

Detection of Sleep Apnea from surface ECG based on features extracted by an Autoregressive Model

Martin O. Mendez, Davide D. Ruini, Omar P. Villantieri, Matteo Matteucci *IEEE Member*,
Thomas Penzel, Sergio Cerutti *IEEE Member*, Anna M. Bianchi *IEEE Member*

Abstract—This study proposes an alternative evaluation of Obstructive Sleep Apnea (OSA) based on ECG signal during sleep time. OSA is a common sleep disorder produced by repetitive occlusions in the upper airways. This respiratory disturbance produces a specific pattern on ECG. Extraction of ECG characteristics, as Heart Rate Variability (HRV) and peak R area, offers alternative measures for a sleep apnea pre-diagnosis. 50 recordings coming from the apnea *Physionet* database were used in the analysis, this database is part of the 70 recordings used for the Computer in Cardiology challenge celebrated in 2000. A bivariate autoregressive model was used to evaluate beat-by-beat power spectral density of HRV and R peak area. K-Nearest Neighbor (KNN) supervised learning classifier was employed for categorizing apnea events from normal ones, on a minute-by-minute basis for each recording. Data were split into two sets, training and testing set, each one with 25 recordings. The classification results showed an accuracy higher than 85% in both training and testing. In addition it was possible to separate completely between Apnea and Normal subjects and almost completely among Apnea, Normal and Borderline subjects.

I. INTRODUCTION

Sleep apnea (SA) is a common sleep disorder characterized by repetitive cessation of breathing during sleep time. In clinics, SA is classified in three classes, obstructive, central and mixed. Obstructive sleep apnea is the very common in the population with a prevalence between 2 and 4 % and can be related to aging and obesity. Obstructive apnea consists in an interruption in the airflow to the lungs caused by an occlusion at the upper airways level. Independently of the type, an apnea event is typically accompanied by a reduction in the blood oxygen and arousal events that restore respiration. An apneic event generates well defined

This work is supported by the MyHeart project IST 507816 of the European Community and by CONACYT

M.O. Mendez is with Dept. of Biomedical Engineering, Politecnico di Milano, Pzza. Leonardo da Vinci 32, Italy
martin.mendez@polimi.it.

D.M. Ruini is with Dept. of Electronics and Information, Politecnico di Milano, Pzza. Leonardo da Vinci 32, Italy
ruini.davide@gmail.com.

O.P. Villantieri is with Dept. of Biomedical Engineering, Politecnico di Milano, Pzza. Leonardo da Vinci 32, Italy
omar.villantieri@polimi.it.

A.M. Bianchi is with Dept. of Biomedical Engineering, Politecnico di Milano, Pzza. Leonardo da Vinci 32, Italy
annamaria.bianchi@polimi.it.

M. Matteucci is with Dept. of Electronics and Information, Politecnico di Milano, Pzza. Leonardo da Vinci 32, Italy
matteucc@elet.polimi.it.

T. Penzel is with Das Schlafmedizinische Zentrum der Charit, Berlin, Germany
thomas.penzel@charite.de.

S. Cerutti is with Dept. of Biomedical Engineering, Politecnico di Milano, Pzza. Leonardo da Vinci 32, Italy
sergio.cerutti@polimi.it.

oscillations of tachycardia-bradycardia in the heart rate. During obstructive sleep apnea (OSA), respiratory muscles do not stop working, and mechanical efforts are generated in order to overcome the occlusion. If these efforts are not sufficient, oxygen in the blood begins to decrease, muscle efforts increase by the hypoxia, until an arousal takes place to reactivate all the systems and restore the respiration [1]. Severe sleep apnea could have hundreds of respiratory events during sleep with serious health and social consequences ranging from daytime sleepiness up to heart failure [2]. Sleep apnea diagnosis is expensive since requires dedicated personnel, infrastructure and special systems. Diagnosis requires different signals that are generally obtained by the polysomnography procedure and at least one night must be spent in the specialized sleep center. Due to the limited number of sleep centers and specialized personal, diagnosing sleep apnea becomes no accessible to the population and this pathology is underestimated. These difficulties generate the necessity of investing efforts in order to obtain a simpler and reliable sleep apnea diagnosis. It could be achieved by assessing the sleep apnea by the projections onto peripheral systems of easier measurement such as heart rate and pulse oximetry. In the past decade, some researches have focused on this concept, and many studies have been presented using a variety of signal processing and pattern recognition techniques. Some studies used features extracted from the ECG as RR intervals, R amplitude, T duration, R area and peripheral tonometry [3] [4]. Results obtained by those studies showed good levels of classification between normal respiration time and apneic time. Important results were obtained during a competition in year 2000 conducted by the organizers of Computer in Cardiology conference. This competition was focused on an apnea screening based on ECG [3]. Motivated by the results of the competition and following very close the criteria established, we decided to test a time-variant feature extraction algorithm and some classifiers to detect sleep apnea. The aim of this study is dedicated to evaluate a reliable sleep apnea diagnosis tool based only on measures of superficial electrical activity of the heart in a beat by beat fashion.

