35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Classification of iRBD and Parkinson's patients using a general data-driven sleep staging model built on EEG

Henriette Koch[a], Julie A. E. Christensen[a,b,c], Rune Frandsen[c], Lars Arvastson[b], Soren R. Christensen[b], Helge B. D. Sorensen[a] and Poul Jennum[c]

Abstract— **Sleep analysis is an important diagnostic tool for sleep disorders. However, the current manual sleep scoring is time-consuming as it is a crude discretization in time and stages. This study changes Esbroeck and Westover's [1] latent sleep staging model into a global model. The proposed data-driven method trained a topic mixture model on 10 control subjects and was applied on 10 other control subjects, 10 iRBD patients and 10 Parkinson's patients. In that way 30 topic mixture diagrams were obtained from which features reflecting distinct sleep architectures between control subjects and patients were extracted. Two features calculated on basis of two latent sleep states classified subjects as "control" or "patient" by a simple clustering algorithm. The mean sleep staging accuracy compared to classical AASM scoring was 72.4 % for control subjects and a clustering of the derived features resulted in a sensitivity of 95 % and a specificity of 80 %. This study demonstrates that frequency analysis of sleep EEG can be used for data-driven global sleep classification and that topic features separates iRBD and Parkinson's patients from control subjects.**

I. INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative disorder, where a progressive loss of structure or function of neurons lead to severe morbidity and even death. There exist no cure for PD and the treatment is symptomatic. If a neuprotective agent becomes available it is essential to identify the patients as early as posible, ideally in the presymptomatic interval before the clinical onset. Research has shown that idiopathic rapid-eye-movement sleep behavior disorder (iRBD) is a potential preclinical marker for neurodegenerative diseases including PD [2]. Polysomnography (PSG) is an important diagnostic tool when diagnosing sleep disorders [3] and many studies treat sleep electroencephalography (EEG), electromyography (EMG) and electrooculography (EOG) in the search for biomarkers [4].

The American Academy of Sleep Medicine (AASM) has defined the present standard for manually sleep scoring using PSG [5]. The AASM sleep stages are divided into rapid eye movement (REM) and non-REM, which is further divided into three stages N1-N3 according to the level of drowsiness. Despite the common use of AASM, manually scoring is criticized for several weaknesses such as oversimplification of the sleep structures by discretizing sleep stages into defined stages. Another weakness is that the inter-rater reliability is low (reliabilities for healthy subjects lie between 67-91 % [3]). Focusing solely on brain activity during sleep requires a robust and general method for analysing EEG. Many proposed automatic sleep classifiers uses EEG frequency measures to obtain performances with reliabilities between 70-90 % (methods are summarized in [3]).

In [1] an automatic and unsupervised sleep staging model is presented using EEG spectral features from the standard frequency bands (delta: <4 Hz, theta: 4-7 Hz, alpha: 8-13 Hz and beta: 14-30 Hz). The model uses the Latent Dirichlet Allocation (LDA) method for sleep staging and each epoch is described as a mixture of topics, where each topic is evaluated as a latent sleep structure. All topics are defined in a data-driven way and are therefore not biased by former knowledge of sleep staging. The model presented in [1] had a mean performance of 70.1 % compared to manually scored Rechtschaffen and Kales hypnograms.

In [1] a new model for each subject was conducted, gaining high flexibility but compromising direct comparison of topics between subjects. This study uses control subjects to train an automatic and unsupervised LDA model and applies the trained general model to new subjects.

The hypothesis is that this global sleep staging model will make the sleep staging general and objective in contrast to manual sleep staging based on the AASM standard. It has been hypothesized that iRBD and PD patients express a changed sleep architecture compared to control subjects [6][7]. We hypothesize that the brain activity of control subjects will exhibit similar frequency characteristics, that they differ from the characteristics of patients and that this diversity will be captured by the general topic model.

In [8] a similar approach is carried out to classify iRBD and PD patients by using EOG features reflecting eye movements.

