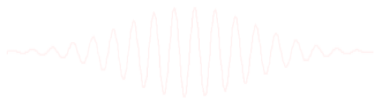
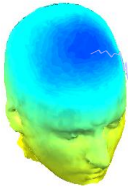


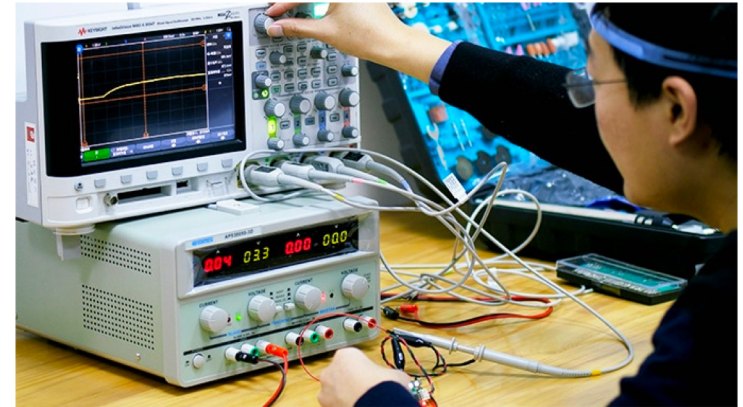
EEG Preprocessing in EEGLAB



Why preprocess data?



EEG data out of the recording device is a continuous unprocessed signal. It is like measuring a difference of potential on an oscilloscope.

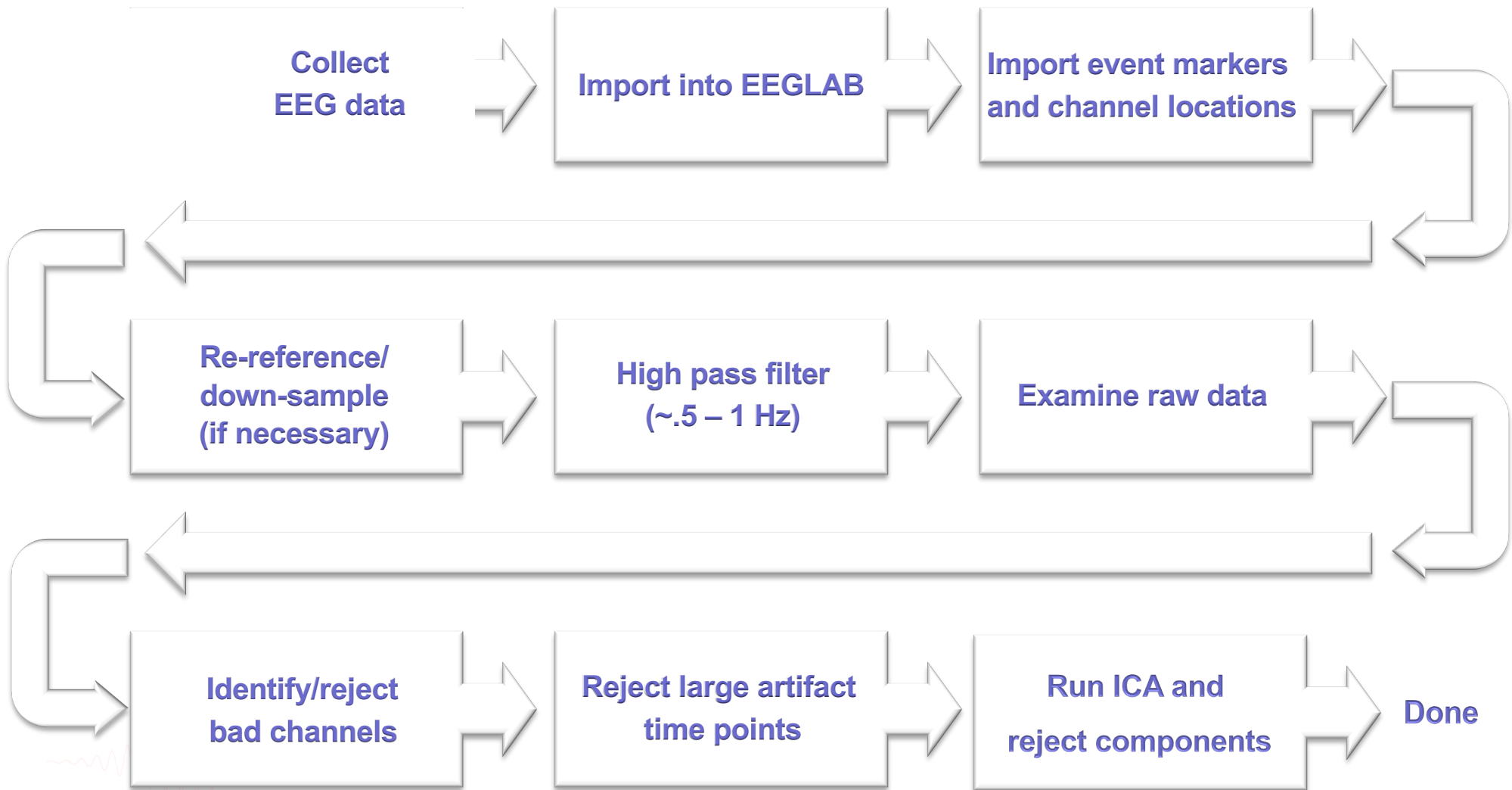
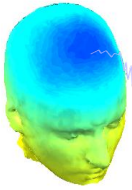


To make sense of the data, we need to:

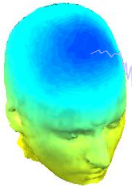
- Extract meaningful measures from it (such as brain oscillations; brain source activations)
- Compare brain data in different conditions
- Assess reliable changes due to external stimuli (event-related potentials)

Before we do all that, we apply a series of transformation to the data.

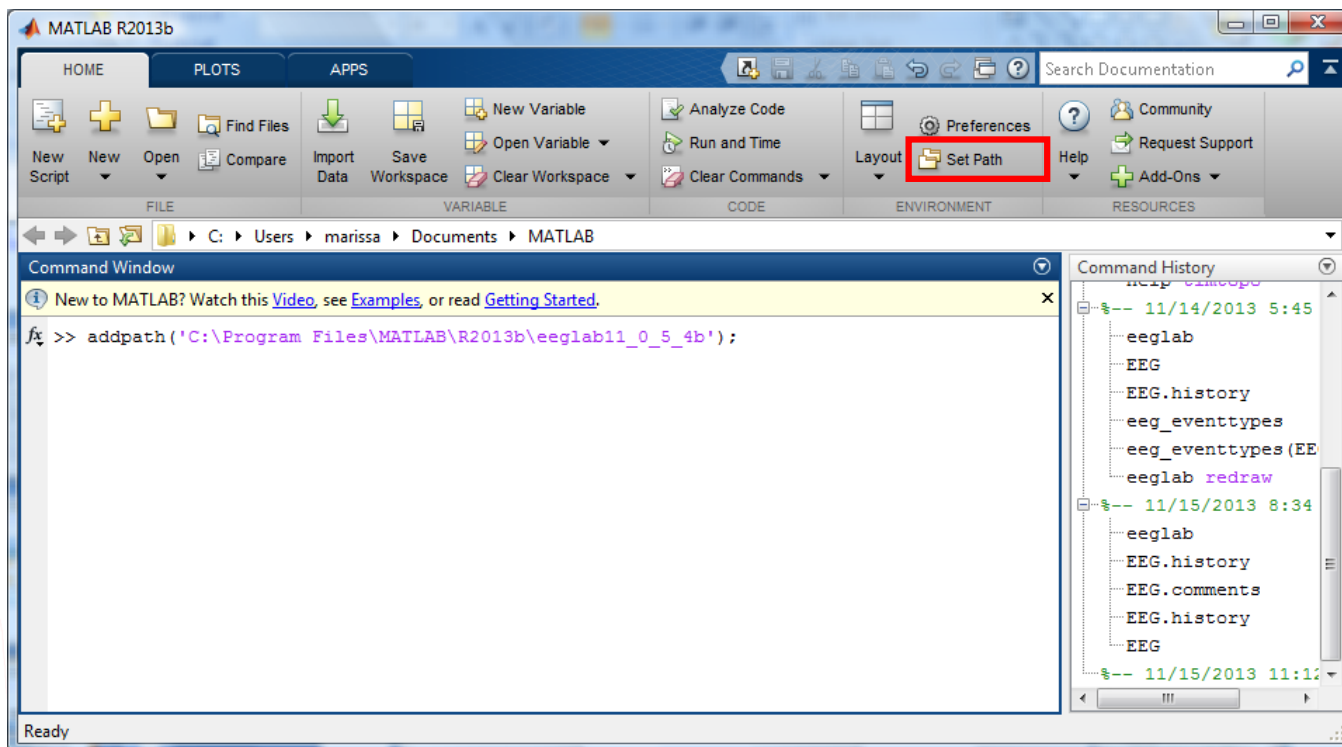
Pre-processing pipeline



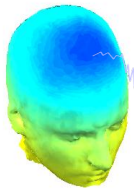
Installing EEGLAB and data folder



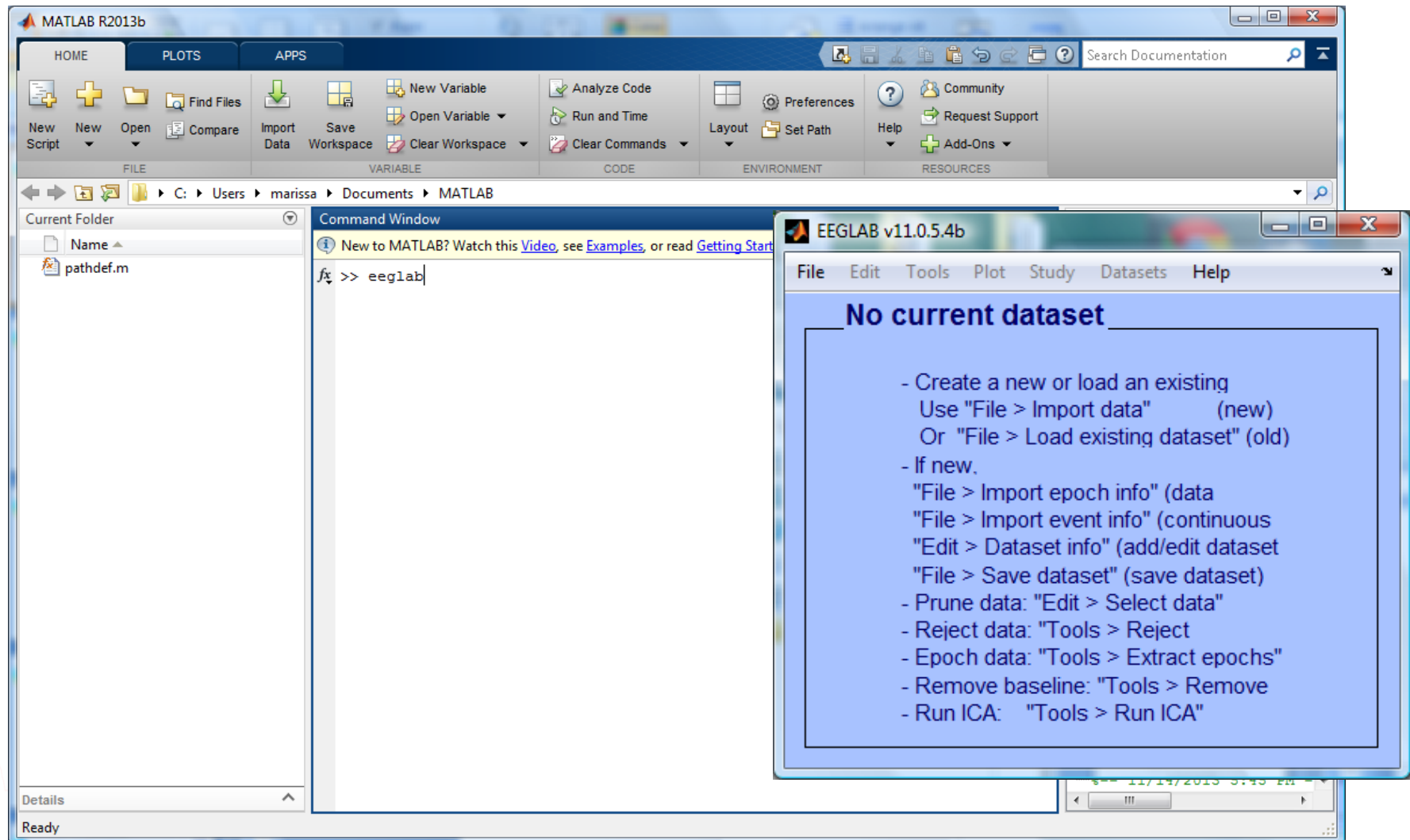
- Download and install Matlab (2008b or later)
- Download EEGLAB (<http://www.sccn.ucsd.edu/eeglab>)
- Unzip EEGLAB
- Add the EEGLAB folder to your Matlab path:



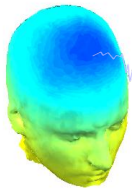
The EEGLAB Matlab software



main graphic interface



Importing a dataset

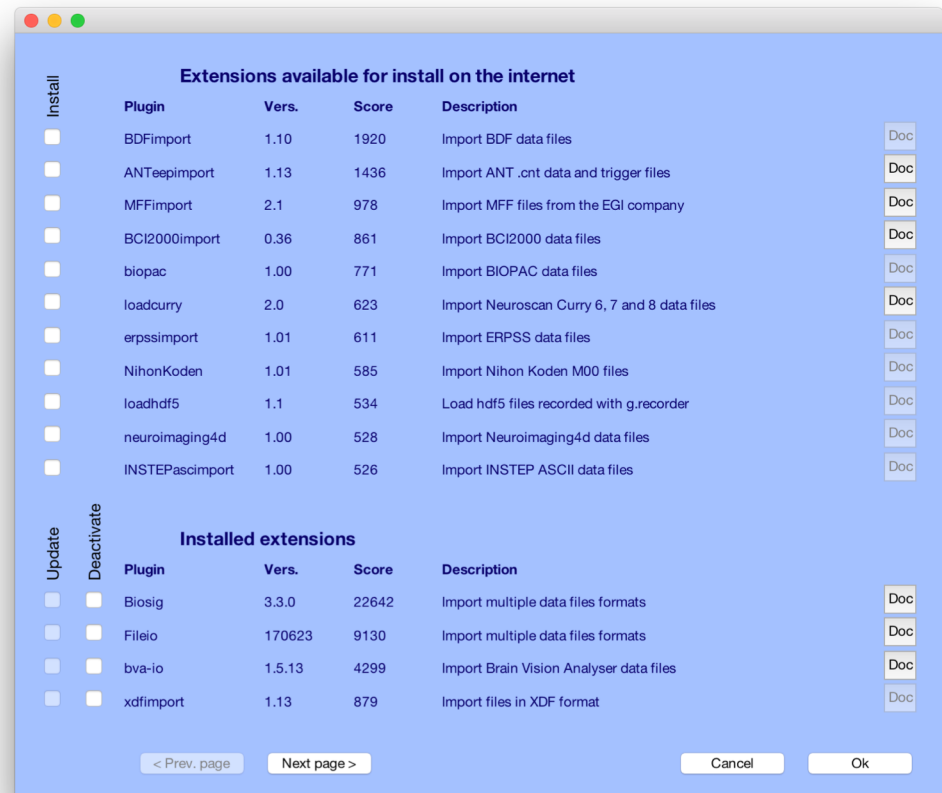
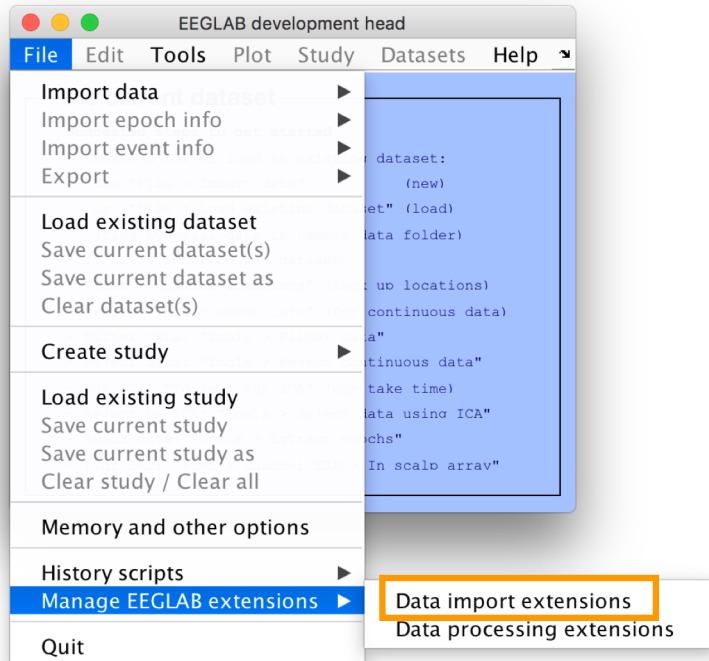
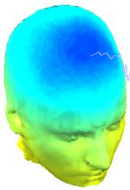


The screenshot shows the EEGLAB v11.0.5.4b window. The 'File' menu is open, displaying the following options:

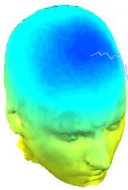
- Import data
 - Using EEGLAB functions and plugins
 - From ASCII/float file or Matlab array
 - From Netstation .mff (FILE-IO toolbox)
 - From Netstation binary simple file
 - From Multiple seg. Netstation files
 - From Netstation Matlab files
 - From BCI2000 ASCII file
 - From Snapmaster .SMA file
 - From Neuroscan .CNT file
 - From Neuroscan .EEG file
 - From Biosemi BDF file (BIOSIG toolbox)
 - From Biosemi BDF and EDF files (BDF plugin)
 - From EDF/EDF+/GDF files (BIOSIG toolbox)
 - From ANT EEProbe .CNT file
 - From ANT EEProbe .AVR file
 - From BCI2000 .DAT file
 - From BIOPAC MATLAB files
 - From Brain Vis. Rec. .vhdr file
 - From Brain Vis. Anal. Matlab file
 - From CTF folder (MEG)
 - From ERPSS .RAW or .RDF file
 - From INStep .ASC file
 - From 4D .m4d pdf file
 - From Procom Infinity Text File
 - Using the FILE-IO interface
 - Using the BIOSIG interface
 - Troubleshooting data formats...
- Import epoch info
- Import event info
- Export
- Load existing dataset
 - Save current dataset(s)
 - Save current dataset as
 - Clear dataset(s)
- Create study
 - Load existing study
 - Save current study
 - Save current study as
 - Clear study
- Memory and other options
- History scripts
- Quit



Install extension for importing data files



Supported data formats



EEGLAB tutorial: https://sccn.ucsd.edu/wiki/A01:_Importing_Continuous_and_Epoched_Data

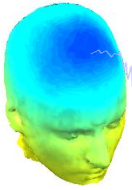
Supported Data Formats

File Format	File Extension	File type	Events	Channel Labels	EEGLAB	Biosig	File IO	Support
ANT EEPProbe	.avr	—	—	—	—	—	—	Comments
ANT EEPProbe	.cnt	—	—	—	y	y	y	Comments
ASCII	.txt	—	—	—	y	y	—	Comments
BCI2000	.bci2000	continuous	—	—	p	—	—	Comments
BCI2000	.gdf	continuous	—	—	p	—	—	Comments
Biologic	.eeg	—	—	—	—	—	—	Comments
Biopac	.mat/.acq	—	—	—	p (see comments)	—	—	Comments
Biosemi	.bdf	continuous	Channel	—	y	y	y	Comments
Blackrock	.NEV/.NSx	—	—	—	see comments	—	—	Comments
Brain Vision Analyzer	.mat	continuous & segmented	Embedded	—	y	y	n	Comments
Brain Vision Analyzer	.vhdr	—	file	—	y	y	n	Comments
BrainStorm	.vsm	—	—	—	—	—	—	Comments
Cogniscan	—	—	—	—	p	—	—	Comments
Compumedics Profusion	.raw	—	—	—	see comments	—	—	Comments
CTF/BrainStorm	.ctf	—	—	—	y	y	y	Comments
EGI/Netstation	.RAW	continuous & segmented	Channel	—	y	y	y	Comments
Elekta (MEG)	.fif	—	—	—	n (see comments)	y	n	Comments
Emotiv	.edf	—	—	—	y (see comments)	y (see comments)	y (see comments)	Comments
ERPSS	.raw	—	—	—	y	n	n	Comments
ERPSS	.rdf	—	—	—	y	n	n	Comments
European Data Format (16-bit)	.edf	—	Channel	—	y	y	n	Comments
EDF+	.edf	—	Channel	—	y	y	n	Comments
INSTEP	.asc	—	—	—	y	n	n	Comments
Matlab Array	.mat	—	Channel	—	y	y	n	Comments
Micromed	—	—	—	—	p	—	—	Comments
Neuroimaging4D	.m4d	—	—	—	y	n	n	Comments
Neuromag	.fif	—	—	—	see comments	n	see comments	Comments
Neuroscan	.avg	—	—	—	—	—	—	Comments
Neuroscan	.CNT	—	Embedded (see comments)	—	y	y	y	Comments
Neuroscan	.eeg	continuous	—	—	y	y	y	Comments
Nihon Kohden	.eeg	continuous	—	—	—	y (see comments)	—	Comments
Profusion	.slp	—	—	—	—	—	—	Comments
Snapmaster	.SMA	—	Channel	—	y	y	n	Comments
Spike2	.mat	—	—	—	y (see comments)	n	n	Comments
Tucker-Davis Technology	.tdt	—	—	—	p	—	—	Comments

BIOSIG: <http://pub.ist.ac.at/~schloegl/biosig/TESTED>

File-IO: <http://www.fieldtriptoolbox.org/development/fileio>

Sample data: basic P300 paradigm



File

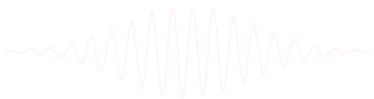
SimpleOddball.set

Data

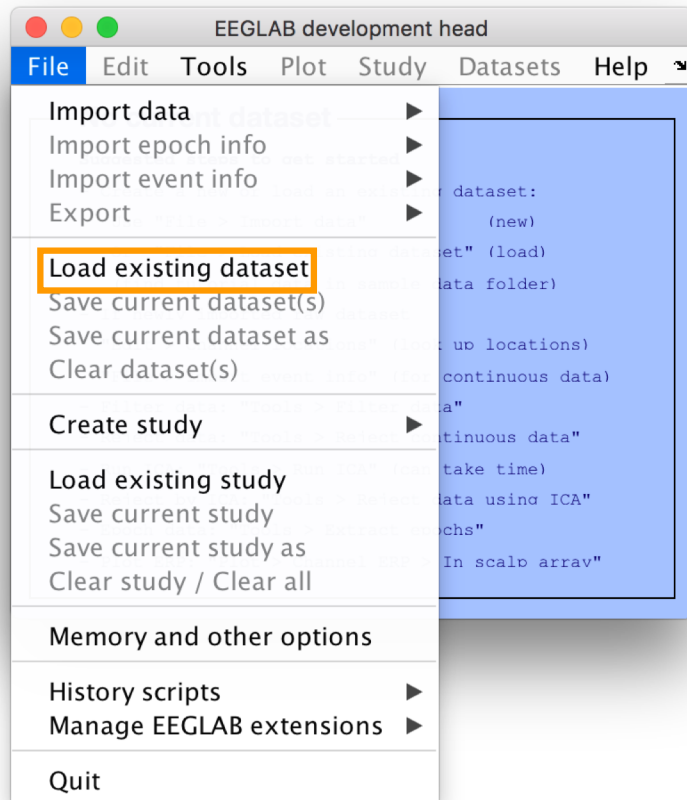
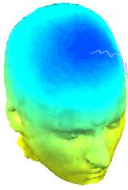
68 channel EEG, 256 Hz sampling rate, Biosemi system, re-referenced during import to averaged left and right mastoid electrodes

Task

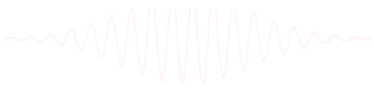
speeded button press response to star shape (no response to circle shape), 100 ms presentation duration, 200 trials



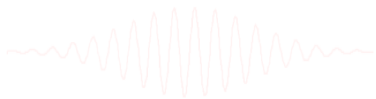
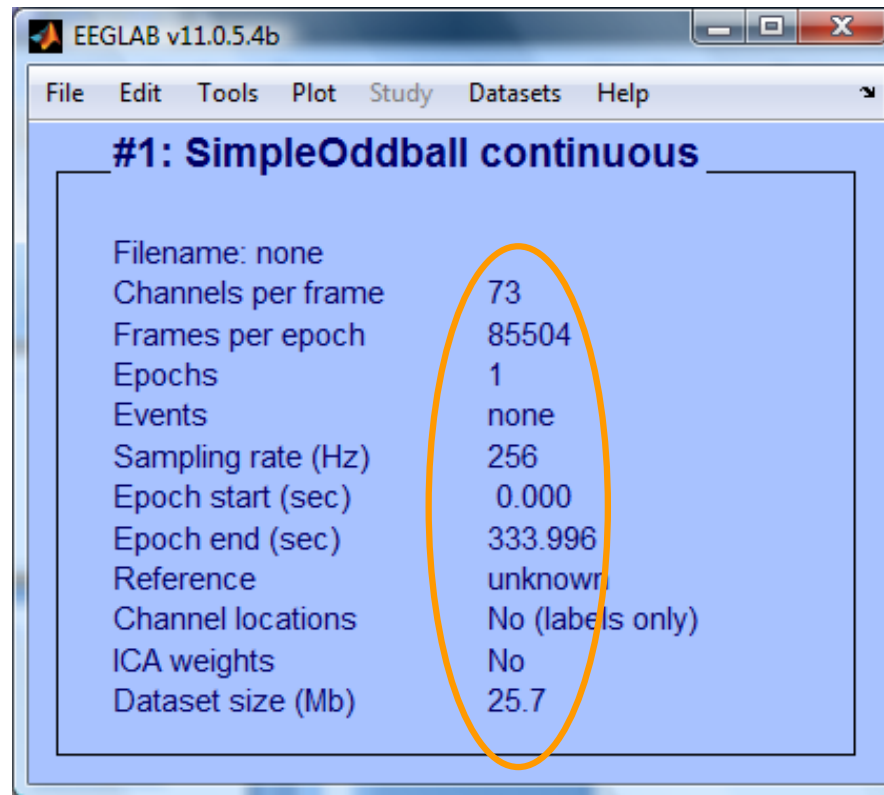
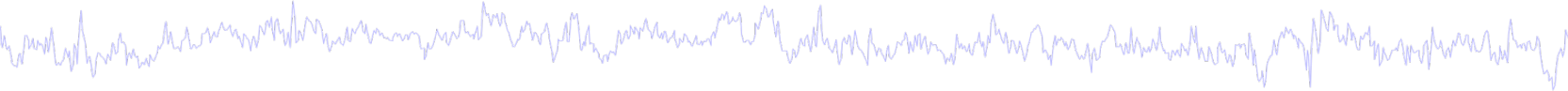
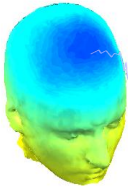
Load a dataset



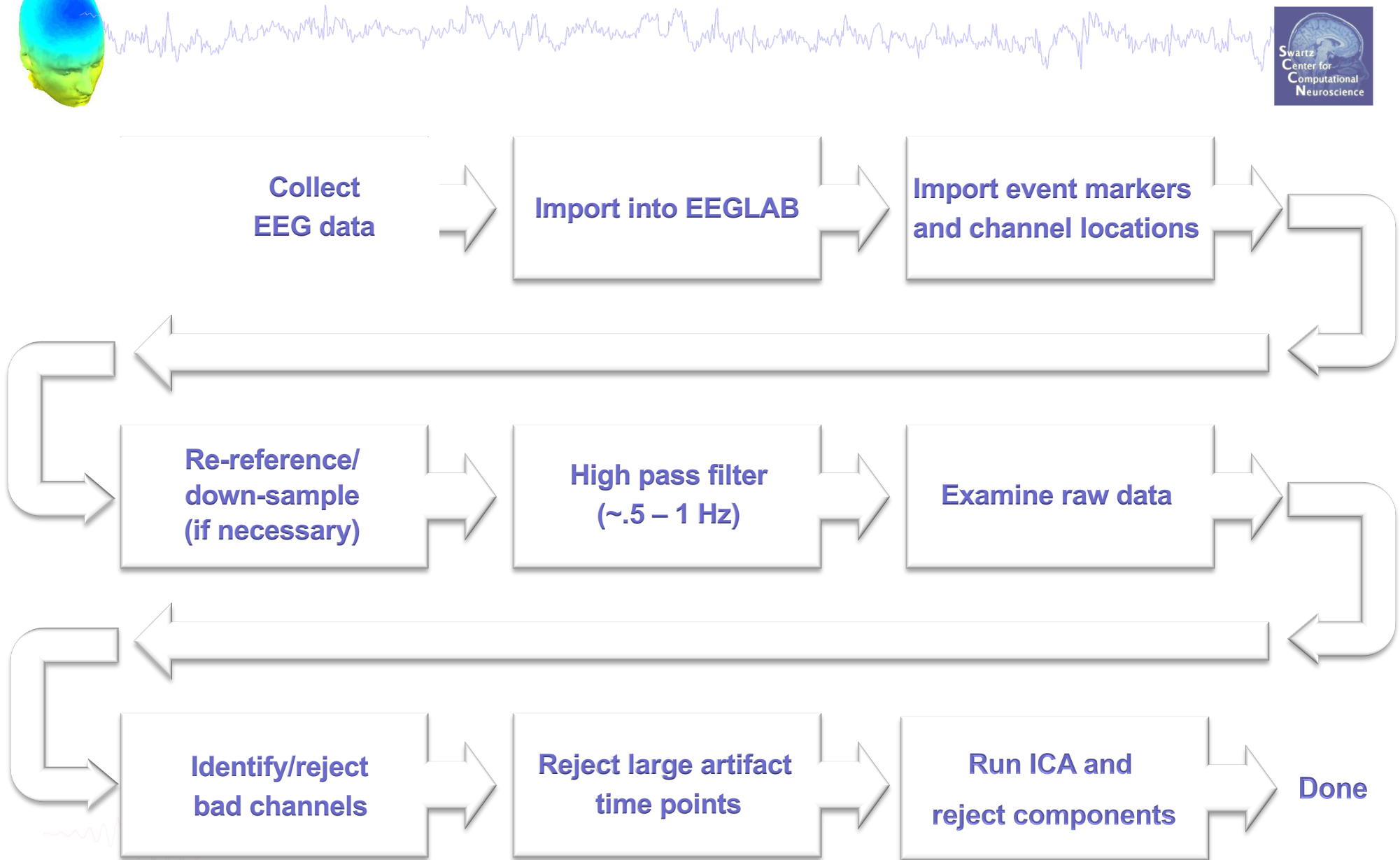
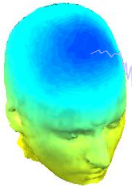
Load "SimpleOddball.set"



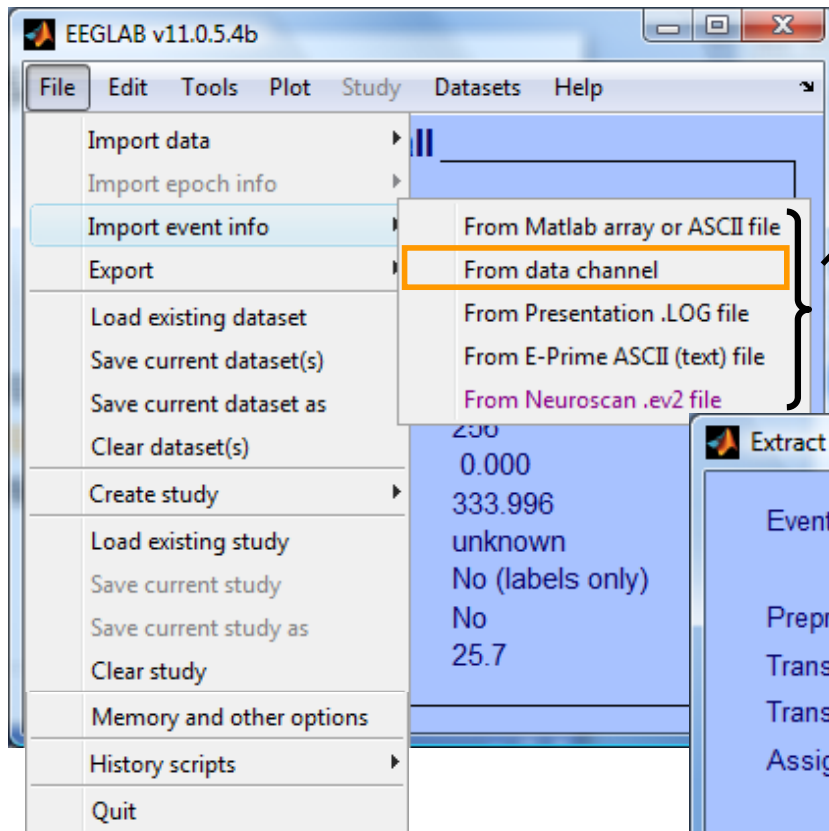
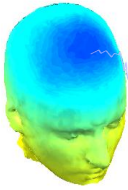
Imported EEG data



