

# Topographic ERP Analyses: A Step-by-Step Tutorial Review

Micah M. Murray · Denis Brunet · Christoph M. Michel

Accepted: 13 February 2008 / Published online: 18 March 2008  
© Springer Science+Business Media, LLC 2008

**Abstract** In this tutorial review, we detail both the rationale for as well as the implementation of a set of analyses of surface-recorded event-related potentials (ERPs) that uses the reference-free spatial (i.e. topographic) information available from high-density electrode montages to render statistical information concerning modulations in response strength, latency, and topography both between and within experimental conditions. In these and other ways these topographic analysis methods allow the experimenter to glean additional information and neurophysiologic interpretability beyond what is available from canonical waveform analyses. In this tutorial we present the example of somatosensory evoked potentials (SEPs) in response to stimulation of each hand to illustrate these points. For each step of these analyses, we provide

the reader with both a conceptual and mathematical description of how the analysis is carried out, what it yields, and how to interpret its statistical outcome. We show that these topographic analysis methods are intuitive and easy-to-use approaches that can remove much of the guesswork often confronting ERP researchers and also assist in identifying the information contained within high-density ERP datasets.

**Keywords** Electroencephalography (EEG) · Event-related potentials (ERPs) · Topography · Spatial · Reference electrode · Global field power · Global dissimilarity · Microstate segmentation

## Introduction

This tutorial review has been predicated by a growing interest in the use of EEG and ERPs as a neuroimaging technique capable of providing the experimenter not only with information regarding *when* experimental conditions differ, but also *how* conditions differ in terms of likely underlying neurophysiologic mechanisms. There is an increasing appreciation of the fact that EEG and ERPs comport information beyond simply the time course of brain responses or “components” that correlate with a psychological/psychophysical parameter. They can identify and differentiate modulations in the strength of responses, modulations in the latency of responses, modulations in the underlying sources of responses (*vis à vis* topographic modulations), as well as combinations of these effects. Moreover, this information is attainable with sub-millisecond temporal resolution. Our focus here is on providing a tutorial for how to extract such information with minimal experimenter bias and to test such information statistically.

---

M. M. Murray (✉) · D. Brunet · C. M. Michel  
Electroencephalography Brain Mapping Core, Center for Biomedical Imaging of Lausanne and Geneva, Radiologie CHUV BH08.078, Bugnon 46, Lausanne, Switzerland  
e-mail: micah.murray@chuv.ch

M. M. Murray  
The Functional Electrical Neuroimaging Laboratory,  
Neuropsychology and Neurorehabilitation Service, Vaudois University Hospital Center and University of Lausanne,  
46 rue du Bugnon, 1011 Lausanne, Switzerland

M. M. Murray  
The Functional Electrical Neuroimaging Laboratory, Radiology Service, Vaudois University Hospital Center and University of Lausanne, 46 rue du Bugnon, 1011 Lausanne, Switzerland

D. Brunet · C. M. Michel  
Functional Brain Mapping Laboratory, Department of Fundamental and Clinical Neuroscience, University Hospital and University Medical School, 24 Rue Micheli du Crest, 1211 Geneva, Switzerland

Researchers using EEG/ERPs might find themselves daunted by the sheer quantity of data that is now routinely acquired (e.g. 64–256 channels and amplifiers that digitize data simultaneously from all channels at rates from 500 Hz upwards) and perhaps also by the myriad names of ERP components appearing in the literature (e.g. [33] for a recent overview). In the case of a 64-channel ERP investigation with epochs spanning 100 ms pre-stimulus to 900 ms post-stimulus at 500 Hz digitalization, there would be a data matrix containing 32,000 values (which could be further expanded if examined in terms of its spectral decomposition). If one were to assume complete independence of the measurements as a function of time and space/electrode (which is not the case and thus makes this issue all the more problematic because simple Bonferroni correction is inadequate; e.g. [20] for discussion), then by stochastic processes alone 160 of these 32,000 values would meet the 0.05 criterion of statistical significance if the experimenter were to compare *all* of these data from the two experimental conditions. Bearing this in mind, how then should the experimenter choose which of the data to analyze, given the necessity for data reduction in EEG/ERP research, while also avoiding the possibility that the data they analyze are among the 160 values that significantly differ by chance? It should be mentioned that this is not a problem unique to EEG/ERPs. Researchers working with fMRI datasets must also confront this and related issues, which have been most notably addressed by the authors of Statistical Parametric Mapping (SPM; [13]; <http://www.fil.ion.ucl.ac.uk/spm>). A prevailing and even recommended approach in EEG/ERP research has been for the experimenter to *a priori* select time periods or components of interest (often based on hypotheses generated from prior studies) as recorded at a chosen subset of electrodes (e.g. [22, 33]). For example, in a set of published guidelines for conducting ERP research [53, p. 141] proposed that “the simplest approach is to consider the ERP waveform as a set of waves, to pick the peaks (and troughs) of these waves, and to measure the amplitude and latency at these deflections.” Aside from the experimenter bias inherent to this approach, there are several additional weaknesses of analyzing ERP voltage waveforms that render the results arbitrary and of severely limited (neurophysiologic) interpretability. For example, an *a priori* focus on one or a few components of interest leads to the possibility that other (earlier) time periods and effects are overlooked, such as during periods of low-amplitude in a given waveform (e.g. [54, 55]). The spatio-temporal analysis methods that we summarize here can render a far more complete and informative interpretability without any *a priori* bias on certain time periods or scalp locations. Such is not to suggest that these methods cannot be incorporated with purely hypothesis-driven research. In the case of emotion

processing, for example, the experimenter may be testing the hypothesis that negative emotional stimuli are processed more quickly and via a more efficient neural circuit (i.e. different generators) than positive or neutral stimuli. As will be shown below, canonical analyses of voltage waveforms present serious pitfalls and limitations when addressing such questions/hypotheses.

We would be remiss to not immediately acknowledge several prior works that have either introduced or over-viewed many of the methods/issues we shall describe here. Most important among these is the seminal works of Dietrich Lehmann and his scholars (e.g. [3, 27, 31, 32, 37, 38, 60, 61]), though several others are also noteworthy [11, 12, 15].

We have organized this tutorial in the following way. First, we discuss the limitations and pitfalls of canonical waveform analyses, providing some concrete examples. Afterwards we detail the procedures for each step of electrical neuroimaging. In each section, we have attempted to introduce the theoretical basis of and to explain in simple terms and with mathematically simple examples how topographic analyses can be conducted, what information they yield, and how this information can be statistically analyzed and interpreted. While this tutorial discusses the case of somatosensory evoked potentials in order to give the reader an intuitive example based on known underlying neurophysiology, the methods presented here can be readily applied to issues in emotion research and cognitive neuroscience in general (see [55] for an example as well as studies cited throughout this tutorial for applications to specific questions in sensory-cognitive processing). Topographic mapping of the EEG end ERP is often a precursor to source localization. It is therefore of crucial importance that the topography of the scalp electric field is properly analyzed and interpreted before attempting to localize the underlying brain sources. This review focuses on the analysis of the topography of the scalp electric field. Readers interested in a more in-depth coverage of the issue of source localization are referred to [2, 23, 40]. All analyses presented in this tutorial have been conducted using CarTool software (<http://brainmapping.unige.ch/cartool.htm>).

### Data for this Tutorial

The data we use here to illustrate the electrical neuroimaging analyses are a subset from a previously published study demonstrating partially segregated functional pathways within the somatosensory system [6], and full details concerning the paradigm and data acquisition can be found therein. We provide only the essential details here.

Continuous EEG was acquired from six healthy participants at 512 Hz though a 128-channel Biosemi ActiveTwo

AD-box (<http://www.biosemi.com>) referenced to the common mode sense (CMS; active electrode) and grounded to a passive electrode. Peri-stimulus epochs of EEG (−96 ms pre-stimulus to 488 ms post-stimulus onset) were averaged separately for stimulation of the left and right hand and for each participant. For the present tutorial only non-target, distracter trials from the “what” condition were included (c.f. Table 1 of [6]). In addition to a  $\pm 100\mu\text{V}$  artifact rejection criterion, EEG epochs containing eye blinks or other noise transients were excluded. Prior to group-averaging, data at artifact electrodes from each subject were interpolated using a spherical spline interpolation [52]. Likewise, data were baseline corrected, using the pre-stimulus period, and band-pass filtered (0.68–40.0 Hz).

Tactile stimuli were square wave pulses (300 ms duration; 44,100 Hz sampling) presented through Oticon bone conduction vibrators (Oticon Inc., Somerset, NJ) with  $1.6\text{ cm} \times 2.4\text{ cm}$  surfaces. Two spatial positions (one in the left hemisphere and one in the right hemisphere) and two vibration frequencies (22.5 and 110 Hz) were used. Stimulus delivery and behavioral response recording was controlled by E-prime (Psychology Software Tools, Inc., Pittsburgh, PA; <http://www.pstnet.com/eprime>).

As will become clear below, the specific data we used are not particularly crucial for the points and methods we wish to illustrate here. The utility of comparing SEPs to left and right hand stimulation is that this is an intuitive example with clear neurophysiologic underpinnings in terms of known somatotopic cortical organization.

