31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Detection of obstructive sleep apnea in ECG recordings using time-frequency distributions and dynamic features

A.F. Quiceno-Manrique, J.B. Alonso-Hernández, C.M. Travieso-González,
M.A. Ferrer-Ballester and G. Castellanos-Domínguez

***Abstract*— Detection of obstructive sleep apnea can be performed through heart rate variability analysis, since fluctuations of oxygen saturation in blood cause variations in the heart rate. Such variations in heart rate can be assessed by means of time-frequency analysis implemented with time-frequency distributions belonging to Cohen's class. In this work, dynamic features are extracted from time frequency distributions in order to detect obstructive sleep apnea from ECG signals recorded during sleep. Furthermore, it is applied a methodology to measure the relevance of each dynamic feature, before the implementation of k-nn classifier used to recognize the normal and pathologic signals. As a result, the proposed method can be applied as a simple diagnostic tool for OSA with a high accuracy (up to 92.67%) in one-minute intervals.**

I. INTRODUCTION

Sleep apnea is a disorder in which the patient stops breathing while he is in deep sleeping. There are three different types of sleep apnea: obstructive, central and mixed. The former apnea is produced by an obstruction in the pharynx caused by loss of pharyngeal dilator muscle tone and causes recurrent pharyngeal collapse and temporary cessation of breathing [1]. On the other hand, in central apnea the upper airway is open, but respiratory effort is absent or reduced. In mixed apnea, both central and obstructive aspects are present. A typical mixed apnea may show a period of central apnoea for several seconds, during which the upper airway occludes, followed by increased respiratory effort against the obstruction. In this work, the obstructive sleep apnea (OSA) is studied.

In order to perform a diagnosis of OSA, it is necessary to detect the presence of repetitive episodes of apnea and hypopnea during sleep. This presence is most reliably shown by attended overnight polysomnography in a sleep laboratory. If the patient shows five or more apneas or hypopneas per hour of sleep, is diagnosed with OSA. The severity of disease can be measured by the apnea-hypopnea index (AHI), which indicates the total time of apneas per hour of sleep [1].

The standard polysomnography test consists of recording various physiological parameters including EEG, ECG, EMG of chins and legs, nasal airflow, electro-oculogram (EOG), abdominal and thoracic movements and blood oxygen saturation (SaO2). However, the high cost of the system, discomfort of the electrodes connecting to the body and the high amount of information required to be analyzed are the main disadvantages of this method. The most promising mean for home diagnosis of OSA is the heart rate variability (HRV) analysis [2]. Then, the HRV signal can be estimated through one lead ECG recording followed by QRS detection, and its use to detect obstructive sleep apnea is supported by early researches which concluded that the events of apnea and hypopnoea are accompanied by concomitant cyclic variations in heart rate [3].

The aim of this work is to apply time-frequency analysis to HRV time series extracted from ECG signal (using a QRS detection algorithm [4]), in order to detect OSA in intervals of 1-minute length. The time-frequency representations applied belong to the Cohen's class of quadratic distributions, used previously in HRV analysis [5] to detect OSA. Then, dynamic features are extracted from the representations by means of linear filter cepstral coefficients and spectral subband centroids. Finally, it is applied a method to measure the relevance of each dynamic feature and the results are validated through cross-validation using a k-nn classifier.

II. MATERIALS AND METHODS

A. Time-varying spectral analysis

Time-frequency (t-f) representation (TFR) is the process of estimating the time varying spectral content of non-stationary signals, which are not completely described by a stationary spectral analysis. The result of the TFR is a two-dimensional matrix, $X(t, f)$, representing the joint energy distribution in time and frequency.

