Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica D 177 (2003) 203–232
PHYSICA D
www.elsevier.com/locate/physd

Analysis and modeling of quasi-stationary multivariate time series and their application to middle latency auditory evoked potentials

A. Hutt [a,*], H. Riedel [b]

[a] *Max Planck-Institute of Mathematics in the Sciences, Inselstr. 22-26, 04103 Leipzig, Germany*
[b] *Medical Physics Group, Carl von Ossietzky University of Oldenburg, 26111 Oldenburg, Germany*

Received 1 August 2001; received in revised form 13 March 2002; accepted 24 May 2002
Communicated by E. Bodenschatz

Abstract

A methodological framework for analyzing and modeling of multivariate data is introduced. In a first step, a cluster method extracts data segments of quasi-stationary states. A novel cluster criterion for segment borders is introduced, which is independent of the number of clusters. Its assessment reveals additional robustness towards initial conditions. A subsequent dynamical systems based modeling (DSBM) approach focuses on data segments and fits low-dimensional dynamical systems for each segment. Applications to middle latent auditory evoked potentials yield data segments, which are equivalent to well-known waves from electroencephalography studies. Focussing to wave P_a, two-dimensional dynamical systems with common topological properties are extracted. These findings reveal the common underlying dynamics of P_a and indicate self-organized brain activity.
© 2002 Elsevier Science B.V. All rights reserved.

PACS: 02.50.Sk; 05.45

Keywords: Multivariate analysis; Cluster analysis; Dynamical system based modeling; Evoked potentials

1. Introduction

Multivariate time series are measured in various research fields, ranging from meteorology and geophysics at macroscopic levels to micro-electrode measurements in brain tissue [1]. Measurements in spatially extended systems frequently acquire data from detectors distributed over space. In neuroscience, there are several different approaches to yield information about spatio-temporal activity of the brain. Apart from invasive measurements, which obtain neural activity from brain tissue directly, non-invasive methods are widely applied. We mention the functional magnetic resonance interface (fMRI), positron-emission tomography (PET) and electro- and magnetoencephalogram (EEG/MEG). While the former methods yield high-resolution measurements in space with poor resolution in time,

* Corresponding author.
E-mail address: ahutt@mis.mpg.de (A. Hutt).

0167-2789/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0167-2789(02)00747-9

EEG/MEG, in general, obtains effective electromagnetic potentials and fields on the scalp with good temporal resolution. Since scalp measurements allow only a rough spatial localization, EEG/MEG are mainly applied for investigations of temporal processes [2]. Additional source localization techniques as computation of dipoles or current source densities [3] improve the spatial resolution, which however remains below the resolutions of fMRI or PET.

Data analysis techniques for EEG are based frequently on assumptions about statistical properties of signals or underlying dynamics. Especially models for multivariate signals lead to improved analysis and a better understanding of underlying processes. There are two main modeling approaches in neuroscience [4], which are reflected in the presented work and discussed in the following.

The top-down approach aims at extracting models for brain activity from measured data [5]. For instance, it comprises methods attacking ill-posed inverse problems and modeling of multivariate frequency spectra. Segmentation methods have been developed in the last two decades, which detect quasi-stationary spatio-temporal states in EEG [6–8]. These states are supposed to represent coherent neural activities and are called microstates [9]. The work of Lehmann and Skrandies [7] represents one of the first methods in this field. In the following, we introduce the method in some detail. Starting from the assumption, that coherent neural activity is reflected by global maximum power of the EEG, the global field power (GFP) is computed as the root mean square of the potential deviations from the spatial average. This definition makes the GFP independent of the reference electrode. Since coherent states are observed quasi-stationary in time, transitions between states show changing spatial patterns at short time scales. This means that borders of coherent states are marked by an increased dissimilarity of consecutive spatial activities. This aspect leads to the definition of a global dissimilarity (GD). It is computed as the GFP of potential differences of consecutive intensity maps, which are normalized with respect to their GFP [10]. Plotting GFP and GD with respect to time, peaks of the GD coincide with troughs of the GFP. The corresponding time points mark borders of coherent states, i.e. global dissimilarities of intensity maps show least global intensity of the EEG. These segmented time windows are called microstates and show one prominent peak of GFP. Comparisons of microstates in ERP-data with well-known cognitive components show good accordance [6]. An extension of the method has been developed by Pascual-Marqui et al. [11] by introducing a clustering approach. Here, clusters in multivariate signals are detected by the K-Means algorithm, while the number of clusters is estimated by cross-validation. The present work aims to extend these methods by two additional features. We derive the observed segment structure of the EEG by notions of complex systems and self-organization. Furthermore, we introduce a novel clustering approach, which moves the problem of the right number of clusters to a simpler statistical problem.

The bottom-up approach considers microscopic properties of the investigated system, whose dynamical relations serve as basic model equations [12]. An important concept within this approach is the idea of interacting modes. It is basic in physics of complex systems [13] and has been applied in several works in neuroscience [14–16]. The general concept introduces few interacting modes, which suffice for the description of the systems dynamics. In case of spatially extended systems, spatio-temporal dynamics factorize to few spatial patterns evolving in time with projected amplitudes. This approach has been extended by Uhl et al. [18] by a synchronous fit of optimal projection modes and corresponding differential equations from data. Applications to petitmal epilepsy data [19] and ERP-data [16] identified Shilnikov-attractors and cognitive components, respectively. A recent study of the method shows an improved dipole localization from simulated EEG [20].

The latter dynamical systems based modeling (DSBM) [20] bridges top-down and bottom-up approaches by extracting few interacting modes and a corresponding dynamical system from data. In [16], a two-dimensional model has been extracted from multivariate ERP-data, and fixed points of fitted dynamical systems have been localized in time windows of cognitive components. Additionally, transitions between cognitive components appeared evolving in higher dimensions than the determined two and at smaller time scales. We find that cognitive components represent low-dimensional attractive points in signal space, which are approached and left at changed time scales. We also argue that low-dimensional global models for non-stationary multivariate time series in general do not

catch all local dynamic features, although they might represent optimal models for the average behavior of a signal. Improved models are aimed to describe local behavior by separated models. We introduce an adaptive approach, which replaces a global dynamical model by a sequence of local models separated by transition parts. The chain of models is closed by attaching the single models by Poincaré mappings. However, this model approach is constraint to signals, whose segments show quasi-stability and coherent multivariate behavior.

The present work introduces a methodological framework for modeling quasi-stationary multivariate time series. We introduce a clustering method for data segmentation, which has been developed in previous works [21,22]. The current work introduces the derivation of a novel cluster criterion, which simplifies the problem of optimal cluster numbers. In a second part, we introduce DSBM. Since previous numerical implementations are quite time-consuming for higher dimensions, an analytical derivation of optimal projection modes and corresponding polynomial ordinary differential equations is presented [23]. In practice, the optimal choice of polynomial models depends strongly on properties of the investigated data, which are biased by additional pre-processing steps as lowpass filtering. These aspects are essential for the introduced framework and are examined additionally. Furthermore, we discuss the influence of the applied reference electrode to segmentation and modeling results. First applications to simulated data evaluates derived features of the clustering approach. Applications of clustering and DSBM to middle latent auditory evoked potential (MAEP) data follow. In the time window of wave P_a, we observe similar topologies of determined dynamical systems for three different data sets from different subjects, each with two different intrinsic noise levels. The paper is organized as follows. In Section 2, we introduce the clustering method and DSBM, followed by considerations of model estimations and the influence of re-referencing data. Section 3 contains clustering results from simulated data and measured MAEP-data. Results are discussed in Section 3.3, which is followed by concluding remarks.

2. Methods

2.1. Clustering approach

As mentioned in the previous section, components in event-related potentials exhibit low-dimensional behavior, which is similar to dynamics near fixed points. Following this idea, we assume a signal trajectory in multi-dimensional space, which shows a sequence of segments governed by saddle point dynamics. Data points accumulate close to the fixed points in case of constant sampling rate. Since these regions of increased data density also represent clusters in data space, we treat their detection as a recognition problem in data space [21,22,24]. In the present paper, the K-Means algorithm [25] is applied for cluster detection, but any other unsupervised clustering algorithm is possible.

2.1.1. The clustering method

Let there be an N-dimensional data set $\{\mathbf{q}_i \in \mathfrak{R}^N, i = 1, \ldots, T\}$ containing T data points. A clustering algorithm aims at cluster centers $\{\mathbf{k}_k \in \mathfrak{R}^N, k = 1, \ldots, K\}$ in data space, whose average Euclidean distance to a set of data points is minimized. Here, K denotes the number of detected clusters, which is set a priorily. As a first step, initial cluster centers $\mathbf{k}_k^{(0)}$ are chosen randomly from the set of data points and their Euclidean distances to each data point are computed. The K-Means algorithm defines memberships of data points to a cluster by the smallest Euclidean distance to its center. Thus, data are segmented into K clusters and new cluster centers $\mathbf{k}_k^{(1)}$ are computed as averages of clustered data points. Euclidean distances between data points and centers $\mathbf{k}_k^{(n)}$ are re-estimated and convergence is observed typically after 20 iterations. In the next step, Euclidean distances $d(k, i) = \|\mathbf{k}_k - \mathbf{q}_i\|$ from cluster centers k to data points i are computed [22] and a map

$$\phi : i \to \mathrm{nc}(i) = \min_{2<j\leq k} d(j, i) \quad \forall 1 \leq i \leq T,$$

associates each data point i to its nearest cluster $k = \mathrm{nc}(i)$. Since data points are ordered in time, the inverse map ϕ^{-1} defines time windows, in which the signal approaches a cluster. While the map ϕ is unique, ϕ^{-1} is not

$$\phi^{-1} : j \to \{i|\mathrm{nc}(i) = j\} = \bigcup_{1\le\alpha\le n_j} \mathcal{M}_{\alpha j} \quad \forall 1 \le j \le k,$$

where $\mathcal{M}_{\alpha j} = \{i_{\alpha j}, i_{\alpha j}+1, \ldots, i_{\alpha j}+n_{\alpha j}\}$ are n_j subsets of successive time points. That is, $\mathcal{M}_{\alpha j}$ represents a time window, in which the signal approaches cluster j. These temporal partitions represent the obtained high-dimensional clustering results.

2.1.2. Cluster criterion

Now, the degree of separation of neighboring clusters is quantified by an additional map

$$\theta : i \to \mathrm{snc}(i) = \min_{2<j\le k,\, j\ne \mathrm{nc}(i)} d(j,i) \quad \forall 1 \le i \le T,$$

which associates each data point with its second nearest cluster $\mathrm{snc}(i)$. Let $\mathcal{M}_{\alpha(i)}$ denote the subset that contains data point i. Then

$$A_K(i) = \frac{1}{\mathcal{N}} \sum_{i_{\alpha j}\in\mathcal{M}_{\alpha(i)}} d(\mathrm{snc}(i_{\alpha j})) - d(\mathrm{nc}(i_{\alpha j}), i_{\alpha j}), \tag{1}$$

represents the percental contribution of cluster $\mathrm{nc}(i)$ to the whole signal for K clusters [21] and $\mathcal{N}$ abbreviates the norm $\sum_{i=1}^{T} A_K(i)$.

Eq. (1) reflects the spatial configuration of data point i and its nearest neighbors $i_{\alpha j}$. Data points $i_{\alpha j}$ building a well-separated cluster show large $A_K(i)$, while small $A_K(i)$ reflect rather unstructured data. Additionally, Eq. (1) considers effects caused by different numbers of cluster members: contributions of few far-distance outliers are small, while highly-populated clusters exhibit large values of $A_K(i)$.

Finally, percental contributions are averaged over increasing number of clusters K

$$\Pi : i \to p(i) = \frac{1}{R-2} \sum_{k=2}^{R} A_k(i). \tag{2}$$

This cluster quality measure collects cluster results for $R-2$ different number of clusters and is regarded independent of the number of clusters. It represents an averaged probability distribution and expresses how likely it is that a data point i is member of a cluster. Good convergence is achieved for $R \approx 30$.

An additional hard problem in cluster analysis is the dependence on initial clusters and the question whether a global or local minimum is reached after the iterations. A number L of different initial cluster configurations lead to L measures $p(i)$ and an averaged cluster quality measure and corresponding standard deviation are determined by

$$\bar{p}(i) = \frac{1}{L}\sum_{l=1}^{L} p_l(i), \qquad \Delta p(i) = \sqrt{\frac{1}{L-1}\sum_{l=1}^{L}(p_l(i) - \bar{p}(i))^2},$$

where $p_l(i)$ denote the cluster quality measures for initial condition l. Since first computations of histograms of $p_l(i)$ show no standard distributions as normal or identical distribution, further statistical tests are not applied.

2.2. Dynamical systems based modeling

The proposed framework aims at modeling the dynamics of detected signal partitions by low-dimensional systems $\dot{\mathbf{y}} = \mathbf{f}[\mathbf{y}]$ of ordinary differential equations. Thereby, the amplitudes y_i determine the evolution of corresponding static multivariate modes. The method also seeks to extend principal component analysis (PCA) by a synchronous

fit of dynamical systems [18,23]. Subsequently, we choose a cost function

$$V = V_{\mathrm{PCA}} + \epsilon \cdot V_d(\mathbf{f}), \quad \epsilon \in \Re_0^+.$$

In this ansatz, V_{PCA} represents the cost function for the derivation of static modes, while V_d yields an optimal fit of a dynamical system from data. Increasing the weighting factor ϵ, contributions by the dynamical system also increases. For $\epsilon = 0$, obtained modes are equal to PCA-modes.

2.2.1. *Cost function V_{PCA}*

An N-dimensional signal $\mathbf{q}(t) \in \Re^N$ with $\langle \mathbf{q}(t) \rangle = 0$ can be described as a superposition of static modes $\mathbf{v}_i$ evolving in time with amplitudes $x_i(t)$

$$\mathbf{q}(t) = \sum_{i=1}^{N} x_i(t)\mathbf{v}_i. \tag{3}$$

In the following, $\langle \cdots \rangle$ denotes time average. In case of PCA, modes and amplitudes are chosen orthogonal leading to the well-known relations

$$\mathbf{C} \cdot \mathbf{v}_i = \lambda_i \mathbf{v}_i, \qquad \langle x_i(t) x_j(t) \rangle = \lambda_i \delta_{ij} \tag{4}$$

with $\mathbf{C} = \langle \mathbf{q} \otimes \mathbf{q} \rangle / \langle \mathbf{q}^2 \rangle$, the symbol $\otimes$ denotes the outer product and δ_{ij} represents the Kronecker symbol. From (4), it follows $\sum_{i=1}^{N} \lambda_i = 1$. The error E made by reconstructing the signal by $M \le N$ modes is

$$E = 1 - \sum_{i=1}^{M} \lambda_i = 1 - \sum_{i=1}^{M} \frac{\langle (\mathbf{q} \cdot \mathbf{v}_i)^2 \rangle}{\langle \mathbf{q}^2 \rangle}. \tag{5}$$

This relation is widely used for dimensionality reduction (see, e.g. [14,26,27]). With Eq. (5) and $\mathbf{v}_i^2 = 1$, it is

$$\begin{aligned} E &= 1 - \sum_{i=1}^{M} \frac{\langle (\mathbf{q} \cdot \mathbf{v}_i)^2 \rangle}{\langle \mathbf{q}^2 \rangle} = \frac{\langle \mathbf{q}^2 - 2\sum_{i=1}^{M} (\mathbf{q} \cdot \mathbf{v}_i)^2 + \sum_{i=1}^{M} (\mathbf{q} \cdot \mathbf{v}_i)^2 \mathbf{v}_i^2 \rangle}{\langle \mathbf{q}^2 \rangle} \\ &= \sum_{i=1}^{M} \frac{\langle \mathbf{q}^2 - 2(\mathbf{q} \cdot \mathbf{v}_i)^2 + (\mathbf{q} \cdot \mathbf{v}_i)^2 \mathbf{v}_i^2 \rangle}{\langle \mathbf{q}^2 \rangle} - (M-1) = \sum_{i=1}^{M} \frac{\langle (\mathbf{q} - (\mathbf{q} \cdot \mathbf{v}_i)\mathbf{v}_i)^2 \rangle}{\langle \mathbf{q}^2 \rangle} - (M-1). \end{aligned}$$

Since variations of error function E are invariant under constant shifts, we obtain [23]

$$V_{\mathrm{PCA}} = \sum_{i=1}^{M} \frac{\langle (\mathbf{q} - (\mathbf{q} \cdot \mathbf{v}_i)\mathbf{v}_i)^2 \rangle}{\langle \mathbf{q}^2 \rangle}.$$

This special cost function breaks the degeneration of the solution space for orthogonal modes. Vanishing variations of V_{PCA} with respect to $\{\mathbf{v}_i\}$ lead directly to Eq. (4).

