

# Detection of Obstructive Sleep Apnea in Pediatric Subjects using Surface Lead Electrocardiogram Features

Redmond B Shouldice, BE<sup>1</sup>; Louise M O'Brien, PhD<sup>2</sup>; Ciara O'Brien, BE<sup>1</sup>; Philip de Chazal, PhD<sup>1</sup>; David Gozal, MD<sup>2</sup>; Conor Heneghan, PhD<sup>1</sup>

<sup>1</sup>Digital Signal Processing Research Group, Department of Electronic and Electrical Engineering, University College Dublin, Dublin, Ireland;

<sup>2</sup>Department of Pediatrics, Kosair Children's Hospital Research Institute, University of Louisville School of Medicine, Louisville, KY, USA

**Study Objectives:** To investigate the feasibility of detecting obstructive sleep apnea (OSA) in children using an automated classification system based on analysis of overnight electrocardiogram (ECG) recordings.

**Design:** Retrospective observational study.

**Setting:** A pediatric sleep clinic.

**Participants:** Fifty children underwent full overnight polysomnography.

**Intervention:** N/A.

**Measurements and Results:** Expert polysomnography scoring was performed. The datasets were divided into a training set of 25 subjects (11 normal, 14 with OSA) and a withheld test set of 25 subjects (11 normal, 14 with OSA). Features, calculated from the ECG of the 25 training datasets, were empirically chosen to train a modified quadratic discriminant analysis classification system. The selected configuration used a segment length of 60 seconds and processed mean, SD, power spectral density, and serial correlation measures to classify segments as apneic or normal. By combining per-segment classifications and using receiver-

operator characteristic analysis, a per-subject classifier was obtained that had a sensitivity of 85.7%, specificity of 90.9%, and accuracy of 88% on the training datasets. The same decision threshold was applied to the withheld datasets and yielded a sensitivity of 85.7%, specificity of 81.8%, and accuracy of 84%. The positive and negative predictive values were 85.7% and 81.8%, respectively, on the test dataset.

**Conclusions:** The ability to correctly identify 12 out of 14 cases of OSA (with the 2 false negatives arising from subjects with an apnea-hypopnea index less than 10) indicates that the automated apnea classification system outlined may have clinical utility in pediatric patients.

**Key Words:** Obstructive sleep apnea, automated detection, pediatric, ECG, RR intervals

**Citation:** Shouldice RB; O'Brien LM; O'Brien C et al. Detection of obstructive sleep apnea in pediatric subjects using surface lead electrocardiogram features. *SLEEP* 2004;27(4):784-92.

## INTRODUCTION

OBSTRUCTIVE SLEEP APNEA (OSA) IS A COMMONLY ENCOUNTERED CONDITION CHARACTERIZED BY REPEATED EVENTS OF PARTIAL OR COMPLETE UPPER-AIRWAY OBSTRUCTION DURING SLEEP, RESULTING IN DISRUPTION OF NORMAL VENTILATION, HYPOXEMIA, AND SLEEP FRAGMENTATION. A transient return to the waking state (arousal) restores airway patency. OSA is estimated to affect between 1% and 3% of young children,<sup>1-3</sup> and its potential consequences include behavioral disturbances and learning deficits,<sup>2,4,5</sup> pulmonary and systemic hypertension,<sup>6,7</sup> and growth impairment.<sup>8</sup> Sleep is of particular importance to the pediatric age group, as evidenced by their longer sleep-duration requirements.<sup>9</sup> Furthermore, OSA in children is distinctly different from the disorder seen in adults, in particular with respect to sex distribution, clinical manifestations, and polysomnography (PSG) findings and treatment.<sup>10,11</sup> In adults, OSA is suggested by heavy

snoring, obesity, and excessive daytime sleepiness—especially in adult men. In contrast, OSA in children is equally common in both sexes and not necessarily linked to obesity.<sup>9</sup> OSA is most often seen in the 2- to 8-year-old age range when the tonsils and adenoids are largest relative to the underlying airway size.

The only currently accepted method for diagnosis of OSA is overnight PSG; unfortunately, the scarcity of pediatric sleep laboratories and the relatively high medical costs associated with such tests make it likely that pediatric OSA is widely underdiagnosed.<sup>12</sup> As a consequence, there is a need for simpler measurements to aid in the diagnosis of OSA. Such methods have been explored; these include subject history, otolaryngologic examination, and audio recordings of snoring, and all have been shown to have low diagnostic sensitivity and specificity.<sup>13-15</sup> Methods such as nighttime videotaping,<sup>16</sup> pulse oximetry,<sup>17</sup> and nap PSG<sup>18</sup> have a good true positive detection rate, albeit at the expense of many false negatives, ie, a negative result might still require full PSG to rule out a suspected case.<sup>9</sup> Based on previous preliminary observations in children<sup>19</sup> and our extensive experience in electrocardiogram (ECG)-signal analysis in adults with OSA,<sup>20-22</sup> we explore here the feasibility of ECG time- and frequency-based interval analysis for diagnostic purposes in snoring children suspected of having OSA.

It is useful to briefly review the known clinical observations relating to sleep-disordered breathing and heart rate. In 1984, Guilleminault et al<sup>23</sup> published the first paper in this area, noting that obstructive apneas were often associated with a bradycardia during the apneic period, followed by a tachycardia as breathing is resumed. They termed these patterns cyclical variations in heart rate. Since apneas typically occur on a timescale of 10- to 20-seconds duration, the net effect (on average) is to introduce a frequency component in the RR interval tachogram correspond-

## Disclosure Statement

This work was supported by the Enterprise Ireland Research Innovation Fund under Grant Number IF/2001/006 and by NIH grant HL-65270. Data acquisition (multichannel polysomnographic overnight evaluation) was performed at the Sleep Medicine Center of Kosair Children's Hospital, Louisville, Kentucky. Data analysis and automated classification system development was carried out at the Department of Electronic and Electrical Engineering, University College Dublin, Ireland.