II. DATA

A. Subjects

Sleep EEG from 20 control subjects, 10 iRBD patients and 10 PD patients was enrolled in this study. The control subjects have no history of movement disorder, dream enacting behavior or sleep disorders. Recordings and sleep evaluations were done by the Danish Center for Sleep Medicine at

Corresponding author: HK: henriette_koch@hotmail.com
Contact information: HBDS: hbs@elektro.dtu.dk,
JAEC: jaec@elektro.dtu.dk, RF: RFRA0011@regionh.dk,
LA: larv@lundbeck.com, SRC: srac@lundbeck.com
and PJ: poje@glo.regionh.dk
[a]DTU Electrical Eng., Ørsteds Plads, bldg. 349, DK-2800 Kgs. Lyngby
[b]H. Lundbeck A/S, Ottiliavej 9, DK-2500 Valby
[c]Danish Center for Sleep Medicine, Department of Clinical Neurophysiology, Center for Healthy Aging, Faculty of Health Sciences, University of Copenhagen, Glostrup Hospital, DK-2600 Glostrup

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

TABLE I
DEMOGRAPHICS

Subject group	N^o subjects (male, female)	Age [years] ($\mu \pm \sigma$)
Control (train)	10 (5, 5)	57.2 $\pm$ 8.1
Control (test)	10 (5, 5)	59.8 $\pm$ 8.4
iRBD (test)	10 (8, 2)	59.0 $\pm$ 14.2
PD (test)	10 (6, 4)	63.2 $\pm$ 6.3

Glostrup University Hospital in Denmark. The demographics of the subjects are summarized in Table I.

B. Polysomnographic recordings (PSG)

For all subjects a full night PSG was recorded and evaluated in accordance with the AASM standard [5]. Proper measurement quality was ensured by visual inspection and recordings with presence of artefacts such as electrode disconnections were rejected. All analysis was carried out on the left central, frontal and occipitale EEG channel (C3-A2, F3-A2 and O1-A2) using a sampling frequency of 256 Hz. Three EEG channels were used instead of one as in [1] in order to detect a wider range of brain activities.

III. METHOD

A. Sleep staging model

Initially all data was bandpass filtered forward and reversed by a 4th order Butterworth filter with cut-off frequencies at 0.3 and 35 Hz (3 dB). After filtering, each EEG channel was split into non-overlapping one-second segments and the single sided amplitudes were derived by use of the fast fourier transform. The amplitudes were split into the clinical EEG frequency bands (delta: <4 Hz, theta: 4-7 Hz, alpha: 8-13 Hz and beta: 14-30 Hz) and summed within each band. Fig. 1 shows the method schematically.

The SAX approach [9] to time-series symbolization was used separately on each EEG channel on a per-subject basis to normalize the features and allow comparisons between subjects. The discretization was nessecary for creating "words" used in the topic model. In each frequency band SAX divided the full range of amplitude features into five equiprobable bins and every one-second segment was assigned a discrete number between 1 and 5 according to which quantile the amplitude feature belonged to. The sleep recording for each channel was thereby transformed into four strings (one for each frequency band) consisting of symbols 1-5 with each symbol describing the amplitude level in the corresponding one-second segment.

The symbol strings were divided into 30 seconds non-overlapping epochs to allow comparison with the AASM standard and a sliding window counted "words" defined as three contiguous symbols in each 30 second epoch. In summary each 30 second epoch was described by counts of all 1500 possible words when using five symbols, words containing three contiguous symbols in four frequency bands from three EEG channels.

LDA is a generative model normally applied to text documents to detect an underlying set of topic probabilities by counting words [10]. Input is a matrix containing word counts for each document and the output is mixtures of probabilities that the specific document contains the individual K topics. In this application each 30 second epoch should be consideres as a "document" and each subject's full night sleep recording (from lights off to lights on) as a "corpus". Like [1], this study used variational inference to optimize the model parameters and five topics to allow comparison with the AASM standard.

The topic model was trained on 10 control subjects due to a hypothesis that control subjects express the true sleep characteristics according to brain activity in the different sleep stages. The 10 control subjects corpora were merged into one large corpus and used for training a general topic mixture model, see Fig. 2. The parameters of the topic dirichlet and word dirichlet were saved and these parameters defined the general sleep staging model. Generalization of the model ensures equal "scoring procedure" between subjects and thereby direct comparison between topic mixture diagrams. The trained model was applied to each individual test subject and the posterior probabilities for each topic in each epoch were approximated by variational inference.

Support vector machines (SVM) were used for comparing the topic model with the manual scoring. SVM was extended into a multiclass classifier by training and testing using an one-versus-all approach. This was done on a per-patient basis where the epochs were randomized and distributed equally into 20 dataset ensuring all sleep stages were present in all dataset. Validation was carried out by using all 20 training/test set combinations; training on one set and testing on the other 19. For each subject the mean of the accuracy of the 20 multiclass SVM models was set as the model performance for the specific subject.

