a specialized sleep center. Due to the small number of sleep centers with specialized personnel, the diagnosis of sleep apnea is not accessible to the general population and, as a result, OSA pathology is nowadays generally underestimated. These difficulties motivate new efforts aimed at obtaining simpler and more accessible ways of evaluating sleep apnea.

Although sleep apnea is a respiratory event, its effects can be clearly observable within other peripheral systems such as the cardiovascular system. Due to this relationship, the electrocardiogram (ECG) can provide very valuable information about apneic events and has been broadly studied for apnea detection [4]. One of the most important signals that can be obtained from the ECG is the beat-by-beat series of heart rate. This signal contains fluctuations, commonly named heart rate variability (HRV), which present frequency components between 0 and 0.5 Hz and are linked to the Autonomic Nervous System (ANS) function. Frequency components between 0.15 and 0.5 Hz are generally associated to the vagal tone and are

*M. O. Mendez is with the Department of Biomedical Engineering, Politecnico di Milano, Milano 20133, Italy (e-mail: martin.mendez@biomed.polimi.it).

A. M. Bianchi and S. Cerutti are with the Department of Biomedical Engineering, Politecnico di Milano, Milano 20133, Italy (e-mail: annamaria.bianchi@polimi.it; sergio.cerutti@polimi.it).

M. Matteucci is with the Department of Electronics and Information, Politecnico di Milano, Milano 20133, Italy (e-mail: matteucc@elet.polimi.it).

T. Penzel is with the Sleep Department, Charité Hospital, Berlin 10117 Germany (e-mail: thomas.penzel@charite.de).

known as high frequency components (HF). Frequencies from 0.02 to 0.15 Hz are the manifestation of the activation of both parasympathetic and sympathetic systems and are labeled low frequency components (LF). An increase in the LF power is generally associated to the orthosympatic activation. Finally, frequencies between 0.0033 and 0.02 Hz contain information regarding slow processes such as thermoregulation [5]. The ratio between HF and LF spectral powers is defined as a measure of the sympatho-vagal balance [6]. During sleep, HRV presents specific dynamics [7] and complexity [8] that are characteristics of non-rapid eye movement sleep (NREM) and rapid eye movement (REM) sleep [9]. The power spectral distribution of HRV signals show powers highly concentrated around 0.3 Hz during NREM sleep, in contrast to REM sleep where the high frequency components are less peaked and the low frequency components are prevalent. However, when sleep apnea occurs, there is a reduction in the HRV complexity and frequency components appear around 0.02 Hz as a result of the apnea repetition.

In the last several years, some research has focused on developing automatic screening tools for OSA based on the analysis of the ECG with a variety of signal processing and pattern recognition techniques [4], [10]. Some studies used features extracted from the ECG such as HRV, QRS complex amplitude, T duration, R area, and peripheral tonometry. Results obtained by those studies showed good classification levels between normal and apneic periods. Among the different approaches for classifying apnea, we can find the Hilbert transform, wavelets decomposition, the Fourier transform, time–frequency analysis and other techniques which are mainly used to extract relevant features [4]. Among the apnea screening methods presented in literature, de Chazal *et al.* [10] propose one of the best, it is able to achieve 90% of correct classification on minute-by-minute basis. It extracts spectral features through Fourier transform of both RR series and ECG morphological characteristics. In our opinion, its main weakness is in the high-dimension feature space, i.e., 88 different features. This is probably due to the fact that the features are not chosen taking into consideration physiological mechanisms (this is the so called black-box approach). The aim of this paper was to propose a different approach able to achieve a similar level of performance, but with a smaller number of features for the classification. To reach this objective we focused on two critical aspects: 1) knowledge of the physiological phenomena implied in the OSA episodes to extract features able to better describe the underlying mechanisms; 2) use of time-variant models able to deal with the characteristic nonstationarity of the signals of interest during sleep. In addition, two different classification methods were compared on the same set of features, to clarify the influence of the classification methodology on the final results.

## II. MATERIALS AND METHODS

### A. Protocol

The data analyzed in the present research come from the Physionet website www.physionet.org. The ECG data from complete night polysomnography recordings were acquired with a sampling rate of 100 Hz, 16-bit resolution, and modified lead V2 electrode configuration. Each recording has a duration close to 8 h. Apnea scoring was carried out by expert personnel, utilizing polysomnographic data, according to the standard clinical criteria as reported in [1]. The annotation was provided on a minute-by-minute basis: a minute was labeled as apneic if it contained at least one apneic or hypo-apneic episode, and if not it was defined as nonapneic [11]. The subject age ranged in the dataset between 27 and 63 years ($48 \pm 10.8$ years) and subject weight ranged between 53 and 135 kg ($86.3 \pm 22.2$ kg). In the Physionet data, the recordings are classified into three groups: apnea group (class A, with more than 100 minutes in apnea, mean age of 50 years with a range between 29 and 63 years), Borderline group (class B with total apnea duration more than 5 and less than 99 min, mean age of 46 years with a range between 39 and 53 years) and normal group (class C, with less than 5 min in apnea, mean age of 33 years with a range between 27 and 47 years). The database also contains the QRS complex occurrence times for each recording. From a total of 70 recordings in the database (35 records called withheld set and 35 records called release set), we selected the recordings that satisfied the following criteria:

1) no periods with more than eight consecutive misdetected QRS complexes,
2) no presence of a large quantity of ectopic beats in the recording.

A total of 50 recordings were selected, of which 25 (belonging to physionet release set) were used in the preliminary analysis needed to develop and training our algorithms for feature extraction and classification. This training set consisted in 15 apneic, 2 borderline, and 8 normal recordings with 4950 min with apnea and 7127 without apnea episodes. The other 25 recordings (belonging to physionet withheld set) were used to measure the performance of our algorithms. This testing set contained 13 apneic, 4 borderline and 8 control recordings with 4428 minutes with apnea and 7927 without apnea episodes. The algorithm performance was measured by comparing its minute-by-minute classification to the minute-by-minute annotation provided in the Physionet website. In addition to the Physionet database, eight new recordings coming from Philipps University were used for a further verification of our screening systems. These new recordings were acquired at 100 Hz with 16-bit resolution and modified lead V2 electrode configuration. The subjects had the following characteristics: weight was $76.75 \pm 20.98$ kg, age of $40.75 \pm 16.14$ years and $14.61 \pm 23.44$ as apnea/hypoapena index [(AHI) it is the hour average of apnea and hypoapnea episodes during sleep]. From [1], an apnea/hypoapnea event is defined as a transient reduction in, or complete cessation of, breathing. These episode must fulfill the following criteria,

1) 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 two minutes preceding onset of the event or the mean amplitude of the three largest breaths in the two minutes preceding onset of the event.
2) Clear amplitude reduction of a validated measured breathing during sleep that not reach the above criterion but is

associated with either an oxygen saturation of $> 3\%$ or an arousal.

3) The event last 10 s or longer.

The recordings were annotated by expert personnel as either normal or apneic based on the standard criterion [1].

### B. Methods

From the single lead of the ECG signal, QRS complex area (QRS area) and RR intervals were derived (the algorithm described in [12] was used to find the R peak occurring times in the recordings coming from Philipps University, while for the recordings coming from physionet was not necessary since these recordings already had the QRS occurring times). At this point, a set of features were extracted from this data through time and frequency domain analysis and then the features were used for classification. Fig. 1 shows an example demonstrating the derivation of the RR intervals and the QRS area obtained from a subject during normal (right panel) and apneic (left panel) conditions. The RR intervals time series shows oscillatory patterns of brady-tachycardia during apnea events. These correspond to a very low frequency component (around 0.02 Hz) in the RR intervals spectrum. The same oscillatory patterns occur in the QRS area time series, which is related to the respiratory activity. A high value of coherence does exist between the two time series at very low frequencies. The time series and the related power spectra during normal breathing are also presented in Fig. 1. In such a case the main oscillation, in both the signals, is at the respiration frequency (around 0.25 Hz). This is also true for the coherence function. From this, it is clear that how both time and spectral parameters extracted from RR intervals and QRS area can be considered useful features to be fed into the classifier. Nonparametric (KNN) and parametric (neural networks) supervised classifiers are compared in this paper.

### C. Correction of RR Intervals

From the QRS complex occurrence time provided in Physionet, RR series were extracted. Resulting series were plotted on the PC screen with the related ECG for the manual correction of misdetected beats. After correction, the RR time series were computed. However, after the visual inspection and correction, some RR intervals still presented nonphysiological distances (i.e., ectopic beats). In order to correct those intervals, the following procedure was implemented. The RR mean value was calculated using a ten-beat moving average filter. A search was performed for all RR intervals 30% higher and 20% lower with respect to the mean at each beat. RR intervals that met this condition were substituted by the mean RR value at that time [10].

