Automatic screening of Obstructive Sleep Apnea from the ECG based on Empirical Mode Decomposition and Wavelet Analysis

J. Corthout, S. Van Huffel, M.O. Mendez, A.M. Bianchi, T. Penzel and S. Cerutti

Abstract— **This study proposes three different methods to evaluate Obstructive Sleep Apnea (OSA) during sleep time solely based on the ECG signal. OSA is a common sleep disorder produced by repetitive occlusions of the upper airways, which produces a characteristic pattern on the ECG. Extraction of ECG characteristics as the heart rate variability and the QRS peak area offer alternative measures for cheap, non-invasive and reliable pre-diagnosis of sleep apnea. 50 of the 70 recordings from the database of the Computers in Cardiology Challenge 2000, freely available on Physionet, have been used in this analysis, subdivided in a training and a testing set. We investigated the possibilities concerning the use of the recently proposed method Empirical Mode Decomposition in this application and compared it with the established Wavelet Analysis. From the results of these decompositions the eventual features were extracted, complemented with a series of standard HRV time domain measures and three extra non-linear measures. Of all features smoothed versions were calculated. From the obtained feature set, the best performing feature subset was used as the input of a Linear Discriminant Classifier. In this way we were able to classify the signal on a minute-by-minute basis as apneic or non-apneic with an accuracy of around 90% and to perfectly separate between apneic and normal patients, using around 20 to 40 features and with the possibility to do this in three alternative ways.**

I. INTRODUCTION

Sleep related breathing disorders have a high prevalence in the adult population. Epidemiological studies indicate a high prevalence of 4% in males and 2% in females in the general population [1]. Obstructive sleep apnea (OSA) (Greek: $\alpha\pi\nu o\iota\alpha$, from α-, privative, $\pi\nu\varepsilon\varepsilon\iota\nu$, to breathe), the most common of the different types of sleep-related breathing disorders with about 84% of the cases [2], is characterized by repetitive cessations of respiratory flow during sleep.

The frequent arousals that terminate apneas lead to sleep fragmentation. Sleep fragmentation is characterized by disruption or loss of deep and REM sleep. Consequently sleep loses its restorative function on physical and mental performance. Patients suffering from sleep apnea report excessive daytime drowsiness and non-restorative sleep even when the sleep period is prolonged, while not realizing themselves that they are suffering from it. Besides significant social and emotional problems caused by poor mental performance, patients can also suffer from a series of physiological problems. Frequent findings are obesity, hypertension, and arrhythmia. Hypertension and arrhythmia are often precursors of heart failure, for which a sleep apnea patient has a significantly higher probability of occurrence.

Sleep studies are expensive because they require overnight evaluation in sleep laboratories with dedicated systems and attending personnel. Due to the cost and relative scarcity of diagnostic sleep laboratories, it is estimated that sleep apnea is widely underdiagnosed [3]. Hence, techniques which provide a reliable diagnosis of sleep apnea with fewer and simpler measurements without the need for a specialized sleep laboratory can surely be of benefit. An interesting signal to investigate is the Electrocardiogram (ECG). The importance of the ECG lies in the fact that it can be easily measured, in a non-invasive way and with a high signal-to-noise power ratio. It measures the electrical activity of the heart and has a close relationship with the activity of the Autonomic Nervous System (ANS).

This is why in this study the possibilities have been investigated to provide a reliable diagnostic of OSA based solely on the measurement of the ECG, now using primarily a fairly recent technique: the Empirical Mode Decomposition (EMD). With this EMD any complicated data set can be decomposed into a finite and often small number of components, also known as intrinsic mode functions (IMFs), that are designed so as to admit well-behaved Hilbert transforms (HT). The decomposition method is adaptive, and, therefore, highly efficient. Since the decomposition method is based on the local characteristic time scale of the data, it is applicable to non-linear and non-stationary processes. This decomposition can be linear or non-linear as dictated by the data, and it is complete and almost orthogonal. With the HT the IMFs yield instantaneous frequencies as functions of time that give sharp identifications of embedded structures [4].

