IOP Publishing Physiol. Meas. 39 (2018) 124008 (14pp) https://doi.org/10.1088/1361-6579/aaf5d4

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

RECEIVED
31 July 2018

REVISED
26 November 2018

ACCEPTED FOR PUBLICATION
4 December 2018

PUBLISHED
24 December 2018

PAPER

Automatic sleep stages classification using respiratory, heart rate and movement signals

Maksym Gaiduk[1,3], Thomas Penzel[2], Juan Antonio Ortega[3] and Ralf Seepold[1,4]

[1] HTWG Konstanz, Konstanz, Germany
[2] Interdisciplinary Center for Sleep Medicine of Charité Clinic, Berlin, Germany
[3] University of Seville, Seville, Spain
[4] Department of Information and Internet Technology, Sechenov University, Moscow, Russia

E-mail: maksym.gaiduk@htwg-konstanz.de

Keywords: multinomial logistic regression, sleep stage classification, non-invasive sleep study, heart rate, respiration, body movement, sleep study

Abstract

Objective: This paper presents an algorithm for non-invasive sleep stage identification using respiratory, heart rate and movement signals. The algorithm is part of a system suitable for long-term monitoring in a home environment, which should support experts analysing sleep. *Approach*: As there is a strong correlation between bio-vital signals and sleep stages, multinomial logistic regression was chosen for categorical distribution of sleep stages. Several derived parameters of three signals (respiratory, heart rate and movement) are input for the proposed method. Sleep recordings of five subjects were used for the training of a machine learning model and 30 overnight recordings collected from 30 individuals with about 27 000 epochs of 30 s intervals each were evaluated. *Main results*: The achieved rate of accuracy is 72% for Wake, NREM, REM (with Cohen's kappa value 0.67) and 58% for Wake, Light (N1 and N2), Deep (N3) and REM stages (Cohen's kappa is 0.50). Our approach has confirmed the potential of this method and disclosed several ways for its improvement. *Significance*: The results indicate that respiratory, heart rate and movement signals can be used for sleep studies with a reasonable level of accuracy. These inputs can be obtained in a non-invasive way applying it in a home environment. The proposed system introduces a convenient approach for a long-term monitoring system which could support sleep laboratories. The algorithm which was developed allows for an easy adjustment of input parameters that depend on available signals and for this reason could also be used with various hardware systems.

Introduction

The human body is more fragile than people think. In order to survive, it requires sleep just as much as food, water, or oxygen. This is a basic principle of human physiology that has been borne out by thousands of research studies.

In general, sleep is a state where our bodies and minds rest and rejuvenate (Spriggs 2009). It is obligatory for our normal physiological, mental and emotional functioning during awake hours. The belief that it is possible to have just a couple of hours of sleep a night over a long period of time without suffering any negative consequences is a common misconception (National Heart Lung and Blood Institute (NHLBI) 2011). It is categorically beyond doubt that when sleep contains even slight abnormalities, the aftermath can lead to physical illness, psychological problems or an untimely death (Lee-Chiong *et al* 2012).

A popular misconception is that adults have to sleep at least 7–8 h every night to be rejuvenated properly, while children require far more hours of sleep (National Heart Lung and Blood Institute (NHLBI) 2011). However, this is only standard recommended advice; sleep requirements are individual for every person (National Heart Lung and Blood Institute (NHLBI) 2011). In addition, getting many hours of sleep does not always guarantee a healthy and rested state, because the crucial point here is not the quantity, but the quality (National Heart Lung and Blood Institute (NHLBI) 2011).

In order to get reliable data on the quality of a person's sleep, the sleep stages and their sequence and durations have to be analysed. Usually such studies are executed in sleep laboratories. The established standard procedure for sleep stage quantitative evaluation is the overnight polysomnography (PSG) following the guidelines of the American Academy of Sleep Medicine (AASM) (Chokroverty and Thomas 2013).

Motivation

The execution of PSG generally means that during sleep a human body will be connected to at least 22 laboratory wire attachments (Chokroverty and Thomas 2013), which explains why the PSG is quite a costly and time-consuming procedure (Ettinger and Feldman 2010). Moreover, being in the sleep laboratory is always unfamiliar and disturbing for the subjects, so there could be some discomfort because of the electrodes and the issue of movement limitation. All of these conditions would impact the person's regular sleep behaviour (Le Bon *et al* 2001). It is safe to say that the sleep during the PSG monitoring often has differences compared to sleeping at home.

Moreover, the fact, that different sleep stages have an appreciable effect on heart rate, breathing and movement (Hayet and Slim 2012, Kurihara and Watanabe 2012, Long *et al* 2014), provides good reason to combine these parameters to develop the sleep stage classification algorithm.

Furthermore, there are a limited number of sleep medicine specialists who can provide healthcare and support sleep study (Singh *et al* 2015). Additionally, in order to conduct sleep studies for all the patients who require them without incurring long waiting periods, the number of sleep laboratories and experts would have to be immense. At the same time one of the most important health problems that humanity faces lies in the terribly high amount of undiagnosed sleep disorders (Lee-Chiong *et al* 2012). Besides that, PSG-based sleep scoring is a very expensive procedure (Berry and Wagner 2015). This is why low-cost, non-invasive, home-based diagnostic systems for sleep study, in particular for categorical distribution of sleep stage parameters, would provide substantial additional support for stationary sleep laboratories.

Objective

This project aims to develop a new approach to the classification of sleep stages. The novelty of this approach would be the set of parameters (including the derived parameters) used for the sleep stage recognition. Whereas traditional testing (e.g. PSG) requires examining many parameters, our aim is to develop a system which uses as few appropriate parameters as possible for input. Choosing the parameters would require serious consideration of the two main points:

- the chosen parameters must differ by behaviour patterns during the particular sleep stages;
- these patterns have to be represented by mathematical equations and then converted to algorithms.

As a result, a software application to implement the sleep phase analysis and automatic evaluation will be designed. The foundation for the sleep stages recognition should be provided by mathematical description and transformed into programming code.

The developed algorithm should work with the bio vital parameters that must be available to be collected by using non-invasive methods. In an effort to get recordings of sleep states that are natural and genuine, the experimental subjects must be provided with a feeling of sleeping in their usual environment (Gaiduk *et al* 2018).

The system for sleep phase classification described in this article should be designed to classify the sleep stages using the signals from PSG recordings.

State of the art

Research on the topic of non-invasive sleep stage classification methods has been discussed in a number of scientific publications (Hayet and Slim 2012, Kambayashi and Hagiwara 2012, Kurihara and Watanabe 2012, Long *et al* 2014, Tataraidze *et al* 2015, Penzel *et al* 2016) around the globe. This case study is focused on just those few publications that provide the relevant content for this work. Research studies that were conducted by using heart rate, body movement or respiration as vital signals for sleep classification and sleep cycles investigation are of particular value. The scientific papers listed below mostly dealt with these complex issues, so only the parts significant to this project will be described.

The ECG itself, heart rate and heart rate variability (HRV) can be used to analyse and detect sleep stages (Penzel *et al* 2016).

