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

Sleep Disordered Breathing Detection Using Heart Rate Variability and R-Peak Envelope Spectrogram

Mohammad A. Al-Abed, *Student Member, IEEE*, Michael Manry, *Senior Member, IEEE*, John R. Burk, Edgar A. Lucas, and Khosrow Behbehani, *Senior Member, IEEE*

***Abstract*—We report that combining the interbeat heart rate as measured by the RR interval (RR) and R-peak envelope (RPE) derived from R-peak of ECG waveform may significantly improve the detection of sleep disordered breathing (SDB) from single lead ECG recording. The method uses textural features extracted from normalized gray-level co-occurrence matrices of the time frequency plots of HRV or RPE sequences. An optimum subset of textural features is selected for classification of the records. A multi-layer perceptron (MLP) serves as a classifier. To evaluate the performance of the proposed method, single Lead ECG recordings from 7 normal subjects and 7 obstructive sleep apnea patients were used. With 500 randomized Monte-Carlo simulations, the average training sensitivity, specificity and accuracy were 100.0%, 99.9%, and 99.9%, respectively. The mean testing sensitivity, specificity and accuracy were 99.0%, 96.7%, and 97.8%, respectively.**

I. Introduction

SLEEP- Disordered Breathing (SDB) is estimated to have a prevalence of at least 6% of US adults [1]. SDB has been shown to affect the productivity and quality of life of the patient, and an increase in the mortality risk and congestive heart failure (CHF) for untreated SDB patients. Prevalence is expected to rise for both adults and children due to rise in obesity, a significant risk factor for SDB [1]. The general adult population is widely thought to be under diagnosed, because the present method of diagnosing SDB, nocturnal polysomnography (NPSG), is relatively expensive and not readily accessible. Cost effective and more accessible means to screen the at-risk population for SDB are highly desirable.

Previous studies [2] have shown the possibility of screening for OSA episodes using overnight ECG recordings. Different algorithms have been developed to extract reliable markers from ECG signals including heart rate variability (HRV), R-Peak Envelope (RPE), and power spectral analysis of particular frequency bands of different ECG-derived respiration (EDR) signals. The use of HRV is appealing, as it reflects the autonomic response to apnea episodes. One study has combined different temporal statistical features from HRV and RPE to improve the detection rate, reporting accuracy of about 90% [3]. Different investigators have qualitatively explored spectrograms of HRV for the detection of OSA using ECG [2].

Our previous work have shown similar results in detecting SDB events by combining cross-correlation and scatter plot features from HRV and RPE [4]. In this paper, we combine textural features extracted from the spectrograms of both HRV and RPE to detect SDB events. A multi-layer perceptron (MLP) classifier is used to detect SDB events.

Manuscript received April 7, 2009. This work was supported in part by a grant from the U.S. Department of Energy.

M. A. Al-Abed is a Ph.D. student with the Department of Bioengineering, the University of Texas at Arlington, Arlington, TX 76010 USA (phone: 817-272-2249; fax: 817-272-2251; e-mail: mohammad@uta.edu).

M. T. Manry is professor at the Department of Electrical Engineering, the University of Texas at Arlington, Arlington, TX 76010 USA (e-mail: manry@uta.edu).

J. R. Burk and E. A. Lucas are with Sleep Consultants, Inc., Fort Worth, TX, 76104 USA.

K. Behbehani is a professor and chair of the Department of Bioengineering Department, the University of Texas at Arlington, Arlington, TX 76010 USA (e-mail: kb@uta.edu).

II. Methodology

A. Subjects

Seven volunteers who had no known history of SDB and seven volunteers who had been positively diagnosed as having SDB were recruited for the study. The subjects underwent a full night (6 to 8 hours) of NPSG. The test were performed at Sleep Consultants, Inc., Fort Worth, TX, an accredit sleep laboratory. A certified sleep expert, blind to the aims of this study, scored the NPSG data according to Rechtschaffen & Kales standard. The severity of SDB was measured using the apnea-hypopnea index (AHI). Table I summarizes the subject population demographics.

