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

Detection of Sleep Apnea on a Per-Second Basis Using Respiratory Signals

Nandakumar Selvaraj* and Ravi Narasimhan

Abstract— **There has been a growing interest in out-of-center sleep testing with portable devices for accurate diagnosis of sleep apnea syndrome. This paper presents a new algorithm that extracts features based on filtering and statistical dispersion of the nasal airflow respiration signal and detects apnea events on a per-second basis. The data records were randomly selected from the Sleep Heart Health Study (SHHS-2) database to represent 100 control subjects with Apnea-Hypopnea Index (AHI) less than 5, and 100 apnea subjects with AHI values from 30 to 75. The algorithm was optimized according to the product of sensitivity and positive predictive value of apnea events among a training dataset of 50 apnea subjects with a constraint on the false positive rate among a training dataset of 50 control subjects. From testing of the algorithm on separate datasets, the false positive rate among 50 control subjects was found to be 1.3 events per hour, which corresponds to 100% specificity of classifying apnea subjects. The sensitivity and positive predictive value among 50 apnea subjects were found to be 83.6% and 72.3%, respectively. Among the identified false positive events in the apnea subjects, 64.3% of the events were found to be hypopnea events. Thus, incorporation of hypopnea detection would enhance the performance of the apnea detection algorithm.**

Index Terms— **Obstructive sleep apnea, Apnea-Hypopnea Index, Respiration, Polysomnography.**

I. INTRODUCTION

Sleep Apnea Syndrome (SAS) is a major sleep disorder that causes recurrent episodes of complete (apnea) or partial (hypopnea) blockage of the upper airway during sleep. One of the metrics that quantify the severity of this disorder is the Apnea-Hypopnea Index (AHI) index, which is the number of apnea and hypopnea events per hour (EPH) averaged over the duration of sleep. The prevalence of SAS is approximately 3 to 7% in adult men and 2 to 5% in adult women [1]. Disease prevalence is higher in different population subsets such as obese or older groups. For example, for a specific age group of 30 to 60 years, the prevalence of an AHI ≥ 5 was found to be 24% in men and 9% in women [2]. Moreover, the proportion of undiagnosed SAS patients is found to be much higher. Among a sample of 4,925 employed adults in the general adult population, the undiagnosed were 98% of women and 90% of men with mild SAS, and 93% of women and 82% of men with moderate to severe SAS [3]. Routine clinical visits and blood tests usually do not detect SAS. Furthermore, SAS diagnosis is important since there are adverse health consequences of this disorder, including daytime hypersomnolence, neurocognitive dysfunction, cardiovascular disease, metabolic dysfunction and respiratory failure [4].

Authors are with Vital Connect Inc., Campbell, CA 95008, USA. (*corresponding author's e-mail: nselvaraj@vitalconnect.com)

Polysomnography (PSG) is the gold standard test for diagnosis of SAS. The PSG test involves the monitoring during overnight sleep of multiple physiological signals that include electroencephalography (EEG), airflow, thoracic and/or abdominal respiratory inductive plethysmography (RIP) and oxygen saturation. PSG is a laboratory-based test generally performed at a sleep center that may affect normal sleep patterns; PSG involves high operating costs, in part because of dedicated equipment, facilities and personnel. Moreover, the analysis of PSG records is very time consuming and often varies based on the subjective interpretation of medical experts. Because of the number of disadvantages of PSG, there has been a growing interest in simplified and less expensive out-of-center sleep testing with portable devices for SAS diagnosis.

Among the portable SAS monitor categories, Type-4 monitors utilize at least one channel, usually either oxygen saturation or airflow. Various respiratory and effort sensors alone or their combinations are not sufficiently accurate for SAS diagnosis [5]. One potential problem could be that many algorithms analyze these signals with a poor time resolution ranging from 15 seconds to a few minutes. Such large temporal windows may extend detected apnea/hypopnea events in a single segment to many segments and, hence, may result in several false positive events. Such behavior may result in an overestimation of AHI values.

