Automatic sleep stage classification using two-channel electro-oculography

Jussi Virkkalaa,*, Joel Hasanb, Alpo Värric, Sari-Leena Himanenb, Kiti Müllera

a Sleep Laboratory, Brain Work Research Center, Finnish Institute of Occupational Health, Helsinki, Finland
b Department of Clinical Neurophysiology, Medical Imaging Centre, Pirkanmaa Hospital District, Tampere, Finland
c Institute of Signal Processing, Tampere University of Technology, Tampere, Finland

Received 12 May 2007; received in revised form 21 June 2007; accepted 24 June 2007

Abstract

An automatic method for the classification of wakefulness and sleep stages SREM, S1, S2 and SWS was developed based on our two previous studies. The method is based on a two-channel electro-oculography (EOG) referenced to the left mastoid (M1). Synchronous electroencephalographic (EEG) activity in S2 and SWS was detected by calculating cross-correlation and peak-to-peak amplitude difference in the 0.5–6 Hz band between the two EOG channels. An automatic slow eye-movement (SEM) estimation was used to indicate wakefulness, SREM and S1. Beta power 18–30 Hz and alpha power 8–12 Hz was also used for wakefulness detection. Synchronous 1.5–6 Hz EEG activity and absence of large eye movements was used for S1 separation from SREM. Simple smoothing rules were also applied. Sleep EEG, EOG and EMG were recorded from 265 subjects. The system was tuned using data from 132 training subjects and then applied to data from 131 validation subjects that were different to the training subjects. Cohen’s Kappa between the visual and the developed new automatic scoring in separating 30 s wakefulness, SREM, S1, S2 and SWS epochs was substantial 0.62 with epoch by epoch agreement of 72%. With automatic subject specific alpha thresholds for offline applications results improved to 0.63 and 73%. The automatic method can be further developed and applied for ambulatory sleep recordings by using only four disposable, self-adhesive and self-applicable electrodes.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Sleep; Sleep scoring; Sleep stage; Electro-oculography; Ambulatory; Analysis; Automatic; Automated; Computerized

1. Introduction

Traditionally sleep is monitored using a polysomnography with EEG, EOG, EMG and ECG electrodes (Penzel and Conradt, 2000). Especially in ambulatory use, the limiting aspects of the polysomnography include the use of scalp electrodes and the manual scoring of recordings. The scalp electrode placement is more complicated than the use of self-applicable disposable electrodes on areas outside the hairline.

Recording of the sleep stage is important for clinical diagnosis and treatment of sleep disorders (Carskadon and Rechtschaffen, 2005). In the standard approach, sleep is usually segmented into 30 s epochs of wakefulness (W), movement time (MT), sleep stages SREM, S1, S2, S3 and S4 based on features of EEG, EOG and EMG (Rechtschaffen and Kales, 1968). The main information used is the appearance and quantity (density) of certain features within epochs. Standard sleep scoring is a time consuming manual process requiring central scalp electrode, two EOG electrodes, an EMG electrode pair, a reference electrode and a ground electrode (Rechtschaffen and Kales, 1968). Recently modifications have been suggested for standard rules (Iber et al., 2007; Silber et al., 2007).

There is a demand for easily applied automatic methods which could be used in clinical and experimental ambulatory studies and, for instance, for studying the role of sleep duration and quality in the ethiology of metabolic disorders (Knutson et al., 2007). Placement of electrodes outside the hairline would enable the use of self-adhesive electrodes, which could be a self-applicable task (Ehlert et al., 1998; Poree et al., 2006).

