NIC Technical Document # 9
Bob Frank
Gwen Frishkoff
Kevin Glass

Application of ICA for Blink Removal using the APECS framework: Preliminary Results

Introduction

The present report describes a revision and expansion of the methods outlined in [1]. APECS is a collection of MATLAB m-files, based upon Dr. Joseph Dien's ICA Toolbox [2]. APECS is designed to remove eyeblinks from EEG data using independent component analysis (ICA) and a set of user-specified criteria for classification of components as blink-related [3].

In [3] we described an improved "baseline" (blink-free) dataset, which was used to generate seven sets of "synthetic" EEG data. These datasets consist of the baseline (blinkfree) data, combined with simulated blink streams varying in blink duration, amplitude, and frequency (see [4] for details). FastICA and Infomax algorithms were applied to these data, and the spatial projections of each independent component were correlated with a predefined "blink template" (see [4] and [5] for details). If the correlation exceeded a specified threshold (or "tolerance") for a particular source, and if the spatial topography of the source showed a polarity inversion above and below the eyes, then the source was defined as blink-related and was removed from the data, resulting in an ICA-filtered or "cleaned" dataset.

To evaluate the quality of each ICA decomposition, we then compared the ICA-filtered and the original baseline data. Given that the baseline data was verifiably absent of blink activity [2], it was assumed that lack of correspondence between the ICA-filtered and baseline datasets reflected problems with the decomposition, either due to failure to detect and remove blink activity ("misses") or to removal of non-blink activity that was incorrectly identified as blink-related ("false alarms"). Further procedures, such as spectral and temporal correlations, were then conducted to gain further insight into the nature of the independent components that were classified as blink-related, and to suggest refinements to the APECS framework for separation and classification of independent components.

Methods

General methods are detailed in NIC [3]–[4]. For this report, we compare a single FastICA run (using the hyperbolic tangent function) and a single run with Infomax. The algorithms were run using the default parameters: [Bob fill in…]

Results

In general, the results for this set of decompositions was excellent, even compared with the results from our previous study [1]. Averaging across detectors, across datasets, and across runs, the correlation between the ICA-filtered and original (baseline) data was 98.67%. In many instances, the correlation between the two time series was better than 99%. The best results were obtained with the tolerance (threshold for correlation between the spatial projector for a source and the blink template) was set to .95 (mean correlation = __). For example, Figure 1 shows the average correlation between the baseline data and the ICA-filtered data for each of the seven datasets when the tolerance was set to .95. DS5: Infomax – 1 IC removed. FastICA – 2 components removed [verify]

[revise figure 1 to include other 2 fastica runs]

FastICA vs. Infomax: Correlations (.95 tolerance)

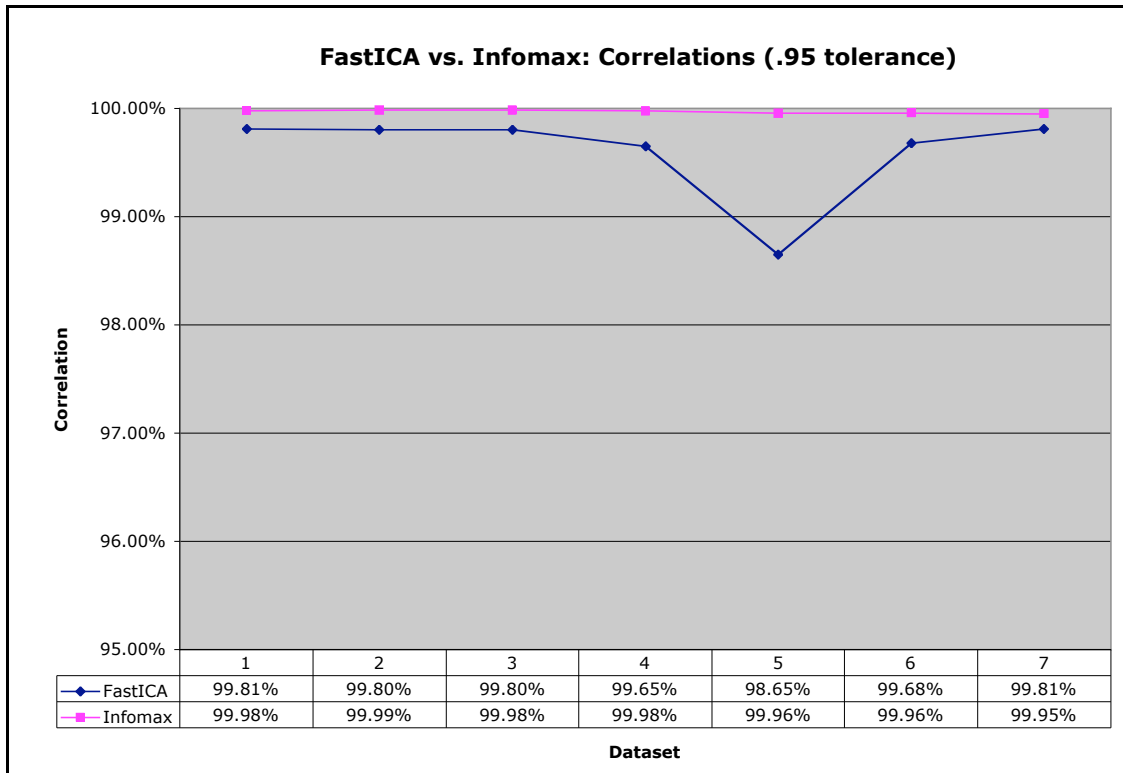| Dataset	1	2	3	4	5	6	7
FastICA	99.81%	99.80%	99.80%	99.65%	98.65%	99.68%	99.81%
Infomax	99.98%	99.99%	99.98%	99.98%	99.96%	99.96%	99.95%

