Contents lists available at SciVerse ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Automatic sleep staging from ventilator signals in non-invasive ventilation

Cristina C.R. Sady [a,*], Ubiratan S. Freitas [b,c], Adriana Portmann [b], Jean-François Muir [b], Christophe Letellier [c], Luis A. Aguirre [a,**]

[a] MACSIN, Laboratório de Modelagem, Análise e Controle de Sistemas Não Lineares, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte MG, Brazil
[b] ADIR, Association, Hôpital de Bois Guillaume, 76031 Rouen, France
[c] CORIA, Université de Rouen, Rouen, France

ARTICLE INFO

ABSTRACT

Non-invasive ventilation (NIV), a recognized treatment for chronic hypercapnic respiratory failure, is predominantly applied at night. Nevertheless, the quality of sleep is rarely evaluated due to the required technological complexity. A new technique for automatic sleep staging is here proposed for patients treated by NIV. This new technique only requires signals (airflow and hemoglobin oxygen saturation) available in domiciliary ventilators plus a photo-plethysmogram, a signal already managed by some ventilators. Consequently, electroencephalogram, electrooculogram, electromyogram, and electrocardiogram recordings are not needed. Cardiorespiratory features are extracted from the three selected signals and used as input to a Support Vector Machine (SVM) multi-class classifier. Two different types of sleep scoring were investigated: the first type was used to distinguish three stages (wake, REM sleep and nonREM sleep), and the second type was used to evaluate five stages (wake, REM sleep, N1, N2 and N3 stages). Patient-dependent and patient-independent classifiers were tested comparing the resulting hypnograms with those obtained from visual/manual scoring by a sleep specialist. An average accuracy of 91% (84%) was obtained with three-stage (five-stage) patient-dependent classifiers. With patient-independent classifiers, an average accuracy of 78% (62%) was obtained when three (five) sleep stages were scored. Also if the PPG-based and flow features are left out, a reduction of 4.5% (resp. 5%) in accuracy is observed for the three-stage (resp. five-stage) cases. Our results suggest that long-term sleep evaluation and nocturnal monitoring at home is feasible in patients treated by NIV. Our technique could even be integrated into ventilators.

1. Introduction

Non-invasive ventilation (NIV) is a recognized treatment for the patient with chronic respiratory failure. It is usually applied during sleep and recently it has been increasingly used at home [1,2]. The main objective of NIV is to control hypoventilation during sleep, thus improving sleep quality. As a consequence, patient's well-being should be improved. However it is quite difficult to confirm objectively this benefit, mainly because NIV may induce undesirable respiratory events disrupting sleep (patient–ventilator interactions [3,4], discomfort due to the mask, etc.) and the sleep may be also influenced by other external physical conditions or clinical reasons. In order to objectively determine whether the patient sleeps sufficiently well or to understand why sometimes the use of NIV is not beneficial, routinely assessing sleep quality during NIV is a crucial step [5].

Polysomnography (PSG) is considered as the gold standard method to evaluate sleep but this is a time consuming procedure that requires expert personnel and which is most often realized in a sleep laboratory with attended recordings. The patient spends the night at the hospital with many sensors connected to the head and other parts of the body. Several physiological signals are recorded such as the neural cortical activity (electroencephalogram, EEG), ocular activity (electrooculogram, EOG), electromyograms (EMG), pulse oximetry (SpO_2), electrocardiogram (ECG) among others cardiorespiratory signals. Special care is required in placing the electrodes and data acquisition should be monitored during the night. Although it is possible to perform home measurements, their quality is very often not sufficient to be of a practical interest [6,7]. Specialized manual intervention is also

* Corresponding author.
** Principal corresponding author at: PPGEE, Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, 31270-010 - Belo Horizonte - MG, Brazil. Tel.: +55 31 3409 4866; fax +55 31 3409 5480.
E-mail addresses: cristinasady@cpdee.ufmg.br, cristinasady@gmail.com (C.C.R. Sady), aguirre@cpdee.ufmg.br (L.A. Aguirre).

necessary for a visual sleep scoring according to standard rules and specifications [8]. All these factors make the PSG a complex and expensive procedure and limit sleep monitoring to sleep clinics and hospitals. Electrodes and sensors used to record these signals may affect sleep quality and structure. Moreover, a more simple way to evaluate sleep at reduced costs is therefore invaluable, especially for home treated NIV patients.

There are many studies that describe non-invasive ways for recording a full night of sleep that could, at least in principle, be applied at home [9–15]. Bed sensors like the ballistocardiography are non-contact measuring systems used to measure heart rate and breathing variations [16,17]. Features related to the cardiovascular system were found to be promising for classifying sleep stages. Sleep stages in adults were estimated using cardiorespiratory information obtained from an air mattress designed as a noninvasive measurement system [18]. Sleep disorders in children were diagnosed using pulse time transit (PTT) in [19,20], thus indicating that this signal gives useful information about sleep structure. The detection of REM sleep stage, which can be understood as a binary classification problem, was performed using only breathing signals in [13]. In other works, respiratory signals were used for detecting apnea events [12,14]. Actigraphy has also been widely used as a portable monitoring method for assessing sleep quality to check whether the patient is quiet during sleep [16]. Nevertheless, in none of such studies it was possible to identify more than three different sleep stages and no sufficient information was thus available to evaluate the sleep structure. Moreover, most of these methods require additional sensors.

