Epilepsia, **47**(12):2058–2070, 2006
Blackwell Publishing, Inc.

Do False Predictions of Seizures Depend on the State of Vigilance? A Report from Two Seizure-Prediction Methods and Proposed Remedies

*†Björn Schelter, *†Matthias Winterhalder, *Thomas Maiwald, ‡Armin Brandt, *Ariane Schad, *†Jens Timmer, and †‡Andreas Schulze-Bonhage

**FDM, Freiburg Center for Data Analysis and Modeling, and †Bernstein Center for Computational Neuroscience Freiburg, University of Freiburg; and ‡Epilepsy Center, University Hospital of Freiburg, Freiburg, Germany*

Summary: *Purpose:* Available seizure-prediction algorithms are accompanied by high numbers of false predictions to achieve high sensitivity. Little is known about the extent to which changes in EEG dynamics contribute to false predictions. This study addresses potential causes and the circadian distribution of false predictions as well as their relation to the sleep–wake cycle.

Methods: In 21 patients, each with 24 h of interictal invasive EEG recordings, two methods, the *dynamic similarity index* and the *mean phase coherence,* were assessed with respect to time points of false predictions. Visual inspection of the invasive EEG data and additional scalp electroencephalogram data was performed at times of false predictions to identify possible correlates of changes in the EEG dynamics.

Results: A dependency of false predictions on the time of day is shown. Renormalized to the duration of the period patients are asleep and awake, 86% of all false predictions occurred during sleep for the dynamic similarity index and 68% for the mean phase coherence, respectively. Combining two reference intervals, one during sleep and one in an awake state, the dynamic similarity index increases its performance by reducing the number of false predictions by almost 50% without major changes in sensitivity. No obvious dependence of false predictions was noted on visible epileptic activity, such as spikes, sharp waves, or subclinical ictal patterns.

Conclusions: Changes in the EEG dynamics related to the sleep–wake cycle contribute to limits of specificity of both seizure-prediction methods investigated. This may provide a clue for improving prediction methods in general. The combination of reference states yields promising results and may offer opportunities to increase further the performance of prediction methods. **Key Words:** Seizure prediction—Seizure anticipation—False predictions—Phase synchronization—Dynamic similarity index.

The daily life of epilepsy patients who cannot be treated successfully with current therapeutic strategies is impaired by recurrent unforeseeable seizures, which potentially lead to life-threatening situations (Buck et al., 1997). The reliable prediction of epileptic seizures in advance of a seizure onset could dramatically improve the quality of life of these patients (Schachter, 1994). For instance, they could either be warned to prevent dangerous situations or be treated in time by implanted devices by using electrical stimulation or the delivery of short-acting drugs (Elger, 2001).

So far the challenge to predict epileptic seizures is approached by searching for characteristic changes in the electroencephalogram before seizure onsets. Such characteristic changes are, for instance, quantified by features taking extreme values when such changes occur. A threshold can be fixed, and whenever the threshold is crossed, an alarm is raised. However, detecting threshold crossings before seizure onsets and using the corresponding alarms to predict seizures are different issues. For instance, if the distribution of alarms before seizures is rather widespread, a reliable seizure prediction is not possible. To assess whether an algorithm is capable of predicting seizures, we suggest the use of the seizure-prediction characteristic (Winterhalder et al., 2003). This assessment method evaluates the prediction performance with particular emphasis on the temporal aspects of a prediction. First, subsequent to an alarm, a time interval is necessary during which no seizure is supposed to start, to be regarded as a correctly predicted. This time interval, accounting for interventions, quantifies the actual prediction time. In the context of the seizure-prediction characteristic, it is called

Accepted May 8, 2006.
Address correspondence and reprint requests to B. Schelter at FDM, Freiburg Center for Data Analysis and Modeling, University of Freiburg, Eckerstr. 1, 79104 Freiburg, Germany. E-mail: schelter@fdm.uni-freiburg.de
doi: 10.1111/j.1528-1167.2006.00848.x

seizure-prediction horizon (SPH). Second, subsequent to the SPH, a time interval is fixed in which a seizure is supposed to start to be regarded as correctly predicted. This time interval, the so-called seizure occurrence period (SOP), accounts not only for the uncertainty in the prediction but also limits the time period the patient is under risk of having a seizure. After fixing a range for both time intervals before any analyses, the seizure-prediction performance must be compared with a random predictor (Schelter et al., 2006). If and only if the performance of the prediction algorithm is superior to a random predictor, we speak of a seizure prediction. Neglecting the SPH and the comparison with a random predictor would correspond to an anticipation of seizures.

To determine promising algorithms for seizure prediction, numerous efforts have been undertaken to detect characteristic changes in the EEG dynamics before seizure onsets over the last few years. Several linear and nonlinear methods have been intensively studied, including accumulated energy (Litt et al., 2001), neuronal complexity based on the correlation dimension (Lehnertz and Elger, 1995; 1998; Osorio et al., 2001), approaches based on neural cell models (Schindler et al., 2002), the largest Lyapunov exponent (Iasemidis et al., 1990), a dynamic similarity index (Le van Quyen et al., 1999; 2000; 2001a; 2001b; Navarro et al., 2002), or synchronization measures (Mormann et al., 2000; Jerger et al., 2001; Mormann et al., 2003a; 2003b; 2005). The current state of the art in the field of seizure anticipation and prediction is summarized in these reports (Litt and Echauz, 2002; Litt and Lehnertz, 2002).

On the one hand, high sensitivity is a desired goal in seizure prediction. A missed seizure in a patient relying on a prediction method could cause life-threatening situations. On the other hand, too many false predictions may be accompanied by potential side effects of interventions or by the loss of the patient's acceptance of seizure-warning devices. Thus seizure-prediction methods have to achieve both, sufficient sensitivity and specificity. An assessment of several algorithms within the framework of the seizure prediction characteristic has shown that low specificity has to be accepted to achieve high prediction sensitivity (Aschenbrenner-Scheibe et al., 2003; Winterhalder et al., 2003; Maiwald et al., 2004).

Possible reasons for false predictions, the number of which is an appropriate measure for specificity in the context of seizure prediction (Osorio et al., 1998), are thus of particular interest. In addition, the distribution of false predictions over the day may be relevant for the practical application of a prediction method. If, for instance, all false predictions occur while patients are asleep, a missed seizure may be less harmful than a seizure during the day, because patients are less vulnerable to accidents or injuries during sleep. Consequently, seizure-prediction methods could be adjusted if false predictions occurred only during particular states of vigilance.

The relation of changes in EEG dynamics during preseizure periods to visual inspection of the EEG has been investigated recently (Navarro et al., 2005). However, examinations into the causes of false predictions and their relations to EEG dynamics have not been the focus of research. False predictions of two previously published algorithms are analyzed in this study. The first one is a univariate approach that quantifies the similarity of patterns in the EEG data, the *dynamic similarity index* (Le van Quyen et al., 1999). More precisely, the dynamic similarity index compares the EEG characteristics in a given time window with the EEG characteristics in a reference interval. The similarity index is a reference-dependent method.

Second, a bivariate method takes into account relations between different electrode contacts, with the *mean phase coherence* measuring synchronization (Mormann et al., 2000; 2003a; 2003b). This method makes no use of reference intervals and is thus a reference-free method.

