

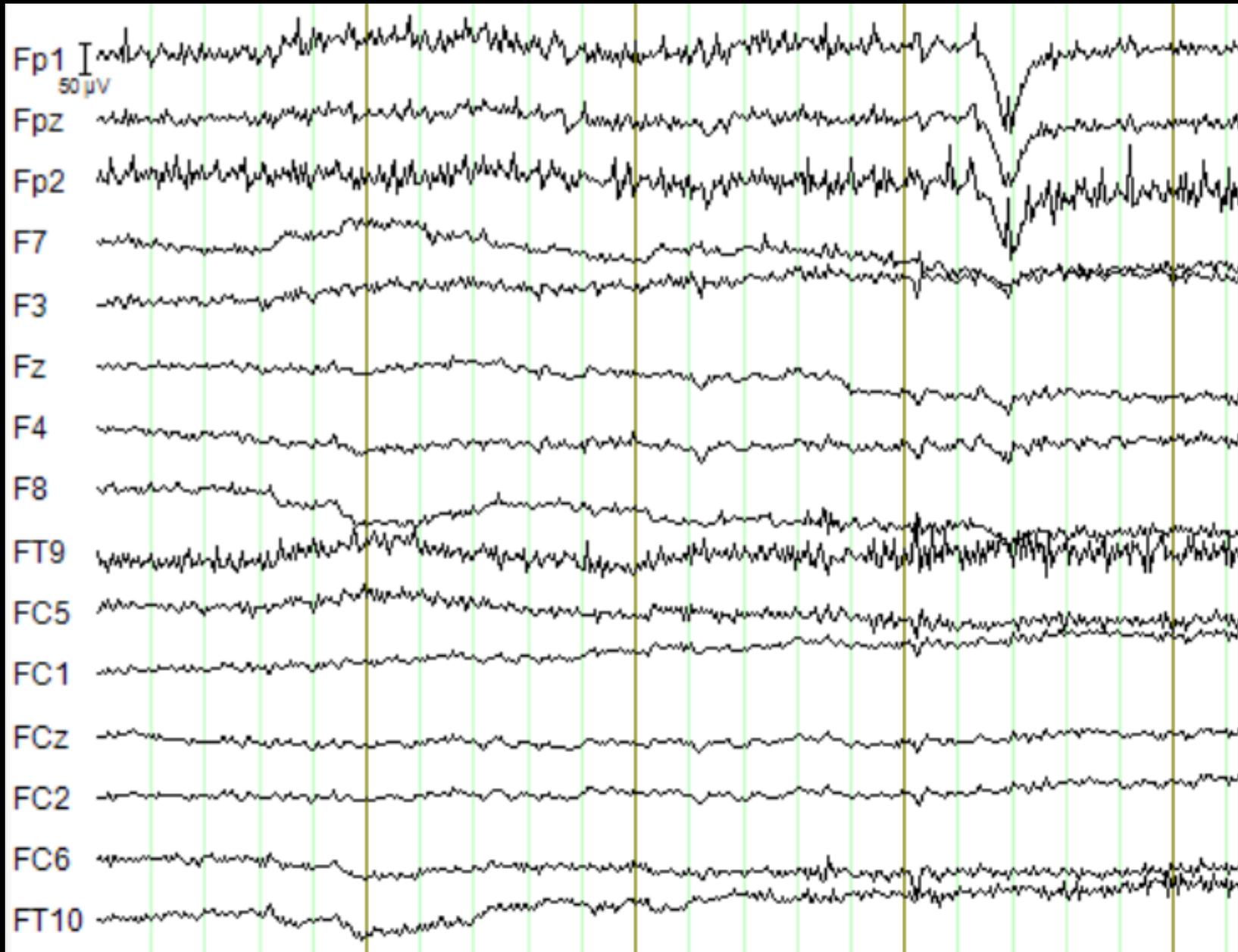
Session 3  
Data Preprocessing I

# Our Goal

In this sessions we will do a full ERP analysis of the data for a single subject.

# Loading Data

# What is in an EEGLAB Data Structure?

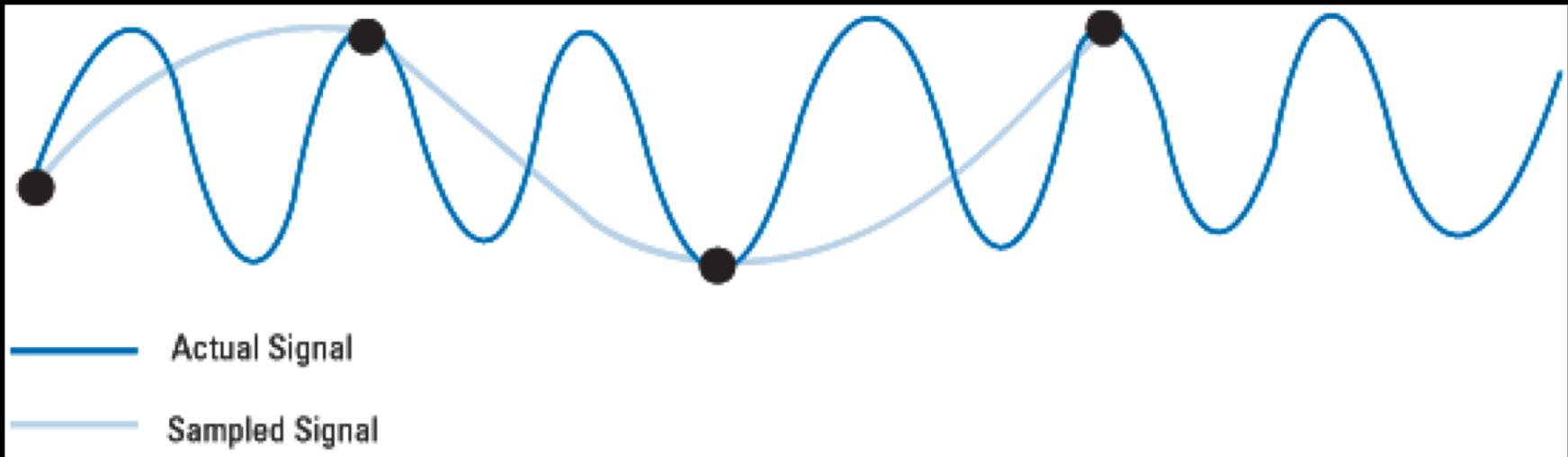


# The Tree (EEG)

1. Downsample
2. Remove Channels
3. New Reference
4. Filtering
5. Segmentation
6. DC Detrend (optional)
7. Ocular Correction (optional)
8. Topographic Interpolation
9. Resegmentation?
10. Baseline Correction
11. Artifact Rejection
12. Resegmentation (into conditions)
13. Average
14. Difference Computation
15. Stats

# Downsampling

# Downsampling



# Downsampling Demos

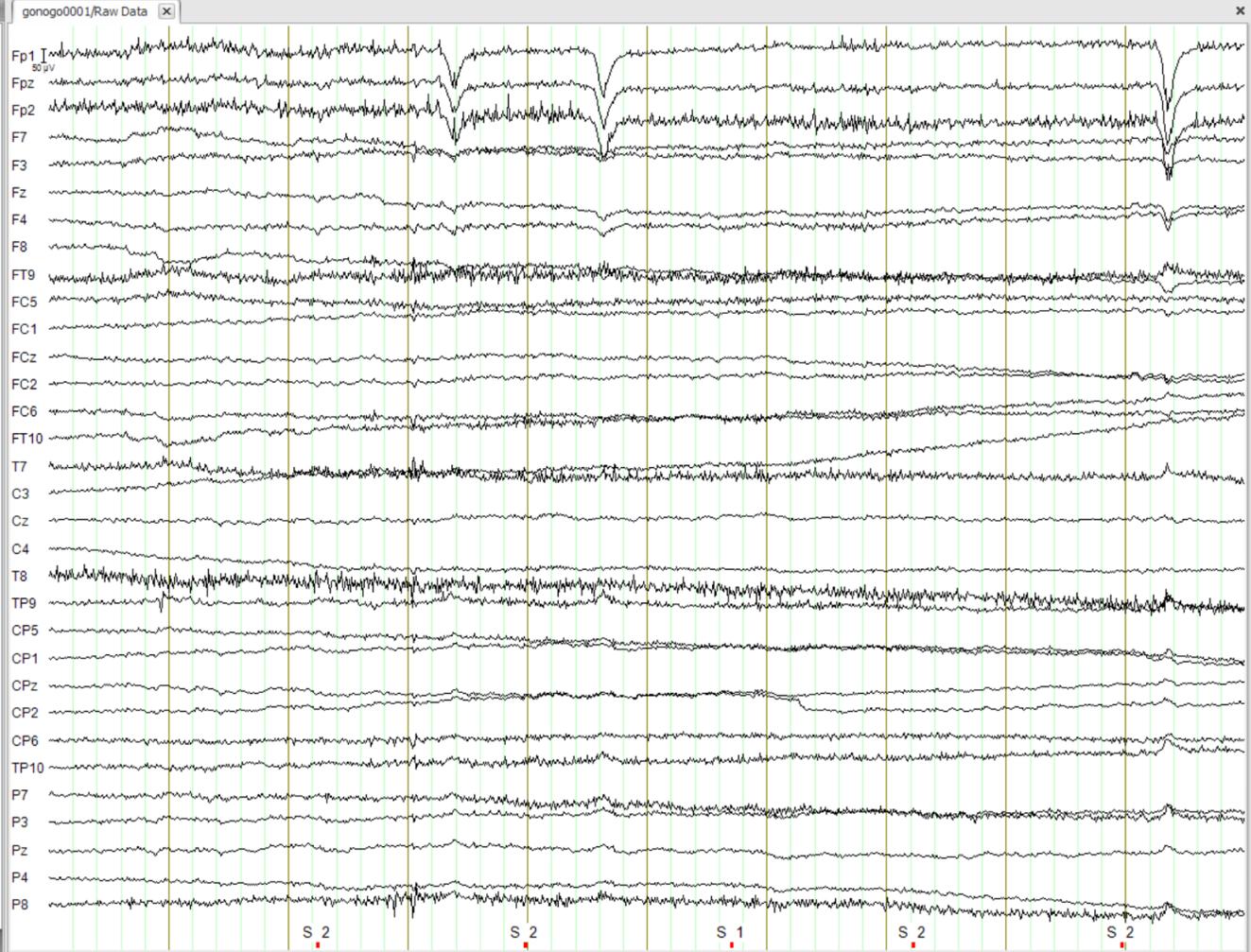
# Raw Data Inspection

Remove Channels

Primary

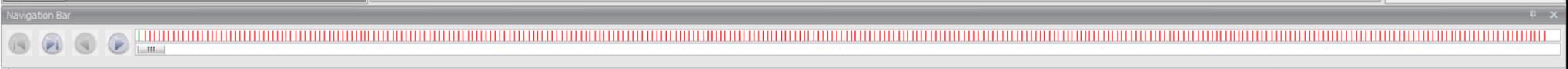
- gonogo0001
  - Raw Data
    - Topographic Interpolation
  - gonogo0002
  - gonogo0003
  - gonogo0004
  - gonogo0005
  - gonogo0006
  - gonogo0007
  - gonogo0008
  - gonogo0009
  - gonogo0010
  - gonogo0011
  - gonogo0012
  - gonogo0013
  - gonogo0014
  - gonogo0015

Primary Secondary History



View Toolbox

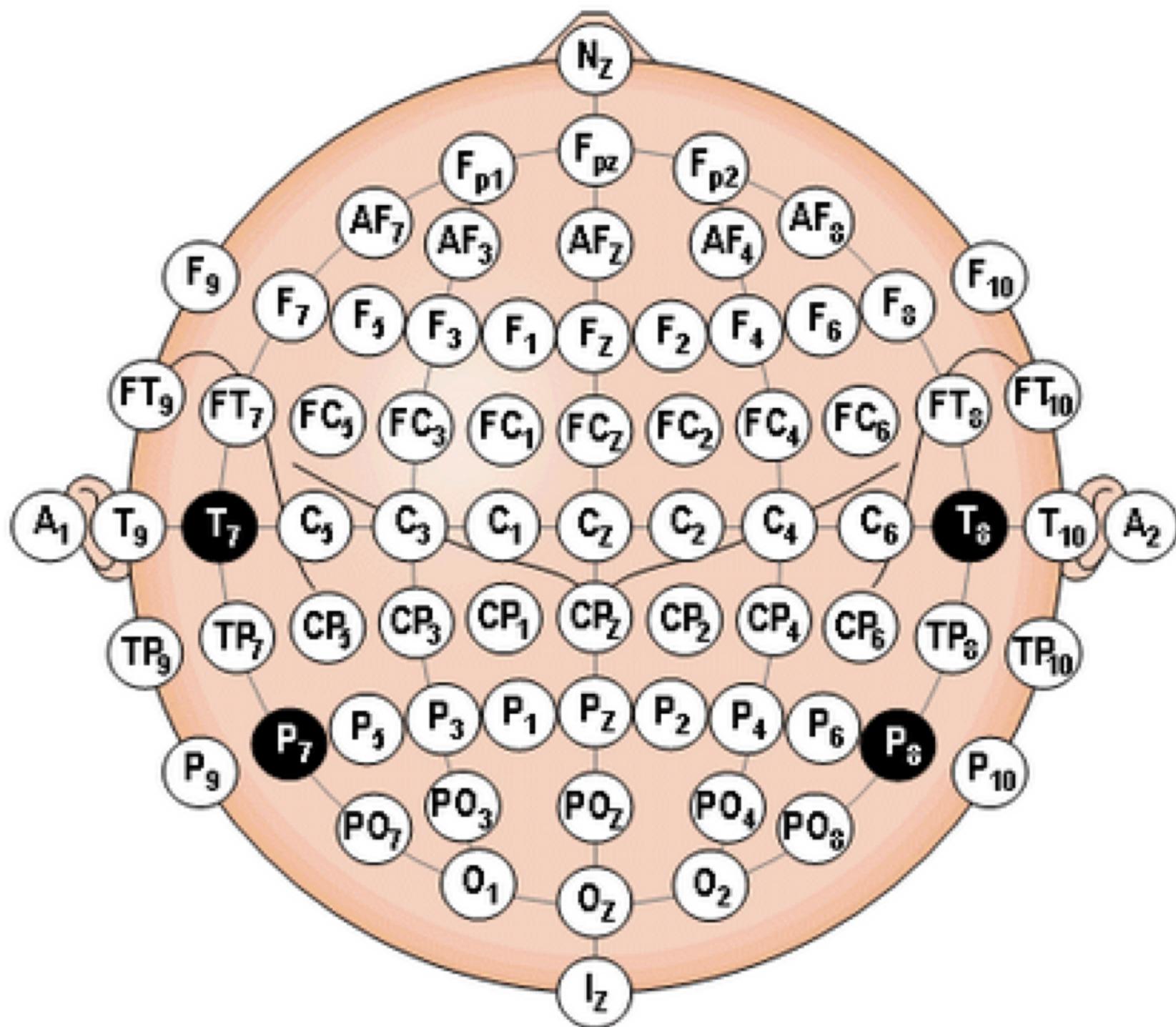
- Channel Selection
- Delta Tool
- Magnifier
- Map Legend
- Mapping
- Scaling Bar
- Text Box
- Value Graphics



# Issues with Removing Channels

1. Too many? (not needed = filtering)
2. Too few? (ICA)
3. We frequently "redo" the analysis and modify this.

# Rereferencing



# The Notion...

First of all, recall that a voltage reference needs to be supplied at the time of data recording.

Why, because a voltage is the potential difference between two points.

However, offline we may wish to re-reference the data again.

# The Important Point

Any voltage you see is the difference between an electrode site and another point...

THUS...

The choice of reference impacts the voltages you are observing

# But Why?

Because you cannot record a voltage at a single electrode site, by definition!

Thus, you need to have a reference, but that reference may not be good from an analysis perspective.

So, we re-reference typically during data analysis.

# The Answer?

The most common reference sites are the EARS and the MASTOIDS.

We use two, one on each side, to avoid a hemisphere bias.

Why not record with these? May not be best for actual recording but it is good for analysis.

# Average References

Not as ideal as one might think

- electrodes not covering the entire surface of the head
- differences in polarities – an average reference can make one site positive and another negative
- reduced effects

# Reference Demos

# Data Filtering

# Filtering is a Complex Business

Why do we do it?

