Combining HRV Features for Automatic Arousal Detection

Jerome Foussier[1], Pedro Fonseca[2,3], Xi Long[2,3], Berno Misgeld[1], Steffen Leonhardt[1]

[1] Chair of Medical Information Technology, RWTH Aachen University, Germany
[2] Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands
[3] Philips Research, Eindhoven, The Netherlands

Abstract

Arousals are vital for sleep as they ensure its reversibility. However, an increased amount of arousals might indicate sleep disturbances or disorders. Since arousal events are similar to wake states but much shorter than the standard annotation epoch length of 30 s, they degrade sleep staging classification performance. Arousals are also related to physiological activities, such as cardiac activation, thus making the detection in a less disturbing way than with polysomnographies in sleep laboratories possible. Therefore, we analyzed 72 features derived from the heart rate variability (HRV) of 15 whole-night polysomnographic ECG recordings to quantify cardiac activation during sleep. After calculating the Mahalanobis distance (MD), ranking the best uncorrelated features and performing MANOVA, we show that combining multiple features increases the discriminative power (MD=1.56, χ^2=33117) to detect arousals during the night compared to the best single feature (MD=1.16, χ^2=16633). A linear mixed model is used to show between-subject effects and to validate the significance of each feature based on Wald test statistics.

1. Introduction

The American Sleep Disorders Association (ASDA) defines arousals as frequency shifts in the electrical activity of the brain in humans that can be acquired with electroencephalography (EEG). Increased arousals lead - same as for reduced sleep length - to daytime sleepiness [1]. In the past decades, arousals were found to be closely related to sleep disorders [2, 3]. However, they are also essential to ensure the reversibility of sleep [2]. It has been found that arousals and short awakenings are very often related to physiological changes, such as heart rate, pulse transit time [4] and blood pressure [2]. Also respiratory events during the night are susceptible to be strongly related to arousals [4]. Indirectly measuring arousals through physiological signals instead of using the EEG is much more convenient and helps transferring laborious clinical setups to the home environment to detect sleep disorders more easily. Additionally, the classification of sleep and wake epochs can be enhanced when knowing the presence of arousals [5, 6].

Sforza *et al.* [7] found the evidence of cardiac activation during arousals. They proposed a simple arousal detector which has been used by Mendez and Matteucci [5] to improve the classification performance of REM (rapid eye movement) and non-REM sleep stages. Other researchers, e.g. Trinder *et al.* [8], also examined the nature on cardiovascular activation during an arousal. The heart rate suddenly increases with the start of an arousal event and drops back to normal values (rebound effect) after a certain time depending on the arousal length. Based on this fact, Basner *et al.* [9] developed an algorithm to detect arousals by calculating beat-by-beat likelihood ratios. These indicate how likely a certain heart beat is to correspond to the start of an arousal.

Some researchers, e.g. Bonnet and Arand [10] and Blasi *et al.* [11] analyzed spectral changes in very low, low or high frequencies (VLF, LF or HF, respectively) of the HRV before and after arousals. More complex computations, such as detrended fluctuation analyses (DFA), entropy calculations or raw ECG examinations, can be found in the work of Vigo *et al.* [12], Noviyanto *et al.* [13], or Penzel *et al.* [14].

In our work we gathered the most relevant features from literature and analyzed their discriminative power. We developed and adapted known features solely based on ECG which allow the automatic and accurate detection of arousals during sleep. All employed features are summarized in Table 1.