### Waveform Analyses: Limitations and Pitfalls

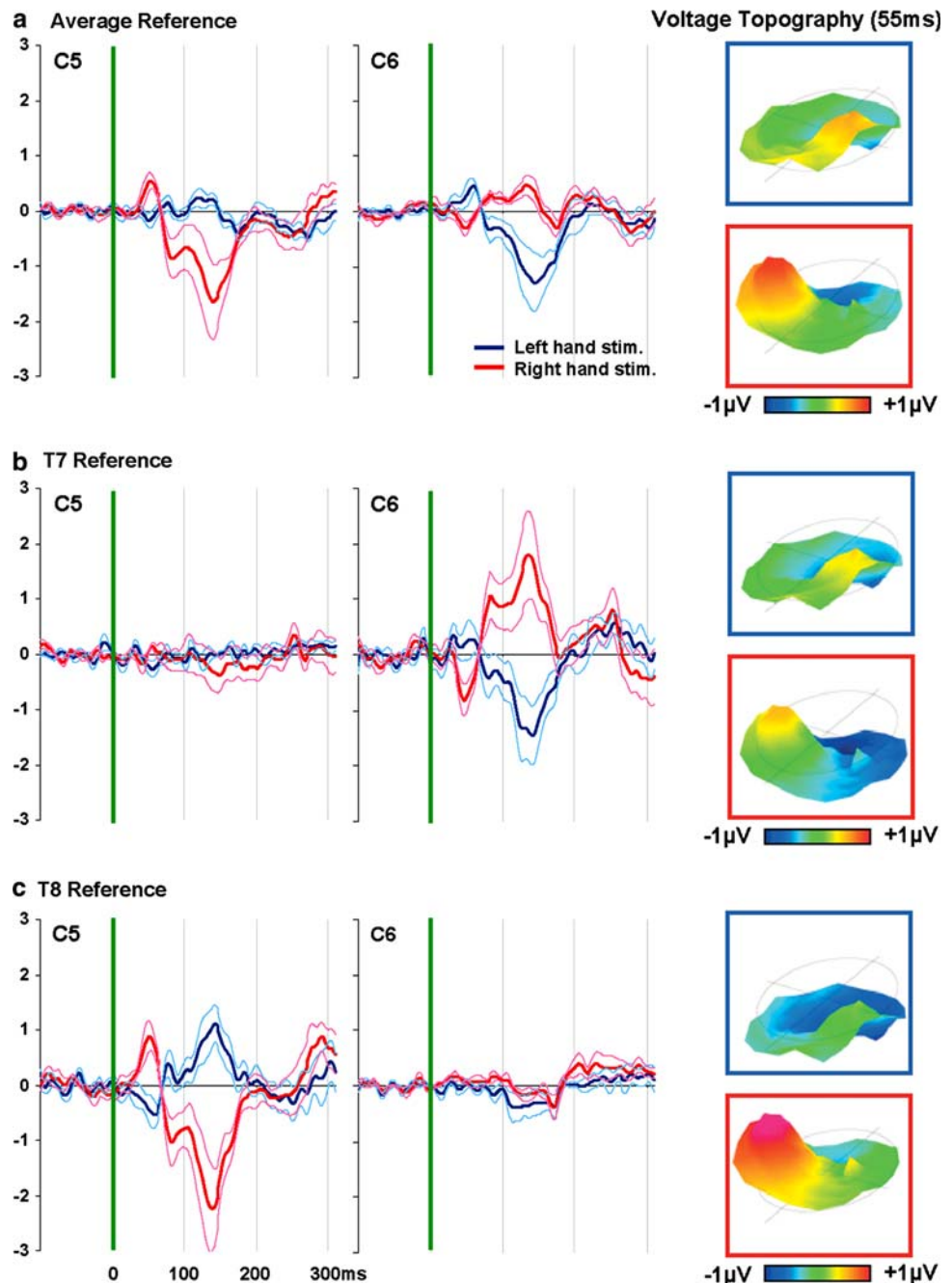
The core limitation (and pitfall) of analyzing voltage ERP waveforms is that they are reference-dependent. Although there is a long history of viewpoints concerning the ‘best’ or ‘appropriate’ reference (e.g. [8, 9, 48], see also [40] for a more recent discussion that includes the role of the reference in source estimations), it will *always* remain a choice and therefore a source of bias introduced by the experimenter. More important is the fact that this choice will critically impact the (statistical) outcome the experimenter observes when analyzing waveform measurements and by extension the interpretation of the data. This section therefore illustrates precisely these points.

Figure 1 displays group-averaged ERPs (s.e.m. indicated) to vibrotactile stimulation of the left and right hand

**Table 1** Some dependent measures obtainable from the ‘fitting’ procedure and their interpretability

Dependent measure	What’s actually being quantified	Interpretation of a map × condition interaction	Considerations & caveats
Number of data points labeled with a given template map (a.k.a. frequency of map presence or map duration)	For each subject, the number of data points, over a specified time period, when a template map yields the highest spatial correlation value over other template maps	Different template maps best represent the experimental conditions. When appropriate, post-hoc contrasts should be conducted	A higher spatial correlation does forcibly not translate into a significantly higher spatial correlation (see below). For example, although 96% is higher than 95%, it does not mean that these values significantly differ. Results of the TANOVA analysis can help address this issue
GEV of a given template map	For each subject, the GEV over a specific time period of a given template map for a given condition (see Appendix I for formula)	Different template maps better “explain” each condition	When this analysis is conducted under competitive fitting conditions (i.e. different maps are vying for labeling of the same time point), a given map might not always label the data from each subject and condition. In this case, missing values should be appropriately handled in the statistical analysis
First onset/last offset of a given template map	The time point in each subject’s data from a given condition when a given template map yields the highest spatial correlation (relative to other template maps being fit) for the first/last time over a specified time period	A latency shift and/or duration difference between conditions in their components or functional microstates, as identified by different template maps	There is no <i>a priori</i> restriction on whether or not the template map onsets and remains the preferentially fit map. That is, only the first onset and last offset are quantified, rather than whether or not there were multiple onsets and offsets over a given time period

**Fig. 1** The effect of the reference electrode. Group-averaged ERP waveforms are displayed in response to left-hand and right-hand somatosensory stimulation (blue and red traces, respectively). Cyan and pink traces indicate the s.e.m. for these group-average ERPs. Panels a–c depict the ERPs as measured from electrodes C5 and C6 when different reference locations are used. The reader should note the change in waveform shape, in the magnitude of the measured s.e.m., and in the presence/absence of differences between conditions. The right-hand side of the figure depicts the voltage topography of these data at 55 ms post-stimulus onset. The reader should note that although the color value ascribed to a given location changes with the choice of the reference (indicated by the projected axis and equator), the form of the topography is reference-independent



(blue and red traces, respectively) as recorded from electrodes at scalp locations C5 and C6, using standard electrode position nomenclature [46]. More specifically, panels a–c of this figure display the ERPs when different reference channels are selected (average reference, T7 reference (to emulate the left mastoid/earlobe), and T8 reference (to emulate the right mastoid/earlobe), respectively). The reader should note several points from this figure. First, the shape of the ERP waveforms changes with different reference electrodes. A given peak/trough might appear or disappear. Second, the variance around the ERP

(and by extension the s.e.m.) changes with different reference electrodes. Third, the latency and electrode(s) at which significant differences are obtained between conditions changes with different reference electrodes.

In Fig. 1a, differences between ERPs to left-hand and right-hand stimuli are observed at both electrodes C5 and C6 beginning at ~40 ms. One might interpret this as a bilateral and approximately equi-opposite effect with larger responses to left-hand stimuli at electrode C6 and to right-hand stimuli at electrode C5. From such, one might further conclude that each hemisphere responds in opposite ways

to stimulation of each hand. In Fig. 1b, with T7 as reference, such differences are observed only at electrode C6. One might interpret this as a right-lateralized effect with no evidence for response differences at electrodes over the left hemiscalp. From such, one might conclude that only the right hemisphere responds to tactile stimuli of the hands. Finally, in Fig. 1c, with T8 as reference, differences are observed only at electrode C5. One might interpret this as a left-lateralized effect with no evidence for response differences at electrodes over the right hemiscalp. From such, one might conclude that only the left hemisphere responds to tactile stimuli of the hands.

Which of the patterns of results shown in Fig. 1 and their subsequent interpretations is ‘correct’? While all are equally ‘correct’ from a statistical perspective, where and when responses to left-hand and right-hand stimulation are differentially processed cannot be unequivocally addressed by the above analyses of ERP waveforms, as the very presence of a given component in an ERP waveform as well as its modulation across experimental conditions is entirely reference-dependent. While the above example shows the data for the average reference and emulations of lateralized references to clearly illustrate our point, the caveats we describe apply to *any* reference location (vertex, nose, etc.). Even if it is “customary” for a given ERP community or lab to use one reference over another, the abovementioned analytical and interpretational pitfalls will remain present. That is, the obtained waveform shape and statistical result only apply for that chosen reference.

A related issue with canonical analyses of voltage waveforms concerns the interpretation of condition  $\times$  electrode interactions observed in an analysis of variance (ANOVA). This issue has been extensively treated by McCarthy and Wood [36] who rightly pointed out how this analysis cannot differentiate modulations in topography from modulations in amplitude when data are not first scaled. In particular, they presented three distinct scaling methods. One involves identifying the instantaneous maximum and minimum for each condition and subtracting the minimum value as well as the difference between the maximum and minimum from each electrode. A second involves scaling by a pre-defined value (see [24]); the shortcomings of which are detailed in [36]. The third, which they (and we) favor, involves dividing the value at each electrode by the instantaneous global field power (GFP; see below); a procedure that they refer to as vector scaling. The methods proposed by McCarthy and Wood [36] are routinely referred to and often applied/recommended [53]. As such, it is worthwhile to mention some important caveats with how this method has been applied (see also [10]). The first is that McCarthy and Wood’s [36] approach is only valid when the data from the *entire* electrode montage is included in the ANOVA; a practice nowadays rarely performed. A second,

intertwined and inescapable caveat is that ANOVAs based on data from voltage waveforms will *always* be reference-dependent even when scaled. We return to the issue of statistically identifying topographic modulations in a reference-independent manner, below.