In order to overcome these limitations, algorithms are needed that classify apnea events with a fine time resolution. This paper presents such an algorithm that extracts features from the nasal airflow respiration (NAR) signals and detects apnea events on a per-second basis.

II. METHODS

A. Apnea Data Selection

The Sleep Heart Health Study (SHHS) is a prospective cohort study held during 1995-1998 to investigate SAS and other sleep-disordered breathing as risk factors for the development of cardiovascular disease. In that study, 6441 individuals aged 40 years or older and not being treated for sleep apnea were recruited to undergo an overnight home PSG, complete several questionnaires, and undergo physical examination. A second PSG database (SHHS-2) was obtained from 3295 individuals during 2001-2003. Full details of the SHHS study designs are found in [6].

The records of the SHHS-2 PSG database have been used in this paper. One hundred PSG records were randomly selected each for the control group with AHI < 5 and for the apnea group with AHI values 30 to 75. Furthermore,

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

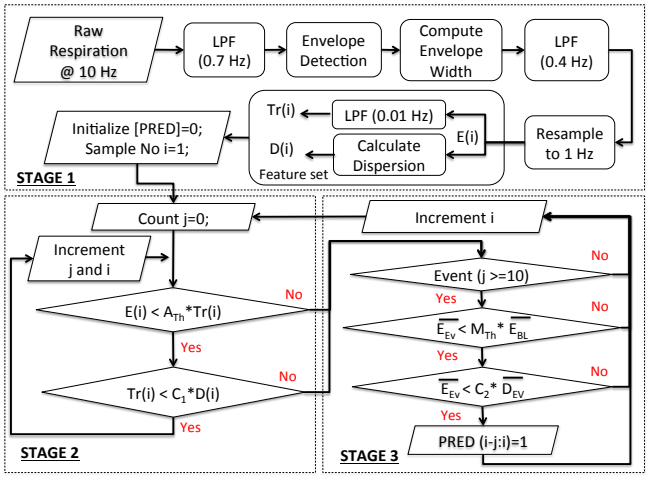

Fig. 1. Flow chart of the apnea event detection algorithm. E, lowpass-filtered envelope width; Tr, very low frequency trend of E; D, statistical dispersion of E for a moving window of 120 seconds.

training and testing datasets (each with sample size $n = 100$) were randomly picked to represent 50 control data and 50 apnea data in each dataset. The present algorithm is based on the NAR signal of the PSG record; the NAR signal, which is sampled at 10 Hz, is generated from an oral-nasal thermocouple (Preotec, Woodinville, WA).

The respiratory event annotations include apnea and hypopnea events with start times and durations. Apnea events were identified if the peak-to-trough amplitude of the airflow signal is less than 25% of the amplitude of the preceding baseline period (with regular breathing and stable oxygen levels) for at least 10 seconds. On the other hand, hypopnea events were identified if the amplitude of any respiratory signal decreases below 70% of the amplitude of the baseline for at least 10 seconds and 2 breaths. The apnea events were further identified as 'obstructive', 'central' or 'mixed' based on whether the effort to breathe (observed as displacement in either chest RIP or abdominal RIP signals) is present during the event. The current algorithm is designed to detect apnea (but not hypopnea) events based on features extracted from the NAR signal and does not distinguish presently between the types of apnea events.

