34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Automatic Detection of REM Sleep in Subjects without Atonia

Jacob Kempfner [a,d], Poul Jennum [b,c,d], Miki Nikolic [c], Julie A. E. Christensen [a] and Helge B. D. Sorensen [a]

Abstract— **Idiopathic Rapid-Rye-Movement (REM) sleep Behavior Disorder (iRBD) is a strong early marker of Parkinson's Disease and is characterized by REM sleep without atonia (RSWA) and increased phasic muscle activity. Current proposed methods for detecting RSWA assume the presence of a manually scored hypnogram. In this study a full automatic REM sleep detector, using the EOG and EEG channels, is proposed. Based on statistical features, combined with subject specific feature scaling and post-processing of the classifier output, it was possible to obtain an mean accuracy of 0.96 with a mean sensititvity and specificity of 0.94 and 0.96 respectively.**

I. Introduction

Rapid-Eye-Movement (REM) Sleep Behavior Disorder (RBD) is characterized by REM sleep without atonia (RSWA) and consequently increased muscle tone and excessive phasic muscle twitch activity of the submental or limb surface electromyographic (EMG) measures [1]. RBD without current sign of neurodegenerative disorder is designated as idiopathic RBD (iRBD). This term is questioned as RBD and other non-motor symptoms and findings often observed in Parkinson's Disease (PD) and atypical PD such as multiple system atrophy and Lewy Body Dementias. Additionally, more than 50% of the subjects diagnosed with iRBD will develop a synnocleinopathy within a time span of 5-10 years [2-7].

Correct detection of RBD is therefore highly important, provided that neuroprotective treatment becomes available. All proposed methods for detecting RSWA assume the presence of a manual scored hypnogram [8-13]. The aim of the current study is to automatically detect the REM sleep stage according to the new international sleep-scoring standard from the American Academy of Sleep Medicine (AASM) [14]. According to the AASM, a sleep stage epoch of 30 seconds must be scored as REM, when the electro-encephalography (EEG) has low amplitude with mixed frequencies (i.e. 4-7 Hz) in the frontal, central and occipital electrodes Furthermore, there should also be relatively low muscle tone in the chin. If there are no indications of another sleep stage between the REM events in the electro-oculography (EOG) channels, it is assumed to be REM sleep. More detailed description can be found in [14].
In a previous study [15] it was concluded, based on the data and method, that the muscle tone had little, if any, influence on the REM sleep detection. However, the method has been enhanced by subject-specific feature scaling, and post processing of the classifier output. The features from each subject were re-scaled by a modified min-max method, where the minimum and maximum values were estimated by percentiles. Furthermore, the binary classifier output was post-processed by averaging the classifier output using a Blackman window and then compared to a user-defined threshold.

[a] Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark.
[b] Danish Center for Sleep Medicine, Glostrup University Hospital, Denmark.
[c] Department of Clinical Neurophysiology, Glostrup University Hospital, Denmark.
[d] Center for Healthy Ageing, University of Copenhagen, Copenhagen, Denmark.

II. Data

A. Subjects and Demographics

A total of 16 subjects from the Danish Center for Sleep Medicine, Department of Clinical Neurophysiology, Glostrup University Hospital, Denmark, were enrolled in this study. The subjects were divided into two groups categorized by their diagnosis. The demographics of the groups are summarized in Table I.

TABLE I. Demographics

	N° subjects (♀, ♂)	**Age ($\mu \pm \sigma$) [years]**
Control	8 (5, 3)	60.8 ± 9.4
iRBD	8 (3, 5)	60.6 ± 8.5

None of the involved subjects were taking any medication which was known to affect the sleep. The presented data did not allow us to balance the two groups in age and gender.