TABLE I
Subject Demographics of the Normal and SDB Subject groups including the Apnea/Hypopnea Index

Subject Group (N)	Number of Males/Females	Age (mean ±std)	BMI (mean ±std)	AHI (mean ±std)
NOR (7)	5/2	43.0±8.6	24.7±4.3	4.4±3.6
OSA (7)	2/5	51.1±9.8	34.2±7.0	38.7±18.9

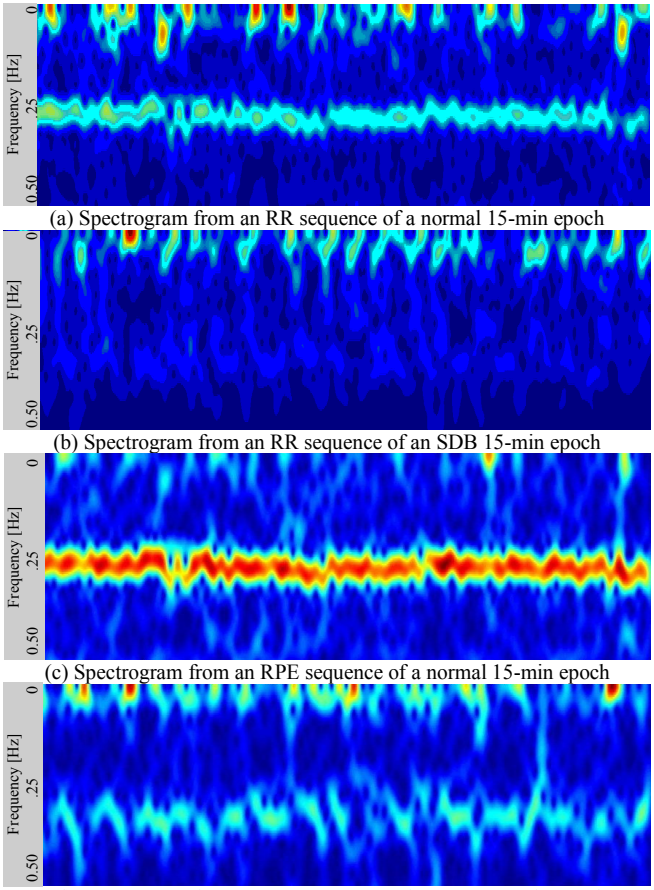
B. Experimental Protocol

The physiological markers were reordered during the NPSG study included ECG, EEG, SaO_2, and airflow. As part of the NPSG recording, ECG Lead I was digitally recorded at 1 kHz sample rate. These recordings were used to extract both the RR and RPE time sequences [3]. The data was parsed into epoch of 900 seconds (15 min) in length to capture very low RR variations (~0.001Hz), as recommended by previous investigators [5]. Fifty six

epochs were used, 4 epochs from each subject. The RR and RPE were interpolated using cubic spline technique, and the resulting function was uniformly sampled at 1 Hz.

C. Image Construction

Short Time Discrete Fourier Transform (STDFT) was separately performed on both the uniformly sampled RR and RPE time series, resulting into spectrograms. Fig. 1 shows a comparison between spectrograms extracted from RR and RPE sequences for a normal clip verses a clip containing SDB events. Visual differences can be observed between normal and SDB clips for both RR and RPE sequences.



(a) Spectrogram from an RR sequence of a normal 15-min epoch

(b) Spectrogram from an RR sequence of an SDB 15-min epoch

(c) Spectrogram from an RPE sequence of a normal 15-min epoch

(d) Spectrogram from an RPE sequence of an SDB 15-min epoch

Fig. 1. Comparison between spectrograms extracted from RR and RPE sequences for a normal clip verses a clip containing SDB events. Visual differences can be observed between normal and SDB clips for both RR and RPE sequences. The described method aims at quantifying these visual differences, and using these features for further classification.

Each spectrogram extracted from both the RR and RPE sequences is encoded to produce four gray level intensity images: First image is generated by finding the magnitude of the complex spectrogram matrix, and quantizing it to 16 equally spaced gray levels (N_g = 16) – darker shades represents lower power intensities and brighter shades represent higher power intensities. The second image was generated using the same method for the first image, but with 32 gray levels (N_g = 32). Third image was generated by computing the log of the magnitude of spectrogram magnitudes before quantizing it to 16 gray levels (N_g = 16). Finally, the fourth image was constructed by finding the histogram of the spectrogram magnitudes, then allocating them to un-equal size quantization bins, so that each gray shade would be represented in the image (N_g = 16) [7].

D. Co-occurrence Matrices and Textural Features

Normalized gray level co-occurrence matrices (NCM) [8] were used to quantitatively analyze the resultant images. Due to the significance of the lower range frequency in reflecting apnea events (0-0.25Hz) [2], two NCMs (NCM-3 and -9) were calculated for the lower half of the images.