Figure 1. Correlation between baseline data and ICA-filtered data (tolerance = .95)

Figure 2 shows the ICA-filtered EEG waveforms for Dataset 5 (Infomax run and 2 of the 3 FastICA runs), overplotted with the baseline data. To emphasize the blink-related activity, the data were segmented into 3-second intervals, centered around the peak of each simulated blink event. The data were then averaged across segments to create a blink-averaged summary of the data (or blink ERP): the logic is that activity that is timelocked to the blinks will become more pronounced, whereas activity that is unrelated will tend towards zero. Thus,

activity that is timelocked to the blinks is exaggerated relative to non-blink activity. As shown in Figure 2, the Infomax-filtered data (blue line) correspond very closely with the baseline data (black line). For the two FastICA runs (red and green lines), however, there is a somewhat stronger divergence from the baseline approximately 100ms before and 100ms after the peak of the blink activity. The divergence is strongest above the eyes (at detectors Fp1 and Fp2), and much less visible at detectors further from the peak of the ocular activity. This suggests that, revaltive to Infomax, FastICA was slightly less effective in removing activity associated with the simulated blinks.
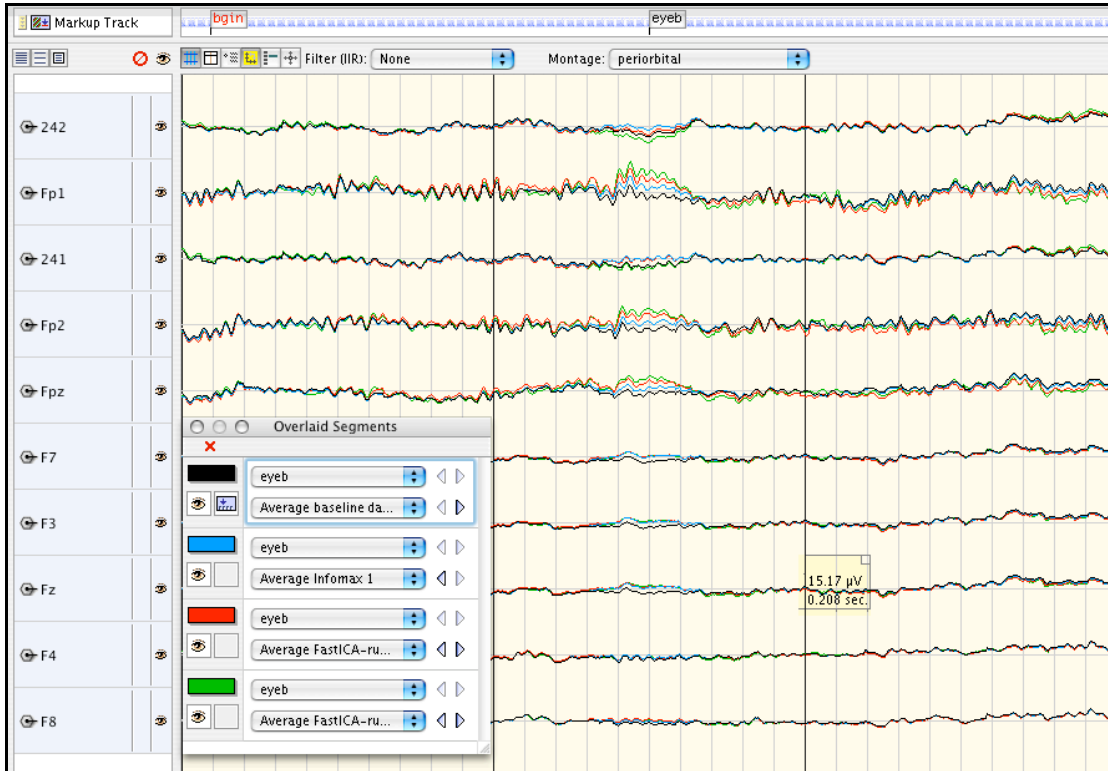

