Mental States as Macrostates
Emerging from EEG Dynamics

Carsten Allefeld   Harald Atmanspacher   Jiří Wackermann

Institute for Frontier Areas of Psychology and Mental Health, Freiburg, Germany

Abstract

Correlations between psychological and physiological phenomena form the basis for different medical and scientific disciplines, but the nature of this relation has not yet been fully understood. One conceptual option is to understand the mental as "emerging" from neural processes in the specific sense that psychology and physiology provide two different descriptions of the same system. Stating these descriptions in terms of coarser- and finer-grained system states (macro- and microstates), the two descriptions may be equally adequate if the coarse-graining preserves the possibility to obtain a dynamical rule for the system. To test the empirical viability of our approach, we describe an algorithm to obtain a specific form of such a coarse-graining from data, and illustrate its operation using a simulated dynamical system. We then apply the method to an electroencephalographic (EEG) recording, where we are able to identify macrostates from the physiological data that correspond to mental states of the subject.

1 Introduction

The existence of correlations between psychological and physiological phenomena, especially brain processes, is the basic empirical fact of psychophysiological research. Relations between mental processes, including modes of consciousness, and those occurring in its physical "substrate", the central nervous system, are generally taken as a matter of course: They form the basis for the use of drugs in the treatment of mental disorders in psychiatry, they are applied as a research tool to shed light on the details of psychological mechanisms, and the explication of the neural structures underlying mental functioning forms the subject of cognitive neuroscience. Still, it remains unclear what the nature of the observed correlations is and what exactly is to be conceived as a neural correlate of a psychological phenomenon.[1]

One way to approach these issues is to interpret the mental as a domain *emerging* from an underlying physiological domain (Broad, 1925; Beckermann et al., 1992). However, despite its long history reaching to recent scientific contributions (e.g. Darley, 1994; Seth, 2008), the term emergence is not very well defined and it is used in a large number of different meanings (cf. Stephan, 2002; O'Connor and Wong, 2006).

In our understanding, emergence is a relation between different *descriptions* of the same system. In this view, the occurrence and correlation of psychological and physiological phenomena is due to the fact that the object of psychophysiological research (the research subject) can be approached and examined in different ways. More specifically, emergence is to be conceived as a relation between different descriptions each of which is useful or adequate in its own manner. The question arises how there can be more than one adequate descrip-

---

[1] While the recent discussion focuses on the notion of "neural correlates of consciousness" (cf. Metzinger, 2000), we are interested in psychophysiological correlations in general.

tion for the same system, and what has to be the nature of their relation in order to permit this.[2]

In this paper we present one possible answer to these questions, motivated by ideas on the emergence of mental states from neurodynamics introduced by Atmanspacher and beim Graben (2007), where the two descriptions take on the form of a dynamical system. We introduce the further specification that the relation between the two associated state spaces is characterized by a Markov coarse-graining (Sec. 2), which leads us to consider metastable states as a particular form of emergent states. In order to demonstrate the practical viability of these ideas, we develop a method to identify metastable states from empirical data (Sec. 3), and illustrate the operation of the algorithm using data from a simulated system (Sec. 4). In the application of the method to a recording of brain electrical activity, we are able to identify states closely corresponding to the mental states of a subject, based on the analysis of the EEG data alone (Sec. 5).

2 Emergence in dynamical systems

A descriptive approach that has proven very fruitful in physics and other fields of the natural sciences is utilizing the concept of a *dynamical system* (Robinson, 1995; Chan and Tong, 2001). Such a description is formulated with respect to the *states* the system can assume, and a dynamical rule that defines the way the state of the system evolves over time. The possible system states form a *state space*, which in the most general case is just a set of identifiable and mutually distinguishable elements.[3]

For a well-defined relation between two such descriptions to hold, it is necessary that the two state spaces can be related to each other.[4] A simple possibility is that the system assumes a particular state in one description exactly if it is in any out of a certain set of states of the other description; that is to say, one state space is a *coarse-graining* or *partition* of the other state space. Because of this asymmetry between the two descriptions one may speak of a higher-level and a lower-level description, and refer correspondingly to *macrostates* and *microstates* of the system. The classic example in physics for this kind of inter-level relation is that between the phenomenological theory of thermodynamics, dealing with the macrostates of extended systems defined in terms of observables such as temperature and pressure, and the theory of statistical mechanics, relating them to microstates defined in terms of the constituents of those systems.[5]

The description of a system is chosen by an observer, but it is also subject to objective constraints insofar as different descriptions may be differently adequate or useful. For a description as a dynamical system, the adequacy of a particular set of system states becomes apparent in the possibility to find a *dynamical rule*, $\Phi_{\Delta t}$, whereby the current state x_t of the system determines its further evolution,

$$x_{t+\Delta t} = \Phi_{\Delta t}\left(x_t\right);$$

here t is a continuous or discrete time variable and Δt a time interval. A particular state space definition may therefore be called *dynamically adequate* if the specification of a state implies all the available information which is relevant for determining subsequent states, that is, if in this description the system possesses the *Markov property* (cf. Shalizi and Moore, 2008). In this sense, the most general model of a dynamical system is the Markov process—a stochastic model which includes deterministic dynamics as a limiting case (cf. Chan and

---

[2]Note that we are concerned with the atemporal or "synchronous" structure of such a relation between descriptions, and do not address the question of how a phenomenon emerges "diachronically", in a process unfolding in time.

[3]This is in accordance with the concept of system states in cybernetics and related disciplines (cf. Ashby, 1962), but is at variance with the use of the term in physics where a state space is generally taken to be spanned by a set of observables (properties that can be precisely quantified). Such a less structured concept of state space is useful because it also covers cases where it is not obvious how to endow that space with a formal structure, for instance mental states. However, as Gaveau and Schulman (2005) point out, introducing into a state space a dynamics in the form of transition probabilities (see below) implicitly provides it with a metric structure.

[4]This of course does not have to be the case; different descriptions of the same system may also be incompatible with each other.

[5]In this context the terms macrostate and microstate derive from the circumstance that they refer to the properties of a "macroscopic" system versus those of its "microscopic" constituents. Though these terms often imply a difference in spatiotemporal scale, the important point is the difference in the amount of detail given by the descriptions.

Tong, 2001).

It is important to note that this criterion for selecting a descriptive level implies a reference back to that same level; while employing a more fine-grained set of states may serve to improve the prediction of the future of a system in general, it will in most cases result in a loss of the Markov property with respect to these finer-grained states themselves. In other words, the Markov-property criterion distinguishes descriptive levels at which the system exhibits a self-contained dynamics ("eigendynamics"), independent of details present at other levels.[6]

This specification of the kind of descriptions sought for leads to a more specific concept of emergence as an inter-level relation.[7] Given a microscopic state description exhibiting the Markov property, an adequate higher-level description or coarse-graining should ideally preserve it. In the context of deterministic nonlinear systems, where the dynamics is defined by a map from a metric space onto itself, such a coarse-graining is called a Markov partition (Adler, 1998; Bollt and Skufca, 2005); for the general case of stochastic dynamics we propose the term *Markov coarse-graining* (cf. Gaveau and Schulman, 2005). Accordingly, states of a higher-level description may be called dynamically emergent states if they correspond to a Markov coarse-graining of a lower-level dynamics.

