Automatic Scoring of Non-Apnoea Arousals Using the Polysomnogram

Nadi Sadr, Philip de Chazal

School of Electrical and Information Engineering, University of Sydney, Australia

Abstract

In this study, we developed a system that identifies the respiratory effort related arousals (RERA) and the non-RERA, non-apnoea arousals by processing four signals of the chin EMG signal, two channels of EEG signals (C4-M1 and C3-M2) and the oximetry signal. The 2 EEG signals were processed identically. Firstly, preprocessing was applied to remove the baseline wander, unwanted low frequency components and the abrupt changes. Then, the EEG signals were divided into non-overlapping epochs and a power spectral decomposition was calculated resulting in 5 PSD features per epoch. The Chin EMG signals were processed in the same fashion and resulted in 5 PSD features. The artefact signal of the SaO2 signal was removed and the square root of the standard deviation of the signal was calculated. The features were combined into a 16-element epoch. Following this, features from surrounding epochs were combined with the current epoch. We compared the performance of combining features from one to four epochs either side of an epoch. The 10-fold cross validation results of three classifiers including linear discriminant analysis (LDA), logistic regression (LR) and single hidden layer feed-forward neural networks (SHLN). The performance of our best system was an AUC 0.82 and an AUPRC of 0.24 using the 10 hidden units feed-forward neural network.

1.   Introduction

Sleep is vital for physical and mental health affecting neurocognitive, physiological and psychopathology functions and performance [1], [2]. Arousals are linked with sleep and interrupts the sleep states, forming a sleep/arousal loop [1]. Spontaneous arousals are part of a normal sleep/wake cycle [3], [4]. There are also different clinical conditions causing sleep fragmentation and arousals including respiratory sleep disorders and nocturnal myoclonus such as snoring, sleep apnoea (obstructive, central and mixed apnoea), hypopnoea, and periodic leg movement [1], [3], [4]. The interruptions in sleep and respiratory effort arousals may lead to serious health issues such as neurocognitive disorders, mood and mental conditions and cardiovascular disease [5].

Sleep staging and detecting arousals can be achieved using different physiological signals by extracting time and frequency domain information from the recorded signals during sleep [6]. A commonly used definition of arousal is provided by the American Academy of Sleep Medicine (AASM) indicating an abrupt minimum of 3-second increase in EEG frequency [7], [8]. Different algorithms have been studied in the literature for automatic sleep staging and detection of arousals using a variety of signals including EEG, EMG, and electrooculogram (EOG) [3], [6], [8]–[16]. Different features and methods were used including temporal and statistical features, time-frequency features, spectral decomposition of the EEG signals using a decision tree [9], and wavelet transform of the EEG data [3], [13]. The studies mostly proposed automated sleep scoring and detected arousals without discriminating different types of arousals including RERA, apnoea and non-apnoea types [3], [6], [8]–[16].

This study aims to detect the respiratory effort related arousals (RERA) and the non-RERA, non-apnoea arousals [17]. In this paper, we consider different signal processing methods to identify the two types of arousals using overnight recorded physiological signals including electroencephalography (EEG), chin electromyography (EMG), and oxygen saturation (SaO_2).

2.   Input Data

The dataset was provided by the PhysioNet/ Computing in Cardiology Challenge 2018 [17]. It was comprised of overnight polysomnographic (PSG) studies of 994 recordings as training set and 989 hidden test set data from 1985 subjects at the Massachusetts General Hospital (MGH) labs. There are nine signals including electrocardiogram (ECG), electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), and oxygen saturation (SaO_2). The signals were measured in microvolts and sampled at 200 Hz, except SaO_2 which was resampled to 200 Hz and measured in percentage. The MGH clinical staff annotated the sleep stages using the AASM manual for sleep scoring. The sleep scoring was provided for 30 second epochs with labels including wakefulness, stage 1, stage 2, stage 3, rapid eye movement (REM) and undefined. Also, the arousals were annotated by MGH sleep technologists into labels of spontaneous
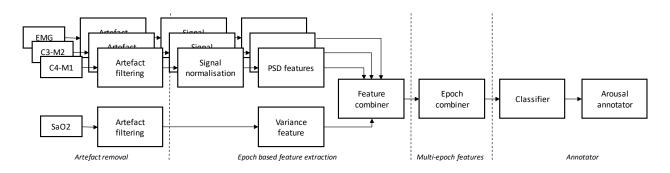
Figure 1. The block diagram of the proposed automatic arousal detection system.

