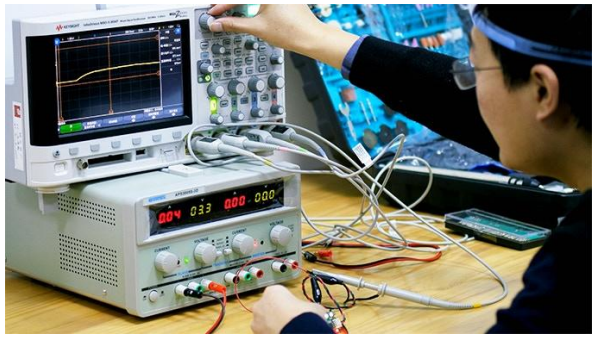
Importing, cleaning, preprocessing data

Romain Grandchamp, PhD
Johanna Wagner, PhD
Ramon Martinez-Cancino, PhD
Arnaud Delorme, PhD

Practical MEEG 2025, Aix-Marseille Université, France, October, 2025

Why preprocess data?

Measuring EEG data out of the recording device is like measuring a difference of potential on an oscilloscope.

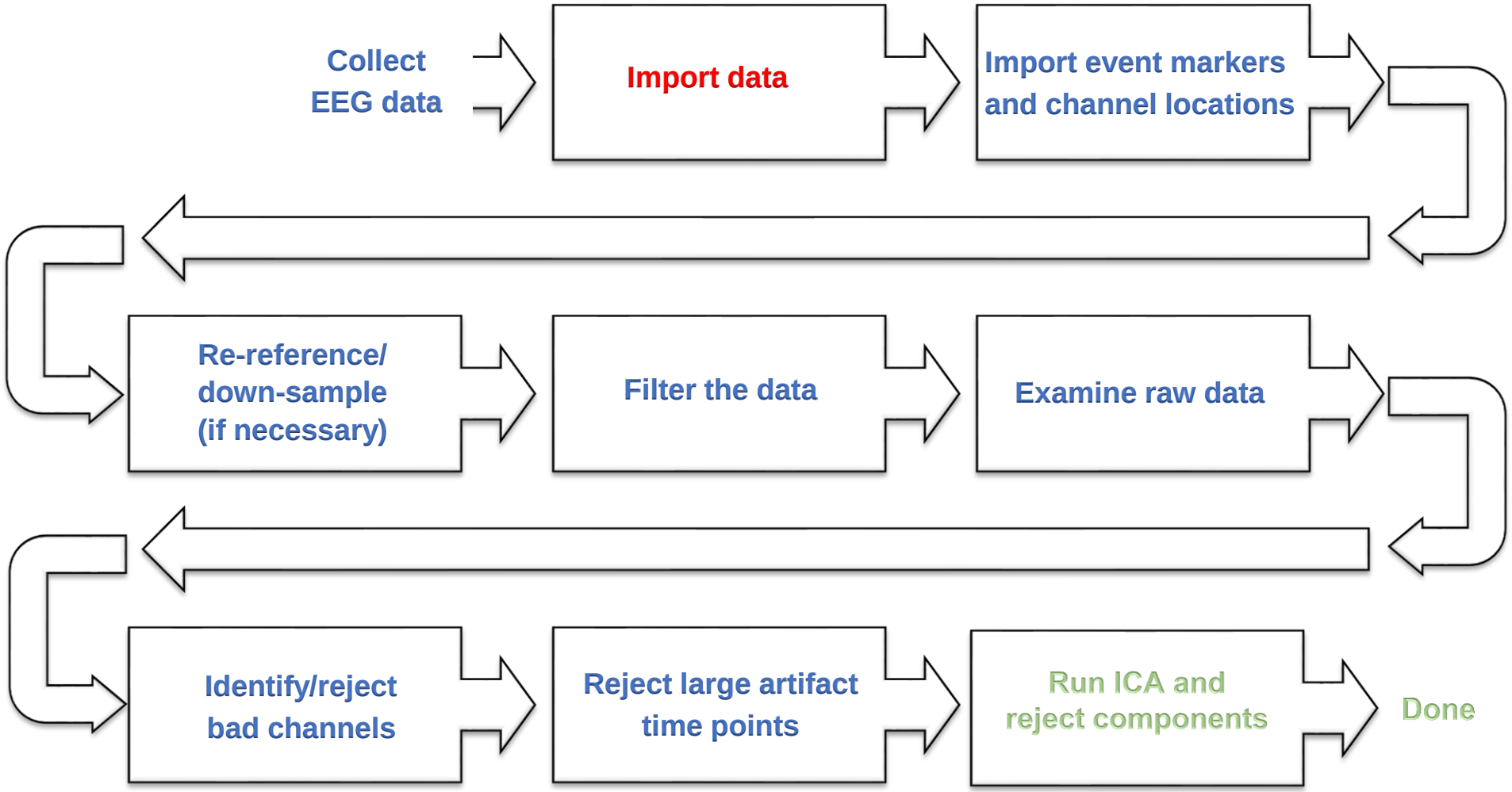


To make sense of the data, we need to:

- 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 transformations to the data

Pre-processing pipeline



What is BIDS?

BIDS is a way to organize your raw data

- To improve consistent and complete documentation
- To facilitate re-use by your future self and others

BIDS is NOT

- A new file format
- A search engine
- A data sharing tool

BIDS in popular open-source tools

EEGLAB

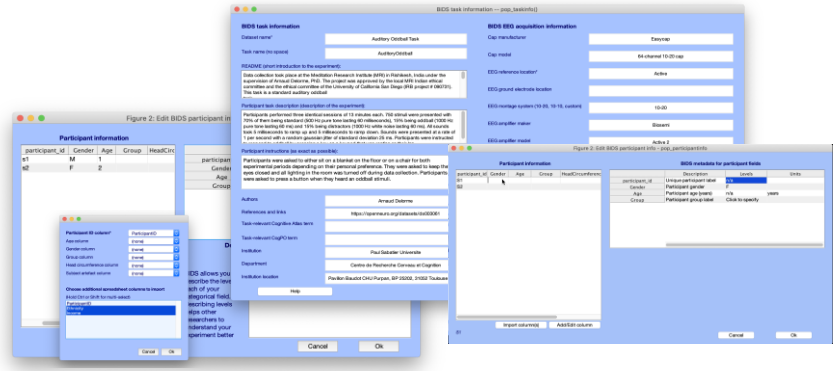
Brainstorm

Have dedicated tools for processing group-level data

FieldTrip

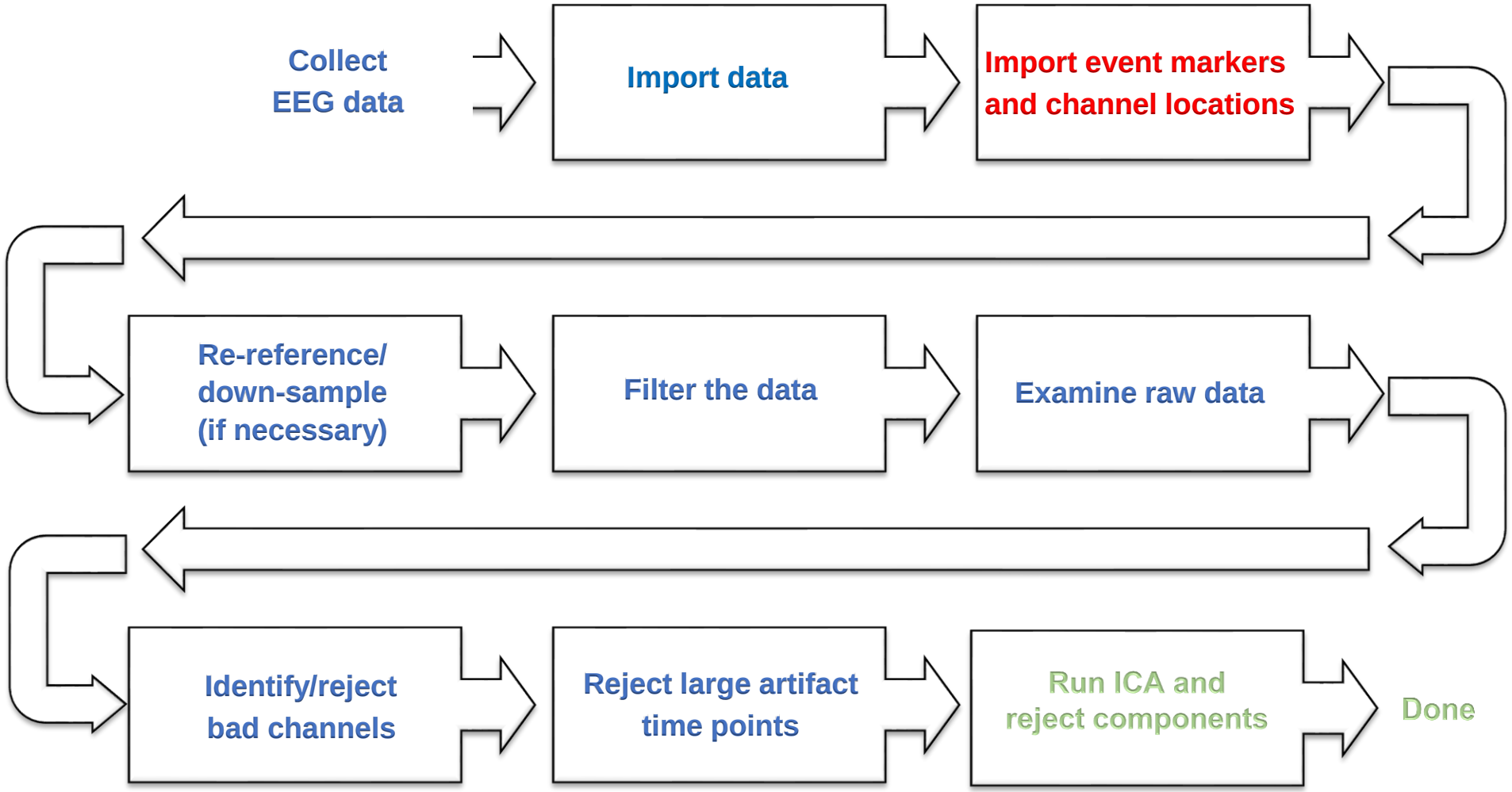
MNE
MEG + EEG ANALYSIS & VISUALIZATION

EEGLAB *bids-matlab-tools* menus and script functions (1450 downloads)

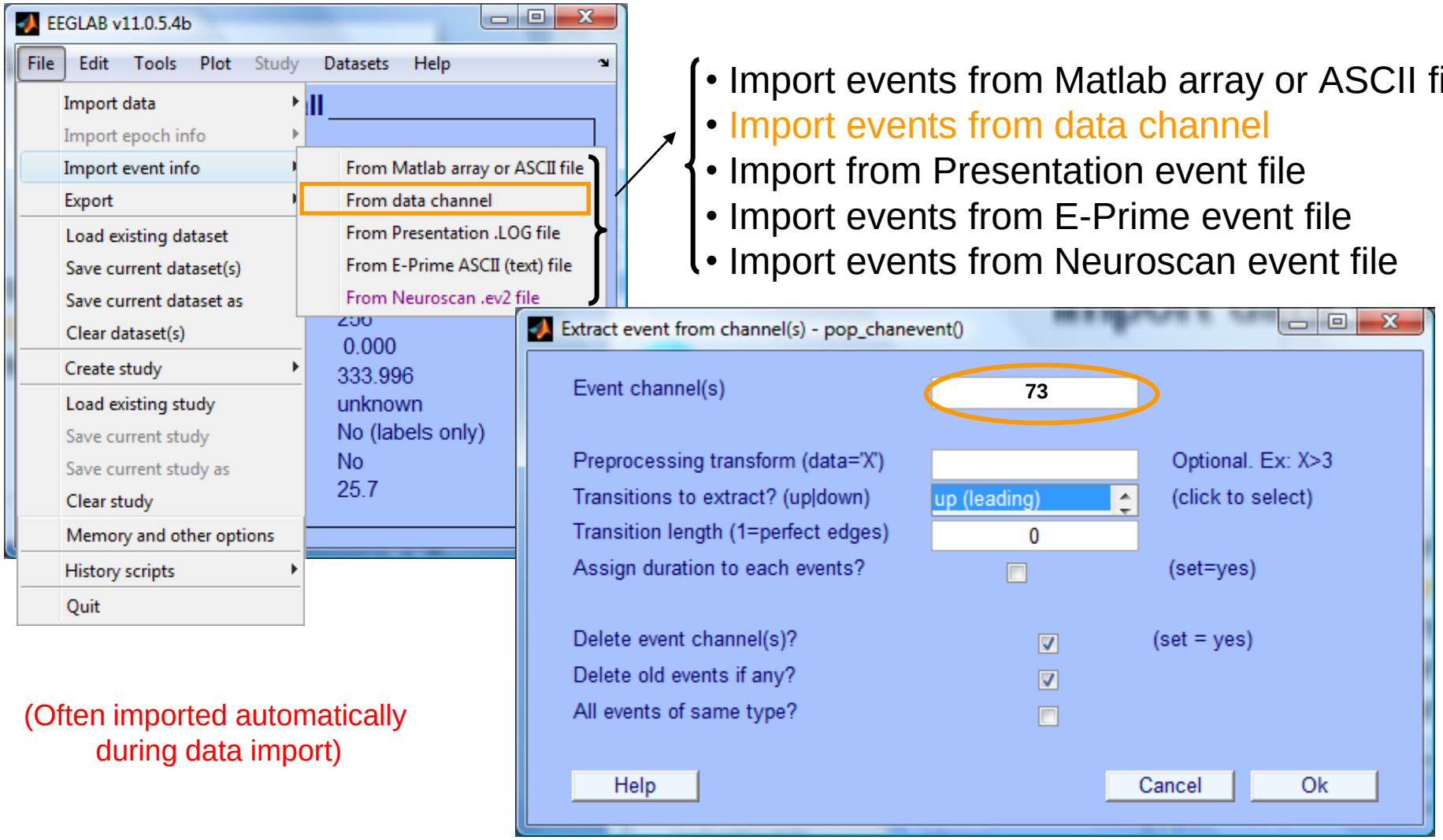


- All tools can **import** BIDS single subject data (FieldTrip and EEGLAB support BIDS sidecar event files)
- All tools can **export** BIDS data (GUI available in Brainstorm and EEGLAB)
- Brainstorm and EEGLAB have dedicated BIDS tools for group analysis