Intracranial EEG data of 21 patients with pharmacoresistant focal epilepsy form the basis of the present study. First, seizure-prediction performance is presented for both algorithms and compared with the performance of a random predictor to illustrate that both seizure-prediction methods are reasonable candidates that may be used for seizure prediction in the future. Second, time points of false predictions and their distribution over the day are analyzed for all patients. To determine states of vigilance during the emergence of false predictions, a subgroup of 10 patients for whom simultaneous scalp EEG recordings are available is studied.

Third, the relation of false predictions to potential events in the invasive EEG is analyzed. Based on the results for circadian dependencies of false predictions for the reference-dependent seizure-prediction method (i.e., the dynamic similarity index), a strategy is investigated to reduce the number of false predictions. Although the specificity of the method is improved considerably, the sensitivity hardly changes, which leads to an improved overall seizure-prediction performance.

The article is structured as follows. In the next section, the patients' characteristics and the invasive and scalp EEG data are described. Furthermore, the seizure-prediction characteristic is briefly summarized and the two investigated seizure-prediction methods are introduced. Results are followed by the discussion. Some of the results have been reported previously in abstract form (Schelter et al., 2004).

METHODS

The EEG database used for this investigation and the two prediction methods are briefly introduced in this section. The method for evaluation of the prediction performance is summarized.

TABLE 1. *EEG data and patient characteristics*

Patient	Sex	Age	Seizure type	H/NC	Origin	Electrodes	# Seizures analyzed	Interictal duration [h]
1	F	15	SP,CP	NC	Frontal	g,s	5	24
2	M	38	SP,CP,GTC	H	Temporal		3	24
3	M	14	SP,CP	NC	Frontal	g,s	5	24
4	F	26	SP,CP,GTC	H	Temporal	d,g,s	5	24
5	F	16	SP,CP,GTC	NC	Frontal	g,s	3	24
6	F	31	CP,GTC	H	Temporo-occipital	d,g,s	3	24
7	F	42	SP,CP,GTC	H	Temporal	d	2	25
8	F	32	SP,CP	NC	Frontal	g,s	5	24
9	M	44	CP,GTC	NC	Temporo-occipital	g,s	5	24
10	M	47	SP,CP,GTC	H	Temporal	d	4	24
11	F	10	SP,CP,GTC	NC	Parietal	g,s	4	24
12	F	42	SP,CP,GTC	H	Temporal	d,g,s		25
13	F	22	SP,CP,GTC	H	Temporo-occipital	d,s	2	24
14	F	41	CP,GTC	H and NC	Frontal-temporal	d,s	4	24
15	M	31	SP,CP,GTC	H and NC	Temporal	d,s	4	24
16	F	50	SP,CP,GTC	H	Temporal	d,s	5	24
17	M	28	SP,CP,GTC	NC	Temporal	s	5	24
18	F	25	SP,CP	NC	Frontal	s	5	25
19	F	28	SP,CP,GTC	NC	Frontal	s	4	24
20	M	33	SP,CP,GTC	NC	Temporoparietal	d,g,s	5	26
21	M	13	SP,CP	NC	Temporal	g,s	5	24
						total	88	509
						mean	4,2	24,2

Seizure types: SP, simple partial; CP, complex partial; and GTC, generalized tonic–clonic seizure origin; H, hippocampal and NC, neocortical; g, electrodes grid; s, strip; d, depth.

For each patient, either all or five seizures (mean, 4.2; total, 88) and $\geq$24 h of interictal EEG recordings (mean, 24.2 h/total, 509 h) are examined.

EEG recordings and patient characteristics

Data from invasive long-term EEG recordings from 21 patients undergoing presurgical epilepsy monitoring were investigated. The retrospective evaluation of the data received prior approval by the Ethics Committee, Medical Faculty, University of Freiburg. Informed consent was obtained from each patient. The EEG data were recorded using a sampling rate of 256 Hz or 512 Hz, respectively, and were bandpass-filtered between 0.5 and 120 Hz. A 50 Hz notch-filter was used to eliminate possible line noise.

A restriction of the analysis to few electrode contacts is necessary because the performance of a random predictor is strongly dependent on the number of electrode contacts investigated (Schelter et al., 2006). To be able to achieve a prediction performance superior to a random predictor, six contacts of all implanted grid, strip, and depth electrodes were selected before any analyses by visual inspection of the raw data by a certified epileptologist (A.S.).

Three contacts were chosen from the seizure-onset zone (i.e., from areas involved early in ictal activity). The remaining three electrode contacts were selected as not involved or involved latest during seizure spread. This procedure ensures that electrodes from the seizure focus as well as out-of-focus areas of the brain are well represented.

At least 24 h of continuous interictal recordings were available for 13 patients. For the remaining eight patients, interictal EEG data consisting of $<$24 h were joined such that $\geq$24 h per patient (mean, 24.2 h) and all times of day were included. In total, the time period between 9 pm and 9 am comprises 265 h of the total interictal recording, and the time period between 9 am and 9 pm, 243 h. The median time between the last seizure and the beginning of the interictal period was 5 h 18 min, and the median time between the end of the interictal period and the next seizure was 9 h 36 min. The minimum time between the end of the interictal period and the next seizure was 1 h 18 min for one patient (number 14). For the remaining patients, this time interval was $\geq$3 h. In a subgroup of 10 patients, additional scalp EEG recordings including electrooculographic (EOG) electrodes and submental electromyographic (EMG) electrodes were available for classification of sleep stages according to Rechtschaffen and Kales (1968).

To evaluate the sensitivity of both methods, 88 seizures with 50 min preictal invasive EEG recordings were examined. The number of seizures per patient varied from two to five, with a mean value of 4.2 seizures per patient. Further details of the patients' characteristics and the investigated EEG data base are given in Table 1.

The seizure-prediction characteristic

The seizure prediction performance is evaluated by using the recently introduced seizure-prediction characteristic S(FPR_{max}, SOP, SPH) (Winterhalder et al., 2003). The seizure-prediction characteristic estimates sensitivity based on three factors characterizing a prediction.

First, it depends on the maximum false prediction rate FPR_{max}, which is the maximum number of false predictions allowed in a certain time interval. This limit is necessary because of the possible side effects of intervention

systems or of the loss of the patient's acceptance of a seizure-warning device after too many false predictions.

For a prediction, the time interval after an alarm during which the seizure is expected to occur must be specified. To account for the impossibility of a perfect prediction and to allow a temporal uncertainty in the occurrence of predicted seizures, the SOP is introduced. A correct prediction is defined by the occurrence of a seizure within the SOP.

Furthermore, any intervention system or seizure-warning device would need some time between being activated and becoming effective. Therefore a time interval between the alarm and the beginning of SOP is required, the SPH. Only seizures starting after the SPH and within the SOP are classified as correctly predicted.

A test to decide about the statistical significance of a certain value of the seizure-prediction characteristic is defined by the "prediction performance" of an unspecific random prediction (Schelter et al., 2006). For an unspecific random prediction, alarms are triggered randomly according to a Poisson process in time without using any information from the EEG. Critical sensitivity values of the random prediction can be calculated for an individual patient based on the parameters FPR_{max}, SOP, the number of preseizure periods investigated, and the number of features analyzed. Because it is unknown whether the data in two different electrode contacts can be assumed to be statistically independent, two critical sensitivity values have to be considered. Sensitivity of a prediction method can be considered superior to a random predictor if it is higher than the upper critical sensitivity. Superiority to a random predictor cannot be clarified finally, if it is between the lower and upper critical value. However, a certain prediction performance can be expected for these seizure-prediction algorithms. For details concerning the statistical test used in this study, see (Schelter et al., 2006).

