In IEEE Engineering in Biology and Medicine Conference, Osaka, Japan 2013

Comparison of Averaging and Regression Techniques for Estimating Event Related Potentials

Matthew D. Burns, Nima Bigdely-Shamlo, Nathaniel J. Smith, Kenneth Kreutz-Delgado, Scott Makeig

***Abstract*— The traditional method of estimating an Event Related Potential (ERP) is to take the average of signal epochs time locked to a set of similar experimental events. This averaging method is useful as long as the experimental procedure can sufficiently isolate the brain or non-brain process of interest. However, if responses from multiple cognitive processes, time locked to multiple classes of closely spaced events, overlap in time with varying inter-event intervals, averaging will most likely fail to identify the individual response time courses. For this situation, we study estimation of responses to all recorded events in an experiment by a single model using standard linear regression (the *rERP technique*). Applied to data collected during a Rapid Serial Visual Presentation (RSVP) task, our analysis shows: (1) The rERP technique accounts for more variance in the data than averaging when individual event responses are highly overlapping; (2) the variance accounted for by the estimates is concentrated into a fewer ICA components than raw EEG channel signals.**

I. Introduction

The Event Related Potential (ERP) averaging method for electroencephalographic (EEG) data [1] is one way to gain insight into how specific cognitive processes are related to brain electrical activity. Traditionally, the way of increasing the signal to noise ratio (SNR) of an ERP estimate is to average epochs time-locked to a stimulus class of interest. This technique places severe restrictions on the experimental protocol: only a small number of stimulus categories can be used, stimulus events must be well separated in time and all other cognitive processes must be held constant. Violating the latter conditions will cause the ERP to be estimated sub-optimally. Here we study using multiple regression as a way to overcome this limitation, extending the work of N. J. Smith [2]. In [3], Hinrichs et al. have suggested a highly similar approach for deconvolving fMRI responses. Hauk et al. and Pernet et al. [4, 5] have suggested using separate regression models for each individual latency, such as massive univariate general linear analyses. Hendrix et al. in [6] have proposed using Generalized Additive Models (GAMs). In [2], author Smith offers a unified conceptual framework for ERP regression and shows how these different techniques relate to averaging for the purposes of ERP estimation.

We continue this discussion by applying linear regression and averaging to a real EEG dataset and exhaustively comparing the results of the two approaches. The goal is to make clear that in practice, regression can offer a significant performance increase compared to averaging. Indeed, as EEG experiments become more sophisticated, with many (intermittent or continuous) processes being monitored simultaneously, averaging ceases to be an effective option. Independent Component Analysis (ICA) [7] has become a popular and often effective method for separating EEG sources [8, 9]. Thus, we also compared how regression and averaging compare with one another in both ICA component activations (ICs) and EEG channels.

II. Background

A. A Problem With Averaging

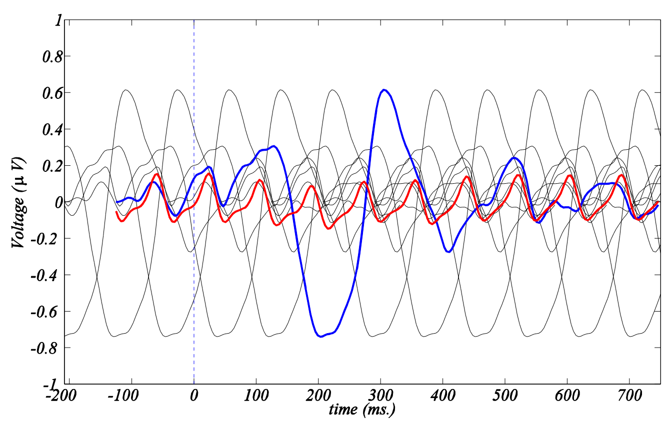

Figure 1. Illustrating how averaging can produce an incorrect ERP estimate in the presence of overlapping activity due to closely spaced cognitive events. The latency window is a typical EEG epoch in a 12/s rapid serial visual presentation (RSVP) experiment. An ERP of interest (*blue*), is produced following each visual stimulus every 83 ms (*black*). These ERPs combine additively, giving a misleading (*red*) averaged Steady State Response (SSR) ERP estimate. Regression considers all the experimental events in a single additive model, taking into account this overlap.

*Research supported by a gift from The Swartz Foundation (Old Field NY).