Interpreting psychophysiological correlations as reflecting a relation of emergence between two levels of description as a dynamical system, the lower-level description is stated in terms of physiological, neural states, the higher-level description in terms of mental states. At both levels a wide variety of descriptive approaches is possible, depending on the experimental methods used to assess the brain state on the one hand (electrophysiology, imaging methods, brain chemistry, etc.) and the chosen set of psychological categories on the other hand (conscious/unconscious, sleep stages, moods, cognitive modes, etc.).[8] Applying the dynamical specification of emergence outlined above, emergent macrostates that are defined via a Markov coarse-graining of the neural microstate dynamics are candidates for a further characterization as mental states. In order to empirically substantiate these ideas, macrostates obtained from the dynamics that has been observed in neurophysiological data are to be related to mental states of subjects that have been determined by other means, such as behavorial assessment or verbal reports.

In the following we undertake first steps towards this program. Since a general algorithm for finding Markov coarse-grainings is not known, we focus on the special case of *metastable states*. Because a system stays in such a state for prolonged periods of time and only occasionally switches into another, antecedent states provide practically no information on the subsequent evolution beyond that implied in the current state, so that the macrostate dynamics is approximately Markovian. The following section describes an algorithm to obtain metastable states from the microstate dynamics observed in empirical data.

3 Identifying metastable macrostates from data

Metastable states correspond to the "almost invariant sets" of a dynamical system, i.e. subsets of the state space which are approximately invariant under the system's dynamics. Since we are dealing with empirical data where there is generally no precise theoretical knowledge of the dynamics, it has to be determined from the data.

Via a finite set of microstates resulting from a discretization of the state space (Sec. 3.1), the time

---

[6]This concept is akin to the idea of operational closure or autonomy in the theory of autopoietic systems (Maturana and Varela, 1980; Varela, 1979), which alongside the separation from the environment also refers to the indifference of system operations towards the internal "microscopic" complexity of system elements (cf. Luhmann, 1996). However, the topic of self-defined system boundaries is not addressed in this paper and accordingly, the term "system" is used in the unspecific sense of a section of reality which has been chosen for observation.

[7]The stability conditions of Atmanspacher and beim Graben (2007) are here realized by the Markov property, while contextual constraints (Bishop and Atmanspacher, 2006) can be seen effective in the selection of a particular descriptive level out of those admissible.

[8]Since each mental state allows for multiple realizations at the neural level, mental states may be said to "supervene on" brain states (cf. Kim, 1993)—but this alone does not provide a sufficient characterization of their relation. Moreover, contrary to assumptions prevalent in the discussion (cf. Chalmers, 2000) a neural correlates need not necessarily be realized in a particular neural subsystem of the brain.

evolution operator $\Phi_{\Delta t}$ is estimated in the form of a matrix of transition probabilities P (Sec. 3.2). Metastable states are then determined using an algorithm to find the almost invariant sets of a Markov process (Sec. 3.3). Additionally, an estimate of the optimal number of macrostates is obtained via an analysis of the characteristic timescales of the dynamics (Sec. 3.4). Our algorithm builds on work by Deuflhard and Weber (2005), Gaveau and Schulman (2005), and Froyland (2005), and re-uses an idea of Allefeld and Bialonski (2007).

3.1 Discretization of the microstate space

In order to represent the observed microstate dynamics as a finite-state Markov process, the state space defined by K variables $(x_1, x_2, \ldots, x_K) = x$ has to be discretized, resulting in a set of compound microstates which forms the basis for further analysis. Since the data set may be high-dimensional and of varying density in different areas of the state space, we need a flexible algorithm which adapts the size and shape of microstate cells to local properties of the distribution of data points.

This procedure has to meet two competing goals: It should capture as much detail as possible in order to faithfully represent the underlying continuous dynamics within its discretized version; but since transition probabilities between cells are to be estimated, the number of data points per cell should not fall below a certain minimum. Moreover, the extensions of the cells in the directions of the different variables should be of roughly the same size.[9]

To achieve this, we use a recursive bipartitioning approach (Fig. 1): For a given set of n data points $S = \{x_m\}$, $m = 1 \ldots n$, the direction of maximal variance is determined, i.e. a unit vector e, $|e| = 1$, such that $\text{var}_m (x_m \cdot e)$ obtains its maximum value. Using the median M of the data points' positions along this direction as a threshold

Figure 1: Discretization by recursive bipartitioning, illustrated with a set of data points drawn from a two-dimensional normal distribution stretched out along the main diagonal. Cuts occurring earlier in the procedure are indicated by thicker lines.

old value, the set is divided into two subsets,

$$S_1 = \{x_m \mid x_m \cdot e \leq M\},$$
$$S_2 = \{x_m \mid x_m \cdot e > M\}.$$

The procedure is repeated for each of the resulting subsets, up to a recursion depth of b steps. This algorithm leads to a practically identical number of data points per cell (either $\lfloor n/2^b \rfloor$ or $\lceil n/2^b \rceil$) which can be adjusted via the parameter b. It provides a high level of detail in those areas of the state space where the system spends most of the time, and it avoids too elongated cells by applying cuts perpendicular to the current main extension.

3.2 Microstate dynamics

Via the bipartitioning procedure, each data point x_m ($m = 1 \ldots n$) is assigned to one out of a finite set of microstates, identified by an index $\mu \in \{1 \ldots N\}$ ($N = 2^b$). The observed sequence of data points (where the index m enumerates samples taken at consecutive time points) is thereby transformed into a sequence of microstate indices μ_m. Considering this sequence of compound microstates as a

---

[9]We assume at this point that the variables spanning the state space permit a comparison of distances along different directions. Where this is not the case it is advisable to map all variables onto the same range of values before performing the discretization.
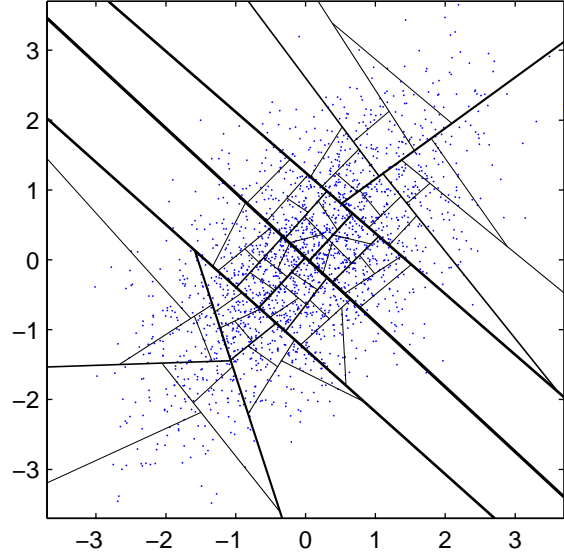
realization of a finite-state Markov process, the underlying dynamics is described by a discrete transfer operator P, an $N \times N$-matrix of transition probabilities between states,

$$P_{ij} = \Pr(\mu_{m+1} = i \mid \mu_m = j),$$

which may be estimated according to

$$\hat{P}_{ij} = \frac{c_{ij}}{\sum_{i'} c_{i'j}}$$

where

$$c_{ij} = \#(\mu_{m+1} = i \wedge \mu_m = j)$$

is the number of observed transitions from state j to state i.

We assume that the Markov process described by P is irreducible, i.e. that it is possible to reach any state from any other state.[10] If this is not the case, the system has not only almost invariant but proper invariant sets, each forming an irreducible process of its own which itself may be subjected to a search for almost invariant subsets.[11] We assume moreover that the process is aperiodic, which is already the case if only one diagonal element P_{ii} is different from zero. For a finite-state Markov process these two properties amount to ergodicity, which implies that there exists a unique invariant probability distribution π over microstates, with $P \pi = \pi$, which is also the limit distribution approached from every initial condition.

