Machine learning on electrophysiology EEG/MEG signals

Alexandre Gramfort
http://alexandre.gramfort.net

@agramfort

with MNE

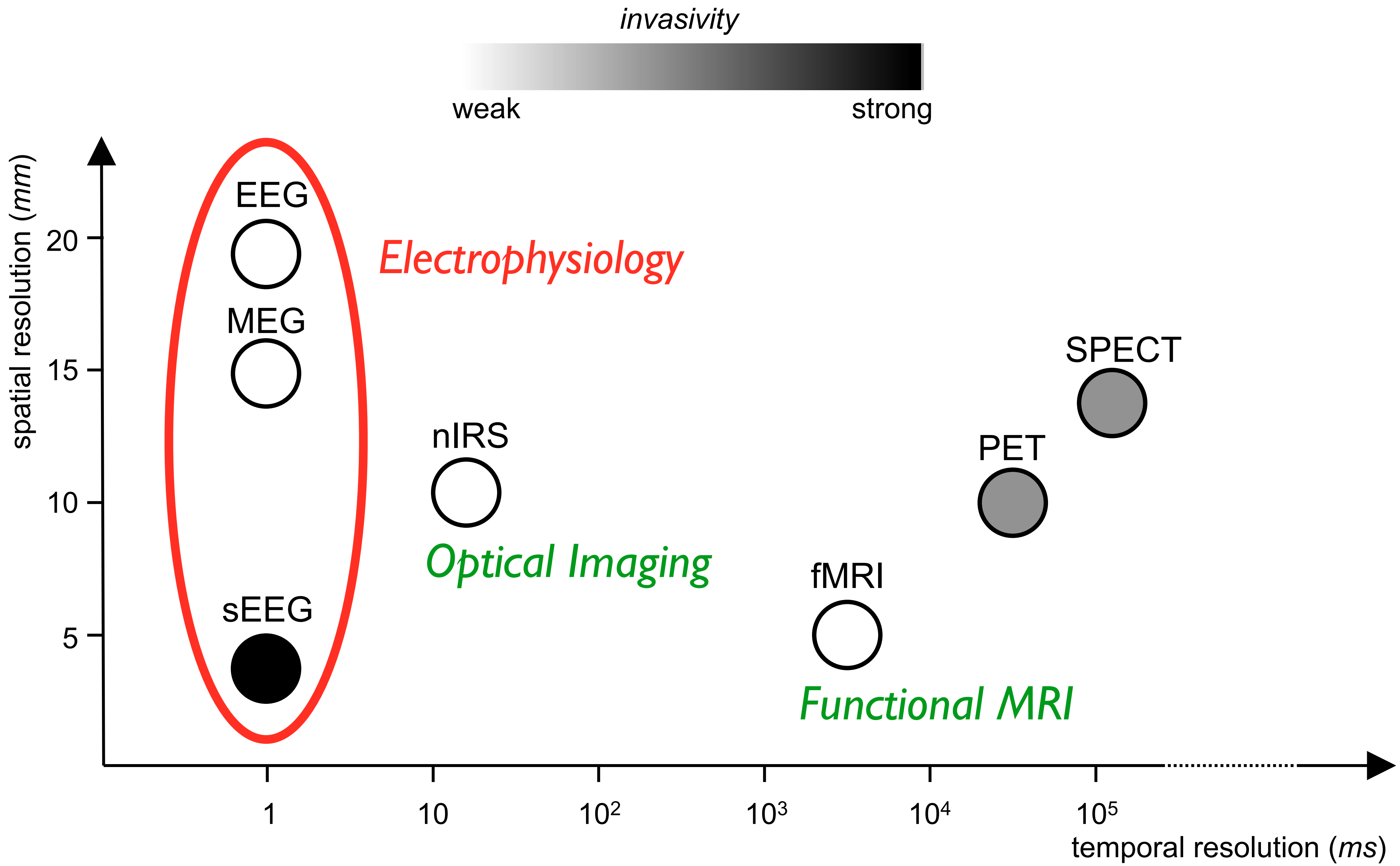
http://mne.tools

scikit learn

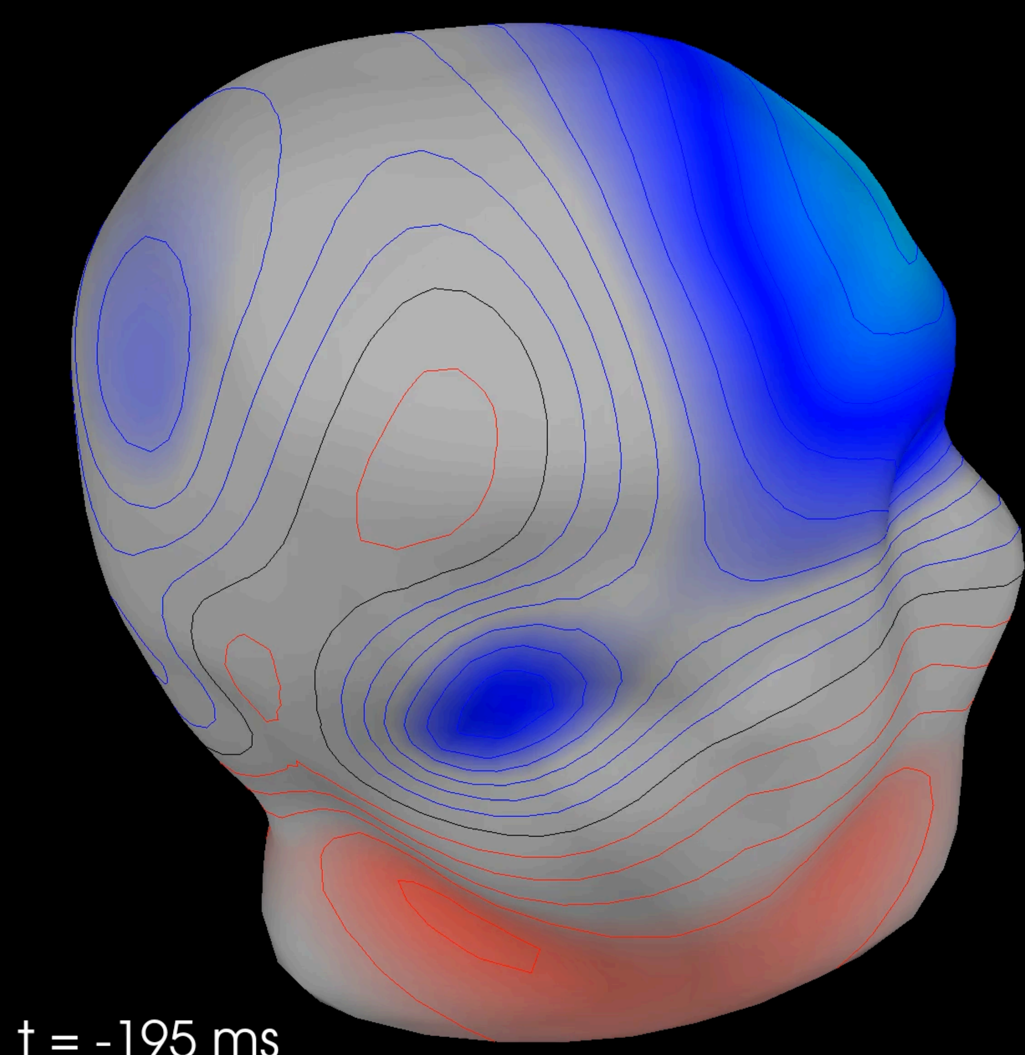
https://scikit-learn.org

Functional neuroimaging

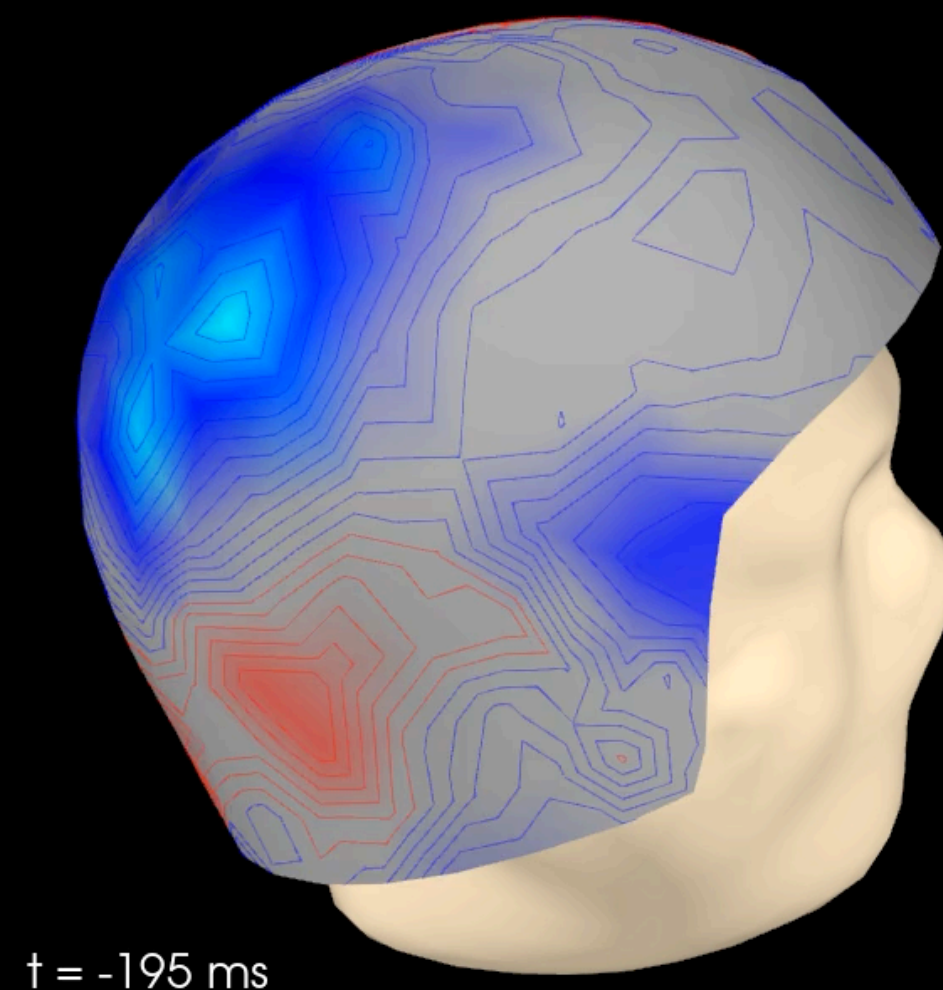
It’s the study of the **brain activity** through **functional imaging devices**



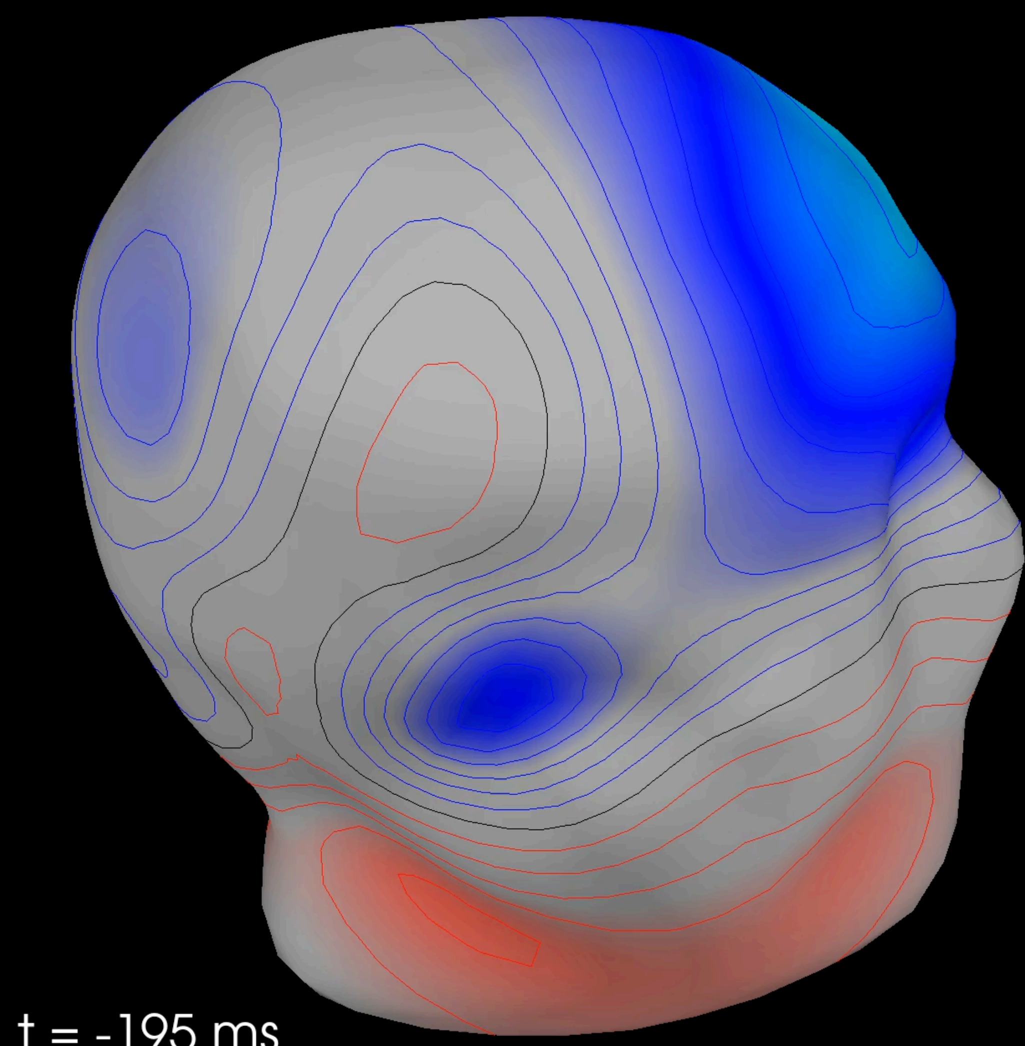
Electroencephalography (EEG)
Electric Potential [V]



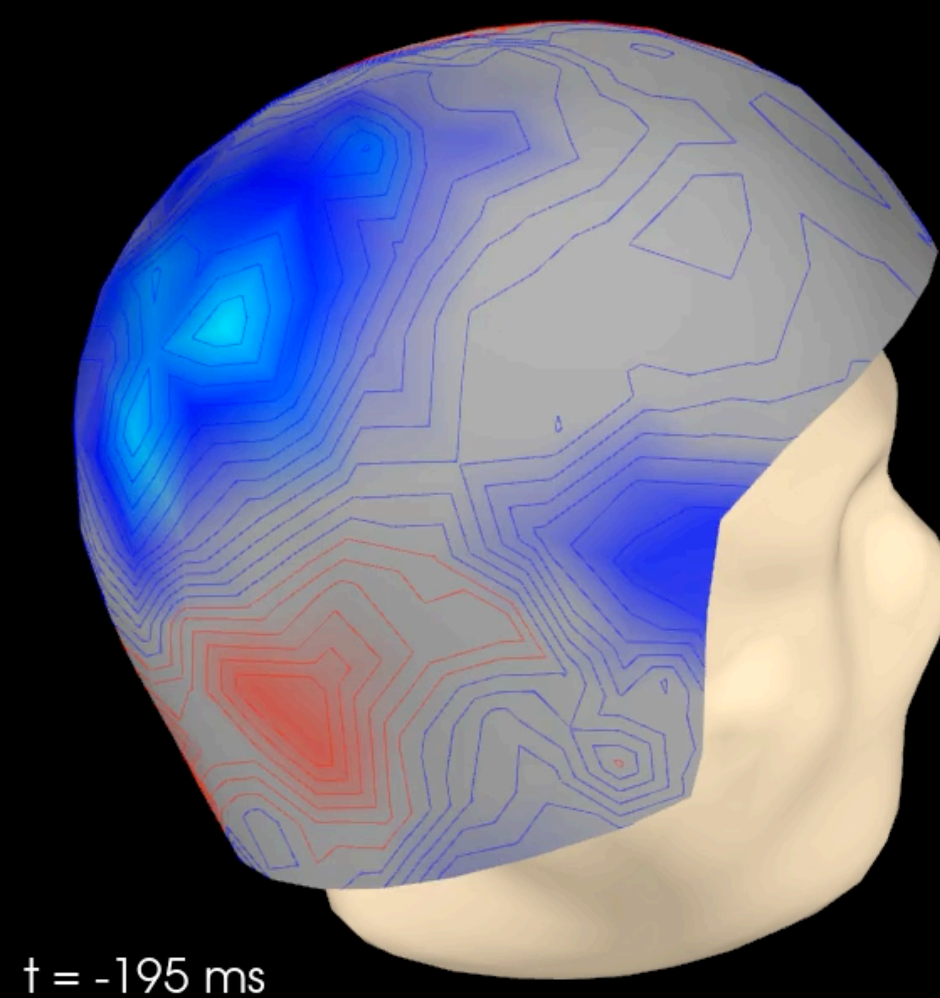
Magnetoencephalography (MEG)
Magnetic field [T]



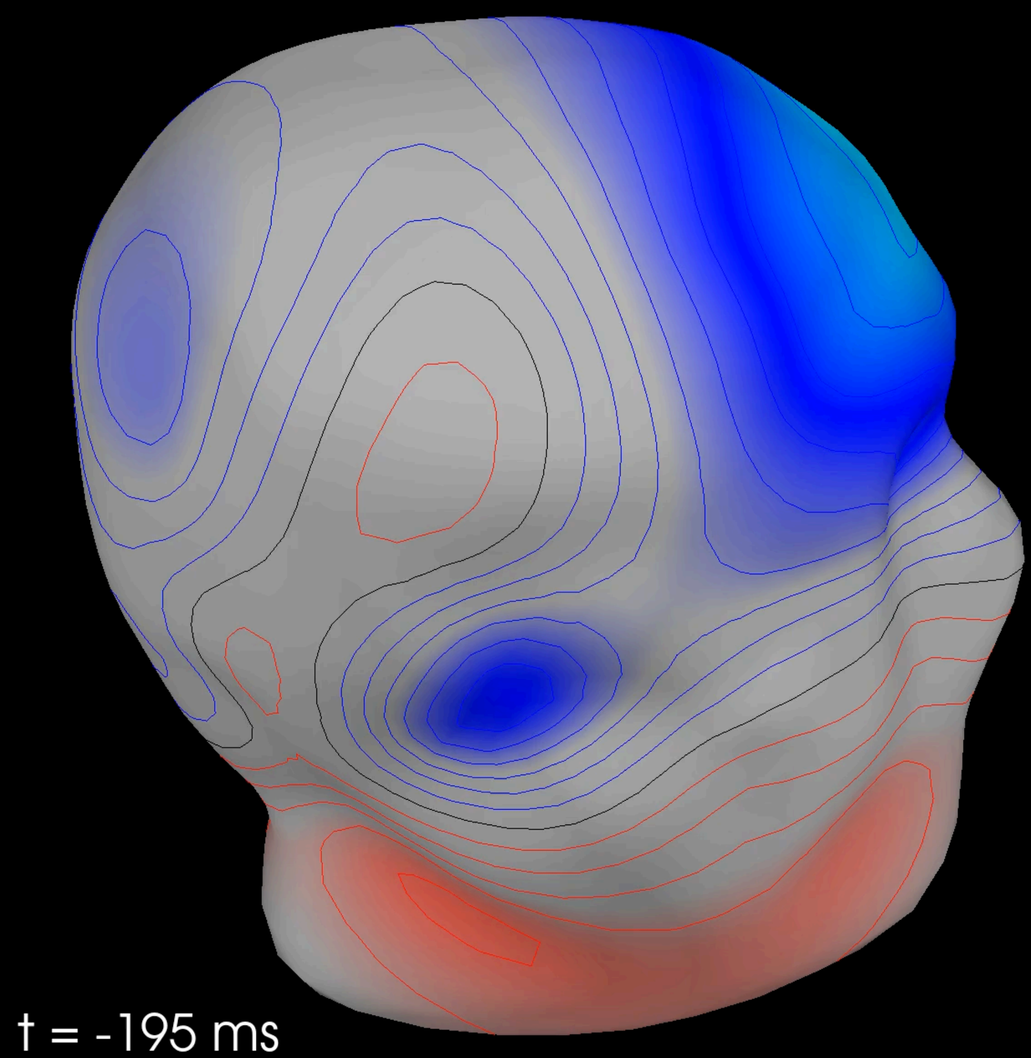
Electroencephalography (EEG)
Electric Potential [V]



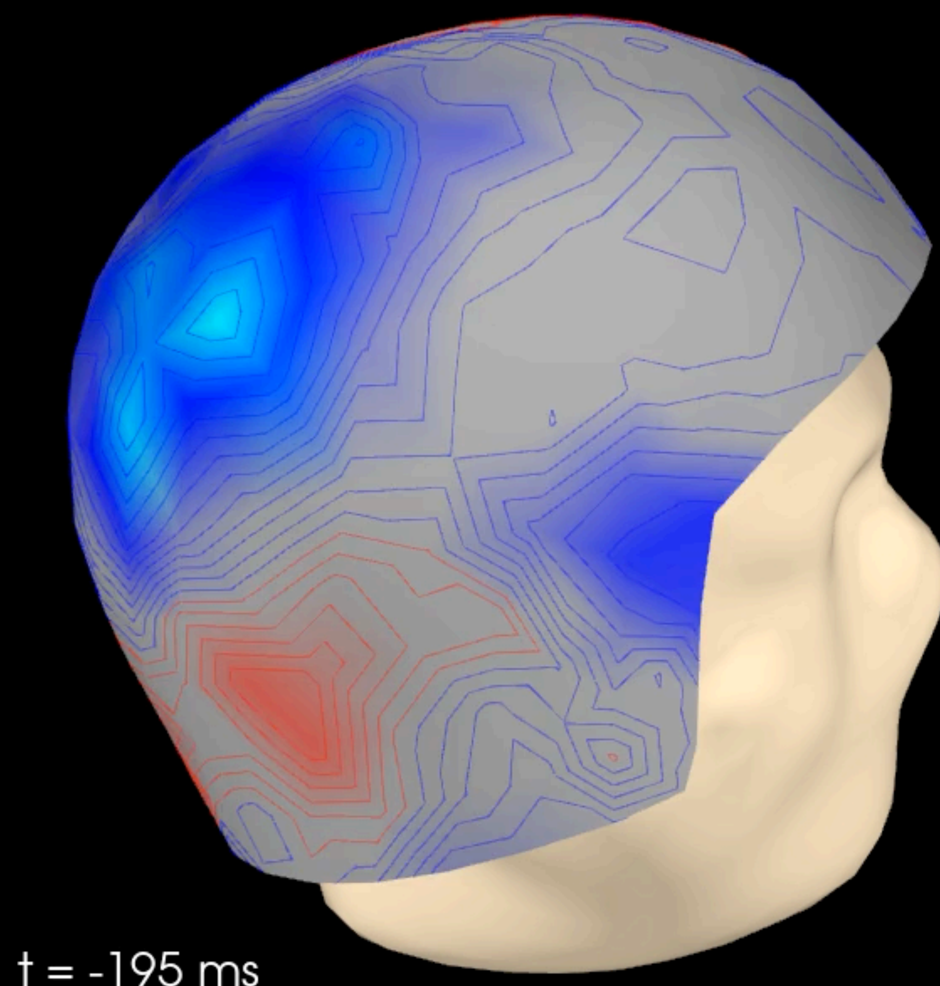
Magnetoencephalography (MEG)
Magnetic field [T]

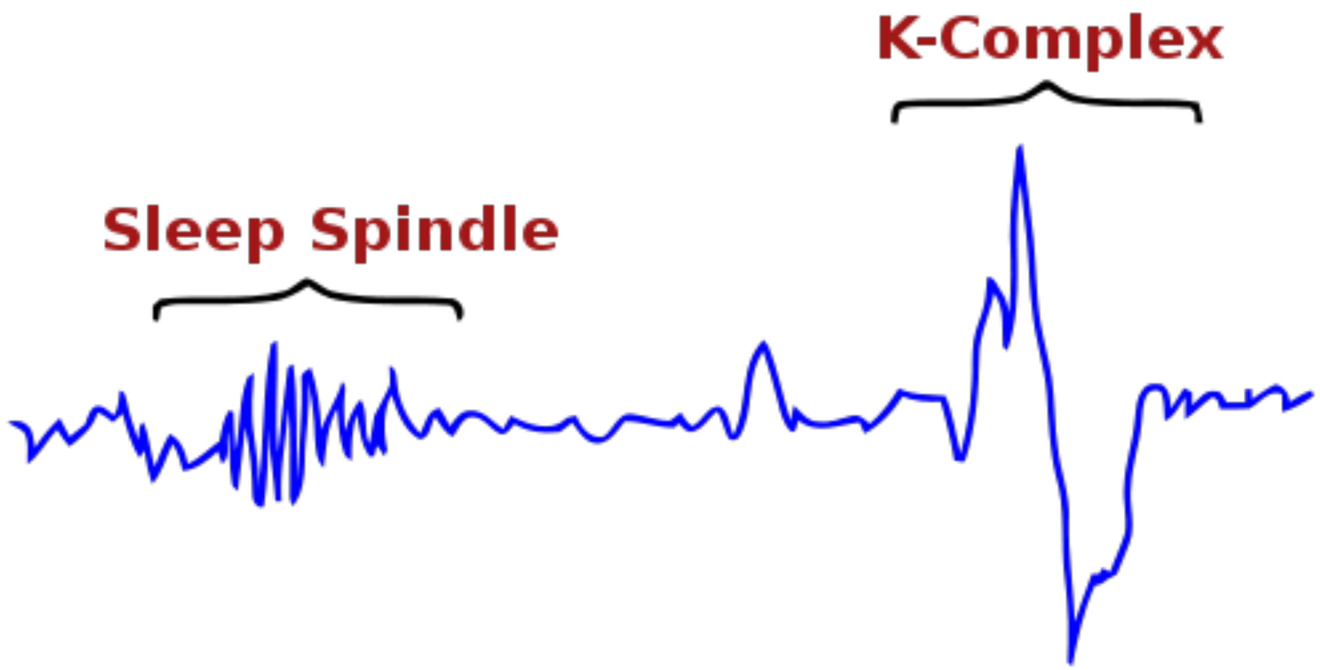


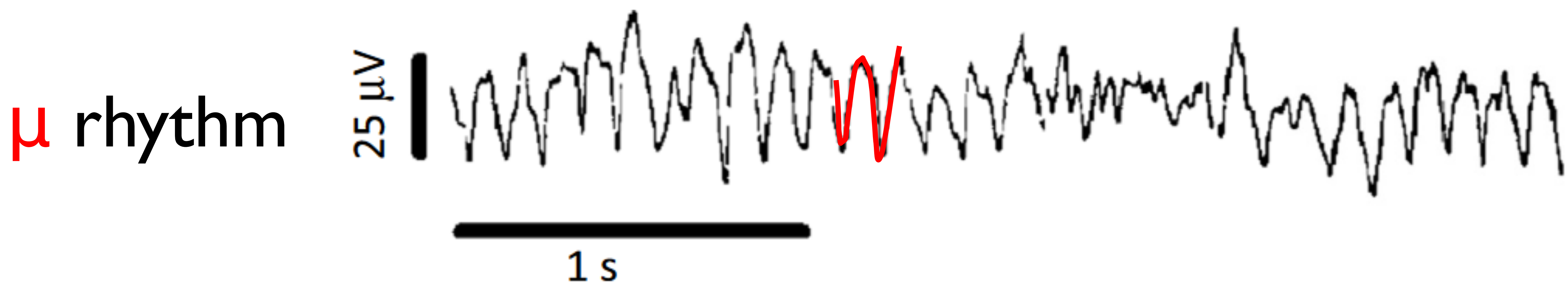
Electroencephalography (EEG)
Electric Potential [V]



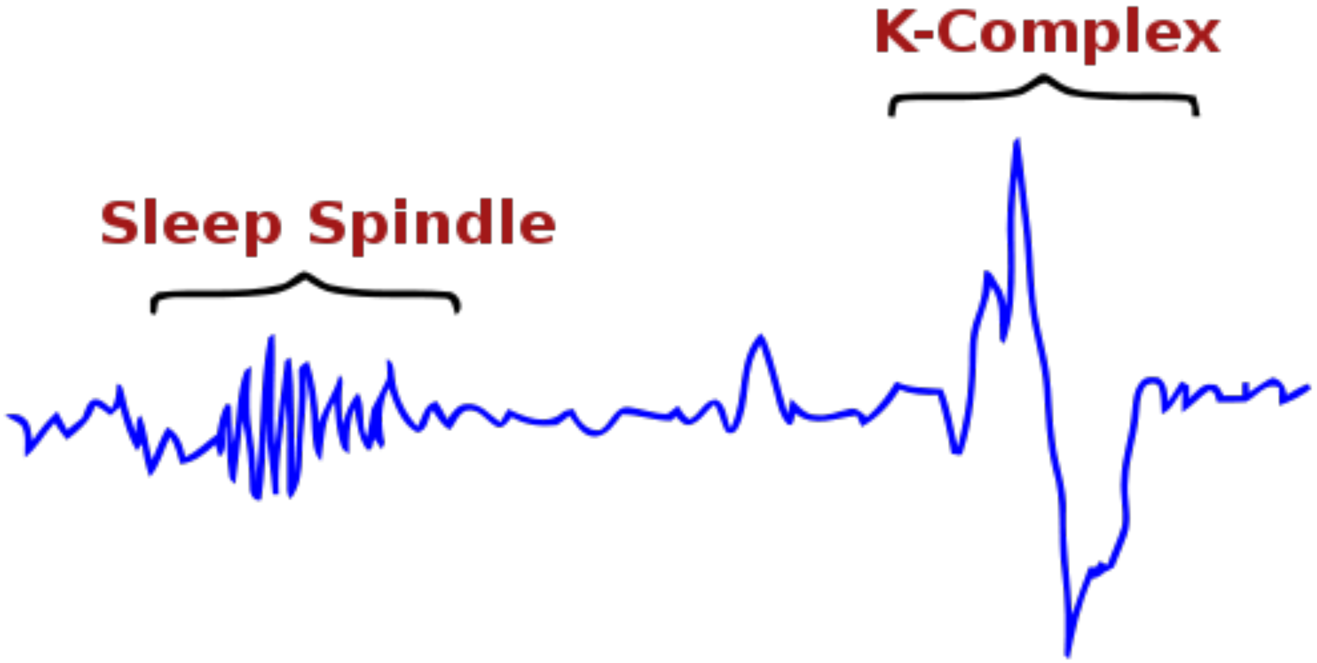
Magnetoencephalography (MEG)
Magnetic field [T]

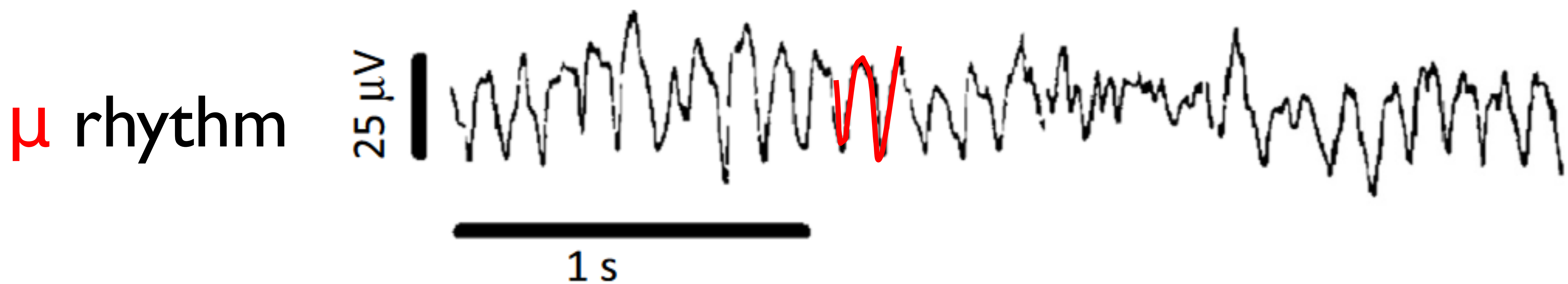






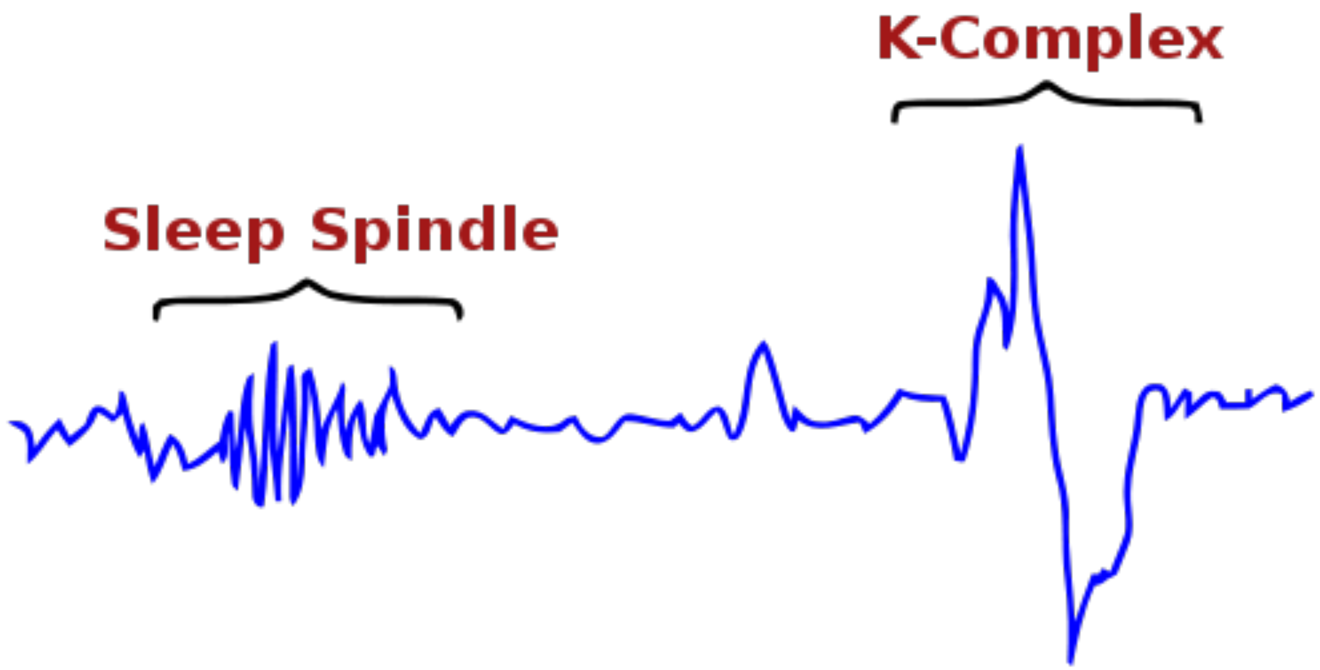
[T. Dupré la Tour, L. Tallot, L. Grabot, V. Doyère, V. van Wassenhove, Y. Grenier, A. Gramfort, (2017) PLOS Computational biology]



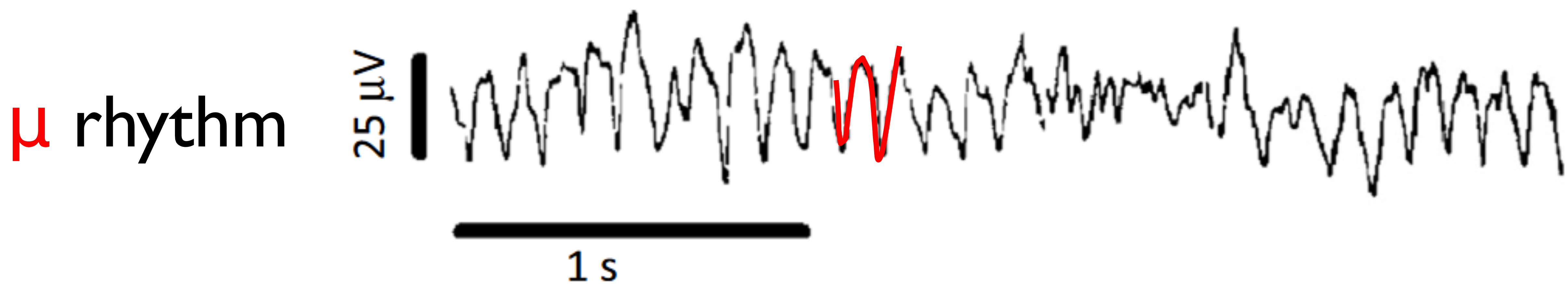


CFC: High frequency bursts coupled with slow waves

[T. Dupré la Tour, L. Tallot, L. Grabot, V. Doyère, V. van Wassenhove, Y. Grenier, A. Gramfort, (2017) PLOS Computational biology]



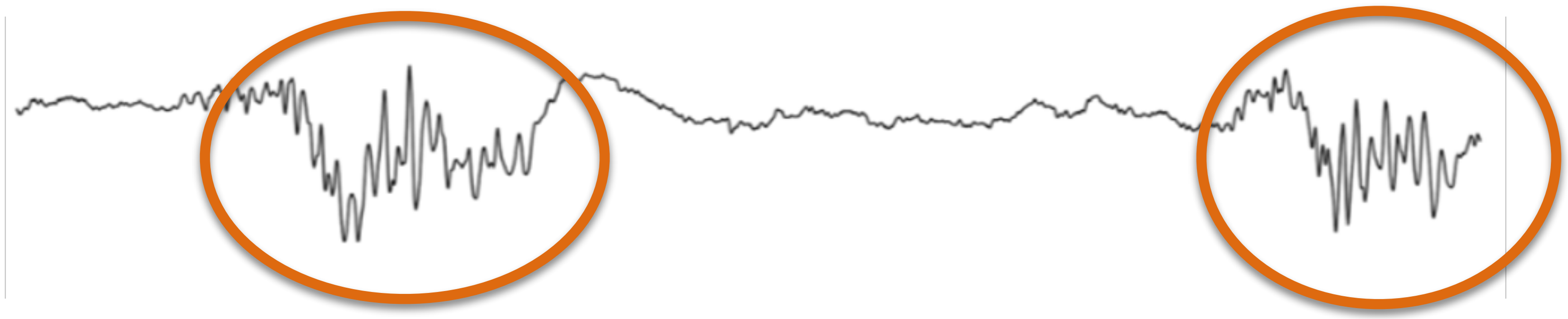
Neural signals exhibit diverse and complex morphologies



CFC: High frequency bursts coupled with slow waves

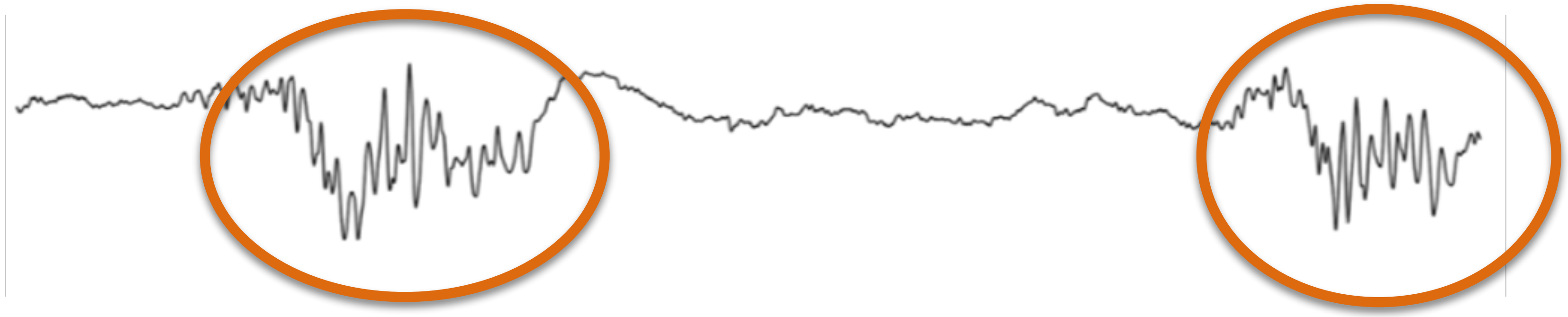
[T. Dupré la Tour, L. Tallot, L. Grabot, V. Doyère, V. van Wassenhove, Y. Grenier, A. Gramfort, (2017) PLOS Computational biology]

(abnormal) EEG during during general aneasthesia (GA)

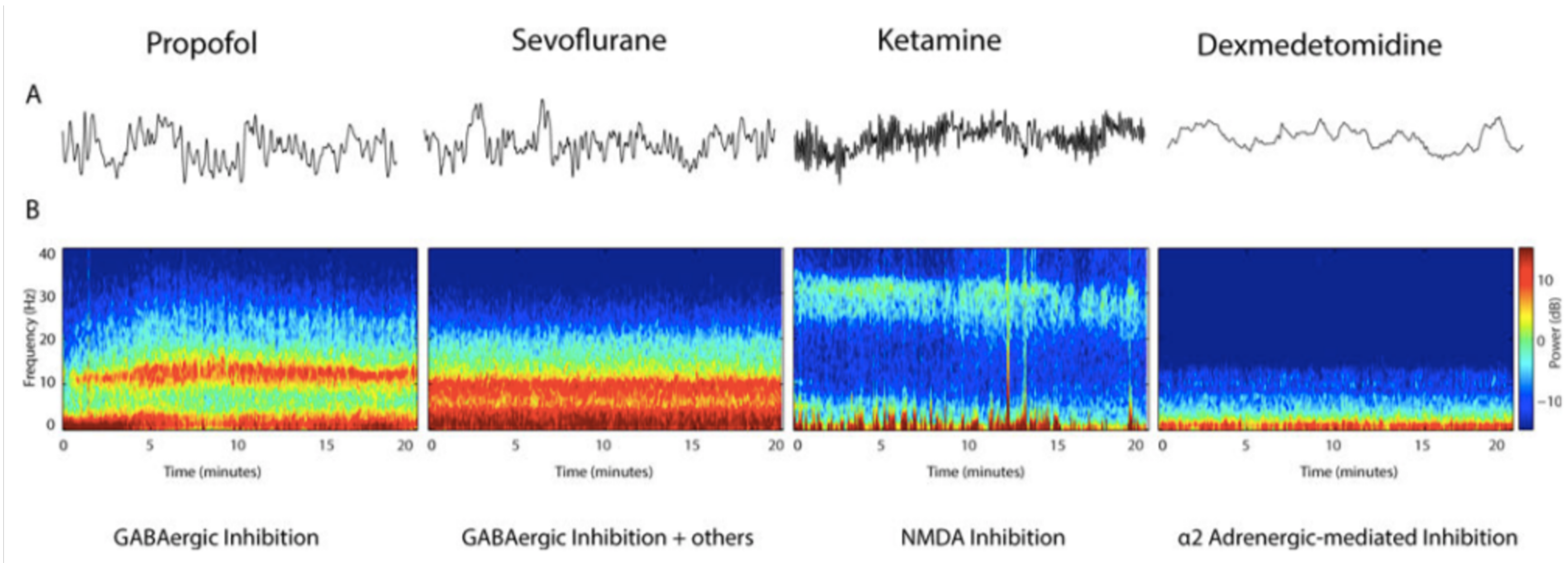


Burst suppression

(abnormal) EEG during during general aneasthesia (GA)

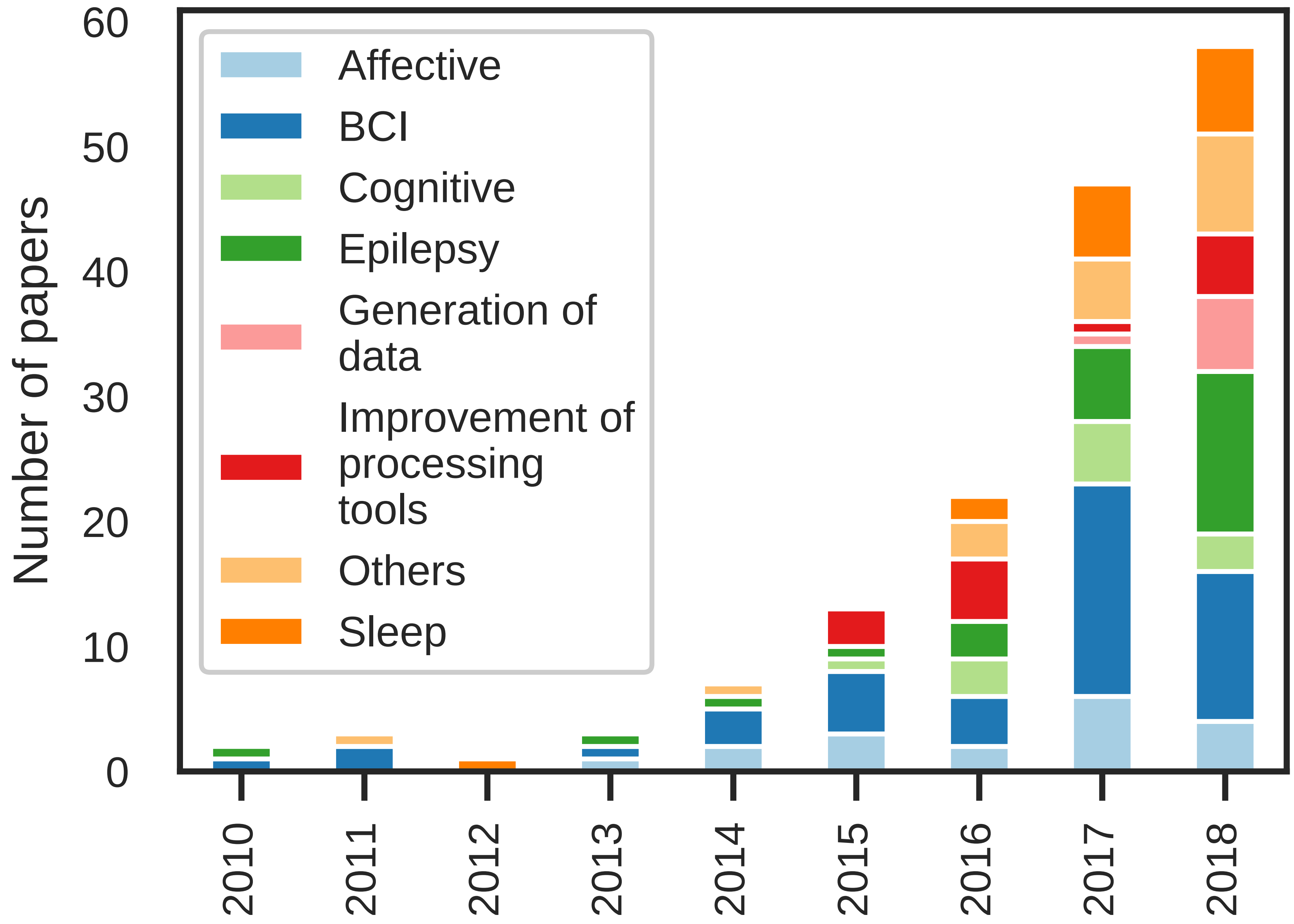


Burst suppression



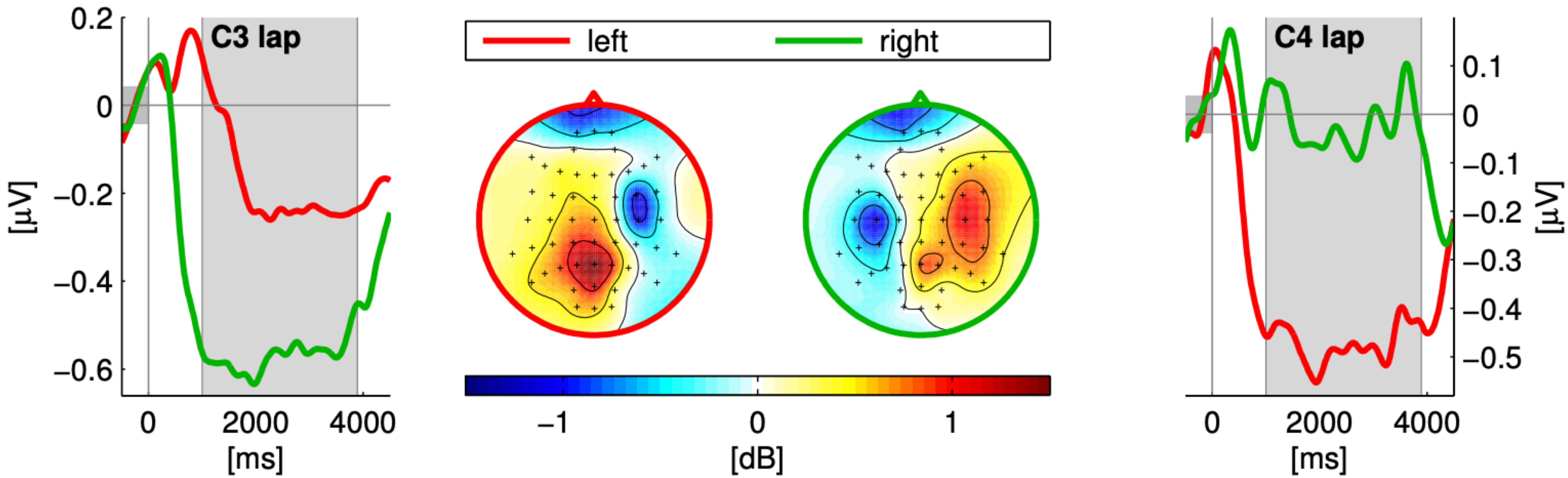
[Purdon, Anesthesia 2015]

Deep Learning papers on EEG



[Deep learning-based electroencephalography analysis: a systematic review
Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. and Faubert, J. (2019)
Journal of Neural Engineering 16: (051001).]