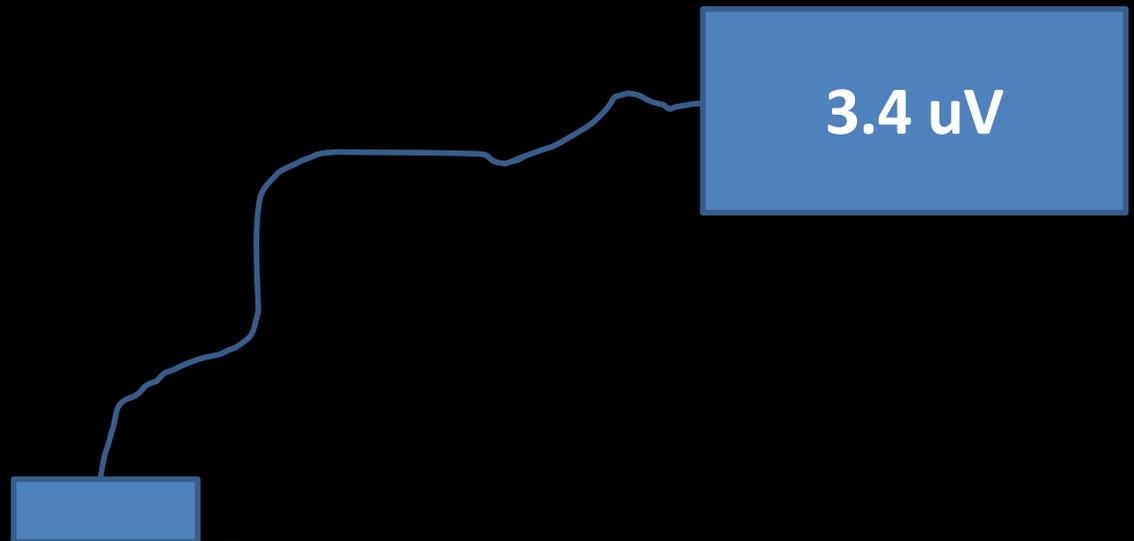
- reduce noise in the data
- prevent distortion in the data

The Problem

- filters can substantially distort your data

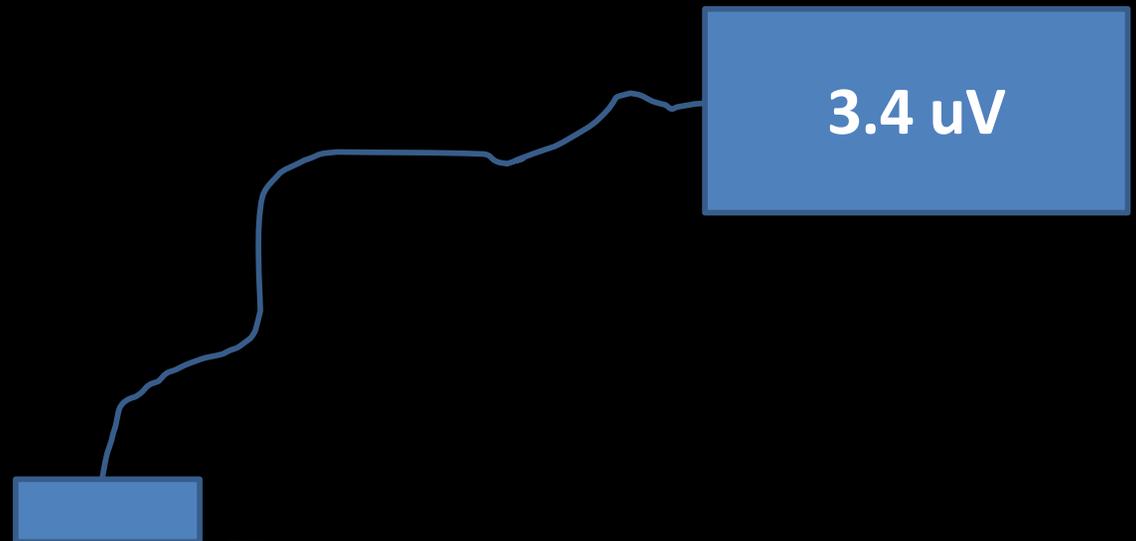
# Voltage as a Waveform

At any given point in time, an electrode is measuring a voltage, or potential difference, that is the net result of the underlying neural activity.



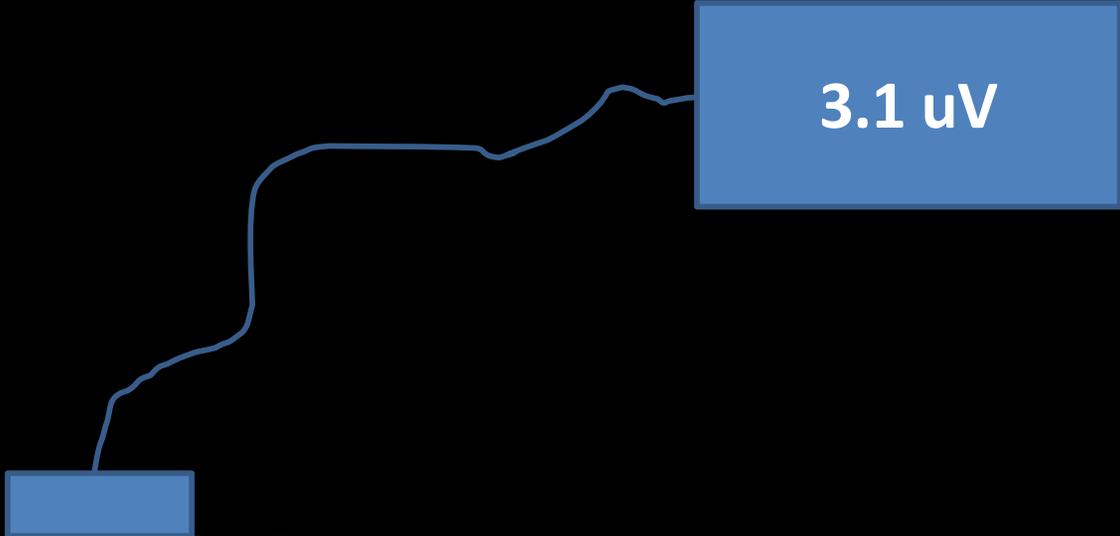
# Voltage as a Waveform

However, if you take sequential measurements you will have a series of voltages from the sensor.



# Voltage as a Waveform

However, if you take sequential measurements you will have a series of voltages from the sensor.

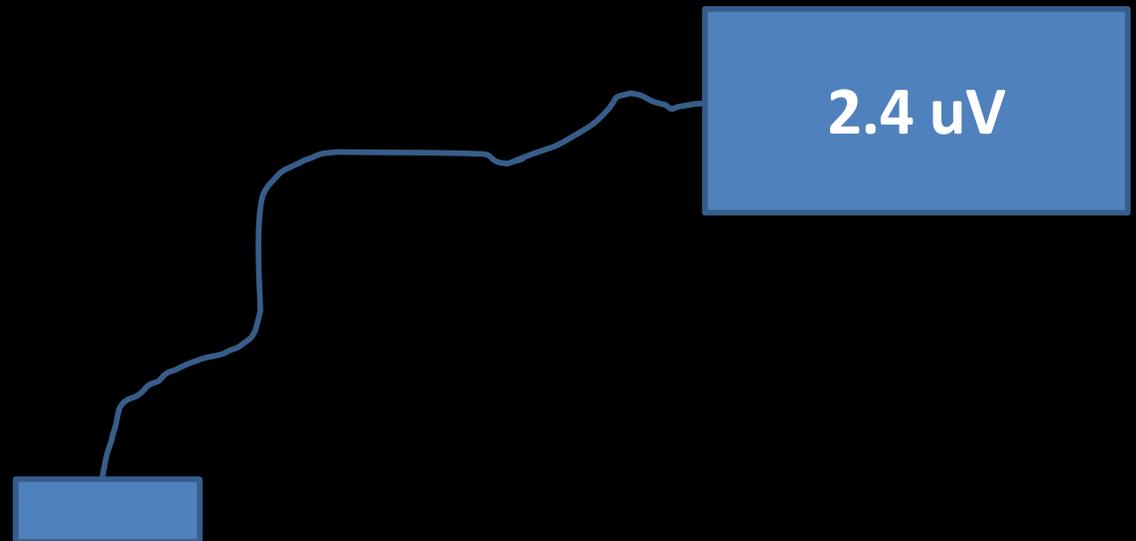


3.1  $\mu\text{V}$

The diagram illustrates a sensor (represented by a small blue rectangle) placed on a curved, light-colored surface. A blue line representing a voltage waveform starts at the sensor and rises in a step-like fashion, ending at a value of 3.1  $\mu\text{V}$ . A blue box on the right side of the waveform contains the text "3.1  $\mu\text{V}$ ".

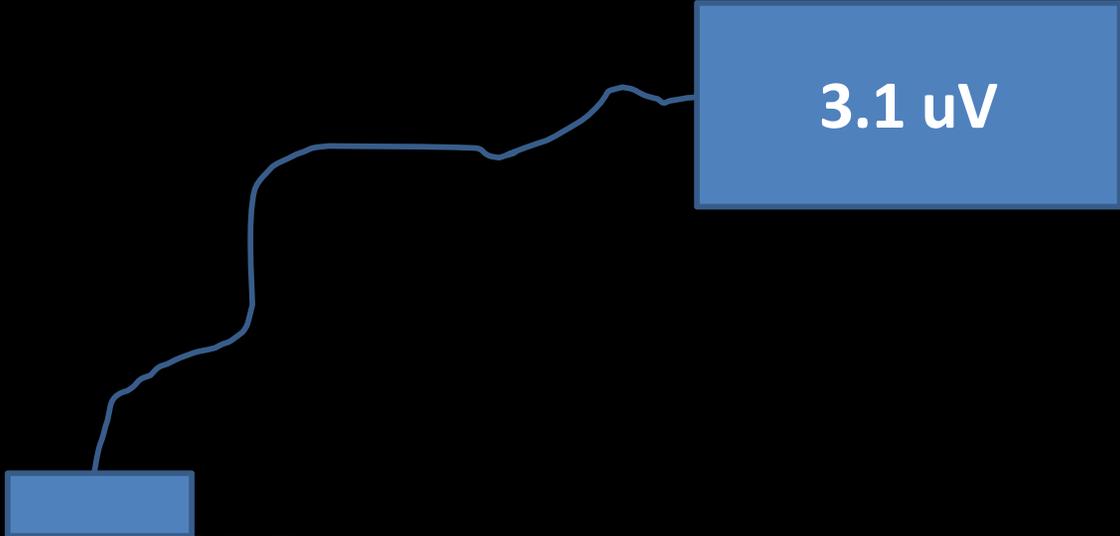
# Voltage as a Waveform

However, if you take sequential measurements you will have a series of voltages from the sensor.



# Voltage as a Waveform

However, if you take sequential measurements you will have a series of voltages from the sensor.

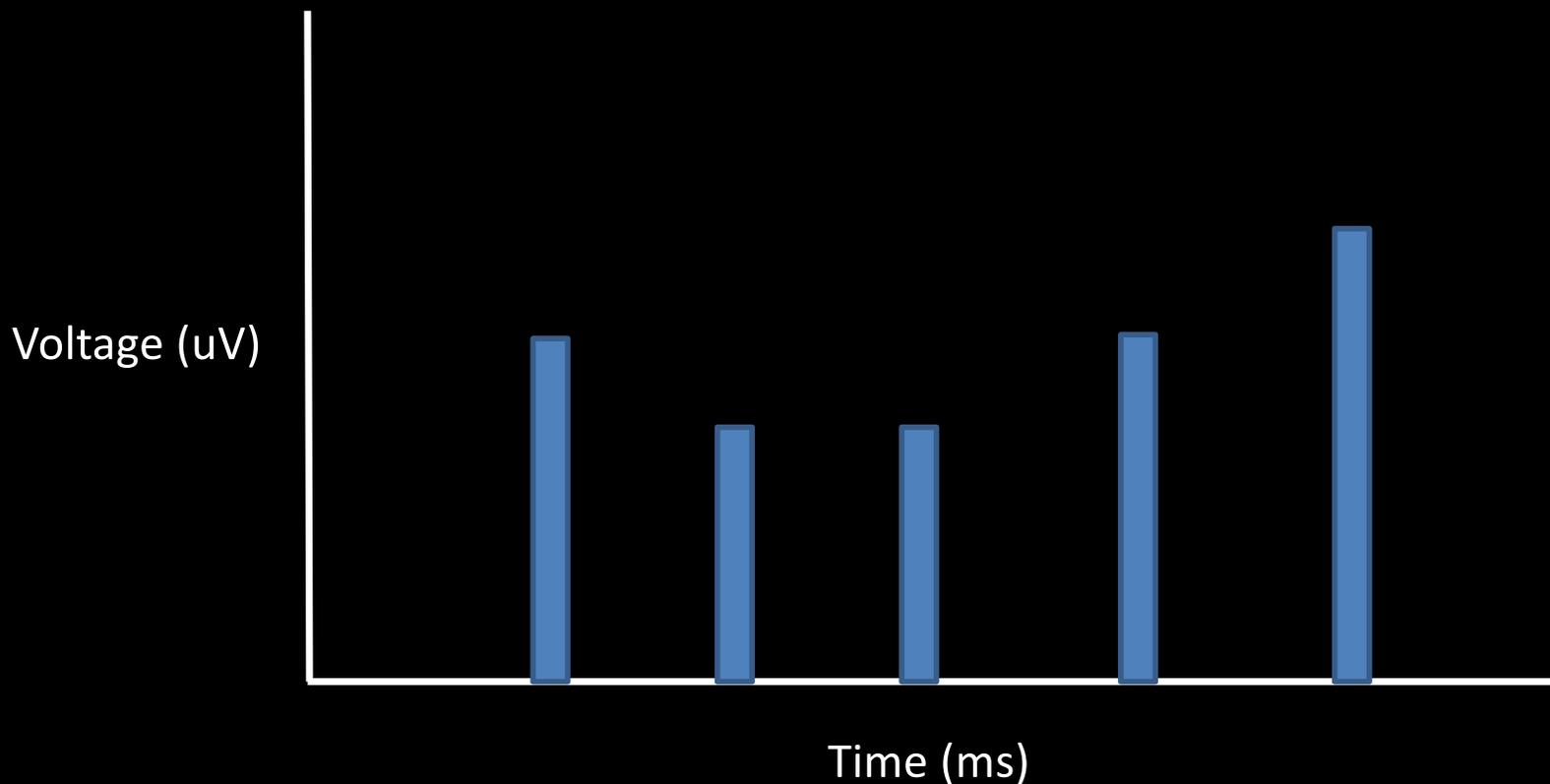


3.1  $\mu\text{V}$

The diagram illustrates a sensor (represented by a small blue rectangle) placed on a curved, light-colored surface. A blue line representing a voltage waveform starts at the sensor and rises in a step-like fashion, ending at a value of 3.1  $\mu\text{V}$ . A blue box on the right side of the waveform contains the text "3.1  $\mu\text{V}$ ".

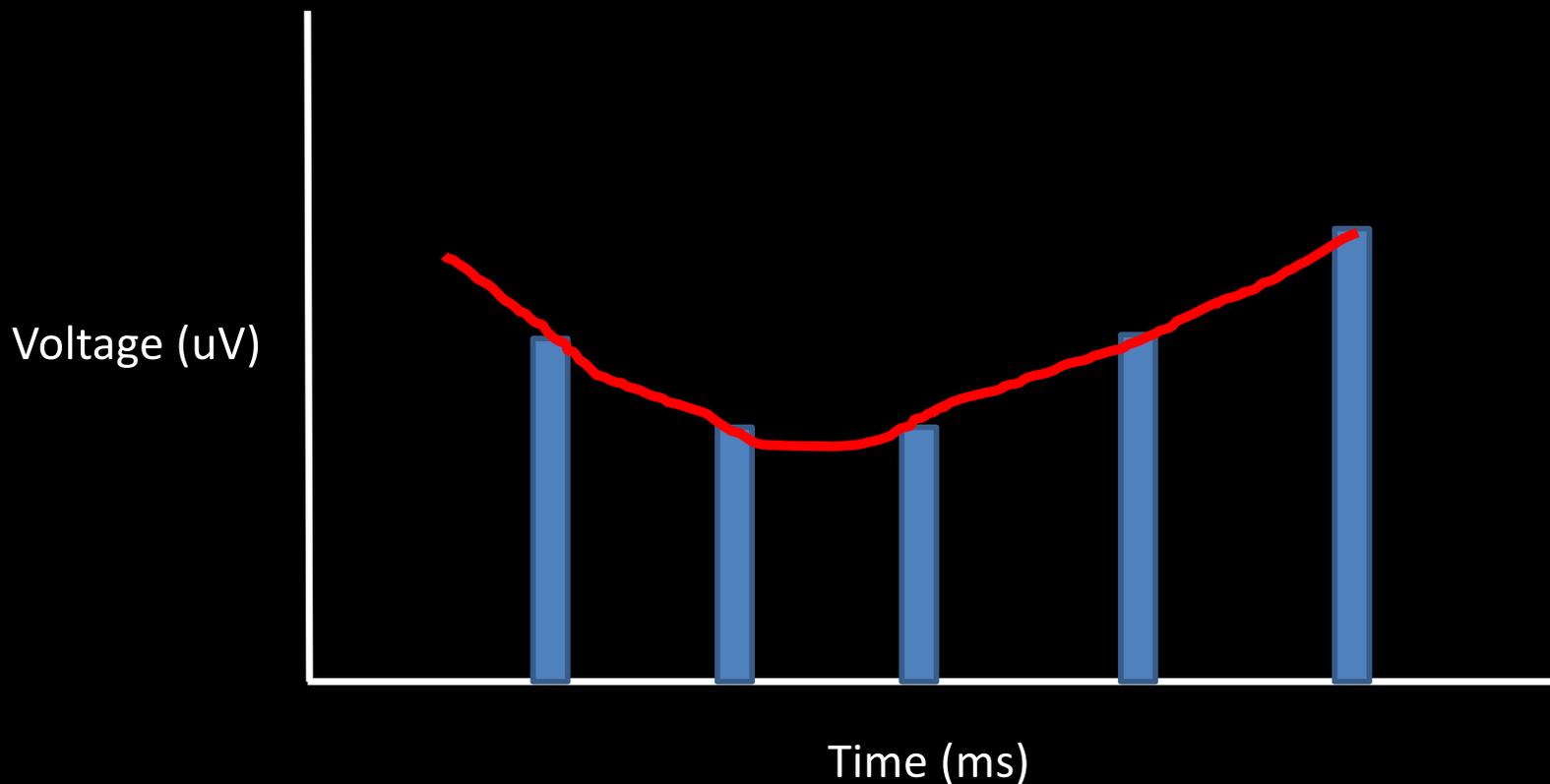
# Voltage as a Waveform

However, if you take sequential measurements you will have a series of voltages from the sensor.



