34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Detection of Breathing Segments in Respiratory Signals*

Carlos A. Robles-Rubio, *Student Member, IEEE*, Karen A. Brown, and R. E. Kearney, *Fellow, IEEE*

Abstract— **The typical approach for analysis of respiratory records consists of detection of respiratory pauses and elimination of segments corrupted by movement artifacts. This is motivated by established rules used for manual scoring of respiratory events, which focus on pause segmentation and do not define criteria to identify breathing segments. With this strategy, breathing segments can only be inferred indirectly from the absence of abnormalities, yielding an unclear and ambiguous definition. In this work we present novel detectors for synchronous and asynchronous breathing, and compare them with AUREA, a novel system for Automated Unsupervised Respiratory Event Analysis, which performs indirect classification of breathing. Results from analysis of real infant respiratory data show an improvement in the identification of synchronous and asynchronous breathing of 9% and 27% respectively, demonstrating that direct detection of breathing enhances the classification performance.**

I. Introduction

Studies of respiratory activity typically require the analysis of long data records acquired in sleep laboratories, postoperative recovery rooms, or in the home [1, 2]. Respiratory Inductive Plethysmography (RIP) is frequently used in these studies to record the breathing pattern, because it provides robust and noninvasive monitoring of respiratory effort in the ribcage and abdomen.

The most common method of analysis for these data is manual scoring, performed off-line by trained experts using a set of predefined rules. The focus of this analysis is on pause segmentation, consequently, the established rules for manual scoring (e.g., [3]) lack of definitions for other respiratory states, such as normal breathing, whose identification is only inferred indirectly from the absence of abnormalities. This unclear and ambiguous identification of breathing results in large intra and inter-scorer variability.

Manual classification is labor intensive, expensive and suffers from low inter and intra-scorer agreement [4]. Attempts to automate the analysis have encountered difficulties when applied to clinical data, mainly because of problems dealing with segments corrupted by movement artifacts [5]. For this reason, we developed an automated, supervised method that uses uncalibrated RIP signals to classify the respiratory state into one of four categories: pauses, movement artifacts, asynchronous (ASB) and synchronous breathing (SYB) [2]. This effectively separates artifact corrupted segments, while still identifying the remaining states. The method defined detection statistics for pause, movement artifact and ASB, and compared them to thresholds to identify the corresponding events. Segments not assigned to one these three categories were classified as SYB (i.e., this last category was determined as the absence of other states). This approach is not robust to bad signal segments (e.g., high noise, absent data, broken leads, technical problems), which are not detected by the movement, pause or ASB detectors and so are considered SYB by default.

Although fully automated, the supervised classification method in [2] requires a sample of manual scores provided by an expert to determine the optimum thresholds for the statistics. This is again labor intensive, time consuming, and limited by the subjective judgment of the expert and the intra-scorer reliability. To overcome this, we recently developed AUREA [6], a novel system for Automated Unsupervised Respiratory Event Analysis that requires no human intervention. It uses the detection statistics developed for the supervised method as inputs, to classify the respiratory records into one of the four basic states, using a modified implementation of k-means clustering that corrects for the unbalanced sampling observed in respiratory data (e.g., SYB is usually > 50% of the recordings, while ASB is only < 10%). Even though it exploits better the multivariate relationship of the inputs, AUREA still suffers from the lack of a SYB detection statistic, which limits its ability to detect bad signal data. To overcome this problem, we have developed a detector for SYB that fully distinguishes it from all other types of activity.

The paper is organized as follows: Section II describes the statistics developed for detection of breathing in RIP signals; Section III describes the acquisition of infant data and analysis procedures used to evaluate the new statistics; Section IV reports the performance results; and Section V provides concluding remarks.

II. Methods

A. Detection of Synchronous Breathing

Previously, we observed that for infants most respiratory-related power lies in a frequency band (0.4 – 2.0 Hz) higher than that for pauses and movement artifacts (0 – 0.4 Hz) [2, 7]. We used this observation to design detection statistics to distinguish accurately between breathing and non-breathing segments as follows.