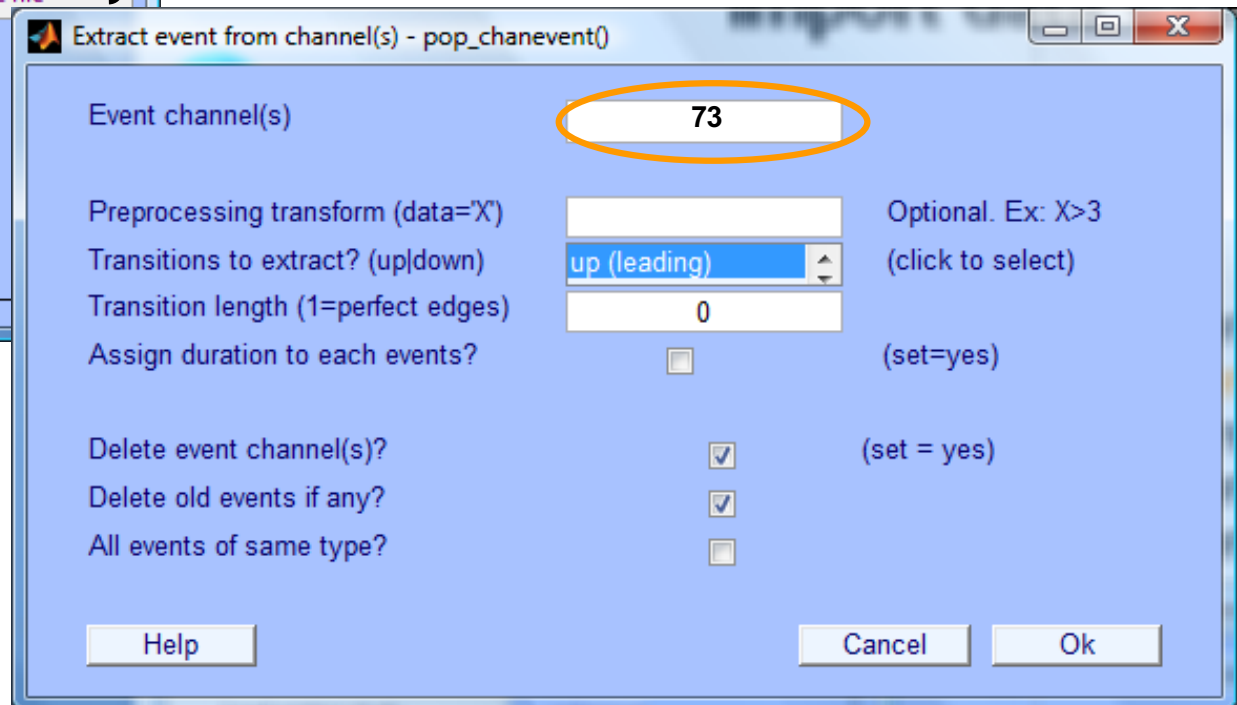
Pre-processing pipeline



Import data events

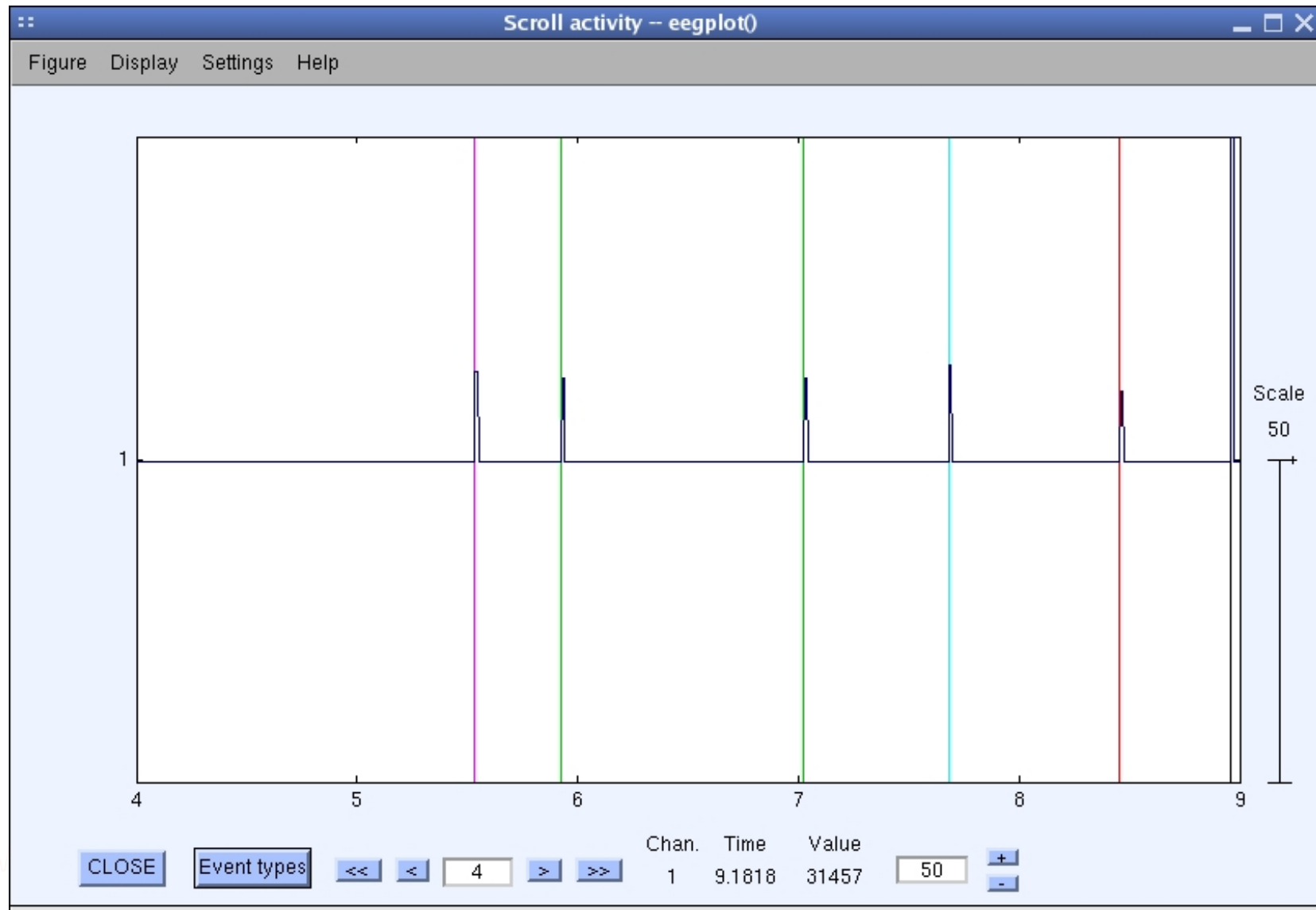
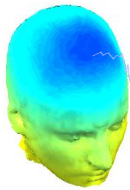


- Import events from Matlab array or ASCII file
- **Import events from data channel**
- Import from Presentation event file
- Import events from E-Prime event file
- Import events from Neuroscan event file

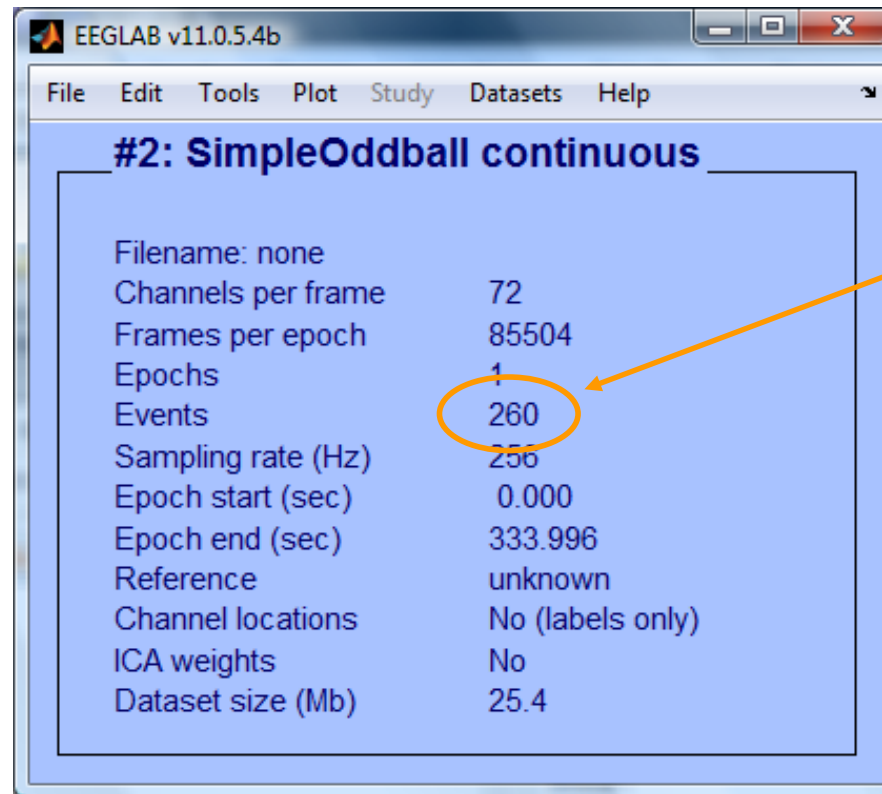
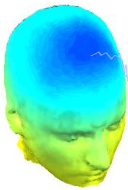


(Often imported automatically
during data import)

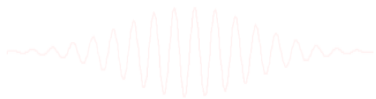
Appearance of an event channel in raw data



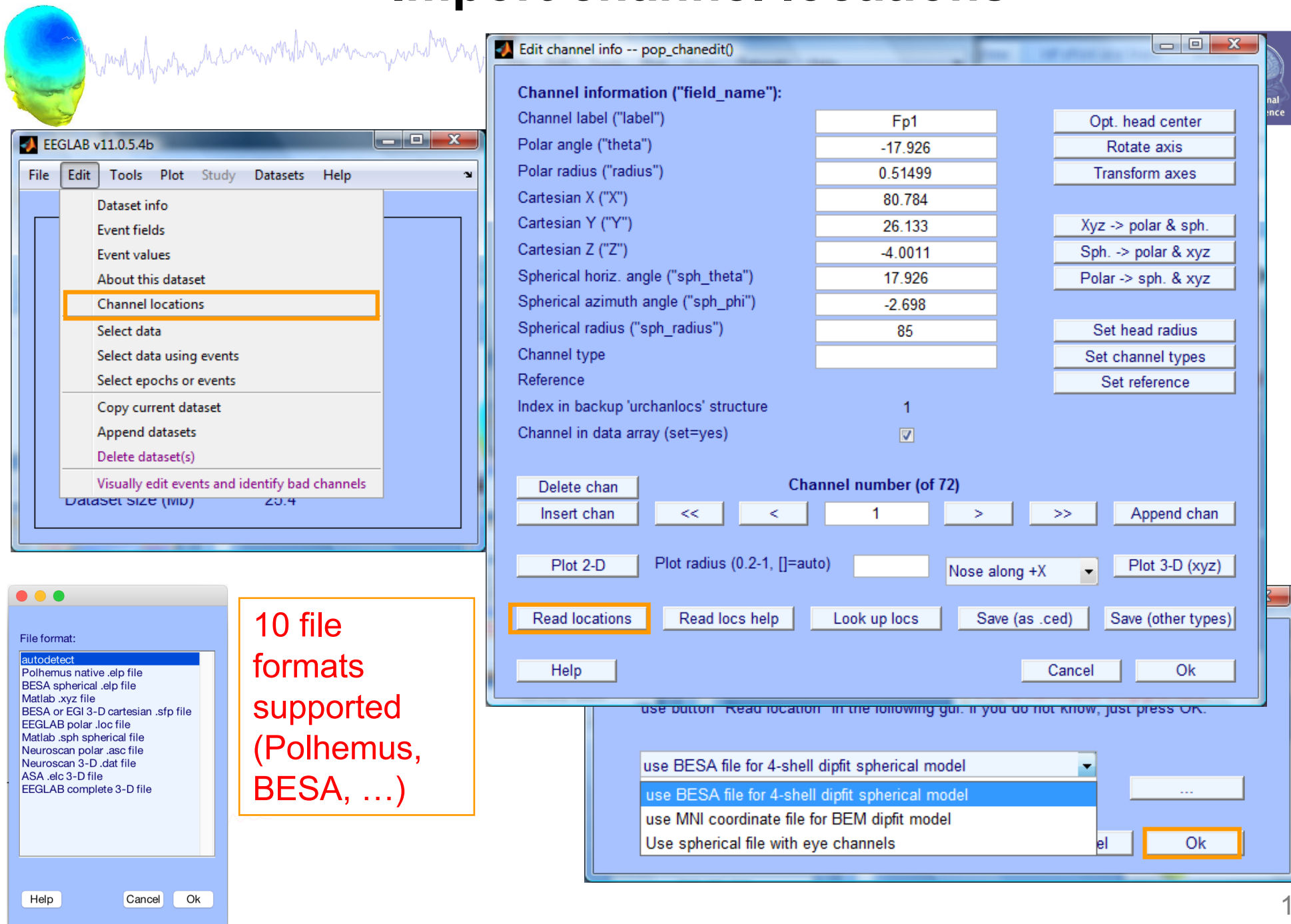
Imported data events



If event import was successful, you will see an appropriate number here



Import channel locations



EEGLAB v11.0.5.4b

File Edit Tools Plot Study Datasets Help

Dataset info
Event fields
Event values
About this dataset
Channel locations
Select data
Select data using events
Select epochs or events
Copy current dataset
Append datasets
Delete dataset(s)
Visually edit events and identify bad channels

Dataset Size (MB) 23.4

Edit channel info -- pop_chanedit()

Channel information ("field_name"):

Channel label ("label")	Fp1	Opt. head center
Polar angle ("theta")	-17.926	Rotate axis
Polar radius ("radius")	0.51499	Transform axes
Cartesian X ("X")	80.784	XYZ -> polar & sph.
Cartesian Y ("Y")	26.133	Sph. -> polar & xyz
Cartesian Z ("Z")	-4.0011	Polar -> sph. & xyz
Spherical horiz. angle ("sph_theta")	17.926	Set head radius
Spherical azimuth angle ("sph_phi")	-2.698	Set channel types
Spherical radius ("sph_radius")	85	Set reference
Channel type		
Reference		
Index in backup 'urchanlocs' structure	1	
Channel in data array (set=yes)	<input checked="" type="checkbox"/>	

Delete chan Channel number (of 72) Append chan

Insert chan << < 1 > >>

Plot 2-D Plot radius (0.2-1, []=auto) Nose along +X Plot 3-D (xyz)

Read locations Read locs help Look up locs Save (as .ced) Save (other types)

Help Cancel Ok

File format:

- autodetect
- Polhemus native .elp file
- BESA spherical .elp file
- Matlab .xyz file
- BESA or EGI 3-D cartesian .sfp file
- EEGLAB polar .loc file
- Matlab .sph spherical file
- Neuroscan polar .asc file
- Neuroscan 3-D .dat file
- ASA .elc 3-D file
- EEGLAB complete 3-D file

Help Cancel Ok

10 file formats supported (Polhemus, BESA, ...)

use BESA file for 4-shell dipfit spherical model

use BESA file for 4-shell dipfit spherical model

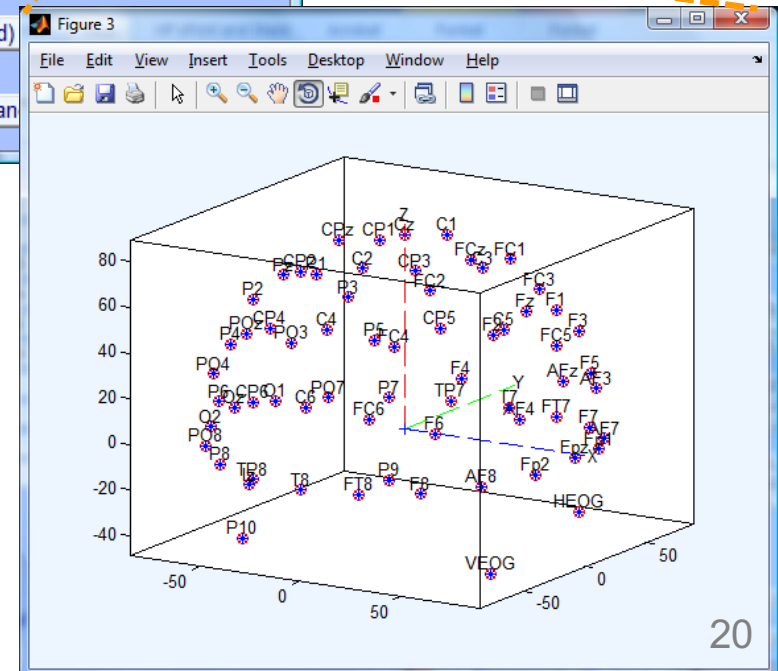
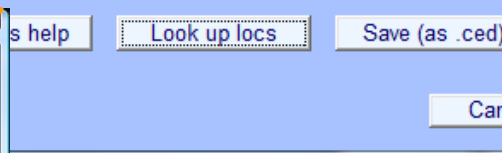
use MNI coordinate file for BEM dipfit model

Use spherical file with eye channels

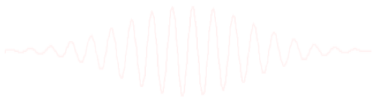
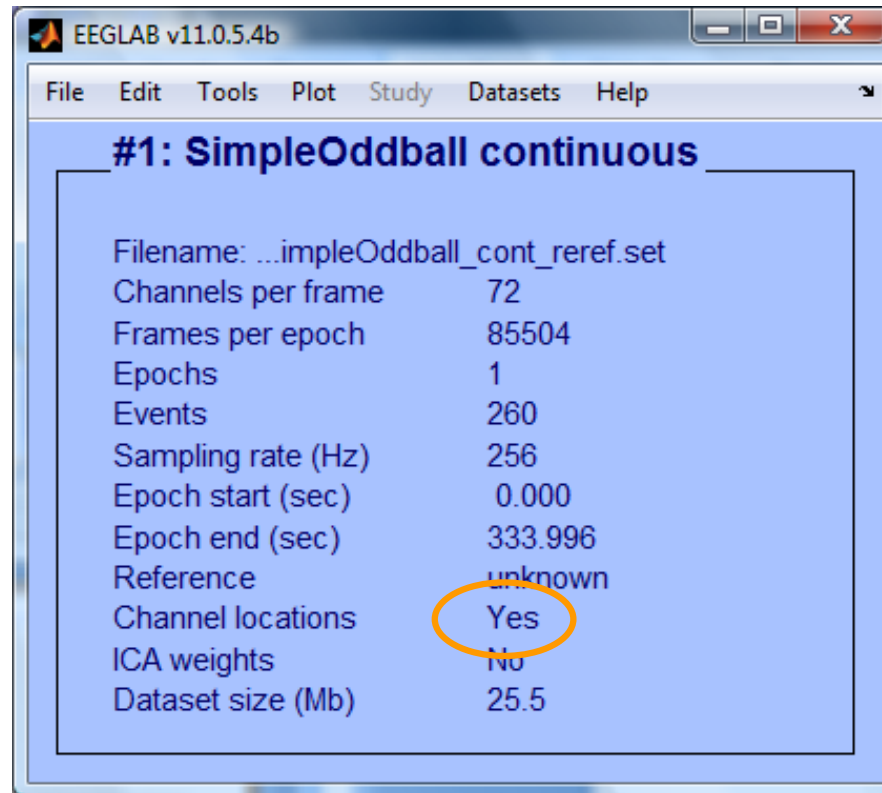
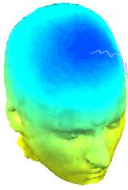
Ok



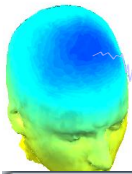
This screenshot shows the 'Look up locs' dialog box, which is a modal window for searching locations. It features a title bar with standard window controls and a menu bar with 'File', 'Edit', 'View', 'Insert', and 'Tools'. The main area contains a text input field with the placeholder text 'Enter location name'. Below this is a 'Look up' button. At the bottom, there are 'OK' and 'Cancel' buttons. The dialog is overlaid on a background window titled 'Figure 3'.



Imported channel locations



Comments and dataset history



The image shows the EEGLAB v11.0.5.4b software interface. The 'Edit' menu is open, and the 'About this dataset' option is highlighted. A dialog box titled 'Read/Enter comments -- pop_comments()' is displayed, showing the 'About this dataset' section. The dialog box contains the following text:

About this dataset

Data recorded by Marissa Westerfield
Recording date: Oct. 14, 2011

Paradigm:
-Participant looked at fixation box in center of screen
-Two types of stimuli (outline of a circle, outline of a star) were presented in the fixation box in random order
-Participant pressed a button in response to the star

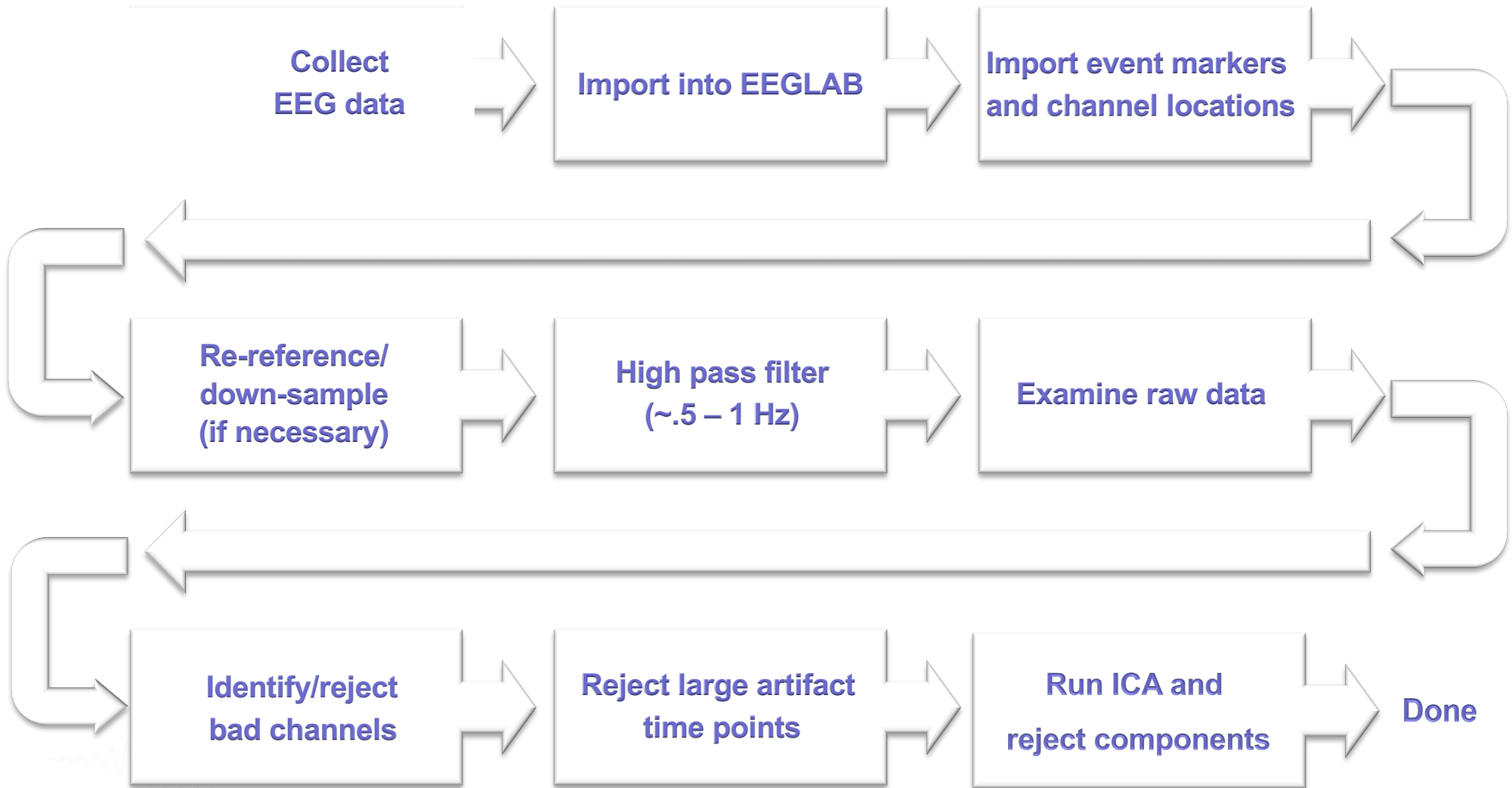
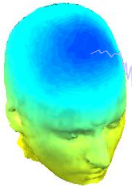
Stimulus codes:
1 = circle
2 = star
3 = button press

Recording information:
-reference electrodes were placed on right and left mastoids (data has already been referenced and the mastoid channels have been removed)

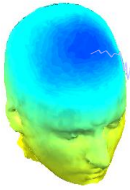
Processing steps:
high-pass filter - 0.5 Hz
Cleanline applied to 60, 120 Hz

The dialog box has 'CANCEL' and 'SAVE' buttons at the bottom.

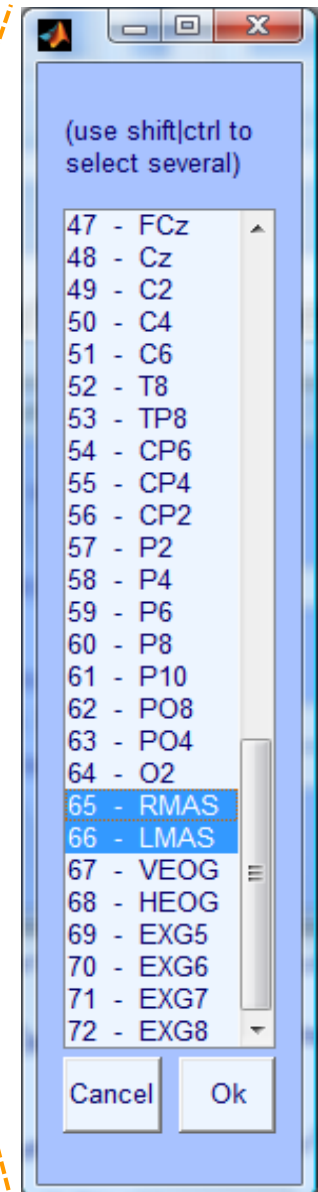
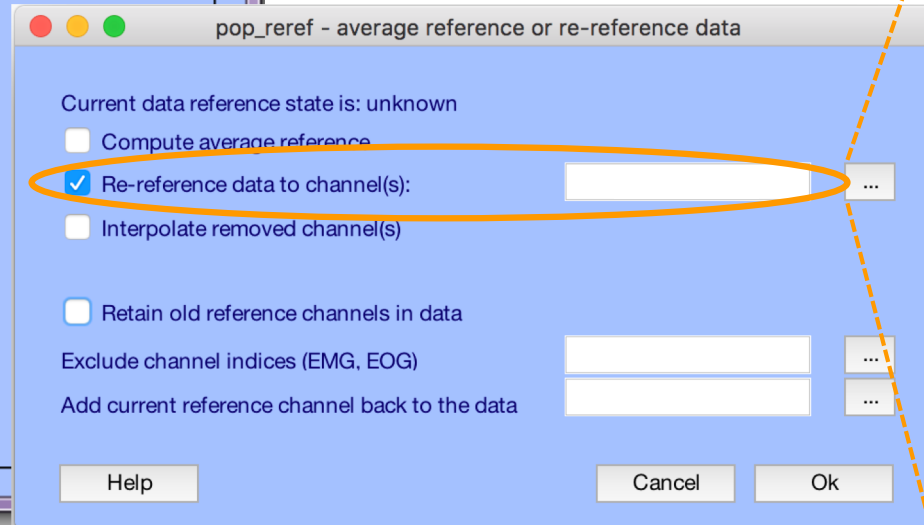
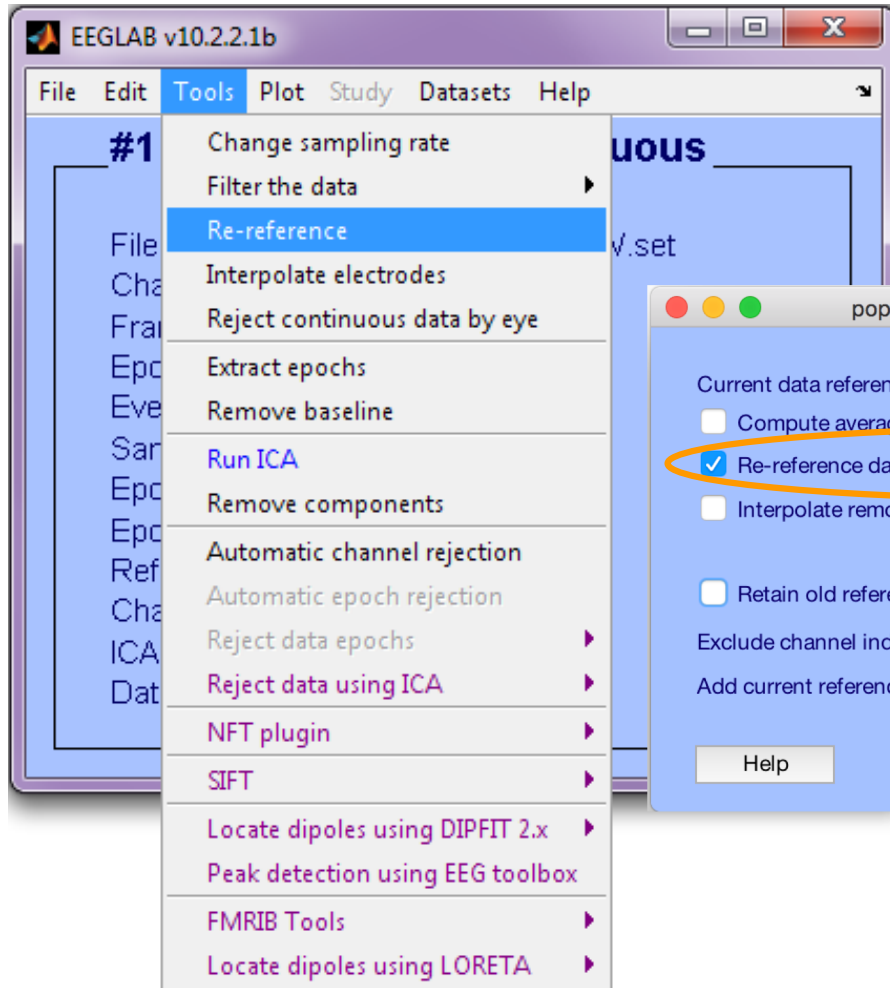
Pre-processing pipeline



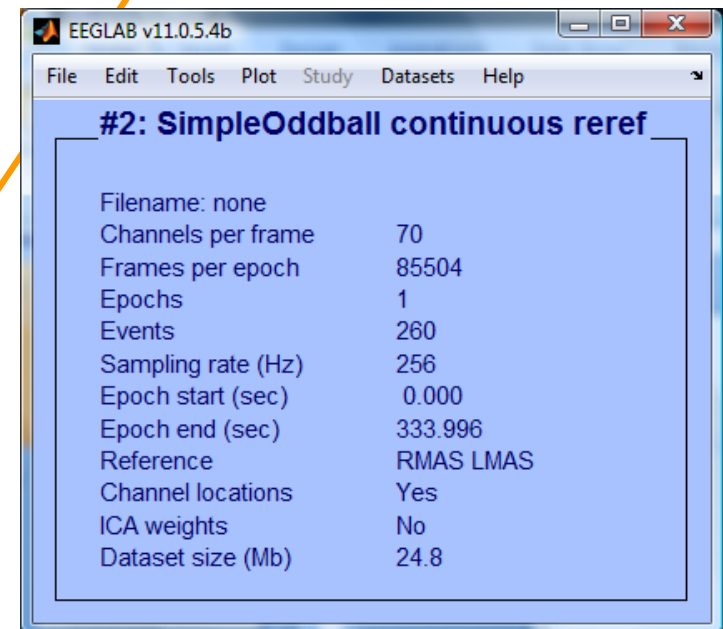
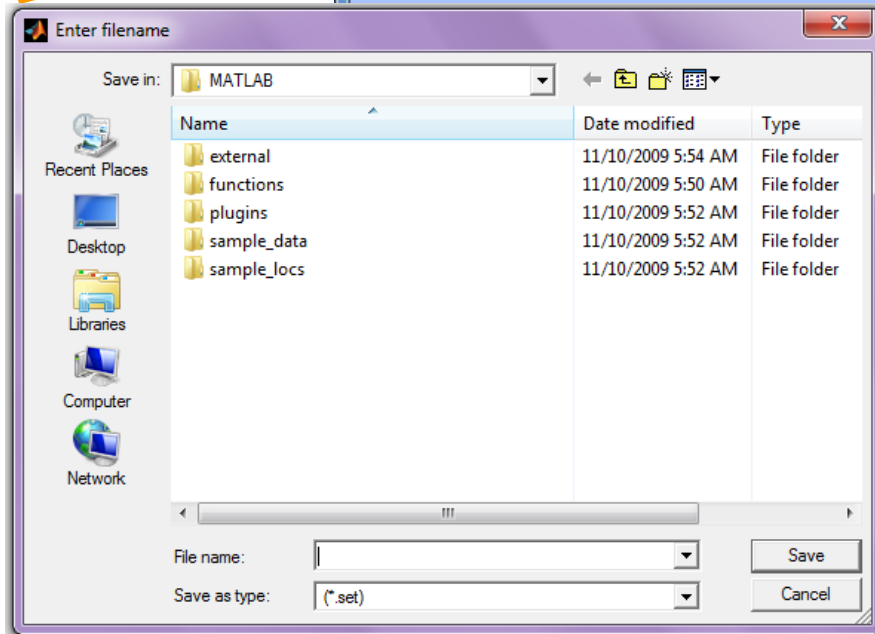
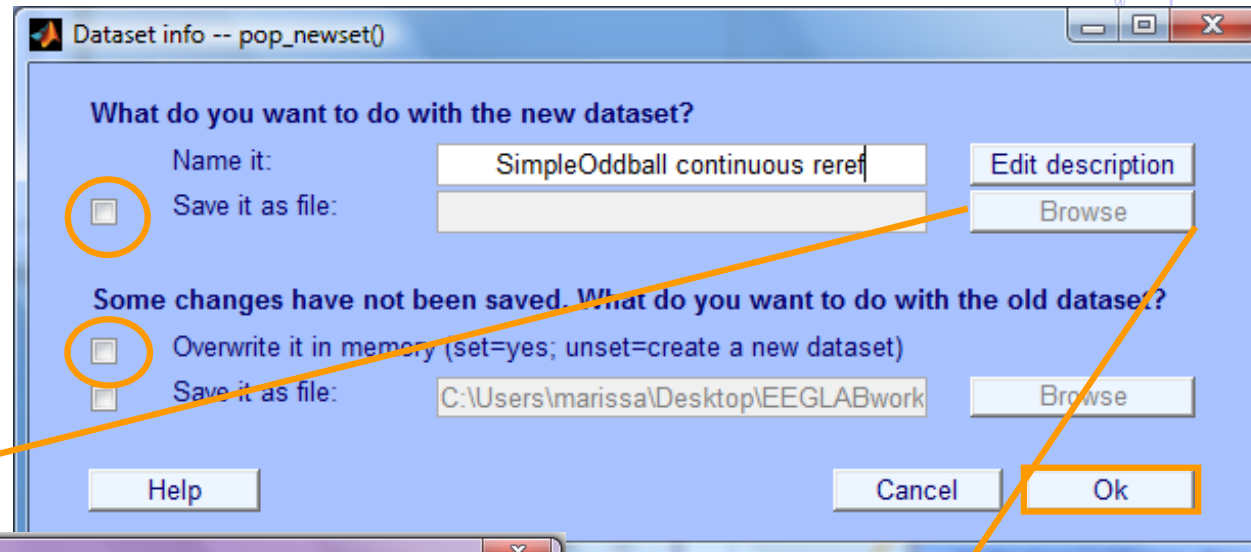
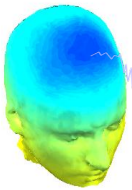
Re-reference data (if necessary/desired)



