NicTR2004-016 09August2004
Team NIC

Summary of Work Toward NIC Journal Paper: "A Framework for Evaluating ICA Methods of Artifact Removal from Multichannel EEG: Reloaded"

Introduction:

The purpose of this summary is to outline the work that has been accomplished so far regarding the journal paper.

This work, or work flow, can be broken into a series of steps that must be completed in essentially the following temporal order:

1. Construction of the baseline (blink-free) EEG data,
2. Generation of the simulated blink-contaminated data sets,
3. Decomposition of the blink-contaminated data using ICA, both FastICA and Infomax,
4. Identification and extraction of the blink activity,
5. Error analysis: Comparison of filtered EEG to baseline and extracted blink activity to synthetically generated blink activity,
6. Actually writing the journal paper,
7. Miller Time!

Step 1. The Baselines:

Conference Paper:

The current baseline is a refinement of the one initially produced for the conference paper, referred to here as the "original" dataset.

This “original” dataset was created by sampling real EEG at 250Hz, using a 0.1Hz highpass and 100Hz lowpass filter. To remove line noise, a 30Hz digital lowpass filter was further applied to the data offline. Before adding the artificially generated blinks, the actual, physiologically generated blinks were manually identified and removed. These resulting “blink-free” segments were then concatenated to form a continuous time series of hopefully blink-free data.

However, as demonstrated in NicTR2004-007, the dataset contained a low-frequency ICA component, possibly reflecting amplifier recovery related to the real (physiologically generated) blinks that had been previously removed. This discovery necessitated the construction of a second, refined baseline, produced from the first by applying a 0.5Hz - 20Hz bandpass filter to further remove any possible traces of activity related to the original, physiological blinks This refined baseline did not contain the low-frequency ICA component (NicTR2004-007), and these two datasets, original and refined, ultimately formed the baselines used in the conference paper.

Journal Paper:

For the journal paper, we went back to the "original" dataset but chose a different technique to remove the suspected amplifier recovery contaminant: ICA.

We first decomposed the dataset using Infomax ICA. We then removed the EEG activity corresponding to the independent component whose spatial projector had the highest correlation to the blink template, a correlation of 0.874, and met the ocular polarity constraint.

To verify that this new potential baseline was free from blink activity, we decomposed it using Infomax ICA and again looked at the correlations of its spatial projectors to the blink template. The highest correlation of any projector to the blink template was now 0.328, while the greatest correlation of any projector that also met the ocular polarity constraint was 0.261.

As shown in NicTR2004-008, the independent component subtracted from the original data by the above process had the characteristic shape of what we believe to be amplifier recovery remaining from the original physiological blinks. This observation and the low blink template correlation of any remaining activity gave us confidence that this new dataset is "clean" of blinks, and so it forms the new journal paper baseline.

It should be noted that both the refined baseline of the conference paper and the new baseline of the journal paper are derived from two different filtering approaches: frequency filtering via a bandpass filter versus statistical filtering via ICA. Since the refined baseline of the conference paper was only used with blink type # 5, a future comparison of the two baselines, across all the blink types, may be worthwhile.

<u>Step 2. Generation of Simulated Blinks:</u>

The generation of the simulated blink datasets is detailed in NicTR2004-002, the details of which will be incorporated in part. In addition, I have modified the original blink-generation code to produce an array of pointers to the peaks of the simulated blinks, as well as a cell array of pointers to segments time-locked to the peaks. These arrays are now used by APECS to automatically segment and average the EEG data about the blinks, allowing APECS to generate blink event related potentials for subsequent analysis, just like NetStation, only not as pretty.

Step 3. ICA Decomposition (APECS) / Step 4. Blink Identification & Extraction:

The ICA decomposition process, along with the identification and extraction of blink related activity, is now completely automated within APECS. NicTR2004-004 is a fairly detailed account of the APECS structure, and will be incorporated in part into the journal paper. A mathematical description of ICA, in terms of orthogonal rotations (NicTR2004-012) will also be incorporated here.

Step 5. Error Analysis:

To compare the effectiveness of the different ICA algorithms blink removal abilities on blinks of varying morphology, we have the following metrics:

1. ***B****link* ***E****vent* ***R****elated* ***P****otential****s****:* BERPS compare the filtered EEG to the baseline, by segmenting the data on the blink peaks and then averaging across the segments.

