

Improving Actigraph Sleep/Wake Classification with Cardio-Respiratory Signals

Walter Karlen, Claudio Mattiussi and Dario Floreano, *Senior Member, IEEE*

Abstract—Actigraphy for long-term sleep/wake monitoring fails to correctly classify situations where the subject displays low activity, but is awake. In this paper we propose a new algorithm which uses both accelerometer and cardio-respiratory signals to overcome this restriction. Acceleration, electrocardiogram and respiratory effort were measured with an integrated wearable recording system worn on the chest by three healthy male subjects during normal daily activities. For signal processing a Fast Fourier Transformation and as classifier a feed-forward Artificial Neural Network was used. The best classifier achieved an accuracy of 96.14%, a sensitivity of 94.65% and a specificity of 98.19%. The algorithm is suitable for integration into a wearable device for long-term home monitoring.

I. INTRODUCTION

Monitoring sleep behavior over several days or weeks is essential for sleep and circadian rhythm studies. Clinical polysomnography (PSG) [1] can rarely be used in these studies, because it is expensive and the EEG measurement is not reliable over longer time periods [2]. Further, placing electrodes and sensors for PSG is complicated, and wearing them is uncomfortable. For this reason, actigraphy is used by many medical examiners and researchers as an ambulatory sleep detection method.

Actigraphy is a method which uses accelerometers worn on the body to measure its acceleration and estimate the level of human activity. In sleep research applications the periods of low activity are considered as sleep. Actigraph devices are very practical, because of their low power consumption and the small size of the electronic components. Actigraphs for sleep research are composed of a multi-axis accelerometer, a processing unit, a storage memory and a power supply. In general these are integrated into a watch-like frame and worn on the wrist. Different ways of measuring activity from accelerometer data by the processing unit have been used [3]. Most often, the activity is calculated for 1-minute segments and stored every minute in an internal memory. Current devices can store approximately 45 days of continuous data. After having downloaded the data for visualization to a PC, automatic classification algorithms can be used to estimate sleep and wake times. Cole's model [4] is commonly used for the estimation. This model based on a weighted sum of the activity of the current (A_0), the 4 previous (A_{-x}) and 2

future (A_{+x}) 1-minute activity epochs (1).

$$D = w_{-4}A_{-4} + w_{-3}A_{-3} + w_{-2}A_{-2} + w_{-1}A_{-1} + w_0A_0 + w_{+1}A_{+1} + w_{+2}A_{+2} \quad (1)$$

This kind of algorithm is optimized with logistic regression by updating the weights $w_{-4...+2}$ and using a constant threshold for D . It is designed for offline use, as it makes use of future data. Other algorithms like Artificial Neural Networks (ANN) [5] or statistical methods [6] have also been suggested.

It has been shown by [7] that actigraphy is not as accurate as PSG, especially for activities with low movement, like reading or watching television. This kind of activities are misclassified by most existing actigraphy algorithms. In particular, sleep time is overestimated for subjects with low daily activity patterns, which may lead to wrong diagnosis. Further, even if actigraphy gives a good estimation of global sleep patterns, it is not accurate enough to detect sleep related disorders [2], [8].

Alternatively, cardio-respiratory signals have been previously used for sleep/wake classification. Redmond and Henegan [9] used ECG and respiration data to separate the three states of wake, non-REM sleep and REM sleep and obtained 79% accuracy compared to PSG. However, their algorithm was not intended for long-term studies and relied on controlled clinical data.

In our previous work we used the power spectral density features of cardio-respiratory signals to separate wake from sleep using a single-layer ANN model [10], [11]. The achieved accuracy of 89.4% was comparable to the accuracy reported for actigraphy studies. However, we were able to reduce the amount of wrongly classified wake intervals considerably (from 58% in [7] to 8%).

