Comparison of PSG signals and Respiratory Movement Signal via 3D Camera in Detecting Sleep Respiratory Events by LSTM Models

Carmina Coronel^{1*}, Christoph Wiesmeyr¹, Heinrich Garn¹, Bernhard Kohn¹, Anahid Naghibzadeh-Jalali¹, Alexander Schindler¹, Markus Wimmer², Magdalena Mandl², Martin Glos^{3,4}, Thomas Penzel^{3,4,5}, Gerhard Kloesch⁶, Andrijana Stefanic-Kejik⁶, Marion Boeck⁶, Eugenijus Kaniusas⁷, Stefan Seidel⁶

¹AIT Austrian Institute of Technology GmbH, Vienna, Austria

² Kepler University Hospital, Department of Neurology 2, Linz, Austria

³ Advanced Sleep Research GmbH, Berlin, Germany

⁴ Interdisciplinary Sleep Medicine Center, Charité-Universitätsmedizin Berlin, Berlin, Germany

⁵ Saratov State University, Saratov, Russia

⁶ Medical University of Vienna, Department of Neurology, Vienna, Austria

⁷ Vienna University of Technology (TU Wien), Institute of Electrodynamics, Microwave and Circuit Engineering, Vienna, Austria

Abstract— A contactless system to detect respiratory events during sleep may be advantageous because it enables normal sleeping pattern and eliminates the need for constant monitoring of contact sensors. To detect sleep respiratory events, a 3D timeof-flight (TOF) camera is placed above the bed to measure a respiratory movement signal. Using this signal, we trained a long-short term memory (LSTM) model to detect respiratory events. In addition, we trained LSTM models based on SpO₂, and and abdomen thorax respiratory inductance plethysmography (RIPsum). LSTM models were trained on 8 folds using 61 synchronized 3D video and polysomnography (PSG) recordings of patients with suspected sleep apnea to classify 30-second segments as either respiratory event or normal breathing. Manual PSG annotations served as ground truth. The LSTM model based on 3D TOF camera achieved a mean accuracy of 0.79 and mean Cohen's kappa of 0.54. SpO₂ based model scored a mean accuracy of 0.86 and mean Cohen's kappa of 0.68 while RIPsum based model scored a mean accuracy of 0.82 and mean Cohen's kappa of 0.61. The 3DRespMvt performance can be improved by combining it with SpO₂. resulting in a mean accuracy of 0.87 and mean Cohen's kappa of 0.71.

I. INTRODUCTION

Sleep respiratory disorders are a subtype of sleep disorders that concern with the disruption of respiration. The main feature of these disorders is the presence of apneas and hypopneas [1], where an apnea is defined as a decrease in airflow signal of at least 90% for at least 10 seconds [2]. Hypopneas are similar to apneas but with at least 30% decrease in airflow signal with associated oxygen desaturation or arousal [2]. The severity of a sleep respiratory syndrome is described in the index called apnea-hypopnea index (AHI). AHI is the number of apneas and hypopneas per hour of sleep. The AHI can be measured via a sleep study called a polysomnography (PSG). PSG is an overnight test that monitors and analyzes sleep using several contact sensors. In particular for respiratory disorders, the specific relevant PSG



Fig. 1 3D TOF Camera and PSG setup in a sleep laboratory

signals are airflow signals via oronasal thermistor and nasal pressure cannula, SpO₂ (level of oxygen in blood at the peripherals) via pulse oximetry, and respiratory effort via respiratory inductance plethysmography (RIP) belts. Decrease in airflow, respiratory effort, and SpO₂ could be indicators of respiratory events wherein the specific rules concerning apnea and hypopnea scoring are set by the American Academy of Sleep Medicine scoring manual [2]. Another method of measurement is via the home sleep apnea testing (HSAT). HSATs are tests taken by patients in their homes and include four signals: nasal pressure, SpO₂, pulse rate and respiratory effort via RIP belts.