In Aktaruzzaman *et al* (2017) wrist and chest actigraphy were compared in combination with HRV for sleep classification. In this study a support vector machine (SVM) was used for the automatic identification of sleep and wake stages. The achieved accuracies for the group of healthy adult 18 subjects are nearly equal—78% for chest and 77% for wrist actigraphy jointly with HRV.

One study Hayet and Slim (2012) describes a method for the sleep and wake stages classification based on just the ECG signals and a neural network algorithm. To develop and analyse the obtained results, they used 16 PSG

Table 1. Characteristics of REM sleep (Kurihara and Watanabe 2012).

N	Characteristics
1	Brainwaves similar to those in NREM1 and WAKE are found
2	Decreasing of incidence ratios of delta wave and spindle
3	Disappearing of the tension of anti-gravity muscles
4	Appearing of rapid eye movement
5	Increasing of frequency of heartbeat and respiration while simultaneously becoming less rhythmic, blood pressure becomes high
6	REM sleep occurs once every 90–100 min on average (adults)
7	Concentration of body movement before and after REM stage

recordings from the MIT-BIH polysomnographic database. First of all, to pre-process the data they built the RR series from the QRS annotation files to help calculate the HRV. At the next stage, the extreme learning machine (ELM) neural network algorithm with a single hidden layer was used to classify the wake and sleep phase (Guang-Bin *et al* 2006). In order to distinguish between wake and sleep phases, the algorithm was trained on two scenarios:

- subject-specific classification;
- subject-independent classification.

In the first scenario, the time periods were selected from every single night's recording and fed into the ELM algorithm. Two-thirds of the whole sample recording was used for that. Next, one-third of the night recordings were used to test the algorithm's performance. As for the subject-independent classification, the researchers used epochs from all subjects to train the ELM. After training the algorithm, a recognition rate of approximately 90 percent was achieved. Predictably, the subject-specific rates are a bit better (93.33%) than the subject-independent ones (90.03%).

An automatic algorithm to determine REM sleep on the basis of the autonomic activities reflected in heart rate variations was developed in Yoon *et al* (2017). The HRV was calculated from ECG using *R–R* intervals and an adaptive threshold was applied to ascertain the REM stage. The average accuracy was 87% for the evaluation of 25 healthy and OSA subjects.

In another scientific study the concept of using data about body movement for sleep cycles estimation was investigated (Kambayashi and Hagiwara 2012). The experiment was conducted by studying 16 healthy people who slept through the night with electrodes attached to their bodies. A detection device (NapVIEW) with a NaPiOn infrared motion sensor was used to detect body movements. It was placed on the bedside 50 cm from the subject's head. Along with this, the body movement density (BMD) was defined as the amount of body movements per time unit as an index of sleep transition. In sum, this study showed that BMD cycle is strongly linked to sleep cycle. And to evaluate sleep cycles we can use body movement data instead of brain activity or other parameters. Moreover, it was concluded that a BMD cycle is less affected by individual variations and it is therefore a better index than the absolute value of BMD.

One particular research study Kurihara and Watanabe (2012) describes another two algorithms and methods for sleep stages estimation. Their classification experiment was conducted with the help of ten healthy adults with a mean age of 22.2 years. The main goal of this project was to find a non-invasive method to obtain the vital parameters and classify the subjects' sleep stages into the categories WAKE, REM and NREM1—NREM4. To achieve this, the scientists developed an air mattress with a highly responsive pressure sensor. It allows the measuring of respiration, heartbeat and body movements. They created an algorithm based on the idea that all sleep stages have certain characteristics regarding body activity and that body signals behave differently in the different sleep stages. In tables 1 and 2 the different characteristics of REM and NREM sleep, which are important for this article's purposes because of using similar features, are listed.

In the end, this experiment got 51.6% agreement with the classification results done by sleep experts within six stages (WAKE, REM, NREM1,2,3,4) and 77.5% within three stages (WAKE, REM, NREM).

Furthermore, there are various studies (e.g. Long *et al* (2014) and Tataraidze *et al* (2015)) where respiration behaviour during sleep hours has been examined. One study Chung *et al* (2007) tried to classify REM sleep based on respiration rates. The researchers aimed to develop a new user-friendly method for the sleep-wake-monitoring called 'bed actigraphy' (BACT), which was presented in an earlier study (Choi *et al* 2007). To achieve the goal, a system with four load cells fixed at the bottom of the bed legs was built. In order to assess the methods, the researchers studied three healthy participants—two males and one female between 27 and 32 years old. Their sleep cycles were analysed and classified in a specialized sleep center with PSG, according to the scoring manual of Rechtschaffen and Kales. There are peaks between expiration and inspiration during a fixed period of time called *window size*. It was determined to further compute the average respiration rate, after which the maximum

Table 2. Characteristics of NREM sleep (Kurihara and Watanabe 2012).

N	Characteristics
1	The incidence ratio of delta waves is more frequent with the deeper sleep
2	In the NREM2 stage spindle waves are recognized
3	When sleep deepens from the WAKE stage, body movements become smaller and less frequent
4	The deeper the sleep, the less frequent the heart rate
5	NREM1 occasionally is found after NREM3, NREM4, or REM stages with large body movement

peak inside this window was outlined as a peak point. The comparison clearly illustrates that the respiration rate is more irregular and increased during REM sleep stage than in other categories. Using an appropriate threshold level, the researchers determined the REM stages for all three participants. The percentages of correct predictions for these stages were 87.9%, 69.1% and 69.7%. It was concluded that respiration is a suitable parameter to classify REM and NREM sleep.

All the studies explored above proved that the examination of the bio vital parameters, particularly respiration and heartbeat, and additionally body movement during sleep has a potential to evaluate several sleep stages. However, the scientists tried to obtain multiple patterns of HRV and respiration during special sleep phases using different mathematical approaches. Moreover, one study has shown that the occurrence of body movement during sleep provides the ability to make conclusions about the sleep cycles.

To evaluate the sleep stages, special sleep phases and sleep cycles, simple mathematical algorithms along with a more complicated neural network algorithm were used. The resulting findings were promising. But they have also indicated that the estimation of sleep stages without recording brain activity and eye movements (as is done in PSG) is too imprecise. In addition, a perfect match with estimated results of the experts is not possible because the judgment standards of the *R–K* method include ambiguities. Thus, it can lead to different sleep stage classifications in some parts of sleep recordings due to the subjective interpretations of particular sleep-stage evaluators (Kurihara *et al* 2010).

Nevertheless, it is still vital to develop a non-invasive and as accurate as possible solution for sleep stage recording and estimation, in order to support the complex PSG procedure in the future.

Statistical methods

To calculate the strength of correlation between one dependent variable and a series of other changing independent variables, regression analysis is used as a statistical measure. It is known that regression is the most widely used statistical evaluation and its applications occur in almost every field (Douglas *et al* 2012). In this method, the dependent variable, which has to be analysed or predicted, is usually denoted by Y. In order to do so, one or more other independent variables related to variable Y and denoted by $X_1, X_2, ..., X_n$ must be estimated (Alexander Von and Christof 1998).

Though, it is not necessary for these independent variables to be statistically independent of each other (Belsley 1991). Therefore, the information about the dependent variable carried by the independent variables has to be presented through a mathematical function, which will define the existing relationship as accurately as possible (Alexander Von and Christof 1998). Such a mathematical function must be found.