B. Features and classification

Specific features were defined in accordance with the hypothesis that the sleep architecture between control subjects

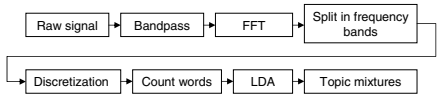

Fig. 1. Flow chart of data preparation and model application.

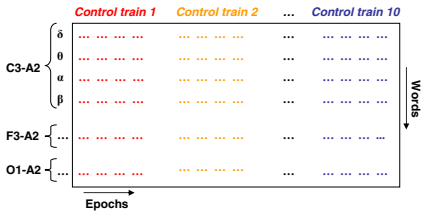

Fig. 2. Word counts for all training control subjects were merged into a single matrix used for training the topic mixture model. The parameters describing the topic and word dirichlet were applied when topic mixtures for test subjects were derived.

and iRBD/PD patients differs. The certainty, C, of different topics were used as features and extracted from the topic mixture diagrams from all test subjects (10 control subjects, 10 iRBD patients, 10 PD patients) as

$$C(m) = \frac{\sum_{n=1}^{N} logical\{p_n(m) > Q_m\}}{\sum_{n=1}^{N} p_n(m)} \quad (1)$$

where $n = 1,2,...,N$ defines the epoch number, $m = 1,2,...,5$ specifies the topic with corresponding $p_n(m)$ topic probability and Q_m defines the threshold for when an epoch is counted as certain for that specific topic.

Receiver operating characteristic (ROC) analysis evaluated by area under curve (AUC) values using the leave-one-subject-out approach was performed individually on each topic feature. This was done to derive the certainty threshold that best separated the controls and patients. To examine a potential cluster separation between control subjects and iRBD/PD patients a simple K-means algorithm with $K = 2$ was applied on the certainty feature from two topics (topic 2 and 4) using the leave-one-subject-out validation approach.

The third feature extracted was the number of transitions T normalized with the subject-specific total number of epochs. A transition was defined as a change in dominating topic between subsequent epochs (all topics included) expressed as

$$T = \sum_{n=1}^{N-1} logical\left\{\arg\max_m p_n(m) \neq \arg\max_m p_{n+1}(m)\right\} \quad (2)$$

This extra feature was included in the K-means algorithm creating a 3D distance separation.

IV. RESULTS AND DISCUSSION

A. Topic mixtures for sleep staging

This study shows a clear visual similarity between the obtained topics and the manually scored hypnograms. Fig. 3 shows the trained model applied to a control test subject, a moving average smoothed version using ten epochs and the corresponding hypnogram. Each epoch is represented as a coloured vertical bin where the amount of each colour is the individual topic probabilities $p_n(m)$.

There is a distinct concordance between topic 2 (light blue) and REM as well as topic 4 (orange) and deep sleep. Less pronounced but valid for the full sleep recording is the correlation between topic 1 (dark blue) and N1, topic 3 (green) and N2, topic 5 (red) and awake. Table II summarizes the visual concordances and the same concordance was present in all test subjects independent of test group. The temporal course of the topics are smooth and physiological this indicates continuous shifts between sleep stages in contrast to the abrupt shifts in manual scoring.

Fig. 4 shows the model applied to an iRBD patient. The iRBD patient expresses less topic 2 and 4 compared to the control subject. The AASM stage N3 is highly influenced by topic 3 indicating that the majority of deeper sleep might be dominated by N2 and rarely by N3. The REM periods

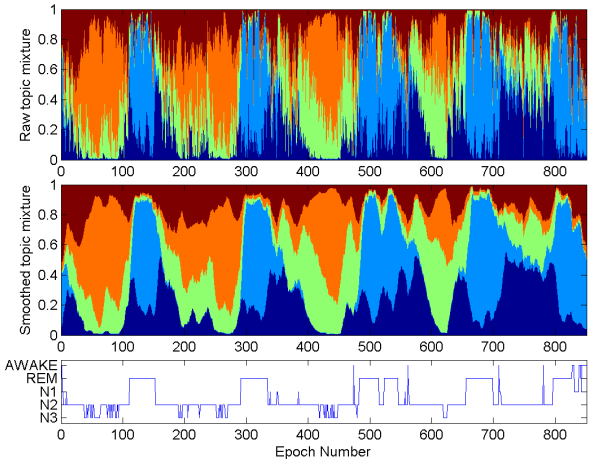

Fig. 3. Control test subject. Top: the trained model applied to a control test subject. Middle: moving average of the top plot using ten epochs. Bottom: corresponding manually scored hypnogram.

