A Tutorial on Data-Driven Methods for Statistically Assessing ERP Topographies

**Thomas Koenig · Maria Stein · Matthias Grieder ·
Mara Kottlow**

Abstract Dynamic changes in ERP topographies can be conveniently analyzed by means of microstates, the so-called *"atoms of thoughts",* that represent brief periods of quasi-stable synchronized network activation. Comparing temporal microstate features such as on- and offset or duration between groups and conditions therefore allows a precise assessment of the timing of cognitive processes. So far, this has been achieved by assigning the individual time-varying ERP maps to spatially defined microstate templates obtained from clustering the grand mean data into predetermined numbers of topographies (microstate prototypes). Features obtained from these individual assignments were then statistically compared. This has the problem that the individual noise dilutes the match between individual topographies and templates leading to lower statistical power. We therefore propose a randomization-based procedure that works without assigning grand-mean microstate prototypes to individual data. In addition, we propose a new criterion to select the optimal number of microstate prototypes based on cross-validation across subjects. After a formal introduction, the method is applied to a sample data set of an N400 experiment and to simulated data with varying signal-to-noise ratios, and the results are compared to existing methods. In a first comparison with previously employed statistical procedures, the new method showed an increased robustness to noise, and a higher sensitivity for more subtle effects of microstate timing. We conclude that the proposed method is well-suited for the assessment of timing differences in cognitive processes. The increased statistical power allows identifying more subtle effects, which is particularly important in small and scarce patient populations.

Keywords Microstates · Timing · Statistics ·
Randomization · Topography · Model selection

T. Koenig (✉) · M. Stein · M. Grieder · M. Kottlow
Department of Psychiatric Neurophysiology, University Hospital
of Psychiatry, University of Bern, Bern, Switzerland
e-mail: thomas.koenig@puk.unibe.ch

M. Stein
Department of Clinical Psychology and Psychotherapy, Institute
of Psychology, University of Bern, Bern, Switzerland

M. Kottlow
Institute of Pharmacology and Toxicology, University of Zurich,
Zurich, Switzerland

Introduction

Scalp recorded evoked potentials permit the non-invasive mapping of human brain functions at an excellent temporal resolution. This allows for the decomposition of complex cognitive processes into a sequence of processing stages, each with a different functional significance (Lehmann 1990; Murray et al. 2008). Importantly, an unequivocal distinction of ERP components originating from different brain regions can be obtained by comparing the topographies of scalp electromagnetic fields of the ERP (McCarthy and Wood 1985; Michel et al. 2009). By identifying and comparing ERP scalp topographies, it is thus possible to track changes of brain functional states, where a state is defined globally by a specific distribution of one or several simultaneously active brain regions. Spatial analysis of scalp electromagnetic fields (Lehmann and Skrandies 1984) has moreover the advantage of being reference independent, as topographic configurations are not influenced by a reference electrode (Lehmann 1987).

A commonly used way to compare multichannel ERP data between groups or conditions is to quantify the difference of the topography in a given time range and to test it for significance. Various such methods exist and have been proven to allow for a sound assessment of topographic differences in ERPs (Koenig et al. 2011; Lehmann 1987; Lehmann et al. 1993; Nishida et al. 2013; Strik et al. 1998). If the topography of a certain process is known, it is also possible to quantify the amount of ERP variance that can be attributed to this process and compare different datasets based on this quantifier (Brandeis et al. 1992).

Another approach to multichannel ERP analyses are various kinds of data driven spatio-temporal factor analyses, such as principal component analysis (PCA), independent component analysis (ICA), or as discussed in more detail below, cluster analysis. Factor analyses of multichannel ERP data describe an ERP as composed of a limited set of constant topographies, each with a specific time course. The comparison of ERPs among different groups or conditions is then primarily based on a comparison of the time-course of selected factors. A good overview of spatial factor analysis methods (PCA, ICA, microstates) in comparison to classical ERP approaches is provided by Pourtois et al. (2008).

While PCA and ICA were primarily based on statistical arguments such as independence among the factors, the rationale for using cluster analysis emerged from the observation of periods of stable field configurations typically separated by brief moments of rapid transitions (Lehmann 1990; Wackermann et al. 1993). These periods of quasi-stable field configurations were called microstates (Lehmann and Skrandies 1980). They offered a natural, data-driven and bottom-up definition of a brain functional state as a period where a quasi-stable field configuration was observed. Meanwhile, microstate analysis has become a widely accepted tool for the assessment of the sequence of functional states in ERPs (see Murray et al. 2008, for a review). Microstates could also be observed in the electrocorticogram of mice (Megevand et al. 2008). In addition, it is also possible to identify microstates in the ongoing resting EEG (Koenig et al. 2002; Lehmann 1990) and microstate analyses of single trial ERP data have been proven to be a sensitive and unique tool to track cognitive processes on a single subject level (De Lucia et al. 2010, 2012; Tzovara et al. 2012a, b, 2013).

Technically, ERP microstate analysis based on spatial clustering identifies a small set of prototypical ERP topographies that can be observed in the measured data (so called microstate class maps) and assigns each time period of the ERP to exactly one of these microstate class maps based on a best fit criterion (Murray et al. 2008; Pascual-Marqui et al. 1995). Whereas the microstate maps correspond to the forward solution of all sources contributing to a microstate class, the assignment step yields the time of the on- and offset of the microstates in the ERP. If this algorithm is used to identify microstates in data consisting of several experimental conditions or groups, the assignment can be used to identify differences in the timing of a given microstate class (i.e. onset, offset and duration), which is a very elegant and efficient way to exploit the information yielded by the high temporal resolution of the data.

On the level of statistics, the microstate analyses performed so far have been done by identifying the microstate maps in ERP datasets averaged over a group of subjects (grand mean ERPs), but the assignment was then done in the ERPs of the individuals. From this individual assignments, several parameters were extracted for a given microstate map, such as the variance explained by the map, the time when the first or last assignment to the map was observed, or the total number of time points assigned to the map. These individual parameters were then entered into classical, usually parametric, univariate test statistics such as t tests or ANOVAs (Michel et al. 2009).

While this approach has been applied successfully in a series of studies (Arzy et al. 2007; Chouiter et al. 2013; Darque et al. 2012; Knebel and Murray 2012; Kottlow et al. 2011; Kovalenko et al. 2012; Laganaro and Perret 2011; Overney et al. 2005; Pannekamp et al. 2011; Pegna et al. 1997; Perret and Laganaro 2012; Pourtois 2011; Spierer et al. 2007; Stevenson et al. 2012; Taha et al. 2013), it appeared to the authors that the method can still be improved to increase statistical power and decrease the effects of individual variance. Our criticism is that in the above described approach, the microstate maps are compared to data that has not been directly available to the clustering algorithm, which obviously impoverishes the amount of variance explained by the microstate maps. Furthermore, the individual data contains individual variance that is usually of little interest, but reduces the topographic similarity to the microstate maps. We suspect that this loss of similarity resulting from comparing microstate maps obtained in grand mean data to individual ERPs may negatively affect the resulting statistical power.

