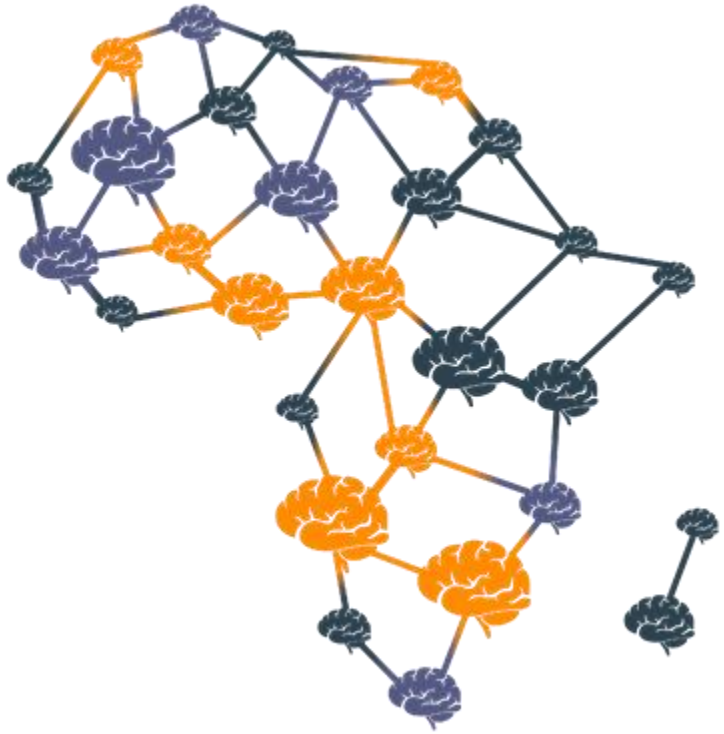
Preprocessing of EEG data

Robert Oostenveld

Mikkel C. Vinding

9-14 June 2025
Port Harcourt, Nigeria



African Brain Data Network

Outline and topics to cover

What is pre-processing, versus processing, versus analysis

Aligning the EEG with external timed events

Common artefacts

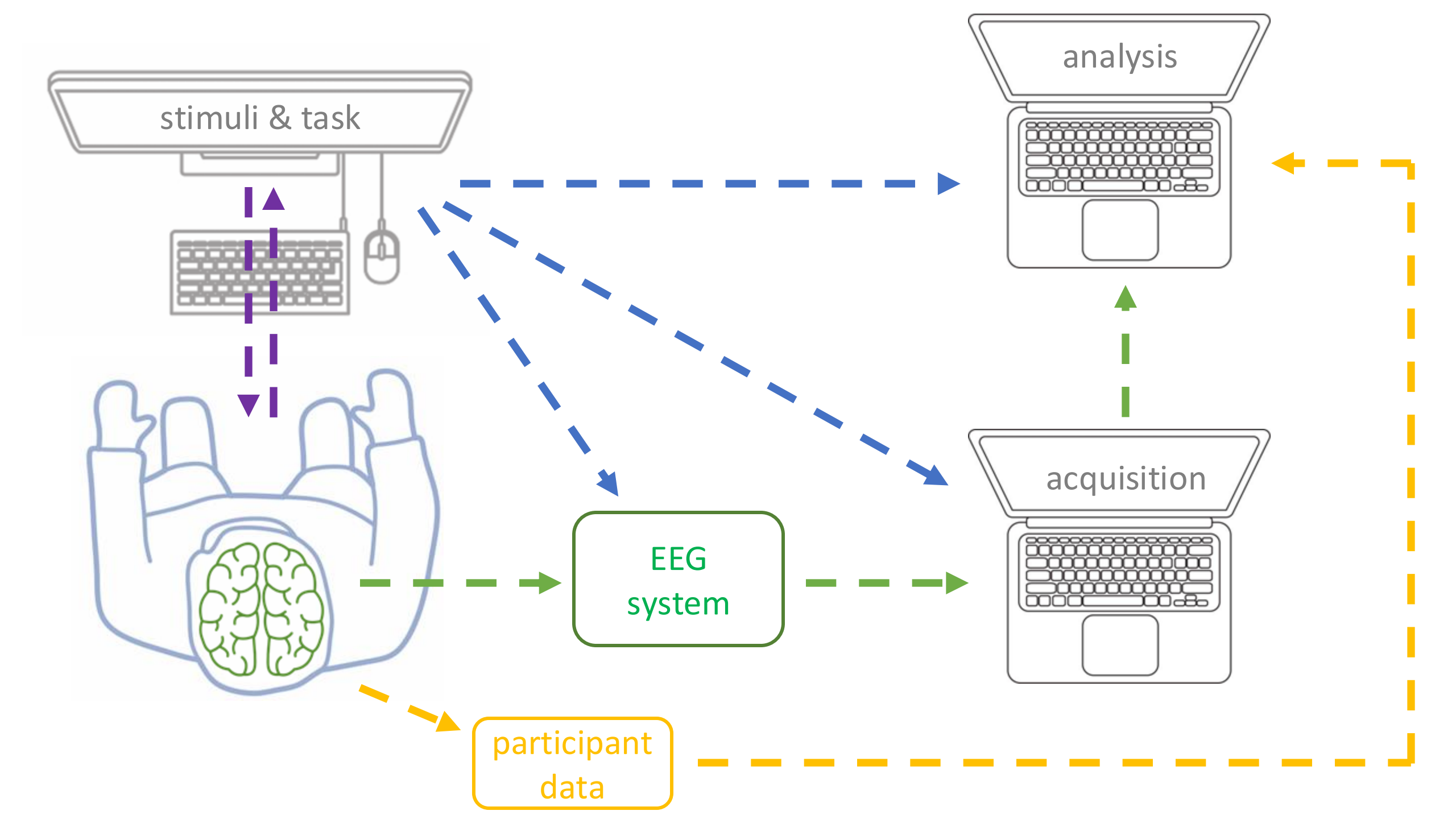
Artifacts in the data versus not doing the task properly

