PAPER

Automatic classification of apnea/hypopnea events through sleep/wake states and severity of SDB from a pulse oximeter

IOP Publishing | Institute of Physics and Engineering in Medicine Physiological Measurement

Physiol. Meas. **36** (2015) 2009–2025 doi:10.1088/0967-3334/36/9/2009

Automatic classification of apnea/hypopnea events through sleep/wake states and severity of SDB from a pulse oximeter

Jong-Uk Park[1], Hyo-Ki Lee[2], Junghun Lee[3], Erdenebayar Urtnasan[1], Hojoong Kim[4] and Kyoung-Joung Lee[1]

[1] Department of Biomedical Engineering, Yonsei University, Wonju, Gangwondo, Korea
[2] Department of Surgery, College of Medicine, Interdisciplinary Consortium on Advanced Motion Performance (iCAMP), University of Arizona, Tucson, AZ 85721, USA
[3] Department of Emergency Medicine, Wonju College of Medicine, Yonsei University, Wonju, Gangwondo, Korea
[4] Division of Pulmonary and Critical Care Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

E-mail: lkj5809@yonsei.ac.kr

Received 6 March 2015
Accepted for publication 15 June 2015
Published 10 August 2015

Abstract

This study proposes a method of automatically classifying sleep apnea/hypopnea events based on sleep states and the severity of sleep-disordered breathing (SDB) using photoplethysmogram (PPG) and oxygen saturation (SpO_2) signals acquired from a pulse oximeter. The PPG was used to classify sleep state, while the severity of SDB was estimated by detecting events of SpO_2 oxygen desaturation. Furthermore, we classified sleep apnea/hypopnea events by applying different categorisations according to the severity of SDB based on a support vector machine. The classification results showed sensitivity performances and positivity predictive values of 74.2% and 87.5% for apnea, 87.5% and 63.4% for hypopnea, and 92.4% and 92.8% for apnea + hypopnea, respectively. These results represent better or comparable outcomes compared to those of previous studies. In addition, our classification method reliably detected sleep apnea/hypopnea events in all patient groups without bias in particular patient groups when our algorithm was applied to a variety of patient groups. Therefore, this

method has the potential to diagnose SDB more reliably and conveniently using a pulse oximeter.

Keywords: sleep apnea/hypopnea, photoplethysmogram, oxygen saturation, sleep/wake states, severity of SDB

1. Introduction

Sleep-disordered breathing (SDB) collectively refers to a variety of breathing disorders that occur during sleep, which can be broadly divided into sleep apnea and sleep hypopnea. When sleep apnea or hypopnea occurs, oxygen cannot be sufficiently supplied to the human body, thus hypoxia and reduced sleep quality may occur due to sleep fragmentation. These disorders also increase the risk of comorbid vascular diseases, such as high blood pressure, cardiac arrhythmia, and stroke in severe cases (Collop 2007).

A standard method for diagnosing SDB is polysomnography (PSG). PSG provides objective indicators to assess the severity of SDB by simultaneously recording a variety of biological signals during sleep: electroencephalogram (EEG); electrooculogram (EOG); electrocardiogram (ECG); respiratory airflow; photoplethysmogram (PPG); and oxygen saturation (SpO_2) (Kushida *et al* 2005). However, PSG requires data interpretation by clinical experts at a specialized sleep center, which is associated with considerable costs related to facilities, human resources, and equipment. For patients, this translates into time, space, and economic burdens (Su *et al* 2004).

To reduce the inconveniences associated with PSG, many researchers have proposed using a pulse oximeter (Fietze *et al* 2004, Marcos *et al* 2010, Koley *et al* 2014). A pulse oximeter can measure PPG and SpO_2 while checking functions in the autonomic nervous system through pulse rate variability (PRV) analysis (Rauh *et al* 2004); it can also diagnose SDB by measuring the level of oxygen desaturation (Kushida *et al* 2005). Heart rate variability (HRV) presents fluctuations related to autonomic nervous system (ANS) activity (Hossen *et al* 2011). Rauh *et al* (2004) compared PRV parameters with HRV parameters; they concluded that the PRV method could be used more easily as an ANS screening tool. Fietze *et al* (2004) verified that the oxygen desaturation index (ODI), which is the total sum of oxygen desaturation events per hour, was highly correlated with SDB severity; they concluded that ODI could be used for SDB diagnosis. In addition, Koley *et al* (2014) extracted SpO_2 characteristics, which changed according to apnea or hypopnea, and classified sleep apnea or hypopnea events using a support vector machine (SVM) classifier.

However, existing studies using a pulse oximeter have failed to differentiate apnea and hypopnea, while methods using only SpO_2 could result in overestimation or underestimation of incidents, depending on the patient group (Chung *et al* 2012). This misclassification results from the fact that patient sleep/wake states cannot be identified using only SpO_2; in addition, the pattern of oxygen desaturation might differ according to SDB severity. Apnea and hypopnea occur only during sleep, and incidents of oxygen desaturation that occur during waking hours may be classified incorrectly as sleep apnea/hypopnea events. The frequency of apnea/hypopnea events also differs according to severity, and these events also have varying SpO_2 characteristics. Under such circumstances, results may be biased due to the concentration on severe patients, thereby misclassifying the sleep apnea/hypopnea events of mildly affected patients.

To improve on previous studies, this research classified sleep apnea and hypopnea events while taking sleep/wake states and severity of SDB into consideration. A change in the autonomic nervous system occurs according to sleep/wake states, and this change is reflected in PRV. In addition, SDB severity can be determined via the ODI. Therefore, in this study, we propose an

Table 1. Summary of demographic and anthropometric information for the patient groups.

