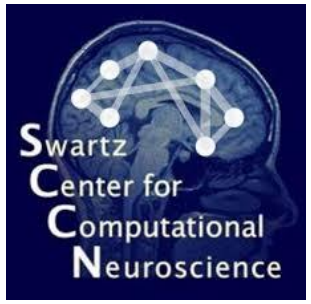
From raw data to robust connectivity analysis

Arnaud Delorme

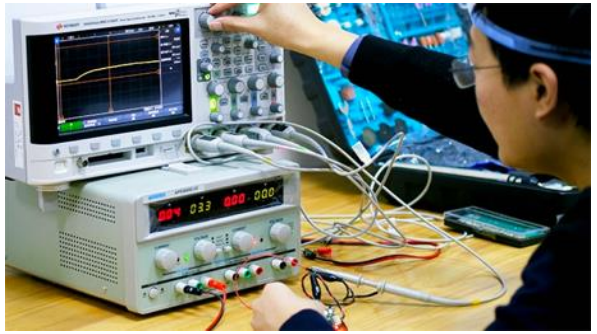
Swartz Center for Computational Neuroscience, UCSD, La Jolla, CA, USA



UC San Diego

Why preprocess data?

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

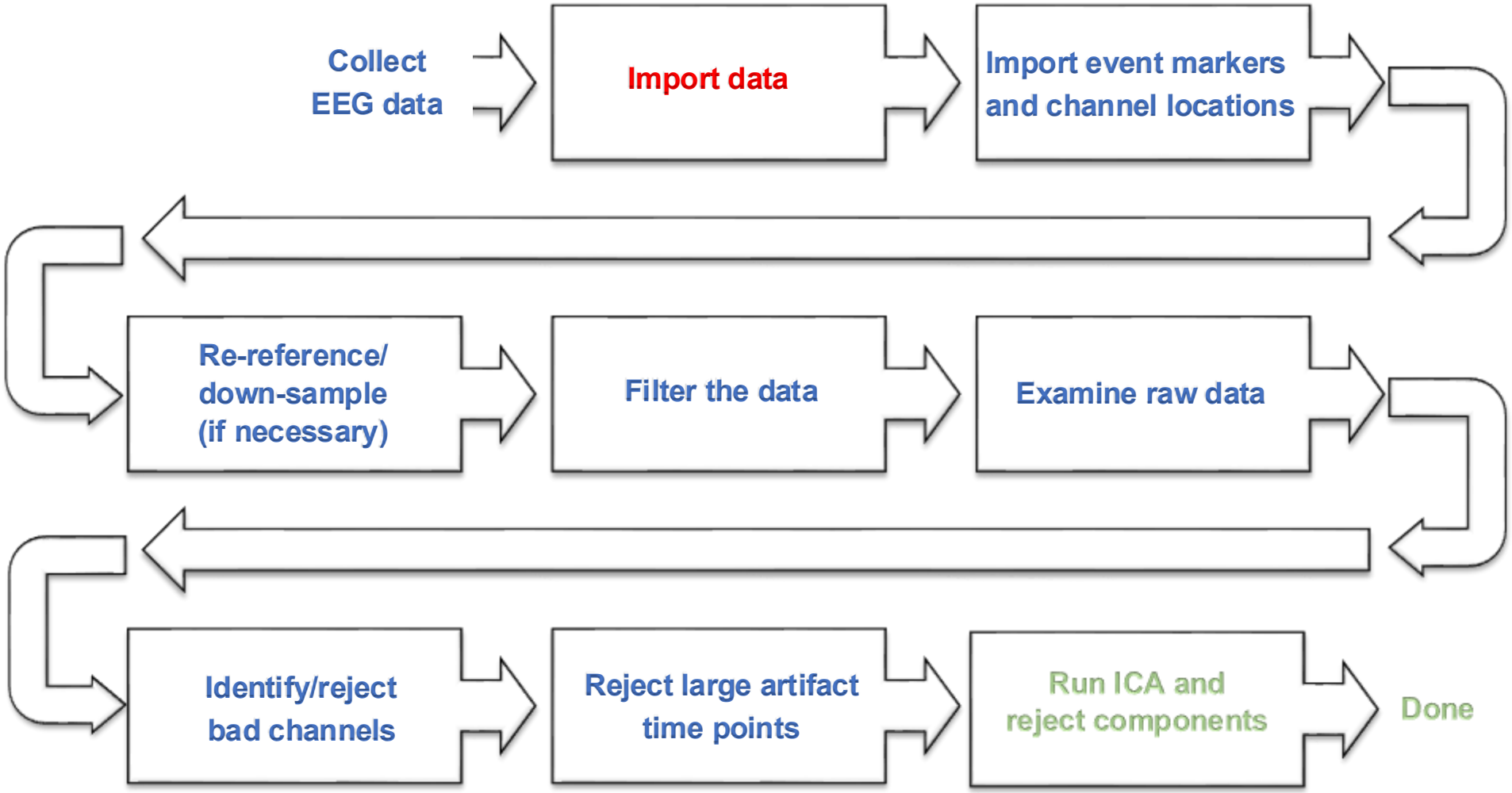


To make sense of the data, we need to:

- 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



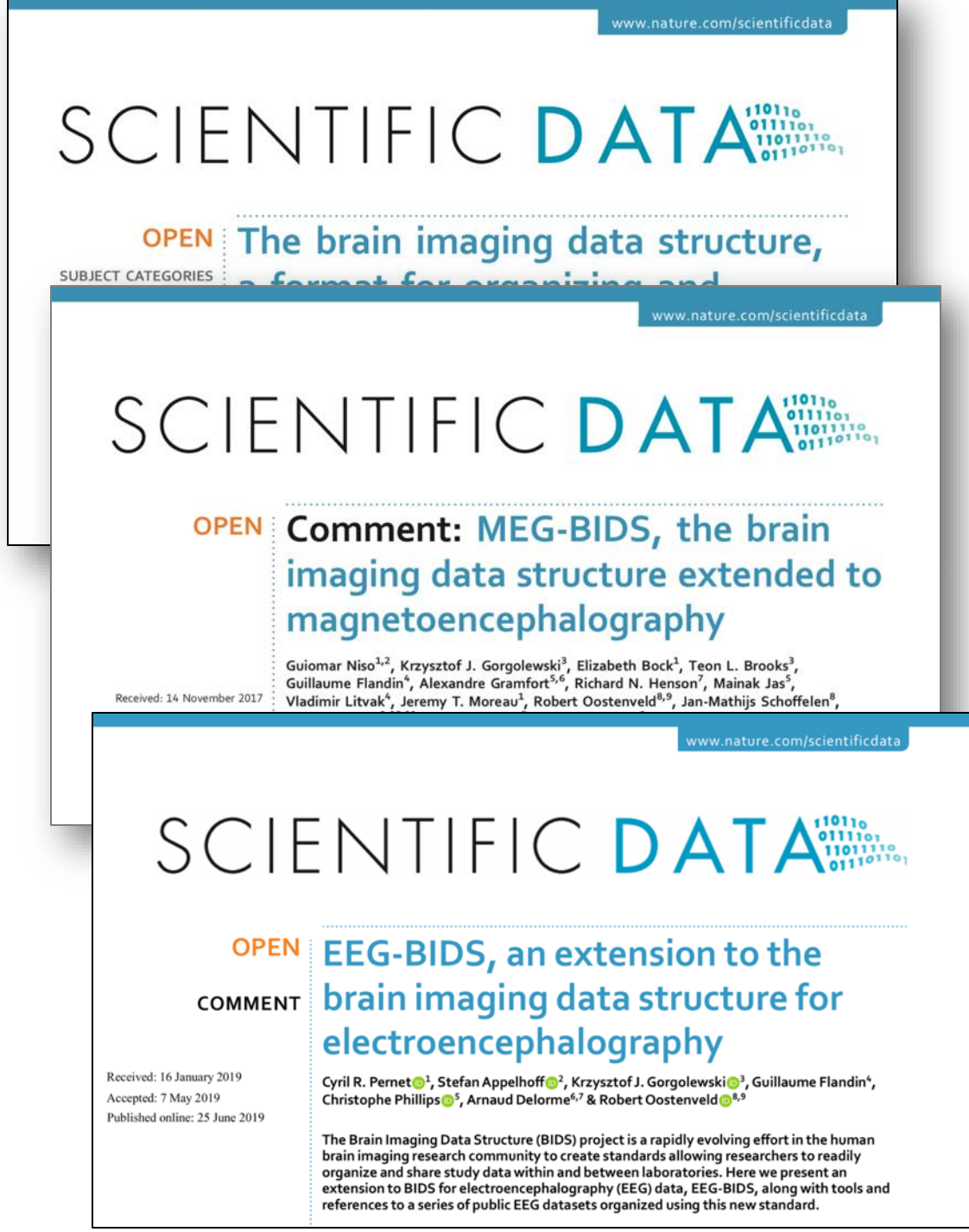
What is BIDS?

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

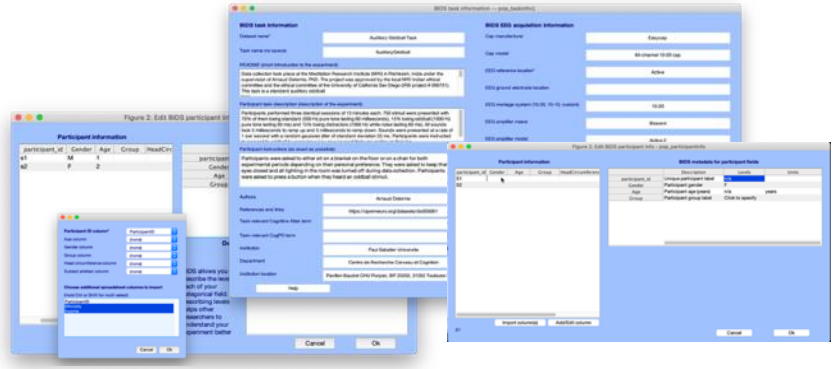
Brainstorm

Have dedicated tools for processing group-level data

FieldTrip

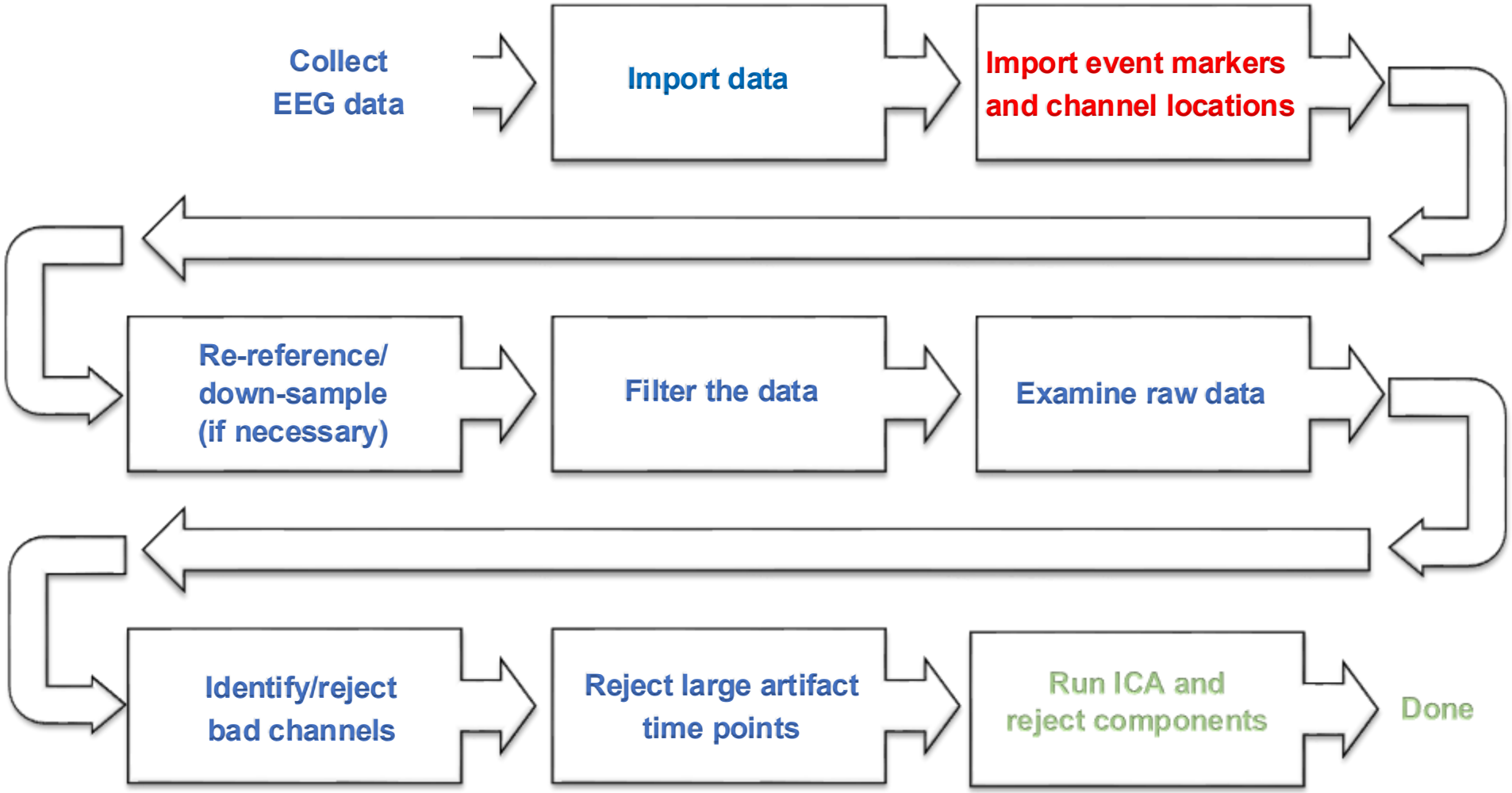
MNE
MEG + EEG ANALYSIS & VISUALIZATION

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

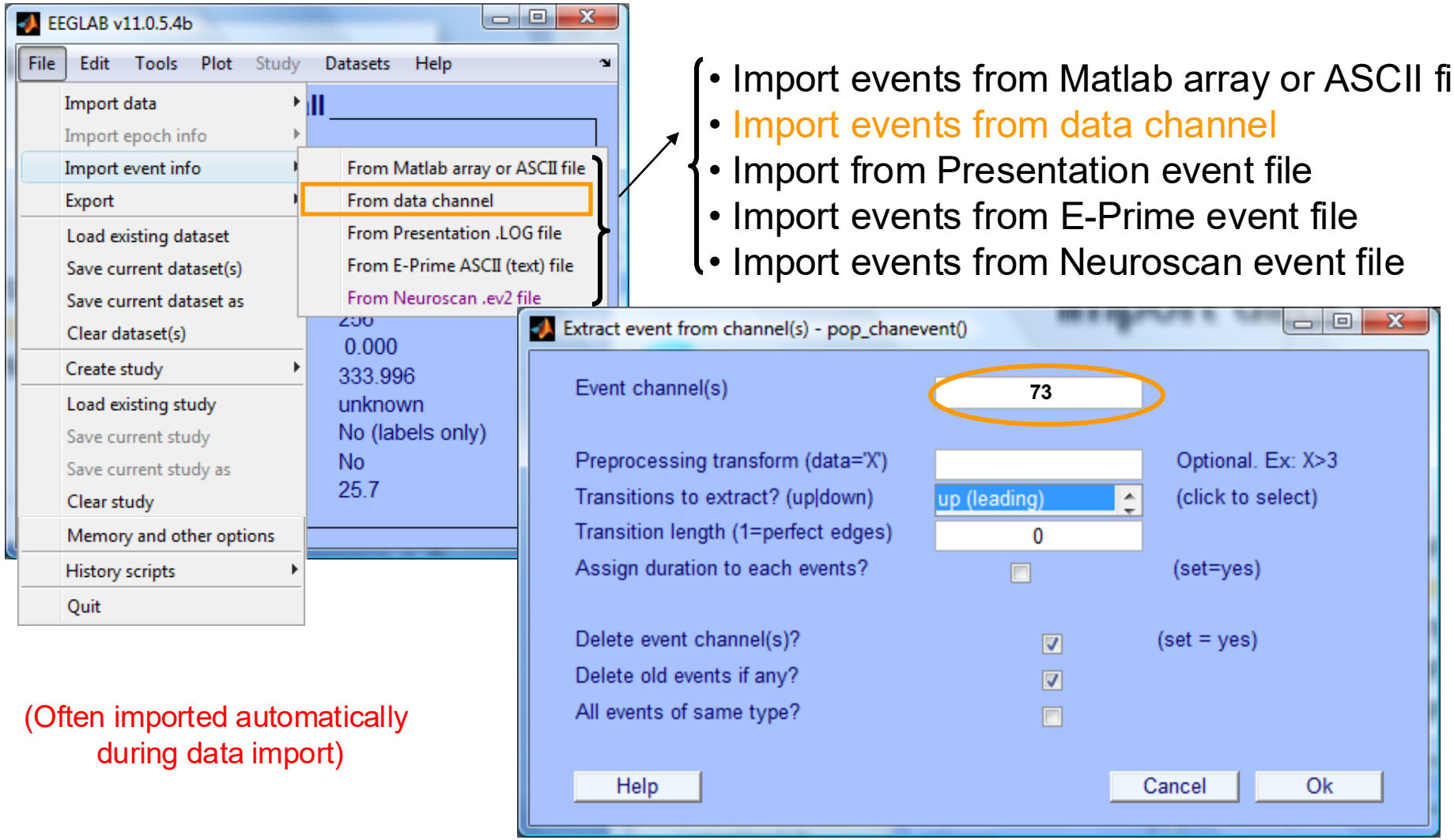


- All tools can **import** BIDS single subject data (FieldTrip and EEGLAB support BIDS sidecar event files)
- All tools can **export** BIDS data (GUI available in Brainstorm and EEGLAB)
- Brainstorm and EEGLAB have dedicated BIDS tools for group analysis

Pre-processing pipeline



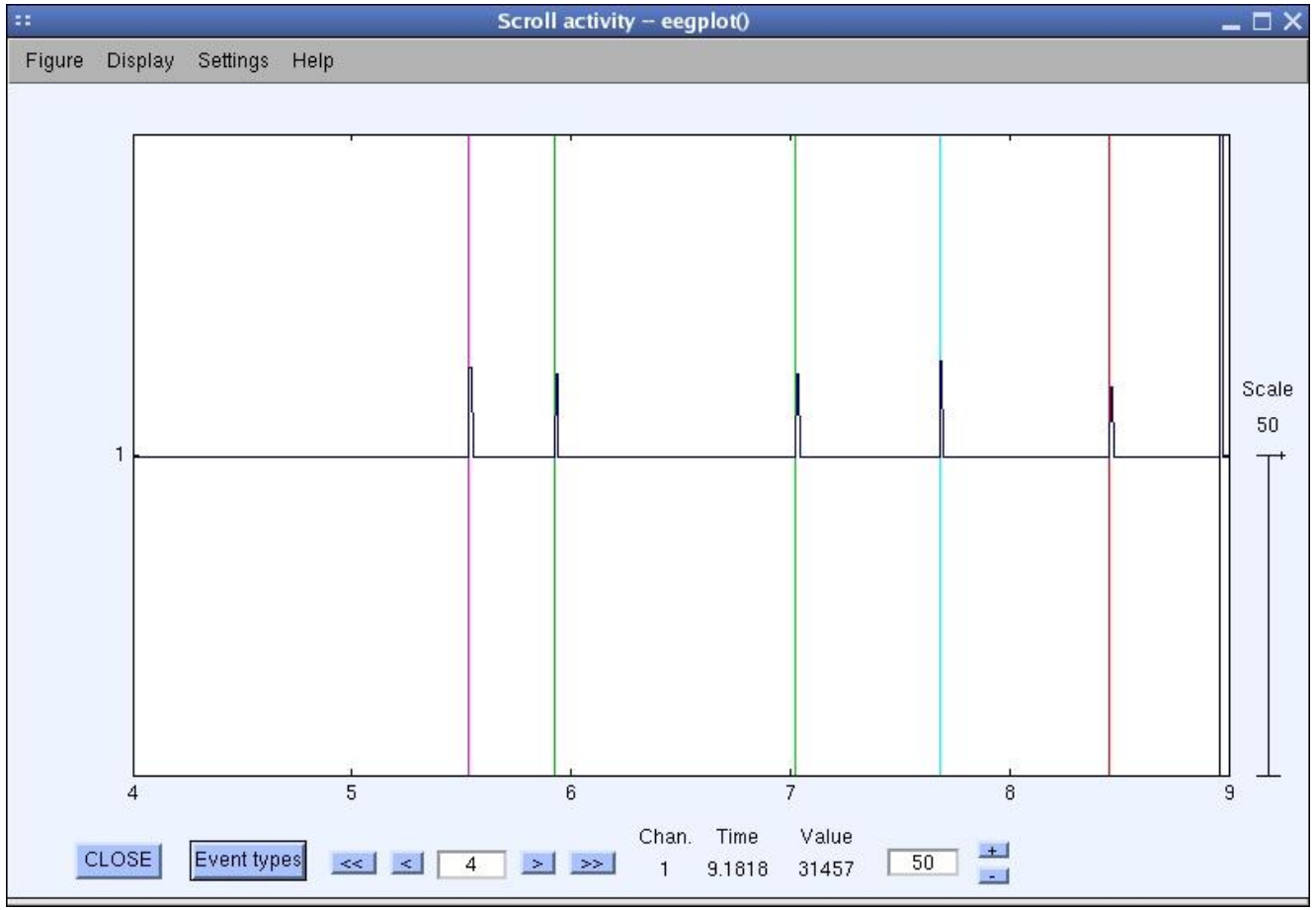
Import data events



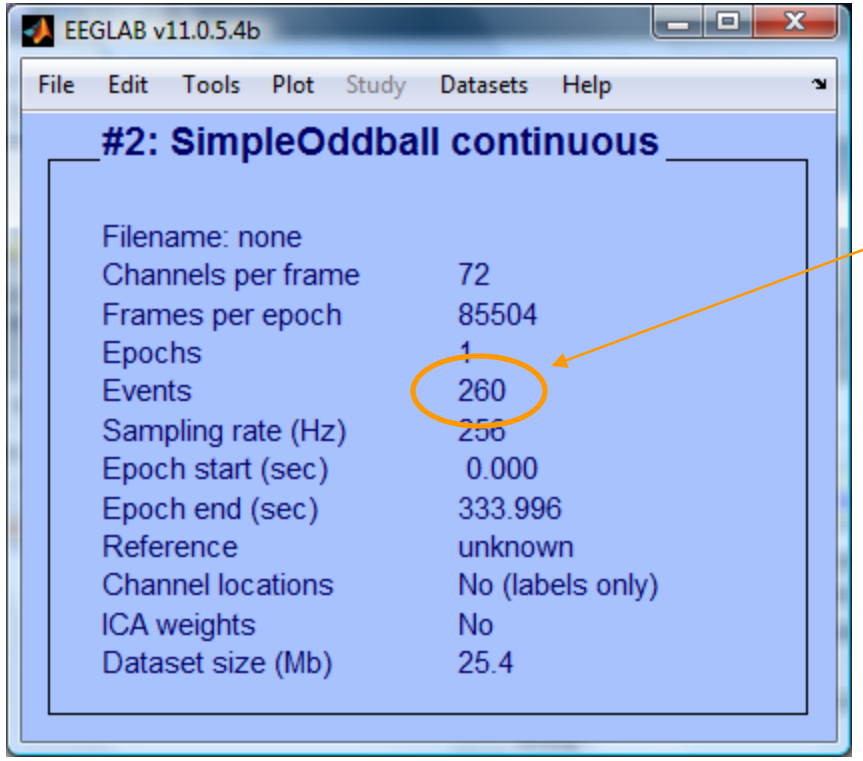
- Import events from Matlab array or ASCII file
- Import events from data channel
- Import from Presentation event file
- Import events from E-Prime event file
- Import events from Neuroscan event file

(Often imported automatically during data import)