First, to remove low frequency trends and have a scale-independent representation, the uncalibrated ribcage and abdomen signals (termed *rc* and *ab* respectively) were converted to binary signals. To do so, we first computed the moving average value of *rc* over a window of length N_b as follows:

*Research supported in part by the Natural Sciences and Engineering Research Council of Canada. The work of CARR was supported in part by the Mexican National Council for Science and Technology.

CARR and REK are with the Department of Biomedical Engineering, McGill University, Montreal QC, H3A 2B4, Canada (e-mail: carlos.roblesrubio@mail.mcgill.ca; robert.kearney@ mcgill.ca).

KAB is with the Department of Anesthesia, McGill University Health Center, Montreal QC, H3A 2B4, Canada (e-mail: karen.brown@mcgill.ca).

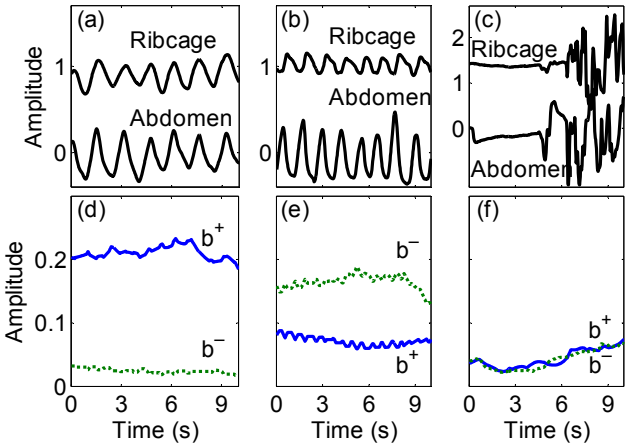

Figure 1. Top: Representative signals for (a) synchronous breathing, (b) asynchronous breathing, and (c) non-breathing (pause and movement artifact). Bottom: Values of the synchronous (b^+) and asynchronous (b^-) breathing statistics computed from the representative segments of (d) synchronous breathing, (e) asynchronous breathing, and (f) non-breathing.

$$\mu^{rc}[n] = \frac{1}{N_b} \sum_{k=n-(N_b-1)/2}^{n+(N_b-1)/2} rc[k]. \quad (1)$$

Additionally, to reduce the effect of noise, a smoothed version of the original ribcage signal (termed rc_s) was obtained using (1) with a window length $N_{avg} << N_b$. These signals were then used to convert rc to a binary signal as $rc_b[n] = 1$ if $rc_s[n] > \mu^{rc}[n]$ and $rc_b[n] = 0$ otherwise. The abdomen signal was converted similarly to $ab_b[n]$.

During completely synchronous breathing periods (i.e., when the phase, θ, between rc and ab is $= 0^o$), $rc_b[n]$ and $ab_b[n]$ are expected to be equal for most n values. In contrast, during paradoxical asynchronous breathing (i.e., $\theta = 180^o$) these signals are expected to have opposite values. We used this principle to design a detection statistic to distinguish between SYB and all other respiratory states, including ASB.

To this end, we defined the sum signal $su[n] = (rc_b[n] + ab_b[n]) / 2$. During SYB the value of this signal should oscillate between 0 and 1 at a frequency in the breathing band. In contrast, during paradoxical ASB, $su[n]$ should remain nearly constant at 0.5, while during pauses and movement artifacts it would vary at frequencies lower than the breathing band (i.e., at 0 – 0.4 Hz), given the slow changes on the signals during these states.

To extract the power associated with synchronous breathing only, $su[n]$ was high-pass filtered at a cutoff frequency at the lower edge of the breathing band ($f_c = 0.5$ Hz), using zero-phase forward-backward filtering with a digital high-pass elliptical filter of order 4. The peak-to-peak ripple in the filter pass band was set to 0.1 dB, and the minimum attenuation in the stop band was 50 dB.

