Computers in Biology and Medicine 91 (2017) 47–58

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Heart rate variability feature selection in the presence of sleep apnea: An expert system for the characterization and detection of the disorder

Sofía Martín-González [a,*], Juan L. Navarro-Mesa [a], Gabriel Juliá-Serdá [b], Jan F. Kraemer [c], Niels Wessel [c], Antonio G. Ravelo-García [a]

[a] Institute for Technological Development and Innovation in Communications, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria 35017, Spain
[b] Pulmonary Medicine Department, Hospital Universitario de Gran Canaria Dr. Negrín, Las Palmas de Gran Canaria, 35010, Spain
[c] Department of Physics, Humboldt-Universität zu Berlin, Berlin 10115, Germany

ARTICLE INFO

Keywords:
Sleep apnea
Single-lead ECG
Heart rate variability
Cepstrum
Filter bank
Detrended Fluctuation Analysis
Feature selection

ABSTRACT

We introduce a sleep apnea characterization and classification approach based on a Heart Rate Variability (HRV) feature selection process, thus focusing on the characterization of the underlying process from a cardiac rate point of view. Therefore, we introduce linear and nonlinear variables, namely Cepstrum Coefficients (CC), Filterbanks (Fbank) and Detrended Fluctuation Analysis (DFA). Logistic Regression, Linear Discriminant Analysis and Quadratic Discriminant Analysis were used for classification purposes.

The experiments were carried out using two databases. We achieved a per-segment accuracy of 84.76% (sensitivity = 81.45%, specificity = 86.82%, AUC = 0.92) in the Apnea-ECG Physionet database, whereas in the HuGCDN2014 database, provided by the Dr. Negrín University Hospital (Las Palmas de Gran Canaria, Spain), the best results were: accuracy = 81.96%, sensitivity = 70.95%, specificity = 85.47%, AUC = 0.87. The former results were comparable or better than those obtained by other methods for the same database in the recent literature.

We have concluded that the selected features that best characterize the underlying process are common to both databases. This supports the fact that the conclusions reached are potentially generalizable. The best results were obtained when the three kinds of features were jointly used. Another notable fact is the small number of features needed to describe the phenomenon. Results suggest that the two first Fbanks, the first CC and the first DFA coefficient are the variables that best describe the RR pattern in OSA and, therefore, are especially relevant to extract discriminative information for apnea screening purposes.

1. Introduction

The Sleep Apnea-Hypopnea Syndrome (SAHS) is a widespread sleep respiratory disorder characterized by repetitive breathing pauses during sleep. There are three types of sleep apnea: obstructive, central, and mixed, depending on the breathing efforts during the apnea events. In central sleep apnea (CSA), the respiratory drive is absent or inhibited, and the upper airway is open. However, in obstructive sleep apnea (OSA), the upper airway collapses and the airflow is obstructed, but the respiratory effort continues. In mixed sleep apnea both types are present (CSA and OSA). CSA is often seen in OSA patients, but pure CSA is unusual.

OSA is the most common in the general population with a prevalence of 14% in men and 5% in women (USA) [1]. During an OSA episode, the respiratory muscles try to overcome the obstruction of the upper airway. If the efforts are unsuccessful, the blood oxygen level decreases and consequently, muscle effort increases until an arousal takes place to reestablish normal breathing. In the last few years, many studies [2,3] have shown its major health implications, ranging from daytime drowsiness to important cardiac and vascular diseases that lead to increased mortality rates. The criterion [4] used to decide whether a patient suffers from OSA is the mean number of breathing pauses per hour (Apnea-Hypopnea Index, AHI). To distinguish between apnea and hypopnea it is necessary to know whether the cessation of breathing is complete or partial, respectively. An AHI lower than 5 is regarded as normal, an AHI ranging [5,15], as mild sleep apnea, an AHI ranging [15, 30] as moderate sleep apnea, and an AHI greater than 30, as severe sleep apnea.

The gold standard for OSA diagnosis is polysomnography. This method is expensive, time-consuming and unpleasant for the patients,

* Corresponding author.
E-mail address: sofia.martin@ulpgc.es (S. Martín-González).

https://doi.org/10.1016/j.compbiomed.2017.10.004
Received 23 May 2017; Received in revised form 6 October 2017; Accepted 6 October 2017

due to the amount of physiological signals registered throughout the whole night [5,6]. These drawbacks hinder large-scale population studies. That is why a major effort is being made to create automatic sleep apnea screening methods based on a smaller number of physiological signals and portable systems. The ECG signal has turned out to be especially interesting for screening purposes as it can be easily recorded with wearable devices, in particular the single-lead ECG signal.

Most of the methods proposed in the literature for sleep apnea detection using the ECG, introduce features derived from the heart rate variability (HRV) and the ECG derived respiratory signal (EDR) [7–18].

The normal HRV is based on the autonomic neural regulation of the heart and the circulatory system, and changes in the HRV mirror the effects of the physiological factors modulating the normal heart rhythm [19]. To analyze this HRV, it is often necessary to use signals derived from the original ones, as they are more robust and reliable. This is the case of the RR signal. It is derived from the electrocardiogram and is used to analyze HRV. It is constructed by measuring the delay between two consecutive R-peaks of the electrocardiogram. The sequence of consecutive delays forms the RR series.

The reason why heart rate information is valuable for OSA detection is that sleep apnea affects HRV [19–22]. There are periodic variations in the RR series due to breathing phases (cardio-acceleration during inhalation and cardio-deceleration during exhalation) in a healthy person at rest that are called respiratory sinus arrhythmia (RSA). This periodic information can be observed around 0.25 cycles/beat. However, the way sleep apnea affects HRV is called 'cyclical variation of heart rate' (CVHR) [20]. When apnea occurs, the heart rate tends to decrease and normally rises once the apnea has ended. This pattern of brady- and tachycardia has been attributed to a parasympathetic control of heart rate during sleep, interrupted by sympathetic activation that ends with an arousal [19–21]. These periodic variations of the heart rate (HR) produce frequency components around 0.02 cycles/beat because of the apnea repetition [22]. Other authors, like Acharya et al., indicate a range for this spectral information between 0.01 and 0.05 cycles/beat [19].

Therefore, it is necessary to find feature extraction techniques that enable us to obtain information about the periodicities taking place in the RR series. Linear analysis is a valid candidate for this purpose. It allows us to interpret the regular structure of the associated cardiovascular system, such as the frequency contents. Particularly, the use of power spectrum analysis can help us to understand the relationship between HRV and the neurocardiological system.

On the other hand, it is also known that the cardiovascular system and, by extension, the HRV, is considered dynamic, nonlinear, and nonstationary [23–25]. Hence, there are other aspects, such as the complex autonomic and respiratory control mechanisms that interact in the regulation of the cardiac function, that are better analyzed by nonlinear methods [23,25].

Several studies over the last 20 years have tried to determine the presence of OSA using features extracted from single-lead ECGs. Shinar et al. [8] studied the QRS complex changes and the spectral abnormalities in the HRV caused by apneas. De Chazal et al. [9] introduced a wide variety of time and spectral features based on heartbeat intervals and EDR. Mendez et al. [10] used a bivariate autoregressive model to evaluate beat-by-beat power spectral density of HRV and R-peak area. Khandoker et al. [11] employed wavelet based features of HRV and EDR as inputs to a support vector machine (SVM). Mendez et al. [12] investigated the possibility of using the empirical mode decomposition (EMD), comparing the results with those obtained through wavelet analysis (WA). Le et al. [13] employed SVM to determine the sleep apnea events based on the power spectral density (PSD) and Recurrence Quantification Analysis (RQA) features of the RR intervals. Nguyen et al. [14] applied RQA to HRV data, and for classification purposes they used SVM, neural network (NN), and a soft decision fusion rule to combine the results of these classifiers. Varon et al. [15] employed the standard deviation and the serial correlation coefficients of the RR interval time series, the principal components of the QRS complexes, and a feature that extracts the information shared between respiration and heart rate using orthogonal subspace projections. Gutiérrez et al. [16] assessed the ability of spectral entropy (SE) and multiscale entropy (MsE) to characterize SAHS in HRV recordings and evaluated the differences in the analysis depending on the gender. Song et al. [17] proposed an OSA detection approach based on ECG signals by considering temporal dependence within segmented signals using a discriminative hidden Markov model (HMM). Sharma et al. [18] approximated each QRS complex of the ECG signal using a linear combination of the lower order Hermite basis functions and the coefficients of the Hermite expansion were used for classification purposes.

Despite the fact that a lot of work has been done using the analysis of HRV in the context of apnea detection, until now there is no defined model that describes the complex dynamics in the cardiovascular system when apnea occurs. That is why we consider there is a margin for further studies to reach a better understanding of the underlying process and to explore new features which extract as much information as possible from the ECG.

