Postprint of: Ciołek M., Niedźwiecki M., Sieklicki S., Drozdowski J., Siebert J., Automated Detection of Sleep Apnea and Hypopnea Events Based on Robust Airflow Envelope Tracking in the Presence of Breathing Artifacts, IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, Vol. 19, Iss. 2 (2015), pp. 418-429, DOI: 10.1109/JBHI.2014.2325997

Automated Detection of Sleep Apnea and Hypopnea Events Based on Robust Airflow Envelope Tracking in the Presence of Breathing Artifacts

Marcin Ciołek, Maciej Niedźwiecki, *Senior Member, IEEE*, Stefan Sieklicki, Jacek Drozdowski and Janusz Siebert

***Abstract*—The paper presents a new approach to detection of apnea/hypopnea events, in the presence of artifacts and breathing irregularities, from a single channel airflow record. The proposed algorithm, based on a robust envelope detector, identifies segments of signal affected by a high amplitude modulation corresponding to apnea/hypopnea events. It is shown that a robust airflow envelope - free of breathing artifacts - improves effectiveness of the diagnostic process and allows one to localize the beginning and the end of each episode more accurately. The performance of the proposed approach, evaluated on 15 overnight polysomnographic recordings, was assessed using diagnostic measures such as accuracy, sensitivity specificity and Cohen's coefficient of agreement; the achieved levels were equal to 96%, 91%, 96% and 0.85, respectively. The results suggest that the algorithm may be implemented successfully in portable monitoring devices, as well as in software-packages used in sleep laboratories for automated evaluation of sleep apnea/hypopnea syndrome.**

***Index Terms*—Sleep apnea and hypopnea, breathing artifacts, median filters, envelope detector**

I. Introduction

SLEEP apnea/hypopnea syndrome (SAHS) is a sleep-breathing disorder characterized by repetitive episodes of complete obstruction (sleep apnea event) or partial obstruction (hypopnea event) of the upper airway, often resulting in a blood oxygen desaturation and arousals leading to sleep fragmentation. The usual daytime manifestation is excessive sleepiness, fatigue, and poor concentration, which can escalate to traffic accidents, depression, and memory loss. The major risk factors for the disorder include obesity, male gender, and age [1]. Untreated SAHS may lead to ischemic heart disease, cardiovascular disfunction, and stroke [2]. SAHS is a noticeable problem of social and health life, affecting 3% of children [3], 2% of female adults and 4% of male adults worldwide [4]. In fact, still up to 80% of cases of moderate or severe SAHS have gone undiagnosed despite adequate access to health care [5], [6].

Currently, in sleep laboratories, there are carried out overnight polysomnographic studies (PSG) aimed at early detection and assessment of the severity of SAHS in patients. PSG study is considered as the "gold standard" method for SAHS diagnosis [7]. It involves recording and studying simultaneously many signals such as electrocardiogram (ECG), nasal airflow (NAF) and blood oxygen saturation (SaO2). To reach the final conclusion, the recorded signals are analysed by a physician experienced in the field of pulmonology. The final diagnosis is based on calculation of the apnea/hypopnea index (AHI) which reflects the number of sleep apnea/hypopnea (SAH) events per hour of sleep. It is assumed that accurate and reliable identification of SAH events is critical for case identification and for quantifying disease severity classified as: mild when $5 \leq$ AHI < 15, moderate when $15 \leq$AHI < 30, and severe when AHI ≥ 30 (events per hour of sleep) [7], [8]. Currently the clinical routine is based on manual correction of the results obtained by automated analysis, which is an extremely tedious and time consuming task [8]. Unfortunately, the drawbacks of the full PSG study, such as high cost, a long list of patients waiting to be tested, and a feeling of discomfort due to a large number of sensors placed on the patient's body during the overnight test, suggest the need for developing alternative methods of diagnosis, based on information from selected channels of PSG, which could be implemented in portable monitoring devices for evaluation of SAHS [1].

Numerous methods exist, based on the evaluation of various signals originating from PSG, for detection of SAH events. This includes methods based on analysis of an ECG signal [9], [10], [11], [12], a SaO2 signal [13], [14], or a combination of the two [15]. Both signals mentioned above provide only supportive evidence of SAH events and do not allow one to localize precisely their beginning and end. It often happens that the primary evidence (significant reduction in NAF signal) is not observed. In the case of a SaO2 signal, the supportive evidence, in the form of a blood oxygen desaturation, is delayed in time in relation to the moments of occurrence of SAH events. Additionally, it is highly dependent on many factors, such as calibration of the measuring device, physiological conditions of the patient, presence of artifacts, etc. [8]. In the recent years several studies have focused on detection of SAH events based exclusively on the analysis of the airflow signal [16], [17], [18], [19], [20]. Almost all of these studies are

M. Ciołek, M. Niedźwiecki and S. Sieklicki are with the Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology, Narutowicza 11/12, Gdańsk, Poland (e-mails: marcin.ciolek@pg.gda.pl, maciekn@eti.pg.gda.pl, sieklick@eti.pg.gda.pl).

J. Drozdowski is with the Department of Pneumology and Allergology, Medical University of Gdańsk, Gdańsk, Poland (e-mail: ??????@???.???).

J. Siebert is with the Department of Family Medicine, Medical University of Gdańsk, Gdańsk, Poland (e-mail: kmr@amg.gda.pl).

based on a “black box” approach and use such techniques as artificial neural networks [16], spectral analysis [17], feature selection [18] or support vector machines [19], [20].

The main contribution of this paper is demonstration that SAH events can be effectively identified, in the presence of artifacts and breathing irregularities, based on analysis of a robust airflow envelope.

II. Breathing Artifacts

The standard morphological criteria, given by the American Academy of Sleep Medicine (AASM) [1], describe SAH events as a significant reduction of airflow amplitude lasting at least 10 seconds. The reduction of airflow amplitude is observed in relation to the level of breathing amplitude preceding and succeeding the respiratory event, further called baseline value. Two thresholds of airflow reduction, 50% and 90%, were accepted to represent partial and complete obstruction of the upper airway, respectively. The baseline was not defined clearly in an analytical way. Often pulmonology specialists adopt for the baseline the value the local maximum peak, or the average value of peaks from the last 100 seconds [1]. In the presence of abnormally large peaks, further called artifacts, none of these recommendations is appropriate for baseline value tracking. Irregular breathing artifacts are usually associated either with the patient’s motion during sleep - see Fig. 1A (rapid body movements, changing of sleep position), or with a sudden opening of the upper airway succeeding a sleep apnea event - see Fig. 1B. Such a rtifacts in airflow measurements can lead to incorrect identification of SAH events by automated sleep scoring methods, which in turn may result in incorrect diagnosis of the SAHS syndrome. For this reason, the physician localizes, based on visual inspection, signal segments corrupted by artifacts and manually marks them as the ones that should be ignored during automated sleep scoring. This is a very time-consuming process, and a subjective judgment is required to complete the job. Unsatisfactory quality of automated analysis based on NAF signal is often caused by problems with adaptive tracking of the baseline value, with respect to which SAH events are identified. Due to the presence of artifacts, the correct morphological description of the baseline value is not a trivial task [8], [21]. In [22] the authors propose to track the baseline value in adaptive way, based on an exponentially weighted average of past peaks. To eliminate impact of large positive and negative peaks, only peaks that remain within 40% of the current baseline value are considered. This approach seems to work well until a sudden change of breathing rhythm appears, leading to problems with fast updating of the baseline value. The approach proposed in [23] is based on airflow signal modeling. The model is used to reconstruct fragments corrupted by artifacts.