Submitted for publication October 2003

Accepted for publication January 2004

Address correspondence to: Conor Heneghan, PhD, Digital Signal Processing Research Group, Department of Electronic and Electrical Engineering, University College Dublin, Dublin, Ireland. Tel: 353 1 7161925. Fax: 353 1 2830921; E-mail: conor.heneghan@ucd.ie

ing to this time scale, ie, at frequencies in the range 0.05 to 0.1 Hz. These variations can usually be more easily seen in the corresponding RR spectrum. In a recent paper, Stein et al<sup>24</sup> gave a useful graphical interpretation of this observation. In adult patients, they were able to identify episodes of obstructive apnea solely through visual inspection of the RR-interval tachogram by recognizing the characteristic cyclical variations in heart-rate patterns. Other researchers have also noted the low-frequency fluctuations introduced by apneas and have developed a range of possible systems for using heart rate to recognize apneas.<sup>25-27</sup>

An additional common clinical observation that we have investigated in the development of our automated system is the fact that respiration modulates the amplitude of the recorded surface ECG, due to mechanical motion of the electrodes relative to the heart and impedance changes in the thorax as we breathe. Again, several researchers have conducted a more careful study of this phenomenon and have proposed various algorithms for the calculation of an ECG-derived respiratory (EDR) signal.<sup>28,29</sup> Such a signal has obvious potential utility in apnea monitoring because absence, or diminution, of respiratory effort is a canonical feature of sleep-disordered breathing.

In our previous research on adult subjects, we have shown that combining the information provided by the cyclical variations in heart rate with that from EDR signals can provide useful diagnostic insight. Indeed, we have previously shown that by applying an automated system for detection of OSA in adults, based solely on ECG measurements, it is possible to detect more than 90% of obstructive events.<sup>20</sup> Such a system is capable of providing a sequence of minute-by-minute classifications of either normal or obstructive respiration and also indicates the presence or absence of clinically significant apnea. However, this methodology has yet to be tested in the pediatric arena. Thus, we hypothesized that the ECG would be a useful tool in identifying children with OSA.

## METHODS

### PSG Study

Children referred for evaluation for OSA underwent a standard multichannel PSG study at the Sleep Medicine Center of Kosair Children's Hospital, Louisville. Control children without a history of snoring also underwent PSG assessment as part of an ongoing research project. The study was approved by the University of Louisville Institutional Review Board, and informed parental consent was obtained. For children aged 7 years and above, child assent, in the presence of a parent, was obtained. Children were studied for up to 12 hours in a quiet darkened room with an ambient temperature of 24°C in the company of 1 of their parents. All children were in bed with lights out between 9:00 and 9:30 PM and were awakened at 7:00 AM (unless already awake). No drugs were used to induce sleep. Measurements included chest and abdominal wall movement by respiratory impedance or inductance plethysmography; heart rate by ECG; and air flow with a sidestream end-tidal capnograph, which also provided breath-by-breath assessment of end-tidal carbon dioxide levels (PETCO<sub>2</sub>; BCI SC-300, Menomonee Falls, Wisc), a thermistor, or both. Arterial oxygen saturation (SpO<sub>2</sub>) was assessed by pulse oximetry (Nellcor N 100; Nellcor Inc., Hayward, Calif), with simultaneous recording of the pulse waveform. The bilateral electroocu-

lograms, 8 channels of electroencephalogram, chin and anterior tibial electromyograms, and analog output from a body-position sensor (Braebon Medical Corporation, NY) were also monitored. Digital signal acquisition was performed using a commercial PSG system (Medcare Diagnostics, Amsterdam, The Netherlands); lead I ECG data was sampled at 256 Hz. Tracheal sound was monitored with a microphone sensor (Sleepmate, Midlothian, Virginia), and a digital time-synchronized video recording was also performed.

### Apnea-Hypopnea Index Scoring

Sleep architecture was assessed using standard techniques by one of the authors (LMO).<sup>30</sup> The apnea index was defined as the number of apneas per hour of total sleep time. Central, obstructive, and mixed apneic events were counted. Obstructive apnea was defined as the absence of airflow with continued chest wall and abdominal movement for at least 2 breaths.<sup>31</sup> Hypopneas were defined as a 50% or more decrease in nasal flow with a corresponding decrease of 4% or more in SpO<sub>2</sub>, an arousal, or both<sup>31</sup> and were scored if the duration was at least 2 breaths. The apnea-hypopnea index (AHI) was defined as the number of apneas and hypopneas per hour of total sleep time. The mean oxygen saturation, as measured by pulse oximetry (SpO<sub>2</sub>), together with SpO<sub>2</sub> nadir, was determined. Subjects were classified as having OSA if their obstructive AHI was  $\geq 1$  and were further divided into mild (AHI < 10), moderate (AHI 10-15), or severe (AHI > 15). Subjects were classified as controls if their obstructive AHI was < 1. In our clinical population, similar to others,<sup>17</sup> we have an approximate probability of 60% to 65% for detecting OSA based on history and physical examination alone.

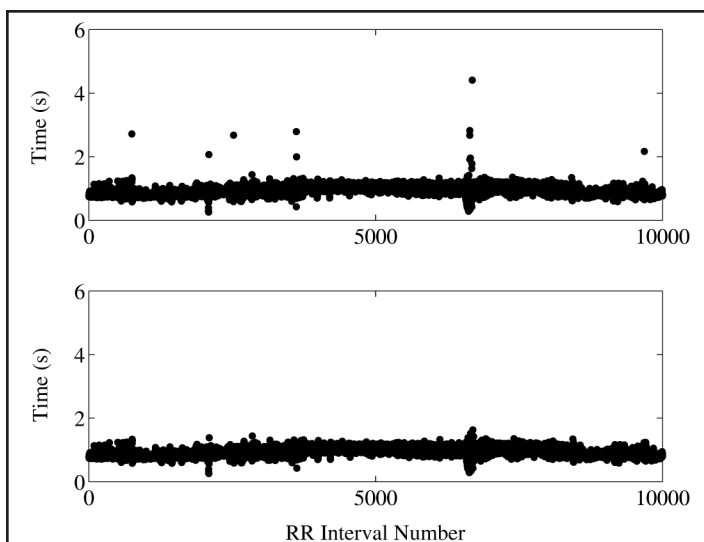
### Classification Methodology

Our aim was to design an automated pediatric apnea screening system that could correctly assign a subject to 1 of 2 classes: "normal" or "apneic," based solely on the given surface ECG recording. This system was designed using a supervised classifier methodology, ie, we had initially provided the classifier with both ECG recordings and correct annotations on a training dataset as judged by a human expert. Data analysis was carried out using MATLAB (The Mathworks, Inc., Natick, Mass). Our overall approach was to carry out classification of the recordings at relatively short timescales (eg, 30 seconds), and then combine the classifications at short timescales into a single subject classification. Hence, the automated pediatric apnea detection system implemented was composed of 3 sections. The first section was concerned with the accurate extraction of ECG fiducial points—in particular, successive QRS complexes. The second stage consisted of segmenting the ECG characteristic points into nonoverlapping fixed-length time segments and calculating features to best represent changes in the segments due to apnea. Thirdly, a pattern recognition approach was utilized to classify normal and apneic data segments, ie, a module capable of discriminating normal and apneic segments based upon a limited set of known normal apnea training data.

