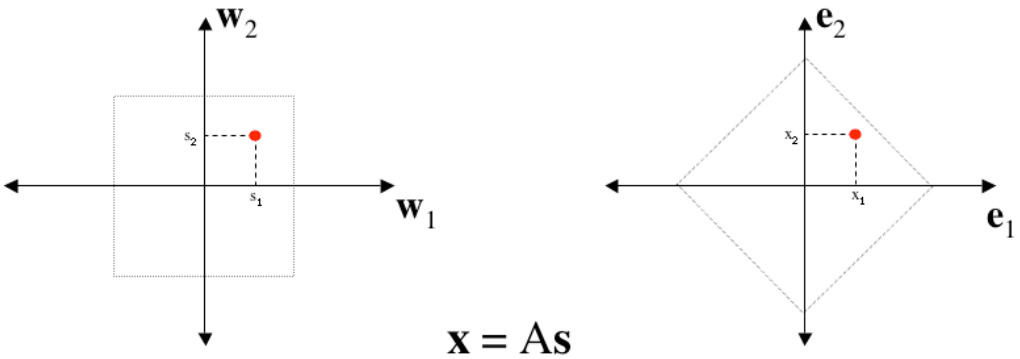
Independent Component Analysis (ICA):

Independent component analysis (ICA) is a mathematical technique[1] that combines aspects of both linear algebra and statistics to represent EEG data as a linear mixture of a set of statistically independent streams, called the independent components (IC).

A consequence of the central limit theorem in statistics states that a linear mixture of a set of statistically independent random variables has a distribution that is more gaussian than any of the random variables which comprise the sum. So, in ICA theory, maximizing non-gaussianity, as measured by the data's higher order moments, is equivalent to separating out the independent random variables (the IC) from their linear mixture (the EEG data).

ICA achieves this separation by computing a new set of basis vectors for the data, such that the components of the data's coordinate vectors, with respect to this new basis, have a distribution that is maximally non-gaussian. Essentially, ICA is an orthogonal rotation applied to the data, whereby the nature of the rotation, its yaw, pitch and roll, is determined by the application of statistics.

Figure 1. ICA as a Change of Basis



Distribution of probability is more Gaussian for coordinate vector (x_1, x_2) than (s_1, s_2).

(x_1, x_2) is coordinate vector for basis $\{\mathbf{e}_1, \mathbf{e}_2\}$: $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$
(s_1, s_2) is coordinate vector for basis $\{\mathbf{w}_1, \mathbf{w}_2\}$: $\mathbf{x} = s_1\mathbf{w}_1 + s_2\mathbf{w}_2$

In ICA theory, the data is often assumed to be whitened by the prior application of a sphering matrix, which is derived from either the SVD of the

data or the eigenvector decomposition of its correlation matrix. The whitened data have zero mean, unity variance and are linearly decorrelated, and the subsequent rotation of ICA transforms this linear decorrelation into statistical independence. In this view, ICA can be seen as a variation of basic PCA, which seeks only to linearly decorrelate the data.

Application of ICA to Blink Removal:

Eye blinks, which occur naturally during the course of EEG data collection, are a source of serious contamination to the recorded EEG and so must be cleanly removed. Previously, eye blinks were often dealt with by instructing the subject to blink only during certain times, or by simply removing from the data all the blink contaminated segments. However, instructing the subject on when to blink is not always practical and can impose an additional constraint on the experiment which skews the result, whereas removing blink-contaminated segments may not leave sufficient blink-free data with which to draw valid conclusions.

With respect to blink removal, eye blinks ideally represent a stream of activity statistically independent from other cerebral activity and so amenable to capture by a single independent component using ICA. For this reason, the NIC is investigating the effectiveness of different ICA algorithms for blink removal and is developing high-speed implementations of these algorithms capable of running on multi-processor systems.

To systematically study ICA requires a framework that can receive pre-processed (cleaned of bad channels, saturated observations, filtered to remove line noise, etc...) EEG data in EGI .raw format, that can apply ICA, and that can extract activity which appears blink like, as specified by the user. Furthermore, it necessitates the creation of a blink-free baseline, rich in cortical activity, to which simulated blinks can be added and subsequently removed via ICA. By knowing a priori the nature of the blinks and the baseline, we can effectively determine the extent to which the various ICA algorithms are capable of cleaning the data, i.e. removing blinks. To achieve this end, we have developed APECS as part of a testing framework.

