available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

EEG gamma frequency and sleep–wake scoring in mice: Comparing two types of supervised classifiers

Jurij Brankačk[a,*], *Valeriy I. Kukushka*[b], *Alexei L. Vyssotski*[c], *Andreas Draguhn*[a]

[a]*Institute for Physiology and Pathophysiology, University Heidelberg, Im Neuenheimer Feld 326, 69120 Heidelberg, Germany*
[b]*Laboratory of Biophysics and Bioelectronics, Dniepropetrovsk National University, 49050 Dniepropetrovsk, Ukraine*
[c]*Institute of Neuroinformatics, University of Zürich/ETH, Winterthurerstr. 190 CH-8057 Zürich, Switzerland*

ARTICLE INFO

Article history:
Accepted 26 January 2010
Available online 1 February 2010

Keywords:
Sleep stage scoring
EEG frequency
Gamma activity
Theta activity
Period-amplitude analysis
Linear discriminant analysis

ABSTRACT

There is growing interest in sleep research and increasing demand for screening of circadian rhythms in genetically modified animals. This requires reliable sleep stage scoring programs. Present solutions suffer, however, from the lack of flexible adaptation to experimental conditions and unreliable selection of stage-discriminating variables. EEG was recorded in freely moving C57BL/6 mice and different sets of frequency variables were used for analysis. Parameters included conventional power spectral density functions as well as period-amplitude analysis. Manual staging was compared with the performance of two different supervised classifiers, linear discriminant analysis (LDA) and Classification Tree. Gamma activity was particularly high during REM (rapid eye movements) sleep and waking. Four out of 73 variables were most effective for sleep–wake stage separation: amplitudes of upper gamma-, delta- and upper theta-frequency bands and neck muscle EMG. Using small sets of training data, LDA produced better results than Classification Tree or a conventional threshold formula. Changing epoch duration (4 to 10 s) had only minor effects on performance with 8 to 10 s yielding the best results. Gamma and upper theta activity during REM sleep is particularly useful for sleep–wake stage separation. Linear discriminant analysis performs best in supervised automatic staging procedures. Reliable semi-automatic sleep scoring with LDA substantially reduces analysis time.

1. Introduction

Sleep research in human subjects and animals is of increasing importance, mainly due to the assumed role of sleep in memory consolidation (for recent reviews see: Axmacher et al. (2009); Brankačk et al. (2009) and Walker (2009)) and to the expanding need for sleep phenotyping of genetically modified mice (Pang et al., 2009). The main two sleep stages NREM (non rapid eye movements) and REM (rapid eye movements) can be easily distinguished from each other and from waking by visual inspection of the EEG and EMG. In rodents, reliable staging can be achieved from epidural recordings of cortex

* *Corresponding author*. Fax: +49 6221 546364.
E-mail address: jurij.brankack@physiologie.uni-heidelberg.de (J. Brankačk).

Abbreviations: ANOVA, analysis of variance; EEG, electroencephalography; EMG, electromyography; FFT, Fast Fourier transform; LDA, linear discriminant analysis; NPV, negative predictive value; N, NREM, non rapid eye movements; R, REM, rapid eye movements; SEM, standard error of the means; PAA, period-amplitude analysis; PCA, principal component analysis; PPV, positive predictive value; W, waking

0006-8993/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2010.01.069

areas overlaying the dorsal hippocampus (allowing for the detection of theta rhythms) together with EMG activity from neck muscles (Timo-Iara et al., 1970; Brankačk and Buzsáki, 1986; see Fig. 1A and for details see Experimental procedures section). Manual sleep scoring by visual inspection, however, is extremely time-consuming. As a rule of thumb, analysis time almost equals recording time. Therefore, during the past three decades numerous attempts of automatic or semi-automatic sleep scoring have been made, using a large variety of methods (reviews in Robert et al. (1999); in mice: Veasey et al. (2000); recently: Crisler et al. (2008)). Despite of these enormous efforts, the accuracy of most commercially available sleep scoring programs is dissatisfying. Besides being costly, these programs are fixed and have been often designed for a particular experimental condition performing poorly under differing conditions.

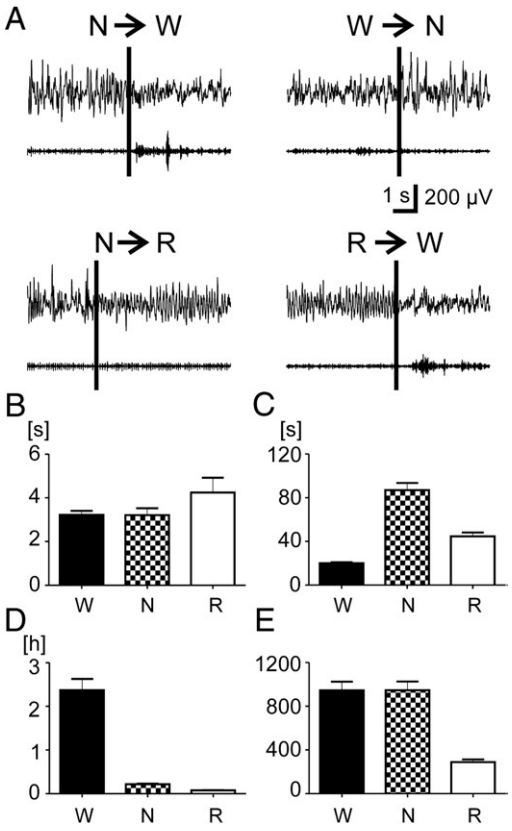

Fig. 1 – A: Examples of EEG and EMG recordings during representative transitions (transitions points: vertical markers) of sleep–wake stages: from NREM sleep to waking (N→W); from waking to NREM sleep (W→N); from NREM to REM sleep (N→R) and from REM sleep to waking (R→W). The upper traces show EEG recordings from the lateral parietal cortex overlying hippocampus, the lower traces are recordings of neck muscle EMG. Time and amplitude scales are identical for all traces. Note the increase in EMG activity and a drop in EEG amplitude during awakening (N→W, R→W). Please note also the high-amplitude slow wave activity during NREM, the lower amplitude irregular activity in quiet waking (W→N shortly before N) and the regular theta (7–8.5 Hz) rhythm during REM sleep (R). These changes in EEG and EMG activity were used as criteria for the manual sleep–wake staging. B–E: Means and S.E. (N=10 animals) of minima (B), of individual medians (C), of maxima (D) and of the number (E) of manually classified stage durations for waking (W), NREM (N) and REM (R).

We have developed a custom-made program for reliable and flexible sleep–wake stage scoring with supervised learning algorithms. We compared three different sleep scoring methods with results of visual inspection: i) a conventional threshold formula as often implemented in commercially available programs (Mochizuki et al., 2004); ii) linear discriminant analysis (LDA); and iii) Classification Tree (Tree). The latter two solutions are supervised classifiers (MacLachlan, 1992; Fielding, 2006), allowing for flexible adaptation to different experimental conditions and parameters. After processing small sets of training data, the program reliably differentiated waking, NREM and REM, using EMG, delta, upper theta and upper gamma amplitudes as discriminating parameters.

2. Results

2.1. *Manual sleep–wake scoring by visual inspection*

Continuous EEG recordings of at least 72 h were recorded from ten mice and were manually scored into three stages: wakefulness (W), NREM (N) and REM (R) according to the criteria described in Experimental procedures. Fig. 1A shows representative transitions between the three stages indicated with vertical markers (vertical lines in Fig. 1A). The term "transition" is used in its direct meaning, i.e. the end of one stage corresponding to the beginning of the next stage and not in the meaning that one stage may be separated from the next by a transitional period not belonging to either of the two neighbouring stages. High-amplitude EMG activity indicated wakefulness when combined with low-amplitude irregular EEG (quiet waking) or with regular theta rhythm (active waking; see "W" in Fig. 1A). Lower EMG activity with high-amplitude delta (1–4 Hz) waves was classified as NREM (see "N" in Fig. 1A) and low EMG activity with regular theta (4.5–8.5 Hz) rhythm corresponded to REM sleep (see "R" in Fig. 1A). The criteria of placing transitional markers are demonstrated in Fig. 1A: W→N: first large-amplitude slow wave; N→W: beginning of EEG desynchronisation; N→R: beginning of regular theta rhythm, regardless of its amplitude; R→W: end of regular theta rhythm, regardless of EMG. The duration of manually scored stages varied between 1.6 s and 218 min. Averaged (mean±SEM) over ten animals, minimal stage durations (Fig. 1B) corresponded to 3.2±0.2 s in waking, 3.2±0.3 s in NREM and 4.2±0.7 s in REM sleep. The means of individual medians (W: 20.1±0.8 s, N: 87.0±6.6 s, R: 44.6±3.5 s) are shown in Fig. 1C and the mean maximal stage durations (W: 8539.0±925.0 s, N: 778.5±41.3 s, R: 272.6±9.2 s) are shown in Fig. 1D. Fig. 1E illustrates the mean numbers of stages (W: 946±80, N: 946±80, R: 289±24). No stage duration differences were found between female (N=7) and male (N=3) mice.

