



Grand average ERP-image plotting and statistics: A method for comparing variability in event-related single-trial EEG activities across subjects and conditions



Arnaud Delorme^{a,b,c,d,*}, Makoto Miyakoshi^a, Tzyy-Ping Jung^a, Scott Makeig^a

^a Swartz Center of Computational Neuroscience, Institute of Neural Computation, University of California San Diego, La Jolla, San Diego, CA 92093-0559, USA

^b Institute of Noetic Sciences, Petaluma, CA, USA

^c CerCo, CNRS UMR5549, Toulouse, France

^d Centre de Recherche Cerveau et Cognition, Université de Toulouse III, UPS, Toulouse, France

HIGHLIGHTS

- The ERP-image plotting method visualizes a collection of event-related EEG data epochs sorted on some trial variable of interest.
- We demonstrate and assess extending the ERP-image plotting method to sets of data epochs from multiple subjects.
- We demonstrate the use of 'grand' ERP-image plotting and associated inferential statistics on data collected in a visual attention task.

ARTICLE INFO

Article history:

Received 11 August 2014

Received in revised form 2 October 2014

Accepted 3 October 2014

Available online 22 October 2014

Keywords:

EEG

Event related potential

ERP

Single-trial

ERP-image

ABSTRACT

Background: With the advent of modern computing methods, modeling trial-to-trial variability in bio-physical recordings including electroencephalography (EEG) has become of increasingly interest. Yet no widely used method exists for comparing variability in ordered collections of single-trial data epochs across conditions and subjects.

New method: We have developed a method based on an ERP-image visualization tool in which potential, spectral power, or some other measure at each time point in a set of event-related single-trial data epochs are represented as color coded horizontal lines that are then stacked to form a 2-D colored image. Moving-window smoothing across trial epochs can make otherwise hidden event-related features in the data more perceptible. Stacking trials in different orders, for example ordered by subject reaction time, by context-related information such as inter-stimulus interval, or some other characteristic of the data (e.g., latency-window mean power or phase of some EEG source) can reveal aspects of the multifold complexities of trial-to-trial EEG data variability.

Results: This study demonstrates new methods for computing and visualizing 'grand' ERP-image plots across subjects and for performing robust statistical testing on the resulting images. These methods have been implemented and made freely available in the EEGLAB signal-processing environment that we maintain and distribute.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Since the early 1960s, EEG research has been dominated by the study of averaged event-related potentials (ERPs) whose waveforms capture perturbations in EEG signals that are both time- and phase locked to some set of experimental events. Over the past

40 years, the response-averaging method has given many insights into brain processing (Luck, 2005). However, response averaging ignores the rich trial-to-trial variability that is a basic and important characteristic of EEG data. Over the past decade, the focus of modeling interest in the neuroscientific community has gradually moved beyond trial averaging to analysis of trial variability. Early evidence for single-trial correlation between P300 ERP amplitude and behavior (motor reaction time) was reported for example by Ritter et al. (1972) who manually identified single-trial latencies of ERP peaks, a method later criticized as a 'laborious technique impractical for experiments involving hundreds of trials' (Kutas et al., 1977). Today,

* Corresponding author at: SCCN, INC, UCSD, 9500 Gilman Drive, La Jolla, CA 92093, USA. Tel.: +1 858 822 7542; fax: +1 858 822 7556.

E-mail addresses: arno@cerco.ups-tlse.fr, arno@ucsd.edu (A. Delorme).

an intuitive visualization approach to visualizing systematic relationships between event-related EEG dynamics across a collection of event-related trials is readily available in the EEGLAB signal-processing environment we have developed and distributed over the past dozen years (Delorme and Makeig, 2004; Delorme et al., 2011).

EEGLAB allows exploring trial-to-trial differences in event-related EEG epochs using a visualization method we dubbed ‘ERP-image plotting’ (Jung et al., 2001; Makeig et al., 1999). A precursor to this visualization scheme is spike raster plots (MacPherson and Aldridge, 1979), a standard plotting scheme in the field of single-unit recording studies. ERP-image plots differ from such plots in that they use color coding to represent some continuous ERP activity variable (potential itself, or some measure derived from it) instead of the times of occurrence of neural spike events as in spike raster plotting.