Pre-processing pipeline



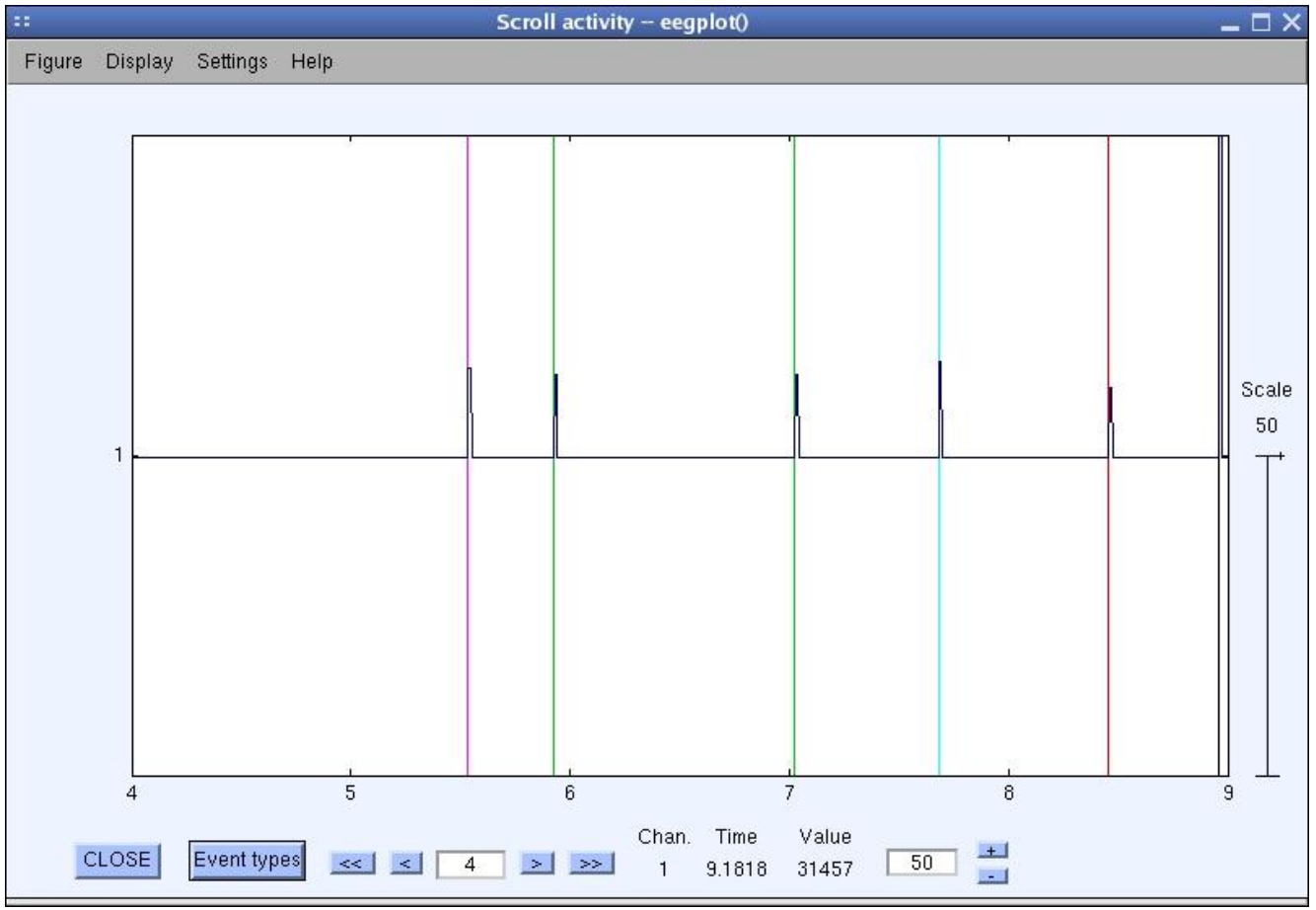
Import data events



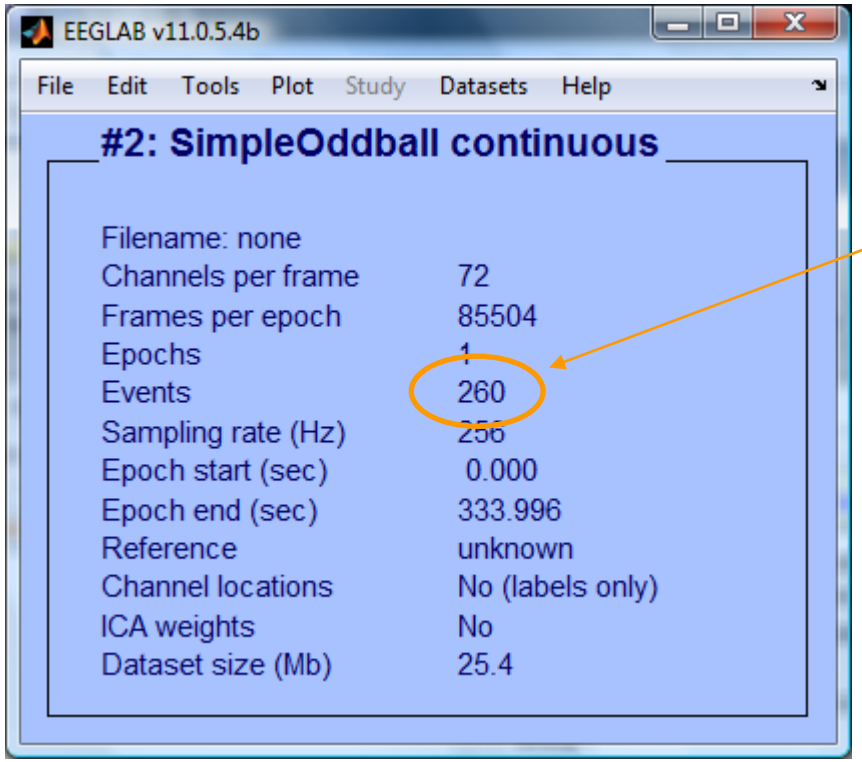
- Import events from Matlab array or ASCII file
- Import events from data channel
- Import from Presentation event file
- Import events from E-Prime event file
- Import events from Neuroscan event file

(Often imported automatically during data import)

Appearance of an event channel in raw data

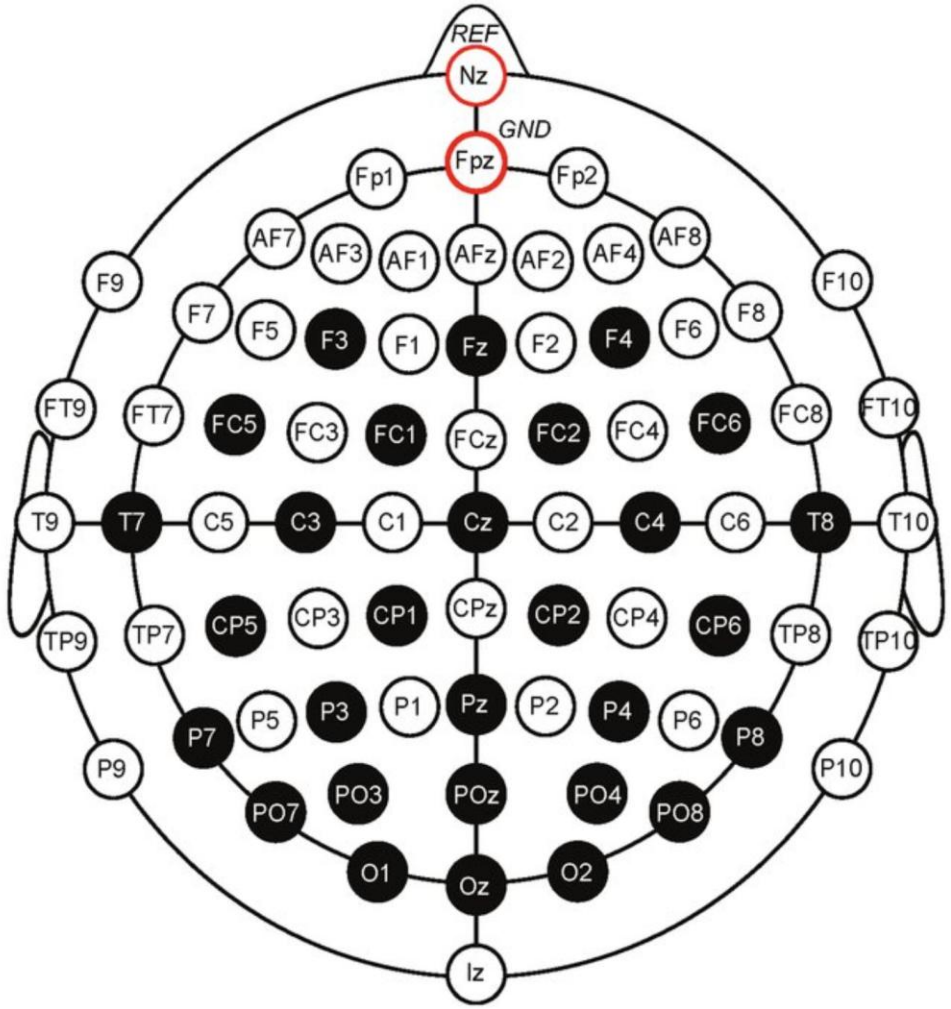


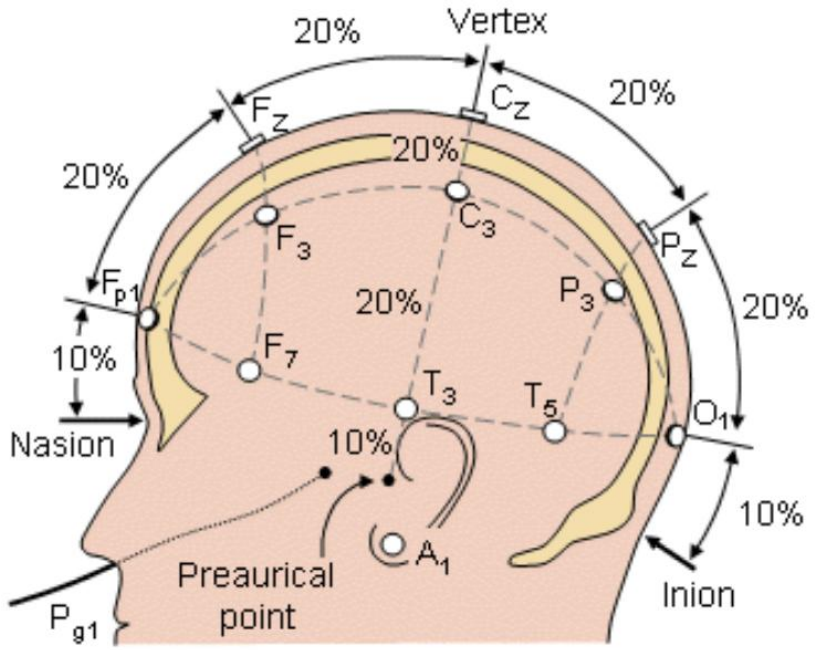
Imported data events

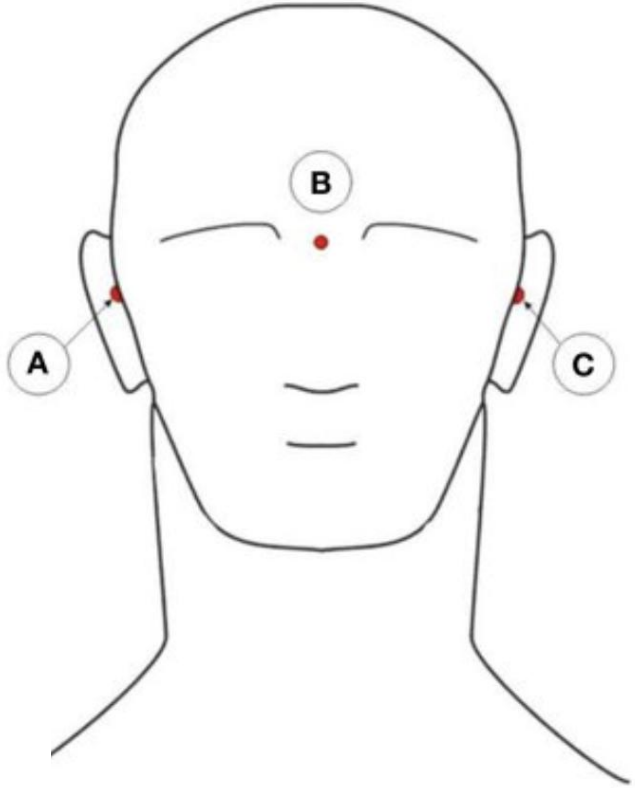


If event import was successful, you will see an appropriate number here

Channel labels & locations





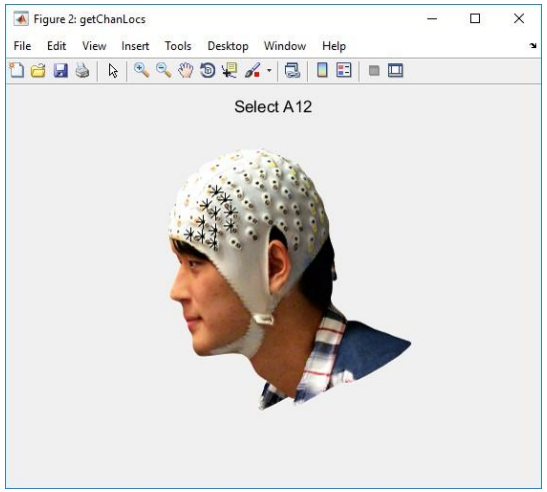


Scanning electrode position will become standard

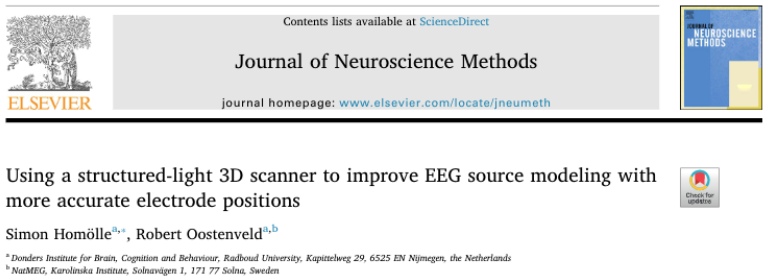
Get_chanlocs EEGLAB plugin interfacing Fieldtrip's functions

