DOI 10.1007/s11055-016-0369-8
Neuroscience and Behavioral Physiology, Vol. 47, No. 1, January, 2017

Dependence of the Accuracy of Automatic Identification of Sleep and Waking States in Mice on the Spectral Characteristics of the Electroencephalogram

A. I. Manolov,[1] V. M. Koval'zon,[2] Yu. V. Ukraintseva,[1] L. S. Moiseenko,[2] and V. B. Dorokhov[1]

UDC 612.821.6

Translated from Zhurnal Vysshei Nervnoi Deyatel'nosti imeni I. P. Pavlova, Vol. 65, No. 5, pp. 635–640, September–October, 2015. Original article submitted November 26, 2014. Accepted March 4, 2015.

Computer programs for the automatic analysis of the electroencephalogram (EEG) from humans and animals have entered wide use and are successfully employed in many areas of physiological research. They are of particular value for sleep studies, as traditional expert analysis of polysomnograms (PSG) is very laborious. The aim of the present work was to investigate the relationship between the accuracy of automatic staging and the EEG spectral parameters characterizing activity specific for the waking and sleep states. This relationship can be used as an objective measure of the quality of PSG traces, i.e., the extent of signs providing for the identification and differentiation of waking and the various sleep phases. We found a statistically significant relationship (including linear) between the accuracy of automatic staging and various spectral EEG characteristics in mice. This approach to automated PSG analysis provides objective criteria for trace quality with a priori assessment of the accuracy of automatic staging.

Keywords: EEG, automatic analysis, polysomnography, mice.

Analysis of brain electrical activity is important in various areas of neurobiological research. It has a particular role in somnology, where it is one of only a small number of objective measures allowing calm waking to be discriminated from sleep and sleep structure, i.e., the substitution of phases and stages over time, to be identified. The appearance of digital polysomnographs means that paper polysomnograph (PSG) recordings are no longer needed, though investigators still come up against the need for visual analysis of digital PSG recordings, "manually" extracting epochs of waking and the various phases and stages of sleep – a process termed staging.

Although many algorithms achieving approximately 90% accuracy between the results of automated staging and expert assessments have been developed [Veasey et al., 2000], we have identified a number of significant gaps in automatic PSG staging in laboratory animals. Firstly, there is only a small number of investigations comparing the accuracies of different algorithms. Secondly, in methodological articles discussing sleep staging algorithms, all traces are assumed to be of adequate quality for automatic staging. Trace selection criteria are not given and there are no comparisons of the effectivenesses of the different algorithms on traces of different quality. Trace quality refers to the extent to which features characteristic of waking and the different stages of sleep are expressed. Thus, slow-wave sleep is characterized by the presence of high-amplitude, low-frequency (<4 Hz) waves on the EEG; REM sleep is characterized by a marked hippocampal θ rhythm (5–8 Hz) on the background of an absence of motor activity; waking is typified by a low-amplitude, desynchronized EEG and a high level of motor activity. In practice, investigators have to deal with the findings of extensive interindividual variability in electroencephalograms (EEG), including the main

[1] Institute of Higher Nervous Activity and Neurophysiology, Russian Academy of Sciences, Moscow, Russia.
[2] Institute of Problems in Ecology and Evolution, Russian Academy of Sciences, Moscow, Russia; e-mail: paraslonic@gmail.com.

markers of sleep and waking (both in humans [Huiskamp, 2008; Sannita et al., 1999; Buckelmüller et al., 2006] and rodents [Veasey et al., 2000]). In addition, trace quality can suffer from various artifacts which are unavoidable after chronic implantation of electrodes in animals and nighttime recording of the PSG in humans.

The present study addressed the following tasks:

1) to select an algorithm allowing automatic analysis of the EEG to identify and differentiate waking and the individual sleep phases in the PSG in mice;

2) to evaluate the accuracy of automatic staging using this algorithm;

3) to study the relationship between the accuracy of staging and the extents of a number of EEG spectral features characterizing activity specific for the states of waking and sleep.