arousals, RERA, bruxisms, hypoventilations, hypopnoeas, apnoeas (obstructive, central and mixed), vocalisations, snores, periodic leg movement, Cheyne-Stokes breathing and partial airway obstructions. The aim of the challenge was to identify the target arousals including RERA arousals and non-RERA, non-apnoea arousals. For the labelling the targets associated with a RERA arousals, the target began 2 seconds before the arousal up to 10 seconds after it ended. For the non-RERA, non-apnoea arousals, the targets began 2 seconds before a non-RERA arousal, up to 2 seconds after it ended. It should be noted that the apnoea or hypopnoea arousals that occurred within 10 seconds of a period of wakefulness were not scored. The target arousals were labelled as "1". There were also targets associated apnoea-arousals that were marked as "-1", which were not considered for scoring.

3.　Methods

Figure 1 depicts our automatic arousal annotator. It processes the chin EMG signal, two channels of EEG signals (C4-M1 and C3-M2) and the oximetry signal. It outputs an estimated probability of an arousal occurrence every $1/200^{th}$ of a second. Each stage will be discussed below.

3.1.　Mapping annotations to epochs

We used an epoch-based process system as all of our features required a time window of data in order to find an estimated value. The arousal annotations were event based and provided as a signal sampled at 200Hz. The event-based annotations were first mapped to our epochs using the following rules:
- If all event-based annotations were non-arousal for an epoch then epoch was labelled as 'non-arousal',
- if the duration of non-apnoea arousals in the epoch was greater than the apnoea arousals, the epoch was labelled as 'non-apnoea arousal',
- otherwise the epoch was labelled as 'apnoea arousal'.

3.2.　EEG and EMG signals

One of the conditions of the 1992 American Sleep Disorders Academy (ASDA) for identifying an arousal [8] is an abrupt shift in EEG signal which may include theta, alpha and/or frequencies above 16Hz but not spindles. The ASDA recommends using central or occipital EEG channels. Inspired by these scoring rules, we identified features that captured the power in the similar frequency bands and used these values as inputs to our classifier.

The 2 channels of central EEG signal (C4-M1, C3-M2) were processed in an identical fashion. Firstly, the signals were high-pass filtered with a median filter of width 0.25 seconds to produce a high-pass filtered signal. This removed baseline wander and unwanted low frequency components. Visually inspections of the high-pass signals revealed that there were many sections when signal abruptly changed. To correct these sections, we differentiated the high-pass filtered signal and where the absolute value of the differentiated signal exceeded a threshold, the high-pass filtered signal was replaced with a zero value. A threshold of 100 was found to provide satisfactory performance. These steps had the effect of replacing the sudden changes in the signal with zeroes. After these processing steps, the EEG signals were divided into non-overlapping epochs. A power spectral decomposition (Welch's periodogram method [18]) the EEG signal at 1 Hz resolution was calculated. Frequency bins were then combined as follows, 2-4Hz, 5-8Hz, 9-12Hz, 13-16Hz, and 17-32Hz. The values in the combined bins were then divided by the sum of the PSD for 2-32Hz and after applying a log transform these values used for features. Thus, for each EEG signal there were 5 features per epoch.

The Chin EMG signal was processed in an identical fashion to the EEG signals to yield 5 features per epoch.

3.3.　SaO2 signals

Artefact signal in the SaO2 signal was identified by finding all samples with 0 value. These samples were replaced with not a number (NaN) tag. A single feature

Table 1. Results for different levels of feature combining using an LDA classifier

| | Number of Epochs combined | | | | |
	± 0	± 1	± 2	±3	±4
AUROC	0.672	0.747	0.765	0.772	0.776
AUPRC	0.101	0.150	0.163	0.169	0.172

was calculated for each epoch calculated as the square root of the standard deviation of the non-NaN SaO2 values in the epoch.