Due to the long waiting times, high prevalence of apnea in the population, and for economic reasons, the authors previously introduced contactless 3D time-of-flight (TOF) camera to derive a respiratory movement signal [3] - [4]. In addition, using a contactless sensor can improve sleep quality, enable usual sleep mobility, and reduce constant monitoring of sensors. Fig. 1 illustrates the setup of our system where the camera was placed above the bed. Through the depth information from the 3D TOF camera of the upper body region a respiratory movement signal we call 3DRespMvt is derived. In [3], we compared the decrease in 3DRespMvt to

^{*}E-mail: carmina.coronel@ait.ac.at Tel: +43-50550-4155



Fig. 2 Signal 3DRespMvt from 3D TOF camera compared to SpO₂ and RIPsum from PSG during respiratory events



Fig. 3 Signal 3DRespMvt from 3D TOF camera compared to SpO₂ and RIPsum from PSG during normal breathing

the respiratory effort signal measured by RIP belts in a standard PSG during sleep respiratory events. In [4], we introduced a rule-based detection algorithm using 3DRespMvt and SpO₂. In this paper, we want to evaluate further the use of 3DRespMvt by developing a neural network detection model and comparing to other PSG signals to evaluate performance.

Detection methods for apneas and hypopneas via neural networks have previously been evaluated. These networks either used a combination of signals already present in a PSG [5]–[7], or used a single signal such as electrocardiogram (ECG) or SpO₂ [8]–[11], or used clinical information such as age, gender, and body mass index (BMI) [12], [13]. In this paper, we show the use of 3D TOF camera to detect respiratory events via neural networks. In addition, we also show the detection performance of SpO₂ and RIP belts, as both are used in standard PSG and HSAT. We show the performance comparison between 3DRespMvt to the other two signals.

II. METHODOLOGY

A. Subjects

61 3D video and PSG recordings of patients with suspected obstructive sleep apnea were recruited in this study. PSG recordings were performed at the sleep laboratories of Kepler University Clinic (JKU), Linz, Austria or at the Advanced Sleep Research GmbH (ASR), Berlin, Germany. The mean



age of the patients was 56 (\pm 12.8) and the mean BMI was 29.8 (\pm 5.8). Forty one of the patients were men.

The following ethical committees approved this study: ethical committees of the state of Upper Austria (B-130-17) and of the Charité – Universitätsmedizin Berlin (EA1/127/16). Prior to the inclusion in the study, written consents from the subjects were acquired.

B. Collection of Data

The 3D TOF camera used in this study was the Kinect II. More information on the recording setup is further explained in [4]. The sleeping setup followed the usual protocol: lights were turned off and blankets were provided. The JKU clinic used the Somnoscreen Plus with Domino software (Somnomedics, Randersacker, Germany) while the ASR clinic used the EMBLA N7000 system with RemLogic 3.4.1 software (Embla Systems, Broomfield, CO, USA).

Recordings were annotated manually based on the AASM scoring manual [2]. These manual annotations served as reference in this study. Apneas and hypopneas were scored individually and grouped together in this study and are referred to as respiratory events.

C. Training Data

Consecutive respiratory events are shown in Fig. 2, where there is a decrease in signal or lack of respiratory movements shown in 3DRespMvt. The same can be observed from RIPsum. On the other hand, SpO₂ desaturates by more than 3% although with a delay. These changes in the signals enable identification of respiratory events. Therefore, we used 3DRespMvt, SpO₂, and RIPsum (sum of abdomen and thorax RIP belts), to derive features for the models. On the other hand, Fig. 3 shows the behavior of the signals when there are no respiratory events that can be observed.

The following time features were computed: mean, minimum, maximum, average difference, standard deviation.

Samples were created by dividing the recording to 30-s segments in each recording. Then another 30s segments succeeding and preceding were combined to each 30s segment to create 90s segments. Each sample of 90s segment will have 60 seconds of overlap. As shown in Fig. 4, each sample of data is a 90s segment where features were computed on 3s segments. This resulted in a (30, N_F) input sample size where N_F is the number of features for the LSTM model. If a respiratory event is present in the center 30s segment, the sample is labelled as P for respiratory event. Otherwise, it is labelled as N for normal breathing.

D. LSTM Model

In this study, we made use of long-short term memory (LSTM) neural networks. LSTM is a subclass of recurrent

Table I 8-fold cross validation results of LSTM models.