B. Algorithm Development

The flow chart of the apnea detection algorithm is given in (Fig. 1). The algorithm consists of three distinct stages. The feature extraction is the first stage where, the raw NAR signal is low-pass filtered at a passband corner frequency (fc) of 0.7 Hz (the normal bandwidth of the NAR signal during sleep). Next, all local maxima and local minima points are interpolated using piecewise cubic Hermite interpolation to obtain upper and lower envelopes. The width of the envelope is obtained as the difference signal between upper and lower envelopes and would reflect the changes in instantaneous respiratory amplitudes. The width of the envelope is lowpass filtered at fc=0.4 Hz and later resampled to 1 Hz to obtain E as the first feature for apnea event detection. The feature E retains the variability of the respiratory instantaneous amplitudes up to 0.4 Hz, since the power beyond 0.4 Hz is found to be negligible among all the 200 subjects. The signal E is again low-pass filtered at fc=0.01 Hz to obtain the second feature Tr as an adaptive trend that exhibits very low frequency variability of respiratory instantaneous amplitudes. Note that the spectrum of NAR signal exhibits increased power in the range of 0.01-0.1 Hz during apnea events [7]. All the above three filters are 4^{th}-order elliptic filters with a passband ripple of 0.5 dB and a stopband attenuation of 30 dB. In addition, the statistical dispersion in the E signal is calculated as the difference between 90^{th} and 10^{th} percentile values for a moving window of duration 120 seconds and denoted as the third feature D. The window length is chosen in order to minimize the false positive events that may arise from long apnea episodes. The feature D together with Tr detects transitions between normal and apnea episodes.

At the second stage of the algorithm, the event detection signal on a per-second basis is initially set to 0, which indicates the absence of an apnea event. The feature signals are analyzed for each sample index i (i.e., on a per-second basis). The algorithm checks if the following two conditions are satisfied on a per-sample basis to detect a candidate event:

1) The instantaneous amplitude of the NAR signal (E) is less than A_{Th} times the trend mean (Tr), where A_{Th} refers to an amplitude threshold, and
2) The transient decrease in E amplitudes during apnea skew its probability distribution and increase the range of its statistical dispersion (D) such that the Tr is less than C_1 times D.

As long as these two conditions are satisfied on a per-second basis, a count j is incremented. When a sample does not satisfy either of these conditions, the count is not updated. If the current count $j < 10$, the count j is reset to 0 and the next sample $i \leftarrow i + 1$ is processed. Otherwise, a candidate event is detected. The intuition behind the above logic is that during a normal respiratory event, the signal E may decrease below the trend mean briefly but may not stay lower for at least 10 seconds. Secondly, the dispersion metric D would also typically be lower than the trend mean during a normal respiratory event, but an apnea event would cause D to increase above the trend mean.

At the third stage, once a candidate event is detected that consists of the past $j + 1$ samples, the algorithm compares two conditions on a per-event-basis with respect to a baseline, which refers to a window consisting of the preceding 10 samples to the candidate event. This step exploits the linear relationships between the features and helps to reduce the probability of detecting false positive events. The first condition is that the mean of E for an apnea event ($\overline{E_{Ev}}$) should be less than M_{Th} times the mean of E during baseline ($\overline{E_{BL}}$), where M_{Th} refers to the mean threshold, and the subscripts "Ev" and "BL" denote "event" and "baseline," respectively. The second condition is that $\overline{E_{Ev}}$ should also be less than C_2 times the mean dispersion metric of the same

event ($\overline{D_{Ev}}$). If the above two conditions are satisfied, the candidate window is declared as an apnea event; otherwise, no event is detected.

The intuition behind this logic is that the respiratory signal decreases to less than 25% of the baseline values for at least 10 seconds during an apnea event; thus the $\overline{E_{Ev}}$ should be less than 25% of $\overline{E_{BL}}$. Instead of selecting an arbitrary value of 25% as the mean threshold, the optimal value of M_{Th} is selected from the training data. Moreover, the difference between the $\overline{D_{Ev}}$ and the $\overline{E_{Ev}}$ will be larger for an apnea event than for a hypopnea event. Therefore, by also optimizing the constant C_2, the probability of false positives originating from hypopnea events can be reduced. Thus, the parameters A_{Th}, C_1, M_{Th} and C_2 are determined to maximize the product of the sensitivity (Se = TP / (TP+FN) * 100) and the positive predictive value (PPV = TP / (TP+FP) * 100) for the apnea subjects of the training set. TP, FP and FN are referred to as true positives, false positives and false negatives, respectively.

