31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Sleep Staging Classification Based on HRV: Time-Variant Analysis

M.O. Mendez, M. Matteucci, S. Cerutti, F. Aletti and A.M. Bianchi

Abstract— **An algorithm to evaluate the sleep macrostructure based on heart rate fluctuations from ECG signal is presented. This algorithm is an attempt to evaluate the sleep quality out of sleep centers. The algorithm is made up by a) a time-variant autoregressive model used as feature extractor and b) a hidden Markov model used as classifier. Characteristics coming from the joint probability of HRV features were used to fed the HMM. 17 full polysomnography recordings from healthy subjects were used in the current analysis. When compared to Wake-NREM-REM given by experts, the automatic classifier achieved a total accuracy of 78.21±6.44% and a kappa index of 0.41±.1085 using two features and a total accuracy of 79.43±8.83% and kappa index of 0.42±.1493 using three features.**

I. Introduction

During last years, the sleep evaluation has become an important issue due to a considerable number of pathologies linked to the sleep. The performance of basic activities in the normal life such as memorization, learning, productivity and concentration, are closely connected to a good sleep quality. Until a few years ago, the sleep evaluation was not possible out of sleep centers; this mainly occurs for technological limitations. However, these limitations have been partially overcome with the introduction of new technologies that allow the acquisition of physiological signals with high precision in different environments. In addition to the sociological and physiological consequences produced by the low sleep quality, sleep evaluation is a time consuming task that has to be done by expert clinicians. This evaluation consists in defining differents sleep stages through visual scoring of the polysomnography (PSG). PSG includes the recording of many signals such as electroencephalography, electromyography, respiration, electrooculagram, etc. These signals are recorded during a whole night in a specialized sleep center. With the PSG procedure it is possible to observe some sleep characteristics based on specific rules related to sleep quality [1] and sleep disorders [2]. Although, the PSG is an accurate procedure, some inconveniences rise; for instance, we can cite the need of specific equipment, dedicated sleep centers and specialized and trained personnel. All these PSG characteristics have generated underestimation of the sleep pathologies and low accessibility to the general population. Then, the development of new automatic systems, for sleep screening, able to interact with the new acquisition technologies at home could be of great help.

Previous studies have shown that, during sleep the heart rate variability (HRV) presents characteristic oscillations connected to REM and NREM sleep stages [3-5], and with high discriminatory characteristics [6]. To evaluate the HRV characteristics linked to the sleep stages, it is necessary the selection of mathematical approaches suitable to deal with the nonstationarity and temporal information that HRV presents during sleep. One of the most suitable approaches is the Time-Variant Autoregressive Model (TVAM). It has fine properties to deal with non-stationary time series [7]. TVAM presents a shift moving window across the data, which permits to TVAM to characterize a time series in a variant polynomial at each signal sample. On the other hand, normal sleep presents a well-defined dynamic pattern into its macrostructure, which is characterized for the alternating between REM and NREM stages. This characteristic can be well modeled by models that are able to recognize the dynamic of a pattern in the time. One of the most suitable is hidden Markov model (HMM), which uses the pattern temporality of a time series to define the most probable present state based on the previous ones [8].

The goal of this study is to present a system that recognizes automatically Wake, REM and NREM during sleep time based only on heart rate fluctuations. The sleep stages are automatically detected using a time varying autoregressive model as feature extractor, and a hidden Markov model as probabilistic classifier.

II. Methodology

A. Protocol

17 recordings coming from healthy subjects were used in this study. These were used to develop the classifier system based on standard ECG signal. Age of the subjects ranges between 40 and 50 years. Subjects have a body mass index less than 29 kg/m^2. All subjects have an Apnea-Hypoapnea Index of zero and were drug-free. Each subject participated with one night recording. Mean sleep efficiency was 85%. All experiments were conducted at the sleep clinic of the San Raffaele Hospital. Sleep evaluation was done by expert personnel and assessed following the standard PSG procedure [1]. The acquisition system was a Heritage Digital

Manuscript received April 7, 2009. This work was supported in part by the HeartCycle Project ICT FP7 216695 of the European Community.

