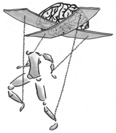
M/EEG Decoding and Brain-Computer Interfacing

Moritz Grosse-Wentrup

Max Planck Institute for Intelligent Systems
Department Empirical Inference
Tübingen, Germany

June 8, 2014

MAX-PLANCK-GESELLSCHAFT



Brain-Computer Interfacing in ALS

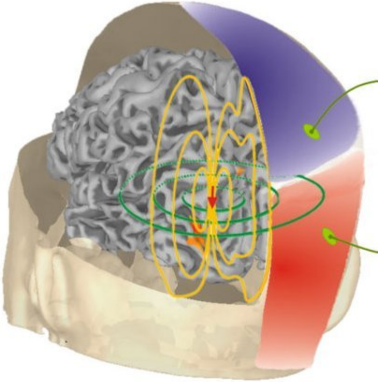
Outline

Outline

1. M/EEG Decoding Models

2. Brain-Computer Interfacing

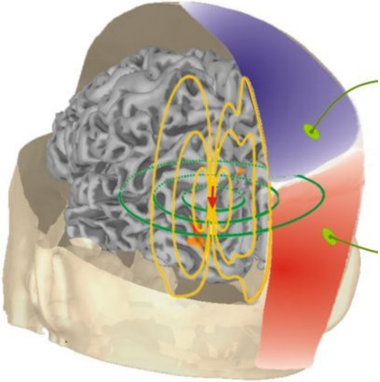
The brain's electromagnetic field (EMF)



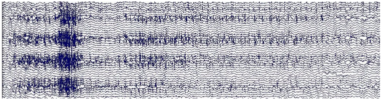
http://www.canada-meg-consortium.org/

(P.L. Nunez. Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, 2005.)

The brain's electromagnetic field (EMF)



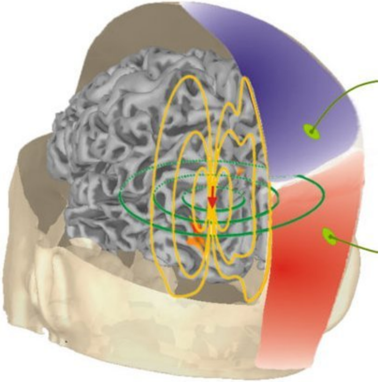
http://www.canada-meg-consortium.org/



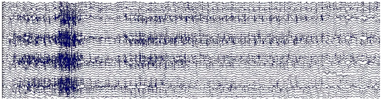
Time

(P.L. Nunez. Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, 2005.)

The brain's electromagnetic field (EMF)



http://www.canada-meg-consortium.org/

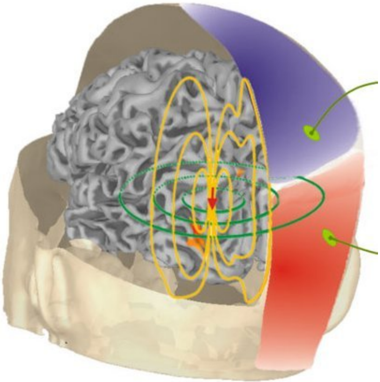


Time

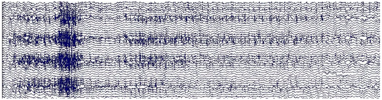
Notation:

(P.L. Nunez. Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, 2005.)

The brain's electromagnetic field (EMF)



http://www.canada-meg-consortium.org/

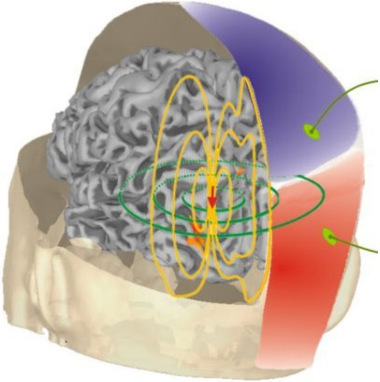


Time

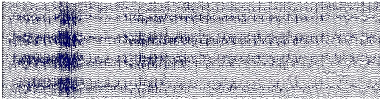
Notation:

- M/EEG signal $\mathbf{x}_i[t] \in \mathbb{R}^N$

(P.L. Nunez. Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, 2005.)



http://www.canada-meg-consortium.org/



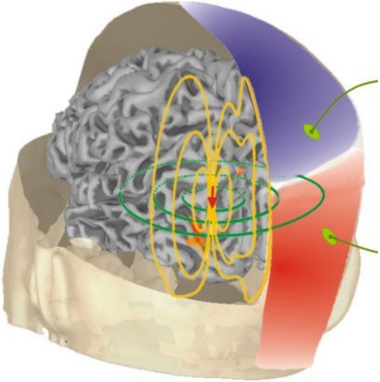
Time

Notation:

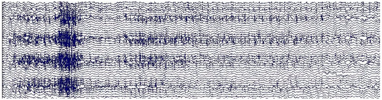
- M/EEG signal $\mathbf{x}_i[t] \in \mathbb{R}^N$
- N electrodes (32–256)

(P.L. Nunez. Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, 2005.)

The brain's electromagnetic field (EMF)



http://www.canada-meg-consortium.org/



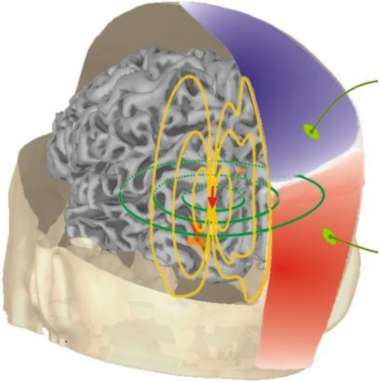
Time

Notation:

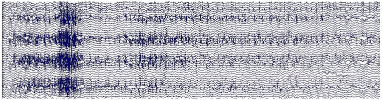
- M/EEG signal $\mathbf{x}_i[t] \in \mathbb{R}^N$
- N electrodes (32–256)
- Time $t = 1, \ldots, T$

(P.L. Nunez. Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, 2005.)

The brain's electromagnetic field (EMF)



http://www.canada-meg-consortium.org/



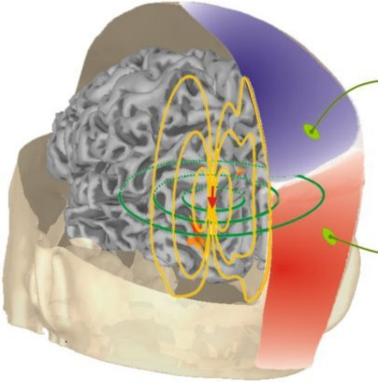
Time

Notation:

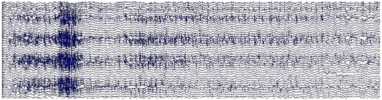
- M/EEG signal $\mathbf{x}_i[t] \in \mathbb{R}^N$
- N electrodes (32–256)
- Time $t = 1, \ldots, T$
- Trial $i = 1, \ldots, M$

(P.L. Nunez. Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, 2005.)

The brain's electromagnetic field (EMF)



http://www.canada-meg-consortium.org/

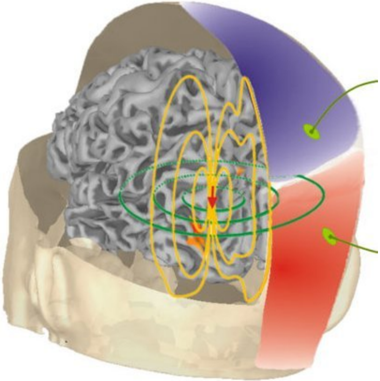


Notation:

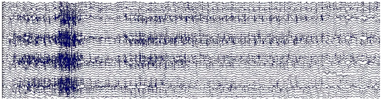
- M/EEG signal $\mathbf{x}_i[t] \in \mathbb{R}^N$
- N electrodes (32–256)
- Time $t = 1, \ldots, T$
- Trial $i = 1, \ldots, M$
- Exp. condition $c_i \in \{-1, +1\}$

(P.L. Nunez. Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, 2005.)

The brain's electromagnetic field (EMF)



http://www.canada-meg-consortium.org/



Time

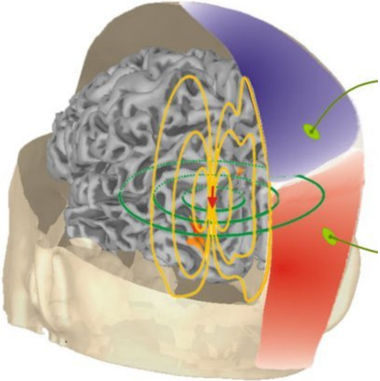
Notation:

- M/EEG signal $\mathbf{x}_i[t] \in \mathbb{R}^N$
- N electrodes (32–256)
- Time $t = 1, \ldots, T$
- Trial $i = 1, \ldots, M$
- Exp. condition $c_i \in \{-1, +1\}$

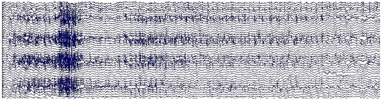
Experimental data:

(P.L. Nunez. Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, 2005.)