Ten NCMs were extracted from the four gray encoded images described earlier [7]. Here, d represents the distance of the pairing of the pixels, and θ is the orientation. From the first image, four NCMs were selected:

- NCM-1 (d = 5, θ = 90º)
- NCM-2 (d = 1, θ = 90º)
- NCM-3 (d = 5, θ = 90º)
- NCM-4 (d = 5, θ = 0º)

From the second image, three NCMs were selected:

- NCM-5 (d = 5, θ = 90º)
- NCM-6 (d = 3, θ = 90º)
- NCM-7 (d = 1, θ = 90º)

From the third image, two NCMs were selected:

- NCM-8 (d = 5, θ = 90º)
- NCM-9 (d = 5, θ = 90º)

From fourth image, one NCM was selected:

- NCM-10 (d = 5, θ = 90º)

Nine textural features [9–10] were computed from these NCMs. The Matrix Mean and Matrix Variance are defined as follows [11]:

$$\text{Matrix Mean, } \mu = \sum_{i}^{N_g} i \sum_{j}^{N_g} M(i,j) \tag{1}$$

$$\text{Matrix Variance, } \sigma^2 = \sum_{i}^{N_g} (i-\mu)^2 \sum_{j}^{N_g} M(i,j) \tag{2}$$

Matrix Variance (2) and the following textural features were used for this study [9 –11]:

$$\text{Entropy: } ENT = -\sum_{i}^{N_g}\sum_{j}^{N_g} M(i,j)\cdot\log(M(i,j)) \tag{3}$$

$$\text{Angular Second Moment: } ASM = \sum_{i}^{N_g}\sum_{j}^{N_g} [M(i,j)]^2 \tag{4}$$

$$\text{Contrast: } CON = \sum_{i}^{N_g}\sum_{j}^{N_g} M(i,j)\cdot(i-j)^2 \tag{5}$$

$$\text{Correlation: } COR = \sum_{i}^{N_g}\sum_{j}^{N_g} \frac{M(i,j)\cdot(i-\mu)\cdot(j-\mu)}{\sigma^2} \tag{6}$$

$$\text{Dissimilarity: } DIS = \sum_{i}^{N_g}\sum_{j}^{N_g} M(i,j)\cdot|i-j| \tag{7}$$

Inverse Difference: $$IND = \sum_{i}^{N_g}\sum_{j}^{N_g}\frac{M(i,j)}{1+(i-j)} \quad (8)$$

Inverse Difference Moment: $$IDM = \sum_{i}^{N_g}\sum_{j}^{N_g}\frac{M(i,j)}{1+(i-j)^2} \quad (9)$$

Inverse Recursivity: $$INR = -\sum_{i}^{N_g}\sum_{j=i+1}^{N_g} 2\cdot M(i,j)\cdot\log(2\cdot M(i,j)) - \sum_{i}^{N_g} M(i,i)\cdot\log(M(i,i)) \quad (10)$$

where N_g is the number of gray levels used in the image, M is an $N_g \times N_g$ gray level co-occurrence symmetric matrix with $M(i,j)$ as its ith, jth element for $i = j = 1, 2, 3, \ldots, N_g$.

E. Optimum Feature Selection

For each of the 56 15-min epochs, a total of 180 textual features were computed. That is, 90 textural features were extracted from the RR sequence, and 90 textural features were extracted from the RPE sequence. It is noted that 180 features are generated by computing the textural feature (2) – (10) above, for each of the 10 NCMs defined earlier. A Piecewise Linear Networks (PLN) algorithm was used [12] to find the optimum feature subset. The PLN utilizes a piecewise linear orthonormal least square procedure. It selects the feature subset that is optimal for use with the MLP classifier in a computationally efficient fashion, as it requires only one pass of the data. The features extracted from the data are divided into an appropriate number of clusters and auto- and cross-correlation matrices are calculated only once. Then, it finds combinations of features that have high potential for classification of signal source [12]. The training vectors for the classifier will include only the optimum features selected.