### ECG Characteristic Point Extraction

ECG processing was aimed at reliable extraction of the RR interval. This task can be considered as QRS complex detection,

followed by accurate measurement of the location and amplitude of the R peaks. A Hilbert transform-based detector was employed<sup>32</sup> in this work. This methodology has been applied in our research group to a variety of digital ECG signals sampled at anywhere between 100 Hz and 1 kHz and can cope with severe electromyogram noise and lead drop out. The detector is suitable for processing all common electrode lead positions that provide a distinct R peak and can also be modified to detect the S-wave minimum if required. The ECG signals were band-pass filtered between 8 Hz and 18 Hz; the Hilbert transform was then taken of the first derivative of the signal to emphasize the R peaks. A moving-window peak search was carried out with adaptive threshold. This extraction methodology was applied to the 50 complete datasets, which represented a total of 24,835 minutes. A number of R peaks were missed due to electromyographic artifact and electrode detachment; in addition, some spurious noise peaks were incorrectly detected as R peaks. Of the R peaks that were analyzed, 1.1% required data conditioning in the form of interpolation or deletion. No attempt was made to distinguish normal sinus beats from aberrant beats, on the basis that such beats were rare in the 50 subjects considered. The 2 data-conditioning steps that were incorporated into the RR interval extraction technique included (1) the removal of physiologically unreasonable intervals and (2) RR series interpolation to help correct for undetected QRS complexes. Given the set of detected R peaks from each subject's ECG signal, the corresponding RR intervals series were derived. To correct for spurious or missed detections, the following methodology was applied. For the case of an RR interval  $RR(n)$  being smaller than 150 milliseconds in length, the QRS onset  $R(n+1)$  was removed from the series. If a particular RR interval was greater than double the trimmed mean at 10% (ie, largest 5% and smallest 5% of intervals removed) of the entire RR series, an estimate of the missing RR interval based upon the mean of the preceding 5 intervals was inserted. In the case where the suspect interval length was more than 2.2 times the estimate, an integer number of replacement RR intervals were generated for insertion into the series using a sliding window based on the preceding 5 intervals. This postprocessing step ensured a set of physiologically reasonable RR intervals, as per the example shown in Figure 1.



**Figure 1**—A segment of an RR-interval series for a subject before (top) and after (bottom) deletion/interpolation postprocessing was applied.

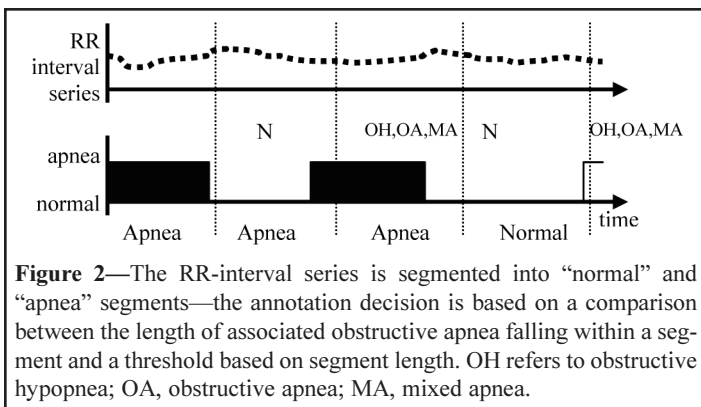
## Interval Segmenting and Feature Choice

The second stage of the detection process focused on the segmentation of the RR series and subsequent feature estimation. In this work, the RR intervals were split up into 15-second, 30-second, or 1-minute nonoverlapping segments. Each segment was labeled as either normal or 1 of the available preclassified and verified apnea events: OA (obstructive apnea), OH (obstructive hypopnea), CA (central apnea), or MA (mixed apnea). This labeling was carried out from the respiratory-event list generated during the scoring process, which gave the onset and offset times of all significant respiratory events. A segment label was determined by whether it contained greater than 10% of the segment length of 1 of the annotated respiratory events listed above. The segment-labeling method is shown in Figure 2. Note that this potentially led to a 5-class problem. This was avoided by collapsing OA, OH, and MA into a single “apnea” class and by not including CA events in the classification process (since these were rare in this data set—40 CA events occurred in total for the 50 subjects considered). It was then desirable to choose the temporal and/or frequency features of the ECG that most clearly represented differences between normal and apneic respiration. Many different features of the ECG characteristic point series can be considered. The main features considered in this study were mean, SD, serial correlation coefficients, and the interval-based power spectral density (PSD) of the RR intervals,<sup>20,33</sup> as well as features obtained from an EDR signal.<sup>28</sup>

For each segment, the temporal features investigated included the mean and square root of the SD of the RR intervals within that segment and the first 3 serial correlation coefficients. The serial correlation coefficients were derived from the cross-correlation between a subset of the segmented series and a shifted version of same; this provided an indication of the persistence in a time series. As an example, the first serial correlation coefficient of a series of RR intervals  $x$  of length  $n$  was estimated by:

$$r_1 = \frac{\sum_{k=1}^{n-1} (x_k - \bar{x})(x_{k+1} - \bar{x})}{\sqrt{\sum_{k=1}^{n-1} (x_k - \bar{x})^2 \sum_{k=1}^{n-1} (x_{k+1} - \bar{x})^2}}$$

The frequency-based measures considered included PSD estimates of the RR-interval series directly and of an EDR signal. The EDR signal was obtained by looking for the modulation of the R peak by respiration (due to mechanical motion of the ribcage during ventilation, which causes chest electrodes to move



**Figure 2**—The RR-interval series is segmented into “normal” and “apnea” segments—the annotation decision is based on a comparison between the length of associated obstructive apnea falling within a segment and a threshold based on segment length. OH refers to obstructive hypopnea; OA, obstructive apnea; MA, mixed apnea.

relative to the heart). To derive the EDR signal, it was necessary to first filter the ECG signal twice with median filters of length 172 milliseconds and 586 milliseconds, respectively, and subtract the resulting baseline wander signal from the ECG. By integrating over the range 10 milliseconds before and 10 milliseconds after the value of the R peak, an estimate of respiration at a time corresponding to the R peak was generated.