The automatic sleep stage scoring using only a two-channel EOG was developed and compared to the standard visual scoring based on EEG, EOG and EMG by using the sleep data of 263 subjects. Algorithms used in our previous studies for the automatic detection of slow wave sleep (Virkkala et al.,

* Corresponding author at: Sleep Laboratory, Brain Work Research Center, Finnish Institute of Occupational Health, Topeliuksenkatu 41 a A, FIN-00250 Helsinki, Finland. Tel.: +358 40 5680360; fax: +358 9 5884759.
E-mail address: jussi.virkkala@ttl.fi (J. Virkkala).
URL: www.neuropdate.com/jussivirkkala (J. Virkkala).
2007a) and for unintentional sleep onset detection (Virkkala et al., 2007b) were combined and extended for separating wakefulness, SREM, S1, S2 and SWS. Extensions included also simple smoothing rules of sleep stages (Baumgart-Schmitt et al., 1998) and automatic subject specific alpha threshold for offline applications.

2. Materials and methods

The study design has been reported in detail earlier by Härmä et al. (2002). The cross-sectional, population-based, random sample study was approved beforehand by the local ethics committee. A total of 265 randomly selected train drivers and railway traffic controllers were recorded for a single night. Polysomnographic recordings were sorted by the amount of visually scored slow wave sleep (SWS). From the sorted list, entries with an odd order number were assigned to the training group and entries with even order number were assigned to the validation group. The subjects and recordings are identical to our previous study, where only slow wave sleep was analysed (Virkkala et al., 2007a). The mean age of subjects was 43 in the training group and 44 in the validation group. Two subjects, one in the training and one in the validation group, had EOG electrode artefacts for a whole night and were excluded from all analyses.

The recording equipment included a digital 16-channel Embla A10 (Medcare Flaga, Reykjavik, Iceland) with a sampling rate of 200 Hz and a bandwidth of 0.5–90 Hz. The visual scoring of the recordings was done based on recorded EOG L-M1, EOG R-M1, C3-M2, O1-M2, and submental EMG. The study employed the standard EOG locations: EOG Left (EOG L) slightly lateral and 1 cm up from the outer canthus and EOG right (EOG R) slightly lateral and 1 cm down from the outer canthus referenced to left mastoid M1 (Rechtschaffen and Kales, 1968). A ground electrode was placed on the forehead. The scoring was done by an experienced sleep technologist according to the standard criteria (Rechtschaffen and Kales, 1968). In this study, movement time (MT) was classified as wakefulness (W), and sleep stages 3 and 4 were called slow wave sleep (SWS). The recordings of every tenth subject in the sorted validation list were rescored by another experienced sleep technologist to obtain inter-scorer agreement. For the automatic analysis, only the channels EOG L-M1, EOG R-M1 and the calculated EOG L–R were used.

Similarly to our previous studies, the analysis was run in 0.5 s steps (Virkkala et al., 2007a,b). Two-second Hann-windowed segments were used, which resulted in a 75% overlap. The segments were filtered using a discrete Fourier transform (DFT) and an inverse discrete Fourier transform (IDFT) from 0.5 to 6, 1 to 6 and 1.5 to 6 Hz. In each segment the cross-correlation between the filtered channels EOG L-M1, EOG R-M1 and peak-to-peak amplitude differences from EOG L–R were calculated. The difference between the cross-correlation of the 1–6 Hz band and the cross-correlation of the 0.5–6 Hz band was used as an indicator of slow eye movements (Virkkala et al., 2007b). If eye movements recorded by EOG are restricted to the 0.5 Hz band and have an opposite phase, this difference is close to 1, and if in addition there is synchronous activity in the 1–6 Hz band, the difference is close to 2. If eye movements are restricted to the 1–6 Hz band and have an opposite phase without any other activity, this difference is close to −1. This slow eye-movement feature is amplitude independent in noise-free measurements.

Alpha power was calculated from the 8 to 12 Hz band from the DFT power spectrum of EOG L-M1. Beta power was calculated from the 18 to 30 Hz band from the DFT power spectrum of EOG L–R. The data flow and basic analysis steps are described in Fig. 1.