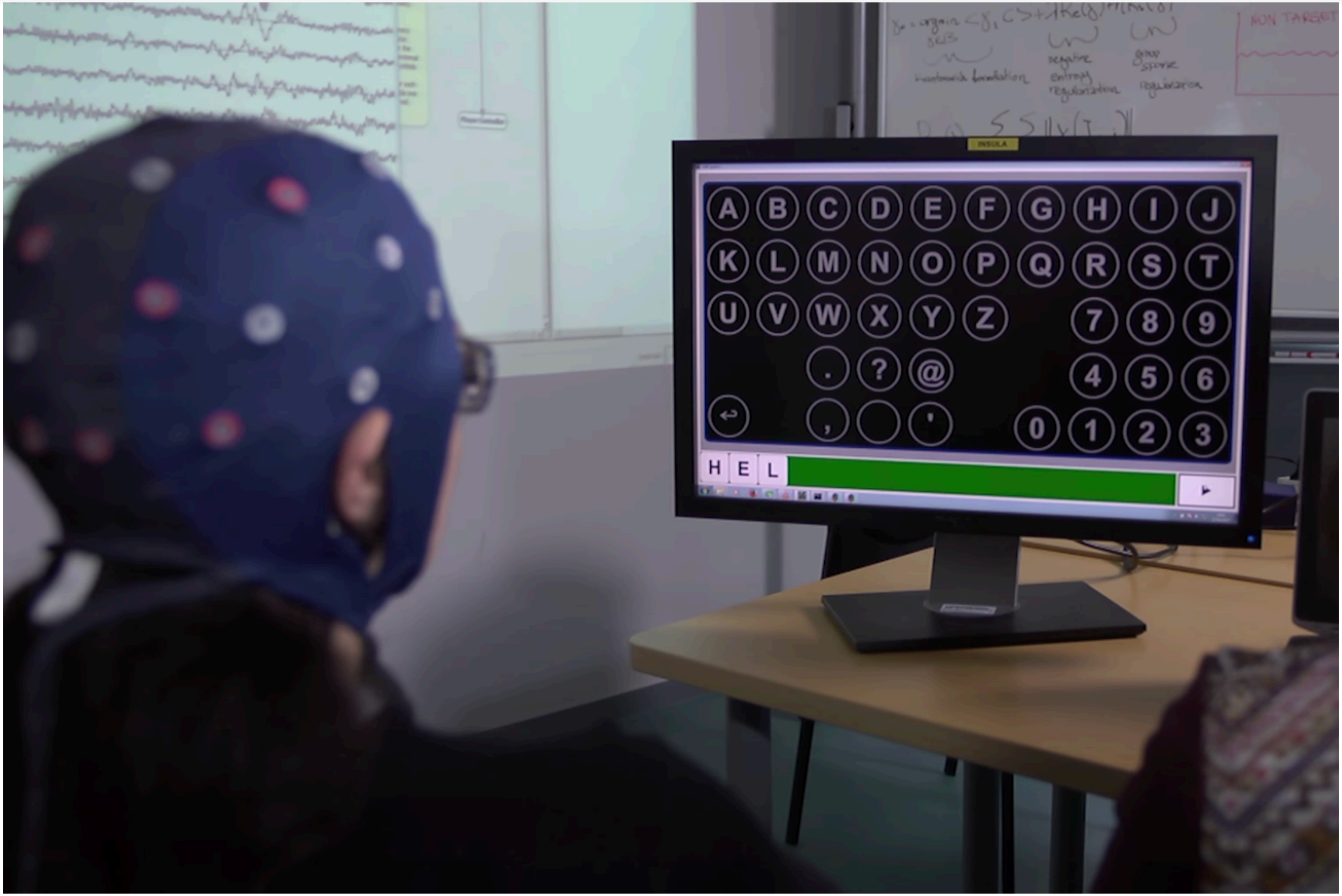
Example: Motor Imagery



Event-Related Desynchronization (ERD) during motor imagery of the left and the right hand. Averaged signal Hilbert enveloppe between 9 and 13Hz on C3 and C4 (After Laplace / Current Source Density transform)

[Blankertz et al. 2008]

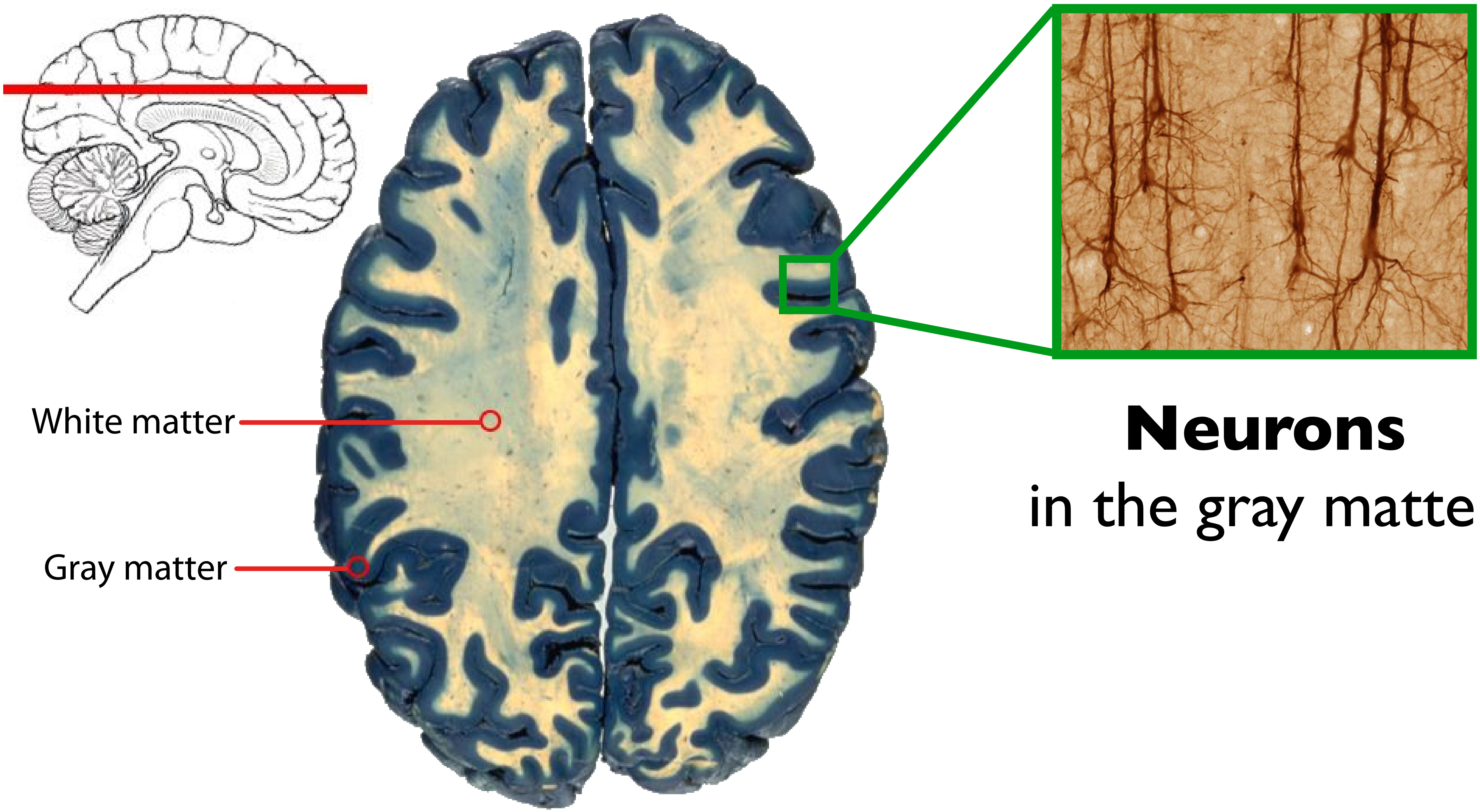
Example: P300 Speller



https://www.youtube.com/watch?v=1WzU_8_qA6w

Part 1:
Electrophysiology:
Origin of the signals

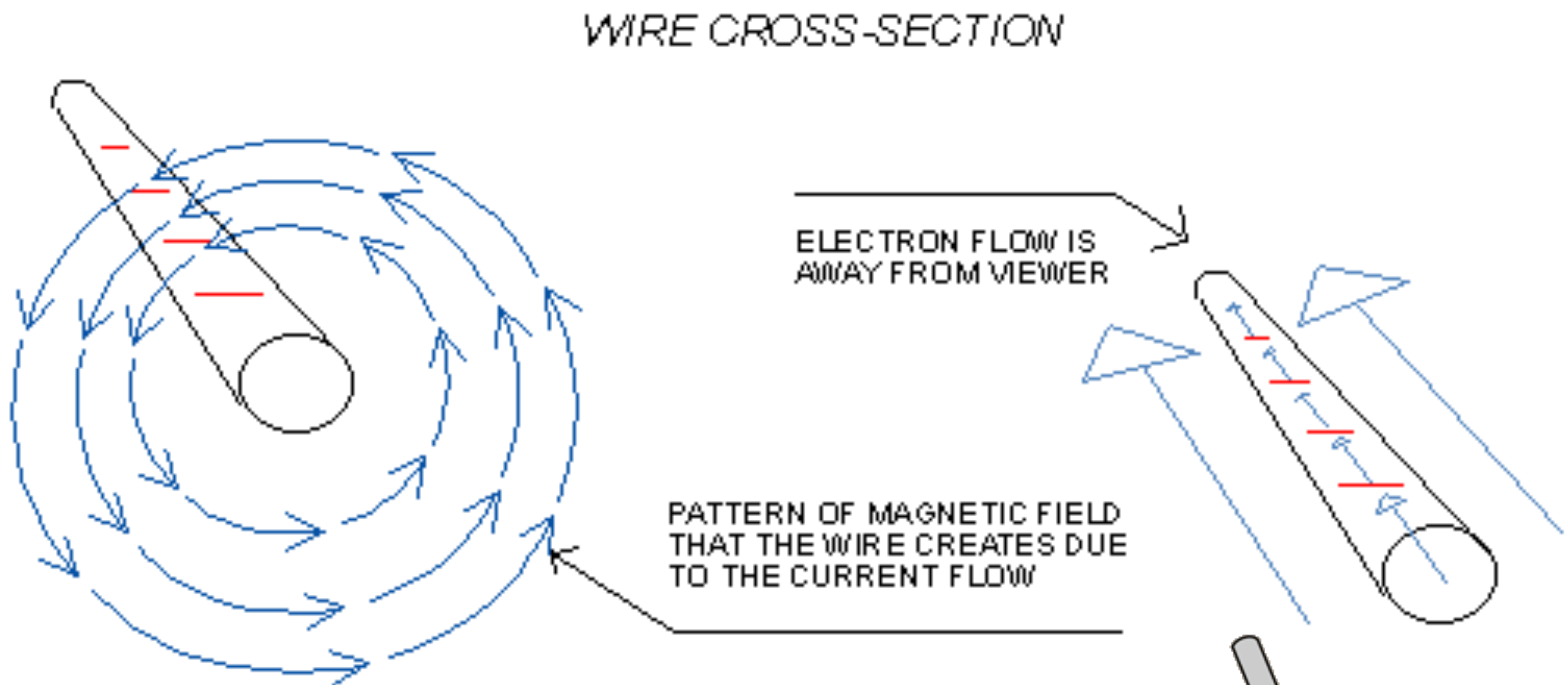
Brain anatomy

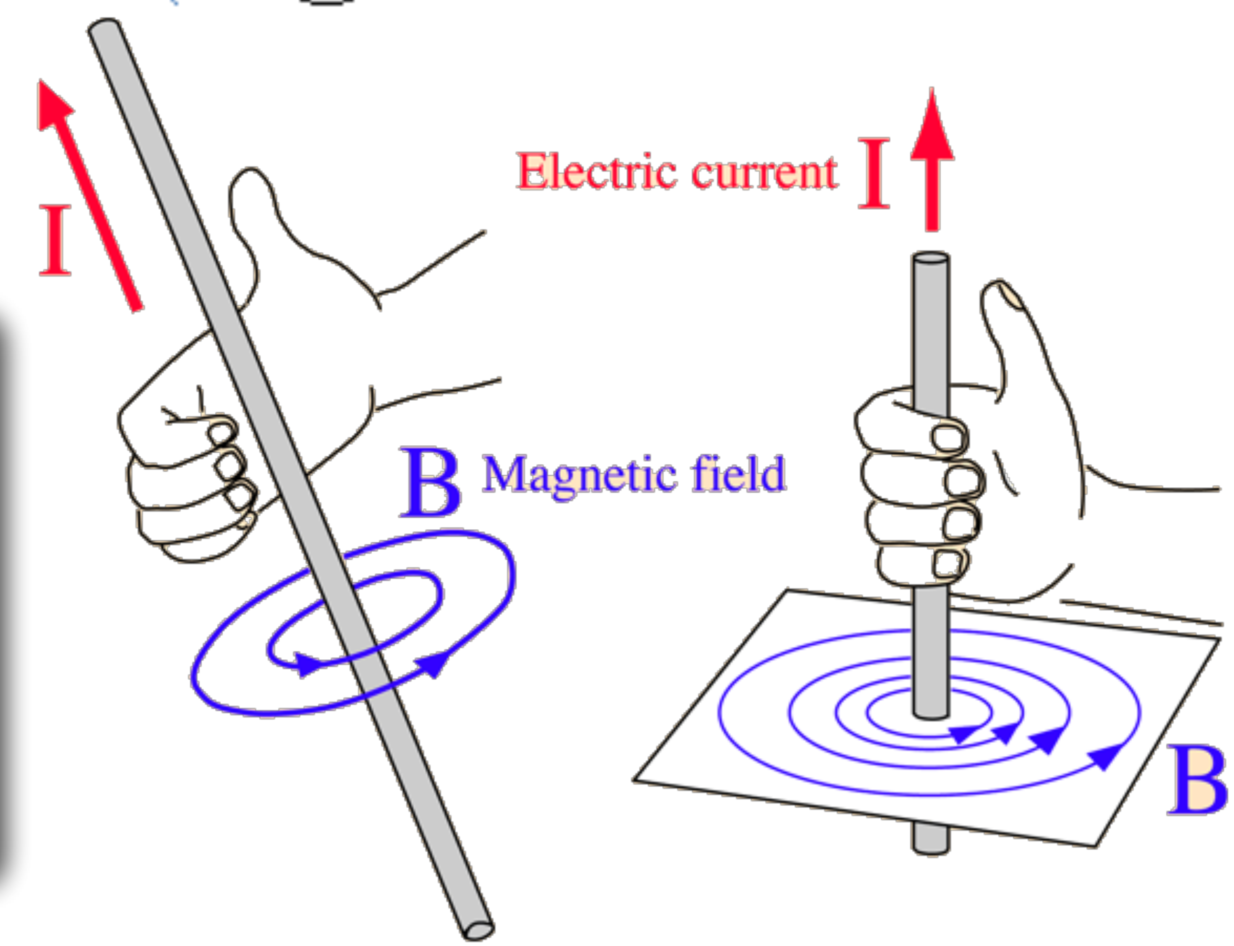


Neurons
in the gray matter

Source: dartmouth.edu

Background on electromagnetism

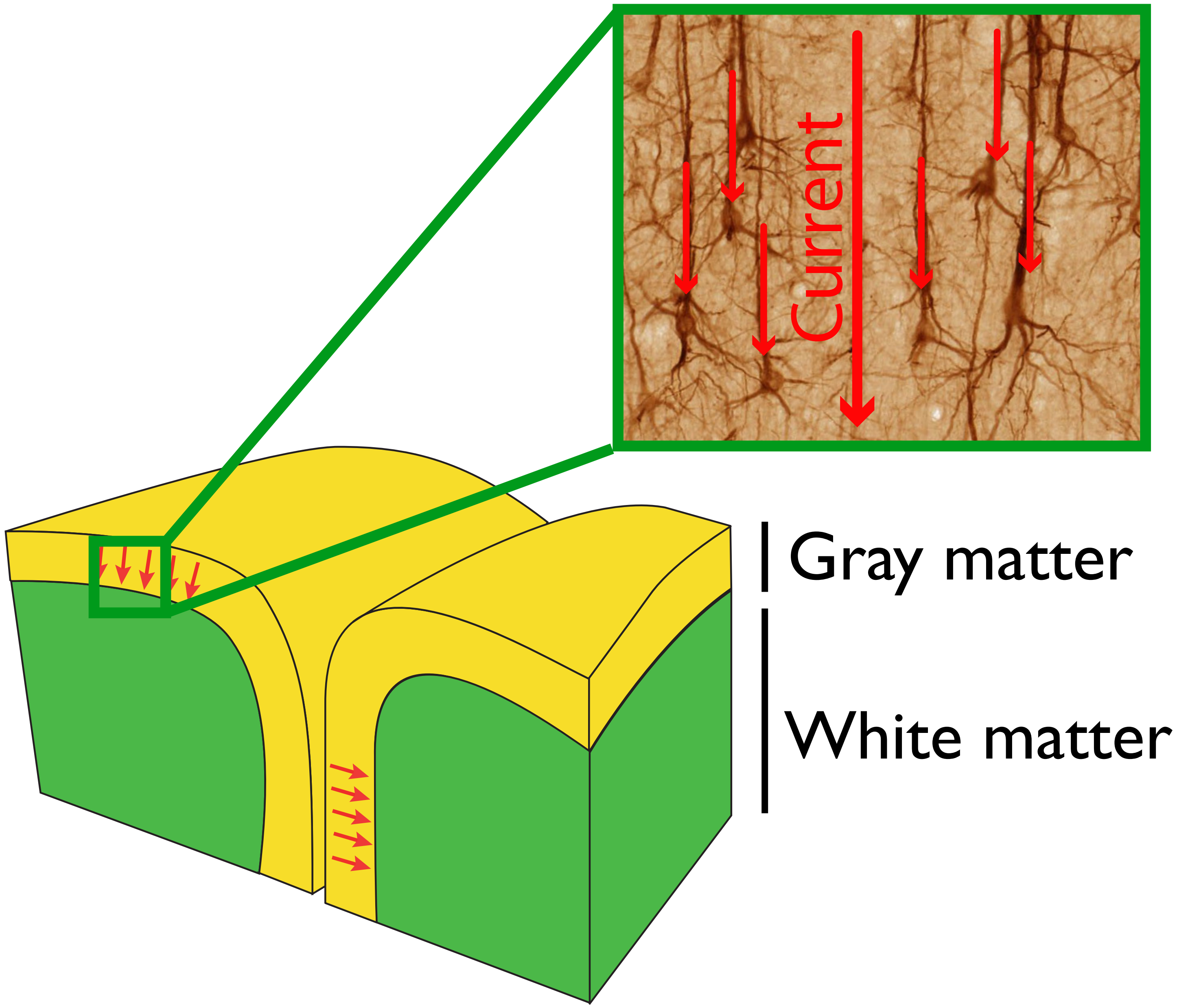




moving charges on a wire induces an electo-magnetic field

Neurons as current generators

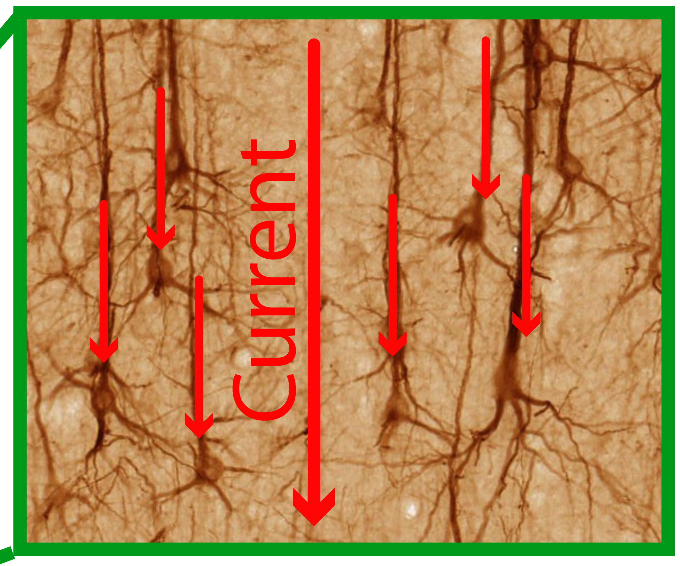
Large cortical pyramidal cells organized in macro-assemblies with their **dendrites normally oriented to the local cortical surface**

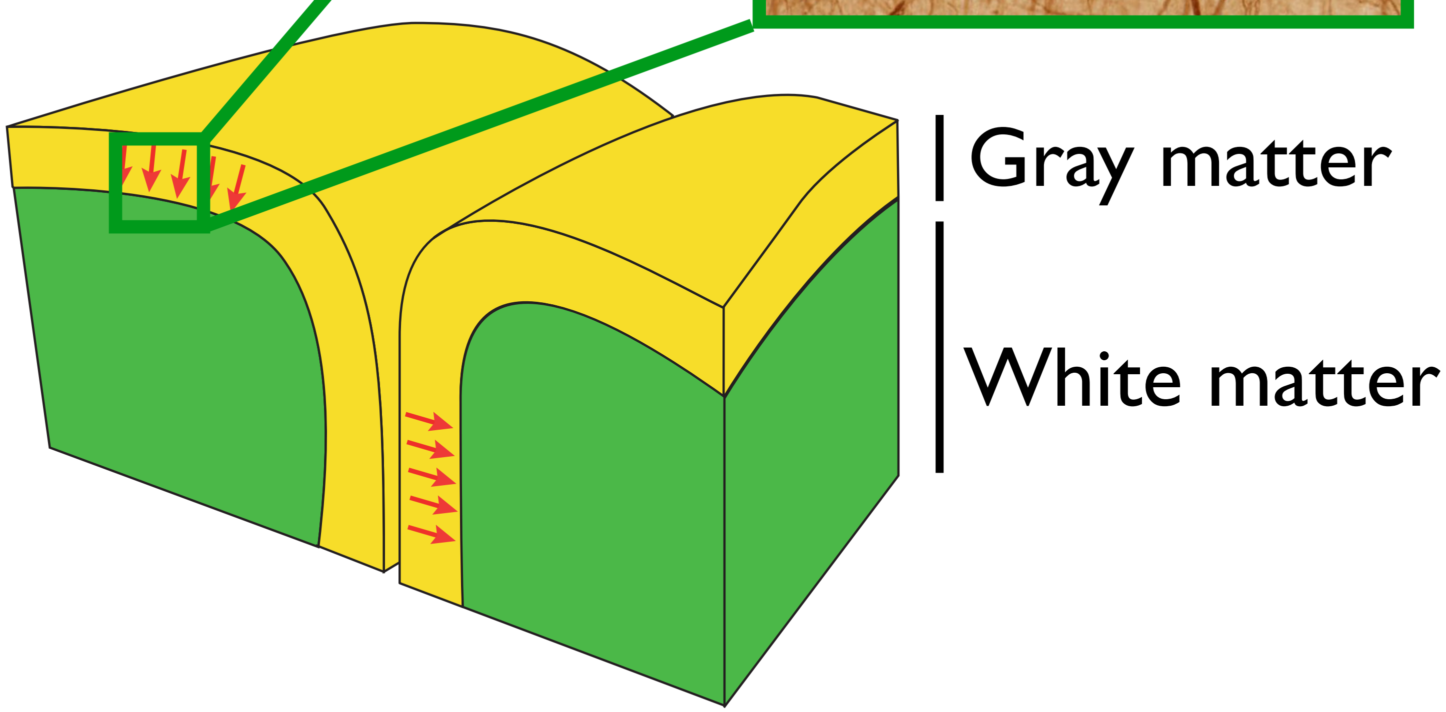


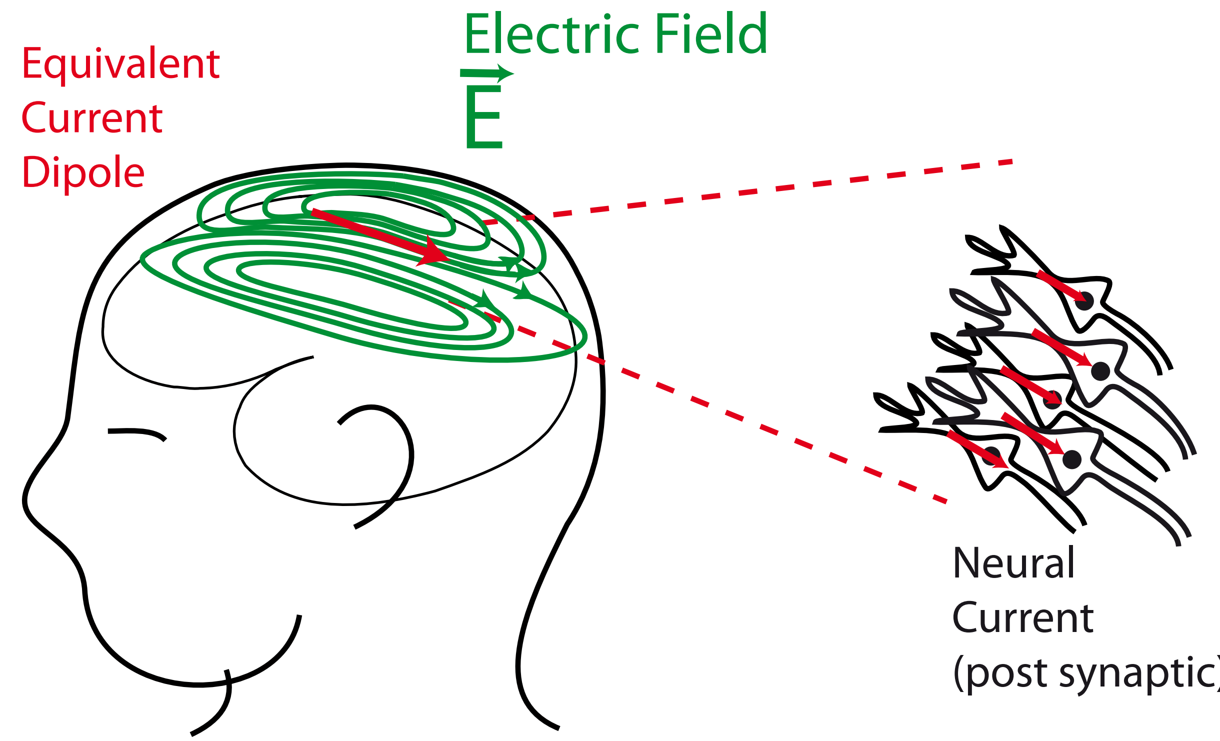
Neurons as current generators

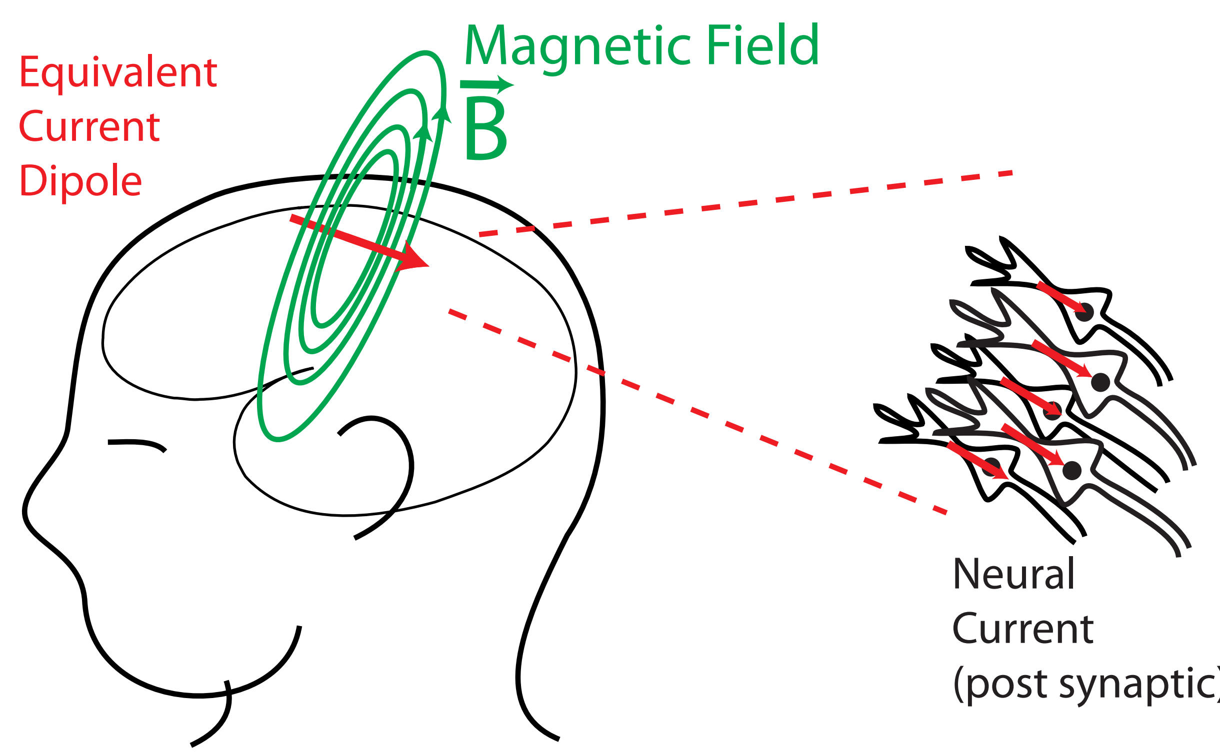
Large cortical pyramidal cells organized in macro-assemblies with their **dendrites normally oriented to the local cortical surface**

Q = I × d
(10 to 100 nAm) with the electric current dipole (ECD) model

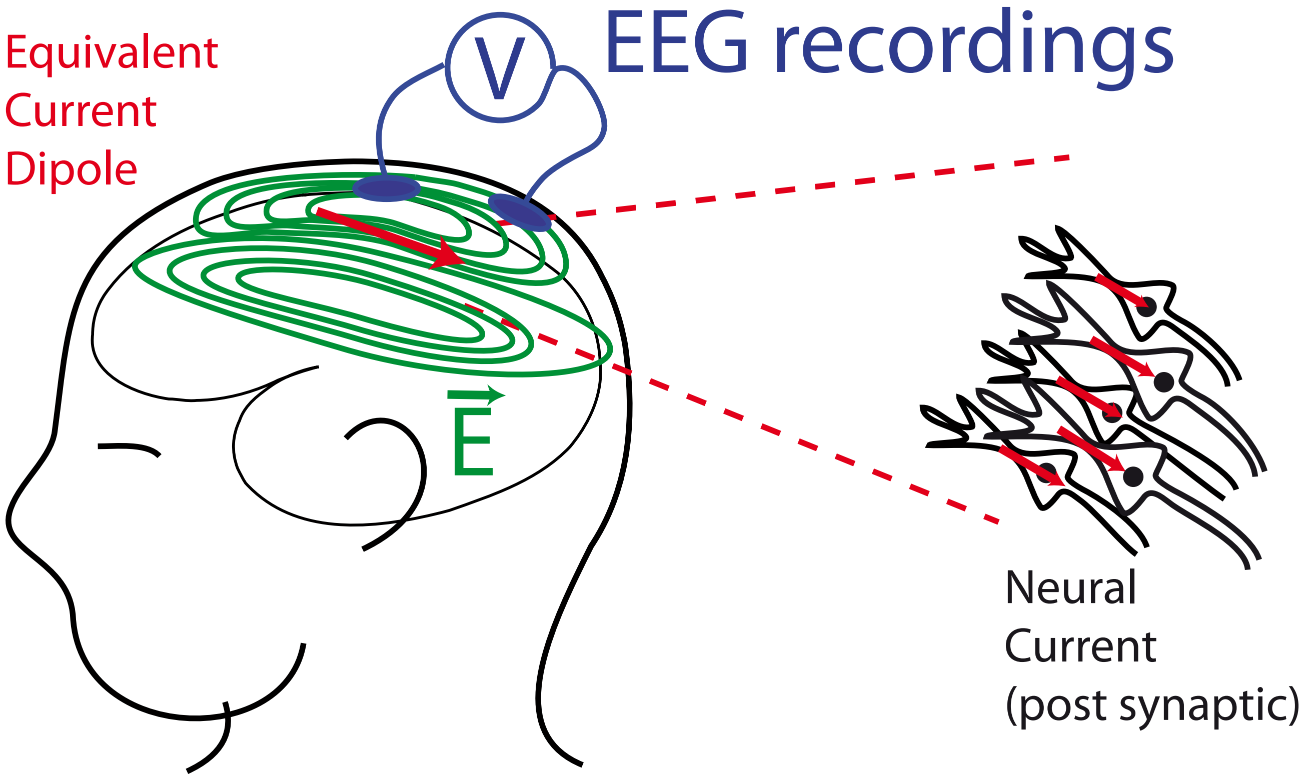


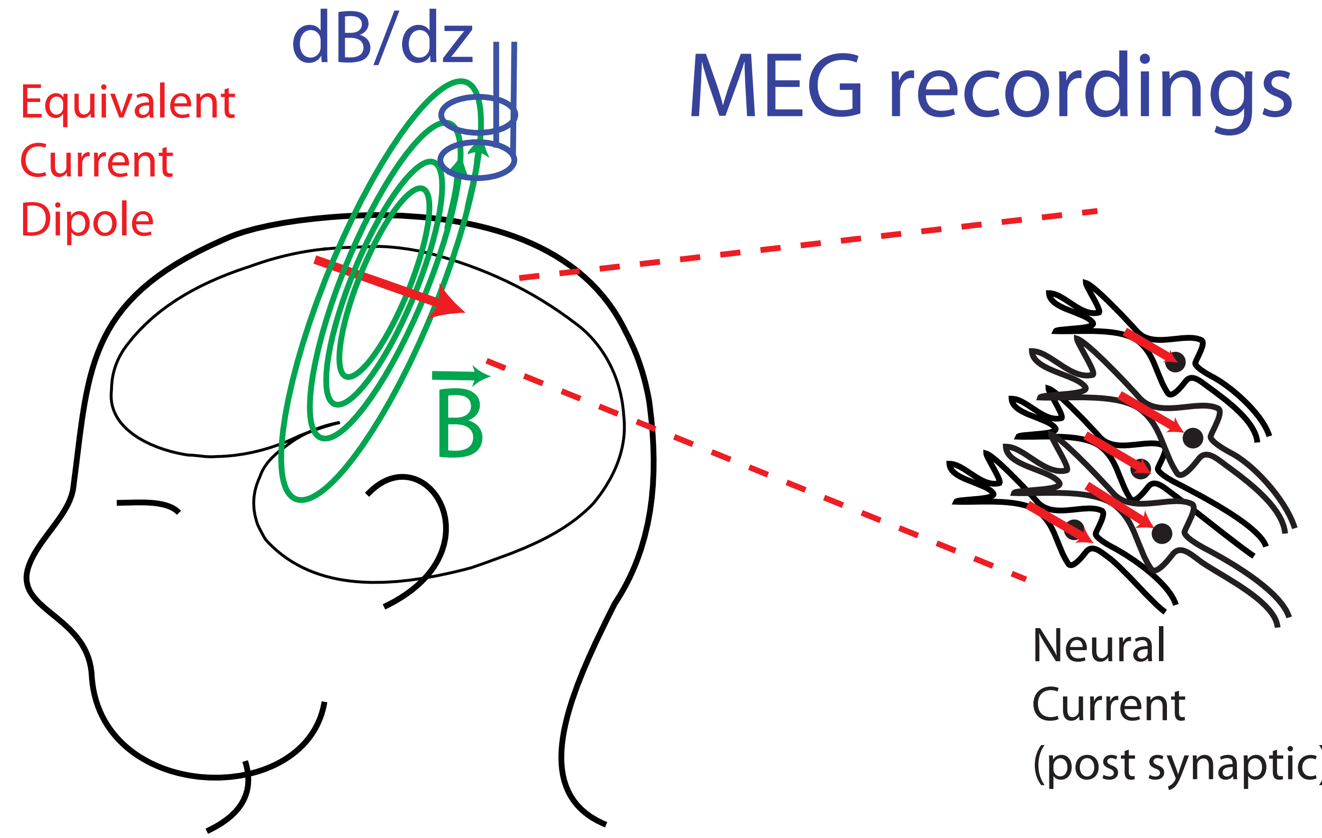




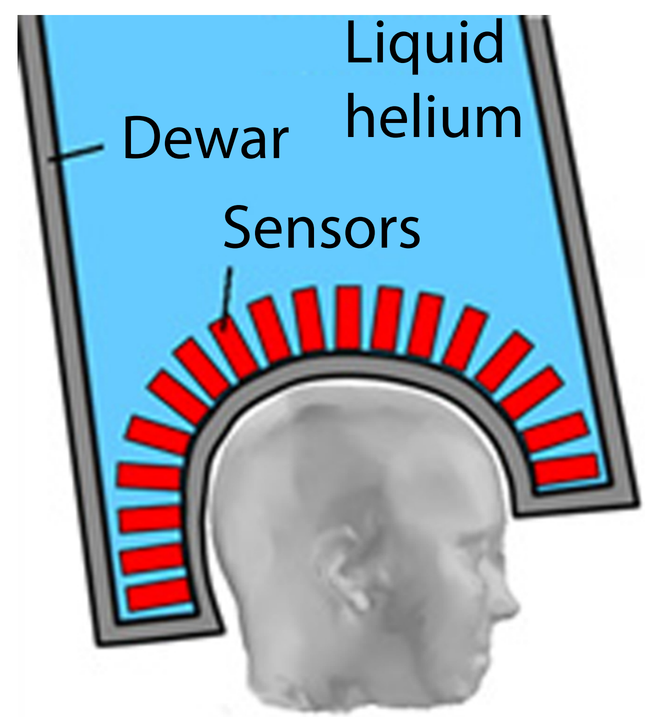


EEG & MEG systems



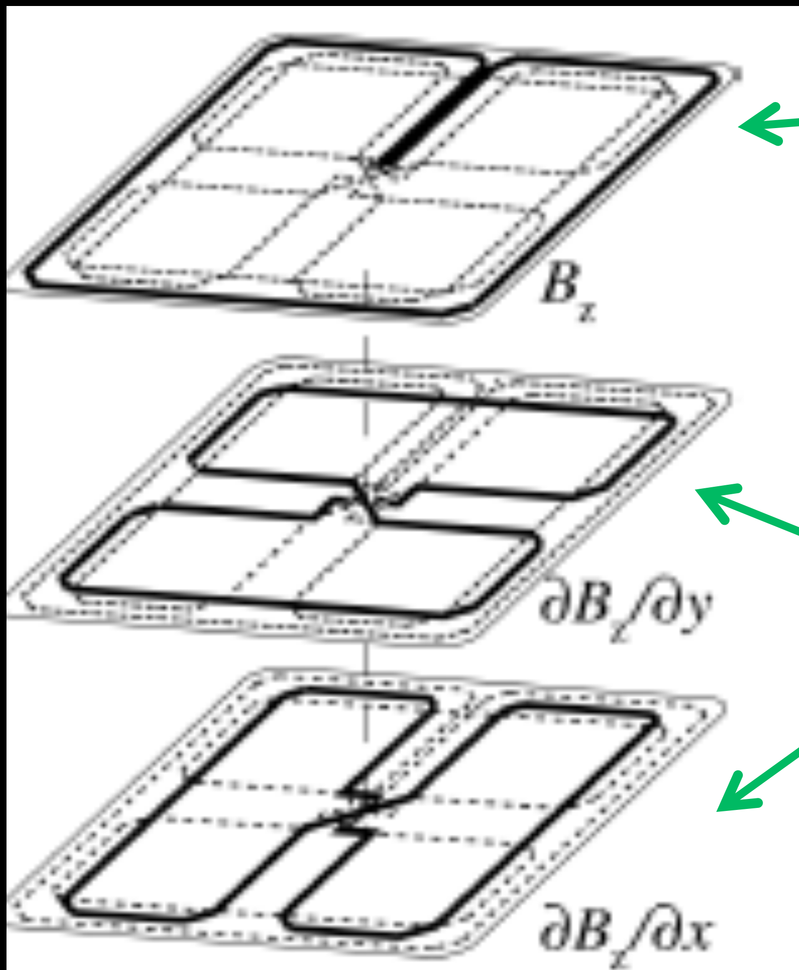


First EEG recordings in 1929 by H. Berger



Hôpital La Timone Marseille, France

MEG sensors

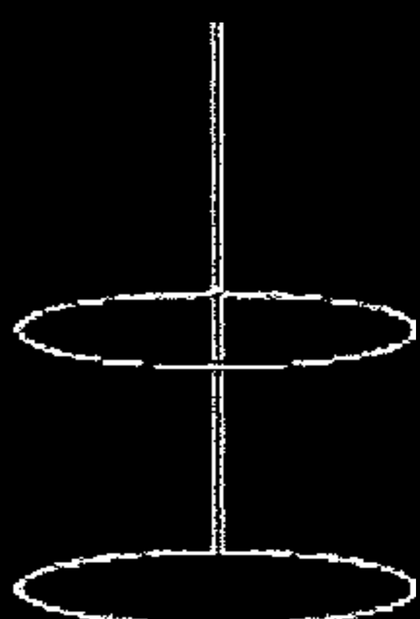


Magnetometer

- General magnetic fields
- Very sensitive overall, **noisy**

Planar Gradiometer

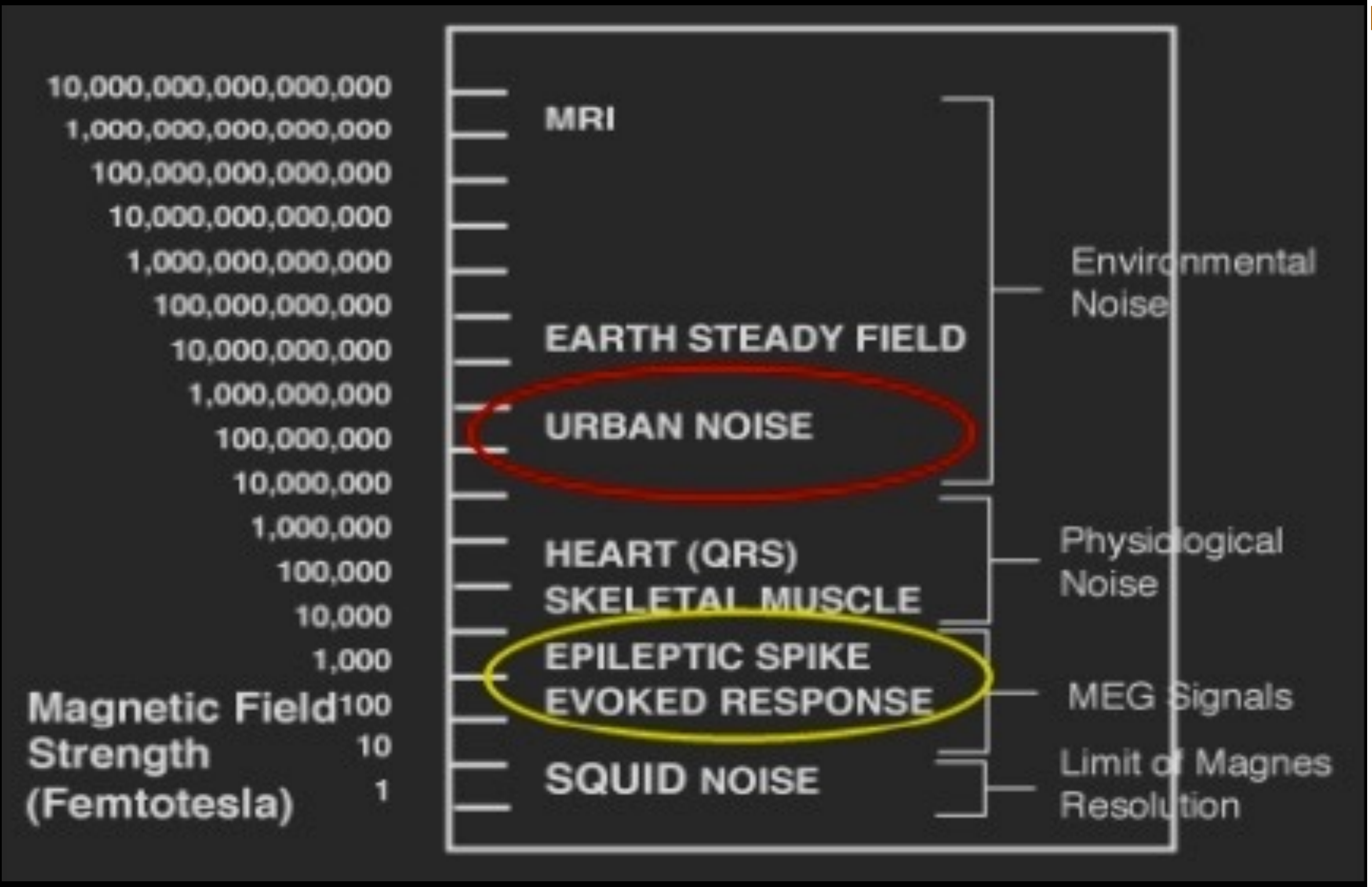
- Focal magnetic fields
- Most sensitive to fields directly underneath

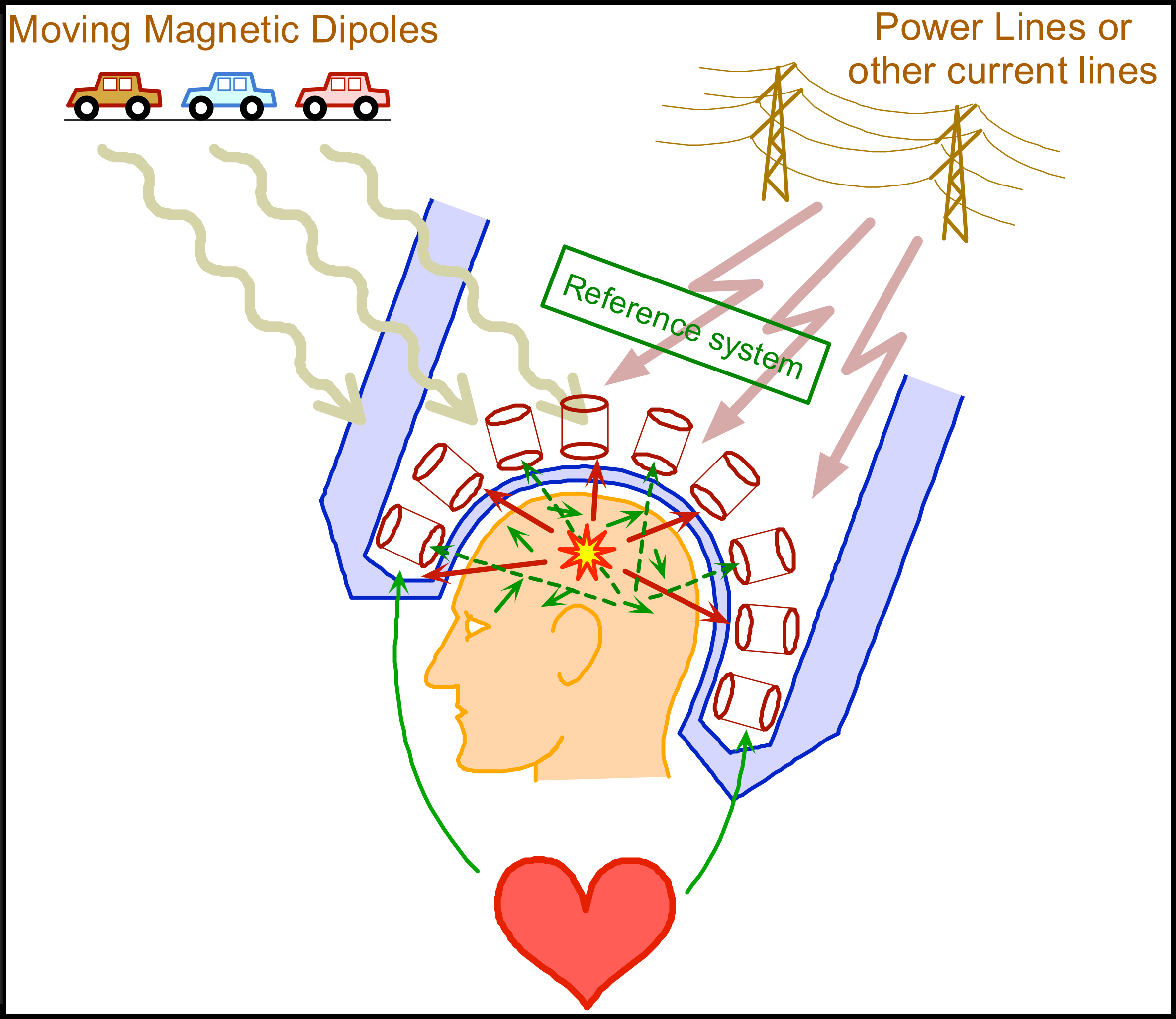


Axial Gradiometer

- Focal magnetic fields
- Most sensitive to fields directly underneath it

Magnetic shielding

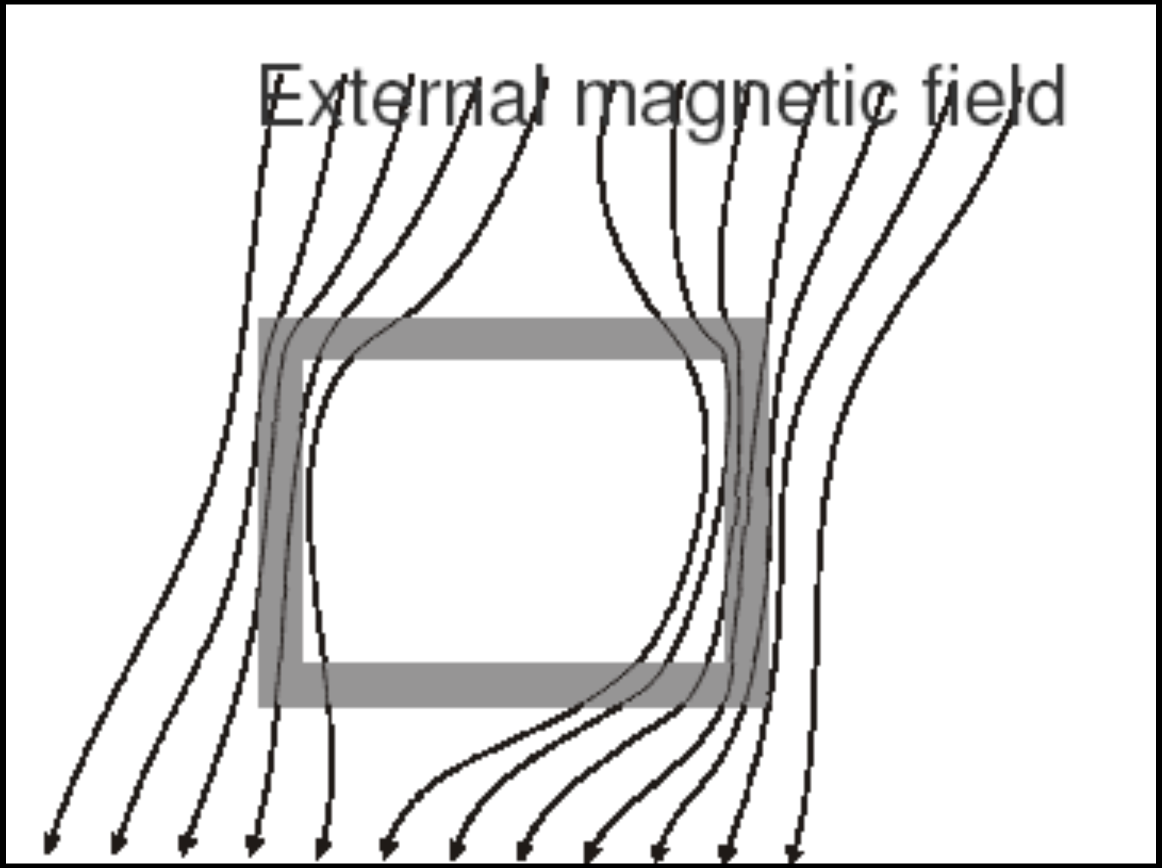




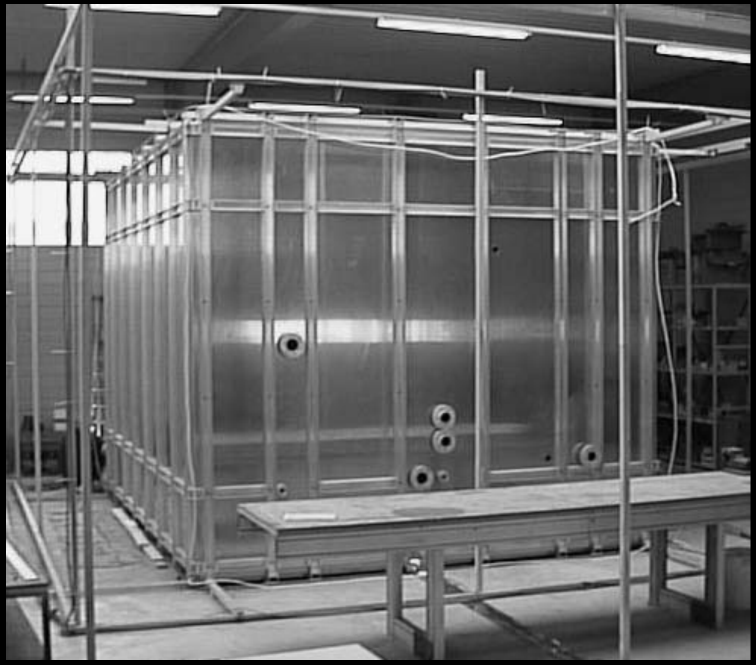
Hence the importance of shielding...

