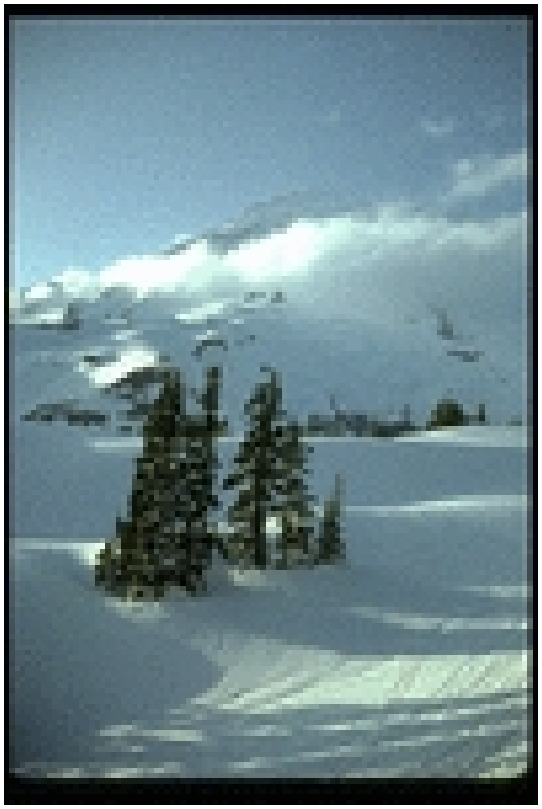
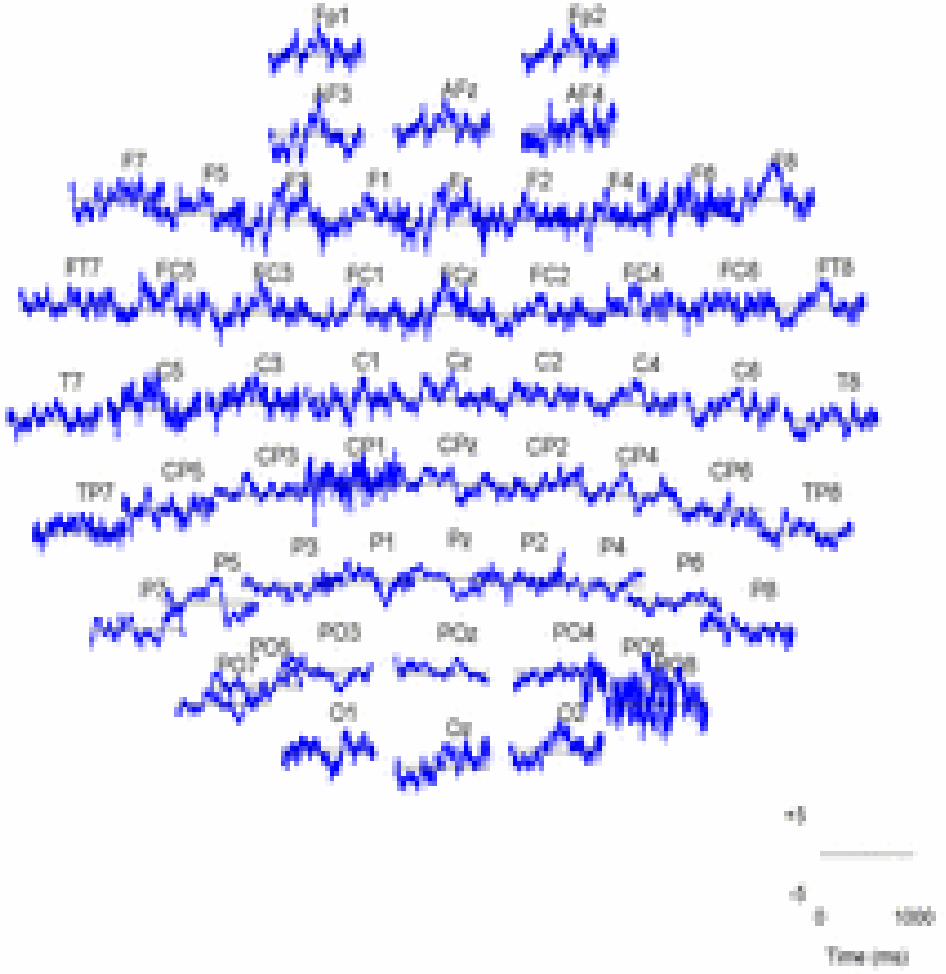
Tutotial: EEG linear decoding and spatial filtering

Lucas C. Parra
parra@ccny.cuny.edu
http://parralab.org

Neuro-Engineering Laboratory
City University of New York

Event related potential (ERP)





Content

Preliminaries
Evoked responses
Anatomical models
Spatial filters and components

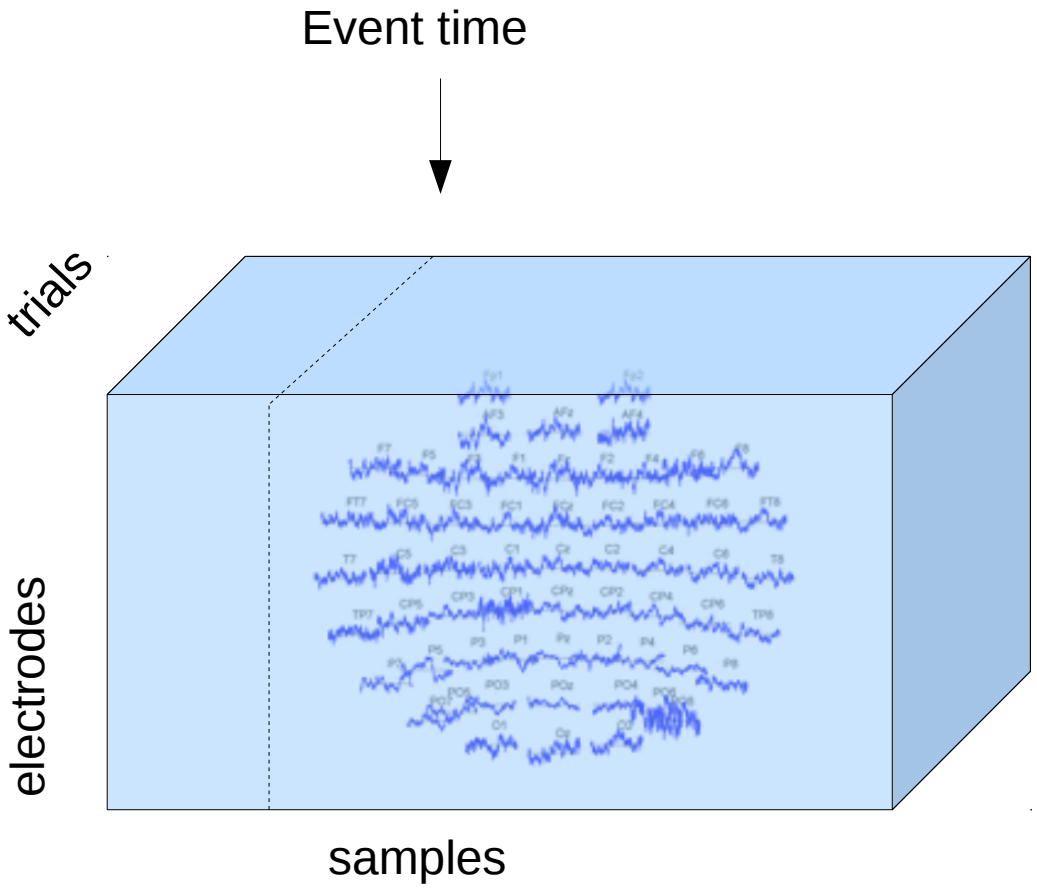
Single component models
Matched filter
Cross validation*
Maximum ERP effect size
ERP difference
Linear discriminat
Logitic regression
Temporal filter (encoding model)
EOG removal (noise canceling)*

Multi-component models
Maximum power (PCA)
Maximum power ratio (CSP)
Maximum power difference
Independent components (ICA)
Maximum SNR in power (DSS)
Maximum inter-subject correlation (CCA-1)
Maximum stimulus-repose correlation (CCA-2)

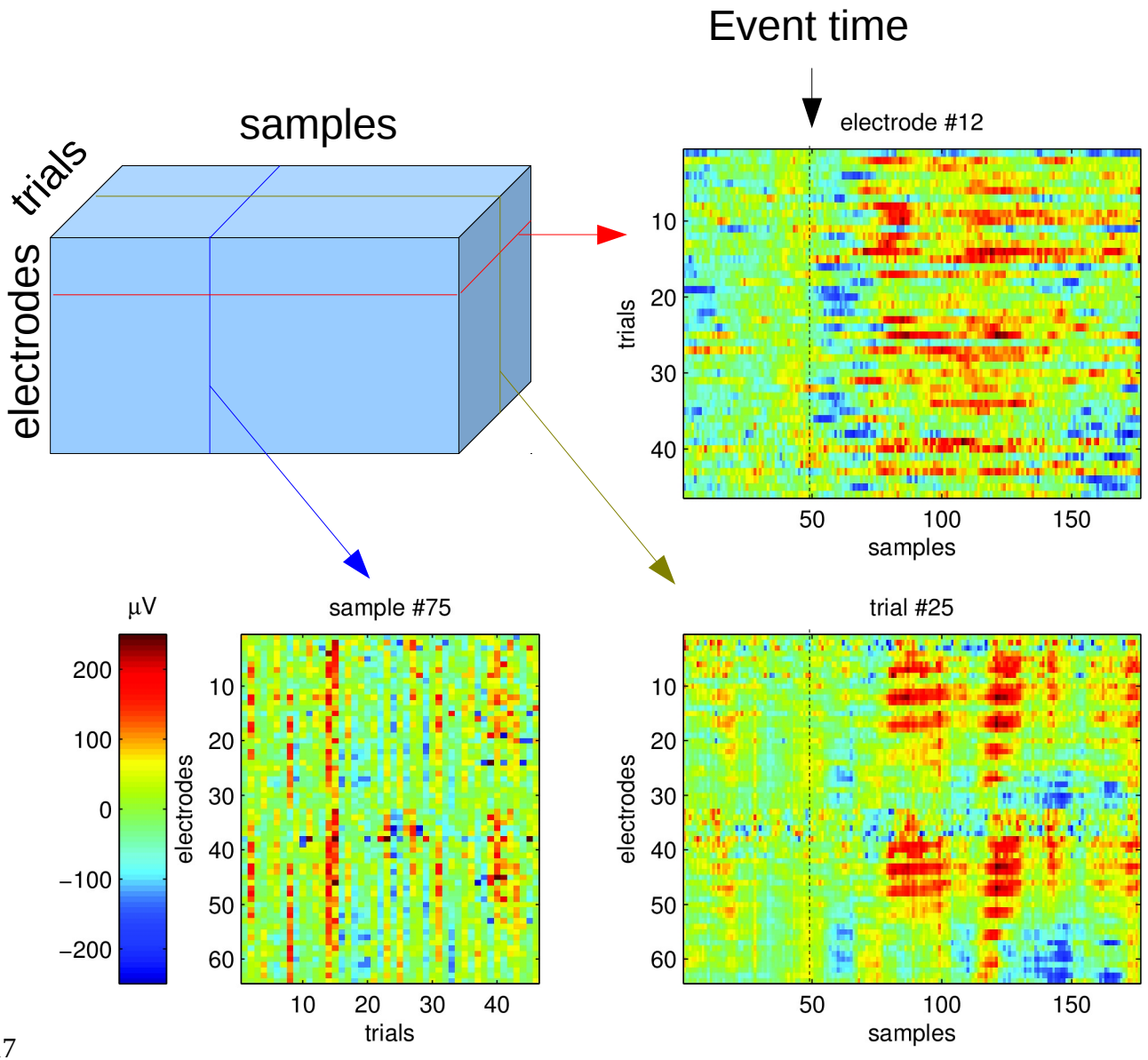
Robust PCA*
Shrinkage regularization*
Shuffle statistics*

*** Other useful stuff**

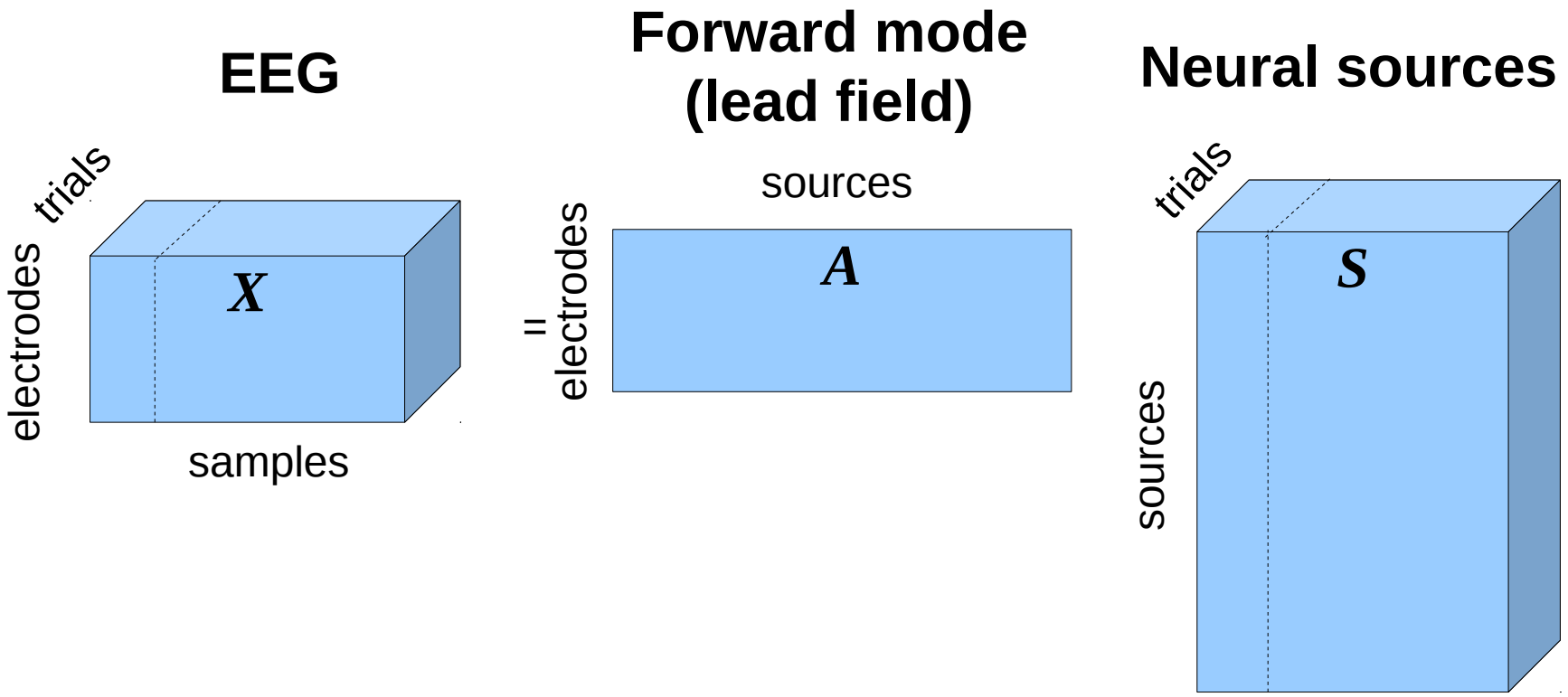
Event related potential (ERP)



Event related potential (ERP)



Where EEG comes from



$X = A * S$

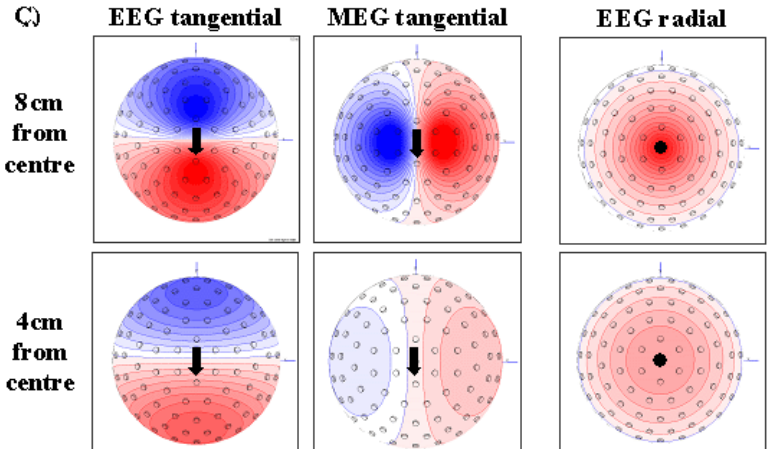
Electrical potentials on the skull surface

Electric tissue resistance

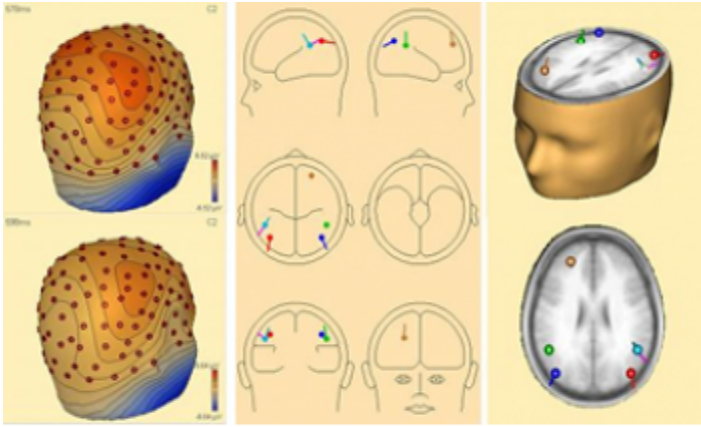
Neuronal current sources in the brain

Anatomical models

- One can make assumptions about anatomical origin of currents and compute $\boldsymbol{A}$.
- Simple forward models assume dipoles in a spherical head.
- Modern techniques assume dipoles or distributed activity in realistic 3D anatomy.