Appearance of an event channel in raw data



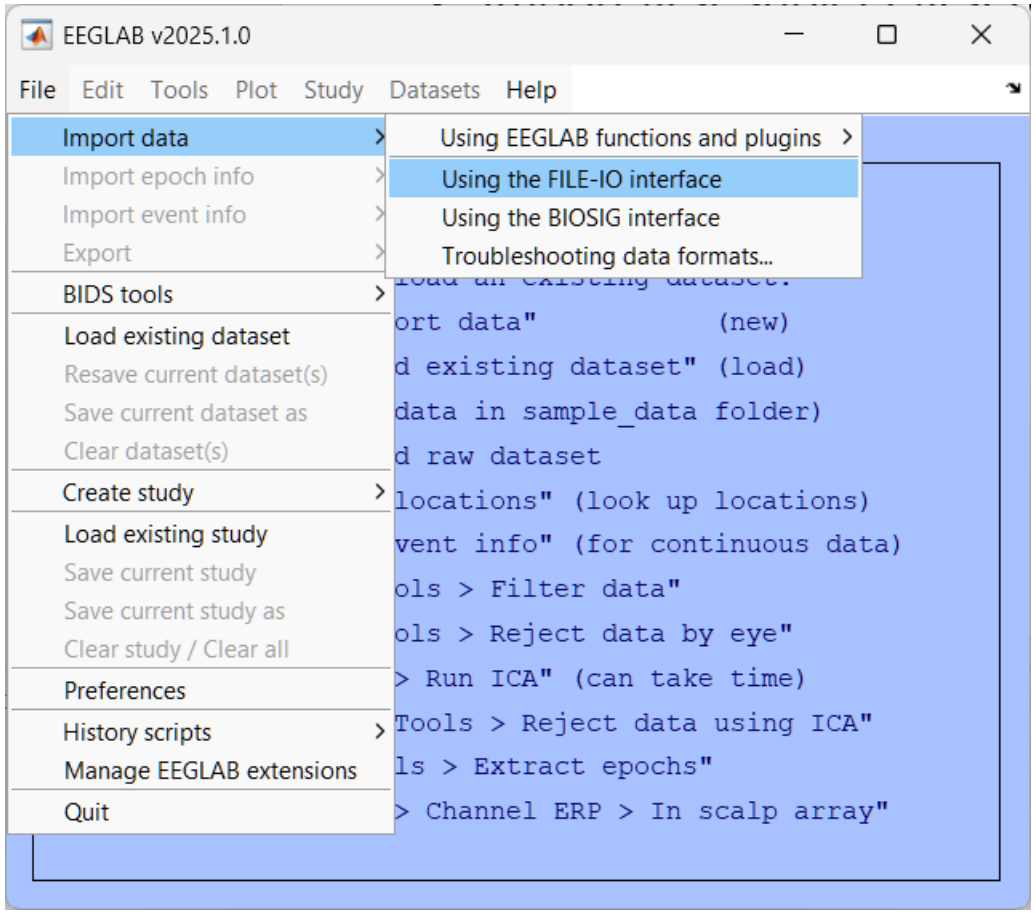
Imported data events



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

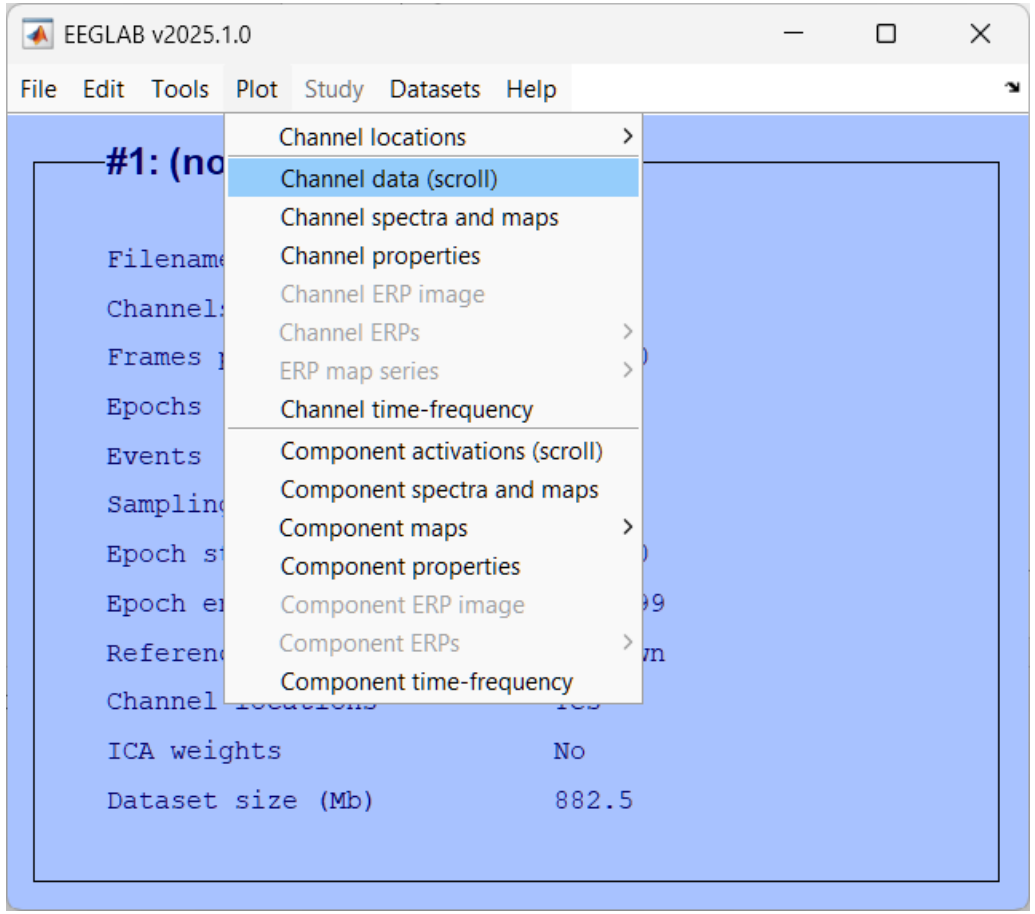
Import data

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

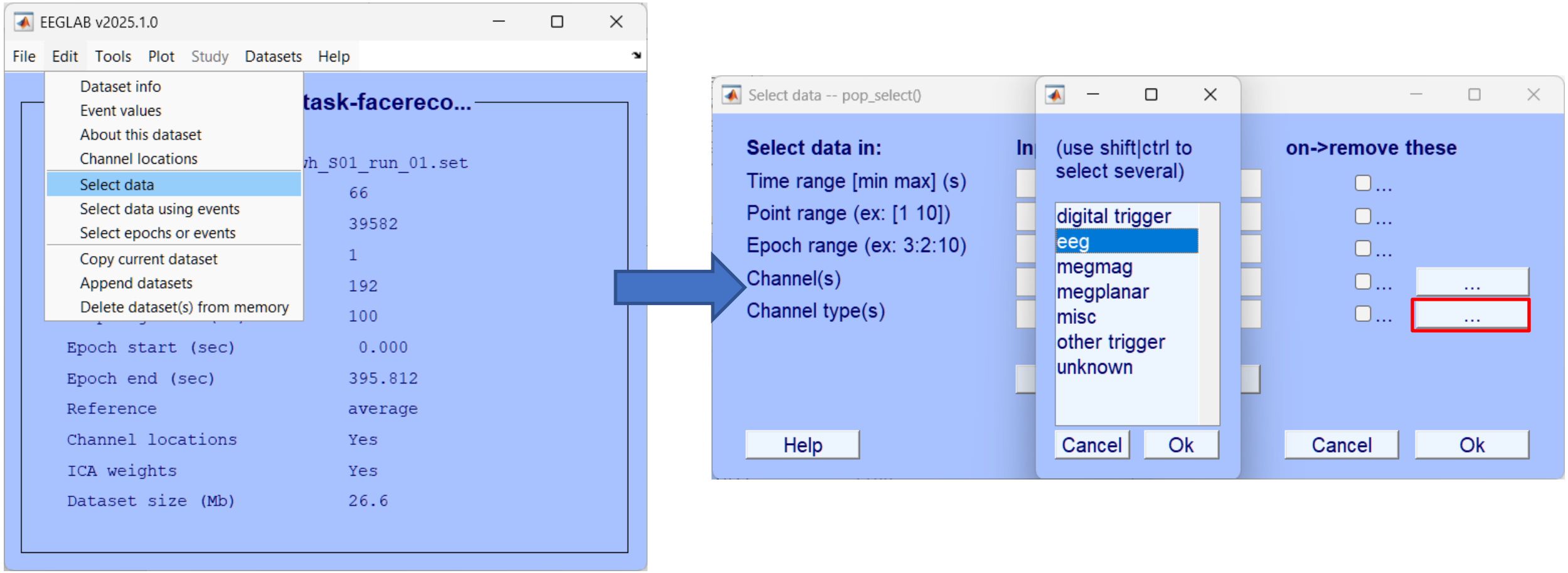


***Path** : ds000117_pruned\derivatives\meg_derivatives\sub-01\ses-meg\meg*
***File** : sub-01_ses-meg_task-facerecognition_run-01_proc-sss_meg.fif*

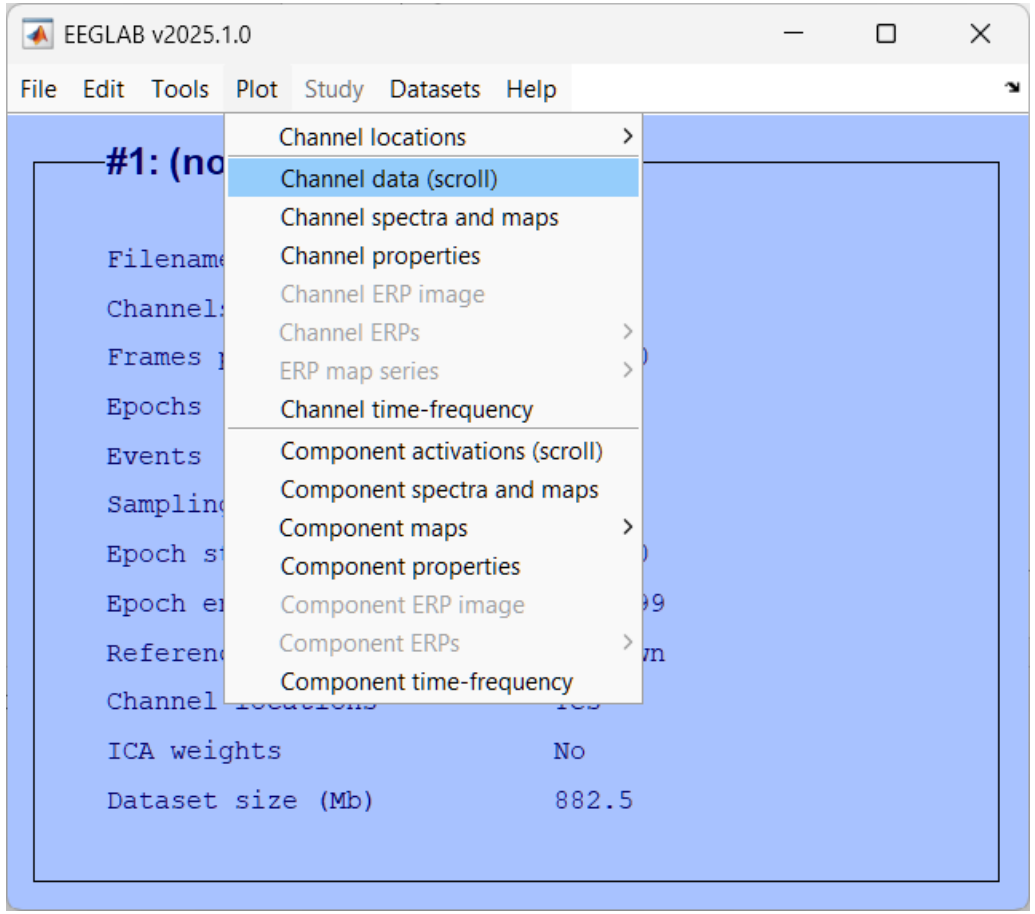
Plot channels timecourse



Select EEG channels



Check again signal timecourse



Hands on: Basic data loading and plotting

- Start EEGLAB
- Import FIF file using menu item **File > Import data > Using File-IO**
  ds000117_pruned/derivatives/meg_derivatives/sub-01/ses-meg/meg/sub-01_ses-meg_task-facerecognition_run-06_proc-sss_meg.fif
- Use menu item **Edit > Select data** to select all the channels of type EEG
- Use menu item **Plot > Scroll** data to scroll the data. Use menu **Edit > Channel locations** to inspect channel locations
- **Note:** Because the data was collected in an MEG machine, it is complex and is preprocessed by the *script script_01_import_data.m* including renaming and realigning events. In future lectures, we will use the result of that script. If it has not already been executed for you on that folder, you will need to execute it.

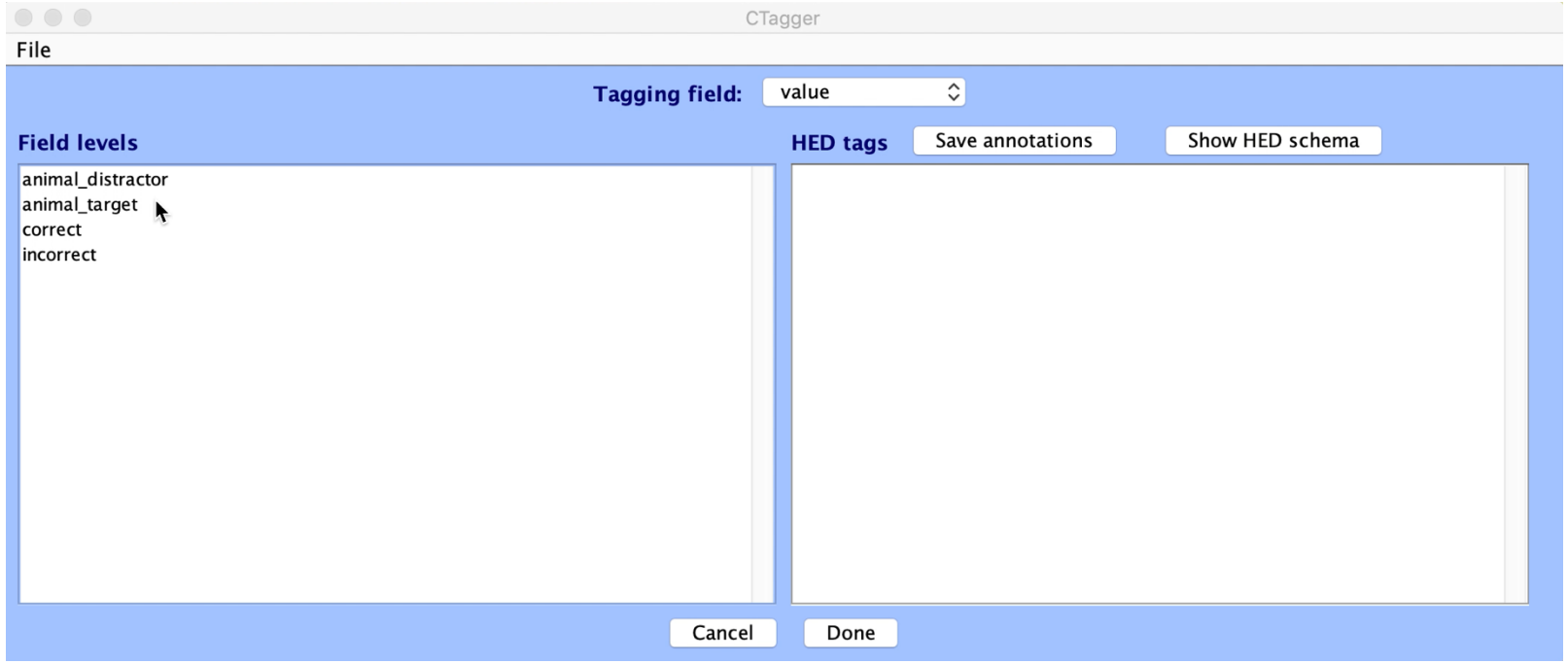
Contents lists available at ScienceDirect

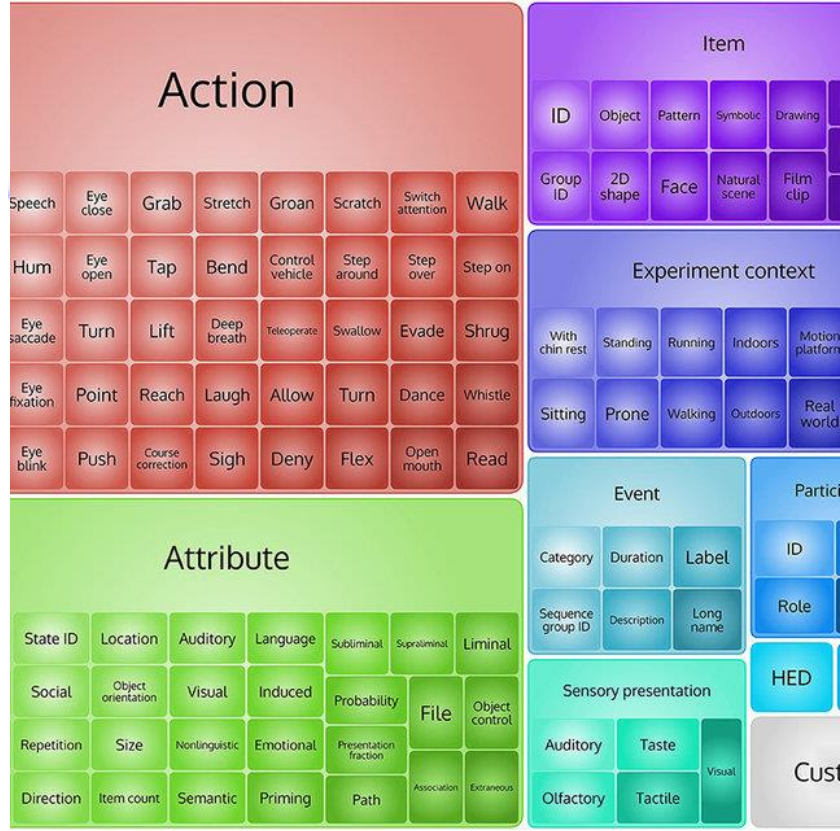
NeuroImage

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

Capturing the nature of events and event context using hierarchical event descriptors (HED)

Kay Robbins[a,*], Dung Truong[b], Stefan Appelhoff[c], Arnaud Delorme[b,d], Scott Makeig[b]



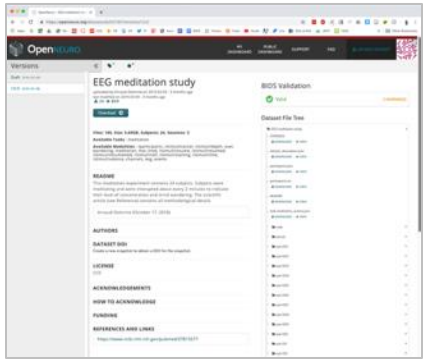


HED is integrated in

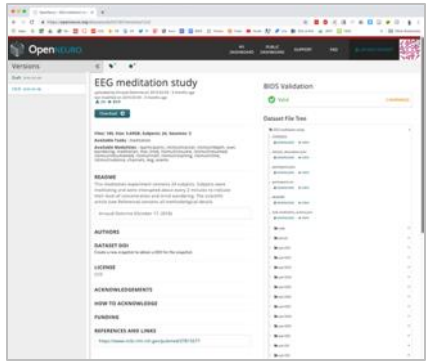
BIDS

BRAIN IMAGING DATA STRUCTURE

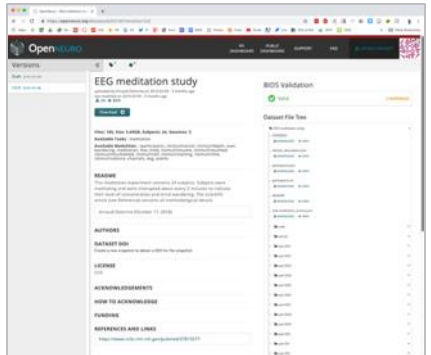
NEMAR BIDS experiment 1



NEMAR BIDS experiment 2



NEMAR BIDS experiment 3



HED event selection

Automated pipelines

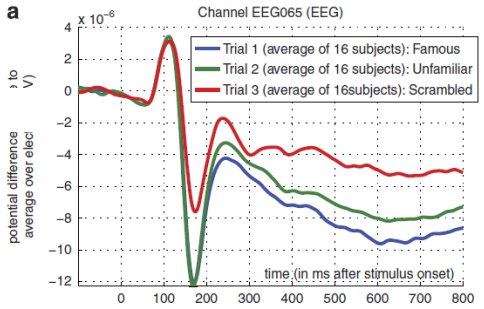
EEGLAB

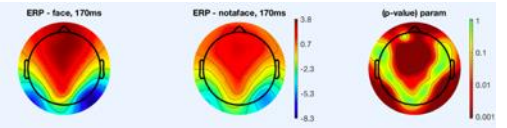
Brainstorm

FieldTrip

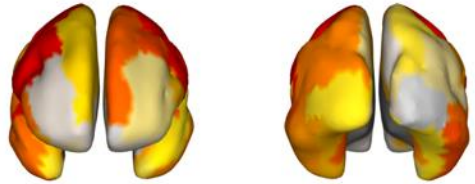
MNE
MEG + EEG ANALYSIS & VISUALIZATION

ERPs and ERSPs

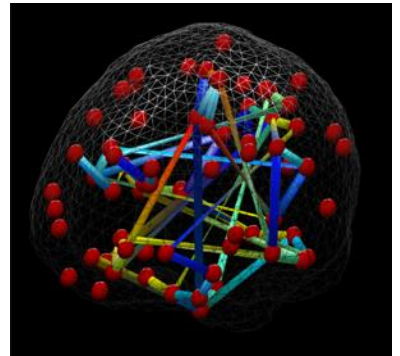




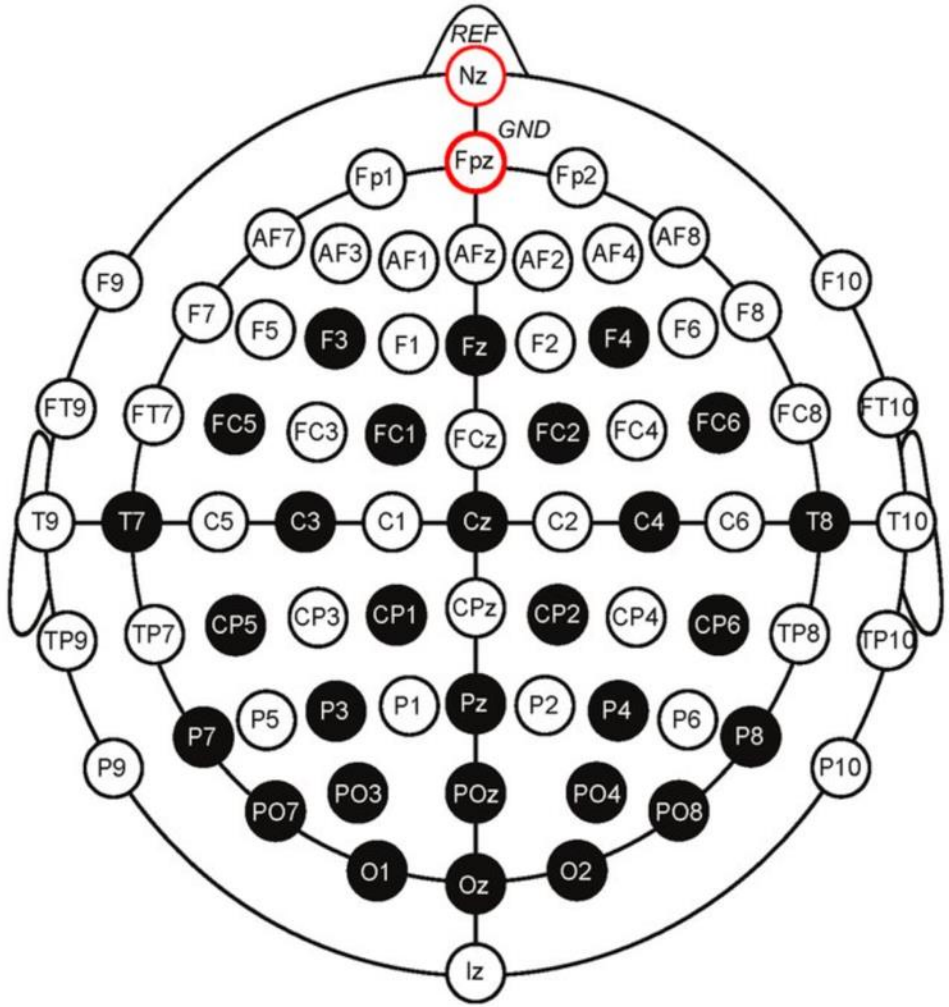
Source localization

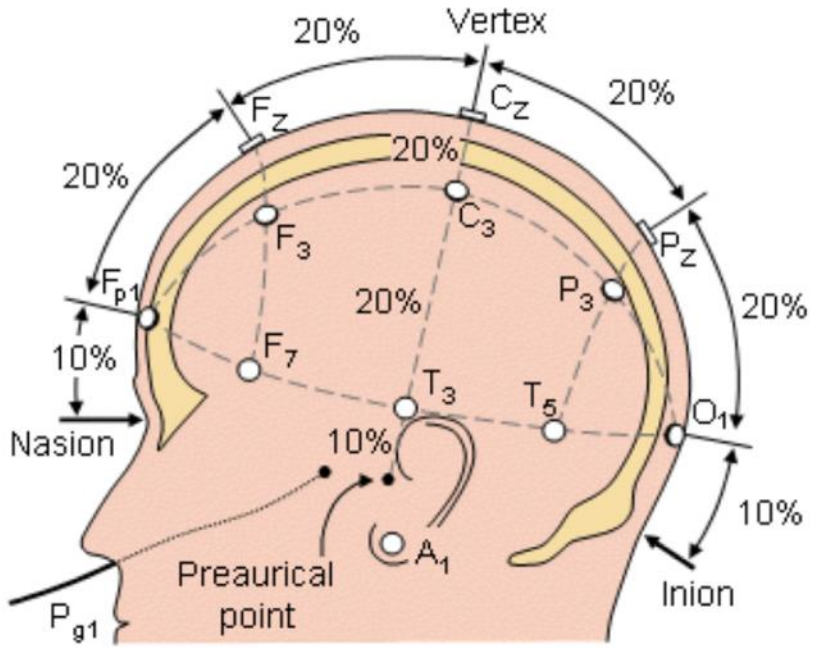


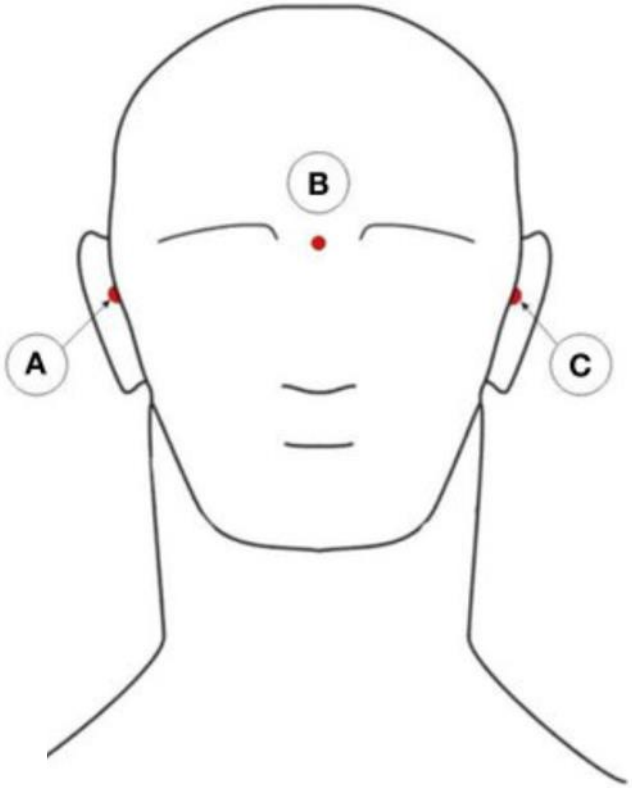
Connectivity



Channel labels & locations







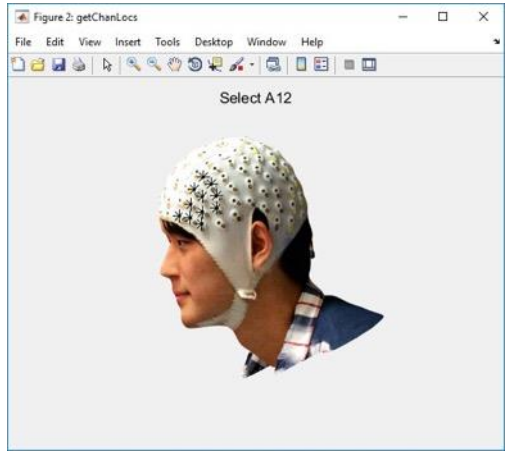
Scanning electrode position will become standard

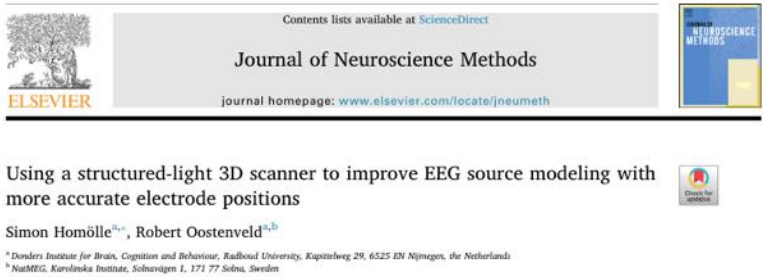
Get_chanlocs EEGLAB plugin interfacing Fieldtrip's functions