In summary, we use the term "prediction" if sensitivity is calculated in dependence on FPR_{max}, SOP, and SPH; the range of values for both time intervals are fixed with respect to a clinical application; and the prediction algorithm is superior to a random predictor. In contrast, we use the term "anticipation" if the SPH was ignored and a comparison with a random predictor was missing.

In this study, the focus was on false predictions. To obtain a sufficient amount of false predictions, we used a maximum false-prediction rate of 0.5 FP/h. The random predictor will yield a rather high upper critical value. Thus it will hardly be possible to obtain higher prediction performance than the upper critical sensitivity of the random predictor. To assess the prediction methods that achieve some prediction performance, we require the prediction algorithm to be, at least, comparable with the upper critical value of the random predictor and above the lower critical value. Strictly speaking, we cannot really prove superiority to a random predictor in these cases. However, some prediction performance is guaranteed by our approach, and the parameters are a trade-off between showing predictability of seizures and obtaining a sufficient amount of false predictions to be able to assess them.

In the following, the SPH has been fixed to 2 min, which is an appropriate value to enable, for instance, administration of short-acting drugs or warning of the patient. The SOP has been chosen to last for 30 min. When the seizure-prediction methods have been shown to work sufficiently well, a much smaller SOP might be eligible, for example, for a seizure-warning device. However, thinking of automatic seizure prevention, an SOP on the order of half an hour is reasonable, too, if the treatment effect lasts for this period. This is, for example, usually expected for antiepileptic drugs (AEDs).

Two seizure-prediction methods

A seizure-prediction method analyzing similarity patterns in the EEG data as well as a prediction method based on synchronization theory is analyzed in this study.

Dynamic similarity index

The dynamic similarity index compares the dynamic behavior of the EEG signals within sliding time intervals S_t with a fixed reference interval S_{ref} (Le van Quyen et al., 1999). The reference interval is chosen at the beginning of the interictal period and is 5 min long. The sliding time window has a duration of 25 s (Le van Quyen et al., 1999). Specific features of the invasive EEG data are captured by constructing time intervals between positive zero crossings, which are delay-embedded and projected on the principal axes calculated for the reference interval, subsequently leading to $X(S_t)$. A random selection $Y(S_{ref})$ of the reference interval is compared with the sliding time intervals by using the cross-correlation sum

$$C(S_{ref}, S_t) = \frac{1}{N_{ref} N_t} \sum_{i=1}^{N_{ref}} \sum_{j=1}^{N_t} \Theta(r - \|\vec{Y}_i(S_{ref}) - \vec{X}_j(S_t)\|). \quad (1)$$

The Heaviside function is denoted with Θ, the Euclidean norm with $\|\cdot\|$, and the numbers of vectors with N_{ref} and N_t, respectively. Finally, the dynamic similarity index

$$\gamma(S_t) = \frac{C(S_{ref}, S_t)}{\sqrt{C(S_{ref}, S_{ref}) C(S_t, S_t)}}. \quad (2)$$

is estimated. An alarm is triggered by lower values of the dynamic similarity index than a given threshold value in one electrode.

Mean phase coherence

Nonidentical, self-sustained, chaotic oscillators have been shown to synchronize their phases if the oscillators are weakly coupled. Amplitudes stay uncorrelated in the case of this phase synchronization. An almost constant

phase difference (Rosenblum et al., 1996)

$$\left|n\Phi^{(1)} - m\Phi^{(2)}\right| = \left|\Phi_{n,m}\right| < \text{const} \qquad (3)$$

leads to the notion of phase synchronization, where $\Phi^{(1,2)}$ denotes the phase signals of the two time series investigated, and *n,m* are two integers. In presence of additional stochastic influences, particular values of $\Psi_{n,m} = \Phi_{n,m} \mathrm{mod} 2\pi$ are preferred. The corresponding distribution can be quantified (Mormann et al., 2000; Rosenblum et al., 2001)

$$R^2_{n,m} = \left\langle cos\,\Psi_{n,m}(t)\right\rangle^2 + \left\langle \sin\Psi_{n,m}(t)\right\rangle^2 . \qquad (4)$$

Perfect synchronization is represented by a value of one, whereas values close to zero occur in the absence of any phase synchronization. When applied to EEG data from epilepsy patients, a decrease of the mean phase coherence R: = R_1,1 has been reported in advance of seizure onsets (Mormann et al., 2000; 2003a). An alarm is raised once a threshold is crossed in the mean phase coherence estimated for a certain pair of electrodes.

Originating from nonlinear dynamics estimating the mean phase coherence based on the Hilbert transform is reasonable for narrow-band signals only. In terms of seizure prediction, the method, however, is used to predict seizures and not to estimate synchronization. Therefore any procedure using filters or not is reasonable as long as a proper statistical assessment of the prediction performance is ensured. Moreover, no easy rule can be given for in which frequency band the signals should be band-pass filtered, especially for nonictal data. Using too many frequency bands would increase the performance of a random predictor, leading to hardly any significant superiority of the seizure-prediction performance of the proposed algorithms.

RESULTS

In the following investigations, the maximum false-prediction rate has been fixed to $\mathrm{FPR}_{\max} = 0.5$ FP/h. As our study focused on the analysis of false predictions, parameters had to be chosen to allow a sufficient number of false predictions to render a statistical evaluation possible. Nevertheless, to ensure that we do not discuss arbitrary results, more false-prediction rates have been evaluated but are not reported, as they show very similar results. The SOP has been set to half an hour, and the SPH, to 2 min. This was done to ensure that we could use the same parameter set for all patients. Averaged for all 21 patients, a sensitivity of 82% is obtained for the dynamic similarity index. For the mean phase coherence, an average sensitivity of 89% is achieved.

Patient individual performance is shown in Fig. 1. The central bar represents the detected sensitivity, whereas the left and right bars represent 5% significance levels. The significance levels are calculated on the basis of an unspecific random predictor, which has been adapted to be suitable for the seizure-prediction characteristic (Schelter et al., 2006). Two critical sensitivity values exist, as it is usually unknown whether the data in two different electrode contacts can be assumed to be statistically independent. The left bar, representing the lower critical sensitivity value, is obtained for complete dependence. The right bar, representing the upper critical sensitivity value, is obtained for complete independence. As discussed, sensitivity can definitely be considered superior to a random predictor, if it is higher than the upper critical sensitivity. Because a sufficient amount of false predictions is required and to apply a common parameter set for all patients to ensure comparability of the results, the desired prediction performance in this study is achieved when sensitivities are at least above the lower critical value and comparable to the upper critical value. The localizations of the electrode contacts for the dynamic similarity index and the electrode combinations for the mean phase coherence thereof selected by the algorithms as yielding highest prediction performance are reported in Table 2.

To examine the false predictions, their times of occurrence and especially circadian dependencies are investigated. Furthermore, to analyze the relation of false predictions to epileptic activity and vigilance, a visual inspection of scalp and invasive EEG data is performed. Finally, the choice of the reference interval for the dynamic similarity index and its effects on the emergence of false predictions is analyzed.

