Automated Processing of the Single-Lead Electrocardiogram for the Detection of Obstructive Sleep Apnoea

Philip de Chazal*, *Member, IEEE*, Conor Heneghan, *Member, IEEE*, Elaine Sheridan, Richard Reilly, *Member, IEEE*, Philip Nolan, and Mark O'Malley, *Senior Member, IEEE*

Abstract—A method for the automatic processing of the electrocardiogram (ECG) for the detection of obstructive apnoea is presented. The method screens nighttime single-lead ECG recordings for the presence of major sleep apnoea and provides a minute-by-minute analysis of disordered breathing. A large independently validated database of 70 ECG recordings acquired from normal subjects and subjects with obstructive and mixed sleep apnoea, each of approximately eight hours in duration, was used throughout the study. Thirty-five of these recordings were used for training and 35 retained for independent testing. A wide variety of features based on heartbeat intervals and an ECG-derived respiratory signal were considered. Classifiers based on linear and quadratic discriminants were compared. Feature selection and regularization of classifier parameters were used to optimize classifier performance. Results show that the normal recordings could be separated from the apnoea recordings with a 100% success rate and a minute-by-minute classification accuracy of over 90% is achievable.

Index Terms—Electrocardiogram, estimated respiration, heart rate variability, pattern recognition, sleep apnoea.

I. INTRODUCTION

SLEEP apnoea [1] is commonly defined as the cessation of breathing during sleep. If breathing does not stop but the volume of air entering the lungs with each breath is significantly reduced, then the respiratory event is called a hypopnoea. Clinicians usually divide sleep apnoea into three major categories—obstructive, central, and mixed apnoea. Obstructive sleep apnoea (OSA) is characterized by intermittent pauses in breathing during sleep caused by the obstruction of the upper airway. The airway is blocked at the level of the tongue or soft palate, so that air cannot enter the lungs in spite of continued efforts to breathe. This is typically accompanied by a reduction in blood oxygen saturation and leads to wakening from sleep in order to breathe. Central sleep apnoea (CSA) is a neurological condition which causes the loss of all respiratory effort during sleep and is also usually marked by decreases in blood oxygen saturation. With CSA the airway is not necessarily obstructed. Mixed sleep apnoea combines components of both CSA and OSA, where an initial failure in breathing efforts allows the upper airway to collapse.

Currently, a definitive diagnosis of sleep apnoea is made by counting the number of apnoea and hypopnoea events over a given period of time (e.g., a night's sleep). Averaging these counts on a per-hour basis leads to commonly used standards such as the apnoea/hypopnoea index (AHI) or the respiratory disturbance index (RDI) [1]. Standard criteria have been developed to categorize the severity of the apnoea. Accurate identification of an apnoea or hypopnoea event requires direct measurement of upper airway airflow and of respiratory effort. Definitive measurements of these quantities require a subject to use an esophageal balloon and wear a full-faced mask which, being invasive, disturbs sleep. Consequently, this method is not often used and instead the less invasive polysomnogram [1] is used.

The polysomnogram requires the recording of electroencephalogram, electrooculogram, and electromyogram to determine sleep stages, oronasal airflow, and chest wall abdominal wall movements for respiratory effort, and oxygen saturation to monitor the effect of respiration and the electrocardiogram (ECG) for heart rate monitoring and arrhythmia screening. Typically, a full night's sleep is observed before a diagnosis is reached and in some subjects a second night's recording is required. The polysomnogram provides indirect evidence of apnoea and hypopnoea events and studies have shown there is a significant intraobserver and interobserver variability when used to identify these events [2].

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 apnoea is widely under diagnosed [3]. Hence, techniques which provide a reliable diagnosis of sleep apnoea with fewer and simpler measurements and without the need for a specialized sleep laboratory may be of benefit.

This paper outlines a scheme for providing a reliable diagnostic measure of OSA based solely on measurement of the ECG. The motivation for this study arose from a joint initiative of *Physionet* and the organizers of the *2000 Computers in Cardiology Conference*, in which they conducted a competition

*P. de Chazal is with the Department of Electronic and Electrical Engineering, University College Dublin, Belfield Dublin 4, Ireland (e-mail: philip@ee.ucd.ie).

C. Heneghan, R. Reilly, and M. O'Malley are with the Department of Electronic and Electrical Engineering, University College Dublin, Belfield Dublin 4, Ireland.

E. Sheridan is with the Motorola, Cork, Ireland.

P. Nolan is with the Physiology Department, University College Dublin, Earlsfort Terrace Dublin 2, Ireland.
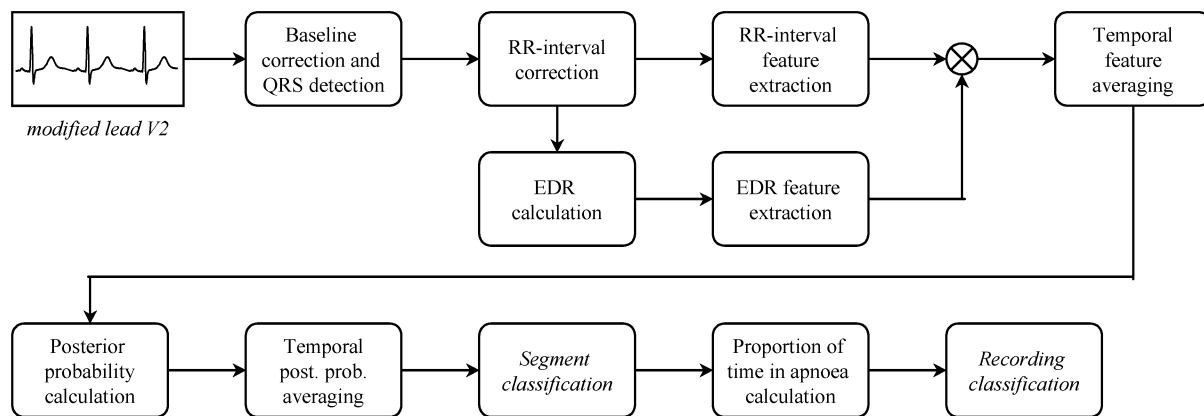
Fig. 1. Schematic representation of an automated system for the detection of OSA using single-lead ECG measurements. The processing steps include detection of QRS complexes leading to an RR-interval data set and the calculation of an ECG-derived respiratory (EDR) signal. Using these RR-intervals and the EDR signal, features were constructed for each 1-min segment of recording. Using a variety of classifier techniques, each minute can then be assigned a probability of representing apnea or normal breathing, to produce a "*Segment classification.*" Combining these classifications over a longer period of time, results in a "*Recording classification,*" which can be used by a physician as a diagnostic measure.

[4], [5] to determine the efficacy of ECG-based methods for apnoea detection using a large, well-characterized and representative database. Computers in Cardiology is an annual IEEE-sponsored conference. PhysioNet is a web-based library of physiologic data and analytic software sponsored by the US National Institutes of Health's National Center for Research Resources [6] and it provided free access to the database of ECGs.

Our group participated in the competition and arising from this work we have developed an automated system for detection of OSA using ECG measurements (Fig. 1). The system provides two outputs. The first output is a minute-by-minute sequence of classifications of "normal" or "apnoea." The second output provides an overall summary of the presence of clinically significant apnoea and it is derived on the basis of the minute-by-minute annotation sequence.

