Adaptive Sleep/Wake Classification Based on Cardiorespiratory Signals for Wearable Devices

Walter Karlen, Claudio Mattiussi, Dario Floreano
Laboratory of Intelligent Systems, Ecole polytechnique fédérale de Lausanne (EPFL), Switzerland
{**name.surname**}**@epfl.ch**

***Abstract*—In this paper we describe a method to classify online sleep/wake states of humans based on cardiorespiratory signals for wearable applications. The method is designed to be embedded in a portable microcontroller device and to cope with the resulting tight power restrictions. The method uses a Fast Fourier Transform as the main feature extraction method and an adaptive feed-forward Artificial Neural Network as a classifier. Results show that when the network is trained on a single user, it can correctly classify on average 95.4% of unseen data from the same user. The accuracy of the method in multi-user conditions is lower (89.4%). This is still comparable to actigraphy methods, but our method classifies wake periods considerably better.**

***Index Terms*—biomedical signal analysis, wearable computing, sleep and wake classification, electrocardiography, respiratory effort, neural classifier.**

I. Introduction

Increased sleepiness over daytime has been identified as an important cause of accidents in transportation and factory plants [1]. It is therefore a major health interest to continuously monitor and report the sleepiness level of high risk persons such as pilots, truck drivers or shift workers. Continuously updated information about the persons' "need for sleep" could help these persons to schedule their breaks and sleep times.

We are currently developing a wearable adaptive device that monitors sleepiness. Different mathematical models to estimate sleepiness have been suggested [2]. In this paper we describe a method for sleep/wake classification which could be used with such a model in a wearable device. The device should be self-contained, low-power and light-weight. This puts tight restrictions not only on the choice of signals for the classification task, but also on the signal recording, processing and on the computational requirements of the classifier.

The gold standard method for assessing sleep in humans is the analysis of brain wave patterns (EEG) first described by Rechtschaffen and Kales [3]. The most common sleep analysis method is called polysomnography (PSG), which combines EEG recordings with different physiological signals like electromyography (EMG), electrooculography (EOG), respiratory effort, blood oxygen saturation, electrocardiograms (ECG) and video analysis. In PSG, 30-second epochs of the signals are used for decision making. The method is normally carried out in a controlled hospital environment and needs medical assistance for setting up sensors, monitoring and analysis. Although the analysis is typically computer-assisted [4], it still requires a sleep expert and is therefore expensive and time consuming. It is difficult to integrate polysomnographic sensors into a wearable device, as they are rather bulky, power-consuming and highly susceptible to noise. Furthermore, EEG recordings require many electrodes to be glued to the scalp, which makes it very cumbersome and uncomfortable.

In home environments where PSG is typically not available, physicians rely on actigraphy for sleep monitoring [5]. In this method, the acceleration of extremities (typically arms) are recorded over several days with a watch-like device using miniature accelerometers and a storage medium. Periods of low activity are later classified as sleep by offline computer processing. Many different classification algorithms have been suggested for actigraphy [6], [7], but often they cannot cope with the problem of mis-classifying low activity tasks like reading, watching TV, or the case where the sensor band is not worn [5], [7]. Recently, alarm clocks using accelerometers were commercialized [8], [9]. The activity is used to detect the best sleep phase for easy wake-up in a given time window (10 to 30 minutes). However, the accelerometers are only active at night and the clocks do not calculate sleep duration.

Changes in the Autonomic Nervous System (ANS) during sleep/wake transitions have been successfully identified as a reliable source of information [10]. Changes in activity of the ANS are reflected in various physiological signals such as heart rate, blood pressure, skin conductance, etc. The main focus of current research is on fluctuations of heart rate variability (HRV) during sleep [11]–[13]. However, the way of calculating the HRV is not uniform [14] and therefore results contradict each other. Further, HRV measures are very susceptible to noise. A wearable application of this technique is therefore difficult. Recently, Redmond and Heneghan [15] have added respiratory signals to the HRV to show the feasibility of using cardiorespiratory signals for discriminating sleep stages in subjects with obstructive sleep apnea. The advantage of cardiorespiratory signals is that they are easy to measure and the sensors can be applied by non-experienced users.

We are using cardiorespiratory signals together with an Artificial Neural Network (ANN) for sleep detection, but instead of filtering, signal reconstruction or artifact rejection with sophisticated algorithms, we treat the artifacts not as noise in the system, but as relevant information in the signal. Taken from that perspective, movement artifacts may give an indication of the activity of the user as actigraphy would do, but without the need of using an additional sensor. Contrary to all other studies, we rely on day and night recordings obtained in a non-hospital environment to have more realistic data.

