An Actuarial Approach to the Identification of Artifacts in Large Data Sets of Frontal EEG Asymmetry

David N. Towers and John J.B. Allen
University of Arizona

Abstract and Overview

The superiority of actuarial approaches over human judgment in classification is well documented, suggesting their utility in classifying artifact-laden segments of EEG data. In studies of frontal alpha asymmetry, relevant scores are derived from mean alpha-band power across the EEG record. Artifacts in the EEG record will bias such estimates, sometimes drastically. An ongoing study of frontal alpha asymmetry has, to date, collected over 100 hours of 64-channel EEG data. The size of this dataset requires that the process of manually rejecting artifacts be distributed over a number of scorers. Estimates of between-rater agreement on duplicate files are typically high, suggesting that many artifacts are easily recognized and reliably rejected across the pool of scorers. Despite this consistency, initial analyses suggest that occasional artifacts still remain, in some cases severely biasing the overall estimate of asymmetry. Manual identification of alpha-band outliers (e.g. via scatter-plots) is not practical, again due to the size of the dataset. Several algorithm-based methods to identify artifacts are discussed, as is the relative performance between methods with respect to identifying known artifacts. These methods can in turn provide an actuarial data-driven set of criteria for artifact rejection.

The following presents two approaches to artifact detection: detection of bad channels within an EEG record, and detection of artifacts within otherwise good data.

General Methods

Subjects

- 104 paid undergraduate students (77 female)
- 19 met criteria for current Major Depressive Disorder (MDD), 41 met criteria for past MDD
- Beck Depression Inventory scores ranged from 0 to 43 (mean=10.7, median = 10)

EEG Procedure

- EEG data were recorded on 4-6 separate occasions, with two 8-minute resting periods on each occasion, comprised of eyes open and closed recordings in one of two orders (COOCOCCO, OCCOCOOC).
- Scalp EEG recorded from 64 channels (including mastoids), with an online reference located between Cz and CPz. Vertical and horizontal EOG were recorded using bipolar leads. All data were low-pass filtered at 200Hz, and digitized at 1000Hz.
- Data were initially reviewed for bad channels and artifacts by a number of different scores (blinks and most other ocular artifacts were not removed, as an automated rejection algorithm was to be applied at a later stage).

Bad Channel Detection

Method

- EEG data consisted of all data within each one-minute rest period, excluding segments rejected during the initial scoring.
- Remaining raw EEG data (i.e. within each 1-minute rest period, excluding rejected segments) were digitally filtered (FIR band-pass, -3dB at 0.5 and 57Hz). This band was selected so that the derived metrics would be applicable to a variety of EEG studies (e.g. ERP and non-alpha band oscillations)
- Raw EEG files were again assessed by a single rater (first author), to establish a "ground truth" regarding whether channels were bad.
- Two metrics were derived from the raw EEG data:

Z Metric

The rationale for this metric was that activity at a good site should be highly correlated with neighboring sites. The Z metric was derived by fitting a regression between the EEG time series of the channel of interest and its four nearest neighbors. A Fisher z-transform was applied to the resulting correlation coefficient (*r*), then normalized with respect to that site across all records.

$$Z = \frac{z(r,c_i^j) - \overline{z(r,c_i)}}{\sigma[z(r,c_i)]}$$

$z(r,c_i^j) =$ Z-transformed correlation coefficient for channel *i*, in file *j*

R Metric

The rationale for this metric was that bad sites will show excessive deviations from baseline. The R metric is the root-mean-square amplitude estimate of the filtered EEG data, again normalized by site across all records.

$$R = \ln\left(\frac{rms_i^j - \mathrm{median}(rms_i)}{\mathrm{IQR}(rms_i)}\right)$$

$rms_i^j =$ rms amplitude of channel *i*, in file *j*

- For both metrics, cut-points were chosen such that both sensitivity and specificity were maximized, and sensitivity was greater than specificity. Posterior probability of bad channel detection was calculated for separate and combined metrics