III. Envelope Detection

During a routine analysis physician can easily track the “true” signal envelope using visual inspection - even in the presence of artifacts. Such an envelope corresponds to a smooth curve (without ripples) that matches, in a way that is robust to breathing artifacts, the main peaks of the waveform,

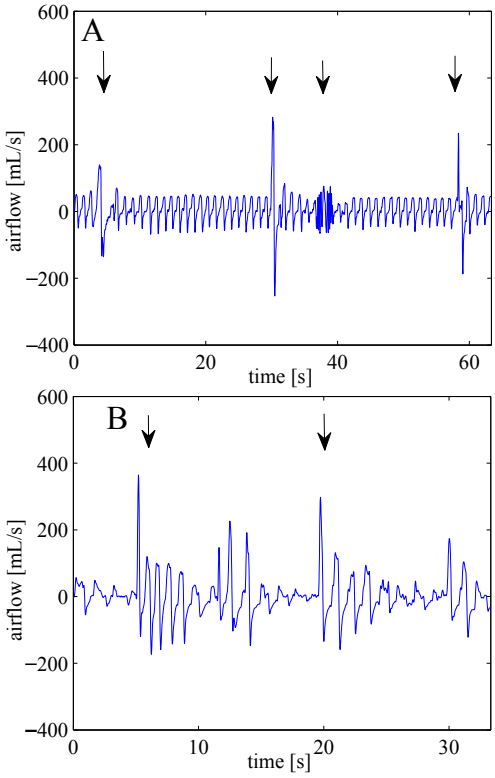

Fig. 1: Breathing artifacts : A - rapid body movements and changing of sleep position, B - sudden opening of the upper airway succeeding a sleep apnea/hypopnea event. Artifacts are marked with arrows.

and follows closely sudden variations in signal amplitude. If a partial or complete reduction of airflow takes place, then SAH events take the form of characteristic “valleys” visible in the signal envelope. A physician identifies and classifies these valleys as hypopnea or sleep apnea events, based on the standard morphological criteria and his/her own experience. In the proposed approach we try to reproduce such a procedure.

Envelope detection has numerous applications in the field of signal processing and communications [24], one of which is demodulation of amplitude-modulated (AM) signals governed by

$$s(t) = A[1 + \beta m(t)] \cos \omega_{\mathrm{c}} t \tag{1}$$

where $t = \ldots, -1, 0, 1, \ldots$ denotes normalized (dimensionless) discrete time, $m(t) \geq 0$ denotes the baseband message, ω_{c} denotes carrier (angular) frequency, A denotes the carrier amplitude and $\beta > 0$ is the so-called amplitude sensitivity of the modulator. When $m(t)$ is a lowpass signal with bandwidth W much smaller than the carrier frequency ω_{c}, the amplitude envelope of the signal $s(t)$ is defined as

$$e(t) = A[1 + \beta m(t)] \geq 0. \tag{2}$$

Envelope can be “extracted” from the AM signal using devices known as envelope detectors. The two most frequently used envelope detectors that will be briefly described below are those based on the square-law (full rectification) principle, and on the Hilbert transform, respectively. Due to irregularities in

the breathing rhythm, the airflow signal only approximately fits the AM model (1) which adversely affects performance of the classical envelope detectors. The situation becomes even more complicated in the presence of breathing artifacts. We will show that both problems indicated above can be taken care of if the classical detection schemes are suitably modified.

A. Square-law envelope detector

The flowchart of the square-law detector [24] is shown in Fig. 2. When $s(t)$ is an AM signal governed by (1), the scaled

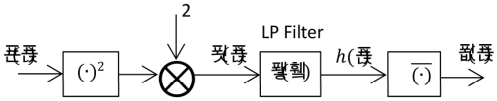

Fig. 2: The square-law envelope detector.

output of the squaring device can be written down as a sum of two components

$$f(t) = A^2[1+\beta m(t)]^2 + A^2[1+\beta m(t)]^2 \cos 2\omega_c t. \quad (3)$$

and half is shifted down towards DC. The first term on the right-hand side of (3) is a lowpass signal with the cutoff frequency $2W$, and the second one is a bandpass signal with spectrum confined to the frequency bands $(-2\omega_c - 2W, 2\text{-}wc+2W)$ and $(2\omega_c-2W, 2\omega_c+2W)$, centered around $\pm 2\omega_c$. Hence, when the condition $\omega_c > 2W$ is met, the lowpass component of $f(t)$ can be extracted using a lowpass FIR filter $L(\omega)$ with cutoff frequency $2W$

$$h(t) = L[f(t)] = \sum_{i=-k}^{k} l_i f(t-i) \cong A^2[1+\beta m(t)]^2 \quad (4)$$

where $l_i = l_{-i}, i = -k, \ldots, k$, denote impulse response coefficients of the filter. The estimated value of the envelope can be obtained from

$$\widehat{e}(t) = \sqrt{h(t)} \quad (5)$$

B. Envelope detector based on the Hilbert transform

The second classical method of envelope detection, based on the Hilbert transform [24], is depicted in Fig. 3. The Hilbert transform of an analog real-valued finite energy signal $x(t_c), -\infty < t_c < \infty$, is defined as [24]

$$s_H(t_c) = \mathcal{H}[s(t_c)] = \int_{-\infty}^{+\infty} \frac{s(\tau)d\tau}{\pi(t_c-\tau)} \quad (6)$$

where t_c denotes the continuous-time variable. Envelope detection involves creation of the complex-valued analytic signal, defined as

$$y(t_c) = s(t_c) + j s_H(t_c). \quad (7)$$

The analytic counterpart of a discrete-time signal $s(t)$ can be evaluated either directly - using the FFT-based frequency-domain approach [25], or indirectly - by computing an "analytic" signal $s_H(t)$ (using the discrete-time FIR approximation of the Hilbert transform) and combining it with $s(t)$:

$$y(t) = s(t) + j s_H(t). \quad (8)$$

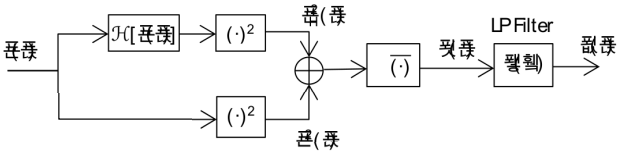

Fig. 3: Envelope detector based on the Hilbert transform.