The spectral features were obtained from the RR and EDR intervals using Welch's averaged, modified periodogram method. The resulting interval-based PSD was scaled and made to approximate a count-based PSD by scaling the interval axis by the mean RR value.<sup>33</sup> For the RR-interval data, the spectrum was obtained by zero padding to length 1024; the features of interest were derived from the log values of eighteen 0.03-Hz-wide frequency bins in the range 0 to 0.54 Hz. The spectrum of the EDR was obtained by zero padding to 512; 11 features were obtained from the log values of 0.05-Hz-wide bins in the range 0 to 0.55 Hz. 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, a log transformation was applied to the PSD features in order that the histogram of the transformed feature might more closely approximate a Gaussian distribution. Features were chosen by systematic search. To summarize, each segment produced 34 features. For convenience, features were divided into 2 groups: RR features and EDR features.

### Classification Stage

A supervised training technique was used to derive the classifier used in our study.<sup>34</sup> In this study, a quadratic discriminant (QD) model was used with Mahalanobis distance.<sup>35</sup> (Details relating to the derivation and use of this quadratic discriminant classifier are described in the Appendix.) The use of an ad hoc postprocessing step was also investigated; if a segment was scored as "apnea" but had neighboring segments that were both scored "normal," it was reclassified as "normal." This had the effect of removing apneic events isolated to a particular segment—ie, some false positives would be filtered but at the expense of reduced sensitivity.

The performance of the classifier scheme on a particular dataset was initially calculated on a per-segment basis. A true-positive segment was one that both the human expert and automated classifier labeled as "apnea," a true-negative was one labeled as "normal" by both, a false-positive was one labeled as "apnea" but that was actually "normal," and so forth. Specificity, sensitivity, positive predictive value, and negative predictive value were all defined in a standard fashion. Accuracy was defined as the total number of correctly classified segments divided by the total number of segments observed. After training a classifier using training data, the classification performance can be estimated by processing the training data with the classifier. However, such classification performance estimates tend to be optimistic, as the classifier would have already been adapted to that particular dataset. In this study, more realistic estimates of classifier performance were obtained using a leave-1-out cross-validation scheme.<sup>36</sup> In this scheme, 1 subject's data were withheld, and classifier parameters were obtained by using all the other available training data. The classification performance of

the resulting classifier can then be estimated on the 1-withheld set. Overall estimated classification performance can then be obtained by averaging the results over all possible withheld sets. In this way, test specificity, sensitivity, and accuracy figures may be produced on a per-segment basis. From the per-segment classification of the training set, a simple per-subject decision was possible by counting the average number of detected apnea segments per hour of sleep (as defined by sleep staging) and choosing an appropriate threshold. An optimum choice of threshold was inferred from the receiver operating characteristic (ROC) curve<sup>37</sup>—ie, the threshold that gives the point nearest the (0,1) coordinate. By selection of this threshold, the sensitivity and specificity coordinates of the overall clinical classifier were then generated. Segment lengths of 15, 30, and 60 seconds were considered. Previous work on adult ECG classification focused on a database of containing 60-second segments.<sup>22</sup> The shorter segment lengths were also investigated in case the pediatric obstructive events might be captured more easily, although risking that the low frequency fluctuations of interest might be incompletely described.

In order to determine the clinical value of using the ECG, we also calculated likelihood ratios. These ratios indicate how much a diagnostic test result raises or lowers the pretest probability of a subject being classified as OSA or normal, given a positive or negative outcome to the test. The segments-per-hour, based on the best performing classifier model, are cross-correlated with observed AHIs to produce a parametric *r* value and associated *P* value.

### RESULTS

In this work, PSG data from 50 subjects were considered. Expert scoring of the data yielded 22 control subjects and 28 apneic OSA subjects. Subject characteristics are given in Table 1. Of the apneic subjects, 7 were classified as having mild OSA, 10 as moderate OSA, and 11 as severe OSA.

The 50 subjects were divided into 2 sets: a training set of 25 subjects (11 normal, 14 OSA), and a withheld test set of 25 subjects (11 normal, 14 OSA). The withheld data were never used for training of classifier models and, thus, represented independent test data. Table 2 gives the per-segment and per-subject classification performances that were achieved using the classifier model described above and with various feature sets at different timescales. Results are presented both with and without postprocessing of segment annotations. The 60-second segment with RR60 features performed best on a per-subject basis across the

**Table 1**—Demographic Information of 50 Children

Parameter	Normal (n = 22)	Apneics (n = 28)
Age, y	9.1 ± 3.3	8.5 ± 4.8, (1.2–16)
Sex, % male	59	50
AHI	0.06 ± 0.13*	20.6 ± 21.2*
BMI	20 ± 9 (13.7–42.4)	25.6 ± 10.4 (13.8–51.8)

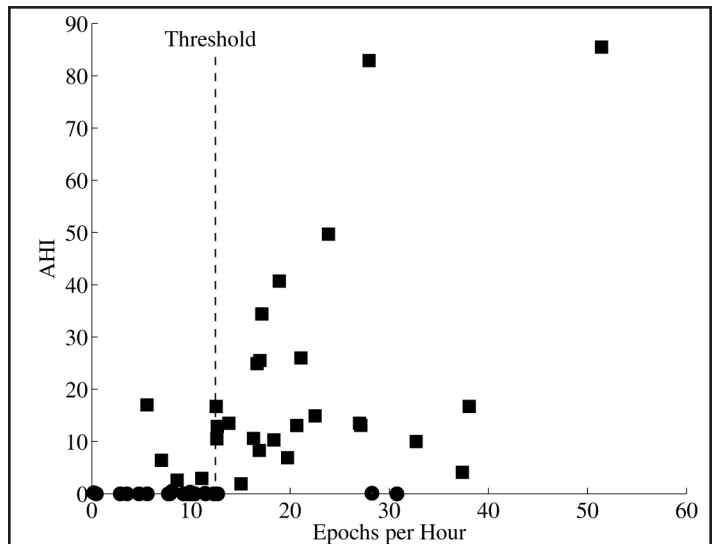
Data are presented as mean ± SD (range), unless otherwise noted. AHI refers to apnea-hypopnea index; BMI, body mass index.