Automatic sleep stage classification was done in a hierarchical manner (Fig. 2) and was based on calculating the density of features SW2, SW3, S and S1 indicating the activities of sleep stages S2 or SWS (SW2), SWS (SW3), any sleep (S) and

![Fig. 1. Description of the algorithm data flow and basic analysis steps. DFT indicates the discrete Fourier transform and IDFT the inverse discrete Fourier transform. Alpha power is obtained by summing the 8–12 Hz bins of the DFT of EOG L-M1. Beta power is obtained by summing the 18–30 Hz bins of DFT of EOG L–R. Synchronized activity is calculated in 0.5–6 Hz and in 1.5–6 Hz bands.](image1)

![Fig. 2. Decision tree. Four binary decisions rules SW2T, SW3T, ST and S1T were used to separate S2, SWS, wakefulness, SREM and S1.](image2)
S1 (S1), respectively, during each epoch (Table 1). For each human scored 30 s epoch, the percentages of overlapping 2 s segments fulfilling the following four requirements (rules) were calculated:

- Rule (1). SW2 T for separating S2 and SWS from wakefulness, SREM and S1: correlation and peak-to-peak amplitude difference in 0.5–6 Hz had to be above selected thresholds. Beta power and SEM feature had to be below selected thresholds.
- Rule (2). SW3 T for separating SWS from S2: correlation and peak-to-peak amplitude difference in 0.5–6 Hz had to be above selected thresholds. Beta power and SEM feature had to be below selected thresholds.
- Rule (3). ST for separating SREM and S1 from wakefulness: beta and alpha power had to be below selected thresholds and correlation had to be either below defined threshold or above another threshold.
- Rule (4). S1 T for separating S1 from SREM: correlation and peak-to-peak amplitude difference in 1.5–6 Hz band had to be above selected thresholds with beta power below threshold. Additionally maximum of 1.5–6 Hz band eye-movement (correlation < −0.50) peak-to-peak amplitude in 90 s window had to be below the selected threshold.

A decision tree using these four simple rules is shown in Fig. 2. The percentage of slow wave (SW) segments in each 30 s epoch were calculated using two different sets of thresholds (SW2 and SW3) to separate S2 and SWS from wakefulness, SREM and S1. Wakefulness was detected by counting the number of 2 s segments with alpha and beta power above threshold. Rapid eye-movements (REM) or synchronized 1.5–6 Hz EEG activity with beta power indicated SREM or S1. Separation between S1 and SREM was based on synchronized EEG activity with beta power and maximum of eye movement (defined as correlation < −0.50) peak-to-peak amplitude in a 90 s window (S1 T).

All two-second segments were evaluated for SW2, SW3, S and S1 (rules 1–4). If the percentage (density) of these segments in a 30 s epoch, SW2 T, SW3 T, ST and S1 T was above the selected threshold the corresponding binary decision was taken (Fig. 2). For S1 T, maximum peak-to-peak amplitude of 1.5–6 Hz eye movement (correlation < −0.50) in the 90 s window was also used. High amplitude difference values indicated rapid eye-movement (REM) sleep (SREM). After estimating the sleep stage using the decision tree (Fig. 2), the following two smoothing rules were also applied to provide the final estimate of sleep stages:

- Smoothing (1): Three consecutive epochs of SREM, S2, SREM were replaced with the sequence SREM, SREM, SREM. Similarly, consecutive epochs of S2, S1, S2 were replaced with S2, S2, S2 and Wake, SREM, Wake were replaced with Wake, Wake, Wake.
- Smoothing (2): Any SREM epochs before the very first appearance of sleep stage 2 (S2) were replaced with wakefulness epochs.