### D. ECG and QRS Area

Previous studies analyzed the QRS area variations during both normal respiration and respiratory maneuvers such as Valsalva. They reported variations on the QRS area with frequencies close to those of the mechanical movements produced by respiration (see Fig. 1). These changes in the ECG are mainly generated by the variations in the relative distance between the electrode
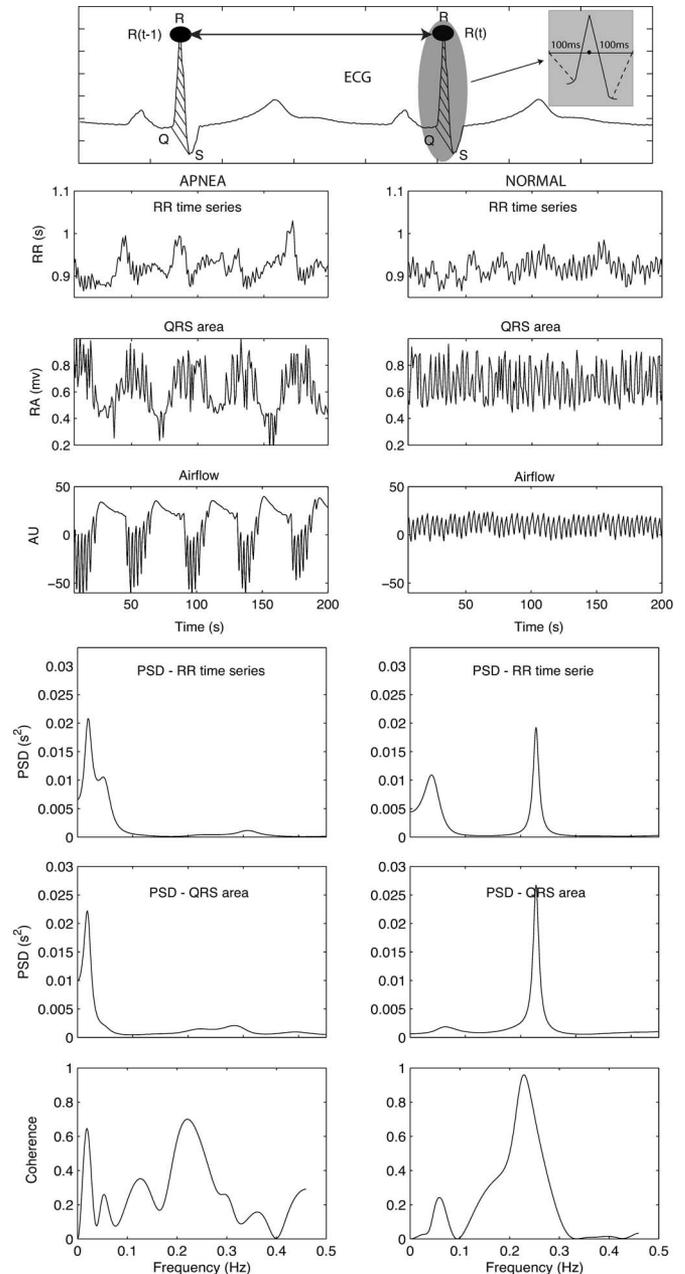


Fig. 1. Temporal and spectral relationship between RR intervals and QRS area during apneic and normal sleep. From top to bottom: ECG, RR intervals, QRS area, airflow, power spectrum of the RR intervals, power spectrum of the QRS area and coherence. Left column is related to the apneic breathing, while right column is related to normal breathing during sleep.

located on the chest surface and the heart, and by the changes in the thoracic impedance produced by the inflation and deflation of the lungs. For this reason, the QRS area could be used as a rough estimation of the respiratory activity. The QRS area was calculated by subtracting the ECG baseline from the original ECG. The baseline was calculated by a median filter of 200 ms width[1] [10], [13]. Using the resulting ECG, the minimum value

---

[1]This is a nonlinear technique that slides a window to a sequence, replacing the center value in the window with the median value of the window samples.

was found within a temporal window 100 ms before and after the maximum R peak value. Then, the area inside this region was calculated by summing up all of the sample values.

### E. Time-Varying Autoregressive Model (TVAM)

Frequency analysis of the signals provides useful spectral parameters that have important physiological meanings, as is the case with HRV. Parameters are in general estimated under the hypothesis of stationarity, but when the signal to be analyzed is nonstationary, as is the case of HRV during sleep time, it is necessary to use techniques such as TVAMs, which are based on a filter able to adapt its parameters with each new sample. An adaptive filter self-adjusts its transfer function according to the characteristics of the input signal, and thus seems particularly suitable for this purpose. Let $x$ be a row vector with $N$ values defined as $x = |x(1), x(2), \ldots, x(N)|$. The prediction (approximation) of the current output at time $n$ for an all pole system is defined as follows:

$$\tilde{x}(n) = -\sum_{s=1}^{S} a_s x(n-s) \tag{1}$$

where the $a_s$ values are coefficients of the filter (Model), $S$ the filter order, and $\tilde{x}(n)$ denotes the prediction output. Thus the prediction error is evaluated as

$$e(n) = x(n) - \tilde{x}(n) = x(n) + \sum_{s=1}^{S} a_s x(n-s). \tag{2}$$

Each time the system moves toward a new input sample $x(n+1)$, the vector $\mathbf{a}$ of the coefficients is updated in order to follow the variation in the time series. This update can be achieved in a recursive way by

$$\mathbf{a}(n+1) = \mathbf{a}(n) + \mathbf{K}(n)e(n) \tag{3}$$

where $\mathbf{K}(n)$ is the analogue of the Kalman gain, which is updated sample by sample [14]. The previous relations can be extended to the bivariate case, where the prediction can be written as

$$\tilde{\mathbf{x}}(n) = -\sum_{s=1}^{S} \mathbf{A}_s \mathbf{x}(n-s) \tag{4}$$

where

$$\mathbf{A}_s = \begin{vmatrix} a_{11}(s) & a_{12}(s) \\ a_{21}(s) & a_{22}(s) \end{vmatrix} \tag{5}$$

$$\mathbf{x}(n) = \begin{vmatrix} x_1(n) \\ x_2(n) \end{vmatrix}. \tag{6}$$

We can observe that for a specific time $n$, the characteristics of the model system are described by the following matrix:

$$\mathbf{A}(n) = |\, \mathbf{A}_1(n) \quad \mathbf{A}_2(n) \quad \cdots \quad \mathbf{A}_s(n)\,| \tag{7}$$

and, for a multichannel AR model, the power spectral densities are given by [15]

$$\mathbf{P}_n(f) = (\mathbf{A}_n(f)^{-1} \mathbf{V}_n \mathbf{A}_n(f))^{-H} T \tag{8}$$

where $T$ is the sampling period, $H$ denotes the Hermitian transpose, $\mathbf{V}$ is the covariance matrix of the multichannel noise input process at time $n$ (covariance of the prediction errors) and, being $\mathbf{I}$ the identity matrix, $\mathbf{A}(f)$ is given by

$$\mathbf{A}_n(f) = \mathbf{I} + \sum_{s=1}^{S} \mathbf{A}_n(s)\exp^{-2\pi i s f T}. \tag{9}$$

As a result, we obtain the following Hermitian power spectral density matrix

$$\mathbf{P}_n(f) = \begin{vmatrix} P_{11}(f) & P_{12}(f) \\ P_{21}(f) & P_{22}(f) \end{vmatrix} \tag{10}$$

where the diagonal elements contain the spectra of individual channels and the off-daigonal elements contain the cross-spectra. Finally, we can evaluate a measure of the spectral dependence between two channels. This measure is called complex *coherence* and it is computed as

$$\gamma_{12_n}(f) = \frac{P_{12_n}(f)}{\sqrt{P_{11_n}(f)}\sqrt{P_{22_n}(f)}} \tag{11}$$

the coherence spectrum between two channels is given by

$$\eta_{12_n}(f) = |\gamma_{12_n}(f)|^2 \tag{12}$$

while the phase spectrum between two channels is given by

$$\theta_{12_n}(f) = \tan^{-1}\left[\frac{\text{Im}(\gamma_{12_n}(f))}{\text{Re}(\gamma_{12_n}(f))}\right]. \tag{13}$$

For further details see [16].

