

Symbolic dynamics of event-related brain potentials

Peter beim Graben,* J. Douglas Saddy, and Matthias Schlesewsky
Institute of Linguistics, Universität Potsdam, P.O. Box 601553, D-14415 Potsdam, Germany

Jürgen Kurths
Institute of Physics, Nonlinear Dynamics Group, Universität Potsdam, P.O. Box 601553, D-14415 Potsdam, Germany
 (Received 13 January 2000)

We apply symbolic dynamics techniques such as word statistics and measures of complexity to nonstationary and noisy multivariate time series of electroencephalograms (EEG) in order to estimate event-related brain potentials (ERP). Their significance against surrogate data as well as between different experimental conditions is tested. These methods are validated by simulations using stochastic dynamical systems with time-dependent control parameters and compared with traditional ERP-analysis techniques. Continuous EEG data are cut into *epochs* according to stimuli events presented to the subjects. These ensembles of time series can be considered as ensembles of trajectories given by some dynamical systems. We employ a statistical mechanics approach motivated by the Frobenius-Perron equation and apply it to coarse-grained symbolic descriptions of the dynamics. We develop time-dependent measures of complexity founded on running cylinder sets and show that these quantities are able to distinguish simulated data obtained by different control parameters as well as experimental data between different experimental conditions. As a first finding, our approach restores the well-known ERP components and it reveals additionally qualitative changes in the EEG that cannot be detected by means of the traditional techniques. We criticize the prerequisites of the traditional approach to ERP analysis and propose to consider ERP instead in terms of dynamical system theory and information theory.

PACS number(s): 87.19.Nn, 02.50.-r, 05.45.Tp

I. INTRODUCTION

The structural components of the brain, neurons, and their ionic channels situated in the cell membranes are known to be nonlinear devices. The resistivities of ionic channels are generally nonlinear functions of voltage or the chemical concentration of transmitter substances, and neurons fire only if the somatic potential at the axon hillock cross a certain activity threshold [1,2]. The activity of large formations of neurons is macroscopically measurable as the electroencephalogram (EEG) at the human scalp. The sources of the EEG are assumed to be extracellular ionic currents driven by voltage gradients between excitatory postsynaptic potentials at the apical dendrites and inhibitory postsynaptic potentials at the neurons somata [3,4]. Hence, it seems to be promising to apply nonlinear techniques of data analysis to EEG data records in order to detect qualitative changes of the brains dynamics according to Haken [5] and Kelso [6].

This has been attempted by several authors for different brain states such as α -activity, sleep stages, epileptic seizures, diseases, and cognitive tasks by computing the correlation dimension [7–18]. But some of these findings reporting low-dimensional chaotic attractors have been criticized. Layne *et al.* argued that EEG time series are essentially nonstationary, but concepts of dynamical system theory such as attractors or fractal dimensions are only well defined for stationary behavior [11]. Further objections are due to Osborne and Provenzale, who showed that colored noise also entails low values of correlation dimensions [19], and due to Rapp

et al., who demonstrated that filtered noise can spuriously exhibit low dimensions [20]. Theiler *et al.* applied the technique of surrogate data to correlation dimensions of EEG and reported that there is no evidence for low-dimensional chaos but significance for nonlinearity [21].

More recently, the concept of pointwise dimensions has been utilized and applied to EEG data, which are nonstationary by definition: these are event-related potentials (ERP), i.e., EEG recordings triggered by external events such as different beeps or words occurring at a monitor [22]. In the beep paradigm (the so-called oddball experiment), subjects are presented beeps of two different frequencies with different probabilities. After averaging all EEG trials corresponding to the rare event, there is a significant positive voltage deflection 300 ms after the stimulus, called the P300 ERP component. Molnár *et al.* reported a drop of the pointwise dimension as a function of time corresponding to the P300 component [22].

Another approach to analyzing natural data is based on the concepts of symbolic dynamics [23,24] and complexity measures [25,26]. They have been successfully applied to physiological data such as cardiorespiratory time series [27,28], movement control [29–31], or bone structure [32], but also to neuronal spike trains obtained by electrophysiological measurements [33]. Symbolic dynamics is able to tackle nonstationarity as has been shown by Schwarz *et al.*, who applied these techniques to spectrograms obtained from astrophysical data [34], and, more recently, by Buchner and Zebrowski, who discussed magnetic systems [35]. In both cases the authors deal with time-dependent measures of complexity, such as Shannon entropy and algorithmic complexity, where symbolic dynamics are given either by the spectral powers at certain times [34] or by vectors constructed from delay embeddings [35].

*Electronic address: peter@ling.uni-potsdam.de; also at Inst. of Physics, Nonlinear Dynamics Group, Universität Potsdam, D-14415 Potsdam, Germany.

In this study we present an approach that combines the concepts of statistical mechanics of transient behavior and of symbolic dynamics to analyze ensembles of ERP time series. We show that the information theoretic notions of cylinder sets and their probability measures are appropriate tools in order to deal with ensembles of nonstationary data. The calculated running cylinder entropies and word statistics are able to capture the coherency and disorder of ERP. Large entropy drops correspond with traditional ERP voltage averages just as the pointwise dimensions reported above. But on the other hand, measures of complexity are able to detect phase transitions towards higher disorder that cannot be revealed by averaging voltages.

In order to validate our approach, we perform simulations with nonlinear maps with time-dependent control parameters, uncertainty of initial conditions, and disturbed by noise. We choose these conditions because it is more plausible that ERP data are given by nonstationary stochastic dynamical systems rather than by deterministic signals disturbed by purely observational stationary and ergodic noise.

The organization of the paper is as follows. In the second section we briefly report the technique and requirements of the traditional approach to analyze ERP data. In the third section we present our data-analysis technique, which is based on the statistical mechanics of transient dynamics and provides the basic concepts of symbolic dynamics and measures of complexity. The fourth section is devoted to results of our simulations of noisy and nonstationary dynamical systems. Finally, we analyze data from a language-processing ERP experiment and compare the traditional approach of ERP analysis with results of our techniques.

II. TRADITIONAL ANALYSIS OF ERP DATA

Continuous EEG data are multivariate time series recorded as real-valued matrices $X_{ij} \in \mathbb{R}$, $1 \leq i \leq K+1$, $1 \leq j \leq L$, where K is the number of recording electrodes at the subject's scalp (the additional channel $K+1$ contains the *trigger codes* which mark the stimuli events in time) and L is the number of total recording samples. After digital filtering of this data set in certain pass bands and after manual rejection of artifacts, i.e., cutting eye blinks, saccades, voltage drifts, or electromyograms scattered into the EEG, the continuous EEG is split into *epochs* according to the time marks given by the trigger codes in channel $K+1$. Each epoch is a short time series consisting of a *prestimulus interval* $[-T, 0]$ and a *poststimulus interval* $[0, T_p]$, where $t=0$ is centered at the stimulus event. These epochs are collected into ensembles corresponding to different experimental conditions, say $x_i^{c_1}(t)$, $x_j^{c_2}(t)$, where i, j ($1 \leq i \leq N^{c_1}, 1 \leq j \leq N^{c_2}$) are the epoch indices and t is the (discrete) time index ranging from $-T$ to T_p . N^{c_1}, N^{c_2} are the numbers of epochs for both ensembles. The indices c_1 and c_2 refer to experimental conditions.

In order to explain the next steps of data analysis, we shall give a brief theoretical description of the method [36–38]. The classical approach of ERP data analysis requires that each ERP epoch $x_i(t)$ is a realization of a stochastic process [84],

$$x_i(t) = s(t) + \eta_i(t), \quad (1)$$

where $s(t)$ is assumed to be the invariant ERP signal that should be uncorrelated to the background EEG activity (a rather unrealistic assumption) and that obeys

$$s(t) = 0 \quad \text{for } t < 0, \quad (2)$$

i.e., there is no event-related potential in the prestimulus interval at all. For each epoch, $\eta_i(t)$ is a stationary and ergodic stochastic process mimicking the spontaneous EEG activity and observational noise so that the different $\eta_i(t), \eta_j(t)$ must be stochastically independent for $i \neq j$ and all $\eta_i(t)$ will have the same family of distribution functions $F_i(x)$, $F_{t_1, t_2}(x_1, x_2)$, $F_{t_1, t_2, t_3}(x_1, x_2, x_3)$, etc.

The first step in the further analysis is a correction of the baseline because it cannot be assumed that the time averages of the processes $\eta_i(t)$ will vanish generally. One therefore computes the averages of the prestimulus intervals

$$\beta_i = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^0 x_i(t) dt. \quad (3)$$

According to Eq. (2), this leads to the random numbers

$$\beta_i = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^0 \eta_i(t) dt. \quad (4)$$

Next, one subtracts these baseline values from the corresponding ERP epochs,

$$\zeta_i(t) = x_i(t) - \beta_i = s(t) + \eta_i(t) - \beta_i.$$

From these processes one computes the (empirical) means

$$\overline{\zeta(t)} = \frac{1}{N} \sum_{i=1}^N \zeta_i(t) = s(t) + \frac{1}{N} \sum_{i=1}^N [\eta_i(t) - \beta_i]$$

obtaining a stochastic process again. These mean values are estimators of the expectation values

$$E(\overline{\zeta(t)}) = E\left(s(t) + \frac{1}{N} \sum_{i=1}^N [\eta_i(t) - \beta_i]\right) = s(t),$$

which makes use of the stationarity obtaining $E(\eta_i(t)) = e_i$, of the fact that the processes $\eta_i(t)$ are identically distributed, which yields $e_i = e$, and finally of the ergodicity property providing $e_i = e = E(\beta_i)$ due to Birkhoff's ergodic theorem (see, e.g., [39], p. 313).

In order to improve the signal-to-noise ratio (SNR, [40]) of the ERP signal $s(t)$, one needs many trials per condition and per subject and many subjects per experiment. The number of trials is just the ensemble size of measurement epochs N . The improvement in SNR with increasing ensemble size is given by \sqrt{N} [38] when the signal $s(t)$ remains invariant from trial to trial. But this assumption is also never met in real EEG data: there are changes in amplitude as well as in latency time (e.g., the signal onset time) from trial to trial,

e.g., caused by habituation, learning, or changes in attention [41–43,40]. Thus, it is more realistic to describe the signal content of EEG at least by the ansatz [44]

$$s_i(t) = s(t + \tau_i) \quad (5)$$

in order to counter the latency jitter problem, where $s(t)$ is the invariant signal as before but it is randomly shifted in time by some discrete stochastic process τ_i with mean $E(\tau_i) = 0$ and variance $D^2(\tau_i) = \sigma_\tau^2$. As before, one assumes that all τ_i are independent and identically distributed. Note that, if $s(t)$ is analytic, we may expand it into a Taylor series,

$$s(t + \tau_i) = s(t) + \dot{s}(t)\tau_i + \frac{1}{2}\ddot{s}(t)\tau_i^2 + \dots \quad (6)$$

For the expectation value of the empirical mean, we get then in second order

$$E\left(\frac{1}{N}\sum_{i=1}^N s_i(t)\right) = s(t) + \frac{1}{2}\ddot{s}(t)\sigma_\tau^2.$$

The latency jitter, therefore, smears the ERP signal out in the course of time and makes the SNR worse [41,38,45].

Finally, the averaged epochs of all subjects will be averaged again in order to obtain the *grand average* of the whole experiment. Since one assumes stochastic independence of the single trial epochs, the subject averages are approximately Gaussian distributed according to the central limit theorem. Hence, the grand averages can be treated by classical statistical tests, such as running *t*-tests or multivariate analyses of variance (MANOVA) in order to decide whether means are significantly different between channels, time windows, subjects, and experimental conditions [38]. Note that the analysis of variance requires stationarity of variances. This approach yields typical wave forms of averaged voltages which are classified according to their polarity and latency time (generally the time points of extremal values), such as N100 (a negative peak after 100 ms that indicates a shift of the subject's attention towards the stimulus), P300 (a positive peak after 300 ms indicating surprise), and language-processing-related components such as N400 (negativity after 400 ms arising from semantically implausible words) or P600 (positive wave after 600 ms elicited by unexpected grammatical properties) [46–48,4,49,50]. These averaged voltage waves can be subjected to further statistical analyses such as principal component analysis for revealing superimposed subcomponents and their factorial loads [51,46,38,4], or techniques from linear system analysis in order to compute transient response frequency characteristics of the neural tissue [52].

III. SYMBOLIC DYNAMICS OF ERP DATA

A. The dynamical approach to event-related potentials

It is important to summarize the prerequisites of the traditional ERP averaging technique: the EEG epochs are supposed to be superpositions of an event-related signal $s(t)$ that should be invariant from trial to trial and also from

subject to subject and realizations of independent and identically distributed stationary and ergodic stochastic processes $\eta_i(t)$. These conditions are usually not met in real ERP data [41–43,40,45]. The signal $s(t)$ at least jitters in time, and correlations between the signal and the background EEG and within the background EEG across trials provide a deterioration of the signal-to-noise ratio. Thus, averaging of signals that are not time-locked according to the stimulus yields a damping of wave forms. Note that ansatz (1) states that there is no impact of the noise on the dynamics. It is assumed to be purely observational noise. Next, we develop a data-analysis technique that overcomes these strong assumptions.

Our approach to ERP data analysis is related to ideas of Kelso and Başar, who consider brain activity in terms of *synergetics* where experimental manipulations are *control-parameters* of the system that undergoes *phase transitions* at critical parameter values [6,53–55]. Başar also suggests describing the coherency of brain activity by entropy as an *order parameter* like the magnetization in ferromagnets ([52], p. 193). Thus, we are going to describe event-related potentials in terms of dynamical system theory and propose a way to compute measures of complexity from the EEG data reflecting qualitative changes or phase transitions of brain dynamics depending on experimental conditions considered as control parameters.

In the following subsections we discuss the theoretical issues of statistical mechanics of deterministic and stochastic dynamical systems exhibiting transient behavior. We derive an evolution equation of their probability measures in the state space, the Frobenius-Perron equation, and show how these probability measures can be related to a coarse-grained description of the dynamical system by means of cylinder sets. Then, we apply classical measures of complexity, such as Shannon and Rényi entropies, to time-dependent probability distributions of words given by a symbolic dynamics. Finally, we provide statistical tests — surrogate data and a running χ^2 test — in order to decide whether changes in the complexity measures are significant either against the null hypotheses of Bernoulli processes or between different experimental conditions.

B. Statistical mechanics of transient dynamics

In order to explain the physical ideas of our approach, we will discuss noisy one-dimensional and time discrete dynamical systems due to Crutchfield and Packard [56,57]. A deterministic dynamical system, like the logistic map $x_{t+1} = rx_t(1 - x_t)$, $x_t \in [0,1]$, $r \in [0,4]$, is a pair (X, Φ_r) , where X is the state space (sometimes called phase space) and $\Phi_r : X \rightarrow X$ is a map that assigns to a state x_t at a certain time t its successor x_{t+1} . The logistic map lives in the unit interval $[0, 1]$ as its state space and the map is just given by $\Phi_r(x) = rx(1 - x)$, where r is the control parameter of the system. Since ERP data are nonstationary by definition (otherwise there would be no average voltage deflections in the course of time), we need a statistical formalism of transient behavior of dynamical systems. This can be provided by a statistical description of initial conditions even if the system is purely deterministic. Let us prepare an ensemble of initial conditions according to a probability measure μ_0 . In the following, we shall assume the measure to be absolutely con-

tinuous, so there is a probability density function $\rho_0(x)$ defined on the state space and the measure of a set $A \subset X$ is given by

$$\mu_0(A) = \int_A \rho_0(x) dx. \quad (7)$$

Now let the system evolve. What is the image of the set A under the action of the map Φ_r after time t ? In the deterministic case, the image $B = \Phi_r^t(A)$, where Φ_r^t is recursively defined by iterating the map ($\Phi_r^t = \Phi_r \circ \Phi_r^{t-1}$), must have the same probability as A . Therefore, it holds that $\mu_0(A) = \mu_t(\Phi_r^t(A))$ with μ_t as the probability measure after time t . This identity can be reformulated as

$$\mu_t(B) = \mu_0(\Phi_r^{-t}(B)) \quad (8)$$

for any measurable subset $B \subset X$ with $\Phi_r^{-t}(B) = (\Phi_r^t)^{-1}(B)$ as the preimage of B after time t . This is the Frobenius-Perron equation of the dynamics [58]. One can also introduce a Frobenius-Perron operator acting on the probability density functions by

$$\rho_t(x) = (L_{\Phi_r^t} \rho_0)(x) = \int dy \delta(x - \Phi_r^t(y)) \rho_0(y). \quad (9)$$

This operator mediates the map Φ_r^t , acting on the states, to a function at the space of probability distribution densities. Next, we are going to introduce some external noise into the system that influences the dynamics by giving the equation of motion

$$x_{(t+1)} = \Phi_r(x_t) + \xi_{t+1}, \quad (10)$$

where ξ_t is a discrete stochastic process. For the sake of simplicity, let us assume ξ_t to be a Markov process. Then, ξ_t is completely described by its transition probability density functions $f(t_1, x_1; t_2, x_2)$ determining the probability for obtaining the state x_2 at time t_2 given the state x_1 at time t_1 . Note that we assume neither stationarity nor ergodicity of this process. In analogy to [59,57], a Frobenius-Perron operator of this noisy dynamics can be constructed if we set

$$\rho_{t+1}(x) = \int dy \rho_t(y) f(t, \Phi_r(y); t+1, x). \quad (11)$$

This leads by recursion to the evolution equation of the density ρ_t ,

$$\begin{aligned} \rho_t(x) = & \int d^t y \rho_0(y_1) f(0, \Phi_r(y_1); 1, y_2) \\ & \times f(1, \Phi_r(y_2); 2, y_3) \cdots f(t-1, \Phi_r(y_t); t, x). \end{aligned} \quad (12)$$

In order to include nonstationarity of the dynamics, we also allow the control parameter r to depend on time: $r = r(t)$.