Model	Mean (Standard Deviation)			
	Accuracy	Sensitivity	Specificity	Cohen's kappa
А	0.79 (0.03)	0.76 (0.07)	0.80 (0.06)	0.54 (0.05)
В	0.82 (0.04)	0.82 (0.07)	0.82 (0.07)	0.61 (0.09)
С	0.86 (0.03)	0.85 (0.07)	0.86 (0.04)	0.68 (0.06)
D	0.87 (0.03)	0.86 (0.06)	0.87 (0.04)	0.71 (0.05)
Е	0.87 (0.02)	0.86 (0.06)	0.87 (0.04)	0.71 (0.05)



Fig. 5 LSTM Model using 3RespMvt and SpO2 as training data

neural networks introduced in [14]. LSTM is capable of processing sequences of data such as time series.

Five different model and signals combination were trained to compare the signals and to find the best signals to detect respiratory events. Every model used different signals or set of signals: (A) 3DRespMvt, (B) RIPsum (C) SpO₂, (D) 3DRespMvt and SpO₂, and (E) RIPsum and SpO₂.

The model outputs a number between 0 and 1, indicating probability of a respiratory event in an input. We used a cutoff of ≥ 0.5 to classify segments as P (with respiratory event), otherwise it was classified as N (no respiratory event). Our models consisted of an input layer, bidirectional LSTM layers, fully connected layers, and an output layer. Bidirectional LSTM layers are two LSTM layers where the input is reversed in one. We added a dropout layer to avoid overfitting as shown in Fig. 5.

E. Model Evaluation

Evaluation of the models was performed by 8-fold cross validation. The folds were defined based on the AHI severity of recordings. The severity groups were mild (5-15), moderate (15-30), and severe (\geq 30) [15]. This was to ensure that a similar distribution between mild, moderate, and severe was present in each fold. Performing the division of data on the

recording level ensured that data from a single recording were not subdivided into training and validation set. In addition, this approach enabled evaluation of the inter-personal generalization of the model.

The average sample size of training data across 8 folds was $39,214 (\pm 530)$ and while for the test data it was $6,536 (\pm 362)$.

F. Statistical Analysis

The model outputs a number between 0 and 1 where anything above 0.5 is classified as P (with respiratory event), and N (no respiratory event) if not. Accuracy, specificity, sensitivity, and Cohen's kappa were used to evaluate the performance of the models. Cohen's kappa is a measure of inter-rater agreement. A Cohen's kappa score ranges between 0 to 1, where a value between 0.41-0.60 suggests a moderate agreement, 0.61-0.80 suggests substantial agreement and 0.81-1.00 suggests a perfect agreement [16]. Sensitivity is the ratio of the correctly classified P over all true P instances while specificity is the measure of correctly classified N over all true N instances.

AHI estimates the severity of a sleep respiratory condition and is an index calculated after scoring to determine the number of respiratory events per hour of sleep. The equation of AHI (# events / hour) according to AASM is (1) where it is based on the total sleep time (TST) in minutes, and A and H are the number of apneas and hypopneas respectively. Here, we introduce the LSTM-based Respiratory Index (LRI) (# of positive segments / hour) to be compared to the AHI. LRI is calculated by assuming that for each input classified as P by the model is a single respiratory event. We calculated LRI per recording using the formula in (2) where P_n is the number of predicted P segments and T is the total number of input samples of a particular recording. LRI is therefore based on the total recording time (TRT). Pearson's correlation was then used to compare LRI to the AHI.

$$AHI = \frac{(A+H) \cdot 60}{TST_{\rm mins}} \tag{1}$$

$$LRI = \frac{P_{\rm n} \cdot 60}{(T \cdot 30)/60}$$
(2)

III. RESULTS

As shown in Table I, model A based on 3DRespMvt performed with Cohen's kappa of 0.54, suggesting moderate agreement. Model B based on RIPsum performed better with a Cohen's kappa of 0.61 while Model C using SpO_2 obtained 0.68.

