Robust Detection of Sleep Apnea from Holter ECGs

Joint Assessment of Modulations in QRS Amplitude and Respiratory Myogram Interference

C. Maier[1,2]; H. Wenz[3]; H. Dickhaus[1]

[1]Institute of Medical Biometry and Informatics, Heidelberg University Hospital, Heidelberg, Germany;
[2]Department of Medical Informatics, Heilbronn University, Heilbronn, Germany;
[3]Thoraxklinik, Heidelberg University Hospital, Heidelberg, Germany

Keywords
ECG, ECG-derived respiration, heart rate, respiratory myogram interference, sleep related breathing disorders

Summary
Introduction: This article is part of the Focus Theme of *Methods of Information in Medicine* on "Biosignal Interpretation: Advanced Methods for Studying Cardiovascular and Respiratory Systems".
Objectives: Detect presence of sleep-related breathing disorders (SRBD) in epochs of 1 min by signal analysis of Holter ECG recordings.
Methods: In 121 patients, 140 synchronized polysomnograms (PSGs) and 8-channel Holter ECGs were recorded. The only excluded condition was persistent arrhythmias. Respiratory events as scored from the PSGs were mapped to a 1 min grid and served as reference for ECG-based detection. Moreover, 69/70 recordings of the Physionet Sleep Apnea ECG Database (PADB) were included. We performed receiver operating characteristics analysis of a single, novel time-domain feature, the joint local similarity index (jLSI). Based on cross-correlation, the jLSI quantifies the time-locked occurrence of characteristic low-frequency (LF) modulations in ECG respiratory myogram interference (RMI), QRS amplitude (QRSA) and heart rate.
Results: Joint oscillations in QRSA, RMI and the envelope of RMI identified positive epochs with a sensitivity of 0.855 (PADB: 0.873) and a specificity of 0.86 (PADB: 0.88). Inclusion of heart rate did not improve detection accuracy.
Conclusions: Joint occurrence of LF-modulations in QRSA and RMI is a characteristic feature of SRBD that is robustly quantified by the jLSI and permits reliable and reproducible detection of sleep apnea in very heterogeneous settings.

Correspondence to:
Christoph Maier
Institute of Medical Biometry and Informatics
Department of Medical Informatics
Heidelberg University Hospital
Im Neuenheimer Feld 305
69120 Heidelberg
Germany
E-mail: christoph.maier@med.uni-heidelberg.de

Methods Inf Med 2014; 53: 303–307
doi: 10.3414/ME13-02-0043
received: October 22, 2013
accepted: April 22, 2014
epub ahead of print: July 31, 2014

1. Introduction

Sleep-related breathing disorders (SRBD) are a very prevalent, yet severely underdiagnosed health problem. They significantly reduce quality of life and life expectation [1]. Because effective therapy is available for most cases, it is highly desirable to identify patients as early as possible. A particularly promising and attractive approach in this context is the idea to screen Holter ECGs for SRBD.

Respiratory modulations in the ECG with respect to QRS amplitude (QRSA), respiratory myogram interference (RMI), and RR-interval have already been described by Einthoven. Based on beat-to-beat quantification of the area under the QRS complex, the ECG-derived respiration (EDR) method [2] is the probably most popular today to extract respiratory information from the ECG. It is applied in numerous studies that deal with ECG-based detection of SRBD, e.g. [3–7]. A second, widely used source of information is the SRBD-related cyclic variation of heart rate [8]. Surprisingly, only few studies have made use of the respiratory myogram interference (RMI) [7, 9].

Most studies follow classical pattern recognition approaches in a sense that they separately quantify features from the different data sources. In a second step, these are aggregated into a feature vector and classified. Only few studies have analyzed the coupling of the different sources of information at the feature level [6].

This article presents a multivariate extension of our previous work [7, 10] and introduces a joint local similarity index (jLSI) as a feature that quantifies the time-locked occurrence of characteristic repetitive oscillations in QRSA and RMI. We show that this index permits robust detection of SRBD on very heterogeneous and different data sets. Moreover, the role of heart rate based features in detection of SRBD is discussed.