Artefact rejection vs. correction

Filtering, ICA, rereferencing

Segmenting

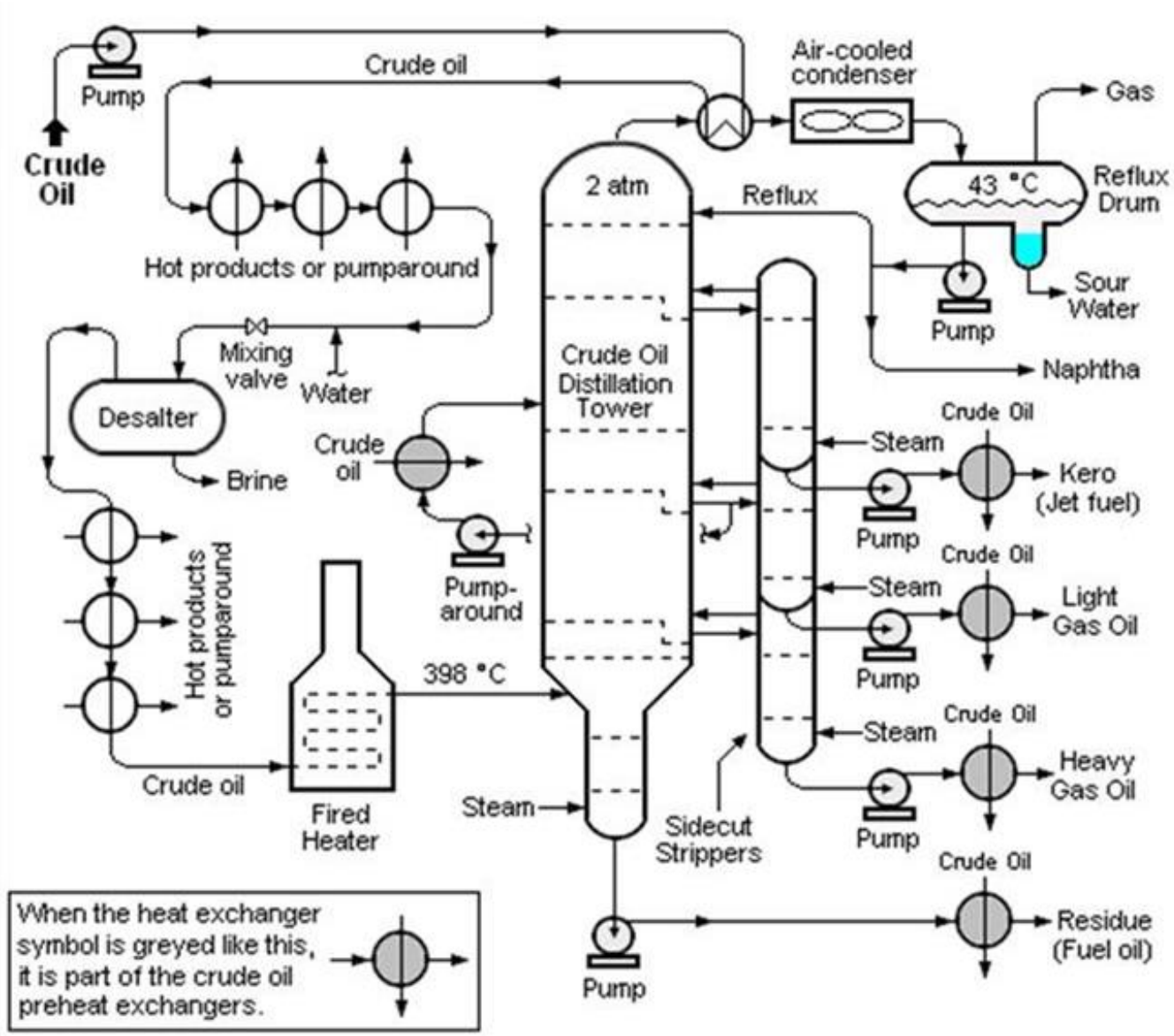
Confounds


stimuli & task
analysis
acquisition
EEG
system
participant
data

What is preprocessing – compare oil refinery

Crude oil goes in

Gasoline comes out



EEG preprocessing – refining the data

Common and unavoidable artifacts are due to physiological activity

- EMG = muscle activity
- ECG = heart activity
- EOG = eye activity, note that we try to avoid these by means of the experimental design

Artifacts can be due to the electrical contact between scalp and electrode, especially when the impedance (=resistance) is high and/or whether electrodes or cables move

Artifacts can be due to the environment

- 50 or 60Hz line noise, especially when touching something
- RF interference from bluetooth, wifi, or a cell-phone

EEG artifacts

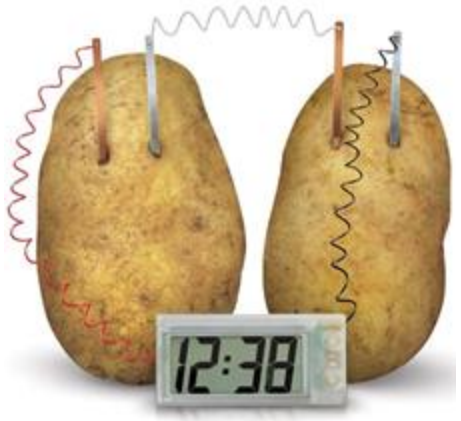
~1.000.000 Volt

1,5
Volt

1-10 micro Volt

EEG artifacts

Power line noise, 50/60 Hz electrical equipment, radio signals



Electrochemical noise (sweating)

Electrostatic noise (e.g., rubbing feet over the carpet)

- Mostly common-mode, i.e., similar on all channels

Physiological bioelectricity

- Muscle (EMG)
- Heart (ECG)
- Eye movements (EOG)

What is preprocessing, or processing, or analysis

"Pre-processing" is that part of the data analysis that does not depend strongly on the research question.

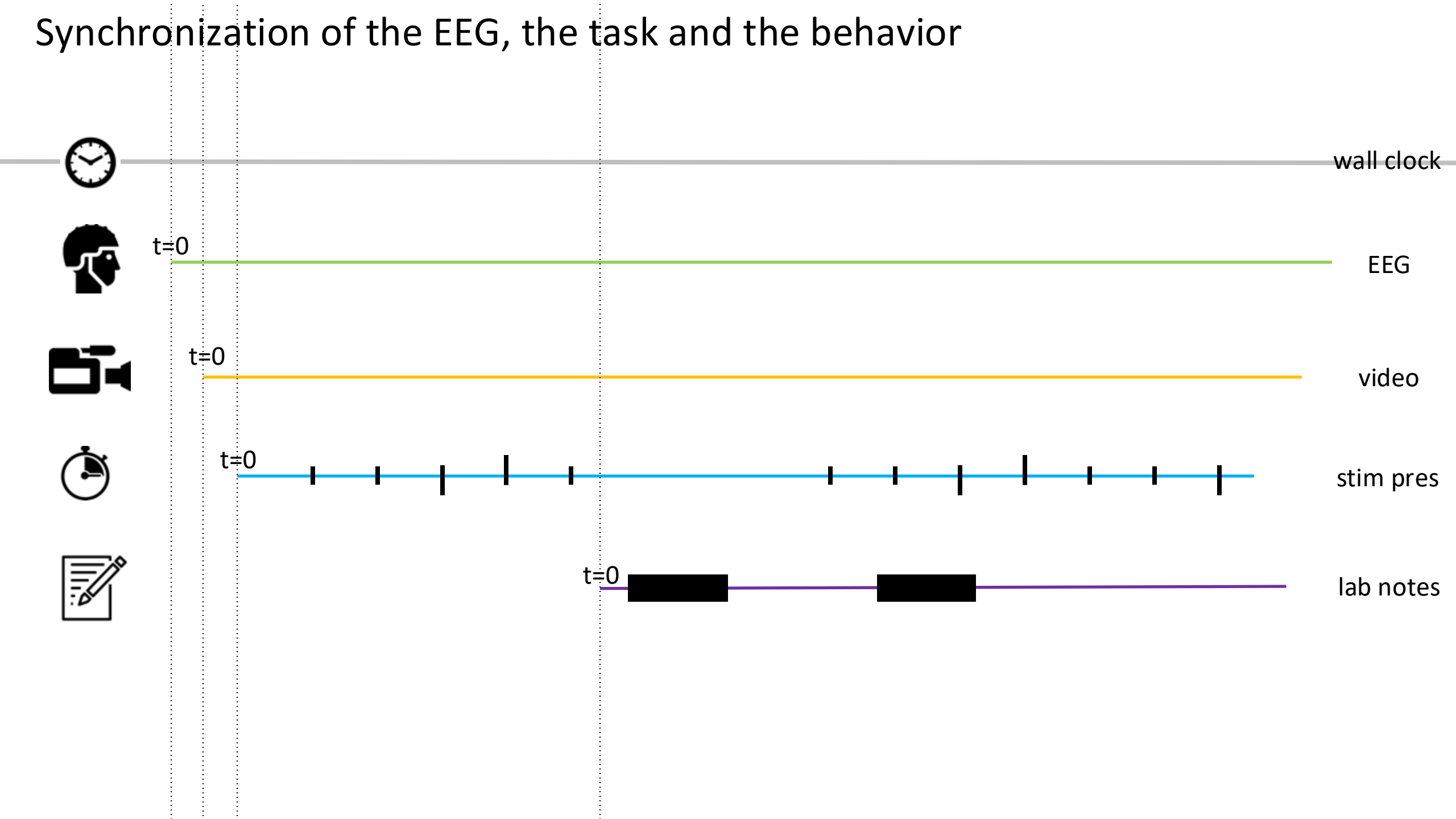
However, it is not totally independent of the research question or experimental design.