3-D camera mounted on iPad

EEGLAB plugin interface



iPhone 13 pro and above
& Samsung Galaxy S20 Ultra
can do depth scanning



Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

ELSEVIER

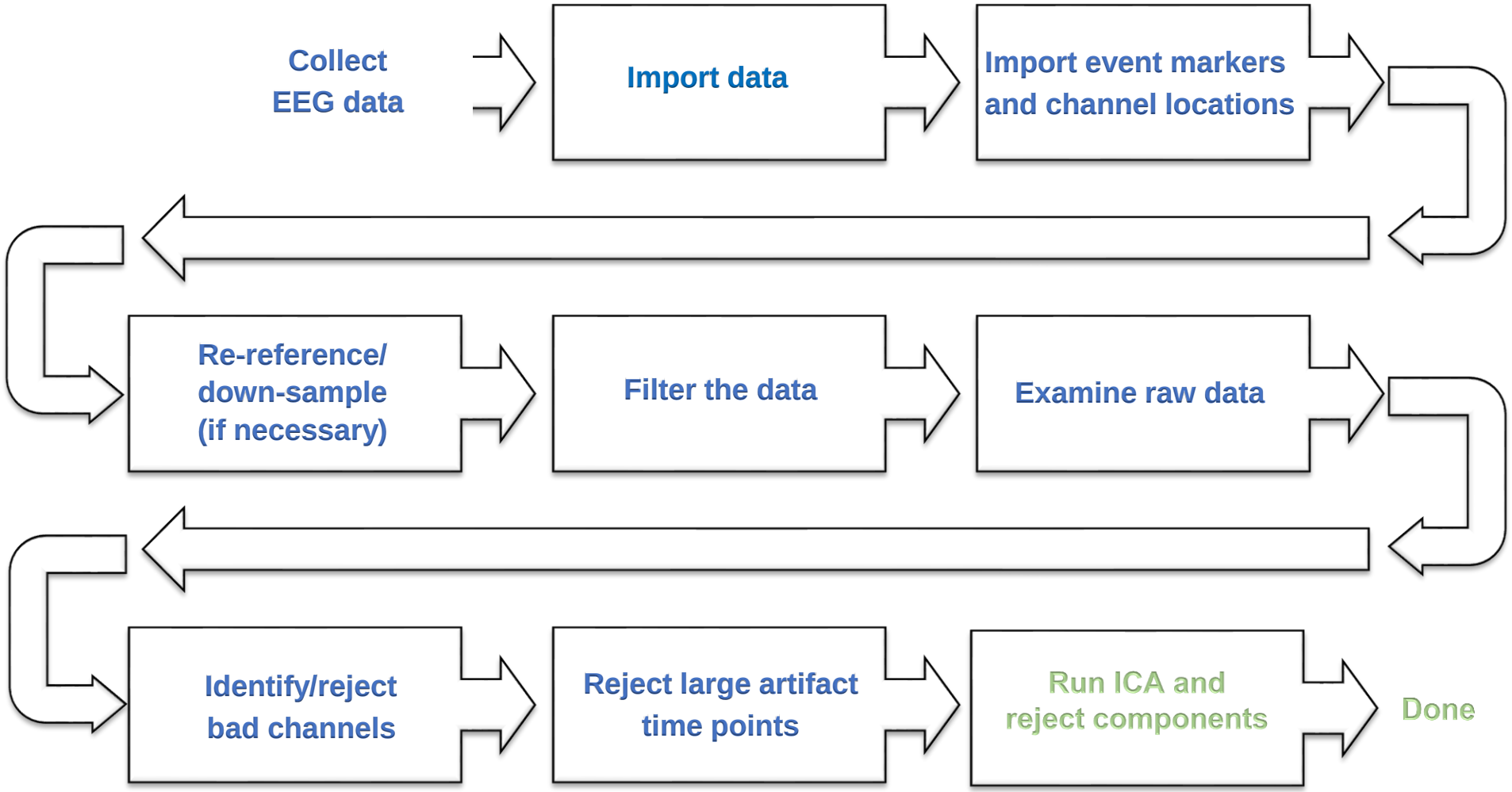
Using a structured-light 3D scanner to improve EEG source modeling with more accurate electrode positions

Simon Homölle[a,*], Robert Oostenveld[a,b]

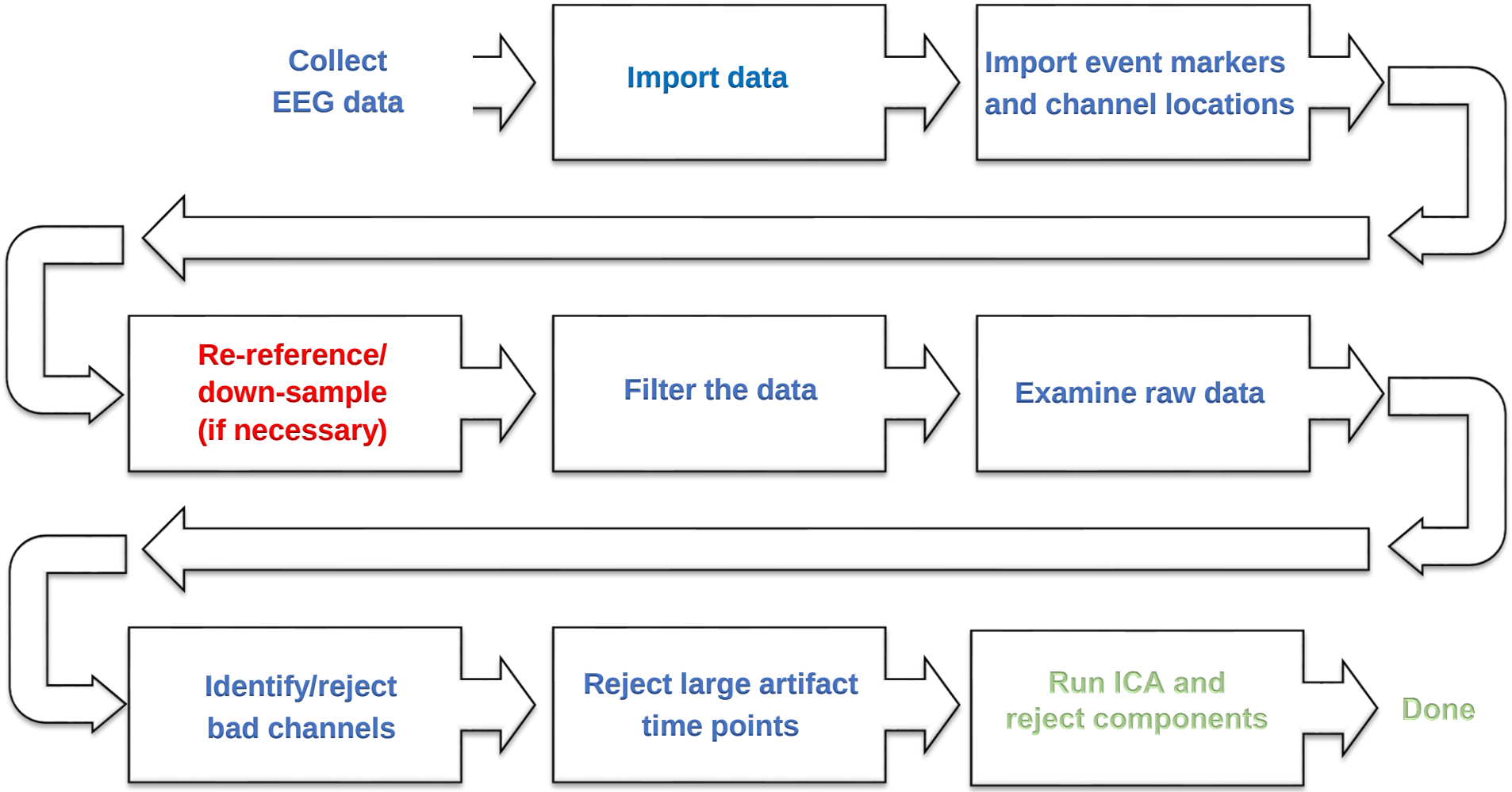
[a] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Kapittelweg 29, 6525 EN Nijmegen, the Netherlands
[b] NatMEG, Karolinska Institute, Solnavägen 1, 171 77 Solna, Sweden

- The EEGLAB and FieldTrip teams plan to work together to automatically align scans with templates and speed up the manual electrode labeling process (after training, currently 15 minutes per subject for 64 channels)

Pre-processing pipeline

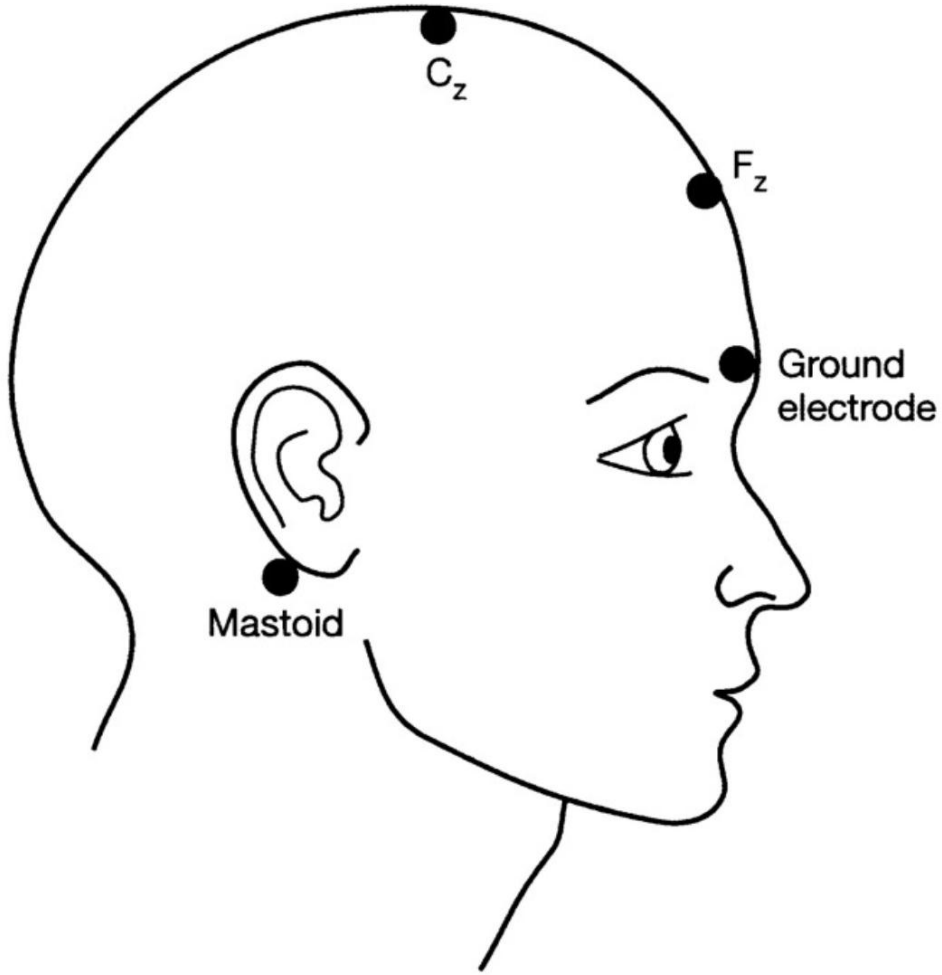


Pre-processing pipeline

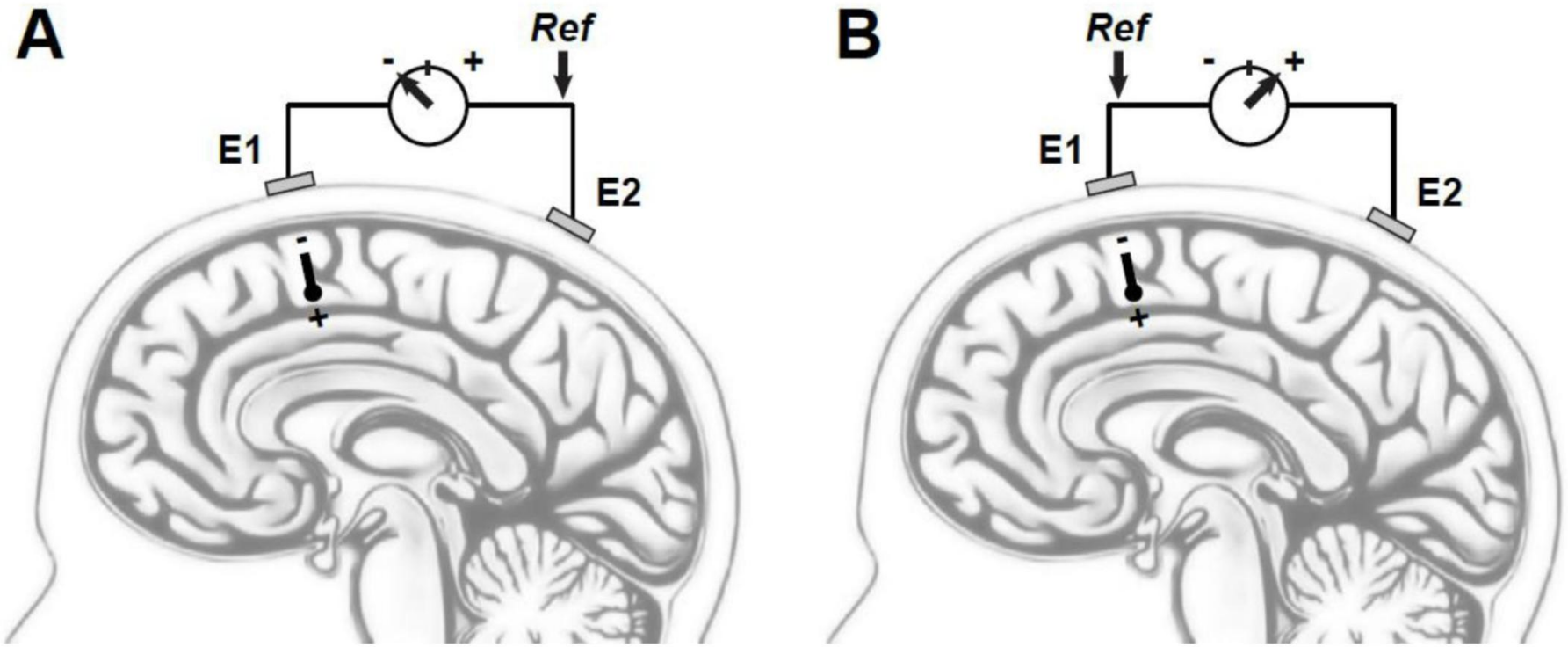


Referencing

- Earlobes
- Nose
- Average
- Mastoids
- Vertex (Cz), scalp electrode
- Bipolar
- Infinity reference