False predictions depending on the time of day

Time points of false predictions for both prediction methods during 24 h are shown for patient 15 in Fig. 2. For the maximum number of allowed false predictions of 12 within 1 day, altogether six false predictions emerge for the mean phase coherence and 10 false predictions for the dynamic similarity index.

To statistically evaluate circadian dependencies for all patients, including those without scalp EEG recordings, nighttime is considered to be the period between 9 pm and 9 am. Patients are usually asleep during major parts of this period. For the patient in question, all false predictions for the mean phase coherence occur during night. For the dynamic similarity index, two seizures are falsely predicted at ~1 pm and at ~6 pm. The remaining eight false predictions occur again during the night.

In Fig. 3, histograms of the occurrences of false predictions for a given interval of 1 h duration are shown for all 21 patients, for the dynamic similarity index in (a) and the mean phase coherence in (b). Again, the period between 9 p.m. and 9 a.m. during which patients are usually sleeping is marked by the gray area. The total number of false predictions amounts to 103 during the night compared with 38 during the day for the dynamic similarity index (in total, 141) and 78 compared with 40 for the mean phase coherence (in total, 118).

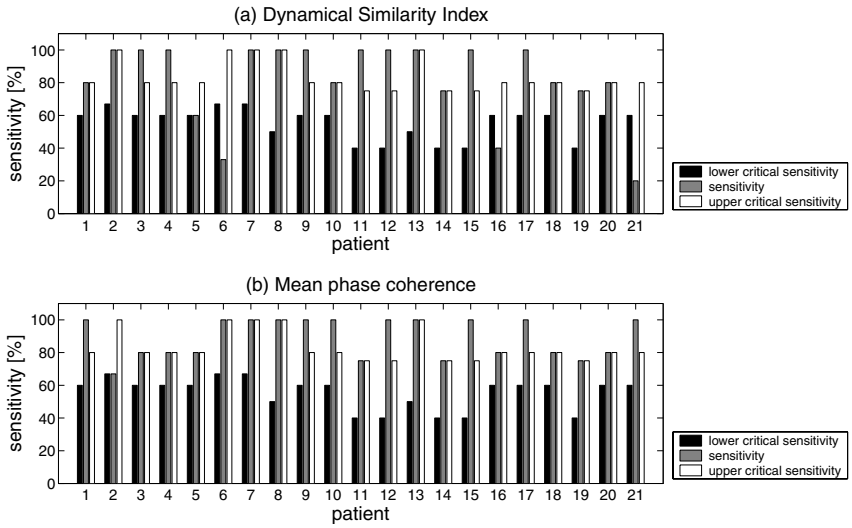

FIG. 1. Patient individual seizure-prediction performance. The central bar represents the detected sensitivity of the dynamic similarity index and the mean phase coherence. The left bars correspond to the lower critical sensitivity values calculated for a 5% significance level, assuming complete dependence between the features. The right bars correspond to the upper critical sensitivity values calculated for a 5% significance level, assuming complete independence between the features. The critical sensitivity values are calculated on the basis of an unspecific random predictor.

The difference in the number of false predictions during night and daytime is statistically significant for the dynamic similarity index ($p < 0.01$), validated by an exact two-sided Wilcoxon signed rank test with respect to each patient. The difference is not statistically significant for the mean phase coherence ($p = 0.07$).

Visual inspection of scalp EEG data

Based on the circadian dependency of false predictions of one of the two algorithms under investigation (i.e., the dynamic similarity index), we hypothesized that different stages of vigilance are associated with changes in the extracted feature. As the period between 9 pm and 9 am, which was considered "nighttime" in the previous investigation, is rather long compared with the approximate 8 h patients sleep on average, a more refined analysis of the state of vigilance at the times of false predictions is performed in the following. Sleep stages are classified in the subgroup of 10 patients.

In Fig. 4, the numbers of false predictions occurring during rapid eye movement (REM) sleep, non-REM sleep, and wakefulness, respectively, are shown for each patient. For the dynamic similarity index, 24% of all 72 false predictions occurred while the patient was awake, 65% during non-REM sleep, and 8% during REM sleep. For the mean phase coherence, 4% of all 49 false predictions occurred during REM sleep. False predictions emerge with the same frequency during non-REM sleep and while the patients

TABLE 2. *Electrode contacts for the dynamic similarity index and electrode combinations for the mean phase coherence yielding highest prediction performance*

			Dynamical similarity index		Mean phase coherence		
Patient	H/NC	Origin	Focal contact	Extra-focal	Focal/focal	Focal/extra-focal	Extra-focal/extra-focal
1	NC	Frontal	X			X	
2	H	Temporal	X			X	
3	NC	Frontal	X		X		
4	H	Temporal	X			X	
5	NC	Frontal		X		X	
6	H	Temporo-occipital		X		X	
7	H	Temporal		X		X	
8	NC	Frontal		X		X	
9	NC	Temporo-occipital	X			X	
10	H	Temporal		X			X
11	NC	Parietal		X		X	
12	H	Temporal	X				X
13	H	Temporo-occipital		X	X		
14	H and NC	Fronto-temporal	X			X	
15	H and NC	Temporal		X		X	
16	H	Temporal				X	
17	NC	Temporal		X		X	
18	NC	Frontal		X		X	
19	NC	Frontal		X	X		
20	NC	Temporoparietal	X			X	
21	NC	Temporal	X		X		
Total			10	11	4	15	2

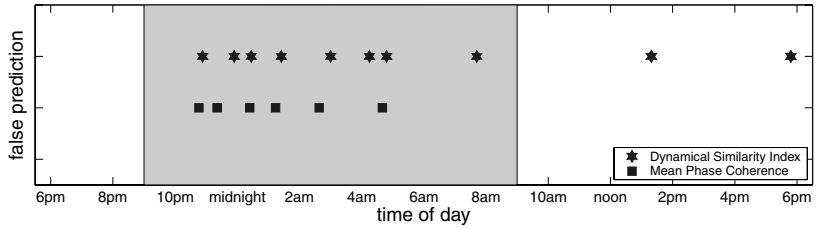

FIG. 2. Times of false predictions during a continuous interictal EEG recording of 24 h duration. For both seizure-prediction methods investigated, times of false predictions are shown for patient 15 for FPR_{max}, 0.5 FP/h; SOP, 30 min; and SPH, 2 min. The gray area marks the period between 9 pm and 9 am and is defined as nighttime, comprising the period during which patients are usually asleep. For the mean phase coherence, all six false predictions occurred during this time period. Only two of 10 false predictions occurred between noon and 6 pm for the dynamic similarity index.

are awake (47% each) for the mean phase coherence. Because of missing scalp EEG data at the corresponding time points, states of vigilance were not assessable for two false predictions for the dynamic similarity index and one false prediction for the mean phase coherence, respectively. The results for both algorithms and a detailed classification of sleep stages are reported in Tables 3 and 4. The majority of false predictions in non-REM sleep are associated with sleep stage II.

Considering that, on average, patients are asleep 8 hours per day and are awake for the remaining time, the a priori chance of false predictions occurring during a period when patients are awake is twice as high as during a period when patients are asleep. Renormalization to the different duration of the time periods "asleep" (REM plus non-REM) and "awake" leads to the result that 86% of all false predictions emerge during REM and non-REM for the dynamic similarity index and 68% for the mean phase coherence.

