Recognition and Quantification of Sleep Apnea by Analysis of Heart Rate Variability Parameters

C Maier, M Bauch, H Dickhaus

University of Heidelberg / University of Applied Sciences Heilbronn, Heilbronn, Germany

Abstract

This study was performed within the scope of the CinC-2000 challenge on detection and quantification of obstructive sleep apnea from single channel ECGs. Established statistical time domain heart rate variability (HRV) measures as well as parameters based on time delay embedding and correlation analysis are investigated for their diagnostic quality by means of receiver operating characteristics (ROC) analysis on a training set of 35 ECGs and cross validated on an independent tests set of equal size. Moreover, several feature combinations are evaluated with a second order polynomial classifier. The results indicate, that at least most of the information necessary to recognize sleep apnea is contained in the ECG. Recognition rates up to 93% for screening of apnea patients and 85.55% for minute by minute quantification are achieved.

1. Introduction

Obstructive sleep apnea (OSA) is a common health concern with an estimated prevalence of about 4 % in middle-aged males. It is associated with a wide range of health implications and increased cardiovascular morbidity and mortality. The gold standard in diagnosing OSA is polysomnography, an inconvenient, expensive and time consuming procedure which includes an overnight multi channel recording of respiratory and other vital parameters in a specialized sleep laboratory. To reduce the number of polysomnographies in patients without or with mild OSA, an effective and inexpensive screening with both, high sensitivity and specificity is highly desirable.

The CinC Challenge 2000 aims at answering the question whether screening as well as quantification of OSA is possible based on information available from the electrocardiogram (ECG) alone. Many studies in recent years have hinted at this possibility. As soon as 1984, Guilleminault [1] suggested a HRV pattern - 'cyclic variation of heart rate' (CVHR) - as screening tool for OSA. It consists of a phase of bradycardia followed by abrupt tachycardia, both mediated via the autonomic nervous system (ANS).

The aim of this study is to assess the suitability of several established time domain HRV measures, which are known to reflect ANS control [2], and other HRV parameters based on time delay embedding and correlation analysis for screening and minute by minute quantification of OSA. In a first step, the quality of single features is investigated by means of ROC analysis and the best results are cross-validated by thresholding on an independent tests set. Moreover, several combinations of features are evaluated with a second order polynomial classifier.

2. Material and methods

The data set under investigation consists of 70 ECG signals (1 channel) of ca 8 h duration, recorded overnight with a sampling frequency of 100 Hz and amplitude resolution of 12 bits. It is divided into two groups (training set and tests set) each containing records of 20 patients suffering from OSA, 5 borderline cases and 10 control probands. The record of a patient contains at least one hour with an apnea index of 10 or more, and at least 100 minutes with apnea during the recording. Controls have fewer than 5 minutes with apnea. All apneas are either obstructive or mixed. Hypopneas are also counted as apneas. More details can be found in [3].

In the training set, file by file information on the proband's status (patient/borderline/control) as well as minute by minute annotations on the occurrence of apnea at the beginning of this minute are available. The annotations were made by human experts on the basis of simultaneously recorded respiration signals. For the data in the tests set, no further information is given.

The goal within the apnea screening task (AST) is the correct identification of the 20 patients and 10 controls in the tests set, (the 5 borderline cases are neglected). With respect to quantification of sleep apnea, a minute by minute classification of the data in the tests set is attempted.

All parameters investigated within the scope of this study quantify heart rate variability and are calculated from the sequence of RR intervals of the ECG. In order to increase the time-resolution of the original data, the ECG signal is first interpolated using cubic splines and then resampled with 1000 Hz. A median highpass filter (width 501 ms) to reduce baseline wander is then applied. After R peak detection, a classification of QRS-

0276-6547/00 $10.00 © 2000 IEEE

Computers in Cardiology 2000;27:741-744.

morphology [4] and -timing is performed to identify artefacts and ectopic beats and exclude them from further processing. Gaps in the RR-sequence resulting from rejected or missing beats are interpolated by means of a nonlinear algorithm described in [5]

Data Analysis is performed on two different timescales (figure 1) In the apnea quantification task (AQT), successive segments of one minute in duration are constructed from the corrected series of RR intervals, and for each of this segments, one feature value is calculated. A smoothing median filter (width 13) to reduce the temporal variability of the results of adjacent segments is finally applied.