Instead of performing trial averaging, brain activity in each trial for some source or scalp channel is coded as a row of values in an ERP-image matrix. Row vectors for the single trials are stacked and may then be smoothed vertically, forming an ERP-image matrix of dimensions (number of epoch latencies \times number of trials). The ERP-image matrix can be color-coded so that the color of each pixel represents some measure of activity at each epoch time point, forming an ERP-image plot (Fig. 1A–C). Another innovation was to encourage ERP-image plotting in which the single trials are sorted by some trial-wise measure of interest, for example subject reaction time (Jung et al., 2001; Makeig et al., 1999). Colored ERP-image plotting is most useful for revealing patterns in the data visible only by sorting the trials based on a measure that systematically affects, in some way, the size, latency, or nature of the event-locked portion of the data trials. For example, a result of interest often appears when trials are sorted on the latencies of subject manual responses, i.e., subject reaction times (Delorme et al., 2007; Jung et al., 2001). Other sorting orders of interest include the mean potential of the imaged EEG source in a given latency window, or amplitude or phase in some spectral band in a given latency/frequency window. See Onton and Makeig (2009) for a detailed example.

2. Method

One method for aggregating ERP-image matrices across subjects is to simply concatenate similar event-related trials (for some set of equivalent sources) across subjects, then make an ERP-image matrix and plot for the whole trial ensemble. In this case, to avoid the large variability in the overall size of the EEG across subjects, possibly allowing a few subjects to dominate the smoothed ‘grand’ ERP-image, it may be preferable to normalize each trial by dividing by the subject’s pre-stimulus standard deviation, i.e., converting the single trials from each subject to z-score values (Makeig et al., 2004b). This study presents an alternative method that guarantees equal contributions from all subjects to a grand ERP-image matrix. The current method offers the advantage of allowing computation of standard statistics on subject or condition differences as detailed below.

Assuming data epochs have been extracted, our method first generates an ERP-image matrix, EM_{S_i, C_j} for each subject S_i (out of N total subjects) and each condition C_j . The grand ERP-image \bar{EM}_{C_j} for condition C_j is then $\bar{EM}_{C_j} = \sum_{i=1}^N EM_{S_i, C_j}$. This method assumes that all the ERP-image matrices have the same number of rows (m) and columns since to perform averaging m must be constant across subjects and conditions. The vertical smoothing window across trials is also best kept constant. To make all the ERP-image matrices have the same number of rows after applying vertical smoothing, we may decimate trials for some subjects and/or conditions. Trial decimation – involving smooth trial interpolation or omitting some

smoothed trial averages – is usually used to reduce unnecessary computation (why ask the computer to display 10,000 ERP-image lines on a display that can only show a few hundred?). Decimation can also be used to make constant the number of smoothed rows in each subject ERP-image matrix. Fig. 1D illustrates the ERP-image trial decimation process with a simple example.

Inferential statistics on ERP-image matrices consists in independently comparing each row and column of a set of ERP-image matrices for different subjects. In so doing, $k \times n$ statistical estimates are calculated. Since neighboring points in ERP-image matrices are not independent, using an appropriate method to correct for multiple comparisons is necessary. Neighboring rows and columns of an ERP-image matrix may be correlated with each other—because of the vertical smoothing across trials (ERP-image rows) and the low pass nature of continuous EEG signals that correlates adjacent time points (ERP-image matrix columns). Although standard correction methods, such as False Discovery Rate (Benjamini and Yekutieli, 2001) may be used to correct for such correlations (Groppe et al., 2011), in the following section we propose use of a cluster correction method (Maris and Oostenveld, 2007) that correctly estimates family wise error rate (Pernet et al., 2015).