Results

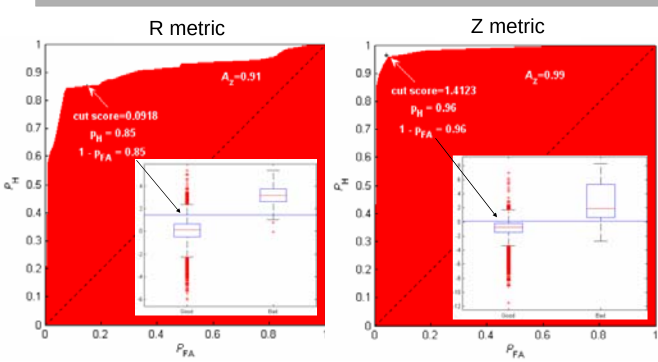

Figure 1. ROC Curves for Z and R metrics in identifying bad channels. "Ground Truth" is ratings made by first author.

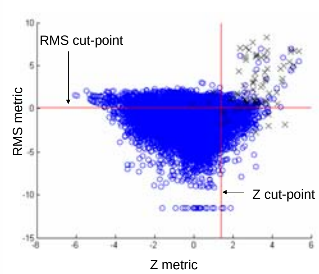

Figure 2. Distribution of Z and R metrics (X indicates a bad channel)

p(B)=	0.002		
p(Z|B)=	0.962	p(B|Z) =	0.034
p(~Z|~B)=	0.957	p(~B|~Z) =	0.957
p(R|B) =	0.848	p(B|R) =	0.009
p(~R|~B) =	0.848	p(~B|~R) =	0.848
p(B|Z,R) =	0.165	p(~B|~Z,~R) =	1.000
p(B|~Z,~R) =	0.000	p(~B|~Z,R) =	1.000
p(B|Z,~R) =	0.006	p(~B|Z,~R) =	0.994
p(B|~Z,R) =	0.000	p(~B|Z,R) =	0.835

Table 1. Posterior probabilities of bad channel detection, for separate and combined metrics (Z = Z metric > cut-score, R =metric > cut-score, B = bad channel).

Discussion

- The overall probability of detecting bad channels is low despite the reasonable sensitivities and specificities of the above metrics, largely due to the low base rate of bad channels
- In contrast, the combined metrics perform well in determining whether a channel is acceptable.
- Performance may be improved by using a "successive-hurdles" approach, where additional metrics could be derived to discriminate good from bad channels in cases where the above metrics indicate the channel is bad. As an example, a number of the "false-alarm" channels show high EMG activity; as such, it may be useful to employ additional metrics to detect whether EMG increases overall amplitude, or decreases the correlation in signal between adjacent sites.

David N. Towers and John J.B. Allen

Artifact Detection

- Development of epoch artifact detection metrics focused on finding those metrics that were robust with respect to differences between subjects (e.g. blink rates, global cortical activity) as well as between channels (e.g. variation in amplitude across sites due to the position of the online reference, differences in ocular activity). The following describes several candidate metrics.

Method and Discussion

- For all the metrics described below, complex FFT data was calculated for both the original signal and its derivative (via first-order difference between samples).
- Data segmented into 2.048 sec epochs overlapping by 75%
- Complex FFT from 0.1 to 30Hz derived for each epoch after application of hamming window to both the original and derivative signal.
- Metrics evaluated for blink and step artifact examples, as well as a segment as posterior alpha burst. The goal is to find metrics that are sensitive to artifacts but insensitive to large changes in actual EEG signal.