Generation Of Simulated Blink Data:

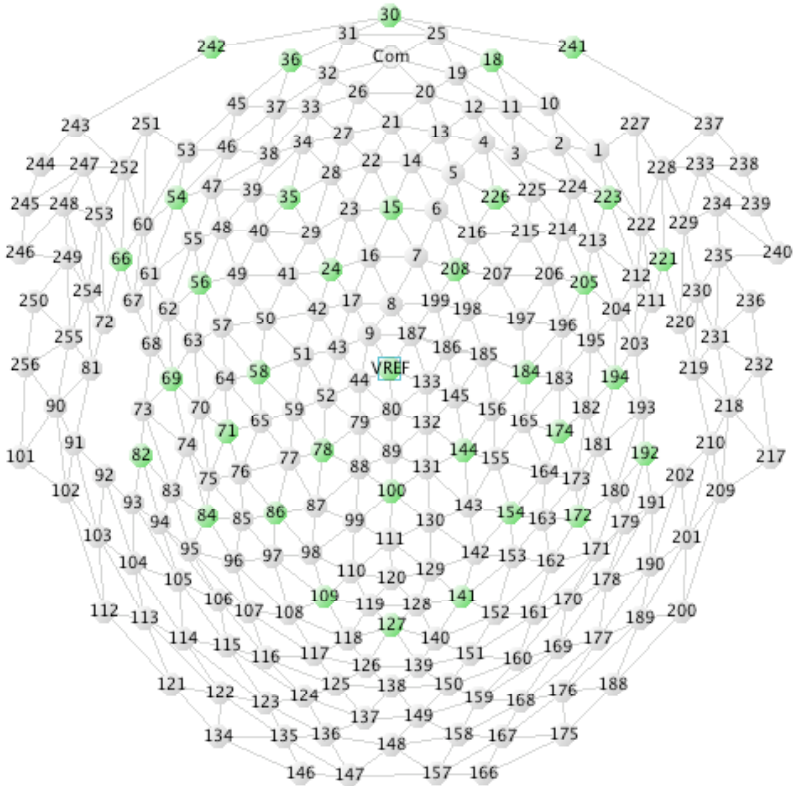
The objective here is to create realistic sets of blink-contaminated EEG data which can then be used to test the efficacy of ICA for blink removal with APECS or the signal cleaning toolbox.

Note: The results of some our early investigations have already been published in a conference paper[6] and will shortly be expanded upon in a subsequent journal paper now being written.

1. EEG data acquisition and preprocessing

EEG data were acquired from 256 scalp electrodes using a Geodesic Sensor Net, with vertex recording reference, at a sampling rate of 250 Hz and bandpass filtering from 0.1 to 100 Hz. To remove high frequency line noise, the data were further low pass filtered. All trials with blinks were then manually detected and eliminated, creating a "blink-free" 256 channel data set. Since the data contained more channels than were necessary for the purpose of measuring blink removal effectiveness, the 256-channel montage was down-sampled to the 34 highlighted channels shown below.

Figure 2. Geodesic Sensor Net, Down-Sampled Channels Highlighted



A variation of this data set was created by applying additional bandpass filtering from 0.5 Hz to 20 Hz to eliminate the possible presence of amplifier recovery related to the original physiological blinks and not completely eliminated through the deletion of blink contaminated trials.