We then averaged the percentages of different states for the two recorded dark/light periods in each animal. No difference was found between the seven female and three

male mice. Averaged over ten mice, sleep occupied 64.79% of the light period (NREM: 54.50±0.98%, REM: 10.29±0.24%; means±SEM) and 35.54% of the dark period (NREM: 31.77±1.17%, REM: 3.78±0.17%). During the light period, REM sleep covered 15.9±0.5% of the total sleep time (range: 13 to 19). During darkness, the percentage of REM sleep was 10.7±0.4% (10 to 12) of total sleep. The sleep diurnal ratio (light period/dark period) corresponded to 1.84±0.07 (1.61 to 2.24), the NREM diurnal ratio was 1.74±0.07 (1.54 to 2.11) and REM diurnal ratio: 2.77±0.15 (2.20 to 3.45). The wake diurnal ratio was 0.55±0.02 (0.47 to 0.66). Fig. 4 shows the resulting percentages of each state during the diurnal cycle. These data reflect the typical circadian rhythm of mice, indicating that the recording did not disturb normal sleep-activity cycles.

2.2. *Gamma activity and other sleep scoring parameters*

For comparison with different automated scoring algorithms, we divided manually scored recordings into 10 s epochs, after testing different lengths between 1 and 20 s (but see Section 2.5 below). Power spectral density as well as period-amplitude analysis (PAA) was performed on EEG recordings from the lateral parietal cortex for each of the ten frequency bands listed in Table 1. This EEG channel overlaying the hippocampus was sufficient because it contained the essentials of sleep–wake EEG changes at delta, theta and gamma bands. In addition, we used the integrated amplitude of neck muscle EMG, two ratios, one between theta and delta power and another between theta and delta amplitude as well as rhythm scores for each of the ten frequency bands. In total, these calculations resulted in 73 different parameters (see Experimental procedures for details). Principal component analysis (PCA) revealed parameters which contributed most strongly to the total variance of data, i.e. the integrated amplitude of the sigma band 9–14 Hz. The principal components did not discriminate well between sleep wake stages neither did the most contributing parameters. We therefore empirically verified the efficacy of pairs or triples of parameters to differentiate between the sleep–wake stages (see, for example, Fig. 3A). Comparing automatic scoring with visually identified stages, we found that four parameters revealed the largest differences between W, NREM and REM. Three of these parameters were integrated EEG amplitudes calculated with PAA, namely delta (1–4 Hz), theta2 (7–8.5 Hz) and gamma2 (52–70 Hz). In addition, we used integrated amplitudes of neck muscle EMG.

Table 1 – Definition of frequency bands.

Frequency band	Frequency analysed [Hz]	Band pass filter for PAA [Hz]
Delta	1–4	1–4.5
Theta	4.5–8.5	4–10
Theta1	4.5–6.5	4–10
Theta2	7–8.5	4–10
Sigma	9–14	7–20
Beta1	14.5–18.5	10–40
Beta2	19–30	10–40
Gamma1	30.5–48	25–80
Gamma2	52–70	25–80
Total	1–70	1–80

The upper and lower limits were used for power spectral density and period-amplitude analysis (PAA). Prior to PAA, raw data were band pass filtered between the frequencies indicated.

Table 2 – Coincidences and errors of semi-automated analysis.

	LDA100%	Tree100%	LDA5%	Tree5%
Percentage coincidence of automated with manual scoring				
W Power	81.2±1.1	90.2±0.8	81.0±1.1#	85.8±1.1+
W PAA	83.2±0.8*	91.1±0.7*	83.5±1.0*#	87.0±1.1+
N Power	93.1±0.9	94.5±0.4*	93.0±1.0	89.8±1.4+
N PAA	94.2±0.5	94.1±0.5	93.6±1.0	89.4±0.4+
R Power	80.5±2.1	80.7±0.8	81.0±2.8	73.6±2.4+
R PAA	90.2±1.0*	83.4±1.0*	89.4±0.8*#	79.2±2.9*
Percentage of total errors (false positives+false negatives)				
W Power	20.0±1.2	11.3±0.9	20.0±1.1	20.0±1.6+
W PAA	17.6±0.9*	10.1±0.8*	17.3±0.9*	18.0±1.1+
N Power	16.8±0.9	7.6±0.6	16.0±0.6	16.7±1.8+
N PAA	12.9±0.9*	7.6±0.6	14.8±2.3	16.4±1.1+
R Power	53.4±5.9	22.4±1.4	55.4±6.0#	39.9±6.5+
R PAA	32.5±6.2*	18.5±1.2*	31.5±6.4*	32.0±5.8+

Scores obtained by LDA and Tree, respectively, are compared to manual sleep–wake scoring. Upper part indicates means±SEM of coincidences. Lower part indicates total amount of errors. Averaged data from 7 mice are given in % of all epochs analysed. Symbols for significances ($p<0.05$): *analysis of power spectrum (Power) versus period-amplitude analysis (PAA); +training with 100% versus 5% of data; #LDA versus Tree (both with 5% training data). W: waking, N: NREM, R: REM.

We calculated integrated gamma2 amplitudes by PAA analysis from all seven mice. This parameter was significantly larger in REM ($p<0.0001$, *T*-test) and waking ($p<0.0001$, *T*-test) as compared to NREM. It could not differentiate, however, between REM and waking ($p>0.05$, *T*-test). In contrast, the power spectral density (band power) for the same frequency band revealed smaller differences between sleep–wake stages and was significantly less successful when used with the supervised classifiers LDA and Tree (see significances marked with "*" in Table 2). Sleep–wake transitions of gamma2 amplitudes were closely paralleled by similar changes of integrated EMG amplitudes. However, gamma2 amplitudes were larger in REM sleep compared to NREM whereas EMG amplitudes were lower in REM. In addition to gamma2, EEG amplitudes of upper theta (7 to 8.5 Hz) and delta (1 to 4 Hz) and the EMG amplitude were found to discriminate well between sleep–wake stages. Upper theta amplitude was significantly larger during REM sleep as compared to NREM sleep ($p<0.05$, *T*-test) and waking ($p<0.005$, *T*-test). Fig. 2 displays changes of all four amplitudes with sleep wake alternations. Note that none of the EEG-derived parameters seems to be well suited for threshold settings to discriminate among sleep–wake stages.

2.3. *Semi-automated sleep scoring using supervised classifiers*

Threshold-based algorithms of sleep–wake staging require the time-consuming determination of optimal thresholds for each dataset and animal. This can only be done with extensive manual data inspection and does frequently still yield

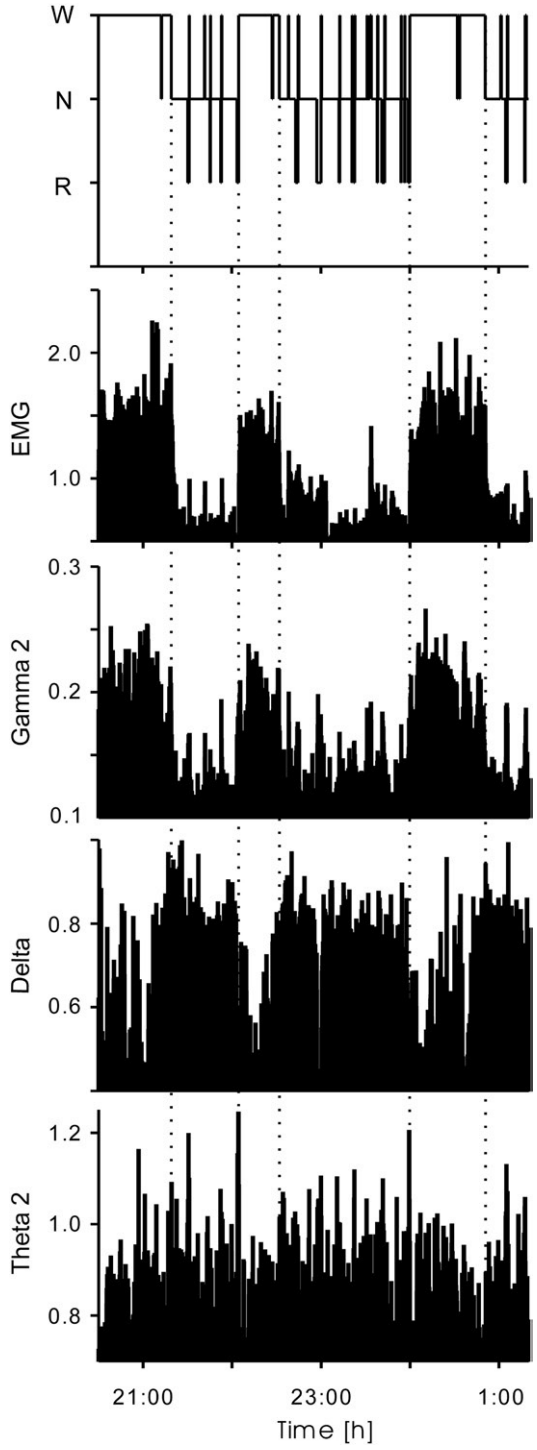

Fig. 2 – Amplitude changes of EMG, gamma2, delta and theta2 during several sleep–wake alterations in a representative animal. Note the similarity between EMG and gamma2 amplitude variations with waking and sleep. Note also that in spite of the correlative amplitude variations with sleep and waking the threshold settings to discriminate between the three stages W (waking), N (NREM) and R (REM) are not unambiguous.