This paper is organized as follows. Section II describes the database. Section III details the methodology of the paper including preprocessing of QRS data, derivation of an ECG-derived respiratory signal, choice of ECG feature sets, choice of classifier model and methods for improving classifier performance. In Section IV, we detail the results of our automated system. In Section V, a discussion of our results is presented and we draw some conclusions regarding the potential usefulness of this system in Section VI.

II. DATABASE

The database was provided by the competition organizers [4], [7]. It contains 70 nighttime recordings of a single continuous ECG signal of approximately 8 hours duration. The ECG recordings were extracted from a larger database of simultaneously recorded polysomnogram measurements provided by Philipps-University, Marburg, Germany. The ECG signal was sampled at 100 Hz, with 16-bit resolution, with one sample bit representing 5 μV. The standard sleep laboratory ECG electrode positions were used (modified lead V2).

The database does not contain episodes of pure central apnoea or of Cheyne–Stokes respiration; all apnoeas in these recordings are either obstructive or mixed. The subjects of these recordings were men and women between 27 and 63 years of age (mean: 43.8 ± 10.8 years) with weights between 53 and 135 kg (mean: 86.3 ± 22.2 kg). The sleep recordings originated from 32 subjects (25 men, 7 female) that were recruited for previous studies on healthy volunteers and patients with OSA. Four subjects contributed a single recording each, 22 subjects contributed two recordings each, two subjects contributed three recordings each and four subjects contributed four recordings each. The duration of the recordings varied between 401 and 578 min (mean: 492 ± 32 min). The AHI ranges from 0 to 93.5 in these recordings. The initial scoring of apnoeas and hypopnoeas in the recordings was done according to standard criteria [1]. One expert rescored the database as follows: The polysomnogram recordings were divided into a set of 1-min segments. Each segment was annotated on the basis of respiration and oxygen saturation signals using amplitude criteria for airflow and desaturation and if at any time during that minute there was evidence of sleep apnoea the segment was classified as "apnoea;" otherwise it was classified as "normal." The total number of recorded minutes classified in this manner was 34 313. This provided a minute-by-minute sequence of gold-standard annotations for the ECG recordings. Segments containing hypopnoeas (defined as intermittent drops in the volume of air entering the lungs on each breath below 50% of normal, accompanied by drops in oxygen saturation of at least 4% and followed by compensating hyperventilation) were also classed as apnoea.

In addition, each *recording* was screened for the presence of clinically significant apnoea and classified as "apnoea," "borderline," or "normal." The assessment was made on the basis of the number of apnoeas and the apnoea index. The following criteria were used in classifying the recordings:

- Recordings classed as apnoea contained at least 60 min with an apnoea index of ten or more and at least 100 min with apnoea during the recording. There were 40 recordings in this class.
- Recordings classed as borderline contained at least 60 min with an apnoea index of five or more and between five and

99 min with apnoea during the recording. There were ten recordings in this class.

- Recordings classed as normal (or control) contained fewer than 5 min with apnoea during the recording. There were 20 recordings in this class.

The database was divided into two sets each containing 35 recordings. The first set (released-set) was used to develop our classification algorithms and the second set (withheld-set) was used by the competition organizers to provide an independent performance assessment. The recordings were split so each set had 20 apnoea, 5 borderline, and 10 normal recordings. The 1-min segments were almost equally balanced between the released-set ($n = 17\,045$) and the withheld-set ($n = 17\,268$). Seventeen subjects contributed recordings to both the released and withheld-set. Eight subjects' recordings were only in the withheld-set, while the recordings of the remaining seven subjects were in the released-set only.

III. METHODS

A. Classification Philosophy

Rather than trying to use our knowledge of the physiological underpinnings of OSA, we adopted a standard "black-box" pattern recognition approach using supervised learning. We considered various representations of the ECG using different features. Previous studies had shown that features based on the timing of QRS complexes [8]–[11] and the amplitude of the ECG [12]–[14] might be useful for apnoea identification. Both types of features were considered for classification. Two different classifier methods were explored and the classifier performance was optimized using the available ECG data and our selected features. The classifier methods used in this study were linear and quadratic discriminants [15].

B. QRS Detection

All the features used in this study required QRS detection times. A "QRS detection time" is loosely defined as the time of occurrence of the QRS complex in an ECG signal. QRS detection times were generated automatically for all recordings using an algorithm described in [16]. This algorithm provides detection times, which occur close to the onset of the QRS complex. A second set of QRS detections was formed after manual verification and correction of the first set. The two sets were used separately to generate features to determine the importance of QRS detection on classification accuracy.

C. RR-Interval Correction

RR-intervals were defined as the interval between successive QRS detection points. Due to poor signal quality and errors in the automatically generated QRS detections, the RR-interval sequences generated from both sets of QRS detection times contained physiologically unreasonable times. A first preprocessing step prior to calculating the ECG features was to calculate a corrected RR-interval sequence where all intervals were physiologically reasonable. The following automatic algorithm was developed for this purpose.

Suspect RR-intervals were found by applying a median filter of width five to the sequence of RR-intervals. This provided a robust estimate of the expected value for each RR-interval. Significant variations from this expected value led to it being flagged as a suspect RR-interval. Suspect RR-intervals could be due to either spurious QRS detections, or missed QRS complexes.

Spurious QRS detections were found by comparing the sum of adjacent RR-intervals with the robust RR-interval estimate. If this sum was numerically closer to the robust estimate than either of the individual RR-intervals then a spurious detection was deemed to be present. The two RR-intervals were merged to form a single RR-interval.

Conversely, we determined heuristically that if an RR-interval was a factor of 1.8 times or greater than the robust estimate then it was probable that one or more QRS complexes were missed. To estimate (interpolate) the times of the missing QRS complexes the RR-interval was divided by the sequence of integers 2, 3, 4, ..., until it best matched the robust estimate of the RR-interval. The single RR-interval was then subdivided by the appropriate integer to form a series of new detections.

D. ECG-Derived Respiratory Signal

During the breathing cycle, the body-surface ECG is influenced by electrode motion relative to the heart and by changes in thoracic electrical impedance as the lungs fill and empty with air. The effect is most obviously seen as a slow modulation of the ECG amplitude at the same frequency as the breathing cycle [12]–[14]. To access this signal the original ECG signal was filtered with two median filters to remove the baseline wander. The original ECG signal was processed with a median filter of 200-ms width to removed QRS complexes and P waves. The resulting signal was then processed with a median filter of 600-ms width to remove T waves. The signal resulting from the second filter operation contained the baseline of the ECG signal, which was then subtracted from the original signal to produce the baseline corrected ECG signal.

A sample point of an ECG-derived respiratory signal (EDR) was then obtained by calculating the area enclosed by the baseline corrected ECG in the region 100 ms beyond the QRS detection point.

E. Feature Sets

The processing steps outlined above resulted in discrete index sequences of the RR-intervals and the EDR signal. Based on these, a large set of features that could potentially be used for classification were considered. Since the database classifications were provided for 1-min segments of data, features were generated for each of these 1-min segments. The features considered in this study were as follows.

