Detecting Obstructive Apnea Episodes using Dynamic Bayesian Networks and ECG-based Time-Series

Daniel Romero and Raimon Jané, *Senior Member, IEEE*

***Abstract*— In this study, we proposed an automatic detector for obstructive apnea episodes using only ECG-based time-series from a single-ECG channel. Several obstructive apnea episodes were provoked for different separated sequences of 15 minutes in anesthetized Sprague-Dawley rats. In this recurrent obstructive sleep apnea (OSA) model, each episode lasted 15 s, while the number of total events per sequence was randomly selected. The beat-to-beat interval (RR) and the R-wave amplitude (Ra) time-series were extracted and processed for each sequence, and used to train Dynamic Bayesian Networks with different lags. An optimal trade-off between the lag ($\mathcal{L}$) and $RMSE$ values was considered to select the best model to be used when detecting apnea episodes. The selected models were then used to estimate the occurrence probability of apnea episodes, $p(A_t)$, by using a filtering approach. Finally, the time-series of the estimated probabilities were post-processed using non-overlapped 15-s epochs, to determine whether they are classified as apneic or non-apneic segments. Results showed that those lagged models with orders greater than 5, presented suitable RMSE values and become more sensitive as the order increased. A detection threshold of 0.2 seems to provide the best apnea detection performance overall, with Acc=0.81, Se=0.83 and Sp=0.79, using two ECG parameters and $\mathcal{L} = 10$.**

***Clinical relevance*— Dynamic Bayesian Networks represent a powerful tool to develop personalized models for apnea detection and diagnosis in OSA patients.**

I. INTRODUCTION

Recurrent apnea during patients' sleep are caused by repetitive obstructive sleep apnea (OSA) episodes, resulting in sustained exposure to intermittent hypoxia (IH). This condition has been linked to some cardiovascular consequences, including among others, systemic hypertension, heart failure, coronary artery disease and stroke [1], [2], [3]. On the other hand, patients suffering from OSA usually present excessive daytime drowsiness and non-restorative sleep, tiredness, reduced learning capabilities, and significant social problems caused by poor mental performance.

Although the links between cardiovascular diseases and IH are unclear, an elevated sympathetic tone of autonomic control has been suggested as an important contributing factor. Moreover, during OSA, the muscles of the respiratory system increase the mechanical effort to overcome the occlusion. This additional respiratory effort can influence other peripheral systems such as the cardiovascular system. In particular, the surface electrocardiogram (ECG) can be very informative about apneic events and has been widely used for apnea detection. In general, patientS' breathing information can be extracted from: the fluctuations observed in the beat-to-beat interval time-series, named heart rate variability (HRV) and; the morphological ECG changes caused by breathing modulation, through the analysis of ECG-derived respiratory (EDR) signals.

Several studies have proposed automatic approaches for OSA patients' diagnosis and sleep apnea detection. These works usually employ a large set of features extracted from polysomnographic (PSG) recordings, and apply machine learning, neural networks, or deep learning techniques to detect and classify apnea episodes in the night-sleep recordings [4]. Other studies have used only ECG-based features for the same purpose [5], but most of them use segmented intervals or epoch, typically of one-minute duration, to train the models and performing the prediction task in an off-line fashion. However, designing a more simple predictive model, with higher time resolution and potentially useful for online predictions, could be of great interest for home monitoring systems, wearable devices, and healthcare in OSA patients.

The study aimed to detect the occurrence of apnea episodes based on the effects caused by respiratory obstruction on different ECG markers. Specifically, we exploited the subtle changes occurring in the heart rate and ECG amplitude during apnea episodes using an experimental rat model. To this end, beat-to-beat time-series of the aforementioned markers along with apnea information were used to train individual Dynamic Bayesian Networks (DBN), able to detect apnea segments using short epochs of 15 s.