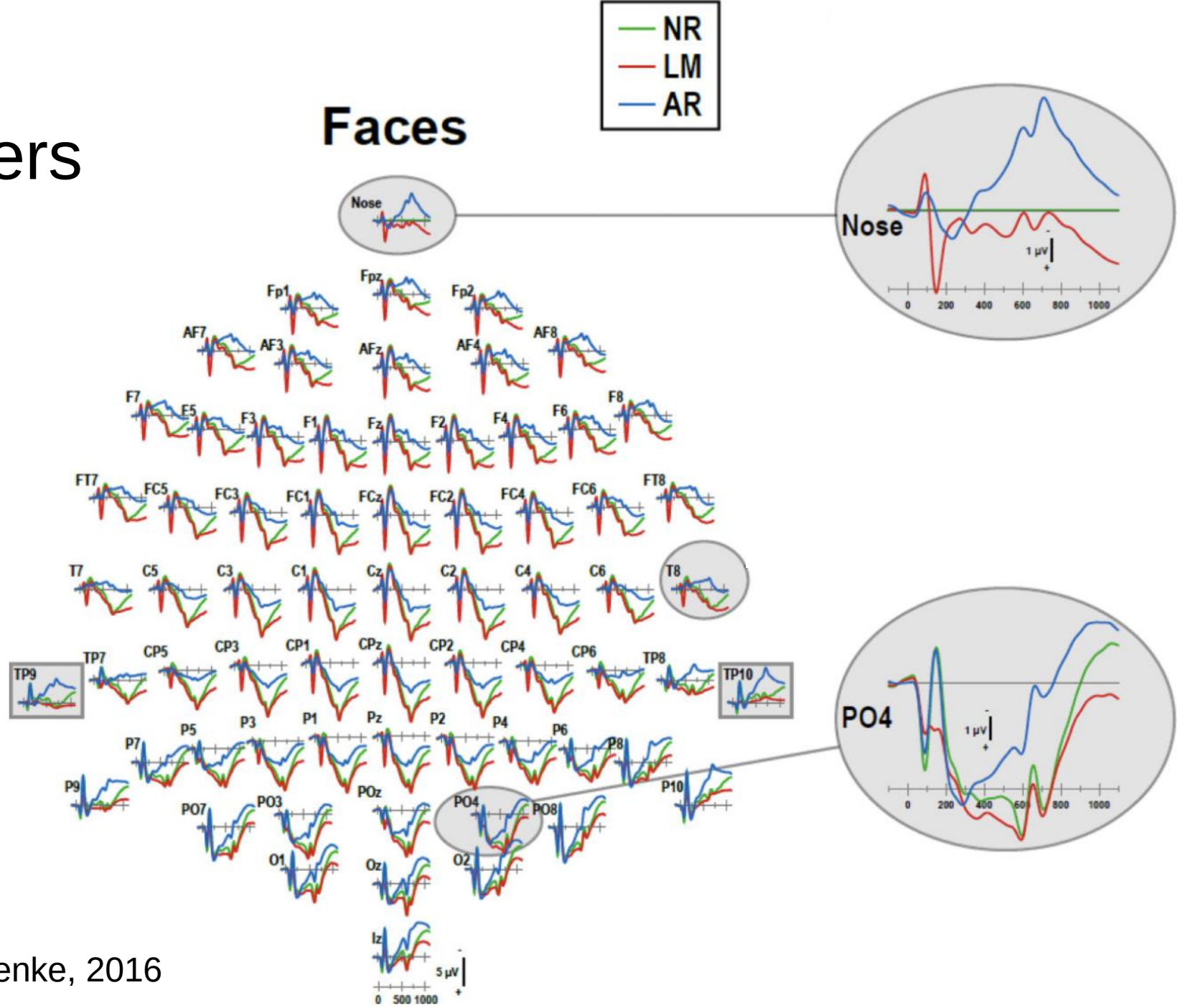


The location of the Reference matters



Kayser & Tenke, 2016

The location of Reference matters

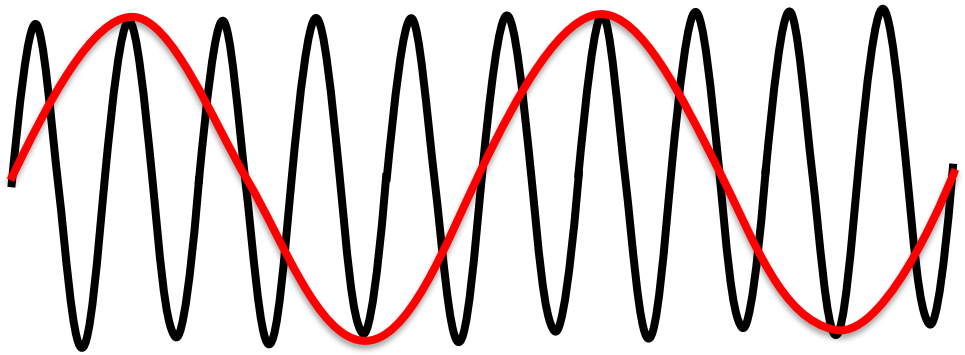


Kayser & Tenke, 2016

Downsampling data → 500 Hz is usually enough

Reasons for lowering sampling rate

- Reduce time and computational cost
- Most MEEG processes are below 100 Hz. Maybe no need to keep a sample rate at 500Hz (beware of the Nyquist frequency)

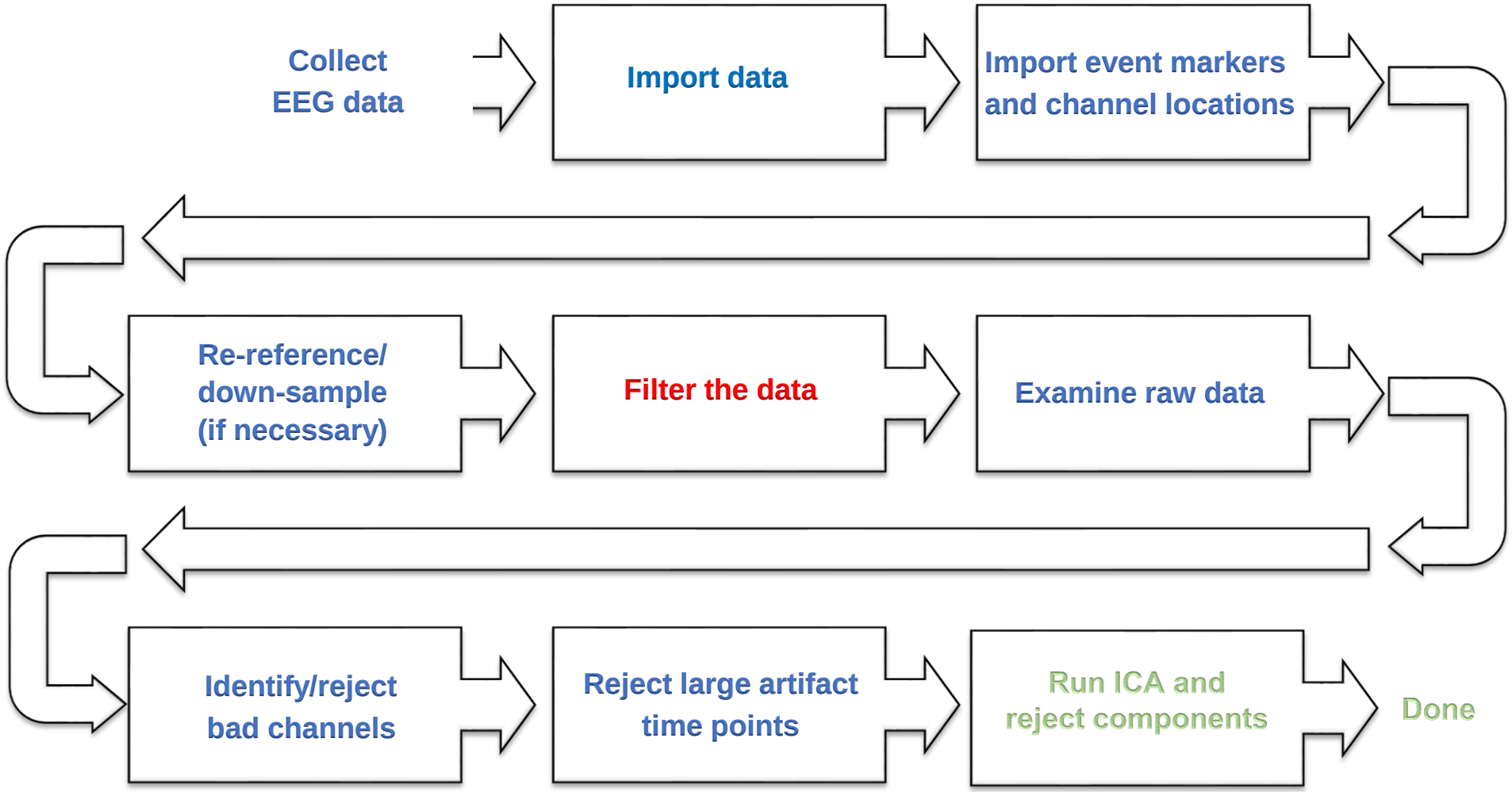


Reasons for NOT lowering sampling rate

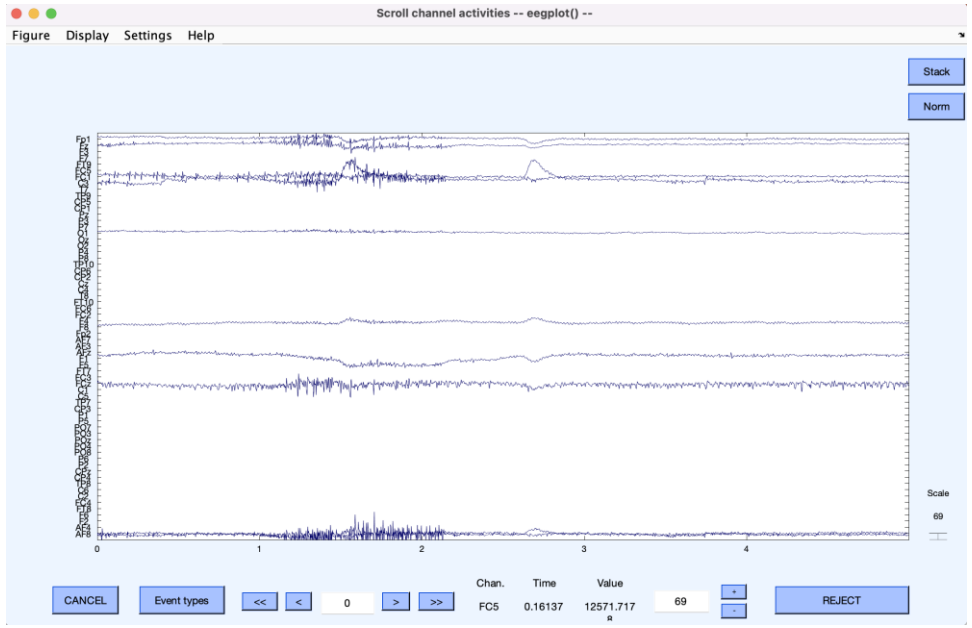
- Even with MEEG amplifiers noise, information above 250 Hz (500 Hz sampling rate) might be useful for some algorithm (e.g., Independent Component Analysis)
- Behavioral responses are measured on the order of millisecond – keep MEEG at the same time scale?

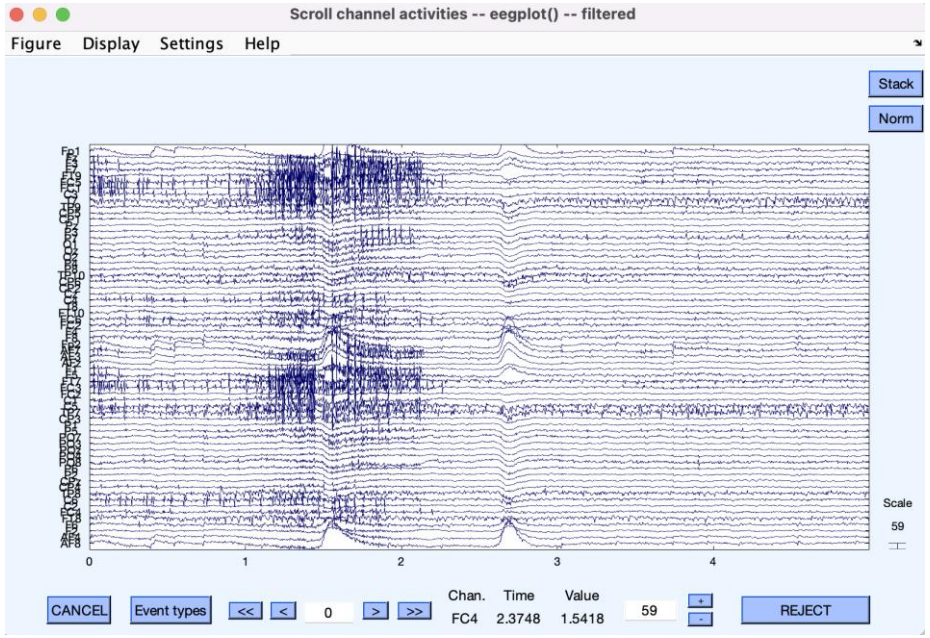
Pre-processing pipeline

01.



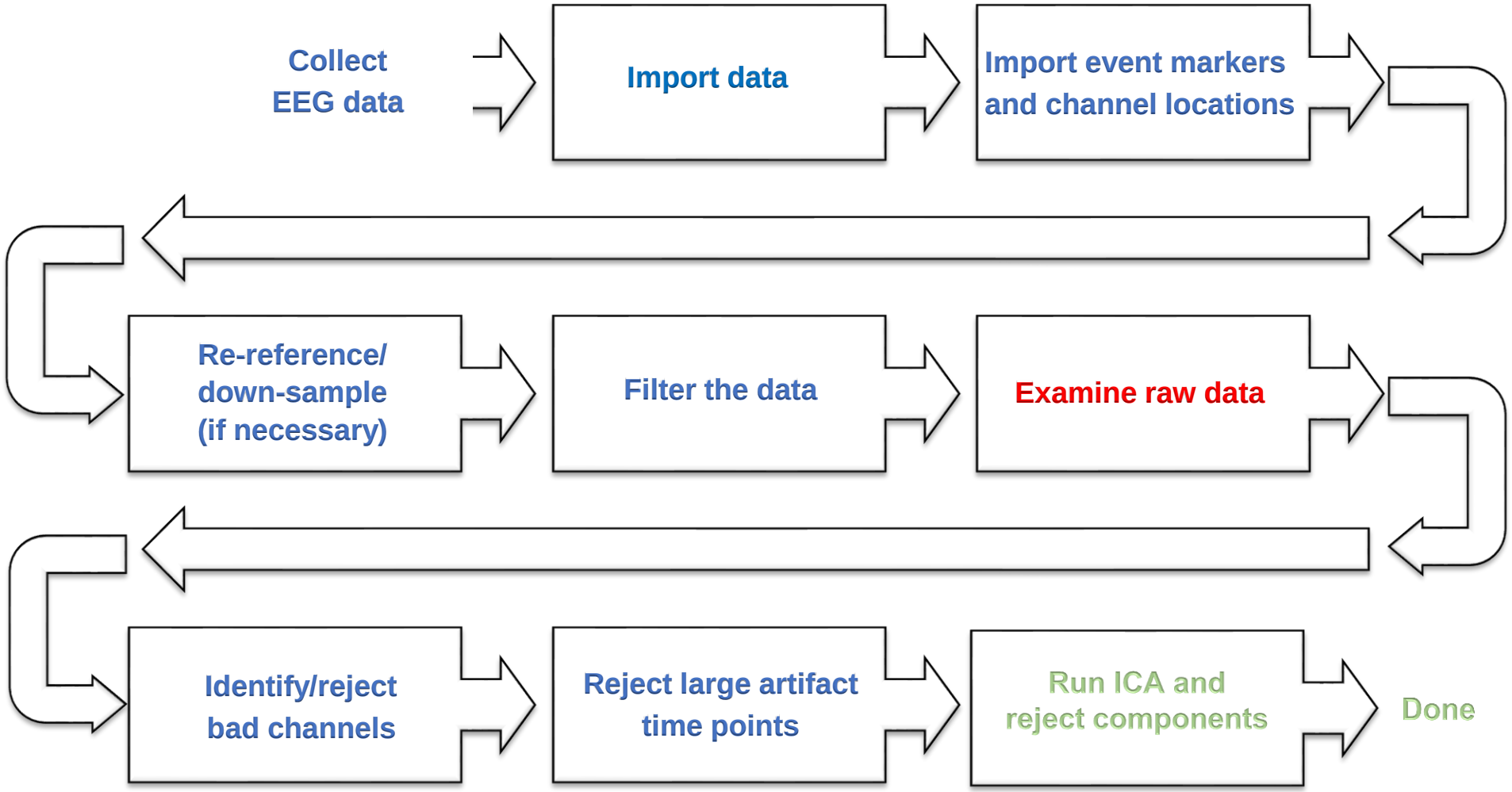
Highpass filter the data





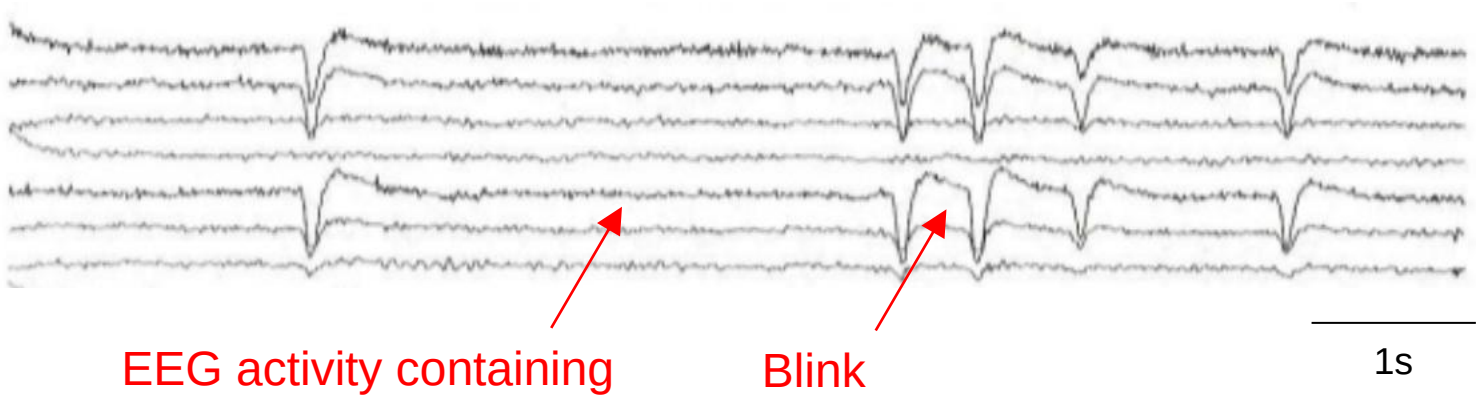
Pre-processing pipeline

01.