C. Symbolic dynamics

Next, we shall introduce symbolic dynamics of transient behavior. A coarse-grained description of the dynamics that lets ‘‘robust’’ properties such as periodicity of the system investigated invariant can be gained by a partition covering the state space of the dynamical system [60,23]. Let $\{A_i | i = 1, 2, \dots, I\}$ be a family of I pairwise disjoint subsets covering the whole state space X , i.e., $\cup_{i=1}^I A_i = X$, $A_i \cap A_j = \emptyset$, $i \neq j$. The index set $A = \{1, 2, \dots, I\}$ of the partition can be interpreted as a (finite) alphabet of letters $a_i = i$. The expression $A^{\mathbb{Z}}$ refers to the set of all bi-infinite strings of letters $s = \cdots a_{i-1} a_{i_0} a_{i_1} a_{i_2} \cdots$ from A . Given a time-discrete and invertible deterministic dynamics Φ_r [85] we construct a map $\pi: X \rightarrow A^{\mathbb{Z}}$ that assigns initial conditions $x_0 \in X$ to infinite symbol strings $s \in A^{\mathbb{Z}}$ by the rule $\pi(x_0) = s$, if $x_t = \Phi_r^t(x_0) \in A_{i_t}$, $t \in \mathbb{Z}$. Thus, π maps initial conditions x_0 in the state space onto symbolic strings regarded as coarse-grained trajectories starting at x_0 . A sequence $s \in A^{\mathbb{Z}}$ is called admissible by the dynamics Φ_r if there is an initial condition $x_0 \in X$ mapped onto s by π . The theory of symbolic dynamics deals with sets of admissible sequences $\Sigma \subset A^{\mathbb{Z}}$. Note that the map π might not be invertible. If π is invertible, the partition is called generic and every string of symbols corresponds to exactly one initial condition generating it [58].

Nevertheless, we can apply the map π^{-1} at subsets of $A^{\mathbb{Z}}$, namely at strings of finite length, looking for their preimages in X . In order to do this, we introduce now the basic notions of this paper.

Let $t \in \mathbb{Z}$, $n \in \mathbb{N}$, and $a_1, \dots, a_n \in A$. The set

$$[a_1, \dots, a_n]_t = \{s \in A^{\mathbb{Z}} | s_{t+k-1} = a_{i_k}, k = 1, \dots, n\} \quad (13)$$

is called *n-cylinder* at time t . The symbol sequence $w = (a_1, \dots, a_n) \in A^n$ is called *n-word*, where A^n denotes the set $A^{n-1} \times A$ of n -tuples of symbols.

This definition goes back to McMillan [61], though he did not use the term ‘‘cylinder set.’’ For recent references, see [56–58,25]. For an instructive example, consider the set of all books ever printed. Then the cylinder $[\text{book}]_{100}$ is the subset of all books having the word ‘‘book’’ beginning with the 100th letter. Note that this cylinder is different from the set $[\text{book}]_{250}$ containing all books that have the word ‘‘book’’ beginning with the 250th letter. Since cylinders are subsets of $A^{\mathbb{Z}}$ of infinite strings coinciding in a (discrete) time interval $\{t, t+1, \dots, t+n-1\}$, we can determine their preimages under the deterministic dynamics Φ_r ,

$$\begin{aligned} \pi^{-1}([a_1, \dots, a_n]_t) = & \{x_t \in X | x_t \in A_{i_1} \wedge x_{t+1} \in \Phi_r^{-1}(A_{i_2}) \wedge \\ & \cdots \wedge x_{t+n-1} \in \Phi_r^{-n+1}(A_{i_n})\}, \end{aligned} \quad (14)$$

where Φ_r^{-1} denotes the preimages of a set.

Finally, let us endow the state space with a probability measure μ_0 of initial conditions again. Due to the Frobenius-Perron equation, we obtain a family of measures μ_t attached to each point in time t .

Let $t \in \mathbb{Z}$, $n \in \mathbb{N}$, and $a_{i_1}, \dots, a_{i_n} \in A$ as before. The measure

$$\begin{aligned} p(a_{i_1}, \dots, a_{i_n} | t) &= \mu_t([a_{i_1}, \dots, a_{i_n}]_t) \\ &= \int dx \rho_t(x) \chi_{[a_{i_1}, \dots, a_{i_n}]} t(x) \end{aligned} \quad (15)$$

is called *word probability* of the n -word $(a_{i_1}, \dots, a_{i_n}) \in A^n$ at time t . The distribution of all word probabilities for a given length n is called *word statistics*. According to the example given above, the measure of the cylinder $[\text{book}]_{100}$ is just the cardinality of this set normalized to the total amount of books ever printed. Again, the measures of the sets $[\text{book}]_{100}$ and $[\text{book}]_{250}$ are certainly different.

In the definition above, $\chi_{[a_{i_1}, \dots, a_{i_n}]} t(x)$ refers to the characteristic function of the cylinder $[a_{i_1}, \dots, a_{i_n}]_t$ considered as a subset of X ,

$$\chi_{[a_{i_1}, \dots, a_{i_n}]} t(x) = \begin{cases} 1: & x \in \pi^{-1}([a_{i_1}, \dots, a_{i_n}]_t) \\ 0: & x \notin \pi^{-1}([a_{i_1}, \dots, a_{i_n}]_t). \end{cases} \quad (16)$$

For a stochastic nonlinear dynamics such as Eq. (10), the maps π, π^{-1} are no longer defined and the distinction between admissible and nonadmissible sequences will be generally destroyed. However, the fluctuations change the word probabilities. Hence, we take a rather pragmatic point of view regarding the cylinder sets as generated by a dynamics which is assumed to be deterministic, while their probabilities have been changed due to the impact of noise (a similar consideration has been made by Tang and Tracy [24]). Doing so, we can apply the Frobenius-Perron equation obtained for the noisy dynamics (12) in order to compute the word probabilities (15),

$$\begin{aligned} p(a_{i_1}, \dots, a_{i_n} | t) &= \int dx \int d^t y \rho_0(y_1) \\ &\quad \times f(0, \Phi_r(y_1); 1, y_2) \cdots f(t-1, \Phi_r(y_t); t, x) \\ &\quad \times \chi_{[a_{i_1}, \dots, a_{i_n}]} t(x). \end{aligned} \quad (17)$$

To this end, we supposed a deterministic dynamics to compute the characteristic functions $\chi_{[a_{i_1}, \dots, a_{i_n}]} t(x)$ and then we introduced noise in order to determine the probabilities of these maps under the influence of fluctuations.

To describe effects such as the latency jitter of ERP epochs, we introduce a further random variable τ that is discretely distributed according to $\{(t, p_t) | t \in \mathbb{Z}\}$ with vanishing mean $E(\tau)$ and finite variance $D^2(\tau)$. By means of τ , we shift initial conditions x_0 in time to $x'_0 = x_\tau$. In order to determine the measures of cylinder sets from shifted initial conditions, we have to convolute the word statistics with the distribution p_t leading to

$$\tilde{p}(a_{i_1}, \dots, a_{i_n} | t) = \sum_{k=-\infty}^{\infty} p_k p(a_{i_1}, \dots, a_{i_n} | t+k). \quad (18)$$

Thus, the coarse graining transforms a noisy nonlinear dynamics with uncertainty about initial conditions in state space as well as in onset time into a nonstationary symbolic stochastic process described by all n -word statistics $\tilde{p}(a_{i_1}, \dots, a_{i_n} | t)$ or in other words into an information source in the sense of information theory [62,61].

D. Measures of complexity

The most important quantities of information theory introduced by Shannon and Weaver [62] are entropy and rate of information transmission. These are called *classical* complexity measures in order to distinguish them from recent complexity measures such as machine complexity or renormalized entropy [25,27]. In the following we describe how measured ensembles of time series from a dynamical system can be transformed into a symbolic dynamics in order to estimate word statistics and to calculate entropies.

We have introduced a symbolic dynamics of a dynamical system by providing a partition of the state space X . For experimental data the state space is generally unknown and has to be reconstructed from measured time series by embedding techniques [63]. This could be done also as a first step for obtaining a symbolic dynamics of the system investigated [35]. But symbolic dynamics does not need the application of these methods, since every partition of the set of measurement values yields a partition of the state space automatically. For proof, see Appendix A 1.

Let us consider two ensembles of time series $x_i^{c_1}(t), x_j^{c_2}(t)$ of the dynamical system (X, Φ_r) obtained by a real-valued observable h by $x_i^{(c_1, c_2)}(t) = h(y_i^{(c_1, c_2)}(t))$, where i, j ($1 \leq i \leq N^{c_1}, 1 \leq j \leq N^{c_2}$) are the ensemble indices and t is the (discrete) time index. N^{c_1}, N^{c_2} are the cardinalities of both ensembles. The indices c_1 and c_2 refer to different dependences of the control parameter on time, say $r^{c_1}(t), r^{c_2}(t)$. In our notation, $x_i^{(c_1, c_2)}(t)$ means either $x_i^{c_1}(t)$ or $x_i^{c_2}(t)$. The $y_i^{(c_1, c_2)}(t) \in X$ are ensembles of trajectories in the state space for conditions c_1 and c_2 , respectively. After choosing a partition $S_i, i = 1, 2, \dots, I$ of the set $H = h(X)$ leading to a partition of X (see Appendix A 1), we decide whether the values $x_i^{(c_1, c_2)}(t)$ belong to the sets S_j in order to assign a symbol $a_{i;k}^{(c_1, c_2)}$. Thus, the ensembles of time series will be mapped onto ensembles of symbolic sequences

$$E^{c_1} = \{s_i^{c_1} | s_t^{c_1} \in A^L, 1 \leq i \leq N^{c_1}\} \quad (19)$$

$$E^{c_2} = \{s_j^{c_2} | s_j^{c_2} \in A^L, 1 \leq j \leq N^{c_2}\}, \quad (20)$$

where L is the length of the time series, i.e., the number of samples. Each string $s_i^{(c_1, c_2)}$ is a sequence $a_{i;k_1}^{(c_1, c_2)} a_{i;k_2}^{(c_1, c_2)} \dots a_{i;k_L}^{(c_1, c_2)}$ of L letters from the alphabet A . The cylinder sets (13) will be defined for measured data as $[a_{k_1}, \dots, a_{k_n}]_t^{(c_1, c_2)} = \{s \in E^{(c_1, c_2)} | s_{t+l-1} = a_{k_l}, l = 1, \dots, n\}$

and the word probabilities will be estimated by the counting measures of the ensembles E^{c_1}, E^{c_2} as the relative frequencies,

$$\begin{aligned} \bar{p}^{(c_1, c_2)}(a_{k_1}, \dots, a_{k_n} | t) &= \mu_t([a_{k_1}, \dots, a_{k_n}]_t^{(c_1, c_2)}) \\ &= \frac{|[a_{k_1}, \dots, a_{k_n}]_t^{(c_1, c_2)}|}{N^{(c_1, c_2)}}, \end{aligned} \quad (21)$$

where “ $|\cdot|$ ” denotes the set theoretic cardinality function.

Next, we shall quote the definitions of Shannon’s and Rényi’s entropies. The Shannon entropies [62] of order n at time t of the ensembles $E^{(c_1, c_2)}$ will be accommodated as

$$\begin{aligned} H_n^{(c_1, c_2)}(t) &= - \sum_{(a_{k_1}, \dots, a_{k_n})} \bar{p}^{(c_1, c_2)}(a_{k_1}, \dots, a_{k_n} | t) \\ &\quad \times \log \bar{p}^{(c_1, c_2)}(a_{k_1}, \dots, a_{k_n} | t). \end{aligned} \quad (22)$$

The quantities

$$H^{(c_1, c_2)}(t) = H_n^{(c_1, c_2)}(t)/n \quad (23)$$

measure the information per letter and are called *relative entropies*. The quantities

$$I_{n; q}^{(c_1, c_2)}(t) = \frac{1}{1-q} \log \sum_{a_{k_1}, \dots, a_{k_n}} \bar{p}^{(c_1, c_2)}(a_{k_1}, \dots, a_{k_n} | t)^q \quad (24)$$

are called n -order *Rényi entropies* depending on the parameter q [64]. The base of the logarithm in the formulas above is arbitrary. But it is recommended to use either the logarithm dualis $\text{ld} \equiv \log_2$ which measures information content in binary digits (bits) or (what we shall do) the \log_I , where I is the cardinality of the letter alphabet. This choice has the advantage that relative entropy will always be normalized to the range $[0, 1]$.

Entropy is a measure of uncertainty of a given probability distribution. It reaches its maximum value $+1$ for uniformly distributed events. It takes its minimum 0 if there is only one certain event with probability 1 . For uniform distributions all q Rényi entropies have the same value $+1$. Estimating probabilities from relative frequencies might be deceptive, especially in the case of small samples. Entropies will be systematically underestimated, if the number of possible words I^n is of the order of the ensemble size N . Several solutions to this problem have been proposed in the literature [65, 66, 34, 67]. In our analyses, this issue has only minor importance, because we need only short words ($n=1, \dots, 4$) and small alphabets ($I=2, 3$) while the grand ensembles of ERP data have cardinalities about $N=300-1000$.

Whether changes in entropy are significant or not can be tested using the method of surrogate data suggested by Theiler *et al.* [21]. Surrogate data are time series that have been generated artificially either by phase randomization of the measurements [21] or by linear stochastic models, such

as autoregressive processes [68]. They possess the same linear characteristics, such as autocorrelation function or power spectrum, as the real data and serve, therefore, as null hypotheses about the nature of the real processes. We apply this technique on the symbolic dynamics by shuffling the order of symbols within each sequence $s_i^{(c_1, c_2)}$ randomly many times. This provides realizations of Bernoulli processes. Computing and averaging the running (i.e., time-dependent) cylinder entropies for the surrogates yields an estimate of the cylinder entropies of the uncorrelated stochastic symbolic process obtained by the shuffling procedure. As a measure of significance, Theiler *et al.* give the quantity

$$S_D = \frac{Q_D - \overline{Q_H}}{s_H}, \quad (25)$$

where Q_D refers to an (arbitrarily chosen) statistic of the data. We shall use the Shannon and Rényi entropies here. Q_H stands for the same statistics estimated from the surrogates, $\overline{Q_H}$ stands for their empirical mean $(1/M) \sum_{m=1}^M Q_{H_m}$, and s_H stands for their empirical standard deviation $((1/M - 1) \sum_{m=1}^M (Q_{H_m} - \overline{Q_H})^2)^{1/2}$ for an ensemble of M surrogates [21]. The measure of significance so introduced can be seen as similar to the t -scores of the classical t -test. $S_D \geq 2$ is assumed to indicate significance about the 95% confidence level [69] provided the data are Gaussian distributed.

In order to compare the two ensembles E^{c_1}, E^{c_2} corresponding to different control parameters or experimental conditions, we apply the χ^2 -test to the word statistics $p^{c_1}(a_{k_1}, \dots, a_{k_n} | t), p^{c_2}(a_{k_1}, \dots, a_{k_n} | t)$ obtaining the χ^2 -scores and their error probabilities as functions of time according to [[70], p. 622]. Finally, we encounter the problem of latency jitter. In our theoretical framework given above and in the simulations, we introduced latency jitter as an uncertainty of observation time. The symbol sequences forming the cylinder sets will thus be shifted in time according to some discrete random variable τ_i . For strings obtained by simulated or measured time series, we use a sliding window technique collecting all word frequencies that belong to a window of width D at time t ,

$$\hat{p}^{(c_1, c_2)}(a_{k_1}, \dots, a_{k_n} | t) = \sum_{r=t-\Delta}^{t+\Delta} \frac{|[a_{k_1}, \dots, a_{k_n}]_r^{(c_1, c_2)}|}{(D-n+1)N^{(c_1, c_2)}}, \quad (26)$$

where $\Delta = (D-n+1)/2$. In the next section, we will discuss how the sliding window measures influence the entropies and the word statistics. The sliding window method is also useful in order to improve the quality of probability estimates and hence to improve the significance of the χ^2 test between conditions.

IV. SIMULATIONS

In order to validate our technique, we analyze control data given by stochastic dynamical systems with transient, non-stationary behavior, and uncertainty about initial conditions as well as about observation times. We have chosen a noisy

logistic map because its symbolic dynamics is well understood and it is easy to obtain a binary partition of its state space by the critical value $x_c=0.5$ [71,56,72,60,23]. In this section, we generate ensembles of time series from a stochastically disturbed logistic map by preparing initial conditions randomly in some interval $[0,P]$. We varied systematically the impact of noise σ_ξ and allowed the control parameter r_t to depend continuously on time ranging between fixed limits $[r,R]$. Additionally, we introduced uncertainty about the observation time by shifting the latency of the change of the control parameter t_0 randomly with variance σ_τ^2 . These free parameters of the model, $r,R,\sigma_\xi,\sigma_\tau$, might be related to the unknown parameters of ERP recordings, namely experimental conditions, the many degrees of freedom given by the spontaneous activity of the brain, and at least to latency jitter.