2. Blink Generation

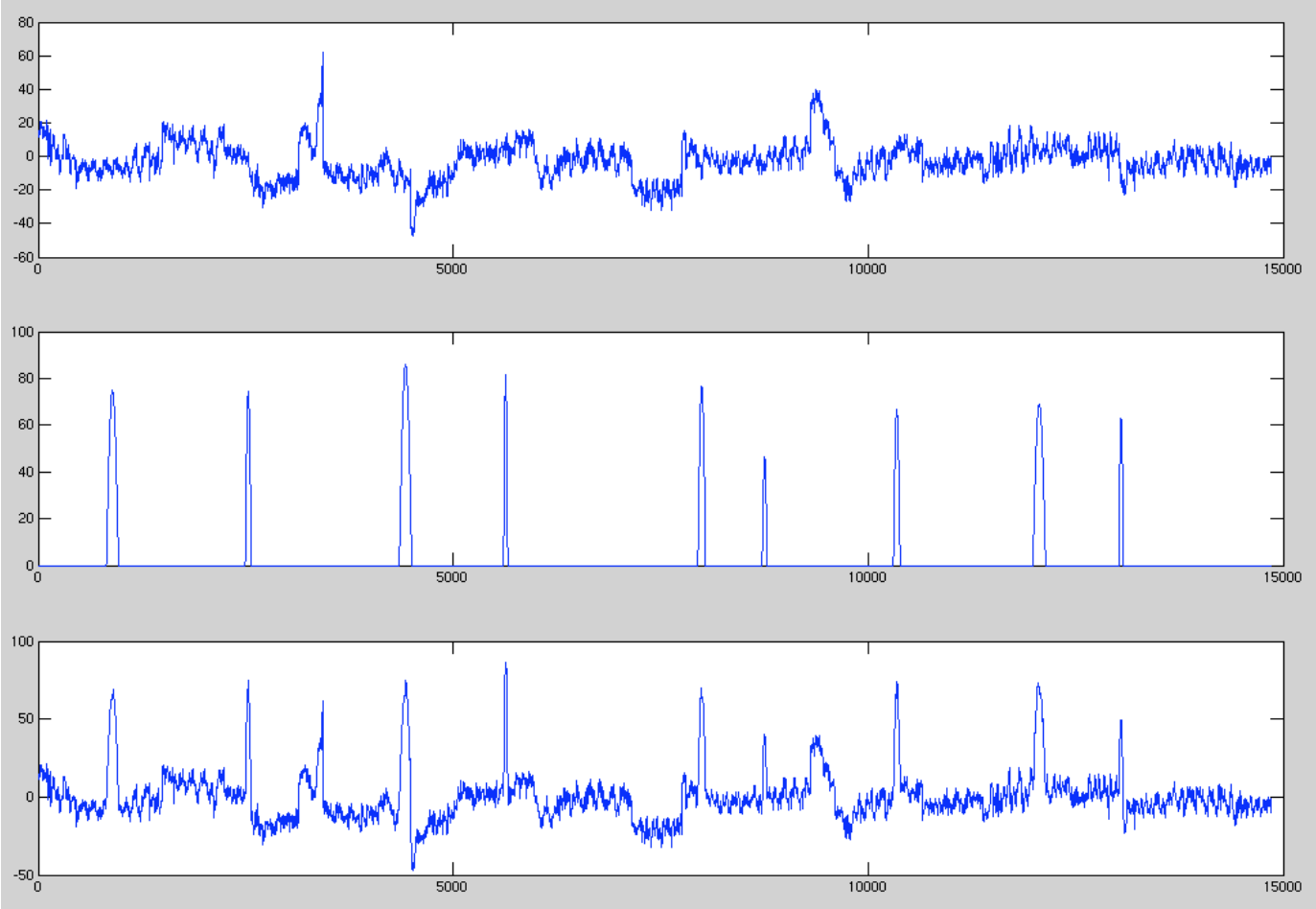
Blink activity was simulated by first generating a single positive sinusoidal pulse, $y = \sin(x)$, $0^{o} <= x <= 180^{o}$. A single time course of blink activations was then constructed by concatenating together this sequence of positive sinusoidal pulses. Pulse duration and inter-pulse spacing were both specified by the user, and could be set to either a fixed or varying integer number of data sample points, where each sample point corresponded to a 0.004 second time interval. The height of each pulse could also be set to a fixed or varying quantity to simulate blinks of different intensities. By varying these parameters, in part or in whole, we created a stream of blink activity in the independent component space.

This time course of blink activations, represented in MATLAB by a row vector of size 1 x NumberOfSamples, then right multiplied the blink template, a NumberOfChannels x 1 column vector, thus projecting the blink activations onto the detectors. The result is a matrix of size NumberOfChannels x NumberOfSamples, in which each row represents the entire time course of blink activity at its corresponding detector, and each column represents one instant of blink activity at all the detectors. This matrix was then added to the array of "blink-free" EEG data, essentially superimposing the totality of blink activity onto the blink-free EEG data to construct the array of blink contaminated EEG data.

$$x_{Blinky} = x_{BlinkFree} + Template_{Noise} * BlinkActivation$$

(Continued Next Page)

Figure 3. Blink Type # 6, EEG Channel # 1 (15000 samples ~ 60 sec)



From Top to Bottom:

- Clean, Blink-Free, Data
- Blink Stream
- Clean, Blink-Free Data + Blink Template * Blink Stream

The overriding objective in creating these data sets was to develop a suite of data with which we could test the ability of different ICA algorithms to remove blinks from EEG recordings. To this end, we first constructed the "blink-free" EEG data set, rich in cortical activity yet free of blink contamination. This was the baseline or gold-standard to which filtered EEG, after blink removal, would be subsequently compared. We then constructed seven distinct sets of blinks, which were then added to the "blink-free" EEG, resulting in seven sets of "blinky" EEG, each with a unique blink morphology. Now, in subsequent studies, some or all of the following comparisons can be made:

- Compare filtered data to the "blink-free" data;
- Compare extracted blinks to the simulated, inserted blinks;
- Compare corresponding projections of extracted blink activations to the blink template.

Since we know not only the structure of the simulated blinks but also at which sample points they were inserted, detailed comparisons can and have been made employing a variety of metrics.

