Real-Time Respiratory Event Detector Based on a Gated Recurrent Unit

Pierre Hayek[1], Jérémy Beaumont[1], Jean-Louis Pépin[2], Virginie Le Rolle[1], Alfredo Hernandez[1]

[1] Univ Rennes, Inserm, LTSI - UMR 1099, Rennes, France
[2] Grenoble Alps University, HP2 laboratory, Inserm, U1042, Grenoble, France

Abstract

Aims: Sleep-related breathing disorders (SBD) are found in 6-17% of the global adult population and are associated with a high risk of complications such as strokes or type-II diabetes. Previous studies have proposed a sample-wise real-time finite state machine (FSM) SBD event detector using as input the nasal pressure signal. This study aims at improving this SBD event detection by using a light Gated Recurrent Unit (GRU) neural network.

Methods: This work is based on data from a retrospective 5 centers study including 8 hours manually annotated polysomnography of 29 severe obstructive sleep apnea patients. The GRU was trained to classify in real-time each sample of an 8 Hz nasal pressure signal as either a normal or respiratory event (apnea or hypopnea). The GRU performances were compared to those reported for the FSM detector.

Results: Apnea and hypopnea detection provided by the GRU detector shows improved performances on the testing database on averaged sensitivity (apnea: 80.2% vs 66.9%, hypopnea: 67.2% vs 47.1%) while preserving a similar averaged positive predictive value (apnea: 79.0% vs 80.7%, hypopnea: 33.7% vs 25.0%).

Discussion: The GRU architecture appears as a promising tool for real-time detection of sleep apnea events. Further experiments on a larger database are warranted.

1. Introduction

Sleep-related breathing disorders (SBD) are found in 6% to 17% of the global adult population [1]. Sleep apnea is associated with a high risk of complications such as strokes, high blood pressure or heart failure [2]. The gold-standard SBD diagnosis is polysomnography (PSG), a multi-channel monitoring of cardio-respiratory and electroencephalography signals. The number of SBD events (apneas and hypopneas) per hour determines the SBD severity and helps driving the SBD treatment choice. Apneas are defined as a cessation of airflow for at least 10 seconds. Hypopneas are defined as a 50% minimal reduction in airflow magnitude for at least 10 seconds, followed by a 3% minimum oxygen desaturation or a sleep arousal.

Continuous Positive Airways Pressure (CPAP) is a SBD treatment that re-opens the upper airway and is currently considered as the first line therapy. Despite its efficiency, CPAP is associated with a 15% rate of initial refusal. Additionally, CPAP is characterized by a low adherence in the long term for 20-30% of the patients [3]. Recently, a new SBD treatment based on kinesthetic stimulation was proposed with the aim of improving the patient adherence to the treatment [4]. SBD treatments based on kinesthetic stimulation require a real-time detection of apnea and hypopnea events, which is the focus of the current study.

In a previous study conducted by our team, Feuerstein et al. proposed a sample-wise real-time detector based on an embedded Finite State Machine (FSM) that is dedicated to the detection of SBD events from a nasal pressure signal only [5]. Multiple studies reported in the literature showed that Recurrent Neural Networks (RNN) are associated with high performances for SBD event detection [6, 7]. To the best of our knowledge, the use of RNN for the sample-wise real-time detection of SBD events from a nasal pressure signal only has not been investigated. The current study aims at determining whether the Gated Recurrent Unit (GRU), a light RNN architecture, is a good candidate for this type of SBD event detection.

2. Method description

2.1. Data and patient population

The patient population consisted in 29 obstructive sleep apnea patients who underwent an in-lab PSG, gathered from a retrospective 5 centers study (HYPNOS). For each patient, 5 signals were recorded: the nasal pressure, an electrocardiogram, a photoplethysmogram, the oxygen saturation and the cardiac frequency. The PSGs were manually annotated by an expert, blinded core-lab (CHU Grenoble, France). The manual annotations followed the rules defined by the AASM [8]. The PSG recordings lasted on average 8 hours, representing an overall total of 4.6M recorded samples and 7846 recorded SBD events (34 events recorded per hour on average).

2.2. GRU real-time detection

The nasal pressure signal, recorded at 8 Hz, was preprocessed as in [5] to improve the signal to noise ratio (3 samples moving average) and remove the signal offset (exponential moving average). This preprocessed signal was then feed sample by sample to the GRU. The GRU was trained to classify each received sample as either a normal breathing, an apnea or an hypopnea. The network training was performed using the AdamW optimizer with a 0.01 starting learning rate and a standard cross-entropy loss function, with the PSG manual annotations used as a ground truth. The preprocessed nasal pressure signal of each patient was split in batches of even size, with 25% of the batches from the training database randomly selected for training at each epoch. The training phase lasted for 100 epochs, with the total number of epochs determined as a function of the results obtained on the training and validation sets used for the hyperparameter optimization.