### F. KNN Algorithm

KNN is a nonparametric technique for classification, where it is assumed that we have no *a priori* parameterized knowledge about the probability structure of the data [17] and data samples are directly used instead. In short, the KNN algorithm assigns to a new observation the most frequent label among the samples in the training set that are closest to it. This simple rule implements a nonparametric Bayesian classifier that maximizes the *a posteriori* probability of label class $w \in w_1, \ldots, w_M$ given the observed data $\mathbf{x}$ in the classical form

$$p(w|\mathbf{x}) = \frac{p(\mathbf{x}|w)p(w)}{p(\mathbf{x})}. \tag{14}$$

The theory behind this method can be summarized in the implementation of a nonparametric estimate of the unknown probability density $p(\mathbf{x})$, for the data. This estimate is based on the fact that the probability $P$, for each vector $\mathbf{x}$, to fall in a (possibly small) region $\Re$ around it can be evaluated by

$$P = \int_{\Re} p(\mathbf{x}')d\mathbf{x}'. \tag{15}$$

$P$ is the expected value of $p(\mathbf{x})$ in a given region around a sample; one can estimate a smoothed value of $p(\mathbf{x})$ for the data by locally estimating the probability $P$ around each possible sample. Assuming $p(\mathbf{x})$ is constant in a (possibly small) region

$\Re$ around a sample $\mathbf{x}$, we can write

$$P = \int_{\Re} p(\mathbf{x}')d\mathbf{x}' \simeq p(\mathbf{x})V \tag{16}$$

where $V$ is the volume enclosed by the region $\Re$ around $\mathbf{x}$.

Let us now assume we have $n$ independent and identically distributed samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ drawn from the unknown probability distribution $p(\mathbf{x})$. Using a maximum likelihood approach, we can estimate the probability $P = k/n$ as the number $k$ of samples falling in a given region of volume $V$ around a sample. This gives us an estimate of the data density

$$p(\mathbf{x}_n) \simeq \frac{k/n}{V}. \tag{17}$$

A KNN classifier specifies $P$ taking, for each record, a surrounding region capable of containing a fixed number of samples $K$; this region, depending on the specific record of the datasets it is referred to, will have a different volume $V_n$ and the data density will be

$$p(\mathbf{x}_n) \simeq \frac{k/n}{V} = \frac{K}{V_n} \tag{18}$$

where $V_n$ is the volume of the region around $\mathbf{x}_n$, i.e., its neighborhood, containing $K$ samples. We can then estimate the *a posteriori* probabilities of the label for a new sample $P(w_i|\mathbf{x}_{\text{new}})$ from a set of labeled samples using this result.

Let us place a cell of volume $V_{\text{new}}$ around $\mathbf{x}_{\text{new}}$ so as to capture $K$ samples, of which $K_i \leq K$ subsequently labeled $w_i$. Doing so, the estimate for the joint probability $p(\mathbf{x}_{\text{new}}, w_i)$ becomes

$$p(\mathbf{x}_{\text{new}}, w_i) = \frac{K_i}{V_{\text{new}}} \tag{19}$$

and thus a reasonable estimate for the *a posteriori* probability over the possible classes $P(w_i|\mathbf{x})$ is computed as

$$P(w_i|\mathbf{x}_{\text{new}}) = \frac{p(\mathbf{x}_{\text{new}}|w_i)p(w_i)}{p(\mathbf{x}_{\text{new}})} = \frac{p(\mathbf{x}_{\text{new}}|w_i)p(w_i)}{\sum_{j=1}^{c} p(\mathbf{x}_{\text{new}}|w_j)p(w_j)} \tag{20}$$

$$= \frac{p(\mathbf{x}_{\text{new}}, w_i)}{\sum_{j=1}^{c} p(\mathbf{x}_{\text{new}}, w_j)} = \frac{K_i}{K} \tag{21}$$

where $c$ is the number of classes.

All this reasoning boils down to the simple fact that the *a posteriori* probability of (14) of class label $w_i$ for a given record $\mathbf{x}$ is merely the fraction of the samples within the cell around $\mathbf{x}$ containing $K$ other samples that are labelled $w_i$. For minimum error rate, we need to select the category most frequently represented within the cell, and it can be proved that, if there are enough samples and if the cell is sufficiently small, this will approach the performance of the best possible classifier [17]. In practical implementations, the local region enclosed by the nearest neighbor training samples of $\mathbf{x}_{\text{new}}$ is commonly defined in terms of the Euclidean distance in the $n$-dimensional feature space $\Re^n$ [18].

### G. Artificial Neural Networks

Artificial neural networks (ANNs) are generic nonlinear function approximators extensively used for pattern recognition and
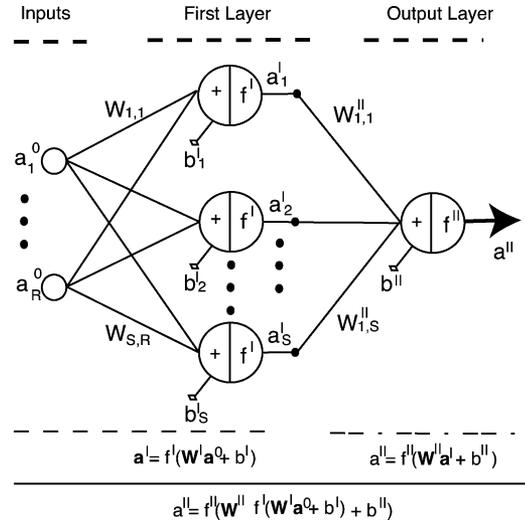


Fig. 2. Schematic representation of a feedforward neural network. $a^0$ are the input neurons, $a^I$ represent the neurons of the hidden layer and $a^{II}$ is the output neuron. $f$ is the neuron function and $b$ represents the neuron bias.

classification [19], [20]. A neural network is a collection of basic units, called neurons, computing a nonlinear function of their input. Every input has an assigned weight that determines the impact that this input has on the overall output of the node. In Fig. 2, a schematic representation of such an artificial network in shown. In this diagram, the label $w_{sr}$ is the weight of the connection from neuron $r$ to neuron $s$, and $a_s$ is the activation or output of neuron $s$. Unit $s$ output is obtained by ideally following a two step procedure. First the total weighted input $a_s$ is computed using the formula

$$a_s = \sum_r w_{sr} a_r \tag{22}$$

where $a_r$ is the activity level of the $r$th unit in the previous layer and $w_{sr}$ is the weight of the connection between the $r$th and the $s$th unit. Then, neuron output is obtained as a nonlinear function (e.g., sigmoid or hyperbolic tangent) of the total weighted input $a_s$ minus a bias term.

By interconnecting a proper number of nodes in a suitable way and by setting the weights to appropriate values, a neural network can approximate any nonlinear function with arbitrary precision [21]. This structure of nodes and connections, known as network topology, together with the weights of the connections determines the final network behavior. Fig. 2 describes a simple feed-forward topology (i.e., no loops are present) with a single hidden layer (i.e., a layer of neurons neither connected to the input nor to the output). Given a neural network topology and a training set, it is possible to optimize the values of the weights in order to minimize an error function by means of any back-propagation algorithm [22], standard optimization technique [23] or randomized algorithm [24].

The topology of a neural network plays a critical role in whether or not the network can be trained to learn a particular dataset [40]. A simple topology will result in a network that cannot learn to approximate a complex function, while a complex topology is likely to result in a network losing its generalization

capability. This loss of generalization is the result of overfitting the training data instead of approximating a function present in the data, a neural network that has an overly complex structure may have the ability to memorize the training set resulting in inaccurate predictions on future samples. In this paper, in order to improve generalization, we use the early stopping technique [25]. This consists of using a validation set to stop the training algorithm before the network starts learning noise in the data as part of the model. Error resulting on the validation set can be used also as an estimate of the generalization error and thus can be used to select a proper number of hidden neurons.

### H. Feature Set

Preprocessing (RR detection and QRS area calculation) provided two time series with physical and physiological information about the ANS and the respiratory system. Based on these time series, it is possible to extract characteristics that could be of physiological and clinical relevance. In addition to those features that have a physiological meaning, it is possible to extract related features without an explicit or direct physiological meaning, but with discrimination power that allows for class separation, as it is needed for the problem under study. To this end, from RR intervals and QRS area (QRS area was normalized with its mean value index to maintain a range value comparable with the RR intervals) phase space, two time series were derived. The first one was the module $\text{Mod}(n)$, which is defined as

$$\text{Mod}(n) = \sqrt{\text{QRSarea}(n)^2 + \text{RR}(n)^2} \qquad (23)$$

where $n$ is the beat number. The second one was called phase $\text{Pha}(n)$ and is defined as

$$\text{Pha}(n) = \tan^{-1}\left(\frac{\text{QRSarea}(n)}{\text{RR}(n)}\right). \qquad (24)$$

These latter features were generated with the expectation to highlight the relationships between RR value and respiration (that are well known in physiology, see for example [26]), through the use of simple geometrical relations.