Behavior of the participant can relate to the experiment and cause artifacts that need to be taken into account -> possible confounds.

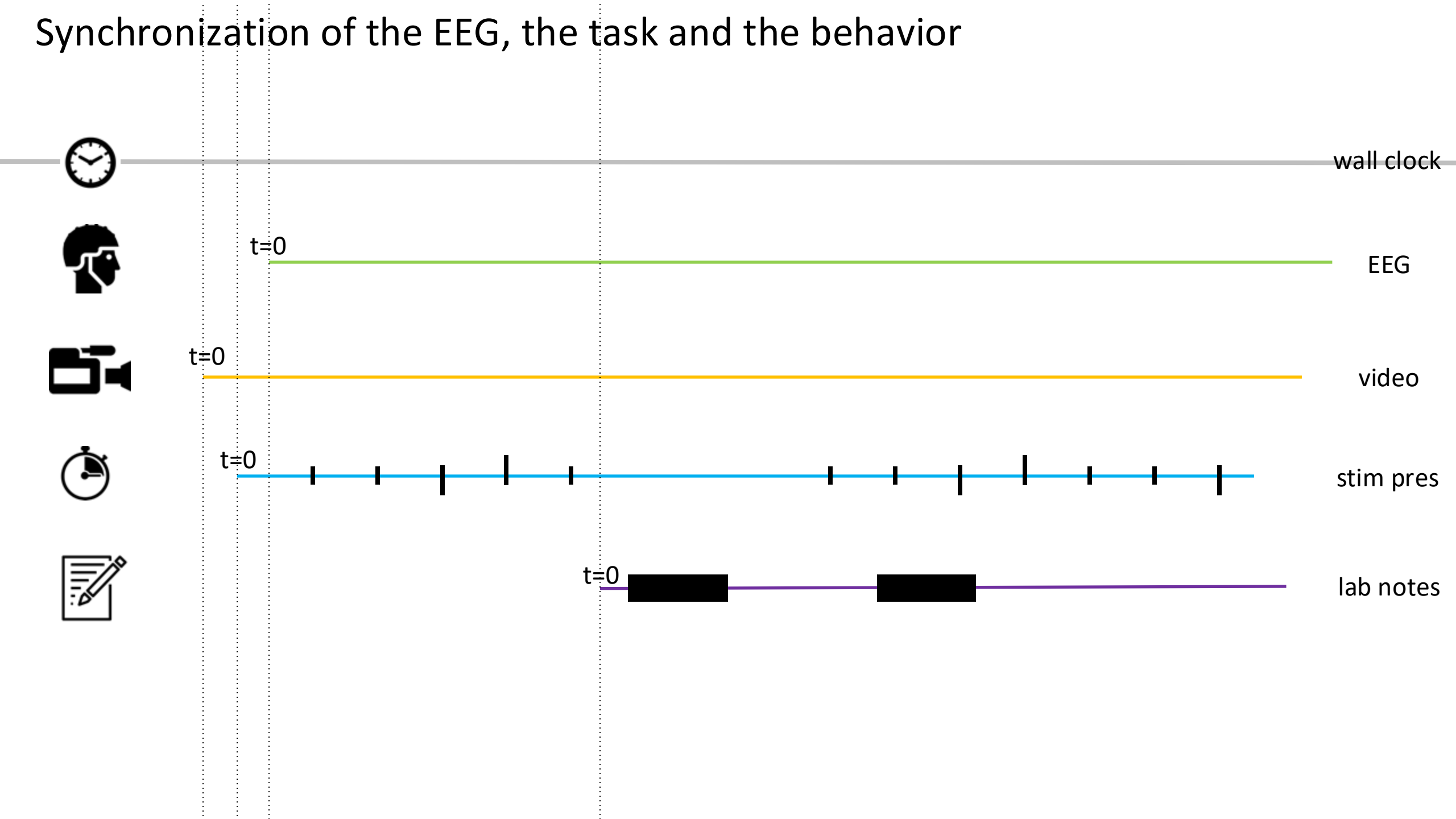
Preprocessing - synchronization

Collect all data and figure out how it relates to each other

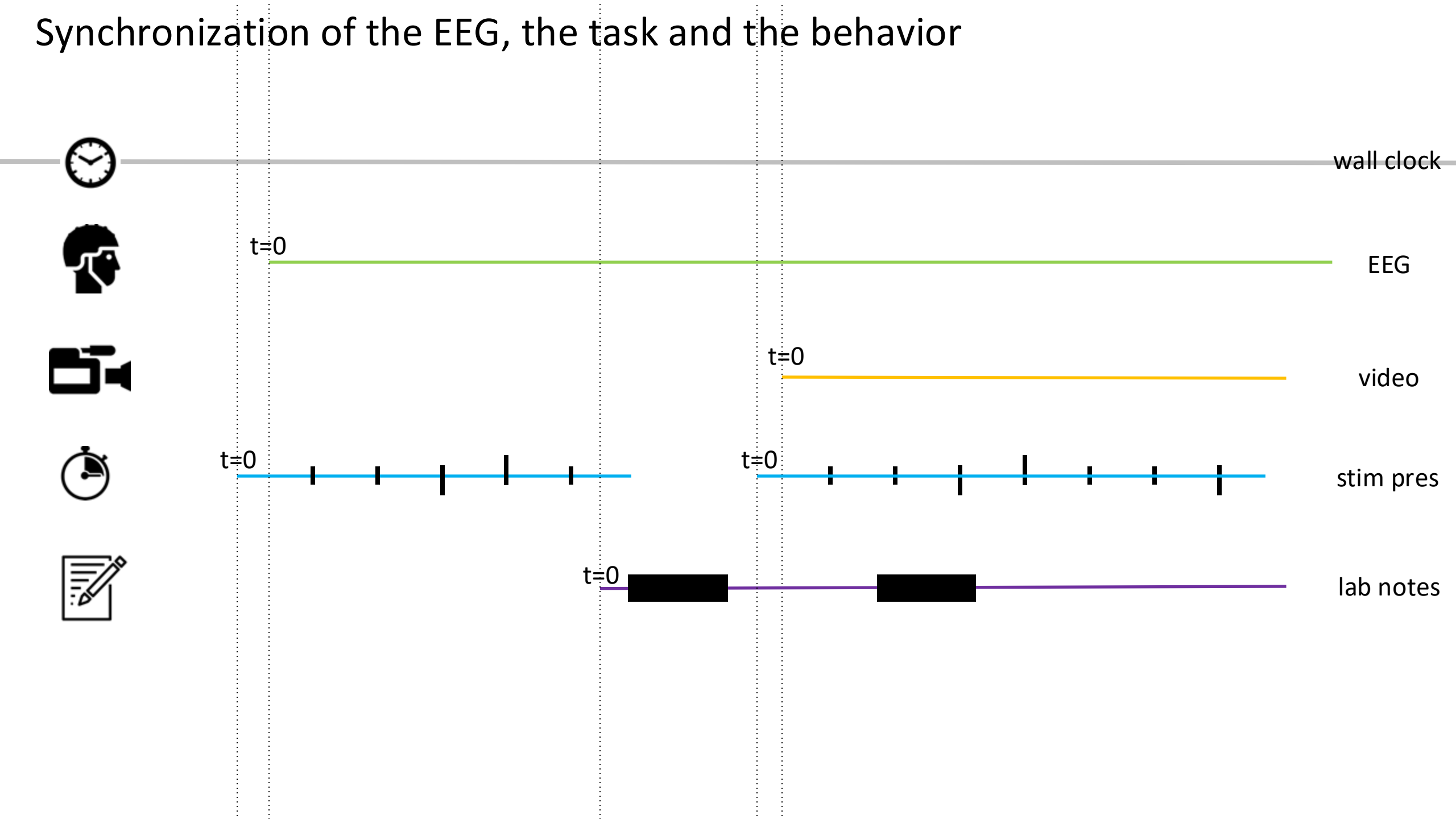
Synchronization of the EEG, the task and the behavior



Synchronization of the EEG, the task and the behavior



Synchronization of the EEG, the task and the behavior



Preprocessing - synchronization

Most of the time we have triggers (or events, or markers) in the EEG recording.

If those triggers “tell the whole story”, then we do not have to look at the stimulus presentation log files. Sometimes the stimulus presentation log files contain information that is not coded in the triggers.

