33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Detection of arousals in Parkinson's disease patients

Gertrud Laura Sorensen[a,b,c], Jacob Kempfner[a,b,c], Poul Jennum[b,c] and Helge B. D. Sorensen[a]
[a]*Department of Electrical Engineering, Technical University of Denmark, Kongens Lyngby, Denmark*
[b]*Danish Center for Sleep Medicine, Glostrup University Hospital, Glostrup, Denmark*
[c]*Center for Healthy Ageing, University of Copenhagen, Copenhagen K, Denmark*

***Abstract*— Arousal from sleep are short awakenings, which can be identified in the EEG as an abrupt change in frequency. Arousals can occur in all sleep stages and the number and frequency increase with age. Frequent arousals during sleep results in sleep fragmentation and is associated with daytime sleepiness. Manual scoring of arousals is time-consuming and the inter-score agreement is highly varying especially for patients with sleep related disorders. The aim of this study was to design an arousal detection algorithm capable of detecting arousals from sleep, in both non-REM and REM sleep in patients suffering from Parkinson's disease (PD). The proposed algorithm uses features from EEG, EMG and the manual sleep stage scoring as input to a feed-forward artificial neural network (ANN). The performance of the algorithm has been assessed using polysomnographic (PSG) recordings from a total of 8 patients diagnosed with PD. The performance of the algorithm was validated using the leave-one-out method resulting in a sensitivity of 89.8 % and a positive predictive value (PPV) of 88.8 %. This result is high compared to previous presented arousal detection algorithms.**

I. INTRODUCTION

Manual scoring of arousals in polysomnographic (PSG) recordings is a time-consuming and tedious task. Furthermore the inter-score agreement between two or more sleep specialists is highly varying, especially for patients with sleep disorders [1], [2], [3]. Therefore several methods have been proposed for automatic detection of arousals, in order to make the arousal scoring time-saving and standardized.

According to the American Academy of Sleep Medicine (AASM), arousals from sleep are defined as an abrupt shift of electroencephalography (EEG) frequency that lasts at least 3 seconds [4], [5]. Furthermore if the arousals are from rapid-eye movement (REM) sleep, at least 1 second of an increase in chin electromyography (EMG) must be present. The scoring of arousals is thus independent of the 30-seconds epochs considered in conventional sleep scoring, and therefore the EEG and EMG signals have to be analysed continuously for significant changes. The first adaptive segmentation approach was suggested by [2] who detected changes in EEG frequency and transient increase in submental EMG to define arousals. From 11 PSG recordings, the algorithm achieved a sensitivity of 88.1 % and a positive predictive value (PPV) of 74.5 %.

A few other adaptive segmentation approaches have been proposed [6], [7], [8], [9], [10] however none of these have achieved a better sensitivity than [2] but an improved PPV has been reported in a few studies.

The aim of this study was to design an arousal detection algorithm capable of detecting arousals in both non-REM and REM sleep for PD patients. The definition of arousals is not dependent on a certain frequency content but simply a frequency shift, and the detection of arousals should therefore still be possible independent of specific frequency and power changes in the EEG which has been detected in EEG from PD patients.

II. METHOD

A. Subjects

Eight PD patients (age 59.8 ± 8.8) were selected from the patient database at the Danish Center for Sleep Medicine, Department of Clinical Neurophysiology, Glostrup University Hospital, Denmark. Patients taking any anti-depressant drug, sleep medicine or other medication known to affect sleep were excluded from the study. Furthermore, subjects with sleep apnea, bruxism, epilepsy, or other abnormalities known to effect sleep recordings, were also excluded from this study.

B. Data acquisition

All subjects underwent one night of PSG including six leads of EEG (F3-A2, F4-A1, C3-A2, C4-A1, O1-A2, O2-A1), surface EMG of the left and right anterior tibialis muscle and the submental muscle, electrocardiography (ECG) and vertical and horizontal electrooculography (EOG). Impedances were kept below 10 kΩ, and all signals were lowpass filtered and subsequently decimated to a joined sampling frequency of 256 Hz. Subjects were instructed not to consume any caffeinated or alcoholic drinks for at least 6 hours prior to the recordings. Sleep stages and arousals from sleep were manually scored according to standard criteria [4], [5].

C. Biomedical signal processing

The AASM manual for scoring of sleep and associated events states that either the central or occipital EEG channels should be used when scoring arousals [4]. Furthermore the submental EMG channel should be used to score arousals during REM sleep. Therefore, it was decided to use only the two central (C3-A2, C4-A1) and two occipital (O1-A2, O2-A1) EEG channels and the submental EMG channel.