For offline applications, automatic subject specific alpha threshold was determined by calculating mean alpha power during segments that had positive cross-correlation in the 1.5–6 Hz band. This value was used for calculating new threshold for rule 3. For reporting agreement, concordance between the two scorings, Cohen’s Kappa (Cohen, 1960) and epoch-by-epoch agreement were used. Cohen’s Kappa $\kappa$ is the proportion of agreement after change agreement $p_c$ is removed from consideration. Using probabilities $p_{ij}$ from agreement matrix the $\kappa$ can be defined as (Cohen, 1960):

$$\kappa = \frac{p_o - p_c}{1 - p_c}, \quad p_o = \sum_i p_{ii}, \quad p_c = \sum_i \left[ \frac{\sum_j p_{ij}}{\sum_j p_{ij}} \right]$$

Cohen’s Kappa values greater than 0.80 represent almost perfect agreement (Landis and Koch, 1977). Cohen’s Kappa values between 0.61 and 0.80, 0.41 and 0.60, 0.21 and 0.40 and 0 and 0.20 represent substantial, moderate, fair, and slight agreement, respectively (Landis and Koch, 1977).

### Table 1
**Description of the used features**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description of feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEM</td>
<td>Slow eye movements indicated as difference between cross-correlations of EOG L-M1 and EOG R-M1 and between bands 1–6 and 0.5–6 Hz</td>
</tr>
<tr>
<td>SW2</td>
<td>Synchronous 0.5–6 Hz activity with low SEM, 8–12 Hz alpha and 18–30 Hz beta power between EOG L-M1 and EOG R-M1</td>
</tr>
<tr>
<td>SW3</td>
<td>Synchronous 0.5–6 Hz activity with low SEM, 8–12 Hz alpha, and 18–30 Hz beta power between EOG L-M1 and EOG R-M1</td>
</tr>
<tr>
<td>S</td>
<td>Low 8–12 Hz alpha and 18–30 Hz beta power with either between EOG L-M1 and EOG R-M1 synchronous 1.5–6 Hz activity or eye movement</td>
</tr>
<tr>
<td>S1</td>
<td>Synchronous 1.5–6 Hz activity with low 8–12 Hz alpha and 18–30 Hz beta power without eye movements between EOG L-M1 and EOG R-M1</td>
</tr>
</tbody>
</table>

### Table 2
**Number of subjects, age, number of 30 s epochs and sleep parameters in the training and validation group**

<table>
<thead>
<tr>
<th>Subjects (males)</th>
<th>Age mean (range)</th>
<th>Epochs</th>
<th>Wake (%)</th>
<th>SREM (%)</th>
<th>S1 (%)</th>
<th>S2 (%)</th>
<th>SWS (%)</th>
<th>AHI (SD) h⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>132 (117)</td>
<td>43 (26–61)</td>
<td>134,744</td>
<td>16.0%</td>
<td>17.0%</td>
<td>12.5%</td>
<td>41.5%</td>
<td>12.4% 8.3 (12.4)</td>
</tr>
<tr>
<td>Validation</td>
<td>131 (120)</td>
<td>43 (28–60)</td>
<td>134,887</td>
<td>15.7%</td>
<td>17.7%</td>
<td>12.1%</td>
<td>41.7%</td>
<td>12.1% 5.8 (12.1)</td>
</tr>
</tbody>
</table>

AHI data of five training and one validation subject was unavailable.
3. Results

Descriptions of the training and validation groups are presented in Table 2. The total number of the scored 30 s epochs was 269,631 corresponding to about 2247 h of data. In the training group there were 16% wakefulness, 17% SREM, 13% S1, 42% S2, and 12% SWS epochs. In the validation group the values were within one percent identical. The sleep efficiency index in the training and in the validation groups was 84%. The mean apnoea–hypopnoea index (AHI) was 8.3 h^{-1} in the training group and 5.8 h^{-1} in the validation group.

The median, mean and standard deviation of the SEM feature of every segment in wakefulness and in different stages using the data of 132 training subjects are presented in Fig. 3(a). The median, mean and standard deviation of the maximum 1.5–6 Hz band eye movement (defined as cross-correlation < -0.50) in a 30 s epoch are presented in Fig. 3(b).