1-4244-1525-X/07/$25.00 © 2007 IEEE

II. METHOD

The work presented in this paper consisted of recording ECG and respiratory effort signals of different subjects over day and night periods. Additionally, video, EMG and EOG were recorded for labeling the users' state as wake or sleep by a technician. This information will be used for training the neural classifier. The pre-processing consisted of calculating an estimation of the power spectral density (PSD) for the raw ECG and respiratory signals with the help of an Fast Fourier Transform (FFT). The obtained periodogram is then pruned. Three classifier architectures were designed, each having as input either the PSD values of ECG, respiration or their combination (Fig. 1). For each architecture, an ANN is trained and tested using data from a single user only. To investigate the capability of these networks to generalize for other users, another series of networks are then trained and tested using data from multiple users.

A. Data Recordings

For this study, we conducted home recordings with 4 healthy male subjects, between 23 and 29 years old. ECG and respiration effort were recorded with a Heally system (Koralewski Industrie Elektronik, Celle, Germany). The Heally is a wearable recording system that uses an inductive belt sensor for ribcage respiratory effort measurement and gel electrodes for recording ECG. We have chosen the sampling frequencies f according to the requirements for digitalized PSG [4]. The respiratory signal is sampled at $f_{Resp} = 50$ Hz and the 1-lead ECG at $f_{ECG} = 100$ Hz. Additionally, the Heally offers the possibility to measure the reference signals EMG (recorded from the right shoulder muscle (trapezius) at 200 Hz) and EOG (recorded at 200 Hz). EOG was only measured during the night, in order not to disturb the subjects too much during daily activities. During night-time a video of the upper part of the body was recorded.

The subjects wore the Heally for 16 hours per session. The recording started approximatively 4 hours before the regular bed time of the subject. A total of 14 recording sessions were carried out, 8 sessions for one subject (subject A) and 2 sessions for each of the other subjects (subject B, C and D). Each session contained an average of 7.18 hours ($\pm$ 1.46 SD) of sleep and 9.17 hours ($\pm$ 2.15 SD) of waking. A total of 100.48 hours of sleep and 128.42 hours of wake were analyzed. In case of sensor failure or detachment, the corresponding data were discarded.

Manual analysis of the video was done to determine if the subject was asleep or not. The video was divided into segments of 10 seconds and each segment was evaluated following precise criteria derived from [6], [10], [16]:

1) The person is considered to be awake if his eyes are open or body movements occur for more than 10 seconds.
2) If the eyes are closed, the subject is considered to be asleep when muscle tonus is released or slow eye movements are present. If segments of the video analysis were uncertain, the EOG and EMG signals were examined.
3) In doubtful cases, where neither EOG nor EMG signals could help to clearly identify sleep, the state was set to awake.

To avoid undetected wakefulness with closed eyes, the subjects were asked to open the eyes if they woke up during night.

B. Preprocessing and Feature Extraction

The feature extraction step consisted of calculating power spectral density (PSD) estimates of raw ECG and respiratory signals. PSD estimation methods are widely used for this purpose in biomedical signal analysis [17]. The periodogram method that we have chosen cuts the original signals into equally sized segments s. Each segment s is windowed with a Hamming window w to reduce the effects of spectral leakage at the first side lobe. For each segment a periodogram $\hat{\mathbf{S}}$ is then calculated with an FFT, as follows

$$\hat{\mathbf{S}} = |FFT\{s(m)w(m)\}|^2 \qquad m = 0, 1, \ldots, N-1 \qquad (1)$$

where N is the number of samples inside a segment s. The obtained pattern $\hat{\mathbf{S}}$ gives an estimation of the frequency content in the given time segment. Computationally efficient FFT calculation in a digital signal processing (DSP) microcontroller requires $N = 2^l$ where l is a positive integer. The size of N ($N_{ECG} = 4096, N_{Resp} = 2048$) was selected according to this criterion and corresponds to a segment length of 40.96 seconds. It is the next larger possible segment size compared to the traditional segment length of 30 seconds of PSG. Increasing the segment size would increase the resolution of the PSD, but also considerably increase the computational and memory costs in a microcontroller.