Our proposal is thus to develop a statistical test for microstate features where the assignment procedure remains on the level of the grand mean data. This is expected to improve the similarity between the microstate maps and the data these maps are assigned to, and thus increase the statistical power of the results. For this purpose, we will employ randomization techniques, which (although computationally expensive) allow custom-tailoring statistical tests to such specific problems.

A further aim of the paper is to propose a solution to the problem of selecting the appropriate number of microstate maps. This selection has so far been made on criteria

extracted from grand mean data (Pascual-Marqui et al. 1995), and the individual variance has been neglected. In general, the aim of model selection procedures (such as selecting a number of microstate maps) is to choose a model that captures as much of that part of the data that follows some generalizable rules, while it is oblivious to random noise. Our proposal is that in ERP microstate models, the generalizability of the model can be assessed by testing it's consistency across subjects; the parts of the data that can be observed independently of the individual subjects belong to the optimal microstate model, while those parts of the data that depend on the individual subjects should not be part of the model. The optimal model (i.e. the optimal number of microstate maps) should thus maximize the amount of explained variance that is independent of individual attributes. This criterion can be evaluated using cross-validation procedures across subjects (Devijver and Kittler 1982).

In the following methods and results sections, we will give a detailed explanation of the procedures and apply it to a real sample dataset and a series of simulated datasets with defined signal to noise ratios (SNRs). We will then also analyze the same dataset with the established methodology and compare the results.

As sample data set we chose data of healthy US American subjects staying in Switzerland for a German language exchange. EEG was measured while subjects performed a sentence reading task once at the beginning and once at a later phase of their stay (Stein et al. 2006). These sentences ended with semantically correct or incorrect endings. Incorrect versus correct sentence endings have been found to induce a so-called N400 effect which was described by (Kutas and Hillyard 1980).

Methods

Selection of the Optimal Microstate Model

As outlined in the introduction, we aimed to identify a microstate model that is sufficiently complex to accommodate the part of the data variance that is common across subjects, while avoiding to account for variance that appears to be tied to individual attributes. This type of problems is typically addressed using cross-validation, where models of different complexity are constructed based on a subset of the available data, and the resulting models are then used to make predictions for the remaining data. Therein, the optimal model is the one that minimizes the prediction error (Devijver and Kittler 1982).

In the context of microstate modeling, we propose to implement microstate model selection through cross-validation by computing microstate models with different numbers of microstate classes based on ERPs averaged over a subset of the subjects (training data). These microstate models are then tested for their predictive value (mean correlation) in the ERP's averaged over the subjects not included in the construction of the microstate model (test data). Since the mean correlation of the test data with a model will depend on the division of the data into training- and test-sets, this procedure has to be repeated with different, randomly created subsets of training and test data. For each number of microstates, the mean correlation of the test data with the model is then averaged across the results obtained in the different subsets. The optimal number of microstates is selected where this grand mean correlation is maximal.

Note that this procedure contains no measures to minimize the total number of microstates per se, but only minimizes the number of microstates that cannot be found consistently across subjects. The encountered number of microstates therefore does not represent something that necessarily generalizes across studies, but rather something that is optimally suited for a dataset with a limited size.

Computationally, the procedure is illustrated in Fig. 1 and is as follows:

1. The algorithm randomly subdivides the subjects into a training and a test dataset. If the subjects belonged to different groups, each dataset must contain members of all groups.

2. Grand mean ERPs are computed in the training and test datasets as a function of group and condition.

3. Spatio-temporal microstate models with different numbers of microstate maps are computed in the grand means of the training dataset. This model contains both the
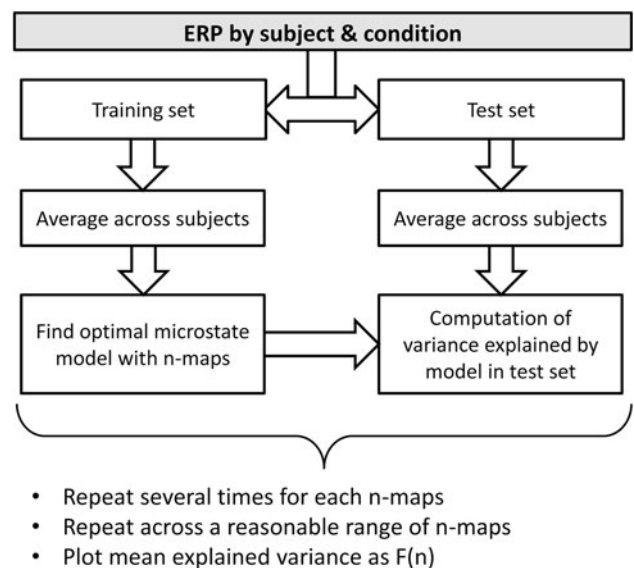

Fig. 1 Flow-chart illustrating the procedure for the selection of the optimal microstate model

topographies of the microstate maps as well as the time instances when these microstate maps are observed.

4. The mean correlation of the test data with each microstate model is computed (Eq. 1)

$$\text{Mean correlation} = \frac{\sum_{t=1}^{nt} Corr(V_t, T_t)}{nt} \tag{1}$$

where t is time, nt is the number of time points, $Corr$ is the correlation function, V_t is the voltage vector of the test data at time t, and T_t is the voltage vector of the microstate class observed in the training data at time t. If several conditions or groups are available, the mean correlation is computed in each condition and group and averaged.

5. Steps 1–4 are repeated for a sufficient number of times, and the mean correlations from each run are retained.

6. The mean correlations are averaged across repetitions and the number of microstate classes yielding the maximum mean correlation is identified. This represents the optimal number of microstate classes for the analysis of the given dataset.

7. The microstate templates with the optimal number of classes are now computed using the grand mean ERPs of all available subjects and conditions.

Once the optimal microstate model has been identified, we can proceed to the statistical evaluation of the experimental manipulations in the entire dataset.

Statistical Testing of Differences in Microstate Models

As in any statistical testing, an analysis of ERP microstate features needs to compare an effect (e.g. a difference in the onset of a given microstate class in the ERPs of two groups) against the distribution of this effect under the null-hypothesis. While in classical statistics, this distribution is estimated based on the variance of the individual data, and on assumptions about the nature of the distribution, randomization statistics determine this distribution based on simulations of the effect under the null hypothesis. For our purposes, the important point here is that with randomization statistics, we can simulate ERP data under the null-hypothesis and still compute grand mean ERPs, and therefore still assess microstate effects based on these grand means while the null-hypothesis is true.