The maximum of Cohen’s Kappa was sought by a systemic variation of thresholds for each of the four rules using only the data of the training group. The optimal thresholds resulting in the highest Cohen’s Kappa in each four binary decisions in separating wakefulness and sleep stages are given in Table 3. When combined together using the decision tree (Fig. 2), Cohen’s Kappa of 0.63 (epoch-by-epoch agreement 72%) was obtained in separating wakefulness, SREM, S1, S2 and SWS. Smoothing improved agreement to 73% and Cohen’s Kappa to 0.64. With individual alpha thresholds, the results were 74% and 0.65.

When applying the obtained optimal thresholds to the data of the validation group, the epoch-by-epoch agreement was 71% and Cohen’s Kappa was 0.60 (Table 4). By using the two smoothing rules, the agreement was increased to 72% with Cohen’s Kappa of 0.62. With additional automatic subject specific alpha thresholds, the results were 73% and 0.63. Fig. 4 shows an example of calculated SW2T, SW3T, ST, S1T percentages during 30 s together with automatic and visual sleep stage classification.

The visual inter-scorer agreement using standard montage was 82% with Cohen’s Kappa of 0.76 for the subset of subjects. For the same subjects automatic classification resulted in an agreement of 73% and Cohen’s Kappa of 0.63.
4. Discussion

Using only two-channel EOG data and automatic analysis, the epoch-by-epoch agreement was 72% and Cohen’s Kappa was 0.62 in separating S2, SWS, wakefulness, SREM, and S1 when compared to the standard visual scoring with the full montage. If online capability of the algorithm is not needed, the results increase to 73% and 0.63 using automatic determination of subject specific alpha threshold. The inter-rater agreement was evaluated for a subset of data. As expected, it was better than the one obtained with the automatic method. Consensus scoring was not available for the current study. With consensus scoring, we would expect the man–machine agreement to improve (Anderer et al., 2005). The cost-benefit of this laborious task would, however, have been low.

We used a total of 20 different thresholds (Table 3). The thresholds were optimized for four different binary decision rules 1–4 for separating wakefulness, S1, SREM, S2 and SWS (Fig. 2). All the thresholds, apart from subject specific alpha threshold for offline applications, were fixed across subjects. There are various other automatic sleep classifiers that in addition to EOG also use EEG and EMG (Agarwal and Gotman, 2001; Anderer et al., 2005; Hasan, 1983; Martin et al., 1972; Park et al., 2000). Use of automatic and semi-automated scoring of polysomnographic recordings has been recently discussed by Svetnik et al. (in press). Standard visual and automatic scoring also requires the use of central scalp EEG and EMG. With more electrodes and more complicated algorithms better agreement results have been observed. Some devices only use single channel EEG (Baumgart-Schmitt et al., 1998; Berthomier et al., in press; Davies et al., 1999; Flexer et al., 2005). Using, for instance, neural networks or linear discriminator analysis might have resulted in slightly better results, but we wanted to keep the system as simple as possible. All thresholds used are reported in this study. The thresholds were optimized for one scorer and for the used hardware and may have to be changed for replication with different laboratory practise and devices. Previously used sleep epoch smoothing rules (Baumgart-Schmitt et al., 1998) proved to increase the agreement also in this study.

Two subjects out of 265 were excluded due to EOG L artefacts. No other manual artefact handling was used and it is possible that some misclassifications have resulted from artefacts that remained undetected by the used cross-correlation and beta thresholds. Alpha activity was measured from the channel EOG L–M1 and it is more likely that the activity originates from the reference electrode M1 than from the EOG L electrode (Flanigan et al., 1995). We have selected the EOG L–R for calculation of beta power and amplitudes to reduce the effect of, for instance ECG, artefacts and to make the system less dependent on the location of the reference electrode. Change of reference electrode location does not affect the beta power or amplitude calculations. In a previous study, we used the same slow eye-movement feature to separate wakefulness and sleep onset (Virkkala et al., 2007b). In this study the mean of SEM feature was quite similar between wakefulness and S1 (Fig. 3). This difference between studies is likely to be the result of subjects having eyes closed during night-time wakefulness epochs.
We observed SEM features during SREM (Fig. 3) as has been observed by Magosso et al. (2007).