2.2.2. *Cost function V_d*

Since static modes evolve in time with amplitudes $x_i(t)$, multivariate dynamics obeys ordinary differential equations

$$\dot{x}_i(t) = f_i[x_j(t)], \tag{6}$$

$$\dot{x}_i(t) = \sum_{j=1}^{M} \Gamma_{ij}^1 x_j + \sum_{j=1}^{M} \sum_{k=1}^{j} \Gamma_{ijk}^2 x_j x_k + \sum_{j=1}^{M} \sum_{k=1}^{j} \sum_{l=1}^{k} \Gamma_{ijkl}^3 x_j x_k x_l + \cdots \tag{7}$$

for M nonlinear interacting modes, which are equivalent to a linear regression model

$$\dot{x}_i(t) = \sum_\alpha a_{i\alpha}\xi_\alpha(t).$$

Here, nonlinear terms are collected in a time-dependent vector

$$\{\xi_\alpha\} = \{x_1, x_2, \ldots, x_M, x_1^2, x_1x_2, \ldots, x_M^2, x_1^3, x_1^2x_2, \ldots, x_M^3\},$$

and model coefficients $a_{i\alpha}$ are determined by minimizing

$$V_d = \frac{1}{M}\sum_{i=1}^{M}\frac{\langle(\dot{x}_i - \sum_\alpha a_{i\alpha}\xi_\alpha)^2\rangle}{\langle\dot{x}_1^2\rangle}.$$

The denominator scales the temporal derivatives and reduces effects caused by different amplitude magnitudes.

2.2.3. Final cost function, V

The orthogonality of PCA-modes limits an optimal fit to data structures. This severe constraint is abolished by introducing biorthogonal modes [18]

$$\mathbf{w}_i^\dagger \cdot \mathbf{w}_j = \delta_{ij}, \quad i = 1, \ldots, M, \tag{8}$$

which leads to new definitions of amplitudes and cost functions. Introducing this relation and $\mathbf{w}_i^2 = 1$ by Lagrange multipliers, the final cost function reads

$$V = \sum_{i=1}^{M}\frac{\langle(\mathbf{q} - (\mathbf{q}\cdot\mathbf{w}_i^\dagger)\mathbf{w}_i)^2\rangle}{\langle\mathbf{q}^2\rangle} + \epsilon\cdot V_d(\dot{y}_j, y_k, a_{l\gamma}) + \sum_{i,j=1}^{M}\tau_{ij}(\mathbf{w}_i^\dagger\cdot\mathbf{w}_j - \delta_{ij}) + \sum_{i=1}^{M}\alpha_i(\mathbf{w}_i^2 - 1) \tag{9}$$

with

$$V_d = \frac{1}{M}\sum_{i=1}^{M}\frac{\langle(\dot{y}_i - \sum_\alpha a_{i\alpha}\xi_\alpha(y_j))^2\rangle}{\langle\dot{y}_i^2\rangle}, \quad y_i = \frac{\mathbf{q}\cdot\mathbf{w}_i^\dagger}{\sqrt{\langle\mathbf{q}^2\rangle}}.$$

For $M < N$, low-dimensional descriptions of the signal are gained. Minimizing (9) with respect to $\{\mathbf{w}_i\}$, $\{\mathbf{w}_i^\dagger\}$ leads to $2M$ coupled nonlinear vector equations [23]. Since these equations cannot be solved rigorously, a perturbational approach is applied with the following properties:

- For $\epsilon = 0$, V describes PCA. Biorthogonal modes $\{\mathbf{w}_k^\dagger\}$, $\{\mathbf{w}_k\}$ reduce to orthogonal PCA-modes $\{\mathbf{v}_k\}$, no dynamical system is fit and the Lagrange multipliers $\{\tau_{kl}\}$, $\{\alpha_{kl}\}$ vanish.
- For $\epsilon > 0$, minimizing V leads to biorthogonal spatial modes $\{\mathbf{w}_k^\dagger\}$, $\{\mathbf{w}_k\}$, a dynamical system $\mathbf{f}[\mathbf{a}]$ and non-vanishing Lagrange multipliers $\{\tau_{kl}\}$, $\{\alpha_{kl}\}$.
- Increasing ϵ from zero, the biorthogonal basis grows from the orthogonal basis, angles between basis vectors change from $\pi/2$ to larger or smaller values.

2.2.4. Derivation of static modes and the dynamical system

Similar to perturbation theory in non-relativistic quantum mechanics, modes and Lagrange multipliers are expanded in perturbation series in ϵ and spatial modes $\{\mathbf{w}_k^\dagger\}$, $\{\mathbf{w}_k\}$ are set as superpositions of orthogonal modes, here chosen as PCA-modes

$$\mathbf{w}_k^\dagger = \sum_{j=0}^{\infty}\sum_{l=1}^{N} \epsilon^j c_{kl}^{(j)} \mathbf{v}_l, \qquad \mathbf{w}_k = \sum_{j=0}^{\infty}\sum_{l=1}^{N} \epsilon^j d_{kl}^{(j)} \mathbf{v}_l, \quad k = 1, \ldots, M.$$

The expansion coefficients $c_{kl}^{(n)}$, $d_{kl}^{(n)}$ are non-quadratic matrices and represent the contribution of PCA-mode l to new mode k. In first perturbation order, definition (8) and the normalization constraints for $\{\mathbf{w}_k\}$ lead to vanishing self-contributions $c_{kk}^{(1)} = d_{kk}^{(1)} = 0$. In case $k \neq m$, expansion coefficients for $k, m \in [1, \ldots, M]$ are

$$c_{km}^{(1)} = \frac{-(1/2)(\lambda_m + \lambda_k)(\partial V_d/\partial \mathbf{w}_m^\dagger)|_0 \cdot \mathbf{v}_k + \lambda_k (\partial V_d/\partial \mathbf{w}_k^\dagger)|_0 \cdot \mathbf{v}_m}{(\lambda_k - \lambda_m)^2}, \qquad d_{km}^{(1)} = -c_{mk}^{(1)}, \tag{10}$$

and for $k \in [1, \ldots, M]$, $m \in [M+1, \ldots, N]$

$$c_{km}^{(1)} = -\frac{1}{2(\lambda_k - \lambda_m)} \left.\frac{\partial V_d}{\partial \mathbf{w}_k^\dagger}\right|_0 \cdot \mathbf{v}_m, \qquad d_{km}^{(1)} = \left(\frac{\lambda_m}{\lambda_k}\right) c_{km}^{(1)}. \tag{11}$$

The terms $(\partial V_d/\partial \mathbf{w}_m^\dagger)|_0$ represent the influence of the dynamics fit in absence of a perturbation and reads explicitly

$$\left.\frac{\partial V_d}{\partial \mathbf{w}_k^\dagger}\right|_0 = -2\frac{\langle \dot{\mathbf{q}} f_k^{(0)}\rangle}{\langle \dot{x}_k^2\rangle} + 2\frac{\langle \dot{\mathbf{q}}\dot{x}_k\rangle\langle (2\dot{x}_k - f_k^{(0)}) f_k^{(0)}\rangle}{\langle \dot{x}_k^2\rangle^2}$$

with

$$f_i^{(0)} = \sum_\alpha a_{i\alpha}^{(0)} \xi_\alpha^{(0)}, \qquad \left.\frac{\partial f_i}{\partial \mathbf{w}_k^\dagger}\right|_0 = \sum_\alpha a_{i\alpha}^{(0)} \left.\frac{\partial \xi_\alpha}{\partial \mathbf{w}_k^\dagger}\right|_0, \qquad \xi_\beta = y_r: \quad \left.\frac{\partial \xi_\beta}{\partial \mathbf{w}_k^\dagger}\right|_0 = \mathbf{q}\delta_{rk},$$

$$\xi_\beta = y_r y_s, \quad r \le s: \quad \left.\frac{\partial \xi_\beta}{\partial \mathbf{w}_k^\dagger}\right|_0 = \mathbf{q}(\delta_{kr} x_s + \delta_{ks} x_r),$$

$$\xi_\beta = y_r y_s y_t, \quad r \le s \le t: \quad \left.\frac{\partial \xi_\beta}{\partial \mathbf{w}_k^\dagger}\right|_0 = \mathbf{q}(\delta_{kr} x_s x_t + \delta_{ks} x_r x_t + \delta_{kt} x_r x_s).$$

Additional variations with respect to the parameter matrix $\mathbf{a}$ yield

$$\mathbf{a}^{(0)} = \langle \dot{\mathbf{x}} \otimes \xi(x_j)\rangle \cdot \langle \xi(x_j) \otimes \xi(x_j)\rangle^{-1}.$$

Eqs. (10) and (11) implicate the non-degenerative case with $\lambda_k \neq \lambda_m$. In case of large k and m, PCA-eigenvalues are very small, $\lambda_m - \lambda_k \to 0$ and degeneration emerges [24]. However, it is not necessary to discuss this case for small k and $\lambda_k - \lambda_m \gg 0$, i.e. for low-dimensional descriptions. Final projection modes are found by substituting (10) and (11) into

$$\mathbf{w}_k^\dagger = \mathbf{v}_k + \epsilon \cdot \sum_{l=1, l\neq k}^{N} c_{kl}^{(1)} \mathbf{v}_l. \tag{12}$$

The corresponding dynamical system is determined by

$$A_{kl} = \langle \dot{\mathbf{y}} \otimes \xi(y_i)\rangle \langle \xi(y_i) \otimes \xi(y_i)\rangle^{-1}, \qquad y_k(t) = x_k(t) + \epsilon \cdot \sum_{l=1}^{N} c_{kl}^{(1)} x_l(t). \tag{13}$$

Now, minimal errors V_d with respect to weights ϵ define optimal projections and dynamics [23].

2.3. Model estimation and corresponding problems

Eq. (7) represents a multivariate polynomial expansion at point $\mathbf{x} = 0$. It converges to measured data in finite order, for sufficiently small magnitude of $x_i(t)$. Therefore, the magnitude of amplitudes determines the least polynomial order. While it is known that increasing the order reduces estimation errors, in many cases one is interested in an optimal polynomial order. We would like to mention cross-validation methods and information criteria [28] in this context. Since the presented work shows low estimation errors for cubic polynomials and drastic error increases for lower orders, polynomials are fixed to cubic order in the following applications.

Let us consider an N-dimensional model of polynomial order M. Its number of parameters is given to $p \cdot N$ with

$$p = N + \frac{1}{2!}N(N+1) + \cdots + \frac{1}{M!}N(N+1)\cdots(N+M-1).$$

When this number exceeds the number of data points, the solution space of model parameters is degenerated and solutions are ambiguous. For exceeding numbers of data points, regression methods yield unique data models. Thus, there is a least number of data points for a reliable model of fixed order. For instance, two-dimensional cubic models necessitate at least 18 data points. Summarizing, both the number and the magnitude of data points decide whether data can be modeled by polynomials. For sufficiently large data sets, as EEG obtained from sleep studies or motor-behavior studies, no such problems occur, while modeling of data in short time windows have to take care about these aspects. Examples are event-related components or early evoked potentials, which contain usually only tens of data points.

Increased sampling rates at experimental data acquisition and a subsequent application of lowpass filters support modeling of short time windows. However, lowpass filtering generates artificial correlations between data points and bias extracted models. According to the sampling theorem [29]

$$n_{\text{thr}} = \frac{\nu_s}{2\nu_f}$$

is the number of points that are sufficient to resolve oscillations with signal frequency ν_f at a sampling rate ν_s. Since n_{thr} also corresponds to the least lag of vanishing autocorrelation of the investigated signal, data points with lag of n_{thr} points are statistical independent. We point out that statistical dependencies do not remove the reliability of dynamical models, but might questions common properties of obtained models. This aspect arises in Section 3.2.

2.4. Re-referencing to average reference

Several works argue that the choice of the reference electrode of zero-potential influences coherence and dipole calculations [7,30]. In general, these effects are reduced off-line by re-referencing the data to an artificial reference. Here, we present the widely used referencing to the spatial average

$$\tilde{q}_{N+1}(t) = 0, \qquad \tilde{q}_i(t) = q_i(t) - \frac{1}{N+1}\sum_{j=1}^{N+1} q_j(t), \quad i = 1, \ldots, N.$$

That is, the new signal $\tilde{\mathbf{q}}$ is obtained by an additional time variation for N channels. Formulating temporal dynamics by autonomous differential equations $\dot{q}_i(t) = f(q_j)$, the correction emerges as an additional global driving force $F(t)$

$$\dot{\tilde{q}}_i(t) = \tilde{f}(\tilde{q}_i, t) - F(t), \quad F(t) = \frac{1}{N+1}\sum_{j=1}^{N+1} \dot{q}_j(t) \tag{14}$$

in a non-autonomous differential equation system. For signals with large time scales and spatial distributions of $\dot{\mathbf{q}}$ with vanishing mean effects to the dynamical behavior will be small. Subsequently, contributions of average re-referencing are large in transient signal parts, while dynamics $\dot{\mathbf{q}}$ of functional components and clusters evolving at large time scales are hardly influenced.