The power spectral density (PSD) and coherence, for each RR intervals and QRS area series, were obtained by a bivariate time-varying autoregressive model at each beat. From the beat-by-beat spectra and coherence, three time series were obtained by integrating the following bands: very low frequency (VLF) = 0.003–0.04 Hz, LF = 0.04–0.15 Hz and HF = 0.15–0.5 Hz. The same spectral indexes were extracted from Mod and Pha, but in a monovariate way.

When dealing with physiological series, it is important to normalize the time series in order to eliminate the intersubject variability, which is produced by subject specific physiological limits and conditions. In this way, some inconsistencies are eliminated and it is assured that noise caused by the bio-diversity of subjects is reduced or damped. We applied two different normalization procedures to the data; each feature, coming directly from RR intervals and QRS area time series, was normalized to zero mean and unit standard deviation where the corresponding mean and standard deviation were calculated from the corresponding whole night recording. HF, LF, and VLF were normalized in two different ways: zero mean and unit standard deviation with respect to the night recording and beat-by-beat band power percentage with respect to the total power (VLF + LF + HF). Since coherence values range between 0 and 1, no normalization was applied. At this point, all the features presented a temporal beat resolution. For all the features, mean and variance were evaluated minute-by-minute in order to obtain the same temporal resolution given by the apnea annotations in the Physionet database. For RR intervals, QRS area, Mod, and Pha, kurtosis and skewness were also evaluated minute-by-minute. A total of 72 features were extracted.

### I. Selection and Transformation of the Features

To prevent the course of dimensionality, avoiding classification errors produced by domination of irrelevant features in estimating the *a posteriori* distribution for the classification, feature selection has been applied. Feature selection can be performed in different ways: it can be evaluated by statistical analysis of features, wrapper methods, principal component analysis or factor analysis [17]. Wrapper methods select the features based on the classifier performance for each possible subset in order to predict the benefits of adding or eliminating a feature from the feature set (cross-validation). Wrapper methods might have a high computational cost due to the exhaustive procedure to evaluate all possible combinations of the input features to find the best subset; to this regard, some greedy approaches have been proposed to overcome this drawback and maintain high performance on wrapper methods. Some of the most common wrapper methods used in the paper include: sequential forward selection (SFS), sequential backward selection, and bidirectional selection [27], [28]. In this study, sequential forward selection with KNN wrapper evaluation was used to find the best feature subset to classify sleep apnea.

*1) Wrapper Approach Through Leave-One-Out (LOO):* LOO cross-validation (LOOCV) is a loop procedure to compute some statistics of a classifier performance. LOOCV evaluates the classifier performance using a validation set consisting of a single observation from the whole dataset and the rest of the observations as training set. Then the evaluation of the performance is repeated such that each observation in the dataset is used once as the validation set. Finally, mean and variance of the classifier performance are computed, giving a statistical evaluation of the classification.

SFS has been used in this work to find the best feature subset that maximizes the classifier performance based on LOOCV. SFS begins evaluating the classifier performance for each single feature. Afterwards SFS selects the best individual feature as the feature that maximizes the classifier performance. Then SFS evaluates the LOOCV consisting of two components. The first of this is the one that maximized the classifier performance in the previous phase and the second is one of the features from the remaining feature subset. This allows for the selection of the best pair of features. Consecutively, SFS evaluates feature subsets with three, four, or more features. As a result, SFS gives
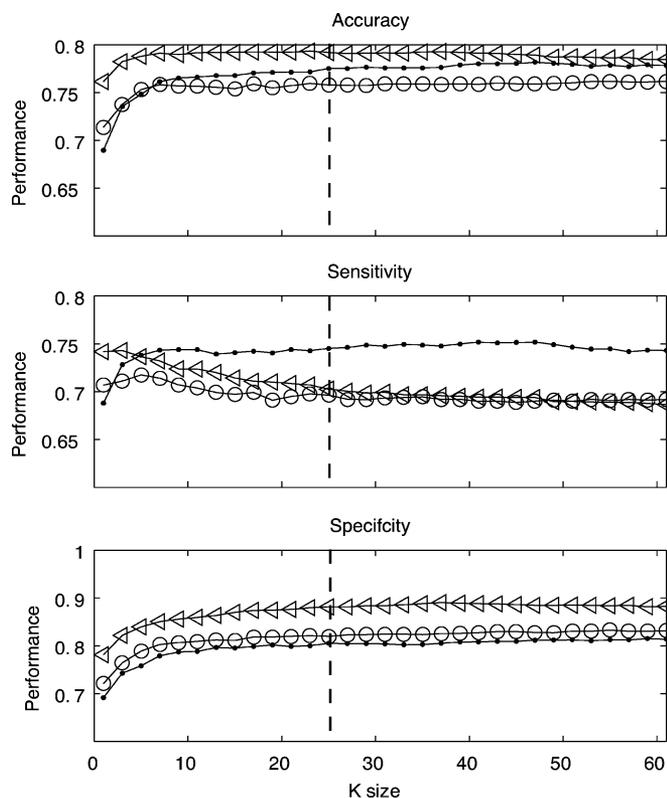
Fig. 3. KNN performance at different Ks and different feature sets selected by statistical significance. From top to bottom, accuracy, specificity and sensitivity. Dot is for one feature, circle for ten features and triangle represents the performance of 30 features. Dashed line shows $K = 25$ that was selected for classification.

the best feature subset consisting of the combination of best features selected at each step.

*2) KNN-Model Selection:* A KNN$(K, F, S)$ model classifier is defined by three variables: $(K)$, which represents the number of samples nearest to the target to be considered for class decision, $(F)$ the feature set, and $(S)$ the training set. In our case $(S)$, consisted of the training set that contained 7127 nonapneic and 4959 apneic minutes. In order to find a proper $K$ to apply SFS, we selected three features subsets with the highest statistical difference between apnea and no apnea classes by using the Wilcoxon test, which is a nonparametric statistical analysis. These subsets were formed, respectively, by the single feature with the highest $p$ value, the ten features with the highest $p$ value, and the 30 features with the highest $p$ value, respectively, according to the Wilcoxon test. For each subset, accuracy, specificity, and sensitivity were evaluated for $K = 1, 3, 5, \ldots, 61$ by LOOCV. Fig. 3 shows the behavior of performance measures for the three subsets at different $K$ values. The KNN performance at a $K$ value of 25 is identified by the dotted line. This $K$ value was selected for our analysis because the performance measures of accuracy, sensitivity, and specificity did not show clear improvements with higher $K$ values.

After selecting the best $K$, it is necessary to define the set of features $F$ that has separation power for apneic and nonapneic events. In order to define the best set with nonredundant information, the SFS previously described was used. The SFS
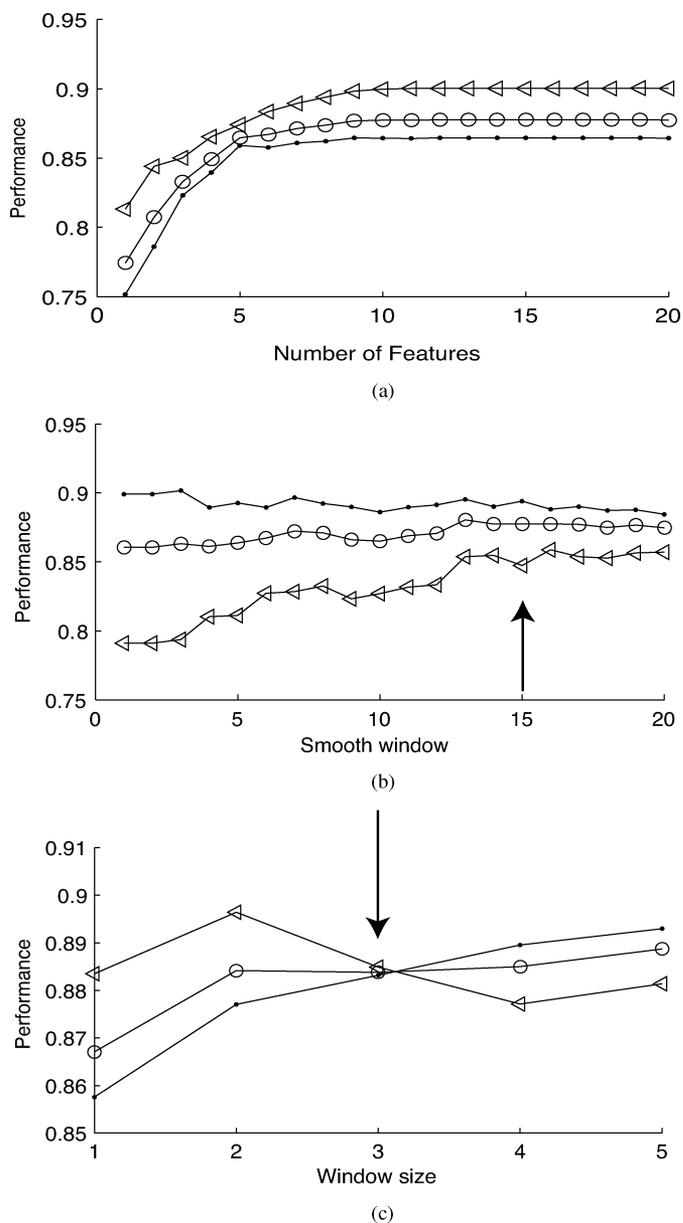