3-D camera mounted on iPad

Smartphone 3-D scanners

EEGLAB *get_chanlocs* plugin





Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

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

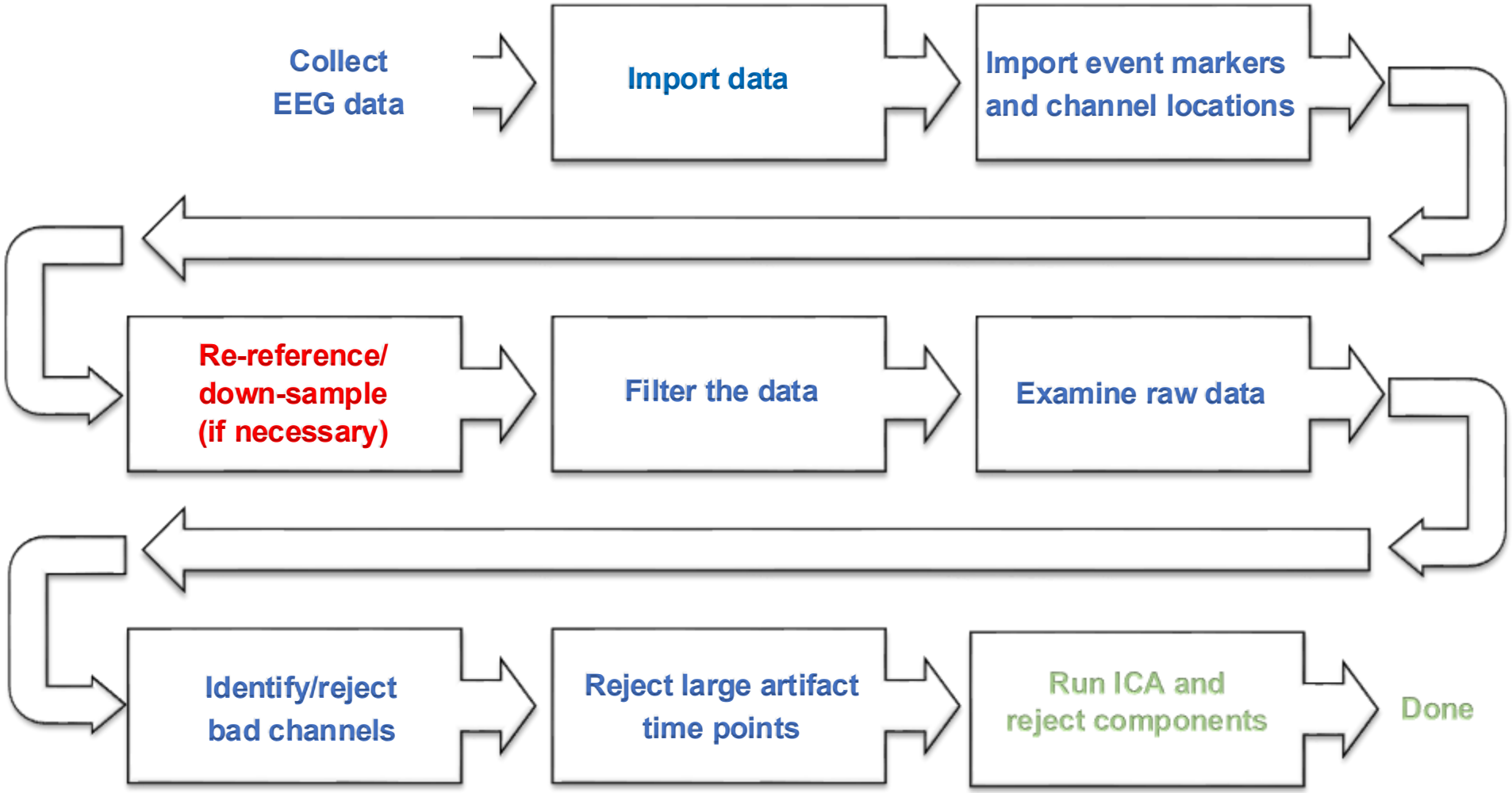
Simon Homölle[a,*], Robert Oostenveld[a,b]

[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)

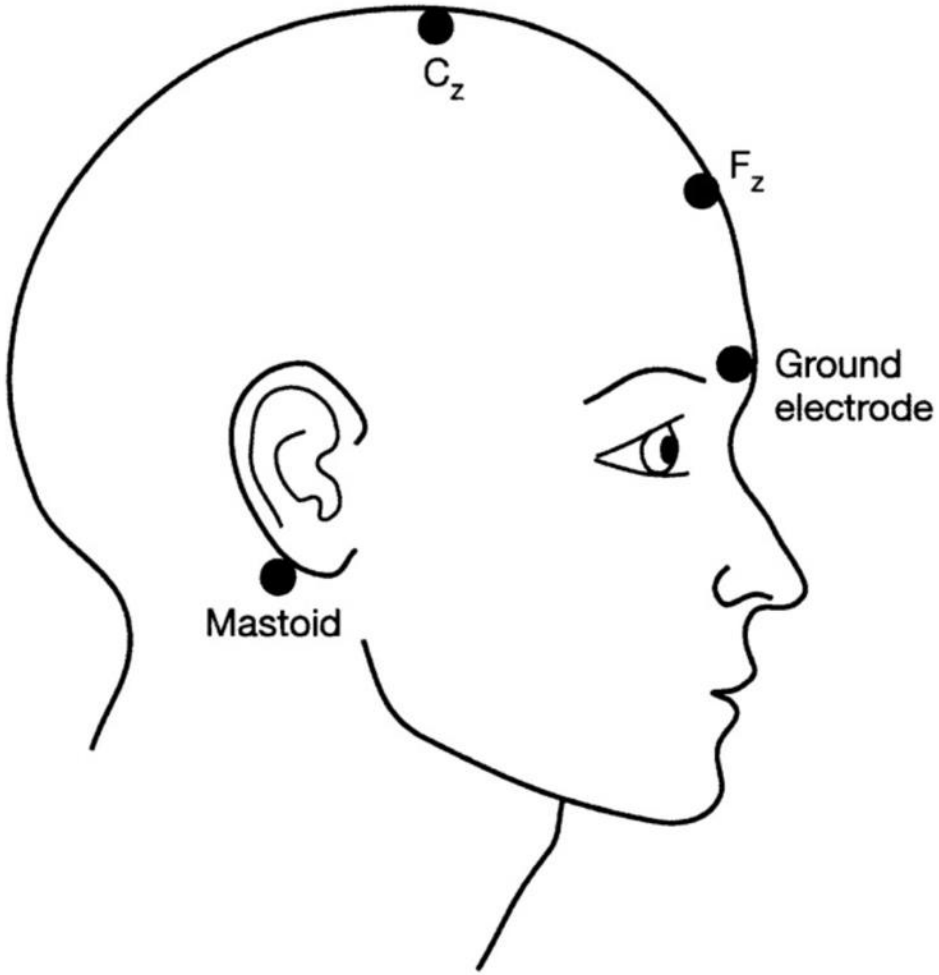
Pre-processing pipeline

01.

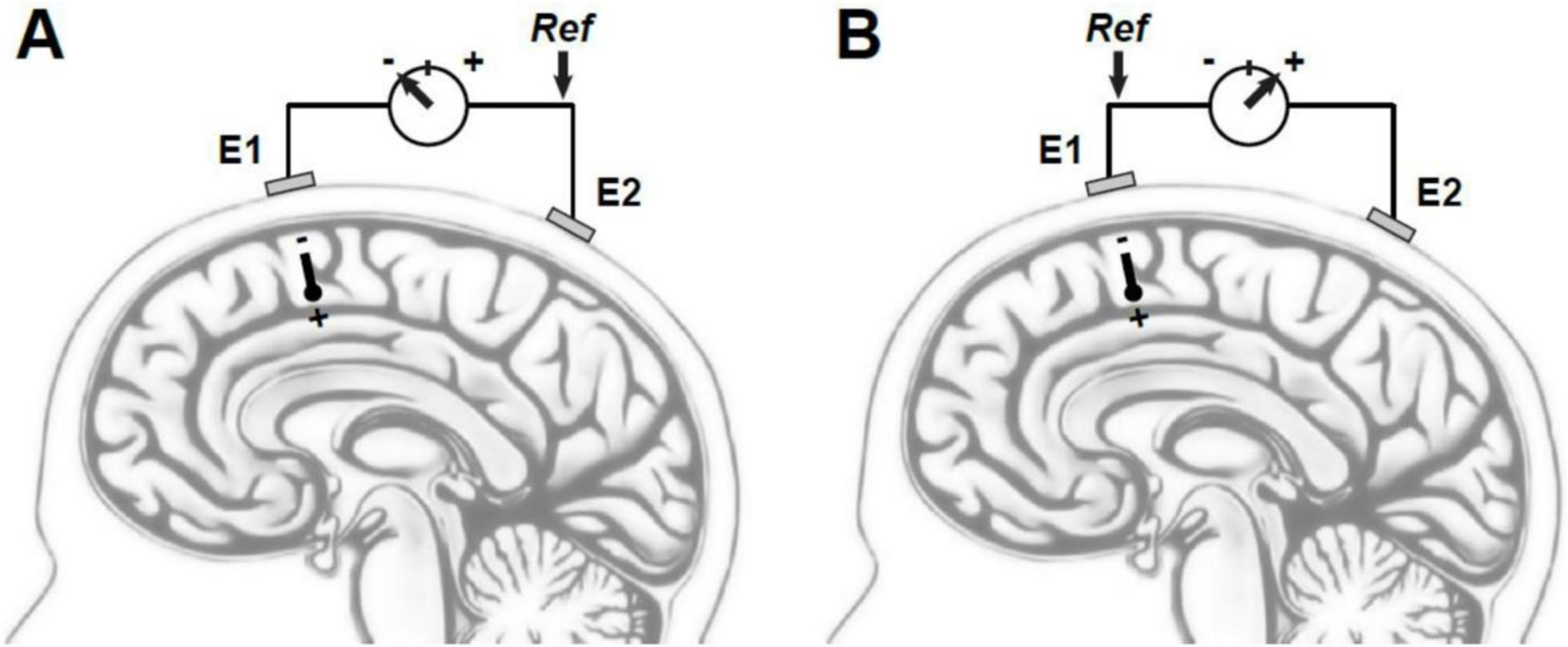


Referencing

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

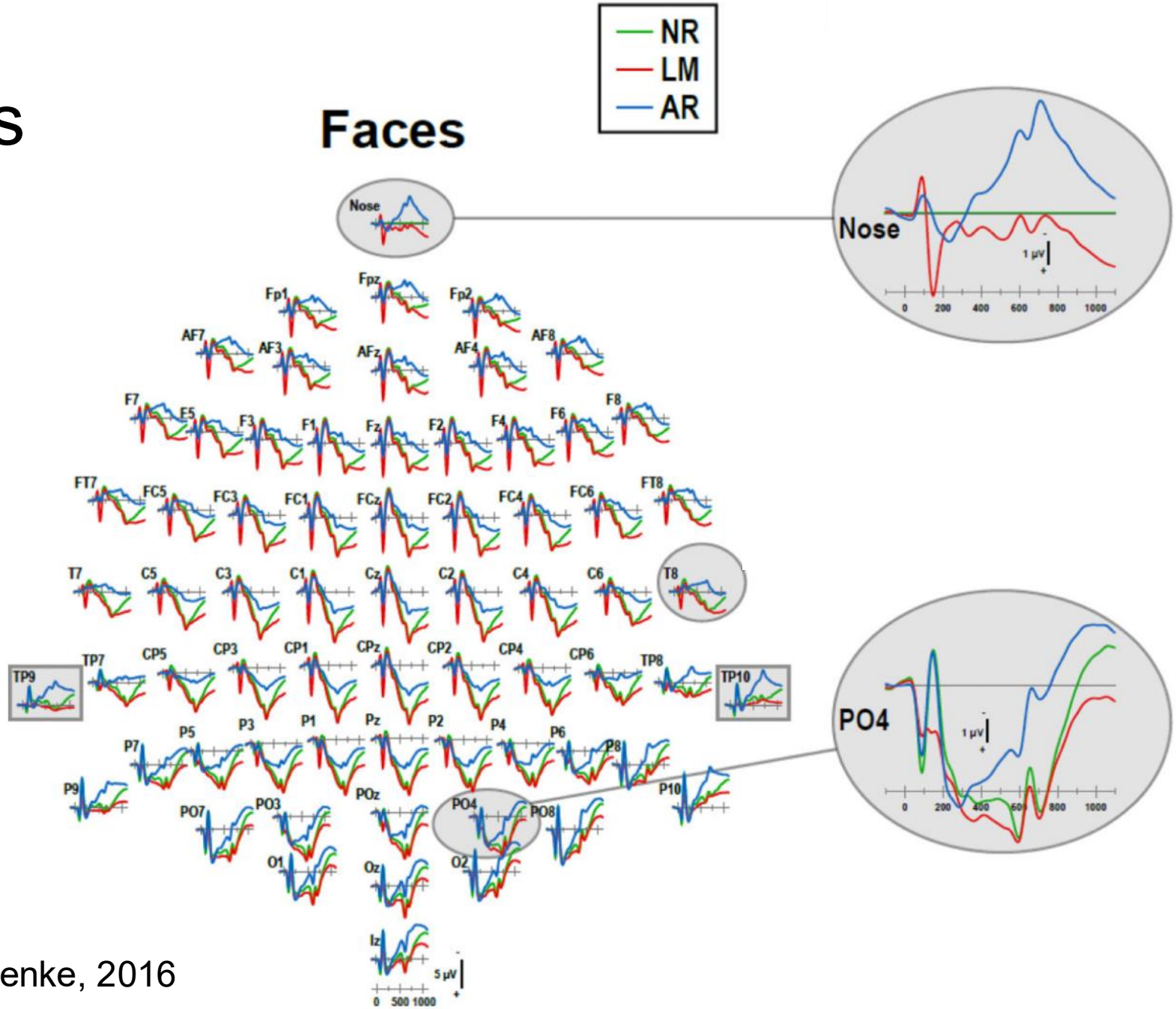


The location of the Reference matters



Kayser & Tenke, 2016

The location of Reference matters



Kayser & Tenke, 2016

scientific reports

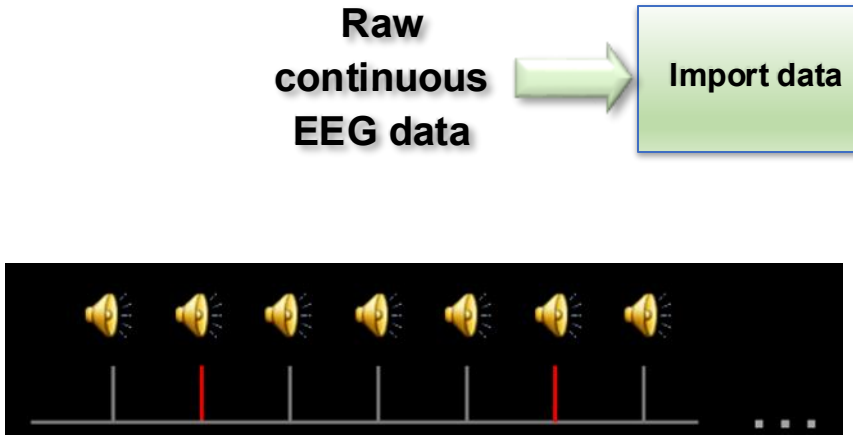
Article | Open access | Published: 09 February 2023

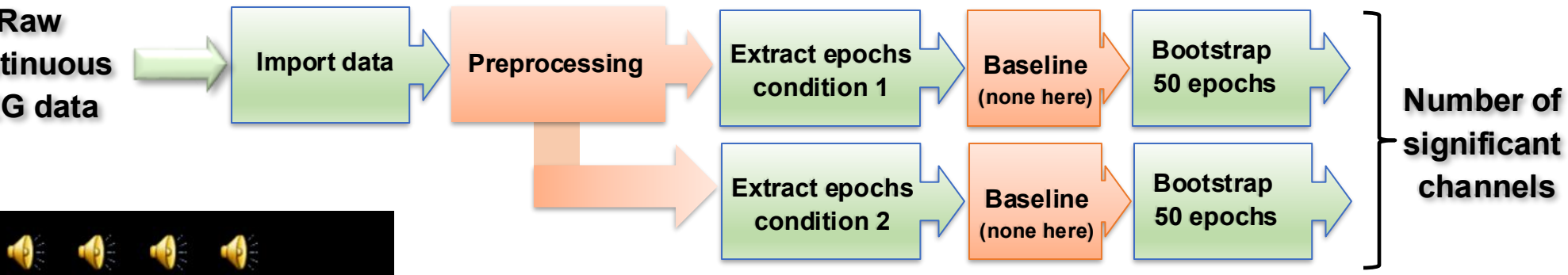
EEG is better left alone

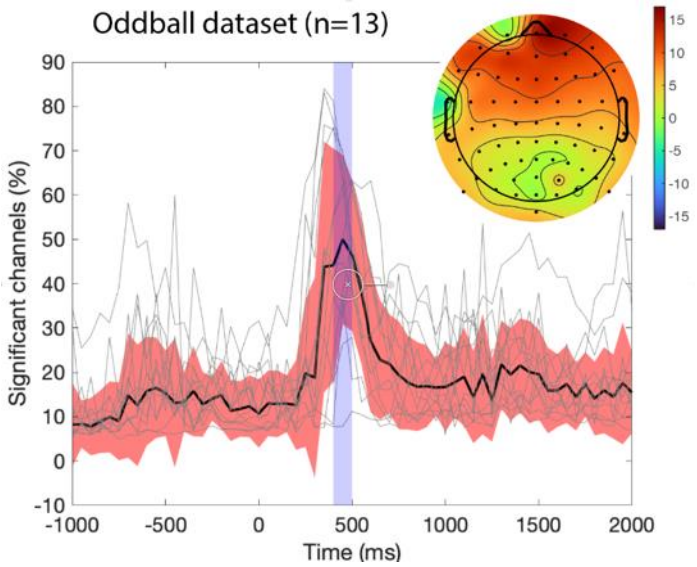
Arnaud Delorme

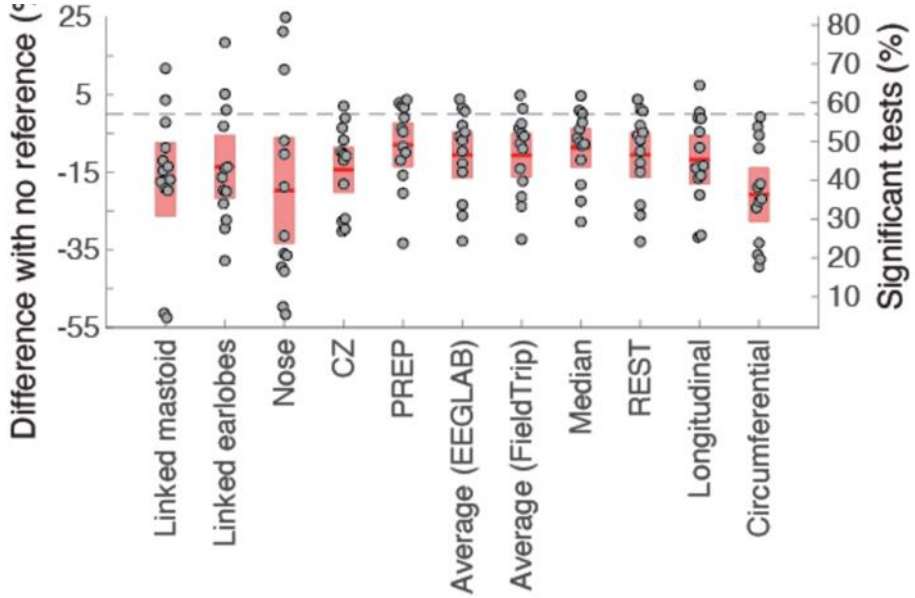
Scientific Reports **13**, Article number: 2372 (2023) | Cite this article

Influence of the reference on ERPs

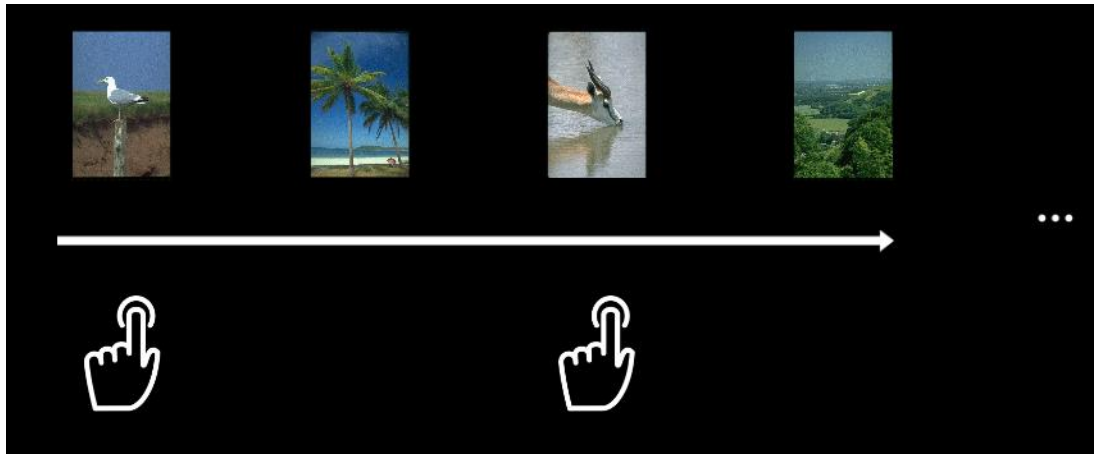


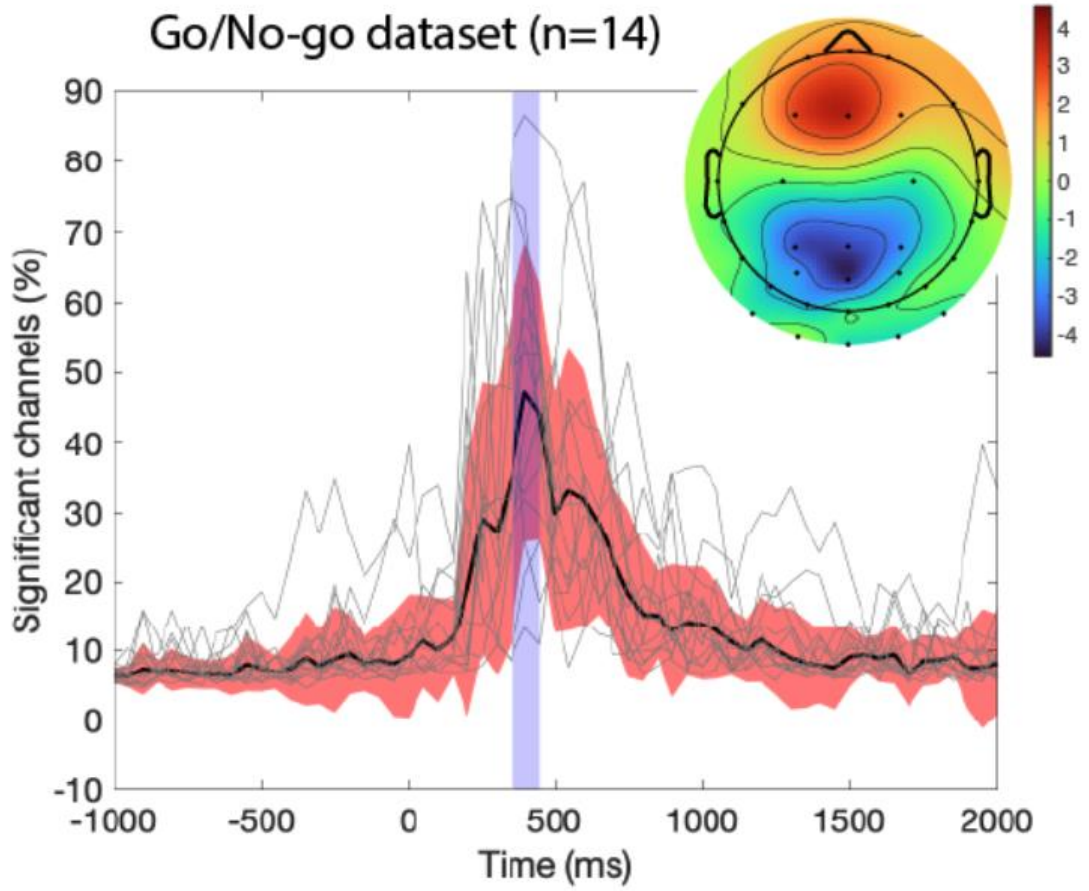


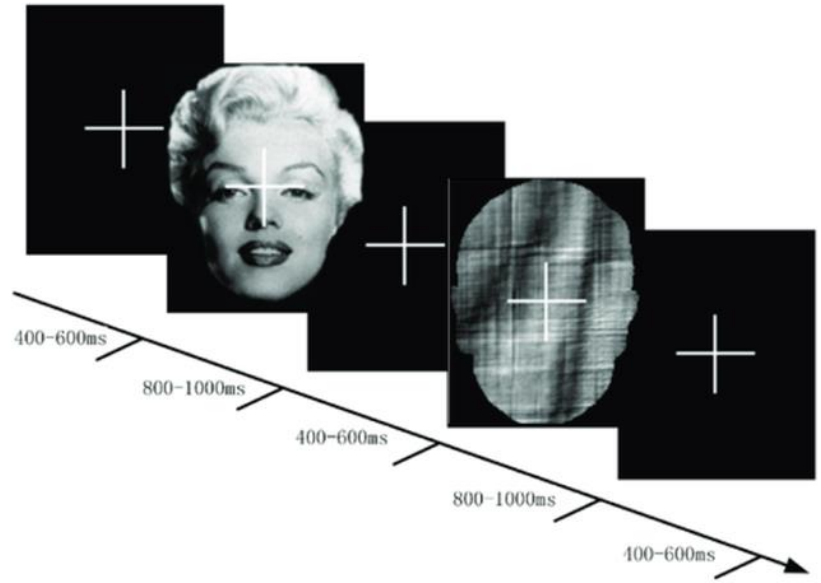


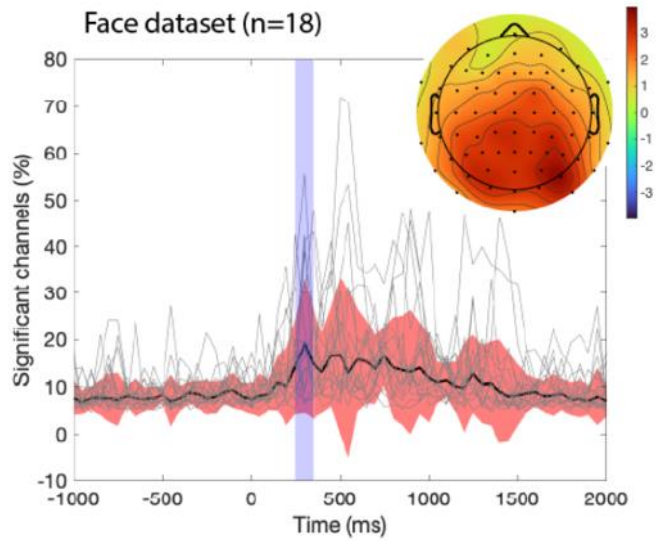


Influence of the reference on ERPs



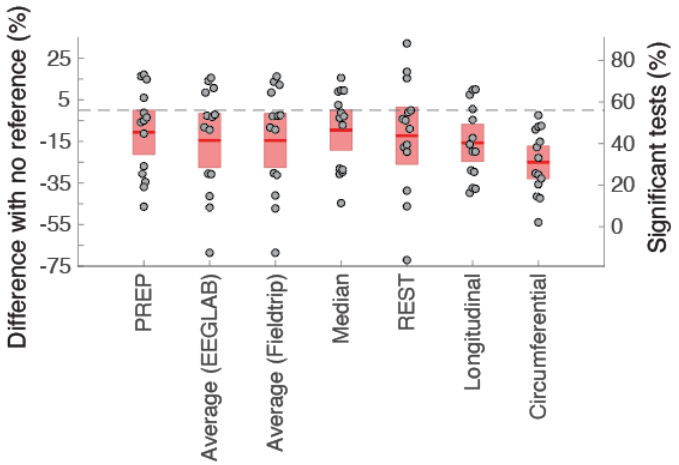



400-600ms
800-1000ms
400-600ms
800-1000ms
400-600ms

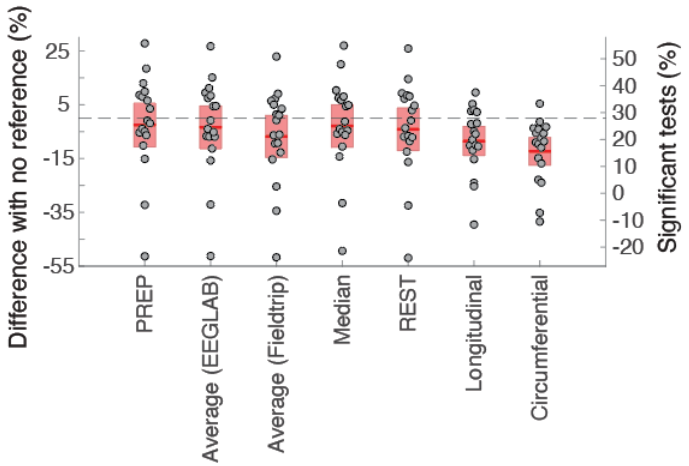

Face dataset (n=18)
Significant channels (%)
Time (ms)
80
70
60
50
40
30
20
10
0
-10
-1000
-500
0
500
1000
1500
2000
3
2
1
0
-1
-2
-3