The analysis of the dynamical properties of a Markov process leading to the identification of its metastable states is strongly facilitated if it is reversible, i.e. if the dynamics is invariant under time reversal: $P_{ij} \pi_j = P_{ji} \pi_i$ for all i, j. This property cannot usually be assumed for an arbitrary empirically observed process. But since the property of metastability, the tendency of the system to stay within certain regions of the state space for prolonged periods of time, is itself indifferent with respect to the direction of time (Froyland, 2005), we can base the search for the corresponding almost invariant sets on the transition matrix for the

---

[10]We use the terminology and results of Feller (1968), Ch. XV.

[11]Another possible problem is that there may be states a transition into or out of which is never observed, because they only occur at the beginning or end of the given data segments. Along with the general possibility of transient states, this issue is resolved in a natural way by the reversibilization step described below.

reversibilized process R,

$$R_{ij} = \frac{1}{2} \left(P_{ij} + \frac{P_{ji} \, \pi_i}{\pi_j} \right),$$

instead. This operator can be directly estimated according to

$$\hat{R}_{ij} = \frac{c_{ij} + c_{ji}}{\sum_{i'} (c_{i'j} + c_{ji'})}$$

i.e. by counting transitions forwards and backwards in time, and the corresponding invariant probability distribution determined as

$$\hat{\pi}_i = \frac{\sum_j (c_{ij} + c_{ji})}{\sum_{i'} \sum_j (c_{i'j} + c_{ji'})}.$$

In the following we will use the symbols R and π to denote these estimated quantities.

3.3 Almost invariant sets

To identify almost invariant sets we employ the PCCA+ algorithm which was developed by Deuflhard and Weber (2005) to find metastable states in the conformation dynamics of molecules. In this section we outline the main ideas of that approach which are necessary to understand the operation of the method, while for further details on the implementation and mathematical background the reader is referred to their paper.

Our starting point is the ergodic and reversible Markov process characterized by the $N \times N$-transition matrix R along with its invariant probability distribution π. Due to the reversibility of the process (symmetry of the stationary flow $R_{ij} \pi_j$), the left and right eigenvectors A_k, p_k and eigenvalues λ_k of R,

$$A_k R = \lambda_k A_k, \quad R p_k = \lambda_k p_k, \quad k = 1 \dots N,$$

are real-valued. Resolving the scaling ambiguity left by the orthonormality relation $A_k p_l = \delta_{kl}$ by the choice $p_{ik} = \pi_i A_{ki}$ leads to the normalization equations

$$\sum_i \frac{p_{ik}^2}{\pi_i} = 1 \quad \text{and} \quad \sum_i \pi_i A_{ki}^2 = 1,$$

5

and the transition matrix can be given the spectral representation

$$R = \sum_k \lambda_k \, p_k A_k.$$

We assume that the eigenvalues are sorted in descending order, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$. The unique largest eigenvalue $\lambda_1 = 1$ belongs to the invariant probability distribution, $p_1 = \pi$, while the corresponding left eigenvector has constant coefficients, $A_{1i} = 1$.

If the process R possesses q almost invariant sets, it can be seen as the result of a perturbation of a process $\bar{R}$ that possesses q perfectly invariant sets. Any (normalized) element of the right eigenvector subspace of $\bar{R}$ belonging to eigenvalues $\bar{\lambda}_1 = \ldots = \bar{\lambda}_q = 1$ gives an invariant probability distribution of the process. Moreover, if the invariant sets are described by characteristic functions $\bar{\chi}_l$ ($l = 1 \ldots q$), such that $\bar{\chi}_l(i) = 1$ if state i belongs to invariant subset l, 0 otherwise, then any linear combination of them is a left eigenvector of $\bar{R}$ for eigenvalue 1. Conversely, from any linearly independent set of left eigenvectors for eigenvalue 1, $\{\bar{A}_1, \bar{A}_2, \ldots, \bar{A}_q\}$, the characteristic functions of the invariant sets can be recovered via suitable linear combinations.

Through the perturbation, the multiple eigenvalue 1 becomes a cluster of large eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_q$ close to $\lambda_1 = 1$, and the invariant sets become almost invariant sets. They are described by *almost characteristic functions $\chi_l(i)$*, attaining values in the range $[0, 1]$ which may be interpreted as quantifying the degree to which state i belongs to almost invariant set l. In analogy to the unperturbed case, these functions are constructed as linear combinations of the left eigenvectors belonging to the q large eigenvalues,

$$\chi_l(i) = \sum_{k=1}^q \alpha_{kl} \, A_{ki}, \quad l = 1 \ldots q,$$

defined by coefficients $\alpha = (\alpha_{kl})$. Admissible are those regular transforms that conform to the constraints

- partition of unity: $\sum_l \chi_l(i) = 1$ for all i, and

- non-negativity: $\chi_l(i) \geq 0$ for all i, l.

The PCCA+ algorithm optimizes the transform α with respect to an objective function to be maximized; we here choose the *maximum scaling* function

$$I(\alpha) = \sum_l \max_i \chi_l(i),$$

which favors attributions of states i to almost invariant sets l that are as clear-cut as possible.

The input data for the optimization are the dominant left eigenvectors A_k, $k = 1 \ldots q$. The first eigenvector is trivially $A_1 = (1, \ldots, 1)$, but the remaining eigenvector coefficients can be geometrically interpreted as attributing to each microstate i a position in a $(q-1)$-dimensional left *eigenvector space* with position vectors

$$o(i) = (A_{ki}), \quad k = 2 \ldots q.$$

Within this space, the optimization procedure appears as fitting a q-simplex as closely as possible around the microstate points. In a system with pronounced metastable macrostates each of them appears as a cluster of microstates located at the boundary of the point cloud, and the optimization procedure matches these q clusters to one of the vertices of the q-simplex. The values of the almost characteristic functions $\chi_l(i)$ then attain the geometric meaning of barycentric coordinates of the data points with respect to the locations of the simplex vertices v_l:

$$o(i) = \sum_{l=1}^q \chi_l(i) \, v_l.$$

Finally, metastable macrostates corresponding to almost invariant sets of microstates are identified by attributing each microstate i to that macrostate $l \in \{1 \ldots q\}$ for which the almost characteristic function $\chi_l(i)$ attains the highest value (or, to whose defining vertex it is closest in terms of barycentric coordinates).

3.4 Macrostates and timescales

If no prior information on the number of metastable states to be identified is available, it is desirable to obtain an estimate from the data set itself. Since the existence of q almost invariant sets leads to q large eigenvalues, a criterion based on gaps in the eigenvalue spectrum is the natural choice.

However, the concrete values in the spectrum of R depend on the step size of the underlying discrete time, which is implicitly given with the input data. Changing the timescale from 1 to τ steps, the process has to be described by the transition matrix R^τ, whose spectral representation

$$R^\tau = \sum_k \lambda_k^\tau \, p_k A_k$$

is essentially the same as that of R, but with eigenvalues raised to the power τ.

The question which eigenvalues or which gap in the eigenvalue spectrum is to be considered "large" therefore depends on the chosen timescale. As Gaveau and Schulman (2005) note, a coarse-graining of the state space always implies a corresponding "coarse-graining" or rather change of scale with respect to the time axis.