- mean RR-interval;
- standard deviation of the RR-interval;
- the first five serial correlation coefficients of the RR-intervals [17];
- the NN50 measure (variant 1), defined as the number of pairs of adjacent RR-intervals where the first RR-interval exceeds the second RR-interval by more than 50 ms;

- the NN50 measure (variant 2), defined as the number of pairs of adjacent RR-intervals where the second RR-interval exceeds the first RR-interval by more than 50 ms;
- two pNN50 measures, defined as each NN50 measure divided by the total number of RR-intervals;
- the SDSD measures, defined as the standard deviation of the differences between adjacent RR-intervals;
- the RMSSD measure defined as the square root of the mean of the sum of the squares of differences between adjacent RR-intervals;
- the Allan factor $A(T)$ evaluated at a time scale T of 5, 10, 15, 20, and 25 s where the Allan factor is defined as $A(T) = E\{[N_{i+1}(T) - N_i(T)]^2\}/2E\{N_{i+1}(T)\}$, $N_i(T)$ is the number of QRS detection points occurring in a window of length T stretching from iT to $(i + 1)T$ and E is the expectation operator [18];
- interval-based power spectral density (PSD) of the RR-intervals [19];
- mean EDR amplitude;
- standard deviation of the EDR amplitude;
- the PSD of the EDR signal.

A useful summary of RR-interval based measures and their uses is contained in [20]. It is worth noting that none of the measures listed above consider the morphology of the ECG. It is implicitly assumed that the processes leading to apnoea occur at a location external to the heart and, thus, do not directly affect the generated cardiac potentials.

Since our results show that the interval-based RR-interval and the EDR PSDs are among the most useful features for classification, it is worth carefully defining how these quantities are calculated. The interval-based RR-interval PSD was calculated in the following way. A sequence of RR-intervals was associated with each 1-min segment. The index for this sequence was beat number, not time. The mean RR-interval for that segment was removed from each value, to yield a zero-mean sequence. The sequence was zero-padded to length 256 and the fast Fourier transform (FFT) was taken of the entire sequence. The magnitudes of the FFT coefficients were squared to yield a periodogram estimate of the PSD, which had high variance. Averaging of four adjacent frequency bins yielded a 64-point PSD estimate of which only the first 32 points were used as features (due to the symmetry of the upper and lower PSD point estimates). The x axis has units of cycles/interval. The EDR PSD was calculated in an identical fashion, with the spectral variable also defined as cycles/interval.

The features have been divided into labeled sets. The RR set contains 52 features derived from the RR-intervals comprising the following: 32 features from the interval-based PSD, the mean and standard deviation of the RR-intervals per segment, the mean and standard deviation of the RR-intervals per recording, the first five serial correlation coefficients, one SDSD feature, one RMSSD feature, two NN50 features, two pNN50 features and five Allan variance measures calculated at different time scales. The EDR set contains 36 features derived from the EDR signal. The PSD provided 32 features, two features were the mean and standard deviation of the EDR areas per segment and two features were the mean and standard deviation of the EDR areas per recording. Both of these sets were combined to form the $\mathrm{RR} + \mathrm{EDR}$ set.

F. Classifier Models

A supervised training technique was used to derive all classifiers [21]. In this study, linear discriminant (LD) and quadratic discriminant (QD) models were considered. Optimization of both models was achieved by the method of maximum likelihood (ML). Both models provide a posterior probability estimate of each class. A final classification is obtained by choosing the class with the highest posterior probability estimate.

Training of the LD and QD classifiers using "plug-in" ML estimates of classifier parameters proceeded as follows [21]. Let $\mathbf{x}$ be a column vector containing d feature values. Assume that we wish to assign $\mathbf{x}$ to one of c possible classes. A total of N feature vectors are available for training the classifier. The number of feature vectors available for training for class k is N_k and, hence

$$N = \sum_{k=1}^{c} N_k. \qquad (1)$$

The nth feature vector for training in class k is designated as $\mathbf{x}_{nk}$. Training of both models involves determining the class-conditional mean vectors $\boldsymbol{\mu}_k$ using

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N_k} \mathbf{x}_{nk}. \qquad (2)$$

For an LD classifier the common covariance matrix Σ is calculated using

$$\boldsymbol{\Sigma} = \frac{1}{N - c} \sum_{k=1}^{c} \sum_{n=1}^{N_k} (\mathbf{x}_{nk} - \boldsymbol{\mu}_k)(\mathbf{x}_{nk} - \boldsymbol{\mu}_k)^T. \qquad (3)$$

For a QD classifier the class-conditional covariance matrices, Σ_k, are calculated using

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k - 1} \sum_{n=1}^{N_k} (\mathbf{x}_{nk} - \boldsymbol{\mu}_k)(\mathbf{x}_{nk} - \boldsymbol{\mu}_k)^T. \qquad (4)$$

To classify a feature vector $\mathbf{x}$, values are assumed for the prior probabilities π_k and the discriminant value y_k for each class is calculated using

$$y_k = -\frac{1}{2}\boldsymbol{\mu}_k^T \Sigma^{-1} \boldsymbol{\mu}_k + \boldsymbol{\mu}_k^T \Sigma^{-1} \mathbf{x} + \log(\pi_k) \qquad (5)$$

for LD classifiers and

$$y_k = -(\mathbf{x} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) + 2\log(\pi_k) - \left|\Sigma_k^{-1}\right| \qquad (6)$$

for QD classifiers.

In this paper, the estimated posterior probabilities, $P(k \mid \mathbf{x})$, are used. They are easily calculated from the discriminant values using

$$P(k \mid \mathbf{x}) = \frac{\exp(y_k)}{\sum_{l=1}^{c} \exp(y_l)} \qquad (7)$$

The final class is the class with the highest posterior probability or, equivalently, the highest discriminant value.

G. Transformations

The classifier models considered in this study implicitly assume that the feature data have a class-dependent Gaussian distribution. Classifier performance will be degraded when the actual feature statistics differ significantly from this assumption. Therefore, where appropriate, a transformation was applied to the features so that the histogram of the transformed feature more closely approximated a Guassian distribution. A logarithmic transform was applied to all standard deviation, PSD, RMSSD, and SDSD features. A square-root transform was applied to the Allan variance features.

H. Classification Performance Estimation

Classifier performance has been estimated using two methods in this paper. In the first method, the cross-validation scheme [21], [22] was applied to the 35 recordings in the released-set. This data was divided into 35 folds with each fold containing the data from one recording. This method was used to assess the performance of different candidate classifiers and identify the best performing classifiers for a second performance estimation.

The second method of performance estimation was to process the 35 recordings in the withheld-set and submit the annotations to the competition organizers for assessment. They responded with a single figure representing the total number of correct classifications. Fig. 2 shows how the data was divided in this study into separate sets for performance assessment.

In this study, we considered the overall classification accuracy, sensitivity and specificity defined as follows. The overall accuracy is the percentage of total segments correctly classified. The sensitivity is the percentage of apnoea segments correctly classified and the specificity is the percentage of normal segments correctly classified.

I. Improving the Classifier Performance

The performance of most classifier training algorithms is degraded when there are too many parameters of the classifier to be estimated relative to the size of the training data. This occurs because of the so-called "curse of dimensionality" [23] which, applied to classification problems, refers to the exponential growth of training data required to successfully train a classifier with the number of classifier parameters. In these cases, techniques are available that may improve classification performance. Two techniques applicable to LD and QD classifiers were used in this study. The first was to use feature selection to find a subset of the available features with good classification performance for the classifier model used. The second method was to regularize the covariance matrices.