Modern ventilators used in NIV commonly measure and record data in the ventilatory circuit (at least airflow and pressure). Several models are also equipped with a pulse oximeter to study the patient's cardiorespiratory system. It is important to investigate whether these data could allow sleep scoring and, consequently, sleep assessment of patients ventilated at home. Moreover, this sleep evaluation could potentially be much less disturbing compared to PSG, since the patient already sleeps with the ventilator and no extra sensor would be required, apart from a pulse oximeter which is already used in some cases.

The main goal of this work is to develop and to assess the performance of a sleep classification technique for patients under NIV that only uses signals available to ventilators. The key point is to consider an amplitude signal derived from pulse plethysmography (PPG). This type of signal so far has not been fully exploited in spite of its apparent potential as mentioned in [21]. In order to be able to accurately assess the sleep structure, up to five sleep stages will be scored. Sleep staging (classification) is accomplished by a multi-class support vector machine (SVM) scheme. The technique described in this paper could eventually be integrated to the ventilators themselves, thus allowing for continuous sleep monitoring at home for patients treated by NIV.

The paper is organized as follows. Section 2 describes the protocol and the preprocessing of the three signals used to train a classifier based on SVMs, which is described in Section 3. Results are presented in Section 4 and discussed in Section 5, where the main conclusions are provided.

2. Methods: data and signal processing

The data used in this work were collected during an observational study conducted at the Pulmonary and Respiratory Unit of the Rouen University Hospital (France) [22].[1] This study was authorized as part of a protocol for routine care by the Committee

to Protection People (CPP North West 1, approval dated October 16, 2009) and the patients gave informed consent to the work. The aim of this study was to evaluate patient–ventilator interactions and changes in sleep structure during the first nights after initiation to NIV. Some details of 13 patients that suffer from chronic respiratory failure and selected from the aforementioned study will be described in the sequel. Data of a 14 patient (P_5), originally included in [22], were not used in the present study.

2.1. Patients

Clinical characteristics of 13 patients are listed in Table 1. Seven patients (P_3, P_4, P_7, P_8, P_{11}, P_{12}, P_{13}) had obesity hypoventilation syndrome (OHS) associated with obstructive sleep apnea syndrome (OSAS) of variable severity (Apnea/Hypopnea Index (AHI) ranging from 34 to 142 events/hour). Four patients (P_2, P_6, P_{10}, P_{14}) had amyotrophic lateral sclerosis (ALS) among which the first three had a peripheral presentation and OSAS, the fourth having a bulbar presentation. Two other patients presented restrictive pathologies: one patient (P_1) suffers from tuberculosis sequel and another one (P_9) from kyphoscoliosis associated with OSAS.

All patients performed three full PSG. The first diagnostic PSG was carried out under spontaneous breathing (night 1) and the two others (night 2 and night 15, that is, two weeks later) were made under NIV. Only the latter two PSG recordings were considered in the present work due to our interest in investigating sleep in patients under NIV [22]. One PSG recording (night 15) is missing for patient P_{13}.

2.2. Recorded data

The recorded signals included electrophysiological signals (EEG, EMG, EOG and ECG) and ventilation physiological data (pressure, airflow, oxygen saturation, among others). The acquisition system used was a CID102-L8D (CIDELEC SA, France). Each data set was analyzed by a neurologist and sleep was coded in 30-s epochs according to the guidelines of the American Association of Sleep Medicine [8]. Five stages were distinguished: Wake (W), Rapid Eye Movements (REM) sleep, sleep stages N1 and N2 and slow wave sleep (N3).

From all signals recorded in PSG, only three were used: (i) photo-plethysmogram (PPG); (ii) hemoglobin oxygen saturation (SpO_2) and (iii) airflow. The first two were measured by a sensor placed at patient's index finger. The sampling frequency of PPG was 64 Hz and that of SpO_2 was 1 Hz. Airflow was measured by a pneumotachograph inserted in the ventilation circuit and was sampled at 128 Hz. Although the ventilator has an internal pneumotachograph like most modern ventilators, its raw measurements are not easily accessible. Unlike the oxygen saturation signal SpO_2, the PPG and airflow signals required preprocessing as described in the next two subsections.

2.3. Photo-plethysmogram preprocessing

All features related to the cardiac activity were determined by analyzing photo-plethysmogram signal (ECG signal was not used in this work). Although the use of PPG to determine heartbeat intervals for sleep classification was suggested in [21] to avoid ECG measurement and the required electrodes, it was not yet implemented. This is the key-signal used in this work which enables sleep stage classification for patients under NIV, possibly at home and without the need of any electrode.

The PPG signal was first preprocessed by applying a third order band-pass Butterworth filter with corner frequencies at 0.667 Hz and 5 Hz, which correspond to 40 and 300 beats per minute, respectively. The filtered signal is then used to detect heartbeats

---

[1] ClinicalTrials.gov identifier: NCT01255111.

Table 1
Clinical characteristics of the 13 patients (6 females and 7 males) included in the study. From [22].