We propose[12] a measure of the size of spectral gaps that is invariant under a rescaling of the time axis. This is achieved by transforming eigenvalues into associated characteristic timescales,

$$T(k) = -\frac{1}{\log |\lambda_k|},$$

and introducing the *timescale separation factor* as the ratio of subsequent timescales:

$$F(k) = \frac{T(k)}{T(k+1)} = \frac{\log |\lambda_{k+1}|}{\log |\lambda_k|}.$$

Substituting λ_k^τ for λ_k in this equation, the resulting factors cancel out, so that $F(k)$ provides a measure of the spectral gap between eigenvalues λ_k and λ_{k+1} that is independent of the timescale.

Using this measure, the number of macrostates q is estimated as the value of k for which $F(k)$ becomes maximal. The choice $q = 1$ leading to a single macrostate comprising all microstates has thereby to be excluded, because it is always associated with the largest timescale separation factor, $F(1) \to \infty$.

If several larger gaps exist, a ranking list of possible q-values may be compiled, where each value leads to a different possible coarse-graining of the system into macrostates. This way different layers of the system's dynamical structure are recov-

---

[12]See Allefeld and Bialonski (2007) for a very similar approach in a different context.

ered, which (extending Deuflhard and Weber's approach) may be considered as the result of multiple superimposed perturbations. An example of this is given in the following section, where the method is illustrated using data from a simulated system.

4 Example: A system with four metastable macrostates

To illustrate the operation of the algorithm we apply it to data from a simulated system, where we can interpret the analysis results with respect to our precise knowledge of the underlying dynamics. We use a discrete-time stochastic system in two dimensions, (x_1, x_2), where the change over each timestep is given by

$$\Delta x_i = a\left(x_i - 2x_i^3\right) + b_i \xi_i,$$

with $a = 0.01$, (ξ_1, ξ_2) standard normal two-dimensional white noise, $b_1 = 0.03$, and $b_2 = 0.05$. The first term of the right hand side of this equation describes an overdamped movement within a double-well potential along each dimension, leading to four attracting fixed points at $(x_1, x_2) = (\pm 1/\sqrt{2}, \pm 1/\sqrt{2})$. Without the stochastic second term, the system would be decomposable into four invariant sets, separated by the two coordinate axes. But due to the noise the system performs a random walk, staying for prolonged periods of time in the vicinity of one of the attracting points, but occasionally wandering into another point's basin of attraction. These switches occur more frequently along x_2 because the noise amplitude is larger in that direction, $b_2 > b_1$.

Data resulting from a simulation run of this system are shown in Fig. 2. A section of the connecting trajectory illustrates how the system state moves through the state space, entering and leaving the cells of the microstate partition. Counting these transitions between cells leads to an estimate of the reversibilized transition matrix R.

The largest eigenvalues of R are plotted in Fig. 3a, revealing a group of four large eigenvalues (> 0.995), which itself is subdivided into two groups of two eigenvalues each. This picture becomes clearer after the transformation into
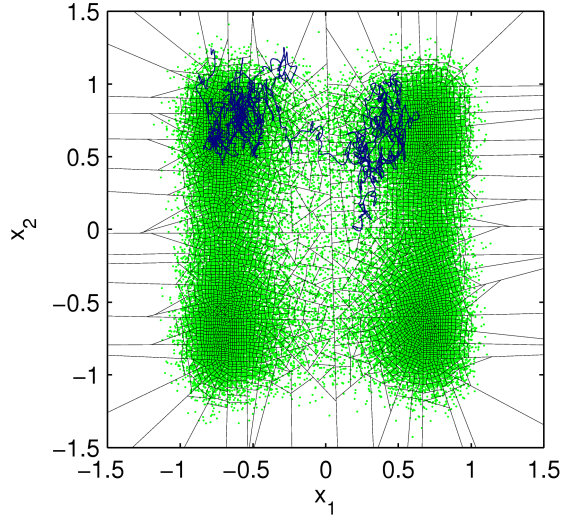
7

Figure 2: Data points from a simulation run of a system with four metastable macrostates over 10^6 time steps, and a part of the connecting trajectory. Straight lines indicate the cell borders of the partition into 4096 microstates obtained via the bipartition algorithm.
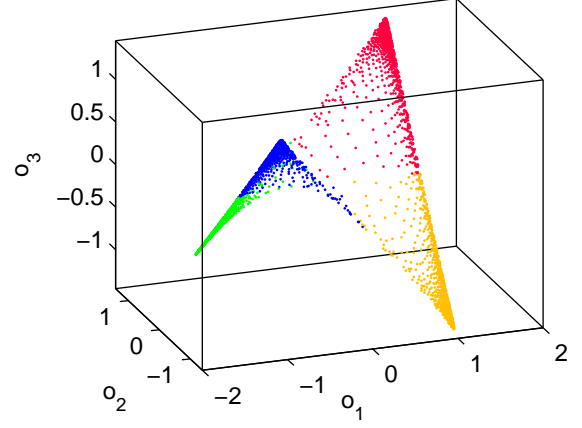

Figure 4: Eigenvector space (o_1, o_2, o_3) of the system with four metastable macrostates for $q = 4$. Each dot representing a microstate is colored according to which vertex of the enclosing tetrahedron is closest, defining the four metastable macrostates.
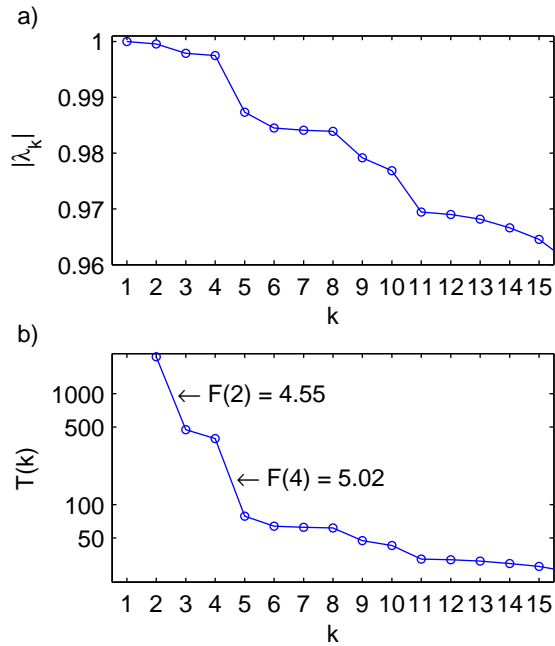
a)

b)

Figure 3: Eigenvalue spectrum of the transition matrix R of the system with four metastable macrostates. (a) The eigenvalues of largest magnitude. (b) Logarithmic timescales; locations and values of the two largest timescale separation factors are indicated.
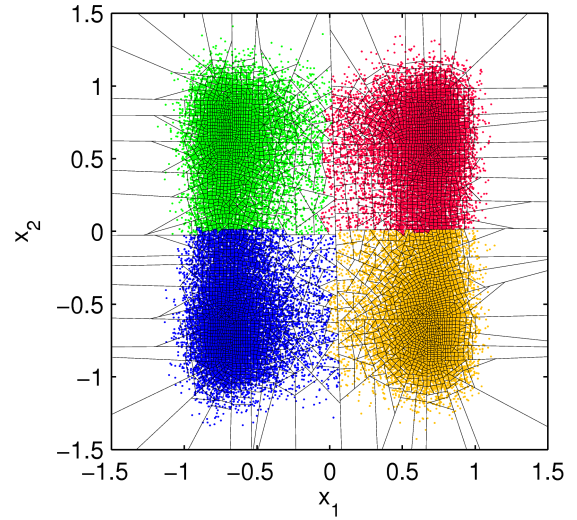

Figure 5: Micro- and macrostates of the system with four metastable macrostates. Lines indicate the cell borders of the partition of the state space into microstates, while the coloring of data points shows the attribution of microstates to one of the four metastable macrostates.