3. Application

As an illustration, we present data from an attention task discussed in earlier reports (Delorme et al., 2007; Makeig et al., 2004b; Makeig et al., 1999). Fifteen right-handed subjects (age 19–53 years old; mean 30 years) with normal or corrected to normal vision participated in the experiment. Filled white disks appeared briefly inside one of five empty squares. Subjects were instructed to attend to one of the five squares. In each block, 100 stimuli (filled white disks) were displayed for 117 ms within one of the five empty squares in a pseudo-random sequence with inter-stimulus intervals (ISIs) of 250–1000 ms (in four equiprobable 250 ms steps). Thirty blocks of trials were collected from each subject, yielding 120 targets – when the filled disk appeared at the location the subject was attending to – and 480 non-target trials at each location. In the present report, we consider only the correctly responded target trials in which subjects responded between 150 ms and 1000 ms following target onset (these amount to >95% of all target trials). More details about the experimental protocol can be found in previous publications (Delorme et al., 2007; Makeig et al., 2004b; Makeig et al., 1999).

Data were preprocessed as described in Delorme et al. (2007) including removal of six independent component clusters (not shown) accounting for eye movement and muscle tension artifacts. Fig. 2 shows grand ERP-image plots for artifact-cleaned scalp channels C3 and C4, as well as the difference between the ERP-image plots for these two channels. For each of the two electrodes, we computed the null distribution of grand ERP-image matrix values under the null hypothesis of no difference given a common mean-zero baseline. To assess the null ERP-image distribution, we computed a grand ERP-image after randomly designating the polarity of each subject ERP-image – meaning the probability of switching ERP-image polarity for each subject was $\frac{1}{2}$. This procedure was repeated 10,000 times (sampling from 32,768 possibilities, as there were 15 subjects). This gave a surrogate distribution of values for each ERP-image matrix value that was used to assess significance. Correction for multiple comparisons was assessed using a voxel clustering method (Maris and Oostenveld, 2007). To compare event-related activity for (left central scalp) C3 and (right central scalp) C4 electrode channels, we used paired statistics. For each subject, the ERP-image matrix for C3 was subtracted from the ERP-image matrix for C4. Under the null