As the input values are real numbers, the PSD output is symmetrical around the DC component and it is sufficient to use half of the output points ($N/2$). The DC component is eliminated, because it contains mainly the offset of the uncalibrated sensors. High frequency components of the signals ($>$10 Hz for ECG and $>$8 Hz for Resp) did not show variations in the spectrogram and are removed to reduce the size of the input vector from $N/2$ to n ($n_{ECG} = 409, n_{Resp} = 327$).

C. Neural Classifier

We used a feed-forward ANN with no hidden layers and one single output unit (see Fig. 1, ANN classifier). To train the ANN and update the synaptic weights we use the Levenberg-Marquardt backpropagation algorithm [18][1]. We studied three different architectures, which differed in the type of input signal. The input vector of the first architecture was composed of the logarithm of the periodograms $\hat{\mathbf{S}}_{ECG}$ and $\hat{\mathbf{S}}_{Resp}$ (Fig. 1). The other two architectures use only the frequency content of one of the two signals, ECG or respiratory effort, respectively.

Initialization of the weights is done with the Nguyen-Widrow method [19]. The output of the neuron is thresholded so that $y(x) \geq 0$ is mapped to sleep and $y(x) < 0$ is mapped to wake. To train the networks, the data were divided into

[1]The parameters are: μ: 0.001; μ increase: 10; μ decrease:0.1; μ max: 10^{10}; min gradient: 10^{-10}; max validation failures: 30.

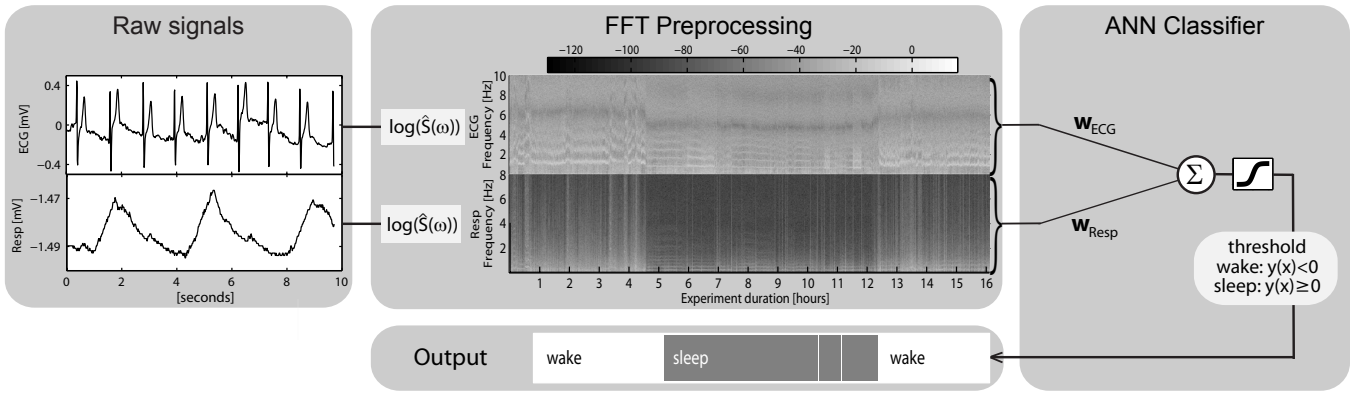

Fig. 1. Overview of the sleep/wake classification system. Raw ECG and respiratory effort signals are projected to the frequency space with the help of a FFT. The resulting frequency data (represented here by a spectrogram) are fed to a feed-forward, single-layer ANN with a threshold.

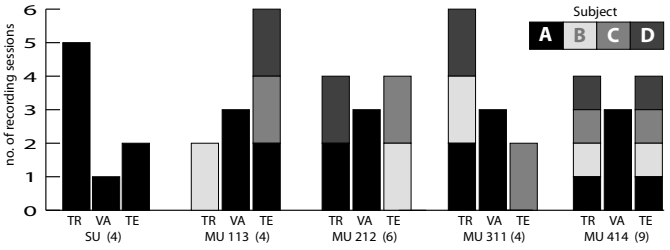

Fig. 2. Experimental design for training the neural classifier. SU = trained and tested on single user. MU = trained and tested on multiple users. Numbers indicate users in TR, users in VA, users in TE (nr. of repetitions with different combinations of users/sessions in training and testing).

three sets: training, validation and test. The training set (TR) contains the data used to update the synaptic weights. The performance of the network is evaluated on the validation set (VA) after each iteration and the training is stopped if the performance decrease of the validation pattern exceeds the maximal validation criterion. The test set (TE) is used to measure the performance of the network after the training.