It is interesting that in separating wakefulness (see rule 3) the optimal threshold percentage of segments was almost 50\% like in visual scoring. In SWS the optimal percentage was 28\% in contrast to 20\% in visual scoring. It was also observed that calculated slow wave (SW) feature could also separate sleep stage 2 (S2) from wakefulness, S1 and SREM. Traditionally, sleep spindle or sigma activity are used for sleep stage scoring (Anderer et al., 2005). In this study no spindle or sigma activity detection was used. Calculated percentages of SW2T, SW3T, ST and S1T also indicate microstructure of sleep as shown in Fig. 4. The usefulness of these indicators of sleep microstructure should be studied further.

The detection of S1 was the most problematic of the sleep stages. It is not easy to find a simple feature that would separate S1 clearly from S2, SREM and wakefulness. This has been observed previously by many authors (e.g. Anderer et al., 2005). This could be due to S1 being a transition phase between wakefulness and different sleep stages as discussed by Himanen and Hasan (2000). In disturbed sleep S1 can comprise a large proportion of the night.

EOG and frontal EEG electrodes were used for delta detection already in 1972 by Hilbert and Naitoh (1972). Later they have been used for visual sleep stage detection (Dyson et al., 1984; Lapinlampi and Himanen, 2004; Werth and Borbely, 1995). Automatic epoch-by-epoch classification has however been implemented only for single sleep stages (Virkkala et al., 2007a,b). Eyelid movements have been combined with body movement for NREM and REM evaluation (Kayed et al., 1979; Mamelak and Hobson, 1989). It is also common to use cross-correlation to separate eye movements for synchronized delta activity (Drewes et al., 2000; Virkkala et al., 2007a). There are also systems where electrodes can be placed on the forehead (Fischer et al., 2004; Poree et al., 2006).

The aim of the present study was to derive an estimate of sleep stage defined using central and occipital EEG, EOG and EMG by analysing only two-channel EOG. Beside EOG L and EOG R electrodes, also reference M1 and ground electrodes are needed. There are probably central EEG features that remain undetected when analysing only EOG electrodes. However, the advantage is that the electrodes can be disposable and self-adhesive, which makes electrode placement an easy, self-applicable task compared to the placement of standard central and occipital electrodes. In recent scoring updates (Iber et al., 2007; Silber et al., 2007), slightly different EOG montage (Häkkinen et al., 1993) is recommended. Scoring of this study was based on original guidelines (Rechtschaffen and Kales, 1968).

Most portable sleep devices do not enable diagnosis of non-breathing-related sleep disorders. In the future the developed method combined with currently used cardiorespiratory recorders could be used for home screening of sleep disor-

### Table 4
Obtained agreement matrix, agreement percentage, and Cohen’s Kappa in validation data without smoothing, with smoothing, and with automatic determination of subject specific alpha threshold