Fig. 4. Performance of the KNN Classifier during (a) feature selection by SFS, (b) with smoothing versions of the feature set in training where x axis represent the smoothing window length, and (c) post-processing window selection. Circle is accuracy, dot is specificity, and triangle represents sensitivity.

is stopped when it finds either a desired number of combination of features or there are no more features to combine. In our case, we stopped the SFS when the feature subset consisted of 20 features. Fig. 4(a) shows the accuracy behavior at each step when using SFS.

### J. Preprocessing and Postprocessing

In order to boost the classification performance after feature selection, the classifiers were trained again with the best feature subset and complemented with their respective smoothed versions. This was done since it is assumed that the smoothed features contain information at different time scales. First for

each classifier, LOOCV was evaluated using its best ten feature subset and complemented with the same best ten features, but smoothed by a moving average filter of window size equal to 2 min. Then, the same procedure was applied but using a moving average filter of window size equal to 3 min. This procedure was repeated with time windows of 3, 4 until 20 min. Fig. 4(b) shows the performance of the KNN classifier during training with the best feature subset of ten features and their respective smoothed versions. The window width selected was 15 min (identified with an arrow) since, with this value, accuracy, specificity, and sensitivity are maximized.

Generally, when dealing with time series, the output of a classifier can be postprocessed to eliminate spurious misclassification events. Spurious events are produced when the values of the features are located in the proximity of the decision boundary or by the presence of outliers. In order to produce an automatic post-processing procedure, a median window running across the classification sequences for each recording was used. This procedure boosted classification performance. Different windows were tested in order to boost the classification and to adapt the classifier to the data. Fig. 4(c) shows the KNN performance on the training set at different size window lengths (size is given in minutes). The selected window was 3 min since accuracy, sensitivity and specificity are maximized at this window length.

*1) NN Model Selection:* A feedforward neural network was selected to be compared with KNN in classifying apnea and non apnea epochs. A back-propagation algorithm was used to adjust the weight of the net during the training process. Neural networks are powerful models prone to overfitting the data; so, we followed a training procedure aimed at reducing this issue. The same training set used in KNN was further divided in three groups: the training group (60% of the training epochs), the validation group (20% of training epochs) and the testing group (20% of the training epochs). The training group has been used by the Levenberg–Marquardt training procedure in order to optimize the weights, while the validation group has been used to perform early stopping. The NN architecture was then selected by an interactive bottom-up procedure for the hidden layer, using classification error in the test group. In principle, the validation error could have been used for network structure optimization, but we aimed at maximum generalization in the training procedure and thus we have decided to use the test set instead.

The features used for the NN training were the best ones found by the KNN algorithm together with the best smoothed versions and post processing median filter selected for the KNN. The input layer dimension was chosen depending on the feature set dimension. The output layer had 1 neuron (linear neural function) and the hidden layer (logarithmic Sigmoid neural function) was varied from 3 to 30 neurons. For each net, the training process was stopped when classification in the validation set began to increase (early stopping procedure). Each net was restarted 20 times and the NN architecture with the lowest classification error in the test group was selected for further analysis. Table I presents the selected architectures (in terms of number of neurons in the hidden layer) for different best feature subsets and the corresponding classification errors.

TABLE I
MEAN SQUARE ERROR OF NN DURING TRAINING

| | | Number of Neurons | | |
|---|---|---|---|---|
| No. Features | Error | Input | Hidden layer | Output |
| 4 | 0.09 | 4 | 13 | 1 |
| 6 | 0.071 | 6 | 17 | 1 |
| 10 | 0.064 | 10 | 16 | 1 |

TABLE II
PERFORMANCE OF KNN AND NN ON THE TESTING SET

| Average of KNN | | | |
|---|---|---|---|
| Features | Accuracy | Sensitivity | Specificity |
| 4 | 0.86 | 0.86 | 0.87 |
| 6 | 0.87 | 0.84 | 0.88 |
| 10 | 0.88 | 0.85 | 0.90 |
| Average of NN | | | |
| Features | Accuracy | Sensitivity | Specificity |
| 4 | 0.87 | 0.87 | 0.85 |
| 6 | 0.87 | 0.88 | 0.84 |
| 10 | 0.88 | 0.89 | 0.86 |

All results using the testing set, 25 recordings for training and 25 recordings for testing.

## III. RESULTS

A training set with 25 recordings was used to develop the KNN and NN classifiers, this set was composed of eight normal, 15 apneic, and two borderline subjects. The training set contained 4950 min annotated as apneic and 7127 min labeled as normal. KNN, NN, and the best set of features with the best smoothing windows in the pre- and post-processing procedure were used to classify apneic and nonapneic minutes in the test set. A different set of 25 recordings formed the test set. This set contained 13 apnea, four borderline, and eight control recordings with 4428 apneic and 7927 nonapneic minutes. The best features selected by KNN using the SFS during the training process were the following features (see Fig. 4(a) for KNN performance) given in order of importance:

1) VLF power of RR intervals normalized to total power;
2) coherence at VLF between RR intervals and QRS area;
3) LF/HF ratio of QRS area;
4) VLF power of module normalized to total power;
5) VLF power of QRS area series normalized to total power;
6) HF power of phase normalized to total power;
7) LF power of phase normalized to total power;
8) coherence at HF between RR intervals and QRS area;
9) LF power of module normalized to total power;
10) LF power of QRS area normalized to total power.

Table II shows the KNN and NN performance when the testing set is analyzed using feature subsets of 4, 6, and 10 features.

One can observe from Table II that the performance of the NN algorithm is very similar to that of the KNN algorithm, even though NN was trained with the features that maximized the KNN performance, thus confirming the generality of the SFS approach with the KNN wrapper. In addition, the performance of both classifiers is up to 88% accurate with respect to correct classification for all different dimensions of the feature set analyzed. Fig. 5(a) shows the automatic classification by KNN of the number of minutes that each subject, in the test set, spent in apnea during sleep, while Fig. 5(b) presents the
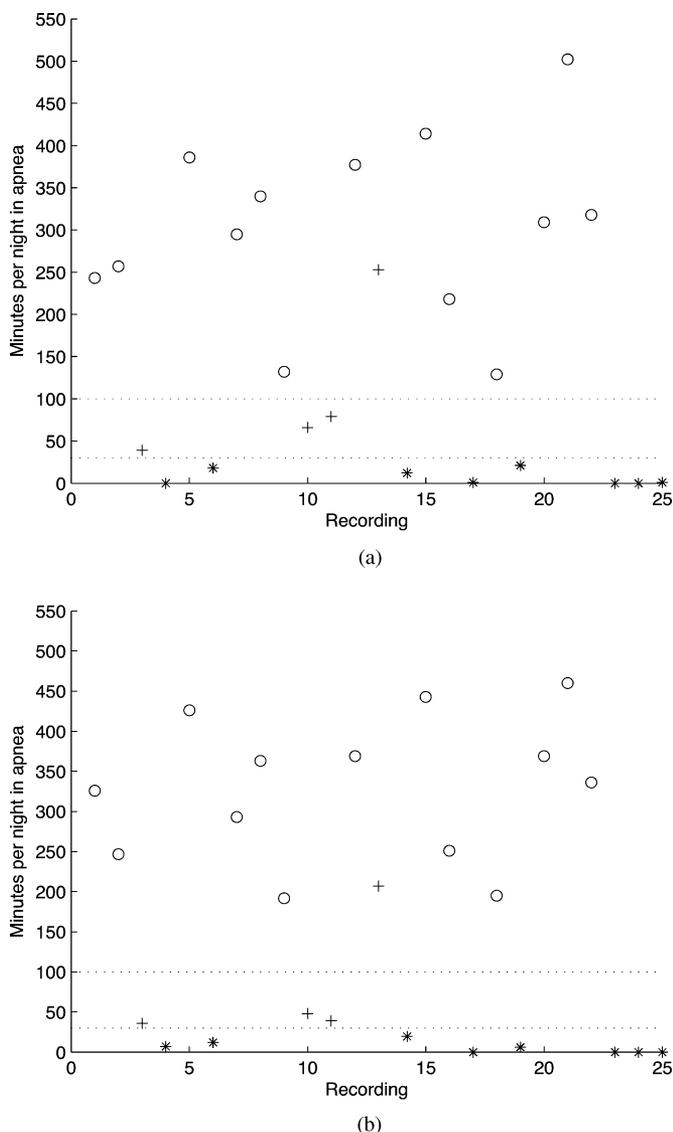
(a)