II. MATERIAL AND METHODS

A. Database Description

The database used for this research is free and available on the Physionet web-site (www.physionet.org). This database was used during the annual *Computers in Cardiology Challenge* in 2000 [3]. It consists of 70 ECG whole night

recordings with a duration close to eight hours. Data were collected by the sleep laboratory at the Philipps University in Marburg, Germany. Standard ECG recordings were acquired with a sampling frequency of 100 Hz, and 16 bits resolution. Recordings were separated in three groups: Apnea group (class A), Borderline group (class B) and Normal group (class C). Apnea scoring was carried out based on standard criteria by sleep expert personal with a minute-by-minute resolution. A minute was defined as apnea if at least one apnea or hypo-apnea episode happens. Otherwise, that minute was defined as normal breathing. This procedure was used for the total sleep time for each subject. The database is divided into two groups each containing 35 subjects: release-group for training and withheld-group for testing. Each group contains 20 apnea, 5 borderline and 10 normal subjects. From the release-group we selected 25 subjects randomly to train our classification algorithm. A second group with 25 recordings chosen randomly from the withheld-group, was used to measure the performance of our algorithm. We selected only 50 recordings for evaluating the feasibility of the algorithm to classify correctly apnea and normal conditions. Future studies will include the whole database.

B. Methods

Based on previous knowledge about physiological effects of OSA on ECG signal, we derived signals as R area [4] and RR intervals [3]; then we extracted a set of features that were used for our pattern recognition algorithm. RR intervals show characteristic oscillations (brady-tachycardia) during an apnea event, this pattern produces a very low frequency in the signal spectrum that could help to find an apnea. Time and spectral parameters extracted from RR intervals and R area by a time-varying autoregressive model are used as input for the classifier. The non parametric supervised classifier *K-Nearest Neighbor* was used for this study.

C. RR intervals correction and ECG-Derived Respiratory signal (EDR)

The database has the QRS complex points for each recording. We used an automatic algorithm to search the R peaks into ECG signal corresponding to each QRS point. Resulting series were plotted with the respective ECG for correcting by hand misdetected beats. After correction, the RR time series were computed. The baseline was subtracted from the original ECG to obtain an estimation of the ECG Derived Respiratory signal (EDR). Baseline was calculated by a median filter of 200 ms width. From the resulting ECG, interval between the minimum value 100 ms before and after to the maximum R peak value was searched. Then, the area was calculated inside the region enclosed between those two points. Methods for the calculation of EDR are in [6].

D. Autoregressive Models

When the signal to be analyzed is nonstationary as the HRV during sleep time, it is necessary to develop a filter able to adapt its parameters at each sample. An adaptive filter self-adjusts its transfer function according to the prediction error.

Self adjustment involves a cost function which determines the algorithm performance, the objective is to minimize this cost function (it is minimized in the sense of least square) at each time n as:

$$CF_P(n) = \sum_{k=1}^n \lambda^{n-k} |e(n)|^2 \quad (1)$$

where the observation interval is k and λ is the forgetting factor and $1/(1 - \lambda)$ represents the memory of the algorithm. λ weights the vector of prediction error giving more importance to the recent error, λ^{n-k} is always less than one. Self adjustment and forgetting factor are desirable when we work with non stationary sequences. When we deal with non stationary signals RLS is the most adequate algorithm to update the filter parameters [5].

E. Features Sets

Preprocessing (RR detection and ECG Derived Respiratory signal evaluation) gave as result two time series with physical and physiological information about the Autonomic Nervous System (ANS) and the respiratory system. Based on those time series, it is possible to extract characteristics that could be of physiological and clinical interest. These characteristics are potential features to be considered for classification. The Power Spectral Density (PSD) for each RR interval and DRS series, and the interrelation between them were obtained by a bivariate time-varying autoregressive model at each beat. The total spectrum at each beat was separated in the classical spectral indexes of HRV for the resulting PSD, VLF = 0.003-0.04 Hz, LF = 0.04 - 0.15 Hz, HF = 0.15-0.5 Hz. Further, coherence was evaluated in order to obtain the frequency interrelation between both time series at each frequency band. When we deal with physiological series it is important to normalize the time series in order to eliminate the inter-variability between subjects, which is produced by the personal physiological limits and conditions. In this way, some inconsistencies are eliminated and it is assured that noise caused by the diversity of subjects is reduced or damped. We applied two separate normalization procedures to the data. Time features coming from RR intervals, and DRS were normalized to zero mean and unit standard deviation. HF, LF and VLF were normalized in two different ways: zero mean and unit standard deviation plus power percent at each beat. For Coherence no normalization was applied since these value ranged between 0 and 1. A total of 52 features were extracted.