Influence of the reference on ERPs

Animal vs non-animal task
(BIDS 2680)



Face scrambled vs. famous
(dataset used at this workshop)

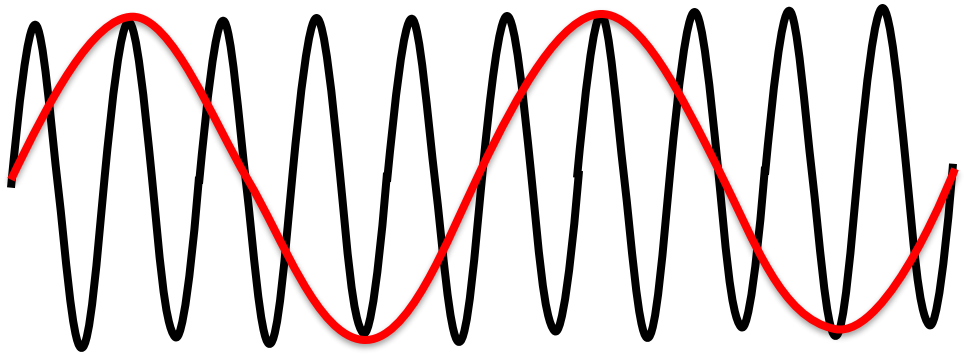


- It is unknown why using a reference decreases the number of significant channels when comparing 2 conditions for these 3 datasets
- No reference wins (REST not superior)

Downsampling data → 500 Hz is usually enough

Reasons for lowering sampling rate

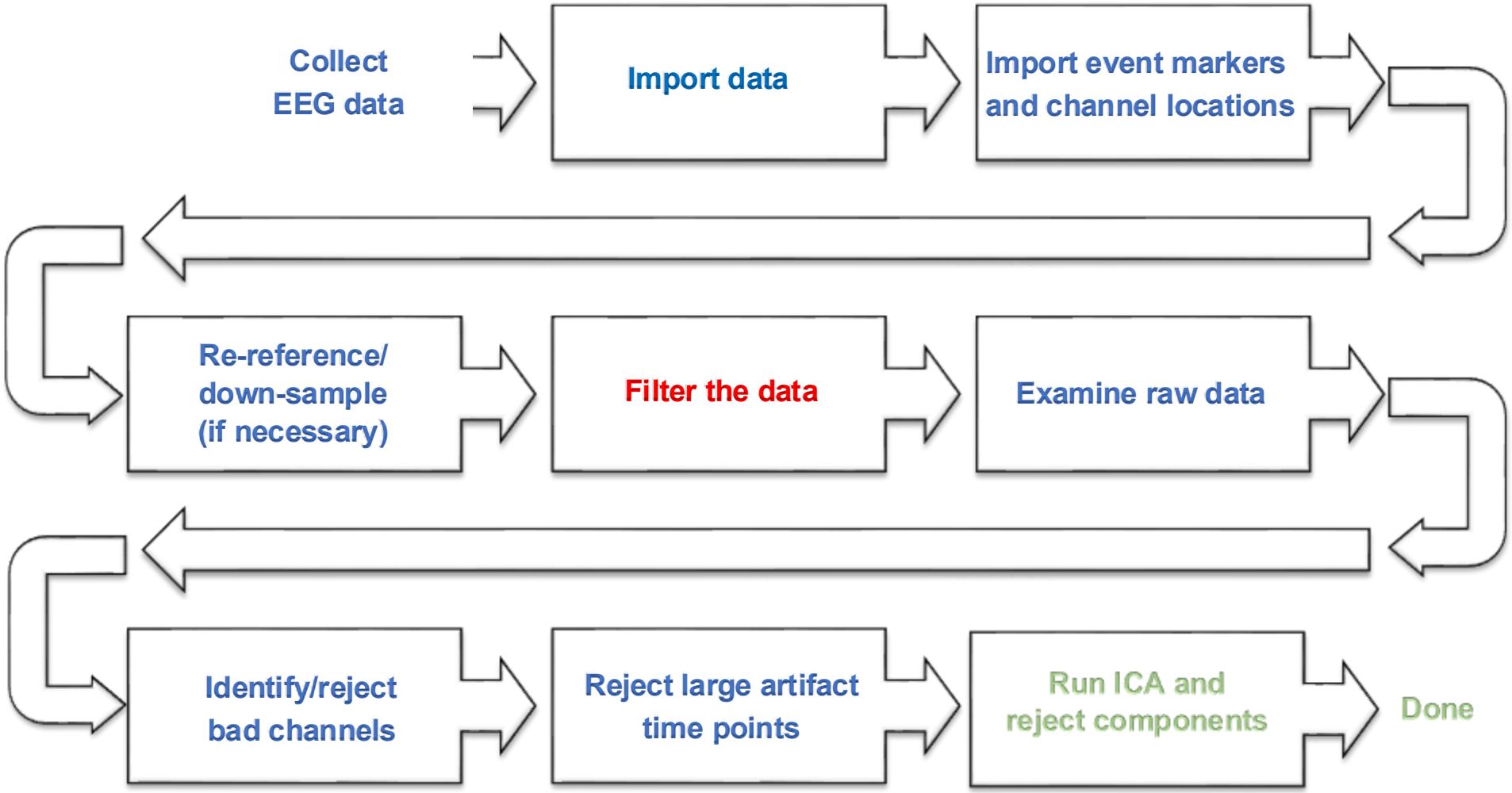
- Reduce time and computational cost
- Most MEEG processes are below 100 Hz. Maybe no need to keep a sample rate at 500Hz (beware of the Nyquist frequency)



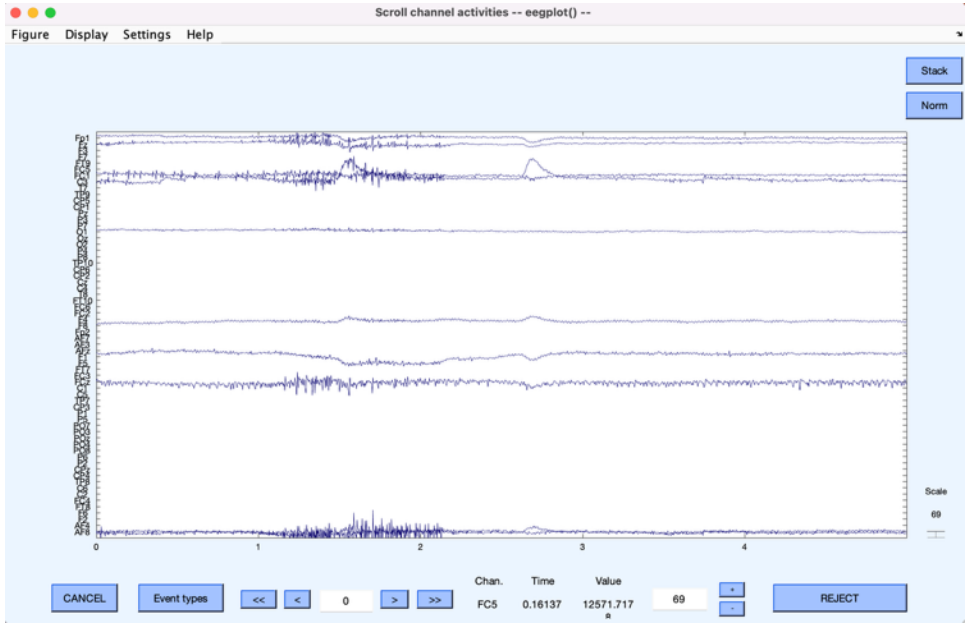
Reasons for NOT lowering sampling rate

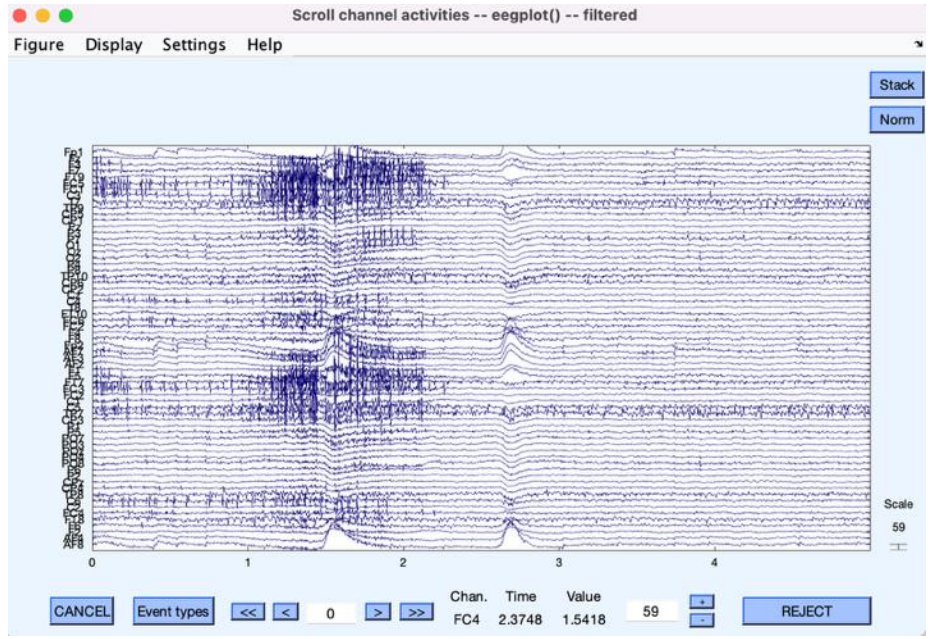
- Even with MEEG amplifiers noise, information above 250 Hz (500 Hz sampling rate) might be useful for some algorithm (e.g., Independent Component Analysis)
- Behavioral responses are measured on the order of millisecond – keep MEEG at the same time scale?

Pre-processing pipeline

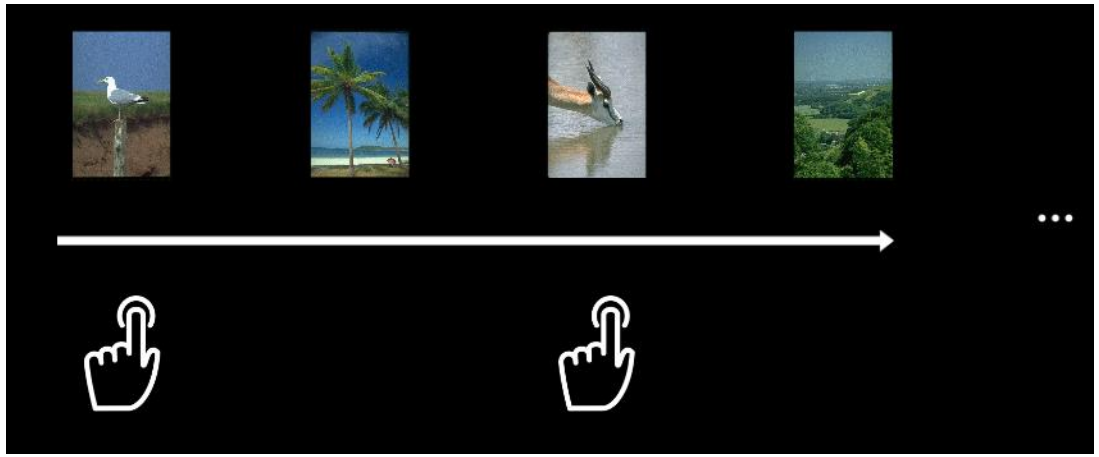


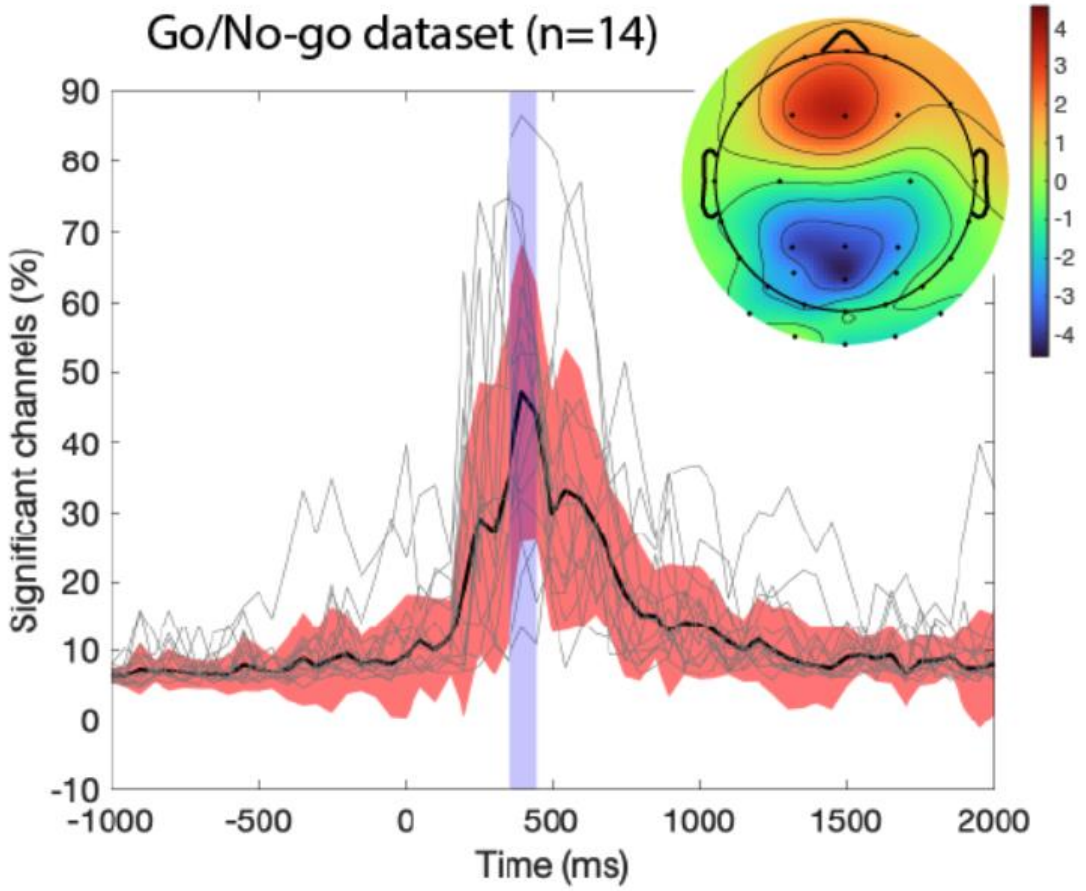
Highpass filter the data



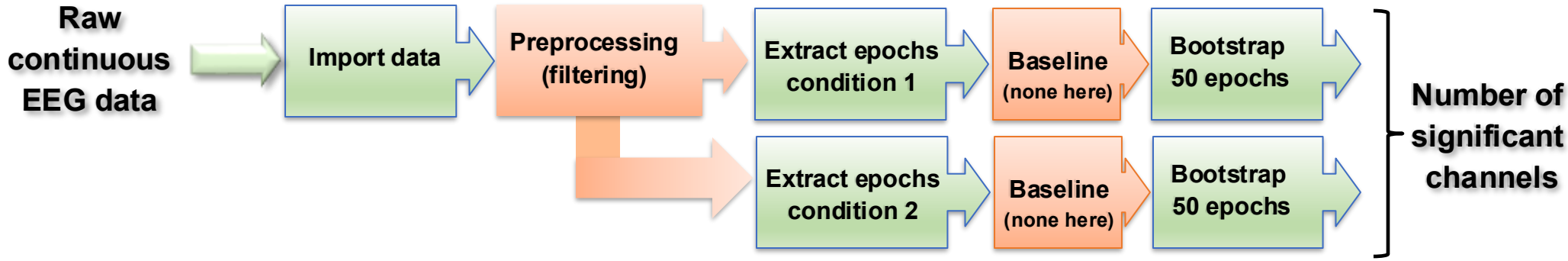


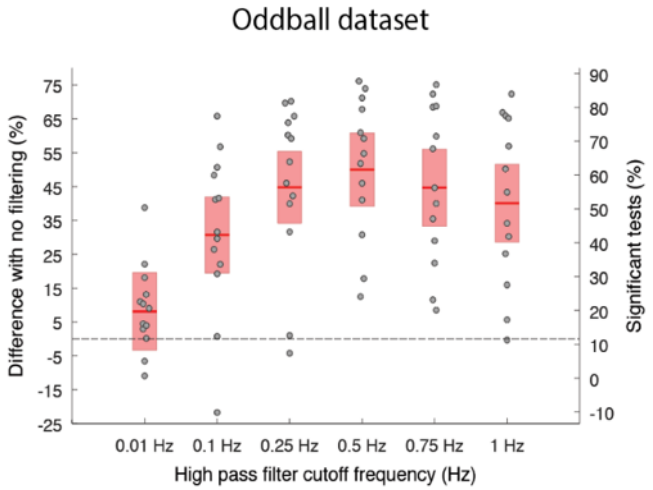
Influence of HP filtering on ERPs

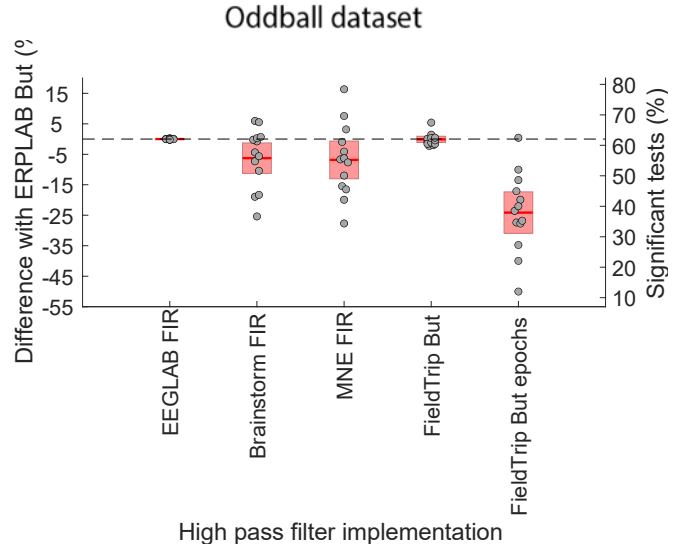




Influence of HP filtering on ERPs





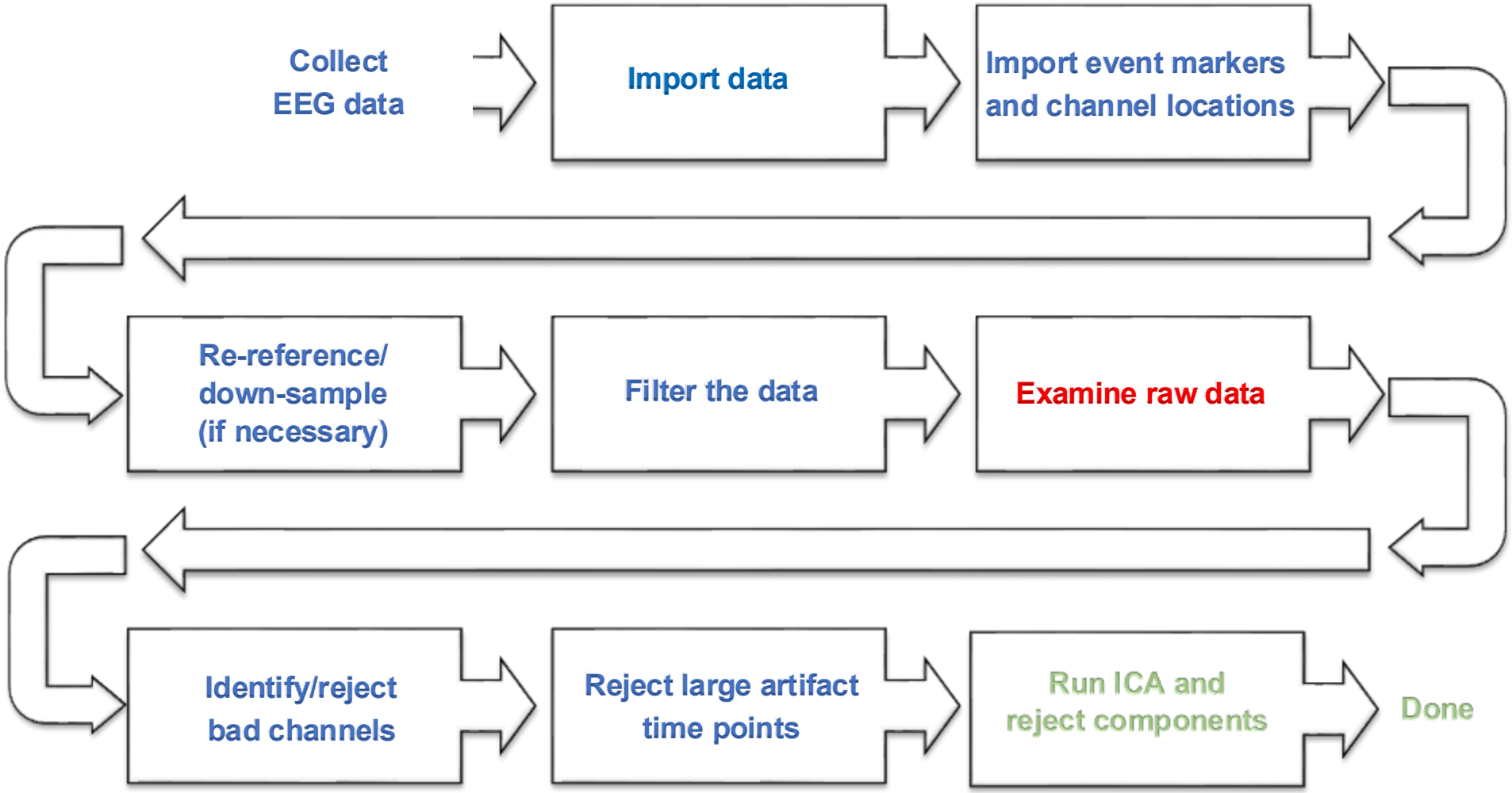


- The frequency cutoff is important
- Usually 0.5Hz is a good value

- The type of filter FIR or IIR is not critical
- Raw data should be filtered not epochs (e.g., FieldTrip preprocessing)

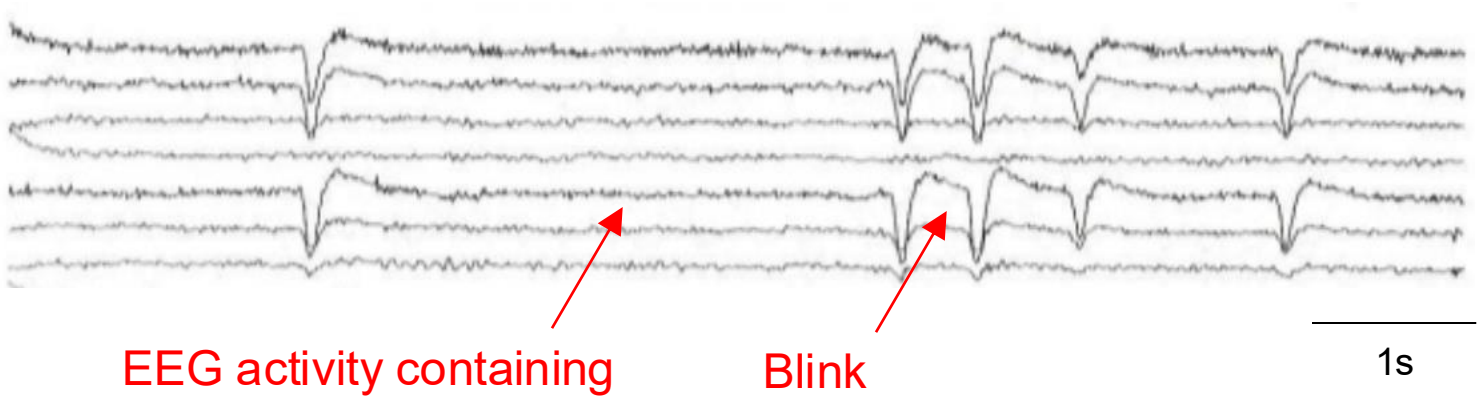
Pre-processing pipeline

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

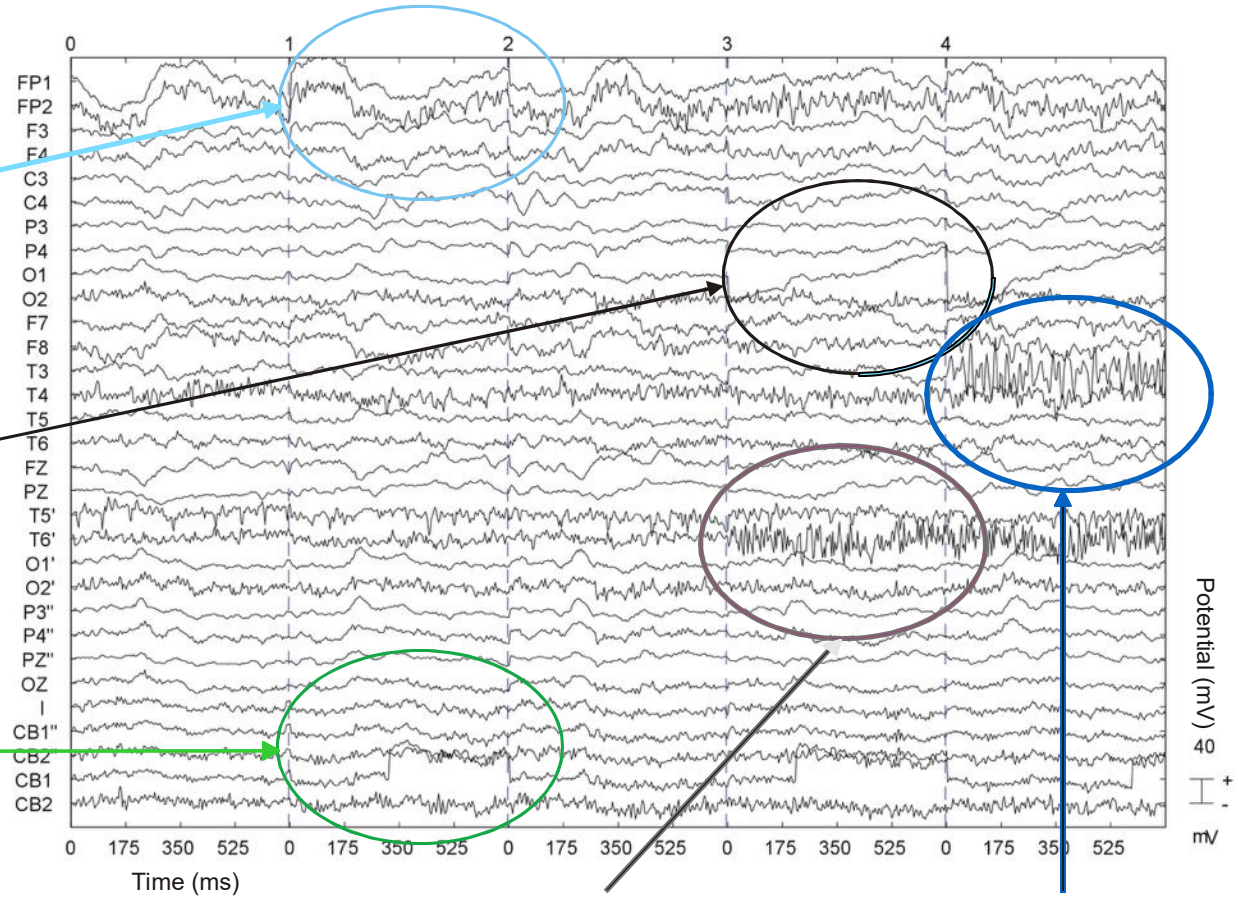


Type of artifacts

2 - Low frequency event (eye movements)