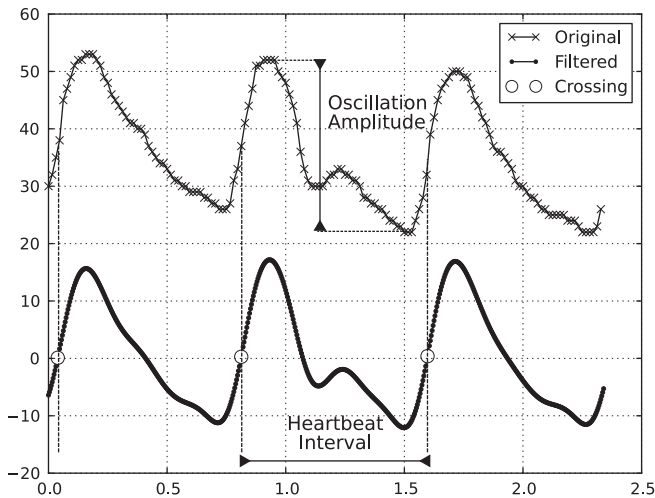| Clinical characteristics	Mean $\pm \sigma$	n
Age (years)	65.4 ± 10.9	–
Body mass index (kg m^{-2})	32.8 ± 8.62	–
Apnea/hypopnea index (per hour)	53.4 ± 41.6	–
OSAS+OHS	–	7
OSAS+ALS	–	3
ALS	–	1
Sequelæof tuberculosis	–	1
OSAS + thoracic deformation	–	1

Fig. 1. Amplitudes of the PPG oscillations are computed using maxima and minima of the original non-filtered, non-interpolated PPG signal (top). Heartbeat intervals are computed using the zero-crossings of the bandpass filtered and interpolated PPG signal (bottom).

and to compute inter-beat intervals. The relatively low sampling frequency of PPG signal (64 Hz) induces a high *quantization* noise in the heartbeat intervals, since any computed interval must be an integer multiple of the PPG sampling period, that is, 15.625 ms in the present case. In order to have a smaller quantization error for the heartbeat intervals, the filtered PPG signal was interpolated. The Fourier transform of the filtered signal was zero-padded and then transformed back to the time domain in such a way that the new signal was sampled at 256 Hz. This type of interpolation does not change the frequency content of the original signal.

The heartbeats were detected by looking for zero-crossings of the band-pass-filtered and interpolated PPG signal (Fig. 1). Only zero crossings switching from negative to positive values were considered. This proved to be numerically more efficient than the analogous procedure for detecting peaks of the PPG signal. The detected heartbeat positions were then used to compute a tachogram. The amplitude of PPG oscillations was computed using the difference between the maximum and the minimum of the original *raw* (non-filtered, non-interpolated) PPG signal inside each detected beat (Fig. 1), thus producing a time series of the fluctuations of PPG oscillations. This procedure resulted in two time series: (i) one corresponding to a tachogram (from the processed PPG signal) and (ii) one containing *amplitude* of PPG oscillations for each heartbeat (from the raw PPG signal).

2.4. Flow preprocessing

The flow signal was filtered with a third-order band-pass Butterworth filter with corner frequencies of 0.0833 Hz and 0.667 Hz, which correspond to 5 and 40 breaths per minute. One time series made of the time duration of breathing cycles and one made of the amplitudes of the airflow were generated using the same unidirectional zero-crossings method as used with PPG (Fig. 1). The amplitudes of the airflow oscillations are also rarely used because such measurements require the use of a mask during all the night, and this may disturb sleep. However, since patients under NIV are necessarily equipped with such a mask, this procedure comes, in our case, without extra cost.

3. Methods: classifier design and feature generation

The main goal of the classification phase is to develop from the data a mathematical algorithm, referred to as the classifier, that should be able to classify sleep stages. Instead of trying to build the classifier directly from the three recorded signals, it is known to be more convenient to determine some key features from the data and then to train a classifier from such information [23]. Examples of classifiers include Support Vector Machines (SVM's), Hidden Markov Models (HMM's), Artificial Neural Networks (ANN's) [24], genetic-fuzzy models [25] and others. Examples of classification features include model coefficients [26], Discrete Fourier Transform coefficients, wavelets coefficients, features based on artifact-free signals [24], and so on. In the development of classifiers, the choice of which features to use is vital, and far from trivial. The outcome of this important step will be described below, before discussing the classifier.

3.1. Feature extraction

After a thorough investigation 14 features were chosen to train the classifier. Such features can be divided into four groups depending on the data from which they were estimated (Table 2). Four features were estimated from the PPG-signal: HRV_{med}, the median of heartbeat intervals; HRV_{sd}, the interquartile interval of heartbeat intervals; amp_{med}, the median of PPG oscillation amplitudes, and amp_{sd}, the interquartile interval of PPG oscillations. Four other features were estimated from the airflow: tot_{med}, the median of respiratory cycle intervals (tot); tot_{sd}, the interquartile interval of tot; $ampf_{med}$, the median the amplitudes of flow oscillations, and $ampf_{sd}$, the interquartile interval of the amplitudes of flow oscillations. Two features were estimated from the SpO$_2$ signal for which no preprocessing was required, as mentioned in Section 2.2: sat_{med}, the median of SpO$_2$ values, and sat_{sd}: the interquartile interval of SpO$_2$ values. Median and interquartile values were used instead of average and standard

Table 2
List of features computed for each 30 s epoch.

Index	Feature	Source
1	HRV_{med}	PPG
2	HRV_{sd}	
3	amp_{med}	
4	amp_{sd}	
5	tot_{med}	Flow
6	tot_{sd}	
7	$ampf_{med}$	
8	$ampf_{sd}$	
9	sat_{med}	O$_2$ saturation
10	sat_{sd}	
11	age	Other
12	BMI	
13	EPAP	
14	IPAP	

deviation for their lower sensitivity to outliers. Finally, the following four features were obtained from "other" sources: age of the patient; body mass index (BMI); expiratory positive airway pressure (EPAP) and inspiratory positive airway pressure (IPAP).

The first three groups of features (Table 2) were computed from the three time series listed in Section 2.2 for each epoch. The last group, of four features, is patient-related and was taken to be constant over each PSG night but could vary from patient to patient or from one PSG recording to the next.

