35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

# Automated Detection of Sleep Apnea in Infants using Minimally Invasive Sensors

Gregory Cohen, *Student Member IEEE* and Philip de Chazal, *Member IEEE*

*Abstract*— **To address the difficult and necessity of early detection of sleep apnea hypopnea syndrome in infants, we present a study into the effectiveness of pulse oximetry as a minimally invasive means of automated diagnosis of sleep apnea in infants. Overnight polysomnogram data from 328 infants were used to extract time-domain based oximetry features and scored arousal data for each subject. These records were then used to determine apnea events and to train a classifier model based on linear discriminants. Performance of the classifier was evaluated using a leave-one-out cross-validation scheme and an accuracy of 68% was achieved, with a specificity of 68.6% and a sensitivity of 55.9%.**

## I. Introduction

Sleep apnea hypopnea syndrome (SAHS) is a highly prevalence and under-diagnosed sleep-related breathing disorder in which the cessation of breathing occurs repeatedly during sleep. It can be caused due to an upper airway obstruction (obstructive sleep apnea), a neurological condition in which respiratory function ceases during sleep (central sleep apnea) and a combination of the two (mixed sleep apnea). Obstructive sleep apnea is estimated to affect up to 4% of the adult population [1] and it is also prevalent in very young children, with obstructive sleep apnea estimated to affect between 1% and 4% of infants [2].

Sleep-related breathing disorders have also been linked to several negative effects in the health and development of infants and young children, including behavioral effects such as depression, cognitive impairment and attention-deficit/hyperactivity disorder [3]. It has also tentatively been linked to Sudden Infant Death Syndrome (SIDS) [4]. To exacerbate the problem, studies have shown that young children and infants tend to suffer more severe episodes of sleep apnea [5].

It is estimated that over 80% of cases of sleep-related breathing disorders go undiagnosed [6], primarily due to the limited availability and reliability of appropriate recording and monitoring equipment. The gold standard for the detection and diagnosis of sleep-related breathing disorders is an overnight in-hospital polysomnogram [7], in which a patient's sleep is monitored through a multitude of sensors under controlled laboratory conditions.

Unfortunately, the costs of performing such tests, which require expensive equipment and specially trained staff, makes it prohibitively expensive for widespread diagnostic use and completely unavailable in many countries. The high degree of intrusiveness required by the sensors employed also limit the practicality and effectiveness of the technology. It often does not produce good results when used on young children and infants [8]. As a result, there is a clear need for a less invasive and lower cost means for detecting and diagnosing apnea events, especially in infants and young children for both diagnosis and therapeutic purposes.

One such method of detecting apnea events is through monitoring the levels of oxyhemoglobin saturation (SaO2) in the patient as a desaturation may often result from an apnea event. Direct measurements of SaO2 are difficult to obtain directly, but pulse oximetry (SpO2) provides an indirect and rapid means of measuring SaO2 saturation [9].

This study seeks to determine the potential and effectiveness of using oximetry readings (SpO2), obtained from a minimally-invasive finger sensor, for the automated detection of apnea events in infants.

Oximetry data has been studied as a potential indicator for apnea detection due to its low cost and ease of use. Cyclic desaturations in overnight oximetry recordings is a potential indicator of sleep apneas, but it appears to be limited by the negative predictive value [10] as certain apnea events may not lead to an obvious and detectable desaturation [11]. As oximeter sensors are attached peripherally to the body, they are subject to noise resulting from motion and poor perfusion. However, the majority of this work has been performed on adult subjects, and far less research has been conducted on infants and young children, and as a result, this study investigates the classification properties of a single oximeter reading as an automated classifier specifically in infants.

The study draws upon the polysomnogram recordings found inside the National Collaborative Home Infant Monitoring Evaluation (CHIME) database, collated by the National Institute of Health (NIH), which studied the effectiveness of home monitoring for apnea and bradycardia in infants [12].