A. Changing control parameters

In our simulations we generate ensembles of N time series each consisting of L points in time by iterating a noisy logistic map $x_{t+1}=r_t x_t(1-x_t)+\xi_{t+1}$, where $x_t \in [0,1]$, $r_t \in [0,4]$, and ξ_t are independent identically and uniformly distributed random numbers with zero mean and standard deviation σ_ξ . An important point in our simulation is that the control parameter depends continuously on time obeying

$$r_t = \begin{cases} R: & |t-t_0| > d, \\ \frac{R-r}{2} \cos\left[\pi\left(1+\frac{t-t_0}{d}\right)\right] + \frac{R+r}{2}: & |t-t_0| \leq d, \end{cases} \quad (27)$$

where R is the maximum and r the minimum value of the control parameter r_t . t_0 is the latency time of the simulated time series and $2d$ is its duration. The latency time t_0 was allowed to jitter according to $t_0+\tau$, where τ was a uniformly distributed random number with zero mean and standard deviation σ_τ . Figure 1 shows an example of the dependency of the mean control parameter (averaged over all simulated trials) for the case where $\sigma_\tau=0$, i.e., without any latency jitter. This function causes the logistic map to become nonstationary at t_0-d until t_0+d . We used a latency $t_0=100$ and as the duration of the transient regime $2d=30$. Thus, the following figures should show some conspicuous behavior around $t=100$ where something is going on.

The ensembles have been generated by choosing N initial conditions randomly and uniformly distributed in an interval $[0,P]$, where P stands for the precision of the preparation of initial conditions. We ran simulations for $N=30$, which is a typical size of a single subject ERP ensemble, and for $N=1000$ as a typical size of a grand ERP ensemble of all subjects collected together. The sample size was always $L=300$. In the former we chose $P=0.001$ when $\sigma_\xi=0$ and $P=0.01$ when $\sigma_\xi \neq 0$. In the latter, $P=0.2$ was chosen for all values of σ_ξ .

For comparison between our symbolic dynamics technique and the traditional averaging approach of ERP, we calculated the ensemble averages of the simulated time series. The symbolic dynamics has been performed using the critical point $x_c=0.5$ of the logistic map as the dividing point

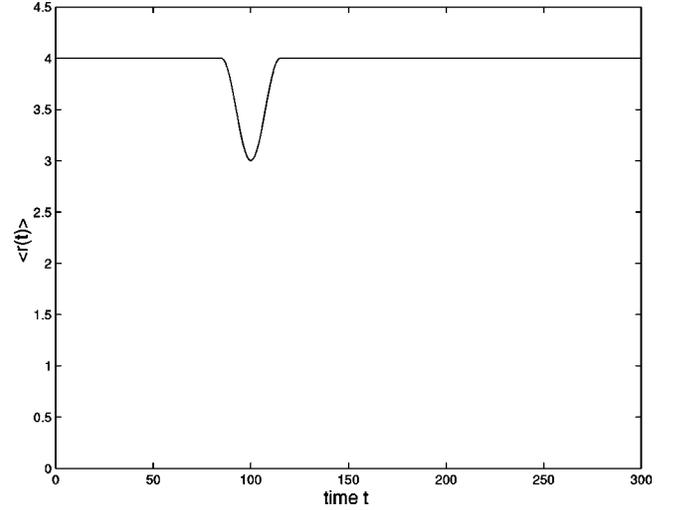


FIG. 1. Dependency of mean control parameter on time for an ensemble of logistic maps without noise ($\sigma_\xi=0,\sigma_\tau=0$) but with control parameter varying in the range $R=4.0$, $r=3.0$ [see Eq. (27)]. The nonstationarity happens around $t=100$.

for a binary partition of the unit interval $[0,1]=A_0 \cap A_1$, $A_0=[0,0.5]$, and $A_1]=[0.5,1]$ [23]. Furthermore, we created binary Bernoulli processes of the same ensemble size and the same number of samples in order to compare their entropies and word statistics with those given by the logistic map. We did three numerical experiments: (a) with different maxima and minima of the control parameter but without any noise and any latency jitter, (b) varying only the dispersion of the noise leaving the range of the control parameter and the latency invariant, and (c) investigating the impact of latency jitter on ensemble averages and symbolic dynamics using a fixed range of the control parameter and no noise of the dynamics.

In the first series of simulations, we chose the parameters of the model as $N=30$, $L=300$, $P=0.001$, $R=4.0$, $\sigma_\xi=\sigma_\tau=0$, and $r \in \{3.0,3.8,3.85,3.95\}$. The maximum $R=4$ corresponds to fully developed chaos, while for $r=3.0$ the map possesses an unstable fixed point at $x^*=\frac{2}{3}$. At values between $r_\infty \approx 3.56999\dots$ (the Feigenbaum attractor) and $R=4.0$, the map shows very complicated behavior changing from chaotic bands to periodic windows and back to chaotic bands (see, e.g., [72]). Figure 2 shows (a) the averaged time series $\langle x(t) \rangle$ and (b) the symbolic dynamics for the setting $R=4.0$, $r=3.0$, $\sigma_\xi=0$, and $\sigma_\tau=0$ with the change of the control parameter shown in Fig. 1. Black pixels denote the symbol ‘‘0’’ ($x_t \in [0,0.5]$) and white pixels denote the symbol ‘‘1’’ ($x_t \in]0.5,1]$). Both Figs. 2(a) and 2(b) show clearly the change of the dynamics around $t=100$. At the very beginning, the systems behavior is transient while it settles at the chaotic attractor for $R=4.0$. In the averaged time series (a) this nonstationarity is given by a fast, almost exponential increase of $\langle x(t) \rangle$. In the symbolic dynamics, this transient behavior is indicated by a black vertical stripe, while the states evolve in $A_0=[0,0.5]$. At time $t=85$, the control parameter starts to depend on time, and the systems become nonstationary. This is indicated by the averages beginning to oscillate until the control parameter becomes constant at $t=115$. In the symbolic dynamics, the transient regime is a

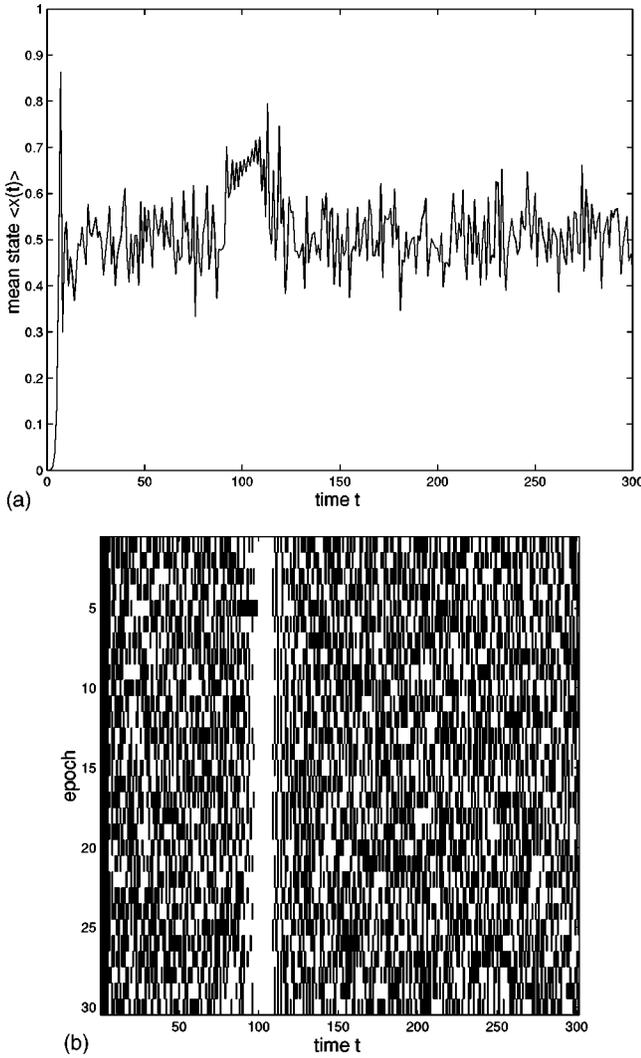


FIG. 2. Time series from the logistic map (a) Ensemble averages, (b) symbolic dynamics. Black pixel, “0” ($x_t \in [0, 0.5]$); white pixel, “1” ($x_t \in]0.5, 1]$). Parameter settings: $N=30$, $L=300$, $P=0.001$, $R=4.0$, $r=3.0$, $\sigma_\xi=0$, and $\sigma_\tau=0$. The nonstationarity happens around $t=100$.

white vertical stripe, because the fixed point $x^* = \frac{2}{3}$ belongs to $A_1 =]0.5, 1]$. Figure 2 illustrates the differences of averaging time series and symbolic dynamics. One could expect that, by lowering the amplitude of the change in control parameter $R-r$, the jump of the averaged states will decrease and eventually be hidden by the chaotic dynamics. This will not happen in the symbolic dynamics unless the fixed point leaves the interval $A_1 =]0.5, 1]$.

For the same example, Fig. 3 gives the statistics of words of length 1, i.e., the distribution of symbols over time. As in the visualization [Fig. 2(b)], the word statistics reveal clearly the phase transition from chaotic to periodic behavior at time $t=85$. The large white stripe in Fig. 2(b) corresponds to a degeneration of the word distribution where only the “word” “1” remains.

Next, we shall tune the control parameter towards the fully developed chaotic regime. We present the ensemble averages of the time series generated by the logistic map for minimal control parameter settings $r=3.8$ (Fig. 4). It is ob-

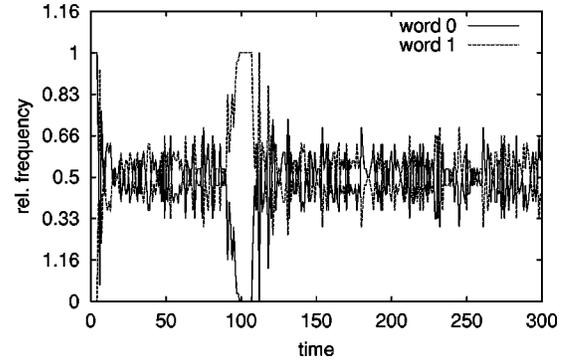


FIG. 3. One-word statistics of the ensemble of logistic maps. Parameter settings: $N=30$, $L=300$, $P=0.001$, $R=4.0$, $r=3.0$, $\sigma_\xi=0$, and $\sigma_\tau=0$. Parameters are the same as Fig. 2.

vious that for diminishing distances $R-r$ the phase transition at time $t=85$ becomes harder to recognize in the averages.

But in contrast, Fig. 5 reveals that there is a change of the dynamics visible in the running cylinder entropies [shown are Shannon entropies ($q=1$) for word length $n=2$, Fig. 5(a)]. Though also vanishing for $r \rightarrow R$, the distinction remains significant by means of surrogate data [Fig. 5(b)], where $M=1000$ surrogates have been generated. There are some further peaks of the significances at later times at 150 and 250. These could be explained by the fact that the logistic map is not fully chaotic for a parameter value $r=3.8$ and therefore significantly different from the Bernoulli process represented by the surrogates.

Using the χ^2 -test to compare the data from the logistic map with those from the Bernoulli process fails for small ensembles because the χ^2 -test demands at least five occurrences of each word. Thus, we collected words within sliding windows according to Eq. (26). This has been done for different word lengths in different windows. We present the results of the four-word statistics in sliding windows of length $D=20$ samples in Fig. 6.

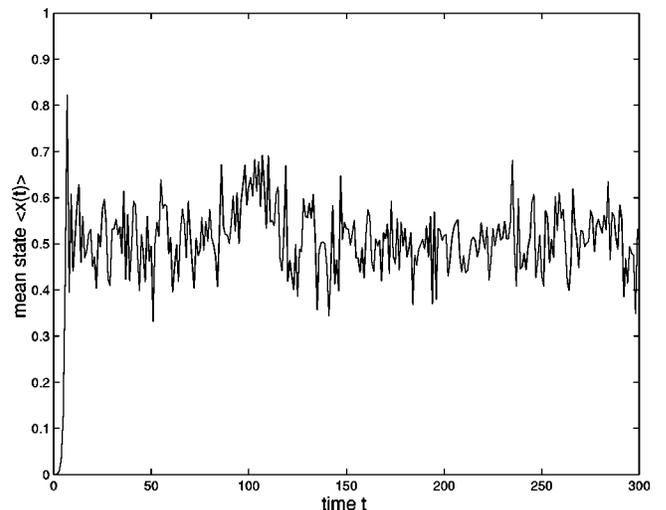


FIG. 4. Ensemble averages of time series of the logistic map for minimal value of the control parameter: $r=3.8$. $N=30$, $L=300$, $P=0.001$, $R=4.0$, $\sigma_\xi=0$, and $\sigma_\tau=0$. Again, the nonstationarity happens around $t=100$.

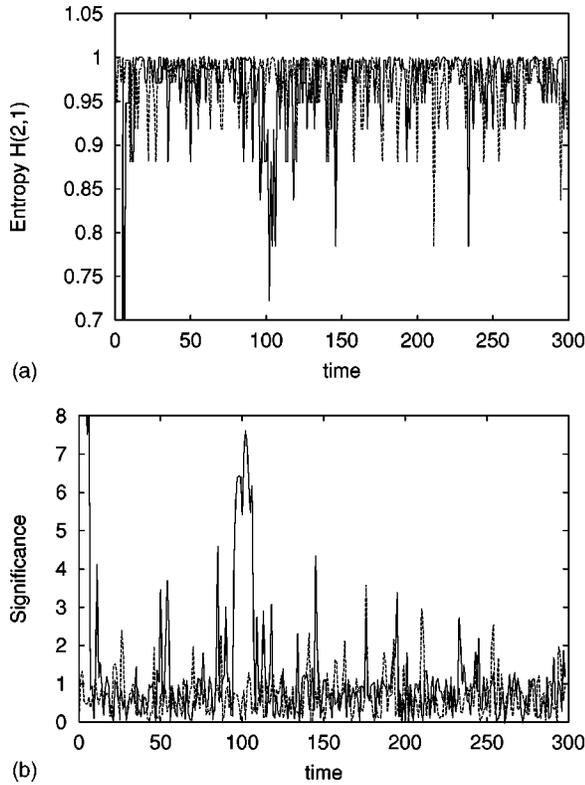


FIG. 5. Running normalized Shannon cylinder entropies [Eqs. (22) and (23)] (a) and their significances (b) from $M=1000$ surrogates [Eq. (25)] of the symbolic dynamics in comparison to a Bernoulli process. Solid, logistic map; dashed, Bernoulli process. Word length: $n=2$, $q=1.0$. (a),(b) $r=3.8$, $N=30$, $L=300$, $P=0.001$, $R=4.0$, $\sigma_\xi=0$, and $\sigma_\tau=0$. Note the nonstationarity around $t=100$.

Finally, we show the Shannon entropies obtained from these word statistics in a sliding window. Figure 7 shows the result for $r=3.8$. In contrast to the running cylinder entropies that drop in the periodic regime of the dynamics, the entropies of the sliding window cylinders increase at the beginning and at the end of the phase transition while they drop in between. This increase in entropy is explained by the fact that the dynamical behavior becomes richer when the sliding window crosses the onset of the phase transition: On the left the dynamics remain chaotic but on the right side of the window the dynamics become periodic. Between both peaks, these entropies drop when the sliding window is entirely contained within the transient regime of the dynamics. This is the case in the simulation because the sliding window length $D=20$ is smaller than the duration $2d=30$ of the nonstationarity.

The pattern shown in Fig. 7 should be the same even if there is an uncertainty in latency. This will be addressed in Sec. IV C.

B. Changing noise level

Next, we study simulations where $N=1000$ initial conditions in the range $[0,0.2]$ have been prepared. We chose $R=3.56999$, the parameter value near the Feigenbaum attractor, and $r=3.4$ in the periodic regime again. At the Feigenbaum attractor, the logistic map is very sensitive to small

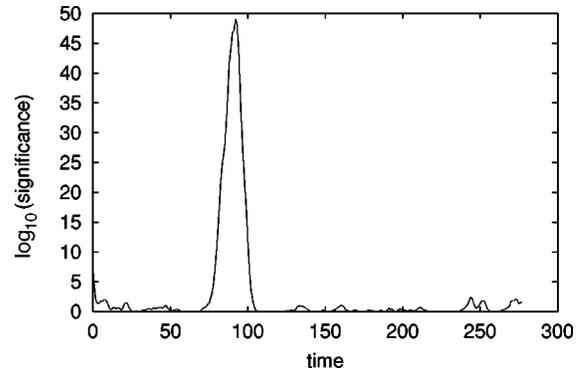


FIG. 6. Significance $[-\log_{10}(p_{\text{error}})]$ of the χ^2 -test between word statistics of symbolic dynamics of the logistic map and Bernoulli process in sliding windows. $r=3.8$, word length: $n=4$, sliding window length $D=20$, $N=30$, $L=300$, $P=0.001$, $R=4.0$, $\sigma_\xi=0$, and $\sigma_\tau=0$.

changes in control parameter and, in the presence of noise, it is also sensitive to the noise level. In the following, we increase the standard deviation of noise according to $\sigma_\xi \in \{0.0005, 0.001, 0.002, 0.0021, 0.0022, 0.00221\}$. Figure 8 shows the fourth-order Shannon entropies for the simulations with (a) $\sigma_\xi=0.0$, (b) $\sigma_\xi=0.001$, (c) $\sigma_\xi=0.002$, and (d) $\sigma_\xi=0.00221$, while Fig. 9 presents the χ^2 -significances of the four-word statistics between logistic maps for time-dependent and fixed values of the control parameters of the same ensemble size. The parameter of the control condition has been fixed to $R=r=3.56999$. While dynamical noise destroys any distinction in the averaged time series at the critical value r_∞ , the symbolic dynamics remains a reliable detector of phase transitions even in the presence of noise. The only impact of noise on the running entropies is that the differences of the dynamics before and after the nonstationary regime [oscillations at time $t=150$, Fig. 8(a)] are smeared out.