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

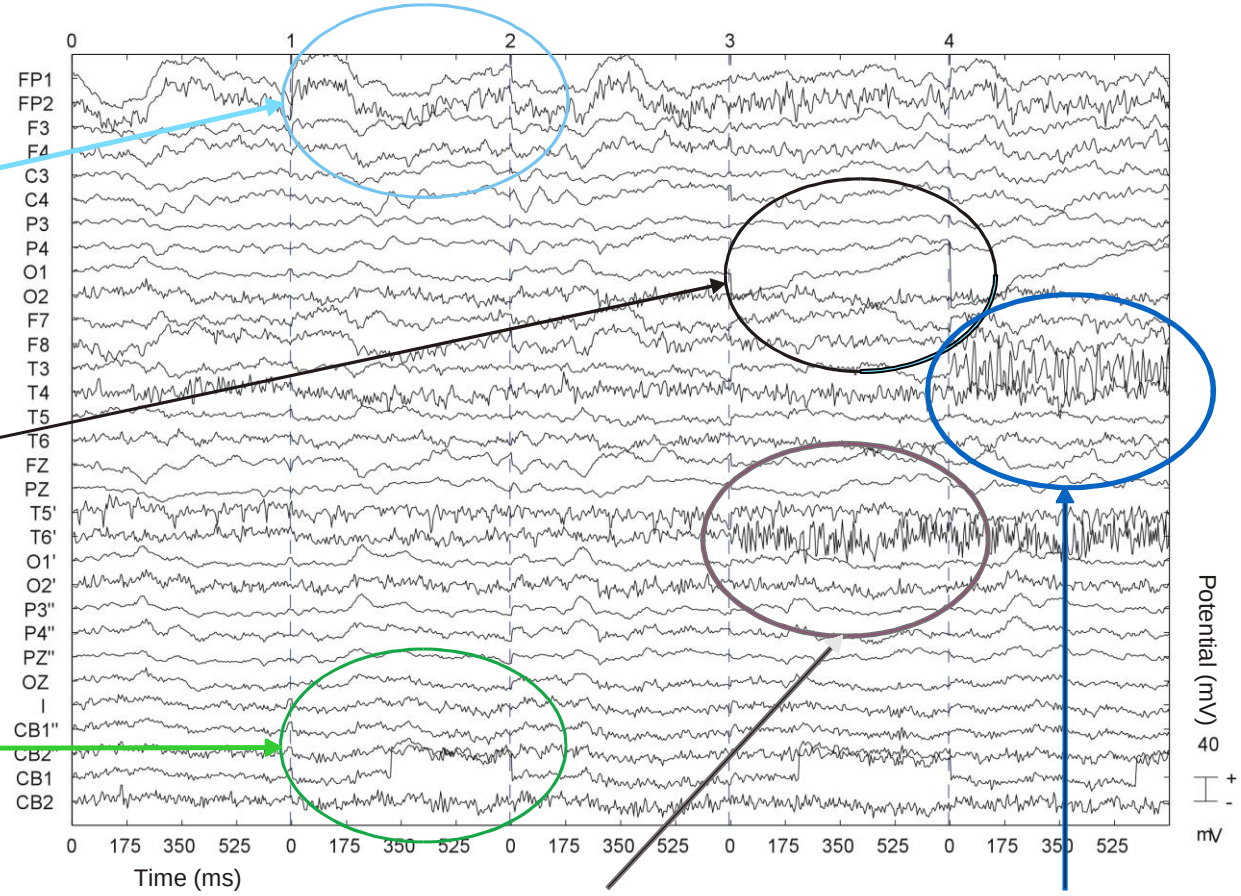


Type of artifacts

2 - Low frequency event (eye movements)

5 - Linear trend

3 - Discontinuity

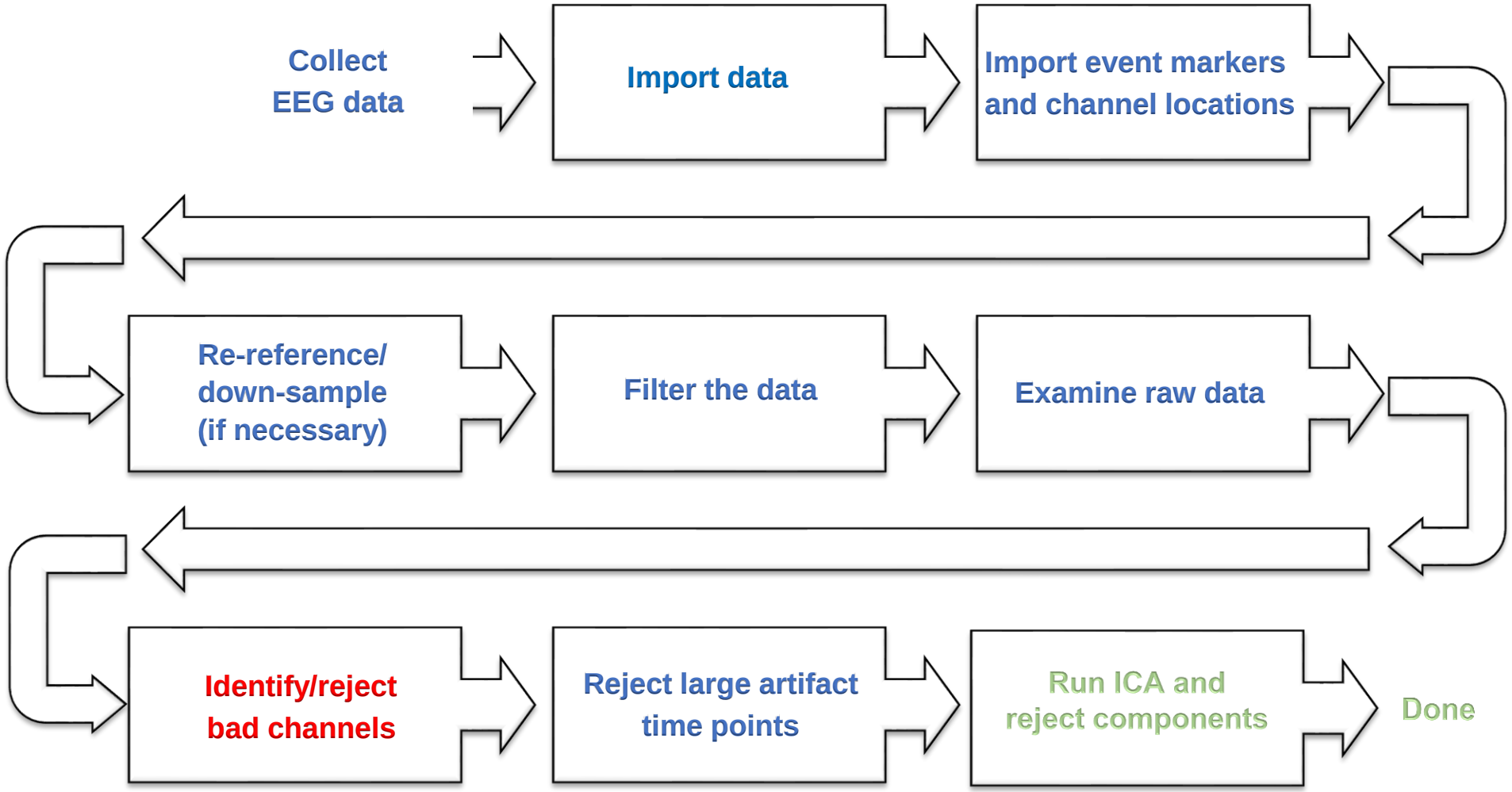


4 - High noise

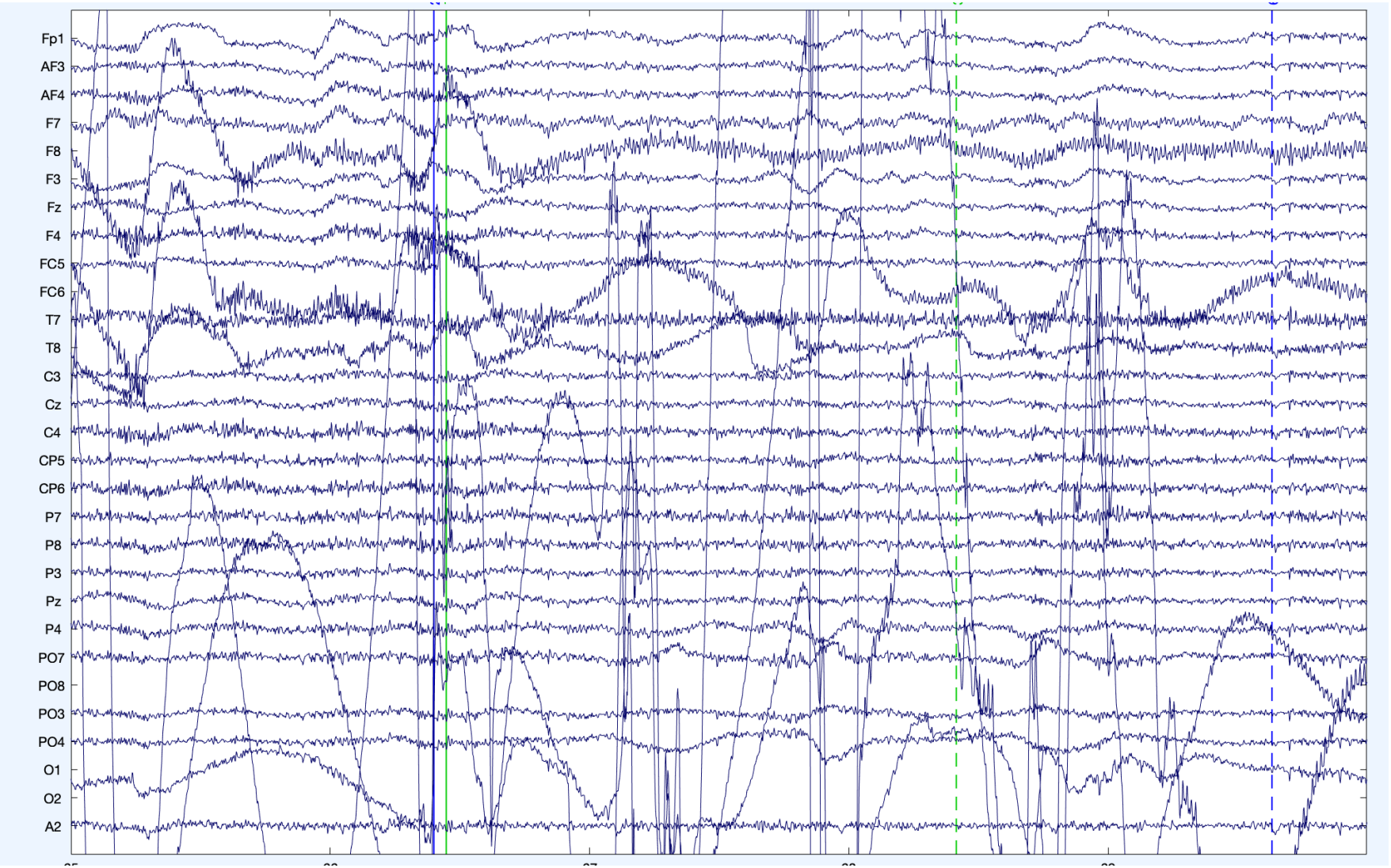
1 - Transient high frequency event (muscle)

Pre-processing pipeline

01.