In previous contributions [2,26–29], we introduced features, like Cepstrum Coefficients (CC) and Filterbanks (Fbank), to extract information related to frequency contents, and Detrended Fluctuation Analysis (DFA), to uncover nonlinear characteristics of the physiological process associated with sleep apnea. In this article we want to go a step further in the characterization of the OSA phenomenon, carrying out a systematic study combining these features (CC, Fbank and DFA), that, to the best of our knowledge, has never been done by other authors. This study allows us to discover the relative importance of each of the extracted features, as the main goal is the characterization of the underlying system from a cardiac rate point of view. This would allow us to enter this information into expert systems and, ultimately, contribute to describing the mechanisms regulating sleep apnea.

We also use a forward feature selection technique to detect the features that best discriminate between apneic and nonapneic minutes. Finally, we enter them into three different classifiers. In particular, we propose three statistical classification methods: Logistic Regression (LR), Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA). All the possible combinations of feature extraction and classification strategies allow us to pinpoint the one that performs best in the quantification process, thus revealing the features with the highest discriminant capacity for apnea screening purposes in a multivariate analysis.

In this work, we classify in two ways: per-segment and per-recording. In the per-segment classification, also called quantification, the binary classifier decides whether the evaluated minute shows apnea or not. In the per-recording classification, a global score is defined for each subject to establish whether it is OSA-diagnosed or healthy, based on the results obtained in the quantification.

In the experiments we use two databases to give the results a more generalizable character, so that conclusions would not be limited by database variability.

2. Methods

2.1. Databases

Two databases have been used to carry out all the experiments: the Apnea-ECG Physionet database [30], hereinafter referred to as Physionet database, provided by Prof. Dr. Thomas Penzel for Computers in Cardiology Challenge 2000, and the HuGCDN2014 database, provided by the Dr. Negrín University Hospital (Las Palmas de Gran Canaria, Spain). Neither of them discriminates between apnea and hypopnea, all are considered apneas.

Physionet is a widely used database, made up of 70 electrocardiogram recordings, each containing a single ECG signal and its corresponding score, which indicates the presence or absence of apnea every minute. It is digitized at 100 Hz with 12-bit resolution, and lasts about 8 h (from

401 to 578 min). The labelling process was done by a human expert based on other respiratory signals recorded simultaneously.

The recordings are divided into two equally sized sets, a learning set (L) and a test set (T). Based on the amount of minutes with apnea of each subject, the database is divided into 3 groups: 1) GROUP A: includes the recordings with at least 100 min with apnea. L-set and T-set both contain 20 recordings of this group. 2) GROUP B: includes recordings with a number of apnea minutes between 5 and 99. L-set and T-set both contain 5 recordings of this group. 3) GROUP C: the recordings included in this group have fewer than 5 min with apnea. L-set and T-set both contain 10 recordings of this group.

The other database used in the experiments is HuGCDN2014. It was provided by the sleep unit of the Dr. Negrín University Hospital and contains recordings of 77 subjects, each containing the ECG signal, digitized at 200 Hz. Every minute was labeled by an expert based on the simultaneous polysomnography, indicating the presence or absence of apnea. There are two groups: 1°) CONTROL: Forty healthy subjects with an AHI lower than 5.2°) APNEA: Thirty-seven OSA patients with an AHI higher than 25. The learning set (L) is made up of the recordings of 20 control subjects and 18 OSA patients. The rest belong to the test set (T).

2.2. Preprocessing of the signal

The Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology defined, in order to standardize physiological and clinical studies, the guidelines for HRV analysis [31]. They suggest using short-term recordings of 5 min taken under physiologically stable conditions. In particular, for very low frequencies (VLF) components, like the CVHR, they recommend avoiding segments of less than 5 min. So, the single ECG signal is divided into 5-min frames, that are shifted in time in increments of 1 min. This frame length allows us to balance stationarity and good spectral resolution. The sliding step of 1 min is suitable to identify the CVHR which ranges from 20 to 60 s [13]. The quantification obtained for each segment is assigned to the central minute.

During the preprocessing each frame is band-pass filtered between 20 and 35 Hz and then it is full wave rectified and low-pass filtered to 12.5 Hz. An interval of values is set around a relative maximum of the output signal, locating the R-peak as the maximum value of the ECG signal within that range. The RR interval series is constructed as a series of time differences between the successive heart beats.

Once RR series are obtained, an adaptive filtering procedure for automatic artifact removal is applied [32]. It is based on the interval filter described by Wessel et al. [33]. The filtering algorithm consists of three sub-procedures: (i) the removal of obvious recognition errors, (ii) the adaptive percent-filter, and (iii) the adaptive controlling filter. This is necessary, as artefacts and ectopic values often corrupt the analysis of HRV. The advantage of this method is the spontaneous adjustment of the system coefficients to sudden changes in the series.

2.3. Features obtained from RR series

After this processing, we extract the features of each frame following different strategies. The goal is to compare the results obtained when the quantification of sleep apnea is performed by Cepstrum Coefficients (CC), Filter Banks (Fbank), Detrended Fluctuation Analysis (DFA) or a combination of the three. The main concepts of the different feature extraction techniques used in the article are briefly explained below.

2.3.1. Cepstrum coefficients

Real cepstrum coefficients have been obtained from the original cepstrum equation (equation (1)). It is the result of taking the inverse Fourier transform of the logarithm of the magnitude of the RR spectrum. The real cepstrum uses only the information of the magnitude of the spectrum and makes it possible to obtain information about the spectrum envelope and the harmonic components.

$$c(\tau) = real\left(F^{-1}(log(|F(x(n))|))\right) \quad (1)$$

In the experiments, once the cepstrum coefficients are obtained from the RR intervals, we take the first 20 elements (Fig. 1), as they seem to contain enough relevant information of the underlying system [28]. As Fig. 1 shows, the dynamic range of the first CC is much greater than in the other CC's. To avoid it, we use standardization on the data set and transform it to have zero mean and unit variance.

2.3.2. Filter banks

A Filter Bank is an example of a frequency-domain analysis of a signal. First, it is necessary to obtain the periodogram (equation (2)), that is a good estimator of the power spectral density of the signal.

$$\widehat{P}(k) = \frac{1}{N}\left|X_N(k)\right|^2 \quad (2)$$

$X_N(k)$ is the Discrete Fourier Transform (DFT).

After estimating the periodogram, the signal is introduced into M equally-spaced filters. The outcomes of the filters are an estimation of the power in each frequency band, which cover a range defined by a lower limit [(0.5/34)*(nº Fbank-1)] and an upper limit [(0.5/34)*(nº Fbank)]. That is, we obtain as many parameters as filters are in the bank. In our experiments, we decided to work with 34 filters, as we consider it is sufficient to analyze the very low frequency (VLF, 0.003–0.04 Hz), the low frequency (LF, 0.04–0.15 Hz) and the high frequency (HF, 0.15–0.4 Hz), with enough resolution. In this way, we are able to study the spectral information related to CVHR, and the components related to the RSA [29].

2.3.3. Detrended Fluctuation Analysis (DFA)

DFA is a useful technique to study signals or derived time series that appear to be long-memory processes. It is a scaling analysis method that provides a simple quantitative parameter to represent the autocorrelation properties of a signal [34]. Peng et al. first introduced it in 1995 [35]. An important consideration is that it can be applied to non-stationary signals so that it can be useful in non-stationary segments of signals. Another important characteristic of DFA is the capacity of avoiding spurious detection of long-range correlations in the heartbeat time series that are an artifact of non-stationarity [21]. Due to these properties, we think that this variable can be useful to characterize the physiological process associated with OSA and to detect breathing pauses. In previous studies, several authors have introduced DFA in the analysis of biomedical signals [21,34–39].

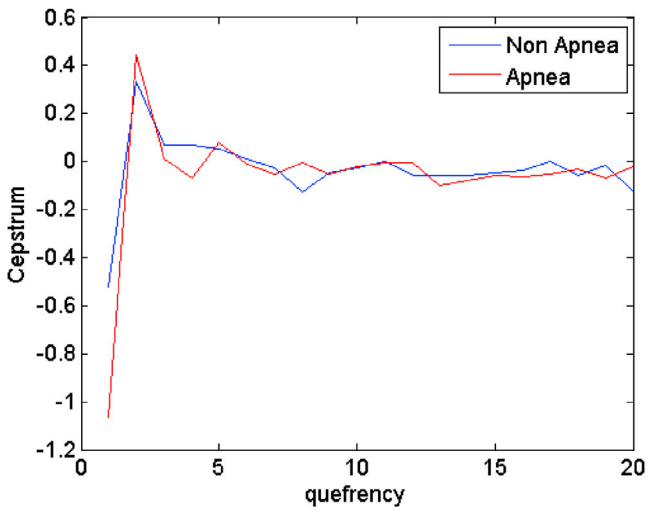

Fig. 1. Cepstrum coefficients obtained for two 5-min data windows (nonapnea/apnea minute).

The DFA was applied to the RR series, and the algorithm used to calculate the DFA parameters was as follows:

1 Determine the profile of the RR interval of length *N, B(k)*, where $\overline{B}$ is the mean value:

$$Y(i) = \sum_{k=1}^{i} \left(B(k) - \overline{B}\right) ; i = 1, \ldots, N \tag{3}$$

2 Divide the profile into $N_t = int\ (N/t)$ segments of equal-length *t*. As it is possible that the length *N* is not a multiple of time scale *t*, so that some samples of the end of the profile could not be included, the segmentation is performed twice, from the beginning to the end and vice versa, so that $2N_t$ segments are finally calculated.