(c) Olaf Hauk , MRC Cognition and Brain Sciences Unit, http://www.mrc-cbu.cam.ac.uk/research/eeg/eeg_intro.html

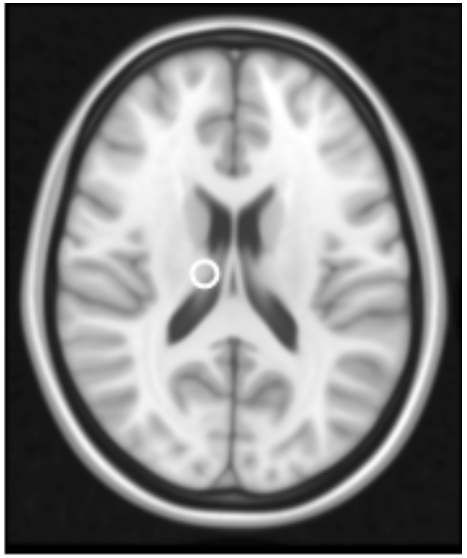


http://www.besa.de/

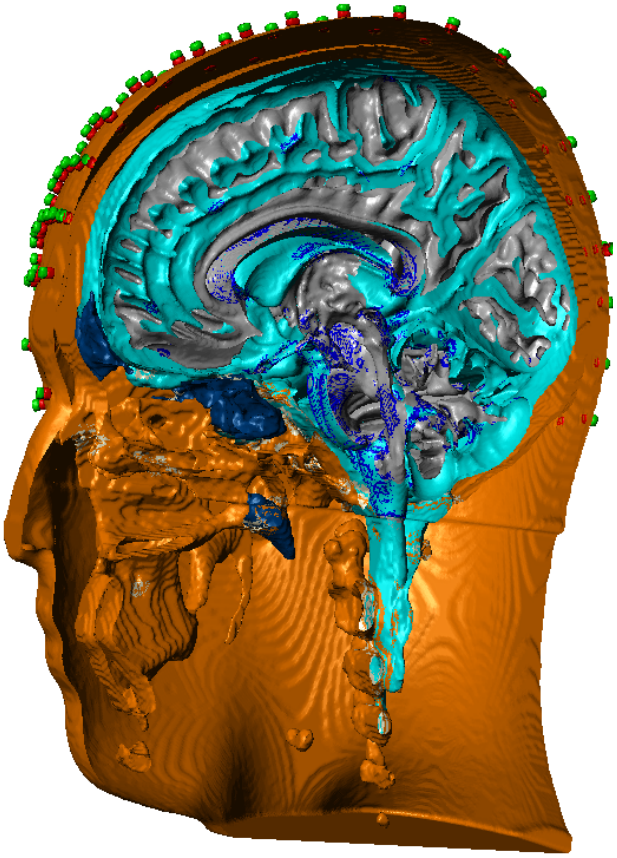
The New York Head

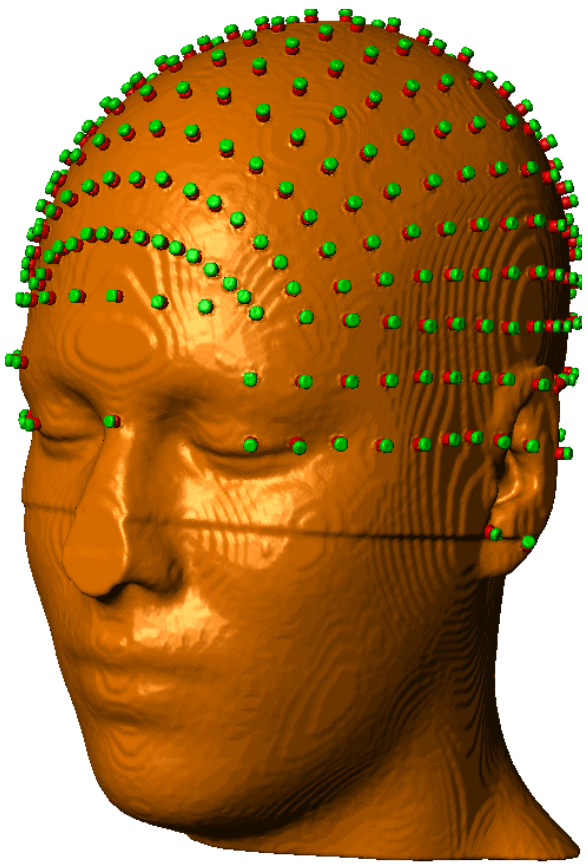
MNI-152 with extended field of view

MNI-152 (2009b)



CABI-25
(Chris Rorden)





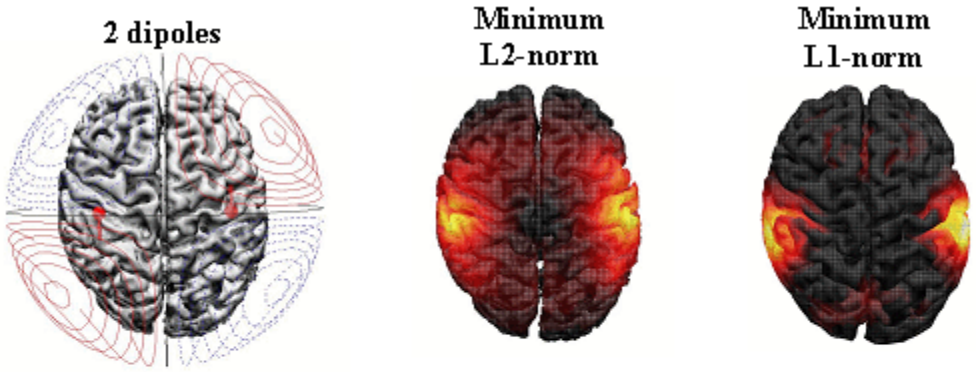
http://www.parralab.org/nyhead/

Huang, Neuroimage, 2016

Inverse modeling

$$\boldsymbol{S}(t) = \boldsymbol{A}^{-1}\boldsymbol{X}(t)$$

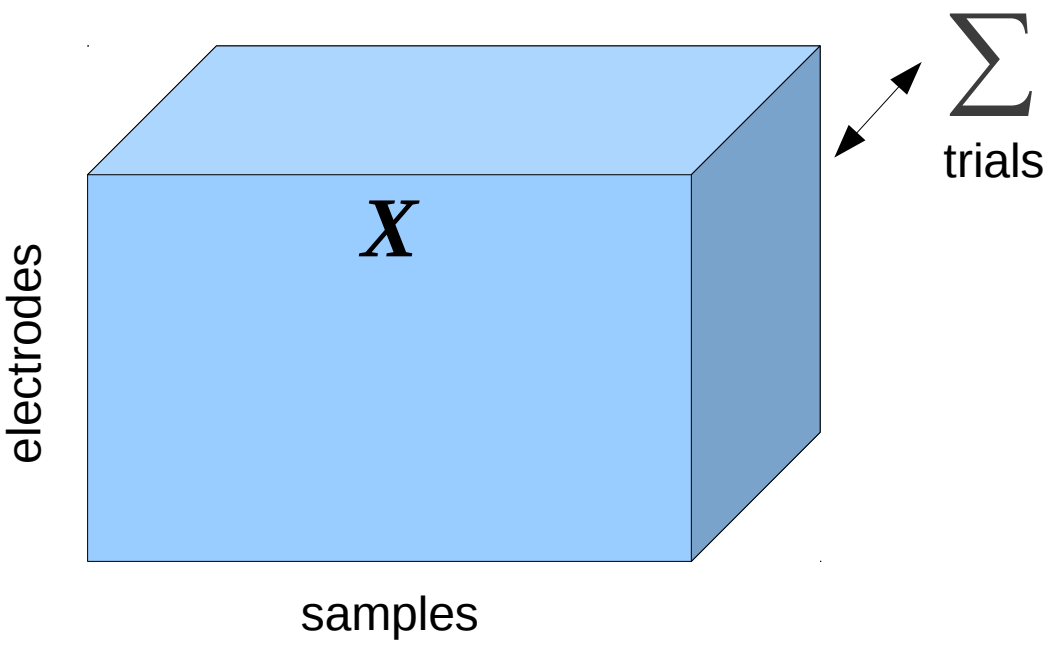
- However, forward model $\boldsymbol{A}$ is not invertible. Computing $\boldsymbol{S}(t)$ from $\boldsymbol{A}$ and $\boldsymbol{X}(t)$ is not possible without additional assumptions on anatomy or sources.

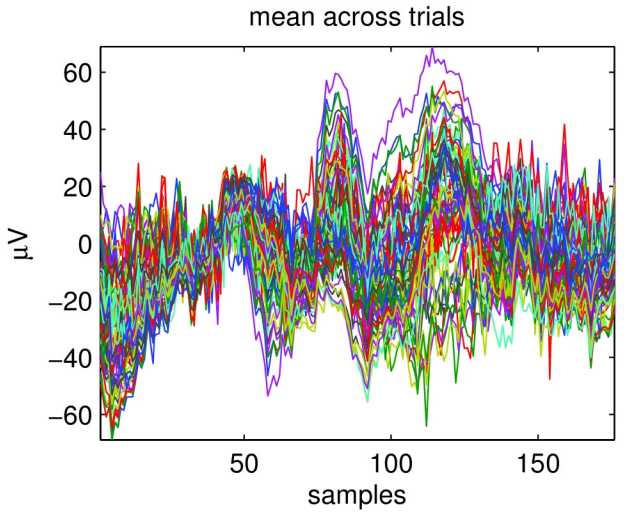


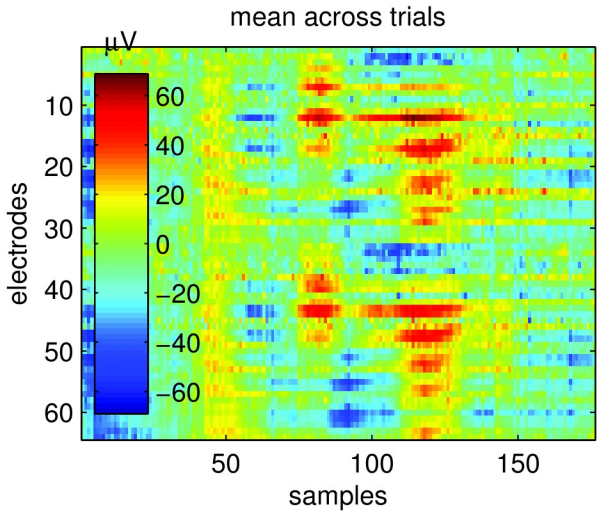
(c) Olaf Hauk , MRC Cognition and Brain Sciences Unit,
http://www.mrc-cbu.cam.ac.uk/research/eeg/eeg_intro.html

- Despite this ambiguity, some feel confident enough to use such ***inverse modeling*** routinely.

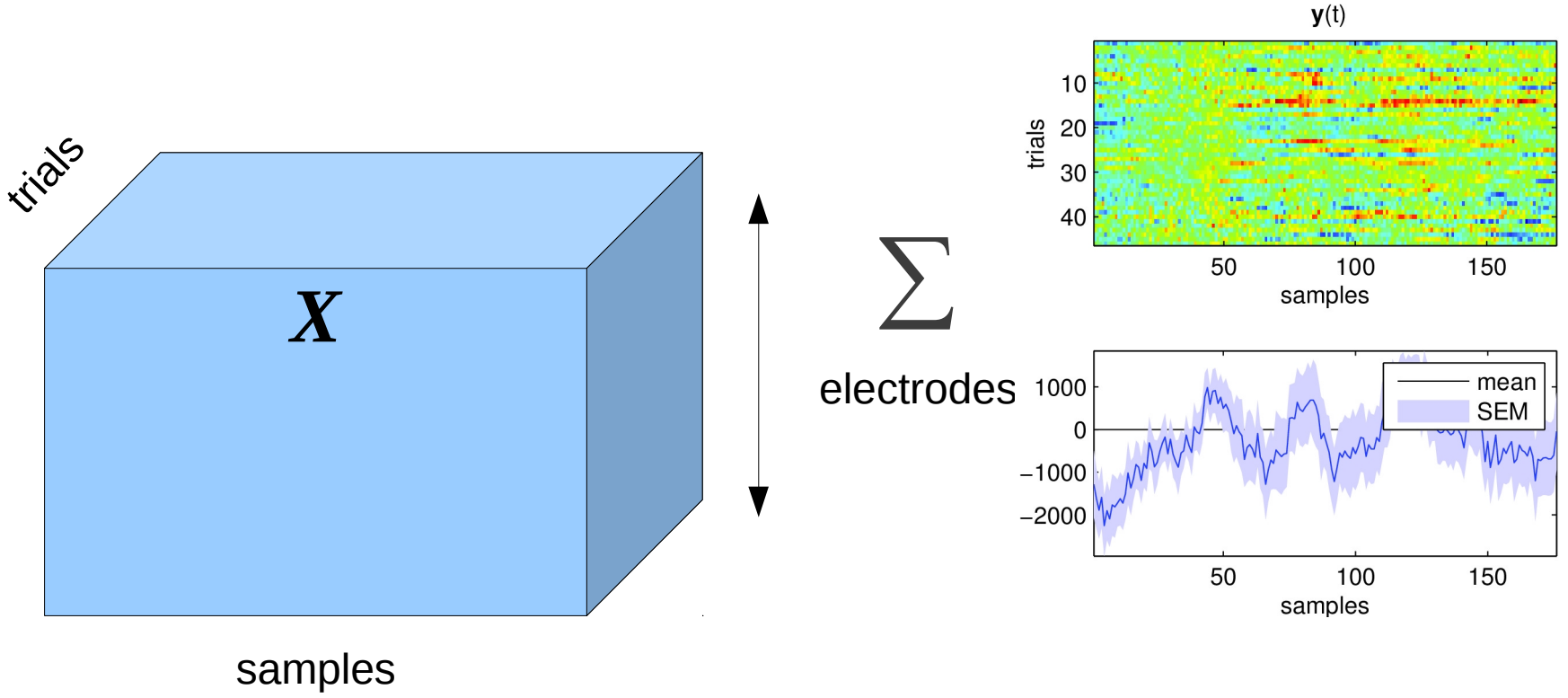
Trial-averaged ERP







Space-averaged ERP

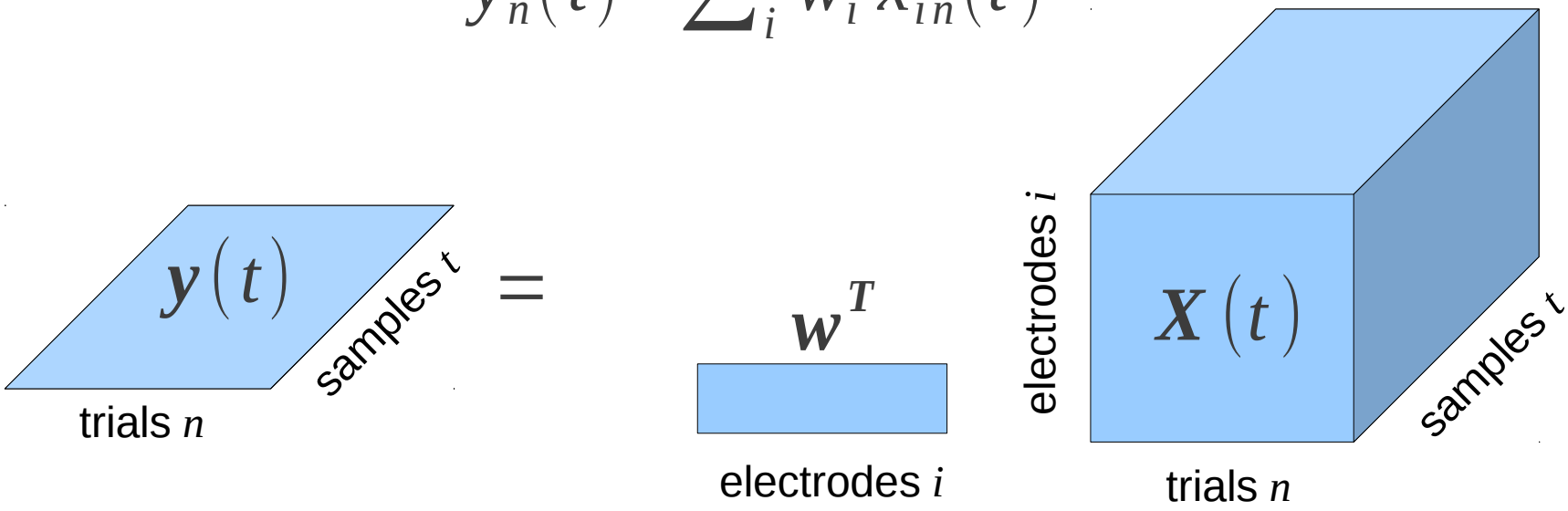


However, not all electrodes have the same sign!