The t-f density of $x(t)$ can be represented by means of the *spectrogram* computed from Short Time Fourier Transform (STFT):

$$|X_\phi(t,f)|^2 = \left| \int_T x(\tau)\phi_a(\tau - t)e^{-j2\pi f\tau} d\tau \right|^2, \quad t, \tau \in T \tag{1}$$

In the STFT-based surface the window length remains constant. Therefore, the extraction of information with fast changes in time (i.e. high frequency components), should be accomplished with short and well-timed localized intervals. In contrast, the low frequency com-

A.F. Quiceno-Manrique and G. Castellanos-Domínguez are with Control and Digital Signal Processing Group, Universidad Nacional de Colombia, sede Manizales. {afquicenom,cgcastellanosd}@unal.edu.co
J.B. Alonso-Hernández, C.M. Travieso-González and M.A. Ferrer-Ballester are with Universidad de Las Palmas de Gran Canaria, España. {jalonso,ctravieso,mferrer}@dsc.ulpgc.es

ponents should involve large time intervals of analysis. As a result, as a nonstationary signal of relatively small time-bandwidth product, the TFR spectrogram-based on STFT is not suitable for revealing the t-f dynamics. This issue is certainly partially solved by using different windowing functions [6].

In practice, the quadratic energy distributions, which distribute the energy of a signal over t-f planes without windowing, are broadly used because of their flexibility, since the time and frequency resolutions can be adapted independently to suit the particular signal and cross terms [7]. One of the most commonly studied high resolution TFR is the generalized bilinear class, or *distributions of Cohen's class*, $X_C(t,f)$ which for a time t and frequency f variables are stated as follows:

$$X_C(t,f) = \int\int h_C(t-u,\tau)x(u+\tau/2)x^*(u-\tau/2) \\ e^{-j2\pi f\tau}dud\tau, \quad X_C(t,f), \quad h_C(t,\tau) \in L^2(\mathbb{R}) \tag{2}$$

where the 2-D function $h_C(t,\tau)$ is a time-lag kernel, that defines the particular TFR and can be stated as follows:

$$h_C(t,\tau) = h(t)\,g(\tau); \;\; h(t),\; g(\tau) \in L^2(\mathbb{R}) \tag{3}$$

being $h(t)$ a function of time, and $g(\tau)$ the window function in lag.

The variety of TFRs and their properties are determined by the choice of the *kernel function*, which is basically a two dimensional smoothing filter that reduces the interference terms inherent for the Cohen's class of time-frequency distributions, caused by its quadratic nature and the presence of multi-component signals. For instance, by setting the time-lag kernel equal to 1, the distribution will be recognized as the Wigner-Ville distribution (WVD), which provides a high resolution in both t-f planes, but only for mono-component signals. In multi-component cases, and due to the cross-term artifacts caused by aliasing, the TFRs present worse performance. To overcome this problem, any widely accepted window function such as the Gaussian, Hamming or Hanning type can be applied to the WVD to smooth the cross-terms. So, for the smooth windowed WVD or smoothed pseudo WVD (SPWVD), the functions $h(t)$ and $g(\tau)$ of the kernel (3) are chosen to be time-lag smoothed, eliminating the interferences by setting a parameter. Thus, improving the WVD by modifying the kernel function can suppress the cross-terms while maintaining a good t-f resolution.

B. Linear frequency cepstral coefficients

This method uses a filter bank whose triangular filters are linearly spaced in the frequency domain. The outputs of the M band-pass filters can be calculated by a weighted summation of each filter response set $\{H_m[k] : m = 1,\ldots,M\}$ and the energy spectrum $|X[n,k]|^2$ [8]:

$$c_m[n] = \sum_{k=1}^{K} |X[n,k]|^2 H_m[k], \quad 1 \le m \le M \tag{4}$$

where m, n and k are indexes for filter ordinal, time and frequency axes, respectively; K is the number of samples in the frequency domain.

Next, a Discrete Cosine Transform is taken over the log filter bank energies, so the final linear frequency cepstral coefficients (LFCC), C_p, for the desired number of cepstral features P, can be written by:

$$C_p[n] = \sum_{m=1}^{M} \log(c_m[n]) \cos\left[p\left(m-\frac{1}{2}\right)\frac{\pi}{P}\right], \quad 1 \le p \le P \tag{5}$$