In general, randomization based statistics consist of the following three steps (Manly 2007):

1. Quantification of an effect of interest in the measured data.
2. Creation of cases of the same quantifier compatible with the null hypothesis. This is achieved by repeatedly applying the quantifier to the measured data after randomizing it in a way that eliminates the suspected structure in the data.
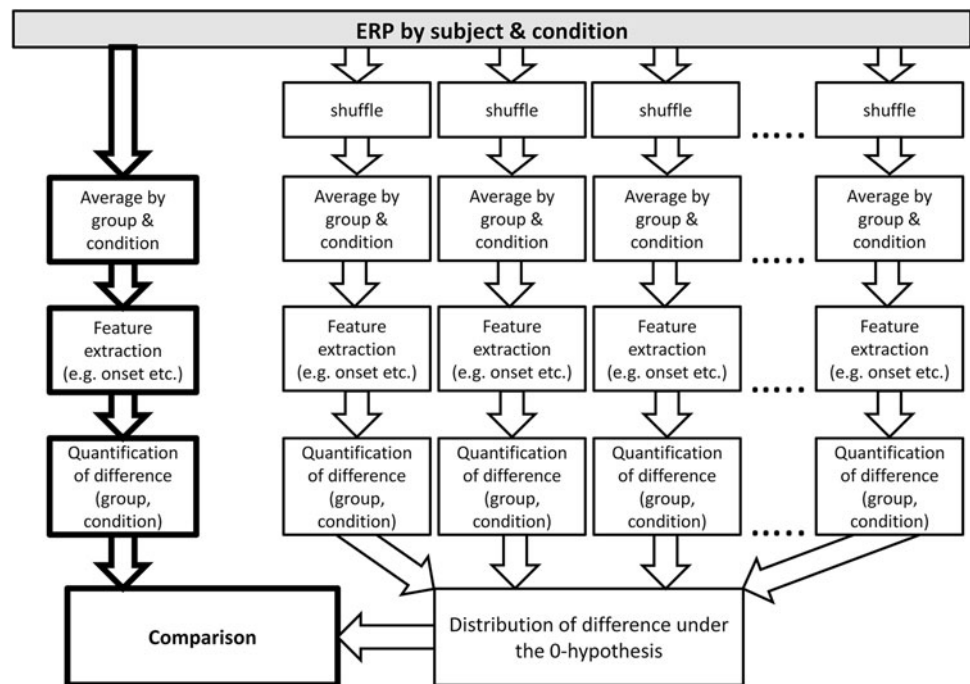3. Comparison of the distribution of the quantifier obtained in the real data with the distribution of the quantifier under the null-hypothesis.

We will follow this scheme for our microstate statistics, with the constraint that the assignment procedure shall always be applied on the level of the grand mean data. The proposed procedure is also illustrated in Fig. 2.

To quantify the effect of interest (step 1), we propose to use the previously employed features extracted from the established microstate assignment procedures (Murray et al. 2008). These features are specific for a given microstate map and for the given ERP and include, among others, the amount of variance explained by the map, the time point of the first (onset) or last (offset) assignment of the ERP to that map, or the count of time-points assigned to the maps. The important difference to the previously proposed method is that in our procedure, these features are extracted after the microstate maps have been assigned to group and/or condition specific grand mean data and not to the individual data. The quantifier of the effect of interest is then defined by the variance of the feature extracted from the different groups and/or conditions. For example, in an analysis of the onset of a language related microstate under two different conditions, the quantifier of the effect of interest could be defined as the difference of onset of the first occurrence of the language related microstate map between the two conditions (the difference here is equivalent to the variance of the two onsets). If we would hypothesize that the language related microstate systematically differs between three groups of subjects, our quantifier could for example be the variance among the onsets obtained from the grand means of each of the three groups.

For the creation of instances of the chosen quantifier under the null hypothesis (step 2), we propose to randomize the ERP data such that the possible suspected structure of interest in the data is eliminated. For example, if we suppose that semantically expected and unexpected sentence endings systematically lead to different responses in a group of subjects, we would construct data with two random conditions R1 and R2 and randomly assign, in each subject, the ERPs of expected sentence endings to either R1 or R2, and the ERPs of unexpected sentence endings to the remaining random condition. If we expected that two groups of subjects (e.g. good and weak learners) differ systematically, we would randomly shuffle the ERPs of each subject among the two groups. Once this randomization has been done, the random group and/or condition "specific" grand means ERPs can be computed, and the quantifier of interest can again be computed as above. The

Fig. 2 Flow-chart depicting the proposed statistical testing of the microstate models

important difference to the previously employed procedure is again that the microstate assignment necessary for the feature extraction is computed in grand mean data.

Finally, the quantifier obtained in the measured data in step 1 is compared to the distribution of the quantifier obtained under the null hypothesis (step 3). This is done by simple rank statistics, and the probability of the data being compatible with the null hypothesis is defined by the proportion of quantifiers obtained under the null-hypothesis that were larger or equal to the quantifier obtained in the real data. As an example, let us assume that our first example above, the difference of onset obtained from the randomized data was larger than the difference obtained from the real data in 7 out of 500 cases. The probability p that the observed difference is compatible with the null hypothesis is then $7/500 = 0.014$, which would (given an alpha-level of 0.05) indicate that it is significant. If the variance of the onset of the three groups obtained after randomizing the data would be larger than the variance obtained in the real data in 1,293 out of 5,000 randomization runs, the probability p that the observed group differences were obtained by chance is estimated to be $1,293/5,000 = 0.259$, which would typically be considered as not-significant. Note that the distribution of the quantifier under the null-hypothesis depends on the precise random permutations and assignments and may thus vary. The resulting p value is thus not an exact value, but an estimate.

The literature suggests that for a reliable rejection of the null-hypothesis on a 5 % level, 1,000 randomization runs are necessary, and for an estimate at the 1 % level, 5,000

randomization runs are recommended (Manly 2007). In contrast to parametric methods, statistical tests as the one described here are ultimately based on rank statistics. Therefore, they can be expected to be more robust against false positive results due to biases and outliers in individual data.

Sample Data Analysis and Simulations

The sample data and analysis are based on an experiment that has previously been used to demonstrate statistical procedures of the analysis of ERPs (Koenig et al. 2008, 2011). These data consist of ERPs recorded in 16 healthy young English-speaking exchange students that spent a year in the German-speaking part of Switzerland and that participated in a larger study on the neurobiology of training-related changes of the language system (Koenig et al. 2008; Stein et al. 2006). Participants passively viewed on a computer screen word-by-word presented German sentences with semantically expected or unexpected sentence endings. This is a typical setup to elicit the so-called N400; an ERP component that is associated with the violation of semantic expectancy and characterized by a parietal negativity peaking around 400 ms after stimulus presentation (Brandeis et al. 1995; Kutas and Hillyard 1980). Subjects were recorded twice, once at the beginning of their stay, and once after having lived about 3 months in Switzerland. The aim of the experiment was to track the progress of semantic integration in the acquired foreign language using an N400 paradigm. The measured data

consisted of 72 channel ERPs (0–1,000 ms time window, 250 Hz sampling rate 0.5–70 Hz band-pass) in four conditions, completing a 2×2 (expected vs. unexpected sentence ending by day 1 vs. day 2) within-subject factorial design.

To identify the optimal number of microstate prototype maps, the proposed cross validation procedure was applied 250 times, each time randomly splitting the 16 subjects into training and learning datasets of 8 subjects each, testing between 3 and 20 microstate classes. Every microstate identification run used the k-mean algorithm with 50 random initializations each. The resulting optimal model was used for the comparison of the existing statistical method and the novel randomization based microstate statistics. While it has been argued that other algorithms such as hierarchical clustering are somewhat better suited for the identification of microstate maps (Murray et al. 2008), we used the traditional k-mean algorithm here because other clustering algorithms were not yet implemented and tested in our software environment, and because all the procedures proposed here can be applied to any clustering algorithm.