All EEG signals were bandpass filtered with a 3 dB cut off frequency between 0.3 and 35 Hz and the EMG signal was bandpass filtered with a 3 dB cut off frequency between

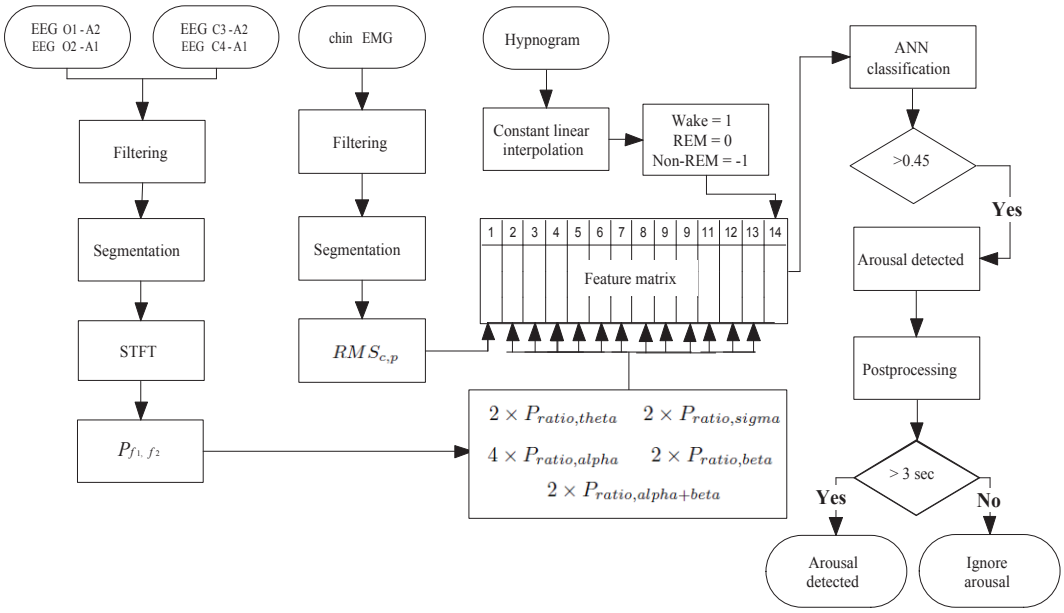

Fig. 1. Illustration of the biomedical signal processing performed by the arousal detection algorithm. Features are extracted from the EEG signals (12 features), the chin EMG signal (1 feature) and the manually scored hypnogram (1 feature) and given as input to an ANN. Output from the ANN is interpreted as probabilities, so that all outputs above 0.45 are classified as an arousal being present. A postprocessing step combines arousals detected within 10 seconds and removes arousals with a duration less than 3 seconds.

30 and 100 Hz. Both bandpass filters were implemented as linear-phase equiripple FIR filters using the Parks-McClellan algorithm [11]. Also a 50-Hz notch filter were applied for all signals, which attenuated the 50 Hz component with more than 100 dB.

1) Feature extraction: The frequency content of the EEG signals was assessed by calculating the short time fourier transform (STFT) of the EEG signals. By using STFT, the Fast Fourier Transform (FFT) of small signal segments was calculated to get the frequency spectrum for a specific time of the recording. Before the FFT was calculated, each segment was multiplied by a Hanning window to avoid large sidelobes in the frequency spectrum.

The EEG signals were segmented into a past and current time window defined simultaneously throughout the EEG signals as illustrated in Fig. 2. The current window was the one which should detect if an arousal was present, and since the frequency shift should last at least 3 seconds, the length of this window was set to 3 seconds. The length of the past window was determined through trial and error to 10 seconds. The windows were moved 1 second at a time along the EEG signals.

The two temporal windows represented two consecutive segments of the EEG signals, which were transformed by FFT to the frequency domain, and the power spectrum of each segment was calculated by squaring the magnitude of the spectra. The mean band power was measured in different frequency bands as:

$$P_{f_1,f_2} = \frac{1}{N}\sum_{n=0}^{N-1} |X(f_1 + n\Delta f)|^2, \qquad (1)$$

where X is the filtered EEG signal segment transformed to the frequency domain by FFT, Δf is the frequency resolution and N is the number of samples in the fourier transformed segment. The frequencies f_1 and f_2 denote the first and last frequency in the frequency band, thus:

$$f_1 + (N-1)\Delta f = f_2. \qquad (2)$$

The term $1/N$ in (1) makes the mean band power of the two segments of different time lengths comparable.