C. Subband spectral centroids

Another parameterization framework that efficiently combines frequency and magnitude information from the short-term power spectrum of PCG signals is achieved through the computation of TFR spectral subbands. An effective method of combining the frequency and magnitude information from the power spectrum can be achieved through computation of subband spectral centroid histograms [9]. The computation of the centroid is achieved by:

$$\kappa_m[n] = \frac{\sum_{k=1}^{K} kH_m[k]\,|X^\gamma[n,k]|^2}{\sum_{k=1}^{K} H_m[k]\,|X^\gamma[n,k]|^2} \tag{6}$$

where γ is a parameter that decides the dynamic range of the spectrum used in the computation of the centroid. The energy around each centroid in the fixed bandwidth Δk is computed by means of:

$$\hat{E}_m[n] = \sum_{k=\kappa_m[n]-\Delta k}^{\kappa_m[n]+\Delta k} |X[n,k]|^2, \quad 1 \le m \le M \tag{7}$$

D. Analysis of dynamic relevance

Let $Y = \{y_{ij}[k] \in \mathbb{R}^m : j = 1,\ldots,p; i = 1,\ldots,n\}$ be the input training set, where $y_{ij}[k]$ is the j-th dynamic feature changing over discrete time k, that belongs to i-th observation, p is the number of features, and n the number of observations. Each observation vector, noted as $\mathbf{y_i}$, can be represented by a supervector of size $pm \times 1$, i.e., $\mathbf{y_i} = [y_{i1}[1], y_{i1}[2], \ldots, y_{i1}[m], y_{i2}[1], \ldots, y_{i2}[m], \ldots, y_{ip}[m]]^\top$. After centering each of the observation supervectors by subtracting the mean vector over the entire observation set, the respective estimation of the covariance matrix can be computed as $\mathbf{\Sigma_y} = \mathbf{Y}\mathbf{Y}^\top/n$, where each column of $\mathbf{Y}$ corresponds to $\mathbf{y_i}$. The eigenvectors associated with the q largest eigenvalues of $\mathbf{\Sigma_y}$ are selected as *principal directions*, which span an orthonormal basis for a subspace containing most of the information given by observations.

The eigenvalues λ_i and eigenvectors $\overline{\lambda}_i$ can be used to form the weighting feature vector, defined as follows:

$$\mathbf{w} = \sum_{i=1}^{q} |\lambda_i \overline{\lambda}_i| \tag{8}$$

so, that its larger values, no matter what sign they carry out, typify the most significant dynamic features. Then, the following weighting supervector is obtained:

$$\mathbf{w}_{pm\times 1} = [w_{11}, w_{12}, \ldots, w_{1m}, w_{21}, \ldots, w_{2m}, \ldots, \quad (9)$$

$$w_{p1}, \ldots, w_{pm}]^{\top}, w_{ij} \geq 0 \quad (10)$$

Rearranging $\mathbf{w}$, in such a way that a weighting value is allocated for each of the dynamic features:

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{21} & \cdots & w_{p1} \\ w_{12} & w_{22} & \cdots & w_{p2} \\ \vdots & \vdots & & \vdots \\ w_{1m} & w_{2m} & \cdots & w_{pm} \end{bmatrix} \quad (11)$$

the j-th scalar value $\overline{w}_j = \sum_{k=1}^{m} w_{jk}, \quad j = 1, \ldots, p$ can then be achieved, which is the sum of the elements for every j column of weighting matrix. In consequence, the main assumption is that the largest values of $\overline{w}_j$ point out to the best input attributes, since they exhibit higher overall correlations with principal components.

E. Database

The database used in this study is formed by 35 recordings, containing a single continuous ECG signal of approximately 8 hours duration [10]. The ECG recordings were extracted from a larger database of simultaneously recorded polysomnogram measurements. The ECG signal was sampled at $100Hz$, with 16 bit resolution, with one sample bit representing 5 μV.

Although the database is labeled by one-minute intervals, the analysis intervals taken during this work are 3 minutes in order to include the low-frequency components of the signal. The automatic diagnosis is given for the central minute, and the output of the classifier is compared against the label.

There were extracted 8928 intervals (4464 per class, in order to have a balanced class problem), each one of 3 minutes. The estimation of HRV might fail due to noise. Therefore, the best 4000 intervals were selected, and two tests were performed: the first test uses all 8928 observations, while the second test uses only the 4000 best observations.