For the evaluation of the statistical power of the new and the existing method, simulated data with different SNR were constructed. For this purpose, the condition-wise grand average ERP across all subjects was taken as signal, whereas the individual deviations from this grand average ERP were considered as noise. To obtain a dataset with a predefined SNR, this noise was scaled such that the ratio of the standard deviations of the signal and the scaled noise were equal to the chosen SNR. The dataset with the predefined SNR was then obtained by adding the scaled noise to the signal (Eq. 2).

$$V_{s,c,t,e} = M_{c,t,e} + \frac{(N_{s,c,t,e}) * sd(M_{c,t,e})}{sd(N_{s,c,t,e}) * R} \tag{2}$$

where $V_{s,c,t,e}$ is the simulated data with a SNR of R, $M_{c,t,e}$ is the grand mean across subjects, $N_{s,c,t,s}$ is the noise, sd is the standard deviation, s is the subject, c is the condition, t is time and e is the sensor. Simulated datasets were constructed with the following SNRs: 0.01, 0.1, 0.2, 0.5, 1, 2, 5, and 10. Note that these SNR values represent the strength of the grand mean ERPs against the variability of individual averaged ERPs, and not the strength of the individual averaged ERPs against the single trials, where SNRs are typically much higher.

For the sake of simplicity, the subsequent statistical analyses of these simulated datasets were limited to microstate classes associated to the well-established N400 effect. As dependent variables, we chose those the most common microstate features; these were microstate onset (time when the selected microstate map is observed first), microstate offset (time when the selected microstate map is observed last), map count (total number of time points a

selected microstate map is observed), mean GFP (mean Global Field Power of all time points assigned to the selected microstate map) and the center of gravity (center of gravity in time of the GFP of all time points assigned to the selected microstate map). To avoid a priori choices, no distinction was made if a microstate class was observed during a single or during several separate time periods, and microstate assignments were always based on the complete set of microstate class topographies.

The statistical analyses of the simulated data were done using two existing software packages where the new and the existing methodology have been implemented. For the computations based on the existing methodology, the CARTOOL program (Brunet et al. 2011) was used. The novel method has been integrated in the RAGU program (Koenig et al. 2011) developed by the authors. The rule for the microstate assignment is common for both methods; in an ERP map series the topography of each moment in time is compared with all microstate class topographies using the correlation coefficient across sensors. Each time point is then assigned to the microstate class that fitted best, i.e. that had the highest spatial correlation with the momentary ERP topography. However, while in CARTOOL, this is done independently for each subject and condition, the new methodology does this assignment based on data averaged across subjects.

In order to reduce statistical noise in the microstate parameters, very briefly occurring microstates were suppressed by a smoothing algorithm for categorical variables (10 point smoothing with a penalty factor of 3, for details, see Pascual-Marqui et al. (1995)). To exclude potential confounders from differences in the clustering algorithms, identical microstate prototype maps were used in both analyses.

The results of the CARTOOL assignment were tested for significance using repeated measures ANOVAs as implemented in Matlab, separately for each microstate map. When in a subject a microstate map was not encountered in the analysis window in all conditions, the data of that subject and map was excluded from the analysis for those features that could not be computed (i.e. those that included time in reference to the stimulus onset).

The implementation of microstate statistics in RAGU is based on the algorithm described above and directly yields the *p*-values. Therefore, no further statistical analysis was necessary. In addition, the estimation of significance is based on rank statistics, such that no tests for normality were necessary. To compare the statistical power of both methods, the obtained *p*-values were plotted as function of the SNR.

In order to illustrate the extent of correspondence of the proposed analysis approach with more conventional topographical ERP analyses, additional electrode-wise

time-point by time-point t tests on one effect of interest were conducted. Moreover, a topographic ANOVA (TANOVA) with the above described 2×2 within-subject-design (semantic expectation×day) as well as a randomization test using GFP with the same design were computed. These analyses compared, time-point by time-point, the topography (TANOVA-analysis) and the strength (GFP-analysis) of the momentary grand-mean maps of the different conditions. For the topographic analysis, a generalized measure of map differences was used (Koenig et al. 2011), the GFP analysis employed the difference of GFP of the same maps. These indices of differences of topography and strength were then tested for statistical significance by comparing them against the distribution of those measures under the null-hypothesis as obtained by randomizing the individual ERPs across conditions 5000 times. For further details, see also Koenig et al. (2011).

Results

The resulting values of the mean correlations in the test datasets are shown in Fig. 3. It becomes apparent that there is a considerable variation in mean correlation, which is driven by the random divisions of the subjects into learning- and test-sets. It is also apparent that the mean correlation first increases with increasing number of microstates to reach a plateau where no further predictive power is gained by more complex models. The maximum grand mean correlation was reached when using 10 microstate classes, which was therefore the number of microstate classes chosen for the remaining analyses.

The resulting optimal 10-microstate model and its assignment to the grand-mean data are shown in Fig. 4. The model shows many of the typical ERP components expected in a visual N400 experiment (Brandeis et al. 1995), and substantial differences between expected and unexpected sentence endings. For the assessment of the N400 effect, the microstate classes 6 and 8 were used, since their latencies and topographies corresponded best to the N400. As Fig. 4 illustrates, microstate class 6 was observed during unexpected sentence endings, while microstate class 8 was observed during expected sentence endings.

The results of the statistical analyses are shown in Fig. 5. As the figure shows, both methods yielded the predicted significant main effects of expectancy with SNR ratios of at least 1. However, the new method appeared to be in general less sensitive to noise, because it also yielded significant main effects at lower SNRs, and also it detected interactions of day and expectancy in microstate class 8 at SNRs of 1 and below. The SNR of the real data was 0.77, such that the new method would clearly have detected more differences than the old one at a p-value of 0.05.
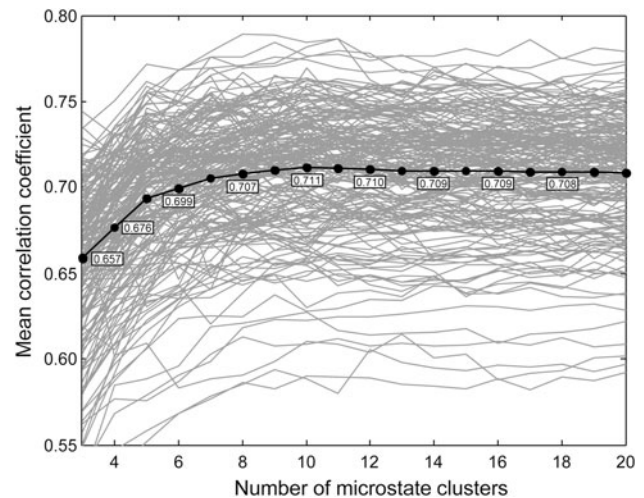

Fig. 3 Results of the cross-validation of microstate models with different numbers of clusters. The *grey lines* indicate the mean correlation coefficients (*vertical axis*) of the spatio-temporal models constructed with different training dataset, applied to their complementary test-dataset as function of the number of microstate classes (*horizontal axis*). The *bold black line* indicates the grand mean correlation across 250 cross-validation runs. The best grand mean correlation was obtained using 10 microstate classes

Furthermore, with the new method, the p-value always declined with an increasing SNR, which was not the case with the existing method. This may suggest that the assessment of significance is more robust with the new method. Another important difference between the two methods is that in some cases, the new method gave no statistical output. In this analysis, this was the case for the interaction of expectancy and day in microstate class 6. This happens when in the grand-mean data, the microstate class of interest was not observed in one or more condition and no differences between conditions could be computed.