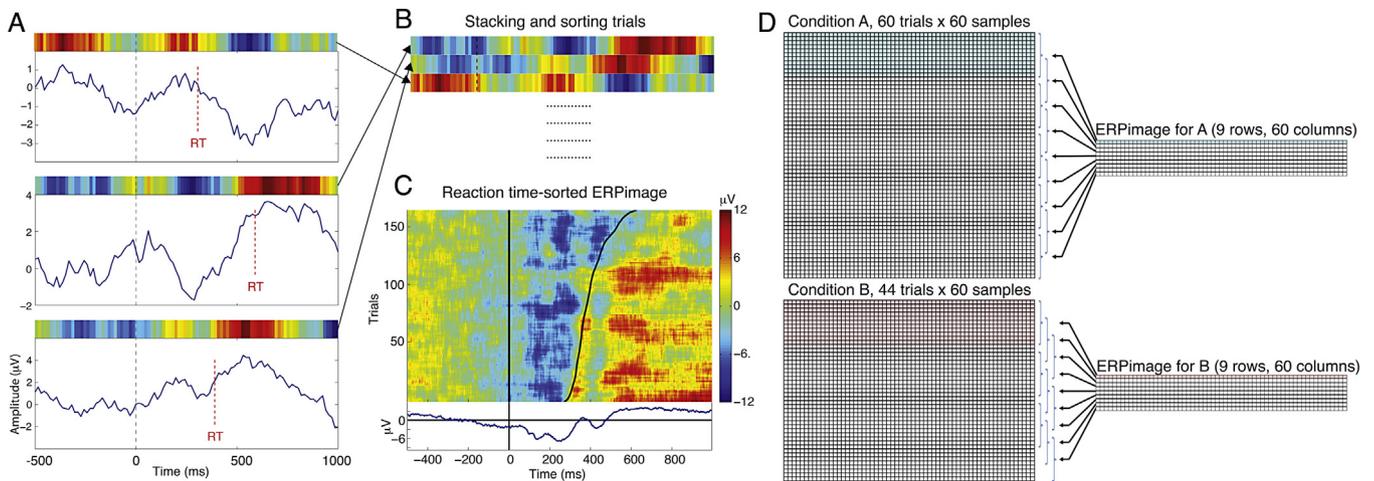


Fig. 1. Example of ERP-image and grand ERP-image matrix and plot construction. (A) In ERP-image plots each single trial is encoded as a colored line, warm colors representing positive activity and cool colors, negative activity. (B) Conceptually, the colored lines for the trials are stacked after (re)ordering them with respect to some variable of interest (reaction time for example—shown in (C) by the curving black trace). The ordered trials may then be smoothed (vertically) using a multi-trial smoothing window. The trace below the ERP-image panel in (C) shows the trial-averaged ERP. Note that the distinction between stimulus- and response-locked features of the data, well represented in the ERP-image plots, are lost in the trial average ERP. (D) Trial decimation to create ERP-image matrices with constant row and column size. Illustration for 9 output rows, with a trial smoothing width of 11 trials. Trial decimation: The number of trials in condition A is 60, so to make the numbers of output lines in the ERP-image matrix the same for each subject, eight smoothed estimates are computed centered on trial indices $(6i)$, $i = 1, 2, \dots, 9$. For condition B with 44 trials, eight smoothed estimates are computed centered on trial indices $(6+4i)$, $i = 0, 1, \dots, 8$. In both cases, the ERP-image matrices will then contain 9 rows. Boundary trial indices are determined by the smoothing factor. The interval between the boundary trials is then divided the number of remaining output rows. In case the resulting trial indices are not integers, the closest trial indices may be used.

hypothesis of no difference between the two conditions ERP-image matrices, we performed random sign inversion for some of the ERP-image matrix differences—this is equivalent to switching the condition for each subject with probability 1/2. The resulting surrogate distribution was used to assess significance after correction for multiple comparisons.

Fig. 2 shows that event-related EEG activities reaching the scalp electrode channels C3 and C4 immediately follow the subject's manual response. Though this difference is not visible from inspection of the two grand ERP-images for channels C3 and C4, computing the difference between the ERP-image matrices for the two channels cancels (by phase cancellation) some of the trial-to-trial and wideband variability ('noise') not accounted for by subject reaction time differences alone. The result reveals that the motor event-locked activity begins earlier at right-central scalp site C4 than at left-central C3. Note that it is difficult to interpret such single scalp channel data in terms of brain processes since these signals sum many cortical sources (Makeig et al., 2004a). The power

of ERP-image plotting and statistics, and the interpretability of the results, are typically stronger when source-resolved activity is imaged (Onton and Makeig, 2009).

4. Discussion

The amount of useful information gleaned from EEG studies designed for extracting event-related potentials has been limited for years because ERP averaging methods do not allow visualizing and studying the rich variability in cortical dynamics occurring across EEG epochs (trials) time-locked to the set of experimental events of interest. ERP-image plotting and statistics allow visualization of differences between single trials that may be revealed when they are ordered by a trial-wise variable of interest. We showed that it is possible to generalize the concept of the 'grand average' ERP across subjects to the new concept of the 'grand average' ERP-image matrix and visualization. This result is important for field of EEG research in particular and for human neuroscience