The authors thank to the hospital San Raffael for the polysomgraphy data and the scoring sheets.

M. O. Mendez is with the Politecnico di Milano, Milano, IT 20133 Italia (phone: 0039 02 3399 3322; fax: 0039 02 3399 3360; e-mail: martin.mendez@ biomed.polimi.it).

A. M. Bianchi is with the Politecnico di Milano, Milano, IT 20133 Italia (e-mail: annamaria.bianchi@polimi.it).

S. Cerutti is with the Politecnico di Milano, Milano, IT 20133 Italia (e-mail: sergio.cerutti@polimi.it).

F. Aletti is with the Politecnico di Milano, Milano, IT 20133 Italia (e-mail: federico.aletti@polimi.it).

M. Matteucci is with the Politecnico di Milano, Milano, IT 20133 Italia (e-mail: matteucc@elet.polimi.it).

PSG Grass Telefactor and all data was acquired with 128 Hz as sampling rate. Polysomnographic data were scored in epochs of 30 seconds each. The hypnogram was obtained as a result of this procedure. Only hypnogram and the ECG signal were used for developing the classifier system. The ECG was extracted from the polysomnography data; R peaks were detected from the ECG using a derivative built and tested algorithm, and parabolic interpolation was added in order to overcome the low R peak accuracy due to the low sampling rate. Distances between consecutive R peaks were evaluated. This procedure gave as result the tachogram (RRI). Some R peaks were misdetected and some ectopic beats were found in the ECG. Then, ECG and RR series were plotted together in order to observe clearly the erroneous detections. Where a beat or a series of beats were misdetected, these were manually corrected and the new RR series recalculated.

B. Feature extraction and classification procedure

1) Feature extractor algorithm: To deal with the different conditions that the RRI could presents across sleep stages, a TVAM was used. This is suitable to analyze signals which change their statistical properties in the time. TVAMs extract the signal characteristics in their polynomial coefficients which are self-adjusted beat-by-beat in relation to the prediction error. The prediction at each beat is obtained as follows:

$$\hat{y} = -\sum_{k=1}^{P} a_k y(n-k), \quad (1)$$

where n denotes the time, a_k are the coefficients of the model, P is the filter order (here eight) and $\hat{y}(n)$ denotes the prediction output. The prediction error is computed as:

$$e(n) = y(n) - \hat{y}(n) = y(n) + \sum_{k=1}^{P} a_k y(n-k) \quad . \quad (2)$$

The goal in adaptive models is to minimize this cost function at each time n:

$$CF_P(n) = \sum_{k=1}^{n} \lambda^{n-k} \mid e(n) \mid^2, \quad (3)$$

where k is the observation interval and λ (here 0.98) is the forgetting factor. Recursive Least Square algorithm was used to update the filter parameters [7].

2) Features: From the beat-by-beat filter coefficients the following spectral indexes were computed:

- TP; Total Power (0.003–0.5 Hz).
- LF; low-frequency component (0.02–0.15 Hz).
- HF; high-frequency component (0.15–0.5 Hz).
- pLF; frequency peak to the maximum power in the low frequency band.
- pHF; frequency peak to the maximum power in the high frequency band.

In addition to the classical spectral indexes, the modulus and phase of the representative pole in the high frequency band were extracted [9]. Spectral features were normalized with respect to the beat-by-beat total power for each recording. No normalization was applied to the pole features since these already ranged into the interval *[0 1]*. From the RRI, its mean value each 30 seconds was computed in order to have the same hypnogram resolution. The same procedure was carried out for all spectral indexes as well as for the phase and modulus of the high frequency pole.

3) Transformation and quantization: Since features did not present Gaussian distributions, a transformation procedure was applied. Logarithmic transformation was used for spectral features and square root for the modulus of the pole. Discretization process, with M equal values (here $M=10$) ranging from the minimum to the maximum feature value, was applied to each feature in each recording.

