

Automated Sleep Arousal Detection Based on EEG Envelopograms

Filip Plesinger, Ivo Viscor, Petr Nejedly, Petr Andrla, Josef Halamek, Pavel Jurak

The Czech Academy of Sciences, Institute of Scientific Instruments, Brno, Czech Republic

Abstract

Background: Sleep arousal is basically described as a shift in EEG activity in frequencies > 16 Hz for a duration of > 3 sec (by the American Sleep Disorders Association – ASDA). The number of these arousals during sleep is a reflection of sleep quality. In accordance with the PhysioNet/CinC Challenge 2018, we present a method for automatic detection of arousals in polysomnographic recordings.

Method: Each file in the training dataset ($N=994$) has defined “Target Arousal Regions” (TAR, median length 33 seconds); however, arousals were usually located in the right half of these TARs. We built a method detecting EEG frequency shift to locate arousals inside ARs: envelopograms (14-20, 16-25 and 20-40 Hz) were inspected in a 3-sec floating window for an increase against a 10-sec background. We then extracted 133,573 blocks with such a shift from TARs ($N=38,628$) as well as outside TARs ($N=94,945$). We extracted 23 features from these blocks (how many EEG channels/frequency bands EEG frequency shift; heart rate before/during arousal; airflow and EMG changes) and trained a bagged tree ensemble model (70/30 % hold-out).

Results: The method showed AUPRC 0.27 on a training set and AUPRC 0.20 on a testing set ($N=989$).

1. Introduction

Sleep arousals disturb sleep and, therefore, affect sleep quality. Sleep arousals were defined by the American Sleep Disorders Association (ASDA) in 1992 as a shift in EEG activity in frequencies above 16 Hz for more than 3 seconds [1]. Multiple other rules were defined, for example that at least 10 seconds of continuous sleep must precede such an EEG frequency shift to score this as an arousal. General rules for scoring arousals were further improved by the American Academy of Sleep Medicine (AASM) [2].

According to the AASM guidelines, arousals are usually scored in a 30-second window. Because an arousal must be preceded by at least 10 seconds of continuous sleep, this setup allows a wide baseline area for consideration of EEG frequency shift. However, how intense such a frequency shift must be or how long this baseline area should be are not exactly defined.

The PhysioNet Challenge 2018 [3] was aimed at detection of target arousal regions (TAR). A TAR, according to the definition of the organizers, should be an arousal extended by 2 seconds prior to its start and prolonged by 2 or 10 seconds (depending on the kind of arousal). In this paper, we propose an automated method to recognize TARs in polysomnographic data.

2. Method

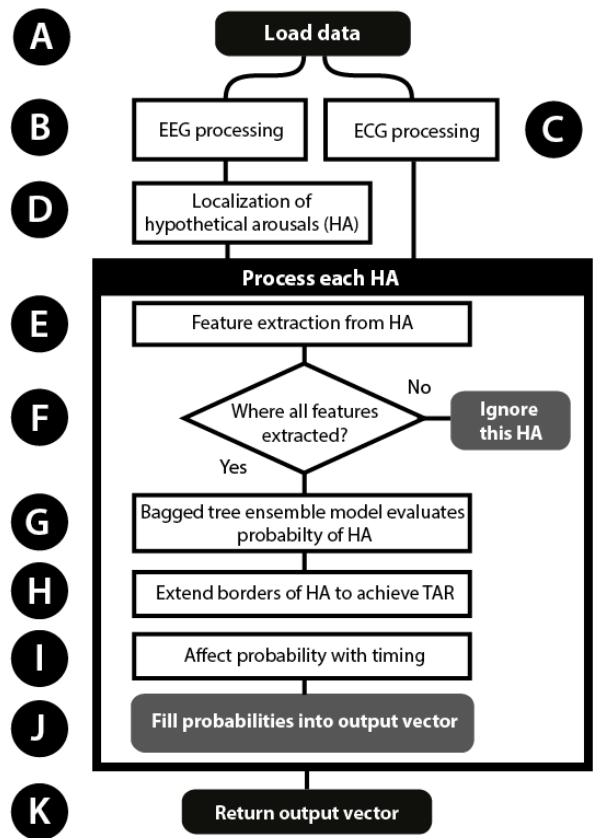


Figure 1. Method flowchart. EEG, EMG, ECG and airflow signals are loaded (A); envelopograms are prepared from EEG signals (B); QRS complexes are detected in ECG (C). Hypothetical arousals (HA) are detected in places of EEG frequency shift (D). For each HA, features are extracted and processed with a bagged tree ensemble (E-G). HA is extended to “Target Arousal Region” (H), modified by time (I) and added to output vector (J).