F. Computation of dynamic features

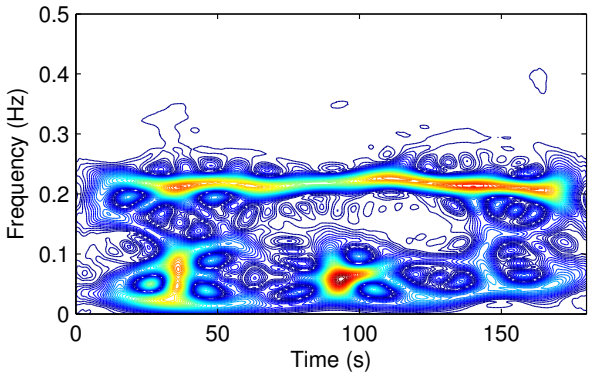
According to the methods explained previously, there were computed 30 dynamic features distributed in the following way:

1) Spectral centroids (1-10)
2) Energy of spectral centroids (11-20)
3) Cepstral coefficients (21-30)

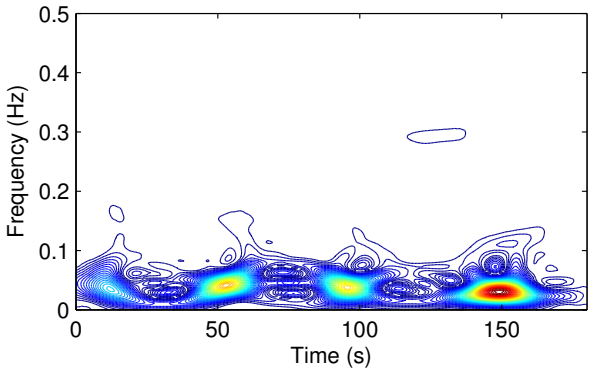
The features were estimated using 10 filters linearly distributed in the band between 0 and $0.5Hz$.

III. RESULTS

Figure 1 shows the time-frequency representations of normal and pathologic HRV signals. It can be seen that the normal signal has a constant frequency component around $0.23Hz$ due to the modulation caused by respiration, whereas the pathologic representation does not have that component, and its energy is concentred around the low frequencies.



(a) Normal



(b) Apnea

Fig. 1. Time-frequency representations (SPWVD) of normal and pathologic signals

The results of the analysis of dynamic relevance are shown in the Figure 2, where each dynamic contour has a relative weight associated, the larger the weight, the larger its relevance.

According to the relevance shown in Figure 2, there were selected the 10 most relevant contours, and a cross-validation process was performed, using 70% of the observations to train the classifier, and the remaining 30% for validation purposes. There was used a k- nearest neighbor classifier with $k = 3$, and principal component analysis (PCA) in order to reduce the feature space. The parameter k of the classifier was obtained performing a sweep between 1 and 11 neighbors, then, the value which gave the best performance was selected. The results of the classification stage are shown in Table I.

The results in Table I show that the quadratic time-frequency distribution SPWVD is more adequate to

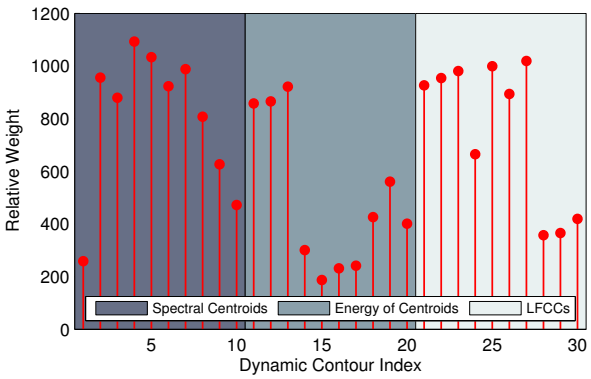


Fig. 2. Relevance of dynamic contours

TABLE I
PERFORMANCE OF CLASSIFIER – ACCURACY

	database - 8928		database - 4000	
	Mean	Std	Mean	Std
SPEC	86.44%	5.91%	92.66%	2.21%
SPWVD	89.02%	4.27%	92.67%	2.49%

represent HRV signals with the aim of detecting OSA, evidencing an improvement of 2.5 percentage points over the spectrogram, using 8928 observations to perform the cross-validation. On the other hand, when the test is performed with the best 4000 observations, the accuracy obtained with both representations is similar.

