Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

SaEP1.6

Detecting Behavioral Microsleeps from EEG Power Spectra

Malik T. R. Peiris, Richard D. Jones*, Paul R. Davidson, and Philip J. Bones

Abstract— **EEG spectral power has been shown to correlate with level of arousal and alertness in humans. In this paper, we assess its usefulness in the detection of behavioral microsleeps (BMs). Eight non-sleep-deprived normal subjects performed two 1-hour sessions of a continuous tracking task while EEG and facial video were recorded. BMs were identified independent of tracking performance by a human rater by viewing the video recordings. Spectral power, normalized spectral power, and power ratios in the standard EEG bands were calculated using the Burg method on 16 bipolar derivations to form an EEG feature matrix. PCA was used to reduce the dimensionality of the feature matrix and linear discriminant analysis used to form a classifier for each subject. The 8 classifiers were combined using stacked generalization to create an overall detection model and N-fold cross-validation used to determine its performance (Φ = 0.30±0.05, mean±SE). While modest, the detection of BMs at such a high temporal resolution (1 s) has not been achieved previously other than by our group.**

I. Introduction

Microsleeps are brief episodes of sleep during which there is a marked reduction in behavioral responsiveness [1]. While they have been shown to occur frequently in tired subjects [2], there is often no advance warning of microsleep onset. Microsleeps have traditionally been defined as short bursts of sleep in the EEG/EOG [1]. While associated with drowsiness, EEG signatures of sleep in the transitional phases (sleep stages 1 and 2) are not well correlated with behavioral sleep [3]. Therefore, we identify microsleeps from a behavioral perspective and emphasize our usage is distinct from an EEG-defined microsleep. We refer to such events as *behavioral microsleeps* (BMs).

The occurrence of such BMs may be critical in occupational groups with a need to maintain a high level of alertness for extended periods, such as truck drivers, locomotive drivers, pilots, air traffic controllers, and surgeons, as well as the general population when performing a task such as driving [4]. BMs in these occupations can have disastrous consequences, including multiple fatalities.

Previous research has shown correlations between EEG power spectra and levels of arousal and alertness [5-7]. In this paper, we investigate the relationship between spectral power of the EEG and BMs using data from an earlier study [8], and design and test the performance of a spectral-power-based BM detector.

Manuscript received April 24, 2006. This research was financially supported by the Foundation for Research Science and Technology and the University of Canterbury, Christchurch, New Zealand.

All authors are members of the Van der Veer Institute for Parkinson's and Brain Research, Christchurch, New Zealand.

In addition, M. T. R. Peiris is with the Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand (malik.peiris@vanderveer.org.nz).

*R. D. Jones is with Medical Physics and Bioengineering, Christchurch Hospital, and Medicine, University of Otago, and Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand (phone: 643-378-6077; fax: 643-378-6080; e-mail: richard.jones@vanderveer.org.nz).

P. R. Davidson is with Medical Physics and Bioengineering, Christchurch Hospital, Christchurch, New Zealand. (e-mail: paul.davidson@vanderveer.org.nz).

P. J. Bones is with Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand (e-mail: P.Bones@elec.canterbury.ac.nz).

II. Methodology

A. Subjects

Data were recorded from 15 normal, healthy, non-sleep-deprived, male volunteers aged 18–36 years (mean = 26.5). Subjects attended two sessions held on different days and refrained from consumption of stimulants/depressants for 4 hours prior to each session. All sessions were held between 12.30 p.m. and 5.00 p.m.

The characteristics of lapsing were explored in an earlier analysis via full-head EEG, eye movements, and video from 15 subjects undertaking a 1-D continuous tracking task over two 1 hr sessions [8]. For the current research, we selected those subjects who had had at least one video-based BM for spectral analysis (N=8).

B. Tracking, Video, and EEG

The tracking task required subjects to use a steering wheel to track a pseudo-random target (bandwidth 0.164 Hz, period 128 s) which scrolled down the screen at 21.8 mm/s with an 8-s preview [9].

A video camera was used to record head and facial features during the session (25 Hz frame rate).

EEG was recorded from 16 scalp locations and digitized at 256 Hz (bandwidth 0.5–100 Hz). Bipolar derivations were Fp1–F7, F7–T3, T3–T5, T5-O1, Fp2–F8, F8–T4, T4–T6, T6–O2, Fp1–F3, F3–C3, C3–P3, P3–O1, Fp2–F4, F4–C4, C4–P4, and P4–O2.

