Machine Learning for Multimodal Neuroimaging

Felix Bießmann

Max-Planck-Institute for Biological Cybernetics

Acknowledgements

- Experiments

Nikos K. Logothetis
MPI Tübingen

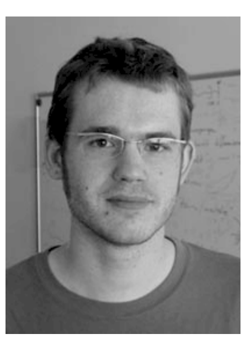
Yusuke Murayama
MPI Tübingen

Lei Wu
University of
New Mexico

Tom Eichele
University of Bergen

Michael Gaebler
Charite Berlin

- Data Analysis



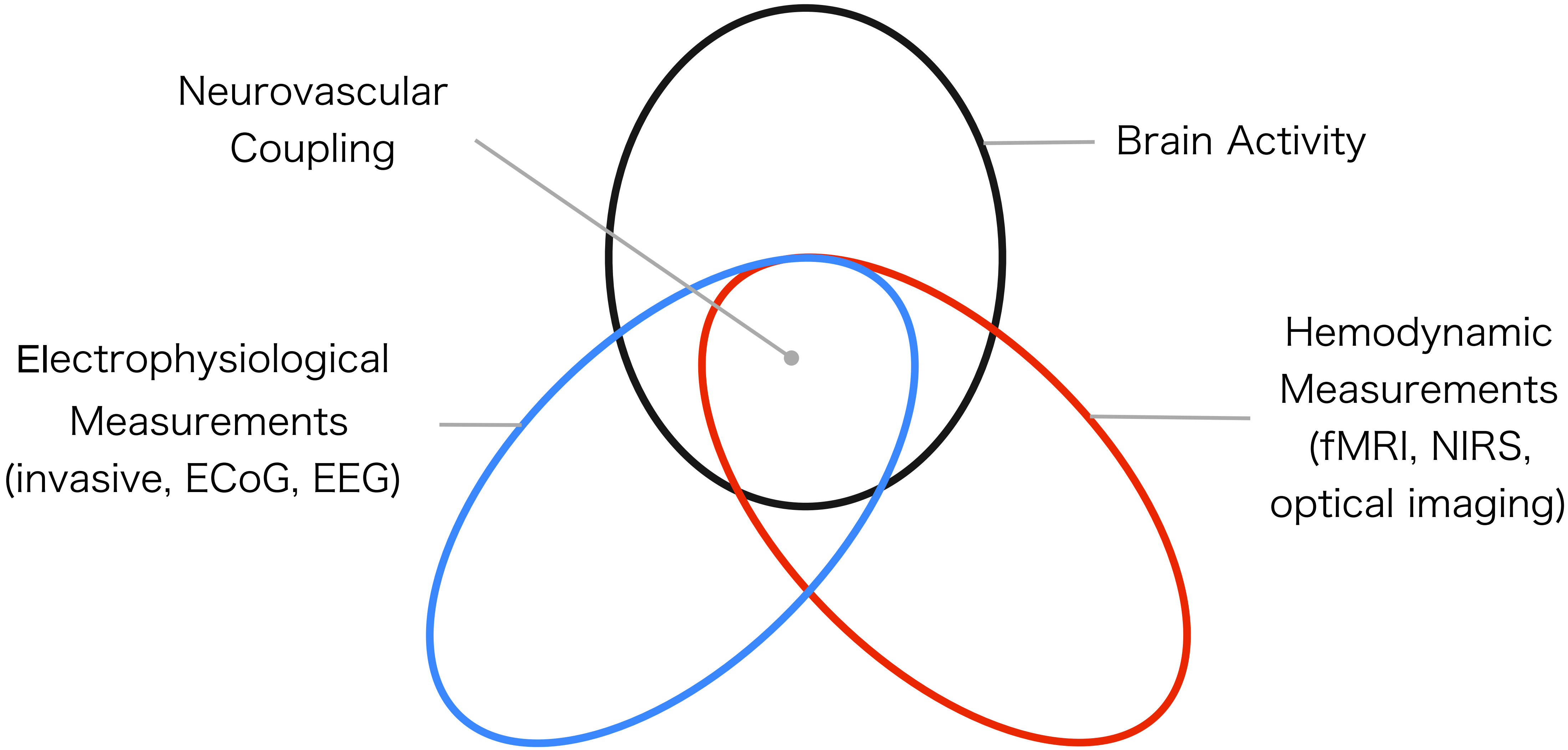
Frank C. Meinecke
TU Berlin

Arthur Gretton
Gatsby Unit, UCL

Stefan Haufe
TU Berlin

Klaus-Robert Müller
TU Berlin

Multimodal Neuroimaging



Motivation for Multimodal Neuroimaging

1. Neural activation is reflected in
   - Electromagnetic field changes
   - Hemodynamic Activity

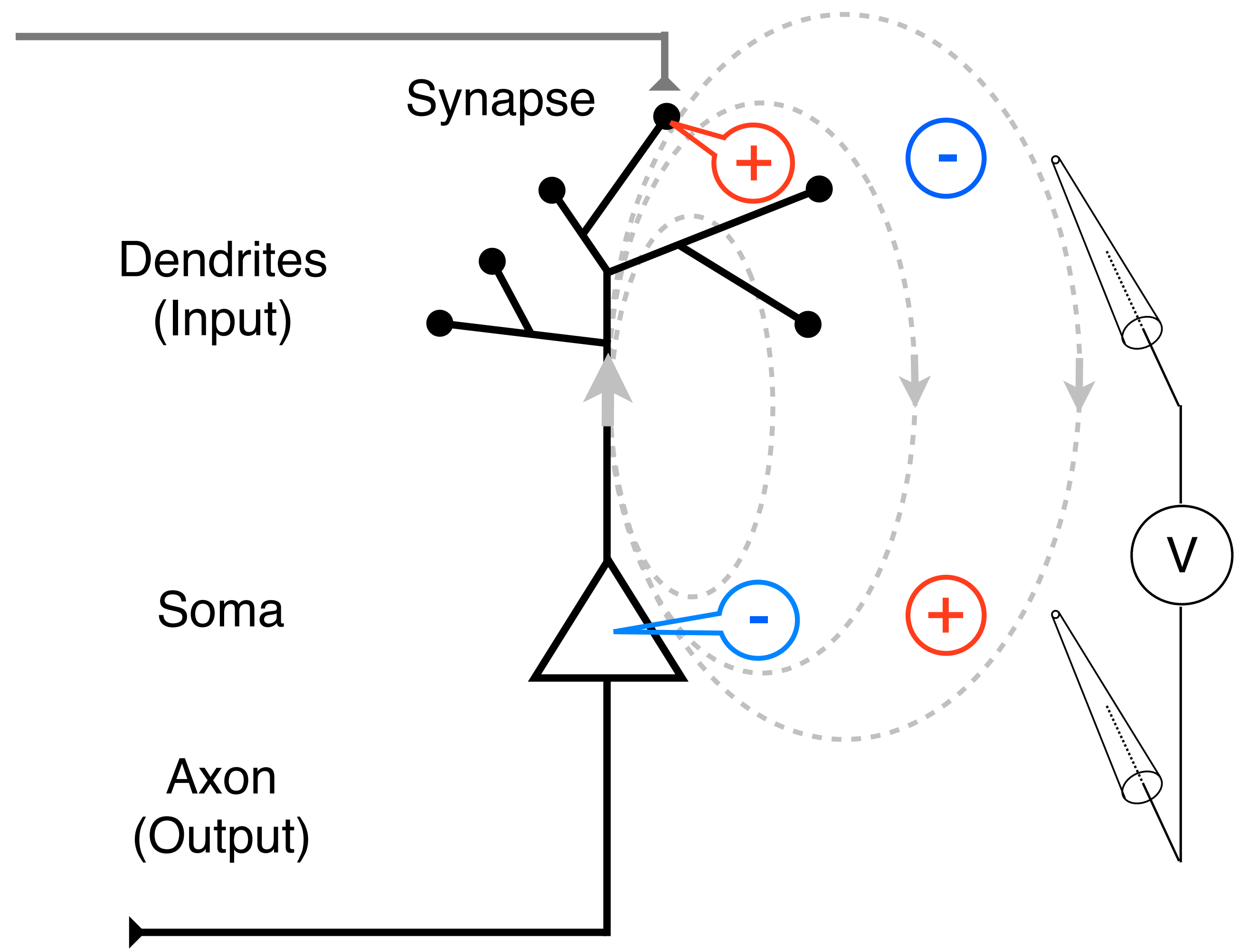
2. Modalities are complementary

3. Multimodal is better than single modality

4. Understanding the coupling is essential

Electromagnetic Field Changes

Neural activation is reflected in electromagnetic field changes



Electromagnetic Field Changes

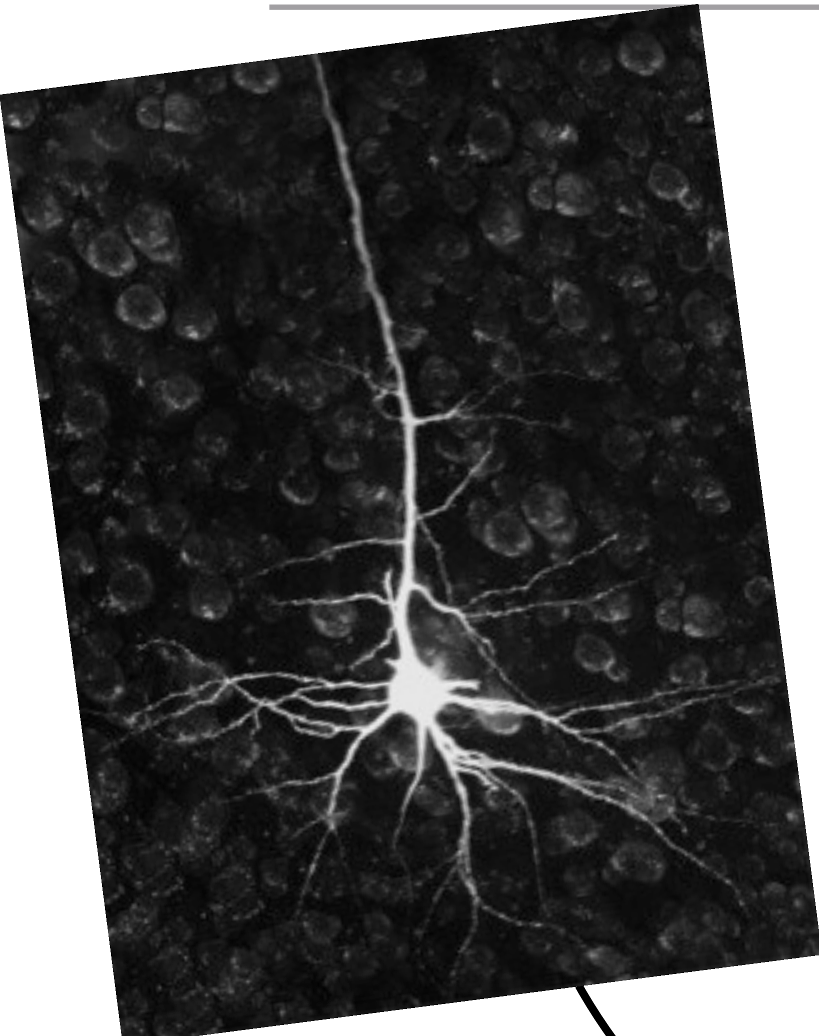
Neural activation is reflected in electromagnetic field changes

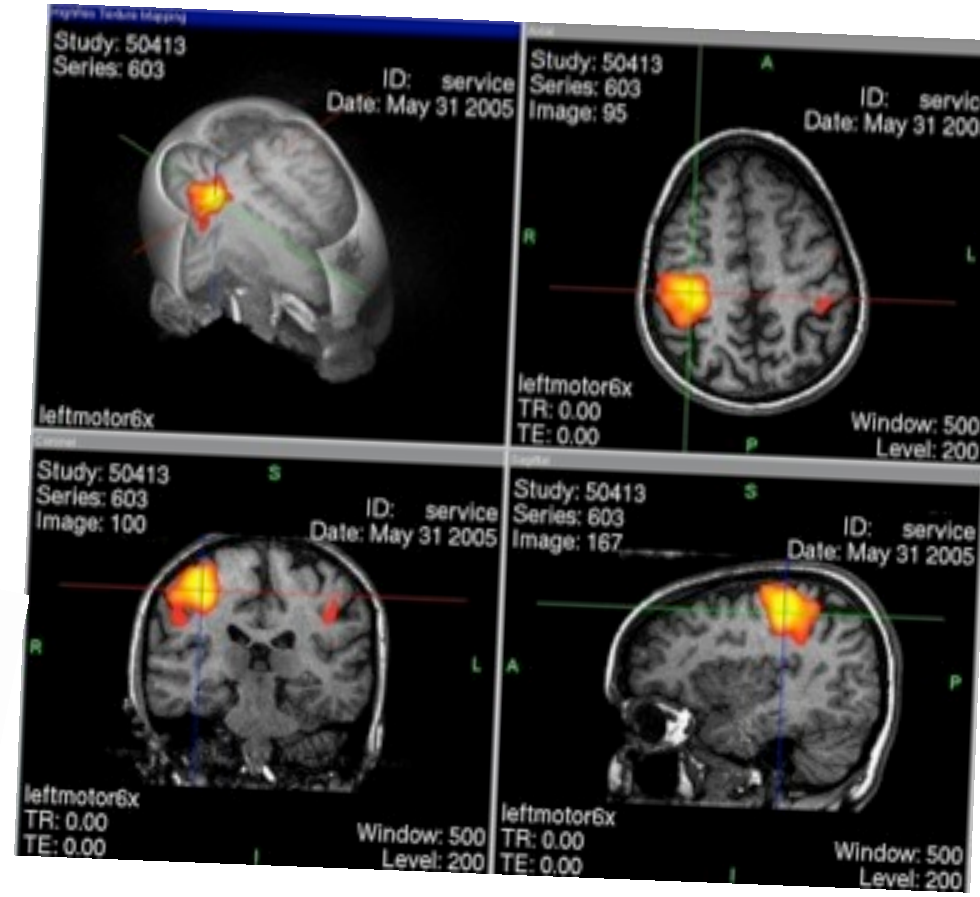
Measured by:

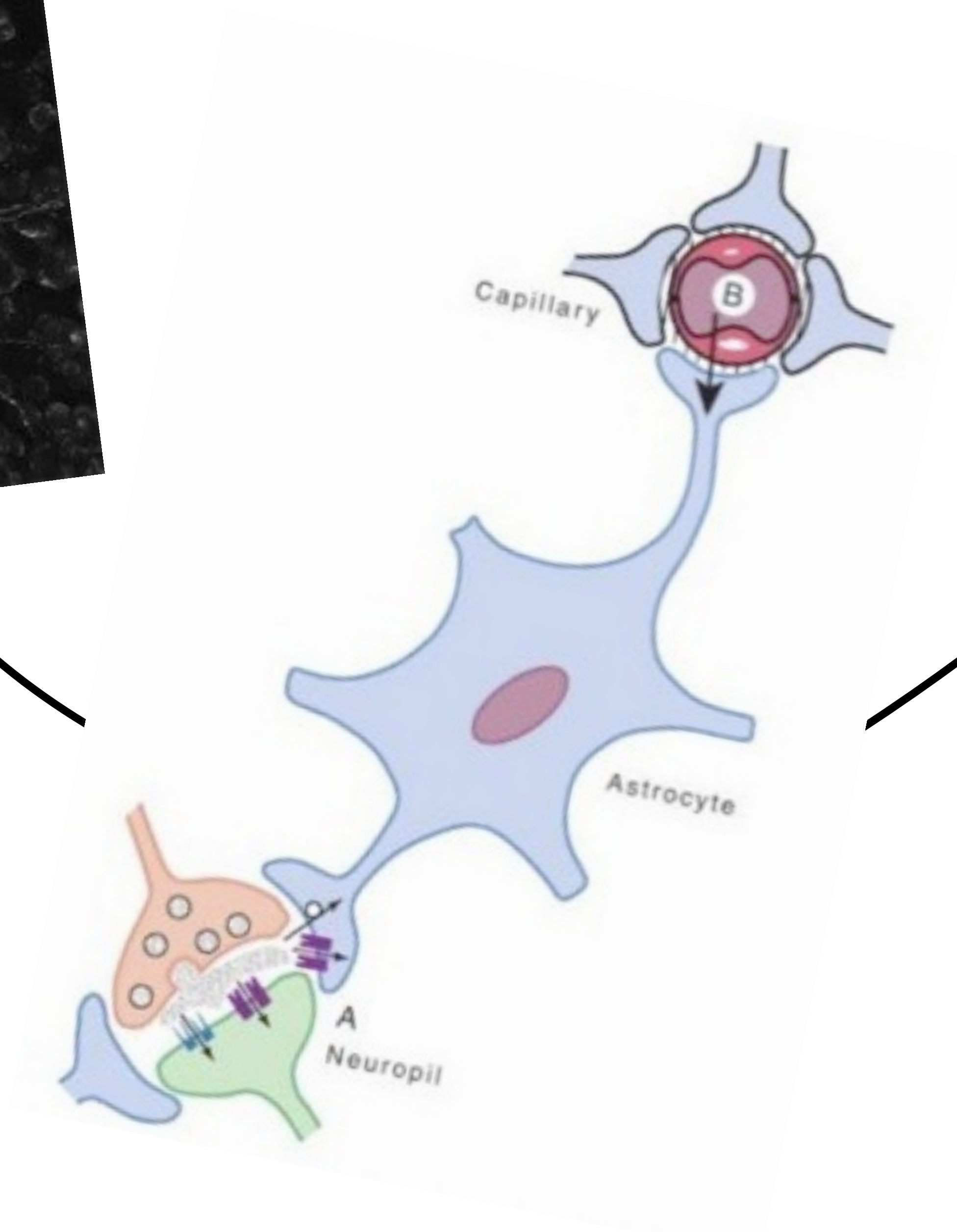
- intracranial electrodes
- Electrocorticograms (ECoG)
- Electroenchephalography (EEG)
- Magnetoencelography (MEG)

Data Specs:

- High temporal resolution
- Low spatial resolution







Neurovascular
Coupling

Blood-Oxygen Level Dependent Signal

Neural activation is reflected in BOLD contrast

Single whisker deflection in rats

Optical measurements:

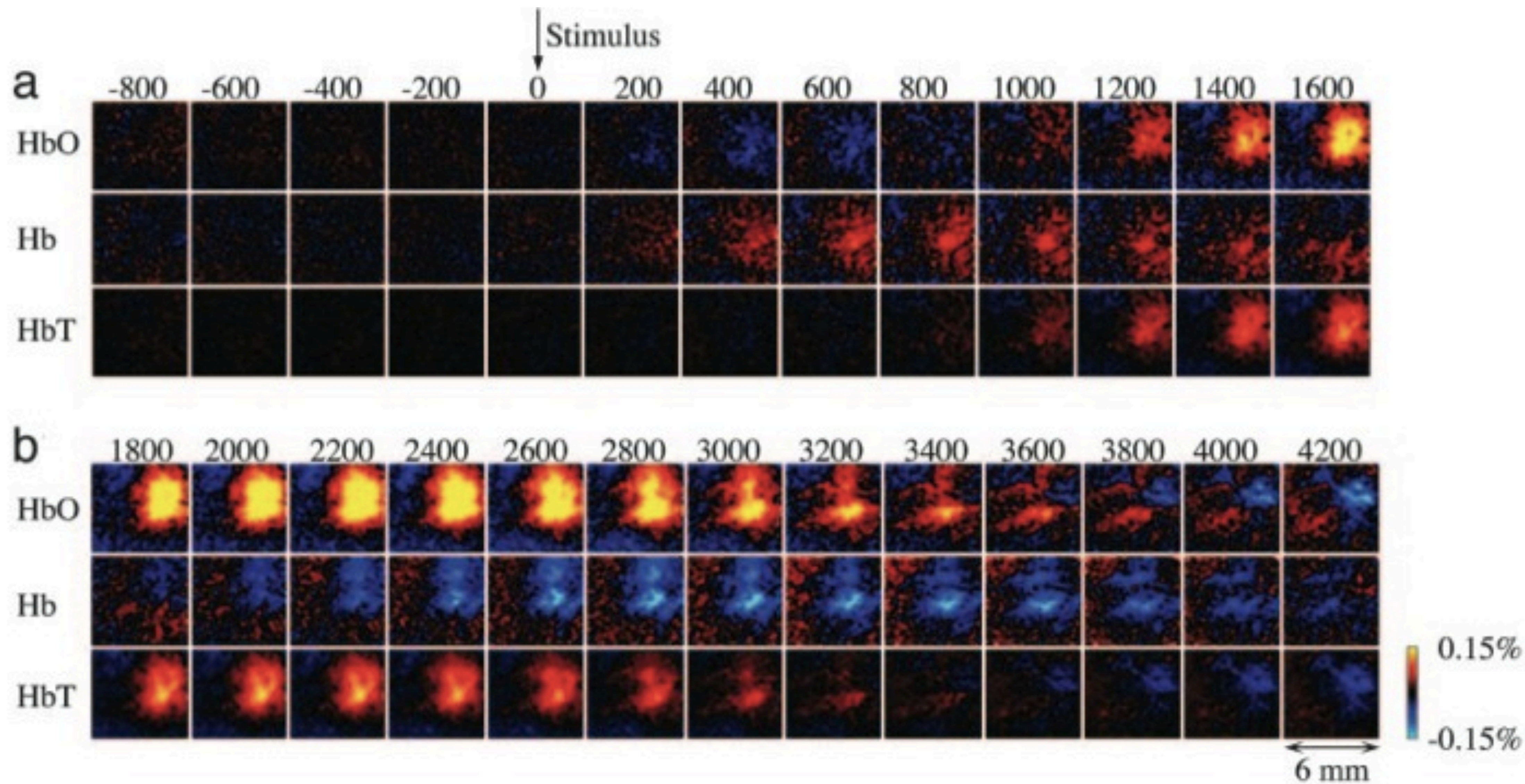
deoxyhemoglobin (Hb)

oxyhemoglobin (HbO)

total hemoglobin (HbT)

Blood-Oxygen Level Dependent Signal

Neural activation is reflected in BOLD contrast



Devor, PNAS, 2005

Blood-Oxygen Level Dependent Signal

Neural activation is reflected in BOLD contrast

Measured by:

- Intrinsic optical imaging
- Near-infrared Spectroscopy (NIRS)
- functional Magnetic Resonance Imaging (fMRI)

Data Specs

- High spatial resolution
- Low temporal resolution

Single whisker deflection in rats

Optical measurements:

deoxyhemoglobin (Hb)

oxyhemoglobin (HbO)

total hemoglobin (HbT)

Multimodal Neuroimaging: Benefits and Challenges

Benefits

- Clinical Applications
  - Better Diagnosis (e.g. Epilepsy)
  - Therapy: Hybrid Brain-Computer Interfaces
- Basic Research
  - Better Understanding of Single Modalities
  - Better Understanding of Modality Coupling

Challenges

- Recording Setups (Artifacts)
- Analysis Approaches (No Gold Standard)

Outline

1. Analysis of Multimodal Neuroimaging Data
    - Supervised Learning Approaches
    - Unimodal Unsupervised Approaches
    - Multimodal Unsupervised Approaches

2. Applications
    - Hybrid BCIs: NIRS and EEG
    - Cleaning Artifacts from Multimodal Recordings
    - Estimating the Neural Information in fMRI Signals
    - Multisubject Analyses

Multimodal Neuroimaging

Most of what will be discussed today is described in detail in

IEEE REVIEWS IN BIOMEDICAL ENGINEERING

Analysis of Multimodal Neuroimaging Data

Felix Bießmann[◇], Sergey Plis[§], Frank Meinecke[◇], Tom Eichele[*], Klaus-Robert Müller[◇,‡]

Biessmann et al., IEEE Reviews in Biomedical Engineering, 2011

Matlab Code, toydata examples and real data examples available at
http://www.user.tu-berlin.de/felix.biessmann/mmreview/

Analysis of Multimodal Neuroimaging Data

Analysis of Multimodal Neuroimaging Data

1. No gold standard analysis for Multimodal Neuroimaging
2. Analysis approaches are difficult to categorize
3. Many analyses combine different methods
4. Most of these methods are based on simple algebra
5. Use these tools to tailor your analyses to your needs

Analysis of Multimodal Neuroimaging Data

- Supervised Analyses $$w_m^\top x_m(t) = y(t) + \epsilon$$
- Unimodal Unsupervised Analyses $$x_m(t) = A_m s_m(t) + \epsilon$$
- Biologically Inspired Models $$x(t) = m_\Phi(t) + \epsilon$$
- Multimodal Unsupervised Analyses $$x_m(t) = A_m s(t) + \epsilon$$

Supervised Analyses

Supervised Analyses

$$w_m^\top x_m(t) = y(t) + \epsilon$$

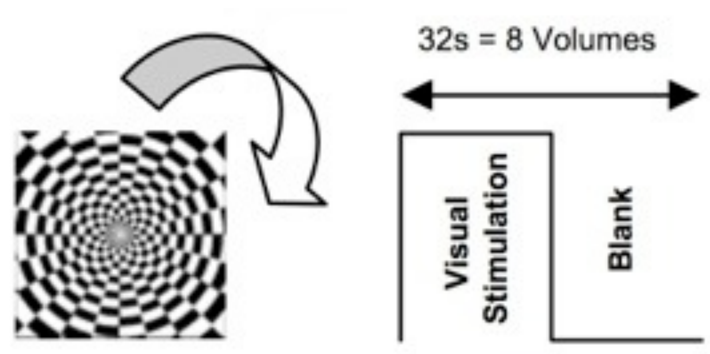
Supervised analyses fit data $x_m(t)$ to some label $y(t)$

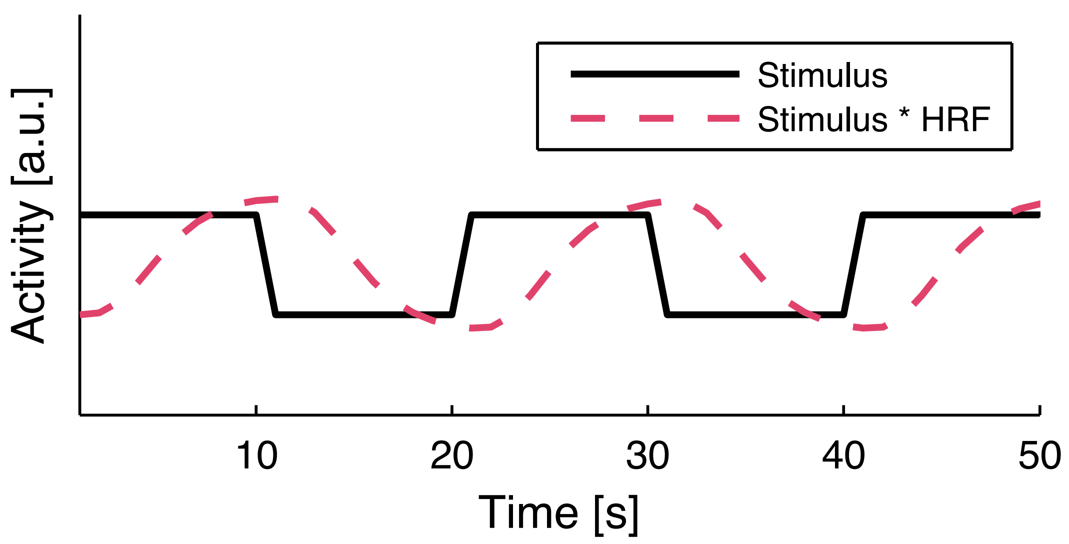
$x_m(t), y(t)$ Data from either modality / Stimulus

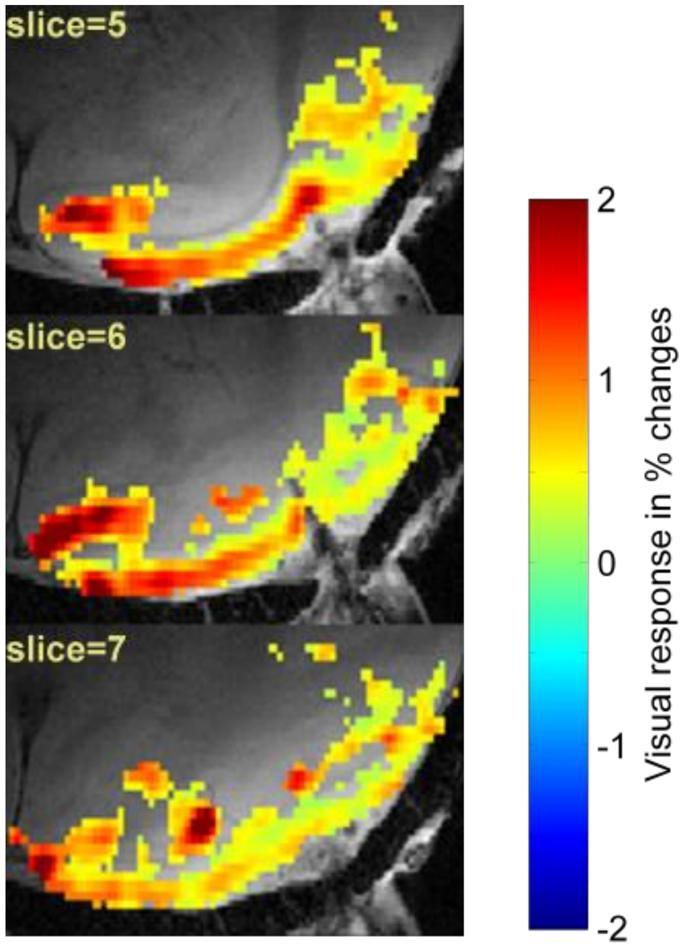
Examples:

GLMs, Linear Discriminant Analysis, Support Vector Machines

Supervised Analysis: General Linear Models







EEG Bandpower as Label for fMRI GLM

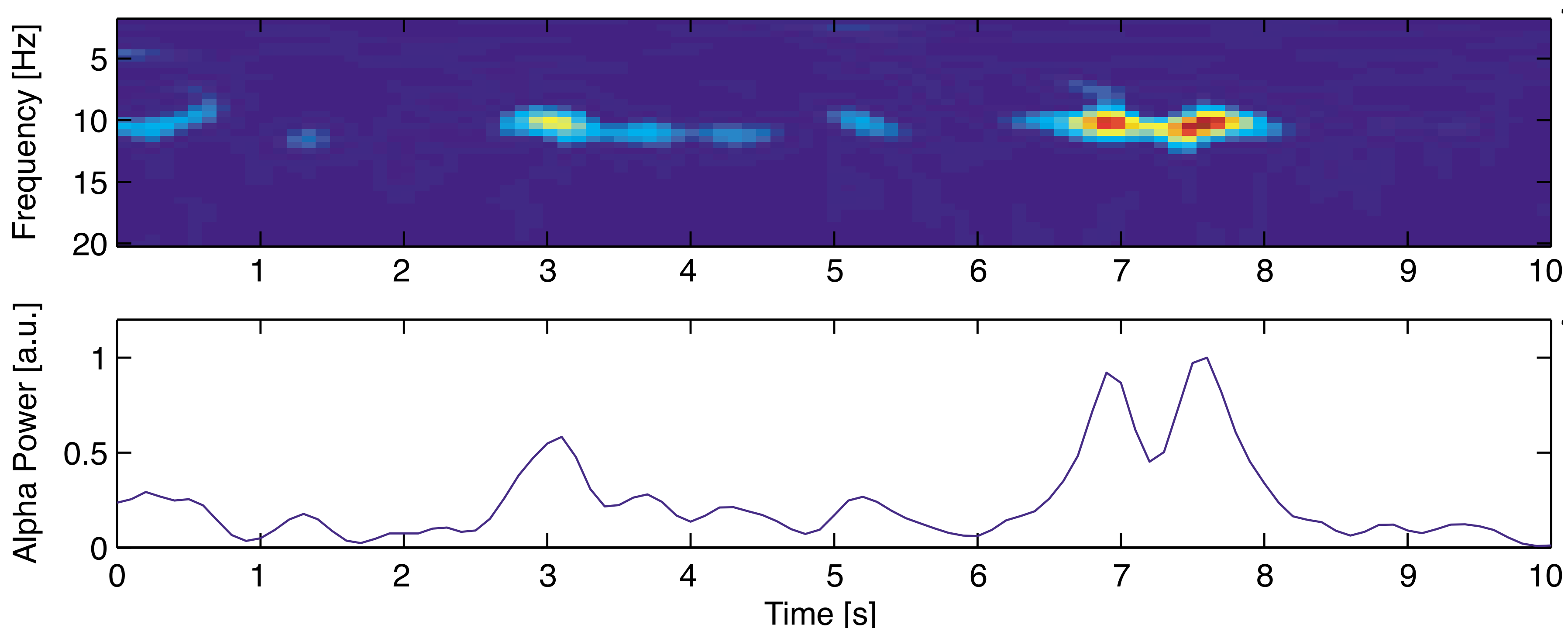
$$w_m^\top x_m(t) = y(t) + \epsilon$$

EEG features as regressor

e.g. Moosmann et al., Neuroimage, 2003

$x_m(t)$ EEG alpha bandpower

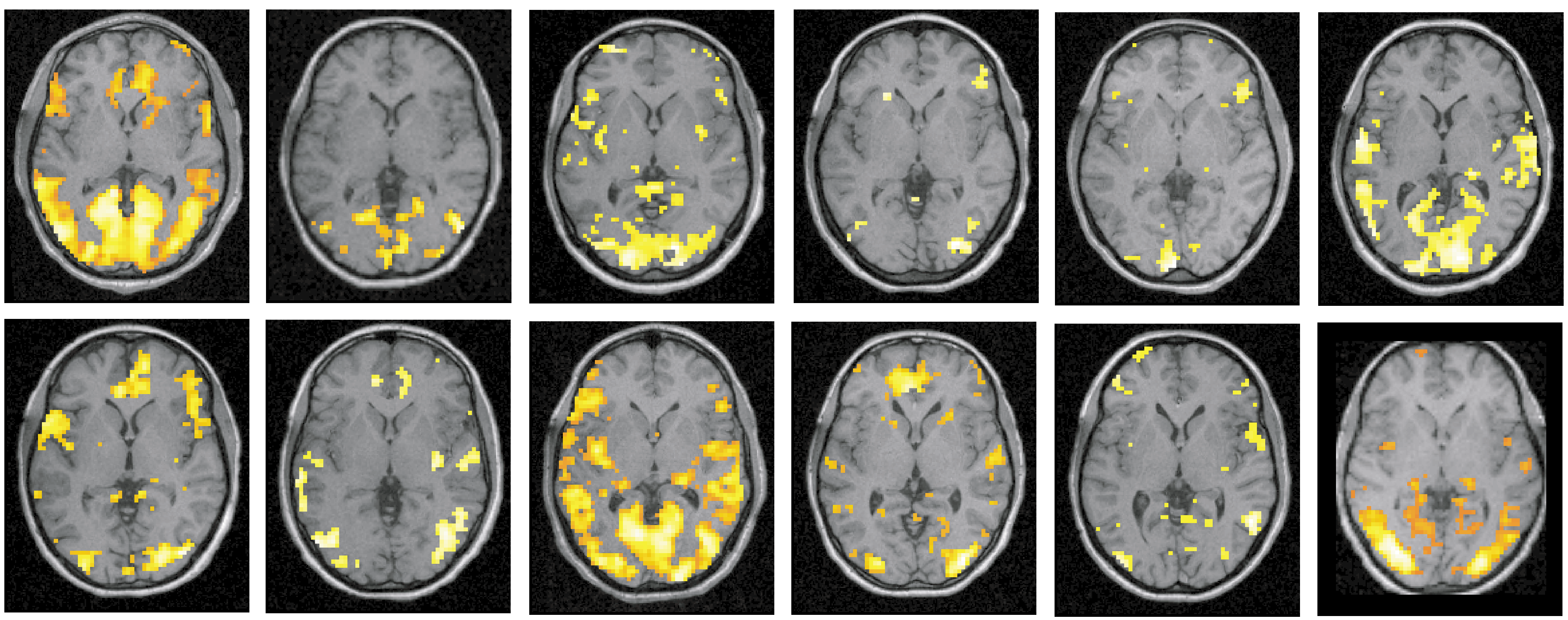
$y(t)$ fMRI voxel time series



$$w_m^\top x_m(t) = y(t) + \epsilon$$

EEG features as regressor

e.g. Moosmann et al., Neuroimage, 2003

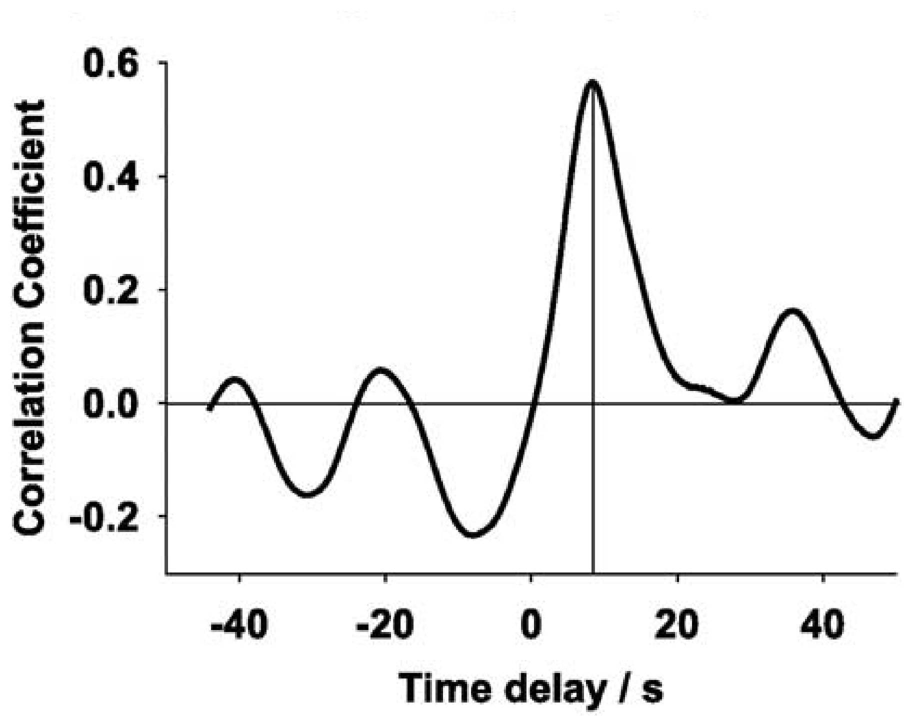


SPM negative activation patterns

$$w_m^\top x_m(t) = y(t) + \epsilon$$

EEG features as regressor

e.g. Moosmann et al., Neuroimage, 2003



Cross-correlogram with NIRS

$$w_m^\top x_m(t) = y(t) + \epsilon$$

Supervised analyses fit data $x_m(t)$ to some label $y(t)$

Problems

Choice of target variable difficult (often there is none)

EEG Bandpower regressor incompatible with physics

When regressing onto stimulus, all variance related to label (not necessarily neural activation) is used

Unimodal Unsupervised Analyses

Unsupervised Analyses

$$x_m(t) = A_m s_m(t) + \epsilon$$

Many unsupervised analyses learn mapping A_m

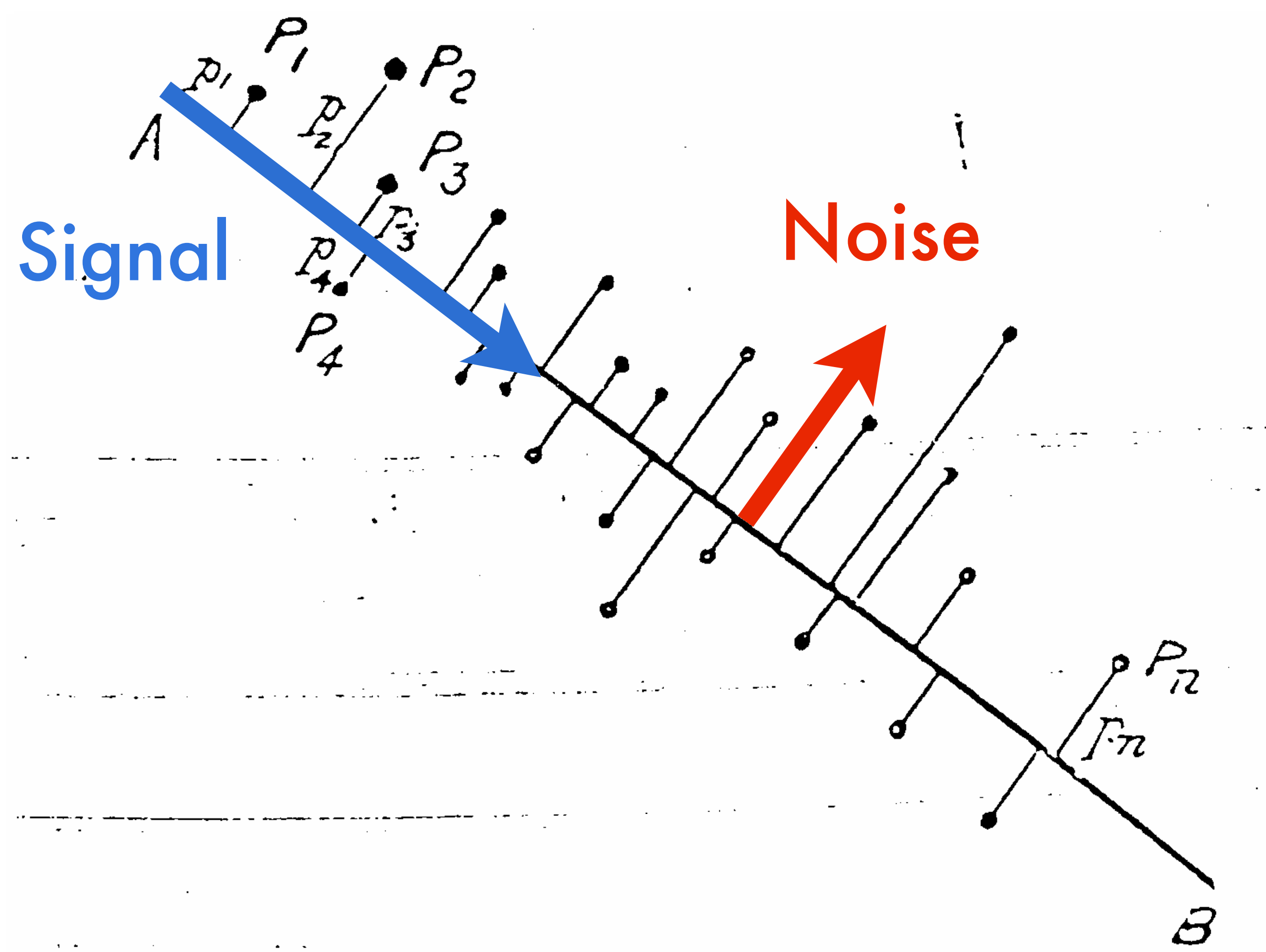
from (modality specific) neural sources $s_m(t)$

to unimodal measurements $x_m(t)$

Examples:

Principal/Independent Component Analysis, Clustering

Which line fits this data best?



The line that **maximizes the variance** after projecting the data onto it

Pearson, Philosophical Magazine, 1901

We store the data in a matrix

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \in \mathbb{R}^{D \times N}$$

PCA finds a line $\mathbf{w}^* \in \mathbb{R}^D$

such that the variance of the data projected onto the line is maximized

$$\mathbf{w}^* = \operatorname*{argmax}_{\mathbf{w}} \mathbf{w}^\top \mathbf{X}\mathbf{X}^\top \mathbf{w}$$

$$\|\mathbf{w}\|^2 = \mathbf{w}^\top \mathbf{w} = 1$$

Pearson, Philosophical Magazine, 1901

Setting up the Lagrangian and
rearranging terms in its first derivative …

$$\frac{\partial \mathcal{L}}{\partial w} = 2\mathbf{X}\mathbf{X}^\top \mathbf{w} - 2\lambda \mathbf{w} = 0$$

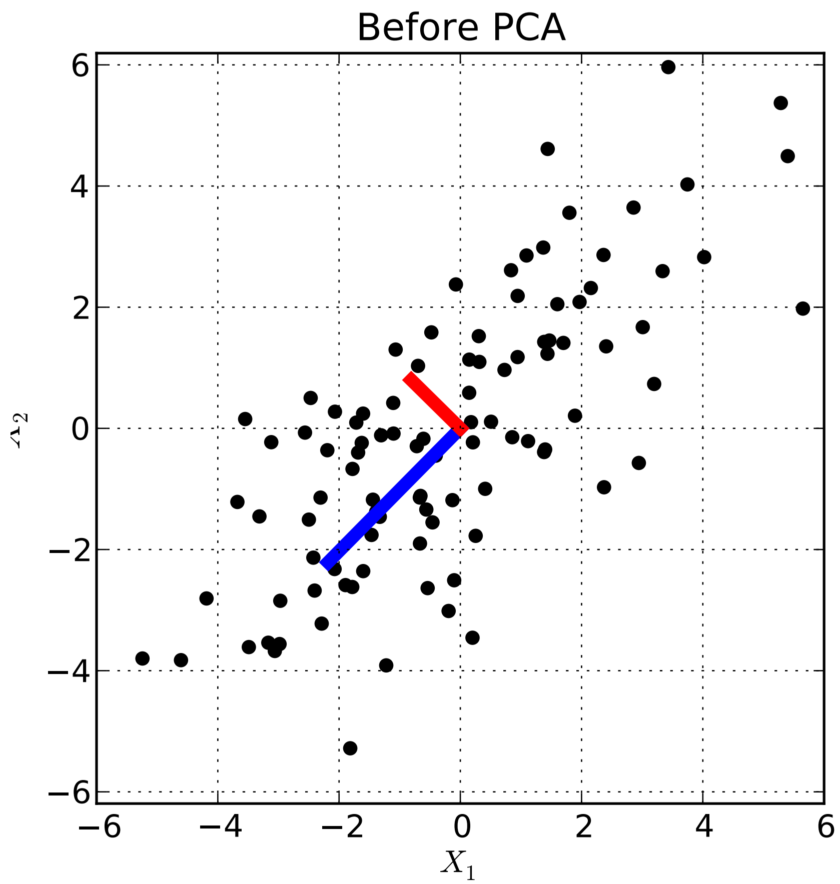
$$\Rightarrow \mathbf{X}\mathbf{X}^\top \mathbf{w} = \lambda \mathbf{w}$$

We see that the best line for describing the data is
the strongest eigenvector of the covariance matrix

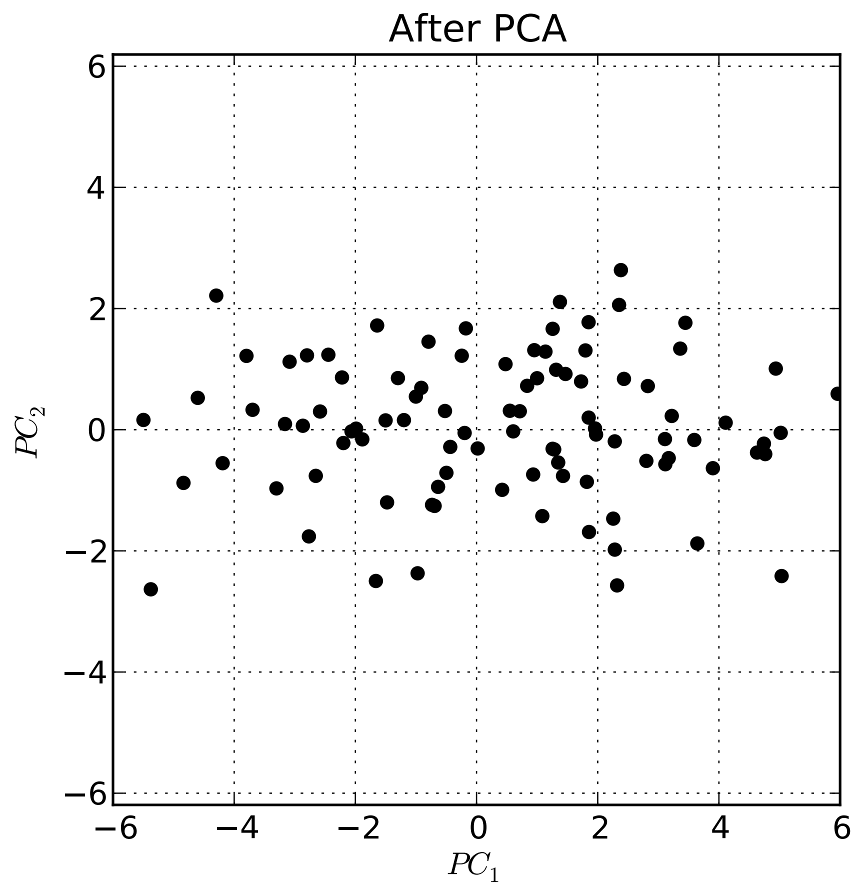
Pearson, Philosophical Magazine, 1901

Principal Component Analysis

$\mathbf{X}$

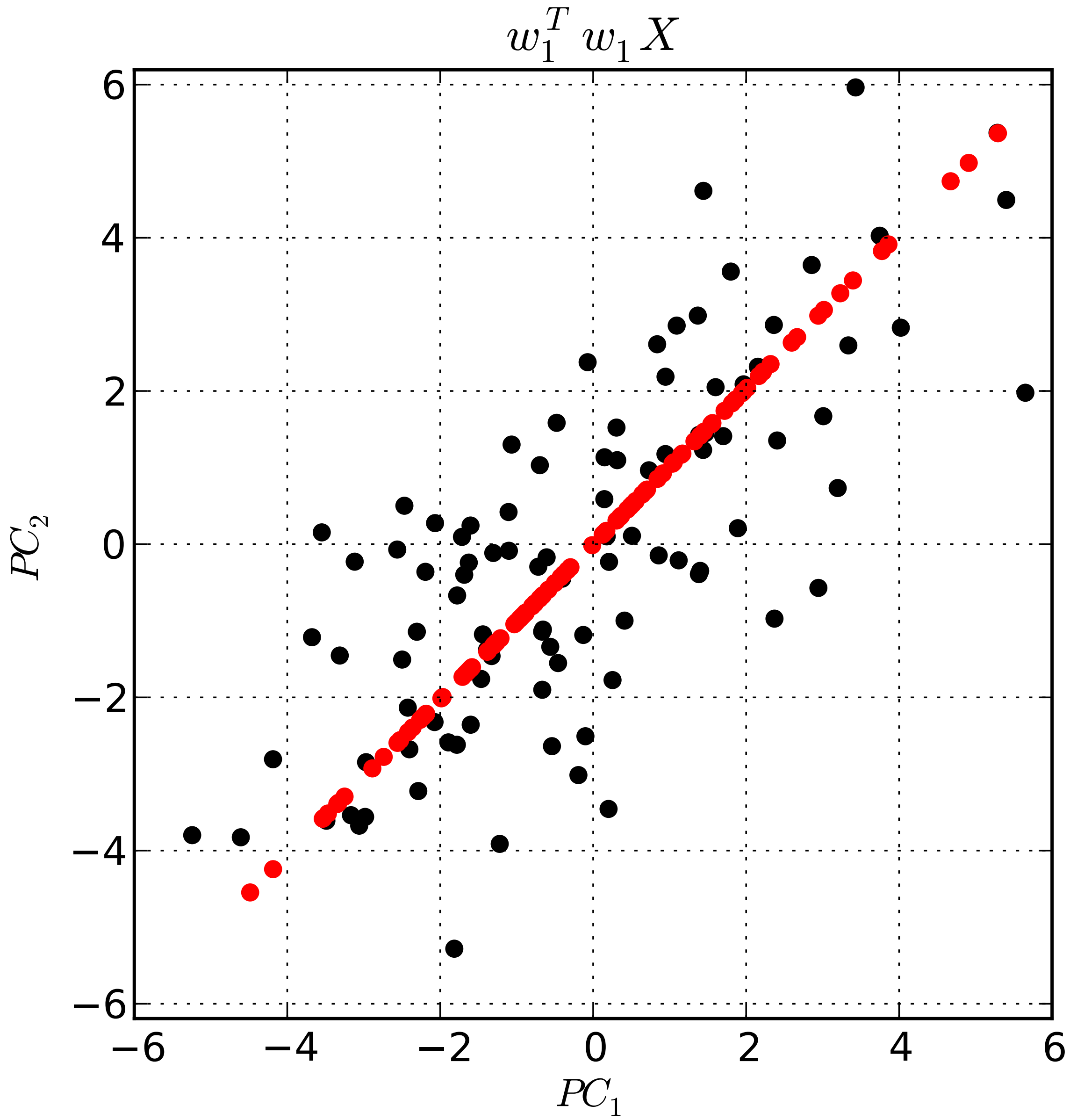


$\mathbf{W}^\top \mathbf{X}$

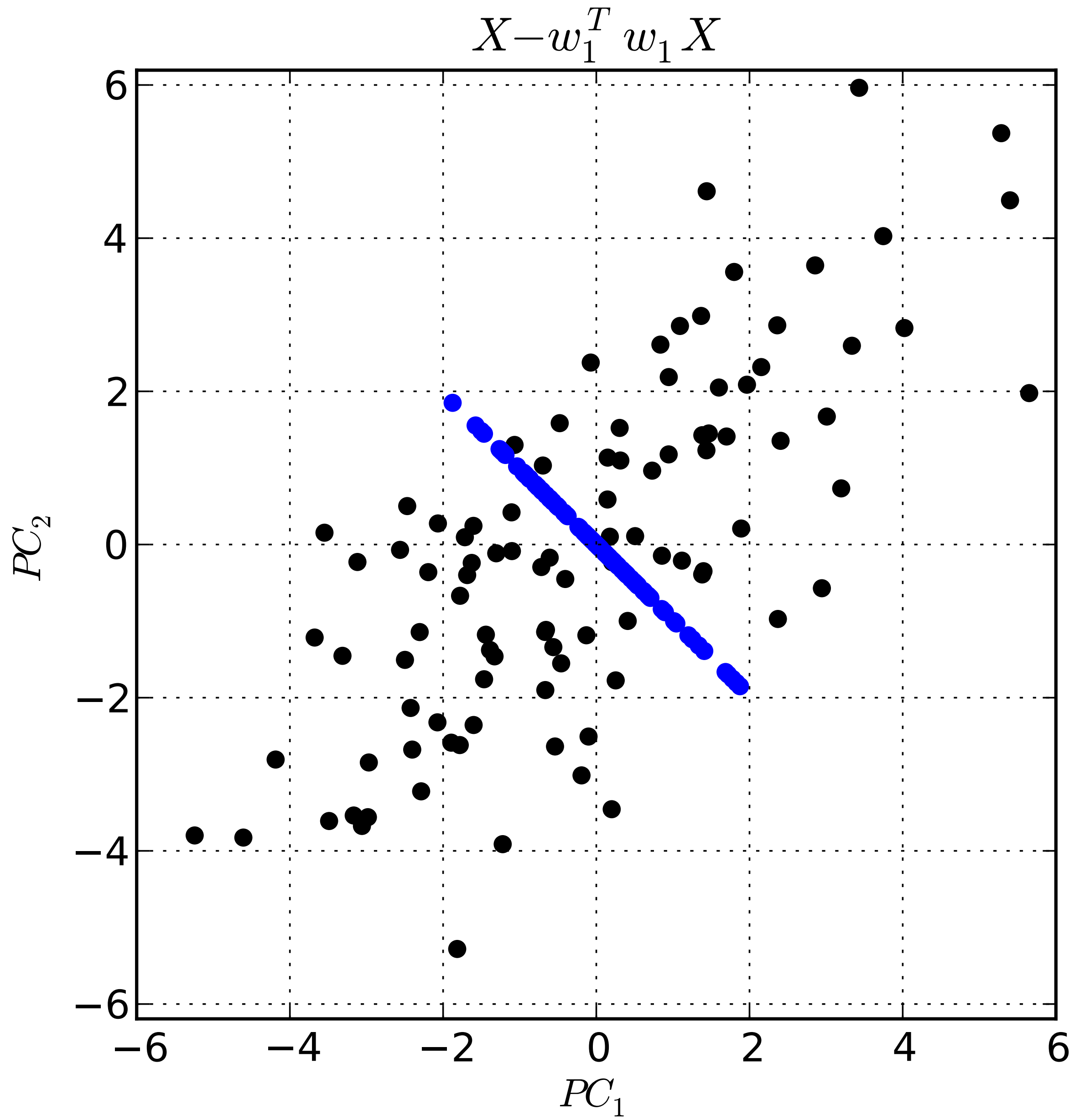


PCA aligns maximum variance directions with standard basis
Now we can remove each dimension separately

$$\mathbf{X}_{\mathbf{w}_1} = \mathbf{w}_1 \mathbf{w}_1^\top \mathbf{X}$$



$$\mathbf{X} - \mathbf{w}_1\mathbf{w}_1^\top\mathbf{X} = (\mathbf{I} - \mathbf{w}_1\mathbf{w}_1^\top)\mathbf{X}$$



What if there are (many) more dimensions than samples?

We can use the **kernel trick** (here's just the linear one)

$$\mathbf{w} = \mathbf{X}\mathbf{a}$$

$$\mathbf{X} \underbrace{\mathbf{X}^\top \mathbf{X}}_{\text{Kernel } \mathbf{K}_X} \mathbf{a} = \lambda \mathbf{X}\mathbf{a}$$

Schoelkopf et al, Neural Computation, 1998

What if there are (many) more dimensions than samples?