From this example the reader should gain a sense of the severe limitations and pitfalls of analyzing ERP voltage waveforms. We have provided the example of somatosensory processing, but this could readily be extrapolated to more cognitive functions such as attention, language, and emotion, as well as the components typically ascribed to them. For some, this “reference-dependent” attribute has been viewed as the principal shortcoming of EEG versus magnetoencephalography (MEG; [50, 69, 70]). However, as will be shown throughout this tutorial and elsewhere in this special issue [55], alternative and easy-to-use analyses can be performed on EEG/ERP data (as well as MEG/MEF data) that do allow researchers to address fundamental neurophysiologic questions.

The first step for making EEG/ERP analyses more informative is to identify a reference-independent measure. We would direct the reader to the right-sided portion of Fig. 1 where the voltage topography at 55 ms post-stimulus onset is shown in response to stimulation of the left hand and right hand (blue and red frames, respectively). The projected axis and equator indicate the 0  $\mu$ V plane (i.e. the reference). As before, the reader should note several points by comparing topographies when different reference channels are used. First, changing the reference shifts vertically the position of the 0  $\mu$ V plane. Second and of far greater importance, the shape of the topography remains constant even though the color ascribed to a given position changes (see also Fig. 3 in [40]). That is, the configuration of the electric field at the scalp (i.e. the topographic map) is reference-independent [14, 31]. To provide the reader with a more everyday example, the shape of a mountain range remains constant even if the altitude of sea level (i.e. the reference elevation) were to change [31]. A third point that should be taken from this figure is that the topographies of responses to left-hand and right-hand stimuli differ and differ in a reference-independent manner. As will be shown below, the extent of topographic similarity or dissimilarity can be quantified and statistically tested. Of more importance is the fact that topographic differences have direct neurophysiologic interpretability. Changes in the topography of the electric field at the scalp can only be caused by changes in the configuration of the underlying intracranial sources (given the exclusion of artifacts such as eye movements, muscle activity, etc.), though the converse need not be the case [12, 31, 67].

We would be remiss not to mention that in an effort to obtain reference-free waveforms some have advocated the analysis of so-called current source density (CSD) or

Laplacian waveforms<sup>1</sup> over their voltage counterparts (e.g. [41, 45, 59, 68]). This procedure is undoubtedly beneficial in that it indeed eliminates the reference-dependent problem inherent to voltage waveforms (as well as contributions of volume conduction within the plane of the scalp) and is a suitable alternative for those researchers more accustomed to handling waveform data. However, CSD waveforms considered in isolation do not in and of themselves provide information concerning the underlying neurophysiologic mechanism(s) giving rise to a modulation between experimental conditions. In addition, the CSD is not readily calculated at the border of the electrode montage, and the CSD is generally more sensitive to the level of noise in the data. Finally, the experimenter would still be faced with the choice of which CSD waveforms and time periods to analyze.

### Why We Use the Average Reference

In the above we have highlighted the caveats of reference-dependent measurements. However, EEG/ERP requires the use of a reference. So, which one should be used? We advocate the use of a common average reference [32] for the following reason. Inverse solution methods (i.e. methods to reconstruct the intracranial sources of surface-recorded data) recalculate the data to a common average reference. This is because of the biophysical assumption of quasi-stationarity—i.e. that the net source activity at each instant in time within the brain sums to zero. Because the reference electrode adds a constant potential value to the value recorded at each electrode and instant in time, a “re-centering” of the data (i.e. a removal of this constant value) is necessary before applying an inverse solution so as to avoid violating the above quasi-stationarity assumption. Mathematically, this is equivalent to calculating the average reference of the surface-recorded EEG/ERP [48].

When using the average reference, it is therefore important to have adequate sampling of the electric field at the scalp. Discussions of how many electrodes and appropriate inter-electrode distances are outside the scope of this tutorial review (see e.g. [30, 65] for treatments of this issue). However, the relatively low cost of EEG equipment makes high-density montages accessible to most laboratories. Another important issue when using the

average reference, performing the analyses detailed here, and estimating intracranial sources, is how to cope with artifact-contaminated channels. This applies to both the single-subject and group-averaged data. Values at such channels are typically interpolated (see [40] for discussion for different methods). Likewise, group-averaged data require normalization to the same electrode configuration/positions before averaging [53].

It is also worthwhile to mention a common misunderstanding in how the average reference should be computed. Under typical experimental conditions, the recording reference is often discarded. However, the data at this location (provided it is near the brain and not elsewhere on the body surface or even off the body) is nevertheless a valid sampled value of the brain’s electric field at that location and as such should be included in the electrode montage and data analyses, being ascribed a value of 0  $\mu\text{V}$  as a function of time in all the formulae, including in the calculation of the average reference (see Appendix I).<sup>2</sup> Once the data have been recalculated to the average-reference, the reference electrode is just another electrode within the montage with a measurement of potential varying as a function of time (see topographic representations in Fig. 1).

### Global Field Power: A Single, Reference-Independent Measure of Response Strength

We now return to the kinds of neurophysiologic information we wish to extract from the EEG/ERP data, beginning with response strength. In the above, we detailed the pitfalls and limitations of analyzing ERP voltage waveforms due to their being dependent on the choice of the reference electrode(s). Global Field Power (GFP), by contrast, constitutes a single, reference-independent measure of response strength. GFP was first introduced by Lehmann and Skrandies [32] and has since become a commonplace measure among MEG users. Mathematically, GFP equals the root mean square (RMS) across the average-referenced electrode values at a given instant in time. More simply,

<sup>1</sup> For readers less familiar with CSD derivations, it is perhaps worthwhile to briefly describe what is being calculated that makes them reference-independent. The CSD or Laplacian derivation involves calculating the 2nd spatial derivative across the electrode montage (i.e. the degree of change of the degree of change in the voltage measured at electrode  $X$  relative to its neighbors). In this way, CSD derivations are intrinsically based on *spatial* gradients in the electric field at the scalp.

<sup>2</sup> Including the reference in the average reference calculation is very much desired, but becomes problematic if the reference has been placed far away from the brain. In some montages, a nose reference is already somewhat problematic because its inclusion in the average reference calculation will add an extreme point without any surrounding electrodes. In such cases, the experimenter might prefer to take it out of the montage and calculate the average reference from the remaining electrodes, particularly when a small number of electrodes are used. The reason is that the influence of the one “unknown” value (i.e. the reference) on the other known values is  $1/(\# \text{ of electrodes})$ . Of more critical importance is that the reference itself is not artifact-contaminated (e.g. eye movements, cardiac activity, etc.).

GFP is the standard deviation of all electrodes at a given time (see Appendix I). In the case of ERPs, the resultant GFP waveform is a measure of potential ( $\mu\text{V}$ ) as a function of time. GFP can be assessed statistically using approaches common to ERP research (e.g. time point by time point; area measures, peak measures, etc.). It is important to recall, however, that because GFP is a non-linear transformation, the GFP of the group-average ERP is *not* equivalent to the mean GFP of the single-subject ERPs. Thus, experimenters should exercise caution when visually inspecting and/or displaying GFP waveforms for group-averaged data.