2. Methods

2.1 Data

The data analyzed in this study stems from two independent sources. Sample I consists

of 140 8-channel Holter-ECG recordings (Mortara H12+, sample frequency 1000 Hz/ch) from 121 patients who were referred to the sleep medicine center of the Thoraxklinik at Heidelberg University Hospital owing to suspected sleep-related breathing disorders. The respiratory annotations of simultaneously registered polysomnograms (PSGs, Alice 4) were synchronized to the Holter ECGs and mapped to a grid of adjacent one-minute intervals. A minute was scored as 'apnea positive' when it overlapped with a respiratory event of the PSG. In order to approximate a typical SRBD collective, we did not exclude any confounding medication or comorbidity except for persistent supraventricular arrhythmias. In particular, sample I includes 16 diabetes (type II) patients and 7 patients with a history of myocardial infarction.

Sample II consists of 69 out of the 70 ECG recordings of the Physionet sleep apnea ECG database (PADB) [11]. One record (b04) was excluded because of its bad ECG signal quality. We increased the original sampling rate of 100 Hz rate by means of cubic spline interpolation to 1000 Hz. The PADB provides annotations on the occurrence of SRBD during each minute of the recordings. In contrast to sample I it only contains obstructive or mixed events, but not purely central apneas. Based on the cumulative time spent in disordered breathing, the PADB is divided into three groups. Group A consists of 40 severe cases with more than 100 min spent in apnea. Group B is a borderline group with 10 recordings (9 of which were used in our study) containing between 10 min and 100 min of apnea. Group C comprises 20 control recordings with less than 5 min in apnea.

2.2 Preprocessing

QRS complexes were detected and classified using self-developed software. The series of RR intervals was derived by taking the difference between the reference timepoints of adjacent heart beats. For each beat in each lead a surrogate measure for the area under the QRS complex was calculated as the mean absolute amplitude value in a window of ± 60 ms centered at the QRS fiducial. Finally, the myogram interference was assessed for each beat in each lead as the RMS value of the residuals in the ECG signal after high-pass filtering with 60 Hz in a temporal window extending from 80 ms to 430 ms after the QRS complex. For sample II, the cutoff-frequency was 30 Hz owing to the low original sample rate of only 100 Hz. In practice, the cutoff is not critical. Details on the calculation of these respiratory surrogate data series can be found in [7]. All beat-to-beat series were interpolated using cubic splines and equidistantly re-sampled at 3 Hz. In the further text, these quantities are referred to as R (RR-series), Q (QRSA) and M (RMI). Since SRBD also modulates the intensity of respiratory efforts, we additionally calculated signals related to the envelope of R, Q and M. These were derived by high-pass filtering, rectification and subsequent low-pass filtering [10]. The identifiers used for these quantities are eR, eQ, and eM.

2.3 Classification Feature

Clinically relevant respiratory events occur repetitively and elicit typical low-frequency (LF) modulations on a time scale between 20s and 100s in the base data series (▶Figure 1). The idea of the local similarity index (LSI) is the quantification of local recurrences of such LF-oscillations with a stable morphology [7]. This is performed separately for each of the data types R, Q, M, eR, eQ and eM. The LSI is calculated from the band-pass filtered data series in temporal windows of 5 min duration. Within the central minute of each 5 min window, a candidate prototype SRBD pattern with a length of 30s is identified dependent on the location of the most dominant signal extremum (in ▶Figure 1 this would correspond to one of the 'bumps'). Then, the normalized cross-correlation between this candidate pattern and the 5 min window is calculated for each time shift. Correlation values are in the range [–1, 1], where values close to 1.0 indicate a recurrence of the prototype pattern. Values that exceed a predefined positive threshold θ are additively accumulated to form the LSI feature value. It is logically assigned to the central minute of the 5 min window. Then, the window is shifted by 1 min resulting in an effective time resolution of 1 min for the LSI values. A detailed description of the LSI is given in [7].