M. D. Burns, N. Bigdely-Shamlo and K. Kreutz-Delgado are with Dept. of Electrical and Computer Engineering, University of California, San Diego, San Diego, CA 92093 USA (phone: 858-822-7534; fax: 858-822-7556; email: mdburns@ucsd.edu, nima@sccn.ucsd.edu, kreutz@eng.ucsd.edu).

N.J. Smith is at the School of Informatics, University of Edinburgh, Informatics Forum, 10 Crichton Street, Edinburgh, EH8 9AB, Scotland, United Kingdom (email: nathaniel.smith@ed.ac.uk).

S. Makeig directs the Swartz Center for Computational Neuroscience (SCCN) of the Institute for Neural Computation (INC), University of California San Diego (email: smakeig@ucsd.edu).

If events in an experiment occur sufficiently close in time to one another, the EEG brain responses to these events will overlap. Taking an average of these event time-locked epochs will produce a summed and/or blurred ERP estimate.

B. Data

The experiment is fully described in [10]. 127-channel EEG data were collected during a Rapid Serial Visual Presentation (RSVP) task involving satellite picture presentation. The subject was shown bursts of 49 satellite images in 4.1 seconds (12/s.). In 60% of the bursts, a (flying airplane) target feature was randomly added to one image. At each burst end, the participant indicated by button press whether or not that burst contained the target feature. During training, they were told whether they were correct or not.

There were nine recorded event types in the experiment, listed by *(event-code) event-description*: (1) non-target stimulus, (2) target stimulus, (4) “no targets” button press, (5) “one target” button press, (6) trial block start, (16) trial start, (32) “correct” feedback given, (64) “incorrect” feedback shown, and (129) image burst start.

III. METHODS

We calculated ERP estimates for a seven subject/12 session study across nine different events using averaging and linear regression with Ordinary Least Squares (OLS). This analysis was repeated for all 127 channels of EEG data and again for all 127 ICs, derived by extended Infomax ICA [11]. We used five-fold cross-validation to obtain our performance figures: the ERPs were calculated with training data and validated on test data [12].

A. Preprocessing

First, we addressed the issue of outliers and artifacts. We identified outlier data portions by two methods: Low Probability and Mutual Information Reduction (MIR). For the probability method we first whitened the data and performed a rank transform to obtain a two-tailed significance value for each sample. We then found 200 ms windows where the average log significance over all the sphered dimensions and time-frames was higher than 2.1 and marked them as outliers. For the MIR method, we first calculated the mutual information reduction index [13] in 2s windows with 80% overlap using the sphering matrix. Then we found regions with MIR Z score of lower than −1.5 and marked them as outliers. We discarded events occurring during or near outlier periods. Out of 23,477 events, 1,654 were identified as contaminated and discarded. The data were highpass filtered (−3 dB at 1 Hz) to reduce DC bias.

All ERPs were estimated using the same maximum length, heuristically set for this analysis at 1 second (256 samples), from −125 ms to 875 ms around each event. This defined 256 variables per event. For nine event types, each regression or averaging model thus contained 2304 ERP parameters for each EEG channel or IC.

B. Regression Framework

First we looked at the case of only one event type, E_1, producing an ERP response β_1. The observed signal (IC or channel) y is then a linear transformation of β_1 plus a Gaussian noise term, $\varepsilon \sim N(0, \sigma I)$.

$$\beta_1 = [\ \beta_{11}\ \beta_{12} \ldots \beta_{1N}\ ]^T \quad (1)$$

$$y = \boldsymbol{A_1}\beta_1 + \varepsilon \quad (2)$$

We position y and β_1 as a column vectors of length M (the length of the data) and N (= 256) respectively. $\boldsymbol{A_1}$ is the M x N matrix of *predictors*, x_{mn}, constructed from latency recordings. x_{mn} has a value of 1 when the n^{th} sample of ERP β_1 is predicted to occur at latency m.