\*Statistically significant differences between control and apneic datasets at *P* < .01

training data. While the results on a per-segment basis were relatively modest (features had been empirically tuned to maximize per-subject results), the real potential clinical utility lies in whether combinations of the per-segment classifications could yield an overall reliable clinical classification for individual subjects.

A simple per-subject decision was made by counting the average number of detected apnea segments per hour of sleep from the output of the automated classification system. Figure 3 shows that the segments-per-hour (using the RR60 classifier model) are reasonably correlated with observed AHIs (parametric  $r$  value was 0.5,  $P < .0001$ ). It was apparent that a threshold of 12.5 segments per hour could be used to separate the “normals” from the “apneics,” though not with 100% accuracy. By selection of this threshold, the sensitivity and specificity coordinates of the overall clinical classifier were generated, as per the ROC plot of Figure 4 (top) using a range of hard decision thresholds. An optimum choice of threshold from this ROC plot yields a sensitivity of 85.7%, a specificity of 90.9%, and accuracy of 88%. The area under the curve (AUC) using trapezoidal numerical integration was 0.92. When the same thresholds were applied to the test data, 12 out of 14 apnea cases and 9 out of 11 normals were correctly identified (sensitivity of 85.7%, specificity of 81.8%, and accu-

racy of 84%). The AUC was 0.83. The positive and negative predictive values were 85.7% and 81.8% respectively. The positive likelihood ratio ( $LR^+$ , defined as  $Se/[1-Sp]$ ) was 4.71, increasing



**Figure 3**—Agreement between the detected segments per hour and the apnea-hypopnea index (AHI) for the best-performing classifier in Table 2. The ● and ■ symbols denote control and apneic cases, respectively.

**Table 2**—Classification Performance on the Training Set (n = 25) and Test Dataset (n = 25)

Feature set	No. feat	Average training set results		Training set results per-subject decision		Average of test set results		Test set results per-subject decision	
		Acc.	Sens, Spec	Sens	Spec	Acc	Sens, Spec	Sens	Spec
RR15	23	89.1%	33.7%, 92.7%	78.6%	72.7%	83.3%	34.9%, 86.7%	92.9%	81.8%
		92.3%	22.6%, 96.8%	78.6%	72.7%	88.5%	24.7%, 92.9%	85.7%	81.8%
EDR15	11	44.2%	68.6%, 42.6%	78.6%	72.7%	43.5%	65.7%, 42.0%	71.4%	72.7%
		46.2%	64.2%, 45.0%	78.6%	72.7%	45.6%	60.8%, 44.6%	71.4%	72.7%
RR15+ EDR15	23+11	85.2%	38.1%, 88.2%	92.9%	72.7%	74.5%	37.2%, 77.2%	71.4%	72.7%
		88.1%	27.3%, 92.0%	78.6%	72.7%	80.3%	27.9%, 83.9%	71.4%	72.7%
RR30	23	81.1%	61.0%, 82.6%	85.7%	81.8%	62.3%	78.2%, 60.9%	100%	72.7%
		86.5%	50.5%, 89.3%	78.6%	81.8%	67.7%	73.3%, 67.2%	85.7%	81.8%
EDR30	11	42.9%	56.6%, 41.8%	57.7%	81.8%	50.8%	57.3%, 50.3%	71.4%	72.7%
		45.7%	49.8%, 45.4%	57.7%	81.8%	54.4%	51.1%, 54.6%	64.3%	81.8%
RR30+ EDR30	23+11	79.7%	58.1%, 81.3%	92.8%	72.7%	60.0%	75.3%, 58.7%	92.9%	63.7%
		84.3%	48.7%, 87.1%	92.8%	72.7%	65.3%	70.0%, 64.9%	71.4%	72.7%
RR60	23	82.5%	62.5%, 84.5%	92.9%	81.8%	66.9%	72.4%, 65.5%	78.6%	81.8%
		87.5%	52.7%, 90.9%	<b>85.7%</b>	<b>90.9%</b>	72.1%	66.2%, 72.9%	<b>85.7%</b>	<b>81.8%</b>
EDR60	11	55.7%	48.5%, 56.4%	64.3%	72.7%	56.0%	50.3%, 56.7%	64.3%	81.8%
		59.7%	41.7%, 61.5%	64.3%	72.7%	60.1%	41.6%, 62.3%	64.3%	81.8%
RR60+ EDR60	23+11	82.9%	56.9%, 85.5%	100%	81.8%	64.3%	61.9%, 64.6%	64.3%	72.7%
		87.1%	47.6%, 91.0%	92.9%	81.8%	69.7%	54.6%, 71.4%	78.5%	72.7%

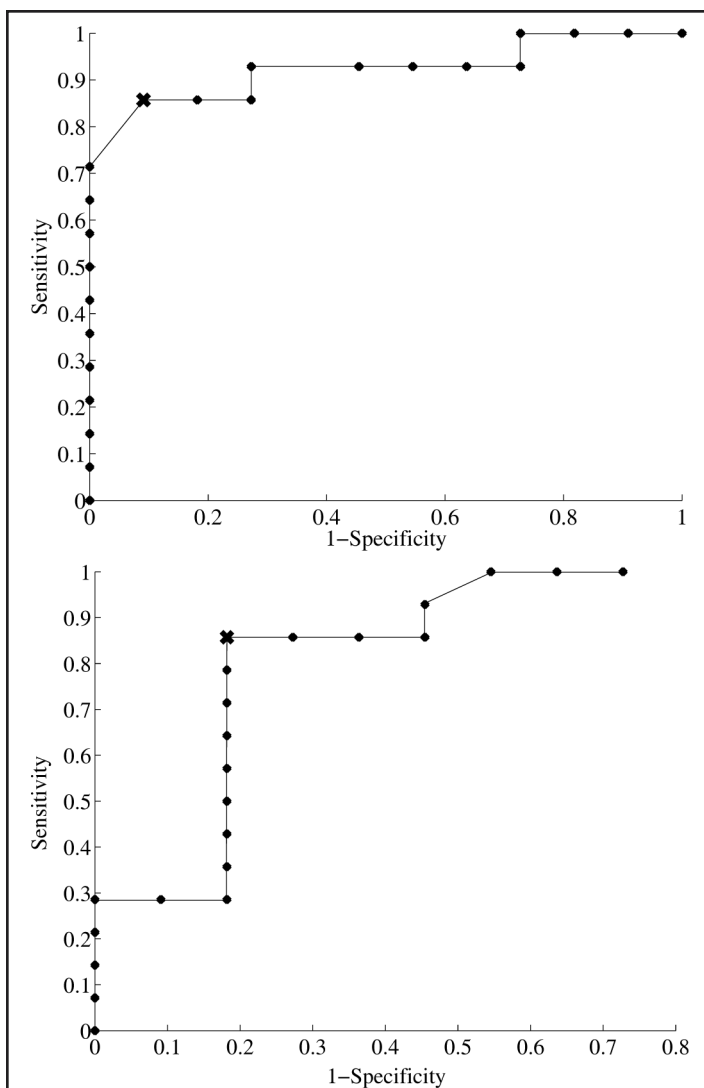
Feat refers to features; Acc, accuracy; Sens, sensitivity; spec, specificity; RR, RR interval; EDR, electrocardiogram-derived respiratory signal. The quadratic discriminant classifier model was employed to generate this table, as discussed in the text. RR15 is the set of 23 RR-based features calculated using a 15-second segment, EDR15 is the set of 11 EDR-based features calculated using 15-second segments, and so forth. In each cell of this table, the upper figure gives a performance without segment postprocessing, the lower is when the postprocessing scheme is used. The best-performing system on the training set is highlighted in bold.