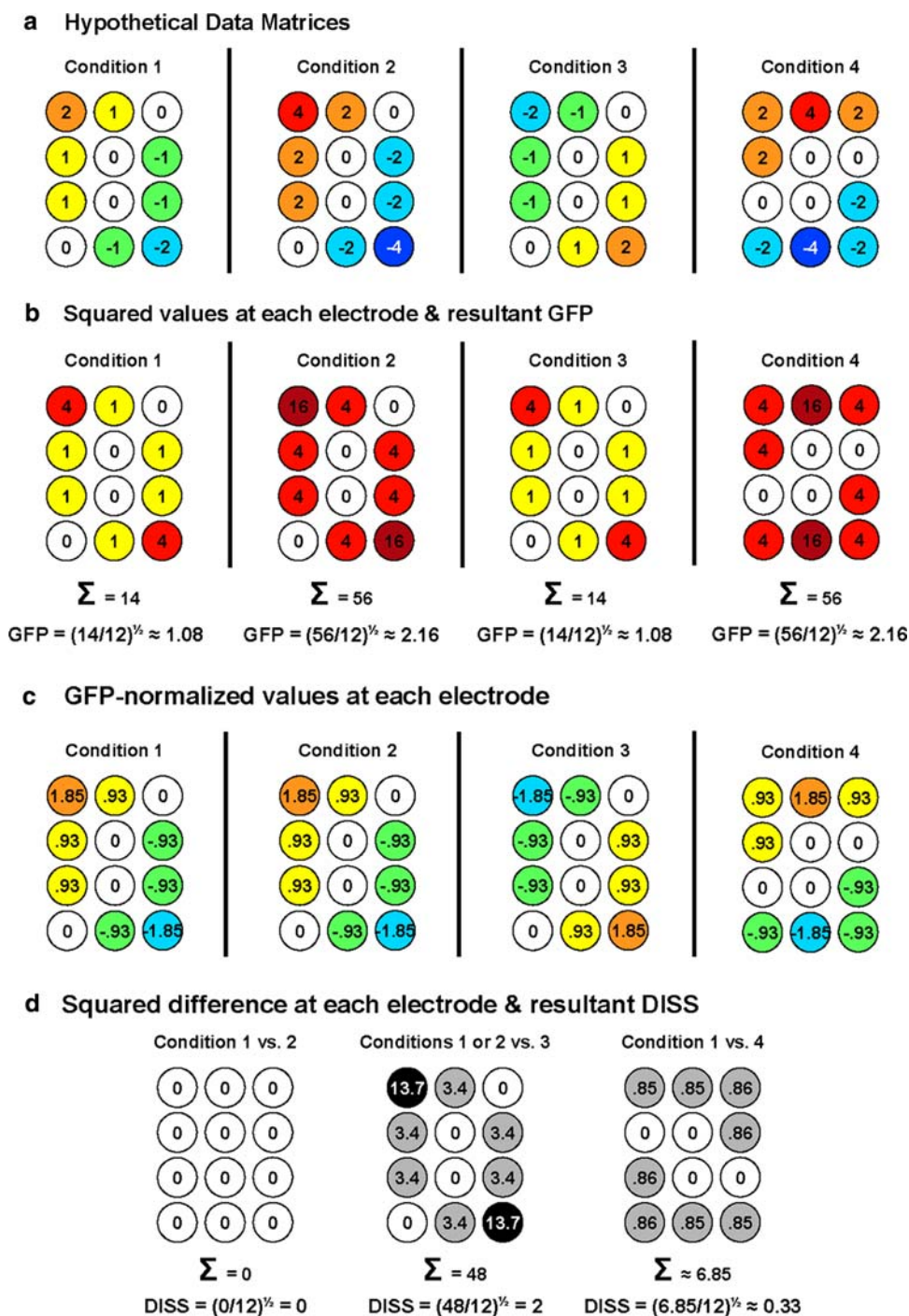
What GFP tells the researcher is on average across the electrode montage how strong a potential is being recorded. What GFP does *not* tell the researcher is any information about how this potential is distributed across the electrode montage—i.e. where large and small potentials were measured. These points are illustrated in Fig. 2, which displays four hypothetical data matrices (i.e. the potential values recorded from 12 electrodes at a given latency). The four conditions differ in the follow ways. Condition 2 is precisely twice that of Condition 1 at each electrode, resulting in an identical spatial distribution of values that are simply stronger in Condition 2. Condition 3 is the mathematical inverse of Condition 1 (i.e. the value at each electrode was multiplied by  $-1$ ), thereby resulting in a different spatial distribution (i.e. topography) of the same values (i.e. the frequency of each value/color is identical). Note that Condition 3 is included to illustrate an extreme case that is unlikely under typical experimental conditions. Condition 4, by contrast, represents a more typical observation in that it varies in both strength and topography from Condition 1. Figure 2b displays the squared value of these potentials at each electrode, the sum of these values across electrodes, and the resultant GFP. Note that while Conditions 1/3 have the same GFP and Conditions 2/4 have the same GFP, Conditions 1/3 have a GFP half that of Conditions 2/4. As such, it is important to note that the observation of a GFP modulation does not exclude the possibility of a contemporaneous change in the electric field topography. Nor does it rule out the possibility of topographic modulations that nonetheless yield statistically indistinguishable GFP values. For example, in the case of the somatosensory ERPs presented above, there is no evidence of a reliable GFP difference between responses to left-hand and right-hand stimulation (Fig. 3a). However, we should add that the observation of a GFP modulation in the absence of a topographic modulation would most parsimoniously be interpreted as a modulation of the number of synchronously activated but statistically indistinguishable generators across experimental conditions [62]. Next, we present methods for identifying and quantifying topographic modulations.

### Global Dissimilarity: A Single, Strength-Independent Measure of Response Topography

Global dissimilarity (DISS) is an index of configuration differences between two electric fields, independent of their strength. Like GFP, DISS was first introduced by Lehmann and Skrandies [32]. This parameter equals the square root of the mean of the squared differences between the potentials measured at each electrode (versus the average reference), each of which is first scaled to unitary strength by dividing by the instantaneous GFP (see Appendix I). To provide a clearer sense of the calculation of DISS, consider again the data in Fig. 2. As already mentioned in the above section, Conditions 1 and 2 have the same topography but different strengths, whereas Conditions 1 and 3 have the same strength but different (inverted) topographies. Finally Conditions 1 and 4 differ in both their strength and topography. Figure 2a shows the original data, whereas the data in Fig. 2c have been GFP-normalized. Having thus re-scaled all four conditions to have the same GFP, the topographic similarities and differences between conditions becomes readily apparent. As shown in Fig. 2d, DISS can range from 0 to 2, where 0 indicates topographic homogeneity and 2 indicates topographic inversion.

Unlike GFP, however, the statistical analysis of DISS is not as straightforward, in part because DISS is a *single* measure of the distance between two vectors (each of which represents one electric field topography), rather than a separate measure for each condition about which a mean and variance can be calculated. Consequently, a non-parametric statistical test has to be conducted, wherein the dependent measure is the DISS between two maps at a given point in time,  $t$ . We and others have colloquially referred to this analysis as topographic ANOVA or TANOVA (e.g. [6, 7, 42, 44, 54, 63, 71], see also [29]), though we would immediately remind the reader that no analysis of variance is being conducted. Instead, TANOVA entails a non-parametric randomization test [34].<sup>3</sup> To do this for a within-subjects design, an empirical distribution of possible DISS values is determined by (1) re-assigning single-subject maps to different experimental conditions at a within-subject level (i.e. permutations of the data), (2) recalculating the group-average ERPs, and (3) recalculating the resulting DISS value for these ‘new’ group-average ERPs. The number of permutations that can be made with a group-average ERP based on  $n$  participants is  $2^n$ , though

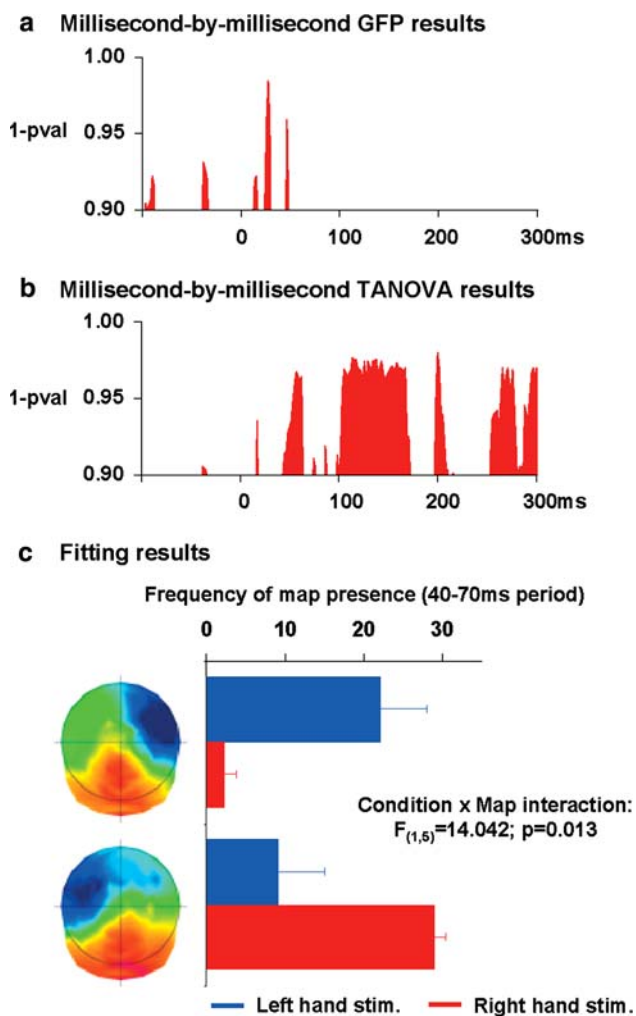
<sup>3</sup> Other methods for determining whether two electric fields differ have been proposed [25, 36, 64, 65]), criticized [21], and defended [58]. Some of these methods use average referenced and/or normalized values (like the  $T_{\text{sum}}^2$ ), but not all, which disqualifies them as true topographic analyses. Also, note the impossibility of using the Hotelling  $T^2$  multivariate statistic, as it requires more samples than electrodes and it precludes the use of the average reference.