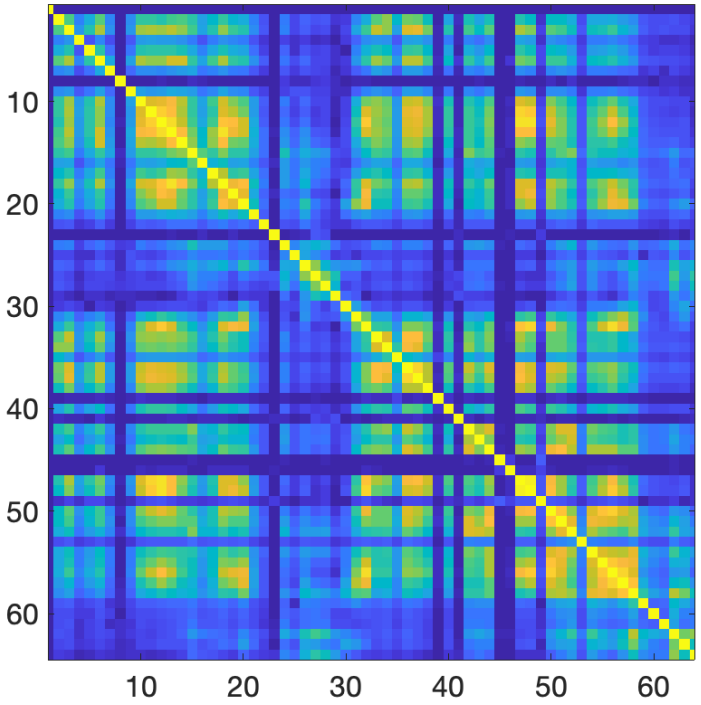


Looking for bad channels

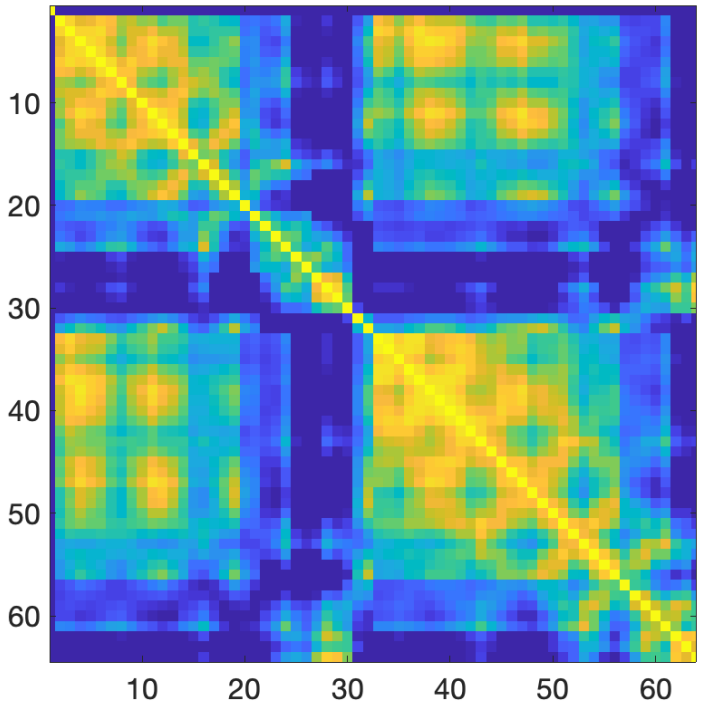


Pairwise correlation to find bad channels

Bad data

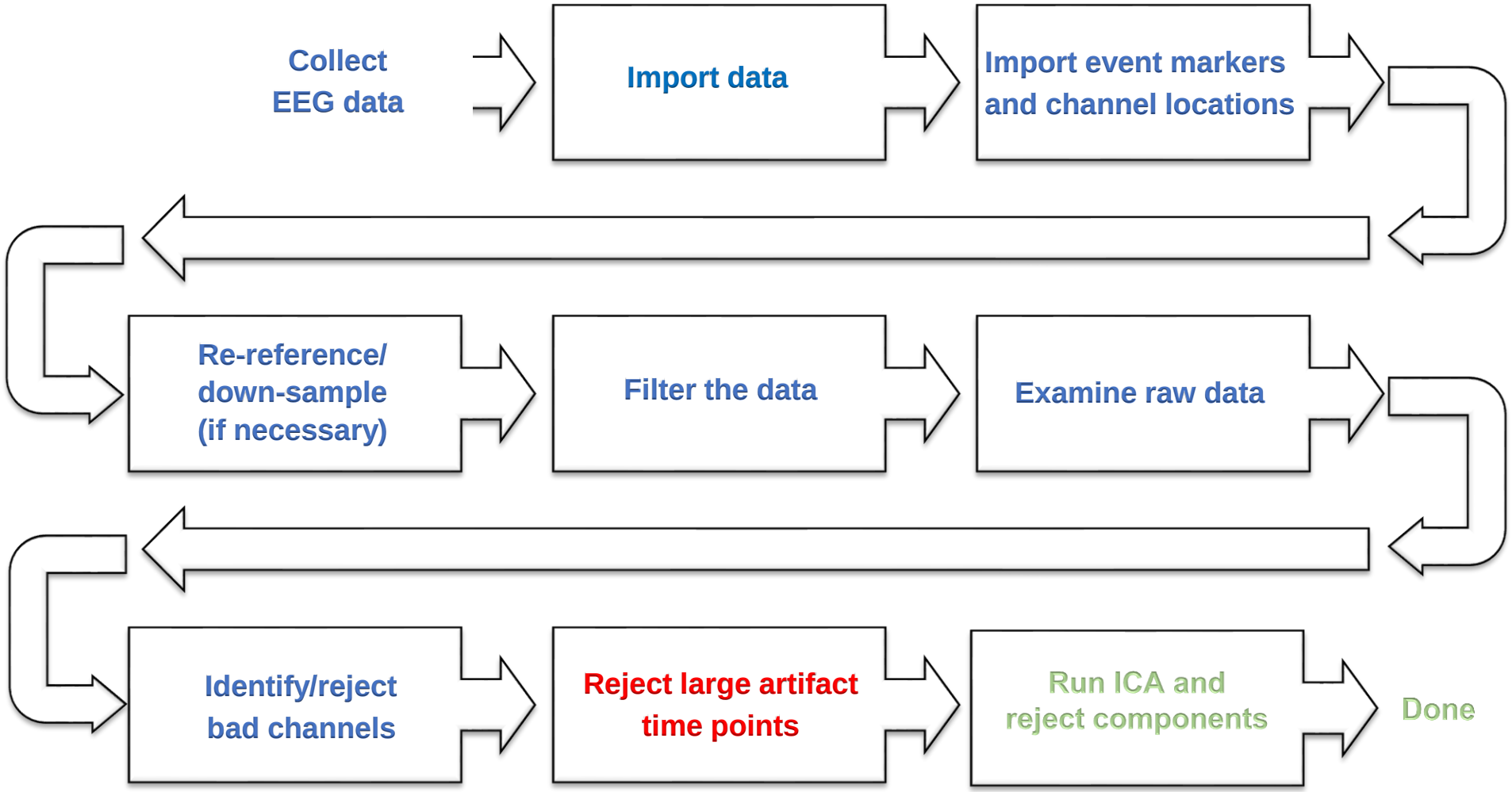


Good data

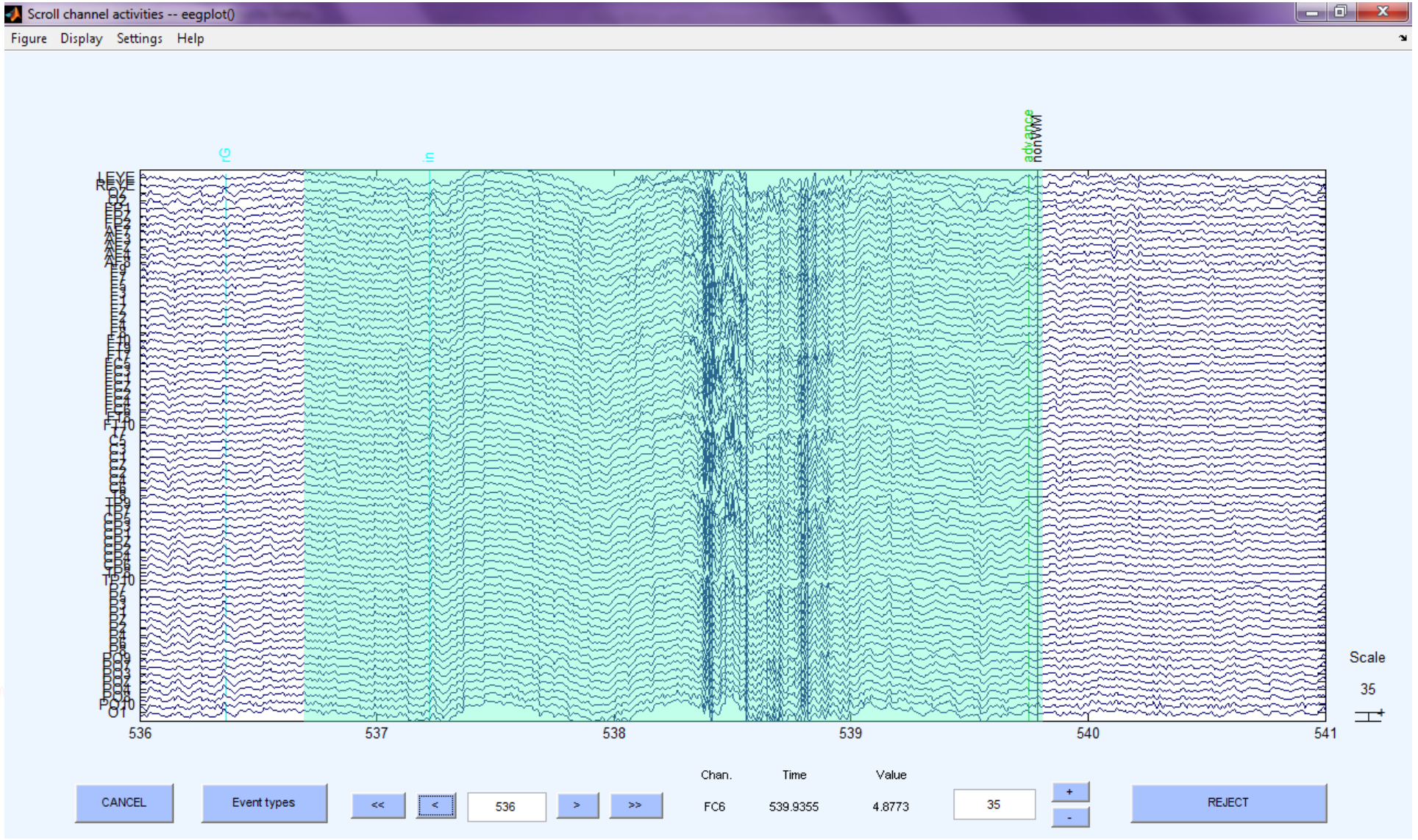


Tim R. Mullen, Christian Kothe, et al.(2015) Real-time neuroimaging and cognitive monitoring using wearable dry EEG. IEEE Transactions on Biomedical Engineering. DOI:10.1109/TBME.2015.2481482

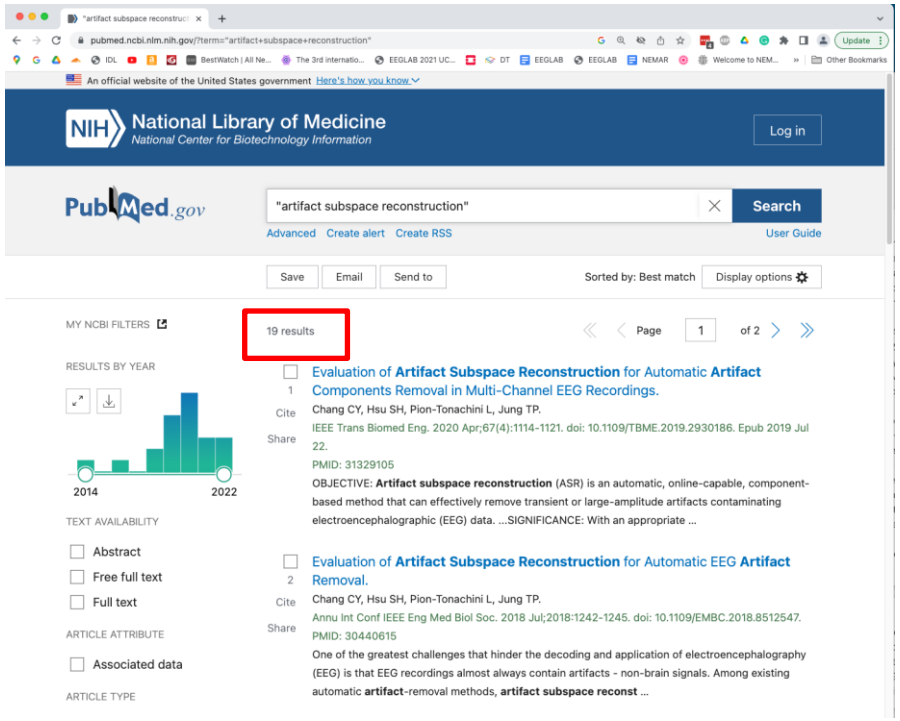
Pre-processing pipeline

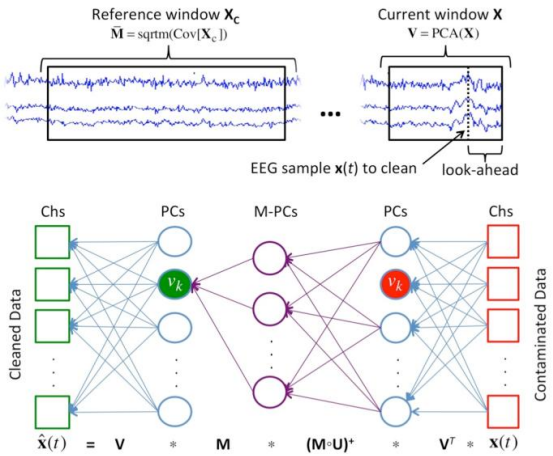


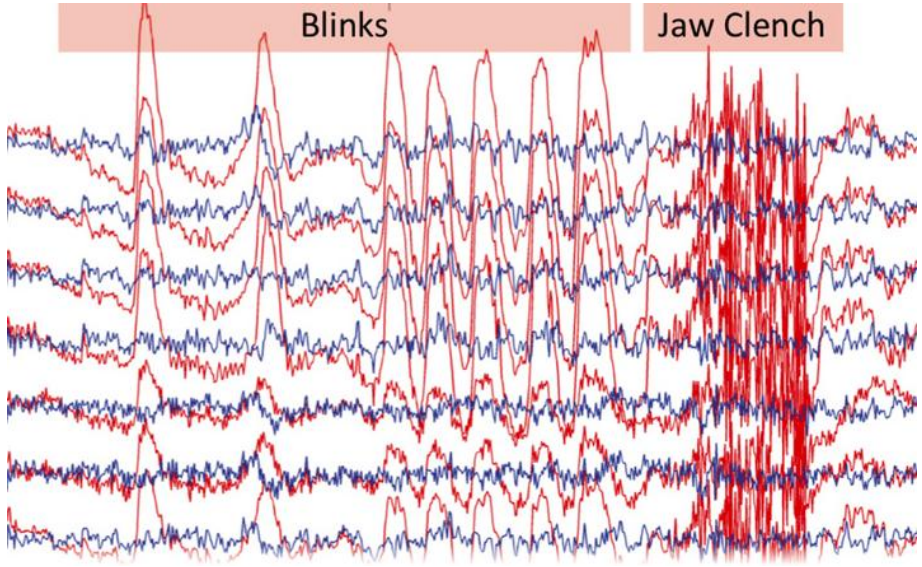
Rejecting continuous data



The powerful ASR method







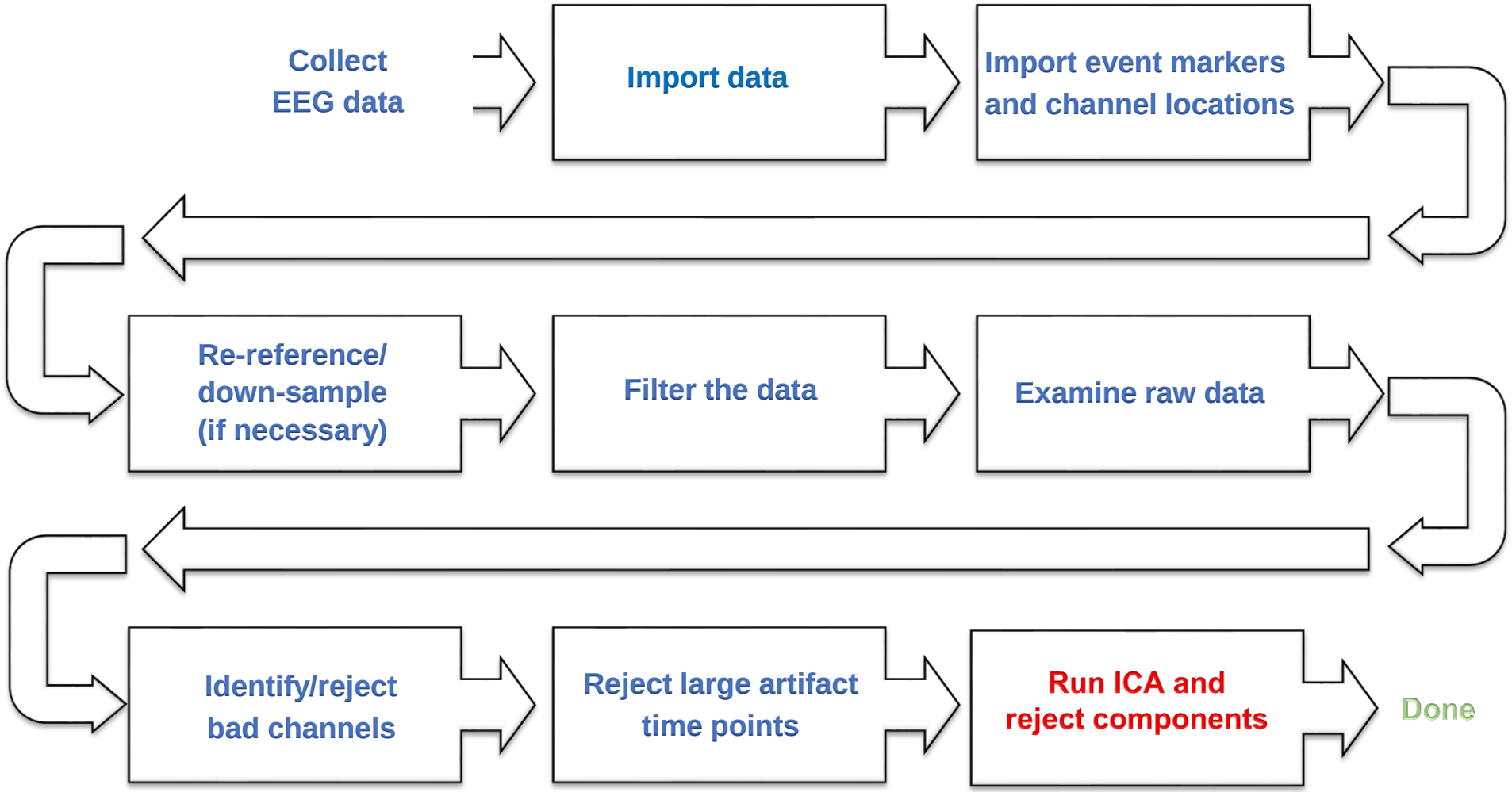
- For offline analysis, ASR is usually used to **detect and remove** but not correct data