The network structure consisted in stacked GRU layers followed by a 3-neurons dense layer with a log-softmax activation function, which was used to obtain the final classification. The number of GRU layers, the size of the input and output vectors, as well as the batch length used for splitting the patient data, were considered as hyperparameters that were optimized with a grid search approach. The grid search optimization was performed by maximizing the sum of the F1-score of each class, with each hyperparameter optimized sequentially and taking its values within a predefined range presented in Table 1. The grid search optimization was repeated until the choice of the hyperparameter values was stable. The hyperparameter optimization was conducted after randomly splitting the database in a training set (16 patients), a validation set (4 patients) and a testing set (9 patients), with the F1-score maximization performed on the validation set. The final structure of the network is presented in Figure 1.

Table 1. Range of GRU hyperparameter values investigated during the grid search optimization. The values selected after performing the optimization are also reported.

Hyperparameter	Range	Selected
Number of layers	1 - 10*	6
Batch length (s)	1 - 20*	10
Input vector length (s)	1 - 10*	6
Output vector size	1 - 256**	32

*step size: 1; **step size: power of 2

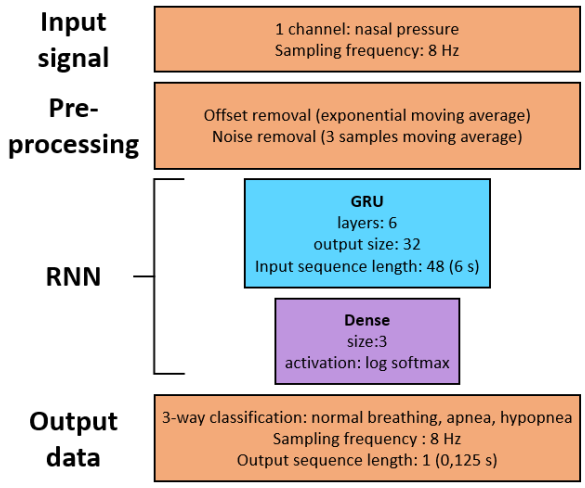

Figure 1. Architecture of the optimized neural network used in this study.

2.3. Data analysis

To assess the stability of the GRU detector as a function of the data used for training, a Monte-Carlo approach was applied to generate 100 realizations of the training and testing phases. Each training phase used 20 randomly chosen patients, while the testing phase used the remaining 9.

The performances of the GRU detection were assessed by computing the sensitivity (SE), the positive predictive value (PPV) and the specificity (SP), with the PSG manual annotations used as the ground truth. The performances were assessed sample-wise by converting the manually annotated PSG events to a digital scoring at the sampling frequency of the nasal pressure signal. To allow for a fair comparison between the results reported in the literature for the FSM detector and the results obtained with the GRU detector, the definition of a true positive was the same as in [5]. Specifically, a sample manually annotated in the PSG as an apnea (resp. hypopnea) and classified as an hypopnea (resp. apnea) by the GRU detector was considered as a true positive. This definition of the true positive was motivated in [5] by the aim of obtaining a global assessment of the respiratory event detection.

A receiver-operating characteristic (ROC) curve of the apnea detection was built for each Monte-Carlo realization of the GRU by increasing and decreasing the bias of the neuron from the dense layer that was linked to the apnea class. To facilitate the comparisons between the FSM and GRU performances, the GRU performances were reported for a specific point of the ROC curve that was selected to match the PPV of the FSM detector. The apnea PPV was selected as the matching metric with the aim of obtaining at least a similar rate of apnea false alarms for the GRU than the one reported for the FSM.

3. Results

The hyperparameters obtained by the grid search optimization are presented in Table 1. The resulting GRU detector contains 39648 parameters. The GRU detector provided high performances in terms of apnea classification for both the validation dataset (SE: 92%; PPV: 73%) and the testing dataset (SE: 90%; PPV: 75%). Moderate performances were obtained for the classification of hypopnea events on the validation dataset (SE: 55%; PPV: 54%) and on the testing dataset (SE: 46%; PPV: 48%). Please note that the aforementioned SE and PPV values were computed prior to the modification of the bias used to match the FSM detector PPV performances, while all the remaining performances reported were computed after PPV matching (Table 2).