The multivariate extension of this approach, which defines the joint LSI (jLSI), is straightforward. In SRBD, the source of the LF-oscillations that we are interested in, is a common trigger: the respiratory event. This common root causes the oscillations to occur strictly synchronized and time-locked in the different base data types (cf. dotted vertical lines in ▶Figure 1). This means that if an apnea event is truly present in a given minute, the same location of a prototype pattern that has been selected in one data type should also identify a meaningful pattern in any of the

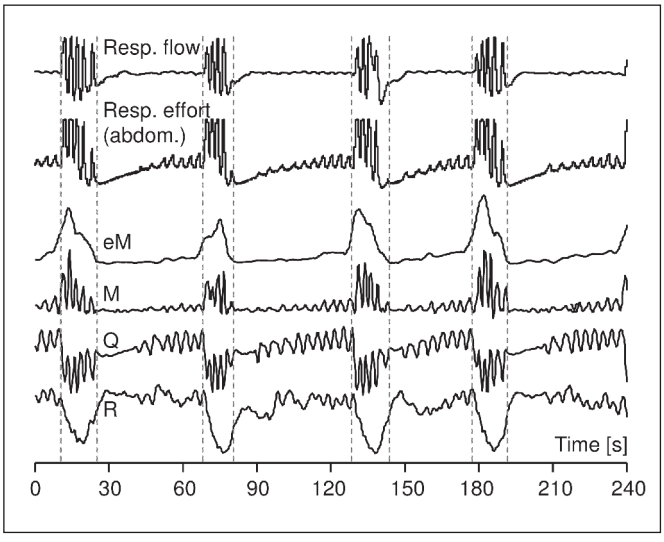

Figure 1 Exemplary time courses of derived time series eM, M, Q and R (four lower traces) and respiratory measurements in the PSG (two upper traces) during repetitive mixed apnea events

other data types. This keynote is used in calculation of the jLSI.

In our realization, the series Q (which yielded the best LSI results) always serves as basis for the selection of the prototype pattern location in all jLSI calculations (▶Figure 2 step A). In exactly the same location, the patterns for all other base data types used in the jLSI calculation are extracted (▶Figure 2 step B). Comparable to the LSI, a normalized correlation function is calculated separately for each base data type using its own specific pattern (▶Figure 2 step C). If local recurrences of the patterns are truly due to apnea events, their locations again should be time-locked between the data types, meaning that the maxima of correlation should occur in the same positions. This is simply quantified by multiplying the correlation functions element-wise (▶Figure 2 step D). In a last step, all values of the product correlation function that exceed a predefined positive threshold θ are accumulated. Their sum defines the feature value of jLSI (▶Figure 2 step E). We analyzed different combinations of up to four of the base data types with the jLSI.

It should be noted that in sample I, each base data type actually comprises eight time series, one for each ECG lead. The LSI and jLSI were calculated separately for each ECG lead. Then, for each minute, the maximum value found in any lead was kept as the final feature value for this minute.

2.4 Classification

Since both the LSI and the jLSI represent scalar values, our classification procedure utilizes a threshold which is selected according to a receiver operating characteristics (ROC) analysis. A median filter (width: 9) was applied prior to the ROC analysis to suppress spurious outliers in LSI resp. jLSI.

3. Results

▶Table 1 lists the results for the LSI applied separately to all base data types R, M, Q and their envelopes eM, eR, eQ. Moreover, the results for the jLSI are given for selected combinations of the base data types. In its last row, ▶Table 1 shows the result that was obtained when the best jLSI combination as found on sample I was applied to sample II (PADB). For the LSI, Q, M, eQ and eM show comparable results with sensitivity and specificity ranging from 0.80 to 0.83, and an AUC around 0.88. R and especially eR perform significantly worse with sensitivity below 0.684 and specificity below 0.755. This extends into the results of the jLSI where inclusion of heart rate information turns out to be detrimental. Whereas the combination of Q and M rises sensitivity and specificity to 0.849/0.852, and even further to 0.855/0.860 for Q, M, eM, the additional combination with R even decreases the performance slightly to 0.851. More evident is the decrease when only Q and R are combined, resulting in a drop of almost 4% in sensitivity and 3% in specificity compared to the LSI of Q alone. When the best combination as seen on sample I – jLSI(Q, M, eM) – is evaluated on the independent sample II, the results of sample I are even surpassed resulting in a sensitivity of 0.874 and a specificity of 0.880. This combination achieves perfect separation between groups A and C of the PADB. For group B only one record overlaps with group A and two records overlap with group C.