II. MATERIALS AND METHODS

A. Experimental data

The study dataset comprises three male Sprague-Dawley rats anesthetized with urethane (1g/1kg) and with an average weight of 437±27 g. The animals were connected to a system, where a nasal mask with two tubes was used to induce obstructive apneas. One of these tubes was open to the atmosphere, and the other was connected to a positive pressure pump to prevent the animal from rebreathing. Recurrent apnea was simulated by closing the airways in the tubes through controlled electrovalves. The nasal mask

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 846636. This work was also supported by the CERCA Program/Generalitat de Catalunya, in part by the Secretaria d'Universitats i Recerca de la Generalitat de Catalunya under grant GRC 2017 SGR 01770 and in part by the spanish grant RTI2018-098472-B-I00 (MCIU/AEI/FEDER, UE).

D. Romero is with the BIOSPIN group at the Institute for Bioengineering of Catalonia (IBEC), the Barcelona Institute of Science and Technology (BIST), Barcelona, Spain (email: dromero@ibecbarcelona.eu).

R. Jané is with the ESAII Department, Universitat Politècnica de Catalunya (UPC - BarcelonaTech), the Institute for Bioengineering of Catalonia (IBEC-BIST), and Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Barcelona, Spain (email: rjane@ibecbarcelona.eu).

and electrodes were placed on the anesthetized animals after shaving the specific positions. The experimental model was approved by the Institutional Animal Care and Ethics Committee of the Hospital Clínic, Barcelona.

During the experiments, several physiological signals were recorded including among others, two ECG channels (leads I and II) using Biopac® Systems, and two respiratory-related signals such as flow and pressure [6], [7].

B. Experimental Protocol

The experiment setup consisted of several recurrent apnea sequences provoked for periods of 15 min, preceded, and followed by 15-min periods of normal respiration. Apnea episodes were simulated at fixed rates for each sequence, using 20, 40, or 60 events/hour. Each individual episode lasted 15 s, while the order of the applied rates was randomly selected for each rat. Therefore, a total of 30 apnea episodes were induced in each animal (90 for the dataset) during those 45 min, although the entire record had a total of 115 min when including the normal respiratory periods.

C. ECG analysis

ECG signals analysis included baseline drift attenuation, low-pass filtering at 45 Hz (bidirectional 4^{th} order Butterworth filter) to remove high-frequency noise, and QRS complex detection using a wavelet-based technique [8] followed by a visual inspection to exclude abnormal beats. Then, the R-wave peak amplitude (Ra) and RR interval time-series were extracted for further analysis and modeling.

D. Bayesian networks to detect apnea episodes

Obstructive apnea episodes affect several physiological variables from different interacting systems including the cardiac, respiratory, and neural systems. Such variables reflect different responses through amplitude variations and time duration and may help to detect apnea occurrence if their relationships can be modeled. In this study, these variables are represented as univariate time-series that can affect each other via unknown relationships. To model these relationships, hybrid DBNs that combine continuous and categorical variables were used to approximate their temporal interactions during normal respiration and apnea episodes.

BNs are a type of probabilistic graphical model representing the conditional independences among random variables with directed acyclic graphs (DAGs) [9]. They can manage both discrete and continuous variables simultaneously (hybrid BNs), but also only discrete (multinomial BNs) or continuous nodes (Gaussian BNs). Each node in the DAG has an associated conditional probability distribution (CPD) that defines the probability distribution of the node given its parents in the DAG. In general, for a BN with N variables $\mathbf{X} = \{X_1, ..., X_N\}$, the joint distribution factorizes as:

$$P(\mathbf{X}) = \prod_{i=1}^{N} P(X_i|pa(X_i)) \tag{1}$$

where $pa(X_i)$ denoting the configuration of the set of parents of X_i in the network.