C. Video Rating

Video recordings for all sessions were conservatively rated by one of the authors (MP), without knowledge of the corresponding tracking performance. Video was rated on a 6-level scale: 1 = alert, 2 = distracted, 3 = forced eye closure while alert, 4 = light drowsy, 5 = deep drowsy, and 6 = sleep (including microsleep). Samples with a video rating of 6 were defined as BMs and became the gold standard binary metric – *BM Index* (BMI) – for this study. Criteria similar to Weirwille *et al.* [10] were used to define our video rating scale. A BM observed on video typically contained prolonged eye lid closure and head nodding as the subject enters a BM, and a head jerk at the termination of the event.

Overall, subjects had 65.1±16.8 BM/h and a mean BM duration of 3.4±0.5 s.

D. EEG Spectral Features

The EEG was pre-processed using ICA to remove blink artifacts [11]. The blink-free signal from each derivation was then filtered to remove 50 Hz mains activity using an IIR notch filter with a Q-factor of 35. The signal was then transformed into z-scores relative to the first 2 min (baseline) of the signal.

Frequency spectra were calculated using a window size of 512 samples (2 s) and an overlap of 50% between successive windows giving a temporal resolution of 1 s. Windows containing samples with an absolute z-score >30 were rejected as artifacts and excluded from further analysis. Data in each 2 s window were detrended and the spectrum estimated using a 40^{th}-order Burg model. This parametric-model method was selected to estimate power spectra due to its high frequency resolution for short data records [12].

Spectral features were calculated (Table 1). For a given window, the average spectral power in each EEG band was calculated, then normalized by dividing it by the mean power over the entire spectrum. In addition, power ratios θ/β, θ/α, α/β, δ/θ, α/δ, β/δ, β_1/α, β_2/α, β_1/β_2 were calculated. This resulted in 13 features for the mean power in each band (including total power over the entire spectrum), 12 normalized power features, and 9 power ratios for each bipolar derivation; i.e., a total 544 spectral features (16 derivations x 34 features per derivation).

TABLE I
EEG BANDS AND FREQUENCY DIVISIONS

EEG Band	Frequency
Delta (δ)	1.0 – 4.5 Hz
Theta (θ)	4.5 – 8.0 Hz
Alpha 1 (α_1)	8.0 – 10.5 Hz
Alpha 2 (α_2)	10.5 – 12.5 Hz
Alpha (α)	8.0 – 12.5 Hz
Beta 1 (β_1)	12.5 – 15.0 Hz
Beta 1 (β_2)	15.0 – 25.0 Hz
Beta (β)	12.5 – 25.0 Hz
Gamma 1 (γ_1)	25.0 – 35.0 Hz
Gamma 2 (γ_2)	35.0 – 45.0 Hz
Gamma (γ)	25.0 – 45.0 Hz
High	> 45.0 Hz

E. Reducing Redundancy of Input Data

As spectral data from 16 EEG derivations were used, there was a high likelihood of redundant information. Hence, PCA was used to reduce the dimensionality of the feature space by generating a subset of principal components (PCs).

From the 544 features, 30 PCs were selected on the arbitrary basis that they explained > 90% of the mean variance in the data.

F. BM Classification Model

The next step was to form a model capable of detecting BMs from the EEG. The process involved forming a classification model using linear discriminant analysis (LDA) from the subset of 30 PCs extracted from the feature matrix as predictive variables and the BMI as grouping variable. LDA generates a linear discriminant function which can be used to classify future cases.

Fig. 1 shows a block diagram of the steps followed to create an LDA classification model from the feature matrix and BMI for a given subject: (i) the mean of each feature vector across the entire record was calculated and subtracted from vector (a prerequisite for PCA), (ii) PCA was undertaken on the mean-subtracted feature matrix to obtain PCs, (iii) a subset of the first 30 PCs and BMIs were taken from both sessions of a subject and concatenated to make a single feature matrix and BMI, (iv) each PC was converted to z-scores, and (v) the PCs and BMI were used to form a linear discriminant classification model based on the subject via MATLAB®'s discriminant analysis toolbox. The resultant model comprised linear discriminant function coefficients and a PCA-generated transformation matrix.