When attempting to improve the classifier performance two nested loops of cross-validation were used to obtain unbiased estimates of classifier performance (see Fig. 2). The inner loop was used to measure the accuracy of the different classifier configurations. Although these performance figures used test-set values, these values are optimistically biased as they were used during the configuration selection process.
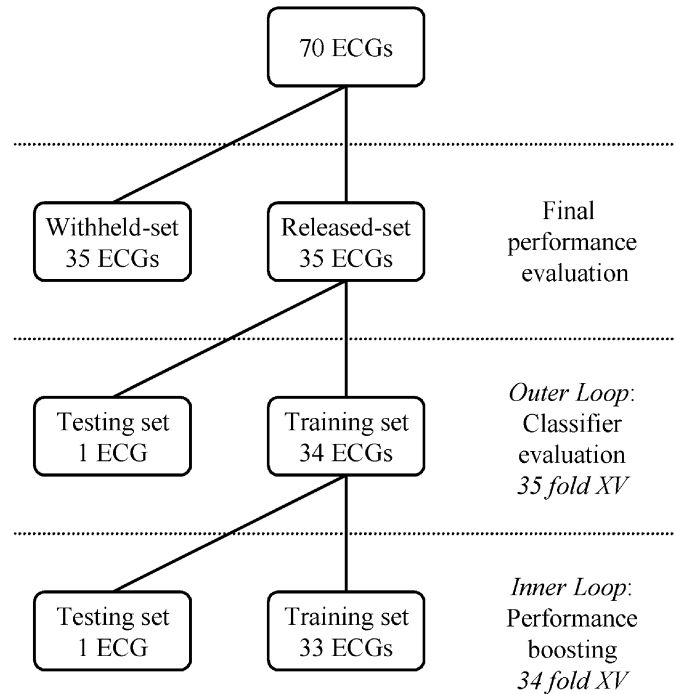

Fig. 2. Division of the data into training and testing sets for performance boosting and classifier evaluation using the cross-validation scheme. Final performance evaluation is performed on an independent set (the withheld-set).

The outer cross-validation loop was used to test the classifier configuration that was selected by the inner loop. As before, the outer loop used 35-fold cross-validation. The inner-loop of cross-validation further divided the 34 recordings available for training from the outer loop into 34 folds each containing data from one subject.

1) Feature Selection: Given a set of features it is often possible to find a subset with better classification performance than the full set of features. However, the interaction of features is difficult to determine so, in general, the only way to optimize the classification performance of a set of features is to determine the performance of all possible subsets and choose the best performing subset. This approach is often computationally intractable as feature sets of size d generate 2^d subsets. To ease the computational burden, a heuristic search is used to search the feature set space so that subsets with probable good performance are evaluated first. The simplest procedures are the stepwise searches [24] but they are not guaranteed to find the optimal subset. In this paper, the best-first search procedure [22] has been used, which, although slightly more complicated than the stepwise searches will (eventually) find the optimal subset.

2) Regularizing the Covariance Matrices: For an LD classifier the covariance matrix may be "shrunk" toward the identity matrix $\mathbf{I}$ using [21]

$$\Sigma(\alpha) = (1 - \alpha)\Sigma + \alpha\mathbf{I} \text{ with } 0 \leq \alpha \leq 1 \qquad (8)$$

and substituting $\Sigma(\alpha)$ for Σ in (5). However, this is only appropriate if the training data has been rescaled so that the variance of each feature is equal to unity. When $\alpha = 1$, then $\Sigma(\alpha) = \mathbf{I}$
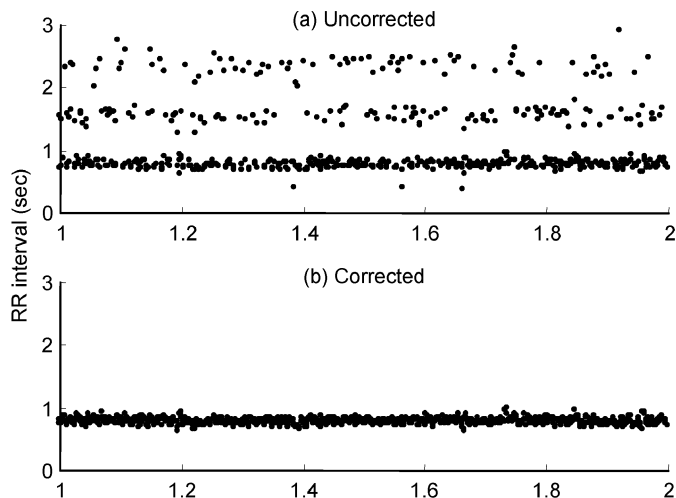
Fig. 3. An example of RR-intervals calculated from the (a) uncorrected and (b) corrected automatically determined QRS detection times for the first to second hour of recording 2 of the released-set.

TABLE I
THE CLASSIFICATION ACCURACY (%) FOR DIFFERENT LEVELS OF TEMPORAL INPUT FEATURE AND POSTERIOR PROBABILITY AVERAGING

		Posterior probabilities: *number of segments averaged*									
		1	2	3	4	5	6	7	8	9	10
Features: *number of segments averaged*	1	85.1	87.6	87.7	88.4	88.4	88.5	88.3	88.2	88.1	88.0
	2	87.6	88.7	88.8	89.2	89.3	89.3	89.2	89.2	89.0	88.9
	3	88.8	89.4	89.5	89.8	89.9	90.0	90.0	89.7	89.6	89.5
	4	89.2	89.6	89.6	89.7	89.8	89.8	89.7	89.6	89.6	89.5
	5	89.4	89.6	89.7	89.8	89.8	89.8	89.8	89.7	89.7	89.5
	6	89.4	89.5	89.6	89.7	89.7	89.6	89.5	89.5	89.5	89.3
	7	89.4	89.5	89.4	89.5	89.5	89.5	89.5	89.5	89.4	89.3
	8	89.3	89.3	89.2	89.2	89.3	89.3	89.3	89.2	89.1	89.1
	9	89.1	89.2	89.2	89.2	89.1	89.2	89.1	89.1	89.0	89.0
	10	88.7	88.7	88.8	88.8	88.8	88.8	88.8	88.8	88.7	88.7

All results are the test-set performance determined using 35-fold cross-validation on the released-set for the LD classifier processing the RR+EDR set. Features derived from the manually verified QRS detection times.

and a special case occurs in which the features are assumed to be statistically independent (and, hence, have no covariance).

For the QD classifier each covariance matrix can be "shrunk" toward the common covariance matrix [21] using

$$\Sigma_k(\alpha) = \frac{(1-\alpha)N_k\Sigma_k + \alpha N\Sigma}{(1-\alpha)N_k + \alpha N} \text{ with } 0 \leq \alpha \leq 1, \quad (9)$$

and substitute these new matrices in for the Σ_ks in (6). When $\alpha = 1$ then a LD classifier results as all $\Sigma_k(\alpha) = \Sigma$. In practice, for both LD and QD classifiers, various values of α in the range 0 to 1 can be evaluated and the classifier performance determined. A value of α that optimizes classifier performance can then be chosen.

J. The Temporal Relationship of Segments

Observation of sequences of the minute-by-minute classifications revealed that there was a temporal association of the classifications. For example, if the 11th–13th 1-min segments were classed as normal, then there was a high probability that 10th and 14th 1-min segments were also classed as normal. In light of this observation, two techniques were found to boost classification performance. The first was to average the posterior probabilities calculated for a segment with the posterior probabilities obtained from chronologically close segments. The second was to average the input features of a segment with the input features from chronologically close segments.

K. Overall Screening For Clinically Significant Apnoea

The screening for clinically significant apnoea for each recording was achieved by calculating the average number of minutes per hour (min/h) spent in apnoea from the minute-by-minute classifications. A threshold value was determined to best separate the normal and apnoea recordings.

IV. RESULTS

A. QRS Detections and RR-Interval Correction

The two sets of QRS detection times were compared to determine the accuracy of the automatically generated set. If an automatically generated QRS detection time was within 100 ms of a manually verified QRS detection time it was regarded as correct. Using this method, 98.6% of the automatically generated QRS detections were correct.