Specifically, in a hybrid BN, the models are constructed as a set of Conditional Linear Gaussian (CLG) distribution models [10]. Here, discrete nodes are not permitted to have continuous parents, and their conditional distribution given its parents are multinomial. In the case of continuous nodes, the conditional distribution $Z \in X_C$ with discrete parents $Z_D \subseteq X_D$ and continuous parents $Z_C \subseteq X_C$, is given by:

$$f(z|pa(z) = \{z_D, z_C\}) = \mathcal{N}(z; \alpha(z_D) + \beta(z_D)^T z_C, \sigma^2(z_D)) \tag{2}$$

for all $Z_D \subseteq X_D$ and $Z_D \subseteq X_D$, where α and β are the coefficients of a linear regression model of Z given its continuous parents. Note that this model can differ for each configuration of the discrete variables Z_D. After fixing any configuration of the discrete variables X_D, the joint distribution of any set of continuous variables X_C is a multivariate Gaussian.

1) DBNs: DBNs usually assume a first-order Markovian (i.e., future states are independent of the past, given the present) for the underlying process they model. However, the nature of many biological processes requires relaxing this assumption to include additional past information in order to improve model prediction of future states.

In DBN, the time is discretized into slices for a given period. For each time slice, there is a static BN that has parents in the previous slices in addition to those of the actual slice. In our study, each time slice is associated with each heartbeat occurrence. Here, the joint probability distribution accounts for all time slices from a certain time T:

$$p(X_{0:T}) = p(X_0) \prod_{t=0}^{T-1} (X_{t+1}|X_{0:t}) \tag{3}$$

where $X_t = X_{1t}, X_{2t}, \ldots, X_{nt}$ represents all the nodes in a time slice t for $t = 0, 1, \ldots, T$. In this equation, it is required all the previous time slices to be taken into account to calculate the product. This can be simplified by using the Markov assumption [9]. On the other hand, the Markovian order defines the number of time slices required to assume that the present is independent of the past. Increasing the Markovian order implies more arcs appearing from earlier lags to the present, and thus, a greater complexity when learning the network structure [11].

E. Structure and parameter learning in DBNs

The structure of the DBNs can be learned in two steps: (1) the intra-slice arcs (static structure) of the network are learned with the max-min hill-climbing (MMHC) algorithm; (2) followed by learning the inter-slice arcs (transition structure). MMHC is a hybrid learning method that searches for possible network structures with a local search and then directs the arcs and scores the networks with the Bayesian information criterion (BIC) [12]. Specifically, a modified version of the MMHC algorithm named dynamic MMHC (DMMHC) was applied [13]. Once the network structure is learned, the next step consists in estimating the network parameters for each node. This is performed based on the

maximum likelihood estimator (MLE), whose specific form depends on the parents having each node, and the assumed distribution for each variable.

F. Statistical inference with DBNs

After learning the parameters of the network, it is possible to make inferences about any unobserved nodes or system states by providing some evidence to the DBN. For instance, we can predict the most likely state of the system (i.e., apnea episode or not) over the actual interval. The evidence provided from past time slices should be used to predict the next time slice for a particular node, but also to estimate the actual value for some unobserved parameter. For Markovian orders higher than one, the prediction task consists in providing some evidence of the observed nodes for the preceding time slices and predicting the state of the desired nodes at t:

$$p(X_{i,t+1}|Y_{:,0:t}) \tag{4}$$

where $X_{i,t+1}$ represents the predicted variable at time $t+1$ and $Y_{:,0:t}$ represents the evidence (observed or known variables) from $t=0$ to T. A special case like filtering, $p(X_{i,t}|Y_{:,0:t})$, can also be of particular interest and was applied in this study [14]. In this case, $X_{i,t}$ represents the i-th hidden variables at time t.

The inference tasks usually consist of two types of queries, conditional probability (CP) and the maximum a posteriori (MAP) [15]. A CP query is performed when some evidence $E=\{X_{i1},\ldots,X_{ik}\}$ from the set X is available, allowing to estimate the conditional probability of an event involving other variables $Q=\{X_{j1},\ldots,X_{jk}\}$. In the case of MAP queries, the goal is to find the best combination of values q^* for some variables in the BN defined by Q that has the highest probability given some evidence E.