The results have been tested for their statistical significance by using an exact two-sided Wilcoxon signed rank test with respect to each patient. The results for the dynamic similarity index are statistically significant ($p < 0.05$). For the mean phase coherence, the results are non-significant ($p > 0.05$).

Furthermore, a short interval of 1 min in advance of each false prediction has been examined with respect to changes in the state of vigilance, which are, for example, arousals or long-lasting changes from the sleep stage to awake stage or vice versa. For the mean phase coherence, 22% of false predictions are associated with a change of the state of vigilance versus 21% for the dynamic similarity index (cf. Tables 3 and 4).

Visual inspection of invasive EEG data

Besides investigations of dependencies on different states of vigilance, invasive EEG data are examined at the times of false predictions with respect to interictal epileptic activity, such as interictal spikes and sharp waves, subclinical electroencephalographic ictal activity, and artifacts that might potentially lead to false predictions.

A 1 min time interval in advance of each false prediction is inspected, as artifacts and interictal activity within this time interval might affect a subsequent prediction. Only 5% of false predictions could possibly be related to artifacts for the dynamic similarity index, and 6.7% for the mean phase coherence, respectively.

Interictal epileptic activity was present during a major part of the segments in which false predictions have been observed. This interictal activity is not necessarily restricted to the periods associated with false predictions, because interictal activity is also observed in periods not related to false predictions.

Finally, time intervals after each false prediction of 32-min duration, corresponding to the sum of SPH and SOP,

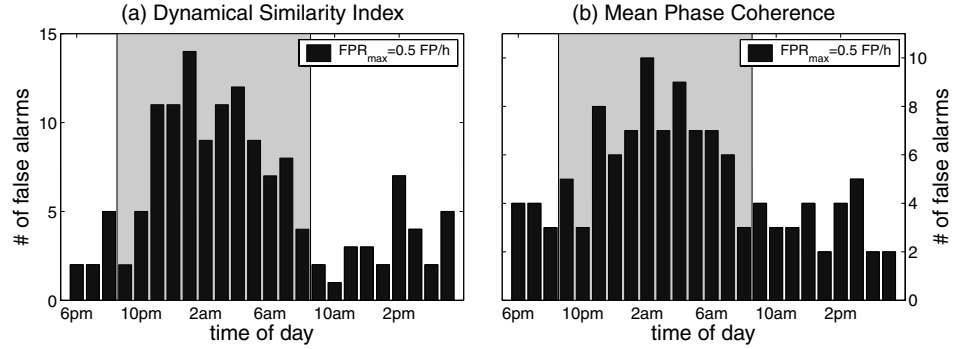

FIG. 3. Histograms of the circadian dependencies for number of false predictions for all 21 patients and both prediction methods. The gray area marks the period between 9 p.m. and 9 a.m., considered as nighttime. For the dynamic similarity index, 73% of all false predictions, and for the mean phase coherence, 66% of all false predictions occurred during the period between 9 p.m. and 9 a.m.

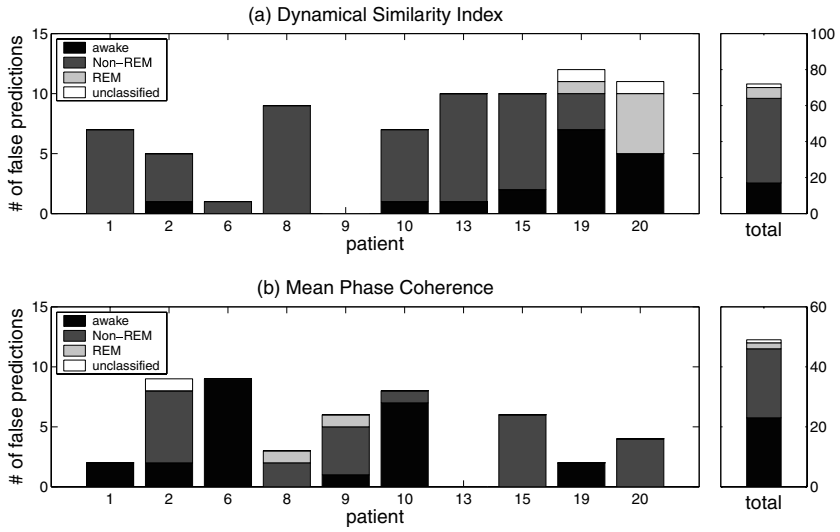

FIG. 4. Incidence of false predictions depending on the states of vigilance for 10 patients with simultaneous scalp EEG recordings, which allow assessment of sleep stages. For the dynamic similarity index, 24% of all 72 false predictions occurred while the patient was awake, 65% during non-REM sleep, and 8% during REM sleep. For the mean phase coherence, 47% of all 49 false predictions occurred while the patient was awake, 47% during non-REM sleep, and 4% during REM sleep. The state of vigilance could not be determined for one false prediction for the mean phase coherence and for two false predictions for the dynamic similarity index because of missing scalp EEG data at the corresponding time points.

are investigated with respect to subclinical epileptic activity. The results are given in Fig. 5 for the dynamic similarity index (a) and the mean phase coherence (b) for all 21 patients individually. Of all 141 false predictions, 14 (9.9%) are followed within the seizure occurrence period by a subclinical seizure for the dynamic similarity index and 20 (16.9%) of all 118 false predictions for the mean phase coherence. Additionally, seven of 14 subclinical seizures belong to multiple and repeated subclinical activity for the dynamic similarity index, and six of 20 subclinical seizures for the mean phase coherence.

Dynamic reference intervals

Analysis of the states of vigilance has shown that a vast majority of false predictions for the dynamic similarity index occur during non-REM sleep (see section entitled "Visual inspection of Scalp EEG data"). The calculation of the dynamic similarity index is based on a comparison of the dynamics of the EEG data within a fixed reference interval. The reference interval has been chosen at the beginning of the interictal recording period according to Le van Quyen et al. (1999). In the following analyses, the effect of the state of vigilance during the reference interval on the emergence of false predictions is investigated.

For the subgroup of 10 patients with additional scalp EEG recordings, two reference intervals of 5-min duration each were selected. One reference interval was chosen during a period while the patient was awake. The second reference interval was chosen during a period while the patient was asleep (non-REM). For both reference intervals, the dynamic similarity index was calculated and assessed by using the seizure prediction characteristic for $FPR_{max} = 0.5$ FP/h; SOP = 30 min; and SPH = 2 min. By visual inspection of scalp EEG recordings, states of vigilance were determined at the time points at which false predictions emerged.