2. Methods and materials

Data set

For this work, arousal events and sleep stages in polysomnographic data of 15 subjects without any known sleep disorder were annotated by sleep technicians according to the American Academy of Sleep Medicine (AASM) guidelines. Nine of the subjects were measured in Boston (USA), at the Sleep Health Center and the others in Eind-

Table 1. Description of all employed features

Features (Total 72)	#	References
VLF, LF, HF, LF/HF ratio, module and phase of HF pole	6	[5, 10, 12, 13, 15, 16]
short time LF, HF, LF/HF ratio, module and phase of HF pole	5	[5, 10, 12, 13, 15, 16]
mean, standard deviations and ranges of (detrended) heart rates and RR intervals	7	[5, 12, 13, 16]
10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles of (detrended) heart rates and RR intervals	20	[13]
Mean absolute deviation of (detrended) heart rates and RR intervals	4	[13]
pNN50	1	[13]
Likelihood ratios (mean, median, min, max)	4	[9]
Sample entropy 1 and 2 with scales 1-10	20	[12]
DFA parameters (α_1 , α_2, α , α_{al} and windowed DFA)	5	[12–14]

hoven (The Netherlands), at the sleep laboratory of the High Tech Campus.

Feature extraction and ranking

Based on the observation that heart rate changes are related to arousals, 72 HRV features known from literature were extracted. The instant heart rate was derived from the ECG signal with a QRS detector. Features have been computed in the time and frequency domain with a time step interval of one second. This short interval was chosen because arousals are annotated with a higher time resolution since they last for 3-15 s and can occur several times during regular sleep stages, normally annotated in 30 s epochs [1]. The trend removal procedure of the RR peaks intervals and heart rates described by Redmond *et al.* [16] was applied to reduce between-subject variability .

Furthermore, a feature selection (FS) algorithm based on Mahalanobis distance (MD) ranking was used to reduce the feature space and remove highly correlated features. A similar MD ranking algorithm has been proposed in earlier work for the sleep/wake classification task [17]. The MD is a distance metric, but unlike the Euclidean distance, it takes the covariances of multivariate data into account and is scale invariant. In this work the covariance matrices of the "arousal" and "no arousal" class are averaged to form the pooled covariance matrix.

The ranking process starts sorting the features by their MD values in descending order. Since a classification task is not performed in this work and hence no classification performance metric is available as in [17], we set the maximum allowed between-feature Pearson correlation to a fixed value of 0.9. As highly correlated features do not contain new information they can be discarded to ensure a maximum data diversity of the input data. Using non-correlated features within supervised learning also reduces the training bias of classifiers and helps preventing overfitting.

Feature evaluation

After having identified the power of each feature in discriminating the presence or absence of arousals, the multivariate analysis of variances (MANOVA) was applied to test whether the null-hypothesis - the two groups have the same mean values - can be rejected at a certain significance level p. Since we have repeated measures on the same subjects and MANOVA expects normally distributed, uncorrelated data, we also applied linear mixed model analysis to emphasize the discriminative power of each feature using the Matlab (The Mathworks, MA, USA) `nlme` package. The mixed model using between-subject difference analysis with fixed slopes [18], neglecting within-subject effects, is defined as:

$$y_{ij} = \beta_0 + \beta_1 \cdot \bar{x}_j + \beta_2 \cdot l_{ij} + u_{0j} + e_{0ij}, \qquad (1)$$

with $\bar{x}_j$ as the mean feature value and l_{ij} as the numeric class label of measurement i for subject j. The intercept β_0 and the estimated fixed slopes for each variable are represented by β_1 and β_2. The random intercept u_{0j} and residual error e_{0ij} are supposed to be normally distributed having zero mean and between-subject and within-subject variances $\sigma^2_{u_{0j}}$ and $\sigma^2_{e_{0ij}}$, respectively. The statistical significances of fixed and random effects are based on Wald statistical (WS) tests with one degree of freedom [18].

