

# **ERP PCA Toolkit 2.23 Tutorial**

Joseph Dien  
Center for Advanced Study of Language  
University of Maryland

Thanks to Tim Curran for being a patient, long-suffering beta-tester, to Scott Miller for his extensive, helpful editing of the Tutorial document, to Dennis Molfese for his unstinting support for the development of the 2.00 version of the Toolkit, and Don Tucker for first suggesting to me that it might make some sort of mad sense to visualize the ERP PCAs by multiplying the factor loadings by the factor scores.

# Contents

<b>ABOUT THE ERP PCA (EP) TOOLKIT</b>	<b>4</b>
USES	4
CITATION	5
WEBSITE	5
MAILING LIST	5
SUGGESTIONS AND BUG REPORTS	5
SYSTEM REQUIREMENTS	5
SUPPORTED FILE TYPES	5
CREDITS	6
EXAMPLE DATASET	8
<b>SET UP</b>	<b>9</b>
INSTALLATION	9
1. INSTALL MATLAB	9
2. INSTALL EP TOOLKIT	9
3. INSTALL EEGLAB	10
4. INSTALL FIELDTRIP	11
5. INSTALL SATIMAGE OSAX	12
AVOIDING COMMON PROBLEMS	13
MONITOR RESOLUTION	13
MEMORY ISSUES	13
ENERGY SAVER SETTINGS	15
FILE PREPARATION	15
EP TOOLKIT FILE TYPES	15
ELECTRODE COORDINATES	16
REFERENCE CHANNELS	17
NET STATION FILES	17
OTHER FORMATS	18
<b>RUNNING THE TOOLKIT</b>	<b>19</b>
STARTING UP	19
PREFERENCE FILE & EPWORK DIRECTORY	21
RESTARTING	22
QUITTING	22
<b>TOOLKIT FUNCTIONS</b>	<b>23</b>
ARTIFACT CORRECTION	23
CHUNKING	23
DEFAULT PROCEDURES	23
SETTING PREFERENCES	25
MANUALLY FORMING A BLINK TEMPLATE	29
AVERAGING	42
POSTPROCESSING	44

<b>READ</b>	<b>45</b>
<b>EDIT</b>	<b>46</b>
<b>VIEW DATA</b>	<b>52</b>
<b>PRINCIPAL COMPONENTS ANALYSIS (PCA)</b>	<b>55</b>
<b>WINDOWING</b>	<b>66</b>
WINDOWING SUBJECT AVERAGE DATA	66
WINDOWING PCA DATA	74
<b>ROBUST ANOVA</b>	<b>75</b>
ABOUT ROBUST STATISTICS	75
GENERAL PROCEDURES	76
BEHAVIORAL DATA	85
<b>SAVE FILES</b>	<b>85</b>
IMPORTING TO EEGLAB	86
IMPORTING TO NET STATION	87
<b>REFERENCES</b>	<b>88</b>
 <b>APPENDICES</b>	 <b>90</b>
<b>EP TOOLKIT FILE FORMAT</b>	<b>90</b>
<b>PUBLICATIONS USING THE EP TOOLKIT</b>	<b>92</b>
<b>BIBLIOGRAPHY</b>	<b>94</b>

# About the ERP PCA (EP) Toolkit

## Uses

The EP Toolkit is a collection of software tools for enhancing analysis of ERP data. It runs in MATLAB and is meant to be an adjunct to the various ERP software suites currently used by researchers. The Toolkit contains functions for improved artifact correction, trial averaging, file editing, data viewing, factor analysis, and inferential analysis. It has three major goals:

- 1) Facilitating multivariate decomposition of ERP data in order to better characterize their components. This is done by:
  - a) providing automated routines for conducting principal components analysis (PCA), including the two-step sequential PCA procedure.
  - b) providing a range of rotation options, including facilitated access to EEGLAB's independent components analysis (ICA) Infomax rotation.
  - c) translating PCA results back into voltage space so that they can be transparently evaluated as ERP waveforms and exported into various file formats for additional analysis.
- 2) Facilitating the use of robust statistics for descriptive and inferential analyses. This is accomplished by:
  - a) allowing averaging (including trial averaging) using trimmed means that are more robust against outliers than conventional means.
  - b) implementation of robust ANOVA statistics that are more resistant to noisy data than conventional ANOVA statistics.
  - c) providing automated implementation of vector tests for electrode effects and follow-up ANOVAs of interaction effects.
  - d) allowing application of these statistics to traditional ERP measures, the results of multivariate decomposition, and non-ERP data, including behavioral data.
  - e) providing direct access to data transformations, including various combinations of cells generated by ANOVAs, regional electrode averages, and trimmed means, all of which can be viewed as waveforms and exported for further analysis.
- 3) Facilitating analysis of noisy data, including data from developmental and clinical research populations. This is made possible by:
  - a) multivariate techniques for automated removal of blink and motion artifacts from session data using a combination of ICA and PCA.
  - b) use of trimmed means for trial averages and robust ANOVAs for inferential analyses, as outlined above.
  - c) provision of quality control (QC) measures that allow for easier detection of problematic data files.

The EP Toolkit is freely available under the GNU General Public License version 3. It is intended to be a research tool only and must not be used in clinical practice.

## Citation

When used for a publication, please provide credit in the methods section and let me know that you have used it. I am hoping to eventually get some grant funding to support continued development of this toolbox. The appropriate citation is:

Dien, J. (2010). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods* 187(1), 138-145.

## Website

You can always download the latest version of the EP Toolkit from the following address: <http://sourceforge.net/projects/erppcatoolkit/>

## Mailing List

If you wish to be informed of new releases, please be sure to sign up for the mailing list: <https://lists.sourceforge.net/lists/listinfo/erppcatoolkit-support>

## Suggestions and Bug Reports

Please feel free to send me suggestions or bug reports at: [jdien07@mac.com](mailto:jdien07@mac.com).

## System Requirements

This package has been tested for use on a PowerPC Macintosh under OS X 10.4.11 using Version 7.3 (2006b) of MATLAB and an Intel Macintosh under OS X 10.6.6 using version 7.11 (2010b) of MATLAB. It may not work on earlier versions of MATLAB and I cannot afford the effort to support them. The toolbox has also been used somewhat in Windows XP with MATLAB 7.8 (2009a) but has not been fully tested so caveat emptor. Note that MATLAB versions 2006 and 2007 do not support the *Table* element in the user interface and so the Toolkit will substitute cruder looking controls. I would be happy to help users troubleshoot platform incompatibilities if they are willing to work through them with me. Likewise, I mostly work with EGI files and so the Toolkit has mostly been tested with them, although it has received some use with other file formats. I would be happy to work with users to better support the existing list of file formats or to add new ones. The Toolkit functions also need significant amounts of RAM. I would suggest a minimum of 2GB and preferably at least 4GB, depending on the size of the datafiles. Also, the faster the computer the better as the algorithms it implements are quite computationally intensive.

## Supported File Types

The EP Toolkit currently reads EGI Simple Binary (EGI .raw), EGI EGIS (BESA .raw), EGI Matlab, EEGLAB set, Neuroscan average, text, and EP Toolkit format files. There

is an attempt to support reading .EDF files but it needs to be tested and an initial attempt failed. I would be happy to complete this implementation if someone with access to .EDF files could help me work out the remaining issues. It currently can write out EGI Simple Binary (EGI .raw), EGI EGIS (BESA .raw), EEGLAB set, text, and EP Toolkit format files.

Note that NetStation (as of 4.1.2) loses the trial-specific information when writing out continuous (non-segmented) simple binary files. It also does not seem to read in event data at all from such files.

## Credits

Thanks to Dennis Molfese for helping support the development of the 2.0 version of the Toolkit.

Also, credit should be given to the algorithms used for the rotations:

### VMAX - Varimax

Kaiser, H. F. (1959). Computer program for varimax rotation in factor analysis. *Educational and Psychological Measurement*, 19(3), 413-420.

### PMAX - Promax

Hendrickson, A. E. & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *The British Journal of Statistical Psychology*, 17, 65-70.

### IMAX - ICA Infomax

Delorme, A. & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134, 9-21.

Gradient Projection method used for the following rotations:

QMAX - Quartimax

QMIN - Quartimin

OMIN - Oblimin (gamma=0 tends to be recommended)

CRFE - Crawford-Ferguson family

MINE - minimum entropy

IPSC - Bentler's invariant pattern simplicity criterion

TIIC - Comrey's tandem II criterion

GMIN - Geomin

MMER - McCammon minimum entropy ratio

Jennrich, R. I. (2001). A simple general procedure for orthogonal rotation. *Psychometrika*, 66, 289-306.

Jennrich, R. I. (2002). A simple general method for oblique rotation. *Psychometrika*, 67(1), 7-20.

Also, credit for the rotational criteria themselves:

#### VMAX - Varimax

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.

#### PMAX - Promax

Hendrickson, A. E. & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *The British Journal of Statistical Psychology*, 17, 65-70.

#### IMAX - ICA Infomax

Bell, A. J. & Sejnowski, T. J. (1995). An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7(6), 1129-1159.

#### QMAX - Quartimax

Carroll, J. B. (1953). An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, 18, 23-38.

#### QMIN - Quartimin

Carroll, J. B. (1953). An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, 18, 23-38.

#### OMIN - Oblimin

Carroll, J. B. (1957). Biquartimin Criterion for Rotation to Oblique Simple Structure in Factor Analysis. *Science*, 126(3283), 1114-1115.

#### CRFE - Crawford-Ferguson family

Crawford, C. B. & Ferguson, G. A. (1970). A general rotation criterion and its use in orthogonal rotation. *Psychometrika*, 35, 321-332.

#### MINE - minimum entropy

Jennrich, R. I. (2004). Rotation to simple loadings using component loss functions: The orthogonal case. *Psychometrika*, 69, 257-273.

#### IPSC - Bentler's invariant pattern simplicity criterion

Bentler, P. M. (1977). Factor simplicity index and transformations. *Psychometrika*, 42, 277-295.

#### TIIC - Comrey's tandem II criterion

Comrey, A. L. (1967). Tandem criteria for analytic rotation in factor analysis. *Psychometrika*, 32, 277-295.

#### GMIN - Geomin

Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis*. Albany, New York: State University of New York Press.

#### MMER - McCammon minimum entropy ratio

McCammon, R. B. (1966). Principal components analysis and its application in large-scale correlation studies. *Journal of Geology*, 74, 721-733.

## **Example Dataset**

The accompanying example dataset (downloadable from the same site as the Toolkit itself) was kindly provided by Dennis Molfese's lab, which supported the development of the 2.0 version upgrade. This example dataset consists of ten children from a larger study of speech perception. The ERPs reflect the perception of the phonemes "da" and "ga", which vary in the point of articulation, and which in turn are divided into a normal version (da9+ and ga11) and an unrealistic version (da7+ and ga13). Collection of ERP data from such young children (2-4 years old) can be quite challenging as they are very distractible and will move constantly. These data are therefore filled with movement artifacts. An additional feature of the data is that the lower VEOG channels were not placed below the eyes as very young children do not tolerate them well.



# Set Up

## Installation

In order to run the EP Toolkit, you will need to install the following programs:

- (1) MATLAB
- (2) EP Toolkit
- (3) EEGLAB
- (4) FieldTrip
- (5) Satimage osax (for Macs only).

All of these are freely available online except for MATLAB, which has affordable options for student and academic licensing.

### ***1. Install MATLAB***

MATLAB may be purchased directly from The Mathworks, Inc. ([www.mathworks.com](http://www.mathworks.com)), but check with your academic institution first as it may already have a volume license or reduced pricing. Only the core MATLAB program is needed to run the EP Toolkit, so it is not necessary to purchase any MATLAB add-ons from Mathworks. For those who intend to use EEGLAB extensively, it will also run on the core MATLAB program, but the EEGLAB authors recommend that you purchase the Signal Processing Toolbox and Statistics Toolbox to improve the efficiency of some EEGLAB functions. These add-ons are already included in the current student version of MATLAB.

Follow the instructions that come with MATLAB to install it.

At this point, I recommend creating a single folder on your computer to contain the EP Toolkit, EEGLAB, and FieldTrip, as well as any other MATLAB programs you may wish to run on your computer. For example, on my Mac, I have created a folder within the Applications folder named *Matlab\_Programs*. Simply create the new folder, name it, and make sure it is placed where you want it. This keeps the programs organized and provides a common path within MATLAB for accessing applications.

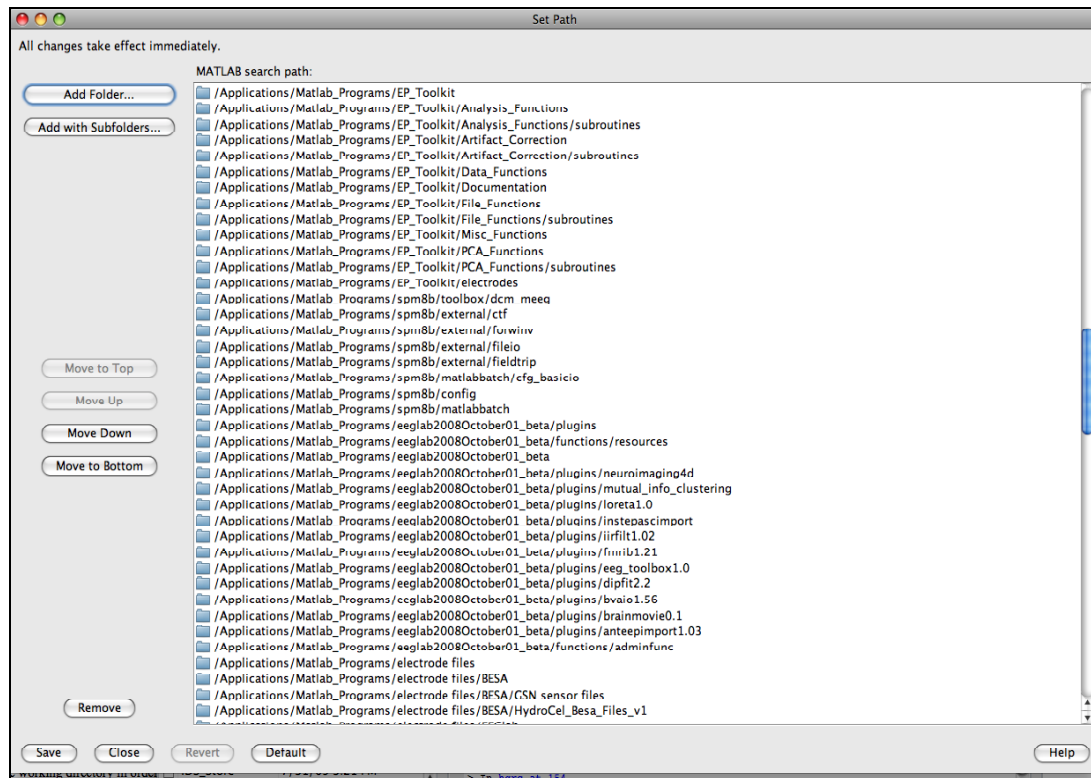
### ***2. Install EP Toolkit***

If you haven't already done so, download the latest version of the EP Toolkit from the project website listed above, and unzip the file. This will produce a folder named *EP\_Toolkit* with several files and folders inside of it. On non-Mac platforms you will find a number of files starting with "\_" which contain Mac-specific information and can be ignored.

Put this EP\_Toolkit folder into the desired directory. On my computer, I place it in the Matlab\_Programs folder that I created and placed in the Applications folder.

Now open MATLAB and add this folder to the MATLAB path. To do this, click on

“File > Set Path...”, then click the "Add with Subfolders" button. A browser window will open. Navigate to the EP\_Toolkit folder (e.g. on my Mac, I click on *Applications*, then *Matlab\_Programs*, then *EP\_Toolkit*). With the EP\_Toolkit folder selected, click the “Open” button. The path list for the Toolkit should end up looking like this, with several lines, one for each subdirectory in the Toolkit folder:



Now save these changes by clicking “Save” at the bottom of the Set Path window. Many new users of MATLAB forget this critical step. If you close the window without saving, MATLAB will discard the path changes and you will have to repeat the process.

### 3. Install EEGLAB

You will also need Delorme and Makeig's excellent EEGLAB toolbox (<http://scn.ucsd.edu/eeglab/index.html>) if you wish to run the InfoMax rotation (ICA). They recommend using the latest development version ("most current"). As of "eeglab8\_0\_3\_3b" this worked fine.

**EEGLAB 6.03b (34MB)**  
[Download link](#)

**EEGLAB6.03b.** Generated Wednesday the 01 of October, 2008.  
(tutorial demo STUDY, separately available in downloads below).  
[Link to EEGLAB revision history](#)  
[Link to previous EEGLAB versions](#)

**EEGLAB most current (34MB)**  
[Download link](#)

**EEGLAB most current.** Recommended. Use the latest generated development version in the folder link (latest is at the bottom 7.x.x.x). It is usually more stable than the version since it contains the most recent bug fixes but can sometimes contains folder or file inconsistencies. Instead of downloading the ZIP, it is also possible to CVS (Concurrent Revision Systems). See [the download wiki page](#) for more information.

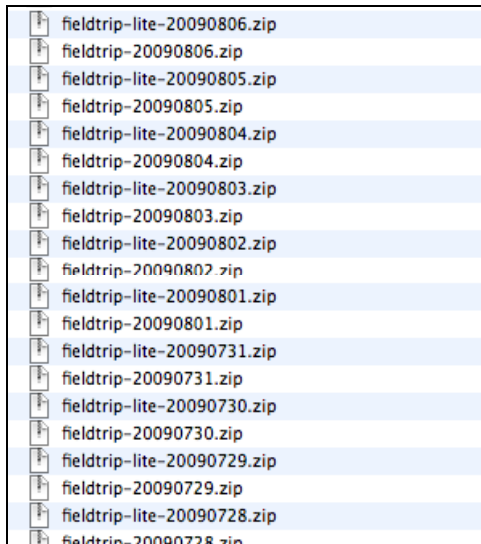
After downloading and unzipping EEGLAB, move the folder to the desired directory (e.g. my Matlab\_Programs folder). Then open MATLAB and add the EEGLAB folder to the path (remember to add its subfolders and save the changes too).

If you install the binary version of EEGLAB's *runica* (<http://sccn.ucsd.edu/wiki/Binica>) it may run the ICA routine much, much faster. The caveat is that the current binary code only uses one core whereas the current version of MATLAB uses all available cores. With the current version of MATLAB and a computer with multiple cores it may be better not to use the binary version. My suggestion is to try it both ways and see which is faster for your configuration. Unfortunately, it is at present only available for Linux, FreeBSD, and OS X. Results are not exactly the same as the current MATLAB version but seem to be quite close (in a test run the largest divergence in the FacPat results was .01).

To install the binary version of *runica* (called *binica*), follow the instructions on the website above. Here is an example of how to do it on a Mac. Download the correct version for your operating system. Unzip the file, and place the resulting folder (e.g. *binica\_osx\_fat*) in the desired directory (e.g., Matlab\_Programs). Then open MATLAB and add the new folder (with subfolders) to the path. Inside this new folder is a file named *ica\_osx*. Open an OSX Finder window to locate *ica\_osx*, and write down the full path to this file. Do a web search if you are uncertain how to view the full path on your version of OSX. Then open MATLAB and use the File menu to open the *functions/sigprocfunc/icadefs.m* file, which can be found in the EEGLAB folder. Edit the line after "ICABINARY =" so that it specifies the full path to the *ica\_osx* file, as in: *ICABINARY = '/Applications/Matlab\_Programs/binica\_osx\_fat/ica\_osx'*. There cannot be any spaces in the path (e.g., "Macintosh HD") to either the binary ICA file or to the working directory in order for it to work. Use the File menu to save this change. Each time you download an update to EEGLAB, you will need to remember to edit this line again.

#### **4. Install FieldTrip**

You will also need Robert Oostenveld's excellent FieldTrip toolbox (<http://www.ru.nl/fcdonders/fieldtrip/>) in order to be able to read in files. Go to the website and register. They will email you an ftp site to download the file from. The site will contain a list that will look like the figure below, if sorted by date. The number is the date of the file (YYYY-MM-DD). Take the latest version (MAKE SURE IT REALLY IS THE LATEST VERSION! If you just take the top file and your computer is set to sort them in order of name, not date, then you will be taking the oldest one, not the newest one). There are two versions for each, the full FieldTrip and the Lite version. Either will work fine. If disk space or internet speed is an issue, choose the Lite version. If you plan to use FieldTrip on its own, consider trying the full version.



Download the file, unzip it, place it in the desired directory (e.g. Matlab\_Programs). Remember to add it to the MATLAB path too.

EEGLAB adds its own stripped-down copy of FieldTrip as part of the installation (in its *external* folder). I would remove this copy (just delete the *fieldtrip* subfolder) and rely instead on the most current version of FieldTrip, as it is continually updated with bugfixes and so forth.

The current version of FieldTrip also installs some EEGLAB files in its *external* folder. I would delete the *eeglab* subfolder found there to avoid conflicts with the EEGLAB installation.

In order for the dipole analysis feature to work, it will also be necessary to have a copy of the gcc compiler installed on your computer. On the Mac, this is installed as part of the optional (free) Developer Tools installation, available on the OS X installation disk. If you hadn't done it during the initial OS X installation, you can still go back and have this added without causing any disruption to your disk drive (make sure to back it up first though). I'm not really certain how to install gcc on Windows but it should be available from the internet for free as it is the open source compiler.

## 5. Install Satimage osax

On a Mac, you will also want to install *Satimage osax* to allow the EP Toolkit to automatically set the montage information in EGIS files. This function will not be available on other platforms, since the implementation relies on AppleScript. Download the software from this address

([http://www.satimage.fr/software/en/downloads/downloads\\_companion\\_osaxen.html](http://www.satimage.fr/software/en/downloads/downloads_companion_osaxen.html)).

No other programs listed at this address are needed, just Satimage osax. After downloading, quit out of all programs, and install the package. Installing creates a file named "Satimage.osax" in the "ScriptingAdditions" directory in your main Library directory. This makes the scripts available "for all users" on the computer; but the option to install for the "current user only" (i.e. in the Library directory under your home

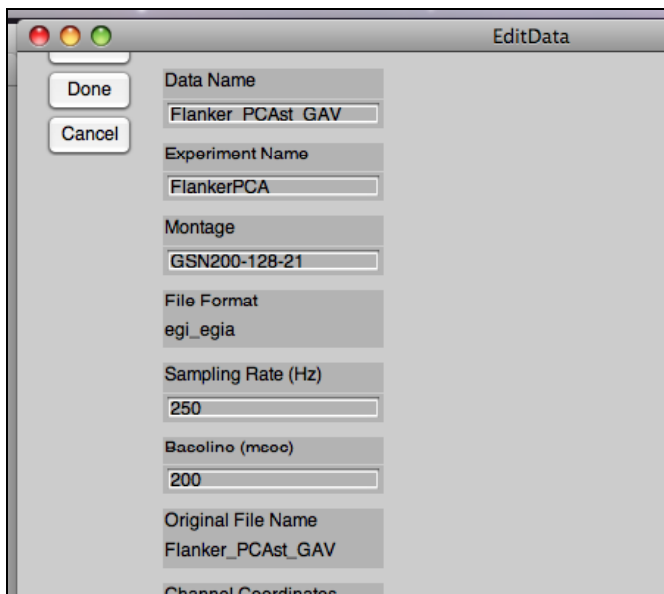
directory) should also work. If the "ScriptingAdditions" directory doesn't exist yet, just create it.

Note that the current version of Satimage osax does not fully operate under OS X 10.6. There appears to be a bug in OS X itself that is keeping it from properly reading the montage information in files but it seems to work fine for inserting the montage information into files that it creates. There is currently no solution for this problem. Hopefully Apple will fix it soon. Bottom line for the user is that, as long as you are running the EP Toolkit on a Mac, you will still be able to open the .egis and .sbin files from NetStation without any problem (if that is the EEG software you are using) but when you open these files from the EP Toolkit it will have to ask you what the electrode montage was rather than doing it automatically. Other file formats will not be affected.

## Avoiding Common Problems

### *Monitor Resolution*

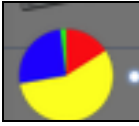
On a Mac with a smaller monitor, make sure to set your Dock to appear on the right or left side of the screen. If it is set to be at the bottom and is abutting onto the EP Toolkit pane then it may throw off its contents, as in this picture. Note how the contents of the pane have been shifted upwards, resulting in some things disappearing off the screen. The Task Bar can cause the same problem on Windows machines.