The mean band power in the current segment, $P^c_{f_1,f_2}$, was divided with the mean band power of the past segment, $P^p_{f_1,f_2}$, to yield a measure of the shift in frequency content. For each frequency band, a new signal was then obtained, describing

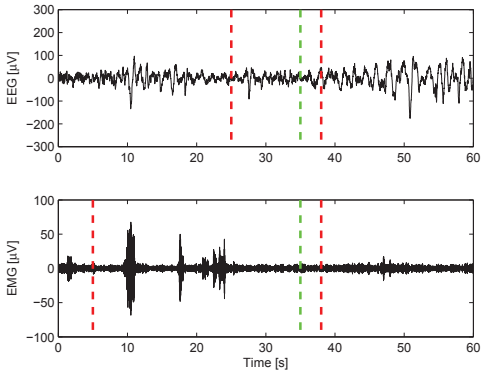

Fig. 2. EEG signals are segmented in a current window of 3 seconds and a past window of 10 seconds. The EMG signal is segmented in a current window of 3 seconds and a past window of 30 seconds. The current and past windows are demarcated by the red dotted lines and separated by the green dotted line.

the change in the frequency content of each band:

$$P_{ratio,band} = \frac{P^c_{f_1,f_2}}{P^p_{f_1,f_2}}, \quad (3)$$

where *band* refers to the frequency band $f_1 - f_2$. For each frequency band, $P_{ratio,band}$ describes a feature vector for each EEG signal. Through a feature selection process, a combination of 10 features from the central EEG signal and two features from the occipital EEG signal were chosen. For the two central EEG signals, the features were the power ratio given in (3) derived from the 4-7 Hz, 8-12 Hz 12-14 Hz, 13-30 Hz and 8-30 Hz band. For the occipital EEG signals, the power ratio features derived from the 8-12 Hz band were used.

To make sure that outliers did not influence the classification process, all EEG feature vectors with values above a certain threshold were truncated. Through trial and error, this threshold value was chosen to 10, thus:

$$P_{ratio,band}(l) = \begin{cases} P_{ratio,band}(l) & \text{if } P_{ratio,band}(l) < 10 \\ 10 & \text{if } P_{ratio,band}(l) > 10 \end{cases},$$
$$l = 1,2,...,L, \quad (4)$$

where L is the length of the features, which corresponds to the length of the PSG recording in seconds. Afterwards all features were normalized with their maximum value to get the same dynamic range for all patients.

The objective of the EMG feature was to detect amplitude changes in the EMG signal. Therefore a comparison between the amplitude of a past window of 30 seconds and a current window of 3 seconds was performed, as illustrated in Fig. 2. Since muscle activity sometimes starts several seconds before the onset of arousals, it was necessary to consider the EMG signal several seconds prior to any possible arousal in order to detect a change in amplitude.

To estimate the EMG amplitude, the root-mean-square (RMS) indicator was calculated for each window as:

$$RMS(x) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x^2(i)}, \quad (5)$$

where x is the filtered EMG signal segment in the time domain and N the number of samples in this segment. The EMG signal was considered second by second as with the EEG signals, measuring the RMS for both a signal segment of 3 seconds and a signal segment of 30 seconds. Thus in (5), $N = f_s \cdot \Delta t$, where Δt was either 3 seconds or 30 seconds. The RMS value in the current segment, RMS^c, was divided with the RMS-value in the past segment, RMS^p, to yield a measure of the amplitude change of the EMG signal:

$$RMS_{c,p} = \frac{RMS^c}{RMS^p}. \quad (6)$$

The dynamic range of the EMG feature in (6) varied from subject to subject. Therefore it was not possible to determine a fixed threshold independent of the subjects. Instead a threshold value was based on a percentile of the EMG feature. Through trial and error, it was decided to truncate all values above 1.5 times the 99$^{\text{th}}$ percentile:

$$RMS_{c,p}(l) = \begin{cases} RMS_{c,p}(l) & \text{if } RMS_{c,p}(x) < 1.5 \cdot p_{99} \\ 1.5 \cdot p_{99} & \text{if } RMS_{c,p}(x) > 1.5 \cdot p_{99} \end{cases},$$
$$l = 1,2,...,L, \quad (7)$$

where p_{99} is the 99$^{\text{th}}$ percentile of the EMG feature.