8

timescales $T(k)$. In Fig. 3b they are displayed on a logarithmic scale, such that the magnitude of timescale separation factors $F(k)$ becomes directly visible in the vertical distances between subsequent data points. The largest separation factor is $F(4) = 5.02$, closely followed by $F(2) = 4.55$, indicating that a partitioning of the state space into $q = 4$ macrostates is optimal, while searching for two different macrostates may also yield a meaningful result.

The identification of almost invariant sets of microstates defining the metastable macrostates is performed within the 3-dimensional eigenvector space (o_1, o_2, o_3). Fig. 4 reveals that the points representing microstates are located on a saddle-shaped surface stretched out within a 4-simplex or tetrahedron. The algorithm identifies the vertices of the tetrahedron and attributes each microstate to that macrostate whose defining vertex is closest, resulting in the depicted separation into four sets.

In Fig. 5 this result is re-translated into the original state space of Fig. 2, by coloring the data points of each microstate according to the macrostate it is assigned to. The identified metastable states coincide roughly with the basins of attraction of the four attracting points, i.e., the almost invariant sets of the system's dynamics.

From Fig. 4 we can also assess which macrostate definitions would be obtained by choosing $q = 2$, the next-best choice for the number of metastable states according to the timescale separation factor criterion. In this case the eigenvector space is spanned by the single dimension o_1, along which the two vertices of the tetrahedron on the left and right side, respectively, coincide. This means that the two resulting macrostates each consist of the union of two of the macrostates obtained for $q = 4$. With respect to the state space, these two macrostates correspond approximately to the areas $x_1 > 0$ and $x_1 < 0$.

This result can be understood from the system's dynamics, since because of the smaller probability of transitions along x_1 these two areas of the state space form almost invariant sets, too. As can be seen from this example, the possibility to select different q-values of comparably good rating may allow to recover different dynamical levels of a system, giving rise to a hierarchical structure of potential macrostate definitions.

5 Application to EEG data

For the purpose of a first application of the algorithm to neurophysiological data, we chose an electroencephalographic (EEG) recording from a patient suffering from petit-mal epilepsy, a condition characterized by the occurrence of frequent short (several seconds) epileptic episodes, during which the patient becomes irresponsive (cf. Niedermeyer, 1993). This kind of data is favorable for our methodological approach because we can expect two clearly distinct states to be present—"normal" EEG / mentally present and paroxysmal episodes / mentally absent—, and because it is possible to observe many transitions between these states in a recording of moderate size.

The data set consists of a section of 89 min length from the patient's monitoring EEG. It was recorded from the 19 electrode positions of the international 10-20 system (American Electroencephalographic Society, 1991) at a sampling rate of 250 Hz, digitally bandpass-filtered (2–15 Hz), and transformed to the average reference. Due to artifact removal by visual inspection the amount of data available for analysis was reduced to 71 min total length (1 064 435 data points).

In the preceding simulation example we know by definition that the given values of the system variables immediately specify its dynamical state, and therefore can be directly processed by the algorithm for the identification of metastable states. With measurement data like EEG the situation is not so clear. The data "as is" may be accepted as a specification of the system state, but any further processed version of them fulfills this function as well and may for some reason be even more suitable. This means that empirical data pose the problem of how to define the input state space for the analysis.

For low-dimensional nonlinear deterministic dynamical systems techniques have been developed to reconstruct the state space of the system, or a higher-dimensional space comprising it, from scalar time series via the method of time-delay embedding (Takens, 1981; Kantz and Schreiber, 1997). However, these techniques are not appropriate for our purposes. Firstly, the data-set is already multi-dimensional and using the embedding approach we would have to either blow up the dimensionality even more, thereby introduc-

ing a high amount of redundance, or discard many of the input data channels, possibly loosing crucial information. And secondly, previous attempts to demonstrate low-dimensional nonlinear structure in EEG data had only limited success (cf. Theiler and Rapp, 1996; Paluš, 1996).

Instead, we pursue the following strategy: In a first step, we use the original 19-dimensional data space as the input state space. Guided by the results obtained in this way as well as by independent observations on the behavior of multichannel EEG, in a second step we develop a preprocessing procedure defining a more abstract input state space.

5.1 Original data state space

Using the recursive bipartitioning algorithm, the data points were assigned to 32 768 different compound microstates (32 or 33 points in each cell). The resulting timescale spectrum (Fig. 6a) exhibiting a large separation factor $F(3) = 2.15$ suggests a search for three metastable macrostates in a two-dimensional eigenvector space (Fig. 6b). This is supported by the 3-simplex shape of the distribution of microstate positions within this space.

The arrangement of the areas belonging to the identified macrostates in the input data space is shown in Fig. 7, where the 19-dimensional space is represented using the first three PCA components of the data. The most prevalent state accounting for about 99% of the data points appears here as a centrally located spherical area, with the two other states forming handle-like appendices at opposite sides.

The role of these three macrostates becomes clearer considering the transitions between them over time, in comparison with the underlying EEG time series (Fig. 8a). Within periods of normal electroencephalographic activity the system stays within the "main" macrostate, while during seizures switches between all three states occur regularly, corresponding to an oscillation along the PC1 axis of Fig. 7. This macrostate dynamics reflects the spike-wave oscillatory activity visible in the EEG channels shown in the lower panel of Fig. 8, which are characteristic for paroxysmal episodes.

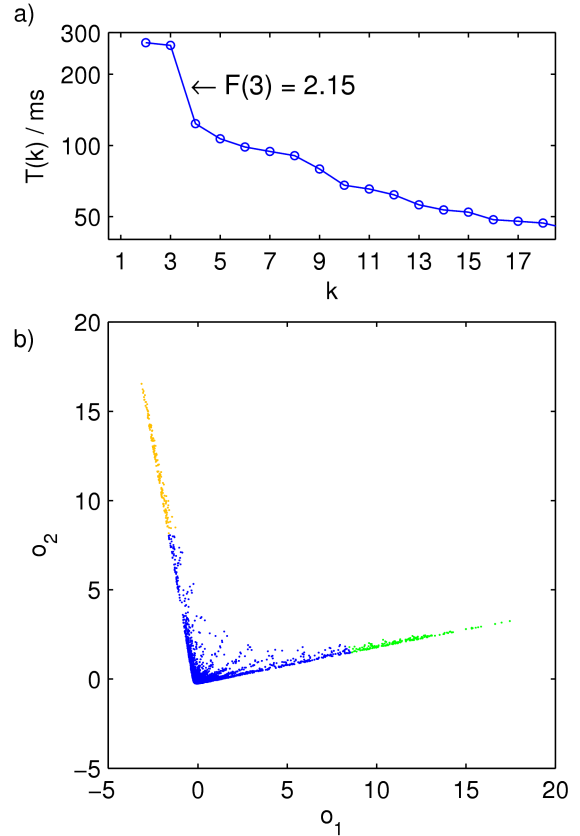With this first result, the attempt at identifying emergent macrostates using the EEG data space is