3.2. Classifier

The classifiers were trained from data that were divided into epochs of 30 s following current practice for sleep scoring [8]. Each epoch has its own set of 14 features whose indices are reported in Table 2, and a sleep stage (a class label) attributed by a neurologist. After training, the classifier automatically determine the sleep stage corresponding to the set of 14 features associated with each epoch (Fig. 2). This following procedure is based on a multi-class SVM with a Gaussian kernel (Radial Basis Function, RBF) and training was accomplished using the LIBSVM open source library [27]. Details on SVM-based classification schemes can be found in [23,28].

3.2.1. Support vector machine

SVMs provide good generalization performance on pattern classification problems in the case of separable patterns [28]. It is assumed that a training set is composed of instance-label pairs $(\mathbf{x}_i, y_i)$, $i = 1,\ldots,\ell$, where ℓ is the number of epochs used for training. Vectors $\mathbf{x}_i \in \mathbb{R}^n$ are composed of n features (Table 2) and $y_i \in \{1, -1\}$ is the class label (sleep class) of $\mathbf{x}_i$.

An input data vector $\mathbf{x} \in \mathbb{R}^n$ can be represented in an N-dimensional space ($N > n$) using a vector function $\phi : \mathbb{R}^n \longrightarrow \mathbb{R}^N$ where classification can be performed by means of a hyperplane. The SVM produces a nonlinear classification boundary in the original $\mathbb{R}^n$ space by constructing a linear hyperplane in a transformed $\mathbb{R}^N$ space [29]. This is achieved by taking the sign of the function

$$f(\mathbf{x}) = \mathbf{w} \cdot \phi(\mathbf{x}_i) + b. \tag{1}$$

Fig. 2. Sketch of the classification procedure.

Therefore, for linearly separable patterns, $\exists(\mathbf{w}, b)$ such that the inequalities

$$f(\mathbf{x}_i) \geq +1 \text{ if } y_i = 1$$
$$f(\mathbf{x}_i) \leq -1 \text{ if } y_i = -1 \tag{2}$$

are valid for all elements of the training set [30,31]. The vectors for which the *equalities* hold are the so-called *support vectors*. Clearly, by definition, a SVM is a binary classifier. The multi-class classification problem will be mentioned shortly. The computations are simplified by formulating the functional to be minimized with respect to the weight vector $\mathbf{w}$ as [28]

$$\Phi(\mathbf{w}, \epsilon) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{\ell}\epsilon_i, \tag{3}$$

where $C > 0$ is the *regularization parameter* and $\epsilon_i \geq 0$ are the *slack variables*. After some algebra, function f can be written as [28]

$$f(\mathbf{x}) = \sum_{i=1}^{\ell} y_i\alpha_i K(\mathbf{x}, \mathbf{x}_i), \tag{4}$$

where $\mathbf{x}_i$ is the ith training vector, $\alpha_i \geq 0$ are the Lagrangian multipliers and $K : (\mathbb{R}^N \times \mathbb{R}^N) \rightarrow \mathbb{R}$ the kernel function defined as the inner product $K(\mathbf{x}, \mathbf{x}_i) = \phi(\mathbf{x}) \cdot \phi(\mathbf{x}_i)$.

3.2.2. Multi-class SVM

The "one-against-one" approach [32] is implemented in [27] for multi-class classification problems. The number of binary classifiers needed to classify among k classes is $k(k-1)/2$. Each (binary) classifier is trained using data from only two different classes. Every data vector $\mathbf{x}$ is submitted to the set of binary classifiers. At the end, a vector is designated to be in the class where it received the maximum number of votes. In the case of a tie, choose the class that appears first in the storage array [27].

3.2.3. Unbalanced data

Due to the nature of sleep, the classification problem is unbalanced. The amount of time a patient stays in a sleep stage can vary significantly from one stage to another, and the number of epochs for one class can be very different from another. Such training with unbalanced data set can lead to biased classifiers. The approach chosen to deal with this problem is random subsampling [33] by which it is guaranteed that each class (sleep stage) has the same number of occurrences in the training data. This was done for all training data sets used in this work.

3.2.4. Model selection—RBF Kernel and cross-validation

The next step is the choice of the kernel and SVM parameters. The RBF kernel is a common first choice

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{\ell}\alpha_i \exp\left\{\frac{|\mathbf{x}-\mathbf{x}_i|^2}{\sigma^2}\right\}\right). \tag{5}$$

The main parameters to be chosen are the pair (C, σ) in Eqs. (3) and (5), respectively. Instead of fixing *a priori* values for these parameters, LIBSVM was used to perform a search over a 2D parameter space for the most adequate values to our problem. An important problem is to choose a good pair of values for (C, σ). This was done by testing a grid of values and cross-validating the corresponding SVMs. The one with greatest accuracy was selected. Logarithmically spaced values were used, hence $C \in \{2^{-1}, 2^{-0.25}, 2^{0.5}, 2^{1.25}, 2^2\ldots, 2^8\}$ and $\sigma \in \{2^5, 2^{4.25}, 2^{3.5}, 2^{2.75}, 2^2\ldots 2^{-2}\}$). Once the best pair (C, σ) was found, the final classifier was trained using such a pair and the whole data set.
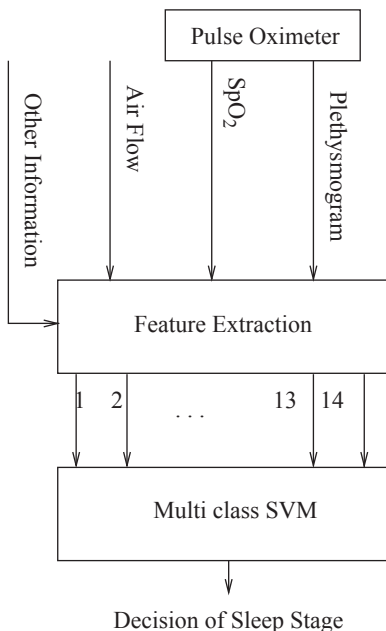
3.3. Patient-dependent classification