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

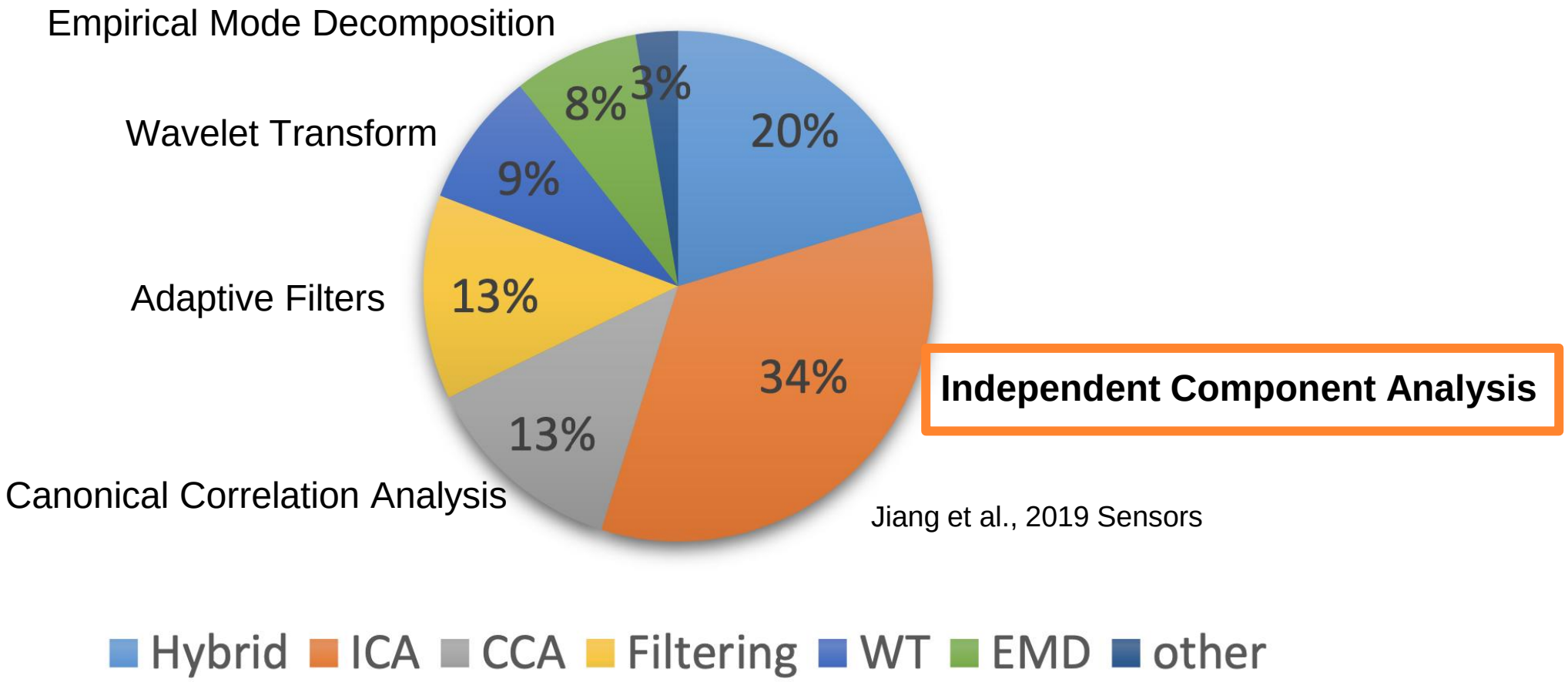
A. Delorme and J. A. Martin, "Automated Data Cleaning for the Muse EEG," *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2021, pp. 1-5, doi: 10.1109/BIBM52615.2021.9669415.

Pre-processing pipeline

01.



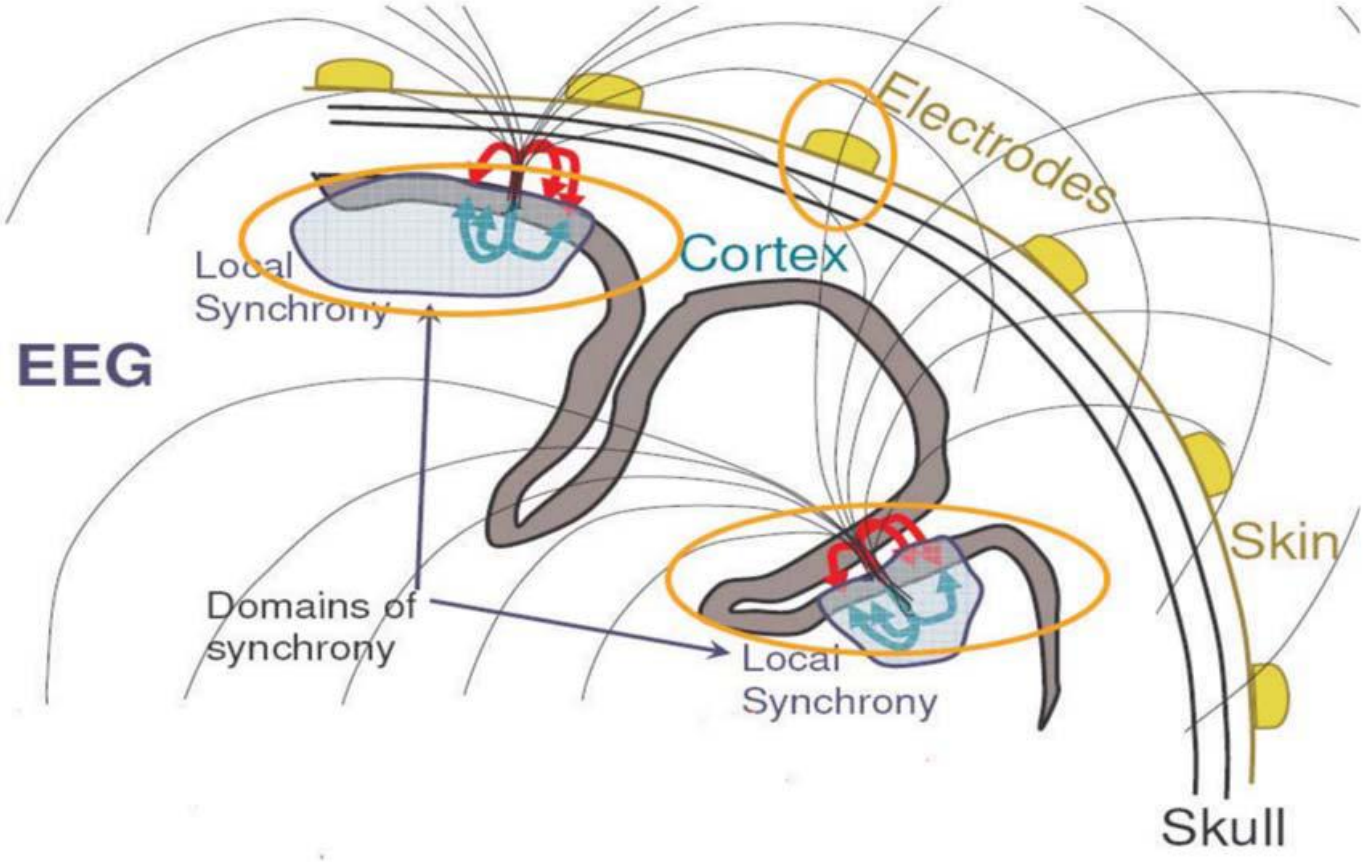
Artefact correction methods



Independent Component Analysis

Jiang et al., 2019 Sensors

Independent Component Analysis



ICA – separates the EEG in temporally independent source signals (Makeig, 1996)

ICLabel Website and Label Collection

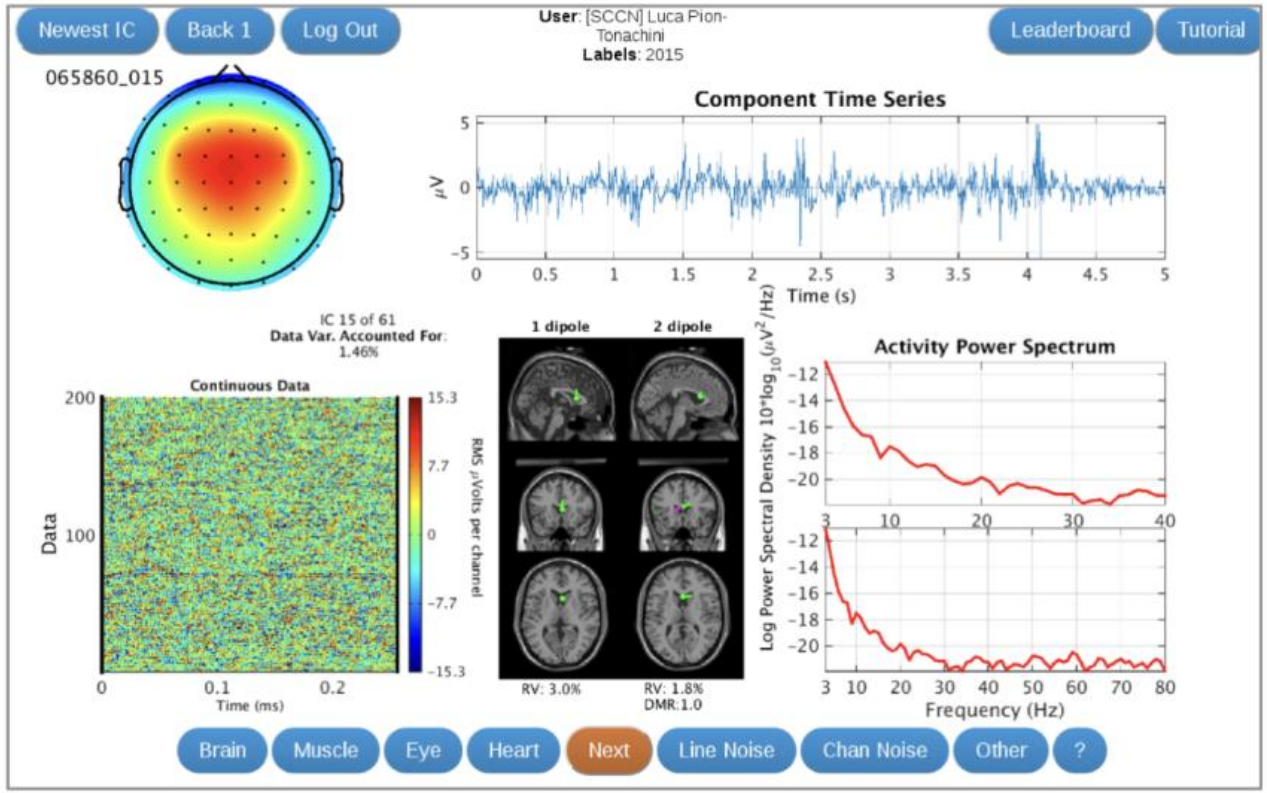
Summary: ask others to help out with labeling EEG components.

Website: labeling.ucsd.edu

Have experts from the SCCN and elsewhere label a subset.

Ask the EEGLAB community to help label a larger subset.

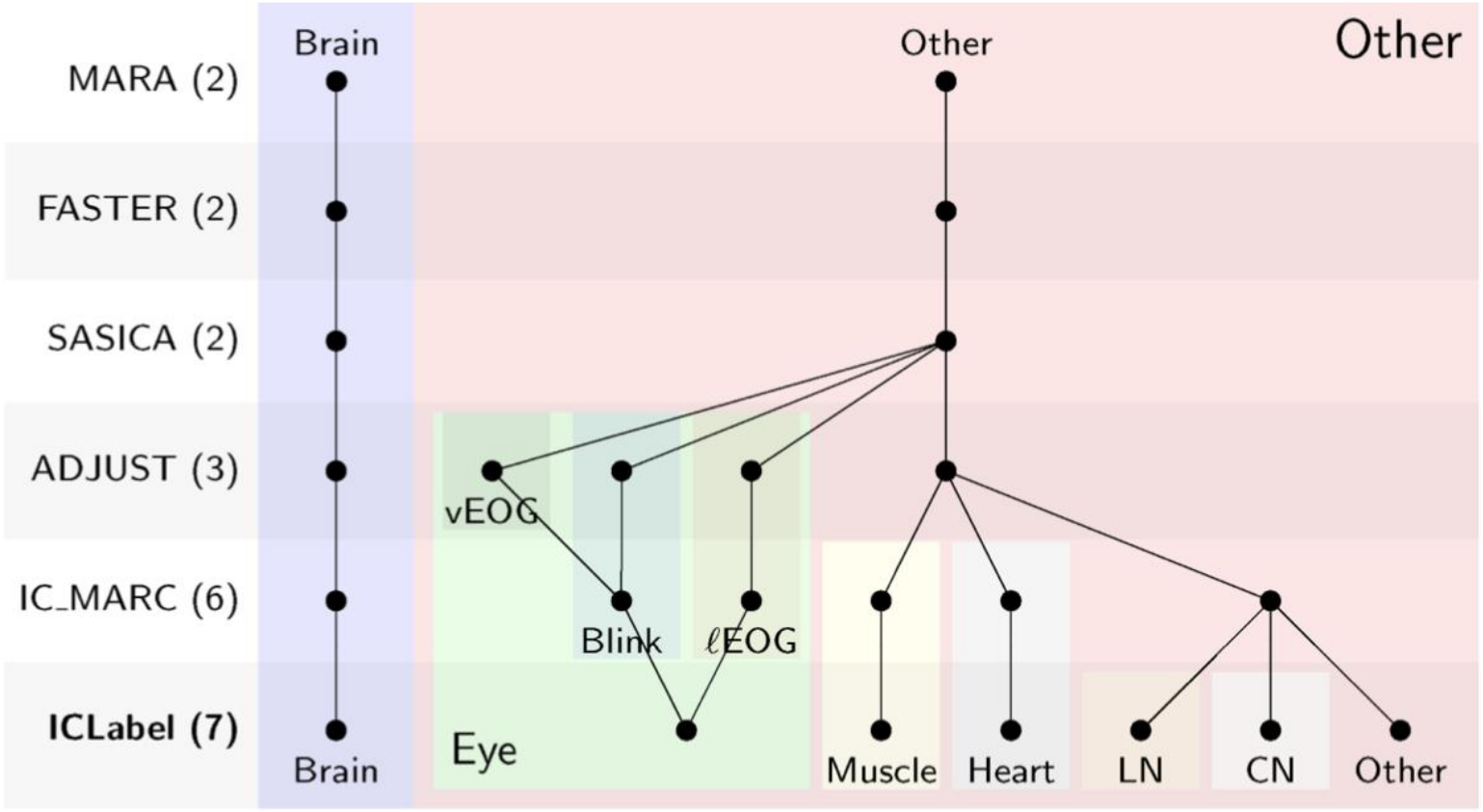
Currently 328 contributing users and 34,000+ submitted labels.