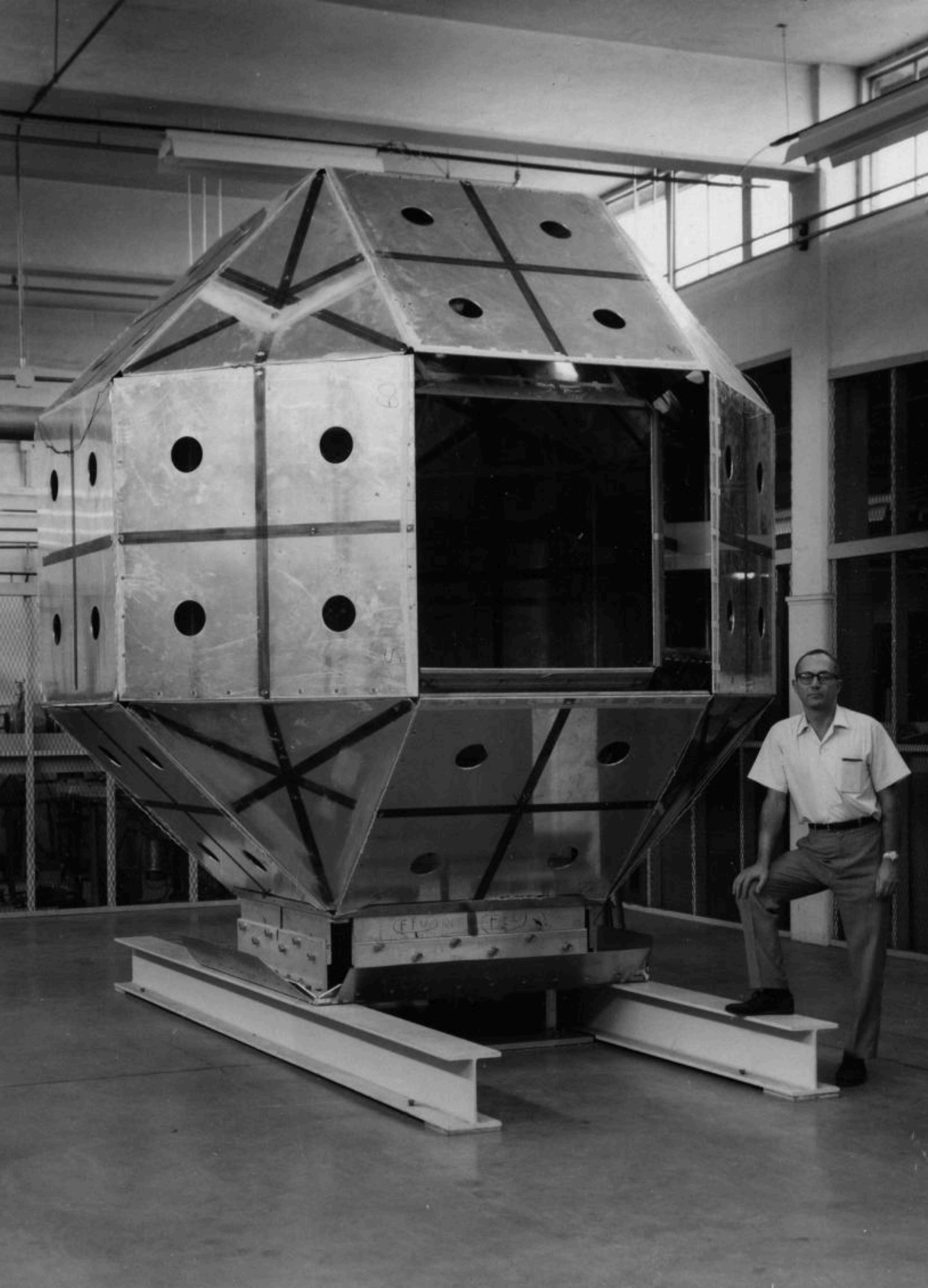
Magnetic shielding

Magnetically Shielded Room (MSR)



3-ply µ-metal room





In late 1969, David Cohen and James Zimmerman ushered in a new era of biomagnetism when they measured the magnetic field from a human heart using a superconducting quantum interference device, or SQUID. The first use of a SQUID with a human subject, the experiment threw open the doors to medical applications of the technology, the possibilities of which researchers are still exploring today.

Source: https://www.nmr.mgh.harvard.edu/news/141119/45-years-of-biomagnetism-and-squid

Cohen 1971
First unaveraged MEG recording

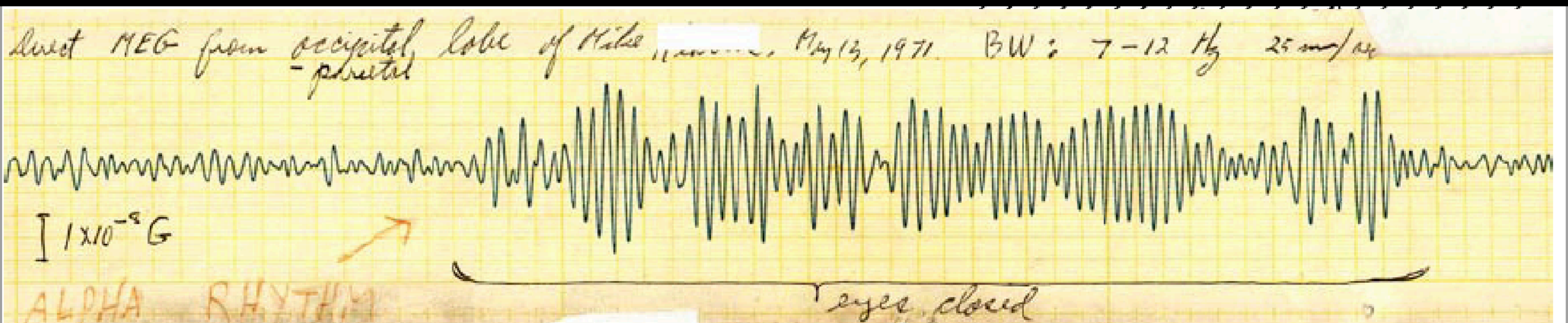


Figure 3. First MEG measured with a SQUID, in the MIT shielded room (May, 1971), using an early commercial SQUID. The subjects eyes were open at the beginning of the trace, then closed, resulting in the large alpha rhythm, then open again. The MEG was now as clear as the conventional EEG.

A machine (Neuromag vectorview)

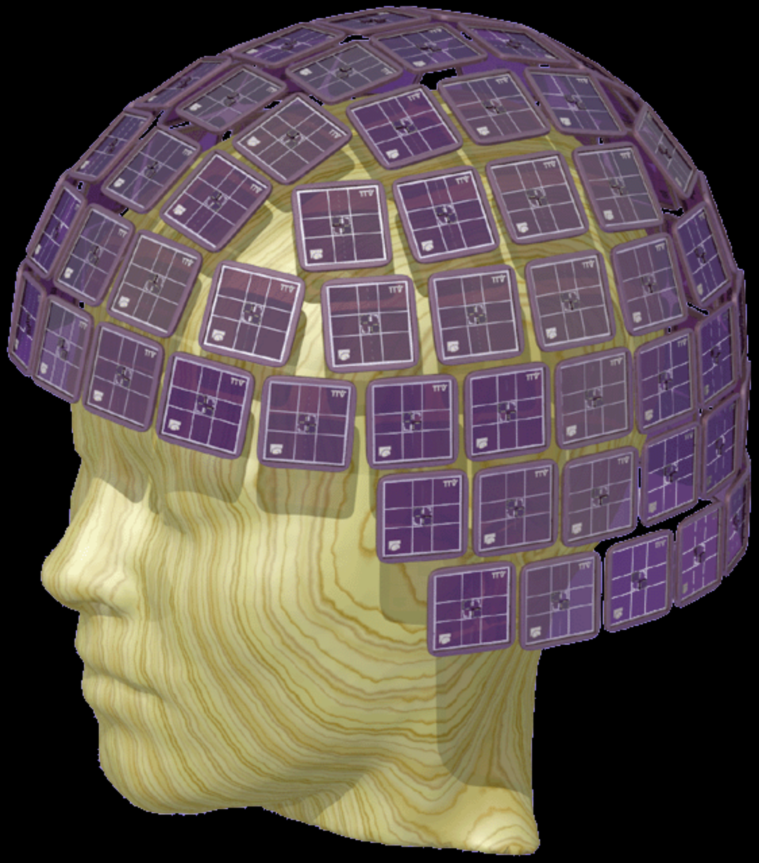
No Magnet

Quiet

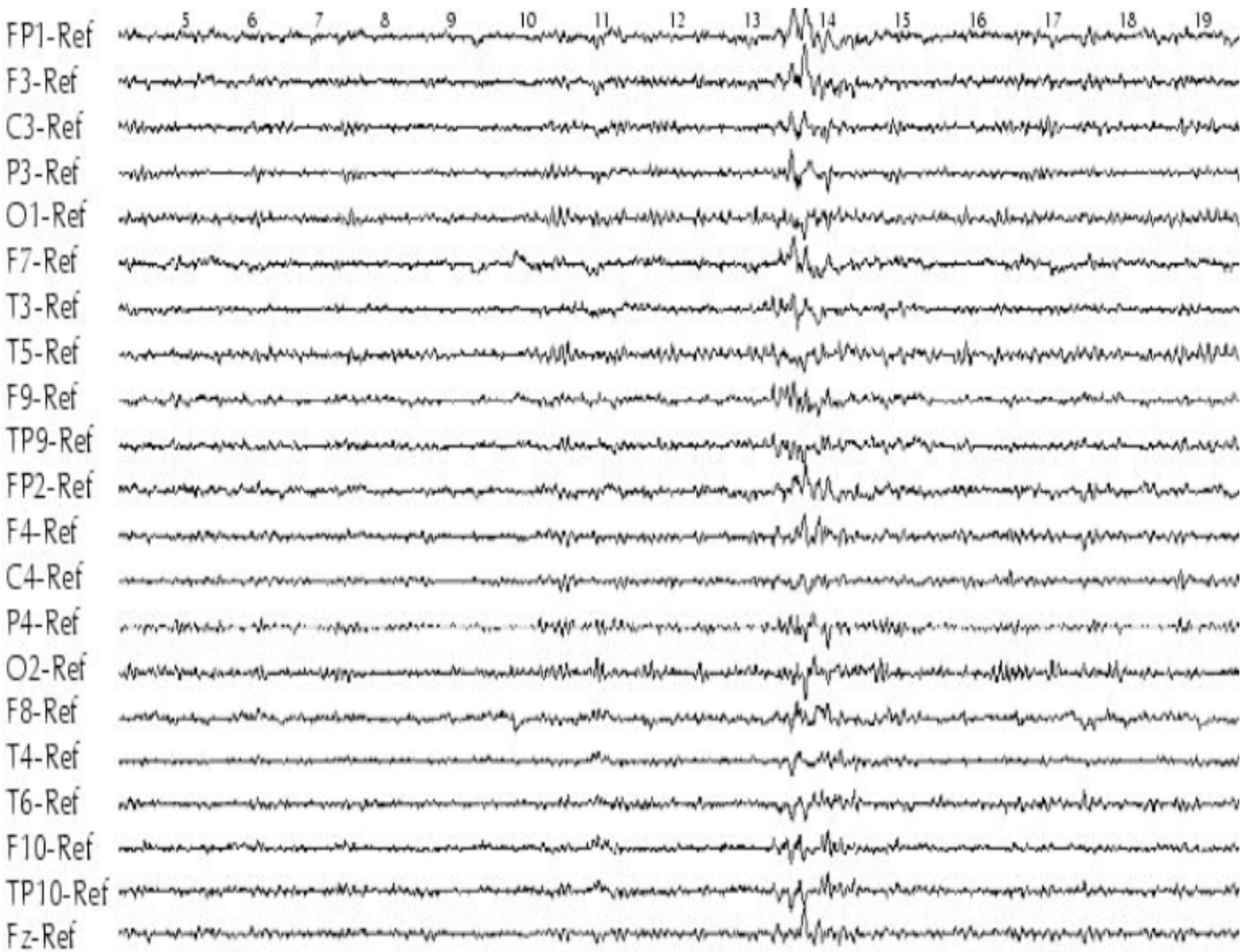
Machine makes no noise

Participant can sit or lay down

Can record 128 EEG simultaneously



M/EEG Measurements



Sample EEG measurements

EEG :

- ≈ 100 sensors

MEG :

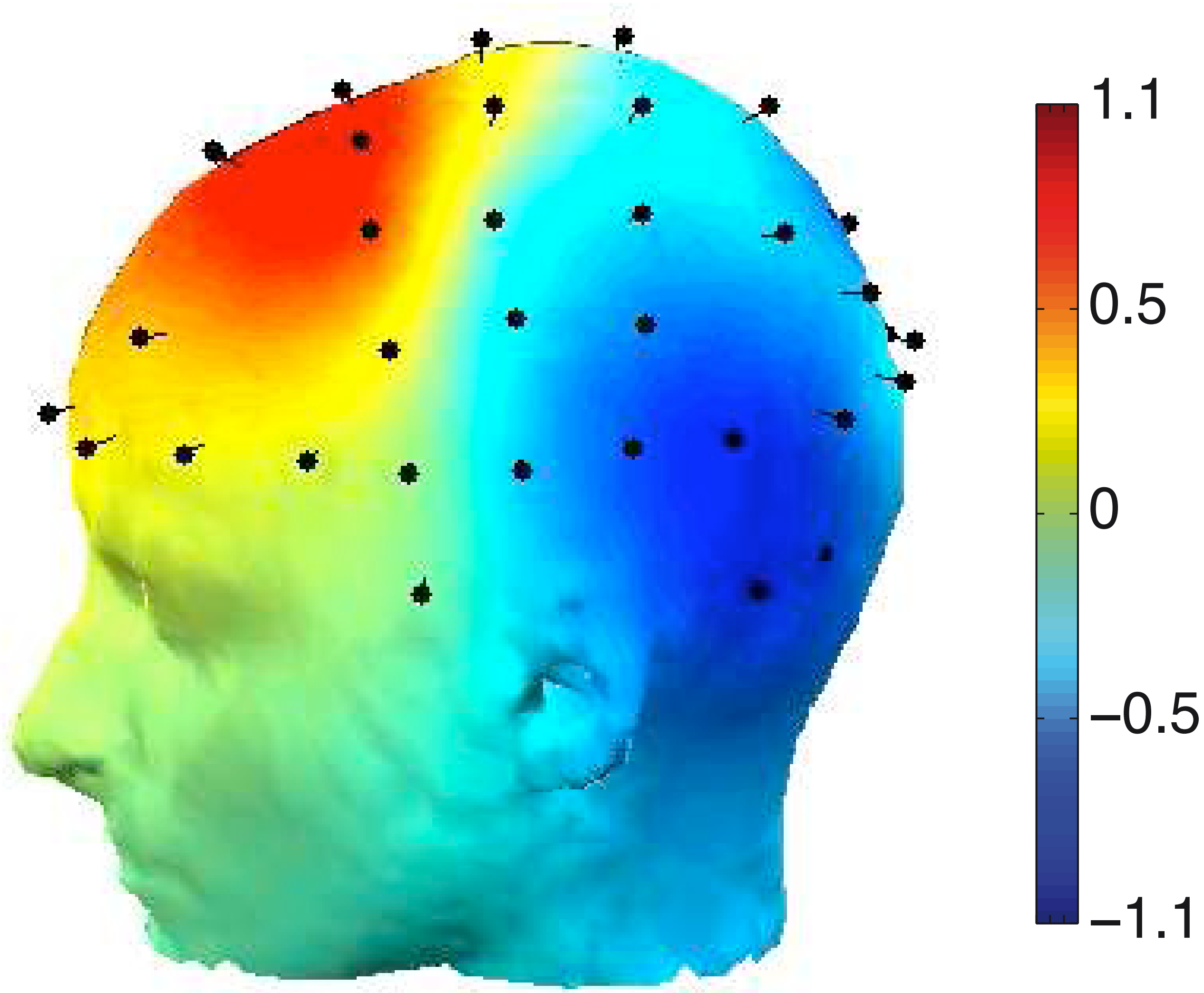
- ≈ 150 to 300 sensors

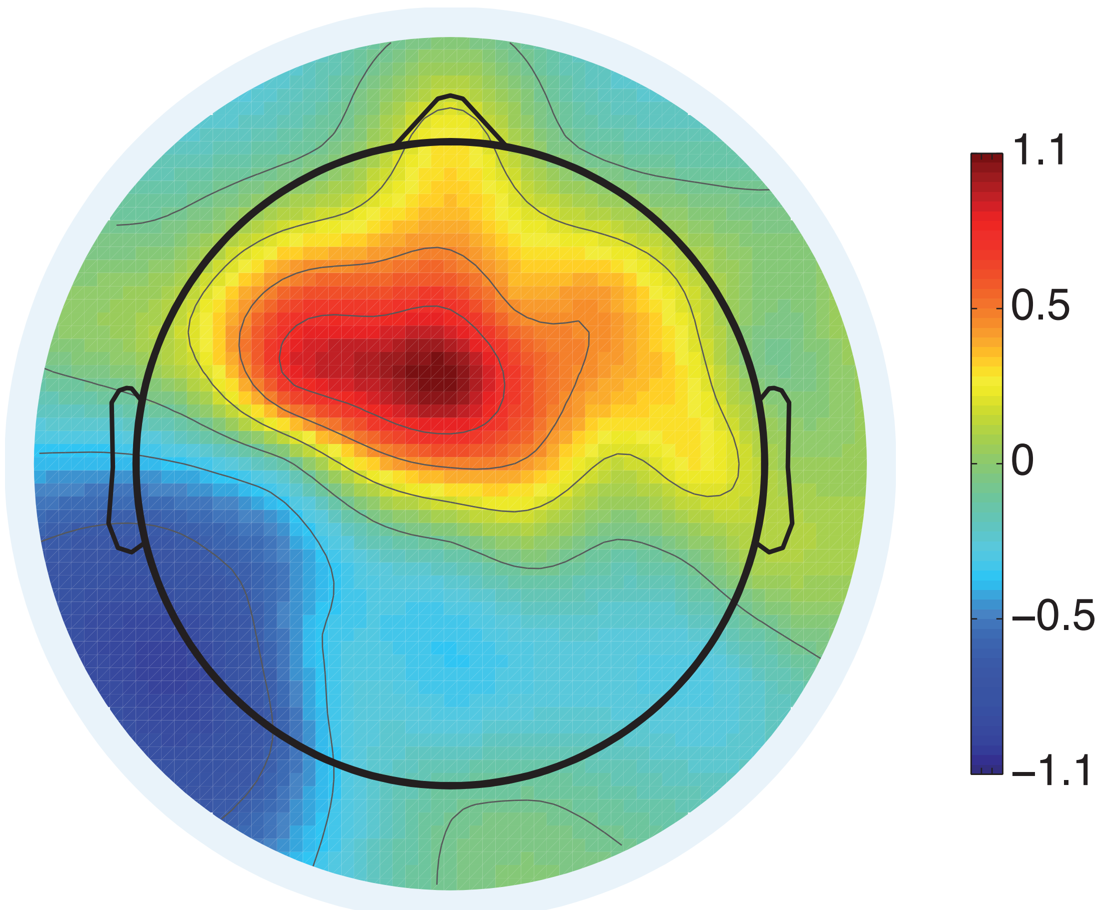
Sampling between 250 and1000 Hz

High temporal resolution but what about spatial resolution?

M/EEG Measurements

At **each time instant** EEG sensors measure a potential field



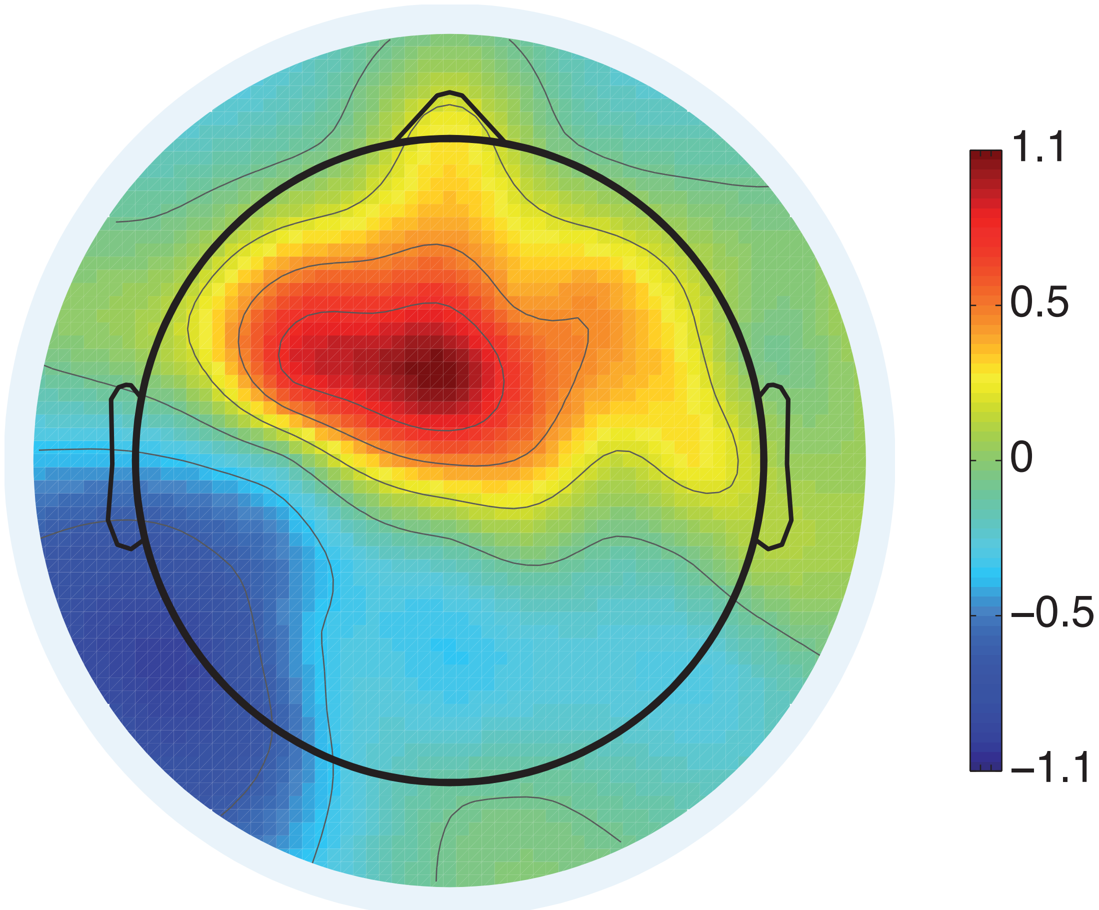


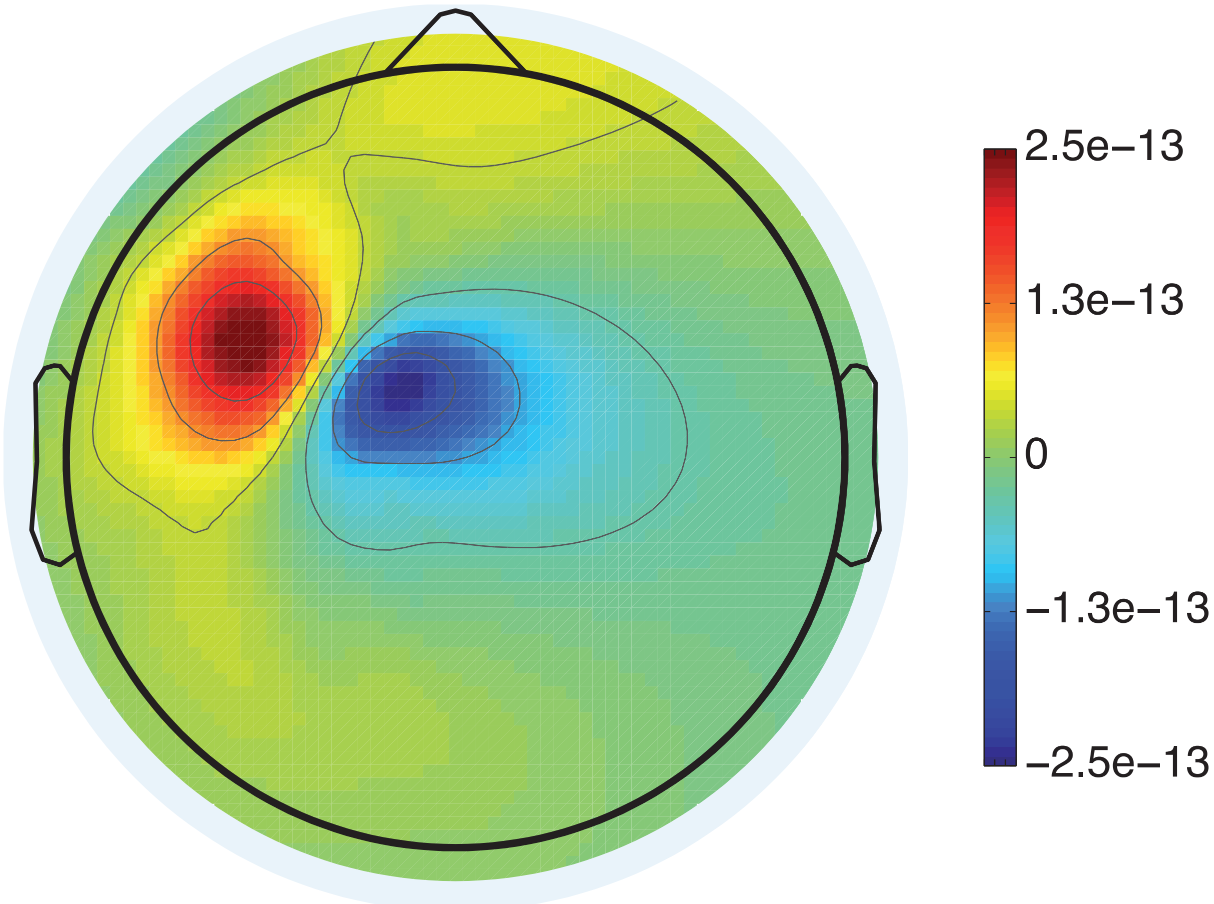
EEG topography

Remark: Such a smooth potential field confirms the presence of **current generators** within the head

M/EEG Measurements

EEG topography vs. MEG topography

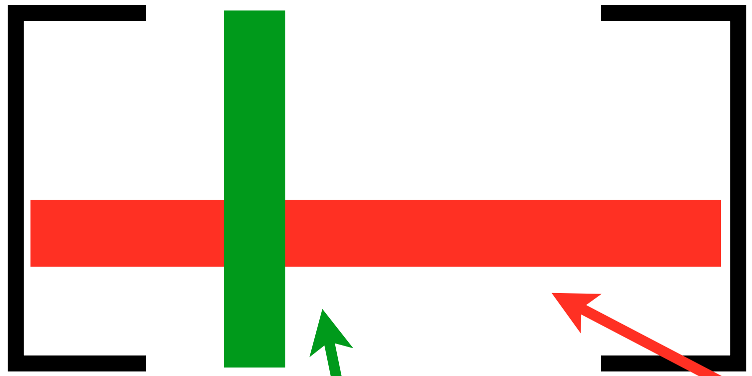




CTF system with 151 axial gradiometers

MEG topography exhibits also a dipolar field but MEG has a **better spatial resolution**

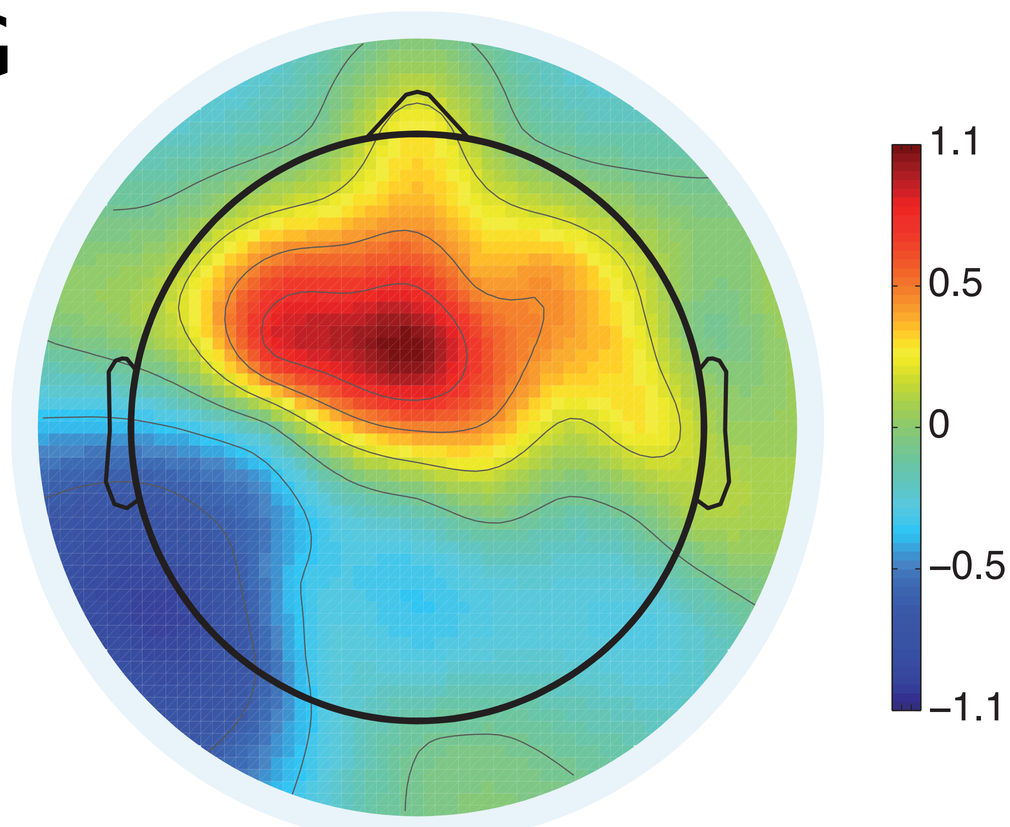
M/EEG Measurements: Notation

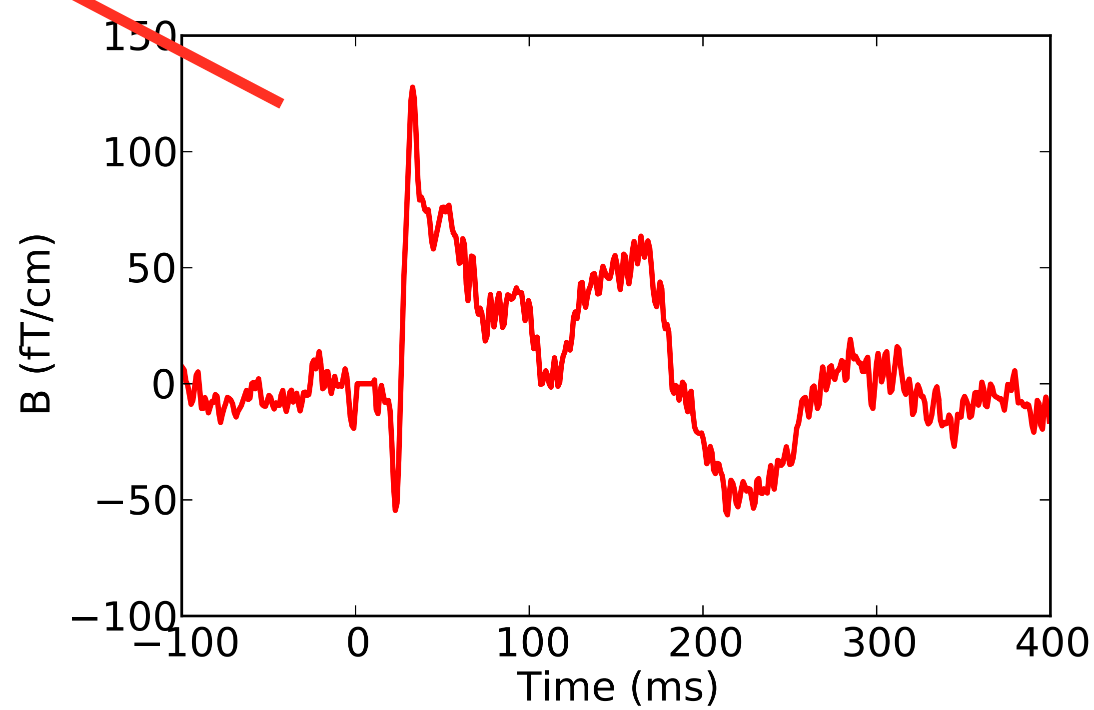
M =  $\in \mathbb{R}^{d_m \times d_t}$

MEG
and/or
EEG

d_m : Nb of sensors

d_t : Nb of time points



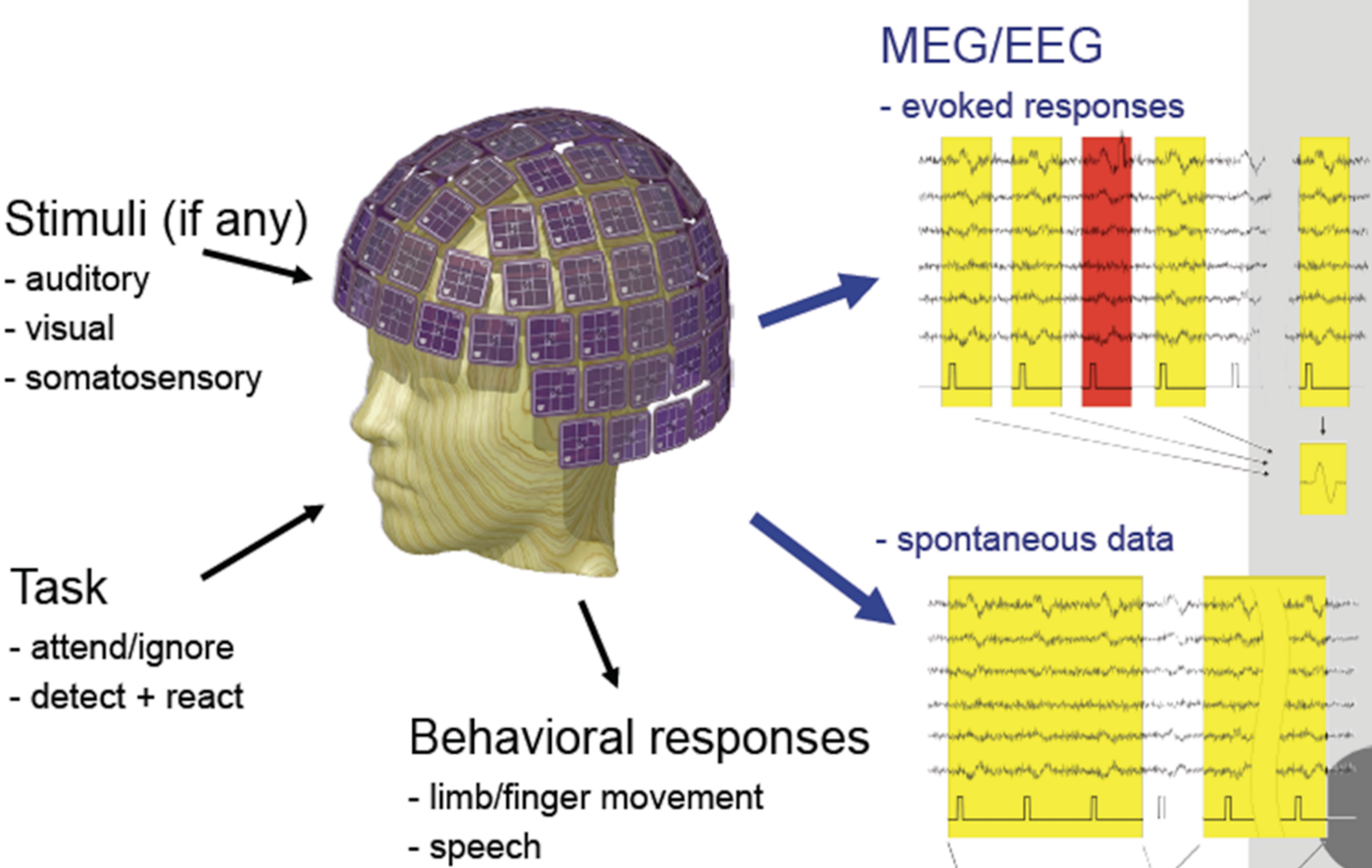


Time

1 column = 1 topography

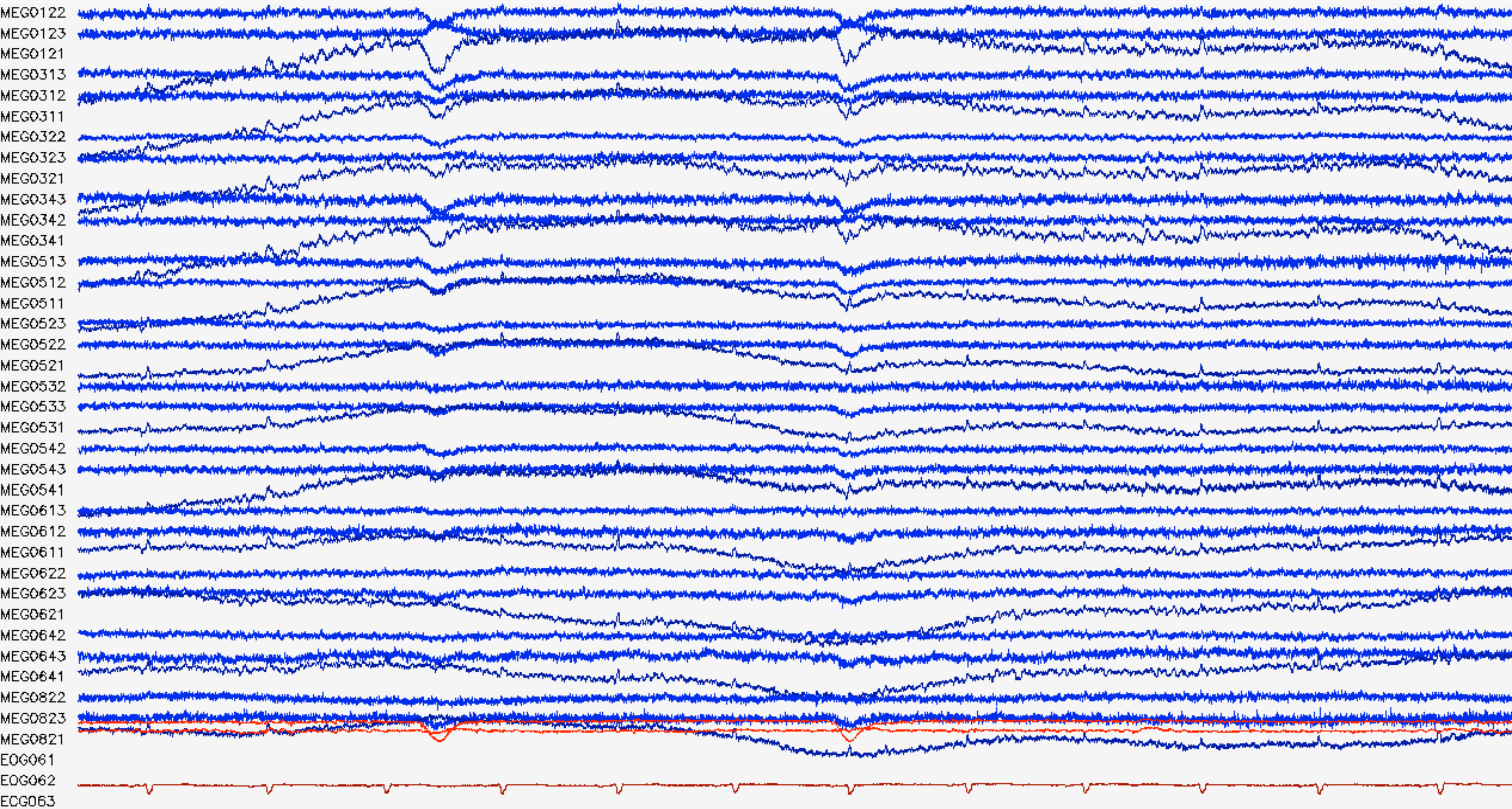
1 row = 1 time series on 1 sensor

Data acquisition examples



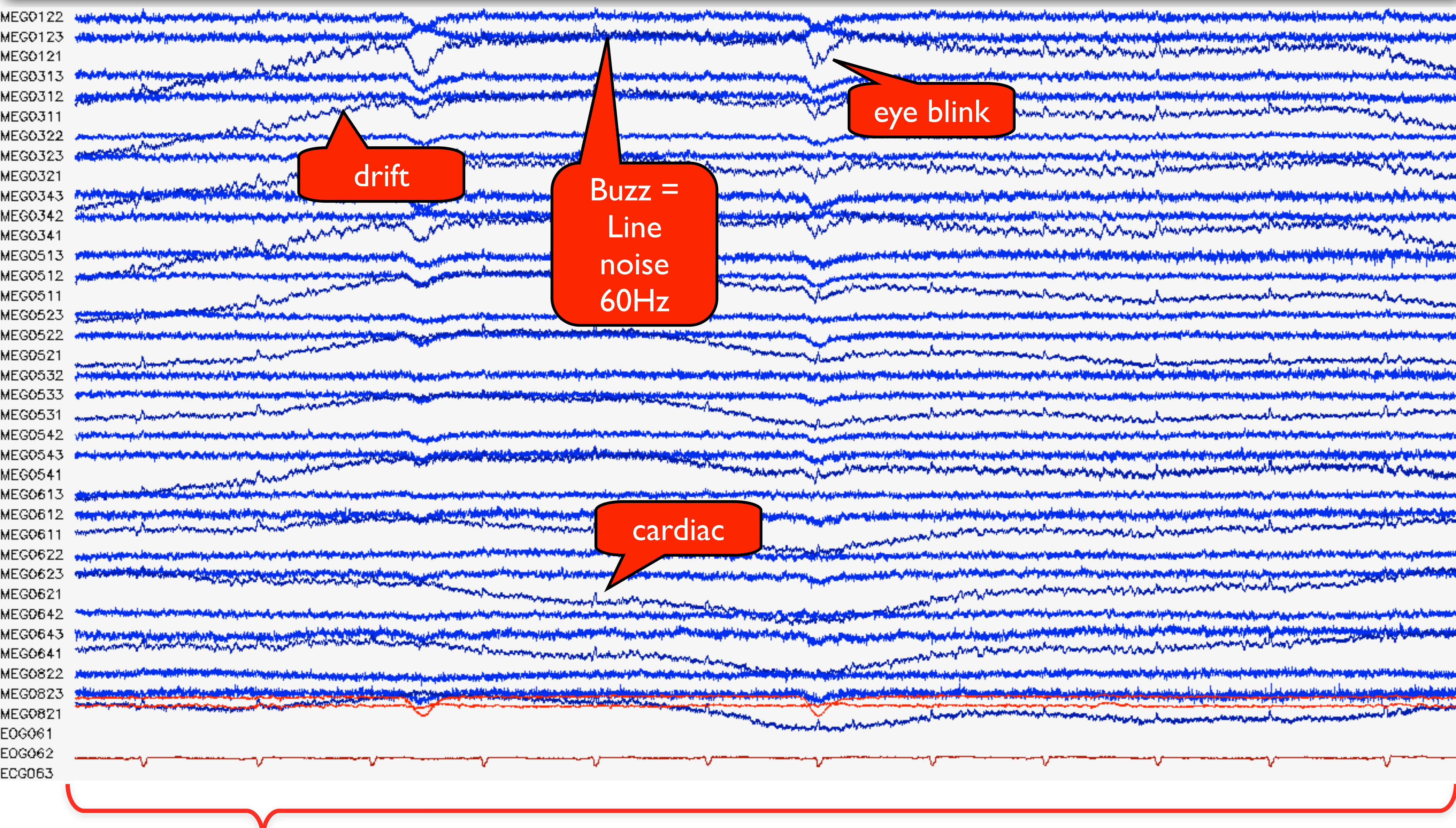
Preprocessing

Raw continuous data



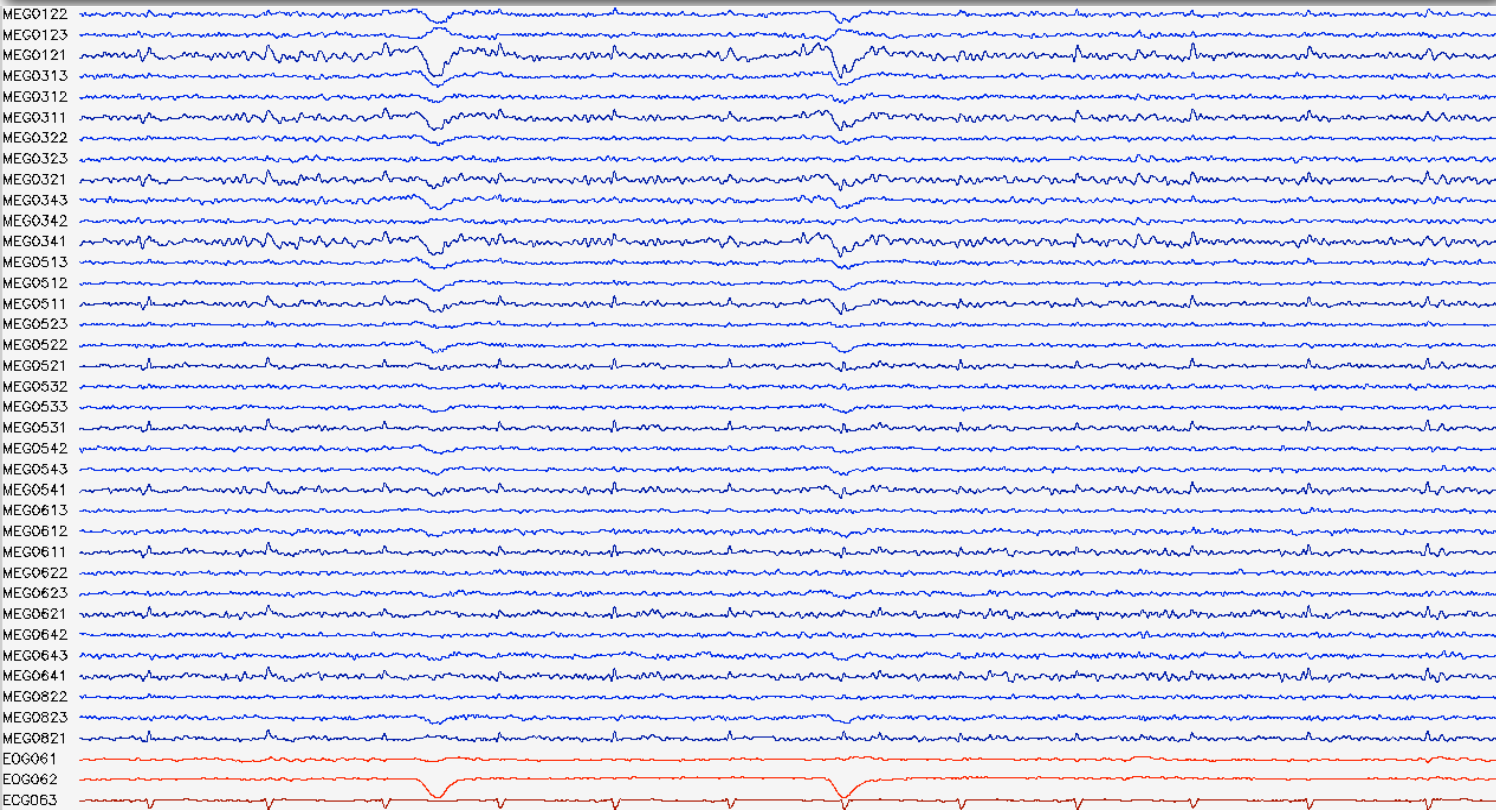
Time frame: 10 seconds

Artifacts



Time frame: 10 seconds

Filtered 1-40Hz



Time frame: 10 seconds

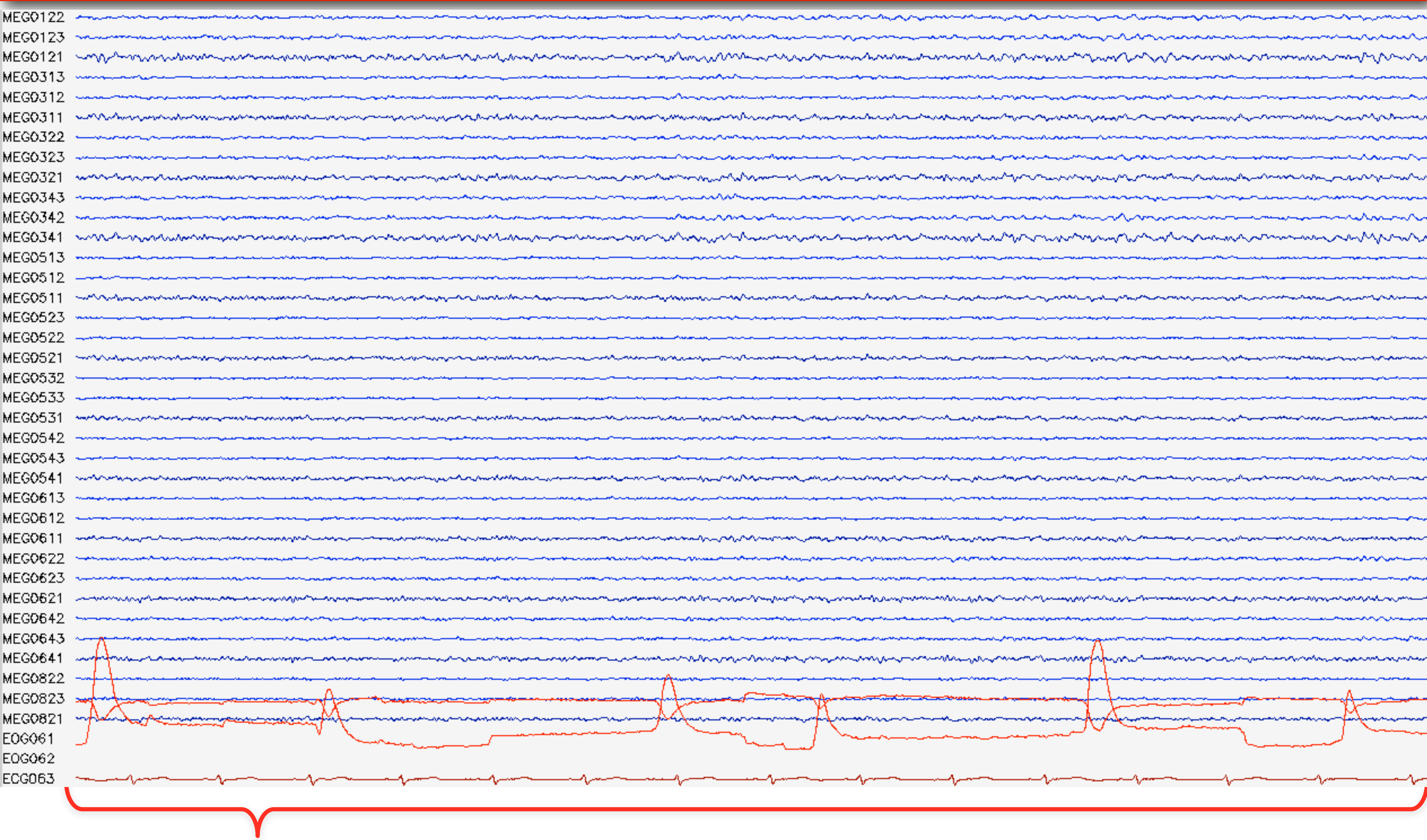
Artifact correction

- Signal Space Projections (SSP)
- Independent component analysis (ICA)
- ...

https://mne.tools/dev/auto_tutorials/preprocessing/plot_45_projectors_background.html

https://mne.tools/dev/auto_tutorials/discussions/plot_background_ica.html

To get clean data...