Fig. 3(a) shows the RR-intervals calculated from the automatically generated QRS detections determined from a very noisy recording. Fig. 3(b) shows the same recording after RR-interval correction. A visual inspection of the RR-intervals associated with the corrected QRS detections for all recordings showed a significant reduction in the number of physiologically unreasonable RR-intervals.

B. Classifier Performance Measured Using the Released-Set

There were 17 045 segments in the released-set with 10 531 (61.78%) labeled normal and 6514 (38.22%) labeled apnoea. All results in this section were calculated using the cross-validation scheme. Features were derived from the manually verified QRS detection times.

1) Temporal Averaging: Different length equally weighted temporally averaging filters were applied to the input features and posterior probabilities and the effect on classification performance observed. All input features and the LD classifier were used. The filter length was varied between one and ten segments and the results are shown Table I. The accuracy was maximized when using temporal averaging of three segments of features and six segments of posterior probabilities. The temporal averaging was set at these values for the rest of the study.

2) Temporal Averaging—Performance of Different Feature Sets: Table II(a) shows the classification results of different feature sets using the standard classifiers. Table II(b) shows the classification results after feature selection was applied to the three feature sets. For feature selection the best first search was
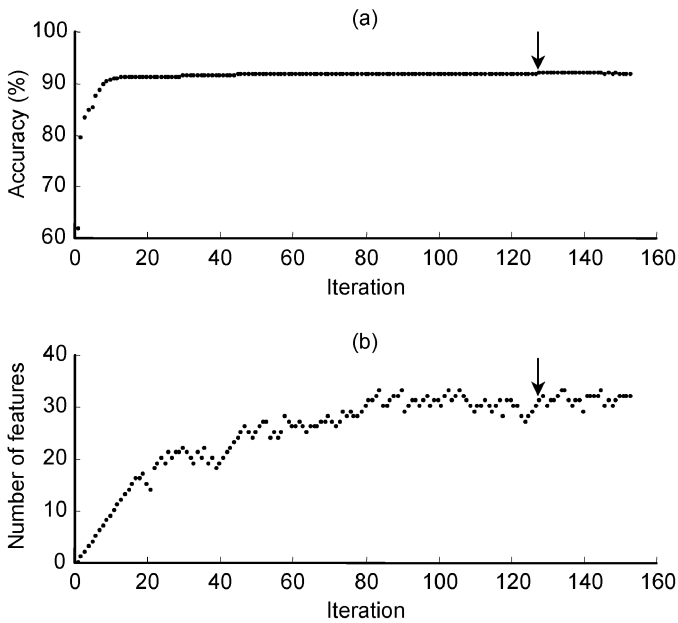
Fig. 4. Best-first search algorithm applied to feature selection using the LD classifier with the $RR + EDR$ set using the 35 ECG recordings of the released-set. (a) Accuracy of the best performing features set versus iteration number. (b) Number of features in the chosen subset. The search algorithm was terminated after 25 consecutive iterations with no improvement in accuracy. The point of maximum classification accuracy is shown with an arrow.

terminated after 25 iterations with no improvement in accuracy. The starting point was the empty feature set. Fig. 4(a) shows an example plot of the accuracy of the best set at each iteration step and Fig. 4(b) shows the number of features in the best set. Note that the values in the third column of Table II(b) are noninteger as each is the average of individual results from the different data splits of the cross-validation process. Table II(c) shows the classification results obtained through regularization of the covariance matrices.

C. Classifier Performance Measured On the Withheld-Set

The assessment of classifier performance on the withheld-set was provided by the competition organizers and contains only overall accuracy assessments (i.e., breakdown by sensitivity and specificity is not known).

The best performing classifiers (as measured with cross-validation on the released-set) were retrained using all recordings of the released-set and then tested using all recordings of the withheld-set. Four classifiers were validated [see Table III(a)] and all processed the $RR + EDR$ set. The features for the first three classifiers were calculated using the manually verified QRS detection times. These classifiers were the best performing standard classifier (an LD classifier), the best performing classifier after feature selection (an LD classifier) and the best classifier after regularization of the covariance matrices (a QD classifier). The fourth classifier was the best of the above three classifiers (as measured on the withheld-set) retrained using features calculated using the automatically generated QRS detection times. This last classifier is a fully automatic system.

D. Automatic Screening of Recordings For the Presence of Clinically Significant Sleep Apnoea

Using the fully automatic system described above the number of minutes per hour spent in apnoea calculations were generated for the released-set. Fig. 5 shows the results of these calculations. A threshold of 5 min/h, allows complete separation of the subjects classed as "normal" from those classed as "apnoea." Table III(b) presents the classification results on the released- and withheld-sets for discriminating between normal and apnoea subjects using the above threshold value. The withheld-set accuracy assessment was provided by competition organizers.

E. PSD Features

The PSD features made up the majority of our features and contributed significantly to the overall classification performance of our system. Fig. 6(a) shows the class-dependent PSD values of the RR signals over the normal and apnoea segments of the released-set. For each frequency bin the median value is shown along with vertical bars indicating the range of the 25%–75% values. As both spectra have a log-normal amplitude distribution a log scale has been used on the y axis. Fig. 6(b) shows the class-dependent PSDs of the EDR signal. Again a log scale has been used on the y axis.

V. DISCUSSION

A. Released-Set

In the following, all the performance figures discussed are the test-set figures from Tables I and II.

1) Temporal Averaging: Temporal averaging significantly boosted the classification performance. Using no temporal averaging, the classification accuracy was 85.1%. Introducing temporal averaging of the posterior probabilities resulted in a maximum accuracy of 88.5% and occurred when six segments were averaged. Temporal averaging of features produced a maximum accuracy of 89.4% and occurred when averaging six segments. The maximum accuracy of 90.0% occurred when using both types of temporal averaging simultaneously. It occurred when features were averaged over three segments and posterior probabilities were averaged over six segments.

2) Feature Selection and Covariance Regularization: The results in Table II(b) and (c) show feature selection and covariance regularization slightly reduced the classification accuracy of the best performing standard classifier. The best overall accuracy (90.0%) was the standard LD classifier processing the $RR + EDR$ set. The sensitivity was 86.4% and the specificity was 92.3%. Feature selection applied to this classifier reduced the average number of features selected to 27.0 and the overall accuracy to 89.5%. When covariance regularization was applied the overall accuracy was 89.8%.