**Fig. 2** Measurement of GFP and DISS. The basis for the reference-independent measurement of response strength and topography is shown. Color values throughout this figure denote polarity, with warmer colors indicating positive values and cooler colors negative values. (a) Hypothetical data from four different conditions from an array of 12 electrodes. The reader should note that Condition 2 is precisely twice the value of Condition 1 at each electrode and that Condition 3 is the inverse of the values of Condition 1 (i.e. the value at each electrode has been multiplied by -1). Finally, Condition 4 is a spatial re-arrangement of the values of Condition 2, making it differ in both strength and topography from Condition 1. (b) The squared value at each electrode and the summed value across electrodes as

well as the resulting GFP. The reader should note that Conditions 1 and 3 have the same GFP, even though their topographies are inverted, and that Conditions 2/4 have twice the GFP of Conditions 1/3. (c) The GFP-normalized values of the original data displayed in panel a. The reader should note that once strength differences are normalized, Conditions 1 and 2 have the same topography, whereas the topography of Condition 3 is the inversion of Conditions 1 and 2 (i.e. the extreme case) and the topography of condition 4 is slightly different from that of the other conditions. (d) The squared difference of the values in panel c at each electrode as well as the resulting DISS. The reader should note that DISS ranges from 0 to 2, with the former indicating identical topographies and the latter inverted topographies





**Fig. 3** Results of applying the methods of this tutorial to somatosensory ERPs. **(a)** The results of the analysis of GFP across time, which was based on the variance across subjects. This analysis failed to reveal any differences between responses to stimulation of each hand. **(b)** The results of the TANOVA analysis. ERPs to stimulation of each hand first topographically differed over the  $\sim 40$ – $70$  ms post-stimulus period. **(c)** The results of the fitting after having conducted a topographic pattern analysis based on AAHC (see text for details). Template maps are displayed for the 40–70 ms period. The bar graph shows that one template map better represents responses to left-hand stimulation and another template map better represents responses to right-hand stimulation

Manly [34] suggests that 1,000–5,000 permutations is sufficient. The DISS value from the actual group-average ERPs is then compared with the values from the empirical distribution to determine the likelihood that the empirical distribution has a value higher than the DISS from the actual group-average ERPs. This procedure can then be repeated for each time point. The results of this analysis for the somatosensory ERPs presented above is shown in Fig. 3b, which displays the occurrence of significant topographic differences between responses to left-hand and right-hand stimuli initially over the  $\sim 40$ – $70$  ms post-stimulus period

(as well as at subsequent time periods). For a between-subjects design, the analysis is generally identical, except that the permutations are performed by first putting all participants' data into one pool irrespective of experimental condition/group. Then new conditions/groups are randomly drawn and group-average ERPs are calculated for determining the empirical distribution.

At a neurophysiologic level, because electric field changes are indicative of changes in the underlying generator configuration (e.g. [12, 31, 67]) this test provides a statistical means of determining if and when the brain networks activated by the two conditions differ. In this way, the reader should note how response strength (GFP) and response topography (DISS) can be measured and analyzed orthogonally and in a completely reference-independent manner without the necessity of a priori selecting time periods or electrodes for analyses. Moreover, these two attributes can (and in our view should *always*) be analyzed as a function of time without the necessity of the experimenter *a priori* choosing time periods or components of interest. Still, some considerations when interpreting results of analyses with DISS are worth mentioning. Primary among these is that although a significant effect is unequivocal evidence that the topographies (and by extension configuration of intracranial generators) differ, this analysis does not in and of itself differentiate between several alternative underlying causes. For example, a significant difference may stem from one condition having one single and stable ERP topography during a given time period and the other condition another single and stable ERP topography over the same time period. That is, representing the electric field topography at a given time point by a letter, one condition might read “AAAAAA” and the other “BBBBBB”. Alternatively, each condition may be described by either single or multiple stable ERP topographies over the same time period (i.e. “AAABBB” versus “CCDDDD” or “AAAAAA” versus “BBCCDD”). Topographic differences might likewise stem from a latency shift between conditions (“ABCDEF” versus “BCDEFG”). Because all of these alternatives could result in highly similar (if not identical) patterns of statistical outcomes, additional analyses have been devised to determine the pattern of topographies both within and between conditions.

### Topographic Pattern Analysis & Single-Subject “Fitting”

An important issue, parallel to those already outlined above, in the analysis of EEG/ERPs is how to define the temporal intervals of a component, the temporal intervals for statistical analyses, and the temporal intervals to subject to source estimation. This becomes increasingly challenging when

high-density electrode montages are used. Also, the approach of averaging the measured potentials over a fixed and/or experimenter-defined time interval assumes that the electric field configuration is stable; an assumption that is seldom empirically verified. Our approach derives from the principle of functional microstates, which was first introduced by Dietrich Lehmann (e.g. [31]; reviewed in [38–40]). This principle is based on the empirical observation in both continuous EEG and ERPs that the electric field configuration at the scalp does not vary randomly as a function of time, but rather exhibits stability for tens to hundreds of milliseconds with brief intervening intervals of topographic instability. Similar findings have been observed in intracranial microelectrode recordings in non-human primates (e.g. [57]).

Here, we overview analysis procedures for identifying the periods of topographic stability within and between experimental conditions (Other approaches based on principal component analysis or independent component analysis are also frequently used; see e.g. [55, 56]). To return to the example in the preceding section, these analyses serve to identify the sequence of “letters”. The overarching procedure is the following. A clustering algorithm is applied to the collective group-averaged data across all experimental conditions/groups. This clustering does not account for the latencies of maps, but only for their topographies. This is done as a hypothesis generation step wherein the sequence of template maps that best accounts for the data is identified. The hypotheses generated at the group-average level are then statistically tested by means of a fitting procedure based on the spatial correlation between template maps obtained from the group-average ERPs and the single-subject ERP data [4, 51]. Several different dependent measures can be obtained from this fitting procedure; the advantages and disadvantages of which are presented in Table 1.

Two clustering algorithms will be presented whose implementation in the dedicated software, CarTool (<http://brainmapping.unige.ch/cartool.htm>), simultaneously treats both the spatial and temporal features of the data. One is based on k-means clustering [49], and the other on hierarchical clustering [66] that has been renamed “AAHC” for Atomize and Agglomerate Hierarchical Clustering. An intuitive way of understanding the main difference between these approaches is that the k-means approach operates independently for each number of clusters, whereas the hierarchical clustering approach operates in a bottom-up manner wherein the number of clusters is initially large and progressively diminishes. Both approaches yield generally comparable results, though some important differences are noteworthy. First, as will be made clearer below, because the k-means approach is based on the random selection of data points from within the dataset as seed clusters, its results can in principle vary from one run to the next, even though the same dataset is being analyzed.

This can be reasonably overcome by ensuring a high number of randomizations in the procedure (see below).<sup>4</sup> By contrast and because the AAHC approach is completely driven by the quantification of GEV (see below), its results will not vary from one run to another with the same dataset. Second, whereas the k-means approach is blind to the instantaneous GFP of the data being clustered, the AAHC approach takes such into consideration when calculating which clusters to retain. Because higher GFP is observed when signal quality is higher, the AAHC preferentially considers as robust clusters time periods with higher signal quality. The downside, however, is that the AAHC would be less effective if one were to perform the clustering analysis on normalized maps (i.e. where the GFP were constant across time). Additional material, including a tutorial film on the topographic pattern analysis based on k-means clustering, can be viewed and/or downloaded from the following URL (<http://brainmapping.unige.ch/docs/Murray-Supplementary.pps>).

### K-means Clustering

First, a concatenated dataset is defined using the group-averaged ERPs across all conditions/groups of the experiment. In the case of the example in this tutorial, there are two experimental conditions (left-hand and right-hand vibrotactile stimulation) that each contains 300 time points of data (i.e. all 600 time points of data). Second,  $n$  data points (where the term “data point” refers to the ERP from all scalp electrodes at a given instant in time) from this concatenated dataset (hereafter, template maps) are randomly selected from the concatenated dataset. The number of data points can range from 1 to the number of total data points. Third, the spatial correlation (Appendix I) between each of the  $n$  template maps and each time point of the concatenated dataset is calculated. This gives a spatial correlation value for each template map as a function of time, and for any given time point one of the  $n$  template maps yields highest spatial correlation value. As alluded to above, what is empirically observed in ERP data is that a given template map will yield the highest spatial correlation for a sustained period of time after which another and different template map will yield the highest spatial correlation, and so on. In addition, the experimenter can optionally constrain the minimal duration over which a given template map must yield the highest spatial correlation, thereby automatically rejecting short periods of comparatively unstable topography. From these spatial correlation values, the Global Explained Variance (GEV) of these template maps is calculated (Appendix I). GEV gives a metric of how well these  $n$  template maps

<sup>4</sup> A tangential side-effect of this need for randomizations during the implementation of the k-means clustering approach is that it is computationally longer than the AAHC method.

describe the whole dataset. Each of the  $n$  template maps is then redefined by averaging the maps from all time points when the  $i$ th template map yielded the highest spatial correlation versus all other template maps. Spatial correlation for each of these redefined template maps and the resultant GEV are recalculated as above. This procedure of averaging across time points to redefine each template map, recalculating the spatial correlation for each template map, and recalculating the GEV is repeated until the GEV becomes stable. In other words, a point is reached when a given set of  $n$  template maps cannot yield a higher GEV for the concatenated dataset. Because the selection of the  $n$  template maps is random, it is possible that neighboring time points were originally selected, which would result in a low GEV. To help ensure that this procedure obtains the highest GEV possible for a given number of  $n$  template maps, a new set of  $n$  template maps is randomly selected and the entire above procedure is repeated. It is important to note that the number of these random selections is user-dependent and will simply increase computational time as the number of random selections increases.<sup>5</sup> The set of  $n$  template maps that yields the highest GEV is retained. Finally, the above steps are now conducted for  $n + 1$  template maps and can iterate until  $n$  equals the number of data points comprising the concatenated dataset. The above steps provide information on how well  $n$ ,  $n + 1$ ,  $n + 2$  ... etc. template maps describe the concatenated dataset. An important issue for this analysis is the determination of the optimal number of template maps for a given dataset. We return to this below after first providing an overview of hierarchical clustering of EEG/ERPs.

