# Detecting Respiratory Effort-Related Arousals in Polysomnographic Data Using LSTM Networks

Sven Schellenberger[1], Kilin Shi[2], Melanie Mai[2], Jan P Wiedemann[2], Tobias Steigleder[3], Björn Eskofier[4], Robert Weigel[2], Alexander Kölpin[1]

[1] Chair of Electronics and Sensor Systems, Brandenburg University of Technology, Cottbus, Germany
[2] Institute for Electronics Engineering, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany
[3] Department of Palliative Care, Universitätsklinikum Erlangen, Erlangen, Germany
[4] Machine Learning and Data Analytics Lab, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany

## Abstract

*To diagnose sleep disorders, hours of sleep data from lots of different physiological sensors have to be analyzed. To do so, experts have to look through all the data which is time-consuming and error-prone. Automatic detection and classification of sleep related breathing disorders and arousals would significantly simplify this task. This years Physionet/CinC Challenge deals with this topic. This paper examines the use of a Long Short-Term Memory network for automatic arousal detection. On the test set, an AUPRC score of 0.14 was achieved.*

## 1. Introduction

Sleep is undoubtedly of great importance for the overall health and well-being. One of the well-studied sleep disorders that cause a bad sleep are obstructive apnea syndromes. Apneas are defined as sequences with a reduction in airflow greater then $90\%$ of baseline [1]. Obstructive apnea means that a respiratory effort is recorded throughout the signal. One of the causes of obstructive apnea syndrome are respiratory effort related arousals (RERAs). RERAs are a breathing disorder characterized by obstructive upper airway airflow reduction, which does not meet the criteria of apnea [1]. However, also other non-RERA, non-apnea kind of arousals such as bruxism might lead to the syndrome. RERAs as well as non-RERAs lead to a shallower sleep which means less deep recuperation. As part of the Physionet/CinC Challenge 2018 [2], a variety of physiological signals shall be evaluated to detect those non-apnea arousals during sleep. Amongst others, these include electroencephalography (EEG), electrooculography (EOG), electromyography (EMG) and electrocardiography (ECG) data.

Since a change in various physiological parameters usually occurs in the period prior to the arousals, Long Short-Term Memory (LSTM) networks promise to be a good approach for automatic detection and classification of arousals. LSTMs recently evolved as a strong algorithm for classification of time series data. This paper presents a concept of using a LSTM network and several features derived from different physiological sensors for automatic detection of RERA and non-RERA arousals.

## 2. Long Short-Term Memory

Learning from past information is a crucial part when analyzing time series data. In principle, recurrent networks (RNNs) are able to store information over extended time intervals. However, using the conventional Back-Propagation Through Time (BPTT) algorithm, error signals either tend to get very big or vanish over time since they exponentially depend on the value of the weights. Long Short-Term Memory networks are also based on a recurrent network structure which is designed to overcome the before mentioned problems during training [3]. The network can learn to bridge large time intervals while still keeping short time lag capabilities. This can be used in many areas like translation, text prediction and generation, natural language processing, audio and image analysis [4, 5]. In general, LSTMs achieve particularly good results in applications with time dependencies.

LSTM are built by artificial neural networks (ANNs) as core elements. Fig. 1 shows the structure of a simple feedforward ANN consisting of an input layer with two neurons, a hidden layer with three neurons and an output layer with one neuron [6]. To calculate the output $y_1$ the weighted outputs of all previous layers are added and fed
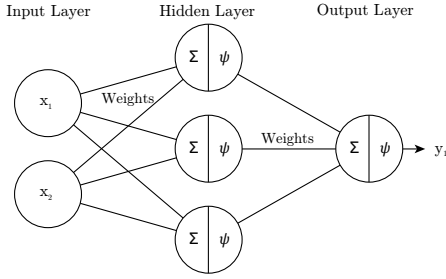
Figure 1. Structure of a simple ANN with summation $\Sigma$ and activation $\psi$. [6]

into an activation function $\psi$. Some often used activation functions are the sigmoid function, tangens hyperbolicus (tanh) or the rectified linear unit (ReLu) [7].
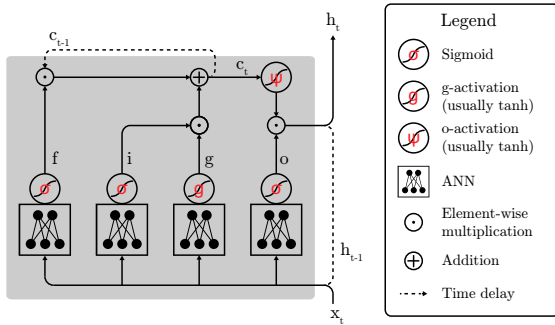