(NetStation & Conference Paper Baseline, from Conference Paper)

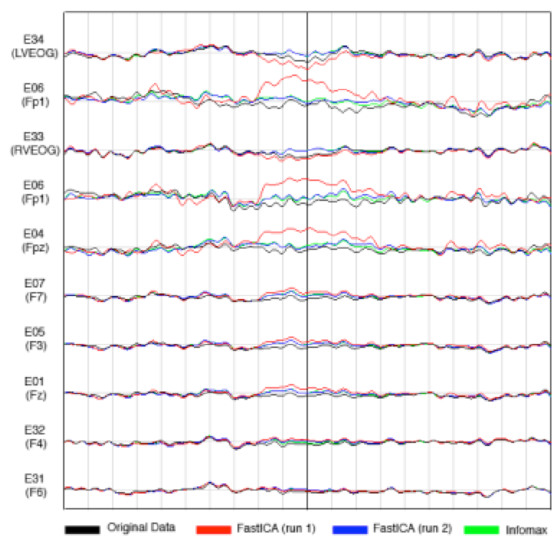


Fig. 6. EEG waveforms, averaged to the peak of the blink activity. Note residual blinks in run 1 for FastICA, where more than one source was strongly correlated with the blink template, and the source activations revealed misallocation of variance (cf. Fig 5).

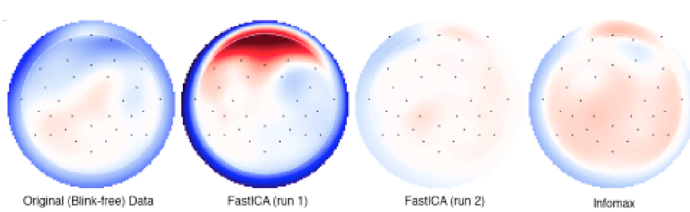
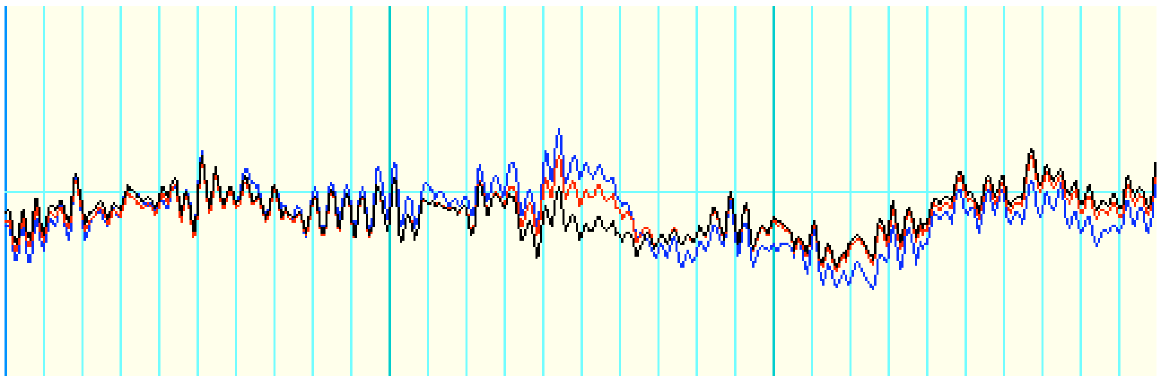


Fig. 7. Topography of blink-averaged data, centered at peak of blink activity. Red, positive voltage. Blue, negative voltage. FastICA run1 is the less successful decomposition. Note the remaining blink activity at this time point.

(NetStation & Journal Paper Baseline from Gwen)