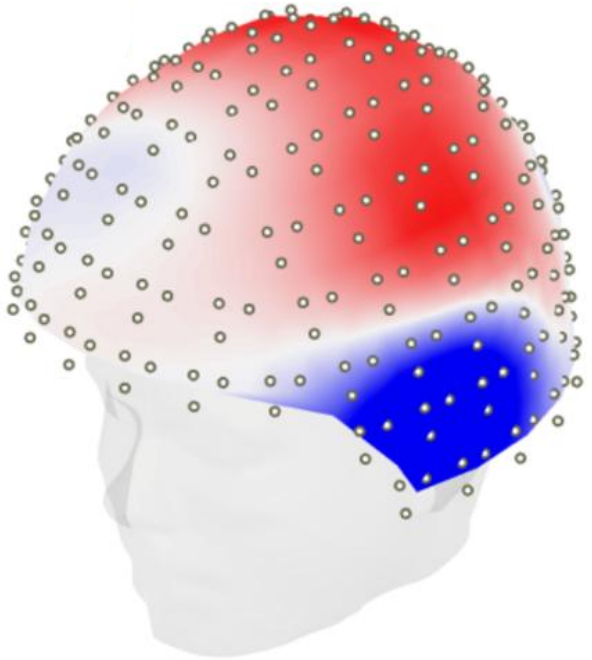


Has been adapted for educational use as well.

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). The ICLabel dataset of electroencephalographic (EEG) independent component (IC) features. *Data in Brief*, *25*. https://doi.org/10.1016/j.dib.2019.104101

Automated ICA classification methods





Do not use with MEG

Every method listed above is available as separate EEGLAB plugins

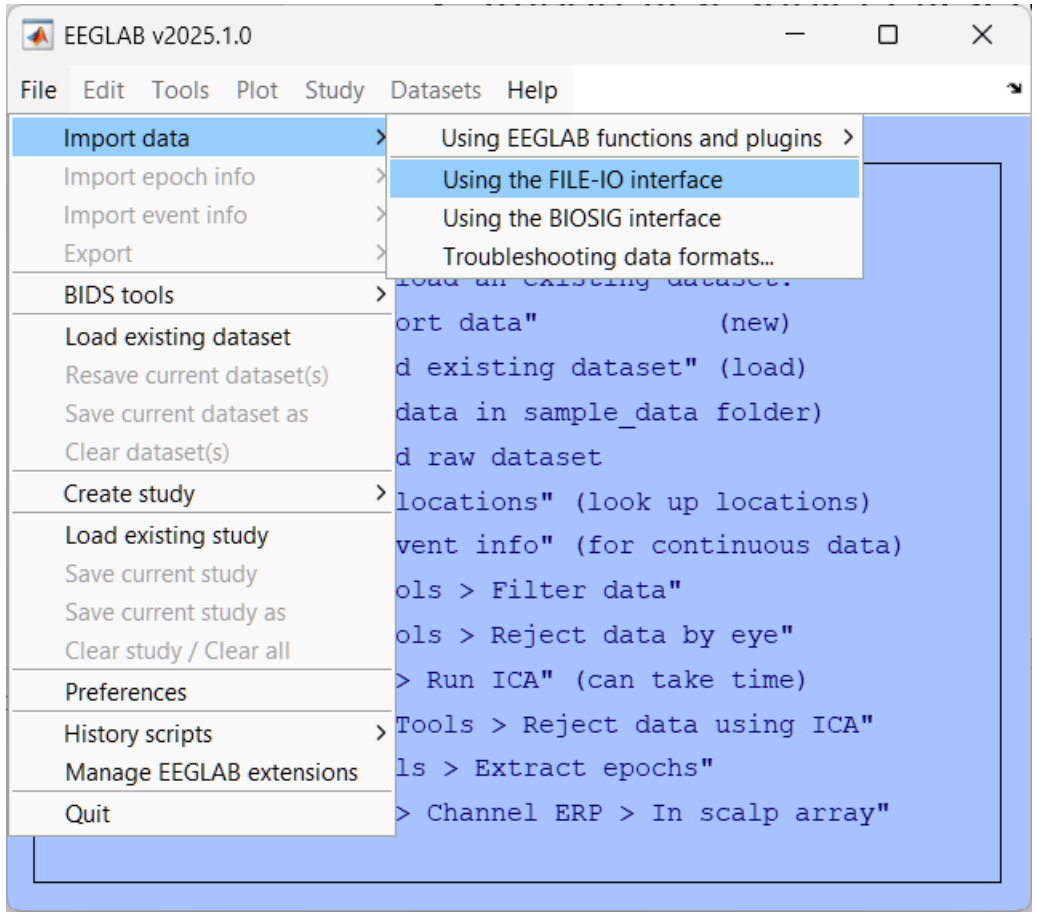
FieldTrip, MNE and Brainstorm can apply ICA (but automated classification are not available yet)

Hands on !

Follow along

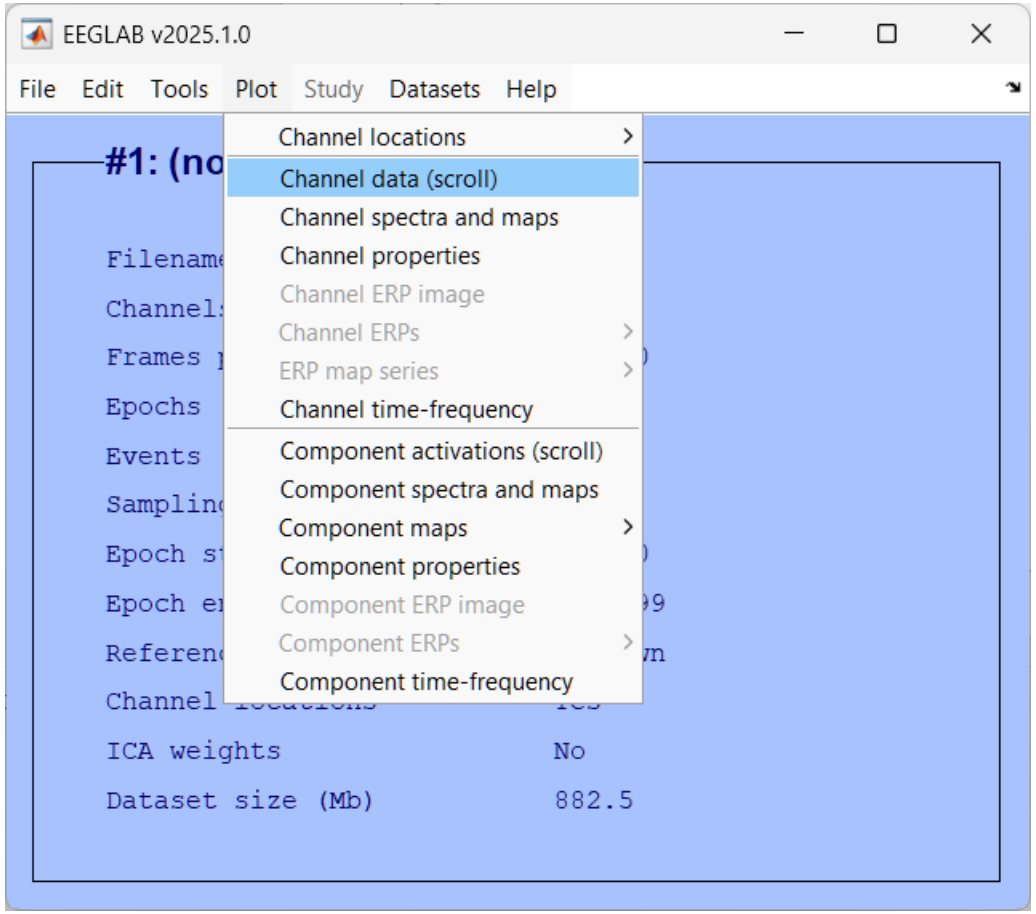
Import data

- Open matlab
- Start eeglab
- Import first subject first run meeg .fif datafile from ds000117 dataset :

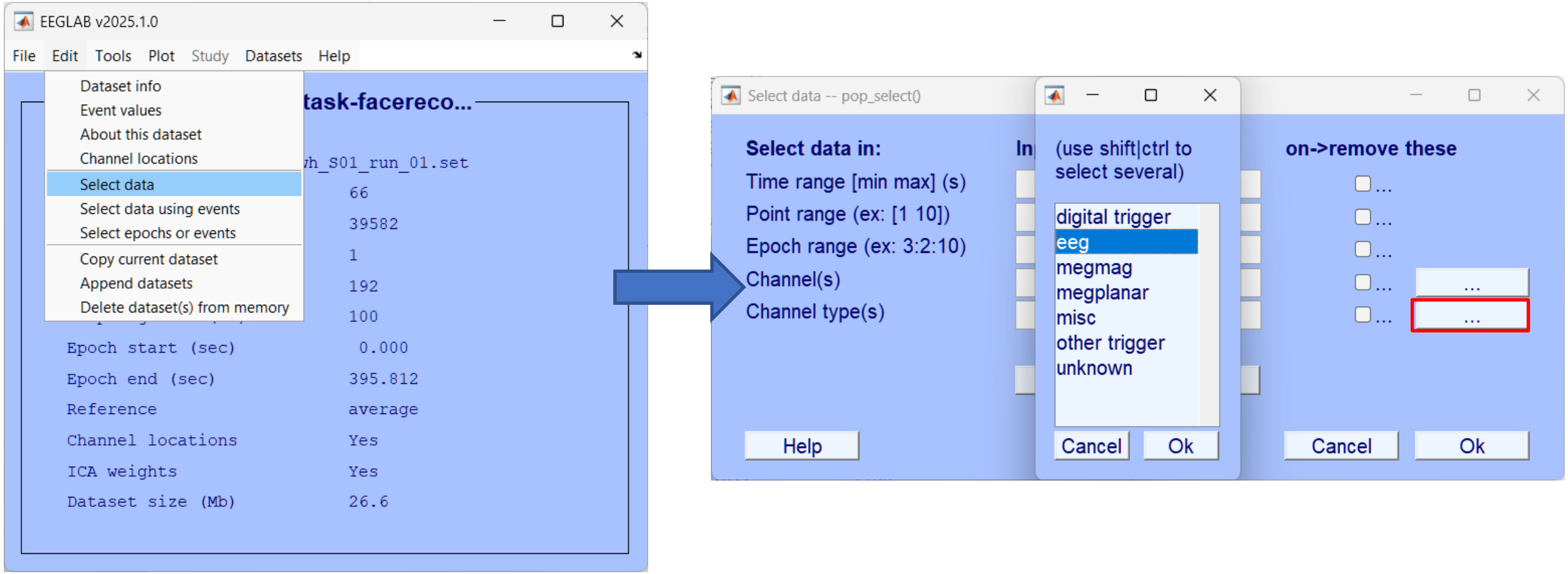


***Path** : ds000117_pruned\derivatives\meg_derivatives\sub-01\ses-meg\meg*
***File** : sub-01_ses-meg_task-facerecognition_run-01_proc-sss_meg.fif*

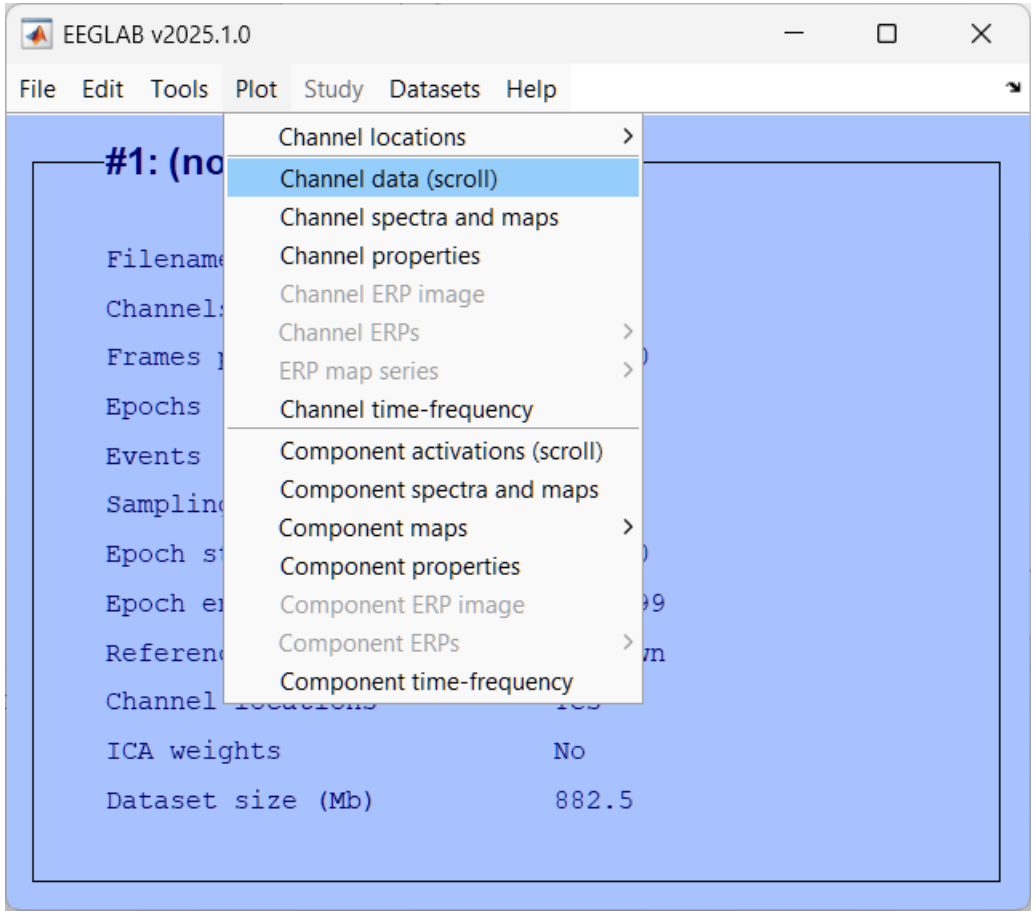
Plot channels timecourse



Select EEG channels



Check again signal timecourse



Execute full script

- PracticalMEEG_Session_1_Import_Data.m

More hands on !

Prepocessing pipeline on your own

Preprocess data

- Open script PracticalMEEG_Session_1_Preprocess_Data.m
- Start eeglab and try to reproduce script operations using GUI
- Check with *eegh* matlab command if you get it right