dissatisfying results. In contrast, supervised classifiers like the linear discriminant analysis (LDA) and the Classification Tree (Tree) learn to discriminate between classes by a limited set of training data. LDA delineates boundary areas between the classes in *n*-dimensional parameter space. Fig. 3 provides a three-dimensional illustration of such a parameter space, using manually scored epochs (W = black, NREM = blue, REM = red). The dimensions are the four parameters which produced best results with both LDA and Tree: integrated amplitudes of EMG; delta EEG frequency (1–4 Hz); theta2 (7–8.5 Hz); gamma2 (52–70 Hz). Obviously, NREM and REM epochs are well separated by the three integrated EEG amplitudes (Fig. 3A; waking epochs omitted for clarity). The importance of the neck muscle EMG amplitude for differentiating waking from both sleep stages is apparent in Fig. 3B. By using manually defined epochs of the entire dataset (100%) as training data, the percentage of coincidences between automated and manual scored epochs is slightly superior in Tree compared to LDA. The percentage of errors (sum of false positives and false negatives divided by the total number of manually scored epochs) was about 50% lower for Tree as compared to LDA (Table 2). While these findings indicate superior performance of the Classification Tree algorithm, the use of 100% of

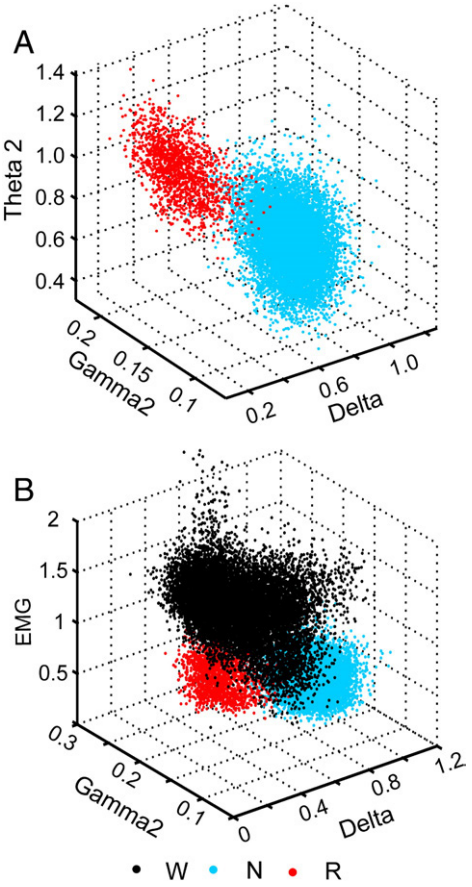

Fig. 3 – Three-dimensional distribution of waking, NREM and REM sleep epochs in LDA parameter space. A: NREM (N, blue) and REM (R, red) epochs are well discriminated by the integrated amplitudes of delta (1–4 Hz), gamma2 (52–70 Hz) and theta2 (7.5–8 Hz). Waking epochs are omitted for clarity. B: NREM and REM sleep epochs are well separated from waking (W, black) by the integrated amplitudes of EMG in combination with those of delta and gamma2.

predefined training data is, of course, useless in practical applications. We therefore repeated the analysis with 5% of training data, randomly distributed throughout the dataset. Under these conditions, performance of the Tree classifier decreased significantly (significant changes marked with "+" in Table 2). However, performance of LDA was not different between analyses with 100% and 5% training data, respectively. Table 2 shows means (±SEM) of coincidences between automated and manual scoring and the corresponding errors for LDA and Tree, averaged for all seven animals. For 5% training data the coincidences between automated and manual scoring for REM epochs were significantly larger with LDA (89.4±0.8%) compared to Tree (79.2±2.9%; $p=0.0082$, paired *T*-test). In contrast, coincidence values for waking epochs were significantly larger in Tree (87.0±1.1%) compared to LDA (83.5±1.0%; $p=0.0095$, paired *T*-test). Coincidences for NREM and error rates for all three stages are not significantly different between LDA and Tree both with 5% training data. Fig. 4 shows the averaged circadian sleep–wake cycles from ten mice with manual (black circles) and semi-automated (grey circles) scoring using LDA and 5% randomly distributed training data. Two-way ANOVA followed by the Bonferroni post-test group comparison revealed no significant differences between manual and semi-automated scoring for any of the time points and stages, except for REM (Fig. 4).

2.4. *Comparing supervised classifiers with a threshold formula*

Finally, we tested data scoring by the threshold formula: "if EMGIntegral>X, Stage='W';, else, if Ratio<Y, Stage='N';, else, Stage='R'; end, end" (for details see Experimental procedures). Coincidences between automated and manual scoring for waking epochs were significantly smaller with the threshold formula (77.4±2.0%) compared to LDA (83.5±1.0; $p=0.0168$, paired *T*-test) and Tree (87.0±1.1; $p=0.0005$, paired *T*-test). Coincidences for REM sleep were also significantly smaller with the formula (73.3±3.6%) compared to LDA (89.4±0.8 %; $p=0.0048$, paired *T*-test) but were not different from Tree (79.2±2.9). The coincidences for NREM were significantly larger with the threshold formula (95.7±0.4) compared to Tree (89.4±0.4; $p=0.0001$, paired *T*-test) but there was no difference to LDA (93.6±1.0). Error rates resulting from classification with the formula were larger for waking (24.0±2.0%) and REM (42.2±6.0%) epochs when compared to LDA (W: $p=0.0024$; R: 0.0118, both paired *T*-test) and Tree (W: $p=0.0026$, paired *T*-test; R: not significant). For NREM epochs the error rate with the threshold formula (21.4±1.8%) was significantly larger compared to Tree ($p=0.0275$, paired *T*-test) only. In total, classification by the threshold formula proved to be far less reliable and efficient than that by both supervised classifying algorithms.

2.5. *LDA scoring performance, stage transitions and epoch duration*

Next the effect of epoch duration on semi-automated scoring with LDA (5% random training data) was examined. As described

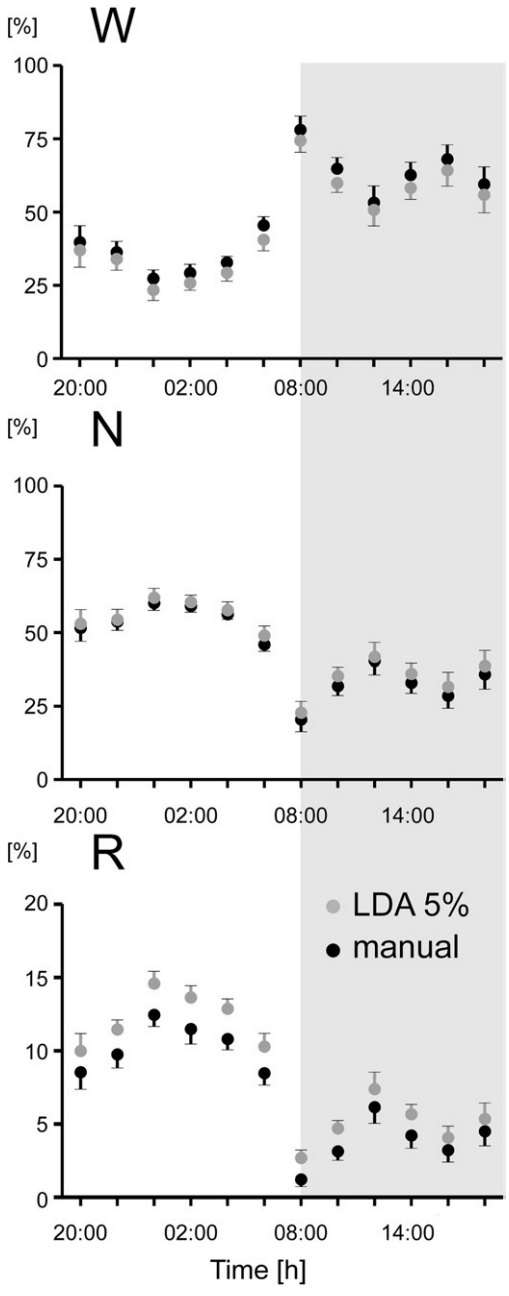

Fig. 4 – Circadian sleep–waking cycle based on manual (black filled circles) and semi-automated (grey open circles) sleep scoring. The graphs represent means and SEM of 10 animals for waking (upper graph), NREM (middle graph) and REM sleep (lower graph) averaged for two hour bins over two light/dark periods in each of the seven animals. Manual sleep scoring by visual inspection was based on the criteria shown in Fig. 1. Semi-automated sleep scoring was carried out with the supervised classifier linear discriminant analysis (LDA) using four integrated amplitudes (EMG, delta, theta2 and gamma2) and 5% of manually scored epochs as training data. The light period started at 20:00 h, the dark period at 8:00 h (shaded area). For waking and NREM semi-automated scoring does not differ significantly from manual sleep scoring (two-way ANOVA followed by Bonferroni post tests, $p>0.05$ for all time points). For REM sleep the results of semi-automated scoring revealed significantly higher values for all time points with the exception of 16:00 and 18:00 h (Bonferroni post tests, $p<0.05$).