All subjects	Total	Mild ($5 \leqslant$ AHI < 15)	Moderate ($15 \leqslant$ AHI < 30)	Severe (AHI $\geqslant 30$)	p-value
Subjects (n)	30	10	10	10	
Age (years)	55.7 ± 12.6	57.3 ± 17.2	54.3 ± 11.04	55.6 ± 9.5	*NS*
Male (%)	60%	60%	60%	60%	*NS*
BMI (kg m^{-2})	25.2 ± 2.9	25.1 ± 1.9	23.9 ± 2.1	26.6 ± 3.9	*NS*
TRT (h)	7.4 ± 0.8	7.7 ± 0.7	7.1 ± 1.2	7.4 ± 0.6	*NS*
TST (h)	5.8 ± 0.9	5.5 ± 0.8	5.9 ± 1.0	5.9 ± 0.8	*NS*
AHI (e/h)	26.6 ± 17.2	9.5 ± 2.7	23.0 ± 4.2	47.3 ± 10.8	< 0.001
Training set	Total	Mild ($5 \leqslant$ AHI < 15)	Moderate ($15 \leqslant$ AHI < 30)	Severe (AHI $\geqslant 30$)	p-value
Subjects (n)	15	5	5	5	
Age (years)	55.1 ± 7.0	56.8 ± 8.9	54.4 ± 6.3	54.0 ± 6.8	*NS*
Male (%)	60%	60%	60%	60%	*NS*
BMI (kg m^{-2})	25.1 ± 2.5	25.0 ± 1.3	23.5 ± 1.5	26.6 ± 3.5	*NS*
TRT (h)	7.4 ± 0.6	7.5 ± 0.8	7.5 ± 0.7	7.0 ± 0.5	*NS*
TST (h)	5.7 ± 0.6	5.5 ± 0.3	5.9 ± 0.7	5.7 ± 0.8	*NS*
AHI (e/h)	25.6 ± 16.2	9.5 ± 3.1	23.4 ± 4.4	43.9 ± 11.9	< 0.001
Test set	Total	Mild ($5 \leqslant$ AHI < 15)	Moderate ($15 \leqslant$ AHI < 30)	Severe (AHI $\geqslant 30$)	p-value
Subjects (n)	15	5	5	5	
Age (years)	56.4 ± 16.7	57.8 ± 24.2	54.2 ± 15.3	57.2 ± 12.3	*NS*
Male (%)	60%	60%	60%	60%	*NS*
BMI (kg m^{-2})	25.3 ± 3.3	25.1 ± 2.5	24.2 ± 2.6	26.6 ± 4.7	*NS*
TRT (h)	7.4 ± 1.0	7.8 ± 0.6	6.8 ± 1.5	7.8 ± 0.4	*NS*
TST (h)	5.8 ± 1.1	5.5 ± 1.2	5.8 ± 1.4	6.2 ± 0.7	*NS*
AHI (e/h)	27.6 ± 18.7	9.5 ± 2.7	22.6 ± 4.5	50.6 ± 9.8	< 0.001

Note: data are presented as mean ± SD, numbers (n), or percentages (%). kg m^{-2}: kilogram per square meter. e/h: events per hour. BMI: body mass index. TRT: total record time. TST: total sleep time. AHI: apnea/hypopnea index. *NS*: no significant difference among patient groups (p-value > 0.05).

algorithm that can classify sleep apnea/hypopnea by applying a different classifier according to SDB severity, calculating the ODI automatically and classifying sleep/wake states through PRV.

2. Material

2.1. Populations

The study subjects included 30 patients diagnosed with SDB via PSG after admission to the Samsung Medical Center in Seoul, Korea, from March 2009 to March 2012. Table 1 provides the patient characteristics. The Institutional Review Board of the Samsung Medical Center authorized the study, which included only those who provided written consent. The 30 patients were divided into a training set and a test set with 15 in each set to implement and

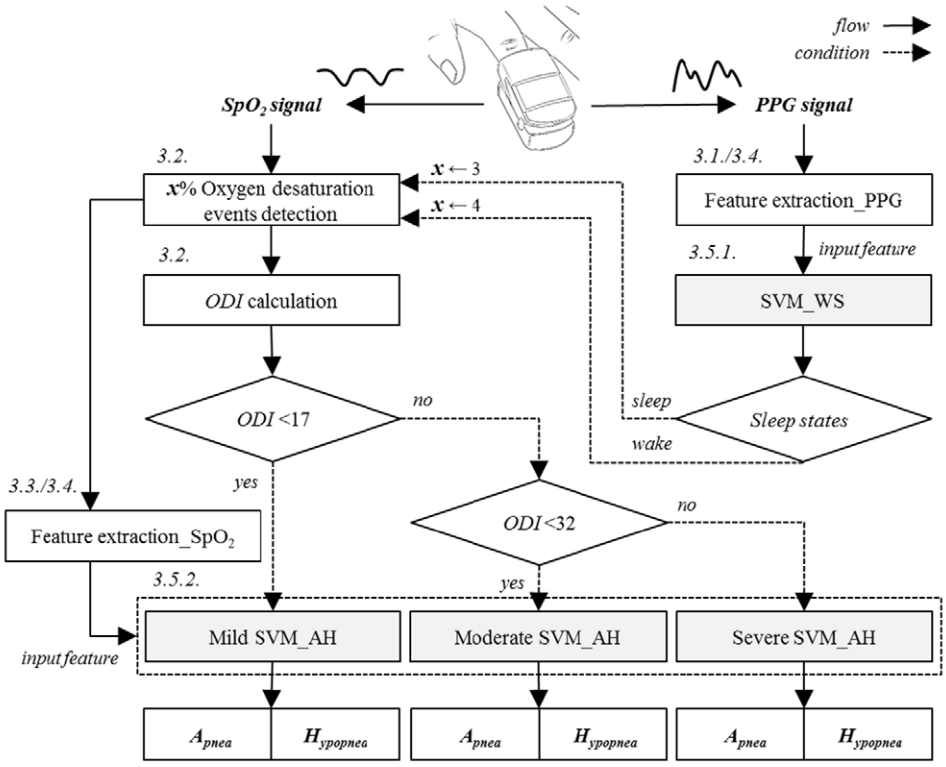

Figure 1. Classification algorithm of sleep apnea/hypopnea considering sleep/wake states and SDB severity.

calculate an algorithm. The training and test sets were not significantly different with regard to clinical characteristics ($p > 0.05$).

2.2. Instrumentation and measurement

Data were acquired using a PSG device (Embla N7000 PSG Amplifier, Embla Systems LLC, USA). Each examination took at least 6 h. The polysomnogram included an electroencephalogram (C3/A2 and O2/A1 electroencephalography), left and right eye movement (electrooculography), chin and leg electromyography, rib cage and abdomen movement (piezoelectric belts), body position, nasal pressure, nasal thermistor, electrocardiography (ECG), and finger pulse oximetry (Oximeter Flex Sensor 8000J, Embla Systems LLC, USA). The PPG obtained using a pulse oximeter was recorded at 16 bits/sample and 100 samples/second, while the SpO_2 signal was 8 bits/sample and 2 samples/second. The data were stored in a personal computer. Clinical experts analyzed the sleep stages and sleep apnea/hypopnea events using RemLogic PSG Software (Embla Systems LLC, USA) according to the 2007 American Academy of Sleep Medicine (AASM) publication (Iber *et al* 2007).

3. Methods

Figure 1 presents the design of the proposed algorithm used in this study. The solid arrow line and dotted line in the flow diagram represent the flow and condition, respectively. Sleep/

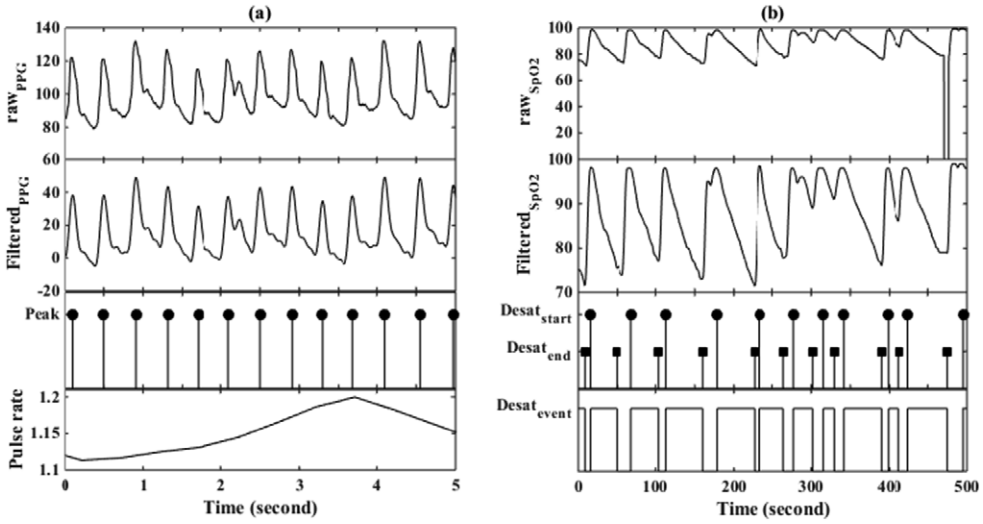

Figure 2. Pre-processing of (a) PPG and (b) SpO_2.