A classification model was generated for each of the 8 subjects.

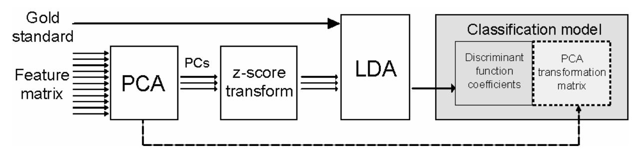

Fig. 1. A block diagram depicting the creation of a BM classification model using EEG spectral features.

G. Classifier Performance

In general, combining the output of several models increases predictive performance over a single combined model [13]. Consequently, stacked generalization was used to combine the predictions of the individual within-subject classification models. A block diagram of stacked generalization is depicted in Fig. 2. Stacking, incorporating a simple meta-learner algorithm, was used to determine how much each level-0 model should be trusted in making the final prediction based on the w_k weights of the level-1 model.

When the full stacked model was used for classification, each new case was fed into all level-0 models, producing 8 predictions, and these were fed into the level-1 model which scaled them according to w_k, summed them, and applied a threshold (=0.5) to provide a final binary output (BM or no-BM).

The weights of the level-1 model were determined via an algorithm which assessed the generalization performance of each level-0 model. Data from the 7 subjects not used to form the level-0 model were used as a "test set". Each level-0 model's classification accuracy was estimated by finding the mean Φ-correlation-squared between the classifier output and BMI across all test set subjects. The mean Φ^2 values for each level-0 model were then normalized to obtain a level-1 model weight w_k for each. Thus, the level-1 weights indicate the relative generalization performance of each within-subject model.

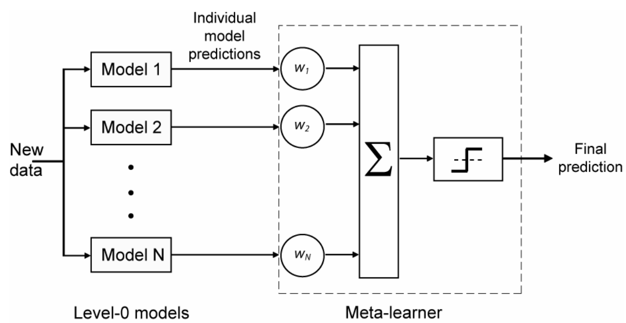

Fig. 2 Schematic diagram of stacked generalization

H. Validating Overall Detection Model

Performance was validated using standard n-fold cross-validation as follows

1. Reserve one of the 8 subjects as the validation subject and leave their data aside.
2. Create classification models (level-0 models) using data from the 7 remaining subjects.
3. Determine the generalization of the 7 level-0 models and obtain model weights (w_i) as outlined above ('Classifier Performance').
4. Feed the validation subject's data to all 7 level-0 models in the stacked generalization system and obtain the final prediction from the meta-learner output.
5. Calculate the correlation between the validation subject's BMI and the meta-learner output (Φ_v).
6. Repeat steps '1' to '5' and obtain Φ_vs for all 8 subjects.
7. The overall detector performance is the mean across all Φ_vs.

III. Results

A. Classifier Performance

Fig. 3 shows the generalization performance of the 8 classification models. The normalized model weights indicate how well each model generalized to the other subjects, with subject 814 providing the best between-subject classification model.

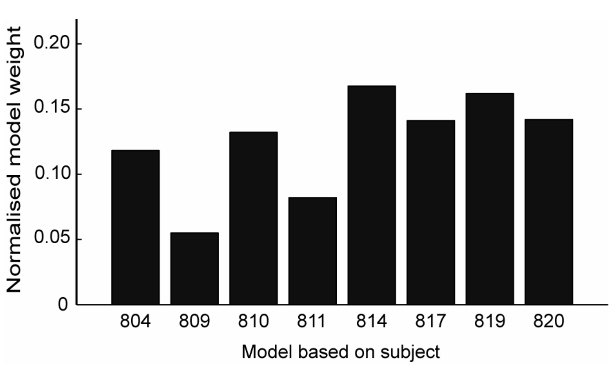

Fig. 3. Model weights calculated according to the generalization performance of each model obtained by classifying the data from the remaining subjects.