## II. Methodology

### A. Data

The training and test data used in this study were obtained from the CHIME dataset. The CHIME database contains extensive recordings from home-based monitors for over 1000 infants, ranging in age from newborns to approximately 27 weeks.

*Research supported by Australian Research Council grant number FT110101098.

G. Cohen is with the MARCS Institute, University of Western Sydney, Australia (Phone: +61 2 4736 0447; fax: +61 2 4736 0833; e-mail: g.cohen@uws.edu.au).

P. de Chazal is with the MARCS Institute, University of Western Sydney, Australia (e-mail: p.dechazal@uws.edu.au).

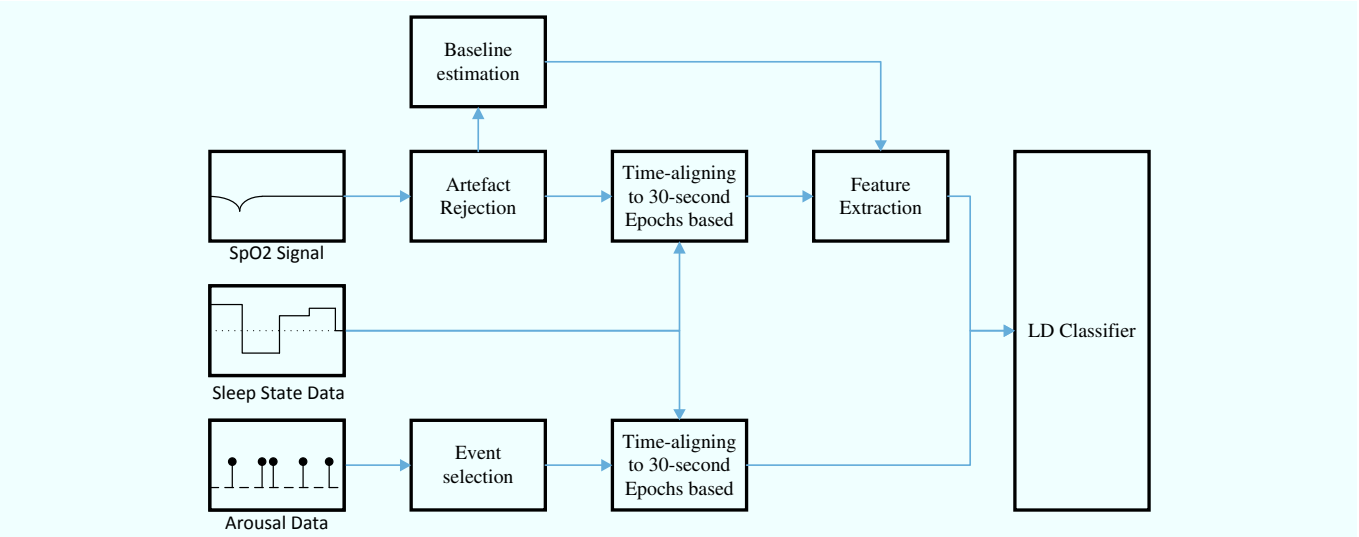


Figure 1: Oximetry-based pre-processing and feature extraction steps

In addition to healthy infants, the subjects also included pre-term infants, infants who have lost sibling to SIDS, and infants exhibiting severe apnea events. A breakdown of the genders and screening conditions for the subjects used in this study are presented in Table I.

The CHIME database also includes 700 polysomnogram recordings, each containing data from 17 different sensors and recorded using a Healthdyne ALICE3 system. The collators of the CHIME database also scored these recordings, determining sleep states and arousal events. This sleep state data was provided in both a raw event format and in a smoothed format.