3. Applications

3.1. Simulated data

In this section, features of the introduced cluster criterion are computed by application to artificial multivariate data. Cluster centers, Euclidean distances to data and percental contributions $A_K(i)$ are computed for K clusters and two different time windows. Beyond previous works [21,24,31], a statistical assessment of cluster quality measures $p(i)$ is applied for varying parameters. The investigated data set $\mathbf{q}(t)$ represents a superposition of three interacting modes

$$\mathbf{q}(t) = x(t)\mathbf{v}_x + y(t)\mathbf{v}_y + z(t)\mathbf{v}_z,$$

where amplitudes $x(t)$, $y(t)$, $z(t)$ obey a three-dimensional dynamical system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = x - x(x^2 + 4y^2) + \Gamma(t), \quad \frac{\mathrm{d}y}{\mathrm{d}t} = y - y(y^2 + 4z^2) + \Gamma(t), \quad \frac{\mathrm{d}z}{\mathrm{d}t} = z - z(z^2 + 4x^2) + \Gamma(t) \tag{15}$$

with identically-distributed noise $\Gamma(t) \in [-0.1; 0.1]$. Dynamics described by Eq. (15) arises in various physical systems, e.g. in rotating fluids at large Taylor numbers [32]. In the current context, spatial modes $\mathbf{v}_x$, $\mathbf{v}_y$, $\mathbf{v}_z$ represent artificial (75×75)-patterns (Fig. 1), i.e., the signal $\mathbf{q}(t)$ lives in a 5625-dimensional space. It is generated by 2200 integration steps with initial conditions $(x(0), y(0), z(0))^{\mathrm{t}} = (0.03, 0.2, 0.8)^{\mathrm{t}}$ and its trajectory passes saddle points at $\mathbf{x}_3 = (0, 0, 1)^{\mathrm{t}}$, $\mathbf{x}_1 = (1, 0, 0)^{\mathrm{t}}$, $\mathbf{x}_2 = (0, 1, 0)^{\mathrm{t}}$ and $\mathbf{x}_3 = (0, 0, 1)^{\mathrm{t}}$ in this sequence. Fig. 2 shows a sampled time series of $\mathbf{q}(t)$.

3.1.1. Clustering approach

In the following, we investigate cluster results for various number of clusters. Fig. 3(a) and (b) shows Euclidean distances $d(k, i)$ and percental contributions $A_K(i)$ for the full time window and number of clusters $K = 2, 3, 5$ and 15. For $K = 2$, data points $i \in [0; 330]$ show smaller Euclidean distances to cluster 1 (solid line) than to cluster 2 (dashed line). This means that they are located near cluster 1. At $i = 330$, data move from cluster 1 to cluster 2, leaves cluster 2 again at $i = 1100$ and returns to cluster 1. An additional cluster ($K = 3$) yields three cluster transitions at $i = 330$, 1050 and 1600. Here, cluster 3 (dotted line), cluster 2 (dashed line), cluster 1 (solid line) and cluster 3 are approached in this sequence. For $K = 4$, only two approached clusters (solid and dashed lines) are detected, while two others are located too far away from data. Further increases of K yield three transitions between

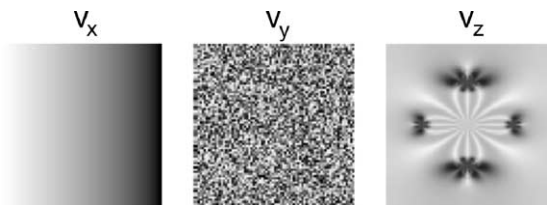

Fig. 1. Spatial modes of simulated data. Pattern v_z is adapted from Thaler [37] by permission.

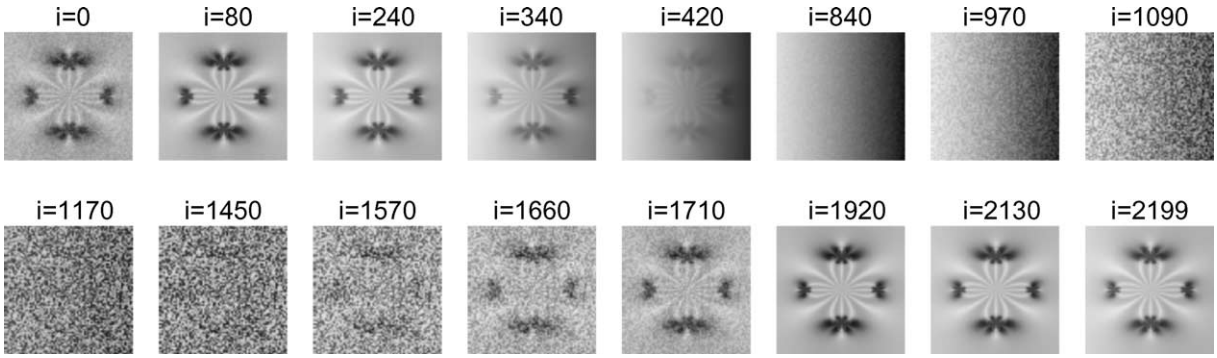

Fig. 2. Sampled time series of spatial maps of simulated data. Quasi-stationary states emerge at $i \approx 240$, 840, 1450 and 1920.

different clusters at similar time points (see plots for $K = 5$ and 15). In the following, we refer to time windows between cluster transitions as clustered windows.

Eq. (1) introduces subsets $\mathcal{M}_\alpha$ and percental contributions $A_K(i_j)$ of clustered windows. They are illustrated for $K = 2$ with $i_1 \in [0; 330]$, $i_2 \in [330; 1100]$ and $i_3 \in [1100; 2199]$ in Fig. 3(a) and plotted for different K in Fig. 3(b). Finally, computed quality measures $p(i)$ show plateaus, troughs and steep rises (Fig. 3(c)) with clusters in the same time windows. Large values of $p(i)$ originate from large values of $A_K(i)$ and indicate clustered windows, whereas rapid changes suppose points with changing cluster memberships for different K. Thus, plateaus represent clustered time windows, while troughs and rises mark their upper and lower borders, respectively.

Since spatio-temporal clusters are specified in temporal and spatial domains, averages of data in clustered windows yield cluster centers, i.e. (75×75)-patterns. Fig. 4 shows the computed cluster centers in the intervals [0; 314] (cluster 1), [416; 929] (cluster 2), [1150; 1450] (cluster 3) and [1710; 2130] (cluster 4). These patterns show good accordance to original patterns (Fig. 1).

Since the choice of time windows is crucial to results and interpretation, we focus to a smaller time window $i \in [0; 500]$ and re-apply the clustering method. Plotted Euclidean distances (Fig. 5(a)), percental contributions $A_K(i)$ (Fig. 5(b)) for $K = 2, 3, 4, 5$ and 15 and the quality cluster measure $p(i)$ (Fig. 5(c)) show a smeared cluster border at $i = 230$ and a sharp border at $i = 370$. This indicates rather spread and packed data points $i \in [0, 230]$ and $i \in [370, 499]$, respectively. The interval between the clustered windows represents the transition part between clusters.

Besides the effects of time windows, the number of clusters R also affects clustering results. In Fig. 6, $p(i)$ is plotted with respect to index i for $R = 20$, 30 and 40. We observe drop-offs and rises at the same time steps i for all three curves. Plots for $R = 20$ and 40 are shifted by a constant value of p, while clustering with $R = 30$ pronounces the last clustered window and diminishes the cluster quality in the first window. Since cluster borders are determined only by drop-offs and rises, all detected clustered windows coincide.

In the following, we aim to assess the presented method by some statistical tests. Figs. 3 and 5 indicate both temporal and spatial cluster structures, which should vanish for time-randomized signals. Fig. 7 shows plots of $p(i)$ for the artificial temporally-disordered data set. No distinguished temporal windows are observed as before, while values of $p(i)$ are small.

Since the computed cluster centers depend on initial clusters, averaged cluster results and corresponding standard deviations are computed for $L = 10$ different initial cluster configurations. Fig. 8 shows $\bar{p}(i)$ and error bars $\Delta p(i)$ for the time window [0; 500]. The transition part $i \in [260; 380]$ shows slightly increased variances compared to the clustered time windows. Points $i \in [0; 150]$ appear statistically independent from points $i > 180$ and, subsequently, originate from a different data structure. While error bars for $i \in [180; 240]$ and $i \in [240; 370]$ overlap, the increase at $i \approx 370$ does show only few overlapping error bars.

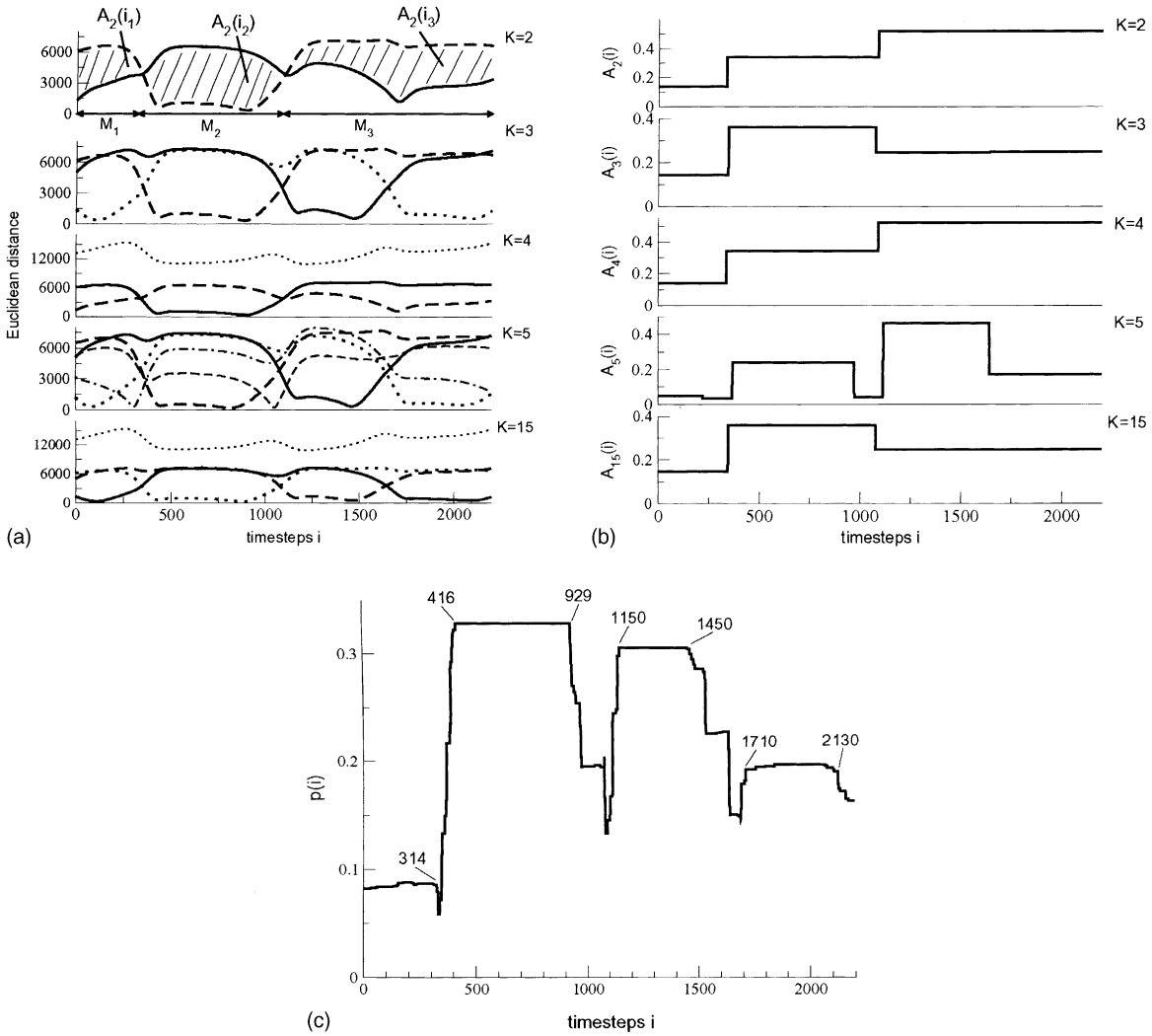

Fig. 3. Clustering results of simulated data in the time window [0; 2199]. (a) Euclidean distances $d(k, i)$ from data to K cluster centers are plotted with respect to time points i for different K. At $K = 2$, areas between the lowest and the second lowest curve are denoted by $A_K(i_j)$, corresponding to definition (1). (b) Percental contributions A_K are plotted with respect to time points i for different K. (c) Cluster quality measures p are plotted with respect to time points i. Plateaus at [0; 314], [416; 929], [1150; 1450] and [1710; 2130] are detected.

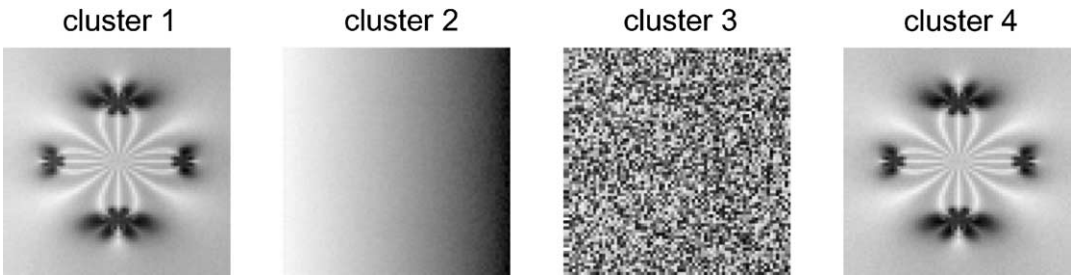

Fig. 4. Spatial averages of clustered windows in corresponding sequence.

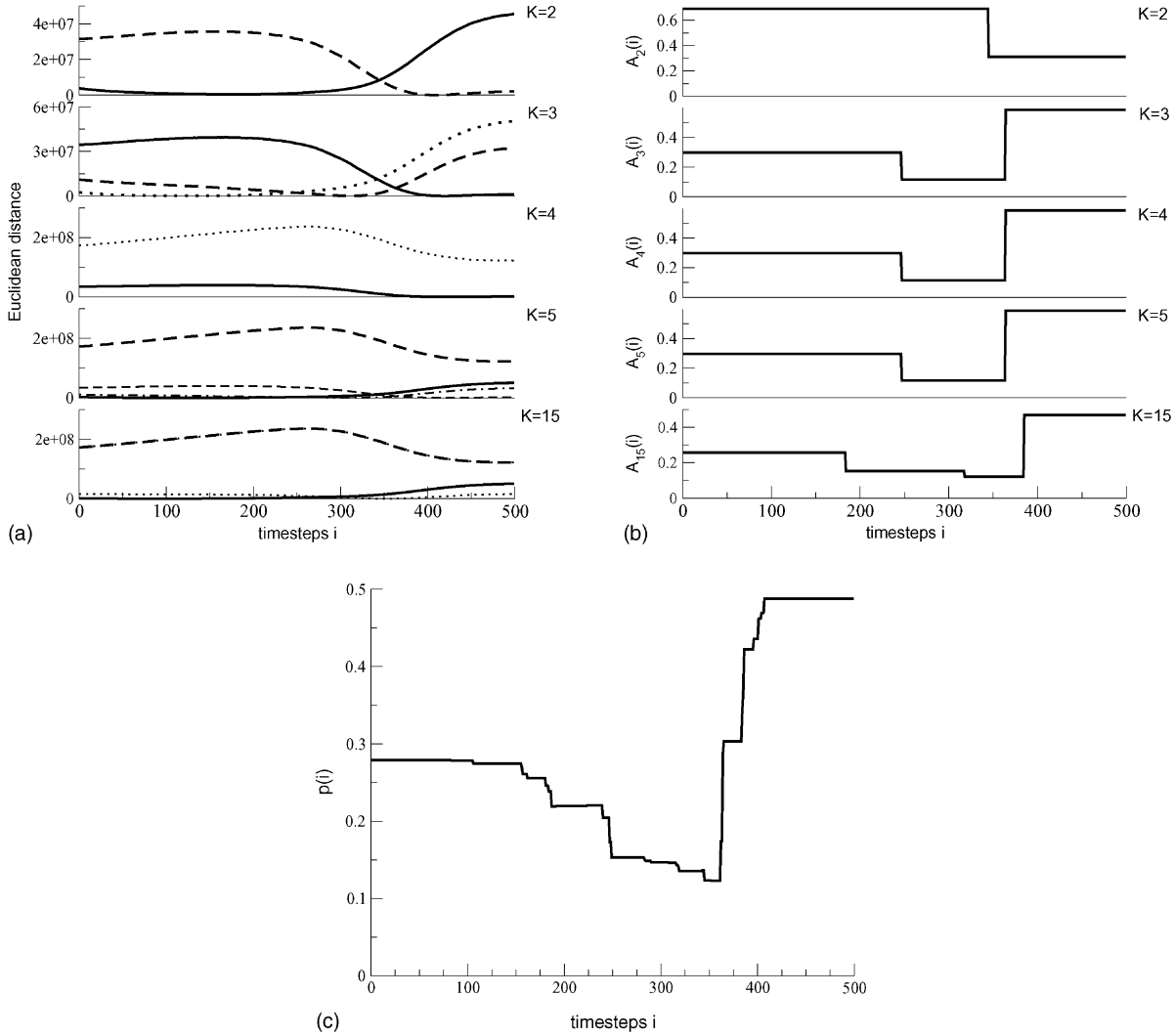

Fig. 5. Clustering results of simulated data in time window [0; 499]: (a) Euclidean distances; (b) percental contributions; (c) cluster quality measures p are plotted with respect to time points, respectively.