### *Memory Issues*

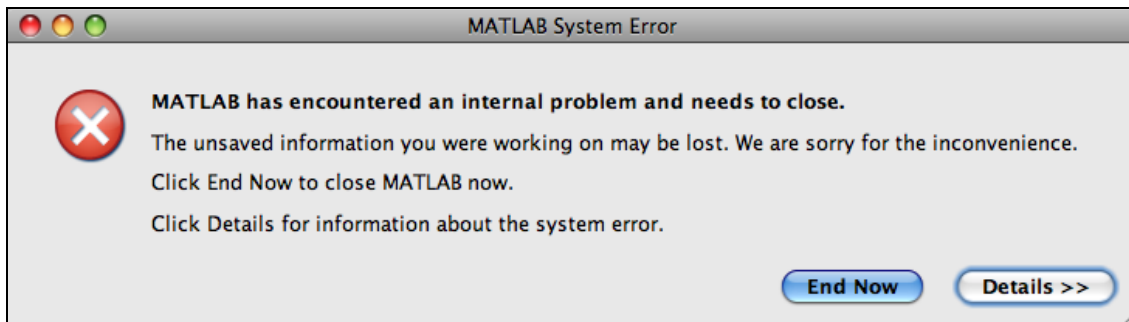
On OS X, it is highly recommended that you use the Activity Monitor (Linux and Windows should have their own equivalent). Go to the Utilities folder of the Applications folder and drag Activity Monitor to the Dock. Click on it to start the

program. Go to View>Dock Icon>Show Memory Usage. This will change the Dock icon to a pie chart divided into green, red, blue, and yellow. Green means free space. Red means in use. Blue means in use by the system (I believe). Yellow means "wired", meaning that it is free but temporarily reserved for some program; this status allows things to go a bit quicker if that program needs it again.



The problem is that MATLAB has very poor memory management. You can think of memory like a large room where objects are placed for later use. Objects are then removed when no longer needed. The problem is that when a new object is added, MATLAB cannot rearrange the existing objects to make efficient use of the floor space. If it tries to store a large object and there is no existing space large enough to accommodate it, then the program will crash. You will start getting alarming red messages about being OUT OF MEMORY in the command line and the EP Toolkit will stop working. When there is enough overall memory but it has simply been cut into too many small pieces to handle the new object (like a new data file), then this is called "memory fragmentation."

Examples of memory errors are:



Or you may see the text:

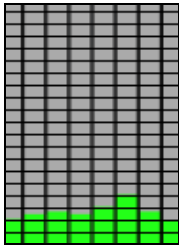
??? Error using ==> zeros

Out of memory. Type HELP MEMORY for your options.

Once memory fragmentation has occurred, the only way to deal with it is to quit out of MATLAB and then restart it. If you see the green slice of the pie in the Activity Monitor shrink to zero, this is an indication that you may soon run into memory problems (although it can happen even when there is still plenty of green). To avoid this situation, start quitting programs to free up the "wired" memory space until there is a substantial slice of green memory space available again. Another symptom of memory fragmentation, short of an actual crash, is that MATLAB starts slowing down to a crawl. Again, the best solution is to just restart MATLAB at this point. The active set of data will still be present when you restart the Toolkit. It's also not a bad idea to set the

Activity Monitor to show the floating window with CPU usage so you can have a better sense of what is going on when you're waiting for the Toolkit to finish a function.

In principle, these memory issues should be greatly ameliorated by the advent of 64-bit computing, which allows for the use of much more memory. On the Mac, this requires OS X 10.6 (Snow Leopard); and on the PC, a 64-bit version of Windows. This also requires MATLAB 2009b and a computer with 64-bit hardware (a recent computer) and lots of RAM.



### ***Energy Saver Settings***

An issue specific to OS X is that MATLAB does not register as an active program to the Energy Saver controls. Thus, if the Energy Saver is set to put the Mac to sleep after a certain amount of time, it will put it to sleep in the middle of an ongoing MATLAB run. Thus, if you are planning on running long MATLAB sessions, such as eyeblink correction runs, it would be best to at least temporarily set Energy Saver to Never Sleep.

## **File Preparation**

### ***EP Toolkit File Types***

The EP Toolkit recognizes five types of files:

1. *continuous* - is a single stretch of unsegmented raw data recording. Artifact correction has been implemented for these files but most other functions will not work.
2. *single\_trial* - is a segmented raw data recording. Most of the EP Toolkit functions are designed to work with these files, including *preprocessing*, *averaging*, and *postprocessing* functions. In principle, it is possible to conduct PCA on these files, but this has not been extensively tested.
3. *average* - is one or more subject averages. The Toolkit's PCA functions are designed to work on combined subject averages in which a single file contains an experiment's entire set of subject averages in the form of separate waveforms. If you have separate subject average files, you can use the "append" function on the subject pane of the Edit function to glue them together into a single combined subject average file or use the single file mode to merge them together with the



Read function.

4. *grand average* - is the average over an entire set of subject averages. It is generally treated the same as an average file, except for some details of how the Toolkit computes the trials-per-average figures.

5. *factor* - is a file that was generated by a PCA. It will consist of a set of reconstructed waveforms, each corresponding to a single factor.

A full EP data file may have all of the following dimensions: channels, time points, cells, subjects, and factors. This structure can be perused via the Edit function and the waveforms may be examined directly via the View function.

The EP file format itself is described in the appendix. While it is really a .mat file, the suffix for EP files is .ept (to help distinguish it from normal .mat files). If using an EP file created previous to the implementation of the .ept suffix, you may just rename the file from ".mat" to ".ept". Note that on the Mac, you should do so using Get Info via a right click or you may find that the renaming isn't working properly (the suffix can be invisible so you haven't actually changed the suffix when you thought you changed it).

As a side note, while it is true that various software packages can produce .mat files, such files are not actually useable by the EP Toolkit. The reason is that basic information like whether channels are the rows or the columns are not specified in such files and basic information like the sampling rate will be entirely missing. It's not worth my effort to add support to the Toolkit for them and it's not worth the effort you would have to make to manually add the missing information every time you wanted to use such a file.

### ***Electrode Coordinates***

Many EP Toolkit routines, especially the artifact correction function, require a file that provides the 3D electrode locations in order to locate the EOG channels and to determine which channels are neighboring. EEGLAB's .ced format is used for this purpose. The EP Toolkit already includes .ced files for the EGI montages and they will be automatically accessed as needed. To create a .ced file from an existing coordinate file in some other format type the following into the Matlab command window:

1) pop_chanedit([]); 2) click on Read Locations. 3) click on Save (as .ced)
---

The .ced file will be treated as the definitive statement on what channels should be in the data file and their proper ordering, except for the EGI EGIS and simple binary formats where all channels are always present and in the same fixed order. Except for these two formats, if the data file has channels that are not present in the .ced file, then they will be used but will not be utilized for operations that need electrode coordinates like bad channel replacement. If there are EEG channels in the .ced file that are not present in the data file, it will be assumed that the channels were bad channels and had been dropped



(normal practice for Neuroscan files) and they will be added back in as zeroed out channels and they will be internally marked as being bad channels.

### ***Reference Channels***

Also, note that data files often have the reference site present only implicitly. Reference channels are by definition zero voltage and so to save disk space the software leaves them out entirely. If the reference channels specified in the .ced file are missing from the data file, they will be added as explicit flat channels. Note also that for a mean mastoid reference, if both mastoid channels are missing then they cannot be reconstituted as only their average is zero; they should both have non-zero waveforms that are mirror images of each other that split the difference between the two sites (for more on reference issues, see Dien, 1998).

### ***Net Station Files***

If you are using Net Station, use either EGIS or Simple Binary formats (the native Net Station format cannot be read by the Toolkit). I recommend using EGIS over Simple Binary unless the file is very large (in which case the file format may not be able to accommodate the data and will generate an error message when you try to save it). Each format has its drawbacks.

The EGIS session format cannot represent baseline period information and so that has to be input manually using the Edit function (to be described). EGIS average files can represent baseline period information but when Net Station saves them, it leaves it out. Also, it leaves out the subject number information. Both types of information will need to be added in using the Edit function. EGIS files are only accurate up to about two decimal places but that is normally sufficient. EGIS files, unlike Simple Binary, can support Trial Specific information but Net Station does not support this aspect of the file format. EGIS files also do not preserve the full event information the way Simple Binary files do but as Net Station does not import the event information in Simple Binary files, this is largely moot. **Note: NetStation sometimes exports EGIS files with the data scaled about three times larger than they should be, causing most of the data to be dubbed bad data. If a data file is labeled as being excessively bad, verify data scaling against the original data file.**

Simple Binary format has the drawback that it does not represent the baseline period information but the FieldTrip import routine (which I wrote) will try to guess what it is. You will therefore need to verify that it deduced it correctly (using the Edit function). More seriously, Simple Binary format does not represent the montage information (the information needed by NetStation to display the data). The Toolkit will ask you what it is when you import the file. On non-Mac computers you will need to wait for the Toolkit to ask you which montage is to be used even with EGIS files. If you are not using NetStation, the montage information is irrelevant.

EGI is trying to phase out the EGIS format but given that Simple Binary is not fully supported yet, I can't recommend using it, except when the file is too large for EGIS to work.

Net Station can generate EGIS and Simple Binary files by either using the Save a Copy... command in the File Menu or the File Export tool. If using the former method make sure to deselect the option to apply viewer transformations.

Text files can also be used as a data export option from NetStation. A separate file will be generated for each segment (for single trial data) or each average (for averaged files). In this case, use the Single File mode to merge these files back together again. **Note: NetStation sometimes exports text files with the data scaled several times larger than they should be, causing most of the data to be dubbed bad data. If a data file is labeled as being excessively bad, verify data scaling against the original data file.**

Another option for data exporting is Matlab file format. In this case, use the File Export Tool, not the Save as menu item. Choose the "cell array" option. Depending on the version of NetStation, this will have either a .mat or a .nsf file suffix. Note that the EP Toolkit does not support importing of event data due to problems in how NetStation exports them (if the "export good segments only" option is chosen, the bad segments are dropped only from the data, not the events, making it impossible to determine which events go with which data).

If you wish to skip the initial steps and go straight to PCA, you can just export an EGIS average file. Make sure that it is a combined subject average file. You can use the Combine Files Tool to generate such a file if you normally keep your subject average files separate. Combining the subject average files makes them easier to work with and is the default approach for Net Station as well. If you subsequently decide to drop a subject or add a subject, you can use EP's Edit function to do so.

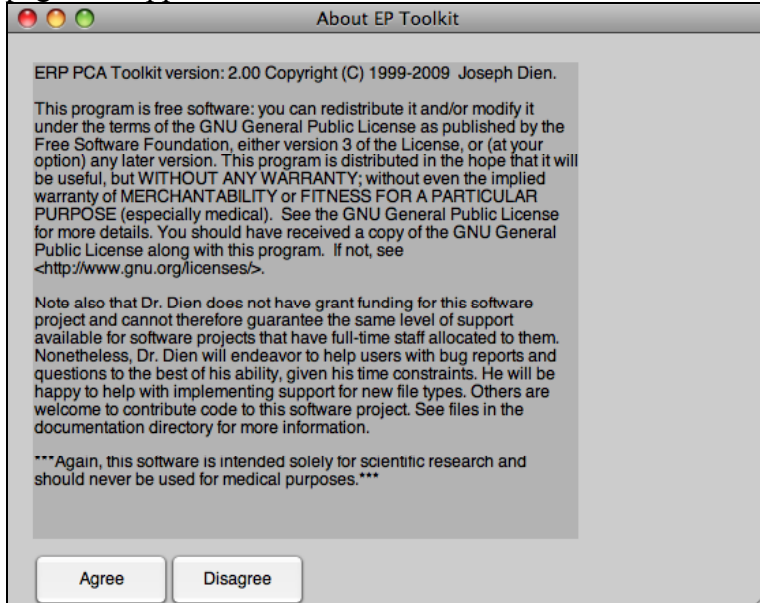
### ***Other Formats***

For average files other than EGI EGIS and EGI Simple binary formats, if the cell names all start with "Sub" and are followed by three numbers then the Toolkit will interpret this situation as meaning that this suffix (Sub####) refers to subject numbers and will arrange the data accordingly. Any text following the three numbers will be interpreted as the actual cell name (e.g., "Sub001target" would be interpreted as the cell "target" for subject "Sub001"). This is a workaround for a limitation of the FieldTrip I/O code.

# Running the Toolkit

## Starting Up

To start the Toolkit, type `ep` at the command line of MATLAB. The following splash page will appear:



Read it and indicate whether you agree with the stipulations. If you disagree then the Toolkit will quit (sorry).

If you are using a version of MATLAB prior to 2008, you may get the following error message:

```
java.lang.UnsatisfiedLinkError: no rxtxSerial in java.library.path thrown while loading gnu.io.RXTXCommDriver
```

Don't worry about it.

You may also periodically get the following alarming message (at least as of MATLAB 2009b):

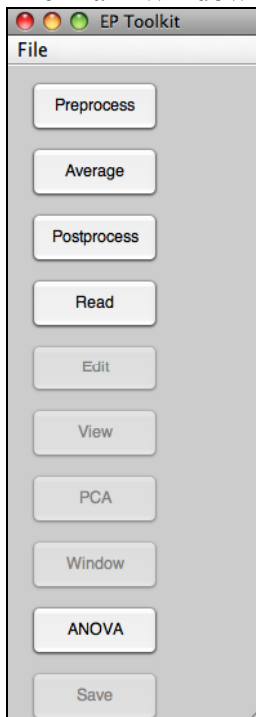
```

Exception in thread "AWT-EventQueue-0" java.lang.ArrayIndexOutOfBoundsException: 2
    at sun.font.FontDesignMetrics.charsWidth(FontDesignMetrics.java:492)
    at javax.swing.text.Utilities.getTabbedTextOffset(Utilities.java:381)
    at javax.swing.text.Utilities.getTabbedTextOffset(Utilities.java:302)
    at javax.swing.text.Utilities.getTabbedTextOffset(Utilities.java:286)
    at javax.swing.text.PlainView.viewToModel(PlainView.java:403)
    at javax.swing.text.FieldView.viewToModel(FieldView.java:263)
    at javax.swing.plaf.basic.BasicTextUI$RootView.viewToModel(BasicTextUI.java:1541)
    at javax.swing.plaf.basic.BasicTextUI.viewToModel(BasicTextUI.java:1090)
    at javax.swing.text.DefaultCaret.moveCaret(DefaultCaret.java:311)
    at javax.swing.text.DefaultCaret.mouseDragged(DefaultCaret.java:565)
    at java.awt.AWTEventMulticaster.mouseDragged(AWTEventMulticaster.java:303)
    at java.awt.Component.processMouseEvent(Component.java:6396)
    at javax.swing.JComponent.processMouseEvent(JComponent.java:3273)
    at java.awt.Component.processEvent(Component.java:6117)
    at java.awt.Container.processEvent(Container.java:2085)
    at java.awt.Component.dispatchEventImpl(Component.java:4714)
    at java.awt.Container.dispatchEventImpl(Container.java:2143)
    at java.awt.Component.dispatchEvent(Component.java:4544)
    at java.awt.LightweightDispatcher.retargetMouseEvent(Container.java:4618)
    at java.awt.LightweightDispatcher.processMouseEvent(Container.java:4299)
    at java.awt.LightweightDispatcher.dispatchEvent(Container.java:4212)
    at java.awt.Container.dispatchEventImpl(Container.java:2129)
    at java.awt.Window.dispatchEventImpl(Window.java:2475)
    at java.awt.Component.dispatchEvent(Component.java:4544)
    at java.awt.EventQueue.dispatchEvent(EventQueue.java:635)
    at java.awt.EventQueue.dispatchEvent(EventQueue.java:635)
    at java.awt.EventDispatchThread.pumpOneEventForFilters(EventDispatchThread.java:296)
    at java.awt.EventDispatchThread.pumpEventsForFilter(EventDispatchThread.java:211)
    at java.awt.EventDispatchThread.pumpEventsForHierarchy(EventDispatchThread.java:201)
    at java.awt.EventDispatchThread.pumpEvents(EventDispatchThread.java:196)
    at java.awt.EventDispatchThread.pumpEvents(EventDispatchThread.java:188)
    at java.awt.EventDispatchThread.run(EventDispatchThread.java:122)
Exception in thread "AWT-EventQueue-0" java.lang.ArrayIndexOutOfBoundsException
Exception in thread "AWT-EventQueue-0" java.lang.ArrayIndexOutOfBoundsException
Exception in thread "AWT-EventQueue-0" java.lang.ArrayIndexOutOfBoundsException
Exception in thread "AWT-EventQueue-0" java.lang.ArrayIndexOutOfBoundsException
Exception in thread "AWT-EventQueue-0" java.lang.ArrayIndexOutOfBoundsException

```

Again, don't worry about it.

The Main Window should also appear, showing the Main Pane:



Note that some of the options will be grayed out until appropriate data have been loaded.

## Preference file & EPwork directory

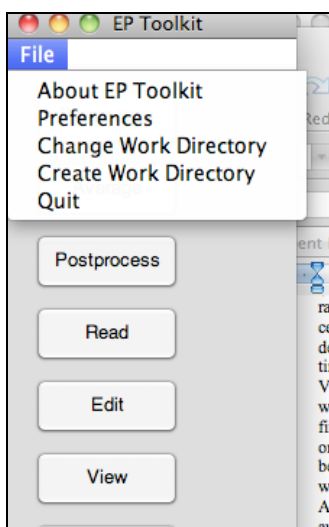
As of EP Toolkit version 2.22, a working directory called EPwork will be placed in Matlab's default user directory. On a Mac, this is in the Documents/Matlab folder. Other operating systems have their own user locations. A permanent preference file called EPprefs will be placed inside the EPwork folder. If you already had an EPwork or EPpreferences folder installed from a prior version of the Toolkit, it will continue to use them.

EPwork contains a working copy of each data file that you have loaded into the Toolkit. DO NOT modify the contents of the EPwork directory. It is meant for the internal workings of the Toolkit.

The EPwork working directory makes it possible for the Toolkit to maintain a working set of data. This set is listed on the Read, Edit, View, PCA, Window, and Save panes. You can add files or remove files from the working set using the Read function. The resulting data file from PCA is also added to the working set. When using these functions, you just select the data to be processed from this working set. When you are done, you must use the Save function to make a permanent copy of the file in the working set.

For the top three functions (prior to Read) the working set is not used. For these three functions (Preprocess, Average, and Postprocess) you will select the files to be processed directly from the disk and the resulting files will be saved directly to the disk. The one exception is when you need to manually construct a blink template.

You can maintain multiple EPwork directories as long as they are located in different directories. When you start up the EP Toolkit, the EPwork directory in the current directory will be used if present. You can switch to a different EPwork directory using the Change Work Directory command in the File menu of the Toolkit and you can create a new one by using the Create Work Directory command.



## Restarting

If the EP Toolkit is stopped in mid session (as by memory fragmentation), you can continue where you left off by restarting MATLAB (on a non-Mac, EPwork will need to be in the active directory when you type ep into the command line). If you have encountered some other type of error state (such as due to a bug in the Toolkit) it may be sufficient to just click on the close box of the Main Window and then just type ep at the command line again. Sometimes MATLAB has been left in an odd state and you may need to type in ep a second time, as in the following:

```
Caught std::exception Exception message is:
St9bad_alloc
Unexpected error status flag encountered. Resetting to proper state.
>> ep
??? Error using ==> set
Invalid handle object.

Error in ==> ep at 1731
    set(EPmain.handles.view.view,'enable','off');

>> ep
```

(The second time the initial splash window appeared as normal.)

## Quitting

When you Quit out of the Toolkit, you will be asked if you wish to delete the working directory. Since it is keeping a working copy of the files, it may take up some amount of disk space so make sure that the disk you start on has sufficient space to accommodate them.

# Toolkit Functions

## Artifact Correction

The Toolkit can automatically eliminate both blink artifacts and movement artifacts from session files, prior to averaging. It is recommended for the data to have already have been segmented into discrete trials if it is meant for ERP analysis as using the full continuous data takes more time and disk space than is needed.

### *Chunking*

The artifact correction function reads in a session file and segments it into manageable chunks which are temporarily stored (since an ICA of an entire session file can often exceed the available RAM). Since the final chunk (the remainder) would often be too small for an effective ICA run, it is appended to the previous chunk and the total chunk is divided into two chunks that are smaller than the others but large enough to process. The chunking is not allowed to divide any trials in two. Then it processes each chunk individually and stores the result. Then it glues them back together into a new session file with a “\_e” appended.

**Note: If the data are large enough to be chunked and you have more than 1GB of RAM, you should go to the preferences menu and see if you can increase the chunking size without running into out of memory errors. About 100,000 per GB of available RAM seems to generally work well.**

### *Default Procedures*

Preprocessing parameters can be changed (see Setting Preferences section below), but the default procedures of the Toolkit are as follows:

1) *Bad channel identification* - The first step is to check for any channels whose best absolute correlation with neighboring channels falls below .4. These channels are considered to be globally bad and are excluded from further analyses. If it has more than ten percent bad channels then it will be considered a bad subject and it will not be further processed. A warning will be provided if any of the globally bad channels are neighbors. Additionally, flat channels are identified as being bad channels (unless it was identified as being the reference channel). Additionally, channels that correlate perfectly with a reference channel will be marked bad (this happens when a file with a flat bad channel is rereferenced). Channels that are shorted together (perfectly correlated) will generate a warning in the log but will not be touched.

2) *Blink correction* - The next step is to perform the blink correction process. The artifact detection routine uses Makeig and Delorme's runICA routine to decompose the data into basic scalp topography components. These components are then compared to a blink template which specifies the topography of a typical blink. Components which correlate highly with this blink topography are deleted from the data. Since this process can

increase noise in trials without a blink (experience shows this), this subtraction is implemented on a trial by trial basis, using the criterion that the subtraction must not increase the overall variance of the trial data. The removal of the blink topography will apply to blink recoveries as well (which have the inverse scalp topography and are caused by the amplifiers recovering from the sudden blink amplitudes).

There are two options for the blink template. The first is to use a file with a manually generated template of blink topography. The second is to generate automatically an individualized blink template for each individual chunk. It requires at least two good VEOG channels (one upper and one lower), however. It operates by identifying trials where the upper and lower VEOG pairs covary within each pair and negatively vary with the other pair (if all four VEOG channels are present). The maximum amplitude timepoints of all such trials are averaged together as the blink template.

Normally, the EP Toolkit is able to automatically determine which channels are the VEOG channels if it is given the electrode coordinates information. If the automatic determination process is not working, it is also possible to set the EOG channels manually via the preferences settings. Just in general, it is best to check what channels are being used for the blink correction and confirm that they are the correct ones.

3) *Detrending* - The next step can be to detrend the data, although the default is not to do this. Detrending can be helpful for salvaging noisy data with strong DC drifts but will also affect the final ERP waveform. Detrending is accomplished by subtracting the mean across all time points in an epoch from each individual time point in the same epoch. The mean calculation and subtraction are performed in a trial specific and channel specific manner. You can think of this like baseline correction with the baseline set to be the entire epoch. Subsequent baseline correction will undo this procedure.

4) *Movement artifact removal* - Using an experimental approach (not yet peer-reviewed), movement artifacts can then be removed. Whereas ICA is highly effective at removing eye blinks, it is not suited for movement artifacts because they are typically highly variable between trials so ICA can't be applied to them at the level of a dataset (there may not even be enough ICA factors available to account for all of them, let alone the ERPs themselves, since there cannot be more factors than there are variables, which in this case is the number of channels). ICA can't be applied to isolated trials because it needs more observations than are available from a typical trial. In contrast, PCA (using the Promax rotation) works even with small numbers of observations and seems to provide adequate results for individual trials. For this reason, a temporal PCA (using the Promax rotation) is used to identify factors which account for activity with an amplitude difference between minimum and maximum values greater than 200  $\mu\text{v}$  and this activity is removed.