3.3. Feature combiner

The features from each signal were combined into a single vector of 16 elements per epoch. The 2 EEG and the EMG signals resulted in 5 features per epoch per signal. The SaO2 signal resulted in 1 feature per epoch.

3.4. Epoch combiner

Adjacent arousal classifications are not independent. Hence, using information in surrounding epochs to classify an epoch, results in successful classification. A boost in the performance of a classification system was achieved by combining features from surrounding epochs with the current epoch. We tried combining features up to four epochs either side of an epoch.

3.5. Classifier

Before applying the data to the classifier training module, all epochs that were associated with apnoea-arousals were removed from the analysis. Three types of classifiers were trialed - linear discriminant analysis (LDA) [19], logistic regression [19] and single hidden layer feed-forward neural networks (FFNN) [19] with varying number of hidden units. All classifiers provide a probability estimate for each epoch of the occurrence of a non-apnoea arousal.

The FFNN used a 'tanh' transfer function for the hidden units and a softmax output stage. A cross-entropy cost function was used. Training was performed iteratively using an extended Levenberg Marquardt algorithm for cross-entropy cost functions [20]. Twenty percent of the available training data was set aside and used as a validation set.

During training, the performance of the system was monitored on the validation set and the training algorithm finished when there was no improvement on the validation set for 10 iterations. For this dataset, our training method arrived at a solution in 10-40 iterations.

3.6. Arousal annotator

The outputs of the classifier were probability estimates of an arousal occurring during the epochs. To map this to a sequence equivalent to the original arousal annotation (200Hz sampling) we up-sampled the epoch-probabilities to a 200 Hz signal using a first order hold filter.

3.7. Performance estimation

Performance was estimated on the 994 records of the training set using 10-fold cross validation. Each fold had either 99 or 100 records, with 9 folds used for training and 1-fold used for testing. This was repeated 10 times to use all recordings as the test set in 10 iterations.

3.4. Performance measures

Performance was measured using the area under the receiver operator curve (AUROC) [21] and the area under the precision-recall curve (AUPRC) [22].

4.   Results

We used a 15 second epoch as the basic signal window for our processing.

Table 1 shows the cross-validation performance of the system on the training set using an LDA classifier and

Table 2. Performance results for different classifiers using feature combining set at ± 4 epochs

| | | | Single Hidden Layer FFNN | | | |
	LDA	Logistic	3HUs	6HUs	10 HUs	20 HUs
AUROC	0.776	0.779	0.806	0.815	0.819	0.819
AUPRC	0.172	0.173	0.210	0.229	0.237	0.235

varying the level of feature combining. Performance was maximised when features were combined using ± 4 epochs

Table 2 shows the cross-validation performance of the system on training set using different classifiers processing feature combining using ± 4 epochs.

5.   Discussion and Conclusion

Table 1 shows that the highest performance for an LDA classifier was obtained when the features of every 8 adjacent epochs (4 epochs from either side) were combined. As this result was obtained at the maximum range of combining trialed, wider epoch combining may yielded greater performance.

The best performing feature-extraction system as identified in Table 1 was then used as input to five other classifiers including logistic regression and single hidden layer feed forward neural network with 3, 6, 10 and 20 hidden layer units. The two linear systems (LDA and logistic regression) returned very similar performance with an AUROC of 0.78 and an AUPRC of 0.17. Introducing a non-linearity into the classifier benefitted the system. The performance of our best system was an AUC 0.82 and an AUPRC of 0.24 and was achieved with a FFNN using 10 non-linear hidden units.

Processing other signals of the PSG and combining the features may enhance the results. Also, the performance of other machine learning algorithms (e.g. support vector machines) may improve the algorithm performance.

References

[1]   R. E. Dahl, "The regulation of sleep and arousal: Development and psychopathology," *Dev. Psychopathol.*, vol. 8, no. 01, p. 3, 1996.

[2]   P. Sleep and D. Functions, "Associations Between Sleep Duration Patterns and Behavioral / Cognitive Functioning at School Entry," no. c, pp. 2–8, 1998.