For the LD classifier, feature selection resulted in slightly lower accuracy figures in all feature sets when compared to the standard classifier results. The reduction ranged from 0.3% for the RR set to 1.3% for the EDR set. In all cases, the average number of features was much less after feature selection. The

TABLE II
CLASSIFICATION PERFORMANCE ON THE RELEASED-SET FOR THE LD AND QD CLASSIFIERS AFTER
(A) NO OPTIMISATION, (B) FEATURE SELECTION, AND (C) COVARIANCE REGULARISATION

	Classifier	Feature Set	Number of Features	Regularisation Value (α)	Average of training set results			Average of testing set results		
					Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
(a)	LD	RR	52	0	89.2	83.2	92.9	85.8	76.9	91.3
	LD	EDR	36	0	87.7	81.9	91.3	86.2	79.6	90.4
	LD	RR+EDR	88	0	92.3	89.6	94.0	90.0	86.4	92.3
	QD	RR	52	0	83.0	94.1	76.2	80.2	87.9	75.5
	QD	EDR	36	0	89.1	86.2	90.9	83.6	74.1	89.5
	QD	RR+EDR	88	0	89.9	94.6	86.9	84.8	82.3	86.4
(b)	LD	RR	23.1[a]	0	89.3	84.2	92.5	85.5	76.1	91.3
	LD	EDR	9.7[a]	0	87.7	82.6	90.8	84.9	78.0	89.2
	LD	RR+EDR	27.0[a]	0	92.2	89.7	93.7	89.5	84.1	92.8
	QD	RR	18.4[a]	0	90.0	90.0	90.0	86.1	82.8	88.1
	QD	EDR	13.9[a]	0	88.8	84.9	91.3	84.3	76.1	89.4
	QD	RR+EDR	22.7[a]	0	93.2	92.8	93.4	89.5	86.1	91.6
(c)	LD	RR	52	0.02	88.9	82.8	92.6	85.9	76.5	91.8
	LD	EDR	36	0.00	87.7	81.9	91.3	86.2	79.6	90.3
	LD	RR+EDR	88	0.06	92.0	89.6	93.6	89.8	86.4	91.9
	QD	RR	52	0.86	89.3	86.2	91.3	86.3	80.0	90.3
	QD	EDR	36	0.74	88.2	83.5	91.2	86.3	79.9	90.2
	QD	RR+EDR	88	0.84	92.5	91.4	93.1	89.8	87.9	91.0

All results determined using 35-fold cross-validation. Features derived from the manually verified QRS detection times.
[a]Value is non-integer as it is the average of the results from the different data splits of the cross-validation process.

TABLE III
CLASSIFICATION PERFORMANCE ON THE RELEASED-SET AND WITHHELD-SET OF THE (A) BEST PERFORMING
MINUTE-BY-MINUTE CLASSIFIERS AND (B) SCREENING CLASSIFIER

	Classifier	Feature Set	Optimisation technique	Number of Features	Regularisation Value (α)	Released-set			Withheld-set
						Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
(a)	LD	RR+EDR[a]	Unoptimised	88	0	92.3	89.6	94.0	90.4
	LD	RR+EDR[a]	Feature selection	31	0	92.2	89.8	93.7	90.5
	QD	RR+EDR[a]	Covariance regularisation	88	0.85	92.5	91.4	93.1	90.6
	QD	RR+EDR[b]	Covariance regularisation	88	0.91	91.2	88.5	92.9	90.5
(b)	QD	RR+EDR[b]	-	-	-	100[c]	100[c]	100[c]	100[c]

All results determined by training each classifier on all data in the released-set.
[a]Features derived from the manually verified QRS detection times.
[b]Features derived from the automatically generated QRS detection times.
[c]Performance figure based on the 30 non-borderline recordings.

EDR set required an average of just 9.7 features and the RR set required an average of 23.1 features.

Feature selection applied to the QD classifier improved the performance of all feature sets relative to the standard QD classifier. The highest gain was for the RR set where the accuracy increased by 5.9% to 86.1%. The average number of features chosen with the QD classifier was lower than for the LD classifier for the RR (18.4) and $RR + EDR$ (22.7) sets, whereas the EDR set alone required more features than the LD classifier. The QD classifier has more parameters to be estimated than the LD classifier for a given feature set size. Therefore, for a fixed size training set and feature set, the QD classifier can use fewer features than the LD classifier before the "curse of dimensionality" becomes a problem.

For the LD classifier, regularization of the common covariance matrix led to optimal regularization parameters between 0.00 and 0.06, i.e., the regularized covariance matrices in all cases were very similar to the common covariance matrices from the unregularized LD classifier. For all feature sets, covariance regularization made little difference to the classification performance relative to the unregularized LD classifier. The biggest change was a reduction in performance of 0.2% for the $RR + EDR$ set.

However, for the QD classifier, covariance regularization significantly boosted performance for all feature sets relative to the unregularized classifier. The optimal regularization parameter varied between 0.74 and 0.86, thus, the common covariance matrix heavily weighted the covariance matrices.
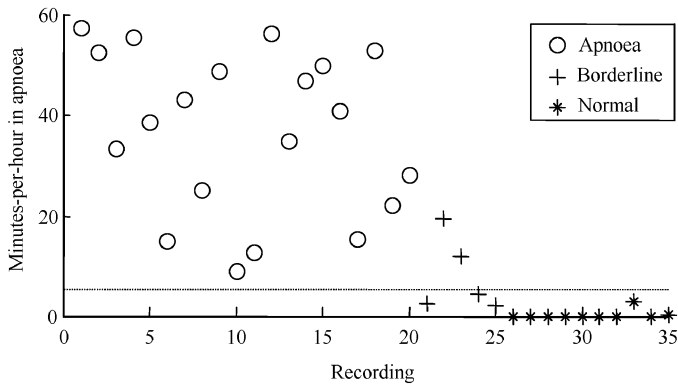
Fig. 5. The number of minutes per hour spent in apnoea as estimated by the regularized QD classifier processing all features for all recordings of the released-set. Applying a threshold of 5 min/h (as shown by the dotted line) completely separates the normal and apnoea records.

These results obtained here demonstrate that feature selection and covariance regularization can lead to an improvement in classification performance. Although they introduced a significant amount of extra numerical processing during the training of a classifier, we believe these methods are worth investigating.

3) Comparison of Feature Sets: Covariance regularization of the QD classifier resulted in the overall best performing classifier for both the RR and EDR sets. The overall accuracy figure for both these feature sets was 86.3%. It is a coincidence that these two feature sets resulted in the same overall performance as both feature sets provide separate diagnostic information for detection of apneas. The RR set provides information on heartbeat timing whereas the EDR set provides information on chest wall movement. The combined $RR + EDR$ set feature results in better classification performance than either set alone, suggesting that the RR and EDR sets provide independent classification information.

From a practical viewpoint, the result for the RR set is significant as it shows that by considering interbeat timing information *alone*, a classification rate of over 85% is achievable. Deriving interbeat timing intervals from an ECG is computationally straightforward, relatively insensitive to electrode positioning and amenable to real-time processing.

B. Withheld-Set

The independent validation of performance of the best three performing classifiers on the withheld-set resulted in similar accuracy figures to those determined on the released-set. The best accuracy result of 90.6% was from the QD classifier processing all features with regularized covariance matrices. The regularizing parameter was 0.85 so the common covariance matrix heavily weighted the covariance matrices. Feature selection reduced the $RR + EDR$ set to 31 features and resulted in the simplest classifier. The overall accuracy was 90.5%. The overall accuracy of the standard LD classifier was 90.4%.

When the QD classifier processing all features with regularized covariance matrices was retrained using the features derived from the automatically generated QRS detection times the withheld-set accuracy was 90.5%. The significance of this result is that it estimates the performance of a fully automatic system for minute-by-minute identification of disordered breathing. This result also demonstrates that the method

presented here is insensitive to imperfect QRS detection times as using the automatically generated QRS detection points in place of the manually verified QRS detection points has resulted in 0.1% decrease in overall accuracy. It is worth noting that other QRS detection schemes exist (e.g., [25]–[27]) that claim to provide better detection performance than the 98.6% we obtained with our QRS detector, although our results suggest that these improved QRS detector schemes would make little difference to the classification performance.

C. Screening

The 30 nonborderline recordings in the withheld-set were all successfully classified by our system. This is a significant result. Using only the ECG, the method presented here has successfully distinguished between subjects with no apnoea and subjects with clinically significant apnoea. This result suggests that adoption of ECG-based diagnostic tests for OSA such as proposed here could lead to a significant simplification of the process of screening for OSA.