4. Discussion and Conclusions

This paper presents an accurate and traceable method to robustly detect SRBD by analysis of the nocturnal Holter-ECG based on modulations in QRSA and RMI. It naturally extends the idea of the LSI as an adaptive matched filter that identifies characteristic LF-oscillations related to SRBD [7]. The jLSI realizes a significant improvement of sensitivity and specificity by quantifying the joint, time-locked occurrence of such fluctuations in more than one base-data type. At the same time, the simplicity and traceability of the original LSI ap-

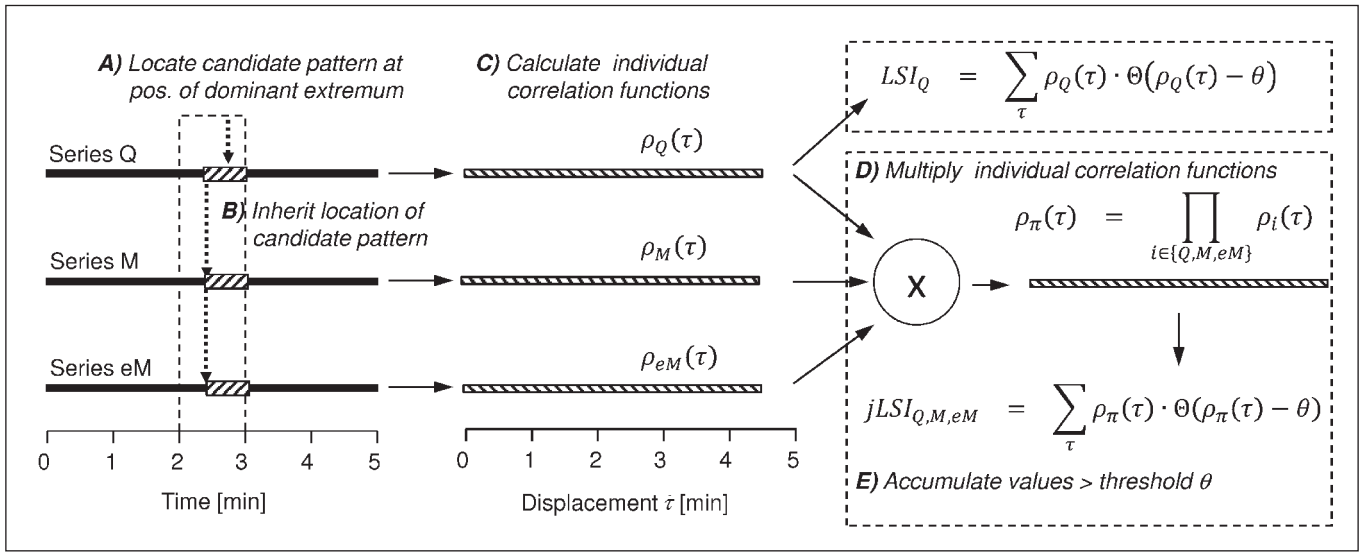

Figure 2 Calculation of the jLSI (see text), here for three base data types Q, M and eM. The shaded region in the left figure part indicates the central minute of the 5 min segment, to which the jLSI value is logically assigned.

proach is kept. A comparison with other multivariate approaches in the literature shows that an increase in accuracy in the order of magnitude of several percent is not easily achieved by feature combination [4–6]. This holds especially in the light of the already high base level of more than 0.82 accuracy that is given by LSI(Q) (▶Table 1).