Spatially filtered ERP

- "Averaging" should at least respect the sign.
- More generally, could use "filter" with weights $\boldsymbol{w}$:

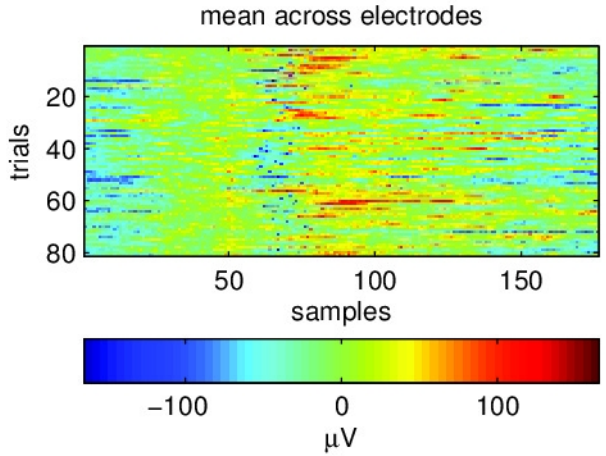
$$y_n(t) = \sum_i w_i \, x_{in}(t)$$

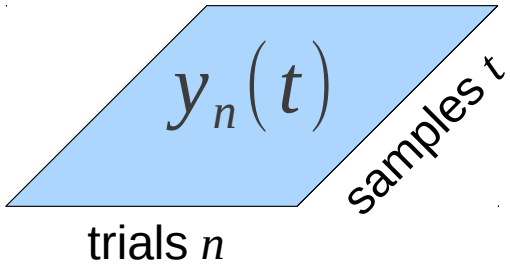


$$\boldsymbol{y}(t) = \boldsymbol{w}^T \, \boldsymbol{X}(t)$$

Parra, *Neuroimage*, 2005

Spatially filtered ERP





Advantages:

- Improved SNR increases statistical power.
- Improved SNR may allow single-trial analysis.
- Single component solves multiple comparison problem.
- Different criteria for picking $\boldsymbol{w}$ may capture different “sources” in the brain.

Techniques to picking weights w

$$y(t) = w^T * X(t)$$

Single w:

Mean or mean difference (Matched Filter)
Maximum effect size (Fisher Linear Discriminant)
Discriminant robust to outliers (Logistic regression)

Temporal filter $w(t)$:

Conventional regression (VESPA)

Several W:

Maximum power (Principal Component Analysis, PCA)
Maximum power ratio (Common Spatial Pattern, CSP)
Maximum correlation across repeats (CCA)
Independence (Independent Component Analysis, ICA)
Sources with less temporal noise (Denoising SS)

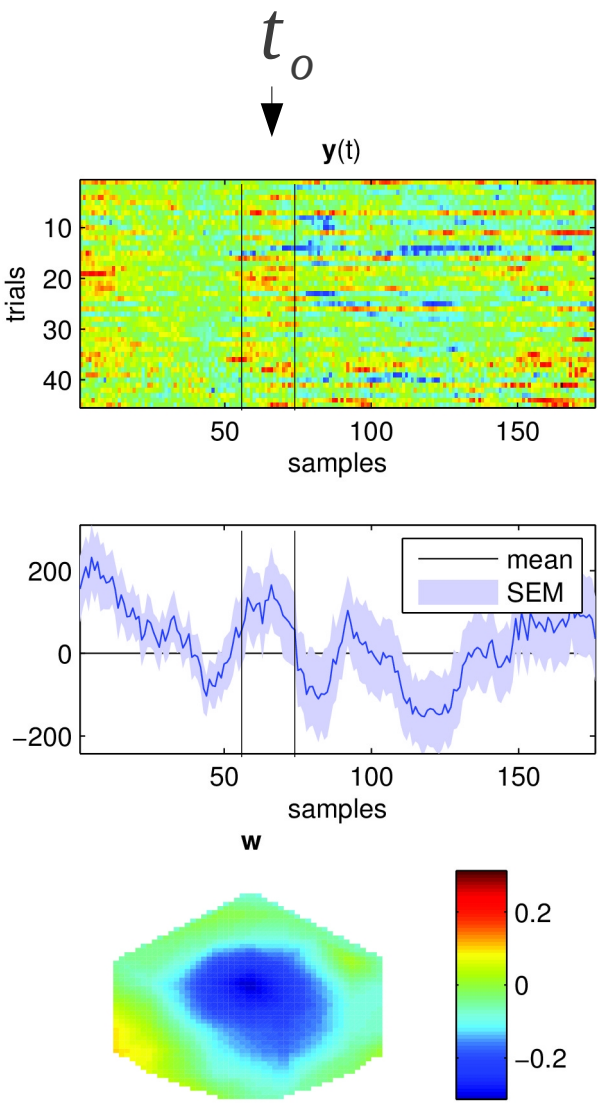
Matched filter

Pick the weights to be the activity at a given time t_o averaged over trials n:

$$w_i = \frac{1}{N}\sum_{n=1}^{N} x_{ni}(t_o)$$

$$\boldsymbol{w} = \bar{\boldsymbol{x}}(t_o)$$

Electrodes with a positive or negative mean, both contribute positively to the weighted spatial average, and their contribution is stronger if the mean is strong.

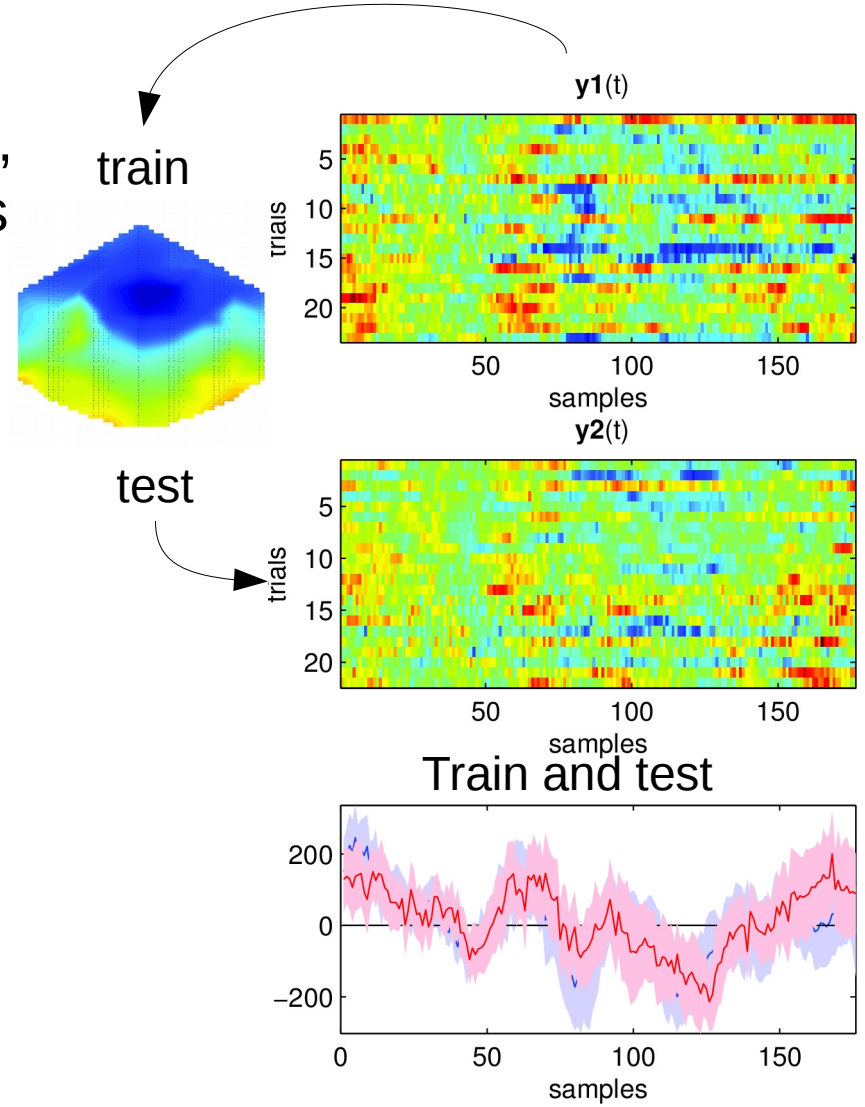


Parra, *Neuroimage*, 2005

Cross Validation

Weights are picked from the data, Is this not just highlighting what is already in the data, e.g. if it was noise, would we not just emphasize noise?

This is a well known problem called “over training”. Is can be simply addressed with cross-validation: $\boldsymbol{w}$ is formed from one part of the data, and significance (effect/variance) is tested on the rest of “unseen data”.



Cross validation can be used to validate all subsequent methods.

Maximum effect size

When looking for an effect often one evaluates the t-statistic which measures mean over std error (Student t-test).

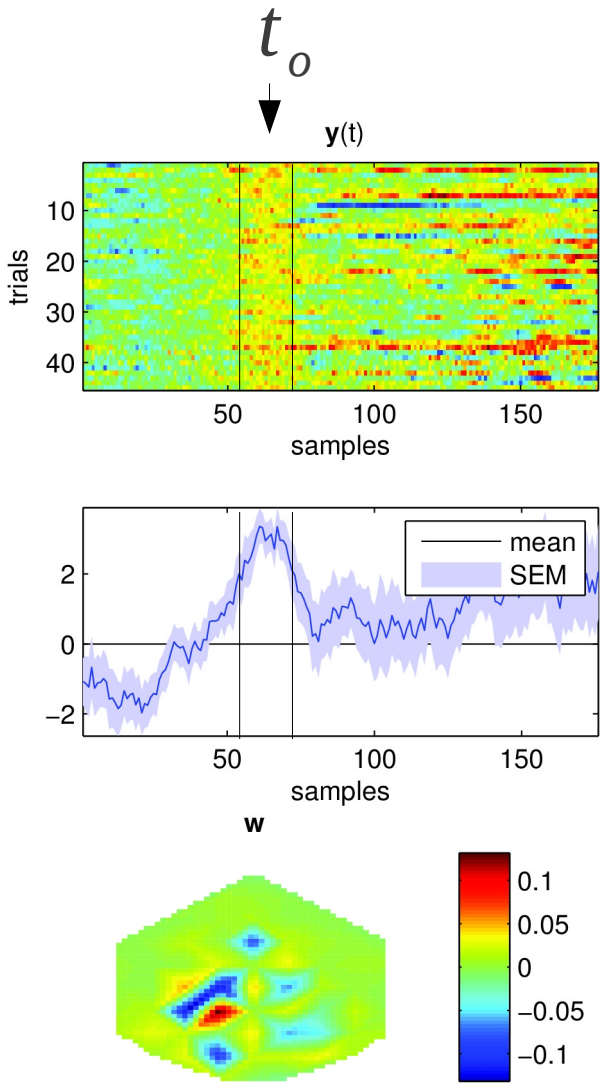
$$t = \sigma_{\bar{y}}^{-1} \bar{y}$$

Maximal t-statistic is achieved with

$$w = \boldsymbol{R}_{xx}^{-1} \bar{x}$$

where $\bar{x}$, $\boldsymbol{R}_{xx}$ are the mean and covariance of the activity of interest.

This maximizes the effect size!



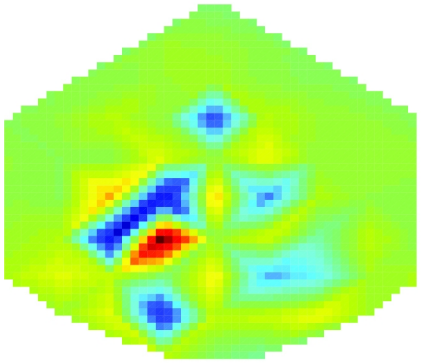
Parra, *Neuroimage*, 2005

Forward from backward model

Backward model: Projection ***w*** takes one from the sensor data to a putative source *y*(*t*) in the brain.

$$\boldsymbol{S}(t) = \boldsymbol{A}^{-1}\boldsymbol{X}(t)$$

backward model



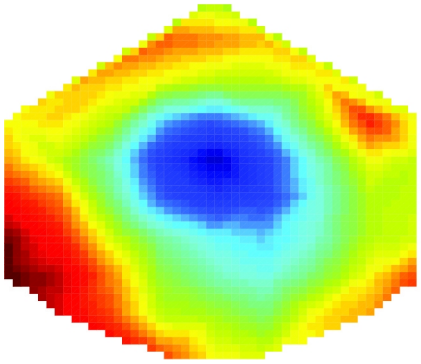
Forward model: To know how the activity in the brain looks on the scalp one needs the "forward" model. Namely, the projection ***a*** that take a current source ***y***(*t*) in the brain and "generates" the measurement ***X***(*t*):

$$\boldsymbol{X}(t) = \boldsymbol{a}\,\boldsymbol{y}(t)$$

Very loosely speaking that can be estimated as

$$\boldsymbol{a} = \boldsymbol{X}(t) / \boldsymbol{y}(t)$$

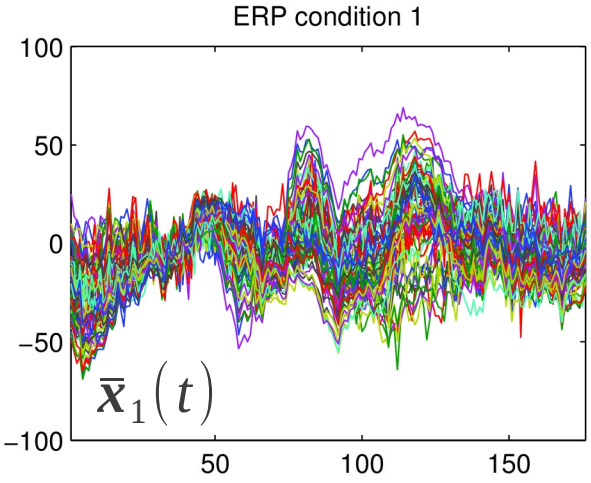
forward model

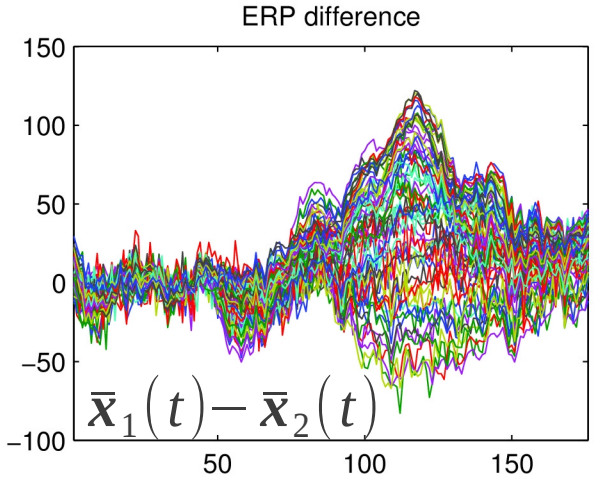


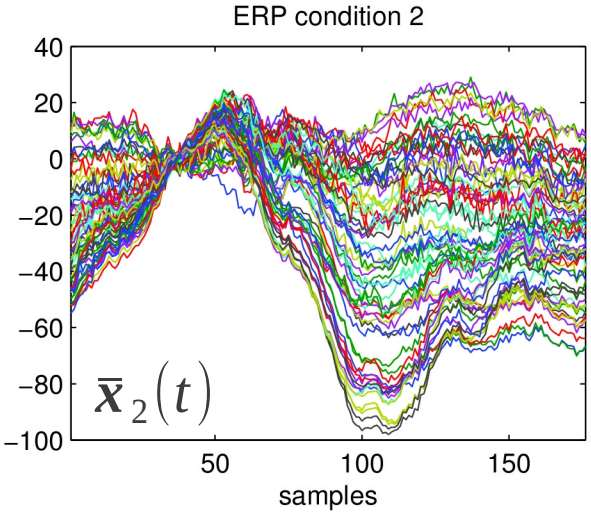
Parra, *Neuroimage*, 2005

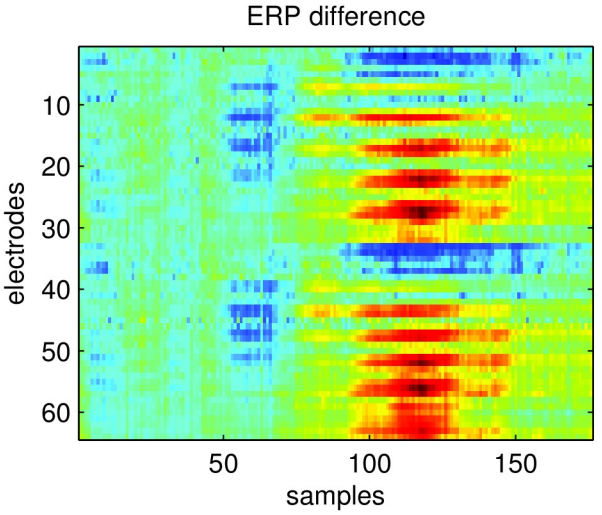
Difference of two conditions

Sometimes experiments consist of two conditions and we are only interested in the activity that is different.