3.2. Middle latent auditory evoked potentials

The investigated data sets were obtained from a study in which auditory brain stem responses (ABR) and auditory evoked potentials of middle latency (MAEP) were investigated simultaneously. Stimuli were diotic rarefaction clicks of 100 μs duration at a level of 60 dB nHL. The interstimulus interval was chosen to vary randomly and equally distributed between 62 and 72 ms, yielding an average stimulation rate of approximately 15 Hz. The EEG was recorded with 32 electrodes which were placed according to an extended 10–20-system, Cs served as common reference electrode.

Before digitization, raw data were passed through an analog anti-aliasing lowpass filter with a cutoff frequency 2 kHz. Data were sampled at a rate of 10 kHz, the recording interval comprised 600 samples in the time inter-

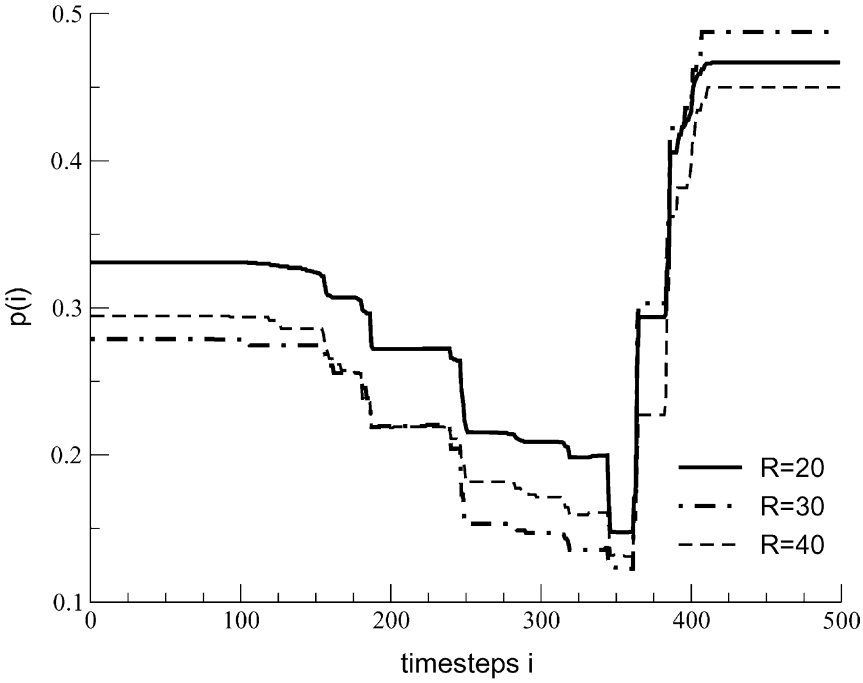

Fig. 6. Cluster quality measure p plotted with respect to time for different parameters R.

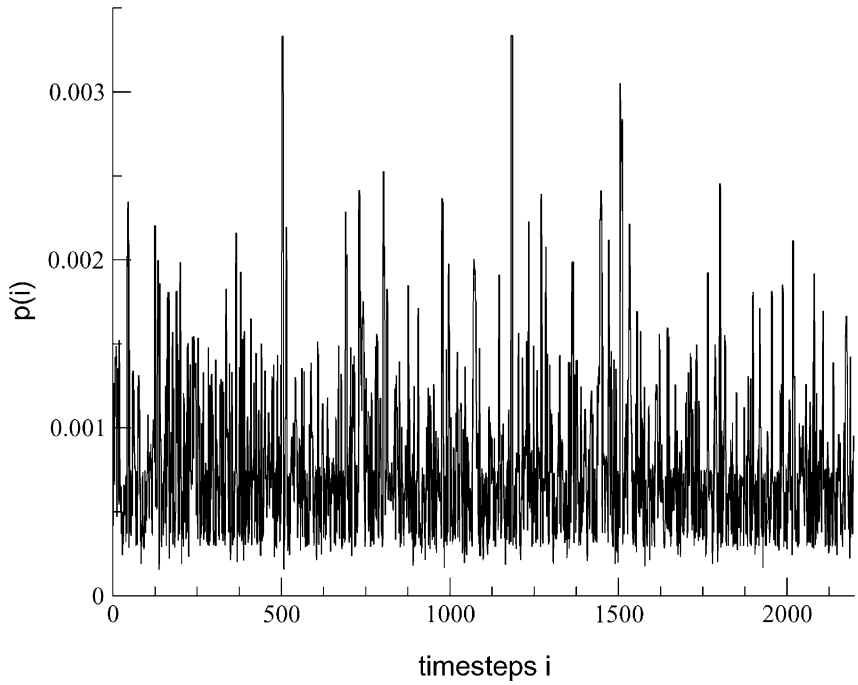

Fig. 7. Cluster quality measure $p(i)$ for time-randomized simulated data.

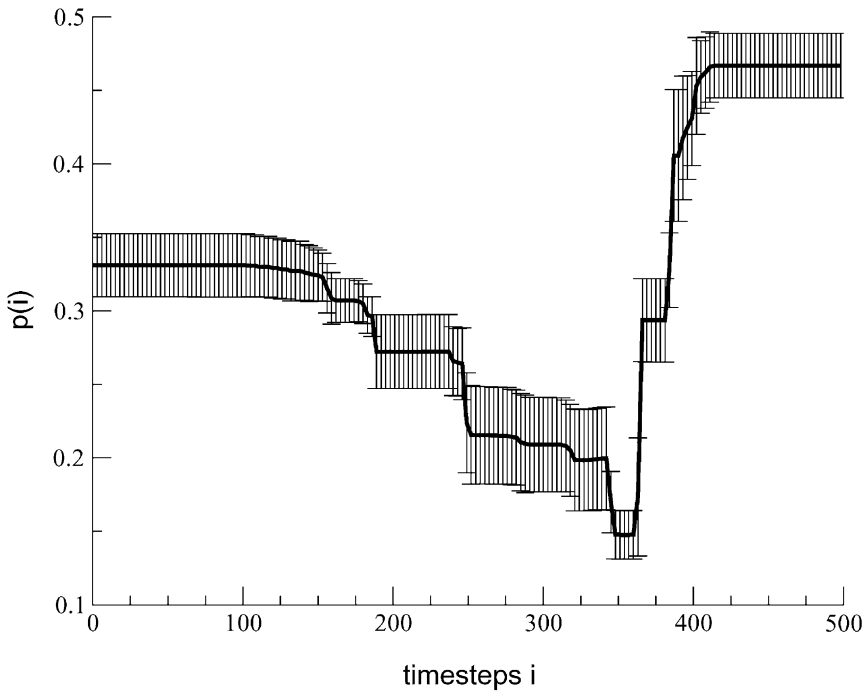

Fig. 8. Average cluster quality measures $\bar{p}$ and corresponding standard deviations Δp (error bars) plotted with respect to time. For reasons of visualization, only every third time point is plotted.

val from −15 to 45 ms relative to stimulus onset. The 10,000 single trials were recorded and stored to hard disk for off-line analysis. They were filtered by a zero-phase forward–backward bandpass filter with lower frequency $\nu_l = 20$ Hz and upper frequencies $\nu_h = 300$ Hz and 2000 Hz, respectively. From the filtered trials a weighted average was computed using the inverse powers of the noise in the trials as weightings [17]. For each channel, trials show Gaussian amplitude distributions with varying means over time, but nearly time-constant variances. Thus, averaging over trials extracts evoked responses as means, and averaged variances over time represent the noise level for each channel. The present work analyzes three data sets from different subjects showing low (subject dj), middle (hr) and higher (rh) noise levels. Figs. 9–11 show the electric activity measured on the scalp for the three subjects. Single channel plots illustrate the temporal dynamics, while the temporal sequences of spatial activities (equipotential maps) exhibit the spatio-temporal dynamics. Additional data sets were calculated for every subject and corner filter frequencies $\nu_l = 20$ Hz, $\nu_h = 300$ Hz by re-referencing to spatial average.

3.2.1. Cluster results with $\nu_h = 300$ Hz

Applying the first clustering step, Euclidean distances from data points to cluster centers are obtained. Fig. 12 shows results for three clusters. The lowest curves reflect smallest distances from data to clusters and indicate approached cluster centers. In data sets dj and hr, clusters are approached in several similar time windows. Cluster nc = 3 is approached until ∼4 ms (dashed line), as data points near cluster 1 (data set dj) and cluster 2 (data set hr) in the interval around 30 ms. In subject rh, cluster 1 (solid line) is approached in larger time intervals with $i \in \mathcal{M}_{11} = [-15\,\text{ms}; 13\,\text{ms}]$ and $i \in \mathcal{M}_{21} = [22\,\text{ms}; 45\,\text{ms}]$, while clusters 2 (dotted line) and nc = 3 (dashed line) are reached in between. Since increased numbers of clusters change the borders of time intervals, the cluster

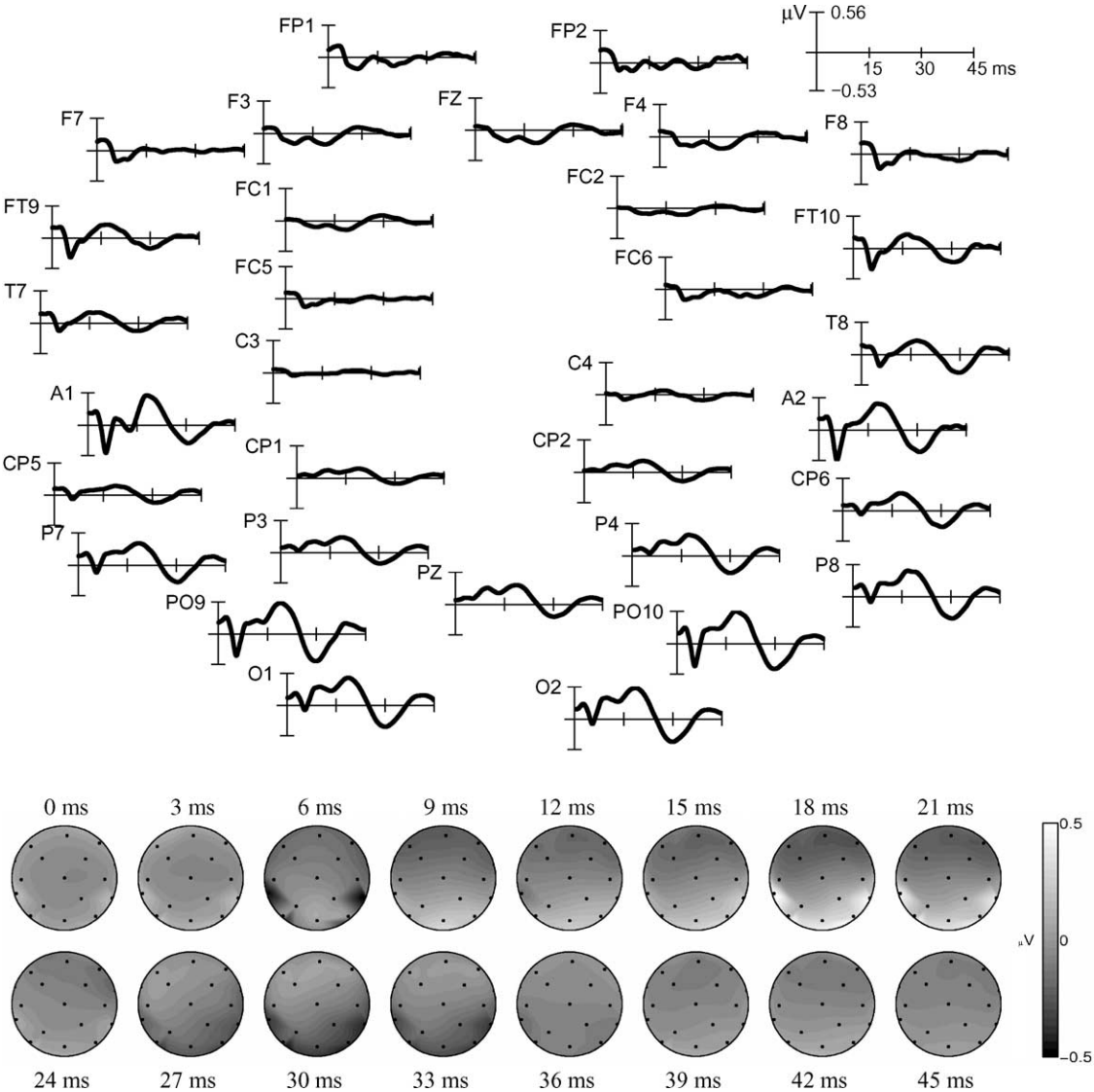

Fig. 9. Measured electric potentials on the scalp of subject dj. The top plots show activities in the single channels, at the bottom a time series of spatial distributions is shown.

quality measure $p(i)$ is calculated and plotted with respect to time in Fig. 13. At 5 ms, descents in p are observed in all data sets followed by a sharp rise. Following plateaus last until 7 ms, when p drops again. As in the previous section, this structure indicates a data cluster in the corresponding time interval. Clusters also are found around 17 and 29 ms for all data sets. These results show good accordance with clustered windows in re-referenced data.

Re-applying the cluster method to cropped time intervals for original and re-referenced data around 5 and 30 ms confirms these findings (Figs. 14 and 15). The corresponding maps on the left side represent signal averages over plateaus for single-referenced data. They confirm clustered windows by good accordance to quasi-stationary states in the data.