5 - Linear trend

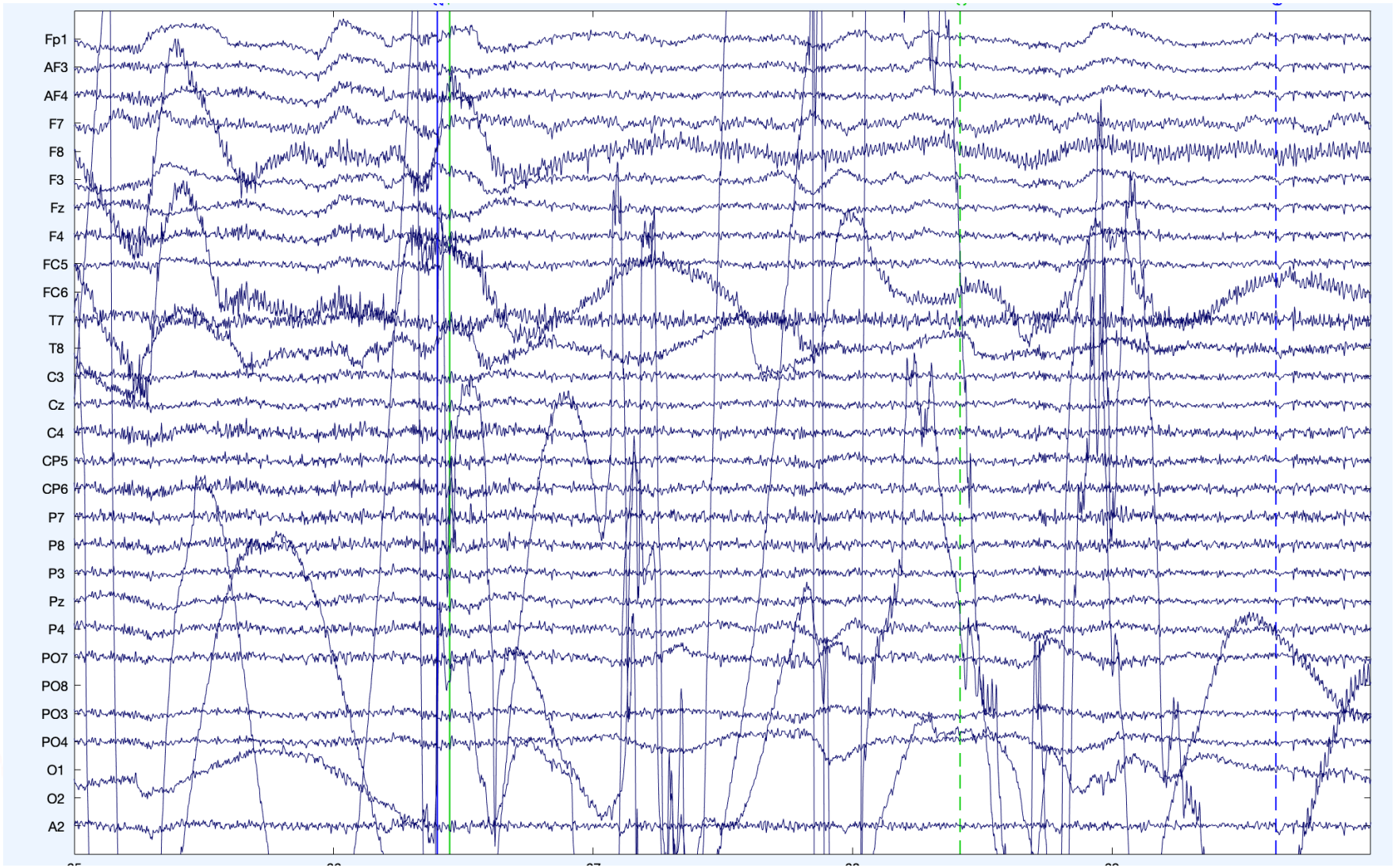
3 - Discontinuity



4 - High noise

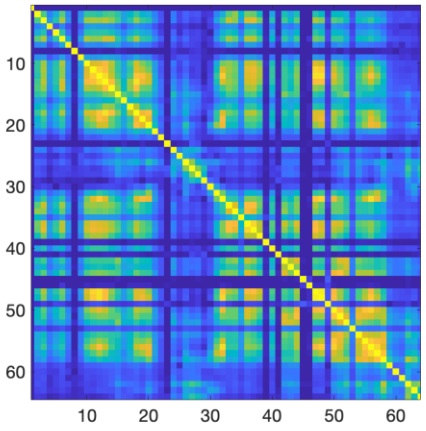
1 - Transient high frequency event (muscle)

Looking for bad channels

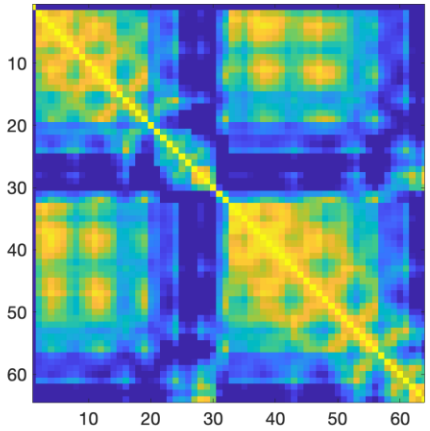


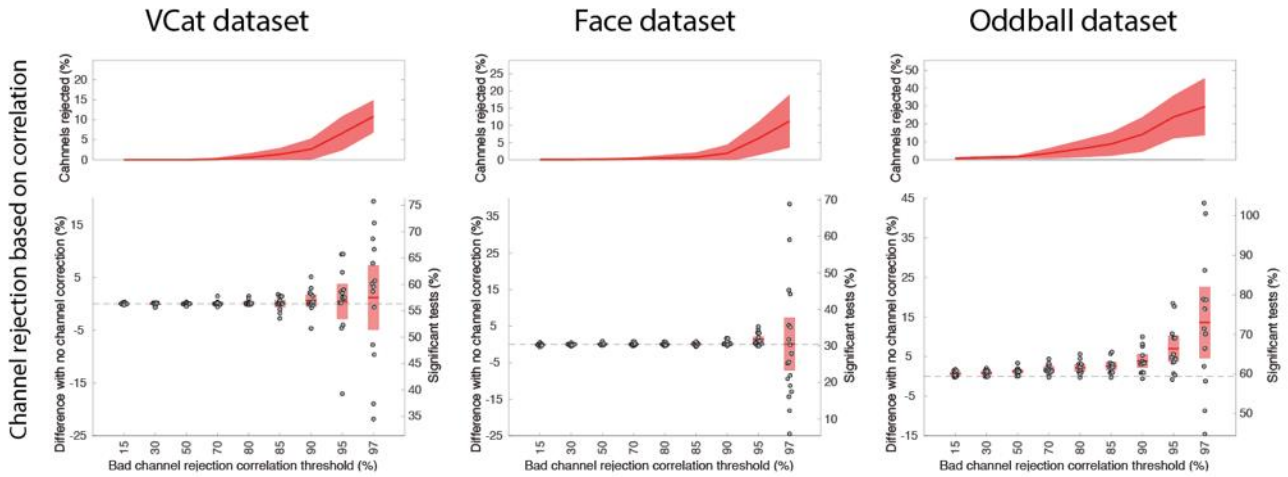
Pairwise correlation to find bad channels

Bad data
(some uncorrelated channels)



Good data
(all channels highly correlated)

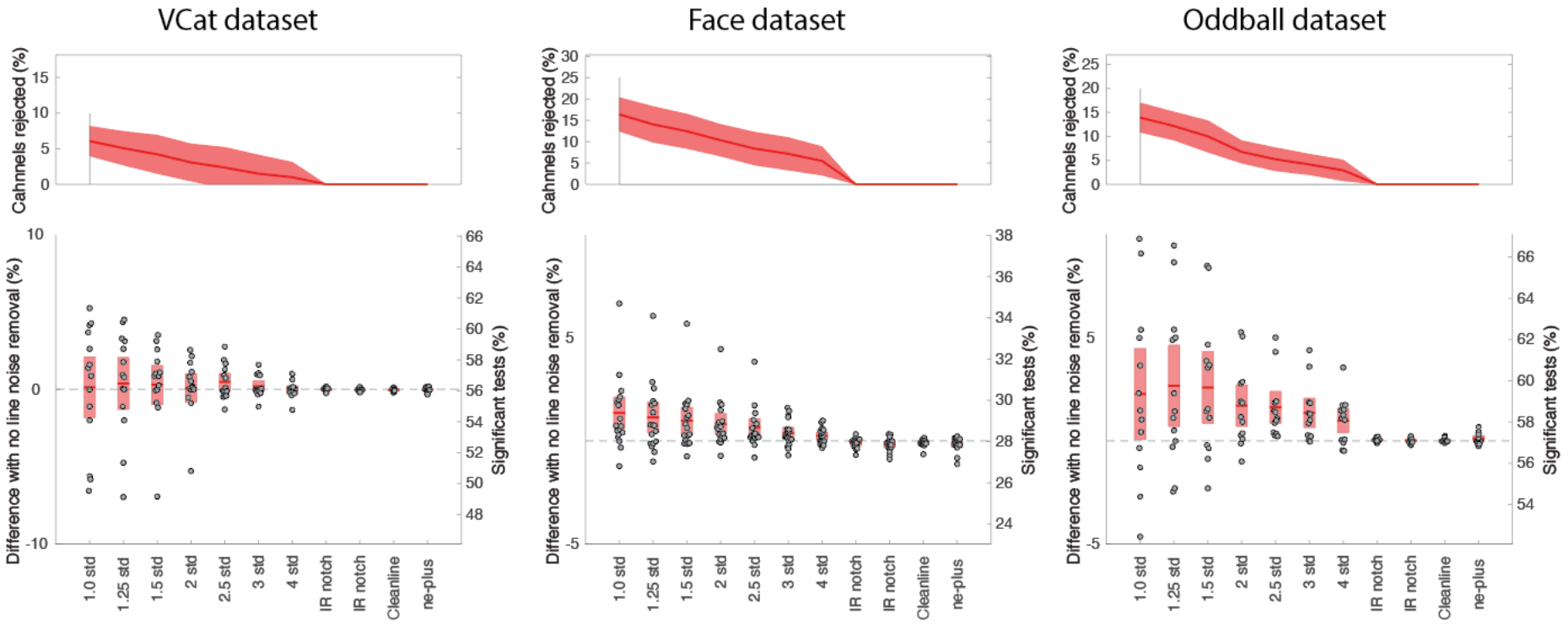




Tim R. Mullen, Christian Kothe, et al.(2015) Real-time neuroimaging and cognitive monitoring using wearable dry EEG. IEEE Transactions on Biomedical Engineering. DOI:10.1109/TBME.2015.2481482

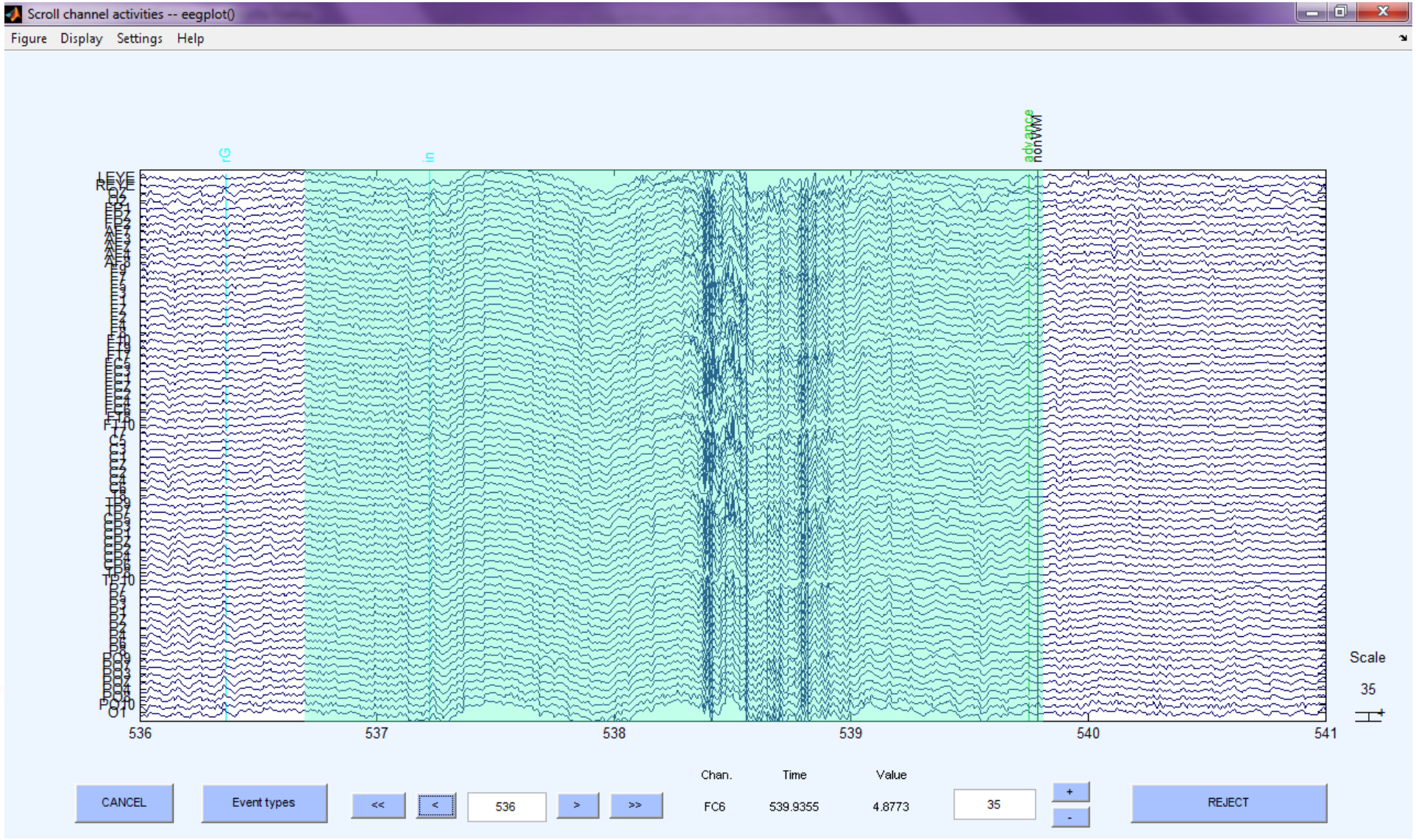
- Interpolating bad channels does not increase significance for reasonably clean data

What about line noise?

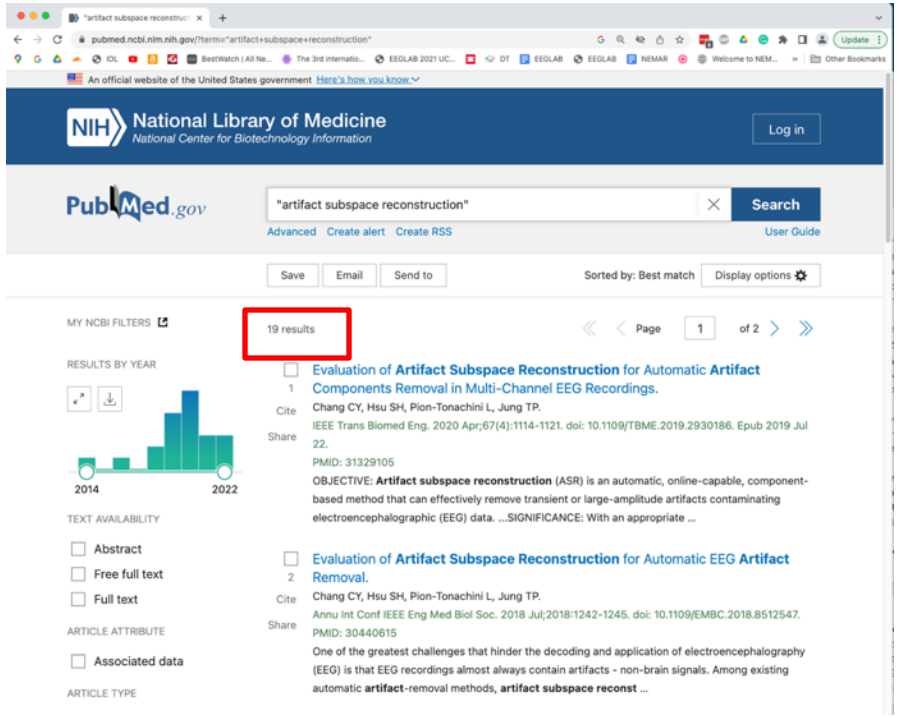


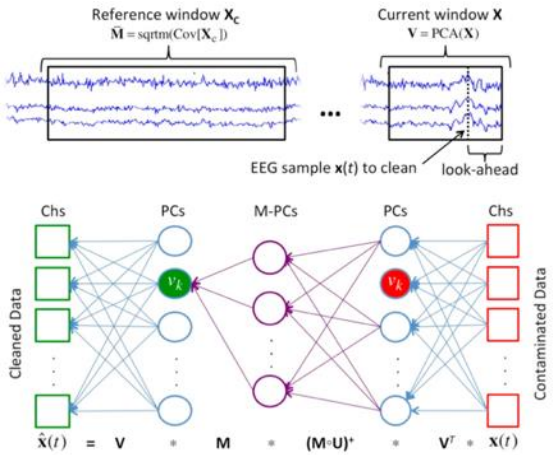
- Removing and interpolating noisy channel helps
- Notch filtering or using line-noise removal methods does not (for these datasets)

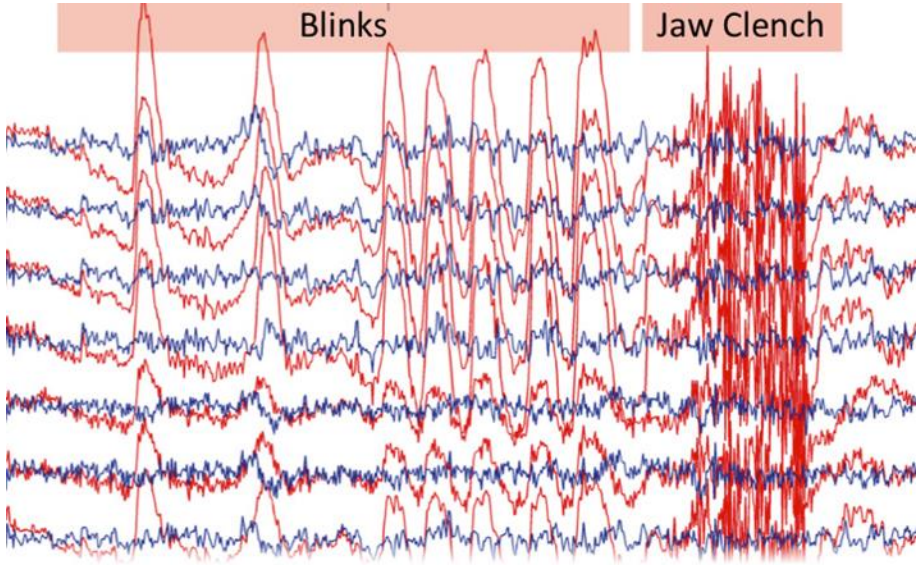
Rejecting continuous data



The powerful ASR method







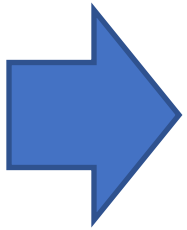
- For offline analysis, ASR is usually used to **detect and remove** but not correct data

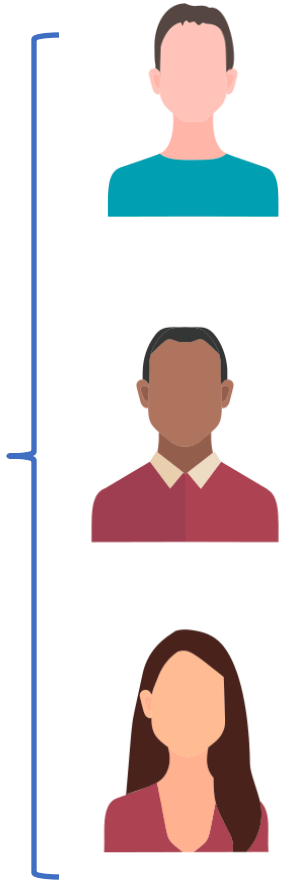
Tim R. Mullen, Christian Kothe, et al.(2015) Real-time neuroimaging and cognitive monitoring using wearable dry EEG. IEEE Transactions on Biomedical Engineering. DOI:10.1109/TBME.2015.2481482

100 files of 12 min

9a8ebd54-0550-476e-97d5-a2cd411d9ee2.fdt
9a8ebd54-0550-476e-97d5-a2cd411d9ee2.set
9b61960a-77fa-4221-a056-ef986e03d25e.fdt
9b61960a-77fa-4221-a056-ef986e03d25e.set
9e594c8d-8bae-41f5-b7bd-60e3dcdec0fc.fdt
9e594c8d-8bae-41f5-b7bd-60e3dcdec0fc.set
09f74f47-4ad0-40ee-89af-c8b42fed7638.fdt
09f74f47-4ad0-40ee-89af-c8b42fed7638.set
24a6c932-feb0-42c3-b734-ca4aeb397e6f.fdt
24a6c932-feb0-42c3-b734-ca4aeb397e6f.set
30a975b2-8858-480a-b54b-c420cd1192e9.fdt
30a975b2-8858-480a-b54b-c420cd1192e9.set
35ce3ada-48a1-4591-82c9-b890ac521878.fdt
35ce3ada-48a1-4591-82c9-b890ac521878.set
38e9e984-9976-45ee-b1ed-aaf01fb5ebb9.fdt
38e9e984-9976-45ee-b1ed-aaf01fb5ebb9.set
42b69861-11b5-4670-afc3-388d9e082798.fdt
42b69861-11b5-4670-afc3-388d9e082798.set
53c88ea0-2cdf-4ea7-9d7f-982a426af4b3.fdt
53c88ea0-2cdf-4ea7-9d7f-982a426af4b3.set
53efe3e6-0cff-48bc-9b88-ce3050d6040a.fdt
53efe3e6-0cff-48bc-9b88-ce3050d6040a.set
054f0410-a222-481c-b466-e6bb9fdac6f2.fdt
054f0410-a222-481c-b466-e6bb9fdac6f2.set
58f1f29c-e0b4-4e18-88ba-1fb9a5be9b96.fdt
58f1f29c-e0b4-4e18-88ba-1fb9a5be9b96.set
60a711a1-f357-426c-934f-5306ead30649.fdt
60a711a1-f357-426c-934f-5306ead30649.set
060b360c-c219-46b4-a45b-a86e74c6f36f.fdt
060b360c-c219-46b4-a45b-a86e74c6f36f.set
60fa6dc1-e7fc-42fd-880a-82c0f182c6df.fdt
60fa6dc1-e7fc-42fd-880a-82c0f182c6df.set
63e1d59a-c18d-443c-a64c-7c07d52f991b.fdt
63e1d59a-c18d-443c-a64c-7c07d52f991b.set
67c40959-54ce-43ed-bcc9-b6b72e343041.fdt
67c40959-54ce-43ed-bcc9-b6b72e343041.set
69a09140-7010-4fda-b23c-cbc5ed02ff63.fdt
69a09140-7010-4fda-b23c-cbc5ed02ff63.set
70a73ae3-530b-4026-a1d6-8289819e751e.fdt
70a73ae3-530b-4026-a1d6-8289819e751e.set
73eceb88-5d63-4d22-837a-9170e9a0e587.fdt
73eceb88-5d63-4d22-837a-9170e9a0e587.set
80d785af-dcb2-42f1-9cad-79120336bdce.fdt
80d785af-dcb2-42f1-9cad-79120336bdce.set
85bcc377-73b7-4636-a54c-e8801e2c3848.fdt
85bcc377-73b7-4636-a54c-e8801e2c3848.set
88eb6986-8515-4bb2-90af-2cb445b0414f.fdt
88eb6986-8515-4bb2-90af-2cb445b0414f.set
89bdc282-02b1-404d-bd3e-8a376e31f9df.fdt
89bdc282-02b1-404d-bd3e-8a376e31f9df.set
92add86d-708f-4da0-bd79-0d1c0dab4ccb.fdt
92add86d-708f-4da0-bd79-0d1c0dab4ccb.set
92cf9c1c-9f6f-40d8-9835-05f3f823f49b.fdt
92cf9c1c-9f6f-40d8-9835-05f3f823f49b.set
96c5274f-38ef-4be9-b39f-7c5fc567ae87.fdt
96c5274f-38ef-4be9-b39f-7c5fc567ae87.set