However, when using the minutes per hour in apnoea calculation it was not possible to additionally provide perfect separation of the class labeled "borderline," although Fig. 5 shows it is apparent that the borderline recordings do mostly lie in the region between the "apnoea" and "normal" recordings.

D. PSD Plots

With reference to Fig. 6(a) the median values of all frequencies of the two RR spectra are very similar, except for the frequency range 0–0.06 cycles/interval. In this range, there is more power in the PSD of the apnoea segments than in the PSD of the normal segments and the two spectra show a reduced amount of overlap of the 25%–75% ranges. There is a small peak at 0.2 cycles/interval of the normal segment PSD which probably corresponds to the average breathing frequency. The apnoea spectrum does not show the same peak.

The median values of the EDR spectrum of the apnoea segments shown in Fig. 6(b) have more power at every frequency than the normal segments. The significance of this is that the EDR signal has, on average, higher variance from the apnoea segments than normal segments. Inspection of the 25%–75% values shows that there is a significant amount of overlap of the two spectra for the frequency range 0.06–0.5 cycles/interval. The 25%–75% values for the frequency range 0–0.06 cycles/interval shows a reduced amount of overlap. There are small peaks in both spectra, probably associated with average breathing frequency. They occur at 0.2 cycles/interval for the normal segments and 0.25 cycles/interval for the apnoea segments.

These results suggests that most of the information in the ECG relating to the presence of sleep apnoea is found at relatively low frequencies (i.e, less than 0.06 cycles/interval), corresponding approximately to time-scales on the order of 10–15 s or longer. Since apnoea and hypopnoea events by definition last for longer than 10 s, it is not surprising that the effect of apnoea is observed at these low frequencies.

VI. Conclusion

This paper has presented a method for screening single-lead ECG recordings from sleep-study patients for the presence of significant OSA and for providing a minute-by-minute
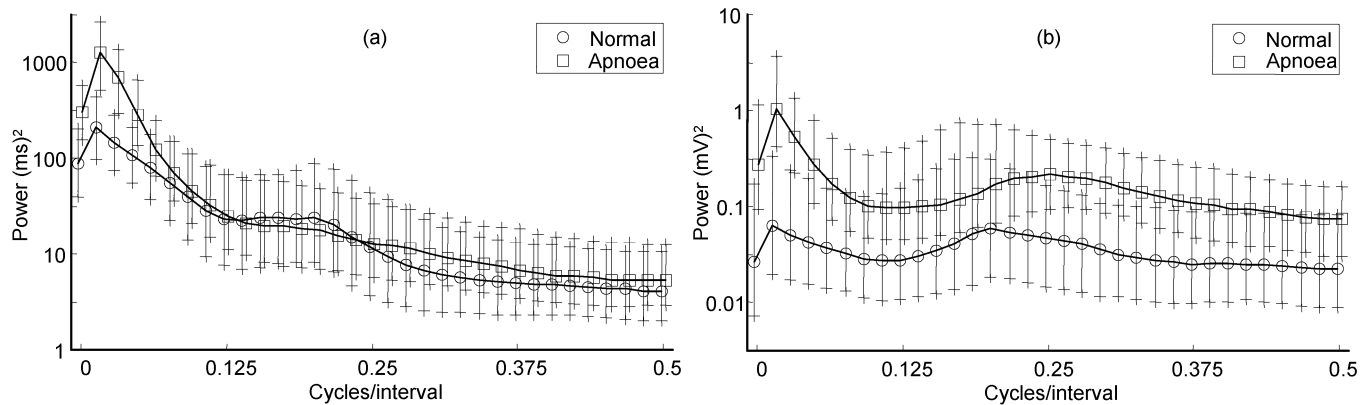
Fig. 6. (a) Median PSD derived from the RR-intervals of the normal and apnoea 1-min segments over the 35 ECG recordings of the released-set. Along the y axis, a log scale is used and the median and 25%–75% of the power is shown for each frequency bin. Along the x axis the frequency bins of the apnoea data are offset by 0.004 cycles/interval to facilitate comparison with the frequency bins of the normal data. (b) Median PSD derived from the EDR for the normal and apnoea 1-min segments over the 35 ECG recordings of the released-set. As in (a), a log scale is used on the y axis, the mean and 25%–75% of the power is shown for each frequency bin and the frequency bins of the apnoea data are offset by 0.004 cycles/interval to facilitate comparison.

analysis of disordered breathing. Classifiers based on linear and quadratic discriminants were used. Classifier performance was optimized by considering feature selection and regularization of classifier parameters. Both the time of occurrence of the QRS complex and the EDR were found to be useful features for classification.

Independent testing has shown that the method separated the normal recordings from the apnoea recordings with a 100% success rate and achieved a minute-by-minute classification rate of over 90%. By considering features based only on RR-intervals a minute-by-minute classification rate of over 85% is achievable. The final structure for the proposed classifier is computationally efficient and lends itself to real-time implementation.

The method was shown to be insensitive to imperfect QRS detections. The difference in the minute-by-minute classification accuracy when using manually verified QRS detection points compared to automatically generated QRS detection points was 0.1%.

The significance of this study is that it provides a simple scheme for diagnosis of OSA, yet its overall diagnostic accuracy is comparable with full polysomnography. ECG-based diagnostic tests for sleep apnea are ideally suited to Holter monitors as no additional hardware is required. Ideally, Holter recordings would be routinely screened for apnoea and this may allow a significant reduction in the costs associated with the detection of OSA. This would be a step toward addressing the serious public health issue caused by under-diagnosis of OSA.

REFERENCES

[1] Amer. Acad. Sleep Med. Task Force, "Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research," *Sleep*, vol. 22, no. 5, pp. 667–689, Aug. 1999.

[2] C. W. Whitney, D. J. Gottlieb, S. Redline, R. G. Norman, R. R. Dodge, E. Shahar, S. Surovec, and F. J. Nieto, "Reliability of scoring respiratory disturbance indices and sleep staging," *Sleep*, vol. 21, pp. 749–757, 1998.

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

[4] T. Penzel, "The apnea-ECG database," in *Computers in Cardiology*. Piscataway, NJ: IEEE Press, 2000, vol. 27, pp. 255–258.

[5] G. B. Moody, R. G. Mark, A. L. Goldberger, and T. Penzel, "Stimulating rapid research advances via focused competition: The computers in cardiology challenge 2000," in *Computers in Cardiology*. Piscataway, NJ: IEEE Press, 2000, vol. 27, pp. 207–210.

[6] A. L. Goldberger, A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit and physionet," *Circulation*, vol. 101, pp. e215–e220, 2000.

[7] T. Penzel, J. McNames, P. de Chazal, B. Raymond, A. Murray, and G. Moody, "Systematic comparison of different algorithms for apnea detection based on ECG recordings," *Med Biol Eng Comp*, vol. 40, pp. 402–407, 2002.

[8] C. Guilleminault, S. J. Connolly, R. Winkle, K. Melvin, and A. Tilkian, "Cyclical variation of the heart rate in sleep apnoea syndrome. Mechanisms and usefulness of 24h electrocardiography as a screening technique," *Lancet*, vol. I, pp. 126–131, Jan. 1984.

[9] T. Penzel, G. Amend, K. Meinzer, J. H. Peter, and P. von Wichert, "Mesam: A heart rate and snoring recorder for detection of obstructive sleep apnea," *Sleep*, vol. 13, pp. 175–182, 1990.

[10] 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 the sleep apnoea syndrome," *Med. Biol. Eng. Comput.*, vol. 37, no. 6, pp. 760–769, 1999.