When there is a need to analyse and adjust a nominal outcome for multiple independent variables, the multinomial logistic regression (MLR) is used. It is a special form of logistic regression that falls under the multiple regressions category. The usage of MLR presumes that one of the categories has to be designated as the reference. Which category is chosen as the reference will affect the procedure for reporting the results, but not the mathematical answer itself. The MLR model compares all categories of the dependent variable to the reference category for each of the independent variables. The number of such comparisons is equal to one minus the number of categories of the dependent variable. Basically, the MLR models the logit (logarithm of an odd) of being in one of the outcome categories compared to being in the reference group (Katz 2011).

In this case, the odds ratio is equal to the antilogarithm of the logistic regression coefficient. In multinomial regression, the odds ratio presents the information on how the probability of being in one category versus being in the reference group is influenced by the change of independent variable. If a value of the odds ratio is greater than 1, then as an interval-independent variable increases, the value of probability increases too. In other case, if the odds ratio is less than 1, then there is an opposite scenario in place—the value of probability decreases when an interval-independent variable increases (Katz 2011).

The outcomes of categorical dependent random variables denoted as $Y \in \{0, 1, 2 \ldots k\}$, are modelled by MLR. According to the model, a conditional mean of the dependent categorical variables is the logistic function of an affine combination of independent variables, which is usually denoted as x and is defined as

$$E[Y|x] = \sigma\left(c^T x\right)$$

for some unknown vector of c coefficients, and where

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

is the logistic function. Then, MLR finds the vector of coefficients c that maximizes the probability of observations.

In turn, maximizing the probability, which is defined as

$$\prod_{i=1}^{n} Pr(Y = Y_i | X_i)$$

is equal to maximizing the log-probability, which is estimated as

$$\sum_{i=1}^{n} \log Pr(Y = Y_i | X_i).$$

This leads to the simplified function for maximizing the probability:

$$l(c) = -\sum_{i=1}^{n} \log\left(1 + e^{-1^{y_i} * c^T x_i}\right).$$

The standard error of coefficient i is calculated as

$$se(c_i) = ((X^T A X) - 1)_{ii},$$

where the Hessian of $H = -X^T A X$ and $A = \text{diag}(a_1, \ldots, a_n)$ is the diagonal matrix with $a_i = \sigma(c^T x) * \sigma(c^T x)$ (Garson 2016).

Finally, the so-called *Wald z-test* defined as

$$z_i = \frac{c_i}{se(c_i)}$$

is necessary to test the significance of predictors in the logistic regression. In this case, the Wald p-value for coefficient i gives the probability of seeing a value at least as extreme as the one observed, considering that the null hypothesis ($c_i = 0$) is true.

The formula for Wald p-value of i coefficient is

$$p_i = Pr(|Z| \geqslant |z_i|) = 2 * (1 - F(|z_i|))$$

where F is the cumulative density function of a standard normal distribution (Garson 2016).

Methods

The existing research studies have confirmed that there is, indeed, a correlation between the various behaviour patterns of bio vital signals—respiration, heartbeat and body movement—and the different sleep stages. This correlation can be used as a base for sleep stage recognition with an appropriate algorithm. Not only does this include the aforementioned parameters themselves, but also parameters derived from them are able to be used in the algorithm for calculations. The selection of a convenient method and its application will now be described.

Approach

In order to build on the successful studies and promising results of several publications presented in part earlier in the section 'State of the art', this project aims to combine existing methods for the purpose of creating a new algorithm which will be able to analyse sleep phases based only on the following parameters of the human body:

- heartbeat;
- respiration;
- body movement.

The choice of these parameters is not accidental. Each of them can potentially be measured using non-invasive methods. This method will help to ensure that the sleep recording process does not jeopardize the accuracy of results or disturb the subjects' sleep.

All three signals mentioned below are taken from PSG records of the Charité dataset[5]. In this study, the heartbeat signals were substituted by ECG and, for the respiration signals, a thoracic (VTH) inductive plethysmogra-

[5] The Charité dataset consists of 230 h of sleep recordings from the Sleep Medicine Center at Charité Universitätsmedizin Berlin (Germany) Center of Sleep Medicine, in which 30 individuals (an equal number of male and female participants) with no significant health disorders were studied.

phy record was used. The body movement signal was replaced by the signal called *Beweg*, which is monitored by a 3D acceleration sensor in a recording device placed on the chest of the subject. For that, a 3D actigraphy sensor integrated into a PSG-System 'SOMNOscreen PLUS' is used. It has the following properties: amplitude resolution = 16 bit, accuracy = ±5%, Sampling-rate = 32 Hz, frequency range = 0.07–1 kHz. Baseline wandering should be excluded by the lower frequency range and also not significant for the end results as only the changes of the signal and not the absolute value will be analysed and furthermore, baseline wander has a much lower amplitude than significant movements. For every axis the mean value of movement for every second is calculated and stored:

$$X = \frac{1}{n}\sum_{i=1}^{n}|x_i - x_{i-1}|$$
$$Y = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_{i-1}|$$
$$Z = \frac{1}{n}\sum_{i=1}^{n}|z_i - z_{i-1}|$$

with $n = 32$ (sampling-rate), and after that the activity was computed as

$$\text{Body}_i = \sqrt{X^2 + Y^2 + Z^2}.$$

The part of the software which is responsible for the classification procedure is based on the statistical regression method—MLR. Therefore, both dependent (Y) and independent ($X_1, X_2, \ldots, X_n$) variables have to be estimated. The dependent variable Y signifies the several sleep stages. As for the independent parameters, they must be defined as derived parameters from the vital parameters mentioned earlier—heartbeat rate, respiration and body movements. In other words, HRV is a derived parameter from the heartbeat rate parameter and could be used as a predicting variable for parameter Y.

In this study, to implement the MLR model, ten derived parameters were selected, which would reflect the behaviour of a number of sleep phases. All derived parameters were calculated for 30 s time periods k_i of the particular sleep recording.

The chosen derived parameters from the heartbeat signals are the following:

- Heart beat interval—mean RR interval between successive heart beats for every 30 s epoch (HBI).
- Number of heartbeats per 30 s (HB).
- HRV—mean difference of successive RR interval lengths per 30 s (HRV).
- *R*—algorithm (RA).

The RA algorithm stands for the $R(k)$-algorithm (Kurihara and Watanabe 2012). The values of $R(k)$ increase during REM stage. The reason for this behaviour is that the frequency of respiration and heartbeat accelerate during REM sleep (see table 1), but at the same time these signals become less rhythmic (Kurihara and Watanabe 2012). It is calculated as

$$R(k) = \frac{1}{2q+1}\sum_{i=-q}^{q}\left|H_{k+i}^{former} - H_{k+i}^{latter}\right|,$$

where the discrete time for every 1 min (starting from the first minute of the record and ending at the last one) is defined as k, H_{k+i}^{former} and H_{k+i}^{latter} are the heart rate values from the former and latter 30 s of the time interval $(k + i)$ and i represents the movement inside the window (moving average) with the size $2q$. In this work according to (Kurihara and Watanabe 2012) $q = 10$ was used

One parameter derived from body movement is the mean value of body movement (BM) signal. It is defined as

$$\text{BM}(k) = \frac{1}{n}\sum_{i=0}^{n-1}\text{Body}_i$$

with n equal 30 (the number of body-movement records for one 30 s epoch—one movement record per second—see above) and Body_i is calculated as the square root of $(X * X + Y * Y + Z * Z)$ with X, Y and Z the values of signals per axis of a 3D actigraphy-sensor as described above.