5) *Bad trial identification* - Bad channels are then detected on a trialwise basis. They are defined as either having a difference of more than 100  $\mu\text{v}$  from the minimum and the maximum values in that trial or as differing at some point by more than 30  $\mu\text{v}$  from all the neighbors (the six closest channels). Flat channels are also marked as being bad, unless they were identified as being a reference channel AND the channel is flat over the



entire dataset. If more than 10% of the channels (including globally bad channels) are marked as bad then the entire trial is marked bad. If a channel is marked as bad on more than 20% of the trials, then it is marked as globally bad.

6) *Horizontal eye movement routine* - The EP Toolkit now can remove saccade artifacts. Since saccade artifact topographies are not as invariant as those of blink artifacts (the topography depends on whether the saccade was exactly horizontal or whether it was at some other angle), it was necessary to implement it on a trialwise basis. Since it is applied on a trialwise basis, there are not enough observations in a typical trial to utilize ICA. A PCA with Promax rotation is therefore used (spatial since the time course will not have a distinctly identifiable signature, even though Promax is not as effective in the spatial domain). Since saccades artifacts are much smaller (about 20  $\mu\text{v}$  rather than about 200  $\mu\text{v}$ ) they are harder to detect. It has not therefore been possible to implement an automatic saccade template option. Instead, the saccade template must be formed manually. The EP Toolkit will also note the onset latency of the saccades. It is also not clear to me yet how well this procedure works so it should be considered experimental and used carefully.

7) *Bad channel replacement* - If the "replace" option was chosen, channels that were marked bad are replaced via interpolation from the good channels of the scalp voltage field (the entire scalp is interpolated and then the missing channel voltage is obtained from this interpolated scalp surface) and bad trials are zeroed out.

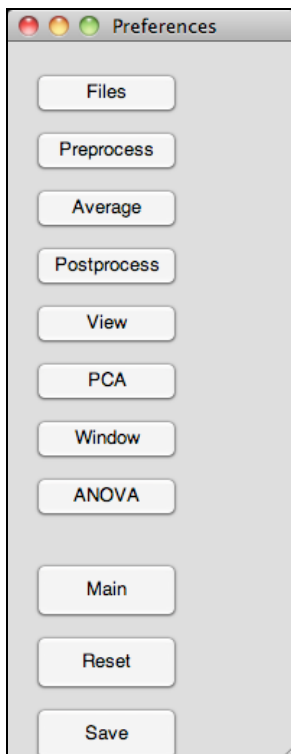
8) *Log file* - A log file is generated to summarize the results of the session.

### ***Setting Preferences***

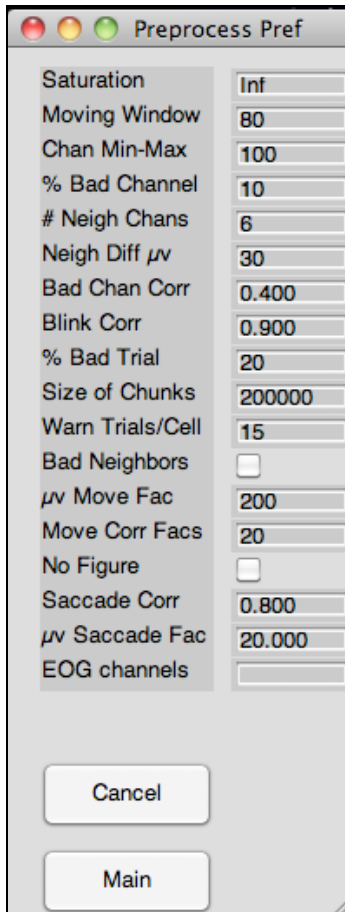
Before preprocessing data, you might want to adjust the preferences. Choose Preferences from the File menu.



Then click on Preprocess:



The following window should appear:



1) *Saturation* is the voltage level at which the amplifier reaches the maximum number that it is capable of reporting. Time points reaching this value are treated as bad data during the blink correction process and ignored. If this is not an issue, then just leave it at the default value of "inf", meaning infinity.

2) *Moving Window* is the number of milliseconds over which the artifact correction routines average the data in a form of low pass filtering. The larger the number, the less sensitive it is to high frequency spikes.

3) *Chan Min-Max  $\mu\text{v}$*  is the maximum allowed change in voltage levels for a channel during a trial before it is deemed to be a bad channel for that trial.

4) *% Bad Channel* is the maximum percentage of channels allowed to be bad in a trial before it is deemed to be a bad trial.

5) *# Neigh Chans* is the number of channels considered to be a neighbor for purposes of the artifact correction algorithms. The electrode coordinates are then used to figure out which channels are to be used.

6) *Neigh Diff  $\mu\text{v}$*  is the maximum voltage difference allowed between a channel and its neighbors before it is deemed to be a bad channel.

7) *Bad Chan Corr* is the correlation criterion for determining whether a channel is a globally bad channel over the entire session. If its best correlation with any neighbor is lower than this criterion then it is judged to be globally bad and dropped from any further consideration.

8) *Blink Corr* is the correlation criterion for determining whether an ICA factor matches the blink template and should therefore be subtracted from the data.

9) *% Bad Trial* is the maximum number of trials a channel is allowed to be judged bad before it is deemed to be globally bad.

10) *Size of Chunks* is the number of time points that are read into each chunk (about 100,000 per GB of available RAM seems to generally work). If there is sufficient memory, it is best to process a datafile as a single chunk so be sure to set this number as high as is possible. Unfortunately, on the Mac it is not possible to have MATLAB set this parameter automatically so the user has to set this. If the data files are already being processed as a single chunk then raising this number will have no effect.

11) *Warn Trials/Cell* is the minimum number of good trials that is considered to be sufficient for a cell. Any cells dropping below this number will trigger a warning in the artifact correction log. There is no other effect of this setting.

12) *Bad Neighbors* is an option where if two neighboring channels are marked as being locally bad then the trial is also marked bad (because this typically means that a movement artifact of some sort is present, as opposed to isolated bad channels). This is a very stringent criterion and should not be used for developmental data.

13)  *$\mu$ v Move Fac* is the maximum voltage difference (maximum-minimum) allowed by a factor by the movement artifact correction step. Factors exceeding this limit are deemed to reflect artifacts and are subtracted from the data.

14) *Move Corr Facs* is the number of factors to be retained by the movement artifact correction routine. A larger number results in a more accurate but slower process.

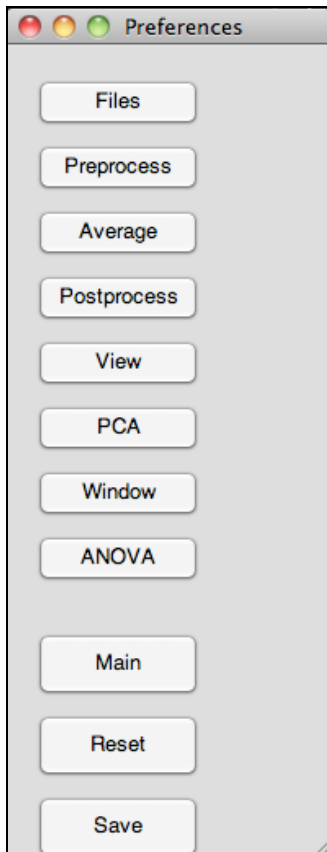
15) *No Figure* is an option to not provide a summary figure for the artifact correction process. While a very useful figure, it requires substantial memory and so dropping it can be helpful when encountering recalcitrant memory problems.

16) *Saccade Corr* is the correlation criterion for determining whether an PCA factor matches the saccade template and should therefore be subtracted from the data.

17)  *$\mu$ v Saccade Fac* is the maximum voltage difference (maximum-minimum) allowed by a factor by the movement artifact correction step. Factors exceeding this limit are deemed to reflect artifacts and are subtracted from the data.

18) *EOG channels* are the EOG channels. If left blank, the Toolkit will try to automatically determine which are the EOG channels based on electrode coordinates. If it fails to properly identify them, then fill in this field to specify them manually: [LUV RUV LLV RLV LH RH]. Specify missing EOG channels with a -1.

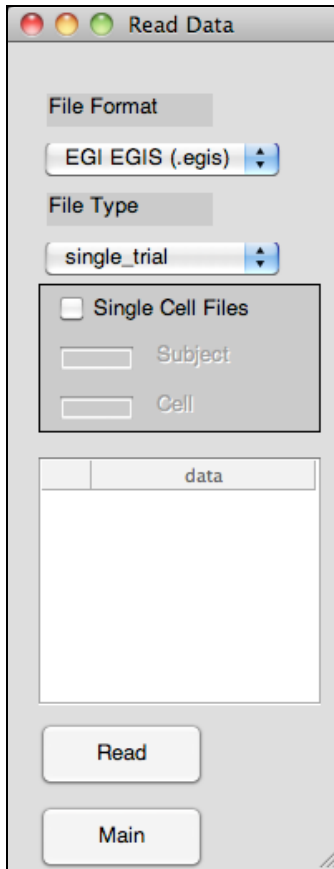
If any changes are made, then click Main, otherwise either Cancel or Main is fine.



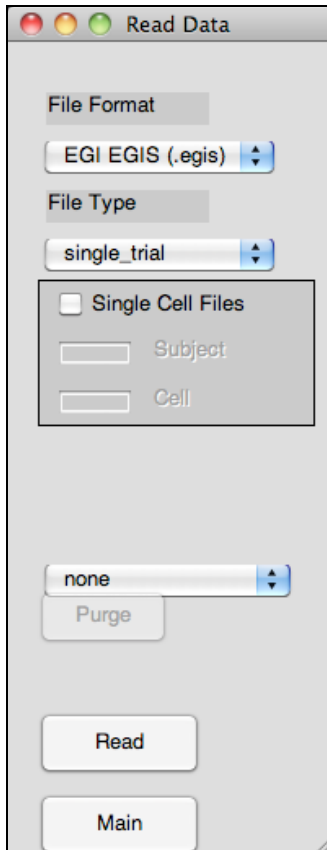
Click Main to return to the Main Pane. Click Save to make any changes permanent (applying to future EP Toolkit sessions). Click Reset if you wish to return all the preference settings to the original default values.

### ***Manually Forming a Blink Template***

The following example uses the files included with this tutorial (see *Example Dataset* in the *About the ERP PCA (EP) Toolkit* section above for instructions on how to download these files). In this dataset, the lower VEOG channels were purposely not placed in the normal location (as very young children do not tolerate them well) and the blinks cannot be identified automatically (the lower and upper VEOG channels do not diverge from each other in the usual manner). The first step is therefore to generate a blink template manually. First click on Read in the Main Pane to load in the session files. This will bring up the Read Data Panel. Change the File Format to the appropriate type. The File Type should be "single\_trial" since these are session files.

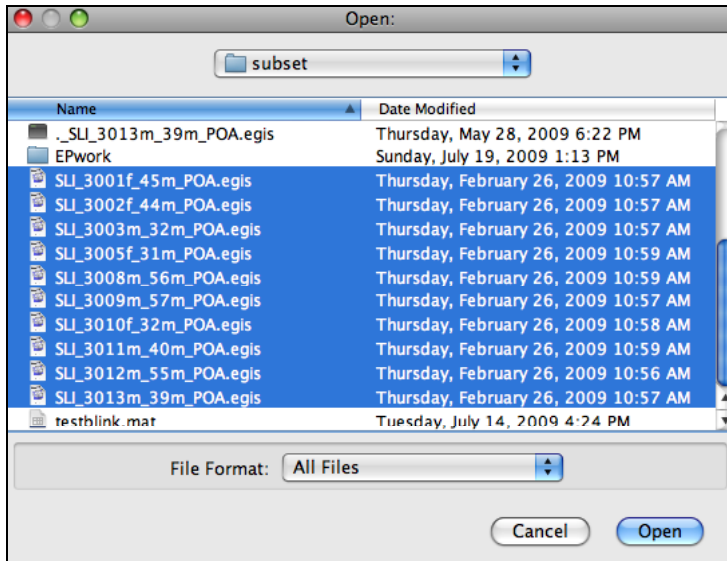


Note that MATLAB versions 2006 and 2007 do not support the Table element in the user interface and so the Toolkit will substitute cruder looking controls.



For your own data, if the data take the form of separate files for each cell then it would be necessary to click on the Single Cell Files option and to specify in the Subject and Cell fields which characters in the file names correspond to the subject and cell labels (as in 4:6 and 7:9 for APP001TAR.txt for subject #001 in the TAR condition). Note that it will be necessary for each file to have an appropriate suffix in order to be recognized (listed along with the file format name). For text files, the General Preferences pane lets you indicate which rows and columns to import and whether the channels are in the rows or the columns of the files.

Then click on Read and choose all ten example files and then click on Open:



On a Mac, it may then ask you to identify the montage if the Satimage osax is not fully working (as noted earlier). If so, choose "Adult GSN200 128-channel 2.1". On a non-Mac, it may ask you to identify the Electrode Coordinate file, which would be "GSN129.ced".

For each file, the MATLAB command line will say something like:

**Read in 4 cells with a total of 180 trials.**

**osascript**

**'/Applications/Matlab\_Programs/EP\_Toolkit/File\_Functions/subroutines/readSLAY.spt'**

**'/Applications/Matlab Programs/EP Toolkit/File Functions/subroutines/SLAY.rtf'**

### 136: Segmentation fault

**Reading file (lines): 10 20 30 40 50 60 70 80 90 100 110 120 130 132**

The name of the experiment is:  Sp48  P

ø□□4~È□49" O}d□ segm OÆ^ yyy y O}d

**The pre-stimulus period of the data is: 0 msec.**

**Assuming the electrode information is the same as for the last file.**

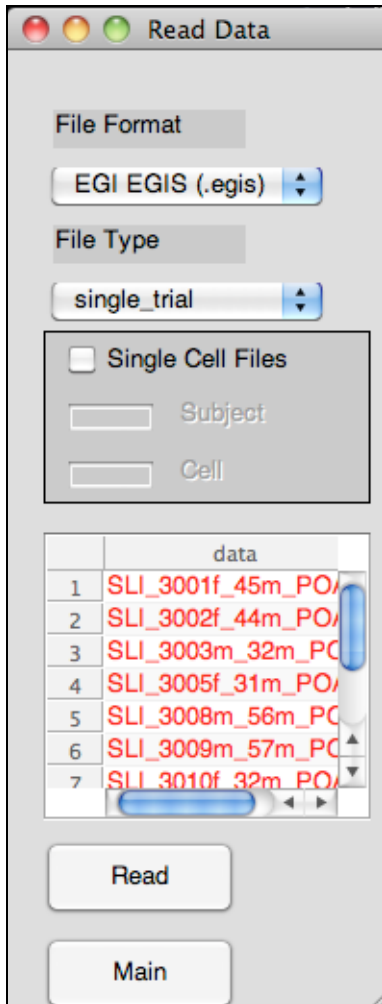
**Assuming the montage information is the same as for the last file.**

You can fix the baseline and experiment name information later on if you'd like. The segmentation fault error is an indication that the Satimage osax function is not working properly (as noted earlier, it doesn't fully work with OS 10.6), requiring it to ask you what montage to use.

The files that you loaded in will now be listed in red. A copy of them have been loaded into the working directory (EPwork) but will not be taking up space in the computer memory. If you need to purge any from the set of work files, click on them and they will be immediately removed (the red is a reminder not to click on them unless you really want to delete your working copy of the file). Also, a star will be displayed in front of



every file that contains unsaved changes (changes following the initial file import but not including anything that occurred during the import itself).

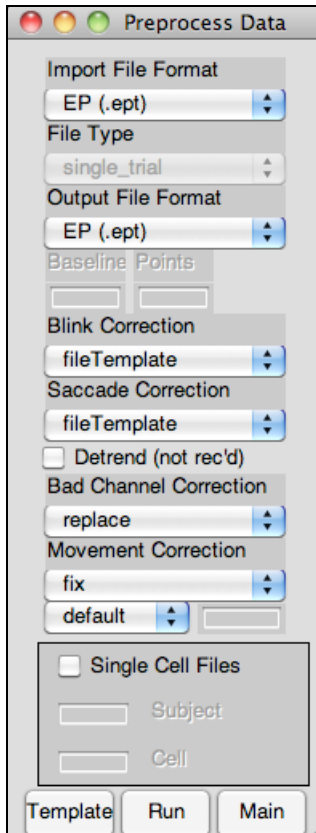


The 'Read Data' dialog box is a vertical window with a title bar containing three colored buttons (red, yellow, green) and the text 'Read Data'. It contains several sections: 'File Format' with a dropdown menu showing 'EGI EGIS (.egis)'; 'File Type' with a dropdown menu showing 'single\_trial'; a section with a checkbox 'Single Cell Files' and two input fields labeled 'Subject' and 'Cell'; a table with a single column 'data' containing seven rows of file names; and two buttons at the bottom labeled 'Read' and 'Main'.

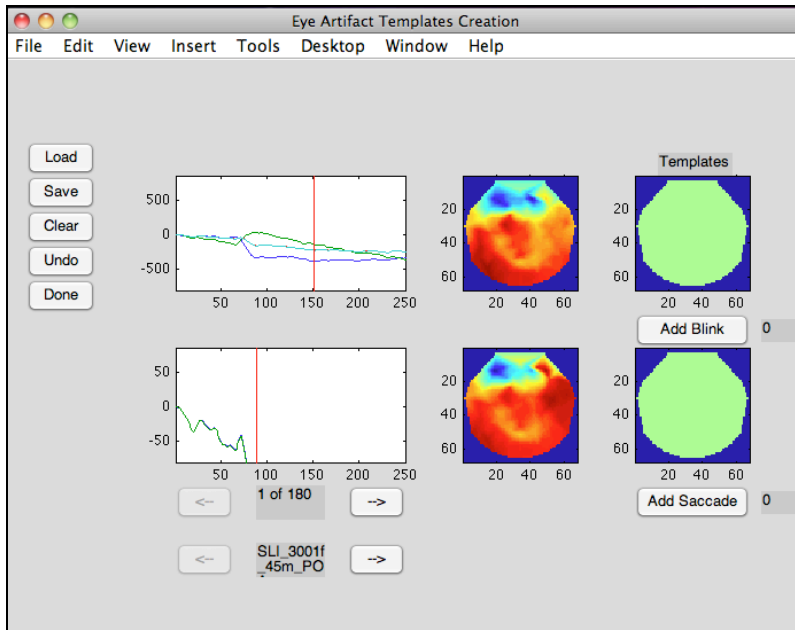
	data
1	SLI_3001f_45m_PO/
2	SLI_3002f_44m_PO/
3	SLI_3003m_32m_PC
4	SLI_3005f_31m_PO/
5	SLI_3008m_56m_PC
6	SLI_3009m_57m_PC
7	SLI_3010f_32m_PO/

Now click on Main to go back to the Main Pane.

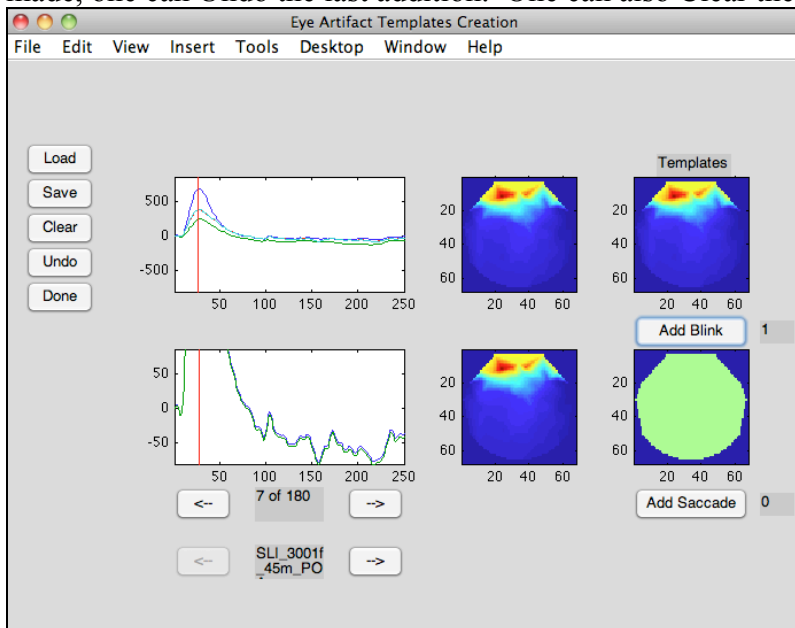
Next, click on Preprocess.



Click on Template to bring up the Template Creation Window. In the first row, it shows the VEOG waveforms of the first data file. It also marks the time point with the maximum voltage (across all the channels, including the ones not plotted so it may not always be apparent why it's chosen the time point) with a red line. To the right of the waveforms is the scalp topography at that point of maximum voltage (across all the channels). To the right of that is the scalp topography of the new blink template. Since nothing has been added to it yet, it is a blank green. The second row shows the two HEOG channels and the scalp topography at the point of maximal divergence between the two HEOG channels.



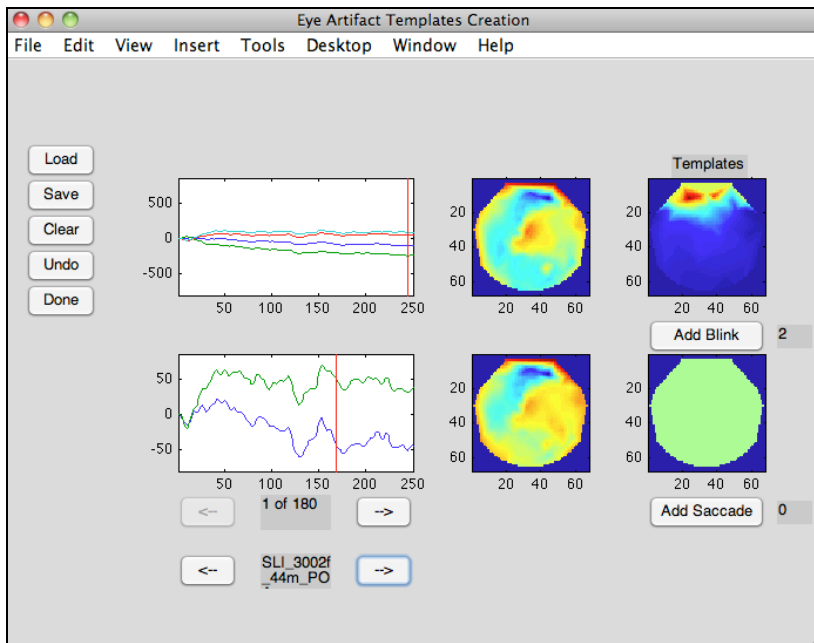
Clicking on right arrow below the waveforms brings one to the first blink (in the 7th trial). It can be recognized as such because of the smooth, high amplitude bell-shaped curve which the scalp topography image shows to be tightly centered around the eyes. If the lower VEOG channels were correctly placed, they would be seen to be going negative even as the others went positive. Click on the Add Blink button to add this blink to the template. The counter below the blink template image increases to one. If a mistake was made, one can Undo the last addition. One can also Clear the template formed thus far.



There is no hard and fast rule on how many blinks to use in a template but I find that having ten of them (two each from five subjects) tends to yield acceptable results. Using two each helps remind one to confirm that a putative blink really is a blink as blinks

should be quite consistent across a session. Using blinks from multiple subjects helps ensure that the template is not overly specific to a single subject. The template doesn't have to be exact, though, since the eyeblink correction process only uses it to determine which ICA factors correspond to blinks and so the match to an individual's blink does not need to be exact (the default setting in the preferences is to correlate .9).

Once the second blink has been added to the template, click on the lower right arrow to bring up the next active dataset. The trial counter will be reset to one.



The second subject in the example dataset has no apparent blinks. Proceeding on to the rest of the session files, the blinks are reasonably consistent across the datasets, resulting in a blink template quite similar in appearance to the first blink found. Avoid examples of blinks that are noisy as they will degrade the effectiveness of the template.

If some of the subjects had the lower VEOG channel placed in the normal fashion (they aren't), it would be best to test them separately as their blinks would be different. One could add only the displaced lower VEOG blinks to the template and then use the "bothTemplate" option with the expectation that the manual template would catch the displaced VEOG blinks and the automatic template would catch the normally placed VEOG blinks. Alternatively, one could form a separate manual blink template for each subject and run them individually (usually not necessary).

Although this dataset is not appropriate for demonstrating saccades, they are handled in much the same fashion. They have an asymmetric topography with the HEOG in the direction of the saccade becoming positive. The HEOG waveforms should start diverging at about 200 msec after the stimulus onset.