Time frame: 10 seconds

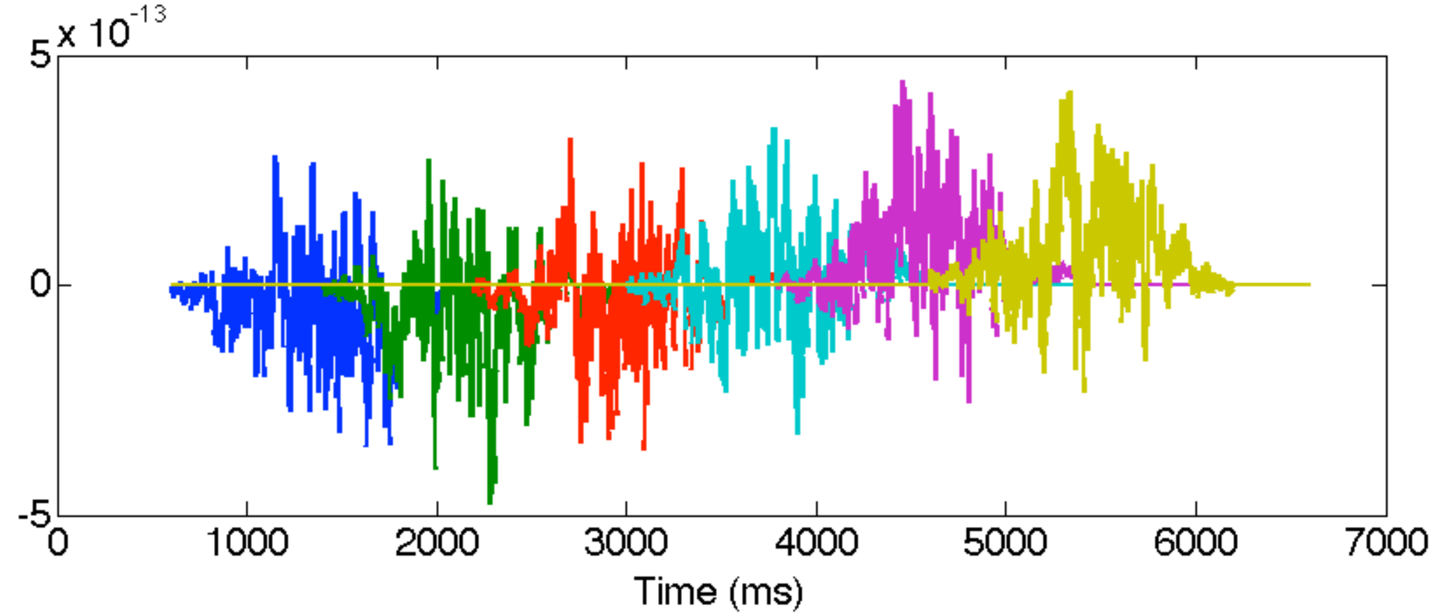
Power Spectra estimation

How to estimate the PSD of 5 mins of MEG signal at 1000 Hz?

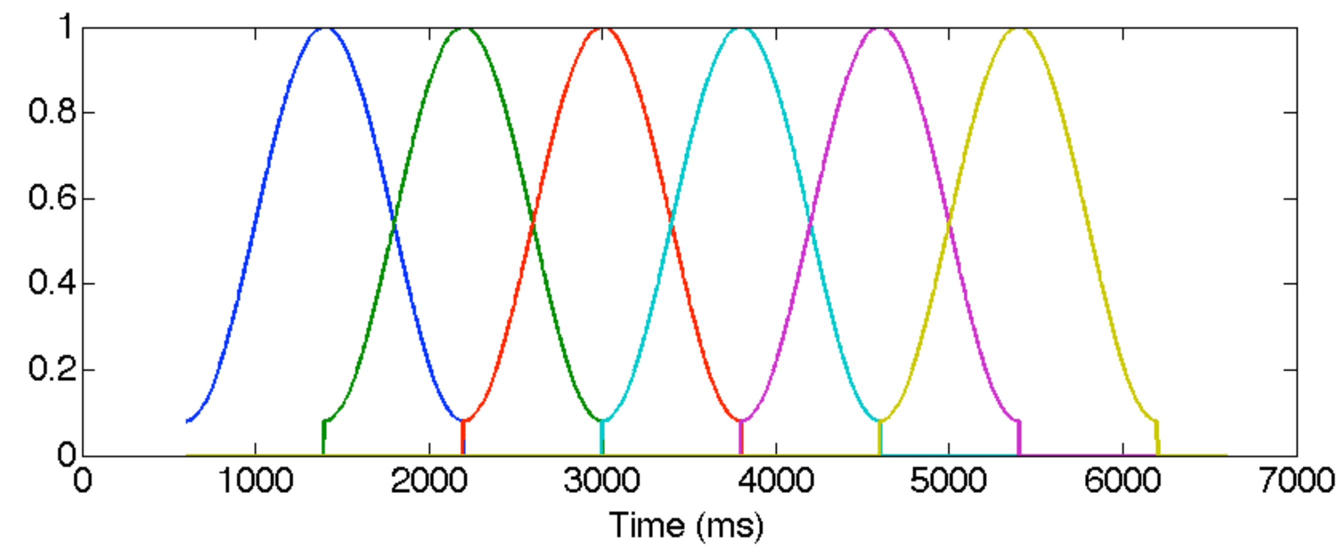
Power Spectrum Density

On raw data (typically > 5 minutes)

Approach: Sliding-Overlapping windows



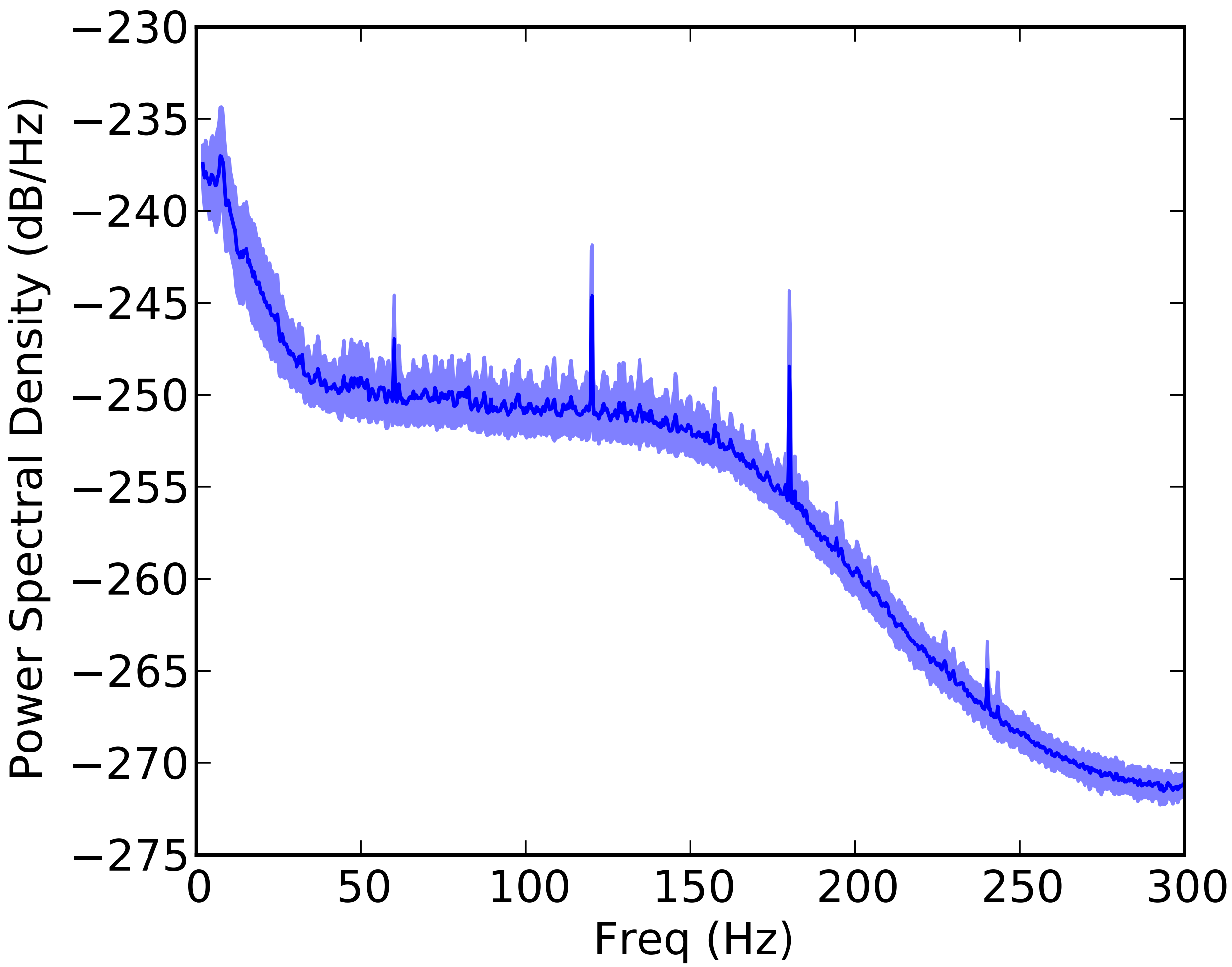
Pro: Presence of artifacts at some time instants so the windows can be discarded

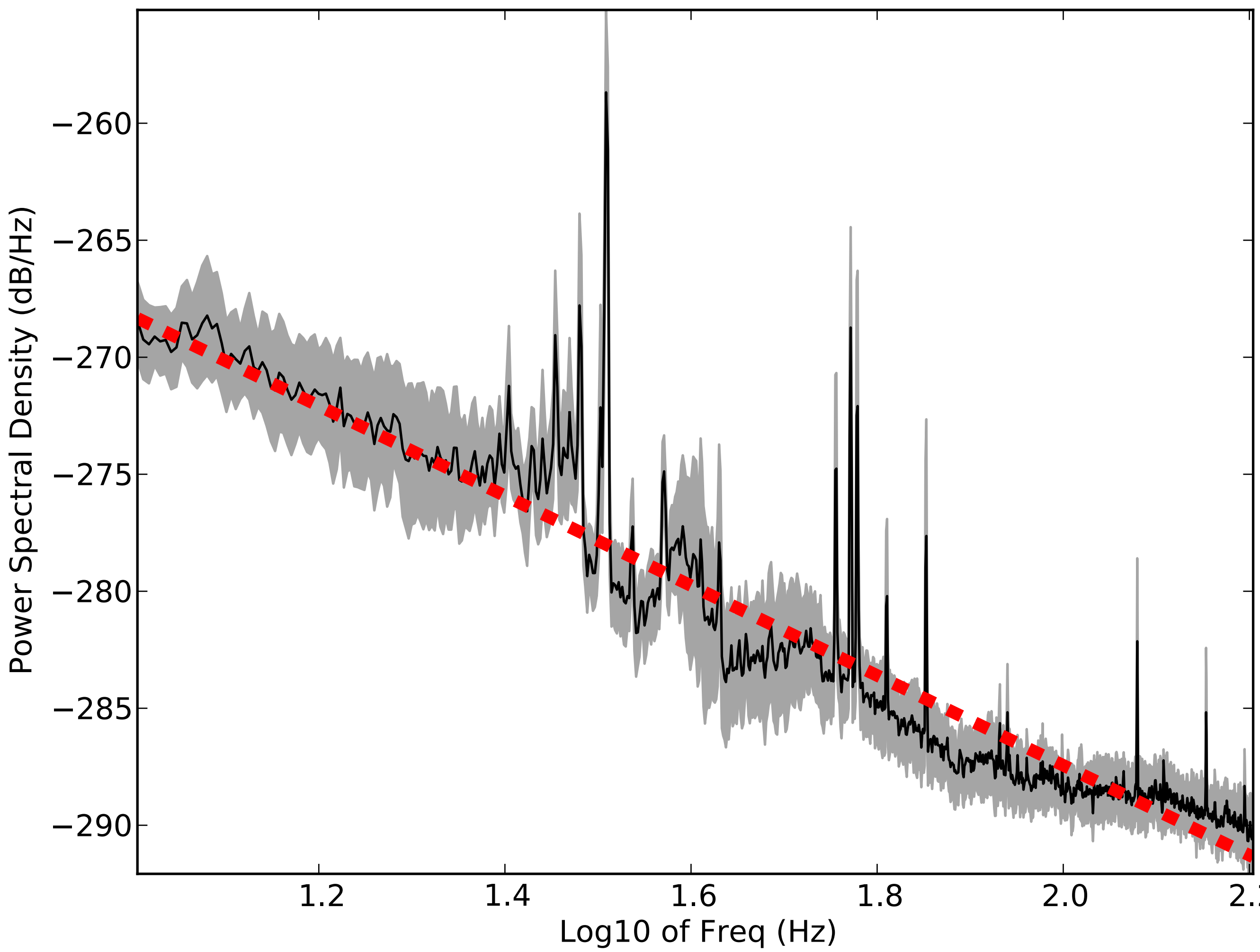


Welch, P.D, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms," IEEE Trans. Audio Electroacoustics, Vol. AU-15 (June 1967), pp.70-73.

Power Spectrum Density

On raw data (typically > 5 minutes)

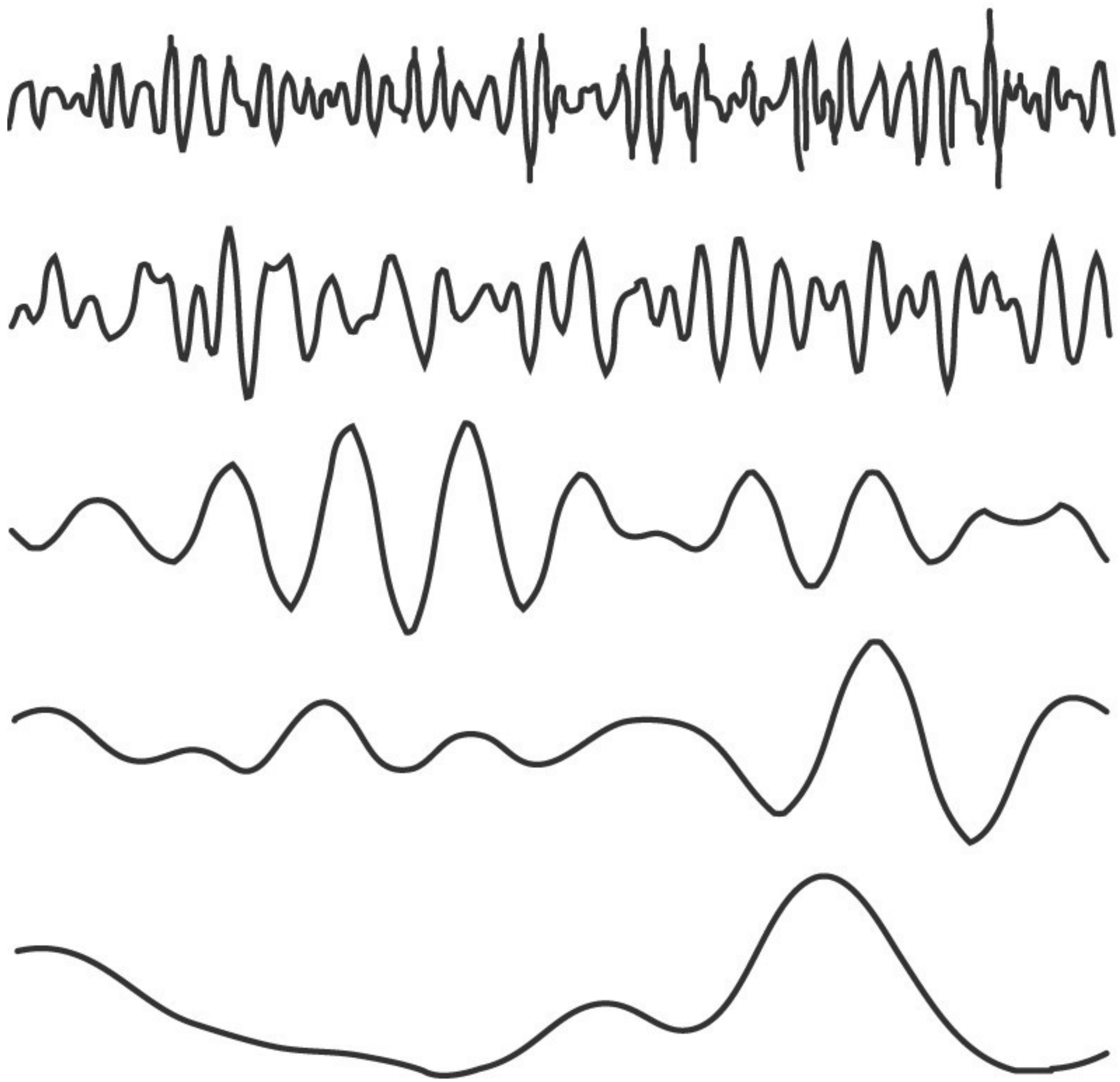




Presence of peaks that are artifacts

1/f PSD

“Textbook” brain rythms



Gamma
(> 25 Hz)

Beta
(12-25 Hz)

Alpha
(8-12 Hz)

Theta
(4-8 Hz)

Delta
(1-4 Hz)

Part 2:
Machine learning on EEG/MEG (a.k.a. decoding)

Encoding and Decoding Neuronal Dynamics: Methodological Framework to Uncover the Algorithms of Cognition

Jean-Rémi King, Laura Gwilliams, Chris Holdgraf, Jona Sassenhagen, Alexandre Barachant, Denis Engemann, Eric Larson, Alexandre Gramfort

https://hal.archives-ouvertes.fr/hal-01848442/document

IOPscience Journals Books Publishing Support Login Search IOPscier

Journal of Neural Engineering

TOPICAL REVIEW • OPEN ACCESS

Deep learning-based electroencephalography analysis: a systematic review

Yannick Roy[1,5], Hubert Banville[2,3,5], Isabela Albuquerque[4], Alexandre Gramfort[2], Tiago H Falk[4] and Jocelyn Faubert[1]

Published 14 August 2019 •

Journal of Neural Engineering, Volume 16, Number 5

Article PDF

Figures References

https://arxiv.org/abs/1901.05498

Machine Learning on EEG/MEG
"big picture"

Machine Learning

Machine Learning on EEG/MEG "big picture"

Machine Learning

Unsupervised learning

denoising

data exploration

Methods:

Principal Component Analysis (PCA)
Independent Component Analysis (ICA)
Convolutional Sparse Coding (CSC)
Spatio-Spectral Decomposition (SSD)
etc.

Machine Learning on EEG/MEG "big picture"

- Machine Learning
  - Unsupervised learning
    - denoising
    - data exploration
  - Supervised learning
    - Cognitive Neuroscience
    - Brain Computer Interface (BCI)
    - Diagnosis Personalised Medicine

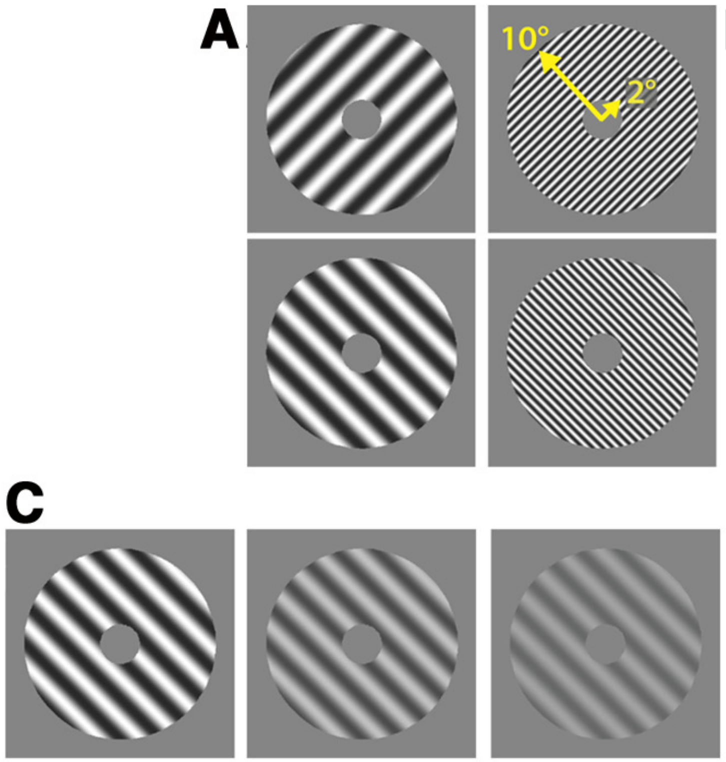
Methods:
Principal Component Analysis (PCA)
Independent Component Analysis (ICA)
Convolutional Sparse Coding (CSC)
Spatio-Spectral Decomposition (SSD)
etc.

Methods:
Linear Discriminant Analysis (LDA)
Logistic Regression (LR),
Support Vector Machine (SVM)
Common Spatial Filters (CSP)
Convolutional Neural Network (CNN)
Riemann on covariance matrices,
Random Forest, etc.

Outline

- Examples of EEG/MEG papers you can replicate with MNE scikit learn
- Evoked vs. Induced activity
- ML on EEG/MEG: Encoding vs. Decoding
- Why doing machine learning on EEG/MEG?
- Algorithms
    - Background on linear models
    - Algorithms for evoked activity (Time decoding, xDAWN)
    - Algorithms for induced activity (CSP, Riemann)
- How to run cross-validation right !
- Choosing your metric

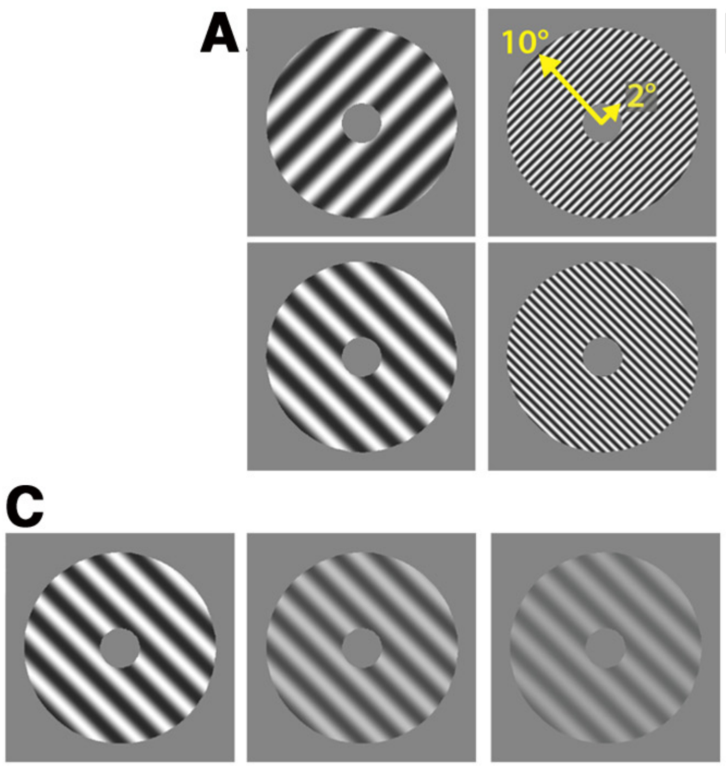
Examples of ML on EEG and MEG

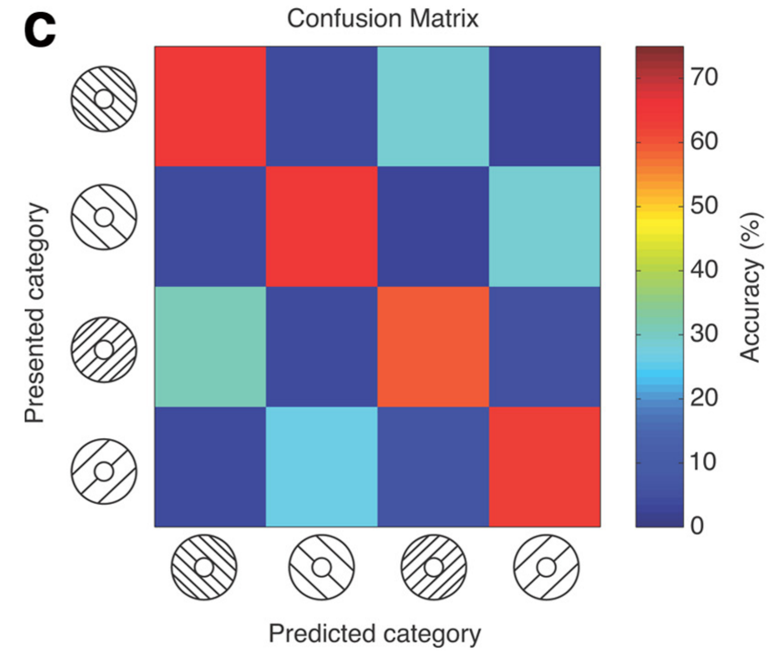


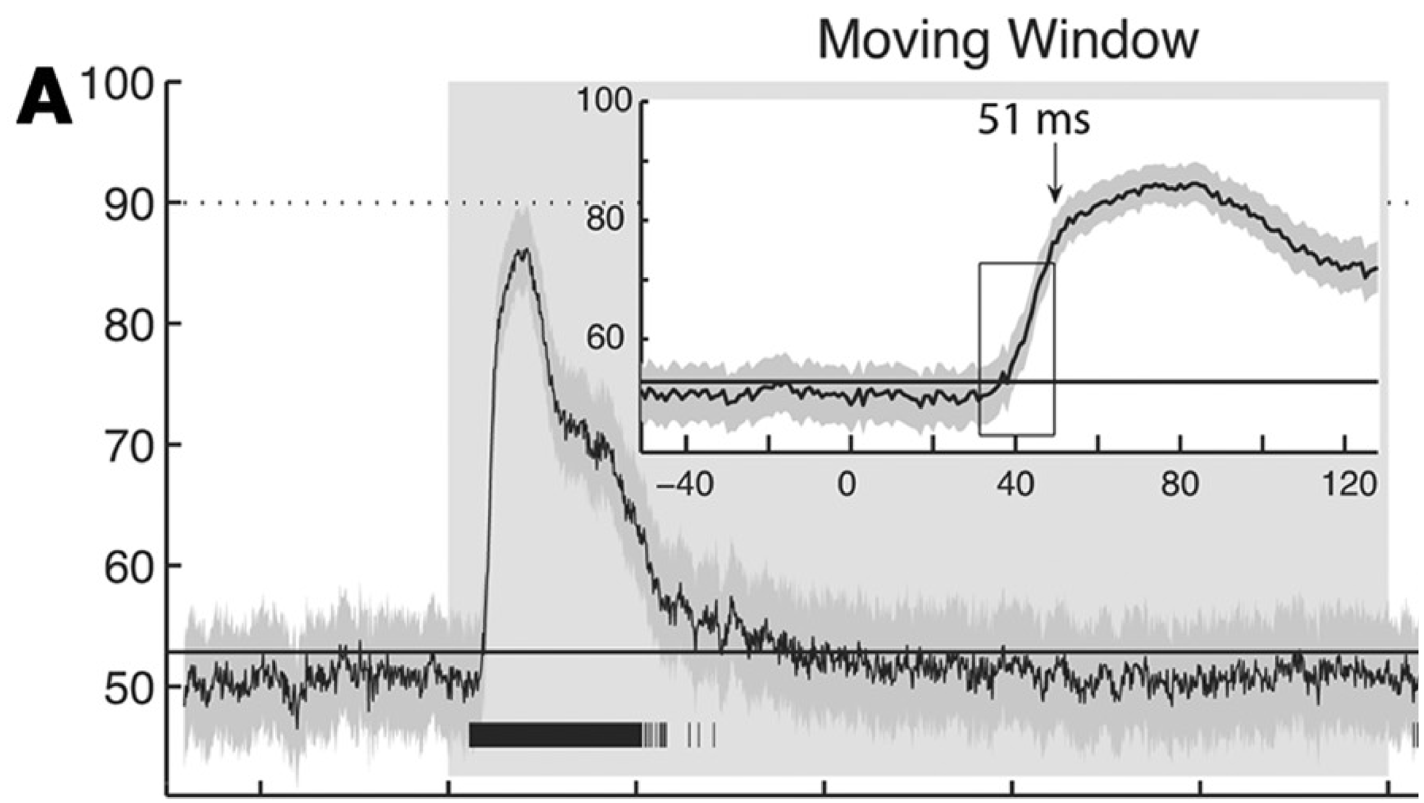
[Ramkumar, Jas, Pannasch, Hari, Parkkonen, J. Neuroscience (2013)]

https://mne.tools/stable/generated/mne.decoding.SlidingEstimator.html

MNE
MEG + EEG ANALYSIS & VISUALIZATION





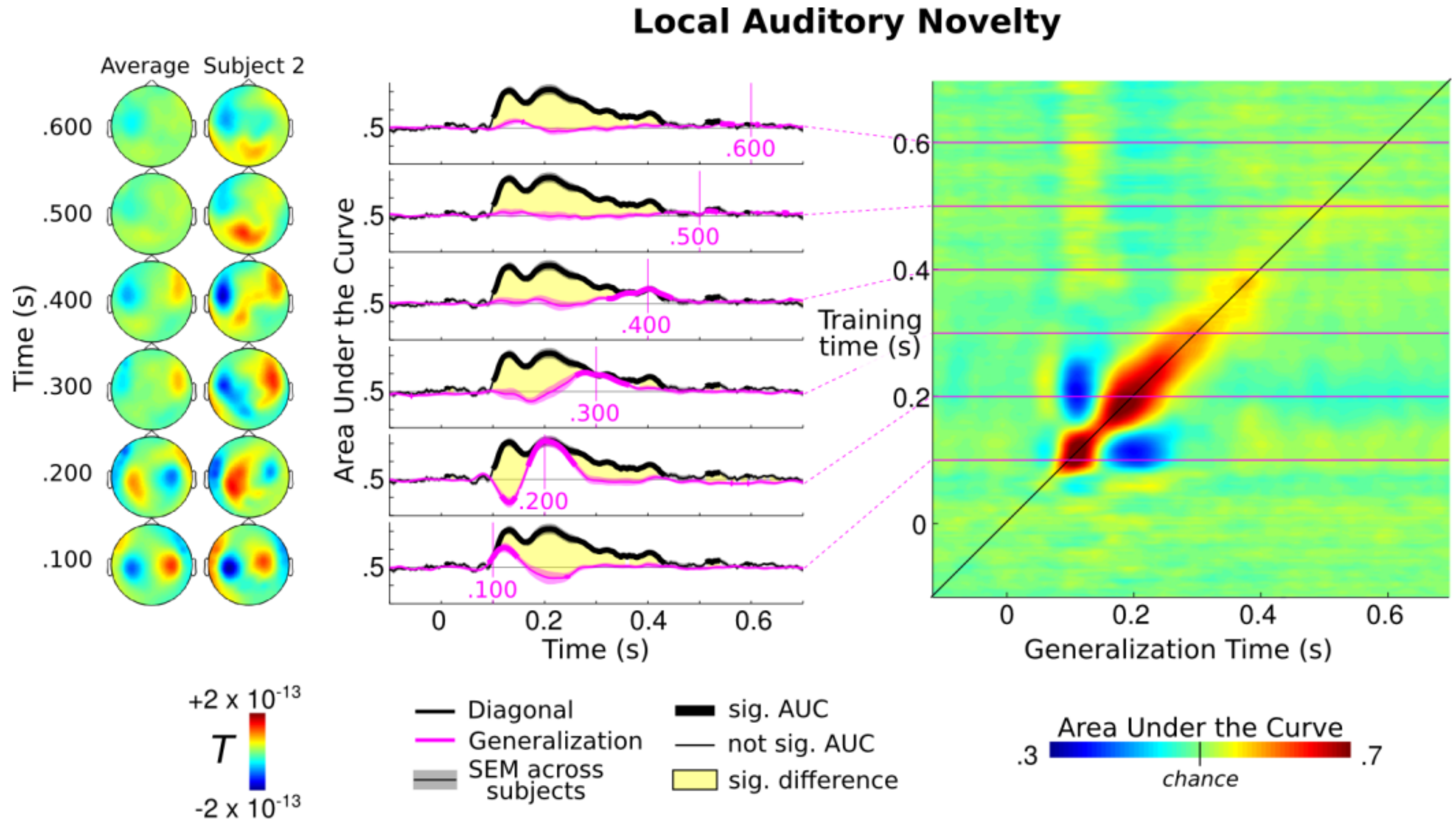


“Time-resolved classifiers using a 20 ms moving window exceeded chance level at 51 ms (the later edge of the window) for spatial frequency, 65 ms for orientation, and 98 ms for rotation direction.”

[Ramkumar, Jas, Pannasch, Hari, Parkkonen, J. Neuroscience (2013)]

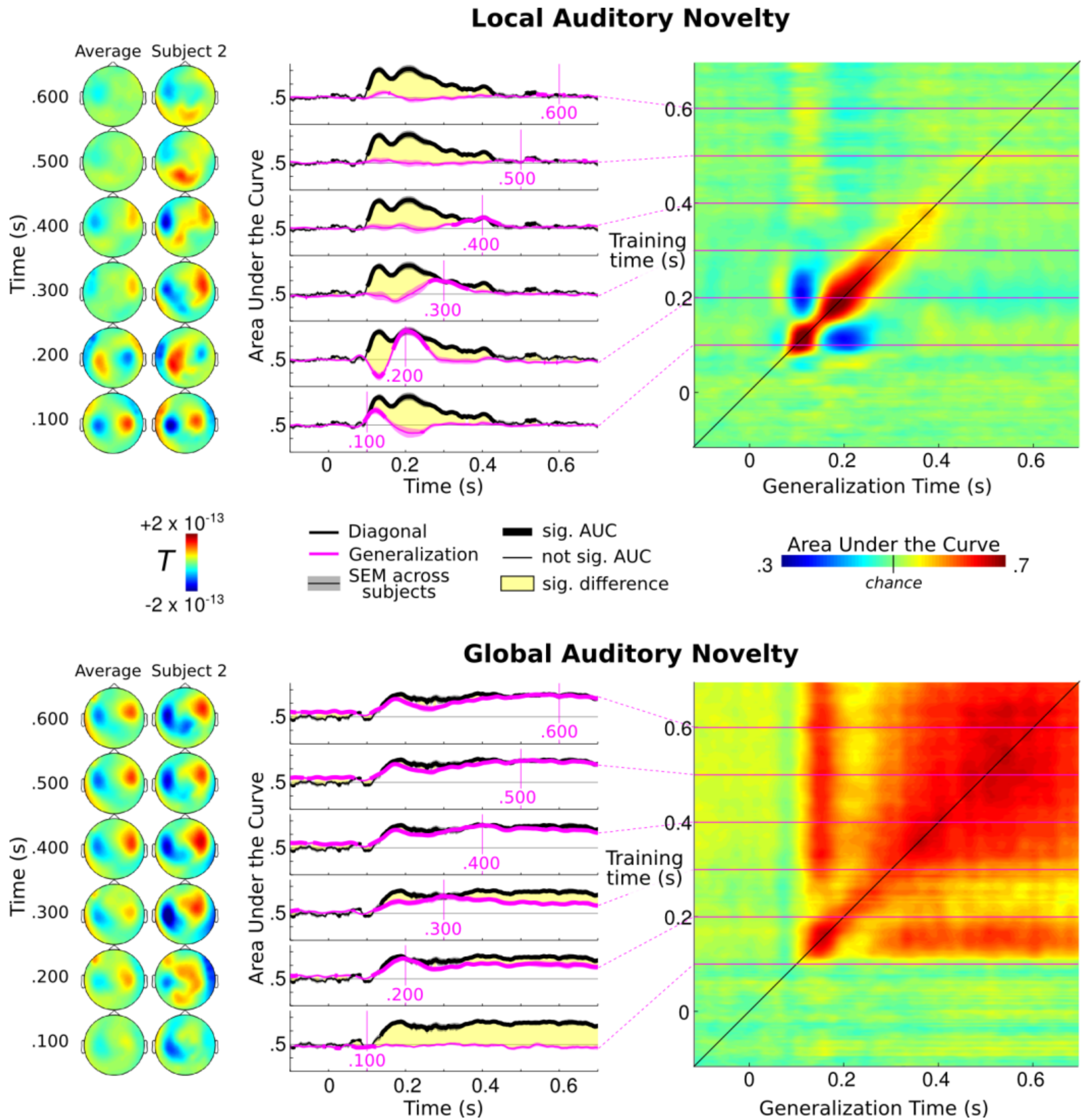
https://mne.tools/stable/generated/mne.decoding.SlidingEstimator.html

MEG + EEG ANALYSIS & VISUALIZATION



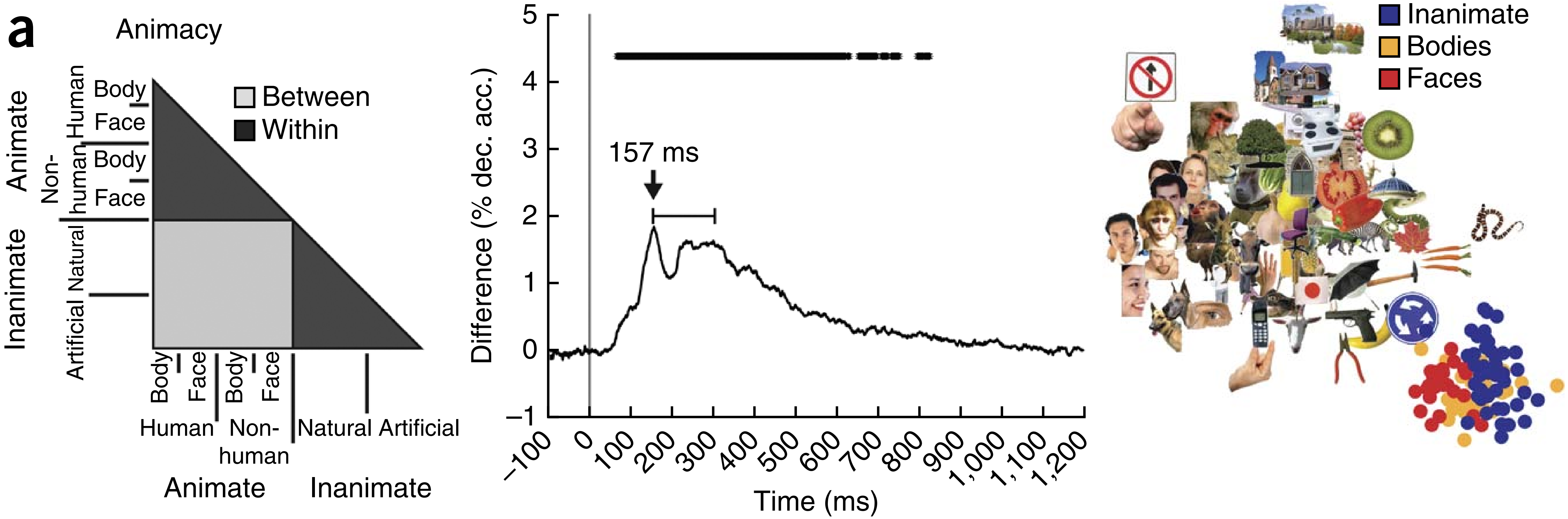
[King, Gramfort, Schurger, Naccache, Dehaene (2014) PLoS One]

https://mne.tools/stable/generated/mne.decoding.GeneralizingEstimator.html



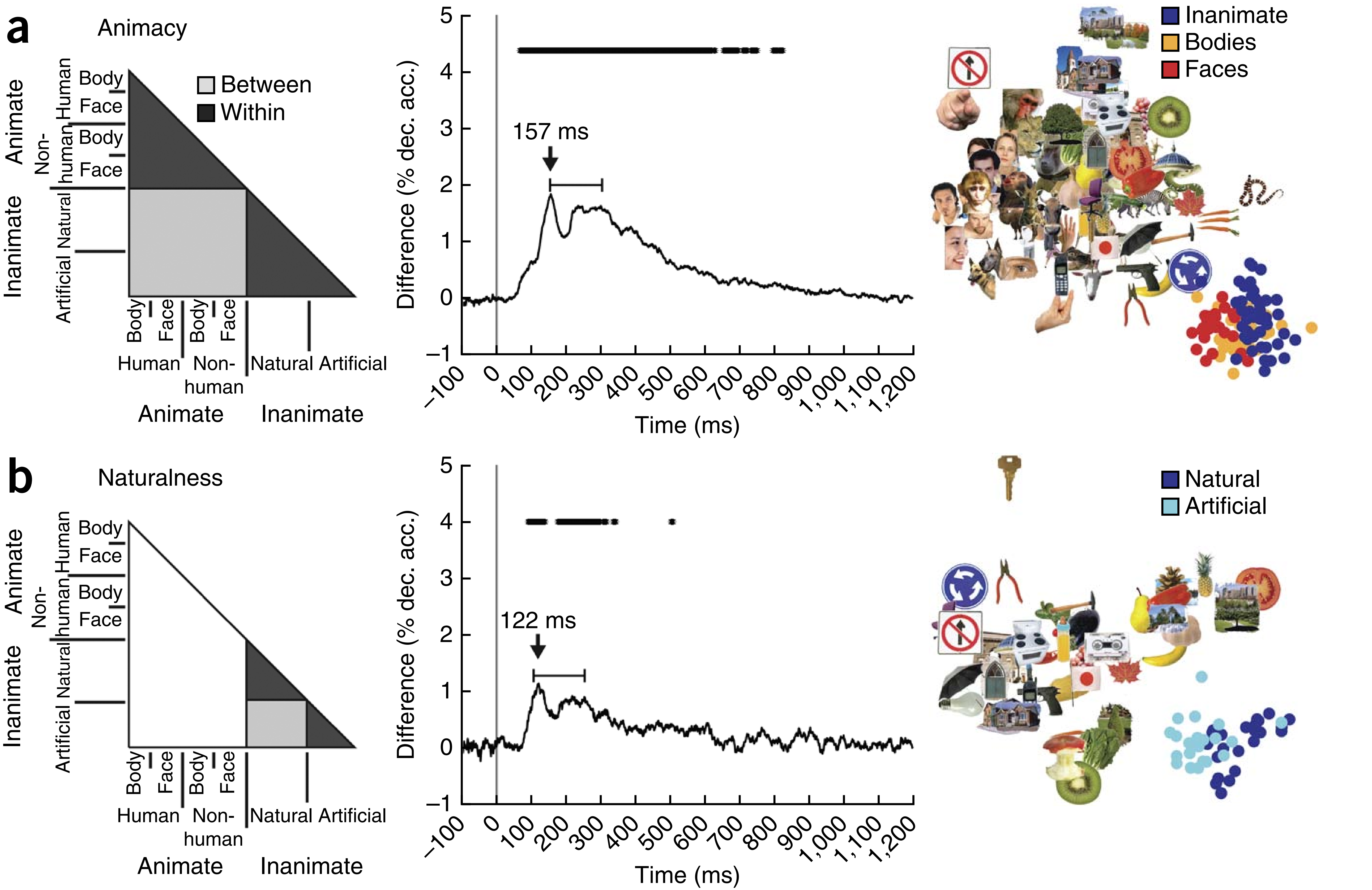
[King, Gramfort, Schurger, Naccache, Dehaene (2014) PLoS One]

https://mne.tools/stable/generated/mne.decoding.GeneralizingEstimator.html



[Resolving human object recognition in space and time, Cichy, Pantazis & Oliva, Nat Neuro 2014]

https://mne.tools/stable/auto_examples/decoding/decoding_rsa.html



[Resolving human object recognition in space and time, Cichy, Pantazis & Oliva, Nat Neuro 2014]

https://mne.tools/stable/auto_examples/decoding/decoding_rsa.html

American Physiological Society | Journal of Neurophysiology

PUBLISHED ARTICLE
ARCHIVES
SUBSCRIPTIONS
SUBMISSIONS
CONTACT US

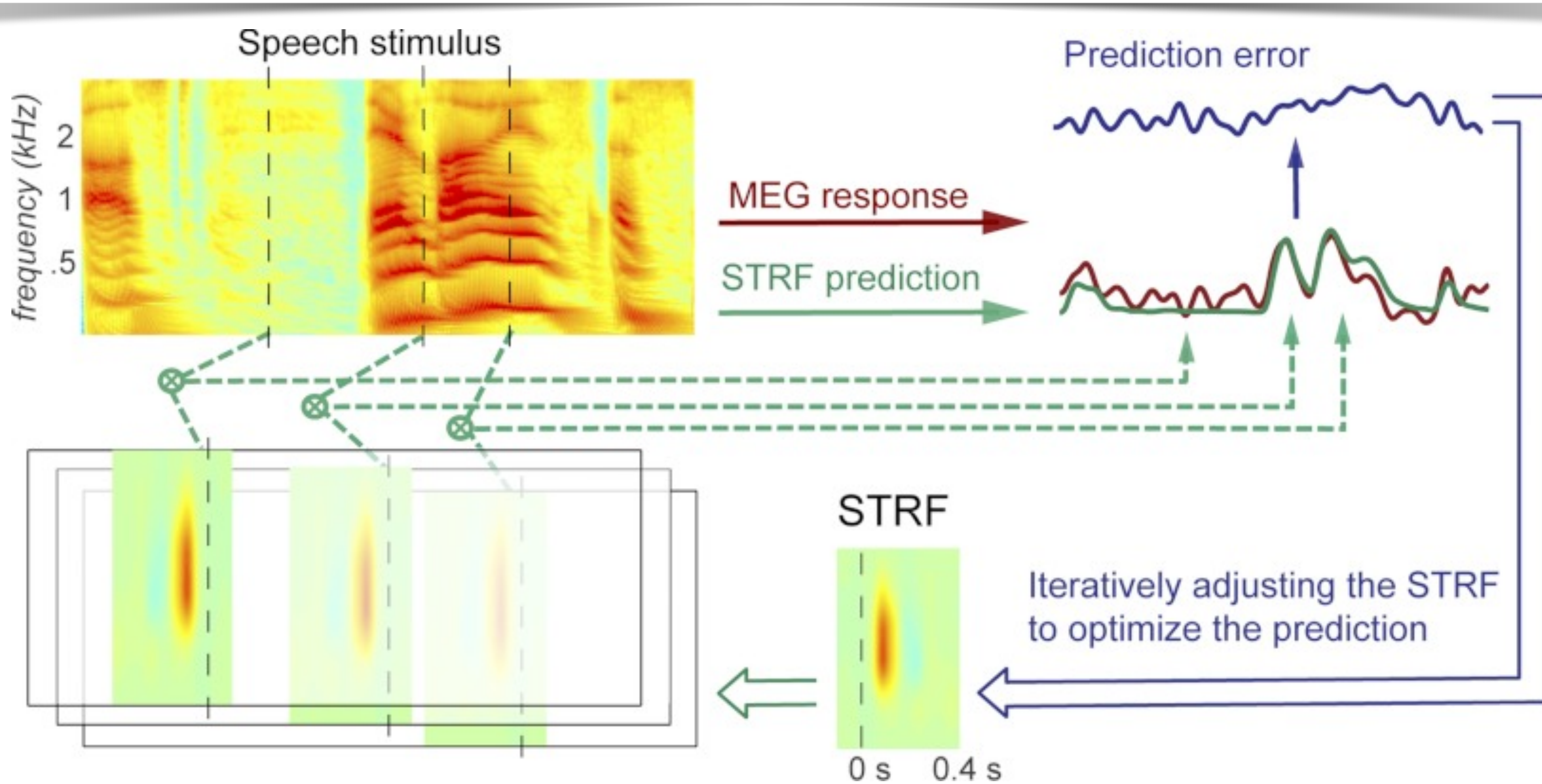
J Neurophysiol. 2012 Jan 1; 107(1): 78–89.
Published online 2011 Oct 5. doi: 10.1152/jn.00297.2011

PMCID: PMC3570829
PMID: 21975452

Neural coding of continuous speech in auditory cortex during monaural and dichotic listening