The brain's electromagnetic field (EMF)



http://www.canada-meg-consortium.org/



Notation:

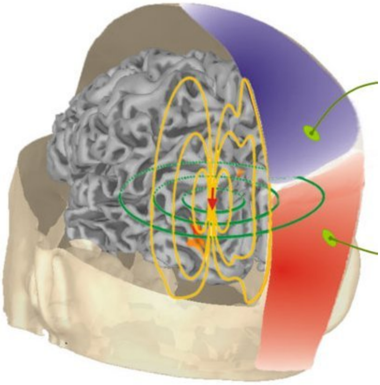
- M/EEG signal $\mathbf{x}_i[t] \in \mathbb{R}^N$
- N electrodes (32–256)
- Time $t = 1, \ldots, T$
- Trial $i = 1, \ldots, M$
- Exp. condition $c_i \in \{-1, +1\}$

Experimental data:

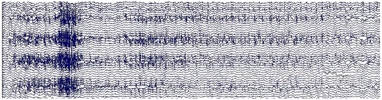
- $(c_i, X_i) \sim p(c, X)$

(P.L. Nunez. Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, 2005.)

The brain's electromagnetic field (EMF)



http://www.canada-meg-consortium.org/



Time

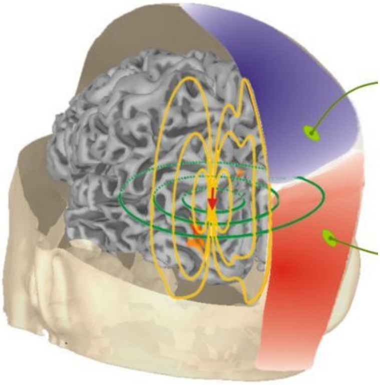
Notation:

- M/EEG signal $\mathbf{x}_i[t] \in \mathbb{R}^N$
- N electrodes (32–256)
- Time $t = 1, \ldots, T$
- Trial $i = 1, \ldots, M$
- Exp. condition $c_i \in \{-1, +1\}$

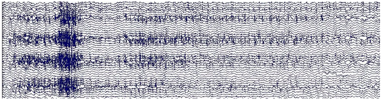
Experimental data:

- $(c_i, X_i) \sim p(c, X)$
- $\mathcal{D} = \{(c_1, X_1), \ldots, (c_M, X_M)\}$

(P.L. Nunez. Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, 2005.)



http://www.canada-meg-consortium.org/



Notation:

- M/EEG signal $\mathbf{x}_i[t] \in \mathbb{R}^N$
- N electrodes (32–256)
- Time $t = 1, \ldots, T$
- Trial $i = 1, \ldots, M$
- Exp. condition $c_i \in \{-1, +1\}$

Experimental data:

- $(c_i, X_i) \sim p(c, X)$
- $\mathcal{D} = \{(c_1, X_1), \ldots, (c_M, X_M)\}$
- Typically i.i.d. sampling is assumed

(P.L. Nunez. Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, 2005.)

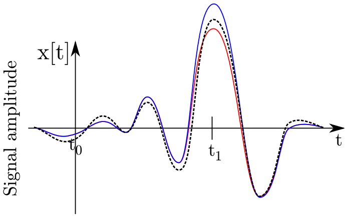
Decoding models

Decoding models

- Given a certain brain-state, what is the probability of an experimental condition: $p(c|X)$?

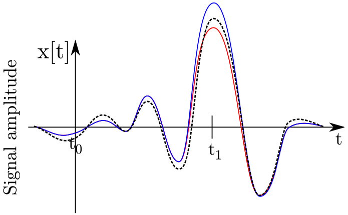
Decoding models

- Given a certain brain-state, what is the probability of an experimental condition: $p(c|X)$?



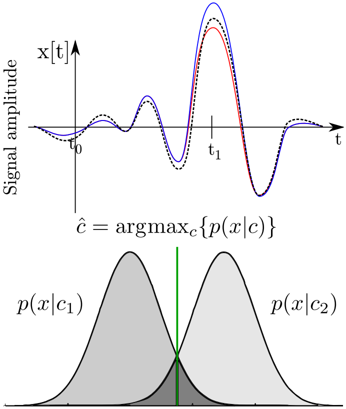
Decoding models

- Given a certain brain-state, what is the probability of an experimental condition: $p(c|X)$?
- The (optimal) Bayes classifier: $\hat{c}_i = f(X_i) = \text{argmax}_c\{p(X_i|c)\}$



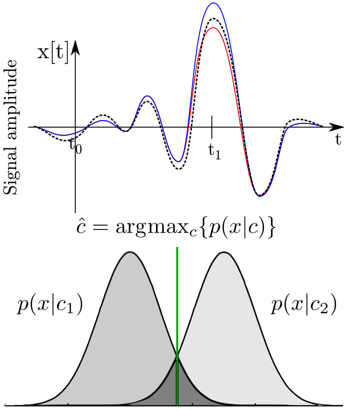
Decoding models

- Given a certain brain-state, what is the probability of an experimental condition: $p(c|X)$?
- The (optimal) Bayes classifier: $\hat{c}_i = f(X_i) = \text{argmax}_c\{p(X_i|c)\}$



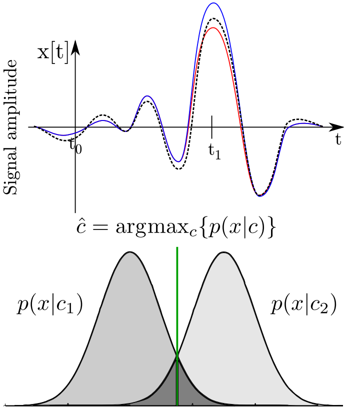
Decoding models

- Given a certain brain-state, what is the probability of an experimental condition: $p(c|X)$?
- The (optimal) Bayes classifier: $\hat{c}_i = f(X_i) = \text{argmax}_c\{p(X_i|c)\}$
- Examples of classification methods:



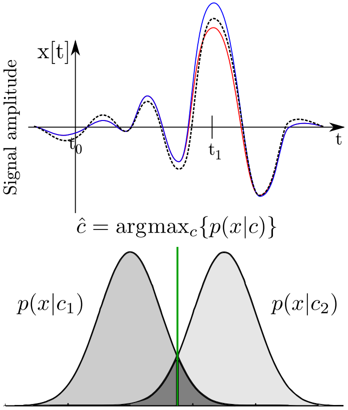
Decoding models

- Given a certain brain-state, what is the probability of an experimental condition: $p(c|X)$?
- The (optimal) Bayes classifier: $\hat{c}_i = f(X_i) = \text{argmax}_c\{p(X_i|c)\}$
- Examples of classification methods:
  - Linear discriminant analysis (LDA)
  - Support Vector Machines (SVM)
  - Gaussian Processes (GP)



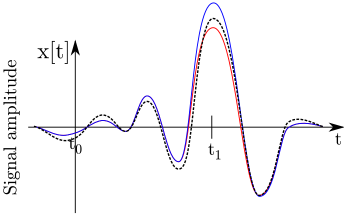
Decoding models

- Given a certain brain-state, what is the probability of an experimental condition: $p(c|X)$?
- The (optimal) Bayes classifier: $\hat{c}_i = f(X_i) = \text{argmax}_c\{p(X_i|c)\}$
- Examples of classification methods:
  - Linear discriminant analysis (LDA)
  - Support Vector Machines (SVM)
  - Gaussian Processes (GP)
- Prediction error $P_e := 1 - P(\hat{c} = c)$

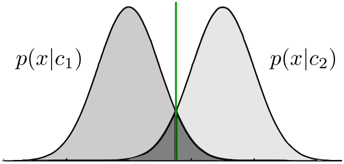


Decoding models

- Given a certain brain-state, what is the probability of an experimental condition: $p(c|X)$?
- The (optimal) Bayes classifier: $\hat{c}_i = f(X_i) = \text{argmax}_c\{p(X_i|c)\}$
- Examples of classification methods:
  - Linear discriminant analysis (LDA)
  - Support Vector Machines (SVM)
  - Gaussian Processes (GP)
- Prediction error $P_e := 1 - P(\hat{c} = c)$
- Advantage: Robust against (measurable) noise.