B. Data Acquisition and Scoring of Hypnograms

All subjects underwent one full night polysomnography in accordance to the AASM. This is equivalent to approximately eight hours of sleep per subject when using data from lights-off to light-on. In this study only the left and right EOG channel combind with the F_3-A_2, C_3-A_2 and O_1-A_2 EEG channel were used. A_2 denotes the right mastoid. To ensure the quality of each recording, visual inspection of all the recordings were conducted, and corrupted recordings in which the analysis channels were disconnected or continuously contaminated with artifacts were rejected. The sampling frequncy of the analyzed sleep data was 256 Hz. The sleep data were analyzed in MATLAB (R2010b, 64-bit, The MathWorks, Natick, MA, USA)).

C. Sleep Stage Distribution

The total number of recorded epochs with a duration of 30 seconds are summarized in Table II.

TABLE II. Total number of Manual scored 30-second epochs

	Wake (%)	**NREM (%)**	**REM (%)**	**∑(100%)**
Control	1645 (20)	5159 (61)	1588 (19)	8392
iRBD	1985 (25)	4696 (60)	1182 (15)	7863

The non-REM sleep, which is denoted NREM in Table II, corresponds to the intersection of the three non-REM stages NREM-1, NREM-2 and NREM-3 respectively.

III. METHOD

In REM sleep the eyes tend to move rapidly sideways under closed eyelids. This produces rapid conjugated eye movements, and since the eye acts as a dipole it appears as "out-of-phase" EOG channel deflections. Furthermore, the EOG channel will also measure the muscle activity surrounding the eye. It is assumed that the "lower" frequencies characterize the rapid eye movements, while the "higher" frequencies characterize the surrounding muscle activity. In [17] a 4th order Butterworth bandpass filter with cutoff frequencies at 1 and 5 Hz respectively was successfully used to separate the REM from Slow-Eye-Movement, baseline drift and EMG activity. This superposition property may be exploited to measure the increased muscle activity during wake. According to [14] it is recommended that the EOG signals should be viewed in the frequency band 0.3-35 Hz, while the EMG should be viewed in the frequency band 10-100 Hz. In this study a less narrow band was used. The left and right EOG channel, denoted EOGL and EOGR respectively, were band-pass filtered by 4th order Butterworth bandpass filters using the zero-phase filtering approach. The frequency bands are defined in Table III.

TABLE III. FREQUENCY BANDS OF THE EOG CHANNELS

Modality	Band	Low-Cut (3dB)	High-Cut (3dB)
EOGL	ρ^{L}	1 Hz	10 Hz
EOGL	φ^{L}	10 Hz	45 Hz
EOGR	ρ^{R}	1 Hz	10 Hz
EOGR	φ^{R}	10 Hz	45 Hz

The ρ-band (1-10 Hz) is assumed to characterize REM, while the φ-band (10-45 Hz) is assumed to contain EMG activity. The data was recorded over a period of three years, using different amplifier systems, where the lowest cut-off frequncy of the anti-aliasing filters was 65 Hz. During REM sleep the EEG has low amplitude with mixed frequency content, also known as the background sleep EEG. The frequency range is typically 4-7 Hz. However, this may also occur in the NREM sleep [14]. The EEG channels (F_3-A_2, C_3-A_2, O_1-A_2) was filtered into the clinical bands by 4th order Butterworth bandpass filters, using the zero-phase filtering approach. The frequency bands are defined in Table IV.

TABLE IV. FREQUENCY BANDS OF THE EEG CHANNELS

Modality	Band	Low-Cut (3dB)	High-Cut (3dB)
EEG	δ	1 Hz	4 Hz
EEG	θ	4 Hz	8 Hz
EEG	α	8 Hz	13 Hz
EEG	β	13 Hz	30 Hz
EEG	γ	30 Hz	45 Hz

It was not necessary to filter the power-line noise in both modalities, due to the selcted bandwidth.

A. Data Segmentation

The preprocessed sleep data, from a given subject, was then segmented into mini-epochs of 3-second length, which is widely used in the sleep community. Notice, in sleep scoring an epoch is normally 30 seconds. From each mini-epoch a total of 18 characterizing features were extracted from the EEG and EOG channels. This was conducted on all 16 subjects.