If we want to estimate the response to more than one event type, we stack the β_n in a column vector, and concatenate their corresponding $\boldsymbol{A_n}$ along the second dimension

$$\boldsymbol{A} = [\ \boldsymbol{A_1}\ \boldsymbol{A_2} \ldots \boldsymbol{A_n}\ ] \quad (3)$$

$$\beta = [\ (\beta_1)^T\ (\beta_2)^T \ldots (\beta_n)^T\ ]^T \quad (4)$$

and subsequently

$$y = \boldsymbol{A}\beta + \varepsilon \quad (5)$$

with least squares solution

$$\beta_{reg} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T y \cong \beta \quad (6)$$

C. Performance Metrics

We subtract the ERP estimates from the original signal to obtain a residual noise signal. The difference between the variance (power) of the original signal and the variance of the noise signal represents the variance accounted for by that ERP. We use this Reduction of Variance (ROV), as our metric, with higher ROV corresponding to better performance.

$$ROV \equiv (P_{Data} - P_{Noise}) \quad (7)$$

The reason for using ROV instead of Signal to Noise Ratio (SNR)

$$SNR \equiv P_{Signal} / P_{Noise} \quad (8)$$

is that we aren’t especially interested in maximizing the size of the ERP estimate (the “signal” in this case). ROV measures to what extent the estimate accounts for overall variance in the data.

For each event type in the experiment, we computed the ROV for averaging by extracting each epoch (y_i) and subtracting the averaged ERP estimate β_{av} from it.

$$\varepsilon^{av}{}_i = y_i - \beta_{av} \quad (9)$$

$$ROV_{av} = \langle\ var[y_i] - var[\varepsilon^{av}{}_i]\ \rangle \quad (10)$$

where the mean is taken across all the events of that type. For regression, we computed a signal estimate

$$y^{reg} = \boldsymbol{A}\beta_{reg} \tag{11}$$

then, extracted each epoch from the estimated signal ($y^{reg}{}_i$) and the original signal (y_i).

$$\varepsilon^{reg}{}_i = y_i - y^{reg}{}_i \tag{12}$$

$$ROV_{reg} = < var[y_i] - var[\varepsilon^{reg}{}_i] > \tag{13}$$

We use normalized estimates from (14) to identify which channels/ICs have the highest ROV percentage.

$$ROV' = ROV/var[y_i] \tag{14}$$

ROV′ of the top 20 channels/ICs from each dataset are averaged to obtain the final estimates for each event type. For significance testing we applied a two-sample *t*-test ($p < 0.01$) to the cross-validation folds of all twelve datasets.

IV. Results

A. ERP Estimates (Figure 2)

Due to the 83-ms stimulus onset asynchrony, simple average responses to non-target type 1 events had significant confound from overlapping responses. As shown in *Figure 2*, the averaged estimate does not reflect the ERP associated with a single non-target frame (*Figure 1* shows graphically how this occurs). In this case, regression recovered a plausible visual response to each non-target stimulus event. Event type 6 did not usually occur near any other experiment events. Here, as expected, regression and averaging gave similar results.

B. Performance as Measured by ROV (Figure 3)

*Event type 1 (*Figure 3*, top panel) shows the most significant difference between the two methods for both ICs and channels. For the most frequent event type 1,* regression has the advantage for both channel and IC measures. Compare the difference in the regression versus the average ERP (here, SSR) estimates in *Figure 2* (top panel). The averaging method clearly did not estimate the ERP for this event type. For the other event types, which