Ideally, an automatic sleep classifier should be able to work with data from a new patient without additional training, that is, it should be *patient-independent* (PI). The training of such a classifier will therefore require data from a representative set of patients. In this way, when data of a new patient is used, as long as his sleep patterns are somehow accounted for in the original training data, an adequate corresponding classification can be expected. On the other hand, a *patient-dependent* (PD) system is based on a classifier trained with data from a specific patient. Of course, this single patient is far less representative and unless other patients with very similar sleep patterns are in view, one should *not* expect good performance of this classifier applied to any *other* patient. In spite of its limitations, a patient-dependent classifier is also useful. First, it serves as a proof for the feasibility of the technique and also serves as an upper bound for the accuracy one may expect to obtain using a patient-independent classification. Second, a classifier can be trained for a specific patient whose sleep is to be monitored through his/her private ventilator.

Sleep classifiers trained with data from a single patient were analyzed (i) with data from a single night and, (ii) with data from two nights, as discussed in the sequel.

3.3.1. One-night classification

One SVM multi-class classifier was trained for each night of each Patient (two classifiers for each patient – with the exception of P_{13} for whom data corresponding to a single night are available). Two scenarios were evaluated: a five-classes classifier using W, REM, N1, N2 and N3 sleep stages and a three-class classifier using W, REM and non-REM (in which N1, N2 and N3 stages were grouped) sleep stages. Thus 25 classifiers were trained for the three-class type and 25 for the five-class type.

Each classifier was trained with typically 12% of the available epochs and was subsequently tested on all the data. The SVM classification was considered as being correct when it coincided with that of the neurologist. Hence, for a given patient and a full-night recording with M epochs, the percentage accuracy is given by

$$A_{cc} = \frac{n_a}{M} \times 100\%, \tag{6}$$

where n_a indicates the number of correctly classified epochs. A_{cc} is an estimate of the classifier overall accuracy. For the one-night classification study, in the end there were 25 classifiers with the respective A_{cc}. The mean and standard deviation of such a sample are reported.

3.3.2. Two-night classification

The same procedure mentioned above was followed for the two-night records of each patient for which two nights were recorded (they are only 12). The epochs from both nights were grouped to form a single data set of greater generality and a single classifier for each patient was trained. Tests were performed using all data from the two nights. Training data were always balanced. Once again the percentage accuracy A_{cc} in (6) was computed for each patient. The mean and standard deviation of these 12 values are also reported.

3.4. Patient-independent classification

Since features 11–14 (Table 2) vary from one patient to the other, they are useful in training patient-independent classifier. For the construction of a such a classifier, epochs of the 13 patients were grouped to form the training and test data. In fact, two classifiers were trained: one to discriminate between three classes

and one to classify among five sleep stages. The 14 features (Table 2) were extracted for each patient separately. The features for all patients were then arranged in a feature array. Subsequently, 12 different balanced training data sets of the same size were randomly chosen from the feature array. This number was chosen in order to make performance comparable to the case of two-night *patient-dependent* classification, where 12 classifiers were obtained, one for each patient. All the data in the feature array were used for testing. After the percentage accuracy A_{cc} was computed for each of these 12 realizations, mean and standard deviation were computed and reported.

3.5. Feature pertinence test

Two sets of features used in this work are not standard in automatic sleep classification. They are the features associated with the amplitudes of PPG oscillations and flow oscillations. In order to test their influence on the classifier performance, specific tests were performed in which such features were excluded for training different classifiers. The performance of these classifiers was then compared with that obtained with the full set of features. Three cases were studied: (i) only the features related to PPG amplitude were excluded; (ii) only the features related to flow amplitude were excluded, and (iii) both the features related to PPG amplitude and to flow amplitude were excluded. To evaluate the impact of such changes, a non-parametric paired difference test (Wilcoxon) was used to compare accuracies before and after removing the features.

4. Results

In Table 3 (Table 4) are reported the results for five-stage (three-stage) classification for both patient-dependent and independent scenario. The hypnogram produced by a five-stage one-night patient-dependent classifier is shown in Fig. 3. The accuracy rate in this case is about 92%. It can be observed that the automatically scored hypnogram is more fragmented than the one produced by the sleep specialist. Unlike the visual/manual scoring, where the continuity from one epoch to the other is subjectively taken into account, the SVM classifies each epoch independently.

Table 5 (Table 6) shows the results that concern the automatic sleep classification in five (three) stages. The first third part of the table corresponds to the classification without features associated with plethysmogram and flow amplitude (features 3, 4, 7 and 8 in

Table 3
Results for 5 sleep stages (classes). $\overline{A}_{cc}$ stands for the average value of A_{cc} defined in (6). $\overline{k}$ stands for the Mean Cohen's Kappa Coefficient. PD: patient-dependent; PI: patient-independent, and in parenthesis is the number of nights recorded.