3. Results and conclusions

The results of the best 20 selected features are presented in Table 2. The first column indicates the ranking (1 means best). The MD between the "arousal" and "no arousal" class is listed in the second column. Each selected feature is also analyzed with a MANOVA, represented by the χ^2 value, and the estimated mixed model parameters of equation (1). The last column gives a short descriptive text of the feature. All other columns are the estimates from the mixed model in equation (1). The last line of this

Table 2. MD, χ^2, WS values of the mixed model (1) for the selected features. Standard deviations in parenthesis.

pos	MD	χ^2	WS β_0	WS β_1	WS β_2	$\sigma^2_{u_{0j}}$	$\sigma^2_{e_{0ij}}$	description
1	1.16	16633*	7.9	6446.2†	25072.1†	3.30 (0.03)	4167.86 (0.11)	10th percentile of detr. RR intervals
2	1.07	15173*	0.0	3517.5†	23161.3†	3.39 (0.03)	4533.14 (0.12)	10th percentile of RR intervals
3	0.97	35973*	11.9†	180.6†	42402.4†	0.00 (0.04)	0.03 (0.00)	max of likelihood ratios
4	0.70	6977*	9.5	15993.8†	16532.6†	0.00 (0.02)	0.00 (0.00)	mean absolute deviation of heart rate
5	0.63	3425*	2.8	1238.5†	4092.7†	0.00 (0.01)	0.92 (0.00)	windowed DFA
6	0.55	2261*	10.2	100610.7†	3484.9†	0.00 (0.01)	0.24 (0.00)	norm. LF band in PSD (short window)
7	0.55	3579*	1.1	19110.4†	3651.7†	0.00 (0.01)	0.00 (0.00)	module of HF pole (short window)
8	0.46	1646*	6.0	221264.2†	1957.3†	0.00 (0.00)	0.18 (0.00)	norm. VLF band in PSD
9	0.43	1549*	2.6	121205.4†	1739.5†	0.00 (0.00)	0.07 (0.00)	sample entropy 2 with scale 1
10	0.41	2067*	1.3	61511.5†	2066.1†	0.00 (0.00)	0.00 (0.00)	module of HF pole
11	0.41	1546*	0.1	97674.1†	1902.7†	0.00 (0.00)	0.04 (0.00)	scaling exponent α_1 of DFA
12	0.40	1835*	0.5	54487.8†	2740.1†	0.00 (0.01)	0.02 (0.00)	scaling exponent α_{al} of DFA
13	0.37	1264*	3.7	184113.6†	2087.7†	0.00 (0.01)	0.10 (0.00)	norm. HF band in PSD
14	0.36	1062*	1.3	80712.1†	1271.0†	0.00 (0.00)	0.08 (0.00)	sample entropy 2 with scale 2
15	0.35	1235*	0.8	33004.1†	1892.5†	0.00 (0.00)	0.04 (0.00)	scaling exponent α of DFA
16	0.29	648*	0.3	100638.8†	1126.1†	0.00 (0.00)	0.10 (0.00)	sample entropy 2 with scale 3
17	0.27	607*	0.1	58522.4†	998.2†	0.00 (0.00)	0.10 (0.00)	sample entropy 2 with scale 4
18	0.26	541*	0.0	39100.6†	943.4†	0.00 (0.00)	0.11 (0.00)	sample entropy 2 with scale 5
19	0.26	713*	6.0	529776.3†	356.6†	0.00 (0.00)	353.87 (0.03)	pNN50
20	0.24	453*	0.1	37937.8†	896.1†	0.00 (0.00)	0.13 (0.00)	sample entropy 2 with scale 6
P	**1.56**	**33117***						**Top 20 features pooled**

overview represents the results when pooling the best features. Values denoted with $*$ and † have $p \leq 0.001$ based on MANOVA and Wald statistical (WS) test, respectively.