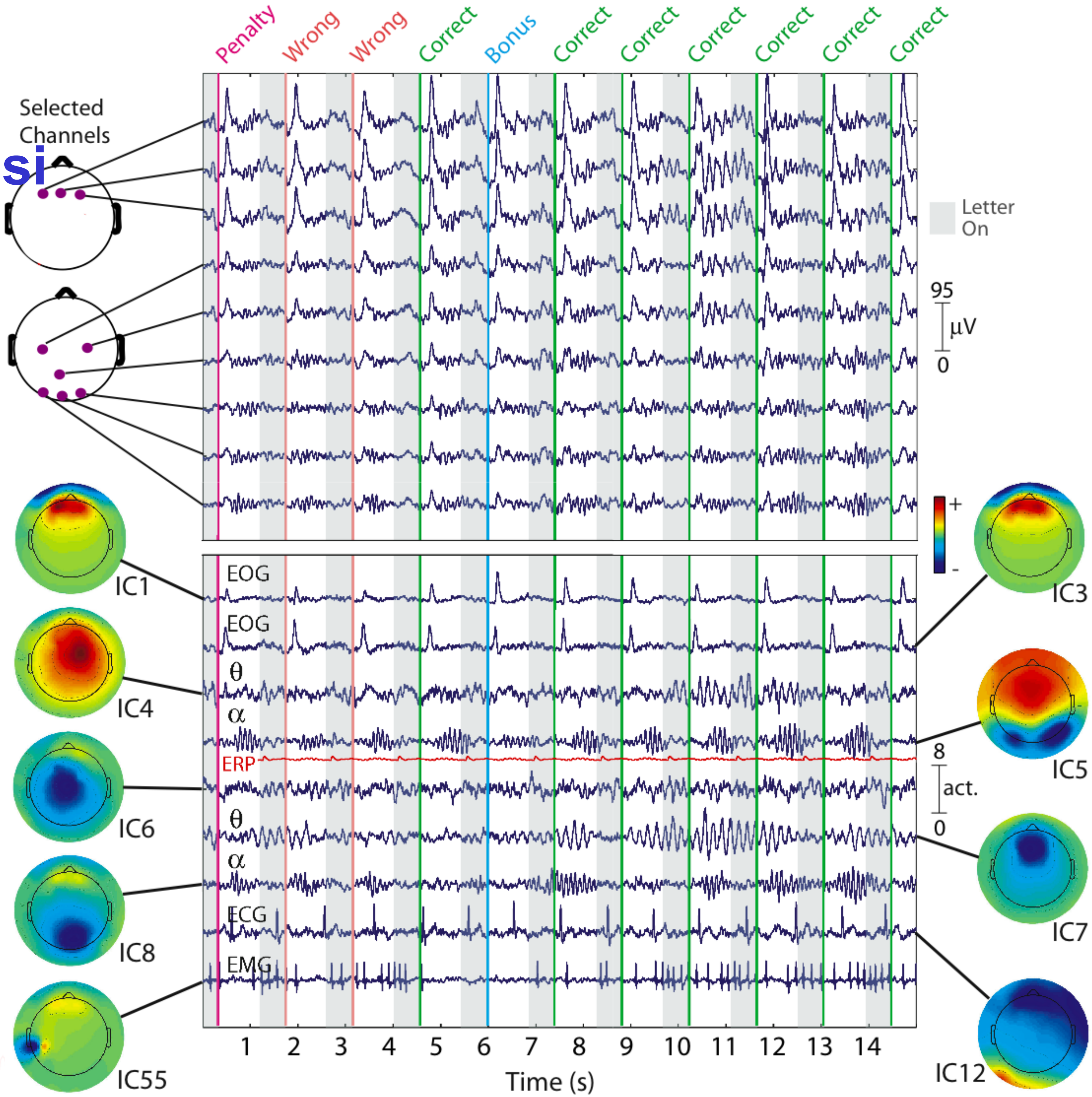
Nai Ding[1] and Jonathan Z. Simon[1,2]

[1]Department of Electrical and Computer Engineering and
[2]Department of Biology, University of Maryland, College Park, Maryland
Corresponding author.
Address for reprint requests and other correspondence: J. Z. Simon, Univ. of Maryland, College Park, MD 20742 (e-mail: jzsimon@umd.edu).



https://mne.tools/stable/auto_tutorials/machine-learning/plot_receptive_field.html

Sample EEG Decomposition

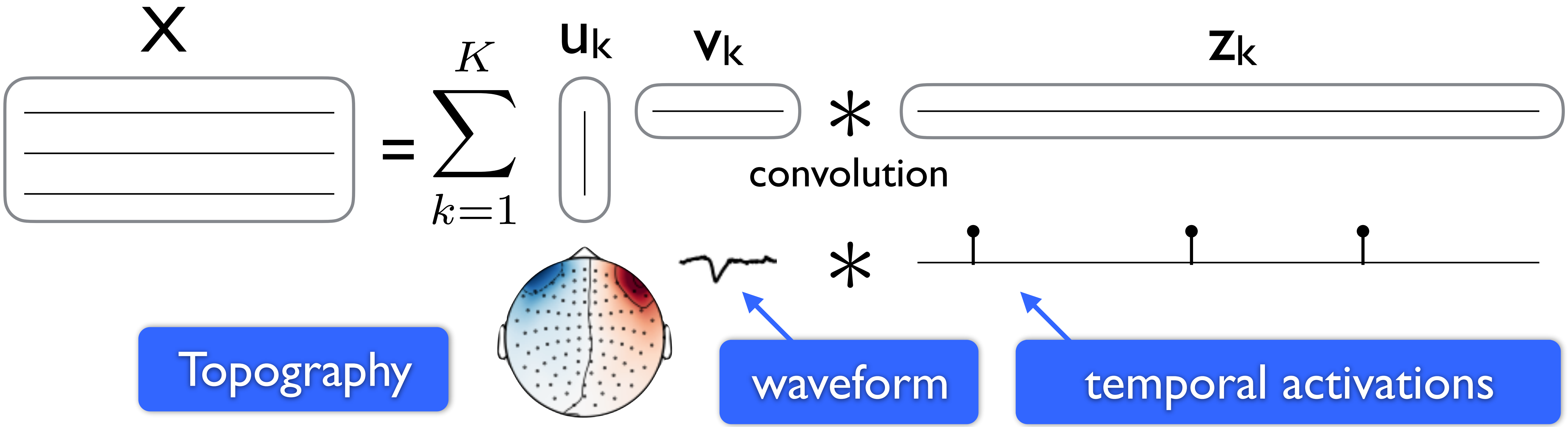


[Makeig et al. 1996, Jung et al. 1998, Jung et al. 1999 etc.]

https://mne.tools/stable/generated/mne.preprocessing.ICA.html

Credit: J. Onton

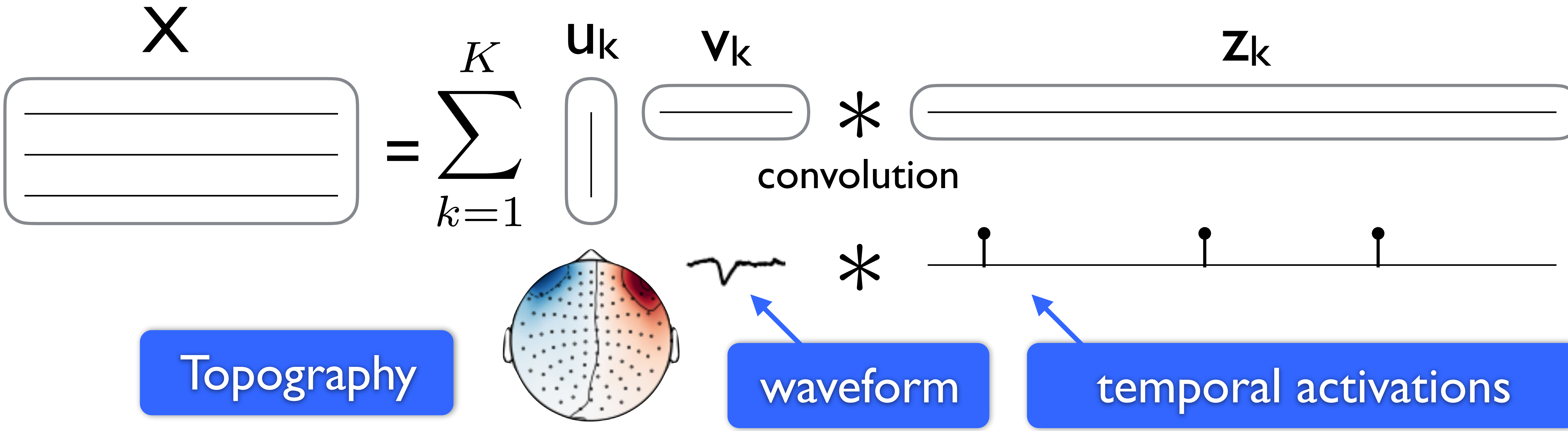
Convolutional Sparse Coding

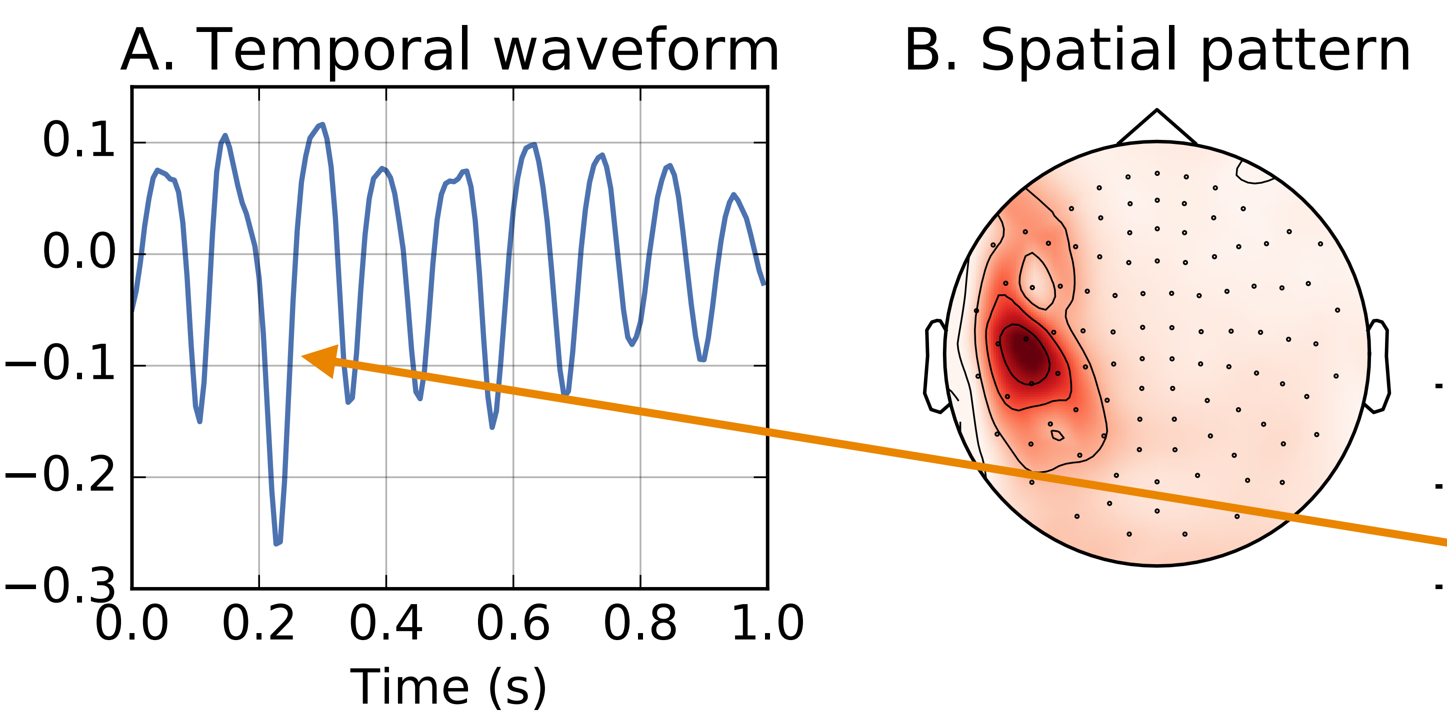


[Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals, (2018), T. Dupré la Tour, T. Moreau, M. Jas, A. Gramfort, Proc. NeurIPS Conf.]

https://alphacsc.github.io

Convolutional Sparse Coding





- MEG vectorview
- Median nerve stim.

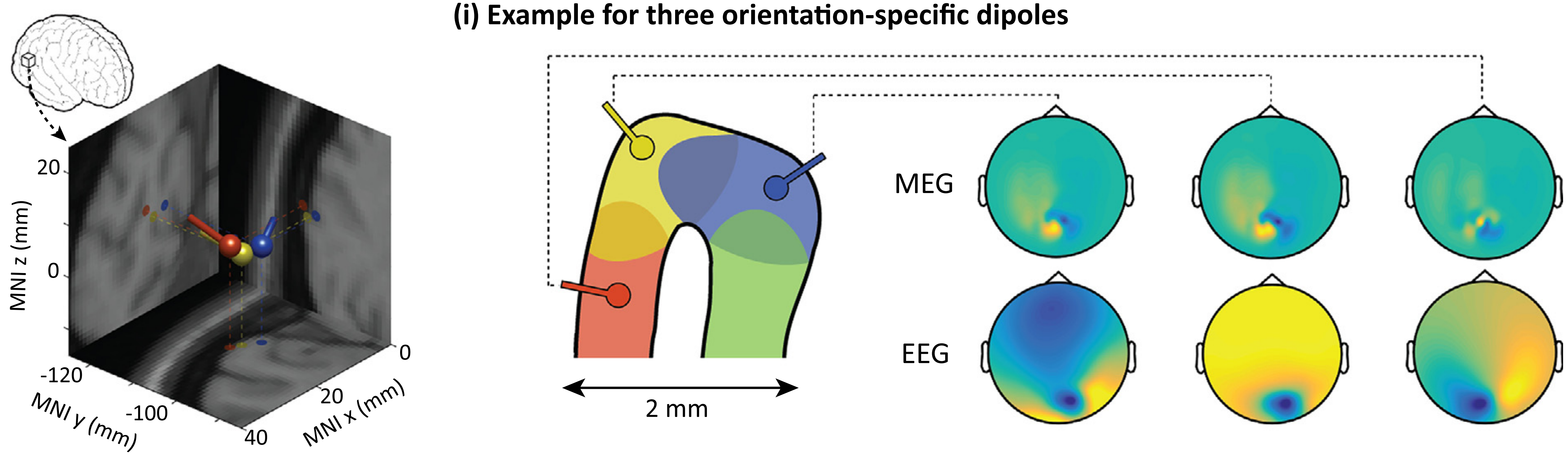
CSC reveals mu-shaped waveforms

[Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals, (2018), T. Dupré la Tour, T. Moreau, M. Jas, A. Gramfort, Proc. NeurIPS Conf.]

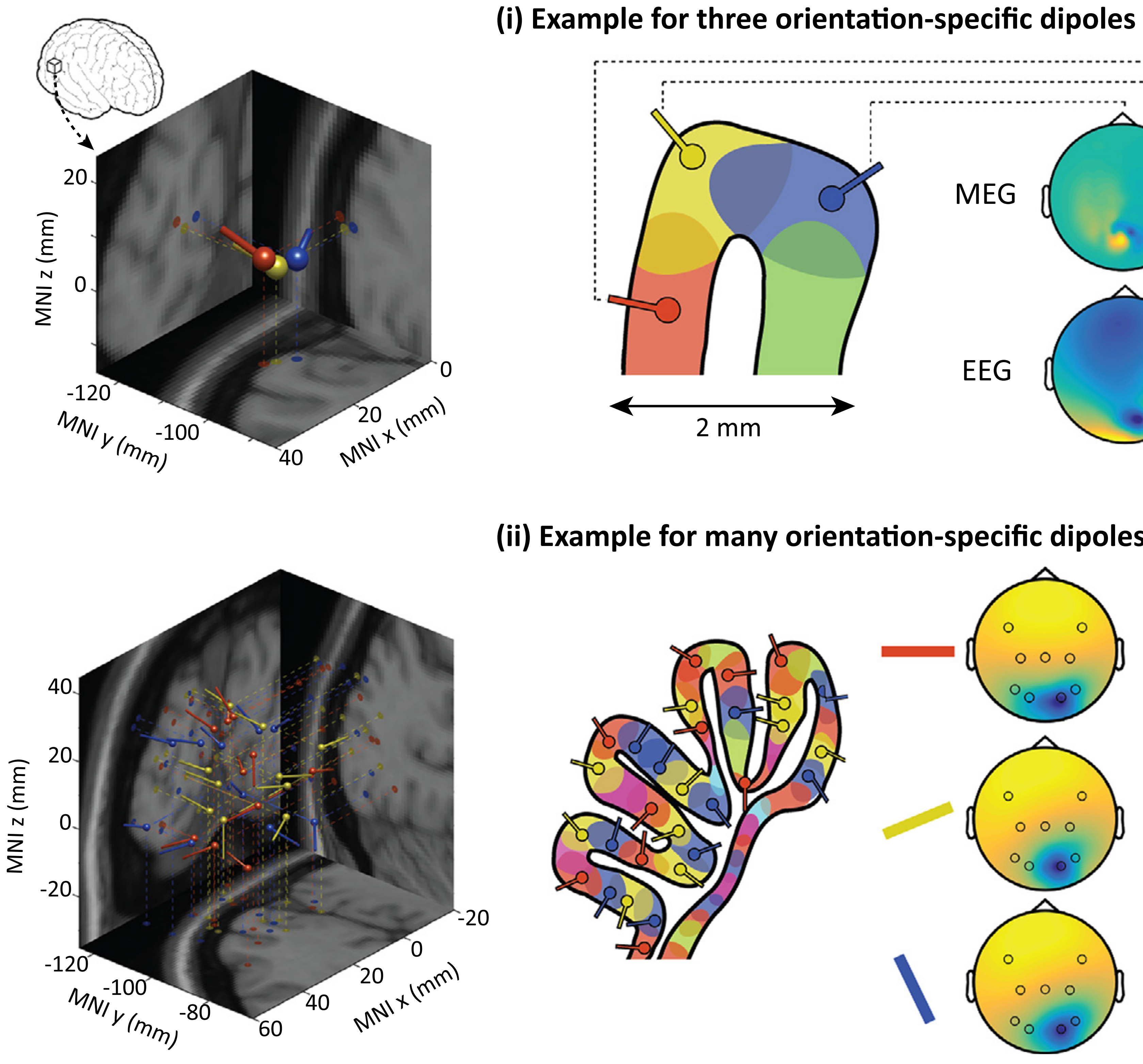
https://alphacsc.github.io

Why it works

(i) Example for three orientation-specific dipoles



[Decoding Rich Spatial Information with High Temporal Resolution, Stokes, Wolff & Spaak, TICS 2015]



New research suggests that magnetoencephalography (MEG) contains rich spatial information for decoding neural states. Even small differences in the angle of neighbouring dipoles generate subtle, but statistically separable field patterns. This implies MEG (and electroencephalography: EEG) is ideal for decoding neural states with high-temporal resolution in the human brain.

[Decoding Rich Spatial Information with High Temporal Resolution, Stokes, Wolff & Spaak, TICS 2015]

Multivariate models boost SNR

Consider 2 EEG signals such that:

$$x_1(t) = 2s(t) + n(t)$$
$$x_2(t) = s(t) + n(t)$$

where $s(t)$ is the source and $n(t)$ the additive noise, which is common.

Multivariate models boost SNR

Consider 2 EEG signals such that:

$$x_1(t) = 2s(t) + n(t)$$
$$x_2(t) = s(t) + n(t)$$

where $s(t)$ is the source and $n(t)$ the additive noise, which is common.

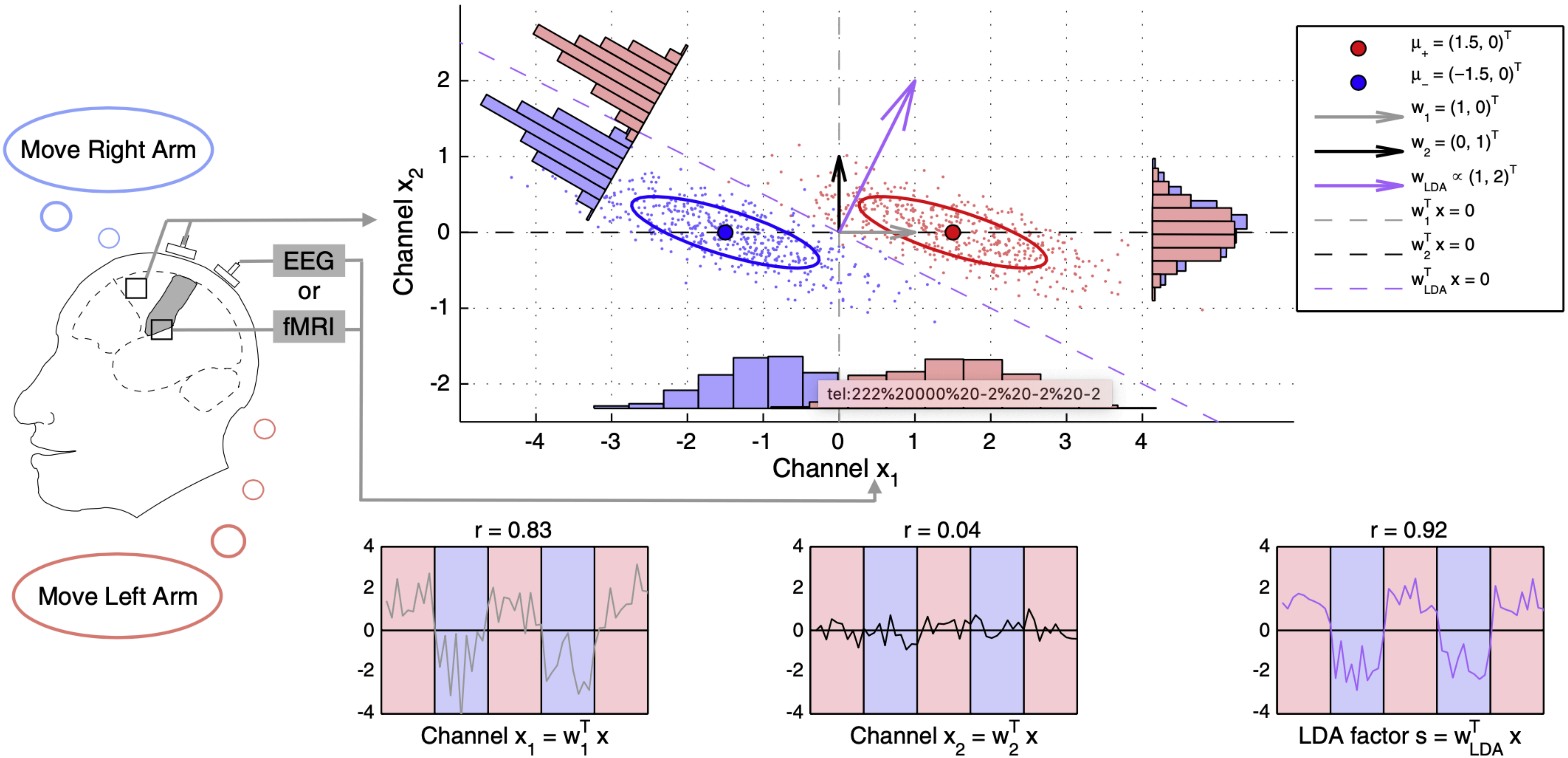
This implies that:

$$x_1(t) - x_2(t) = s(t)$$

Only signal remains !

THM: By subtracting the signals on can improve the SNR even if noise is huge

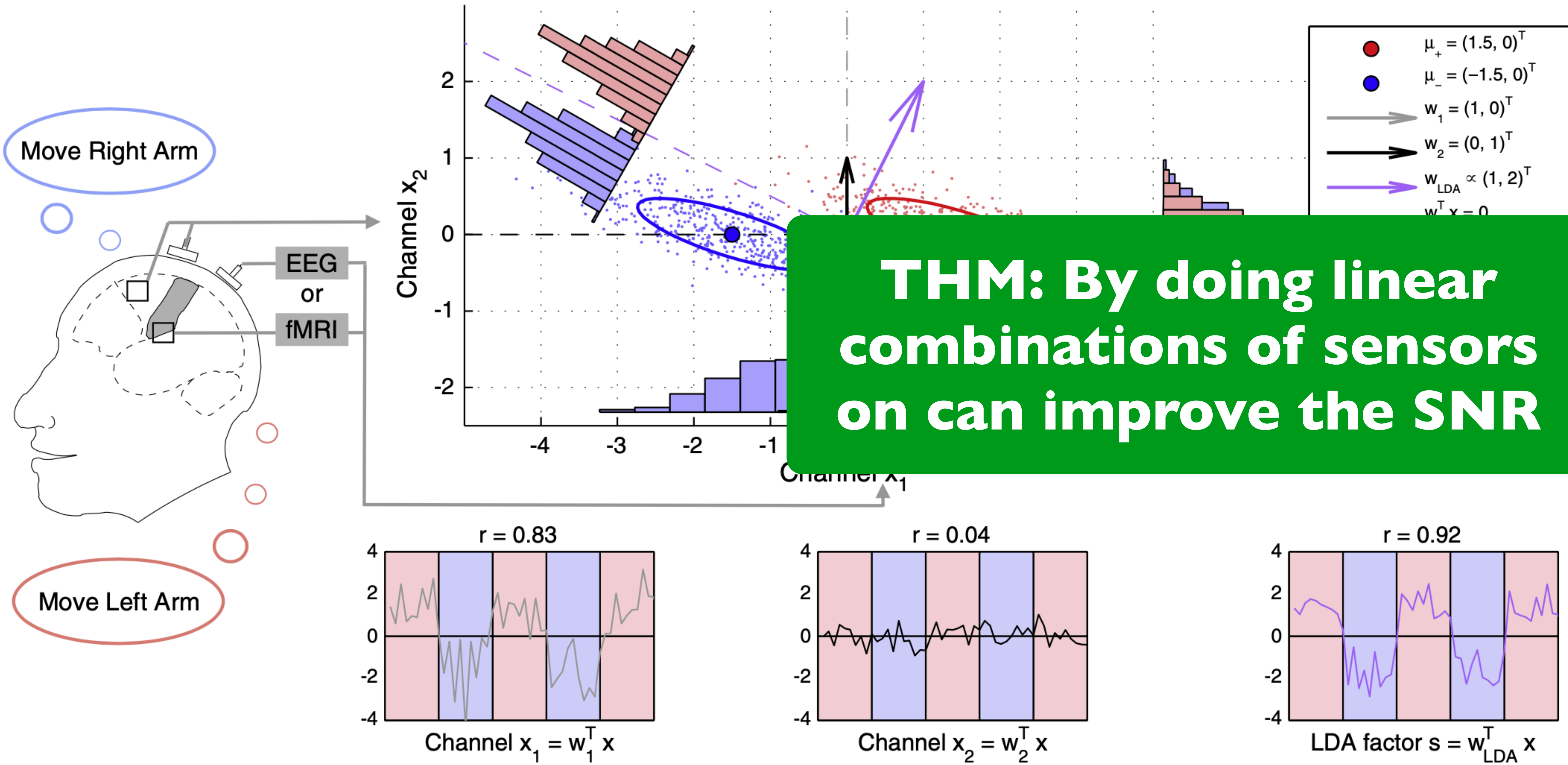
Multivariate models boost SNR



[Haufe et al. (2014) On the interpretation of weight vectors of linear models in multivariate neuroimaging, NeuroImage]

https://mne.tools/stable/auto_examples/decoding/plot_linear_model_patterns.html

Multivariate models boost SNR



THM: By doing linear combinations of sensors on can improve the SNR

[Haufe et al. (2014) On the interpretation of weight vectors of linear models in multivariate neuroimaging, NeuroImage]

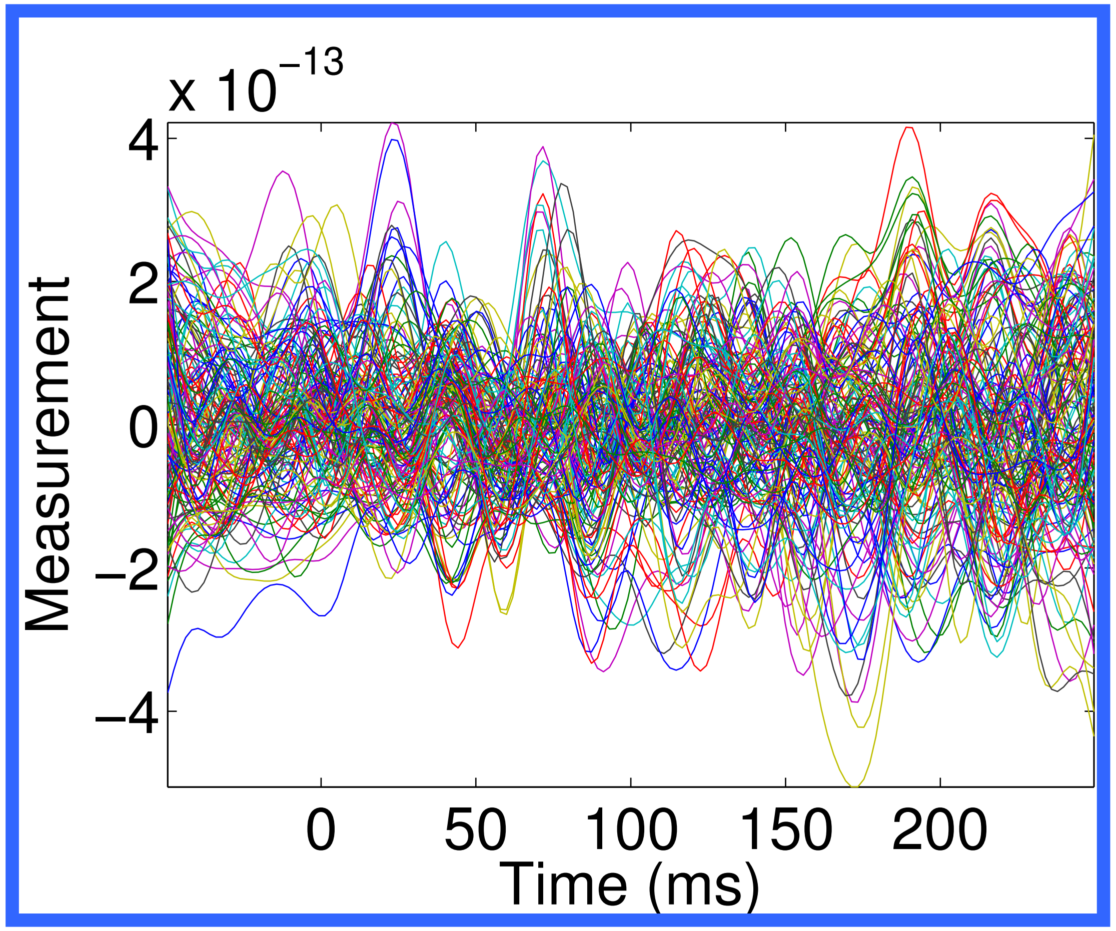
https://mne.tools/stable/auto_examples/decoding/plot_linear_model_patterns.html

Induced and Evoked M/EEG activity

requires different ML approaches

Evoked activity

Evoked activity is revealed by trial averaging:

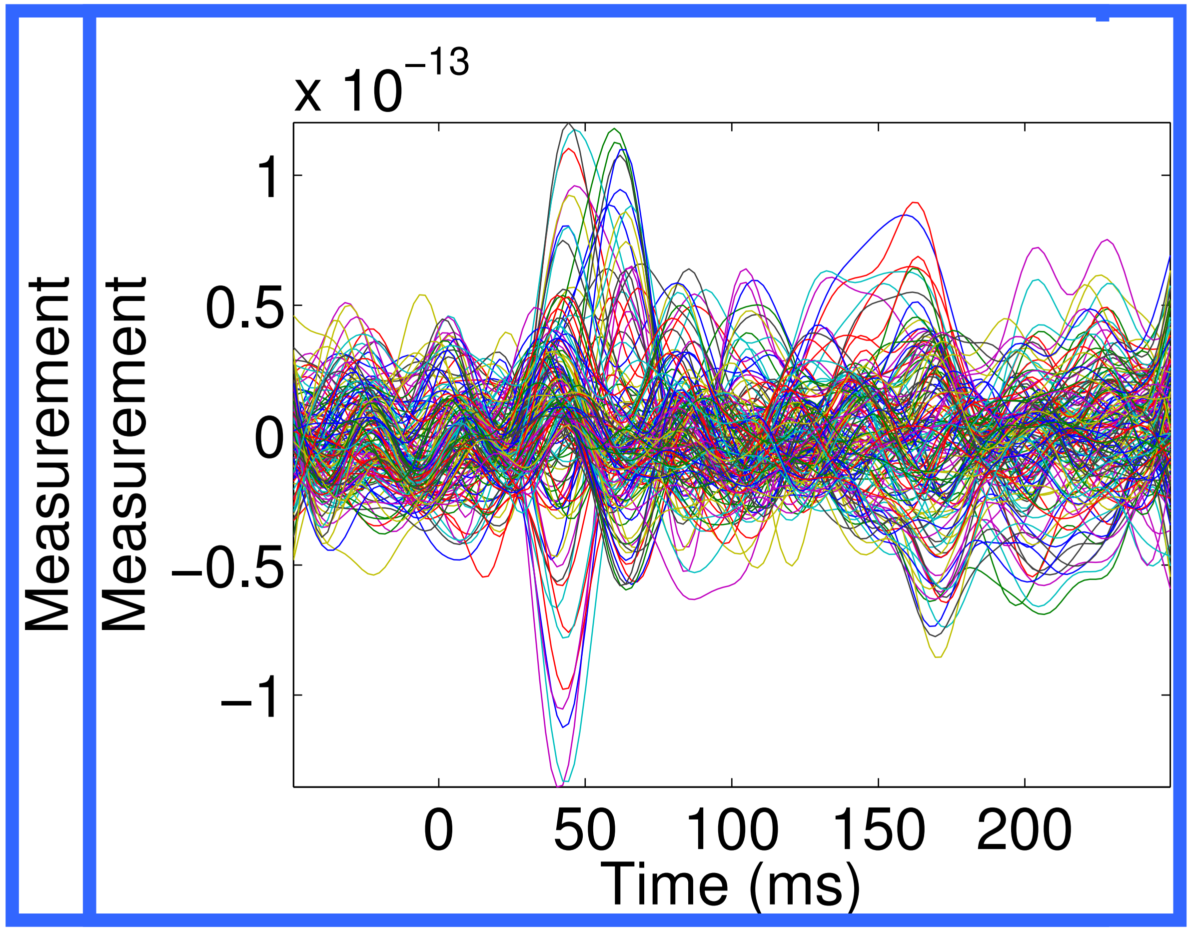


1 trial

Signal on 151 MEG Channels

Evoked activity

Evoked activity is revealed by trial averaging:

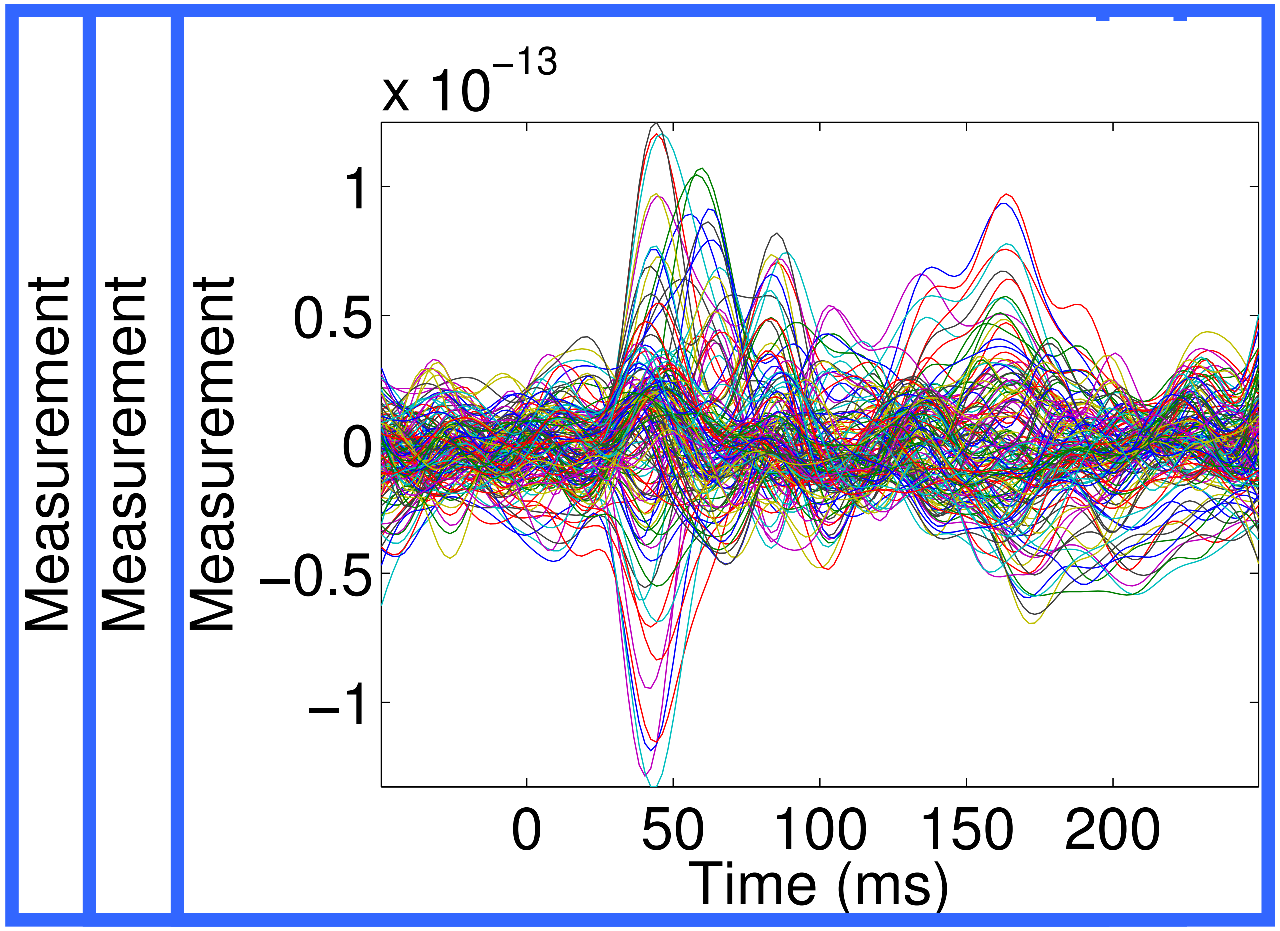


50 trials

Signal
on
151
MEG
Channels

Evoked activity

Evoked activity is revealed by trial averaging:

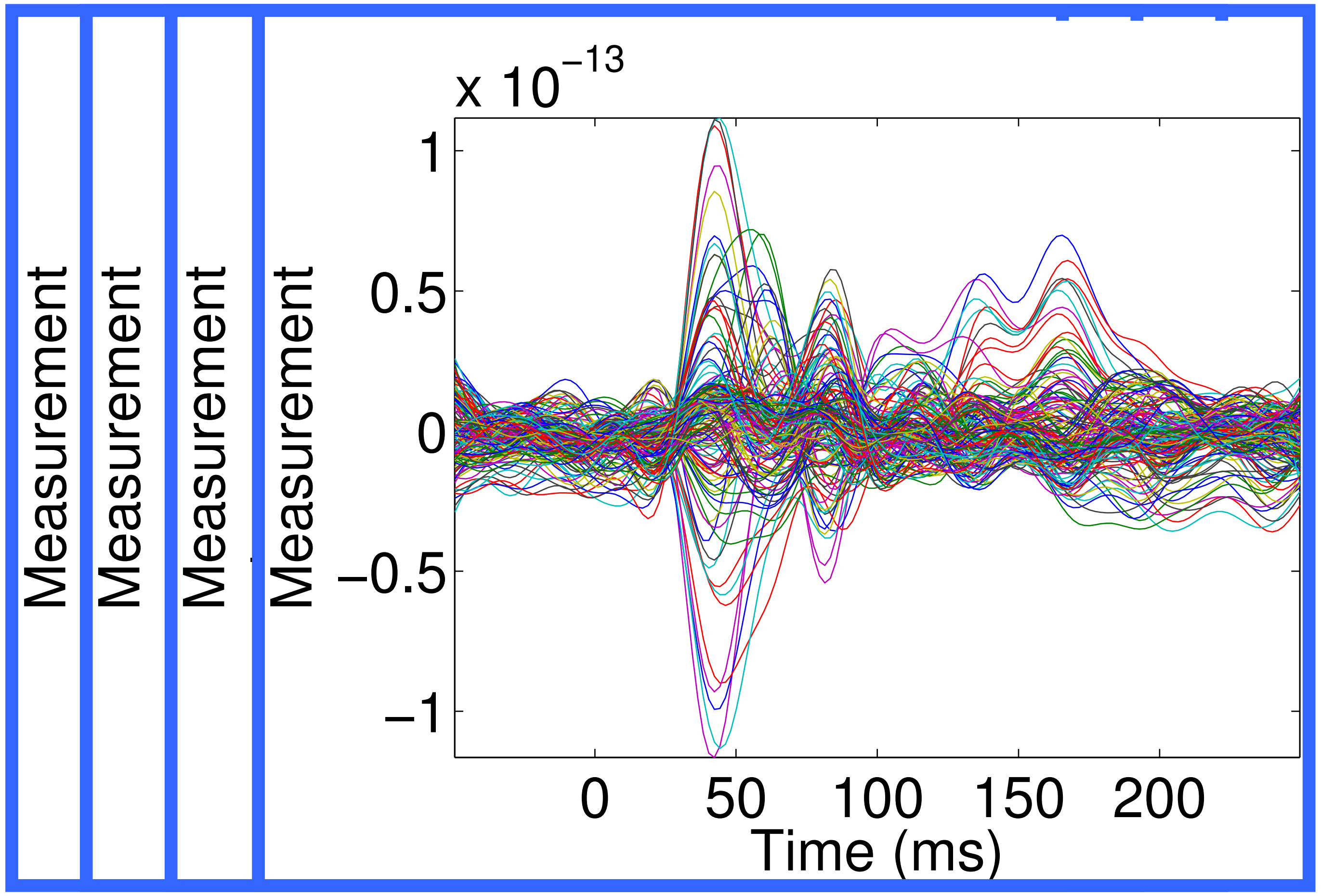


Signal
on
151
MEG
Channels

100 trials

Evoked activity

Evoked activity is revealed by trial averaging:

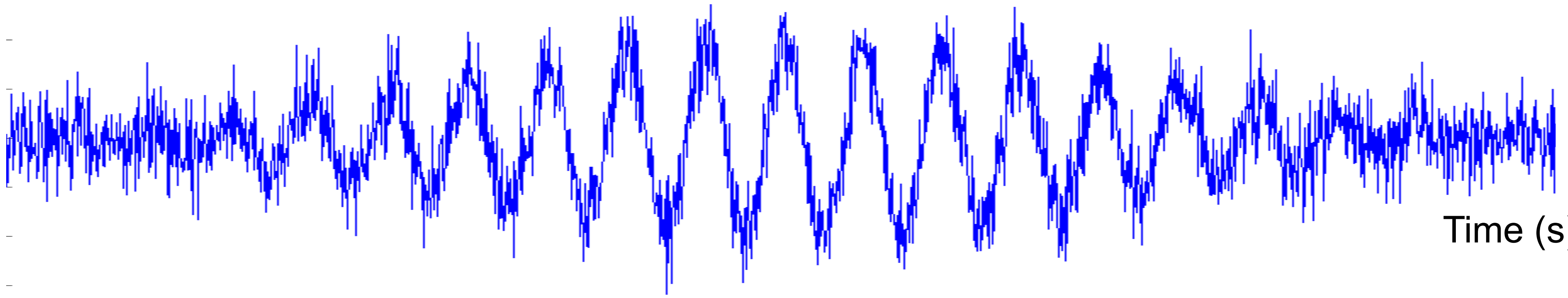


Signal on 151 MEG Channels

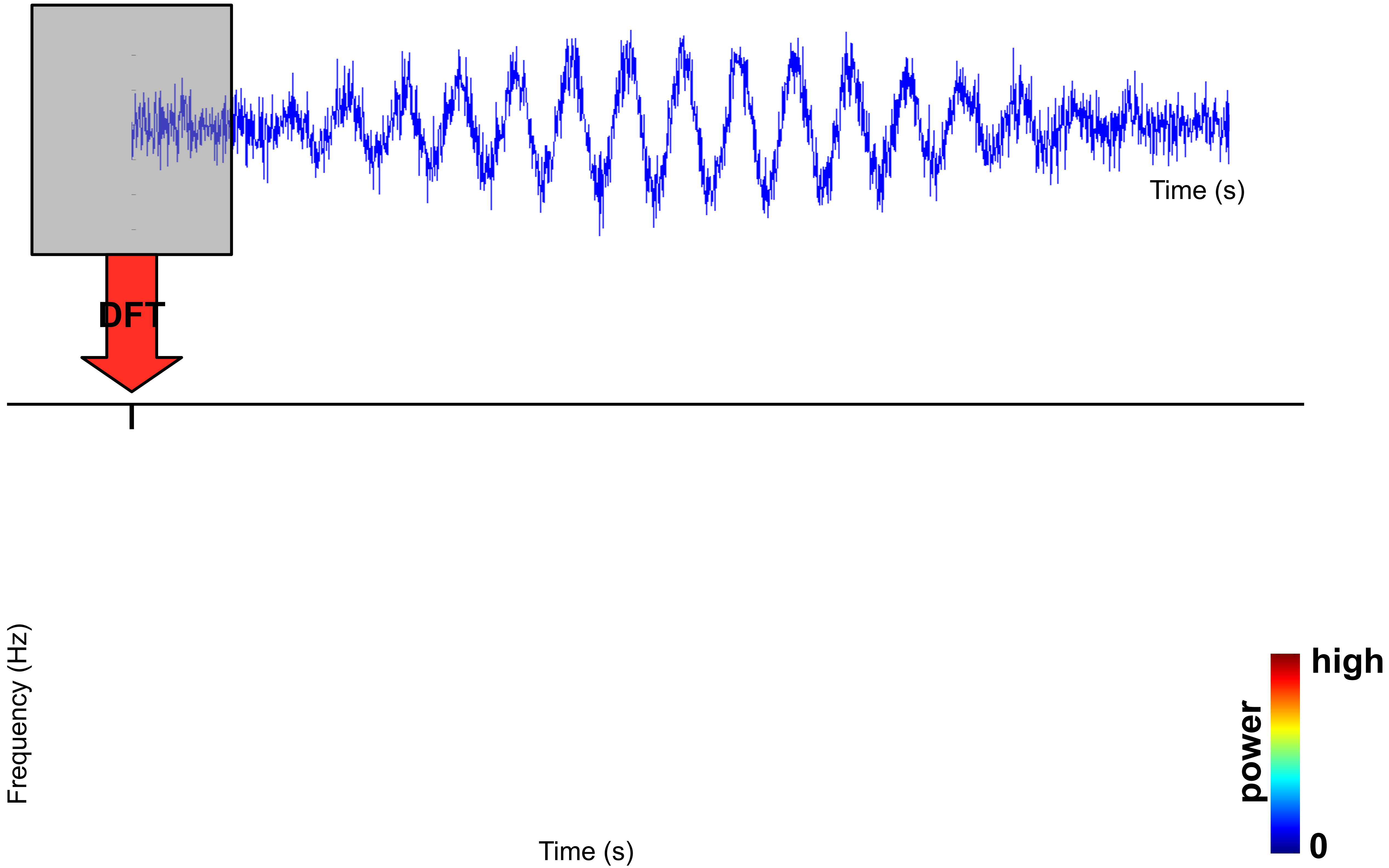
400 trials

BUT: Does not work if the phase/latency are different over trials

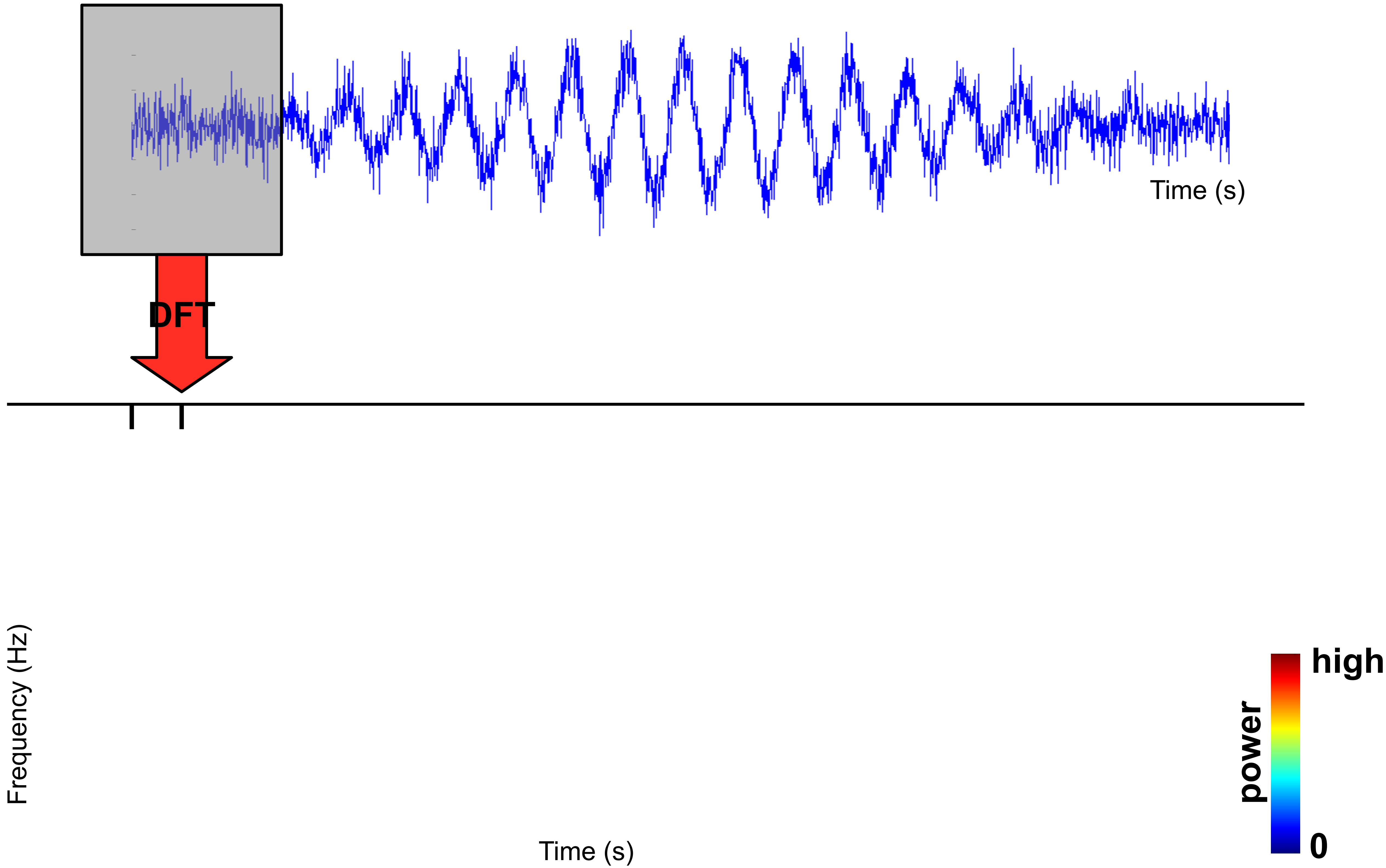
Time frequency analysis



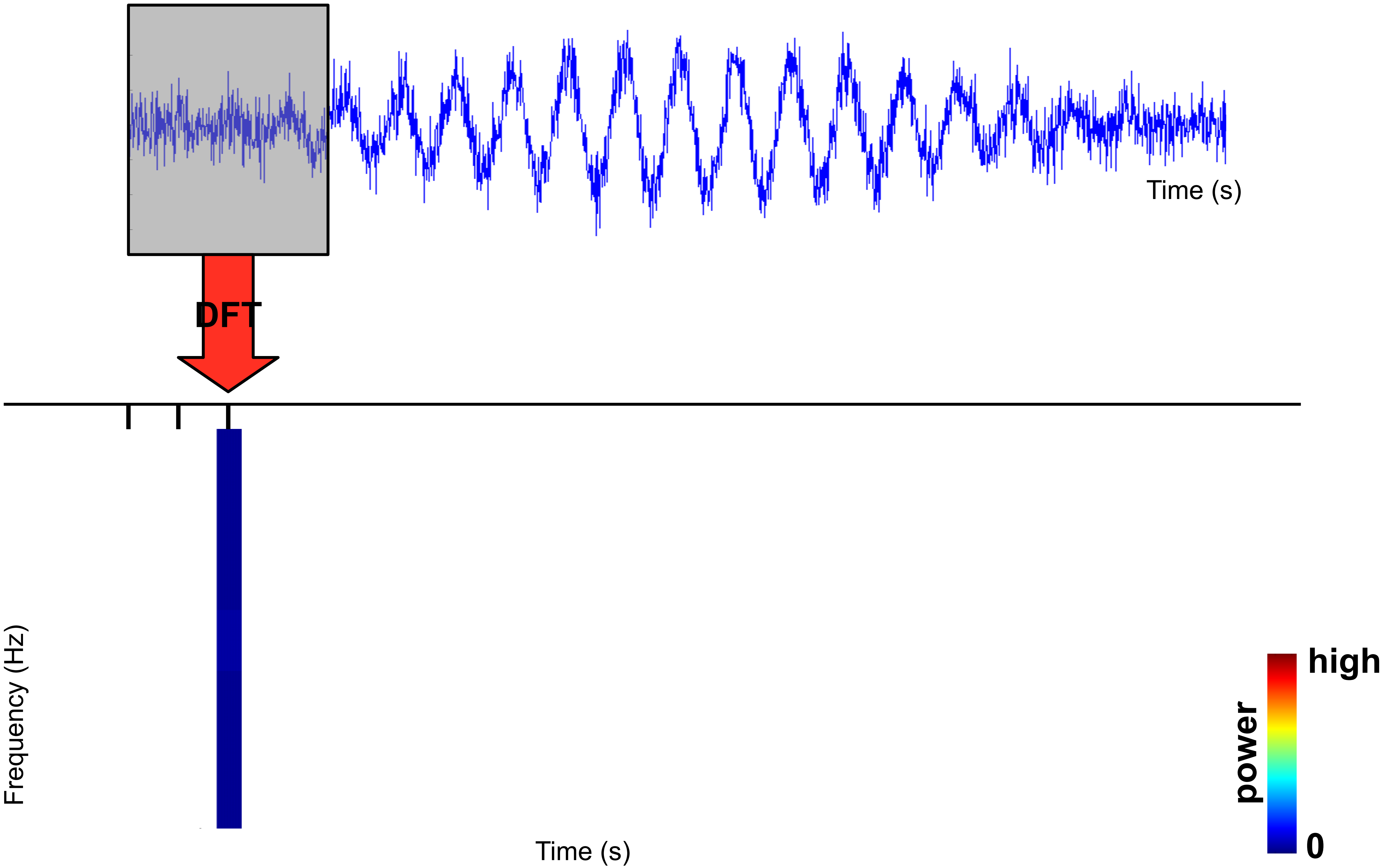
Time frequency analysis



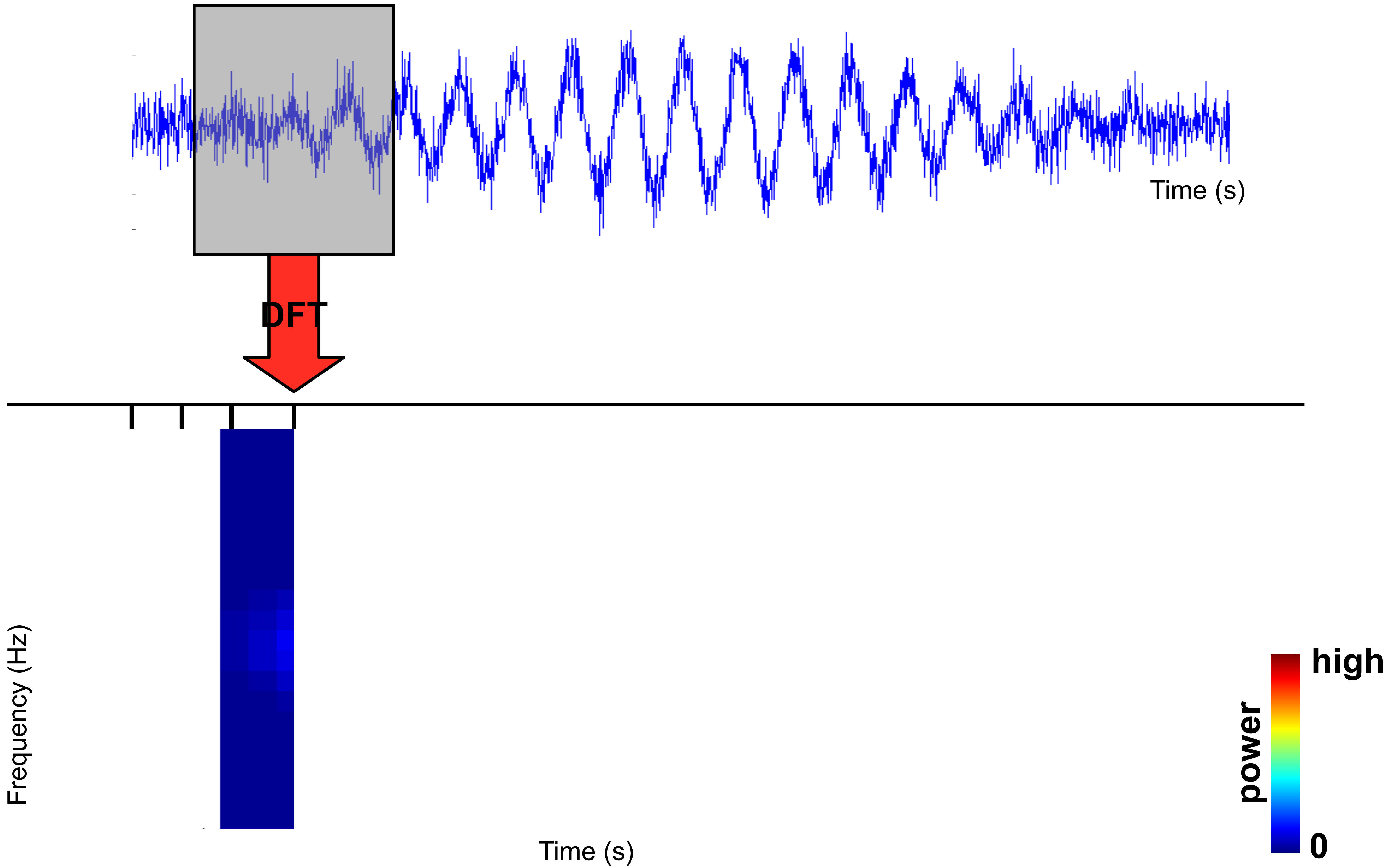
Time frequency analysis



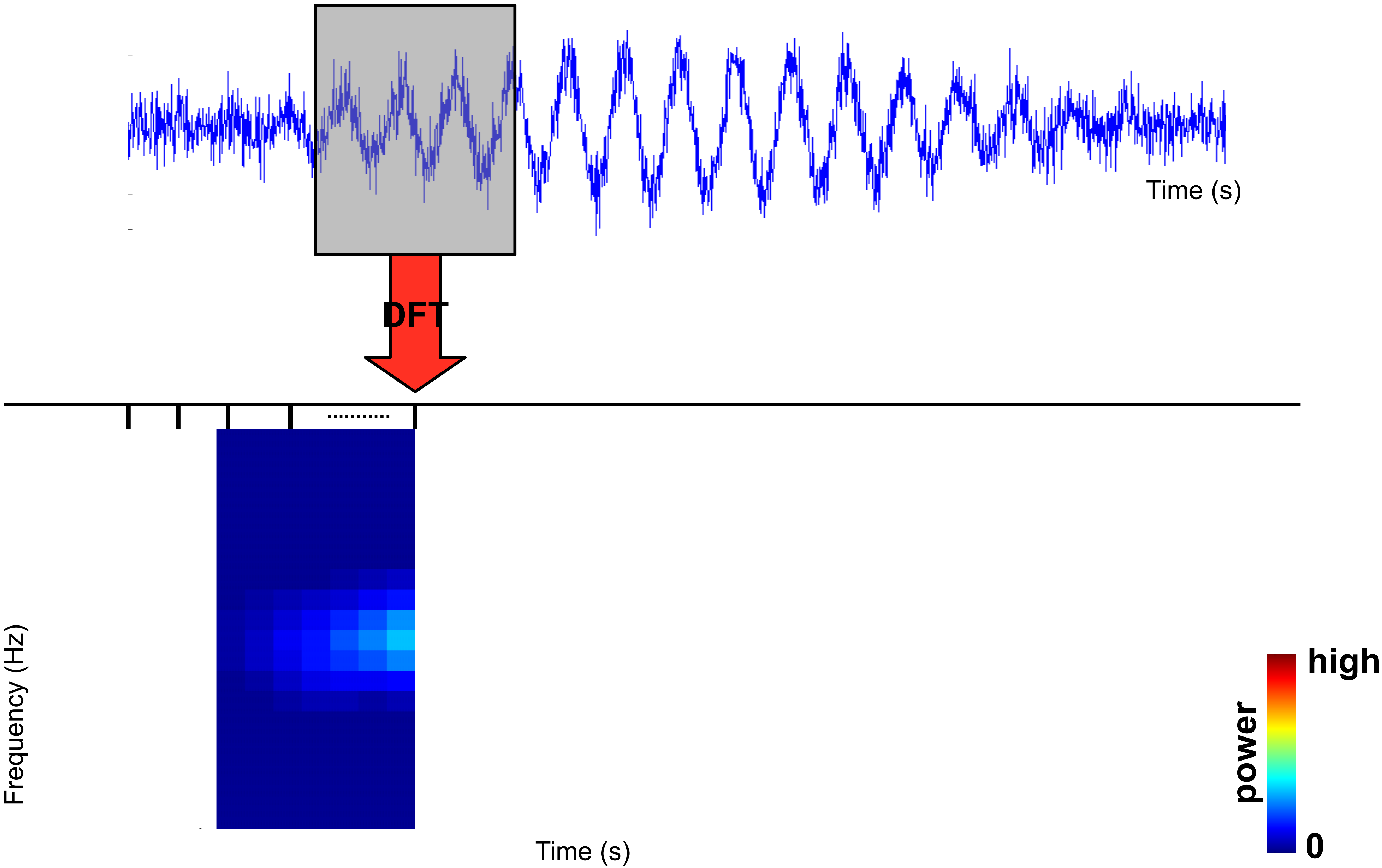
Time frequency analysis



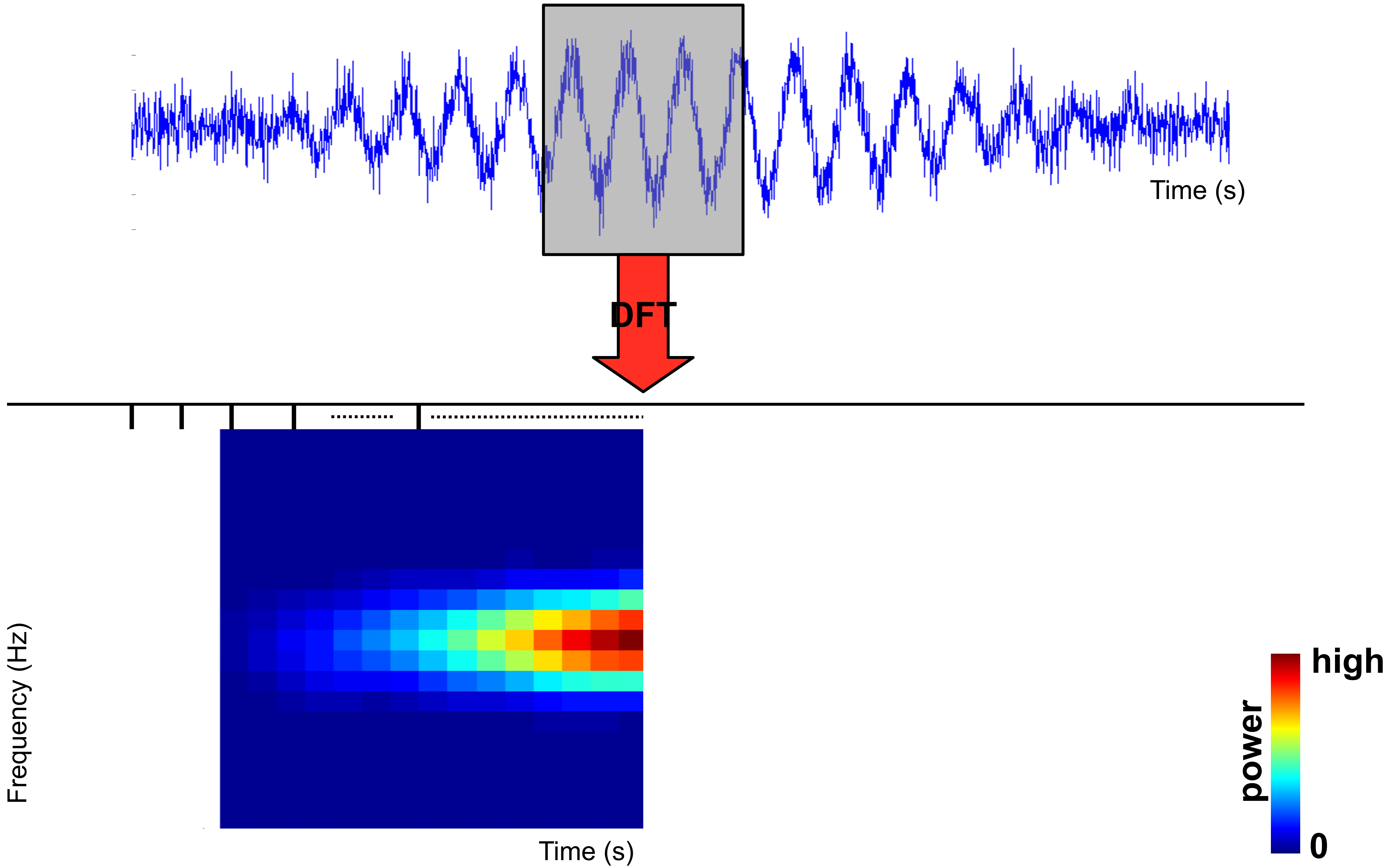
Time frequency analysis



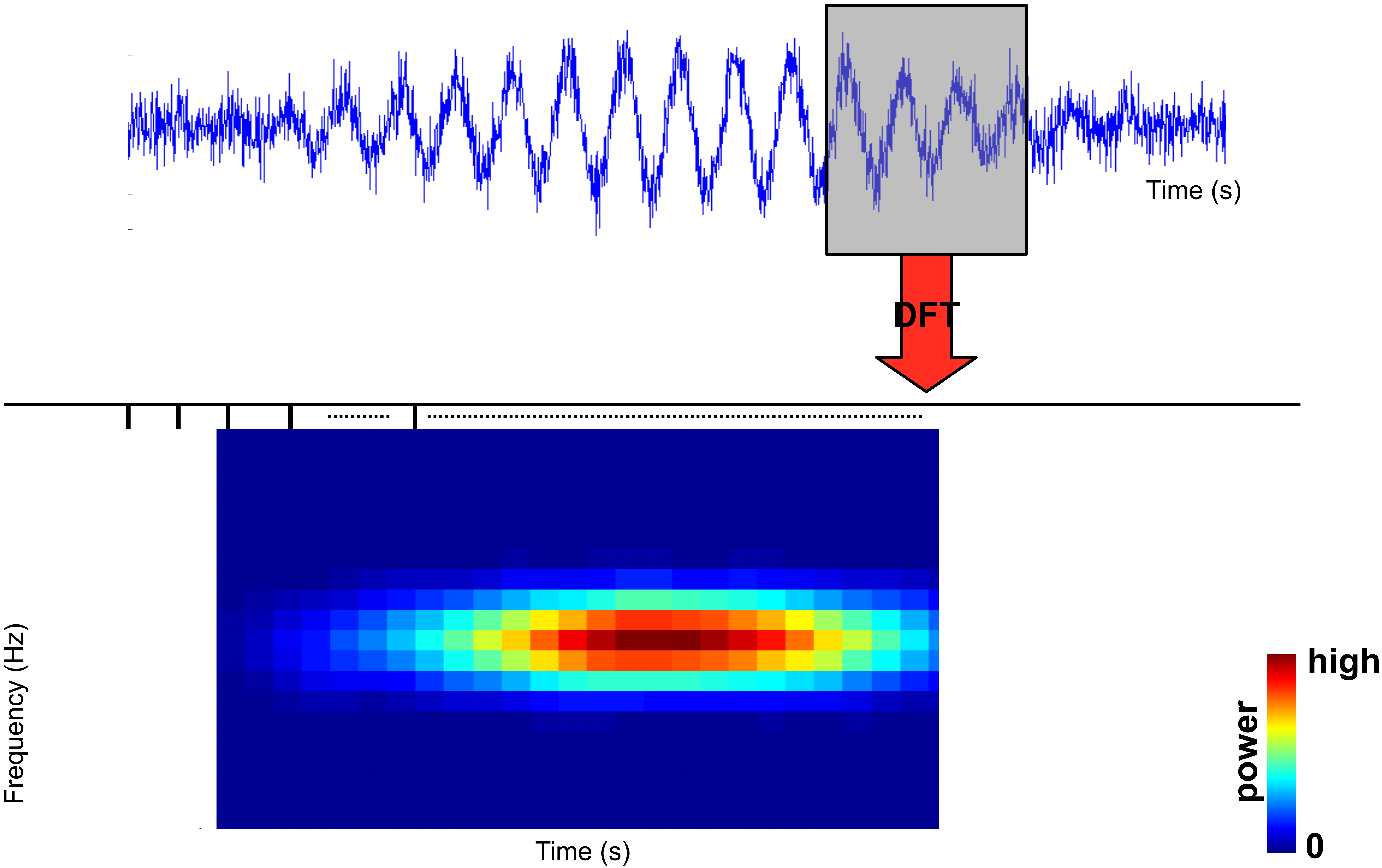
Time frequency analysis



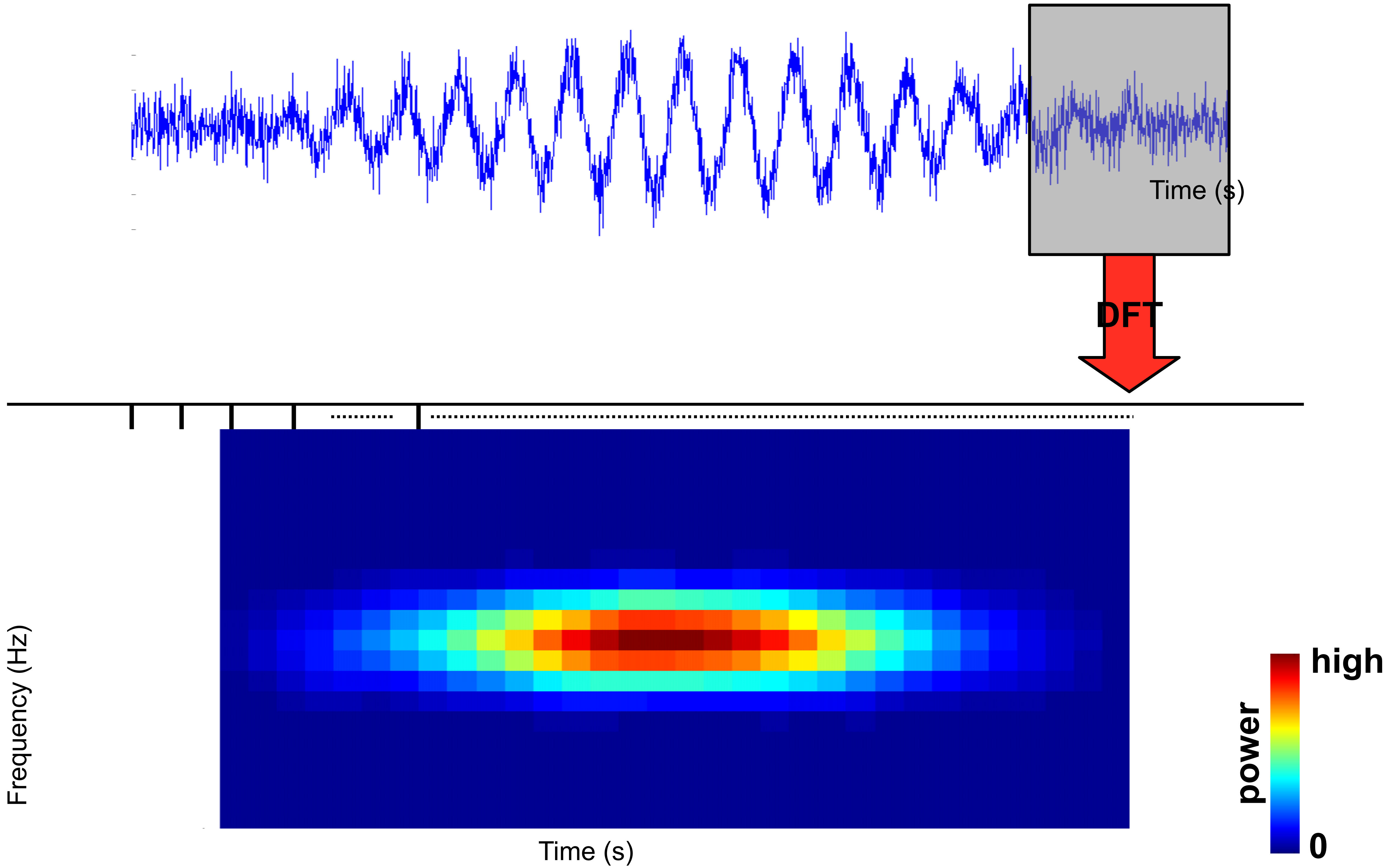
Time frequency analysis



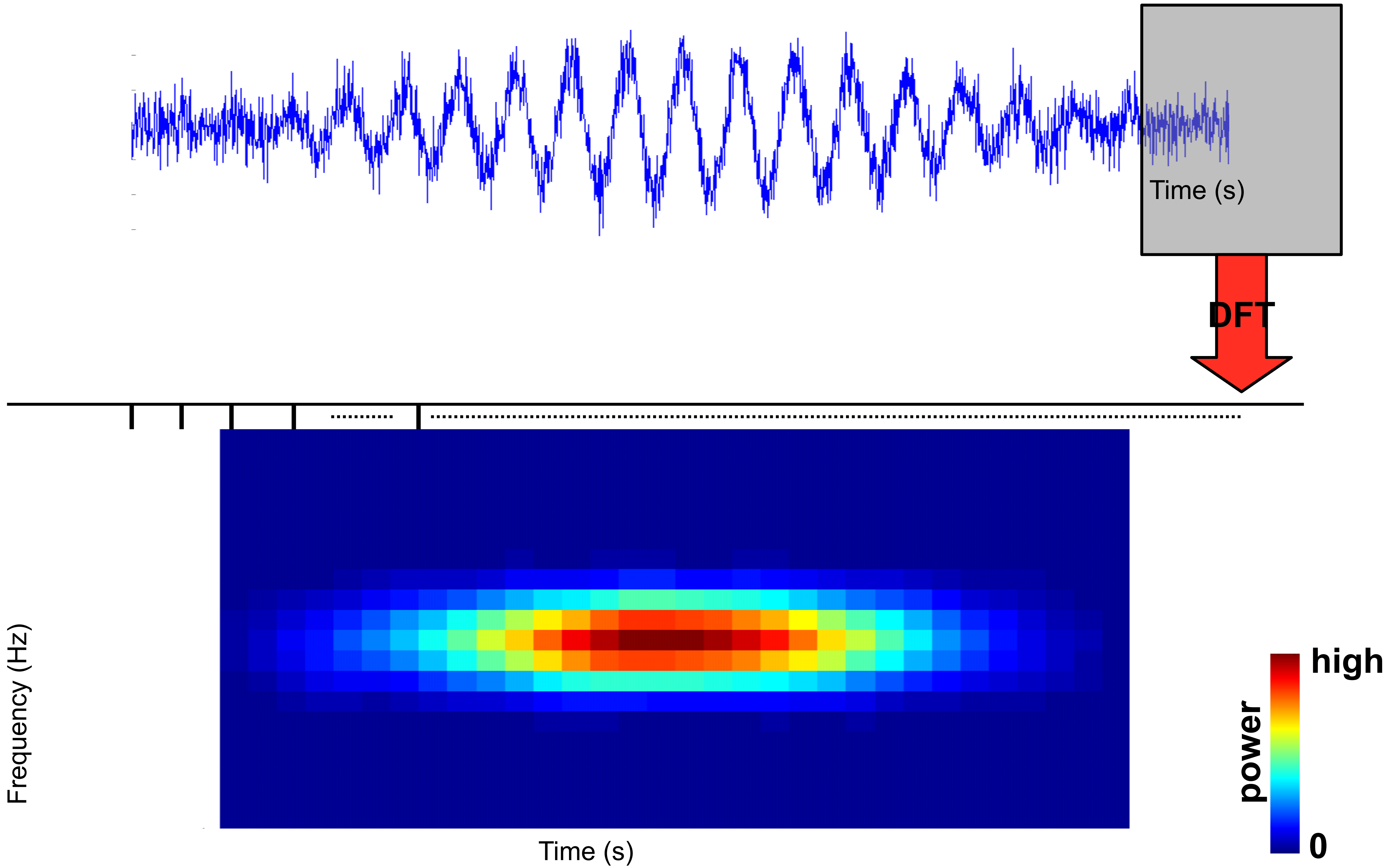
Time frequency analysis



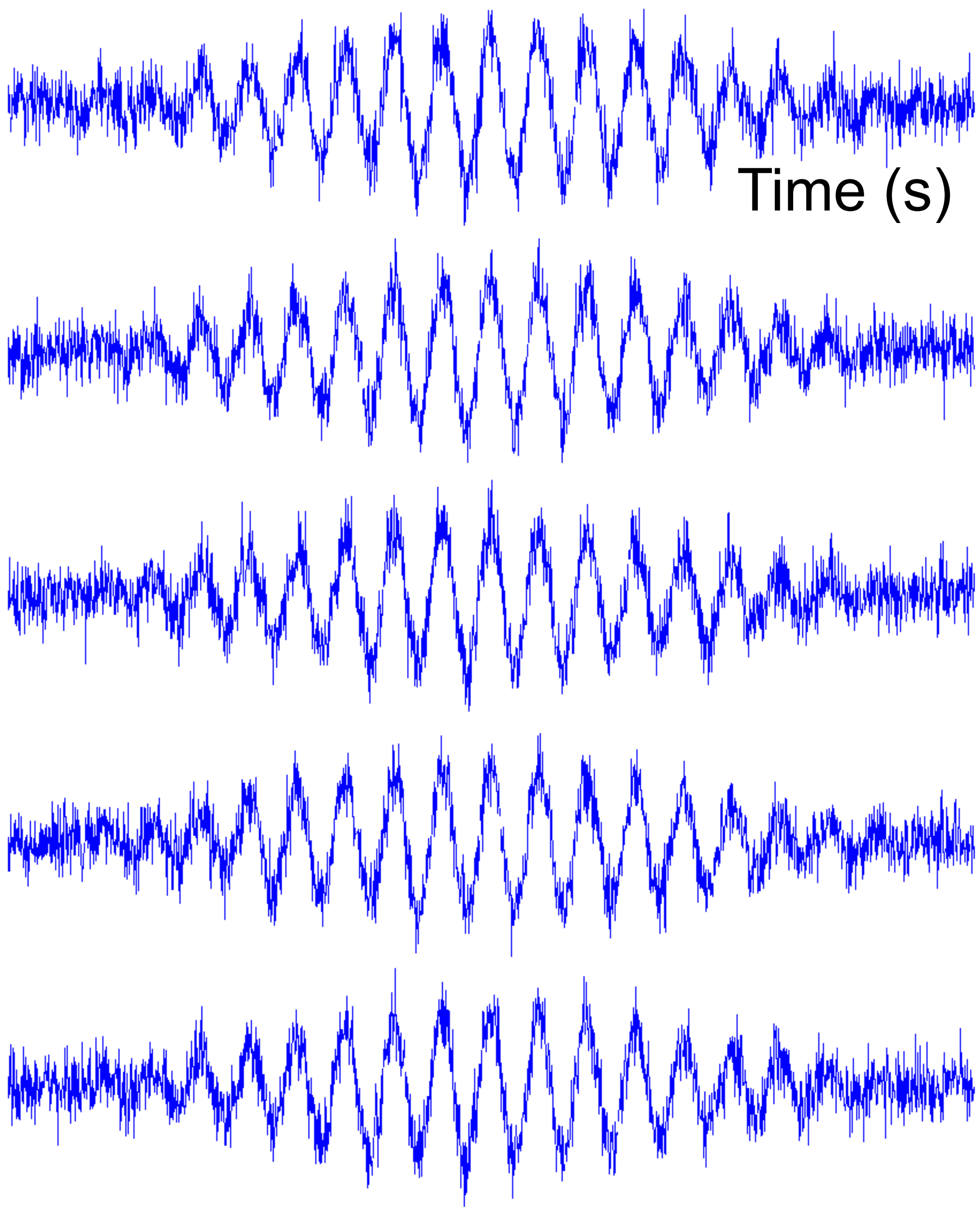
Time frequency analysis



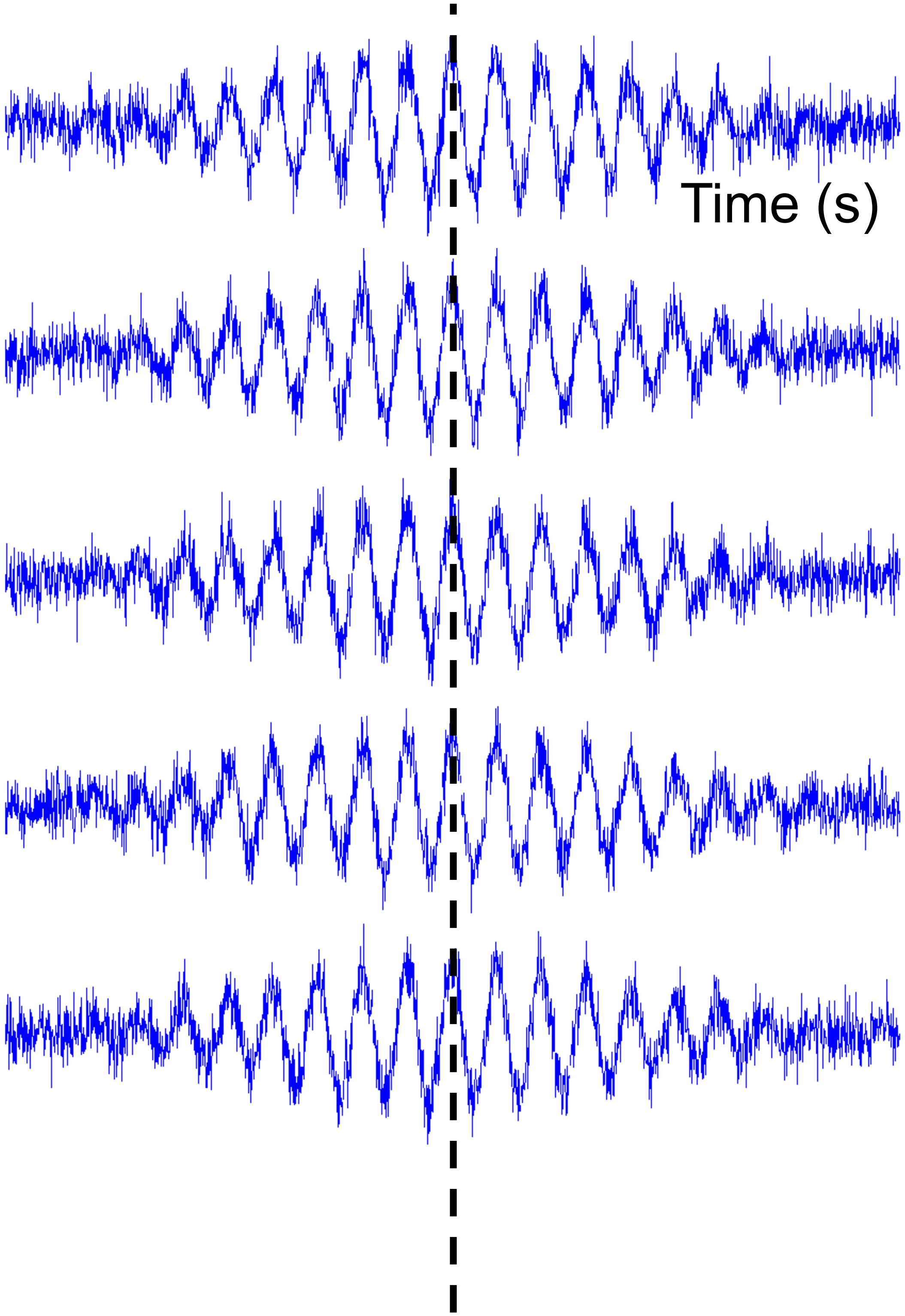
Time frequency analysis



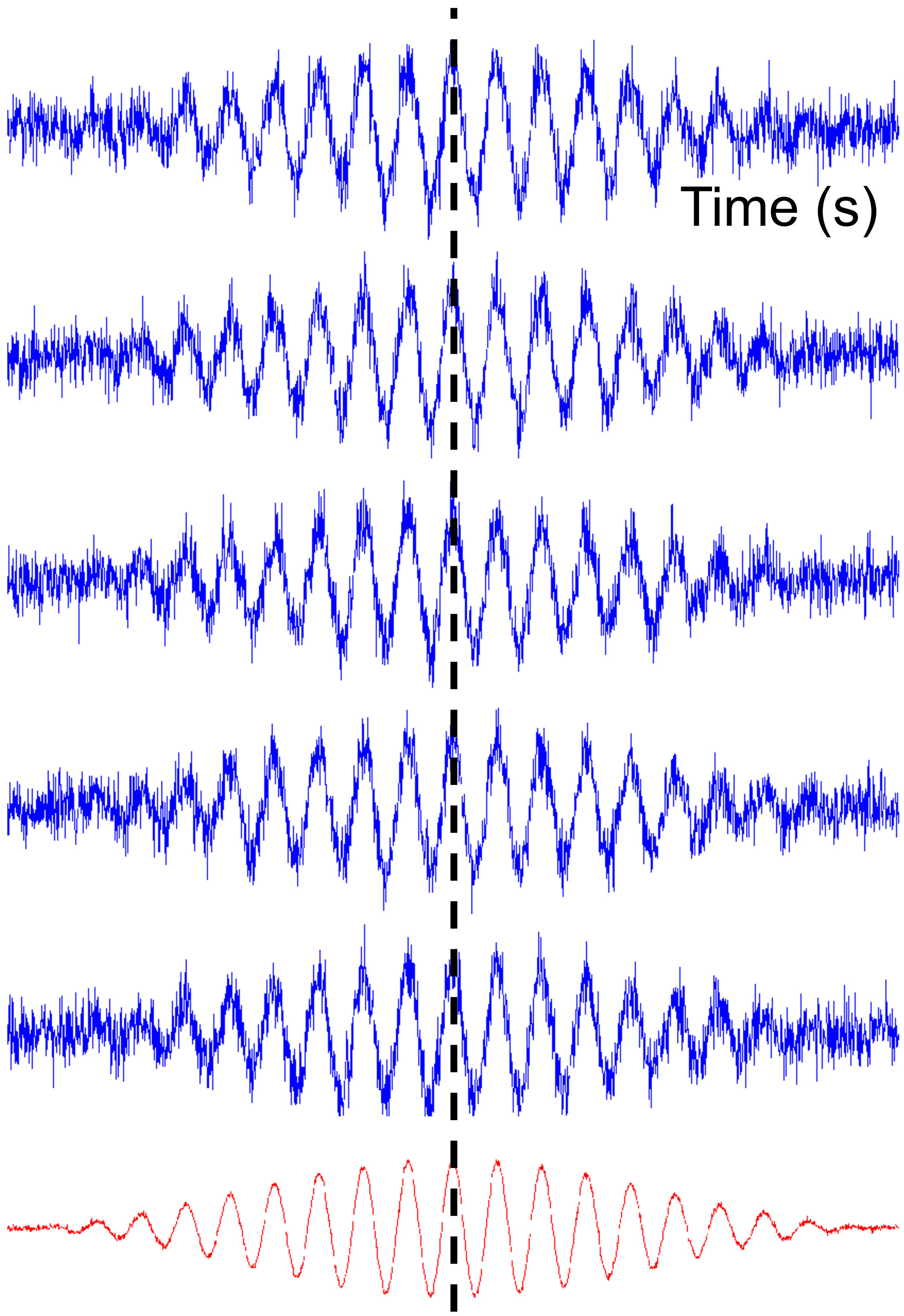
Evoked vs. induced activity



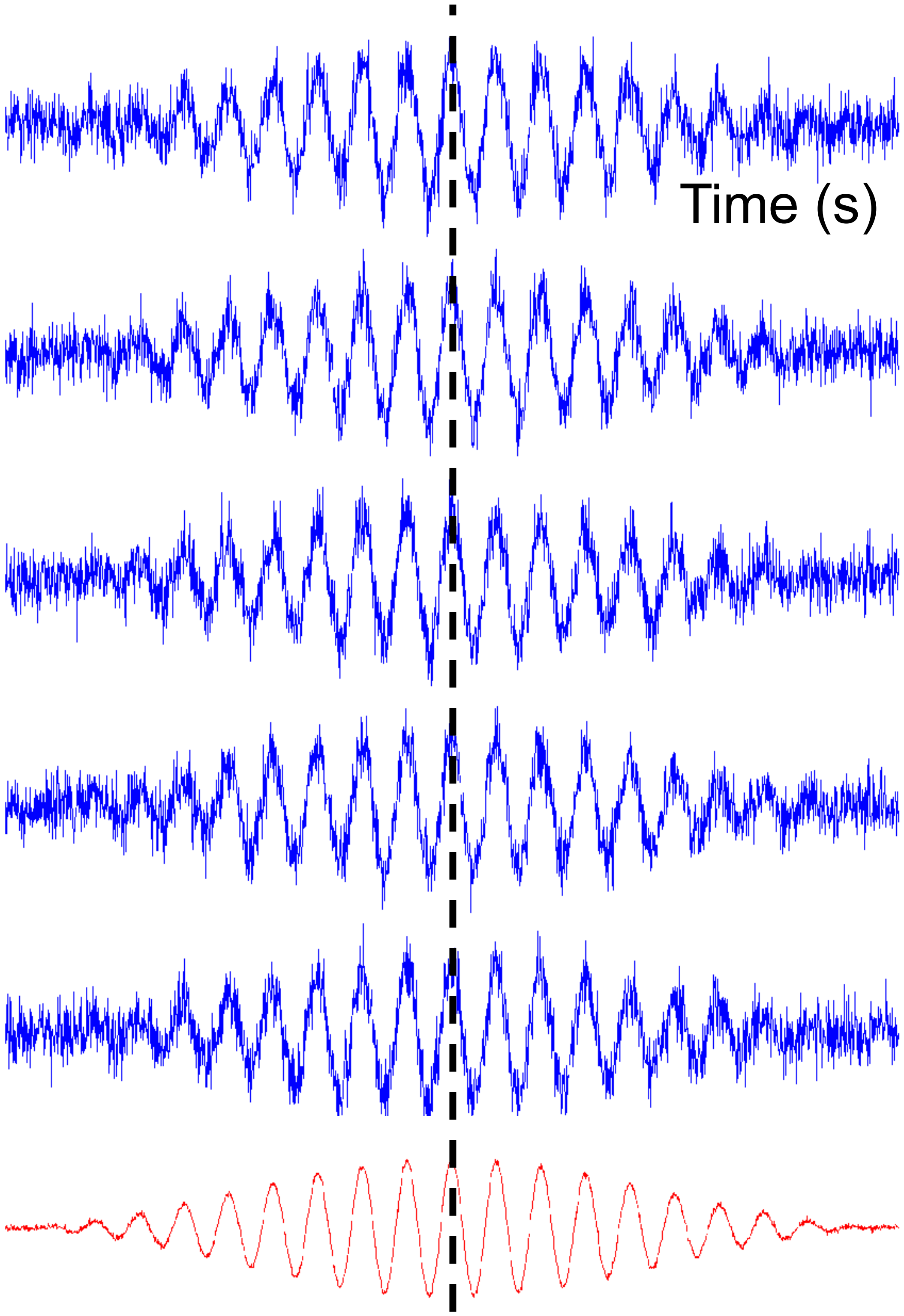
Evoked vs. induced activity

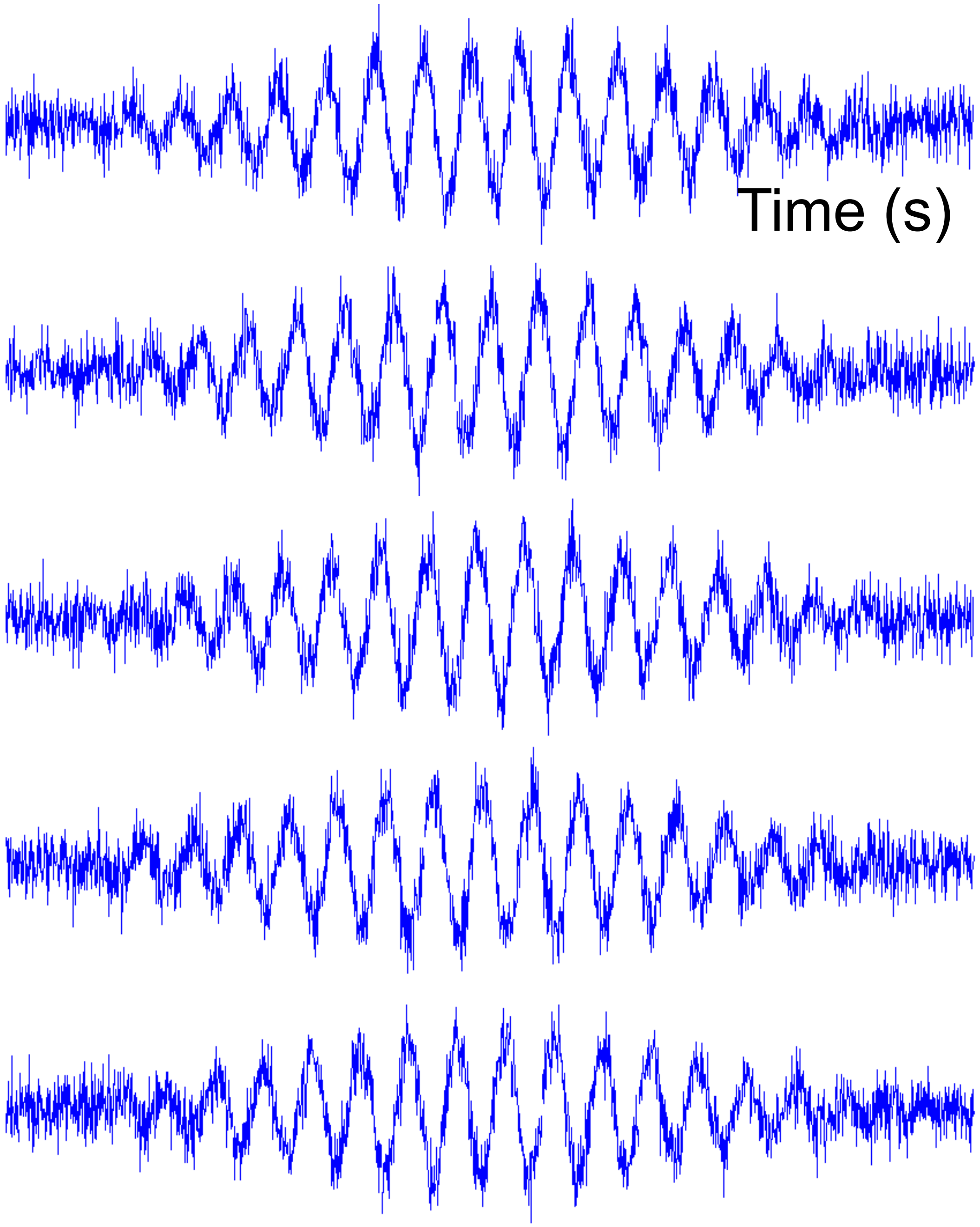


Evoked vs. induced activity

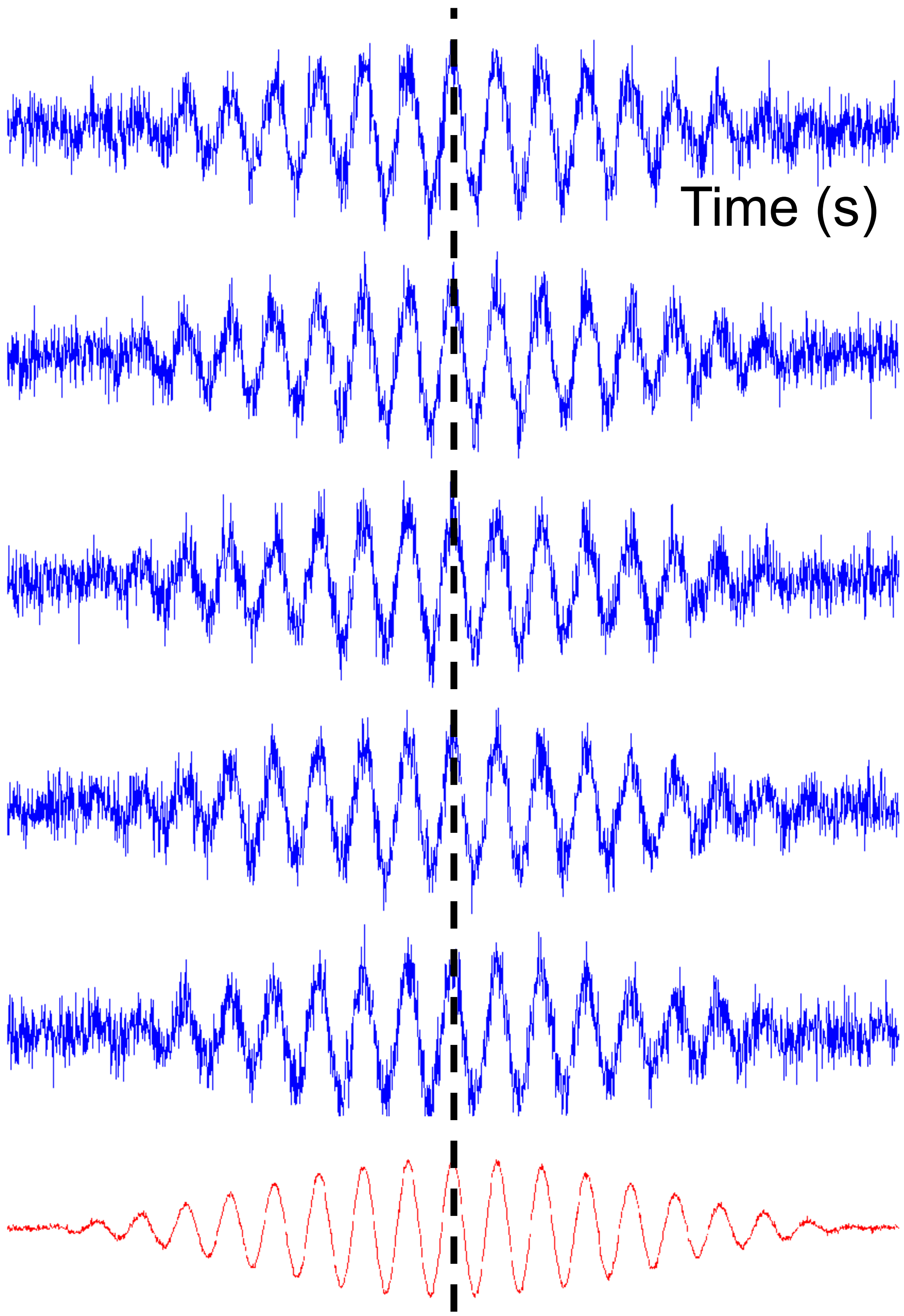


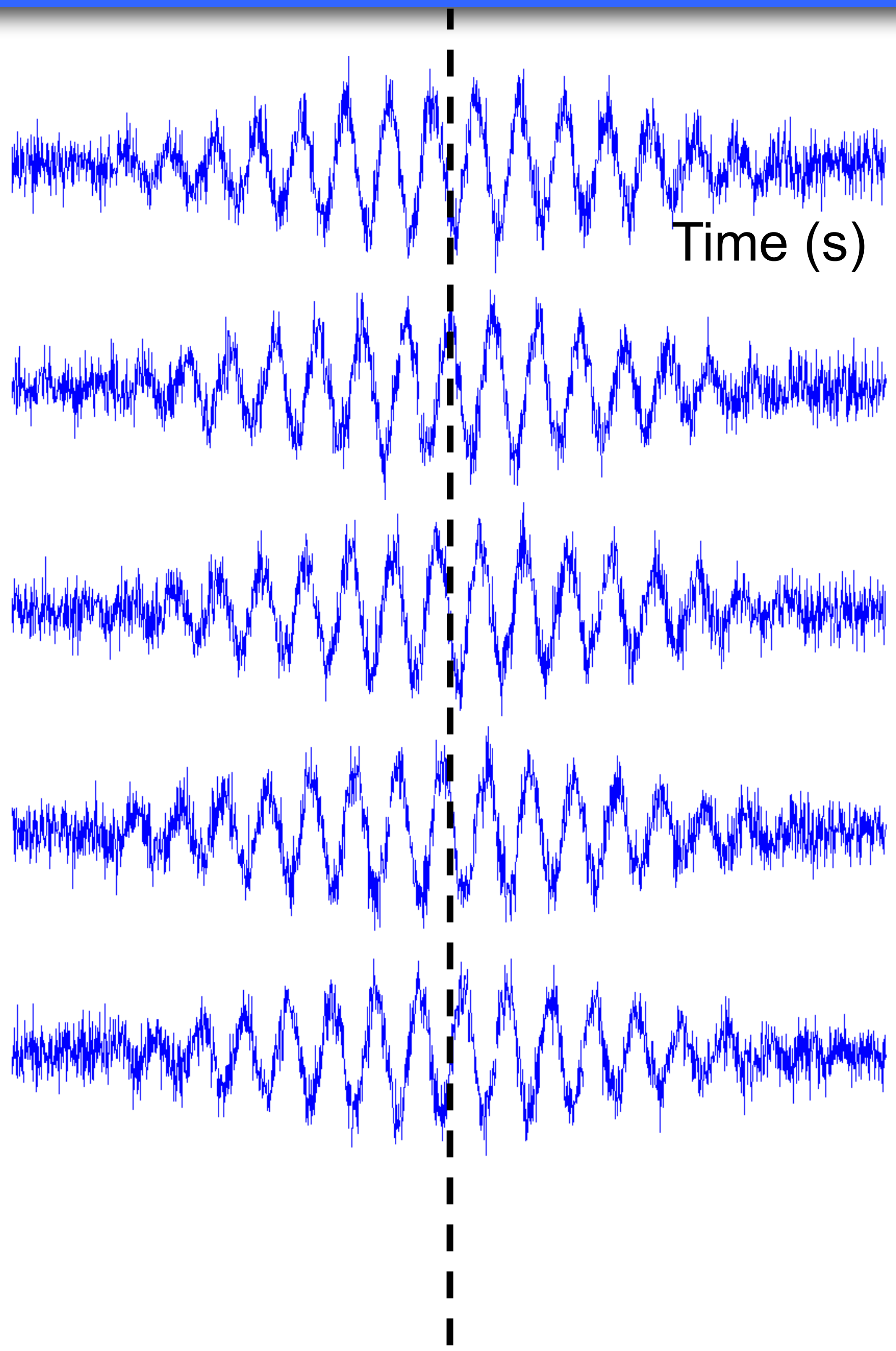
Evoked vs. induced activity



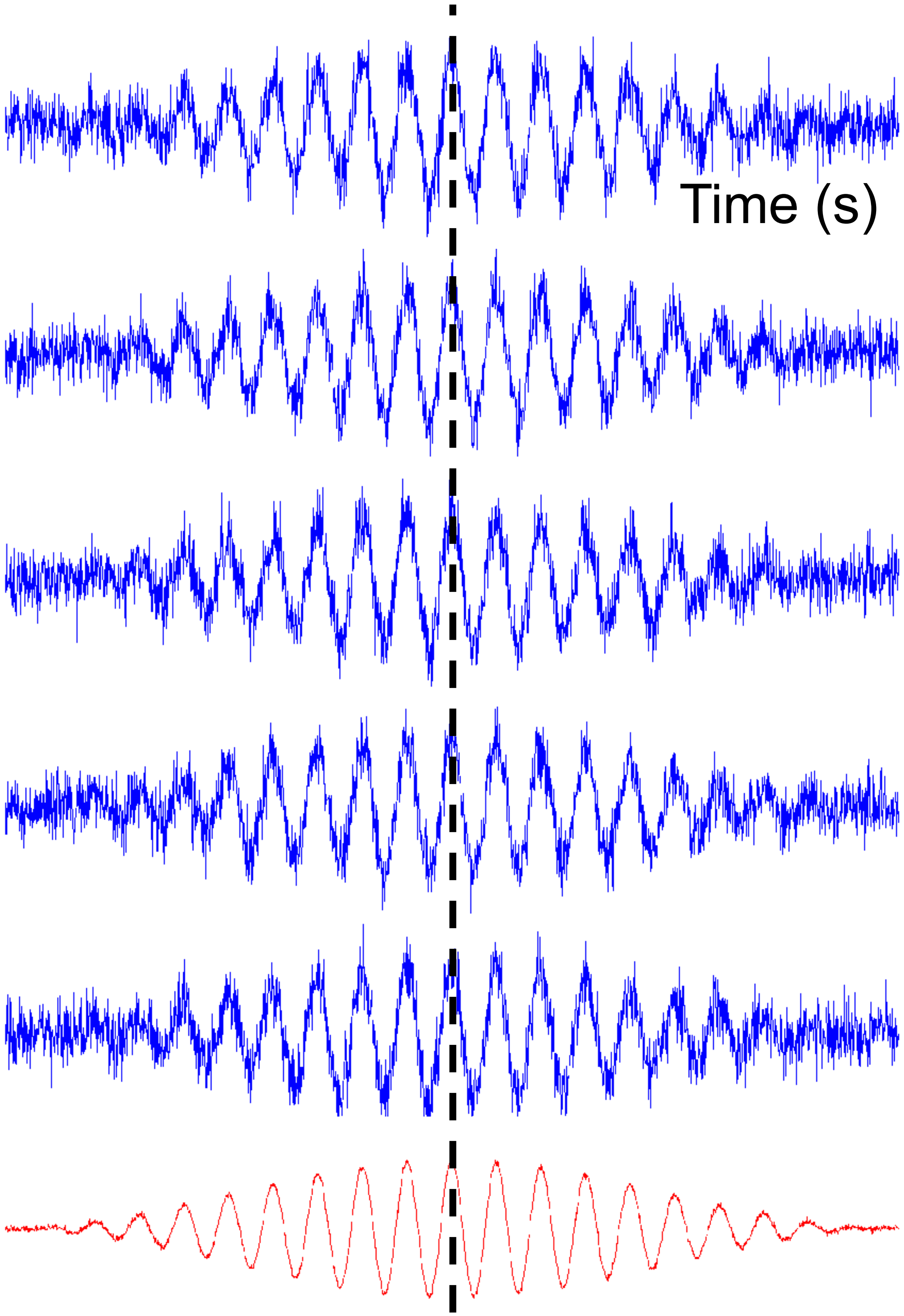


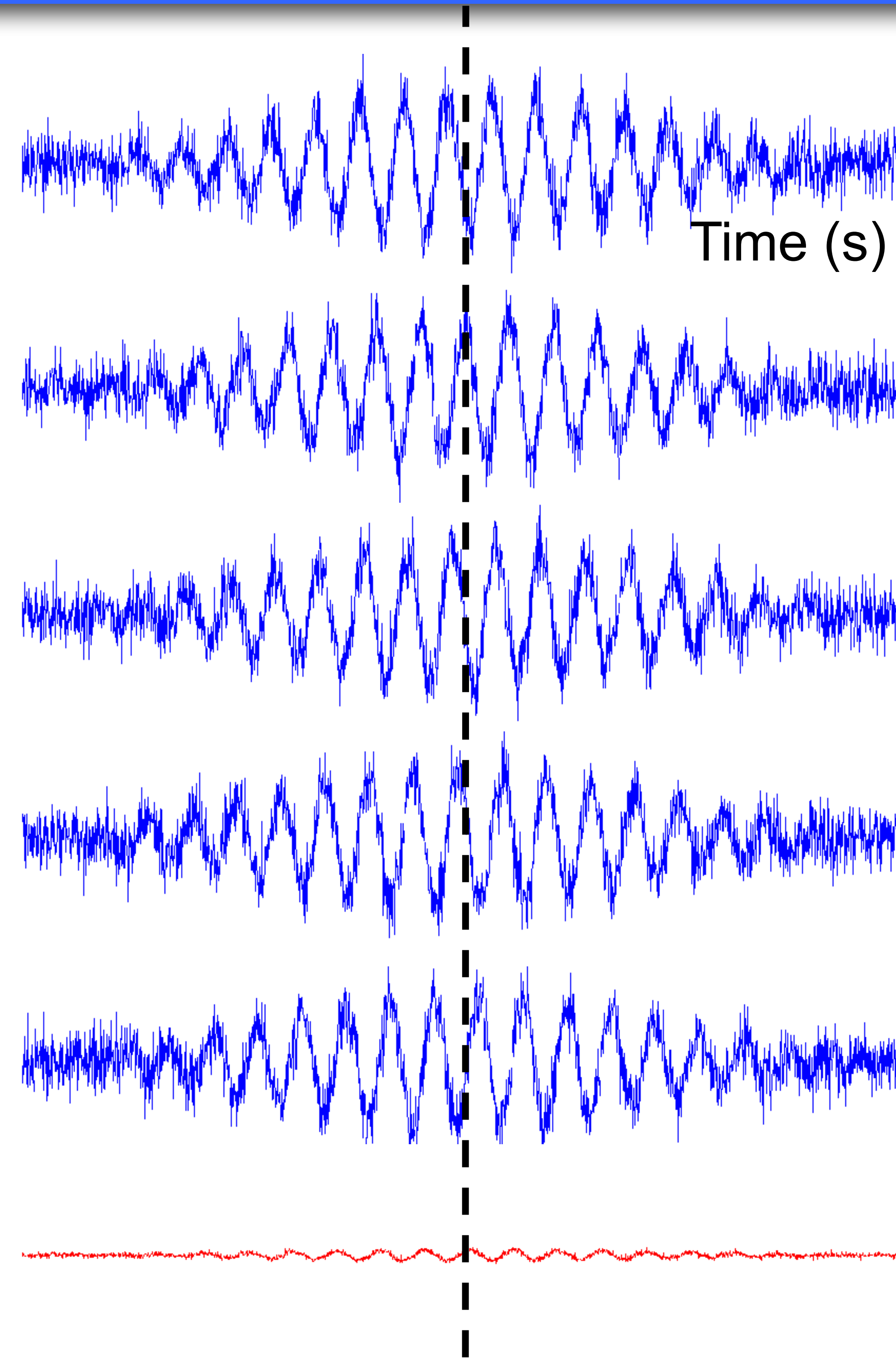
Evoked vs. induced activity



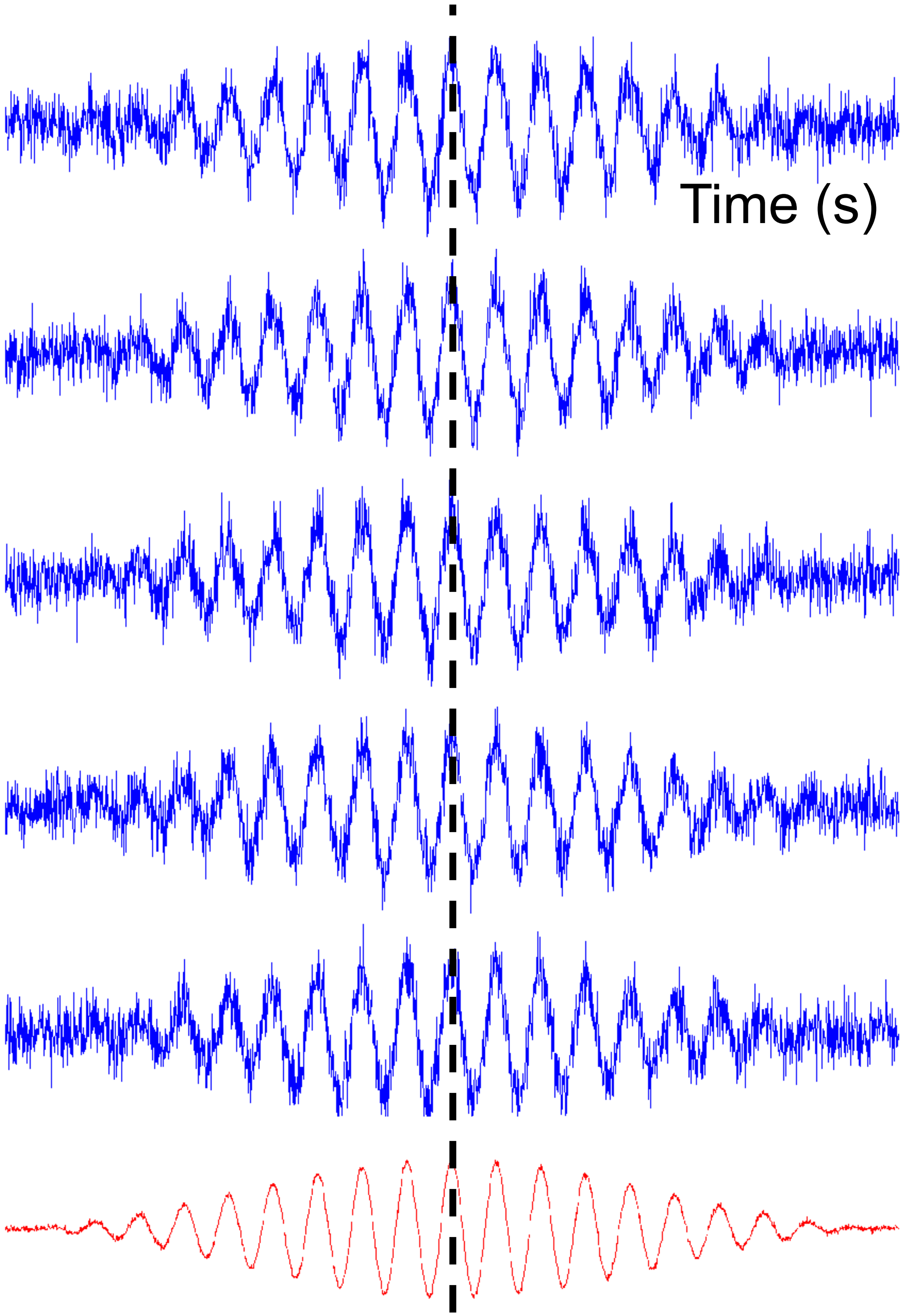


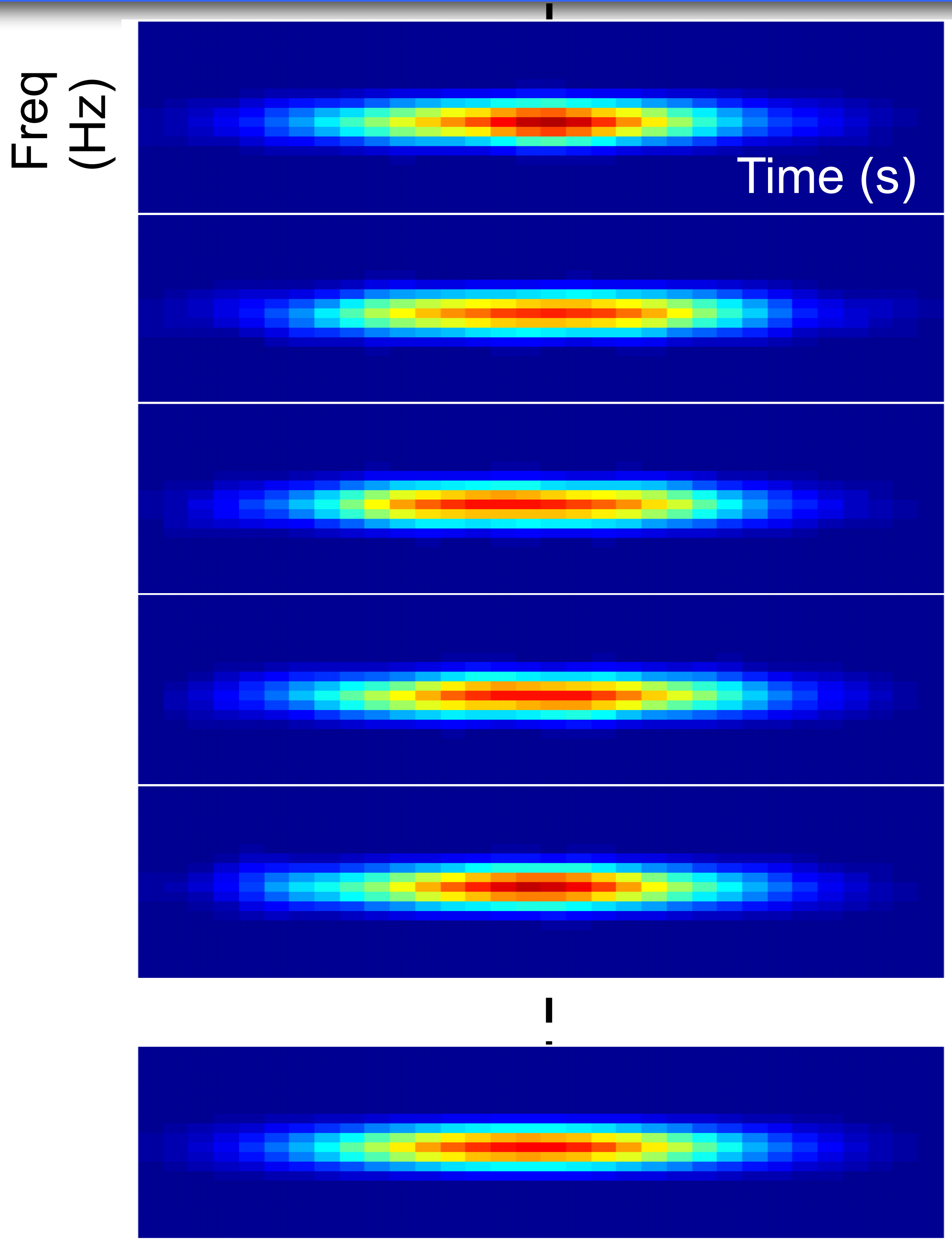
Evoked vs. induced activity



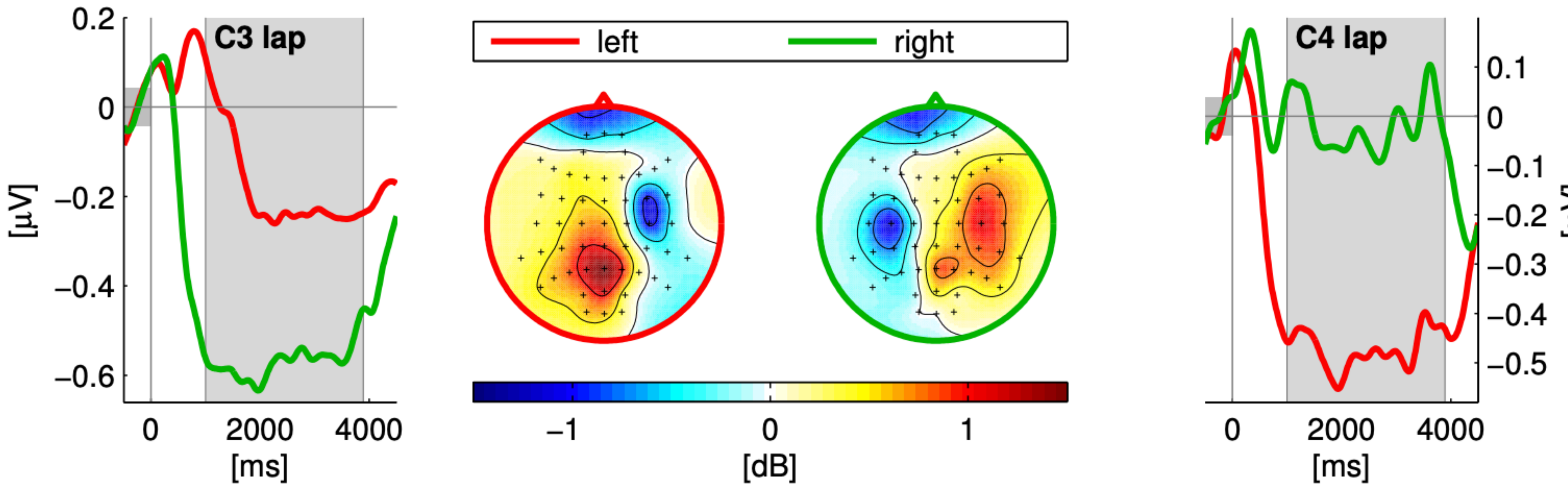


Evoked vs. induced activity





Example of induced: Motor Imagery



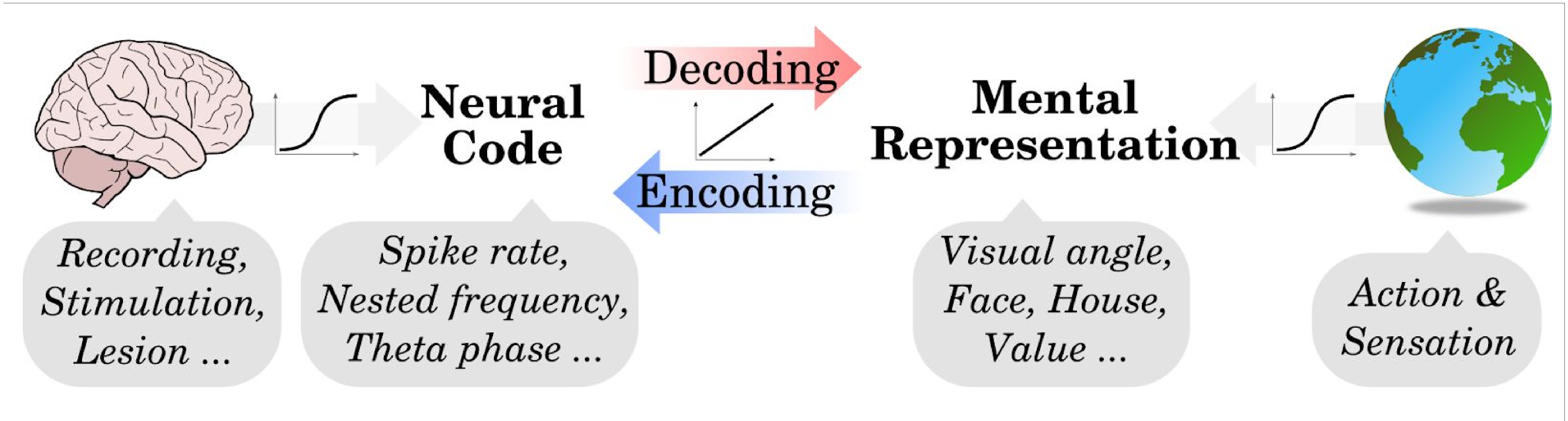
Event-Related Desynchronization (ERD) during motor imagery of the left and the right hand. Averaged signal Hilbert enveloppe between 9 and 13Hz on C3 and C4 (After Laplace / Current Source Density transform)

[Blankertz et al. 2008]

Decoding vs. Encoding

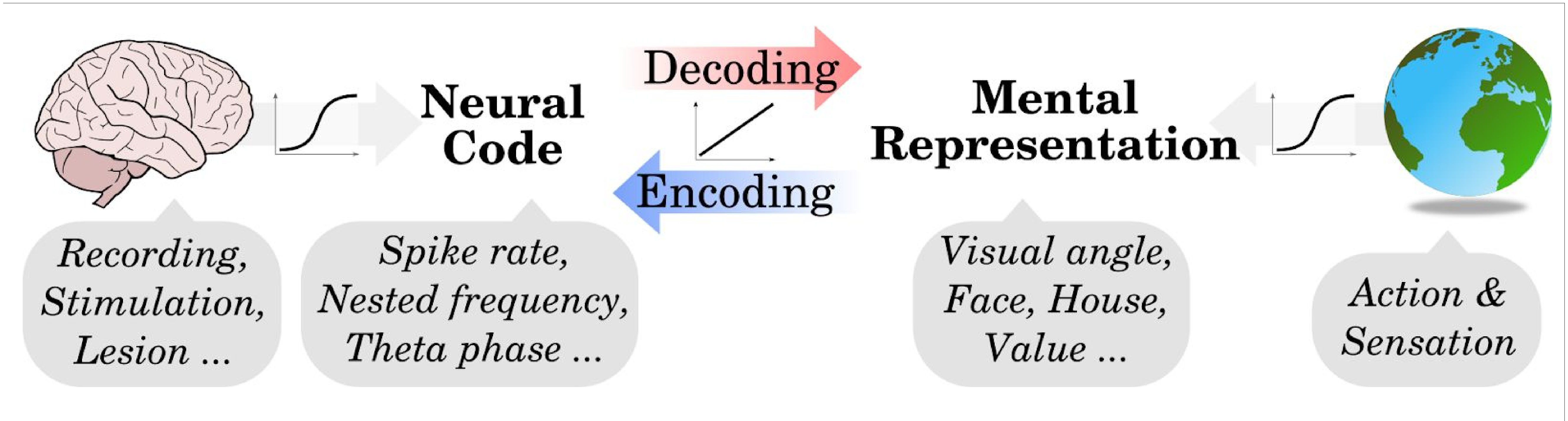
… introducing some maths notations

Neural code & mental representations