Figure 2. ICA-filtered data for Dataset #5 for four runs (Infomax and 3 FastICA runs).

When the tolerance was set to .90, there were __ blink matches for Infomax, ___ matches for FastICA run 1 and ___ matches for FastICA run 2. As shown in Figure 3, the correlation drops somewhat when these additional sources were removed from the data (mean correlation = __).

[NEED TO ADD TO XLS SPREADSHEETS LOG WITH NUMBER OF BLINK MATCHES FOR EACH RUN]

FastICA vs. Infomax: Correlations (.90 tolerance)

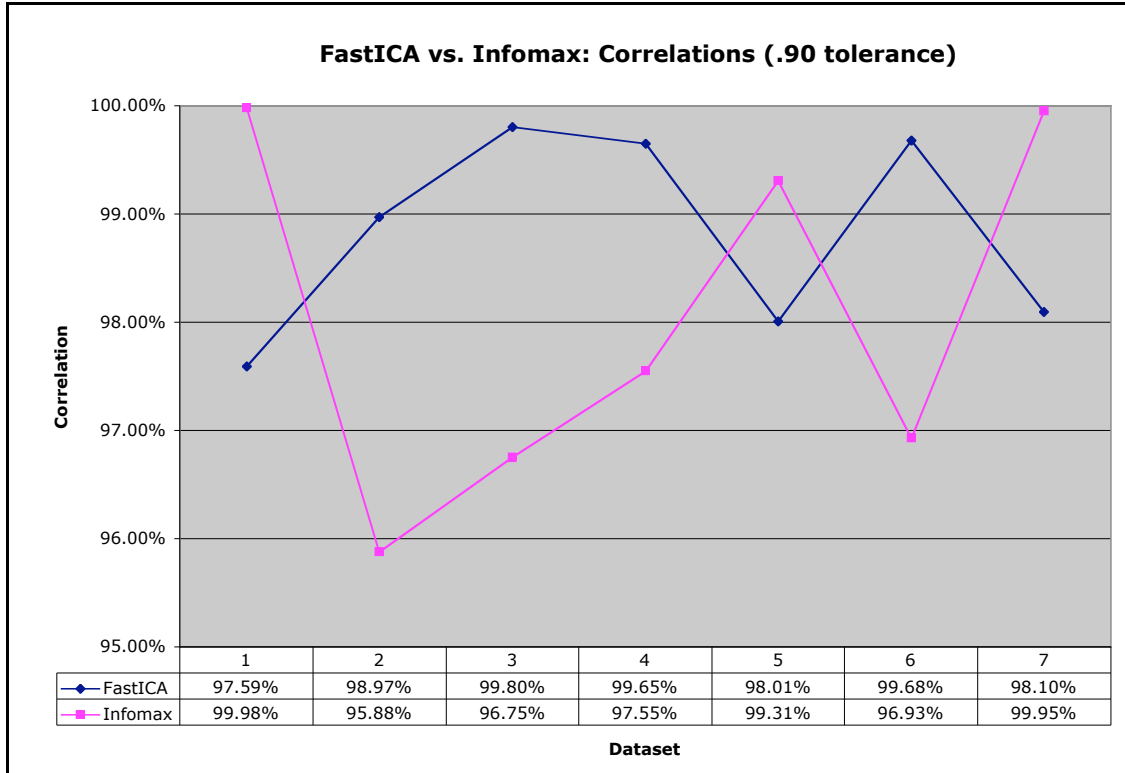| Dataset	1	2	3	4	5	6	7
FastICA	97.59%	98.97%	99.80%	99.65%	98.01%	99.68%	98.10%
Infomax	99.98%	95.88%	96.75%	97.55%	99.31%	96.93%	99.95%

