DTU Library

REM Behaviour Disorder Detection Associated with Neurodegerative Diseases

Kempfner, Jacob; Sørensen, Gertrud Laura; Zoetmulder, Marielle; Jennum, Poul; Sørensen, Helge Bjarup Dissing

Link back to DTU Orbit

REM Behaviour Disorder Detection Associated with Neurodegenerative Diseases

Jacob Kempfner[†] [*], Gertrud Sorensen[†], Marielle Zoetmulder[‡] [*], Poul Jennum[‡] [*] and Helge B.D. Sorensen[†]

[†]Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark
[‡]Danish Centre for Sleep Medicine, Glostrup University Hospital, Glostrup, Denmark
[*]Center for Healthy Ageing, University of Copenhagen, Copenhagen, Denmark

Abstract— Abnormal skeleton muscle activity during REM sleep is characterized as REM Behaviour Disorder (RBD), and may be an early marker for different neurodegenerative diseases. Early detection of RBD is therefore highly important, and in this ongoing study a semi-automatic method for RBD detection is proposed by analyzing the motor activity during sleep. *Method*: A total number of twelve patients have been involved in this study, six normal controls and six patients diagnosed with Parkinsons Disease (PD) with RBD. All subjects underwent at least one ambulant polysomnographic (PSG) recording. The sleep recordings were scored, according to the new sleep-scoring standard from the American Academy of Sleep Medicine, by two independent sleep specialists. A follow-up analysis of the scoring consensus between the two specialists has been conducted. Based on the agreement of the two manual scorings, a computerized algorithm has been attempted implemented. By analysing the REM and non-REM EMG activity, using advanced signal processing tools combined with a statistical classifier, it is possible to discriminate normal and abnormal EMG activity. Due to the small number of patients, the overall performance of the algorithm was calculated using the leave-one-out approach and benchmarked against a previously published computerized/visual method. *Results*: Based on the available data and using optimal settings, it was possible to correctly classify PD subjects with RBD with 100% sensitivity, 100% specificity, which is an improvement compared to previous published studies. *Conclusion*: The overall result indicates the usefulness of a computerized scoring algorithm and may be a feasible way of reducing scoring time. Further enhancement on additional data, i.e. subjects with idiopathic RBD (iRBD) and PD without RBD, is needed to validate its robustness and the overall result.

I. INTRODUCTION

Idiopathic RBD (iRBD) may be an early marker for different neurodegenerative diseases, such as Parkinsons disease (PD) and atypical Parkinsons disease, i.e. multiple system atrophy (MSA) and dementia with Lewy bodies (DLB) [1] [2]. Approximately 50% of the subjects diagnosed with iRBD will develop a PD [3]. Detection of iRBD in patients, who later on develop a neurodegenerative disease, is therefore highly important, provided that neuro protective treatment eventually becomes available. In normal subjects the chin electromyographic (EMG) activity in non-REM (N1, N2 and N3) is higher than in REM. This normal balance tends to be distorted in subjects with RBD. According to the International Classification of Sleep Disorders, RBD is characterized by the intermittent loss of REM sleep electromyographic atonia, and by the appearance of elaborate motor activity associated with dream mentation [4]. The scoring criterias [5] may be interpreted individually, which is a challenge when trying to develop a computerized method for RBD detection. No acceptable computerized method for RBD detection exists, but several attempts have been proposed [6] [7] [8] [9] [10], where [6] has been focusing on neurodegenerative diseases, using a relatively large number of subjects, with promising results. For that reason, [6] has been chosen as a reference algorithm, and benchmarked against this new proposed algorithm. This was obtained by implementing the reference algorithm, using the available data. The reference algorithm calculates the variance of the chin EMG mini-epochs (three seconds), and compares the calculated variance in REM sleep to a fixed threshold defined by the variance during non-REM sleep. The percentage of all REM mini-epochs above the threshold defines the motor activity during REM sleep, also called STREAM (Supra-Threshold-REM-EMG-Activity-Metric). The reference method correlates well with previously published visual RBD detection methods, and it was able to detect RBD with 100% sensitivity, 71% specificity and an area under the ROC curve of 0.84. Furthermore, a recent study concludes that the submental REM EMG baseline activity plays an important role in predicting PD, whereas the phasic activity did not [3]. Based on the study above, a new semi-automatic method for RBD detection is proposed.