APECS and The Signal Cleaning Toolbox:

APECS, a component of the more general signal cleaning toolbox, is a collection of MATLAB m-files, originally based upon Dr. Joseph Dien's ICA Toolbox[2,3], designed to remove eye blinks from EEG data using independent component analysis (ICA). APECS is also implemented as a stand alone MATLAB application, outside the toolbox, and can be run with or without a graphical user interface on a single or multi-processor system.

In ICA, as viewed from a matrix algebra perspective, EEG data (x) is represented by a linear mixture of a set of statistically independent streams (s), called the temporal activations of the independent components, or independent components for short. Mathematically,

$$x = As$$

where:

- A = m by m mixing matrix
- s = m by n matrix of independent components, one per row
- x = m by n matrix of EEG data, where each row corresponds to the output of a single detector and each column represents one time sample from all the detectors

The i^{th} column vector of the matrix A projects the temporal activation of the i^{th} independent component onto the array of EEG detectors and thus determines the spatial contribution of that component to each detector. Mathematically speaking, the spatial-temporal contribution of the i^{th} component to the EEG data is the outer product of the i^{th} column of A with the ith row of s:

$$x_{IC} = A_i * s_i.$$

The eye blink activity at the detectors ($x_{EyeBlink}$) is then computed as the outer product of its independent component(s) ($s_{EyeBlink}$) with its corresponding spatial projector(s) ($A_{EyeBlink}$)

$$x_{EyeBlink} = A_{EyeBlink} * s_{EyeBlink},$$

and subsequently removed from the EEG data by a matrix-matrix subtraction to yield blink-free EEG data ready for further analysis

$$x_{BlinkFree} = x_{Original} - x_{EyeBlink}.$$

At present, APECS can use one of three algorithms to ICA decompose the data: FastICA[4], with a hyperbolic tangent contrast function, Infomax[5], via

either MATLAB or its faster C implementation[5] and Par-FastICA, the NIC's parallel implementation of FastICA.

To remove blink activity from the EEG data, the relative polarity of the elements representing the EOG detectors in each column of the mixing matrix A is checked to ascertain if it corresponds to either horizontal or vertical eye movements. A normalized covariance may also be computed between each column of matrix A and the blink template. Then, in accordance with user specifications, all of those columns that meet the VEOG polarity requirement, correlate sufficiently to the blink template, or both, are flagged as describing a spatial topography at the detectors consistent with blink activity.

As stated previously, each column vector of the mixing matrix A determines the spatial distribution of its corresponding independent component, and can be thought of as the spatial projector for that component. Since we now know which column vectors of matrix A correspond to blink activity, we therefore know that their corresponding independent components contain the temporal activations of those blinks. The blink activity at all detectors for all time is then computed as the outer product of these spatial projectors (m x 1 column vectors, m = # of detectors) with their independent components (1 x n row vectors, n = # of samples), resulting in an m x n matrix of blinks, which is subtracted from the original m x n EEG data matrix, to give an m x n blink-free EEG data matrix:

$$EEG_{BlinkFree} = EEG_{Original} - EEG_{PureBlinks}.$$

APECS writes this data to disk in the following EGI raw format files:

- Mixing Matrix A: <FileName_Mix>
- Independent Components s: <FileName_Ica>
- Blink-Free EEG Data: <FileName_Fltrd>
- Extracted Blinks: <FileName_Blnks>

Note: FileName is the name of the original raw format EEG data file, sans extension.

If the user specified more than one blink template correlation, APECS writes one blink-free EEG data file and one set of extracted blinks for each of the tolerances.

These raw files can then be read back into NetStation for post-processing, segmentation, and averaging to aid the analysis of the blink removal procedure.

References:

- [1] Hyvarinen, A. and E. Oja. (1999). "Independent Component Analysis: A Tutorial." http://www.cs.helsinki.fi/u/ahyvarin/papers/index.html. March 12, 2004.

- [2] ICAToolBox download site. http://people.ku.edu/~jdien/. March 12, 2004.

- [3] Dien, J. (1998). Issues in the application of the average referance. Review, critiques, and recommendations. Behave. Res. Methods, Instruments, and Computers, 30(1), 34 - 43.

- [4] FastICA download site. http://www.cis.hut.fi/projects/ica/fastica/. March 12, 2004.

- [5] EEGLab download site. http://www.sccn.ucsd.edu/eeglab/downloadtoolbox.html. March12, 2004.

- [6] Kevin A. Glass, Gwen A. Frishkoff, Robert M. Frank, Colin Davey, Joseph Dien, Allen D. Malony, Don M. Tucker[1,2] (2004) " A Framework for Evaluating ICA Methods of Artifact Removal from Multichannel EEG"

- Team NIC. (2004). NIC Technical Reports: NicTR2004-001 through NicTR2004-010.