In addition, three of the most representative histograms out of the top 20 features are shown in Fig. 1. To enhance readability of the graphs, the "no arousal" values have been mirrored along the x-axis, which shows the feature values. The ordinate is the relative occurrence in % in the total class distribution and the circles indicate the median values of each class. The histograms visually confirm the results obtained in Table 2. Especially for the first features it is noticeable that the median values of both classes are different. Most distributions follow approximately a normal or log-normal distribution. Special case is feature 3 (and also 19, but not shown), where zero-inflated distributions can be found. This is due to the fact that the likelihood ratios nearly have a binary behavior, so either 0 or 1 as output value. The ratios have already been processed by statistical means (see [9]). For the "arousal" class, more than 20% of the values are around 1. Between 0 and 1, about 30% for the "arousal" class and 10% for the "no arousal" class are present. Considering the MD, χ^2 and differences in median values in Fig. 1 and Table 2 it can be seen that the discriminative power diminishes rapidly towards the last feature of the list. Therefore the feature ranking based on the MD seems to be appropriate for this type of analysis.

Finally, the overall discriminative power significantly increases when combining the top 20 features. The combination improves the results to an MD of 1.56 and χ^2 of 33117, whereas the maximum obtained with the single most discriminating feature (10^{th} percentile of detrended RR intervals) is an MD of 1.16 and χ^2 of 16633. The WS

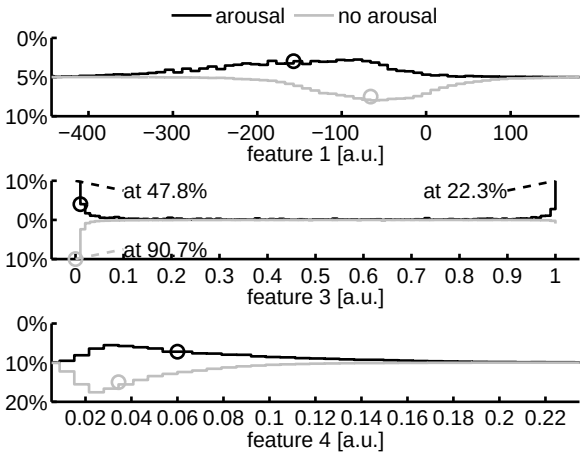

Figure 1. Representative histograms with relative occurrence in % and median values (circles) (black = "arousal" class, gray = "no arousal" class).

test of β_2 confirms the significance of discriminating between the "arousal" and "no arousal" classes. It is noticed that all are significant without considering between-subject centering effect (represented by β_1) on features. It also confirms the existence of between-subject effects, which are suggested to be removed, e.g. by normalization.

To conclude, we show in this work that the combination of multiple HRV features known from literature and extracted on a one second basis considerably improves the arousal detection capability compared to a single feature. Being able to detect arousals automatically only with the use of ECG measurements, which is more convenient than standard EEG acquisitions in sleep laboratories, it has the potential to introduce new techniques for screening and diagnosis of sleep related disorders.

Since arousals annotated with EEG signals might occur before or after the cardiac activation, it still has to be evaluated how large the time differences between the EEG and the onset of the different features are. In future work, this evaluation should be included in a training procedure for automatic arousal classification. Furthermore to improve the classification performance, post-processing criteria as defined by Mendez and Matteucci [5] or Bonnet *et al.* [10] (e.g. limiting the length of arousals) should be included to match the original ASDA definition of arousals [1].

References

[1] Bonnet M, Carley D, Carskadon M, Easton P, Guilleminault C, Harper R, Hayes B, Hirshkowitz M, Keenan S, Consultant MP, Roehrs T, Smith J, Weber S, Westbrook P, Jordan B. EEG arousals: scoring rules and examples. Sleep 1992; 15(2):173–184.

[2] Halász P, Terzano M, Parrino L, Bódizs R. The nature of arousal in sleep. Journal of sleep research March 2004; 13(1):1–23. ISSN 0962-1105.

[3] Bulckaert A, Exadaktylos V, De Bruyne G, Haex B, De Valck E, Wuyts J, Verbraecken J, Berckmans D. Heart rate-based nighttime awakening detection. European Journal of Applied Physiology May 2010;109(2):317–322.

[4] Poyares D, Guilleminault C, Rosa A, Ohayon M, Koester U. Arousal, EEG spectral power and pulse transit time in UARS and mild OSAS subjects. Clinical Neurophysiology October 2002;113(10):1598–1606. ISSN 1388-2457.