We used the high-pass filtered sum signal $su_{hp}[n]$ to define the SYB detection statistic:

$$b^+[n] = \frac{1}{N_b} \sum_{k=n-(N_b-1)/2}^{n+(N_b-1)/2} su_{hp}^2[k], \quad (2)$$

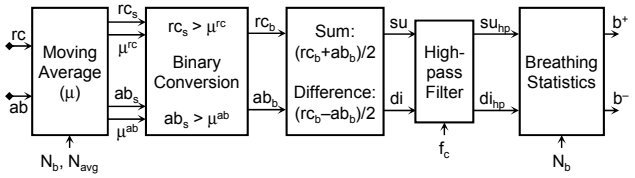

Figure 2. Block diagram for the computation of the synchronous (b^+) and asynchronous (b^-) breathing statistics.

where N_b is the length of the window used to estimate b^+. During SYB, b^+ is expected to have higher values than during ASB and non-breathing segments.

B. Detection of Asynchronous Breathing

We also observed that the difference signal, $di[n] = (rc_b[n] – ab_b[n]) / 2$, oscillated between -0.5 and 0.5 at a frequency in the breathing band during paradoxical asynchronous breathing, stayed constant at 0 during complete synchrony, and varied at low frequency during non-breathing. The difference signal was high-pass filtered to extract the power associated with asynchronous breathing, and the resulting high-pass filtered difference signal $di_{hp}[n]$ used to define the ASB detection statistic:

$$b^-[n] = \frac{1}{N_b} \sum_{k=n-(N_b-1)/2}^{n+(N_b-1)/2} di_{hp}^2[k], \quad (3)$$

During ASB, b^- is expected to have higher values than during all other respiratory states. Figs. 1(d), 1(e), and 1(f) show examples of b^+ and b^- during SYB, ASB and non-breathing respectively.

The complete procedure for the computation of b^+ and b^- is summarized in the block diagram in Fig. 2.

III. Evaluation Procedures

A. Subjects and Data Acquisition

To evaluate the detection statistics, we acquired data from 16 infants (12 males), postconceptional age 42.8±2.1 weeks, weight 3.7±1.0 kg, in the postoperative period after elective inguinal herniorrhaphy with general anesthesia. Written informed parental consent was obtained and the study was approved by the Institutional Ethics Review Board of the Montreal Children's Hospital (MCH). These subjects have been previously reported in [6].

Upon arrival at the postoperative recovery room, infant respibands were placed around the ribcage and abdomen and interfaced with RIP (Ambulatory Monitoring Inc., Battery Operated Inductotrace, Ardsley, NY, USA). An infant oximeter (Nonin 8600 Portable Digital Pulse Oximeter, Plymouth, MN, USA) was taped to a hand or foot. We recorded ribcage and abdominal RIP, as well as oxygen saturation and photoplethysmography for 6 to 12 hours in accordance with the MCH practice guidelines for apnea monitoring in term and former preterm infants. The analogue signals were low-pass filtered (cutoff frequency 10 Hz) with an 8-pole Bessel anti-aliasing filter (Kemo, Jacksonville, FL, USA), digitized and sampled at $F_s=50$ Hz.

B. Manual Classification

To provide a "gold" standard for comparison with the presented methods, one of the investigators (KAB) used an in-house interactive, graphical, manual classification tool, to manually classify the respiratory state for the first 11 infant data sets acquired. Only a subset of the 16 recordings was manually classified due to the time consuming and labor intensive nature of the task. The scorer collaborated with experience and clinical expertise to the design of objective manual scoring rules for RIP data and the understanding of the various respiratory states, but was independent from the development of the automated methods. Segments were classified into one of three categories:

Synchronous Breathing: Quasi-sinusoidal waveforms in both ribcage and abdomen and a phase difference $\theta \leq 90^o$, assessed by a breath-by-breath visual inspection (Fig. 1(a)).
Asynchronous Breathing: Quasi-sinusoidal waveforms in both ribcage and abdomen with a $\theta > 90^o$, (Fig 1(b)).
Non-breathing: Segments in the respiratory records which do not present a quasi-sinusoidal waveform and are associated with events other than breathing (e.g., pause, movement artifact, bad signal) (Fig. 1(c)).