Methods

Animal study protocols were approved by the ethics committee of the Institute of Problems in Ecology and Evolution, Russian Academy of Sciences, where the experiments were performed.

A total of 22 C57Bl/6 mice aged 2.5–3 months and weighing 25–30 g were anesthetized with Avertin and underwent epidural implantation of EEG recording electrodes in the frontal and parietal neocortex. Animals were placed in individual soundproofed boxes with a constant 12/12 light regime (09:00–21:00 white light (150 lx) and 21:00–09:00 weak red light (15 lx)). After a one-week post-operative recovery period, day-round recording of the polysomnogram was started, including two EEG channels and recording of a motor activity mechanogram and video recording of the animals' behavior. Each animal was connected to the input of a miniature digital biopotentials amplifier (DBA, size 30 × 25 × 4 mm and weight 5 g) containing an accelerometer via a flexible cable and rotating connector attached to the chamber ceiling and not restricting freedom of movement. This construction allowed the DBA plate to move freely in three planes for recording of the mechanogram of even small movements by the mouse. The EEG was recorded with a sampling frequency of 250 Hz and motor activity at 50 Hz. PSG analysis epochs lasted 20 sec.

As the tasks addressed here required inclusion of traces known to be of different quality, PSG obtained on the first day of were analyzed, these generally being good-quality traces, along with recordings made on day 7, these being of worse quality, containing more artifacts and a lower signal:noise ratio. In six mice, PSG recorded on day 7 were excluded from analysis because they contained too many artifacts. Overall, 38 12-h recordings were analyzed, each of which was divided into 2160 12-sec epochs. Each trace was staged by extracting epochs of waking, slow-wave sleep, and REM sleep. A total of 35 of these PSG were staged by a single expert and a further three were staged by two independent experts. All 38 PSG were also staged using an automatic staging algorithm. Staging was performed using one of two EEG channels, whichever showed the markers of waking and the individual sleep stages more clearly (see description of trace quality indicators, above). The accuracy of automatic staging was assessed using both channels when present.

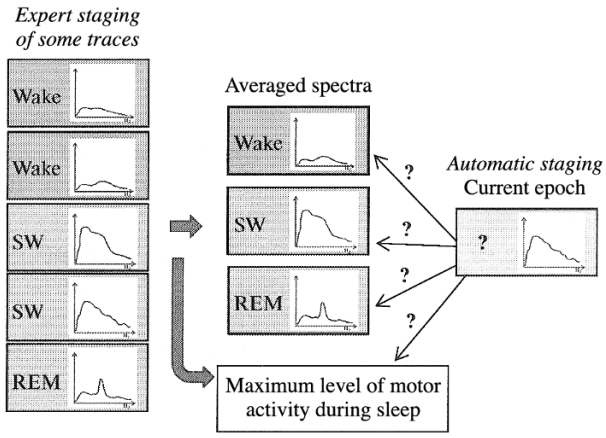

Fig. 1. Simplified scheme of the automatic staging algorithm. See text for explanation.

Description of the automatic staging algorithm. We selected a simple and reliable staging algorithm (as demonstrated in a comparative study of human EEG recordings [Becq et al., 2005]), based on spectral characteristics and comparison of distances between the spectrum of the ongoing epoch and the averaged spectral characteristics of typical traces for epochs of waking, slow-wave sleep, and REM sleep.

Figure 1 shows a simplified diagram of the process of semiautomatic staging, consisting of two stages: 1) expert staging of a small number of epochs (30 20-sec epochs for slow-wave sleep and waking and 10 epochs for REM sleep, which was used as a training set; 2) automatic staging of the remaining traces on the basis of features obtained from the training set.