### Hierarchical Clustering

The version of hierarchical clustering that has been devised by our group is a modified agglomerative hierarchical clustering termed “AAHC” for Atomize and Agglomerate Hierarchical Clustering. It has been specifically designed for the analysis of EEG/ERPs so as to counterbalance a side-effect of classical hierarchical clustering. Ordinarily, two clusters (i.e. groups of data points, or in the case of EEG/ERPs groups of maps) are merged together to proceed

from a total of  $n$  clusters to  $n - 1$  clusters. This leads to the inflation of each cluster’s size, because they progressively aggregate with each other like snow balls. While this is typically a desired outcome, in the case of EEG/ERPs it is potentially a major drawback when short-duration periods of stable topography exist (e.g. in the case of brainstem potentials). Following classical hierarchical agglomerative clustering, such short-duration periods would eventually be (blindly) disintegrated and the data would be designated to other clusters, even if these short-duration periods contribute a high GEV. In the modified version that is described here, clusters are given priority, in terms of their inclusion as one progresses from  $n$  to  $n - 1$  clusters, according to their GEV contribution. In this way, short-duration periods can be (conditionally) maintained.

Given this modification the AAHC procedure is then the following. As in the case of the k-means clustering, a concatenated dataset is defined as the group-averaged ERPs across all conditions/groups of the experiment. Initially, each data point (i.e. map) is designated as a unique cluster. Upon subsequent iterations, clusters denote groups of data points (maps), whose centroid (i.e. the mathematical average) defines the template map for that cluster. This is akin to the averaging across labeled data points in the k-means clustering described above. Then, the “worst” cluster is identified as the one whose disappearance will “cost” the least to the global quality of the clustering. Here, such is done by identifying the cluster with the lowest GEV (see Appendix I). This “worst” cluster is then atomized, meaning that its constituent maps are then “freed” and no longer belong to any cluster. One at a time, these “free” maps are independently re-assigned to the surviving clusters by calculating the spatial correlation between each free map and the centroid of each surviving cluster. The “free” map is then assigned to that cluster with which it has the highest spatial correlation (see Appendix I). The method then proceeds recursively by removing one cluster at a time, and stops when only 1 single final cluster is obtained (even though the latter is useless). Finally, for each level, i.e. for each set of  $n$  clusters, it is then possible to back-project the centroid/template maps onto the original data. This gives an output whose visualization is much like what is obtained via k-means clustering. As is the case for k-means clustering, an important next step will be to determine the optimal number of template maps (clusters).

### Identifying the Optimal Number of Template Maps

To this point, both clustering approaches will identify a set of template maps to describe the group-averaged ERPs. The issue now is how many clusters of template maps are optimal. Unfortunately, there is no definitive solution. This is because there is always a trade-off between the facts

<sup>5</sup> Clearly, the more variable the dataset is, the more random selections should be made to ensure the ‘best’  $n$  template maps are identified. However, this variability is often not known *a priori*. As the only ‘cost’ for more random selections is the experimenter’s time, in theory one could/should conduct  $(d!)/(n!(d - n)!)$  random selections, where  $d$  is the number of data points in the concatenated dataset and  $n$  is the number of template maps being randomly selected. In our experience, however, the results converge when  $\sim 100$  random selections are performed. The reason that computational time increases is that for each selection of  $n$  template maps from the original group-averaged data, all of the processing steps need be completed.

that the more clusters one identifies the higher the quality of the clustering (vis a vis GEV) but the lower the data reduction, and the converse. On one extreme, if the number of cluster is low then the explained variance will remain low, and the dataset itself will be highly compressed because it will now be represented by a small number of template maps. On the other extreme, if the number of clusters is high then the explained variance will also be high, but the dataset itself will not be compressed. The goal is to determine a middle-ground between such extremes. Here we present two methods: one based on Cross Validation (CV) and the other on the Krzanowski-Lai (KL) criterion.

Cross Validation criterion (CV) was first introduced by Pascual-Marqui et al. [49] as a modified version of the predictive residual variance (see Appendix I). Its absolute minimum gives the optimal number of segments. However and because CV is a ratio between GEV and the degrees of freedom for a given set of template maps, this criterion is highly sensitive to the number of electrodes in the montage. In our experience, the results actually become less reliable (i.e. there is less often an absolute minimum) when montages of more than 64 channels are used. That is, a unique CV minimum is more often obtained if the same 128-channel dataset is later down-sampled to a 32-channel dataset. Clearly, CV does not benefit from the added information of high-density electrode montages. Moreover, CV is also undefined in case there are more segments than electrodes.

Given these considerations with CV, another criterion has been developed that is based on the Krzanowski-Lai criterion [66]. It works by first computing a quality measure of the segmentation, termed Dispersion ( $W$ ).  $W$  trends toward 0 as the quality of the clustering results increases, in much the same manner that the GEV itself trends towards 1 as the quality of the clustering improves. The shape of the resulting  $W$  curve is then analyzed by looking for its L-corner; i.e. the point of highest deceleration where adding one more segment will not increase much the quality of the results. The KL measure has been slightly adapted to be a relative measure of curvature of the  $W$  curve (see Appendix I). As a consequence, its highest value should in principle indicate the optimal clustering. In practice, however, the KL will nearly all the time peak for three segments due to the very nature of the data we analyze. That is, there is systematically a steep deceleration of the  $W$  curve when progressing from 1 and 2 clusters (which are unsurprisingly “very bad” in terms of their overall quality in accounting for the concatenated dataset) to 3 clusters (which therefore always appears to then be “far better”). Though this peak at three segments can theoretically be of some interest, we advise considering the subsequent highest peak as the one indicating the optimal number of template maps, though additional peaks may also

ultimately be of interest if they lead to statistically significant results.

### Spatial Correlation-based Fitting & Its Dependent Measures

Irrespective of which clustering approach is used (and despite the abovementioned differences between these approaches), the experimenter is now confronted with the question of how to statistically assess the validity of the hypothesis that emerges from the clustering algorithm performed on the group-average dataset. The method we present here, like the above clustering algorithms, is based on calculating the spatial correlation between maps. In the case of the clustering algorithms this was performed on group-average ERPs and template maps. Here, the calculation is between single-subject ERPs and template maps that were identified by the clustering algorithm applied to the group-averaged ERPs (see also [4]). We colloquially refer to this calculation as “fitting”. Several different dependent measures from this fitting procedure can be obtained and statistically analyzed. We list a subset of these and their interpretability in Table 1. In addition, these dependent measures can in turn be correlated with behavioral measures (e.g. [1, 43, 63]), behavioral/mental states (e.g. [26, 28]), and/or parametric variations in stimulus conditions (e.g. [47, 51]). In Fig. 3c we present the outcome of the AAHC clustering and fitting procedure when applied to the somatosensory ERPs presented throughout this tutorial. In particular, we show the two template maps identified over the 40–70 ms period in the group-average ERPs and the incidence with which each of these maps yielded a higher spatial correlation with individual subjects’ data from each condition. The output shown in the bar graph is a mean value in time frames (milliseconds) that can then be statistically analyzed to reveal whether one map is more representative of one condition and another map is more representative of another condition (vis a vis a significant interaction between experimental condition and map). In the present example, one map is more representative of responses to stimulation of the left hand and another map is more representative of responses to stimulation of the right hand.

### Conclusions, Future Directions & Outlook

This tutorial review provides the details of both the rationale for as well as the implementation of a set of topographic analyses of multi-channel surface-recorded event-related potentials. A key advantage of these methods is their independence of both the reference and also *a priori* selection of certain electrodes or time points. These measures render

statistical information concerning modulations in response strength, latency, and topography both between and within experimental conditions. In these and other ways topographic analysis techniques allow the experimenter to glean additional information and neurophysiologic interpretability beyond what is available from canonical waveform analysis.

In addition to the progress in analysis tools and data interpretability, multi-channel EEG systems have become readily affordable for nearly all clinical and research laboratories. However, a potential risk of this ease-of-access to the equipment is that it may not be paralleled by researchers fully understanding or appropriately applying these analysis tools. As a result, EEG/ERPs as a research field risks becoming divided between those who apply only a minimal level of analysis and those who seek to more fully capitalize on the interpretational power of the technique. One goal of this tutorial was to show even to newcomers to the field that information-rich analyses can also be easy-to-use.

A final step that we have not addressed in this review is the application of source estimation techniques. This topic has been treated in several comprehensive reviews [2, 23, 40]. The relevance of the analyses presented in this tutorial to source estimations is the following. Analyses of the electric field at the scalp must be conducted that serve as the basis for

estimating the sources underlying these fields. That is, analysis of the surface-recorded data helps inform the researcher of specific time periods of interest for source estimations. Without such and if the experimenter were to arbitrarily select time periods, the resulting source estimation would have little (or more likely no) neurophysiologic meaning (c.f. [53] for discussion).

We would end by mentioning some additional approaches under development that are promising for providing a closer translational link across brain imaging methods and across studies conducted in different species. Among these are the application of clustering algorithms to single-subject and single-trial data [5, 17] and the direct analysis of single-subject and single-trial source estimations [16, 19], including within the time-frequency domain [18, 35].

**Acknowledgements** We thank Laura De Santis for assistance with data collection. Cartool software is freely available at (<http://www.brainmapping.unige.ch/Cartool.htm>) and is supported by the Center for Biomedical Imaging (<http://www.cibm.ch>) of Geneva and Lausanne. MMM receives financial support from the Swiss National Science Foundation (grant #3100AO-118419) and the Leenaards Foundation (2005 Prize for the Promotion of Scientific Research). CMM receives financial support from the Swiss National Science Foundation (grant #320000-111783).