Furthermore, the mixed derived parameter from body movement and heartbeat signals, known as DA, was chosen. It represents the $D(k)$-algorithm (Kurihara and Watanabe 2012). This algorithm was developed based on the knowledge that when moving from the wake stage to deeper sleep, body movement decreases and at the same time its frequency drops (see table 2). DA represents the combination of changes in movement and heartrate and is calculated as

$$\mathrm{DA}\,(k) = \log_2 \frac{A_k^{body}}{A_k^{heart} + A_k^{body}}.$$

A_k^{body} and A_k^{heart} are the mean amplitudes of the body movement and heartbeat signals for the time k. The proposed formula uses a logarithm to reduce the effect of large fluctuations in body movements and to enhance slighter movements of the body.

The derived parameters for the respiration signal are the following:

- Mean respiratory depth of inhalation (P_{sdm}).
- Mean respiratory depth of exhalation (T_{sdm}).
- Median respiratory volume during breathing cycles (V_{br}).
- Median respiratory volume during inhalation (V_{in}).

These parameters were chosen because these two variants of respiration characteristics are included in the sleep stage calculation. The P_{sdm} and T_{sdm} features take into account the mean respiratory depth and its variability at the same time, in terms of inhalation and exhalation. The V_{br} and V_{in} parameters are the volume-based features that should reflect the changes of respiratory effort signals (Long *et al* 2014).

Here is the mathematical representation of the derived respiration parameters:

$$P_{sdm}\,(k) = \frac{\mathrm{median}(p_1,p_2,\ldots p_n)}{\mathrm{IQR}(p_1,p_2,\ldots p_n)}$$
$$T_{sdm}\,(k) = \frac{\mathrm{median}(t_1,t_2,\ldots t_n)}{\mathrm{IQR}(t_1,t_2,\ldots t_n)}$$
$$V_{br}\,(k) = \mathrm{median}\left(\sum_{s_x\in\Omega_1^{br}} s_x, \sum_{s_x\in\Omega_2^{br}} s_x, \ldots, \sum_{s_x\in\Omega_k^{br}} s_x\right)$$
$$V_{in}\,(k) = \mathrm{median}\left(\sum_{s_x\in\Omega_1^{in}} s_x, \sum_{s_x\in\Omega_2^{in}} s_x, \ldots, \sum_{s_x\in\Omega_k^{in}} s_x\right),$$

where $p = p_1, p_2, \ldots, p_n$ and $t = t_1, t_2, \ldots, t_n$ are the peak and trough sequences from a chosen time window (in this case—30 s), the kth breathing cycle is declared with Ω_k^{br}, the kth inhalation and exhalation cycles with Ω_k^{in} and Ω_k^{ex} with k consecutive breathing cycles ($k = 1, 2, \ldots, K$).

The possibility that all derived parameters could in some way depend on each other is not problematic since there is no need for the independent variables to be statistically independent of each other (Belsley 1991).

Exploratory data analysis (EDA) was used to identify the significant features among 10 proposed derived parameters. According to its results, several classification attempts were executed with different sets of features. However, reducing the number of features to 7 or lower (e.g. excluding RA, VBR and HRV according to results of EDA) led to a reduction of accuracy (about 10%–30%). Excluding 1 or 2 parameters (e.g. VBR and RA) did not have any significant effect on accuracy (with the current amount of test persons). As a result, a decision was made to include all ten of the derived parameters because their calculation did not make the algorithm significantly more difficult, but did retain the high accuracy of the results. Furthermore, to accurately evaluate the influence of derived parameters on the system work, it wold be necessary to increase the number of test subjects in the study substantially. In other case, because of the individuality of each person and just minimal differences by excluding of 1–2 features, it is not possible to estimate the importance of every feature for the overall results.

To increase the accuracy of classification results, the transition patterns between different sleep stages were taken into consideration by the implementation of the algorithm described in this paper. As Schlemmer *et al* (2014) has proved, some transitions between sleep stages are much more probable than others and at the same time some of transitions have a probability of nearly zero. In the first implementation of the algorithm the emphasis was placed on the not probable transitions. As the algorithm works with probabilities, it allows the inclusion of additional adjustments, which increases accuracy without having to adhere to strict rules. In this case the algorithm decreased the probability of the examined sleep stage by 10% to 15% (a higher percentage would have a negative effect if the previous stage was incorrectly classified), if its appearance probability according to the transition pattern is almost equal to zero. This approach has led to an increase of the accuracy of the proposed algorithm by up to 3% compared to the implementation without considering the transition probabilities.

Implementation

In order to calculate the derived heartbeat parameters HB, HBI, HRV and RA, the exact time (in respect to the start of the record) of R-peaks, which occurred within 30 s epochs, must be identified and extracted. Therefore, to remove the disturbances (e.g. baseline wander) from the ECG, data filter techniques have to be used along with methods to remove artefacts. To achieve this goal, a linear high-pass filter (HPF) algorithm (Chen and Chen 2003) was implemented in the project. Due to the usage of HPF, the low-frequency noise sources of the ECG signals, such as P and T waves, and the baseline wander will be suppressed. The HPF-method used here is represented mathematically as

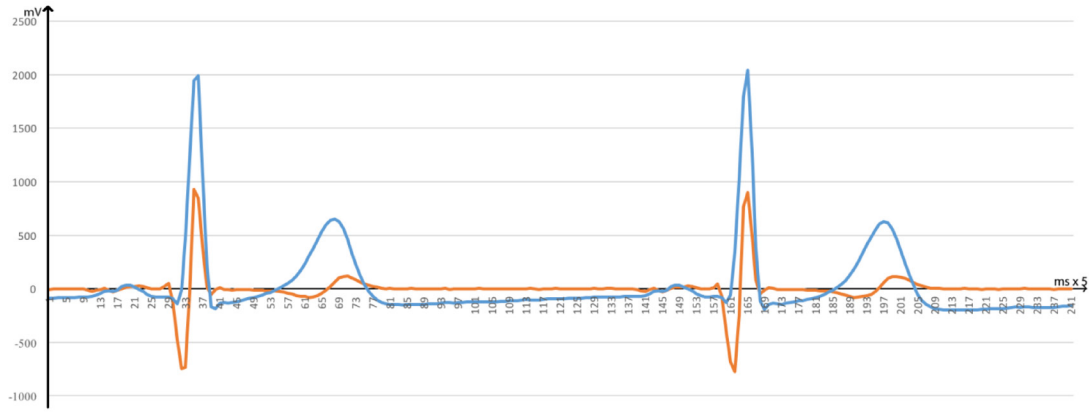

Figure 1. Overlap of the same ECG signal interval.

$$y_1[n] = \frac{1}{M}\sum_{m=0}^{M-1} x[n-m],$$

where $x[n]$ is the input data and M represents the filter length. For the presented implementation the value of parameter M was determined to be 5 according to Chen and Chen (2003).