C. Changing latency jitter

Finally, we study the impact of latency jitter on symbolic dynamics of simulated data and measures of complexity. We created 30 time series per condition for the control parameter

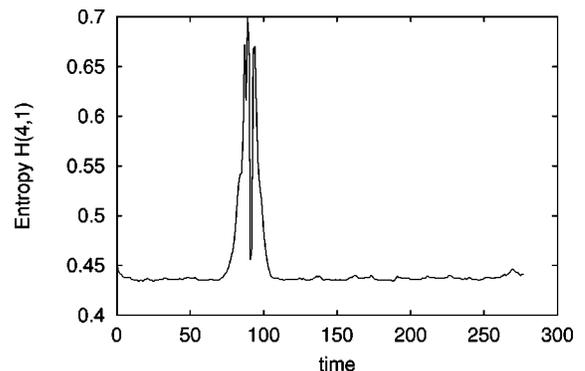


FIG. 7. Shannon entropies ($q=1$) of the word statistics of symbolic dynamics of the logistic map in sliding windows. $r=3.8$, word length: $n=4$, sliding window length $D=20$, $N=30$, $L=300$, $P=0.001$, $R=4.0$, $\sigma_\xi=0$, and $\sigma_\tau=0$.

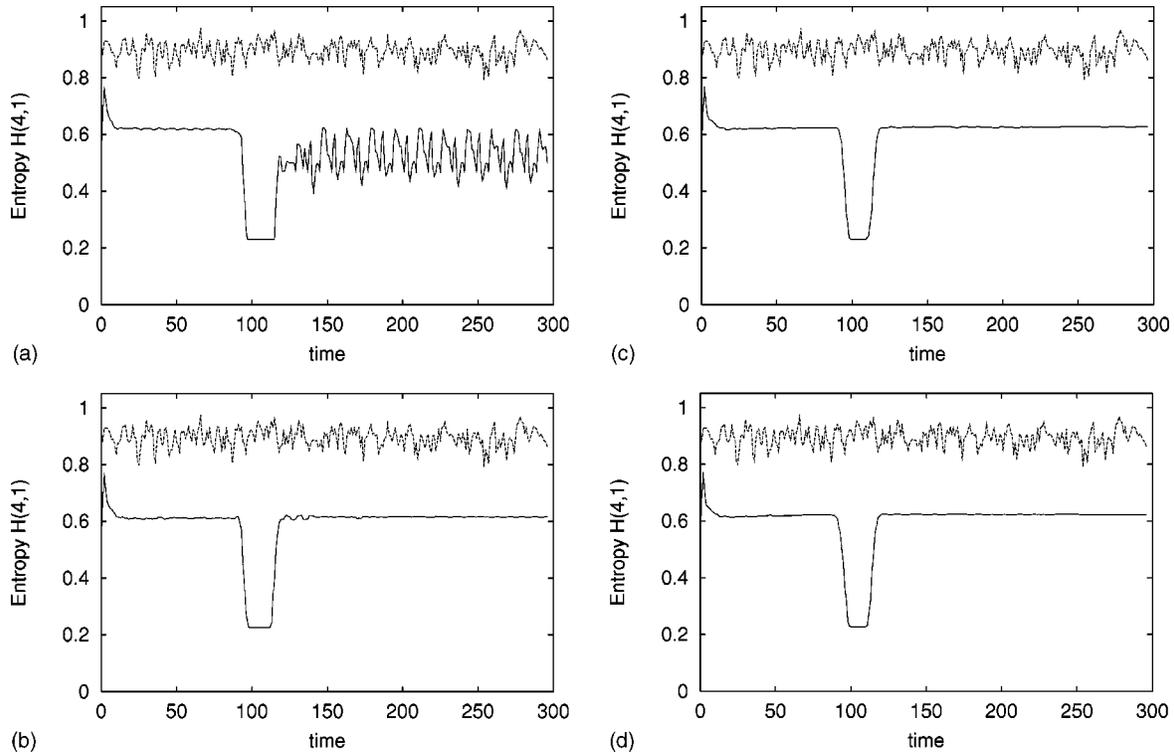


FIG. 8. Running Shannon cylinder entropies of the symbolic dynamics of the noisy logistic map for different additive noise strength in comparison to a Bernoulli process. Solid, logistic map; dashed, Bernoulli process. Word length: $n=4, q=1.0$. (a) $\sigma_\xi=0.0$, (b) $\sigma_\xi=0.001$, (c) $\sigma_\xi=0.002$, (d) $\sigma_\xi=0.002\ 21$. (a)–(d) $N=1000, L=300, P=0.02, R=3.569\ 99, r=3.4$, and $\sigma_\tau=0$.

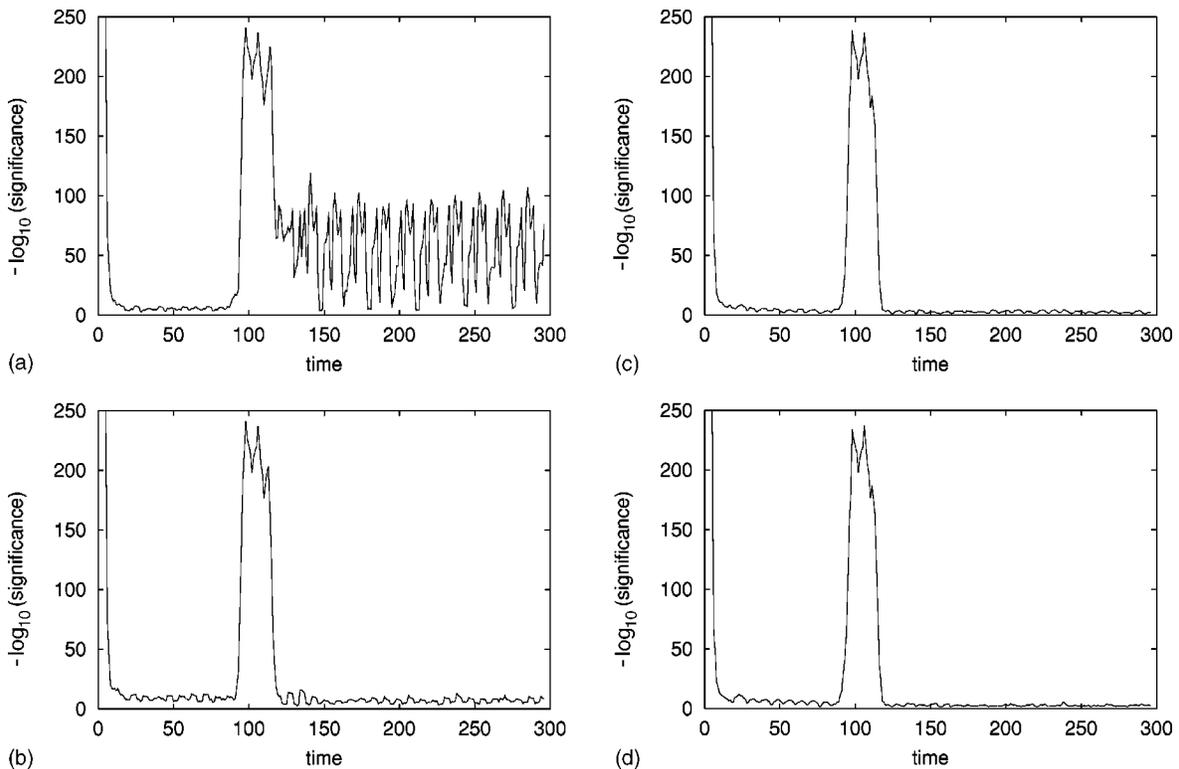


FIG. 9. Significance $[-\log_{10}(p_{\text{error}})]$ of the χ^2 test between word statistics of symbolic dynamics of the noisy logistic map for different additive noise strength against the undisturbed logistic map with constant control parameter $r=R=3.569\ 99$. Word length: $n=4$. (a) $\sigma_\xi=0.0$, (b) $\sigma_\xi=0.001$, (c) $\sigma_\xi=0.002$, (d) $\sigma_\xi=0.002\ 21$. (a)–(d) $N=1000, L=300, P=0.02, R=3.569\ 99, r=3.4, \sigma_\tau=0$.

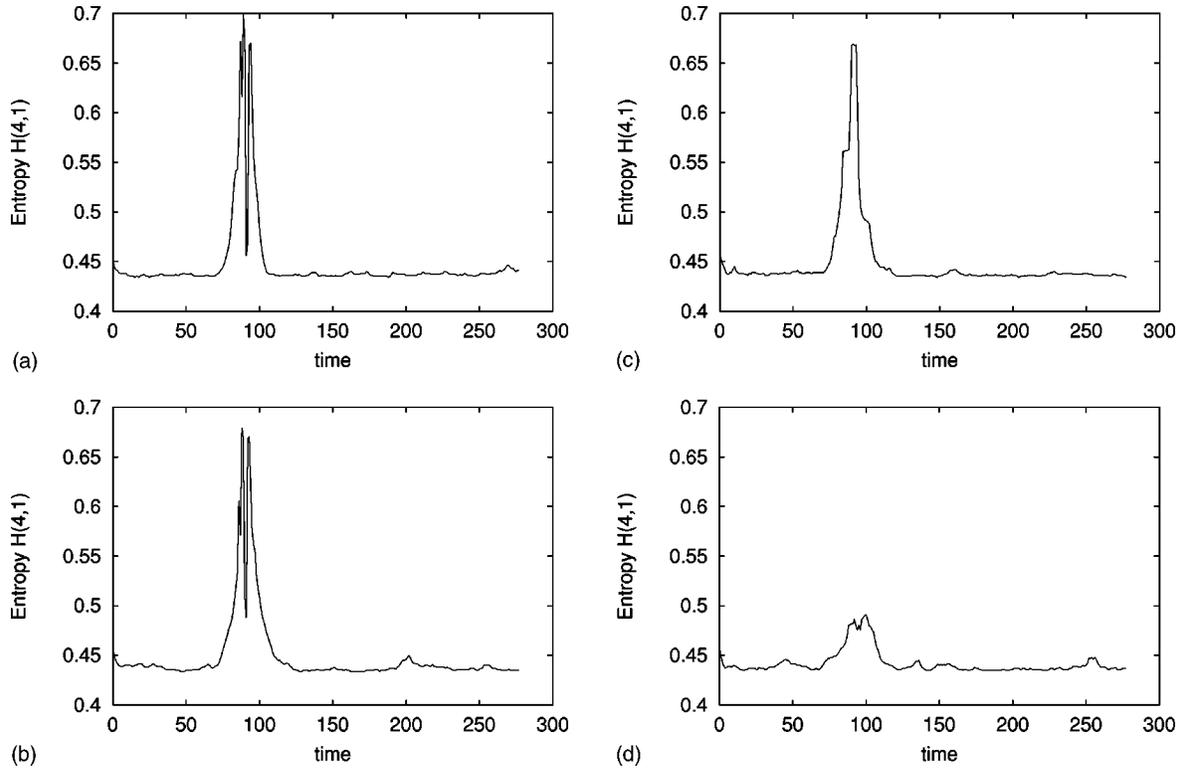


FIG. 10. Shannon entropies ($q=1$) of the four-word statistics of the symbolic dynamics of the logistic map with varying latency jitter in sliding windows. (a) $\sigma_\tau=0$, (b) $\sigma_\tau=2$, (c) $\sigma_\tau=6$, (d) $\sigma_\tau=10$. (a)–(d) Sliding window length $D=20$, $R=4.0$, $r=3.8$, $N=30$, $L=300$, $P=0.001$, $R=4.0$, and $\sigma_\xi=0$.

values $R=4.0$, $r=3.8$, and increased the standard deviation of latency jitter according to $\sigma_\tau \in \{0, 1, \dots, 15, 20\}$. First, we show the Shannon entropies of four-words collected in a sliding window of length $D=20$ (Fig. 10) for (a) $\sigma_\tau=0$, (b) $\sigma_\tau=2$, (c) $\sigma_\tau=6$, and (d) $\sigma_\tau=10$, whereas Fig. 11 presents the χ^2 significances of these four-word statistics between the logistic map and a Bernoulli process for (a) $\sigma_\tau=5$ and (b) $\sigma_\tau=10$.

The phase transition of the dynamics remains significant for small as well as for large dispersions of latency time even if there is no difference visible in the ensemble averages. For a latency jitter of $\sigma_\tau=20$ we obtained significance better than the 10^{-5} confidence level. The spurious peaks at times $t=50$ and about $t=260$ can be neglected by choosing a significance threshold smaller than 10^{-5} .

Our numerical experiments with a noisy logistic map and four freely controlled parameters — $r, R, \sigma_\xi, \sigma_\tau$ — mimicking some properties of ERP data, such as different experimental conditions, spontaneous brain activity, and latency jitter, have shown that symbolic dynamics and measure of complexity are able to distinguish transient regimes of behavior from stationary dynamics much better than averages of time series can do. Also symbolic dynamics provide many parameters such as different partitions, word lengths, sliding windows, and q values of Rényi entropies. These parameters must be appropriately chosen in order to obtain the best result of data analysis. We varied these parameters quite arbitrarily to illustrate the dependence of results from these choices. For example, Figs. 5 and 8 demonstrate the impact of the word length on running entropies and it seems to be the case that longer words will provide better

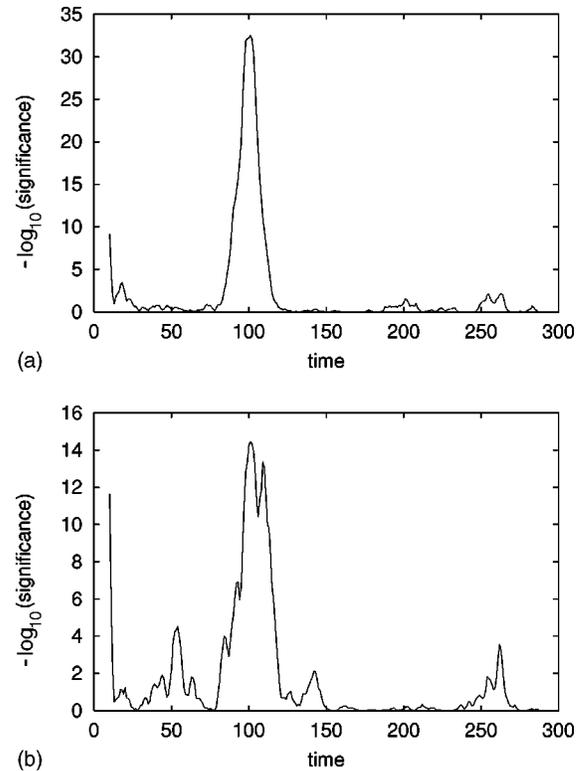


FIG. 11. Significance $[-\log_{10}(p_{\text{error}})]$ of the χ^2 -test between word statistics of the symbolic dynamics of the logistic map with varying latency jitter and Bernoulli process in sliding windows. (a) $\sigma_\tau=5$, (b) $\sigma_\tau=10$. (a),(b) Sliding window length $D=20$, $R=4.0$, $r=3.8$, $N=30$, $L=300$, $P=0.001$, $R=4.0$, and $\sigma_\xi=0$.