# Voltage as a Waveform

However, if you take sequential measurements you will have a series of voltages from the sensor.



A quick note on sampling  
theory...

# The Nyquist Theorem

Recall what we said yesterday.

To capture your EEG signal you need to have a sampling rate such that its at least twice the highest frequency of the signal.

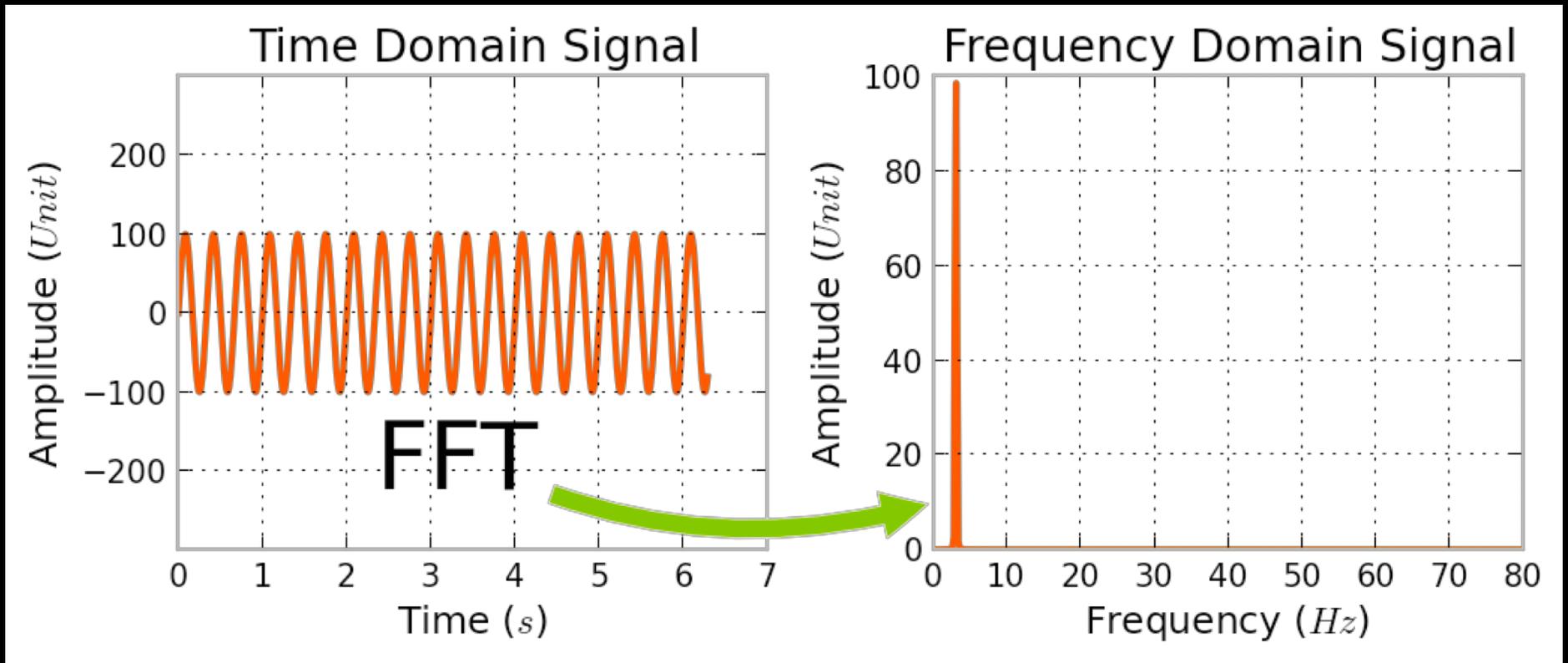
Typically, the signals we want to look at are not greater than 100 Hz (for ERPs 40 Hz).

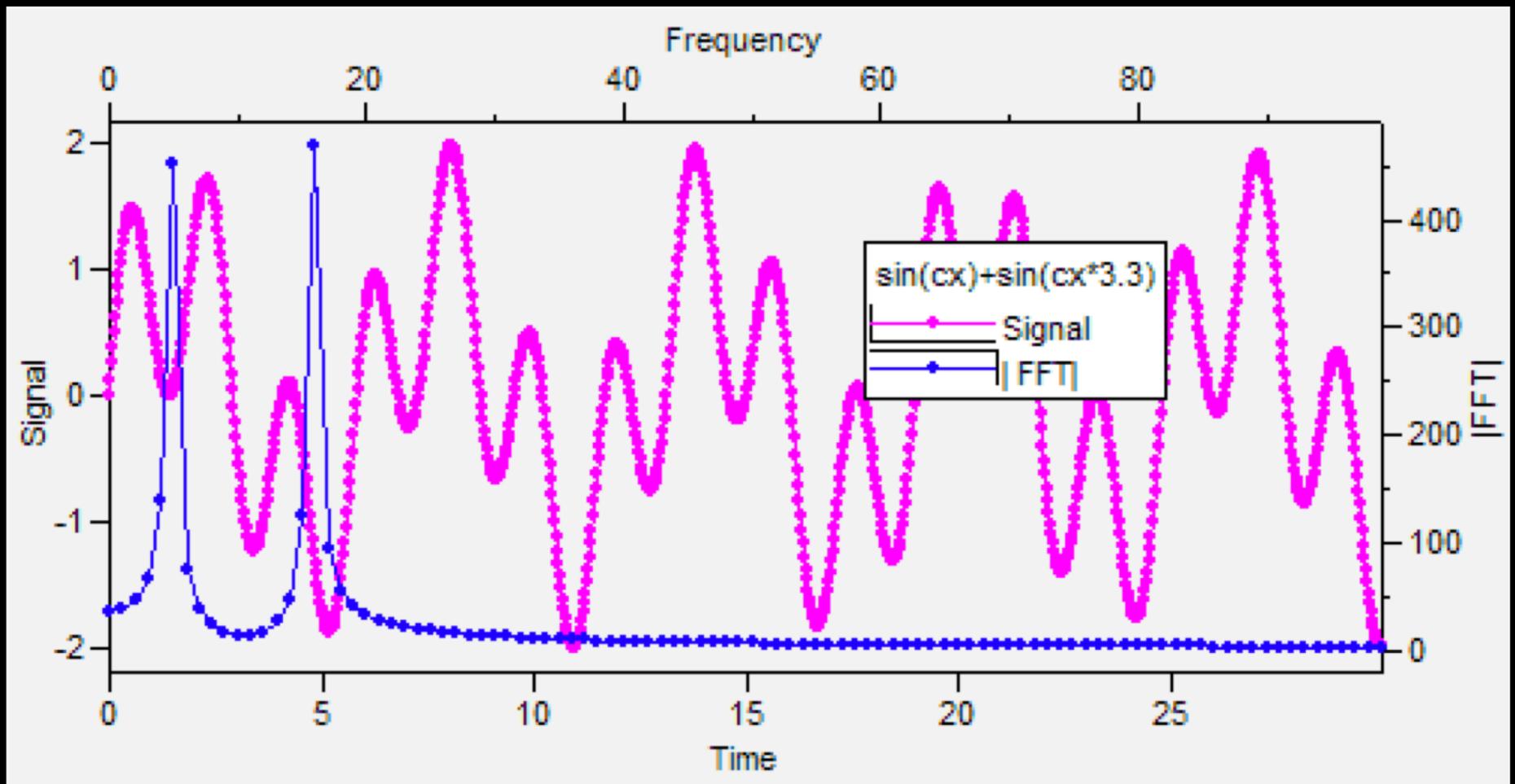
So you can think of a series of voltages as a waveform, so what?

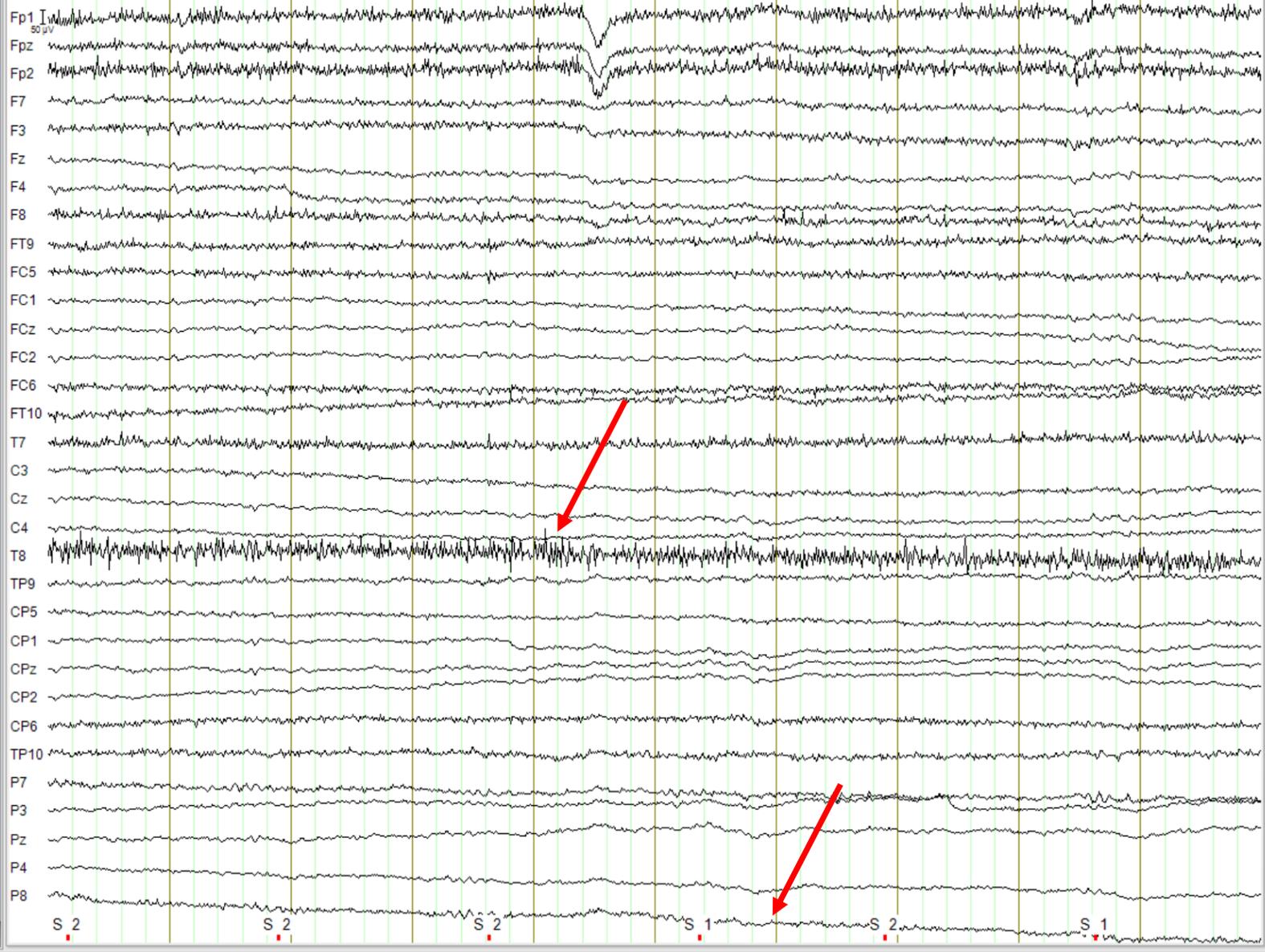
# The Fourier Theorem

Simply put, any complex waveform can be decomposed into a series of sine waves, that when added back together, reconstruct the original waveform.

This process can be used to subtract out unwanted frequencies of data.