wake states were classified using the characteristics of the PPG acquired via a pulse oximeter, and the oxygen desaturation events of SpO_2 were detected with different weights depending on sleep/wake states. The ODI was calculated using the detected oxygen desaturation events to estimate severity; a different classifier (Mild SVM_AH, Moderate SVM_AH, or Severe SVM_AH) was selected according to the ODI value. A classifier categorized sleep apnea and hypopnea using a feature vector for oxygen desaturation events. The following subsections provide a detailed explanation of the processes depicted in figure 1.

3.1. Feature extraction from PPG signal for classifying sleep/wake states

3.1.1. Pre-processing. Figure 2(a) shows the pre-processing for feature extraction from PPG. To calculate PPG pulse-to-pulse intervals, baseline wander and power noise were removed using a band-pass filter of 0.1–5 Hz, and a maximum point of pulse was detected using an adaptation threshold algorithm (Shin *et al* 2009). By changing the threshold slope according to the previous pulse amplitude, this adaptation threshold algorithm shows strong performance in detecting the maximum point even in rapidly changing amplitudes.

An initial value (TH_{init}) of the threshold was estimated according to formula (1), while the threshold (TH_n) was updated by applying a weight of ($-0.6/f_s$) to the maximum value of the previous pulse according to formula (2). Then, a maximum point (n_{MAX_i}) was detected as the largest value among the values above the threshold, while the pulse rate (PR) was calculated according to formula (3).

$$TH_{\text{init}} = 0.2 \underset{n}{\text{argmax}} \{x_{\text{PPG}}(n)\},\ n \in [0,\ 5f_s] \tag{1}$$

and

$$TH_n = TH_{n-1} - \frac{0.6}{f_s} \cdot x_{\text{PPG}}(n_{\text{MAX}_{i-1}}) \tag{2}$$

where $x_{\text{PPG}}(n)$ is the PPG signal at point n and f_s is the sampling frequency of PPG.

$$\mathrm{PR}(n) = \sum_i \frac{1}{f_s} \frac{1}{n_{\mathrm{MAX}_i} - n_{\mathrm{MAX}_{i-1}}} \delta(n - n_{\mathrm{MAX}_i}) \tag{3}$$

3.1.2. PRV feature extraction. PRV is the result of continuous alteration of the autonomic nervous regulation (Rauh *et al* 2004, Hossen *et al* 2011). When falling asleep, the parasympathetic nervous system activity increases and the sympathetic nervous system activity decreases, and these changes cause a decrease of pulse rate. Therefore, we calculated four time-domain parameters and six frequency-domain parameters in order to determine a change in PR according to sleep/wake states. An ectopic beat of PR in formula (3) was removed to extract the time-domain parameters, and this signal was defined as an *NN* signal (normal-to-normal PR). In addition, characteristics of the *NN* signal were calculated using a 30 s unit measurement in order to update sleep/wake states every 30 s according to AASM criteria (Iber *et al* 2007).

The four time-domain characteristics are as follows:

- *mean_NN*: mean of the *NN*
- *median_NN*: median of the *NN*
- *SDNN*: standard deviation of the *NN*
- *RMSSD*: root-mean-square of successive differences for the *NN*

The *NN* calculated to extract the frequency-domain parameters was interpolated equidistantly and then resampled at 4 Hz. A fast Fourier transform (FFT) was performed with the resampled signal, and the power spectrum density was calculated by taking the square of the FFT. To calculate the frequency-domain features, we used the following frequency bands: VLF (very low frequency: 0–0.04 Hz), LF (low frequency: 0.04–0.15 Hz), HF (high frequency: 0.15–0.4 Hz), and TF (total frequency: 0–0.4 Hz). The six frequency-domain features include the following:

- P_{VLF}: the power in the VLF band components
- P_{LF}: the power in the LF band components
- P_{HF}: the power in the HF band components
- $P_{\mathrm{LF}}/P_{\mathrm{TF}}$ ratio: LF band power/TF band power
- $P_{\mathrm{HF}}/P_{\mathrm{TF}}$ ratio: HF band power/TF band power
- $P_{\mathrm{LF}}/P_{\mathrm{HF}}$ ratio: LF band power/HF band power

3.2. Oxygen desaturation event detection and ODI calculation

Figure 2(b) shows the pre-processing for feature extraction from SpO_2. The spike noise was removed using a 10 point median filter to detect oxygen desaturation events. An oxygen desaturation event included the following criteria: a region with a decrease of more than 3% below the baseline during sleep or a decrease of more than 4% below the baseline during wake state, according to the sleep/wake states. The ODI was calculated as the total sum of oxygen desaturation events per hour while sleep/wake states were estimated using the PPG.

To determine whether the ODI was appropriate for assessing SDB severity, linear regression analysis was performed between the ODI estimates for a training set and the apnea/hypopnea index (AHI), as evaluated by clinical experts. As a result, a linear correlation of $f(x) = 0.9x + 0.9$ ($r^2 = 0.98$, $p < 0.01$) was found between the estimated ODI and AHI, as shown in figure 3(a). Figure 3(b) also illustrates the Bland–Altman verification that the mean difference between the two variables was 2.43. Accordingly, this study set the criteria of severity using the ODI from 15 and 30 to 17 and 32.

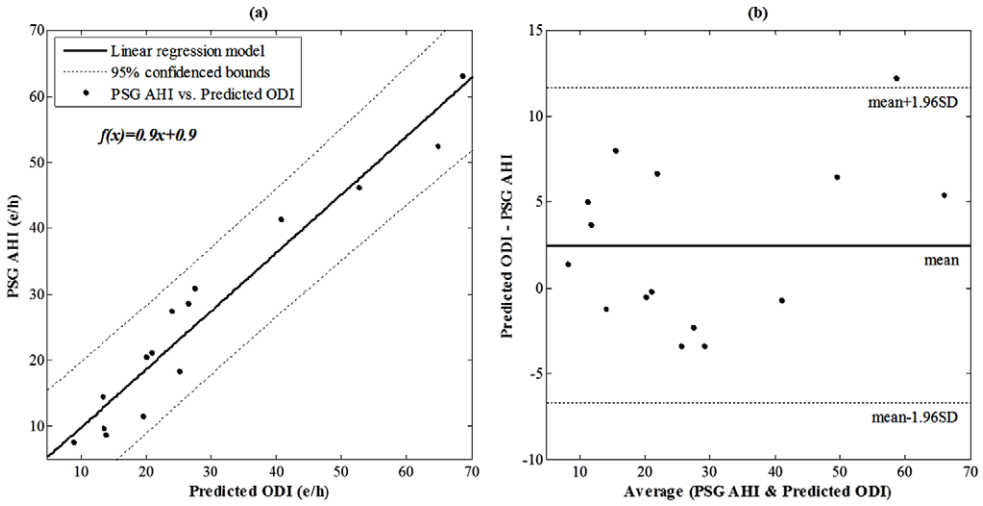

Figure 3. Statistical analyses of the relationship between PSG AHI and predicted ODI. (a) Regression analysis between estimated ODI and AHI assessed by clinical experts; (b) Bland–Altman plot.