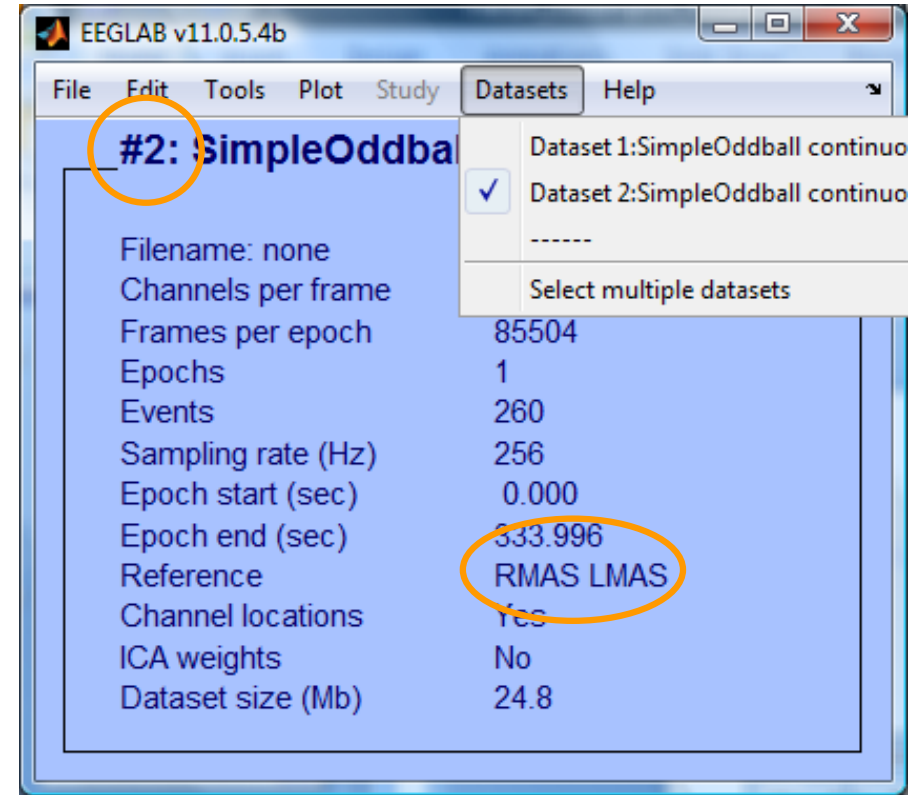
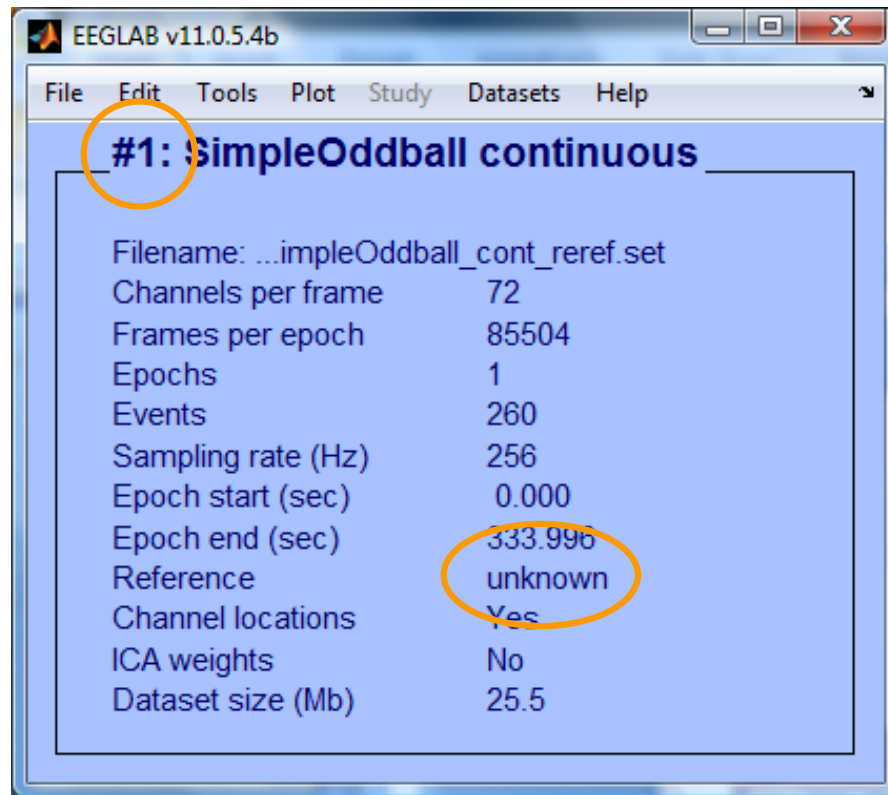
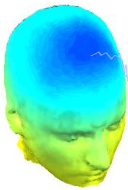
Re-reference to
(e.g.) 'linked mastoids'



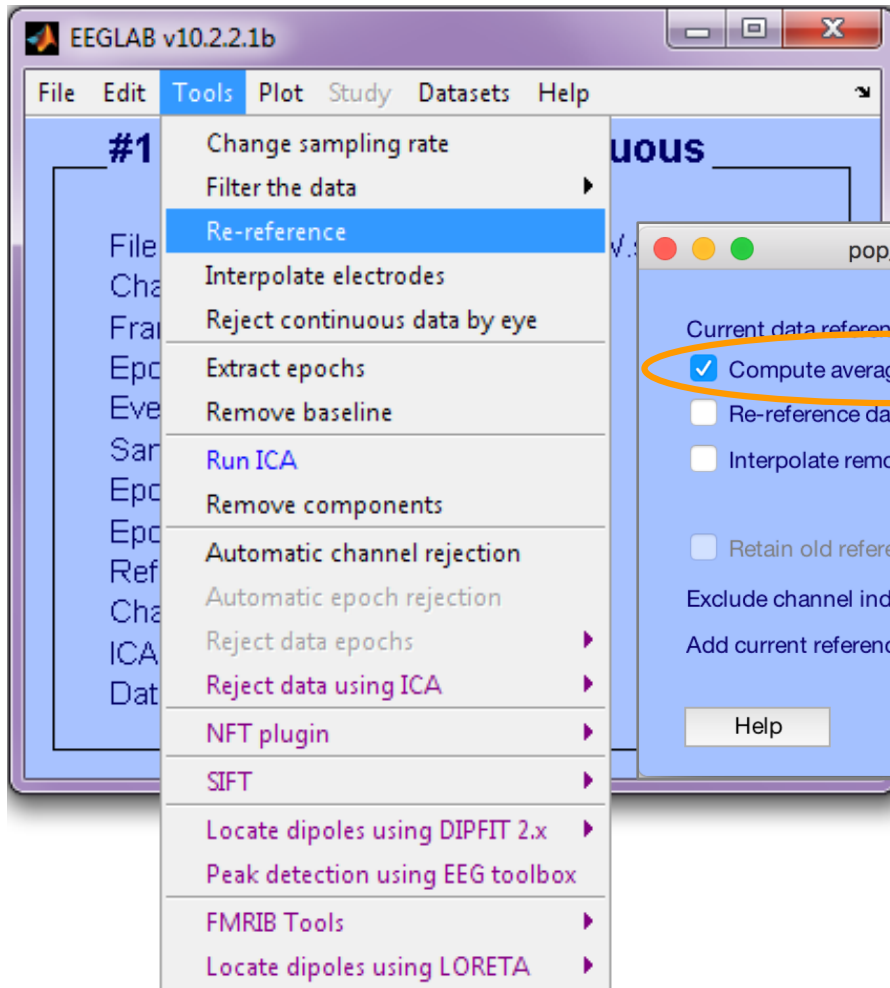
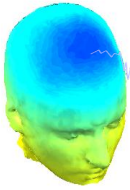
Save new dataset, keep old one



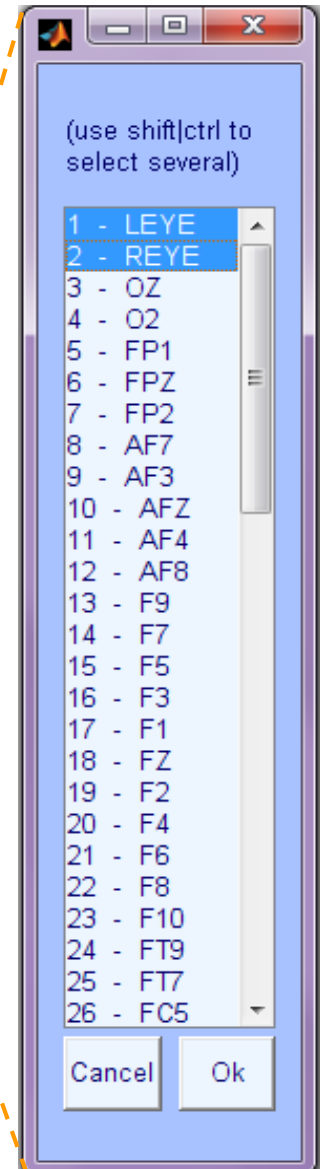
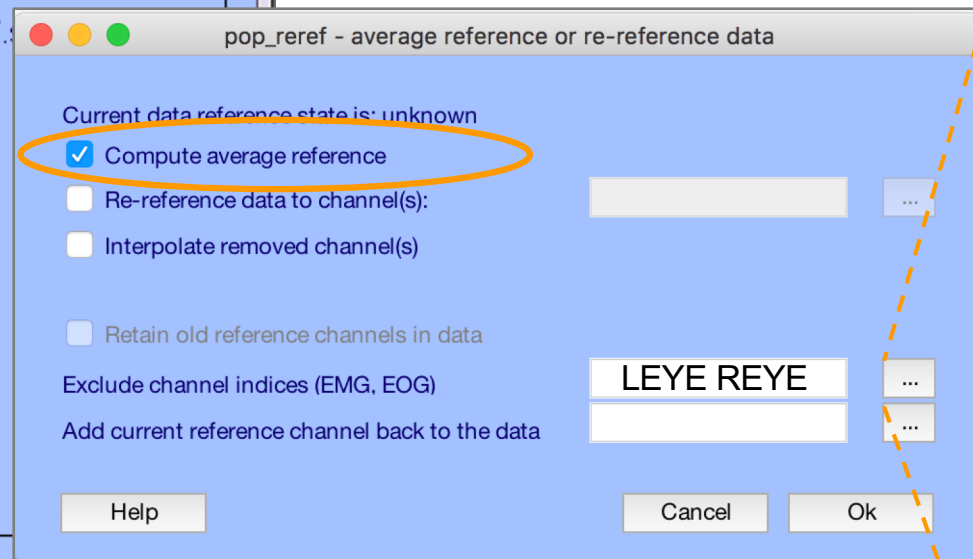
Multiple active datasets



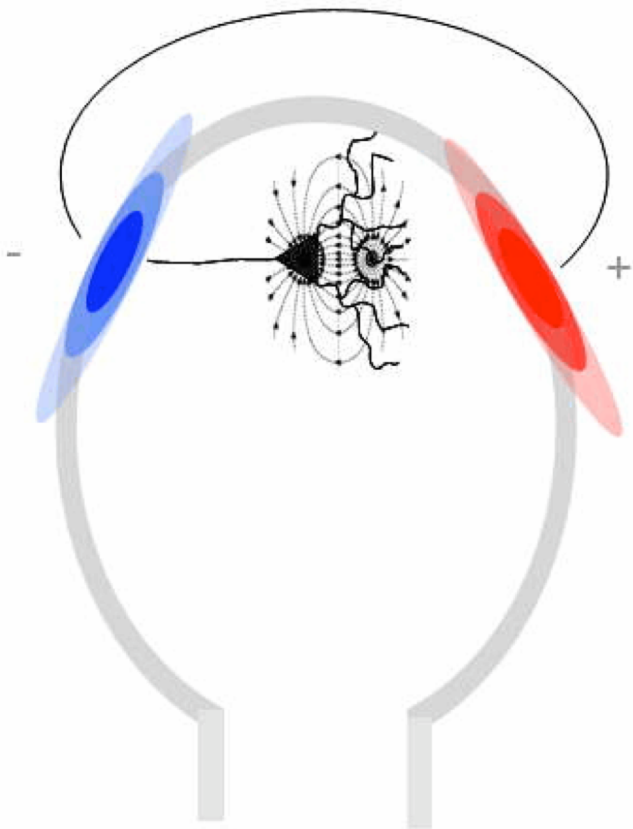
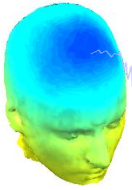
Re-reference data (if necessary/desired)



Or,
average reference

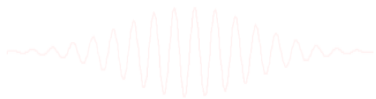


On Average Referencing

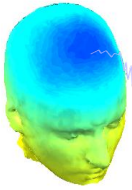


In theory, positive and negative current across entire head should balance—no net current source or sink: Average referencing enforces this.

In practice, depends on distribution of electrodes.

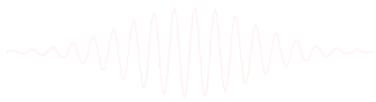
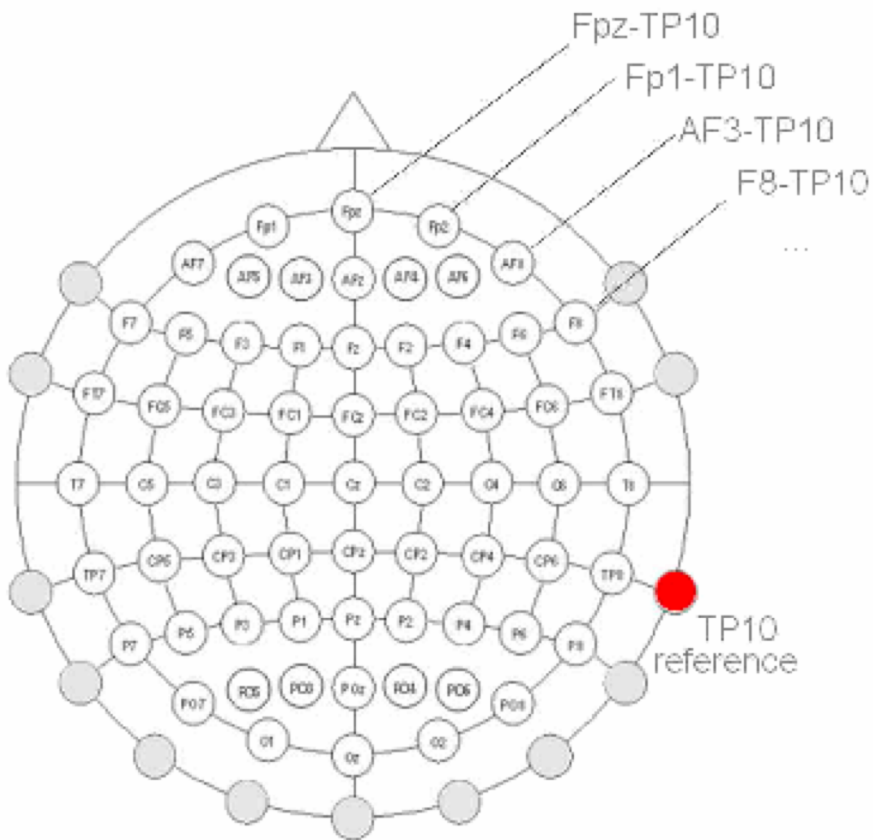


Average reference

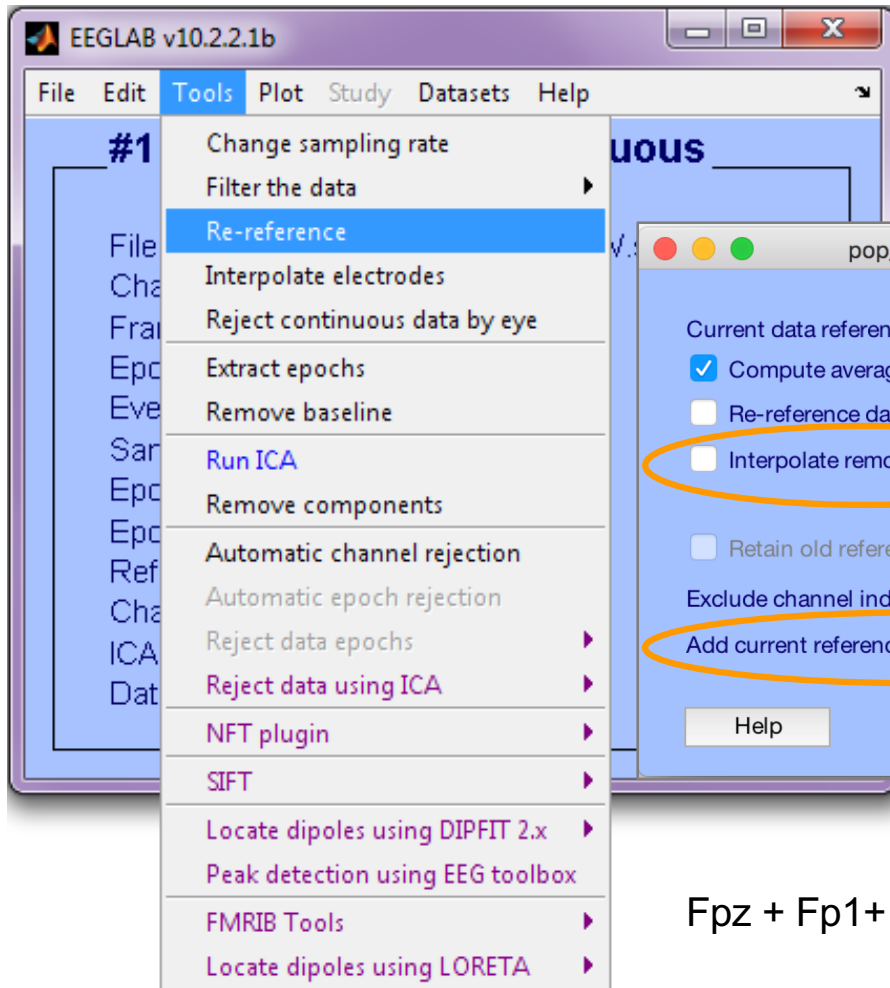
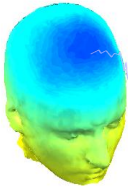


1. Average Reference assumption

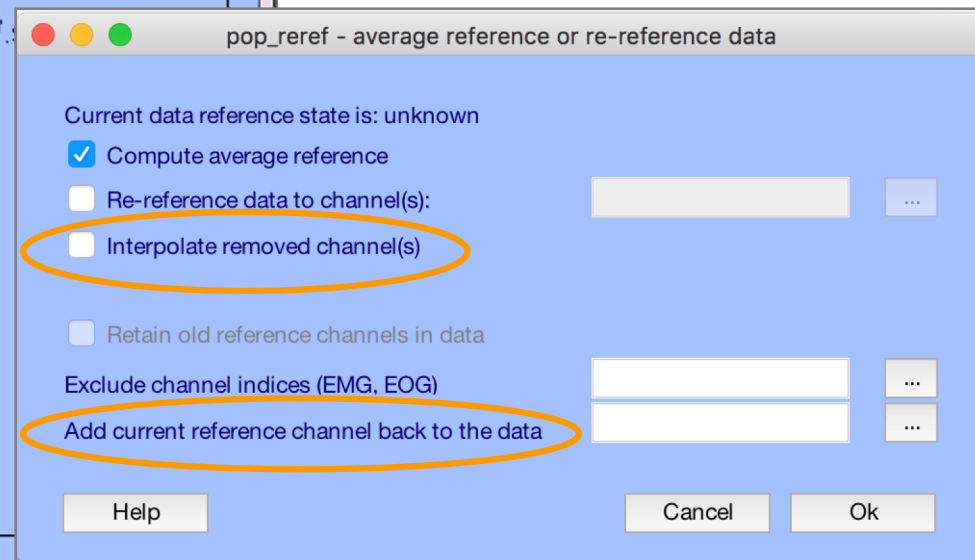
$$Fpz + Fp1 + AF3 + F8 + FT8 + \dots + TP10 = 0$$



Re-reference data (if necessary/desired)



average reference

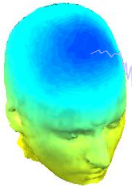


$Fpz + Fp1 + AF3 + F8 + FT8 + \dots + REF = 0$ (REF is a scalp channel)

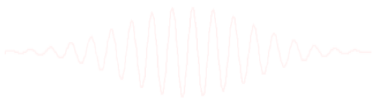
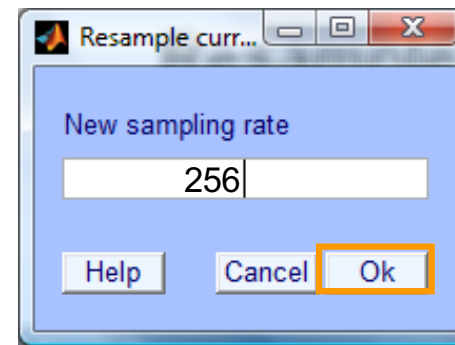
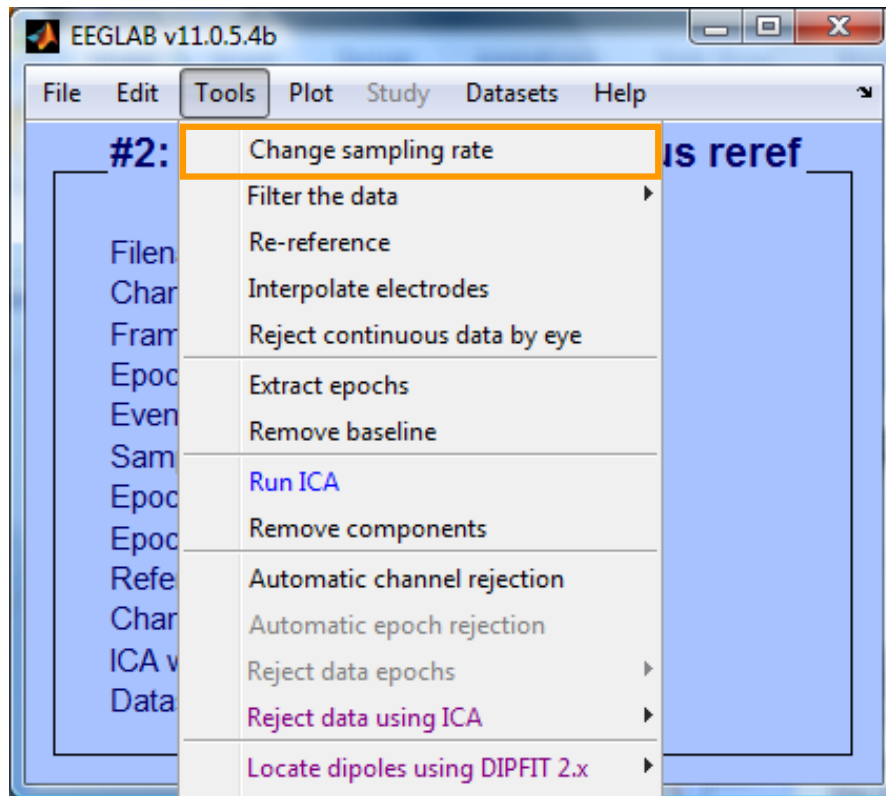
OR

$Fpz + Fp1 + AF3 + F8 + FT8 + \dots = 0$ (REF is not on the scalp, ear reference, ...)

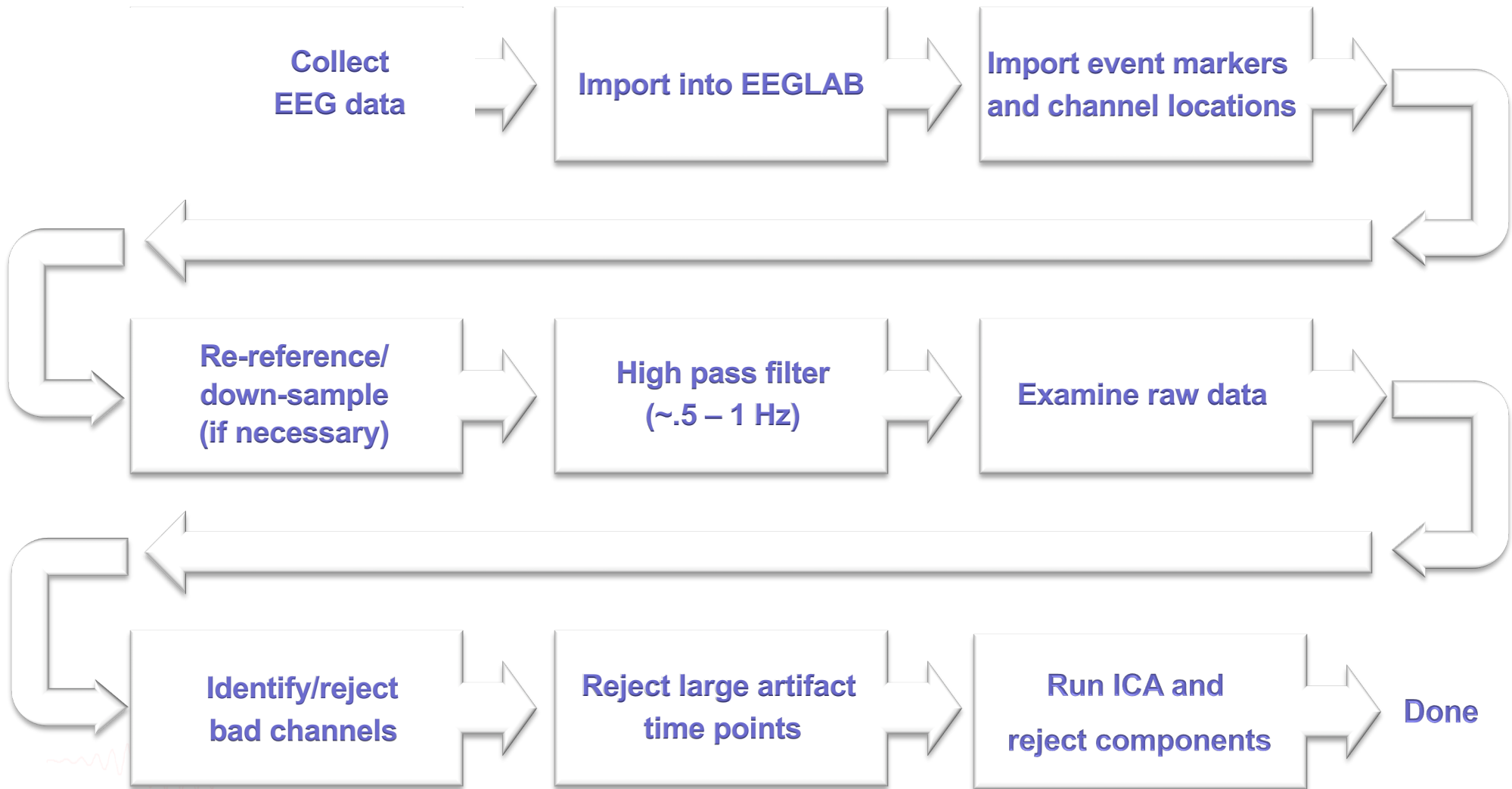
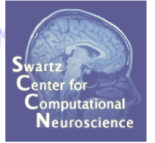
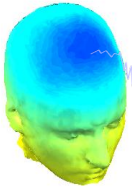
Resample data (if desired)



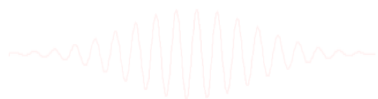
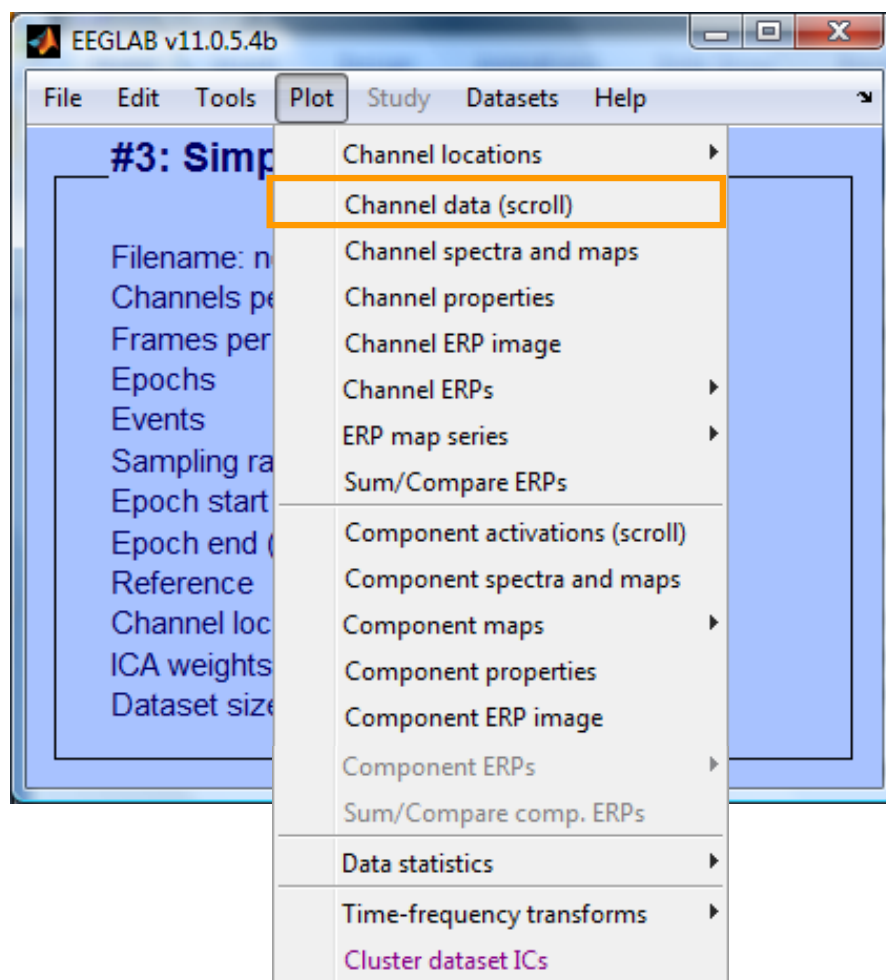
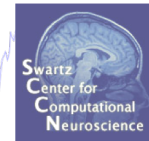
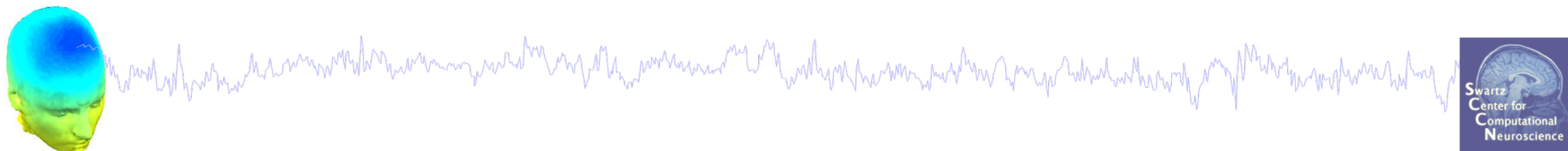
Reason: Reduce space, time.