C. Data Analysis

The data analysis was carried out offline using Matlab. The three feature signals and the reference annotations and were obtained for all the 100 training subjects and stored. Selection of optimal values for the variables A_{Th}, C_1, M_{Th} and C_2 were carried out in a two step process process. In the first step, the four parameters were coarsely chosen as arbitrary vectors (0 to 2 with an increment of 0.2), and for every combination of these parameters, the false positive rate (FPR_C in EPH) was obtained for all 50 control subjects in the training set. FPR_C is calculated as the total number of detected false positive events over the total number of hours. In addition, for all the 50 apnea subjects in the training set, Se_A, PPV_A and FPR_A were obtained. The grid search was carried out for each FPR_C between 1 and 4 with an increment of 0.1 (since the target FPR in control subjects should be less than 5) to find the highest probability (product of Se_A and PPV_A) of apnea detection. With the identified course values of four parameters, another grid search was carried out by repeating the above process with fine values of four parameters. The performance of the apnea event detector algorithm was evaluated on the test data using the optimized parameters. On a subject-by-subject basis as well as overall, the Se_A and PPV_A in %, and FPR_A in EPH were computed from apnea test group, and FPR_C in EPH was from the control test group. These values are given in mean ± SE (standard error).

III. Results

The optimal parameters for the apnea detector algorithm are found to be A_{Th}=1.42, C_1=0.8, M_{Th}=0.92, and C_3=0.22; the performance on the training data set is FPR_C=1.3 EPH, Se_A=84.1%, and PPV_A=72.6%. An example of the algorithm processing is shown in Fig. 2 for an overnight apnea test record, where the respiration signal is given in the top panel. The derived features, reference and detected annotations for each second (marked as o and +, respectively) are given in the bottom panel. This record shows an example of high sensitivity of 98% for apnea detection. Fig. 3 is a zoomed segment of Fig. 2 that illustrates the transition between apnea events followed by a control segment with the derived features. In this example, there is good agreement of the locations and durations of apnea events compared to the reference annotations. Boxplots of the performance of the apnea detection algorithm over all the test subjects are given in (Fig. 4). Analysis of the 50 control test subjects revealed FPR_C = 1.1 ± 0.1 EPH (range of 0 to 3.1 EPH) on a subject-by-subject basis, and 1.3 EPH overall. This result indicates good performance of the algorithm on 50 control subjects whose actual AHI values are less than 5. The range of FPR_C implies that all the 50 control subjects are correctly classified as controls, corresponding to a specificity of 100%.

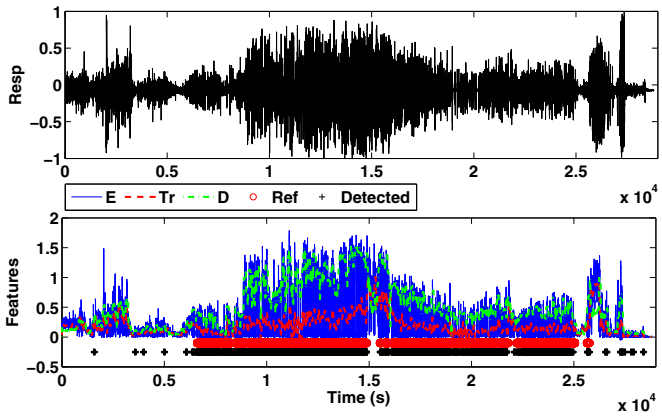

Fig. 2. Example of apnea event detection on a per-second basis. Nasal airflow respiration signal (top panel), the derived features, reference 'o' and detected apnea '+' annotations (bottom panel) are given for an overnight recording. The x and y axes are given in seconds and arbitrary units respectively.