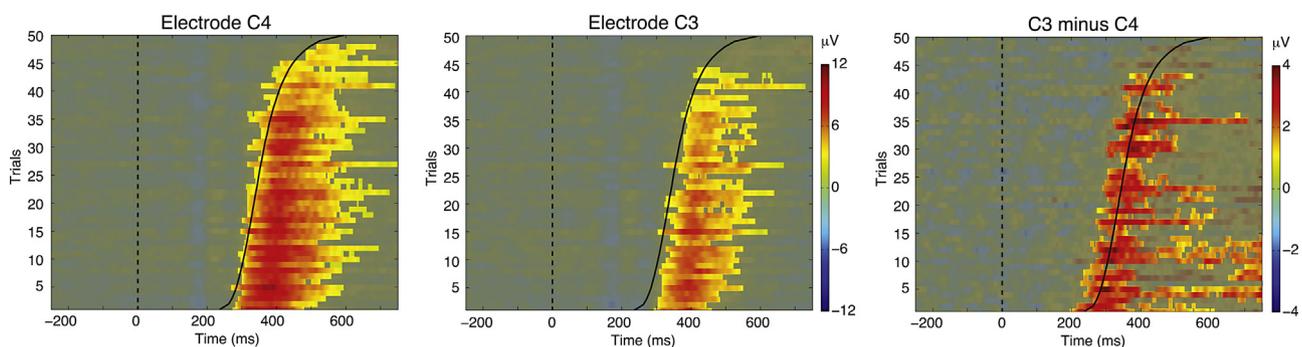


Fig. 2. Example of grand ERP-image plots for scalp electrode channels C3 and C4 (left), and the difference between the two electrode recordings (right), each averaged across 15 subjects performing an attention task (see text). Trials are sorted (vertically) by reaction time; the thick black traces indicate moving-window mean reaction time (smoothing width, 11 trials/subject). Vertical dotted lines indicates stimulus onset. Obscured (grayed) areas correspond to (trial, latency) areas of non-significant difference after correction for multiple comparisons. The μV color scales are shown for the both the significant and the obscured regions. The difference ERP-image shows clearly that the consistent response, time-locked to the subject button press, is consistently larger and begins consistently earlier in the right central scalp channel C4 than in the left central scalp channel C3.

in general as it opens new avenues for the study of relationships between ongoing EEG dynamics, behavior, and experience. These methods are implemented and distributed in the EEGLAB signal-processing environment, which we develop, maintain, and make freely available to the EEG research community (Delorme and Makeig, 2004).

The grand ERP-image averaging method has some limitations. When comparing conditions with a different number of trials, obtaining a constant number of lines in the ERP-image matrices across subjects and conditions means that the correlation between neighboring rows in the ERP-image matrices might vary across subjects and conditions. The smaller the number of trials, the stronger the correlation between the rows of the (vertically) smoothed ERP-image matrix. Using the same smoothing kernel across conditions and subjects seems the simplest solution since it means that the rows of the ERP-image matrices for individual subjects are all computed in the same way and contribute equally to the grand ERP-image matrix. For all subjects, a given pixel of the ERP-image matrix represents an average of the same number of trials. Assuming uniform noise across subjects and conditions, the amount of noise in a given pixel is constant. This is the method we chose to present in this report. Note also that the pixel clustering method for correcting for multiple comparisons takes into account correlations across columns and rows of ERP-image matrices. This is because such correlations are preserved in the surrogate data distribution.

An alternative method for computing ERP-image across subjects would be to adjust smoothing kernels for each ERP-image matrix to preserve trial-to-adjacent-trial variability. A third alternative would be to select the same number of trials at random for each subject, or to construct a mean ERP-image matrix containing the required number of rows by averaging bootstrap ERP-image matrices each leaving out the needed number of trials at random. The best choice among these solutions is likely to depend on the experimental question.

A second concern relates to the time course of event-related activities during the trials. For example, ERP-images sorted by subject reaction time may clearly bring out portions of the epochs dominated by stimulus-locked and/or response-locked features. The question of how to create a grand average ERP-image matrix and plot in this case is, however, ambiguous. Simply averaging the ERP-image matrices means that the latency of the second defining event in such trials (the motor response) is not the same across subjects – the first rows of the grand ERP-image matrix will contain the averages of the fastest responses for all subjects but not the fastest absolute responses of all subjects pooled together. In this case, the simplest method might be to concatenate all trials across subjects and then form a single ERP-image matrix using whatever desired across-trials smoothing width (Makeig et al., 2004b). This would adapt to overall reaction time range differences between subjects, but would blur the response-locked features in the data, even those that were clearly revealed in the single-subject ERP-image plots. A next option would be to perform multiple regression of the trial data on the two classes of events (e.g., stimulus onsets and manual responses) and then to create ERP-image plots showing the contribution of the multiple regression model to each trial and the unaccounted trial residuals, respectively. Such an option is available in the rERP EEGLAB extension (Burns et al., 2013).