The automatic staging algorithm was as follows:

1. Collection of information on the training set:

a) averaged spectra were calculated for each type of epoch (waking, slow-wave sleep, REM sleep). Averaged spectra were median spectra (to ensure stability to possible outliers);

b) the maximum levels of motor activity during slow-wave and REM sleep were determined.

2. The following rules were applied to all epochs not in the training set:

a) if the level of motor activity was greater than the maximum level of motor activity during slow-wave and REM sleep identified by expert staging, this epoch was regarded as a waking epoch;

b) the distances between the spectrum of the ongoing epoch and the median spectra calculated at step 1a were cal-

culated. Distances were calculated using the so-called Canberra metric (Lance and Williams, 1967] calculated as:

$$d(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{n} \frac{|p_i - q_i|}{|p_i| + |q_i|},$$

where p and q are the two spectra being compared, i is the number of the frequency component, and n is the number of frequency components;

c) if the distance to the slow-wave sleep spectrum was less than the distances to the REM sleep and waking spectra, the epoch was assigned to slow-wave sleep;

d) if the distance to the REM sleep spectrum was also less than the distances to the spectra of other stages and the proportion of slow-wave sleep during the preceding minute was at least 75%, then the epoch was identified as REM sleep;

e) if the preceding condition was not fulfilled, the epoch was regarded as waking.

A program was written to record, visualize, and process PSG using the semiautomatic staging described above. This add-on was developed in EDFbrowser with open-source initial code.

The EEG spectral characteristics for individual epochs was the power spectrum in the range 2–30 Hz with a frequency resolution of 0.5 Hz, calculated by fast Fourier transformation (a standard EDFbrowser function).

Assessment of the relationship between staging accuracy and EEG trace parameters. EEG recordings were evaluated using the following parameters:

a) the distances between the median EEG spectra for the waking, slow-wave, and REM sleep epochs;

b) the widths of sleep and waking cluster silhouettes.

Silhouette width is a method for assessing the reliability of data clustering. It is calculated as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}},$$

where i is some point in the dataset; $a(i)$ is the mean distance between point i and all points in the cluster to which it belongs; $b(i)$ is the mean distance between i and all points in the closest-lying cluster (a cluster not containing i, with the smallest mean distance to its points) [Rousseeuw, 1987].

Clusters were waking, slow-wave sleep, and REM sleep identified by experts. Distances were Euclidean distances between the spectra of the corresponding epochs.

The accuracy of the algorithm described above was assessed by comparing the results of expert and computerized staging of 35 PSG traces. The accuracy of automatic staging was evaluated as the ratio of the number of epochs for which the expert and computerized evaluations coincided to the total number of epochs.

Three traces analyzed by two independent experts were used to assess the extent of agreement of expert staging. The level of agreement was assessed as the ratio of the number of epochs for which the two expert evaluations agreed to the total number of epochs.

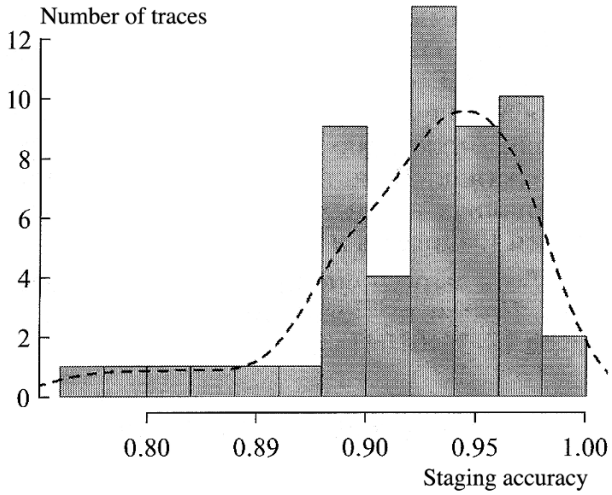

Fig. 2. Histogram of the distribution of the accuracy of agreement between automatic staging and expert staging. The abscissa shows the proportion of agreements and the ordinate shows the number of epochs.