(b)

Fig. 5. Class separation based on minutes per night calculated by (a) KNN classifier and (b) NN classifier, processing 10 features for 25 recordings of the testing. Note that applying a threshold of 30 min of apnea per night patients and controls classes are separated. To compare with real classification, ($\circ$) represents records with apnea, ($+$) is borderline and ($*$) is for normal records.
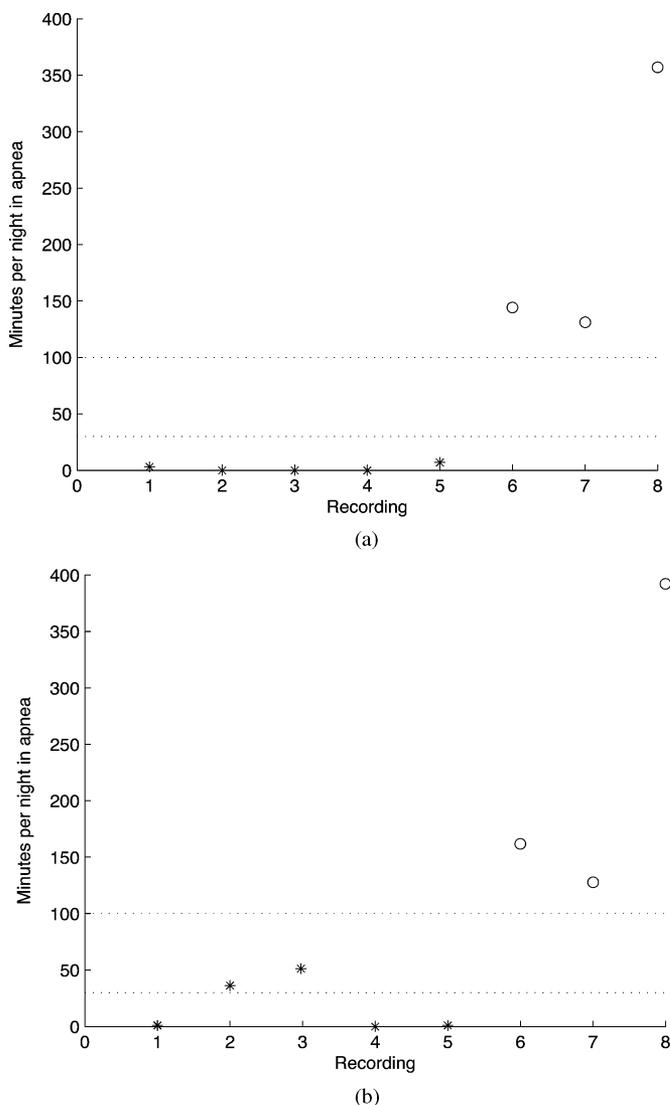


(a)



(b)

Fig. 6. Class separation based on minutes per night calculated by (a) KNN classifier and (b) NN classifier, processing ten features of eight recordings. Note that applying a threshold of 50 min per night apnea and normal classes are completely separated. ($\circ$) represents records with apnea and ($*$) is for normal records.

results obtained with NN. For both KNN and NN classifiers, a threshold of 30 min in apnea per night allows a total separation of the recordings classified as *normal* and the ones classified as affected by *apnea* [see Fig. 5(a) and 5(b)]. Furthermore, most of the borderline recordings had between 30 and 110 apnea minutes per night. The chosen threshold was the one that maximized the class separation.

Finally, both classifiers were tested on eight recordings not taken from the Physionet database. The record classification given by the physician was apneic or nonapneic. KNN and NN classifiers were used in order to define the class. Fig. 6(a) and (b) shows the results for KNN and NN respectively. Note that for KNN the normal recordings remain below the separation line of 30 apneic minutes. Taking this in consideration, one can

note that pathological and normal subjects are separated completely. In contrast, NN overestimates apneic minutes. However, a complete separation between apneic and nonapneic subjects could be achieved using NN with a line separation at 50 apneic minutes.

## IV. DISCUSSION

This paper deals with the automatic classification of obstructive apnea during sleep based on features from the ECG signal. TVAMs were used for feature extraction and KNN and NN were used as classifiers to identify the time spent in apnea during sleep. Our main claims are: 1) high capacity to distinguish between normal and apneic sleep is achieved when sleep apnea is classified using the RR intervals and QRS area parameters coming from the ECG signal; 2) the best feature which can be

used to separate apneic and no apneic conditions is the very low frequency component of RR intervals and QRS area, since this is the component that reflects the apnea repetition frequency; 3) regardless of the classifier, similar class separation is achieved; and 4) TVAMs seem to be valuable tools for the analysis of the variations of the heart rate in normal and pathological conditions during sleep, overcoming the stationary requirement of the signals to be analyzed.

With respect to the selected features, it is worth noting that one of the best performing features is the coherence between the RR intervals and the QRS area. This suggests that respiration is not only related to the respiratory sinus arrhythmia that influences the heart rate, but that it also modifies the ECG morphology. In addition, this feature directly represents the relation between the respiratory nature of the pathology and the synchronization with the heart rate. From a practical point of view, it seems that the selection of features related to QRS area by the SFS could be a shrewd choice, since QRS area can be seen as a rough estimation of the respiration signal [29], [30], which is the one used by clinicians for the identification of apneic episodes. The evaluation of spectral parameters through TVAMs has a very low computational cost and thus the extraction of spectral features is at the same time fast and simple. The high time resolution (beat-by-beat) overcomes the problems of nonstationarity of the signal and it is suitable for developing a real time apnea detection system. In the present research the classification was obtained on a minute-by-minute basis. This was done to allow for a comparison with the annotation in the database. However, the beat-by-beat calculation of the features makes the classification on other time bases possible. Power spectral features extracted by TVAMs represent the most robust features for the evaluation of the apnea condition. This is especially true for the spectral component in very low frequency which reflects the rhythm of apnea repetition. Since our algorithm is based on this component, isolated apneas are difficult to detect. Other physiological and pathological events during sleep, such as cyclic alternating pattern [31], [32] and periodic leg movements [33], could constitute confounding effects for apnea detectors. This could explain the apnea overestimation obtained by the classifiers and the necessity for our selection of a threshold of 30 min instead of the 5 min suggested in Physionet for separating normal from pathological recordings. However, the use of QRS area, strictly related to respiration, makes the method more robust. Furthermore, we have observed that two different classifiers, KNN (nonparametric) and NN (parametric), perform very well for the sleep apnea classification even when using a reduced number of features. This suggests that the features extracted using TVAMs present good discriminatory power independently of the classifier.

In the present paper, respiration was recorded by a nasal thermistor. Nevertheless nasal thermistor is one of the simplest ways to record the nasal airflow, the quantitative description of the dependence of thermistor on room temperature is an open argument. This issue unfortunately is not trivial because the thermistor signal depends on room temperature and the temperature near the nose that may differ from room temperature. As long as room temperature is in a range of industrial bed rooms, such as 15 °C–20 °C the dependence may be small. But if temperature raises above, say 25 °C, the dependency becomes more severe. No systematic studies on this dependency are published in journals so far. The dependency has been reported by several sleep laboratories that tried to make a quantitative correlation between pneumotachographic determined airflow and thermistor in order to replace the heavy pneumotachograph. However, even if nasal airflow presents some problems, it is used by the physicians for patient classification in apneic or nonapneic according to their well-accepted clinical practice and experience. Finally, we would like to remark that no one of the parameters proposed in the present paper comes from the respiratory signal.