A total of 100 hours and 48 minutes were analyzed in this way, yielding 41,691 segments (SYB: 22,574 (68 hrs); ASB: 5,400 (7 hrs); Non-breathing: 13,717 (26.6 hrs)).

C. Individual Detection Performance

A detection problem consists of selecting between two hypotheses, one defined by the presence of the event of interest (H_1), and another defined by its absence (H_0). A detector is the combination of a detection statistic and a decision rule, values above a given threshold (γ) are assigned to one of the hypotheses, and the rest to the other. The statistic's performance as a function of the threshold may be summarized in its Receiver Operating Characteristics (ROC), which shows the probability of detection (P_D) as a function of the probability of false alarm (P_{FA}) (see Fig. 3(b)).

We used the manual classifications to estimate two nonparametric probability density functions (PDFs) for each detection statistic: one for samples classified as the event of interest (i.e., H_1), and one for samples classified as any other event (i.e., H_0) (see Figs. 3(a) and 4(a)). These PDFs were then used to generate the ROC curves. Each P_D and P_{FA} pair in the ROC curve corresponds to a unique threshold value. For each detector, we selected the optimal threshold γ_{opt} as the point farthest from the chance line ($P_D = P_{FA}$), as the best tradeoff between P_D and P_{FA} [2] as shown in Figs. 3(b) and 4(b). We also computed the area under the ROC curve (AUC).

D. Agreement with Expert Scorer

We assessed the overall performance of the breathing detection statistics by combining them to determine the respiratory state as follows:

1) Samples where $b^+[n] > \gamma_{b^+opt}$ were labeled SYB.
2) From the remaining samples, those where $b^-[n] > \gamma_{b^-opt}$ were labeled ASB.
3) All other samples were labeled non-breathing.

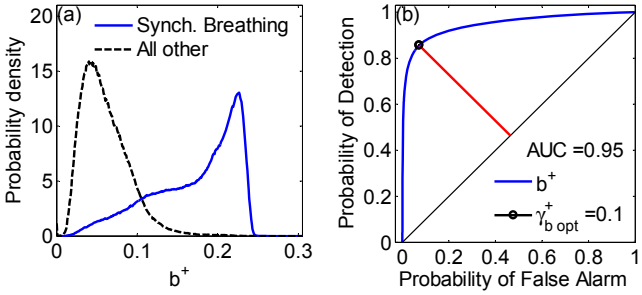

Figure 3. Synchronous breathing detection. (a) Probability densities of the synchronous breathing detection statistic (b^+) for all samples manually identified as Synchronous Breathing or any Other. (b) Receiver operating characteristics (ROC) of b^+, with optimum threshold (γ_{b^+opt}) and area under the ROC curve (AUC).

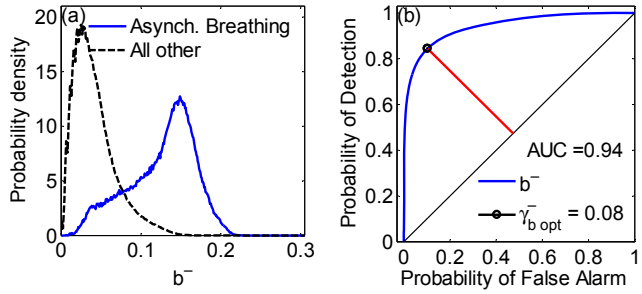

Figure 4. Asynchronous breathing detection. (a) Probability densities of the asynchronous breathing detection statistic (b^-) for all samples manually identified as Asynchronous Breathing or any Other. (b) Receiver operating characteristics (ROC) of b^-, with optimum threshold (γ_{b^-opt}) and area under the ROC curve (AUC).

