Is the nonREM–REM sleep cycle reset by forced awakenings from REM sleep?

Michael Grözingera,*, Domien G.M. Beersmab, Jürgen Fella, Joachim Röschkea

aDepartment of Psychiatry, University of Mainz, Untere Zahlbacher Str. 8, 55101 Mainz, Germany
bDepartment of Zoology, University of Groningen, Groningen, The Netherlands

Received 3 January 2002; received in revised form 27 February 2002; accepted 30 July 2002

Abstract

In selective REM sleep deprivation (SRSD), the occurrence of stage REM is repeatedly interrupted by short awakenings. Typically, the interventions aggregate in clusters resembling the REM episodes in undisturbed sleep. This salient phenomenon can easily be explained if the nonREM–REM sleep process is continued during the periods of forced wakefulness. However, earlier studies have alternatively suggested that awakenings from sleep might rather discontinue and reset the ultradian process. Theoretically, the two explanations predict a different distribution of REM episode duration.

We evaluated 117 SRSD treatment nights recorded from 14 depressive inpatients receiving low dosages of Trimipramine. The alarms were triggered by an automatic mechanism for the detection of REM sleep and had to be canceled by the subjects themselves. The REM episodes were determined as in undisturbed sleep—they had to include the remaining REM activity and were separated by 30 min without REM epochs. The frequency histogram of REM episodes declined exponentially with episode duration for each of the first four sleep cycles. The duration of nonREM intervals revealed bimodal distributions. These results were found consistent with the model assuming a reset of the ultradian cycle upon awakening. Whether REM or nonREM activity is resumed on return to sleep can be modeled by a random decision whereby the probability for REM sleep might depend on the momentary REM pressure.

© 2002 Elsevier Science Inc. All rights reserved.

Keywords: Selective REM sleep deprivation; REM episode; Ultradian process; Sleep cycle; Sleep onset REM episode

1. Introduction

Normal sleep in humans is composed of four to six similar intervals that remind of cycles. Some 75 min of nonREM sleep are followed by about 15 min of REM activity in the first and 25 min in later cycles [1,2]. Wakening paradigms were applied in various experiments to challenge and study the mechanisms regulating this ultradian organization of sleep [3–7]. The impact of the interventions on the nonREM–REM alternations has been modeled according to two basic ideas. The first assumes an ongoing ultradian process that is maintained during the periods of forced wakefulness by an underlying generator (sleep dependent or independent) and is resumed on return to sleep. The second postulates that each intervention discontinues the ultradian process, which then has to be renewed on return to sleep. For awakenings from nonREM sleep, a study in healthy subjects seemed to rather support the second hypothesis [5].

There is no empirical evidence, however, that this result can easily be accepted for interventions from REM sleep. To fill this gap, we evaluated a series of sleep profiles recorded in an investigation on selective REM sleep deprivation (SRSD) [8]. During the procedure, occurrences of stage REM are repeatedly interrupted by short awakenings. SRSD was first introduced as dream deprivation, was later attributed antidepressive properties and has been applied in animals [9,10], healthy subjects [11,12] and psychiatric patients [13,14].

Various physiological responses following successive treatment nights are consistent with the concept of REM sleep homeostasis and are commonly summarized by the term REM pressure. These include an increasing number of awakenings required to prevent REM sleep, a decreasing REM latency, a rise of the wakening threshold and an increase in REM sleep immediately after SRSD [14]. In
the course of SRSD nights, the REM pressure increases, but still varies with the circadian phase and decreases during the remaining REM activity to an extent depending on the effectiveness of the procedure [15,16]. Typically, the interventions are not equally distributed in the course of SRSD recordings. During certain time periods, the awakenings are followed by immediate reoccurrences of REM sleep. Then, all of a sudden this behavior is abandoned and nonREM sleep prevails. Treatment nights therefore exhibit clusters of interrupted REM sleep activity alternating with bouts of nonREM sleep [15,16].