Figure 1 shows an overlap of the typically unfiltered (blue) and filtered (red) ECG signals for the same time interval. As the figure indicates, the key pillar of the HPF-method lies in the fact that the filter does not shift the signals on the axis X, which otherwise could cause the occurrence of incorrect time values of the R-peaks. Therefore, the lower frequency noise sources of the ECG signal and the baseline wander will be suppressed with the help of the HPF-method.

After removing the artefacts from the ECG signal with a HPF-filter, the next stage of calculation is to identify the timestamps of all R-peaks. The following step is to estimate the derived parameters and to store, for further computations, the lists of all derived parameters for every subject in a CSV-File. The values from each list are calculated in ratio to the derived parameter average signal. Each list is sorted in chronological order and every entry contains the values for a 30 s epoch. For instance, in the column HRV are determined values for every epoch, calculated as the difference (in ms) between HRV per actual 30 s and the average HRV of the full sleep record. The first 10 min of the recording of one subject are presented in table 3 to illustrate the possible deviation of derived parameters. As can be seen in the table, the deviations can be very diverse even for the same sleep stage, but the results of current research confirm that it is still possible to find some dependencies with appropriate mathematical model. This makes sleep stage recognition based on these parameters a challenging, yet promising method.

The list column 'Index' represents the consecutive number of sleep epochs, then values of derived parameters are listed. 'AllStages' contains the sleep stages obtained by the experts, according to the RK-method. The column 'Four-Stages' shows more results of those specialists; however, the categorization is divided into the stages WAKE (1), Light- and Deep Sleep (2 and 3) and REM (4). The column 'ThreeStages' represents the experts' results which are categorized into the stages WAKE (1), NREM (2) and REM (3).

Regression analysis is a complex process and difficult to implement in a research project. Furthermore, algorithms have already been implemented in various programming environments (Bochkanov 2018). Therefore, there was no need to develop completely new algorithms or implementation strategies. To achieve the goals of this project, free software and the programming language *R* (v.3.3.1) for statistical computing and graphics were utilized. The use of *R* was justified because of its ability to provide a wide range of statistical variety (classical statistical tests, linear and nonlinear modelling, time-series analysis, clustering, classification, etc) and various graphical techniques. Also, *R* is compatible with and runs on a wide variety of UNIX platforms, Windows and MacOS. Integration of *R* into the developed software was done using the corresponding free access libraries.

The merged derived parameter list of five files (representing five subject recordings) represents the training data, which then is used as a training set for the MLR model in *R*. In addition, category variables have to be set as factor variables. 'Wake' was chosen as the reference category for the MLR analysis because it is the only 'stage' not included in the bigger category 'Sleep' containing REM, Light Sleep and Deep Sleep stages. The *multinom* method of *R software* is used to fit the multinomial log-linear models to maximum probability estimations. Finally, the significance of the parameters can be analyzed with the 2-tailed Wald-z test. The most important part of this project is the *R* prediction-method, called *predict*. It requires the fitted model and the data frame to predict the probability of each stage belonging to a particular category. The predicted results are supposed to be stored in a separate .xls file.

Table 3. Extract of the file Param.csv, including the results of the first 10 min of a subject.

Index	HB	HRV	HBI	BM	TSDM	PSDM	VBR	VIN	RA	DA	AllStages	FourStages	ThreeStages
1	19.234	19.87	−18.031	361.135	−88.419	−85.571	−22.398	0.974	−26.582	−78.349	1	1	1
2	3.121	3.222	−2.095	−7.022	−3.863	38.145	−27.179	−32.812	−6.312	−17.725	1	1	1
3	−0.101	−0.108	−2.352	−12.769	7.2	0.854	−23.869	−13.121	−11.14	−13.394	1	1	1
4	3.121	3.222	−3.133	−18.379	78.726	−33.214	−21.294	−5.255	19.782	−4.734	1	1	1
5	−0.101	−0.108	−3.37	−29.668	56.493	−28.998	−13.939	14.367	58.445	8.257	2	2	2
6	3.121	3.222	−2.122	−20.911	−35.676	−27.311	−20.559	−19.075	40.053	−0.404	2	2	2
7	−0.101	−0.108	−2.627	−29.736	34.384	−34.394	−15.778	14.78	−27.579	8.257	2	2	2
8	−0.101	−0.108	−0.757	−20.5	−1.307	−11.514	−11.732	13.721	−86.397	−4.734	2	2	2
9	−0.101	−0.108	−1.115	−24.263	−64.183	−63.253	−12.468	−14.441	−88.39	−0.404	1	1	1
10	−3.324	−3.438	2.523	−33.636	−18.537	−52.984	−4.744	25.011	−83.562	12.587	2	2	2
11	−3.324	−3.438	1.897	−37.057	7.611	46.315	5.554	29.109	−8.266	16.918	2	2	2
12	−6.546	−6.768	2.766	−35.552	37.512	−8.966	2.244	31.983	81.59	8.257	3	2	2
13	−0.101	−0.108	1.361	−35.825	26.288	−33.008	−1.434	32.808	18.786	16.918	3	2	2
14	−0.101	−0.108	−2.16	−20.158	42.802	−8.722	−2.17	27.197	−30.529	−4.734	3	2	2
15	25.679	26.53	−20.673	137.134	−84.719	−26.599	−34.535	11.507	−6.427	−61.027	1	1	1
16	−3.324	−3.438	0.36	−23.237	−9.815	−11.252	−15.41	−0.374	−12.098	−4.734	1	1	1
17	−3.324	−3.438	3.234	−31.447	−40.002	−40.316	−12.468	15.027	−50.685	8.257	2	2	2
18	−3.324	−3.438	1.897	−37.673	−50.225	−54.089	−8.79	15.206	−99.004	16.918	2	2	2
19	−3.324	−3.438	1.641	−37.741	−45.882	−47.399	−2.17	18.921	−29.533	16.918	2	2	2
20	−3.324	−3.438	3.519	−34.936	2.249	−23.919	10.335	29.673	−55.628	12.587	2	2	2

Table 4. Classification results. Four stages.

		Stage (developed SW-system)			
		WAKE	LS	DS	REM
Stage (expert)	WAKE	2649	1146	480	219
	LS	1686	10 023	1700	942
	DS	168	3041	2104	58
	REM	628	1328	151	1339

Table 5. Classification results. Three stages.

		Stage (developed SW-system)		
		WAKE	NREM	REM
Stage (expert)	WAKE	2472	1642	380
	NREM	2060	16 004	1658
	REM	748	1101	1597

Results

This project is focused on analysing the sleep records from the Charité clinic in Berlin[6]. In the main study, about 230 h of recordings from the Sleep Medicine Center at Charité were analysed. The data originates from the sleep recordings of 30 individuals (the amount of male and female participants is equal) with no significant health disorders. The average age was 38.5 $\pm$ 14.5 years old and the BMI of participants averaged 24.4 $\pm$ 4.9 kg m^{-2}. The sleep stages of each PSG recording were manually measured in 30 s time intervals based on the Rechtschaffen and Kales method. As a result, the classification was made for three stages: WAKE, NREM and REM, and for four stages: WAKE, Light Sleep (i.e. NREM 1 and NREM 2), Deep Sleep (i.e. NREM 3 and NREM 4) and REM.