II. METHODOLOGY

A. Subject Selection

A ongoing comprehensive medical history and database search for usable candidates were conducted, and a total of twelve subjects, from the Danish Centre for Sleep Medicine, Department of Clinical Neurophysiology, Glostrup University Hospital, Denmark, were selected for this study. Six subjects (2 females, 4 males, age: 62.5 ± 9.1 years) were

J. Kempfner, Electrical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark `jke@elektro.dtu.dk`

G. L. Sorensen, Electrical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark `s052926@student.dtu.dk`

H. B. D. Sorensen, Electrical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark `hbs@elektro.dtu.dk`

M. Zoetmulder, Center for Healthy Ageing, University of Copenhagen, Copenhagen, Denmark `marielle_zoetmulder@hotmail.com`

P. Jennum, Danish Centre for Sleep Medicine, Department of Clinical Neurophysiology, Glostrup University Hospital, Denmark `poje@glo.regionh.dk`

diagnosed with Parkinson's disease (PD) with RBD, and the other six subjects (3 females, 3 males, age: 59.5±12.3 years) were normal controls. Due to the current stage of the research project no iRBD subjects nor PD subjects without RBD were included. Specialized neurologists conducted the diagnoses of all candidates. Subjects with snoring, apnea, bruxism, epilepsy, or other abnormalities that could have affected the sleep recording, were excluded from this study. Furthermore, none of the normal subjects showed any signs of abnormal sleep pattern, or was taking any medication, which was known to affect the sleep. The PD subjects could not be asked to withdraw their prescribed medicine before getting monitored. However, to reduce the affect of the medicine, highly medicated subjects, were excluded from this study.

B. Polysomnography Montage and Data Extraction

All subjects underwent at least one ambulant full night PSG monitoring. The bipolar EMG of the chin was analysed in this study. The impedance was kept less than 10kΩ, and the recorded digital signals were decimated to a joint sampling frequency of 200 Hz for computational reasons. Visual inspection of all the recordings was conducted by two sleep specialists, and corrupted recordings, where the analysis channels were flat, not connected or continuously contaminated by artifacts were rejected. Furthermore, all recordings were pruned such that only the sleep, i.e. from lights-off to lights-on, were analysed. The recordings were scored according to the new standard from the American Academy of Sleep Medicine (AASM), using a 30 second window without overlap [5]. However, due to the individual interpretation of the scoring definition, and the difficulties of scoring the PD group, two sleep specialists scored the PSG recordings independently, and a follow-up analysis of the scoring consensus between the two specialists was conducted. The raw sleep data and hypnograms were extracted from Somnologica Studio® (V5.1, Embla, Broomfield, CO 80021, USA) or Nervus® (V5.5, Cephalon DK, Norresundby, Denmark), using the build-in export data tool, and saved on a hard-disk in the European Data Format (EDF) [11]. The exported data was then converted into the used analysis software MATLAB® (R2008b, The MathWorks, Natick, MA., USA) for further processing.

C. Scoring Consensus

To verify the overall consensus between the two manual sleep stage scorings, two confusion matrices (TABLE I and TABLE II), have been created. The wake, REM and Non-REM stages are labeled as W, R, N1, N2 and N3 respectively, while the sleep stage occurrence is labeled as occ. The overall agreement of the control and PD group is approximately 83% and 81% respectively. It is clear that the scoring task of the PD group is not easy, compared to the control group. Especially the N1 sleep stage seems to be a difficult stage to score in both groups. Furthermore, it seems like the PD group is more awake and has less REM compared to the normal group, which may affect the overall performance of the algorithm. This will be addressed later.

TABLE I

CONFUSION MATRIX OF THE CONTROL GROUP IN PERCENTAGE

| | | Scoring 1 | | | | | |
		W	R	N1	N2	N3	Occ
Scoring 2	W	**89.7**	0.2	5.7	4.1	0.2	15.1
	R	0.8	**90.7**	3.6	4.5	0.5	17.9
	N1	14.2	16.5	**54.4**	14.9	0.0	9.5
	N2	0.2	0.4	2.8	**88.5**	8.1	35.1
	N3	0.3	0.0	0.4	9.7	**89.6**	22.3
	Occ	15.2	18.0	7.8	36.0	23.0	

TABLE II

CONFUSION MATRIX OF THE PARKINSON GROUP IN PERCENTAGE

| | | Scoring 1 | | | | | |
		W	R	N1	N2	N3	Occ
Scoring 2	W	**85.3**	2.4	1.0	10.4	0.8	29.4
	R	0.7	**85.9**	2.4	9.9	1.0	9.8
	N1	18.2	2.4	**65.2**	14.0	0.3	5.8
	N2	1.8	0.9	1.8	**91.9**	3.6	25.2
	N3	0.3	0.2	0.3	23.7	**75.5**	29.5
	Occ	27.0	9.6	4.8	35.0	23.5	

To increase the overall performance of the algorithm, only the epochs, i.e. a 30 second window without overlap, were the two specialists agreed, were analysed. The total amount of recorded sleep data was 77.1 hours (Control: 41.5 hours, PD: 35.6 hours).