3 Subtract the local trend for each segment. Local trend is calculated using a least-square fit of the data. Then, determine the variance of each segment v, $v = 1, \ldots, N_t$.

$$F_t^2(v) = \frac{1}{t} \sum_{t=1}^{t} [Y((v-1)t + i) - p_v(i)]^2 \tag{4}$$

4 Finally, the fluctuation function *F(t)* is calculated averaging over all segments and then taking the square root. Steps 2 to 4 are repeated for different time scales *t*, in order to establish a relationship between *F(t)* and the length of the segment (*t*).

$$F(t) = \sqrt{\frac{1}{2N_t} \sum_{v=1}^{2N_t} F_t^2(v)} \tag{5}$$

In Fig. 2, there is an example of application of the DFA. It represents the RR interval series for a frame, its integrated version corresponding to expression (3), and the trends estimated in each segment for two different time scales *t* ($t = 40$ and $t = 150$).

DFA parameters of RR-series are approximated by power-law [34,35,37,39] and if the data are long-range power-law correlated, *F(t)* increases for large values of *t*, as a power law:

$$F(t) \sim t^{\alpha} \tag{6}$$

where α is called the scaling exponent. The scaling exponent is also the slope of the line relating log *F(t)* to log *t*. For uncorrelated data, $\alpha = 0.5$. In short-term correlations, $\alpha > 0.5$ for small *t*, whereas in long-range correlations, $\alpha > 0.5$ for large *t*.

In the experiments, we wanted to distinguish between short-range and long-range correlations. That is why we use two scaling exponents: α_1 measures the short-range correlations, for time scales *t* between 10 and 40 beats and α_2 detects long-range correlation, using time scales *t* between 70 and 194. These boundaries have been used in previous work [21,29] with only one exception, the upper boundary for *t* in α_2. We calculate both α values in interbeat interval (RR) series of 5-min frames and their length does not allow us to use segments longer than 194 in all cases. Penzel et al. [21] point out the reasoning associated to the selection of these boundaries. The differences in the time scaling properties can help us to discriminate healthy and pathologic data sets. In both cases, data were fitted to a second-order polynomial.

In Fig. 3, a double logarithmic graph log*(F(t))* vs log*(t)* is shown, where the fluctuation function *F(t)* is plotted as a function of segment size *t*. The representation corresponds to the scaling analysis of the interbeat interval (RR) series in two 5-min frames, for an apneic and a nonapneic minute. The slope of the lines determines the scaling exponent (short-range scaling exponent α_1 and long-range scaling exponent α_2).

2.4. Classifiers

Binary classification is done on a minute by minute basis, where the classifier decides whether the evaluated minute shows apnea or not. The methods proposed for the quantification of apnea minutes are based on Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and Logistic Regression (LR), as these classifiers balance performance, complexity and interpretation.

2.4.1. LDA and QDA

LDA [40] is based on a parametric model with adjustable parameters that are obtained with the learning set. Once this first step is fulfilled, we are able to map the input features to the output classes.

A class-dependent multivariate Gaussian distribution is assumed for the features:

$$f_k(x) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)} \tag{7}$$

where μ_k and Σ_k are the mean vector and covariance matrix of each class *k* (apnea and nonapnea).

In the case of LDA, it is considered that both classes have the same covariances ($\Sigma_k = \Sigma$) and it is possible to define a linear boundary

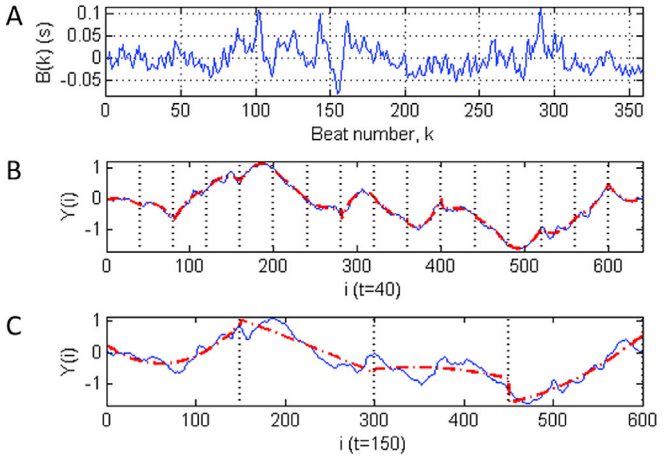

Fig. 2. Steps followed in the Detrended Fluctuation Analysis. (A) The interbeat interval (RR) series for a 5-min frame. (B) The integrated time series corresponding to equation (3), where B(k) is the RR series shown in (A). The vertical dotted lines show the limit between adjacent segments ($t = 40$). The dashed curves represent the trend estimated in each segment. (C) The same as (B) with $t = 150$.

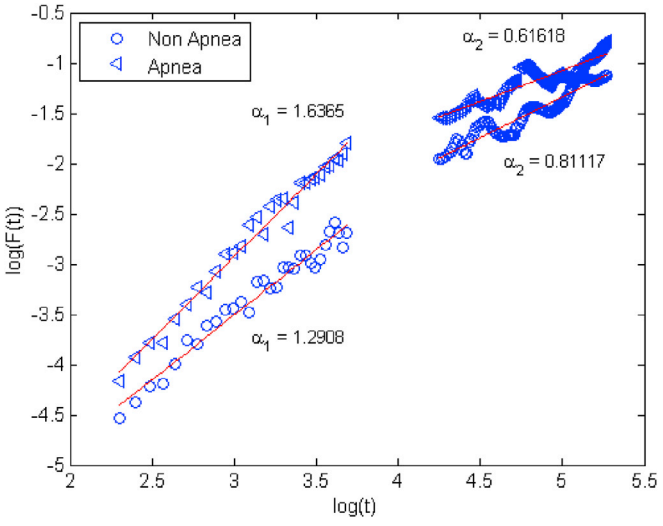

Fig. 3. Comparison of the scaling analysis for the interbeat interval (RR) series of a 5-min frame in presence/absence of apnea.

between the classes as:

$$\delta_k(x) = x^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + log\pi_k \quad (8)$$

where π_k is the prior probability of class k.

The parameters of the Gaussian distributions will be estimated using the learning data:

$$\widehat{\pi}_k = N_k/N \quad (9)$$

where N_k is the number of class-k observations.

$$\widehat{\mu}_k = \Sigma_{g_i=k}x_i/N_k \quad (10)$$

$$\widehat{\Sigma} = \Sigma_{k=1}^{K}\Sigma_{g_i=k}(x_i-\widehat{\mu}_k)(x_i-\widehat{\mu}_k)^T/(N-K) \quad (11)$$

The LDA will classify 1 min as apnea if:

$$x^T\widehat{\Sigma}^{-1}(\widehat{\mu}_{ap}-\widehat{\mu}_{nap}) > \frac{1}{2}\mu_{ap}^T\widehat{\Sigma}^{-1}\widehat{\mu}_{ap} - \frac{1}{2}\mu_{nap}^T\widehat{\Sigma}^{-1}\widehat{\mu}_{nap} + log\left(\frac{N_{nap}}{N}\right) - log\left(\frac{N_{ap}}{N}\right) \quad (12)$$

where N_{ap} and N_{nap} are the number of apnea and nonapnea observations.

If the covariance matrices (Σ_k) are not considered to be the same, we get quadratic discriminant functions. The boundary region is defined as:

$$\delta_k(x) = -\frac{1}{2}log|\Sigma_k| - \frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) + \log\pi_k \quad (13)$$

2.4.2. Logistic Regression

The third and last classifier proposed is Logistic Regression (LR), as it is considered a good predictor. In LR, the probability of apnea class is defined by p_{ap}:

$$p_{ap} = \frac{1}{1+e^{-(\beta_0+\beta_1X_{i1}+\ldots+\beta_pX_{ip})}} \quad (14)$$

where $X_{i1}, \ldots, X_{ip}$ represent the input feature vector in the instant i and $\beta_{0,\ldots,p}$ are the coefficients that fit the model. In order to optimize the classifier performance, we establish a threshold based on this probability.

2.5. Feature selection technique

In order to obtain a feature subset that reduces the dimensionality and maximizes the accuracy of the learning algorithm, a 250 repeated random sub-sampling validation is used. The number of repetitions allows us to obtain stable results in all cases under study. This process uses only the learning set (L). On each iteration a random partition of half of the feature vectors is used for training, and the rest for validation. The algorithm avoids feature vectors from one patient being simultaneously in both the training set and the validation set.

The feature selection process is performed in two phases. In a first step, for each iteration, a sequential forward feature selection method returns (based on the classifier performance) the optimal feature set, corresponding to the maximum accuracy in the validation. We rank the features according to the number of times they are included in the selected feature subset.

In the second step, also repeated 250 times, the error rate is obtained for an increasing number of features, that are entered in the same order as they appear in the ranking created in the first step. This allows us to analyze the evolution of the average misclassification error, obtained for the validation data, according to the number of features. The final selected features in the process will be those that produce the minimum average misclassification error. In all cases, the number of features selected is smaller than the original number of features. This process allows us to extract the variables that best describe the RR pattern in OSA.

