Amplitude Variability Extraction from Multi-Lead Electrocardiograms for Improvement of Sleep Apnea Recognition

C Maier[1], H Dickhaus[1,2], P Laguna[3]

[1]Heilbronn University, Heilbronn, Germany
[2]University of Heidelberg, Heidelberg, Germany
[3]University of Zaragoza, Zaragoza, Spain

Abstract

This study is in the context of sleep apnea recognition from multi-lead ECGs. In 38 patients, 8-channel ECGs were recorded simultaneously to a polysomography (PSG). The ECG was classified in segments of one minute for occurrence of sleep apnea events by quantification of the regularity of characteristic oscillations in either heart rate or ECG amplitude. Diagnostic accuracy is compared by ROC-analysis against the expert annotations of the PSG, and its reproducibility was tested on the Physionet apnea ECG database. Whereas amplitude variations yield consistent results on both data sets (ROC-area 89.0% vs. 88.3%), a remarkable loss in performance is observed for heart rhythm (89.8% vs. 77.9%). Reasons for this difference are discussed and it is shown that factors like diabetes have a confounding influence on heart rate. With respect to aggregation of multi-lead information, simple averaging (89.3%) seems to be as appropriate as more complex PCA-based methods (87.2%).

1. Introduction

In the CinC Challenge 2000 on recognition of sleep apnea from the ECG, participants mainly focussed on two well-known ECG manifestations of apnea [1]: First, the cyclic variation of heart rate (CVHR) [2], a regular pattern of alternating phases of bradycardia and tachycardia mediated via the autonomous nervous system (ANS). And second, the modulation of ECG amplitude by respiratory activity [3]. However, the authors often combined both approaches into one classifier so that the individual merits were hard to judge. Moreover, only a single ECG channel sampled at 100 Hz was available, and little was known about how representative the sample is.

In this context, this study addresses four questions: Is there a significant difference in apnea recognition accuracy from heart rate and ECG amplitude variations as sources of information or are they equivalent? How reproducible are the recognition rates on different data sets? What improvement is achievable when high resolution multi-lead ECG data is available, and how can amplitude variations from multiple leads best be aggregated into one single series?

2. Methods

The ECG signals used in this study are 38 8-channel (I, II, V1 to V6) holter recordings (Mortara H12+) registered with a sampling rate of 1 kHz during sleep in parallel to a full night polysomnography (Respironics Alice 4). After 50 Hz reduction, the baseline of the ECG was corrected to zero in the isoelectric interval prior to the QRS complex using a heart rate dependent highpass filter. QRS detection and classification was performed and carefully confirmed using self-developed software.

The dataset used for comparison is the Physionet apnea ECG database (AECGDB) which is available on the internet and has also been used in [1, 4].

The series of RR-intervals between consecutive beats served as basis for CVHR detection. For robust quantification of variations in amplitude, we use the mean absolute value (MAV) of the baseline corrected ECG signal, calculated for each lead i (i=1 to 8) in a 120 ms window, centred at the sample index θ_n of the nth QRS fiducial point:

$$m_i(n) = \frac{1}{121}\sum_{k=-60}^{+60}\left|ecg_i(\theta_n + k)\right| \quad (1)$$

This results in eight series $\{m_i(n)\}$. Ectopic and strongly corrupted beats were excluded from MAV-calculation; the corresponding values were later interpolated to keep time relations.

Fig. 1 gives an example for the MAV-series in leads II, and V2 during a phase of recurring severe apneas. Moreover, the corresponding traces of respiratory movements and flow as well as the series of RR-intervals are shown. In the latter, a pronounced CVHR pattern is easily detectable. The start of the reflex tachycardia coincides with the resumption of air flow in consequence of an arousal. Upon cessation of air flow due to absence of respiratory movements, bradycardia restarts and persists during the obstructed breathing cycles. In the