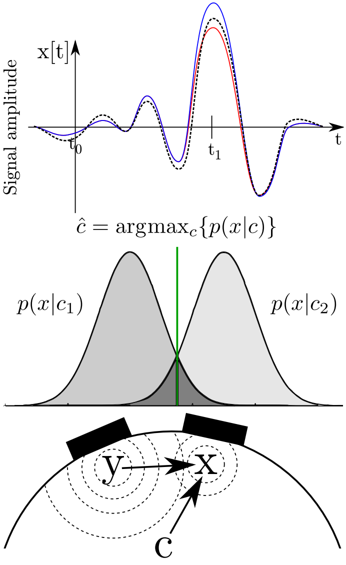


$\hat{c} = \text{argmax}_c\{p(x|c)\}$



Decoding models

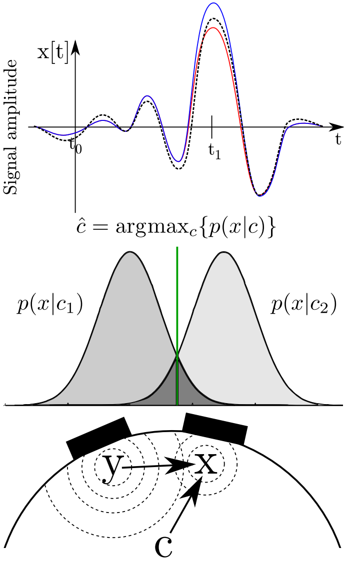
- Given a certain brain-state, what is the probability of an experimental condition: $p(c|X)$?
- The (optimal) Bayes classifier: $\hat{c}_i = f(X_i) = \text{argmax}_c\{p(X_i|c)\}$
- Examples of classification methods:
  - Linear discriminant analysis (LDA)
  - Support Vector Machines (SVM)
  - Gaussian Processes (GP)
- Prediction error $P_e := 1 - P(\hat{c} = c)$
- Advantage: Robust against (measurable) noise.



$x_i[t_1] = \alpha c_i + \beta y_i[t_1]$

Decoding models

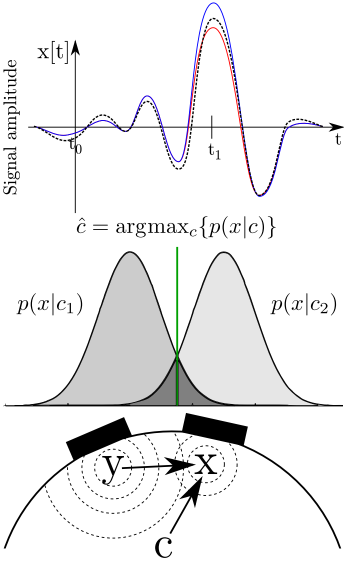
- Given a certain brain-state, what is the probability of an experimental condition: $p(c|X)$?
- The (optimal) Bayes classifier: $\hat{c}_i = f(X_i) = \text{argmax}_c\{p(X_i|c)\}$
- Examples of classification methods:
  - Linear discriminant analysis (LDA)
  - Support Vector Machines (SVM)
  - Gaussian Processes (GP)
- Prediction error $P_e := 1 - P(\hat{c} = c)$
- Advantage: Robust against (measurable) noise. $\hat{c}_i = \underbrace{\text{sign}\{x_i[t] - \beta y[t_1]\}}_{f} = c_i$



$$x_i[t_1] = \alpha c_i + \beta y_i[t_1]$$

Decoding models

- Given a certain brain-state, what is the probability of an experimental condition: $p(c|X)$?
- The (optimal) Bayes classifier: $\hat{c}_i = f(X_i) = \text{argmax}_c\{p(X_i|c)\}$
- Examples of classification methods:
  - Linear discriminant analysis (LDA)
  - Support Vector Machines (SVM)
  - Gaussian Processes (GP)
- Prediction error $P_e := 1 - P(\hat{c} = c)$
- Advantage: Robust against (measurable) noise. $\hat{c}_i = \underbrace{\text{sign}\{x_i[t] - \beta y[t_1]\}}_{f} = c_i$
- Disadvantage $f : \mathbb{R}^{N \times T} \mapsto \{-1, +1\}$ needs to be learned from $\mathcal{D}$.



$x_i[t_1] = \alpha c_i + \beta y_i[t_1]$

Learning decoding models I

(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

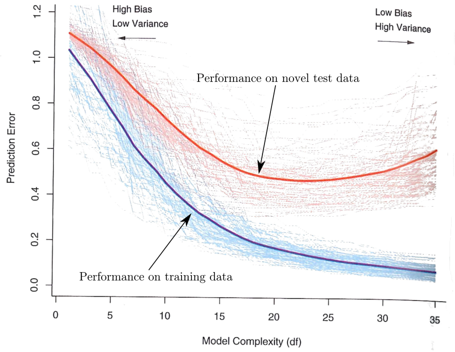
Learning decoding models I

The decoder f has to be chosen from a model class $\mathcal{F}$. How to choose $\mathcal{F}$?

(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

Learning decoding models I

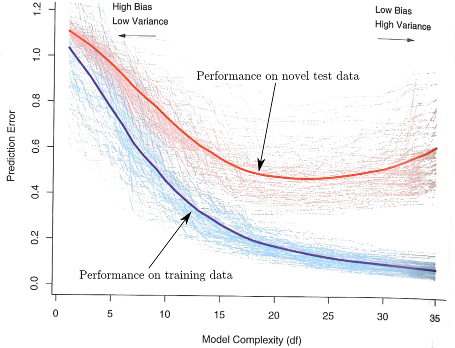
The decoder f has to be chosen from a model class $\mathcal{F}$. How to choose $\mathcal{F}$?

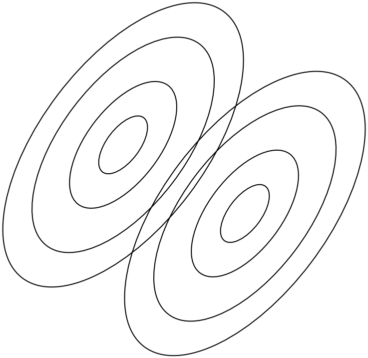


(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

Learning decoding models I

The decoder f has to be chosen from a model class $\mathcal{F}$. How to choose $\mathcal{F}$?

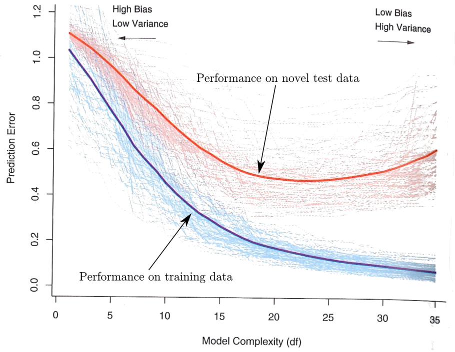


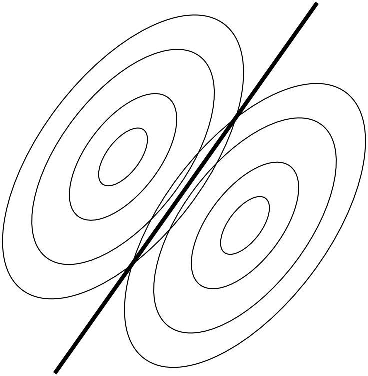


(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

Learning decoding models I

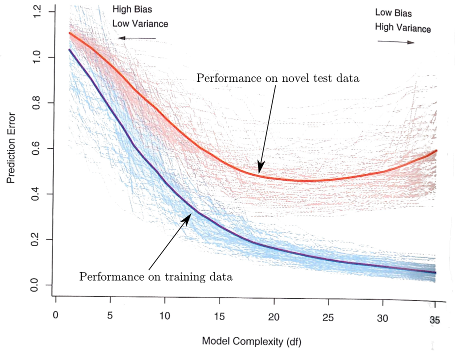
The decoder f has to be chosen from a model class $\mathcal{F}$. How to choose $\mathcal{F}$?

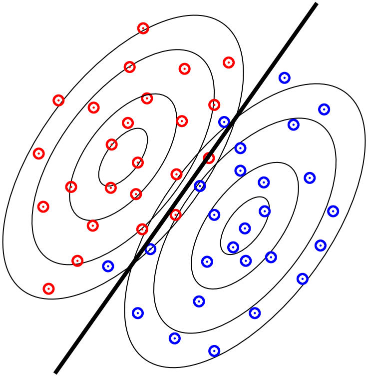




(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

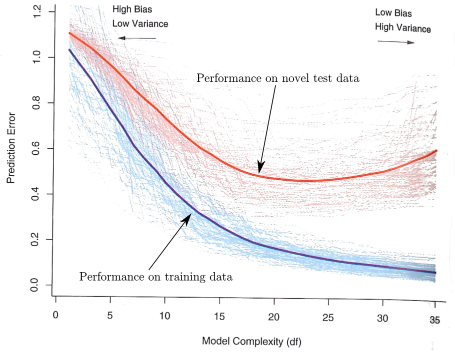
The decoder f has to be chosen from a model class $\mathcal{F}$. How to choose $\mathcal{F}$?