3. Results

The general goal of all data analysis presented in this article is to evaluate the performance of different strategies to carry out the extraction of features from the RR-series obtained from single-lead ECG signals, and the classification of every feature vector, each corresponding to 1 min, as apnea or normal. In this way, we try to further our knowledge of the underlying physiological system.

3.1. Per-segment classification and feature selection

As we outlined in the section on feature extraction, we use features obtained from the spectral (Fbank) and cepstral (CC) analysis, and DFA variables. As a last test, all variables are put together. In particular, we use: CC, Fbank, DFA, CC + DFA, Fbank + DFA and Fbank + CC + DFA. In general, the aim is to evaluate the system's performance when different types of features individually quantify apnea minutes, and to analyze whether their combination enriches the feature vector and indirectly improves the system's performance. The features obtained in this way are entered into three different classifiers: LDA, QDA and LR. All the combinations and the results obtained for each of them in the test sets (T), are shown in Tables 1 (Physionet) and 2 (HuGCDN2014). Tables 3 and 4 contain the first 10 selected features, shown in order of appearance in the ranking vector described in Section 2.5, in Physionet and HuGCDN2014, respectively. In Physionet, Group A and C training recordings (L-set) have been used during the learning process, leaving out group B. The classification accuracy and the area under the receiver operating characteristic (ROC) curve, known as AUC, are used to evaluate the system's performance. Results were obtained for the ROC curve's cut-off point, where the sum of sensitivity and specificity was maximized. Fig. 4 shows, for the best combinations of feature extraction and classification strategy, the ROC curves for test in both datasets, obtained when the detection threshold is changing.

Results show that, regardless of the database used in each case, the best quantification is always achieved when we combine the three types of features: CC, Fbank and DFA. The only difference is that in Physionet we reach the best values (Acc.: 84.76%; sensitivity: 81.45%; specificity: 86.82%, AUC: 0.92) using the QDA classifier, whereas in HuGCDN2014, the best results (Acc.: 81.96%; sensitivity: 70.95%; specificity: 85.47%; AUC: 0.87) are obtained by the LDA classifier. These results suggest that there is relevant spectral and cepstral information in the signal, as well as short-term correlations, that play an important role in the classification task.

As far as the feature selection is concerned, in Section 2.5 we explained the method used to reduce the final number of features to improve the system's performance. Fig. 5 relates the evolution of the averaged misclassification error to the number of selected features for the best cases in the corresponding databases. The horizontal dotted line represents the misclassification error obtained without feature selection. We highlight the points with the minimum classification error, and the number of features in the x-axis of these points is the number that appears in the third column of the result tables (Tables 1 and 2).

But the decrease of features differs greatly depending on the case. In Tables 1 (Physionet) and 2 (HuGCDN2014), we can see the original number of features (N(or.)) and the reduced ones (N(red.)). The only cases where no reduction is proposed by the feature selection process is when the feature vector is only made up of the DFA variables. In these cases, we can find accuracy and AUC values that are below the results obtained when these variables (α_1 and α_2) are combined with the Fbank and CC coefficients.

The last three cases show the highest decrease of feature numbers, when all variables (Fbank, CC and DFA) are combined in the feature vector. There we find the highest mean reduction (89.9% in Physionet

Table 1
Per-segment and per-recording performance in terms of feature types, classifiers and number of features, in Physionet. The per-segment analysis was performed using the whole T set (35 recordings), whereas for the per-recording analysis, the five subjects of group B were discarded.

	N(or.)	N(red.)	Per-segment				Per-recording		
			ACC.	SENS.	SPEC.	AUC	ACC.	SENS.	SPEC.
CC + LDA	20	16	77.01	69.85	81.46	0.84	76.67	80.00	70.00
CC + QDA	20	15	79.27	77.54	80.36	0.88	93.33	95.00	90.00
CC + LR	20	18	78.16	73.04	81.35	0.85	83.33	85.00	80.00
Fbank + LDA	34	13	80.01	78.44	80.99	0.87	90.00	90.00	90.00
Fbank + QDA	34	17	82.42	81.10	83.25	0.91	96.67	95.00	100.00
Fbank + LR	34	13	79.98	77.98	81.23	0.87	93.33	90.00	100.00
CC + DFA + LDA	22	5	82.90	81.03	84.07	0.91	90.00	85.00	100.00
CC + DFA + QDA	22	7	82.48	76.44	86.24	0.90	90.00	85.00	100.00
CC + DFA + LR	22	5	82.62	77.49	85.81	0.90	86.67	85.00	90.00
Fbank + DFA + LDA	36	6	81.66	79.41	83.07	0.89	93.33	90.00	100.00
Fbank + DFA + QDA	36	4	80.24	81.98	79.18	0.89	93.33	90.00	100.00
Fbank + DFA + LR	36	6	81.69	78.86	83.46	0.89	93.33	90.00	100.00
DFA + LDA	2	2	80.08	83.92	77.73	0.88	93.33	90.00	100.00
DFA + QDA	2	2	79.30	83.35	76.81	0.88	96.67	95.00	100.00
DFA + LR	2	2	79.95	84.12	77.39	0.88	93.33	90.00	100.00
*+LDA	56	7	83.80	81.79	85.07	0.92	96.67	95	100
***+QDA**	**56**	**3**	**84.76**	**81.45**	**86.82**	**0.92**	**96.67**	**95**	**100**
*+LR	56	7	83.57	80.80	85.30	0.91	93.33	90	100

* Fbank(1 … 34) + CC(35 … 54) + α_1 (55) + α_2 (56).
(N(or.): original number of features, N(red.): number of selected features, ACC: Accuracy, SEN: Sensitivity, SPE: Specificity, AUC: area under the receiver operating characteristic (ROC) curve).

Table 2
Per-segment and per-recording performance in terms of feature types, classifiers and number of features, in HuGCDN2014. The analysis was performed using the whole T set (39 recordings).

	N(or.)	N(red.)	Per-segment				Per-recording		
			ACC.	SENS.	SPEC.	AUC	ACC.	SENS.	SPEC.
CC + LDA	20	12	73.43	60.59	77.52	0.76	74.36	68.42	80.00
CC + QDA	20	15	75.85	61.29	80.49	0.80	84.62	84.21	85.00
CC + LR	20	14	74.93	57.63	80.45	0.76	74.36	63.16	85.00
Fbank + LDA	34	4	77.61	74.69	78.54	0.84	89.74	94.74	85.00
Fbank + QDA	34	3	76.34	67.89	79.04	0.81	89.74	94.74	85.00
Fbank + LR	34	5	76.10	76.86	75.86	0.84	89.74	94.74	85.00
CC + DFA + LDA	22	3	77.92	72.68	79.59	0.84	76.92	78.95	75.00
CC + DFA + QDA	22	7	77.17	70.03	79.45	0.83	74.36	84.21	65.00
CC + DFA + LR	22	3	78.29	71.97	80.31	0.84	76.92	78.95	75.00
Fbank + DFA + LDA	36	13	76.42	80.32	75.18	0.86	84.62	94.74	75.00
Fbank + DFA + QDA	36	5	75.71	82.13	73.66	0.86	87.18	94.74	80.00
Fbank + DFA + LR	36	15	77.41	79.17	76.84	0.85	87.18	94.74	80.00
DFA + LDA	2	2	75.10	74.93	75.15	0.83	82.05	94.74	70.00
DFA + QDA	2	2	71.16	80.82	68.08	0.83	76.92	94.74	60.00
DFA + LR	2	2	75.02	75.06	75.01	0.83	82.05	94.74	70.00
***+LDA**	**56**	**7**	**81.96**	**70.95**	**85.47**	**0.87**	**87.18**	**78.95**	**95.00**
*+QDA	56	3	77.62	70.79	79.79	0.84	71.79	78.95	65.00
*+LR	56	9	81.48	70.24	85.06	0.86	84.62	78.95	90.00

* Fbank(1 … 34) + CC(35 … 54) + α_1 (55) + α_2 (56).
(N(or.): original number of features, N(red.): number of selected features, ACC: Accuracy, SEN: Sensitivity, SPE: Specificity, AUC: area under the receiver operating characteristic (ROC) curve).

and 88.7% in HuGCDN2014) and also the best results for both databases. In particular, for the best combinations in both databases (highlighted in bold), features are reduced from 56 to 3 in Physionet, and to 7 in HuGCDN2014, which represents a huge drop in both cases. Moreover, α_1 is always included, and α_2 discarded. These results are coherent with the scatter plots shown in Fig. 6, where α_1 and α_2 are represented, for apnea and nonapnea minutes, in both databases. The α values were calculated for all subjects of the datasets' test groups.

For the best cases in each database, boxplots with the selected features are also shown in Fig. 7 for Physionet and Fig. 8 for HuGCDN2014, where it is possible to observe the degree of dispersion of the features for apnea and nonapnea minutes.