B. Feature Extraction

The muscle tone was analyzed by computing the *inter-quartile-range* (IQR) of the φ^{L} and φ^{R} band. The IQR is assumed to be less affected by noise and artifacts, such as e.g. EMG and ECG artifacts. That is, if the difference between the 3rd quartile (75th percentile) and 1st quartile (25th percentile), for vector X is given by:

$$\mathrm{Q}_{31} \stackrel{\text{def}}{=} \mathrm{Q}_3(\mathrm{X}) - \mathrm{Q}_1(\mathrm{X}) \tag{1}$$

Then the IQR features of φ^{L} and φ^{R}, which are more likely to represent the muscle tone, is given by:

$$\mathrm{Y}_n^{\varphi^{\mathrm{L}}} = \mathrm{Q}_{31}(\varphi_n^{\mathrm{L}}) \tag{2}$$

$$\mathrm{Y}_n^{\varphi^{\mathrm{R}}} = \mathrm{Q}_{31}(\varphi_n^{\mathrm{R}}) \tag{3}$$

Where $1 \leq n \leq N$ denotes the mini-epoch number of total N mini-epochs. Several percentile combinations have been tested, but the 25th percentile and 75th percentile did perform best on our data. The characterization of the rapid-eye-movements was analyzed by computing the normalized-cross-correlation-coefficients of lag 0 (NCCC) of the ρ^{L} and ρ^{R} band respectively. The NCCC is given by:

$$\mathrm{C}_n = \frac{\mathrm{cov}(\rho_n^{\mathrm{L}}, \rho_n^{\mathrm{R}})}{\sqrt{\mathrm{var}(\rho_n^{\mathrm{L}})\,\mathrm{var}(\rho_n^{\mathrm{R}})}} \tag{4}$$

where *cov* denotes the sample covariance between the two variables, while *var* denotes the sample variance. The three EEG channels were analyzed with respect to the clinical bands defined in Table IV using the IQR approach defined in (1):

$$\mathrm{Y}_n^{\delta} = \mathrm{Q}_{31}(\delta_n) \tag{5}$$

$$\mathrm{Y}_n^{\theta} = \mathrm{Q}_{31}(\theta_n) \tag{6}$$

$$\mathrm{Y}_n^{\alpha} = \mathrm{Q}_{31}(\alpha_n) \tag{7}$$

$$\mathrm{Y}_n^{\beta} = \mathrm{Q}_{31}(\beta_n) \tag{8}$$

$$\mathrm{Y}_n^{\gamma} = \mathrm{Q}_{31}(\gamma_n) \tag{9}$$

A total of 3 EOG features were computed and 5 EEG features from 3 EEG channels were computed. The features were then merged into a [1x18] feature vector denoted $\mathrm{F}_n^{\mathrm{S}}$, where $\mathrm{S} = \{1,2,\ldots,16\}$ denotes the subject number.

C. Feature Scaling

To avoid features with greater dynamic range dominating those with smaller dynamic range, each feature was re-scaled into the range of approximately [0,1]. That is, the individual features from each subject were re-scaled according to a modified min-max method. The subject-specific feature scaling is given by:

$$\tilde{\mathrm{F}}^{\mathrm{S}} = \frac{\mathrm{F}^{\mathrm{S}} - \min(\mathrm{F}^{\mathrm{S}})}{\max(\mathrm{F}^{\mathrm{S}}) - \min(\mathrm{F}^{\mathrm{S}})} \tag{10}$$

Since the original min-max scaling method is not robust towards outliers, the minimum and maximum were estimated by computing the quartiles of the individual features. The upper and lower boundary is given by:

$$\min(\mathrm{F}^{\mathrm{S}}) \triangleq^{\mathrm{def}} \mathrm{Q}_1(\mathrm{F}^{\mathrm{S}}) \tag{11}$$

$$\max(\mathrm{F}^{\mathrm{S}}) \triangleq^{\mathrm{def}} \mathrm{Q}_3(\mathrm{F}^{\mathrm{S}}) \tag{12}$$

Again, several good percentile combinations have been tested, but the 25$^{\mathrm{th}}$ percentile and 75$^{\mathrm{th}}$ percentile did perform best on our data.