P(N)	$\overline{A}_{cc} \pm \sigma(\%)$	$\overline{k} \pm \sigma$	Features
PD(1)	84 ± 5.0	0.78 ± 0.06	1–10
PD(2)	73 ± 5.2	0.64 ± 0.07	1–10
PI(2)	62 ± 0.85	0.51 ± 0.01	1–14

Table 4
Results for 3 sleep stages (classes). See caption to Table 3 for definitions.

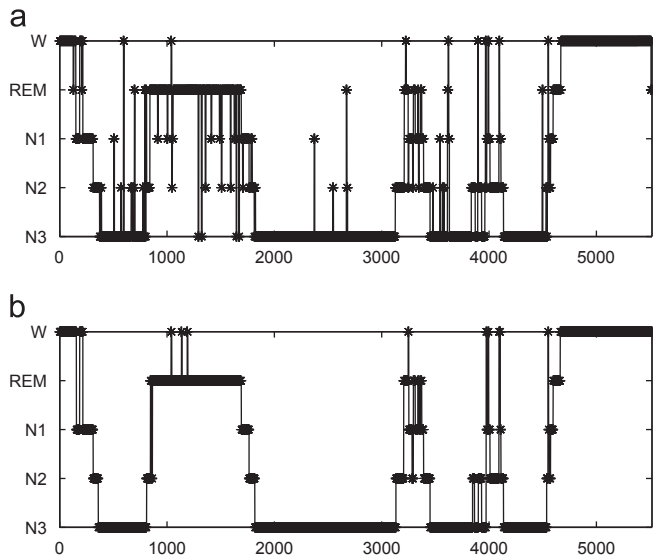| P(N)	$\overline{A}_{cc} \pm \sigma(\%)$	$\overline{k} \pm \sigma$	Features
PD(1)	91 ± 2.5	0.83 ± 0.05	1–10
PD(2)	84 ± 3.5	0.71 ± 0.06	1–10
PI(2)	78 ± 0.74	0.64 ± 0.01	1–14

Fig. 3. Hypnogram produced by a patient-dependent classifier (a) according to the one-night scenario and by a neurologist (b). Case of patient P_{12}. In this case $\overline{A}_{cc} \approx 92\%$.

Table 5
Results for 5 sleep stages (classes) leaving out specific features. See caption of Table 3 for definitions.

P(N)	$\overline{A}_{cc} \pm \sigma(\%)$	$\overline{k} \pm \sigma$	Features
PD(1)	80 ± 5.5	0.73 ± 0.07	1, 2, 5, 6, 9, 10
PD(2)	67 ± 5.6	0.76 ± 0.08	1, 2, 5, 6, 9, 10
PI(2)	57 ± 0.96	0.44 ± 0.03	1, 2, 5, 6, 9–14
PD(1)	83 ± 5.8	0.76 ± 0.07	1, 2, 5–10
PD(2)	71 ± 5.6	0.61 ± 0.06	1, 2, 5–10
PI(2)	59 ± 0.68	0.47 ± 0.01	1, 2, 5–14
PD(1)	83 ± 4.2	0.76 ± 0.05	1 to 6, 9, 10
PD(2)	72 ± 6.2	0.62 ± 0.07	1 to 6, 9, 10
PI(2)	60 ± 0.63	0.49 ± 0.01	1 to 6, 9–14

Table 6
Results for 3 sleep stages (classes) leaving out specific features. See caption to Table 3 for definitions.

P(N)	$\overline{A}_{cc} \pm \sigma(\%)$	$\overline{k} \pm \sigma$	Features
PD(1)	87 ± 4.3	0.76 ± 0.08	1, 2, 5, 6, 9, 10
PD(2)	79 ± 4.0	0.62 ± 0.08	1, 2, 5, 6, 9, 10
PI(2)	74 ± 1.4	0.56 ± 0.02	1, 2, 5, 6, 9–14
PD(1)	90 ± 3.3	0.80 ± 0.06	1, 2, 5–10
PD(2)	82 ± 4.3	0.67 ± 0.07	1, 2, 5–10
PI(2)	76 ± 0.83	0.60 ± 0.01	1, 2, 5–14
PD(1)	90 ± 2.9	0.81 ± 0.05	1 to 6, 9, 10
PD(2)	83 ± 3.6	0.68 ± 0.07	1 to 6, 9, 10
PI(2)	77 ± 0.61	0.61 ± 0.01	1 to 6, 9–14

Table 2). In the second part, the features associated with plethysmogram amplitude (features 3 and 4 in Table 2) were *not* used. In the last (third) part the features associated with flow amplitude (features 7 and 8 in Table 2) were left out.

In order to verify if there are significant differences between results shown in Table 5 (resp. Table 6) compared to Table 3 (resp. Table 4) a Wilcoxon paired test was performed. Results are shown in Table 7 and show that the majority of the differences are statistically significant ($p \leq 0.05$).

Table 7
Wilcoxon paired test for statistical significance of performance differences when specific features are not used. Asterisks indicate values that are *not* statistically significant.