## Appendix

### Appendix I: Formulae

$n$  is the number of electrodes in the montage, including the reference  
 $U_i$  is the measured potential of the  $i$ th electrode, for a given condition  $U$ , at a given time point  $t$  (also including the reference)

$V_i$  is the measured potential of the  $i$ th electrode, either from another condition  $V$ , or from the same condition  $U$  but at a different time point  $t'$

Average reference

$$\bar{u} = \frac{1}{n} \cdot \sum_{i=1}^n U_i$$

$$u_i = U_i - \bar{u}$$

$\bar{u}$  is the mean value of all  $U_i$ 's (for a given condition, at a given time point  $t$ )

$u_i$  is the average-referenced potential of the  $i$ th electrode (for a given condition, at a given time point  $t$ )

Global field power (GFP)

$$GFP_u = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n u_i^2}$$

$$GFP_u = \sigma_u$$

The GFP for a given condition, at a given time point

GFP is equivalent to the standard deviation of the electrode values (at a given time point  $t$ )

Global dissimilarity (DISS)

$$DISS_{u,v} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n \left( \frac{u_i}{GFP_u} - \frac{v_i}{GFP_v} \right)^2}$$

$$DISS_{u,v} = \sqrt{2 \cdot (1 - C_{u,v})}$$

Between two conditions at the same time point, or between two different time points of the same condition

(See below for the definition of  $C$ )

Spatial correlation ( $C$ )

$$C_{u,v} = \frac{\sum_{i=1}^n u_i \cdot v_i}{\|u\| \cdot \|v\|}$$

$$\|u\| = \sqrt{\sum_{i=1}^n u_i^2}, \|v\| = \sqrt{\sum_{i=1}^n v_i^2}$$

$$C_{u,v} = 1 - \frac{DISS_{u,v}^2}{2}$$

Spatial correlation between two conditions at the same time point, or between two different time points of the same condition

( $C$  is equivalent to the Pearson cross-correlation coefficient)

(See above the definition of DISS)

## Appendix I: continued

$n$  is the number of electrodes in the montage, including the reference  
 $U_i$  is the measured potential of the  $i$ th electrode, for a given condition  $U$ , at a given time point  $t$  (also including the reference)  
 $V_i$  is the measured potential of the  $i$ th electrode, either from another condition  $V$ , or from the same condition  $U$  but at a different time point  $t'$

## Segmentation results

$L_{u,t}$  = SegmentIndex

A labeling  $L$ , which holds the index of the segment attributed, for condition  $U$ , at time point  $t$

$T_k$

$T_k$  is the  $k$ th template map (a vector of  $n$  dimensions)

$\bar{T}_k = 0, \|T_k\| = 1$

$T_k$  has a mean of 0, and is normalized

## Global explained variance (GEV)

$$GEV = \frac{\sum_{t=1}^{\max} (GFP_u(t) \cdot C_{u,T_t})^2}{\sum_{t=1}^{\max} GFP_u^2(t)}$$

(This can be computed only after a segmentation)  $t$  is a given time point within the data

$GFP_u(t)$  is the GFP of the data for condition  $U$  at time point  $t$ .  $T_t$  is the template map assigned by the segmentation for condition  $U$  at time point  $t$

$T_t = T_{L_{u,t}}$

$GEV = \sum_{k=1}^q GEV_k$

$C_{u,T_t}$  is the spatial correlation between data of condition  $U$  at time point  $t$ , and the template map  $T_t$ , assigned to that time point by the segmentation

$$GEV_k = \frac{\sum_{t=1}^{\max} (GFP_u(t) \cdot C_{u,T_t})^2 \cdot \gamma_{u,k,t}}{\sum_{t=1}^{\max} GFP_u^2(t)}$$

The GEV can also be broken down into its partial contributions  $GEV_k$  for each of its segment  $k$

$q$  is the number of segments/template maps

$$\gamma_{u,k,t} = \begin{cases} 1 & \text{if } k = L_{u,t} \\ 0 & \text{if } k \neq L_{u,t} \end{cases}$$

$\gamma_{u,k,t}$  is set to 1 only for time points where data have been labelled as belonging to the  $k^{\text{th}}$  segment, and 0 otherwise

## Cross validation criterion (CV)

$$CV = \hat{\sigma}_\mu^2 \cdot \left( \frac{n-1}{n-1-q} \right)^2$$

$q$  is the number of segments/template maps

$n$  is the number of electrodes

$$\hat{\sigma}_\mu^2 = \frac{\sum_{t=1}^{\max} (\|u(t)\|^2 - (T_t \cdot u(t))^2)}{t_{\max} \cdot (n-1)}$$

$(T_t \cdot u(t))$  denotes the scalar product between the template maps  $T_t$  and the data  $u(t)$  at time point  $t$

## Krzanowski-Lai criterion

$$W_q = \sum_{r=1}^q \frac{1}{2 \cdot n_r} \cdot D_r$$

$W$  is the measure of dispersion for  $q$  clusters

$n_r$  is the number of maps for cluster  $r$

$$D_r = \sum_{u,v \in \text{clusterr}} \|u - v\|^2$$

$D_r$  is the sum of pair-wise distance between all maps of a given cluster  $r$ .  $KL_q$  is the Krzanowski-Lai criterion for  $q$  clusters (formula adapted to compute the normalized curvature of  $W$ )

$$KL_q = \frac{d_{q-1} - d_q}{M_{q-1}}$$

$$d_q = M_q - M_{q+1}$$

$$M_q = W_q \cdot q^{2/n}$$

Moreover,  $KL_q$  is set to 0 if  $d_{q-1} < 0$  or  $d_{q-1} < d_q$  (only concave shapes of the  $W$  curve are considered)

## References

- Arzy S, Mohr C, Michel CM, Blanke O. Duration and not strength of activation in temporo-parietal cortex positively correlates with schizotypy. *Neuroimage* 2007;35:326–33.
- Baillet S, Moshier JC, Leahy RM. Electromagnetic brain mapping. *IEEE Signal Process Mag.* 2001;18(16):14–30.
- Brandeis D, Lehmann D. Event-related potentials of the brain and cognitive processes: approaches and applications. *Neuropsychologia* 1986;24:151–68.
- Brandeis D, Lehmann D, Michel CM, Mingrone W. Mapping event-related brain potential microstates to sentence endings. *Brain Topogr.* 1995;8:145–59.
- De Lucia M, Michel CM, Clarke S, Murray MM. Single-subject EEG analysis based on topographic information. *Int J Bioelectromagnet.* 2007;9:168–71.
- De Santis L, Spierer L, Clarke S, Murray MM. Getting in touch: segregated somatosensory ‘what’ and ‘where’ pathways in humans revealed by electrical neuroimaging. *Neuroimage* 2007;37:890–903.
- De Santis L, Clarke S, Murray MM. Automatic and intrinsic auditory ‘what’ and ‘where’ processing in humans revealed by electrical neuroimaging. *Cereb Cortex.* 2007;17:9–17.
- Desmedt JE, Tomberg C, Noel P, Ozaki I. Beware of the average reference in brain mapping. *Electroencephalogr Clin Neurophysiol Suppl.* 1990;41:22–7.
- Dien J. Issues in the application of the average reference: review, critiques, and recommendations. *Behav Res Methods Instrum Comput.* 1998;30:34–43.
- Dien J, Santuzzi AM. Application of repeated measures ANOVA to high-density ERP datasets: a review and tutorial. In: Handy TC, editors. *Event-related potentials: a methods handbook.* Cambridge, MA: MIT Press; 2005. pp. 57–82.
- Duffy FH. Topographic display of evoked potentials: clinical applications of brain electrical activity mapping (BEAM). *Ann N Y Acad Sci.* 1982;388:183–96.
- Fender DH. Source localisation of brain electrical activity. In: Gevins AS, Remond A, editors. *Handbook of electroencephalography and clinical neurophysiology, vol. 1: methods of analysis of brain electrical and magnetic signals.* Amsterdam: Elsevier; 1987. pp. 355–99.
- Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, Penny WD. *Statistical parametric mapping: the analysis of functional brain images.* London: Academic Press; 2001.
- Geselowitz DB. The zero of potential. *IEEE Eng Med Biol Mag.* 1998;17:128–32.