On first sight, this temporal structure appears self-evident because it resembles the REM episodes and the ultradian cycles during undisturbed sleep. This interpretation however takes for granted that the ongoing ultradian process is not completely interrupted by the SRSD awakenings but is instead maintained during the periods of forced wakefulness by some underlying generator. Accordingly, the timing of REM episodes is not determined by the interventions but by an endogenous process. For a sleep-independent generator, the distribution of REM episode duration should then be very similar to uninterrupted nights. For the latter, the histogram of REM episode duration was found gaussian-shaped with a mode occurring at about 15 min for the first and 25 min for the subsequent cycles with a standard deviation of about 10 min [2,5]. REM episodes of very short duration are relatively rare. For a sleep-dependent generator, the periodical interventions in SRSD are expected to extend the REM episodes by a scale factor, which reproduces another gaussian shape. If in conclusion the ultradian process is maintained during SRSD awakenings by a REM sleep generator and is resumed on return to sleep, the distribution of REM episode duration is expected to be gaussian-shaped.

Some authors have alternatively proposed that forced awakenings might discontinue the ongoing ultradian process. On return to sleep a new nonREM–REM cycle is claimed to be initiated starting with a decision for REM or nonREM sleep. Besides the momentary REM pressure, a variety of parameters can influence this choice [17–21]. Once the brain has decided for nonREM sleep, the system usually remains in this condition for at least some 30 min [3,5,22,23].

Obviously, this second model is also able to explain the clustering of interventions in SRSD: chances for stage REM to reoccur after the awakenings are high, because SRSD generates a considerable REM pressure [15,16]. Sleep is therefore repeatedly interrupted and the interventions will accumulate until nonREM sleep activity is chosen for the first time. Afterwards, the interventions will cease until a certain quantity of nonREM sleep has passed.

While the first model predicts a gaussian-shaped distribution of REM episode duration, the second one has a completely different implication: because REM sleep is immediately interrupted by interventions, the REM pressure can be assumed more or less constant during a cluster of interventions. In addition, the successive decisions for REM or nonREM sleep are independent from each other since the awakenings are supposed to completely discontinue the ultradian process. If the probability of entering REM sleep is denoted by \( p_{RP} \) during a cluster of awakenings, the probability that REM sleep can be observed after a series of two interventions is therefore \( p_{RP}^2 \) and after a series of \( n \) interventions, \( p_{RP}^n \). The probability for nonREM sleep to be chosen after exactly \( n \) interventions then equals \((1 - p_{RP})^n p_{RP}^n\). Accordingly, the number of awakenings up to the first decision for nonREM sleep is geometrically distributed. The same is true for the histogram of REM episode duration provided that the interventions occur at approximately equal intervals during the clusters. The chances to find certain REM episode durations are therefore exponentially declining with their duration. As opposed to the gaussian-shaped distribution, very short episodes are expected to be the most numerous here.

To compare both models, the distribution of REM episode duration was evaluated for a large data set of SRSD recordings obtained from depressed patients. For this study, an automatic algorithm for the online detection of REM sleep was developed, validated and continuously improved in our laboratory [24–26]. While only the aspects concerning the regulation of REM sleep are presented here, the therapeutic effects of SRSD are reported elsewhere [8].

2. Methods

SRSD was applied to 14 depressed inpatients, 10 females and 4 males. Data on medical history, physical inspection, blood tests, ECG and EEG were collected in all subjects. Prior to the first treatment night, patients had the following characteristics: age 43.6 ± 11.6 (range: 24–58), diagnosis of unipolar depression, HAMD21 score 23.9 ± 4.2 (range: 18–31), duration of current episode 4.3 ± 3.5 months, number of depressive episodes 1.9 ± 0.9, duration of illness 47.9 ± 73.7 months, no psychotic features, no other relevant psychiatric diagnosis or organic brain disease, no suicidal tendencies, no serious somatic condition, no regular intake of benzodiazepines.

Four of the patients had taken antidepressive medication in a constant dosage for more than 2 weeks without improvement (150 mg Doxepin, 150 mg Opipramol, 200 and 50 mg Trimipramine daily). The prior Trimipramine medication was maintained, but the two other antidepressants were discontinued. Ten remaining patients had not received psychotropic drugs for at least 2 weeks. For ethical concerns, subjects had to be offered an antidepressant. Trimipramine was chosen for its property not to suppress REM sleep [27,28]. The dosage was kept as low as possible and adapted to clinical requirements like agitation, insomnia and side effects. With a mean of 29.5 ± 55.2 mg (range: 0–200 mg), the average daily dosage stayed well below clinical routine treatment. On rare occasions, the adminis-
tation of Chloral Hydrate and Lorazepam (8 mg in total) was unavoidable to manage acute anxiety. No other psychotropic drugs were administered during the investigation.