A high area under the ROC curve was obtained on the testing datasets (average: 0.93, standard-deviation: 0.05; Figure 2). The apnea and hypopnea GRU detection shows improved performances on the testing datasets on averaged SE while preserving a similar averaged PPV when compared to the results reported for the FSM detector (Table 2). However, the GRU detector hypopnea specificity was lower than the one reported for the FSM detector. The Monte-Carlo approach shows a moderate standard-deviation for the apnea and hypopnea GRU detection on the testing datasets (Table 2).

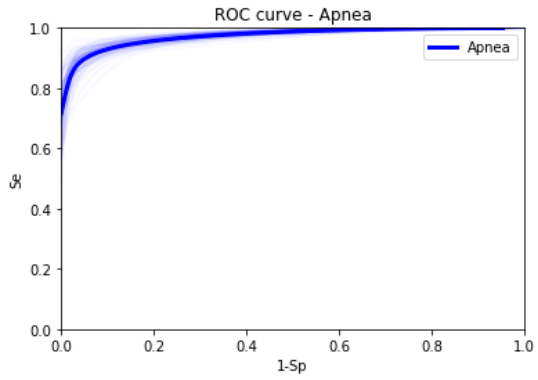

Figure 2. ROC curves of the apnea GRU detection for each Monte-Carlo realization (semi-transparent, thin, blue lines; average: thick blue line).

Qualitative assessment provided promising results, although it was found that the GRU detector did not perform well in terms of accurate differentiation between the apnea and hypopnea events (Figure 3). The qualitative assessment also showed that some of the false positives obtained seemed to correspond to apnea events that lasted for less than 10 seconds (Figure 3, red arrow) or to hypopnea events that preceded an apnea event and that were not annotated on the PSG (Figure 3, blue arrow).

4. Discussion

The current study aimed at providing a new sample-wise real-time SBD event detector, which uses only the nasal pressure signal to perform the detection. Although the aforementioned constraints tend to decrease the SBD event detection performances when compared to other studies found in the literature [6, 7], developing a detector under such constraints may allow for the improvement of SBD treatment performances [4]. Specifically, the triggering of the SBD event detection each time a nasal pressure sample is received by the embedded system may decrease the delay between the onset of an event and the activation of its treatment. In addition, the use of the nasal pressure signal only may improve the long term adherence of the patients to the treatment, which is likely to decrease if additional sensors are placed on the patient to perform the SBD detection.

The grid search optimization of the hyperparameters provided a GRU architecture associated with high apnea detection performances. Moderate performances were obtained for hypopnea detection. This may be explained by the single use of the nasal pressure signal for hypopnea

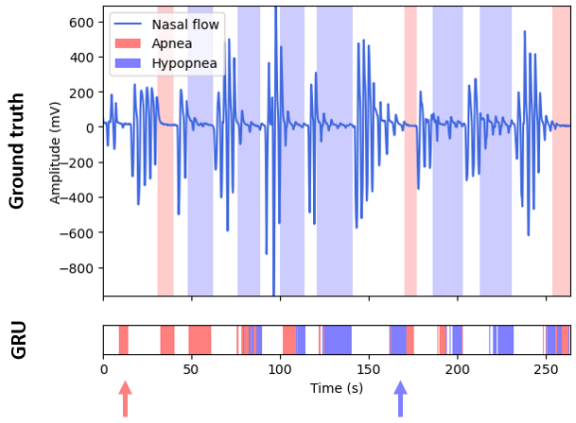

Figure 3. Example of sample-wise GRU detection (red areas: apnea events; blue areas: hypopnea events)

Table 2. Performances of the FSM and GRU detectors, with GRU performances reported as the average and the standard deviation computed over the testing datasets.

	SE (%)	PPV (%)	SP (%)
FSM A[1]	66.9	80.7	97.1
GRU A	**80.2±9.0**	79.0±5.2	98.5±1.7
FSM H[1]	47.1	25.0	**84.3**
GRU H	**67.2±8.0**	33.7±5.9	70.2±7.3

[1] Results reported in [5]; A=Apneas, H=Hypopneas.

detection, while it is advised to also measure the oxygen desaturation and the occurrence of sleep arousal [8]. A low decrease in GRU detection performances was obtained on the testing dataset compared to the validation dataset, which suggests that the hyperparameter optimization was not subject to over-fitting.