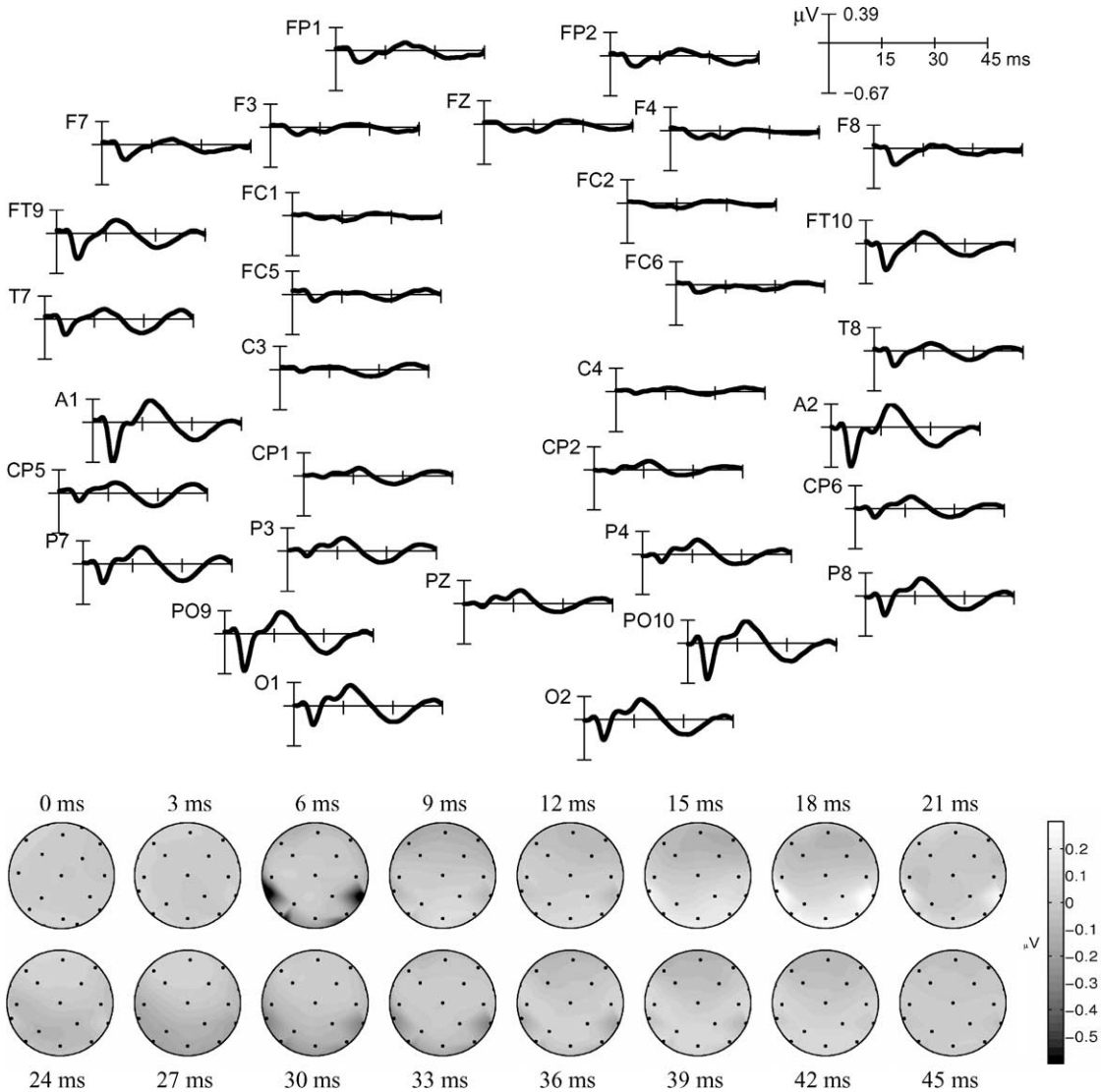

Fig. 10. Measured electric potentials on the scalp of subject hr shown in single channels and as a time series of spatial maps.

3.2.2. *Results from DSBM with* $\nu_h = 300$ *Hz*

This section presents the application of DSBM to a previously detected signal segment at 29 ms for all single-referenced data sets. The segment borders are chosen according to cluster borders in Fig. 15. As a first step, PCA is applied in order to reduce the signal dimensionality with errors $E < 10^{-5}$ according to (5) for five modes. These projections serve as a new five-dimensional signal, which is modeled by DSBM.

Reflecting considerations of Section 2.3 and applying (12), two-dimensional projection planes are obtained for each data set from the first two PCA-modes ($E < 10^{-2}$). Synchronously determined dynamical systems (13) are fitted with minimal errors $V_d(\epsilon = 0.09) = 4 \times 10^{-4}$ (data set dj), $V_d(\epsilon = 0.0) = 4 \times 10^{-4}$ (data set hr) and $V_d(\epsilon = 0.15) = 4 \times 10^{-3}$ (data set rh) applying polynomials of third order. Lower polynomial orders lead to

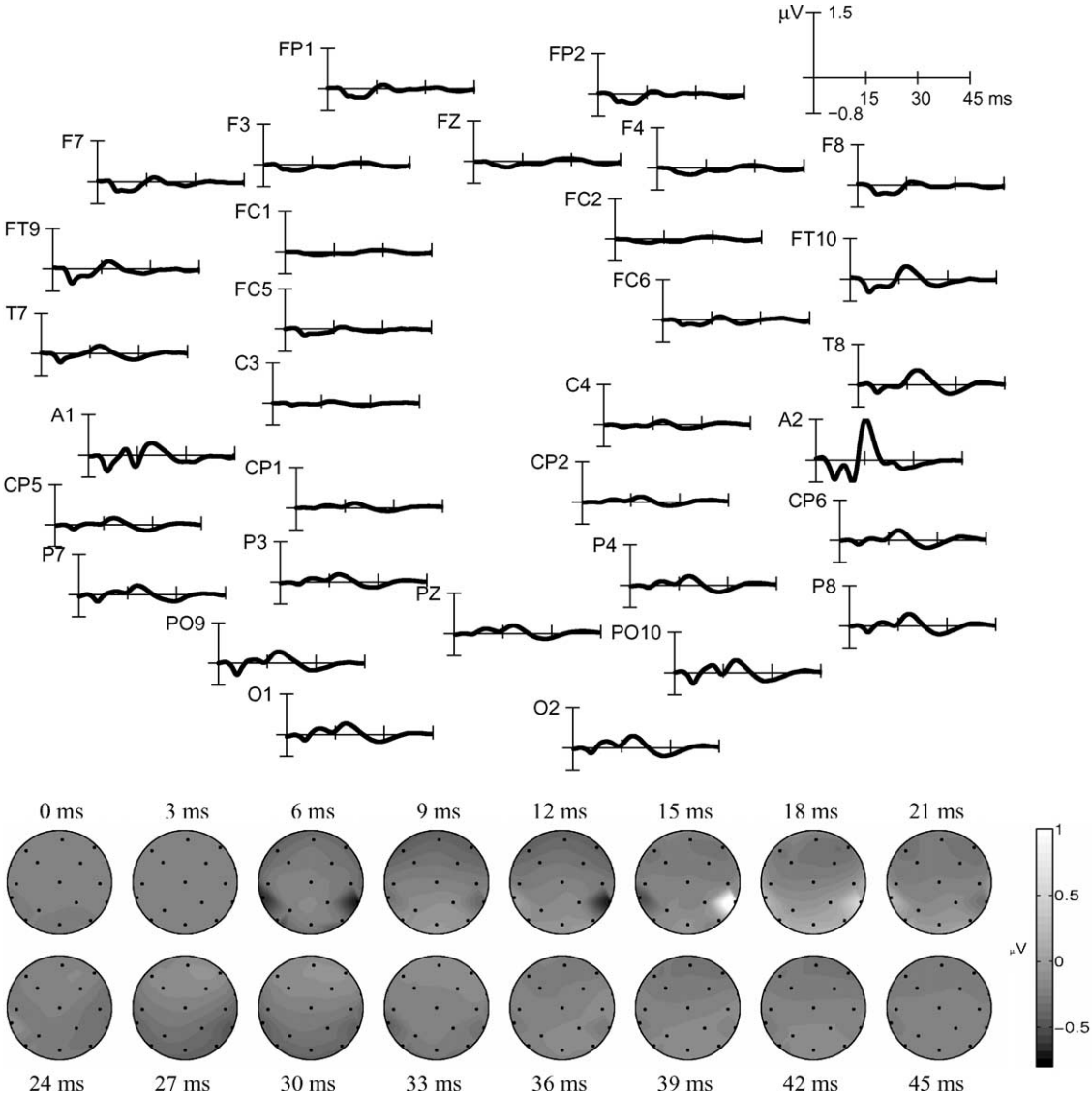

Fig. 11. Measured electric potentials on the scalp of subject rh shown in single channels and as a time series of spatial maps.

worse fits and bad reconstructions of the signal, whereas higher orders do not lead to new dynamical properties. The obtained model parameters are shown in Table 1.

Backward transformations of modes $\{\mathbf{w}_i^\dagger\}$ and amplitudes $\{y_i\}$ to the original data space allow a superposition of the modes according to (3) and the reconstruction of the analyzed signal. Since static modes are determined by their coordinates in data space, they also represent spatial activity distributions. Reconstructed modes $\mathbf{w}_1^\dagger, \mathbf{w}_2^\dagger$ are shown in Fig. 16 with corresponding amplitudes for all subjects. Solutions of the determined dynamical systems are gained by integration and are also plotted in Fig. 16 as dashed lines. They show good accordance with the projected signal.

The basic assumption of attractive fixed points is verified by the topology of the obtained dynamical systems. Integrations of the obtained differential equations with various initial points lead to sets of trajectories and elucidate

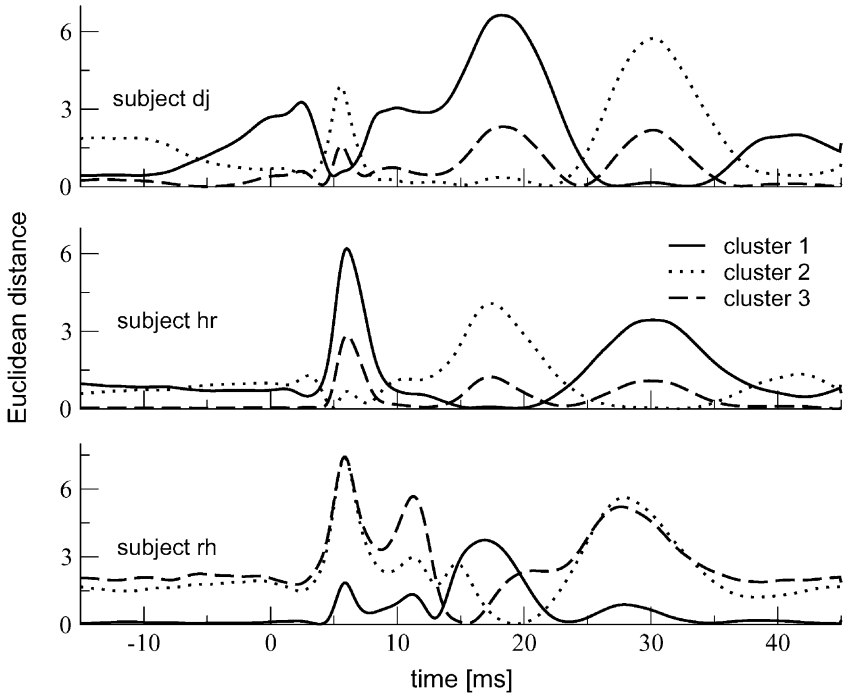

Fig. 12. Plot of Euclidean distances from data to three detected clusters in temporal order for all data sets. The stimulus sets on at 0 ms.

the topology of the temporal dynamics (Fig. 17). All data sets show a saddle node FP3 at $\mathbf{y} = 0$, a stable focus FP4 and another saddle node FP1. The additional stable focus FP2 is detected only in data sets dj and hr.

In Fig. 18, trajectories are shown for re-referenced data. The gray scales in the background code the absolute values of dynamical systems $|\mathbf{f}(y_i)|$. These plots, introduced by Uhl et al. [18], illustrate temporal dynamics quite well: small values are dark-coded and indicate regions of fixed points, i.e. regions of possible attractions or repelling. In case of gradient dynamics, Uhl plots visualize potential landscapes with fixed points located in minima. Here, two regions of attraction are visible in all subjects. They agree with previous results.

3.2.3. *Results for* $\nu_h = 2000$ *Hz*

In this section, we analyze data without off-line lowpass filters. Fig. 19 shows Fourier spectra of four high-energy data channels at component P_a for all subjects. We observe major peaks around 200 Hz for all subjects and

Table 1
Coefficients from Eq. (13) of determined dynamical systems for all subjects

Subject	Mode i	A_{i1}	A_{i2}	A_{i3}	A_{i4}	A_{i5}	A_{i6}	A_{i7}	A_{i8}	A_{i9}
dj	1	0.05	−0.26	0.012	−0.26	−0.07	−0.06	−0.24	−0.34	0.12
	2	0.01	−0.07	−0.02	−0.02	0.12	0.01	0.13	0.38	0.40
hr	1	0.05	0.21	−0.05	0.25	0.13	−0.08	0.04	0.02	−0.06
	2	−0.00	−0.07	0.01	0.11	−0.18	−0.06	0.18	−0.15	0.02
rh	1	0.05	−0.06	0.06	0.04	0.01	−0.07	0.00	−0.00	0.00
	2	−0.03	−0.09	−0.00	−0.17	−0.08	0.23	0.08	0.01	0.01

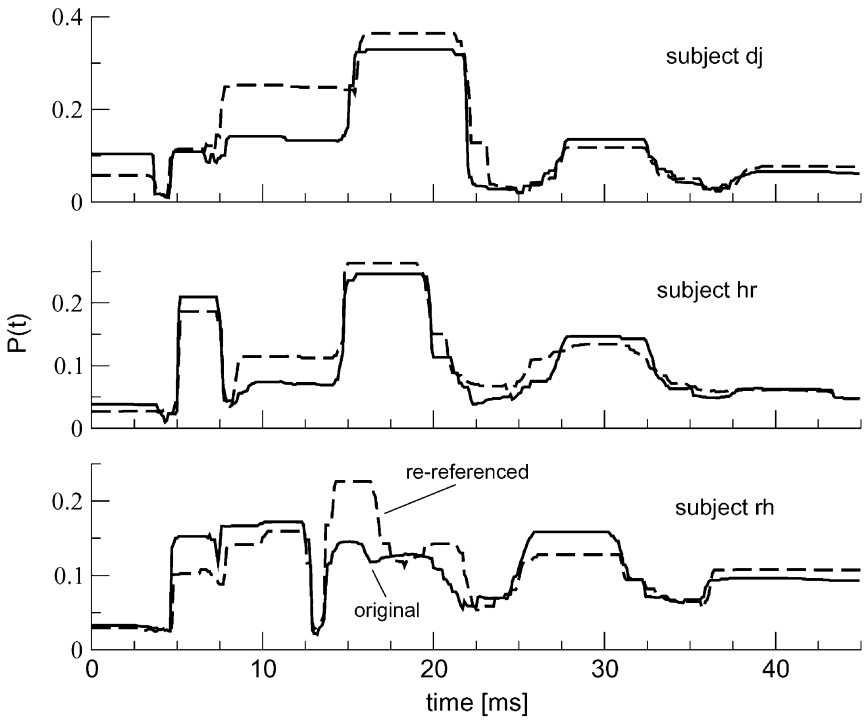

Fig. 13. Cluster quality p plotted with respect to time for original (solid line) and re-referenced (dashed line) data. In all data sets, clusters are recognized at 30, 18, 12 and 5 ms. The stimulus sets on at 0 ms.

non-vanishing contributions up to 1500 Hz. Obtained cluster results $\bar{p}$ and corresponding standard deviations Δp are shown in Fig. 20. We observe similar clustered windows. Focussing to those time intervals, we apply DSBM and obtain optimal projections on two spatial modes. Their corresponding amplitudes and Fourier spectra are shown in Fig. 21. Comparing these results with projections obtained for $\nu_h = 300$ Hz, sign flips are observed in amplitudes of subjects dj and rh, which reflect the numerical non-uniqueness of principal components. We also observe augmented contributions of lower frequencies, which reflect the known lowpass filter properties of DSBM [20]. Finally, we determine optimal polynomial models and integrate the obtained differential equations with various initial values. Increased modeling errors are obtained to $V_d = 0.093$ (subject dj), $V_d = 0.152$ (subject hr) and $V_d = 0.230$ (subject rh), which results from the increased noise level. Fig. 22 presents sets of trajectories for all subjects and we observe again a saddle point FP3 and a stable focus FP4 in all three data sets.

