NeuroImage 158 (2017) 99–111

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Information-theoretical analysis of resting state EEG microstate sequences - non-Markovianity, non-stationarity and periodicities

F. von Wegner [a,b,*], E. Tagliazucchi [b,c], H. Laufs [b,c]

[a] Epilepsy Center Rhein-Main, Goethe University Frankfurt, Schleusenweg 2-16, 60528 Frankfurt am Main, Germany
[b] Department of Neurology and Brain Imaging Center, Goethe University Frankfurt, Schleusenweg 2-16, 60528 Frankfurt am Main, Germany
[c] Department of Neurology, Christian-Albrechts University Kiel, Arnold-Heller-Strasse 3, 24105 Kiel, Germany

ARTICLE INFO

Keywords:
EEG
Resting state
Microstates
Information theory
Non-Markov process
Stationarity
Periodicity

ABSTRACT

We present an information-theoretical analysis of temporal dependencies in EEG microstate sequences during wakeful rest. We interpret microstate sequences as discrete stochastic processes where each state corresponds to a representative scalp potential topography. Testing low-order Markovianity of these discrete sequences directly, we find that none of the recordings fulfils the Markov property of order 0, 1 or 2. Further analyses show that the microstate transition matrix is non-stationary over time in 80% (window size 10 s), 60% (window size 20 s) and 44% (window size 40 s) of the subjects, and that transition matrices are asymmetric in 14/20 (70%) subjects. To assess temporal dependencies globally, the time-lagged mutual information function (autoinformation function) of each sequence is compared to the first-order Markov model defined by the classical transition matrix approach. The autoinformation function for the Markovian case is derived analytically and numerically. For experimental data, we find non-Markovian behaviour in the range of the main EEG frequency bands where distinct periodicities related to the subject's EEG frequency spectrum appear. In particular, the microstate clustering algorithm induces frequency doubling with respect to the EEG power spectral density while the tail of the autoinformation function asymptotically reaches the first-order Markov confidence interval for time lags above 1000 ms. In summary, our results show that resting state microstate sequences are non-Markovian processes which inherit periodicities from the underlying EEG dynamics. Our results interpolate between two diverging models of microstate dynamics, memoryless Markov models on one side, and long-range correlated models on the other: microstate sequences display more complex temporal dependencies than captured by the transition matrix approach in the range of the main EEG frequency bands, but show finite memory content in the long run.

1. Introduction

The most prominent features of resting state surface EEG recordings are ongoing, amplitude-modulated oscillations across the frequency range of approximately 0.5− 70 Hz (Niedermeyer and da Silva, 2005). EEG data is often analyzed with respect to this oscillatory activity. For instance, we may be interested in the spectral power of a given frequency band, phase relationships, or the dynamics of the oscillation's envelope, to name but a few. The spatial distribution of the scalp potential varies over time, showing episodes of stability alternating with short transition episodes between certain quasi-stable EEG topographies (Wackermann et al., 1993). Applying data compression techniques, stable topographies can be clustered into sets of a few maps maximizing the global explained variance (GEV) (Wackermann et al., 1993; Murray et al., 2008). These maps are fitted competitively into the original EEG time series using a maximum correlation criterion at each time step. The resulting microstate sequence at each time step contains the microstate label whose map has the maximum absolute correlation with the EEG topography at that time point. Alternative implementations have been proposed (Koenig et al., 1999). Microstates denote quasi-stable episodes corresponding to a single representative map, with an average duration in the range of 10–100 ms (Koenig et al., 2002; Brodbeck et al., 2012). Many studies have shown a set of $n = 4$ microstates to be optimal (Murray et al., 2008; Brodbeck et al., 2012) although other cluster numbers have been described in healthy resting state (Yuan et al., 2012) and pathological conditions (Koenig et al., 1999; Kuhn et al., 2015). Here, we mostly use $n = 4$ microstates, however, all methods presented here can be used for any number of states.

* Corresponding author. Epilepsy Center Rhein-Main, Goethe University Frankfurt, Schleusenweg 2-16, 60528 Frankfurt am Main, Germany.
E-mail address: fredericvon.wegner@kgu.de (F. von Wegner).

http://dx.doi.org/10.1016/j.neuroimage.2017.06.062
Received 23 January 2017; Received in revised form 19 June 2017; Accepted 22 June 2017
Available online 30 June 2017

The most common approach to microstate analysis is the transition matrix approach. Using this straightforward approach, the transition probabilities between microstate maps are estimated from the empirical sequence of map labels and the resulting matrix is normalized row-wise in order to yield a stochastic matrix. Subsequently, different experimental conditions can be related to changing matrix entries (Koenig et al., 2002; Brodbeck et al., 2012; Kuhn et al., 2015). On a conceptual level, the transition matrix approach implies a (first-order) Markovian model as the information flow over time is summarized by the conditional probability of the future state $x_{t+1} = S_j$, given the current state $x_t = S_i$. Therefore, temporal dependencies more complex than first-order Markov models cannot be captured by this approach. Moreover, calculating a fixed transition matrix for a given data set cannot model transition dynamics changing over time, i.e. non-stationarities. Unfortunately, the Markov property is almost never tested for, an exception being the microstate duration analysis published by (Wackermann et al., 1993). Testing the geometric distribution of microstate durations for short EEG time series up to a duration of 16 s, the Markov property could not be rejected in 22/24 data sets (Wackermann et al., 1993). For longer time series of at least several minutes, as usually recorded in resting state experiments, we are not aware of any formal tests of low-order Markov properties.

As an alternative analysis tool, Hurst exponent estimation has recently been introduced for microstate sequences, with the aim to find possible long-range dependencies (LRD) (Van de Ville et al., 2010; Gschwind et al., 2015). In order to estimate the Hurst exponent, the n-state symbolic sequence has to be mapped to a metric space $\{-1, +1\}$ using a partition of the state space (Van de Ville et al., 2010). The technique is inspired by LRD analyses of 4-state DNA sequences (Peng et al., 1992). However, the following two questions remain unanswered so far: a) which is the correct state space partition, i.e. which EEG topographies should be lumped into one group, and b) how can the technique be applied to arbitrary cluster numbers? Moreover, the resulting Hurst exponents are difficult to interpret as a Hurst exponent of $H > 0.5$ does not automatically imply long-range dependencies, especially in the case of non-stationary signals (McCauley et al., 2007; Riley et al., 2012). We recently reported an excessive proportion of false-positive LRD results when comparing experimental data with short-range correlated Markov models (von Wegner et al., 2016).

The aim of the current study is to systematically analyze temporal dependencies of microstate sequences and to characterize these sequences in the language of stochastic processes. Without making assumptions on symbol distributions and their temporal dependencies, we compute estimates of various information-theoretical quantities. Using this approach, we avoid the need to partition the microstate set and to project the sequences onto a metric space. All quantities used can be computed for arbitrary state spaces, i.e. for any number of microstates.

Our analysis starts on the shortest time scales, assessing low-order Markov properties (order 0– 2) directly, based on well-established statistical tests (Kullback, 1959; Kullback et al., 1962). We perform an additional first-order Markovianity test using a method presented in the context of microstate research (Wackermann et al., 1993). Next, the transition matrix is tested for time-stationarity, also termed conditional homogeneity in (Kullback, 1959; Kullback et al., 1962), and for symmetry. Finally, global temporal dependencies up to time lags of 2000 ms are analyzed using the time-lagged mutual information function (auto-information function). Distinct periodicities found there are further tested for robustness with respect to individual microstate maps and to cluster numbers.

The information-theoretical approach presented here lets us conclude that microstate properties show a behaviour somewhere between memoryless Markov models and possibly long-range correlated random walk models, adding the unique feature of reflecting the underlying EEG periodicities.

2. Material & methods

2.1. Experimental data

A set of EEG recordings from 20 right-handed healthy subjects during wakeful rest (age range: 19– 27, mean age: 23 yrs) was recorded in an eyes-closed, wakeful rest condition. Selection criteria for the recordings used were steady, prominent alpha oscillations in the parieto-occipital EEG channels, and the absence of artefacts (eye blink, muscle, electrode artefacts) or signs of drowsiness. The 30 channel EEG data sets were acquired at a sampling rate of 5 kHz using the standard 10– 10 electrode configuration. Data were band-pass filtered to 1– 40 Hz, down-sampled to 250 Hz and re-referenced to an average reference. Power spectral densities were computed with Welch's method using a segment length of 1024 samples, 50% overlap and a Hanning window. EEG recordings have a total duration ranging from 100 to 312 s, corresponding to lengths of 25000-78000 samples. Written informed consent was obtained from all subjects, and the study was approved by the ethics committee of the Goethe University, Frankfurt, Germany.

2.2. Microstate analysis

EEG microstates were computed using the modified K-means algorithm described in (Murray et al., 2008) and illustrated in Fig. 1. Fig. 1 shows a segment of resting-state EEG for a selection of channels as indicated to the left. In the first step, the global field power time series (GFP, blue line) is computed as the spatial standard deviation of the EEG topography at each given time. At local GFP maxima (red dots), the spatial configuration of the EEG is considered stable and explains most of the variance of the time series (Wackermann et al., 1993). Therefore, K-means clusters are initialized with EEG patterns drawn randomly from

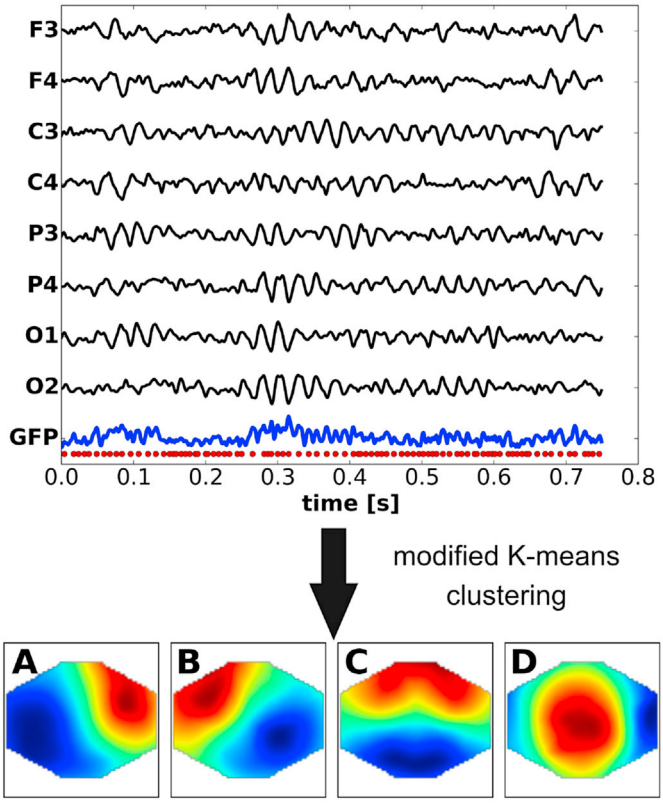

Fig. 1. Microstate segmentation: the top panel shows a section of resting state EEG (1–40 Hz, black lines) along with the resulting global field power (GFP, blue) and the local GFP maxima (red dots). For better visibility, only the EEG channels indicated to the left are shown. EEG topographies at local GFP maxima are clustered by the modified K-means algorithm to obtain the (n = 4) microstate maps labelled A-D (bottom panel).

the set of GFP peaks. During clustering, the algorithm ignores EEG polarity (Wackermann et al., 1993; Brodbeck et al., 2012; Kuhn et al., 2015). As K-means clustering is stochastic, the algorithm was run 25 times, and the optimum set of clusters was selected based on maximum global explained variance (Lehmann et al., 1987; Wackermann et al., 1993; Pascual-Marqui et al., 1995). As reported in numerous studies, an optimum number of 4 microstate maps was obtained (Brodbeck et al., 2012; Kuhn et al., 2015). An exemplary set of microstate maps along with their labels A-D is shown at the bottom of Fig. 1.