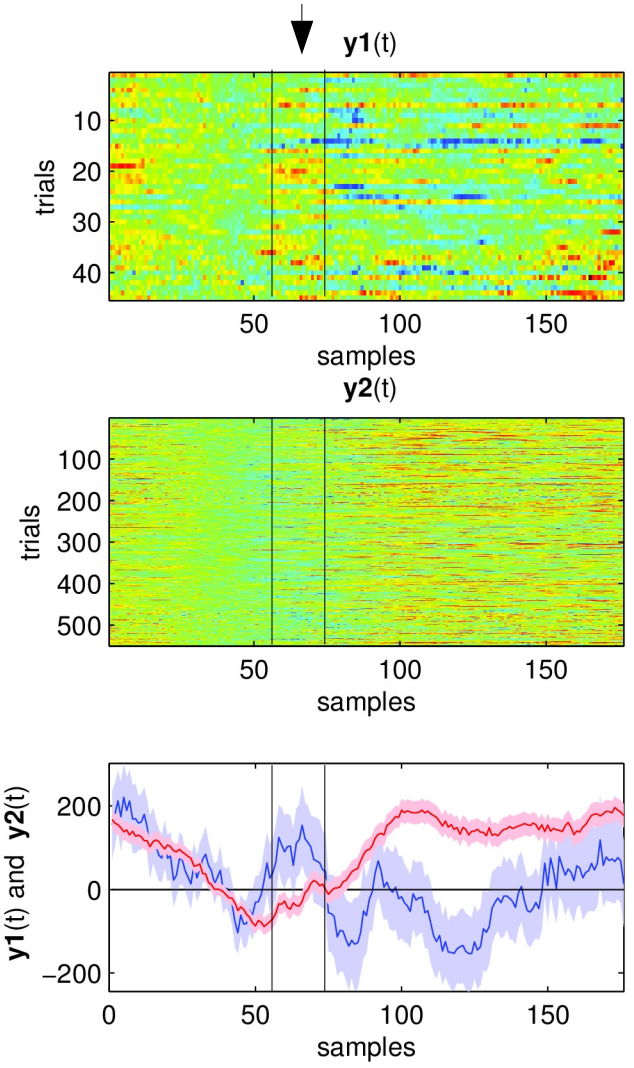




Matched filter for difference

Same as before, but now take the ***difference*** averaged across trials at a time of interest

$$w = \bar{x}_1 - \bar{x}_2$$



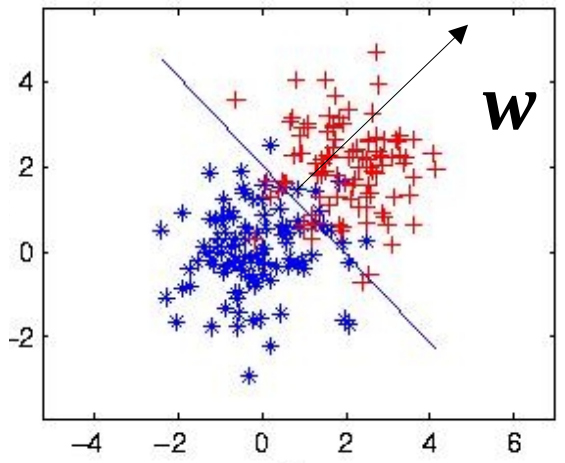
Maximum effect size (FLD)

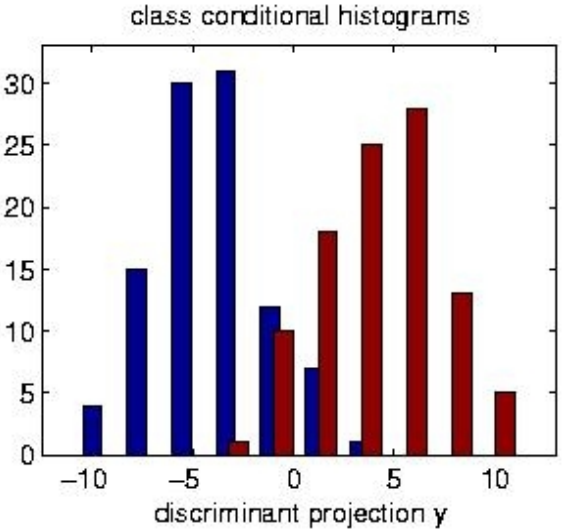
If the effect we are looking for is the difference between conditions then the same criterion of maximum t-statistic is given by the **Fisher Linear Discriminant (FLD)**:

$$w = R_{xx}^{-1}\left(\bar{x}_1 - \bar{x}_2\right)$$

Where now R_{xx} is the "pooled covariance".

FLD gives the projection of the data with the largest effect size!

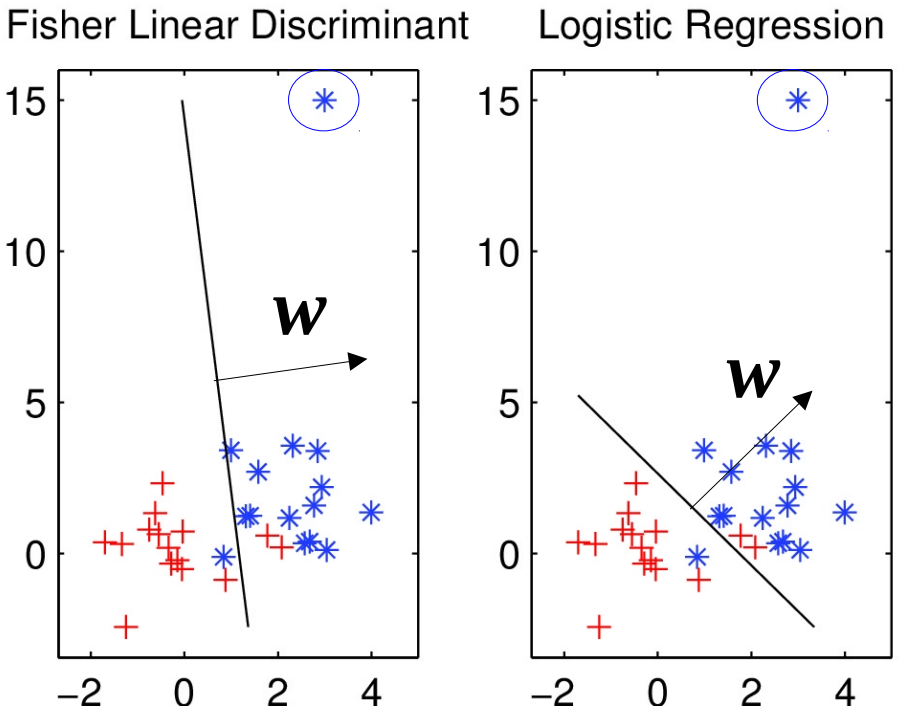




Parra, *Neuroimage*, 2005

Separation robust to outliers

EEG is very noisy, leading to noisy estimates of covariance which is particularly sensitive to outliers. Better to use a technique that finds the direction based on the boundary.



Logistic regression

$$w = logist\left(X_1, X_2\right)$$

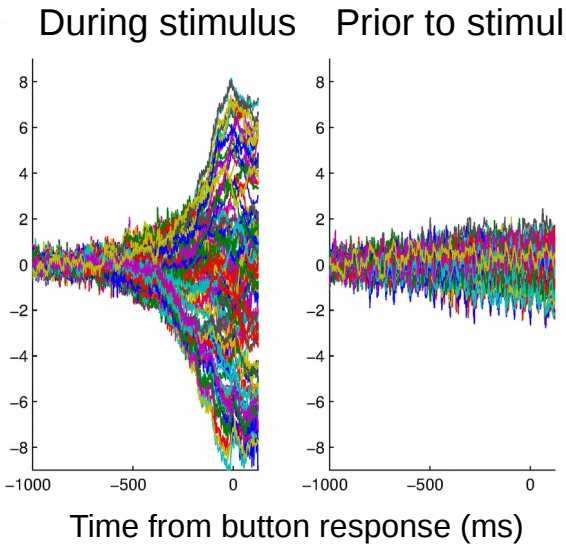
- Insensitive to points far from the boundary.
- Assumes “soft” transition, thus insensitive to noise and boundary.

Parra, *Neuroimage*, 2005

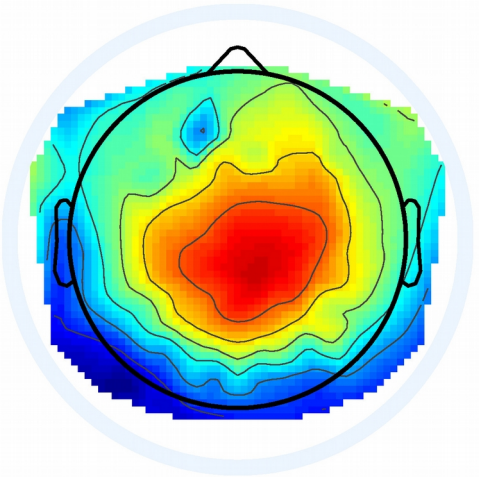
Example: Evidence accumulation

Task: detect a target that is slowly fading-in

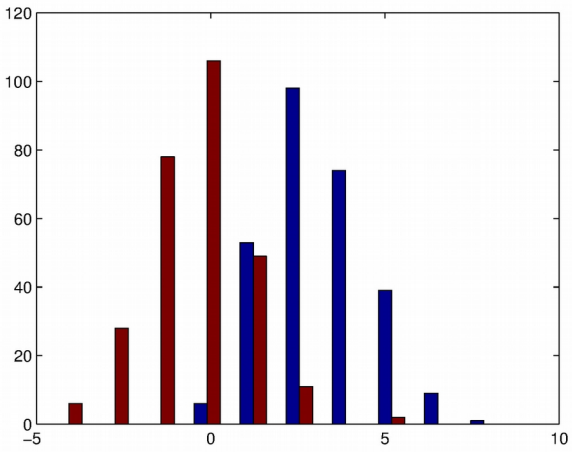
ERP

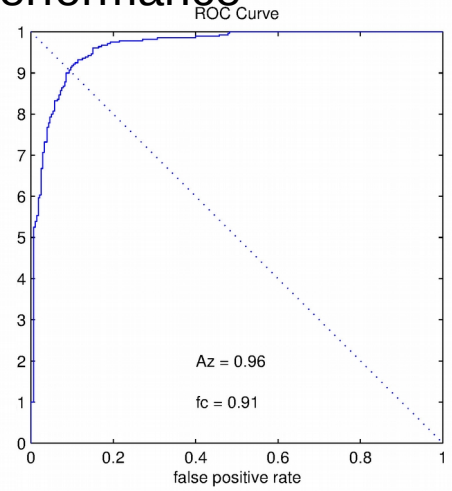


Forward model



Discrimination performance





Techniques to picking weights w

$$y(t) = w^T * X(t)$$

Single w:

Mean or mean difference (Matched Filter)

Maximum effect size (Fisher Linear Discriminant)

Discriminant robust to outliers (Logistic regression)

Temporal filter $w(t)$:

Conventional regression (LMS, VESPA)

Several W:

Maximum power (Principal Component Analysis, PCA)

Maximum power ratio (Common Spatial Pattern, CSP)

Maximum correlation across repeats/subjects (CCA-1)

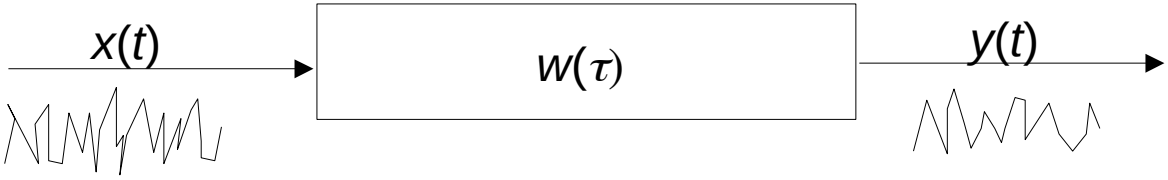
Independence (Independent Component Analysis, ICA)

Sources with less noise (DSS)

Best correlated with stimulus (CCA-2)

Temporal filter – linear encoding model

$$\boldsymbol{y}(t) = \boldsymbol{w}^{\boldsymbol{T}} * \boldsymbol{X}(t) = \sum_{k=0}^{Q} \boldsymbol{w}^{\boldsymbol{T}}[k]\boldsymbol{x}[n-k]$$



EEG

Stimulus

This is textbook “Linear systems identification”

Linear system identification - textbook

In 1 dimensions (single filter)

$$\begin{bmatrix} y[1] \\ y[2] \\ y[3] \\ \dots \end{bmatrix} = \begin{bmatrix} x[1] & 0 & 0 \\ x[2] & x[1] & 0 \\ x[3] & x[2] & x[1] \\ & \dots & \end{bmatrix} \begin{bmatrix} w[0] \\ w[1] \\ w[2] \end{bmatrix}$$

$$\boldsymbol{y} = \boldsymbol{X}\,\boldsymbol{w}$$

$$\hat{\boldsymbol{w}} = \underset{\boldsymbol{w}}{argmin} \,||\boldsymbol{y} - \boldsymbol{X}\,\boldsymbol{w}||^2 = \boldsymbol{R}_{xx}^{-1}\,\boldsymbol{R}_{xy}$$

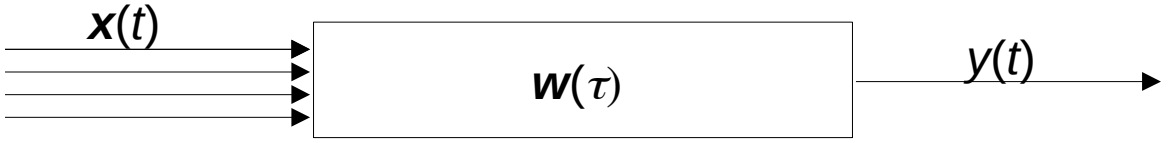
```
w = toeplitz(x,[x(1) zeros(1,Q)])\y;
```

Linear “encoding” model

In D dimensions (D filters)

EEG

Stimulus



$$\begin{bmatrix} y[1] \\ y[2] \\ y[3] \\ \dots \end{bmatrix} = \begin{bmatrix} x_1[1] & 0 & 0 & x_2[1] & 0 & 0 & \dots \\ x_1[2] & x_1[1] & 0 & x_2[2] & x_2[1] & 0 & \dots \\ x_1[3] & x_1[2] & x_1[1] & x_2[3] & x_2[2] & x_2[1] & \dots \\ & \dots & & \dots & & & \dots \end{bmatrix} \begin{bmatrix} w_1[0] \\ w_1[1] \\ w_1[2] \\ w_2[0] \\ w_2[1] \\ w_2[2] \\ \dots \end{bmatrix}$$

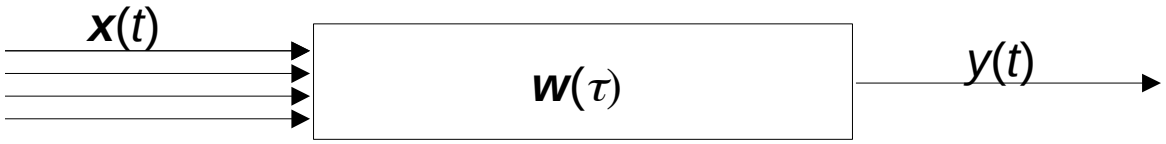
$$y = X\,w$$

$$\hat{w} = R_{xx}^{-1} R_{xy}$$

Lalor et al, Neuroimage 2006, “VESPA”
Sullivan et al, Cerebral Cortex, 2014

Linear “encoding” model

In D dimensions (D filters)