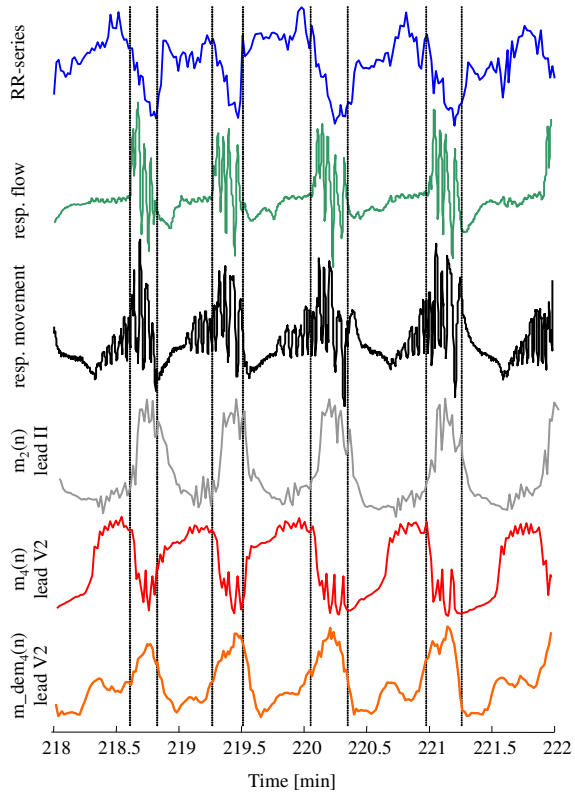

Figure 1. Time course of RR-series, respiratory flow, respiratory movements and series of MAV values of lead II and lead V2 during repetitive apneas. The lowest trace shows the result of the demodulation of MAV of lead V2. Vertical lines delimit ventilatory phases.

traces of the MAV series, two superimposed phenomena are visible. First, a higher frequency (HF) oscillation of comparatively low amplitude that clearly is related to respiratory excursions, and second, a pronounced low frequency (LF) pattern with the periodicity of single apnea events, coincident with the CVHR.

In both MAV-series, the magnitude of the respiratory movements is reflected in an amplitude modulation of the HF-oscillations which permits detection of the absence of activity during the initial central apnea and its presence in the ventilatory phase. Unfortunately, the optimal leads for detection seem to differ from patient to patient, and even within the same patient, dependent on body position. This similarly holds for the LF component.

Our approach for a condensed extraction from multi-lead ECGs consists in the following procedure (fig. 2): First, the MAV-series $\{m_i(n)\}$ of each lead is lowpass-filtered using a 2nd order Savitzky-Golay (SG) Filter of width 9 ms. The resulting series $\{m\_lp_i(n)\}$ is subtracted from the original one which yields the HF respiratory modulation. After rectification, these HF residuals are submitted to a

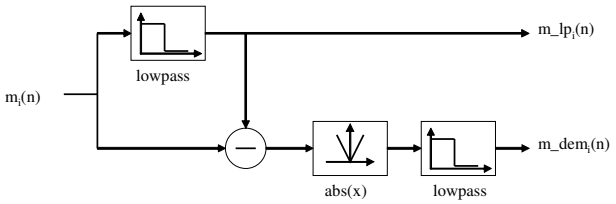

Figure 2. Filtering and demodulation of the MAV series of each lead i. The two resulting series are processed separately.

second SG-filter of width 13 ms. The result of this demodulation step, $\{m\_dem_i(n)\}$, is an estimate of the magnitude of respiratory excursions captured by the corresponding lead i. The lowest trace of fig. 1 shows an example for lead V2.

To summarize the information from all leads, vectors containing the eight m_lp values of a single beat n are defined and submitted to a principal component analysis (PCA). For apnea detection, only the first projection coefficient (PCA_{lp}) is used. The same is done for m_dem resulting in a series $\{PCA_{dem}(n)\}$ To minimize the effect of body position changes, which cause abrupt steps in the time course of coefficients [4], the estimation is performed using an adaptive algorithm (aPCA) as described in [5]. The subtraction of a recursively updated template (forgetting factor: 0.98) from each vector prior to PCA compensates for the time-varying mean caused by postural effects. Initial estimates for template and PC are obtained from the first 50 valid beats.

For the 8-channel ECG data, we compare apnea detection performance on the MAV series of the individual leads as well as on the $aPCA_{lp}$ and $aPCA_{dem}$ series. For the Physionet AECGDB, detection is performed solely on the MAV series since the database contains only one single ECG channel. The results are validated against the minute-by-minute annotations of an expert. For the multi-lead ECGs, minute-by-minute annotations are generated from the polysomnogram. Each minute of the ECG recording is labelled as apnea-positive whenever an apnea (obstructive, central or mixed) or hypopnea event is in progress during that minute; otherwise the label is apnea-negative.

As obvious from fig. 1, repetitive apnea episodes elicit fairly regular low frequent modulations in the RR- and MAV-series and extend frequently over periods of several minutes. To quantify this regularity, we applied a measure of local series similarity (*lSimil*) based on correlation analysis that we already used successfully in earlier studies [4]. Although a statement on presence or absence of apnea is made for each single minute, we use segments of 5 minutes duration for calculation of similarity that are shifted in increments of one minute

over the total ECG signal. Each 5 min segment is band-pass filtered using a low-pass SG filter of width 9 samples, and a high-pass SG filter of width 55 samples, emphasizing the CVHR-related frequency band.

From the band-pass series, we extract the central segment of 1 min duration and shift it over the total 5 min segment, calculating the normalized correlation coefficient for each time shift. The *lSimil* measure is calculated as the sum of all correlation values exceeding the empirically adjusted threshold of 0.55.