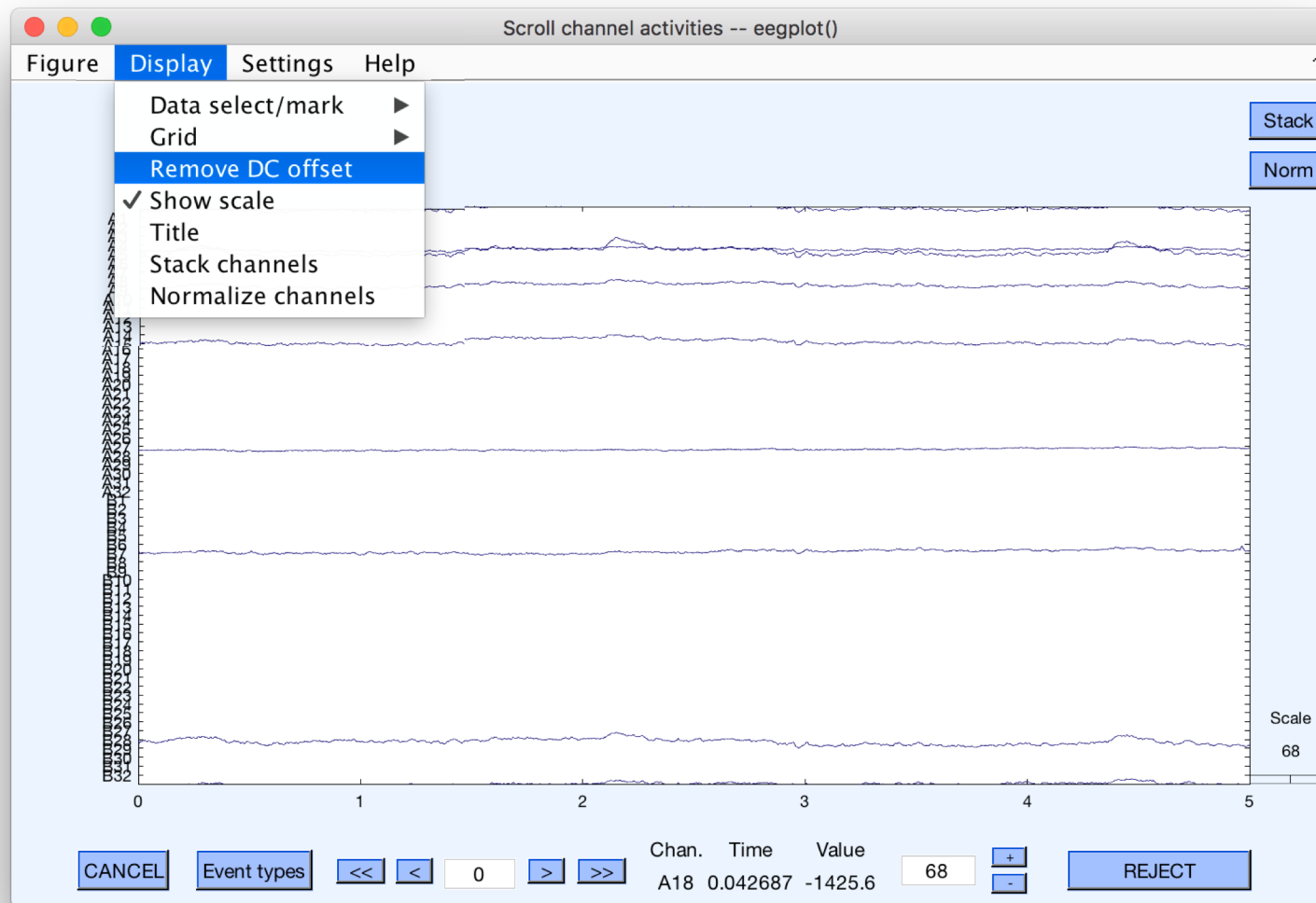
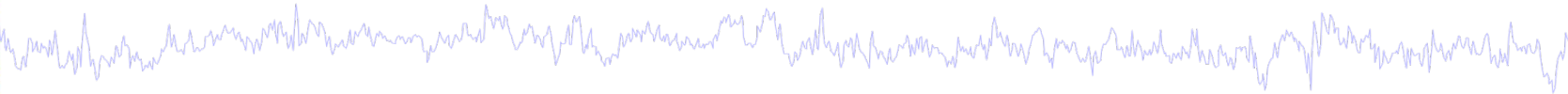
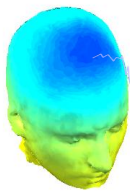


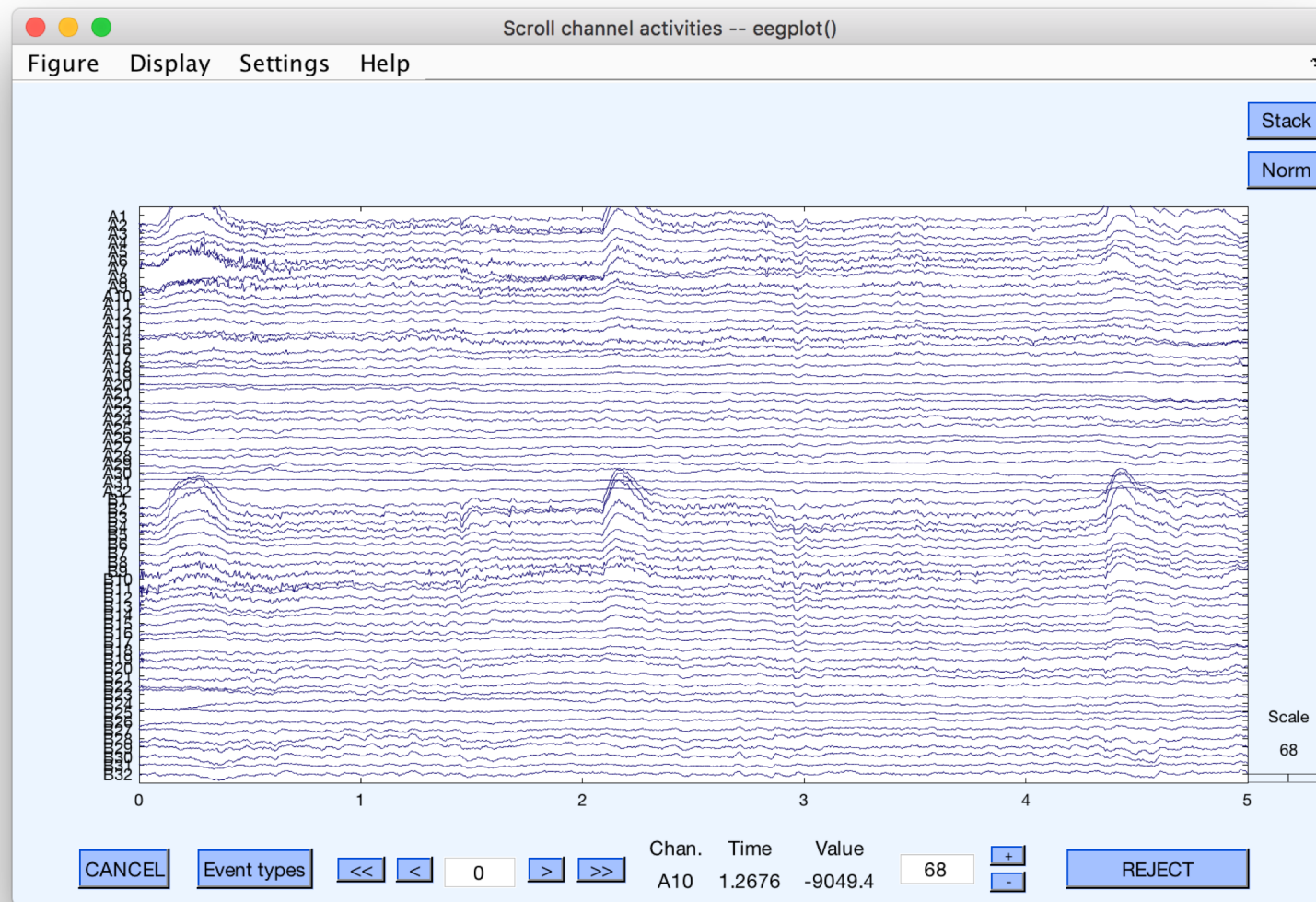
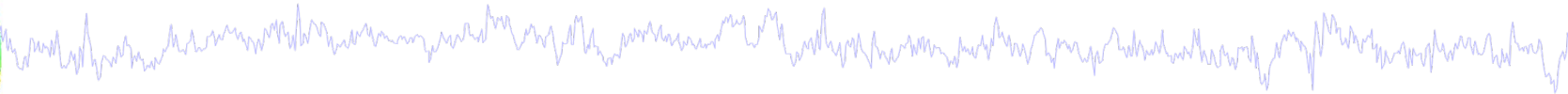
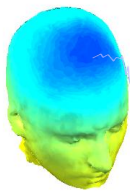
Pre-processing pipeline



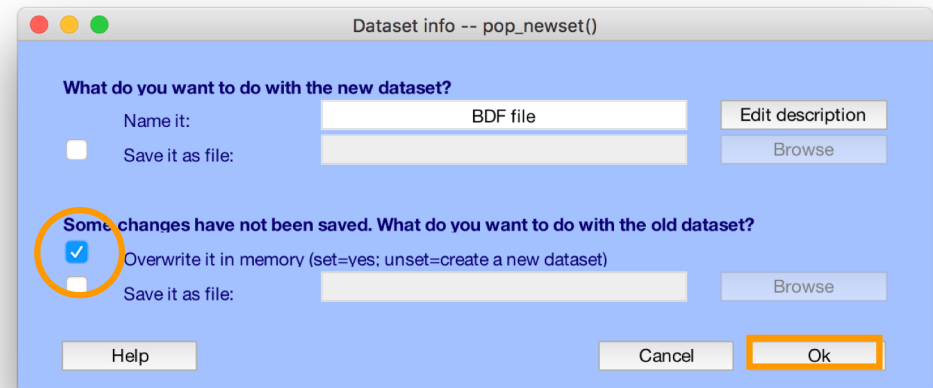
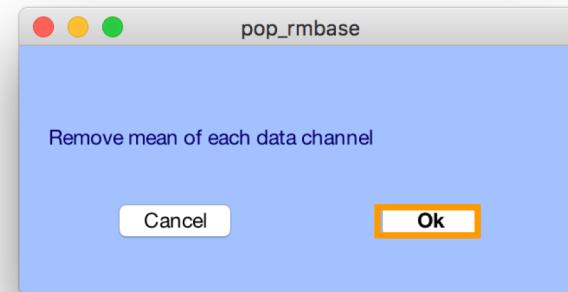
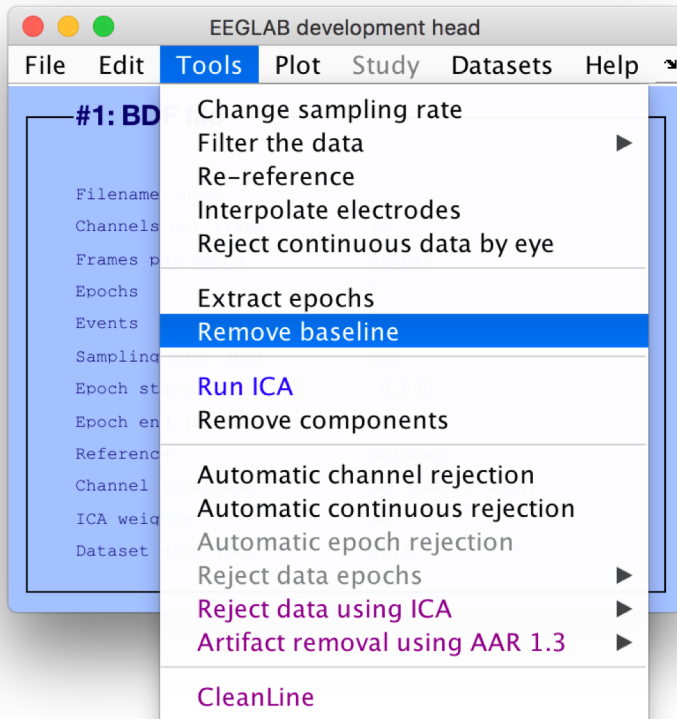
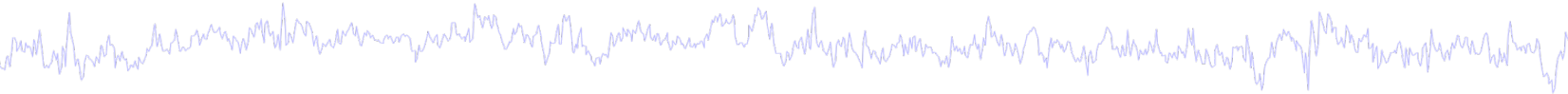
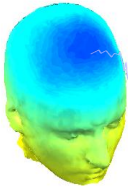
Scroll channel data





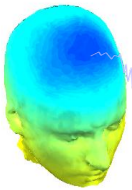


Remove DC offset



DC offsets introduce large filter artifact at signal boundaries, so it better to remove them prior to filter the signal.

High-Pass Filter the data



Reason: remove slow, possibly large amplitude, drift

The screenshot shows the EEGLAB v11.0.5.4b interface. The 'Tools' menu is open, highlighting 'Filter the data'. A sub-menu for 'Basic FIR filter' is visible, listing options: 'ERPLAB Butterworth Filter', 'ERPLAB Polynomial Detrending', and 'Short non-linear IIR filter'. The 'Filter the data -- pop_eegfiltnew()' dialog is open, showing the 'Lower edge of the frequency pass band (Hz)' set to 0.5. A red arrow points to this value with the text 'High-pass needed for ICA'. The 'Channel type(s)' and 'OR channel labels or indices' fields are highlighted with an orange box. The 'Dataset info -- pop_newset()' dialog is also open, showing options for saving the dataset. The 'Overwrite it in memory (set=yes; unset=create a new dataset)' option is checked and circled in orange. The 'Ok' button in the 'Dataset info' dialog is also highlighted with an orange box.

EEGLAB v11.0.5.4b

File Edit Tools Plot Study Datasets Help

#1: Change sampling rate
Filter the data
Re-reference
Interpolate electrodes
Reject continuous data by eye
Extract epochs
Remove baseline
Run ICA
Remove components
Automatic channel rejection
Automatic epoch rejection
Reject data epochs
Reject data using ICA
Locate dipoles using DIPFIT 2.x
Peak detection using EEG toolbox
FMRIB Tools
Locate dipoles using LORETA
CleanLine

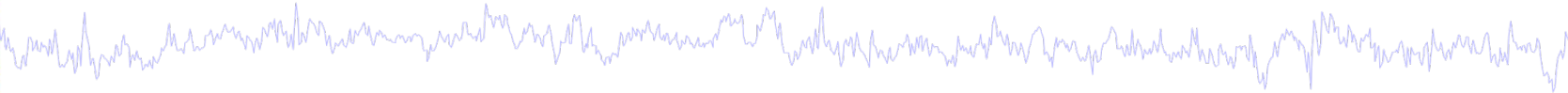
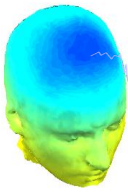
Basic FIR filter
ERPLAB Butterworth Filter
ERPLAB Polynomial Detrending
Short non-linear IIR filter

Filter the data -- pop_eegfiltnew()

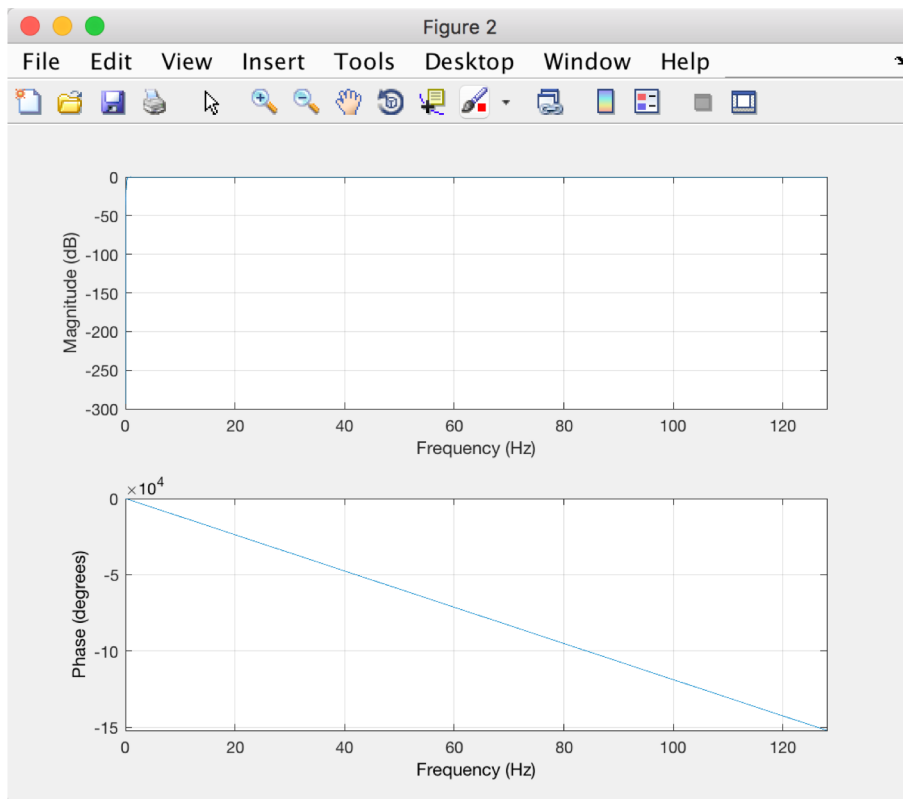
Lower edge of the frequency pass band (Hz) 0.5
Higher edge of the frequency pass band (Hz)
FIR Filter order (Mandatory even. Default is automatic*)
See help text for a description of the default filter order. Manual definition is recommended.
☐ Notch filter the data instead of pass band
☐ Use minimum-phase converted causal filter (non-linear)
☒ Plot frequency response
Channel type(s)
OR channel labels or indices
Help Cancel Ok

Dataset info -- pop_newset()

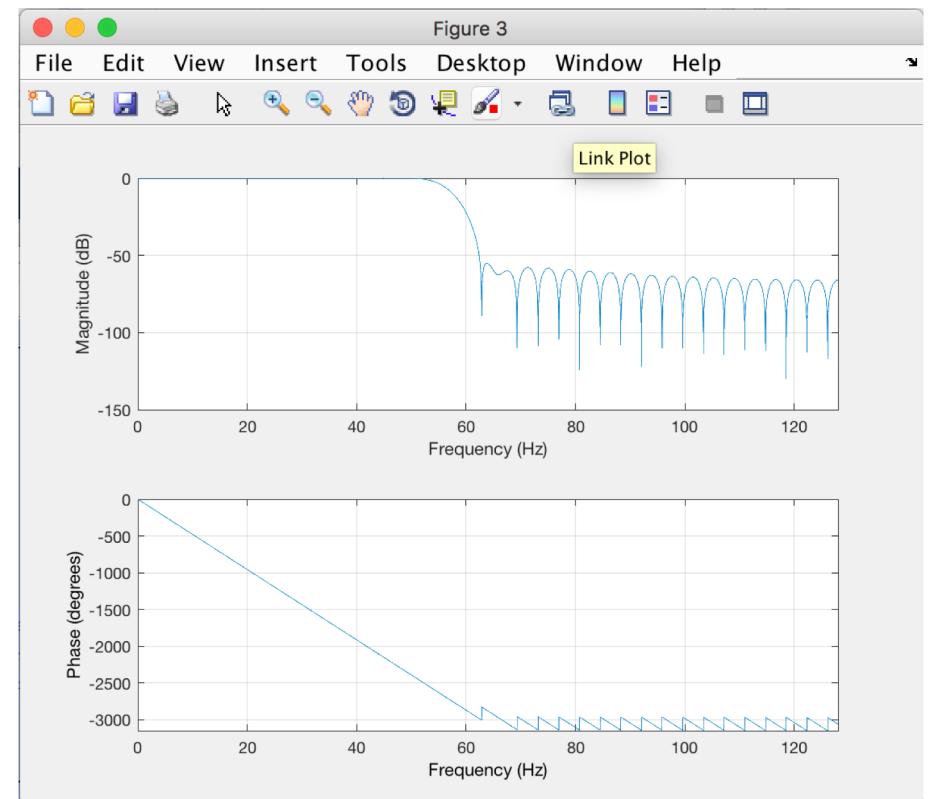
What do you want to do with the new dataset?
Name it: BDF file Edit description
☐ Save it as file: Browse
Some changes have not been saved. What do you want to do with the old dataset?
☒ Overwrite it in memory (set=yes; unset=create a new dataset)
☐ Save it as file: Browse
Help Cancel Ok

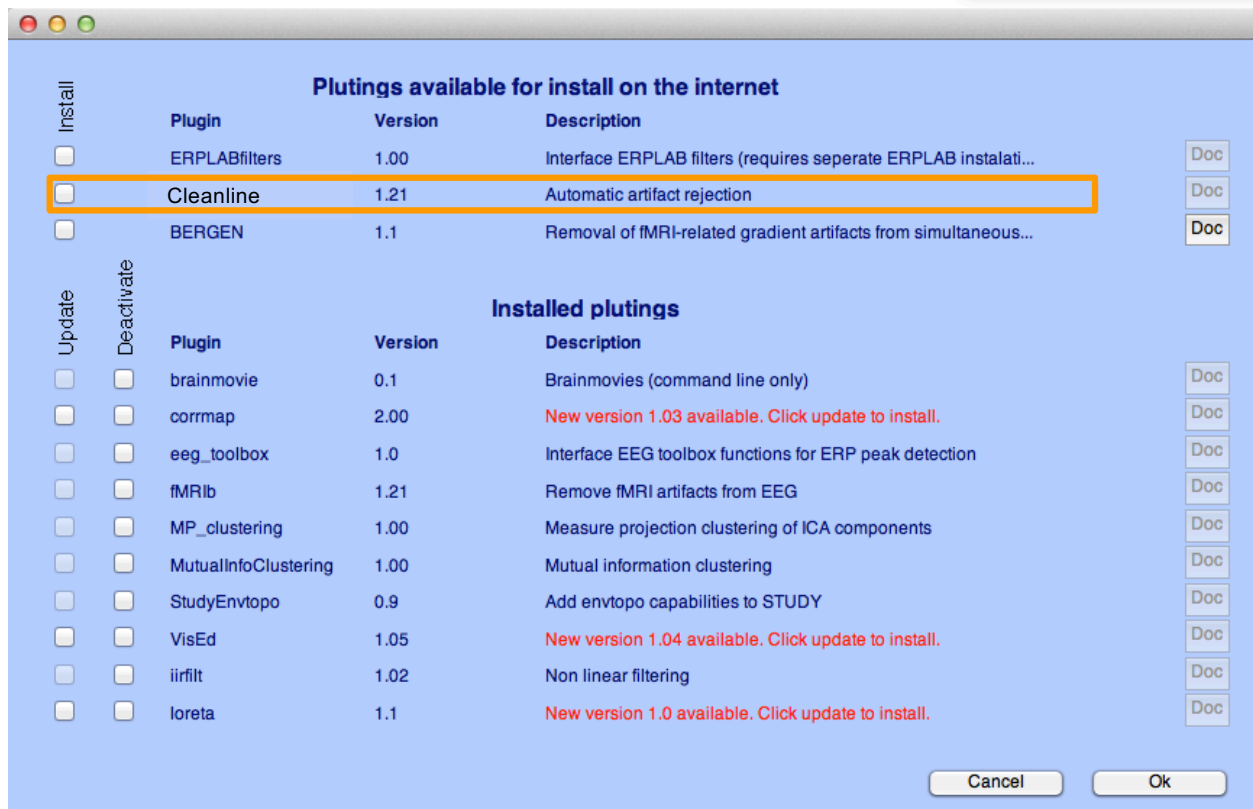
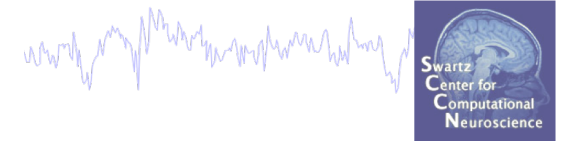
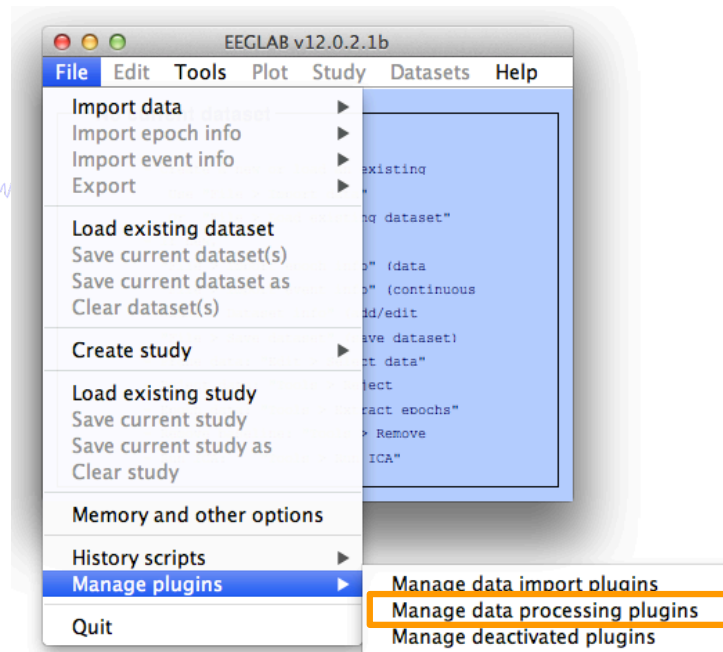
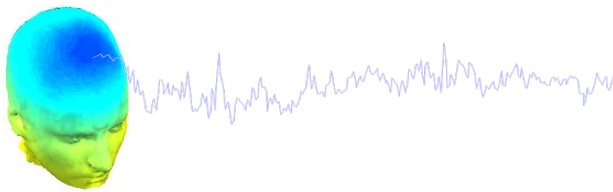


High pass (0.5 Hz)



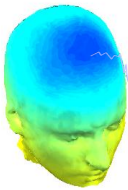
Low pass (50 Hz)





CleanLine uses an approach for line noise removal advocated by Partha Mitra and Hemant Bokil in "Observed Brain Dynamics" (2007), Chapter 7.3.4.

Remove line noise (Cleanline)



EEGLAB v11.0.5.4b

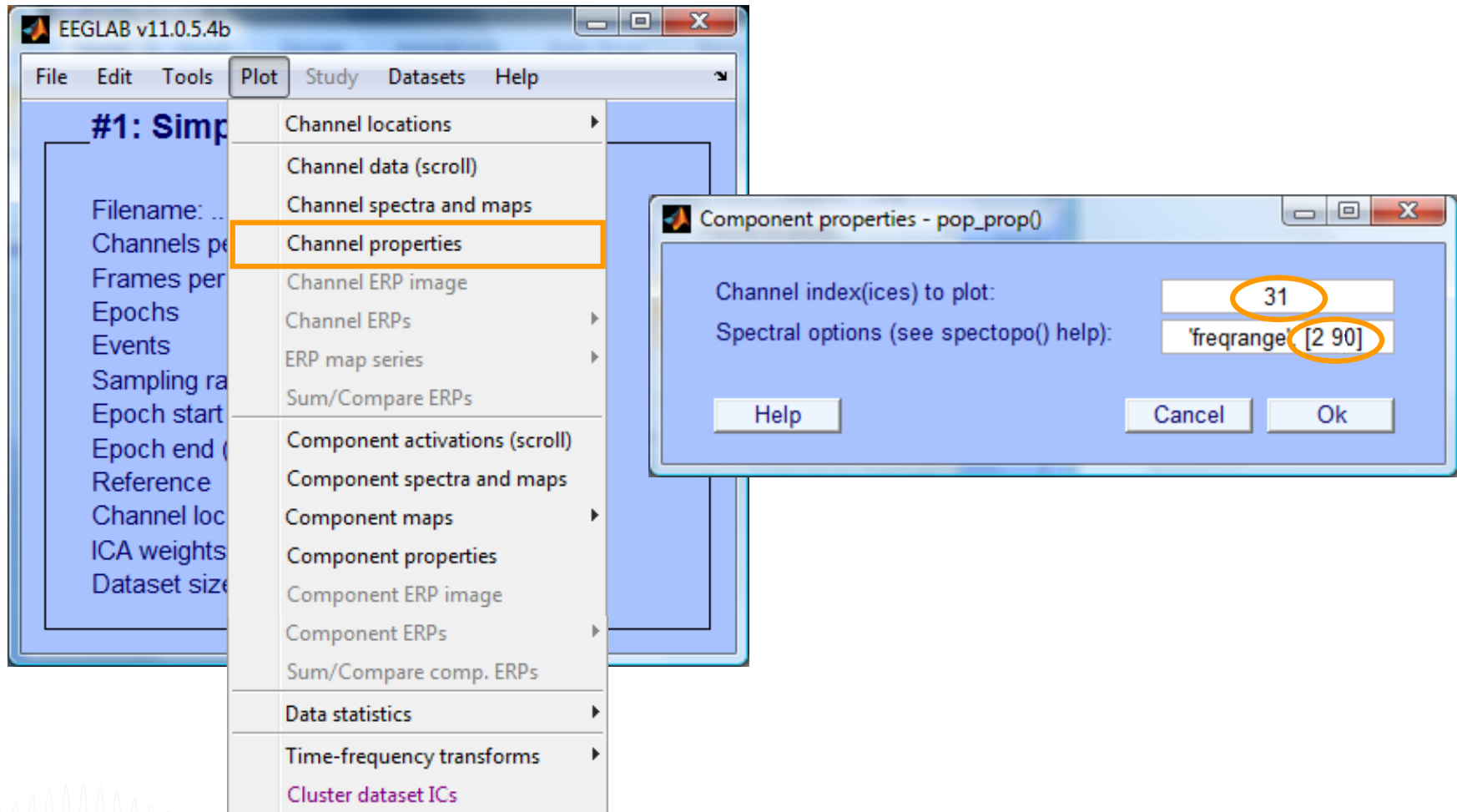
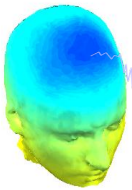
File Edit **Tools** Plot Study Datasets Help

#1: Change sampling rate
Filter the data
Re-reference
Interpolate electrodes
Reject continuous data by eye
Extract epochs
Remove baseline
Run ICA
Remove components
Automatic channel rejection
Automatic epoch rejection
Reject data epochs
Reject data using ICA
Locate dipoles using DIPFIT 2.x
Peak detection using EEG toolbox
FMRIB Tools
Locate dipoles using LORETA
CleanLine

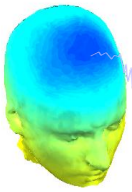
CleanLine Options

Line noise frequencies to remove [60 120]
Scan for line noise ☒ (set)
p-value for detection of significant sinusoid 0.01
Bandwidth (Hz) 2
Type of signal to clean Channels
Indices of Channels/Components to clean '1:66'
Sliding window length (sec) 4
Sliding window step size (sec) 2
Window overlap smoothing factor 100
FFT padding factor 2
Visualize Original and Cleaned Spectra ☐ (set)
Normalize log spectrum by detrending ☐ (set)
Produce verbose output ☒ (set)
Plot Individual Figures ☐ (set)
check
Help Cancel Ok

Plot channel properties



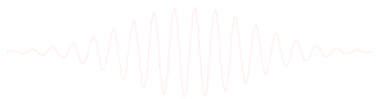
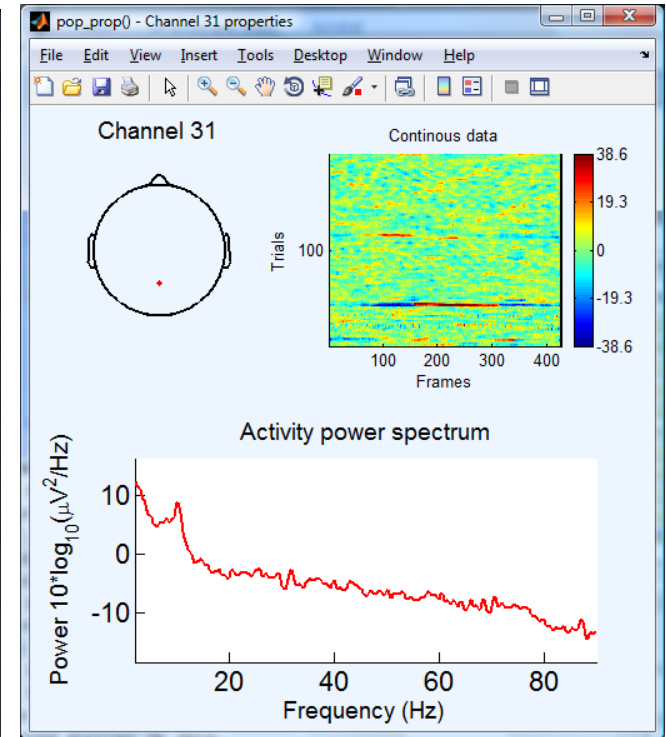
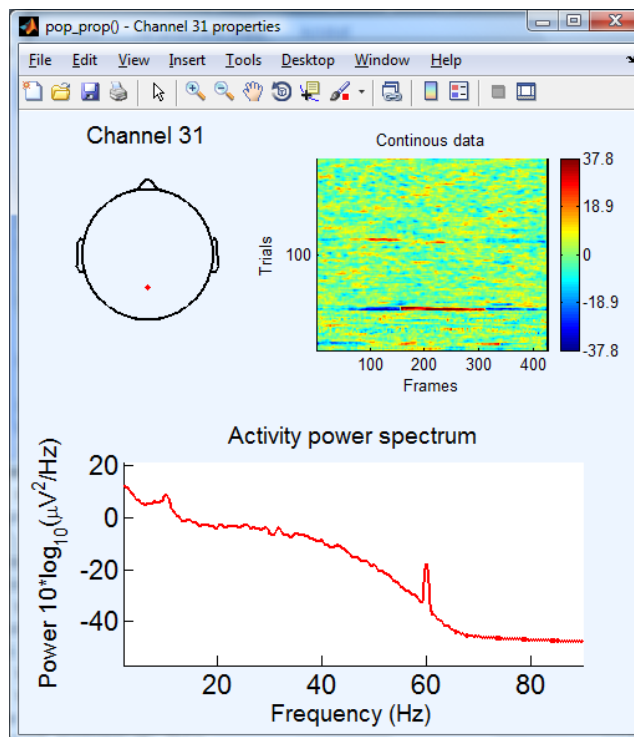
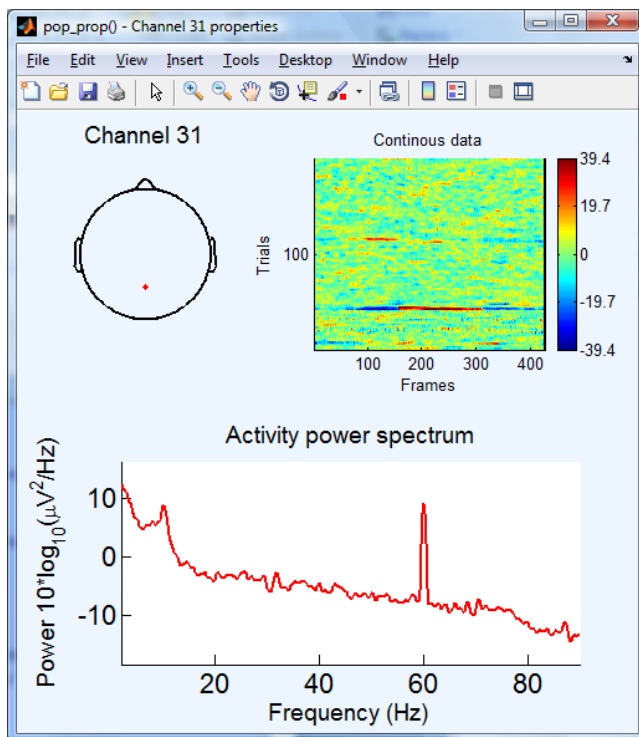
Filter comparisons