The sequence of microstate labels has the same length as the EEG time series. At each time point, the sequence shows the label of the microstate that has maximum spatial correlation with the instantaneous EEG topography (Koenig et al., 1999, 2002). This procedure is called "competitive back-fitting". We implemented the algorithm as detailed by (Murray et al., 2008) in Python 2.7.3. A C++ implementation for Windows is available from (Brunet et al., 2011) (Cartool Software, http://sites.google.com/site/cartoolcommunity). The microstate geometries we obtained and the cross-correlation properties between these maps are highly similar to the results obtained with the original Cartool software. The individual cluster geometries are shown in the supplementary data section (Supplementary Figs. S2–S5). To avoid modifications of the temporal structure of microstate sequences, no minimum microstate duration criteria or other temporal post-processing steps were applied to the microstate sequences.

2.3. Markov surrogate data

Microstate analysis yields a sequence x_t of microstates for each EEG data set. From this sequence, the empirical transition matrix $T_{ij} = P(x_{t+1} = S_j | x_t = S_i)$ and the equilibrium distribution π are calculated. As a first-order Markov process is uniquely defined by an initial distribution and its transition matrix, a Markov process with the quantities T and π identical to the empirical microstate sequence can be synthesized (Häggström, 2002). For each experimentally acquired microstate sequence x_t, we compute $n = 1000$ surrogate Markov sequences of the same length as x_t. The iterative 2-step algorithm contains an initialization function and an updating function. The initialization function yields a random initial state $x_0 \in S_0$, in accordance with the equilibrium distribution π. The updating function yields the subsequent value $x_{t+1} = S_j$ as a function of $x_t = S_i$ and the conditional transition probability T_{ij}. To implement both functions, we use random variables $r_t \sim U[0, 1]$, uniformly and independently distributed on the unit interval. The initial state index j is given by $\sum_{i=0}^{j-1}\pi_i \le r_0 < \sum_{i=0}^{j}\pi_i$. The state index j of $x_{t+1} = S_j$, the successor of state $x_t = S_i$, is given by the random variable r_t and the relation $\sum_{l=0}^{j-1}T_{il} \le r < \sum_{l=0}^{j}T_{il}$.

2.4. Information-theoretical analysis - motivation and basics

The microstate segmentation algorithm yields a set of representative EEG topographies, called microstate maps or simply maps, based on a criterion of maximum explained variance with respect to the underlying EEG data set. In our case of 30-channel EEG data, each of the four microstate maps is a unique array of 30 values, representing a given potential distribution on the skull surface, and labelled with one of the symbols A-D. This seriously limits the set of applicable time series algorithms. The usual linear characteristics computed for time series, such as the autocorrelation function and the power spectral density, are not directly applicable as they involve sums and products of the time series' values. These metric operations are not defined on the discrete set of states A-D, for which not even an ordering can be defined in an unequivocal way. As an alternative, a mapping of the microstate labels to a set of real numbers, e.g. $\{-1, +1\}$, has been proposed (Van de Ville et al., 2010). However, the assignment of a certain real number to each microstate map is a difficult choice. There is no apparent, biologically motivated relation between a microstate class and the real number that should represent it. Thus, the assignment represents an arbitrary choice and the analysis of all possible assignments in parallel results in an explosion of data size for larger numbers of microstates. Information-theoretical methods solve this dilemma by processing the map labels A-D directly, using the statistical properties of their distributions.

Fig. 2 illustrates the first steps of microstate sequence analysis. The sequence of microstate labels (Fig. 2A) yields the distribution of microstate labels across the microstate sequence $P(X_t = S_i)$, i.e. the probability to find the label $S_i \in \{A, B, C, D\}$ at time point t. This distribution can be characterized by the Shannon entropy H of the random variable X, $H = -\sum_i P(X_t = S_i)\log P(X_t = S_i)$ (Fig. 2B) (Kullback et al., 1962). Minimum entropy ($H = 0$) is attained by a delta distribution, assigning a probability of 1 to a given microstate S_i and a probability of 0 to all other microstates $S_{j\neq i}$. In other words, minimum entropy represents perfect predictability (minimum surprise) of the symbol S_i. The other extreme is a uniform distribution, in our case assigning $P(X_t = S_i) = \frac{1}{4}$ to each of the four microstates, resulting in the maximum entropy attainable for a four-state distribution, $H = \log(4)$. We will use the natural logarithm to the base e (Euler's constant) throughout the manuscript, denoting it as log. Information-theoretical quantities based on the natural logarithm are expressed as 'nats', just as the usual 'bits' refer to logarithms to the base 2. One nat denotes the amount of information (surprise) of an event with probability $\frac{1}{e}$.

To measure temporal dependencies within the microstate sequence, the transition matrix T is calculated (Fig. 2C). The matrix component T_{ij} contains the transition probability $P(X_{t+1} = S_j | X_t = S_i)$. To test the statistical significance of temporal dependencies, an independence hypothesis between the random variables X_t and X_{t+1} is tested. In particular, we test the null hypotheses of Markovianity (of order 0, 1 and 2), stationarity of the transition matrix, and symmetry of the transition matrix (Kullback et al., 1962). The statistics under the given null hypotheses are defined and the empirical sequences are tested against these. The discrepancy between the empirical distribution (p_i) and the distribution (q_i) expected under the null hypothesis is quantified by the Kullback-Leibler divergence $D(p, q) = -\sum_i p_i \log\frac{p_i}{q_i}$ (Kullback, 1959; Kullback et al., 1962). It is observed that $D(p, q)$ is not a metric in the mathematical sense as it is not symmetric, $D(p, q) \neq D(q, p)$, however $D(p, p) = 0$ and $D(p, q) \le D(p, r) + D(r, q)$. Statistical significance is assessed using G-tests and χ^2-statistics using classical convergence theorems (Anderson and Goodman, 1957; Kullback, 1959; Billingsley, 1961; Kullback et al., 1962). The test statistics are calculated from the microstate sequences directly and do not require Markov surrogate data.

In notation we follow (Kullback et al., 1962) and denote observed frequencies by f. The estimated probability of symbol S_i, denoted as p_i, is given by f_i, the number of observations of symbol S_i and the sample size n as $p_i = \frac{f_i}{n}$. Indices run over symbols and multiple sums are abbreviated by a single summation sign and the corresponding indices to sum over.

2.5. Zero-order Markov property

The null hypothesis is that information about the current state X_t does not affect the transition probabilities to the next state X_{t+1}, i.e. that $P(X_{t+1}|X_t) = P(X_{t+1})$. The observed number of transitions $X_t = S_i \to X_{t+1} = S_j$ is f_{ij}, f_i is the number of observations $X_t = S_i$ and f_j is the number of observations $X_{t+1} = S_j$. The length of the microstate sequence X_t is n. The G-test statistic

$$G_0 = 2\sum_{ij} f_{ij}\log\frac{nf_{ij}}{f_i f_j} \tag{1}$$

is asymptotically distributed as $G_0 \sim \chi^2_{(n_s-1)(n_s-1)}$ (Kullback et al., 1962).

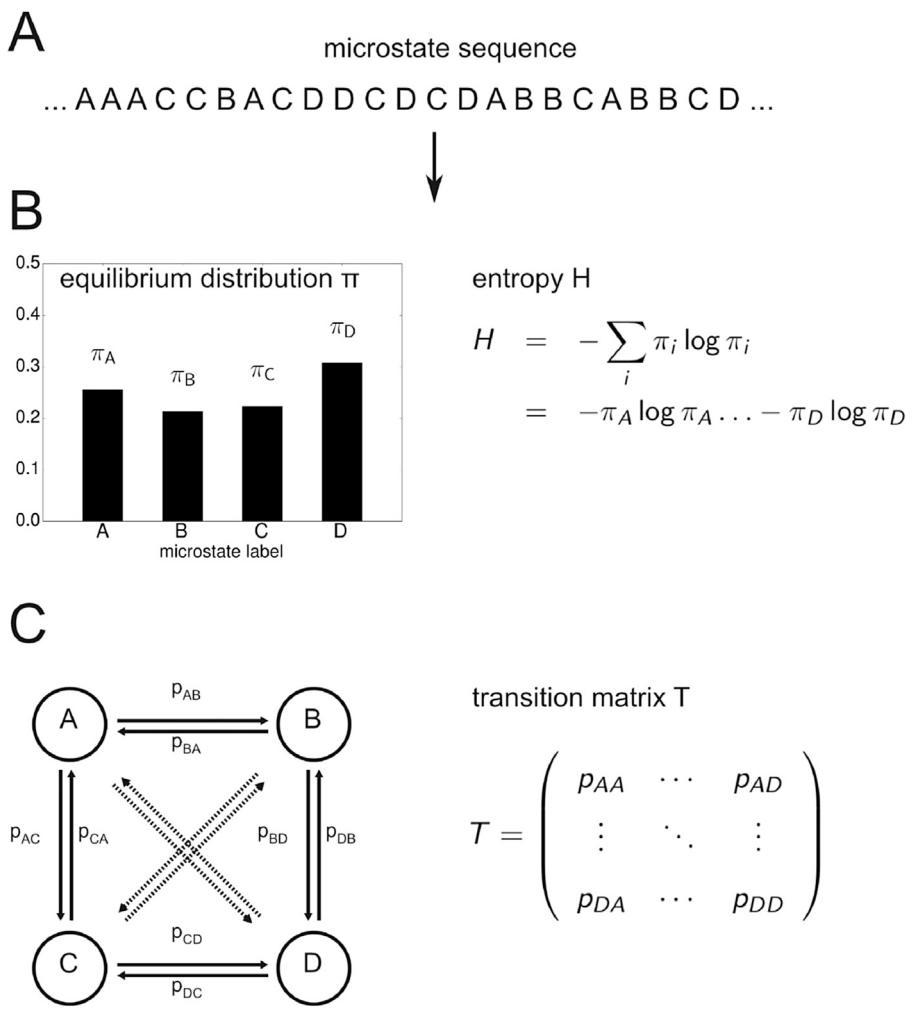

Fig. 2. A: From a given sequence of microstate labels, the equilibrium distribution π is calculated. B: The information content of the distribution π is quantified by the Shannon entropy H. C: For EEG microstates, all transitions between pairs of states including loops (e.g. $A \rightarrow A$) are possible and their transition probability is denoted by p_{ij}. For better visibility, not all p_{ij} terms are shown. First-order transitions $t \rightarrow t+1$ between microstates are summarized by the transition matrix T, higher-order transitions $t \rightarrow t+k$ by the matrix potency T^k.