P(N)	p-value (5 stages)	p-value (3 stages)
PD(1)	1.2×10^{-5}	1.8×10^{-5}
PD(2)	4.88×10^{-4}	4.88×10^{-4}
PI(2)	4.88×10^{-4}	4.88×10^{-4}
PD(1)	0.0397	0.0068
PD(2)	0.0640*	9.7656×10^{-4}
PI(2)	$4.88 \cdot 10^{-4}$	4.8828×10^{-4}
PD(1)	0.0032	0.0042
PD(2)	0.0522*	0.0640*
PI(2)	9.77×10^{-4}	4.88×10^{-4}

5. Discussion and conclusion

Cardiorespiratory signals were used in [9] to verify if measurements obtained from ECG and respiration signals could provide a classification of the wakefulness, REM and non-REM sleep stages. The main goal was to increase the use of cardiorespiratory ambulatory care systems for the detection of sleep-disordered breathing. In that study, a group of 37 patients with obstructive sleep apnea was considered. For each of them, 27 features were extracted from three signals: the RR interval series, electrocardiogram-derived respiration and respiratory effort. Using a Quadratic Discriminant Classifier the attained accuracy was 79% ($\kappa = 0.56$) for the patient-dependent system and 67% ($\kappa = 0.32$) for patient-independent classification. The results obtained in the present paper for a similar classification problem but with different types of patients (Table 4) reveal greater accuracy 91% ($\kappa = 0.83$) and 78% ($\kappa = 0.64$).

Cardiorespiratory measures were also used in [10] to discriminate between the wakefulness, REM and quiet sleep in 35 healthy infants. Seven cardiorespiratory features were used, leading to the following accuracy rates: 80.0% when only respiratory-based features were used, 82.0% using only cardiovascular-based features and 84.8% when both sets of features were employed. ECG and respiratory effort signals (chest plethysmography) of six healthy subjects were used in [11] to discriminate only between wakefulness and sleep (a binary classification problem). Features consisted of the logarithm of power spectral density of raw ECG and respiratory signals. Using artificial neural network classifiers, the following accuracy was reached: 95.4% for patient-dependent cases and 85.3% for patient-independent. In comparing the results presented in Tables 3 and 4 to those in [10,11], it should be remembered that in the present work sleep patterns are classified into 3 and 5 stages using an electrode-free technique.

The inclusion of certain features resulted in statistically significant (Table 7) improvement in classification accuracy. Average improvement was about 1% (resp. 1.7%) when flow-related features were included; 1.5% (resp. 2%) when PPG-related features were included; and 4.5% (resp. 5%) when both classes of features were included for three-stage (resp. five-stage) classification problems. Both sets of features are relevant although PPG-related features being slightly more relevant than flow-related features. Also, the greatest advantage is attained including both sets of features. These results suggest that such features contain information about sleep which cannot be retrieved from other features and, consequently, could lead to more efficient automatic sleep classifiers which do not require electrodes.

The reported results show that it is possible to develop an appropriate procedure for the automatic classification of sleep stages from signals readily available to a ventilator. This opens up

the possibility for a better assessment of the quality of sleep under NIV treatment since it can allow for evaluating sleep at home on a long term basis without the use of electrodes nor loading patients with extra sensors. In the future, it could be even possible for the ventilators themselves to deliver a sleep report together with standard compliance and diagnostic reports such machines already produce. Another possibility would be for the ventilator to use information about the patient's sleep to adapt its settings and to thus improve the quality of treatment.

Summary

Objective: The paper presents a system for automatic sleep scoring targeted at patients treated by non-invasive ventilation (NIV). *Methods*: The method uses only signals available to the ventilator and no other sensor are required apart from the ones already included in ventilator and a pulse oximeter. Particularly no electrodes (e.g. EEG or ECG) are used. Cardiorespiratory (rhythm and amplitude) features are extracted from pulse plethysmography (PPG), airflow and oxygen saturation, which constitutes an important contribution. Classification is achieved by a scheme based on Support Vector Machine (SVM). Three-stage (wake, REM and non-REM) and five-stage (wake, REM, N1, N2 and N3) classification schemes were investigated. Also, patient-dependent and patient-independent classifiers were tested. *Results*: Obtained accuracies for 3 and 5 stage classification are: 91% and 84% with patient-dependent schemes and 78% and 62% with patient-independent schemes, respectively. Such results are quite competitive compared to other electrode-based systems reported in the literature. Also if the PPG-based and flow features are left out, a reduction of 4.5% (resp. 5%) in accuracy is observed for the three-stage (resp. five-stage) cases. *Conclusion*: These results suggest that long-term sleep evaluation and nocturnal monitoring at home could be feasible for patients treated by NIV. It maybe even integrated to the ventilators themselves.

Conflict of interest statement

None declared.

Acknowledgment

References

[1] M. Elliott, A. Simonds, M. Carroll, J. Wedzicha, M. Branthwaite, Domiciliary nocturnal nasal intermittent positive pressure ventilation in hypercapnic respiratory failure due to chronic obstructive lung disease: effects on sleep and quality of life, Thorax 47 (5) (1992) 342–348.

[2] F. Barbé, M.A. Quera-Salva, J. de Lattre, P. Gajdos, A.G. Agustí, Long-term effects of nasal intermittent positive-pressure ventilation on pulmonary function and sleep architecture in patients with neuromuscular diseases, Chest 110 (5) (1996) 1179–1183.

[3] H. Rabarimanantsoa, L. Achour, C. Letellier, A. Cuvelier, J.-F. Muir, Recurrence plots and Shannon entropy for a dynamical analysis of asynchronisms in noninvasive mechanical ventilation, Chaos 17 (1) (2007) 013115-1–013115-10.

[4] L. Achour, C. Letellier, A. Cuvelier, E. VÈrin, J.-F. Muir, Asynchrony and cyclic variability in pressure support noninvasive ventilation, Comput. Biol. Med. 37 (9) (2007) 1308–1320.

[5] M.W. Elliott, Non-invasive ventilation during sleep: time to define new tools in the systematic evaluation of the technique, Thorax 66 (1) (2011) 82–84.