1) Single User Experiments: With this set of experiments, we investigated the performance of the method when trained and tested on the same person. We used subject A, for whom we had the highest number of recording sessions. The 8 available sessions were randomly divided into a training set of 5 sessions, a validation set of 1 session, and a testing set of 2 sessions. 5 independent runs were performed from different initial weight values. In order to prevent performance biases due to the choice of sessions used for training and testing, we repeated the experiment 4 times with different sessions in the training and testing set.

2) Multi User Experiments: With this set of experiments, we investigated the performance of the method when trained on a single person and tested on multiple persons, and when trained on multiple persons and tested on multiple persons. Four experiments were carried out, each with an increasing number of persons in the training set (1 to 4) and all remaining persons in the testing set (in the only case when all 4 persons were in the training set, we made sure that the two sets contained different recording sessions). The validation set used 3 sessions from subject A that were never used for training and testing; the validation set was equal for all four experiments. 5 independent runs of each experiment were performed from different initial weight values.

In order to prevent performance biases due to the choice of sessions, we repeated each experiment with all possible combinations of sessions in the testing and training set, making sure that the same session did not appear both in the training and in the testing set (the number of repetitions for each experiment is indicated between brackets in Fig. 2).

III. Results and Discussion

In the single user condition, the classification accuracy (percentage of correct classifications on the test set) of the networks using both ECG and respiration signals was much better than that of the networks using only ECG signals (95.4% vs. 91.7%, $p<0.001$, t-test) and slightly better than the networks using only respiration signals (95.4% vs. 93.3%, $p<0.01$, t-test) (Table I, top row).

The only other study where both ECG and respiration signals from a single subject were combined, reported an accuracy of 81% [15], but a) that value was measured in the more difficult task of classifying wake, sleep, and REM sleep; b) it used data obtained in controlled hospital environments using PSG equipment; c) it used only data from night recordings; and d) it used a computationally expensive pre-processing algorithm calculating 27 features, which may be difficult to implement in a low-power and wearable system.

In all four multi-user conditions, the accuracies dropped with respect to the single-user conditions. The accuracy drop was the largest for the networks using only ECG signals (Table I, first column). Furthermore, the networks using both ECG and respiration signals (Table I, third column) displayed lower accuracy than the networks using only respiration signals (Table I, second column) ($p<0.001$ for all cases, t-test). Also, the variability of the accuracy across multiple replications was double when the respiration was used in combination with ECG with respect to the condition when respiration alone was

TABLE I
SINGLE USER (SU) AND MULTI USER (MU) TEST CLASSIFICATION ACCURACIES AND STANDARD DEVIATIONS

Experiments	ECG	Resp	ECG + Resp
SU	91.67% ± 2.74	93.27% ± 1.48	95.42% ± 1.61
MU 113	55.46% ± 9.42	86.59% ± 0.73	78.10% ± 4.00
MU 212	67.63% ± 10.45	88.77% ± 1.46	83.53% ± 2.33
MU 311	69.35% ± 14.49	89.41% ± 1.66	86.02% ± 3.48
MU 414	77.66% ± 3.41	90.86% ± 1.08	87.38% ± 2.87

used. These results suggest that ECG signals have unique features that are specific to each individual and do not generalize to other individuals for the purpose of discriminating sleep and wake states.

Our data sets are not sufficiently large to draw conclusions on the observed differences among the four multi-user conditions, although one may notice a positive correlation between accuracy and number of users in the training set.

The accuracies in the multi-user conditions are comparable to the 91% accuracies in actigraphy studies with only accelerometer signals [7] because in that case too, the measures are available for multiple users. However, as we mentioned in the introduction, actigraphy may mis-classify wake periods of low activity (reading, watching TV) as sleep periods. Indeed, the accuracy of the actigraphy methods was only 44% during the wake periods, whereas in our case it was 92.09% ±6.49 when only the respiration signal was used and 93.48% ±4.30 when both ECG and respiration were used (results from condition MU 311, which is the most similar to the condition used in actigraphy studies).

IV. CONCLUSION

The method and results presented in this paper indicate that the combination of ECG and respiratory signals can discriminate with high accuracy between sleep and wake states for an individual user. The choice of signal pre-processing and classification makes it possible to implement the method in a low-power dsPic33 microcontroller from Microchip (16 bit, 40 MIPs, 16 kB RAM, 200 mW). With this microcontroller, the presented calculation can be done with less than 150k instructions (3.75 ms), which fits largely in a 10 ms window between two sampling instructions of ECG.