3.3. Feature extraction from SpO_2 signal for classifying sleep apnea/hypopnea

To classify sleep apnea/hypopnea, the following seven features were calculated in the oxygen desaturation event region of the SpO_2 signals detected in section 3.2.

- *D_percent*: reduction rate of the oxygen desaturation event
- *D_time*: duration of the oxygen desaturation event
- *D_slope*: slope between the starting and ending points of the oxygen desaturation event interval
- *D_slope*80: slope between the starting point and the four-fifths point of the oxygen desaturation event interval
- *D_mean*: mean value of SpO_2 in the oxygen desaturation event
- *D_std*: standard deviation of SpO_2 in the oxygen desaturation event

3.4. Feature selection

Statistical analysis was conducted with the 15 patients assigned to the training set to determine whether there were any significant differences in the 10 features extracted from the PPG signals in section 3.1 for sleep/wake state classification between classes. An independent two-sample *t*-test was conducted for statistical analysis, and if a *p*-value was less than 0.05, the features were considered to be significantly different between classes (Petrie *et al* 2009). In addition, support vector machine-recursive feature elimination (SVM-RFE) was applied to determine the optimum number of features for classification among features with statistically significant differences (Guyon *et al* 2006). SVM-RFE removes a feature first and performs SVM using the rest of the features, thereby calculating a ranking score and removing a feature that was removed previously when the lowest ranking score was calculated. The removed features were set as the lowest ranking score features of the classification performance (Duda *et al* 2012). The above process was iterated until only one feature remained. The last feature

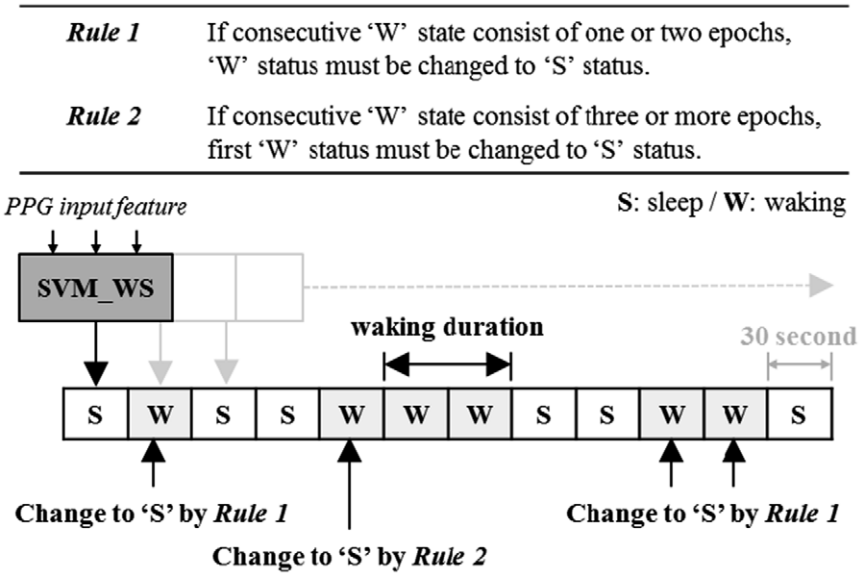

Figure 4. Classification method for sleep/wake states using SVM_WS.

remaining after this process held the highest classification performance, and a feature ranking was given based on this highest classification performance. Then, features were added one by one in order of highest performance to evaluate training and performance iteratively. Similarly, statistical analyses and SVM-RFE were conducted following the same process in order to select the optimum features from the six features extracted from the SpO_2 signals and described in section 3.3 for sleep apnea/hypopnea classification. For sleep apnea/hypopnea classification, the process was conducted to select the optimum features, which differed based on apnea severity (mild, moderate, and severe).

3.5. Classification

3.5.1. Support vector machine (SVM). In this study, SVM was used to classify sleep/wake states and sleep apnea/hypopnea. SVM, a machine learning algorithm, has shown better performance over other classifiers such as linear discriminant, neural network, etc, in sleep states or sleep apnea recognition (Duda *et al* 2012). The objective of SVM is to find the separating hyper-plane with the largest margin. For input data x, the decision function is expressed as formula (4). α_i^*, y_i and $K(x, x_i)$ represents the Lagrange multipliers, label of training data and radial basis function (RBF) kernel, respectively.

$$f(x, \alpha^*) = \sum_{i=1}^{N} \alpha_i^* y_i K(x, x_i) \tag{4}$$

3.5.2. Classification of the sleep/wake states. Figure 4 shows the process of classifying sleep/wake states. The sleep state was used to assign a weight for event detection of oxygen desaturation. A weight of '3' was given for a sleep state and '4' for a wake state. The SVM

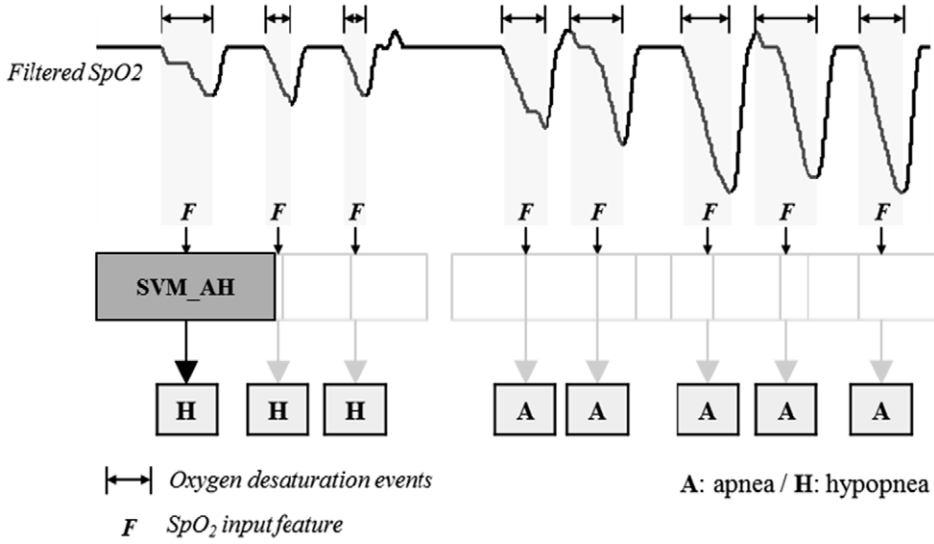

Figure 5. Classification method for sleep apnea/hypopnea using SVM_AH.

classifier for sleep/wake state was called SVM_WS. In SVM_WS, the PRV feature vector extracted from a pulse signal was used as an input, thereby dividing sleep (S) and wake (W) states. After this division, finally sleep/wake states were determined according to the two rules outlined in figure 4.