F. K-Nearest Neighbor(KNN) Classifier

K-Nearest Neighbor (KNN) is a nonparametric classification technique that attempts to divide the data using their natural distribution [6]. The KNN algorithm is based on nearest-neighbor rule which classifies the input by assigning it the label most frequently represented among the k nearest samples in the feature space. There are different measures of distance for KNN such as Minkowsky, Euclidean or Tanimoto, we decided to use the Euclidean Distance for

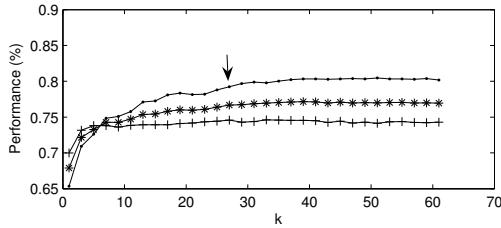


Fig. 1. Classifier performance with one feature at different K 's for one parameter, dot is specificity, * represents accuracy and + is sensitivity. Arrow shows K selected for classification

its simplicity and good classification results. For each test minute sample, the Euclidean distance was measured as:

$$D_j(a,b) = \sqrt{\sum_{n=1}^d (a_n - b_{nj})^2}$$

where d is their number of features, a_n is the vector containing the features of the sample we ant to classify, b_{nj} is vector containing the features of the j -th sample in the training data set and $D_j(a,b)$ is the Euclidean Distance in the d -dimensional feature space between the test sample and the j -th element in the training set. Then to assign a label at the sample we calculate a majority voting between the k nearest points. This process is repeated for all the minutes in a test recording. Details on the KNN classification method are in [7].

G. Selection and Transformation of the Features

To prevent the *course of dimensionality* in estimating the *a posteriori* distribution for the classification performed, feature selection can be addressed in different ways, it can be evaluated by statistical analysis of features or by WRAP methods. WRAP methods consist in selecting the features based on the classifier performance for each group of them. The WRAP method was applied in this work to obtain the best group of features for classification using KNN accuracy as evaluation measure. We followed the next procedure to obtain the best group of features and the best K . First, K was set to 15. Then, the WRAP was then used to find the best parameters for classifying apnea events. Thereafter, we selected the first feature given by the WRAP method and an interactive algorithm was created in order to obtain the classifier performance at different K 's. Fig 1 shows the curves of performance at each K . We observe how the performance of the classifier rises until $K=27$. Then $K = 27$ was chosen. Finally, WRAP method was used as search algorithm for selecting the final best group of features. Figure 2 shows the results obtained by the WRAP method in selecting the best feature group for apnea classification.

H. Post-Processing

Generally, when dealing with time series, the output of a classifier can be post-processed 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 just by outliers. In order to

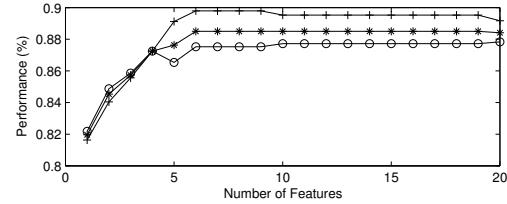


Fig. 2. Classifier performance during feature selection , o is specificity, * represents accuracy and + is specificity.

produce an automatic post processing procedure, a median window running across the classification sequences for each recording was used. This procedure booster our classification performance. Different windows were tested in order to boost the classification and to adapt our classifier to data. A mean window of 7 minutes gave the best results.

I. Classification performance estimation

Classifier performance has been estimated in two ways. First, Leave-One-Out cross-validation technique was used to evaluate classifier performance in the release-group. Second, we classified all recordings of the withheld-group and the overall and individual statistical measures of classification accuracy, sensitivity and specificity were calculated. Specificity gives idea about the percentage of normal events correctly classified. Sensitivity represents the percentage of apnea events classified and accuracy is a measure of the total events correctly classified both normal and pathologic.

III. RESULTS

Out of the 25 subjects for training procedure, we took ten normal, ten apnea subjects and 2 borderline for training. Thereafter we selected randomly 2343 minutes of normal respiration and 2023 minutes with apnea. In this way, we assured a balanced database for training our classifier. The best set of features with the best smoothing window were used to classify apnea and no apnea events on the training set. The best features in order of importance were:

- 1) Percentage of very low frequency in RR intervals
- 2) Coherence in very low frequency
- 3) Very low frequency derived from the RR intervals and Area of the R peak
- 4) Variance of low frequency in RR intervals.

Power spectral components made up the majority of our features. Table I shows the performing classification on the release-set (leave-one-out cross-validation) and on the withheld-set.

Finally, we measured the number of minutes that each subject spent in apnea during sleep time for the withheld set. Figure 3 shows the results obtained automatically by the classifier. A threshold of 50 apneas per night or approximately 6 min/h allows a total separation of the subjects classified as *normal* and those classified as *apnea*. Furthermore, most of the borderline subjects were enclosed between 50 and 110 apneas per night.

TABLE I
CLASSIFICATION PERFORMANCE ON THE RELEASE AND WITHHELD SETS
FOR A KNN CLASSIFIER

	Average of training set results		
Features	Accuracy	Sensitivity	Specificity
4			
4	0.8636	0.8934	0.8337

	Average of testing set results		
Features	Accuracy	Sensitivity	Specificity
4	0.8555	0.8390	0.8850

All results using the release and withheld sets, 25 subjects for creating the balance training set and 25 subjects were used for testing the classifier