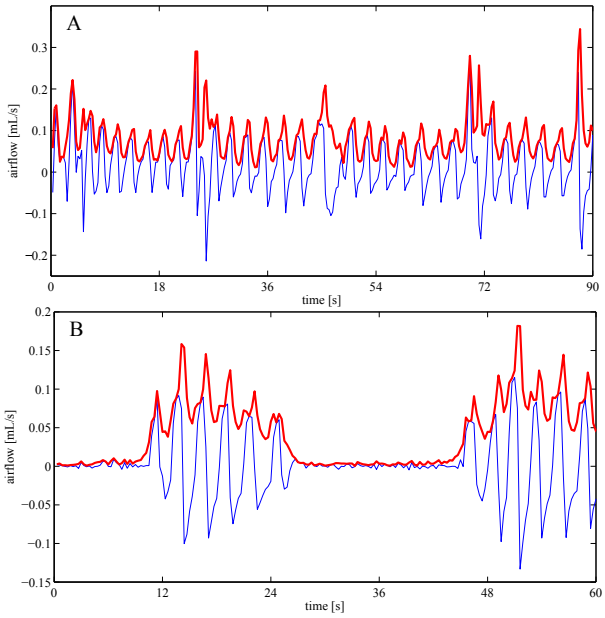

Fig. 4: The output of the Hilbert-transform-based envelope detector *prior* to lowpass filtering (thick line): A) normal breathing in the presence of incidental artifacts, B) sleep apnea patterns. Note the presence of high-frequency fluctuations called ripples. Thin lines show the airflow (input) signal.

Since Hilbert transform shifts the phase of all sinusoidal components by $-\pi/2$, for the "ideal" AM signal (1), one obtains $s_H(t) \cong A[1+\beta m(t)]\sin\omega_c t$ which means that the envelope of $s(t)$ can be obtained by evaluating the magnitude of the analytic signal

$$f(t) = |y(t)| = \sqrt{s^2(t) + s_H^2(t)} \cong A[1+\beta m(t)]. \quad (9)$$

For AM-like signals, such as speech signals or biomedical signals, the envelope extracted in this way suffers from high-frequency fluctuations, called ripples [27] - see Fig. 4. Ripples can be removed by passing the signal $f(t)$ through a lowpass filter, leading to

$$\widehat{e}(t) = L[f(t)]. \quad (10)$$

C. Modification in the envelope detection procedures

The estimation of the airflow envelope obtained using approach based on the square-law or on the Hilbert transform suffers from envelope distortions caused by artifacts which are only partially eliminated at the stage of lowpass filtering. The envelope distortions of short duration and of relatively high amplitude may seriously affect the estimated baseline

values by setting them at too high levels. This may lead to a large number of false-positive decisions, some of which cause erroneous distinction between hypopnea of sleep apnea events. The second problem with analysis based on the classical envelope detection results, is related to the filter-induced time-shift effects, at the beginning and at the end of each apnea episode, namely, in an i increased number of true-negative decisions. Since the sleep apnea event should last at least 10 seconds, wrong localization of its endpoints may result in an erroneous event classification.

To eliminate both drawbacks mentioned above, a cascade made up of a standard median (SM) filter and a recursive median (RM) filter is used instead of the linear lowpass FIR filter $L(\omega)$ in the two envelope detection methods depicted in Figs. 2 and 3.

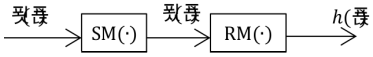

Fig. 5: A cascade made up of standard and recursive median filters.

Median filtering is a popular method of noise removal in applications involving signal and image processing. This nonlinear technique has proven to be a good alternative to linear filtering as it can effectively suppress impulsive noise while preserving the edge information [28].

The output of the SM filter $g(t)$ is the median value of the input data inside the window centered at the point t, and is given by

$$g(t) = \text{med}\{f(t-m), \ldots, f(t), \ldots, f(t+m)\} \qquad (11)$$

where $M = 2m+1$ denotes the window size and $\text{med}\{\cdot\}$ denotes the central value of the ordered sequence of samples. To effectively reduce the influence of artifacts on the airflow envelope, while preserving the sharp envelope "edges" at the beginning and at the end of each apnea episode, the SM window size should be properly selected. If the window size is too small, not all artifacts are suppressed. If the window size is too large, the blurring effect can be observed, similarly as in the case of image processing applications. Therefore the window size should be at least two times larger than the length of the segments affected by artifacts, but smaller than the distance between two neighboring SAH events. Based on the observation that artifacts are usually confined to one breathing cycle, whereas the breathing frequency changes from 0.18 Hz to 0.4 Hz, we suggest that the SM window should cover 12 seconds of the airflow signal.

The RM filter, used to process the median-prefiltred signal samples $g(t)$, is given by

$$h(t) = \text{med}\{h(t-n), \ldots, h(t-1), g(t), \ldots, g(t+n)\} \quad (12)$$

where $N = 2n+1$ denotes the window size. The RM filter is more sensitive to window size than the SM filter. If the window is too wide, it can excessively smooth out the signal leading to deformation of the envelope. It is proposed to set N approximately to $M/2$. According to our experiments, the proposed cascade of two median filters yields better results than those obtained when only one of the filters is applied. The proposed modification in the airflow envelope detection procedures allows one to obtain robust envelopes based on the square-law or on the Hilbert transform, further denoted as RESL and REHT, respectively.

Figs. 6-9 illustrate robustness of the proposed modified envelope detectors in two practically important cases.

Figs. 6 and 7 demonstrate insensitivity of RESL and REHT envelopes to breathing artifacts – unlike the classical ESL/EHT envelopes, which are affected by airflow outliers, the RESL/REHT envelopes are robust to short-lived breathing artifacts. This allows one to keep the baseline (which is set to the local envelope maximum) at a level corresponding to regular, i.e., undisturbed breathing.

Figs. 8 and 9 demonstrate another advantage of nonlinear filtering – preservation of sharp envelope "edges". When linear lowpass filter is used in lieu of the proposed cascade of nonlinear filters, the corresponding envelopes slowly decay/rise at the beginning/end of each reduced-airflow episode. Since the length of such an episode is an important diagnostic factor (see Section 4), its underestimation can result in overlooking or misclasification of SAH events. Median filters do not introduce time shifts mentioned above. Additionally, the recursive median filter is very efficient in smoothing out (without blurring envelope edges) some local signal fluctuations that can be observed at the output of a standard median filter [29]. As a result, the envelope "valleys" corresponding to SAH events usually have only one local minimum. This very much simplifies SAH identification as SAH episodes can be easily localized between the two successive local maxima.