Once finished, click on Save and give the file a name. It will ask to do so twice, once for

the blink template file and once for the saccade template file (if such a template has been constructed). One can also Load a template back for later inspection or to add more blinks to it. Click on Done to go back to the Preprocess Data Window.

Once the template has been formed, it is no longer necessary to have the files in the working set. The preprocessing step operates directly upon the files on the disk (since reading in a large batch of files to the working set can take some time, it's more efficient to just read them in one at a time during the batch session rather than having to wait around). To avoid confusion, one may wish to purge the working set at this point. To do so, click on Main to return to the Main Pane and then click on Read. From the Read Window, one can purge the files by clicking on their names. Once done, return to the Preprocess Window (by clicking on Main and then click on Preprocess).

The next step is to provide the proper settings for the preprocessing. It is recommended to set the Output File Format to EP .mat since this is the native file format for the EP Toolkit program and will preserve information that might otherwise be lost should another file format be used. At the end of the process the data can be saved into a different format as needed.

The Baseline field specifies samples to use for baseline correction (ignoring the baseline field of the data file). If the baseline parameter is set, the function will baseline correct each trial using the time points specified (e.g., 1:25 means use the first twenty-five time points). This ensures that random swings in the baseline do not degrade the quality of the eyeblink correction process. The use of this option is strongly recommended. Note, though, that the output data will still be baseline corrected. For this tutorial dataset, type in "1:25".

The next field is the Points field. Oftentimes only a portion of the epoch is needed for the analysis. If so, it may be desirable to specify the timepoints in the Points field. Not only would doing so save time and disk space, it would also avoid losing trials due to artifacts outside the parts of the epoch of interest. They should be described in terms of samples. Colon means that a range of numbers is being described. For example, "1:250" would indicate that the first 250 time points should be retained. If each sample is 4 ms (i.e., 250 Hz sampling rate), then these samples would correspond to the first second. If the field is left blank then by default the entire trial will be retained. The label of the field is initially grayed out to denote that this control is inactive but if timepoints are typed into the field then the label will no longer be grayed out.

The next setting indicates whether the blink correction step should use a manually generated blink template ("fileTemplate"), an automatically generated template ("autoTemplate"), or both ("bothTemplate"). For the tutorial, use "bothTemplate."

Below this is the Saccade Correction setting. It determines whether saccade correction will be carried out. There is no option for using an autoTemplate.

The next line indicates whether to detrend each trial. In general, this option is not

recommended but may sometimes be of use.

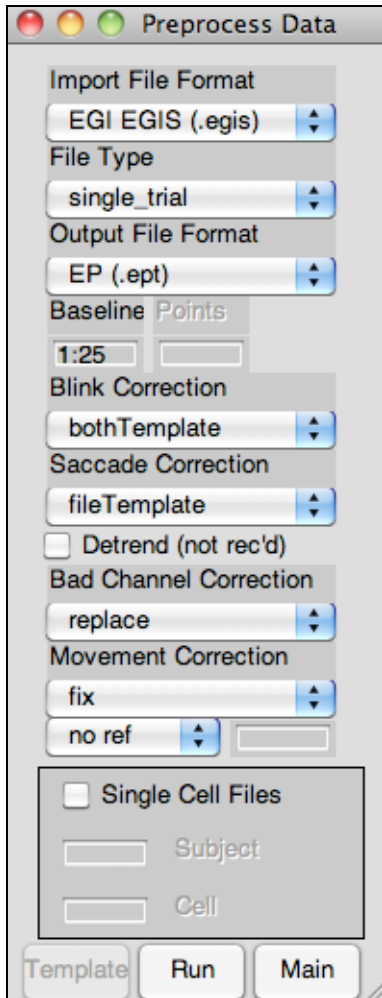
The Bad Channel Correction control indicates whether to "replace" bad channels (via interpolation) and to zero out bad trials, to "mark" the bad channels and bad trials with a flat line and a large spike, or to do nothing except to record their existence in the log. For the tutorial, choose "Replace."

The Movement Correction control indicates whether to use the experimental PCA procedure to "fix" movement artifacts or not. For the tutorial, choose to use it (quite helpful for developmental datasets such as this).

The next control is for indicating which, if any, of the channels are explicit reference channels, meaning that their voltage numbers are included in the data. A single reference channel will by definition be a flat line of zeros. Usually, in order to save on disk space, such a channel is simply left out of the file and is considered to be present implicitly. Mean mastoids reference channels will have a mean of zero and will split the difference, becoming mirror images of each other. Sometimes one of them will be left out of the file as it is redundant with the other one but will still be considered to be present implicitly. The reference channel(s) needs to be indicated when: 1) a single channel reference is explicitly present so that the EP Toolkit will know that it is flat because it is a reference channel rather than because it is a bad channel. 2) when mean mastoids reference is utilized and both channels are explicitly included because the ICA process used to extract eye blinks can be derailed by their presence (since they are mirror images of each other, they will have a perfect correlation of -1, which can end up causing computational problems). "ave ref" means average referenced. "no ref" means no explicit reference channels. "1-2 refs" means one or two explicit reference channels. "default" means that the reference channels indicated by the CED file (when the data file was imported) should be used. These values provided by the CED may not be correct if they indicate the original reference electrode and the data were subsequently rereferenced to some other scheme. In this tutorial dataset Cz was used as the recording reference and is present only implicitly and so "no ref" should be used. It will be added back in later on when the data is rereferenced to average reference. For those using NetStation, it normally represents the reference channel (Cz) implicitly in session files until they are either averaged or rereferenced so you would choose "no ref".

The Single Cell Files operates in the same fashion described under "Read Data".

When done selecting the proper settings, click on the Run button.



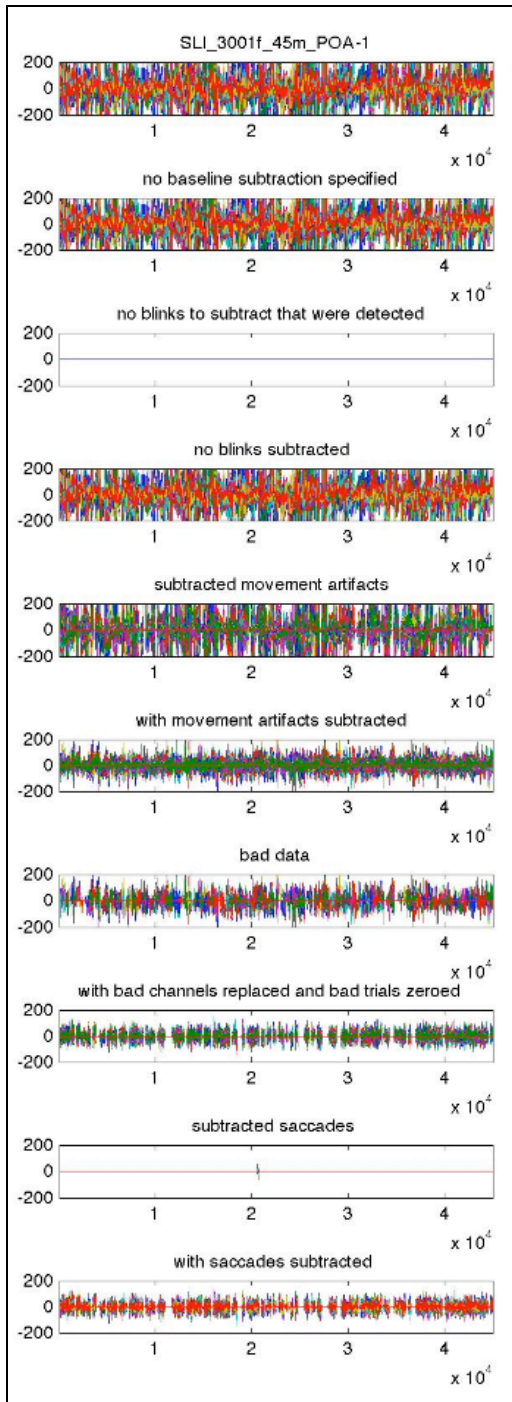
Processing will generally take about 1-3 hours per data file, depending on its size and the speed of the computer. The channel replacement can take up half the total period. The tutorial files take one hour each on a dual 2.8 GHz 8GB Quad-Core Intel Xeon OS X 10.6.2 Mac running MATLAB 2009b (using fully all eight cores, according to the Activity Monitor). It's best to just set up an entire batch and let it run overnight. Remember to set the computer to not sleep if it is a Mac.

During the artifact correction process, a new Artifact Correction figure will be presented illustrating the progression of the process. For each chunk (there will be multiple chunks if the size of the data file is larger than the maximum chunk size that you set earlier on in the preferences), it will chart the state of the data in a butterfly plot, which consists of all the trials laid end to end, with one line corresponding to each channel. Artifacts can be easily detected as high amplitude divergences from the central mass of lines. Several such plots are provided:

- 1) the original data
- 2) the data after baseline correction
- 3) the portion of the data identified as blinks. They should be sharp spikes.

- 4) the data with the blinks subtracted.
- 5) the portion of the data identified as movement artifacts.
- 6) the data with the movement artifacts removed.
- 7) the portion of the data identified as bad channels or bad trials.
- 8) the data with bad channels interpolated and bad trials zeroed out.
- 9) the portion of the data identified as blinks. They should be sharp spikes.
- 10) the data with the blinks subtracted.





The following files will be generated:

- 1) an eyeblink corrected file with a "\_e" suffix added.
- 2) a .jpg file of the butterfly plots. Each one should be examined, principally to determine if the blinks were in fact removed. If not, it may be necessary to rerun the file using a customized blink template formed using blinks just from the one file.
- 3) A file titled Artifact\_Correction\_Log containing the record of the preprocessing

session. It is highly recommended that it be examined carefully for signs of problems.

For the first session file, the log file output is:

#### ARTIFACT CORRECTION SUMMARY

The blink channels are: LUVEOG(26) RUVEOG(8) LLVEOG(127) RLVEOG(126).

The saccade channels are: LHEOG(128) RHEOG(125).

Minimum number of good trials per cell to avoid warning message is: 15.

Moving average window for smoothing during bad channel detection: 80 ms.

Difference from minimum to maximum for bad channel: 100  $\mu$ v.

Percent of bad channels exceeded to declare bad trial, rounding down: 10%.

Difference from minimum to maximum for bad horizontal EOG: 55  $\mu$ v.

Number of electrodes considered to be neighbors: 6.

Minimum predictability from neighbors to not be considered globally bad: 0.4.

Maximum difference from most similar neighboring electrode to be considered bad: 30  $\mu$ v.

Blink template option is: fileTemplate.

Threshold correlation with blink template: 0.9.

Threshold correlation with saccade template: 0.8.

Detrend data: OFF.

Percentage of good trials chan is bad to declare a channel globally bad: 20%.

Number of factors to retain when correcting movement artifacts: 20.

Maximum voltage difference allowed when correcting movement artifacts: 200.

Bad channels and trials: replace.

Do not allow adjacent bad channels (trial declared bad): OFF.

\*\*\*\*\*

Working on: /Applications/Matlab\_Programs/tutorial/SLI\_3001f\_45m\_POA.EGIS.

The channel 129 is marked as being the reference.

Warning: shorted channels: E125-E126; E125-E127; E126-E127; E126-E128; E127-E128;

Global bad Channels: None

No components match blink template so no correction performed.

173 trials corrected for movement activity.

There were 45 bad trials.

Originally there were 0 bad trials.

There were 45 bad trials due to too many bad channels.

There were 0 bad trials due to neighboring bad channels.

For good trials, there was an average of 4.5736 bad channels per trial.

1 trial corrected for saccade activity.

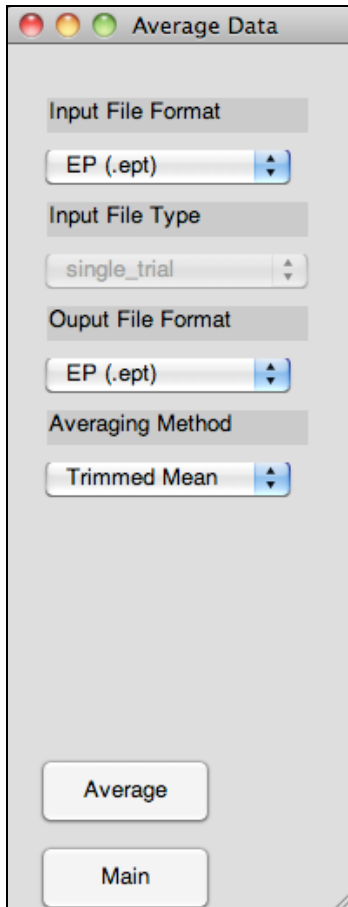
The subject took 71 minutes to process.

Done.

The header gives information about the settings used for the preprocessing run. Then each section after that provides a summary of each file processed. In this case, it is providing a warning that some pairs of channels appear to be shorted together as they are nearly identical.

## Averaging

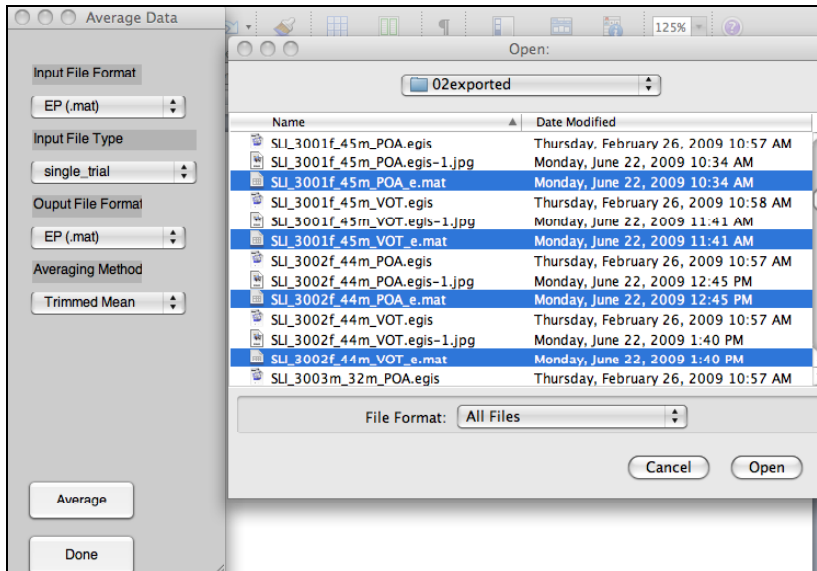
The next step is to average the data. Click on Main to return to the Main Pane. Next, click on Average.



Set the file formats and types to the appropriate values. It is recommended to use the EP .mat format. In the figure, File Type is grayed out because the .ept format includes the information about what type of file it is. The averaging options are:

- 1) mean - which is the conventional averaging procedure.
- 2) median - which uses the median instead of the mean.
- 3) trimmed mean - which uses the mean but first trims off the most extreme values at each time point. The default value (which can be changed under Preferences) is 25% so the highest and lowest 25% will be dropped. This results in a more robust estimate of the central tendency of the event related potentials. The trimmed mean is recommended over the median (Leonowicz, Karvanen, & Shishkin, 2005).

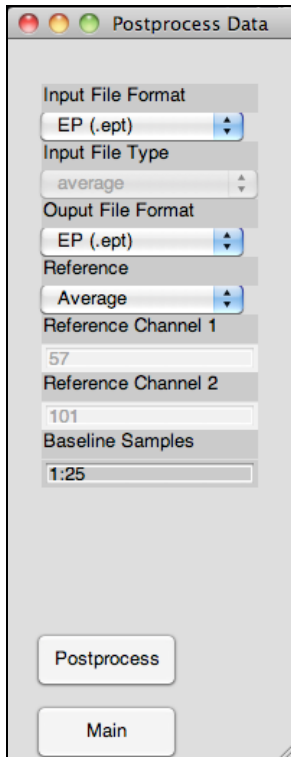
When ready, click on Average. Select the desired session files and they will be combined into a single combined subject average file (it will ask for a name) with separate averages for each subject. The files will be selected from those on the disk and not from those in the working set. This process will take a while. The command window will announce which file is being worked on. When it is finished, the Average and Main buttons will no longer be grayed out. If the session files were generated by the Toolkit's preprocessing step and are in EP format, then trials that were marked as bad will be dropped from the average. Likewise, uncorrected bad channels will be dropped. If a set of subject average files are selected, then a single grand average will be generated from them.



Click Main once it is no longer grayed out to get back to the Main Pane.

## Postprocessing

Next click on Postprocess to rereference and baseline correct the data. It is assumed that filtering was already performed prior to segmentation (using some other software) as otherwise filtering artifacts are produced at the edges of the segmented epochs.



Set the file formats and types appropriately. If reference is set to Traditional then it will be referenced to the channels listed below it. If it is set to Average then those fields will be ignored and it will instead be referenced to the average of the entire dataset (Dien, 1998b).

Finally, the baseline period should be indicated in the Baseline Samples field. In the present case it is being baseline corrected using the first 25 samples (i.e., "1:25" means one through 25). The data were already baseline corrected as part of the blink correction process but the trimmed average could result in it no longer being entirely baseline corrected to zero (if, for example, the original baseline correction included the effects of outlier values that were correctly dropped by the trimming process). Note that the time sample labels and the baseline field will not be affected since one may sometimes wish to use a different set of points for the correction process (in other words, "baseline" has two meanings, the pre-stimulus onset period and the period used to estimate zero voltage).

This step is also performed upon files on the disk rather than on the working set (which is why the preparatory functions of Preprocess, Average, and Postprocess are listed above Read on the Main Pane). Click Postprocess when ready and choose the file generated by the Averaging step. A new file with `_rb` added to the name will be generated.

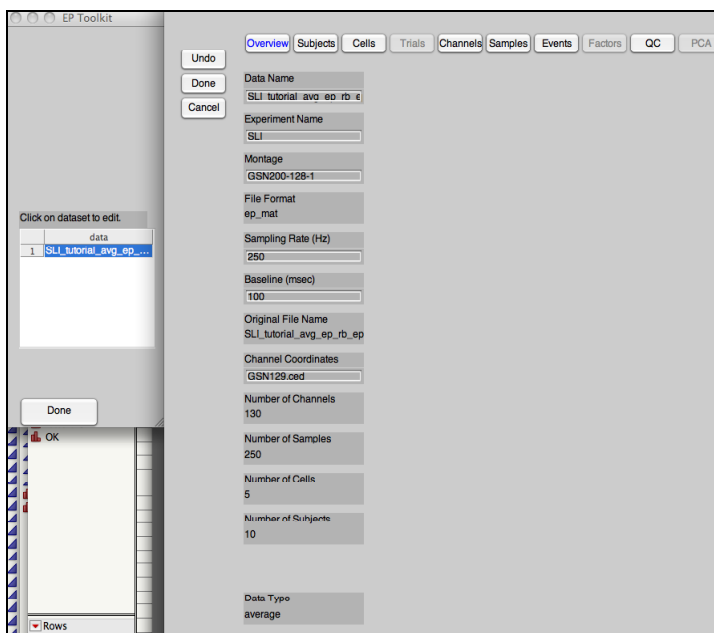
## Read

This function reads files into the EP Toolkit for use with the Edit, View, PCA, Window, and Save functions. It is also used when manually constructing a blink template. See

*Manually Forming a Blink Template under Artifact Correction* above for an example of how to use this function.

## Edit

It is now time to enter the newly processed average file into the working set of data. Click on Main to go back to the Main Pane and then Read. Read in the file as described earlier. Click Done and then click on the Edit button. This will bring up the Edit Window, which allows for editing the data file. If there is only one data file in the working set then it will automatically open, otherwise it will be necessary to click on the name of the one that is to be edited. If you had previously clicked on it and it is still highlighted, it will be necessary (on the Mac) first to command-click it to deselect it so that you can then select it again.



The initial pane of the Edit Window provides the general information about the file. The Data Name is the label by which it is identified in the EP Toolkit. Each data file in the working set needs to have a unique Data Name. The Experiment Name is the name of the overall experiment, rather than the name of the specific file. Take this opportunity to provide an Experiment Name by typing it into the appropriate field. Also, change the baseline field to "100". Note that doing so will not baseline correct the data. In this case, "baseline" refers to the length of the prestimulus period.

If you make a mistake, clicking on Undo will take back the last edit made. Clicking on Undo again will redo the edit. Clicking on Cancel will end the edit session without making any changes to the file. Clicking Done keeps the changes but only in the working copy. The file must still be saved to make them permanent.

Click on the Subjects button to bring up the subject information. You'll want to enter in the correct subject numbers. NetStation tends to lose this information when it exports files into EGIS format. They should be in the order in which they were selected when the averaging step was conducted but it would be best to confirm this.

	order	select	weights	names	type	RunDateMo	Ru
1	1	<input type="checkbox"/>		0 sub001	AVG	7	
2	2	<input type="checkbox"/>		0 sub002	AVG	8	
3	3	<input type="checkbox"/>		0 sub003	AVG	8	
4	4	<input type="checkbox"/>		0 sub005	AVG	7	
5	5	<input type="checkbox"/>		0 sub008	AVG	7	
6	6	<input type="checkbox"/>		0 sub009	AVG	7	
7	7	<input type="checkbox"/>		0 sub010	AVG	6	
8	8	<input type="checkbox"/>		0 sub011	AVG	8	
9	9	<input type="checkbox"/>		0 sub012	AVG	7	
10	10	<input type="checkbox"/>		0 sub013	AVG	8	

Also enter in the sex information as follows:

	yr	RunTimeHr	RunTimeMin	RunTimeSec	SubjID	Handled	Sex
1	17	17	5	0	0	0	F
2	16	36	42	0	0	0	F
3	16	31	39	0	0	0	M
4	15	37	14	0	0	0	F
5	15	38	41	0	0	0	M
6	15	25	35	0	0	0	M
7	10	34	36	0	0	0	F
8	16	10	16	0	0	0	M
9	16	16	17	0	0	0	M
10	15	54	18	0	0	0	M
11							

If using an older MATLAB, the table will look different but can still be used in the same way. Just change one of the column headers to the desired setting and then type in the new values.

we...	na...	type	Sex
0	sub001	AVG	M
0	sub002	AVG	0
0	sub003	AVG	0
0	sub004	AVG	0
0	sub005	AVG	0
0	sub006	AVG	0
0	sub007	AVG	0
0	sub008	AVG	0
0	sub009	AVG	0

This pane can also be used to delete selected subjects, add a grand average, appending more subjects, exporting the data from this pane to a text file, or adding/deleting columns (specs). They can also be reordered by changing the numbers in the order column.

When deleting subjects first click the checkboxes in the select column and then click Delete. The weights are not needed.

To append more subjects, click on Append and then indicate which file is to be added to the present file. It will be assumed that appended file has the same format as the present file. An error message will be generated if the appended file differs too much from the present file.

To add a grand average, indicate in the weights column how much to weigh each subject and then click Add. The checkboxes are irrelevant. When using the weights to generate combinations, a running total of the weights will be maintained on the left side of the subpane. **Note that any weights input into the pane will not take effect until one has clicked elsewhere or pressed enter.** Keep an eye on the running total to verify that each weight value you enter has been recognized. If the weights sum to zero then they will be weighted exactly as entered (e.g., a -2 and a 1 and a 1 one would result in the first subject being doubled and then subtracted from the sum of the other two). If the weights do not sum to zero (e.g., a 1 and a 1 and a 1) then they will be divided by the total weights to provide a mean value (e.g., the total in this case would be  $1+1+1=3$  and so the three subjects' data would be added together and then divided by 3).

The Cells Pane lists the cells in the data. One can delete, reorder, rename, add combined cells, append more cells, or export the information.