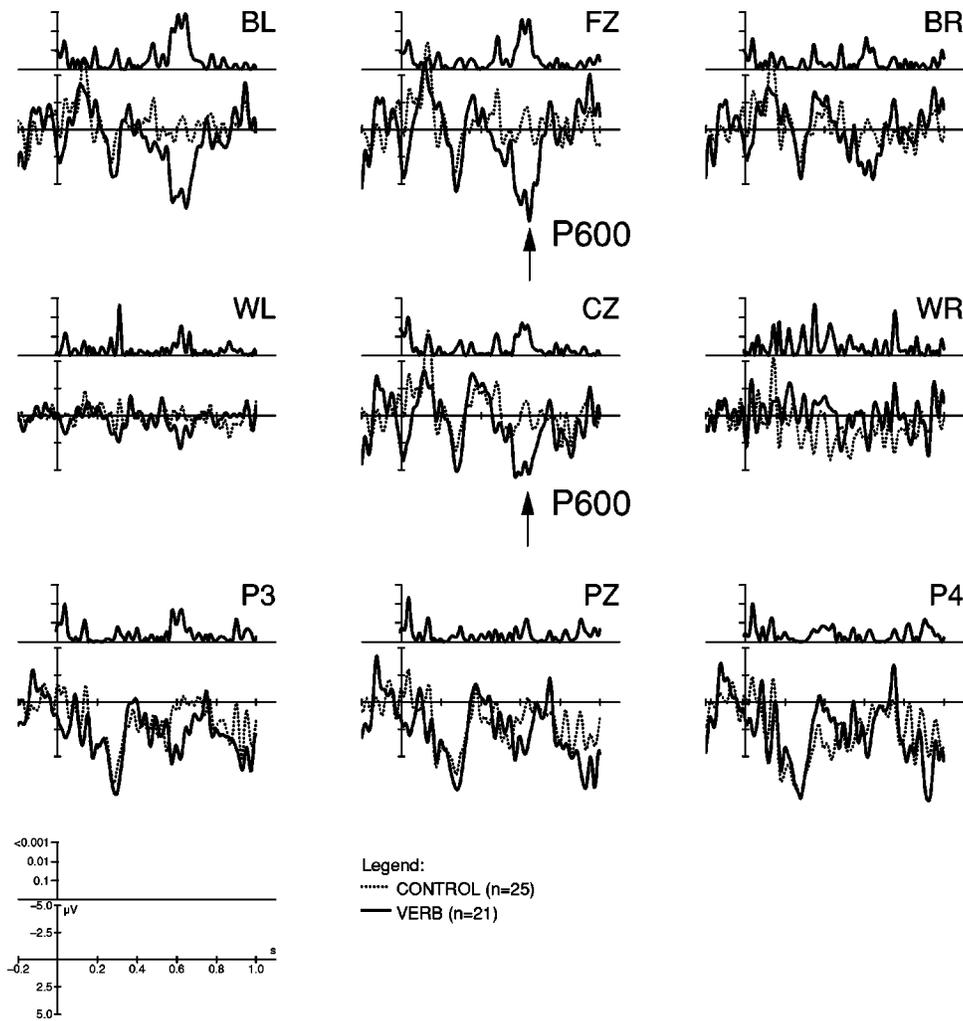


FIG. 12. Single subject voltage averages for the VERB condition (solid lines) compared to the CONTROL-VERB condition (dotted). Lower panels: potential (μV) against time (s), $t=0$: stimulus onset time. Upper panels: t -test error probabilities (logarithmic scale). The ungrammaticality effect, P600 (positive peak after 600 ms), is marked by arrows. For channel labeling see text.

results than shorter words. But for strong noise or small ensemble sizes the opposite holds, as we saw for data obtained from different ERP experiments. However, we have shown that nonstationary dynamics can be detected from ensembles of coarse-grained time series even if there are very small changes in the control parameters and the system is sensitive against changes in initial conditions as well as changes in control parameter. This applies also in the presence of dynamical (and not only observational) noise. In addition, by using sliding windows, our method is able to track uncertainty about the time when changes of the control parameters happen.

V. ANALYSIS OF EXPERIMENTAL DATA

A. Setup

In this section we shall compare the traditional ERP averaging technique with our methods based on symbolic dynamics for ERP data acquired in the Potsdam Language Processing Laboratory. We performed a language-processing experiment in which subjects were seated in front of a monitor for stimulus presentation. They were wired with the EEG amplifier device by a 32-channel electrocap. The EEG data have been recorded in 25 channels according to the international 10-20 system at a sampling rate of 250 Hz with the ScanAmps/NeuroScan recording system against reference at the left mastoid bone and grounded by the electrode Fpz.

Four channels were used for measuring EOG artifacts. ERP data have been digitally band-pass-filtered at 0.2–29 Hz after manual artifact rejection. 150 sentences, 30 sentences for each condition, organized in five blocks, were visually presented to subjects word by word but in randomized order with a presentation time and interstimulus interval of 400 ms. At the end of each sentence, subjects had to indicate whether or not a target word had occurred in the preceding sentence. The experiment was performed in two runs, the first without and the second including distractor sentences in order to mislead subjects about the purpose of the study. We report here the data analysis of 16 subjects (15 female) aged 19–25, mean age 21.78 years. The subjects were volunteers, students of the University of Potsdam, native speakers of German but not familiar with the purpose of the study. The sentences presented to the subjects were initially ambiguous interrogative sentences. The point of disambiguation (i.e., where the meaning becomes clear) was either at the verb or at the second article. For details, see Appendix A 2 or [73]; an introduction into language-related brain activity can be found in [48,4,49].

B. Single subject analysis

First, we discuss the results of traditional averaging analysis and symbolic dynamics obtained from the data of a single subject. Digitally band pass filtering, epoching, averaging,

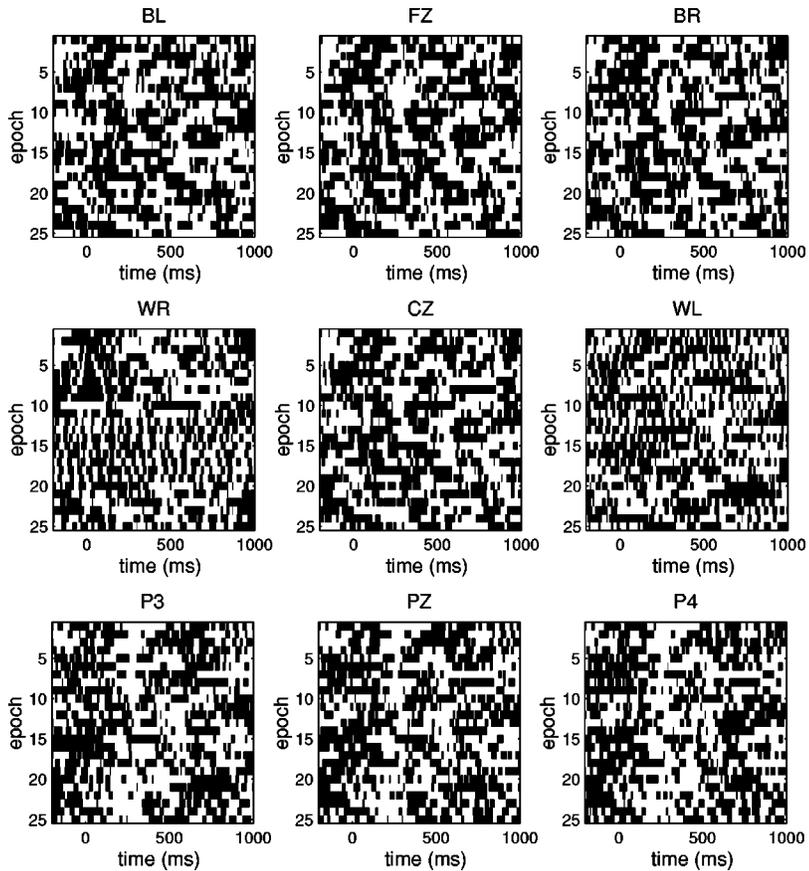


FIG. 13. Single subject symbolic dynamics of the CONTROL-VERB condition. Statically encoded after transform to a uniform distribution. Black pixel, “0” ($r_t \in [0, 0.5]$); white pixel, “1” ($r_t \in]0.5, 1]$).

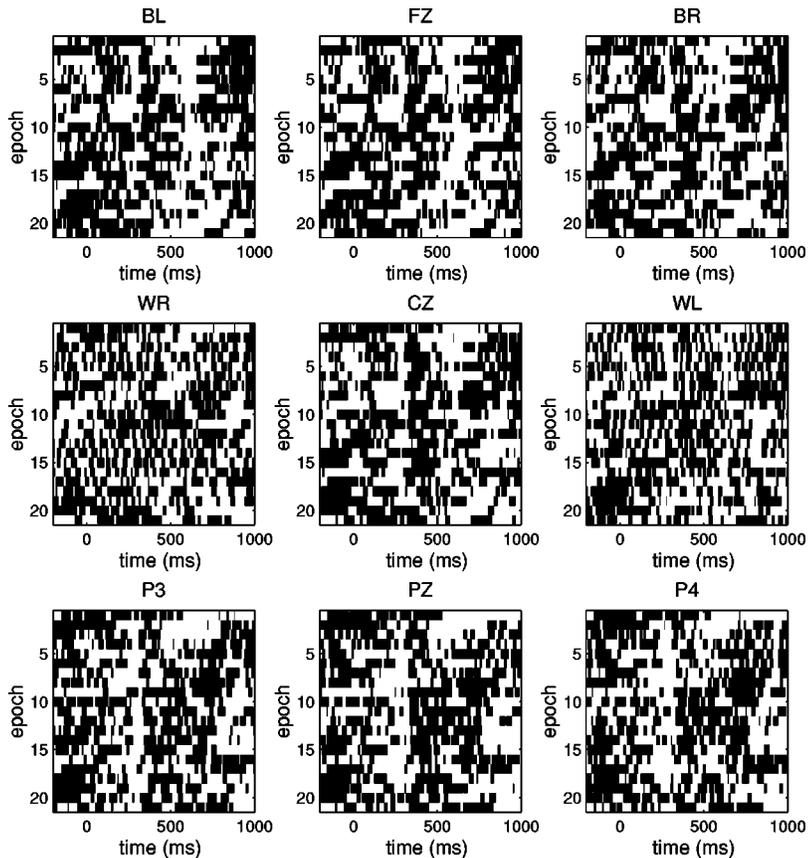


FIG. 14. Single subject symbolic dynamics of the VERB condition. Statically encoded after transform to a uniform distribution. Black pixel, “0” ($r_t \in [0, 0.5]$); white pixel, “1” ($r_t \in]0.5, 1]$).

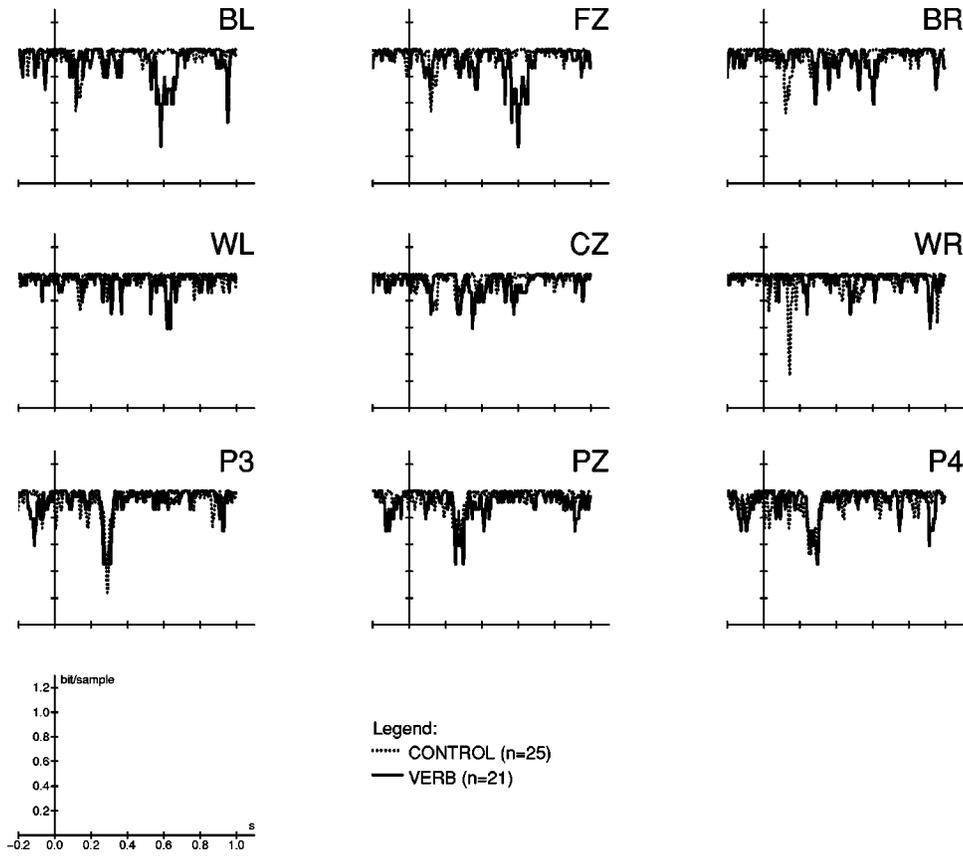


FIG. 15. Running Shannon entropies for the VERB condition (solid lines) compared to the CONTROL-VERB condition (dotted) for single subject data. Panels: normalized Shannon entropy of one word ($n=1; q=1.0$) in bit/sample against time (s), $t=0$: stimulus onset time.

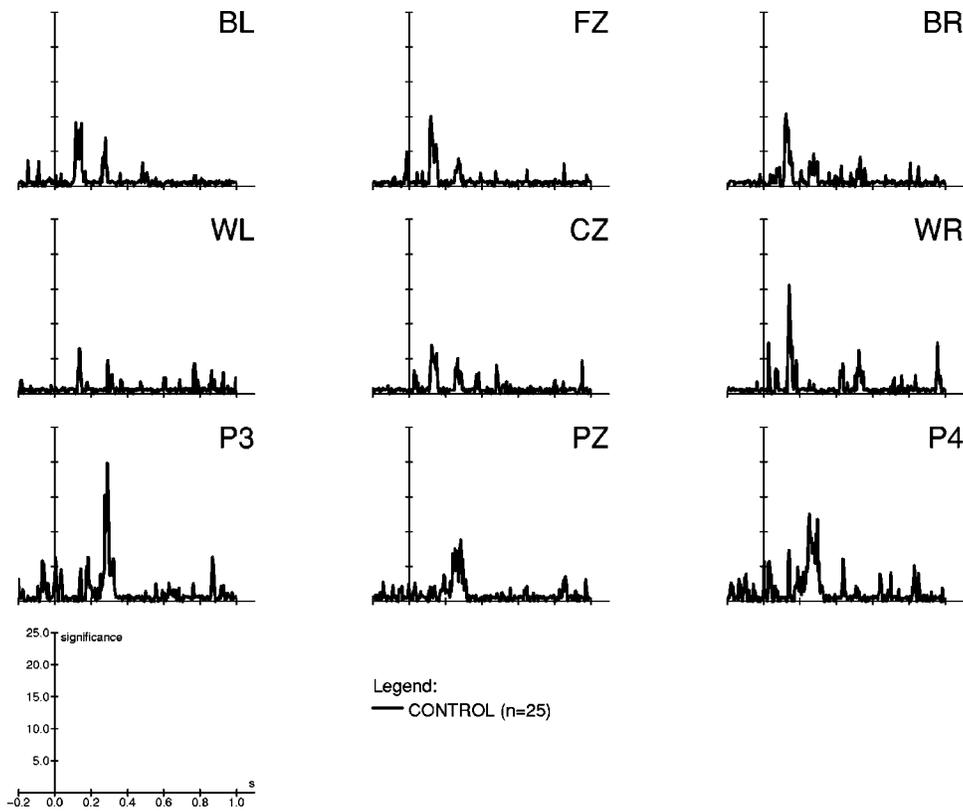


FIG. 16. Significance of the Shannon entropy for the CONTROL-VERB condition of single subject data by means of $M=100$ surrogates simulating a Bernoulli process. Panels: Significance measure S_D (unit σ) against time (s), $t=0$: stimulus onset time.

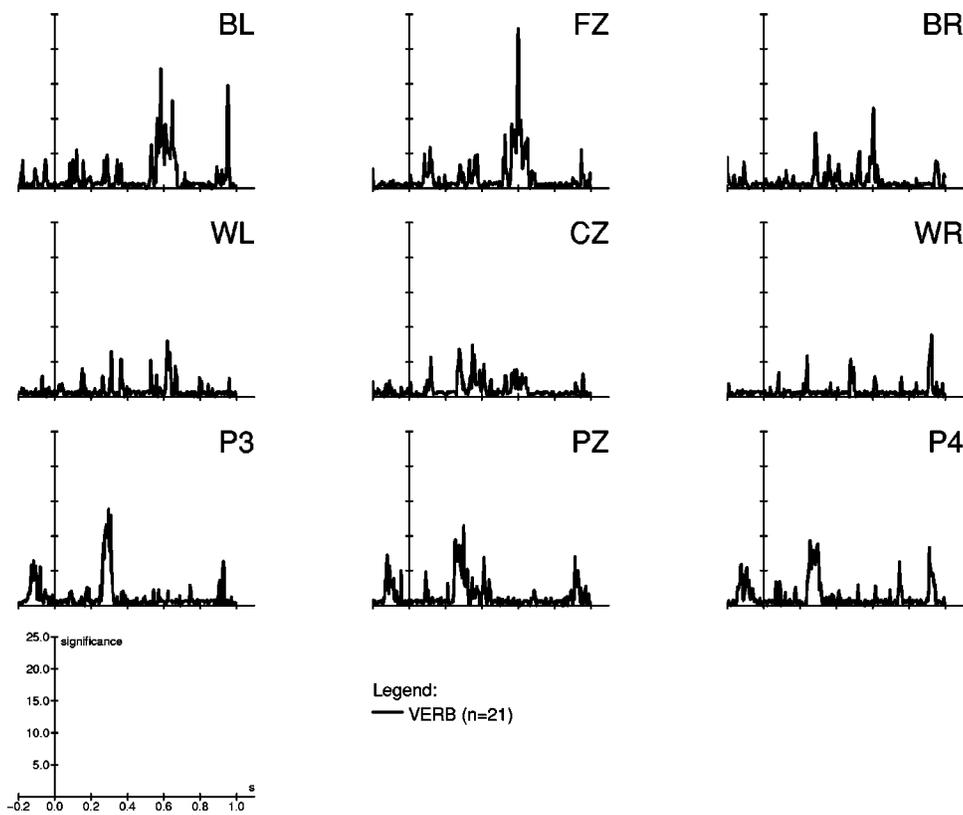


FIG. 17. Significance of the Shannon entropy for the VERB condition of single subject data by means of $M=100$ surrogates simulating a Bernoulli process. Panels: Significance measure S_D (unit σ) against time (s), $t=0$: stimulus onset time.