Figure 3. Correlation between baseline data and ICA-filtered data (tolerance = .90)

[THEN NEED WAY TO CLASSIFY AND INTERPRET REMAINING COMPONENTS]
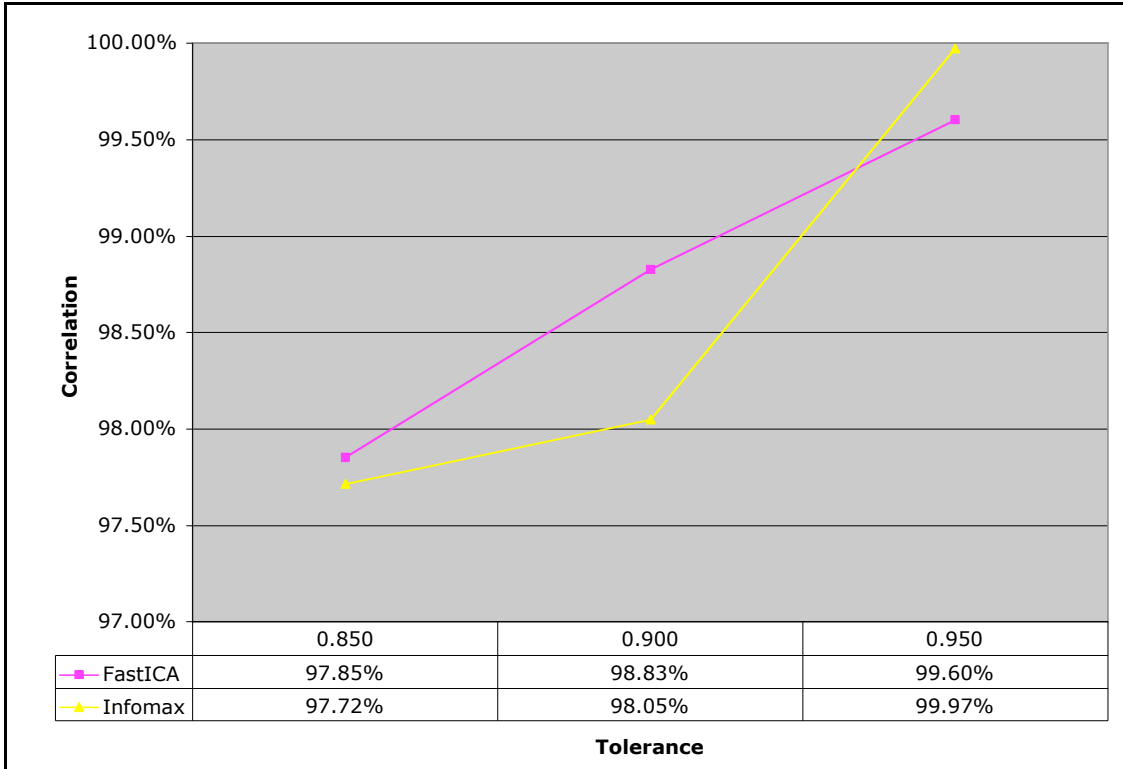Figure 4 shows how the solution is degraded as additional components are removed from the data.