The results of the presented method obtained in the original data yielded, for microstate class 6 (the topography primarily associated with the unexpected sentence ending), a significantly later on- and offset and center of gravity in the correct condition (p always smaller than .05). Microstate class 8 (primarily associated with the correct sentence endings) showed an inverse pattern, with an earlier onset, offset and center of gravity in the correct condition (p-values always smaller than .005). Interestingly, microstate class 8 also showed an interaction of expectancy and day in the onset; after the expected sentence endings, this microstate occurred earlier in day 2 than day 1, which points at a learning effect.

The result of the electrode-wise t tests between the unexpected and expected stimulus condition of day 2 showed mainly that within a time window of 300 and 800 ms after the target word, many of the ERP amplitudes
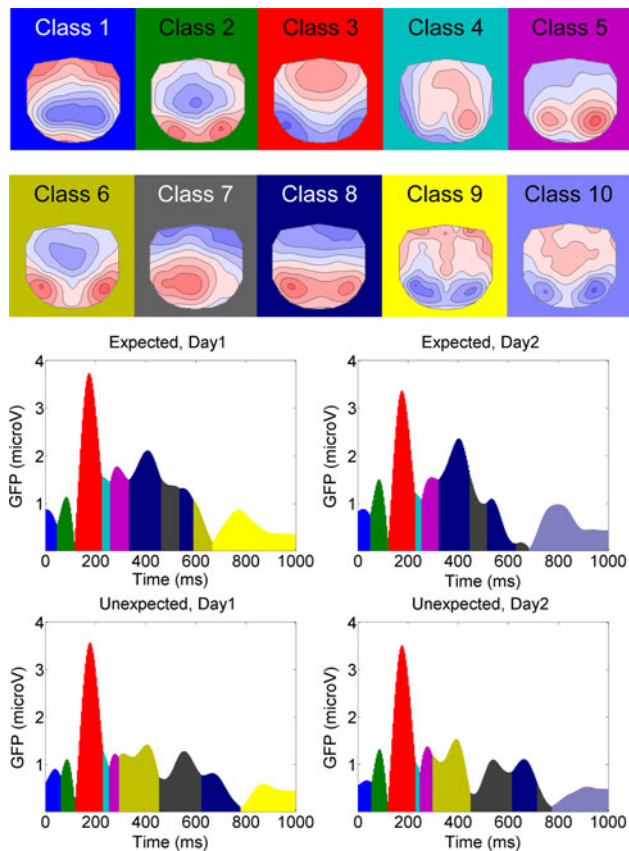
Fig. 4 Upper part: The optimal 10 microstate maps. *Blue map areas* indicate negative, *red areas* positive values; all maps have been scaled to have a GFP of one. Different *background colors* are used to differentiate the microstate classes. In the lower part, the assignment of the 10 microstate classes to the grand mean data of the four conditions is shown as function of time (*horizontal axis*). The color of the areas indicates the microstate class; the height on the *vertical axis* indicates the GFP of the ERP (Color figure online)

differed (Fig. 6). The significant results of the microstate analysis therefore correspond, as expected, to amplitude differences in the ERPs.

As illustrated in Fig. 7, a topographic difference in expectancy was found in the TANOVA within a similar time frame (approx. 250–800 ms). This time frame covers the on- and offset times of microstate classes 6 and 8 that also showed significant main effects of day. Furthermore, a topographic main effect of day (not considered in the microstate analyses) was detected at about 220–250 ms, and there was an interaction of expectancy and day at 600–620 ms. This interaction in the TANOVA is in a period where microstate class 8 that also showed an expectancy by day interaction. The randomization test of GFP showed expectancy differences around 400 and 800 ms and a main effect of day at approximately 560–600 ms. No significant interactions were detected.

Discussion

As expected, the established individual assignment procedure and the introduced randomization based microstate statistics yielded reasonably similar results when investigating the effect of the violation of semantic expectancy under conditions with a high SNR. However, the individual assignment procedure identified a smaller number of significant effects in data simulated with lower SNR ratios. Similar conclusions can be drawn based on the assessment of the interaction of expectancy and day.

The new procedure thus seems to improve the statistical power for the sake of the applicability of the model to the individual data. In general, this gain in statistical power is welcome for two reasons. Firstly, it may help to detect also smaller effects that would not have been detected otherwise. In the sample analysis, this was the case for the interaction, which is probably the most interesting effect in this dataset, because it implies learning. Secondly, the gain in statistical sensitivity allows conducting sufficiently powerful analyses in groups with fewer subjects. This is especially interesting in populations where subjects are difficult to recruit and to measure, namely in psychiatric populations or when studying development and aging (Grieder et al. 2012). This is of major interest as timing and sequence effects of specific microstate classes have been reported in these populations (Kikuchi et al. 2007; Koenig et al. 1999; Lehmann et al. 2005; Nishida et al. 2013) and further evidence for these effects also during tasks would be particularly important.