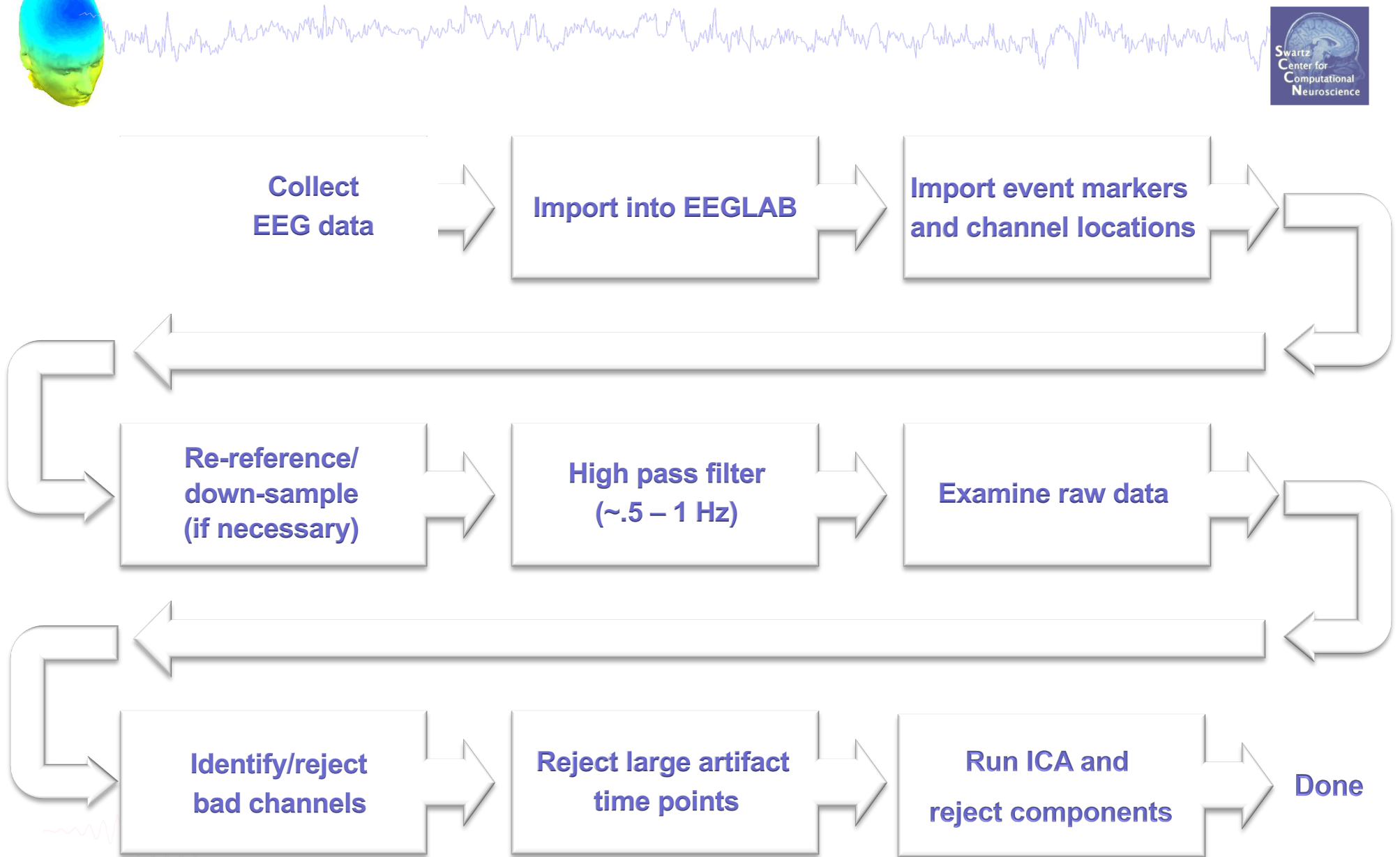
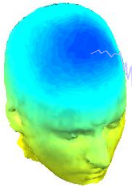
0.5 Hz high-pass filter

0.5 Hz high-pass filter
50 Hz low-pass filter

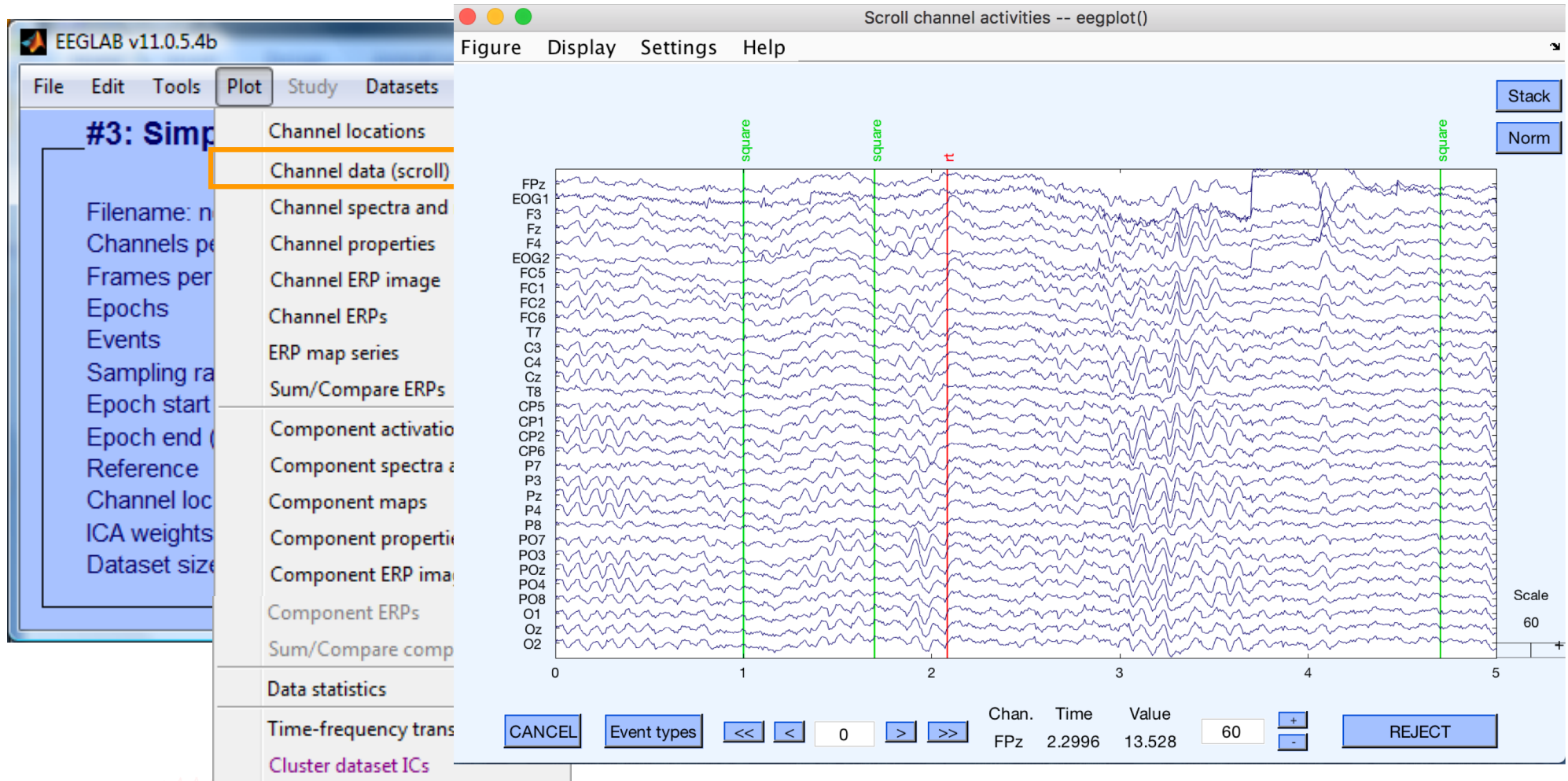
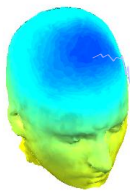
0.5 Hz high-pass filter
Cleanline



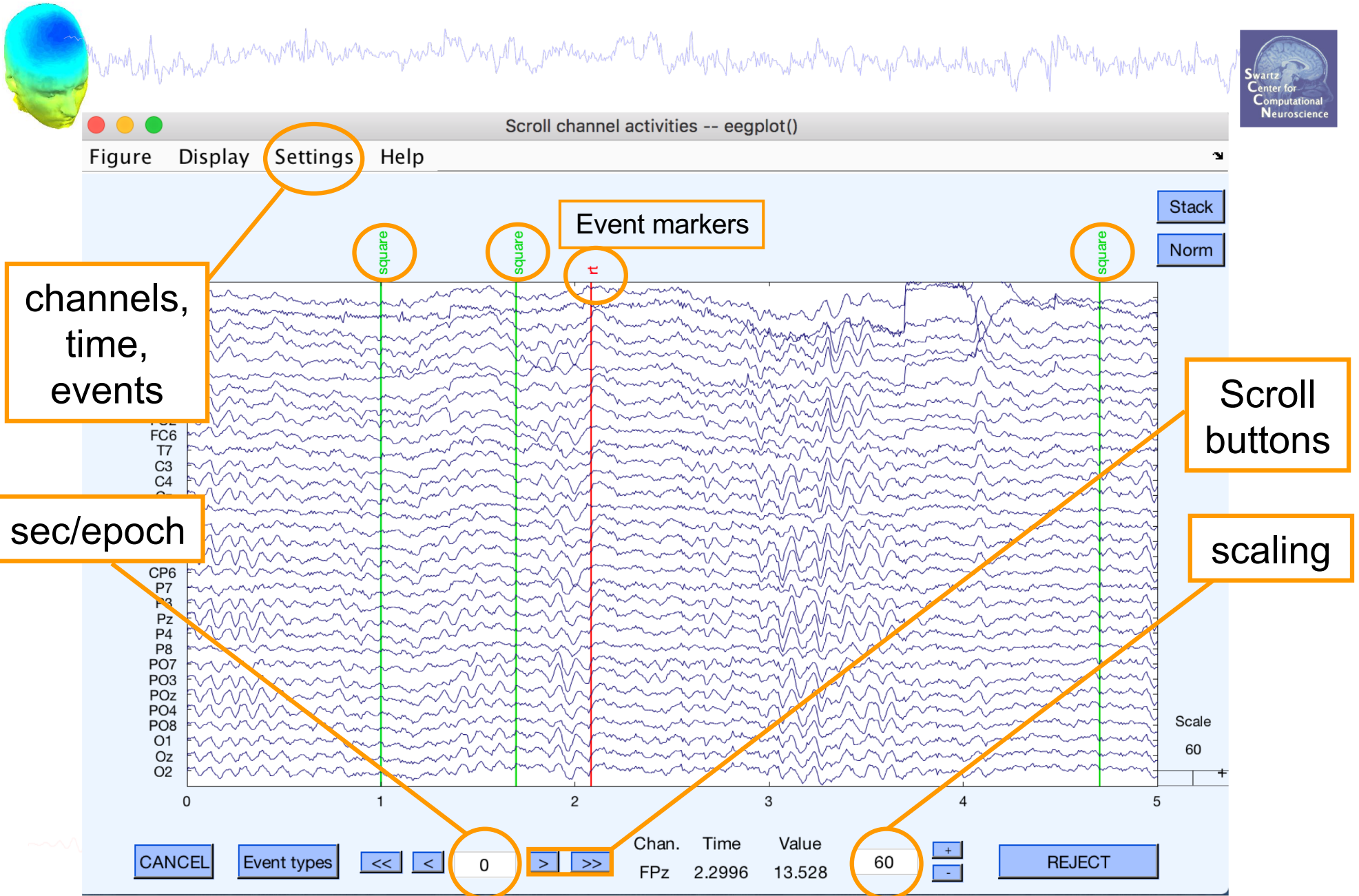
Pre-processing pipeline



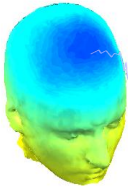
Scroll channel data



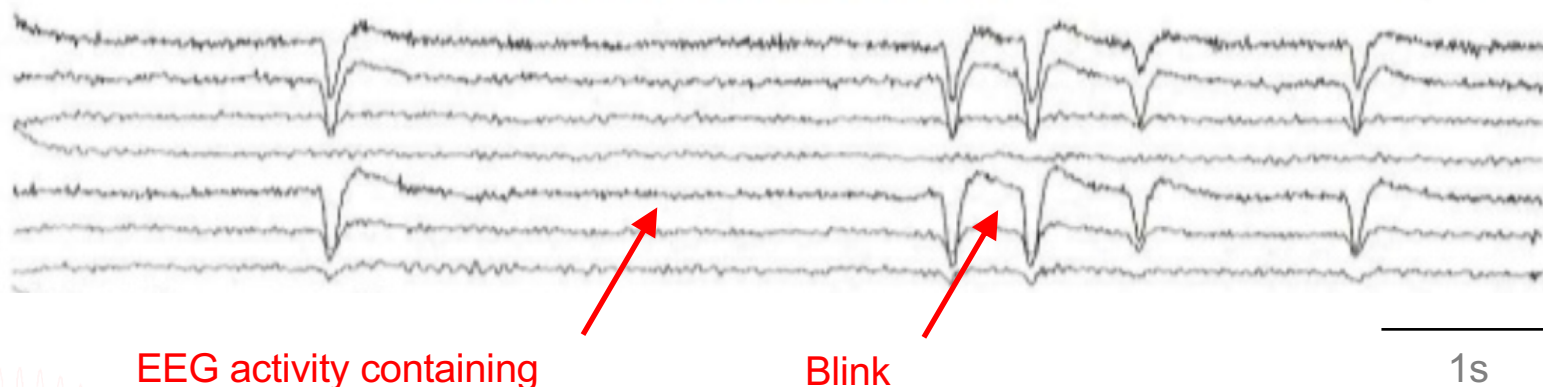
Scroll channel data



EEG artifacts



The amplitude of artifacts (such as eye movements) is often larger than the amplitude of brain data which potentially decrease signal/noise ratio, bias data analysis and potential results

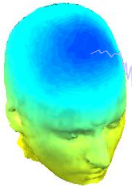


EEG activity containing
potential brain data

Blink

1s

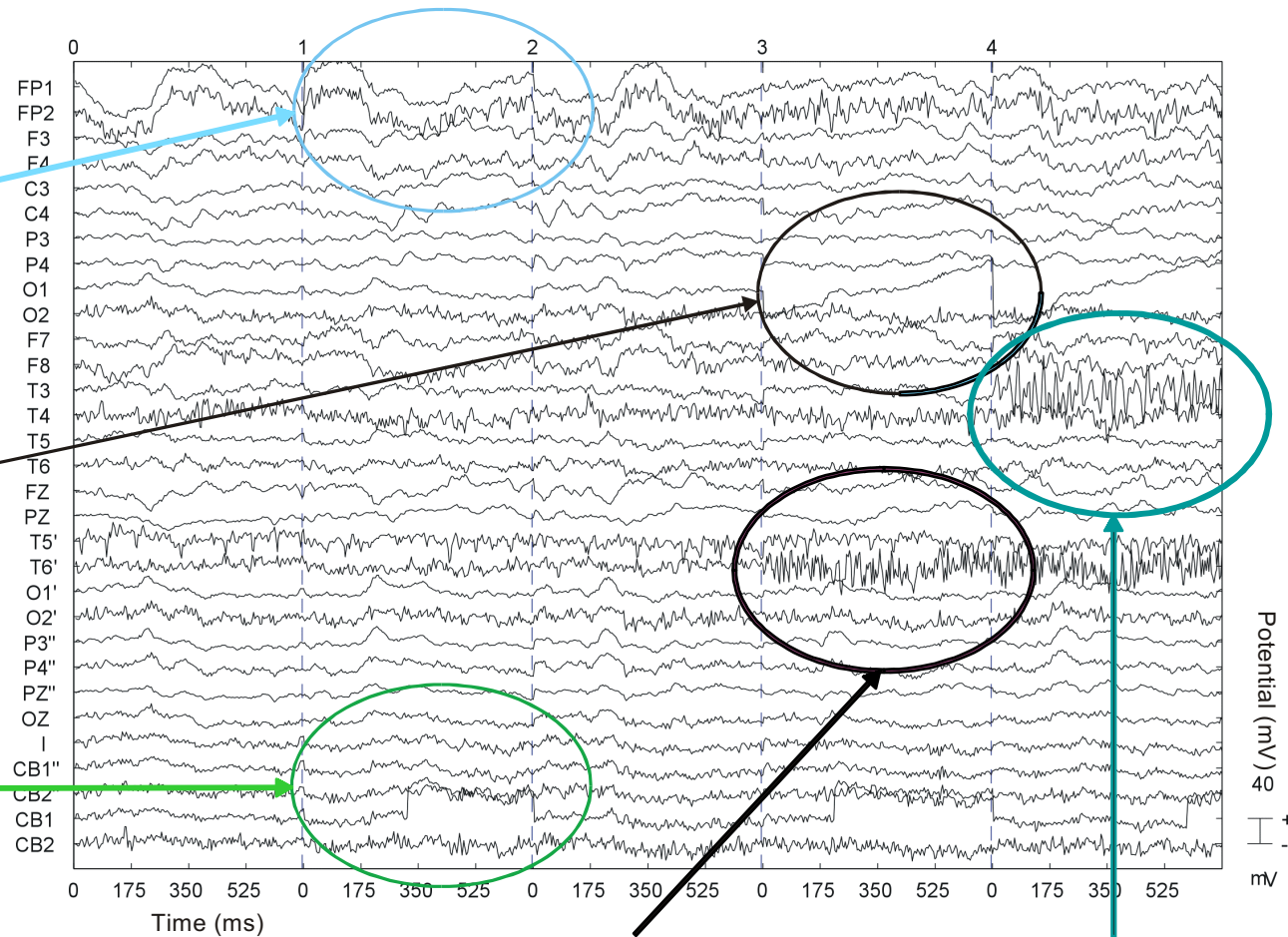
Type of artifacts



2 - Low frequency
event (eye
movements)

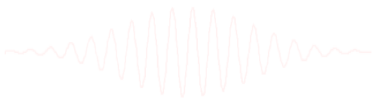
5 - Linear trend

3 - Discontinuity

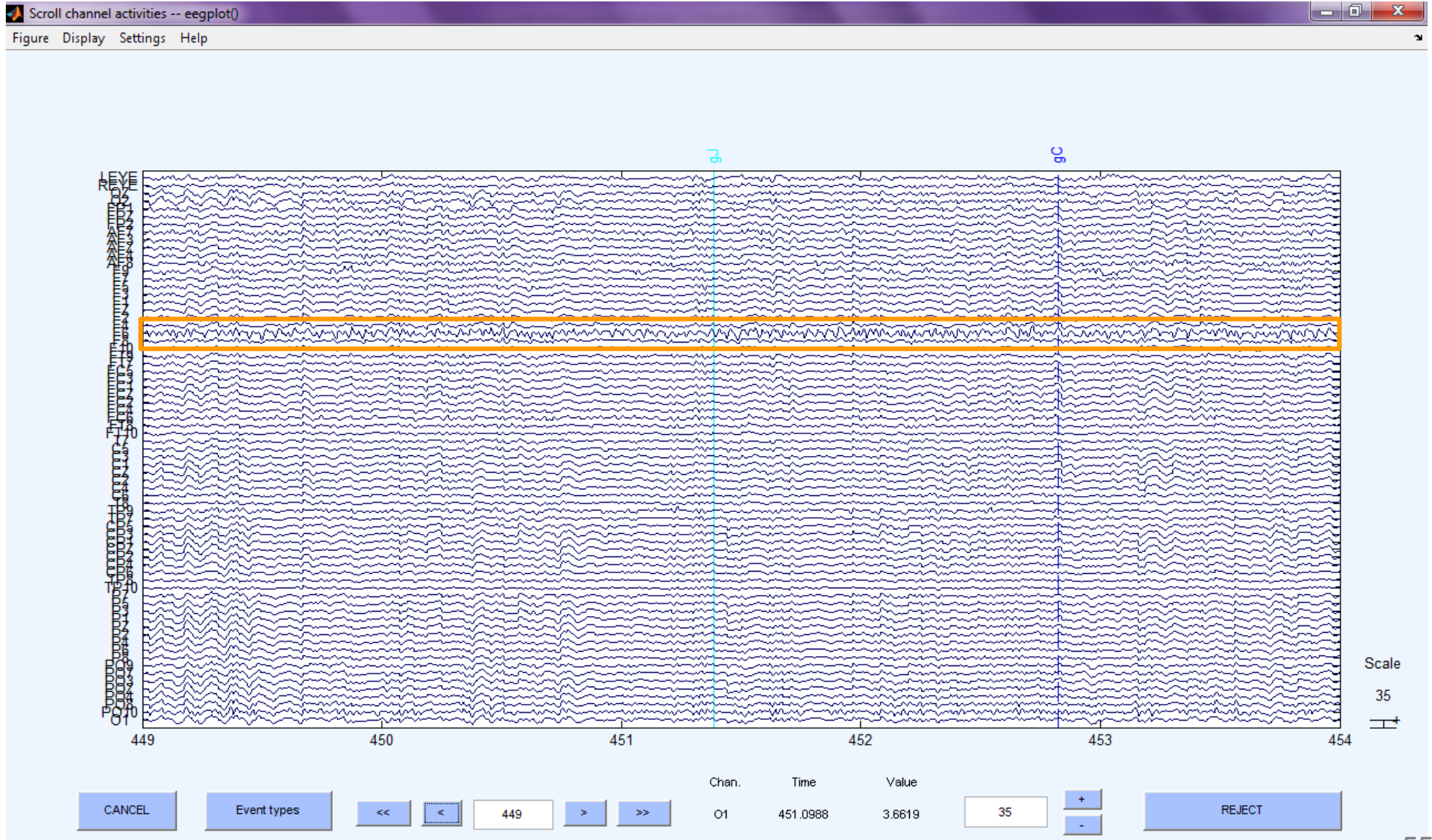
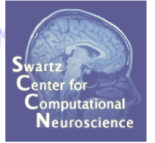
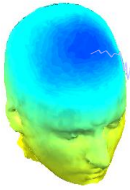


4 - High noise

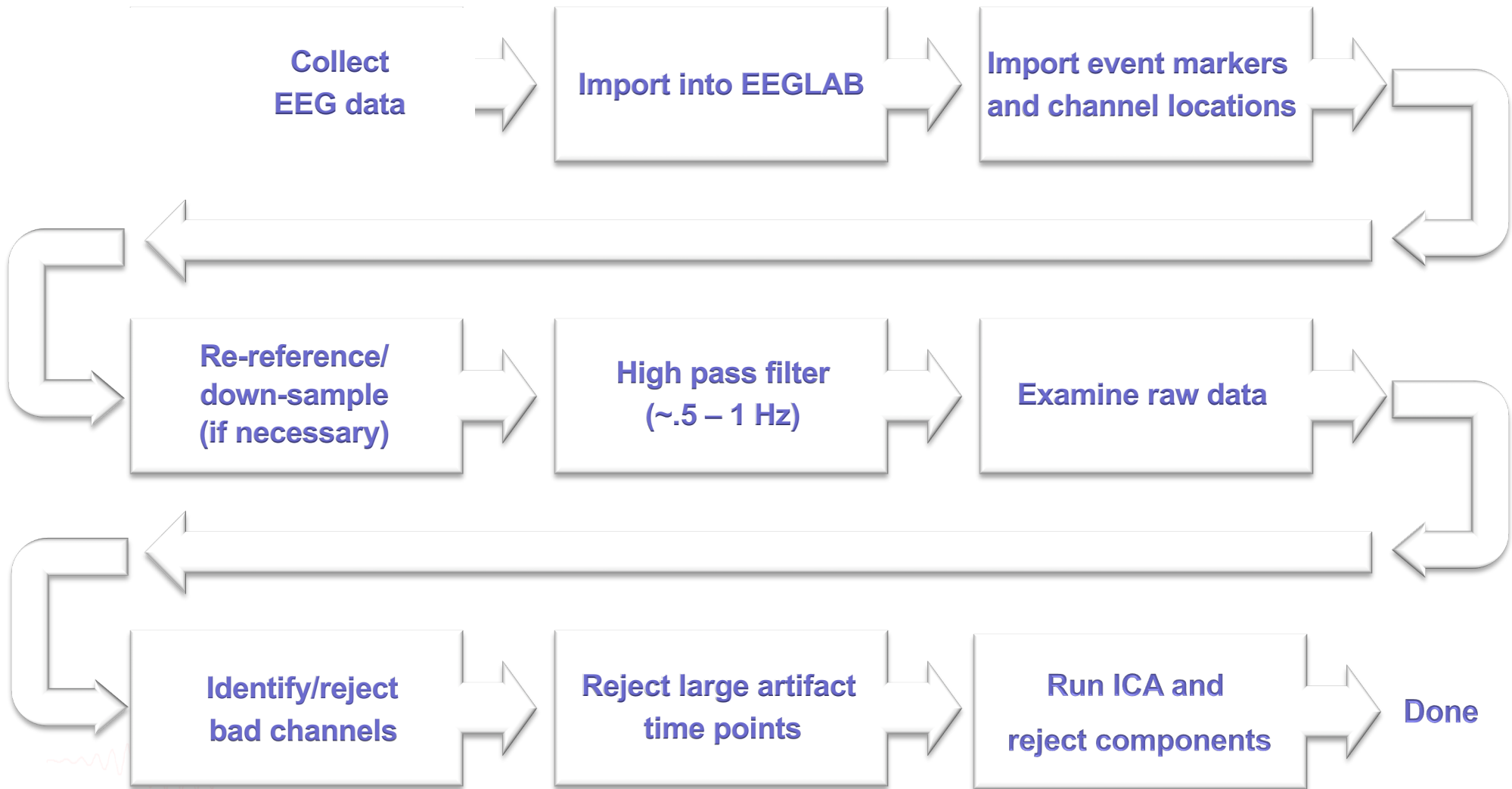
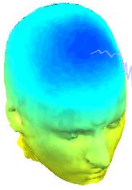
1 - Transient high frequency
event (muscle)



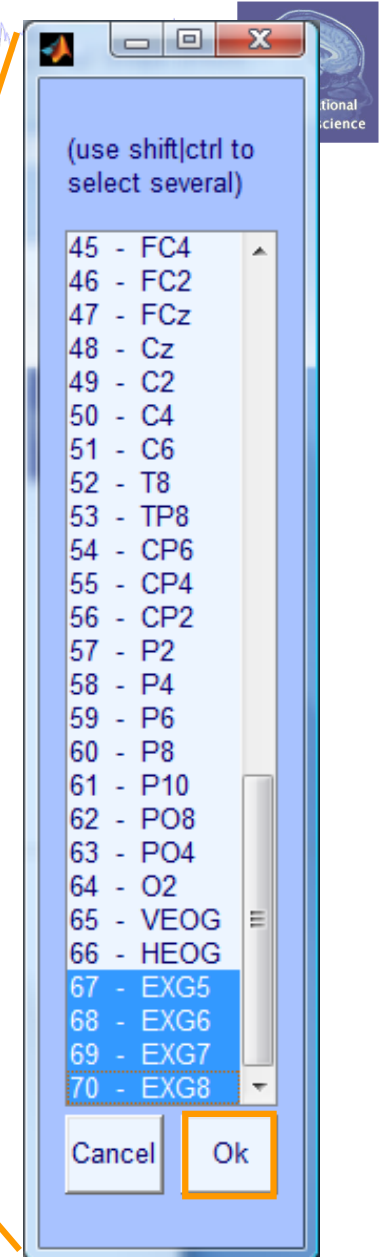
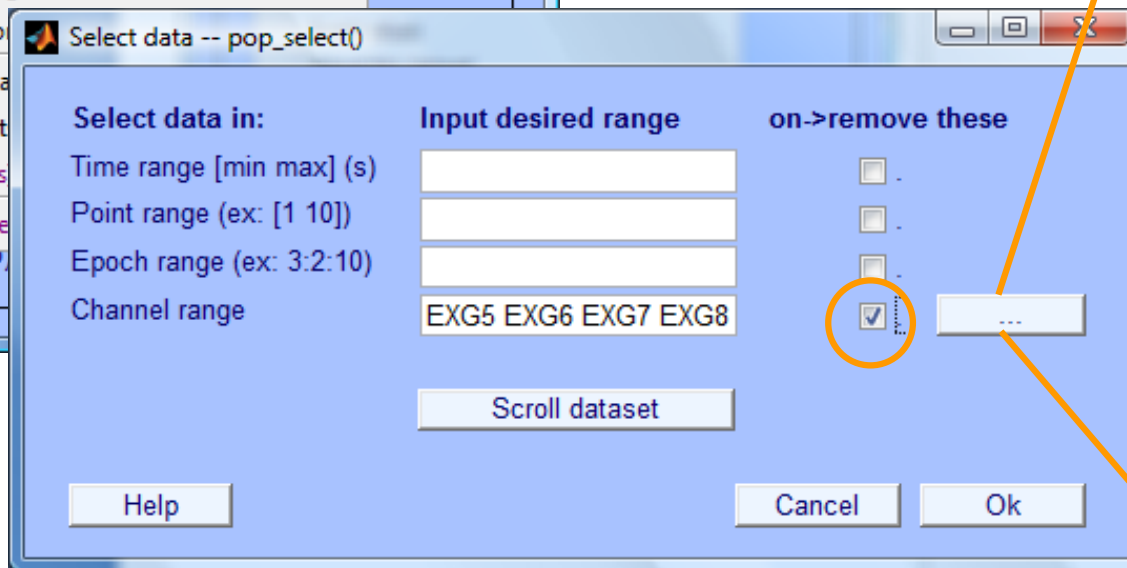
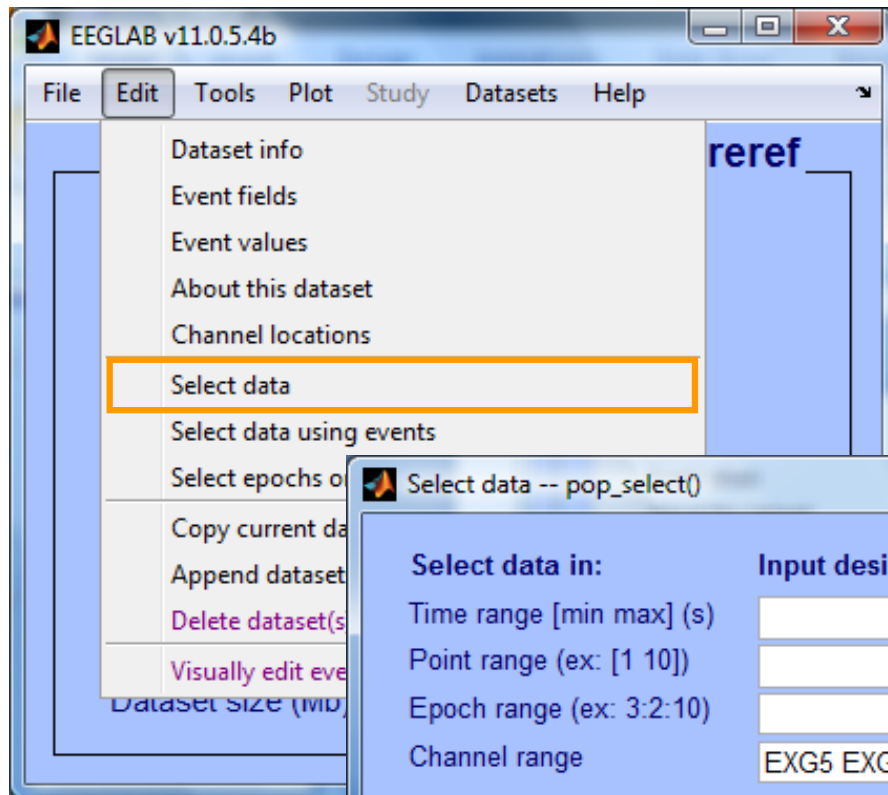
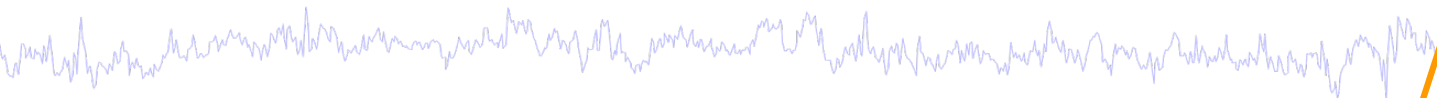
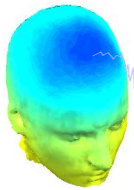
Looking for bad channels