3.5.3. Classification of sleep apnea/hypopnea. As shown in figure 1, a classifier categorizing sleep apnea and hypopnea was called SVM_AH, and the following classifiers were selected depending on severity: mild SVM_AH (ODI $<$ 17), moderate SVM_AH (17 $\leqslant$ ODI $<$ 32), and severe SVM_AH (ODI $\geqslant$ 32). Figure 5 shows the classification process for sleep apnea and hypopnea. An oxygen desaturation feature vector was used for each classifier, thereby allowing classification into sleep apnea (A) and sleep hypopnea (H).

4. Results

4.1. Feature selection

Table 2 summarizes the features used as inputs for each classifier. Features that were significantly different between the two classes were marked with an asterisk (*), and were ranked in order of best classification performance using SVM-RFE. In addition, the features that were finally selected through the iterative process described in section 3.5 are displayed in bold, underlined font. Figure 6 shows the sensitivity (Sen.) calculation results according to formula (5), where significantly different features between classes were added one by one in the order of best classification performance. Feature vectors that showed the optimum performance with the least number of features were selected. Accordingly, SVM_WS had four feature vectors: (P_{HF}/P_{TF} *ratio*, P_{LF}/P_{TF} *ratio*, *RMSSD*, and *median_NN*). Mild SVM_AH had two feature vectors (*D_percent* and *D_amp*). Moderate SVM_AH had three feature vectors (*D_mean*, *D_slope*80, and *D_percent*), and severe SVM_AH had three feature vectors (*D_amp*, *D_std*, and *D_percent*).

Table 2. Features used for the inputs of each classifier.

Rank	SVM_WS (wake/sleep)	Mild SVM_AH (apnea/hypopnea)	Moderate SVM_AH (apnea/hypopnea)	Severe SVM_AH (apnea/hypopnea)
1	$\boldsymbol{P_{HF}/P_{TF}}$ ***ratio***[a]	***D_percent***[a]	***D_mean***[a]	***D_percent***[a]
2	$\boldsymbol{P_{LF}/P_{TF}}$ ***ratio***[a]	***D_min***[a]	***D_slope* 80**[a]	***D_std***[a]
3	***RMSSD***[a]	*D_slope* 80[a]	***D_percent***[a]	***D_min***[a]
4	***median_NN***[a]	*D_mean*[a]	*D_min*[a]	*D_time*[a]
5	*mean_NN*[a]	*D_slope*[a]	*D_time*[a]	*D_slope*
6	P_{HF}[a]	*D_time*[a]	*D_std*[a]	*D_slope* 80[a]
7	P_{LF} / P_{HF} *ratio*[a]	*D_std*[a]	*D_slope*	*D_mean*
8	P_{LF}[a]			
9	*SDNN*[a]			
10	P_{VLF}[a]			

[a]Significant difference between groups ($p < 0.05$), ***feature*** [a]: selected feature.

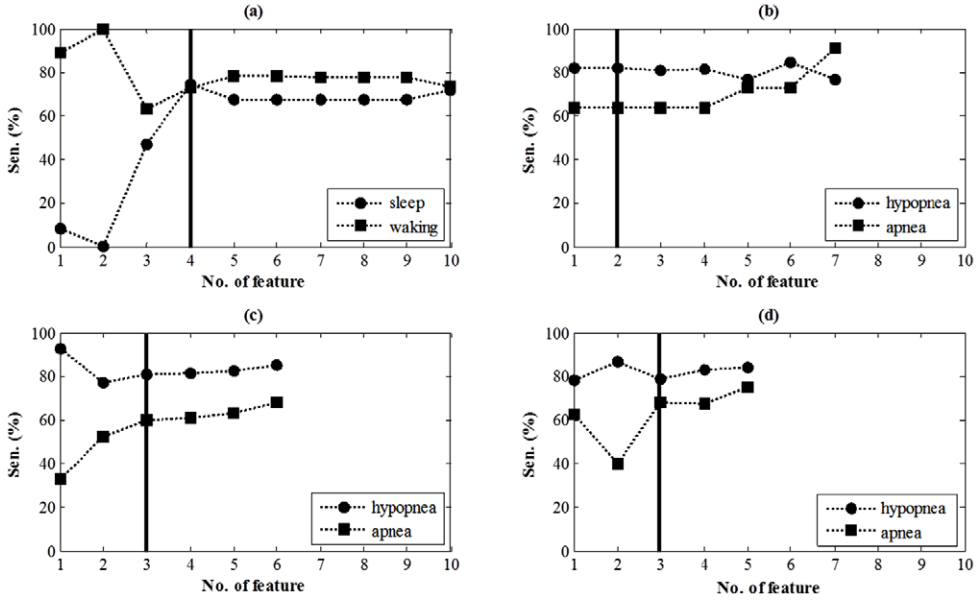

Figure 6. Performance evaluation of classifiers according to the number of features: (a) SVM_WS, (b) SVM_Mild_AH, (c) SVM_Moderate_AH, and (d) SVM_Severe_AH.

4.2. Evaluation of classification performance

To evaluate the classification results of each class, the Sen. and positive predictive value (PPV) were calculated using formulas (5) and (6), respectively, while the performance evaluation was conducted with training and test sets as per severity.

$$\text{Sen. } (\%) = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \tag{5}$$

Table 3. Classification performance for sleep/wake states using PPG.

Training set		Mild ($5 \leqslant$ AHI < 15)	Moderate ($15 \leqslant$ AHI < 30)	Severe (AHI $\geqslant 30$)	Total
Sleep	Sen. (%)	82.6	81.7	85.6	83.3
	PPV (%)	88.3	84.4	89.0	87.2
Wake	Sen. (%)	59.5	30.6	46.4	46.4
	PPV (%)	48.0	26.6	39.0	38.8
Test set		Mild ($5 \leqslant$ AHI < 15)	Moderate ($15 \leqslant$ AHI < 30)	Severe (AHI $\geqslant 30$)	Total
Sleep	Sen. (%)	78.7	85.9	80.1	81.4
	PPV (%)	88.0	83.3	79.7	83.5
Wake	Sen. (%)	60.2	36.1	20.6	39.3
	PPV (%)	43.2	40.8	21.1	36.0

Abbreviations: Sen.: sensitivity, PPV: positive predictive value, A + H: apnea and hypopnea.