below in Experimental procedures, Section 4.4, the boundaries of stages defined by manual scoring may not coincide with the start and end of epochs used with semi-automated scoring. First, the minimal duration of manually scored stages may be shorter than the LDA epoch duration selected (see Fig. 1B for minimal stage durations); second, the manually defined stage durations may not be exactly equal to multitudes of LDA epoch durations. This leads to LDA epochs with more than one manually scored stage, so called transitions. The percentage of transitions increases linearly with epoch duration (Fig. 5A) from 3.3±0.3% for 4 s epochs to 7.8±0.6% for 10 s epochs. Transitions may influence the outcome of semi-automated scoring. To examine the effect of transitions, we calculated coincidences between LDA and manual scoring separately for epochs without (Fig. 5B) and for epochs with transitions (Fig. 5C). In epochs with transitions, coincidence of LDA with manual scoring was achieved if LDA scores corresponded to one of the two manually defined stages (for further explanations see Experimental procedures). Interestingly, in epochs with transitions (Fig. 5C) the coincidence of REM stages (R) was lowest depending little on epoch duration whereas in epochs without transitions (Fig. 5B) coincidence of waking (W) was low depending more strongly on epoch duration (*two-way ANOVA followed by Bonferroni post tests: $p<0.05$, all compared to 8 s). Coincidences in NREM epochs are similar for both conditions. Due to the occurrence of less than 8% of epochs with transitions (Fig. 5A) and the small effect on coincidence rates for waking and NREM sleep we excluded epochs with transitions from further performance measures (Fig. 6, explanations see Statistics and performance tests section in Experimental procedures). Sensitivity of LDA was highest (>95%) for REM and NREM sleep and lowest (>85%) for waking. Comparing LDA sensitivity values between different epoch durations, no differences for epochs from 6 to 10 s were found, except during waking (*two-way ANOVA followed by Bonferroni post tests: $p<0.05$, all compared to 8 s). Specificity was highest (>95%) for REM and waking and lowest (>85%) for NREM. There was no change with epoch duration except for 4 s in REM. Positive predictive value (PPV) was highest for waking (>95%) independently from epoch duration, intermediate for NREM (>80%) increasing significantly with epoch duration (*significances see above) and lowest for REM (>60% except for 4 s). PPV strongly depends on stage prevalence: REM occurs four times less frequently than NREM and W. It also depends on the number of false positives (see formula in Section 4.6 of Experimental procedures) which are relatively high for REM. Negative predictive value (NPV) was highest for REM (≈100%) and NREM (>95%) not depending on epoch duration (except 4 s in NREM). NPV was lowest in waking (>85%) decreasing significantly with epoch durations shorter than 8 s (*two-way ANOVA followed by Bonferroni post tests: $p<0.05$, all compared to 8 s).

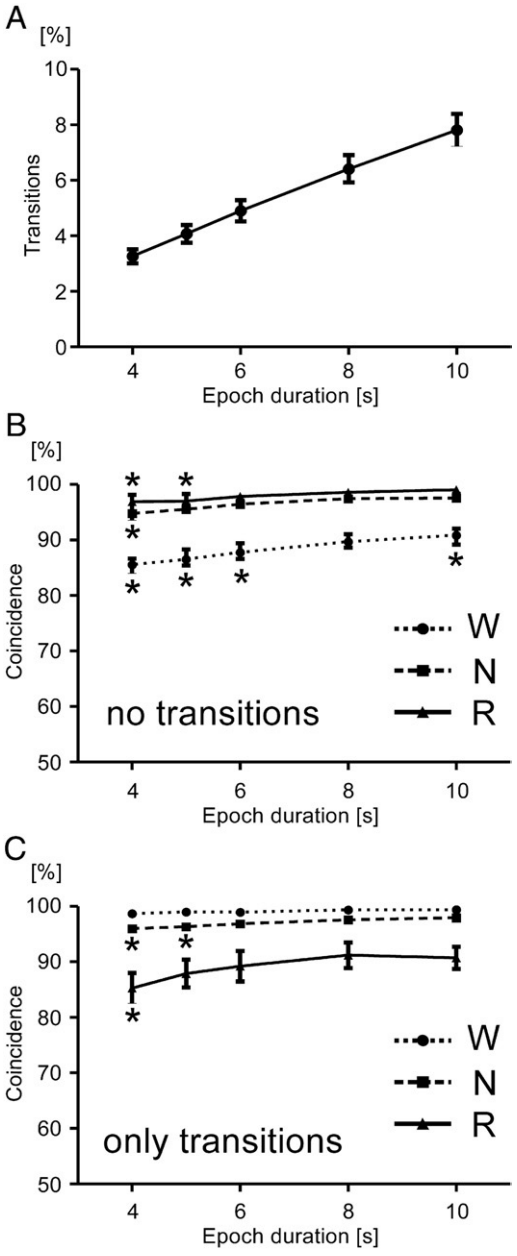

Fig. 5 – Semi-automated scoring with LDA (5% random training) and the effect of different epoch durations. A: The percentage of LDA epochs with more than one manual stage scoring (transitions) decreases with epoch duration from 7.8% (10 s) to 3.3% (4 s). B: Coincidences between semi-automated LDA and manual scoring for different durations of epochs excluding those epochs with transitions. C: The same as in B exclusively for epochs with transitions. LDA scoring corresponding to none of the two manual stages was considered as non-coincidence.

3. Discussion

In the present study, we performed long-term EEG recordings in mice within their home cages, using a miniaturized data logger. Animals exhibited the typical circadian sleep–wake rhythm as previously described for C57BL/6 mice (Huber et al., 2000; Veasey et al., 2000; Koehl et al., 2003). By systematically comparing manual sleep scoring with semi-automated methods of analysis, we found optimal results with the classifier Linear Discriminant Analysis (LDA) and epoch durations of 8 and 10 s. Highly reliable results were obtained after training the classifier with only 5% of manually scored data, randomly distributed over the entire dataset, thereby minimizing effort and time for data analysis. Decreasing epoch duration below 8 s lowered the scoring performance slightly. The automatic

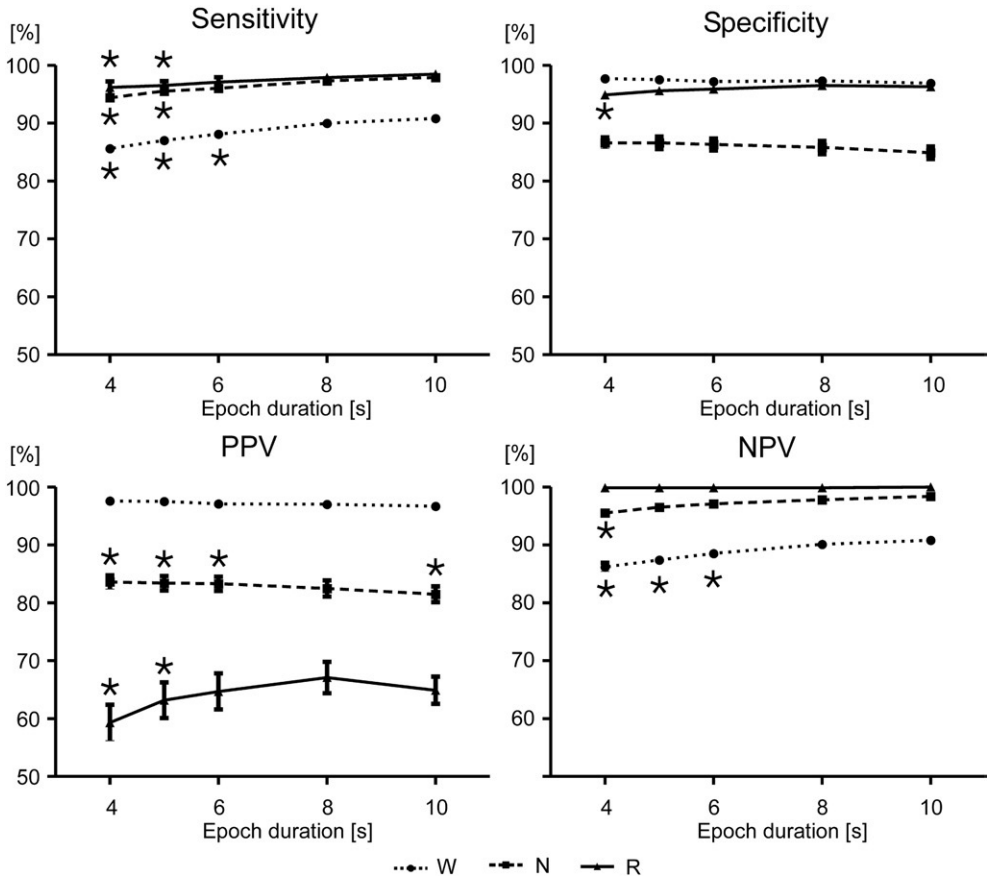

Fig. 6 – Performance measures of the semi-automated scoring with LDA (5% random training) and the effect of different epoch durations. Sensitivity, specificity and negative predictive value (NPV) are between 85 and 100% for all epoch durations and stages. Positive predictive value (PPV) depends on the prevalence of the stage, REM with about 25% prevalence compared to waking and NREM revealed the lowest PPV although it shows the highest sensitivity and NPV and a relatively high specificity. All performance parameters are highest for epochs of 8 and 10 s but change little if the epoch duration is lowered to 6 or even 5 s.

classifier LDA is superior to conventional formulas based on thresholds, because it saves the time required for manual adjustment of thresholds and, most importantly, because its reliability is comparable or higher to previous methods.