Using real data from a sleep laboratory for system evaluation can help to confirm the accuracy of the theoretically developed algorithm in a practical setting. Therefore, firstly, the results of the statistical estimations with regard to the mean and quartiles of the derived parameters were listed to show their behaviour in the four main sleep phases—WAKE (W), Light Sleep (LS), Deep Sleep (DS) and REM (R). In the next step, the MLR model fitting process was described. Finally, the results of the regression analysis in 3 and 4 classification categories were displayed for the provided data.

In this project, the recordings were strictly separated into the types of data—training and tests. Firstly, the recordings of five subjects were entered into the model for the Charité dataset for the training. Then, for classification, the recordings of the remaining 30 subjects were used. Average age, height, weight and state of health of subjects in both datasets had no significant differences. At the same time, the training dataset has also covered the values with deviation (in both directions but without outlier) from the average. The male/female ratio was also similar.

The proposed first version of the algorithm was implemented for the classification of sleep stages of healthy adults. The tests also used adult subjects with no record of sleep disorder.

In total, 27 662 time intervals were recorded, each with a length of thirty seconds. The classification matrices for the Charité subjects in sleep stages 4 and 3 are shown in the tables 4 and 5. In the end, the results for the overall classification rate were

- with 4 stages—accuracy: 58%, Cohen's kappa: 0.50;
- with 3 stages—accuracy: 72%, Cohen's kappa: 0.67.

Each row of the matrix contains the rates for both the classified and misclassified time intervals. For example, the first row in table 4 indicates that 2649 intervals were correctly classified as the phase *WAKE*. The next entry displays the 1146 intervals classified as *LS* phase, but it was actually *WAKE*, and so on.

Tables 6 and 7 present the percentage of correct classification of sleep stages for each stage in particular. It can be seen, that in four stages, WAKE and Light Sleep is recognised with the highest accuracy, whereas the error rate at the Deep Sleep and REM stages is high, so an improvement in accuracy will be required.

[6] Thomas Penzel, Dr. rer. medic. Martin Glos; initial study was carried out in Charité—Universitätsmedizin Berlin Center of Sleep Medicine Charitéplatz 1, D-10117 Berlin (Germany).

Table 6. Correctly classified stages. Four stages.

Rights	58	%
Mistakes	42	%
WAKE	59	%
LS	70	%
DS	41	%
REM	36	%

Table 7. Correctly classified stages. Three stages.

Rights	72	%
Mistakes	28	%
WAKE	56	%
NREM	81	%
REM	42	%

Discussion

Supporting sleep professionals with the help of non-invasive system for long-term sleep study in the home environment is the goal of this research. In particular, this paper discusses the software part of the system.

There are two main points, which represent the novelty of this work: the unique set of derived parameters which can be calculated from bio-vital signals and measured in a non-invasive way, and the use of multinomial logistical regression for the classification (which enables the possibility of easy algorithm adjustment—e.g. the consideration of transition patterns). The following paragraphs will elaborate on these points in more detail.

To implement an automatic sleep stage analysis and evaluation in this project, the appropriate software was developed. This program had the task of automatically classifying a sleep phase in 30 s time intervals based on the vital parameters—respiration, heartbeat and body movement signals from raw PSG recordings. Moreover, for this technique to be transferable into later studies with the sensor set mentioned above, the signals from them would not be translated with the raw values of the recordings. The goal was to create an algorithm which would present the parameter signals through the relation to parameter's average signal. The reason for the development of this method is to make evaluations which consider the differences in individuals. After the experiment was finished, the classification results were compared with the pre-estimated whole-night PSG recordings.

The core of the software is the algorithm for sleep stage recognition using respiratory, heart rate and body movement signals. The results achieved so far are the following: 58% correct recognition rate for Wake, Light (N1 and N2), Deep (N3) and REM stages and up to a 72% hit rate for Wake, NREM and REM stages. Sleep recordings of 30 subjects were used for the evaluation. The outcomes obtained are promising, but improvements are still possible and they will be developed in order to obtain results which are closer to PSG application.

There are several reasons for low accuracy of recognition in the REM stage. One of them is its similarity with WAKE sleep concerning breathing. Another more significant reason is that the REM stage is shorter than other sleep stages. Therefore, every incorrectly recognized epoch of REM sleep has percentagewise a high degree of influence on accuracy of recognition of this stage, whereas accuracy of the whole system will be less affected. Yet another reason for low accuracy of recognition is that the algorithm tends to recognise the next epoch as being the same as the previous one, if its features are similar. That being the case, since REM follows the LIGHT stage, the first 'transitional' epochs are often misclassified due to the presence of similar characteristics (heart rate and breathing).

The primary result of this project is that the selection of parameters (set of derived parameters from respiration, heart rate and body movement) has proved its validity. The theoretical foundation described in the chapter State of the Art has inspired this line of thinking and the results of this study have justified it. It is important to keep in mind, that there are still conceivable improvements to be made to the algorithm which could increase the accuracy of results, but would not change the main concept. Furthermore, these parameters can be obtained in a non-invasive way with a hardware system (e.g. with sensors placed under the bed's mattress) that can be installed by non-experts, which would be important for a home sleep study system.

Reducing the amount of derived parameters by using only three base features (heart rate, respiration rare and movement) will not lead to significant benefit for the proposed algorithm because the calculation of derived parameters does not need a high degree of computational power.

To evaluate the introduction of respiration as a bio vital signal for sleep stage classification, the test was executed with the same dataset but without using the respiratory signal. The following results were achieved: general

accuracy for WAKE/LS/DS/REM decreased to 53% and WAKE/NREM/REM to 64% (compared to 58% and 72% with respiratory signal). More importantly, the recognition rate of the REM stage diminished to about 15% for both sets of stages, which is not acceptable.

The first attempt to use the balanced proportion of classes in the training dataset was made (WAKE/LIGHT/DEEP/REM = 25%/30%/25%/20%), but the total accuracy was decreased by 2% for WAKE/LIGHT/DEEP/REM and was not significantly changed for WAKE/NREM/REM stages. Nevertheless, the increase in the recognition rate of the REM stage has indicated the potential of this approach.

The classification into three (Wake, NREM, REM) and four (Wake, Light, Deep, REM) sleep stages had to be performed according to the research goal. Multinomial logistical regression (MLR) is the classification method that was chosen here because it allows a categorical distribution of values. One dependent variable (in this case—sleep stages) has as input a set of different independent variables-parameters, and they may even overlap. More details are presented in the Statistical Methods chapter. Using MLR as a base for categorical distribution of sleep stage parameters will be used to enhance the process in the future, as it enables an easy and fast adjustment of the algorithm that depends on the requested input parameters. Even if just three bio vital signals are used in the method described, different derived parameters can be calculated when dependent variables are used in the regression model. Moreover, the adaptability of this algorithm allows for quick modifications if other bio vital signals are available for the study.