F. Multilayer Perceptron Classifier

Multilayer perceptron (MLP) is a form of feed-forward neural network (FFNN). Its simple structure allows for relatively easy training, using conventional back-propagation (BP) training algorithms. It has been shown that MLP classifiers are very successful in image classification applications [13]. A classifier is constructed using a three-layer MLP consisting of an input layer, hidden layer and an output layer. The input layer has a number of nodes equal to the input vector length. The output layer consists of one node, accounting for a possibility of only 2 classes to be classified. The number of nodes in the hidden layer, N_h, is selected by an iterative training and validation scheme. Besides changing the number of hidden nodes, each layer of a MLP has two parameters that are selected to achieve maximum detection: node transfer function and weight vector. Both input and output nodes use linear transfer functions. The hidden layer uses a hyperbolic tangent sigmoid function [13].

G. Training and Validation

The epochs in the data set were randomly divided into two sets: a Training Set and a Testing Set. Seventy percent of the epochs are used to train the MLP (39 epochs), while 30% (17 epochs) were kept separate to test the performance of the classifier.

Furthermore, the training set is divided to 3 subsets (13 epochs each). A 3-fold cross-validation (k-fold XV) scheme is used to find the optimum number of hidden nodes, N_h, and the number of training iterations to achieve the maximum validation. Fig. 2. illustrates this process.

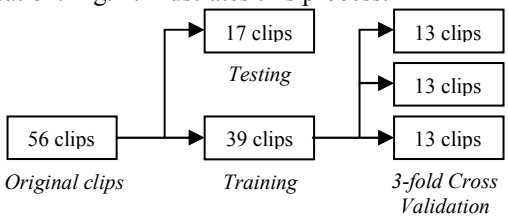


Fig. 2. Schematics showing the assignment of clips for training, validation and testing. The original 56 clips are randomly divided to 39 clips for training and validation, and 17 clips for blind testing. The 39 clips are divided to 3 subset of 13 for a 3-fold cross validation.

H. Testing

Once the optimum topography of the MPL is selected, the MLP is trained using all 39 training vectors. Then, the MLP weights and bias vectors are fixed, and the Testing Set of 17 vectors is run once to test the performance of the classifier [13].

I. Sensitivity, Specificity, and Accuracy

The performance of the system is described by

$$\text{Sensitivity} = \frac{SDB_c}{\text{Total SDB clips tested}}\times 100\% \quad (11)$$

$$\text{Specificity} = \frac{NOR_c}{\text{Total NOR clips tested}}\times 100\% \quad (12)$$

$$\text{Accuracy} = \frac{SDB_c + NOR_c}{\text{Total NOR \& SDB clips}}\times 100\% \quad (13)$$

where SDB_c is the number of correctly detected SDB clips and NOR_c is the number of correctly detected NOR clips [7].

J. Monte Carlo Simulation

Since the optimization process described above is partially dependent on the initial vector assignment to Training and Testing Sets, a Monte Carlo simulation method was devised to estimate the average performance of the MLP classifier. With the MLP topology fixed, the data set is randomly divided into training and testing sets with a 70% to 30% ratio, and the MLP is trained using the training set, then the performance of the system is tested using testing set, not seen by the NNFF. This process is repeated 500 times, and the average training and testing sensitivity, specificity, and accuracy are obtained.

III. Results

A. Feature Selection

Once the Using the PLN algorithm, the following thirteen features comprise the optimal feature subset:

Nine textural features the RR sequence

ASM_2; IDM_4; COR_4; ASM_4;

ENT_4; IND_4; ENT_8; IND_9; ASM_{10}

And four textural features from the RPE sequence

IDM_2; IDM_4; IDM_5; IDM_8

where the subscript indicates the NCM that the feature was extracted from. A vector containing the above optimized features was used as an input to the MLP classifier.

B. Testing and Validation

Using a 3-fold cross validation scheme, and using the Levenberg-Marquardt (LM) training algorithm for the MLP, the resulting optimum training-validation MLP topology was 3-layer MLP, with an input layer of 13 nodes, equal to the length of a training vector, a hidden layer with 8 nodes, and an output layer with one unit. An optimal validation is achieved when the MLP is trained for 46 iterations.

C. Monte Carlo Testing Results

Table II shows summary of the training and testing detection results after a 500-run Monte Carlo simulation.

TABLE II
AVERAGE SENSETIVITY, SPECIFICITY, AND ACCURACY OF A MLP CLASSIFIER FOR A 500-RUN MONTE CARLO SIMULATION

	Sensitivity	Specificity	Accuracy
Training	100.0%	99.9%	99.9%
Testing	99.0%	96.7%	97.8%

IV. Discussion

By studying the images in Figure 1, it is evident that there are notable differences in the power distribution of the spectrograms of RR and RPE for normal epochs compared to epochs containing SDB events. To quantify these differences, image processing techniques were applied to extract textural features from these spectrograms.