The method requires a preliminary stage of labeling recorded data into sleep and wake states, which only few categories of persons (sportsmen, high-endurance workers, pathological cases, etc.) may be willing to do. For a wider use of the system, it would be desirable to pre-train the system on a limited number of persons and use it in generalization mode for several other users.

Although the accuracy of the method in multi-user conditions is lower, it is comparable to actigraphy methods and it is much better than those methods in its consistency during wake and sleep periods. Furthermore, our results indicate that respiratory signals alone are sufficient (and even better then combined respiratory and ECG). Respiratory signals are easier to measure because they do not require electrodes on the skin, and persons may wear them for periods of several days and weeks. Therefore, we think that this method represents a very promising solution for continuous monitoring of sleep and wake states. Our current work consists in the implementation of the algorithm into a DSP that estimates the sleepiness of the person with the help of additional models mentioned in the introduction.

V. ACKNOWLEDGMENTS

The authors thank Jean-Christophe Zufferey and Sara Mitri for several useful discussions. We would also thank all subjects who accepted to participate in this study. A special thank goes to Dr. med Werner Karrer, Dr. med Thomas Rote and Isabelle Arnold of the Luzerner Höhenklinik Montana, Switzerland for their help with physiological signal recording and sleep analysis.

REFERENCES

[1] T. Akerstedt, "Consensus statement: fatigue and accidents in transport operations." *J Sleep Res*, vol. 9, no. 4, p. 395, 2000.
[2] P. Achermann and A. A. Borbely, "Mathematical models of sleep regulation," *Front Biosci*, vol. 8, pp. 683–93, 2003.
[3] A. Rechtschaffen, A. Kales, R. Berger, and W. Dement, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," *Public Health Service, US Government Printing Office*, 1968.
[4] T. Penzel and R. Conradt, "Computer based sleep recording and analysis," *Sleep Med Rev*, vol. 4, no. 2, pp. 131–148, 2000.
[5] A. Sadeh and C. Acebo, "The role of actigraphy in sleep medicine." *Sleep Med Rev*, vol. 6, no. 2, pp. 113–124, 2002.
[6] R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin, "Automatic sleep/wake identification from wrist activity," *Sleep*, vol. 15, no. 5, pp. 461–9, 1992.
[7] L. de Souza, A. A. Benedito-Silva, M. L. Pires, D. Poyares, S. Tufik, and H. M. Calil, "Further validation of actigraphy for sleep studies," *Sleep*, vol. 26, no. 1, pp. 81–5, 2003.
[8] (2007) Sleeptracker. [Online]. Available: http://www.sleeptracker.com
[9] (2007) Axbo shop. [Online]. Available: http://www.axbo.com
[10] R. D. Ogilvie, "The process of falling asleep," *Sleep Med Rev*, vol. 5, no. 3, pp. 247–270, 2001.
[11] M. H. Bonnet and D. L. Arand, "Heart rate variability: sleep stage, time of night, and arousal influences," *Electroenceph clin Neurophysiol*, vol. 102, no. 5, pp. 390–396, 1997.
[12] S. Telser, M. Staudacher, Y. Ploner, A. Amann, H. Hinterhuber, and M. Ritsch-Marte, "Can one detect sleep stage transitions for on-line sleep scoring by monitoring the heart rate variability?" *Somnologie*, vol. 8, no. 2, pp. 33–41, 2004.
[13] Z. Shinar, S. Akselrod, Y. Dagan, and A. Baharav, "Autonomic changes during wake-sleep transition: A heart rate variability based approach." *Auton Neurosci*, vol. 130, no. 1-2, pp. 17–23, 2006.
[14] T. F. of the European Society of Cardiology and the North American Society of Pacing Electrophysiology, "Heart rate variability : Standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.
[15] S. Redmond and C. Heneghan, "Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea," *IEEE Trans Biomed Eng*, vol. 53, no. 3, pp. 485–496, 2006.
[16] S. J. Closs, "Assessment of sleep in hospital patients: a review of methods." *J Adv Nurs*, vol. 13, no. 4, pp. 501–510, 1988.
[17] A. Cohen, *Biomedical Engineering Handbook*, 3rd ed., ser. The electrical engineering handbook series. Boca Raton : CRC Taylor & Francis, 2006, vol. 2, ch. Biomedical Signal Analysis, pp. I1–I22.
[18] M. Hagan and M. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.
[19] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," in *International Joint Conference on Neural Networks*, vol. 3, 1990, pp. 21–26.