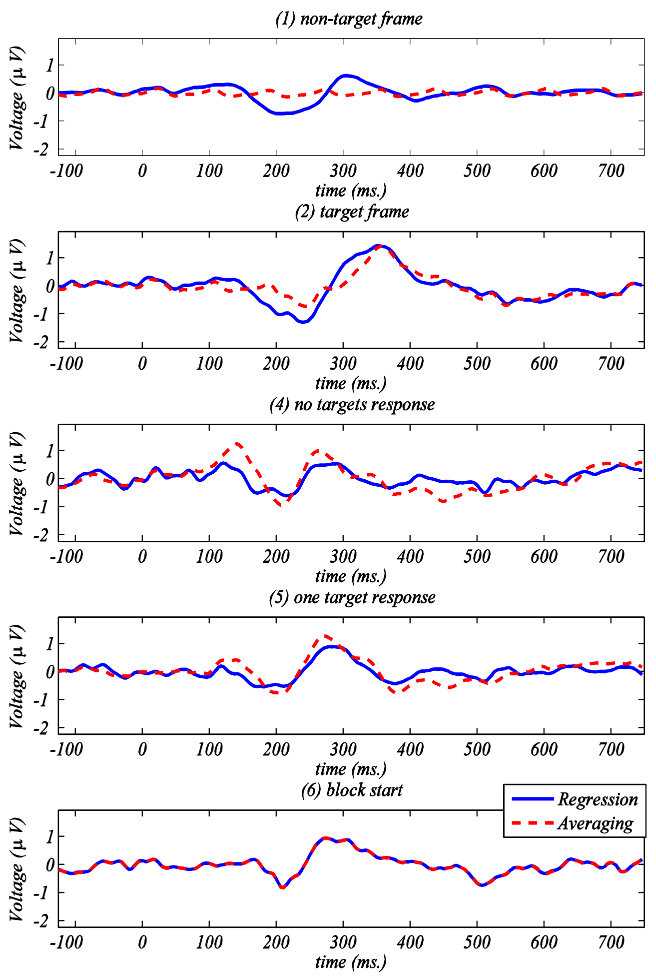

Figure 2. Comparison of ERP estimates by averaging (*red*) and regression (*blue*) for five event types. ROV was statistically higher for regression in event type 1. stimates are for a lateral occipital IC.

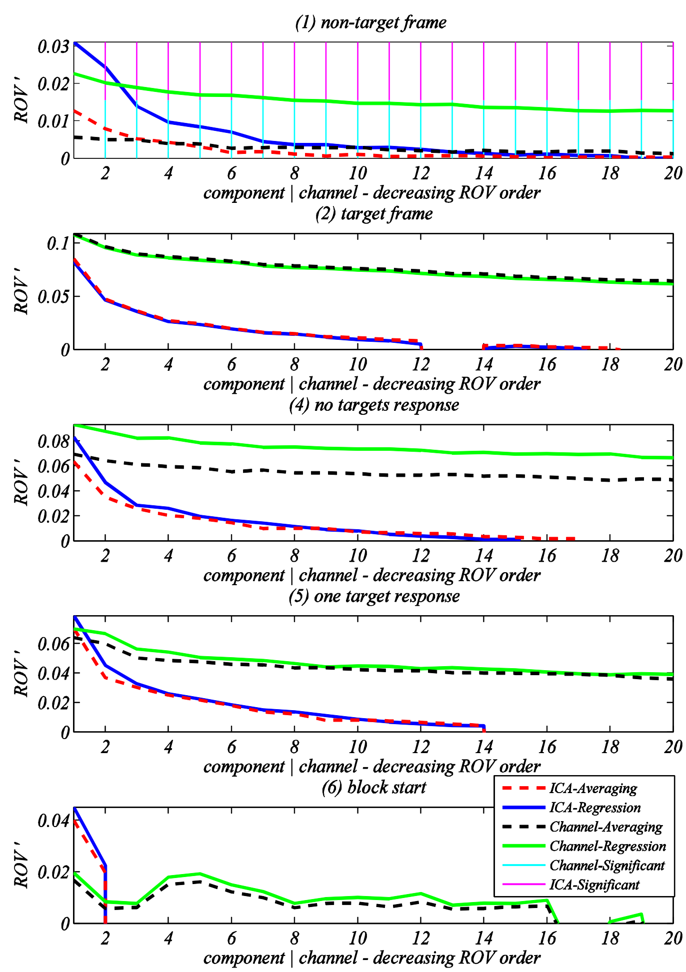

Figure 3. Comparison between averaging and regression ROV for EEG channels and ICA components. The components and channels are sorted and plotted in normalized ROV form.

are less affected by overlap, the two methods performed similarly.

Comparing IC and channel results, we notice a peak of ROV in the first 2-3 ICs for each stimulus type. Since ICs are thought to typically represent the synchronous field activity across a single cortical patch [13, 14], broadly projected to the scalp electrodes by volume conduction, we may expect the regression result to show higher ROV for a smaller number of ICs than scalp channels. The ROV for channels is indeed distributed across a larger number of channels. This is expected, since EEG signals at scalp electrodes that are physically close are highly correlated [8].

Note that normalized ROV is quite low across the board: no more than 12% of the variance in any channel or component signal is accounted for by either method, and usually much less. This is consistent with the frequent observation that most EEG signal variance is not produced by time and phase locked responses to external events.

V. Discussion

As demonstrated in [2], OLS regression can be thought of as a natural extension of event-related potential averaging that can be applied in a larger range of experimental conditions. Mathematically, OLS reduces to averaging when there is no overlap between experimental responses (e.g. *Figure 2*, bottom panel).