In addition to using single sensors, model D and E combined 3DRespMvt and RIPsum with SpO₂ respectively. Model D showed an increase in performance compared to using SpO₂ or 3DRespMvt alone with Cohen's kappa of 0.71. The structure of model D made use of 2 parallel bidirectional LSTM layers, shown in Fig. 5. Model E obtained a Cohen's kappa of 0.71.

We calculated the LRI of model A as shown in Fig. 6, where it has a Pearson's correlation = 0.74 (p < 0.001) when



Fig. 7 AHI vs LRI (model D), line is x = y

compared to the AHI, computed based on the reference annotations. LRI of model D compared to reference AHI, as shown in Fig. 7, scored a Pearson's correlation of 0.82 (p < 0.001).

IV. DISCUSSION AND CONCLUSION

The results show that using 3DRespMvt alone can provide good results with Cohen's kappa = 0.54. In addition, LRI based on 3DRespMvt showed a positive moderate Pearson's correlation with AHI. However, compared to other detection models using single sensors, it performed worse. Our RIPsum-based detection resulted in a Cohen's kappa = 0.61. In addition, our SpO₂ based model scored a Cohen's kappa = 0.68 and an accuracy of 0.86. In the literature, SpO₂ based detection using deep belief network has been reported with an accuracy of > 86% [10]. In comparison to using another single sensor such as ECG to detect sleep disordered breathing (SDB), LSTM and GRU models developed by [8] has a high F1 score of > 98%.

Using 3DRespMvt alone did not provide the same accuracy as using SpO_2 or RIPsum individually. However, by combining 3DRespMvt with SpO_2 , there was a significant improvement in the results with Cohen's kappa of 0.71 and 87% accuracy. This is comparable to the reported results of [5] using recurrent and convolutional neural networks (CNN) with 88.2% accuracy. Another model using CNN reported 79% accuracy in classifying between normal, apnea, and hypopnea [6]. We also showed the possibility of combining 3DRespMvt and SpO₂ in [4] albeit using a rule-based method. In addition, the results of 3DRespMvt and SpO₂-based LSTM also show that it is comparable to the LSTM model using SpO₂ and RIPsum. An advantage of 3D TOF camera over RIP belts is that it's less cumbersome to the patient and it doesn't hinder mobility during sleep. Furthermore, the use of a learning method such as LSTM instead of traditional rulebased methods could be advantageous. In the future, we expect the addition of more data, possibly from different clinics. The update and development of rule-based method requires intensive programming to accommodate all possibilities in which learning methods can cover.

In this study, we used features from 3DRespMvt and SpO₂ to train an LSTM model. LSTM are models that are able to learn such long-term relationships of the data [17]. Therefore, we used LSTM instead of other neural network models or other machine learning methods as it is suitable for time-series data, such as 3DRespMvt and SpO₂. In addition, respiratory events are events that are at least 10 seconds long and its detection is dependent on the pre-event signal. There must be a decrease in signal such as in 3DRespMvt and RIPsum, compared to its pre-event signal. On the other hand, the desaturation in SpO₂ occurs only after the onset of the respiratory event and the amount of delay may vary.

AHI and LRI in Fig. 7 are shown to be correlated although with two outliers. The two outliers can be explained by the difference between TST and TRT. Respiratory events occurring during wake periods were excluded from AHI calculation. However, in the calculation of LRI, predicted events during wake period cannot be filtered out. The LRI estimation can be improved by determining the sleep and wake periods of the recording to calculate a total sleep time. A point of improvement for future work is to utilize the 3D camera depth information of the whole body to perform sleep and wake staging. In addition, as we used uniform 90s segments in the model, it will be a focus on future work to label the recordings per second of data in order to provide a more precise detection of respiratory events. For further improvement of the current models, acquisition of more data is needed.

Our results showed that detection via LSTM of respiratory events via a 3D TOF camera is a promising alternative to current detection methods, and it can be improved by using it in combination with SpO₂.

ACKNOWLEDGMENT

This study was supported by the Austrian Research Promotion Agency (FFG), project ID 859622.

Thomas Penzel was partially supported by the RF Government grant N $^{\circ}$ 075-15-2019-1885.