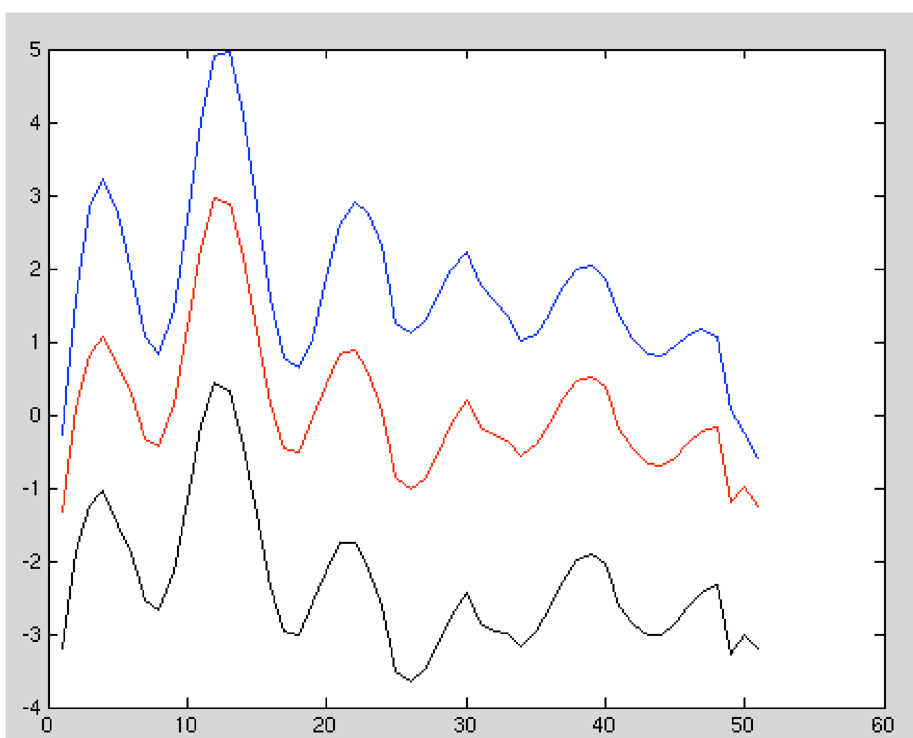
Baseline EEG, Infomax Filtered EEG and FastICA-Run #1 Filtered EEG
(Channel # 6 | Simulated Blink Type # 5)



Horizontal Scale: 3 seconds | Vertical Scale: 0.5 microvolts / mm
Blink-Template Tolerance: 0.95 | #IC Extracted: 1(FastICA) / 1(Infomax)

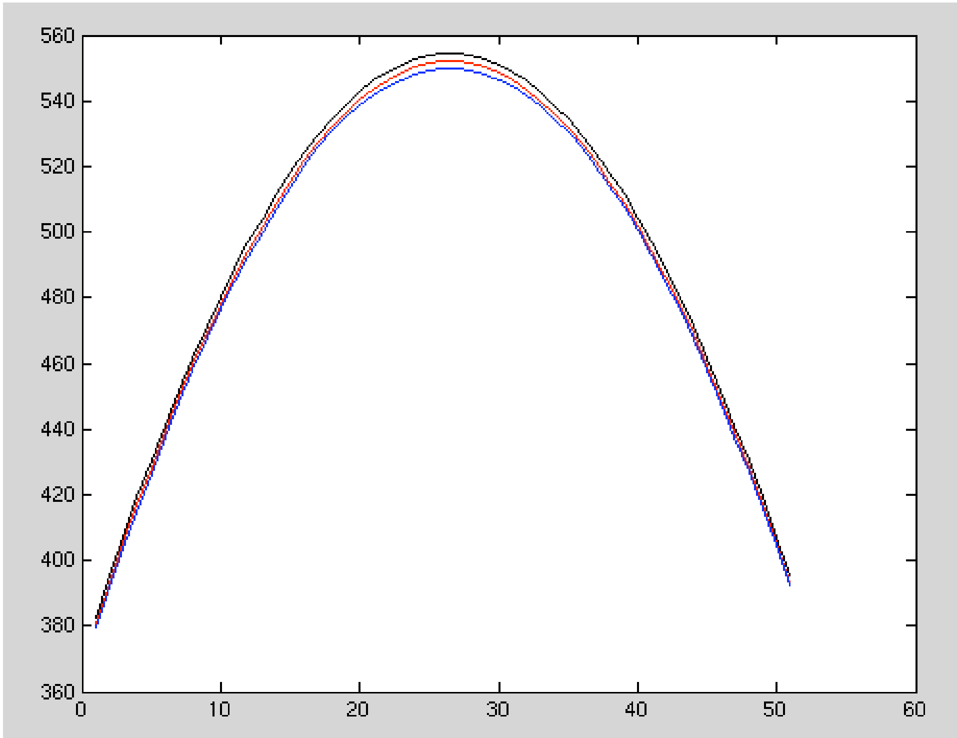
(MATLAB & Journal Paper Baseline from APECS)