The GRU detector provided improved performances when compared to the FSM detector that was proposed in a previous study [5], specifically in terms of apnea and hypopnea sensitivity. However, the GRU detector was characterized by a lower hypopnea specificity than the FSM detector. Qualitative assessments suggest that this decrease in specificity was not necessarily due to a GRU detection failure. Specifically, the GRU model does not wait for an apnea or hypopnea event to end prior to provide its sample-wise classification, which induces the generation of false positives for events that were accurately detected by the GRU but were not annotated in the PSG as they lasted for less than 10 seconds. The FSM detector was implemented so that a delay of 8 seconds (4 seconds in a revised version of the detector for which the sample-wise results were not available [4]) is used to improve the detection [5], which improves the specificity at the cost of increasing the delay between the onset of an event and its treatment.

The current study is subject to some limitations. First, it is difficult to estimate the rate of GRU false positives that correspond to SBD events that were correctly detected and were not annotated in the PSG as they lasted for less than 10 seconds. Therefore, future experiments are required to either aim at improving the GRU detector performances by adding post-processing steps or to better assess its performances by performing a manual scoring of the GRU detection based on the PSG information. In addition, the FSM detector is robust and easy to implement on an embedded system, while 1) the GRU detector performances were subject to moderate variations according to the Monte-Carlo experiments; and 2) the real-time implementation of the GRU detector on an embedded system was not investigated. Further experiments performed on a larger database are warranted to both improve the GRU detector performances and better assess its training stability. Although the low number of parameters of the proposed GRU allows for its real-time implementation on embedded systems, as described in recent works [9], the possibility to use this detector in real-time on an embedded system still has to be validated in future experiments.

To conclude, the results obtained in the current study suggest that a RNN architecture, like the GRU, is a promising tool for the sample-wise real-time detection of SBD event from a nasal pressure signal only. Although further experiments are warranted, the results obtained are in line with previous work from the literature invesigating the interest of RNN achitecture for SBD event detection [6, 7].

5. Acknowledgements

This work was supported by the French National Research Agency (ANR) (ANR-20-THIA-0018) (project AI4SDA) and the French Brittany council (MONITOR project).

References

[1] Senaratna CV, Perret JL, Lodge CJ, Lowe AJ, Campbell BE, Matheson MC, Hamilton GS, Dharmage SC. Prevalence of Obstructive Sleep Apnea in the General Population: A Systematic Review. Sleep Medicine Reviews August 2017; 34:70–81.

[2] Linz D, Woehrle H, Bitter T, Fox H, Cowie MR, Bohm M, Oldenburg O. The Importance of Sleep-Disordered Breathing in Cardiovascular Disease. Clin Res Cardiol 2015;.

[3] Galetke W, Puzzo L, Priegnitz C, Anduleit N, Randerath W. Long-Term Therapy with Continuous Positive Airway Pressure in Obstructive Sleep Apnea: Adherence, Side Effects and Predictors of Withdrawal – A 'Real-Life' Study. Respiration 2011;.

[4] Hernández AI, Guerrero G, Feuerstein D, Graindorge L, Perez D, Amblard A, Mabo P, Pépin JL, Senhadji L. PASITHEA: An Integrated Monitoring and Therapeutic System for Sleep Apnea Syndromes Based on Adaptive Kinesthetic Stimulation. IRBM April 2016;37(2):81–89.

[5] Feuerstein D, Graindorge L, Amblard A, Tatar A, Guerrero G, Christophle-Boulard S, Loiodice C, Hernandez AI, Pepin JL. Real-Time Detection of Sleep Breathing Disorders. In 2015 Computing in Cardiology Conference (CinC). Nice: IEEE, September 2015; 317–320.

[6] Drzazga J, Cyganek B. An LSTM Network for Apnea and Hypopnea Episodes Detection in Respiratory Signals. Sensors January 2021;21(17):5858. Number: 17 Publisher: Multidisciplinary Digital Publishing Institute.

[7] Nikkonen S, Korkalainen H, Leino A, Myllymaa S, Duce B, Leppanen T, Toyras J. Automatic Respiratory Event Scoring in Obstructive Sleep Apnea Using a Long Short-Term Memory Neural Network. IEEE journal of biomedical and health informatics August 2021;25(8):2917–2927.

[8] Iber C, Ancoli-Israel S, Chesson A, Quan SF. The AASM Manual for Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. 1 edition. Westchester Illinois: American Academy of Sleep Medicine, 2007.

[9] Rezk NM, Purnaprajna M, Nordström T, Ul-Abdin Z. Recurrent Neural Networks: An Embedded Computing Perspective. IEEE Access 2020;8:57967–57996.

Address for correspondence:

Jeremy Beaumont

Laboratoire Traitement du Signal et de l'Image, Universite de Rennes, Campus de Beaulieu bat. 22, 35000 Rennes, France.

jeremy.beaumont@univ-rennes.fr