What may be the reason for the discrepancy in the sensitivity of the two methods, and what implications follow? An answer can be found when we consider the microstate assignment procedure as a data reduction step with some loss of information: The full topographic information of the map is being reduced to a labeling, which as a consequence reduces the comparisons among maps from a continuous and parametric range of similarity or dissimilarity to a binary statement of same or different. The essential difference between the two procedures is, in our view, that in the individual assignment procedure, this data reduction takes place at the level of single conditions and subjects, and averaging across subjects is done using the extracted features. In contrast, in the randomization procedure, the data reduction is applied only after all the averaging has been done on the level of topographies. This implies that while the individual assignment procedure becomes "unaware" of the metrics of topographic similarities between maps, the randomization procedure is "unaware" of the presence of the extracted features in the individual data. These different blind spots of the two procedures have specific implications that we briefly discuss here.
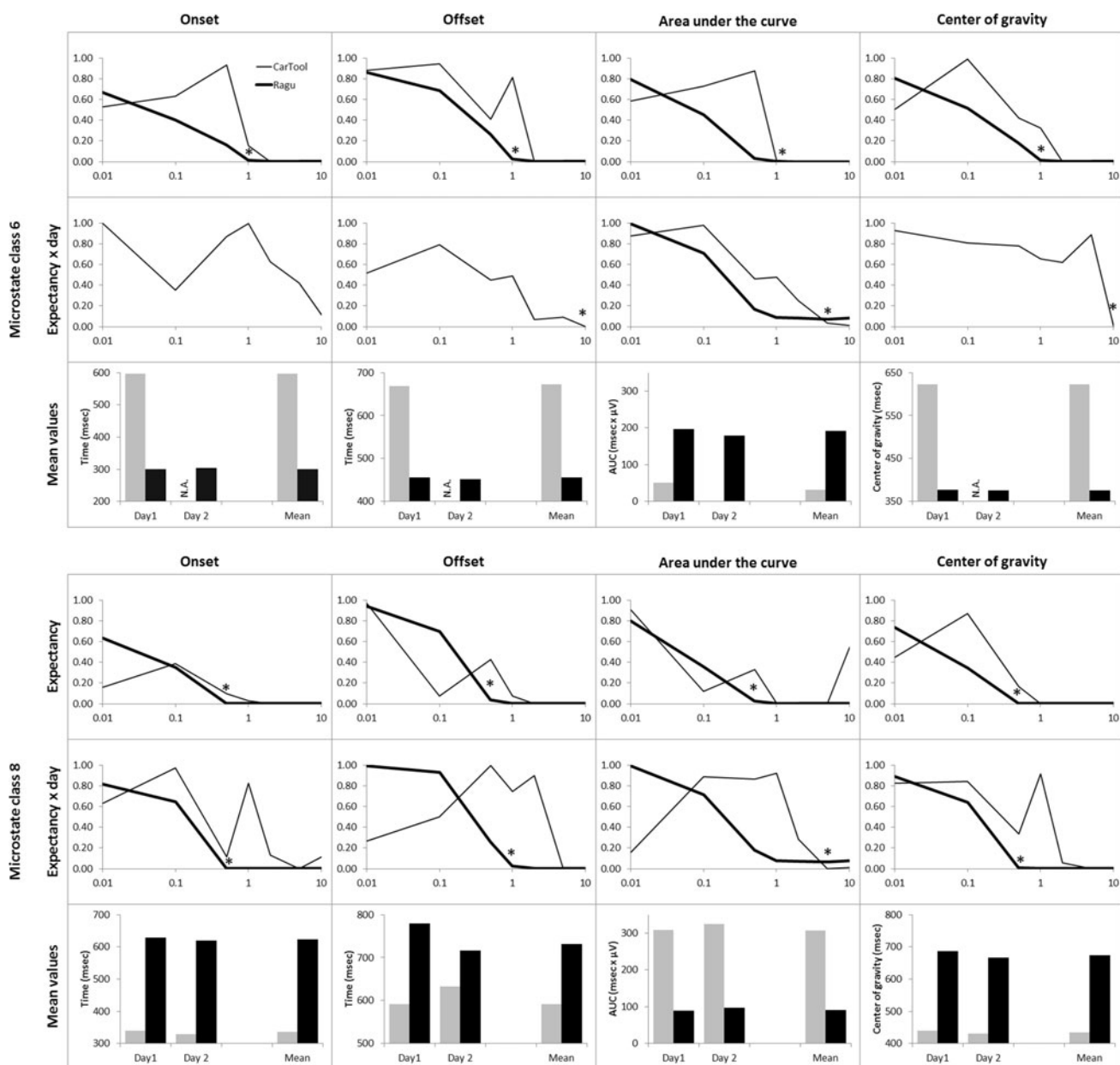
Fig. 5 Comparison of the *p*-values of the different microstate measures obtained with the two procedures in the data with simulated SNRs, as currently implemented in CARTOOL (assignment of individual data) and RAGU (randomization based approach). Results of the main effect of semantic expectancy (N400 effect) and the interaction of day and expectancy are shown for microstate class 6 and 8. The *vertical axes* indicate *p*-values; the *horizontal axes* indicate the SNRs of the data (*logarithmic scale*). The *asterisks* indicate the SNR where one of the methods first reached a *p*-value below 5 %. In addition, there are *bar graphs* showing the mean values for correct (*gray bars*) and false sentence endings for day 1 and 2 and for the average of the 2 days. *NA* not available

In the individual assignment procedure, the loss of quantitative information about map differences before averaging may label two individual maps as different, while that difference is relatively small. Two differently labeled individual maps may thus still have a large amount of communality, and this communality may well contain significant information. The fact that in the sample analysis, the individual assignment method failed to identify the interactions may be explained by this problem, especially

also because these effects occurred in periods of relatively low GFP, where the SNR is typically lower, and common map features are more likely to be obscured by noise. The "blind spot" of the randomization based microstate statistics is of a different type. Because no microstate assignments are made on the individual level, there is no information about the possibility that a difference in microstate assignments can actually be observed in the individual data. In other words, the procedure enhances the
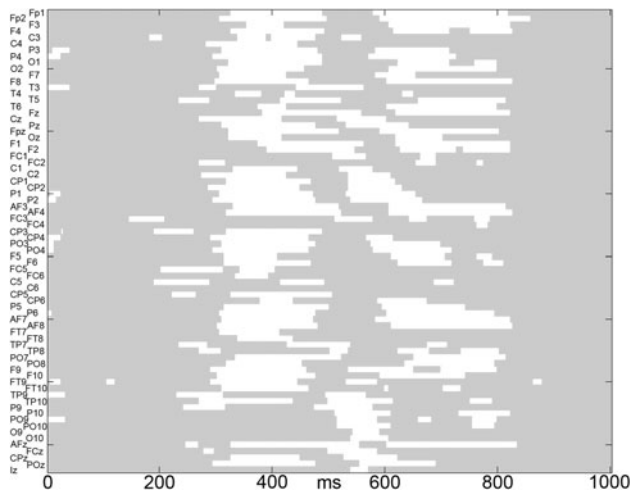
Fig. 6 Plot of the electrode (*vertical axis*) by time (*horizontal axis*) matrix of *t* tests between unexpected and expected target words at day 2. *White* indicates a significant amplitude difference ($p > 0.05$), *grey* reflects no significant difference

sensitivity of the analysis beyond a point that remains reproducible on an individual level. The obtained effect may thus be interesting from a theoretical perspective, but have limited value for each subject. This is a problem if the use of microstate analyses is intended for the classification of single subjects or trials (De Lucia et al. 2010, 2012; Tzovara et al. 2012a, b, 2013). Similarly, if correlations with a continuous variable are to be computed, grand means become meaningless and the method presented here cannot be applied.

The differences between the two approaches become apparent also when we look at what happens if a microstate class is never observed in one condition. Measures containing time information in reference to an event can then obviously not be computed. In the individual assignment procedure, data where this is the case is typically excluded, which is however a manipulation of the data that may have systematic effects on the resulting statistics. In the randomization procedure, the absence of a microstate class in

the grand mean of a relevant condition leads to the result that no statistics on microstate latencies can be computed at all (Fig. 5).

In order to facilitate the usage of the new methodology, it has been integrated into the existing and freely available RAGU software (Koenig et al. 2011). RAGU has been developed under Matlab, allowing for a usage across different operating systems. Furthermore, it implements a series of other analysis methods such as TANOVAs or the topographic consistency test (Koenig and Melie-Garcia 2010) and various tools for visualization, such that users can easily compare and integrate results obtained with the different methods as suggested above.