Baseline EEG, Infomax Filtered EEG and FastICA-Run #1 Filtered EEG
(Channel # 6 | Simulated Blink Type # 5)



Horizontal Scale: Samples (sample # 25 = blink peak) | Vertical Scale: Microvolts
Blink-Template Tolerance: 0.95 | #IC Extracted: 1(FastICA) / 1(Infomax)

Projected Blink, Infomax Extracted Blink and FastICA-Run # 1 Extracted Blink
(Channel # 6 | Simulated Blink Type # 5)



Horizontal Scale: Samples (sample # 25 = blink peak) | Vertical Scale: Microvolts
Blink-Template Tolerance: 0.95 | #IC Extracted: 1(FastICA) / 1(Infomax)

2. *Spreadsheets:* Eighteen Excel spreadsheets quantify the error between the baseline EEG and its filtered approximations for Infomax and each of the 3 FastICA runs for tolerances 0.85, 0.90 and 0.95.

   - *AbsBlnkErrorBI_FastIcaR#:*
     Absolute error between projected blinks and those extracted via Infomax and FastICA Run #, only on the intervals to which the blinks were added (BI).

   - *RltvFltrdError_FastIcaR#:*
     Relative error between baseline EEG and filtered approximations via Infomax and FastICA Run # over the entire data set, both blink and blink-free intervals.

   - *RltvFltrdErrorBI_FastIcaR#:*
     Relative error between baseline EEG and filtered approximations via Infomax and FastICA Run #, only on the intervals to which the blinks were added (BI).

- *RltvFltrdErrorBFI_FastIcaR#:*
  Relative error between baseline EEG and filtered approximations via Infomax and FastICA Run #, only on the intervals to which the blinks were not added (BFI).

- *SegAvgAbsFltrdErr_FastIcaR#:*
  Both the baseline and ICA filtered EEG, via Infomax and FastICA Run #, were first segmented on the blinks and then averaged across the segments. An absolute error was then computed between corresponding sample points of the averaged segments.

- *SegAvgAbsBlnksErr_FastIcaR#:*
  Both the projected blinks and those extracted via Infomax and FastICA Run # were first segmented on the blinks and then averaged across the segments. An absolute error was then computed between corresponding sample points of the averaged segments.

Absolute Error Values Between (Averaged) Actual and (Averaged) Filtered EEG for Channels 2 - 6 of Dataset 5 and Corresponding Correlations

Tolerance: 0.95									
Fast ICA (TanH)					InfoMax				
MaxError	MeanError	MedError	StdError	Correlation	MaxError	MeanError	MedError	StdError	Correlation
1.8838	1.5825	1.5789	0.1886	0.8850	1.6560	1.4910	1.5415	0.1475	0.9315
4.0925	3.3228	3.2796	0.4593	0.8610	1.9295	1.7281	1.7698	0.1536	0.9812
0.9707	0.8155	0.8137	0.0971	0.9716	1.0042	0.9037	0.9336	0.0902	0.9759
3.2363	2.5917	2.5317	0.3827	0.7842	1.3744	1.2282	1.2519	0.1051	0.9811
1.6399	1.3746	1.3742	0.1652	0.9279	1.5450	1.3909	1.4385	0.1379	0.9504
4.9341	4.0343	4.0025	0.5397	0.8921	2.6845	2.4084	2.4816	0.2207	0.9731

Absolute Error Between (Averaged) Actual and (Averaged) Extracted Blinks for Channels 2 - 6 of Dataset 5 and Corresponding Correlations