3.2. Per-recording classification

Although the purpose of the study is the automatic classification of apneic events, the final goal of every PSG simplification approach is sleep apnea diagnosis. Therefore, the per-recording classification is performed according to the estimation of the AHI, derived from the number of apneic minutes per subject. In particular, the AHI is calculated computing the average number of minutes with apnea per hour, i.e. adding the total number of apnea minutes, dividing this number by the total number of minutes of the actual recording and multiplying the result by 60 [41].

The threshold to decide whether the subject belongs to the control or to the apnea group is defined for an estimated AHI of 15, a widely used value in the context of sleep apnea [42–44]. In Physionet, we discarded

Table 3
Chosen features after the feature selection process in Physionet.

CC + LDA	4	13	16	15	14	11	12	18	19	17
CC + QDA	4	16	15	13	17	19	11	12	14	18
CC + LR	4	13	16	14	15	12	19	17	11	18
Fbank + LDA	2	3	4	1	14	23	22	16	24	6
Fbank + QDA	2	3	4	14	1	17	16	12	13	5
Fbank + LR	2	3	4	14	1	23	22	16	24	12
CC + DFA + LDA	21	22	1	2	3					
CC + DFA + QDA	21	1	22	19	4	16	17			
CC + DFA + LR	21	22	1	3	4					
Fbank + DFA + LDA	35	2	1	22	24	23				
Fbank + DFA + QDA	35	1	2	23						
Fbank + DFA + LR	2	35	1	22	24	23				
DFA + LDA	1	2								
DFA + QDA	1	2								
DFA + LR	1	2								
[a]+LDA	55	2	35	1	22	24	42			
[a]+**QDA**	**55**	**1**	**35**							
[a]+LR	55	2	35	1	24	6	22			

[a] Features 1 to 34 correspond to the Fbanks, features 35 to 54 to the CCs, feature 55 to α_1 and feature 56 to α_2.

Table 4
Chosen features after the feature selection process in HuGCDC2014.

CC + LDA	3	15	1	16	13	18	12	11	6	17
CC + QDA	1	13	3	15	2	12	14	16	19	11
CC + LR	3	15	16	1	13	18	12	11	17	6
Fbank + LDA	2	1	3	6						
Fbank + QDA	1	2	6							
Fbank + LR	2	1	4	3	6					
CC + DFA + LDA	22	21	1							
CC + DFA + QDA	1	21	22	5	13	19	8			
CC + DFA + LR	21	22	1							
Fbank + DFA + LDA	2	35	1	3	11	12	36	9	10	14
Fbank + DFA + QDA	1	35	2	16	12					
Fbank + DFA + LR	2	1	35	3	9	10	11	12	36	4
DFA + LDA	2	1								
DFA + QDA	2	1								
DFA + LR	2	1								
[a]+**LDA**	**2**	**35**	**55**	**1**	**11**	**12**	**51**			
[a]+QDA	35	1	55							
[a]+LR	2	35	55	51	3	1	4	52	9	

[a] Features 1 to 34 correspond to the Fbanks, features 35 to 54 to the CCs, feature 55 to α_1 and feature 56 to α_2.

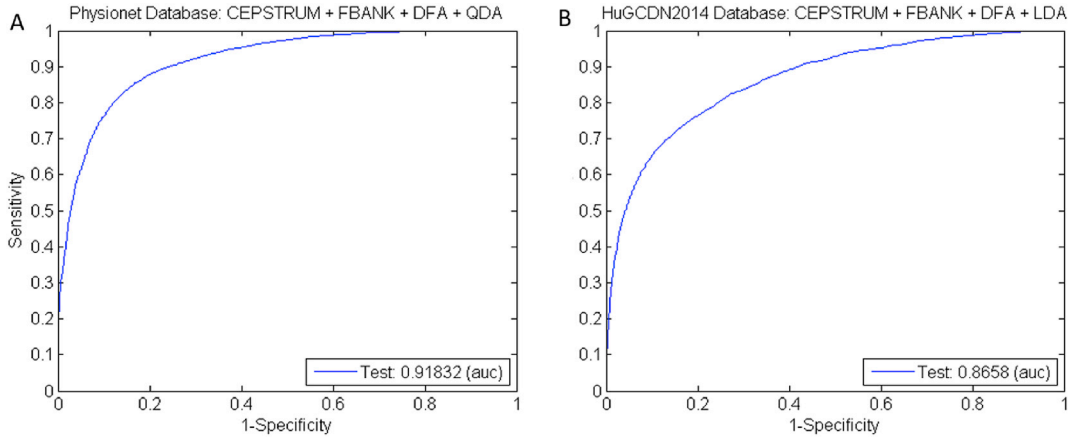

Fig. 4. ROC curves for test in both datasets, for the best combinations of feature extraction and classification strategy, obtained when the detection threshold is changing: (A) Physionet and (B) HuGCDN2014. The values obtained for the area under the curve are also shown in the legend.

the five patients of group B of the test set as indicated in the guidelines of the international competition jointly conducted by Computers in Cardiology and Physionet [7]. Results are shown in Tables 1 (Physionet) and 2 (HuGCDN2014). In the best quantification cases, the number of

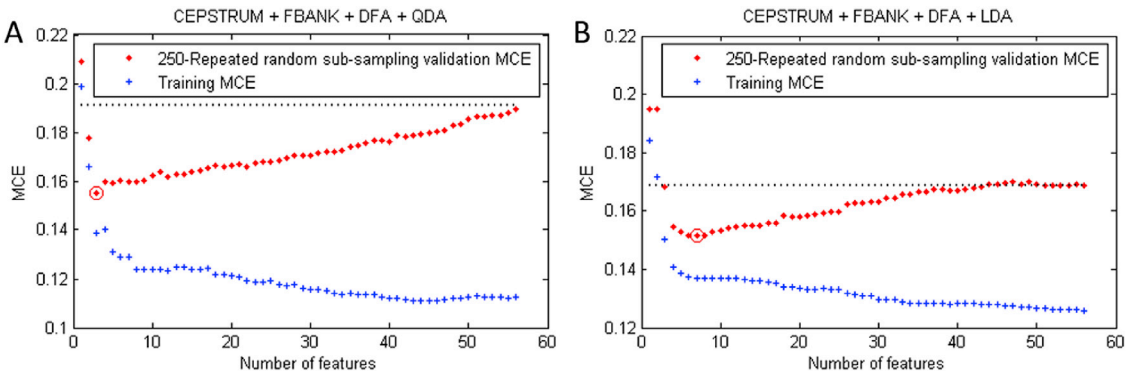

Fig. 5. Evolution of averaged misclassification error to the number of selected features in (A) Physionet and (B) HuGCDN2014.

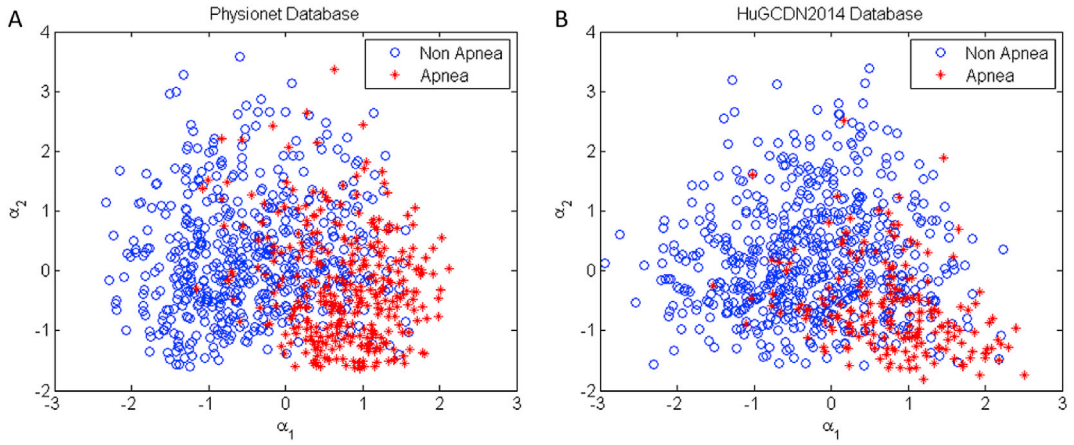

Fig. 6. Scatter plot of scaling components, α_1 and α_2, for apnea minutes and nonapnea minutes of the T sets. (A) Physionet (35 recordings) and (B) HuGCDN2014 (39 recordings).

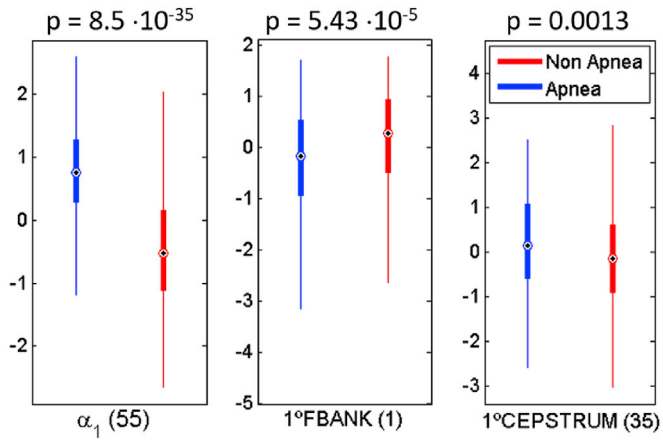

Fig. 7. Boxplot for the features selected in the best combination for Physionet.

global misclassification is one (out of 30 subjects) and five (out of 39 subjects) for Physionet and HuGCDN2014, respectively.