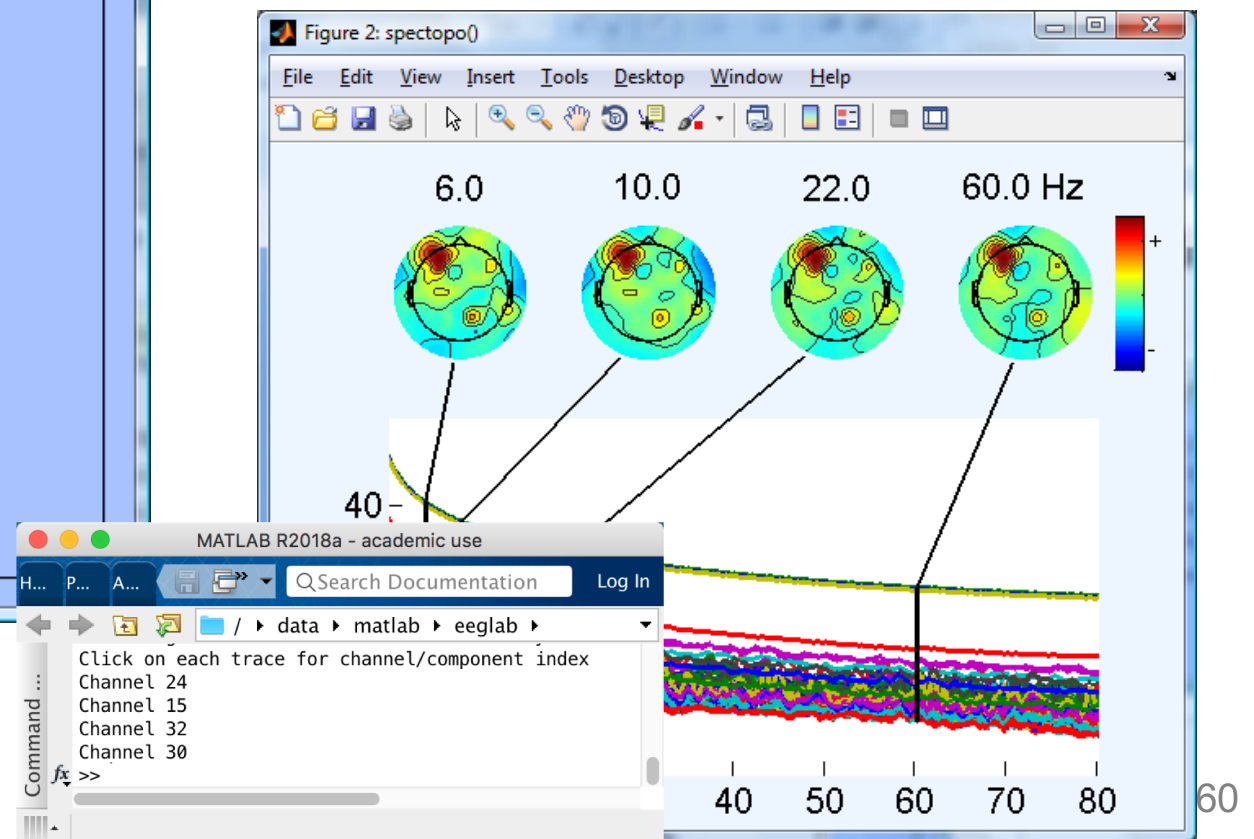
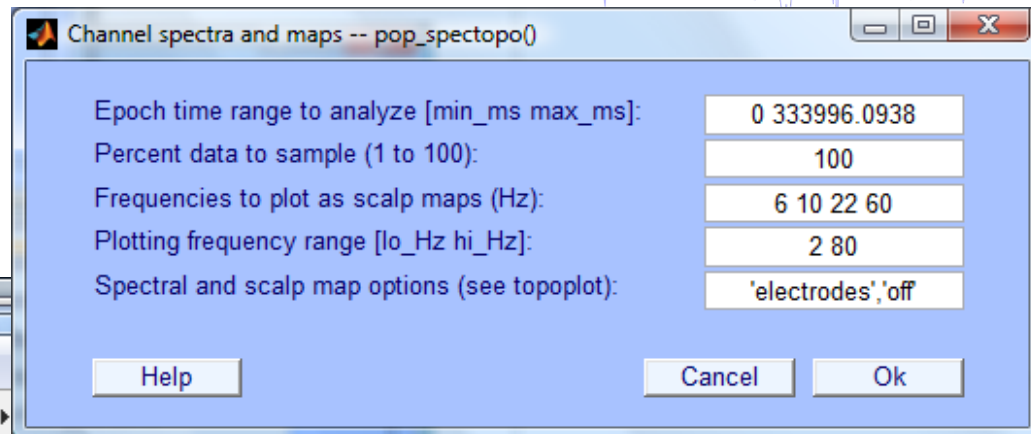
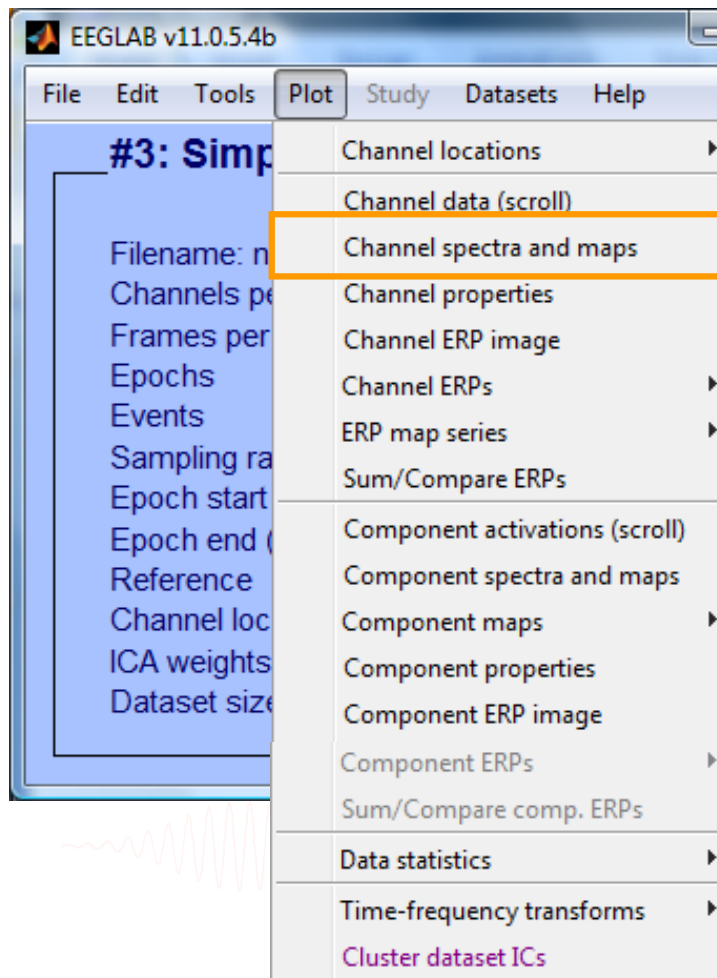
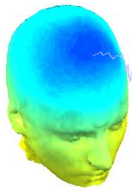
Pre-processing pipeline



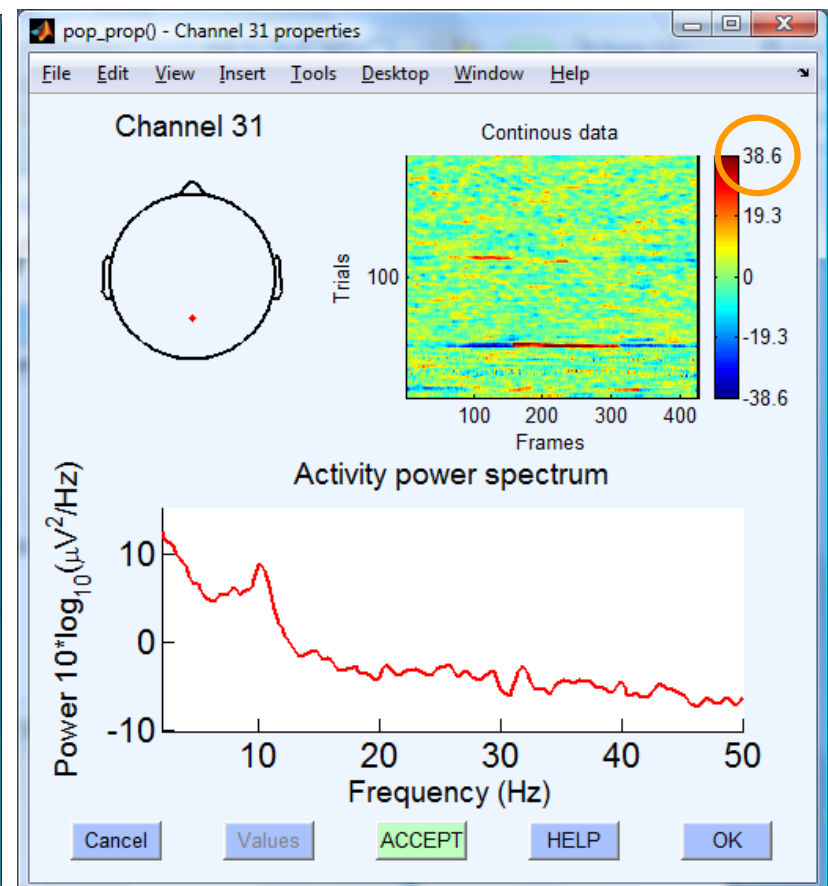
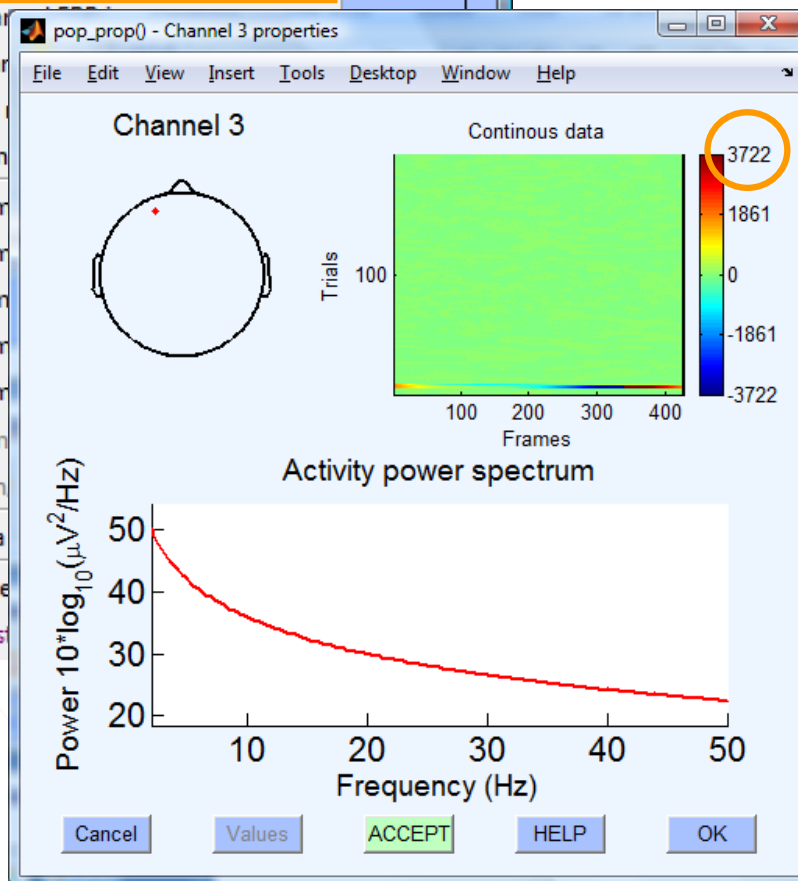
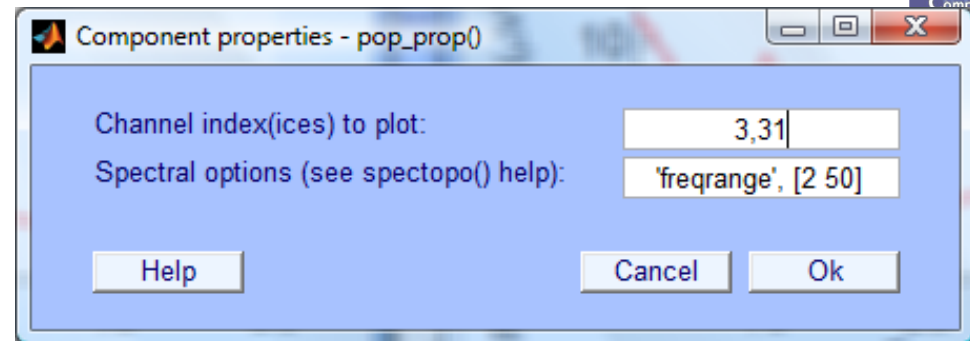
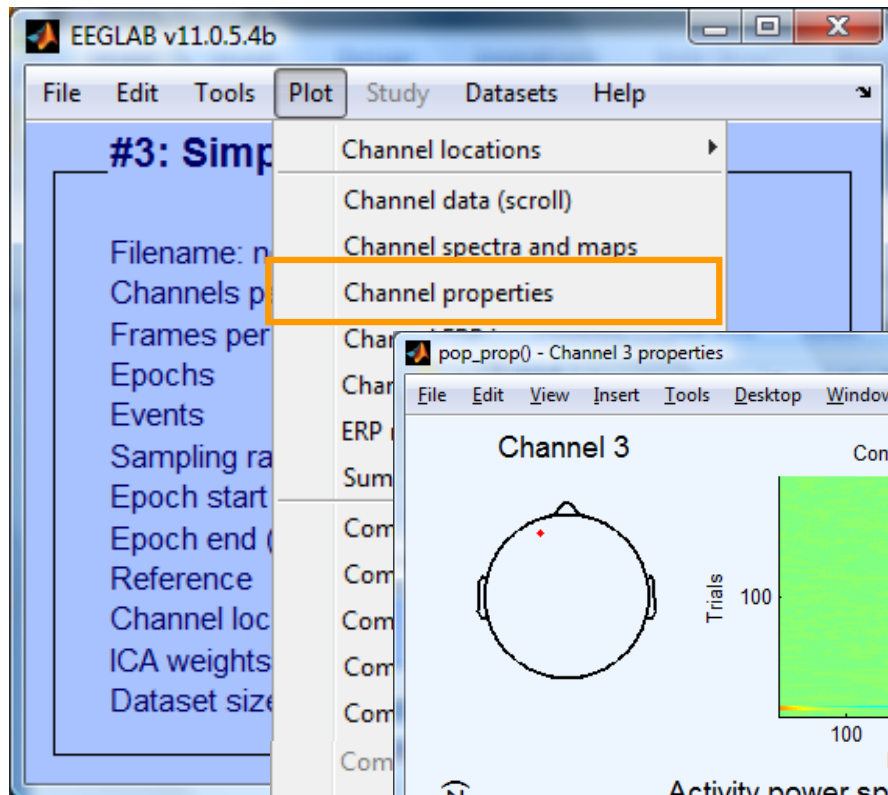
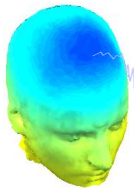
Remove unwanted channels



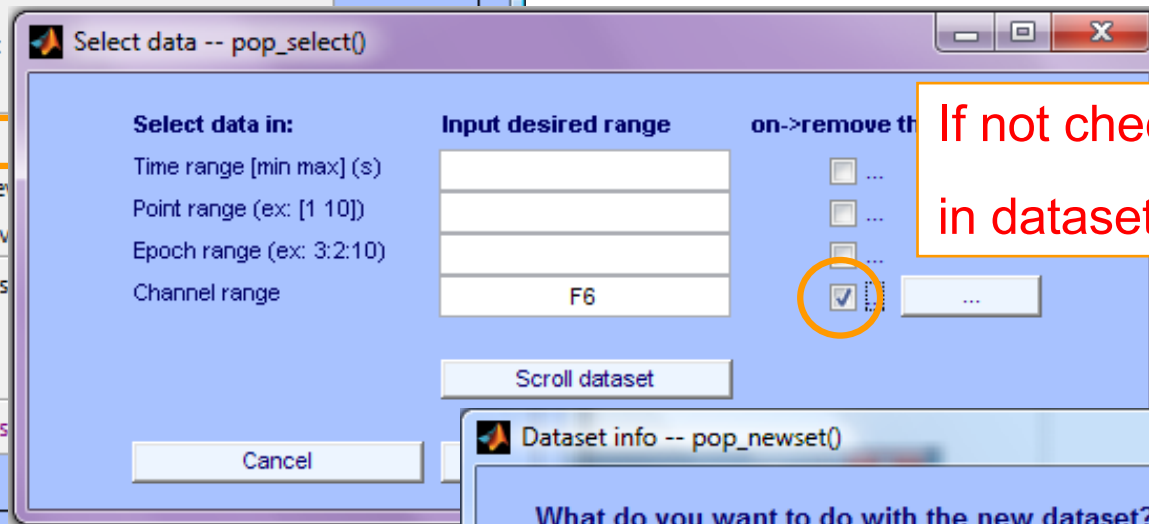
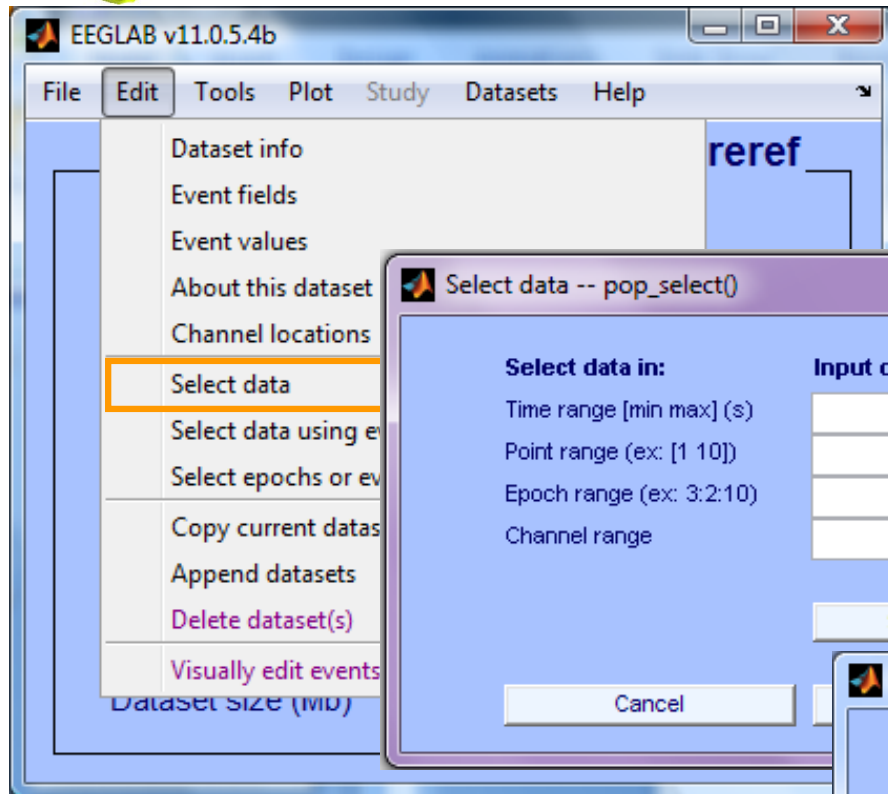
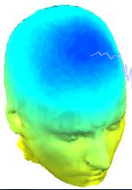
Manually identifying bad channels



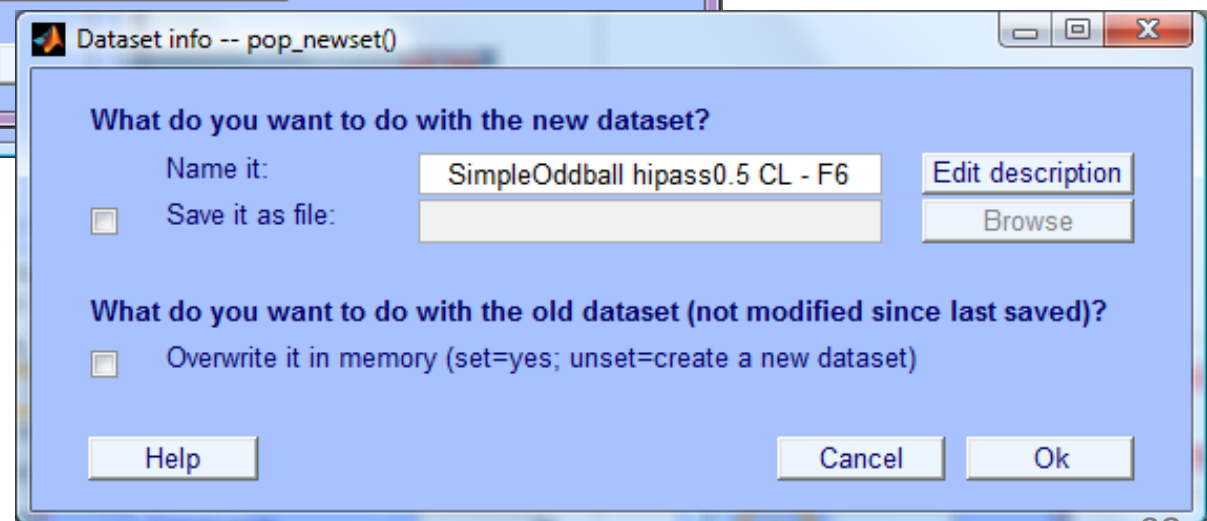
Manually identifying bad channels



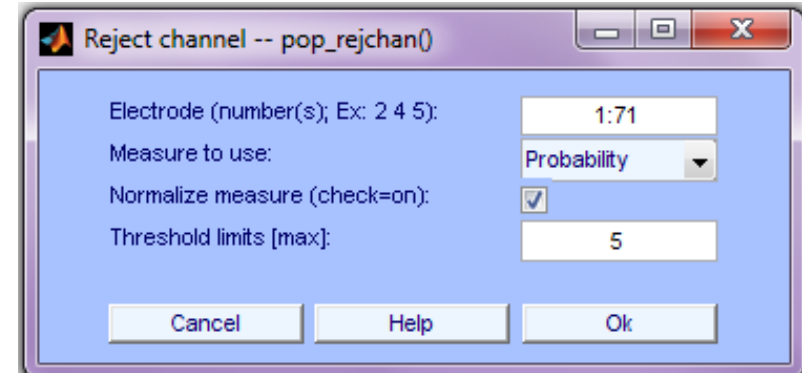
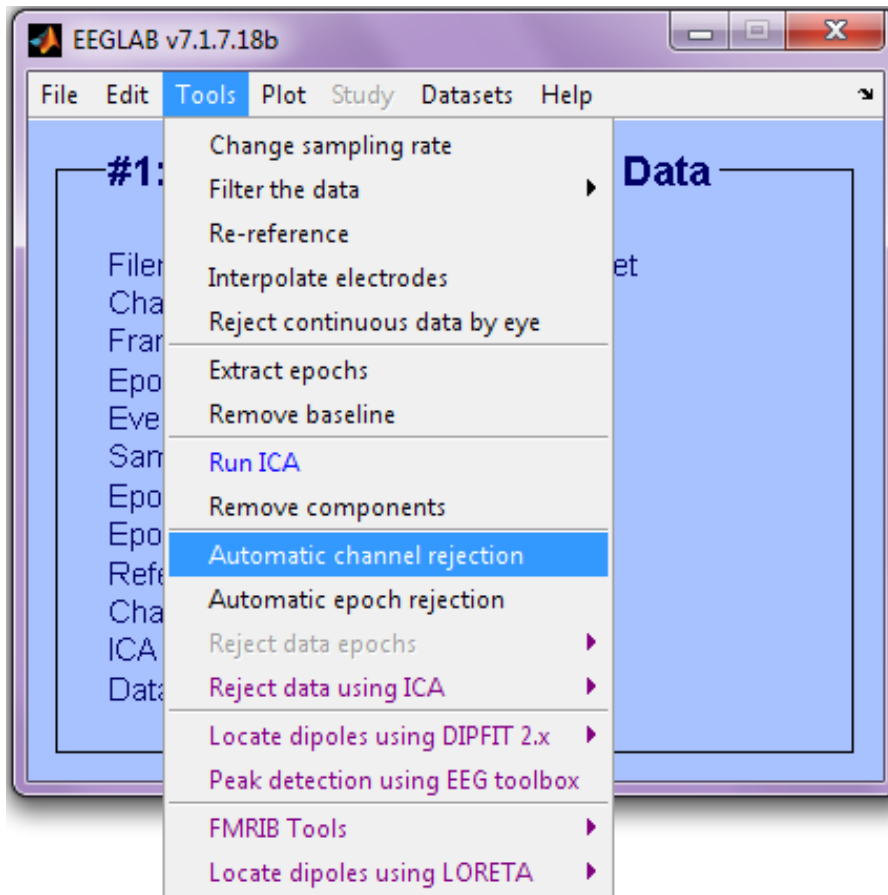
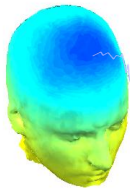
Removing channel(s)



If not checked, will result
in dataset with one channel

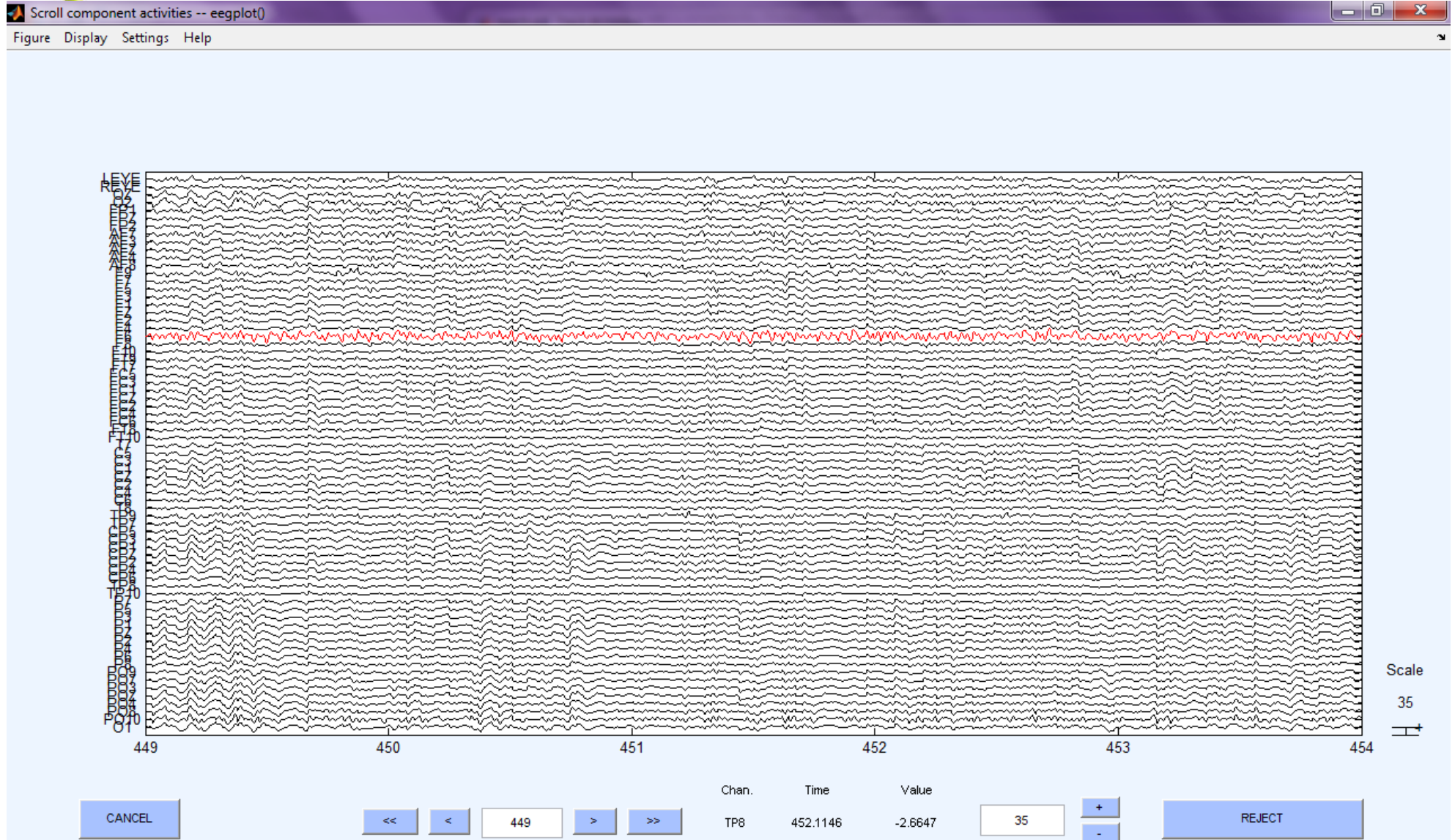
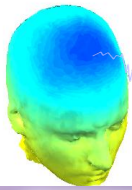


Auto-detection of noisy channels

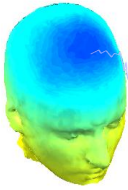


See also `clean_rawdata` plugin of EEGLAB

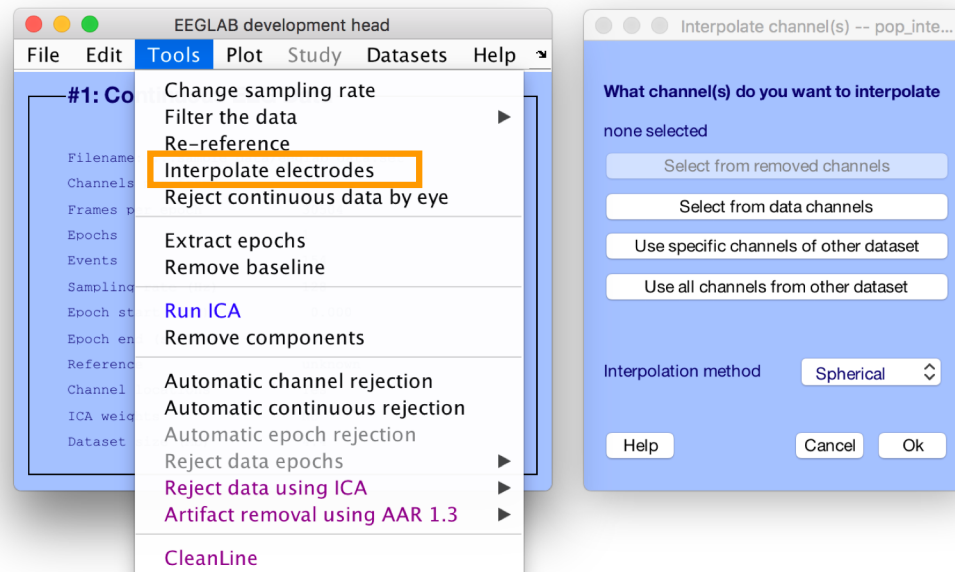
Auto-detected noisy channel



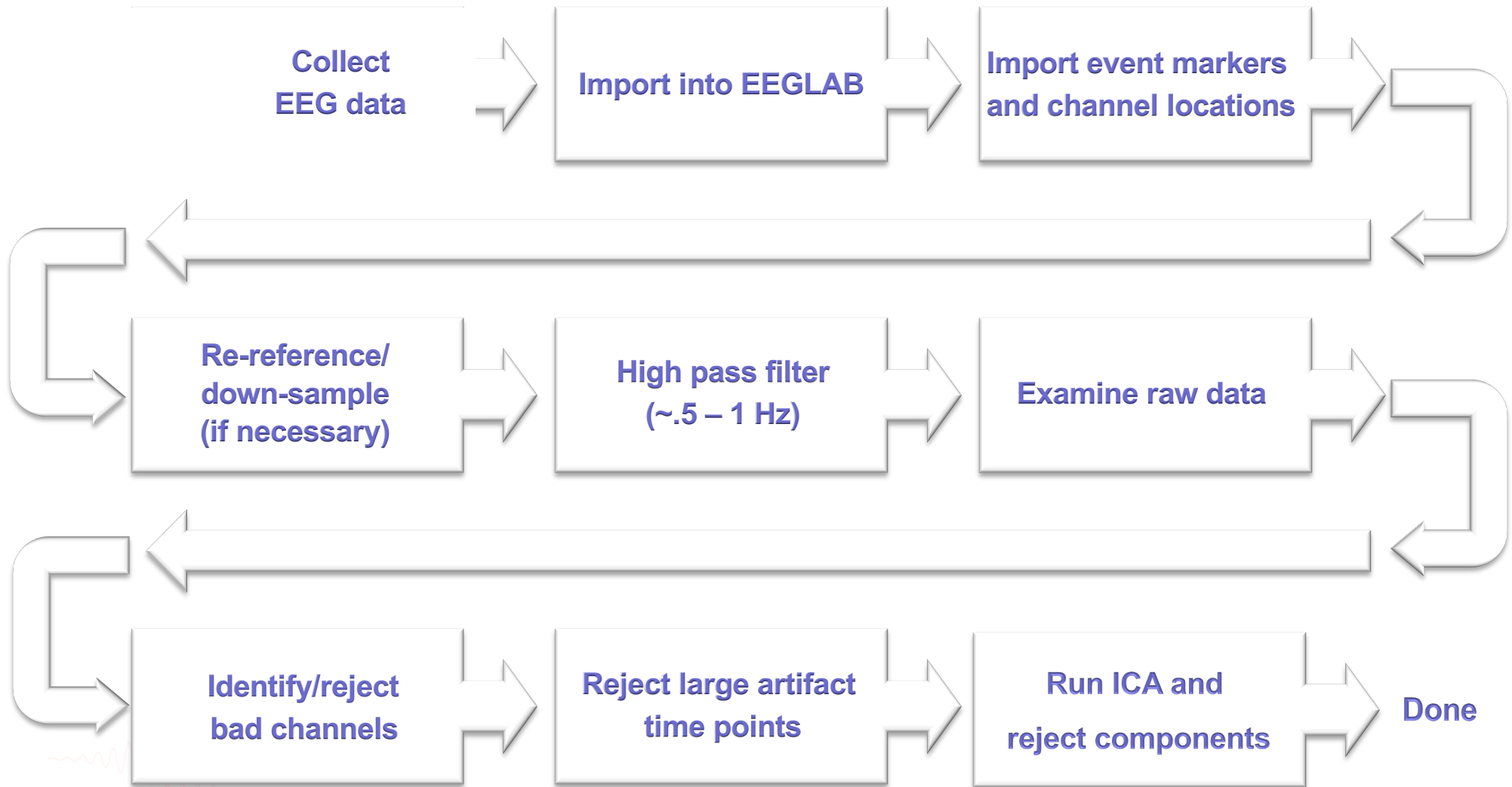
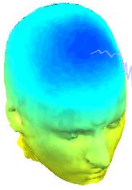
Removed channel(s)



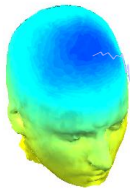
- In EEGLAB, removed channels are not only labeled for rejection, they are actually removed from the data.
- Interpolating channels instead of removing them?