3.3. Discussion

Let us summarize the results. Fig. 3 shows clustering results obtained from the artificial data set. The clustered time windows in Fig. 3(c) coincide with data windows of quasi-stationary states. The equivalence of cluster centers in Fig. 4 and quasi-stationary data patterns confirm the findings. These results propose the clustering approach as an objective algorithm for detecting multivariate quasi-stationary attractors. First successful computations of p for event-related potentials [31] indicate a broad bandwidth of applications. Since there are many methodological features, on which the clustering results depend, additional assessment is applied. One of the most important features

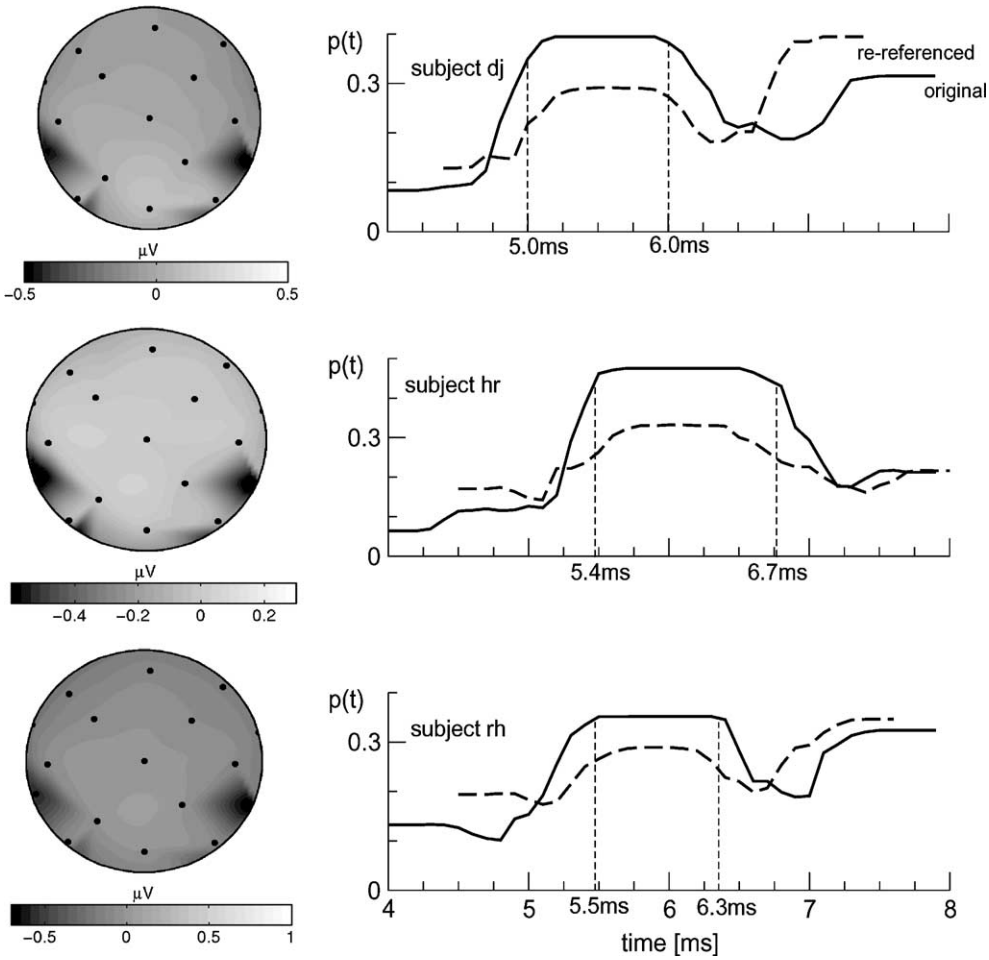

Fig. 14. Cluster results obtained in the time window [4 ms; 8 ms] for all data sets. Probability plateaus indicate clustered data, drop-offs mark their borders (vertical dashed lines). Spatial maps on the left side show signal averages in the marked time windows.

is the time window of clustered data. Fig. 5 shows results obtained from cropped data. We observe similar borders of clustered windows compared to the previous results. The slight deviations of results arise from the focused view to data, which allows a more detailed investigation of the data structure. The previous right cluster border at $i = 416$ is re-gained, while the left border at $i = 314$ has to be shifted to lower values. We point out that the method yields transition windows between clusters with smeared borders rather than fixed borders. This smearing reflects typical data properties of continuous time series. Since the quality measure is computed by averaging R percental contributions, p is regarded to be independent of the number of clusters. A comparison of cluster results from three different R (Fig. 6) confirms this independence. Shifts and scalings of p are observed, but cluster borders remain constant. As in the previous examinations, time points i at drop-offs and rises are invariant and, hence, reflect intrinsic properties. If this is true, a randomized time series should not show any structure. Fig. 7 confirms this aspect. As a last evaluation step, we investigate the dependence on initial clusters. Mean values and standard deviations of p (Fig. 8) illustrate the significance of results. We observe statistically independent mean values of segmented clustered windows [0; 150] and [400; 500] with non-overlapping error bars. We conclude that troughs

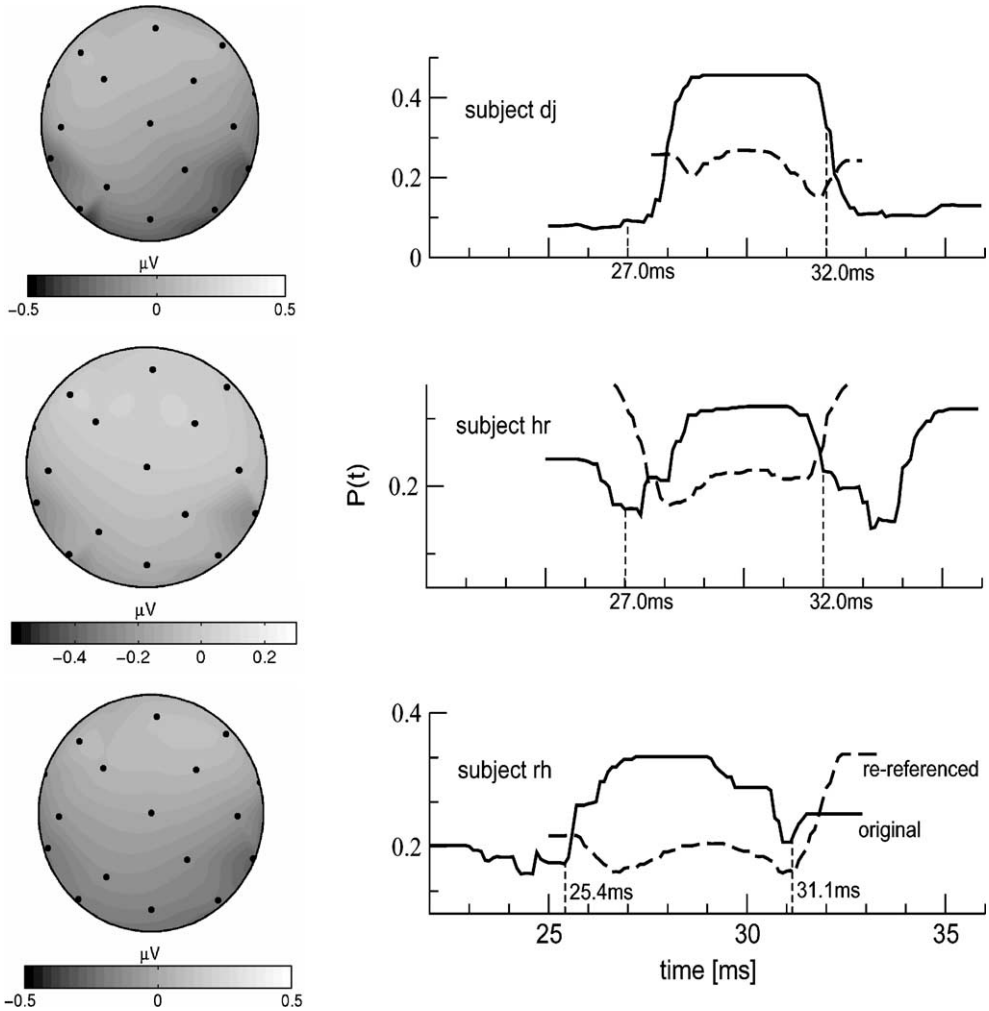

Fig. 15. Cluster results in the time windows of wave P_a. Window borders are obtained from Fig. 13. Vertical dashed lines mark cluster borders for original data (solid line), which coincide to the borders of re-referenced data (dashed line). On the left side, averaged spatial distribution from intervals [28.0 ms; 31.0 ms] (data set dj), [28.0 ms; 31.0 ms] (data set hr) and [26.6 ms; 30.4 ms] are shown.

and sharp rises of p represent borders of clustered windows and are robust against time windowing, number of clusters and the choice of initial clusters.

After these evaluations, measured early evoked potentials are analyzed. First applications to data with applied lowpass filter $\nu_h = 300$ Hz reflect quasi-stationary behavior by drop-offs, sharp rises and plateaus of the quality measures (Fig. 13). The observed clustered windows shows good accordance to quasi-stationary patterns in Figs. 9–11 and the so-called waves [33]. The early wave V from 5 to 7 ms belongs to auditory brainstem responses (ABR) and originates presumably from the lateral lemniscus and inferior colliculus. Components of middle latent potentials are also prominent at 18 and 30 ms as components N_a and P_a, respectively. They are supposed to be generated subcortically in the midbrain and in the temporal lobe [34], respectively. Furthermore, all subjects show rather robust cluster windows for single-electrode reference and average reference. This important aspect originates from the idea of similar sequential patterns and represents a necessary condition for a reliable segmentation method

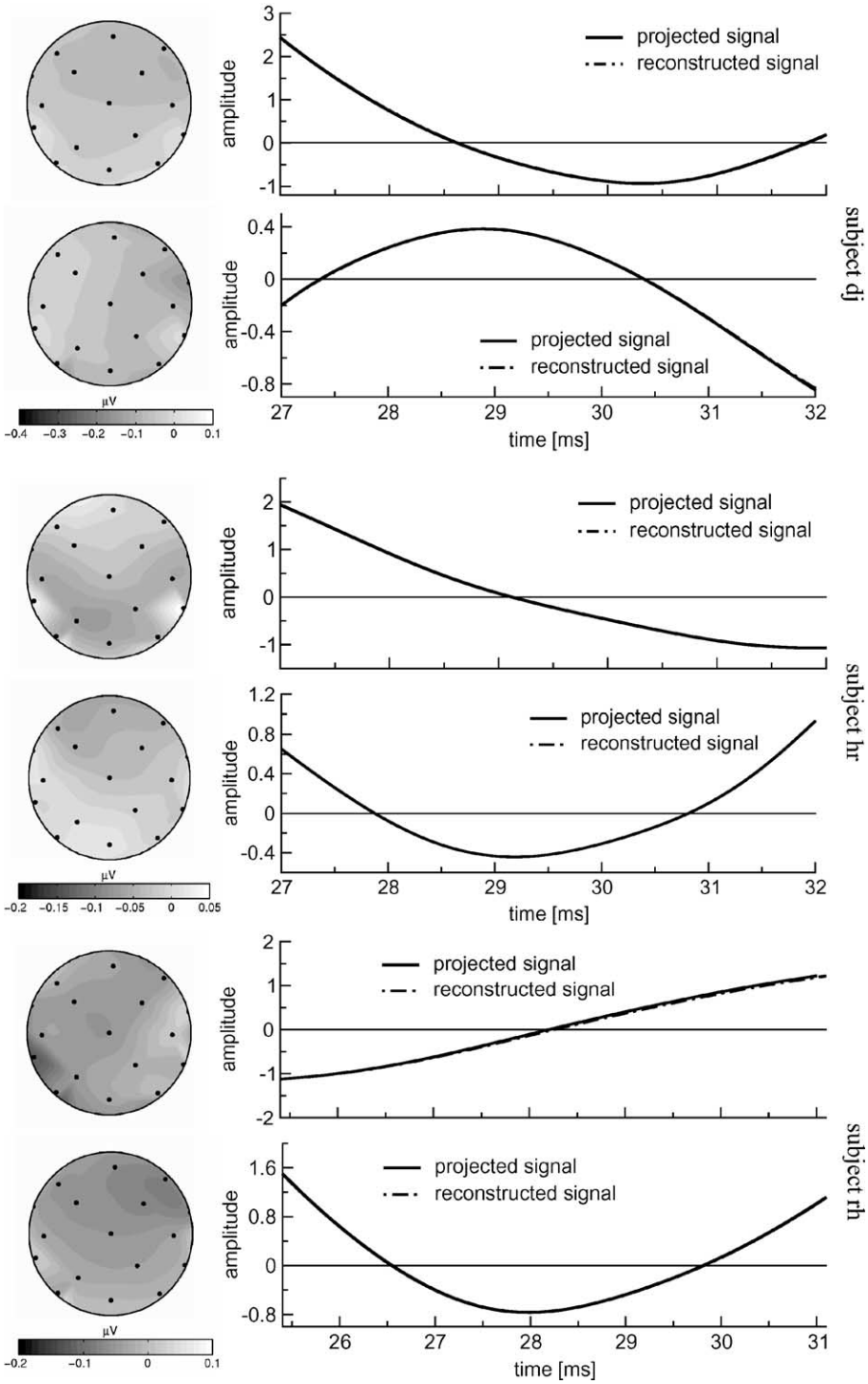

Fig. 16. Results of DSBM for component P_a and all data sets. Right side: amplitudes of the determined low-dimensional projections (solid line) and solutions of the determined dynamical systems (dot-dashed line). Left side: reconstructed spatial modes $\mathbf{w}_1^\dagger$, $\mathbf{w}_2^\dagger$.