D. The Nearest Neighbor Classifier

The objective of a classifier is to classify the feature sampels into the respective classes, i.e. the REM sleep versus every-thing else. The standard Nearest Neighbor classifier (*k*-NN) identify the *k* nearest points from the training data set and then assign a new test point to the class having the largest number of representatives among this set. In this study the Euclidian distance was used to measure the distance between points [18]. The *k*-NN is a widely used supervised learning method and has successfully been applied to different areas, including sleep analysis [19].

E. Training of the k-NN Classifier

The manual hypnogram was modified into a target vector by first labeling the REM sleep epochs '+2' and everything else '+1'. The target vector was then extended by successfully repeating each epoch 10 times. This increases the "sampling rate" from one scoring per 30 second (epoch) to 10 scorings per 30 second, which is equivalent to 1 scoring per 3 second (mini-epoch), cf. Fig. 3. In this study the *k*-NN classifier was used and the *k* variable was found by using a simple grid-search approach combined with the leave-one-out cross validation scheme. Since the samples from each subject may be correlated, a fold consists of whole subjects. A single fold was held out for testing, while the remaining 15 folds were used for training. This is done 16 times, so each fold is used for testing. The individual subject outputs were then post-processed.

F. Post-processing

The agreement between the detected REM sleep and the manually scored REM sleep (i.e. hypnogram) was improved by post-processing the *k*-NN classifier output. Normally, the NREM and REM sleep tend to alternate through the night in cyclial fashion, where REM sleep usually occure in 4-6 discrete episodes each lasting between 5-20 min [20]. This can be seen in Fig. 3. This trend was enhanced by post processing the REM sleep candidates from the output y of the classifier. In this study the post-processing was obtained by filtering the binary output of each individual subject with a normalized Blackman window defined as:

$$w(m) = \frac{1}{\tau}\left(0.42 - 0.5\cos\left(\frac{2\pi m}{M}\right) + 0.08\cos\left(\frac{4\pi m}{M}\right)\right) \tag{13}$$

where $0 \le m \le M$ and M is the estimated duration of the Blackman window. The τ corresponds to the normalization coefficient defined as:

$$\tau = \sum_{m=0}^{M}\left(0.42 - 0.5\cos\left(\frac{2\pi m}{M}\right) + 0.08\cos\left(\frac{4\pi m}{M}\right)\right) \tag{14}$$

Then the post-processed output is given by symmetrical averaging:

$$\tilde{y}(n) \triangleq^{\mathrm{def}} w * y \tag{15}$$

The post-processed output was then classified into the two classes REM and NREM by using an estimated threshold (same threshold for all subjects). The threshold was estimated by testing different threshold values, denoted T_{avg}, in the grid-search approach. The T_{avg} was varied from 0 to 1 in 0.01 steps.

IV. Results and Discussion

A total of two three variables have to be estimated (k_{avg}=31, T_{avg}=0.18, *M=450s*). All variables were estimated by a simple grid-search approach combined with the leave-one-out cross validation scheme. The variable combination with the highest mean accuracy over the test folds was then chosen. The results are summarized in Fig. 1 and Fig. 2

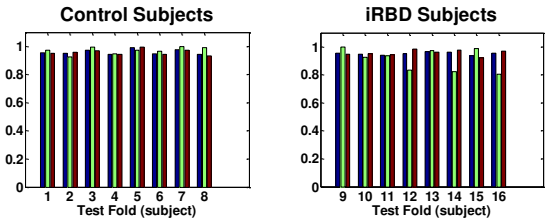

Fig. 1. Accuracy (blue), Sensitivity (green) and Specificity (red) of each test fold (subject) in both classes.