Joint probability distributions for each possible couple of features, from each recording, were computed. From each joint probability distribution, a one-dimensional feature to feed the classifier was obtained. The one-dimensional sequences of symbols were obtained applying the following codebook:

$$o(n) = (f_1(n) \times M) + f_2(n) \quad (4)$$

Where $o(n)$ is the observation to feed in the classifier, $f_1(n)$ and $f_2(n)$ are feature 1 and feature 2 of the joint probability distribution respectively and n is the epoch. Finally, M is the number of possible values, in the features, used in the discretization step.

4) Classification Algorithm: hidden Markov models present high applicability in problems with an inherent temporality. In this case we want to recognize patterns in time such as the WAKE (w_1), REM (w_2) and NREM (w_3) states. In Markov models the sequence generation is described by *transition probabilities* defined as:

$$P(w_r(n+1) \mid w_q(n)) = a_{rq} \quad (5)$$

This means, that a_{rq} (element of the transition matrix) is the time-independent probability of having state w_r at step time $n+1$ given the state w_q at time n. This is a first order discrete time Markov model since particular state w_r at time $n+1$ is a random function that depends only on the state at step n and on the transition probabilities. At any state (sleep stage) the system (central nervous system) emits observable symbols (autonomic nervous system signals) which are characteristics of the system state. In HMM this is described in the following way: in a particular state w there is a probability of emitting a particular observation o_k (feature value), then $P(o_k|w_r)=b_{rk}$, where b_{rk} are called elements of the emission matrix. Such a model is called *hidden Markov model* [8] since we suppose to have only access to the observations and not to the w_r states.

C. Classification Performance

The learning and performance procedure was carried out from the whole database using Leave-One-Out cross-validation technique (LOOCV). Since a supervised learning was applied, the emission matrix was computed from the hypnograms and the emission matrix from the sequence of symbols. Thus the LOOCV procedure consists in: a) holding the symbol sequence (o) and the hypnogram of one

recording out, b) evaluating the emission matrix and transition matrix from the symbol sequences and hypnograms of the remaining recordings, c) decoding the symbol sequence of the recording that was left out and d) calculating the statistical classification measures of accuracy, sensitivity and specificity by comparing the hypnogram and the sequence of states obtained from the HMM decoding. Afterwards, again from the whole set another recording (never the same) was left out and the transition and emission matrices and the measures of classification where again evaluated. This procedure was repeated until each recording was left out once. Finally we obtain the mean performance for the 17 trials. The same procedure was also extended for joint probabilities composed of 3 features. The hypnograms were simplified in: 1) Wake = Wake and stage 1, 2) NREM = stage 2, 3 and 4 and 3) REM. The best performances were obtained with the following combination of features:

- VLF - Modulus pole
- VLF - Modulus pole - TP

Sleep Efficiency was computed from the sequence of stages given by the HMM, defined as the number of epochs in REM-NREM divided by the total number of epochs. This procedure was done for symbol sequences coming from bivariate and trivariate joint probability distributions.

III. Results

Mean performances, of the automatic systems used to identify the sleep macrostructure from heart rate fluctuations, are presented in Table I. Bivariate and trivariate joint distributions showed accuracy close to 80%. However, the agreement level is around 0.41 which means a moderate agreement. Mean sleep efficiency was similar to that given by the physicians with both the bivariate or trivariate joint distributions. Note that the percentage of wake and REM sleep epochs is similar.

From the top to the bottom, Figure 1 shows the hypnogram, the sleep profile obtained with the bivariate probability distribution and the sleep profile obtained with the trivariate probability distribution. One can observe that the dynamic of the hypnogram is maintained by the sleep profiles obtained by HMM, independently if a bivariate or the trivariate probability distribution was used. One can also observe from Figure 1, that fast states transitions (i.e., when there is one wake epoch between two NREM epochs) were not detected by HMM. This is practically driven for the emission and transition matrix, and as a result we obtain a smoothed version of the hypnogram.