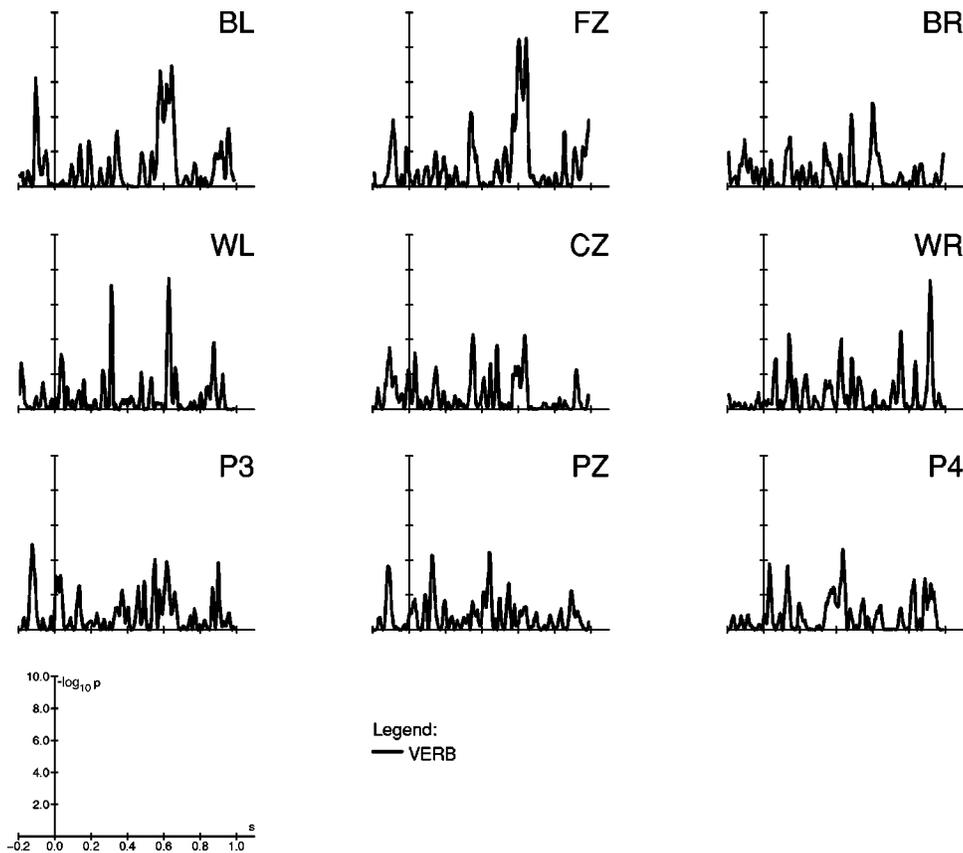


FIG. 18. Significance $[-\log_{10}(p_{error})]$ of the χ^2 -test between one-word statistics of the VERB condition against the CONTROL-VERB condition of single subject data in sliding windows of length $D=20$ ms.

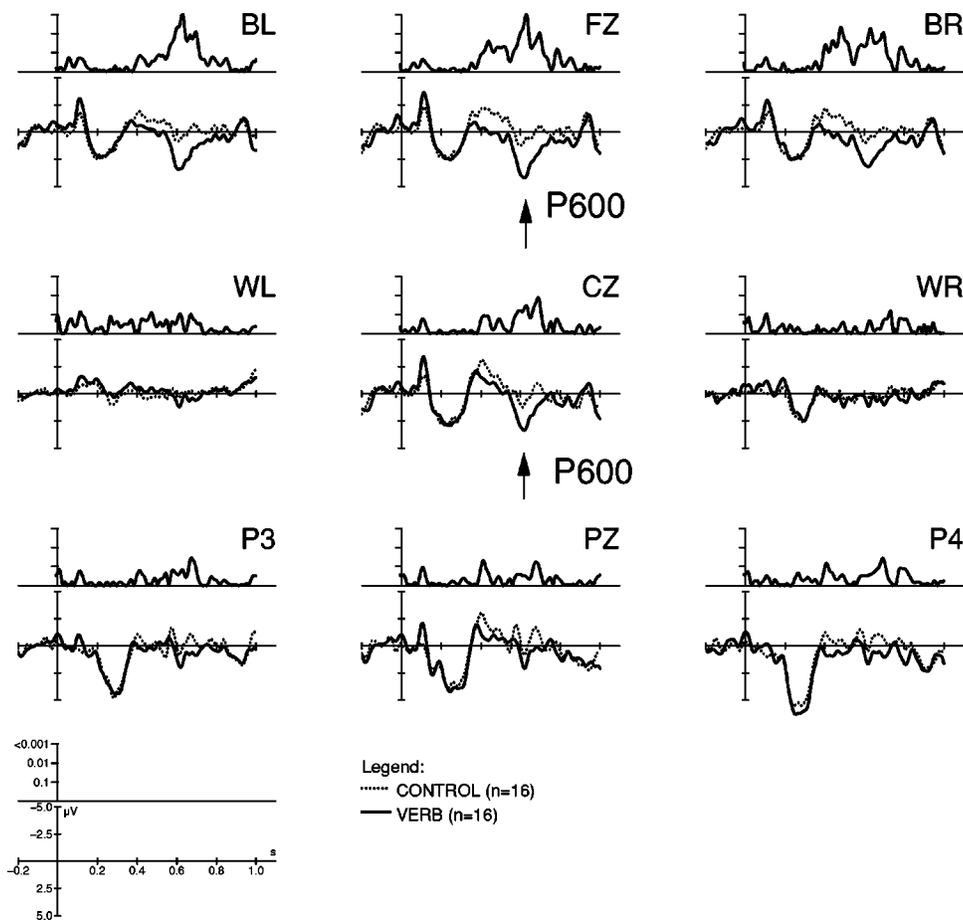


FIG. 19. Voltage grand averages for the VERB condition (solid lines) compared to the CONTROL-VERB condition (dotted). Lower panels: potential (μV) against time (s), $t=0$: stimulus onset time. Upper panels: t -test error probabilities (logarithmic scale). The ungrammaticality effect, P600 (positive peak after 600 ms), is marked by arrows.

and visualization of ERP data have been performed using the EEP (event related potential evaluation package) software developed by the Max Planck Institute for Cognitive Neuroscience [74]. We present first the voltage averages of the VERB condition in comparison to its CONTROL-VERB condition. Figure 12 shows a 3×3 -channel subarray consisting of the electrodes BL, Fz, and BR in the first row; WL, Cz, and WR in the second row; and P3, Pz, and P4 in the last. The labels Fz, Cz, P3, Pz, and P4 are due to the naming convention of the international 10-20 system, where the first symbol denotes the cortical lobes (F , frontal; T , temporal; O , occipital; P , parietal) or the vertex (C , central). The second symbol is either “z” (midline) or a number. Odd numbers correspond to the left hemisphere and even numbers to the right. The labels BL, BR and WL, WR denote language-specific cortical areas (B , Broca’s; W , Wernicke’s areas) at the left (L) or right (R) hemisphere. In the electrode array shown by Fig. 12, two diagrams are assigned to each channel: At the bottom, the time course of the averaged voltages is shown [86]. The solid line represents the experimental condition (VERB) while the dotted line indicates the signal of the control condition (CONTROL-VERB). The latter has been gained by averaging over 21 trial epochs, the former by averaging over 25 trials for the subject chosen here. At the top we show the confidence levels of a running t -test performed pointwise at each sampling point between both conditions. Significant time regions are indicated by error probabilities above the $\alpha=0.01$ (99%) threshold. Obviously, the VERB condition differs significantly from the CONTROL-VERB condition by a huge positive voltage deflection about 600 ms after verb pre-

sentation. This is known as the P600 ERP component (positivity after 600 ms) and has been related to reanalysis processes achieved by the human language-processing system dealing with local ungrammaticalities [48].

Next, we apply our newly developed techniques based on symbolic dynamics on these data. Figures 13 and 14 present the corresponding single subject symbolic dynamics of the CONTROL-VERB and the VERB condition, respectively. According to our discussion in Sec. III D, a partition of the state space or a corresponding partition of the range of the observables that are generated by the dynamical system is needed in order to obtain a symbolic dynamics. Either one encodes the values of time series into symbols by a certain binning (this method is called *static encoding*), or one encodes differences between succeeding samples as estimators of local slopes (this is called *dynamic encoding*) [25]. We decided to use static encoding after we had transformed each voltage epoch to a uniformly distributed time series by a ranking procedure. Ranking, an additional step of preprocessing, has several advantages in signal analysis because it corresponds to a nonlinear but smooth and invertible mapping of data x_t due to their distribution function F . Relative rank numbers $r_t = R_t/L$ depending on time are then obtained by the transform $r_t = F(x_t)$. Since the map F is a homeomorphism, ranking does not disturb the topological properties of state space and thus leaves basic quantities such as Lyapunov exponents or Kolmogorov entropies invariant [75]. A numerical problem linked to ranking methods arises when some values occur several times in the data set. Using sorting algorithms leads to consecutive rank numbers for the same value of

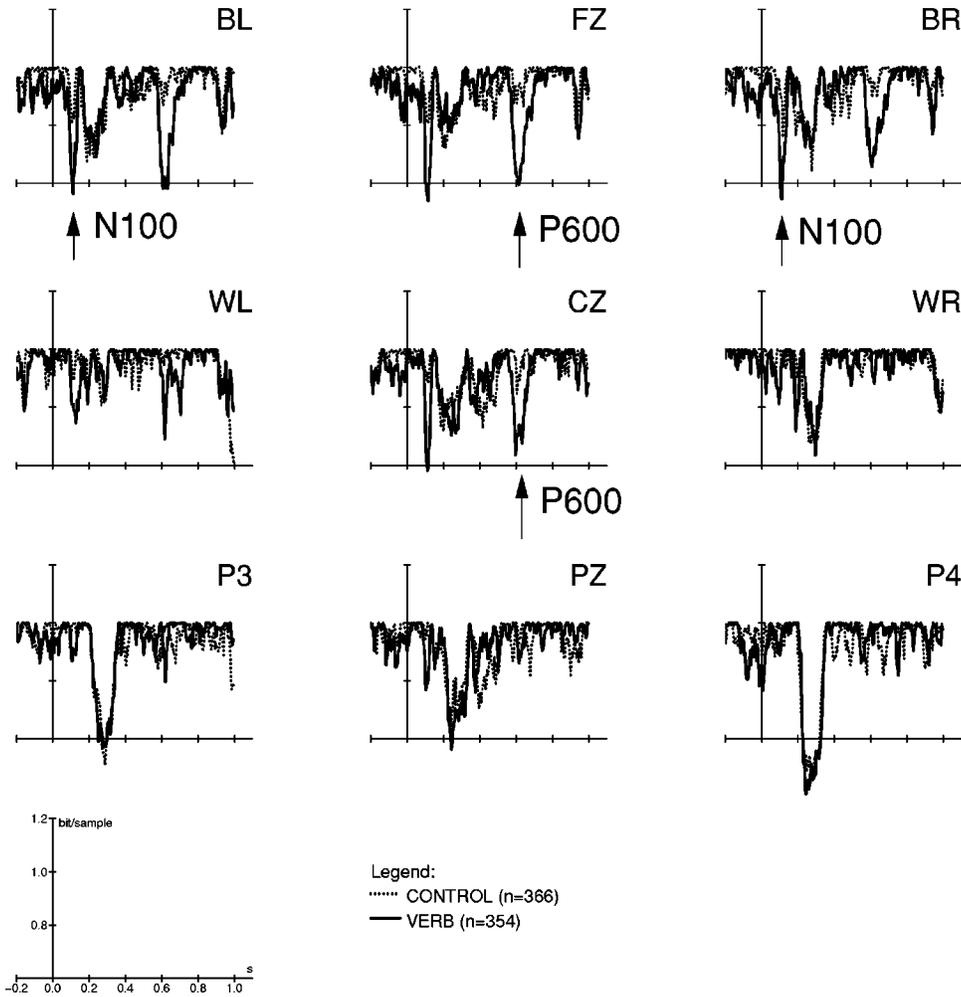


FIG. 20. Running normalized Rényi entropies for the VERB condition (solid lines) compared to the CONTROL-VERB condition (dotted) of the grand ensemble. Panels: Rényi entropy of one word ($n=1; q=10.0$) in bit/sample against time (s), $t=0$: stimulus onset time. The ungrammaticality effect, P600 (positive peak after 600 ms), is marked by arrows. Additionally, the attention shift indicator (N100) is marked by arrows.

data. Pompe recommends solving this problem of “tied ranks” by replacing the measured values x_i by some disturbed values $x_i + a_i$ before doing the ranking, where $0 \leq a_i \leq a_{i+1} \leq 1$ ([75], p. 85). We used white noise of different variances added to the voltages to tackle this problem, but we found no significant differences compared to tied ranks which have been admitted.

Performing a binary static encoding after the ranking procedure according to the median [$F(x_{\text{med}}) = 0.5$] of each epoch separately as the decision point yields a symbolic sequence containing the same number of “0” ($r_i \leq 0.5$) and “1” ($r_i > 0.5$). Thus, ranking maximizes the entropy of each sequence. However, the visualized data in Figs. 13 and 14 show distinct patterns of more or less vertical stripes. The P600 ERP component is again clearly visible as a wide white band occurring mainly at 600 ms in the channels BL, Fz, and Cz (Fig. 14). The static encoding strategy provides, therefore, an interface between ERP voltage averages and symbolic dynamics techniques that allows for comparisons between both methods. Simple static encodings by assigning, e.g., “0” to negative and “1” to positive voltage deflections have already been proposed in the early 1970s by Lehmann [76]. Counting symbol occurrences from all epochs at a certain point in time [in our terms one-word statistics, see Eq. (21)] is well known as polarity histograms in the ERP literature [41]. However, simple static encoding of voltages requires some baseline correction.

Next, we are going to quantify the patterns visible in Figs. 13 and 14. Our concepts of running cylinder entropies [Eqs. (22)–(24)] are appropriate tools for quantifying vertical stripes because every vertical band of the symbolic dynamics can be considered as a family of cylinder sets. Figure 15 presents the Shannon entropies calculated from these running cylinders of length $n=1$ for the VERB condition (solid line) and the CONTROL-VERB condition (dotted line), respectively. In order to determine whether these changes in entropy are significant or not, we computed the average Shannon entropies from an ensemble of $M=100$ surrogate data according to Eq. (25) generated by shuffling the symbols of each epoch in time. Figures 16 and 17 show these results for the CONTROL-VERB condition and the VERB condition, respectively.

Figures 15 and 17 demonstrate a highly significant entropy drop in the VERB condition around 600 ms after stimulus presentation, whereas the entropy curve remains flat in the CONTROL-VERB condition in this time domain. There are also some significant events in both conditions in channels BL, BR, and WR at time 100–200 ms as well as in channels P3 and P4 in the 300-ms time period shown in Figs. 15, 16, and 17. The early event is the N100 attention shift mentioned in Sec. II. The electrodes P3 and P4 are situated near the primary visual cortex, so they indicate optical information processing at time 300 ms.

Cylinder measures are the appropriate tool for studying patterns such as vertical stripes. But as mentioned above, the

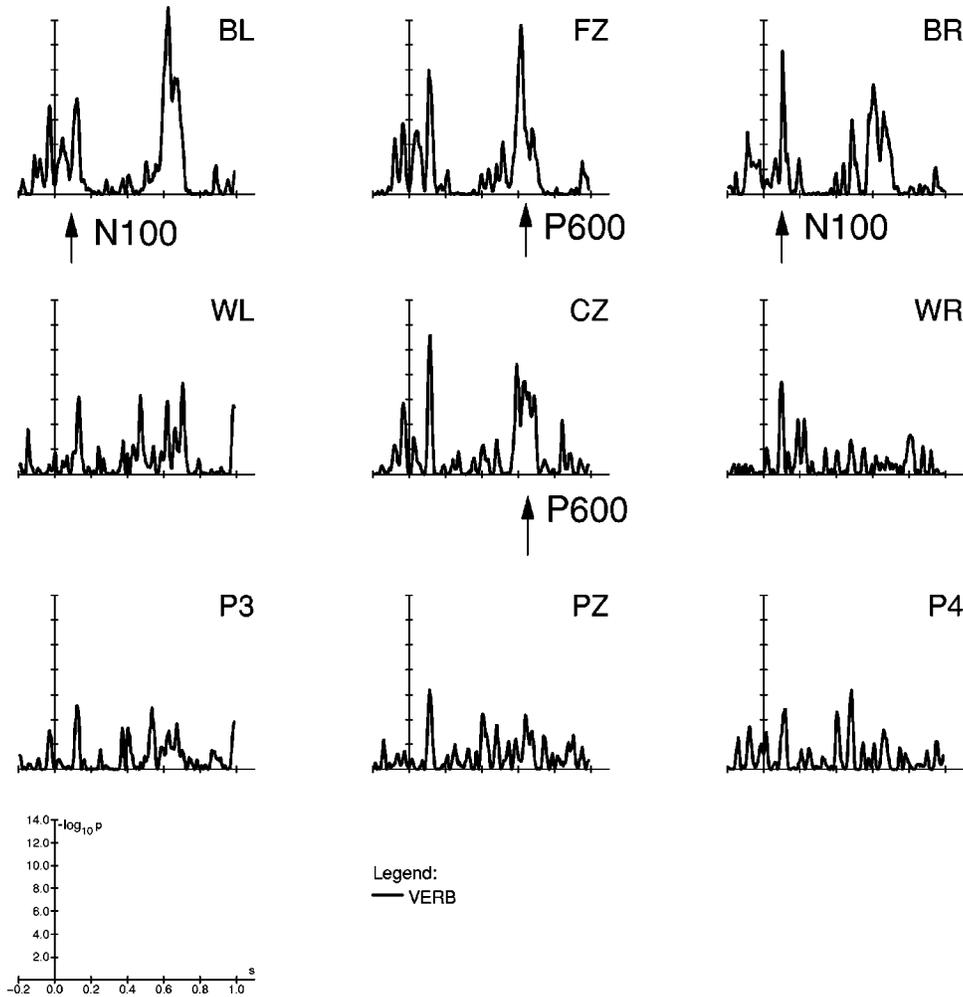


FIG. 21. Significance $[-\log_{10}(p_{\text{error}})]$ of the χ^2 test between one-word statistics of the VERB condition against the CONTROL-VERB condition of the grand ensemble in sliding windows of length $D=20$ ms. Again, P600 and N100 are marked by arrows.

symbolic dynamics of ERP are meandering bands rather than vertical stripes due to the latency jitter. Thus, we need some measures of “stripiness.” These are provided by word statistics collected in sliding windows and their corresponding entropies given by Eq. (26). For single subject investigations, collecting words in sliding windows is also necessary in order to fulfill the requirements of the χ^2 -test, where at least five events should be counted into one bin. Finally, Fig. 18 shows the result of the χ^2 -test comparing the VERB condition against the control condition by their one-word statistics in a sliding window of 20 ms. The level of confidence for the P600 at electrode Fz is better than 10^{-8} .