During the last years, a series of papers has been published which addresses sleep apnea evaluation from peripheral signals [34]–[36]. These papers applied diverse techniques of signal processing and pattern classification [4]. Some of them reached the same classification performance as we did with our models, in terms of accuracy, sensitivity and specificity, but with a higher number of features used. This suggests that apnea classification based on peripheral signals such as HRV and QRS area provides this level of performance regardless of the approach utilized. The main differences between the current paper and [10] are: 1) the philosophic approach, since we use a gray-box approach with basis in the well-known cardio-respiratory parameters accepted in physiology (normally used for the evaluation of ANS status), while they applied a black box philosophy using tuples of the cardio-respiratory frequencies obtained by Fourier transform as independent features; 2) we propose the use of a time-variant autoregressive model as feature extractor, since this technique is able to follow and adapt itself to the nonstationarieties of the signals overcoming the hypothesis required for the correct application of Fourier transform; and 3) we focused on feature space reduction, since it is always important to avoid course of dimensionality and reduce computation cost. In addition De Chazal *et al.* [37] studied the use of different time resolutions in order to explore possible classification improvements with temporal and spectral features obtained by Fourier transform. Their results showed that the best performance was obtained with 1 min resolution or higher. This could be due to the temporal dynamic of the apnea phenomenon which is in the range of the VLF (corresponding to periods ranging between 20 and 40 s) or due to the intrinsic limitation of the Fourier transform that is not able to recognize very low frequencies in short time windows. It could be interesting to reproduce this study with TVAMs, since it presents a higher time resolution and it does not require the precondition of stationarity.

It is worth noting that the developed algorithms are highly dependent on the R peak detector performance. In this study, we used the well-known Pan–Tompkins algorithm [12]. In literature many different algorithms have been proposed, and a comparison is presented in [38]. The Pan–Tompkins algorithm has been widely used and is one of the most standard algorithms for R peak detection. Pan and Tompkins reported a failed detection percentage of 0.675 % (from 116 137 beats, the results showed 507 false positive beats and 277 false negative beats), which represents a high performance in the current task. However, it is important to stress that such a result is connected to the quality

of the signal, and, especially for long-time recordings during the night, some periods may occur during which the signal is really bad or absent. When the ECG presents a low signal to noise ratio, any R peak detector could fail and then alternatives 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 [10], when consecutive misdetected beats occur (from three to ten misdetected beats). In addition, when more than ten consecutive misdectected 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, since the TVAM is able to evaluate finely the feature characteristics where the signal does exist. However, for future clinical applications, it will be necessary to define the minimal amount of the total night recording needed for a reliable diagnosis.

The algorithms, presented in this work, slightly overestimate apnea. This overestimation could be reduced by the calculation of other features of different nature such as the ones obtained with nonlinear approaches, for instance measurements of time series complexity [39]. In addition, other characteristics coming from the ECG signal such as ST segment and T wave duration as well as other peripheral signals (i.e., photopletysmography (PPG) [41]) could provide information for improving apnea screening. However, it is worth noting that when classifying patients as apneic or nonapneic, the presented method is very accurate. This says, the possible presence of a few false positive apneic periods would not likely create significant problems if this methodology was used as a screening tool in the general population. Finally, the inter-rater variability of visual computer screen based scoring among experts presents an agreement level of around 85% [42]. This suggests that the presented algorithms are inside the confidence interval of agreement among experts. Similar performances levels of the automatic approaches for sleep apnea detection [4], including the ones presented in this paper, confirm the hypothesis that automatic screening is robust and repeatable. This idea reinforces the possible use of automatic screening for supporting the medical decision making and evaluation of sleep apnea in different environments, including home setting. The main advantage, of the algorithms described in the present paper, is their high performance (similar to the performance levels of the algorithms presented in literature) but with a smaller feature space. This makes the whole system particularly suited for home screening devices based on one lead ECG signals. A patient suspected for OSA can be easily identified and directed to a sleep center for standard clinical evaluation.

## V. Conclusion

The bivariate time-varying autoregressive model, proposed in this study to analyze the interrelation of time series during sleep, seems to allow for fast computation, simple implementation, high signal variation adaptability, high resolution, and extraction of features with large classification power between normal and apneic sleep periods. Independently from the classifier we used (i.e., KNN or NN), the results of the classification showed an accuracy of 88%, which is in line with previous studies but with a reduced number of features utilized. The performance of the classification models in this paper suggests the high applicability of TVAMs in possible home health systems for sleep apnea screening in the general population.

## References

[1] W. W. Flemons, D. Buysse, S. Redline, A. Pack, K. P. Strohl, J. Wheatley, T. Young, N. Douglas, P. Levy, W. McNicholas, J. Fleetham, D. White, W. Schmidt-Nowarra, D. Carley, and J. Romanjuk, "Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research: The report of american academy of sleep medicine task force," *Sleep*, vol. 22, no. 5, pp. 667–689, Aug. 1999.

[2] D. P. White, "Sleep Apnea," *Proc. Amer. Thorac. Soc.*, vol. 3, pp. 124–128, 2006.

[3] T. Young, P. E. Peppard, and D. G. Gottlier, "Epidemiology of obstructive sleep ppnea, a population health perspective," *Amer. J. Respir. Crit. Care. Med.*, vol. 165, pp. 1217–1239, 2002.

[4] T. Penzel, J. McNames, A. Murray, P. de Chazal, G. Moody, and B. Raymond, "Systemantic compratition of different algorithms for apnoea detection based on electrocardiogram recordings," *Med. Biol. Eng. Comput.*, vol. 40, no. 4, pp. 402–407, Jul. 2002.

[5] A. J. Camm, M. Malik, J. T. Bigger, Jr., G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy, R. E. Kleiger, F. Lombardi, A. Malliani, A. J. Moss, J. N. Rottman, G. Schmidt, P. J. Schwartz, and D. H. Singer, Task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation, interpretation and clinical use," *Circulation*, vol. 93, pp. 1043–1065, 1996.

[6] A. Malliani, "The pattern of sympathovagal balance explored in the frequency domain," *News Physiol. Sci.*, vol. 14, pp. 111–117, 1999.

[7] M. H. Bonnet and D. L. Arand, "Heart rate variability: Sleep stage, time of night, and arousal influences," *Electroencephalogr. Clin. Neurophysiol.*, vol. 102, no. 5, pp. 390–396, May 1997.

[8] A. L. Goldberger, "Complex Systems," *Proc. Amer. Thorac. Soc.*, vol. 3, pp. 467–472, 2006.

[9] M. O. Mendez, M. Matteucci, V. Castronovo, L. Ferini-Strambi, S. Cerutti, and A. M. Bianchi, "Sleep staging from heart rate variability: Time-varying spectral features and hidden Markov Models," *IJBET*, to be published.

[10] P. de Chazal, C. Heneghan, E. Sheridan, R. Reilly, P. Nolan, and M. O'Malley, "Automated processing of the single lead electrocardiogram for the detection of obstructive sleep Apnoea," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 6, pp. 686–696, Jun. 2003.

[11] R. G. Mark, A. L. Goldbergers, J. H. Peter, T. Penzel, and G. B. Moody, "The apnea-ecg database," *Comput. Cardiol.*, vol. 27, pp. 255–258, 2000.

[12] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.

[13] L. Sörnmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. New York: Academic, 2005.

[14] A. M. Bianchi, L. T. Mainardi, E. Petrucci, M. G. Signorini, and S. Cerutti, "Time-variant power spectrum analysis for the detection of transient episodes in HVR signal," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 2, pp. 136–144, Feb. 1993.

[15] S. Marple, *Digital Spectral Analysis with Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[16] L. T. Mainardi, A. M. Bianchi, R. Furlan, S. Piazza, R. Barbieri, V. di Virgilio, A. Malliani, and S. Cerutti, "Multivariate time-variant identification of cardiovascular variability signals: A beat-to beat spectral parameter estimation in vasovagal syncope," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 10, pp. 978–989, Oct. 1997.

[17] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley, 2001.

[18] T. Mitchell, *Machine Learning*. New York: McGraw-Hill, 1997.

[19] C. M. Bishop, *Neural Networks for Pattern Recognition*. London, U.K.: Oxford Univ. Press, 1995.

[20] S. Haykin, *Neural Networks. A Comprehensive Foundation*. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2009.

[21] K. M. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, pp. 359–366, 1989.

[22] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.

[23] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge, U.K.: Cambridge Univ. Press, 1992.

[24] D. J. Montana and I. Davis, "Training feedforward neural networks using genetic algorithms," in *Proc. 3rd Int. Conf. Genetic Algorithms*, 1989, vol. 3, pp. 762–767.

[25] R. Caruana, S. Lawrence, and C. L. Giles, "Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping," *NIPS*, pp. 402–408, 2000.