[3]   F. De Carli, L. Nobili, P. Gelcich, and F. Ferrillo, "A method for the automatic detection of arousals during sleep," *Sleep*, vol. 22, no. 5, 1999.

[4]   R. Mathur and N. J. Douglas, "Frequency of EEG arousals from nocturnal sleep in normal subjects.," *Sleep*, vol. 18, no. 5, pp. 330–333, 1995.

[5]   G. S. Hamilton, P. Solin, and M. T. Naughton, "Obstructive sleep apnoea and cardiovascular disease.," *Intern. Med. J.*, vol. 34, no. 7, pp. 420–426, Jul. 2004.

[6]   S. Motamedi-Fakhr, M. Moshrefi-Torbati, M. Hill, C. M. Hill, and P. R. White, "Signal processing techniques applied to human sleep EEG signals - A review," *Biomed. Signal Process. Control*, vol. 10, no. 1, pp. 21–33, 2014.

[7]   Atlas Task Force of the American Sleep Disorders Association, "EEG arousals: scoring rules and examples: a preliminary report," *Sleep*, vol. 15, no. 2, pp. 173–184, 1992.

[8]   M. H. Bonnet *et al.*, "The scoring of arousal in sleep:

[9]   Reliability, validity, and alternatives," *J. Clin. Sleep Med.*, vol. 3, no. 2, pp. 133–145, 2007.

[9]   D. Popovic, M. Khoo, and P. Westbrook, "Automatic scoring of sleep stages and cortical arousals using two electrodes on the forehead: Validation in healthy adults," *J. Sleep Res.*, vol. 23, no. 2, pp. 211–221, 2014.

[10]   M. H. Bonnet and D. L. Arand, "Clinical effects of sleep fragmentation versus sleep deprivation," *Sleep Med. Rev.*, vol. 7, no. 4, pp. 297–310, 2003.

[11]   M. J. McKeown, C. Humphries, P. A. Achermann, A. A. Borbely, and T. J. Sejnowski, "A new method for detecting state changes in the EEG: exploratory application to sleep data.," *J. Sleep Res.*, vol. 7 Suppl 1, pp. 48–56, 1998.

[12]   Jobert, Tismer, Poiseau, and Schulz, "Wavelets-a new tool in sleep biosignal analysis.," *J. Sleep Res.*, vol. 3, no. 4, pp. 223–232, Dec. 1994.

[13]   A. Tsanas and G. D. Clifford, "Stage-independent, single lead EEG sleep spindle detection using the continuous wavelet transform and local weighted smoothing," *Front. Hum. Neurosci.*, vol. 9, no. April, pp. 1–15, 2015.

[14]   C. Cajochen, D. P. Brunner, K. Krauchi, P. Graw, and A. Wirz-Justice, "Power density in theta/alpha frequencies of the waking EEG progressively increases during sustained wakefulness," *Sleep*, vol. 18, no. 10, pp. 890–894, 1995.

[15]   I. G. Campbell, "EEG recording and analysis for sleep research.," *Curr. Protoc. Neurosci.*, vol. Chapter 10, p. Unit10.2, Oct. 2009.

[16]   V. Lawhern, S. Kerick, and K. A. Robbins, "Detecting alpha spindle events in EEG time series using adaptive autoregressive models," *BMC Neurosci.*, vol. 14, no. 1, p. 1, 2013.

[17]   M. M. Ghassemi *et al.*, "You Snooze , You Win : the PhysioNet / Computing in Cardiology Challenge 2018," *Comput. Cardioilogy*, vol. 45, pp. 1–4, 2018.

[18]   P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, Jun. 1967.

[19]   T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*. 2009.

[20]   N. N. Schraudolph, "Fast curvature matrix-vector products for second-order gradient descent," *Neural Comput.*, vol. 14, no. 7, pp. 1723–1738, 2002.

[21]   J. a Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating (ROC) curvel characteristic," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[22]   K. Boyd, K. H. Eng, and C. D. Page, "Area under the precision-recall curve: Point estimates and confidence intervals," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8190 LNAI, no. PART 3, pp. 451–466, 2013.

nadi.sadr@sydney.edu.au
School of Electrical and Information Engineering
University of Sydney, NSW 2006, Australia