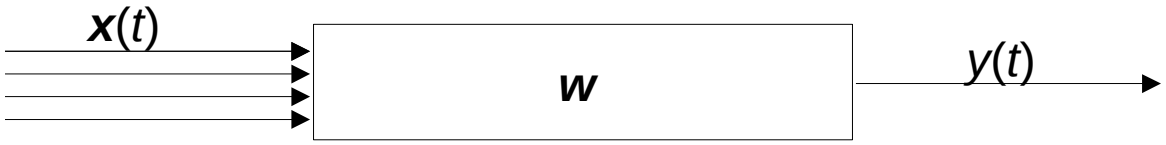


Potential problem: Too many parameters

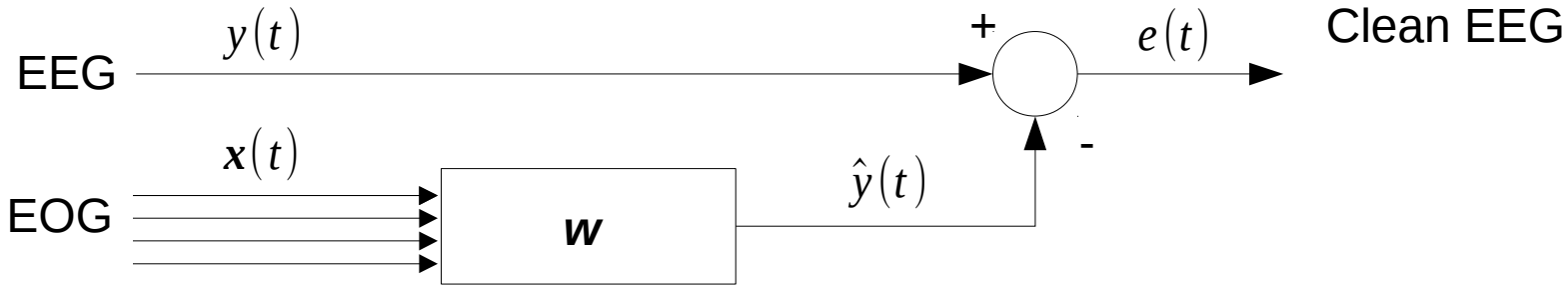
Standard Solutions:
L2 constraint: “Ridge regression”
L1 contraint: “LASSO”

Noise canceling – EOG removal

In D dimensions, but instantaneous



Works great for EOG removal



$$e = y - w\,x \qquad\qquad w = R_{xx}^{-1} R_{xy}$$

If data is arranged as samples by channels this line will generate clean version of EEG with EOG “regressed out”:

```
>> EEG = EEG - EOG * EOG\EEG;
```

Techniques to picking weights w

$$y(t) = w^T * X(t)$$

Single w:

Mean or mean difference (Matched Filter)

Maximum effect size (Fisher Linear Discriminant)

Discriminant robust to outliers (Logistic regression)

Temporal filter $w(t)$:

Conventional regression (LMS, VESPA)

Several W:

Maximum power (Principal Component Analysis, PCA)

Maximum power ratio (Common Spatial Pattern, CSP)

Maximum correlation across repeats/subjects (CCA-1)

Independence (Independent Component Analysis, ICA)

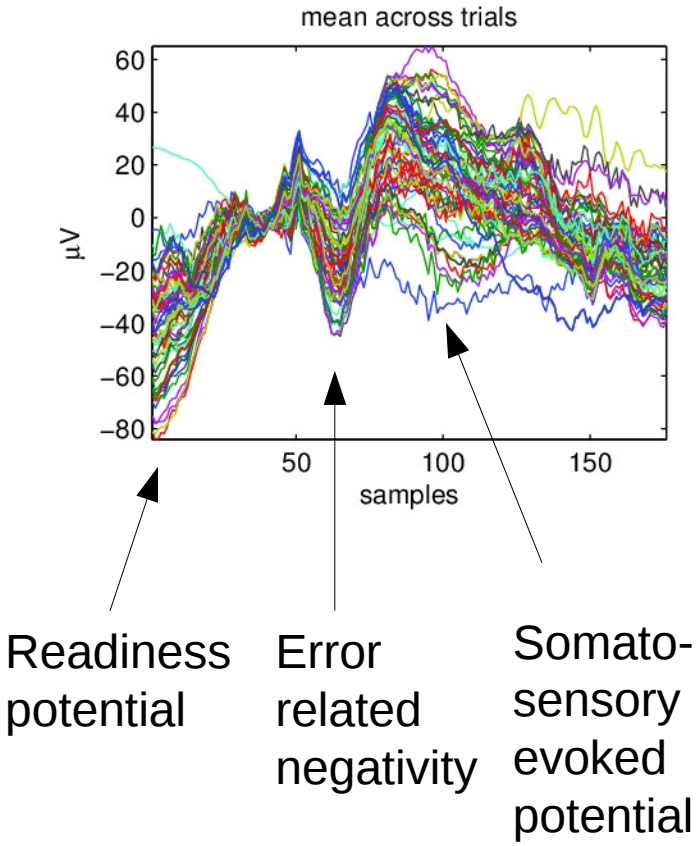
Sources with less noise (DSS)

Best correlated with stimulus (CCA-2)

Maximum power: PCA

- Typical recordings have more than one component with different spatial profile.
- They may be temporally overlapping.
- A common technique to capture all the “action” is to find component ***w*** which maximizes variance in source *y*(*t*).

$$\max_{\boldsymbol{w}} \sigma_y^2 = \max_{\boldsymbol{w}} \boldsymbol{w}^T \boldsymbol{R}_{xx} \boldsymbol{w}$$



Maximum power: PCA

- Once extracted, there may be other components that have still a lot of variance.
- To get spatial distributions that as different as possible one can assume that these components are *spatially orthogonal*.
- With that assumption they can all be found in a single step as solutions of an eigenvalue equation

$$\boldsymbol{R}_{xx}\,\boldsymbol{w} = \sigma^2 \boldsymbol{w}$$

$$\boldsymbol{W}^{-1} = \boldsymbol{W}^{T}$$

$$\boldsymbol{W} = eig\left(\boldsymbol{R}_{xx}\right)$$

Forward from backward model

Recall the “loose” definition of the forward model*

$$\boldsymbol{a} = \boldsymbol{X}\left(t\right) / \boldsymbol{y}\left(t\right)$$

When there is a set of components arranged as weight matrix ***W*** then the forward model from all “sources” to all sensors is also a matrix ***A***. If there are as many distinct sources as electrodes, then the estimate above simplifies to

$$\boldsymbol{A} = \boldsymbol{W}^{-1}$$

* ***a*** measures the correlation of the putative source activity with the sensors.

This definition has an arbitrary scaling, which may be fixed by setting |w|=1.

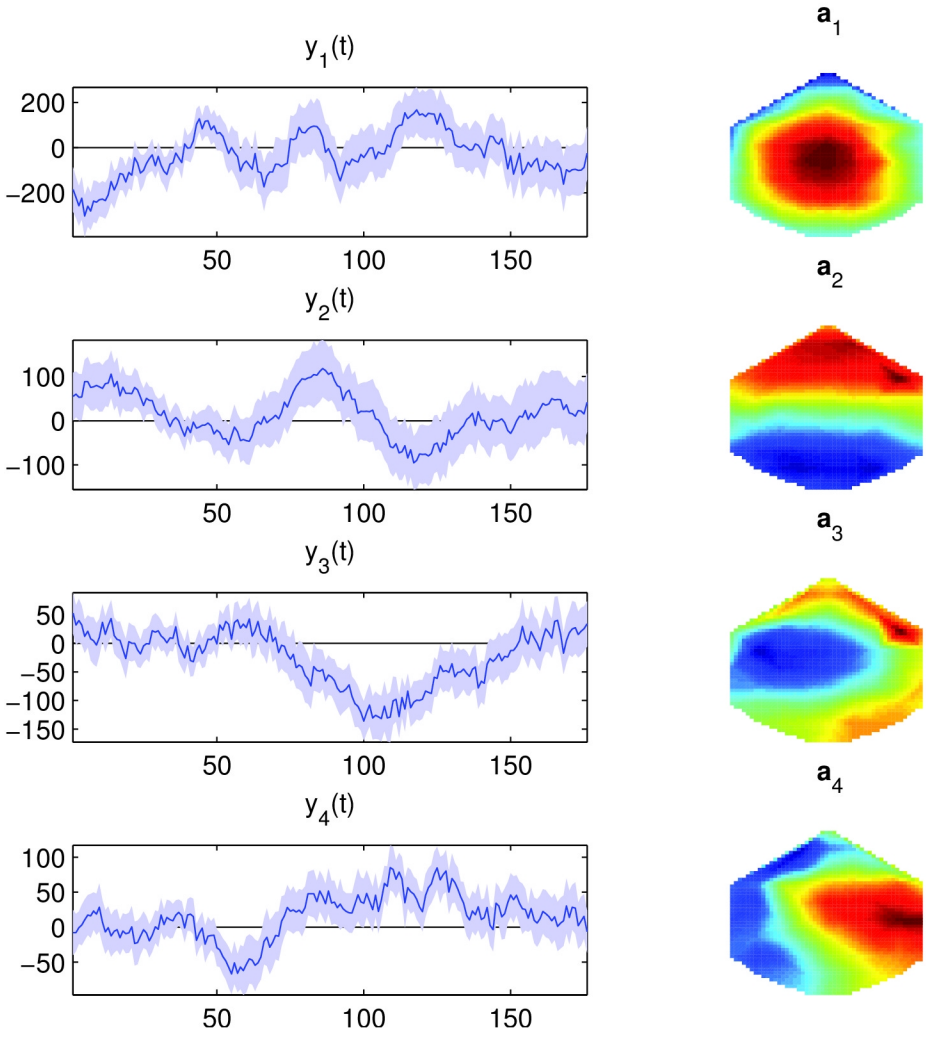
Parra, *Neuroimage*, 2005

Maximum power in decreasing order: PCA

$$W = eig\left(\boldsymbol{R}_{xx}\right)$$

$$y(t) = w^T X(t)$$

$$A = W^{-1}$$



Caveat: spatial orthogonality is meaningless in the brain.

Maximum power ratio: CSP

- Instead of maximum variance (or power) one may be interested in changes of power.
- In particular for oscillatory activity, where sign does not matter, all that once can measure is power of oscillation.
- One may be interested in components that change power in time, e.g. alpha “de-synchronization”

$$\max_{w} \frac{\sigma_y^2(t_1)}{\sigma_y^2(t_2)} = \max_{w} \frac{\boldsymbol{w}^T \boldsymbol{R}_{xx}(t_1)\boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{R}_{xx}(t_2)\boldsymbol{w}}$$

de Cheveigne, Parra, *Neuroimage*, 2014

Maximum power ratio: CSP

- Maximum power ratio is again given by an eigenvalue equation:

$$\boldsymbol{R}_{xx}^{-1}(t_2)\,\boldsymbol{R}_{xx}(t_1)\,\boldsymbol{w} = \lambda\,\boldsymbol{w}$$

- And again, after extracting the strongest, there are other components that also give a large ratio.

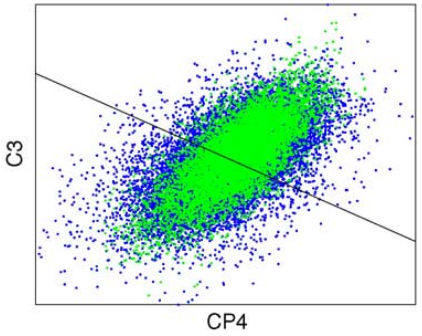
$$\boldsymbol{W} = eig\left(\boldsymbol{R}_{xx}(t_1), \boldsymbol{R}_{xx}(t_2)\right)$$

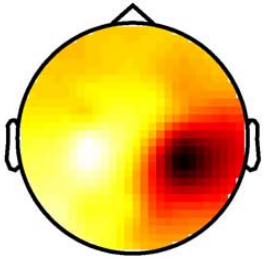
- They have been called “common spatial pattern” (CSP), because they are meaningful for both time intervals.

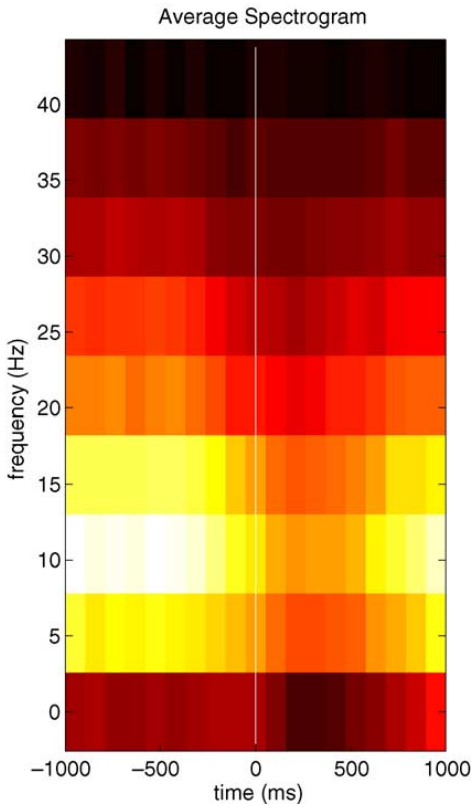
de Cheveigne, Parra, *Neuroimage*, 2014

Maximum power ratio: CSP

- CSP have been used for single-trial analysis of power as it leads to components with strong changes in power.
- Interestingly they do not need to be orthogonal.
- The approach is very similar to ICA. In fact, it can the thought as one version of blind source separation.







Parra, *Neuroimage*, 2005

Maximum power difference

Recall that covariance estimates can be noisy, so CSP are often very noisy. Dividing by a noisy estimate may be a bad idea. When we are looking for very small effects on power so that $\boldsymbol{R}(t_1)$ and $\mathbf{R}(t_2)$ are very similar one can use the difference in power instead of the ratio

$$\max_{\boldsymbol{w}} \left(\sigma_y^2(t_1) - \sigma_y^2(t_2) \right)$$

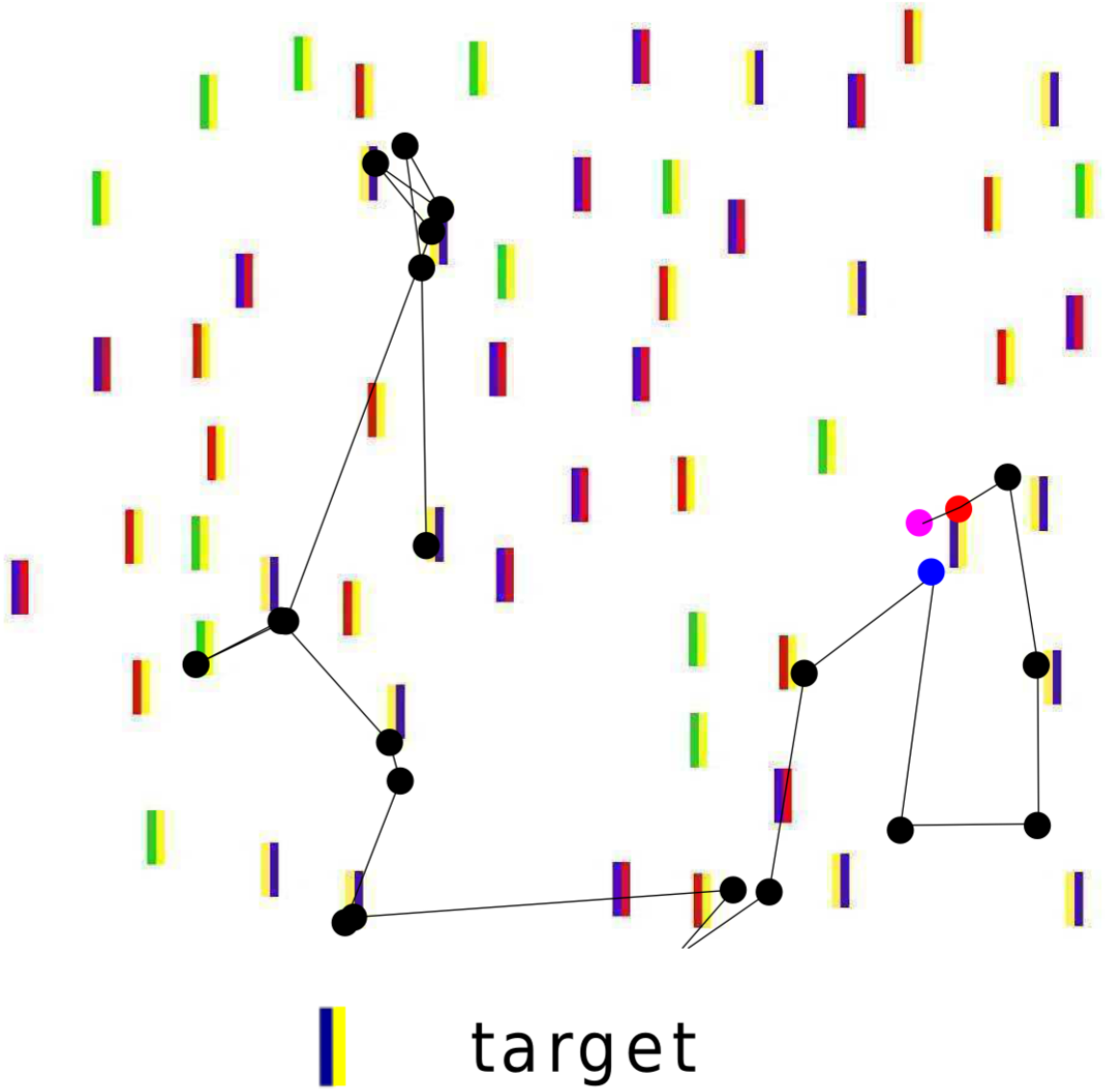
The solution of which is given by

$$\boldsymbol{W} = eig\left(\boldsymbol{R}_{xx}(t_1) - \boldsymbol{R}_{xx}(t_2) \right)$$