[Encoding and Decoding Neuronal Dynamics: Methodological Framework to Uncover the Algorithms of Cognition, King et al. 2019]

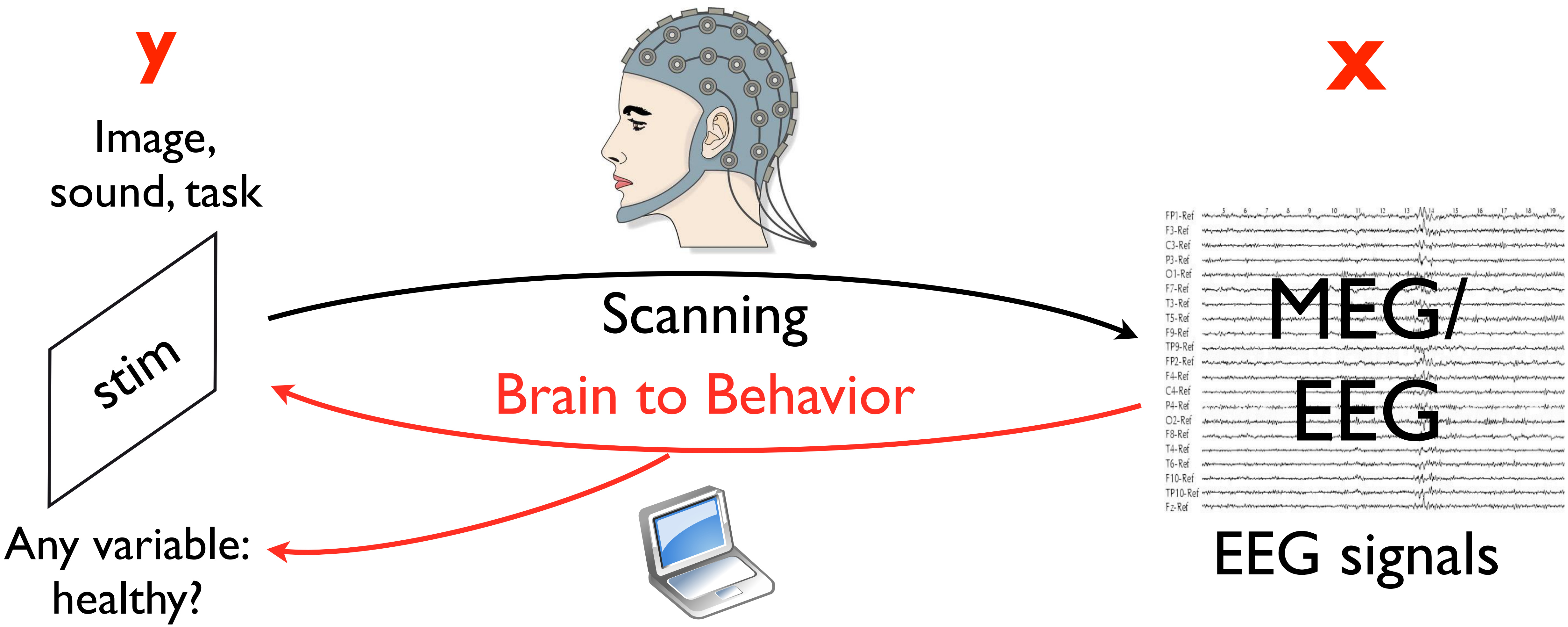
Neural code & mental representations



Decoding predicts experimental factors (e.g. the speed of a hand movement, the luminance of an flashed image etc) from specific features of brain activity (e.g. spike rate, electric field etc), whereas encoding predicts the reverse.

[Encoding and Decoding Neuronal Dynamics: Methodological Framework to Uncover the Algorithms of Cognition, King et al. 2019]

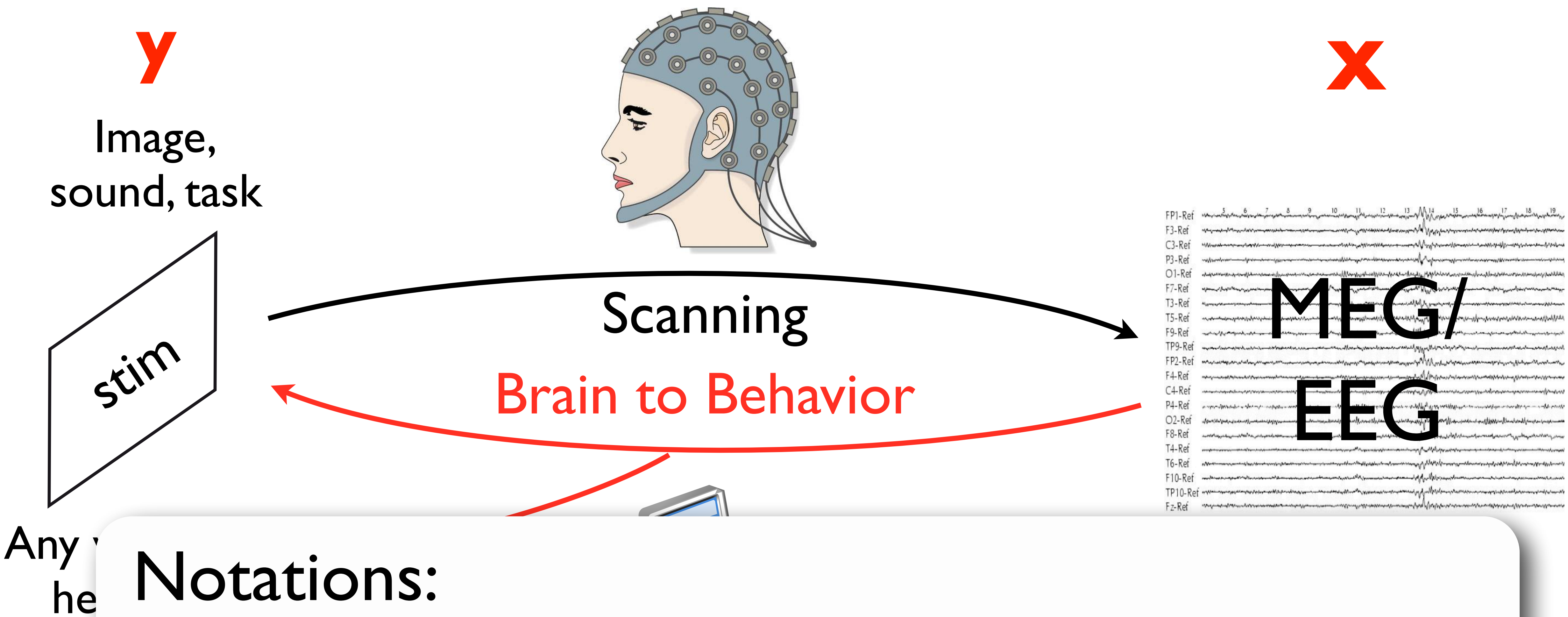
Decoding



Challenge: Predict a behavioral variable from the EEG/MEG data

Objective: Predict y given X or learn a function f : X -> y

Decoding



Challeng

Objectiv

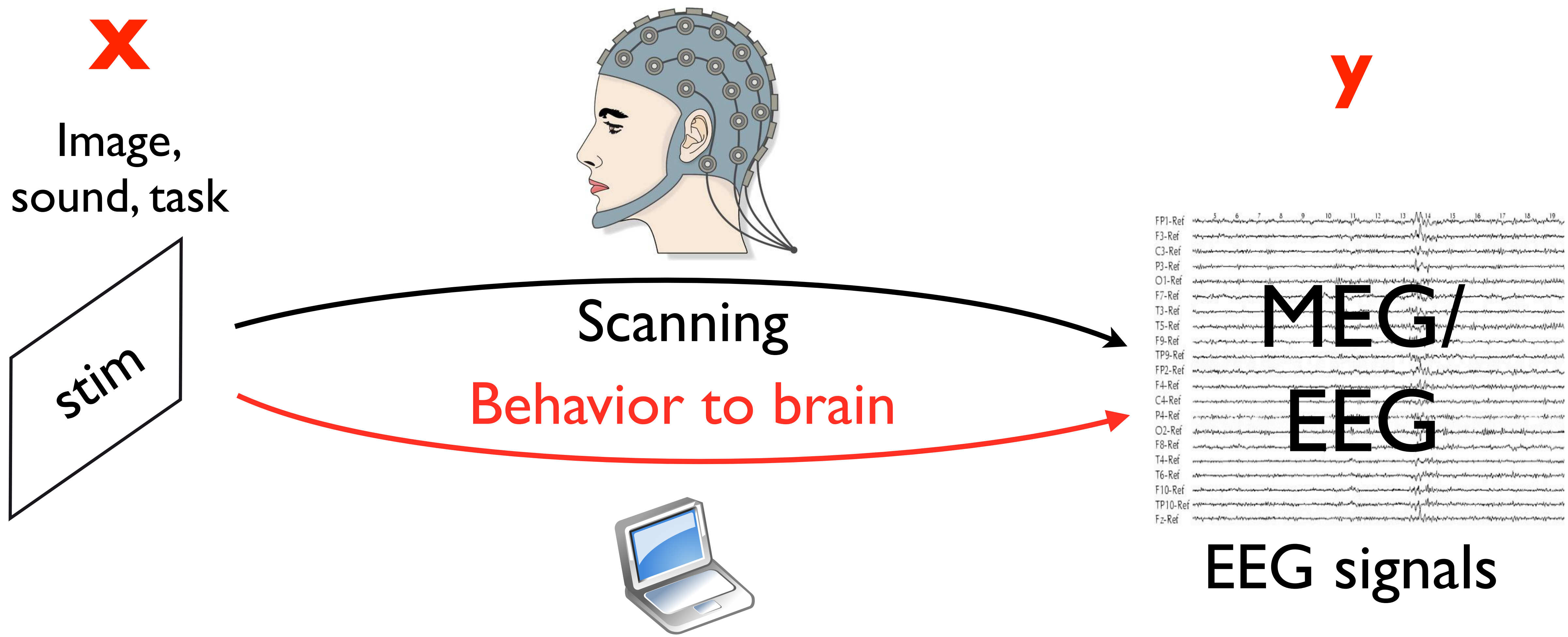
Notations:

$x_i \in \mathbb{R}^P$: vector of features for sample i

$y_i \in \mathbb{R}$: target to predict for sample i

$f : \mathbb{R}^P \rightarrow \mathbb{R}$: prediction function (to learn)

Encoding

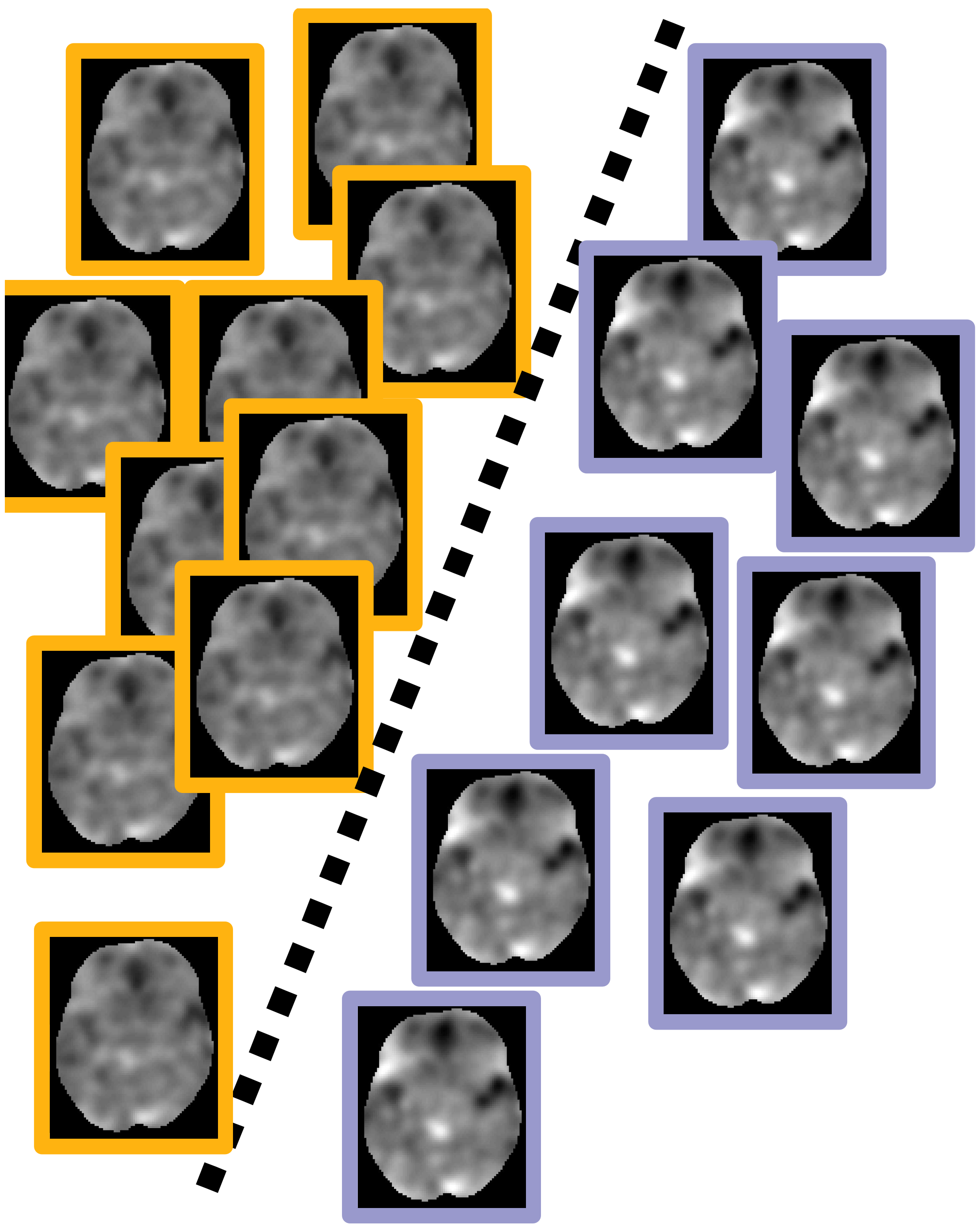


Challenge: Predict a EEG/MEG data from behavioral variable / stimuli

Objective: Predict y given X or learn a function f : X -> y

Algorithms

101 Classification example



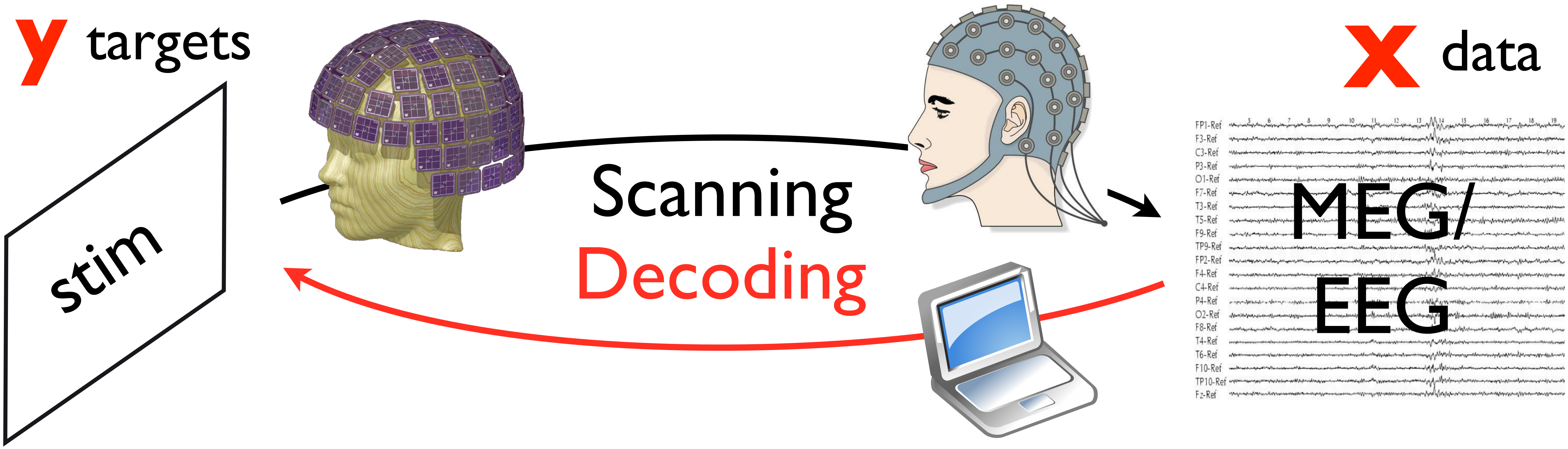
The **objective** is to be able to **predict** or given a brain activation map

Patient	vs.	Controls
Faces	vs.	Houses
...	vs.	...
1	vs.	-1

ie. $y = \{-1, 1\}$

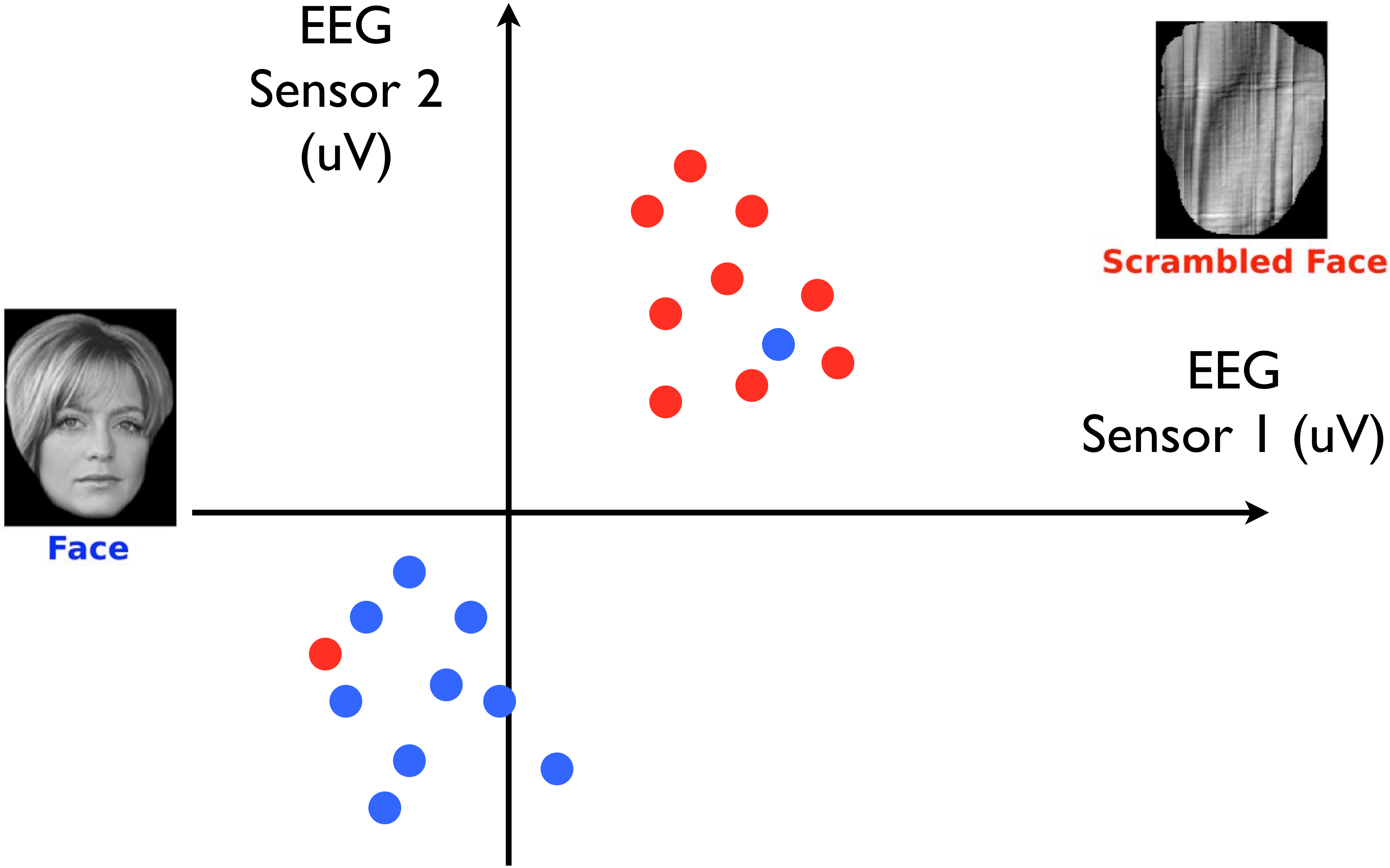
objective: Predict $y = \{-1, 1\}$ given $x \in \mathbb{R}^p$

My first "decoder"

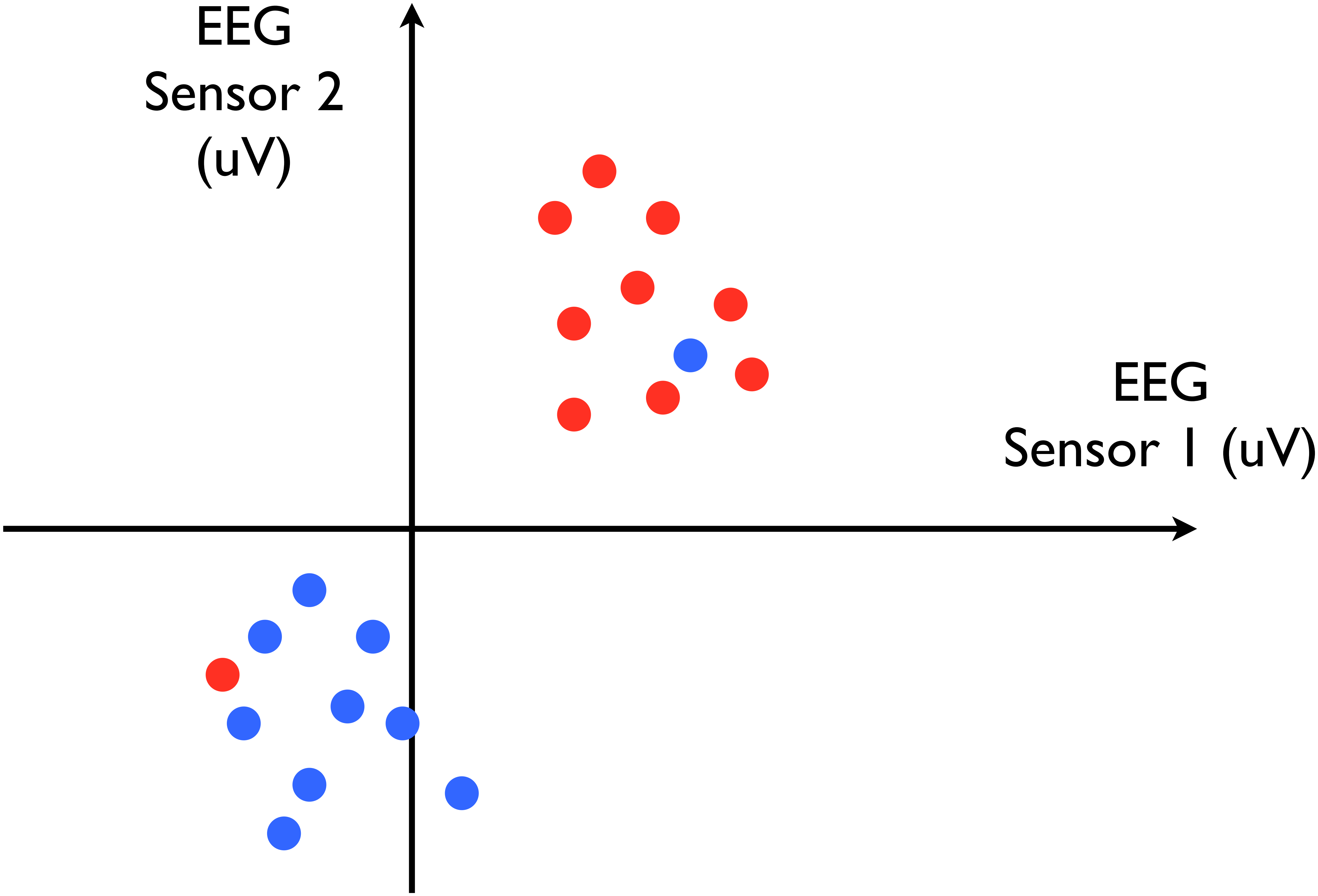


- Visual paradigm with 2 conditions: **faces** or **scrambled faces**
- **Questions:**
  - Can we learn from M/EEG signals to predict on a single trial if the subject saw a face or a scrambled face?
  - When does this happen? How is it encoded in the brain?
  - How to link "neural codes" to "mental representations"?