Figure 2. Basic structure of a LSTM block. [8]

As already mentioned, LSTMs belong to the group of RNNs. RNNs differ from ANNs in that they have a "hidden" state $h_t$ that feeds in data from the last time step $t-1$. This way, RNNs can look at past data, however, they have the disadvantages mentioned in the beginning of this chapter. LSTMs solve this issue by using a so-called constant error carousel (CEC) which corresponds to the loop of $c_t$ and $c_{t-1}$ in Fig. 2. $c_t$ ist the state of the cell which allows information to flow through easily. Errors are not exponentially degraded by going through the same weights of the RNN in each time step but remain in the CEC. This way, errors can flow back for an almost unlimited time whereby long-term dependencies can be modeled. What is stored in the CEC is controlled by two operations. The forget gate $f$ controls which data are remembered or can be forgotten by filtering $c_{t-1}$ multiplicatively. The second operations decides, which data are remembered newly. At first all possible entries are created by the input gate $i$, which are then passed through the gate $g$ and then added to $f \odot c_{t-1}$. These operations describe how the new cell state is calculated. Since this state should be invisible from outside, $c_t$ is filtered by an activation function and the ouput gate $o$. LSTM cells can also be concatenated to understand more complex patterns and deliver better results [9].

## 3. Methods

The following subsections briefly describe the recorded polysomnographic data which provide the basis for feature extraction and training of the designed LSTM model.

### 3.1. Data

The polysomnographic data provided for the challenge consists of $1985$ datasets which were split in $994$ sets for training and $989$ sets for test. A variety of sensors recorded physiological signals of subjects throughout the night, an overview of these signals is given in the first two columns of Tab. 1. The mean length of the recordings is $7.7\,\mathrm{h}$, the sensors are sampled with $200\,\mathrm{Hz}$. In each dataset arousals and also sleep stages were labeled by professionals whereby these information are only available for the training set. [2]

Table 1. Overview of recorded signals and the corresponding extracted features.

| Sensor | Signal | Features |
|---|---|---|
| Electroencephalography | O2-M1 | Power |
| | C4-M1 | Power |
| | C3-M2 | Power |
| | F3-M2 | Power |
| | F4-M1 | Power |
| | O1-M2 | Power |
| Electrooculography | E1-M2 | Peak Height, RMS |
| Electrocardiography | ECG | Heartrate |
| Respiratory Airflow | Airflow | Peak Height |
| Electromyography | Chin1-Chin2 | Onsets, RMS |
| | Chest | Onsets |
| | Abdomen | Onsets |
| Oxygen Saturation | SaO2 | Percentile, Diff. |

### 3.2. Feature Extraction

For data reduction and compact LSTM network configuration a set of features is extracted. To represent a physiological interpretation of the individual signals, different features were used. A chosen window size of one feature value for one second of data reduces the resolution for arousal detection to one second which is considered sufficient. An overview of the information extracted from each signal is given in the feature column of Tab. 1. Some of the features are calculated with the open source biosignal processing toolbox *BioSPPy* for Python [10]. A short description of the signals and their extracted values is given in the following listing:

*Electroencephalography* The main sensor for sleep stage estimation is the EEG. One indication of arousals are sudden changes in the sleep phase, which can be identified by the different frequency bands of the EEG's signals: alpha, beta, gamma, theta and delta. For this purpose the power in a given frequency band from $f_1$ to $f_2$ of a discrete signal $X$ in frequency domain can be calculated by:

$$P_{f_1-f_2} = \sum_{f=f_1}^{f_2} |X_f|. \quad (1)$$

In particular the relative power of a band is used for better comparison, which is normed to the power of the entire signal. For all of the given EEG signals the relative power of each of the five bands is calculated.