Most of the time we do not record detailed behavior, since the participant can be assumed to behave well. Except for young children, or when the task is complex.

Sometimes we record video, or we record eye gaze with an “eye tracker”.

EEG preprocessing

Visual inspection

Bad channel detection

Filtering

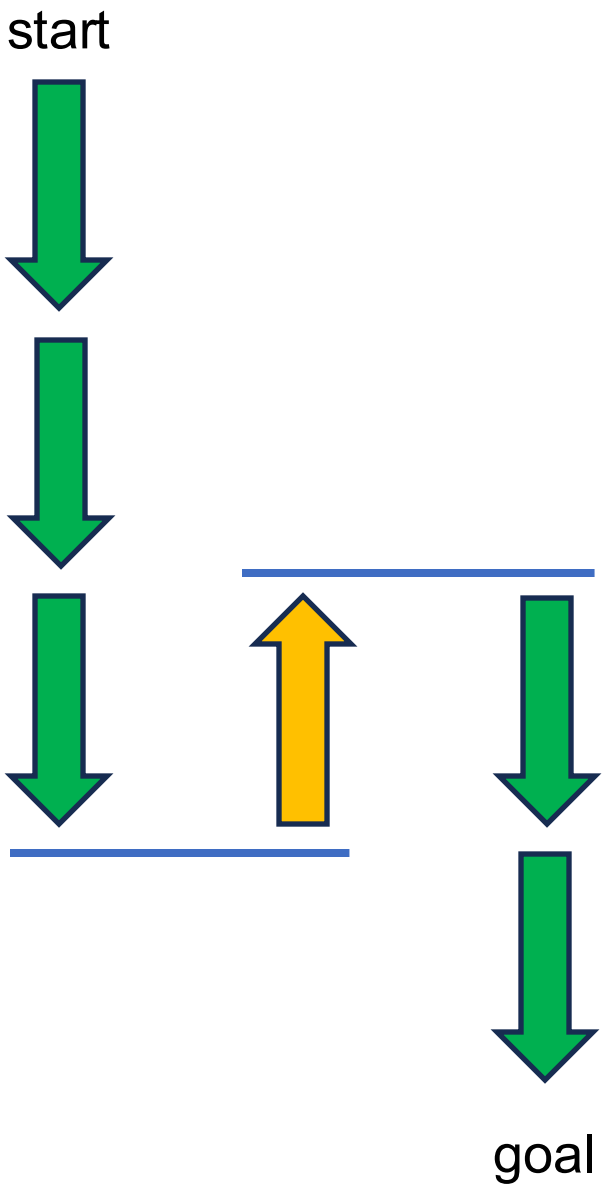
Re-referencing

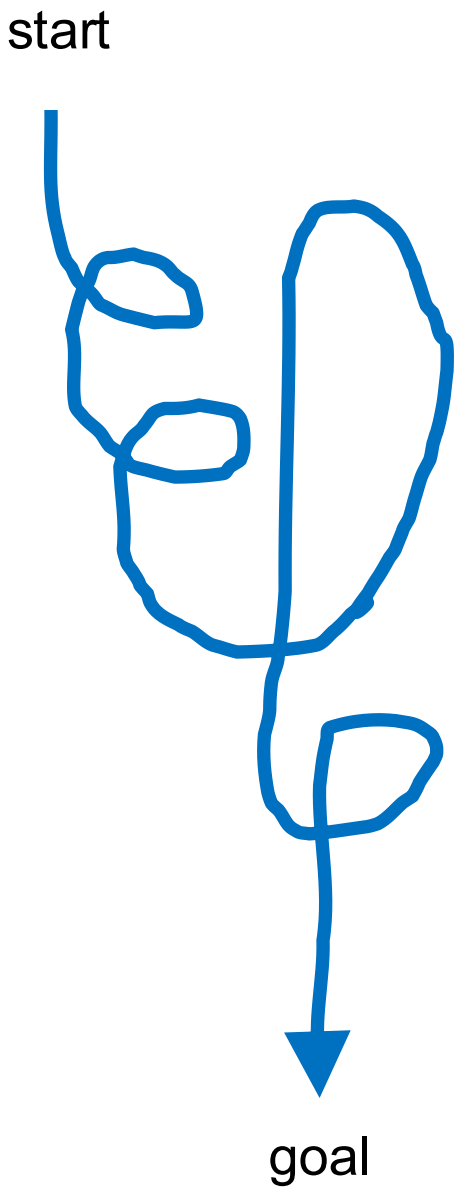
Independent component analysis

Segmenting

Confounds

Go through multiple iterations of EEG preprocessing





Visual inspection

Loading EEG data into memory

Check that it looks ok - not that it is clean, but that the data exists

Loading the events into memory

Check that they existin the correct amount and with the expected values

Sometimes we first do a “behavioral analysis”. If the subject did not do the task properly, there is no reason to spend efforts on the EEG data.