In Fig. 6, results are given for the reference interval during wakefulness (a) and during non-REM sleep (b). For instance, for patient 10 and the reference interval during the awake state, all seven false predictions occurred during non-REM sleep (cf. Fig. 6a). This number is reduced to two false predictions for the reference state "non-REM"

TABLE 3. *Detailed results for the classification of sleep stages for the dynamic similarity index*

Patient	Total number of false predictions	Number of false predictions during								# of fps associated with change in vigilance
		Awake	Non-REM 1	Non-REM 2	Non-REM 3	Non-REM 4	Sum Non-REM	REM	No scalp	
1	7	0	0	5	1	1	7	0	0	2
2	5	1	0	1	3	0	4	0	0	0
6	1	0	0	0	0	1	1	0	0	0
8	9	0	2	7	0	0	9	0	0	2
9	0	0	0	0	0	0	0	0	0	0
10	7	1	1	5	0	0	6	0	0	0
13	10	1	0	6	2	1	9	0	0	2
15	10	2	1	7	0	0	8	0	0	1
19	12	7	0	3	0	0	3	1	1	1
20	11	5	0	0	0	0	0	5	1	1
Total	72	17	4	34	6	3	47	6	2	15
Percentage		24%	6%	47%	9%	4%	65%	9%	3%	21%

TABLE 4. *Detailed results for the classification of sleep stages for the mean phase coherence*

Patient	Total number of false predictions	Number of false predictions during								# of fps associated with change in vigilance
		Awake	Non-REM 1	Non-REM 2	Non-REM 3	Non-REM 4	Sum Non-REM	REM	No scalp	
1	2	2	0	0	0	0	0	0	0	0
2	9	2	1	5	0	0	8	0	1	7
6	9	9	0	0	0	0	0	0	0	1
8	3	0	0	2	0	0	2	1	0	0
9	6	1	3	0	1	0	4	1	0	2
10	8	7	1	0	0	0	1	0	0	1
13	0	0	0	0	0	0	0	0	0	0
15	6	0	0	6	0	0	6	0	0	0
19	2	2	0	0	0	0	0	0	0	0
20	4	0	1	1	2	0	4	0	0	0
Total	49	23	6	14	3	0	23	2	1	11
Percentage		47%	12%	29%	6%	0%	47%	4%	2%	22%

(cf. Fig. 6b). Rather an opposite dependency is observed, as six false predictions occur during wakefulness for the reference state "non-REM."

For the reference state "awake," 53 of 70 false predictions emerge during non-REM sleep, 12 during wakefulness and two during REM sleep. The state of vigilance could not be determined for three false predictions, as no scalp EEG recordings were available at the corresponding time points. For the reference state "non-REM," 40 of 72 false predictions emerged during non-REM sleep, 23 during awake and seven during REM sleep. No scalp EEG recordings were available at the time points of two false predictions.

Considering the difference in the duration of the periods patients are asleep and awake, 88% of all false predictions occurred during sleep (non-REM plus REM) for the reference state "awake." For the reference state "non-REM," 79% of all false predictions emerged during sleep (non-REM plus REM).

The choice of the reference interval influences the emergence of false predictions. The following analysis investigates whether the number of false predictions can be reduced by combining the algorithms for both reference states at a minimal loss of correct predictions.

The combination follows the rule that, within a specific combination period, two alarm events (i.e., one raised by the algorithm based on the reference state awake and one raised by the algorithm based on the reference state non-REM) are required to trigger a correct or false prediction. In this study, the combination period was chosen with 30-min duration, according to the SOP.

Sensitivity values are shown in Fig. 7a, and the number of false predictions is shown in Fig. 7b for the individual algorithms based on the two reference states "awake" and "non-REM," as well as for the combination of both. The averaged sensitivity for the 10 patients investigated decreased from 88% (reference state awake) and 77% (reference state non-REM) to 65% for the combination.

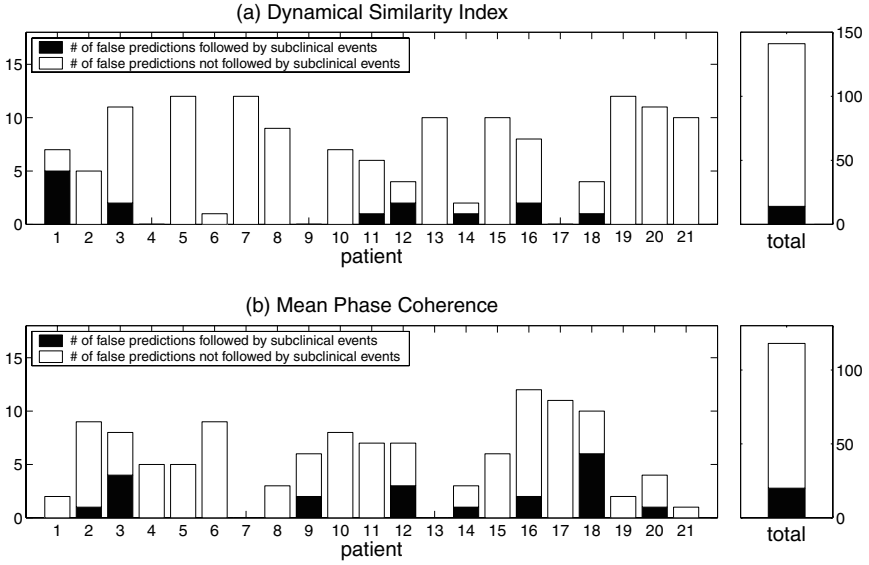

FIG. 5. Subclinical events within the seizure occurrence period of 30-min duration after each false prediction and a seizure prediction horizon of 2-min duration. For the dynamic similarity index, 10% of all 141 false predictions are followed by subclinical activity, and 17% of all 118 false predictions, for the mean phase coherence.

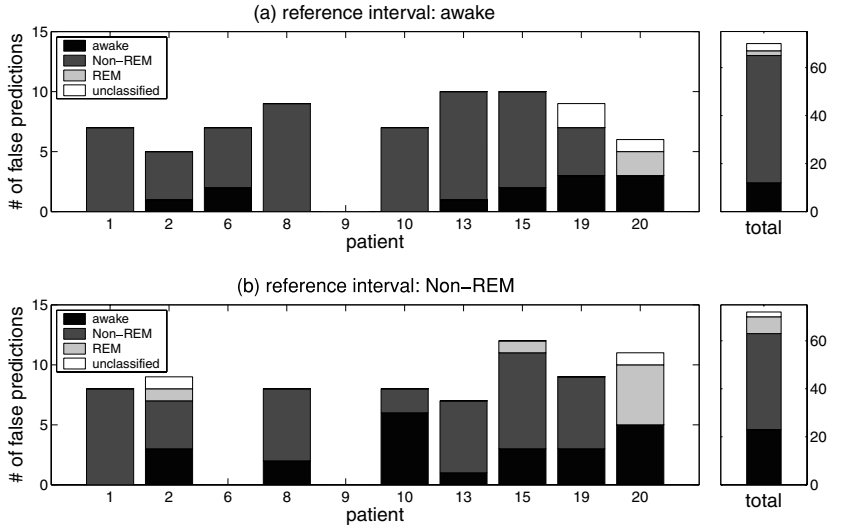

FIG. 6. Dynamic reference intervals. States of vigilance at the occurrence of false predictions for the dynamic similarity index with different reference intervals. In (a), a reference interval was chosen in an interictal period while the patient was awake. In (b), a reference interval was selected during an interictal period while the patient was asleep (non-REM).

For the combination of both reference states, the total number of false predictions decreased to 41. Compared with the algorithm based on the reference state awake, 29 false predictions were avoided. Compared with the algorithm based on the reference state non-REM, 31 false predictions were prevented.

Combining two reference intervals, one during sleep and one during the awake state, the dynamic similarity index increases its performance by reducing the number of false predictions by almost 50% without major changes in sensitivity.