Following an adaptation night, SRSD was applied to patients in the sleep laboratory for a series of 10 consecutive nights from 23:00 p.m. until 7:00 a.m. A commercial halogen floor lamp of constant and strong intensity was turned on to overcome the strong habituation towards the disturbances. Subjects had to press a button to cancel the alarm clock with slowly incrementing loudness was triggered on each occurrence of REM sleep. Simultaneously, an alarm with a loud tone was played.

Sleep was strictly prohibited by the nurses. Daytime conditions, all subjects stayed on the same ward. Daytime sleep was electronically registered. To provide for similar conditions, all subjects stayed on the same ward. Daytime sleep was strictly prohibited by the nurses.

The protocol was approved by the local ethics committee. The procedures used were in compliance with the Declaration of Helsinki, and informed consent was obtained from all subjects. Two patients stopped participating after two intervention nights, one patient after 3 nights and 11 completed the series of 10 treatments. As a result, 117 recordings were evaluated.

Prior to the study, an automatic algorithm based on artificial neural networks had been developed for the automatic detection of REM sleep [24,26]. The EEG channel Cz/A1 served as a single source of input. For each 20-s epoch, the signal was preprocessed by calculating the power in seven frequency bands and in eight adjacent time segments of 2.5 s. Based on these 56 input values, the neural network decided on the presence of stage REM. Manually, evaluated sleep profiles served as examples during training sessions.

In certain situations, the sleep stage definitions according to Rechtschaffen and Kales must be correctly classified as REM or nonREM sleep [24–26]. Most of the errors were due to confounding the sleep stage I and REM.

Sleep onset was determined by the first occurrence of stage II, SWS or REM sleep. The time interval from turning off the lights to sleep onset was regarded as sleep latency and from sleep onset to the first REM epoch as REM latency.

An alarm or REM epoch was considered to terminate a REM episode if no alarm or REM sleep occurred within the following 30 min of the recording. Wakefulness was allowed to contribute to this time interval. The first REM epoch of the recording respectively the first REM epoch following a terminated REM episode marked the beginning of a new REM episode. Time periods from sleep onset to the first REM episode or in between two REM episodes were considered nonREM intervals. If not specified otherwise, wakefulness was subtracted when the duration of REM episodes and nonREM intervals was quoted.

The histogram of REM episode duration was approximated for each sleep cycle by a geometrical distribution.

![Fig. 1. The number of REM epochs that precede the alarms. The diagram shows the percentage of intervals between adjacent interventions containing less than a given number of REM epochs. In total, 4776 intervals were evaluated from 117 EEG recordings.](image-url)
The corresponding decay parameters were estimated using the maximum likelihood method \( p_{est} = 1 - n/\sum x_j \) for \( n \) REM episodes with duration \( x_j \). To statistically appreciate, this procedure the range of REM episode duration was divided in six bins and chi-square goodness of fit tests were applied.

### 3. Results

For each of the 10 consecutive treatment nights, the variables characterizing the sleep profiles were averaged across subjects and are demonstrated in Table 1.

The number of REM epochs in between adjacent interventions was evaluated for the present study and illustrated in Fig. 1 to indicate the sensitivity of the REM sleep detection algorithm. Eighty-six percent of the intervals between adjacent interventions included less than five REM epochs. Only 2.5% of these intervals included more than 12 REM epochs. Therefore, the algorithm seemed well suited for the detection of REM sleep.

As an example the sleep profile of a fifth treatment night is illustrated in Fig. 2. The alarms are indicated by super-

![Fig. 2. Example of the sleep profile of a treatment night. The automatic mechanism triggered 41 alarms. The remaining REM sleep was 41 min and 20 s corresponding to 9.65% of SPT.](image)

imposed markers. Although the sleep process is repeatedly interrupted by wakefulness, the remaining REM sleep and the interventions still aggregate in clusters reminding of the REM episodes in undisturbed nights.

Statistically, the tendency of the interventions to aggregate in clusters can be recognized from Fig. 3. During the first minutes after an alarm chances for another one to occur are high. In 84% of the cases, two alarms are less than 10 min apart. The diagram also indicates a second mode of the distribution at about 50 min and a minimum occurring between 20 and 40 min.

The correlation between the REM episode duration and the number of alarms during that episode was characterized by coefficients of .80–.86 for the different sleep cycles. The corresponding regression lines revealed slopes of 214–261 s per intervention. This indicates rather equidistant alarms in the course of each cluster.