4. Discussion

This is the first study which systematically compares the system's performance for sleep apnea detection when spectral, cepstral and DFA variables, extracted from the HRV, are combined. In this section, we will focus on the physiological significance of the selected features, the limitations of the proposed method and the comparison with prior work.

4.1. Physiological significance of the selected features

The most important aspect, as far as the physiological interpretation is concerned, is the analysis of the feature selection done for each of the proposed combinations.

4.1.1. CC and CC + DFA

Focusing first in the CCs, it is remarkable that this analysis technique is suitable for detecting periodic structures in signals. Moreover, there is another key property of the cepstral analysis, often used in the context of speech, that allows us to separate the excitation from the transfer function of the system. The cepstral analysis is useful in this context as the low-frequency periodic excitation from the vocal cords and the formants of the vocal tract, are additive and in different regions in the quefrency domain. We can find parallels in the behavior of the RR-series [29]. There is a low frequency periodicity during apnea and a high frequency periodicity called Respiratory Sinus Arrhythmia (RSA) during normal breathing. We have applied this technique with good results in previous studies [26,27,29] to analyze the RR-series to detect periodicities during apnea episodes.

Analyzing the features selected using only CCs, we can observe that most of them are chosen, always including the first CC. If we focus on the coefficients that first appear in the list, we can see in Physionet that the 4th coefficient, related to the RSA component (0.25 cycles/beat), is especially important as it is always selected in first place. Next in ranking are coefficients 13, 14, 15 and 16, that are related to the low frequency fluctuation in breathing sleep pauses. These results agree with those obtained by Ravelo-García et al. [27]. In HuGCDN2014, outcomes are similar. In this case the 3rd coefficient and coefficients 15 and 16 are, in most cases, top. The interpretation would be the same as in Physionet.

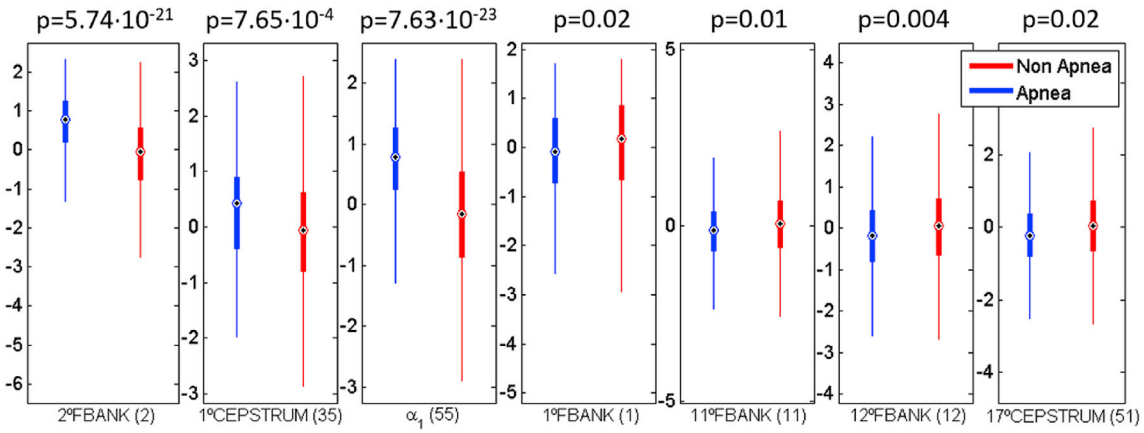

Fig. 8. Boxplot for the features selected in the best combination for HuGCDN2014.

We can deduce that CCs around 4th and 15th contribute valuable information about the presence or absence of breathing pauses during sleep.

Things change significantly when we combine CCs with the DFA variables. The feature selection algorithm leaves out most of the CCs, except the first ones, which appear top, as well as both DFA variables, that are always chosen.

4.1.2. *Fbank and Fbank + DFA*

On the other hand, in the combinations where we introduce only Fbank, we can see in Physionet that the first selected Fbanks (2, 3 and 4) cover from 0.015 to 0.05 cycles/beat, which is in the range Acharya indicates [19], (0.01–0.05 cycles/beat) for the CVHR. Next in ranking, and almost as important, are, Fbanks 1 (up to 0.015 cycles/beat), the lowest part of the already appointed Acharya range, and 14 (around 0.2 cycles/beat), representing the RSA component. In HuGCDN2014, the two first Fbanks (spectral information up to 0.03 cycles/beat) are always selected first.

However, when Fbank and DFA variables are put together, in contrast to the behavior of the system with the CCs, α_1 is always chosen, but α_2 appears only twice in the HuGCDN2014 database, and not in the first positions. As far as the selected Fbank is concerned, the two first Fbanks occupy, in both databases, a privileged position in the list. This means that, when combining DFA features with Fbank, we reach the highest discriminant capacity using the information of the very low frequency (VLF), which is related to the frequency fluctuation in sleep apneas.

4.1.3. *DFA*

In Section 2.3.3, we showed (see Fig. 3) the scaling analysis of two 5-min frames, representing an apneic and a nonapneic event. In the graphs, there is a difference in the scaling behavior when apnea occurs, but in both cases, the scaling exponents (α_1 and α_2) are greater than 0.5, the result of having short- and long-range correlations. Moreover, α_1 values are larger than α_2. This means that there are stronger short-range than long-range correlations. These results are consistent with those obtained by Penzel et al. [21]. They applied DFA to HRV for sleep apnea detection, but considering the sleep stages. In particular, they also combined spectral analysis with DFA, but the spectral parameters they worked with were not the same as in our studies. Unlike in our analysis, they concluded that changes in HRV are better quantified by scaling analysis than by spectral analysis. They highlight that α_1 is related to the effect of breathing on heart rate and α_2 could be related to effects of slower brain functions such as sleep regulation on heart rate. In our work, this would justify the selection of α_1 as one of the most important features in the quantification process, as it is based on HRV.

When comparing the values, we can see that for the short-range correlations, represented by α_1, the values during apnea minutes are greater than for nonapnea minutes. That is, during apnea, there are higher correlations in the signal, which suggest a loss of complexity and more regularity in the system. These results are in line with other authors' findings, not only in the context of the cardiovascular system: Mendez et al. [22] refer to the HRV loss of complexity in sleep apnea, Javorka et al. [45] point out the complexity loss and simplification of heart-rate dynamics in patients with diabetes mellitus and Puthanmadam et al. [46] report an increasing degree of structural complexity in the EEG of normal subjects compared to those of patients with epilepsy. This suggests, in general, a tendency towards less complexity in unhealthy physiological systems.

We can also observe differences in the long-range scaling exponent α_2 between apnea and nonapnea minutes, and in this case we find the 'crossover phenomena' suggested by Peng et al. [35], associated with a change in short and long-range scaling exponents. The values obtained are consistent with those presented by Ivanov et al. [47] for sleeping subjects.

4.1.4. *CC + Fbank + DFA*

Finally, we focus on the last three lines of the respective tables, where all the available information (CC, Fbank and DFA variables) is included and the best results are obtained. The feature selection process yielded quite robust results as similar features were selected, regardless of the database. The three variables that appear to carry significant information for the quantification process are: the first Fbank, the first CC and α_1. The second Fbank is also chosen when LDA and LR are the classifiers. This fact reinforces the previous theory where we set out that each of the proposed feature extraction techniques (Fbank, CC and DFA) contributed to the increase of classification accuracies and AUCs with valuable and complementary information of the heart rhythm behavior in the context of apnea. This result was foreseeable after analyzing the cases where we put together Fbank with DFA, and CC with DFA, and support the idea that the system generating the analyzed signal presents linear and nonlinear characteristics.

The fact that in the spectral features the first elements were selected, reinforces the theory that, in the very low spectrum frequencies (around 0.02 cycles/beat), we can find important information for classification purposes. That is, after combining linear and nonlinear features, the information of RSA (0.25 cycles/beat) is discarded. This highlights the superiority of the very low frequency spectral component, which reflects the rhythm of apnea repetition. This, in turn, agrees with the conclusions reached by Mendez et al. [12,22]. Therefore, we can conclude that the information carried by these variables is relevant for the description of the physiological process.

The fact that in all cases where α_1 is contained in the feature vectors, it is selected after the feature selection process, agrees with the results obtained by Ravelo-García et al. [29]. However, α_2 was always discarded.

This could be due to its poor discriminant capacity in the context of HRV, but we could also presume that the frames are not long enough to reflect long range correlations. These results agree with those obtained by Penzel et al. [21], who stated that short-range correlations appear to be much more discriminant than long-range correlations, and with results shown in Fig. 6.