*Electrooculography* Eye movements are an indication of the REM phase. Therefore, by looking at the mean peak height in a window of one second, the motion activity is to be concluded. Additionally the root-mean-square (RMS) value should reflect the intensity of a movement. It is calculated for a given window of signal $x$ with length $N$ by:

$$x_{RMS} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} |x_n|^2}. \quad (2)$$

*Electrocardiography* A QRS detector by P. Hamilton [11] is used to calculate the instantaneous heart rate from the ECG's R-peaks, which is then smoothed by averaging $10\,\text{s}$ windows.

*Airflow* One characteristic feature of RERAs is airflow reduction. That is why, similar to the EOG, the mean peak height is calculated for every second of data to weight the airflow intensity.

*Electromyography* Muscular activity is recorded with the help of the EMG. Using the toolbox *BioSPPy*, onsets (muscle activation intervals) can be derived in these signals. In addition, the RMS value of the Chin1-Chin2 signal is determined.

*SaO2* Airflow restrictions are reflected in a decrease of the oxygen saturation. Therefore, the difference of the last $10\,\text{s}$ is computed to determine the change in time. Also, using the first and 99th percentile of the signal, all values below and above the percentiles are marked.

After extraction, each of the features is weighted with an empirically determined value in order to obtain figures of a similar magnitude between 0 and 10. In this way a total of 42 feature values were generated for every second of data to train the model.

### 3.3. Classification

In search for the optimal model to detect the arousals a configurable LSTM network was developed with *TensorFlow* in Python [12]. The number of time steps for the LSTM cells, hidden cells and layers of the network were kept variable. Input of the model is a matrix with the 42 feature values for every second of data times the amount of time steps and for training the labelled arousal information. Arousals are assigned the value one and using the rounded mean of a window one value per second is given. The training can be done with variable batch sizes, tested were sizes of 16 and 32 datasets.

A softmax layer was connected to the output of the LSTM network to receive the probability of an arousal. Using the RMSprop optimizer, which is a built in function of *TensorFlow*, a decaying learning rate with a momentum of 0.2 is utilized to overcome plateaus while training. Also a dropout of 0.2 was defined to reduce overfitting.

To get an overview of the models performance while training, metrics like the area under precision-recall curve (AUPRC) and the area under receiver operating characteristics (AUROC) were calculated for the current batch next to the cross entropy which is the objective function to be minimized.

### 4. Results

The LSTM network was trained with different configurations. The results of three of the configurations will be discussed briefly. They have the following parameters in common: two LSTM layers with each 256 hidden cells. Two networks have 1024 time steps but were trained with different batch sizes, model one with 16 and model two with 32. The third model has 2048 time steps and was trained with a batch size of 32. The performance metrics while training were recorded and shown in Fig. 3-5. The results calculated for each newly trained batch fluctuate, so they were smoothed with a moving average over 100 points.
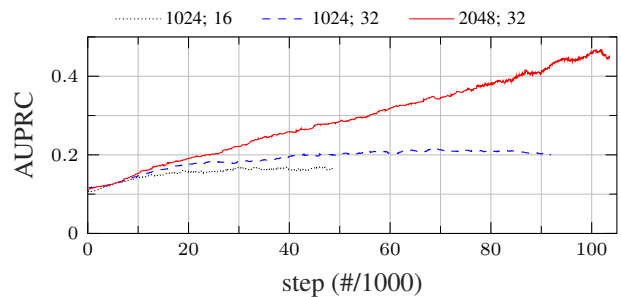


Figure 3. Comparison of the AUPRC of different model configurations (time steps; batch size) with increasing training time.

If you compare model one to model two in Fig. 3 the second model has a higher AUPRC of 0.21 over 0.17. Also, this model has a larger AUROC and the cross entropy is with 0.25 smaller. In the end the larger batch size

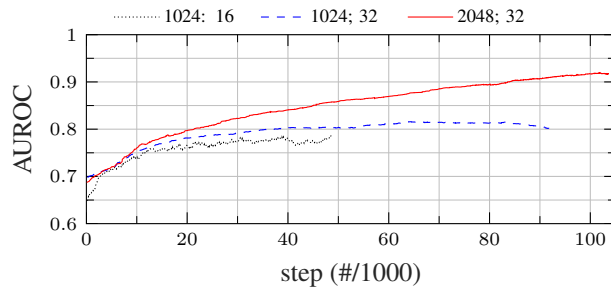results in a small performance increase.