For comparison, we used AUREA, our automated unsupervised method for respiratory state classification, to analyze the same data and estimate the respiratory state at each time [6]. Samples classified as pause or movement artifact were assigned to the non-breathing category.

Automated classification was performed on all "gold" standard scored data sets using both the breathing detection statistics combination and AUREA. We computed the agreement of each method with the respiratory state manually determined by the expert scorer. This was measured on a sample-by-sample basis using Fleiss' kappa (κ) statistic for inter-rater reliability [8]. A value of $\kappa = 1$ indicates perfect agreement, while $\kappa = 0$ reflects the performance expected by chance. We computed the overall κ value, as well as the category specific agreement for each category: non-breathing, synchronous and asynchronous breathing. To quantify the improvement in classification performance, we computed the difference of agreement (κ_{dif}) by subtracting the κ values obtained with AUREA to those from the new method.

The performance of the methods was evaluated using cross-validation. To ensure that data from the same infant was not used to train and test the methods, the complete data set was split into two subsets: (1) testing, containing the data from one of the subjects; and (2) training, with the remaining data sets. The training subset was used to tune the new detectors (obtain γ_{b^+opt} and γ_{b^-opt}), and to determine the parameters of the classification rules for AUREA. These results were then used to classify the respiratory state in the

testing subset and evaluate the κ values. This was repeated iteratively until all subjects were part of the testing set. We computed the mean and standard deviation of the agreement values as weighted averages based on the total number of samples assigned by the expert to each category on each subject. The values of κ_{dif} from each iteration were used on a Wilcoxon signed-rank test [9] to assess for significant changes in agreement (the null hypothesis of this test corresponds to a mean $\kappa_{dif} = 0$).

IV. Results

A. Detection Performance

We computed the breathing detection statistics b^+ and b^- on all data sets with parameters $N_b = 251$, $N_{avg} = 21$. The SYB statistic b^+ was used to classify between H_1: SYB, and H_0: all other states. Fig. 3(a) shows the corresponding PDFs, and Fig. 3(b) the ROC curve. This statistic performed extremely well with an $AUC = 0.95$ and optimum threshold $\gamma_{b^+opt} = 0.1$. We also used b^- to distinguish between H_1: ASB, and H_0: all other states. Fig. 4(a) shows the PDFs for H_1 and H_0 and Fig. 4(b) shows the ROC curve which had an $AUC = 0.94$ and optimum threshold $\gamma_{b^-opt} = 0.08$.

B. Evaluation of Agreement

The agreement of the two automated methods with the manual classification from the expert is shown in Table I. The largest improvement was observed in ASB ($\kappa_{dif} = 0.13 \pm 0.04$, $p < 0.001$). Agreement on SYB also had a slight improvement ($\kappa_{dif} = 0.06 \pm 0.06$, $p = 0.03$), together with the overall agreement ($\kappa_{dif} = 0.04 \pm 0.06$, $p = 0.12$). Agreement on non-breathing remained mostly equal ($\kappa_{dif} = 0.00 \pm 0.08$, $p=0.64$).

TABLE I. Cross-Validation Results of Agreement (κ) between Automated Methods and Expert Scorer

Method	SYB[a]	ASB[a]	NB[a]	Overall
AUREA	0.67±0.06	0.49±0.08	0.68±0.07	0.65±0.05
New detectors	0.73±0.06	0.62±0.10	0.68±0.07	0.69±0.05
Difference[b]	0.06±0.06[c]	0.13±0.04[d]	0.00±0.08	0.04±0.06

a. SYB = Synchronous Breathing; ASB = Asynchronous Breathing; NB = Non-breathing.
b. Positive difference indicates improvement of the new detectors.
c. P-value < 0.05. d. P-value<0.001.