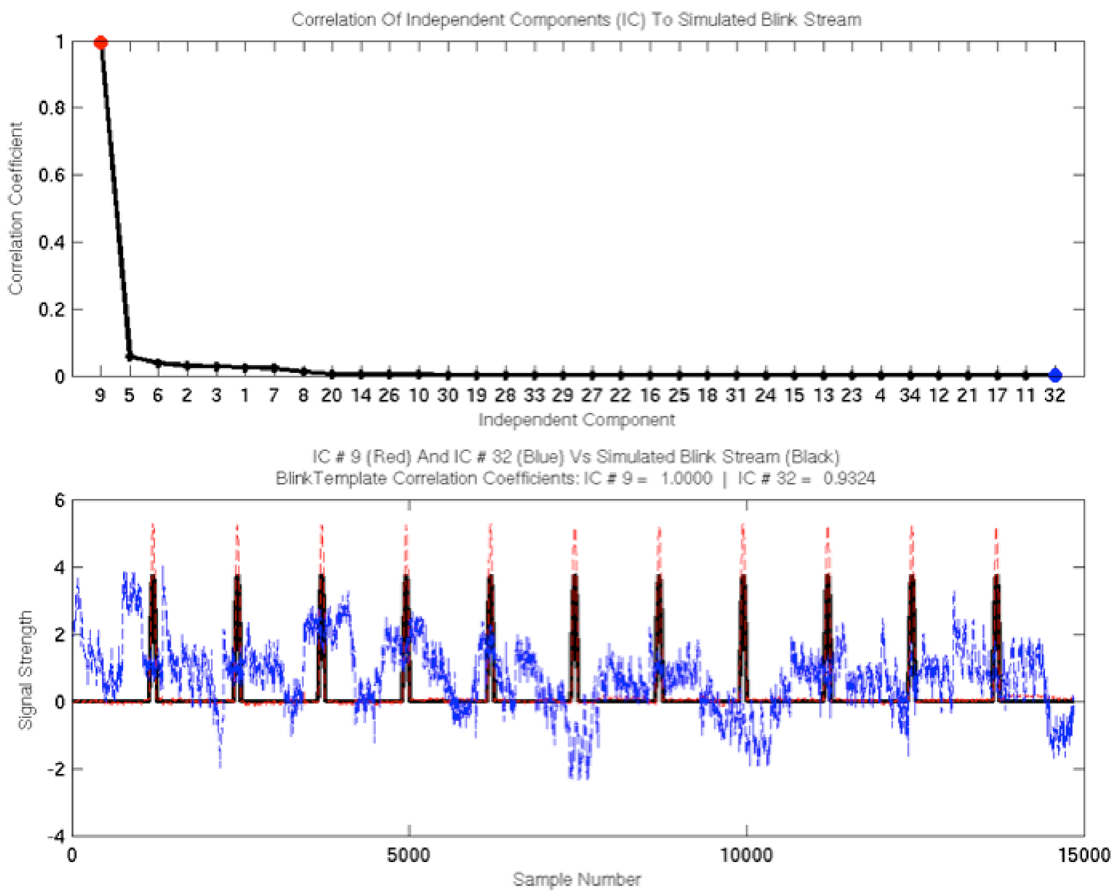
Tolerance: 0.95									
Fast ICA (TanH)					InfoMax				
MaxError	MeanError	MedError	StdError	Correlation	MaxError	MeanError	MedError	StdError	Correlation
1.8838	1.5825	1.5789	0.1886	1.0000	1.6560	1.4910	1.5415	0.1475	1.0000
4.0925	3.3228	3.2796	0.4593	1.0000	1.9295	1.7281	1.7698	0.1536	1.0000
0.9707	0.8155	0.8137	0.0971	1.0000	1.0042	0.9037	0.9336	0.0902	1.0000
3.2363	2.5917	2.5317	0.3827	1.0000	1.3744	1.2282	1.2519	0.1051	1.0000
1.6399	1.3746	1.3742	0.1652	1.0000	1.5450	1.3909	1.4385	0.1379	1.0000
4.9341	4.0343	4.0025	0.5397	1.0000	2.6845	2.4084	2.4816	0.2207	1.0000

The maximums of the absolute errors, computed over the 50 sample points that comprise the averaged blinks, are 4.09 for channel 2 and 4.93 for channel 6 (FastICA) and 1.93 and 2.68, respectively, for Infomax. These values reflect the separation between the curves in the averaged segments. The correlation in all four cases, however, is 1.0000, which does not capture this error.

Notes: The production of the BERPs and error analysis spreadsheets is now fully automated in MATLAB 7.0, and can be incorporated into the APECS framework for post-ICA analysis.

The spreadsheet and BERP data show that Infomax was superior overall to Fast ICA across the blink types, while blink types # 5 and # 7, which have the greatest amplitudes, gave both Infomax and FastICA the most trouble. These results are consistent with the conference paper.