REFERENCES

- [1] American Academy of Sleep Medicine, *International classification of sleep disorders*. Darien, Ill.: American Acad. Of Sleep Medicine, 2014.
- [2] R. B. Berry et al., 'The AASM Manual for the Scoring of Sleep and Associated Events', Rules, Terminology and Technical Specifications. Version 2.5, Darien, Illinois, American Academy of Sleep Medicine, 2018.
- [3] C. Coronel *et al.*, 'Measurement of respiratory effort in sleep by 3D camera and respiratory inductance plethysmography: A comparison of two methods', *Somnologie*, May 2019.
- [4] C. Coronel *et al.*, '3D Camera and Pulse Oximeter for Respiratory Events Detection', *IEEE J. Biomed. Health Inform.*, pp. 1–1, 2020 (early access).
- [5] S. Biswal, H. Sun, B. Goparaju, M. B. Westover, J. Sun, and M. T. Bianchi, 'Expert-level sleep scoring with deep neural networks', *J Am Med Inform Assoc*, vol. 25, no. 12, pp. 1643– 1650, 01 2018.
- [6] L. Cen, Z. L. Yu, T. Kluge, and W. Ser, 'Automatic System for Obstructive Sleep Apnea Events Detection Using Convolutional Neural Network', *Conf Proc IEEE Eng Med Biol Soc*, vol. 2018, pp. 3975–3978, 2018.
- [7] O. Aydoğan, A. Öter, K. Güney, M. K. Kıymık, and D. Tuncel, 'Automatic Diagnosis of Obstructive Sleep Apnea/Hypopnea Events Using Respiratory Signals', *Journal of Medical Systems*, vol. 40, no. 12, p. 274, Oct. 2016.
- [8] E. Urtnasan, J.-U. Park, and K.-J. Lee, 'Automatic detection of sleep-disordered breathing events using recurrent neural networks from an electrocardiogram signal', *Neural Comput & Applic*, vol. 32, no. 9, pp. 4733–4742, May 2020.
- [9] E. Urtnasan, J.-U. Park, E.-Y. Joo, and K.-J. Lee, 'Automated Detection of Obstructive Sleep Apnea Events from a Single-Lead Electrocardiogram Using a Convolutional Neural Network', *J Med Syst*, vol. 42, no. 6, p. 104, Apr. 2018.
- [10] S. S. Mostafa, F. Mendonça, F. Morgado-Dias, and A. Ravelo-García, 'SpO2 based sleep apnea detection using deep learning', in 2017 IEEE 21st International Conference on Intelligent Engineering Systems (INES), Oct. 2017.
- [11] D. S. Morillo and N. Gross, 'Probabilistic neural network approach for the detection of SAHS from overnight pulse oximetry', *Med Biol Eng Comput*, vol. 51, no. 3, pp. 305–315, Mar. 2013.
- [12] H. Karamanli, T. Yalcinoz, M. A. Yalcinoz, and T. Yalcinoz, 'A prediction model based on artificial neural networks for the diagnosis of obstructive sleep apnea', *Sleep Breath*, vol. 20, no. 2, pp. 509–514, May 2016.
- [13] A. A. el-Solh, M. J. Mador, E. Ten-Brock, D. W. Shucard, M. Abul-Khoudoud, and B. J. Grant, 'Validity of neural network in sleep apnea', *Sleep*, vol. 22, no. 1, pp. 105–111, Feb. 1999.
- [14] S. Hochreiter and J. Schmidhuber, 'Long Short-Term Memory', *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [15] American Academy of Sleep Medicine Task Force, 'Sleep-Related Breathing Disorders in Adults: Recommendations for Syndrome Definition and Measurement Techniques in Clinical Research', *Sleep*, vol. 22, no. 5, pp. 667–689, Aug. 1999.
- [16] J. R. Landis and G. G. Koch, 'The measurement of observer agreement for categorical data', *Biometrics*, vol. 33, no. 1, pp. 159–174, Mar. 1977.
- [17] Goodfellow, I., Bengio, Y. and Courville, A., 2016. Deep Learning. Cumberland: MIT Press.