To assess the suitability of the MAV and RR-series for apnea detection, ROC analysis is performed on the *lSimil* measure calculated separately from those series. From the ROC-curve, its area (AUC) as well as sensitivity and specificity for the point on the curve geometrically closest to 100% sensitivity and specificity are determined.

As an alternative method to condense the information from multiple leads, we averaged the time courses of *lSimil* over all eight channels and compared it to the results obtained from the adaptive PCA.

3. Results

Tables 1 and 2 summarize the results of the ROC-analysis after application of the *lSimil* measure to the RR- and the MAV -series. For the clinical data set (table 1), MAV of individual leads as well as averages of *lSimil* over all leads are given. Moreover, the first PCA coefficient of the lowpass-filtered and demodulated MAV series is listed.

From table 1, it is obvious that the results based on heart rhythm (RR-series, AUC 77.9) are worse compared to the MAV and PCA series. Moreover, the results of MAV differ considerably between the ECG channels, with best performance for leads I (AUC 89.0) and II (AUC 85.6) that slightly increases when the *lSimil* values are averaged over all channels prior to ROC-analysis (AUC 89.3). The aPCA_lp-coefficients perform slightly worse (AUC 87.6) than MAV from lead I. Similarly, averaging *lSimil* of the individually demodulated MAV-series (AUC 84.5) outperforms the first $aPCA_{dem}$ coefficient (AUC 84.5). In table 2, the results for the Physionet AECGDB are listed. Here, RR-series and MAV from the available ECG lead perform comparable with slight advantages for the heart rhythm. The comparison of the two data sets shows that the MAV values from channel I (AUC 89.0) and Physionet (AUC 88.3) yield comparable results. However, there is a remarkable difference in the performance of the RR-series (AUC 89.8 vs. 77.9) which is much better in the Physionet AECGDB. In search of reasons for this different behaviour, we analyzed the clinical data set in two sub¬groups: patients with (n=4) and without (n=24) diabetes, which is known to affect the ANS and HRV. The prevalence of apnea minutes was comparable in both subgroups (26.7% vs. 28.7%). Before, all patients taking beta blockers were excluded (n = 10), since in our sample 9 out of 10 beta blocker patients had a body mass index (BMI) ≥35 (7 out of 10 BMI ≥40), and generally showed a higher degree of ectopy in the ECG. Because of these dependencies, a conclusion would hardly have been possible.

Table1. ROC-results in % for the clinical dataset (*lSimil* calculated from different data sources).

data source	ECG lead	Sens.	Spec.	AUC
RR-series		67.6	76.2	77.9
MAV	I	80.7	83.1	89.0
MAV	II	77.8	79.5	85.6
MAV	V1	67.3	76.9	77.2
MAV	V2	74.3	77.9	81.7
MAV	V3	77.3	78.4	84.4
MAV	V4	72.4	80.0	82.4
MAV	V5	74.0	75.8	81.9
MAV	V6	74.9	77.2	82.7
$aPCA_{lp}$	all	80.5	81.1	87.6
MAV	avrg(all)	83.6	81.5	89.3
$aPCA_{dem}$	all	72.3	76.8	81.6
dem(MAV)	avrg(all)	77.4	77.3	84.5

Table 2. ROC-results in % for the Physionet Apnea ECG database (*lSimil* calculated from different data sources)

data source	Sens.	Spec.	AUC
RR-series	81.3	82.8	89.8
MAV	79.8	84.9	88.3

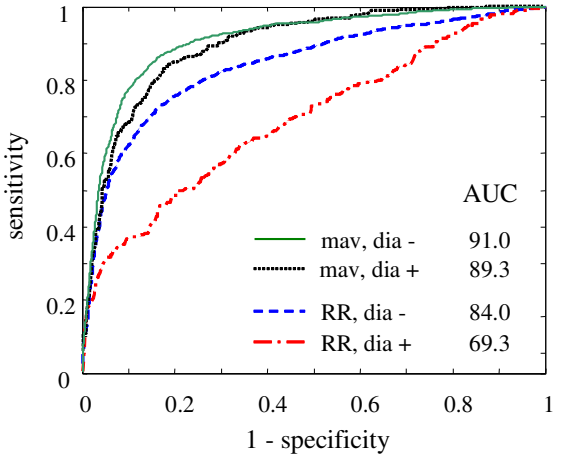

Figure 3. ROC-curves and AUC for non-diabetes (dia -) patients (solid line, dashed line), and diabetes (dia +) patients (dot, dash-dot). In each subgroup, *lSimil* of the RR-series (dash and dash-dot lines) and *lSimil* of MAV, averaged over all ECG leads, (solid, dot) are calculated. Patients with beta blocker medication were excluded.