TABLE I. GENDER AND SELECTION CRITERIA BREAKDOWN

| **Screening Criteria** | **Gender** | | |
|---|---|---|---|
| | ***Male*** | ***Female*** | ***Total*** |
| Apnea of Infancy | 29 | 30 | **59** |
| Healthy Term | 39 | 33 | **72** |
| Premature | 68 | 62 | **130** |
| Sibling of SIDS | 33 | 34 | **67** |
| Total | 169 | 159 | **328** |

The recommendations of the Sleep Disorders Atlas Task Force of the American Sleep Disorders Association were used as a basis for identifying arousal events. The raw signals from the polysomnogram were scored by hand and assigned sleep states. A smoothing algorithm was then applied to the data.

A strict protocol governing both the scoring and sleep-state smoothing process ensured that the methodology was consistent across all subjects.

The home recording data from the CHIME monitors contained six sensors, including an Aequitron ECG/Impedance channel and an Ohmeda pulse-oximeter. As this data does not include expert sleep scoring annotations, it was applicable, and only the polysomnogram data was used in this study.

Both the arousal events and the sleep state data were used in the classification system, along with the raw oximetry data from the polysomnogram recordings. Of the 700 recordings available, only 328 subjects were found to contain complete recordings, annotations and sleep state data, all of which was required for training and testing purposes. As a result, only these records were used in the training and testing of the classifier.

### *B. Pre-processing*

The physiological data, annotations and sleep state information are all located within different parts of the CHIME database and are linked by a uniquely assigned patient identifier. These three separate datasets were combined and processed as shown in Figure 1. The SpO2 readings, smoothed sleep state data and arousal events were extracted from the CHIME database for each patient, and were time-aligned to 30-second epochs, based on the smoothed sleep state data.

The sleep state data was also used to determine the first period of active sleep and the final waking event in order to discard the periods before and after the subject fell asleep. This step was performed to ensure that the classifier was only trained on periods during which the subject was in a state of sleep and to exclude any artifacts or false signals that may have arisen during the setup and teardown of the procedure.

Arousal events were sorted according to their annotated type and were time-aligned to their appropriate epoch. In the case where apnea events spanned multiple epochs, all the affected epochs were marked as containing an epoch event.

Arousal events consisted of central, obstructive and mixed apneas and hypopneas, and were labeled accordingly. Four different sleep states (wake, quiet, non-quiet and indeterminate) were assigned to each epoch using the smoothed sleep state data.

The oximetry data was recorded at a frequency of one reading per second. This data was extracted from the polysomnogram recordings data files and underwent a pre-processing step which iterated over each epoch and removed any obvious artifacts.

Any oximetry values below 65% saturation were automatically excluded, and all values representing a change in oxygen saturation greater than 4% per second were also excluded. These values were removed from their respective epochs, and the epoch itself was only discounted if it contained fewer than 30 values.

In order to approximate the long-term trend within the SpO2 data, an estimated baseline value was calculated to a resolution of one second, in order to match that of the oximeter sensor. These values were then paired up with the corresponding SpO2 value. The baseline was generated using a 5-minute rolling average situated symmetrically about the respective SpO2 value.

### C. Feature Extraction

There has been extensive research done on the subject of feature generation from pulse oximetry sensors. Most of these methods employ time-based statistical properties and are affected by limitations arising from physiological effects, variations in sensor location, and sensor specificity [13].

Standard time-based features include statistical properties such as mean, median value, minimum value, threshold values and various inter-measurement interval calculations [14]. Many of these features suffer from a lack of standardized limits, which make comparisons with other results and studies difficult.

A number of methods have been developed that explore approaches that break away from the standard time-based approach. These include frequency-based features [15], non-linear features [16] and features based on multivariate regression [17].

The bulk majority of the work has been performed on adult subjects and as a result, this study limits itself to well-established, time-based features. Seven different time-based features were calculated for each epoch using both the pre-processed SpO2 values and the associated baseline SpO2 value.