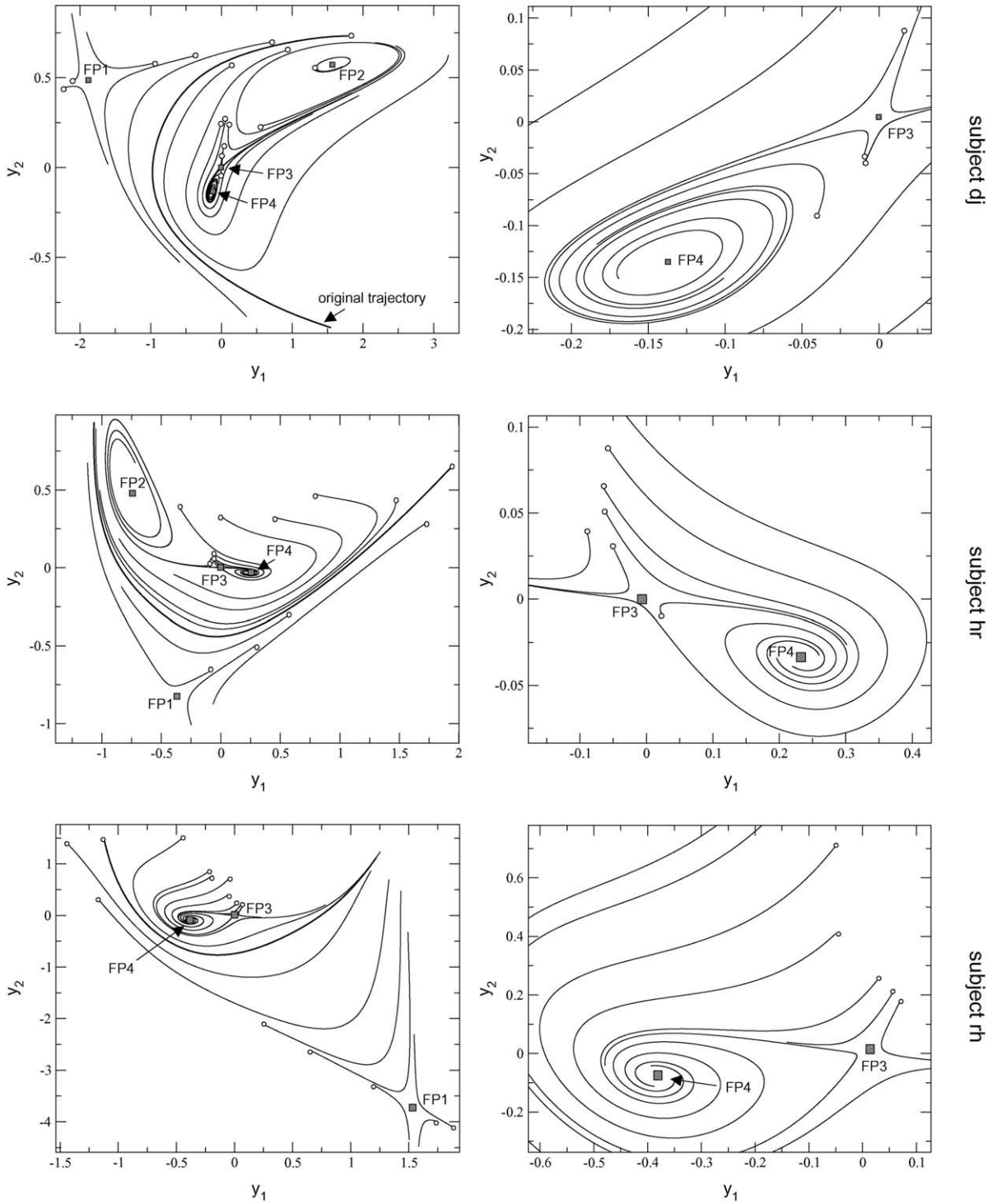

Fig. 17. Sets of trajectories of component P_a for $\nu_h = 300$ Hz and all subjects. The right side shows focused plots of attractive areas on the left side. Circles mark initial values of integration, while boxes denote fixed points.

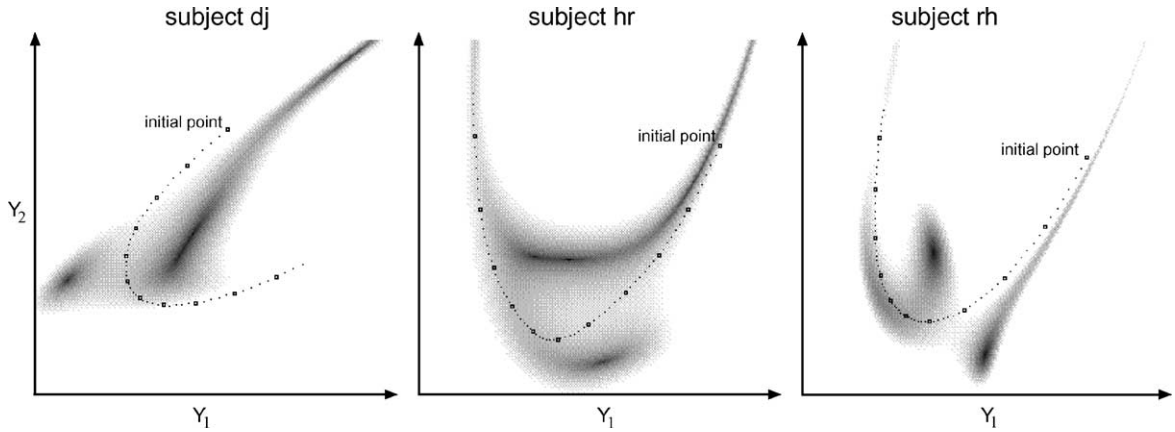

Fig. 18. Uhl plots for all re-referenced data sets. An Uhl plot shows the projected signal trajectory and its corresponding dynamical landscape $|\mathbf{f}(y_1, y_2)|$ in the background as gray scales. Low values are dark-coded and represent regions near fixed points.

[7]. Focussing to wave V and P_a (Figs. 14 and 15, respectively), spatial maps are obtained. We observe a good accordance to quasi-stationary patterns in Figs. 9–11 at ∼6 and ∼30 ms, respectively. We point at the latency variability of components V and P_a between subjects, which reflects the physiological diversity of human brains even at such early and functional rudimentary stages.

Applications of DSBM to wave P_a yields optimal projective spatial modes and corresponding amplitudes, shown in Fig. 16 for all subjects. These amplitudes obey dynamical systems, whose corresponding vector fields are shown

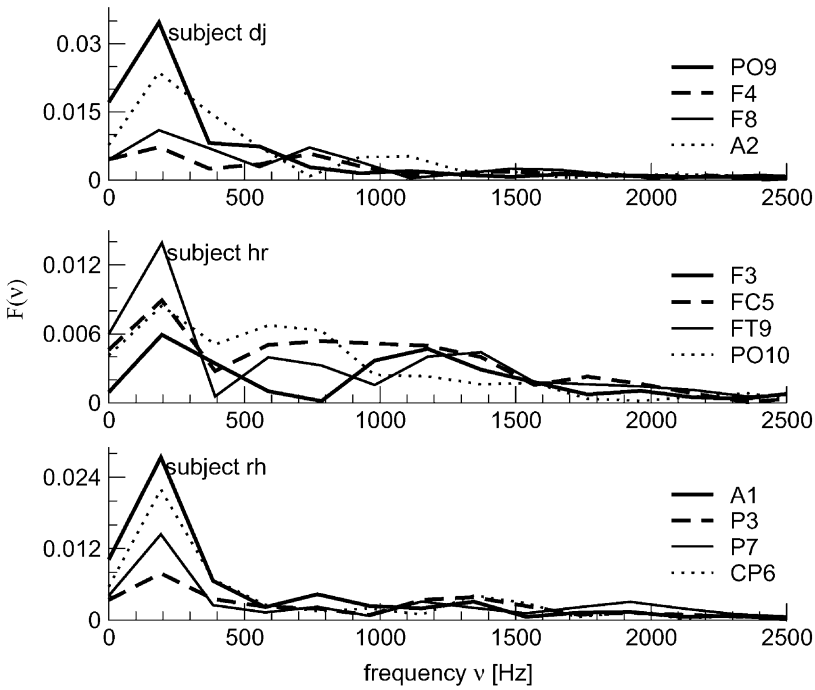

Fig. 19. Frequency spectra of all data sets for $\nu_h = 2000$ Hz. For each data set, spectra of four high-energy channels are computed.

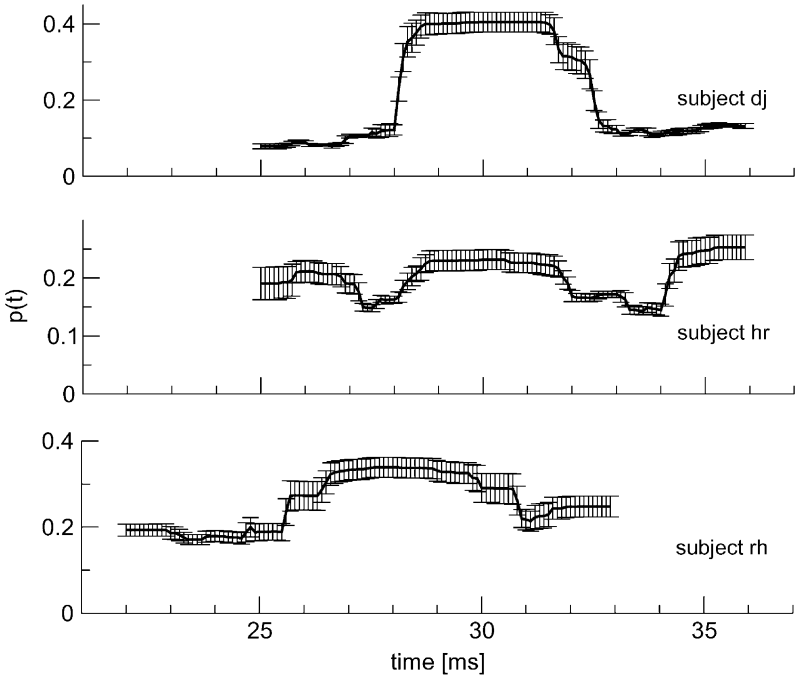

Fig. 20. Average cluster quality measure $\bar{p}$ and corresponding standard deviations Δp (error bars) plotted with respect to time for $\nu_h = 2000$ Hz and all data sets.

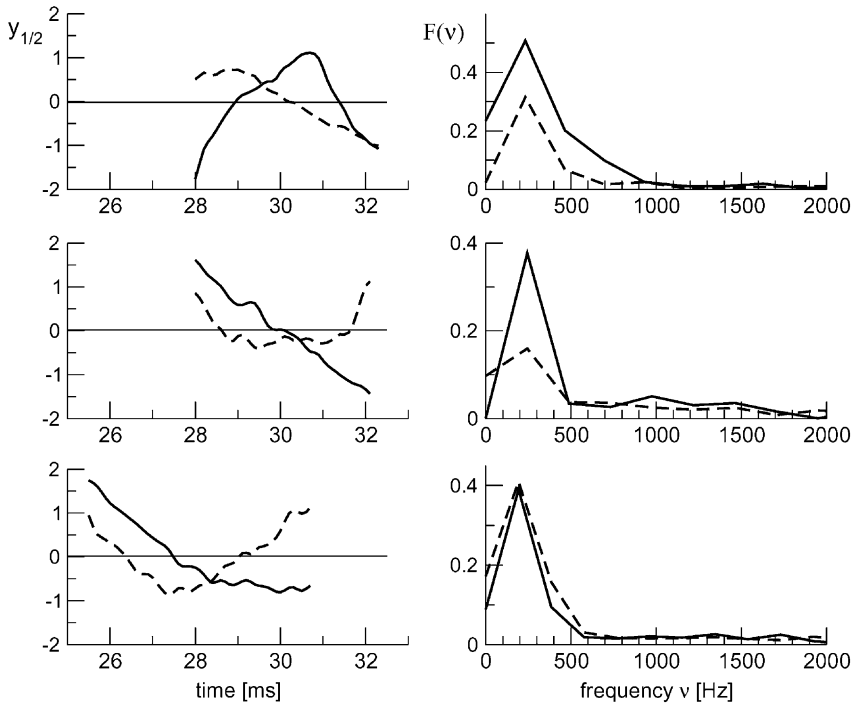

Fig. 21. Amplitudes $y_{1,2}(t)$ and frequency spectra $F(\nu)$ of obtained low-dimensional models for $\nu_h = 2000$ Hz and all data sets.

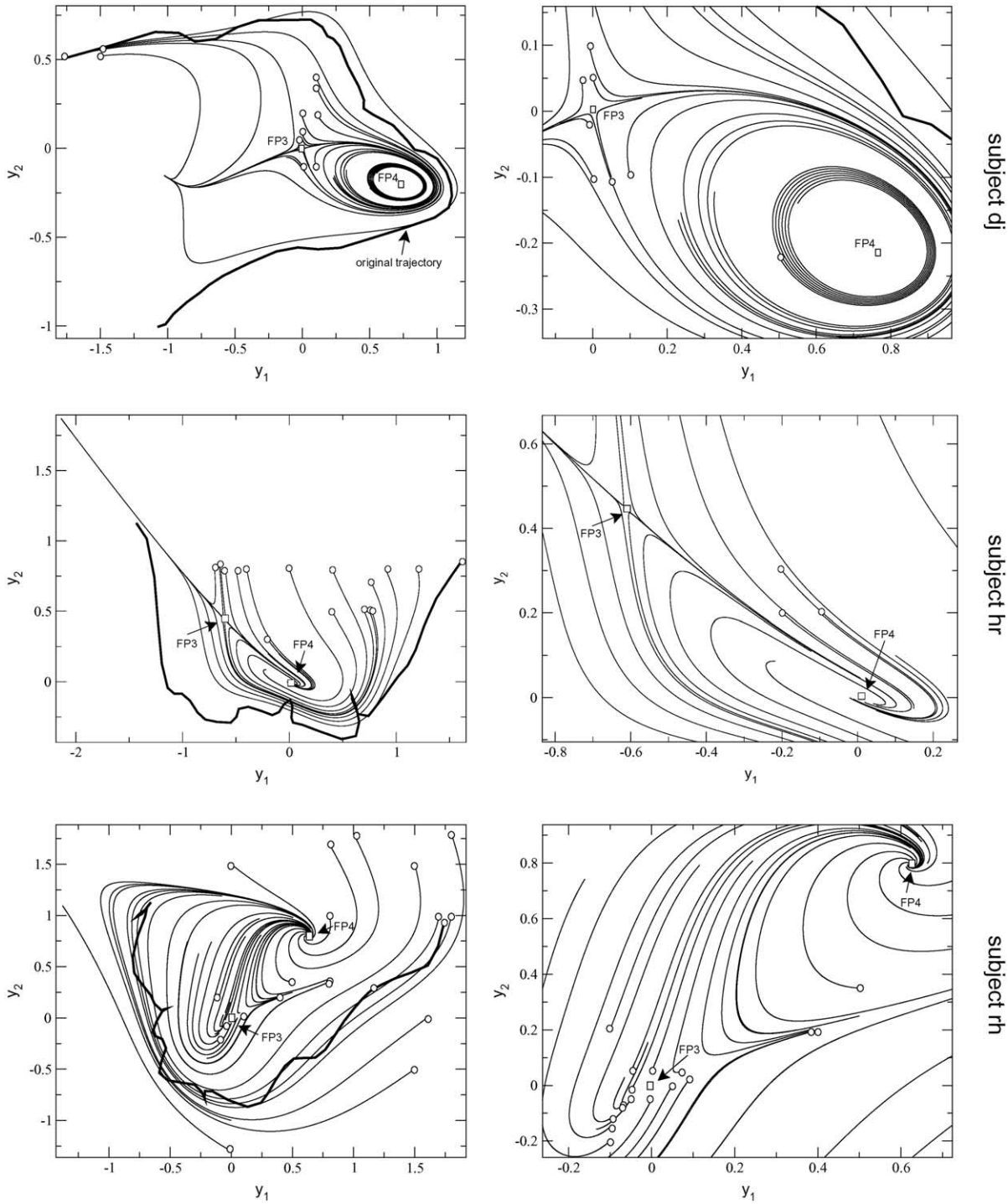

Fig. 22. Sets of trajectories of component P_a for $\nu_h = 2000$ Hz and all subjects. The right side show focused plots of attractive areas on the left side. Circles mark initial values of integration, while boxes denote fixed points.

in Fig. 17. We observe similar topologies for all subjects, while the stable focus FP2 is not detected in data set rh. Since the length of the time window and the amplitude scale decides how far and how fast trajectories evolve from the origin, the observed dynamics appears more local in some cases. We argue that FP2 results from scaling differences, whereas fixed points FP1, FP3 and FP4 contribute mainly to the dynamics of component P_a and define its dynamical behavior. Attractive regions in re-referenced data agree with the detected topology. However, investigations of regions near fixed points, dark-coded in Fig. 18, show slightly different stability properties. Since corresponding investigations are still in progress, we refer to [35].