We also aimed at deriving optimal parameters for sleep–wake staging from the original EEG and EMG recordings. Following a fine separation of EEG frequencies into 10 bands, we found that estimation of integrated amplitudes with amplitude-period analysis was superior to the commonly used power spectra. It was also important to recruit parietal EEG recording which clearly contains hippocampal theta activity (for parietal cortex in rat see: Brankačk et al. (1996)) volume-conducted from dorsal hippocampus (Bland and Whishaw, 1976; Brankačk et al., 1993). We identified four parameters which were specifically well suited for stage classification. Conventionally used variables were i) integrated amplitude of neck muscle EMG (differentiating waking from sleep) and ii) delta frequency band (separating NREM from REM sleep). In addition, we found two new highly efficient parameters for differentiation between the two sleep stages, namely the amplitudes of the upper gamma frequency band (52–70 Hz) and of the upper theta band (7–8.5 Hz). Both upper gamma and upper theta frequencies were stronger during REM sleep as compared to NREM sleep. In contrast, the lower theta band (4.5–6.5 Hz) did not differentiate well between REM, NREM sleep and waking. Previous sleep–wake classifications did not separate between different theta frequencies (e.g. Maloney et al., 1997). Our data indicate that low and high frequency theta are differently involved in rhythmic brain activity during different sleep–wake states.

3.1. *Data logger versus cable tethering and telemetry*

As described in Experimental procedures, the total weight of the data logger including head implant and batteries for long (>30 h) EEG recordings was 4.6 g, more than 10% of the body weight. In spite of this relatively heavy burden the movement of the mice was not visibly altered. The animals still climbed regularly on lid bars of the home cage as they typically do. Unlike in larger animals, the behaviour of mice tends to be heavily influenced by cable tethering due to their smaller physical size, faster and more vertical movement and climbing. It has been reported that cable-based recordings in mice lead to severe movement restrictions as compared to implantable telemetry, due to twisting forces necessary for rotational movements (Tang and Sanford, 2002; Tang et al. 2004). On the other hand, intraperitoneal implantation of telemetric devices leads to substantial surgical trauma with

average mortality rates of 9% and significantly reduced horizontal and vertical movements after surgery (Tang and Sanford, 2002). In spite of its relatively heavy weight, we therefore found the data logger technique less invasive compared to intraperitoneal implantable telemetry and behaviourally less restrictive compared to cable recordings. The most important advantage of the data logger is that it allows for recording in the home cage. In some mouse strains, motor activity varies considerably between home cage and open field conditions (Tang and Sanford, 2002).

3.2. *Diurnal sleep–wake pattern in female C57BL/6 mice*

The overall percentages of sleep and wakefulness are very close to those reported from cable recordings in male C57BL/6 mice (Huber et al., 2000), despite some methodological differences (we used mainly females and data logger recording in the home cage). The sleep diurnal ratio reported by Huber et al. (2000) was 1.94 compared to 1.84 in our recordings, the wake diurnal ratio 0.54 compared to 0.55 and the REM diurnal ratio 2.69 compared to 2.77 in the present study. The percentage of REM sleep compared to total sleep time (10.7% of total sleep during dark and 15.9% during light) is also almost identical with that of Huber et al. (2000): 10.9% and 15.1%. In order to reduce variability between individual animals, we recorded mainly from female mice but for comparison used males. We found no differences between the seven female and three male mice. Koehl et al. (2006) reported some sex differences between sleep patterns of cable tethered C57BL/6 mice. Males slept more than females, with higher percentages of NREM during light and dark periods and more REM sleep during the light but not during the dark periods. Changes in sleep architecture during the estrous cycle are, however, very subtle (Koehl et al., 2003).

3.3. *EEG gamma frequencies and REM sleep*

High levels of gamma activity during REM sleep have been first reported by Parmeggiani and Zanocco (1963), later confirmed in humans (Itil, 1970; Llinás and Ribary, 1993; Mann et al., 1993) and rats (Franken et al., 1994). The frequency characteristics of the data logger permitted recordings with unaltered power up to 70 Hz, serving as our upper boundary for "high" gamma frequencies. Other definitions include frequencies reaching 100 Hz into the gamma range (Bragin et al., 1995; Csicsvari et al., 2003). Nevertheless we found that amplitudes of the gamma frequency band between 52 and 70 Hz were significantly larger in REM sleep compared to NREM, and similar between REM and waking. In rats Maloney et al. (1997) found increased gamma EEG at frequencies from 30.5 to 58 Hz in REM and W as compared to NREM. This frequency range corresponds mostly to our lower gamma frequency band (30.5 to 48 Hz) which, in our hands, showed less reliable correlations with waking and REM sleep compared to our upper gamma frequency range (52 to 70 Hz). Montgomery et al. (2008) found higher gamma synchrony in phasic and tonic REM sleep in the dentate gyrus (but not in CA1 and CA3), as compared to waking. Interestingly, in a recent study on human subjects, gamma-band EEG responses were found time locked to the onset of involuntary miniature eye movements (Yuval-Greenberg et al., 2008). It would be important to test for this relation during REM sleep in mice where miniature eye movements can be precisely measured (Oommen and Stahl, 2008).

3.4. *Power spectrum versus period-amplitude analysis*

Power spectral density function (simply: power spectrum) shows the strength of the variations of a time series as a function of frequency. It is a useful tool to identify oscillatory signals and their amplitude in time series data and is widely used in EEG analysis. With unvarying signal amplitude, the power spectrum delivers a good estimation of the frequency distribution. However problems arise with analysis of narrowband EEG activity if amplitudes change significantly over time. Due to the inherent averaging process in power spectrum generation (Ktonas and Gosalia, 1981), such spectra cannot differentiate between low-amplitude-high incidence and high-amplitude-low incidence EEG activity. A different approach is the period-amplitude analysis (PAA) which separates frequency incidence from amplitude and has been successfully used by various groups (Roessler et al., 1970; Feinberg et al., 1978; Ktonas and Gosalia, 1981; Uchida et al., 1999). Using PAA in rats, Campbell and Feinberg (1993) found parallel changes in amplitude and incidence during the light period but differences between both parameters across the dark period. Our present data confirm that PAA "seems to offer more resolution than the power spectrum in detecting electrographic details in amplitude and incidence within relatively narrow frequency bands" (Ktonas and Gosalia, 1981). It may also agree more closely with visual EEG analysis than power spectra (Schenk, 1976). This would explain the superior results with LDA in differentiating sleep and waking using PAA-derived amplitudes as compared to power spectra. The deficiencies of PAA in earlier studies have been corrected in the eighties by adding appropriate band pass filters (Ktonas, 1987). Without band pass, PAA can fail to detect the presence of fast EEG waves riding on top of large-amplitude slow activity. Appropriate band pass filtering before application of PAA ensures that both fast and slow waves are detected. Signal distortion due to a too narrow band pass filter can be avoided by making the band pass broader than the analysed frequency range (Ktonas, 1987; see Table 1).

3.5. *Choosing the optimal parameters*

Combining systematic and empirical methods, we found an optimal combination of parameters for semi-automated sleep scoring. The use of the integrated amplitude of high pass filtered neck muscle EMG as a sleep–wake scoring parameter is generally accepted and does not need further explanations. However, recently it has been shown that EMG can be substituted by microelectro-mechanical system (MEMS) accelerometers measuring the animal's movement (Sunderam et al., 2007). Here, we found a strong similarity of higher frequency gamma amplitudes with the EMG. Potentially, this correlation may be used to differentiate waking from NREM sleep within one-channel EEG recordings. Many sleep scoring methods use power of delta and theta bands or a ratio between both (Robert et al., 1999). In agreement with earlier reports we found integrated amplitudes of delta half-waves superior to delta band power (Ruigt et al., 1989; Robert et al., 1999). A new

finding of the present analysis is, however, that the upper theta band (7–8.5 Hz) is better suited to separate between NREM and REM sleep than the entire theta band. We also found that amplitudes of our upper gamma-band are useful for differentiation between NREM and REM sleep. In contrast, ratios of theta and delta power as well as theta and delta amplitudes did not improve the performance of supervised classifiers (although they worked well with threshold formulas). In contrast to our empirical approach, the attempt to use PCA for systematic parameter reduction and selection of the strongest stage indicators was not successful in the context of supervised scoring algorithms (but see: Jobert et al. (1994) and Martinez and Kak (2001)).