Channel Selection

Delta Tool

Magnifier

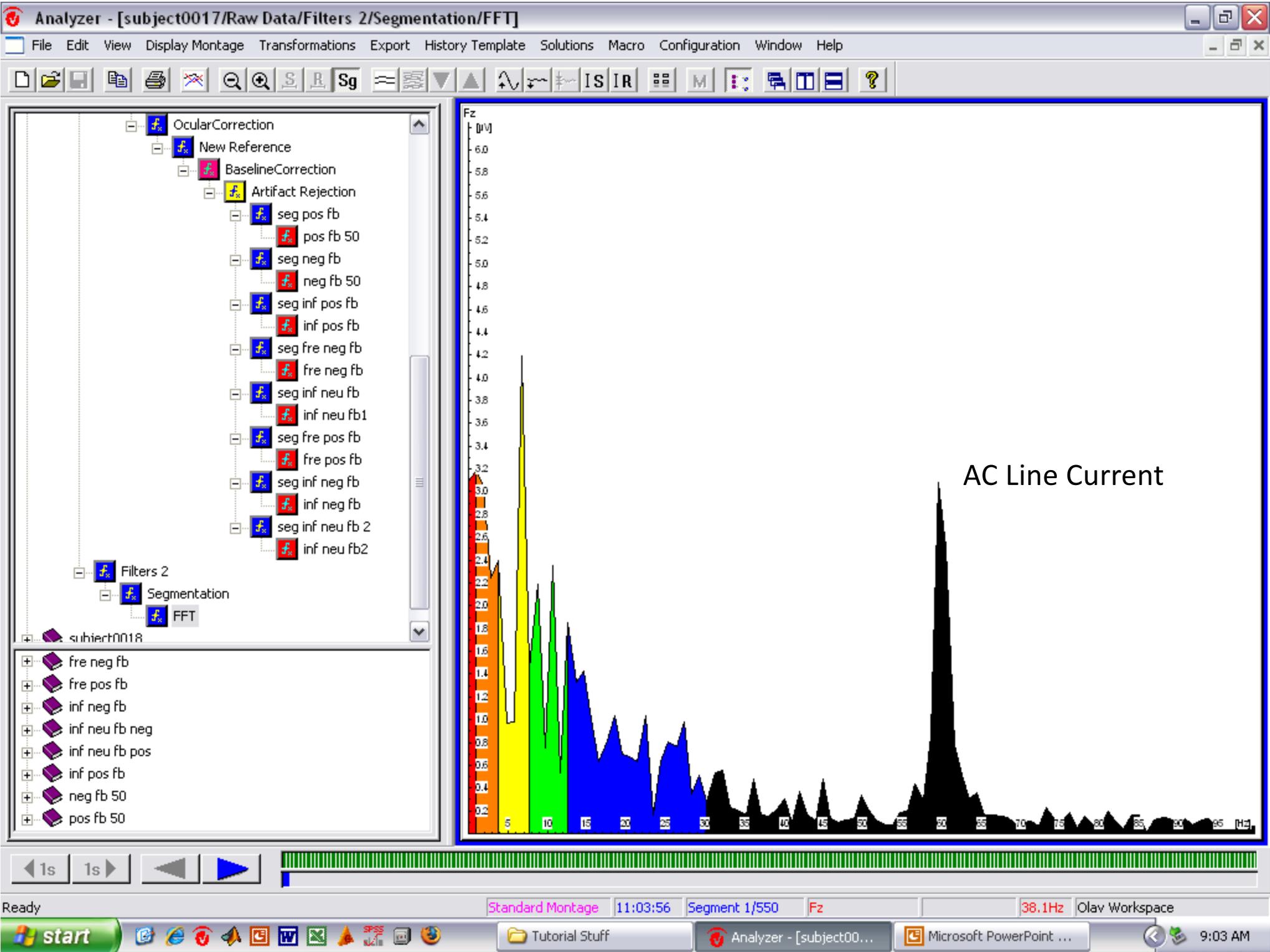
Map Legend

Mapping

Scaling Bar

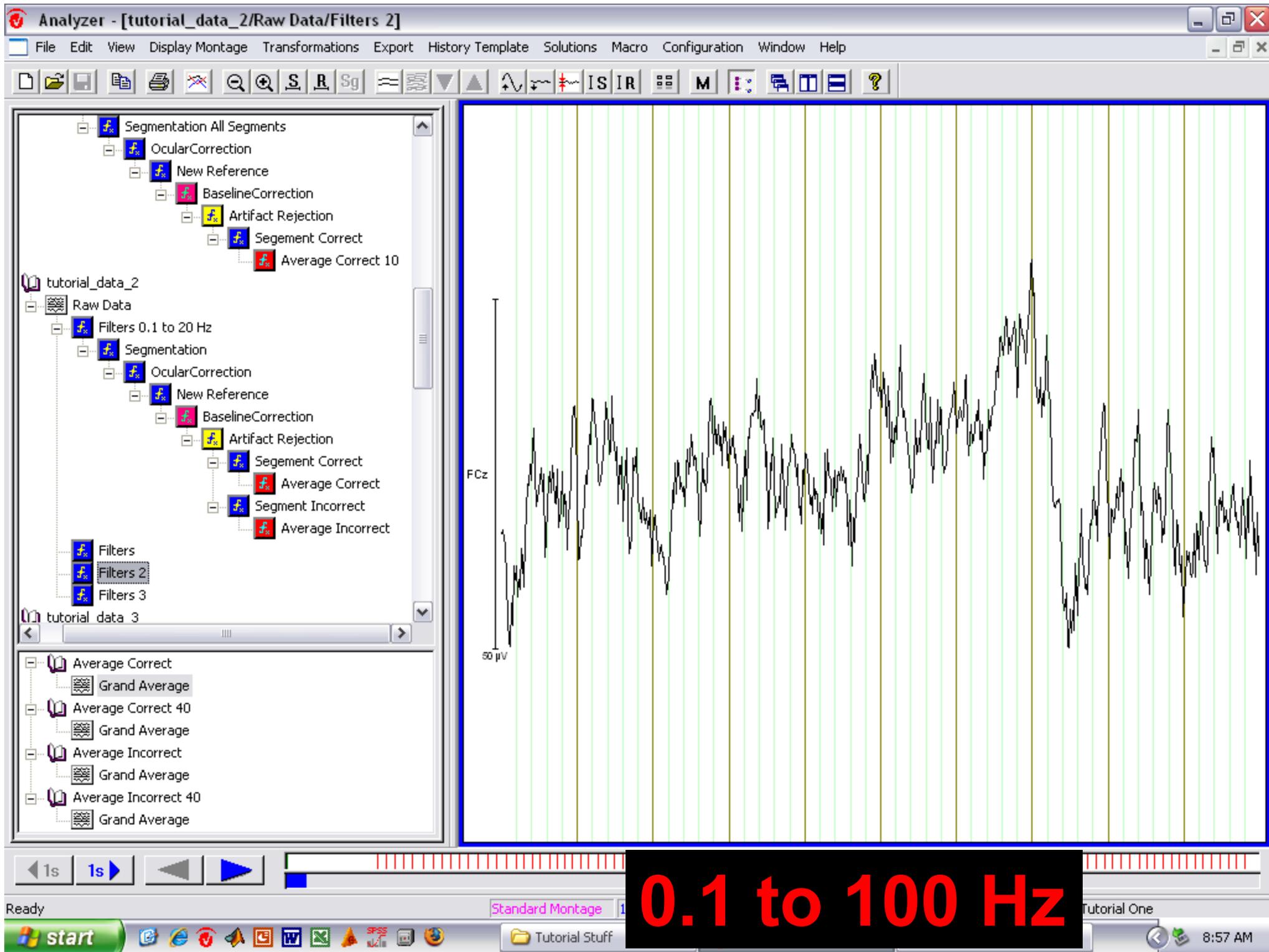
Text Box

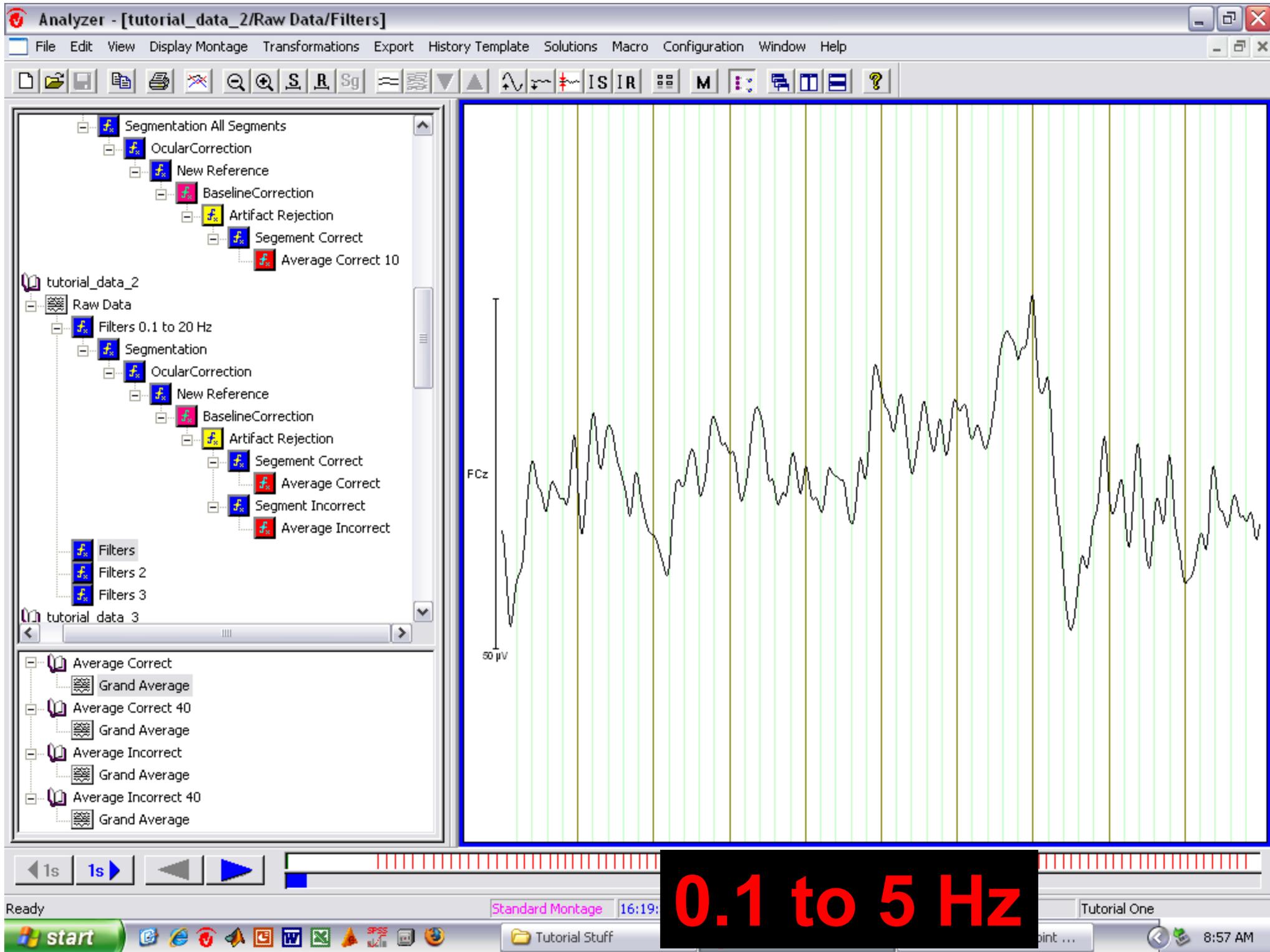
Value Graphics

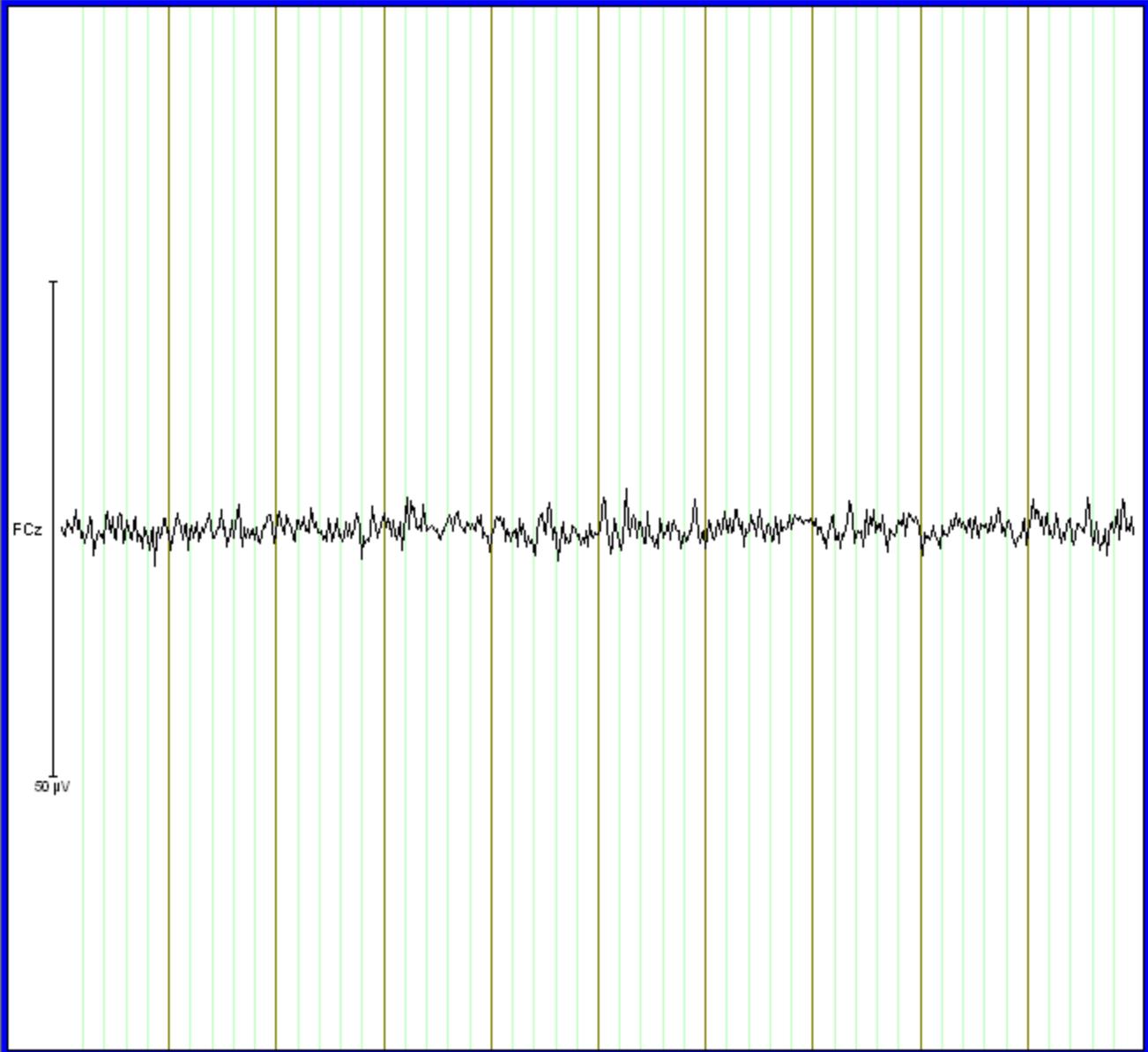
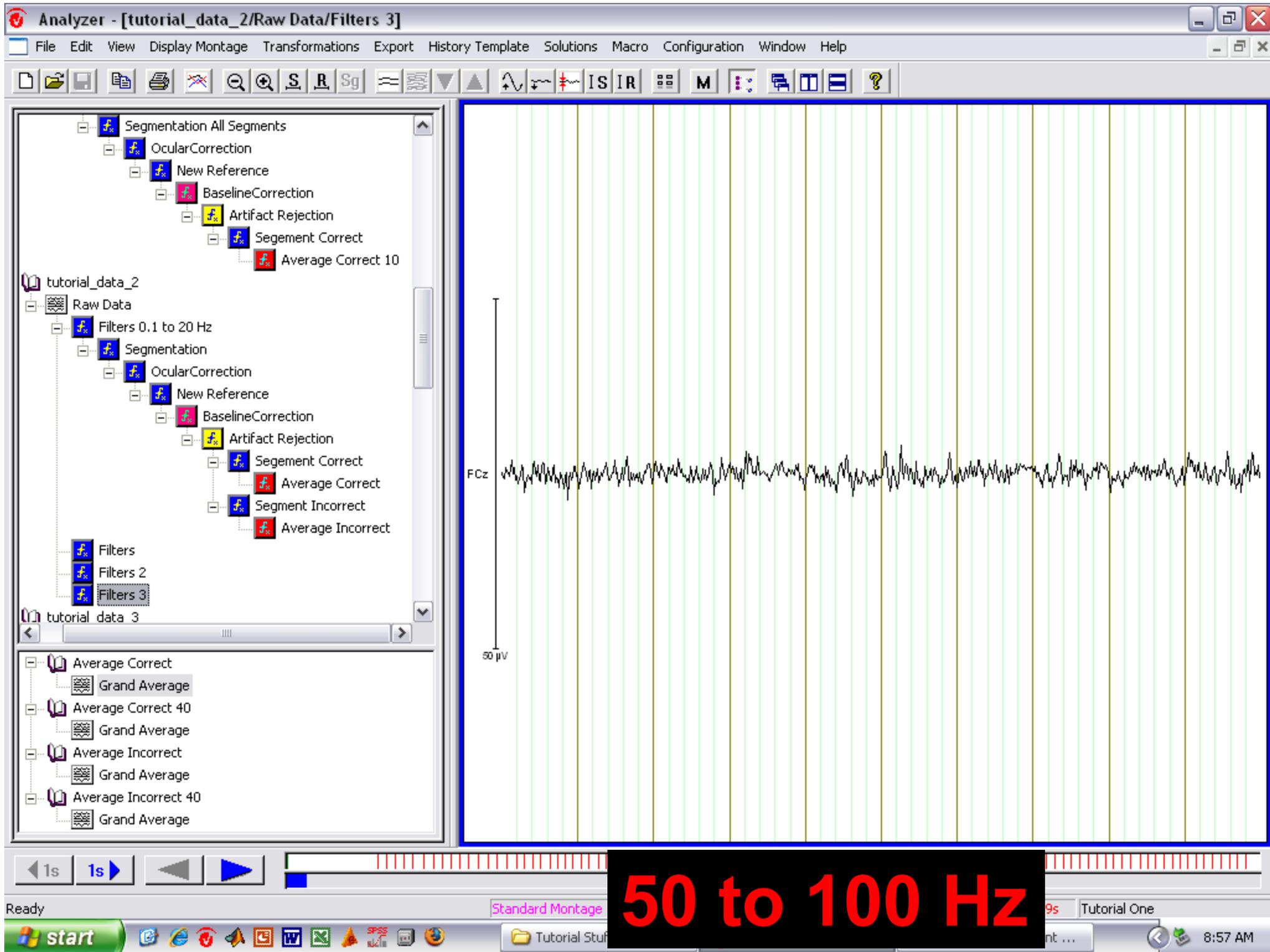


# Types of Filters

1. Low Pass (30 Hz – removes above this)
2. High Pass (0.1 Hz – removes below this)
3. Bandpass (0.1 to 30 Hz – removes below and above)
4. Notch Filters (remove a specific frequency)