Figure 6: Analysis results for the original EEG data state space. a) Timescale spectrum; a large separation factor indicates three metastable states. b) Microstate positions in two-dimensional eigenvector space forming a triangular structure, and the resulting three metastable macrostates.
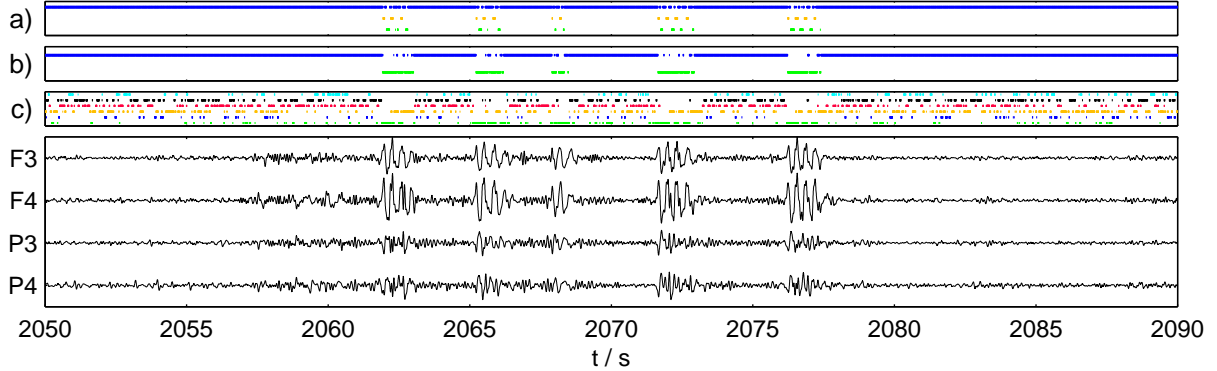
Figure 8: Analysis results illustrated using a segment of 40 s length. Upper panels: Macrostate dynamics over time, resulting from different input state space definitions. a) Original EEG data state space. b) Amplitude vector state space. c) As in (b), but using normalized amplitude vectors. Lower panel: EEG timeseries at four selected recording sites. Paroxysmal episodes are characterized by short bursts of spike-wave activity.
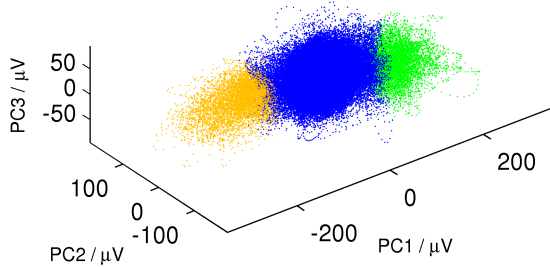

Figure 7: Analysis results for the original EEG data state space: Location of data points belonging to the three identified metastable states. The 19-dimensional state space is represented using the first three PCA components of the data.

only partially successful: The occurrence of states correlates strongly with those features of the underlying process which are psychophysiologically most important, and also most prominent in visual inspection of the data. However, the two states expected are not directly recovered by the EEG analysis. Instead of one persistent state during paroxysmal episodes, we find rapid oscillatory changes between states including the one associated with normal EEG. This indicates that the input state space is not yet optimally defined.

5.2 Amplitude vector state space

This finding can be understood from the fact that electroencephalographic activity in general is so strongly shaped by a predominant oscillatory layer of the dynamics—not only during epileptic episodes but also in normal EEG, particularly in the form of the alpha rhythm—that it is hard to discern more subtle dynamical features. To recover those features, we need a preprocessing step that eliminates the oscillatory character of the data but retains the more slowly changing parameters of the oscillation.

As observed by Wackermann (1994), the trajectory formed by multichannel EEG within the data space can be approximated by a movement along an elliptical orbit with slowly changing orientation and shape (Fig. 9). By locally matching ellipses to the data, a global instantaneous phase
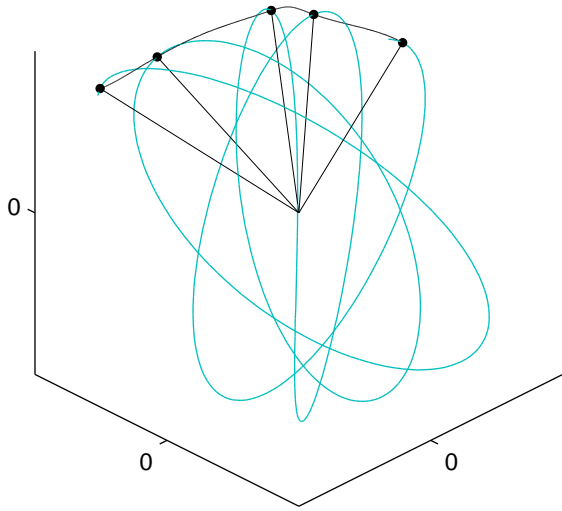
11

Figure 10: Analysis results for the amplitude vector state space. a) Timescale spectrum indicating the presence of two metastable states. b) Microstate positions in one-dimensional eigenvector space and the resulting two metastable macrostates.

Figure 9: Amplitude vector state space. The trajectory corresponding to a multivariate oscillatory signal like EEG takes on the form of an elliptical orbit with slowly varying parameters. Locally matching ellipses to the trajectory, the instantaneously dominant oscillatory component can be characterized by the major semiaxis vectors (straight radial lines), resulting in a description of the system's oscillatory state which itself evolves in a non-oscillatory way (black curve).
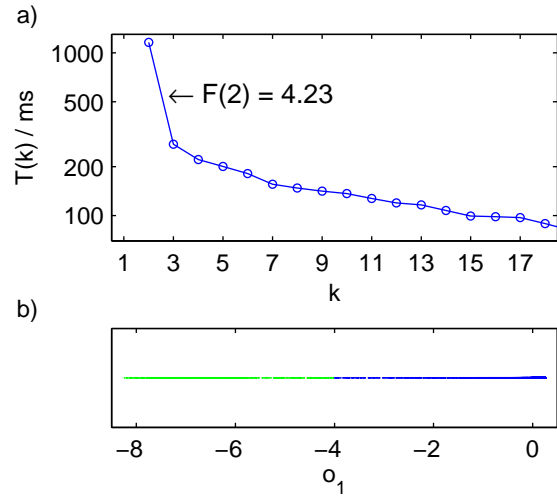
and amplitude can be defined, where the amplitude corresponds to the two main semiaxis vectors of the ellipse. (For a full account of the calculation see App. A.) For simplicity we only use the major semiaxis vector, which specifies the direction and strength of the momentarily dominant oscillatory component, to define an amplitude vector state space as the input state space for the algorithm.[13]

With the specification of the system state via the major amplitude vector an ambiguity arises, because vectors of opposite orientation are equivalent. This is resolved by enforcing positive sign for the first vector component during the assignment of data points to microstates. For visualization (Fig. 11) the axis vectors are used as they come out of the calculation described in App. A, that is with basically random orientation.

The resulting timescale spectrum is shown in Fig. 10a. The largest separation factor of $F(2) = 4.23$ now gives a more definite indication of the number of macrostates than for the original data

---

[13]This approach is similar to one of the strategies employed in the "spatial analysis" of EEG (Lehmann, 1987), to select only those EEG potential maps (data vectors) which occur at local maxima of the "global field strength" (the norm of the data vectors).
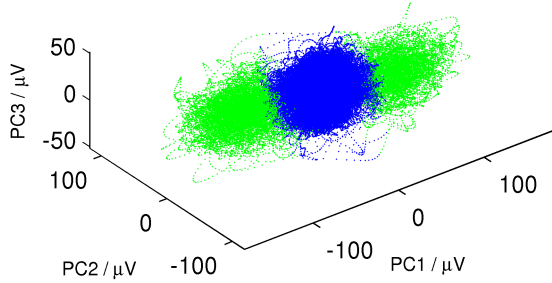
Figure 11: Analysis results for the amplitude vector state space: Location of data points belonging to the two identified metastable states in a representation of the state space using the first three PCA components.