Remark 1: Note that median windows centered at instants t and $t+1$ overlap. Therefore, since evaluation of $g(t)$ given by (11) has already sorted most of the samples that are required for evaluation of $g(t+1)$, the computation can be made much more efficient. Using the indexable skip list technique [30] the computational complexity of a median filter can be reduced from $O(M^2)$ comparisons to only $O(\log M)$ comparisons per time update [31]. The same technique can be used for realization of a fast recursive median filter.

Remark 2: Similarly as in the case of classical envelope detectors, the computational load of the proposed methods can be further reduced by downsampling the signal $f(t)$ prior to nonlinear filtration. To avoid aliasing effects, prior to downsampling the signal $f(t)$ should be passed through a linear lowpass filter with appropriately chosen cutoff frequency.

IV. Identification of SAH events

Once the airflow envelope has been obtained, identification of SAH events is easy: they can be localized (if present) between the succeeding local maxima of the robust envelope. Each time a new local maximum appears, the baseline value is updated and set to the value of this local maximum. Consider a sequence of samples $\{\widehat{e}(t_1), \ldots, \widehat{e}(t_n)\}$ corresponding to a segment of robust envelope, where $\widehat{e}(t_1)$ and $\widehat{e}(t_n)$ represent two subsequent local maxima. The baseline value is set to $\widehat{e}(t_1)$ and two thresholds are computed, corresponding to 50%

TABLE I: Quantitative comparison of scores provided by an expert with the results yielded by two variants of the proposed method: Approach I based on the square-law, and Approach II based on the Hilbert transform. The scores correspond to the number of detected apnea/hypopnea events (SAH), only apnea events (Apnea) and only hypopnea events (Hypopnea). The AHI index reflects the number of sleep apnea/hypopnea events per one hour of sleep. The number of artifacts present in airflow recordings, based on subjective scoring of abnormally large peaks, ranged from 13 to 210 per recording.

		Expert				Approach I				Approach II			
Patient	Artifacts	SAH	Apnea	Hypopnea	AHI	SAH	Apnea	Hypopnea	AHI	SAH	Apnea	Hypopnea	AHI
1	20	33	5	28	4.41	36	3	33	4.81	34	2	32	4.54
2	70	36	2	34	4.81	45	2	43	6.01	49	2	47	6.55
3	13	46	13	33	6.15	45	9	36	6.01	53	11	42	7.08
4	70	50	5	45	6.68	53	1	52	7.08	58	1	57	7.75
5	78	106	17	89	14.16	127	8	119	16.97	130	8	122	17.37
6	17	111	28	83	14.83	107	17	90	14.30	113	16	97	15.10
7	59	126	75	51	16.84	136	67	69	18.17	142	70	72	18.98
8	36	167	126	41	22.32	147	100	47	19.64	147	100	47	19.64
9	76	169	19	150	22.58	168	13	155	22.45	178	15	163	23.79
10	64	216	133	83	28.86	222	23	199	29.67	201	20	181	26.86
11	55	217	170	47	29.00	229	167	62	30.60	231	165	66	30.87
12	210	241	74	167	32.20	247	22	225	33.01	241	27	214	32.20
13	69	290	196	94	38.75	295	179	116	39.42	295	184	111	39.42
14	206	311	247	64	41.56	312	197	115	41.69	330	203	127	44.10
15	91	315	177	138	42.09	314	154	160	41.96	301	154	147	40.22
overall	1134	2434	1287	1147		2483	962	1521		2503	978	1525	

TABLE II: Event-by-event analysis. Comparison of scores obtained for two variants of the proposed method: Approach I based on the square-law, and Approach II based on the Hilbert transform. The scores that are higher than or equal to those yielded by the competing approach are shown in boldface. Equal scores are marked with asterisks.

	Approach I					Approach II				
Patient	SAH det. (%)	Apnea det. (%)	Hypopnea det. (%)	Misclass.	False det.	SAH det. (%)	Apnea det. (%)	Hypopnea det. (%)	Misclass.	False det.
1	**84.85**	**40.00**⋆	**82.14**	3	**8**	75.76	**40.00**⋆	75.00	**2**	9
2	**97.22**⋆	**50.00**⋆	**94.12**⋆	**2**⋆	**10**	**97.22**⋆	**50.00**⋆	**94.12**⋆	**2**⋆	14
3	**93.48**	69.23	**93.94**	**3**⋆	**2**	91.30	**76.92**	87.88	**3**⋆	11
4	**88.00**	**20.00**⋆	**86.67**	**4**⋆	**9**	78.00	**20.00**⋆	75.56	**4**⋆	19
5	**93.40**	**23.53**⋆	**88.76**	**16**⋆	**27**	89.62	**23.53**⋆	84.27	**16**⋆	34
6	**89.19**	**57.14**	**86.75**	**11**	**8**	88.29	53.57	85.54	12	16
7	**98.41**	89.33	**98.04**	**7**	**13**	96.83	**90.67**	90.20	8	21
8	**91.62**	**87.30**⋆	**95.12**	**4**⋆	**4**	90.42	**87.30**⋆	90.24	**4**⋆	6
9	**96.45**	68.42	**96.00**	6	**5**	93.49	**78.95**	92.67	**4**	20
10	**91.67**	**17.29**	**90.36**	**100**	24	82.87	15.04	63.86	106	**22**
11	**99.54**	**95.29**	**93.62**	**10**⋆	**12**	97.70	94.71	87.23	**10**⋆	18
12	**95.44**	28.38	**93.41**	53	**17**	89.63	**33.78**	85.03	**49**	25
13	**93.79**	**88.78**⋆	**84.04**	**19**	**25**	91.72	**88.78**⋆	75.53	21	32
14	95.82	78.54	**90.63**	**46**⋆	**15**	**96.14**	**80.57**	84.38	**46**⋆	32
15	**95.56**	**87.01**	**92.75**	**19**	**16**	90.79	84.18	81.88	24	18
overall	**93.63**	60.02	**91.09**	**20.20**	**13.00**	89.99	**61.20**	83.56	20.73	19.80

and 10% of the baseline value for hypopnea and apnea events, respectively. The following decisions are made

$$d_s(t_i) = \begin{cases} 1 & \text{if } \widehat{e}(t_i) \leq \frac{1}{2}\,\widehat{e}(t_1) \\ 0 & \text{otherwise} \end{cases}$$

$$d_b(t_i) = \begin{cases} 1 & \text{if } \widehat{e}(t_i) \leq \frac{1}{10}\,\widehat{e}(t_1) \\ 0 & \text{otherwise} \end{cases}$$

where $d_s(t)$ and $d_b(t)$ denote sequences of binary values indicating which samples in the analysed segment can be classified as hypopneic/apneic activity. If the segment includes less than 10 seconds of continuous hypopneic/apneic activity, it is classified as normal breathing. Otherwise hypopnea or apnea is detected. When both types overlap, only the sleep apnea episode is scored – see Fig. 10.