The use of the EMD in this application had not yet been thoroughly investigated before. Consequently three alternative methods are presented for apnea screening starting from the ECG: one based on EMD and HT, one based on EMD and reassigned spectrogram (RAS) and one based on wavelet analysis (WA) and HT, the latter to allow comparison with a more commonly used decomposition.

J. Corthout is student with the Dept. of Electrical Engineering - ESAT, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium jeroen.corthout@telenet.be.

S. Van Huffel is Professor with the Dept. of Electrical Engineering - ESAT, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium sabine.vanhuffel@esat.kuleuven.be.

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

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

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

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

II. USED MATERIAL AND METHODOLOGIES

A. Database Description

The used database has been made freely available via PhysioNet, initially to support the Computers in Cardiology Challenge 2000. Data was collected by the sleep laboratory at the Philipps Universität in Marburg, Germany. Thirty-five of the seventy ECG recordings were provided together with apnea annotations, and eight of these were accompanied by respiration and oxygen saturation signals. The apnea scoring was carried out based on standard criteria by one sleep expert on a minute-by-minute resolution for these recordings. For the other 35 recordings, apnea annotations were withheld for the duration of the challenge [5]. Recordings were subdivided in three groups: Apneic patients (class A), Borderline patients (class B) and Control or normal patients (class C).

Both the provided training and test set contain 20 apneic, 5 borderline and 10 normal subjects. From the released training set we randomly selected 25 subjects to train our classification algorithm. A second group with 25 recordings chosen randomly from the released test set, was used to measure the performance of our algorithm. We selected only 50 recordings in which we manually corrected the QRS peak detection, i.e. the detection of the heart beat. Future studies would preferably include the whole database.

B. Used methodologies

Based on previous knowledge regarding the physiological effects of OSA on the ECG signal, we extracted signals like RR intervals between subsequent QRS peaks, further called the RR signal, and QRS peak area, further called the AR signal. The RR signal shows a characteristic oscillation (brady-tachycardia) during periods with apneic events, producing a very low frequency apnea respiration peak in the signal spectrum. The same frequency component can be found in the AR signal. During periods without apneic events a high frequency peak coupled with respiration activity is being observed. From both signals after preprocessing an EMD has been calculated, then a HT or a RAS, from which the eventual features are extracted, complemented with a series of standard heart rate variability (HRV) time domain measures and three extra non-linear measures. Of all features smoothed versions were calculated. From the obtained feature set, the best performing feature subset was used as the input of a Linear Discriminant Classifier. To allow comparison with established methods, this whole process has been repeated using a wavelet analysis instead of an EMD. An illustration of both decompositions applied to an apneic segment of a random RR signal is presented in Fig. 1.

C. Manual RR signal correction and ECG-Derived respiratory signal (AR signal)

In the provided database for each ECG signal a series of automatically detected QRS complexes is included. We used our own automatic algorithm to search the QRS peaks in the ECG signal and later plotted these detections together with the ECG to allow manual correction of the misdetected beats. After correction, the RR signal could be computed more accurately. Also the AR signal could be calculated by subtracting the baseline from the original ECG, detecting the minimum value within 100*ms* before and after the maximum value of the R peak and integrating the area under the R peak in the interval between these two points. The baseline itself was calculated using a median filter of 200*ms* width.

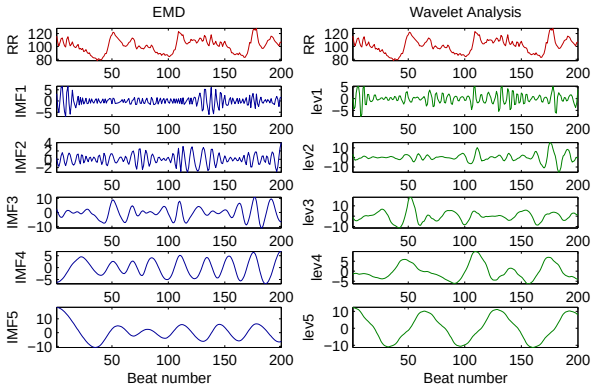


Fig. 1. Empirical Mode Decomposition (EMD) and Wavelet Analysis (WA) of an apneic RR signal segment.