In this paper we propose a system that uses a combination of accelerometer and cardio-respiratory data to reduce classification errors. The system is small, low power and easy to wear and permits onboard classification, enabling the display of the estimated sleep/wake state to the wearer, who might want to correct the classifier decision. We make use of a Fast Fourier Transformation (FFT) for pre-processing the signal, and a feed-forward, single-layer ANN classifier model. With the input data size listed in Table I, both the FFT and the ANN output can be efficiently calculated on modern low-power DSP microcontrollers (more details on an actual DSP platform are given in [11]). Beside the online classification, the presented architecture offers the possibility of recording respiratory parameters with a single integrated

Manuscript received April 16, 2008; revised June 25, 2008. This work was supported by the Solar Impulse project grant of Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

The authors are with the Laboratory of Intelligent Systems, Institute of Micro-engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (e-mail: {name.surname}@epfl.ch).

system. This can be helpful for clinicians to diagnose sleep-related breathing disorders like obstructive sleep apnea in ambulatory examination.

II. METHOD

A. Recording Device

The Equivital physiological monitoring system (Hidalgo Ltd, UK) was used as a recording device. It measures a 2-lead electrocardiogram (ECG), respiration effort, 3D acceleration and temperature [12]. This device has been chosen because of the high integration of sensors in a single belt worn across the upper chest area, which facilitates the correct wearing and use of the system by inexperienced users. The system uses an open communication protocol which facilitates interface design. The belt contains 3 dry ECG electrodes which are sampled at 256 Hz and in 10-bit resolution. The resistive strain gauge sensor for respiratory effort and the accelerometers are sampled at 25.6 Hz and data is also available in 10-bit resolution.

Additionally, during bedtime, an electrooculogram (EOG) and a video of the upper part of the body were recorded with other devices in order to label the users' state as wake or sleep by a technician.

B. Recording Sessions and Subjects

A recording session consisted of wearing the recording device for approximately 12 hours during daily activities including sleeping. Three volunteer healthy male subjects with a mean age of 27 (± 2) years participated in the study. Subject 1 wore the device for 7 sessions, the two others for 2. The total sleep time recorded was 72.21 hours and wake 63.24 hours.

C. Data analysis

When the recording device failed to give any data, the missing data segments were discarded. However, signals from movement artifacts or other task-dependent disturbances were not discarded, since they might contain useful information for the classification task.

To determine if the subject was asleep or not, a manual video analysis was performed. The video was divided into segments of 10 seconds and each segment was evaluated using the following precise criteria derived from [4], [13], [14]:

- 1) The person is considered to be awake if his eyes are open or body movements occur for more than a segment of 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 signal was examined.
- 3) In doubtful cases, where the EOG signal could not help to clearly identify sleep, the state was set to awake.

To avoid undetected wakefulness with closed eyes, the subjects were asked to open their eyes if they woke up during the night. The video and EOG data were synchronized with the Equivital signals using timestamps from the recordings

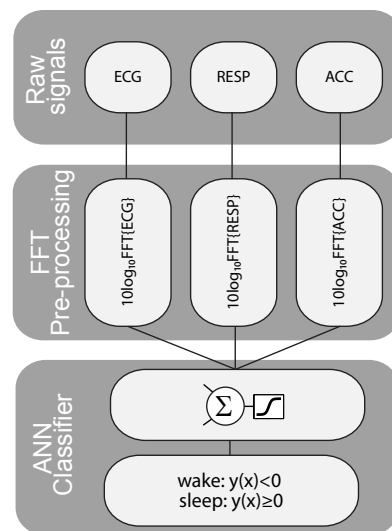


Fig. 1. Architecture of the classification algorithm A2. The FFT of ECG, respiratory effort (RESP) and acceleration (ACC) is fed to a single-layer ANN.

and by comparing the accelerometer orientation changes to the changes in position of the subject on the video.

D. Pre-processing

The feature extraction method for ECG and respiration signals has been adapted from [11]. We calculated the logarithm of the power spectral density (PSD) estimates \hat{S} of ECG and respiratory signals with a FFT according to (2).

$$\hat{S} = 10 \log_{10} |FFT \{s(m)w(m)\}|^2 \quad (2)$$

$$m = 0, 1, \dots, N - 1$$

The size of the segment s has been chosen so that the number of samples inside a segment corresponds to $N = 2^l$, where l is a positive integer. The ECG signal is subsampled to 51.2 Hz to reduce the segment size and decrease the computational time. The segment length was chosen to be 20 seconds. This corresponds to 1024 datapoints for ECG and 512 datapoints for respiration in one segment¹. w is a Hamming windowing function to reduce the effects of spectral leakage at the first side lobe [15]. As the signal is real-valued, the \hat{S} is symmetrical around the DC component and it is sufficient to use half of the spectrum points ($N/2$), as shown in Table I.