If we analyze Tables 3 and 4 globally, we can deduce that the relative importance of the different features in the classification task depends on their association in the vector, but finally, when we combine all the variables, the results are conclusive. Regardless of the database, the combination of spectral and cepstral information (Fbank and CC) and information about short-range correlations, provided by α_1 (DFA), gives us the best discriminant capacity and best describes the underlying physiological process.

4.2. Limitations of the proposed method

Despite its contributions, some limitations need to be pointed out in this study. One of the databases employed for the analysis was the widely used Physionet database, which presents some drawbacks, e. g. a restricted number of subjects, whose age ranges from 27 to 63, or a small number of women included in the dataset (7 women and only one OSA diagnosed). The latter is especially important as, according to recent studies [16], there are potential gender differences in HRV sleep apnea information. In future studies, this consideration should be taken into account for a differentiated learning and validation process. Moreover, it would be desirable to include in the database apneic patients with various cardiovascular disorders, what would probably have an impact on the system's performance. Thus, for a clinical validation of the proposed approach, a larger database, including older participants, a higher number of women, and cardiac patients, would be necessary. As far as the HuGCDN2014 is concerned, the lack of subjects representing mild and moderate OSA patients in the database is also a limitation that should be taken into consideration.

4.3. Comparison with prior work

The performance of the OSA classification approach proposed in this article is compared with existing works in the literature. Table 5 shows a selection of the most representative methods that employ the widely used Physionet database with the results obtained for per-segment and per-recording classification, as for meaningful comparison, results obtained from the same database have to be compared.

Unlike other comparative studies, particular relevance is given to the number of features and the number of recordings used in the experiments. As shown in Table 5, Mendez et al. [12,22] obtain high accuracies. However, they discard in the process 20 recordings that do not satisfy certain criteria of data quality. So, their method requires high-quality datasets, that are not normally available, as physiological signals are, by nature, noisy. In this regard, our method is more robust because it does not require a preselection of high quality data.

Other remarkable outcomes are those reached by Schrader et al. [48], de Chazal et al. [9], Karandikar et al. [41] and Nguyen et al. [14]. In all these studies, the main weakness is the high-dimension feature space, over 20. From Table 5 we can conclude that the proposed method gives better per-segment accuracies than the most recent literature. There is only one exception, the method presented by Song et al. [17]. They obtained a per-segment accuracy slightly higher than in the proposed approach, namely 86.2%, but applying 9 features. Moreover, in their method they derive two signals from the ECG, RR and EDR, like Varon et al. [15]. However, in our approach, only the RR interval series is obtained from the ECG, thus simplifying the preprocessing stage. So, we consider our results very promising. Equivalent or better performance is reached compared to other methods existing in the recent literature at reduced computational cost, as only 3 features are required, i.e. to the best of our knowledge, the least amount of features ever used in this context.

As far as the per-recording accuracies are concerned, our proposed algorithm provides results comparable to those obtained by different methods in the state of the art. In particular, only one recording out of 30 is misclassified.

In summary, results indicate that the use of single-lead ECGs can reach good accuracies in the detection of sleep apnea. Nevertheless, there is still margin for further improvement of the global system performance, e.g. applying other pattern recognition methods and classifier combination rules, trying other feature selection algorithms and comparing the results, or introducing more powerful feature selection methods, such as Genetic Algorithms, Ant Colony Algorithm or Simulated Annealing.

5. Conclusions

This paper presents a methodology for the automatic detection of sleep apnea from single – lead ECG. Only 3 features, derived from RR interval series, allow us to reach performances that are comparable or better than those reported in the literature for fully automated algorithms. Thus, we have been able to characterize the underlying process from a cardiac rate point of view combining a very small number of features, originating from different domains. In particular, results suggest that the first two Fbanks, the first CC and the first DFA coefficient are of special interest when extracting relevant discriminative information for apnea screening purposes. Moreover, according to the used classifiers, our system shows a good performance/complexity ratio that makes it interesting for HW implementation. In particular, in Physionet, the accuracy achieved by the proposed algorithm is 84.76% for per-segment classification and 96.67 for per-recording classification. In HuGCDN2014, results are 81.96% and 87.18%, respectively. According

Table 5
Comparison of per-segment and per-recording OSA detection results on Physionet database.

Method	Year	Nº of recordings	Nº of features	Per-segment			AUC	Per-recording Acc.(%)
				Acc(%)	Sen(%)	Spe(%)		
Spectral features and LDA [48]	2000	70	30	88.31	–	–	–	93.3
Temporal and spectral RR and EDR features and LDA [9]	2003	70	88	90	86.4	92.3	–	100
Sample entropy, spectral features [49]	2007	70	6	72.9	72.2	73.3	–	90
Temporal and spectral features from RR and QRS area and kNN [22]	2009	50	10	88	85	90	–	100
WA and QDA [12]	2010	50	10	89.07	90.37	86.37	–	100
RQA of HRV and EDR, and Autoeural model [41]	2013	70	21	88.06	91.93	85.84	–	–
RQA and soft decision fusion rule (SVM and NN) [14]	2014	70	72	85.26	86.37	83.47	–	–
			33	84.19	85.81	81.69		
Principal components of QRS and orthogonal subspace projections (LS-SVM) [15]	2015	70	6	84.74	84.71	84.69	0.88	100
HMM and SVM [17]	2016	70	9	86.2	82.6	88.4	0.94	97.1[a]
Hermite basis functions and LS-SVM [18]	2016	70	5	83.8	79.5	88.4	0.83	97.14[a]
Filterbank, Cepstrum, DFA and QDA (our approach)	2017	70	3	84.76	81.45	86.82	0.92	96.67

[a] Group B of the test set is included.

to the accuracy rates obtained in our experiments, we can say they are in the confidence interval of agreement among experts (85%). The fact that in the experiments the results obtained for two different databases were very similar, supports that the conclusions reached are potentially generalizable, and would not be limited by database variability, something we consider a contribution with respect to other state of the art articles. Hence, these results indicate that the use of single-lead ECG signals can achieve good accuracies in the detection of sleep apnea and encourages the development of home-based OSA screening devices.

Conflict of interest statement

None declared.

Acknowledgements

We would like to thank the staff of the Dr. Negrín University Hospital Sleep Unit for their support in the acquisition of the data, as well as to Ubay Casanova for his collaboration in the first field experiments. This work was partially supported by Cátedra Telefónica, Universidad de Las Palmas deGran Canaria 2013.