<table>
<thead>
<tr>
<th></th>
<th>Wake</th>
<th>SREM</th>
<th>S1</th>
<th>S2</th>
<th>SWS</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No smoothing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wake</td>
<td>16,032</td>
<td>3105</td>
<td>1466</td>
<td>714</td>
<td>62</td>
<td>75.0%</td>
</tr>
<tr>
<td>SREM</td>
<td>1719</td>
<td>17,298</td>
<td>2281</td>
<td>2704</td>
<td>19</td>
<td>72.0%</td>
</tr>
<tr>
<td>S1</td>
<td>2393</td>
<td>3850</td>
<td>5514</td>
<td>4619</td>
<td>61</td>
<td>33.5%</td>
</tr>
<tr>
<td>S2</td>
<td>1645</td>
<td>1079</td>
<td>5171</td>
<td>4,4212</td>
<td>4488</td>
<td>78.1%</td>
</tr>
<tr>
<td>SWS</td>
<td>213</td>
<td>20</td>
<td>49</td>
<td>3685</td>
<td>12,488</td>
<td>75.9%</td>
</tr>
<tr>
<td>Specificity</td>
<td>72.9%</td>
<td>68.2%</td>
<td>38.1%</td>
<td>79.0%</td>
<td>73.0%</td>
<td></td>
</tr>
<tr>
<td>Agreement</td>
<td>70.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohen’s Kappa</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Smoothing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wake</td>
<td>16,701</td>
<td>2443</td>
<td>1394</td>
<td>779</td>
<td>62</td>
<td>78.1%</td>
</tr>
<tr>
<td>SREM</td>
<td>1889</td>
<td>17,564</td>
<td>2197</td>
<td>2352</td>
<td>19</td>
<td>73.1%</td>
</tr>
<tr>
<td>S1</td>
<td>3019</td>
<td>3279</td>
<td>4990</td>
<td>5088</td>
<td>61</td>
<td>30.4%</td>
</tr>
<tr>
<td>S2</td>
<td>1821</td>
<td>955</td>
<td>4205</td>
<td>4,5126</td>
<td>4488</td>
<td>79.7%</td>
</tr>
<tr>
<td>SWS</td>
<td>215</td>
<td>18</td>
<td>36</td>
<td>3698</td>
<td>12,488</td>
<td>75.9%</td>
</tr>
<tr>
<td>Specificity</td>
<td>70.6%</td>
<td>72.4%</td>
<td>38.9%</td>
<td>79.1%</td>
<td>73.0%</td>
<td></td>
</tr>
<tr>
<td>Agreement</td>
<td>71.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohen’s Kappa</td>
<td>0.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subject specific alpha threshold and smoothing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wake</td>
<td>17,035</td>
<td>2329</td>
<td>1182</td>
<td>771</td>
<td>62</td>
<td>79.7%</td>
</tr>
<tr>
<td>SREM</td>
<td>1273</td>
<td>18,153</td>
<td>2235</td>
<td>2341</td>
<td>19</td>
<td>75.6%</td>
</tr>
<tr>
<td>S1</td>
<td>2800</td>
<td>3465</td>
<td>5023</td>
<td>5088</td>
<td>61</td>
<td>30.6%</td>
</tr>
<tr>
<td>S2</td>
<td>1689</td>
<td>995</td>
<td>4299</td>
<td>4,5124</td>
<td>4488</td>
<td>79.7%</td>
</tr>
<tr>
<td>SWS</td>
<td>183</td>
<td>17</td>
<td>65</td>
<td>3702</td>
<td>12,488</td>
<td>75.9%</td>
</tr>
<tr>
<td>Specificity</td>
<td>74.1%</td>
<td>72.7%</td>
<td>39.2%</td>
<td>79.1%</td>
<td>73.0%</td>
<td></td>
</tr>
<tr>
<td>Agreement</td>
<td>72.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohen’s Kappa</td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sensitivity and specificity of automatic detection for each sleep stage separately is also reported. Rows are by human scorer and columns by automatic method.
Acknowledgements

We thank Riitta Velin, Susan Pihl and Nina Lapveteläinen for their work in recording and scoring the material and Hanna Liikala for language editing. The study was financially supported by the Finnish Work Environment Fund project ‘Computerized testing of railway traffic safety workers, assessment of vigilance and cognitive performance’ and the Finnish Funding Agency for Technology and Innovation project ‘MICROSLEEP Development of analysis methods for the dynamic structure of sleep’.

References

Dyson RJ, Thorton C, Dore CJ. EOG electrode positions outside the hairline to monitor sleep in man. Sleep 1984;7:180–8.