All 38 traces were used to evaluate the relationship between staging accuracy, i.e., the proportion of epochs for which independent assessments (computerized and expert or two expert assessments) coincided and the following EEG characteristics:

a) the distance between the median EEG spectra for slow-wave sleep and waking epochs, slow-wave sleep and REM sleep epochs, and REM sleep and waking epochs, as well as the mean distance between the median spectra of all three states;

b) the width of the cluster silhouette in terms of the distances between the spectra for slow-wave sleep, REM sleep, and waking and the total cluster silhouette widths for all three states.

Statistical processing was performed in R version 3.0.2 (http://r-analytics.blogspot.ru). Linear approximation and assessment of its statistical significance were performed using the standard function lm.

Results

Overall, the process described above for automated staging showed high accuracy: the percentage agreement with expert staging was 93% (median). The histogram of the distribution of accuracy is shown in Fig. 2.

Analysis of the extent of agreement between staging by two independent experts on three traces showed that coincidence levels amounted to 98.3, 98.6, and 94.2% (mean coincidence levels between expert and automated staging for these PSG were 96.3, 97.6, and 85.3%, respectively).

Analysis of the influences of EEG parameters on staging accuracy identified a statistically significant relationship between the proportion of agreement between automatic and expert staging and the distance between the median

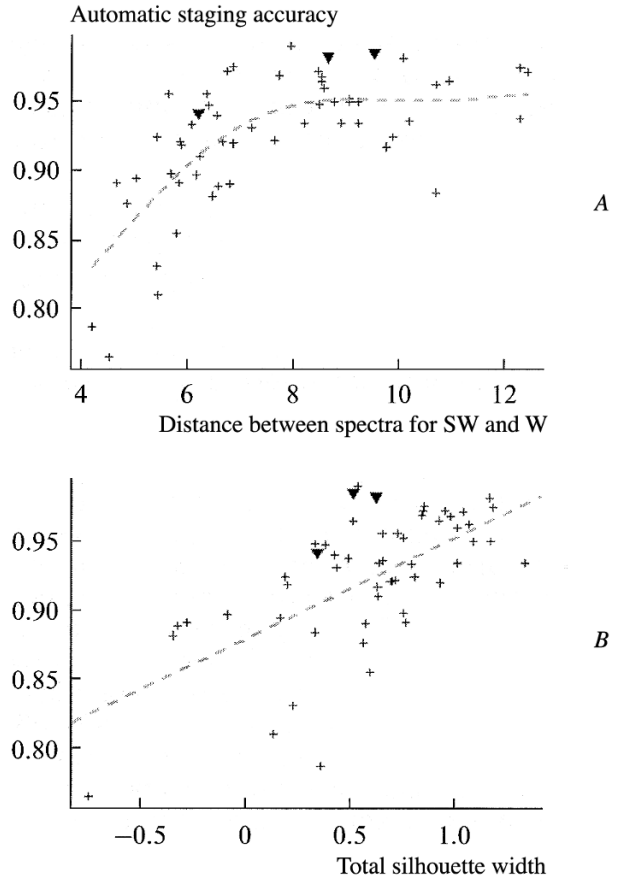

Fig. 3. Relationship between the accuracy of automatic staging (crosses) and extent of agreement of staging by two independent experts (triangles) of *A*) distances between mean spectra for slow-wave sleep (SW) and waking (W) and *B*) total width of cluster silhouettes for slow-wave sleep, REM sleep, and waking. The abscissas show EEG frequency characteristics and the ordinates show staging accuracy.

EEG spectra for slow-wave sleep and waking (rank correlation coefficient 0.61, Fig. 3, *A*); the mean distance between median spectra for waking and the two sleep phases (rank correlation coefficient 0.49, data not shown); and the total cluster silhouette width for slow-wave sleep, REM sleep, and waking (rank correlation coefficient 0.65, Fig. 3, *B*). The relationship between the accuracy of automatic staging and total silhouette width was close to linear ($R^2 = 0.43$; $F(1,54) = 38.4$, $p < 0.0001$).