D. Biomedical Data Processing

The surface detected signals are often affected by noise from slow movement and instability of the electrode-skin interface which typically can be seen in the frequency range of 0-20 Hz [12]. Furthermore, many signals are also affected by power-line noise (i.e. 50 Hz in Europe, Denmark). To reduce these interferences, all the signals were pre-processed using a band-pass filter and a notch filter. The applied bandpass filter was an 8th. order Butterworth filter with a 3dB cut off frequency of 20 Hz and 70 Hz respectively. The last filter was a Butterworth notch filter of 2nd order with a 3dB cut off frequency at 48 Hz and 52 Hz. The filtered REM data was then divided into three second segments (600 samples per segment) without overlap. The pre-processing process was followed by a feature extraction process, were each segment were analysed using four RBD characterizing features:

1) Kurtosis: The kurtosis feature of the nth segment is estimated by following equation, where N is the total amount of samples per segment, M is the number of REM segments, μ corresponds to the mean and finally, the $x_{i,n}$ is the ith EMG amplitude in the nth REM segment:

$$K(n) = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_{i,n} - \mu_n)^4}{(\frac{1}{N}\sum_{i=1}^{N}(x_{i,n} - \mu_n)^2)^2} - 3 \quad where \ n = 1,2..,M \quad (1)$$

The kurtosis measures the "peakedness" or "sharpness" of the distribution. For an ideal normal distribution the kurtosis is equal to three, and subtracting three would remove the

offset and make the kurtosis of the normal distribution equal to zero. It is expected that during movement, the EMG activity would be less normal distributed.

2) Percentile Ratio: The percentile P_q is the value below which P_q percent of the observation may fall. The percentile ratio is given by (2), where $P_{75}(|X_n(REM)|)$ corresponds to the 75th percentile of the *absolute* amplitudes of the *n*th REM segment, while $P_{75}(|X(nonREM)|)$ corresponds to the 75th percentile of *all* the absolute non-REM amplitudes:

$$P_r(n) = \frac{P_{75}(|X_n(REM)|)}{P_{75}(|X(nonREM)|)} \quad where \ n = 1,2..,M \quad (2)$$

In (2) the *M* corresponds to the total number of REM segments for the specific subject. Notice, that the non-REM is only used as a subject specific normalization factor (scalar). As described above, the overall EMG activity in non-REM is typically higher than the EMG activity in REM. If this is the case, then the ratio is expected to be below one, while for abnormal activity it would be one or above. In this study the percentile was set to 75. Different percentile values have been tested (result not shown).

3) Log-Power: The power is a widely used method for estimation of activity in signals. However, high variation of the power in the PD group has been observed. A way to map down high values could be by taking the natural logarithm of the power:

$$E(n) = ln\left(\frac{1}{N}\sum_{i=1}^{N} x_{i,n}^2\right) \quad where \ n = 1,2..,M \qquad (3)$$

As above, the $x_{i,n}$ is the *i*th EMG amplitude in the *n*th REM segment, while N is the total amount of samples per segment and M is the number of REM segments. It is expected that the overall power in REM would be higher in the PD group compared to the controls.

4) Median Frequency: The median frequency is defined as that frequency, that divides the power spectrum density (PSD) in two equally sized parts and is given by:

$$\sum_{f=0}^{f_\xi(n)} PSD_n(f) = 0.5 \sum_{f=0}^{f_s/2} PSD_n(f) \quad where \ n = 1,2..,M \quad (4)$$

In equation (4), the $PSD_k(f)$ is the amplitude EMG PSD of the *n*th segment. The median frequency f_ξ provides some general information regarding the spectrum of the EMG signal versus time, and it is also less affected by additive noise compared to the mean frequency. It is expected that the PSD of the EMG signals would change during movements compared to no, or sustained, activity.