[11] F. Roche, J. M. Gaspoz, I. Court-Fortune, P. Minini, V. Pichot, D. Duverney, F. Costes, J. R. Lacour, and J. C. Barthelemy, "Screening of obstructive sleep apnea syndrome by heart rate variability analysis," *Circulation*, vol. 100, no. 13, pp. 1411–1415, 1999.

[12] G. B. Moody, R. G. Mark, A. Zoccola, and S. Mantero, "Derivation of respiratory signals from multi-lead ECG's," in *Computers in Cardiology*. Los Alamitos, CA: IEEE Computer Society Press, 1985, vol. 12, pp. 113–116.

[13] ——, "Clinical validation of the ECG-derived respiration (EDR) technique," in *Computers in Cardiology*. Los Alamitos, CA: IEEE Computer Society Press, 1986, vol. 13, pp. 507–510.

[14] A. Travaglini, C. Lamberti, J. DeBie, and M. Ferri, "Respiratory signal derived from eight-lead ECG," in *Computers in Cardiology*. Piscataway, NJ: IEEE Press, 1998, vol. 25, pp. 65–68.

[15] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley, 2001.

[16] W. A. H. Engelse and C. Zeelenberg, "A single scan algorithm for QRS-detection and feature extraction," in *Computers in Cardiology*. Los Alamitos, CA: IEEE Computer Society Press, 1979, vol. 6, pp. 37–42.

[17] A. Ambardar, *Analog and Digital Signal Processing*. Boston, MA: PWS, 1995.

[18] M. C. Teich, S. B. Lowen, B. M. Jost, K. Vibe-Rheymer, and C. Heneghan, "Heart rate variability: Measures and models," in *Nonlinear Biomedical Signal Processing*, M. Akay, Ed. Piscataway, NJ: IEEE Press, 2000, vol. II.

[19] R. W. DeBoer, J. M. Karemaker, and J. Strackee, "Comparing spectra of a series of point events particularly for heart rate variability data," *IEEE Trans. Biomed. Eng.*, vol. BME-31, pp. 384–387, 1984.

[20] Task Force Eur. Soc. Cardiol. and North Amer. Soc. Pacing and Electrophysiology, "Heart rate variability—Standards of measurement, physiological interpretation and clinical use," *Eur. Heart J.*, vol. 17, pp. 354–382, 1996.

[21] B. D. Ripley, *Pattern Recognition and Neural Networks*. Cambridge, U.K.: Cambridge Univ. Press, 1996.

[22] R. Kohavi, "A study of cross validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artificial Intelligence*, 1995, pp. 1137–1143.

[23] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York: Oxford Univ. Press, 1995.

[24] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. San Diego, CA: Academic, 1999.

[25] V. X. Afonso, W. J. Tompkins, T. Q. Nguyen, and S. Luo, "ECG beat detection using filter banks," *IEEE Trans. Biomed. Eng.*, vol. 46, pp. 556–564, Feb. 1999.

[26] S. Kadambe, R. Murray, and G. F. Boudreaux-Bartels, "Wavelet transformed-based QRS complex detector," *IEEE Trans. Biomed. Eng.*, vol. 46, pp. 838–848, July 1999.

[27] C. Li, C. Zheung, and C. Tai, "Detection of ECG characteristic points using wavelet points," *IEEE Trans. Biomed. Eng.*, vol. 42, pp. 21–28, Jan. 1995.

Elaine Sheridan received the B.E. and M.Eng.Sc. degrees in electronic engineering from the National University of Ireland, Dublin, in 1999 and 2001, respectively.

She is currently a Design Engineer with Motorola SPS, Cork, Ireland. Her research interests include biomedical engineering, signal processing, and communication systems.

Richard Reilly (M'92) received the B.E., M.Eng.Sc., and Ph.D. degrees from the National University of Ireland in 1987, 1989, and 1992, all in electronic engineering.

In 1988, he joined Space Technology Ireland and the Dept. de Recherche Spatiale (CNRS group) in Paris, France, developing DSP-based on-board experimentation for the NASA satellites. In 1990, he joined the National Rehabilitation Hospital, Dublin, Ireland, and in 1992 became a Postdoctoral Research Fellow at University College Dublin, focusing on signal processing for speech and gesture recognition. Since 1996, Dr. Reilly has been on the academic staff in the Department of Electronic and Electrical Engineering at University College, Dublin. He is currently Senior Lecturer and researches into neurological signal processing and multimodal signal processing.

Philip de Chazal (M'94) received the B.E. degree in electronic engineering, and the M.Biomed.E and Ph.D. degrees in biomedical engineering from the University of New South Wales, Sydney, Australia, in 1989, 1995, and 1999, respectively.

He was a Research Scientist with the Ultrasonics Laboratory, CSIRO, Sydney, from 1990 to 1992, and a Biomedical Engineer with MedCare Systems, Sydney, from 1993 to 1999. Since June 1999, he has been with the University College Dublin. He spent one year as a Postdoctoral Scholar with the Conway Institute of Biomolecular and Biomedical Research and now works as a Research Fellow for the Department of Electronic and Electrical Engineering. His research interests include signal processing and pattern recognition for biomedical applications and image processing for multimodal applications.

Dr. de Chazal received an award at the Computers in Cardiology 2000 conference for the work described in this paper. He is a member of the IEEE Engineering in Medicine and Biology Society. He is a reviewer for IEEE TRANSACTIONS ON MULTIMEDIA.

Philip Nolan received the B.Sc. Degree (with honors) in physiology and the primary medical degree (M.B., B.Ch., B.A.O.), from University College, Dublin, Ireland, in 1998 and 1991, respectively.

He held the posts of Health Research Board Fellow (1992–1994) and Newman Scholar (1994–1996) at University College Dublin before being appointed Lecturer in Physiology in 1996. He is currently Senior Lecturer in Physiology at University College Dublin. His major research interest is in the physiological control of breathing and the circulatory system, with a particular interest in the neural control and physiological properties of upper airway muscles, their role in preventing upper airway obstruction during sleep, and the pathophysiology of the obstructive sleep apnoea syndrome.


Conor Heneghan (M'90) was born in Dublin, Ireland, in 1968 and received the B.E. degree in electronic engineering from University College Dublin, in 1990 and the Ph.D. degree in electrical engineering from Columbia University, New York, NY, in 1995.

He is currently a Senior Lecturer in the Department of Electronic and Electrical Engineering at University College Dublin. He was previously Director of Tele-Informatics at the New York Eye and Ear Infirmary and a Postdoctoral Research Scientist at Boston University, Boston, MA. His research interests include signal and image processing for biomedical applications and signal processing for communications.

Dr. Heneghan is also a member of the Institution of Electrical Engineers (IEE) and the Institution of Engineers of Ireland (IEI). He is a member of the IEEE Engineering in Medicine and Biology Society, the Signal Processing Society, and the Communications Society. He is a reviewer for IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING and IEEE TRANSACTIONS ON SIGNAL PROCESSING.
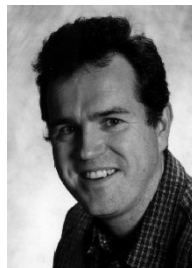
Mark O'Malley (M'86–SM'96) was born in Dublin, Ireland, in 1962, he received the B.E. and Ph.D. degrees from the National University of Ireland, Dublin, in 1983 and 1987, respectively.

He is currently a Professor in the National University of Ireland, Dublin with research interests in power systems, control theory, and biomedical engineering.