The gold standard of artifact rejection: inter-rater agreement

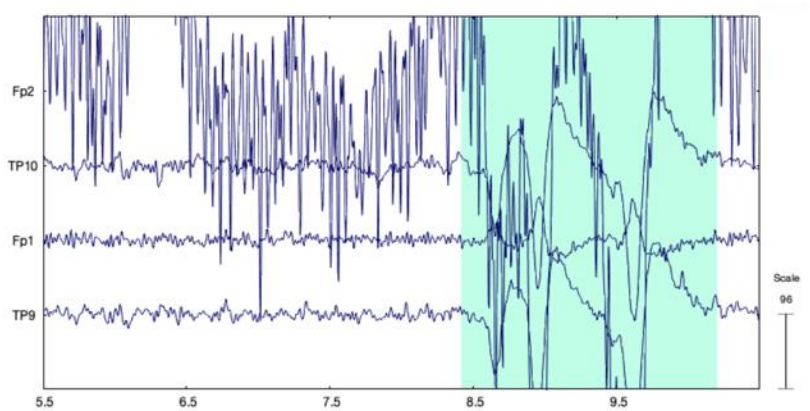




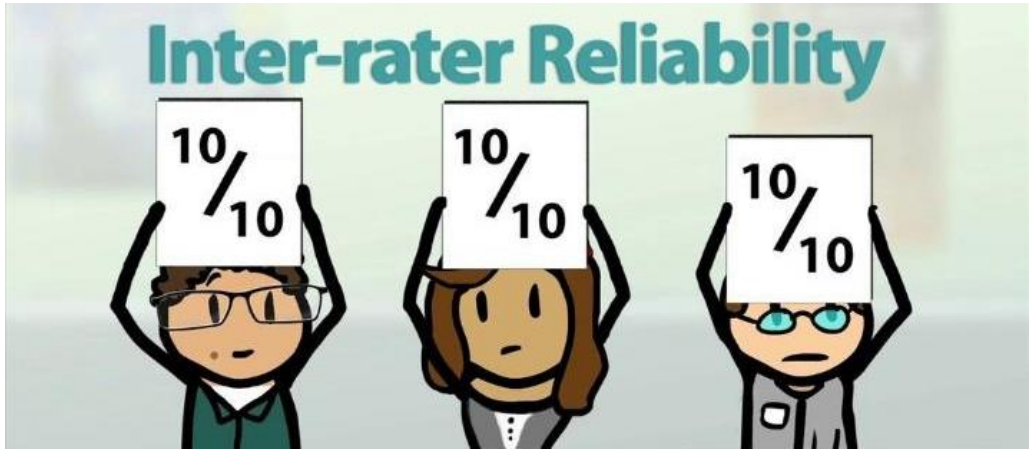
89 retained

50% finding the best algorithm

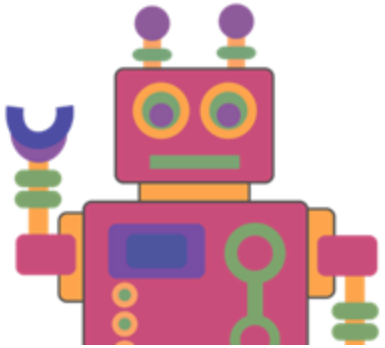
50% testing



BAD portion of data



R1 **R2** **R3**



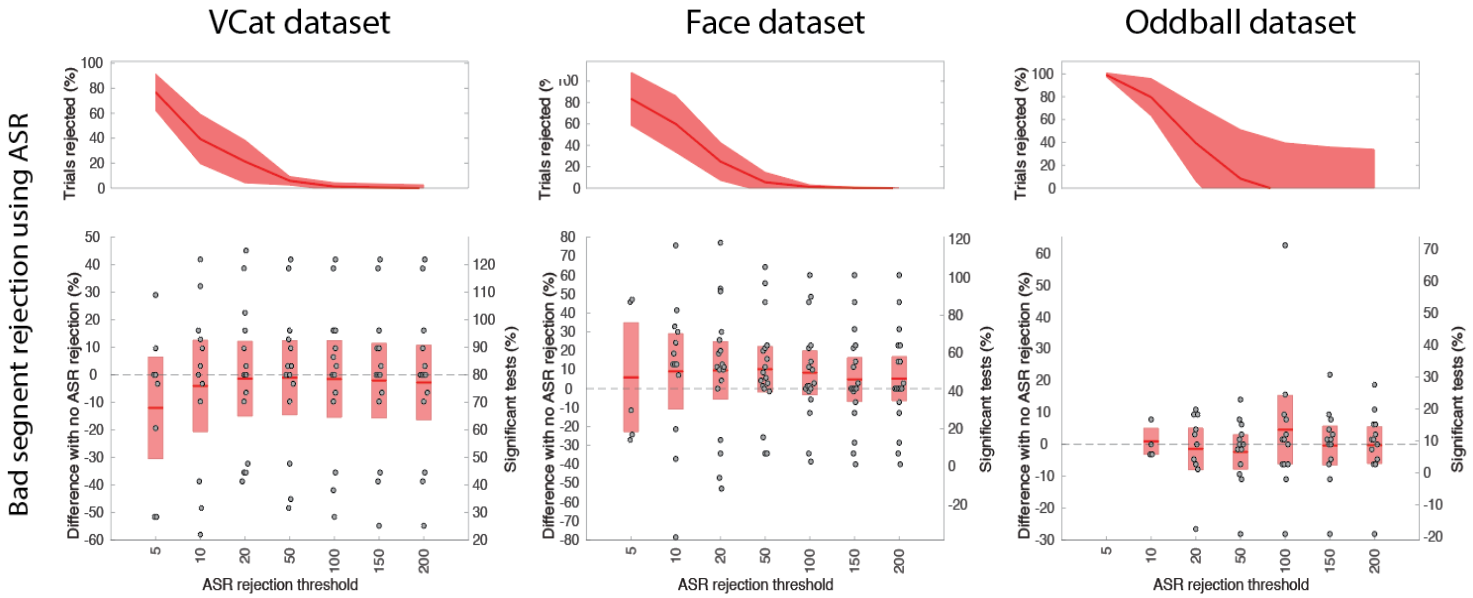
ASR

	% rejection	vs R1	vs R2	vs R3	Accuracy
R1	37	n/a	80.7	79.7	80.2
R2	37	80.7	n/a	78	79.4
R3	30	79.7	78	n/a	78.8
ASR (11)	31	81.9	82.4	80.6	81.6 (79.5 83.6)

- ASR with threshold 11 is closer to each human rater than raters are with each other on the test data

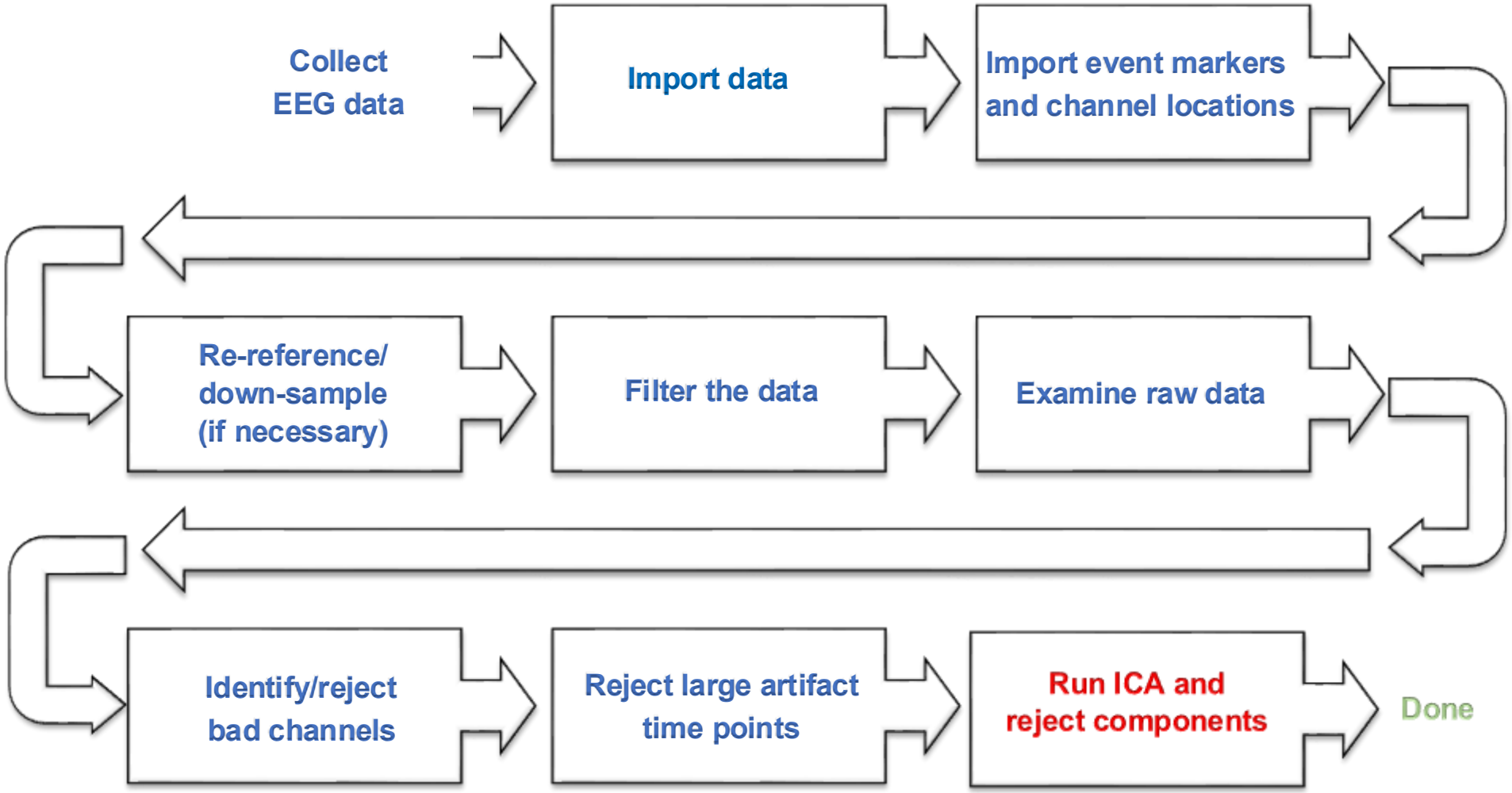
A. Delorme and J. A. Martin, "Automated Data Cleaning for the Muse EEG," *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2021, pp. 1-5, doi: 10.1109/BIBM52615.2021.9669415.

And yet…



- For **reasonably clean data, ASR brings no gain**
- The removal of bad trials does not compensate for the loss of statistical power due to the lower trial count
- Brainstorm, MNE, Fieldtrip, and other EEGLAB other automated rejection methods show no gain either

Pre-processing pipeline



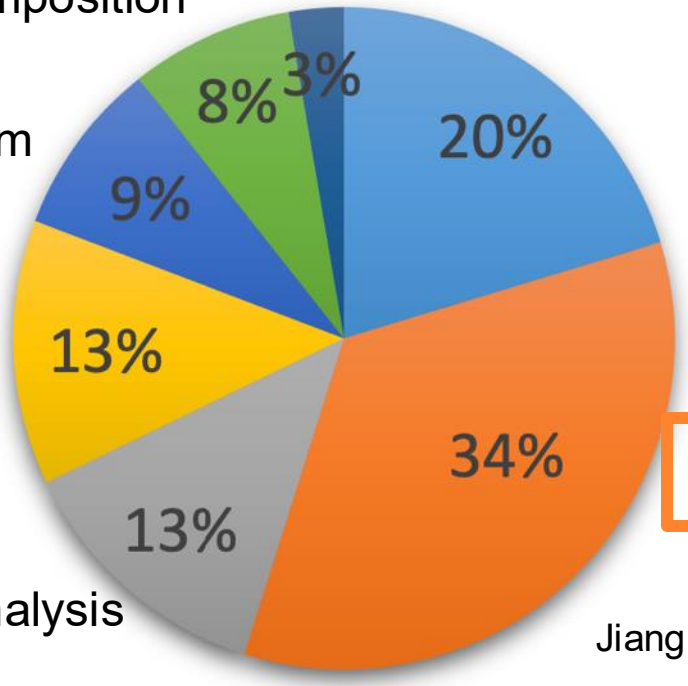
Artefact correction methods

Empirical Mode Decomposition

Wavelet Transform

Adaptive Filters

Canonical Correlation Analysis

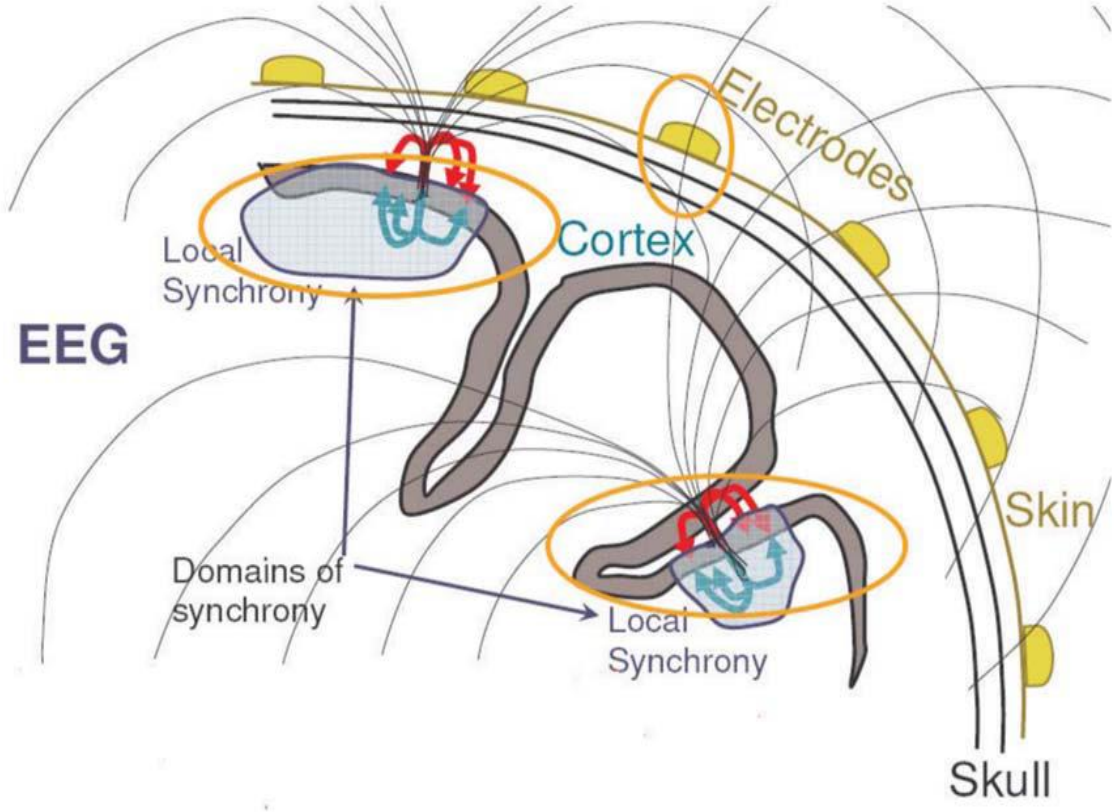


Independent Component Analysis

Jiang et al., 2019 Sensors

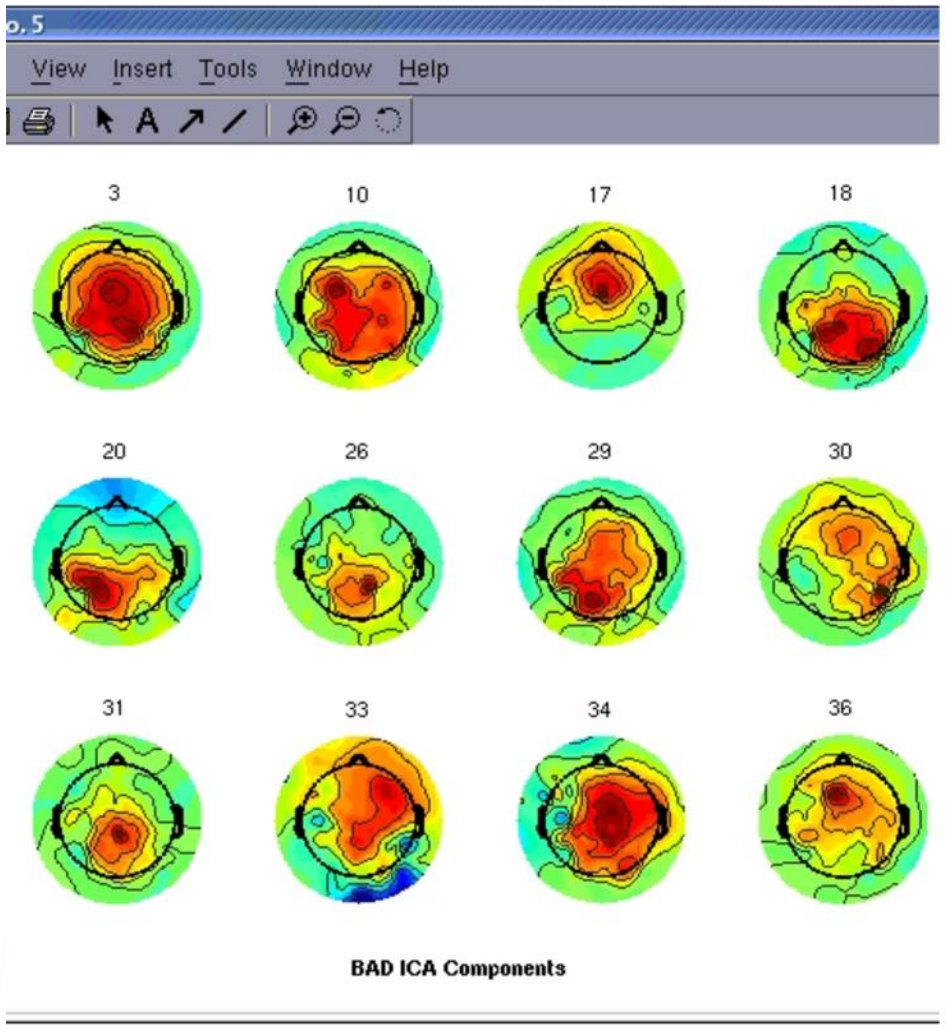
Hybrid ICA CCA Filtering WT EMD other

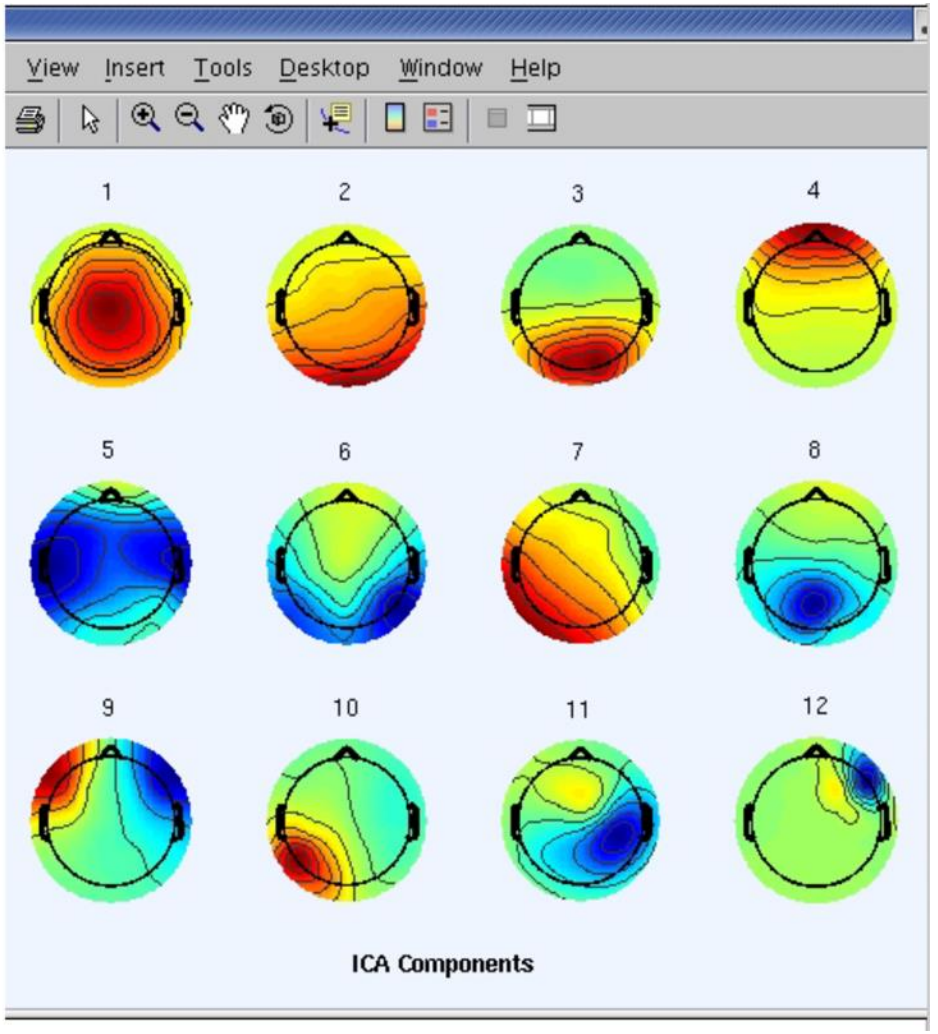
Independent Component Analysis

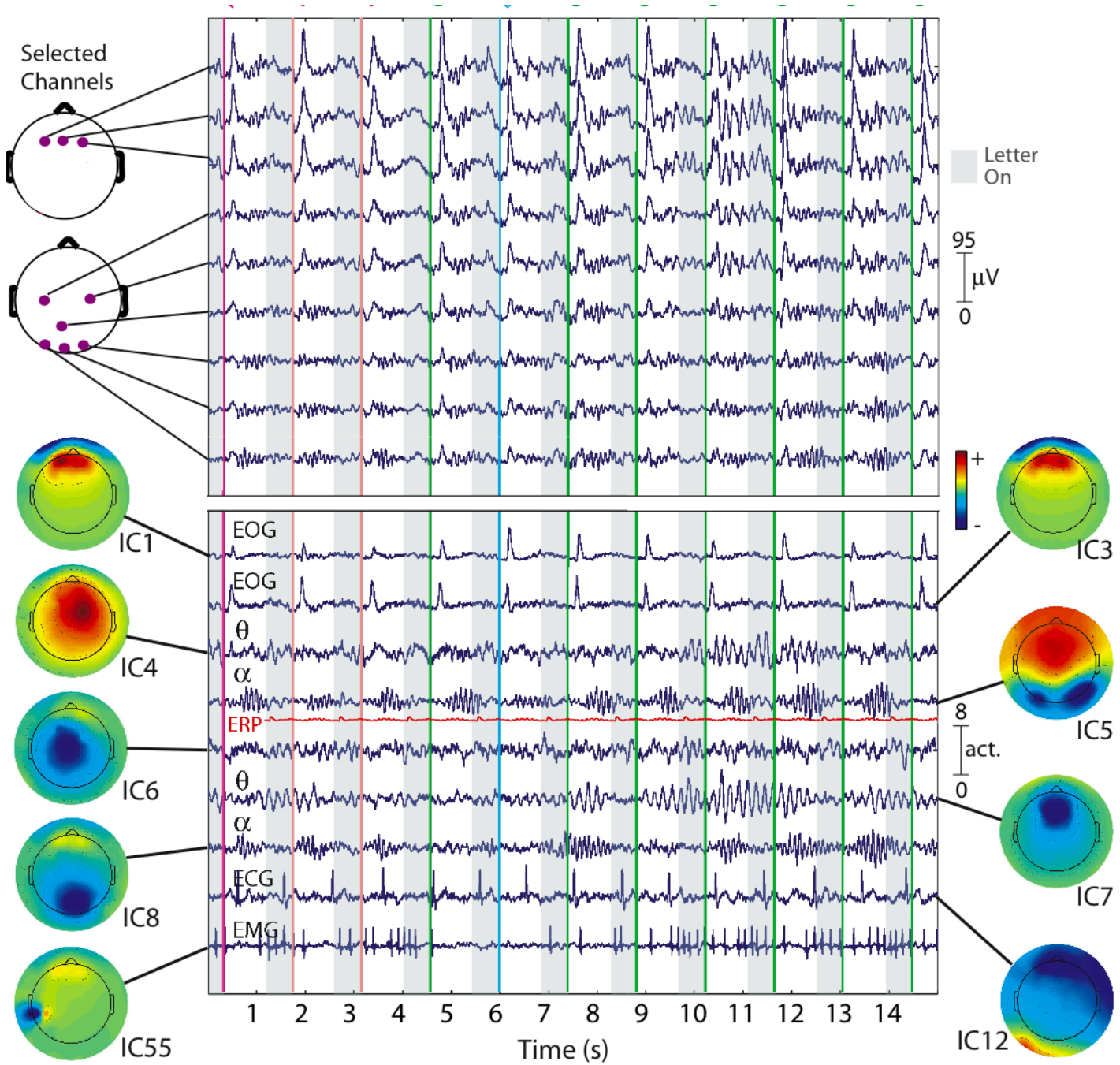


ICA – separates the EEG in temporally independent source signals (Makeig, 1996)

Quality of Decomposition






Selected Channels
Letter On
95
µV
0
+
-
EOG
EOG
θ
α
ERP
θ
α
ECG
EMG
8
act.
0
IC1
IC4
IC6
IC8
IC55
IC3
IC5
IC7
IC12
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Time (s)

ICLabel Website and Label Collection

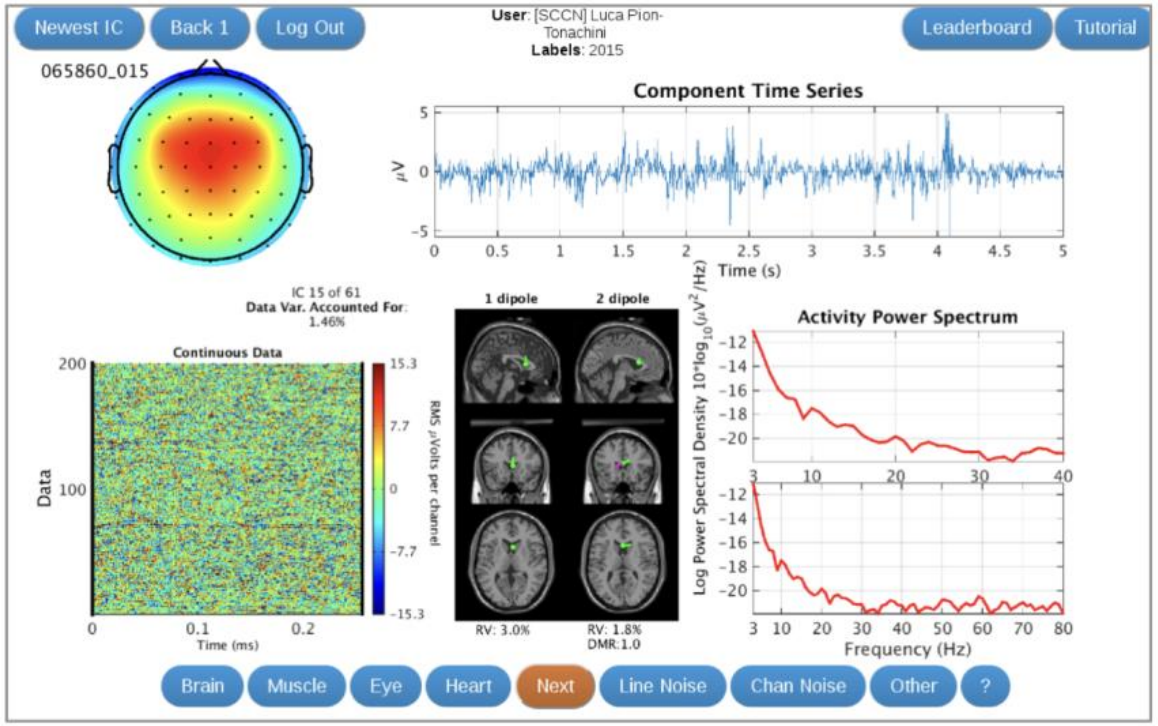
Summary: ask others to help out with labeling EEG components.

Website: labeling.ucsd.edu

Have experts from the SCCN and elsewhere label a subset.

Ask the EEGLAB community to help label a larger subset.