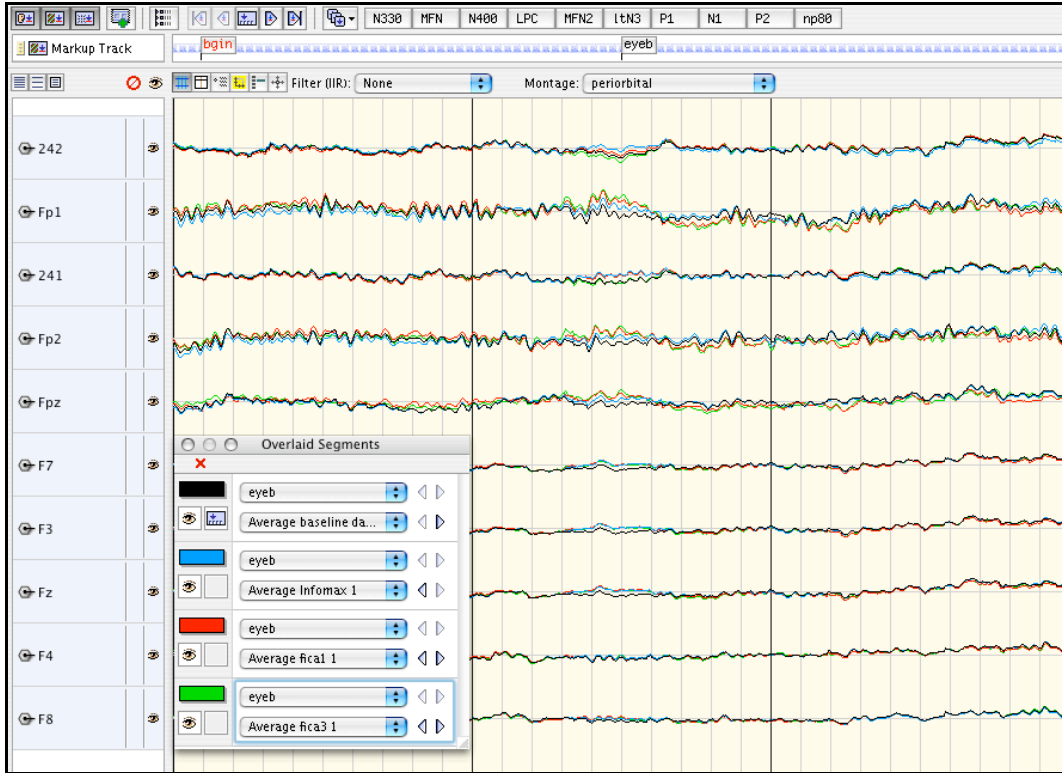Figure 4. Correlation between filtered and baseline data at each tolerance (averaging across datasets).

Figure 5. ICA-filtered data for Dataset #5 for four runs (Infomax and 3 FastICA runs), removing all matches when tolerance = .85

Figure 6. ICA-filtered data for Dataset #2 for four runs (Infomax and 3 FastICA runs), removing all matches when tolerance = .85

Summary & Conclusions
Given the improved baseline, the differences between FastICA and Infomax appear less pronounced than we observed in the first set of experiments [1]. Both algorithms proved fairly successful when the threshold for correlation with the blink template (or "tolerance") was set to .95. In this case, the activity corresponding to only a single independent component was removed from the data: the average correlation between the filtered data and the baseline data was >99%. The correlation dropped slightly (~97%) when the tolerance was set to .90, and more than one component was removed. The same pattern was observed for both the blink and interblink intervals. This leads us to consider how best to classify and interpret these other components, which match our blink criteria (correlation with blink template and vertical polarity inversion). We suggest the need for temporal as well as spatial criteria for IC classification, to avoid false positives and to help distinguish between the following outcomes (reasons for multiple matches to the blink template):

1. Splitting of blink factor (e.g., because resulting components are not truly independent)
2. False positives, i.e., matches due to nonblink activity
3. Matches due to ocular activity of a different nature from phasic blinks (need for more refined understanding of blink activity).

[Need for temporal and spectral criteria….]

question about whether the baseline was really blinkfree. Ack!! See notes.

References

[1] Glass, K., Frishkoff, G., Frank, R., Davey, C., Dien, J., Malony, A., & Tucker, D. (2004, September 2004). *A Framework for Evaluating ICA Methods of Artifact Removal from Multichannel EEG.* Paper accepted to the ICA 2004 Conference, Grenada, Spain.

[2] Dr. Jospeh Dien's ICA ToolBox download site. http://people.ku.edu/~jdien/. March 12, 2004.

[3] Frank, R., Frishkoff, G., & Glass, K. (2004). NIC Technical Report 04-008: Construction of Blink-free "Baseline" Data to be Used in Evaluating Results of ICA applied to Synthethetic EEG Datasets (TR No. 008). Eugene, Oregon: Neuroinformatics Center.

[4] Frank, R., & Frishkoff, G. (2004). NIC Technical Report 04-002: Generation of Simulated Blink Data (TR No. 002). Eugene, Oregon: Neuroinformatics Center.

[5] Frishkoff, G. (2004). NIC Technical Report 04-003: Procedure for creating a "blink template" to use in identification of blink components after EEG data decomposition (TR No. 003). Eugene, Oregon: Neuroinformatics Center.