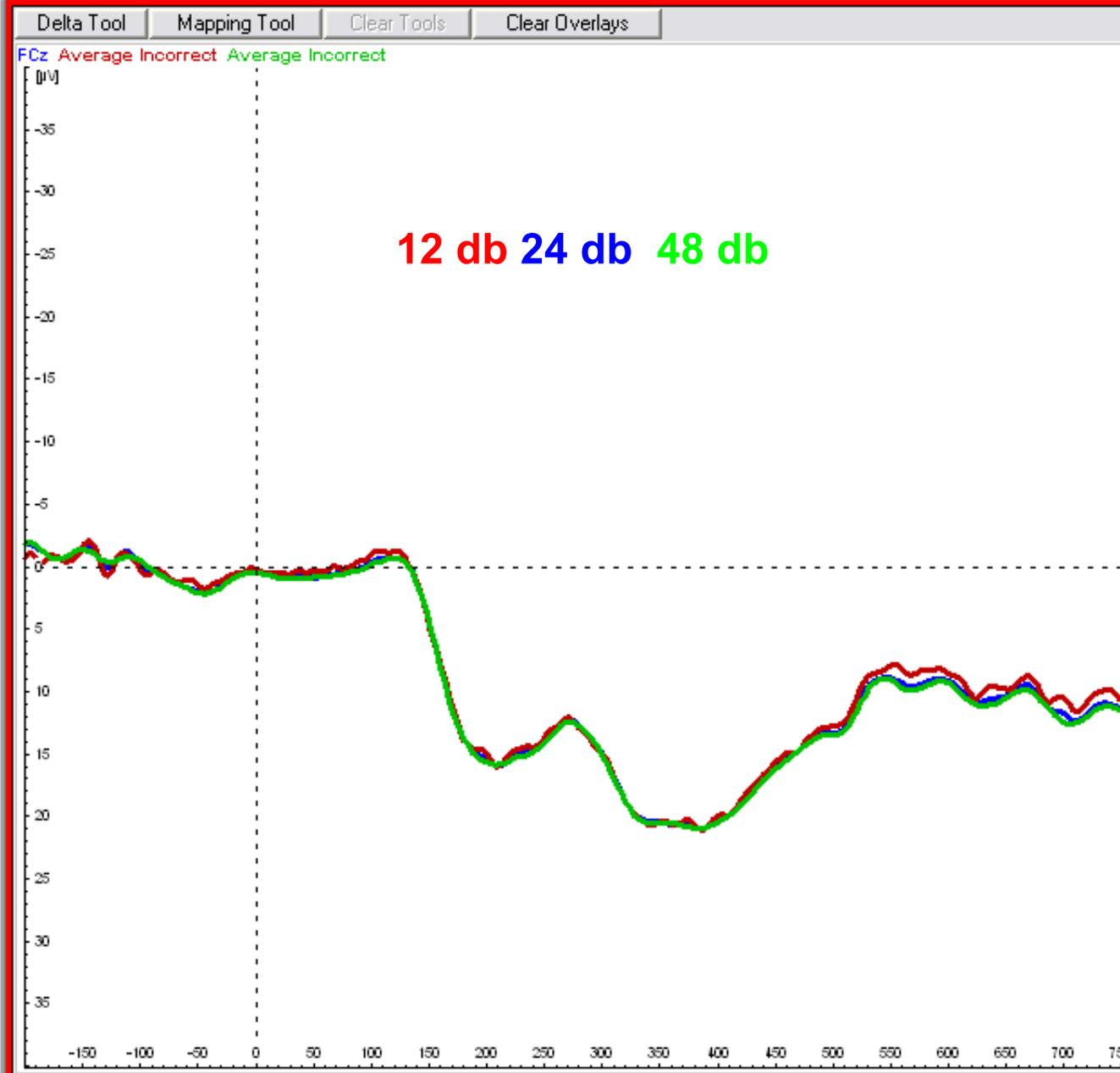
**50 to 100 Hz**

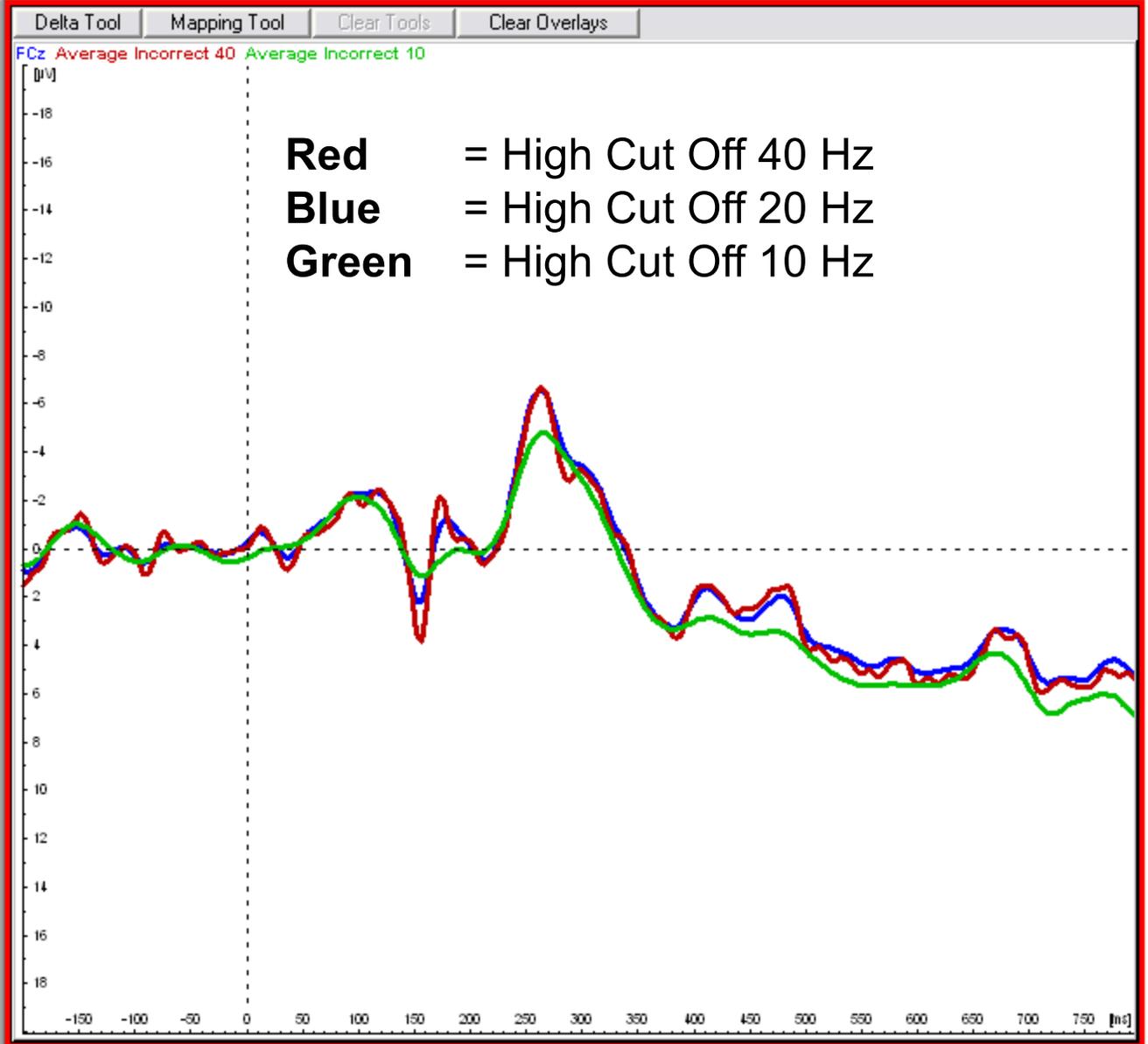
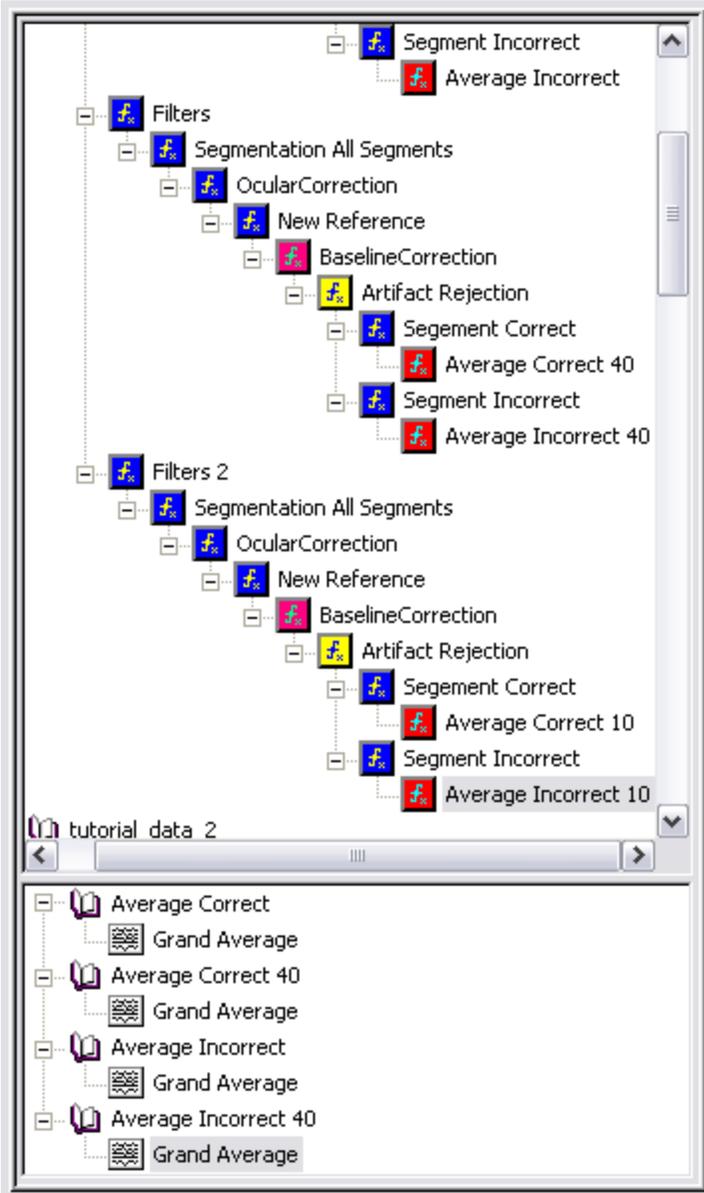


- New Reference
  - BaselineCorrection
    - Artifact Rejection
      - Segment Correct
        - Average Correct
      - Segment Incorrect
        - Average Incorrect
- Filters 2
  - Segmentation
    - OcularCorrection
      - New Reference
        - BaselineCorrection
          - Artifact Rejection
            - Segment Correct
              - Average Correct
            - Segment Incorrect
              - Average Incorrect

tutorial\_data\_3

- Raw Data
  - Filters 0.1 to 20 Hz
    - Segmentation
      - OcularCorrection
        - New Reference
- Average Correct
  - Grand Average
- Average Correct 40
  - Grand Average
- Average Incorrect
  - Grand Average
- Average Incorrect 40
  - Grand Average





# Filter Distortion

Filters can distort the AMPLITUDE and LATENCY of components.

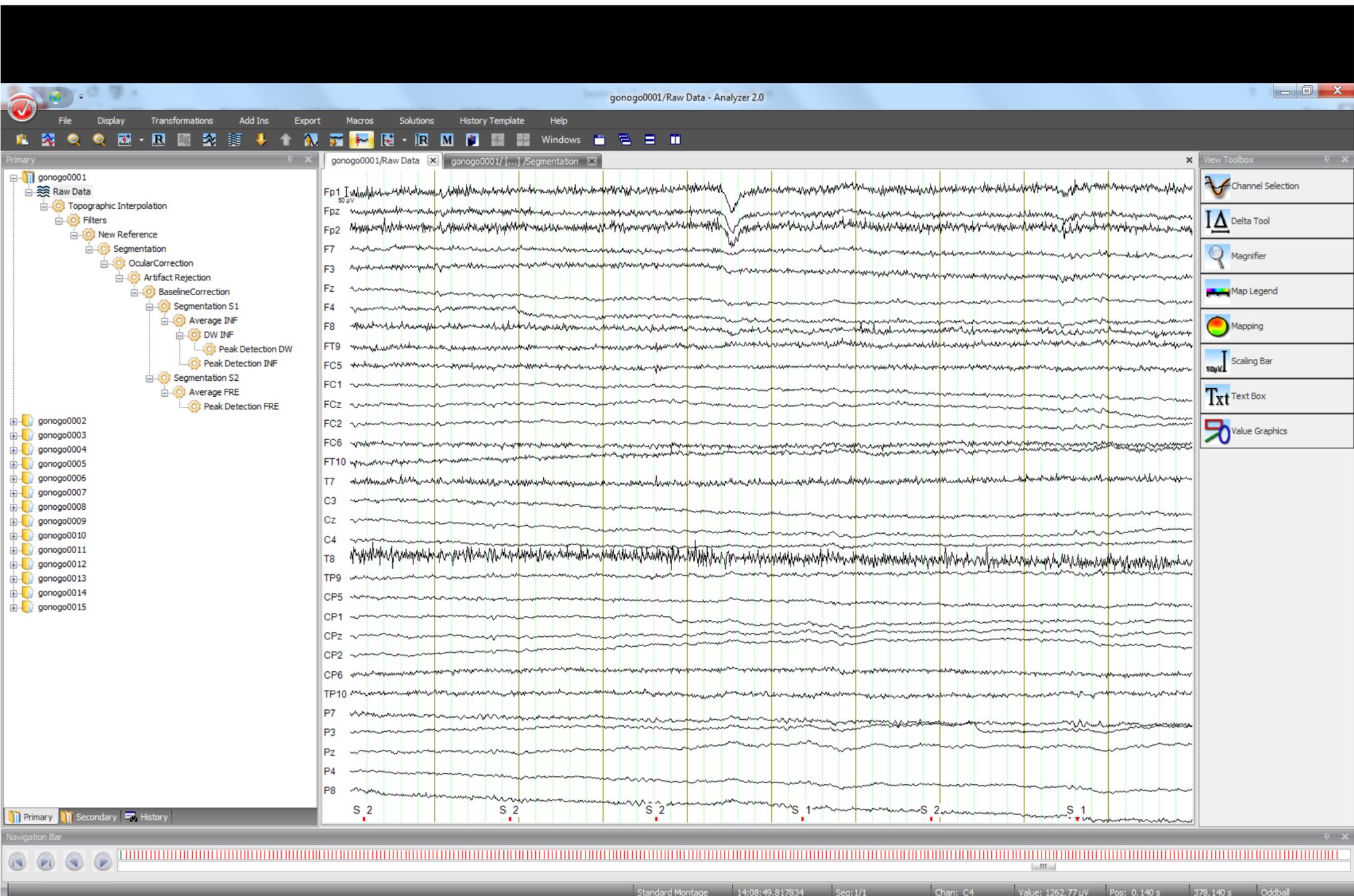
Use sensible filters – 0.1 to 30 Hz

Use sensible slopes – 24 dB

See Luck, 2014, for considerably more detail.

# Fourier Demonstrations

# Segmentation (Epoching)



Continuous Data

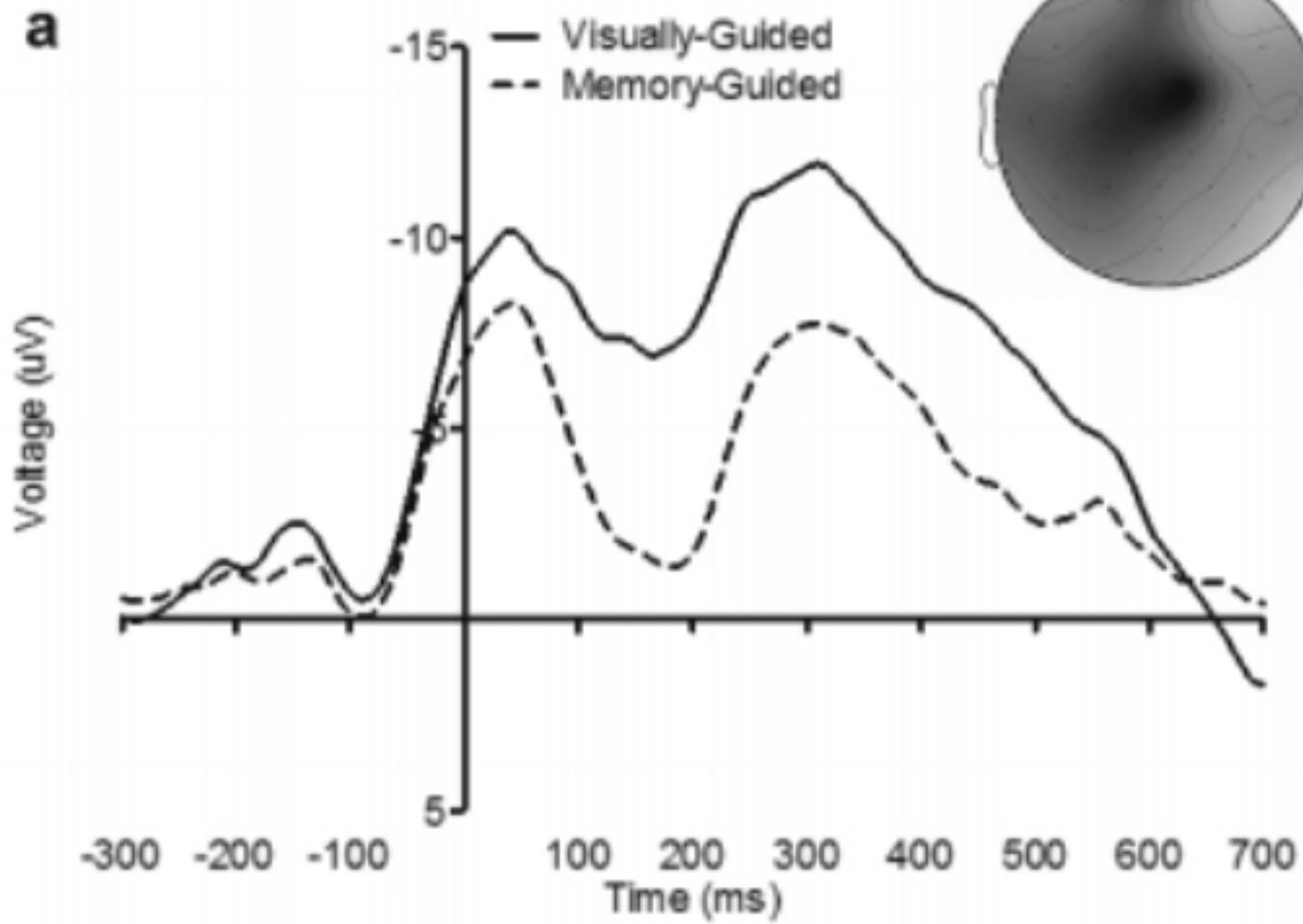


# Segmentation (Epochs)

Essentially, identifying a window of a given trial length

## Considerations

- is the baseline common
- how large of a baseline
- how long of a window (ideally as short as possible but depends on ICA and or FFT analysis)
- other considerations