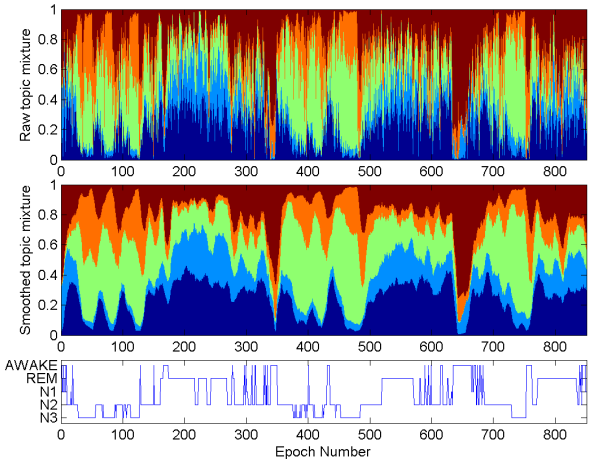

Fig. 4. iRBD test patient. Top: the trained model applied to an iRBD test patient. Middle: moving average of the top plot using ten epochs. Bottom: corresponding manually scored hypnogram. Compared to control subjects iRBD patients show less certainty of topic 2 and 4.

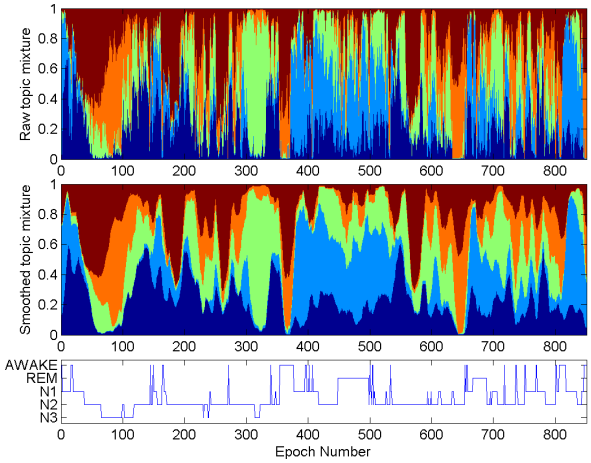

Fig. 5. PD test patient. Top: the trained model applied to a PD test patient. Middle: moving average of the top plot using ten epochs. Bottom: corresponding manually scored hypnogram. Compared to control subjects PD patients show shorter sleep stages, more abrupt transitions and less systematic sleep achitecture.

TABLE II

CONCORDANCE BETWEEN AASM SLEEP STAGE AND TOPIC

Topic	1	2	3	4	5
Colour	dark blue	light blue	green	orange	red
AASM	N1	REM	N2	N3	Awake

are shared between topic 1 and 2 and not only topic 2. No iRBD subjects' sleep contain topics with large certainty which can be caused by the fact that the model was trained on control subjects and therefore fits the control sleep EEG better. Alternatively, the difference is due to a physiological phenomenon indicating that iRBD patients lack the ability to lock in a single sleep topic or the sleep stages simply have different brain activity characteristics.

For PD patients the certainty of topic 2 and 4 is large in some short periods whereas other periods show low certainty as was the case for iRBD, seen in Fig. 5. In general, PD patients show shorter sleep stages, more abrupt transitions and less systematic sleep achitecture. This corresponds well with insomnia which is common for PD patients [7].

The accuracies of the sleep scoring compared to the manually scored hypnograms are derived by multiclass SVM on a per-patient basis. This is the best way of validating the automatic method but it is important to remember that the manual scored sleep stages are not necessarily the truth and the low inter-rater reliability in manual scoring will affect the accuracy of the SVM. This study's model shows accuracies of $(\mu \pm \sigma)$ 72.4±4.7 for control subjects, 65.2±7.0 for iRBD patients and 58.6±11.3 for PD patients. This is considered good performance for automatic sleep stager models [3]. The standard deviation of the performance between control subjects is low which supports that the model is general. Accuracies for iRBD and PD patients have lower means and larger standard deviations. This is most likely caused by greater uncertainty of manually scoring in diseased patients and/or lack of fit of this study's model on these groups.