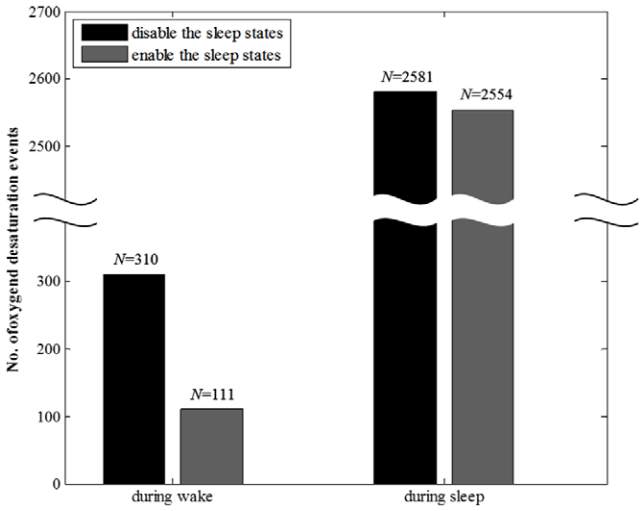

Figure 7. The number of oxygen desaturation events detected while considering sleep/wake states.

$$\text{PPV}\ (\%) = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100 \tag{6}$$

where TP represents the number of true positives, FN is the number of false negatives, and FP is the number of false positives.

4.2.1. Classification of the sleep/wake states. Table 3 shows results from the classification of sleep and wake using four PRV feature vectors extracted from a pulse signal. The Sen. and PPV in a total test set showed the following performance: 83.8% and 87.2%, respectively, in the case of sleep and 39.3% and 36.0%, respectively, in the case of wake state.

Figure 7 shows a bar graph representing the number of oxygen desaturation events detected depending on enabling or disabling a sleep state. When sleep was disabled, the number of oxygen desaturation events was detected as follows: 310 events during wake and 2581 events during sleep. When sleep was enabled, 111 events during wake and 2554 events during sleep were detected. No oxygen desaturation events that occurred during wake were related to sleep

Table 4. Classification performance for sleep apnea/hypopnea in consideration of sleep/wake states and SDB severity.

Training set		Mild ($5 \leqslant$ AHI < 15)	Moderate ($15 \leqslant$ AHI < 30)	Severe (AHI $\geqslant 30$)	Total
Apnea	Sen. (%)	81.8	66.0	70.3	69.0
	PPV (%)	32.1	82.8	84.2	82.1
Hypopnea	Sen. (%)	86.7	83.2	84.9	84.8
	PPV (%)	77.3	63.0	63.3	68.0
A + H	Sen. (%)	72.6	91.3	93.4	90.2
	PPV (%)	82.5	97.1	96.4	95.1
Test set		Mild ($5 \leqslant$ AHI < 15)	Moderate ($15 \leqslant$ AHI < 30)	Severe (AHI $\geqslant 30$)	Total
Apnea	Sen. (%)	64.0	63.2	77.0	74.2
	PPV (%)	55.2	77.1	90.6	87.5
Hypopnea	Sen. (%)	95.8	88.4	81.3	87.5
	PPV (%)	66.5	68.2	56.9	63.4
A + H	Sen. (%)	74.0	86.7	97.7	92.4
	PPV (%)	73.7	93.0	96.0	92.8

Abbreviations: Sen.: sensitivity, PPV: positive predictive value, A + H: apnea and hypopnea.

Table 5. Classification results for severity of sleep disordered breathing.

Proposed system \ PSG	Mild (*N*)	Moderate (*N*)	Severe (*N*)	Total
Mild (*N*)	4	1	0	5
Moderate (*N*)	1	4	0	5
Severe (*N*)	0	0	5	5
Total (*N*)	5	5	5	15

Abbreviations: PSG: polysomnography, *N*: number.

apnea/hypopnea. In addition, 199 events that occurred during wake by enabling a sleep state were not detected. In contrast, all 27 events that were not detected during sleep were oxygen desaturation events due to apnea/hypopnea.

4.2.2. Classification of the sleep apnea/hypopnea events. Table 4 shows the classification results of sleep apnea/hypopnea using the oxygen desaturation feature vectors. The classification results of the entire test set showed Sen. and PPV values, respectively, of 74.2% and 87.5% with respect to apnea, 87.5% and 63.4% with respect to hypopnea, and 92.4% and 92.8% with respect to apnea + hypopnea.

We calculated AHI and classified SDB severity. Table 5 shows the classification performance for SDB severity. 13 of the 15 patients were correctly classified, 2 others were incorrectly classified as mild to moderate patient (AHI: 11.5 → 19.7) and moderate to mild patient (AHI: 17.6 → 13.6), respectively.

5. Discussion

5.1. Comparison between the conventional method and the proposed method

Additional experiments were conducted to determine the differences between the conventional method and our proposed method. For this comparison, a system lacking the consideration of sleep/wake states and SDB severity was called SYS_CON, while this study's proposed system

Table 6. Comparison of classification performance for sleep apnea/hypopnea between SYS_CON and SYS_PRO.

SYS_CON (test set)		Mild ($5 \leqslant$ AHI < 15)	Moderate ($15 \leqslant$ AHI < 30)	Severe (AHI $\geqslant 30$)	Total
Apnea	Sen. (%)	12.0	48.0	82.4	75.7
	PPV (%)	15.0	69.8	90.9	87.0
Hypopnea	Sen. (%)	97.0	88.9	76.3	86.2
	PPV (%)	60.3	64.2	45.9	55.8
A + H	Sen. (%)	71.2	85.5	97.3	91.4
	PPV (%)	57.1	87.5	91.6	86.2
SYS_PRO (test set)		Mild ($5 \leqslant$ AHI < 15)	Moderate ($15 \leqslant$ AHI < 30)	Severe (AHI $\geqslant 30$)	Total
Apnea	Sen. (%)	64.0	63.2	77.0	74.2
	PPV (%)	55.2	77.1	90.6	87.5
Hypopnea	Sen. (%)	95.8	88.4	81.3	87.5
	PPV (%)	66.5	68.2	56.9	63.4
A + H	Sen. (%)	74.0	86.7	97.7	92.4
	PPV (%)	73.7	93.0	96.0	92.8

Abbreviations: Sen.: sensitivity, PPV: positive predictive value, A + H: apnea and hypopnea.

was called SYS_PRO. Table 6 shows the performance evaluation from applying SYS_CON and SYS_PRO to a test set. The total classification performance of SYS_CON and SYS_PRO was comparable, but the performance of SYS_PRO was relatively better when considering severity. In particular, the Sen. and PPV with respect to the apnea classification of mildly affected patients showed a big difference between SYS_CON and SYS_PRO: 12.0% and 15.0% respectively in SYS_CON, and 64.0% and 55.2% respectively in SYS_PRO. These results indicated that the false positive classification of apnea/hypopnea for mildly affected patients might occur due to biased training results. The occurrence of apnea/hypopnea events may be more frequent in severely affected patients, and a classifier may be trained without consideration of severity. In addition, no significant differences were found in total performance between SYS_PRO and SYS_CON despite SYS_PRO's superior performance in considering severity, because the detection performance of sleep apnea/hypopnea events for severely affected patients impacted total performance the most. However, when classification algorithms are applied to a variety of patient groups, it is more important to show reliable and unbiased performance without effects from a particular patient group than to have better total performance. Table 7 shows the performance per patient in terms of average classification performance ± standard deviation (SD) when SYS_CON and SYS_PRO were applied to a test set. In most cases, SYS_PRO showed better performance and lower standard deviation than SYS_CON. This result reveals that the method proposed in this study can classify sleep apnea/hypopnea more reliably and with better performance.