Generate a da7+-da9+ difference wave by entering a weight of 1 by the first and a weight of -1 to the second.

	order	select	weights	names	types
1	1	<input type="checkbox"/>	1	da7+	SGL
2	2	<input type="checkbox"/>	-1	da9+	SGL
3	3	<input type="checkbox"/>	0	ga11	SGL
4	4	<input type="checkbox"/>	0	ga13	SGL



Then click Add. Note that the new condition is marked as having the type CMB to denote that it is a combined cell and is therefore not to be treated as a separate cell in the succeeding analyses (it will be ignored when doing PCAs or windowing data for example).

	order	select	weights	names	types
1	1	<input type="checkbox"/>		0 da7+	SGL ▼
2	2	<input type="checkbox"/>		0 da9+	SGL ▼
3	3	<input type="checkbox"/>		0 ga11	SGL ▼
4	4	<input type="checkbox"/>		0 ga13	SGL ▼
5	5	<input type="checkbox"/>		0 +da7+-da9+	CMB ▼

Finally, rename the resulting new condition. Note that conditions must have unique names. In a session file, renaming a cell to have the same name as another cell results in their being combined.

	order	select	weights	names	types
1	1	<input type="checkbox"/>		0 da7+	SGL ▼
2	2	<input type="checkbox"/>		0 da9+	SGL ▼
3	3	<input type="checkbox"/>		0 ga11	SGL ▼
4	4	<input type="checkbox"/>		0 ga13	SGL ▼
5	5	<input type="checkbox"/>		0 da7+-da9+	CMB ▼

The Trials Pane is available only for session files and so its button is grayed out.

The Channels Pane provides similar functionality for channels. One can, for example, generate regional channel averages in the same manner.

123	123	<input type="checkbox"/>	0 e123	EEG
124	124	<input type="checkbox"/>	0 e124	EEG
125	125	<input type="checkbox"/>	0 e125	EEG
126	126	<input type="checkbox"/>	0 e126	EEG
127	127	<input type="checkbox"/>	0 e127	EEG
128	128	<input type="checkbox"/>	0 e128	EEG
129	129	<input type="checkbox"/>	0 Cz	EEG
130	130	<input type="checkbox"/>	0 combined	REG
131	131	<input type="checkbox"/>	0 FidNz	FID
132	132	<input type="checkbox"/>	0 FidT9	FID
133	133	<input type="checkbox"/>	0 FidT10	FID

The Samples Pane provides some control over the temporal aspects of the data, including a button (1/2 Rate) to halve the sampling rate. Doing so also reduces the data itself by averaging consecutive pairs of time samples (e.g., 1&2, 3&4, etc.). This option is useful when the data was sampled at a higher rate than is now needed and reducing the size of the data file would be useful.

The Events Pane provides a listing of the events in the file. The EGIS file format (that was used prior to translating the data over into EP format) does not represent events and so nothing is listed.

The Factors and the PCA buttons are only for factor data and so are grayed out.

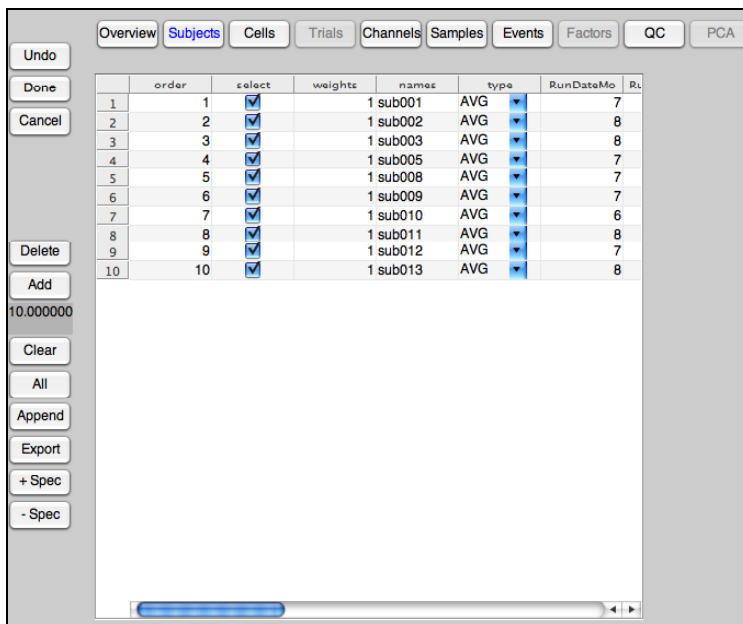
The QC Pane provides quality control information by which problematic data can be detected. This information is generally only available if the EP Toolkit was used for the averaging process and the EP file format was used to store the results.

	names	type	da7+	da9+	gal1	gal3	d
1	sub001	AVG	31	33	36	41	
2	sub002	AVG	27	22	25	17	
3	sub003	AVG	28	30	31	29	
4	sub005	AVG	39	39	38	38	
5	sub008	AVG	44	42	45	43	
6	sub009	AVG	42	46	44	44	
7	sub010	AVG	39	37	40	38	
8	sub011	AVG	34	36	33	29	
9	sub012	AVG	32	31	33	31	
10	sub013	AVG	34	35	34	36	

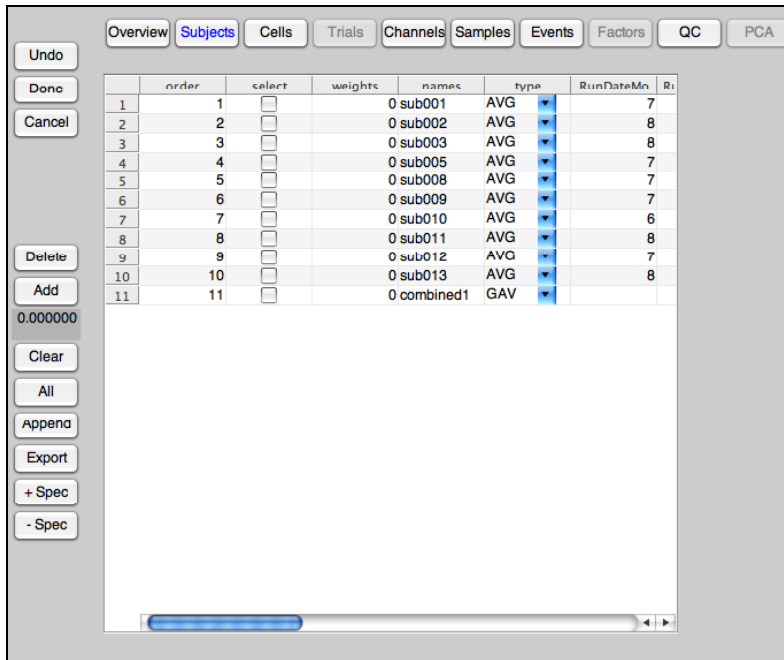
The first subpane (accessed by the buttons at the bottom of the pane) provides the number

of trials going into each cell. The Subs Subpane provides the number of subjects going into each average. The Blinks Subpane indicates the proportion of trials that contained blinks, according to the artifact correction step. The Move Subpane does likewise for movement artifacts. The BadTrials Subpane provides the proportion of bad trials in the cell. The BadChans Subpane provides the same information on uncorrected bad channels (at this point in the tutorial, there are no such bad uncorrected bad channels so the button is invisible). The RepChans Subpane provides this information for corrected bad channels. The Noise Subpane provides the magnitude of the noise in the cell, as estimated by the +/- reference (Schimmel, 1967) in which every other trial is inverted during averaging, resulting in an average in which the ERP has been cancelled out, leaving only noise. The numbers in the table are the root mean square (RMS) measures across all the channels and time points in those cells (meaning that the numbers were squared to eliminate negative signs, the mean was taken of the resulting squared numbers, and then the square root was taken of this final number in order to get back to the original metric, without any minus signs - basically just an elaborate way to get the average of the absolute values). Thus, larger numbers are higher levels of noise. The Std Subpane provide another measure of noise, the standard deviation of the time points obtained during the averaging process. Again, larger numbers are higher levels of noise. Outliers on these measures should be examined closely to determine if they are problematic. The numbers provided by these QC measures can allow for formal empirical criteria to be formulated and reported if it proves necessary to exclude bad subject data.

If the subjects prove to be acceptable, then a grand average can be generated. To do so, go to the Subjects Pane and click on All. This will give all the subjects a weight of one.



Then click on Add. The new entry is marked as a GAV type to keep it distinct from the subject averages which are AVG type. Rename it "grand"



## View Data

We can now proceed to view the data. Click on Done to return to the Read Pane of the Main Window (and to keep the changes you've made). Then click on Done to return to the Main Pane. Click on View to shift to the View EEG Pane. This pane allows one to view the waveforms directly. Four colors are available corresponding to four waves in the plots. For each color, there are controls indicating the data in the active set, the cell, the subject, the trial, and the factor (where appropriate). In this case, set the controls to show the grand average for the four experimental cells.

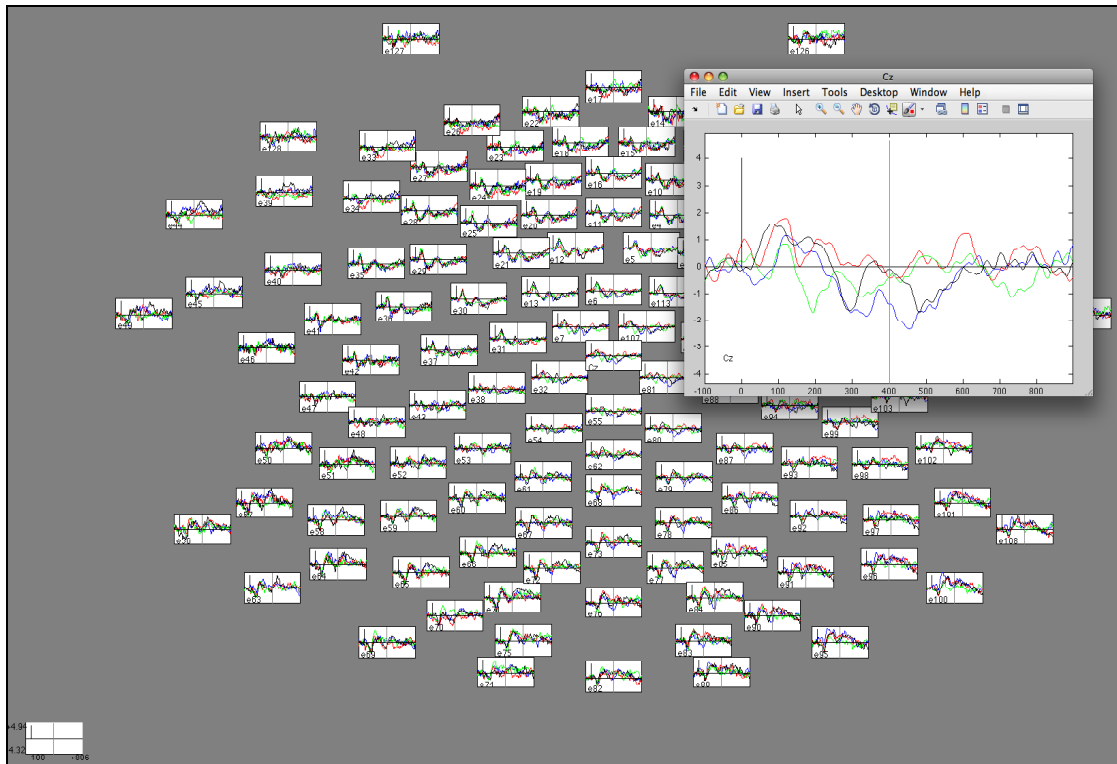
The 'View EEG' window displays four data series, each with its own set of controls:

- Blue Series:** SLI\_POA\_avg\_rb, da7+, combined1, No Trials, No Factors.
- Red Series:** SLI\_POA\_avg\_rb, da9+, combined1, No Trials, No Factors.
- Green Series:** SLI\_POA\_avg\_rb, ga11, combined1, No Trials, No Factors.
- Grey Series:** SLI\_POA\_avg\_rb, ga13, combined1, No Trials, No Factors.

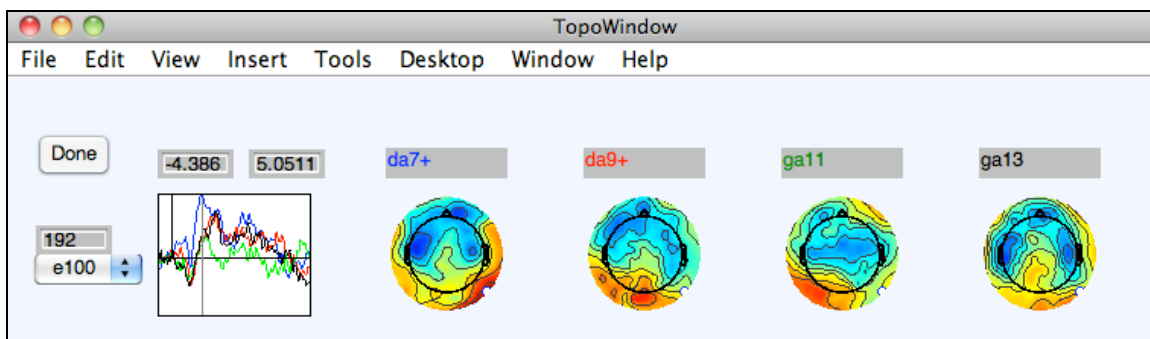
At the bottom, there are three tabs: 'Waves', 'Topos', and 'Main'. Below the tabs is a table for voltage and epoch settings:

Voltage	Epoch	Markers
5.05	-100	
-4.39	900	

There are two options for viewing the data, Waves and Topos. The controls within the box apply only to the Waves option. Note that the maximum and minimum voltages across all four colors is listed at the bottom under Voltages and these numbers are updated as the settings are changed. These provide the scaling for the waveform plots once Waves is clicked. One can also change these settings manually if desired. The Epoch numbers show the ms to be displayed, representing the time points shared in common if different datasets with different sized epochs are to be displayed together. The marker fields allow for optional marker lines to be included on the waveforms. Enter 400 to indicate that one should be placed at 400 ms. Then click Waves. After a pause the waveforms should be displayed in a new window. An expanded view of a waveform can be obtained by clicking on it.



The Topos option provides an alternative viewing format in which the waveforms from one representative channel plus the scalp topography maps are provided. The settings in the box are ignored for this option. Whatever dataset is in the first color (blue) is used as the reference dataset. The waveforms for each line will be from the channel with the largest absolute voltages (in the reference dataset) and each topography plot will be from the time point with the largest absolute voltages (in the reference dataset). The gray line in the waveform figure indicates the time point corresponding to the topographical plots and the white dot in the topographical plots indicates the electrode corresponding to the waveform figure. You can change which channel and time point is being displayed using the controls on the left side. The voltage range of the figures can be set using the controls just above the waveform figure. If desired, you can obtain a color bar via the insert menu, although it will shrink one of the topographical plots. Press the Done button to return control to the View pane.



## Principal Components Analysis (PCA)

The next step is to obtain a PCA of the data. I typically recommend using a two-step sequential PCA (Spencer, Dien, & Donchin, 1999; Spencer, Dien, & Donchin, 2001) in which the first step is a temporal Promax rotation and the second step is a spatial Infomax (ICA) rotation {Dien, 2010, #6222}.

First click on Main to return to the Main Pane. Then click on PCA. I generally recommend using Promax with a covariance relationship matrix (Kayser & Tenke, 2003) and Kaiser weighting (Dien, Beal, & Berg, 2005) for this first step. The kappa for the Promax is generally set at 3 but does not make much difference {Dien, 2010, #6222}. You will need to determine how many factors to retain for the initial step. Unfortunately, there question of how to best proceed remains unsettled at this point. Indeed, one recent paper concluded it is best to keep all the factors (Kayser & Tenke, 2003). At this point I find the data insufficient to convince me to change my analysis procedures, given the potential drawbacks of such an approach such as multiple comparison problems (Dien, 2006), but agree that it is a reasonable position and the data provided are intriguing. I therefore consider the question open and in need of further investigation. Leave the # Factors as zero to indicate that a Scree chart is needed. Click on the name of the dataset to begin.

PCA

Mode  
temporal

Rotation  
Promax

Rotation Parameter  
3

Relationship matrix  
covariance

Loading Weighting  
Kaiser

☐ Parametric Analysis

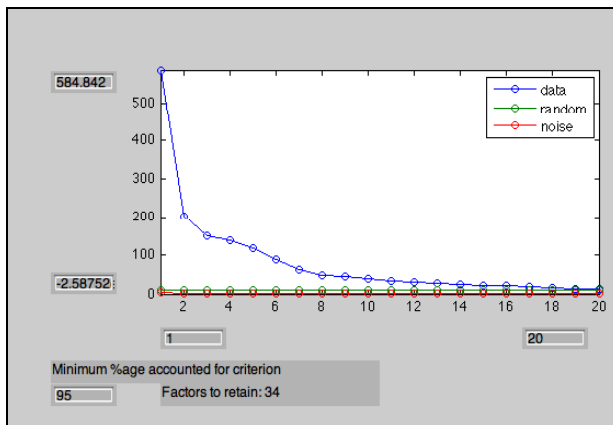
# Factors (n=scree)  
0

Title of PCA

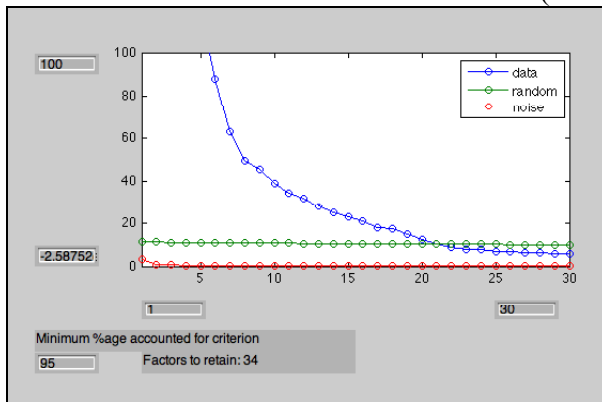
	data
1	SLI_tutorial_avg_ep_...

Done

A Scree plot (Cattell, 1966; Cattell & Jaspers, 1967) is now displayed. The normal criterion is to choose the elbow but unfortunately, at least with ERP data, there are typically multiple elbows. The initial plot suggests 8 factors, for example, but if one expands the scaling then other elbows appear.



For this reason, I currently suggest the use of a parallel test (Horn, 1965), which compares the Scree of the dataset to that obtained from a fully random dataset. Adjust the scaling (on the left) and the number of points (below) to reveal the point at which the green line intersects with the blue line. In this case, the intersection point is 21 factors and so 20 factors is the number indicated (the last point above the line).



The red line represents the Scree plot of the noise data produced by the +/- reference (Schimmel, 1967). While I once recommended using it as the comparison Scree (Dien, 1998a), I no longer recommend its use as an effective PCA needs to represent coherent noise sources as well as the ERP components and so the random noise dataset is a more appropriate comparison.

For completeness sake, the minimum %age accounted criterion is also available. Once finished, use the red close button in the upper left hand corner.

Set the # of factors to 20 and provide a title for the PCA, as in "SLItemp" and then click on the name of the dataset. It will first announce that it is computing jack-knife PCA Loadings. Basically, it is computing a set of PCAs in which each of the subjects in turn is let out of the sample, in order that one might later on determine how stable the PCA solution is to changes in the composition of the sample (see the Topos pane section of the tutorial for a further explanation). The PCA solution should be shortly added to the list of active datasets. It will have a star in front of the name to denote that it hasn't been saved yet.



PCA

Mode: temporal

Rotation: Promax

Rotation Parameter: 3

Relationship matrix: covariance

Loading Weighting: Kaiser

☐ Parametric Analysis

# Factors (0=scree): 20

Title of PCA: SLItemp

	data
1	SLI_POA_avg_rb
2	*SLItemp

Done

Next, set it up for the second step by changing the settings to that of a spatial Infomax rotation. For Infomax, the rotation parameter, relationship matrix, and loading weighting does not apply. Make sure the number of factors is set at zero. Now click on the temporal PCA dataset.

PCA

Mode  
spatial

Rotation  
Infomax

Rotation Parameter  
3

Relationship matrix  
covariance

Loading Weighting  
Kaiser

☐ Parametric Analysis

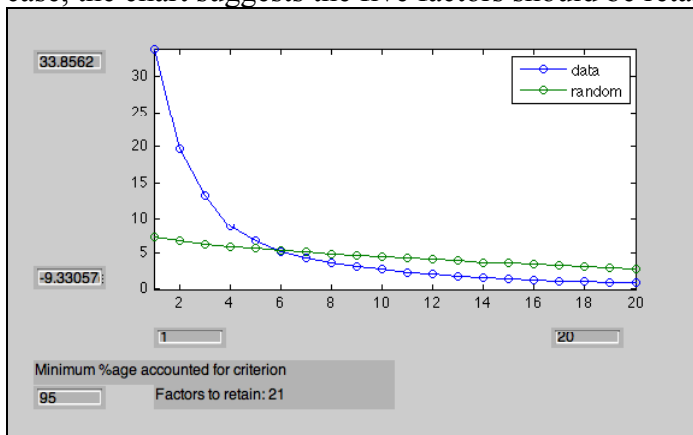
# Factors (0=scree)  
20

Title of PCA

	data
1	SLI_POA_avg_rb
2	*SLItmp

Done

The EP Toolkit generates a separate spatial PCA for each individual temporal factor, based on the argument that, for a given temporal factor, the other factors have nothing to contribute and can only distort the results (Dien, Spencer, & Donchin, 2003). This means that in the current example twenty separate spatial PCAs are run. Currently, for simplicity's sake the EP Toolkit retains the same number of factors for each of the spatial PCAs. The scree chart is the average of all twenty scree charts. This is obviously an oversimplification and the intention is to address this issue in a future update. In this case, the chart suggests the five factors should be retained in this second step.



Enter five for the # of factors and a title for this temporospatial PCA, such as SLIts and then click on the name of the temporal PCA.

At this point there was a memory fragmentation message:

??? Error using ==> zeros

Out of memory. Type HELP MEMORY for your options.

Error in ==> ep\_PCAoutput at 434

```
reconFacs = zeros(numChans, numPoints, numOutputCells, numSubs, NUM_FAC);
```

Error in ==> ep>pickPCAdat at 3715

```
[PCAoutput, peakLatency, peakSamp, peakChan] = ep_PCAoutput(FactorResults,  
cnames, cellcoll,  
exclChan, parametricData.data, parametricData.colheaders);
```

??? Error while evaluating uitable CellSelectionCallback

I therefore quit out of MATLAB without bothering to quit out of the EP Toolkit itself by going to MATLAB's File menu and choosing Quit. I then changed the current directory back to the one I had been using and typed ep at the command line. The EPwork directory retained the datasets that were being analyzed and so all that was needed was to go back to the PCA Pane and try the spatial step again as if nothing had happened. This time it worked fine. To purge a dataset from the working set, return to the Read Pane and click on the name of the dataset.

PCA

Mode  
spatial

Rotation  
Infomax

Rotation Parameter  
3

Relationship matrix  
covariance

Loading Weighting  
Kaiser

☐ Parametric Analysis

# Factors (0=scree)  
5

Title of PCA  
SLIts

	data
1	SLI_POA_avg_rb
2	*SLItmp
3	*SLIts

Done

If you click on Main and then use the Edit function to examine the temporo-spatial PCA data, you will find that the Factors and the PCA panes are now active. The PCA Pane for the temporo-spatial PCA shows the basic results for the PCA. The Summary subpane displays the basic parameters of the PCA. Note how the second step accounted for less total variance (64% vs. 90%) since additional noise variance was dropped from each of the spatial PCAs conducted on each of the temporal factors. The variance accounted for by each spatial PCA of each temporal PCA is provided in the table. Factor Variance is how much of the temporal factor was retained and Total Var is how much of the total original variance was retained by that second step.

The factor matrices are also available for inspection. For example, the FacPat Subpane displays the factor pattern matrix (one type of factor loading) from the initial temporal PCA step. The FacPatST displays it from the second spatial step. FacStr is the factor structure matrix, FacScr is the factor scores, FacCof is the factor scoring coefficients, and FacCor is the factor correlations. I don't have time to provide tutorials on factor analysis by e-mail but you're welcome to read my tutorial chapter (Dien & Frishkoff, 2005) or to consult books on factor analysis (Gorsuch, 1983; Tabachnick & Fidell, 1989).