For the accelerometer data, two different pre-processing methods have been considered. The first method consisted of the same procedure as for the respiration signal described previously ($N = 512$). In the second method we applied a 0.5-11Hz bandpass filter on the accelerometer signal to remove slow (gravitational) movements and excessive accelerations. From the filtered signal the feature extraction algorithm counted the number of times the signal crossed the zero acceleration line within each 1-minute segment (zero-crossing, [3]). This type of pre-processing has been used for creating the inputs A_x of (1) by [7].

¹On a dsPic33 from Microchip (16-bit DSP microcontroller) the FFT of 1024 (512) datapoints takes approximately 113K (52K) instruction cycles.

TABLE I
ARCHITECTURE COMPOSITION AND TEST SET CLASSIFICATION RESULTS

Architecture	Signal and pre-processing				number of inputs ANN	normalized training time	Experiment 1 Accuracy [%]			Experiment 2 Accuracy [%]		
	ECG FFT	RESP FFT	ACC FFT	ACC Zero-crossing			Best	Mean	SD	Best	Mean	SD
A1	X	X			770	34	95.67	93.08	1.77	95.24	86.29	5.12
A2	X	X	X		1027	50	96.14	94.23	1.65	96.5	88.59	6.66
A3	X	X		X	777	36	95.93	93.48	2.14	95.54	85.68	7.22
A4			X		257	19	94.58	92.41	1.68	93.97	86.97	6.38
A5				X	7	1	91.14	84.81	5.77	88.07	73.87	15.25
size of N (2)	1024	512	512	7*1536								

E. Neural classifier

We studied 5 different classifier architectures. All classifiers were based on single-layer, feed-forward ANNs, but they differed in the type of input structure. The architecture A1 had as input the ECG and respiration PSD as done previously in [10]. Architecture A2 additionally used the PSD of the corresponding accelerometer segment (Fig. 1). Architecture A3 used the seven zero-crossing values $A_{-4\dots+2}$ from (1) instead of the accelerometer PSD in A2. The inputs of architecture A4 and A5 were only the frequency content of the accelerometer, or the zero-crossing values respectively. The size of the ANN input for each architecture and its normalized computational time for the training is shown in Table I.

The output unit of the ANN was a single neuron with a tangent-sigmoid transfer function that is thresholded so that $y(x) \geq 0$ is mapped to sleep and $y(x) < 0$ is mapped to wake. To train the ANN and update the synaptic weights we used the Levenberg-Marquardt backpropagation algorithm [16]². Initialization of the weights for each training run was done with the Nguyen-Widrow method [17].

F. Training

To train the networks, the data was divided into three sets: training, validation and test. The training set (TR) contained the data used to update the synaptic weights. The performance of the network was evaluated on the validation set (VA) after each iteration and the training was stopped if the performance of VA did not increase for more than 15 training iterations or the minimal gradient was reached². The test set (TE) was used to measure the performance of the network with the best VA performance.

G. Experimental Setup

1) *Single user experiments*: This set of experiments was intended to analyze the performance of the different architectures. Two recording sessions from the subject with the highest number of sessions was chosen as TE, another as VA. The remaining four sessions from subject 1 were used for training. The sessions were reshuffled 7 times until every recording session of subject 1 was twice in the TE. For each permutation we made 10 independent runs, each with different initial weight values for the ANN. For each

architecture and permutation the mean, standard deviation and the best accuracy (3) of the TE was calculated (Table I).

$$accuracy = \frac{true\ sleep + true\ wake}{all\ sleep + all\ wake} \quad (3)$$

2) *Generalization experiments*: In a second experiment we investigated the generalization capabilities of the different architectures. We used the same set configurations as in the previous experiments and added the sessions of subject 2 and 3 to the TR, VA and TE, respectively. 11 permutations were generated until every recording session was twice in TE. Again we made 10 independent runs with different initial weight values for each permutation.