This is much more stable but works well only when the two are similar (not useful for single-trial classification where the difference is expected to be large)

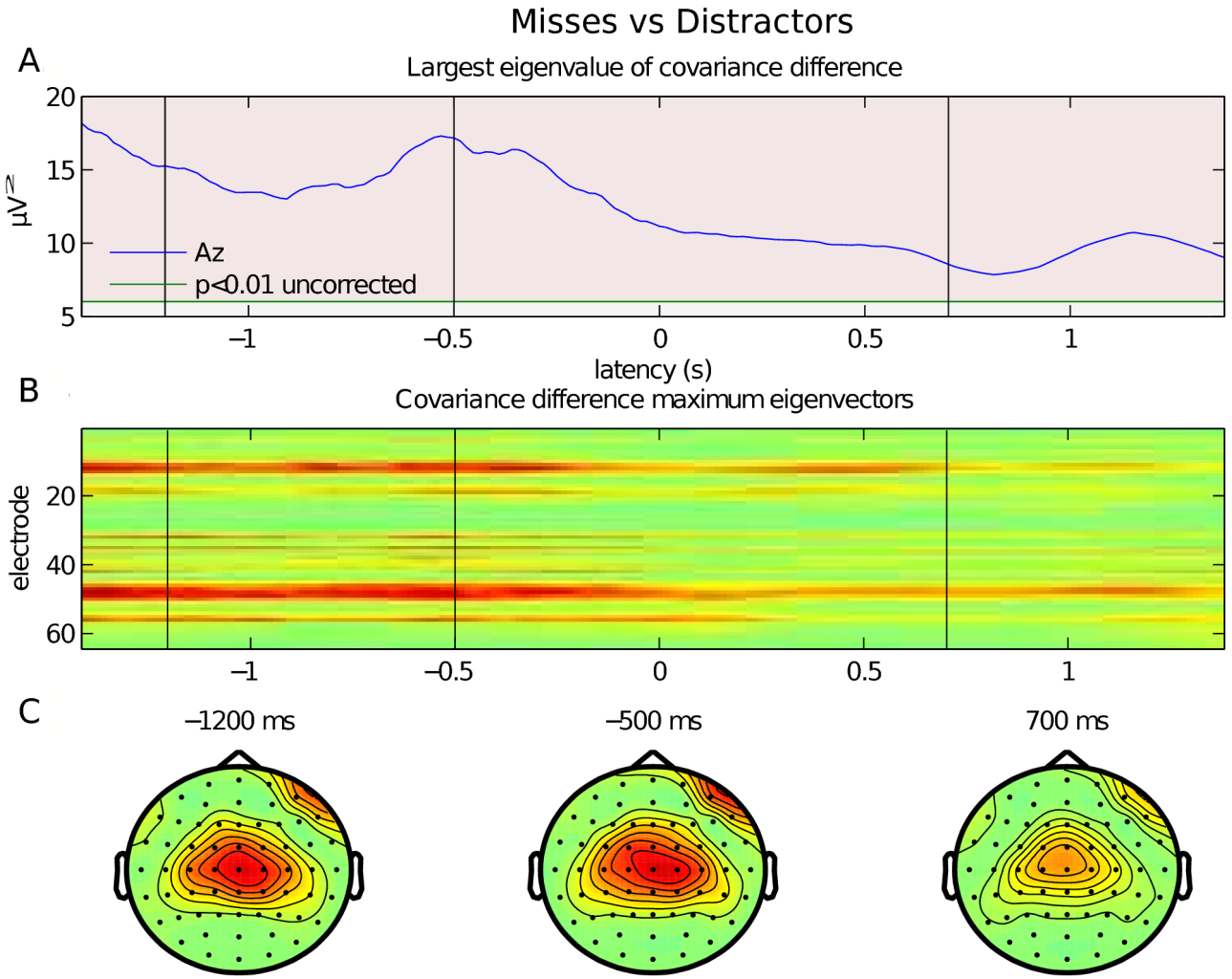
Dias et al. *Journal of Vision*, 2013.

Maximum power difference



Dias et al. *Journal of Vision*, 2013.

Maximum power difference



Dias et al. *Journal of Vision*, 2013.

Blind Source Separation problem

$$\boldsymbol{X} = \boldsymbol{A}\,\boldsymbol{S}$$

Question: Given ***X***, can one tell what ***A*** and ***S*** is?

Answer: Yes! Provided some ***prior information*** on ***S***.

Bind Source Separation

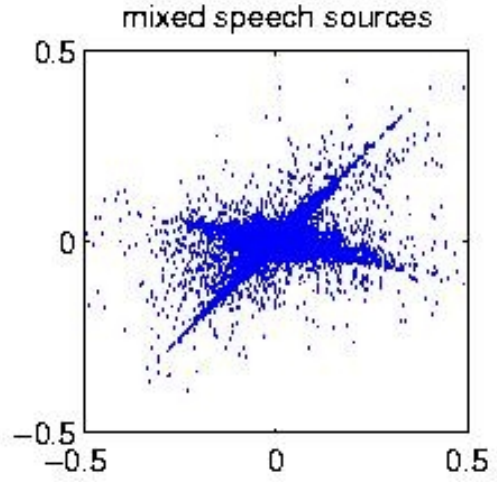
Prior information: Statistical independence. It implies that expected values of product of different sources $i \neq j$ factorize:

$$\overline{y_i^n(t)\, y_j^m(t+l)} = \overline{y_i^n(t)}\ \ \overline{y_j^m(t+l)}$$

For M sources and N sensors each t,l,n,m gives $M(M\text{-}1)/2$ equations. Thus, they provide $M(M\text{-}1)/2$ conditions on the NM unknowns in $\boldsymbol{A}$. We have sufficient conditions if we use multiple:

use	sources assumed	resulting algorithm
n, m	non-Gaussian	ICA
t	non-stationary	CSP
l	non-white	TDSEP, DSS

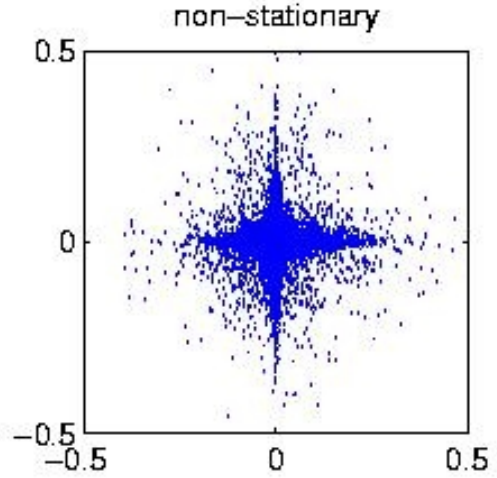
BSS in two lines of matlab

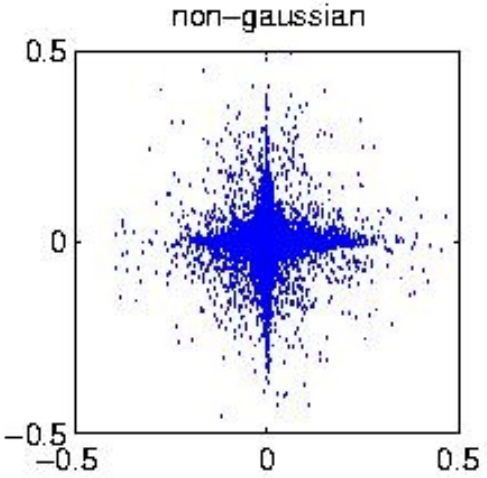


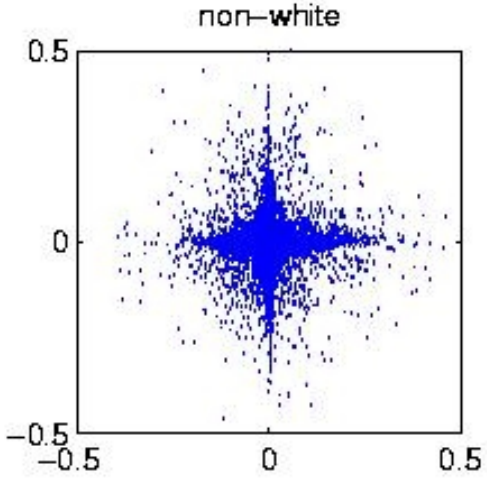

```
% linear mix of sourses S
X=A*S;

% Separation based on Generalized Eigenvalues
[W,D]=eig(X*X',Q);
S=W' *X;
```

Results with Q assuming:

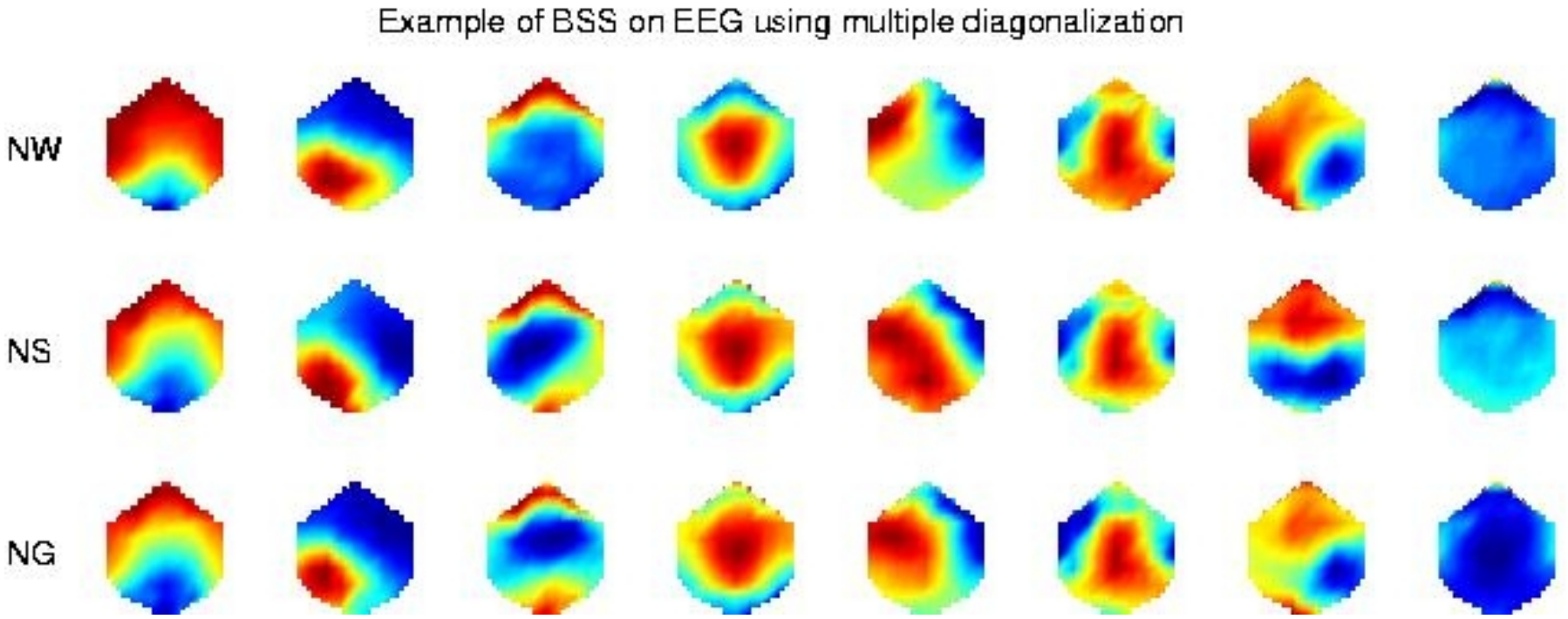






Parra, Sajda, Journal of Machine Learning Research, 2003.

Blind Source Separation – Multiple diagonalization



Blind Source Separation – discussion

Caveats

- Sources in the brain are not independent. Hence it is better to talk about ***components*** **and not** ***sources***.
- Which of the many source should one look at?
- Which BSS criterion/algorithm to use?
- → Problem of multiple comparison is aggravated.

Solution:

Use BSS algorithms that are not only consistent with independent sources but also optimize a meaningful objective criterion.

Examples

Maximum power ratio → CSP

Maximum evoked response → version of DSS

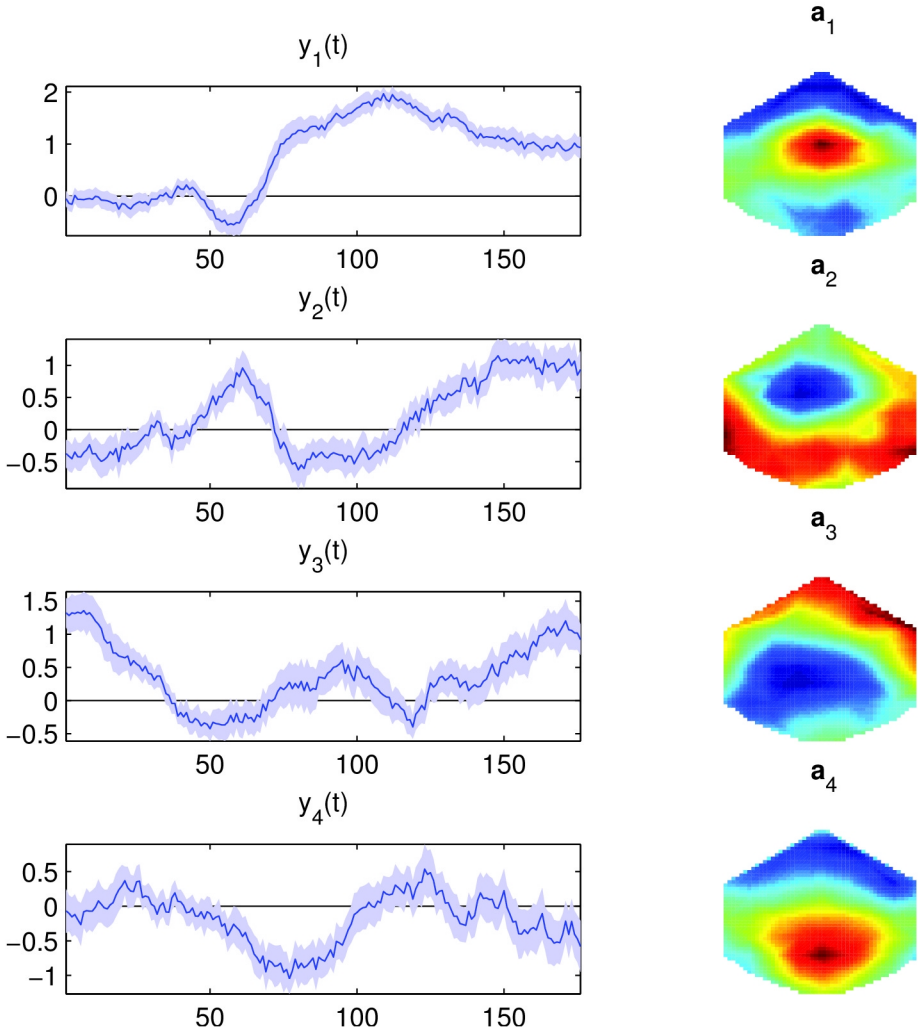
Maximum repeat correlation → CCA

Maximum evoked response

Evoked response is the mean activity $\bar{x}$. Maximize its variance relative to the total variance in the data.

$$F = \frac{var\left(mean(y)\right)}{mean\left(var(y)\right)} = \frac{\boldsymbol{w}^T \boldsymbol{R}_{\bar{x}\bar{x}} \boldsymbol{w}}{\boldsymbol{w}^T \overline{\boldsymbol{R}_{xx}} \boldsymbol{w}}$$

$$\boldsymbol{W} = eig\left(\boldsymbol{R}_{\bar{x}\bar{x}}, \overline{\boldsymbol{R}_{xx}}\right)$$