2.6. First-order Markov property

The null hypothesis is that the transition probability $P(X_{t+1}|X_t)$ does not change with information about one state further in the past of the sequence, i.e. $P(X_{t+1}|X_t, X_{t-1}) = P(X_{t+1}|X_t)$. Thus, we aim to measure the divergence between the distributions $P(X_{t+1} = S_i|X_t = S_j, X_{t-1} = S_k)$ and $P(X_{t+1} = S_i|X_t = S_j)$. The G-test statistic

$$G_{1a} = 2\sum_{ijk} f_{ijk} \log \frac{f_{ijk} f_j}{f_{ij} f_{jk}} \tag{2}$$

is asymptotically distributed as $G_{1a} \sim \chi^2_{n_s(n_s-1)(n_s-1)}$ (Kullback et al., 1962).

Alternatively, first-order Markovianity can also be assessed based on the equivalence of the first-order Markov property (the memoryless property) with a geometric distribution of state durations (Feller, 1971). For each symbol, its lifetime distribution contains the lengths of contiguous segments containing the given symbol. Under the first-order Markovian null hypothesis, the expected frequency of a segment of length k containing symbol i follows the geometric distribution

$$q_i(k) = (1 - T_{ii}) \times T_{ii}^{k-1}$$

where T_{ii}^{k-1} is the $(k-1)$-th potency of the i-th diagonal element of the transition matrix T. The G-test statistic for each symbol is given by

$$G_{1b} = 2\sum_{i=1}^{m} p_i \log \frac{p_i}{q_i} \tag{3}$$

where m is the maximum lifetime and p_i is the empirical lifetime distribution. The test statistic is asymptotically distributed as $G_{1b} \sim \chi^2_{m-1}$. This test was applied to short (16 s) microstate sequences by (Wackermann et al., 1993).

2.7. Second-order Markov property

The null hypothesis is that the transition probability $P(X_{t+1}|X_t, X_{t-1})$ does not change with more information from the past of the sequence, i.e. $P(X_{t+1}|X_t, X_{t-1}, X_{t-2}) = P(X_{t+1}|X_t, X_{t-1})$. The G-test statistic

$$G_2 = 2\sum_{ijkl} f_{ijkl} \log \frac{f_{ijkl} f_{jk}}{f_{ijk} f_{jkl}} \tag{4}$$

is asymptotically distributed as $G_2 \sim \chi^2_{n_s n_s (n_s-1)(n_s-1)}$ (Kullback et al., 1962).

2.8. Stationarity of the transition matrix

We test the null hypothesis that the transition matrix is stationary over time. To this end, the microstate sequence is partitioned into r non-

overlapping blocks of length L each. Then, the transition matrix for each data block $k = 0, \ldots, r-1$ is calculated. If the transition matrix is stationary, the observed frequency f_{ijk} of the transition $X_t = S_i \rightarrow X_{t+1} = S_j$ in the data block k is independent of the block index k, i.e. not dependent on time. Thus, the null hypothesis is $P(X_{t+1}^{(k)} \big| X_t^{(k)}) = P(X_{t+1}|X_t)$, where $X_t^{(k)}$ refers to the k-th data block and X_t to the non-blocked data. The G-test statistic is calculated as:

$$G_3 = 2\sum_{ijk} f_{ijk} \log \frac{f_{ijk} f_j}{f_{ij} f_{jk}} \tag{5}$$

and is asymptotically distributed as $G_3 \sim \chi^2_{(r-1)(n_s-1)n_s}$ (Kullback et al., 1962).

2.9. Symmetry

If T is a symmetric transition matrix, the null hypothesis can be formulated as $P(X_{t+1} = S_i | X_t = S_j) = P(X_{t+1} = S_j | X_t = S_i)$. The G-test statistic

$$G_4 = 2\sum_{i \neq j} f_{ij} \log \frac{2f_{ij}}{f_{ij} + f_{ji}} \tag{6}$$

is asymptotically distributed as $G_4 \sim \chi^2_{n_s(n_s-1)/2}$ (Kullback et al., 1962).

2.10. The autoinformation function - time-lagged mutual information

For symbolic (non-metric) sequences, e.g. microstate sequences, the autocorrelation function cannot be calculated because notions such as sums, means or variances cannot be applied to non-metric symbols. However, it is possible to compute an analogous quantity that measures the dependence between different time points with time lag τ. To this end, we compare the symbol distributions at time points t and $t+\tau$ by means of the Kullback-Leibler divergence.

We define the autoinformation function (AIF) for time lag τ as:

$$I(\tau) = H(X_{t+\tau}) - H(X_{t+\tau}|X_t) \tag{7}$$

In words, the AIF is the mutual information between the random variables $X_{t+\tau}$ and X_t, and can be calculated as the difference between the entropies of the distributions $P(X_{t+\tau})$ and $P(X_{t+\tau}|X_t)$. Equivalently, $I(\tau)$ measures the amount of information about $X_{t+\tau}$ contained in X_t. For the null hypothesis of a stationary Markov chain, we get $P(X_{t+\tau} = i) = P(X_t = i) = \pi_i$ and $P(X_{t+\tau} = j | X_t = i) = T_{ij}$. Thus, the entropy terms read as $H(X_{t+\tau}) = -\sum_i \pi_i \log \pi_i$ and $H(X_{t+\tau}|X_t) = -\sum_i \pi_i \sum_j T_{ij}^\tau \log T_{ij}^\tau$. Here, T_{ij}^τ denotes the element (i, j) of the τ-th potency of the transition matrix T. Diagonalization of the transition matrix T_{ij} as

$$T^n = M \Lambda^n M^{-1} \tag{8}$$

defines a coordinate transform M, M^{-1} for which the transition matrix T is given by the diagonal matrix Λ containing the eigenvalues $\lambda_{1..n}$ of T in descending order. The autoinformation function (AIF) for the first-order Markov process with equilibrium distribution π and transition matrix T is then given by

$$I(\tau) = -\sum_i \pi_i \log \pi_i + \sum_i \pi_i \sum_j T_{ij}^\tau \log T_{ij}^\tau \tag{9}$$

$$= -\sum_i \pi_i \log \pi_i + \sum_i \pi_i \sum_j \left[\sum_k \lambda_k^\tau M_{ik} M_{kj}^{-1} \log \sum_k \lambda_k^\tau M_{ik} M_{kj}^{-1} \right] \tag{10}$$

using the definition of the matrix potency T^τ. This is the analytical form of the AIF in terms of the constant matrices M, M^{-1} and the eigenvalues $\lambda_{0..3}$.

It is observed that the expression for $I(\tau)$ consists of a sum of terms corresponding to each microstate S_i. The individual terms correspond to the autoinformation function of each microstate S_i. Using $P(X_t = S_i) = \pi_i$ and $P(X_{t+\tau} = S_j | X_t = S_i) = T_{ij}^\tau$, we get:

$$I_i(\tau) = -\pi_i \log \pi_i + \pi_i \sum_j T_{ij}^\tau \log_{ij}^\tau \tag{11}$$

The expression $I_i(\tau)$ measures the contribution of the microstate with index i to the global AIF. It is calculated for each microstate map of each subject.

3. Results

3.1. Basic microstate properties

First, we will review some basic microstate properties, as shown in Table 1. The first two columns contain the subject index and the length of the recording, respectively. The third column contains the number of GFP peaks per second (PPS), across the data sample. The next column contains four percentage values, normalized to one, representing the ratio of time covered by each of the microstates A-D, divided by the total duration of the sample (RTT). Next, column 5 gives the mean duration of each microstate (MMD) in milliseconds. Durations refer to the lengths of contiguous blocks of a specific symbol within the microstate sequence. The last two columns quantify the global explained variance (GEV), column 6 the contribution of each microstate, column 7 the total explained variance. All values lie well within the range of previously published microstate studies (Brodbeck et al., 2012). The ordering of microstates follows the geometries shown in the supplemental Figs. S2–S5.

To test the quality of Markov surrogates, we calculated a distance metric between the empirical transition matrix T of the microstate sequence and the transition matrix $\tilde{T}$ of the Markov surrogates. We used the maximum norm $d = \max_{ij} |T_{ij} - \tilde{T}_{ij}|$. Averaged across all subjects and surrogates, the mean transition probability error was 0.007, the maximum error 0.01.

In the following, the statistical tests defined above are applied to the 20 experimental data sets and to $n_{surr} = 1000$ first-order Markov surrogates matched with each subject. All test results are corrected for multiple comparisons using Bonferroni correction. Statistical significance is assessed at the $\alpha = 0.01$ level.

Table 1
Basic microstate properties: sample length T, GFP peaks per second (PPS), ratio of total time covered (RTT), mean map duration (MMD), global explained variance per map (GEV_{A-D}) and total global explained variance (GEV_T).

Subject	T [s]	PPS	RTT	MMD [ms]	$GEV_{A,B,C,D}$	GEV_T
1	266	28.95	.24, .25, .29, .22	47.5, 49.8, 56.4, 43.0	.15, .16, .27, .09	.67
2	270	27.74	.21, .29, .26, .24	45.5, 63.4, 52.0, 48.7	.10, .31, .17, .12	.70
3	120	31.21	.25, .24, .25, .26	44.1, 39.6, 45.1, 47.9	.17, .12, .18, .19	.66
4	100	31.01	.24, .26, .29, .21	45.4, 47.0, 50.9, 40.0	.15, .19, .25, .08	.67
5	300	27.12	.26, .29, .23, .22	59.9, 64.1, 52.7, 47.1	.20, .25, .12, .11	.68
6	192	24.67	.25, .23, .30, .22	61.3, 57.2, 78.3, 58.9	.18, .11, .29, .12	.70
7	303	29.05	.23, .26, .27, .24	45.4, 51.8, 59.5, 47.0	.10, .17, .29, .11	.67
8	180	31.99	.27, .18, .28, .27	45.8, 37.9, 48.3, 46.8	.17, .08, .20, .18	.63
9	180	30.69	.26, .23, .27, .24	44.3, 44.0, 52.9, 46.2	.13, .13, .26, .14	.66
10	180	34.67	.27, .22, .27, .24	41.8, 37.6, 44.7, 38.7	.18, .09, .22, .12	.61
11	312	28.64	.32, .27, .20, .21	63.5, 56.7, 43.8, 45.2	.30, .25, .09, .11	.75
12	289	29.27	.27, .21, .25, .27	54.1, 42.4, 45.8, 49.2	.26, .10, .12, .17	.65
13	200	39.09	.17, .25, .30, .28	28.2, 36.1, 41.7, 37.9	.08, .12, .25, .14	.59
14	298	30.17	.24, .25, .27, .24	45.8, 45.5, 55.2, 43.7	.12, .14, .27, .12	.65
15	150	26.33	.22, .26, .27, .25	53.9, 59.3, 62.5, 54.0	.13, .21, .24, .14	.72
16	165	29.02	.24, .23, .27, .26	46.6, 45.7, 56.5, 49.4	.13, .13, .25, .14	.65
17	103	24.46	.26, .24, .28, .22	57.4, 57.7, 70.1, 54.6	.18, .13, .27, .12	.70
18	120	28.56	.27, .22, .27, .24	54.6, 44.7, 52.9, 46.0	.19, .12, .22, .13	.66
19	121	23.97	.31, .24, .26, .19	78.6, 54.8, 62.3, 46.5	.33, .16, .22, .07	.78
20	300	27.85	.24, .25, .27, .24	49.7, 50.0, 59.0, 47.1	.14, .13, .28, .12	.67

3.2. Low-order Markov properties

When testing the zero-order Markov property (Eq. (1)), the null hypothesis is rejected for all EEG microstate sequences. In other words, subsequent elements of the sequence are not statistically independent and there is at least some temporal dependency in the microstate sequence. The first-order Markov surrogates, as expected, show the same behaviour.