We can use the **kernel trick** (here's just the linear one)

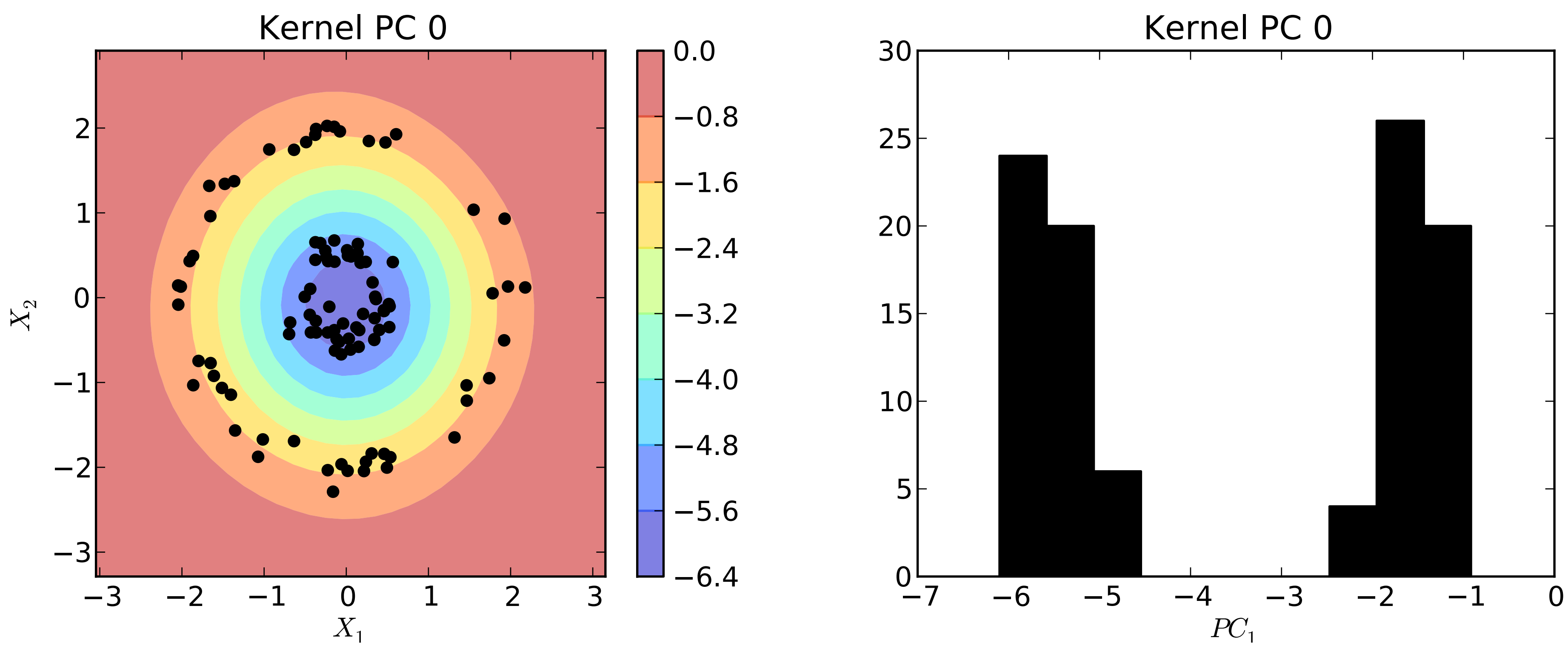
$$\mathbf{w} = \mathbf{X}\mathbf{a}$$

$$\mathbf{X} \; \underbrace{\mathbf{X}^\top \mathbf{X}}_{\text{Kernel } \mathbf{K}_X} \; \mathbf{a} = \lambda \mathbf{X}\mathbf{a}$$

Schoelkopf et al, Neural Computation, 1998

Kernel Principal Component Analysis

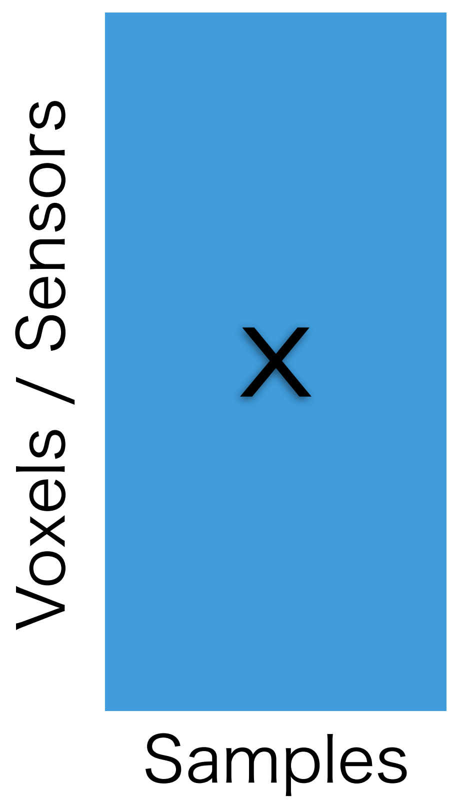
Using nonlinear kernels we can fit arbitrary manifolds



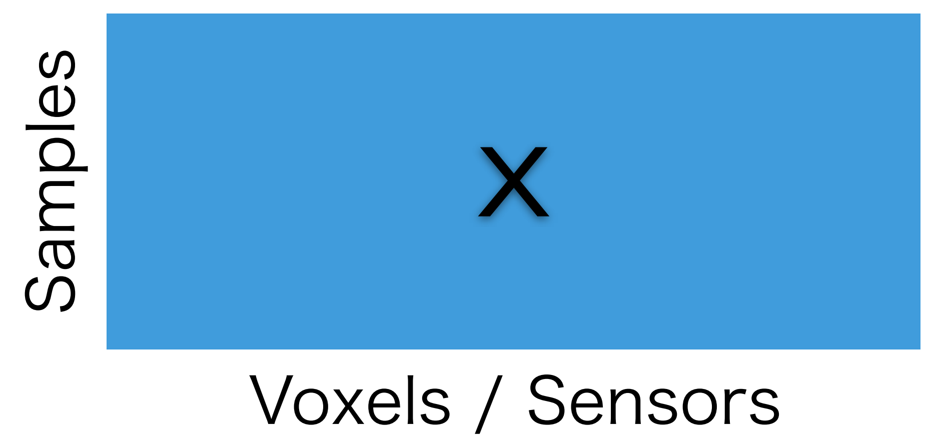
Schoelkopf et al, Neural Computation, 1998

Principal Component Analysis

Temporal PCA



Spatial PCA



For fMRI data linear Kernel PCA

can be thought of as Spatial PCA

Unimodal Unsupervised Analyses

$$x_m(t) = A_m s(t) + \epsilon$$

Many unsupervised analyses learn mapping A_m

from neural sources $s_m(t)$

to unimodal measurements $x_m(t)$

Problems

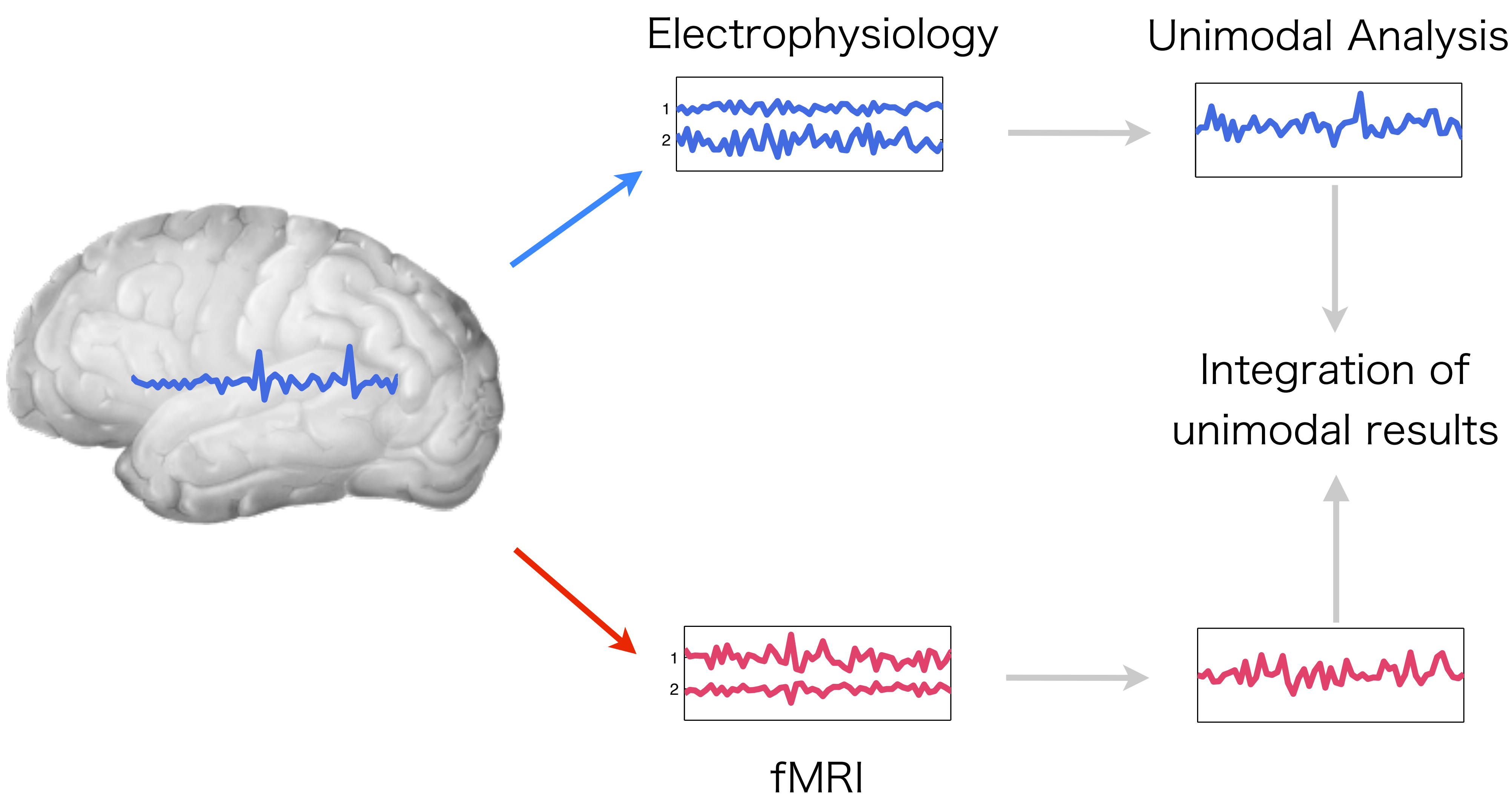
Correspondence of unimodal components

One component is only present in one modality

One component corresponds to multiple components in other modality

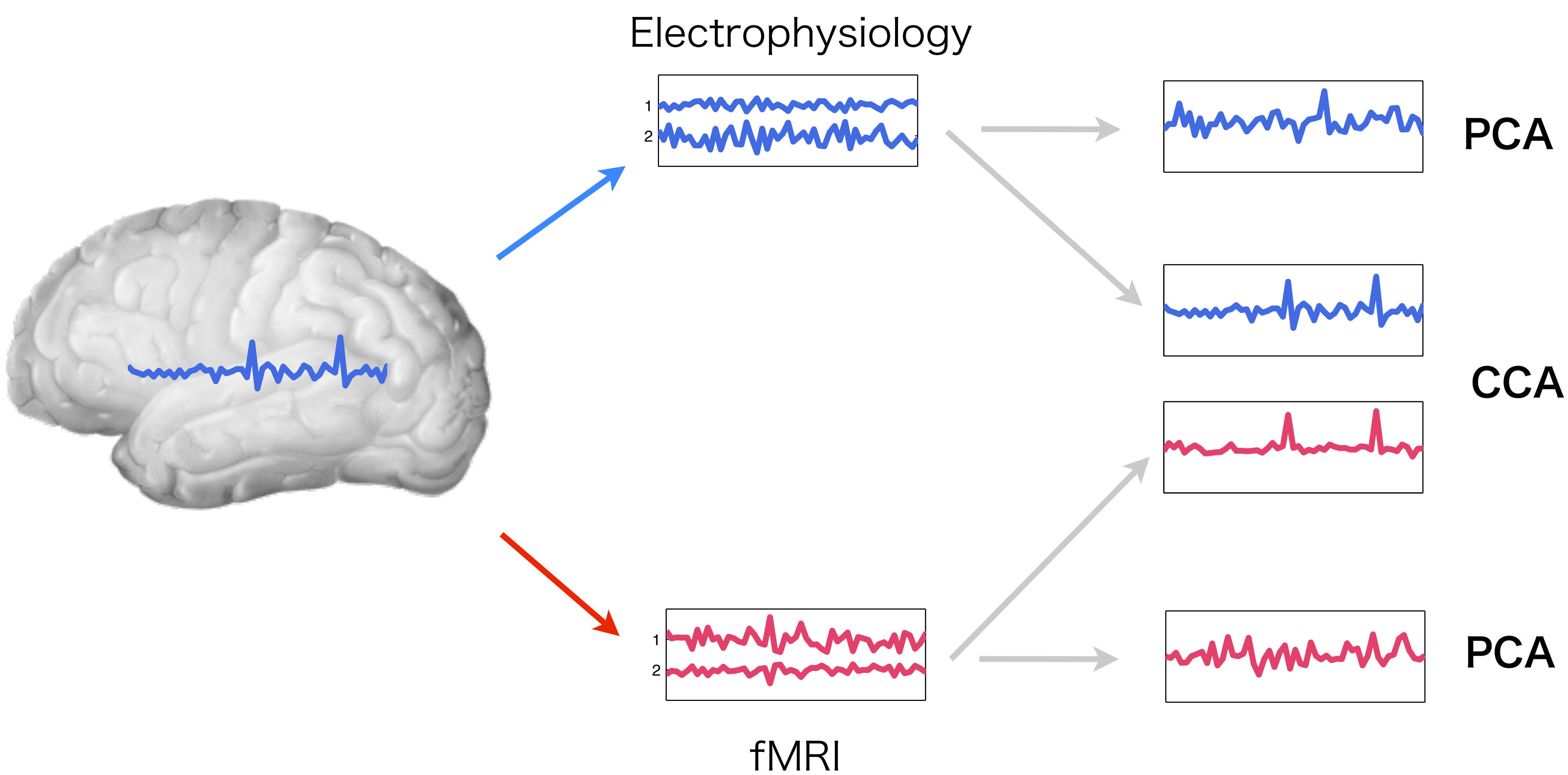
Multimodal Unsupervised Analyses

Classical Approach To Multimodal Neuroimaging



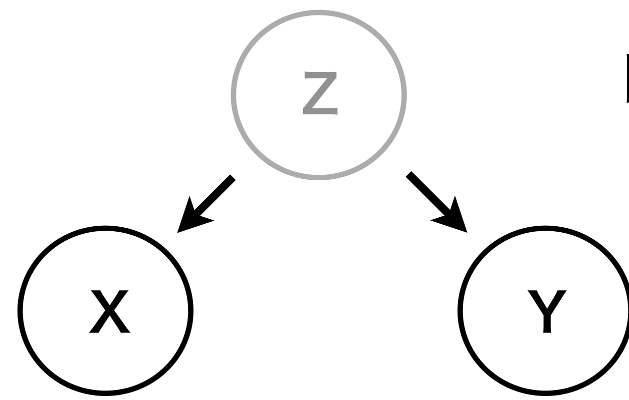
- Unimodal preprocessing might discard relevant information

When Unimodal Methods Fail: CCA vs. PCA



- Unimodal preprocessing might discard relevant information
- CCA recovers the **common** underlying variable

Canonical Correlation Analysis



Hidden Variable

Measured Variables

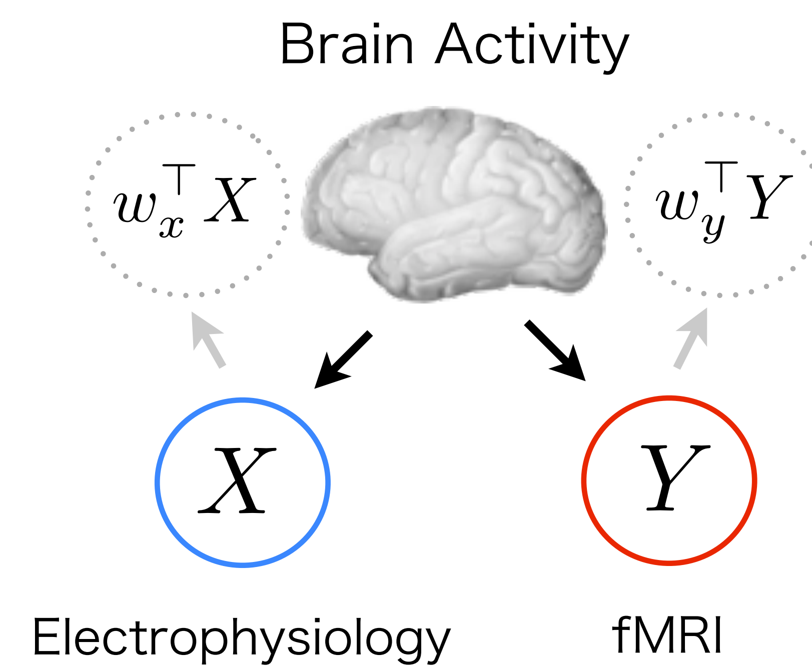
Canonical Correlation Analysis

Given Data $X \in \mathbb{R}^{F \times T}$

$Y \in \mathbb{R}^{S \times T}$

CCA finds canonical directions $w_x \in \mathbb{R}^{F \times 1}$

$w_y \in \mathbb{R}^{S \times 1}$



such that the correlation between X and Y is maximized:

$$\underset{w_x, w_y}{\operatorname{argmax}} \left(\frac{w_x^\top X Y^\top w_y}{\sqrt{w_x^\top X X^\top w_x w_y^\top Y Y^\top w_y}} \right)$$

Jordan 1875, Hotelling 1936

$$\underset{w_x, w_y}{\operatorname{argmax}} \, w_x^\top C_{xy} w_y \qquad \text{subject to} \qquad \begin{aligned} w_x^\top C_{xx} w_x &= 1 \\ w_y^\top C_{yy} w_y &= 1 \end{aligned}$$

$$\mathcal{L} = w_x^\top C_{xy} w_y - \frac{1}{2}\lambda_x \left(w_x^\top C_{xx} w_x - 1 \right) - \frac{1}{2}\lambda_y \left(w_y^\top C_{yy} w_y - 1 \right)$$

$$\underbrace{\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix}}_{A} \begin{bmatrix} w_x \\ w_y \end{bmatrix} = \underbrace{\begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix}}_{B} \begin{bmatrix} w_x \\ w_y \end{bmatrix} \lambda$$

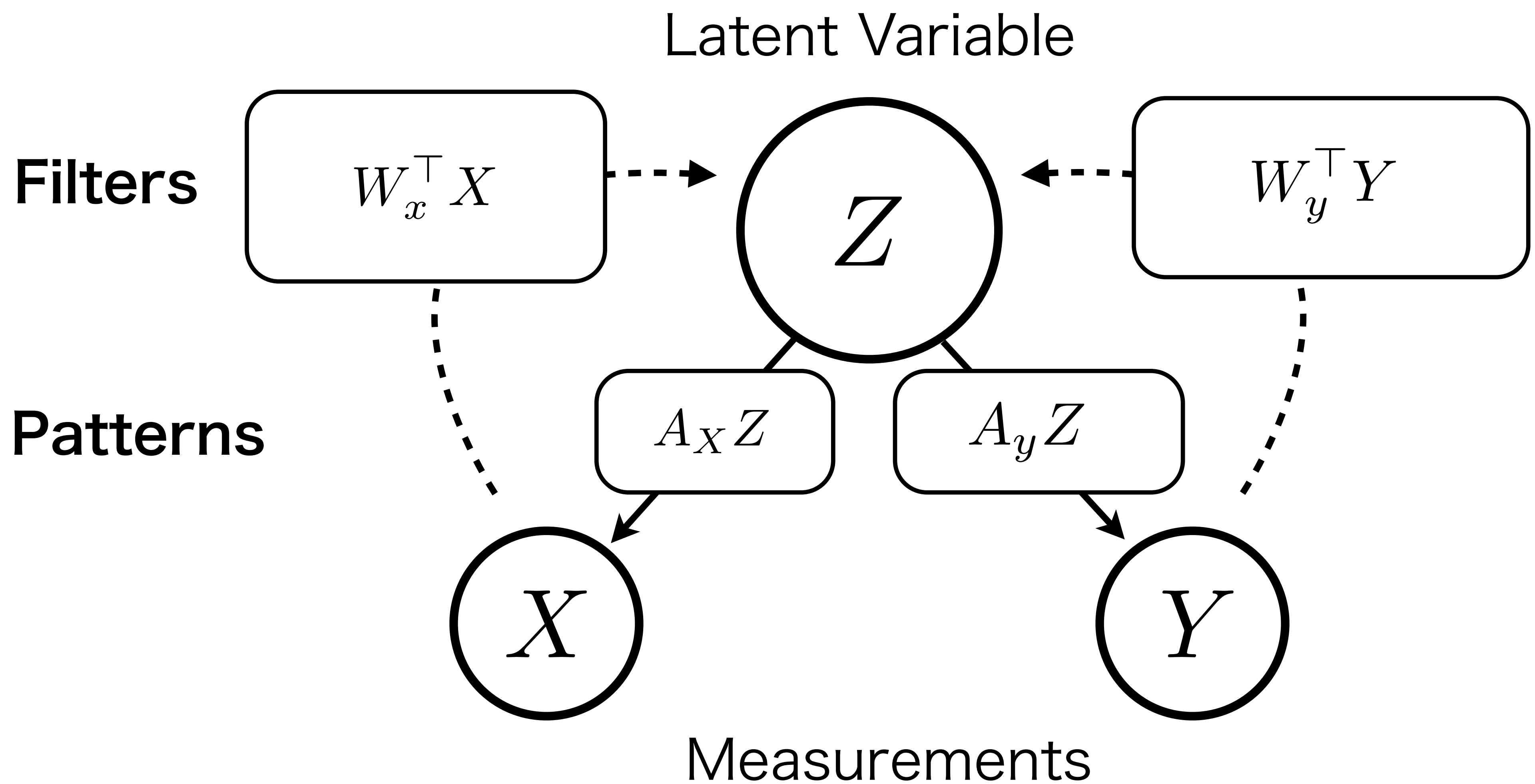
CCA and Other Projection Methods

$$Aw = Bw\lambda$$

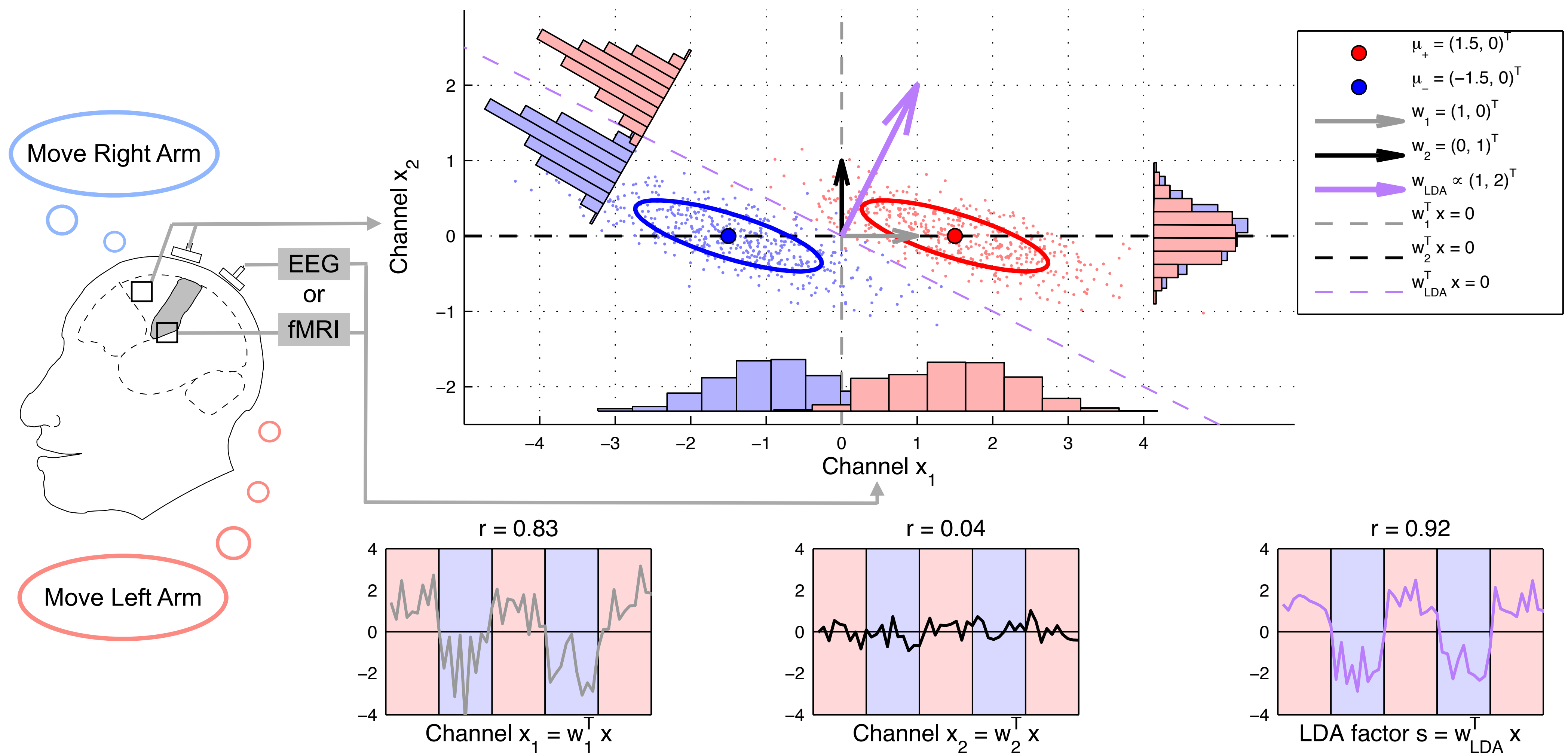
		$\mathbf{A}$	$\mathbf{B}$
Maximizes Variance	PCA	$\mathbf{C}_{xx}$	$\mathbf{I}$
Maximizes Covariance	PLS	$\begin{pmatrix} \mathbf{0} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{0} \end{pmatrix}$	$\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$
Maximizes Correlation	CCA	$\begin{pmatrix} \mathbf{0} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{0} \end{pmatrix}$	$\begin{pmatrix} \mathbf{C}_{xx} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{yy} \end{pmatrix}$
	MLR	$\begin{pmatrix} \mathbf{0} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{0} \end{pmatrix}$	$\begin{pmatrix} \mathbf{C}_{xx} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$

Borga, PhD Thesis, 1998

CCA Filters and CCA Patterns



Why You Should Not Interpret Filter Coefficients



Haufe et al, 2014

Transforming Filters into Patterns

If the sources are uncorrelated

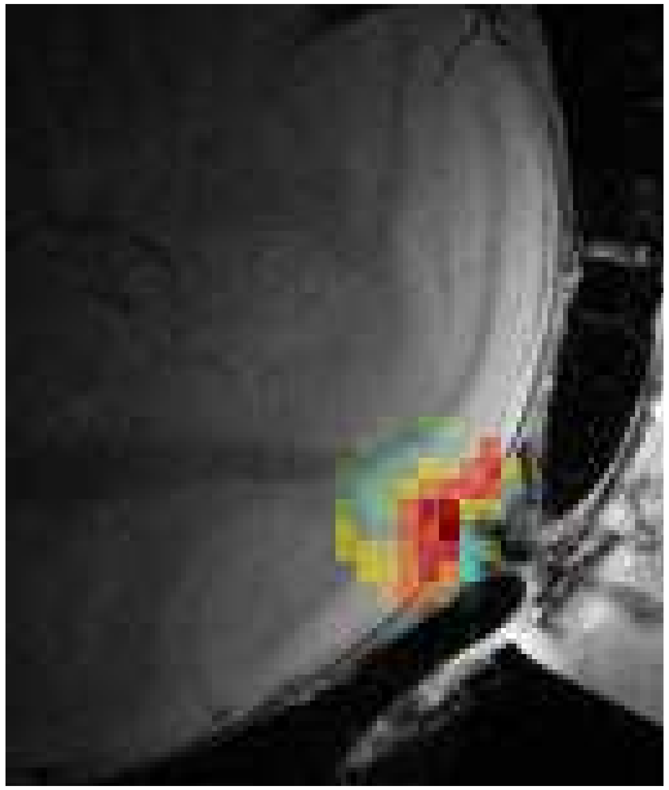
(they are for PCA/ICA/CCA and all regression/classification)

$$A \propto XX^\top W$$

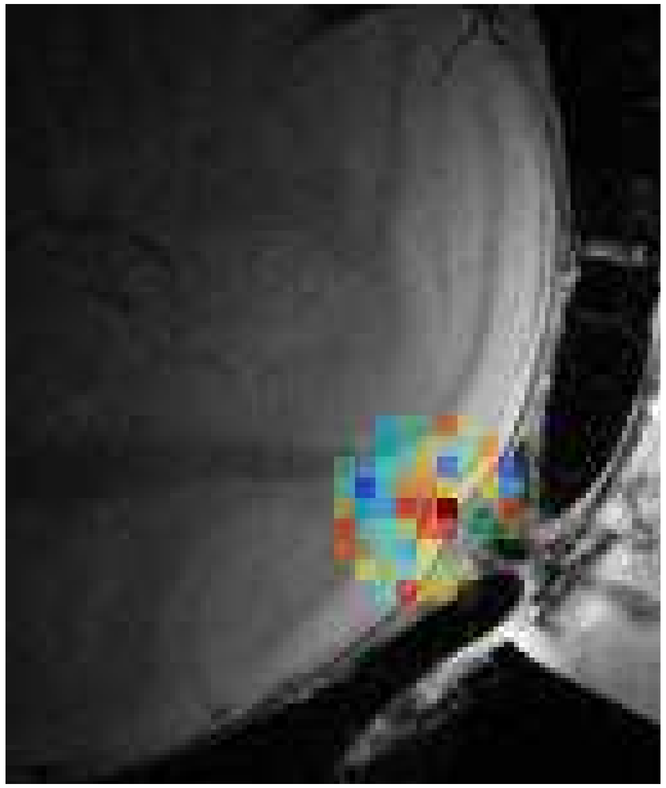
Parra et al. 2005, Blankertz et al. 2011, Haufe et al, 2014

CCA Filters and CCA Patterns

Pattern



Filter



Haufe et al, 2014

Information Theoretic Measures

Mutual Information between two variables x, y

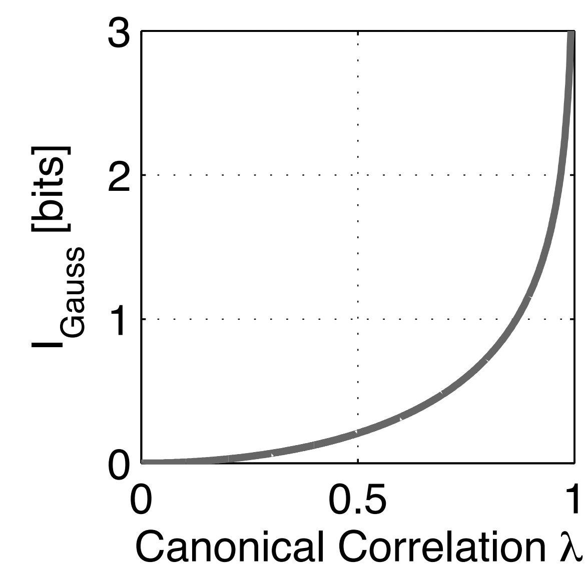
$$I(X,Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

- Almost all neuroimaging analyses use 2nd order statistics
- If data is described by mean/variance

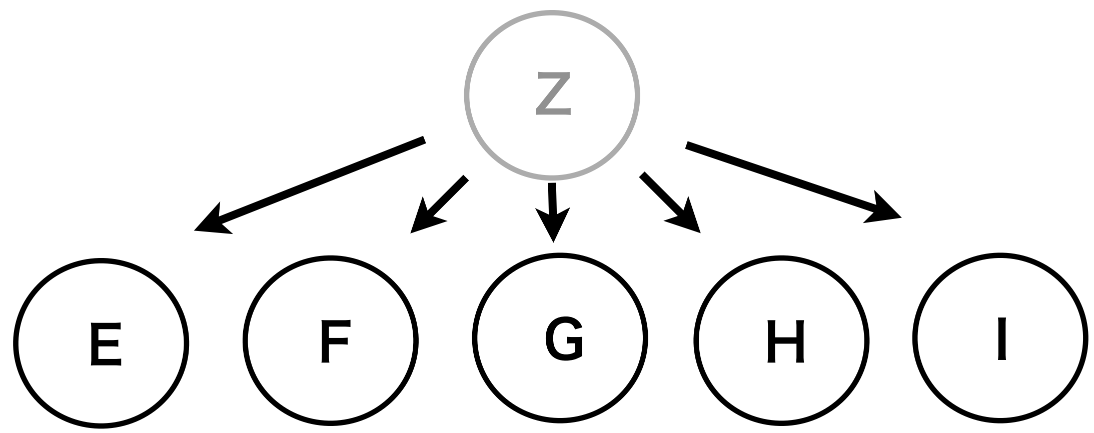
  Canonical Correlation

  $\sim$ Mutual information

$$I_{\text{Gauss}}(X,Y) = \frac{1}{2} \sum_i \log \left(\frac{1}{(1-\lambda_i^2)} \right)$$