Since the definition of arousals is different in non-REM and REM sleep stages, another feature had to distinguish between wake, non-REM and REM. Both the EEG and EMG signals were processed second by second, an therefore a piecewise constant interpolation of the hypnogram was performed to expand from the 30 seconds epochs to sleep stages described second by second. This feature vector consisted of 1 for wake stages, 0 for REM stages and -1 for non-REM stages and represented the last feature for the arousal detection algorithm.

To summarize 14 features were used in total to describe the frequency shift in EEG, the muscle activity in the chin EMG and the sleep stages. These feature vectors were placed as columns in an $[L \times 14]$ matrix, where L is the length of the PSG recording in seconds. This is illustrated in Fig. 1.

2) Classification: All features were used as inputs to an artificial neural network (ANN). This type of classifier was chosen, since it has previously shown to perform well in arousal classification [7]. The ANN from the DTU toolbox[1] was chosen, which is a two layer feed-forward ANN with a hyperbolic tangent function for the hidden layer and a logistic sigmoidal function for the output layer.

The hidden layer was comprised of 9 neurons, since this amount of neurons resulted in the best classification error. The output was interpreted as probabilities and therefore the output layer consisted of a single neuron. The probability threshold was 0.45, so that every output above this probability was interpreted as an arousal being present. A postprocessing step was added to combine arousals classified in a certain proximity of each other. By trial and error a limit of 10 seconds was chosen, so that if the output vector had arousals closer than 10 seconds from each other, the arousals were combined to one arousal. Detected arousals lasting less than 3 seconds were removed.

III. Results

In order to measure the performance of the algorithm, the following variables were defined:

True positives (TP):	Number of correct detected arousals by the algorithm.
False positives (FP):	Number of incorrect detected arousals by the algorithm.
False negatives (FN):	Number of arousals missed by the algorithm.

A detected arousal was classified as correct detected, if it overlapped with the manually scored arousals.

[1] *http://cogsys.imm.dtu.dk/toolbox/*

As the output of the classifier was treated as probabilities, a threshold was chosen to satisfy both the sensitivity and the PPV. The threshold resulting in the maximum F-measure was chosen. The F-measure combines the sensitivity and PPV and is defined as:

$$F = 2 \cdot \frac{\text{sensitivity} \cdot \text{PPV}}{\text{sensitivity} + \text{PPV}} \tag{8}$$

The threshold resulting in the maximum F-measure was 0.45.

The algorithm was tested on the 8 PD patients using the leave-one-out method. In table I the sensitivity and PPV are shown for each PD patient, and the average sensitivity and PPV are 89.8 % and 88.8 % respectively.

PD patient	Sensitivity	PPV [%]
1	82.9	87.7
2	86.3	92.6
3	93.9	93.9
4	87.5	87.5
5	91.7	97.5
6	90.2	74.8
7	89.7	86.7
8	96.5	87.4

TABLE I

PERFORMANCE OF THE ALGORITHM WHEN TESTING ON THE 8 PD PATIENTS USING THE LEAVE-ONE-OUT METHOD.

IV. DISCUSSION

This paper describes a method for detection of arousals both in non-REM and REM sleep, only using the EEG, EMG and the manually scored hypnogram. In total 14 features are extracted from these signals and used in an ANN to classify arousals. The algorithm was tested on 8 PD patients resulting in a sensitivity of 89.8 % and a PPV of 88.8 %. This result is better than what has been presented for other arousal detection algorithms to date and it is the first arousal detection algorithm tested on PD patients.

The algorithm is not full-automatic due to the use of the manual scored hypnogram. Therefore a more correct term, would be to denote the algorithm as *semi-automatic*. None of the previous developed arousal detection algorithms detect the sleep stages automatically, however most of the algorithms are still presented as being automatic. To develop a full automatic arousal detection algorithm, it is necessary also to automatically distinguish between wake, non-REM and REM sleep. Several studies have already proposed methods for the automatic detection of sleep stages [12], [13], [14], but especially in patients with sleep disorders, this has seemed to be a difficult task. Furthermore, if a sleep stage algorithm manages to distinguish between non-REM, REM and wake, it will still be necessary to manually score the different non-REM stages to get the conventional hypnogram. Therefore as long as no algorithm manages to distinguish between all sleep stages, a manually scored hypnogram will probably always be available.

Several previous proposed arousal detection algorithms have not considered arousals during REM sleep appropriately. Either they lack to define arousals with the constraint on EMG activity in REM [10], they do not use the EMG-channel [6] or they only look at the non-REM sleep stages [8].