V. Discussion

Existing rules for manual scoring of respiratory data [3] focus on the identification of respiratory pauses, and do not define criteria to identify breathing segments. This limits the study of the complete breathing patterns in recordings where characteristics from all respiratory states (not only pauses) are of interest. This paper presents a set of detection statistics that identify synchronous and asynchronous breathing segments in uncalibrated RIP signals. We successfully applied them to real infant data, and found they performed very well with values of AUC around 0.95.

A simple combination of the statistics was used to assign the respiratory state to one of three categories: non-breathing, synchronous and asynchronous breathing. The results agreed very well with manual classification by an expert scorer. We compared this result with the agreement of an existing method (AUREA), which performs breathing classification by elimination, without explicit breathing detection statistics. The breathing statistic combination compared favorably to AUREA, having higher agreement with the expert in the SYB and ASB categories, as well as overall κ value. The agreement for ASB was substantially higher (≈ 27% increase with respect to AUREA). Note that the moderate improvement in classification of SYB actually translates into many more correct assignments since this state corresponds to > 65% of the complete recordings.

In the clinical environment it is common to encounter problems during data acquisition, leading to bad signal segments. The basic approach to mitigate this problem is to manually inspect the data and/or to design an attended study, where the person supervising the acquisition session logs and troubleshoots any problems. This limits the realization of large studies (e.g., multi-institutional) of respiratory activity involving manual analysis of long recordings. Alternatively, an automated method of analysis capable of discriminating bad signal segments could be used to enable such studies. With our new breathing detectors, bad signal segments are classified as non-breathing. In conjunction with detectors for other respiratory states (such as pause and movement artifact [6]), this can yield a complete method for respiratory state classification able to discriminate bad signal data. Future work will exploit this possibility.

Given the general use of the presented methodology, further studies will aim to analyze a broader data set, including signals from adults (healthy and patients), as well as to validate the methods against an enhanced "gold" standard, generated from manual scores provided by several experts.

References

[1] G. M. Nixon and R. T. Brouillette, "Diagnostic techniques for obstructive sleep apnoea: is polysomnography necessary?," Paediatr. Respir. Rev., vol. 3, pp. 18-24, 2002.

[2] A. Aoude, R. E. Kearney, K. A. Brown, H. Galiana, and C. A. Robles-Rubio, "Automated Off-Line Respiratory Event Detection for the Study of Postoperative Apnea in Infants," IEEE Trans. Biomed. Eng., vol. 58, pp. 1724-1733, 2011.

[3] Anonymous, The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. Welchester, Illinois: American Academy of Sleep Medicine, 2007.

[4] N. A. Collop, "Scoring variability between polysomnography technologists in different sleep laboratories," Sleep Med., vol. 3, pp. 43-47, 2002.

[5] A. D. Groote, J. Groswasser, H. Bersini, P. Mathys, and A. Kahn, "Detection of obstructive apnea events in sleeping infants from thoracoabdominal movements," J. Sleep Res., vol. 11, pp. 161-168, 2002.

[6] C. A. Robles-Rubio, K. A. Brown, and R. E. Kearney, "Automated Unsupervised Respiratory Event Analysis," in Conf. Proc. 33rd IEEE Eng. Med. Biol. Soc., Boston, USA, 2011, pp. 3201-3204.

[7] A. L. Motto, H. L. Galiana, K. A. Brown, and R. E. Kearney, "Automated estimation of the phase between thoracic and abdominal movement signals," IEEE Trans. Biomed. Eng., vol. 52, pp. 614-621, 2005.

[8] J. L. Fleiss, "Measuring nominal scale agreement among many raters," Psychol. Bulletin, vol. 76, pp. 378-382, 1971.

[9] F. Wilcoxon, "Individual Comparisons by Ranking Methods," Biometrics Bulletin, vol. 1, pp. 80-83, 1945.