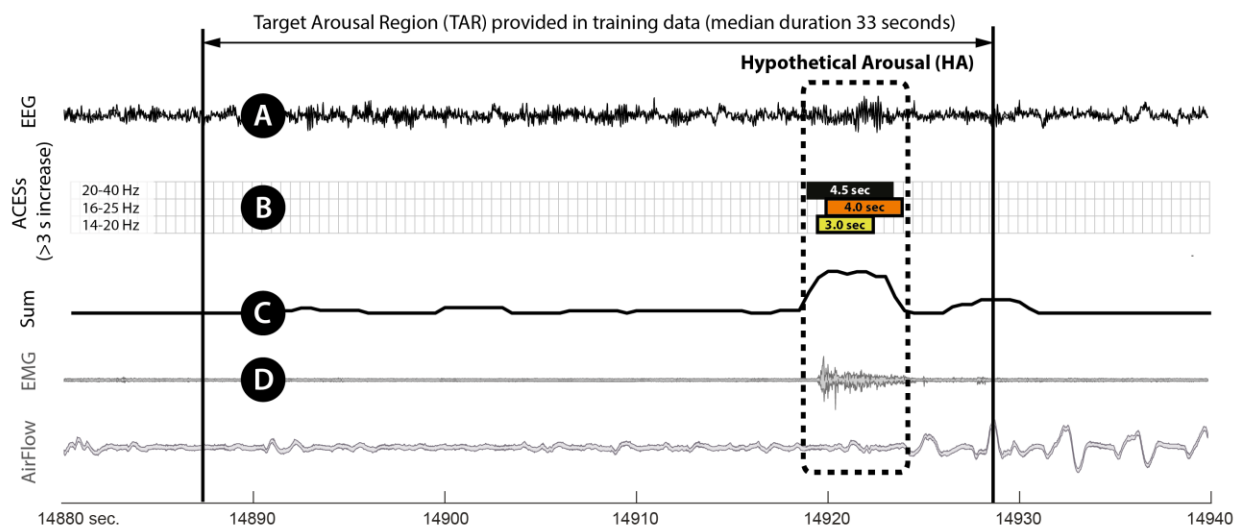


Figure 2. Method to find hypothetical arousals using information related to frequency shift in EEG. The EEG signal (A) was transformed into envelograms. Increases in EEG amplitude longer than 3 seconds were examined in three frequency bands (B). Next, this information is summed together from all envelograms (21) and summed together using increase duration as a weight (C). EMG (D) and Airflow channels are shown for convenience.

We used Matlab® 2016b with the Machine Learning and Statistics Toolbox and Signal Processing Toolbox.

The used public data consisted of a scored training dataset (N=994) and non-scored testing dataset (N=989) of whole-night polysomnographic data (fs=200 Hz). Both datasets were supplied by the Challenge organizers. Datasets contained 7x EEG channels, 1x EMG channel, 1x ECG channel, respiratory curve and SaO2 saturation curve.

The method flowchart is shown in Fig. 1. Since the elementary condition for sleep arousals is a shift in EEG activity, we decided to find such hypothetical arousals (HA) using EEG signals only. We then trained a bagged-tree-ensemble model computing arousal validity; this bagged tree ensemble used additional information from EMG, ECG and airflow. We extended arousal borders for compliance with the PhysioNet Challenge 2018 and, finally, the probability received from the bagged tree ensemble was modified by the general probability of an arousal at a specified time of sleep. The method output is a vector of the same length as the input data; this vector indicates the probability of a TAR.

2.1. Data pre-processing

Each EEG channel (N=7, Fig. 2A) was transformed into 3 envelograms (14-20 Hz, 16-25 Hz and 20-40 Hz) using Fast Fourier and Hilbert transformations. All envelograms (N=21) were down-sampled to 50 Hz. Next, for each envelogram we computed continuous change in

EEG amplitudes comparing a 3-s window to a 10-s window of preceding signal (step 0.5 s). Continuous changes higher than 1.1 were found; this threshold is arbitrary since no specific value was found in the literature. If such continuous change lasted at least 3 seconds, it was accepted as an arousal-compliant EEG shift (ACES, Fig. 2B). Finally, for each 0.5-second window we evaluated the sum of registered ACES from all envelograms (each ACES was multiplied by its duration). This sum creates a single curve with maxima pointing to hypothetical arousal (Fig. 2C).