3.6. *Sleep scoring with linear discriminant analysis*

LDA (Molinari et al., 1984; Sunderam et al., 2007) and Tree (Hanaoka et al., 2001) and combinations of both (Anderer et al., 2005) have been used previously for sleep–wake scoring in humans. To our knowledge, however, they have not been adapted for sleep scoring in mice or rats. The advantage of both supervised classifiers over threshold formulas is their flexibility and time-saving applicability. The number of classes and their characteristics are predefined by manual scoring of small portions of the dataset which serve as training data. With LDA, there is no need for extensive data examination and threshold setting. The sleep scoring performance of LDA was stable even with very low amounts of training data. Using the Classification Tree algorithm, in contrast, scoring performance decreased already with 50% of the full data being used as training set. For systematic comparison and averaging of data from seven mice we used 5% of manually scored epochs as training data, randomly distributed over the entire dataset. Similar methods have been previously used for object recognition (Martinez and Kak, 2001).

The number of transitions, i.e. epochs with more than one manually scored stage due to transition points between stages, decreased linearly with epoch duration from (<8% at 10 s to ≈3% at 4 s) with little impact on LDA performance. Coincidence of LDA to manual scoring was similar for transition epochs compared to epochs without transitions although affecting differently waking and REM. Varying epoch duration down to 6 s had no or little effects on LDA scoring performance. The positive predictive value or precision rate for REM was surprisingly low in contrast to the very high sensitivity, specificity and negative predictive value. This seems a disadvantage of the present method requiring additional semi-automatic or manual error corrections which are however less time-consuming for REM as for the other two stages.

Based basically on the same set of parameters commonly used with other scoring methods, other recording techniques or other rodent species, we are confident that our LDA method will perform similarly well under different conditions. Nevertheless LDA should be tested in the future with cable recording techniques commonly used in most laboratories and with other species of animals. We consider the semi-automatic LDA method described here as one of several steps, followed by error correction routines, e.g. eliminating false positive REM epochs immediately following waking. Additional microelectro-mechanical system accelerators for movement detection may advance semi-automatic sleep–wake scoring in mice.

4. Experimental procedures

4.1. *Animal care and housing conditions*

This study was carried out in accordance with the European Science Foundation Policy on the Use of Animals in Research (2001), the U.S. National Institutes of Health Guide for the Care and Use of Laboratory Animals (1996) and has been approved by the Governmental Supervisory Panel on Animal Experiments of Baden Württemberg at Karlsruhe (35-9185.81/G-30/08).

Either 28 or 45 day old female and male C57/Bl6 mice were purchased from Charles River (Germany). They were housed in groups of four to five inside a ventilated Scantrainer (Scanbur BK A/S Denmark) on an inverted 12/12-h light/dark cycle with light on from 20:00 to 8:00 and free access to water and food. After electrode implantation, animals were housed individually throughout the experiment. Feeding and cleaning was done outside the recording time of 72 h which started on Monday and terminated on Thursday, both between 3:30 and 4:30 p.m.

4.2. *Animal preparation*

Eight female and three male C57/Bl6 mice (26–33 g, 87–210 days old) were anesthetized with isoflurane in medical oxygen (5% isoflurane for induction, 1.5–2.5% for maintenance, flow rate: 1 l per min). Anesthetized animals were placed in a stereotactic apparatus with a custom-made inhalation tube. For analgesia, 0.1 mg/kg of buprenorphin was injected subcutaneously prior to and 8 h after surgery. After exposure of the skull bone, three gold plated brass watch screws were fixed permanently into the skull, one over the lateral parietal association cortex (2 mm posterior of bregma, 1.5 mm lateral to the midline) and a second over the primary motor cortex (2 mm anterior of bregma, 1.5 mm lateral to the midline). A third screw over the cerebellum served as ground and reference electrode. In mice, the lateral parietal cortex is overlaying the dorsal hippocampus and permits reliable recordings of theta rhythm with comparable amplitudes and phases between different animals. Three varnish-insulated nichrome wires (100 μm, glued together) where inserted into the neck muscle for EMG recording, one of which served as EMG reference electrode.

4.3. *Electrophysiology*

One week after surgery, continuous monopolar electroencephalographic (EEG) recording began with a 24 h test session followed one week later by a 72 h recording session. Recordings were performed in the animal's home cage by using a miniaturized data logger (Neurologger 2) (Vyssotski et al., 2009). The data logger recorded 4 channels with analogue filtering and amplifying cascades, a microcontroller for A/D-conversion and data management, and a 256 MB onboard memory chip. The dimensions of the logger were 22×14×8 mm including two hearing aid batteries. When

equipped with long lasting batteries for >5 day recording time (Renata Z13; Itingen, Switzerland) the logger weight was 3.8 g. With short-lasting batteries (>1 day; Renata ZA 10; Itingen, Switzerland) the weight was <2.8 g. The head implant (electrodes, wires, contacts and dental acrylic) had a weight of 0.8 g. We recorded two channels of EEG and two channels of EMG recordings at ×2000 amplification (input range ±500 μV). Data were band pass filtered (1 Hz to 70 Hz, –6 dB/octave), sequentially digitized (1600 or 800 samples per second), and stored on the memory chip at sampling rates of 400 or 200 Hz per channel. Maximal continuous recording times were 31 h with four channels at 400 Hz sampling/channel and about 75 h with three channels each at 200 Hz sampling rate. At the end of the experiment, the data was downloaded from the data logger onto a personal computer for further analysis. Recordings were completed in ten out of eleven animals.

4.4. *Data analysis*

The continuous 72 h EEG recordings from seven female and three male animals were imported into a custom-made program based on MATLAB (The Mathworks Inc., Natick, MA). The neck muscle EMG activity was high pass filtered (30–70 Hz). Visual classification of different sleep stages (NREM and REM) or wakefulness (W) was based on: 1) the level of EMG activity (waking [W]>NREM>REM); 2) the amount of high-amplitude-low frequency delta activity in the neocortex (NREM>W and REM; and 3) the amount of regular theta oscillations in the lateral parietal cortex overlaying the dorsal hippocampus (REM and active waking>quiet waking and NREM). Examples of EEG and EMG activities during transitions between W, NREM and REM are shown in Fig. 1A. Waking, NREM and REM sleep were manually scored by placing markers at the transition boundaries between those three stages (see vertical lines in Fig. 1A) and not by using fixed epoch length as in semi-automated scoring. Therefore manual scoring could yield stage durations smaller than the epoch durations used in semi-automated scoring (see Figs. 1B to D for minimal, median and maximal durations of manually scored stages). In addition durations of manually scored stages normally do not equal exactly to multitudes of epoch durations used in semi-automated scoring. Both can lead to epochs which contain transition points between manually scored stages. This may decrease coincidence between manual and semi-automated scoring depending on the epoch duration which was tested (see Fig. 5).

Manually marked time points of such stage transitions were used for evaluation of duration and intervals of individual stages and for the total amount and relative percentage of W, NREM and REM (see Fig. 4). We used two different methods to analyse the EEG content within ten different frequency bands (for limits of these bands see Table 1: 1) power spectral density functions (FFT, Hanning-window, 512 data points, mean band power and power of the dominant frequency); and 2) band pass filtered half-wave period-amplitude analysis (PAA: number of baseline crossings per second, mean duration, integrated amplitude and maximal amplitude of the half-waves). For details of the PAA see Feinberg et al. (1978). Shortly, the half-wave period is calculated as duration from one baseline crossing to the next. The integrated amplitude is an estimation of the area between the half-wave and the baseline. Signal distortions and contamination with superimposed fast rhythms were avoided by appropriate band pass filtering prior to analysis (Ktonas, 1987; parameters see Table 1). Duration and integrated amplitude of each half-wave were averaged over epochs. Table 1 lists the analysed frequency bands and band pass filters applied for period-amplitude analysis (PAA).

Spectral power density functions were used to calculate band power (i.e. mean power of all frequencies within the band) and power maximum (i.e. power of the dominant frequency) for each of the ten frequency bands. We further calculated the ratio of theta band (4.5–8.5 Hz) power divided by delta band (1–4 Hz) power which is frequently used in sleep scoring, including commercial programs based on threshold formulas. From band pass filtered PAA data the following four parameters were calculated for each of the ten frequency bands: 1) number of baseline crossings per second; 2) duration between two consecutive baseline crossings; 3) integrated amplitude of half-waves; and 4) maximal amplitude of half-waves. We also calculated the ratio between integrated half wave-amplitudes in the theta band and in the delta range, respectively. Further, we calculated a parameter for regularity of rhythmic activity for each of the ten frequency bands. This rhythm score was computed as the number of baseline crossings per second multiplied by the interval between consecutive zero crossings. Oscillations with stable frequency during the entire epoch length would lead to a rhythm score of 1. The integrated amplitude of the neck muscle EMG completed the set of parameters, resulting in a total of 73 variables. We tried to select candidate parameters with maximal contribution to variability by applying principal component analysis (PCA). However, it turned out that these parameters did not have the highest discriminative power for the supervised classifiers (see below).