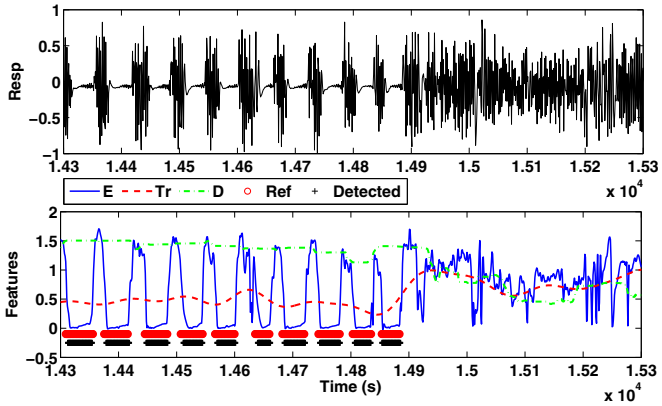

Fig. 3. Zoomed segment of the overnight data given in Fig. 2. In this segment, all ten apnea events are correctly detected. Furthermore, detected and reference markers indicate close correspondence on a per-second basis.

Analysis of the 50 apnea test subjects resulted in Se_A = (83.4 ± 1.8)%, PPV_A = (69.4 ± 1.9)% and FPR_A = 4.4 ± 0.4 EPH on a subject-by-subject basis, and Se_A = 83.6%, PPV_A = 72.3%, FPR_A = 5.0 EPH overall. While

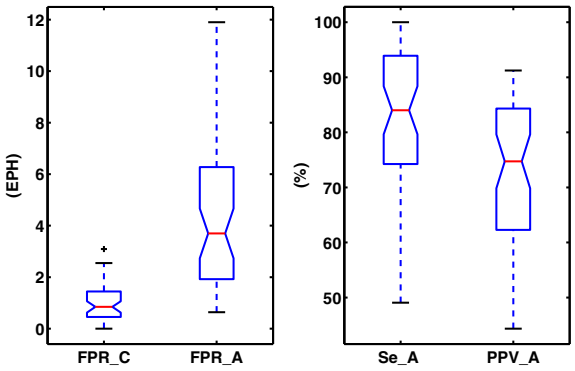

Fig. 4. Boxplots of false positive rate (FPR in EPH, events per hour) obtained in control (denoted with C) and apnea (denoted with A) test subjects (left panel). Boxplots of sensitivity (Se) and positive predictive value (PPV) both in (%) obtained for the apnea group (right panel).

the algorithm has high sensitivity in detecting the apnea events, the PPV is relatively low. A major reason for the low PPV is that 64.3% of the detected false positive events are found to be hypopnea events.

IV. Discussion

This paper discusses a novel approach for the detection of apnea events on a per-second basis using NAR signals alone. The current apnea detection algorithm has a false positive rate of 1.3 EPH in 50 control subjects and a sensitivity of 83.4% in 50 apnea subjects.

From the literature, the cyclic variations in the heart rate that result in bradycardia-tachycardia events have been extensively investigated for the detection of sleep apnea on a per-minute resolution [8]. The cyclic variations in heart rate in part depend upon the sleep stage, degree of desaturation, and aging and are often superimposed on other cardiovascular phenomena such as respiratory sinus arrhythmia, baroreflex and autoregulation mechanisms. Therefore, many individuals with SAS may not demonstrate prominent changes cyclic variations in heart rate. Hence, electrocardiography (ECG)-based SAS diagnosis or screening is yet to be incorporated into clinical practice [9].

The collapse of the upper airways is eventually followed by oxygen desaturation that often leads to an arousal, which is the activation of the central nervous system that changes sleep to a lighter stage and abruptly changes the EEG signal. According to the American Sleep Disorders Association, only 75% of apnea events are terminated by an EEG arousal [10]. Detection of arousals based on EEG features is a complex, time-consuming, and manual process. On the other hand, the oxygen desaturation events are typically delayed by ten or more seconds after the onset of an apnea. Therefore, the current accepted clinical practice for the detection of apnea episodes is using respiration as a primary signal and desaturation events and EEG arousals as secondary features [11].