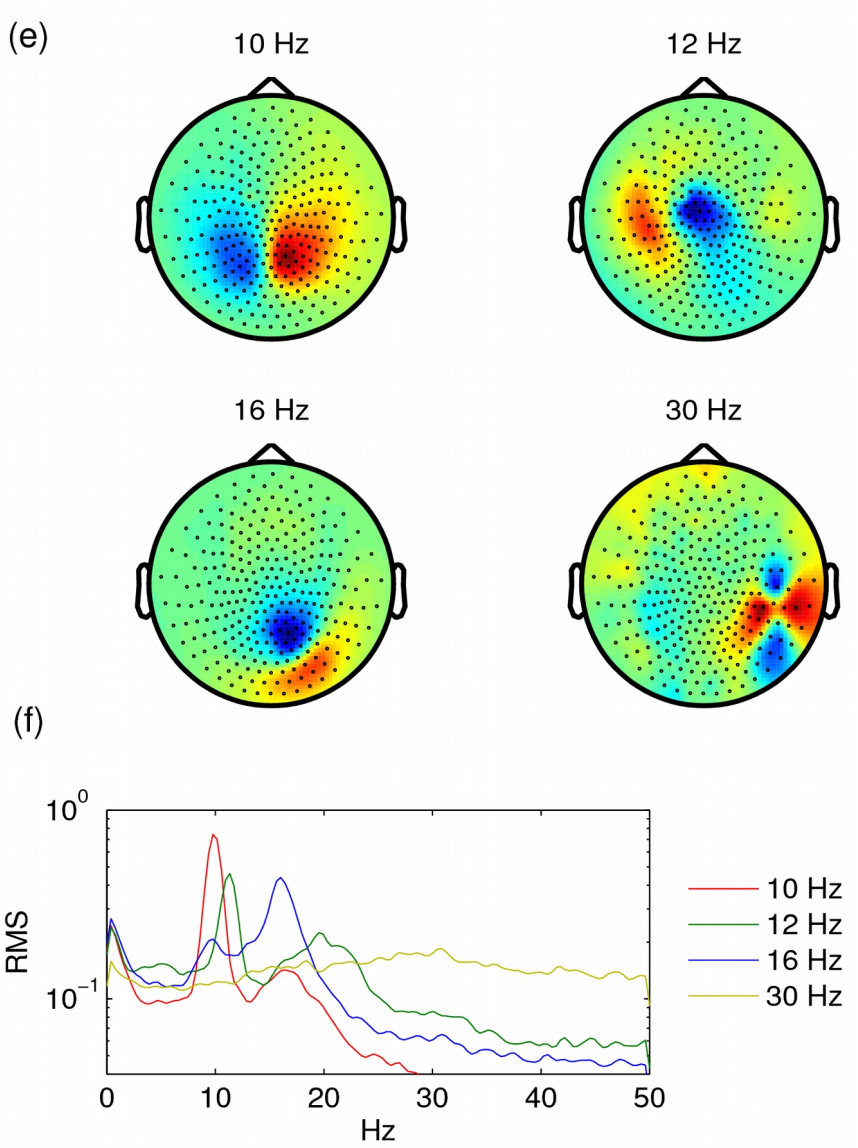
Maximum Signal to Noise Ratio

The previous concepts can be generalized to maximizing SNR where a linear filter enhances signal of interest.

$$\tilde{y}(t) = L[y(t)]$$

$$\max_{w} \frac{\sigma^2_{\tilde{y}}}{\sigma^2_{y}} = \max_{w} \frac{\boldsymbol{w}^T \boldsymbol{R}_{\tilde{x}\tilde{x}} \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{R}_{xx} \boldsymbol{w}}$$

$$\boldsymbol{W} = eig\left(\boldsymbol{R}_{\tilde{x}\tilde{x}}, \boldsymbol{R}_{xx}\right)$$

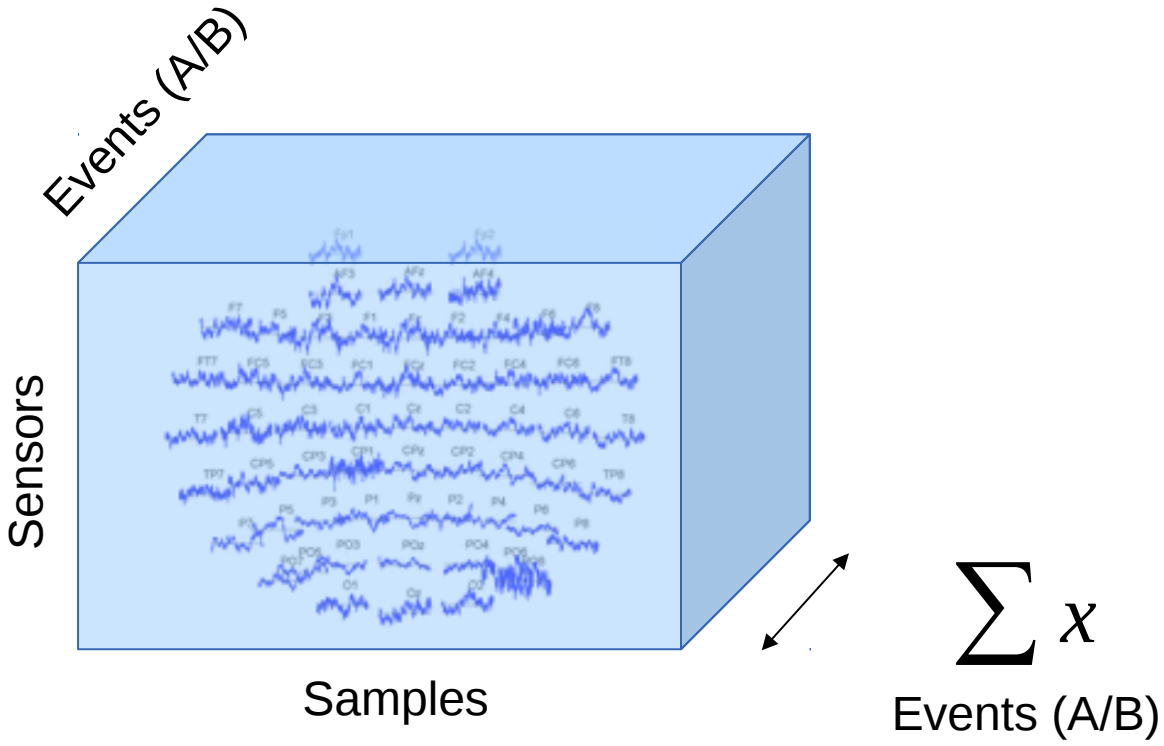


de Cheveigne, Parra, NeuroImage, 2014

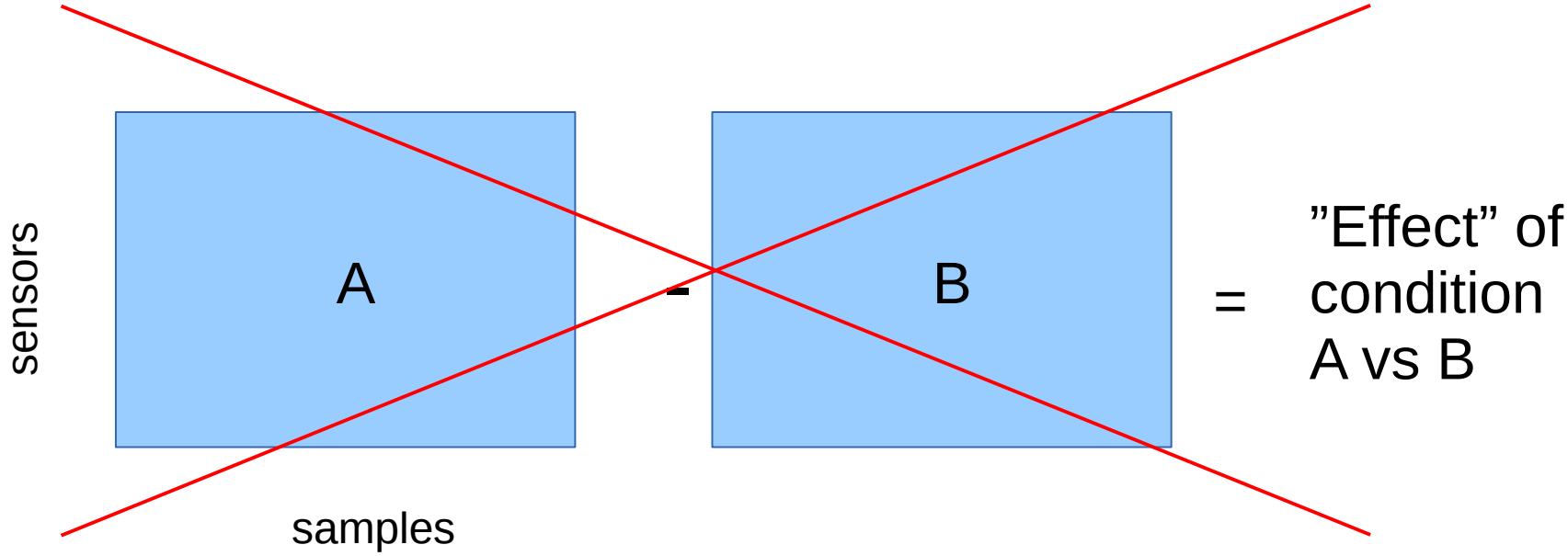
Popular commercials

Most popular commercial Super Bowl 2013

Conventional event-locked analysis



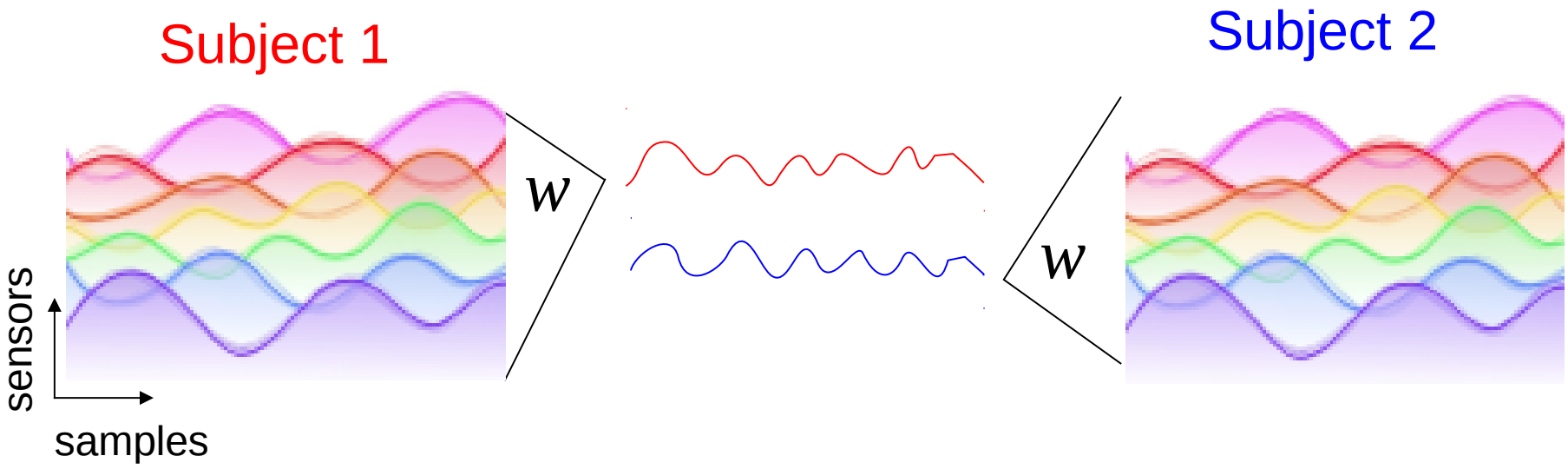
Event-locked evoked response



However! Natural stimuli don't have precise event markers!

Maximal Inter-Subject Correlation (ISC)

We measure ISC in “components” of the EEG:

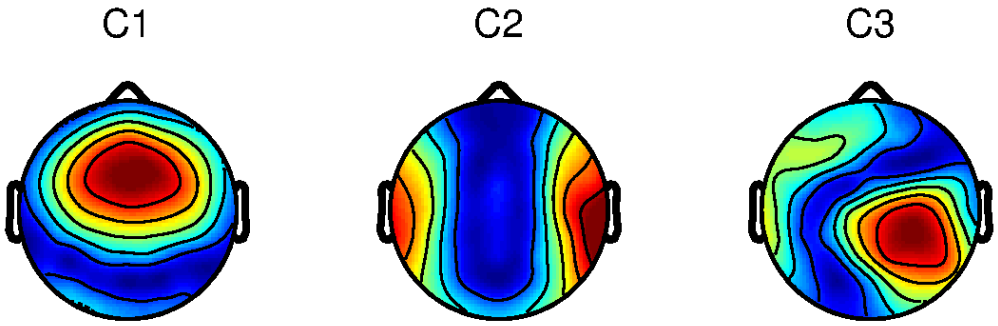


Maximal ISC

"Correlated Component Analysis":

$$\boldsymbol{R}_w^{-1} \boldsymbol{R}_b \boldsymbol{w} = \boldsymbol{w} \lambda$$

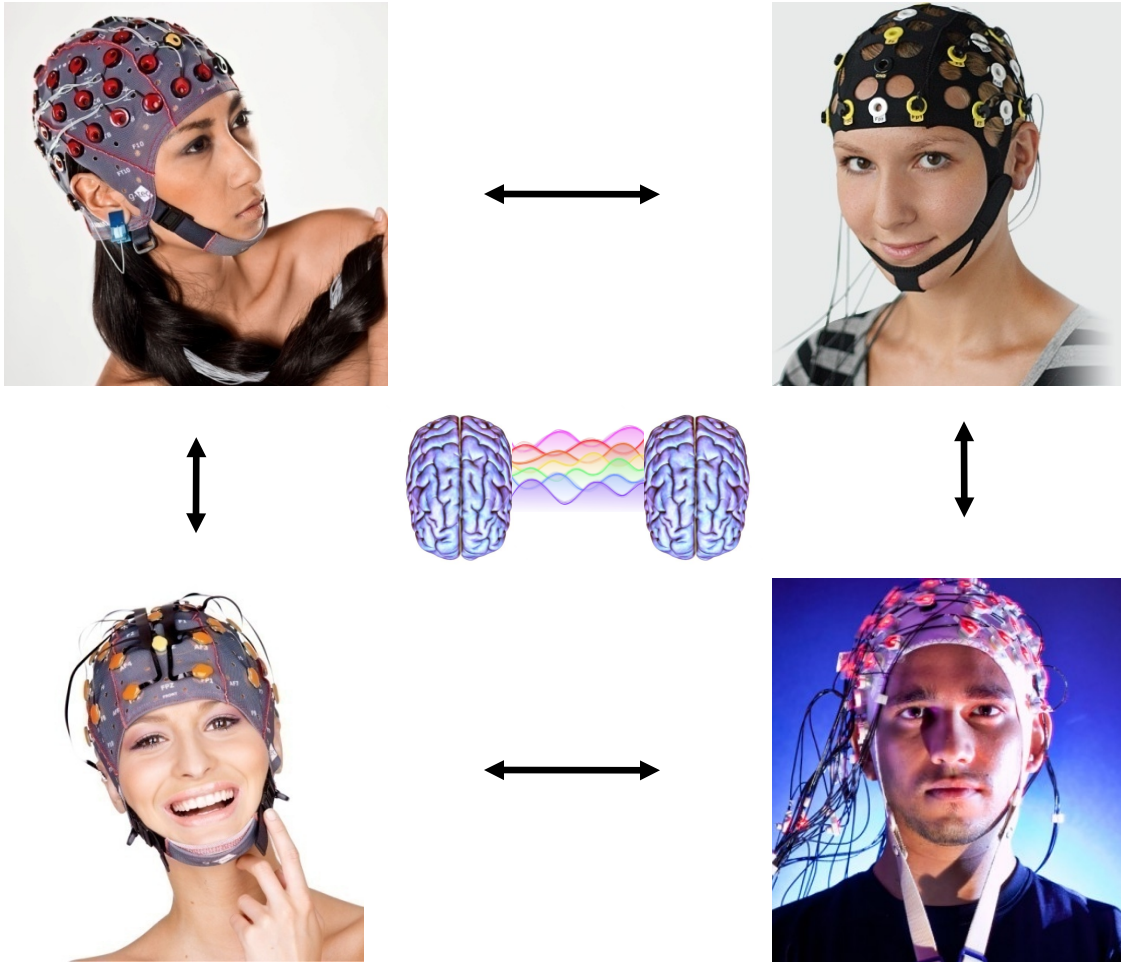
Similar to PCA but instead of maximum variance we capture maximum correlation



Code: www.parralab.org/isc

Dmochowski, *Frontiers in Hum. Neuroscience* 2012

Brains on Video



Dmochowski, Frontiers in Human Neuroscience, 2012.