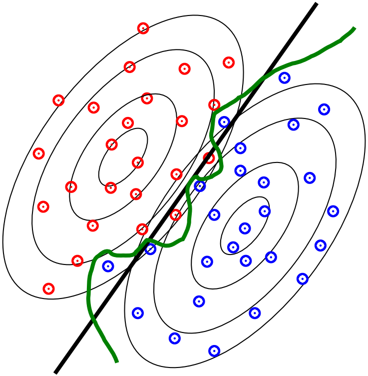




(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

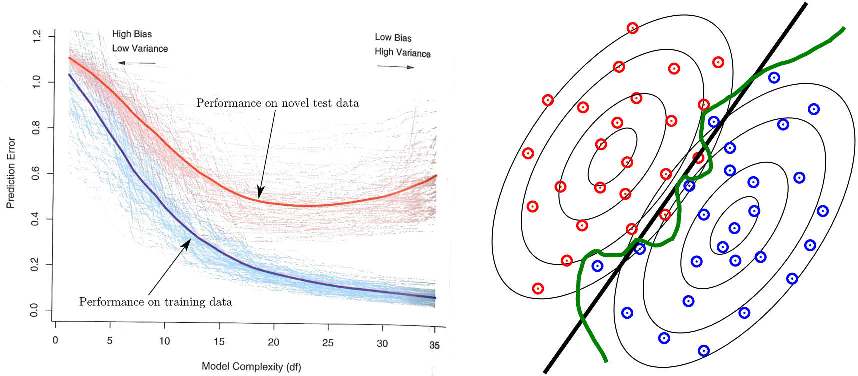
The decoder f has to be chosen from a model class $\mathcal{F}$. How to choose $\mathcal{F}$?





(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

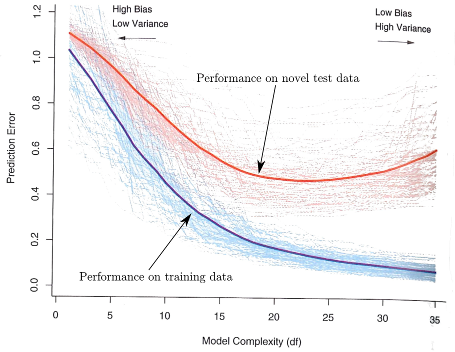
The decoder f has to be chosen from a model class $\mathcal{F}$. How to choose $\mathcal{F}$?

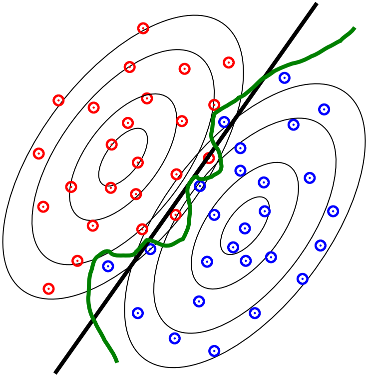


How do we determine which model generalizes best?

(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

The decoder f has to be chosen from a model class $\mathcal{F}$. How to choose $\mathcal{F}$?





How do we determine which model generalizes best?Cross-validation!

(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

Learning decoding models II

(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

Learning decoding models II

$f : \mathbb{R}^{N \times T} \mapsto \{-1, +1\}$ has to learned from training data $\mathcal{D}$.

(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

Learning decoding models II

$f : \mathbb{R}^{N \times T} \mapsto \{-1, +1\}$ has to learned from training data $\mathcal{D}$. The number of training samples needed to find the best $f \in \mathcal{F}$ scales with

(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

Learning decoding models II

$f : \mathbb{R}^{N \times T} \mapsto \{-1, +1\}$ has to learned from training data $\mathcal{D}$. The number of training samples needed to find the best $f \in \mathcal{F}$ scales with

- the model complexity,

(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

Learning decoding models II

$f : \mathbb{R}^{N \times T} \mapsto \{-1, +1\}$ has to learned from training data $\mathcal{D}$. The number of training samples needed to find the best $f \in \mathcal{F}$ scales with

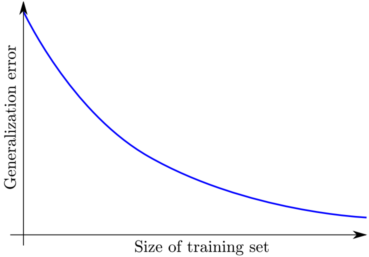
- the model complexity,
- and the number of features $N \times T$.

(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

Learning decoding models II

$f : \mathbb{R}^{N \times T} \mapsto \{-1, +1\}$ has to learned from training data $\mathcal{D}$. The number of training samples needed to find the best $f \in \mathcal{F}$ scales with

- the model complexity,
- and the number of features $N \times T$.

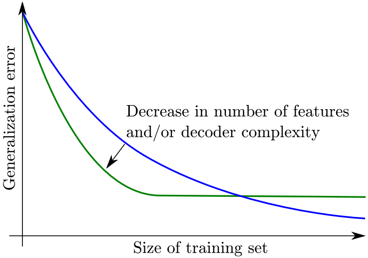


(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

Learning decoding models II

$f : \mathbb{R}^{N \times T} \mapsto \{-1, +1\}$ has to learned from training data $\mathcal{D}$. The number of training samples needed to find the best $f \in \mathcal{F}$ scales with

- the model complexity,
- and the number of features $N \times T$.

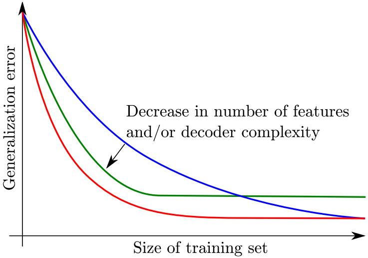


(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

Learning decoding models II

$f : \mathbb{R}^{N \times T} \mapsto \{-1, +1\}$ has to learned from training data $\mathcal{D}$. The number of training samples needed to find the best $f \in \mathcal{F}$ scales with

- the model complexity,
- and the number of features $N \times T$.

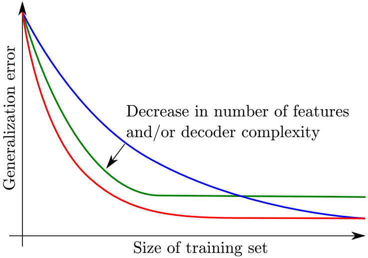


(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

Learning decoding models II

$f : \mathbb{R}^{N \times T} \mapsto \{-1, +1\}$ has to learned from training data $\mathcal{D}$. The number of training samples needed to find the best $f \in \mathcal{F}$ scales with

- the model complexity,
- and the number of features $N \times T$.



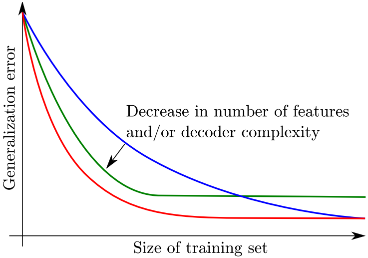
To learn a good classifier with limited training data, we should

(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

Learning decoding models II

$f : \mathbb{R}^{N \times T} \mapsto \{-1, +1\}$ has to learned from training data $\mathcal{D}$. The number of training samples needed to find the best $f \in \mathcal{F}$ scales with

- the model complexity,
- and the number of features $N \times T$.



To learn a good classifier with limited training data, we should

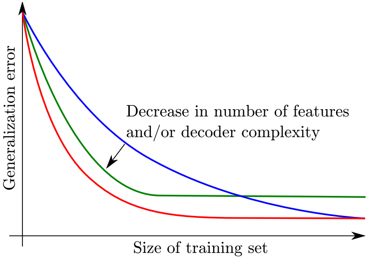
- reduce N and T without discarding information relevant for c,

(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

Learning decoding models II

$f : \mathbb{R}^{N \times T} \mapsto \{-1, +1\}$ has to learned from training data $\mathcal{D}$. The number of training samples needed to find the best $f \in \mathcal{F}$ scales with

- the model complexity,
- and the number of features $N \times T$.



To learn a good classifier with limited training data, we should

- reduce N and T without discarding information relevant for c,
- and find a *simple* representation of X.

(Hastie, Tibshirani, & Friedman. The Elements of Statistical Learning. Springer, 2009)

Reducing the number of M/EEG channels (N)

(Baillet, Mosher & Leahy. Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 2001)

Reducing the number of M/EEG channels (N)

Spatial filtering of M/EEG data:

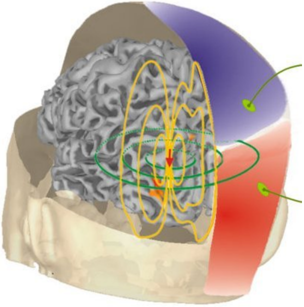
(Baillet, Mosher & Leahy. Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 2001)

Reducing the number of M/EEG channels (N)

Spatial filtering of M/EEG data:

$$\mathbf{x}[t] = \quad L\mathbf{s}[t]$$

- Source vector $\mathbf{s}[t] \in \mathbb{R}^K$
- Leadfield matrix $L \in \mathbb{R}^{N \times K}$



http://www.canada-meg-consortium.org/

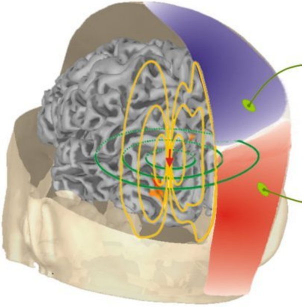
(Baillet, Mosher & Leahy. Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 2001)

Reducing the number of M/EEG channels (N)

Spatial filtering of M/EEG data:

$$y[t] = \mathbf{w}^\mathsf{T}\mathbf{x}[t] = \mathbf{w}^\mathsf{T}L\mathbf{s}[t]$$