Using the G-test for first-order Markovianity, Eq. (2), we find that none of the tested EEG microstate sequences fulfils the first-order Markov property. Equivalently, we can state that the current state of the sequence does not contain the full information about the next future state, as there is additional information stored further in the past of the sequence. The first-order Markov surrogates fulfil the tested property, as expected by construction. Using the alternative test, Eq. (3), we observe that among 80 tests (4 microstate distributions × 20 subjects), the null hypothesis of a geometric lifetime distribution is rejected in all but three cases.

Similarly, none of the EEG microstate sequences fulfils the second-order Markov property (Eq. (4)). This result states that the memory effect in the sequences extends at least 2 time steps into the past. As derived from the construction of the first-order Markov surrogates, these also fulfil the second-order Markov property. The results are summarized in Table 2 where the first three columns contain the results for the Markovianity tests of order 0, 1 and 2 and boldface values indicate statistical significance ($p < 0.01$). The zero-order test produced p-values so small that a value of 0e+ 0 is given. Column 4 contains the results of the alternative test for first-order Markovianity testing the geometric distribution of microstate lifetimes for each microstate label A-D. Overall, low-order Markovianity is rejected for our data sets.

3.3. Stationarity of the transition matrix

To assess time-stationarity of the transition matrix T, the test statistic G_3 (5) is computed using non-overlapping blocks of data. The first step is to determine the optimum block size L. We tested block sizes of $L =$ 250, 500, 1000, 2500, 5000 and 10000, corresponding to durations of 1, 2, 4, 10, 20 and 40 s, respectively. For each subject, we applied the stationarity test to $n_{\mathrm{surr}} = 1000$ first-order Markov surrogates with equilibrium distribution π and transition matrix T identical to the given experimental data set. We find that stationarity is correctly identified in the Markov data set for block sizes $L \geq 1000$. However, the block size L must not be too large compared to the sample size. For our Markovian data sets, we found that a minimum of 3 blocks is necessary to correctly identify stationarity. In our experimental data set, there are 4 subjects with a length of $n = 30000$ samples or shorter. Their Markov surrogates are correctly classified as time stationary for $L = 2500$ and $L = 5000$. However, for $L =$ 10000, the test statistic fails to identify stationarity. Therefore, we excluded the four shorter sequences from the $L = 10000$ analyses. Stationarity tests with a block size of $L = 2500$ and 5000 were applied to all data sets.

As shown in Table 3, for experimental data and a block size of $L =$ 2500 (10 s), the null hypothesis of stationarity is rejected for 16/20 (80%) subjects. For a block size of $L = 5000$ (20 s), 12/20 (60%) subjects show a non-stationary transition matrix. Finally, for $L = 10000$, we found 7/16 (43,8%) non-stationary microstate sequences. These results show that non-stationarity is a frequently encountered property of empirical microstate transition matrices. The results are summarized in columns 2–4 (block sizes $L =$ 2500, 5000, 10000) of Table 3. Each column contains the p-values of the corresponding test and null hypotheses rejected at the $\alpha = 0.01$ significance level are printed in boldface.

3.4. Symmetry

The column 'symmetry' in Table 3 shows the results of the symmetry tests, asking if the off-diagonal elements of the empirical transition matrix are symmetric, i.e. if $T_{ij} \sim T_{ji}$ for $i \neq j$. Only 6 out of 20 subjects display symmetric transition matrices whereas the majority of the matrices are asymmetric. This shows that in 14/20 cases, and for at least one pair of non-identical states (S_i, S_j), the transition $S_i \to S_j$ occurs with a different rate than the transition $S_j \to S_i$. An example of a symmetric transition matrix is given by subject S4, for which the null hypothesis of symmetry is not rejected ($p = 0.563$)

$$T_{\mathrm{sym}} = \begin{pmatrix} 0.760 & 0.090 & 0.083 & 0.067 \\ 0.095 & 0.728 & 0.102 & 0.075 \\ 0.109 & 0.085 & 0.737 & 0.069 \\ 0.092 & 0.088 & 0.081 & 0.739 \end{pmatrix}$$

and the maximum absolute difference in transition probabilities is $|p_{ij} - p_{ji}|_{\max} = 0.026$. Subject S1, in contrast, gives an asymmetric transition matrix ($p < 0.01$) with entries

Table 2
Markovianity tests for resting state EEG microstate sequences. For each subject, the test result for Markovianity of order 0, 1 and 2 is indicated. For significant results ($p < 0.01$), the p-value is given in boldface, non-significant results are marked by n.s. All results are Bonferroni corrected.

Subject	Markov-0	Markov-1	Markov-2	geometric (A, B, C, D)
1	**0e+0**	**1e-189**	**1e-130**	(**1e-38**, **3e-23**, **9e-21**, **3e-35**)
2	**0e+0**	**5e-208**	**3e-137**	(**3e-43**, **2e-70**, **1e-21**, **1e-20**)
3	**0e+0**	**1e-72**	**7e-23**	(1e-02, **2e-06**, **8e-06**, **7e-05**)
4	**0e+0**	**2e-61**	**6e-28**	(**6e-13**, **6e-06**, 1.2e-01, **4e-06**)
5	**0e+0**	**3e-225**	**3e-151**	(**3e-36**, **6e-84**, **1e-44**, **2e-59**)
6	**0e+0**	**2e-106**	**7e-53**	(**2e-37**, **5e-15**, **3e-09**, **6e-15**)
7	**0e+0**	**7e-128**	**1e-46**	(**2e-16**, **2e-42**, **5e-14**, **7e-06**)
8	**0e+0**	**2e-87**	**4e-25**	(**3e-05**, **1e-09**, **4e-28**, **2e-10**)
9	**0e+0**	**2e-116**	**3e-46**	(**7e-08**, **2e-05**, **2e-31**, **7e-20**)
10	**0e+0**	**8e-116**	**1e-42**	(**9e-14**, **7e-04**, **4e-26**, **1e-12**)
11	**0e+0**	**2e-258**	**2e-112**	(**3e-49**, **2e-51**, **1e-04**, **2e-60**)
12	**0e+0**	**5e-156**	**5e-76**	(**6e-62**, **2e-06**, **6e-19**, **5e-08**)
13	**0e+0**	**9e-221**	**6e-48**	(**1e-33**, **1e-84**, **4e-52**, **6e-73**)
14	**0e+0**	**3e-182**	**2e-47**	(**8e-22**, **8e-03**, **1e-36**, **7e-35**)
15	**0e+0**	**1e-105**	**2e-64**	(**2e-05**, **1e-31**, **6e-13**, **1e-23**)
16	**0e+0**	**2e-70**	**9e-27**	(**1e-07**, **3e-14**, **1e-12**, **1e-10**)
17	**0e+0**	**2e-67**	**6e-28**	(**3e-32**, **2e-07**, **2e-13**, **3e-03**)
18	**0e+0**	**3e-48**	**3e-17**	(**2e-13**, **6e-06**, **2e-07**, **3e-03**)
19	**0e+0**	**2e-100**	**7e-54**	(**4e-19**, 5e-01, **2e-40**, **5e-26**)
20	**0e+0**	**7e-284**	**6e-183**	(**5e-63**, **3e-30**, **1e-50**, **2e-59**)

Table 3
Stationarity and symmetry tests for resting state EEG microstate sequences, block lengths $L = 2500$, 5000, 10000. Four recordings were too short to apply $L = 10000$, indicated by n.a. All results are Bonferroni corrected.

Subject	stationarity (L = 2500)	stationarity (L = 5000)	stationarity (L = 10000)	symmetry
1	**2.2e-9**	**9.4e-8**	1.000	**3.3e-4**
2	**1.9e-22**	**1.7e-21**	**5.4e-7**	**2.1e-140**
3	**4.2e-6**	**2.3e-4**	n.a.	0.198
4	0.051	1.000	n.a.	0.563
5	**2.4e-36**	**4.2e-20**	**2.3e-11**	**3.5e-163**
6	**2.9e-6**	0.021	0.274	**7.8e-44**
7	**2.7e-12**	**4.6e-6**	0.106	**3.7e-29**
8	**6.2e-16**	**6.2e-11**	**1.5e-10**	**1.8e-6**
9	**1.5e-8**	**7.4e-8**	**3.6e-5**	0.510
10	**1.1e-3**	0.302	0.0751	1.0
11	**1.0e-47**	**4.3e-25**	**9.9e-15**	**3.8e-305**
12	**2.6e-5**	0.035	0.122	**1.7e-32**
13	0.253	0.227	0.7000	0.278
14	0.115	0.140	0.097	**4.2e-16**
15	**4.4e-3**	0.340	0.0259	**5.8e-12**
16	**7.2e-13**	**2.1e-4**	**3.6e-4**	**1.3e-4**
17	**6.3e-43**	**9.1e-30**	n.a.	**3.6e-151**
18	**1.8e-3**	**1.2e-3**	n.a.	**5.6e-14**
19	**1.6e-3**	**3.5e-4**	**3.7e-7**	**2.7e-80**
20	0.028	0.667	0.340	0.023

Statistical significance was tested for alpha = 0.01.

$$T_{\text{asym}} = \begin{pmatrix} 0.759 & 0.118 & 0.060 & 0.063 \\ 0.076 & 0.768 & 0.096 & 0.060 \\ 0.118 & 0.064 & 0.733 & 0.085 \\ 0.105 & 0.112 & 0.078 & 0.705 \end{pmatrix}$$

and a maximum difference in transition probabilities of $|p_{ij} - p_{ji}|_{\max} = 0.058$.

This property is important in the light of cycling behaviour, i.e. periodic occurrences of symbols. The shown asymmetric transition matrix T_{asym}, for instance, suggests a cycle of $A \to B \to C \to A$. Starting with microstate A and ignoring self-loops (e.g. $A \to A$), the largest transition rate is to microstate B ($p_{ij} = 0.118$), as read from the first row of T_{asym}. From microstate B, the largest transition rate is to microstate C ($p_{ij} = 0.096$), as read from the second row of T_{asym}, and finally, from microstate C, the most probable transition is back to state A ($p_{ij} = 0.118$). The corresponding transition rates in the reverse direction are all smaller than the listed forward rates and therefore, cycling behaviour is anticipated. As we deal with a stochastic process, all other transitions are also possible, but the probability of the indicated cycle is slightly elevated.

3.5. Periodicities in the microstate transform

Before we analyze periodic information content in microstate sequences, we analyze some basic periodic properties of resting state EEG in the frequency and time domains, respectively (Fig. 3). For each subject, the 30 channel EEG is compressed into a single time course using the first principal component (PCA-1) across all EEG channels. Fig. 3A shows the power spectral density of all subjects (thin grey lines) as well as the average spectrum across the subjects (thick black line). As expected during wakeful rest, there is a prominent peak in the alpha frequency band, around 10 Hz. As we will subsequently analyze microstate sequences in the time domain, we also show the EEG time autocorrelation function (ACF) in Fig. 3B. In the time domain, the dominant 10 Hz frequency is represented by the first peak of the ACF around a time lag of 100 ms.