4.1 Properties of the jLSI

One reason may be that – in contrast to 'classic' multivariate approaches – the combination of multi-source information in the jLSI takes place at the level of the feature and not at the level of the classifier. This increases the temporal resolution for synchronization assessment to the sub-epoch scale. The presumably first study in the field that quantifies this type of cross-coupling at the feature level with high temporal resolution was performed by Mendez et al. [6]. They defined phase-space features and used spectral coherence to quantify coupling between the RR interval series and the EDR. At first view, the approach of the jLSI may appear equivalent to that of spectral coherence. But there are important differences. E.g. does the jLSI neither require exactly periodic repetition of the LF oscillations (e. g. apnea events) nor a definition of the frequency range of interest. Because of the time-domain definition of the prototype patterns, not only the energy of the fundamental frequency component but also of higher-order harmonics inherently contributes to the result. In contrast to the coherence function, the jLSI is sensitive to phase shifts of the time series. This may increase its specificity.

Finally, the joint analysis of data with their envelope in the jLSI relates modulating activity at the respiratory frequency to oscillations at the much lower frequency of repetition of the apnea events. Mathematically, this establishes a connection of our analysis to the ideas that underlie higher-order spectral analysis.

4.2 Detection Accuracy and Generalization

It is remarkable that the excellent result of the jLSI as obtained on sample I is not only maintained but even surpassed on sample II (▶Table 1) with entirely different technical data quality. It should be noted that sample II includes 69 out of the 70 recordings of the PADB from both the learning and the test set. Other studies exclude up to 28% of the data [6].

We are aware of only two other studies by de Chazal et al. [4] and Hayano et al. [12], where results have been assessed on both, the PADB and independent data. For de Chazal's method, the perfect *subject-specific* accuracy that is also achieved by our method, dropped by ~20% when applied to different data sets [13]. Hayano's method appears more robust in the sense that the *subject-specific* accuracy of 0.93 (sensitivity 0.9, specificity 1.0) as achieved on the PADB was maintained to a significant extent on their remarkably large sleep-study database (N = 862 out of 1193; accuracy 0.86), however with data exclusions of 27.7%. For *epoch-based* detection the accuracy was 0.83 on the PADB. One likely reason for that better robustness is the straightforward classifier architecture based on a single feature, which it shares with our method. Keeping in mind that, apart from atrial fibrillation, our sample I includes many typical comorbidities as well as concomitant medication, and is comparatively large in size, we can attribute excellent generalization properties to the jLSI approach.

Table 1 Classification results of the LSI and jLSI feature (column 1) in terms of sensitivity ('Sens.'), specificity ('Spec.') and area under the ROC curve ('AUC') for different combinations of base data types as indicated in column 2. The column 'Coll.' specifies the collective (sample I or II).

Feat.	Base data	Coll.	Sens.	Spec.	AUC
LSI	R	I	0.684	0.755	0.784
LSI	M	I	0.806	0.815	0.882
LSI	Q	I	0.815	0.829	0.891
LSI	eR	I	0.636	0.706	0.725
LSI	eM	I	0.820	0.822	0.883
LSI	eQ	I	0.803	0.813	0.874
jLSI	Q R	I	0.776	0.801	0.862
jLSI	Q M	I	0.849	0.852	0.914
jLSI	Q M eM	I	0.855	0.860	0.919
jLSI	Q M eM R	I	0.851	0.851	0.918
jLSI	Q M eM	II	0.873	0.880	0.934

4.3 Base Data Types

The fact that inclusion of heart rate information does not improve but even rather decreases classification accuracy (▶Table 1) seems to contradict the results of the vast majority of other studies, which partly heavily rely on HRV parameters, in particular the large study by Hayano et al. [12]. But in accordance with the results of our current, and a previous study [10], it has been pointed out that the results of HRV-based sleep apnea detection appear to depend strongly on the database used [14]. We suspect that the deeper reason for this observation is the fact that HRV is mediated via physiologic pathways that are modulated by a multitude of confounding factors, as opposed to the rather 'mechanically' or 'physically' mediated oscillations seen in Q and M. Our analysis of subgroups including diabetes (N = 17), myocardial infarction (N = 7) and periodic leg movements (N = 9) clearly confirms a detrimental effect of these conditions on apnea recognition from the RR-series [15]. With respect to periodic leg movements this has also be confirmed by Hayano et al. [12].