G. Data preprocessing

The time-series used in the study were preprocessed before entering the DBN to train the models. A detrend step, based on a moving average filter, was applied for each marker to remove slow oscillations and trends that affect the overall performance of the models. As we are interested in detecting only apnea episodes (two-stage regime), we do not care about the actual amplitude of the markers, which can be quite different from one recording to another. To perform this step, a 500-beat segment length was selected.

H. Model training and selection

1) Univariate conditional Lagged models: The starting point when training the DBNs was the first-order Markovian model ($\mathcal{L}=1$), where only the RR or Ra time-series was used together with the apnea information, $Ap=\{1,0\}$. Then, the Markovian order was gradually increased up to $\mathcal{L}$=15 to obtain more complex, but higher predictive models.

2) Multivariate conditional Lagged models: The same strategy described above was followed to train DBNs using both markers together. In this case, the apnea detection rate can be substantially improved at the expense of more complex networks (more arcs between variables/nodes) but with smaller lags. Figure 1 shows an example of a DBN obtained for $\mathcal{L}=3$.

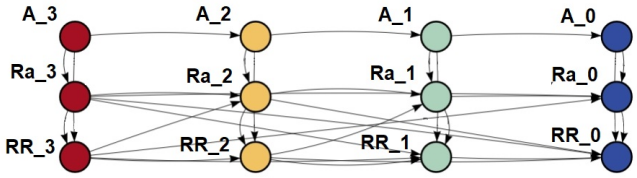

Fig. 1. Dynamic Bayesian network obtained for third-order ($\mathcal{L}$=3) Markov process using both the RR and Ra markers.

The different models described above were trained for each 15-min recurrent apnea sequence simulated in each rat. This allows us to explore the best sequence providing the more predictive model for a particular animal. The $RMSE$ and BIC values served as model's quality measures during training while used to select the optimal lag for the network.

I. Model assessment

After training the obtained DBNs in each animal and sequence, the next step is to compute the filtering density $p(\boldsymbol{x}_t|\boldsymbol{y}_{0:t})$ recursively, where y can be either the RR, the Ra or both markers, while x represent the apnea occurrence at time t. The obtained conditional probability time-series was then smoothed to reduce spurious peaks, using a moving average filter of 40-sample length. Figure 2 shows an example of this methodology for a particular sequence of five apnea episodes. This smoothed probability signal was then segmented in 15-s epochs, to decide whether it is classified as a normal or apneic segment. To do this, the RMS values were computed in each epoch and used as the final decision measure. Finally, the sensitivity (Se), specificity (Sp), and accuracy (Acc) of the models were computed and averaged for all analyzed sequences of recurrent apnea.

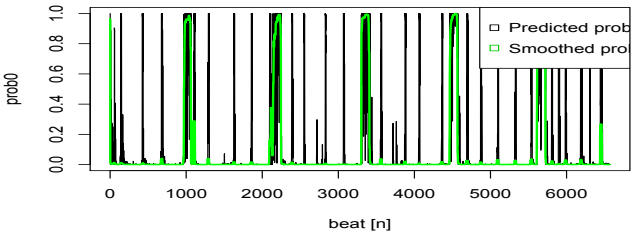

Fig. 2. Example of the conditional probability time-series obtained with a trained DBN ($\mathcal{L}$=10), and considering both the RR and Ra markers as the only observed variables.