Though the selection of suitable parameters for the model was one of the challenging tasks, it has led to satisfactory results. However, some improvements are still possible, for instance the identification of other derived parameters that can be obtained in a non-intrusive way. Selection of the most suitable parameters can be proposed on the basis of theoretical research, but proof is only possible after analysing the results of numerous studies. Even with a distinguished background in qualitative analysis, it would be necessary to test the proposed set of parameters with real data in order to confirm the theoretical assumptions. The reason for this is the complexity of biological processes in the human body which have a variety of effects on bio vital signals, which can be used as inputs for the algorithm.

The consideration of transition patterns between different sleep stages was restricted to only one rule. Introducing other rules in respect to the transition matrix would increase the risk of reducing accuracy because if the previous sleep stage were incorrectly recognised, it would affect the classification of the next stages.

The strategy of this approach has been confirmed as being a highly credible method for yielding successful sleep stage classification with the help of a greatly reduced set of parameters. The number of test subjects could be increased in order to collect and evaluate a higher amount of data. For this reason, further studies with different sets of input parameters and a higher number of subjects are in planning. During the study, several possible algorithm improvements were identified, and amongst others, the implementation of *trust anchors* will be investigated. The chapter Conclusions and Future Work focuses on this and other possibilities.

Comparing the results of this work with other studies, it is important to keep in mind that input signals for the system presented here could be obtained in non-invasive way. Most of the scientific articles on sleep stage classification (e.g. Supratak *et al* (2017) and Karimzadeh *et al* (2018)) describe systems using electroencephalography (EEG) as input. Using EEG can provide higher accuracy, but requires placing the sensors direct on the test subject's head, which would be not acceptable for a non-invasive system. Therefore, the most similar approach to use for comparison would be ballistocardiography or other methods using respiration, heart rate and movement signals as input. In Mendez *et al* (2010) an accuracy rate of about 0.79 was achieved, but only the classification of NREM and REM stages was determined and WAKE stage was excluded. The research system proposed here also includes WAKE stage, which explains the slightly lower total accuracy, but nevertheless provides increased usefulness as a sleep stage identification system.

Another paper Samy *et al* (2014) describes an unobtrusive sleep stage classification system using a pressure-sensitive bed sheet which achieved 70% precision for WAKE, NREM and REM stages. With our system, we obtain a higher rate of accuracy and it was tested on a higher number of overnight PSG recordings (30 compared to 7).

The number of features used in Long *et al* (2014) is higher (14/13 compared to 10 in our method) and the results are equal without subject-specific normalization and just about 3% better with subject-specific normalization, but as mentioned before with a higher number of included features.

In Aktaruzzaman *et al* (2017) an accuracy rate of about 0.78 was achieved with a lower number of features. However, the proposed method can identify only SLEEP/WAKE stages, and in addition, an assumption was made to include NREM1 as a part of the WAKE stage. In the proposed in this paper method not only SLEEP/WAKE, but also WAKE/NREM/REM and WAKE/LIGHT/DEEP/REM classification is done, which is a different and much more complicated task. However, a comparison with the results from Aktaruzzaman *et al* (2017) can be easily done using the already obtained results (having SLEEP = LIGHT + DEEP + REM). In this case even without adjusting the algorithm to recognise only SLEEP/WAKE stages, the approach presented in this paper got better results: accuracy: 0.84, sensitivity: 0.59, specificity: 0.89 and kappa: 0.46 compared to Acc = 0.78, Se = 0.85, Sp = 0.48 and K = 0.30 in Aktaruzzaman *et al* (2017).

The aforementioned results and the abundant opportunities for advancement in sleep stage classification prove the potential of the concept presented here and the necessity for further research on this approach. The comparison with the state of the art articles has confirmed the quality of achieved results, having a better relation 'amount of features'/'correctly classified sleep stages'.

Conclusions and further work

The results obtained in this study (72% accuracy, Cohen's kappa—0.67 for Wake, NREM and REM stages) prove that the algorithm developed for this project represents a truly promising method for sleep stage classification using just a few bio vital parameters: heart rate, respiration and body movement. It is also obvious that there are several ways to improve the classification algorithm. The first possibility is to define more derived parameters and examine them according to their significance in the MLR-model. For instance, the additional inclusion of sample entropy as an independent variable could potentially improve the model.

Also, the classification data could be revised in a second step. If a particular category is predicted with a lower percentage, the time intervals directly preceding and following it could be re-analysed in order to improve the overall classification rate. Besides this, the intervals which were classified with a high-confidence level could act as trust anchors. They could adjust the percentage values for the category membership of their neighbours.

Moreover, to further improve the software system, all the probabilities from the sleep stage transition matrix will be examined and two-step transitions (Schlemmer *et al* 2014) will be introduced.

Furthermore, an investigation is planned to study the effect of a currently unbalanced number of samples of different sleep stages on the accuracy of the method (Tilmanne *et al* 2009). One of the possible solutions could be to consider the balanced proportion of classes (WAKE, NREM, REM) for training dataset with executing of following test classification, but this point should be deeply investigated, as the first results have indicated.

The reduction of used derived features is also planned for the future when executing studies with higher numbers of test subjects would be possible.

Another approach to the classification of sleep phases is the implementation of a multi-layer neural network algorithm with the MLR model in the first layer, though it is still questionable whether the neural network analysis would improve the classification rates. Thus, it should be considered, but thoroughly examined and scrutinized.

A perfect match of the evaluated results with the findings of a sleep medicine professional is not possible due to the *R–K* method's judgment standards, which include ambiguities. However, all in all it is still desirable to develop a more accurate and, more importantly, non-invasive solution for the problem of recording and estimating the sleep stages. It would help to find alternative methods which could, in time, substitute for the more complex PSG procedure.

This research is part of an intended far larger study aiming to develop a monitoring system for home-based sleep analysis. It was particularly focused on designing a software solution that can classify the sleep stages. As mentioned earlier, the ultimate goal is to obtain the signals of particular bio vital parameters via a set of pressure sensitive sensors placed under the mattresses of study subjects. This approach should presumably ensure that the sleep recording process does not disturb the sleep of the study participants. Furthermore, to get proper recordings of a natural sleep state, the subjects must feel as though they are sleeping in a normal bed in familiar conditions.

The planned system will include both software and hardware parts. The hardware portion will provide data collection and pre-processing. The data recording will be done through a set of sensors and necessary filters will be applied directly to the hardware to record movement, breathing and heart rate signals. After that the information will be automatically transferred to a server, where the software will take care of further operations. It will be processed there and analysed by the developed sleep stage algorithm. The interpretation of the acquired results has to be made by medical professionals. After completing the development and integration of all the system parts, the resulting monitoring system could provide useful diagnostic support to sleep professionals and enhance the field of sleep medicine by offering a convenient way to perform preliminary sleep studies in the home environment.

Acknowledgments

We thank the Interdisciplinary Center for Sleep Medicine of Charité Clinic in Berlin and in particular Dr rer. medic. Martin Glos, Beate Diecker and Claudia Biro for supporting the study. We would also like to thank Agnes Klein for her help, which was very important for the development of the system.