From a conceptual point of view, the microstate analysis as presented here is a complement to analysis strategies that compare spatial topographies (Lehmann 1987; Lehmann et al. 1993; Strik et al. 1998) as for example the TANOVA provided by the RAGU software (Koenig et al. 2011). These analyses are usually anchored to invariant time windows and assess an effect in terms of variations of map topographies (Michel et al. 2009; Murray et al. 2008) in a constant time period. Microstate analyses are complementary to this approach in the sense that they are anchored to a constant set of map topographies (the microstate class maps) and search for variations in the time periods where these maps occur. It is thus in our opinion advisable to employ both methods to find the most comprehensible description of an effect. On one side, a simple time shift of a microstate due to an experimental manipulation will yield a quite complicated pattern of topographic differences that may obscure the initial, rather simple nature of an effect. On the other side, small topographic changes may not result in a change of microstate class assignment, and may thus go undetected by a microstate analysis. Thus, the microstate approach complements the TANOVA approach by segregating cognitive processes into different sub-processes represented by specific, constant topographies. Further, it provides information on the sequences and timing (onset, offset and duration) of these

Fig. 7 Left panel: color-coded *p*-values of the TANOVA plotted for each time point (*horizontal axis*) of the main effects of expectancy and day and their interaction. Right panel: Analog to the left panel, but based on the GFP. *Blue to green to yellow colors* reflect non-significant topographic or GFP differences respectively; *orange to red areas* highlight significantly differing topographies or GFP (see *color legend*) (Color figure online)
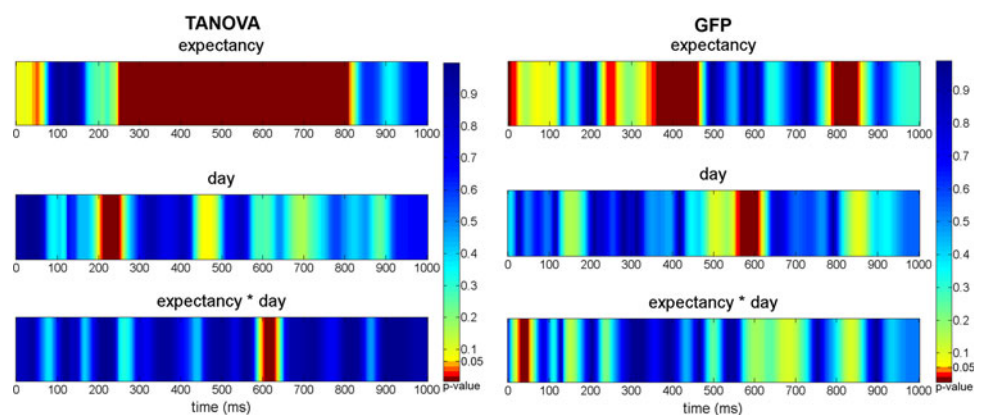
microstate classes. The TANOVA in turn compensates for the simplification introduced by modeling the data by a sequence of non-overlapping and constant topographies. In a comprehensive analysis of an ERP experiment, the two analysis strategies (topographic comparisons and microstates) should however yield compatible and converging conclusions. Finally, since both approaches typically consider exclusively the topography and not the strength of the ERP, an exhaustive analysis of the data should be complemented by an analysis of the GFP, which considers exclusively the strength and not the topography of the scalp field data.

References

Arzy S, Mohr C, Michel CM, Blanke O (2007) Duration and not strength of activation in temporo-parietal cortex positively correlates with schizotypy. Neuroimage 1:326–333. doi:10.1016/j.neuroimage.2006.11.027

Brandeis D, Naylor H, Halliday R, Callaway E, Yano L (1992) Scopolamine effects on visual information processing, attention, and event-related potential map latencies. Psychophysiology 3:315–336

Brandeis D, Lehmann D, Michel CM, Mingrone W (1995) Mapping event-related brain potential microstates to sentence endings. Brain Topogr 2:145–159

Brunet D, Murray MM, Michel CM (2011) Spatiotemporal analysis of multichannel EEG: CARTOOL. Comput Intell Neurosci. doi:10.1155/2011/813870

Chouiter L, Dieguez S, Annoni JM, Spierer L (2013) High and low stimulus-driven conflict engage segregated brain networks, not quantitatively different resources. Brain Topogr. doi:10.1007/s10548-013-0303-0

Darque A, Del ZM, Khateb A, Pegna AJ (2012) Attentional modulation of early ERP components in response to faces: evidence from the attentional blink paradigm. Brain Topogr 2:167–181. doi:10.1007/s10548-011-0199-5

De Lucia M, Michel CM, Murray MM (2010) Comparing ICA-based and single-trial topographic ERP analyses. Brain Topogr 2:119–127. doi:10.1007/s10548-010-0145-y

De Lucia M, Tzovara A, Bernasconi F, Spierer L, Murray MM (2012) Auditory perceptual decision-making based on semantic categorization of environmental sounds. Neuroimage 3:1704–1715. doi:10.1016/j.neuroimage.2012.01.131

Devijver PA, Kittler J (1982) Pattern recognition: a statistical approach. Prentice-Hall, London

Grieder M, Crinelli RM, Koenig T, Wahlund LO, Dierks T, Wirth M (2012) Electrophysiological and behavioral correlates of stable automatic semantic retrieval in aging. Neuropsychologia 50(1):160–171. doi:10.1016/j.neuropsychologia.2011.11.014

Kikuchi M, Koenig T, Wada Y, Higashima M, Koshino Y, Strik W, Dierks T (2007) Native EEG and treatment effects in neuroleptic-naïve schizophrenic patients: time and frequency domain approaches. Schizophr Res 97(1-3):163–172. doi:10.1016/j.schres.2007.07.012

Knebel JF, Murray MM (2012) Towards a resolution of conflicting models of illusory contour processing in humans. Neuroimage 3:2808–2817. doi:10.1016/j.neuroimage.2011.09.031

Koenig T, Melie-Garcia L (2010) A method to determine the presence of averaged event-related fields using randomization tests. Brain Topogr 3:233–242. doi:10.1007/s10548-010-0142-1

Koenig T, Lehmann D, Merlo MC, Kochi K, Hell D, Koukkou M (1999) A deviant EEG brain microstate in acute, neuroleptic-naive schizophrenics at rest. Eur Arch Psychiatry Clin Neurosci 4:205–211

Koenig T, Prichep L, Lehmann D, Sosa PV, Braeker E, Kleinlogel H, Isenhart R, John ER (2002) Millisecond by millisecond, year by year: normative EEG microstates and developmental stages. Neuroimage 1:41–48. doi:10.1006/nimg.2002.1070

Koenig T, Melie-Garcia L, Stein M, Strik W, Lehmann C (2008) Establishing correlations of scalp field maps with other experimental variables using covariance analysis and resampling methods. Clin Neurophysiol 6:1262–1270. doi:10.1016/j.clinph.2007.12.023

Koenig T, Kottlow M, Stein M, Melie-Garcia L (2011) Ragu: a free tool for the analysis of EEG and MEG event-related scalp field data using global randomization statistics. Comput Intell Neurosci. doi:10.1155/2011/938925