[5] Mendez M, Matteucci M. Sleep staging from Heart Rate Variability: time-varying spectral features and Hidden Markov Models. International Journal of Biomedical Engineering and Technology 2010;3:246–263.

[6] Fonseca P, Long X, Foussier J, Aarts R. On the Impact of Arousals on the Performance of Sleep and Wake Classification Using Actigraphy. In 35th Annual International Conference of the IEEE EMBS. Osaka (Japan): IEEE, 2013; 6760–6763.

[7] Sforza E, Jouny C, Ibanez V. Cardiac activation during arousal in humans: further evidence for hierarchy in the arousal response. Clinical Neurophysiology September 2000;111(9):1611–1619. ISSN 1388-2457.

[8] Trinder J, Allen N, Kleiman J, Kralevski V, Kleverlaan D, Anson K, Kim Y. On the nature of cardiovascular activation at an arousal from sleep. Sleep August 2003;26(5):543–551. ISSN 0161-8105.

[9] Basner M, Griefahn B, Müller U, Plath G, Samel A. An ECG-based algorithm for the automatic identification of autonomic activations associated with cortical arousal. Sleep October 2007;30(10):1349–1361. ISSN 0161-8105.

[10] Bonnet MH, Arand DL. Heart rate variability: sleep stage, time of night, and arousal influences. Electroencephalography and clinical neurophysiology May 1997;102(5):390–396. ISSN 0013-4694.

[11] Blasi A, Jo J, Valladares E, Morgan BJ, Skatrud JB, Khoo MCK. Cardiovascular variability after arousal from sleep: time-varying spectral analysis. Journal of Applied Physiology October 2003;95(4):1394–1404. ISSN 8750-7587.

[12] Vigo DE, Dominguez J, Guinjoan SM, Scaramal M, Ruffa E, Solernó J, Siri LN, Cardinali DP. Nonlinear analysis of heart rate variability within independent frequency components during the sleep-wake cycle. Autonomic Neuroscience Basic Clinical April 2010;154(1-2):84–88. ISSN 1872-7484.

[13] Noviyanto A, Isa S, Wasito I. Selecting Features of Single Lead ECG Signal for Automatic Sleep Stages Classification using Correlation-based Feature Subset Selection. International Journal of Computer Science 2011;8(5):139–148.

[14] Penzel T, Kantelhardt JW, Lo CC, Voigt K, Vogelmeier C. Dynamics of heart rate and sleep stages in normals and patients with sleep apnea. Neuropsychopharmacology July 2003;28(Suppl 1):S48–53. ISSN 0893-133X.

[15] Long X, Fonseca P, Haakma R, Aarts RM, Foussier J. Time-frequency analysis of heart rate variability for sleep and wake classification. In IEEE Conference on Bioinformatics and Bioengineering (BIBE), number November. Larnaca, Cyprus: IEEE. ISBN 9781467343589, 2012; 85–90.

[16] Redmond SJ, de Chazal P, O'Brien C, Ryan S, McNicholas WT, Heneghan C. Sleep staging using cardiorespiratory signals. Somnologie October 2007;11(4):245–256. ISSN 1432-9123.

[17] Foussier J, Fonseca P, Long X, Leonhardt S. Automatic Feature Selection for Sleep/Wake Classification with Small Data Sets. In 6th International Conference on Bioinformatics Models, Methods and Algorithms. Barcelona (Spain), 2013; 1–7.

[18] van de Pol M, Wright J. A simple method for distinguishing within- versus between-subject effects using mixed models. Animal Behaviour March 2009;77(3):753–758. ISSN 00033472.

Address for correspondence:

Jerome Foussier
MedIT, RWTH Aachen University
Pauwelsstrasse 20
D-52074 Aachen, Germany
foussier@hia.rwth-aachen.de