V. Experimental Results

The polysomnograms of 15 sleep apnea patients, 9 males and 6 females [age: 53 ± 7 years (mean$\pm$standard deviation), duration of each study: 449 minutes] were used to validate the proposed method. The analyzed sleep studies were drawn from the database of the Medical University of Gdańsk, Gdańsk, Poland. In all studies the airflow signal was recorded with the sampling frequency of 20 Hz. Respiratory events were detected based on analysis of the airflow signal and scored using the criteria proposed by AASM [1]. Hypopnea was defined as an over 50% reduction in airflow from the baseline value, lasting for more than 10 seconds, and associated with a 3% desaturation or an arousal. Sleep apnea was defined as the absence of airflow for more than 10 seconds. The clinical routine was based on manual correction of the results

TABLE III: Epoch-by-epoch analysis. Comparison of scores obtained for two variants of the proposed method: Approach I based on the square-law, and Approach II based on the Hilbert transform. Four quality measures were used to asses the performance of the proposed method: accuracy (Acc), sensitivity (Sens), specificity (Spec) and Cohen's coefficient of agreement (κ). The scores that are better than or equal to those yielded by the competing approach are shown in boldface. Equal scores are marked with asterisks. Evaluation was based on analysis of 30-second airflow epochs classified as positive, if at least 5 seconds of the epoch was affected by hypopneic/apneic activity. For each examined patient the number of epochs was equal to 899. TP: true positive, FN: false negative, TN: true negative, FP: false positive.

Patient	Approach I								Approach II							
	TP	FN	TN	FP	Acc (%)	Sens (%)	Spec (%)	κ	TP	FN	TN	FP	Acc (%)	Sens (%)	Spec (%)	κ
1	33	8	846	12	**97.78**	**80.49**	**98.60**	**0.76**	31	10	843	15	97.22	75.61	98.25	0,70
2	49	3	828	19	**97.55**	**94.23***	**97.76**	**0.80**	49	3	820	27	96.66	**94.23***	96.81	0,75
3	51	5	836	7	**98.67**	**91.07**	**99.17**	**0.89**	48	8	826	17	97.22	85.71	97.98	0.78
4	50	7	827	15	**97.55**	**87.72**	**98.22**	**0.81**	45	12	817	25	95.88	78.95	97.03	0.69
5	136	19	708	36	**93.88**	**87.74**	**95.16**	**0.79**	131	24	697	47	92.10	84.52	93.68	0.74
6	127	23	735	14	**95.88**	**84.67***	**98.13**	**0.85**	127	23	723	26	94.55	**84.67***	96.53	0.81
7	175	9	676	39	**94.66**	**95.11**	**94.55**	**0.85**	173	11	669	46	93.66	94.02	93.57	0.82
8	239	19	630	11	**96.66**	**92.64**	**98.28**	**0.92**	236	22	628	13	96.11	91.47	97.97	0.90
9	181	17	693	8	**97.22**	**91.41**	**98.86**	**0.92**	178	20	677	24	95.11	89.90	96.58	0.86
10	196	17	657	29	**94.80**	**92.02**	**95.77**	**0.86**	180	33	656	30	92.99	84.51	95.63	0.81
11	325	18	536	20	**95.77**	**94.75**	**96.40**	**0.91**	320	23	528	28	94.33	93.29	94.96	0.88
12	296	20	548	35	**93.88**	**93.67**	**94.00**	**0.87**	281	35	537	46	90.99	88.92	92.11	0.80
13	347	28	489	35	**92.99**	**92.53**	**93.32**	**0.86**	340	35	485	39	91.77	90.67	92.56	0.83
14	311	12	555	21	**96.33**	**96.28**	**96.35**	**0.92**	306	17	538	38	93.88	94.74	93.40	0.87
15	428	38	393	40	**91.32**	**91.85**	**90.76**	**0.83**	409	57	390	43	88.88	87.77	90.07	0.78
overall	2944	243	9957	341	**95.67**	**91.08**	**96.36**	**0.85**	2854	333	9834	464	94.09	87.93	95.14	0.80

obtained by automated analysis performed by a commercial PSG software (RemLogic). The common mistakes of the automated analysis are: overlooked episodes, false detections, and misclassification between hypopnea and sleep apnea events. In all studies 2434 SAH events were detected: 1287 apneas and 1147 hypopneas. The apnea/hypopnea index (AHI) for the examined patients ranged from 4.41 to 42.09. The total number of artifacts present in airflow recordings – the result of subjective scoring of abnormally large peaks – was 1134, ranging from 13 to 210 per recording.

The 127-tap FIR filter approximating the Hilbert transform was designed using the Parks-McClellan algorithm. The analytic signal was computed by adding the appropriately time-shifted real signal to its imaginary counterpart generated by the Hilbert filter. To reduce computational complexity, prior to median filtering the signal $f(t)$ was passed through a lowpass FIR filter with a cutoff frequency of 1 Hz, and then downsampled by a factor of $d = 6$. After decimation, the SM window size was set to $M = 51$ and the RM window size was set to $N = 21$.

Table I compares the results of automated detection of SAH events with decisions made by an expert. Patients were ordered according to their apnea/hypopnea index (AHI). All decisions were divided into four categories: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Four quality measures were used to asses the performance of the detectors: accuracy, sensitivity, specificity and Cohen's coefficient of agreement (agreement beyond that expected by chance, usually referred to as kappa statistic) – the corresponding expressions are

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FN}+\text{TN}+\text{FP}} = \text{A}$$
$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP}+\text{FN}}$$
$$\text{Specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}}$$
$$\text{Agreement} = \frac{\text{A}-\text{B}}{1-\text{B}} = \kappa$$

where

$$\text{B} = \frac{(\text{TP}+\text{TN})(\text{TP}+\text{FN})+(\text{FP}+\text{TN})(\text{FN}+\text{FP})}{\text{TP}+\text{FN}+\text{TN}+\text{FP}}$$

Tables II and III show the comparison of two variants of the proposed method: Approach I based on the square-law, and Approach II based on the Hilbert transform.

Table II shows results of the event-by-event analysis. Note that while detection rates of SAH events reach high levels (average detection rate = 93.6%, minimum detection rate = 84.9%), classification of apnea and hypopnea episodes is less successful. In particular, for 4 patients (#4, #5, #10 and #12) the apnea detection rate is lower than 30%. This means that some further work is needed to develop tools capable of distinguishing both types of events in a more reliable way.