Undo

Done

Cancel

Overview

Subjects

Cells

Trials

Channels

Samples

Events

Factors

QC

PCA

First Rotation

Mode

temporal

Rotation

Promax

Matrix Type

Covariance

Loading Weighting

Kaiser

Rotation Option

3

Factors Retained

20

Total Variance

0.90292

Second Rotation

Mode

spatial

Rotation

Infomax

Matrix Type

Covariance

Loading Weighting

None

Rotation Option

3

Factors Retained

5

Total Variance

0.6395

	Factor Var	Total Var
1	0.7564	0.0791
2	0.7371	0.0744
3	0.7307	0.0614
4	0.7032	0.0525
5	0.7037	0.0427
6	0.7002	0.0405
7	0.6340	0.0287
8	0.7027	0.0311
9	0.6568	0.0272
10	0.7296	0.0278
11	0.7242	0.0266

Export

Summary

FacPat

FacStr

FacScr

FacCof

FacCor

FacPatST

FacStrST

FacScrST

FacCofST

FacCorST

The factors pane provides information on the individual factors. Peaklatency is the time point with the greatest absolute voltage (out of all the cells after computing the grand average). PeakChannel is the channel with the greatest absolute voltage (out of all the cells after computing the grand average). PeakPolarity is whether the voltage (of the peak latency) was positive or negative. facVar is the variance accounted for by the factor. facVarQ is the variance that is uniquely accounted for by the factor. Notice that the factors from a two-step PCA are organized with the factors derived from the same first step factor clumped together (so the first five of the example are the five spatial factors derived from the first temporal factor).

Overview

Subjects

Cells

Trials

Channels

Samples

Events

Factors

QC

PCA

Undo

Done

Cancel

Delete

Add

0.000000

Clear

All

Append

Export

	names	type	peakLatency	peakChan	peakPolarity	facVar
1	TF1SF1	SGL	304	e65	-	0.040
2	TF1SF2	SGL	304	e7	-	0.012
3	TF1SF3	SGL	304	e49	+	0.010
4	TF1SF4	SGL	304	e125	+	0.006
5	TF1SF5	SGL	304	e110	+	0.005
6	TF2SF1	SGL	388	e75	-	0.027
7	TF2SF2	SGL	388	e33	-	0.020
8	TF2SF3	SGL	388	e48	-	0.010
9	TF2SF4	SGL	388	e115	-	0.007
10	TF2SF5	SGL	388	e49	-	0.005
11	TF3SF1	SGL	212	e58	-	0.029
12	TF3SF2	SGL	212	e127	+	0.013
13	TF3SF3	SGL	212	e116	+	0.006
14	TF3SF4	SGL	212	e74	-	0.005
15	TF3SF5	SGL	212	e100	-	0.004
16	TF4SF1	SGL	504	e57	-	0.021
17	TF4SF2	SGL	504	e45	+	0.010
18	TF4SF3	SGL	504	e127	-	0.006
19	TF4SF4	SGL	504	e116	+	0.004
20	TF4SF5	SGL	504	e125	+	0.003
21	TF5SF1	SGL	120	e75	+	0.015
22	TF5SF2	SGL	120	e79	+	0.010
23	TF5SF3	SGL	120	e110	-	0.007
24	TF5SF4	SGL	120	e46	-	0.005
25	TF5SF5	SGL	120	e121	+	0.003
26	TF6SF1	SGL	576	e39	-	0.015
27	TF6SF2	SGL	576	e83	-	0.010

One can also view the factors directly by going to the View Pane from the Main Pane of the Main Window. One could, for example, examine what portion of the grand average is accounted for by TFSF1 (the first spatial factor formed from the first temporal factor) by overlaying them using the Waves function.

View EEG

SLI\_POA\_avg...

da7+

combined1

No Trials

No Factors

SLI\_POA\_avg...

da9+

combined1

No Trials

No Factors

SLits

da7+

grand average

No Trials

TF01SF1

SLits

da9+

grand average

No Trials

TF01SF1

Voltage

Epoch

Markers

5.05

-100

-4.39

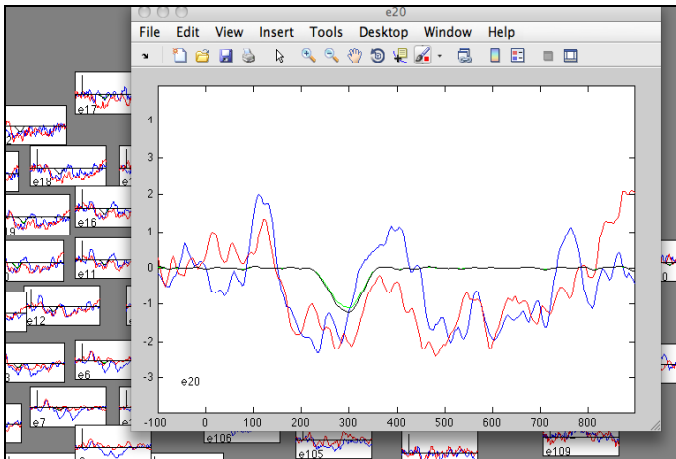
900

Waves

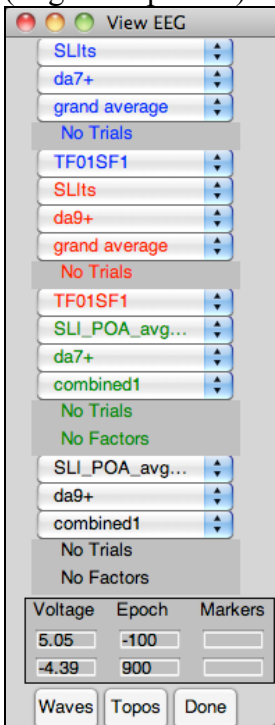
Topos

Done

An expanded view of channel 20 (double click on it) reveals:

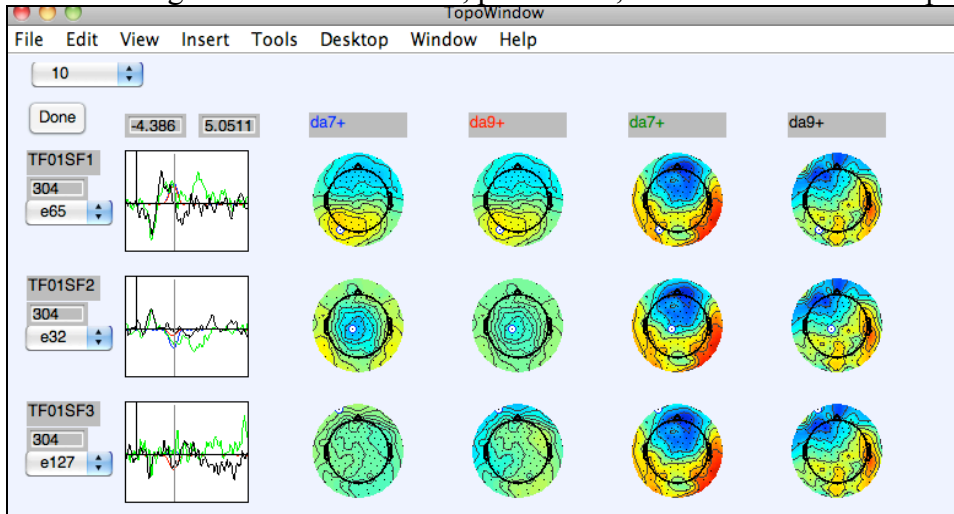


Alternatively, you can use the Topos function to examine the results. If a PCA dataset is examined, instead of only displaying the chosen factor, all of the factors will be displayed, each on a separate line. It is therefore an efficient method for screening for factors of interest. If the first (blue) dataset is the PCA dataset then it will serve as the reference dataset and all the channels and time points will be the factor's most representative (largest amplitude) channel and timepoint.



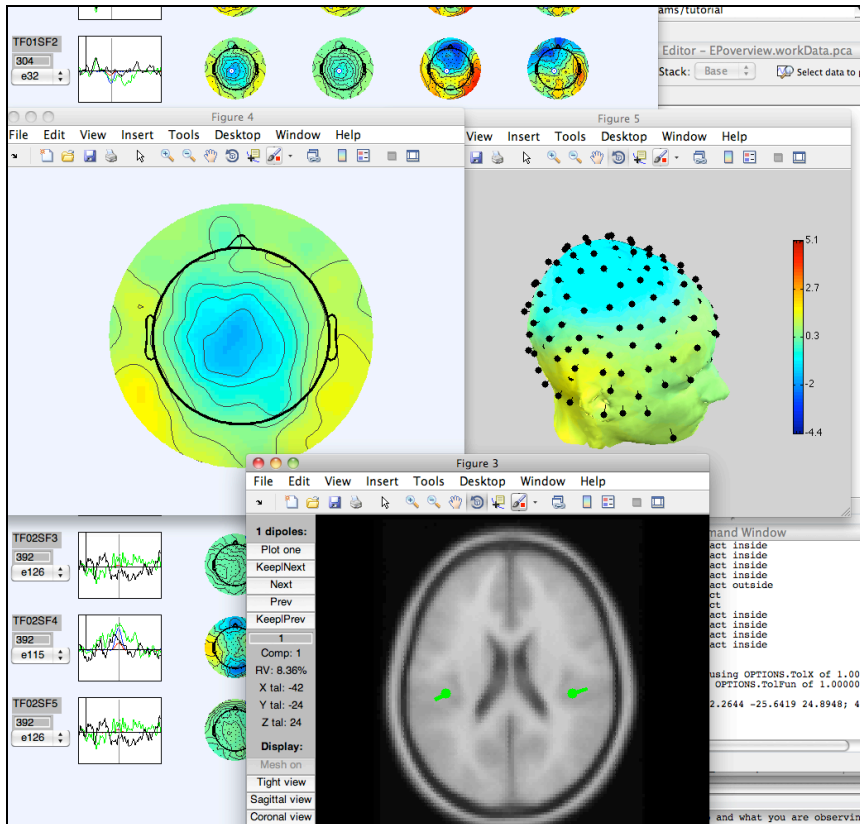
Press the Topos button. Plotting so much data will take some time so be patient. Only ten factors will be displayed on a page. The control at the top of the screen allows one to move between the pages of factors. In the following screenshot, TF01SF01 (the first spatial factor derived from the first temporal factor) had a peak latency of 304 ms and a peak channel at e65. A vertical line in the waveform figure indicates this peak time point, corresponding to the scalp topography maps. A white dot in the scalp topography map indicates the peak channel, corresponding to the waveform figure. You may change

the channel being displayed, either using the pop-up menu or by clicking directly on a scalp topography map. Black dots indicate the locations of the other channels but your click doesn't have to land precisely on the desired electrode location. You may change the voltage range of both the waveform figure and of the scalp topography maps by changing the numbers in the two boxes above the waveform figures. Note that you will need to change the number in a box, press enter, and then wait until the page is updated.

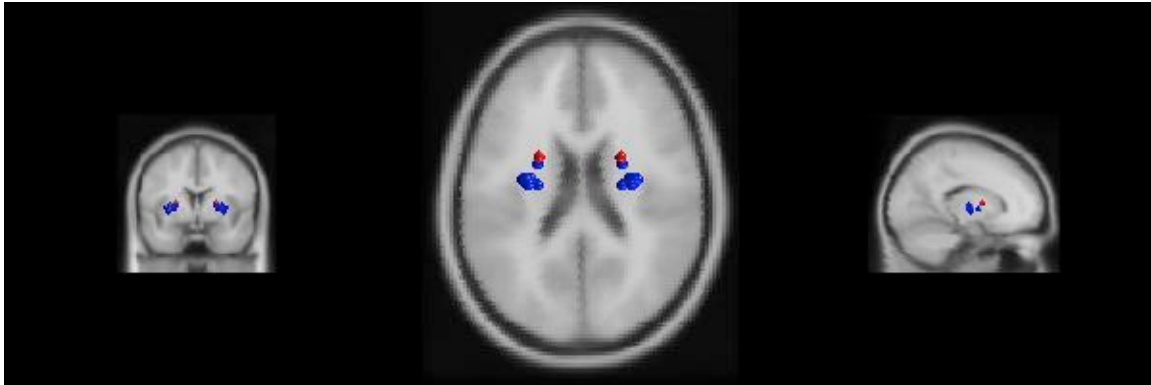


You may also right-click on the scalp topography plots to obtain: 1) an expanded 2D figure, 2) an expanded 3D head that may be rotated as needed, 3) simple dipole analysis, 4) simple dipole analysis of jack-knife PCA results. The former two are courtesy of EEGlab functions and the latter two are courtesy of a FieldTrip function (using an EEGlab function to display the results). The first time you use the 3D head, EEGlab will need to generate a "spline file" which it will automatically place in the same location as the .ced file. You will need to wait while this is done. It only needs to be done once per .ced file that you are using. The dipole analysis is a bit trickier. In order to allow the function to work on multiple platforms, FieldTrip compiles a binary file on the fly. To do so, it needs a gcc compiler to be available (see installation instructions). If this was not done, then this function will not work. The dipole analysis program will attempt to fit two symmetric hemispheric dipoles. If they end up running into each other then it will try again with only one midline dipole. Note that this is a very simple implementation of dipole source analysis and does not seem to provide results as good as that provided by the more refined BESA program. In particular, it seems to have problems with the two hemispheric dipoles running into each other and then exploding into arbitrarily high amplitudes as they interact with each other. BESA provides parameters which help keep the dipoles from doing this. Nonetheless, this implementation of source analysis can be suggestive, especially when the dipoles have not collided. The figure below shows the results from TF01SF01.





With the jack-knife option, what happens is that every time one does a PCA, a jack-knife analysis is automatically conducted in which a set of PCAs are computed in which one subject is left out (so twenty subjects results in twenty PCAs, each with nineteen of the subjects). For two-step PCAs, the jack-knife analysis is only conducted on the second step. The factor loadings from this are stored for later use. When the "jack-knife" option is chosen in the Topos function, a simple dipole analysis is conducted on each of the jack-knife solutions and the resulting cluster of solutions is generated, with the original location in red and the jack-knife solutions in blue. This provides a graphical representation of how stable the original solution was to individual subject variability (Foti, Weinberg, Dien, & Hajcak, submitted). Since this procedure is conducted on the factor loadings, it will only be done for a spatial PCA or for a temporo-spatial PCA. In addition, the jack-knife PCA procedure allows for a significance test to be conducted on the relative strength of the hemispheric dipole amplitudes, using a t-test that has been adapted for jack-knife analyses (Miller, Patterson, Ulrich, 1998), which is then printed out at the command line. Unfortunately, the corresponding test statistic is not yet available using robust statistics.



## Windowing

To test whether there are any statistically significant effects, go to the Window Pane of the Main Window. Let us first consider how subject averages would be analyzed and then address how to analyze PCA results.

Window Data

Dataset: SLI\_tutorial\_av...

Measure: mean

samples: 1 1 ms: -96 -96

Baseline(ms): 100

Channel Groups: none

Channels

Factor: none

	outCells	inCells
1	da7+	da7+
2	da9+	da9+
3	ga11	ga11
4	ga13	ga13

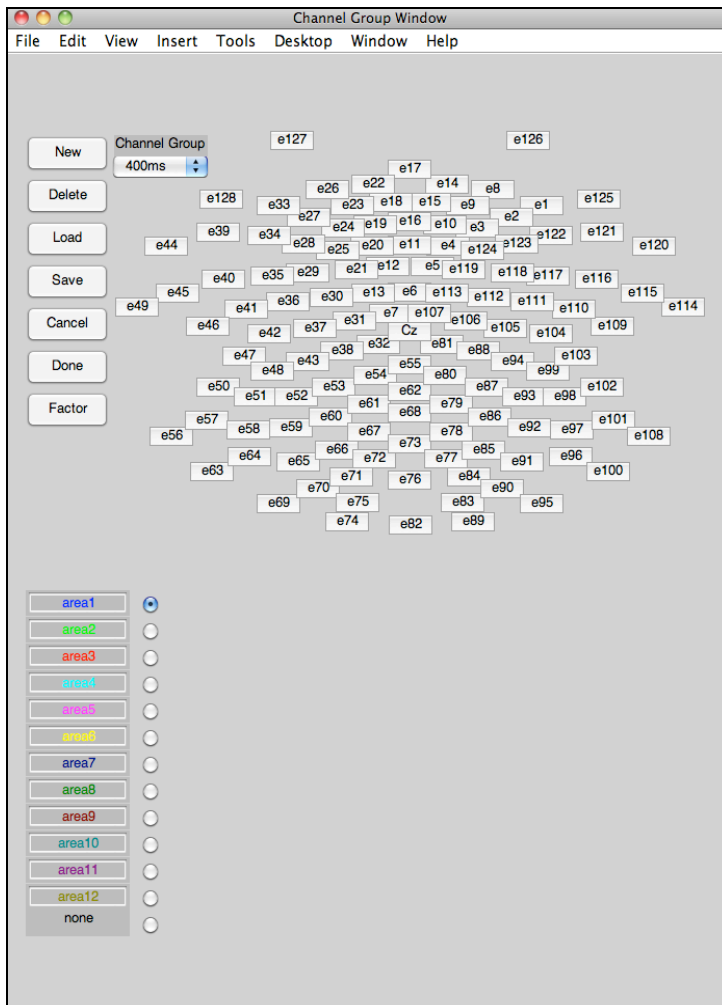
	Select	Subject Specs
1	<input type="checkbox"/>	RunDateMo
2	<input type="checkbox"/>	RunDateDay
3	<input type="checkbox"/>	RunDateYr
4	<input type="checkbox"/>	RunTimeHr

Window AutoPCA Done

### *Windowing Subject Average Data*

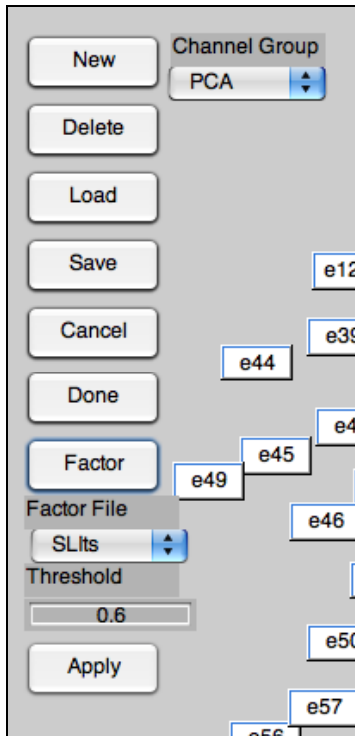
First you need to indicate which of the datasets in the active set to analyze. Then which measure to use within the window. Min and max centroid (Dien, Spencer, & Donchin, 2004) are a way of measuring the latency of the overall area under the curve rather than just the peak, much as a mean measure does for amplitude. Use minimum centroid for positive components and maximum centroid for negative components. The time window also needs to be specified. Let's try a 300-500 ms window (although this tutorial dataset is too small to look for real effects) by entering in samples 100 to 150.

Next, the channels are specified by click on the Channels button. It will then ask for a name for the new electrode grouping. Type in 400ms. It will then bring up the electrode montage.

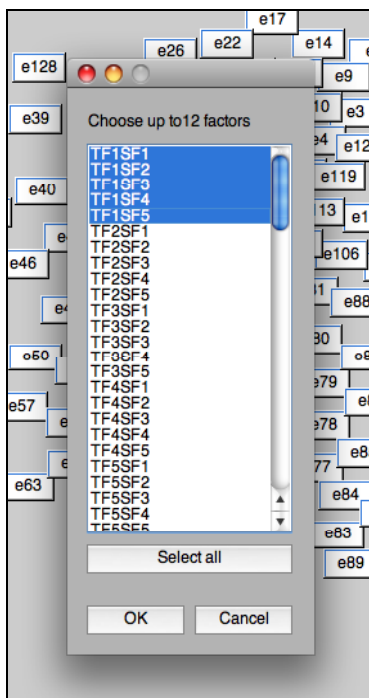


Click on the channels for the first electrode area. You should also give it a name by typing in the "area1" box, say Fz. Let's also add a posterior grouping around Pz by typing in Pz for area 2 and then clicking on the green box to activate the second area. Having clicked on the green box, any channels you click on will be added to the second area (and turned green) rather than the first box (and turned blue).



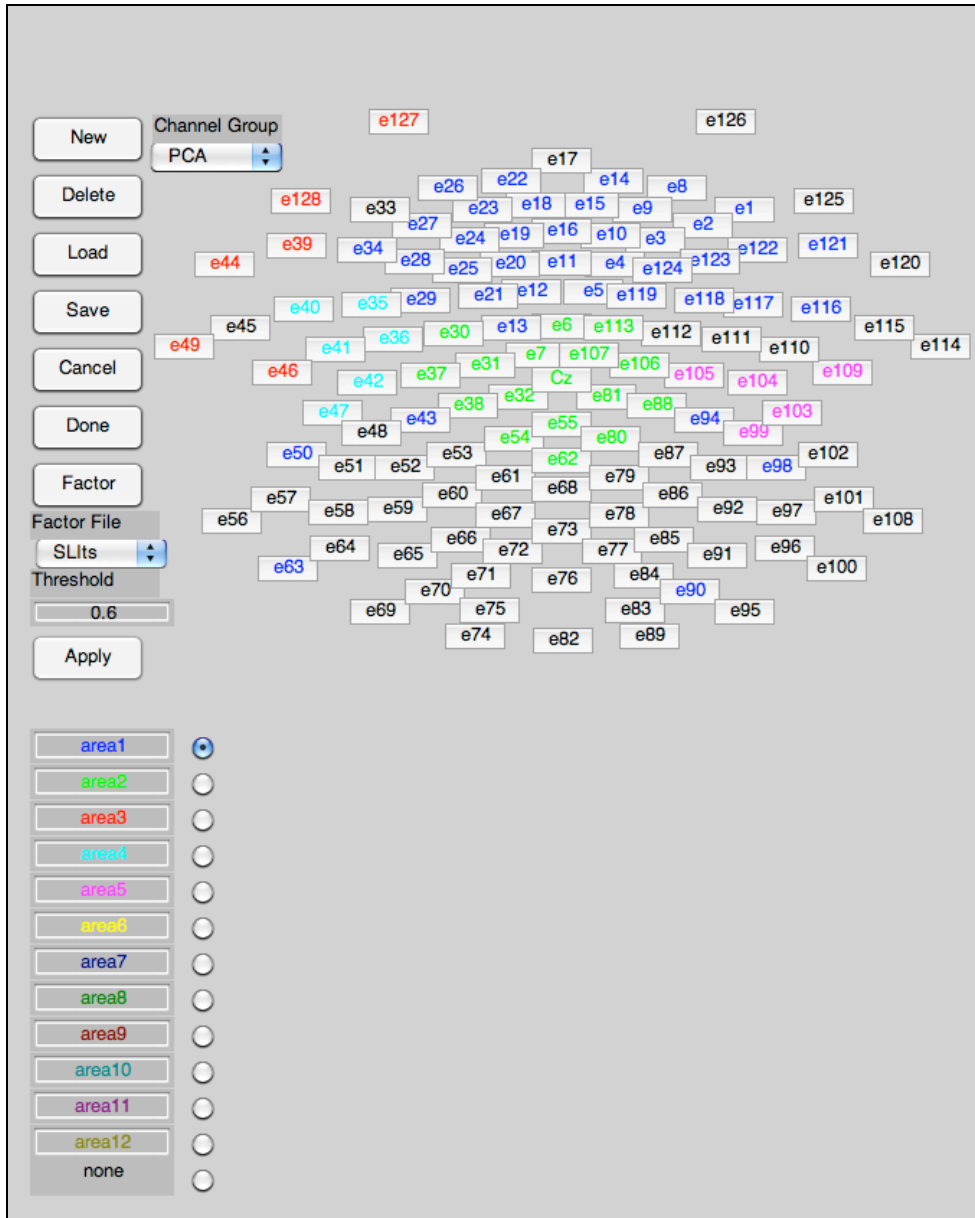


Click on Apply. It will then give you a listing of all the factors, of which you can choose 12. For this example, just choose the first five (on the Mac, hold down the command key or the shift key to choose more than one).



It will now choose label each channel according to which factor had the highest loading, as long as it passed the threshold that you specified (0.6 in this case). If a factor has both

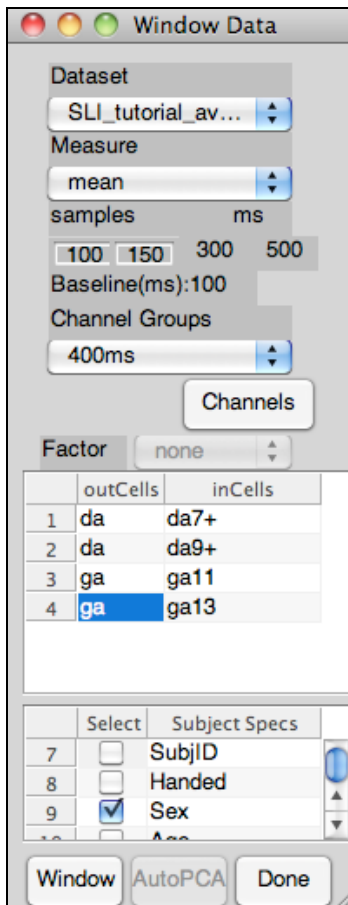
positive and negative loadings that pass the threshold, it will determine which polarity had the highest absolute amplitude and mark only loadings with that sign (otherwise you could end up with a windowed measure with both positive and negative voltages which would then cancel out when averaged together).