4.5. *Semi-automated sleep scoring*

The following highly efficient set of parameters for supervised classifiers (LDA and Tree) was identified by empirical selection: integrated amplitudes from higher gamma EEG frequency band (1), delta EEG (2), higher theta EEG band (3), and amplitude of neck muscle EMG (4). Adding more parameters or reducing to three or less parameters decreased the performance of sleep scoring. Five different epoch durations were used: 4, 5, 6, 8 and 10 s with 8 and 10 s yielding the best performance (Fig. 5).

Sleep scoring was further improved by artefact rejection. We discarded epochs with high-amplitude movement artefacts in the delta frequency range by using thresholds for delta band power and for delta amplitudes larger than those observed during NREM sleep. We used three different methods for semi-automated sleep scoring, the supervised classifiers LDA and Tree and a conventional threshold formula commonly used in commercial sleep scoring programs. The success of sleep scoring was measured as the percentage of coincident epoch classifications (W, NREM, REM) between manual and semi-automated scoring. In addition, the number of false positive and false negative epoch classifications was

calculated for each stage. The total percentage of errors was then calculated as the sum of these two errors divided by the total number of manually defined epochs for each particular stage (see also Section 4.6 for additional performance measures).

4.5.1. *Linear discriminant analysis*

Similar to principal component analysis (PCA), LDA performs dimensionality reduction. However, in contrast to PCA it preserves as much of the class discriminatory information as possible (Fielding, 2006). For any given data set, the ratio of between-class variances over within-class variances is maximized, resulting in an optimal discrimination area between the classes. Original data were divided into 4, 5, 6, 8 or 10 s epochs. For each epoch four empirically chosen EEG/EMG parameters were calculated as described above. The training data consisted of manually scored epochs for each of the three stages, randomly distributed over the entire dataset. After training LDA classified the epochs into waking, NREM or REM according to the parameter values for each epoch. For further details on LDA see MacLachlan (1992) and Fielding (2006).

LDA was applied according to the MATLAB's Statistical Toolbox™ function: "classify" (Statistics Toolbox, The Math Works Inc., 2009, page 11–13). The function "classify" is used in the following example code:

AutoStages = *classify*(*AnalysedParams*, *TrainingParams*, *TrainingStages*)

where *AnalysedParams* contains an array of selected parameters (e.g. integrated amplitude of delta band, EMG etc.) for each epoch into which the dataset was divided. Each row of the array corresponds to one epoch, each column to one of the selected parameters. *TrainingParams* contains an array of training epochs with the corresponding set of parameters. The epochs for each of the stages were randomly chosen throughout the entire dataset of 72 h. The arrays *TrainingParams* and *AnalysedParams* have the same number of columns. Corresponding columns of both arrays correspond to the same kind of parameter. *TrainingStages* contains an array of strings with the stage codes (W: waking, N: NREM, R: REM) for the corresponding rows of the array *TrainingParams*. Both arrays have the same number of rows. *AutoStages* contains an array of strings with stage codes for the corresponding rows of the array *AnalysedParams*. The latter two arrays have the same number of rows. The function "classify" classifies the rows of the array *AnalysedParams* into groups, based on the grouping of the rows in the array *TrainingParams*.

The MATLAB™ program code for LDA is available by request.

4.5.2. *Classification Tree*

A different type of supervised classifier is the Classification Tree (Tree) which is more flexible compared to LDA. This method uses a hierarchy of predictions to sort a particular epoch into given classes using training sets of manually classified stages. For a detailed description of the method see Fielding (2006) and Safavian and Landgrebe (1991). For training of automated scoring algorithms, we used sets of manually predefined epochs. In order to reveal maximal performance of Tree or LDA, respectively, one set of analyses was done after feeding in the entire dataset (100%). This procedure is, of course, useless in real applications. We therefore repeated the analysis with 5% of the original data, taken from randomly distributed sections. Classification Tree was used according to the MATLAB's Statistical Toolbox™ functions: treefit, treeval (Statistics Toolbox, The Math Works Inc., 2009, page 11–19). The function "treefit" is used in the following example code:

ClassificationTree = *treefit*(*TrainingParams*, *TrainingStages*)

where the arrays *TrainingParams* and *TrainingStages* are identical to the arrays with the same name in LDA (Section 4.5.1). *ClassificationTree* is a binary decision tree where each non-terminal node is split based on the values of a column of the array *TrainingParams*. This decision tree is used by the function "treeval" for classification of the array *AnalyseParams* (array identical to LDA, Section 4.5.1).

The function "treeval" is used in the following example code:

AutoStages = *treeval*(*ClassificationTree*, *AnalysedParams*)

Function "treeval" uses the decision tree *ClassificationTree* and the array *AnalysedParams* of predictor values to produce a string array *AutoStages* (identical to arrays for LDA, Section 4.5.1) of predicted response values.

The MATLAB™ program code for Tree is available by request.

4.5.3. *Threshold formulas*

In contrast to supervised and trained classifiers, threshold formulas use predefined settings for each parameter to differentiate between stages. The following formula was used for sleep scoring: "if EMGIntegral>X, Stage='W';, else, if Ratio<Y, Stage='N';, else, Stage='R'; end, end" where X was a threshold of neck muscle EMG (integrated amplitude) and Y corresponded to a threshold value of the ratio between theta and delta powers. While X separated wake states from both different sleep types, Y distinguished between REM (R) and NREM (N) sleep. Both thresholds had to be defined for each dataset and animal.

4.6. *Statistics and performance tests*

Data are given as means and standard error of the means (SEM) calculated from medians or means of individual animals depending on sample distribution. For group comparisons we used two-way ANOVA followed by Bonferroni post tests as well as paired and unpaired *T*-tests as stated in the Results section and Legends.

The classification performance of semi-automatic scoring (LDA with 5% training data, randomly distributed over the dataset) for each of the stages and for different epoch durations was measured with the following statistical methods: sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). The following parameters were used: 1. number of true positives (=coincidences between automatic and manual scoring), 2. number of true negatives (=correct rejections), 3. number of false positives (= incorrectly identified as the particular stage, Type I error) and 4. number of false negatives (= missed cases, incorrectly not identified as the particular stage, Type II error). Sensitivity (also called power or

recall rate) measures the proportion of actual positives which are correctly identified as such:

$$Sensitivity = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ negatives}.$$

Specificity measures the proportion of negatives which are correctly identified:

$$Specificity = \frac{number\ of\ true\ negatives}{number\ of\ true\ negatives + number\ of\ false\ positives}.$$

Positive predictive value (PPV, also known as precision rate) is the proportion of true positives (correctly identified) to all positives (correctly and wrongly identified as the particular stage):

$$PPV = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ positives}.$$

The value of PPV depends strongly on the prevalence of the stage which is four times lower for REM compared to both, W and NREM (see Fig. 1E).

Negative predictive value (NPV) is the proportion of stages with negative results which are correctly identified:

$$NPV = \frac{number\ of\ true\ negatives}{number\ of\ true\ negatives + number\ of\ false\ negatives}.$$

For further details see Campbell et al. (2007).

Acknowledgments

We would like to thank Mrs. Nadine Zuber and Mrs. Cornelia Heuser for their technical assistance. We appreciate helpful discussions with Dr. Thomas Künsting and Mr. Reinhold Wojciechowski. The study was supported by the DFG (SFB 636/B06).

REFERENCES

Anderer, P., Gruber, G., Parapatics, S., Woertz, M., Miazhynskaia, T., Klösch, G., et al., 2005. An e-health solution for automatic sleep classification according to Rechtschaffen and Kales: validation study of the somnolyzer 24×7 utilizing the Siesta database. Neuropsychobiology 51, 115–133.

Axmacher, N., Draguhn, A., Elger, C.E., Fell, J., 2009. Memory processes during sleep: beyond the standard consolidation theory. Cell. Mol. Life Sci. 66, 2285–2297.

Bland, B.H., Whishaw, I.Q., 1976. Generators and topography of hippocampal theta (RSA) in the anaesthetized and freely moving rat. Brain Res. 118, 259–280.

Bragin, A., Jandó, G., Nadasdy, Z., Hetke, J., Wise, K., Buzsáki, G., 1995. Gamma (40–100 Hz) oscillations in the hippocampus of the behaving rat. J. Neurosci. 15, 47–60.

Brankačk, J., Buzsáki, G., 1986. Hippocampal responses evoked by tooth pulp and acoustic stimulation: depth profiles and effect of behavior. Brain Res. 378, 303–314.