D. Automatic preprocessing: cleaning the signals

Based on an estimate of the RR signal calculated with a median filter of width 5 beats, spurious, misdetected and extrasystolic beats were detected and corrected. The first two kinds of inconvenient beats were removed as described before in the paper of de Chazal *et al.* 2003 [6]. We also added a method to remove the extrasystolic beats, by detecting the specific down-up (sometimes up-down) pattern in the RR signal and changing the value to the calculated estimate. The changes in the RR signal were then reflected in the AR signal and a removal of suddenly excessively high values in the AR signal was performed. With this cleaning step the disturbing effects on the higher frequency components of the EMD due to sudden deviations from the normal value levels were minimized, while preserving the characteristics of and the information in the signals as good as possible with reasonable effort.

E. Empirical Mode Decomposition with HT or RAS

The essence of the EMD is obtaining IMFs that obey the necessary conditions for a meaningful HT calculation. An IMF is calculated by subtracting from the original signal $x(t)$ the mean m of the upper and lower envelopes, obtained by a cubic spline interpolation through the local maxima and minima of the signal $x(t)$ respectively:

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

This process is further repeated on C until eventually an IMF is obtained. This IMF is then subtracted from $x(t)$, leaving a residue R:

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

The described procedure can be applied to R over and over again to obtain a set of IMFs, decomposing the signal until the remaining R is no more than a monotonous signal [4].

As the first obtained IMFs exhibited very broadband frequency information in the higher frequencies, which could be problematic in a subsequent HT, as the signal information is reduced to one instantaneous amplitude and also only one instantaneous frequency, furthermore calculation of a reassigned spectrogram was investigated. This method tries to correct the intrinsic blurring in a spectrogram caused by the limited size of the analysis window by reassigning the symmetrical energy distributions due to the blurring back to their original positions, which is supposed to be the center of gravity of the local energy distributions [7].

F. Wavelet Analysis

To allow comparison with an established decomposition method an exactly analogous method was implemented using WA to subdivide the signal in components. Due to its orthogonality and observed resemblance to the HRV oscillations, the discrete Daubechies4 wavelet basis has been chosen [8].

G. Feature set

From the obtained EMD the first five IMFs were retained as they contained the useful information. On a minute-by-minute basis means and standard deviations of ratios between instantaneous amplitudes and frequencies were calculated for the method with HT, and means and standard deviations of self defined band contributions were calculated for the method with RAS for RR and AR signal. The defined bands were: Very Low Frequency (VLF) $0.01-0.08Hz$, Low Frequency (LF) $0.08-0.15Hz$ and High Frequency (HF) $0.15-0.4Hz$. Correlation measures between RR and AR were obtained by either multiplying or dividing corresponding features.

These features were complemented with eight additional feature types. Five features from the standard selected time domain measures [9] were calculated: RMSSD, SDSD, pNN50, HRV triangular index and TINN. Also three extra non-linear measures: form factor, sample entropy and Kolmogorov complexity.

To easily and effectively include information from a larger time scale, also from every feature a smoothed version was obtained.

H. Subset and classifier selection

First a broad set of features was formed, then we defined a limited subset containing the best performing features to minimize computation time and boost classification performance. A best-first search wrapper method was used, which forms a supposedly optimally accurate feature subset by repeatedly expanding the present feature subset with the most promising non-included feature to increase classification accuracy, starting from an empty one. For the evaluation of this classification accuracy both a direct evaluation and a subject-per-subject Leave One Out (LOO) method have been investigated. Direct evaluation means classifying the training set with a classifier built using the training set data itself, LOO means classifying each subject out of the training set with a classifier built using the data of the other subjects. The latter would normally have to provide better results as it includes variability between subjects, but it seemed that this method suffers from overfitting. Also including the smoothed feature versions together with the original ones in the wrapper method appeared to suffer from overfitting. Consequently the wrapper method was performed by direct evaluation on the feature set using only original features. The smoothed versions were added later to their corresponding original feature when the actual classification was performed.