- Source vector $\mathbf{s}[t] \in \mathbb{R}^K$
- Leadfield matrix $L \in \mathbb{R}^{N \times K}$
- Spatial filter $\mathbf{w} \in \mathbb{R}^N$



http://www.canada-meg-consortium.org/

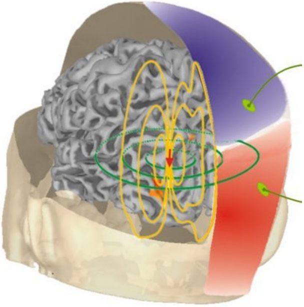
(Baillet, Mosher & Leahy. Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 2001)

Reducing the number of M/EEG channels (N)

Spatial filtering of M/EEG data:

$$y[t] = \mathbf{w}^\mathsf{T}\mathbf{x}[t] = \mathbf{w}^\mathsf{T}L\mathbf{s}[t] = \mathbf{g}^\mathsf{T}\mathbf{s}[t]$$

- Source vector $\mathbf{s}[t] \in \mathbb{R}^K$
- Leadfield matrix $L \in \mathbb{R}^{N \times K}$
- Spatial filter $\mathbf{w} \in \mathbb{R}^N$
- Gain vector $\mathbf{g} \in \mathbb{R}^K$



http://www.canada-meg-consortium.org/

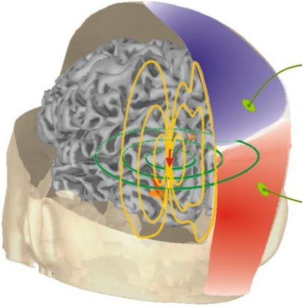
(Baillet, Mosher & Leahy. Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 2001)

Reducing the number of M/EEG channels (N)

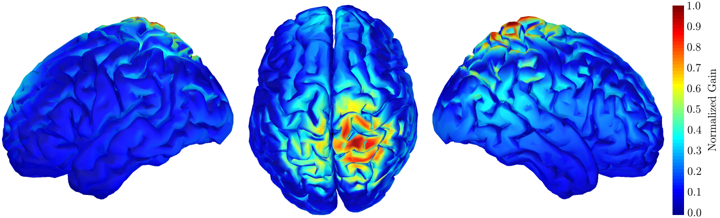
Spatial filtering of M/EEG data:

$$y[t] = \mathbf{w}^\mathsf{T}\mathbf{x}[t] = \mathbf{w}^\mathsf{T}L\mathbf{s}[t] = \mathbf{g}^\mathsf{T}\mathbf{s}[t]$$

- Source vector $\mathbf{s}[t] \in \mathbb{R}^K$
- Leadfield matrix $L \in \mathbb{R}^{N \times K}$
- Spatial filter $\mathbf{w} \in \mathbb{R}^N$
- Gain vector $\mathbf{g} \in \mathbb{R}^K$



http://www.canada-meg-consortium.org/



(Baillet, Mosher & Leahy. Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 2001)

Beamforming

(Van Veen et al. Localization of brain electrical activity via LCMV spatial filtering. *IEEE TBME*, 1997)

Beamforming

Unsupervised method based on a-priori knowledge of the spatial origin of relevant sources:

(Van Veen et al. Localization of brain electrical activity via LCMV spatial filtering. *IEEE TBME*, 1997)

Beamforming

Unsupervised method based on a-priori knowledge of the spatial origin of relevant sources:

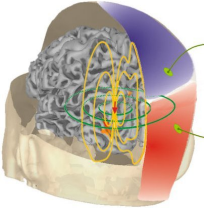
- Pick a cortical target source s^*

(Van Veen et al. Localization of brain electrical activity via LCMV spatial filtering. *IEEE TBME*, 1997)

Beamforming

Unsupervised method based on a-priori knowledge of the spatial origin of relevant sources:

- Pick a cortical target source s^*



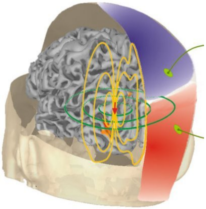
http://www.canada-meg-consortium.org/

(Van Veen et al. Localization of brain electrical activity via LCMV spatial filtering. *IEEE TBME*, 1997)

Beamforming

Unsupervised method based on a-priori knowledge of the spatial origin of relevant sources:

- Pick a cortical target source s^*
- Compute the forward solution $\mathbf{a} = \mathbf{l}s^*$



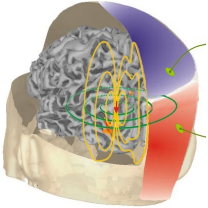
http://www.canada-meg-consortium.org/

(Van Veen et al. Localization of brain electrical activity via LCMV spatial filtering. *IEEE TBME*, 1997)

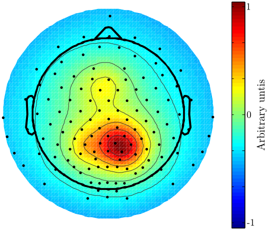
Beamforming

Unsupervised method based on a-priori knowledge of the spatial origin of relevant sources:

- Pick a cortical target source s^*
- Compute the forward solution $\mathbf{a} = \mathbf{l}s^*$



http://www.canada-meg-consortium.org/

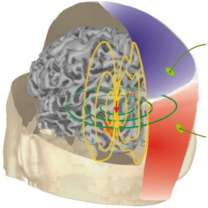


(Van Veen et al. Localization of brain electrical activity via LCMV spatial filtering. *IEEE TBME*, 1997)

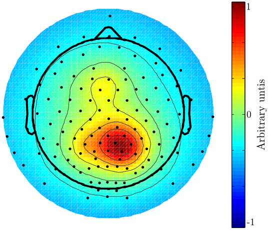
Beamforming

Unsupervised method based on a-priori knowledge of the spatial origin of relevant sources:

- Pick a cortical target source s^*
- Compute the forward solution $\mathbf{a} = \mathbf{l}s^*$
- Compute the M/EEG covariance matrix $\Sigma \in \mathbb{R}^{N \times N}$



http://www.canada-meg-consortium.org/

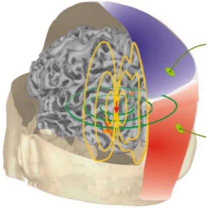


(Van Veen et al. Localization of brain electrical activity via LCMV spatial filtering. *IEEE TBME*, 1997)

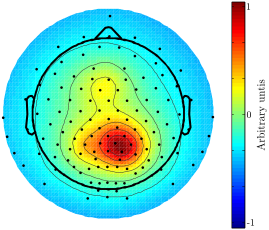
Beamforming

Unsupervised method based on a-priori knowledge of the spatial origin of relevant sources:

- Pick a cortical target source s^*
- Compute the forward solution $\mathbf{a} = \mathbf{l}s^*$
- Compute the M/EEG covariance matrix $\Sigma \in \mathbb{R}^{N \times N}$
- Solve the optimization problem $\mathbf{w} = \text{argmin}_{\mathbf{w}}\{\mathbf{w}^{\mathsf{T}}\Sigma\mathbf{w}\}$ s.t. $\mathbf{w}^{\mathsf{T}}\mathbf{a} = 1$



http://www.canada-meg-consortium.org/

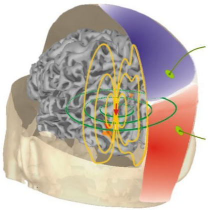


(Van Veen et al. Localization of brain electrical activity via LCMV spatial filtering. *IEEE TBME*, 1997)

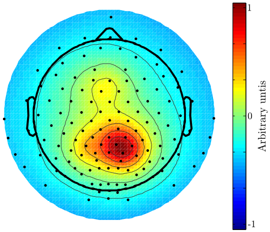
Beamforming

Unsupervised method based on a-priori knowledge of the spatial origin of relevant sources:

- Pick a cortical target source s^*
- Compute the forward solution $\mathbf{a} = \mathbf{l}s^*$
- Compute the M/EEG covariance matrix $\Sigma \in \mathbb{R}^{N \times N}$
- Solve the optimization problem $\mathbf{w} = \text{argmin}_{\mathbf{w}}\{\mathbf{w}^\mathsf{T}\Sigma\mathbf{w}\}$ s.t. $\mathbf{w}^\mathsf{T}\mathbf{a} = 1$
- $\mathbf{w} = \mathbf{a}^\mathsf{T}\Sigma^{-1}/(\mathbf{a}^\mathsf{T}\Sigma^{-1}\mathbf{a})$



http://www.canada-meg-consortium.org/

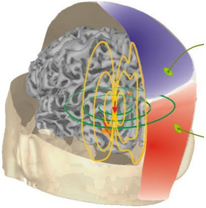


(Van Veen et al. Localization of brain electrical activity via LCMV spatial filtering. *IEEE TBME*, 1997)

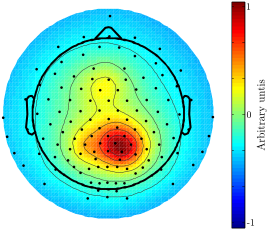
Beamforming

Unsupervised method based on a-priori knowledge of the spatial origin of relevant sources:

- Pick a cortical target source s^*
- Compute the forward solution $\mathbf{a} = \mathbf{l}s^*$
- Compute the M/EEG covariance matrix $\Sigma \in \mathbb{R}^{N \times N}$
- Solve the optimization problem $\mathbf{w} = \text{argmin}_{\mathbf{w}}\{\mathbf{w}^\mathsf{T}\Sigma\mathbf{w}\}$ s.t. $\mathbf{w}^\mathsf{T}\mathbf{a} = 1$
- $\mathbf{w} = \mathbf{a}^\mathsf{T}\Sigma^{-1}/(\mathbf{a}^\mathsf{T}\Sigma^{-1}\mathbf{a})$
- Check the gain vector $\mathbf{g} = \mathbf{w}^\mathsf{T} L$



http://www.canada-meg-consortium.org/

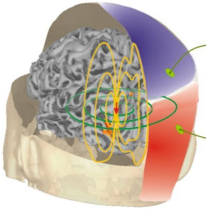


(Van Veen et al. Localization of brain electrical activity via LCMV spatial filtering. *IEEE TBME*, 1997)

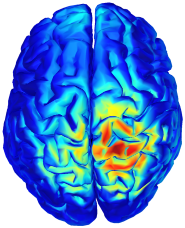
Beamforming

Unsupervised method based on a-priori knowledge of the spatial origin of relevant sources:

- Pick a cortical target source s^*
- Compute the forward solution $\mathbf{a} = \mathbf{l}s^*$
- Compute the M/EEG covariance matrix $\Sigma \in \mathbb{R}^{N \times N}$
- Solve the optimization problem $\mathbf{w} = \text{argmin}_{\mathbf{w}}\{\mathbf{w}^\mathsf{T}\Sigma\mathbf{w}\}$ s.t. $\mathbf{w}^\mathsf{T}\mathbf{a} = 1$
- $\mathbf{w} = \mathbf{a}^\mathsf{T}\Sigma^{-1}/(\mathbf{a}^\mathsf{T}\Sigma^{-1}\mathbf{a})$
- Check the gain vector $\mathbf{g} = \mathbf{w}^\mathsf{T} L$



http://www.canada-meg-consortium.org/

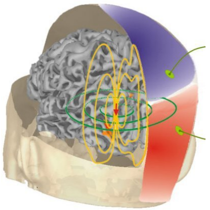


(Van Veen et al. Localization of brain electrical activity via LCMV spatial filtering. *IEEE TBME*, 1997)

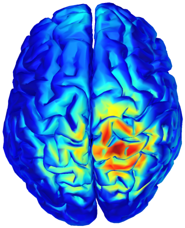
Beamforming

Unsupervised method based on a-priori knowledge of the spatial origin of relevant sources:

- Pick a cortical target source s^*
- Compute the forward solution $\mathbf{a} = \mathbf{l}s^*$
- Compute the M/EEG covariance matrix $\Sigma \in \mathbb{R}^{N \times N}$
- Solve the optimization problem $\mathbf{w} = \text{argmin}_{\mathbf{w}}\{\mathbf{w}^T \Sigma \mathbf{w}\}$ s.t. $\mathbf{w}^T \mathbf{a} = 1$
- $\mathbf{w} = \mathbf{a}^T \Sigma^{-1} / (\mathbf{a}^T \Sigma^{-1} \mathbf{a})$
- Check the gain vector $\mathbf{g} = \mathbf{w}^T L$
- Apply the spatial filter: $y[t] = \mathbf{w}^T \mathbf{x}[t]$



http://www.canada-meg-consortium.org/



(Van Veen et al. Localization of brain electrical activity via LCMV spatial filtering. *IEEE TBME*, 1997)

Common Spatial Patterns (CSP)

(Ramoser et al. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE TBME*, 2000)

Common Spatial Patterns (CSP)

Supervised method to find spatial filters that discriminate between two conditions:

(Ramoser et al. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE TBME*, 2000)

Common Spatial Patterns (CSP)

Supervised method to find spatial filters that discriminate between two conditions:

- Covariance matrices $\Sigma_{c=+1}$ & $\Sigma_{c=-1}$

(Ramoser et al. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE TBME*, 2000)

Common Spatial Patterns (CSP)

Supervised method to find spatial filters that discriminate between two conditions:

- Covariance matrices $\Sigma_{c=+1}$ & $\Sigma_{c=-1}$
- Optimization problem:
  $\mathbf{w} = \text{argmax}_{\mathbf{w}} \left\{ \frac{\mathbf{w}^{\mathsf{T}} \Sigma_{c=+1} \mathbf{w}}{\mathbf{w}^{\mathsf{T}} \Sigma_{c=-1} \mathbf{w}} \right\}$

(Ramoser et al. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE TBME*, 2000)

Common Spatial Patterns (CSP)

Supervised method to find spatial filters that discriminate between two conditions:

- Covariance matrices $\Sigma_{c=+1}$ & $\Sigma_{c=-1}$
- Optimization problem:
  $\mathbf{w} = \text{argmax}_{\mathbf{w}} \left\{ \frac{\mathbf{w}^{\mathsf{T}} \Sigma_{c=+1} \mathbf{w}}{\mathbf{w}^{\mathsf{T}} \Sigma_{c=-1} \mathbf{w}} \right\}$
- Solution: Largest λ for which
  $\Sigma_{c=+1}^{-1} \Sigma_{c=-1} \mathbf{w} = \lambda \mathbf{w}$

(Ramoser et al. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE TBME*, 2000)

Common Spatial Patterns (CSP)

Supervised method to find spatial filters that discriminate between two conditions:

- Covariance matrices $\Sigma_{c=+1}$ & $\Sigma_{c=-1}$
- Optimization problem:
  $\mathbf{w} = \text{argmax}_{\mathbf{w}} \left\{ \frac{\mathbf{w}^{\mathsf{T}} \Sigma_{c=+1} \mathbf{w}}{\mathbf{w}^{\mathsf{T}} \Sigma_{c=-1} \mathbf{w}} \right\}$
- Solution: Largest λ for which
  $\Sigma_{c=+1}^{-1} \Sigma_{c=-1} \mathbf{w} = \lambda \mathbf{w}$
- Matlab: [V,D] = eig($\Sigma_{c=+1}$,$\Sigma_{c=-1}$)

(Ramoser et al. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE TBME*, 2000)

Common Spatial Patterns (CSP)

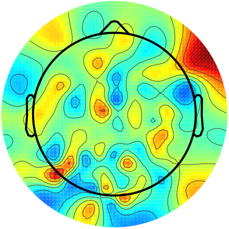
Supervised method to find spatial filters that discriminate between two conditions:

- Covariance matrices $\Sigma_{c=+1}$ & $\Sigma_{c=-1}$
- Optimization problem:
  $\mathbf{w} = \text{argmax}_{\mathbf{w}} \left\{ \frac{\mathbf{w}^\mathsf{T}\Sigma_{c=+1}\mathbf{w}}{\mathbf{w}^\mathsf{T}\Sigma_{c=-1}\mathbf{w}} \right\}$
- Solution: Largest λ for which
  $\Sigma_{c=+1}^{-1}\Sigma_{c=-1}\mathbf{w} = \lambda\mathbf{w}$
- Matlab: [V,D] = eig($\Sigma_{c=+1}$,$\Sigma_{c=-1}$)
- Columns of V are spatial filters

(Ramoser et al. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE TBME*, 2000)

Common Spatial Patterns (CSP)

Supervised method to find spatial filters that discriminate between two conditions:

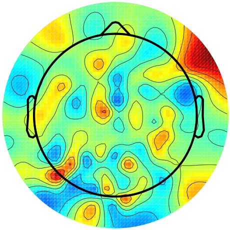


- Covariance matrices $\Sigma_{c=+1}$ & $\Sigma_{c=-1}$
- Optimization problem:
  $\mathbf{w} = \text{argmax}_{\mathbf{w}} \left\{ \frac{\mathbf{w}^{\mathsf{T}} \Sigma_{c=+1} \mathbf{w}}{\mathbf{w}^{\mathsf{T}} \Sigma_{c=-1} \mathbf{w}} \right\}$
- Solution: Largest λ for which $\Sigma_{c=+1}^{-1} \Sigma_{c=-1} \mathbf{w} = \lambda \mathbf{w}$
- Matlab: [V,D] = eig($\Sigma_{c=+1}$,$\Sigma_{c=-1}$)
- Columns of V are spatial filters

(Ramoser et al. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE TBME*, 2000)

Common Spatial Patterns (CSP)

Supervised method to find spatial filters that discriminate between two conditions:



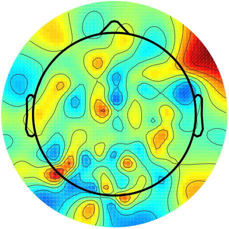
- Covariance matrices $\Sigma_{c=+1}$ & $\Sigma_{c=-1}$
- Optimization problem:
  $\mathbf{w} = \text{argmax}_{\mathbf{w}} \left\{ \frac{\mathbf{w}^{\mathsf{T}} \Sigma_{c=+1} \mathbf{w}}{\mathbf{w}^{\mathsf{T}} \Sigma_{c=-1} \mathbf{w}} \right\}$
- Solution: Largest λ for which $\Sigma_{c=+1}^{-1} \Sigma_{c=-1} \mathbf{w} = \lambda \mathbf{w}$
- Matlab: [V,D] = eig($\Sigma_{c=+1}$,$\Sigma_{c=-1}$)
- Columns of V are spatial filters
- Rows of V are spatial patterns