Our ROV analysis showed that regression is capable of explaining more variance in experimental data than averaging. This overall comparison is limited in the sense that it cannot tell us whether a certain *portion* of an ERP waveform is best represented by either method. In other words, whether or not a specific peak in a response is better estimated by averaging or regression cannot be decided from this analysis alone. The potential benefit of regression is only clear when considering an ERP as a whole, and should depend on its degree of overlap with responses to other experimental events.

Possible extensions: Since OLS is the simplest estimator beyond event-locked averaging, the predictive performance of our model might be expected to increase if it used a more modern estimator. A problem with estimating EEG parameters by OLS is that artifacts can drastically affect its L2-norm error function. The Least Absolute Deviations (LAD) [15] technique uses, instead, an L1-norm error function, and thereby may provide a more robust estimator. The performance of the model is also highly sensitive to its number of parameters. Introducing regularization on the ERP parameters would be a reasonable way to control for this effect and discourage over-fitting.

VI. Conclusion

When overlapping evoked responses are produced by experimental events that are closely spaced in time, multiple stimulus events may contribute to any given average event-related potential (ERP) feature and some additional assumption is necessary to properly segregate this variance. The regression ERP (rERP) technique assumes that ERPs to distinct events sum linearly, even when they are closely spaced in time. In all other ways, the rERP and ERP measures are identical. Yet, as we show here, the rERP approach can account for more total data variance, showing that the rERP assumption is viable for analysis of event-related potential information in rich, complex EEG data sets.

VII. References

[1] S. J. Luck, *An introduction to the event-related potential technique*, 2005.

[2] N. J. Smith. (2011). *Scaling up psycholinguistics*. Available: http://wwwlib.umi.com/cr/fullcit?p3472953

[3] H. Hinrichs, M. Scholz, C. Tempelmann, M. G. Woldorff, A. M. Dale, and H. J. Heinze, "Deconvolution of event-related fMRI responses in fast-rate experimental designs: tracking amplitude variations," *Journal of Cognitive Neuroscience,* vol. 12, pp. 76-89, 2000.

[4] O. Hauk, M. Davis, M. Ford, F. Pulverm ller, and W. Marslen-Wilson, "The time course of visual word recognition as revealed by linear regression analysis of ERP data," *Neuroimage,* vol. 30, pp. 1383-1400, 2006.

[5] C. R. Pernet, N. Chauveau, C. Gaspar, and G. A. Rousselet, "Limo EEG: a toolbox for hierarchical linear modeling of electroencephalographic data," *Computational intelligence and neuroscience,* vol. 2011, p. 3, 2011.

[6] P. Hendrix, W. Tabak, R. Schreuder, and R. Baayen, "nd, Electrophysiological effects in language production: a picture naming study using generalized additive modeling," ed: preparation, 2009.

[7] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks,* vol. 13, pp. 411-430, 2000.

[8] S. Makeig, A. J. Bell, T. P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," *Advances in neural information processing systems,* pp. 145-151, 1996.

[9] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of neuroscience methods,* vol. 134, pp. 9-21, 2004.

[10] N. Bigdely-Shamlo, A. Vankov, R. R. Ramirez, and S. Makeig, "Brain activity-based image classification from rapid serial visual presentation," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on,* vol. 16, pp. 432-441, 2008.

[11] S. Makeig, T. P. Jung, A. J. Bell, D. Ghahremani, and T. J. Sejnowski, "Blind separation of auditory event-related brain responses into independent components," *Proceedings of the National Academy of Sciences,* vol. 94, pp. 10979-10984, 1997.

[12] P. Burman, "A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods," *Biometrika,* vol. 76, pp. 503-514, 1989.

[13] A. Delorme, J. Palmer, J. Onton, R. Oostenveld, and S. Makeig, "Independent EEG sources are dipolar," *PloS one,* vol. 7, p. e30135, 2012.

[14] S. Makeig, M. Westerfield, T. P. Jung, S. Enghoff, J. Townsend, E. Courchesne, *et al.*, "Dynamic brain sources of visual evoked responses," *Science,* vol. 295, pp. 690-694, 2002.

[15] P. Bloomfield and W. L. Steiger, *Least absolute deviations: theory, applications, and algorithms*: Birkhäuser Boston, 1983.