Concerning the choice of the appropriate classifier type, linear, quadratic and Mahalanobis discriminant classifiers were investigated. The linear discriminant (LD) classifier was selected as it outperformed the others in terms of accuracy and even more significantly in terms of sensitivity. The estimates of classification performance were obtained through LOO cross-validation.

TABLE I

CLASSIFICATION PERFORMANCE ON THE TEST SET FOR ALL METHODS

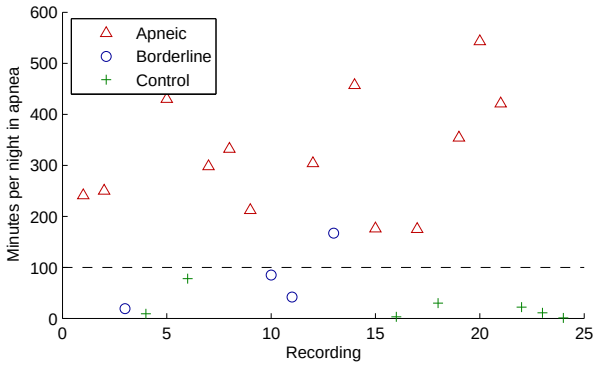
	EMD and HT	EMD and RAS	WA and HT
Used features	20	20	40
Sensitivity	0.8056	0.8182	0.8492
Specificity	0.9444	0.9439	0.9386
Accuracy	0.8888	0.8945	0.9052

III. RESULTS

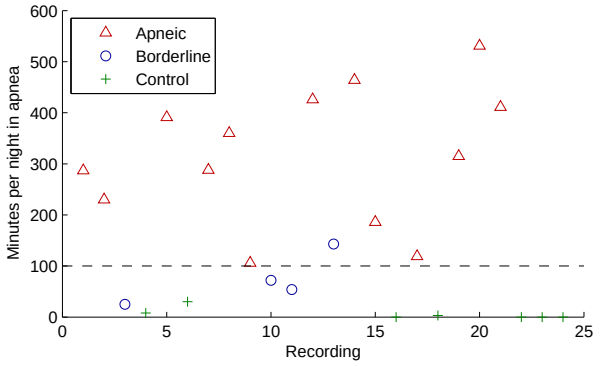
Using features extracted from the respective decompositions, some time domain and non-linear measures, and after having complemented all these features with a smoothed version, all three used methods provided us with classification accuracies around 90% when using a LD classifier on a selected subset. The classification performance measures are presented in Tab. I. Not much difference can be noted between the results obtained with EMD and HT and EMD and RAS; WA and HT seemed to have a sensitivity that outperformed the others a bit, however using the double amount of features. Apneic and normal/control subjects could be separated flawlessly, as shown in Fig. 2 for the methods with EMD and HT, and WA and HT. Borderline patients still couldn't be classified completely. Similar results are obtained using EMD and wavelet analysis.

IV. DISCUSSION

The goal of this study was to develop a method to screen obstructive sleep apnea in a cheap, non-invasive and reliable way and more precisely to investigate the possibilities concerning the use of the recently proposed method EMD on the data obtained in this application from the ECG signal. In addition, one sleep apnea screening method based on EMD and reassigned spectrogram and one based on wavelet analysis and Hilbert transform were proposed. The presented methods produce a minute-by-minute classification with an accuracy equal to the best performing classification algorithms published before, using a generally smaller set of features, while moreover three different alternatives are presented to obtain these results.



(a) EMD and HT



(b) WA and HT

Fig. 2. Subject classification using the methods with Empirical Mode Decomposition (EMD) and Hilbert Transform (HT) and Wavelet Analysis (WA) and HT. Note that applying a threshold of 100 minutes per night, apneic and normal/control patients are separated flawlessly.