2.2. Used features

We detected these HAs in TARs (N=38,628) in areas scored as 1. We also detected HAs outside TARs (Non-TARs, N=94,945) in areas scored as 0. A total of 23 features (Tab. 1) were extracted for each HA, leading to an unbalanced training dataset of 133,573 cases. Since 11 features were extracted from the EEG signal, we also extracted 7 features from ECG activity as heart rate or its changes; the rest were based on EMG and airflow signals. For QRS detection, we used a method designed for 1-lead Holter-ECG processing [4].

2.3. Model training

We trained a bagged tree ensemble model leaving 30 % of training samples for testing. The receiver-operator curve showed AUC 0.95 with values of

sensitivity and specificity of 0.79 and 0.94, respectively and accuracy of 90 %. However, it should be noted that this metric strictly describes ability of the classifier to identify real arousals in HAs; it is not the overall method performance. If any of the input features could not be computed, the HA was excluded from model preparation.

Table 1. List of extracted features. ACES: Arousal-Compliant EEG Shift; EEG: electroencephalogram; ECG: electrocardiogram; EMG: electromyogram; Breath – signal describing airflow.

No.	Feature description	Source
1	Max. weighted ACES sum	EEG
2	Number of ACES	EEG
3	Number of channels with ACES	EEG
4	Heart rate at ACES start	ECG
5	First derivative of #4	ECG
6	Heart rate at ACES end	ECG
7	First derivative of #6	ECG
8	Heart rate 10 s before ACES	ECG
9	Heart rate 10 s after ACES	ECG
10	Heart rate amplitude during ACES	ECG
11	Heart rate standard deviation	ECG
12	Heart rate std. 10 s before ACES	ECG
13	Ratio of std. deviation of heart rate 10 s before and 2 s after ACES maxima	ECG
14	Mean airflow amplitude during ACES	Breath
15	Mean airflow amplitude in 10-sec window before ACES start	Breath
16	Ratio of features #14/#15	Breath
17	Mean amplitude of Chin-Chin EMG signal during ACES	EMG
18	Mean amplitude of Chin-Chin signal 10 seconds before ACES beginning	EMG
19	Ratio of features #17/#18	EMG
20	Maximal ACES length	EEG
21	Number of channels with ACES in band 14-20 Hz	EEG
22	Number of channels with ACES in band 16-25 Hz	EEG
23	Number of channels with ACES in band 20-40 Hz	EEG

2.4. Computing target arousal regions

At this stage, the presented method was able to find and compute the probability of sleep arousals. The task of the Challenge was, however, to find TARs. Measured in the training dataset, the TARs duration median was 33.2 s; the median offset of TAR start measured to the sleep

arousal centre was 23.8 s and the median offset to TAR end was 8.6 s. Because we did not find any other logical link between TAR borders and sleep arousals, we extended each arousal centre to the left and right using these median values. Finally, the probability of each TAR was multiplied by the overall squared probability of sleep arousal (computed in 10-minute steps, Fig. 3).

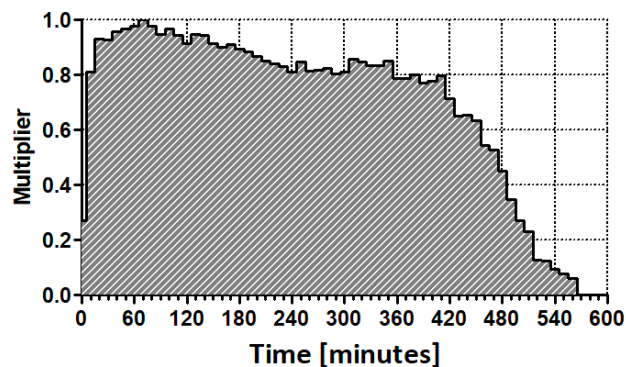


Figure 3. Squared probability of sleep arousal during sleep. These values multiplied the probability of arousal computed by the bagged tree ensemble.

3. Results

The results were evaluated on remote PhysioNet servers using a hidden dataset (Tab. 2). Performance in the Challenge 2018 was evaluated using area under precision-recall curve (AUPRC). During the evaluation on remote servers, it consumed (average) 3.93 % of the quoted running time.