References

[1] P. Peppard, T. Young, J. Barnet, M. Palta, E. Hagen, K. Hla, Increased prevalence of sleep-disordered breathing in adults, Am. J. Epidemiol. 177 (9) (2013) 1006–1014.
[2] A. Ravelo-García, J. Kraemer, J. Navarro-Mesa, E. Hernández-Pérez, J. Navarro-Esteva, G. Juliá-Serdá, et al., Oxygen saturation and RR intervals feature selection for sleep apnea detection, Entropy 17 (5) (2015) 2932–2957.
[3] E. Phillipson, Sleep apnea - a major public health problem, N. Engl. J. Med. 328 (17) (1993) 1271–1273.
[4] R.B. Berry, R. Budhiraja, D.J. Gottlieb, D. Gozal, C. Iber, V.K. Kapur, …, S. Redline, Rules for scoring respiratory events in sleep: updateof the 2007 AASM manual for the scoring of sleep and associated events:deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine, J. Clin. SleepMed. JCSM Off. Publ. Am. Acad. SleepMed. 8 (5) (2012) 597.
[5] M. Bonnett, D. Carley, M. Carskadon, EEG arousal: scoring rules and examples: a preliminary report from the sleep disorders atlas task force of the American Sleep Disorders Association, Sleep 15 (1992) 173–184.
[6] S. Verhulst, First night effect for polysomnographic data in children and adolescents with suspected sleep disordered breathing, Archives Dis. Child. 91 (3) (2006) 233–237.
[7] T. Penzel, J. McNames, P. de Chazal, B. Raymond, A. Murray, G. Moody, Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings, Med. Biol. Eng. Comput. 40 (4) (2002) 402–407.
[8] Z. Shinar, A. Baharav, S. Akselrod, Obstructive sleep apnea detection based on electrocardiogram analysis, in: Computers in Cardiology 2000, IEEE, 2000, pp. 757–760.
[9] P. de Chazal, C. Heneghan, E. Sheridan, R. Reilly, P. Nolan, M. O'Malley, Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea, IEEE Trans. Biomed. Eng. 50 (6) (2003) 686–696.
[10] M. Mendez, D. Ruini, O. Villantieri, M. Matteucci, T. Penzel, S. Cerutti, et al., Detection of sleep apnea from surface ECG based on features extracted by an autoregressive model, in: Engineering in Medicine and Biology Society 2007, IEEE, 2007, pp. 6105–6108.
[11] A. Khandoker, M. Palaniswami, C. Karmakar, Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings, IEEE Trans. Inf. Technol. Biomed. 13 (1) (2009) 37–48.
[12] M. Mendez, J. Corthout, S. Van Huffel, M. Matteucci, T. Penzel, S. Cerutti, et al., Automatic screening of obstructive sleep apnea from the ECG based on empirical mode decomposition and wavelet analysis, Physiol. Meas. 31 (3) (2010) 273–289.
[13] T. Le, C. Cheng, A. Sangasoongsong, S. Bukkapatnam, Prediction of sleep apnea episodes from a wireless wearable multisensor suite, in: Point-of-Care Healthcare Technologies (PHT) 2013, IEEE, 2013, pp. 152–155.
[14] H. Nguyen, B. Wilkins, Q. Cheng, B. Benjamin, An online sleep apnea detection method based on recurrence quantification analysis, IEEE J. Biomed. health Inf. 18 (4) (2014) 1285–1293.
[15] C. Varon, A. Caicedo, D. Testelmans, B. Buyse, S. Van Huffel, A novel algorithm for the automatic detection of sleep apnea from single-lead ECG, IEEE Trans. Biomed. Eng. 62 (9) (2015) 2269–2278.
[16] G. Gutiérrez-Tobal, D. Álvarez, J. Gomez-Pilar, F. del Campo, R. Hornero, Assessment of time and frequency domain entropies to detect sleep apnoea in heart rate variability recordings from men and women, Entropy 17 (1) (2015) 123–141.
[17] C. Song, K. Liu, X. Zhang, L. Chen, X. Xian, An obstructive sleep apnea detection approach using a discriminative hidden Markov model from ECG signals, IEEE Trans. Biomed. Eng. 63 (7) (2016) 1532–1542.
[18] H. Sharma, K. Sharma, An algorithm for sleep apnea detection from single-lead ECG using Hermite basis functions, Comput. Biol. Med. 77 (2016) 116–124.
[19] U. Acharya, K. Joseph, N. Kannathal, C. Lim, J. Suri, Heart rate variability: a review, Med. Biol. Eng. Comput. 44 (12) (2006) 1031–1051.
[20] C. Guilleminault, R. Winkle, S. Connolly, K. Melvin, A. Tilkian, Cyclical variation of the heart rate in sleep apnoea syndrome, Lancet 323 (8369) (1984) 126–131.
[21] T. Penzel, J. Kantelhardt, L. Grote, J. Peter, A. Bunde, Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea, IEEE Trans. Biomed. Eng. 50 (10) (2003) 1143–1151.
[22] M. Mendez, A. Bianchi, M. Matteucci, S. Cerutti, T. Penzel, Sleep apnea screening by autoregressive models from a single ECG lead, IEEE Trans. Biomed. Eng. 56 (12) (2009) 2838–2850.
[23] M. Tayel, E. AlSaba, Review: nonlinear techniques for analysis of heart rate variability, Int. J. Res. Eng. Sci. (IJRES) 4 (2) (2016) 45–60.
[24] D. Christini, K. Stein, S. Markowitz, S. Mittal, D. Slotwiner, M. Scheiner, et al., Nonlinear-dynamical arrhythmia control in humans, Proc. Natl. Acad. Sci. 98 (10) (2001) 5827–5832.
[25] J. Zbilut, N. Thomasson, C. Webber, Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals, Med. Eng. Phys. 24 (1) (2002) 53–60.
[26] A. Ravelo-García, J. Navarro-Mesa, S. Martín-González, E. Hernández-Pérez, P. Quintana-Morales, I. Guerra-Moreno, et al., Cepstrum coefficients of the RR series for the detection of obstructive sleep apnea based on different classifiers, EUROCAST 2013 (2013) 266–271.
[27] A. Ravelo-García, J. Navarro-Mesa, E. Hernández-Pérez, S. Martín-González, P. Quintana-Morales, I. Guerra-Moreno, et al., Cepstrum feature selection for the classification of sleep apnea-hypopnea syndrome based on heart rate variability, in: Computers in Cardiology 2013, IEEE, 2013, pp. 959–962.
[28] A. Ravelo-García, J. Navarro-Mesa, U. Casanova-Blancas, S. Martin-Gonzalez, P. Quintana-Morales, I. Guerra-Moreno, et al., Application of the permutation entropy over the heart rate variability for the improvement of electrocardiogram-based sleep breathing pause detection, Entropy 17 (3) (2015) 914–927.
[29] A. Ravelo-García, U. Casanova-Blancas, S. Martín-González, E. Hernández-Pérez, I. Guerra-Moreno, P. Quintana-Morales, et al., An approach to the enhancement of sleep apnea detection by means of detrended fluctuation analysis of RR intervals, in: Computers in Cardiology 2014, IEEE, 2014, pp. 905–908;
[30] T. Penzel, G. Moody, R. Mark, A. Goldberger, J. Peter, The apnea-ECG database, in: Computers in Cardiology 2000, IEEE, 2000, pp. 255–258.
[31] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, Head rate variability: standards of measurement, physiological interpretation, and clinical use, Circulation 93 (1996) 1043–1065.
[32] N. Wessel, A. Voss, H. Malberg, C. Ziehmann, A. Shirdewan, U. Meyerfeldt, et al., Nonlinear analysis of complex phenomena in cardiological data, Herzschrittmachertherapie Elektrophysiol. 11 (3) (2000) 159–173.
[33] N. Wessel, A. Voss, J. Kurths, P. Saparin, A. Witt, H.J. Kleiner, et al., Renormalised entropy: a new method of nonlinear dynamics for the analysis of heart rate variability, in: Computers in Cardiology 1994, IEEE, 1994, pp. 137–140.
[34] A. Golińska, Detrended fluctuation analysis (DFA) in biomedical signal processing: selected examples. Studies in Logic, Gramm. Rhetor. 29 (2012) 107–115.
[35] C. Peng, S. Havlin, H. Stanley, A. Goldberger, Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, Chaos An. Interdiscip. J. Nonlinear Sci. 5 (1) (1995) 82–87.
[36] J. Lee, D. Kim, I. Kim, K. Park, S. Kim, Detrended fluctuation analysis of EEG in sleep apnea using MIT/BIH polysomnography data, Comput. Biol. Med. 32 (1) (2002) 37–47.
[37] P. Absil, R. Sepulchre, A. Bilge, P. Gérard, Nonlinear analysis of cardiac rhythm fluctuations using DFA method, Phys. A Stat. Mech. Appl. 272 (1–2) (1999) 235–244.
[38] R. Acharya, C. Lim, P. Joseph, Heart rate variability analysis using correlation dimension and detrended fluctuation analysis, ITBM-RBM 23 (6) (2002) 333–339.
[39] E. Rodriguez, J. Echeverria, J. Alvarez-Ramirez, Detrended fluctuation analysis of heart intrabeat dynamics, Phys. A Stat. Mech. Appl. 384 (2) (2007) 429–438.
[40] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, second ed., Springer Verlag, 2009.
[41] K. Karandikar, T. Le, A. Sa-ngasoongsong, W. Wongdhamma, S. Bukkapatnam, Detection of sleep apnea events via tracking nonlinear dynamic cardio-respiratory coupling from electrocardiogram signals, in: 6th International IEEE/EMBS Conference on Neural Engineering (NER) 2013, IEEE, 2013, pp. 1358–1361.
[42] E. Bixler, A. Vgontzas, J. Gaines, J. Fernandez-Mendoza, S. Calhoun, D. Liao, Moderate sleep apnoea: a “silent” disorder, or not a disorder at all? Eur. Respir. J. 47 (1) (2015) 23–26.
[43] N. Shah, D. Hanna, Y. Teng, D. Sotres-Alvarez, M. Hall, J. Loredo, et al., Sex-specific prediction models for sleep apnea from the hispanic community health study/study of latinos, Chest 149 (6) (2016) 1409–1418.
[44] T. Kasai, S. Ishiwata, M. Ohno, K. Narui, Y. Tomita, Relationship between severity of sleep apnea and extent of coronary atherosclerosis. B64, in: BOP Goes the Heart: Cardiovascular Consequences of SDB, American Thoracic Society, 2016. A4195–A4195.
[45] M. Javorka, Z. Trunkvalterova, I. Tonhajzerova, Z. Lazarova, J. Javorkova, K. Javorka, Recurrences in heart rate dynamics are changed in patients with diabetes mellitus, Clin. Physiol. Funct. Imaging 28 (5) (2008) 326–331.

[46] N. Puthanmadam Subramaniyam, J. Hyttinen, Characterization of dynamical systems under noise using recurrence networks: application to simulated and EEG data, Phys. Lett. A 378 (46) (2014) 3464–3474.
[47] P. Ivanov, A. Bunde, L. Amaral, S. Havlin, J. Fritsch-Yelle, R. Baevsky, et al., Sleep-wake differences in scaling behavior of the human heartbeat: analysis of terrestrial and long-term space flight data, Europhys. Lett. (EPL) 48 (5) (1999) 594–600.
[48] M. Schrader, C. Zywietz, V. Von Einem, B. Widiger, G. Joseph, Detection of Sleep Apnea in Single Channel ECGs from the PhysioNet Data Base, in: Computers in Cardiology 2000, IEEE, 2000, pp. 263–266.
[49] H. Al-Angari, A. Sahakian, Use of sample entropy approach to study heart rate variability in obstructive sleep apnea syndrome, IEEE Trans. Biomed. Eng. 54 (10) (2007) 1900–1904.