B. Overall Detector Performance

Table 2 provides a summary of the overall detector performance for various full and restricted feature combinations used to create classification models. Power ratios on their own yielded the highest detector performance (Φ = 0.34±0.06)

TABLE II
SPECTRAL FEATURES USED TO FORM THE MODEL AND THE CORRESPONDING DETECTOR PERFORMANCE

EEG spectral feature	Detector performance (mean Φ ±SE)
Spectral power	0.30±0.07
Normalized spectral power	0.24±0.04
Power ratios	0.34±0.06
Spectral power + normalized spectral power	0.28±0.04
Spectral power + power ratios	0.33±0.06
Normalized spectral power + power ratios	0.29±0.05
Spectral power + normalized spectral power + power ratios	0.30±0.05

IV. Discussion

Our cross-validation analysis showed that the performance of an EEG spectral power based BM detector was, on average, modest. This notwithstanding, detection of BMs at this level of temporal resolution (1 s) has not been achieved previously other than by our group [14, 15].

We have presented a thorough linear analysis which will provide a useful performance baseline for future EEG analysis of the BM phenomenon. In future work we intend to investigate the performance of our stacked between-subjects model with more complex, possibly non-linear, level-0 models.

References

[1] Y. Harrison and J. A. Horne, "Occurrence of 'microsleeps' during daytime sleep onset in normal subjects," *Electroencephalogr. Clin. Neurophysiol.*, vol. 98, pp. 411-416, 1996.

[2] S. K. L. Lal, A. Craig, P. Boord, L. Kirkup, and H. Nguyen, "Development of an algorithm for an EEG-based driver fatigue countermeasure," *J. Safety Res.*, vol. 34, pp. 321-328, 2003.

[3] R. D. Ogilvie, "The process of falling asleep," *Sleep Med. Rev.*, vol. 5, pp. 247-270, 2001.

[4] F. Sagberg, "Road accidents caused by drivers falling asleep," *Accid. Anal. Prev.*, vol. 31, pp. 639-649, 1999.

[5] J. Santamaria and K. H. Chiappa, "The EEG of Drowsiness in Normal Adults," *J. Clin. Neurophysiol.*, vol. 4, pp. 327-382, 1987.

[6] S. Makeig and T. Jung, "Changes in alertness are a principle component of variance in the EEG spectrum," *Neuroreport*, vol. 7, pp. 213-216, 1995.

[7] T. Jung, S. Makeig, M. Stensmo, and T. Sejnowski, "Estimating alertness from the EEG power spectrum," *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 60-69, 1997.

[8] M. T. R. Peiris, R. D. Jones, G. Carroll, J., and P. J. Bones, "Investigation of lapses of consciousness using a tracking task: preliminary results," *Proc. 26th Int. Conf. IEEE Eng. Med. Biol. Soc.*, San Francisco, USA, 2004. 4721-4724.

[9] R. D. Jones, "Measurement of sensory-motor control performance capacities," in *The Biomedical Engineering Handbook*, vol. 2, J. D.

Bronzino, Ed., 2 ed. Boca Raton, Florida: CRC Press, 2000, pp. 149:1-25.
[10] W. W. Wierwille and L. A. Ellsworth, "Evaluation of driver drowsiness by trained raters," *Accid. Anal. Prev.*, vol. 26, pp. 571-581, 1994.
[11] T.-P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, pp. 163-178, 2000.
[12] A. Subasi, "Application of classical and model-based spectral methods to describe the state of alertness in EEG," *J Med Syst*, vol. 29, pp. 473-86, 2005.
[13] I. H. Witten and E. Frank, *Data mining - practical machine learning tools and techniques with Java implementations*: Academic Press, 2000.
[14] P. R. Davidson, R. D. Jones, and M. T. R. Peiris, "Detecting behavioural microsleeps using EEG and LSTM recurrent neural networks," *Proc. 27th Int. Conf. IEEE Eng. Med. Biol. Soc.*, Shanghai, China, 2005. CD-ROM, 4 pages.
[15] M. T. R. Peiris, R. D. Jones, P. R. Davidson, P. J. Bones, and D. J. Myall, "Fractal dimension of the EEG for detection of behavioural microsleeps," *Proc. 27th Int. Conf. IEEE Eng. Med. Biol. Soc.*, Shanghai, China, 2005. CD-ROM, 4 pages.