DISCUSSION

In this study, the performance, temporal aspects, and causes for the emergence of false predictions for two seizure-prediction algorithms, the mean phase coherence and the dynamic similarity index, have been investigated. The performance for both algorithms could be shown to be superior to the lower critical sensitivity values and comparable to the upper critical values for both algorithms and almost all patients. Both algorithms thus showed predictive power for seizures. We mention, though, that different parameter combinations might yield better results, especially if the false-prediction rate is chosen much lower. The performance of the mean phase coherence is slightly better than the performance of the dynamic similarity index, which could be caused by the fact that the mean phase coherence uses the data of two channels simultaneously; the number of channel combinations is more than 2 times higher than the number of channels used by the dynamic similarity index.

We observed a high interindividual variability between the results for the patients for both algorithms. Limitations in the duration of the EEG recordings, different localizations of recording sites, and differences in patients' characteristics, including their AED during the monitoring period, may all contribute to this interindividual variability of the results.

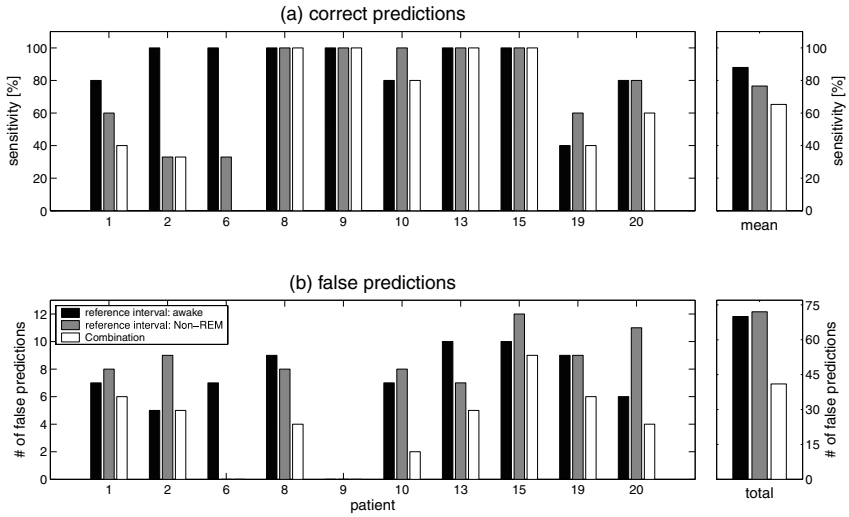

FIG. 7. Combination of two reference intervals. Correct predictions (a) and false predictions (b) for the dynamic similarity index with a reference interval while the patient was awake and while the patient was asleep (non-REM). Combining both reference intervals, a significant reduction of false predictions was possible at the expense of a few no longer correctly predicted seizures.

Interestingly, more false alarms were raised for the dynamic similarity index compared with the mean phase coherence evaluated with the same maximum false prediction rate for both algorithms. The fact that the dynamic similarity index is not reference free may contribute to this effect. Likewise, it could also be caused by the higher number of features for the mean phase coherence than for the dynamic similarity index. For the same sensitivity, the mean phase coherence can optimize the specificity with the aid of more features.

A dependency of false predictions on the time of day has been demonstrated. A statistically significantly higher number of false predictions for the dynamic similarity index occurred during a time period between 9 p.m. and 9 a.m., comprising the night and the period when patients are usually sleeping.

Motivated by this observation, an analysis of the states of vigilance by visual inspection of scalp EEG recordings was performed for a subgroup of 10 patients. Considering the absolute number of false predictions, the majority are associated with non-REM sleep for the dynamic similarity index. For the mean phase coherence, false predictions during non-REM sleep and during periods while the patients are awake emerge with the same frequency. For both prediction methods, false predictions during REM sleep were observed for only two patients.

The background activity may be central in explaining false predictions as opposed to the state of vigilance. Non-REM sleep is accompanied by an increase in background delta power. We thus subclassified non-REM sleep phases that showed that most false prediction occurred during non-REM II but not during deep sleep phases for both algorithms. This strongly suggests that changes in the spectrum of background activity alone, like an increase in delta power, are not the only contributing factor to the generation of false predictions.

The durations of the periods patients are asleep and patients are awake are not identical. Renormalization to the duration of the corresponding periods shows a statistically significant increase in false predictions during sleep (REM plus non-REM) for both algorithms.

Taking dependencies on the state of vigilance of false predictions into account may offer opportunities for an improvement of prediction methods. A seizure during the night may not be as harmful as a seizure during the day, because patients are usually under safe conditions while sleeping. In addition, critical values or parameters of prediction methods can be adapted during non-REM sleep, when more false predictions occur. This leads to lower sensitivity, as, for example, the extracted features are no longer able to cross the adapted thresholds, even in the case of an emerging seizure. However, the number of false predictions would decrease tremendously.

By visual inspection of the corresponding invasive EEG recordings at the time points of false predictions, potential explanations for false predictions have been investigated. It has been shown that artifacts are of minor importance as causes for false predictions. This result may be due to the low rate of artifacts present in invasive EEG recordings or to the robustness of the analyzed algorithms.

A further focus of the visual inspection of the invasive EEG data was used to investigate electroencephalographic ictal but subclinical activity after a false prediction. If false predictions somehow predicted subclinical seizures, it could be argued that the algorithms detect common physiologic changes underlying clinically manifest seizures as well as underlying subclinical seizures. It may therefore be debated whether these predictions should be classified as false predictions. From a therapeutic point of view, for instance, if a seizure-warning device were applied, a distinction between subclinical and clinical ictal activity could be desirable. Again, only a small number of false predictions are associated with subclinical events within a time interval corresponding to the SOP.

The dependency on the state of vigilance has been shown to be more pronounced for the dynamic similarity index than for the mean phase coherence. The mean phase coherence is based on detecting changes of the interactions between the dynamics measured by two electrode contacts simultaneously. In contrast, the dynamic similarity index compares the present dynamics of one electrode contact with a fixed reference interval. The effect of the reference interval on the emergence and circadian dependency of the false predictions was analyzed by applying two different reference intervals. One reference interval was chosen in a period when the patients were awake, and the second reference interval was selected during non-REM sleep.

For both reference states "awake" and "non-REM," the majority of all false predictions raised by the dynamic similarity index occurred during sleep. This effect was more pronounced if the reference interval was chosen in a period when the patient is awake. Although the major number of false predictions occurred during sleep for the reference state chosen during "non-REM" sleep, the number of false predictions during periods while patients were wake increased. This result strongly indicates that a subgroup of false predictions for the dynamic similarity index exists, caused by the difference in the state of vigilance between the reference interval and the current sliding window.

Combining the algorithm of the dynamic similarity index for different reference states prevented many false predictions of this special subgroup. Overall, almost one half of all false predictions could be avoided. However, a minor loss of sensitivity has to be accepted. For a few patients investigated, the results for the combination are promising. For instance, for patient 10, the total number of false predictions decreased dramatically for the combination

at constant sensitivity, compared with the algorithm using only the reference state, "awake."

Future research will be devoted to the combination of several different reference states for the dynamic similarity index. The combination of algorithms introduces more parameters that must be adjusted. For instance, the "combination interval" in our investigation was chosen according to the SOP. As this period is rather long, the effect of shorter "combination intervals" remains to be investigated.