B. Feature extraction and classification

Control subjects showed more percentage of higher certainty of topic 2 and 4. Mean AUC values were used to set the thresholds of 0.67 for topic 2 and 0.60 for topic 4. This means, that a certainty of topic 2 larger that 67 % occurs relatively frequently in all control subjects whereas it is rare for iRBD and PD patients. The same is valid for topic 4 with a certainty threshold at 60 %.

K-means using only these two features showed class separation, see Fig. 6 where all data is used to illustrate the clustering. The cluster analysis classified patients with a sensitivity of 95 % and a specificity of 80 % using the leave-one-subject-out validation scheme. Extending the K-means algorithm with the third feature (number of transitions) did not increase the class separation.

V. CONCLUSION

The topic mixture model derived in this study is a general model for sleep staging performing within the normal range of automatic sleep classifiers. The data-driven approach and mixture representation induced a more detailed discription of sleep compared to classical manual scoring. Using the general topic model, 30 topic mixture diagrams were obtained from test subjects. Feature extraction reflecting certainty of two sleep topics uncovered two clusters: control subjects and iRBD/PD patients and a simple K-means with $K = 2$ classified the subjects with a sensitivity of 95 % and a specificity of 80 %. This shows that the topic mixture modelling is able to automatically identify differences in sleep characteristics. However, the separation method is somewhat supervised and generalization of the result needs to be validated using an increased number of subjects.

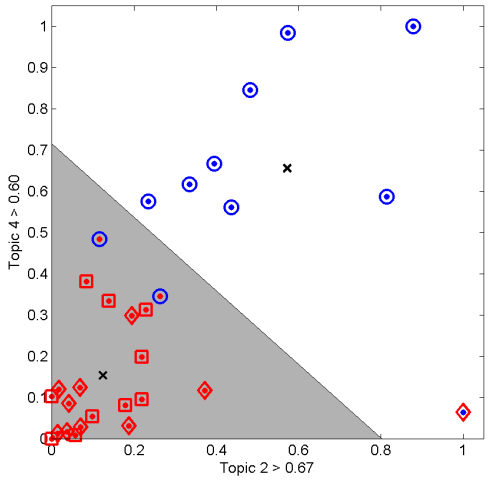

Fig. 6. For visual illustration a K-means algorithm is applied on the certainty features using all data and the decision boundary is drawn by using the position of the centroid centers. Leave-one-out validation showed significant clustering of iRBD and PD patients in one group separated from the group of control subjects. Circle: control subjects, diamond: iRBD patients, square: PD patients. Blue dot: class 1, red dot: class 2, cross: centroid center.

REFERENCES

[1] A. V. Esbroeck, B. Westover, "Data-Driven Modeling of Sleep States from EEG", *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5090-5093, 2012.

[2] R. B. Postuma, J. Montplaisir, "Predicting Parkinson's disease - why, when, and how?", *Parkinsonism and related Disorders*, vol. 15, pp. 105-109, 2009.

[3] B. Ahmed and R. Tafreshi, "Advances in Automatic Sleep Analysis", *IFMBE Proceedings*, vol. 23, pp. 422-426, 2009.

[4] R. B. Postuma et al., "Severity of REM atonia loss in idiopathic REM sleep behavior disorder predicts Parkinson disease", *Neurology*, vol. 74, 2010.

[5] C. Iber, *The AASM Manual for the Scoring of Sleep and Associated Events*. American Academy of Sleep Medicine, 2007.

[6] S. Wilson, D. Nutt, *Sleep Disorders*, Oxford University Press, 2008.

[7] A. H. V. Schapira, *Parkinson's Disease*, Oxford University Press, 2011.

[8] J. A. E. Christensen, H. Koch, J. Kempfner, L. Arvastson, S. R. Christensen, P. Jennum, H. B. D. Sorensen, "Classification of iRBD and Parkinson's patients based on eye movements during sleep", submitted to *2013 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2013.

[9] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A Symbolic Representation of Time Series, with Implications for Streaming Algorithms", *8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, pp. 2-11, 2003.

[10] D. M. Blei, A. Y. Ng, M. I. Jordan, "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, vol. 3, pp. 993-1022, 2003.