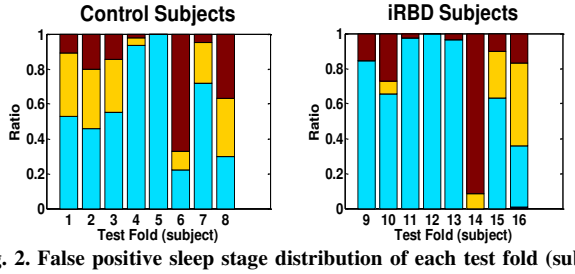

Fig. 2. False positive sleep stage distribution of each test fold (subject) in both classes. NREM-3, NREM-2, NREM-1 and Wake corresponds to blue, cyan, yellow and red respectively.

The performance of each test fold, which corresponds to each subject, is shown in Fig. 1. Furthermore, the

performance is also divided into the two classes control and iRBD. The mean accuracy, sensitivity and specificity of all 16 test folds are 0.96, 0.94 and 0.96 respectively. However, a small, but insignificant, sensitivity drop in the iRBD class can be seen, especially from subject 12, 14 and 16 (Wilcoxon rank sum test, two-sided, 5%, p=0.14). Furthermore, it appears that the primary false positive comes from the NREM-2 sleep stage, cf. Fig. 2. The manual scorings used for training the k-NN classifier are labeled with uncertainties. The difficulty of scoring REM sleep in iRBD patients may have an influence on the performance. In this study the subjects were only scored once. Multi-scorings (at least three) of the REM sleep may reduce some of the uncertainties. A more accurate result may also be obtained if the hypnogram were scored in 3 second mini-epochs instead of the traditional 30 second epochs, especially in the transition regions where epochs may be a mixture of stages. The proposed method does have some disadvantages, especially when REM sleep is highly fragmented, cf. Fig. 3.

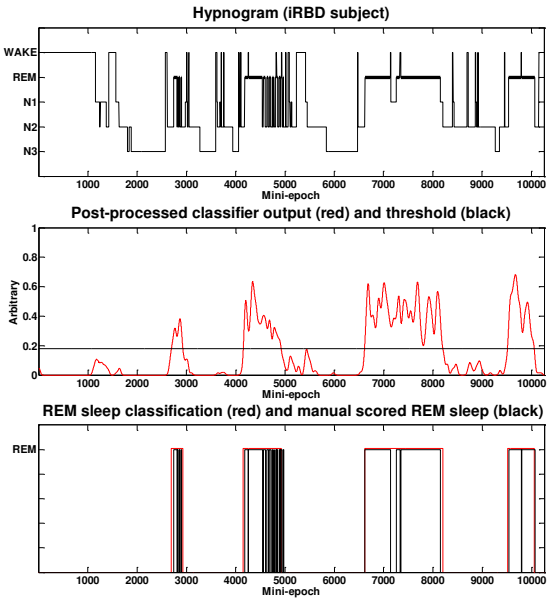

Fig. 3. The first plot shows the manually scored hypnogram, while the second plot shows the outcome of the post-processing scheme. The third plot shows the final classification with respect to the manual scoring.

When the subject frequently "jumps" in and out of REM sleep, the moving average post-processing scheme, due to the large window duration, tends to merge the small fragmentations together. This occurs in the first and second REM sleep period in Fig. 3. The k-NN classifier is highly influenced by the number of training patterns in each class. Hence, decisions in the overlapping regions will be influenced by the dominating class, which in this case corresponds to NREM. To overcome this a Support Vector Machine (SVM) could be used as classifier. Hopefully, this would improve the overall classification, and shorten the moving average window duration.