Figure 12: Timescale spectrum obtained using normalized axis vectors. The largest separation factor (for six macrostates) is only marginally larger than the other occurring values, indicating that no adequate definition of metastable states is possible.

state space. In the corresponding one-dimensional eigenvector space (Fig. 10b) the two macrostates are trivially defined by a cut at the center of the range of values.

In Fig. 11, the location of data points belonging to the two macrostates is shown using the first three PCA components of the data points in the amplitude vector state space. Again, a prevalent macrostate (accounting for 98% of the data points) fills a spherically shaped central area, while two appendices protruding on opposite sites together constitute the second macrostate. Despite the fact that the overall shape of the data cloud is similar to that shown in Fig. 7, the reader should keep in mind that the two diagrams depict differently defined state spaces represented with respect to a different set of dimensions.

Fig. 8b demonstrates that the revision of the input state definition successfully eliminates the oscillatory switching between states during paroxysmal episodes. Starting from the amplitude vector input state space, the algorithm for the identification of metastable states is able to consistently associate normal EEG with one macrostate, and—except for short relapses—epileptic EEG with another macrostate.

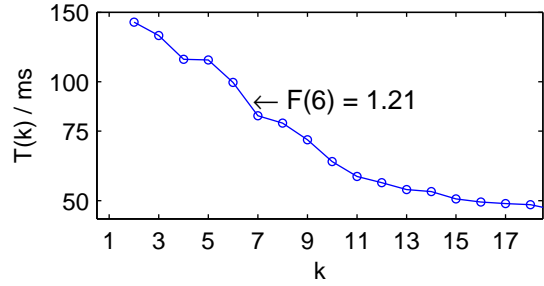The macrostate structure of the amplitude vector state space shown in Fig. 11 suggests that the distinction of the two macrostates relies only on the length of the amplitude vector. To check this, we tested the performance of the algorithm when normalized amplitude vectors are used. The timescale spectrum (Fig. 12), with a maximal separation factor $F(6) = 1.21$ not substantially larger than the rest, indicates that the identification of macrostates is severely impaired under these circumstances. Even so, an examination of the state dynamics over time (Fig. 8c) reveals that there are still two states that are mainly attained during epileptic episodes.

6 Conclusion

Relations between mental (psychological) and neural (physiological) phenomena form the generally accepted basis for work in various disciplines such as psychiatry, psychophysiology, and cognitive neuroscience. While a large body of knowledge has been gathered in these fields, the conceptual question of how mind and brain are related in precise terms is still largely unresolved. Starting from the notion of the mental as "emerging" from neural processes, we argue that this relation of emergence should be understood as one between different descriptions of the same system.

Utilizing concepts from the theory of dynamical systems for the formulation of descriptions, we propose that the relation between descriptive levels should take on the form of a partition or coarse-graining of the state space that is characterized by a preservation of the Markov property. To

13

empirically test the validity of our approach, we turn to a form of such a Markov coarse-graining which can be algorithmically obtained: that of metastable states. We describe how metastable macrostates of a dynamics observed in empirical data can be identified based on the spectral analysis of the transition matrix governing the microstate dynamics, and illustrate its operation with simulation data.

We apply the method to a recording of electroencephalographic (EEG) data from a human subject suffering from petit-mal epilepsy. Combined with a suitable preprocessing procedure, the algorithm is able to automatically identify metastable states from the data which closely correspond to the mental states of the subject (mentally present / absent). This first application substantiates the practical viability of our approach and appears promising for the future application of the method to more challenging forms of data.

Finally we want to point out that the concept of metastable macrostates in the application to EEG data is similar to the notion of "brain functional microstates" introduced by Lehmann and co-workers, which are defined as brief periods of time during which the spatial distribution of the brain's electrical field remains relatively stable (Lehmann, 1971; Lehmann et al., 1987). Transitions between such states are characterized by an abrupt change of the field topography, allowing to decompose the stream of EEG data into segments of the order of magnitude of 10–100 ms duration which can usually be grouped into a small number (< 10) of classes. Note however that the "microstate analysis" of Lehmann et al. results in a coarse-grained description of the brain's electrical activity, i.e. in our nomenclature, a definition of macrostates.

Acknowledgments

A Instantaneous amplitude and phase for multivariate time-series

The local oscillatory behavior of a real-valued univariate signal $x(t)$ is commonly characterized using the corresponding complex-valued analytic signal $z(t)$ (Gabor, 1946). It is obtained by combining $x(t)$ with an imaginary part,

$$z(t) = x(t) + \mathrm{i}\, y(t),$$

which is defined as the Hilbert transform of x,

$$y(t) = \mathrm{H}\, x(t) = \frac{1}{\pi}\, \mathrm{P.V.}\!\int_{-\infty}^{\infty} \frac{x(t')}{t - t'} \mathrm{d}t',$$

where P.V. denotes the Cauchy principal value of the integral. Under the condition that $x(t)$ is dominated by a single frequency component, its instantaneous amplitude $A(t)$ and phase $\phi(t)$ can be determined via the analytic signal according to

$$A(t) = |z(t)|, \quad \phi(t) = \arg z(t),$$

so that

$$x(t) = A(t)\, \cos \phi(t)$$

or

$$z(t) = A(t)\, \exp(\mathrm{i}\, \phi(t)).$$

The terms amplitude A and phase ϕ as they are used here can be interpreted such that they specify the parameters of a strictly periodic sinusoidal oscillation which *locally matches* the behavior of the observed signal $x(t)$ at a given instant t. In particular, $\phi(t)$ attains the value 0 (or equivalently, an integer multiple of 2π) whenever the actual value of $x(t)$ coincides with the associated instantaneous amplitude $A(t)$.

These properties of the analytic signal can also be utilized to determine the parameters of the locally matching oscillation for a multivariate signal $\boldsymbol{x}(t) = (x_i(t))$ ($i = 1 \ldots K$). We assume that each component signal $x_i(t)$ is dominated by a single frequency and that the frequencies of different signals are similar. Using $\boldsymbol{y}(t)$ to denote the channel-wise Hilbert transform of $\boldsymbol{x}(t)$ and $\boldsymbol{z}(t)$ for its channel-wise completion to the analytic signal, the local extension of the signal's oscillatory behavior for instant t is obtained with

$$\boldsymbol{z}_t(\theta) = \boldsymbol{z}(t)\, \exp(\mathrm{i}\, \theta),$$
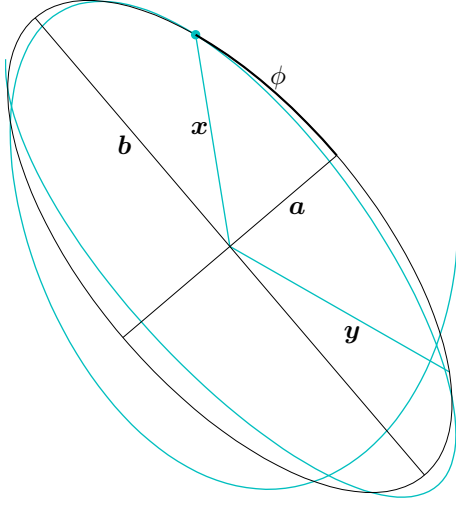
Figure 13: Determination of local ellipse axes. The trajectory formed by the multivariate signal is locally matched to an elliptical orbit, which is defined by the data vector x at a given instant and the corresponding vector y from the signal's channel-wise Hilbert transform as conjugate semiaxis vectors. Main semiaxis vectors of the ellipse, a and b, are obtained using the associated multivariate instantaneous phase ϕ.

parametrized by $\theta \in [0, 2\pi]$. Its real part

$$x_t(\theta) = x(t) \cos \theta - y(t) \sin \theta$$

gives the multivariate oscillation that locally matches the behavior of the signal at instant t; its trajectory is an *elliptical orbit* with conjugate axes specified by the vectors $x(t)$ and $y(t)$.