the probability of having OSA from 60% to 83%, and the negative likelihood ratio (LR-, defined as  $[1-Se/Sp]$ ) was 0.18, decreasing the probability of having OSA from 60% to 15%. Table 3 shows the optimized classification performance for both the training and test datasets.

The mean heart rates for the 60-second segment lengths (from all 50 subjects) classed as either “normal” or “apneic” were found to be 93.5 beats per minute (bpm) and 103.4 bpm, respectively, for subjects aged less than 4 years, 77.8 bpm and 99.8 bpm for subjects aged 4 to 8 years, 79.6 bpm and 82.6 bpm for subjects aged 8 to 12 years, and 77.1 bpm and 78.7 bpm for subjects over 12 years of age.

## DISCUSSION

The per-subject scores obtained from both the training and withheld test sets indicate that ECG-based analysis may have some screening utility for pediatric OSA. Furthermore, a positive ECG result greatly improved the probability of detecting OSA from 60% to 83% over and above clinical impression alone, while a negative result reduced the probability to 15%. Of the 4



**Figure 4**—The receiver operating curve points, based on a range of hard decision thresholds applied to the cross-fold validation results from the 25 training sets (top) provided the decision threshold that, when applied to the 25 withheld datasets (bottom), yielded a specificity of 81.8% and sensitivity of 85.7%.

false negatives exhibited by the system, 3 were for cases that would be considered as “mild” (AHIs of 2.6, 3.0, and 6.4, respectively). The remaining false negative was for a subject with severe OSA, as judged by AHI (the recorded AHI was 17). In order to determine if there is an underlying systematic reason for such an anomalous result, it will be necessary to consider a much larger database of results. A possible cause is that this subject had the largest recorded body mass index BMI value (51.8) in the database. In recent work, it has been shown that obesity in adolescents is associated with changes in heart-rate variability, including an overall reduction in low-frequency heart-rate variability measures as compared to nonobese control subjects, and it is possible that, in some cases, these systemic changes are masking changes in the autonomic response to sleep-disordered breathing.<sup>38</sup> The system also displayed 2 false positives on the test data. In fact, our system characterized these 2 subjects with AHIs of 0.1 and 0 as having segment-per-hour measures of 28.3 and 30.8, respectively, which are well in excess of our threshold value of 12.5. We note in passing that these 2 subjects had some of the lowest BMI values in the database (16.9 and 13.8, as against a lowest recorded BMI of 13.7). At present, we have no plausible mechanism to explain why subjects with low BMI values might be misclassified, and it is perfectly possible that the correspondence between low BMI and false-positive misclassification is by chance alone.

A potential limitation of our method is that we have no direct access to sleep-staging information from an ECG alone and thus cannot, for instance, provide comparative mean heart-rate values for wakefulness during the sleep period. In addition, our technique does not provide any record of SaO<sub>2</sub>, nor does it provide information about body position, which can also be of clinical significance. Finally, the current system makes no attempt to distinguish CA, or to distinguish obstructive and mixed hypopneas from obstructive and mixed apneas. Despite these limitations, however, we believe that this initial study, which correctly classified 21 out of 25 subjects using ECG alone, indicates that our approach may have clinical utility. Our results with automated ECG analysis fall within the reported interrater and intrarater reliabilities for sleep scoring.<sup>39</sup>

To place our work in context, it is pertinent to compare the efficacy of the system outlined in this work with other OSA-screening systems for pediatric subjects. Most of these trials, however, do not utilize independent training and test sets as we have done, making direct comparisons somewhat difficult.

Silvestri et al<sup>40</sup> used a measure of percentage of ideal body weight, presence of adenotonsillar tissue, and the presence of 5 or more sleep-associated breathing-disorder features to predict PSG

**Table 3**—Optimized Classification Performance for 25 Training and 25 Test Datasets.

	Classification Output				
	25 training datasets		25 withheld datasets		
	Normal	Apneic	Normal	Apneic	
<b>Expert</b>	Normal	10 (TN)	1 (FP)	9 (TN)	2 (FP)
<b>Annotation</b>	Apneic	2 (FN)	12 (TP)	2 (FN)	12 (TP)

TN refers to true negative; FP, false positive; FN, false negative; TP, true positive.

disorders in obese children. The reported overall accuracy on a per-subject basis was 81% from a database of 32 subjects (as compared to 84% over 25 test subjects for our method). It should be noted, however, that the probability of OSA is markedly increased by the concomitant presence of obesity and that, therefore, extrapolation from this study to ours is not possible.

Oximetry-based screening has also been widely suggested for both adult and pediatric populations. However, a confounding factor in children is that obstructive events frequently do not lead to significant oxyhemoglobin desaturation<sup>41</sup>; in addition, different averaging times and movement artifact may lead to false desaturation detection.<sup>42</sup> Brouillette et al<sup>17</sup> obtained a positive predictive value of greater than 97% (over 349 subjects) in the case of children suspected of having OSA, using overnight pulse oximetry alone (as compared to a positive predictive value of 86% in our study). However, in order to achieve such a high positive predictive value, their test produced significant numbers of inconclusive results, ie, lowered sensitivity, and required some manual scoring.