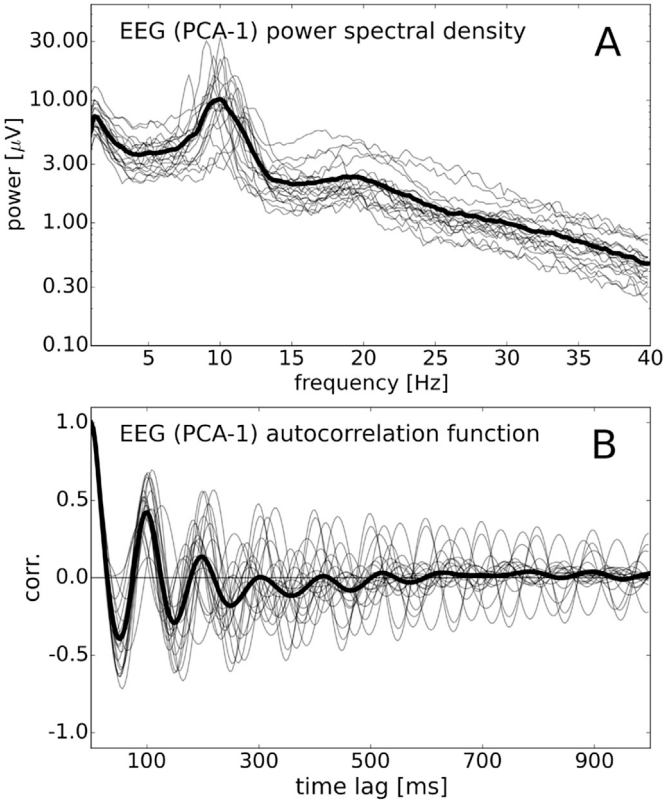

Fig. 3. Oscillatory properties of resting state EEG. For each subject, the 30 channel EEG is represented by its first principal component (PCA-1). A: The power spectral densities of all $n = 20$ subjects (thin grey lines) and the average spectrum across all subjects (thick black line) are shown in log-log coordinates. All spectra show a distinct peak in the EEG alpha frequency band around 10 Hz. B: In the time domain, oscillatory properties are found in the time autocorrelation function (ACF) of each subject (thin grey lines) and in the average ACF across subjects (thick black line). The alpha frequency peak is represented by the first local maximum found at a time lag of approx. 100 ms, or 10 Hz.

Next, we aim to track these oscillatory EEG features as the data are passed through the microstate segmentation algorithm. The results for subject S2 are exemplified in Fig. 4. Fig. 4A shows the autocorrelation function of the EEG global field power over time. We observe that the first peak of the ACF is located near a time lag of 50 ms, or 20 Hz, i.e. at twice the value of the EEG's principal alpha frequency (approx. 10 Hz). In Fig. 4B, we show how the correlations of a given microstate map with the actual EEG topography vary over time. For instance, let C_A represent the correlation of microstate map A with the actual EEG topography over time, and let C_B be the corresponding quantity for microstate B. Fig. 4B shows the time-lagged cross-correlation between C_A and C_B. It is observed that the time series show a preferred phase lag of 50 ms. In other words, the similarity of the EEG topography with each map varies periodically over time and at a frequency around 20 Hz. Fig. 4C shows the plot of C_A vs. C_B. The oscillatory behaviour of EEG-microstate similarity over time is also recognized in the quasi-periodic trajectory described by pairs of (C_A, C_B)-values. In summary, these analyses provide evidence that EEG periodicities might be partially preserved during microstate segmentation and may still be present in the symbolic microstate sequence. This hypothesis is tested in the following paragraph.

3.6. Periodicities of the autoinformation function

Temporal dependencies within microstate sequences for arbitrary time lags are quantified by the autoinformation function. The process is illustrated in Fig. 5. The amount of information (surprise) contained in the microstate sequence at times t and $t+\tau$ is quantified by the entropies $H(X_t)$ and $H(X_{t+\tau})$, respectively (Fig. 5A). The amount of information shared between time points t and $t+\tau$ is given by the time-lagged mutual information $I(\tau) = H(X_{t+\tau}) - H(X_{t+\tau}|X_t)$ and measures the amount of surprise about the microstate map at $t+\tau$, given the microstate label that occurred at time t. In Fig. 5B, these quantities are illustrated using a Venn diagram analogy. The two circles represent the information content at time t and $t+\tau$. Their intersection represents their shared, or mutual information. We observe that the resulting autoinformation function shown in Fig. 5C (subject S5) shows periodic peaks at time lags that are multiples of 50 ms, similar to the autocorrelation function of the GFP time course shown in Fig. 4A. The value at time-lag zero represents the entropy of the whole sequence. As the zero term is significantly larger than the subsequent values, we chose a semi-logarithmic display in order to show the complete function.

Next, we want to test if the periodicities of the AIF are random fluctuations and how they relate to the Markovian picture. Fig. 6 shows the empirical autoinformation function of microstate sequences (black line and circles) as well as the AIF under the first-order Markov null hypothesis (blue line) for two representative resting state recordings. Fig. 6A shows the results for subject S5 and represents the class of subjects for which the AIF has sustained oscillatory peaks up to time lags of ≈500 ms. It is observed that the analytical AIF for the equivalent Markov process (blue line) decays monotonically without any local maxima that would indicate periodic memory content. The Markovian AIF shows the memory content gathered by considering first-order transitions only, i.e. the information captured by the classical transition matrix approach. As the Markovian AIF decays very quickly, the insets show the same curves for small time lags in more detail. In addition to the analytical AIF, the inset also shows the 95% confidence interval for the Markovian AIF (grey shaded area) as obtained from Markov surrogate data. We see that the

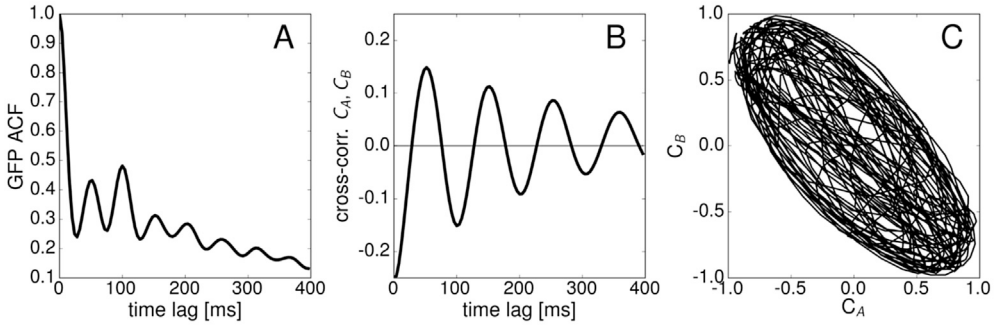

Fig. 4. Periodicities in the microstate transform (subject S2). A: the autocorrelation function (ACF) of the global field power (GFP) time course shows clear oscillations with a first peak at approx. 50 ms, i.e. 20 Hz. B: Time-lagged cross-correlation between C_A and C_B, where C_A represents the spatial correlation of microstate A with the actual EEG topography over time and C_B the analogous quantity for microstate B. The instantaneous correlation between EEG and microstates A resp. B over time has a preferred time lag of approx. 50 ms (20 Hz). C: When C_A and C_B are plotted against each other, we observe a quasi-periodic trajectory. These observations lead to the hypothesis that EEG periodicities may be inherited by the microstate sequence.

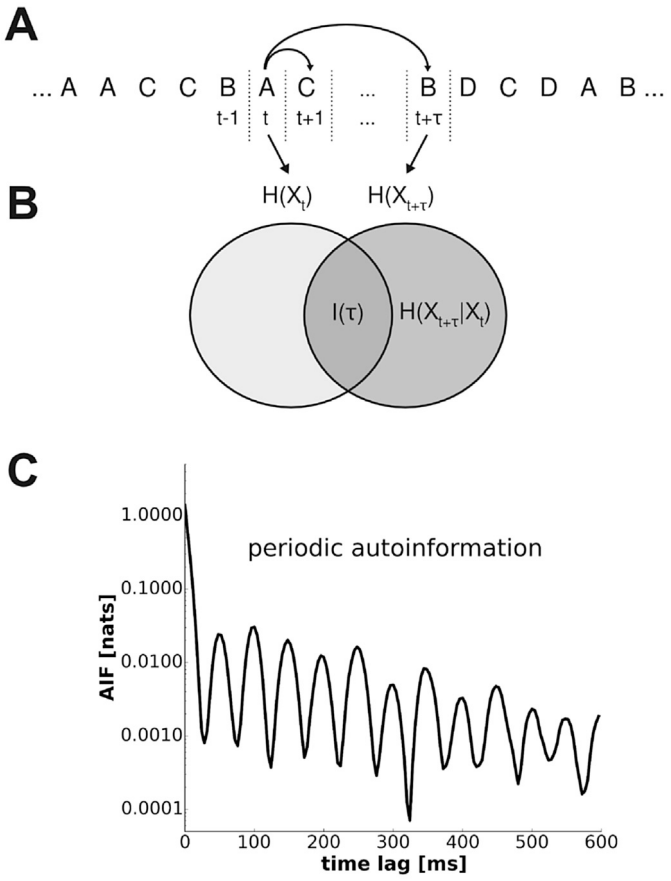

Fig. 5. The amount of information preserved by the microstate sequence while evolving from time t to $t+\tau$ is measured by the autoinformation function (AIF). A: For each time lag τ, the AIF contains the time-lagged mutual information between the random variables X_t and $X_{t+\tau}$. B: The shared information is calculated from the entropies $H(X_t)$, $H(X_{t+\tau})$, and the conditional entropy $H(X_{t+\tau}|X_t)$. It quantifies the amount of surprise about the microstate label at time $t+\tau$, knowing the label at time t. C: The autoinformation function for time lags up to 600 ms. As the numerical values of the function decay quickly, the AIF is plotted in semilog-y coordinates. Oscillatory peaks are observed around the same time lags as in Fig. 4, the first peak representing the 20 Hz range.

analytical AIF and the surrogate data confidence interval coincide very precisely.

Fig. 6B shows the results for subject S2, which represents the group of subjects with a rapidly decaying AIF. In these subjects, the most prominent peaks of the AIF are found at time lags 50 ms and 100 ms, followed by a rapid decay. The deviation from the Markovian null hypothesis however is equally clear as in Fig. 6A, as shown in the inset. Please note that for visualization purposes, the axes in Fig. 6 are clipped.

Frequency doubling is summarized in Table 4. For each subject indexed in column 1, the first peak of the PCA-1 autocorrelation function, corresponding to the subject's alpha frequency in milliseconds, is given in the second column. The corresponding first peak in the AIF is given in column 3. Column 4 shows the ratio of columns 2 and 3 and it is observed that the ratio is centered close to the expected value of 2. Column 5 shows the average microstate duration, averaged across all 4 microstates. While average microstate durations lie in the same range as the first AIF peaks, durations do not correlate linearly with either the location of the first ACF peak ($p = 0.13$) or the first AIF peak ($p = 0.49$).