Table 8 summarizes the existing studies that have classified sleep apnea/hypopnea events using one or two sensors. Koley *et al* (2014) extracted time-domain features of SpO_2 and classified sleep apnea using an SVM classifier. Babaeizadeh *et al* (2010) extracted time-frequency domain features of HRV using ECG and classified the sleep apnea using QDA (quadratic discriminant analysis). Gil *et al* (2008) classified sleep apnea using decreases in the amplitude fluctuation of PPG (DAP). However, these studies failed to differentiate between sleep apnea and hypopnea while performing classifications. De Chazal *et al* (2009) differentiated between sleep apnea and hypopnea using SpO_2 and ECG. However, this study reported a lower performance. Koley *et al* (2013) differentiated among normal breathing, sleep apnea, and sleep hypopnea using airflow, thereby resulting in high performances of 92.4%, 86.0%, and 91.4%,

Table 7. Comparison of classification performance for sleep apnea/hypopnea per patient between SYS_CON and SYS_PRO.

SYS_CON (test set)		Mild ($5 \leqslant$ AHI < 15)	Moderate ($15 \leqslant$ AHI < 30)	Severe (AHI $\geqslant 30$)	Total
Apnea	Sen. (%)	18.0 ± 24.9	40.1 ± 22.1	79.9 ± 17.9	46.0 ± 33.4
	PPV (%)	13.0 ± 18.6	62.7 ± 24.2	88.9 ± 8.2	54.8 ± 36.7
Hypopnea	Sen. (%)	96.5 ± 3.1	90.1 ± 6.1	67.4 ± 23.4	84.7 ± 18.3
	PPV (%)	60.9 ± 14.2	63.6 ± 20.0	41.0 ± 12.2	55.2 ± 18.0
A + H	Sen. (%)	71.5 ± 7.4	84.4 ± 6.9	97.1 ± 3.6	84.36 ± 12.2
	PPV (%)	59.3 ± 23.3	85.8 ± 15.9	90.5 ± 6.7	78.6 ± 21.0
SYS_PRO (test set)		Mild ($5 \leqslant$ AHI < 15)	Moderate ($15 \leqslant$ AHI < 30)	Severe (AHI $\geqslant 30$)	Total
Apnea	Sen. (%)	76.7 ± 24.3	63.4 ± 14.4	73.0 ± 15.5	71.0 ± 18.2
	PPV (%)	49.3 ± 22.9	76.7 ± 15.8	87.7 ± 8.4	71.2 ± 22.8
Hypopnea	Sen. (%)	95.6 ± 3.9	90.3 ± 7.1	75.3 ± 51.6	87.0 ± 13.0
	PPV (%)	67.4 ± 14.2	67.8 ± 19.3	15.9 ± 22.8	62.3 ± 19.3
A + H	Sen. (%)	74.8 ± 8.2	85.6 ± 6.8	97.5 ± 3.4	86.0 ± 11.3
	PPV (%)	72.5 ± 14.9	91.3 ± 9.1	95.4 ± 2.9	86.4 ± 14.0

Note: results are presented as mean ± SD. Sen.: sensitivity, PPV: positive predictive value, A + H: apnea and hypopnea.

respectively. However, it is more convenient to use a pulse oximeter than to rely on a thermal sensor measuring airflow, and a thermal sensor is not suited as a single sensor for detecting a hypopnea (Iber *et al* 2007). In this study, we automatically classified sleep apnea/hypopnea events using SpO_2 and PPG acquired from a pulse oximeter. The classification performance showed Sen. and PPV, respectively, of 74.2% and 87.5% for apnea, 87.5% and 63.4% for hypopnea, and 92.4% and 92.8% for apnea + hypopnea. These results represent better or comparable outcomes compared to those of previous studies. Also, our study is highly significant because it considered sleep/wake states and SDB severity in classifying sleep apnea and hypopnea. Therefore, this method has the potential for more reliably and conveniently screening SDB.

5.2. Oxygen desaturation event detection considering sleep/wake states

To determine a weight during the detection of oxygen desaturation while considering sleep/wake states, the sleep/wake states were estimated using a pulse rate validate feature vector. After this estimate, oxygen desaturation events were detected if oxygen desaturation decreased by more than 3% during sleep or by more than 4% during wake states as compared to the baseline. However, as the introduction of this paper mentioned, apnea and hypopnea did not occur during the wake state. Nonetheless, oxygen desaturation events were detected even in the wake region, because a sleep state can be estimated incorrectly. If a sleep state is estimated as a wake state, the oxygen desaturation events caused by sleep apnea/hypopnea cannot be detected; as a result, false negative classifications will increase, thereby lowering the classification performance. Therefore, oxygen desaturation events were detected by assigning a different weight during the wake state. According to the AASM guidelines, an oxygen desaturation event is determined if the oxygen saturation is decreased to more than 3% from baseline. Figure 8 shows the number of oxygen desaturation events caused by the attenuation of oxygen saturation during wake state. Most of the oxygen desaturation events

Table 8. Comparison between existing related studies and this study.

Authors (year of publication)	No. of subjects (female/male)	AHI	Signal	Methods	Results (class)
Koley *et al* (2013)	Total: 56 (-/-)	—	Air flow	• Time & frequency features	Acc.:
	Mild: 6	2.4 ± 1.6		• SVM classifier	• (normal): 92.4%
	Moderate: 25	9.6 ± 3.2			• (apnea): 86.0%
	Severe: 25	38.6 ± 14.5			• (hypopnea): 91.4%
Koley *et al* (2014)	Total: 34 (-/-)	—	SpO_2	• Time features	Acc.: 93.8% (normal/apnea)
	Mild: 9	2.5 ± 1.4		• SVM classifier	
	Moderate: 11	9.6 ± 3.1			
	Severe: 14	32.5 ± 13.8			
Babaeizadeh *et al* (2010)	Total: 32 (7/25)	⩾0	ECG	• Heart rate variability	Acc.: 84.7%
				• Quadratic classifier	Sen.: 76.7%
					Spc.: 89.6% (normal/apnea)
Gil *et al* (2007)	Total: 44 (16/28)	⩾0	PPG	• DAP detector	Sen.: 76.0%
	OSAS: 21				PPV: 73.0% (normal/apnea)
	Doubt: 4				
	Control: 19				
De Chazal *et al* (2011)	Total: 183 (10/173)	⩾0	SpO_2 + ECG	• EDR & RR features	Sen.:
	Mild: 45			• Time features	• (apnea): 55.0%
	Moderate: 27			• LDA classifier	• (hypopnea): 48.0%
	Severe: 69				• (A + H): 84.0%
This work	Total: 30 (15/15)	26.6 ± 17.2	SpO_2 + PPG (pulse oximeter)	• Sleep state (PRV)	[Sen., PPV]:
	Mild: 10	9.5 ± 2.7		• Severity of SDB (ODI)	• (apnea): [74.2, 87.5]%
	Moderate: 10	23.0 ± 4.2		• Desaturation features	• (hypopnea): [87.5, 63.4]%
	Severe: 10	47.3 ± 10.8		• SVM classifier	• (A + H): [92.4, 92.8]%

Abbreviations: SVM: support vector machine, DAP: decreases in the amplitude of PPG, EDR: ECG-derived respiratory signal, LDA: linear discriminant analysis, Acc.: accuracy, Sen.: sensitivity, Spc.: specificity, PPV: positive predictive value, A + H: apnea and hypopnea.