Nap-based PSG OSA screening is possible, although studies rarely contain rapid eye movement sleep. In addition, many children are unlikely to display spontaneous daytime sleep in an unfamiliar environment.<sup>43</sup> Therefore, such studies require sedation and tend to have good specificity but poor sensitivity. Marcus et al<sup>18</sup> compared 1-hour daytime nap to overnight PSG in 40 children (mean age 5.4 years) and achieved detection sensitivity of 74%, specificity of 100%, positive predictive value of 100%, and negative predictive value of 17% (as compared to 86%, 82%, 86%, and 82% in our study). Again, there is a need for some manual scoring.

Sonography and videography represent further alternative screening techniques. Lamm et al<sup>15</sup> evaluated home audiotapes as an abbreviated test for OSA syndrome in 36 (29 completed study) children yielding a median sensitivity of 71% and specificity of 80%. Sivan et al<sup>16</sup> used home videotape recording in 58 children for OSA screening, with a sensitivity of 94% and specificity of 68%; however, the high sensitivity and low specificity suggest that the parents probably selected the worst 30-minute recording clips for review by physician.<sup>44</sup> Reliable video recording with suitable lighting and recording conditions, and subsequent analysis, requires scoring by trained personnel. Finally, minimal sets of cardiorespiratory signals (cardiorespiratory sleep studies) that typically include 2 or more signals such as chest wall movement, pulse oximetry, heart rate, body movement, or airflow have also been considered.<sup>44</sup> These studies have been shown to be sensitive to OSA, but mostly in adults and frequently in a sleep lab during simultaneous PSG.<sup>45,46</sup>

In summary, existing techniques for screening are expectedly afflicted by several limitations such as low sensitivity or specificity (clinical history, oximetry, nap studies, sonography), or complexity (videography), or insufficient validation on pediatric populations (cardiorespiratory screening), or the need for some manual scoring (all techniques). The sensitivity and specificity of our proposed method is equal to or better than these alternative methods, as judged from this initial database of 50 subjects. Our method, as presented, is a fully automated screening system. A second significant advantage is the ease of recording an overnight ECG signal both in terms of comfort to the subject and reliable signal acquisition. Ambulatory (Holter) ECG monitoring is one of the most frequently carried out diagnostic tests and is

well tolerated in the general population. The ECG detector considered here is capable of handling a wide range of input sampling frequencies and a variety of lead positions containing an R peak acquired via Holter monitors. In general, exact electrode-placement position is not critical and good signal quality can be maintained over long periods of recording. Finally, since the proposed method is in theory capable of piggybacking onto existing Holter tests, it could be combined with a screening test for cardiac problems. We conclude that overnight ECG-based screening for OSA in pediatric patients, using the outlined automated apnea classification system, is technically feasible and appears to have potential clinical utility. However, significantly more clinical experience with the technique, and a detailed cost-benefit analysis, would be required to evaluate its true clinical utility as a screening tool.

## REFERENCES

1. Brouillette R, Hanson D, David R, Klemka L, Szatowski A, Fernbach S, Hunt C. A diagnostic approach to suspected obstructive sleep apnea in children. *J Pediatr* 1984;105:10-4.
2. Ali N J, Pitson D, Stradling J R. Snoring, sleep disturbance and behavior in 4-5 year olds. *Arch Dis Child* 1993;68:360-6.
3. Gislason, T, and B. Benediktsdottir. Snoring, apneic episodes, and nocturnal hypoxemia among children 6 months to 6-years-old. *Chest* 1995;107:963-6.
4. Chervin R, Dillon J, Bassetti C, Ganoczy D, Pituch K. Symptoms of sleep disorders, inattention, and hyperactivity in children. *Sleep* 1997;20:1185-92.
5. Gozal D. Sleep-disordered breathing and school performance in children. *Pediatrics* 1998;102:616-20.
6. Shiomi T, Guilleminault C, Stoohs R, Schnittger I. Obstructed breathing in children during sleep monitored by echocardiography. *Acta Paediatr* 1993;82:863-71.
7. Amin RS, Kimball TR, Bean JA, et al. Left ventricular hypertrophy and abnormal ventricular geometry in children and adolescents with obstructive sleep apnea. *Am J Respir Crit Care Med* 2002;165:1395-9.
8. Everett AD, Koch WC, Saulsbury FT. Failure to thrive due to obstructive sleep apnea. *Clin Pediatr* 1987;26:90-2.
9. Marcus CL. Sleep-disordered breathing in children. *Am J Respir Crit Care Med* 2001;164:16-30.
10. Rosen CL, D'Andrea L, Haddad GG. Adult criteria for obstructive sleep apnea do not identify children with serious obstruction. *Am Rev Respir Dis* 1992;146:1231-4.
11. Carroll JL, McLoughlin GM. Diagnostic criteria for obstructive sleep apnea in children. *Pediatr Pulmonol* 1992;14:71-4.
12. Young T, Evans L, Finn L, Palta M. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep* 1997;20:705-6.
13. Carroll JL, McColley SA, Marcus CL, Curtis S, Loughlin GM. Inability of clinical history to distinguish primary snoring from obstructive sleep apnea syndrome in children. *Chest* 1995;108:610-8.
14. Wang RC, Elkins TP, Keech D, Wauquier A, Hubbard D. Accuracy of clinical evaluation in pediatric obstructive sleep apnea. *Otolaryngol Head Neck Surg* 1998;118:69-73.
15. Lamm C, Mandeli J, Kattan M. Evaluation of home audiotapes as an abbreviated test for obstructive sleep apnea syndrome (OSAS) in children. *Pediatr Pulmonol* 1999;27:267-72.
16. Sivan Y, Kornecki A, Schonfeld T. Screening obstructive sleep apnoea syndrome by home videotape recording in children. *Eur Respir J* 1996;9:2127-31.
17. Brouillette RT, Morielli A, Leimanis A, Waters KA, Luciano R, Ducharme FM. Nocturnal pulse oximetry as an abbreviated testing