The average time span needed to respond to an alarm and the sleep onset latency thereafter are demonstrated in Fig. 4. While the first variable increases significantly in the course of the nights (\( r^2 = 0.63, F = 13.38, df = 8, P < .01 \)), the second did not correlate with the progress of the procedure (\( r^2 = 0.20, F = 2.00, df = 8, P = .19 \)). On grand average, the subjects needed 76.6 s to respond to an alarm. In 2.9% of the cases, it took more than 5 min and in 1% more than 10 min. The mean sleep latency after terminating the alarms was 93.2 s.

![Fig. 3. The time span between adjacent interventions. The lower curve refers to the left vertical axis and shows the percentage of intervals with a given duration. In total, 4776 intervals were evaluated from 117 EEG recordings. The upper curve refers to the right axis and provides the same relationship on a different scaling for the better recognition of small values.](image)

![Fig. 4. The reaction time following an alarm and the sleep onset latency afterwards in the course of succeeding treatment nights. The lower curve refers to the right vertical axis and shows the average time span between the occurrence and the cancellation of an alarm. The upper curve refers to the left axis and shows the time span between canceling the alarm and returning to sleep.](image)

![Fig. 5. Means and standard deviations of the intraindividual averages of interventions during the different recording hours.](image)
In 3.8% of the cases, it was more than 5 min and in 1.7% more than 10 min.

For every subject and every recording hour between 23:00 p.m. and 7:00 a.m., the number of awakenings was averaged across treatment nights. The means and standard deviations of these intraindividual values are demonstrated in Fig. 5. The intervention frequency increases until early in the morning ($r^2 = 0.98, F = 246.9, df = 4, P < 10^{-4}$ for 23:00 p.m. until 4:00 a.m.). The falling slope thereafter was attributed to the growing rate of patients waking up.

The histograms of REM episode duration are illustrated in Fig. 6 for the first to the fourth sleep cycle. On a linear scale (left side), the empirical distributions do not resemble the typical gaussian shape but rather reveal a high frequency of the short episodes. The logarithmic scale (right side) helps to identify a geometrical distribution by searching for a linear relationship. Except for a few values the graphs indeed appear approximately linear on visual inspection with the shortest REM episodes occurring most frequently.

A statistical analysis of this hypothesis is demonstrated in Table 2. No significant misfit was detected at a 5% error level. The 95% confidence intervals of the estimated param-

Table 2

<table>
<thead>
<tr>
<th>Cycle 1</th>
<th>Cycle 2</th>
<th>Cycle 3</th>
<th>Cycle 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obs</td>
<td>Exp</td>
<td>Obs</td>
</tr>
<tr>
<td>Range</td>
<td>1–3</td>
<td>36</td>
<td>29.0</td>
</tr>
<tr>
<td></td>
<td>4–8</td>
<td>24</td>
<td>33.1</td>
</tr>
<tr>
<td></td>
<td>9–15</td>
<td>26</td>
<td>26.3</td>
</tr>
<tr>
<td></td>
<td>16–24</td>
<td>16</td>
<td>15.9</td>
</tr>
<tr>
<td></td>
<td>25–37</td>
<td>11</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td>&gt;38</td>
<td>3</td>
<td>4.1</td>
</tr>
<tr>
<td>$\hat{p}_{\text{est}}$</td>
<td>0.909</td>
<td>0.945</td>
<td>0.968</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.89–0.93</td>
<td>0.94–0.96</td>
<td>0.96–0.97</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>6.03</td>
<td>7.69</td>
<td>7.31</td>
</tr>
<tr>
<td>$P$-value</td>
<td>.30</td>
<td>.17</td>
<td>.20</td>
</tr>
</tbody>
</table>

The decay parameters were estimated using the maximum likelihood method $\hat{p}_{\text{est}} = 1 - \prod x_j$ for $n$ REM episodes with duration $x_j$. The range of REM episode duration was divided in six bins and chi-square goodness of fit tests were applied for $df = 5$.

Fig. 6. Histograms of REM episode duration for the first to the fourth sleep cycle on a linear (left diagrams) and a logarithmic scale (right diagrams).

Fig. 7. Histograms of nonREM interval duration for the first to the fourth sleep cycle on a linear scale.
eters $p_{est}$ were interpolated from statistical tables \[30\] and were found disjunctive for the first to the third sleep cycle. The decay of the geometric distributions can therefore be assumed to decrease with the order of the cycle. This reflects a higher probability for stage REM to be initiated during the later part of the night and is most likely due to the increase of REM sleep propensity in the course of the night. The forth sleep cycle can be distinguished in this regard from the first two but not from the third.