The four features described above were merged into a so-called [Mx4] feature vector, were M is the number of segments and 4 corresponds to each of the four features described above. The feature extraction process was then followed by a feature reduction process, where the reduction was achieved by computing a feature subset using Principle Component Analysis (PCA) [13]. The amount of variance accounted for by the first and second components was approximately 92%, which was assumed to be reasonable. Furthermore, by using the two first components it would be possible to visualize the feature subset. The two classes were discriminated by applying the subset feature vector to a statistical classifier. A commonly used method for classification is the so called Bayes classifier, also known as Quadratic Discriminant Analysis (QDA). By assuming the data to be Gaussian distributed, the Bayes classifier may be a reasonable choice. Notice, this may not always be the case. Other methods are also feasible, but to minimize the number of needed computations and avoid over-fitting the data, a relatively simple method was chosen. The quadratic discriminant function is given by:

$$d_j(\mathbf{x}) = \ln P(C_j) - \frac{1}{2}\ln|\mathbf{C}_j| - \frac{1}{2}[(\mathbf{x} - m_j)^T \mathbf{C}_j^{-1}(\mathbf{x} - m_j)] \quad (5)$$

Where j=1,2 (i.e. the two classes: controls and PD). The classification rule is given by:

$$Class(\mathbf{x}) = \arg\max_j d_j(\mathbf{x}) \qquad (6)$$

In (5) the $P(C_j)$ denotes the prior probability of class *j*, and is estimated by dividing the number of class *j* segments with the total amount of segments in the training set. Furthermore, the term $\mathbf{x}$ is the [Mx2] subset feature vector, where M is the number of segments and 2 corresponds to the first and second principle component. Finally, the $\mathbf{m}_j$ vector and $\mathbf{C}_j$ matrix are the mean and covariance for class j, whereas $|\mathbf{C}_j|$ denotes the determinant of $\mathbf{C}_j$ [14]. To validate the overall performance of the algorithm, the leave-one-out validation approach has been used. This is a widely used method in the field of machine learning. Each of the twelve subjects were put aside as a test set, while the rest, i.e. eleven subjects, were used for training the algorithm. That is twelve training and testing runs in total. This will determine how accurate the algorithm is, when it is presented by unseen data. To compare this study with the previous published method mentioned above, the overall RBD activity in percentage was computed. That is, the total amount of classified RBD activity of the total REM sleep in percentage per subject was calculated, which corresponds to twelve percentage values. The two methods were compared using a so-called receiver operating characteristic curve, or in short, the ROC curve. It provides the correct classification rate for each category by showing the relationship between *sensitivity* and *1-specificity*.

III. RESULTS AND DISCUSSION

One of the training sessions is shown in Fig. 1, where the blue and red cluster corresponds to the control and PD group respectively, while the black hyperbolic line is the decision boundary discriminating between the two clusters. Each coloured dot represents the first and second principle component of the feature subset. Based on Fig. 1, the two groups are separable, but *few* samples (i.e. red and blue dots) are misclassified. The algorithm seems to misclassify when

the overall sleep stage distribution is extremely skewed. That is, when the presence of REM and (or) Non-REM is almost non-existent, while the wake stage is the most dominating stage. However, the majority of the feature subset is correctly classified. To measure the overall performance of the proposed algorithm, it was benchmarked against a previous published algorithm called STREAM, which was described above. Fig. 2 shows the ROC curve of each of the two methods, where the red and blue curve corresponds to the new proposed algorithm and STREAM respectively. Using the available data and optimal settings, the new proposed algorithm yields a 100% sensitivity, 100% specificity and an AUC equal to 1.0, while the STREAM algorithm yields a 100% sensitivity, 33% specificity and an AUC equal to 0.83.
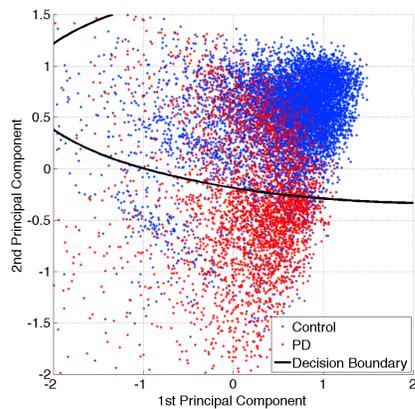

Fig. 1. Scatter plot of the feature subset, using one of the leave-one-out training seasons. The blue and red clusters corresponds to the control and PD group respectively, while the black hyperbolic line is the decision boundary.
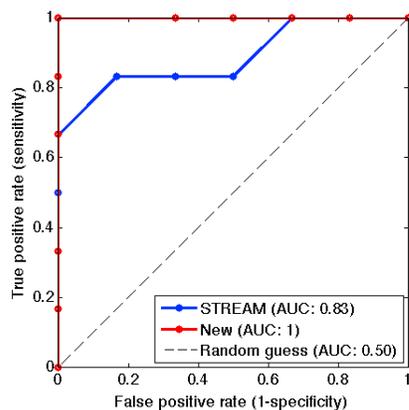

Fig. 2. ROC curve of the new alogrithm and the previous published algorithm called STREAM. The blue and red curves corresponds to the STREAM algorithm and new algorithm respectively.