[6] F. Portier, A. Portmann, P. Czernichow, L. Vascaut, E. Devin, D. Benhamou, A. Cuvelier, J.-F. Muir, Evaluation of home versus laboratory polysomnography in the diagnosis of sleep apnea syndrome, Am. J. Respir. Crit. Care Med. 162 (2000) 814–818.

[7] F. Gagnadoux, N. Pelletier-Fleury, C. Philippe, D. Rakotonanahary, B. Fleury, Home unattended vs hospital telemonitored polysomnography in suspected obstructive sleep apnea syndrome, Chest 121 (3) (2002) 753–758.

[8] C. Iber, S. Ancoli-Israel, A. Chesson, S.F. Quan, The AASM Manual for Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Technical Report, American Academy of Sleep Medicine, 2007.

[9] S.J. Redmond, C. Heneghan, Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea, IEEE Trans. Biomed. Eng. 53 (3) (2006) 485–496.

[10] R.M. Harper, V.L. Schechtman, K.A. Kluge, Machine classification of infant sleep state using cardiorespiratory measures, Electroencephalogr. Clin. Neurophysiol. 67 (1987) 379–387.

[11] W. Karlen, C. Mattiussi, D. Floreano, Sleep and wake classification with ECG and respiratory effort signals, IEEE Trans. Biomed. Circuits Syst. 3 (2) (2009) 71–78.

[12] P. Várady, T. Micsik, S. Benedek, Z. Benyó, A novel method for the detection of apnea and hypopnea events in respiration signals, IEEE Trans. Biomed. Eng. 49 (9) (2002) 936–942.

[13] G.S. Chung, B.H. Choi, J.-S. Lee, J.S. Lee, D.-U. Jeong, K.S. Park, REM sleep estimation only using respiratory dynamics, Physiol. Meas. 30 (12) (2009) 1327–1340.

[14] J.V. Marcos, R. Hornero, D. Alvarez, F. del Campo, C. Zamarrón, Assessment of four statistical pattern recognition techniques to assist in obstructive sleep apnoea diagnosis from nocturnal oximetry, Med. Eng. Phys. 31 (8) (2009) 971–978.

[15] P.I. Terrill, S.J. Wilson, S. Suresh, D.M. Cooper, C. Dakin, Attractor structure discriminates sleep states: recurrence plot analysis applied to infant breathing patterns, IEEE Trans. Biomed. Eng. 57 (5) (2010) 1108–1116.

[16] J.M. Kortelainen, M.O. Mendez, A.M. Bianchi, M. Matteucci, S. Cerutti, Sleep staging based on signals acquired through bed sensor, IEEE Trans. Inf. Technol. Biomed. 14 (3) (2010) 776–785.

[17] M.O. Mendez, M. Migliorini, J.M. Kortelainen, D. Nistico, E. Arce-Santana, S. Cerutti, A.M. Bianchi, Evaluation of the sleep quality based on bed sensor signals: time-variant analysis, in: 32nd Annual International Conference of the IEEE EMBS, 2010, pp. 3994–3997.

[18] T. Watanabe, K. Watanabe, Noncontact method for sleep stages estimation, IEEE Trans. Biomed. Eng. 51 (10) (2004) 1735–1748.

[19] R.P. Smith, J. Argod, J.-L. Pépin, P.A. Lévy, Pulse transit time: an appraisal of potential clinical applications, Thorax 54 (1999) 452–458.

[20] J.-L. Pépin, N. Delavie, I. Pin, C. Deschaux, J. Argod, M. Bost, P. Lévy, Pulse transit time improves detection of sleep respiratory events and microarousals in children, Chest 127 (2005) 722–730.

[21] C.-P. Chua, J. Garvey, S. Redmond, C. Heneghan, W.T. McNicholas, Towards automated sleep state estimation using a Holter-oximeter, in: Proceedings of the 29th Annual International—Conference of the IEEE EMBS, 2007, pp. 3998–4001.

[22] R. Naeck, Evaluation de l'adaptation à la ventilation non invasive chez patients atteints d'insuffisance respiratoire chronique., Ph.D. Thesis, Université de Rouen, France, 2011.

[23] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, NY, 2006.

[24] S. Charbonnier, L. Zoubek, S. Lesecq, F. Chapotot, Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging, Comput. Biol. Med. 41 (6) (2011) 380–389.

[25] H.G. Jo, J.Y. Park, C.K. Lee, S.K. An, S.K. Yoo, Genetic fuzzy classifier for sleep stage identification, Comput. Biol. Med. 40 (7) (2010) 629–634.

[26] C. Lainscsek, P. Rowat, L. Schettino, D. Lee, D. Song, C. Letellier, H. Poizner, Finger tapping movements of Parkinsons disease patients automatically rated using nonlinear delay differential equations, Chaos 22 (1) (2012) 013119-1–013119-13.

[27] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (2011) 27–1-27:27.

[28] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, New Jersey, 1999.

[29] R. Debnath, N. Takahide, H. Takahashi, A decision based one-against-one method for multi-class support vector machine, Pattern Anal. Appl. 7 (2) (2004) 164–175.

[30] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learning 20 (1995) 273–297.

[31] V.N. Vapnik, An overview of statistical learning theory, IEEE Trans. Neural Networks 10 (5) (1999) 988–999.

[32] B. Liu, Z. Hao, E.C.C. Tsang, Nesting one-against-one algorithm based on SVMs for pattern classification, IEEE Trans. Neural Networks 19 (12) (2008) 2044–2052.

[33] C.L. Castro, A.P. Braga, Aprendizado supervisionado com conjuntos de dados desbalanceados, Rev. Controle Autom. 22 (5) (2011) 441–466.