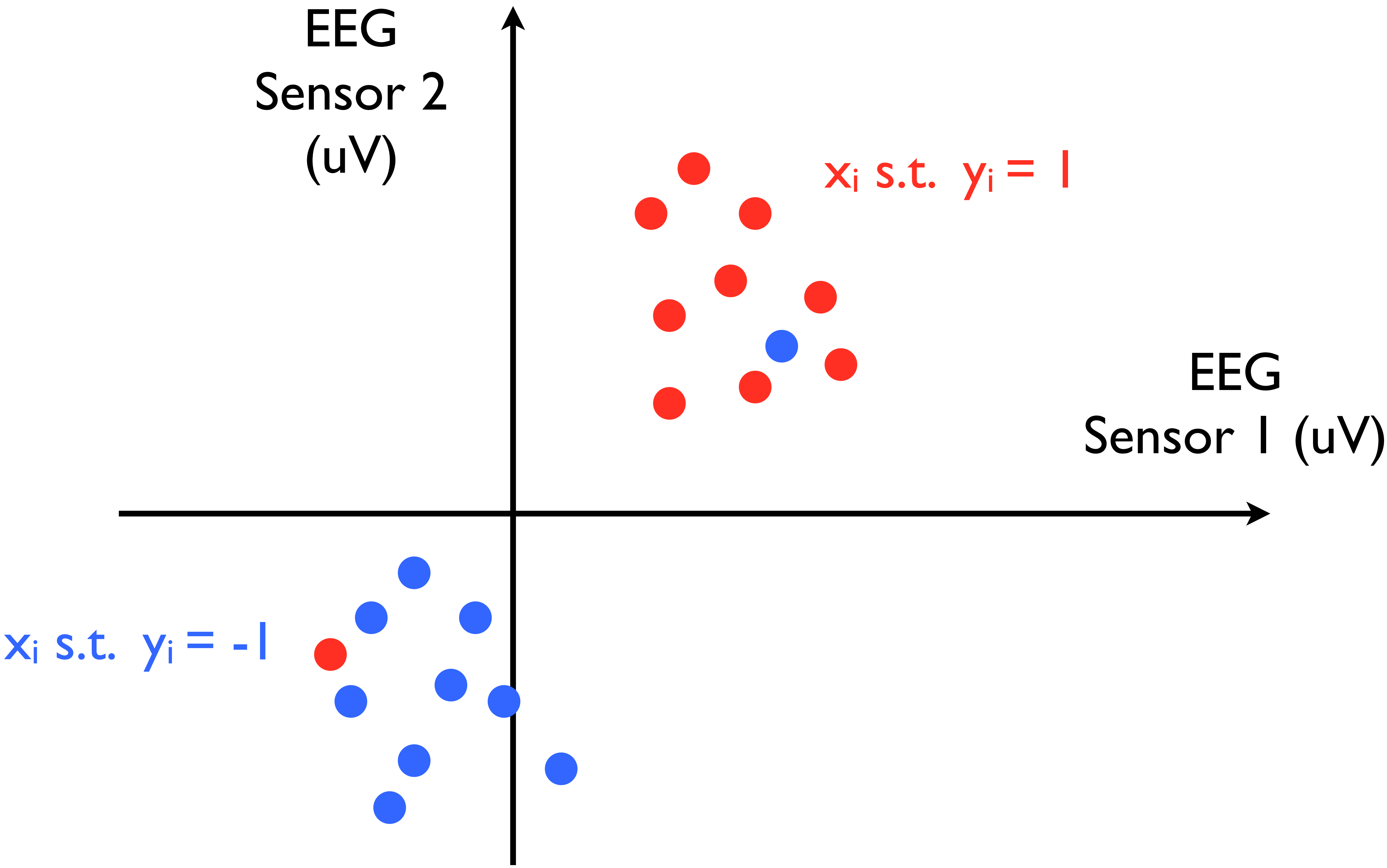
A simple decoder with 2 sensors



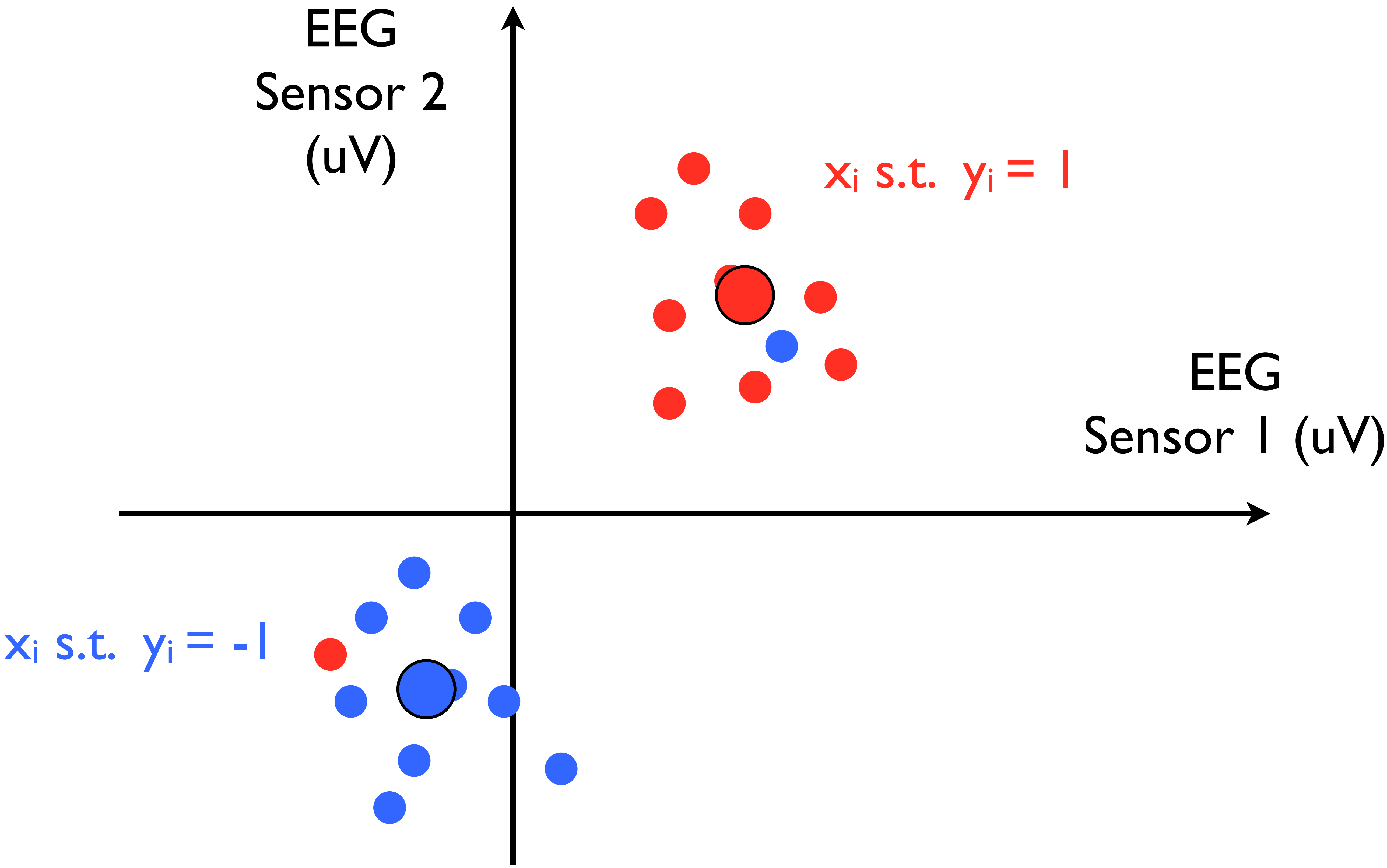
A simple decoder with 2 sensors



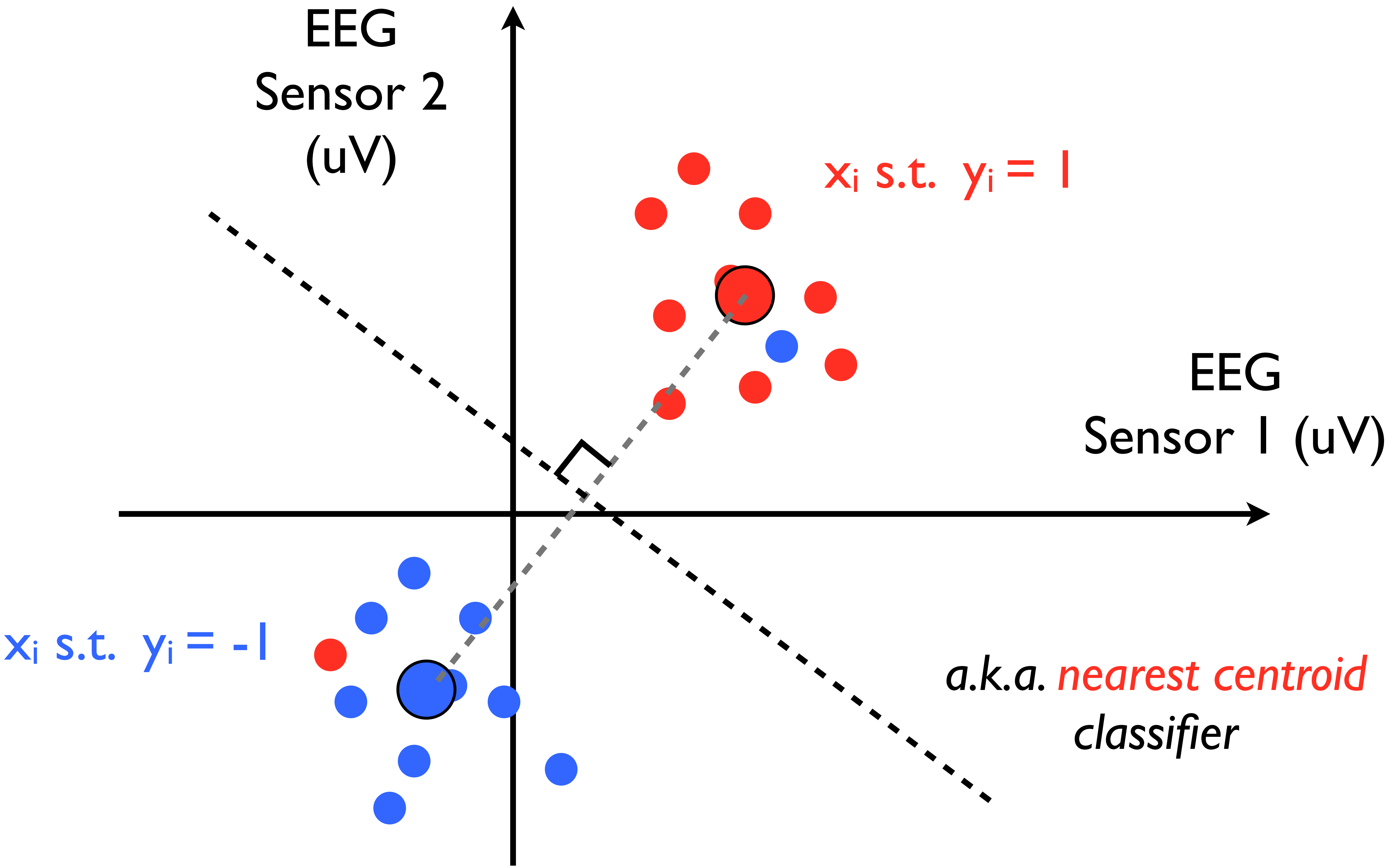
A simple decoder with 2 sensors



A simple decoder with 2 sensors

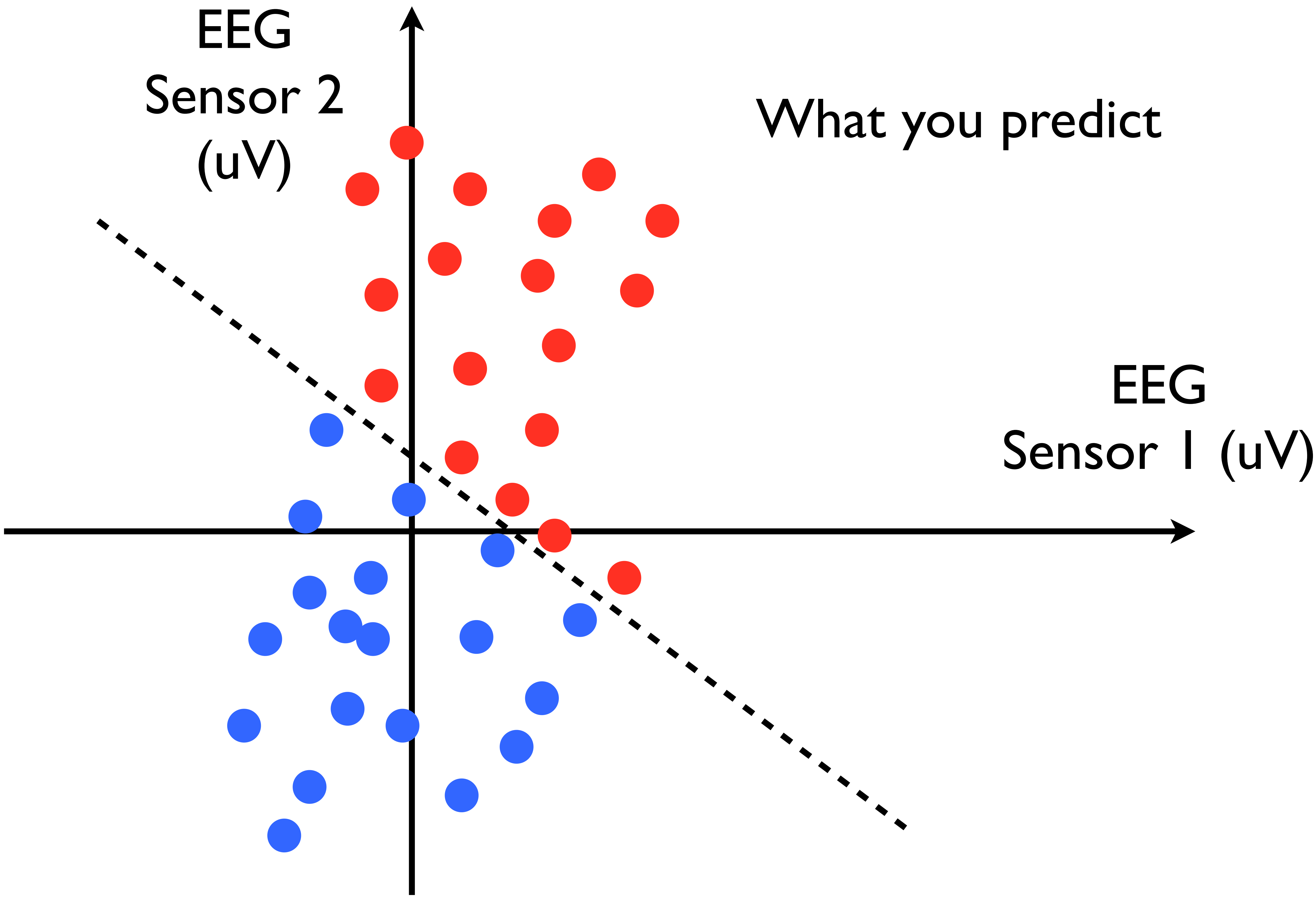


A simple decoder with 2 sensors

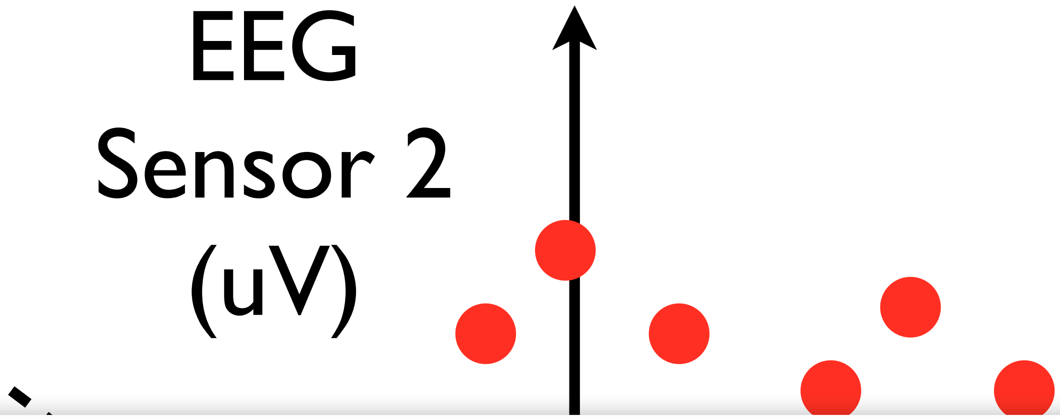


a.k.a. nearest centroid classifier

Testing my first “decoder”



Testing my first “decoder”



What you predict

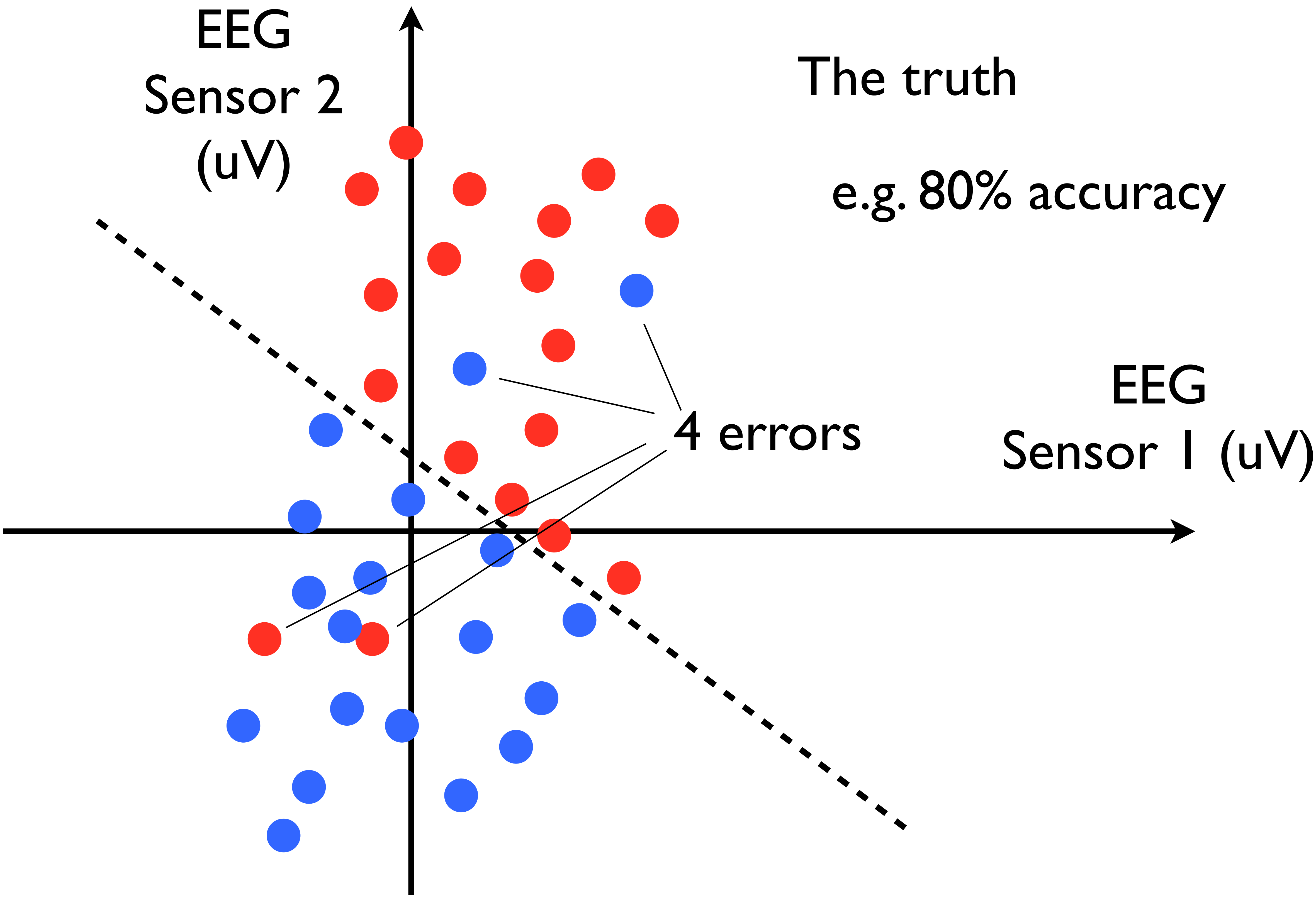
It is a linear model

$$y_i = \mathrm{sign}(\sum_{j=1}^{P} \beta_j x_{ij} + \beta_0) = \mathrm{sign}(\beta^\top x_i + \beta_0)$$

$x_i \in \mathbb{R}^P$: vector of features (here 2)

$\beta \in \mathbb{R}^P, \beta_0 \in \mathbb{R}$: model parameters to estimate

Testing my first "decoder"



Logistic regression

Hypothesis (grounded in the LDA generative model):

$$\log\left(\frac{\mathrm{Prob}(y=1|x)}{\mathrm{Prob}(y=-1|x)}\right) = \beta^\top x + \beta_0$$

Logistic regression

Hypothesis (grounded in the LDA generative model):

$$\log \left(\frac{\mathrm{Prob}(y = 1|x)}{\mathrm{Prob}(y = -1|x)} \right) = \beta^\top x + \beta_0$$

Inference my maximum likelihood:

$$\hat{\beta}, \hat{\beta}_0 = \arg\min_{\beta, \beta_0} \frac{1}{N} \sum_{i=1}^{N} \log(1 + \exp(-y_i(\beta^\top x_i + \beta_0))$$

Logistic regression

Hypothesis (grounded in the LDA generative model):

$$\log\left(\frac{\mathrm{Prob}(y=1|x)}{\mathrm{Prob}(y=-1|x)}\right) = \beta^\top x + \beta_0$$

Inference my maximum likelihood:

$$\hat{\beta}, \hat{\beta}_0 = \arg\min_{\beta,\beta_0} \frac{1}{N} \sum_{i=1}^{N} \log(1 + \exp(-y_i(\beta^\top x_i + \beta_0))$$

$\boxed{+\lambda\Omega(\beta) \quad \text{Regularization}}$

Logistic regression

Hypothesis (grounded in the LDA generative model):

$$\log\left(\frac{\mathrm{Prob}(y=1|x)}{\mathrm{Prob}(y=-1|x)}\right) = \beta^\top x + \beta_0$$

Inference my maximum likelihood:

$$\hat{\beta}, \hat{\beta}_0 = \arg\min_{\beta,\beta_0} \frac{1}{N}\sum_{i=1}^{N} \log(1+\exp(-y_i(\beta^\top x_i + \beta_0))$$

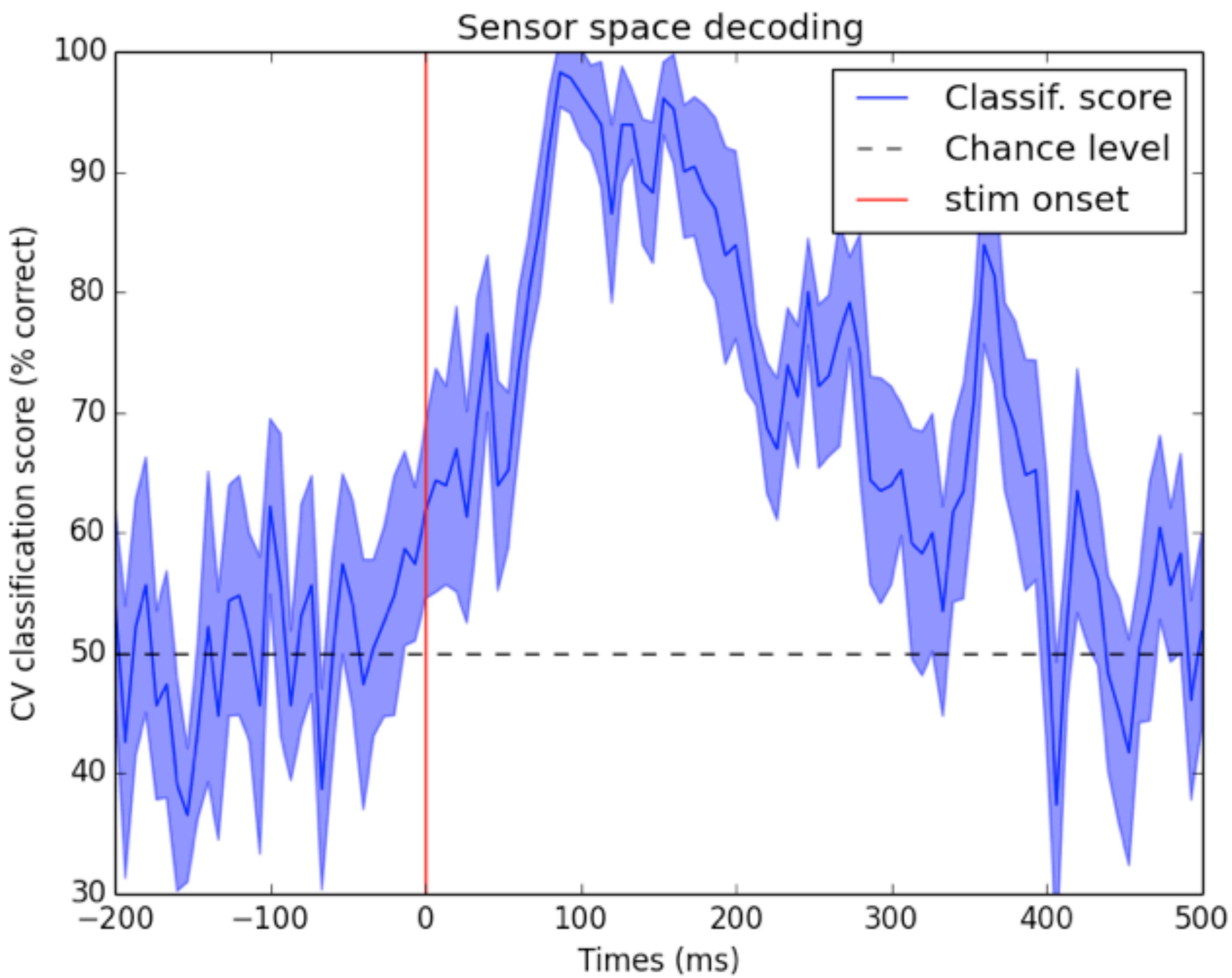
$$\boxed{+\lambda\Omega(\beta) \quad \text{Regularization}}$$

Example: $\Omega(\beta) = \frac{1}{2}\|\beta\|_2^2$ Hyperparameter: λ

Decoding evoked MEG/EEG activity

Decoding over time with MEG/EEG

- **IDEA:** Run a cross-validation loop at each time point



Condition 1 (auditory tone in left ear) vs. condition 2 (visual flash in left hemifield)

- 50 epochs per class
- no baseline
- band pass between 2Hz and 40Hz
- Linear SVM (C=1)
- Monte-carlo cross-val.

https://mne.tools/stable/auto_tutorials/machine-learning/plot_sensors_decoding.html

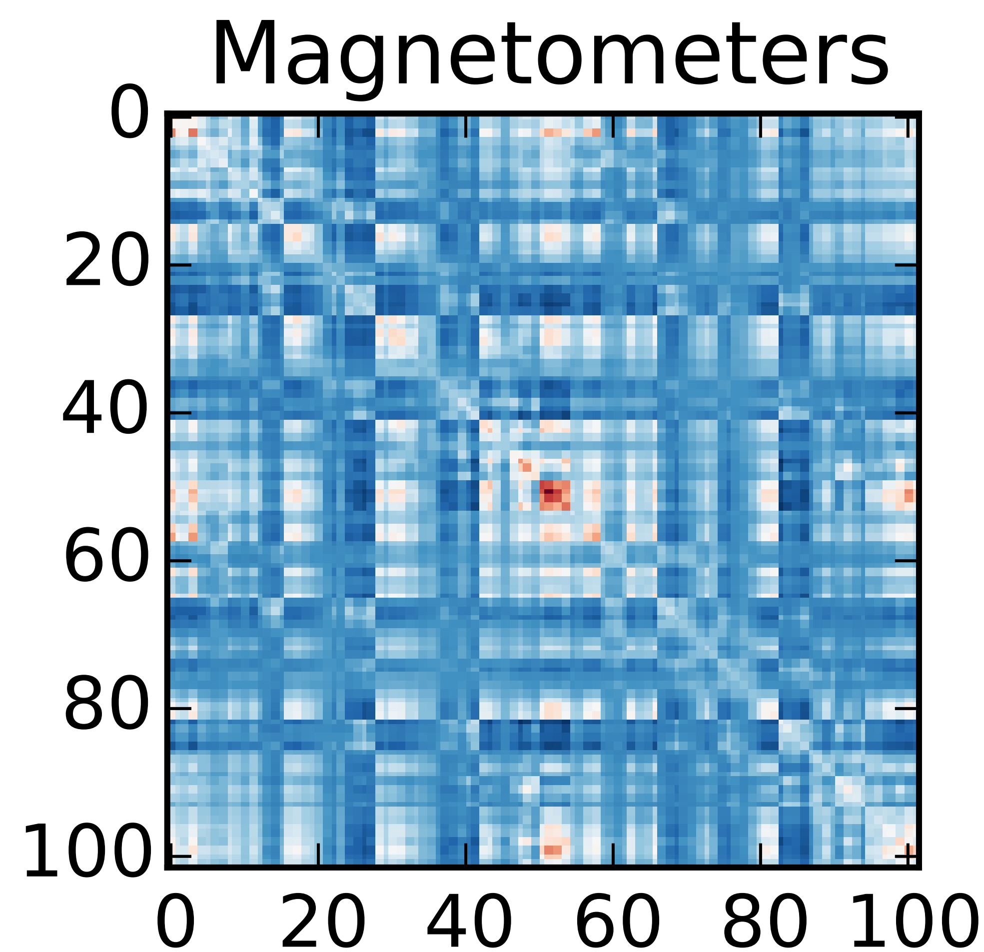
Decoding induced MEG/EEG activity

Covariance

$S_i \in \mathbb{R}^{C \times T}$: MEG/EEG signals

C : number of channels

T : number of time samples



Definition [Covariance]:

$$\Sigma_i = \frac{S_i S_i^\top}{T} \in \mathbb{S}_+^C \subset \mathbb{R}^{C \times C}$$

Set of non-negative symmetric matrices

Remark: Estimation of cov. requires care

[Engemann & Gramfort, Neuroimage 2017]

Common Spatial Pattern (CSP)

The model:

$w_j \in \mathbb{R}^C$: Spatial filter

$x_{ij} = \log(w_j^\top \Sigma_i w_j)$: Feature construction

$y_i = \text{sign}(\sum_{j=1}^{P} \beta_j \log(w_j^\top \Sigma_i w_j) + \beta_0)$: Linear model

`[Koles 1991, Blankertz et al. 2008 etc.]`

Common Spatial Pattern (CSP)

The model:

$$w_j \in \mathbb{R}^C$$: Spatial filter

$$x_{ij} = \log(w_j^\top \Sigma_i w_j)$$: Feature construction

$$y_i = \text{sign}(\sum_{j=1}^{P} \beta_j \log(w_j^\top \Sigma_i w_j) + \beta_0)$$: Linear model

How do you estimate filters?

[Koles 1991, Blankertz et al. 2008 etc.]

Common Spatial Pattern (CSP)

$$\Sigma^c = \frac{1}{N_c} \sum_{i/y_i=c} \Sigma_i \quad c \in \{-1, 1\}$$ Estimate per class covariance

[Koles 1991, Blankertz et al. 2008 etc.]

https://mne.tools/dev/auto_examples/decoding/plot_decoding_csp_eeg.html

Common Spatial Pattern (CSP)

$$\Sigma^c = \frac{1}{N_c} \sum_{i/y_i=c} \Sigma_i \quad c \in \{-1, 1\}$$ Estimate per class covariance

Solve generalized eigenvalue problem:

$$w \in \mathbb{R}^C, \mu \in \mathbb{R} \quad \text{s.t.} \quad \Sigma^1 w = \mu \left(\sum_{c \in \{-1,1\}} \Sigma^c \right) w$$

[Koles 1991, Blankertz et al. 2008 etc.]

https://mne.tools/dev/auto_examples/decoding/plot_decoding_csp_eeg.html

Common Spatial Pattern (CSP)

$$\Sigma^c = \frac{1}{N_c} \sum_{i/y_i=c} \Sigma_i \quad c \in \{-1, 1\}$$ Estimate per class covariance

Solve generalized eigenvalue problem:

$$w \in \mathbb{R}^C, \mu \in \mathbb{R} \quad \text{s.t.} \quad \Sigma^1 w = \mu \left(\sum_{c \in \{-1,1\}} \Sigma^c \right) w$$

Properties:

$$W^\top \Sigma^c W = \Lambda^c \quad \Lambda^c = \mathrm{diag}(\lambda_1^c, \ldots, \lambda_C^c) \quad \sum_{c \in \{1,-1\}} \Lambda^c = I$$

[Koles 1991, Blankertz et al. 2008 etc.]

https://mne.tools/dev/auto_examples/decoding/plot_decoding_csp_eeg.html

Common Spatial Pattern (CSP)

$$\Sigma^c = \frac{1}{N_c} \sum_{i/y_i=c} \Sigma_i \quad c \in \{-1, 1\}$$ Estimate per class covariance

Solve generalized eigenvalue problem:

$$w \in \mathbb{R}^C, \mu \in \mathbb{R} \quad \text{s.t.} \quad \Sigma^1 w = \mu \left(\sum_{c \in \{-1,1\}} \Sigma^c \right) w$$

Properties:

$$W^\top \Sigma^c W = \Lambda^c \qquad \Lambda^c = \mathrm{diag}(\lambda_1^c, \dots, \lambda_C^c) \quad \sum_{c \in \{1,-1\}} \Lambda^c = I$$

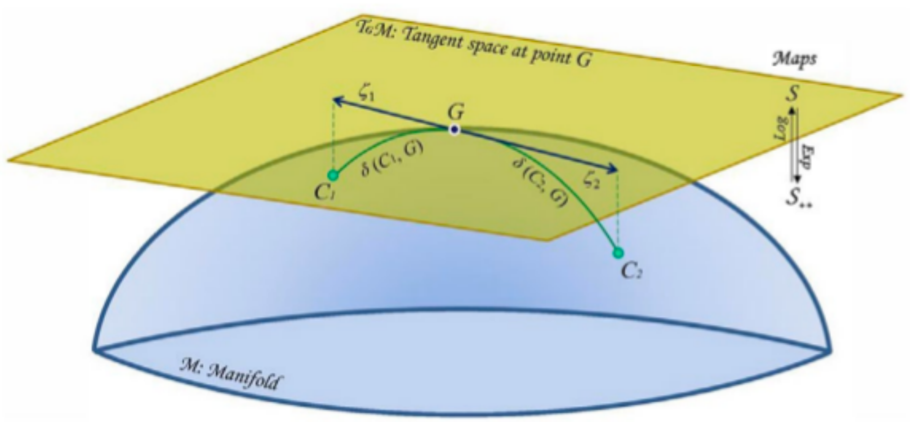
Since $\lambda_j^1 + \lambda_j^{-1} = 1$, a high value for λ_j^1 means that the filter output based on filter w_j yields a high variance for input signals in class 1 and a low variance for signals in class -1 (and vice versa).

[Koles 1991, Blankertz et al. 2008 etc.]

https://mne.tools/dev/auto_examples/decoding/plot_decoding_csp_eeg.html

Riemann

- Summarise a trial by a covariance
- Classification where samples are covariances
- Covariances live a curved manifold $\mathbb{S}_+^C \neq \mathbb{R}^{2C}$



Def. [Geometric distance]:

$$d_G(\Sigma_1, \Sigma_2) = \| \log(\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}}) \|_F$$

$$= \left[\sum_{i=1}^{C} \log^2 \lambda_k \right]^{\frac{1}{2}}$$

[Barachant et al. 2013, Bhatia 2007, Absil et al. 2009]

Nearest Centroid with Riemann

The algorithm:

- Compute centroid for each class

$$\bar{\Sigma}^c = \arg\min_{\Sigma \in \mathbb{S}^C_{++}} \sum_{i / y_i = c} d_G(\Sigma_i, \Sigma)^2$$

"Fréchet Mean"

- Prediction rule:

$$f(\Sigma_i) = \arg\min_{c \in \{1,-1\}} d_G(\Sigma_i, \bar{\Sigma}^c)$$

Remark: Called "Minimum Distance to Mean" (MDM) by Barachant

[Barachant et al. LVA ICA 2010, Barachant et al. 2013]

http://pyriemann.readthedocs.io/en/latest/

Cross-validation done right !

2 setups in neuroimaging

- **Intra-subject decoding:** learn and predict on the same subject

- **Inter-subject decoding:** learn and predict on different subjects

In both cases left-out samples are independent

Training vs. Test vs. Validation error

Full dataset

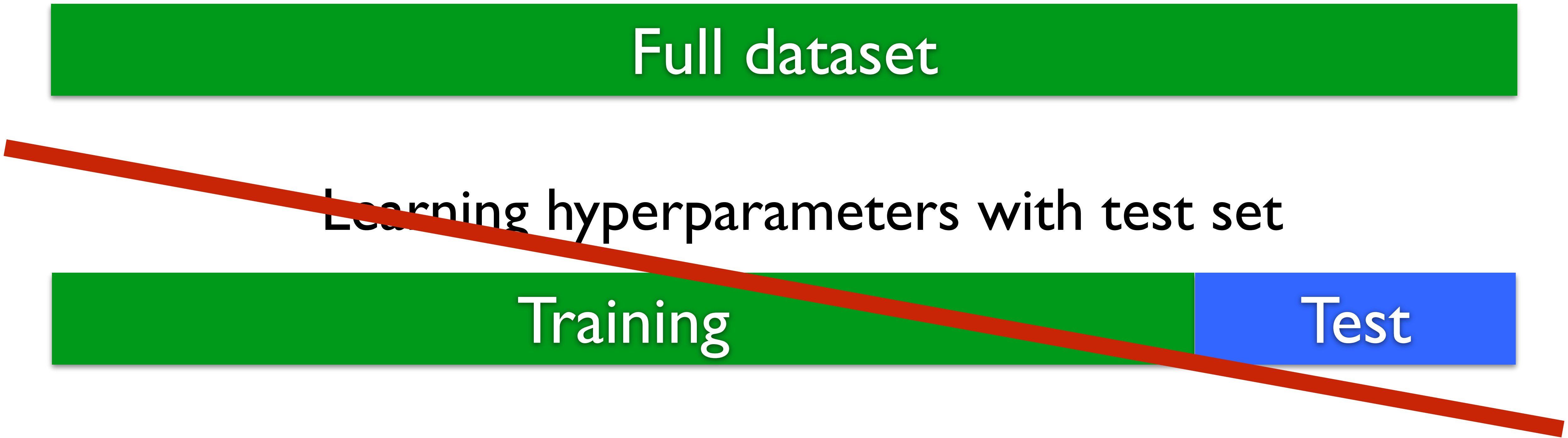
Training vs. Test vs. Validation error

Full dataset

Learning hyperparameters with test set

Training | Test

Training vs. Test vs. Validation error



Training vs. Test vs. Validation error

Full dataset

~~Learning hyperparameters with test set~~

~~Training | Test~~

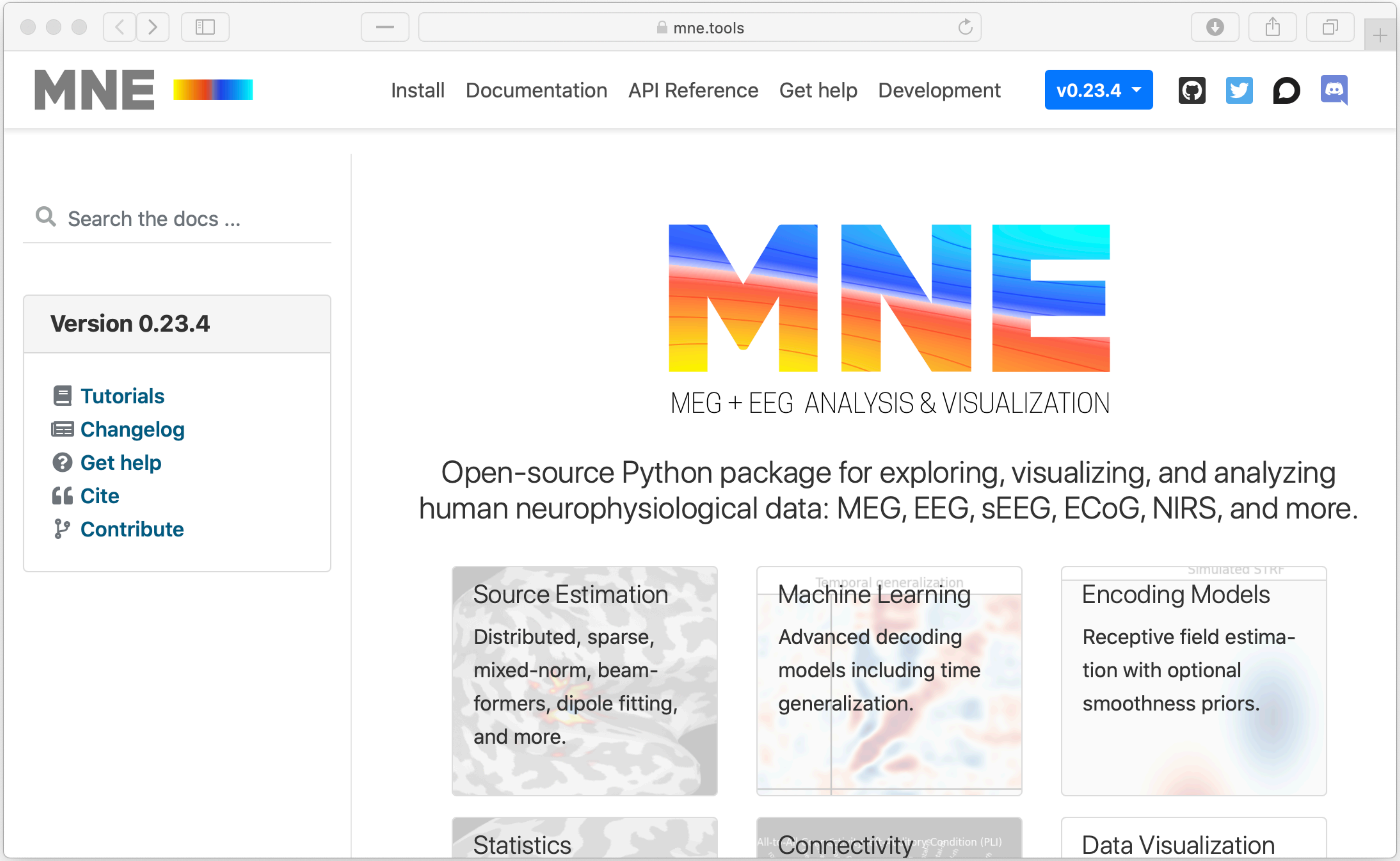
Learning hyperparameters with validation set and use test for test !

Training | Validation | Test

To reliably assess the performance of your model don’t optimize hyperparameters on test set

MNE quick info

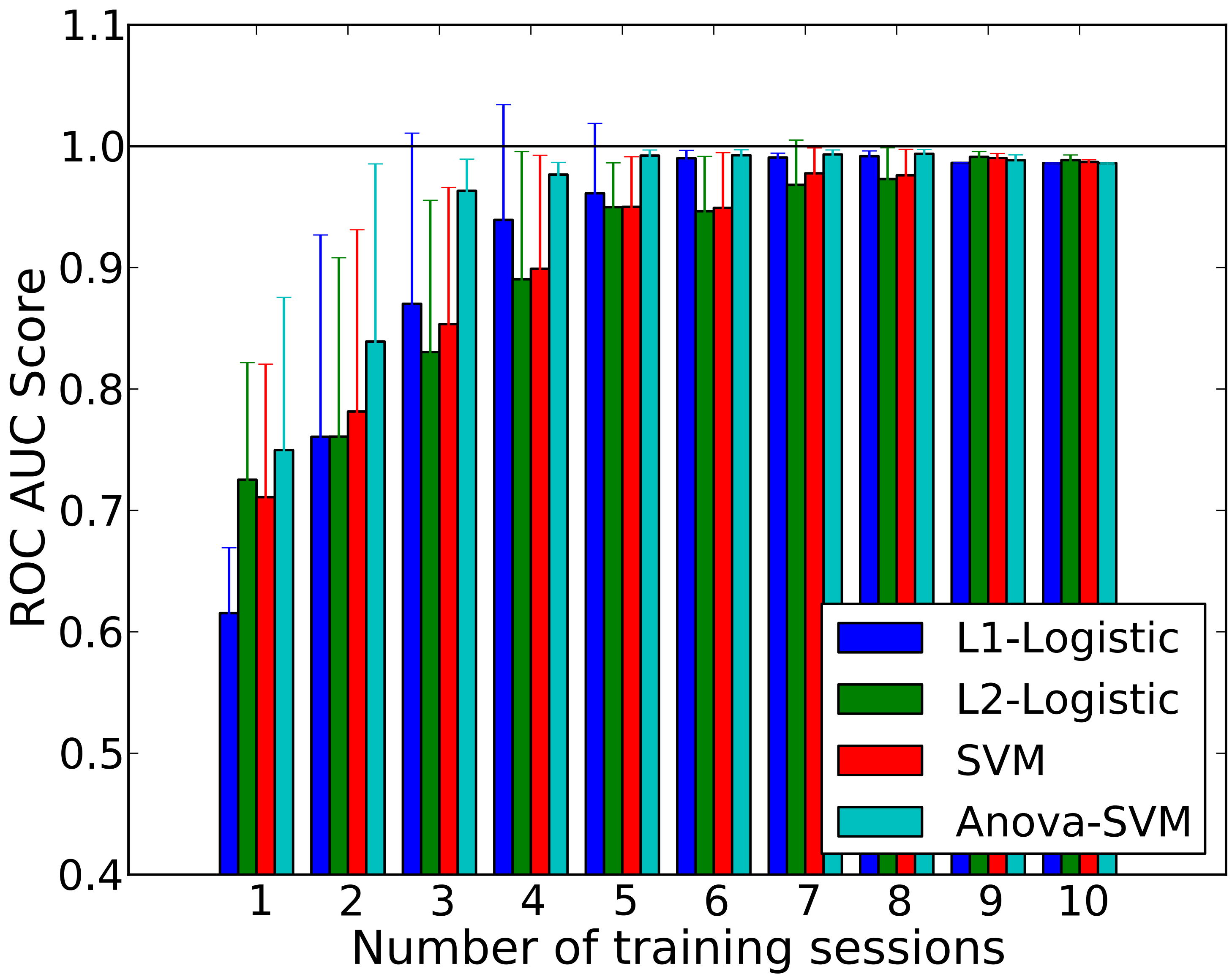
http://mne.tools



MNE software for processing MEG and EEG data, A. Gramfort, M. Luessi, E. Larson, D. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, M. Hämäläinen, Neuroimage 2013

Conclusion

Why more data is better?

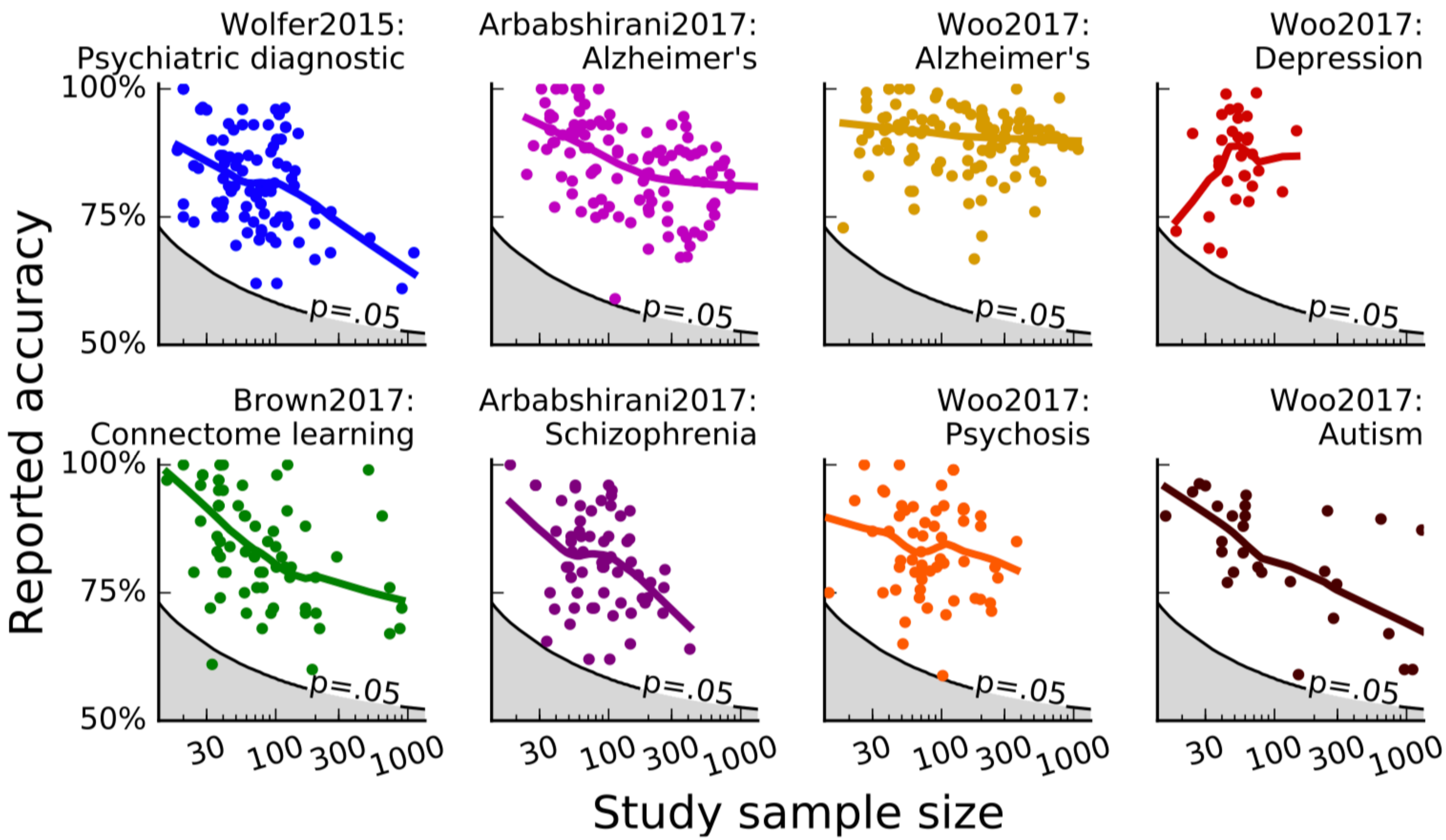


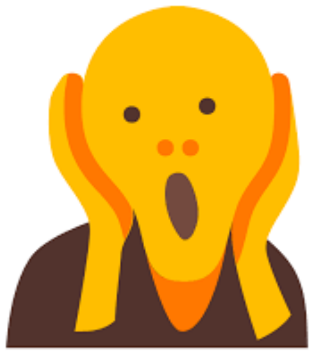
- 5 subjects
- 12 sessions (more than 1000 scans)
- Binary classification
- Test of 2 left-out sessions

- The more data the better
- Almost no noise

Data from *[Haxby et al. 2001]*
Figure from *[Gramfort et al. 2011]*

Cross-validation failure !!!





Prediction accuracy should increase with sample size and not the opposite…

[Varoquaux 2017, Neuroimage]

Now hands on!

To install the bare minimum for this tutorial:

```
pip install mne
```

To install everything using conda (or mamba):

```
conda install mamba -n base -c conda-forge
mamba install mne
```

For more complete install see:

https://mne.tools/stable/install/index.html

Thanks !

Contact http://alexandre.gramfort.net

GitHub : @agramfort *Twitter : @agramfort*

Support ERC SLAB, ANR THALAMEEG ANR-14-NEUC-0002-01
NIH R01 MH106174, ANR Chaire IA BrAIN

"An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem. ~ John Tukey"