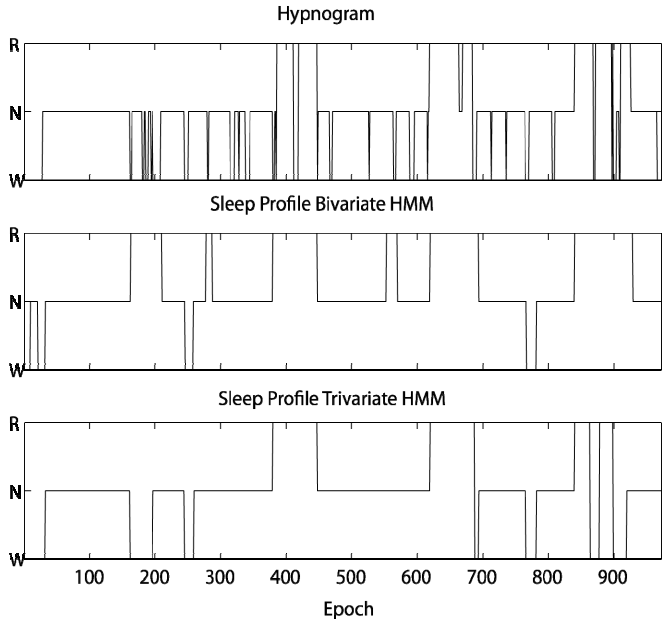

Figure 1. Hypnogram example of a single night. The top plot shows the hypnogram evaluated by an expert based on PSG but simplified to 3 stages. Middle plot shows the sleep profile obtained from bivariate joint probability distributions and bottom panel illustrates the sleep profile obtained from trivariate joint probability distributions. R is REM, N represents non-REM and W is wake.

IV. Discussion

An automatic system that identifies the wake, NREM and REM sleep based only on HRV signal was presented. This system used as feature extractor a time-variant autoregressive model and as classifier a hidden Markov model. Our main claims are: 1) TVAMs seems to be fine tools to extract characteristics with high discriminatory power from the HRV signal to sleep staging. This allows the evaluation of the sleep macrostructure. 2) HMM presents interesting properties that allow to detect the sleep dynamic, 3) Joint probability distributions seem to offer good discriminatory power for sleep staging.

HRV presents slow and fast fluctuations during sleep. This variations produce non-stationarieties that sometimes forbid the application of fundamental techniques such as Fourier Transform. When we deal with this kind of situations, different techniques such as TVAMs offer a different perspective that allows the evaluation of signals even during non-stationary periods. However, even if the features obtained from TVAMs could represent the real dynamic of the time series, these features could not have high discriminatory power for classifying sleep stages. In a first

Table I. Mean and standard deviation of accuracy and agreement measure for the sleep staging obtained by HMM.

	Heart Rate Fluctuations and Movements						
	ACC	**Kappa**	**SeHMM**	**SeHyp**	**%Wake**	**%REM**	**%NREM**
Bivariate	78.21±6.44	0.41±.1085	79.96±14.33	81.86±11.25	18.14±11.25	14.78±5.16	67.09±7.85
Trivariate	79.43±8.83	0.42±.1493	80.46±10.74	81.86±11.25	18.14±11.25	14.78±5.16	67.09±7.85

ACC means general accuracy, kappa is kappa index, SeHMM is the sleep efficiency obtained by the automatic system, Sehyp represent the sleep efficiency obtained from the standard hypnogram. % Wake, %REM and %NREM are the percentage for wake, REM and NREM epochs found in all recordings.

tentative, we tried to separate Wake from REM by feeding single features to HMM. Unfortunately, only good separation between REM and NREM was possible [9]. In the current study, we combined the features by evaluating their join probability distribution (and after one dimensional code sequences was computed to fed mono-dimensional HMM), and from Table I one can observe that good results are obtained using a low dimensional feature space and a low computational cost is maintaining. These results are in line with the results presented in the current literature. Redmon et. al. [6] have presented an interesting algorithm to detect Wake-NREM-REM stages based on features extracted from HRV and respiratory surrogates. Their results were similar to the ones presented here.