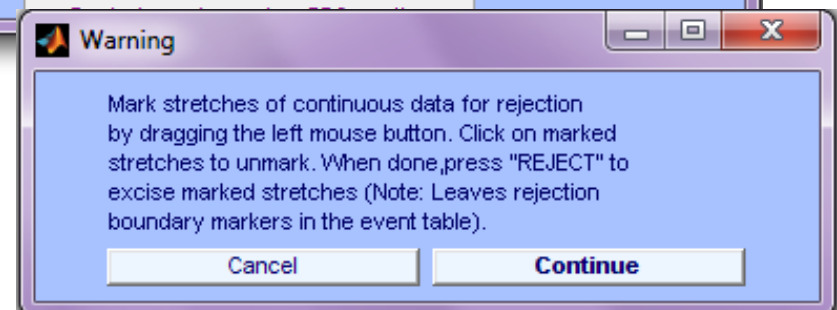
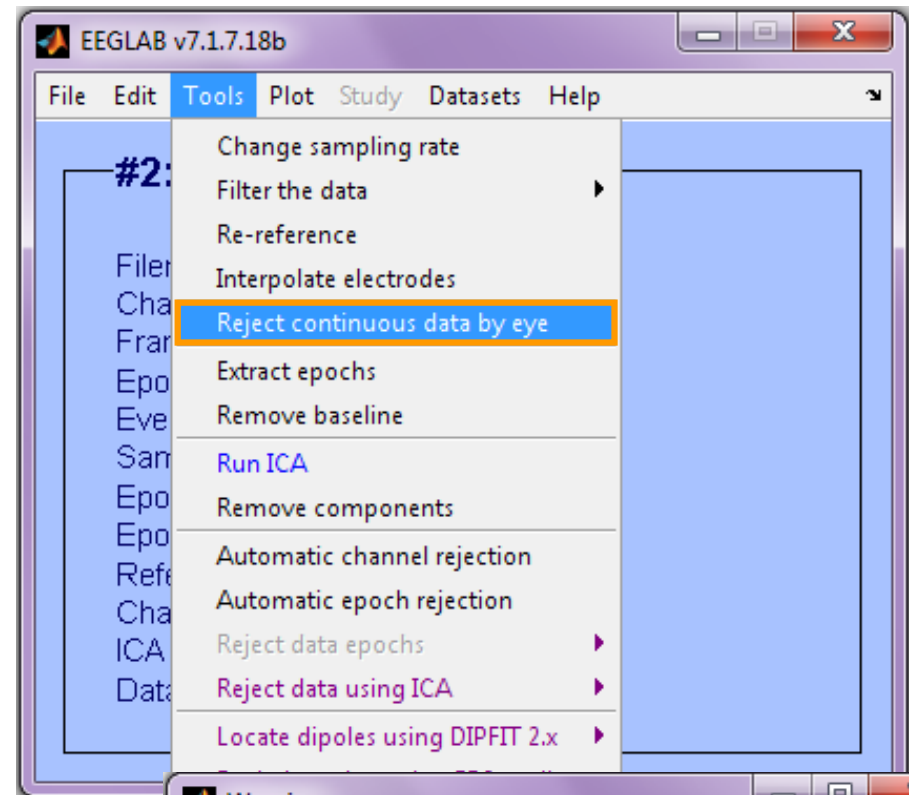
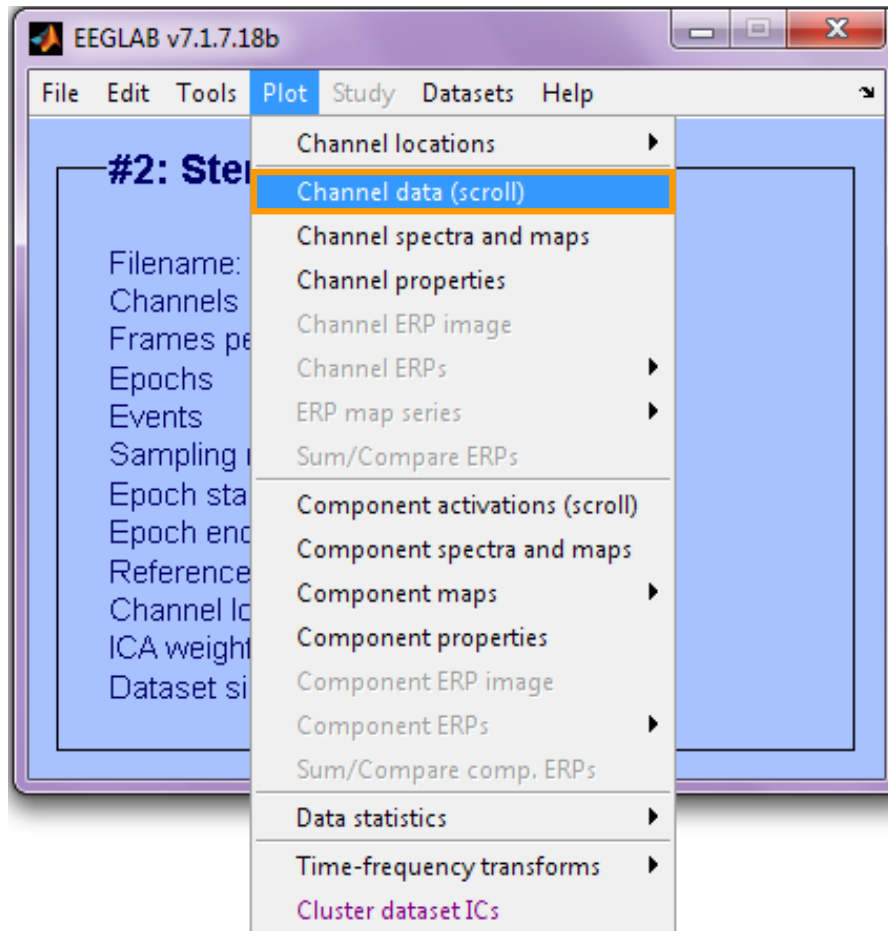
Pre-processing pipeline



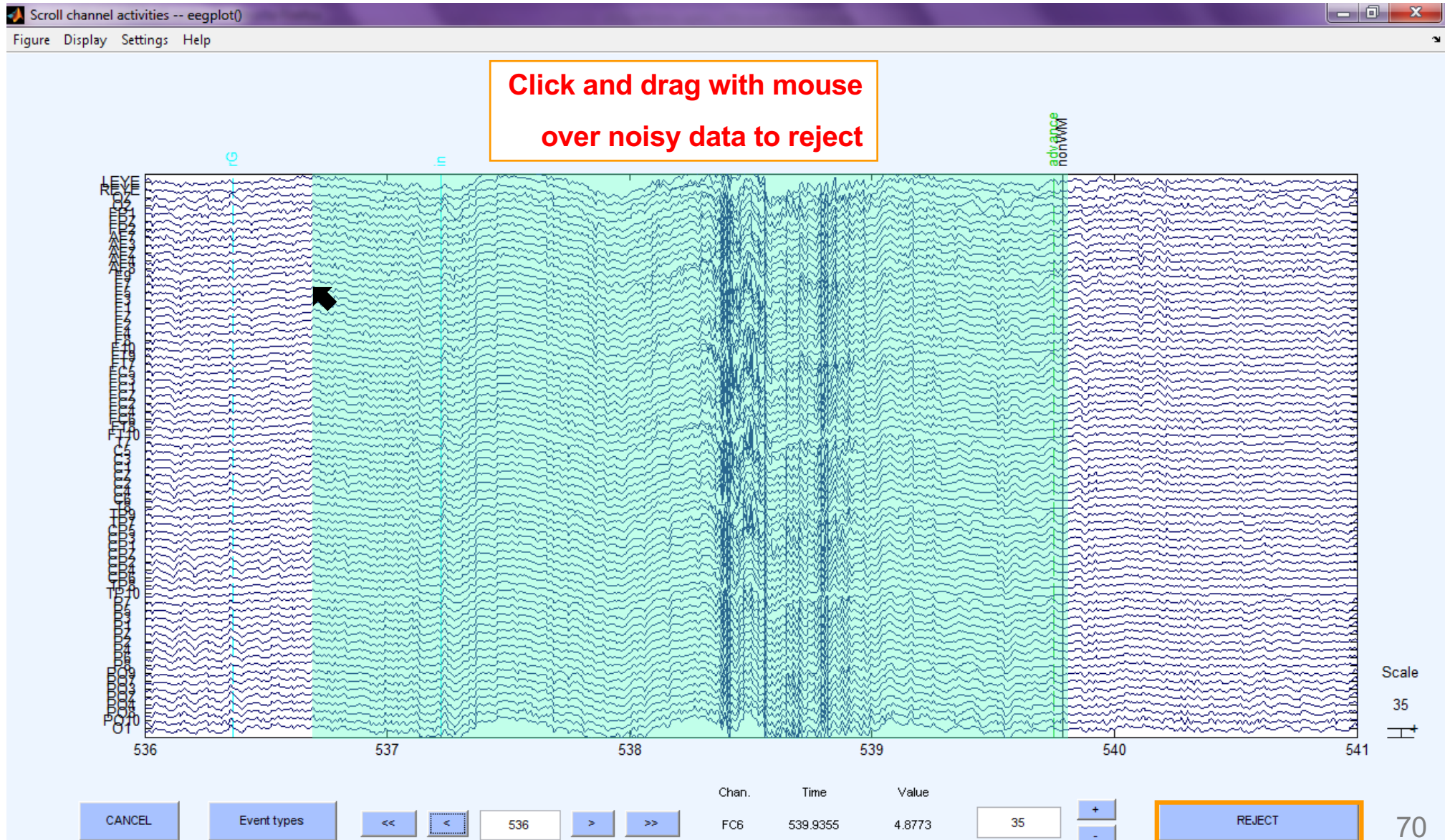
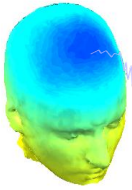
Reject continuous data



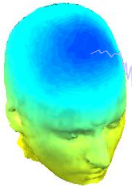
Equivalent



Reject continuous data



Rejecting data for ICA



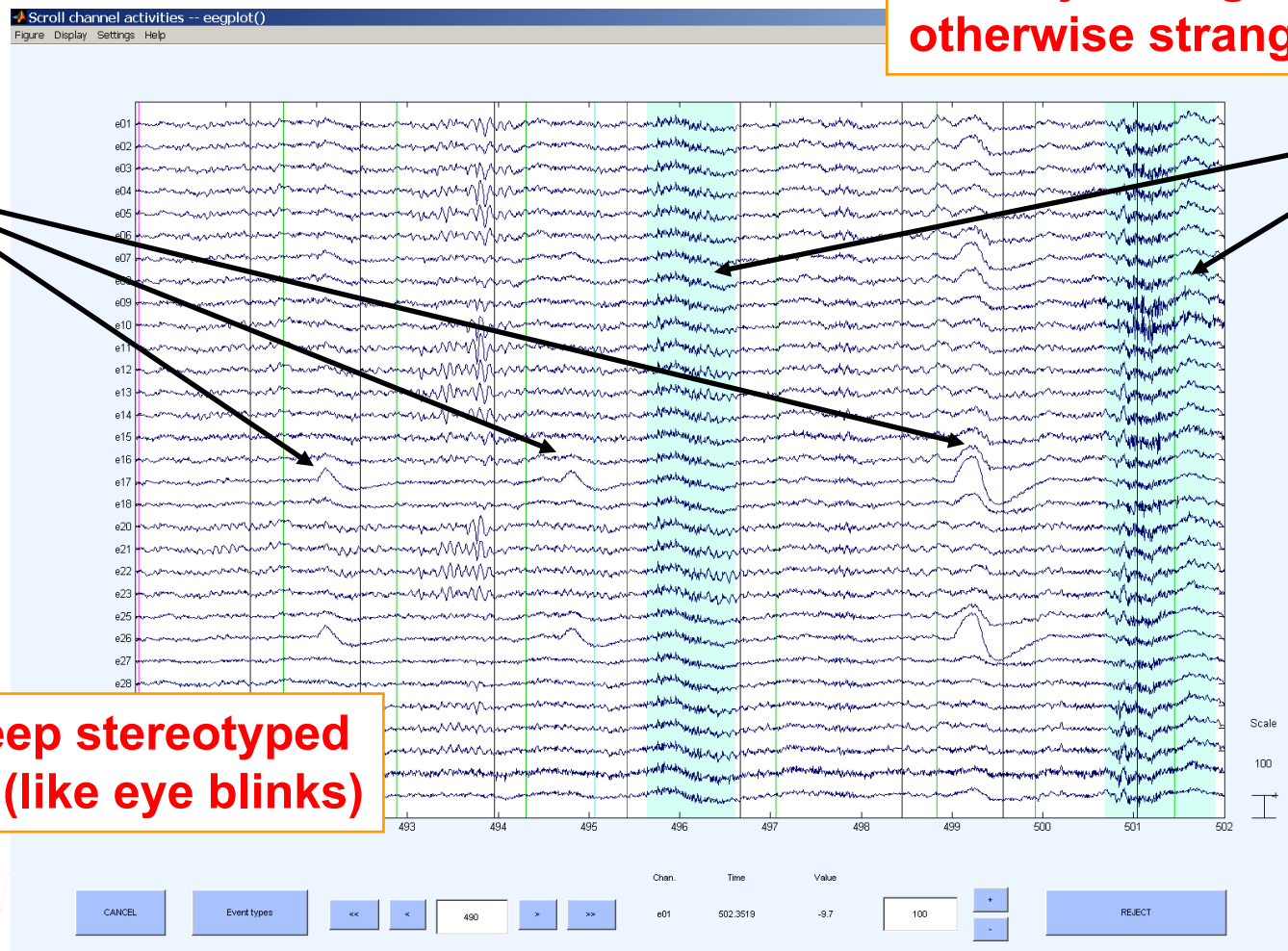
To prepare data for ICA:

Reject large muscle or otherwise strange events...

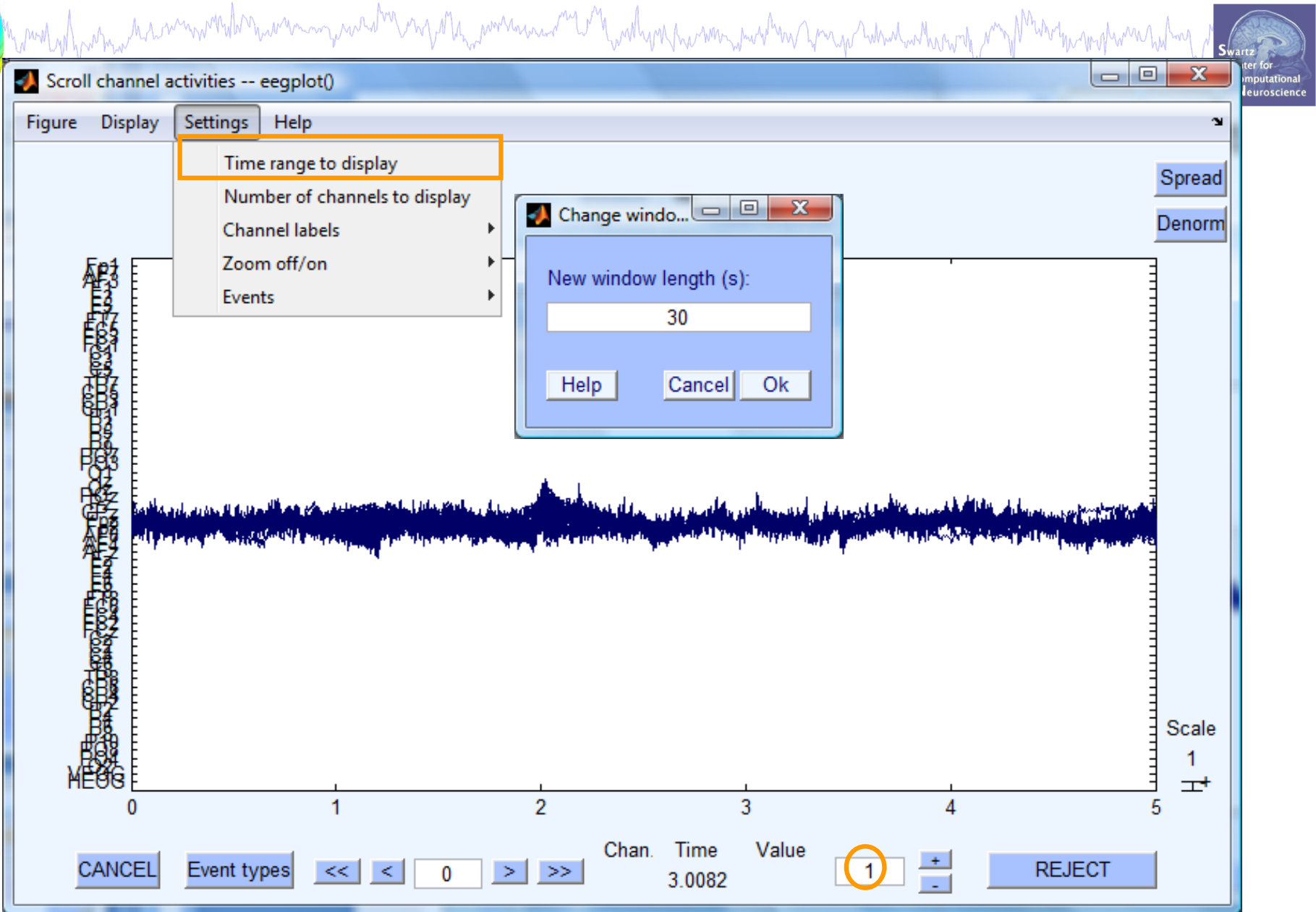
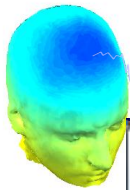
Reject

Keep

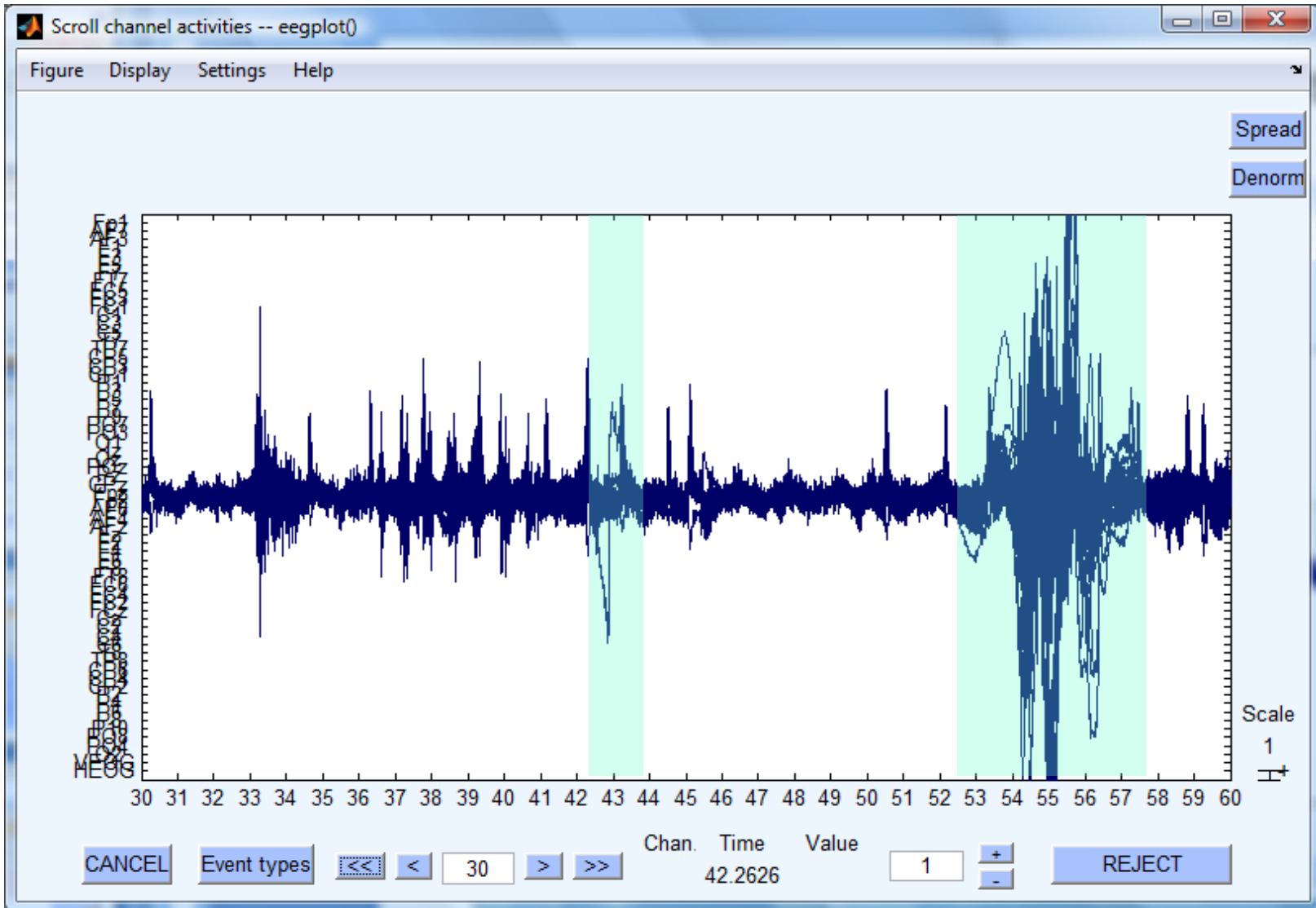
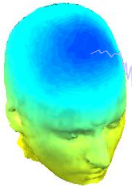
... but keep stereotyped artifacts (like eye blinks)



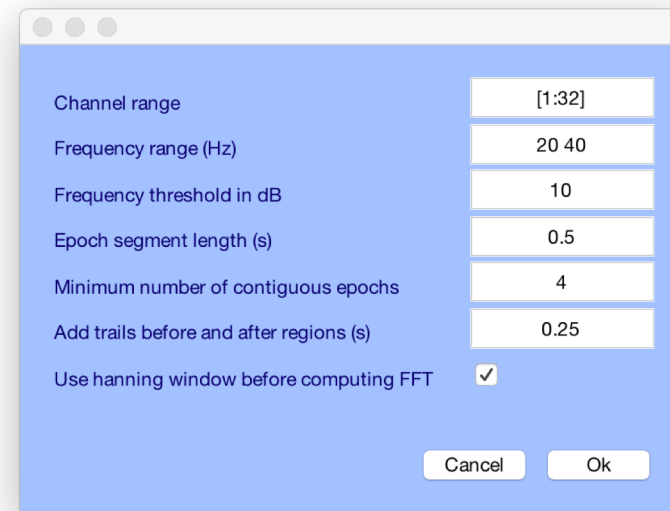
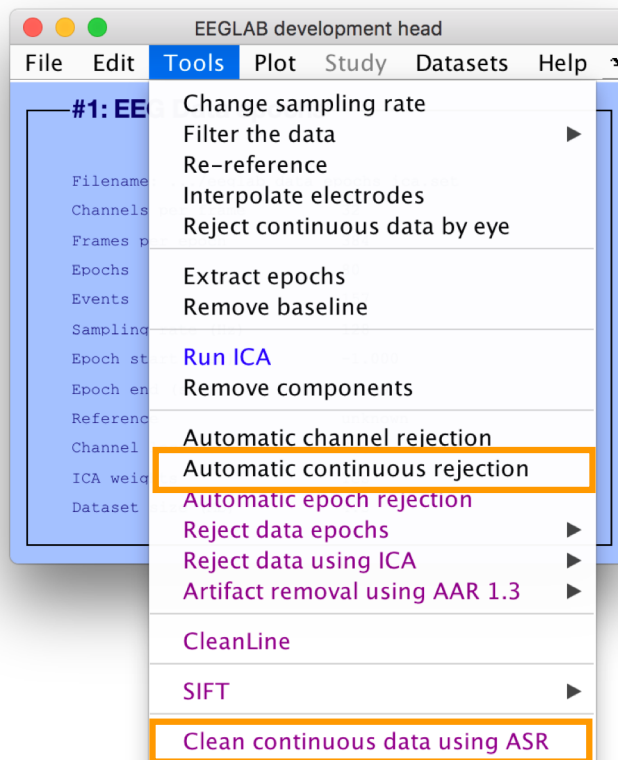
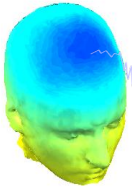
Fast manual artifact rejection

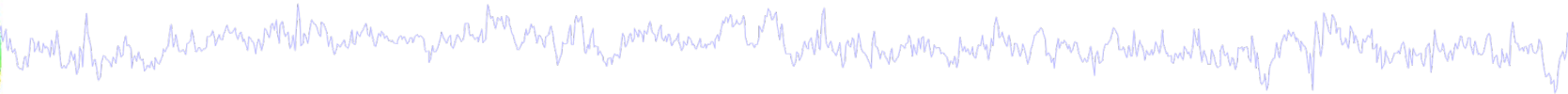
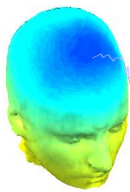


Fast manual artifact rejection



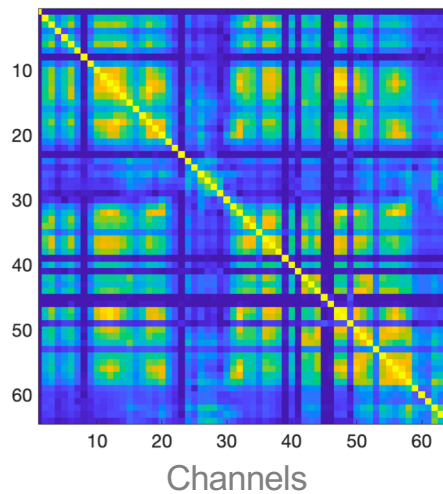
Automatic rejection of continuous data



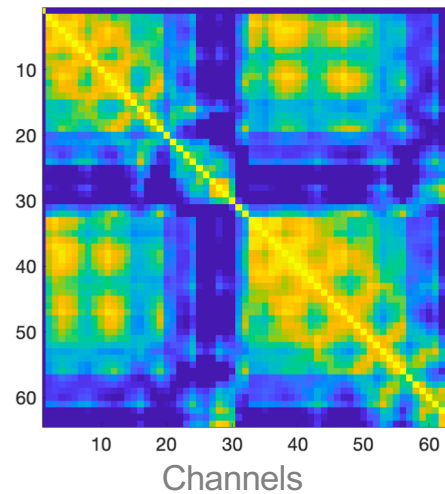


Pairwise correlation to
find bad channels

Bad data



Good data



ASR finds and reconstructs
bad portions of data

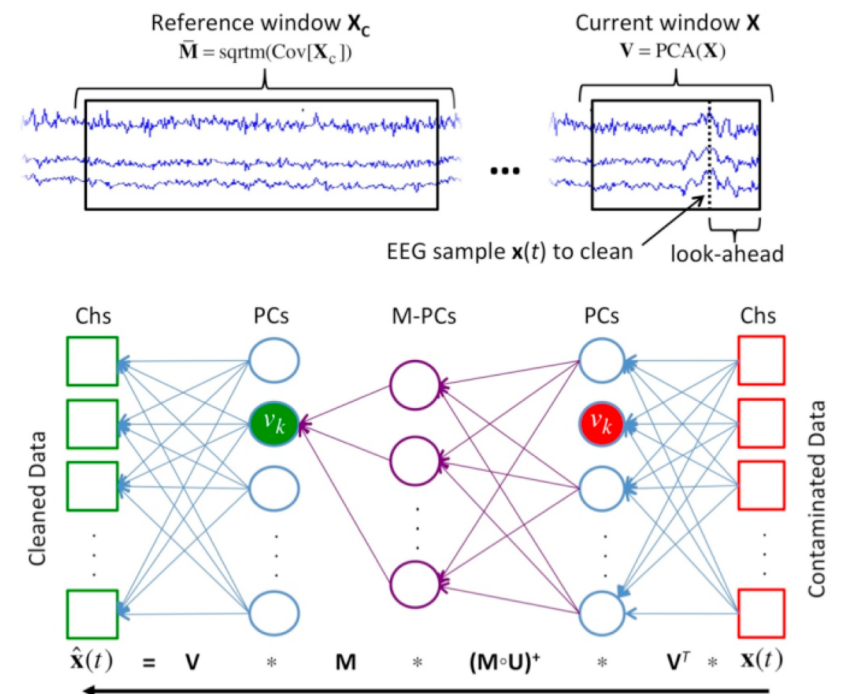
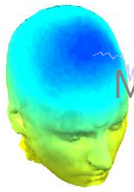
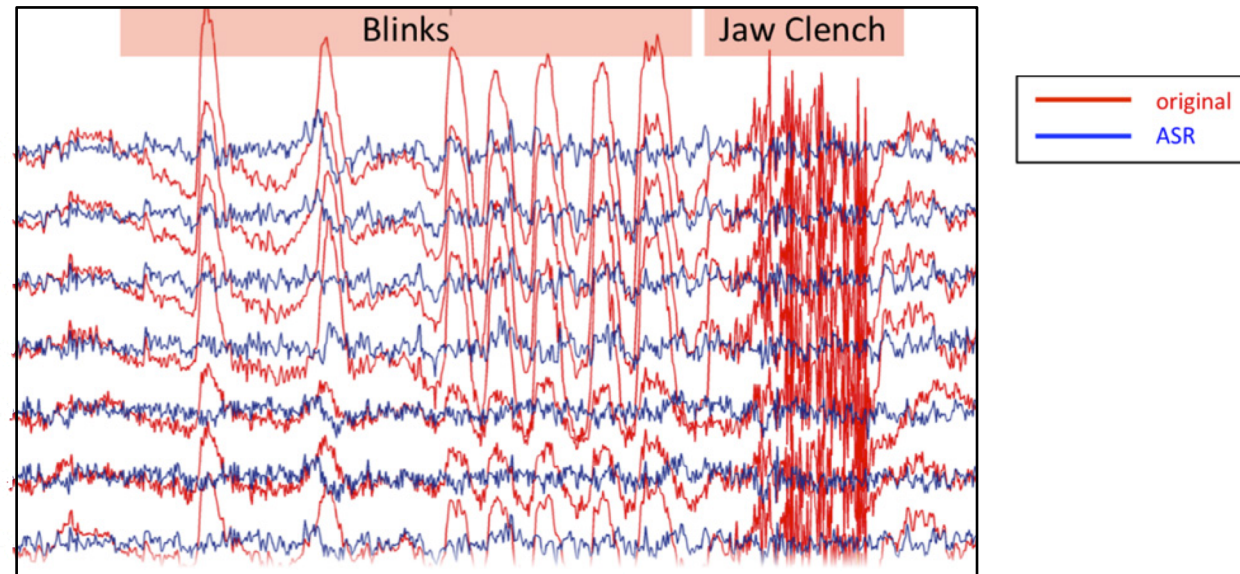


Fig. 3. The Artifact Subspace Reconstruction method. High-variance

Tim R. Mullen, Christian Kothe, et al.(2015) Real-time neuroimaging and cognitive monitoring using wearable dry EEG. Published in IEEE Transactions on Biomedical Engineering.
DOI:10.1109/TBME.2015.2481482