IV. CONCLUSION

In this study a *semi*-automatic method for RBD detection has been conducted, involving twelve subjects and 77.1 hours

of sleep data. Based on the *available* data and using optimal settings, it was possible to correctly classify PD subjects with RBD with 100% sensitivity, 100% specificity, using the leave-one-out approach. This is an improvement compared with previous published studies. The proposed method seems to misclassifying few feature subset samples when the duration of sleep becomes extremely short. The overall result indicates the usefulness of a computerized scoring algorithm, and may be a feasible way of reducing scoring time, and increasing the success-rate of early detection of RBD. Further enhancement on *additional* data, i.e. subjects with iRBD and PD without RBD, is needed to validate its robustness and the overall result. Future work includes development of a full automatic RBD detection algorithm, by using most of the PSG modalities, and an investigation of how well the performance correlates with the semi-automatic algorithm.

REFERENCES

[1] R. Postuma, J. Gagnon, M. Vendette, M. Fantini, J. Massicotte-Marquez, and J. Montplaisir, "Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behavior disorder," *Neurology*, vol. 72, no. 15, pp. 1296–1300, 2009.

[2] A. Iranzo, J. Molinuevo, J. Santamara, M. Serradell, M. Mart, F. Valldeoriola, and E. Tolosa, "Rapid-eye-movement sleep behaviour disorder as an early marker for a neurodegenerative disorder: a descriptive study," *The Lancet Neurology*, vol. 5, no. 7, pp. 572–577, 2006.

[3] R. Postuma, J. Gagnon, S. Rompr, and J. Montplaisir, "Severity of REM atonia loss in idiopathic REM sleep behavior disorder predicts Parkinson disease," *Neurology*, vol. 74, pp. 239–244, 2010.

[4] AASM, *The International Classification of sleep disorders, revised: Diagnostic and coding manual.* Chicago, Illinois: American Academy of Sleep Medicine, 2001.

[5] AASM, *A Technologist's Handbook for Understanding and Implementing the AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specifications.* Westchester, Illinois, 2009.

[6] J. W. Burns, F. B. Consens, R. J. Little, K. J. Angell, S. Gilman, and R. D. Chervin, "EMG Variance During Polysomnography As An Assessment For REM Sleep Behavior Disorder," *SLEEP*, vol. 30, no. 12, pp. 1771–1778, 2007.

[7] J. Fairley, G. Georgoulas, and G. Vachtsevanos, "Sequential Feature Selection Methods for Parkinsonian Human Sleep Analysis," *17th Mediterranean Conference on Control Automation*, pp. 1468–1473, 2009.

[8] R. Ferri, M. Manconi, G. Plazzi, O. Bruni, S. Vandi, P. Montagna, L. Ferini-Strambi, and M. Zucconi, "A quantitative statistical analysis of the submentalis muscle EMG amplitude during sleep in normal controls and patinets with REM sleep behavior disroder," *Journal of Sleep Research*, vol. 17, pp. 89–100, 2008.

[9] G. Mayer, K. Kesper, T. Ploch, S. Canisius, T. Penzel, W. Oertel, and S.-K. K., "Quantification of Tonic and Phasic Muscle Activity in REM Sleep Behavior Disorder," *Journal of Clinical Neurophysiology*, vol. 25, no. 1, pp. 48–55, 2008.

[10] M. Shokrollahi, S. Krishnan, D. Jewell, and B. Murray, "Analysis of the Electromyogram of Rapid Eye Movement Sleep ," *31st Annual International Conference of the IEEE EMBS*, pp. 2659–2662, 2009.

[11] B. Kemp, A. Varrib, A. C. Rosac, K. D. Nielsend, and J. Gaded, "A simple format for exchange of digitized polygraphic recordings," *Electroencephalography and Clinical Neurophysiology*, vol. 82, no. 5, pp. 391–393, 1992.

[12] R. Merletti and P. Parker, *Electromyography: Physiology, Engineering and Noninvasive Applications.* Wiley Interscience, 2004.

[13] J. Jackson, *A User's Guide to Principal Components.* Wiley Blackwell, 2003.

[14] R. M. Rangayyan, *Biomedical Signal Analysis A Case-study Approach.* Wiley Interscience, 2001.