Although the results obtained by Redmon et. al. and those presented here seems motivating, both algorithms suffer of the same problematic. It is to say, that the separation between Wake and REM seems to be a challenging problem when HRV or any peripheral signal (based on heart fluctuations, respiration and vascular activity) is used. Only a moderate agreement is achieved due to difficulties in separating REM and wake. This is because the ECG presents similar characteristics during wake (in rest conditions) and REM. Redmon et. al. [6] had explored different characteristics from the ECC, different features extracted from the HRV signal and higher dimensional feature spaces, unfortunately from our results and those of Redmon et. al. [6], it seems that the HRV signal could be not enough to completely and accurately classify the sleep macrostructure. However, one can correlated the percentage of rapid and low EEG oscillation Wake-REM and NREM respectively, to the time in bed and sleep efficiency. In addition, the temporal dynamic of Wake-REM and NREM can give an idea of the sleep dynamic. On the other hand, probably this kind of system does not offer so accurate detection for sleep, but can be useful for a rough sleep screening in places out of the sleep center. The drawbacks presented by HRV, when Wake and REM needs to be separated, motivate to a deep exploration of different signal processing techniques and a profound research of the characteristics in the ECG signal that could be useful to overcome this limitation. Detection of sleep macrostructure based on ECG is interesting since this signal presents advantages such as low noise/signal ratio and its acquisition is simple.

On the other hand, since the proposed algorithms are based only on the heart rate fluctuations, signals that contain similar information could be used, as is the Ballistocardiogram case. One example, in which these types of algorithms could be applied, is on the bed mattress sensors such as the one proposed by [10]. They obtained the heart rate fluctuations based on multi-channel BCG signals measured by sensors placed in a bed. In addition, these heart rate fluctuations can also come from other type of signals, which are captured by other systems such as the photopletysmography. In the current study, we have applied HMM as classifier; however, it could be interesting to try other methods. This is because probably modeling the sleep dynamic (the a priori probability of the sleep stage) could be not so important to define the sleep stages when the decision is done from the feature characteristics.

V. CONCLUSION

Heart rate fluctuations have valuable information of the sleep macrostructure. This information can be used to develop simple automatic systems and could give an index of sleep quality out of sleep centers. TVAMs seem to be useful to capture the time-variant characteristics of the autonomic nervous system, which are tie linked to the central nervous activity. HMM offers interesting characteristics to evaluate the sleep macrostructure and define sleep stages based on the time series acquired from the peripheral signals.

REFERENCES

[1] A. Rechtschaffen, and A. E. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," *U.S. Government Printing Office*, NIH Publication No. 204, Washington, DC., 1968.

[2] Report of 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, 1999.

[3] U. J. Scholz, S. Cerutti, S. Kubicki, and A. M. Bianchi, "Vegetative background of sleep: spectral analysis of the heart rate variability," *Physiol. Behav.*, vol. 62, No. 5, pp.1037–1043, 1997.

[4] Task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, pp.1043–1065, 1996.

[5] T. Penzel. J. W. Kantelhardt, L. Grote, J. Peter and A. Bunde, "Comparison of detrend fluctuation analysis and spectral heart rate variability in sleep and sleep apnea," *IEEE Trans Biomed Eng*. vol. 50, pp. 1143-1151, 2003.

[6] S. J. Redmond, P, De Chazal, C. O'Brien, S. Ryan, T. W. McNicholas, and C. Heneghan. "Sleep staging using cardiorespiratory signals", *Somnologie*, vol. 11, pp.245–256, 2007.

[7] A. M. Bianchi, E. Petrucci, M. G. Signorini, L. T. Mainardi, and S. Cerutti, "Time-variant power spectrum analysis for the detection of transient episodes in HRV signal," *IEEE Trans. Biomed. Eng.*, vol. 40, No. 2, pp.136–144, 1993.

[8] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp.257–286, 1989.

[9] M. O. Mendez, M. Matteucci, V. Castronovo, L. Ferini-Strambi, S. Cerutti and A. M. Bianchi, "Sleep staging from Heart Rate Variability: time-varying spectral features and Hidden Markov Models," *IJBET*, in press.

[10] Kortelainen J., Virkkala J. "FFT averaging of multichannel BCG signals from bed mattress sensor to improve estimation of heart beat interval," 29th IEEE EMBS Annual International Conference, August 23-26, 2007, Lyon, France.