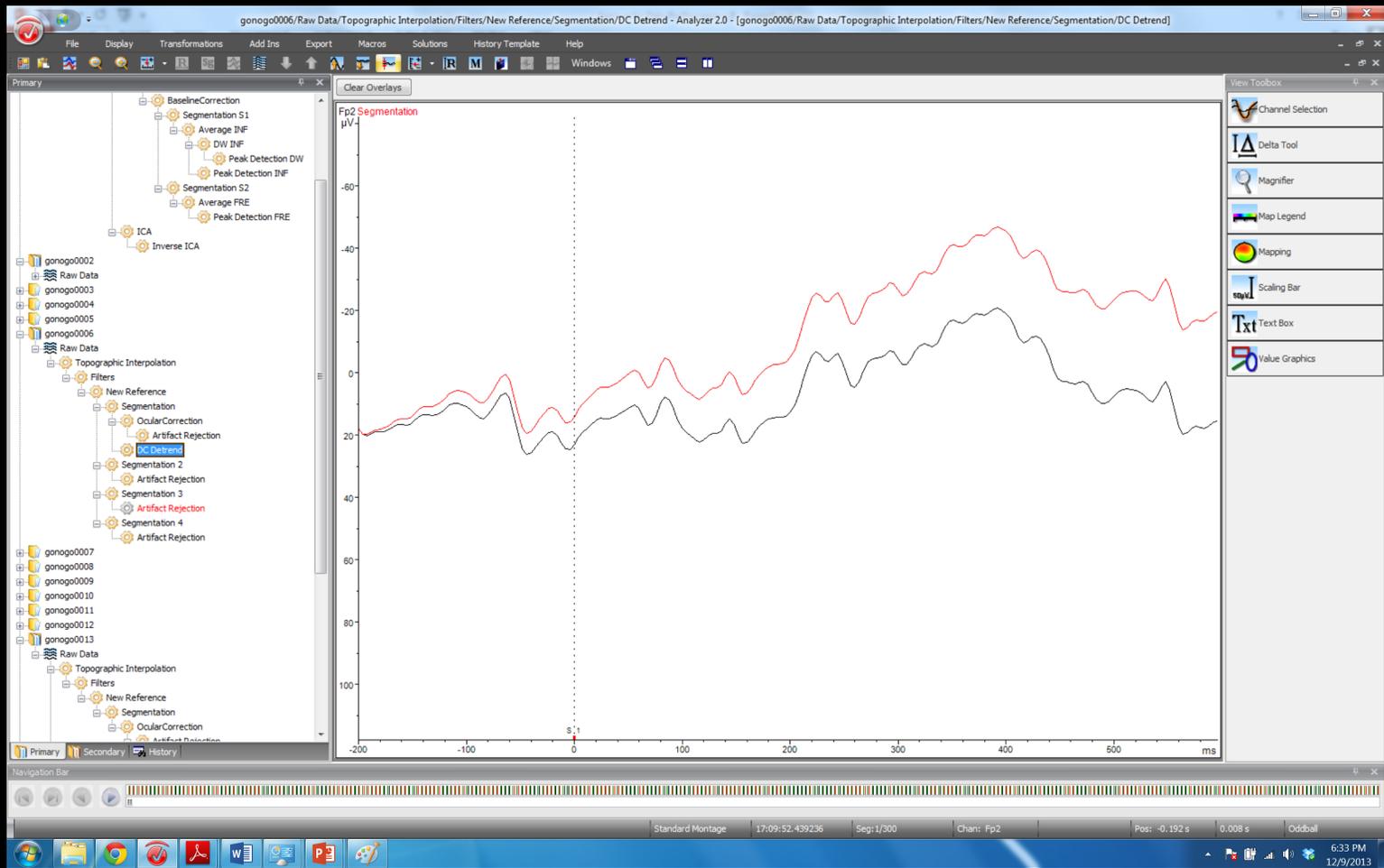
# Segmentation Demos

DC Detrend (optional)

# DC Detrend

Removes a continuous DC slope from each segment.

Note, this step, if used, should be done before ocular correction.



# DC Detrend Demo

# Ocular Correction (Tomorrow)

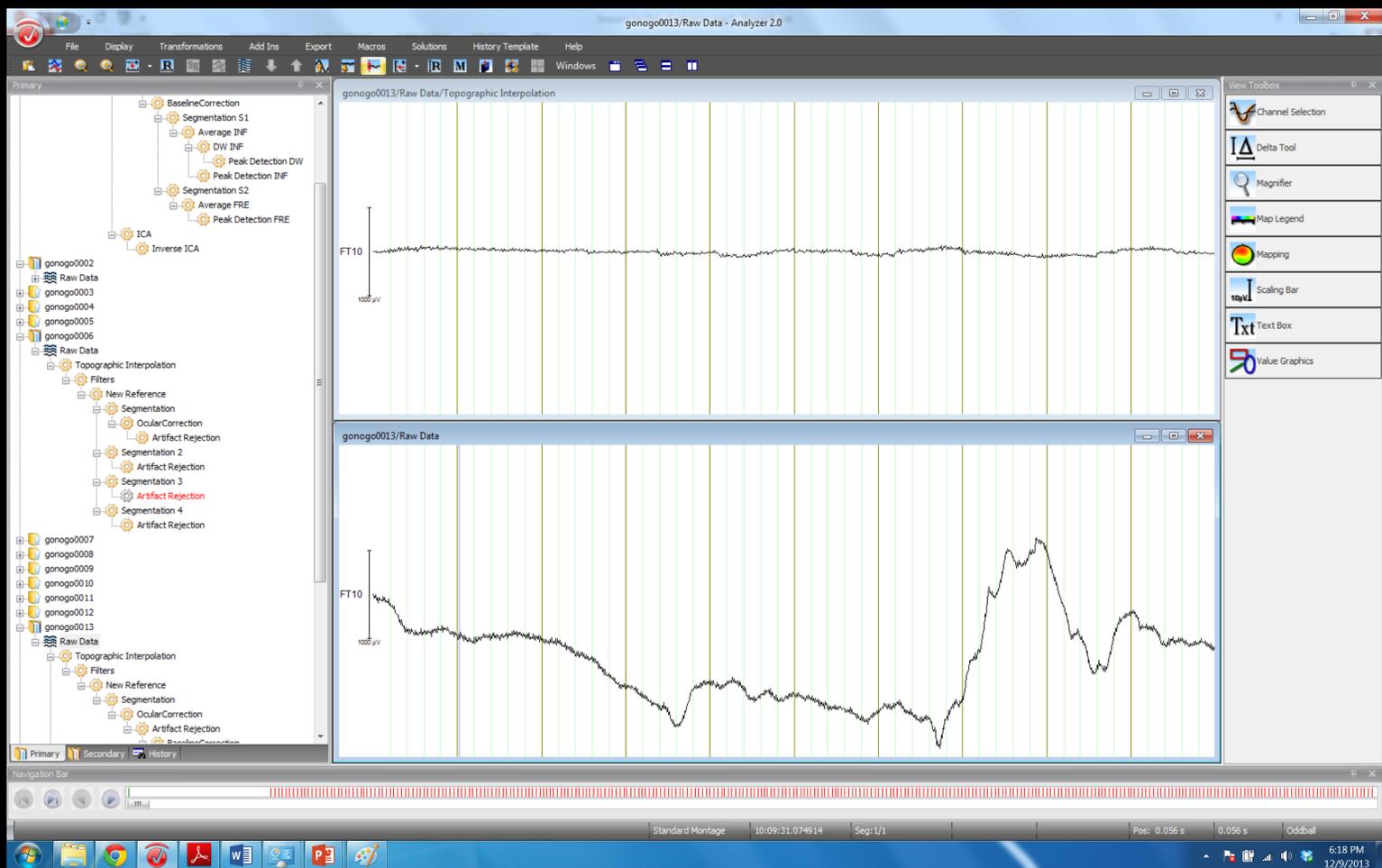
# Channel Interpolation

# Channel Interpolation

Essentially, you replace a channel with data from the surrounding channels.

Given the propagation of the EEG signal this is not a bad idea in principle.

Can be done via a variety of methods – most typically as an average of the surrounding channels or using a method based on spherical splines.



# Channel Interpolation

Never interpolate:

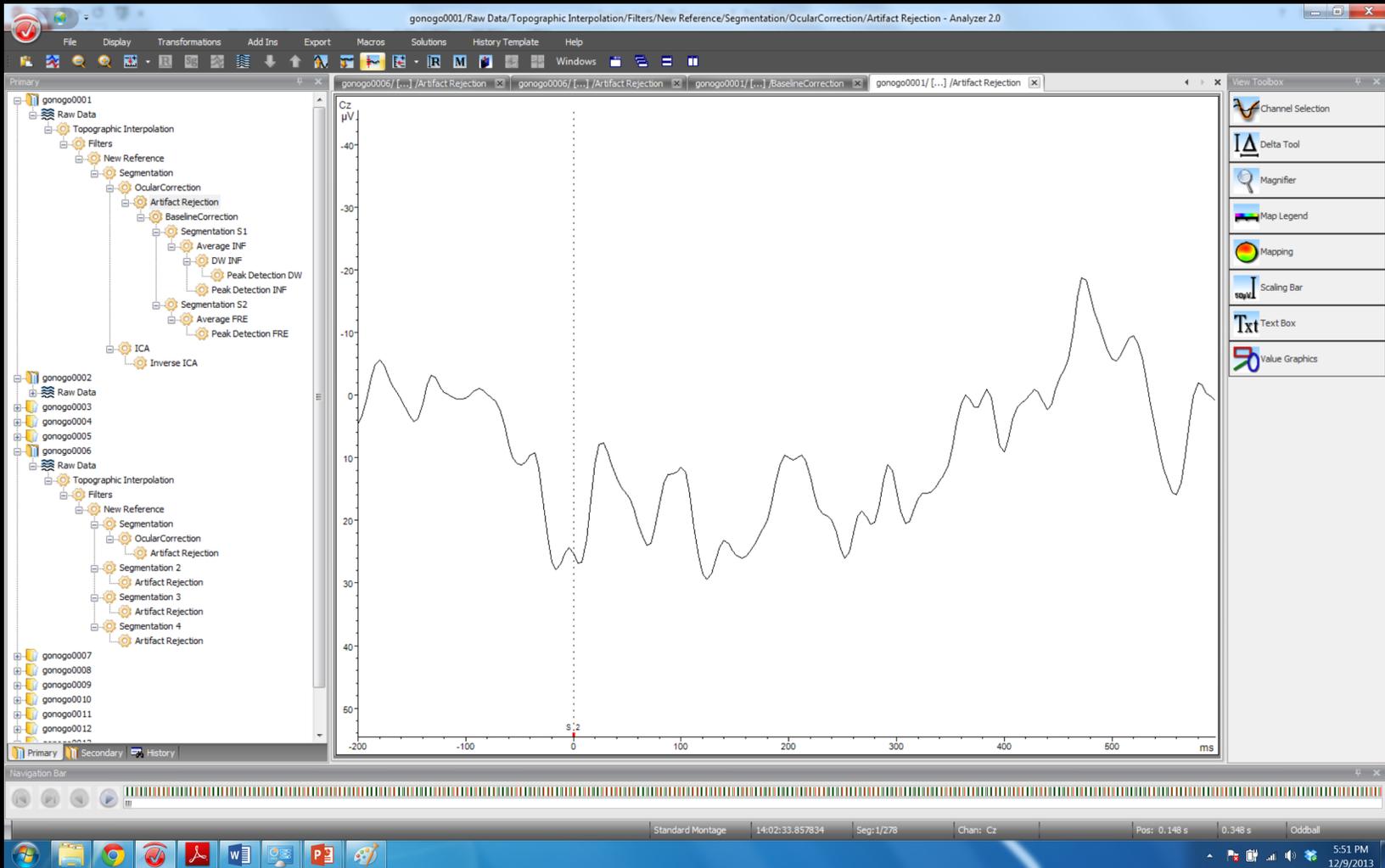
1. Your channel of interest(s)
2. Reference channels
3. EOG channels
4. Too many channels

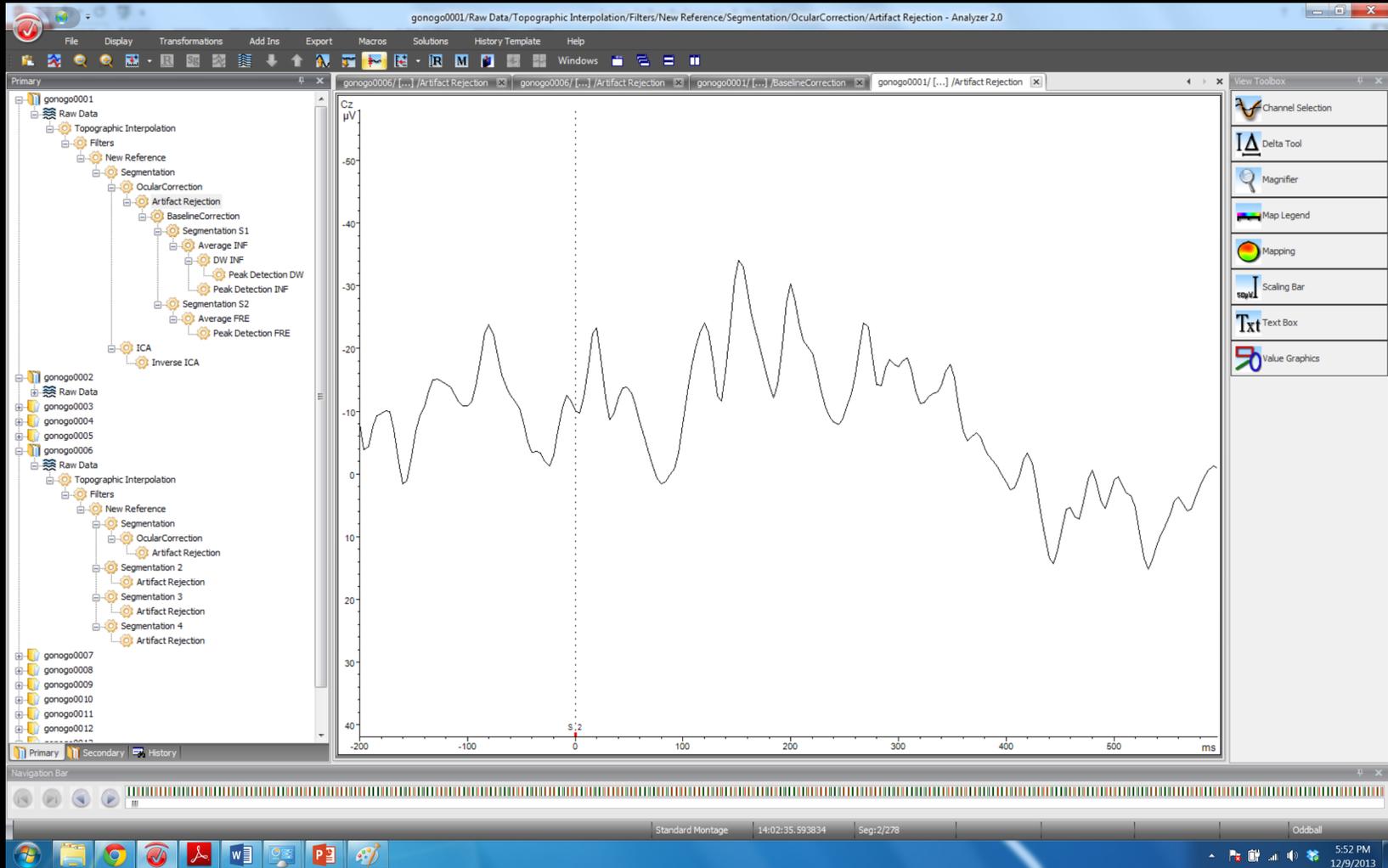
Always visually verify the interpolation looks reasonable.