[26] J. A. Hirsch and B. Bishop, "Respiratory sinus arrhythmia in humans: How breathing pattern modulates heart rate," *Amer. J. Physiol. Heart Circ. Physiol.*, vol. 241, pp. 620–629, 1981.

[27] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, pp. 273–324, 1997.

[28] G. H. John, R. Kohavi, and K. Pfleger, "Irrelevant feaures and the subset selection problem," in *Proc. Mach. Learning: Proc. 11th Int. Conf.*, 1994, pp. 121–129.

[29] G. B. Moody, T. G. Mark, A. Zoccola, and S. Mantero, "Derivation of respiratory signals from Multi–lead ECG's," *Comput. Cardiol.*, vol. 12, pp. 113–116, 1985.

[30] A. Travaglini, C. Lamberti, J. Debie, and M. Ferri, "Respiratory signal derived from eight-lead ECG," *Comput. Cardiol.*, vol. 25, pp. 65–68, 1998.

[31] L. Ferini-Strambi, A. M. Bianchi, M. Zucconi, A. Oldani, V. Castronovo, and S. Smirne, "The impact of cycle alternating pattern on heart rate variability during sleep in healthy young adults," *Clin. Neurophysiol.*, vol. 111, no. 1, pp. 99–101, 2000.

[32] M. G. Terzano, L. Parrino, R. Chervin, S. Chokroverty, C. Guilleminault, M. Hirshkowitz, M. Mahowald, H. Moldofsky, A. Rosa, R. Thomas, and A. Walters, "Atlas, rules and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep," *Sleep Med.*, vol. 2, pp. 2537–2553, 2001.

[33] E. Sforza, V. Pichot, J. C. Barthelemy, J. Haba-Rubio, and F. Roche, "Cardiovascular variability during periodic leg movements: A spectral analysis approach," *Clin. Neurophysiol.*, vol. 116, pp. 1096–1104, 2005.

[34] M. F. Hilton, R. A. Bates, K. R. Godfrey, M. J. Chappell, and R. M. Cayton, "Evaluation of frequency and time-frequency spectral analysis of heart rate variability as a diagnosis marker of the sleep apnoea syndrome," *Med. Biol. Comput.*, vol. 37, no. 6, pp. 760–769, Nov. 1999.

[35] F. Roche, R. V. Pichot, E. Sforza, I. Court-Fortune, D. Duverney, F. Costes, M. Garet, and J. C. Barthelemy, "Predicting sleep apnoea syndrome from heart rate period: A time-frequency wavelet analysis," *Eur. Resp. J*, vol. 26, no. 6, pp. 937–942, Nov. 2003.

[36] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, pp. 1–39, 2007.

[37] P. de Chazal, T. Penzel, and C. Heneghan, "Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram," *Physiol. Meas.*, vol. 25, no. 4, pp. 967–983, Aug. 2004.

[38] G. M. Friesen, T. C. Jannett, M. A. Jadallah, S. L. Yates, S. R. Quint, and H. T. Nagle, "A comparison of the noise sensitivity of nine QRS detection algorithms," *IEEE Trans. Biomed. Eng.*, vol. 37, no. 1, pp. 85–98, Jan. 1990.

[39] H. M. Al-Angari and A. V. Sahakian, "Use of sample entropy approach to study heart rate variability in obstructive sleep apnea syndrome," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 10, pp. 1900–1904, Oct. 2007.

[40] M. Matteucci and D. Spadoni, "Evolutionary learning of rich neural networks in the Bayesian model selection framework," *Int. J. Appl. Math. Comput. Sci.*, vol. 14, no. 3, pp. 423–440, 2004.

[41] E. Gil, M. O. Mendez, J. M. Vergara, S. Cerutti, A. M. Bianchi, and P. Laguna, "Discrimination of sleep apnea related decreases in the amplitude fluctuations of PPG signal in children by HRV analysis," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1005–1014, Apr. 2009.

[42] T. Penzel and R. Conradt, "Computer based sleep recording and analysis," *Sleep Med. Rev.*, vol. 4, pp. 131–148, 2000.

**Martin O. Mendez** received the Engineer degree in electronics from the Tecnologico de Aguascalientes, Aguascalientes, Mexico, in 2001, the M.Sc. degree in bioengineering from the Universidad Autónoma Metropolitana, in 2003, and the Ph.D. degree from the Department of Bioengineering, Politecnico di Milano, Milano, Italy, in 2007.

He is currently with the Department of Biomedical Engineering, Politecnico di Milano, where he is engaged in the analysis and classification of bioelectrical signals during sleep and related pathologies (such as sleep apnea) using parametric and nonparametric approaches as well as pattern recognition techniques.

**Anna Maria Bianchi** (M'93) received the Laurea degree from the Politecnico di Milano, Milano, Italy, in 1987.

From 1987 to 2000, she was a Research Assistant in the Laboratory of Biomedical Engineering, Istituto di Ricovero e Cura a Carattere Scientifico, S. Raffaele Hospital, Milano, where she was engaged in research in connection with the Department of Biomedical Engineering, Polytechnic University, Milano. Since 2001, she has been a Research Assistant with the Department of Biomedical Engineering, Politecnico di Milano, where she is also an Assistant Professor of fundamentals of electronic bioengineering in the Biomedical Engineering School and of biomedical signal and data processing in the Ph.D. course, and since 2004, she has been on the board of the Ph.D. Program in bioengineering. She is the author and coauthor of more than 50 peer-reviewed international papers. She is the Local Coordinator of a national Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) Project and an European IP Project in the area of biomedical signal processing.

Dr. Bianchi is a Member of the IEEE Engineering in Medicine and Biology Society and a reviewer of many international journals on biomedical engineering.

**Matteo Matteucci** (M'07) received the laurea degree, in 1999, the M.S. degree in knowledge discovery and data mining from Carnegie Mellon University, Pittsburgh, PA, in 2002, and the Ph.D. degree in computer engineering and automation from the Politecnico di Milano, Milano, Italy, in 2003.

He is currently with the Department of Electronics and Information, Politecnico di Milano. His research interest includes learning machines (i.e., neural network, decision trees, mixture models, etc.). He has applied learning methods to different industrial and academic applications, becoming a reference source for this with respect to the local research community. In the complex system modeling, his reaearch has been focused on Bayesian approaches to model adaptation and learning, neural models for biological signals interpretation (e.g., age prediction from heart rate variability, sleep staging, obtrusive sleep apnea recognition, lung cancer diagnosis), and augmented and alternative language models.

**Sergio Cerutti** (M'81–SM'97–F'03) graduation in electronic engineering from the Politecnico di Milano University, Milan, Italy, in 1971.

He is currently a Professor in biomedical signal and data processing with the Department of Biomedical Engineering, Politecnico di Milano, Milano, Italy, where he was the Chairman from 2000 to 2006. His current research interests include biomedical signal processing (ECG, blood pressure signal and respiration, cardiovascular variability signals, EEG, and evoked potentials), cardiovascular modeling, neurosciences, and regulation and standards in medical equipments and devices. Since 1983, he has been going through a course at a graduate level biomedical signal processing at Engineering Faculties, Milan and Rome, as well as at the Specialization Schools of Medical Faculties, Milan and Rome. He is the author and coauthor of more than 400 international scientific contributions (more than 160 on indexed scientific journals).

Prof. Cerutti was a Member of the IEEE Engineering in Medicine and Biology Society (EMBS) AdCom (Region 8) from 1993 to 1996. He is a Fellow Member of the European Alliance for Medical and Biological Engineering and Science (EAMBES) and an Associate Editor of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING. He is a Member of the Steering Committee of the IEEE-EMBS Summer School on Biomedical Signal Processing. He was also the Local Organizer of four Summer Schools held in Siena.

**Thomas Penzel** (M'92–SM'06) was born in Hamburg, Germany, in 1958. He received the graduate degree in theoretical physics, in 1986, and the doctorate degree in human biology, in 1991, and in the habilitation in physiology, in 1995.

He was a Professor with the Medical Faculty, University of Marburg, Marburg, Germany, in 2001. He was with the Sleep Laboratory, University of Marburg, in 1982. In 2006 Professor Penzel became Research Director of the Sleep Medicine Center at the Charité University Hospital, Berlin. His research interests include biosignal analysis of electroencephalogram, electrocardiogram, respiration, and telemedicine applications.

Dr. Penzel was a Member of the Board of the German Sleep Society during 1993–2001. Since 2001, he is the President of the International Society on Biotelemetry. In 2001, he received the Bial Award for achievements in Neurotelemedicine in Portugal and in 2008, he received the Bill Gruen Award for Innovations in Sleep Research by the American Sleep Research Society.