Table III shows results of the epoch-by-epoch examination. Evaluation was based on analysis of 30-second airflow epochs classified as positive, if at least 5 seconds of the epoch was affected by hypopneic/apneic activity. For each patient the number of epochs was equal to 899. The obtained results clearly indicate superiority of the square-law-based approach. Note that the Cohen's coefficient of agreement κ evaluated for the robust square-law detector takes pretty large values – for 12 patients it holds that $\kappa \in [0.81, 1]$, which corresponds to

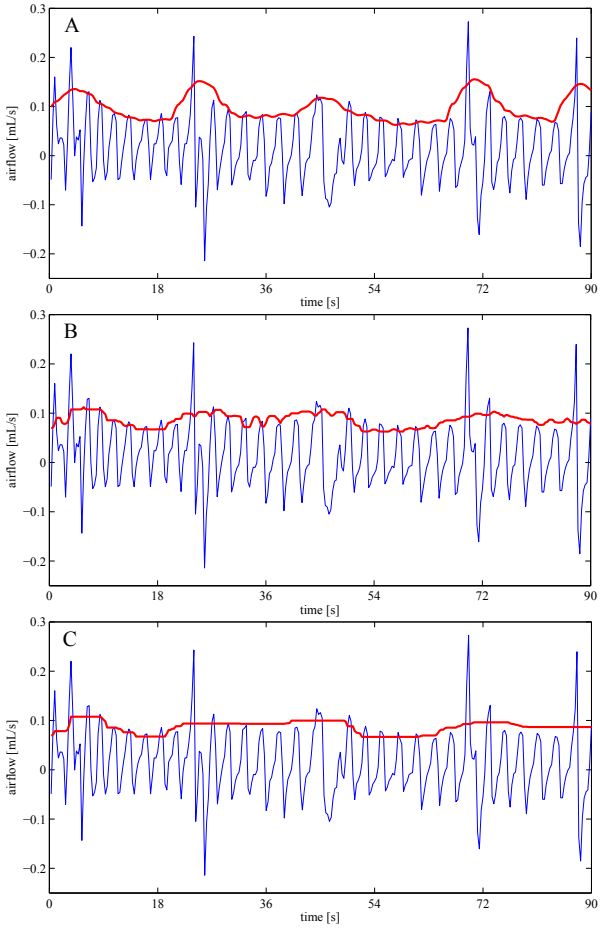

Fig. 6: Performance of the classical square-law envelope detector (A) and its robust version (C) in the presence of breathing artifacts. Middle figure (B) shows the intermediate detection results observed at the output of the standard median filter. Thick line – envelope, thin line – airflow signal.

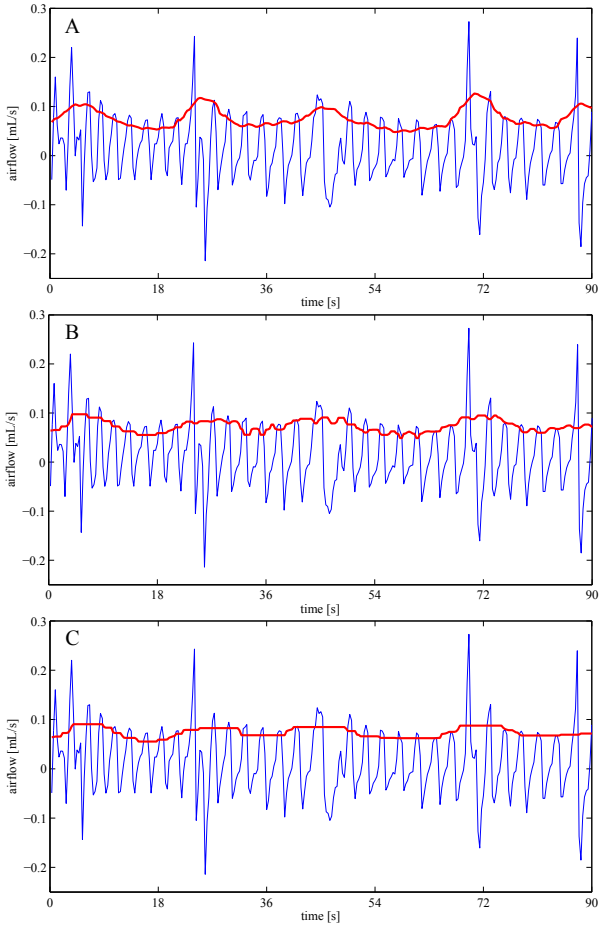

Fig. 7: Performance of the classical Hilbert transform envelope detector (A) and its robust version (C) in the presence of breathing artifacts. Middle figure (B) shows the intermediate detection results observed at the output of the standard median filter. Thick line – envelope, thin line – airflow signal.

the highest qualitative level of agreement strength, interpreted as “almost perfect agreement” [32] or “very good agreement” [33], and for the remaining 3 patients $\kappa \in [0.61, 0.80]$, which is regarded as “substantial agreement” [32] or “good agreement” [33].

Bland-Altman plots comparing AHI scores provided by an expert with those resulting from the automated analysis are shown in Fig. 11 (for Approach I) and Fig. 12 (for Approach II). This is a popular method to compare two scoring techniques. In this graphical method the differences between the two techniques are plotted against their averages. Horizontal lines are drawn at the mean difference, and at the 95% limits of agreement, which are defined as the mean difference ± 1.96 times the standard deviation of the differences. Bland-Altman plots allow one to investigate the existence of any systematic difference (fixed bias) between the scores, and to identify possible outliers. According to the plots shown in Figs. 11 and 12, the proposed methods are nearly unbiased – mean biases are equal to -0.4367 and -0.6153 event/hour of sleep for Approach I and Approach II, respectively. The differences remain close to the bias line, even for an increasing AHI index.

Finally, Table IV shows the results of comparison of 3 approaches to SAH events detection, based on different ways of baseline value tracking: the approach based on the exponentially weighted average of past peaks (EWAPP) [22], the approach adopted in a commercial PSG software (RemLogic), and the proposed approach based on RESL. In this experiment no differentiation between apnea and hypopnea event was made. Evaluation was based on analysis of 30-second airflow epochs classified as SAH events or normal breathing. The epoch was classified as SAH event if at least 5 seconds of the epoch was affected by apneic/hypopneic activity. The total number of epochs for 15 patients was 13485. The total number of events detected by the expert for 15 patients was 2434. The results of the test show clearly superiority of the proposed method.