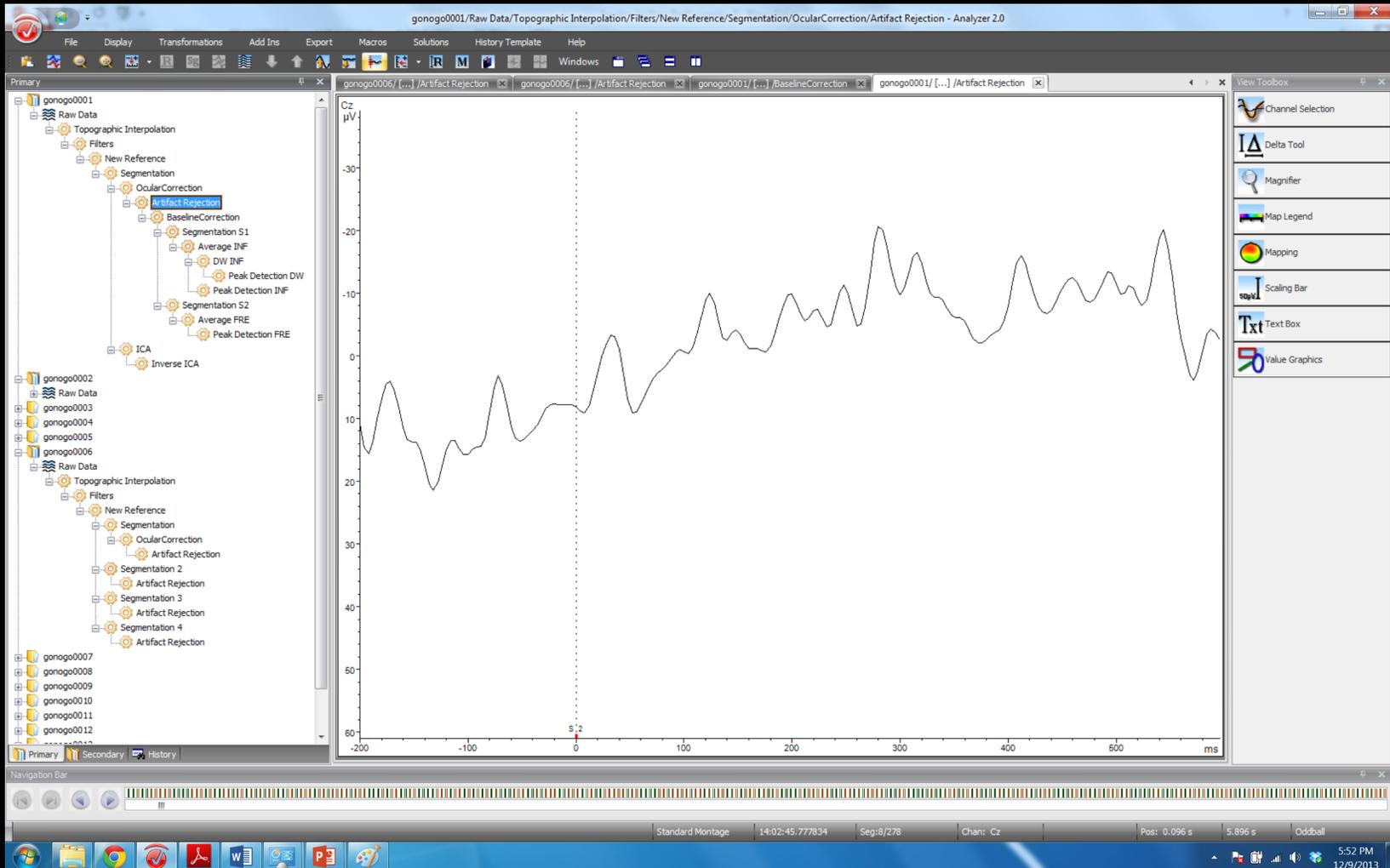
Do not be afraid to throw out bad data!

# Channel Interpolation Demos

# Baseline Correction





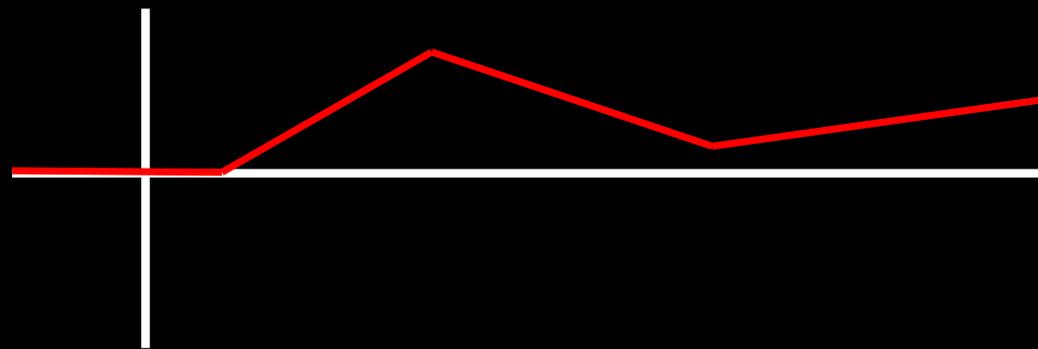
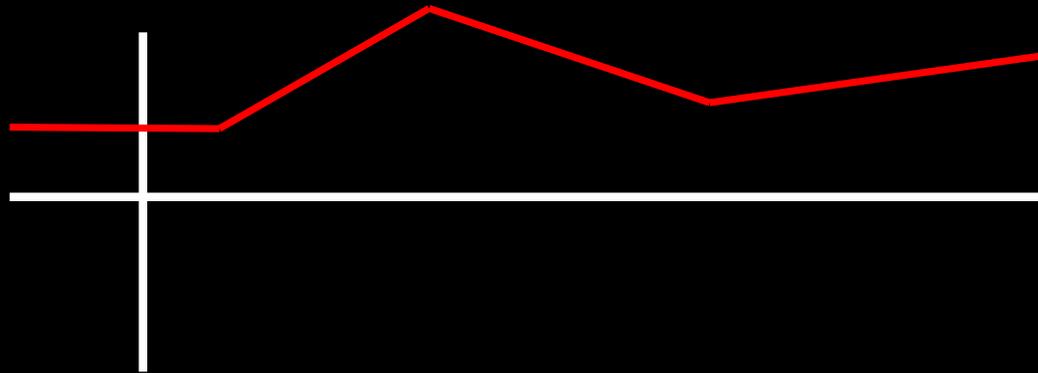


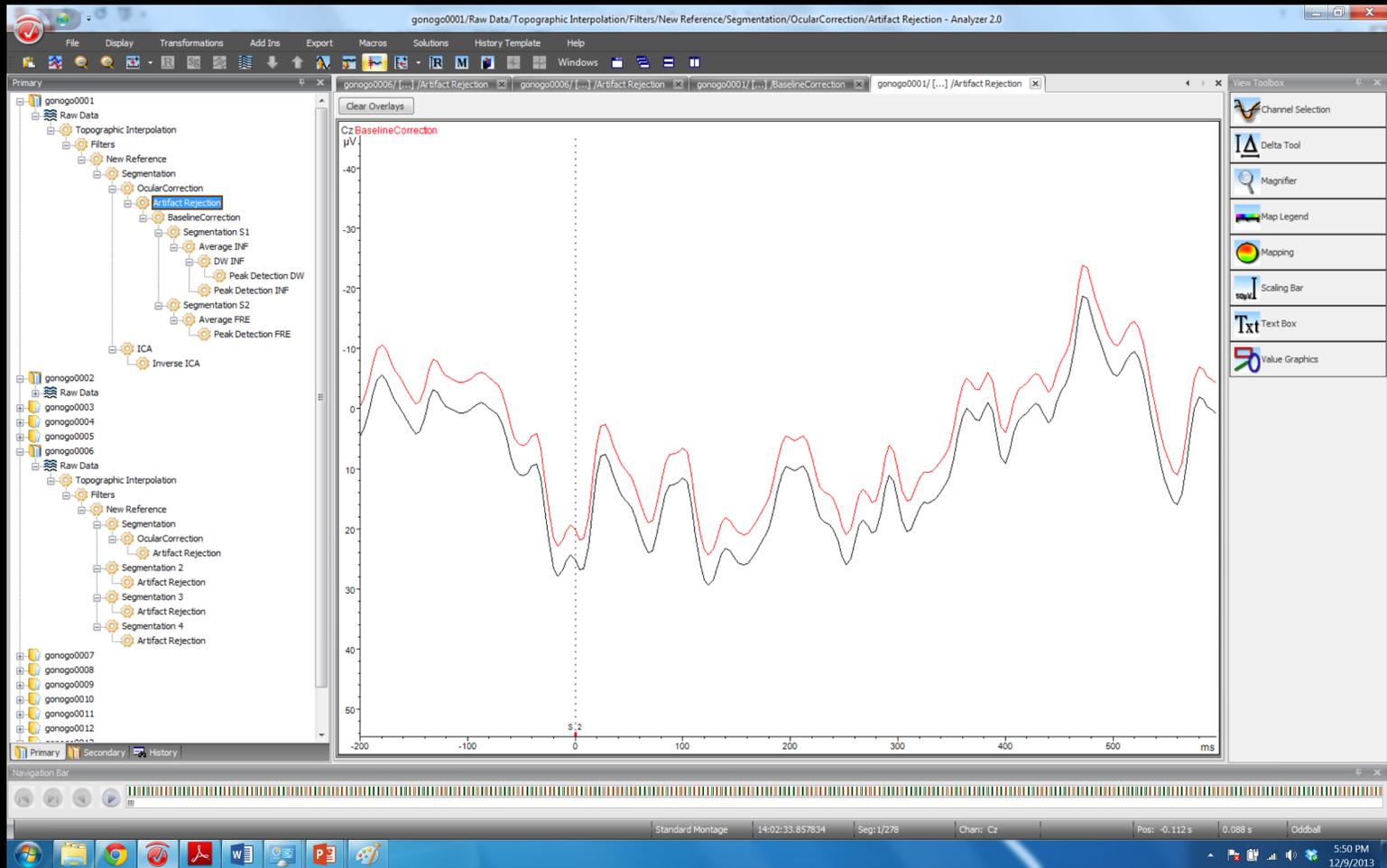
# Baseline Correction

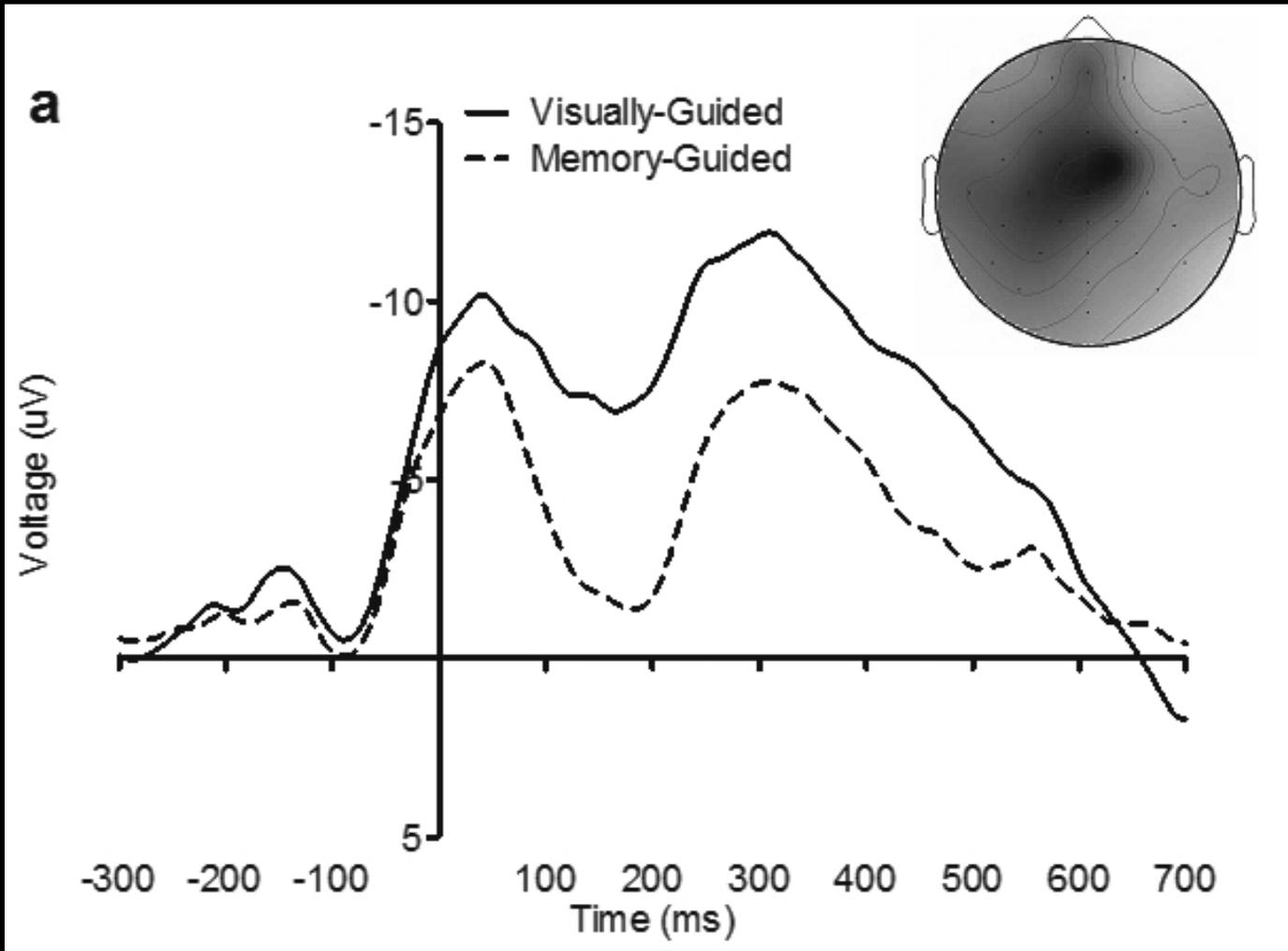
Perhaps the easiest step to understand mathematically.

For each segment, the mean voltage is taken for the baseline period (e.g., -200 to 0 ms) and that is subtracted from every time point.

This is done separately for each segment for each channel.







# Baseline Correction Demos

# Artifact Rejection

# Artifact Rejection

The idea in principle is very simple... use some form of criteria to identify artifacts and then remove those trials from a data set.

You can mark artifacts manually or automatically. I would recommend automatically to avoid introducing bias.

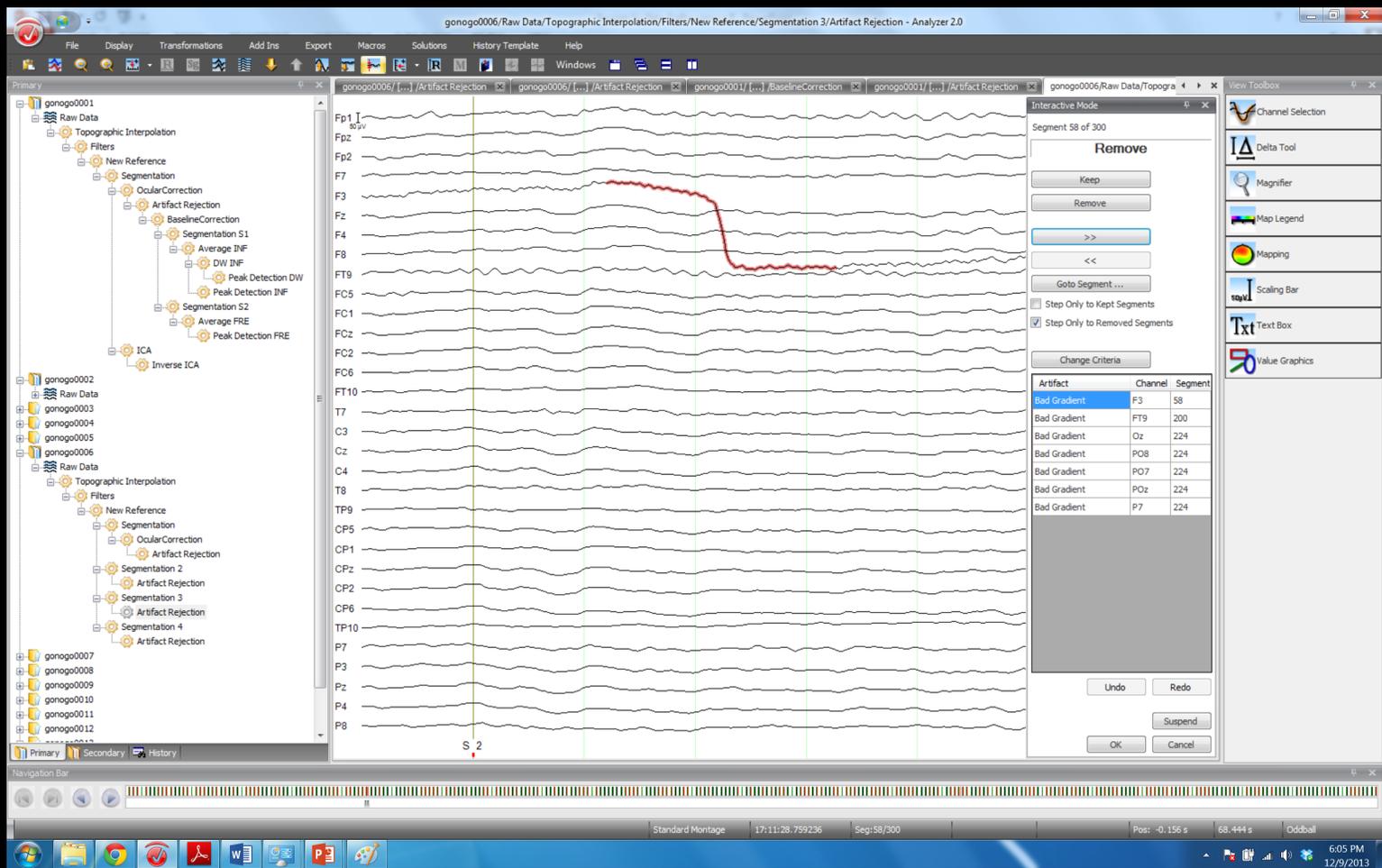
With that said, in some instances gross artifacts should be marked manually (for ICA).

# Artifact Rejection

## Criterion 1: Gradient

Essentially, the difference between any two sequential points in time.

Typically, a value of 10  $\mu\text{V}/\text{ms}$  is enough to catch “bad data” and not lose any “good data”.