If the data is good enough, there might still be some channels that are bad.

Bad channel detection

An electrode that is not properly attached to the scalp will show a noisy or no signal.

- You can remove them from the analysis
- Later you may want to interpolate them again

Electrodes close to the edge of the cap are more often affected, since the cap might not fit optimally and there is less pressure to the scalp.

Electrodes at places where there is a lot of hair are more often affected, because the electrolyte (paste, gel, or salt water) does not make good contact.

A channel may be bad throughout the whole recording, or for a short time.

During movements all channels may be bad for a short time -> remove the segment.

Filtering

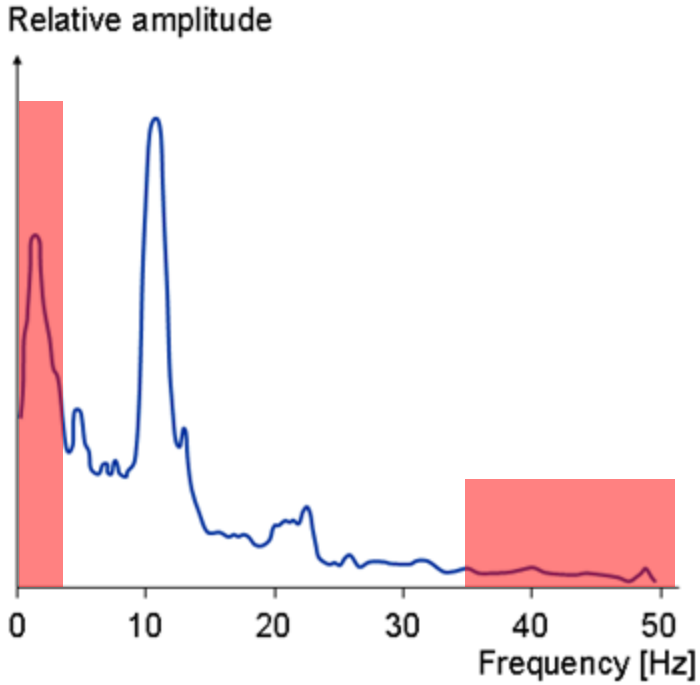
Why **do** you want to filter?

- Low and high frequencies in the signal are often not physiological
- Low frequency signals are often artifacts from sweating, movement, or general instability of the electrode-skin contact
- High frequencies are often noise, can be muscle tension or EMG
- Line frequency at 50 or 60 Hz is not interesting

Why do you **not** want to filter?

- Artifacts are often not only on the low or high frequencies, by filtering you make it more difficult to recognize the artifact
- Filtering can introduce artifacts at the segment edges or around sudden jumps in the signal
- Some later analyses do not benefit from filtering, like frequency analysis
- Filtering can cause ERP component peak latencies to shift in time

Filtering



Low-pass filter (removes the *high* frequencies)

High-pass filter (removes the *low* frequencies)

Band-pass filter (removes both low and high frequencies)

- This combines a low- and high-pass filter

Band-stop or notch filter (removes the 50Hz or 60Hz line noise)

- The opposite of a band-pass filter

Filtering does not completely remove the frequency range, it attenuates (dampens, or weakens) the specific frequency range

Widmann et al. (2015) Digital filter design for electrophysiological data - a practical approach. J Neurosci Methods. https://doi.org/10.1016/j.jneumeth.2014.08.002

Filtering

Filtering

The attenuation means that the signal gets weaker

- The pass-band remains at the original amplitude (1x amplification)
- The stop band gets passed at a weaker amplitude (for example 0.001x amplification)

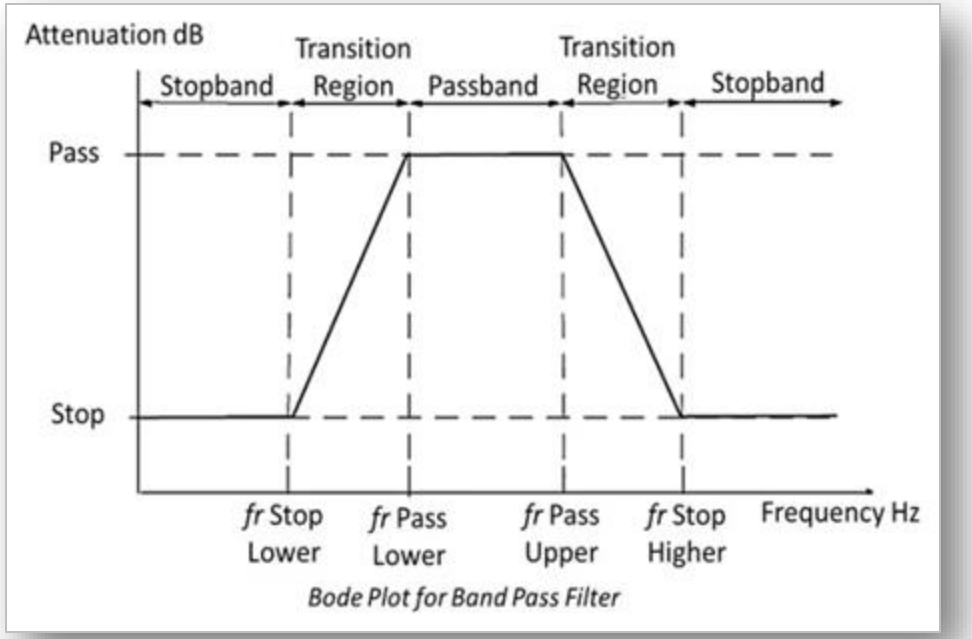
The attenuation in the transition region varies, and phase shifts are introduced

The filter **type** and **order** determine how effective the filter is

The attenuation is often expressed in decibels (dB)