The IDM feature extracted from RPE spectrogram images appears to be the major contributors for these sequences, and might be useful in minimizing the calculation complexity when extracting further features.

Using 3-fold cross validation for the optimal MLP topography produced a 3 layer MLP, with 8 hidden nodes in the hidden layer. A Monte Carlo simulation for 500 runs produced a training accuracy of 99.9%, and a testing accuracy of 97.8%. While these results are highly encouraging, it is recognized that the method needs to be further tested in a larger sample population.

The method described here can readily be implemented for large scale screening of severe SDB cases, and offers a means to health care providers for screening patients suspected of having SDB at reduced cost.

V. Conclusion

By combining optimal HRV and RPE textural features as inputs to an MLP classifier, higher accuracy classification of normal and SDB event in 15-minute epochs was achieved, surpassing the accuracy achieved by RR optimal features alone. The textural features presented here have high potential in increasing the accuracy of the detection of SDB events using a single ECG channel.

References

[1] T. Young, et al, "Sleep Disordered Breathing and Mortality: Eighteen-Year Follow-up of the Wisconsin Sleep Cohort", *Sleep*, 31(8), pp. 1071-1078, 2008.

[2] T. Penzel, J. McNames, P. de Chazel, B. Raymond, A. Murray, and G. Moody, "Systematic Comparison of Different Algorithms for Apnoea Detection based on Electrocardiogram Recordings", *Medical and Biological Engineering and Computing*, 40, pp. 402-407, 2002.

[3] P. de Chazal, et al, "Automated Processing of the Single-Lead Electrocardiogram for the Detection of Obstructive Sleep Apnoea", *IEEE Transactions on Biomedical Engineering*, 50 (6), pp. 686-696, 2003.

[4] M. A. Al-Abed, K. Behbehani, J. Burk, E. Lucas, and M. Manry, "Cross correlation and scatter plots of the heart rate variability and R-Peak Envelope as features in the detection of obstructive sleep apnea," *Conference Proceedings, 30th Annual International Conference of the IEEE EMBS*, pp.3488 – 3491, 2008.

[5] S. Vijendra, K. Behbehani, E.A. Lucas, J.R. Burk, D.N. Burli, and D.H. Dao, "The Use of R-wave morphology in the Detection of Sleep-Disordered Breathing using the Electrocardiogram – A Comparison between Leads," *Conference Proceedings, 26th Annual International Conference of the IEEE EMBS*, 2, pp.3881 - 3884 2004.

[6] M. F. Hilton, R.A. Bates, K.R. Godfrey, M.J. Chappell, and R.M. Cayton, "Evaluation of Frequency and Time-Frequency Spectral Analysis of Heart Rate Variability as a Diagnostic Marker of Sleep Apnoea Syndrome", *Medical and Biological Engineering and Computing*, 37, pp. 760-769, 1999.

[7] M. A. Al-Abed, K. Behbehani, J. Burk, E. Lucas, and M. Manry, "A method to detect obstructive sleep apnea using neural network classification of time-frequency plots of the heart rate variability," *Conference Proceedings, 29th Annual International Conference of the IEEE EMBS*, pp.6101 – 6104, 2007

[8] D. Clausi and R. Jobanputra "Texture Analysis Using Gaussian Weighed Gray Level Co-Occurrence Probabilities," *First Canadian Conference on Computer and Robot Vision*, pp. 51-57, 2004.

[9] J. Shuttleworth, A. Todman, R. Naguib, B. Newman, "Colour Texture Analysis using Co-occurrence Matrices for Classification of Colon Cancer Images", *IEEE Canadian Conference on Electrical and Computer Engineering*, 2, pp.1134-1139, 2002.

[10] D. Clausi and Y. Zhao, "Grey Level Co-occurrence Integrated Algorithm (GLCIA): a Superior Computational Method to Rapidly Determine Co-occurrence Probability Texture Features", *Computers & Geosciences*, 29, pp. 837-850, 2003.

[11] R.M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, pp. 610-619, 1972.

[12] M. Manry. Jiang Li, Pramod L. Narasimha, and Changhua Yu, "Feature Selection Using a Piecewise Linear Network," *IEEE Transactions on Neural Network*, vol. 17, 2006

[13] K. Fukunaga, Introduction to Statistical Pattern Recognition. *Boston*: Academic Press, 1990.