The choice of smoothing kernel is an important free variable in ERP-image analysis. Using too wide an across-trials smoothing window may mask relevant single-trial EEG details. Conversely, when the smoothing window is too narrow, features of the grand average ERP-image matrix might be obscured by irrelevant single-trial ‘noise,’ i.e., variability unrelated to the chosen sorting order. The choice of smoothing kernel will also affect the effective noise level in grand average ERP-image difference matrices. For example, as

mentioned previously, in Fig. 2, taking the difference between the ERP-image matrices for electrode channels C3 and C4 for each subject may cancel some noise (by phase cancellation). Note that this procedure is performed after trial-wise (vertical) smoothing of the individual ERP-image matrices. When no smoothing (or a quite narrow smoothing window) is being used to construct individual subject ERP-image matrices, additional (vertical) across-trials smoothing might be applied to the grand ERP-image matrix itself.

It is important to note that our procedure can be generalized to apply to any experiment design, and to arbitrary statistical measures. In Fig. 2, we have assumed a simple case of two paired conditions for one group of subjects. However, our method may be generalized to arbitrary statistical designs, by using the statistics relevant to the experimental design—to one-way repeated measure ANOVA for designs with more than two conditions, two-way repeated measure ANOVA for designs with two independent variables, and so forth. For example, in a 2×2 design it is perfectly valid to use repeated measure ANOVA on the collection of ERP-image matrices—processing each column and row independently of each other. However, in all cases, methods to correct for multiple comparisons must be used to control the family-wise error rate.

Acknowledgments

This research was supported by a gift from The Swartz Foundation (Old Field, NY) and by a grant from the US National Institutes of Health (R01 NS047293).

References

- Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 2001;29:1165–88.
- Burns M, Bigdely-Shamlo N, Smith M, Kreutz-Delgado K, Makeig S. Comparison of averaging and regression techniques for estimating event related potentials. In: *IEEE Engineering in Medicine & Biology Conference*; 2013.
- Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 2004;134:9–21.
- Delorme A, Mullen T, Kothe C, Akalin Acar Z, Bigdely-Shamlo N, Vankov A, et al. EEGLAB, SIFT, NIFT, BCLAB, and ERICA: new tools for advanced EEG processing. *Comput Intell Neurosci* 2011;2011:130714.
- Delorme A, Westerfield M, Makeig S. Medial prefrontal theta bursts precede rapid motor responses during visual selective attention. *J Neurosci* 2007;27:11949–59.
- Groppe DM, Urbach TP, Kutas M. Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 2011;48:1711–25.
- Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. Analysis and visualization of single-trial event-related potentials. *Hum Brain Mapp* 2001;14:166–85.
- Kutas M, McCarthy G, Donchin E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science (New York, NY)* 1977;197:792–5.
- Luck SJ. *An introduction to the event-related potential technique*. The MIT Press; 2005. ISBN 0-262-12277-4.
- MacPherson JM, Aldridge JW. A quantitative method of computer analysis of spike train data collected from behaving animals. *Brain Res* 1979;175:183–7.
- Makeig S, Debener S, Onton J, Delorme A. Mining event-related brain dynamics. *Trends Cogn Sci* 2004a;8:204–10.
- Makeig S, Delorme A, Westerfield M, Jung TP, Townsend J, Courchesne E, Sejnowski TJ. Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biol* 2004b;2:e176.
- Makeig S, Westerfield M, Jung TP, Covington J, Townsend J, Sejnowski TJ, Courchesne E. Functionally independent components of the late positive event-related potential during visual spatial attention. *J Neurosci* 1999;19:2665–80.
- Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 2007;164:177–90.
- Onton J, Makeig S. High-frequency broadband modulations of electroencephalographic spectra. *Front Hum Neurosci* 2009;3:61.
- Pernet CR, Latinus M, Nichols TE, Rousselet GA. Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: a simulation study. *J Neurosci Methods* 2015;250:85–93.
- Ritter W, Simson R, Vaughan HG. Association cortex potentials and reaction time in auditory discrimination. *Electroencephalogr Clin Neurophysiol* 1972;33:547–55.