Compared to previous studies who have assessed the inter-score variability among manual arousal scorings [3], [15], the algorithm presented in this paper performs even better than what could be expected for an automatic arousal detection algorithm. Future work rely on validation using several data sets of both diseased and healthy subjects and subjects of various ages [16]. Furthermore data sets from other sleep laboratories and scored by other sleep specialists would also be advisable. This will both achieve a more reliable and accurate assessment of the performance of ones algorithm, and allow to compare different algorithms.

REFERENCES

[1] Bonnet M.H., Doghramji K., and Roehrs T., "The scoring of arousal in sleep: reliability, validity and alternatives," *Journal of Clinical Sleep Medicine*, vol. 3, pp. 133–145, 2007.

[2] Carli F.D., Nobili L., Gelcich P., and Ferrillo F., "A method for the automatic detection of arousals during sleep," *Journal of Sleep and Sleep Disorders Research*, vol. 22, no. 5, pp. 561 – 572, 1999.

[3] Drinnan M.J., Murray A., Griffiths C.J., and Gibson G.J., "Interobserver variability in recognizing arousal in respiratory sleep disorders," *American Journal of Respiratory and Critical Care Medicine*, vol. 158, no. 2, pp. 358–362, 1998.

[4] Iber C., Ancoli-Israel S., Chesson A.L., and Quan S.F., *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*, Wastchester, IL: American Academy of Sleep Medicine, 1 edition, 2007.

[5] Bonnet M.H., Carley D., Guilleminault C.G., Harper R., Haydes B., Hirshkowitz M., Ktonas P., Keenan S., Roehrs T., Smith J., Walsh J., Weber S., and Westbrook P., "EEG arousals: scoring rules and examples: a preliminary report from the sleep disorders atlas task force of the american sleep disorders association," *Sleep*, vol. 15, pp. 173–184, 1992.

[6] Agarwal R., "Automatic detection of micro-arousals," *In: 27th annual international conference of the IEEE engineering in medicine and biology society; IEEE-EMBS*, pp. 1158–1161, 2005.

[7] Alvarez-Estevez D. and Moret-Bonillo V., "Model comparison for the detection of EEG arousals in sleep apnea patients," in *Lecture Notes in Computer Science. 10th International Work-Conference on Artificial Neural Networks*, 2009, vol. 5517, pp. 997–1004.

[8] Cho S., Lee J., Park H., and Lee K., "Detection of arousals in patients with respiratory sleep disorders using a single channel EEG," *Conf Proc IEEE Eng Med Biol Soc*, vol. 3, pp. 2733–2735, 2005.

[9] Shmiel O., Shmiel T., Dagan Y., and Teicher M., "Data mining techniques for detection of sleep arousals," *Journal of Neuroscience Methods*, vol. 179, pp. 331–337, 2000.

[10] Sugi T., Kawana F., and Nakamura M., "Automatic EEG arousal detection for sleep apnea syndrome," *Biomedical Signal Processing and Control*, vol. 4, pp. 329–337, 2009.

[11] McClellan J. and Parks T., "A united approach to the design of optimum FIR linear-phase digital filters," *Circuit Theory, IEEE Transactions on*, vol. 20, no. 6, pp. 697 – 701, 1973.

[12] Jensen P.S., Sorensen H.B.D., Leonthin H.L., and Jennum P., "Automatic sleep scoring in normals and in individuals with neurodegenerative disorders according to new international sleep scoring criteria," *Journal of Clinical Neurophysiology*, vol. 27, pp. 296–302, 2010.

[13] Norman R.G., Pal I., Stewart C., Walsleben J.A., and Rapoport D.M., "Interobserver agreement among sleep scorers from different centers in a large dataset.," *Sleep*, vol. 23, no. 7, pp. 901–908, 2000.

[14] Park H.J., Oh J.S., Jeong D.U., and Park K.S., "Automated sleep stage scoring using hybrid rule- and case-based reasoning.," *Computers and Biomedical Research*, vol. 33, no. 5, pp. 330–349, 2000.

[15] Loredo J.S., Clausen J.L., Ancoli-Israel S., and Dimsdale J.E., "Night-to-night arousal variability and interscorer reliability of arousal measurements," *Sleep*, vol. 22, no. 7, pp. 916–920, 1999.

[16] Sorensen G.L., Kempfner J., Jennum P., and Sorensen H.B.D., "Arousal detection in healthy adults and patients with parkinson's disease," *Submitted for Journal of Clinical Neurophysiology*, 2011.