- modality for pediatric obstructive sleep apnea. *Pediatrics* 2000;105:405-12.
18. Marcus CL, Keens TG, Ward SLD. Comparison of NAP and overnight polysomnography in children. *Pediatr Pulmonol* 1992;13:16-21.
  19. Aljadeff G, Gozal D, Schechtman VL, Burrell B, Harper RM, Ward SL. Heart rate variability in children with obstructive sleep apnea. *Sleep* 1997;20:151-7.
  20. 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 Apnea. *IEEE Trans Biomed Eng* 2003;50:686-96.
  21. Shouldice R, de Chazal P, Heneghan C, O'Brien L, Gozal D. Obstructive sleep apnea detection in pediatric subjects using surface lead electrocardiogram characteristic intervals. *Sleep Med* 2003;4:S1-56.
  22. 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;40:402-7.
  23. Guilleminault C, Connolly S, Winkle R, Melvin K, Tilkian A. Cyclical variation of the heart rate in sleep apnea syndrome. Mechanism and usefulness of 24 h electrocardiography as a screening technique. *Lancet* 1984;1:126-31.
  24. Stein PK, Duntley SP, Domitrovich PP, Nishith P, Carney RM. A simple method to identify sleep apnea using Holter recordings. *J Cardiovasc Electrophysiol* 2003;14:467-73.
  25. Shiomi T, Guilleminault C, Sasanabe R, Hirota I, Maekawa M, Kobayashi T. Augmented very low frequency component of heart rate variability during obstructive sleep apnea. *Sleep* 1996;19:370-7.
  26. Roche F, Duverney D, Court-Fortune I, et al. Cardiac interbeat interval increment for the identification of obstructive sleep apnea. *Pacing Clin Electrophysiol* 2002;25:1192-9.
  27. Roche F, Gaspoz JM, Court-Fortune I, Minini P, Pichot V, Duverney D, Costes F, Lacour JR, Barthelemy JC. Screening of obstructive sleep apnea syndrome by heart rate variability analysis. *Circulation* 1999;100:1411-5.
  28. Moody GB, Mark RG, Zoccola A, Mantero S. Clinical validation of the ECG-derived respiration (EDR) technique. In: *Computers in Cardiology*, Piscataway NJ: IEEE Computer Society Press; 1986:507-10.
  29. Travaglini A, Lamberti C, DeBie J, Ferri M. Respiratory signal derived from eight-lead ECG. In: *Computers in Cardiology*, Piscataway NJ: IEEE Computer Society Press; 1998:65-8.
  30. Rechtschaffen A, Kales A. A manual of standardized terminology, techniques and scoring systems for sleep stages of human subject. Washington DC: National Institutes of Health 1968; Pub. No. 204.
  31. American Thoracic Society: Standards and indications for cardiopulmonary sleep studies in children. *Am J Respir Crit Care Med* 1996;153:866-78.
  32. Benitez D, Gaydecki PA, Zaidi A, Fitzpatrick AP. The use of the Hilbert transform in ECG signal analysis. *Computers in Biology and Medicine* 2001;31:339-406.
  33. DeBoer RW, Karemaker JM, Strackee J. Comparing spectra of a series of point events particularly for heart rate variability data. *IEEE Trans Biomed Eng* 1984;31:384-7.
  34. Ripley BD. *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press; 1996.
  35. Duda RO, and Hart PE. *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
  36. Kohavi R. A study of cross validation and bootstrap for accuracy estimation and model selection. *Proc 14th Int Joint Conference on Artificial Intelligence*, 1995;1137-43.
  37. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39:1589.
  38. Rabbia F, Silke B, Conterno A, Grosso T, De Vito B, Rabbone I, Chiandussi L, Veglio F. Assessment of cardiac autonomic modulation during adolescent obesity. *Obes Res* 2003;11:541-8.
  39. Norman RG, Pal I, Stewart C, Walsleben JA, Rapoport DM. Interobserver agreement among sleep scorers from different centers in a large dataset. *Sleep* 2000;23:901-8.
  40. Silvestri JM, Weese-Mayer DE, Bass MT, Kenny AS, Hauptman SA, Pearsall SM. Polysomnography in obese children with a history of sleep-associated breathing disorders. *Pediatr Pulmonol* 1993;16:124-9.
  41. Guilleminault C, Winkle R, Korobkin R, Simmons B. Children and nocturnal snoring: evaluation of the effects of sleep related respiratory resistive load and daytime functioning. *Eur J Pediatr* 1982;139:165-71.
  42. Barker SJ, Shah NK. The effects of motion on the performance of pulse oximeters in volunteers (revised publication). *Anesthesiology* 1997;86:101-8.
  43. Saeed MM, Keens TG, Stabile MW, Bolokowicz J, Davidson Ward SL. Should children with suspected obstructive sleep apnea syndrome and normal nap sleep studies have overnight sleep studies? *Chest* 2000;118:360-5.
  44. Nixon GM, Brouillette RT. Diagnostic techniques for obstructive sleep apnoea: is polysomnography necessary? *Paediatr Respir Rev* 2002;3:18-24.
  45. Flemons WW, Littner MR, Rowley JA, et al. Home diagnosis of obstructive apnea: a systematic review of the literature: an evidence review cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society. *Chest*. 2003;124:1543-79.
  46. Ferber R, Millman R, Coppola M, et al. Portable recording in the assessment of obstructive sleep apnea. *ASDA standards of practice*. *Sleep* 1994;17:378-92.

## APPENDIX

### Quadratic Discriminant Classifier with Mahalanobis Distance

This model provided a posterior probability estimate of the likelihood of a segment belonging to either the normal or apneic class. Training of the quadratic discriminant (QD) classifier using Mahalanobis distance-based estimates of classifier parameters consisted of the following steps. 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:

$$N = \sum_{k=1}^c N_k$$

The  $n^{\text{th}}$  feature vector for training in class  $k$  is designated as  $\mathbf{x}_{nk}$ . Training of the models involves determining the class-conditional mean vectors  $\mu_k$  using:

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N_k} \mathbf{x}_{nk}$$

For a QD classifier the stratified covariance matrices,  $\Sigma_k$ 's, are calculated using:

$$\Sigma_k = \frac{1}{N_k - 1} \sum_{n=1}^{N_k} (\mathbf{x}_{nk} - \mu_k)(\mathbf{x}_{nk} - \mu_k)^T$$



To classify a feature vector  $\mathbf{x}$ , the discriminant value,  $y_k$ , for each class is calculated using:

$$y_k = -(\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)$$

In this study, the estimated posterior probabilities,  $P(k | \mathbf{x})$ , are used. The estimated posterior probabilities can then be calculated from the discriminant values as:

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

The segment is then assigned to the class with the higher posterior probability.