Figure 4. Comparison of the AUROC of different model configurations (time steps; batch size) with increasing training time.

Model three provides the best result: Compared to the shorter networks, temporal relationships over a longer period of time can be observed. The AUPRC in Fig. 3 increases over a long training period to about 0.45, AUROC in Fig. 4 seems to saturate at 0.9. Additionally, the cross entropy shown in Fig. 5 decreases to 0.18.
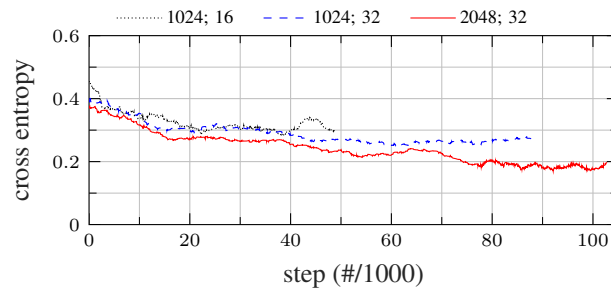


Figure 5. Comparison of cross entropy of different model configurations (time steps; batch size) with increasing training time.

The testing of different configurations and training of the final model had to be discontinued due to time constraints.

## 5.    Conclusion

This paper presents an approach for automatic detection of non-apnea sleep related breathing arousals using a LSTM network. Overall, 42 features from different physiological sensors are obtained. A LSTM network is implemented and different hyperparameters are tested. The best performing network structure was chosen for final classification. It consists of two LSTM cells with an input of 2048 time steps, a hidden layer size of 256 and a batch size of 32 datasets. On the test set, an AUPRC score of $0.14$ was achieved. To enhance the score, bigger hidden layers could be used. Furthermore, more features promise to improve the classification quality.

## References

[1] Tsara V, Amfilochiou A, Papagrigorakis M, Georgopoulos D, Liolios E. Definition and Classification of Sleep Related Breathing Disorders in Adults: Different Types and Indications for Sleep Studies (Part 1). Hippokratia 2009; 13(3):187.

[2] Ghassemi M, Moody B, Lehman L, Song C, Li Q, Sun H, Mark R, Westover B, Clifford G. You Win: the PhysioNet/Computing in Cardiology Challenge 2018. Computing in Cardiology 2018;45:1 – 4.

[3] Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation 1997;9(8):1735–1780.

[4] Marchi E, Ferroni G, Eyben F, Gabrielli L, Squartini S, Schuller B. Multi-Resolution Linear Prediction Based Features for Audio Onset Detection with Bidirectional LSTM Neural Networks. In Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, 2014; 2164–2168.

[5] Sutskever I, Vinyals O, Le QV. Sequence to Sequence Learning with Neural Networks. In Advances in Neural Information Processing Systems. 2014; 3104–3112.

[6] Schmidhuber J. Deep Learning in Neural Networks: An Overview. Neural Networks 2015;61:85–117.

[7] Maas AL, Hannun AY, Ng AY. Rectifier Nonlinearities Improve Neural Network Acoustic Models. In Proc. icml, volume 30. 2013; 3.

[8] Greff K, Srivastava RK, Koutnk J, Steunebrink BR, Schmidhuber J. LSTM: A Search Space Odyssey. IEEE Transactions on Neural Networks and Learning Systems Oct 2017;28(10):2222–2232. ISSN 2162-237X.

[9] Hermans M, Schrauwen B. Training and Analysing Deep Recurrent Neural Networks. In Advances in Neural Information Processing Systems. 2013; 190–198.

[10] Carreiras C, Alves AP, Lourenço A, Canento F, Silva H, Fred A, et al. BioSPPy: Biosignal Processing in Python. `https://github.com/PIA-Group/BioSPPy/`, 2015. [Online; accessed 2018-08-31].

[11] Open Source ECG Analysis Software Documentation. `http://www.eplimited.com/osea13.pdf`. [Online; accessed 2018-08-31].

[12] TensorFlow. `https://www.tensorflow.org/`. [Online; accessed 2018-08-31].

Address for correspondence:

Sven Schellenberger (sven.schellenberger@b-tu.de)
Institute for Electronics Engineering, Wetterkreuz 15, 91058 Erlangen