Remark: When comparing results of automated detection of SAH events with those provided by an expert, one should remember that the latter ones do not necessarily present a 100% correct score. Reports on intrascorer and interscorer reliability for scoring respiratory events were summarized in [8]. Reliability increases when an expert is guided by an

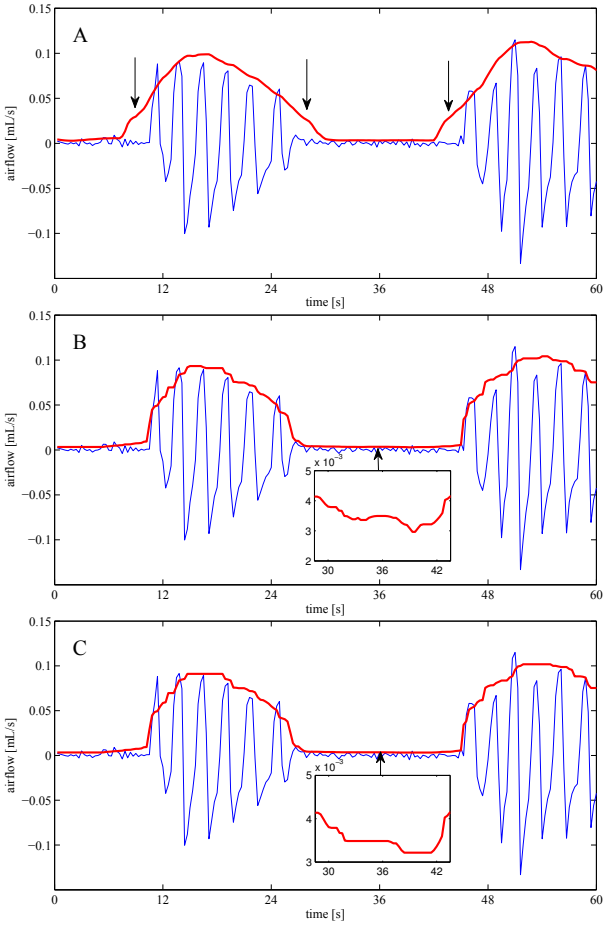

Fig. 8: Performance of the classical square-law envelope detector (A) and its robust version (C) in the presence of sleep apnea/hypopnea events. Middle figure (B) shows the intermediate detection results observed at the output of the standard median filter. Thick line – envelope, thin line – airflow signal.

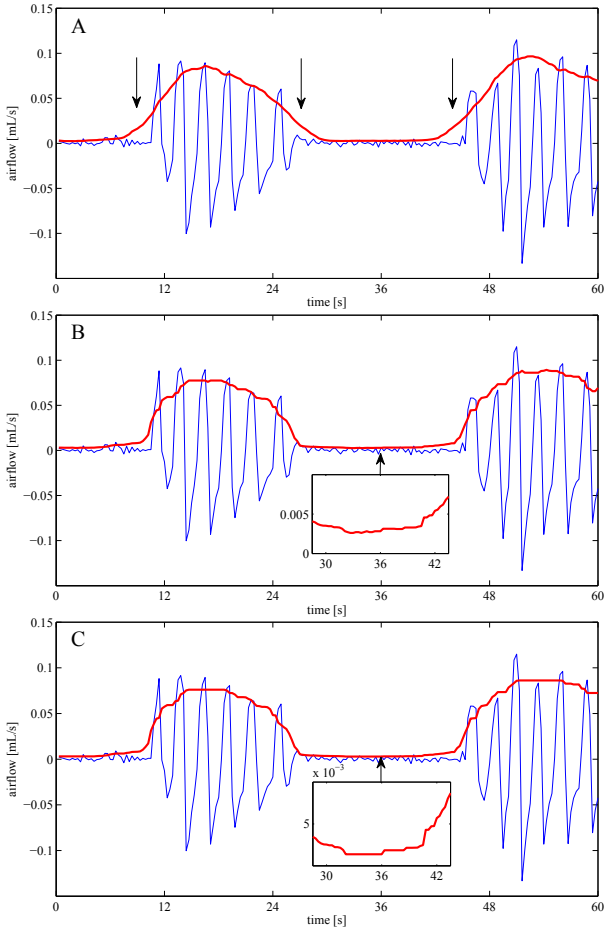

Fig. 9: Performance of the classical Hilbert transform envelope detector (A) and its robust version (C) in the presence of sleep apnea/hypopnea events. Middle figure (B) shows the intermediate detection results observed at the output of the standard median filter. Thick line – envelope, thin line – airflow signal.

TABLE IV: Results of comparison of 3 approaches to SAH events detection, based on different ways of baseline value tracking: the approach based on the exponentially weighted average of past peaks (EWAPP), the approach adopted in a commercial PSG software (RemLogic), and the proposed approach based on RESL. The best scores are shown in boldface.

Approach to SAH det.	Correct det.[%]	False det.[%]	Acc [%]	Sens [%]	Spec [%]	κ
EWAPP	81.53	14.63	90.83	81.24	94.10	0.65
RemLogic	67.40	7.13	90.13	67.70	**97.80**	0.67
RESL	**93.63**	**6.82**	**95.67**	**91.08**	96.36	**0.85**

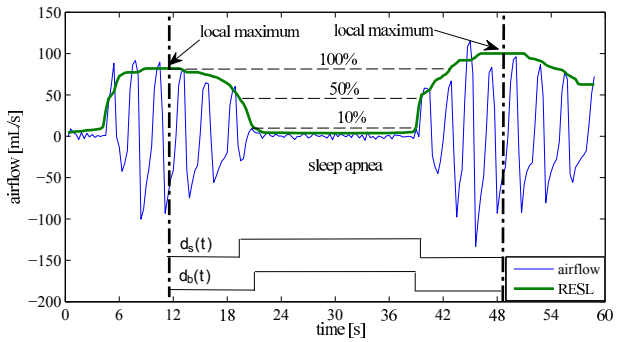

Fig. 10: Identification of SAH events based on the RESL envelope.

automatic detection tool [22]. One of important conclusions of this study is that the proposed methods seem to provide better guidance than the currently available ones.

VI. Conclusion

The widely used classical envelope detection methods are not robust to artifacts present in airflow measurements. The proposed approach is a simple modification of the existing schemes, obtained by replacing the linear lowpass output filter with a cascade of two nonlinear filters - a standard median filter and a recursive median filter. Unlike the methods described in the literature the proposed algorithms do not need

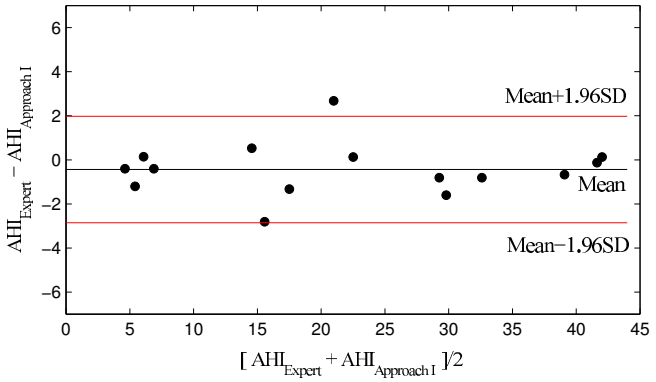

Fig. 11: Bland-Altman plot comparing AHI scores provided by an expert with those resulting from the automated analysis using Approach I (differences between the scores against their averages). Horizontal lines show the mean difference and the 95% limits of agreement (SD = standard deviation).