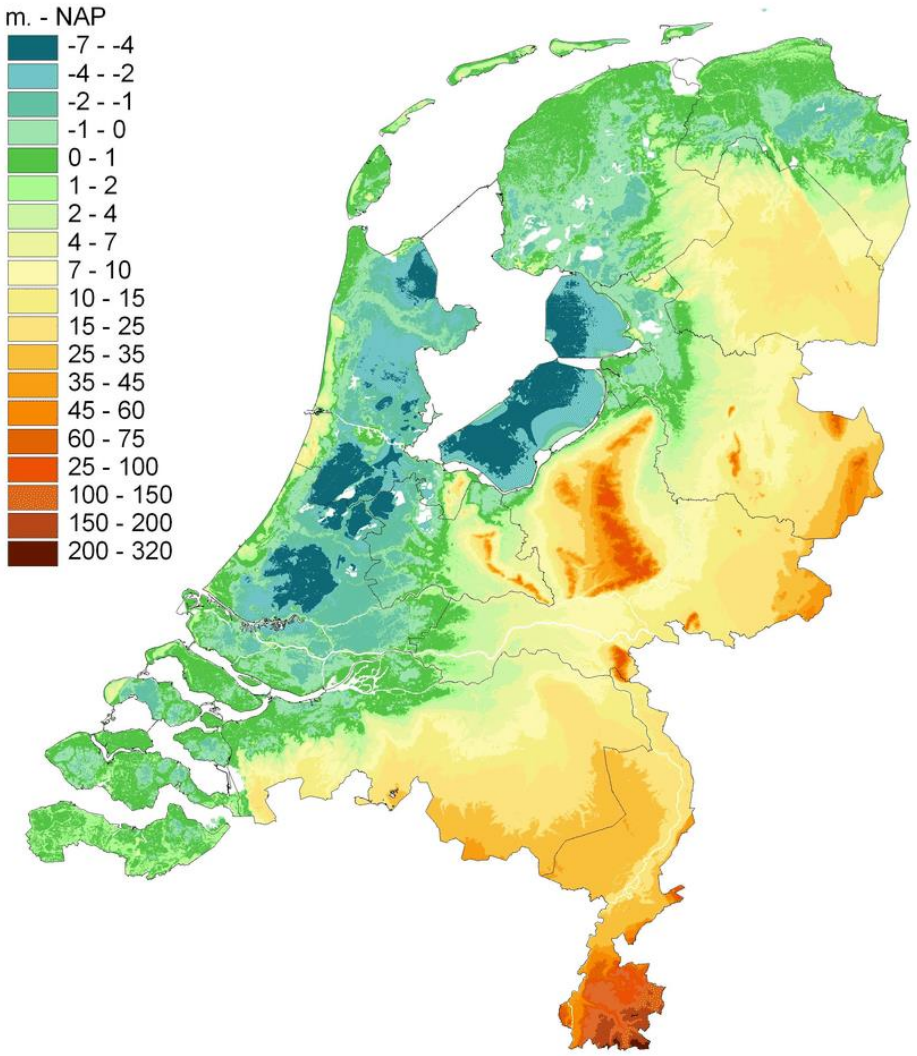
- 0 dB = 1 x attenuation
- 3 dB = ½ x attenuation in amplitude
- 10 dB = 10 x attenuation in amplitude

Note that the attenuation is 2x larger for power than for amplitude



Bode Plot for Band Pass Filter

Re-referencing



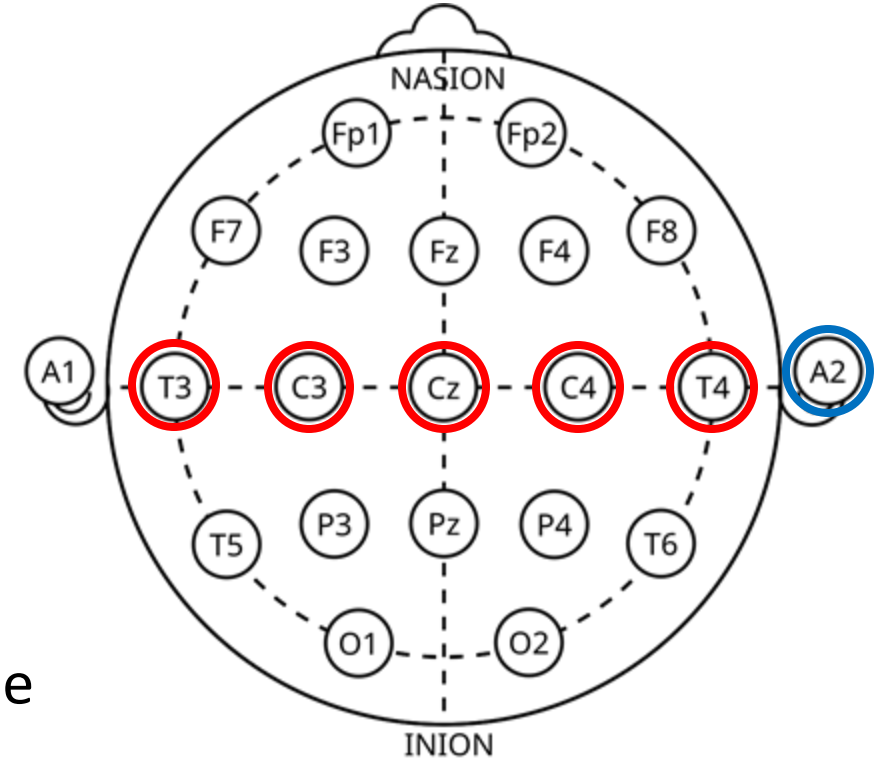
Re-referencing

Each EEG channel is the voltage difference between the electrode of interest and the reference. During recording the reference is “common” to all other eelctrodes.

For example

channel Cz = Cz - A2 (right earlobe)
channel C3 = C3 - A2
channel C4 = C4 - A2
channel T5 = T5 - A2
channel T6 = T6 - A2

Note that the electrodes on the left (odd numbers) are further away from the A2 reference and hence have larger signals.



Re-referencing

Why **do** you want to re-reference?

- The choice of the reference influences the signal on the channel
- The reference during the recording is often not the one that gives the optimal signal
- It can be spatially biasssed
- By choosing your own reference you can optimize the signal
- The reference itself may contain some residual noise (for example with Biosemi CMS/DRL scheme)

Why do you **not** want to re-reference?

- The reference electrode is usually placed such that it has good contact (behind the ear or at FCz or Cz, where there is little hair)
- Another reference channel (or the average of channels) may have a bad signal, which then spreads to all channels

Re-referencing – linked ears

Subtract from each EEG channel the average voltage of the two ears. The right ear A2 is (by definition) zero in the recording, so you basically subtract ½ times A1.