3.7. Periodicity of individual microstates

Analyzing the contribution of individual microstates to the global AIF using Eq. (11), we find two types of behaviour. For one subset of subjects, we find that all microstates show periodic information content. This is represented by subject S5 in the left column of Fig. 7, where for each of the 4 microstate maps (S5 A-D), the empirical AIF (thin black curves, Eq. (11)) and an overlay of the empirical global AIF (light grey curve) are shown. All microstates contribute similar peaks and are qualitatively similar to the global AIF. For another class of subjects, not all microstates have AIFs similar to the global AIF. This case is illustrated in the right column of Fig. 7 for subject S20. We observe that only the third map (S20-C), displaying a horizontal pattern, has a periodic memory trace similar to the global AIF. The other maps show less clear patterns although individual spectral peaks are observed. Interestingly, for the second class of subjects, it was always the horizontal map that followed the global AIF most closely. To focus on the oscillatory part of the AIF, x- and y-axes are clipped. The corresponding microstate maps are shown as pseudo-colored insets.

3.8. Long-range behaviour

In Fig. 8 we look at the long-range behaviour of the AIF up to time lags of 2000 ms. The subject indices are shown next to the label and two subjects are identical to the ones shown in Fig. 6. To the left, the tail of the oscillatory regime is shown, with sustained periodicities in the upper two traces, and a faster decay in the lower two traces. Beyond time lags of ≈1000 ms, the AIF reaches and stays within the first-order Markovian confidence interval (grey shaded area). This short-range memory phenomenon was also observed in the remaining data sets, not shown here. This shows that for large time lags, beyond the oscillatory region, microstate sequences show short-range memory, statistically not different from an equivalent first-order Markov process.

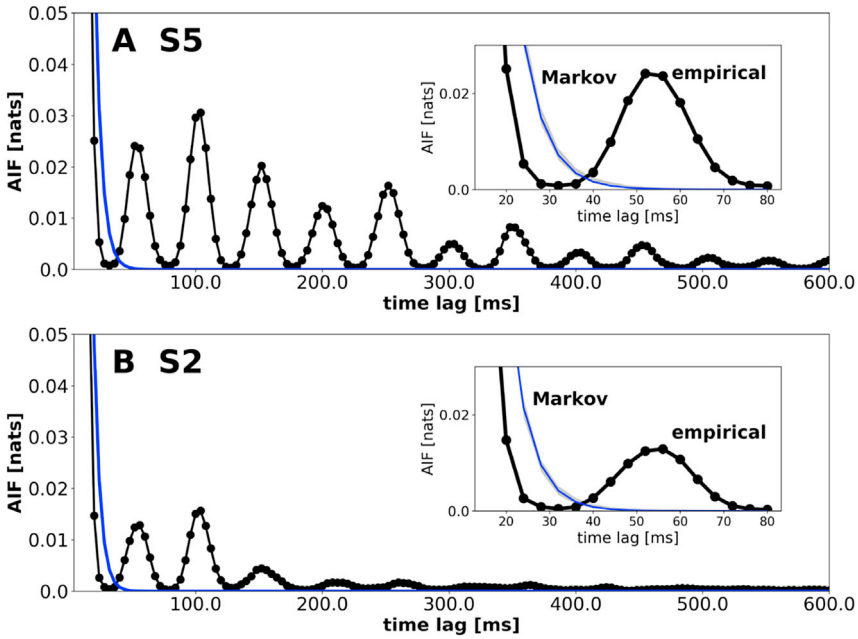

Fig. 6. The autoinformation function for two exemplary subjects. A: Subject S5 represents the group of slowly decaying autoinformation functions (black line and circles) that show regular harmonics of the main 50 ms (20 Hz) peak up to the 500 ms range. B: Subject S2 represents the group of AIFs with clear 50 ms and 100 ms peaks followed by a fast decay. In both cases, the oscillatory behaviour significantly deviates from the first-order Markovian memory decay represented by the theoretical AIF (Eq. (9), blue line) and by the confidence interval constructed from Markov surrogate data (light grey shaded area in the insets). The insets in A and B focus on the time lags around the first AIF peak that show the deviation from Markovianity most clearly. To focus on oscillatory properties, x- and y-axes are clipped.

Table 4
Frequency doubling. For each subject indexed in column 1, temporal dependencies are given as first-peak latencies in milliseconds. Column 2 refers to the autocorrelation function of the EEG's first principal component (ACF PCA-1), and Column 3 to the autoinformation function (AIF) of the microstate sequence. Column 4 shows the ratio of column 2 and column 3, with values close to 2, indicating frequency doubling. Column 5 shows the mean microstate duration in milliseconds, averaged across all microstates.

Subject	ACF PCA-1 [ms]	AIF [ms]	Ratio ACF-1/AIF-1	average ms lifetime [ms]
1	100	48	2.1	49
2	100	52	1.9	52
3	104	52	2.0	44
4	84	52	1.6	46
5	100	48	2.1	56
6	108	52	2.1	64
7	92	48	1.9	51
8	124	60	2.1	45
9	92	48	1.9	47
10	84	52	1.6	41
11	100	52	1.9	52
12	96	52	1.8	48
13	88	40	2.2	36
14	104	56	1.9	48
15	92	48	1.9	57
16	96	52	1.8	50
17	104	56	1.9	60
18	108	56	1.9	50
19	108	52	2.1	60
20	96	40	2.4	52

4. Discussion

Information-theoretical analysis of microstate sequences is presented as a flexible and theoretically well-grounded methodological alternative to other spatio-temporal EEG analyses. Our systematic analysis of temporal dependencies in microstate sequences reveals several previously unrecognized features of these signals. The results can be summarized as follows:

1. All sequences show highly significant non-Markovian behaviour. The Markov properties of low order (orders 0, 1 and 2), are rejected when tested directly and the empirical autoinformation function of microstate sequences is strikingly different from the first-order Markov version.
2. A large proportion of microstate sequences show a non-stationary and asymmetric transition matrix T, i.e. the transition probabilities between microstate maps can change over time and are often not time-reversible.
3. All sequences display periodic information content at frequencies related to the underlying EEG power spectral density. The microstate transform induces frequency-doubling with respect to the EEG spectrum. The speed of decay of these periodicities is analogous in the EEG and microstate sequence spectra.
4. Memory effects in all microstate sequences finally decay to the first-order Markov confidence interval for time lags larger than $\tau \approx 1000$ ms, i.e. microstate sequences display short-range memory.

4.1. Memory effects in microstate sequences

Our analysis sheds new light on the question of temporal dependencies in resting state microstate sequences. Given a sequence of discrete states, one of the first questions that arise is whether the sequence is random or if the states carry some information about the past of the sequence. For continuous data, this property is usually addressed by analyzing the (partial) autocorrelation function. For symbolic data, autocorrelations can only be calculated if the set of symbols is mapped to a metric space, whose elements allow the computation of sums and products. However, if the symbols represent EEG topographies, any such mapping seems arbitrary and difficult to justify biologically. Moreover, when mapping EEG topographies to numbers, a linear order of the elements is created. For EEG topographic clusters as those shown in Figs. 1 and 7, an order relation ($\geq$) is not justifiable. Therefore, we computed the amount of information shared between two time points, using an information-theoretical quantity that only requires the probability distributions of the microstate symbols at these times. Applying statistical

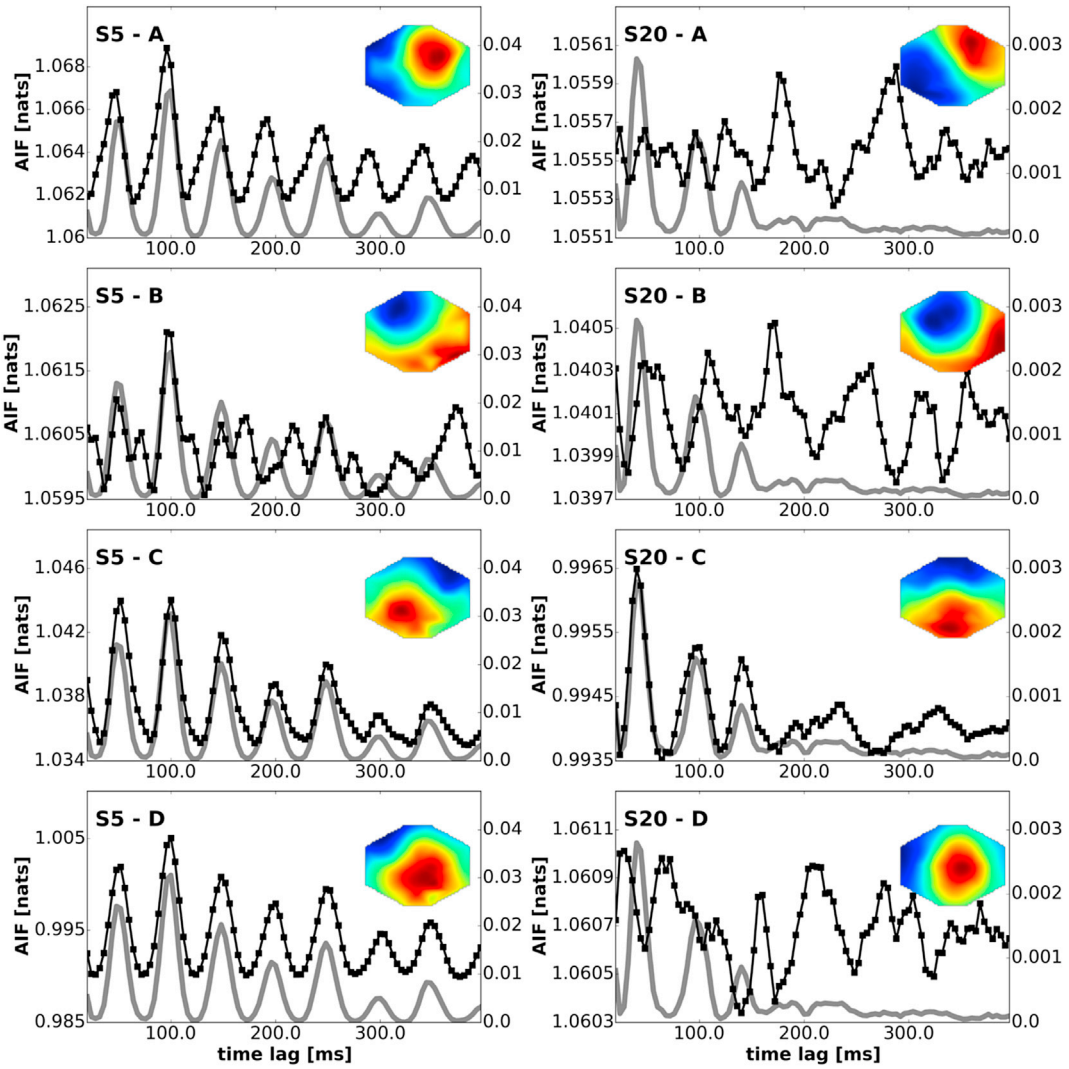

Fig. 7. Individual microstates contribute differently to global periodic information. The left column (S5 A-D) shows the individual AIFs for each microstate in subject S5, the right column (S20 A-D) for subject S20 (see Table 1). Microstate maps are shown as insets. In both columns, the global AIF of the subject is shown as a light grey line. In S5, the individual AIFs (black lines and circles) of all microstates contribute the same peaks to the global AIF, i.e. all microstates re-appear periodically. In S20, microstate C (S20-C) contributes an AIF that closely follows the shape of the global AIF. The remaining microstates (A, B, D) contribute AIFs without these clear oscillatory features. Whenever this pattern was observed, the third microstate type, the 'horizontal map', always resembled the global AIF most closely. For visualization purposes, x- and y-axes are clipped.

tests for the Markov properties of low order first, and looking at the mutual information within sequences subsequently, we find consistent non-Markovianity for time lags up to 1000 ms in the resting state condition. These results show that microstates are not emitted randomly by a memoryless mechanism, they rather show extended memory effects. The same result is obtained when the lifetime distributions of individual microstates are tested. To summarize, we found strong statistical evidence for the non-Markovian nature of the microstate process. We want to emphasize that the same methodology can be applied to any microstate transform (e.g. K-means, hierarchical clustering, independent component analysis) and to decompositions into any number of microstates. An alternative to simple clustering algorithms is the use of Hidden Markov Models (HMM) to construct a set of microstates (Gärtner et al., 2015; Gschwind et al., 2015). In HMMs, the Markov process of state transitions is hidden from observation, and the observable state emitted by the hidden states follows a certain probability distribution. We do not study these models in detail here. However, the analyses presented here can be applied to microstate sequences from HMMs without restrictions. Whether HMM-derived state sequences have a different information-theoretical structure than those discussed here remains a subject of further studies. Another point to mention is the sampling rate during data acquisition. When the sampling rate falls below the autocorrelation time of the measured process, the process may appear to be memoryless, or Markovian. As sampling rates above 100 Hz are common practice in EEG research, this issue should not arise. However, using the same methodological approach in fMRI, for example, may require additional efforts.