In summary, we showed a dependency of the occurrence of false predictions on the time of day and on the state of vigilance for two seizure-prediction methods, by investigating long-term interictal invasive EEG recordings of 24 h duration for each patient. The dependency is more pronounced for the dynamic similarity index. The majority of false predictions occurred during sleep. This may turn out to be advantageous, as false predictions may be less problematic during sleep. Dependencies on circadian changes in the EEG dynamics may offer clues for time-dependent adaptations of thresholds or parameters of the prediction algorithms. A strategy for the reduction of false predictions has been proposed for the dynamic similarity index by combining different reference states. The combination yielded promising results for several patients investigated and may offer opportunities to further increase the performance of seizure-prediction methods.

Acknowledgment: We to thank Carolin Gierschner and the team of EEG technicians at the Epilepsy Center Freiburg for their technical support, especially concerning the extensive data acquisition and handling. This work was supported by the German Federal Ministry of Education and Research (BMBF grant 01GQ0420) and the German Science Foundation (Ti 315/2-1). The data analyzed in this study are available on request: Web page: http://www.fdm.uni-freiburg.de/EpilepsyData E-mail: epilepsydatabase@fdm.uni-freiburg.de.

REFERENCES

Aschenbrenner-Scheibe R, Maiwald T, Winterhalder M, Voss HU, Timmer J, Schulze-Bonhage A. (2003) How well can epileptic seizures be predicted? An evaluation of a nonlinear method. *Brain* 126:2616–2626.

Buck D, Baker GA, Jacoby A, Smith DF, Chadwick DW. (1997) Patients' experiences of injury as a result of epilepsy. *Epilepsia* 38:439–444.

Elger CE, Lehnertz K. (1998) Seizure prediction by nonlinear time series analysis of brain electrical activity. *European Journal of Neuroscience* 10:786–789.

Elger CE. (2001) Future trends in epileptology. *Current Opinion in Neurology* 14:185–186.

Iasemidis LD, Sackellares JC, Zaveri HP, Williams WJ. (1990) Phase space topography and the Lyapunov exponent of electrocorticograms in partial seizures. *Brain Topography* 2:187–201.

Jerger KK, Netoff TI, Francis JT, Sauer T, Pecora L, Weinstein SL, Schiff SJ. (2001) Early seizure detection. *Journal of Clinical Neurophysiology* 18:259–268.

Lehnertz K, Elger CE. (1995) Spatio-temporal dynamics of the primary epileptogenic area in temporal lobe epilepsy characterized by neuronal complexity loss. *Electroencephalography in Clinical Neurophysiology* 95:108–117.

Lehnertz K, Elger CE. (1998) Can epileptic seizures be predicted? Evidence from nonlinear time series analysis of brain electrical activity. *Physical Review and Letters* 80:5019–5022.

Le Van Quyen M, Martinerie J, Baulac M, Varela F. (1999) Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings. *Neuroreport* 10:2149–2155.

Le Van Quyen M, Adam C, Martinerie J, Baulac M, Clemenceau S, Varela F. (2000) Spatio-temporal characterizations of non-linear changes in intracranial activities prior to human temporal lobe seizures. *European Journal of Neuroscience* 12:2124–2134.

Le Van Quyen M, Martinerie J, Navarro V, Boon P, D'Have M, Adam C, Renault B, Varela F, Baulac M. (2001a) Anticipation of epileptic seizures from standard EEG recordings. *Lancet* 357:183–188.

Le Van Quyen M, Martinerie J, Navarro V, Baulac M, Varela F. (2001b) Characterizing neurodynamic changes before seizures. *Clinical Neurophysiology* 18:191–208.

Litt B, Esteller R, Echauz J, D'Alessandro M, Shor R, Henry T, Pennell P, Epstein C, Bakay R, Dichter M, Vachtsevanos G. (2001) Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron* 30:51–64.

Litt B, Echauz J. (2002) Prediction of epileptic seizures. *Lancet Neurology* 1:22–30.

Litt B, Lehnertz K. (2002) Seizure prediction and the preseizure period. *Current Opinion in Neurology* 15:173–177.

Maiwald T, Winterhalder M, Aschenbrenner-Scheibe R, Voss HU, Schulze-Bonhage A, Timmer J. (2004) Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic. *Physical Review D* 194:357–368.

Mormann F, Lehnertz K, David P, Elger CE. (2000) Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Physical Review D* 144:358–369.

Mormann F, Andrzejak RG, Kreuz T, Rieke C, David P, Elger CE, Lehnertz K. (2003a) Automated detection of a pre-seizure state based on a decrease in synchronization in intracranial EEG recordings from epilepsy patients. *Physical Review E* 67:021912.

Mormann F, Kreuz T, Andrzejak RG, David P, Lehnertz K, Elger CE. (2003b) Epileptic seizures are preceded by a decrease in synchronization. *Epilepsy Research* 53:173–185.

Mormann F, Kreuz T, Rieke C, Andrzejak RG, Kraskov A, David P, Elger CE, Lehnertz K. (2005) On the predictability of epileptic seizures. *Clinical Neurophysiology* 116:569–587.

Navarro V, Martinerie J, Le Van Quyen M, Clemenceau S, Adam C, Baulac M, Varela F. (2002) Seizure anticipation in human neocortical partial epilepsy. *Brain* 125:640–655.

Navarro V, Martinerie J, Le Van Quyen M, Baulac M, Dubeau F, Gotman F. (2005) Seizure anticipation: do mathematical measures correlate with video-EEG evaluation? *Epilepsia* 46:385–396.

Osorio I, Frei MG, Wilkinson SB. (1998) Real-time automated detection and quantitative analysis of seizures and short-term prediction of clinical onset. *Epilepsia* 39:615–627.

Osorio I, Harrison MAF, Lai YC, Frei MG. (2001) Observations on the application of the correlation dimension and correlation integral to the prediction of seizures. *Journal of Clinical Neurophysiology* 18:269–274.

Rechtschaffen A, Kales A. (1968) *A manual of standardized terminology, technics, and scoring system for sleep stages of human subjects*. Government Printing Office, Washington, DC.

Rosenblum MG, Pikovsky AS, Kurths J. (1996) Phase synchronization of chaotic oscillators. *Physical Review and Letters* 76:1804–1807.

Rosenblum M, Pikovsky A, Kurths J, Schäfer C, Tass PA. (2001) Phase synchronization: from theory to data analysis. In Moss F, Gielen S (Eds) *Handbook of biological physics,* Vol. 4 *of Neuro-informatics*. Elsevier, Amsterdam, pp. 279–321.

Schachter SC. (1994) *The brainstorms companion: epilepsy in our view*. Raven Press, New York.

Schelter B, Winterhalder M, Maiwald T, Meier-Hirmer C, Brandt A, Schulze-Bonhage A, Timmer J. (2004) Circadian rhythm in epileptic seizure prediction. *Epilepsia* 45:172.

Schelter B, Winterhalder M, Maiwald T, Brandt A, Schad A, Schulze-Bonhage A, Timmer J. (2006) Testing statistical significance of

multivariate time series analysis techniques for epileptic seizure prediction. *Chaos* 16:103–108.

Schindler K, Wiest R, Kollar M, Donati F. (2002) EEG analysis with simulated neuronal cell models helps to detect pre-seizure changes. *Clinical Neurophysiology* 113:604–614.

Winterhalder M, Maiwald T, Voss HU, Aschenbrenner-Scheibe R, Timmer J, Schulze-Bonhage A. (2003) The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods. *Epilepsy and Behavior* 4:318–325.