channel Cz = Cz – (A1+A2)/2

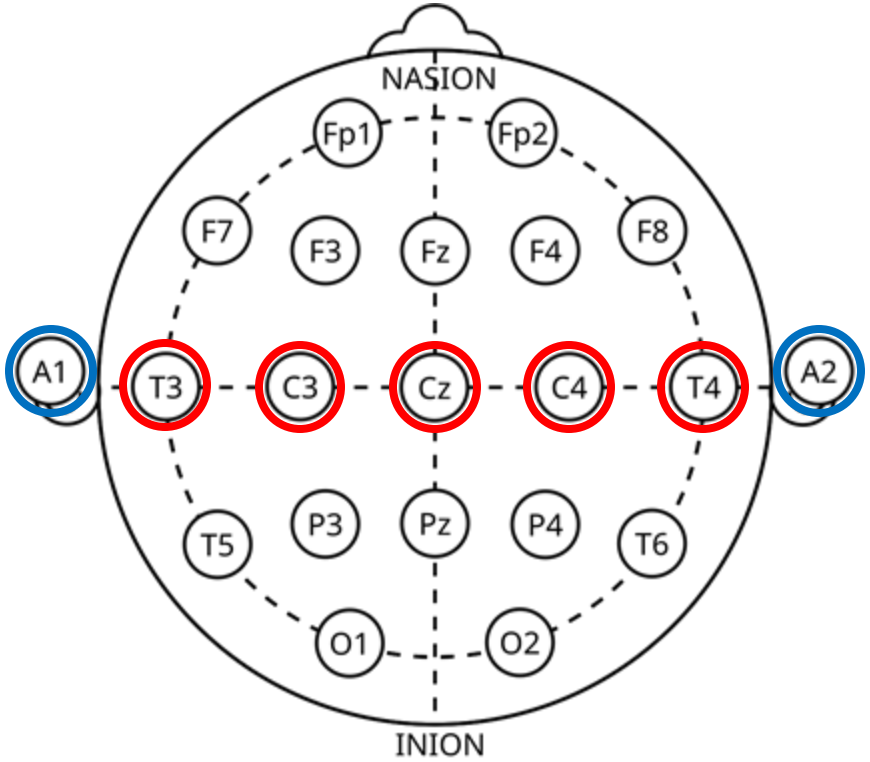
channel C3 = C3 - (A1+A2)/2

channel C4 = C4 - (A1+A2)/2

channel T5 = T5 - (A1+A2)/2

channel T6 = T6 - (A1+A2)/2

The reference electrodes are now (on average) symmetric



Re-referencing – average reference

Subtract from each EEG channel is the average voltage over ALL channels.

channel Cz = Cz – avg

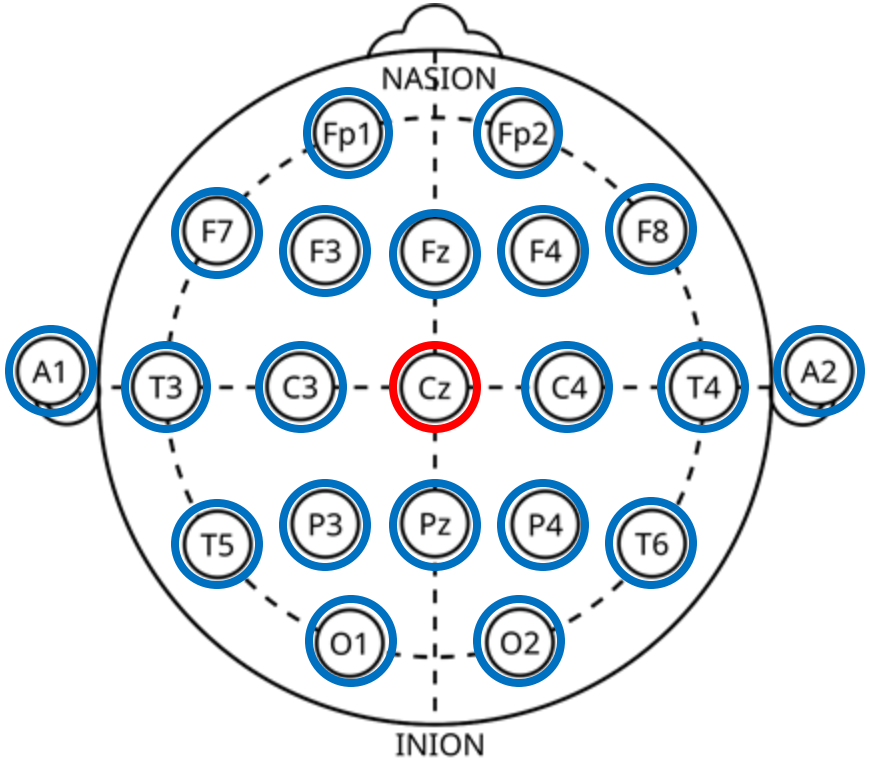
channel C3 = C3 - avg

channel C4 = C4 - avg

channel T5 = T5 - avg

channel T6 = T6 - avg

The reference electrodes are now (on average) symmetric



Re-referencing – the implicit reference

During recording we do not store the voltage of the reference channel, as that is zero anyway (by definition).

After recording we can add the reference electrode as a channel, and give it the value zero.

Since the reference electrode is not explicltly part of the recording as the data is stored on disk, we refer to it as the “implicit reference”.

After adding the implicit reference as a channel with all zeros, it is easier to do the re-referencing as just expolained.

Independent Component Analysis (ICA)

ICA estimates spatio-temporal components from the channel-level data that are statistically maximally independent.

Independent ICA components are more dipolar and hence physiologically more consistent with a small cortical patch.

ICA is also very useful to identify eye movements, blinks, heart, muscle and some other artifacts.

Prior to this we need to remove the atypical artifacts. For example movement artifacts during the breaks, or bad channels.

Delorme, Oostenveld, et al (2012) Independent EEG sources are dipolar.
PLoS ONE https://doi.org/10.1371/journal.pone.0030135

Independent Component Analysis (ICA)

Why **do** you want to use ICA?

- Remove stereotypical artifacts that are present throughout the recording
- To interpret the ICA “source” component topographies and timeseries

Why do you **not** want to use ICA?

- If you do not have enough channels
- If you do not have long enough recordings
- If the components are not spatially stationary, i.e., they change over time
- If electrodes have moved
- If the ICA decomposition fails because the data is too noisy
- …
- If the artifacts are confounded with the experimental task

Segmenting – cutting the data into pieces

Segmenting may not be the last thing to do

- If you start the recording while the participant is still moving, then the first part of the recording is rubbish anway
- Resting state – eyes closed and open - If you start the recording and only then tell the participant what to do.
- If your participant makes movements in the breaks.
- If your participant blinks in-between trials.

Sometimes you want to consider only the EEG data that you are interested in, then you can segment prior to the other preprocessing.

Sometimes you may want to have long segments, other times only the trials of interest. But watch out for filter artifacts at the edges.

Whether you segment early or late in the preprocessing chain is a matter of choice.

Segmenting – cutting the data into pieces

Following segmenting you often want to do a baseline correction.

There is no systematic task-related activity just prior to the stimulus.

The pre-stimulus interval is used to compute the mean voltage (for each channel), which is then subtracted from the whole timecourse.

Low-frequency fluctuations (over many seconds) are often related to electrode instabilities, baseline correction removes this.

Confound

Confounds are unexpected factors that unintentionally may explain your results.

- For example frontal “EEG activity” upon visual stimulation – which may be due to the participant blinking upon every stimulus.
- For example high-frequency gamma-band activity during a cognitively demanding task – which may be due to there being more muscle tension.

Behavioral confounds can show up as EOG, EMG, ECG or movement artifacts.

Removing these artifacts might hide the effect of the confounds but not totally remove it.

Therefore you should check for experimental confounds in your artifacts.

Avoid troubles with preprocessing

Think about what you can fix in the ***acquisition*** rather than trying to fix it in the ***analysis***.

Spend time in the lab and record more data rather than spending time on trying to fix bad data. Getting one good recording might take you a few hours (in total), trying to fix a bad recording might take you days.

Use pilot expetiments, pilot recordings, and pilot analyses.