This research was partially funded by the EU Interreg V-Program 'Alpenrhein-Bodensee-Hochrhein', which means are provided by the European Fund for Regional Development (EFRE) and the Federal State of Switzerland: Project 'IBH Living Lab Active and Assisted Living', grants ABH40 and ABH66.

ORCID iDs

Maksym Gaiduk https://orcid.org/0000-0001-9607-5709
Thomas Penzel https://orcid.org/0000-0002-4304-0112
Juan Antonio Ortega https://orcid.org/0000-0001-9095-7241
Ralf Seepold https://orcid.org/0000-0002-5192-2537

References

Aktaruzzaman Md *et al* 2017 Performance comparison between wrist and chest actigraphy in combination with heart rate variability for sleep classification *Comput. Biol. Med.* **89** 212–21
Alexander V E and Christof S 1998 *Regression Analysis for Social Sciences* (New York: Academic)
Belsley D A 1991 *Conditioning Diagnostics: Collinearity and Weak Data in Regression* (*Wiley Series in Probability and Statistics* (Book 262)) (New York: Wiley)
Berry R B and Wagner M H 2015 Indications for polysomnography, portable monitoring, and actigraphy *Sleep Medicine Pearls* 3rd edn (Philadelphia, PA: Elsevier) (https://doi.org/10.1016/B978-1-4557-7051-9.09946-0)
Bochkanov S 2018 Alglib - a cross-platform numerical analysis and data processing library
Chen H C and Chen S W 2003 A moving average based filtering system with its application to real-time QRS detection *Comput. Cardiol.* **30** 585–8
Choi B H, Choi H J M, Shin B, Lee J Y, Jeong K D U, Seo S J W and Seo J W 2007 Non-constraining sleep/wake monitoring system using bed actigraphy *Med. Biol. Eng. Comput.* **45** 107–14
Chokroverty S and Thomas R 2013 *Atlas of Sleep Medicine* 2nd edn (New York: Elsevier)
Chung G S, Kim K K, Lim Y G, Choi J W, Jeong D-U, Park K S and Choi B H 2007 *REM Sleep Classification with Respiration Rates* (Tokyo: IEEE) pp 194–7
Douglas C M, Geoffrey V G and Peck E A 2012 *Introduction to Linear Regression Analysis* 5th edn (New York: Wiley)
Ettinger S J and Feldman E C 2010 *Textbook of Veterinary Internal Medicine Expert Consult* (*Expert Consult.*) 2nd edn (St Louis, MO: Saunders)
Gaiduk M, Vunderl B, Seepold R, Ortega J A and Penzel Th 2018 Sensor-mesh-based system with application on sleep study *Bioinformatics and Biomedical Engineering IWBBIO 2018* (*Lecture Notes in Computer Science* vol 10814) (Cham: Springer)
Garson G D 2016 *Logistic Regression: Binary & Multinomial* (Asheboro, NC: Statistical Associates Publishers)
Guang-Bin H, Chee-Kheong S, Qin-Yu Z Z and Qin-Yu Z 2006 Extreme learning machine: theory and applications *Neurocomputing* **70** 489–501
Hayet W and Slim Y 2012 Sleep-wake stages classification based on heart rate variability *Int. Conf. on BioMedical Engineering and Informatics (BMEI 2012)* (IEEE) pp 996–9
Kambayashi Y and Hagiwara H 2012 Estimating sleep cycle using body movement density *Biomedical Engineering and Informatics (BMEI)* (IEEE) pp 1081–5
Karimzadeh F, Boostani R, Seraj E and Sameni R 2018 A distributed classification procedure for automatic sleep stage scoring based on instantaneous electroencephalogram phase and envelope features *IEEE Trans. Neural Syst. Rehabil. Eng.* **26** 362–70
Katz M H 2011 *Multivariable Analysis: A Practical Guide for Clinicians and Public Health Researchers* 3rd edn (Cambridge: Cambridge University Press)
Kurihara Y, Hiroshi T and Watanabe K 2010 Sleep-states-transition model by body movement and estimation of sleep-stage-appearance probabilities by Kalman filter *IEEE Trans. Inf. Technol. Biomed.* **14** 1428–35
Kurihara Y and Watanabe K 2012 Sleep-stage decision algorithm by using heartbeat and body-movement signals *IEEE Trans. Syst. Man Cybern.* A **42** 1450–9
Le Bon O, Hoffmann G, Dramaix M, San Sebastian I, Murphy J R, Kentos M, Pelc I, Staner P L L and Staner S L 2001 The first-night effect may last more than one night *J. Psychiatric Res.* **35** 165–72
Lee-Chiong T L, Brooks R and Mattice C 2012 *Fundamentals of Sleep Technology* (Baltimore, MD: Williams & Wilkins)
Long X, Fonseca P, Haakma R, Aarts R M and Foussier J 2014 Analyzing respiratory effort amplitude for automated sleep stage classification *Biomed. Signal Process. Control* **14** 197–205
Mendez M O, Matteucci M, Castronovo V, Ferini-Strambi L, Cerutti S and Bianchi A M 2010 Sleep staging from heart rate variability: time-varying spectral features and hidden Markov models *Int. J. Biomed. Eng. Technol.* **3** 246–63
National Heart Lung and Blood Institute (NHLBI) 2011 *Your Guide to Healthy Sleep* (*NIH Publication No. 11-5271*) (Bethesda, MD: National Heart, Lung, and Blood Institute)
Penzel Th, Kantelhardt J W, Bartsch R P, Riedl M, Kraemer J F, Wessel N, Garcia C, Glos M, Fietze I and Schöbel C 2016 Modulations of heart rate, ECG, and cardio-respiratory coupling observed in polysomnography *Frontiers Physiol.* ()
Samy L, Huang M C, Liu J J, Xu W and Sarrafzadeh M 2014 Unobtrusive sleep stage identification using a pressure-sensitive bed sheet *IEEE Sens. J.* **14** 2092–101
Schlemmer A, Parlitz U, Luther S, Wessel N and Penzel Th 2014 Changes of sleep-stage transitions due to ageing and sleep disorder *Phil. Trans. R. Soc.* **373**
Singh J *et al* 2015 American Academy of Sleep Medicine (AASM) position paper for the use of telemedicine for the diagnosis and treatment of sleep disorders: an American Academy of Sleep Medicine Position Paper *J. Clin. Sleep Med.* **11** 1187–98
Spriggs W H 2009 *Essentials of Polysomnography* (Sudbury, MA: Jones & Bartlett Learning)
Supratak A, Dong H, Wu C and Guo Y 2017 DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 1998–2008
Tataraidze A, Anishchenko L, Korostovtseva L, Kooij B J, Bochkarev M and Sviryaev Y 2015 Sleep stage classification based on respiratory signal *2015 37th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC) (Milan)* pp 358–61 ()
Tilmanne J, Urbain J, Kothare M V, Wouwer A V and Kothare S V 2009 Algorithms for sleep-wake identification using actigraphy: a comparative study and new results *J. Sleep Res.* **18** 85–98
Yoon H, Hwang S H, Choi J-W, Lee Y J, Jeong D-U and Park K S 2017 REM sleep estimation based on autonomic dynamics using R–R intervals *Physiol. Meas.* **38** 631–51