For more than two variables we can extend the generalized eigenvalue problem



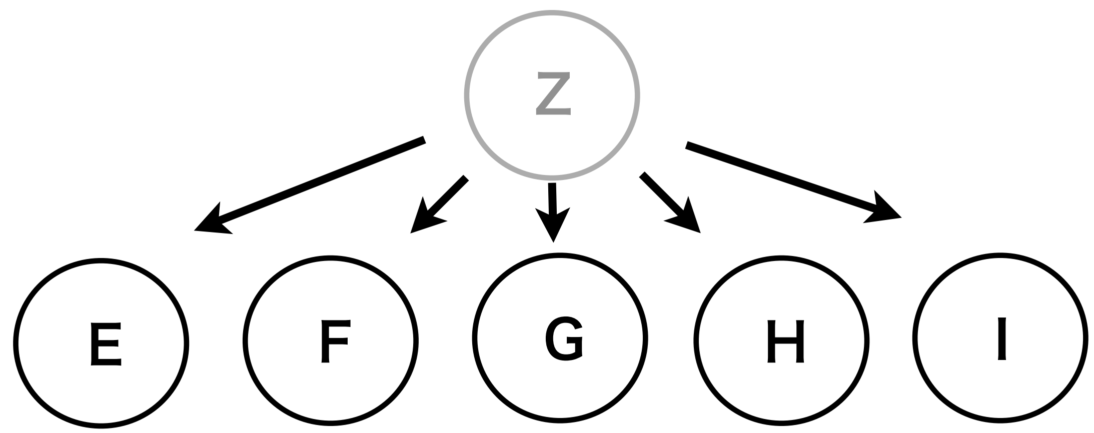
$$\operatorname*{argmax}_{W_i, W_j} \sum_i \sum_j \mathrm{Tr}\left(W_i^\top X_i X_j^\top W_j\right), \quad \forall i, j$$

$$\text{subject to } W_i^\top X_i X_i^\top W_i = \mathbf{I}, \quad \forall i,$$

Kettenring, Biometrika, 1971

CCA for Multiple Modalities

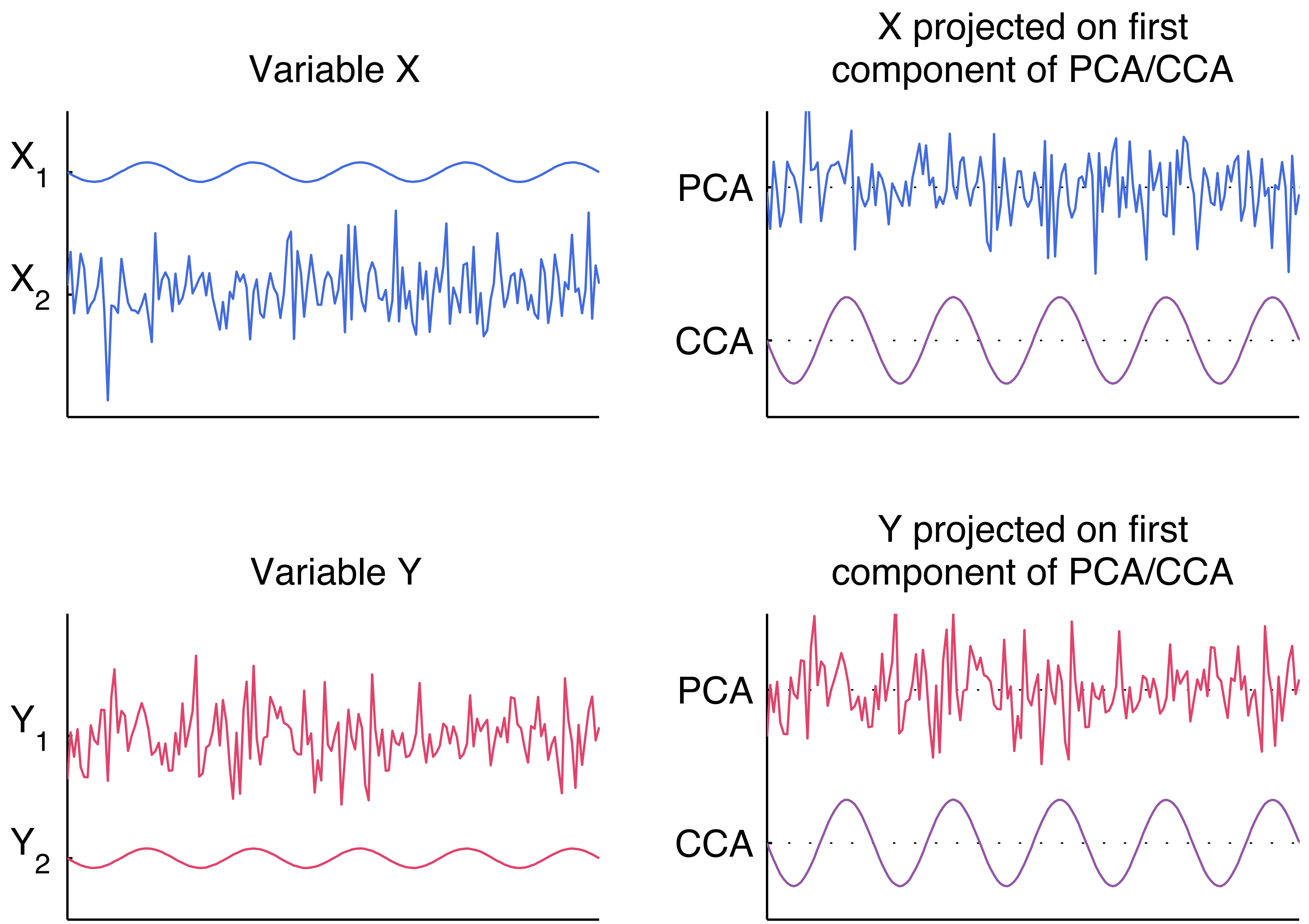
For more than two variables we can extend the generalized eigenvalue problem



$$\begin{bmatrix} 0 & C_{12} & \dots & C_{1N} \\ C_{21} & 0 & \dots & C_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ C_{N1} & C_{N2} & \dots & 0 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_N \end{bmatrix} = \begin{bmatrix} C_{11} & 0 & \dots & 0 \\ 0 & C_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & C_{NN} \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_N \end{bmatrix} \Lambda$$

Kettenring, Biometrika, 1971

Unimodal and Multimodal Analyses



Problems of Standard CCA

- CCA requires computation of covariance matrices $YY^\top$

  ! Computationally infeasible for fMRI

- CCA captures only linear dependencies

  ➡ Solution: Kernel CCA

Project data into a kernel feature space

$$\phi_x : \mathbb{R}^F \to \Phi_X,\ \phi_y : \mathbb{R}^S \to \Phi_Y$$

Optimize

$$\underset{\alpha,\beta}{\operatorname{argmax}} \left(\frac{\alpha^\top K_X K_Y^\top \beta}{\sqrt{\alpha^\top K_X^2 \alpha \beta^\top K_Y^2 \beta}} \right)$$

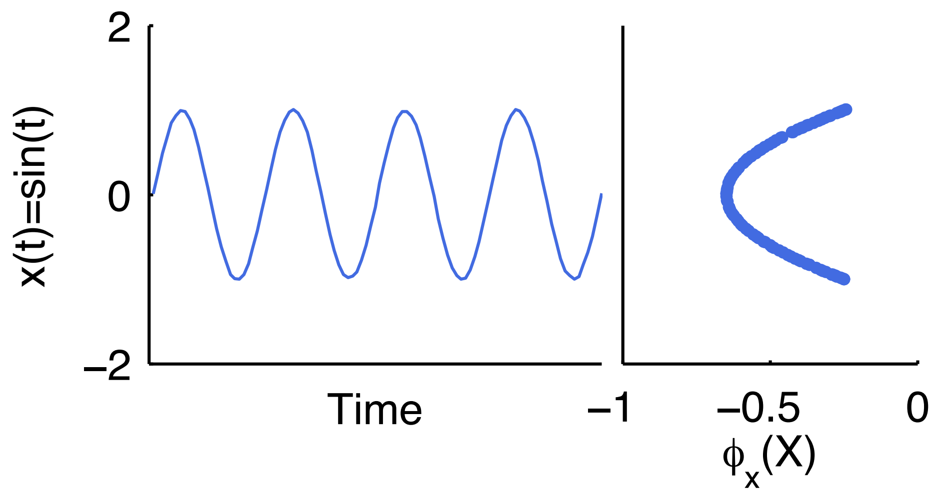
where

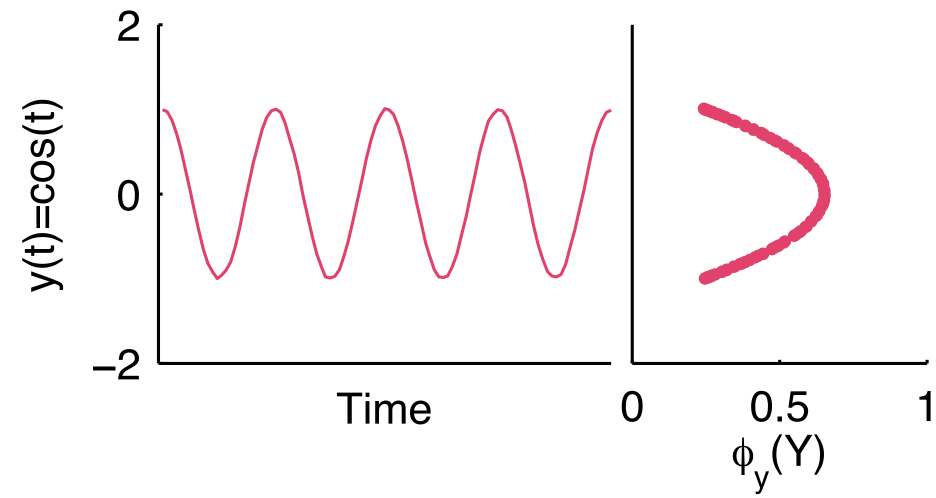
$$\begin{aligned} K_{X,ij} &= \langle \phi_x(x_i), \phi_x(x_j) \rangle_{\Phi_X} \\ K_{Y,ij} &= \langle \phi_y(y_i), \phi_y(y_j) \rangle_{\Phi_Y} \end{aligned}$$

kCCA can be solved efficiently in high (potentially infinite) dimensional spaces because:

$$K_X, K_Y \in \mathbb{R}^{T \times T}$$

Kernel CCA

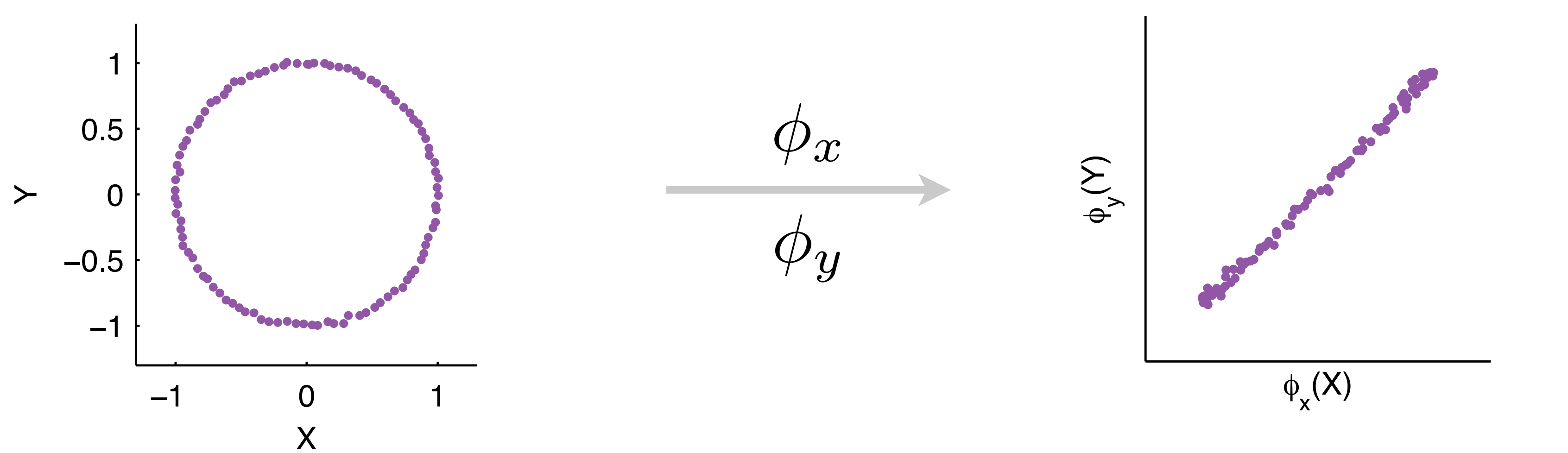




Non-linear dependencies become linear in kernel feature space

Linear Correlation 0

Kernel Correlation 1

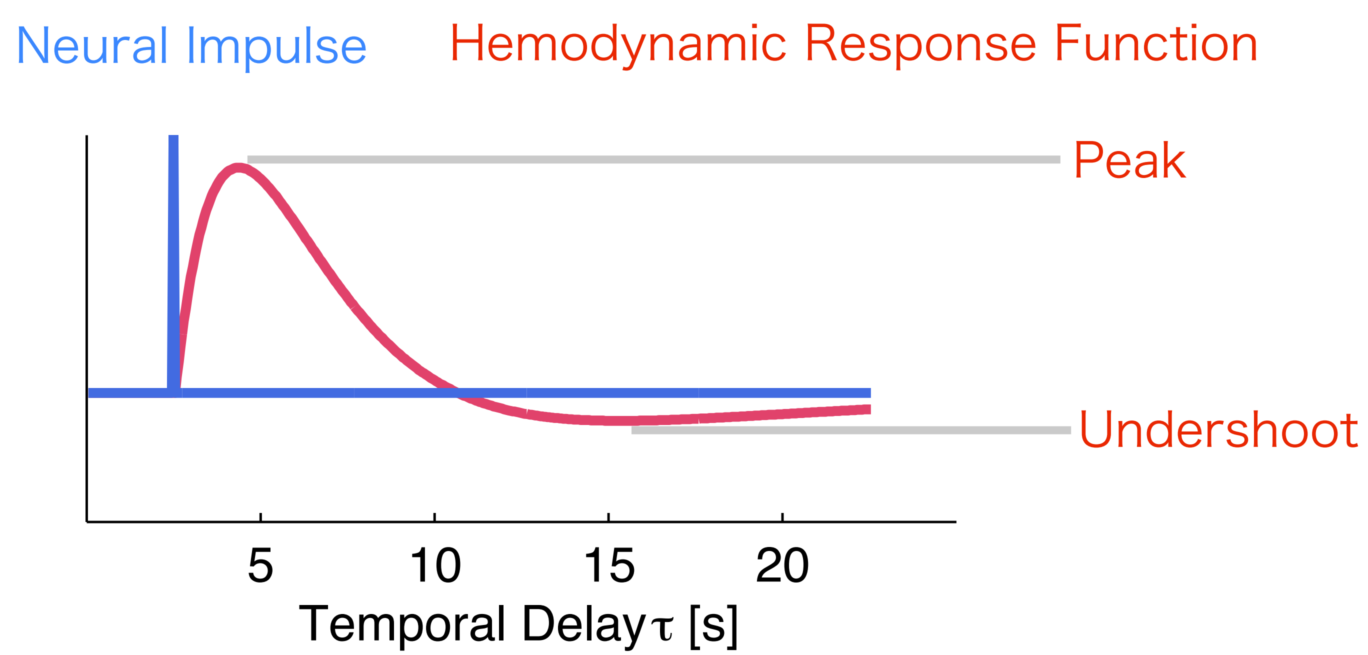


Problems of Standard Kernel CCA

- (K)CCA assumes **instantaneous dependencies**
- Neurovascular coupling is non-instantaneous
- HRF needs to be modeled as convolution

Standard Model of Neurovascular Coupling

Most analyses model temporal dynamics of neurovascular coupling using a **canonical Hemodynamic Response Function** (HRF)



Implications

- Temporal dynamics are the same for all voxels
- Temporal dynamics are **separable** from spatial dynamics

Temporal Kernel CCA

We extend standard kCCA to optimize

$$\underset{\phi_x^\tau, \phi_y}{\operatorname{argmax}} \left(\frac{\sum_\tau (\phi_x^\tau(X))\phi_y(Y)}{\sqrt{\sum_\tau (\phi_x^\tau(X))^\top \sum_\tau (\phi_x^\tau(X)) \ \phi_y(Y)^\top \phi_y(Y)}} \right)$$

where ϕ_x^τ is a temporal convolution in kernel feature space

Bießmann et al, Machine Learning Journal, 2010

Temporal Kernel CCA

$$\underset{w_x(\tau),w_y}{\operatorname{argmax}} \left(\frac{\sum_\tau (w_x(\tau)^\top X_\tau)^\top Y w_y}{\sqrt{\sum_\tau (w_x(\tau)^\top X_\tau X_\tau^\top w_x(\tau)) w_y^\top Y Y^\top w_y}} \right)$$

Temporal Embedding $\quad \tilde{X} = \begin{bmatrix} X_{\tau=-\tau_{\max}} \\ \vdots \\ X_{\tau=\tau_{\max}} \end{bmatrix} \in \mathbb{R}^{F(2\tau_{\max}+1)\times T}$

$$\underset{\tilde{w}_x,w_y}{\operatorname{argmax}} \frac{\tilde{w}_x^\top \tilde{X} Y^\top w_y}{\sqrt{\tilde{w}_x^\top \tilde{X}\tilde{X}^\top \tilde{w}_x w_y^\top Y Y^\top w_y}}$$

Bießmann et al, Machine Learning Journal, 2010

Temporal Kernel CCA

$$\underset{\tilde{w}_x, w_y}{\text{argmax}} \frac{\tilde{w}_x^\top \tilde{X} Y^\top w_y}{\sqrt{\tilde{w}_x^\top \tilde{X} \tilde{X}^\top \tilde{w}_x w_y^\top Y Y^\top w_y}}$$

Kernel Trick

$$\underset{\tilde{w}_x, w_y}{\text{argmax}} \frac{\alpha^\top K_{\tilde{X}} K_Y \beta}{\sqrt{\alpha^\top K_{\tilde{X}}^2 \alpha \cdot \beta^\top K_Y^2 \beta}}$$

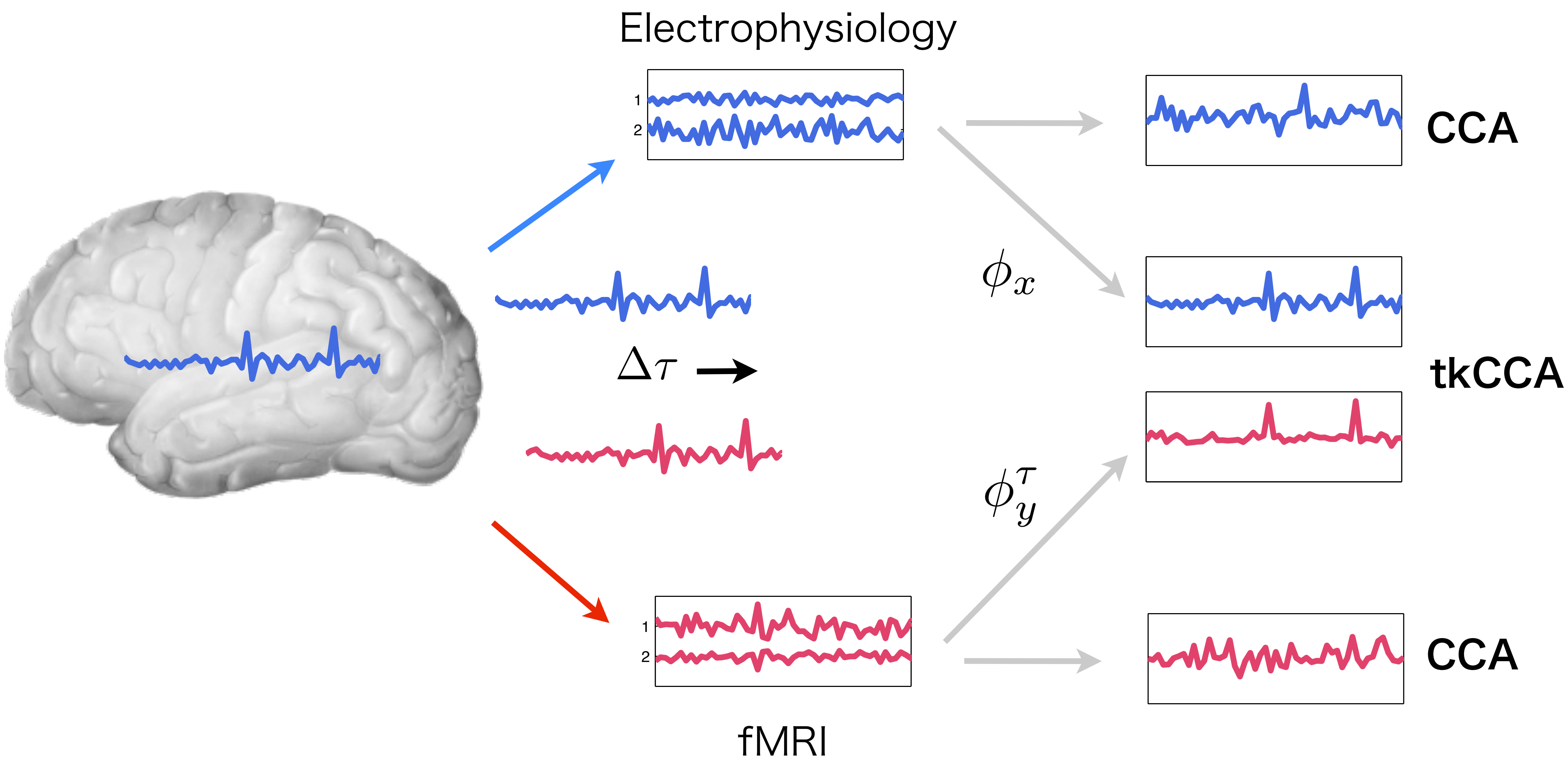
Recover Canonical Convolution

$$\tilde{w}_x = \begin{bmatrix} w_x(\tau = -\tau_{\text{max}}) \\ \vdots \\ w_x(\tau = +\tau_{\text{max}}) \end{bmatrix} = \tilde{X}\alpha$$

$$w_y = Y\beta$$

Bießmann et al, Machine Learning Journal, 2010

Non-instantaneous Coupling: CCA vs. tkCCA

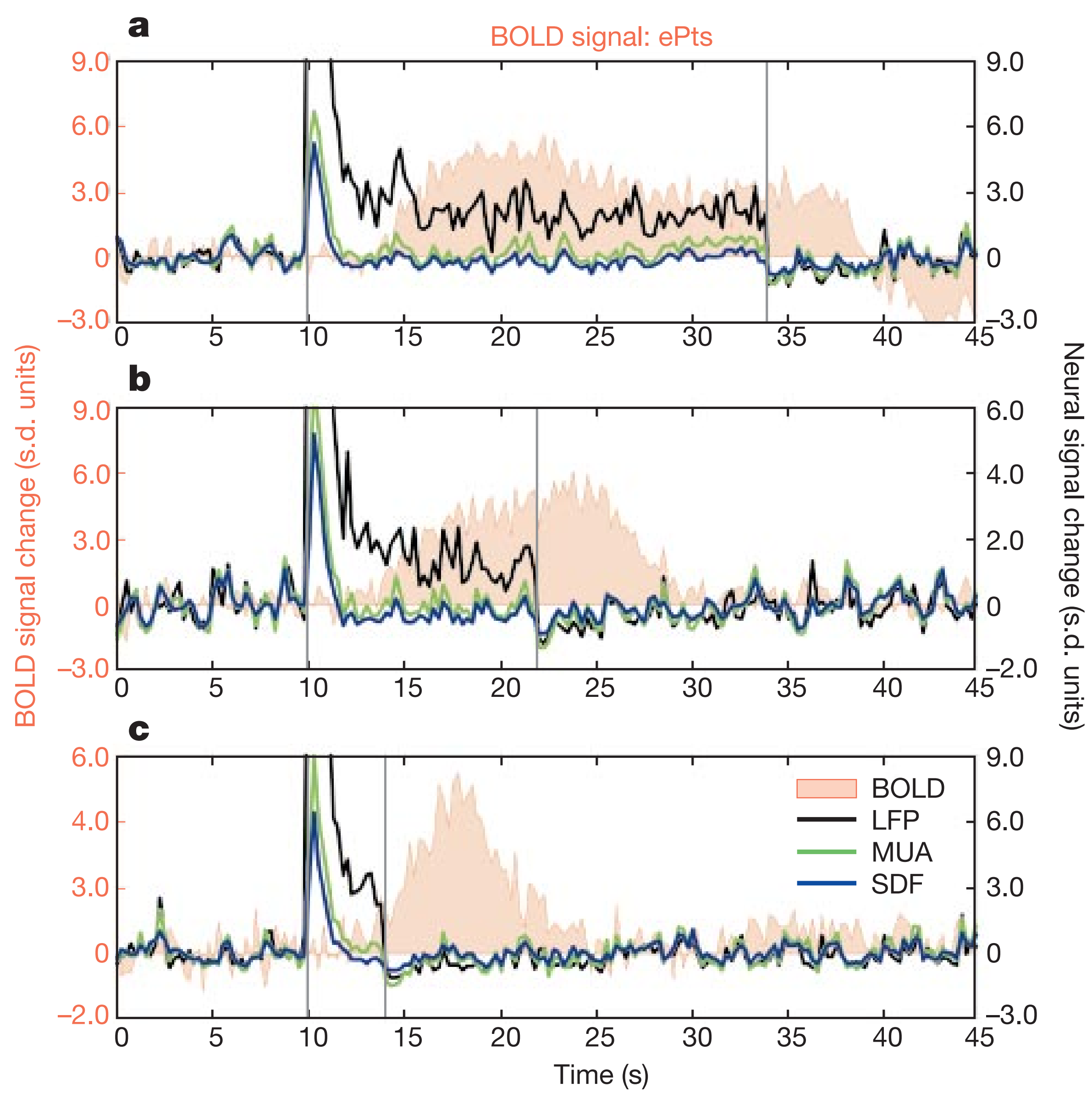


- For non-instantaneous coupling standard CCA fails
- tkCCA recovers the coupling between high dimensional modalities
- ➡ Now we can learn fMRI spatiotemporal dynamics from small data sets

Multimodal Unsupervised EEG-fMRI Source Estimation

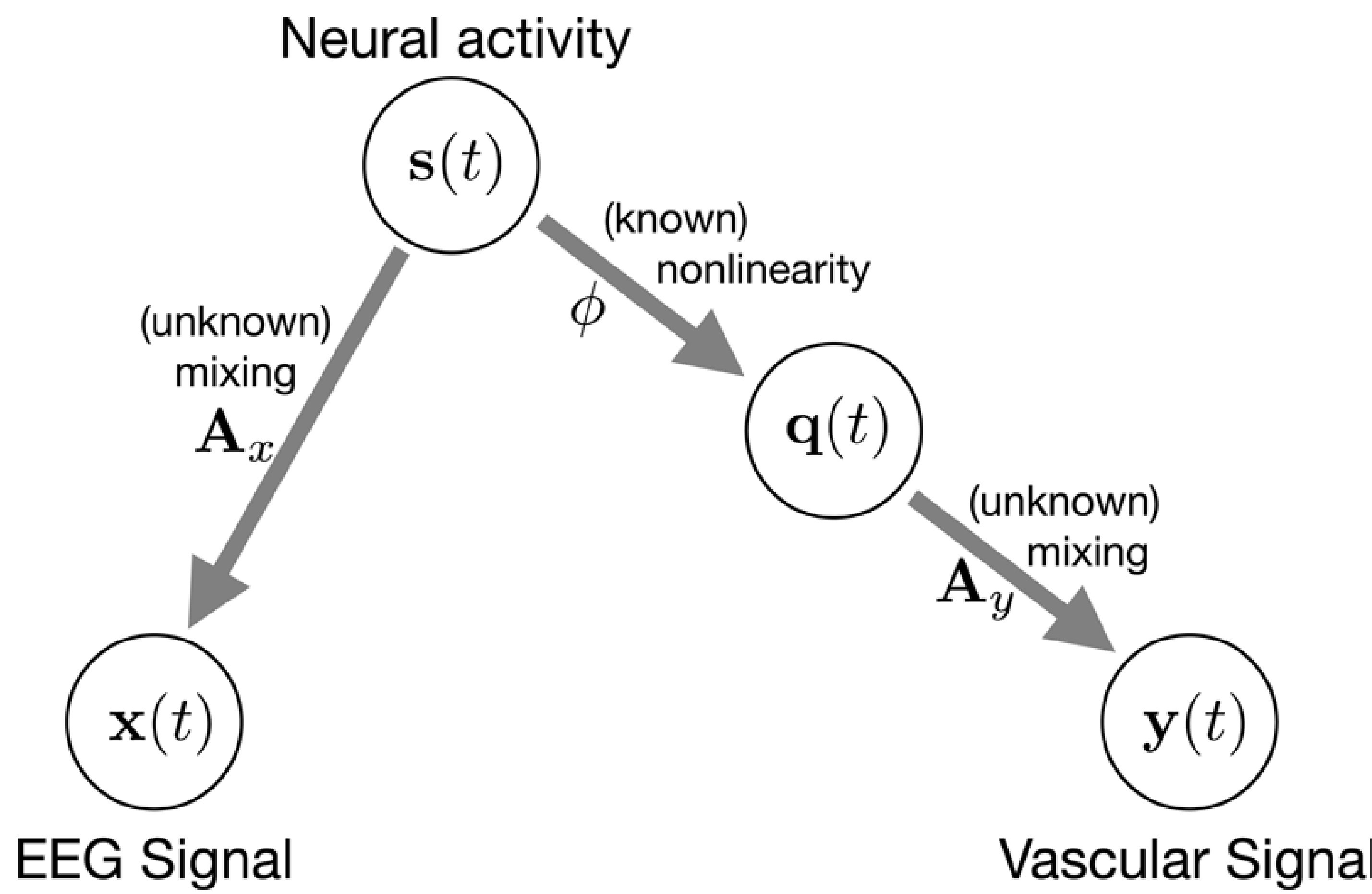
Multimodal Unsupervised EEG-fMRI Source Estimation

fMRI reflects neural bandpower



Logothetis et al., Nature, 2001

- Computing neural bandpower is a nonlinear function



- Knowing this nonlinearity, we can model it explicitly

Daehne et al., IEEE Trans. Multimedia, 2013

- Computing neural bandpower is a nonlinear function
- EEG model is linear $$x_m(t) = A_m s(t) + \epsilon$$
- Computing bandpower in sensor space is problematic
  - Superposition of multiple sources
  - Superposition of neural signals and noise
  - Linear De-mixing cannot undo the nonlinearity