In the AST, two different strategies are evaluated: Firstly, feature values are calculated from all data available for a patient i.e. the segment length is the total signal duration, and secondly, the median of the results obtained from the analysis of the one minute segments is used. In both cases we get one single number per patient.

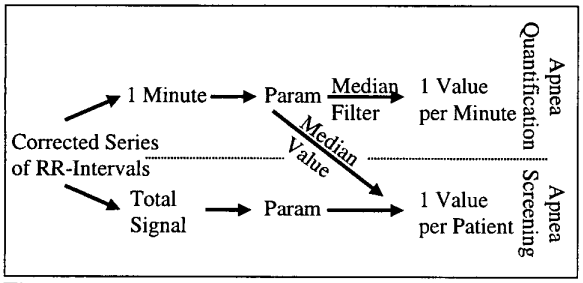

Figure 1: Data processing overview

A part of the parameters under investigation are statistical time domain measures commonly used in HRV analysis [2], such as the standard deviation (SD) of all RR intervals between successive beats of normal origin (NN intervals), (SDNN), the absolute (NN50count) and relative (pNN50) number of successive pairs of NN-intervals that differ more than 50 ms and the square root of the mean of the summed squares of differences between adjacent NN-intervals (RMSSD). For the ADT, moreover the SD of the mean of the NN intervals in all 5-minute segments of the recording (SDANN) and the mean of the SD of all NN intervals for all consecutive 5-minute segments (SDNN index) were calculated.

In addition to those parameters, we investigated two features that have been proposed in EEG processing for brain-computer interfacing [6] and - to the best of our knowledge – are not commonly used in HRV analysis.

Both features are derived from the time delay embedded corrected series of RR intervals. Given that the time segment of analysis contains N RR-intervals x_i (i=1..N), embedding vectors $\vec{x}_i$ of the dimension D are constructed from values x_i that are spaced t RR intervals apart:

$$\vec{x}_i = \left(x_i \quad x_{i+t} \quad \cdots \quad x_{i+(D-1)\cdot t}\right)^T$$

In [6], the vectors $\vec{x}_i$ directly form the columns of the embedding matrix X. In our realisation, we first calculate the mean vector $\vec{m}$ of all embedding vectors $\vec{x}_i$

$$\vec{m} = \frac{1}{N-(D-1)\cdot t}\sum_{i=1}^{N-(D-1)\cdot t}\vec{x}_i$$

and subtract it from each vecor $\vec{x}_i$ prior to the aggregation. So, the embedding matrix X is calculated according to

$$X = \left[\vec{x}_1-\vec{m} \quad \vec{x}_2-\vec{m} \quad \cdots \quad \vec{x}_{N-(D-1)\cdot t}-\vec{m}\right]$$

The sorted eigenvalues l_i of the DxD-Matrix $X \cdot X^T$ are the basis for our parameters

$$l_1 \cdots l_D = Eigenvalues\left(X \cdot X^T\right)$$
$$\text{where} \quad l_i > l_{i+1} \qquad \text{for } i = 1 \cdots D-1$$

The magnitude of each eigenvalue is normalized with respect to the sum of all eigenvalues:

$$\lambda_i = \frac{l_i}{\sum_{i=1}^{D} l_i}$$

and the normalized maximal eigenvalue (NME) serves as classification feature:

$$NME = \lambda_1$$

Since, up to a multiplicative constant, the matrix $X \cdot X^T$ is identical to the covariance matrix of the vecors $\vec{x}_i$, NME reflects the extension of the cluster of the embedded RR series in the direction of its largest extension relative to its 'size' in the directions of the other eigenvectors.

The second parameter is derived from the Entropy H of the embedding space eigenspectrum

$$H = -\sum_{i=1}^{D}\lambda_i \cdot ld(\lambda_i)$$

It is calculated following [6] as

$$EBF = 2^H$$

and quantifies the stochastic 'complexity' (a very badly defined term) of the underlying time series.

In this study, the values for the embedding Dimension D and time delay t were empirically set to 3. It must be noted, that the resulting numbers of NME and EBF do not reflect 'true values' in the sense of the theory of nonlinear dynamics, where an embedding dimension D sufficiently high for the underlying attractor must be guaranteed. Rather, they describe spatial properties of the cluster formed by the embedding vectors. For classification purposes, the most important question is, whether these values have different distributions in the case of apnea segments and non apnea segments, regardless of whether the values are correct in a theoretical sense.