C. Grand ensemble analysis

At the end of this study we compare the results of the grand average analysis and the grand ensemble analysis of the symbolic dynamics gained from all 16 subjects together. Figure 19 gives the voltage grand averages of the VERB condition (solid lines) and the CONTROL-VERB condition (dotted lines). Only the P600 ERP component appears to be highly significant by means of the running t test.

Figure 20 presents the Rényi entropy comparison for word length $n=1$ and $q=10.0$. The VERB condition is plotted solid and the CONTROL-VERB condition dotted, again. This entropy plot shows not only the P600 component but also an early entropy drop symmetrically distributed at the frontal sites BL, Fz, Cz, and BR. This is the N100 attention

shift ERP marker we mentioned before. In contrast to the voltage averages, this event becomes significant in the symbolic dynamics between both conditions.

Figure 21 shows the corresponding χ^2 -error probabilities of one-word statistics collected in 20-ms sliding windows. It comes out that both the P600 as well as the very early entropy drop are significantly better than the 10^{-8} confidence level.

So far we have been only interested in the dynamics at single sites. The behavior of ERP over the whole scalp can be shown using brain-mapping techniques. We shall present brain maps of Rényi entropy ($n=1, q=10$) differences between the VERB and the CONTROL-VERB condition at single instances of time and their corresponding significance maps obtained by the χ^2 -test of one-word statistics collected in 20-ms sliding windows. Figure 22(a) shows the P600 at time 616 ms as a left frontal distributed blacking (the VERB condition has been subtracted from the CONTROL-VERB condition; hence, the entropy drop is shown black). Figure 22(b) shows the significance map of the windowed χ^2 test. Figure 23 gives a legend of the brain map.

Let us take a look at the other experimental conditions that have been tested in the language-processing ERP experiment. Figure 24 shows the voltage grand averages for the ARTICLE (solid line) against the CONTROL-ARTICLE (dotted) condition in a 3×3 array. By means of the running t -test (upper panel) there are no later events that turn out to be

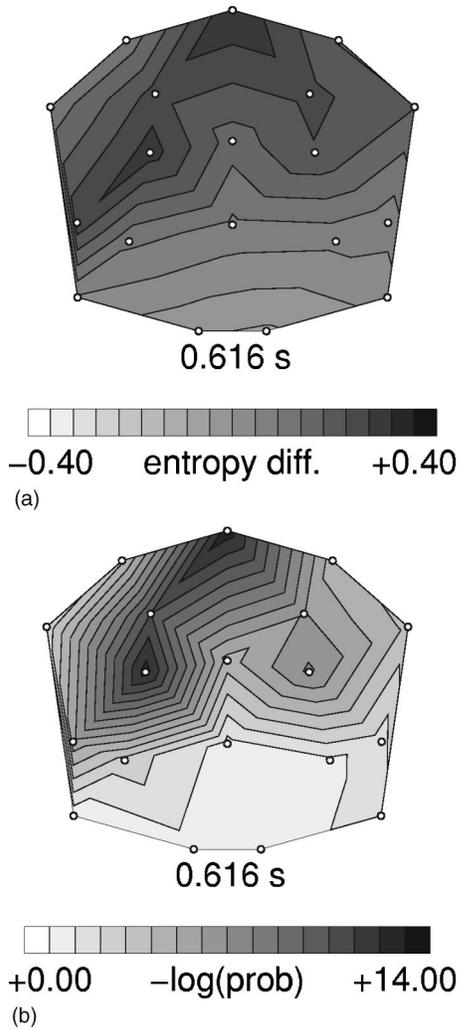


FIG. 22. (a) Rényi entropy difference map of the VERB condition subtracted from CONTROL-VERB condition of the P600 ERP of the grand ensemble ($n=1; q=10.0$). (b) Significance map $[-\log_{10}(p_{\text{error}})]$ of the windowed χ^2 -test of one-word statistics in a sliding 20-ms window.

significant at $p < 0.01$. Both curves are close together at almost all times. Note, there is a very late effect at 800 ms in channels Fz and Cz where the voltage averages of the ARTICLE condition vanish. In the traditional ERP terminology, there is no ERP component, because the voltage average curve does not possess a peak in this time range.

We present brain maps of Rényi entropy differences ($n=1, q=10$) of the ARTICLE and the CONTROL-ARTICLE condition and their corresponding significance map at two instances of time. Figure 25 shows these maps at time 210 ms after stimulus onset. Figure 25(a) reveals an increase in entropy at posterior sites WL and WR represented by white areas, which becomes more significant at WL [Fig. 25(b)]. The late effect is given at time 830 ms in Figs. 26(a) and 26(b). The vanishing voltage average corresponds to an entropy increase distributed at the left anterior areas (electrodes F3 and BL). Thus, we have proven that there is more structure in ERP data than could be discovered by averaging voltages. Our techniques resting on symbolic dynamics and measures of complexity provide access to these hidden structures in the data.

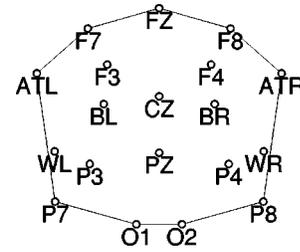


FIG. 23. Map legend of the electrode array.

VI. DISCUSSION

In this study we have criticized the prerequisites of the traditional ERP averaging technique based on stationary and ergodicity properties and intertrial independence of background EEG as well as on intertrial invariance of the assumed ERP signal. Due to these unrealistic assumptions, we have proposed an alternative approach of ERP time series analysis resting on statistical mechanics of dynamical systems and mainly on symbolic dynamics and measures of complexity. We have validated this ansatz by simulations of transient noisy nonlinear dynamics and applied it to real ERP data given by a language-processing ERP experiment. Our method reproduces well-known ERP components such as the P600 of syntactic reanalysis as highly significant entropy drops. We tested data for significance using either the method of surrogates in order to distinguish patterns in the symbolic dynamics indicated by changes in entropies from Bernoulli processes, or by means of a nonparametric χ^2 test. Thus, we should not assume that data were Gaussian distributed or had stationary variances as is necessary for applying t tests or analyses of variance. Employing a χ^2 test requires statistical independence of trial epochs, of course. But this can be approximately fulfilled by the experimental design, e.g., by randomization of stimuli concerning experimental conditions or by using distractor sentences.

We not only regained the well-known ERP component P600, but we also found a very early entropy drop (corresponding to a modulation of the attention shift N100 ERP) that is highly significant in our analysis but not in the comparison of mean voltage differences. Thus, our method is more sensitive to qualitative changes in the ongoing EEG than the averaging technique. While classical ERP components are reflected by drops in entropy of the experimental condition relative to the control condition, the second part of the study shows increasing relative entropies that also become significant. There are no counterparts to these events in the traditional ERP paradigm at all.

These findings together with the theoretical basis of our method suggest viewing ERP components as information sources of a given entropy rate [62] rather than a particular signal hidden by some noisy background that must be extracted from the data. Since symbols and words formed by strings of symbols of a symbolic dynamics correspond to regions of the phase space, periods of time of low cylinder entropy can be seen as bottlenecks in the state space in analogy to channeling in the case of intermittency. On the other hand, periods of time of increased cylinder entropy are periods where a greater volume of phase space is available to the dynamics. This interpretation agrees further with the results reported in [22] concerning the dynamics of pointwise di-

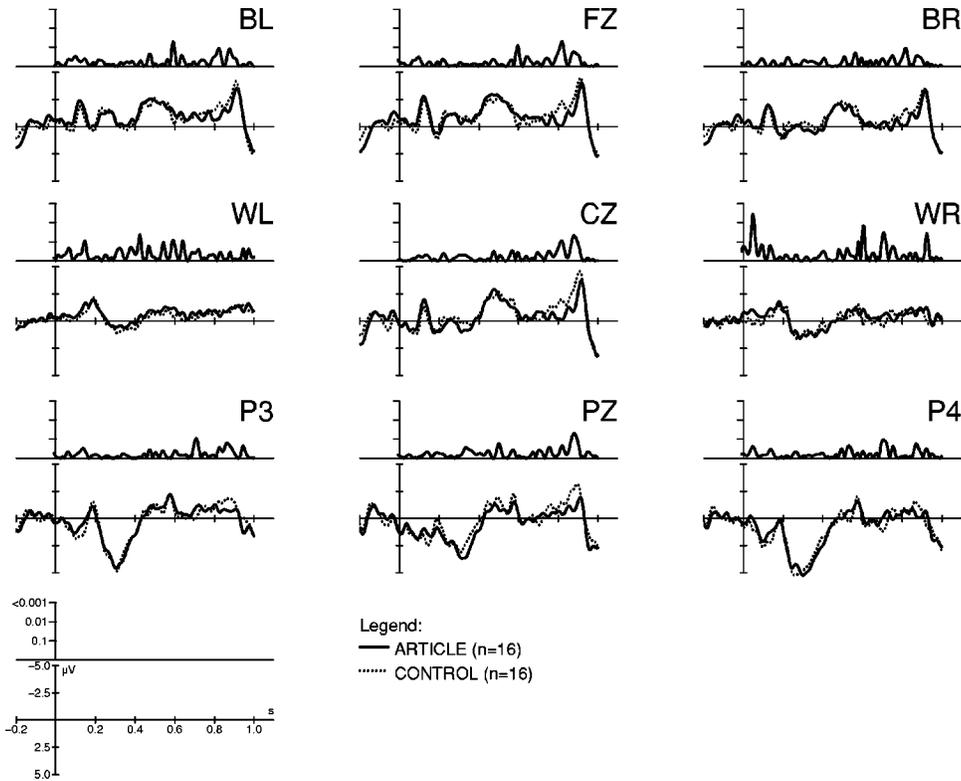


FIG. 24. Voltage grand averages for the ARTICLE condition (solid lines) compared to the CONTROL-ARTICLE condition (dotted). Lower panels: potential (μV) against time (s), $t=0$: stimulus onset time. Upper panels: t -test error probabilities (logarithmic scale).

mensions. Molnár *et al.* showed that the P300 ERP component found in the oddball experiment (see the Introduction) is reflected by a drop in pointwise correlation dimension. Our approach has the advantage that it requires only counting of words and computing entropies and significances. It neither demands the application of embedding techniques nor employs least-square fits to almost linear scaling regions.

So far, we have shown that methods of symbolic dynamics and measures of complexity are able to capture the well-known ERP components. Additionally, our newly developed techniques reveal more structure hidden in the data. However, there remain open questions. Using a static symbolic encoding of data after performing a ranking procedure enables us to compare the results directly with voltage averages provided by the traditional technique. On the other hand, dynamical or mixed strategies of coarse graining may provide better insights into neural information processing since the first derivative of evoked potentials seems to be related to the firing probability of neurons [[52], p. 40]. However, such results might be hard to interpret without an interface to voltage averages.

One central issue is the interpretation of ERP in terms of brain dynamics. The traditional averaging technique assumes the ERP's to be deterministic signals generated by the brain as responses to certain stimuli or cognitive events. Data analysis is thus the job of improving the signal-to-noise ratio [38]. Alternatively, we can view ERP's as indicators of reorganization in the ongoing spontaneous EEG activity. By using methods of spectral analysis, one can show that the distributions of phase spectral values of ERP epochs are different between experimental conditions, while distributions of amplitudes are not [38]. Also symbolic dynamics neglects some information about amplitudes but it restores information about timing. Hence, the question of synchronization and phase locking of neuronal oscillators arises in the con-

text of ERP studies [52,77–81]. A further issue is the role of spatiotemporal correlations of ERP. These could be investigated using the notions of conditional entropy and mutual information in the framework of symbolic dynamics [62,75].

ACKNOWLEDGMENTS

We would like to thank Ralf Engbert, Günter Troll, Thomas Liebscher, and Kelly Sloan for stimulating and helpful discussions and for a critical reading of the manuscript. Wolfgang Jansen's support concerning numerical problems is gratefully acknowledged. We thank Sandra Neumayer and Saskia Kohlen for their work in the laboratory as well as Bryan Jurish for fixing all computer problems. This work has been supported by the Deutsche Forschungsgemeinschaft within the scientists group on conflicting rules in cognitive systems.

APPENDIX

1. Proof of the partitioning lemma

Assume we have a (time discrete) dynamical system (X, Φ_r) together with an observable $h: X \rightarrow \mathbb{R}$ that maps states onto real numbers. Let $h(X) = H \subset \mathbb{R}$ be the image of X under the action of h . Now, we choose a partition S_i , $i = 1, 2, \dots, I$ of H and determine the preimages of the sets S_i in X : $A_i = h^{-1}(S_i)$. According to a basic theorem of set theory for any map $f: X \rightarrow Y$ holds,

$$f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B), \quad (\text{A1})$$

$$f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B) \quad (\text{A2})$$

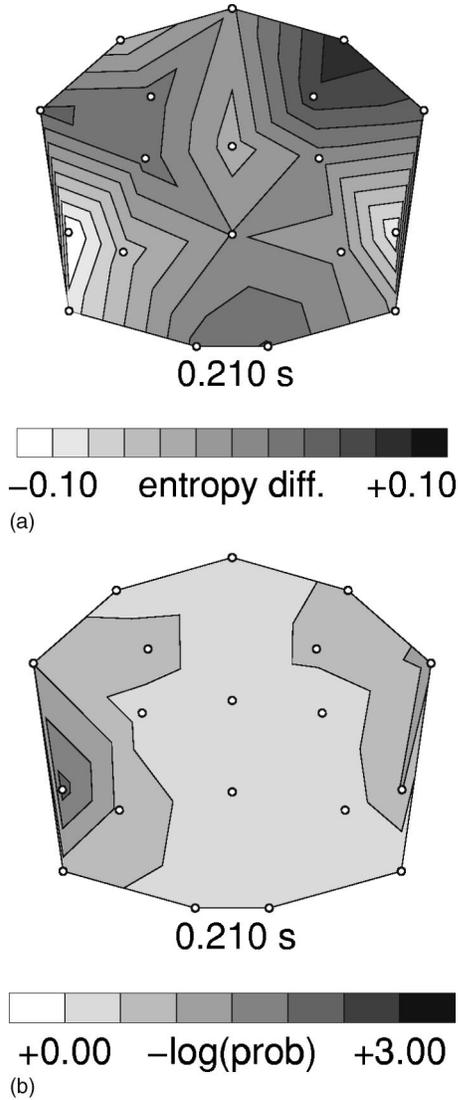


FIG. 25. (a) Rényi entropy difference map of the ARTICLE condition subtracted from CONTROL-ARTICLE condition of the left early entropy rise of the grand ensemble ($n=1; q=10.0$). (b) Significance map $[-\log_{10}(p_{\text{error}})]$ of the windowed χ^2 -test of one-word statistics in a sliding 20-ms window.

for all $A, B \subset Y$. Hence we compute first $A_i \cap A_j A_i \cap A_j = h^{-1}(S_i) \cap h^{-1}(S_j) = h^{-1}(S_i \cap S_j) = \emptyset$, if $i \neq j$. And second $h^{-1}(H) = h^{-1}(\cup_{i=1}^l S_i) = \cup_{i=1}^l h^{-1}(S_i) = \cup_{i=1}^l A_i = X$.