Daehne et al., IEEE Trans. Multimedia, 2013

- Problematic: EEG **sensor bandpower**
  - As regressor for SPM GLM
  - As variable in CCA / Partial Least Squares
  - As data in multivariate regression / classification
- Less problematic: Neural **source bandpower**
  - in Blind Source Separation space
  - of fitted dipole source space
- OR: Explicitly model biophysics of coupling

Daehne et al., IEEE Trans. Multimedia, 2013

Multimodal Source Power Correlation Analysis

Linear Map — Data — Sources

$$\mathbf{w}_{\mathbf{x}}^{\top}\mathbf{x} = \hat{s}_x$$
$$\mathbf{w}_{\mathbf{y}}^{\top}\mathbf{y} = \hat{s}_y$$

Nonlinearity:
Bandpower Computation

$$\hat{p}_{s_x}(e) = \left\langle \left(\mathbf{w}_{\mathbf{x}}^{\top}\mathbf{x}(t)\right)^2 \right\rangle_{t \in T_e} = \mathbf{w}_{\mathbf{x}}^{\top}\mathbf{C}_{xx}(e)\mathbf{w}_{\mathbf{x}}.$$

Temporal Convolution of Bandpower in Source Space

$$\hat{h}(\hat{p}_{s_x})(e) = \sum_{n}^{N_\tau} \mathbf{w}_{\tau n}\hat{p}_{s_x}(e-n)$$

Daehne et al., IEEE Trans. Multimedia, 2013

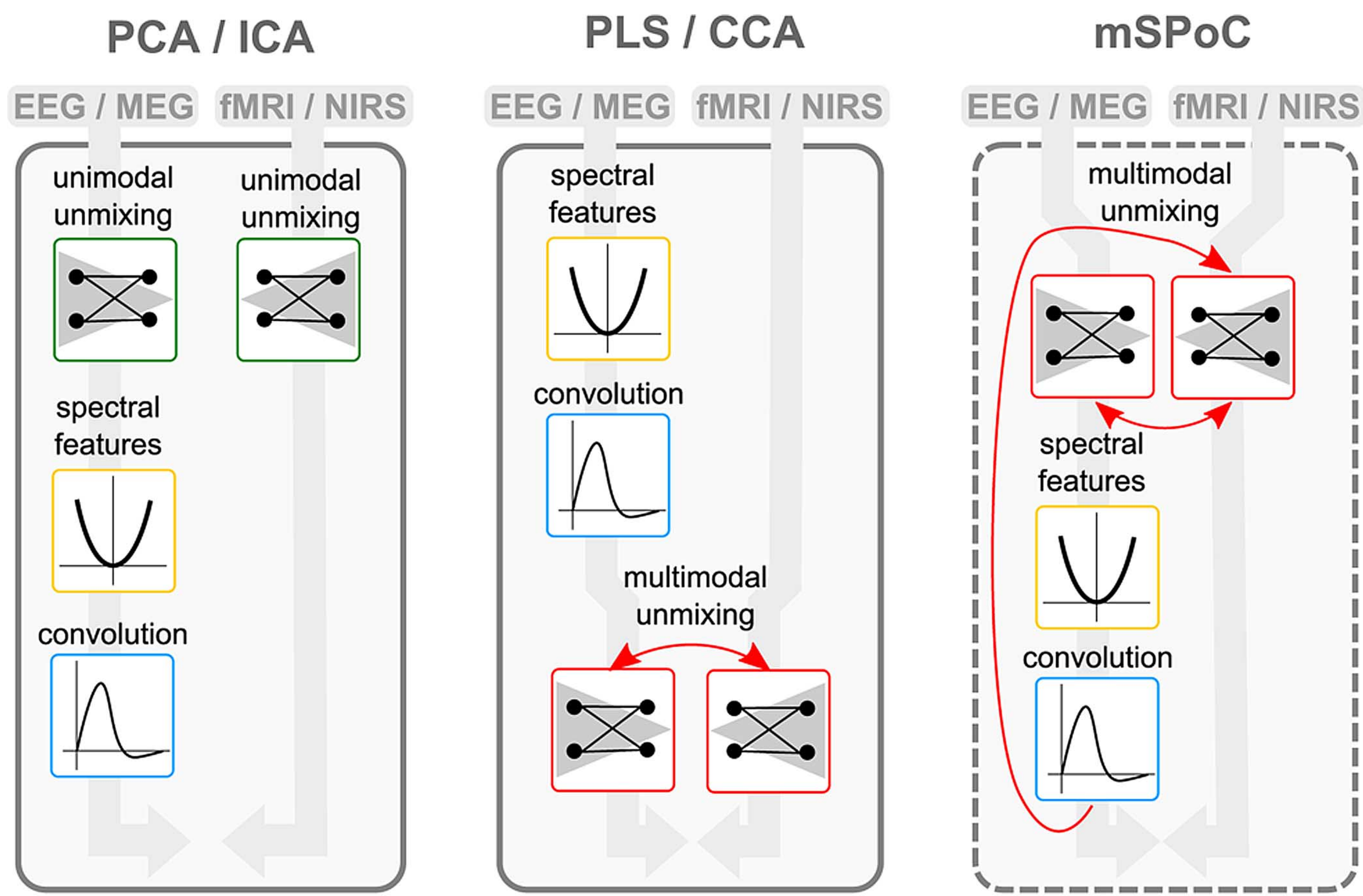
Objective of mSPoC

$$f_{\mathrm{obj}}(\mathbf{w}_{\mathbf{x}}, \mathbf{w}_{\mathbf{y}}, \mathbf{w}_{\tau}) := \mathrm{Cov}\left(\hat{h}(\hat{p}_{s_x}), \hat{s}_y\right)$$

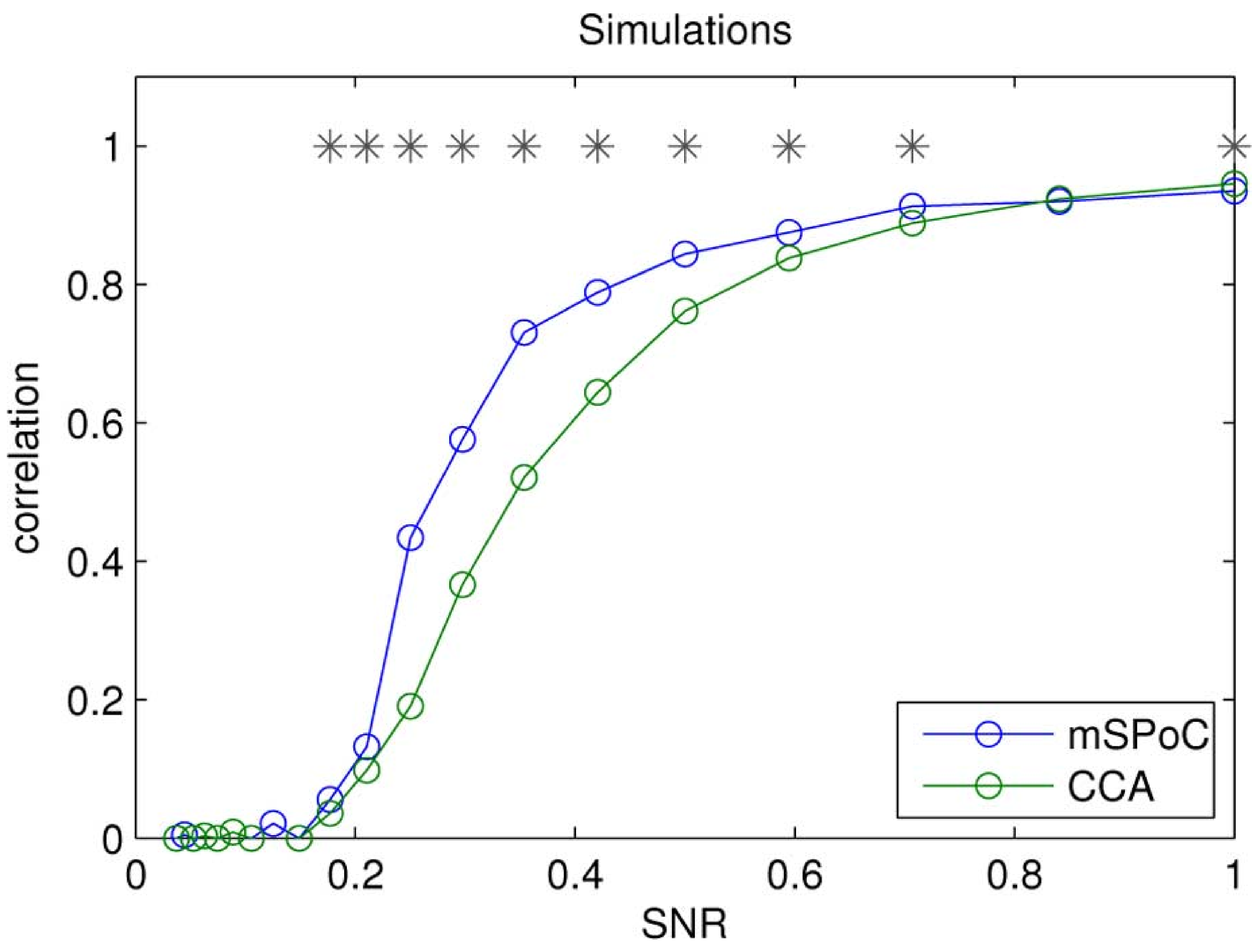
subject to the constraints

$$\begin{aligned}
\|\mathbf{w}_{\mathbf{x}}\|_{\mathbf{C}_{\mathbf{xx}}} &:= \mathbf{w}_{\mathbf{x}}^{\top}\mathbf{C}_{\mathbf{xx}}\mathbf{w}_{\mathbf{x}} = 1, \\
\|\mathbf{w}_{\mathbf{y}}\|_{\mathbf{C}_{\mathbf{yy}}} &:= \mathbf{w}_{\mathbf{y}}^{\top}\mathbf{C}_{\mathbf{yy}}\mathbf{w}_{\mathbf{y}} = 1, \\
\|\mathbf{w}_{\tau}\|_{\mathbf{B}} &:= \mathbf{w}_{\tau}^{\top}\mathbf{B}\mathbf{w}_{\tau} = 1,
\end{aligned}$$

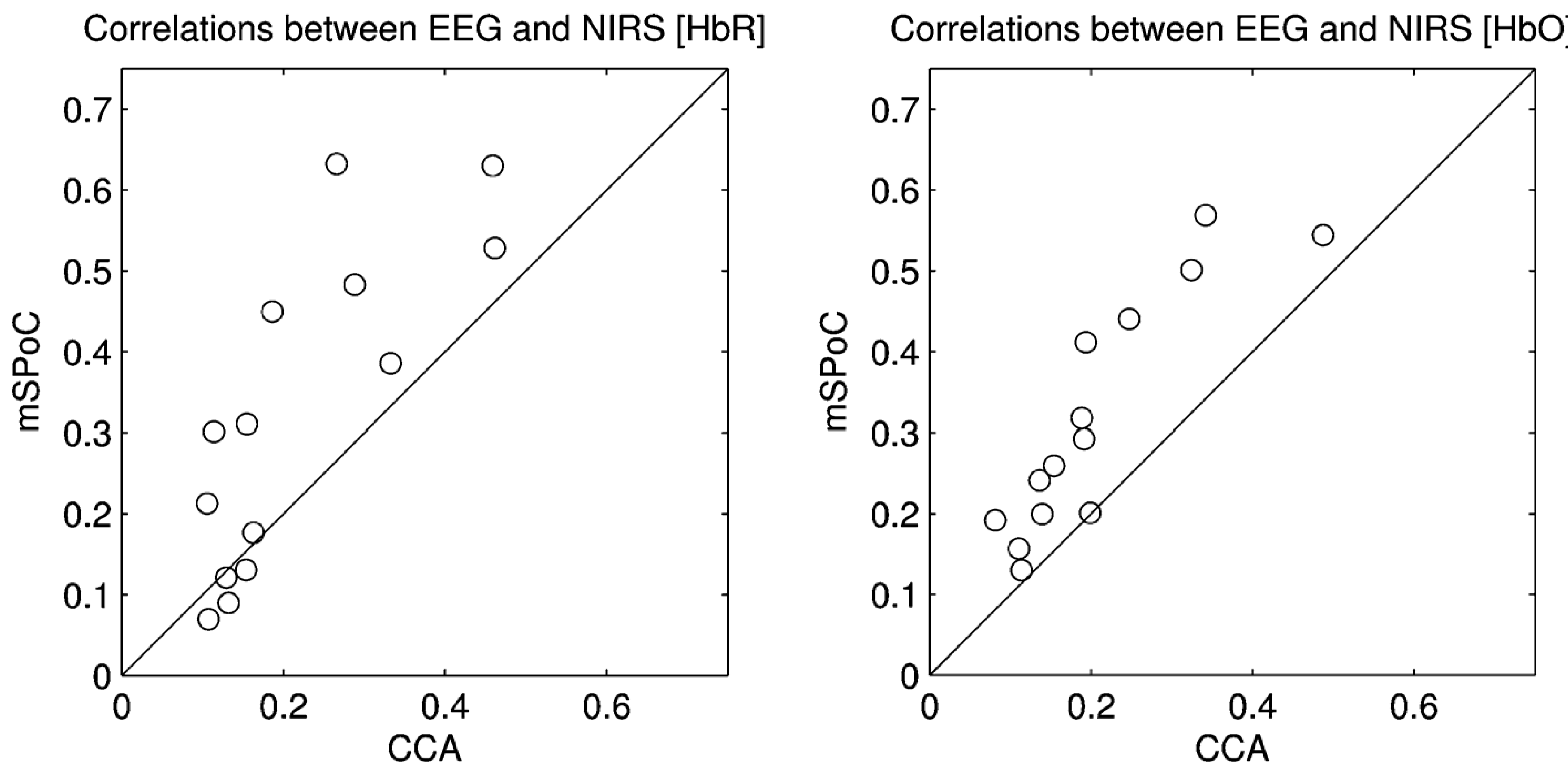
Multimodal Source Power Correlation Analysis



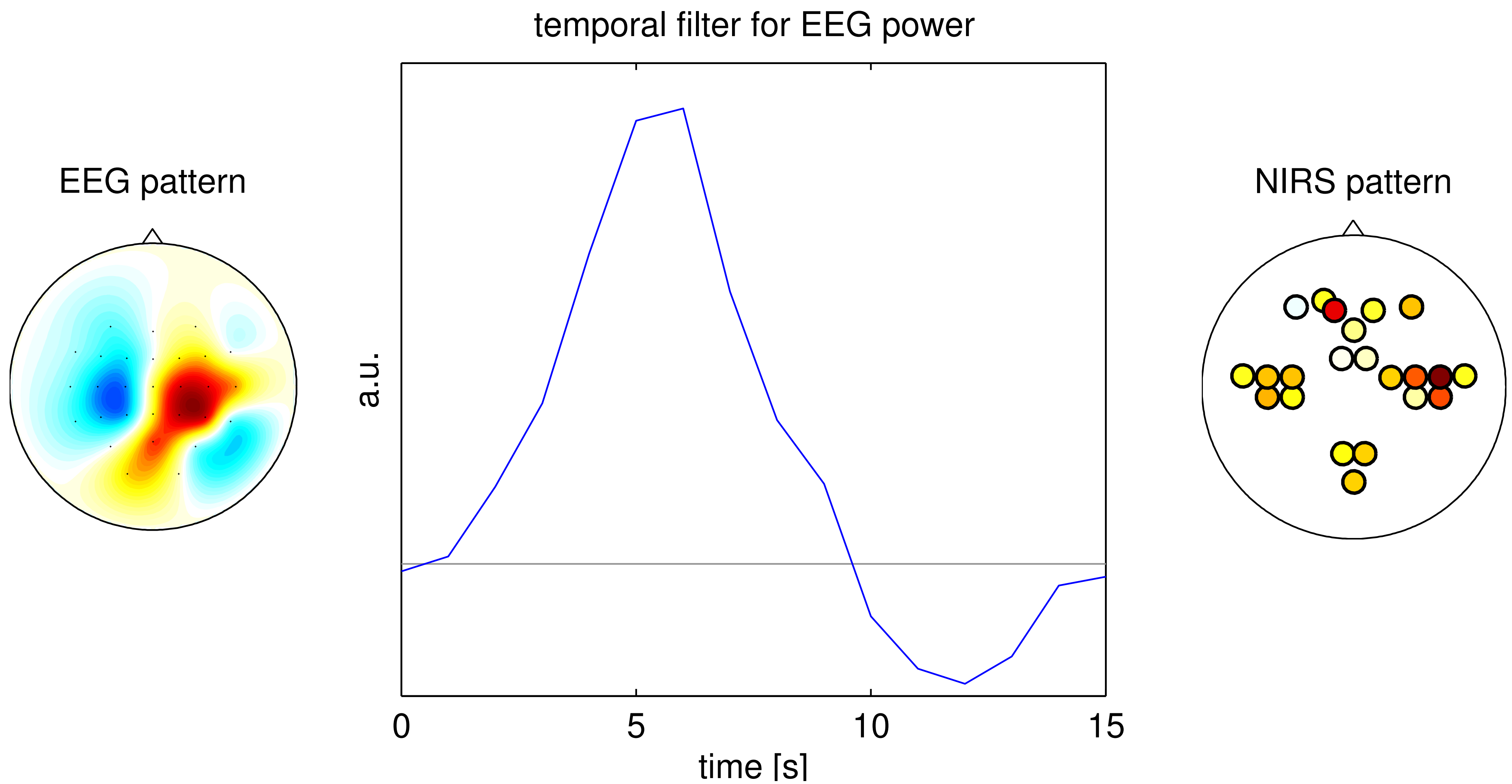
Multimodal Source Power Correlation Analysis



Multimodal Source Power Correlation Analysis



Multimodal Source Power Correlation Analysis

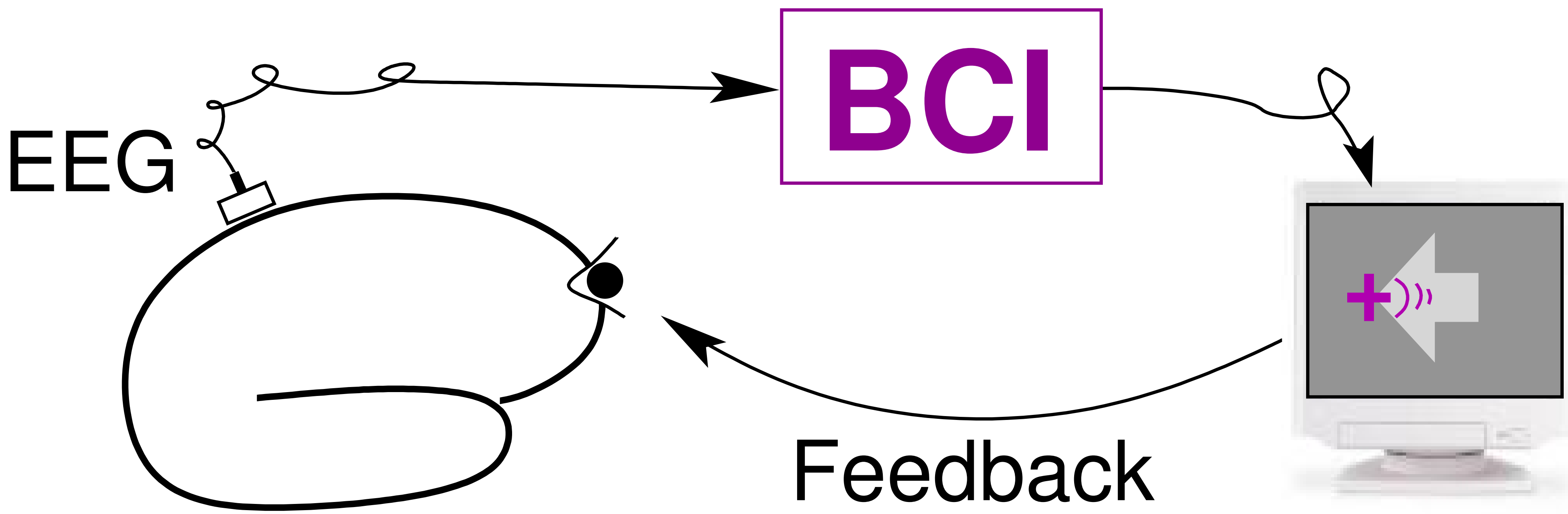


Applications

Applications

- Hybrid BCIs: Combining EEG and NIRS
  - Do multimodal setups increase information transfer rates?
- Cleaning artifacts in multimodal recordings
  - PCA: simple but efficient
- Decoding neural bandpower from fMRI
  - Complex spatiotemporal filters for optimal decoding
- Multisubject Analyses
  - Applications of multimodal methods to hyperscanning

Hybrid BCIs Improve Mental State Detection


EEG
BCI
Feedback

Hybrid BCIs Improve Mental State Detection

EEG and NIRS carry complementary information

Do multimodal BCIs improve information transfer?

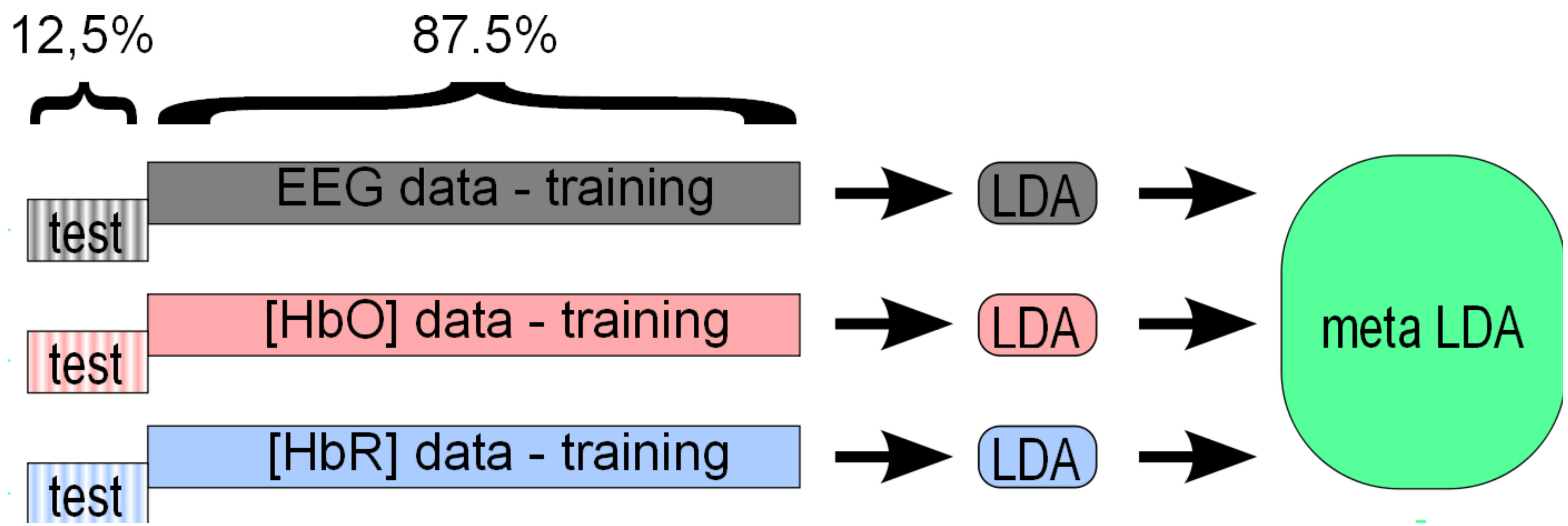
Hybrid BCIs Improve Mental State Detection

Linear Discriminant Analysis (LDA) trained on each modality

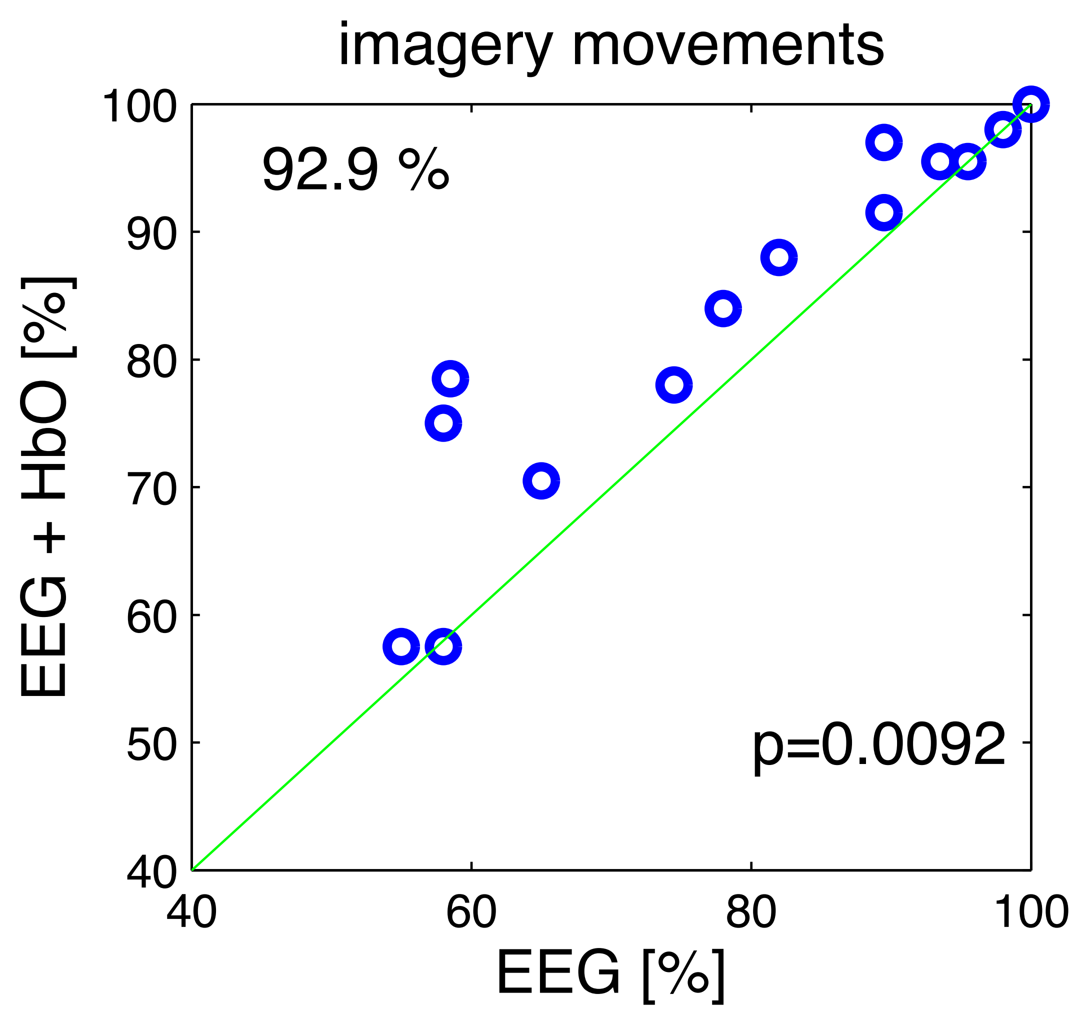
Meta-LDA classifier trained on outputs of single modalities

$$w_{LDA} \propto (X_+ X_+^\top + X_- X_-^\top)^{-1} (\mu_+ - \mu_-)$$

Crossvalidated Analysis Workflow



Hybrid BCIs improve Mental State Detection



Fazli et al., Neuroimage, 2011

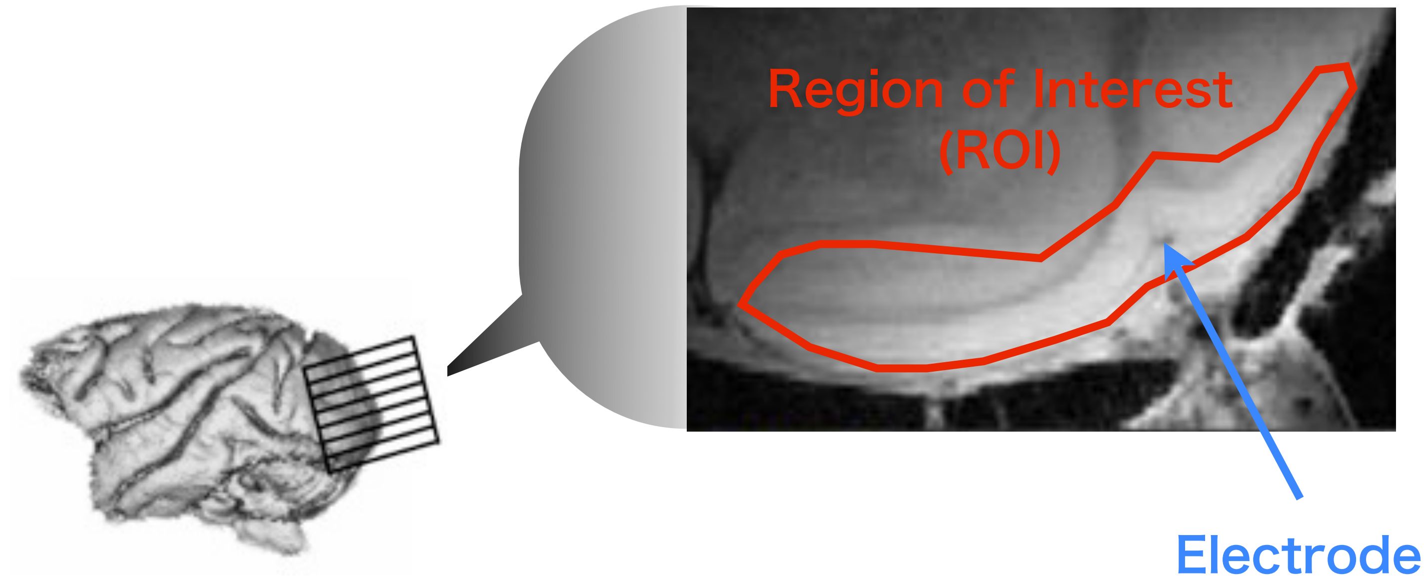
Summary

- EEG and NIRS carry complementary information
- Combining these two modalities leads to improved BCI performances in over 90% of the subjects tested