Table 1: Classification rates for apnea screening (single features). In the 'Thresh' column, > indicates higher values for OSA, < indicates lower values.

	Analysis of total signal			Median of minute segments			
	Training set (without b01-b05)			Training set (without b01-b05)			Tests set
Parameter	Sens	Spec	Thresh	Sens	Spec	Thresh	Total
NME	85	70	> 0.75	95	100	> 0.51	28/30
EBF	75	80	< 1.59	95	100	< 1.98	27/30
CBF				95	100	> 5.58	28/30
pNN50	90	70	< 0.23	90	70	<0.224	
NN50count	75	70	< 5600	80	70	< 13	
SDNN	60	70	< 85.22	55	60	< 53.9	
SDSD	70	70	< 48.7	75	70	< 38.7	
RMSSD	70	70	< 48.9	75	70	< 38.4	
SDANN	60	80	< 48.3				
SDNN-index	60	70	< 72.78				

The last HRV parameter investigated in this study is a correlation based feature (CBF). It is calculated within RR interval segments of 5 minute duration, which are shifted in increments of 1 minute over the whole signal.

From each 5 minute segment, the central window of one minute duration is extracted and cross correlated with the surrounding 5 minute segment. The sum of all normalized correlation values that exceed the threshold 0.45 yield the value of CBF. It aims to identify the cyclical variation of heart rate described in [1].

To assess the quality of the calculated features with respect to the classification task, ROC curves were generated for each measure by plotting sensitivity against (1-specificity) for all possible decision thresholds.

Moreover, different features were combined and the training set served to train a second order polynomial classifier which was used to reclassify the training set. For the best results, a validation was performed on the tests set. In the AQT, a smoothing of the classification results with a median filter was performed additionally

3. Results

The calculation of ROC curves is only possible for data of the training set, since the results for the tests set, received from CinC, contained for obvious reasons only the total number of correct classifications but no information on whether correctly positive or correctly negative. From the ROC plots, the threshold corresponding to the point of the curve closest to the upper right corner (0,1) was considered as value that achieves best separation between the two groups. Only for the best features in the training set, the results on the tests set were submitted to CinC because only a limited number of submissions was allowed for.

Within the AST, the resulting sensitivity and specificity for the different parameters are given in table 1. The left hand side shows the values for the analysis of the total signal duration, the right hand side the results when the median of the results in 1 minute segments was used. Since according to the scoring rules of the CinC Challenge, the classification of the five borderline cases (files b01-b05) did not influence the classification result, they were omitted from the learning set in the AST. Combination of two features did not improve the classification result. Because of the small number of only 30 samples in the training set, we abstained from using a higher dimensional feature space.

For the minute by minute classification in the AQT, the complete training set was used to assess the ROC

Table 2: Classification rates for apnea quantification

	Training set			Tests set
Parameter	Sens	Spez	Thresh	Total
NME	76.73	74.76	> 0.63	
EBF	81.33	72.05	> 1.85	
CBF	81.31	77.16	> 10.25	79.12
pNN50	70.74	37.36	> 0.018	
NN50count	70.76	37.36	> 1	
SDNN	68.57	58.93	> 50.99	
SDSD	52.99	48.17	< 30.93	
RMSSD	52.99	48.16	< 30.72	

Table 3: Classification rates for apnea quantification (selected feature combinations).

No	Combination	Training set		Tests set
		Sens	Spec	Total
1	CBF / NME	72.22	87.38	
2	CBF / EBF	74.11	84.30	
3	CBF / RMSSD	71.98	89.34	
4	NME / SDNN	74.96	87.09	
5	EBF / SDSD	77.54	80.65	
6	1 + SDNN	73.36	89.33	84.79
7	6 + EBF	75.95	88.28	
8	6 + SDSD	72.09	90.94	
9	8 + NN50count	72.12	91.23	85.55
10	7 + NN50count	76.35	88.92	

curves. The results obtained are given in table 2.

Table 3 shows the classification rates achieved with the second order polynomial classifier for several selected feature combinations.