III. RESULTS AND DISCUSSION

Figure 3 illustrates the $RMSE$ values obtained for each lag and sequence in two particular recordings. It can be observed that, from $\mathcal{L}=1$ to $\mathcal{L}=5$, the $RMSE$ values in the upper graphs decreased rapidly, and then remain at very similar values up to $\mathcal{L}=15$. Here, a similar pattern is reflected for the three sequences. However, the bottom graphs show quite different behavior for the second sequence

compared to the others. Notably, depending on the analyzed recording, the sequences with the smaller $RMSE$ values along with the lag parameter, suggest being the best sequence to train the final model. Among several potential factors, the number of apnea episodes and the noise level present in the time-series obtained for each sequence may influence the overall quality of the models.

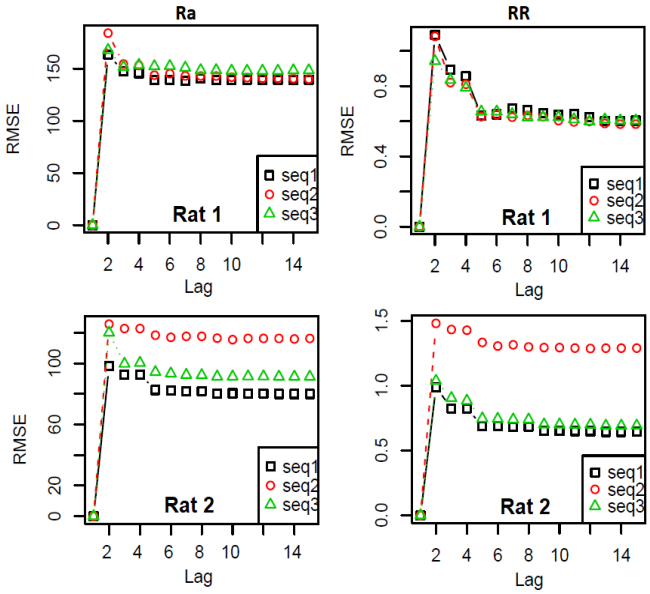

Fig. 3. Root mean square error obtained for Ra and RR when fitting the DBNs with different lags in the three apnea sequences (seq) of two rats.

Figure 4 shows the average results ($\bar{Acc}$, $\bar{Se}$, $\bar{Sp}$) obtained for the models' performance analysis. Different threshold values, $\gamma = \{0.1, 0.2, ..., 0.9\}$, were used to predict the presence/absence of apnea in a particular epoch based on their RMS estimates. A clear and fast drop of the Se values is observed as the threshold increases, unlike the slower increase of Acc and Sp, which is a bit expected. For both cases, the best threshold was 0.2 based on the Euclidean distance between one unit and the average Se and Sp values, with Acc = 0.81, Se = 0.80, Sp = 0.80 for $\mathcal{L} = 5$, and Acc = 0.81, Se = 0.90, Sp=0.78 for $\mathcal{L} = 15$. Here, the higher the Markov order, the better the sensitivity values, at the expense of greater model complexity. Therefore, the smaller lag ($\mathcal{L}$=5) would be a more suitable choice for the model.

Finally, the above results were obtained for DBNs trained with both the RR and Ra markers. When using only one of these parameters, the overall performance deteriorates drastically, suggesting the importance of multivariate models when dealing with prediction problems. Further works are needed in order to obtain similar and personalized models in humans, where OSA detection and diagnosis is much more challenging, as compared to controlled experimental models.

ACKNOWLEDGMENT

The authors are grateful to Ramon Farré, Daniel Navajas, Josep Maria Montserrat, Isaac Almendros, Marta Torres and Puy Ruiz for their excellent help in designing and implementing the animal model.