4.2. *The transition matrix approach*

The aforementioned methods and results extend the information provided by the classical transition matrix approach to microstate analysis. Looking beyond first-order transitions ($t \rightarrow t+1$) seems necessary as the implicit Markov model of the transition matrix approach is rejected when tested directly in empirical data. Although average microstate label distributions and first-order transition probabilities may convey interesting information about general properties of ongoing EEG patterns, we should be aware that the approach does not fit the structure of the data. Not only are microstate processes non-Markovian for short time lags, they also show very clear periodic features. These oscillatory features cannot be reproduced by a first-order Markov model, as shown by the monotonously decaying AIFs in Fig. 6. The case is similar as in autoregressive (AR) modeling of continuous EEG data. The transition matrix

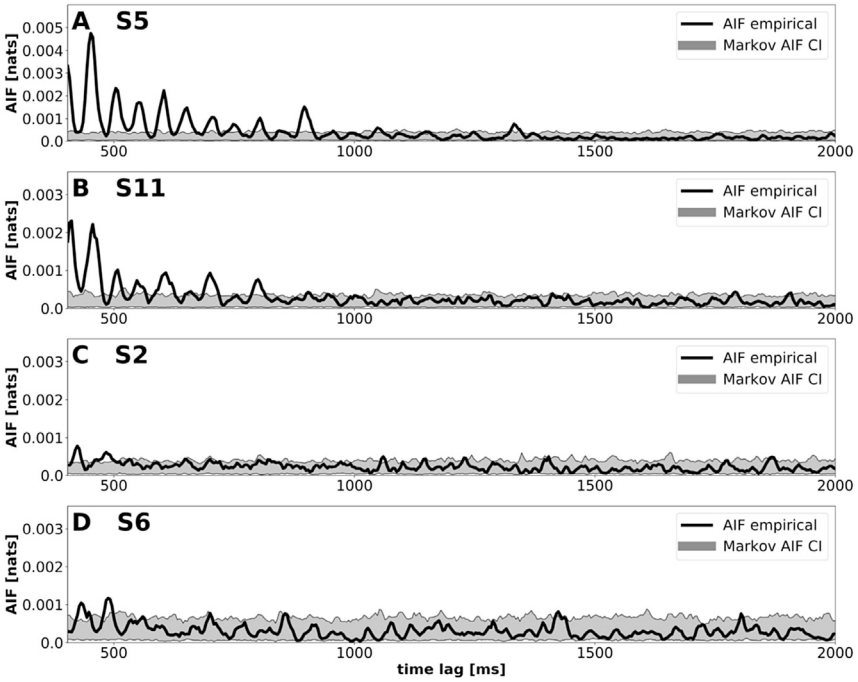

Fig. 8. Short-range memory in microstate sequences. To assess the asymptotic information content of microstate sequences, the plot shows the autoinformation function of four exemplary recordings for long time lags (500–2000 ms). While subjects S5 (A) and S11 (B) represent recordings with strong oscillatory AIFs up to >500 ms, subjects S2 (C) and S6 (D) belong to the group with rapidly decaying AIFs. In all presented cases, the AIF (black lines) decays to the first-order Markov 95% confidence interval (grey shaded area) for time lags >1000 ms, arguing against long-range memory. The same pattern is observed in the remaining subjects.

approach is equivalent to a first-order AR(1) model, whose analytical ACF is exponentially decaying. To model oscillations, at least an AR(2) model is needed (Box and Jenkins, 1976). However, our results show that microstate sequences show temporal dependencies beyond a 2nd order Markov model also (Table 2). This implies a more careful interpretation of transition matrix analyses. If transition matrices are different under variation of the experimental condition, the underlying microstate dynamics are probably different. If transition matrices are equal, however, higher order dynamics may still be different. We propose to use the autoinformation function to scan for altered microstate dynamics across different time scales.

4.3. Periodicities, frequency doubling and symmetry

Already in the early years of microstate research, periodicities of the global field power and its phase-shifted companion, the global dissimilarity measure were reported (Lehmann et al., 1987). Both signals together reflect the spatiotemporal properties of brain electrical activity. Global field power maxima occur at dissimilarity minima and vice versa. This reflects the fact that high spatial variance (global field power) is associated with temporally quasi-stable potential configurations. The stability of the spatial pattern over short time periods is associated with low dissimilarity measures. Interlaced with these stable configurations, unstable, low-variance transition periods are observed. These periods go along with low global field power and large dissimilarity. Furthermore, it has been noted that the potential configurations at subsequent GFP peaks are often inverted with respect to each other (Lehmann et al., 1987; Wackermann et al., 1993). It has been argued that maps with the same potential shape but inverse polarity should be assigned to the same microstate class (Wackermann et al., 1993), and therefore, polarity should be ignored during microstate segmentation (Wackermann et al., 1993; Murray et al., 2008). Finally, it is observed that the GFP signal displays twice the peak frequency of the underlying EEG and that the 50 ms phase lag is preserved during the microstate fitting procedure, as shown in Fig. 4. To summarize, there is an elevated probability to find identical microstate cluster assignments at subsequent GFP peaks, i.e. at half the EEG wavelength. As a consequence, microstates themselves tend to occur with a periodicity of twice the EEG main frequency. Moreover, we show that oscillatory features are invariant with respect to the cluster number as illustrated in Fig. S1. Our analysis also suggests that microstates are not spectrally independent as the sequences inherit their periodicity from the EEG. This means that the dominant EEG frequency band will determine the periodicity of the resulting microstate sequence, and in this sense, microstates do not provide a frequency-independent tool for EEG characterization.

Periodicity appears to be a basic feature of microstate sequences that has not attracted attention by the community so far, possibly due to a lack in methodology for symbolic sequences. We aimed to fill this gap with the methods detailed in this paper. Using a flexible information-theoretical approach, we have provided strong evidence that periodicities are an intrinsic feature of resting state EEG derived microstate sequences. Periodicities represent one of the main classes of non-stationarities that should be identified before further analyzing time series (Box and Jenkins, 1976). Oscillations lead to a very different shape of the AIF and to a much slower decay of the AIF compared to a pure first-order Markov model. However, beyond the oscillatory regime (>1000 ms), the AIF is not distinguishable from a Markov model.

The conditions of periodicity and asymmetry of the transition matrix together suggest that the biophysical process underlying neural activity is in a non-equilibrium state for the following reasons: If the underlying process was in equilibrium, the transition rates between the states would be reversible, and the transition matrix should be symmetric. Asymmetric periodicity however, shows a directed cycling through the set of microstates. This is only possible in the case of a source of free energy, a 'driving force'. The mathematical connection between non-reversible cycling through a discrete set of states and the existence of a free energy source has been detailed in (Rengifo et al., 2002). In the light of neural mass models of cortical activity the driving force would be the constant thalamic input relaying peripheral input (Deco et al., 2008). In these models, action potentials from the thalamus excite cortical oscillators comprised of coupled excitatory and inhibitory neuronal populations giving rise to the oscillatory activity recorded by EEG. Using this

information, it will be interesting to compare groups with and without symmetric transition matrices. In simultaneous EEG-fMRI data sets, the model-derived hypothesis of thalamic input as the discriminating factor could be tested directly.

Future studies should further test the correspondence between EEG and microstate oscillatory activity under different experimental conditions, in health and disease.

4.4. Short-range memory

The range of memory effects in microstate sequences is an interesting question in itself. The presence of long-range temporal dependencies in EEG time series has been investigated and debated. Often, Hurst exponents of $H > 0.5$ were interpreted as a sign of long-range correlations in EEG time series and microstate sequences (Linkenkaer-Hansen et al., 2001; Van de Ville et al., 2010; Gschwind et al., 2015). However, the Hurst phenomenon ($H > 0.5$) does not imply long-range correlations by itself, as it may be reproduced by short-range or even uncorrelated signals with non-stationary increments (Bassler et al., 2006; McCauley et al., 2007). Inferring long-range correlations from the Hurst phenomenon is only valid for sequences with stationary increments. As shown here, non-stationarity is a frequent phenomenon in microstate sequences (Table 3). We therefore addressed long-memory directly (Fig. 8) and find that the AIF decays to the first-order Markov confidence interval, i.e. that microstate sequences are short-range correlated for large time lags. Our results are almost identical to resting state EEG analyses obtained with the diffusion entropy approach (Ignaccolo et al., 2010b, a), where spurious long-range correlations could be resolved as regimes of transient periodicities followed by short-memory behaviour at large time lags, i.e. for time lags > 1000 ms. It should also be noted that the preprocessing filters applied to the EEG do no limit the interpretation of the results. As the length of a contiguous microstate segment is only determined by its spatial similarity with the current EEG topography, 1 Hz high-pass filtered EEG can still yield microstate sequences with autocorrelations up to more than 10 s (Van de Ville et al., 2010; von Wegner et al., 2016). While the Hurst approach is useful for mathematically well characterized processes such as fractional Gaussian noise (Mandelbrot and Van Ness, 1968), a Hurst exponent alone is not enough to infer time-correlation properties of empirical data. We here provided further evidence that information-theoretical analysis is a useful and appropriate tool to understand EEG dynamics.

4.5. Neurobiological interpretation

Microstate analyses allow the characterization of spatio-temporal patterns of ongoing EEG activity. While oscillatory dynamics of individual EEG channels are a prominent and thoroughly investigated feature, the temporal characteristics of ongoing spatial patterns in the resting state are little understood. Our analysis shows that spatial EEG patterns are not generated in a memoryless way, but that these patterns are replayed in recurring patterns, although Fig. 6 shows that the robustness of these oscillating motifs differs between subjects.