Epoch-removed mean power difference

For each epoch *n*, the difference between mean power across all epochs $\left(\sum_{i=1}^{N} \mathrm{Pow}(Epoch_n)/N\right)$ and mean power with epoch removed $\left(\sum_{i=1}^{N} \mathrm{Pow}(Epoch_i)/N-1, i \neq n\right)$

Normalized power

Power in the original signal, with respect to median and inter-quartile power for all epochs within channel

Normal-Derivative Spectrum Correlation

Correlation between original and derivative spectrum amplitude

Original-Derivative Spectrum Euclidian Distance

Euclidian difference between original and derivative signal spectrum amplitude (taken as vectors)

Phase/Amplitude Chi-Square

Normal signal complex phase separated into 10 bins; chi-square evaluated over the observed summed amplitude for each bin, compared to summed amplitude across all frequencies

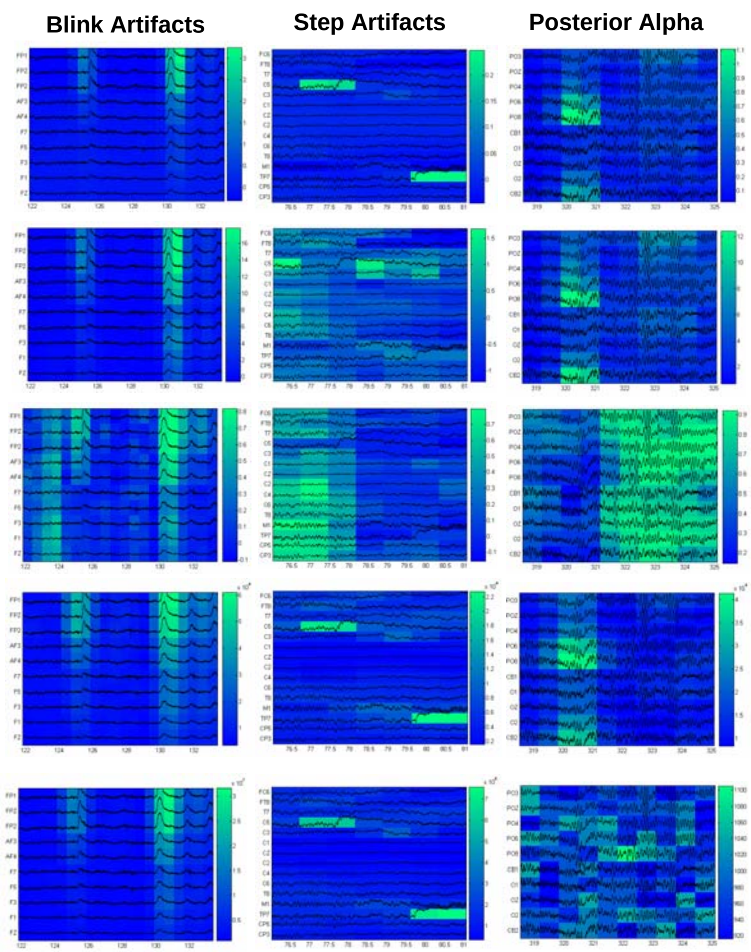

Figure 3. Raw EEG with blinks, step artifacts, and posterior alpha (columns) superimposed on metric value (rows) for the corresponding epoch

- For the above examples, all of the metrics appear sensitive to blink artifacts, and several appear sensitive to the step artifacts. The Phase/Amplitude Chi-Square exhibits some specificity as well, in that the metric is relatively small for the non-artifactual posterior alpha burst.
- Overlapping epochs pose a challenge for developing detection algorithms, in that a relatively short artifact can "spread" across multiple epochs, suggesting that other time-frequency transforms (e.g. wavelets) may perform better.

Reference

Delorme, A. & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics, *Journal of Neuroscience Methods, 134*, 9-12.

The authors wish to thank Dara Halpern, Eliza Ferguson, Jamie Velo, Craig Santerre, Eynav Elgavish Accortt, Andrew Bismark, Jay Hegde, and myriad research assistants for their efforts in recruiting and testing participants, and to thank Jim Coan for his invaluable contributions leading to the receipt of NIMH R01-MH066902, which funded portions of the collection of these data. This work was also funded, in part, by a grant from the NARSAD foundation.

Reprints available from www.psychofizz.org