Fig. 3 shows the ROC-curves for both subgroups. It is obvious that the curves derived from heart rate depend strongly on the presence or absence of diabetes whereas the MAV-series show much more consistent results.

4. Discussion and conclusions

A restricted view, limited to the results on the AECGDB (table 2), suggests the equivalence of heart rhythm and QRS amplitude modulation as sources of information with respect to recognition of sleep apnea from the ECG. Both perform comparably well with an AUC around 89%. Although the the MAV results on the clinical data set (table 1) are in amazingly good agreement with those of the PADB, this interpretation clearly has to be rejected when comparing the performance of the RR-series on both data sets. The decrease of almost 12% in AUC is remarkable (compare first rows in tables 1 and 2), and suggests the existence of non-apnea factors that affect heart rhythm but not morphology and occur with different prevalence in the data sets.

As possible factors, we assumed diseases (i.e. diabetes) and medication (i.e. beta blockers) that are known to affect the ANS and therefore are likely to influence HRV variables but less the ECG amplitude modulation due to respiratory movements. The subgroup analysis clearly supports this hypothesis (fig. 3): Whereas in the non-diabetes group, recognition accuracy of *lSimil* from the RR-series (dashed line) is increased (AUC 84% vs. 78% in table 1), the ROC-curve indicates a breakdown in performance for the diabetes patients (dash-dot, AUC 69.3). So, the neurological side effects of diabetes seem to blur the CVHR pattern, rendering apnea detection from heart rate in this group of patients rather unreliable. It should however be kept in mind, that these results were derived from a rather small sample of patients (n=4).

Contrarily, the ROC-results for the MAV -series are consistent for non-diabetes (AUC 91.0) and diabetes patients (AUC 89.3) and for both subgroups superior to those of the heart rhythm.

We often found cases, where within the same patient, in some apnea phases a prominent CVHR-pattern was observable whereas in other phases the heart-rhythm appeared very unspecific although the apnea was clearly detectable in the MAV -modulations. Most likely, sleep stages are important in this context. Moreover, we observed that periodic leg movements during sleep can elicit a very regular heart rhythm modulation that does not appear in the MAV-series. All these findings indicate a superior robustness of ECG amplitude variations over heart rhythm for sleep apnea detection and lower susceptibility to confounding influences, which is especially supported by the remarkably consistent results on both – very different - data sets.

Table 1 shows that the results of the MAVs are quite sensitive to the lead selected for analysis. The AUC values range from 77.2 % in lead V1 to 89.0% in lead I. Presumably, the optimal lead for detection varies even within the same patient due to body position changes during sleep. This was the main reason for choosing an adaptive algorithm to estimate the PCA coefficients. Nevertheless, it must be stated that our results do not indicate any advantage of the PCA over simple averaging of the local similarities calculated for each lead individually. This holds for both, the lowpass-filtered and the demodulated series, although for the former, averaging all channels only gives slight improvement (AUC 89.3%) over the best individual lead (I, AUC 89.0%). It should be noted that by quantifying *lSimil*, only phases of repetitive apneas (see fig. 1) are detectable which, however, in practice are usually found.

From our results, we conclude that apnea detection relying solely on heart rhythm information should be regarded cautiously and is better complemented by analysis of variability of ECG amplitude and morphology.

In general, the lf-modulation in the MAV-series (AUC 89.3) appears to be more specific for apnea and more robust for its detection than the demodulated MAVs (AUC 84.5). To condense information from multi-lead signals, simple averaging seems to be as appropriate as more complex PCA-based methods.

References

[1] Penzel T, McNames J, de Chazal P, Raymond B, Murray A, Moody G. Systematic comparison of different algorithms for apnoe detection based on electrocardiogram recordings. Med. Biol. Eng. Comput. 2000;40:402-7.

[2] 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 1984:126-31.

[3] Moody GB, Mark RG, Zoccola A, Mantero S. Derivation of respiratory signals from multi-lead ECGs. In Proc. Computers in Cardiology 1985. Washington DC: IEEE Computer Society Press, 1985:113-6.

[4] Maier C, Dickhaus H, Bauch M, Penzel T. Comparison of heart rhythm and morphological ECG features in recognition of sleep apnea from the ECG. In: Proc, Computers in Cardiology 2003. Piscataway: IEEE Press, 2003:311-4.

[5] Weng J, Zhang Y, Hwang WS. Candid covariance free incremental principal component analysis. IEEE Trans. PAMI 2003;25(8):1034-40.

Address for correspondence

Christoph Maier
Department of Medical Informatics, Heilbronn University
Max-Planck-Str. 39, D-74081 Heilbronn, Germany
christoph.maier@hs-heilbronn.de