PCA For Artifact Removal in Multimodal Recordings

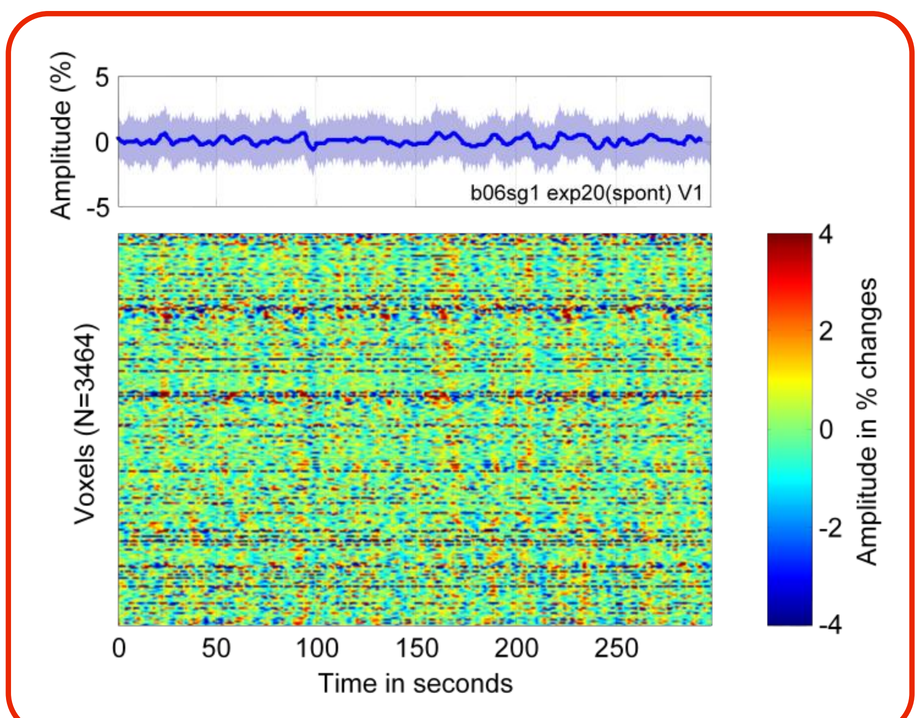
Simultaneous Neural and Hemodynamic Measurements

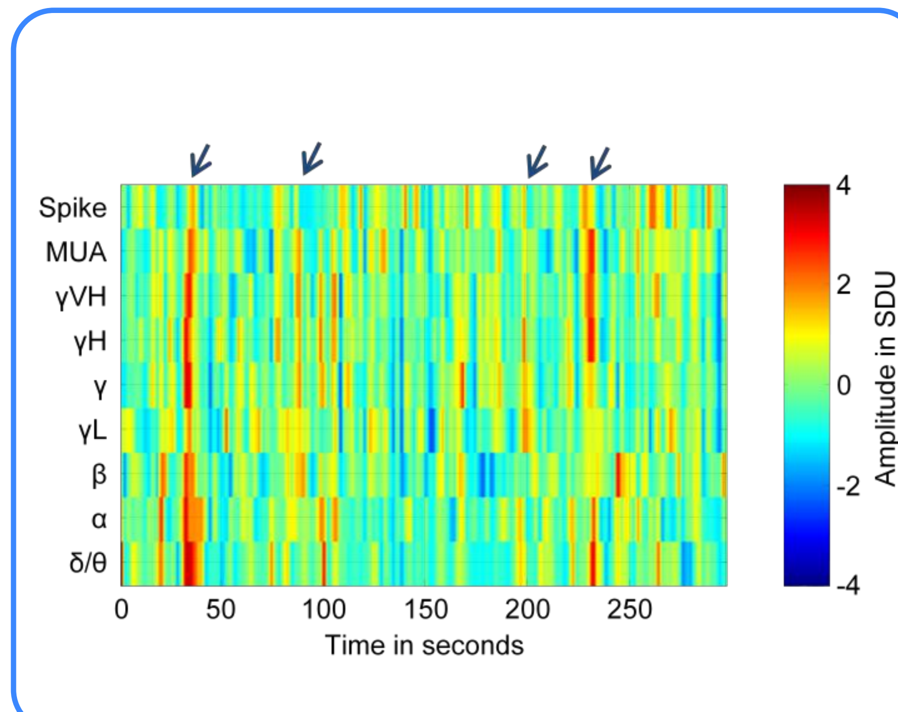
Spontaneous activity in primary visual cortex of the non-human primate



fMRI Y

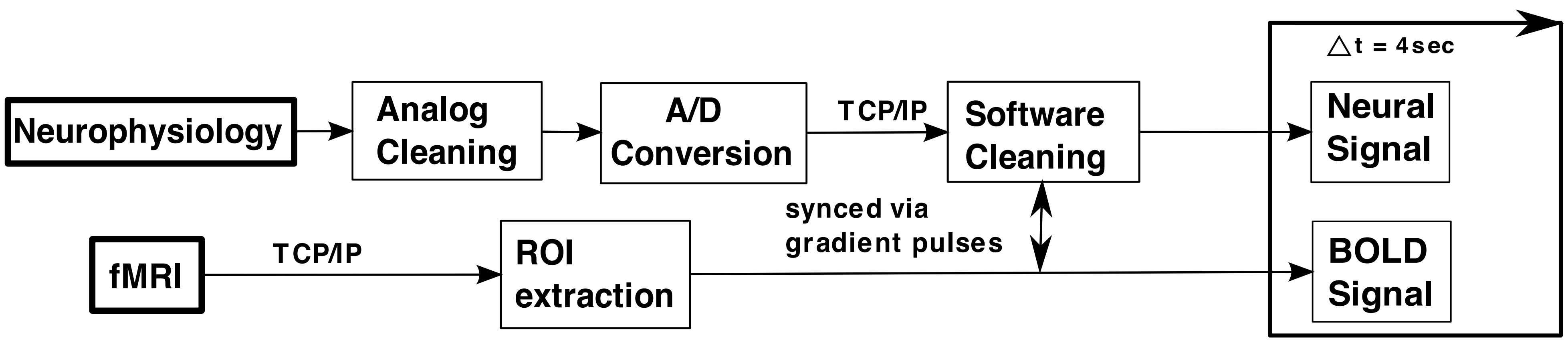
X Neural Bandpower





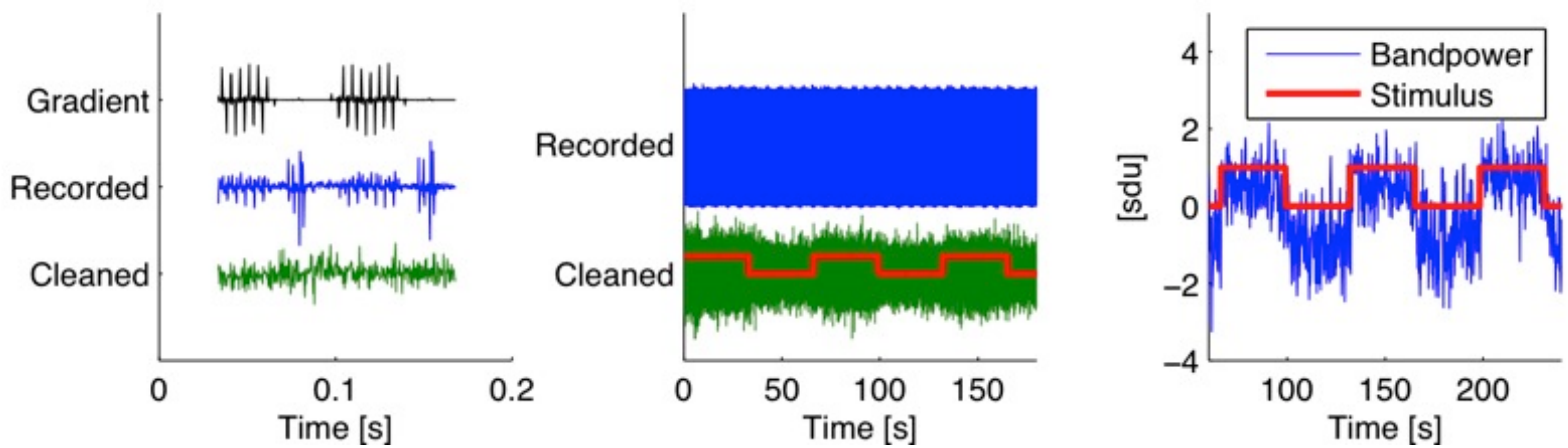
Online Recording System

Integrated recording and artifact removal system

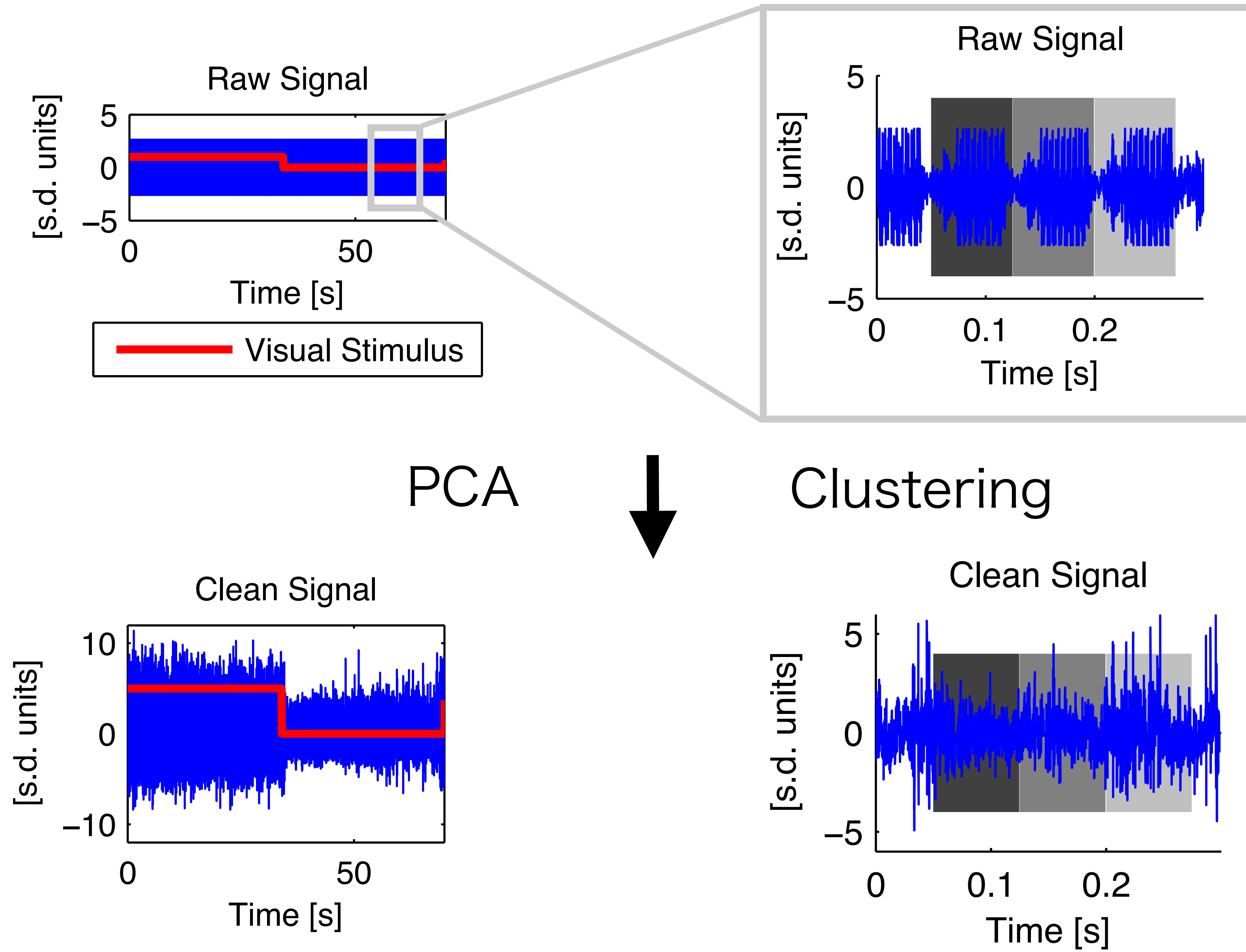


Scanner gradients can be used for signal synchronization

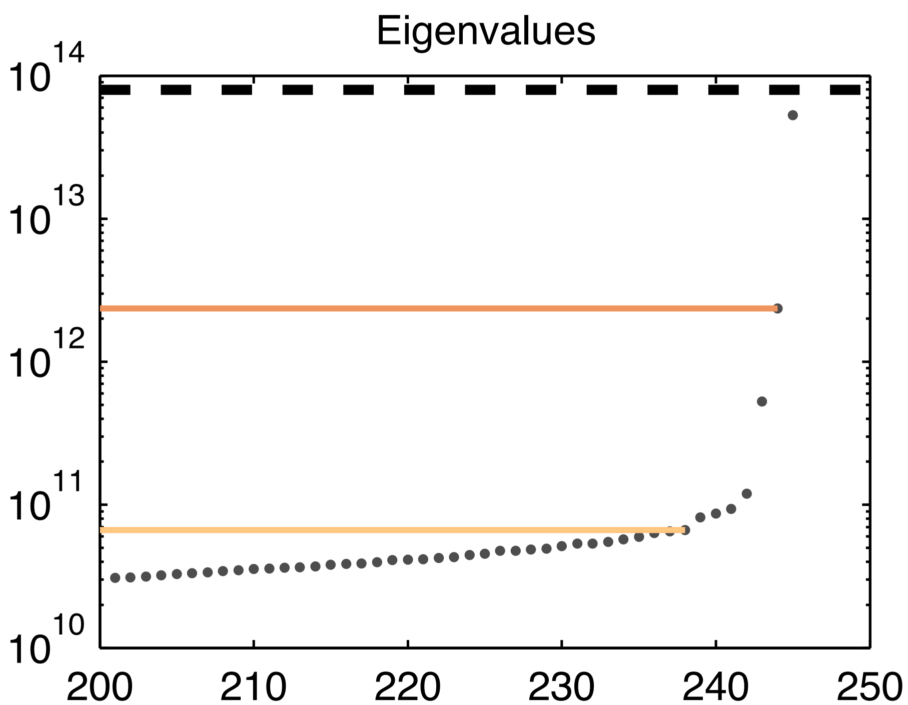
Scanner Gradient Artifact Removal

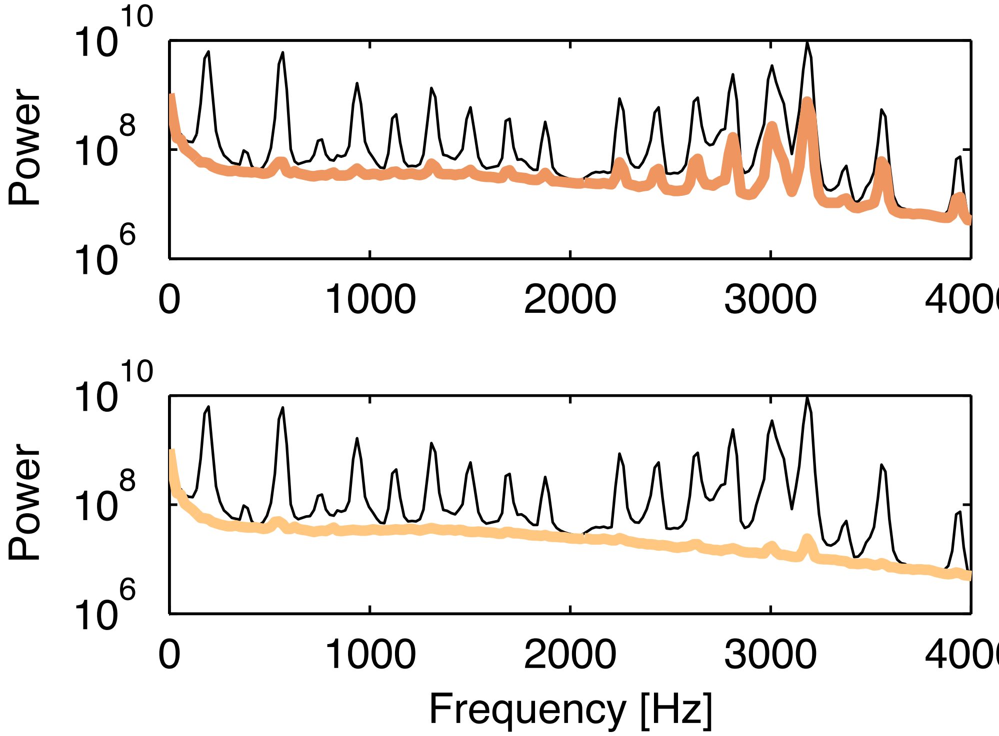


Scanner Gradient Artifact Removal



Empirical Criteria for Artifact Removal





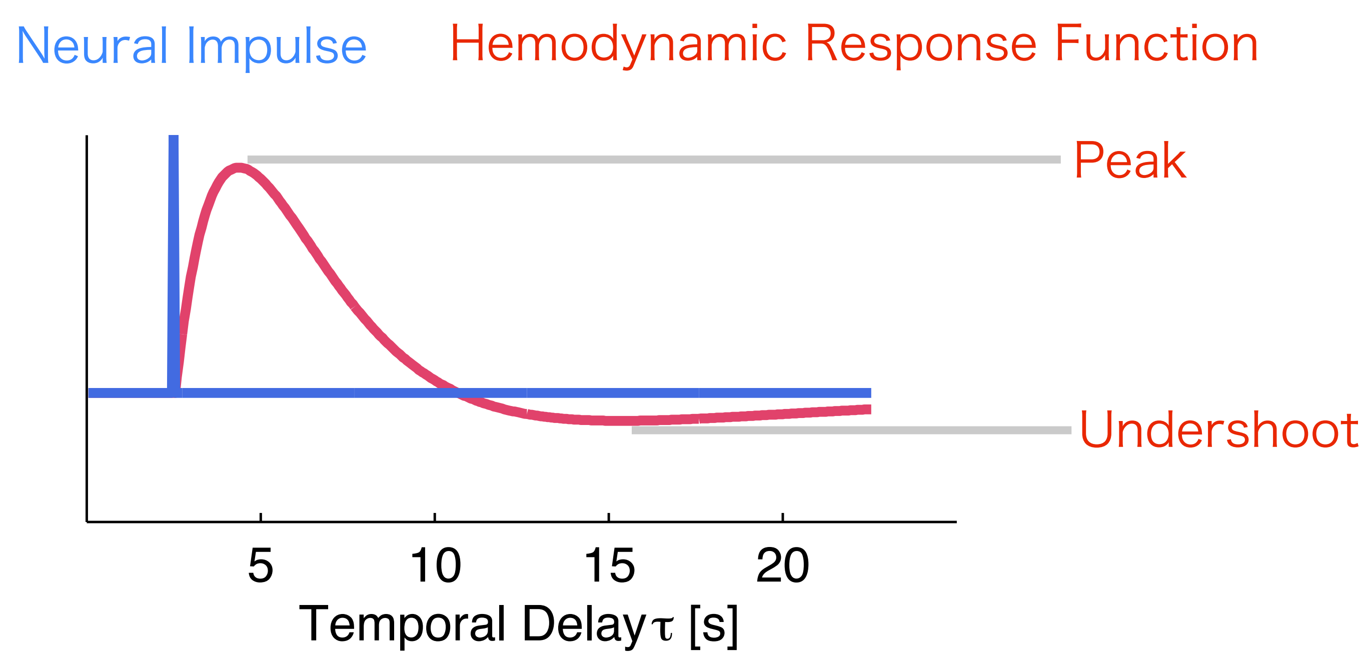
Summary

- Multimodal recordings suffer from artifacts
- Modeling the artifact is difficult
- Reasonable Assumption: artifact has large variance
- Use PCA to remove large variance components

Decoding Neural Information from fMRI Signals

Standard Model of Neurovascular Coupling

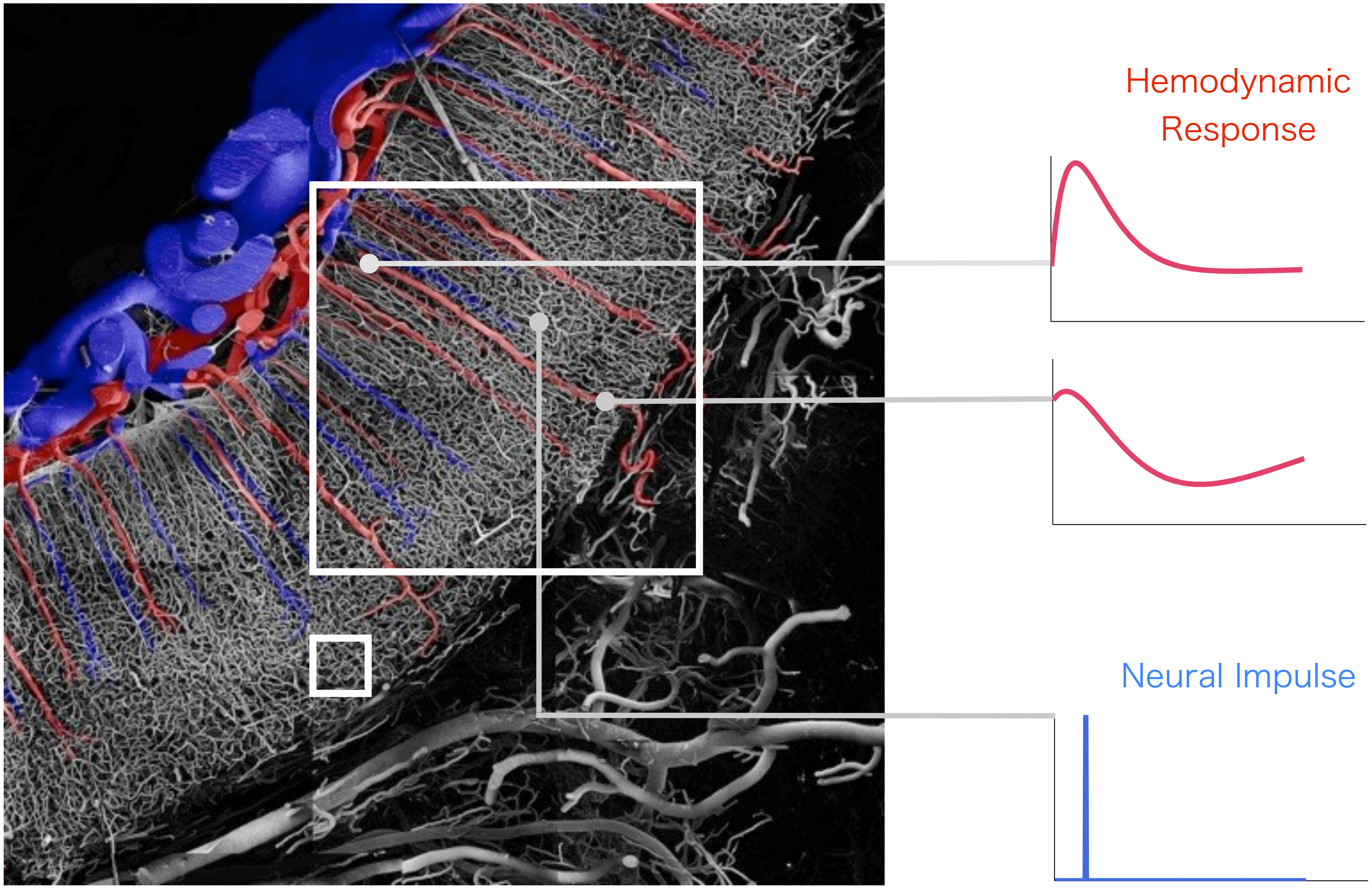
Most analyses model temporal dynamics of neurovascular coupling using a **canonical Hemodynamic Response Function** (HRF)



Implications

- Temporal dynamics are the same for all voxels
- Temporal dynamics are **separable** from spatial dynamics

When Canonical HRF Models Fail



Blood Vessels in Macaque Visual Cortex
(with kind permission of A.-L. Keller, MPI Tübingen)

Is Spatiotemporal HRF Variability Neural Information?

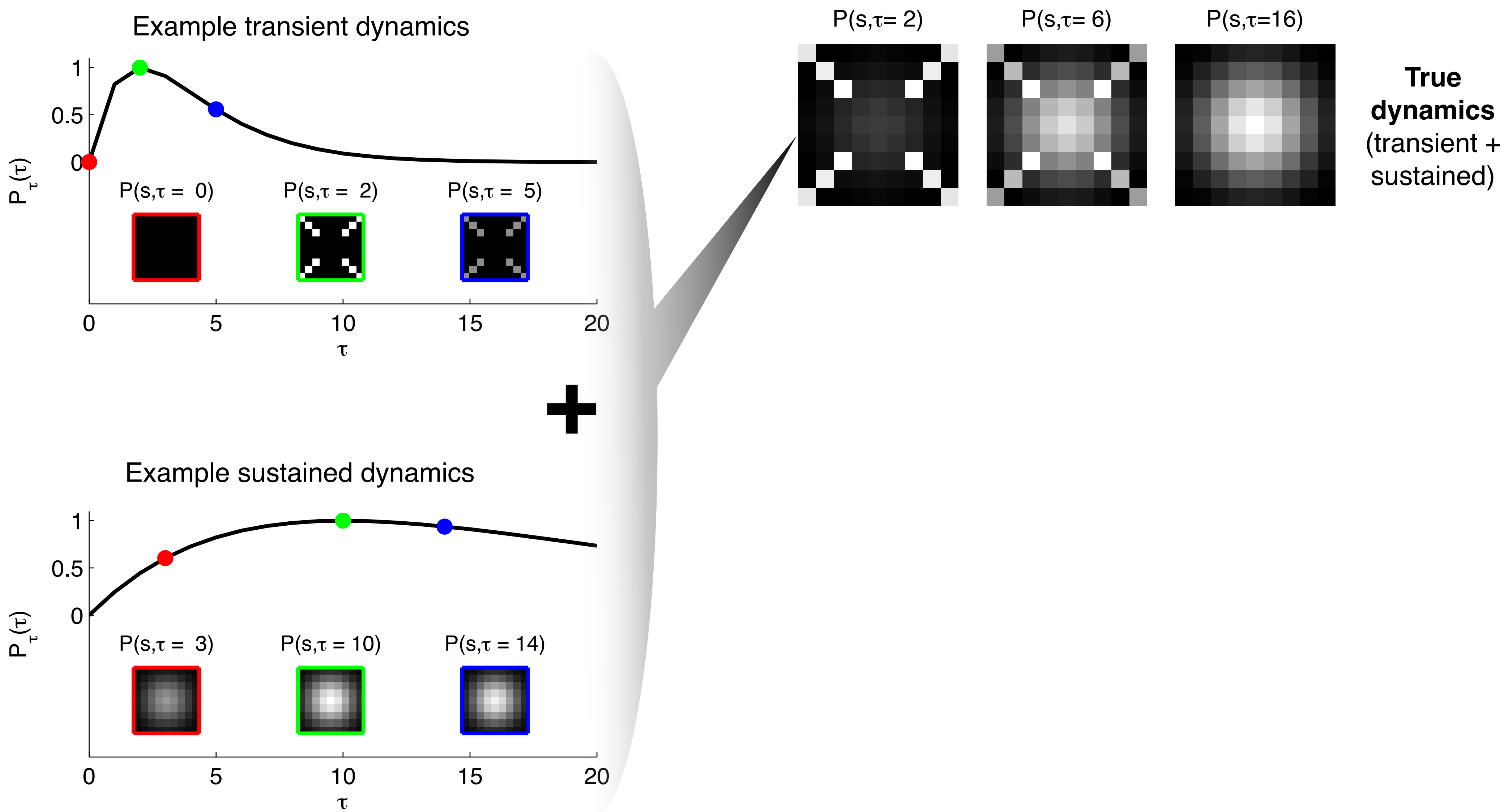
If tkCCA predicts neural signals better

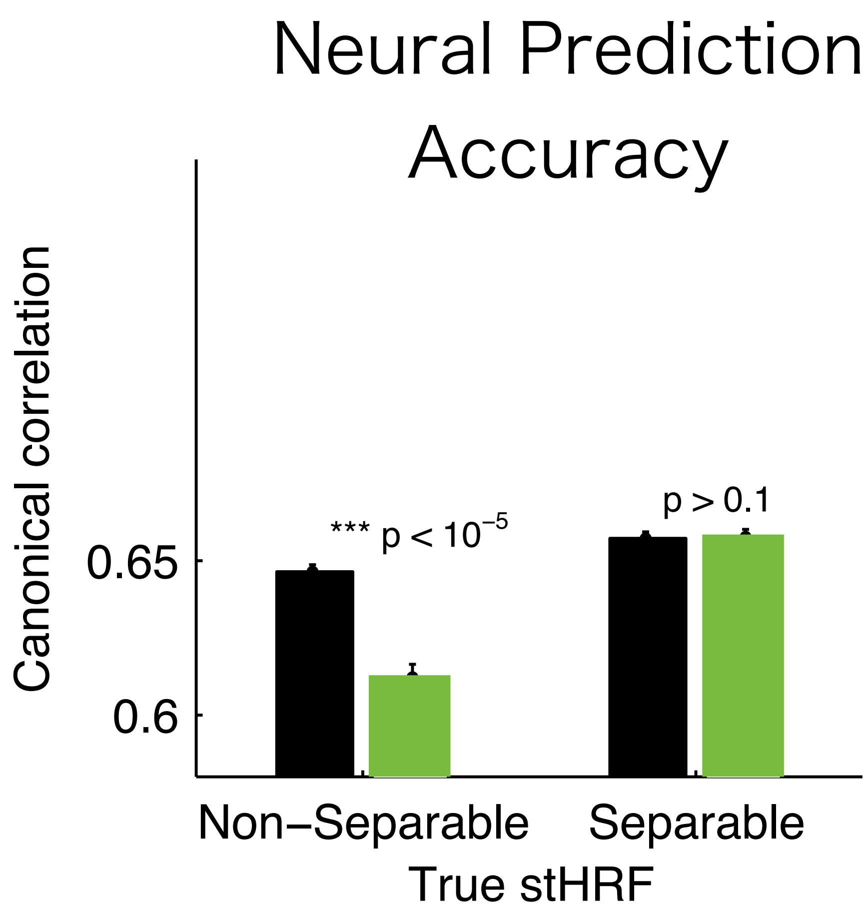
than canonical HRF (i.e. separable) models