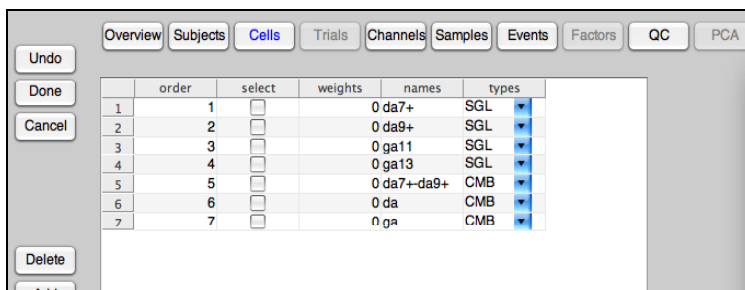
Once done, you can Save these channel groupings for later use and then press Done to go back to the Window Pane of the Main Window. For the tutorial, we'll just use the 400 ms channel grouping.

If you are windowing PCA data, then you will also need to specify the factor, otherwise it will be grayed out.

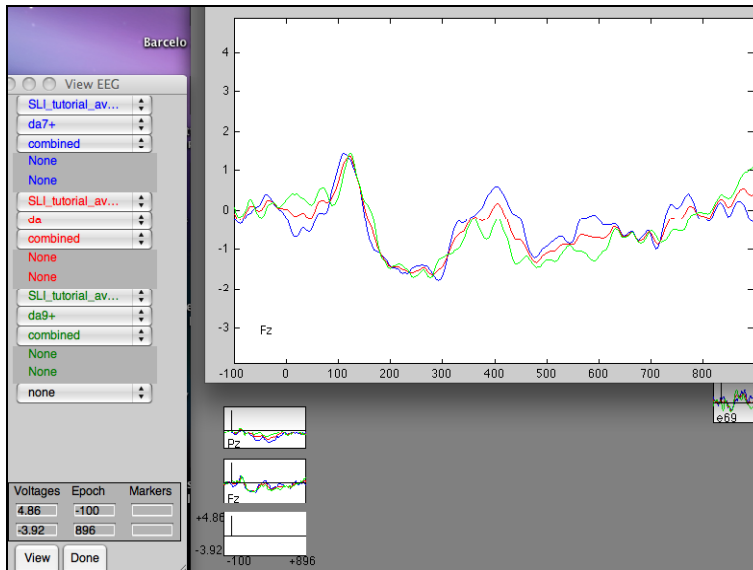
You will also need to specify the cells. It will list all the cells in the dataset (excluding any combined cells that you added). If you want any of them to be combined, rename the outcells accordingly. For this example, let's combine together the two ga cells and the two da cells. Finally, you can specify if you want any of the subject specs (specifications) to be included in the final analysis. Let's click on Sex.



Once ready, click on the Window button. The windowing routine will check to see if the names of the output cells are different from those already in the dataset. If so, then it will add combined cells with those names to the dataset so that you can examine what the waveforms corresponding to the windowed measures looks like. Likewise, if the channel area names are different, then it will add new regional channels to the dataset so that you can again examine them.



If one goes back to the View Pane, the new regional channels show up in the lower left corner, right above the scale figure.



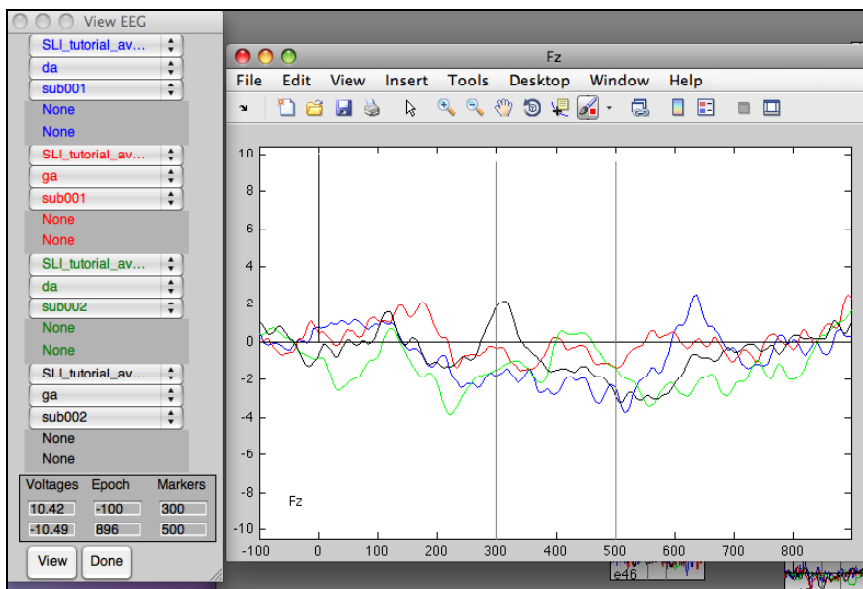
In this example, it can be seen how the red waveform (da) is the average of the blue (da7+) and green (da9+) waveforms.

Turning to the output of the windowing procedure, it generates a tab-delimited text file that can be examined in Excel or similar programs. After some header information it provides the windowed measures in a form suitable for direct use in a statistics program, along with the subjects specs requested (Sex) and the subject names.



New Open Save Print Import Copy Paste Format Undo Redo AutoSum Sort A-Z Sort						
Sheets						
	A	B	C	D	E	F
1	SLI_tutorial_avg_ep_rb					
2	296/496 ms					
3	mean					
4	400ms					
5						
6	da	da	ga	ga	Sex	
7	Fz	Pz	Fz	Pz	spec	
8	-2.298015	0.391224	-1.090105	1.721797	F	sub001
9	0.36576	-0.097886	-0.331143	0.976209	F	sub002
10	-0.813917	-0.125821	0.310715	-0.05587	M	sub003
11	1.880484	-0.38467	1.20445	-2.227916	F	sub005
12	-1.219501	-3.361659	-0.311567	1.046936	M	sub008
13	-0.433242	-0.92818	-1.945169	-2.626063	M	sub009
14	-1.651987	1.447137	-0.267793	0.05043	F	sub010
15	-0.848974	-0.112839	-0.849471	-0.189943	M	sub011
16	-0.694247	0.03444	1.837406	-1.087212	M	sub012
17	-3.197162	-0.149068	-2.727196	0.694948	M	sub013
18						
19						
20						
21						
22						

As verification of the numbers, one can directly compare them with the waveforms. Set markers at 300 ms and 500 ms. See how the waveforms for subjects 1 and 2 correspond to the measures for da and ga in the Fz regional channel average. Note that there is a MATLAB bug that can prevent the baseline, onset, and marker lines from appearing if one clicks directly on the waveform lines. If this happens, just try again but clicking on the white background of the waveform box.



## Windowing PCA Data

Analysis of PCA data is conducted in much the same way. The only difference is that one needs to specify the Factor in addition to everything else. Alternatively, if one is analyzing PCA data, one can click on the AutoPCA button and the Toolkit will run through all the factors, choosing the peak channel and the peak time point for each factor. When using this option, the pane's settings for channel and for window will be ignored. It will only generate a windowed file for each factor whose size (variance accounted for) meets a minimum threshold (set in Preferences, .5% by default) to screen out factors that account for only small clumps of noise. This kind of selectivity is, in turn, helpful when controlling for multiple comparisons as one will not have to be as stringent as if one had included the noise factors. The command line will announce the number of factors that met the criterion.

The output of a PCA is expressed in voltages, corresponding to the voltage accounted for at the chosen time points and channels (the peak time point and channel if the AutoPCA option was chosen). The Toolkit has automatically converted the factor scores into microvolt scaling. One can use the View option to examine the actual factor waveforms, if desired, to see how they correspond.

In order to prepare for the next section, let's redo the cells so that we are not collapsing the ga and da cells together.

Window Data

Dataset: SLIts

Measure: mean

samples: 1 1 ms

Baseline(ms): 100

Channel Groups: 400ms

Channels

Factor: TF1SF1

	outCells	inCells
1	da7+	da7+
2	da9+	da9+
3	ga11	ga11
4	ga13	ga13

	Select	Subject Specs
8	<input type="checkbox"/>	Handed
9	<input checked="" type="checkbox"/>	Sex
10	<input type="checkbox"/>	Age

Window AutoPCA Done

Now run the AutoPCA option. As can be seen in the resulting file for TF1SF1 (temporal factor 1/spatial factor 1), the peak channel is named as being electrode 65 and the peak time point was at 300 ms. The gender is also included as you had clicked on its checkbox.

SLIfac-TF01SF1.txt

Sheets

Charts

SmartArt Graphics

WordArt

	A	B	C	D	E	F	G
1	SLIts						
2	300/300 ms						
3	mean						
4	autoPCA						
5	Factor: 1						
6	da7+	da9+	ga11	ga13	Sex		
7	e65	e65	e65	e65	spec		
8	2.877524	2.865724	-1.409879	-0.03006	F	sub001	
9	0.186353	2.047581	3.916229	-5.566754	F	sub002	
10	0.626646	-0.764784	-0.256201	-0.317984	M	sub003	
11	4.331124	2.902897	2.052944	-1.687073	F	sub005	
12	-4.22317	0.414983	0.432479	-0.688769	M	sub008	
13	-0.76741	-0.655529	-0.038563	1.959017	M	sub009	
14	3.246158	4.710528	7.274915	1.017376	F	sub010	
15	-1.923805	-1.733002	-0.951836	2.134634	M	sub011	
16	8.354988	-3.984057	-3.338107	-5.702075	M	sub012	
17	7.009011	10.903857	7.229362	8.059335	M	sub013	
18							
19							
20							
21							
22							
23							
24							
25							

## Robust ANOVA

### *About Robust Statistics*

The robust statistics function generates inferential statistical tests comparable to ANOVAs that are designed to be more robust against violations of statistical assumptions. The statistical routines were translated from the SAS/IML routines posted by Keselman (<http://www.umanitoba.ca/faculties/arts/psychology/>). The routines were fully described in the following publication (Keselman et al., 2003). Also, the accompanying file "robust.pdf" is also included. A front-end has been added that automatically generates the necessary contrast matrices for omnibus ANOVAs and that

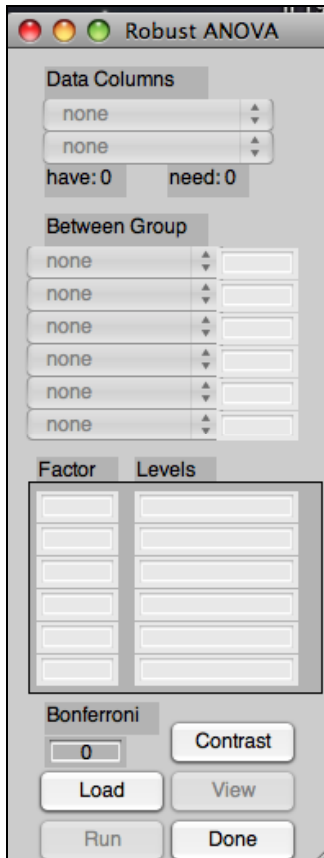
organizes the results in an easy to read manner. Note that the routines take some time to run.

Psychophysiology has recommended the use of these robust statistics (specifically, statistical tests using non-pooled error terms) in its Guidelines to Authors. In brief, this robust statistic has the following three features: 1) trimmed means and winsorized covariances to protect against outliers; 2) bootstrapping routine to estimate the population distribution rather than making the assumption that the data is normally distributed; 3) Welch-James approximate degrees of freedom statistic (resulting sometimes in decimal degrees of freedom) that avoids the assumption that the cells have homogeneous error variances. The latter also makes it unnecessary to use epsilon correction like G-G or H-F since sphericity is not assumed. In practice, it seems to generate results that are largely comparable to normal ANOVAs but that are more robust against violations of assumptions. For a direct comparison of ERP results against univariate and multivariate ANOVAs, see (Dien, Franklin, & May, 2006). A very approachable treatment of this robust statistic is available (Wilcox, 2001). A description of the issues involved in using conventional ANOVAs with ERP data is also available (Dien & Santuzzi, 2005).

The article (Keselman et al., 2003) also provides results of example data analyses. Note that p-value results will differ somewhat from the examples in the article since the random number generators used in the bootstrap routine will differ between programming platforms. Although in principle the bootstrap routine should be robust against changes like that, according to my testing efforts, p-values will only be stable if the number of simulation runs is increased to about 50,000 from the 599 generally used in the published studies.

### ***General Procedures***

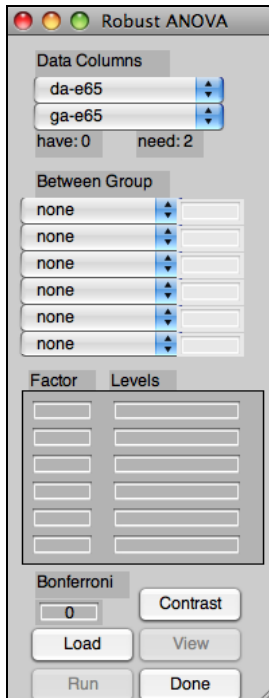
To use the robust statistics, first go back to the Main Pane of the Main Window and then click on the ANOVA button.



The image shows a software dialog box titled "Robust ANOVA". It contains several sections for configuring an ANOVA test:

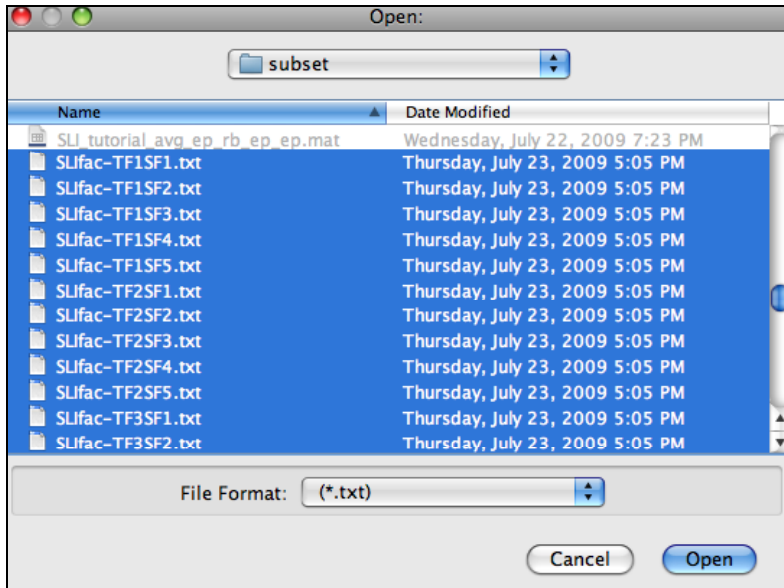
- Data Columns:** Two dropdown menus, both currently set to "none". Below them, it shows "have: 0" and "need: 0".
- Between Group:** A section with six rows, each containing a dropdown menu (all set to "none") and an adjacent empty text input field.
- Factor Levels:** A table with two columns, "Factor" and "Levels". Each column contains six empty text input fields for defining factors and their levels.
- Bonferroni:** A section with a numeric input field set to "0", a "Contrast" button, and a "Load" button.
- Buttons:** At the bottom, there are "View", "Run", and "Done" buttons.

The ANOVA function is set up to assume that there will be a series of windowed files that have a similar structure (in terms of the cells). The first step is therefore to choose one of these files to serve as the template. Click on the Load button and then select one of the windowed factor files that you just generated with the AutoPCA option.



In this file, as you saw earlier, there were four data columns plus the Sex column. The first two menus are the leftmost and rightmost data columns to be analyzed. By default the Toolkit assumes that you will be analyzing all of the data columns but you can adjust the settings if you wish. Note that it indicates that 4 are needed. It means that the factor structure needs to accommodate four total within group cells. Enter in a POA (place of articulation) factor with a D and a G level and a PHN (phoneme) factor with a Y and an N level (yes a phoneme vs. no, not a phoneme). The lowest factor varies fastest as one goes across the columns. Note that the ANOVA factor name needs to have three letters and the ANOVA factor level names need to be one letter each. Also select Sex as a between group factor by selecting it and provide it with a three letter name, in this case "sex". Between group levels also need to be a single letter. If longer labels are provided, only the first letter will be used. It is therefore important that such labels start with different characters. Finally, enter the number of factors that met the variance criterion, 44, into the Bonferroni field (no need to use the Bonferroni if you are using a priori criteria to choose factors, such as their having the latency of prior interest). The pane should end up looking like the following figure:





The Toolkit will start running a full ANOVA on the data (note, advanced options like ANCOVA and nested designs are not currently available). The output goes into an .html file with the name that you specified. The analysis will take some time, even on a fast computer. You can check on its progress by double-clicking on the html file. It should come up in a browser window that will show its current state. Click on the browser's refresh button periodically to follow its progress, if desired.

Some typical output follows. First it provides the basic parameters. Then each combination of between factor effects is presented. For each between factor effect combination, the within factor effects are provided. The section titled "NO BETWEEN EFFECTS" are the effects without interactions with the between group factors. The effect labeled "NO WITHIN EFFECTS" in the "sex MAIN EFFECT" is, conversely, the main effect of the between group effect of sex (thus having no interaction with the within group factors) and the "POA MAIN EFFECT" that follows it is the interaction with the sex main effect. Each effect lists the statistical test numbers (appropriate for directly copying into a manuscript). If the test reaches one-tailed significance (normally .10) then it is listed in green. If it reaches two-tailed significance but not Bonferroni corrected significance then it is listed in orange. If it reaches two-tailed significance even with the Bonferroni correction then it is listed in red. Finally, it lists the trimmed cell means underlying each of these tests.

```
SLIfac-TF01SF1
WELCH-JAMES APPROXIMATE DF SOLUTION
TRIMMED MEANS & WINSORIZED VARIANCES
PERCENTAGE OF TRIMMING: 0.05
BOOTSTRAP CRITICAL VALUE FOR SINGLE TEST STATISTIC
NUMBER OF BOOTSTRAP SAMPLES: 50000
STARTING SEED: 1000
Number of subjects in each group: 4 6
```



Number trimmed from each end of the groups: 0 0

Uncorrected alpha criteria: 0.05

Corrected alpha criteria: 0.0011364

#####

NO BETWEEN EFFECTS

-----  
POA MAIN EFFECT

TWJt/c(1.0,8.0)=1.47, p=0.26

Averaged Trimmed Cell Means:

D    G

+1.60 +0.51

-----  
PHN MAIN EFFECT

TWJt/c(1.0,8.0)=0.82, p=0.38

Averaged Trimmed Cell Means:

Y    N

+1.49 +0.62

-----  
POA \* PHN INTERACTION EFFECT

TWJt/c(1.0,6.5)=0.70, p=0.41

Averaged Trimmed Cell Means:

DY   DN   GY   GN

+1.62 +1.58 +1.35 -0.34

#####

SEX MAIN EFFECT

-----  
NO WITHIN EFFECTS

TWJt/c(1.0,5.8)=1.88, p=0.24

Averaged Trimmed Cell Means:

F    +0.01

M    +2.10

-----  
POA MAIN EFFECT

TWJt/c(1.0,8.0)=0.01, p=0.91

Averaged Trimmed Cell Means:

D    G

F    +0.50 -0.49

M    +2.70 +1.50

-----  
PHN MAIN EFFECT

TWJt/c(1.0,8.0)=0.13, p=0.73

Averaged Trimmed Cell Means:

	Y	N
F	+0.27	-0.26
M	+2.71	+1.49

-----  
POA \* PHN INTERACTION EFFECT

TWJt/c(1.0,6.5)=0.96, p=0.36

Averaged Trimmed Cell Means:

	DY	DN	GY	GN
F	-0.13	+1.14	+0.67	-1.65
M	+3.38	+2.02	+2.04	+0.96

A further feature of the Toolkit's output is that when an interaction is significant (uncorrected), the Toolkit will provide all the possible follow-up tests (i.e., for a three-way interaction, it will provide all the possible two-way interactions, where one level is held constant). If any of those are significant then a the follow-up tests for that test are provided as well. In the following case, a sex by POA by PHN interaction is significant (uncorrected). It is known that this is an interaction with sex since it appears in the sex main effects section. It then follows up with all the possible two-way ANOVAs. It reports that the POA by PHN interaction is significant (uncorrected) for boys (level M of the sex factor). It then follows that up with the tests of the one-way ANOVAs for that interaction, the first of which is included below. Note that follow-up tests are in a smaller font to help set them off from the main tests. Also, follow-ups are indented deeper than the preceding level of results. Once the follow-ups to a test are completed then the results proceed to the next test. Devising a clearer format for these results tables is planned for a future release but it should suffice for now.

POA \* PHN INTERACTION EFFECT

$T_{WJt}/c(1.0,6.7)=13.47, p=0.025$

Averaged Trimmed Cell Means:

	DY	DN	GY	GN
F	-1.15	-1.29	-0.07	+1.61
M	-1.87	+0.31	-0.81	-4.05

Holding level F of factor sex constant.

#####

NO BETWEEN EFFECTS

-----  
POA \* PHN INTERACTION EFFECT

$T_{WJt}/c(1.0,3.0)=1.44, p=0.26$

Averaged Trimmed Cell Means:

DY DN GY GN  
-1.15 -1.29 -0.07 +1.61

#####  
Holding level M of factor sex constant.  
#####

NO BETWEEN EFFECTS

-----  
POA \* PHN INTERACTION EFFECT

$T_{WJt}/c(1.0,5.0)=18.36, p=0.0089$

Averaged Trimmed Cell Means:

DY DN GY GN  
-1.87 +0.31 -0.81 -4.05

Holding level D of factor POA constant.

#####  
NO BETWEEN EFFECTS

-----  
PHN MAIN EFFECT

$T_{WJt}/c(1.0,5.0)=2.55, p=0.34$

Averaged Trimmed Cell Means:

Y N  
-1.87 +0.31

Another feature of the Toolkit is that when an effect involves an electrode factor, it will provide the McCarthy and Wood (1985) vector test (McCarthy & Wood, 1985) to determine if it does indeed indicate a change in the scalp topography rather than being an artifact of the ANOVA model. For an explanation of this issue, see (Dien & Santuzzi, 2005). Such vector tests are set off from the rest of the results tables with italics.

POA \* SIT INTERACTION EFFECT

$T_{WJt}/c(1.0,6.9)=224.44, p=0.0055$

Averaged Trimmed Cell Means:

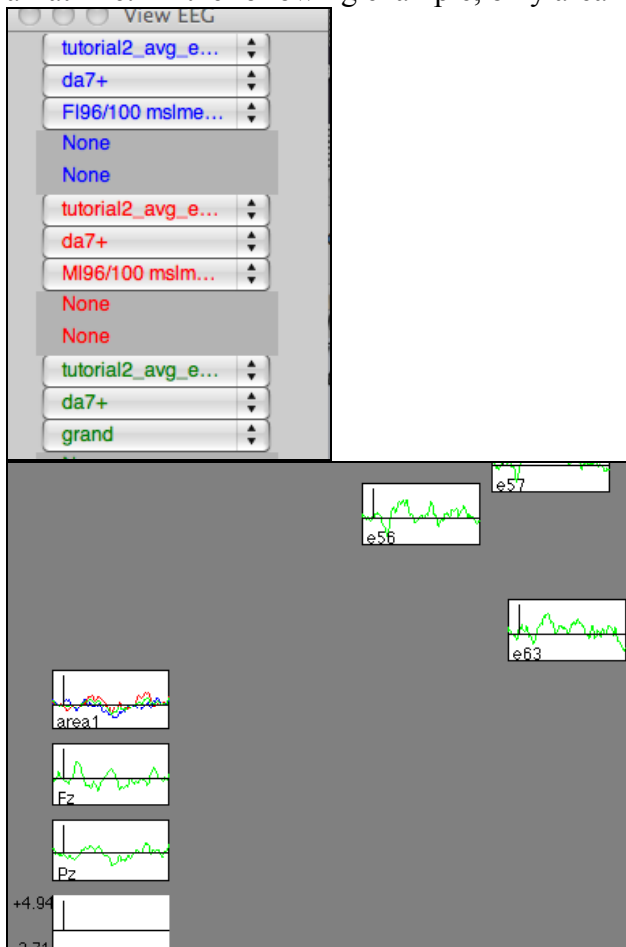
DF DP GF GP  
+4.96 +4.22 +2.54 +4.45

*Vector Scaling Test (is the scalp topography effect genuine?):*

$T_{WJt}/c(1.0,7.8)=295.24, p=0.00012$

The Toolkit will also examine whether the dataset name listed in the ANOVA text file is the same as any of the files in the current working set. If so, new combined subject conditions are added to the dataset with the waveforms (with the name of the corresponding between group cell) of each trimmed average of each cell in the ANOVA. They will only be added if the original dataset already has the corresponding cells and channels. Thus, it will now be possible to compare the ANOVA results to the exact

waveforms that correspond to them. Bear in mind that this will only be done for channels (or regional channel averages) in the actual ANOVA so the other channels will be left as a flat line. In the following example, only area1 was in the ANOVA.