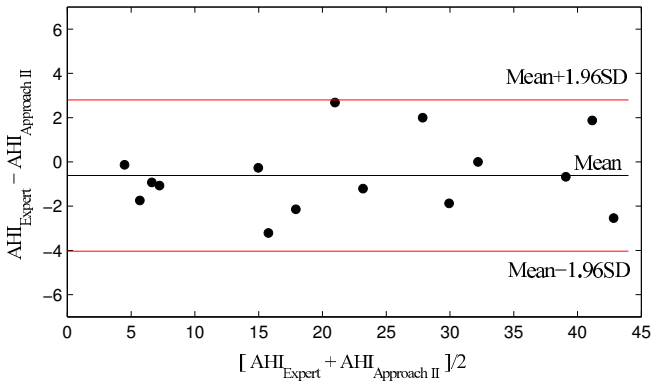

Fig. 12: Bland-Altman plot comparing AHI scores provided by an expert with those resulting from the automated analysis using Approach II (differences between the scores against their averages). Horizontal lines show the mean difference and the 95% limits of agreement (SD = standard deviation).

initial training or optimization.

REFERENCES

[1] American Academy of Sleep Medicine Task Force, "Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research," *Sleep*, vol. 22, pp. 667–689, 1999.

[2] W. Lee, S. Naqubadi, and M.H. Mokhlesi, "Epidemiology of obstructive sleep apnea: A population-based perspective," *Expert Rev. Respir. Med.*, vol. 2, pp. 349–364, 2008.

[3] American Thoracic Society, "Cardiorespiratory sleep studies in children. Establishment of normative data and polysomnographic predictors of morbidity," *American Journal of Respiratory and Critical Care Medicine*, vol. 160, pp. 1381–1387, 1999.

[4] T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, "The occurrence of sleep-disordered breathing among middle-aged adults," *New Engl. J. Med.*, vol. 328, pp. 1230–1235, 1993.

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

[6] V. Kapur, K.P. Strohl, S. Redline, C. Iber, G. O'Connor, and J. Nieto, "Underdiagnosis of sleep apnea syndrome in U.S. communities," *Sleep Breath*, vol. 6, pp. 49–54, 2002.

[7] Standards of Practice Committee of the American Sleep Disorders Association, "Practice parameters for the indications for polysomnography and related procedures," *Sleep*, vol. 20, pp. 406–422, 1997.

[8] S. Redline and et al., "The scoring of respiratory events in sleep: reliability and validity," *J Clin Sleep Med.*, vol. 3, pp. 169–200, 2007.

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

[10] P. de Chazal, C. Henegham, E. Sheridan, R. Reilly, P. Nolan, and M. O'Malley, "Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnea," *IEEE Trans. Biomed. Eng.*, vol. 50, pp.686–696, 2003.

[11] A.H. Khandoker, J. Gubbi, and M. Palaniswami, "Automated scoring of obstructive sleep apnea and hypopnea events using short-term electrocardiogram recordings," *IEEE Trans. on Inforormation Techology in Biomedical*, vol.13, pp. 1057–1067, 2009.

[12] M. Bsoul, H. Minn, and L. Tamil, "Apnea MedAssist: Real-time sleep apnea monitor using single-lead ECG," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, pp. 416–427, 2011.

[13] A. Burgos, A. Goñi, A. Illarramendi, and J. Bermudez, "Real-time detection of apneas on a PDA," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, pp. 995–1002, 2010.

[14] D. Álvarez, R. Hornero, J.V. Marcos, and F. Del Campo, "Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 57, pp. 2816–2824, 2010.

[15] B. Xie and H. Minn, "Real-time sleep apnea detection by classifier combination," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, pp. 469–477, 2012.

[16] P. Várady, T. Micsik, S. Benedek, and Z. Benyó, "A novel method for the detection of apnea and hypopnea events in respiration signals," *IEEE Trans. Biomed. Eng.*, vol. 49, pp. 936–942, 2002.

[17] H. Nakano, T. Tanigawa, T. Furukawa, and S. Nishima, "Automatic detection of sleep-disordered breathing from a single-channel airflow record," *Eur. Respir. J.*, vol. 29, pp. 728–736, 2007.

[18] S.I. Rathnayake, I.A. Wood, U.R. Abeyratne, and C. Hukins, "Nonlinear features for single-channel diagnosis of sleep-disordered breathing diseases," *IEEE Trans. Biomed. Eng.*, vol. 57, pp. 1973–1981, 2010.

[19] B. Koley, "Automated detection of apnea and hypopnea events," in *Proceedings Thrid International Conference on Emerging Applications of Information Technology*. IEEE, pp. 85–88, Kolkata, India, 2012.

[20] B. Koley, and D. Dey, "Adaptive Classification System for Real-Time Detection of Apnea and Hypopnea Events," in *Proceedings Point-of-Care Healthcare Technologies Conference.* IEEE, pp. 44–45, Bangalore, India, 2013.

[21] A. Otero, P. Félix, and M.R. Álvarez, "Algorithms for the analysis of polysomnographic recordings with customizable criteria," *Expert Systems with Applications*, vol. 38, pp. 10133–10146, 2011.

[22] A. Bartolo, B.D. Clymer, R.C. Burgess, and J.P. Turnbull, "Automatic on-line detection of apneas and hypopneas," Proc. 19th Annual EMBS Int. Conf., Chicago, USA, 1997.

[23] S.I. Rathnayake et al., "Modelling of polysomnographic respiratory measurements for artefact detection and signal restoration," *Physiol. Meas.*, vol. 29, pp. 999–1021, 2008.

[24] S. Haykin, *Communication Systems, 5th edition*, Wiley 2009.

[25] S.L. Marple, Jr., "Computing the discrete-time analytic signal via FFT," *IEEE Transactions on Signal Processing*, vol. 47, pp. 2600–2603, 1999.

[26] A.V. Oppenheim and R.W. Schafer, *Discrete-Time Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1989.

[27] A. Potamianos and P. Maragos, "A Comparison of the Energy Operator and the Hilbert Transform Approach to Signal and Speech Demodulation," *Signal Processing* vol. 37, pp. 95–120, 1994.

[28] T.S. Huang,*Two-dimensional digital signal processing II: Transform and median filters.*, Springer-Verlag, New York, 1981.

[29] I. Pitas and A.N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications.*, Springer-Verlag, New York, 1990.

[30] W. Pough, "Skip lists: a probabilistic alternative to balancedtrees," *Communications of the ACM*, vol. 33, pp. 668–676, 1990.

[31] R. Hettinger, "Regaining lost knowledge," `https://rhettinger.wordpress.com/tag/running-median`, 2010.

[32] J.R. Landis and G.G. Koch, "Measurement of observer agreement for categorical data," *Biometrics*, vol. 33, pp. 159–174, 1977.

[33] D.G. Altman, *Practical Statistics for Medical Research.*, Chapman & Hall, London, 1991.