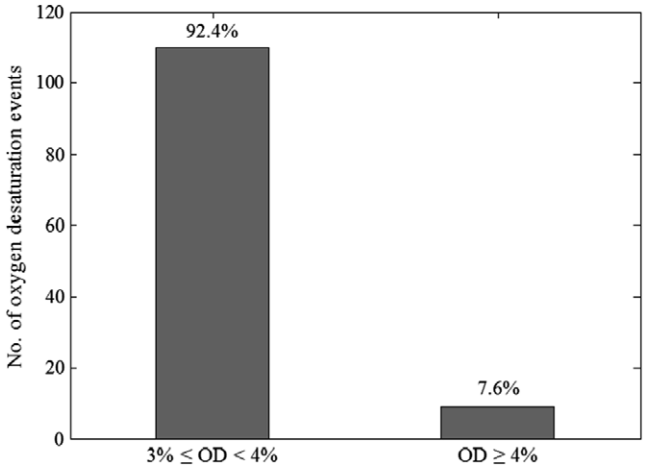

Figure 8. The number of oxygen desaturation (OD) events caused by the attenuation of oxygen saturation during wake state.

during the wake state were decreased to 3–4% in the oxygen saturation. In this study, oxygen desaturation events were detected if the oxygen saturation is decreased to more than 3% during sleep state or more than 4% during wake state.

In addition, as shown in figure 4, sleep/wake states were ultimately decided by applying rules 1 and 2. These rules aimed for a final determination using highly possible wake regions, thereby only preventing the low classification performance of sleep apnea/hypopnea due to incorrectly estimated wake states. Low classification performance results during the wake state were obtained by applying the above criteria. However, it is more important in this study to classify sleep apnea/hypopnea accurately with high reliability rather than to produce high classification performance for the determination of sleep/wake states.

5.3. The limitations of this study and further studies

To estimate the SDB severity, oxygen desaturation events occurring during the total recording time were detected, and ODI were calculated. Because this proposed method detects sleep apnea/hypopnea while considering severity, it can only be done with off-line processing systems. However, it is also necessary to study online processing systems that give feedback to patients by detecting sleep apnea/hypopnea events in real time. To this end, it is necessary to devise a method that estimates severity by only using data in particular regions.

In this study, sleep apnea/hypopnea was classified using feature vectors of oxygen desaturation events. However, these oxygen desaturation events occurred with some delay after oxygen apnea/hypopnea events. Because of this characteristic, it is necessary to compensate for the delay in oxygen desaturation events in order to classify sleep apnea/hypopnea events accurately.

Also, we will conduct an additional experiment with data obtained from more subjects to obtain more reliable results.

6. Conclusion

This study proposed a method for classifying sleep apnea/hypopnea events using a pulse oximeter as a single sensor. Through the analysis of PRV and ODI, sleep apnea/hypopnea

was classified while considering sleep/wake states and SDB severity. As a result, sleep apnea and hypopnea could be classified reliably without being biased by a particular patient group.

Therefore, through this and future research, screening of SDB with reliable performance may be expected through convenient in-home use of a pulse oximeter.

Acknowledgments

This work was supported by the Human Resource Training Program for Regional Innovation and Creativity through the Ministry of Education and National Research Foundation of Korea (NRF-2014H1C1A1063845).

References

Babaeizadeh S *et al* 2010 Automatic detection and quantification of sleep apnea using heart rate variability *J. Electrocardiol.* **43** 535–41

Chung F *et al* 2012 Oxygen desaturation index from nocturnal oximetry: a sensitive and specific tool to detect sleep-disordered breathing in surgical patients *Anesth. Analg.* **114** 993–1000

Collop N 2007 The effect of obstructive sleep apnea on chronic medical disorders *Cleve. Clin. J. Med.* **74** 72–8 (PMID: 17373350)

De Chazal P, Heneghan C and McNicholas W T 2009 Multimodal detection of sleep apnoea using electrocardiogram and oximetry signals *Phil. Trans. R. Soc.* A **367** 369–89

Duda R O, Hart P E and Stork D G 2012 *Pattern Classification* 2nd edn (New York: Wiley)

Fietze I *et al* 2004 Night-to-night variation of the oxygen desaturation index in sleep apnoea syndrome *Eur. Respir. J.* **24** 987–93

Gil E, Maria Vergara J and Laguna P 2008 Detection of decreases in the amplitude fluctuation of pulse photoplethysmography signal as indication of obstructive sleep apnea syndrome in children *Biomed. Signal Process. Control* **3** 267–77

Guyon I *et al* 2006 *Feature Extraction: Foundations and Applications* (Heidelberg: Springer)

Hossen A, Al-Ghunaimi B and Hassan M 2011 The importance of the very low frequency power of heart rate variability in screening of patients with obstructive *Proc. IEEE Symp. on Industrial Electronics and Applications (ISIEA), Sleep Apnea* (*Langkwai, Malaysia*) pp 638–42

Iber C *et al* 2007 *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications* (Westchester, IL: American Academy of Sleep Medicine)

Koley B L and Dey D 2013 Automatic detection of sleep apnea and hypopnea events from single channel measurement of respiration signal employing ensemble binary SVM classifiers *Measurement* **46** 2082–92

Koley B and Dey D 2014 On-line detection of apnea/hypopnea events using SpO2 signal: a rule-based approach employing binary classifier models *IEEE J. Biomed. Health Inform.* **18** 231–9

Kushida C A *et al* 2005 Practice parameters for the indications for polysomnography and related procedures: an update for 2005 *Sleep* **28** 499–521 (PMID: 16171294)

Marcos J V *et al* 2010 The classification of oximetry signals using Bayesian neural networks to assist in the detection of obstructive sleep apnoea syndrome *Physiol. Meas.* **31** 375–94

Petrie A and Sabin C 2009 *Medical Statistics at a Glance* 3rd edn (New York: Wiley)

Rauh R *et al* 2004 Comparison of heart rate variability and pulse rate variability detected with photoplethysmography *Proc. of the Saratov Fall Meeting 2003: Optical Technologies in Biophysics and Medicine V* (*Billingham, USA*) pp 115–26

Shin H S, Lee C and Lee M 2009 Adaptive threshold method for the peak detection of photoplethysmographic waveform *Comput. Biol. Med.* **39** 1145–52

Su S *et al* 2004 A comparison of polysomnography and a portable home sleep study in the diagnosis of obstructive sleep apnea syndrome *J. Otolaryngol. Head Neck. Surg.* **131** 844–50