If needed, one can also perform follow-up contrasts. To do so, click on the Contrast button. From there one can specify up to five contrasts to be done. The numbers in the contrast should add up to zero. If no factors are desired of a certain type (no within or no between) then either put in all ones (the one exception to the add to zero rule) or leave them as zeroes and the Toolkit will change them to all ones.

Run
Done

	Cells	POA	PHN	Con1	Con2	Con3	Con4	Con5
1	da7+-e7	D	Y	1	0	0	0	0
2	da9+-e7	G	Y	-1	0	0	0	0
3	ga11-e7	D	N	0	0	0	0	0
4	ga13-e7	G	N	0	0	0	0	0

	sex	Con1	Con2	Con3	Con4	Con5
1	F	0	0	0	0	0
2	M	0	0	0	0	0

One then clicks on Run. As with the regular ANOVA, one is asked where to output the results and what ANOVA file to use. The results are rather barebones but functional:

SLIfac-TF1SF2

Contrast: 1

Within: 1 -1 0 0

Between: 1 1

$T_{WJt}/c(1.0,5.3)=2.17, p=0.34$

Once finished with contrasts, click on Done to return to the Main Window.

### ***Behavioral Data***

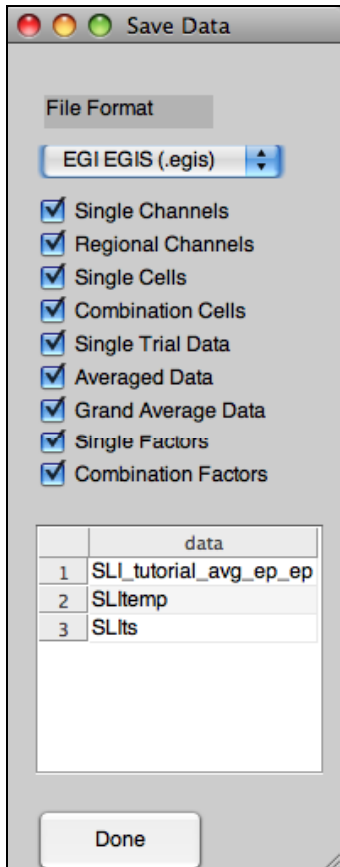
To use the ANOVA pane to analyze non-ERP data, use a spreadsheet program like Excel to have the following format:

- 1) The first line should contain the keyword "behavioral" without the quotation marks.
- 2) There should be four more lines whose contents will be ignored.
- 3) The next line is the column header line and should contain the names of the columns, separated by tabs. The Toolkit will ignore any additional columns that have no column name.
- 4) The next line will be ignored. Normally it would contain the channel region labels.
- 5) The remainder of the file should have the data with the values separated by tabs.

### **Save Files**

The very final function on the Main Pane is the Save function. It allows files in the working set to be saved permanently. Just specify the file format and which types of data are to be saved. By default, all adds to the data are saved (that is, all the ERP data plus all the combined channels, cells, subjects, and factors), although some adds may require separate files. Also, if one saves a factor file in EGIS average format then it will be

necessary to save a separate file for each subject, in addition to the Grand Average Data. If not needed and one wished to avoid the clutter of all the subject average files, one could just deselect "Averaged Data." It would also be advisable to save a copy of the file in EP format as it will retain information that might otherwise be lost in translation to other file formats. When ready, just click on the name of the dataset to be saved.



### ***Importing to EEGLAB***

The next step is to import the file back into EEGLAB format. The .set file format is now an option but incompatibilities between the way the two file formats are organized makes it difficult to provide full compatibility. For example, factor results cannot currently be saved to .set file format and likely never can (EEGLAB .set file format is predicated on the approach of keeping all the single trial data and then generating the averaged data on the fly. Factor matrices are applied to the single trial data and are stored alongside the raw single trial data. There doesn't appear to be a way of having the factor matrices be applied to the averaged data.). If the .set file format is not working for you, then you may wish to try one of the EGI file formats.

EEGLAB calls the EGI simple binary format "EGI .RAW". However, it would be better to use the option "From other formats using FILE-IO" as it will then use the FieldTrip code (which makes an effort to deduce what the baseline period should be since this information is missing from this file format, if you are using the most recent version of FieldTrip rather than the one bundled with EEGLAB). You find the options under "File>

Import Data". You can also import EGI EGIS format files using "File>Import Data>From other formats using FILE-IO". Remember that while this file format keeps track of the baseline period information it does not keep track of events so that information will be missing.

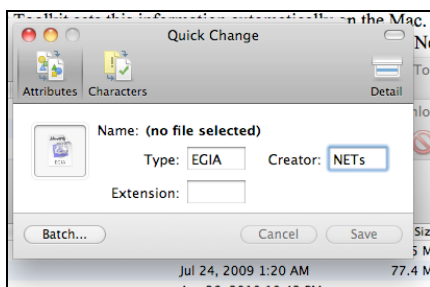
### ***Importing to Net Station***

Use either EGIS or Simple Binary formats. I recommend using EGIS unless the file is very large (in which case the file format may not be able to accommodate the data and will generate an error message when you try to save it). The files will need to have their file type and creator set. ERP PCA Toolkit functions should be able to do this under MATLAB 7+ and OS X 10.5+ should be able to do this automatically. If you get an error message saying it was unable to do so, then you will need to use a utility like Quick Change (free download from a site like VersionTracker.com) and use it to change the file type to "EGIA" and the Creator to "NETs" without the quotation marks.

In order to properly display the contents of an EGIS file, NetStation needs the electrode formats. The EP Toolkit sets this information automatically on the Mac. However, the montage information from NetStation 4.3 differ from prior versions of NetStation. Since only version 4.3 includes Hydrocel net montages, this means that versions of NetStation prior to 4.3 cannot view such data sets. Also, depending on the version you have, NetStation may not give you the full list of possible electrode nets.

There are some incompatibilities in the montages between NetStation 4.3 and earlier versions for Hydrocel nets, with 4.3 unable to open montages from earlier versions and vice versa. Because of this problem, I've added an option under the Files preferences to not add the montage to EGIS files (montage information is not used for other types of files).

If the .sbin or .egis files were originally made on a non-Mac computer or if the file were subsequently moved to such a computer then the metadata normally added by Satimage osx will not be available. In this case, NetStation will not recognize the files. It will be necessary to manually add them. A good program for doing so is Quick Change (<http://www.everydaysoftware.net/quickchange/index.html>). Changing the Creator field to NETs allows the Mac to know that the file belongs to NetStation. Changing the Type field to EGIS for .egis files and eGLY for single\_trial and average (segmented) .sbin files and UGLY for continuous (unsegmented) .sbin files allows NetStation to know which type of file it is. Just type in the fields and then drop the file on the box on the leftmost side or onto the program icon itself (including a Dock icon).



## References

- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- Cattell, R. B., & Jaspers, J. (1967). A general plasmode (No. 3010-5-2) for factor analytic exercises and research. *Multivariate Behavioral Research Monographs*, 67-3, 1-212.
- Dien, J. (2006). Progressing towards a consensus on PCA of ERPs. *Clin Neurophysiol*, 117(3), 699-702; author reply 703-7.
- Dien, J. (in press). Comparison Of Within Evaluating Two-Step PCA Of ERP Data With Geomin, Infomax, Oblimin, Promax, And Varimax Rotations. *Psychophysiology*.
- Dien, J. (1998a). Addressing misallocation of variance in principal components analysis of event-related potentials. *Brain Topography*, 11(1), 43-55.
- Dien, J. (1998b). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, and Computers*, 30(1), 34-43.
- Dien, J., Beal, D. J., & Berg, P. (2005). Optimizing principal components analysis of event-related potential analysis: Matrix type, factor loading weighting, extraction, and rotations. *Clinical Neurophysiology*, 116(8), 1808-1825.
- Dien, J., Franklin, M., & May, C. (2006). Is "blank" a suitable neutral prime for event-related potential experiments? *Brain and Language*, 97, 91-101.
- Dien, J., & Frishkoff, G. A. (2005). Principal components analysis of event-related potential datasets. In T. Handy (Ed.), *Event-Related Potentials: A Methods Handbook*. Cambridge, Mass: MIT Press.
- Dien, J., & Santuzzi, A. M. (2005). Application of repeated measures ANOVA to high-density ERP datasets: A review and tutorial. In T. Handy (Ed.), *Event-Related Potentials: A Methods Handbook*. Cambridge, Mass: MIT Press.
- Dien, J., Spencer, K. M., & Donchin, E. (2003). Localization of the event-related potential novelty response as defined by principal components analysis. *Cognitive Brain Research*, 17, 637-650.
- Dien, J., Spencer, K. M., & Donchin, E. (2004). Parsing the "Late Positive Complex": Mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology*, 41(5), 665-678.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Kayser, J., & Tenke, C. E. (2003). Optimizing PCA methodology for ERP component identification and measurement: Theoretical rationale and empirical evaluation. *Clinical Neurophysiology*, 114(12), 2307-2325.
- Keselman, H. J., Wilcox, R. R., & Lix, L. M. (2003). A generally robust approach to hypothesis testing in independent and correlated groups designs. *Psychophysiology*, 40, 586-596.
- Leonowicz, Z., Karvanen, J., & Shishkin, S. L. (2005). Trimmed estimators for robust averaging of event-related potentials. *J Neurosci Methods*, 142(1), 17-26.
- McCarthy, G., & Wood, C. C. (1985). Scalp distribution of event-related potentials: An



ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, 62, 203-208.

Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, 35(1), 99-115.

Schimmel, H. (1967). The (+/-) reference: Accuracy of estimated mean components in average response studies. *Science*, 157, 92-94.

Spencer, K. M., Dien, J., & Donchin, E. (1999). A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology*, 36, 409-414.

Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal Analysis of the Late ERP Responses to Deviant Stimuli. *Psychophysiology*, 38(2), 343-358.

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics*. New York: Harper & Row, Publishers.

Wilcox, R. R. (2001). *Fundamentals of modern statistical methods: Substantially improving power and accuracy*. New York: Springer.

# Appendices

## EP Toolkit File Format

The EP Toolkit has an internal file format used for analyses that is based on a MATLAB structured variable and is saved as a MATLAB .mat file. Parenthetical notations (e.g., subject,spec) refer to the form of the variable array, where the first term is the rows and the second term is the columns.

EPdata : Structured array with the data and accompanying information.  
.data : 5D data matrix [channels, time points, cells/trials, subjects, factors].  
Trials are grouped by cell in single\_trial format.  
.noise : 4D matrix [channels, time points, cells/trials, subjects] mirroring .data.  
This is the +/- reference average (Schimmel, 1967) which provides an estimate of the noise level  
in averaged data by flipping every other trial. Optional.  
.std : 4D matrix [channels, time points, cells/trials, subjects] mirroring .data.  
This is the standard deviation from averaging trials. Optional.  
.facData : 5D data matrix [channels, time points, cells/trials, subjects, factors] for combined factors (only when facVecS & facVecT are used).  
They are represented separately from .data because they can't be compacted using the facVec mechanism.  
.montage : String with the montage information, if available.  
.fileFormat: The file format.  
.dataType : The type of the data: 'continuous', 'single\_trial', or 'average' (default: average)  
.chanNames : The channel names.  
.timeNames : The msec of the sample onset with respect to the stimulus onset time.  
.subNames : The subject names  
.cellNames : The cell names (once for each trial for single\_trial files).  
.trialNames: The trial number ID per cell (starting from 1). (single\_trial data only)  
.facNames : The factor names (only for factor files)  
.chanTypes : The type of the channel: EEG, MEG, ANS (autonomic), REF (explicit current reference), REG (regional average)  
.subTypes : The type of the subject: RAW (single trial), AVG (subject average), GAV (grand average)  
.cellTypes : The type of the cell: SGL (one cell), CMB (combination of cells)  
.facTypes : The type of the factor: SGL (one factor), CMB (combination of factors)  
.EPver : The EP Toolkit version information.  
.ver : The Matlab version information.  
.date : The date the file was created.  
.Fs : The sampling frequency in Hz.  
.baseline : The number of samples prior to the trigger event (positive number).

Thus, the ms in the timeNames refers to the offset of each sample. A baseline of 50 samples means

that the epoch started 200 ms (50 samples) prior to the stimulus onset. A baseline of 1 sample means

that the epoch started 4 ms (1 sample) prior to the stimulus onset.

- .ename : The name of the experiment.
- .dataName : A descriptive name for the dataset, used to differentiate the active datasets during analysis.
- .trialSpecs : Cell array (trial,spec) of specific information for trials (single\_trial data only)
- .trialSpecNames : Cell array of the name of each trial spec type.
- .subjectSpecs : Cell array (subject,spec) of specific information for subjects. Can be empty if no specs.
- .subjectSpecNames : Cell array of the name of each subject spec type.
- .events : Cell array of event structured variables (subject,cell/trial)
- .type = string (events of "trial" type dropped as they are redundant with cell name information.)
- .sample = expressed in samples, the first sample of a recording is 1
- .value = number or string
- .offset = expressed in samples (0=first sample of epoch, negative means event after first sample of epoch)
- .duration = expressed in samples
- .timestamp = expressed in timestamp units, which vary over systems (optional)
- .avgNum : Number of waveforms going into averages (subject,cell)  
0 means unknown and -1 means bad.
- .subNum : Number of subjects going into averages (subject,cell)  
0 means unknown and -1 means bad.
- .fileName : Name of original file.
- .history : Command used to create current file.
- .ced : The name of the .ced file for electrode coordinates.
- .eloc : The electrode location information, one for each channel (see readlocs header)  
eloc is the same length as the channels, with REG channels having a blank entry.
- .implicit : The electrode information for implicit references and fiducial locations (see readlocs header)
- .FacPat : Optional. Factor pattern matrix from doPCA (or doPCAst if two-step PCA performed). Editable. (rows=variables, cols=factors)
- .facVar : Optional. age of variance accounted for by each rotated factor from doPCA or doPCAst. Editable.
- .facVarQ : Optional. age of variance uniquely accounted for by each rotated factor from doPCA or doPCAst. Editable.
- .facVecT : For temporal PCA factor files, the factor waveform. Used to compress the data. (rows=points,cols=factors)
- .facVecS : For spatial PCA factor files, the factor scalp topography. Used to compress the data.(rows=chans,cols=factors)
- .analysis
- .blinkTrial : Array of blink-corrected trials (subject,cell/trial)

.moveTrial : Array of movement-corrected trials (subject,cell/trial)  
 .badTrials : Array of bad trials (subject,cell/trial)  
 .badChans : Array of corrected bad channels (subject,cell/trial,channel). -1 in a session file means still bad.  
     Negative numbers in an average file means number of still bad channels that went into the average  
     (or rather, were left out of the average).  
 .pca (all optional)  
     fields from ep\_doPCA and ep\_doPCAs steps. See them for documentation.  
     Not affected by editing.

## Publications Using The EP Toolkit

Clementz, B. A., Brahmabhatt, S. B., McDowell, J. E., Brown, R., & Sweeney, J. A. (2007). When does the brain inform the eyes whether and where to move? An EEG study in humans. *Cereb Cortex*, 17(11), 2634-2643.

Curran, T., & Dien, J. (2003). Differentiating amodal familiarity from modality-specific memory processes: An ERP study. *Psychophysiology*, 40, 979-988.

Curran, T., DeBuse, C., Woroch, B., & Hirshman, E. (2006). Combined pharmacological and electrophysiological dissociation of familiarity and recollection. *J Neurosci*, 26(7), 1979-1985.

Curran, T., & Friedman, W. J. (2004). ERP old/new effects at different retention intervals in recency discrimination tasks. *Brain Res Cogn Brain Res*, 18(2), 107-120.

Dien, J. (1999). Differential lateralization of trait anxiety and trait fearfulness: evoked potential correlates. *Personality and Individual Differences*, 26(1), 333-356.

Dien, J., Franklin, M., & May, C. (2006). Is "blank" a suitable neutral prime for event-related potential experiments? *Brain and Language*, 97, 91-101.

Dien, J., Frishkoff, G. A., Cerbone, A., & Tucker, D. M. (2003). Parametric analysis of event-related potentials in semantic comprehension: Evidence for parallel brain mechanisms. *Cognitive Brain Research*, 15, 137-153.

Dien, J., Spencer, K. M., & Donchin, E. (2003). Localization of the event-related potential novelty response as defined by principal components analysis. *Cognitive Brain Research*, 17, 637-650.

Dien, J., Spencer, K. M., & Donchin, E. (2004). Parsing the "Late Positive Complex": Mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology*, 41(5), 665-678.

Dien, J., Tucker, D. M., Potts, G., & Hartry, A. (1997). Localization of auditory evoked

- potentials related to selective intermodal attention. *Journal of Cognitive Neuroscience*, 9(6), 799-823.
- Dien, J., Michelson, C. A., & Franklin, M. S. (2010). Separating the visual sentence N400 effect from the P400 sequential expectancy effect: Cognitive and neuroanatomical implications. *Brain Res*, 1355, 126-140.
- Foti, D., & Hajcak, G. (2009). Depression and reduced sensitivity to non-rewards versus rewards: Evidence from event-related potentials. *Biological Psychology*, 81(1), 1-8.
- Foti, D., Hajcak, G., & Dien, J. (2009). Differentiating Neural Responses to Emotional Pictures: Evidence from Temporal-spatial PCA. *Psychophysiology*, 46(3), 521-530.
- Franklin, M. S., Dien, J., Neely, J. H., Waterson, L. D., & Huber, L. (2007). Semantic Priming Modulates the N400, N300, and N400RP. *Clinical Neurophysiology*, 118(5), 1053-1068.
- Frishkoff, G. A., Perfetti, C. A., & Westbury, C. (2009). ERP measures of partial semantic knowledge: left temporal indices of skill differences and lexical quality. *Biol Psychol*, 80(1), 130-147.
- Frishkoff, G. A., Perfetti, C. A., & Collins-Thompson, K. (in press). Lexical Quality in the Brain: ERP evidence for robust word learning from context. *Developmental Neuropsychology*.
- Hestvik, A., Maxfield, N., Schwartz, R. G., & Shafer, V. (2007). Brain responses to filled gaps. *Brain Lang*, 100(3), 301-316.
- Holroyd, C. B., Pakzad-Vaezi, K. L., & Krigolson, O. E. (2008). The feedback correct-related positivity: sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology*, 45(5), 688-697.
- Krigolson, O. E., & Holroyd, C. B. (2006). Evidence for hierarchical error processing in the human brain. *Neuroscience*, 137(1), 13-17.
- Krigolson, O. E., & Holroyd, C. B. (2007). Hierarchical error processing: different errors, different systems. *Brain Res*, 1155, 70-80.
- Krigolson, O. E., Holroyd, C. B., Van Gyn, G., & Heath, M. (2008). Electroencephalographic correlates of target and outcome errors. *Exp Brain Res*, 190(4), 401-411.
- Lister, J. J., Maxfield, N. D., & Pitt, G. J. (2007). Cortical evoked response to gaps in noise: within-channel and across-channel conditions. *Ear Hear*, 28(6), 862-878.
- Luu, P., Tucker, D. M., & Stripling, R. (2007). Neural mechanisms for learning actions in

context. *Brain Res*, 1179, 89-105.

Macnamara, A., Foti, D., & Hajcak, G. (2009). Tell me about it: neural activity elicited by emotional pictures and preceding descriptions. *Emotion*, 9(4), 531-543.

Matsuda, I., Nittono, H., Hirota, A., Ogawa, T., & Takasawa, N. (2009). Event-related brain potentials during the standard autonomic-based concealed information test. *Int J Psychophysiol*, 74(1), 58-68.

McDonald, C. G., Gabbay, F. H., Rietschel, J. C., & Duncan, C. C. (in press). Evidence for a new late positive ERP component in an attended novelty oddball task. *Psychophysiology*.

Maxfield, N. D., Lyon, J. M., & Silliman, E. R. (2009). Disfluencies along the garden path: brain electrophysiological evidence of disrupted sentence processing. *Brain Lang*, 111(2), 86-100.

O'Hare, A. J., & Dien, J. (2008). The Fear Survey Schedule as a measure of anxious arousal: Evidence from ERPs. *Neurosci Lett*, 441(3), 243-247.

Fishman, I., Goldman, M. S., & Donchin, E. (2008). The P300 as an electrophysiological probe of alcohol expectancy. *Exp Clin Psychopharmacol*, 16(4), 341-356.

Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2005). Less skilled comprehenders' ERPs show sluggish word-to-text integration processes. *Written Language & Literacy*, 8(2), 157-181.

Yang, C. L., & Perfetti, C. A. (2006). Contextual constraints on the comprehension of relative clause sentences in Chinese: ERPs evidence. *Language and Linguistics*, 7(3), 697-730.

Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2007). Event-related potential indicators of text integration across sentence boundaries. *J Exp Psychol Learn Mem Cogn*, 33(1), 55-89.

## **Bibliography**

Dien, J., Tucker, D. M., Potts, G., & Hartry, A. (1997). Localization of auditory evoked potentials related to selective intermodal attention. *Journal of Cognitive Neuroscience*, 9(6), 799-823.

Dien, J. (1998). Addressing misallocation of variance in principal components analysis of event-related potentials. *Brain Topography*, 11(1), 43-55.

Dien, J., Frishkoff, G., Cerbonne, A., & Tucker, D. M. (2003). Parametric analysis of

event-related potentials in semantic comprehension: Evidence for parallel brain mechanisms. *Cognitive Brain Research*, 15, 137-153.

Dien, J., Spencer, K. M., & Donchin, E. (2003). Localization of the event-related potential novelty response as defined by principal components analysis. *Cognitive Brain Research*, 17, 637-650.

Dien, J., Spencer, K. M., & Donchin, E. (2004). Parsing the "Late Positive Complex": Mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology*, 41(5), 665-678.

Dien, J., & Frishkoff, G. A. (2004). Principal components analysis of event-related potential datasets. In T. Handy (Ed.), *Event-Related Potentials: A Methods Handbook*. Cambridge, Mass: MIT Press.

Dien, J., Beal, D. J., & Berg, P. (2005). Optimizing principal components analysis of event-related potential analysis: Matrix type, factor loading weighting, extraction, and rotations. *Clinical Neurophysiology*, 116(8), 1808-1825.

Dien, J., Khoe, W., and Mangun, G. R. (2007). Evaluation of PCA and ICA of simulated ERPs: Promax versus Infomax rotations. *Human Brain Mapping*, 28(8):742-63.

Dien, J. (2010). Evaluating Two-Step PCA Of ERP Data With Geomin, Infomax, Oblimin, Promax, And Varimax Rotations. *Psychophysiology*, 47(1), 170-183.

Dien, J. (2010). The ERP PCA Toolkit: An Open Source Program For Advanced Statistical Analysis of Event Related Potential Data. *Journal of Neuroscience Methods*, 187(1), 138-145.

O'Hare, A. J., Dien, J., Waterson, L., Savage, C. R. (2008). Activation of the Posterior Cingulate by Semantic Priming: A co-registered ERP/fMRI Study. *Brain Research*, 1189(2):97-114.

Spencer, K. M., Dien, J., & Donchin, M. (1999). A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology*, 36, 409-414.

Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal Analysis of the Late ERP Responses to Deviant Stimuli. *Psychophysiology*, 38(2), 343-358.