IV. DISCUSSION AND CONCLUSIONS

In [11] is reported an accuracy of 92.6% using ECG morphology and visual classification; while in [12] it is obtained an accuracy of 92.3% using both visual and automatic classification with wavelet transform and ECG morphology; in addition, in [13] is reported an accuracy of 89.4% with an automatic classifier, using frequency analysis and amplitude of R wave. The method developed in this work presents comparable results with respect to the aforementioned methods, with a complete automatic diagnosis system and using features derived from TFR.

The enhancement of the TFR resolution by means of quadratic energy distributions, such as generalized bilinear class, definitely leads to an improvement of the representation capabilities that determine the details of the HRV signal, and hence to a better detection of OSA (as basic signs of the pathological changes to be identified).

An important issue is the selection of the dynamic features. As commonly known in the state of the art, it was proved that a parameterization framework that efficiently combines magnitude and frequency information from the power spectrum is more suitable for this purpose. In any case, the choice of the number of time-varying contours (related to any of the multivariate dynamic features) for proper representation of HRV signals must be considered. A large number of contours does not always lead to a better classification performance.

V. ACKNOWLEDGMENTS

This study was supported within the framework of "Jóvenes Investigadores" program and "Centro de Investigación e Innovación de Excelencia ARTICA" financed by COLCIENCIAS; and DIMA "Detección de los niveles de compromiso de resonancia en niños con labio y/o paladar hendido" grant.

REFERENCES

[1] T. D. Bradley and J. S. Floras, "Obstructive sleep apnoea and its cardiovascular consequences." *Lancet*, vol. 373, no. 9657, pp. 82–93, Jan 2009.

[2] W. Xu and Z. Pan, "Definition of sleep apnea event by one minute hrv spectrum analysis," in *Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on*, 2008, pp. 2292–2294.

[3] T. Penzel, "Is heart rate variability the simple solution to diagnose sleep apnoea?" *Eur Respir J*, vol. 22, no. 6, pp. 870–871, Dec 2003.

[4] A. Quiceno, E. Delgado, M. Vallverdú, A. Matijasevic, and G. Castellanos-Domínguez, "Effective phonocardiogram segmentation using nonlinear dynamic analysis and high-frequency decomposition," Sept. 2008, pp. 161–164.

[5] M. Mendez, A. Bianchi, O. Villantieri, and S. Cerutti, "Assessment of the heart rate variability during arousal from sleep by cohen's class time-frequency distributions," in *Proceedings 11th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2007*, 2007, pp. 30–33.

[6] L. Cohen, "The uncertainty principle in signal analysis," in *Proc. IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, 25–28 Oct. 1994, pp. 182–185.

[7] B. P. Marchant, "Time-frequency analysis for biosystems engineering," *Biosystems Engineering*, vol. 85, no. 3, pp. 261–281, July 2003.

[8] A. G. Ravelo, C. M. Travieso, F. D. Lorenzo, J. L. Navarro, S. Martín, J. B. Alonso, and M. A. Ferrer, "Application of support vector machines and gaussian mixture models for the detection of obstructive sleep apnoea based on the rr series," in *Proc. of the 8th WSEAS International Conference on Applied Mathematics*. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 2005, pp. 139–143.

[9] B. Gajic and K. Paliwal, "Robust feature extraction using subband spectral centroid histograms," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01)*, vol. 1, 2001, pp. 85–88 vol.1.

[10] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter, "The apnea-ecg database," in *Proc. Computers in Cardiology 2000*, 2000, pp. 255–258.

[11] J. N. McNames and A. M. Fraser, "Obstructive sleep apnea classification based on spectrogram patterns in the electrocardiogram," in *Proc. Computers in Cardiology 2000*, 2000, pp. 749–752.

[12] B. Raymond, R. M. Cayton, R. A. Bates, and M. Chappell, "Screening for obstructive sleep apnoea based on the electrocardiogram-the computers in cardiology challenge," in *Proc. Computers in Cardiology 2000*, 2000, pp. 267–270.

[13] P. de Chazal, C. Heneghan, E. Sheridan, R. Reilly, P. Nolan, and M. O'Malley, "Automatic classification of sleep apnea epochs using the electrocardiogram," in *Proc. Computers in Cardiology 2000*, 24–27 Sept. 2000, pp. 745–748.