The histograms of nonREM interval duration are shown in Fig. 7 for the first to the fourth sleep cycle. The modes of the distributions are located in the range between 45 and 65 min. The diagrams also specify nonREM intervals shorter than 30 min because wake time was subtracted here.

### 4. Discussion

In SRSD, the interventions and the remaining REM sleep activity aggregate in clusters reminding of the REM episodes in undisturbed nights \[15,16\]. This phenomenon was confirmed in the present study. The high sensitivity of the REM detection algorithm seemed appropriate to prevent a rapid accumulation of REM sleep during these clusters. The REM pressure was therefore assumed approximately constant in the course of a REM episode. On a larger time scale, however, the REM pressure varies due to circadian influences and due to the ongoing REM sleep deprivation. This can for example be observed by the decreasing decay of the geometric distributions between the first and the third sleep cycle.

The histograms of REM episode duration did not reveal gaussian distributions for the first four sleep cycles. Instead, short REM episodes were most frequent and the empirical data were found compatible with geometrical distributions. This result is inconsistent with the model assuming a continuation of the ultradian process during the periods of forced wakefulness. It rather suggests that the nonREM–REM cycling is discontinued by the interventions. On the subsequent transition from wakefulness to sleep, a new ultradian process is then initiated starting with a decision for REM or nonREM sleep.

For the first sleep cycle, the histogram of nonREM interval duration reflects the well known bimodal distribution of REM latencies in depressive patients \[31–33\]. Most likely, the phenomenon was enhanced in the present study by the increased REM pressure during SRSD. The modes of the nonREM interval distributions are located in the range between 45 and 65 min. These relatively short values as compared to the undisturbed sleep of depressed patients illustrate the acceleration of the ultradian alternations in SRSD due to the high REM pressure \[34,35\].

When nonREM sleep was interrupted by extended periods of wakefulness, the sleep cycles were found discontinued and the ultradian phase reset \[3,5\]. On return to sleep, new REM–nonREM sleep sequences started either with a sleep onset REM period (SOREMP) or a regular nonREM interval independent of previous sleep-dependent or sleep-independent ultradian processes. This split in REM latency did not correspond to a difference in the prior sleep content \[5\].

The concept that forced awakenings initiate new ultradian alternations starting with either a SOREMP or a nonREM interval can also be transferred to SRSD interventions. The immediate reoccurrence of REM sleep after a SRSD intervention can then be interpreted as a SOREMP and appears in the sleep profile as a continuation of the preceding REM episode. Due to the increased REM pressure in SRSD a higher incidence and a shorter latency of SOREMPS have to be expected as compared to undisturbed nights. This consideration corresponds well with our assumption of a random decision for REM or nonREM activity on return to sleep.

In conclusion, our results from SRSD interventions are in accordance with the general idea that forced awakenings discontinue the ongoing sleep cycle and reset the ultradian phase. On return to sleep a new nonREM–REM sleep sequence starts either with a SOREMP or with a nonREM interval. The probability for these alternatives depends on the momentary REM pressure.

Up to now, all considerations have been limited to the effects of forced awakenings. The above hypothesis might however be speculatively extended to the initiation of REM sleep in general: transitions to stage REM might be allowed during periods of light sleep at the beginning and the end of sleep cycles. Stage REM might then occur randomly with a probability depending on the momentary REM and nonREM tendencies. This concept could explain the sporadic observation of SOREMPS at the beginning of a sleep cycle, could include the discussed phenomena following forced awakenings and might also serve as an explanation for the so-called skipped REM episodes \[23,36,37\].

According to the population studied, the conclusions refer to depressive patients in the first place. There are well-documented differences concerning some REM sleep parameters in depressed patients and in healthy volunteers \[38–40\]. On the other hand, the qualitative reactions to SRSD are not different in both groups of subjects. The clustered pattern of awakenings, the abbreviated ultradian cycles and the increasing REM pressure are demonstrated likewise for both groups. Some more aspects concerning the similarity of REM sleep regulation in depressed patients and healthy subjects have already been suggested in earlier publications \[41,42\]. There is accordingly good reason to assume that the results can be transferred to healthy subjects. Nevertheless, this will have to be confirmed.

### Acknowledgements

This work was supported by Deutsche Forschungsgemeinschaft Ro 809/13-1.
References