The classical microstate interpretation is that of metastable network configurations, where certain scalp potential distributions remain almost constant on time scales of tens to hundreds of milliseconds (Koenig et al., 2002). Our results are not in contradiction to these findings but add extra information about typical microstate dynamics, especially about recurrence. We find average microstate lifetimes and right-skewed lifetime distributions similar to those reported previously, Tables S1–S4 (Brodbeck et al., 2012; Gschwind et al., 2015). In this context it is important to note that the information peak around 50 ms does not mean that the microstate labels change every 50 ms, neither does an average microstate lifetime of 50 ms predict a 20 Hz periodicity. Both points become clear using the Markov null hypothesis as an example. The Markovian lifetime distribution is an exponential that can be parametrized by its mean value, e.g. $\tau = 50$ ms. This means that the probability p of finding a lifetime larger than t is $p = \exp\left(-\frac{t}{\tau}\right)$. For instance, we expect 13.5% of the lifetimes to be longer than 100 ms and 1.8% to be longer than 200 ms. Moreover, we do not expect any periodicities of the autoinformation function, Eq. (9). Looking at our empirical microstate sequences, we find more complex dynamics than predicted by a first-order Markov process. In particular, we find average lifetimes around 50 ms and maximum lifetimes up to several hundreds of milliseconds. We also find that certain microstates, though not all of them, re-appear regularly and the smallest time constant for their re-appearance is 50 ms. Adding the frequency-doubling argument, our results are in line with the classical microstate interpretation of network activity on the alpha-frequency time scale. The new information added is that these networks seem to activate periodically and that transitions between maps do not follow a rigid rule but change over time (non-stationarity). Another observation that asks for further investigation is the contribution of different microstates to global periodicity. In our set of 20 subjects, in most cases (15/20) all microstates contributed to the global AIF. The pattern shown for subject S20 in the right column of Fig. 7, where only the horizontal microstate contributes to global periodicity, was observed in 5/20 cases. As the functional significance of different microstate maps remains still discussed, future studies should investigate the roles and contributions of different topographies to defined experimental conditions.

Interestingly, the peaks of the autoinformation function yield distinct time scales that can be used to characterize the microstate generating mechanism. Conceptually, these time scales stand in contrast to a recently presented scale-free model of microstate sequences (Van de Ville et al., 2010). The latter proposes that the sequence of microstate maps does not display any characteristic time scale, and that long-range (infinite) autocorrelations are present. The differing results can be explained by the analytic approach chosen. As Hurst analysis is focused on fitting an asymptotic scaling exponent, it will not detect oscillations in the intermediate time-scale region. This problem is treated thoroughly in (Ignaccolo et al., 2010b,a), where spurious scaling behaviour in EEG resting state recordings is detected and explained. Our argument against scale-free, long-range correlated microstate dynamics is constructive by approach, providing explicit time scales for each microstate sequence, and showing that their long-range memory content is finite.

The presence of characteristic time scales and oscillations is in line with biophysical models of cortical activity. As these models are based on time constants from physiological experiments, their output can be characterized by a set of time scales. Thus, the periods of alpha and beta oscillations, for instance, can be derived from synaptic time constants and conduction delays (Robinson et al., 2001). Long-range correlated and fractal models of brain activity, on the other hand, propose a scale-free behaviour characterized by a single parameter, the Hurst exponent H. In summary, our findings point more towards classical biophysical models than in the direction of scale-free models of global cortical activity.

5. Conclusion

Microstate sequences represent continuous EEG dynamics based on a spatial variance criterion. These discrete sequences are the result of a complex, non-linear dimensionality reduction transform of the EEG signal. Our results show that EEG microstate sequences often exhibit non-stationarity on the time scale of several tens of seconds. In these cases, some signal processing strategies assuming time-stationarity may give erroneous results. For larger window sizes, e.g. 40 s, the fraction of stationary sequences increases, suggesting a tendency towards stationarity at these time scales. On the other hand, we know that resting state recordings lasting several minutes often show non-stationarity due to vigilance fluctuations, i.e. intrusions of light sleep, changing spectral EEG properties severely (Tagliazucchi et al., 2012).

In conclusion, we recommend to explicitly test the stationarity of microstate sequences shorter than 30 s, and to independently control for

vigilance fluctuations in EEG recordings with a duration larger than five minutes. It is quite remarkable that four (and even fewer) topographic EEG maps capture such important EEG features as periodicity and temporal autocorrelation structure. Using our approach, raw EEG and microstate sequences can be tested for the same properties and thus, microstates can be interpreted in the same way as their continuous EEG source. The methods presented here will hopefully facilitate researchers in both areas, i.e. continuous EEG and discrete microstates, to compare results obtained with both methods and to shift between both modalities while staying within the same theoretical framework.

Disclosure/conflict-of-interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

This work was funded by the Bundesministerium für Bildung und Forschung (grant 01 EV 0703) and LOEWE Neuronale Koordination Forschungsschwerpunkt Frankfurt (NeFF).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.neuroimage.2017.06.062.

References

Anderson, T.W., Goodman, L.A., 1957. Statistical inference about Markov chains. Ann. Math. Statistics 28, 89–110.
Bassler, K.E., Gunaratne, G.H., McCauley, J.L., 2006. Markov processes, hurst exponents, and nonlinear diffusion equations: with application to finance. Phys. A 369, 343–353.
Billingsley, P., 1961. Statistical methods in Markov chains. Ann. Math. Stat. 32, 12–40.
Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control. Holden-Day.
Brodbeck, V., Kuhn, A., von Wegner, F., Morzelewski, A., Tagliazucchi, E., Borisov, S., Michel, C.M., Laufs, H., 2012. EEG microstates of wakefulness and NREM sleep. NeuroImage 62, 2129–2139.
Brunet, D., Murray, M.M., Michel, C.M., 2011. Spatiotemporal analysis of multichannel EEG: CARTOOL. Comput. Intell. Neurosci. 2011, 813870.
Deco, G., Jirsa, V.K., Robinson, P.A., Breakspear, M., Friston, K., 2008. The dynamic brain: from spiking neurons to neural masses and cortical fields. PLoS Comput. Biol. 4, e1000092.
Feller, W., 1971. An Introduction to Probability Theory and its Applications, second ed., vol. II. John Wiley & Sons Inc., New York.
Gärtner, M., Brodbeck, V., Laufs, H., Schneider, G., 2015. A stochastic model for EEG microstate sequence analysis. NeuroImage 104, 199–208.
Gschwind, M., Michel, C.M., Ville, D.V.D., 2015. Long-range dependencies make the difference-comment on "a stochastic model for EEG microstate sequence analysis". NeuroImage 117, 449–455.
Häggström, O., 2002. Finite Markov Chains and Algorithmic Applications. Cambridge University Press.
Ignaccolo, M., Latka, M., Jernajczyk, W., Grigolini, P., West, B.J., 2010a. The dynamics of EEG entropy. J. Biol. Phys. 36, 185–196.
Ignaccolo, M., Latka, M., Jernajczyk, W., Grigolini, P., West, B.J., 2010b. Dynamics of electroencephalogram entropy and pitfalls of scaling detection. Phys. Rev. E Stat. Nonlin Soft Matter Phys. 81, 031909.
Koenig, T., Lehmann, D., Merlo, M.C., Kochi, K., Hell, D., Koukkou, M., 1999. A deviant EEG brain microstate in acute, neuroleptic-naive schizophrenics at rest. Eur. Arch. Psychiatry Clin. Neurosci. 249, 205–211.
Koenig, T., Prichep, L., Lehmann, D., Sosa, P.V., Braeker, E., Kleinlogel, H., Isenhart, R., John, E.R., 2002. Millisecond by millisecond, year by year: normative EEG microstates and developmental stages. NeuroImage 16, 41–48.
Kuhn, A., Brodbeck, V., Tagliazucchi, E., Morzelewski, A., von Wegner, F., Laufs, H., 2015. Narcoleptic patients show fragmented EEG-microstructure during early NREM sleep. Brain Topogr. 28, 619–635.
Kullback, S., 1959. Information Theory and Statistics. Dover Publications, Inc., Mineola, NY.
Kullback, S., Kupperman, M., Ku, H.H., 1962. Tests for contingency tables and Markov chains. Technometrics 4, 573–608.
Lehmann, D., Ozaki, H., Pal, I., 1987. EEG alpha map series: brain micro-states by space-oriented adaptive segmentation. Electroencephalogr. Clin. Neurophysiol. 67, 271–288.
Linkenkaer-Hansen, K., Nikouline, V.V., Palva, J.M., Ilmoniemi, R.J., 2001. Long-range temporal correlations and scaling behavior in human brain oscillations. J. Neurosci. 21, 1370–1377.
Mandelbrot, B.B., Van Ness, J.W., 1968. Fractional Brownian motions, fractional noises and applications. SIAM Rev. 10, 422–437.
McCauley, J.L., Gunaratne, G.H., Bassler, K.E., 2007. Hurst exponents, markov processes, and fractional brownian motion. Phys. A 379, 1–9.
Murray, M.M., Brunet, D., Michel, C.M., 2008. Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr. 20, 249–264.
Niedermeyer, E., da Silva, F.L., 2005. Electroencephalography. Basic Principles, Clinical Applications, and Related Fields, 5 ed. Williams & Wilkins, Lippincott.
Pascual-Marqui, R.D., Michel, C.M., Lehmann, D., 1995. Segmentation of brain electrical activity into microstates: model estimation and validation. IEEE Trans. Biomed. Eng. 42, 658–665.
Peng, C.K., Buldyrev, S.V., Goldberger, A.L., Havlin, S., Sciortino, F., Simons, M., Stanley, H.E., 1992. Long-range correlations in nucleotide sequences. Nature 356, 168–170.
Rengifo, J., Rosales, R., Gonzlez, A., Cheng, H., Stern, M.D., Ros, E., 2002. Intracellular Ca(2+) release as irreversible Markov process. Biophys. J. 83, 2511–2521.
Riley, M.A., Bonnette, S., Kuznetsov, N., Wallot, S., Gao, J., 2012. A tutorial introduction to adaptive fractal analysis. Front. Physiol. 3, 371.
Robinson, P.A., Rennie, C.J., Wright, J.J., Bahramali, H., Gordon, E., Rowe, D.L., 2001. Prediction of electroencephalographic spectra from neurophysiology. Phys. Rev. E Stat. Nonlin Soft Matter Phys. 63, 021903.
Tagliazucchi, E., von Wegner, F., Morzelewski, A., Brodbeck, V., Laufs, H., 2012. Dynamic BOLD functional connectivity in humans and its electrophysiological correlates. Front. Hum. Neurosci. 6, 339.
Van de Ville, D., Britz, J., Michel, C.M., 2010. EEG microstate sequences in healthy humans at rest reveal scale-free dynamics. Proc. Natl. Acad. Sci. U. S. A. 107, 18179–18184.
Wackermann, J., Lehmann, D., Michel, C.M., Strik, W.K., 1993. Adaptive segmentation of spontaneous EEG map series into spatially defined microstates. Int. J. Psychophysiol. 14, 269–283.
von Wegner, F., Tagliazucchi, E., Brodbeck, V., Laufs, H., 2016. Analytical and empirical fluctuation functions of the eeg microstate random walk - short-range vs. long-range correlations. Neuroimage 141, 442–451.
Yuan, H., Zotev, V., Phillips, R., Drevets, W.C., Bodurka, J., 2012. Spatiotemporal dynamics of the brain at rest–exploring EEG microstates as electrophysiological signatures of BOLD resting state networks. NeuroImage 60, 2062–2072.