2. The language-processing experiment

The ERP experiment was performed using 30 structurally equivalent sentences for each condition and five conditions total. We report here the results of two experimental conditions: VERB and ARTICLE against their corresponding control conditions CONTROL-VERB and CONTROL-ARTICLE. As mentioned before, the sentences were presented word by word showing the stimuli for 400 ms followed by a break of 400 ms, too. The sampling of ERP epochs started 200 ms before the critical word occurred and finished 1000 ms after. Thus, at time $t=0$ ms the critical words appeared. The sentences

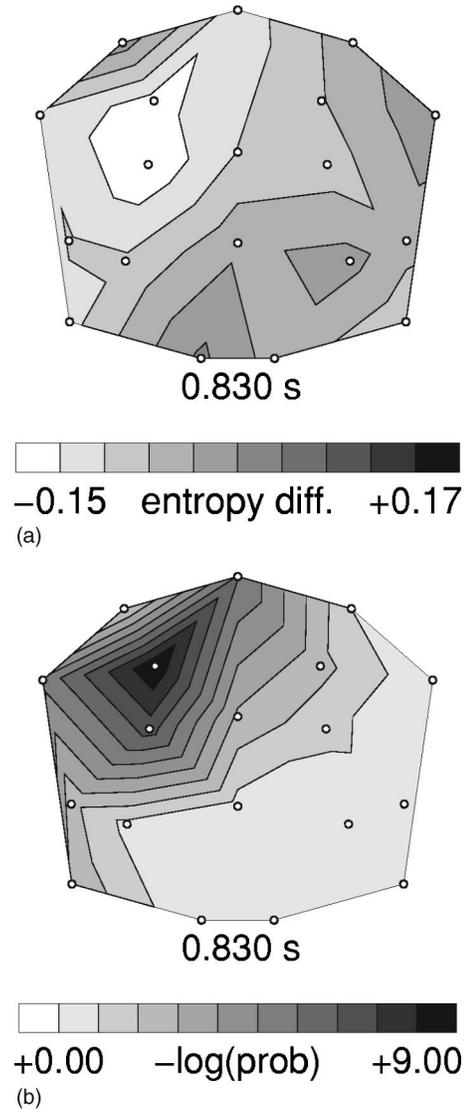


FIG. 26. (a) Rényi entropy difference map of the ARTICLE condition subtracted from CONTROL-ARTICLE condition of the late entropy rise of the grand ensemble ($n=1; q=10.0$). (b) Significance map $[-\log_{10}(p_{\text{error}})]$ of the windowed χ^2 -test of one-word statistics in a sliding 20-ms window.

presented to the subjects were initially ambiguous interrogative sentences. The point of disambiguation was either at the verb (VERB condition) or at the second article (ARTICLE condition). The control sentences were not ambiguous at these points. Table I shows examples of the German sentences, an English paraphrase, and the correct English translation. The critical words where ERP measurement took place are printed in italics.

Interrogative sentences in German are usually expected to be of subject-verb-object order (examples: CONTROL-VERB, CONTROL-ARTICLE). But object-verb-subject order is also possible (examples: VERB and ARTICLE). The first two words in the sentences, “welche Frau,” are ambiguous with respect to nominative or accusative case and are always the same. Subjects expect the sentences to be of the frequent subject-verb-object structure. There are two ways for disambiguating such sentences: either by the number feature, i.e.,

TABLE I. Example sentences of the language-processing ERP experiment. Critical words (where ERP measurements happen) are printed in italics.

Condition	Example	Paraphrase	Translation
VERB	welche Frau <i>sahen</i> die Männer?	which woman (accusative singular) saw (plural) the men (nominative plural)?	which men saw the woman?
CONTROL-VERB	welche Frau <i>sah</i> den Mann?	which woman (nominative singular) saw (singular) the men (accusative singular)?	which woman saw the man?
ARTICLE	welche Frau <i>sah</i> <i>der</i> Mann?	which woman (accusative singular) saw (singular) the man (nominative singular)?	which man saw the woman?
CONTROL-ARTICLE	welche Frau <i>sah</i> <i>den</i> Mann?	which woman (nominative singular) saw (singular) the man (accusative singular)?	which woman saw the man?

subject-noun and verb must agree in number (singular or plural), or by the case feature, i.e., the subject-noun has the nominative case, whereas the object-noun bears the accusative case.

In the VERB condition the disambiguating information comes at the verb that does not agree with the subject-noun in number: the subject noun “Frau” is singular, the verb “sahen” is plural. The sentence seems to be ungrammatical at that point and a reanalysis of the structure built so far must take place. The P600 ERP component indicates this revision

of structure and meaning caused by the violation of expectation.

In the ARTICLE condition the additional information is provided by the article “der” that is nominative case marked (an accusative case marked article would be “den”). As in the VERB condition, the sentence must be reanalyzed, but there is no significant event in voltage ERP averages. We show that in the ARTICLE condition the entropy increases compared with the CONTROL-ARTICLE condition (for further reading, see [73]).

- [1] *Essentials of Neural Science and Behavior*, edited by E. R. Kandel, J. H. Schwartz, and T. M. Jessel (Appleton & Lange, East Norwalk, CT, 1995).
- [2] D. Johnston and S. M.-S. Wu, *Foundations of Cellular Neurophysiology* (MIT Press, Cambridge, MA, 1997), 3rd printing.
- [3] P. L. Nunez, *Electric Fields of the Brain* (Oxford University Press, New York, 1981).
- [4] M. Rugg, in *Cognitive Neuroscience*, edited by M. Kutas and A. Dale (Psychology Press, Hove East Sussex, 1997), pp. 197–242.
- [5] H. Haken, *Principles of Brain Functioning*, Springer Series in Synergetics Vol. 67 (Springer, Berlin, 1996).
- [6] J. A. S. Kelso, *Dynamic Patterns. The Self-Organization of Brain and Behavior* (MIT Press, Cambridge, MA, 1999), 3rd printing.
- [7] A. Babloyantz, J. M. Salazar, and C. Nicolis, *Phys. Lett.* **111A**, 152 (1985).
- [8] P. E. Rapp, I. D. Zimmerman, A. M. Albano, G. C. deGuzman, N. N. Greenbaum, and T. R. Bashore, in *Nonlinear Oscillations in Biology and Chemistry*, edited by H. G. Othmer, Lecture Notes in Biomathematics Vol. 66 (Springer, Berlin, 1986), pp. 175–205.
- [9] A. M. Albano, N. B. Abraham, G. C. Guzman, M. F. H. Tarroja, D. K. Bandy, R. S. Gioggia, P. E. Rapp, I. D. Zimmerman, N. N. Greenbaum, and T. R. Bashore, in Mayer-Kress [82], pp. 231–240.
- [10] A. Babloyantz, in Mayer-Kress [82], pp. 241–245.
- [11] S. P. Layne, G. Mayer-Kress, and J. Holzfuss, in Mayer-Kress [82], pp. 246–256.
- [12] P. E. Rapp, T. E. Bashore, J. M. Martinerie, A. M. Albano, I. D. Zimmerman, and A. I. Mees, *Brain Topogr.* **2**, 99 (1989).
- [13] D. Gallez and A. Babloyantz, *Biol. Cybern.* **64**, 381 (1991).
- [14] W. Lutzenberger, T. Elbert, N. Birbaumer, W. J. Ray, and H. Schupp, *Brain Topogr.* **5**, 27 (1992).
- [15] W. S. Pritchard and D. W. Duke, *Psychophysiology* **29**, 182 (1992).
- [16] E. T. Bullmore, M. J. Brammer, P. Bourlon, G. Alarcon, C. E. Polkey, R. Elwes, and C. D. Binnie, *Electroencephalogr. Clin. Neurophysiol.* **91**, 337 (1994).
- [17] T. Elbert, W. J. Ray, Z. J. Kowalik, J. E. Skinner, K. E. Graf, and N. Birbaumer, *Physiol. Rev.* **74**, 1 (1994).
- [18] C. Besthorn, H. Sattel, C. Geiger-Kabisch, R. Zerfass, and H. Förstl, *Electroencephalogr. Clin. Neurophysiol.* **95**, 84 (1995).
- [19] A. R. Osborne and A. Provenzale, *Physica D* **35**, 357 (1989).
- [20] P. E. Rapp, A. M. Albano, T. I. Schmah, and L. A. Farwell, *Phys. Rev. E* **47**, 2289 (1993).

- [21] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, *Physica D* **58**, 77 (1992).
- [22] M. Molnár, J. E. Skinner, V. Csépe, I. Winkler, and G. Karmos, *Electroencephalogr. Clin. Neurophysiol.* **95**, 118 (1995).
- [23] B.-L. Hao, *Physica D* **51**, 161 (1991).
- [24] X. Z. Tang and E. R. Tracy, *Chaos* **8**, 688 (1998).
- [25] R. Wackerbauer, A. Witt, H. Atmanspacher, J. Kurths, and H. Scheingraber, *Chaos, Solitons and Fractals* **4**, 133 (1994).
- [26] J. S. Shiner, M. Davison, and P. T. Landsberg, *Phys. Rev. E* **59**, 1459 (1999).
- [27] J. Kurths, A. Voss, A. Witt, P. Saparin, H. J. Kleiner, and N. Wessel, *Chaos* **5**, 88 (1995).
- [28] M. Schiek, F. R. Drepper, R. Engbert, H. H. Abel, and K. Suder, in *Kantz et al.* [83], pp. 191–213.
- [29] C. Scheffczyk, A. Zaikin, M. Rosenblum, R. Engbert, R. Krampe, and J. Kurths, *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **7**, 1441 (1997).
- [30] R. Engbert, C. Scheffczyk, R. T. Krampe, M. Rosenblum, J. Kurths, and R. Kliegl, *Phys. Rev. E* **56**, 5823 (1997).
- [31] P. Tass, J. Kurths, M. Rosenblum, G. Guasti, and H. Hefter, *Phys. Rev. E* **54**, R2224 (1996).
- [32] P. I. Saparin, W. Gowin, J. Kurths, and D. Felsenberg, *Phys. Rev. E* **58**, 6449 (1998).
- [33] P. E. Rapp, I. D. Zimmerman, E. P. Vining, N. Cohen, A. M. Albano, and M. A. Jiménez-Montano, *J. Neurosci.* **14**, 4731 (1994).
- [34] U. Schwarz, A. O. Benz, J. Kurths, and A. Witt, *Astron. Astrophys.* **277**, 215 (1993).
- [35] T. Buchner and J. J. Zebrowski, *Phys. Rev. E* **60**, 3973 (1999).
- [36] D. Regan, *Evoked Potentials in Psychology, Sensory, Physiology and Clinical Medicine* (Chapman and Hall, London, 1972).
- [37] D. Regan, *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine* (Elsevier, New York, 1989).
- [38] *Electroencephalography. Basic Principles, Clinical Applications, and Related Fields*, 3rd ed., edited by E. Niedermeyer and F. L. D. Silva (Williams and Wilkins, Baltimore, 1993).
- [39] B. W. Gnedenko, *Lehrbuch der Wahrscheinlichkeitstheorie*, 10th ed. (Harry Deutsch Verlag, Thun, Frankfurt a. M., 1997).
- [40] J. Möcks, T. Gasser, and P. D. Tuan, *Electroencephalogr. Clin. Neurophysiol.* **57**, 571 (1984).
- [41] E. Callaway and R. A. Halliday, *Electroencephalogr. Clin. Neurophysiol.* **34**, 125 (1973).
- [42] R. Coppola, R. Tabor, and M. S. Buchsbaum, *Electroencephalogr. Clin. Neurophysiol.* **44**, 214 (1978).
- [43] T. Gasser, J. Möcks, and R. Verleger, *Electroencephalogr. Clin. Neurophysiol.* **55**, 717 (1983).
- [44] A. Puce, S. F. Berkovic, P. J. Cadusch, and P. F. Bladin, *Electroencephalogr. Clin. Neurophysiol.* **92**, 352 (1994).
- [45] S. Krieger, J. Timmer, S. Lis, and H. M. Olbrich, *J. Neural Transm.* **99**, 103 (1995).
- [46] J. S. Buchwald, in *Event-Related Potentials. Basic Issues and Applications*, edited by J. W. Rohrbaugh, R. Parasuraman, and R. Johnson (Oxford University Press, New York, 1990), pp. 57–75.
- [47] J. E. Hoffman, in Ref. [46], pp. 145–157.
- [48] M. Kutas and C. K. van Petten, in *Handbook of Psycholinguistics*, edited by M. A. Gernsbacher (Academic, San Diego, 1994), pp. 83–133.
- [49] A. D. Friederici, in *Language Comprehension: A Biological Perspective*, 2nd ed., edited by A. D. Friederici (Springer, Berlin, 1999), pp. 265–304.
- [50] F. Rösler, *Hirnelektrische Korrelate Kognitiver Prozesse*, Lehr- und Forschungstexte Psychologie (Springer, Berlin, 1982).
- [51] W. Skrandies, *Brain Topogr.* **2**, 73 (1989).
- [52] E. Başar, *EEG-Brain Dynamics. Relations between EEG and Brain Evoked Potentials* (Elsevier/North Holland Biomedical, Amsterdam, 1980).
- [53] H. Haken, J. A. S. Kelso, and H. Bunz, *Biol. Cybern.* **51**, 347 (1985).
- [54] J. A. S. Kelso, S. L. Bressler, S. Buchanan, G. C. DeGuzman, M. Ding, A. Fuchs, and T. Holroyd, *Phys. Lett. A* **196**, 134 (1992).
- [55] E. Başar, in *Synergetics of the Brain*, edited by E. Başar, H. Flohr, H. Haken, and A. J. Mandell, Springer Series in Synergetics Vol. 23 (Springer, Berlin, 1983), pp. 183–200.
- [56] J. P. Crutchfield and N. H. Packard, *Int. J. Theor. Phys.* **21**, 434 (1982).
- [57] J. P. Crutchfield and N. H. Packard, *Physica D* **7**, 201 (1983).
- [58] C. Beck and F. Schlögl, *Thermodynamics of Chaotic Systems. An Introduction*, Vol. 4 of *Cambridge Nonlinear Science Series* (Cambridge University Press, Cambridge, UK, 1993), reprinted in 1997.
- [59] H. Haken and G. Mayer-Kress, *Z. Phys. B: Condens. Matter* **43**, 185 (1981).
- [60] B.-L. Hao, *Elementary Symbolic Dynamics and Chaos in Dissipative Systems* (World Scientific, Singapore, 1989).
- [61] B. McMillan, *Ann. Math. Stat.* **24**, 196 (1953).
- [62] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949), reprinted in 1963.
- [63] F. Takens, in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L.-S. Young, Lecture Notes in Mathematics Vol. 898 (Springer, Berlin, 1981), pp. 366–381.
- [64] A. Rényi, *Probability Theory* (North-Holland, Amsterdam, 1970).
- [65] P. Grassberger, *Phys. Lett. A* **128**, 369 (1988).
- [66] A. O. Schmitt, H. Herzog, and W. Ebeling, *Europhys. Lett.* **23**, 303 (1993).
- [67] H. Herzog, A. O. Schmitt, and W. Ebeling, *Chaos, Solitons and Fractals* **4**, 97 (1994).
- [68] J. Kurths and H. Herzog, *Physica D* **25**, 165 (1987).
- [69] X. Pei and F. Moss, *Nature (London)* **379**, 618 (1996).
- [70] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, New York, 1992), reprinted in 1996.
- [71] P. Collet and J. -P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems* (Birkhauser, Boston, 1980).
- [72] H. G. Schuster, *Deterministic Chaos* (VCH, Weinheim, 1989).
- [73] J. D. Saddy, P. beim Graben, and M. Schlesewsky, *Cortical Dynamics of Language Processes*, Proceedings of the European Conference on Cognitive Science (ECCS), edited by S. Bagnara (Multimedia & Communication Laboratory, University of Siena, Siena, Italy, 1999), pp. 323–332.
- [74] R. Nowagk and E. Pfeifer, EEP 3.0, Event Related Potential Evaluation Package (1998), Max Planck Institute for Cognitive Neuroscience, Leipzig.
- [75] B. Pompe, in *Kantz et al.* [83], pp. 67–90.
- [76] D. Lehmann, *Electroencephalogr. Clin. Neurophysiol.* **31**, 439 (1971).

- [77] G. Pfurtscheller, in *Rhythms in Physiological Systems*, edited by H. Haken and H. P. Koepchen, Springer Series in Synergetics Vol. 55 (Springer, Berlin, 1991), pp. 289–296.
- [78] W. Singer, *Nature (London)* **397**, 391 (1999).
- [79] E. Rodriguez, N. George, J.-P. Lachaux, J. Martinerie, B. Renault, and F. J. Varela, *Nature (London)* **397**, 430 (1999).
- [80] W. H. R. Miltner, C. Braun, M. Arnold, H. Witte, and E. Taub, *Nature (London)* **397**, 434 (1999).
- [81] P. Tass, M. G. Rosenblum, J. Weule, J. Kurths, A. Pikovski, J. Volkmann, A. Schnitzler, and H.-J. Freund, *Phys. Rev. Lett.* **81**, 3291 (1998).
- [82] G. Mayer-Kress, in *Dimensions and Entropies in Chaotic Systems*, edited by G. Mayer-Kress, Springer Series in Synergetics Vol. 32 (Springer, Berlin, 1986).
- [83] *Nonlinear Analysis of Physiological Data*, edited by H. Kantz, J. Kurths, and G. Mayer-Kress (Springer, Berlin, 1998).
- [84] We allow $x_i(t)$ to be time continuous.
- [85] The logistic map discussed above is an example for a time discrete and noninvertible map. In order to introduce the notions of bi-infinite symbolic strings and cylinder sets subsequently, we shall investigate invertible deterministic dynamics generally.
- [86] Note that due to an electrophysiologists convention, negativity is up.