The features calculated for each epoch were as follows:

1. Mean SpO2 Value over the epoch
2. Minimum SpO2 value in the epoch
3. Number of instances below 92% saturation
4. Average absolute rate of change per second in the epoch
5. The 3rd and 57th value in sorted SpO2 values (corresponding to a 5-95% spread)
6. The number of times the baseline value exceeded the SpO2 value by at least 3%
7. The number of times the SpO2 value exceeded the baseline SpO2 by at least 3%

The baseline value attempts to track the long-term trend within the SpO2 data and the two baseline comparisons attempt to provide a proxy for detecting periods of above and below average saturation, which should represent re-saturation and desaturation periods respectively.

### D. Classification

A linear discriminant classifier was used as means for automatic classification. The training data was used to determine the $\mu_k$-class conditional mean vectors and $\Sigma$-common covariance matrix using 'plug-in' maximum likelihood estimates [18].

### E. Epoch-based Performance Measures

The linear discriminant classifier was trained to discriminate between normal and any type of sleep disordered breathing (SDB). Each epoch was either labelled 'Normal' or 'SDB' by the system and the corresponding expert value determined from the arousal data.

Each epoch label by the system was compared to the "expert" data derived from the arousals information and the outcome determined as one of the following:

- True positive (TP): an epoch is labelled as SDB by the arousal data and labelled as SDB by the system.
- True negative (TN): an epoch is labelled as Normal by the arousal data and labelled as Normal by the system.
- False positive (FP): an epoch is labelled as Normal by the arousal data and labelled as SDB by the system.
- False negative (FN): an epoch is labelled as SDB by the arousal data and labelled as Normal by the system.

The number of outcomes over all the epochs were calculated and used to form the two way confusion matrix shown in Table II.

TABLE II. AGE AND SELECTION CRITERIA BREAKDOWN

| | | **Actual** | |
|---|---|---|---|
| | | ***Normal*** | ***SDB*** |
| **Predicted** | ***Normal*** | TN | FN |
| | ***SDB*** | FP | TP |

Using Table II the following performance measures were then calculated:

- Specificity = TN/(TN+FP)
- Sensitivity = TP/(TP+FN)
- Accuracy = (TN+TP)/(TP+TN+FN+FP)

### E. Performance Evaluation

In order to assess the performance of the classifier, a leave-one-out cross-validation scheme was used to assess the ability of the classifier to handle to independent data. Under this scheme, the classifier is trained on all but one recording,

which is then used to test the predictive nature of the classifier. The record used for testing is rotated through each record in the dataset, and the results from all the tests are combined to produce performance measures shown in Table II.

### *F. Implementation*

The data handling, pre-processing and feature extraction methods were written in python and executed on an 80-node Rocks cluster running pyMPI. The classification and performance evaluation code was written in Matlab and run separately.

## III. RESULTS

The results for the classifier are presented in Table III below. It must be stated that these results are the epoch-classification results, and not AHI-classification results. Therefore these values are not directly comparable to many results presented in other papers.

TABLE III. RESULTS

| | ***Specificity*** | ***Sensitivity*** | ***Accuracy*** |
|---|---|---|---|
| 7 SpO2 Features | 68.5% | 55.9% | 68.0% |

The low sensitivity in the above results may have been caused by the fact that many apnoea events do not necessarily lead to a significant drop in oxyhemaglobin saturation levels. This has been indicated as a potential limiting factor of oximetry data and may have a more pronounced effect in infants, where oxygen regulation may exhibit a more dynamic nature.

The results found in [19] can be used as a benchmark as it includes the intermediate epoch-classification results and utilised similar features and overall design. The authors studied 125 adult patients and trained linear determinant classifiers on the oximetry and ECG data separately. The oximetry epoch classifier yielded a specificity of 94.8%, a sensitivity of 71.4% and an accuracy of 88.7%.

## IV. CONCLUSION

The results of the classifier are somewhat lower than those obtained with adult patients, but it is clear that oximetry readings do possess apnea predictive capabilities, although it may require the addition of other sensor paradigms in order to compensate for the high negative predictive value of the oximetry data alone.

The physiological differences between adults and infants may also affect the outcome of the classification technique as certain standard assumptions, such as epoch duration and threshold values, may need to be adjusted and tuned for use infant apnea detection.