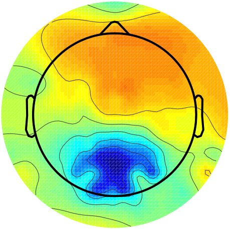
(Ramoser et al. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE TBME*, 2000)

Common Spatial Patterns (CSP)

Supervised method to find spatial filters that discriminate between two conditions:

- Covariance matrices $\Sigma_{c=+1}$ & $\Sigma_{c=-1}$
- Optimization problem:
  $\mathbf{w} = \text{argmax}_{\mathbf{w}} \left\{ \frac{\mathbf{w}^{\mathsf{T}}\Sigma_{c=+1}\mathbf{w}}{\mathbf{w}^{\mathsf{T}}\Sigma_{c=-1}\mathbf{w}} \right\}$
- Solution: Largest λ for which $\Sigma_{c=+1}^{-1}\Sigma_{c=-1}\mathbf{w} = \lambda\mathbf{w}$
- Matlab: [V,D] = eig($\Sigma_{c=+1}$,$\Sigma_{c=-1}$)
- Columns of V are spatial filters
- Rows of V are spatial patterns





(Ramoser et al. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE TBME*, 2000)

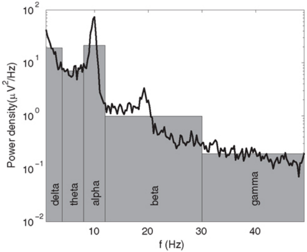
Reducing the number of time points (T)

Reducing the number of time points (T)

The brain cares about oscillations:

Reducing the number of time points (T)

The brain cares about oscillations:

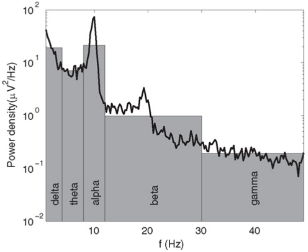


(van Albada & Robinson, *Frontiers in Human Neuroscience*, 2013)

Reducing the number of time points (T)

The brain cares about oscillations:

- $\text{DTFT}(y_i[t], \omega) = 1/T \sum_{t=0}^{T-1} y_i[t] e^{-j\omega t}$

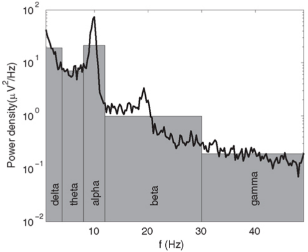


(van Albada & Robinson, *Frontiers in Human Neuroscience*, 2013)

Reducing the number of time points (T)

The brain cares about oscillations:

- $\text{DTFT}(y_i[t], \omega) = 1/T \sum_{t=0}^{T-1} y_i[t] e^{-j\omega t}$
- $z_i[\omega] = \log|\text{DTFT}(y_i[t], \omega)|$

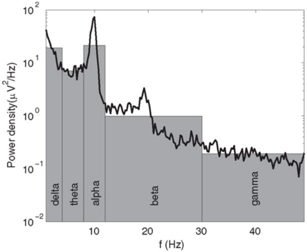


(van Albada & Robinson, *Frontiers in Human Neuroscience*, 2013)

Reducing the number of time points (T)

The brain cares about oscillations:

- $\text{DTFT}(y_i[t], \omega) = 1/T \sum_{t=0}^{T-1} y_i[t] e^{-j\omega t}$
- $z_i[\omega] = \log|\text{DTFT}(y_i[t], \omega)|$
- $\mathcal{Z}_i = \{z_i[\delta], z_i[\theta], z_i[\alpha], z_i[\beta], z_i[\gamma]\}$

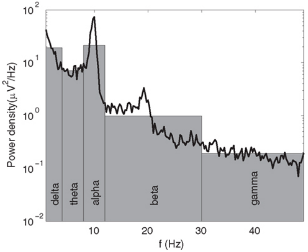


(van Albada & Robinson, *Frontiers in Human Neuroscience*, 2013)

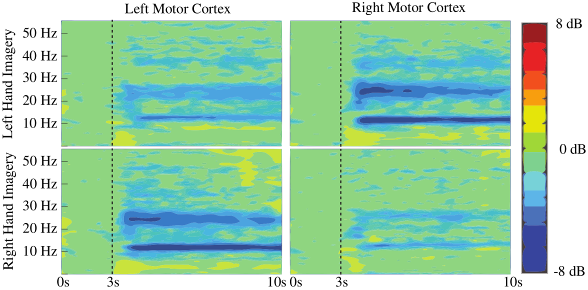
Reducing the number of time points (T)

The brain cares about oscillations:

- $\mathrm{DTFT}(y_i[t], \omega) = 1/T \sum_{t=0}^{T-1} y_i[t] e^{-j\omega t}$
- $z_i[\omega] = \log|\mathrm{DTFT}(y_i[t], \omega)|$
- $\mathcal{Z}_i = \{z_i[\delta], z_i[\theta], z_i[\alpha], z_i[\beta], z_i[\gamma]\}$



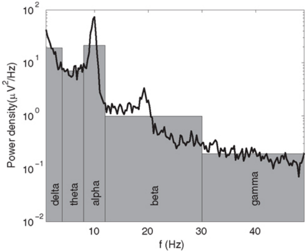
(van Albada & Robinson, *Frontiers in Human Neuroscience*, 2013)



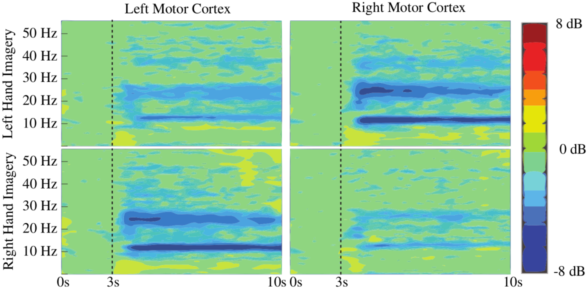
Reducing the number of time points (T)

The brain cares about oscillations:

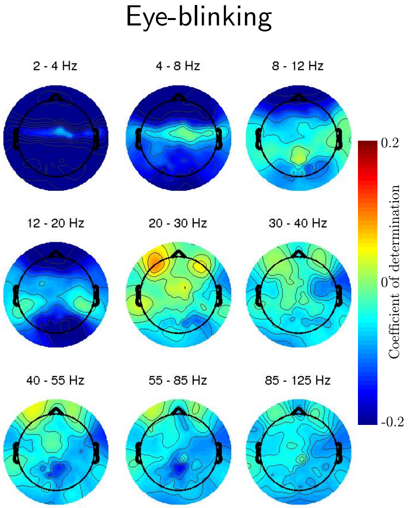
- $\text{DTFT}(y_i[t], \omega) = 1/T \sum_{t=0}^{T-1} y_i[t] e^{-j\omega t}$
- $z_i[\omega] = \log|\text{DTFT}(y_i[t], \omega)|$
- $\mathcal{Z}_i = \{z_i[\delta], z_i[\theta], z_i[\alpha], z_i[\beta], z_i[\gamma]\}$
- For log-bandpower features, linear decoders appear sufficient.

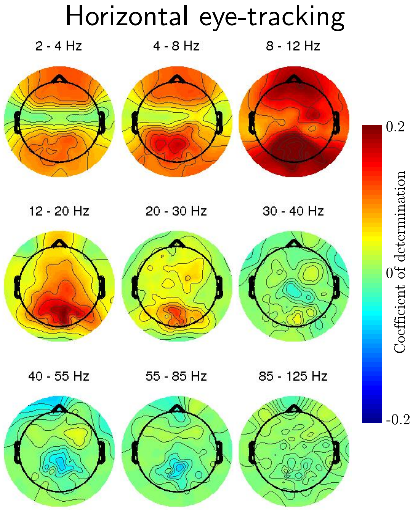


(van Albada & Robinson, *Frontiers in Human Neuroscience*, 2013)