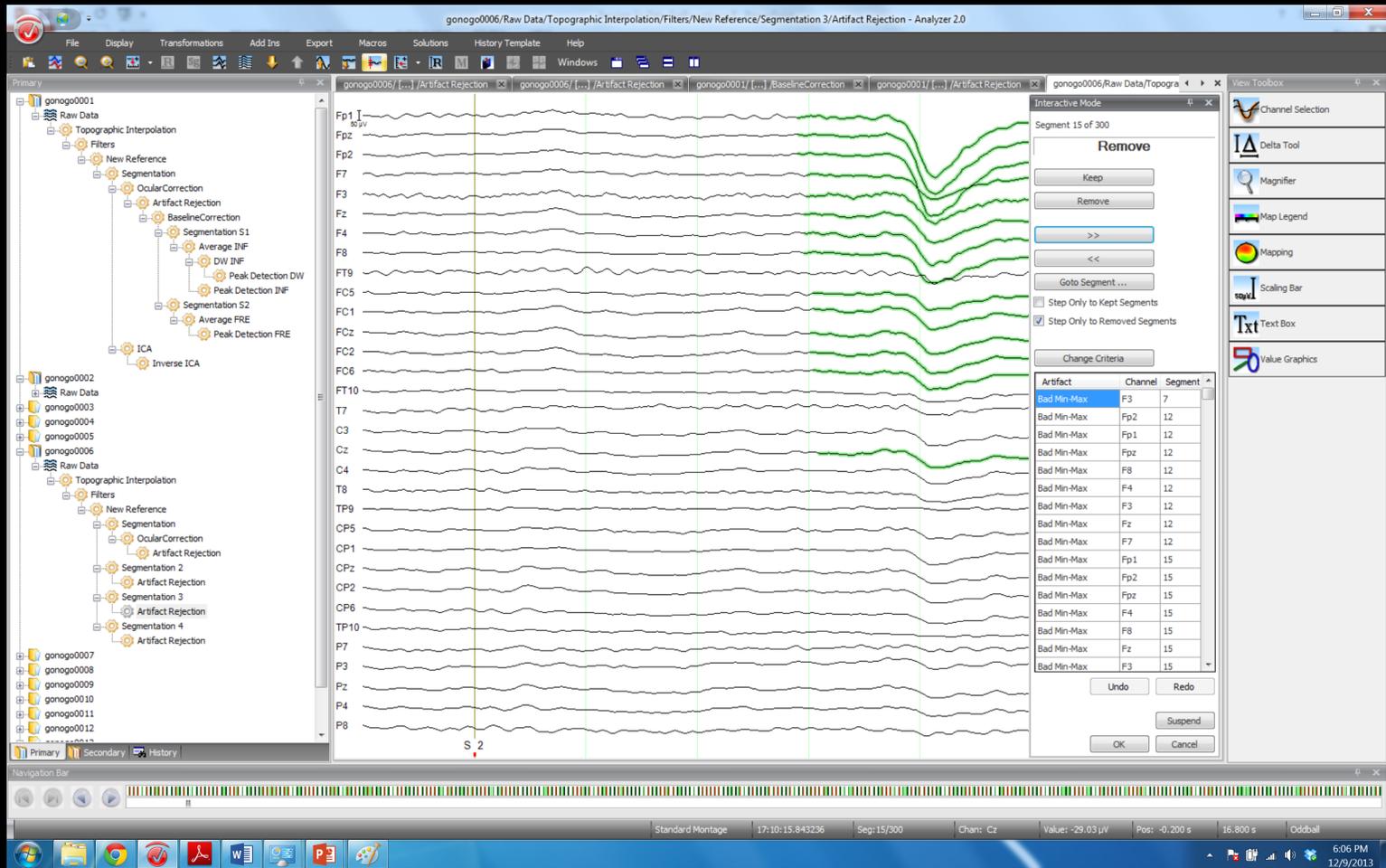


# Artifact Rejection

## Criterion 2: Max Min

Essentially, the difference between the maximum and minimum voltage in a segment,

Typically, a value of 100  $\mu\text{V}/\text{ms}$  is enough to catch “bad data” and not lose any “good data”.

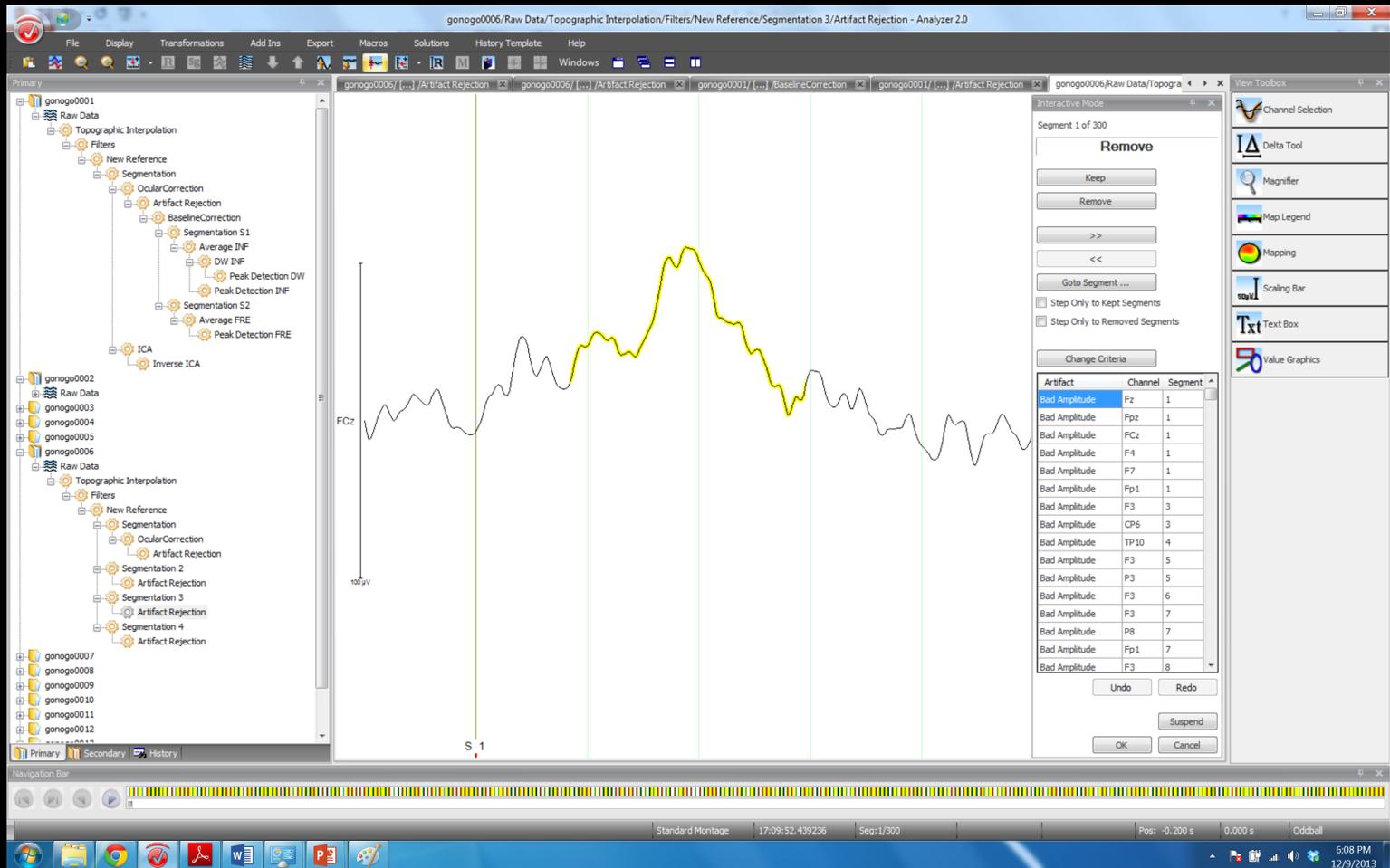


# Artifact Rejection

## Criterion 3: Amplitude

Essentially, the absolute maximum or minimum voltage that is acceptable

There is no need to use this if you use max min and gradient criteria.



# Artifact Rejection

## Criterion 4: Low Activity

Essentially this looks for segments where the deviation in voltage is below a certain range, say 0.5  $\mu\text{V}$  over 800 ms.

This is essentially a check for a flat channel.

Re-referencing will distort flat channels so this is effectively useless.

# Artifact Rejection

Some thoughts...

1. It is important to keep track of the number of artifacts by condition to look for systematic differences.
2. Make sure you eyeball what is being thrown away.
3. Use the most conservative artifact criteria as possible.

And this finally, might be the information needed to remove channels and interpolate them. So... Back to the start if you need to...

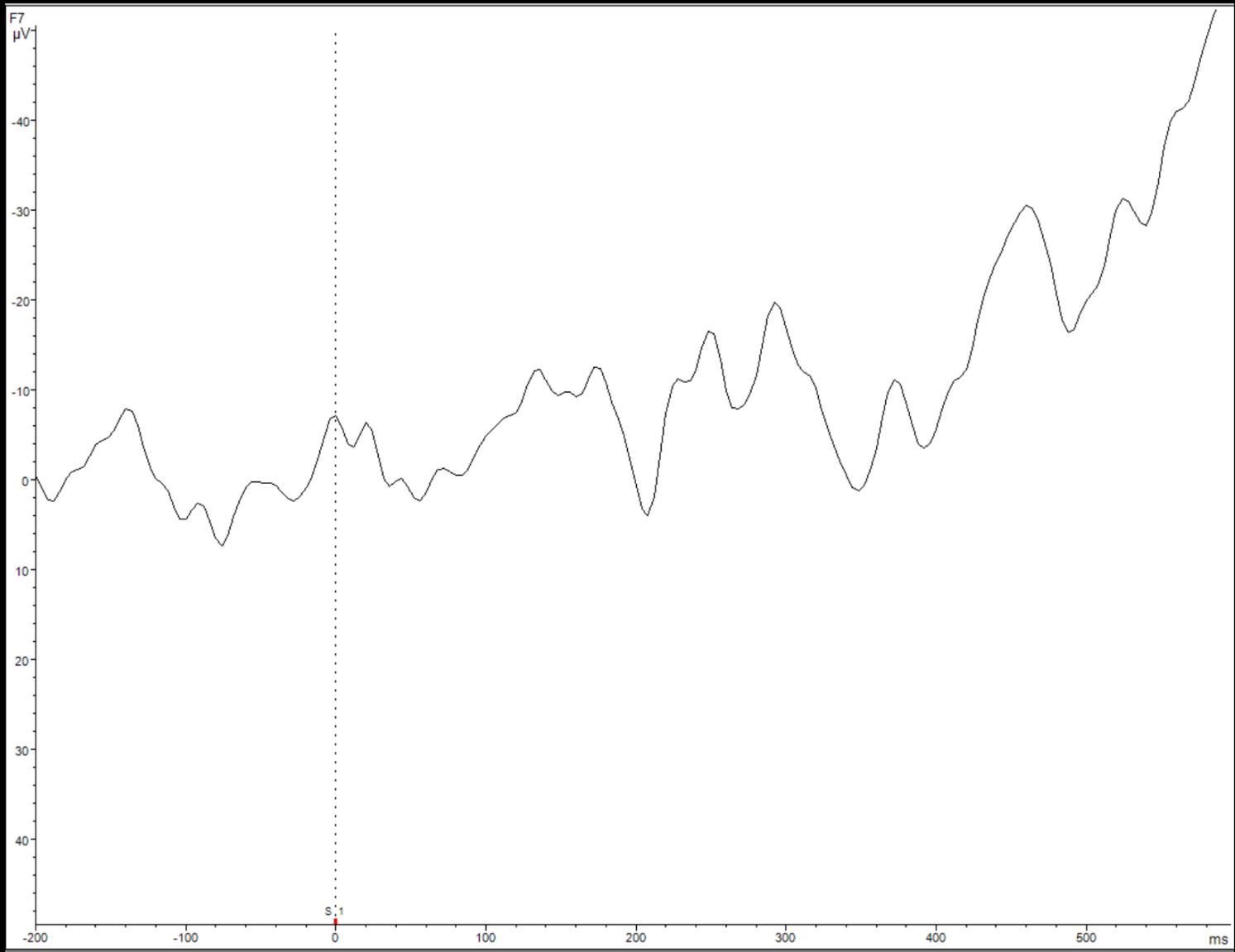
# Artifact Rejection Demos

# Participant Averages

# Participant Averages

A lot of the analysis we have done has been automated, or at least has the potential to be automated.

Before you generate a participant average you should always take a visual inspection of the segments going into the average for anything “weird” and potential remove these segments from analysis.



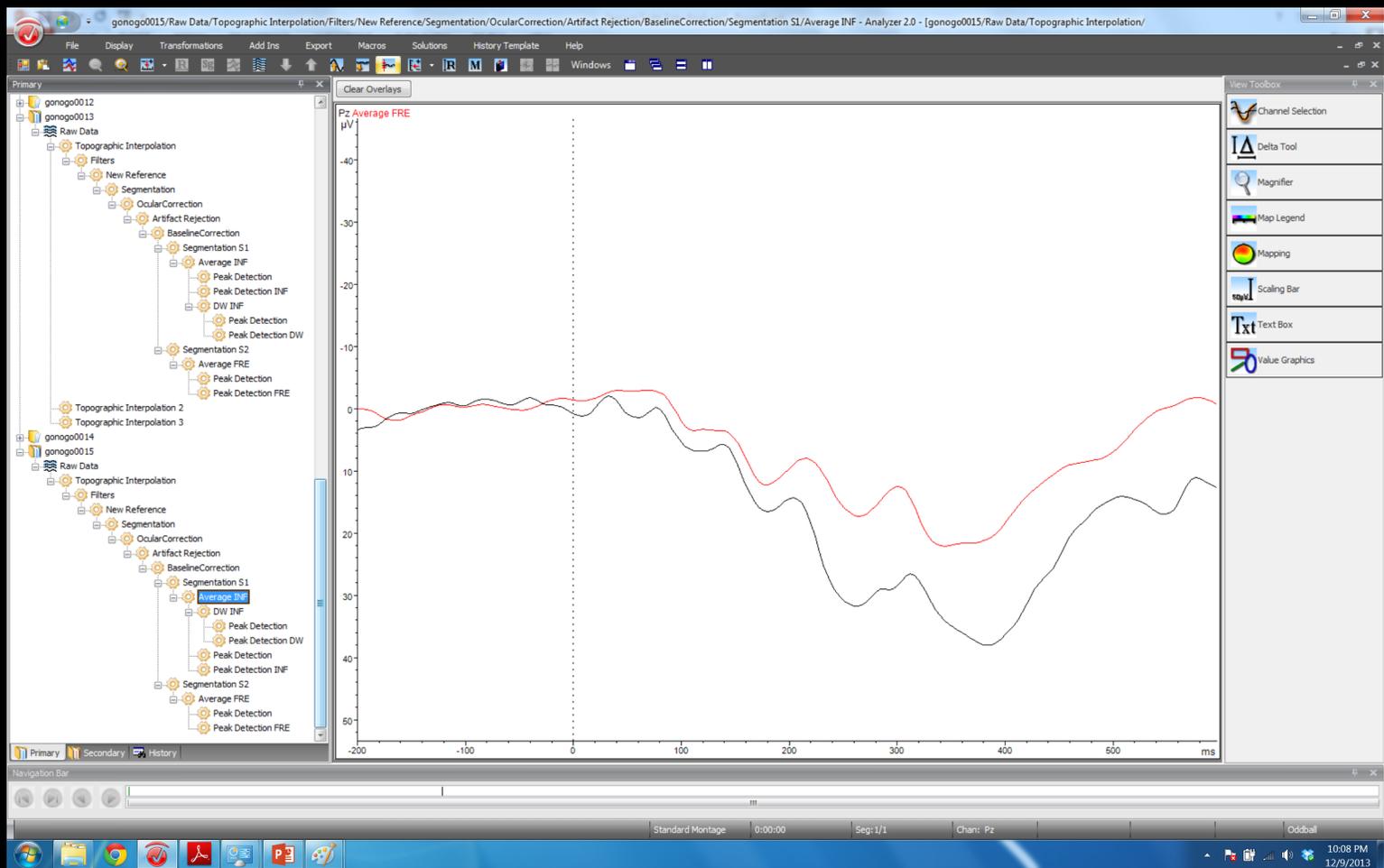
# Participant Averages

A lot of the analysis we have done has been automated, or at least has the potential to be automated.

Before you generate a participant average you should always take a visual inspection of the segments going into the average for anything “weird” and potential remove these segments from analysis.

# When done however...

You will have an average waveform, or ERP, for each participant for each experimental condition.



# Participant Average Demos

# Order of Operations

Most of these steps are linear – they can be interchanged.

For instance, if you are using gradient and max-min criteria for artifact rejection then it does not matter if it comes before or after baseline correction.

There are some exceptions...

# Order of Operations

## Rereference

Always reference on continuous data as soon as possible.

# Order of Operations

## Filtering

Always filter continuous data to avoid edge artifacts.

# Order of Operations

DC Detrend

Must be done before ocular correction.

# Order of Operations

## Ocular Correction

Should be done as early as possible after segmentation and before artifact rejection.

# Order of Operations

## Channel Interpolation

Always interpolate bad channels after ocular correction but before artifact rejection.

# Difference Waveforms