Table 2. Method performance on training and testing datasets. AUROC – area under receiver operating characteristics; AUPRC – area under precision-recall curve.

	Training (N=998)	Testing (N=988)
AUROC	0.81	-
AUPRC	0.27	0.20

4. Discussion

The presented method implements just 23 features, for which reason it could, from the perspective of machine learning, be recognized as light-weight (this could also be measured by the amount of computing time shown in the section Results). An elementary rule requiring a shift in EEG frequency led us to a simple way of predicting areas of hypothetical arousals and we were quite successful in recognizing real arousals in these areas (AUC-ROC 0.95). On the other hand, the results computed when identifying

TARs on the training and testing set in Tab. 2 showed weaker AUROC performance indicating that our approach to recognizing HAs is not as powerful as classification of real arousals in HAs.

Moreover, we found it difficult to associate real arousals and TARs (and finding TARs was the goal of the challenge). Since the term “target arousal region” may be hard to find in the literature related to sleep arousals, definitions in the Challenge state that a TAR should specifically overlap sleep arousal. This overlap should be 2 seconds for arousal beginning and 2 or 10 seconds for arousal end (different for respiratory effort related arousals – RERAs and for non-respiratory effort related arousals – non-RERAs). We showed that the median length of TAR is 33 seconds (derived from the training set). This would mean that median EEG arousal length should not be shorter than 21 seconds which seems unlikely. From our observations, TARs may be associated with the position of the screen used during arousal scoring; this would be in accordance with AASM guidelines suggesting that arousals should be scored using a 30-second window on the screen.

We did not apply additional rules for scoring arousals. For example, we did not use a rule stating that a new arousal should not be scored if the distance from the previous one is shorter than 10 seconds [1]. Also, we did not distinguish between RERAs and non-RERAs; non-RERA arousals used different offsets for TARs. This should be reconsidered because 99.6% of target arousals were RERA (unfortunately, we found this information after the end of the official phase of the challenge). The presented method did not use the SaO2 channel showing oxygen saturation in the blood. The reason for this was that we considered this information too noisy. On the other hand, as shown on the sample entry prepared by the Challenge organizers (and also considering the fact that most arousals were RERA), this information could be successfully used for recognizing target arousals.

Reconsidering these missing points in the future could improve the results of the presented method.

5. Conclusion

We presented a method for automated evaluation of sleep arousals in polysomnographic studies. Although it showed high performance in finding sleep arousals in areas preselected using a shift in EEG frequencies, it achieved below-average performance when finding Target Arousal Regions (which was the task of the PhysioNet Challenge 2018).

Acknowledgements

This research was supported by project MSM 100651602 by the Czech Academy of Sciences and by project LO1212 by MEYS CR.

References

- [1] M. H. Bonnet, D. Carley, M. C. Consultant, P. E. Consultant, C. G. Chairman, R. Harper, B. Hayes, M. Hirshkowitz, S. Keenan, M. P. Consultant, T. Roehrs, J. Smith, S. Weber, P. Westbrook, A. Administrative, and S. Bruce, “EEG arousals: scoring rules and examples: a preliminary report from the Sleep Disorders Atlas Task Force of the American Sleep Disorders Association,” *Sleep*, vol. 15, no. 2, pp. 173–184, 1992.
- [2] Q. S. for the A. A. Iber C, Ancoli-Israel S, Chesson AL Jr., of S. Medicine., C. C. Iber, S. Ancoli-Israel, A. L. Chesson, S. F. Quan, and A. L. Chesson Jr., “The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications,” in *AASM Manual for Scoring Sleep*, 2007, pp. 1–59.
- [3] M. Ghassemi, E. B. Moody, L. H. Lehman, S. Christopher, Q. Li, H. Sun, R. G. Mark, M. B. Westover, and G. D. Clifford, “You Snooze, You Win: the PhysioNet/Computing in Cardiology Challenge 2018,” in *Computing in Cardiology Volume 45*, 2018, pp. 1–4.
- [4] F. Plesinger, P. Nejedly, I. Viscor, J. Halamek, and P. Jurak, “Parallel use of a convolutional neural network and bagged tree ensemble for the classification of Holter ECG,” *Physiol. Meas.*, vol. 39, no. 9, 2018.

Address for correspondence:

Filip Plesinger
Institute of Scientific Instruments of the CAS, v.v.i.
Kralovopolska 147,
Brno 612 64, Czech Republic

E-mail address: fplesinger@isibrno.cz