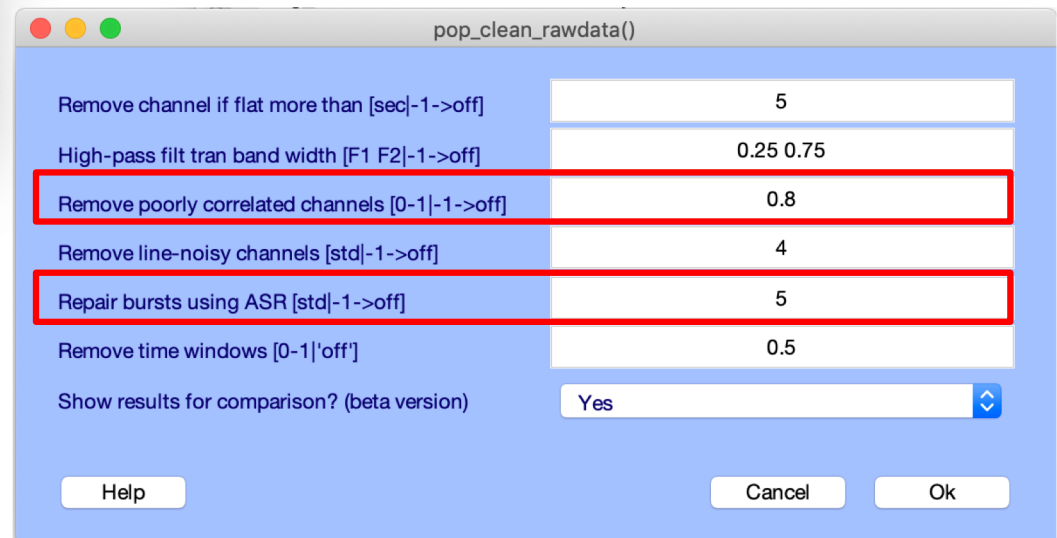
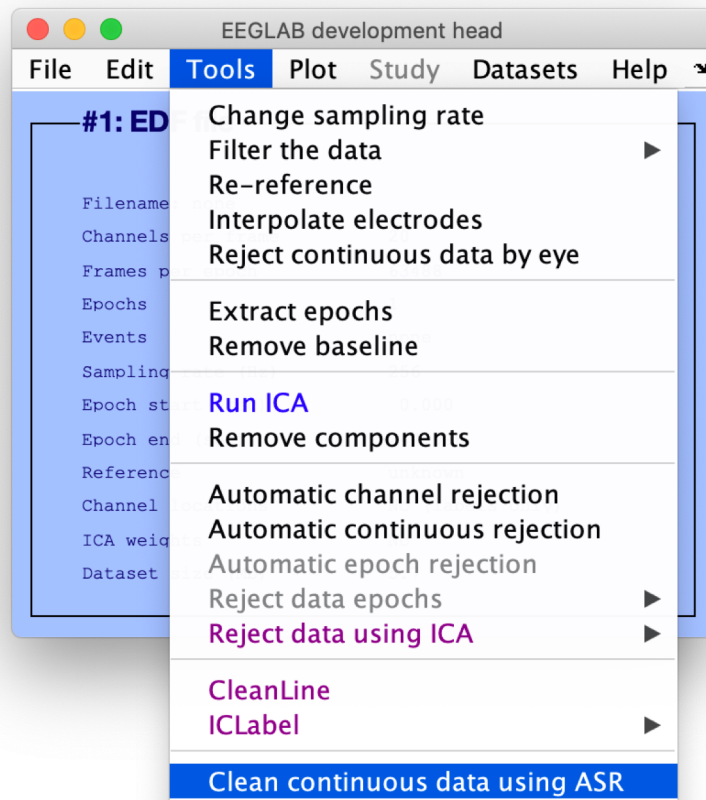
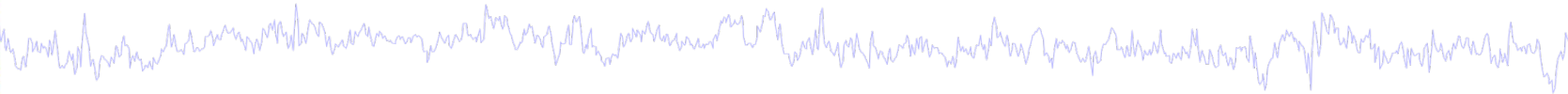
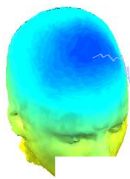


Methods

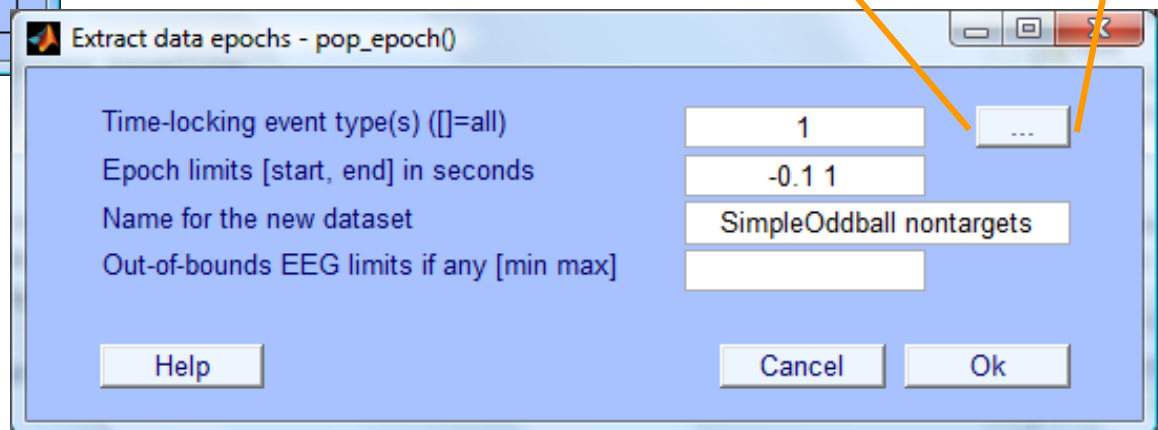
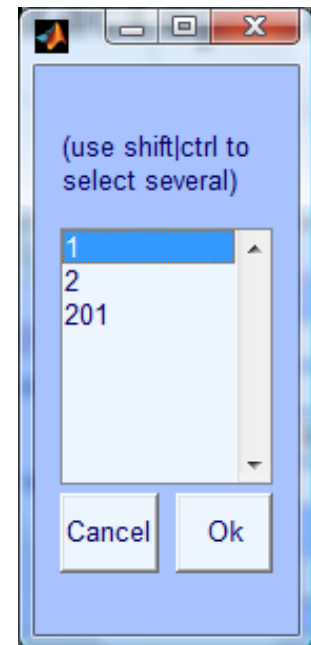
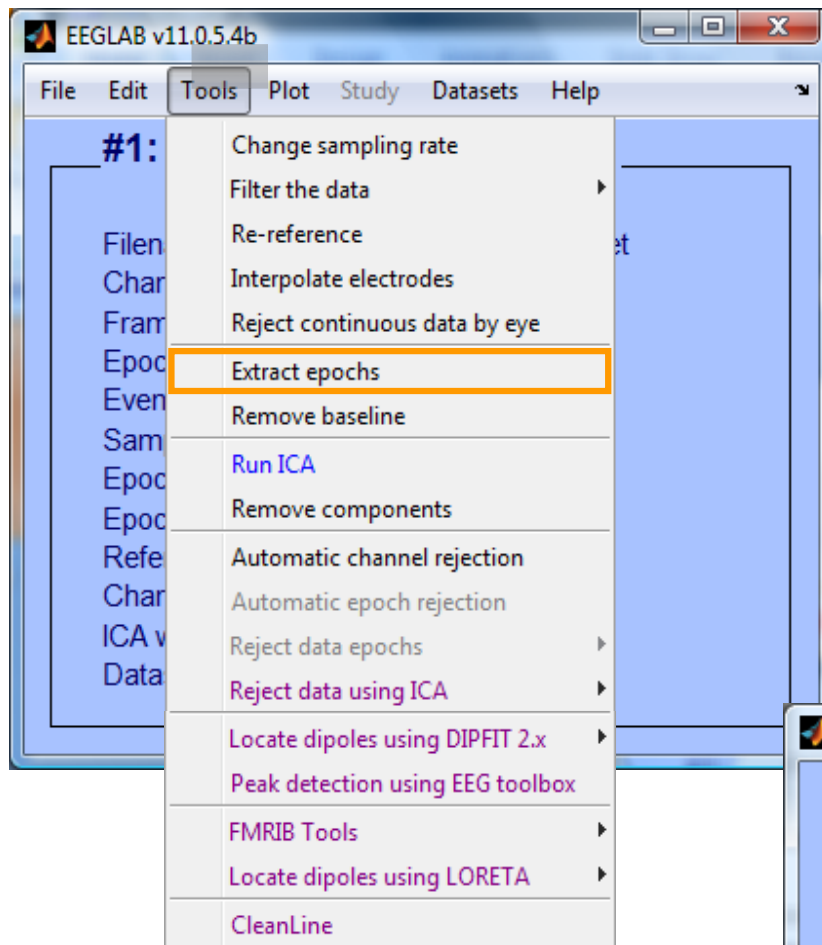
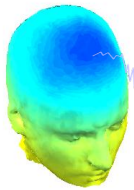


Validation: Chang CY, Hsu SH, Pion-Tonachini L, Jung TP. Evaluation of Artifact Subspace Reconstruction for Automatic EEG Artifact Removal. Conf Proc IEEE Eng Med Biol Soc. 2018 Jul;2018:1242-1245. doi: 10.1109/EMBC.2018.8512547.

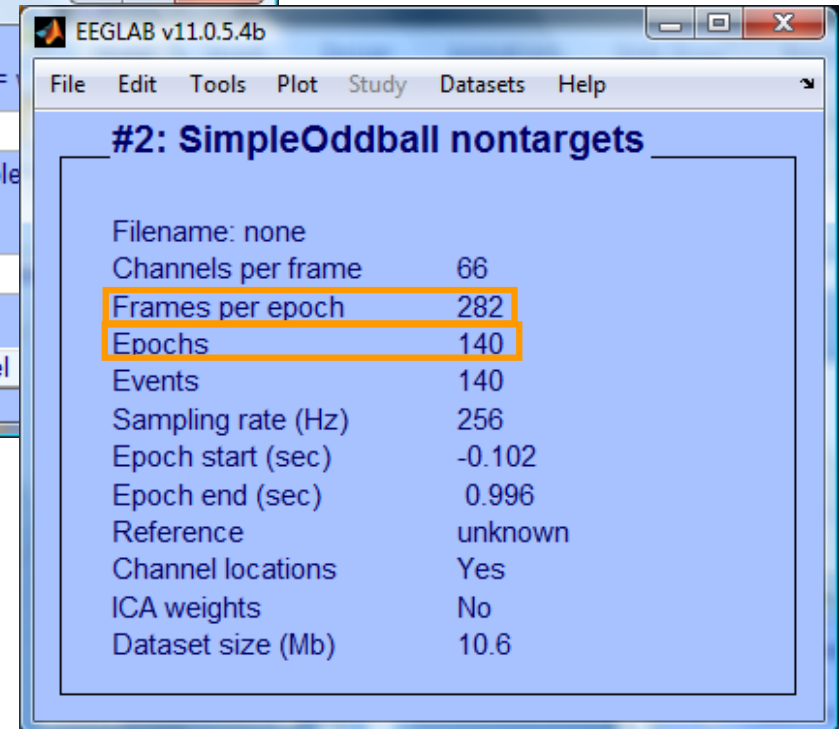
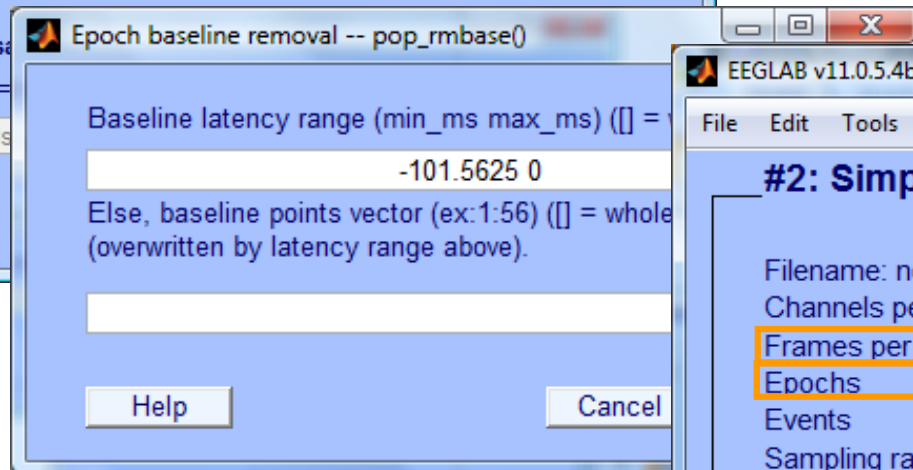
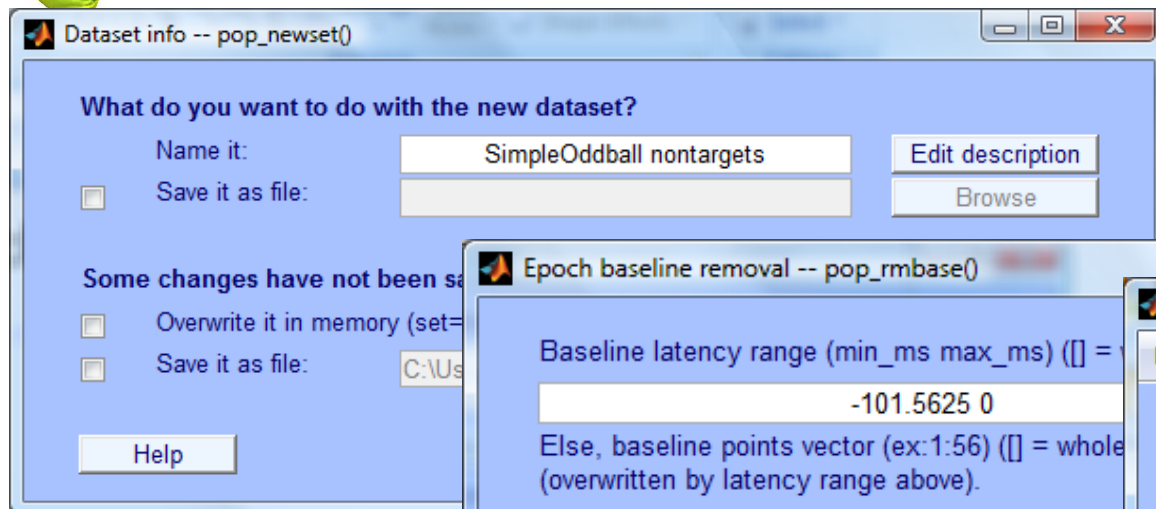
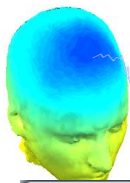
Variation: Sarah Blum*, Nadine S. J. Jacobsen, Martin G. Bleichner and Stefan Debener (2019) A Riemannian Modification of Artifact Subspace Reconstruction for EEG Artifact Handling. Front. Hum. Neurosci., <https://doi.org/10.3389/fnhum.2019.00141>.



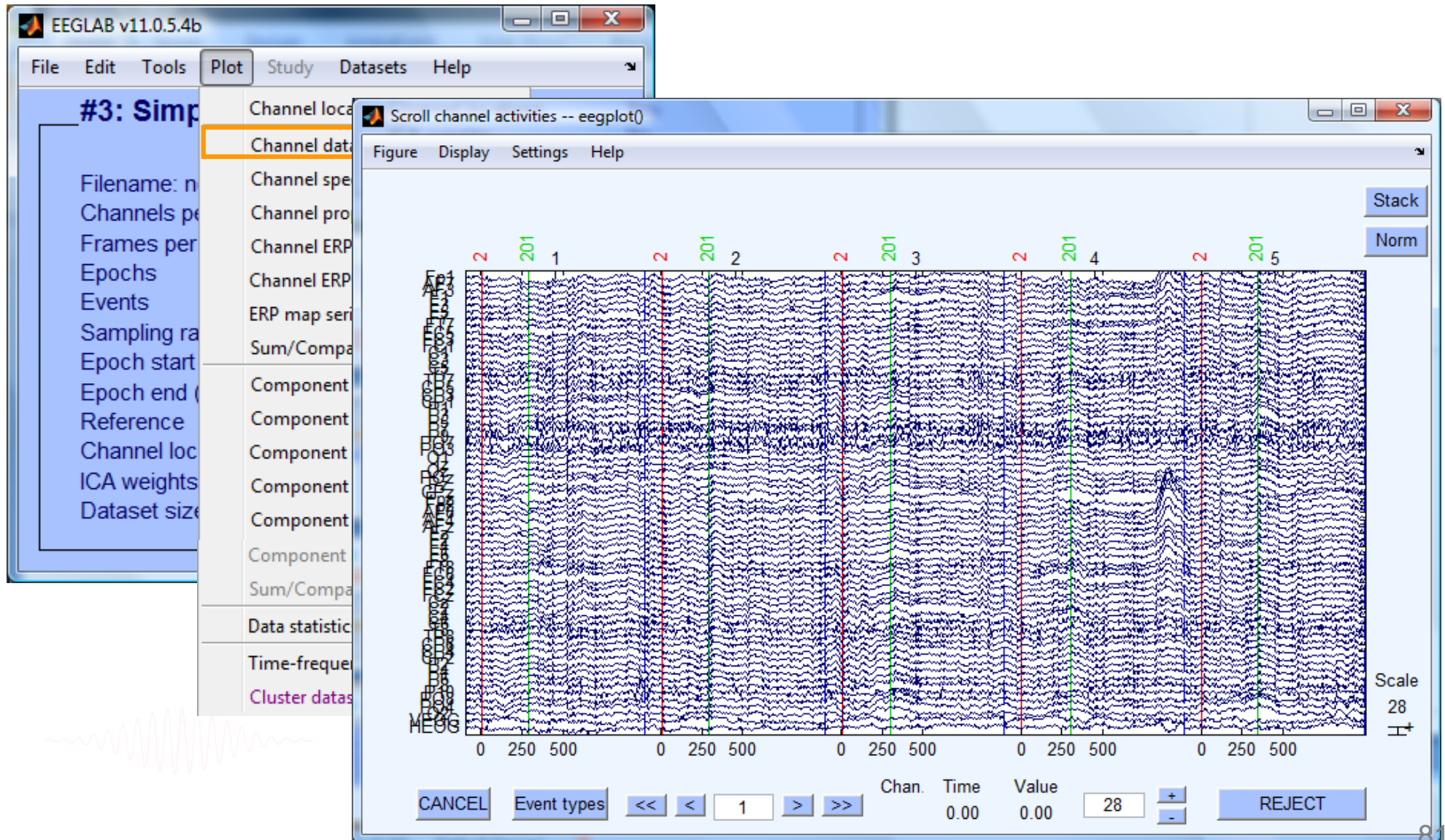
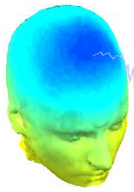
Extract epochs



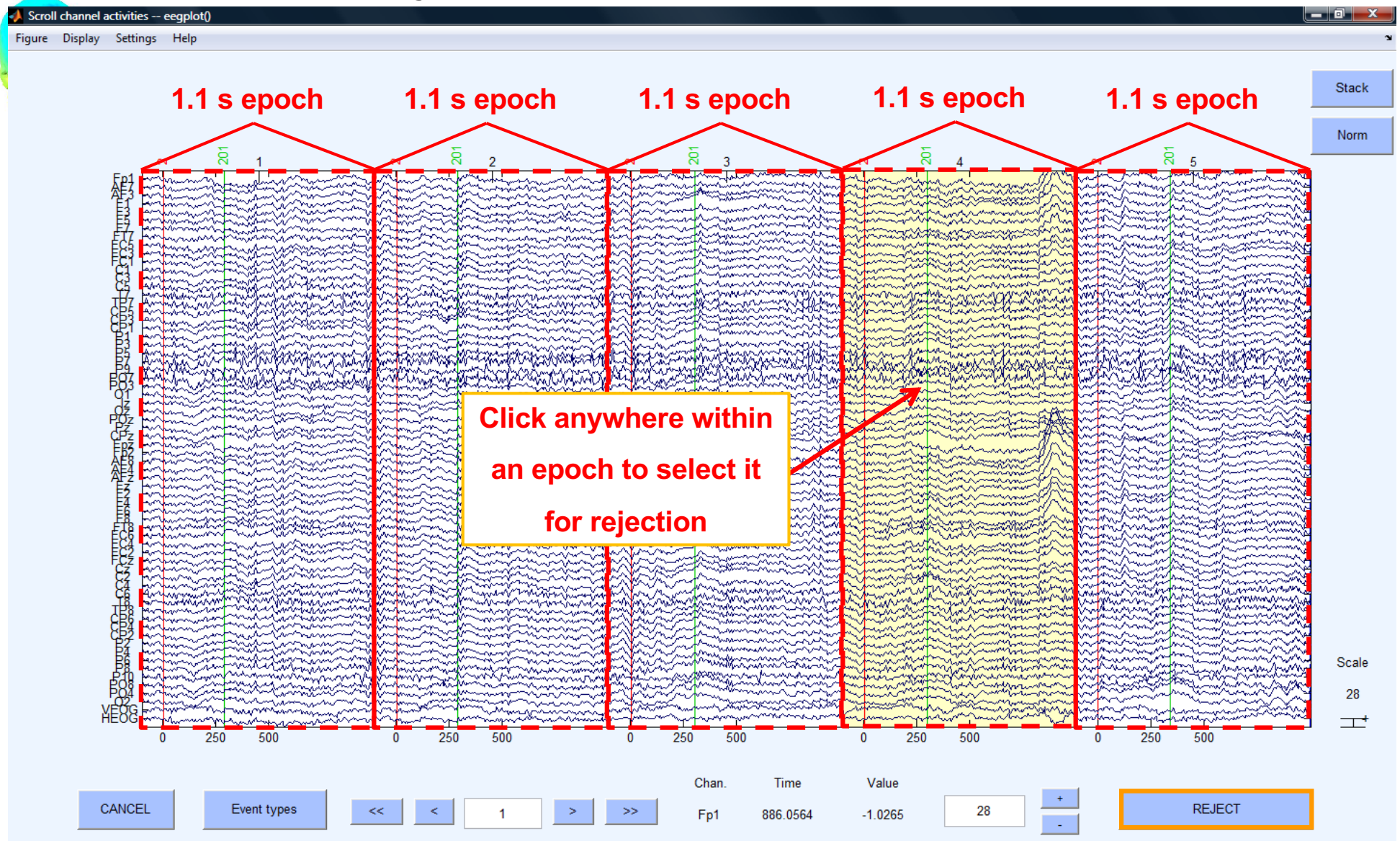
Extract epochs



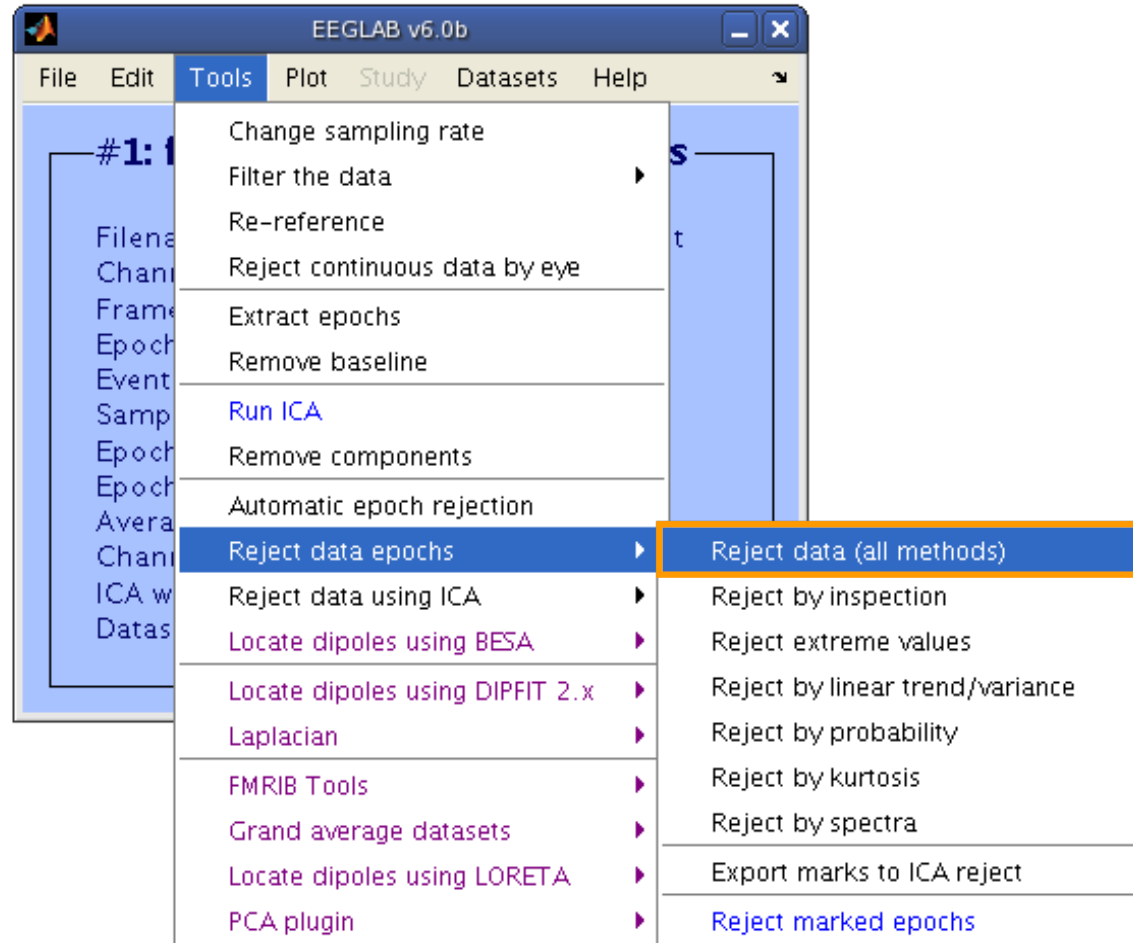
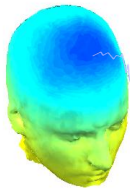
Scroll (epoched) channel data



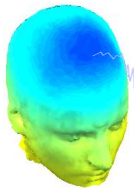
Reject epochs with artifact



Reject data epochs



Reject data epochs



visual
inspection

Reject trials using data statistics - pop_rejmenu()

Mark trials by appearance ☐ Scroll Data Marked trials 0

Find abnormal values

Upper limit(s) (uV) Lower limit(s) (uV)
 Start time(s) (ms) Ending time(s) (ms)
 Electrode(s) Currently marked trials

Find abnormal trends

Max slope (uV/epoch) R-squared limit (0 to 1)
 Electrode(s) Currently marked trials

Find improbable data

Single-channel limit (std. dev.) All channels limit (std. dev.)
 Electrode(s) Currently marked trials

Find abnormal distributions

Single-channel limit (std. dev.) All channels limit (std. dev.)
 Electrode(s) Currently marked trials

Find abnormal spectra (slow)

Upper limit(s) (dB) Lower limit(s) (dB)
 Low frequency(s) (Hz) High frequency(s) (Hz)
 Electrode(s) Currently marked trials

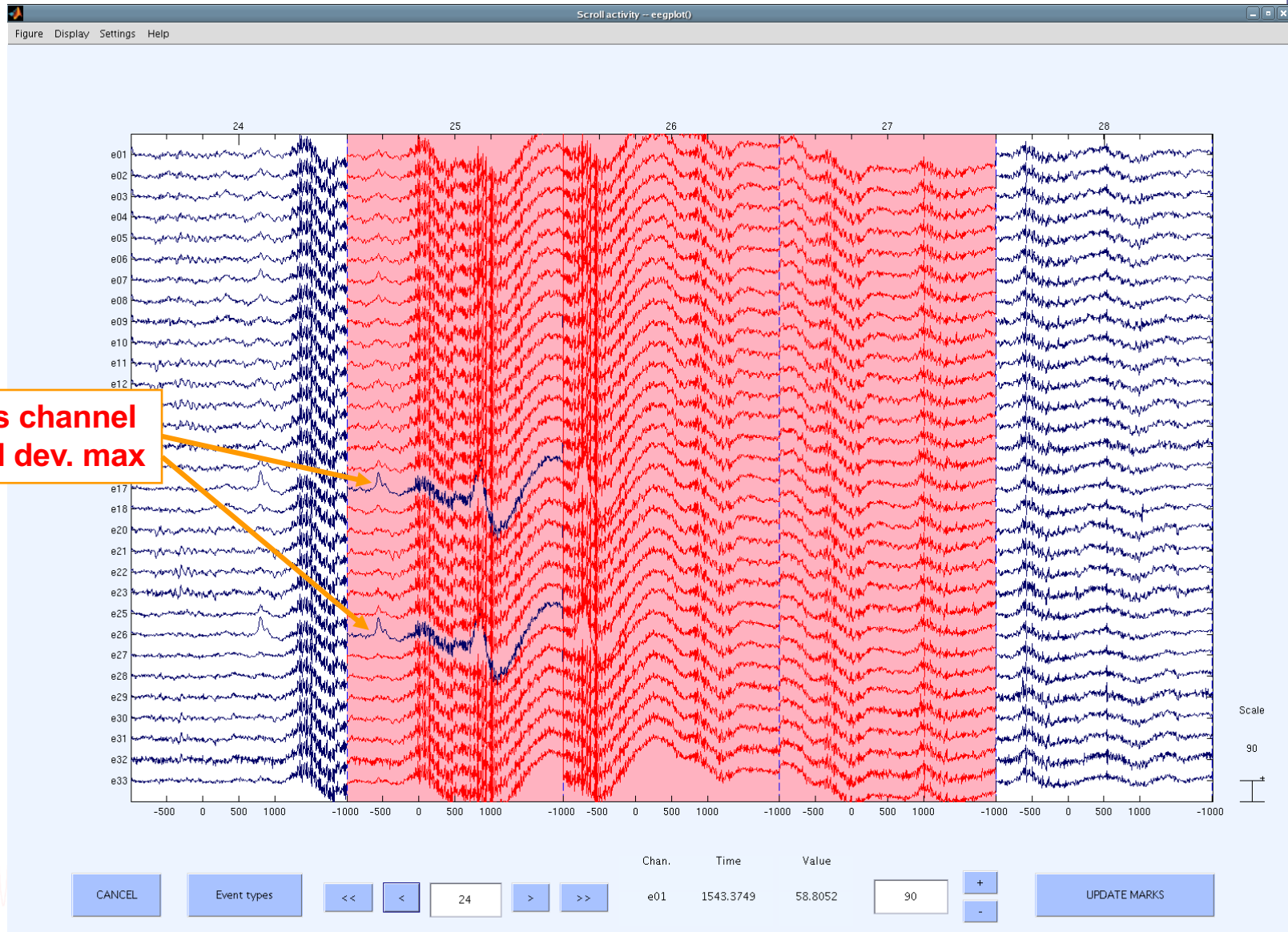
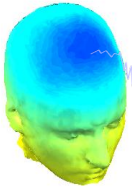
Plotting options

Show all trials marked for rejection by the measure selected above or checked below

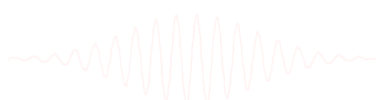
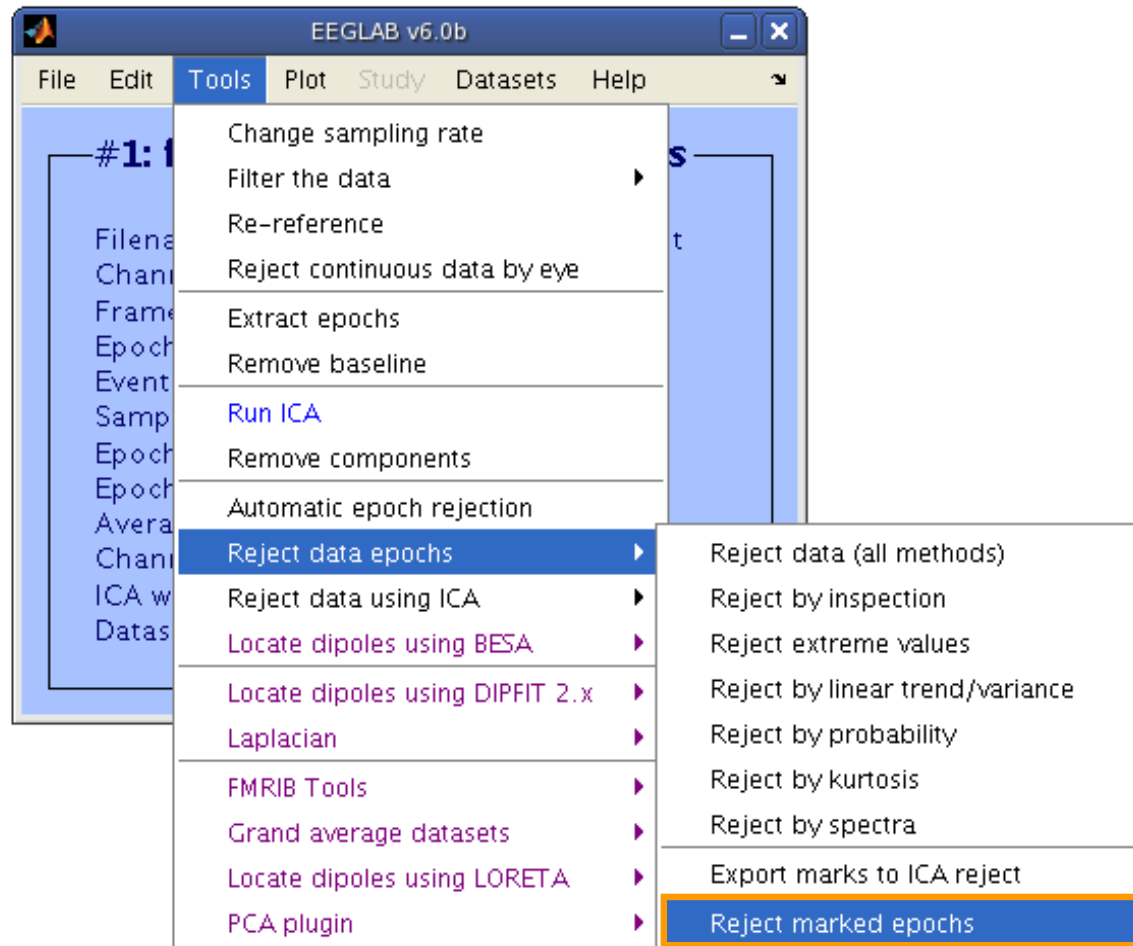
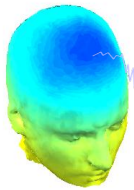
☒ Abnormal appearance ☒ Abnormal values ☒ Abnormal trends
☒ Improbable epochs ☒ Abnormal distributions ☒ Abnormal spectra

probability

Reject data epochs



Reject data epochs



Pre-processing pipeline