3. *Temporal Correlation of the Independent Components & Harmonic Analysis:*

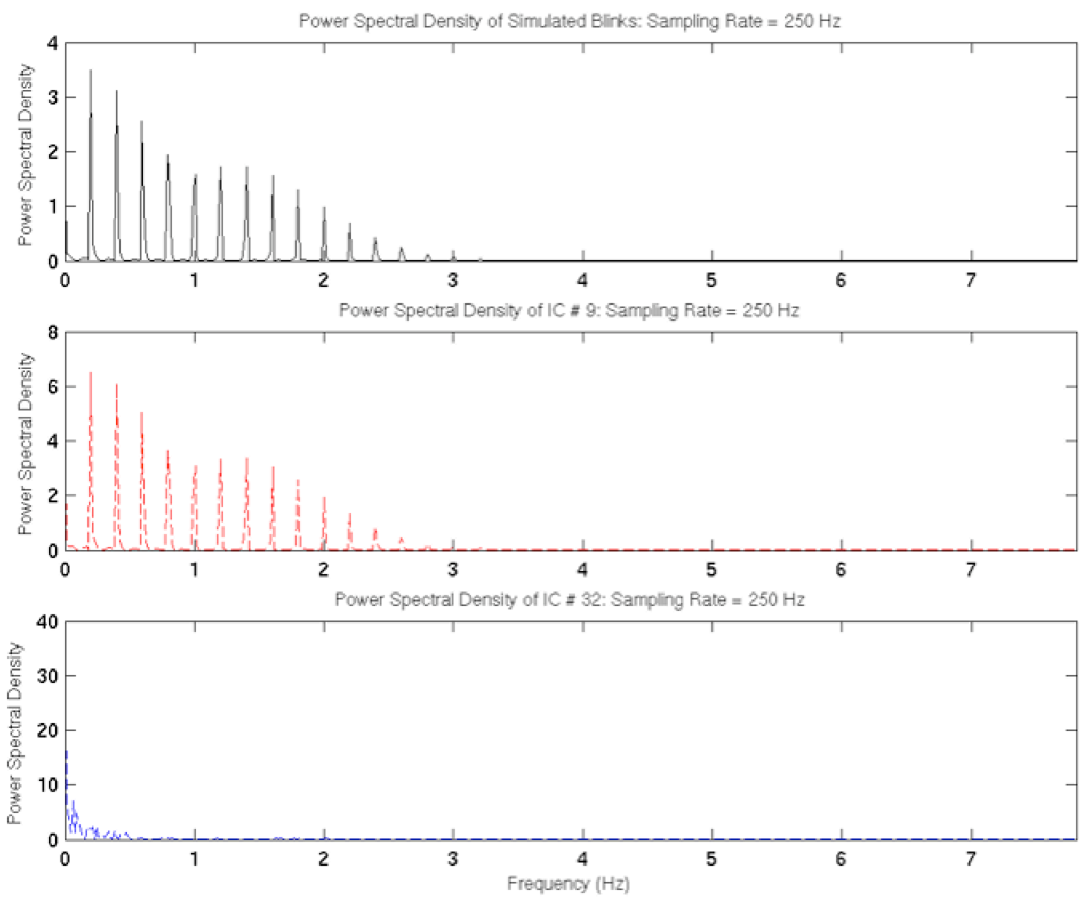


Blink Type # 5, FastICA Run # 1

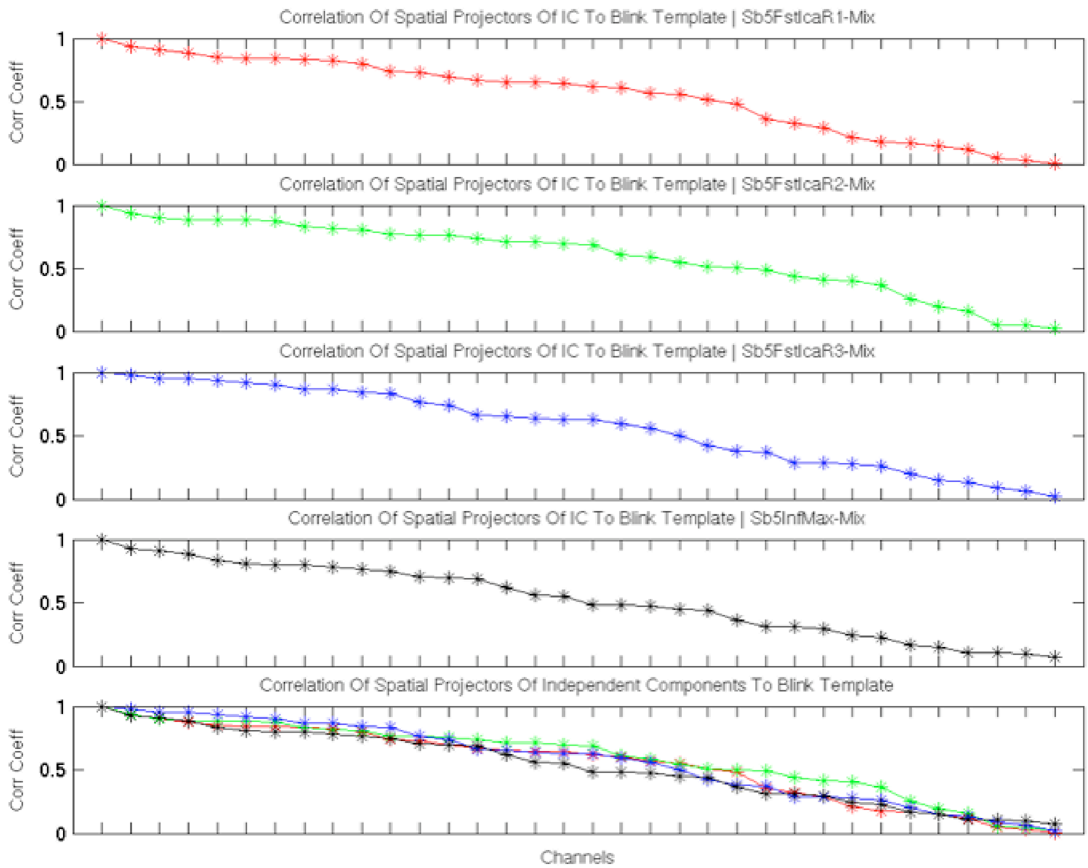
The correlations of each IC to the simulated blink stream are computed and plotted in the top figure. Ideally, 1 IC should correlate to the blink stream at 1.0000, while all the others should be 0.0000. The fact that the plots show this is not actually the case is a nice way of illustrating that we are dealing with numerical approximations to an ideal.

In the bottom figure, the temporal evolutions of the 2 IC with the highest correlations to the blink template, and meeting the ocular polarity constraints, are superimposed over the corresponding temporal evolution of the simulated blink stream. This serves to illustrate that the 1 IC whose corresponding blink template correlation is 1.0000 does indeed capture the majority of the simulated blink activity, while the other IC, which also correlates strongly to the blink template, possibly represents residual activity from the original physiological blinks or cortical activity with a blinky spatial distribution.

Like the BERPs and spreadsheets, the temporal correlation plots are generated automatically, via a MATLAB m-file, and can also be incorporated into APECS. Below are the power spectral densities corresponding to these ICs. Note the close match between the blink stream and IC # 9, while the differences between the blink stream and IC # 32 reinforce the point that blink template correlation and ocular polarity constraint alone are not necessarily sufficient to identify IC representing phasic blink activity.



4. *Correlation of Spatial Projectors of IC to the Blink Template:* These graphs show that several projectors, and thus their IC, can correlate relatively strongly to the blink template, which reinforces the need to consider both the ocular polarity and temporal evolution / harmonic decomposition of the spatial projectors and IC, respectively, when identifying blink activity.



The bottom graph also illustrates that both Fast ICA, from run to run, and FastICA vs. Infomax, produce results which are very similar. This, in my mind, reinforces the idea that they are all seeking the one unique rotation, represented by the inverse of the mixing matrix, that gives truly independent components.

<u>Step 6. The Paper:</u>

- Outline

- Scope

- Editorial / Review Procedures

- To Do

i. Write MATLAB script to compute mean projected variance of IC and then order IC in terms of decreasing mean projected variance for FastICA.
ii. Write MATLAB script to compute kurtosis / negentropy of IC to produce a relative measure of the "independence" of the independent components.
iii. Other:

iv. Write the paper, write the paper, write the paper...