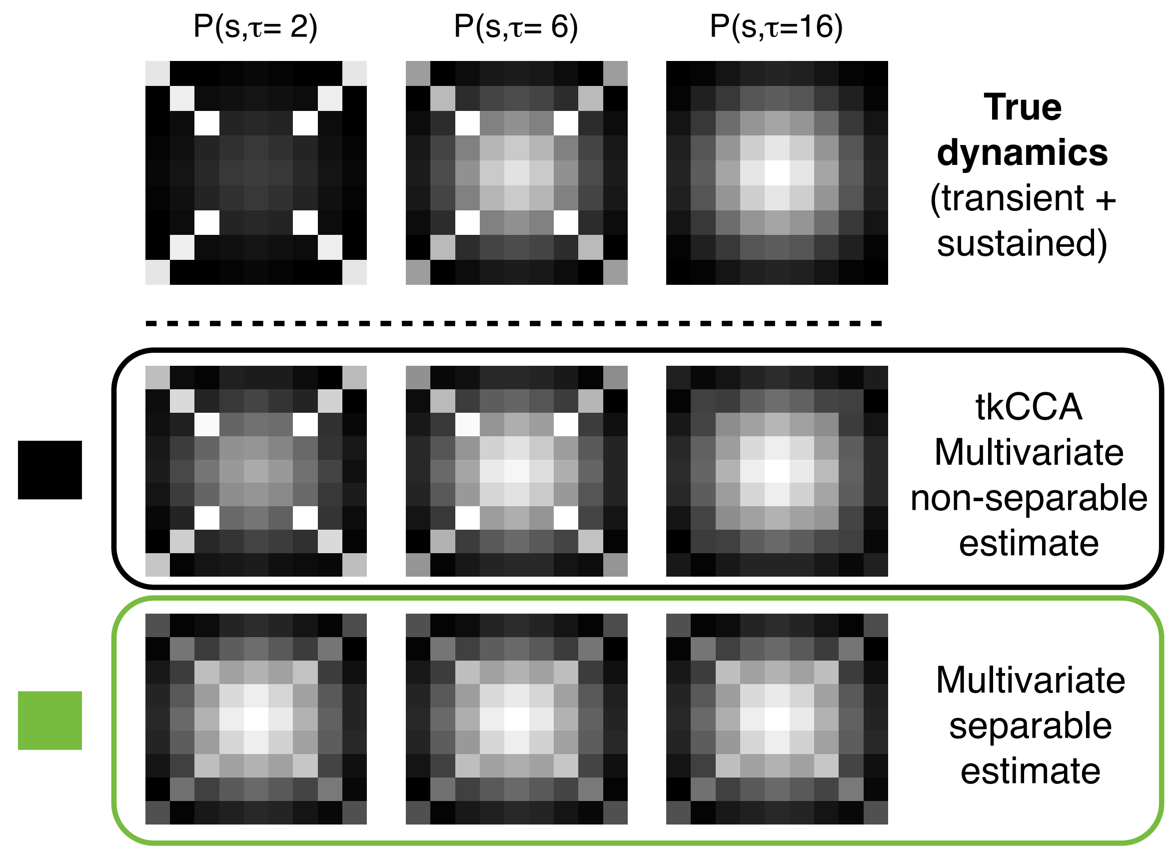
➡ **then** deviations from the canonical HRF model

carry neural information

Non-separable and separable HRFs





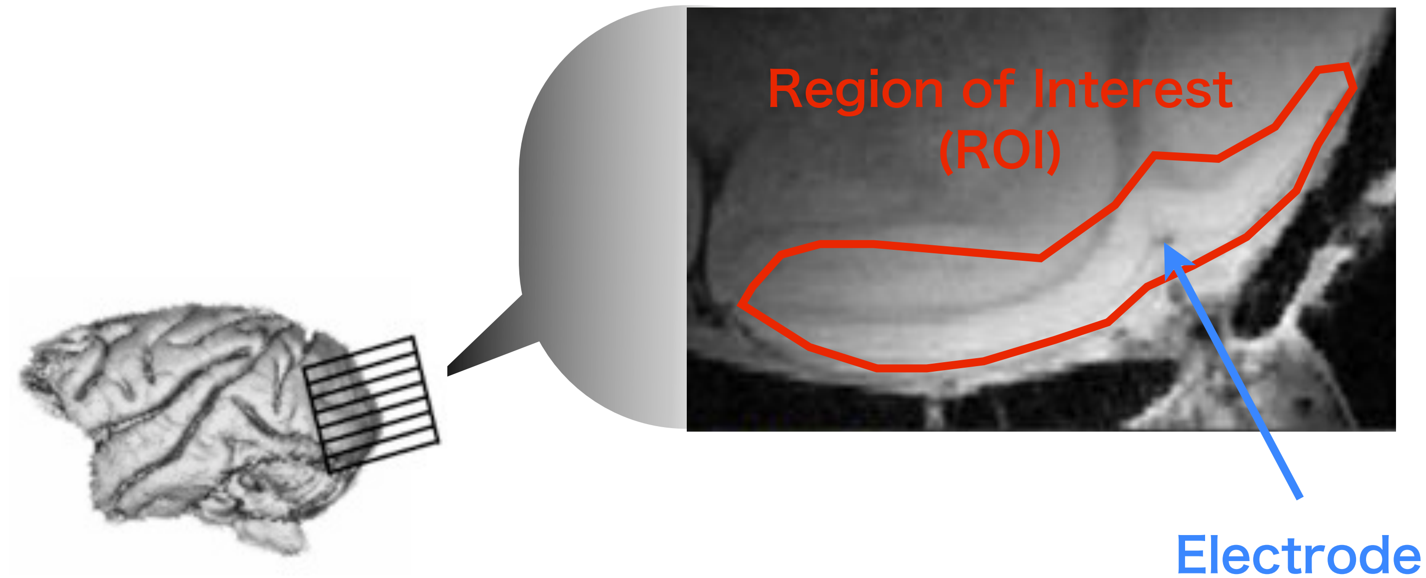


If the HRF is space-time non-separable

➡ canonical HRF models miss neural information

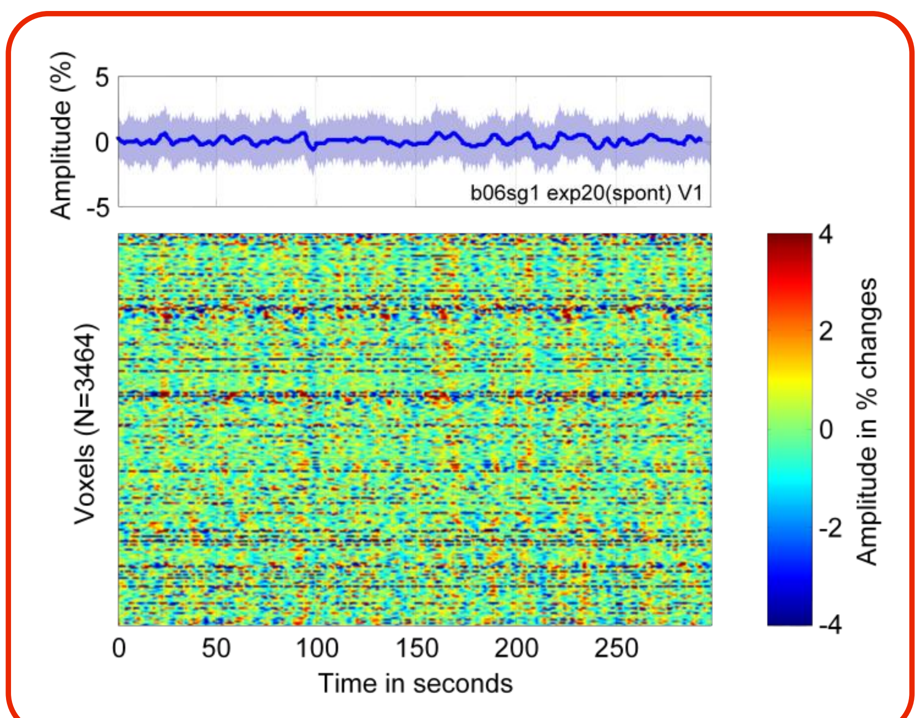
Decoding Neural Information from fMRI Signals

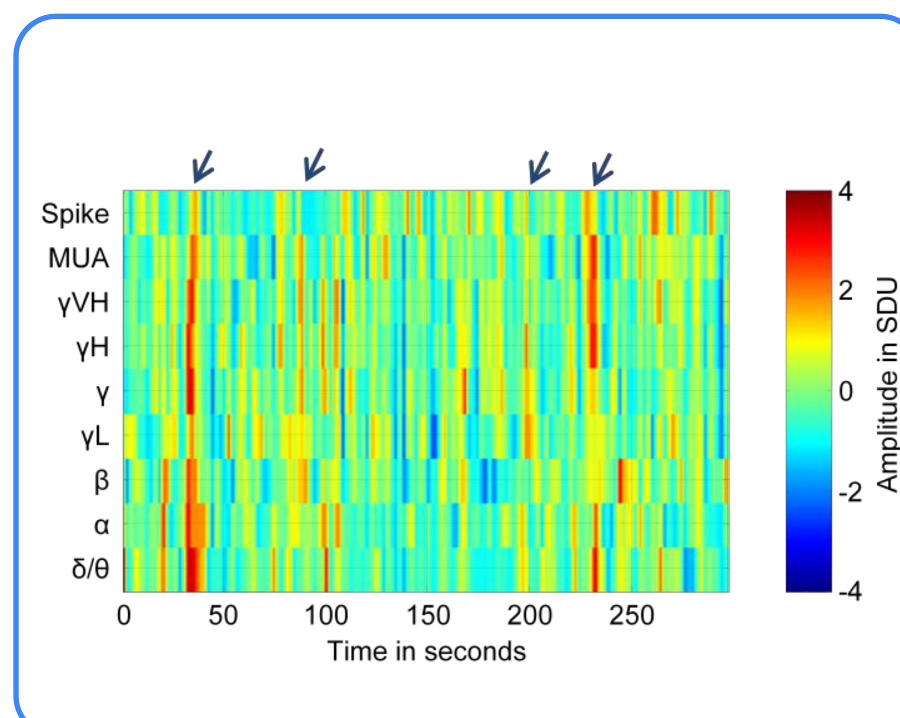
Spontaneous activity in primary visual cortex of the non-human primate