The most important feature information in all cases had to be sought in the lower frequencies of both RR and AR signal, where the apnea repetition peak is present: in IMF5 and IMF4 for the method with HT, in the VLF band for the method with RAS. This was always complemented with the HRV triangular index of the RR signal, probably also changing due to the apnea repetition peak in the RR spectrum, and the pNN50 measure calculated on the AR signal. This feature is in fact designed for an RR signal but indicated HF information content in the AR signal, probably due to the respiration peak in the AR spectrum. This respiration component in both RR and AR spectrum during periods without apneic events was also expressed in a multitude of other included features. For WA the information from the HF component was even the first feature, on the other hand for the EMD it was not included in the feature subset. The calculated non-linear features didn't play any important role in the classification.

The three presented methods seemed to lead to very similar classification performance using quite similar information. The exact nature of the respective frequency separations did not seem to matter that much in the present application, as long as the changes of the spectra concerning apnea repetition and respiration could be quantified. Nevertheless the slightly higher sensitivity of the spectrogram method raised the accuracy of this method a little bit with respect to the method with EMD and HT. But as the computation time needed for calculating the reassignment of the spectrogram is much longer, one may still prefer the method with HT to guarantee a good usability of the algorithm. Also this method seems to make more sense conceptually, as the EMD is primarily conceived to ensure meaningful HT calculation. The method with WA and HT even performed a bit better, but still comparable to the method with EMD and HT.

V. CONCLUSIONS

The obtained results, with comparable accuracies around 90%, showed that apnea screening can be done in a very reliable way using only data obtainable from the ECG signal, and this in multiple alternative ways. Moreover apneic and normal/control subjects could be separated flawlessly. These results further open the way to fast home diagnosis for sleep apnea patients, the goal of the earlier Computers in Cardiology Challenge 2000. Though sounding like a very promising method, EMD did not seem to outperform WA in this application, while the latter even performed slightly better. The VLF information concerning the apnea repetition was one of the main features for the both methods. The importance of the HF component in WA however could indicate a significant difference with EMD, where the information from the corresponding component is not even included in the feature subset. Nevertheless the possible usefulness of the EMD in HRV analysis has been shown and it could be interesting to investigate its use in more different applications.

REFERENCES

[1] Young T., Palta M., Dempsey J., Skatrud J., Weber S., and Badr S., "The Occurrence of Sleep-Disordered Breathing among Middle-Aged Adults", *in New England Journal of Medicine*, Vol. 328 (1993), pp. 1230-1235.

[2] Morgenthaler T.I., Kagramanov V., Hanak V., and Decker P.A., "Complex sleep apnea syndrome: is it a unique clinical syndrome?", *in Sleep*, Vol. 29 (2006), pp. 1203-1209.

[3] Young T., Evans L., Finn L., and Palta M., "Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women", *in Sleep*, Vol. 20 (1997), pp. 705-706.

[4] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", *in Proceedings of the Royal Society*, Vol. 454 (1998), pp. 903-995.

[5] Penzel T., Moody G.B., Mark R.G, Goldberger A.L., and Peter J.H., "The Apnea-ECG Database", *in Computers in Cardiology 2000*, Vol. 27 (2000), pp. 255-258.

[6] de Chazal P., Heneghan C., Sheridan E., Reilly R., Nolan P., and O'Malley M., "Automated Processing of the Single-Lead Electrocardiogram for the Detection of Obstructive Sleep Apnoea", *in IEEE Transactions on Biomedical Engineering*, Vol. 50 (2003), pp. 686-696.

[7] Auger F. and Flandrin P., "The why and how of time-frequency reassignment", *in IEEE International Symposium on Time-Frequency and Time-Scale Analysis*, 1994, pp. 197-200.

[8] Torrence C. and Compo G.P., "A Practical Guide to Wavelet Analysis", *in Bulletin of the American Meteorological Society*, Vol. 79 (1998), pp. 61-78.

[9] Task Force of the European Society of Cardiology and the North American Society of Pacing Electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use", *in Circulation*, Vol. 93 (1996), pp. 1043-1065.