V. Conclusion

An enhanced min-max subject-specific feature scaling in combination with post-processing of the k-NN classifier output, using a moving average approach, was proposed and tested. It was possible to correctly classify REM sleep epochs with a mean accuracy of 0.96. The robust method reflects the potential of detecting REM sleep in subjects with risk of developing PD. In this context, this is the first step of analyzing REM sleep without atonia automatically. The proposed method tends to merge small REM sleep fragmentations together. This may be addressed by using an alternative classifier, such as the SVM, combined with a smaller moving average window.

References

[1] American Academy of Sleep Medicine, *International Classification of Sleep Disorders. 2nd edition: Diagnostic and Coding Manual*, Westchester, IL: American Academy of Sleep Medicine, 2005.

[2] C. H. Schenck, "Delayed emergence of a parkinsonian disorder in 38% of 29 older men initially diagnosed with idiopathic rapid eye movement sleep behavior disorder", Neurology 1996;46(2):388–93.

[3] B. F. Boeve, "Association of REM sleep behavior disorder and neurodegenerative disease may reflect an underlying synucleinopathy", Mov Disord 2001;16(4):622–30.

[4] A. Iranzo, "Characteristics of idiopathic REM sleep behavior disorder and that associated with MSA and PD", Neurology 2005;65(2):247–52.

[5] A. Iranzo, "Rapid-eye-movement sleep behaviour disorder as an early marker for a neurodegenerative disorder: a descriptive study". Lancet Neurol 2006; 5(7):572-77.

[6] R. B. Postuma, "Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behavior disorder", Neurology, 2009;72(15):1296-300.

[7] R. B. Postuma, "Severity of REM atonia loss in idiopathic REM sleep behavior disorder predicts Parkinson disease", Neurology, 2010;74(3):239-44.

[8] J. W. Burns, "EMG Variance During Polysomnography As An Assessment For REM Sleep Behavior Disorder", SLEEP, 2007;30(12):265-71.

[9] G. Mayer, "Quantification of Tonic and Phasic Muscle Activity in REM Sleep Behavior Disorder", J Clin Neurophysiol, 2008;25(1): 48-55.

[10] J. Fairley, "Sequential Feature Selection Methods for Parkinsonian Human Sleep Analysis", 7th Mediterranean Conference on Control & Automation 2009: 1468-73.

[11] M. Shokrollahi, "Autoregressive and Cepstral Analysis of Electromyogram in Rapid Movement Sleep", WC IFMBE Proceedings 2009a;25(4):1580-3.

[12] M. Shokrollahi, "Analysis of the electromyogram of rapid eye movement sleep using wavelet techniques", Conf Proc IEEE Eng Med Biol Soc. 2009:2659-62.

[13] J. Kempfner, "REM Behaviour Disorder detection associated with neurodegenerative diseases", Conf Proc IEEE Eng Med Biol Soc. 2010:5093-6.

[14] Iber C, ed., *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification*, Westchester, IL: The American Academy of Sleep Medicine; 2007.

[15] J. Kempfner J, "Automatic REM Sleep Detection Associated with Idiopathic REM Sleep Behavior Disorder", Conf Proc IEEE Eng Med Biol Soc 2011:6063-6.

[16] F. Gustafsson, "Determining the initial states in forward-backward filtering", Signal Processing, IEEE Transaction on 1996; 44(4): 988-92.

[17] R. Agarwal, "Detection of Rapid-eye movements in sleep studies", Biomedical Engineering, IEEE Transactions on 2005; 52(8): 1390-1396.

[18] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer; 1st ed., 2006, pp. 124-127.

[19] S. Gudmundsson, "Automatic Sleep Staging using Support Vector Machine with Posterior Probability Estimations", CIMCA-IAWTIC'05.

[20] M. H. Kryger, T. Roth, W. C. Dement, *Principles and Practice of SLEEP MEDICINE*, Elsevier Inc; 4th ed., 2005, pp. 13-23.