The current algorithm for apnea detection is fully automated and very robust against noise. Since all the features are derived from the width of the envelope, the algorithm is relatively insensitive to baseline wander of the NAR signal caused by motion artifacts. The algorithm takes less than 20 seconds on a 2.7 GHz Intel Core i7 MacBook Pro laptop computer for the analysis of an overnight PSG recording, and does not require any patient-specific information. The low computational power requirement of the algorithm may allow for real-time analysis as well as implementation in portable devices. The algorithm is also applicable to other types of respiration signals such as RIP effort signals for apnea detection.

Our future research will focus on the improvement of the algorithm by detecting hypopnea events. Inclusion of additional channels such as oxygen desaturation or respiratory effort based on RIP may increase the ability to discriminate between apnea-hypopnea events as well as the types of apnea events. Such an enhanced detector may enable accurate AHI estimation with less obtrusive out-of-center sleep testing.

Acknowledgment

The authors thank A. Chan for helpful discussions.

References

[1] N. Punjabi, "The epidemiology of adult obstructive sleep apnea," *Proceedings of the American Thoracic Society*, vol. 5, no. 2, pp. 136–143, 2008.

[2] T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, "The occurrence of sleep-disordered breathing among middle-aged adults," *New England Journal of Medicine*, vol. 328, no. 17, pp. 1230–1235, 1993.

[3] T. Young, L. Evans, L. Finn, M. Palta *et al.*, "Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women." *Sleep*, vol. 20, no. 9, p. 705, 1997.

[4] S. Patil, H. Schneider, A. Schwartz, and P. Smith, "Adult obstructive sleep apneapathophysiology and diagnosis," *Chest Journal*, vol. 132, no. 1, pp. 325–337, 2007.

[5] W. Flemons, M. Littner, J. Rowley, P. Gay, W. Anderson, D. Hudgel, R. McEvoy, and D. Loube, "Home diagnosis of sleep apnea: A systematic review of the literaturean evidence review cosponsored by the american academy of sleep medicine, the american college of chest physicians, and the american thoracic society," *CHEST Journal*, vol. 124, no. 4, pp. 1543–1579, 2003.

[6] B. Lind, J. Goodwin, J. Hill, T. Ali, S. Redline, and S. Quan, "Recruitment of healthy adults into a study of overnight sleep monitoring in the home: experience of the sleep heart health study," *Sleep Breath*, vol. 7, no. 1, pp. 13–24, 2003.

[7] D. Álvarez, G. Gutie rrez, J. V. Marcos, F. del Campo, and R. Hornero, "Spectral analysis of single-channel airflow and oxygen saturation recordings in obstructive sleep apnea detection," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 847–850.

[8] T. Penzel, J. McNames, P. De Chazal, B. Raymond, A. Murray, and G. Moody, "Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings," *Medical and Biological Engineering and Computing*, vol. 40, no. 4, pp. 402–407, 2002.

[9] P. de Chazal, T. Penzel, and C. Heneghan, "Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram," *Physiological measurement*, vol. 25, no. 4, p. 967, 2004.

[10] N. Douglas and S. Martin, "Arousals and the sleep apnea/hypopnea syndrome," *SLEEP-NEW YORK-*, vol. 19, pp. 196–197, 1996.

[11] C. Kushida, M. Littner, T. Morgenthaler, C. Alessi, D. Bailey, J. Coleman Jr, L. Friedman, M. Hirshkowitz, S. Kapen, M. Kramer *et al.*, "Practice parameters for the indications for polysomnography and related procedures: an update for 2005," *Sleep*, vol. 28, no. 4, pp. 499–521, 2005.