Studies of MAEPs usually apply bandwidth filters with upper frequencies of about 300 Hz (see, e.g. [36]). However, modeling of filtered data implicates a statistical bias (Section 2.3). We compute point lags $n_{\text{thr}} \approx 17$ for $\nu_{\text{h}} = 300$ Hz, which represent time lags of 1.7 ms. This means that the modeled time intervals contain four independent data points for all subjects. One might argue that only these few data points should be analyzed. First of all, fitting of dynamical models is not constrained to independent data, as the temporal dependence itself is modeled by differential equations. Additionally, model dynamics obtained from few data points are not unique, i.e. not reasonable. For verification, we computed models from four temporal equidistant data points for all subjects in corresponding time intervals and found divergent dynamical topologies. Even by increasing the number of points to 5, no common topology was found. This reflects the model ambiguity discussed in Section 2.3. From further analytical investigations, it turns out that the obtained models critically depend on given data points.

Nevertheless, statements about common topological properties of models are not reliable for a single lowpass filter. While each of the models obtained with $\nu_{\text{h}} = 300$ Hz is reliable, common features might be questioned due to the mentioned bias. Therefore, we aim at analyzing data, whose number of statistically independent points exceeds the number of model parameters. The necessary least filter frequency is determined to $\nu_f = 18/(2 \times 5\,\text{ms}) = 1800$ Hz for a time window of 5 ms. Off-line unfiltered data with $\nu_{\text{h}} = 2000$ Hz fulfills this condition and, hence, is analyzed in the following.

Borders of clustered windows turn out to be robust towards the additional noise level (Fig. 20) and define time windows for applications of DSBM. Subsequent optimal dimensionality reductions to a two-dimensional model yield amplitudes and synchronously fit dynamical systems. In Fig. 22, obtained vector fields for all subjects illustrate a good agreement of stability properties of fixed points FP3 and FP4 with the previous results. Since we focus the integration to region near $\mathbf{y} = 0$, we might miss additional fixed points for larger $|y_i|$. While this aspect should be investigated, discussions about scaling effects would exceed the aim of the present paper and we refer the reader to [35]. We find that both in data filtered with $\nu_{\text{h}} = 300$ and 2000 Hz, similar topologies of dynamical systems are found. We recognize invariant dynamical models towards lowpass filtering and noise level. The same fixed points FP3 and FP4 are found around $y_i = 0$ in data sets of three different subjects each with two different noise levels. These findings indicate a general topology of the component P_a, i.e., fixed underlying processes. However, further topological assessment is necessary.

4. Conclusion

We have introduced a methodological framework for analyzing and modeling quasi-stationary multivariate time series. The framework contains a cluster analysis and a subsequent DSBM-approach. In a first step, the clustering method yields data segments of quasi-stationary regions in data space. A novel cluster criterion is derived by introducing a cluster quality measure, which is independent of the number of clusters. Results obtained from simulated data are assessed statistically. Focussing to segmented data, DSBM fits synchronously optimal low-dimensional projections and deterministic dynamical systems.

Applications to MAEP-data from three different subjects yield data segments, which correspond to the well-known evoked potentials. For wave P_a, DSBM extracts two-dimensional dynamical models, while the corresponding amplitudes are interpreted as order parameters of underlying processes. The obtained models show common topological properties, which are invariant under variations of the applied noise filters. These findings indicate an ordered and self-organizing brain state during the emergence of wave P_a.

Future work will apply the proposed combination of methods to experimental data obtained with slightly varied paradigms. We conjecture that every experimental paradigm leads to one corresponding topology for each wave and, hence, can be classified by indices known in dynamical systems theory. This classification scheme for early evoked potentials would be valuable in brain science, especially in clinical research.

Acknowledgements

A. Hutt would like to thank B. Maess, A. Stevens, U. Steinmetz and, especially, F. Kruggel for valuable and fruitful discussions.

References

[1] Y. Kuramoto, Chemical Oscillations, Waves, and Turbulence, Springer, Berlin, 1984;
J.R. Holton, An Introduction to Dynamic Meteorology, Academic Press, New York, 1992;
S.A. Shapiro, P. Hubral, Elastic Waves in Random Media, Springer, Berlin, 1999;
M.S. Gazzaniga (Ed.), The New Cognitive Neurosciences, 2nd ed., MIT Press, Cambridge, MA, 2000.

[2] D. Regan, Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine, Elsevier, New York, 1989.

[3] T.R. Knosche, B. Maess, A.D. Friederici, Processing of syntactic information monitored by brain surface current density mapping based on MEG, Brain Topogr. 2 (12) (1999) 75–87;
C.H. Wolters, R.F. Beckmann, A. Rienaecker, H. Buchner, Comparing regularized and non-regularized nonlinear dipole fit methods: a study in a simulated sulcus structure, Brain Topogr. 1 (12) (1999) 3–18.

[4] P.L. Nunez, Toward a quantitative description of large-scale neocortical dynamic function and EEG, Behav. Brain Sci. 23 (2000) 371–437.

[5] Z.J. Koles, J.C. Lind, A.C.K. Soong, Spatio-temporal decomposition of the EEG: a general approach to the isolation and localization of sources, Electroenceph. Clin. Neurophysiol. 95 (1995) 219–230;
S.K. Law, P.L. Nunez, R.S. Wijesinghe, High-resolution EEG using spline generated surface Laplacians on spherical and ellipsoidal surfaces, IEEE Trans. Biomed. Eng. 40 (2) (1993) 145–153;
P.L. Nunez, B.M. Wingeier, R.B. Silberstein, Spatial–temporal structures of human alpha rhythms: theory, microcurrent sources, multiscale measurements, and global binding of local networks, Human Brain Mapping 13 (3) (2001) 125–164;
A. Fuchs, R. Friedrich, H. Haken, D. Lehmann, The spatio-temporal analysis of multichannel alpha EEG map series, in: H. Haken (Ed.), Computational Systems—Natural and Artificial, Springer, Berlin, 1987.

[6] D. Brandeis, D. Lehmann, C.M. Michel, W. Mingrone, Mapping event-related brain potential microstates to sentence endings, Brain Topogr. 8 (2) (1995) 145–159.

[7] D. Lehmann, W. Skrandies, Reference-free identification of components of checkerboard-evoked multichannel potential fields, Electroenceph. Clin. Neurophysiol. 48 (1980) 609–621.

[8] A. Flexer, H. Bauer, Discovery of common subsequences in cognitive evoked potentials, in: J.M. Zytkow, M. Quafafou (Eds.), Proceedings of the Second European Symposium on Principles of Data Mining and Knowledge Discovery, PKDD'98, Lecture Notes in Artificial Intelligence, vol. 1510, Springer, Berlin, 1998, pp. 309–317;
J. Wackermann, Towards a quantitative characterisation of functional states of the brain: from the non-linear methodology to the global linear description, Int. J. Psychophysiol. 34 (1999) 65–80.

[9] I. Pal, A. Horst, D. Lehmann, EEG segmentation in time using spatial criteria—micro-states and reaction-time, Electroenceph. Clin. Neurophysiol. 61 (3) (1985) 97.

[10] D. Lehmann, Auswertung des evozierten Potential- oder ereigniskorrelierten Potential-Mappings, vol. 23, EEG–EMG, Thieme Verlag, Stuttgart, New York, 1992, pp. 1–11.

[11] R.D. Pascual-Marqui, C.M. Michel, D. Lehmann, Segmentation of brain electrical activity into microstates: model estimation and validation, IEEE Trans. Biomed. Eng. 42 (7) (1995) 658–665;
C.M. Michel, G. Thut, S. Morand, A. Khateb, A.J. Pegna, R.G. de Peralta, S. Gonzales, M. Seeck, T. Landis, Electric source imaging of human brain functions, Brain Res. Rev. 36 (2–3) (2001) 108–118 (special issue).
[12] J.J. Wright, P.D. Bourke, C.L. Chapman, Synchronous oscillation in the cerebral cortex and object coherence: simulation of basic electrophysiological findings, Biol. Cybernet. 83 (2000) 341–353;
W.J. Freeman, Neurodynamics: An Exploration in Mesoscopic Brain Dynamics, Springer, London, 2000;
L. Ingber, Statistical mechanics of neocortical interactions: applications of canonical momenta indicators to electroencephalography, Phys. Rev. E 55 (1997) 4578–4593;
P.L. Nunez, Neocortical Dynamics and Human EEG Rhythms, Oxford University Press, Oxford, 1995;
V.K. Jirsa, H. Haken, Field theory of electromagnetic brain activity, Phys. Rev. Lett. 77 (5) (1996) 960–963;
P.A. Robinson, C.J. Rennie, J.J. Wright, H. Bahramali, E. Gordon, D.L. Rowe, Prediction of electroencephalographic spectra from neurophysiology, Phys. Rev. E 63 (2001) 021903;
T. Wennekers, Orientation tuning properties of simple cells in area V1 derived from an approximate analysis of nonlinear neural field models, Neural Comput. 13 (2001) 1721–1747;
T.D. Frank, A. Daffertshofer, P.J. Beek, H. Haken, Impacts of noise on a field theoretical model of human brain, Physica D 127 (1999) 233–249.
[13] H. Haken, H. Sauermann, Nonlinear interactions of laser modes, Z. Phys. 173 (1963) 261.
[14] A. Fuchs, J.A.S. Kelso, H. Haken, Phase transition in the human brain: spatial mode dynamics, Int. J. Bifurc. Chaos 4 (2) (1992) 917–939;
V.K. Jirsa, R. Friedrich, H. Haken, Reconstruction of the spatio-temporal dynamics of a human magnetoencephalogram, Physica D 89 (1995) 100–122.
[15] T.D. Frank, A. Daffertshofer, C.E. Peper, P.J. Beek, H. Haken, Towards a comprehensive theory of brain activity: coupled oscillator systems under external forces, Physica D 144 (2000) 62–86.
[16] C. Uhl, F. Kruggel, B. Opitz, D.Y. von Cramon, A new concept for EEG/MEG signal analysis: detection of interacting spatial modes, Human Brain Mapping 6 (1998) 137–149.
[17] H. Riedel, M. Granzow, B. Kollmeier, Single-sweep-based methods to improve the quality of auditory brain stem responses. Part II. Averaging methods, Z. Audiol. 40 (2) (2001) 62–85.
[18] C. Uhl, R. Friedrich, H. Haken, Reconstruction of spatio-temporal signals of complex systems, Z. Phys. B 92 (1993) 211–219;
C. Uhl, R. Friedrich, H. Haken, Analysis of spatiotemporal signals of complex systems, Phys. Rev. E 51 (5) (1995) 3890–3900.
[19] R. Friedrich, C. Uhl, Spatio-temporal analysis of human electroencephalograms: petit-mal epilepsy, Physica D 98 (1996) 171–182.
[20] C. Uhl, A. Hutt, F. Kruggel, Improvement of source localization by dynamical systems based modelling (DSBM), Brain Topogr. 13 (3) (2001) 219–226.
[21] A. Hutt, F. Kruggel, Fixed point analysis: dynamics of non-stationary spatiotemporal signals, in: S. Boccaletti, H.L. Mancini, W. Gonzales-Vias, J. Burguete, D.L. Valladares (Eds.), Space–Time Chaos: Characterization, Control and Synchronization, World Scientific, Singapore, 2001.
[22] A. Hutt, M. Svensen, F. Kruggel, R. Friedrich, Detection of fixed points in spatiotemporal signals by a clustering method, Phys. Rev. E 61 (5) (2000) R4691–R4693.
[23] A. Hutt, C. Uhl, R. Friedrich, Analysis of spatio-temporal signals: a method based on perturbation theory, Phys. Rev. E 60 (2) (1999) 1350–1358.
[24] A. Hutt, Methoden zur Untersuchung der Dynamik raumzeitlicher Dynamik, MPI Series in Cognitive Neuroscience, Max Planck Institute of Cognitive Neuroscience, Leipzig, Germany, 2001.
[25] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.
[26] M. Hožič, A. Stefanovska, Karhunen–Loève decomposition of peripheral blood flow signal, Physica A 280 (2000) 587–601.
[27] A. Fuchs, J.M. Mayville, D. Cheyne, H. Weinberg, L. Deecke, J.A.S. Kelso, Spatiotemporal analysis of neuromagnetic events underlying the emergence of coordinate instabilities, NeuroImage 12 (2000) 71–84.
[28] E. Walter, L. Pronzato, Identification of Parametric Models from Experimental Data, Springer, Berlin, 1997.
[29] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C: The Art of Scientific Computing, 2nd ed., Cambridge University Press, Cambridge, 1995.
[30] P.L. Nunez, R. Srinivashan, A.F. Westdorp, R.S. Wijesinghe, D.M. Tucker, R.B. Silberstein, P.J. Cadusch, EEG coherence. I. Statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales, Electroenceph. Clin. Neurophysiol. 103 (1997) 499–515.
[31] A. Hutt, F. Kruggel, R. Friedrich, Dynamische Zustände in raumzeitlichen Signalen, Anwendung auf simulierte Daten und ERPs, Verhandlungen der Deutschen Physikalischen Gesellschaft (IV) 35 (2000) DY46.38;
A. Hutt, F. Kruggel, C.S. Herrmann, Automatic detection of ERP-components and their modelling, J. Cognitive Neurosci. Suppl. 85 (2001);
A. Hutt, Spatiotemporal modelling of EEG/MEG, Brain Topogr. 14 (4) (2002) 347.
[32] F.H. Busse, K.E. Heikes, Science 208 (1980) 173.
[33] T.W. Picton, S.A. Hillyard, H.I. Krausz, R. Galambos, Human auditory evoked potentials. I. Evaluation of components, Electroenceph. Clin. Neurophysiol. 36 (1974) 179–190.

[34] N. Krauss, D.I. Smith, T. McGee, L. Stein, C. Cartee, Development of the middle latency response in animal model and its relation to the human response, Hearing Res. 27 (1987) 165–176.
[35] A. Hutt, Topology of MAEP-component P_a, in preparation.
[36] F. Grandori, M. Hoke, G.L. Romani (Eds.), Auditory Evoked Magnetic Fields and Electric Potentials, vol. 6, Advances in Audiology, Karger-Verlag, Basel, 1990.
[37] B. Thaler, Visual quantummechanics, 1998. http://www.kfunigraz.ac.at/imawww/vqm/pages/colorgallery/index.html.