15. Gevins AS, Morgan NH, Bressler SL, Cutillo BA, White RM, Illes J, Greer DS, Doyle JC, Zeitlin GM. Human neuroelectric patterns predict performance accuracy. *Science* 1987;235:580–5.
16. Gonzalez Andino SL, Murray MM, Foxe JJ, Menendez RGP. How single-trial electrical neuroimaging contributes to multi-sensory research. *Exp Brain Res*. 2005;166:298–304.
17. Gonzalez Andino SL, Grave de Peralta R, Khateb A, Pegna AJ, Thut G, Landis T. A glimpse into your vision. *Hum Brain Mapp*. 2007;28:614–24.
18. Gonzalez SL, Grave de Peralta R, Thut G, Millán Jdel R, Morier P, Landis T. Very high frequency oscillations (VHFO) as a predictor of movement intentions. *Neuroimage* 2006;32:170–9.
19. Grave de Peralta Menendez R, Murray MM, Michel CM, Martuzzi R, Gonzalez Andino SL. Electrical neuroimaging based on biophysical constraints. *Neuroimage* 2004;21:527–39.
20. Guthrie D, Buchwald JS. Significance testing of difference potentials. *Psychophysiology* 1991;28:240–4.
21. Haig AR, Gordon E, Cook S. To scale or not to scale: McCarthy and Wood revisited. *Electroencephalogr Clin Neurophysiol*. 1997;103:323–5.
22. Handy TC. *Event-related potentials: a methods handbook*. Cambridge, MA: MIT Press; 2005.
23. He B, Lian J. High-resolution spatio-temporal functional neuroimaging of brain activity. *Crit Rev Biomed Eng*. 2002;30:283–306.
24. Hanson JC, Hillyard SA. Endogenous brain potentials associated with selective auditory attention. *Electroencephalogr Clin Neurophysiol*. 1980;49:277–90.
25. Karniski W, Blair RC, Snider AD. An exact statistical method for comparing topographic maps, with any number of subjects and electrodes. *Brain Topogr*. 1994;6:203–10.
26. Katayama H, Gianotti LR, Isotani T, Faber PL, Sasada K, Kinoshita T, Lehmann D. Classes of multichannel EEG microstates in light and deep hypnotic conditions. *Brain Topogr*. 2007;20:7–14.
27. Koenig T, Lehmann D. Microstates in language-related brain potential maps show noun-verb differences. *Brain Lang*. 1996; 53:169–82.
28. Koenig T, Prichep L, Lehmann D, Sosa PV, Braecker E, Kleinlogel H, Isenhardt R, John ER. Millisecond by millisecond, year by year: normative EEG microstates and developmental stages. *Neuroimage* 2002;16:41–8.
29. Kondakor I, Pascual-Marqui R, Michel CM, Lehmann D. Event-related potential map differences depend on prestimulus microstates. *J Med Eng Technol*. 1995;19:66–9.
30. Lantz G, Grave de Peralta R, Spinelli L, Seeck M, Michel CM. Epileptic source localization with high density EEG: how many electrodes are needed? *Clin Neurophysiol*. 2003;114:63–9.
31. Lehmann D. Principles of spatial analysis. In: Gevins AS, Remond A, editors. *Handbook of electroencephalography and clinical neurophysiology*, vol. 1: methods of analysis of brain electrical and magnetic signals. Amsterdam: Elsevier; 1987. pp. 309–54.
32. Lehmann D, Skrandies W. Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroenceph Clin Neurophysiol*. 1980;48:609–21.
33. Luck SJ. *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press; 2005.
34. Manly BF. *Randomization and Monte Carlo methods in biology*. London, UK: Chapman & Hall; 1991.
35. Martuzzi R, Murray MM, Meuli RA, Thiran JP, Maeder PP, Michel CM, Menendez RGP, Andino SLG. A new analytical method to investigate frequency- and region-dependant relationships between estimated LFPs and BOLD responses in humans; 2008. In review.
36. McCarthy G, Wood CC. Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. *Electroenceph Clin Neurophysiol*. 1985;62:203–8.
37. Michel CM, Henggeler B, Lehmann D. 42-channel potential map series to visual contrast and stereo stimuli: perceptual and cognitive event-related segments. *Int J Psychophysiol*. 1992;12: 133–45.
38. Michel CM, Seeck M, Landis T. Spatio-temporal dynamics of human cognition. *News Physiol Sci*. 1999;14:206–14.
39. Michel CM, Thut G, Morand S, Khateb A, Pegna AJ, Grave de Peralta R, Gonzales S, Seeck M, Landis T. Electric source imaging of human cognitive brain functions. *Brain Res Rev*. 2001;36:108–18.
40. Michel CM, Murray MM, Lantz G, Gonzalez S, Spinelli L, Grave de Peralta R. EEG source imaging. *Clin Neurophysiol*. 2004;115:2195–222.
41. Murray MM, Foxe JJ, Higgins BA, Javitt DC, Schroeder CE. Visuo-spatial neural response interactions in early cortical processing during a simple reaction time task: a high-density electrical mapping study. *Neuropsychologia* 2001;39:828–44.
42. Murray MM, Michel CM, Grave de Peralta R, Ortigue S, Brunet D, Andino SG, Schneider A. Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage* 2004;21:125–35.
43. Murray MM, Imber ML, Javitt DC, Foxe JJ. Boundary completion is automatic and dissociable from shape discrimination. *J Neurosci*. 2006;26:12043–54.
44. Murray MM, Camen C, Spierer L, Clarke S. Plastic representations of environmental sounds revealed by electrical neuroimaging. *Neuroimage* 2008;39:847–56.
45. Nunez PL, Silberstein RB, Cadusch PJ, Wijesinghe RS, Westdorp AF, Srinivasan R. A theoretical and experimental study of high resolution EEG based on surface Laplacians and cortical imaging. *Electroencephalogr Clin Neurophysiol*. 1994;90:40–57.
46. Oostenveld R, Praamstra P. The five percent electrode system for high-resolution EEG and ERP measurements. *Clin Neurophysiol*. 2001;112:713–9.
47. Overney LS, Michel CM, Harris IM, Pegna AJ. Cerebral processes in mental transformations of body parts: recognition prior to rotation. *Brain Res Cogn Brain Res*. 2005;25:722–34.
48. Pascual-Marqui RD, Lehmann D. Comparison of topographic maps and the reference electrode: comments on two papers by Desmedt and collaborators. *Electroencephalogr Clin Neurophysiol*. 1993;88:530–1, 534–6.
49. Pascual-Marqui RD, Michel CM, Lehmann D. Segmentation of brain electrical activity into microstates, model estimation and validation. *IEEE Trans Biomed Eng*. 1995;42:658–65.
50. Patarraia E, Baumgartner C, Lindinger G, Deecke L. Magnetoencephalography in presurgical epilepsy evaluation. *Neurosurg Rev*. 2002;25:141–59.
51. Pegna AJ, Khateb A, Spinelli L, Seeck M, Landis T, Michel CM. Unravelling the cerebral dynamics of mental imagery. *Hum Brain Mapp*. 1997;5:410–21.
52. Perrin F, Pernier J, Bertrand O, Giard MH, Echallier JF. Mapping of scalp potentials by surface spline interpolation. *Electroencephalogr Clin Neurophysiol*. 1987;66:75–81.
53. Picton TW, Bentin S, Berg P, Donchin E, Hillyard SA, Johnson R Jr, Miller GA, Ritter W, Ruchkin DS, Rugg MD, Taylor MJ. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 2000;37:127–52.
54. Pourtois G, Thut G, Grave de Peralta R, Michel CM, Vuilleumier P. Two electrophysiological stages of spatial orienting towards fearful faces: early temporo-parietal activation preceding gain control in extrastriate visual cortex. *Neuroimage* 2005;26:149–63.
55. Pourtois G, Delplanque S, Michel CM, Vuilleumier P. Beyond conventional event-related brain potentials (ERPs): exploring the time course of visual emotion processing using topographic and principal component analyses. *Brain Topogr*. 2008. doi:10.1007/s10548-008-0053-6.

56. Rigoulot S, Delplanque S, Desprez P, Defoort-Dhellemmes S, Honoré J, Sequeira H. Peripherally presented emotional scenes: a spatiotemporal analysis of early ERP responses. *Brain Topogr*. 2008. doi:10.1007/s10548-008-0050-9.
57. Rolls ET, Tovee MJ. Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proc R Soc Lond B*. 1994;257:9–15.
58. Ruchkin DS, Johnson R Jr, Friedman D. Scaling is necessary when making comparisons between shapes of event-related potential topographies: a reply to Haig et al. *Psychophysiology* 1999;36:832–4.
59. Saron CD, Schroeder CE, Foxe JJ, Vaughan HG Jr. Visual activation of frontal cortex: segregation from occipital activity. *Brain Res Cogn Brain Res*. 2001;12:75–88.
60. Skrandies W. Global field power and topographic similarity. *Brain Topogr*. 1990;3:137–41.
61. Skrandies W. EEG/EP: new techniques. *Brain Topogr*. 1993;5:347–50.
62. Skrandies W. The effect of stimulation frequency and retinal stimulus location on visual evoked potential topography. *Brain Topogr*. 2007;20:15–20.
63. Spierer L, Tardif E, Sperdin H, Murray MM, Clarke S. Learning-induced plasticity in auditory spatial representations revealed by electrical neuroimaging. *J Neurosci*. 2007;27:5474–83.
64. Srebro R. A bootstrap method to compare the shapes of two scalp fields. *Electroenceph Clin Neurophysiol*. 1996;100:25–32.
65. Srinivasan R, Nunez PL, Tucker DM, Silberstein RB, Cadusch PJ. Spatial sampling and filtering of EEG with spline laplacians to estimate cortical potentials. *Brain Topogr*. 1996;8:355–66.
66. Tibshirani R, Walther G, Botstein D, Brown P. Cluster validation by prediction strength. *J Comput Graphical Stat*. 2005;14:511–28.
67. Vaughan HG Jr. The neural origins of human event-related potentials. *Ann N Y Acad Sci*. 1982;388:125–38.
68. Vaughan HG Jr, Arezzo JC. The neural basis of event-related potentials. In: Picton TW, editors. *Human event-related potentials. EEG handbook (revised series) vol. 3*. New York: Elsevier; 1988. pp. 45–96.
69. Wikswa JP, Gevins A, Williamson SJ. The future of EEG and MEG. A review article. *Electroenceph Clin Neurophysiol*. 1993;87:1–9.
70. Williamson SJ, Lu ZL, Karon D, Kaufman L. Advantages and limitations of magnetic source imaging. *Brain Topogr*. 1991;4: 169–80.
71. Wylie GR, Murray MM, Javitt DC, Foxe JJ. Distinct neurophysiological mechanisms mediate mixing costs and switch costs. *J Cogn Neurosci*. 2008. in press.