III. RESULTS AND DISCUSSION

We can observe that the highest accuracy is obtained with the architecture A2 that combines the FFT pre-processing of all three signals (Table I). The architecture A5 which uses the Cole algorithm was not able to reproduce the good results of other studies which used the same method. We have already observed in our previous study that the mean accuracy of all architectures of the generalization experiment is lower and standard deviation is higher than in experiment 1) [11]. However, the adding of more users does not significantly change the accuracy of the best training result.

To compare the results of the classifier in more detail, we calculated the sensitivity (4) and specificity (5) for the best classifier of each architecture and represented them in a Receiver Operating Characteristic (ROC) curve.

$$sensitivity = \frac{true\ sleep}{true\ sleep + false\ sleep} \quad (4)$$

$$specificity = \frac{true\ wake}{true\ wake + false\ wake} \quad (5)$$

ROC curves allow the assessment of the results of classifier data in which the classes are not equally distributed. The ROC curve is obtained when the classification threshold of the ANN output is varied from -1 to 1. The large black symbols in Fig. 2 are the output of TE with the classification threshold equal to 0 as used in the training.

The filled circle in Fig. 2 shows that the architecture A5 has a low specificity, which has been observed before in literature [7]. Although A5 uses a larger time window of seven times 1-minute segments for the classification, many wake segments could not be correctly classified. One explanation might be the different placement of the actigraph (chest instead of wrist) for sensor integration reasons. Studies

²The parameters are: μ : 0.001; μ increase: 10; μ decrease: 0.1; μ max: 10^{10} ; min gradient: 10^{-10} ; max validation failures: 15.

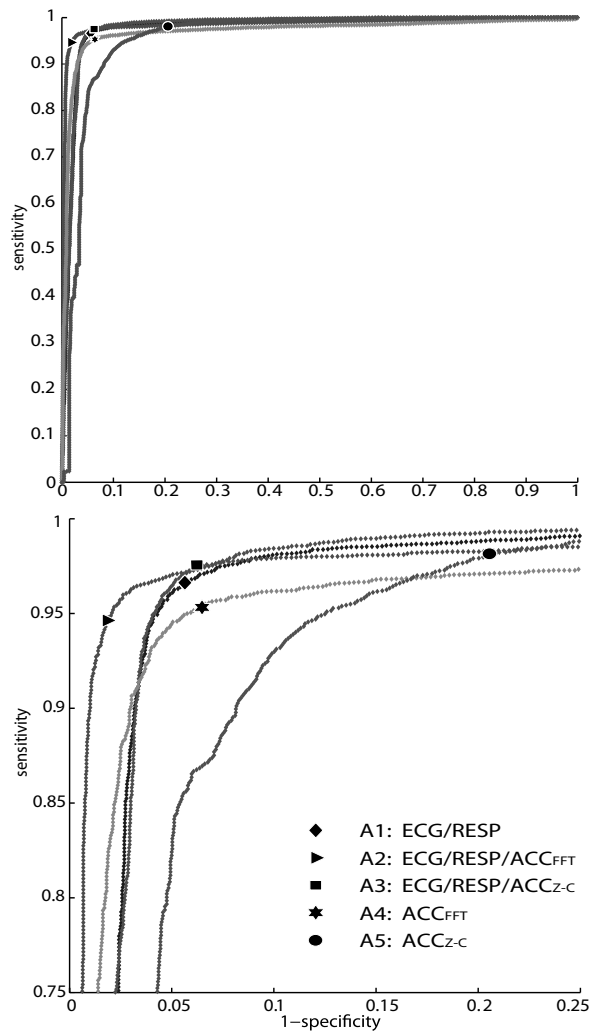


Fig. 2. Receiver Operating Characteristic (ROC) curves for the best classifier of each single user architecture A1 to A5. The ROC curve is obtained by varying the classification threshold of the ANN from -1 to 1. The large black symbols are the ROC points with the same fixed classification threshold used in the training. The output of an ideal classifier, where all wake and sleep states are correctly classified, would be located at (0,1). Bottom: Detailed view.

on different actigraph placements reported differences in actigraphy performance, which supports this hypothesis [18], [19]. Architecture A1 (diamond), which is using ECG and respiration signals is clearly better than A5 in detecting sleep and wake states. The addition of raw accelerometer data to the ECG and respiratory signals improves further the accuracy and also contributes to a better balanced sensitivity and specificity (A2 and A3). From all single user experiments, the best classifier was obtained with A2 and it reached an accuracy of 96.14%, a sensitivity of 94.65% and a specificity of 98.19% (Fig. 2, triangle).