Our results show that – in order to gain insight on the accuracy expected for a realistic SRBD population – it is important to include typical pathologies into the sample, and that accurate recognition is possible without the use of heart rhythm informa-

tion. This even offers the perspective to apply the method in patients with permanent arrhythmia, an important group that hitherto had to be regarded as unaccessible for SRBD-screening from the ECG.

References

1. Young T, Finn L, Peppard PE, Szklo-Coxe M, Austin D, Nieto FJ, et al. Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin sleep cohort. Sleep 2008; 31 (8): 1071–1078.
2. Moody GB, Mark RG, Zoccola A, Mantero S. Derivation of Respiratory Signals from Multi-lead ECGs. In: Murray A, editor. Computers in Cardiology. Vol. 12. Washington, DC: IEEE Computer Society Press; 1985. pp 113–116.
3. Penzel T, McNames J, Murray A, de Chazal P, Moody G, Raymond B. Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings. Med Biol Eng Comput 2002; 40 (4): 402–407.
4. de Chazal P, Heneghan C, Sheridan E, Reilly R, Nolan P, O'Malley M. Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. IEEE Trans Biomed Eng 2003; 50 (6): 686–696. Available from: http://dx.doi.org/10.1109/TBME.2003.812203.
5. de Chazal P, Heneghan C, McNicholas WT. Multimodal detection of sleep apnoea using electrocardiogram and oximetry signals. Philos Transact A Math Phys Eng Sci 2009; 367 (1887): 369–389. Available from: http://dx.doi.org/10.1098/rsta.2008.0156.
6. Mendez MO, Bianchi AM, Matteucci M, Cerutti S, Penzel T. Sleep apnea screening by autoregressive models from a single ECG lead. IEEE Trans Biomed Eng 2009; 56 (12): 2838–2850. Available from: http://dx.doi.org/10.1109/TBME.2009.2029563.
7. Maier C, Wenz H, Dickhaus H. Steps toward subject-specific classification in ECG-based detection of sleep apnea. Physiol Meas 2011; 32 (11): 1807–1819. Available from: http://dx.doi.org/10.1088/0967–3334/32/11/S07.
8. Guilleminault C, Connolly S, Winkle R, Melvin K, Tilkian A. Cyclical variation of the heart rate in sleep apnoea syndrome. Mechanisms, and usefulness of 24 h electrocardiography as a screening technique. Lancet 1984; 1 (8369): 126–131.
9. Bystricky W, Safer A. Identification of individual sleep apnea events from the ECG using neural networks and a dynamic Markovian state model. In: Proc. Computers in Cardiology; 2004. pp 297–308.
10. Dickhaus H, Maier C. Detection of sleep apnea episodes from multi-lead ECGs considering different physiological influences. Methods Inf Med 2007; 46 (2): 216–221.
11. Penzel T, Moody GB, Mark RG, Goldberger AL, Peter JH. The apnea-ECG database. In: Proc. Computers in Cardiology 2000; 2000. pp 255–258.
12. Hayano J, Watanabe E, Saito Y, Sasaki F, Fujimoto K, Nomiyama T, et al. Screening for obstructive sleep apnea by cyclic variation of heart rate. Circ Arrhythm Electrophysiol 2011; 4 (1): 64–72. Available from: http://dx.doi.org/10.1161/CIRCEP.110.958009.
13. Heneghan C, de Chazal P, Ryan S, Chua CP, Doherty L, Boyle P, et al. Electrocardiogram recording as a screening tool for sleep disordered breathing. J Clin Sleep Med 2008; 4 (3): 223–228.
14. Lado MJ, Vila XA, Rodríguez-Liñares L, Méndez AJ, Olivieri DN, Félix P. Detecting sleep apnea by heart rate variability analysis: assessing the validity of databases and algorithms. J Med Syst 2011; 35 (4): 473–481. Available from: http://dx.doi.org/10.1007/s10916–009–9383–5.
15. Maier C. Beitrag zur Detektion schlafbezogener Atmungsstörungen durch Signalanalyse des Elektrokardiogramms. Heidelberg University; 2011.