Brankačk, J., Stewart, M., Fox, S.E., 1993. Current source density analysis of the hippocampal theta rhythm: associated sustained potentials and candidate synaptic generators. Brain Res. 615, 310–327.

Brankačk, J., Seidenbecher, T., Müller-Gärtner, H.W., 1996. Task-relevant late positive component in rats: is it related to hippocampal theta rhythm? Hippocampus 6, 475–482.

Brankačk, J., Platt, B., Riedel, G., 2009. Sleep and hippocampus: do we search for the right things? Prog. Neuropsychopharmacol. Biol. Psychiatry 33, 806–812.

Campbell, I.G., Feinberg, I., 1993. Dissociation of delta EEG amplitude and incidence in rat NREM sleep. Brain Res. Bull. 30, 143–147.

Campbell, M.J., Machin, D., Walters, S.J., 2007. Medical Statistics. A Textbook for the Health Sciences. Wiley, Chichester.

Crisler, S., Morrissey, M.J., Anch, A.M., Barnett, D.W., 2008. Sleep-stage scoring in the rat using a support vector machine. J. Neurosci. Methods 168, 524–534.

Csicsvari, J., Jamieson, B., Wise, K.D., Buzsáki, G., 2003. Mechanisms of gamma oscillations in the hippocampus of the behaving rat. Neuron 37, 311–322.

Feinberg, I., March, J.D., Fein, G., Floyd, T.C., Walker, J.M., Price, L., 1978. Period and amplitude analysis of 0.5–3 c/sec activity in NREM sleep of young adults. Electroenceph. Clin. Neurophysiol. 44, 202–213.

Fielding, A.H., 2006. Cluster and Classification Techniques for the Biosciences. Cambridge University Press, Cambridge.

Franken, P., Dijk, D.-J., Tobler, I., Borbely, A.A., 1994. High-frequency components of the rat electrocorticogram are modulated by the vigilance states. Neurosci. Lett. 167, 89–92.

Guide for the Care and Use of Laboratory Animals, 1996. Institute of Laboratory Animal Research, Commission on Life Sciences. National Research Council National Academies Press, Washington, D.C.

Hanaoka, M., Kobayashi, M., Yamazaki, H., 2001. Automated sleep stage scoring by decision tree learning. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2, 1751–1754.

Huber, R., Deboer, T., Tobler, I., 2000. Effect of sleep deprivation on sleep and sleep EEG in three mouse strains: empirical data and simulations. Brain Res. 857, 8–19.

Itil, T.M., 1970. Digital computer analysis of the electroencephalogram during rapid eye movement sleep state in man. J. Nerv. Ment. Dis. 150, 201–208.

Jobert, M., Escola, H., Poiseau, E., Gaillard, P., 1994. Automatic analysis of sleep using two parameters based on principal component analysis of electroencephalography spectral data. Biol. Cybern. 71, 197–207.

Koehl, M., Battle, S.E., Turek, F.W., 2003. Sleep in female mice: a strain comparison across the estrous cycle. Sleep 26, 267–272.

Koehl, M., Battle, S., Meerlo, P., 2006. Sex differences in sleep: the response to sleep deprivation and restraint stress in mice. Sleep 29, 1224–1231.

Ktonas, P.Y., 1987. Editorial comment: period-amplitude EEG analysis. Sleep 10, 505–507.

Ktonas, P.Y., Gosalia, A.P., 1981. Spectral analysis vs. period-amplitude analysis of narrowband EEG activity: a comparison based on the sleep delta-frequency band. Sleep 4, 193–206.

Llinás, R., Ribary, U., 1993. Coherent 40-Hz oscillation characterizes dream state in humans. Proc. Natl. Acad. Sci. U.S.A. 90, 2078–2081.

MacLachlan, G.J., 1992. Discriminant Analysis and Statistical Pattern Recognition. Wiley, New York.

Maloney, K.J., Cape, E.G., Gotman, J., Jones, B.E., 1997. High-frequency γ electroencephalogram activity in association with sleep–wake states and spontaneous behaviours in the rat. Neuroscience 76, 541–555.

Mann, K., Bäcker, P., Röschke, J., 1993. Dynamical properties of the sleep EEG in different frequency bands. Int. J. Neurosci. 73, 161–169.

Martinez, A.M., Kak, A.C., 2001. PCA versus LDA. IEEE Trans. Pattern Anal. Mach. Intell. 23, 228–233.

Mochizuki, T., Crocker, A., McCormack, S., Yanagisawa, M., Sakurai, T., Scammell, T.E., 2004. Behavioral state instability in orexin knock-out mice. J. Neurosci. 24, 6291–6300.

Molinari, L., Dumermuth, G., Lange, B., 1984. EEG-based multivariate statistical analysis of sleep stages Neuropsychobiology 11, 140–148.

Montgomery, S.M., Sirota, A., Buzsáki, G., 2008. Theta and gamma coordination of hippocampal networks during waking and rapid eye movement sleep. J. Neurosci. 28, 6731–6741.

Oommen, B.S., Stahl, J.S., 2008. Eye orientation during static tilts and its relationship to spontaneous head pitch in the laboratory mouse. Brain Res. 1193, 57–66.

Pang, D.S.J., Robledo, C.J., Carr, D.C., Gent, T.C., Vyssotski, A.L., Caley, A., Zecharia, A.Y., Wisden, W., Brickley, S.G., Frank, N.P., 2009. An unexpected role for TASK-3 potassium channels in network oscillations with implications for sleep mechanisms and anesthetic action. Proc. Natl. Acad. Sci. USA 106, 17546–17551.

Parmeggiani, P.L., Zanocco, G., 1963. A study of the bioelectrical rhythms of cortical and subcortical structures during activated sleep. Arch. Ital. Biol. 101, 385–412.

Robert, C., Guilpin, C., Limoge, A., 1999. Automated sleep staging systems in rats. J. Neurosci. Methods 88, 111–122.

Roessler, R., Collins, F., Ostman, R., 1970. A period analysis classification of sleep stages. Electroenceph. Clin. Neurophysiol. 29, 358–362.

Ruigt, G.S.F., Van Prossdij, J.N., Van Delft, A.M.L., 1989. A large scale, high resolution, automated system for rat sleep staging. I. Methodology and technical aspects. Electroenceph. Clin. Neurophysiol. 73, 52–63.

Safavian, S.R., Landgrebe, D., 1991. A survey of decision tree classifier methodology. IEEE Trans. Syst. Man Cybern. 21, 660–674.

Schenk, G.K., 1976. The pattern-oriented aspect of EEG quantification. Model and clinical basis of the iterative time-domain approach. In: Kellaway, P., Petersén, I. (Eds.), Quantitative Analytic Studies in Epilepsy. Raven Press, New York, pp. 431–461.

Statistics Toolbox™, 2009. User's Guide. Version 7. The MathWorks, Inc.

Sunderam, S., Cherny, N., Peixoto, N., Mason, J.P., Weinstein, S.L., Schiff, S.J., Gluckman, B.J., 2007. Improved sleep–wake and behaviour discrimination using MEMS accelerometers. J. Neurosci. Methods 163, 373–383.

Tang, X., Sanford, L.D., 2002. Telemetric recording of sleep and home cage activity in mice. Sleep 25, 677–685.

Tang, X., Orchard, S.M., Liu, X., Sanford, L.D., 2004. Effect of varying recording cable weight and flexibility on activity and sleep in mice. Sleep 27, 803–810.

Timo-Iara, C., Negrao, N., Schmidek, W.R., Hoshino, K., Lobato de Menezes, C.E., Leme da Rocha, T., 1970. Phases and states of sleep in the rat. Physiol. Behav. 5, 1057–1062.

Uchida, S., Feinberg, I., March, J.D., Atsumi, Y., Maloney, T., 1999. A comparison of period amplitude analysis and FFT power spectral analysis of all-night human sleep EEG. Physiol. Behav. 67, 121–131.

Use of animals in research (2. edition). 2001. European Science Foundation Policy Briefings, ISRN ESF-SPB-01-15, Strasbourg, 1-6.

Veasey, S.C., Valladares, O., Fenik, P., Kapfhamer, D., Sanford, L., Benington, J., Bucan, M., 2000. An automated system for recording and analysis of sleep in mice. Sleep 23, 1–16.

Vyssotski, A.L., Dell'Omo, G., Dell'Ariccia, G., Abramchuk, A.N., Serkov, A.N., Latanov, A.V., Loizzo, A., Wolfer, D.P., Lipp, H.-P., 2009. EEG responses to visual landmarks in flying pigeons. Curr. Biol. 19, 1159–1166.

Walker, M.P., 2009. The role of sleep in cognition and emotion. Ann. N.Y. Acad. Sci. 1156, 168–197.

Yuval-Greenberg, S., Tomer, O., Keren, A.S., Nelken, I., Deouell, L.Y., 2008. Transient induced gamma-band response in EEG as a manifestation of miniature saccades. Neuron 58, 429–441.