From these conjugate axes, the main semiaxis vectors $a(t)$ and $b(t)$ of the local ellipse can be calculated (Fig. 13). It proves useful to do so via introducing a global (channel-independent) instantaneous phase $\phi(t)$, such that for $\phi(t) \in \{0, \frac{1}{2}\pi, \pi, \frac{3}{2}\pi\}$ or equivalents, $x(t)$ coincides with one of the main semiaxis vectors or its negative. This is achieved choosing

$$\phi(t) = \frac{1}{2} \arctan \frac{2\, x(t) \cdot y(t)}{|x(t)|^2 - |y(t)|^2}.$$

Since the resulting values in the range $\left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$ cover only one quarter of a cycle, the outcome may be transformed into an equivalent but more useful representation via a standard "unwrapping" procedure (adding or subtracting $\frac{\pi}{2}$ at discontinuity points) to enforce a smooth evolution of $\phi(t)$.

Using this result, main semiaxis vectors of the locally matching ellipse at instant t are obtained by going backwards along $x_t(\theta)$ by an amount of $\phi(t)$ or forwards by $\frac{\pi}{2} - \phi(t)$, i.e.

$$a(t) = x_t\left(-\phi(t)\right)$$

and

$$b(t) = x_t\left(\frac{\pi}{2} - \phi(t)\right).$$

If $\phi(t)$ has been adjusted for a smooth evolution over time, the same can be expected from the resulting $a(t)$ and $b(t)$. It is, however, not clear from this definition which one of these vectors specifies the major and minor axis of the ellipse, respectively, and it is possible that over the course of time the two vectors change roles. For a specific application of this result, further processing may therefore be necessary.

Complementary to the generalization of the instantaneous phase concept, a multivariate instantaneous amplitude $A(t)$ can be defined such that

$$z(t) = A(t) \exp(\mathrm{i}\,\phi(t)),$$

which is given by

$$A(t) = a(t) - \mathrm{i}\, b(t).$$

The channel-wise modulus of this quantity corresponds to the instantaneous amplitudes $A_i(t)$ of the component signals, while the argument comprises the phase differences between the global and the component signal instantaneous phases, $\phi_i(t) - \phi(t)$.

References

R. L. Adler. Symbolic dynamics and Markov partitions. *Bulletin of the American Mathematical Society*, 35(1):1–56, 1998.

C. Allefeld and S. Bialonski. Detecting synchronization clusters in multivariate time series via coarse-graining of Markov chains. *Physical Review E*, 76:066207, 2007.

The American Electroencephalographic Society. Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, 8(2):200–202, 1991.

W. R. Ashby. Principles of the self-organizing system. In H. von Foerster and G. W. Zopf, Jr., editors, *Principles of Self-Organization: Transactions of the University of Illinois Symposium*, pages 255–278. Pergamon, 1962.

H. Atmanspacher and P. beim Graben. Contextual emergence of mental states from neurodynamics. *Chaos and Complexity Letters*, 2(2/3):151–168, 2007.

A. Beckermann, H. Flohr, and J. Kim. *Emergence or Reduction?* de Gruyter, 1992.

R. C. Bishop and H. Atmanspacher. Contextual emergence in the description of properties. *Foundations of Physics*, 36(12):1753–1777, 2006.

E. M. Bollt and J. D. Skufca. Markov partitions. In A. Scott, editor, *Encyclopedia of Nonlinear Science*. Routledge, 2005.

C. D. Broad. *The Mind and Its Place in Nature*. Kegan Paul, 1925.

D. J. Chalmers. What is a neural correlate of consciousness? In Metzinger (2000).

K.-S. Chan and H. Tong. *Chaos: A Statistical Perspective*. Springer, 2001.

V. Darley. Emergent phenomena and complexity. In R. A. Brooks and P. Maes, editors, *Artificial Life IV*, pages 411–416, 1994.

P. Deuflhard and M. Weber. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra and its Applications*, 398:161–184, 2005.

W. Feller. *An Introduction to Probability Theory and Its Applications*, volume I. Wiley, 3rd edition, 1968.

G. Froyland. Statistically optimal almost-invariant sets. *Physica D*, 200:205–219, 2005.

D. Gabor. Theory of communication. *Journal of the Institution of Electrical Engineers. Part 3, Radio and Communication Engineering*, 93:429–457, 1946.

B. Gaveau and L. S. Schulman. Dynamical distance: Coarse grains, pattern recognition, and network analysis. *Bulletin des Sciences Mathématiques*, 129:631–642, 2005.

H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, 1997.

J. Kim. *Supervenience and Mind*. Cambridge University Press, 1993.

D. Lehmann. Multichannel topography of human alpha EEG fields. *Electroencephalography and Clinical Neurophysiology*, 31:439–449, 1971.

D. Lehmann. Principles of spatial analysis. In A. S. Gevins and A. Rémond, editors, *Methods of Analysis of Brain Electrical and Magnetic Signals*, pages 309–354. Elsevier, 1987.

D. Lehmann, H. Ozaki, and I. Pal. EEG alpha map series: Brain micro-states by space-oriented adaptive segmentation. *Electroencephalography and clinical Neurophysiology*, 67:271–288, 1987.

N. Luhmann. *Social Systems*. Stanford University Press, 1996.

H. R. Maturana and F. J. Varela. *Autopoiesis and Cognition: The Realization of the Living*. Reidel, 1980.

T. Metzinger, editor. *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, 2000. MIT Press.

E. Niedermeyer. Abnormal EEG patterns: Epileptic and paroxysmal. In *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*, pages 217–240. Williams & Wilkins, 3rd edition, 1993.

T. O'Connor and H. Y. Wong. Emergent properties. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Winter 2006. http://plato.stanford.edu/archives/win2006/entries/properties-emergent/.

M. Paluš. Nonlinearity in normal human EEG: Cycles, temporal asymmetry, nonstationarity and randomness, not chaos. *Biological Cybernetics*, 75(5):389–396, 1996.

C. Robinson. *Dynamical Systems. Stability, Symbolic Dynamics, and Chaos*. CRC Press, 1995.

A. K. Seth. Measuring emergence via nonlinear Granger causality. In S. Bullock et al., editors,

Artificial Life XI: Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems, pages 545–553. MIT Press, 2008.

C. R. Shalizi and C. Moore. What is a macrostate? Subjective observations and objective dynamics. preprint, 2008. `arXiv:cond-mat/0303625`.

A. Stephan. Emergentism, irreducibility, and downward causation. *Grazer Philosophische Studien*, 65:77–93, 2002.

F. Takens. Detecting strange attractors in turbulence. In D. A. Rand and L. S. Young, editors, *Dynamical Systems and Turbulence*, pages 366–381. Springer, 1981.

J. Theiler and E. Rapp. Re-examination of the evidence for low-dimensional, nonlinear structure in the human electroencephalogram. *Electroencephalography and Clinical Neurophysiology*, 98(3): 213–222, 1996.

F. J. Varela. *Principles of Biological Autonomy*. North Holland, 1979.

J. Wackermann. Segmentation of EEG map series in n-dimensional state space. *Brain Topography*, 6(3):246, 1994.