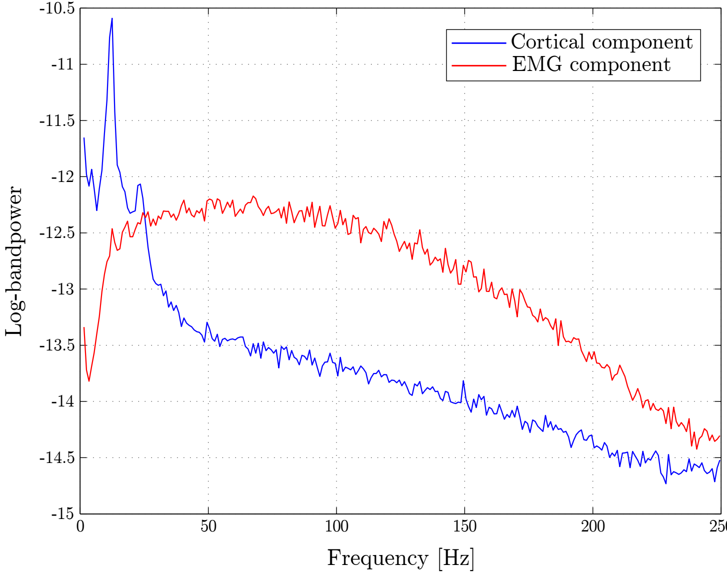


Confounding by EOG-artifacts



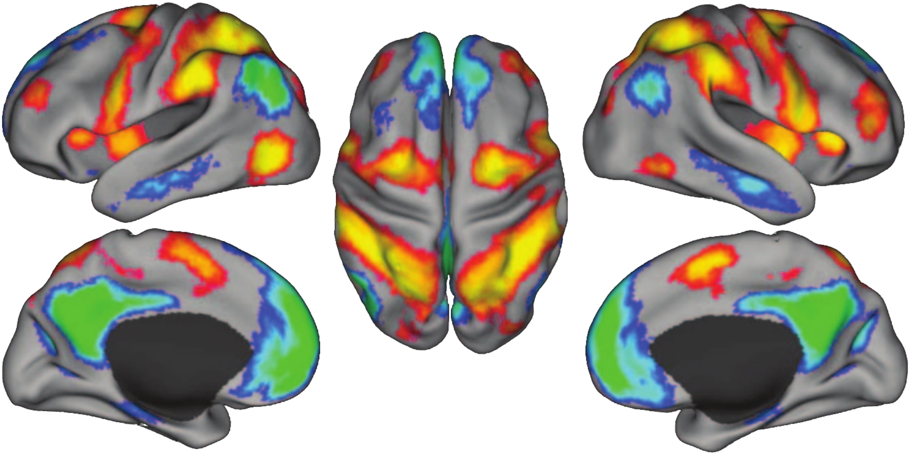


Confounding by EMG-artifacts



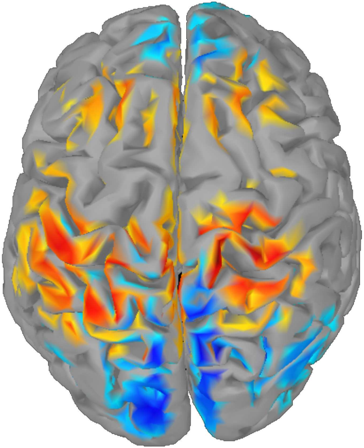
Outline

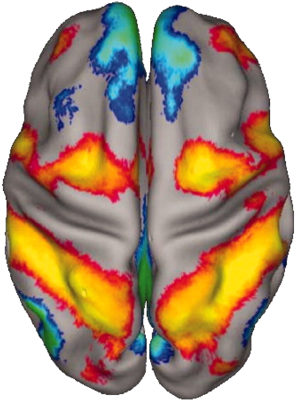
Large-scale brain networks



(Adapted from Fox et al., 2005)

Can large-scale cortical networks be observed in the EEG?



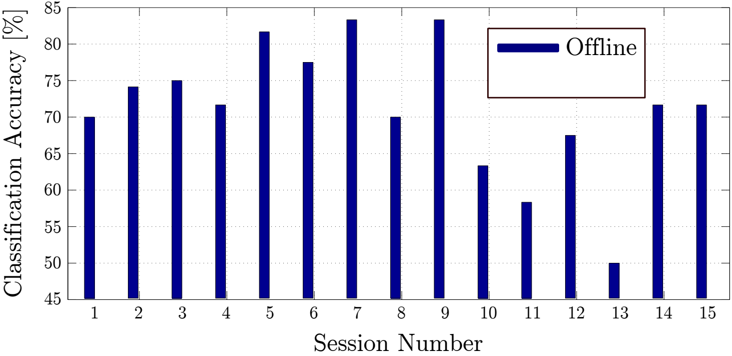


(Fox et al., PNAS, 2005)

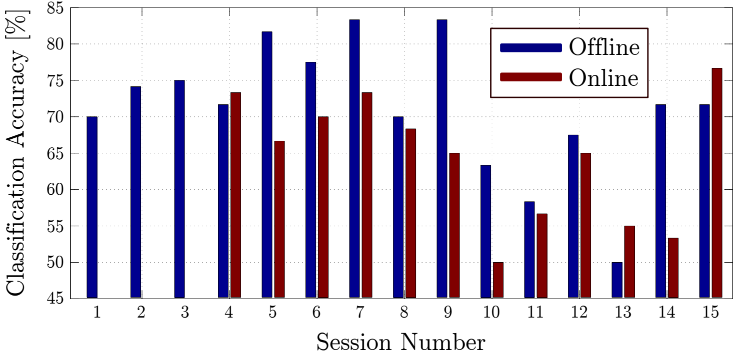
(Grosse-Wentrup & Schölkopf, High Gamma-Power Predicts Performance in SMR BCIs, *Journal of Neural Engineering*, 2012)

Patient GH: Neurofeedback of parietal δ-power

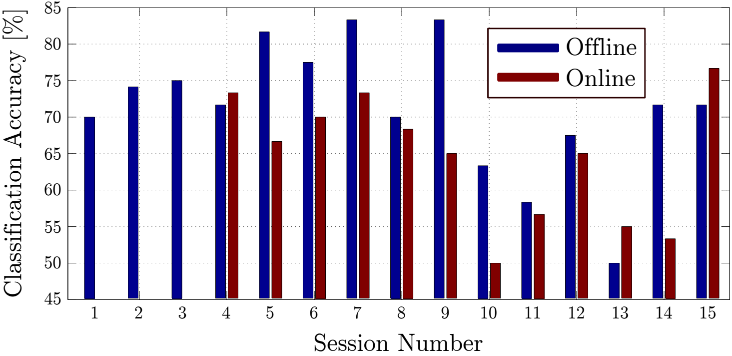
Patient GH: Neurofeedback of parietal δ-power



Patient GH: Neurofeedback of parietal δ-power

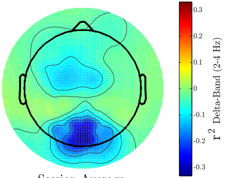


Patient GH: Neurofeedback of parietal δ-power



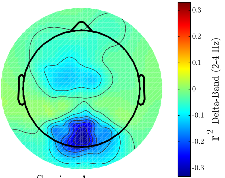
Mean offline/online decoding accuracy: 71.3%/64.4%

Patient GH: fMRI-study

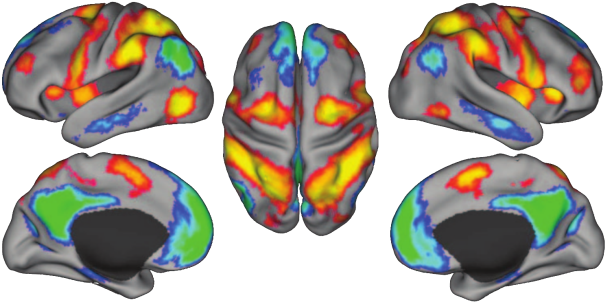


Session Average

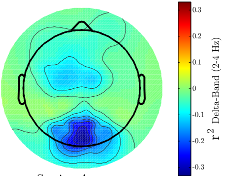
Patient GH: fMRI-study



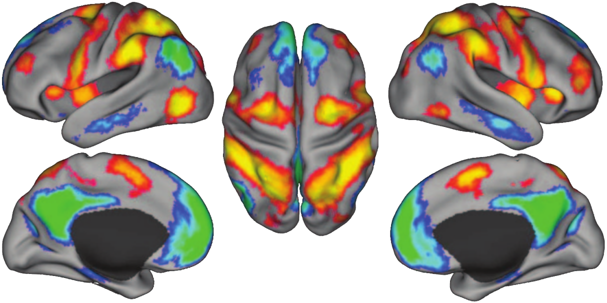
Session Average

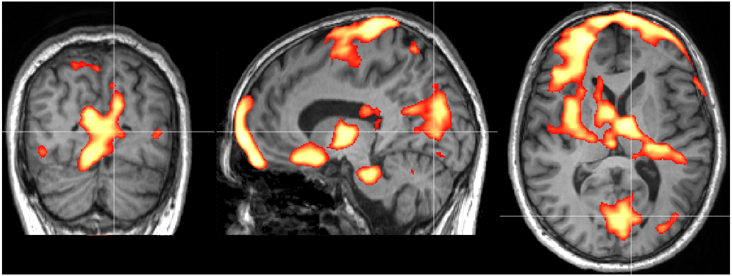


Patient GH: fMRI-study



Session Average





Acknowledgements

Max Planck Institutes:

- Bernd Battes
- Alexander Bretin
- Tatiana Fomina
- Christian Förster
- Marius Klug
- Gabriele Lohmann
- Natalie Widmann
- Bernhard Schölkopf
- Nadine Simon
- Sebastian Weichwald

University of Tübingen:

- Michael Erb
- Thomas Ethofer

http://brain-computer-interfaces.net