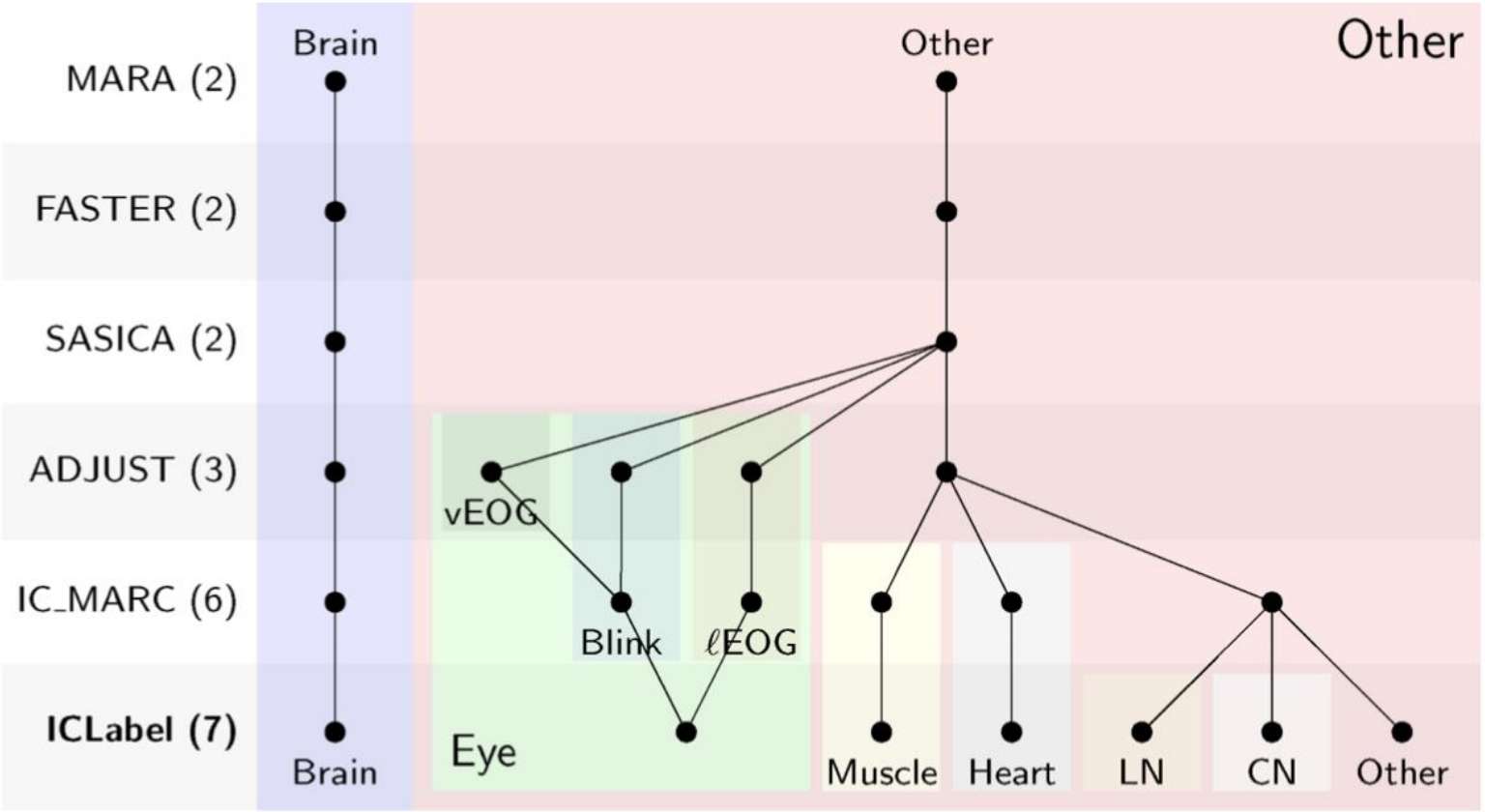
Currently 328 contributing users and 34,000+ submitted labels.



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

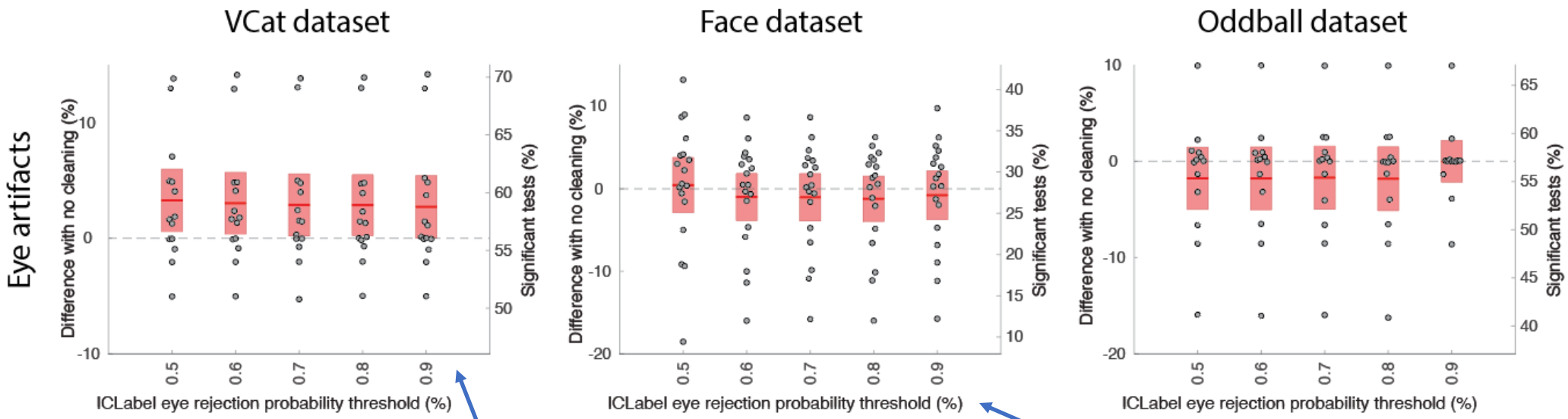
Automated ICA classification methods



Every method listed above is available as separate EEGLAB plugins

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

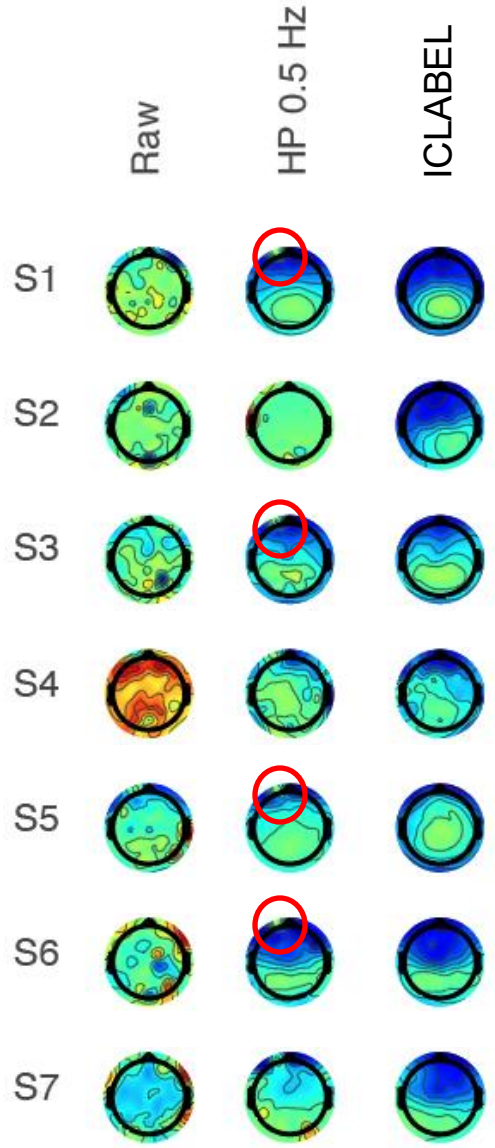
Does ICA help with ERP significance?



- Sometimes a little, but the difference can also be null
- Note that ICA may remove artifacts which contribute to differences between conditions (so this method of assessing performance is not ideal)

Should I even use ICA to remove artifacts, then?

- Yes, because it remove channel artifacts on scalp topographies (good for visualization)
- Yes, because it may improve subsequent source localization



frontiers | Frontiers in Signal Processing

TYPE Original Research
PUBLISHED 03 April 2023
DOI 10.3389/frsip.2023.1064138

OPEN ACCESS

EDITED BY
Rishi Raj Sharma,
Defence Institute of Advanced Technology (DIAT), India

REVIEWED BY
Alina Santillan Guzman,
Universidad Popular Autónoma del Estado de Puebla, Mexico
Anurag Nishad,
Birla Institute of Technology and Science, India

ICA's bug: How ghost ICs emerge from effective rank deficiency caused by EEG electrode interpolation and incorrect re-referencing

Hyeonseok Kim[1], Justin Luo[2], Shannon Chu[1], Cedric Cannard[3,4], Sven Hoffmann[5] and Makoto Miyakoshi[1,6,7]*

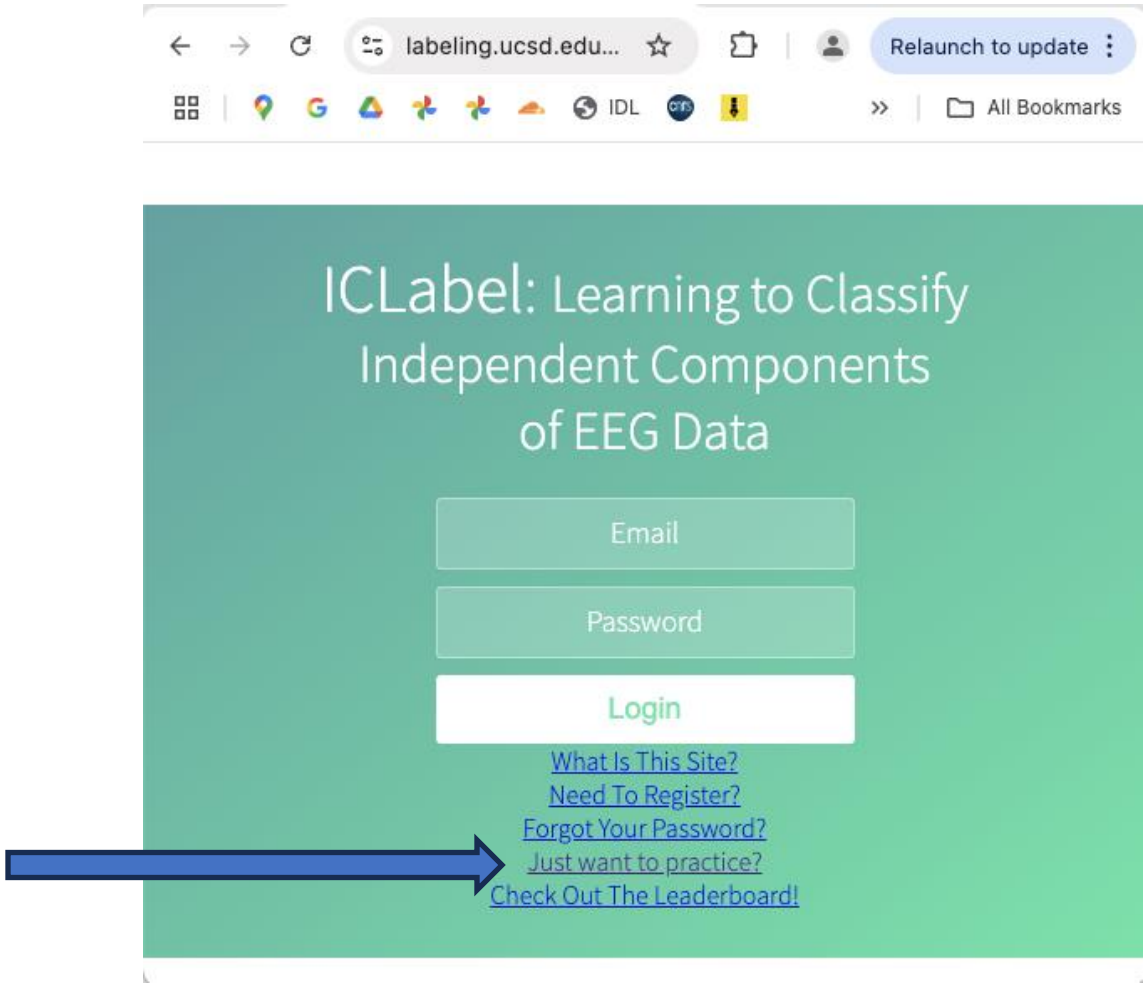
This is no 'ICA bug': Response to the article, "ICA's bug: How ghost ICs emerge from effective rank deficiency caused by EEG electrode interpolation and incorrect re-referencing"

Arnaud Delorme*, Scott Makeig (*corresponding author)

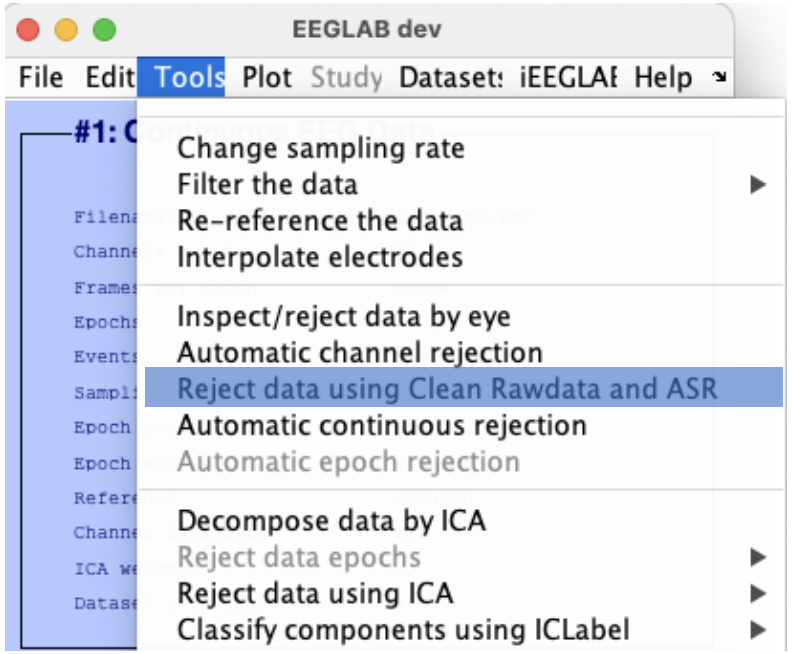
Swartz Center for Computational Neuroscience
Institute for Neural Computation
University of California San Diego 0559

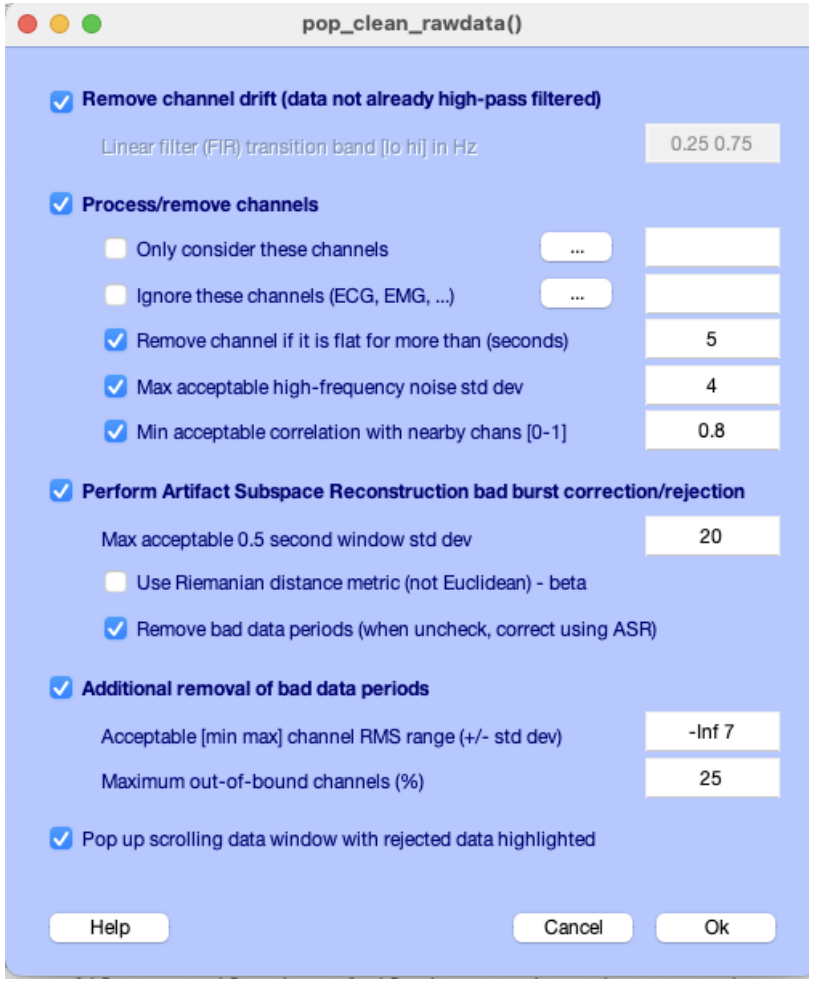
Hands on !

Practicing ICLabel



Clean_rawdata in EEGLAB





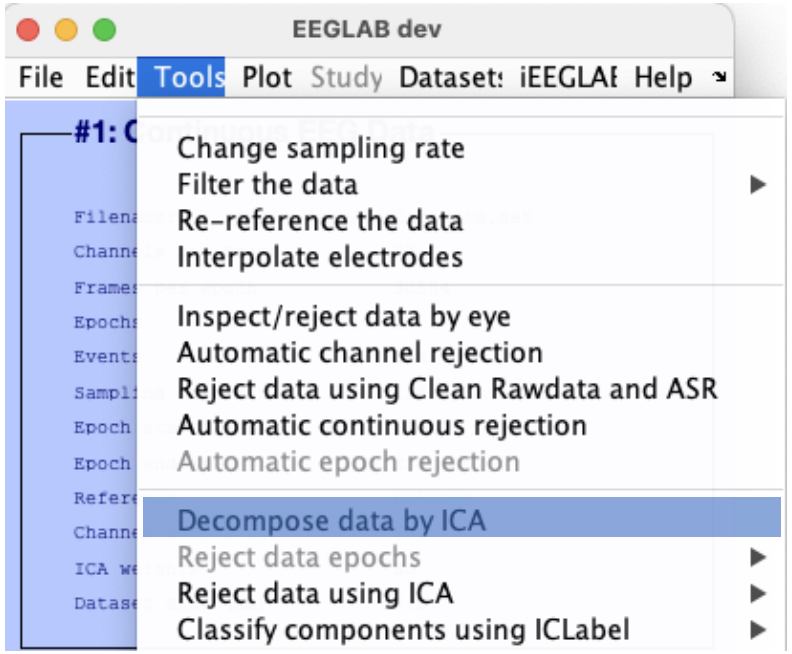
Filter the data

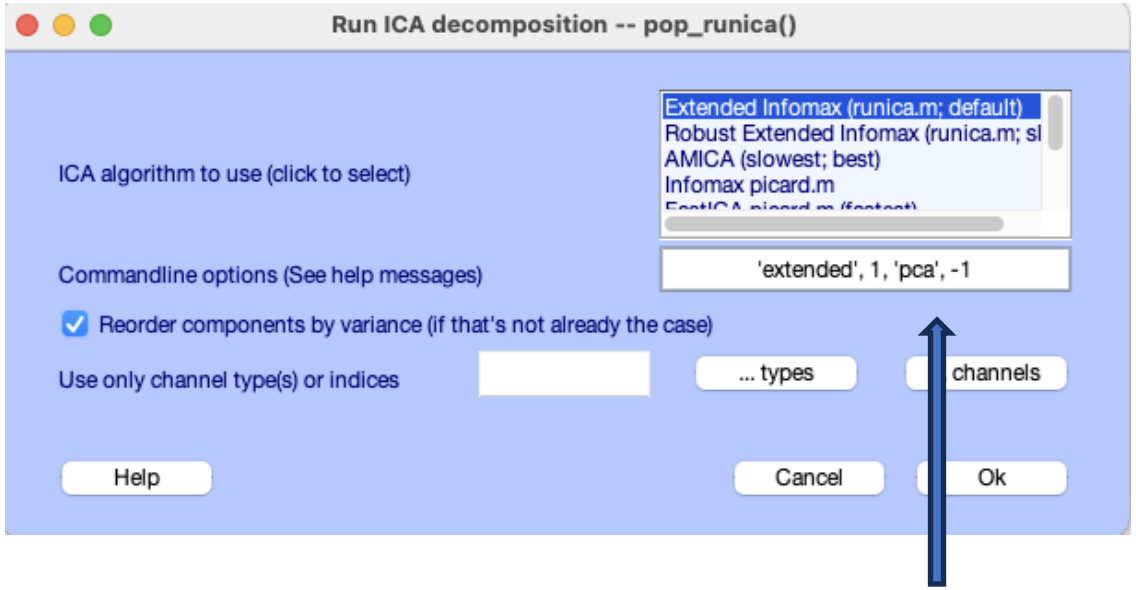
Channel rejection

ASR segment rejection

Final segment rejection based on simple thresholds on signal amplitude

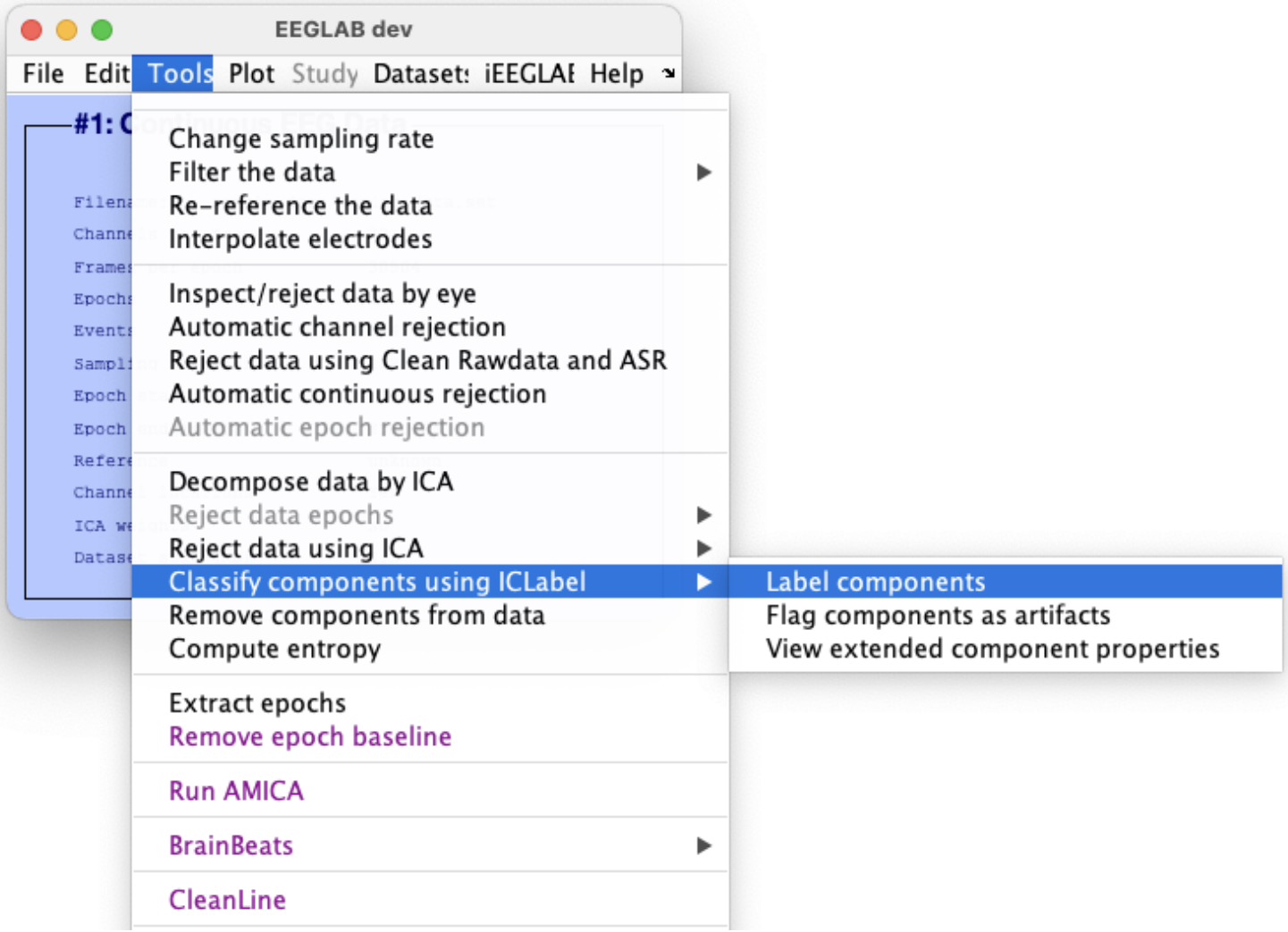
Clean_rawdata in EEGLAB



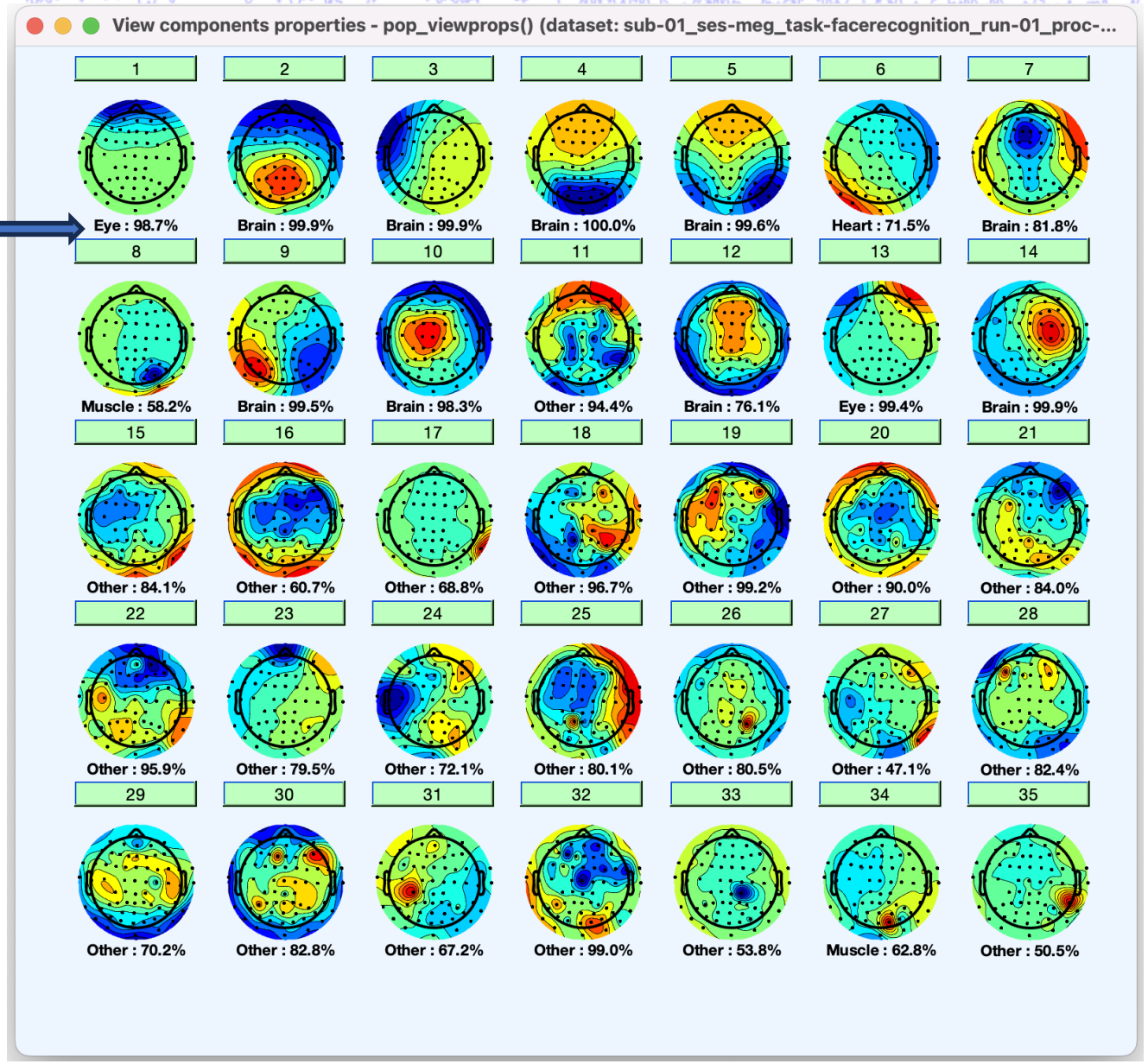


Use ‘pca’, -1 when average reference is calculated to decrease the dimension of the data by 1 before ICA (we will come back to that in the ICA lecture)

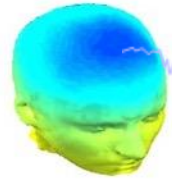
Run ICLabel

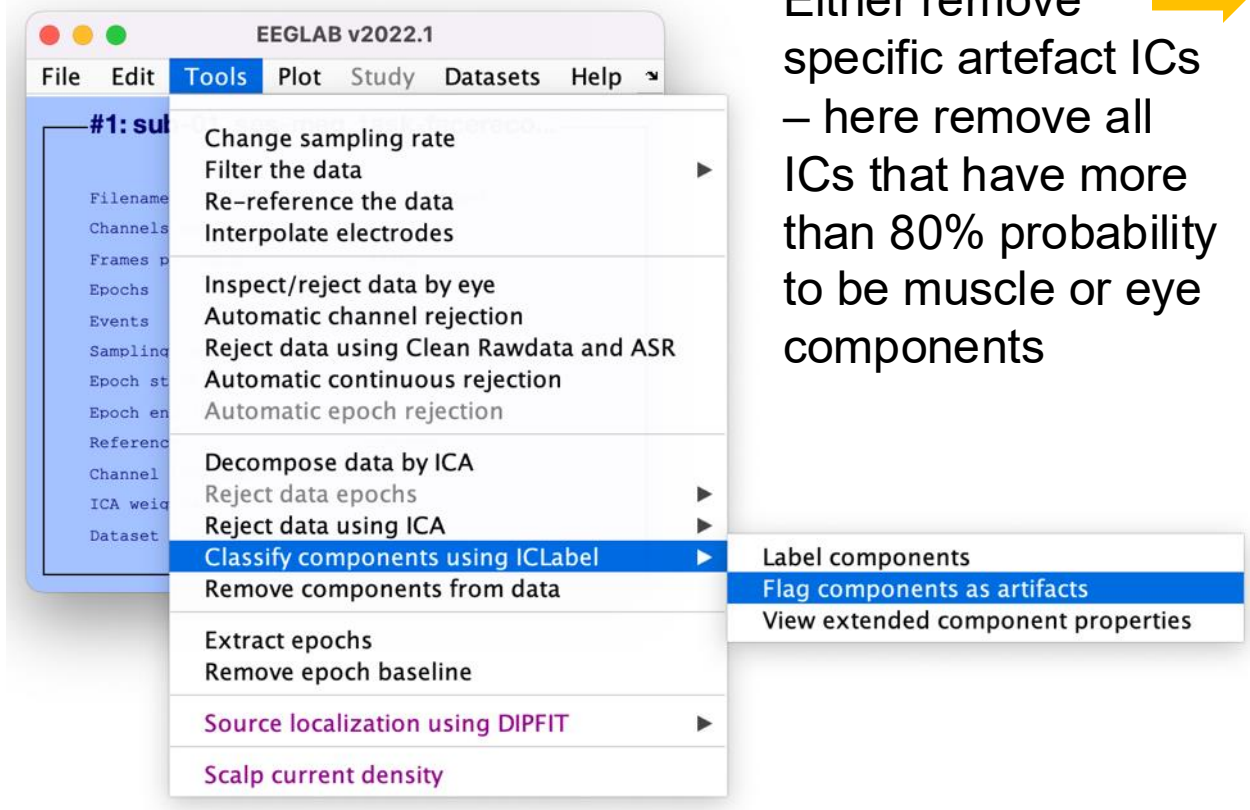

EEGLAB dev
File Edit Tools Plot Study Dataset: iEEGLAB Help
Change sampling rate
Filter the data
Re-reference the data
Interpolate electrodes
Inspect/reject data by eye
Automatic channel rejection
Reject data using Clean Rawdata and ASR
Automatic continuous rejection
Automatic epoch rejection
Decompose data by ICA
Reject data epochs
Reject data using ICA
Classify components using ICLabel
Remove components from data
Compute entropy
Extract epochs
Remove epoch baseline
Run AMICA
BrainBeats
CleanLine
Label components
Flag components as artifacts
View extended component properties

Look at component class and probability of belonging to that class.