Additional features may also be required to better utilize the information present within the SpO2 data.

## V. FUTURE WORK

It is expected that a multimodal approach, using additional minimally invasive sensors, will produce a more accurate and reliable automated apnea classification system.

## REFERENCES


[1] Guilleminault, Christian, Ara Tilkian, and William C. Dement. "The sleep apnea syndromes." *Annual review of medicine* 27.1 (1976): 465-484.

[2] Bixler, Edward O., et al. "Sleep disordered breathing in children in a general population sample: prevalence and risk factors." *Sleep* 32.6 (2009): 731.

[3] Bonuck K, Freeman K, Chervin RD, Xu L, Bonuck K, Freeman K, Chervin RD, Xu L. "Sleep-disordered breathing in a population-based cohort: behavioral out-comes at 4 and 7 years". *Pediatrics*. 129.4 (2012): 857-865.

[4] Steinschneider, Alfred. "Prolonged apnea and the sudden infant death syndrome: clinical and laboratory observations." *Pediatrics* 50.4 (1972): 646-654.

[5] Brouillette, Robert T., Sandra K. Fernbach, and Carl E. Hunt. "Obstructive sleep apnea in infants and children." *The Journal of pediatrics* 100.1 (1982): 31-40.

[6] Vishesh Kapur, M. D., et al. "The medical cost of undiagnosed sleep apnea." *Sleep* 22.6 (1999): 749.

[7] Kushida, Clete A., et al. "Practice parameters for the indications for polysomnography and related procedures: an update for 2005." Sleep 28.4 (2005): 499-521.

[8] Katz, Eliot S., Ron B. Mitchell, and Carolyn M. D'Ambrosio. "Obstructive Sleep Apnea in Infants." *American journal of respiratory and critical care medicine* 185.8 (2012): 805-816.

[9] Morillo, Daniel Sánchez, and Nicole Gross. "Probabilistic neural network approach for the detection of SAHS from overnight pulse oximetry." *Medical and Biological Engineering and Computing* (2012): 1-11.

[10] Chiner, Eusebi, et al. "Nocturnal oximetry for the diagnosis of the sleep apnoea hypopnoea syndrome: a method to reduce the number of polysomnographies?." *Thorax* 54.11 (1999): 968-971.

[11] Sériès, Frédérick. "Interpretation of home oximetry tracings." *Chest Journal* 121.3 (2002): 1006-1007.

[12] Crowell, David H., et al. "Infant polysomnography: reliability. Collaborative Home Infant Monitoring Evaluation (CHIME) Steering Committee." *Sleep* 20.7 (1997): 553.

[13] Morillo DS, Gross N, Leo´n A, Crespo LF (2012) "Automated frequency domain analysis of oxygen saturation as a screening tool for SAHS". *Med Eng Phys* 34(7):946–953

[14] Herer B, Roche N, Carton M, Roig C, Poujol V, Huchon G (1999) "Value of clinical, functional, and oximetric data for the prediction of obstructive sleep apnea in obese patients." *Chest* 116:1537–1544

[15] Zamarron, C., et al. "Oximetry spectral analysis in the diagnosis of obstructive sleep apnoea." *Clinical Science* 97 (1999): 467-473.

[16] Morillo, Daniel S., et al. "Poincaré analysis of an overnight arterial oxygen saturation signal applied to the diagnosis of sleep apnea hypopnea syndrome." *Physiological measurement* 30.4 (2009): 405.

[17] Magalang, Ulysses J., et al. "Prediction of the apnea-hypopnea index from overnight pulse oximetry." *Chest* Journal 124.5 (2003): 1694-1701.

[18] Ripley, Brian D. "Pattern recognition and neural networks." *Cambridge university press*, 2008.

[19] de Chazal, P., et al. "Home-based assessment of sleep apnea using simultaneous electrocardiogram and oximtery signals." *Progress in sleep apnea research* (2007): 115-139.