IV. CONCLUSION

We demonstrated a new method for wearable long-term sleep/wake detection based on cardio-respiratory and accelerometer signals. The proposed algorithm can be implemented into modern low-power microcontrollers. The advantages of the proposed method compared to pure actigraphy

methods are a higher specificity (correct wake detection rate) and the availability of physiological signals for sleep disorder detection. Further validation with data from elderly people and sleep disordered patients is needed.

V. ACKNOWLEDGMENT

The authors would like to thank Dr. med Werner Karrer and his team at the Luzerner Höhenklinik Montana, Switzerland for their help with physiological signal recording and sleep analysis. The authors would also thank all subjects who accepted to participate in this study and Sara Mitri for the feedback.

REFERENCES

- [1] C. Iber, A. Chesson, and S. Quan, *The AASM manual for the scoring of sleep and associated events : rules, terminology and technical specifications*, westchester, il. ed., American Academy of Sleep Medicine, 2007.
- [2] S. Ancoli-Israel, R. Cole, C. Alessi, M. Chambers, W. Moorcroft, and C. P. Pollak, "The role of actigraphy in the study of sleep and circadian rhythms." *Sleep*, vol. 26, no. 3, pp. 342–392, May 2003.
- [3] S. Gorny and J. Spiro, "Comparing different methodologies used in wrist actigraphy," *Sleep Review*, vol. Summer, pp. 40–42, 2001.
- [4] 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.
- [5] S. Edward, S. Nadezhda, S. Stephanie, N. Michael, and C. S. Group, "Activity-based sleep/wake identification in infants," *Physiological Measurement*, vol. 25, no. 5, p. 1291, 2004, 0967-3334.
- [6] G. Jean-Louis, H. von Gizycki, F. Zizi, J. Fookson, A. Spielman, J. Nunes, R. Fullilove, and H. Taub, "Determination of sleep and wakefulness with the actigraph data analysis software (adas)," *Sleep*, vol. 19, no. 9, pp. 739–43, Nov 1996.
- [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] A. Sadeh and C. Acebo, "The role of actigraphy in sleep medicine." *Sleep Med Rev*, vol. 6, no. 2, pp. 113–124, 2002.
- [9] 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.
- [10] W. Karlen, C. Mattiussi, and D. Floreano, "Sleep and wake classification with ecg and respiratory effort signals," *IEEE Trans Biomed Circuits Sys*, 2008, submitted.
- [11] —, "Adaptive sleep/wake classification based on cardiorespiratory signals for wearable devices," *Biomedical Circuits and Systems Conference, 2007. BIOCAS 2007. IEEE*, pp. 203–206, 27-30 Nov. 2007.
- [12] (2008) Equivital physiological monitoring system. [Online]. Available: <http://www.equivital.co.uk>
- [13] R. D. Ogilvie, "The process of falling asleep," *Sleep Med Rev*, vol. 5, no. 3, pp. 247–270, 2001.
- [14] S. J. Closs, "Assessment of sleep in hospital patients: a review of methods." *J Adv Nurs*, vol. 13, no. 4, pp. 501–510, 1988.
- [15] F. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, Jan 1978.
- [16] M. Hagan and M. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE Trans Neural Netw*, vol. 5, no. 6, pp. 989–993, 1994.
- [17] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," in *Int. Joint Conference on Neural Networks*, vol. 3, 1990, pp. 21–26.
- [18] J. J. V. Hilten, H. A. Middelkoop, S. I. Kuiper, C. G. Kramer, and R. A. Roos, "Where to record motor activity: an evaluation of commonly used sites of placement for activity monitors." *Electroencephalogr Clin Neurophysiol*, vol. 89, no. 5, pp. 359–362, Oct 1993.
- [19] H. A. M. Middelkoop, E. M. Dam, D. A. Smilde-Van den Doel, and G. Dijk, "45-hour continuous quintuple-site actimetry: Relations between trunk and limb movements and effects of circadian sleep-wake rhythmicity," *Psychophysiology*, vol. 34, no. 2, pp. 199–203, 1997.