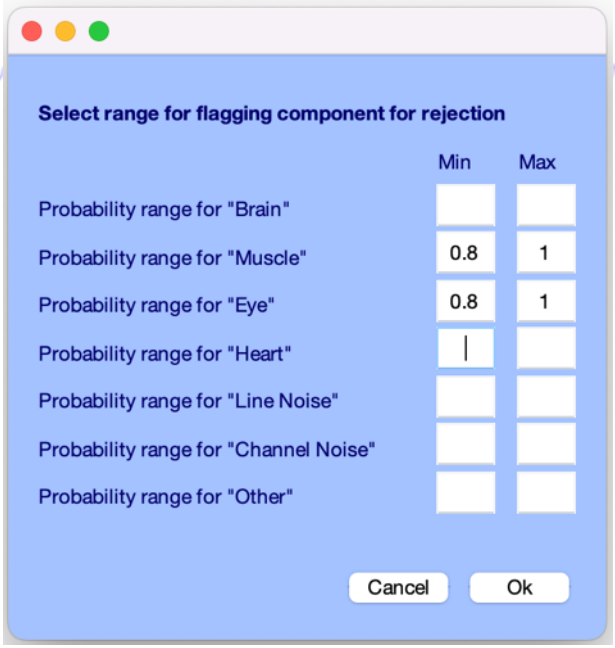


Let's remove artefact ICs using IC Label classification probabilities

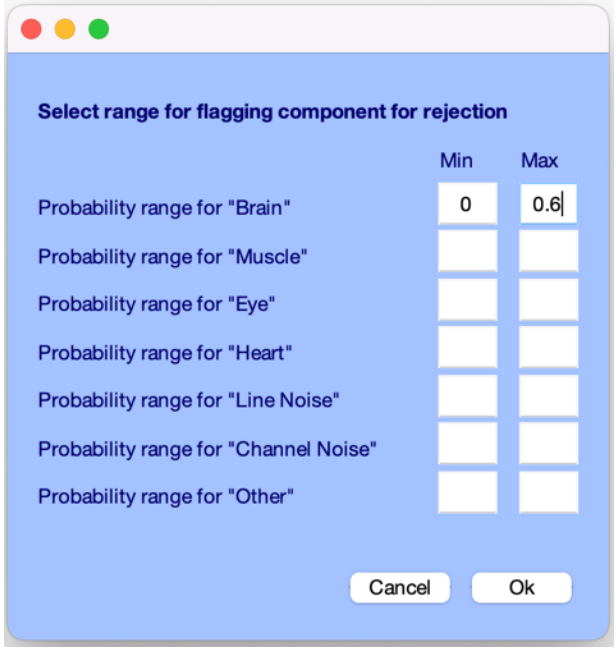




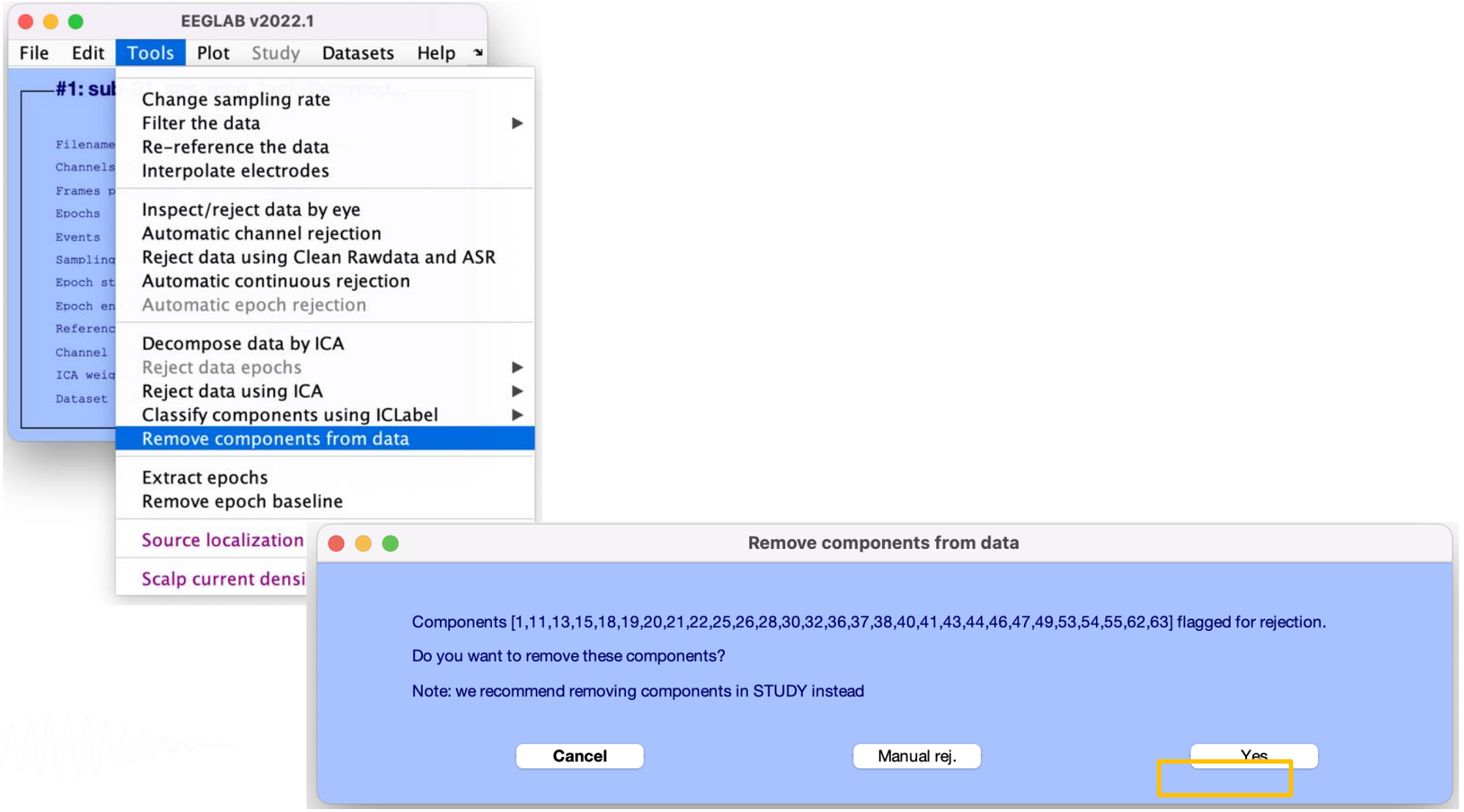
Either remove specific artefact ICs – here remove all ICs that have more than 80% probability to be muscle or eye components



Or remove all ICs that have less than 60% probability to be brain components (you can also choose a different probability – i.e. 70%)



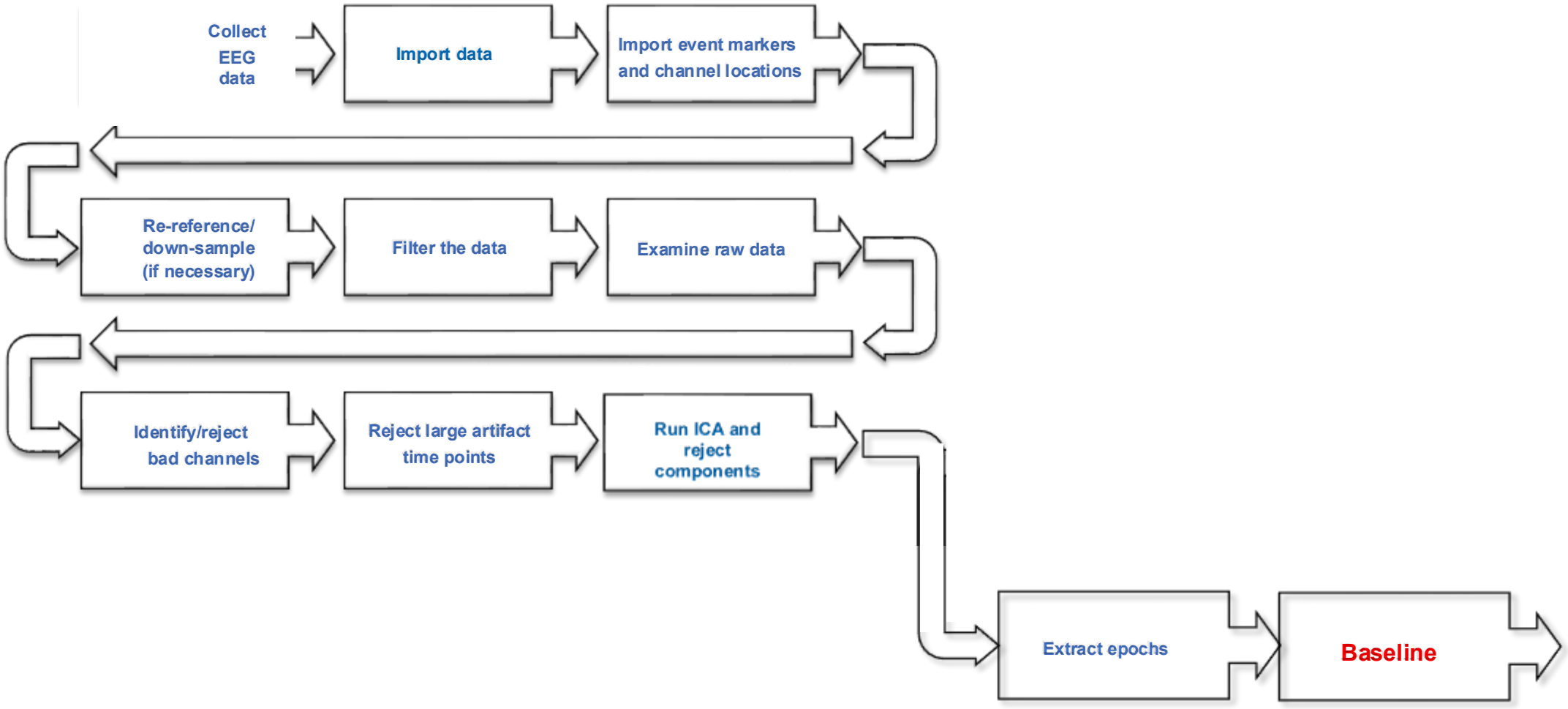
Let’s remove artefact ICs using IC Label classification probabilities



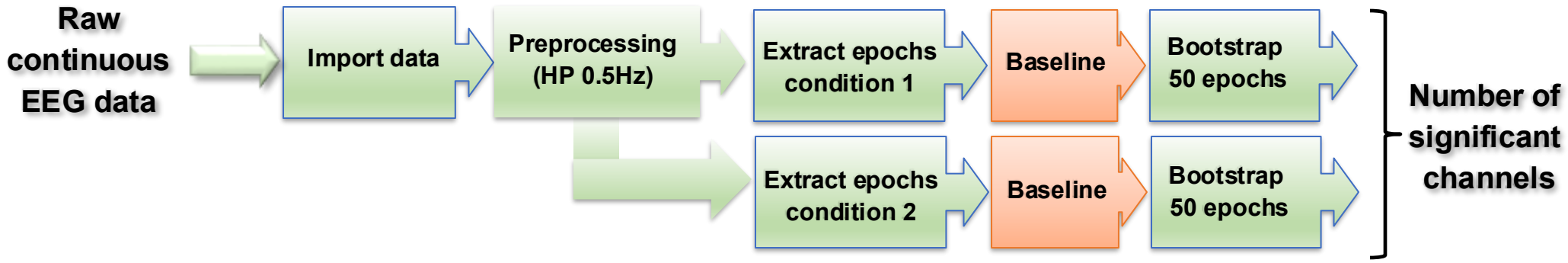
Hands on: Basic preprocessing

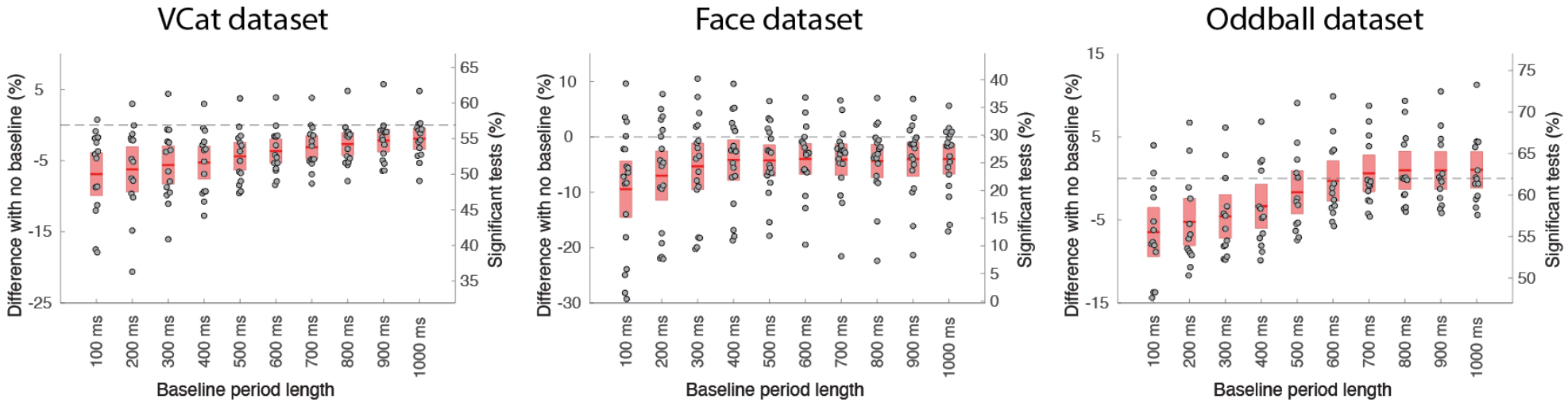
- Use menu item **Tools > Reject data using clean_rawdata and ASR.** Select filtering. Leave all defaults. Inspect the results (**Plot > Scroll data**)
- Use menu item **Tools > Run ICA data to run ICA**
- Use menu item **Tools > Classify components using ICLabel > Label components**, and inspect components
- **Note:** The preprocessing is more involved. It includes referencing, an initial call to **clean_rawdata** to remove bad channels, ICA, then a second call to **clean_rawdata** to remove corrupted portions of data. This order is specific to this dataset, which contains many blink artifacts that we do not want **clean_rawdata** to remove because ICA handles them. The procedure is described in *script_02_preprocess_data.m*. Later lectures will use the output of that script. If it has not already been run for you, you will need to run it.

What about removing the baseline?



Influence of the baseline on ERPs





- Any baseline tend to decrease significance or has no effect (simply high-pass filter the data)
- Short baseline are worst

But wait… what about baseline with minimal High-Pass

Journal of Neuroscience Methods 266 (2016) 166–170

Contents lists available at ScienceDirect

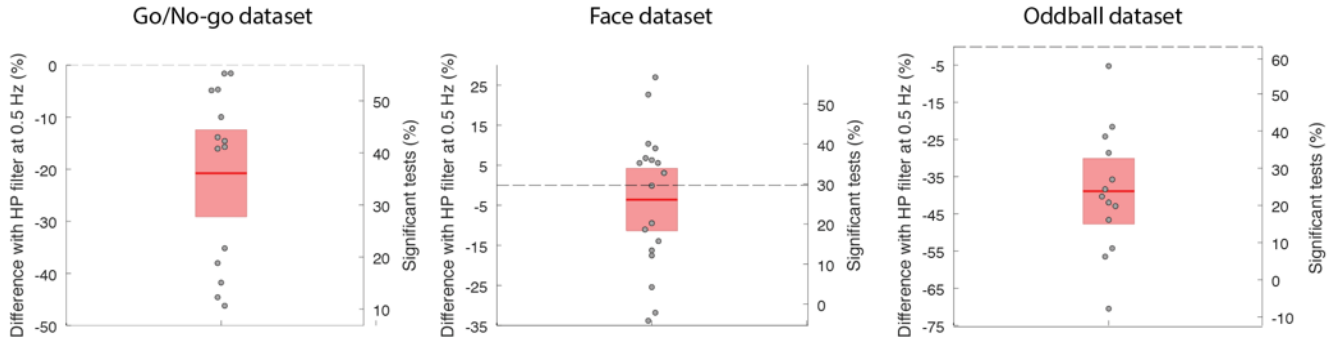
Journal of Neuroscience Methods

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

On high-pass filter artifacts (they're real) and baseline correction (it's a good idea) in ERP/ERMF analysis

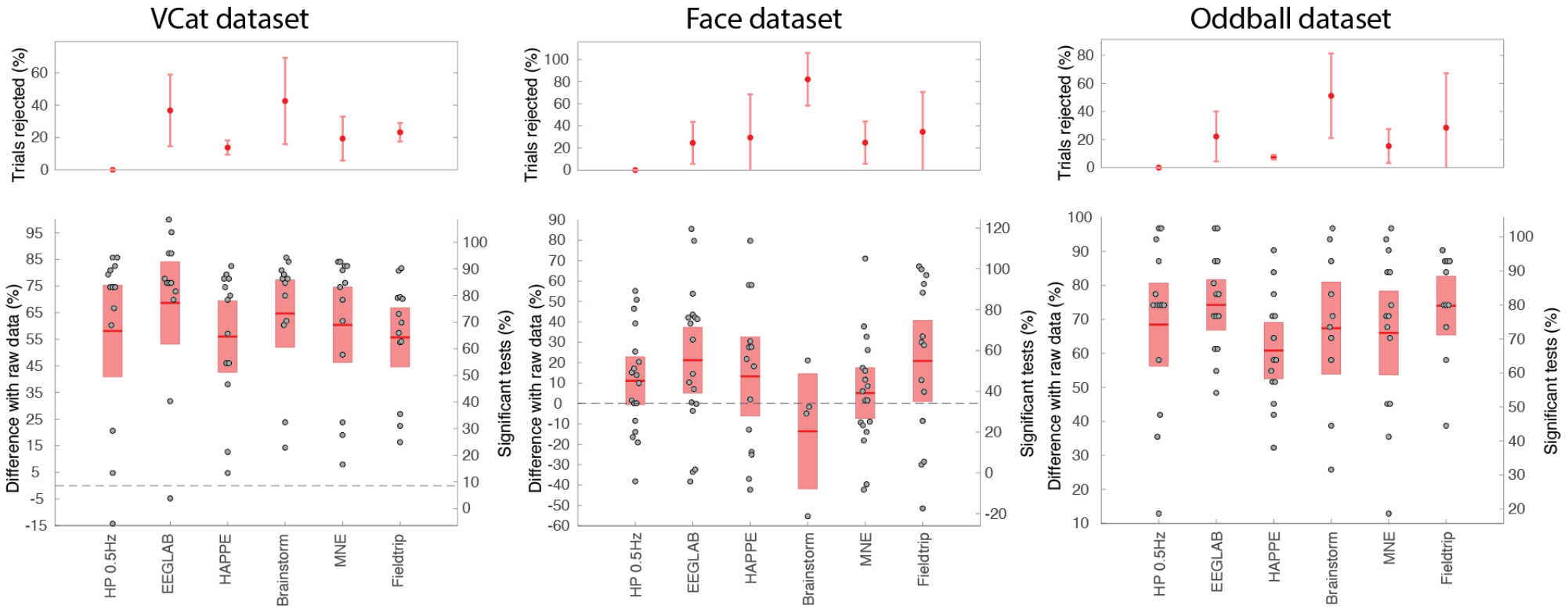
Darren Tanner[a,b,*], James J.S. Norton[b], Kara Morgan-Short[c], Steven J. Luck[d]

- The article above advocate high-pass filter at 0.01Hz and a baseline of 100 ms
- Compared to high-pass at 0.5Hz and no baseline, the number of significant channels is often cut by half
- No baseline is preferable to maximize significance



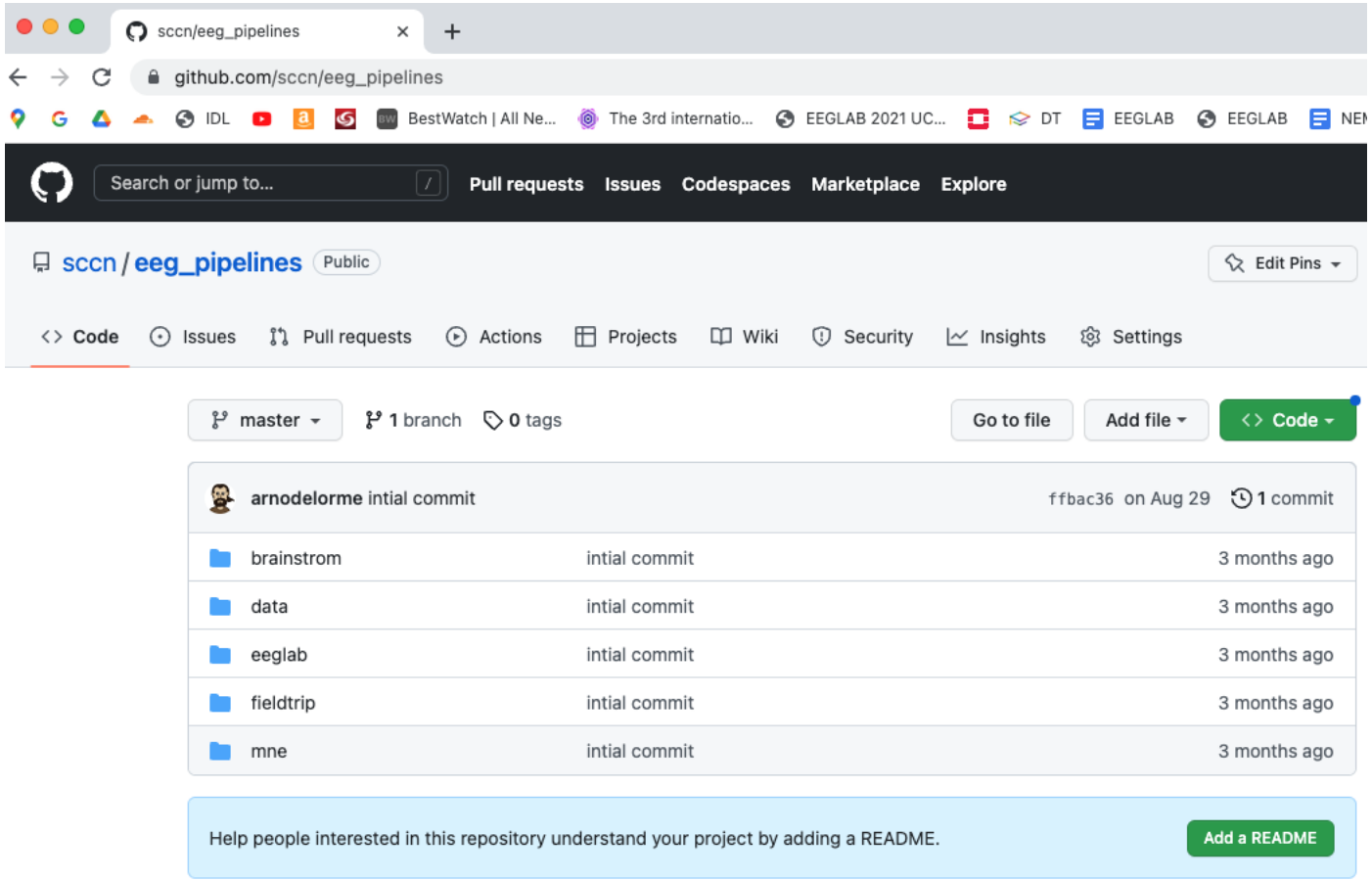
Is there an optimal preprocessing pipeline?

- Each automated artifact rejection optimized (EEGLAB, MNE, Brainstorm, Fieldtrip)



- Compared to high-pass at 0.5Hz, not much improvement of any pipeline (although EEGLAB pipeline the only one significantly superior to 0.5 Hz high pass for all 3 datasets – no other pipeline superior to 0.5 Hz high pass for any dataset)
- EEGLAB minor advantage due to line-noise contaminated channel interpolation and ICA automated rejection

Automated pre-processing pipelines available



- Has test data
- Please comment and help improve the automated pipeline for each software

The **EEGLAB** Youtube Channel