Kottlow M, Praeg E, Luethy C, Jancke L (2011) Artists' advance: decreased upper alpha power while drawing in artists compared with non-artists. Brain Topogr 4:392–402. doi:10.1007/s10548-010-0163-9

Kovalenko LY, Chaumon M, Busch NA (2012) A pool of pairs of related objects (POPORO) for investigating visual semantic integration: behavioral and electrophysiological validation. Brain Topogr 3:272–284. doi:10.1007/s10548-011-0216-8

Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207:203–205

Laganaro M, Perret C (2011) Comparing electrophysiological correlates of word production in immediate and delayed naming through the analysis of word age of acquisition effects. Brain Topogr 1:19–29. doi:10.1007/s10548-010-0162-x

Lehmann D (1990) Brain electric microstates and cognition: the atoms of thought. In: John ER Vol. Machinery of the mind. Birkhäuser, Boston, pp 209–224

Lehmann D (1987) Principles of spatial analysis. In: Gevins A, Remond A (eds) Methods of analysis of brain electrical and magnetic signals: handbook of electroencephalography and clinical neurophysiology, vol 1. Elsevier, Amsterdam, pp 309–354

Lehmann D, Skrandies W (1980) Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroencephalogr Clin Neurophysiol 6:609–621

Lehmann D, Skrandies W (1984) Spatial analysis of evoked potentials in man—a review. Prog Neurobiol 3:227–250

Lehmann D, Wackermann J, Michel CM, Koenig T (1993) Space-oriented EEG segmentation reveals changes in brain electric field maps under the influence of a nootropic drug. Psychiatry Res 4:275–282

Lehmann D, Faber PL, Galderisi S, Herrmann WM, Kinoshita T, Koukkou M, Mucci A, Pascual-Marqui RD, Saito N, Wackermann J, Winterer G, Koenig T (2005) EEG microstate duration and syntax in acute, medication-naive, first-episode schizophrenia: a multi-center study. Psychiatry Res 2:141–156. doi:10.1016/j.pscychresns.2004.05.007

Manly BFJ (2007) Randomization. Bootstrap and Monte Carlo Methods in Biology. Chapman & Hall, Boca Raton

McCarthy G, Wood CC (1985) Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. Electroencephalogr Clin Neurophysiol 3:203–208

Megevand P, Quairiaux C, Lascano AM, Kiss JZ, Michel C (2008) A mouse model for studying large-scale neuronal networks using EEG mapping techniques. Neuroimage 42(2):591–602. doi:10.1016/j.neuroimage.2008.05.016

Michel C, Koenig T, Brandeis D (2009) Electrical neuroimaging in the time domain. In: Michel CM, Koenig T, Brandeis D, Gianotti LRR and Wackermann J, Vol. Electrical Neuroimaging, Cambridge, pp 111–143

Murray MM, Brunet D, Michel CM (2008) Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr 4:249–264. doi:10.1007/s10548-008-0054-5

Nishida K, Morishima Y, Yoshimura M, Isotani T, Irisawa S, Jann K, Dierks T, Strik W, Kinoshita T, Koenig T (2013) EEG microstates associated with salience and frontoparietal networks in frontotemporal dementia, schizophrenia and Alzheimer's disease. Clin Neurophysiol 6:1106–1114. doi:10.1016/j.clinph.2013.01.005

Overney LS, Michel CM, Harris IM, Pegna AJ (2005) Cerebral processes in mental transformations of body parts: recognition prior to rotation. Brain Res Cogn Brain Res 3:722–734. doi:10.1016/j.cogbrainres.2005.09.024

Pannekamp A, van der Meer E, Toepel U (2011) Context- and prosody-driven ERP markers for dialog focus perception in children. Brain Topogr 3–4:229–242. doi:10.1007/s10548-011-0194-x

Pascual-Marqui RD, Michel CM, Lehmann D (1995) Segmentation of brain electrical activity into microstates: model estimation and validation. IEEE Tran Biomed Eng 7:658–665. doi:10.1109/10.391164

Pegna AJ, Khateb A, Spinelli L, Seeck M, Landis T, Michel CM (1997) Unraveling the cerebral dynamics of mental imagery. Hum Brain Mapp 5(6):410–421. doi:10.1002/(SICI)1097-0193(1997)5:6<410:AID-HBM2>3.0.CO;2-6

Perret C, Laganaro M (2012) Comparison of electrophysiological correlates of writing and speaking: a topographic ERP analysis. Brain Topogr 1:64–72. doi:10.1007/s10548-011-0200-3

Pourtois G (2011) Early error detection predicted by reduced pre-response control process: an ERP topographic mapping study. Brain Topogr 4:403–422. doi:10.1007/s10548-010-0159-5

Pourtois G, Delplanque S, Michel C, Vuilleumier P (2008) Beyond conventional event-related brain potential (ERP): exploring the time-course of visual emotion processing using topographic and principal component analyses. Brain Topogr 20(4):265–277. doi:10.1007/s10548-008-0053-6

Spierer L, Tardif E, Sperdin H, Murray MM, Clarke S (2007) Learning-induced plasticity in auditory spatial representations revealed by electrical neuroimaging. J Neurosci 20:5474–5483. doi:10.1523/JNEUROSCI.0764-07.2007

Stein M, Dierks T, Brandeis D, Wirth M, Strik W, Koenig T (2006) Plasticity in the adult language system: a longitudinal electrophysiological study on second language learning. Neuroimage 33(2):774–783. doi:10.1016/j.neuroimage.2006.07.008

Stevenson RA, Bushmakin M, Kim S, Wallace MT, Puce A, James TW (2012) Inverse effectiveness and multisensory interactions in visual event-related potentials with audiovisual speech. Brain Topogr 3:308–326. doi:10.1007/s10548-012-0220-7

Strik W, Fallgatter AJ, Brandeis D, Pascual-Marqui RD (1998) Three-dimensional tomography of event-related potentials during response inhibition: evidence for phasic frontal lobe activation. Electroencephalogr Clin Neurophysiol 4:406–413

Taha H, Ibrahim R, Khateb A (2013) How does arabic orthographic connectivity modulate brain activity during visual word recognition: an ERP study. Brain Topogr 2:292–302. doi:10.1007/s10548-012-0241-2

Tzovara A, Murray MM, Michel C, De Lucia M (2012a) A tutorial review of electrical neuroimaging from group-average to single-trial event-related potentials. Dev Neuropsychol 6:518–544. doi:10.1080/87565641.2011.636851

Tzovara A, Murray MM, Plomp G, Herzog MH, Michel CM, De Lucia M (2012b) Decoding stimulus-related information from single-trial EEG responses based on voltage topographies. Pattern Recogn 6:2109–2122

Tzovara A, Rossetti AO, Spierer L, Grivel J, Murray MM, Oddo M, De Lucia M (2013) Progression of auditory discrimination based on neural decoding predicts awakening from coma. Brain 1:81–89. doi:10.1093/brain/aws264

Wackermann J, Lehmann D, Michel CM, Strik WK (1993) Adaptive segmentation of spontaneous EEG map series into spatially defined microstates. Int J Psychophysiol 3:269–283