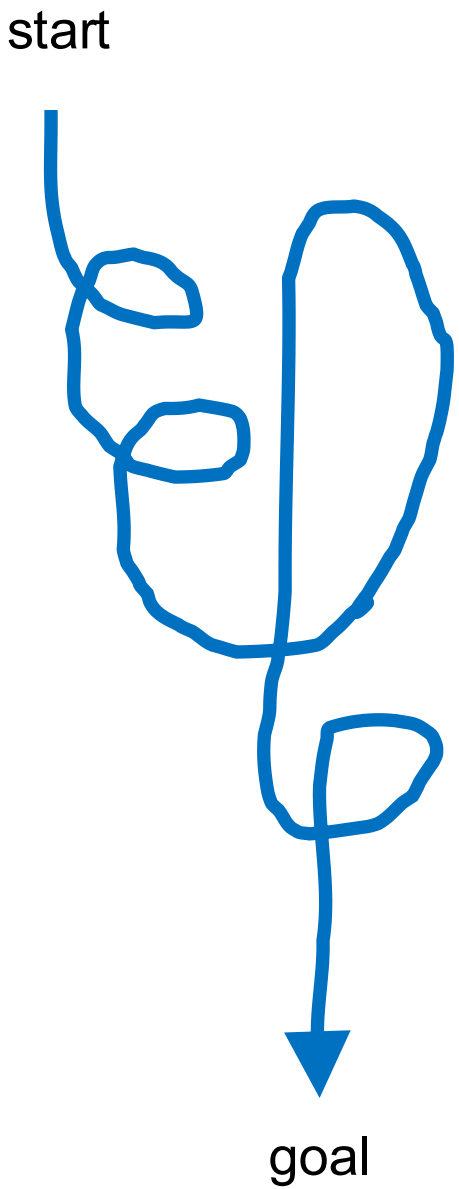
Go through multiple iterations of EEG preprpocessing

Write your analysis as a script

Taking a step back is easy and fast

Store intermediate results

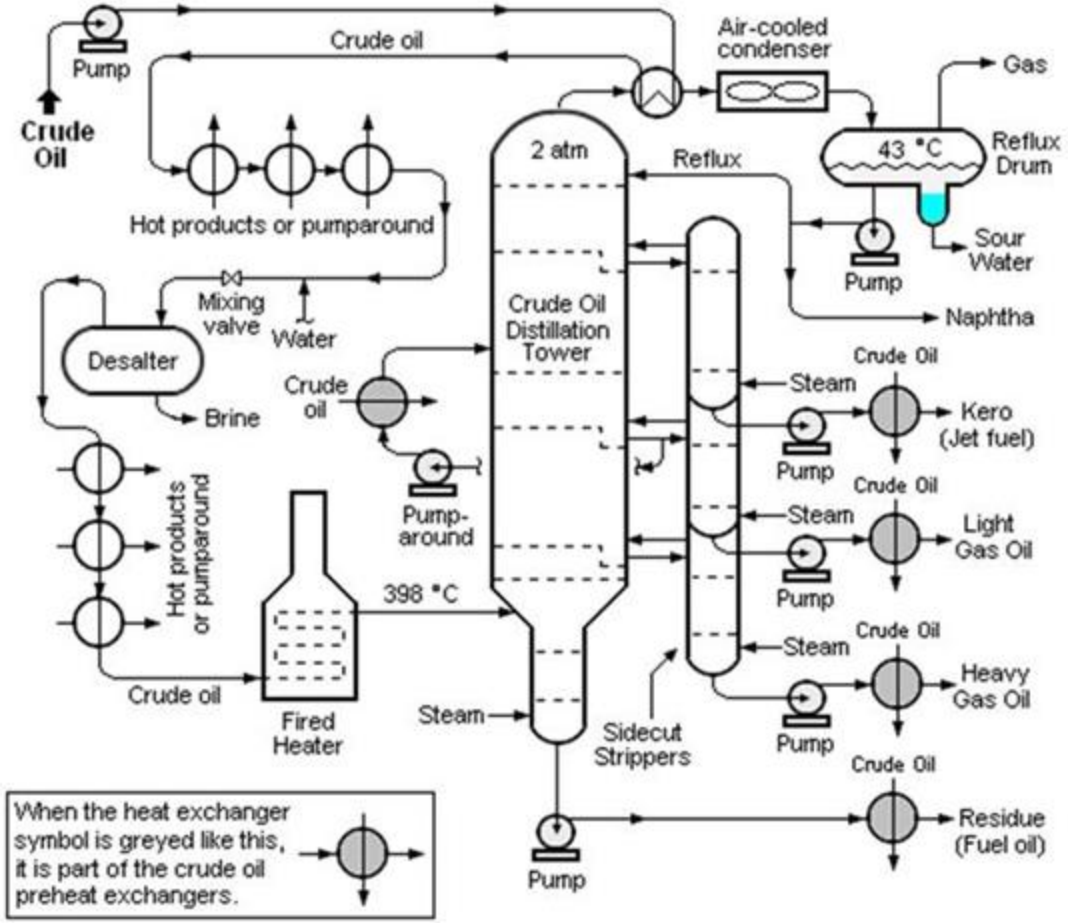
Store time-consuming steps



Summary

Sequence of steps

- Synchronization of all data
- Visual inspection
- Bad channel detection
- Filtering
- Re-referencing
- Independent component analysis
- Segmenting
- Confounds



We often think of preprocessing as a pipeline. There is not a single optimal pipeline with an optimal ordering of the steps, as the EEG quality depends on the interaction between the task/experiment and what the participant is doing.

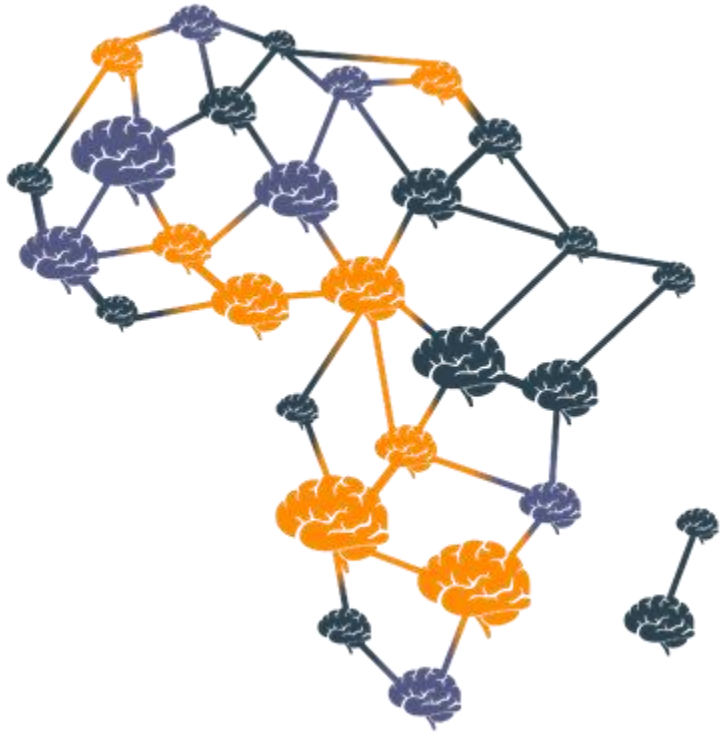
Delorme (2023). EEG is better left alone. Sci Reports https://doi.org/10.1038/s41598-023-27528-0

Preprocessing of EEG data

Robert Oostenveld

Mikkel C. Vinding

9-14 June 2025
Port Harcourt, Nigeria



African Brain Data Network