We also found an essentially linear relationship between the proportion of coincident epochs for the two independent experts and total silhouette width. For the linear approximation, R^2 was 0.77, with a correlation coefficient of 0.87 (see Fig. 3, *B*).

Discussion

Automatic processing of the polysomnograms consisting of the EEG signal and the mechanogram of motor activity is particularly important for analyzing traces from long-term chronic experiments with implanted electrodes, when recording continues for several weeks. Trace quality could suffer from interference and from scar tissue forming around the electrode when present chronically in the brain. These and other causes of decreases in the expression of features discriminating episodes of waking and the different sleep stages give relevance to the question of objective criteria for evaluation of the suitability of PSG traces for expert and automatic analysis.

We demonstrated the existence of relationships between staging accuracy and:

1) the distance between averaged spectra;

2) cluster silhouette widths for waking, slow-wave sleep, and REM sleep.

The first relationship was nonlinear and the second was linear. The linearity of the relationship between the accuracy of automatic staging and total silhouette width makes it convenient for use in evaluating the a priori accuracy of automatic staging. That is, at the initial stage – after a small number of epochs had been staged by an expert for the training set – an assessment of the clarity or ambiguity of expert staging can be obtained, along with a prediction of the accuracy of automatic staging of the remainder of the trace. This in turn allows an objective criterion to be used for inclusion of an EEG trace for processing on the basis of its quality, increasing the level of objectivity in the data analysis.

Differences in the individual extents of EEG phenomena, including markers for waking and the various sleep stages (desynchronization for waking, high-amplitude, slow-wave activity for slow-wave sleep, and the θ rhythm for REM sleep), often significantly decrease the effectiveness of programs for automatic EEG analysis. The semiautomatic method for staging PSG – with a training set consisting of a small number of epochs staged by an expert – avoids this problem, allowing quite fine tuning of the analysis program to a given individual's data.

A programmable system for recording, displaying, and processing PSG by semiautomatic staging method described here has been approved and used successfully in practice. In particular, it has been used for analysis of data during the work reported in Manolov et al. [2014].

These studies were supported by the Russian Humanities Scientific Fund (Project No. 13-36-01041a1).

REFERENCES

Manolov, A. I., Dolgikh, V. V., Ukraintseva, Yu. V., et al., "Changes in motor activity and the sleep-waking cycle in an MPTP model of Parkinson's disease in mice," *Ros. Fiziol. Zh.*, **100**, No. 11, 1252–1260 (2014).

Becq, G., Charbonnier, S., Chapotot, E., et al., "Comparison between five classifiers for automatic scoring of human sleep recordings," in: *Classification and Clustering for Knowledge Discovery*, Springer, Berlin, Heidelberg (2005).

Buckelmüller, J., Landolt, H. P., Stassen, H. H., and Achermann, P., "Trait-like individual differences in the human sleep electroencephalogram," *Neuroscience*, **138**, No. 1, 351–356 (2006).

Huiskamp, G., "Interindividual variability of skull conductivity: an EEG-MEG analysis," *Int. J. Bioelectromagnetism*, **10**, No. 1, 25–30 (2008).

Lance, G. N. and Williams, W. T., "Mixed-data classificatory programs. I. Agglomerative systems," *Australian Comput. J.*, 15–20 (1967).

Rousseeuw, P. J., "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, **20**, 53–65 (1987).

Sannita, W. G., Loizzo, A., Garbarino, S., et al., "Adrenocorticotropin-related modulation of the human EEG and individual variability," *Neurosci. Lett.*, **262**, No. 3, 247–150 (1999).

Veasey, S. C., Valladares, O, Fenik, P., et al., "An automated system for recording and analysis of sleep in mice," *Sleep*, **23**, No. 8, 1025–1042 (2000).