4. Discussion and conclusion

The best results in the AST (95% sensitivity, 100% specificity on the training set, up to 28/30 on the tests set) are obtained from the features CBF, NME and EBF (table 1). Interestingly, these results are only achieved when the median minute by minute values are considered. Calculation over the whole signal duration decreases the performance considerably, because the higher regularity of the cyclic variations of heart rate during periods of apnea is blunted by other fluctuations on this time scale. From the established HRV measures, pNN50 yields the best results (90% sensitivity, 70% specificity). Generally, lower, less complex heart rate variability is found in apnea patients.

The best single parameter in the AQT was found to be CBF (table 2). With a sensitivity of 81,31%, comparable to that of EBF (81.33 %), it has better specificity (77.16%) than all other features and yields on average 79.12% correct classification on the tests set. Slightly worse results are obtained from the embedding based features, nevertheless their superiority to the established time domain HRV measures is clearly visible. Obviously, the comparatively regular structure of the RR intervals during apnea phases [1] is better captured by these features. Especially CBF has the advantage that its magnitude is only based on similarity of the RR intervals on a short timescale (5 min) and therefore allows for variability of the CVHR pattern even in the same patient, largely independent from its amplitude and frequency. The same would be expected for NME and EBF, however only in the limit of a long data sequence and an embedding dimension sufficiently high. The higher specificity of CBF can be explained from the fact that it only detects sequences of several CVHR swings i.e. more pronounced apnea.

It is interesting to note, that for the established HRV measures, the time scale of feature calculation has great influence on the 'sign' of the threshold decision: In table 1, the OSA patients have lower values of SDNN whereas on a minute by minute basis (table 2), SDNN turned out to be even more reduced in OSA patients when apnea was absent, but relatively elevated during phases of apnea. From visual inspection, a quantification of the apnea phases using SDNN seems feasible within one patient, however the high inter-patient variability and the fact that SDNN is generally higher in healthy persons does not allow to use a fixed threshold.

The combination of several features allows to further improve the results (table 3). Using three features – CBF, NME and SDNN – an average classification rate of 84.79% was achieved on the tests set, rising to 85,5 % for five features. Combining improved mainly specificity, probably reflecting the higher prevalence of non apnea phases in the training set.

Generally, the temporal smoothing of the minute by minute values and classification results by means of a median filter yielded a considerable improvement of the classification rates. Best results were achieved using a width between 9 and 15, indicating that apnea phases often extend over several minutes. The filter successfully suppressed spurious short term transgressions of the decision threshold.

Further improvements may be expected from combination of screening and quantification i.e. by attempting a quantification only on patients with a positive screening result. This may also lead to an improvement of quantification sensitivity which is probably too low (< 80%) for practical use. Possibly, this results from missed phases of hypopnea.

Since less is known about the patients in the data set, all conclusions made from the results obtained in this study must be taken with a grain of salt. However, especially in respect to screening of patients suffering from OSA, the results obtained are very promising and indicate, that much – if not all – of the information necessary to diagnose sleep apnea is contained in the ECG signal. Future work will have to include an investigation of stability and specificity in presence of other diseases known to affect the ANS as well as different sleep stages.

References

[1] Guilleminault C, Tilkian A, Dement WC. Cyclical variation of the heart rate in sleep apnoea syndrome. Mechanisms and usefulness of 24hr electrocardiography as a screening technique. Lancet 1984;1:126-131.

[2] Task Force of the European Society of Cardiology and the North Amerian Society of Pacing and Electrophysiology: Heart Rate Variability: Standards of measurement, physiological interpretation, and clinical use. Circulation 1996;93:1043-1065.

[3] www.physionet.org/cinc-challenge-2000.shtml

[4] Maier C, Dickhaus H, Gittinger J: Unsupervised morphological classification of QRS complexes. In: Computers in Cardiology 1999. IEEE Computer Society Press, 1999;26:683-686.

[5] Lippman N, Stein KM, Lerman B. Comparison of methods for removal of ectopy in measurement of heart rate variability. Am. J. Physiol. 1994;267: H411-H418.

[6] Roberts SJ, Penny W, Rezek I. Temporal and spatial complexity measures for electroencephalogram based brain-computer interfacing. Med. Biol. Eng. Comput. 1999;37:93-98.

Address for correspondence.

christoph.maier@fh-heilbronn.de