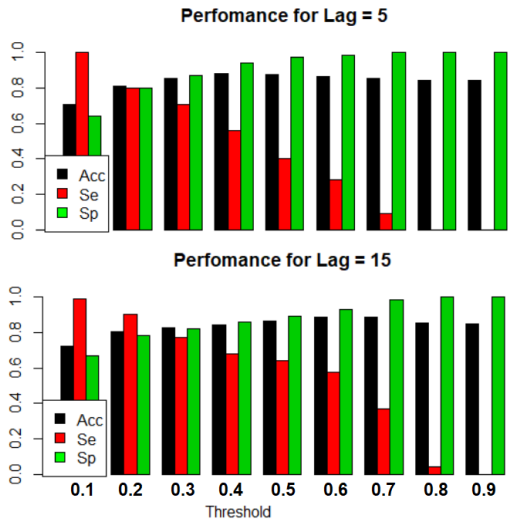

Fig. 4. Average performance metrics (Acc: accuracy, Se: sensitivity, Sp: specificity) obtained for $\mathcal{L} = 5$ (top) and $\mathcal{L} = 15$ (bottom) as a function of the detection threshold.

REFERENCES

[1] N.R. Prabhakar. Invited Review: Oxygen sensing during intermittent hypoxia: cellular and molecular mechanisms. J Appl Physiol, vol. 90, pp. 1986-1994, 2001.

[2] L. Lavie. Obstructive sleep apnoea syndrome — an oxidative stress disorder. Sleep Med Rev, vol. 7, pp. 35-51, 2003.

[3] L. Lavie. Sleep-disordered breathing and cerebrovascular disease: a mechanistic approach. Neurol Clin, vol. 23, pp. 1059–1075, 2005.

[4] F. Mendonça, S. S. Mostafa, A. G. Ravelo-García, F. Morgado-Dias and T. Penzel. A Review of Obstructive Sleep Apnea Detection Approaches. IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 2, pp. 825-837, March 2019.

[5] T. Penzel, J. McNames, P. De Chazal, B. Raymond, A. Murray, G. Moody. Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings. Medical and Biological Engineering and Computing, vol. 40, no. 4, pp. 402–407. 2002

[6] R. Jané R, J. Lázaro, P. Ruiz, E. Gil, D. Navajas, R. Farré, P. Laguna. Obstructive Sleep Apnea in a rat model: Effects of anesthesia on autonomic evaluation from heart rate variability measures. In proc. Computing in Cardiology (CinC), pp. 1011-1014, 2013.

[7] D. Romero, R. Jané. Global and Transient Effects of Intermittent Hypoxia on Heart Rate Variability Markers: Evaluation Using an Obstructive Sleep Apnea Model. IEEE Access, vol. 9, pp. 19043–19052. 2021.

[8] J.P. Martínez, R. Almeida, S. Olmos, A.P. Rocha, P. Laguna. A wavelet-based ECG delineator: evaluation on standard databases. IEEE Trans Biomed Eng, vol. 51, pp. 570-581, 2004.

[9] D. Koller N. Friedman. Probabilistic graphical models: principles and techniques. MIT press. 2009

[10] S.L. Lauritzen, N. Wermuth. Graphical models for associations between variables, some of which are qualitative and some quantitative. Annals of Statistics, pp. 31.-57. 1989.

[11] D. Quesada, G. Valverde, P. Larrañaga C. Bielza. Long-term forecasting of multivariate time-series in industrial furnaces with dynamic Gaussian Bayesian networks. Engineering Applications of Artificial Intelligence, vol. 103, pp. 104301. 2021

[12] I. Tsamardinos, L.E. Brown C.F. Aliferis. The max-min hill-climbing Bayesian network structure learning algorithm. Machine learning, vol. 65, no. 1, pp. 31–78, 2006.

[13] G. Trabelsi, P. Leray, M.B. Ayed, A.M. Alimi. Dynamic MMHC: A local search algorithm for dynamic Bayesian network structure learning. In International Symposium on Intelligent Data Analysis (pp. 392-403). Springer, Berlin, Heidelberg. 2013.

[14] K.P. Murphy. Dynamic bayesian networks: representation, inference and learning. University of California, Berkeley. 2002

[15] R. Nagarajan, M. Scutari, S. Lèbre. Bayesian networks in r. Springer, 122, 125-127. 2013