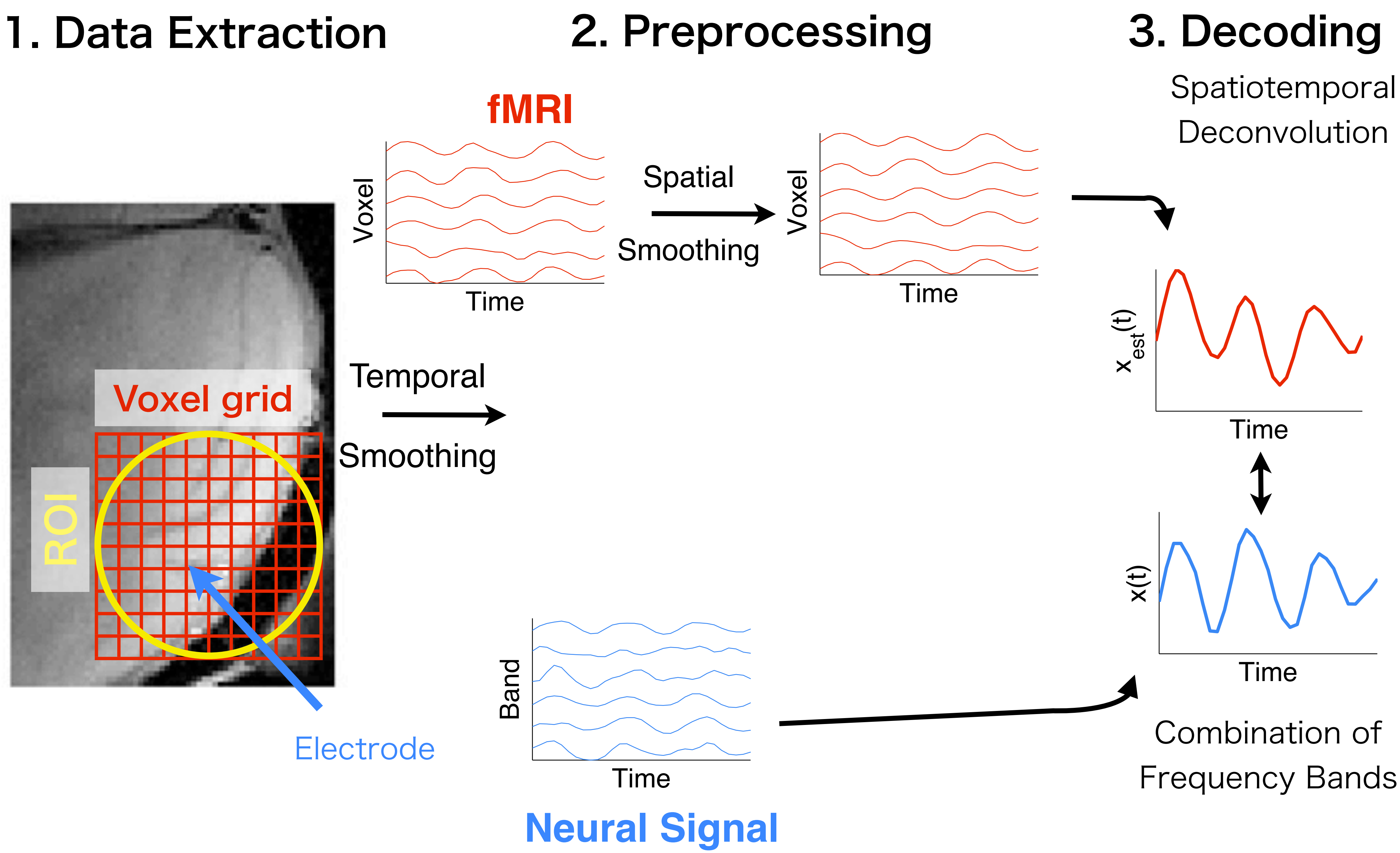
fMRI Y

X Neural Bandpower

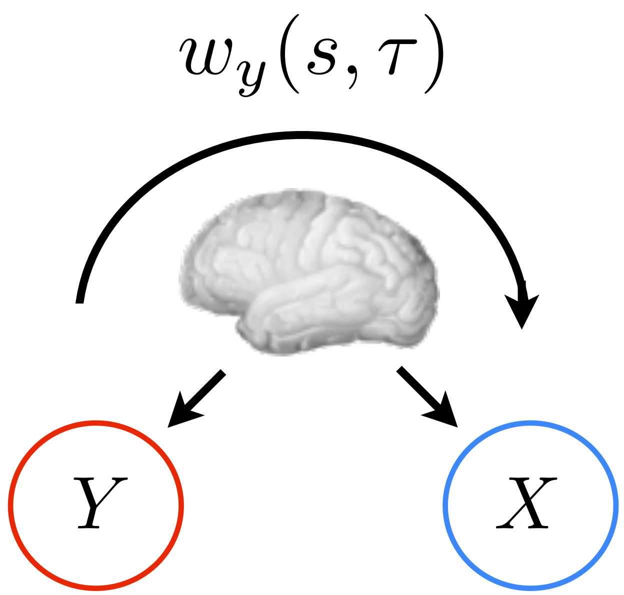




Decoding Neural Information: Workflow



Predicting Neural Amplitude from fMRI



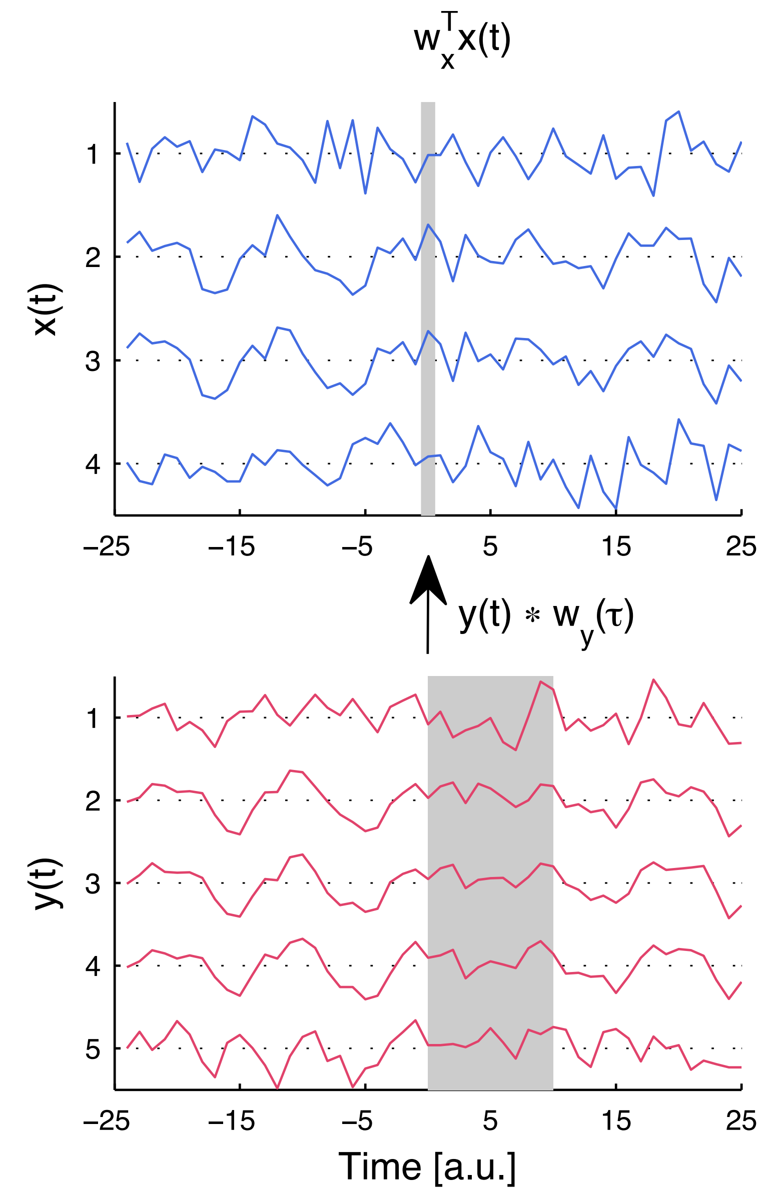
Measured Neural Signal

$$x(t) = \sum_f w_x(f) X(f, t)$$

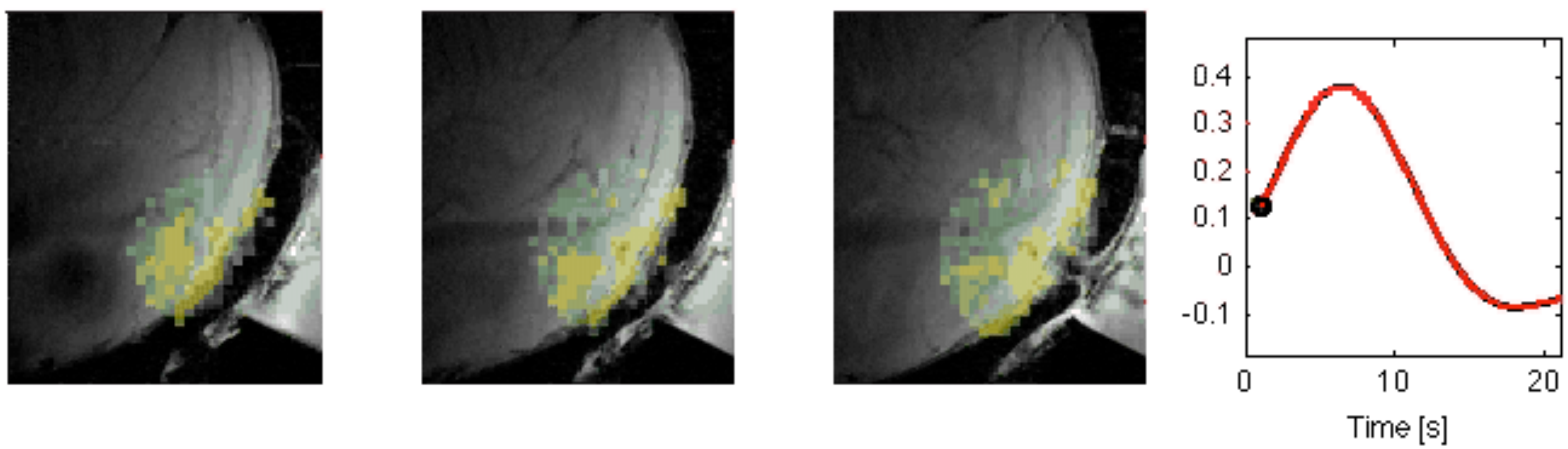
Estimated Neural Signal

$$\hat{x}(t) = \sum_s \sum_\tau w_y(s, \tau) Y(s, t + \tau)$$

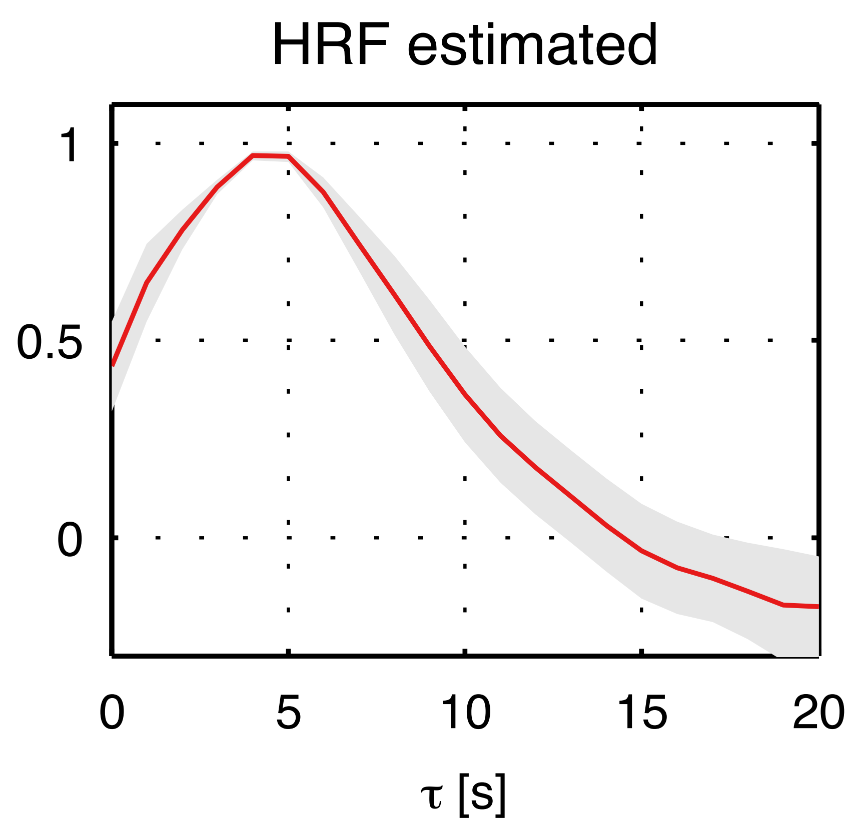
Predicting neural signals from fMRI

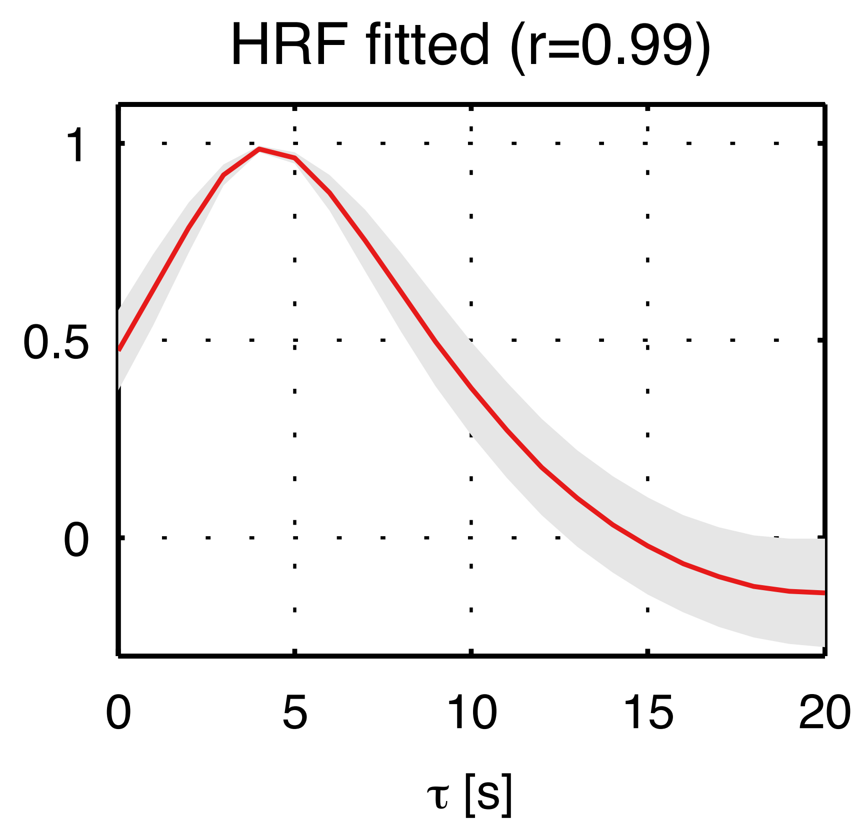


Non-Separable Spatiotemporal deconvolution

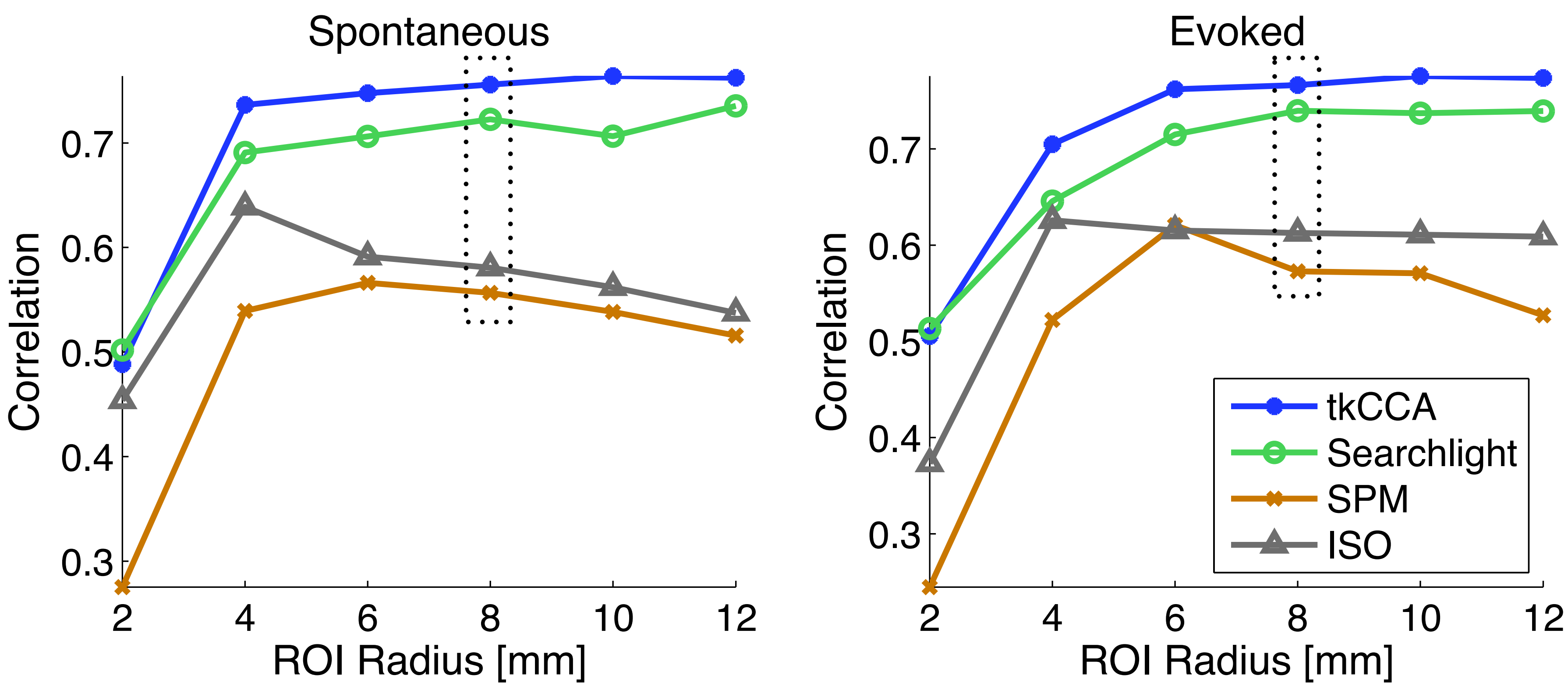


tkCCA HRF estimate vs. Canonical HRF model



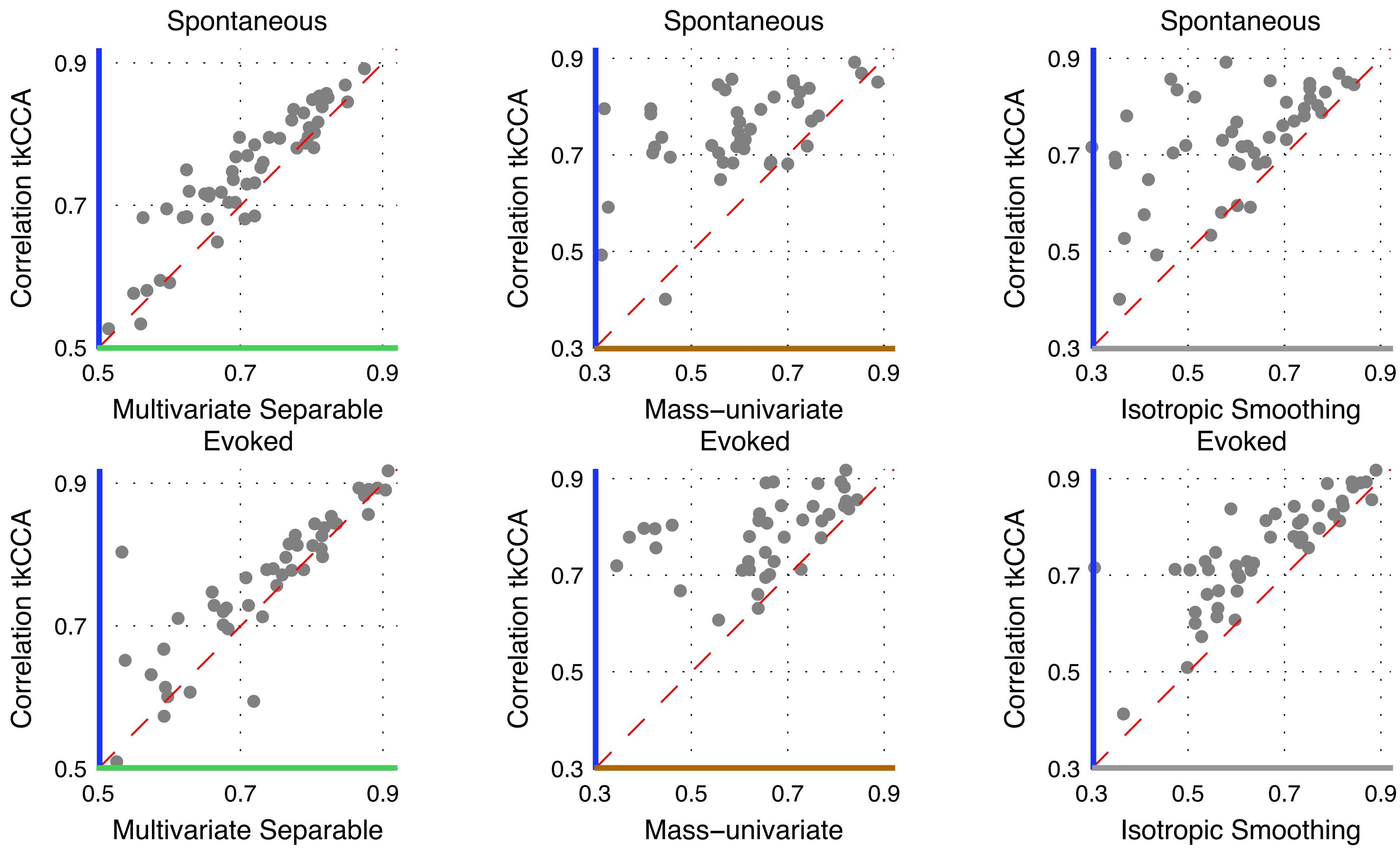


TkCCA vs Canonical HRF Models



Bießmann et al, Neuroimage, 2012

TkCCA vs Canonical HRF Models



Non-separable deconvolutions predict neural signal best

non-separable hemodynamics contain neural information

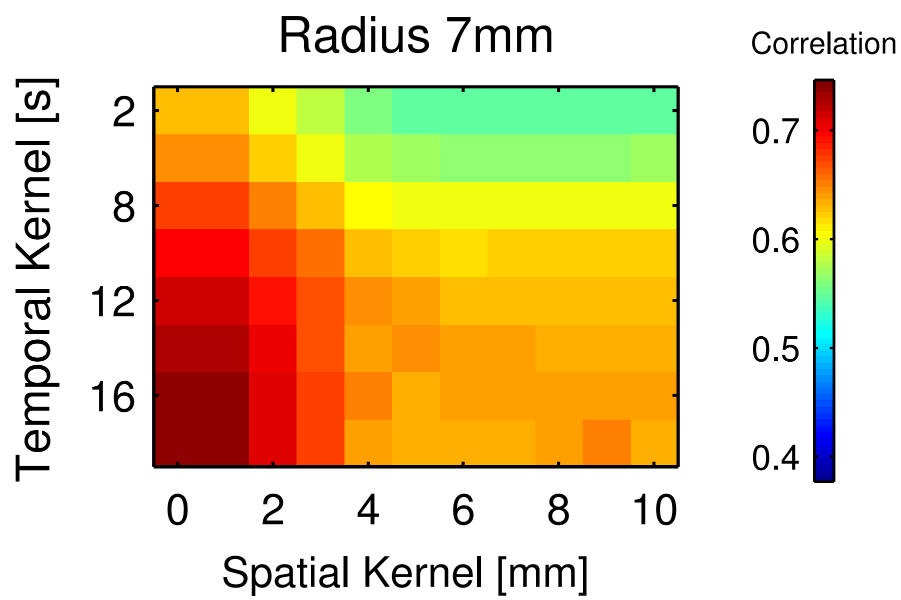
Bießmann et al, Neuroimage, 2012

Optimal Preprocessing for fMRI Decoding

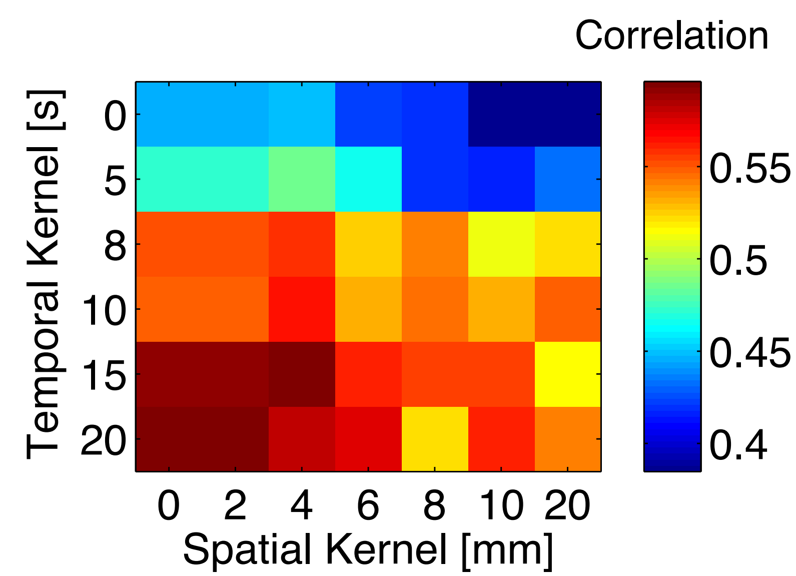
- Temporal smoothing
  - Can we decode neural signals **faster than HRF lowpass**?
- Spatial smoothing
  - **Does it help** for decoding neural information?
- Searchlight decoding
  - What is the **best searchlight radius**?

Effects of Spatial and Temporal Smoothing

Intracranial - fMRI (non-human primate)



EEG-fMRI (human)



Similar effects across species for intracranial and EEG

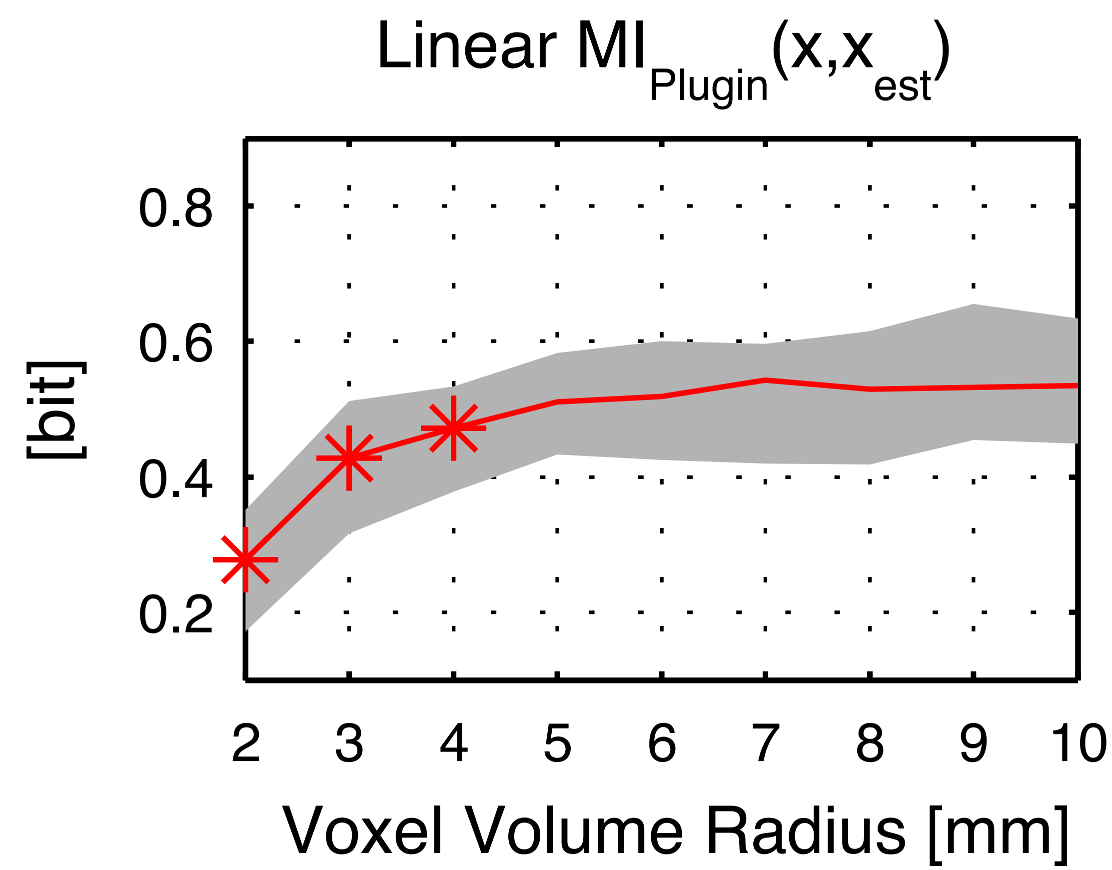
Temporal smoothing ➔ Better decoding:

Fast neural amplitude fluctuations could not be decoded from fMRI

Spatial smoothing ➔ Worse decoding:

Estimation of smoothing inverse difficult with limited amount of data ?

Effect of Searchlight Radius

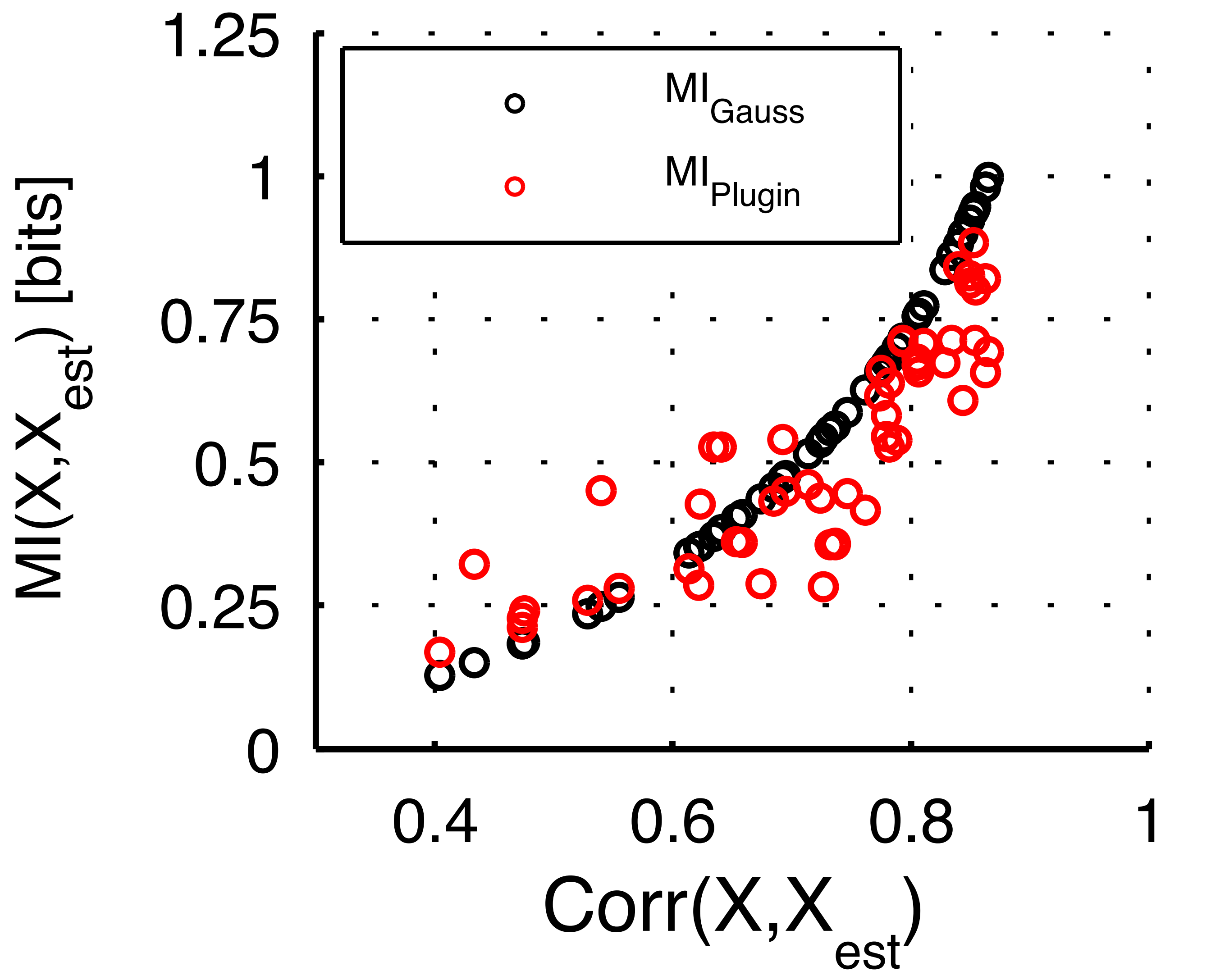


Searchlight radii < 5mm might loose information

Searchlight radii > 8mm add redundant information

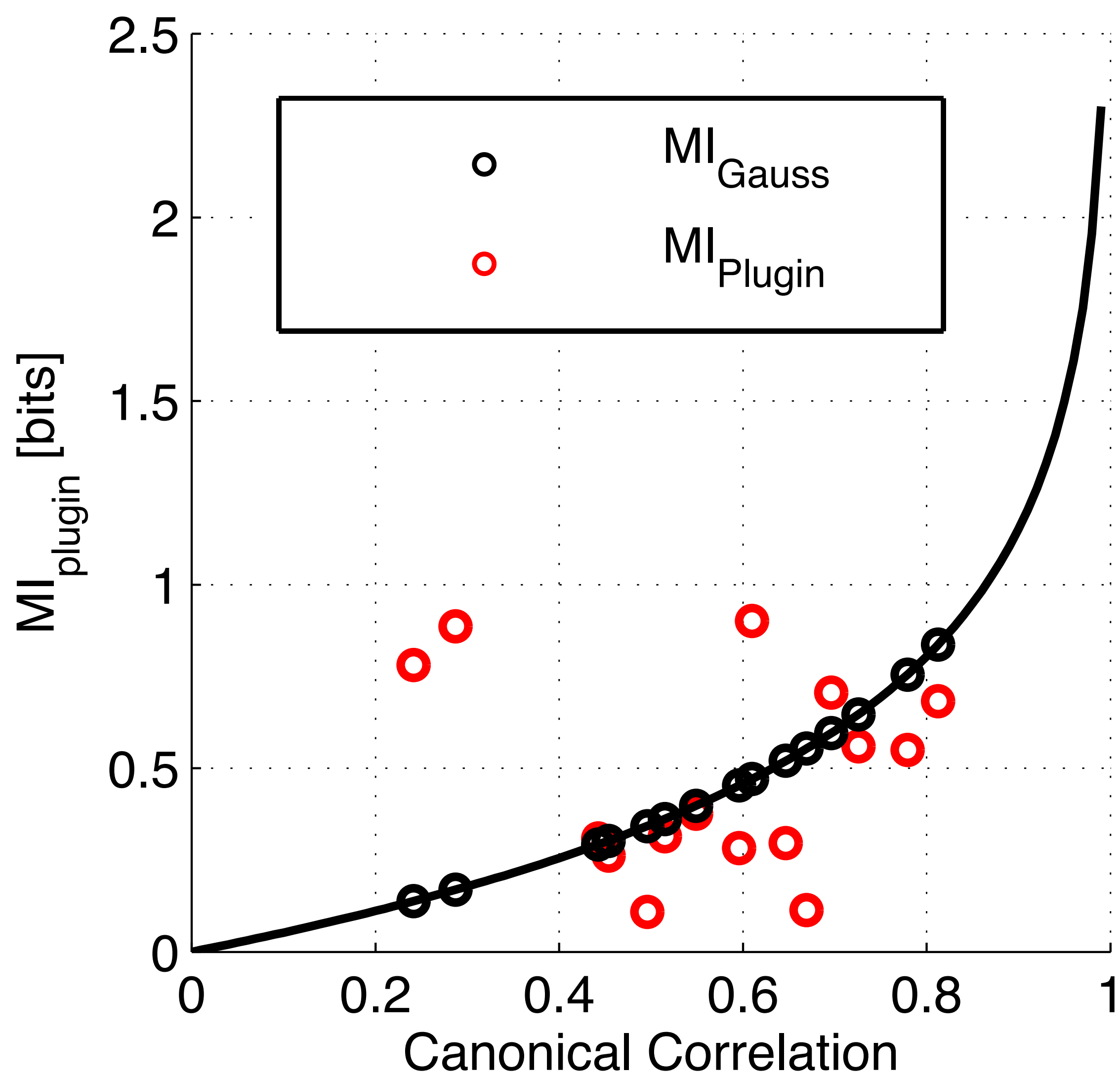
How Much Neural Information is in fMRI signals?

Mutual Information Estimates



Gaussian approximation (cheap and robust)
fits the (bias-corrected) plugin MI estimate well

Mutual Information Estimates: EEG-fMRI



1. Optimal parameters are **the same** as for intracranial data
2. Mutual Information around 0.5 bits

Summary

- There is more information in fMRI than HRF models assume

- Optimal parameters for extracting this information:
  - 4 - 8 mm searchlight radius
    - More includes redundant information
    - Less misses information
  - Temporal smoothing kernel > 15s
  - No spatial smoothing

- An fMRI volume of 5mm radius and 20s duration contains **0.5-0.8 bits of neural information**

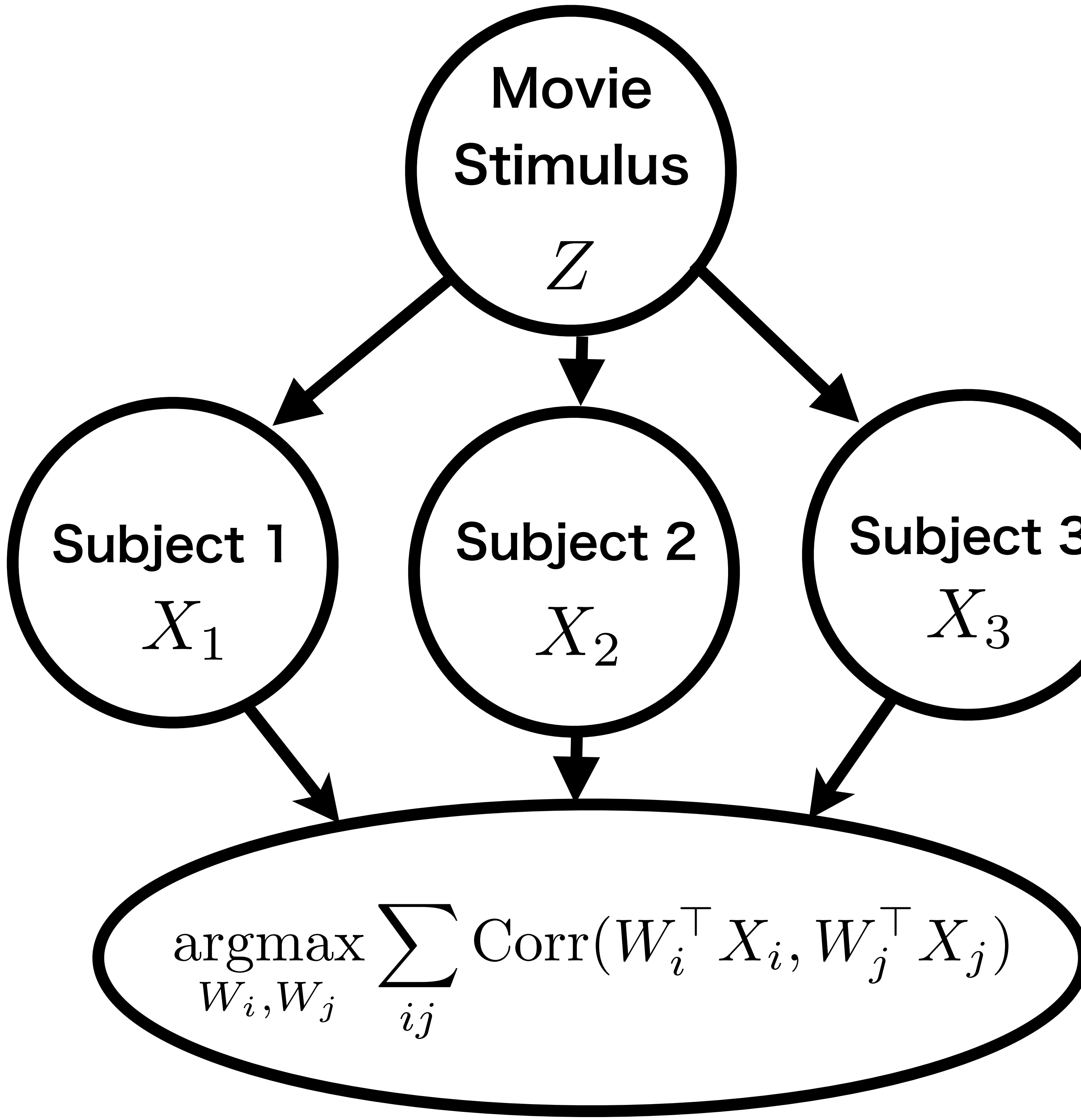
Multimodal Analyses for Hyperscanning

- For many paradigms we do not have stimulus regressors
  - Complex Movie stimuli
  - Resting state data

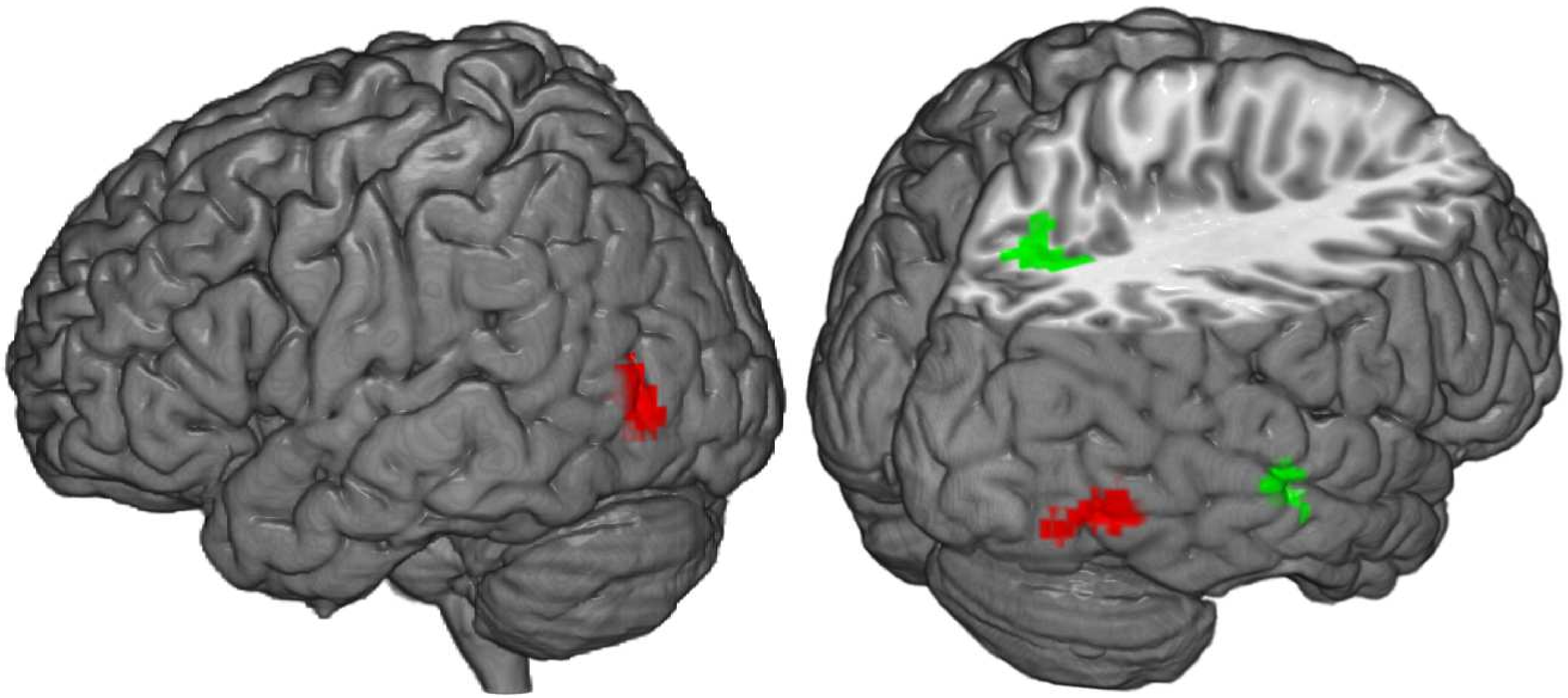
- We can treat subjects as modalities

- If multiple subjects are exposed to the same stimulus …

  - … neural activation **shared across subjects** reflects stimulus (see also [Hasson et al. 2009])


Movie
Stimulus
Z
Subject 1
X_1
Subject 2
X_2
Subject 3
X_3
$\underset{W_i, W_j}{\mathrm{argmax}} \sum_{ij} \mathrm{Corr}(W_i^\top X_i, W_j^\top X_j)$

Common Activation Patterns

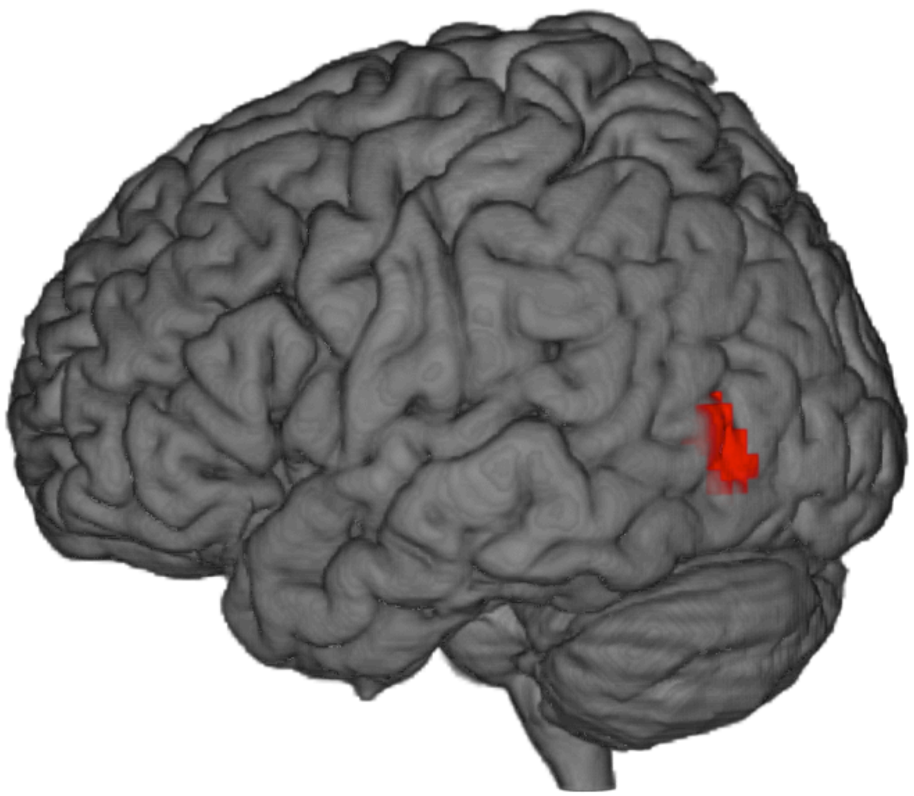


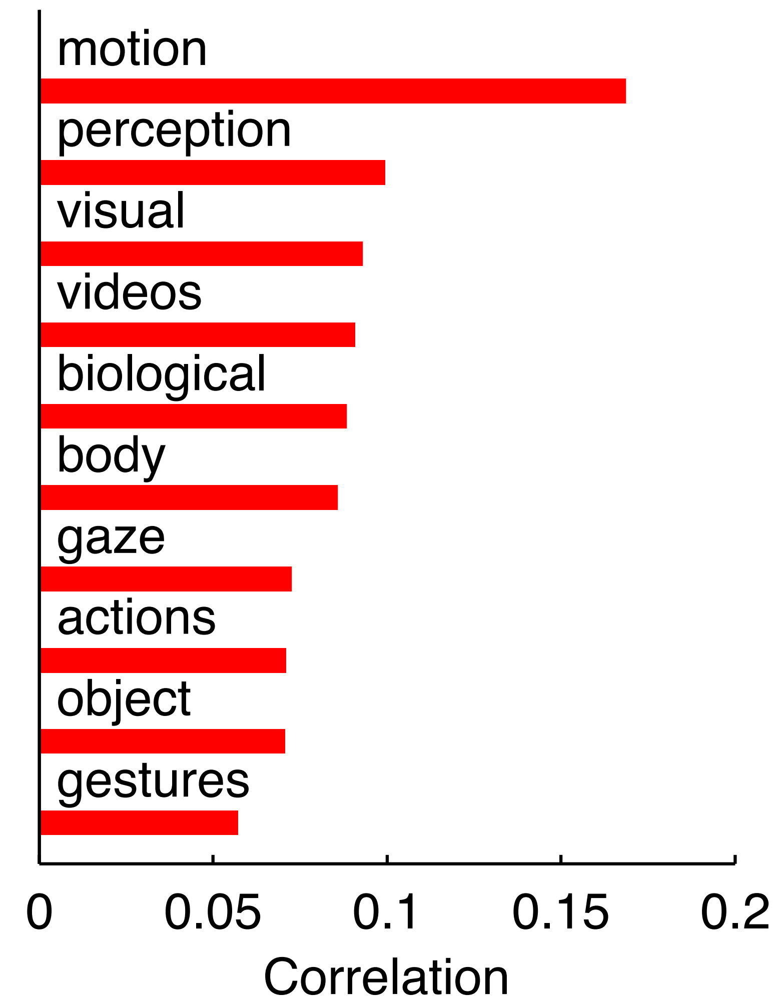
Common activation pattern of 25
subjects while watching the same movie

Gaebler et al, submitted

Common Activation Patterns

Correlations with psychological concepts stored in http://neurosynth.org (Yarkoni et al. Nat. Methods, 2011)





Gaebler et al, submitted

Summary

- Each subject is treated as a separate modality
- We can use multimodal analyses for hyperscanning
  - Multivariate extension of intersubject synchronization studies
  - Multivariate has higher SNR than mass-univariate
  - Allows to study networks rather than single voxels
- Multiple subjects experiencing the same stimulus
  - shared brain activity is likely to be due to the stimulus
  - this allows to analyze complex stimuli **without regressors**

Applications

- Hybrid BCIs: Combining EEG and NIRS
  - Multimodal setups increase BCI information transfer rates
- Cleaning artifacts in multimodal recordings
  - PCA: simple but efficient
- Decoding neural bandpower from fMRI
  - Canonical HRF models might miss information
- Multisubject Analyses
  - More sensitive than mass-univariate intersubject correlations

Thank You