Maximal ISC

Cross-covariance between subjects k and l

$$\boldsymbol{R}_{kl} = \frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{x}_k(t) - \bar{\boldsymbol{x}}_k\right)\left(\boldsymbol{x}_l(t) - \bar{\boldsymbol{x}}_l\right)^T$$

Between subject covariance

$$\boldsymbol{R}_b = \frac{2}{N(N-1)}\sum_{k=1}^{N}\sum_{l=k+1}^{N}\boldsymbol{R}_{kl}$$

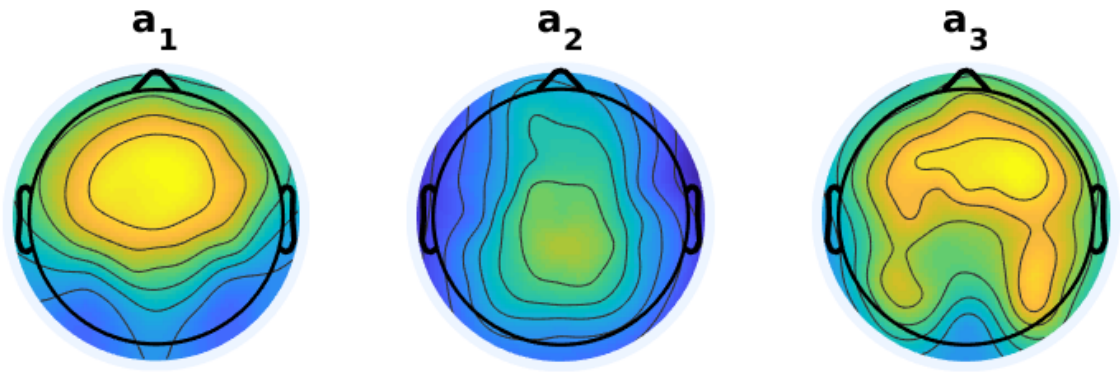
Within subject covariance

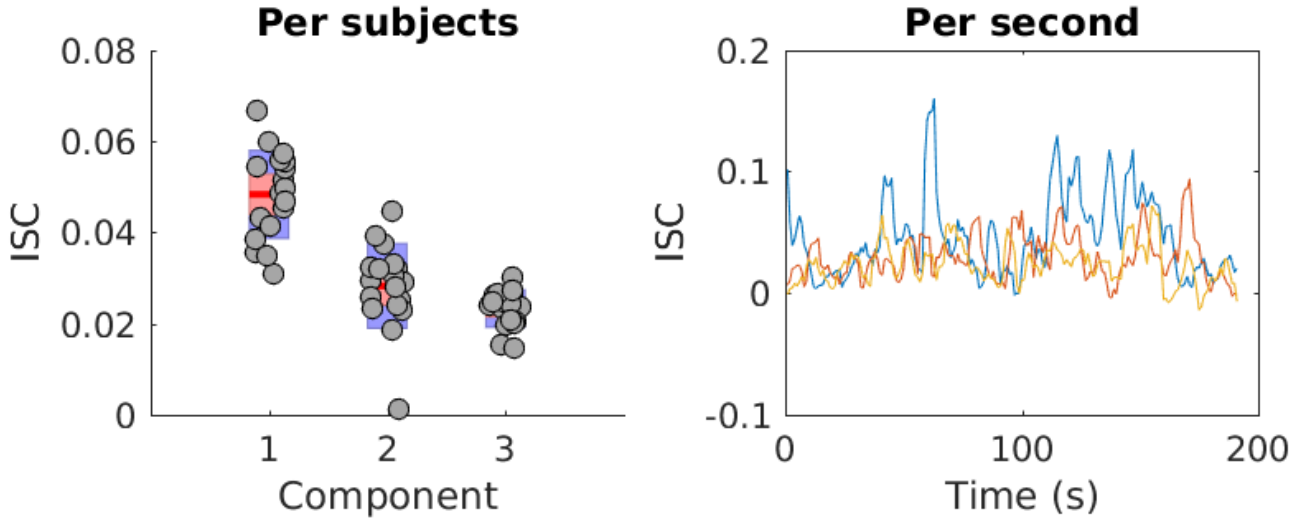
$$\boldsymbol{R}_w = \frac{1}{N}\sum_{k=1}^{N}\boldsymbol{R}_{kk}$$

```
>> W = eig(Rb,Rw)
```

Cohen, Parra, *eNeuro*, 2016

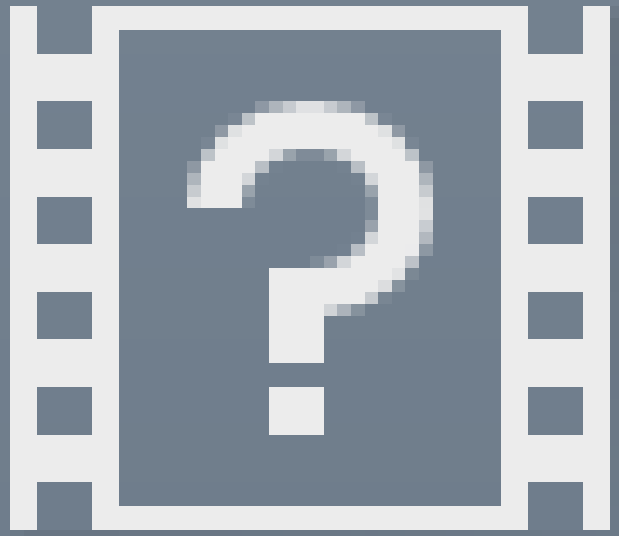
Maximal ISC



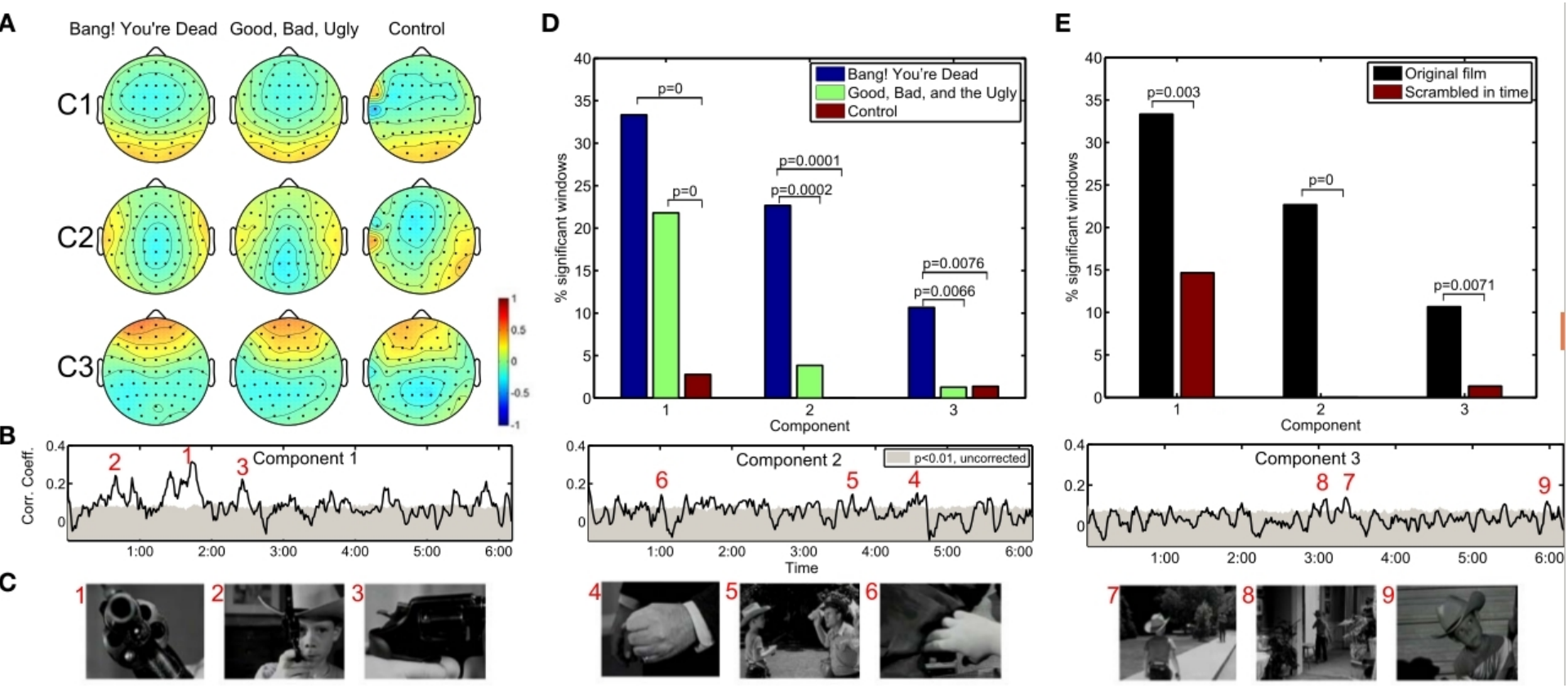


Code: www.parralab.org/isc

Cohen, Parra, *eNeuro*, 2016

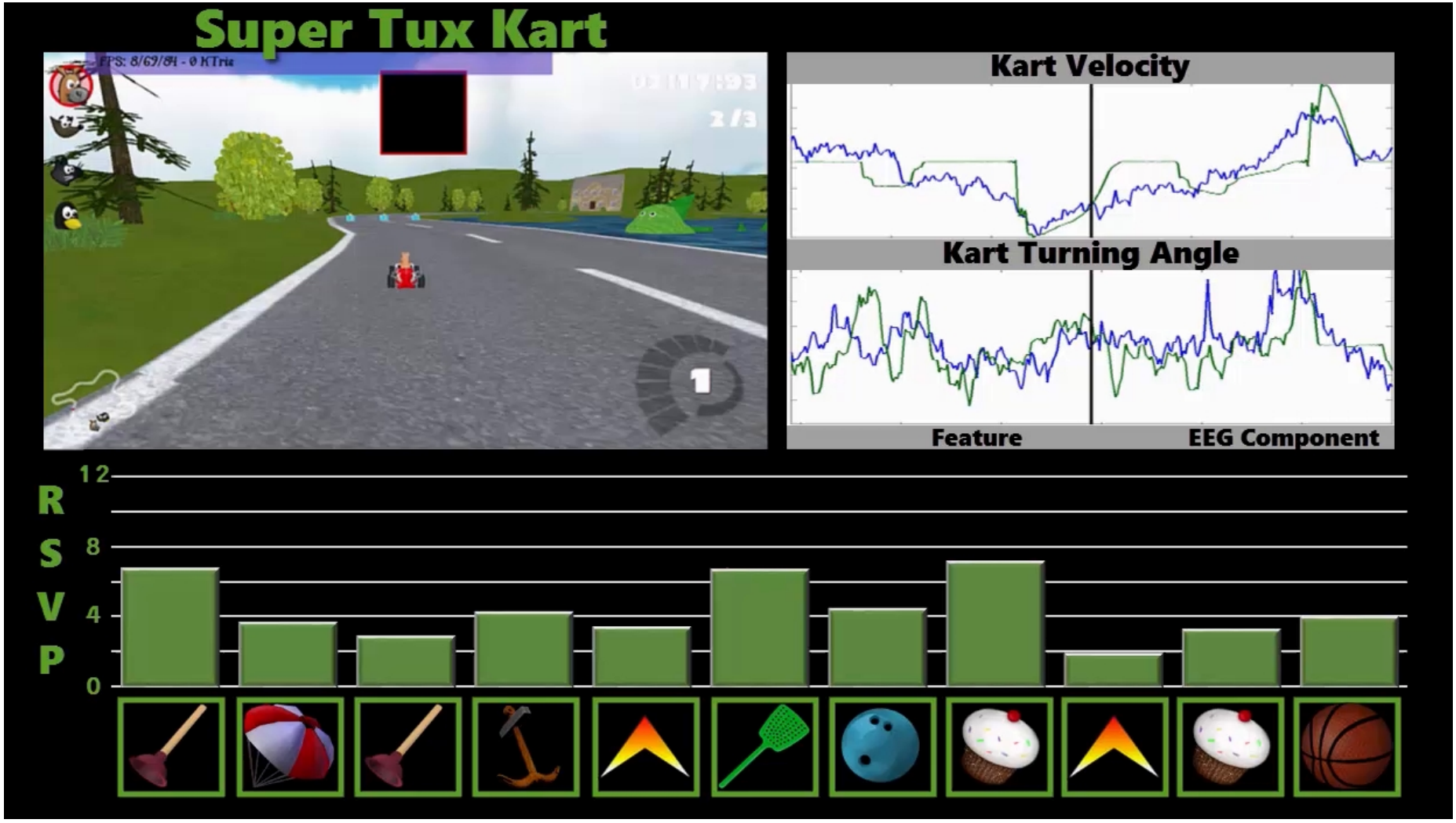


Maximum correlation



Dmochowski, *Frontiers of Human Neuroscience*, 2012.

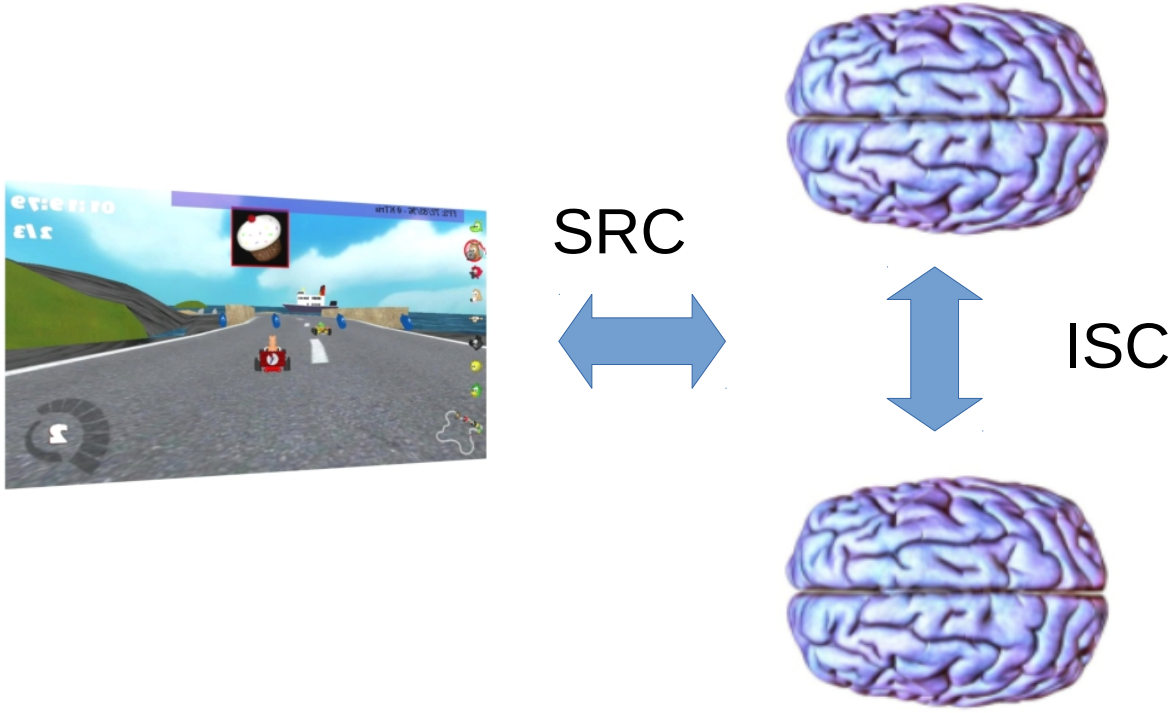
Unique experience: Active video game



Collaboration with Neuromatters (programmed game, collected data)

Unique individual experience

Jacek Dmochowski



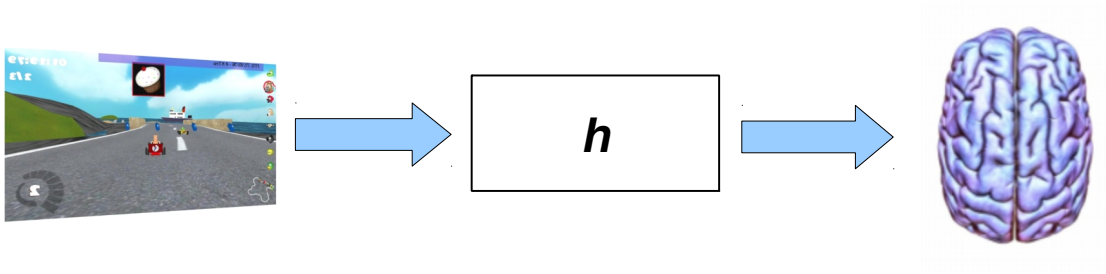
→ Stimulus-Response correlation (**SRC**)

Dmochowski, *NeuroImage*, 2017

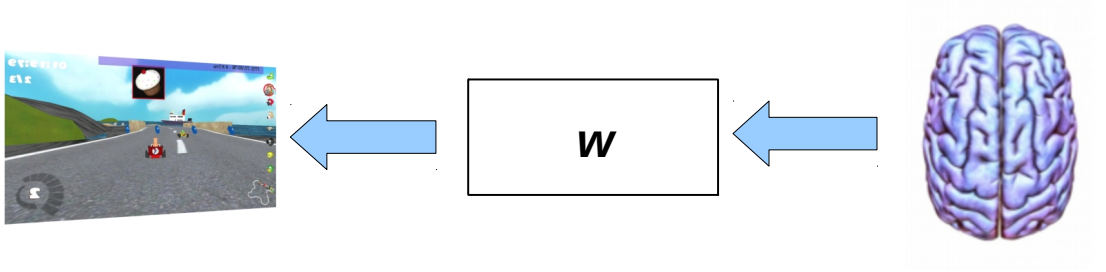
Maximal Stimulus-Response Correlation

Jacek Dmochowski

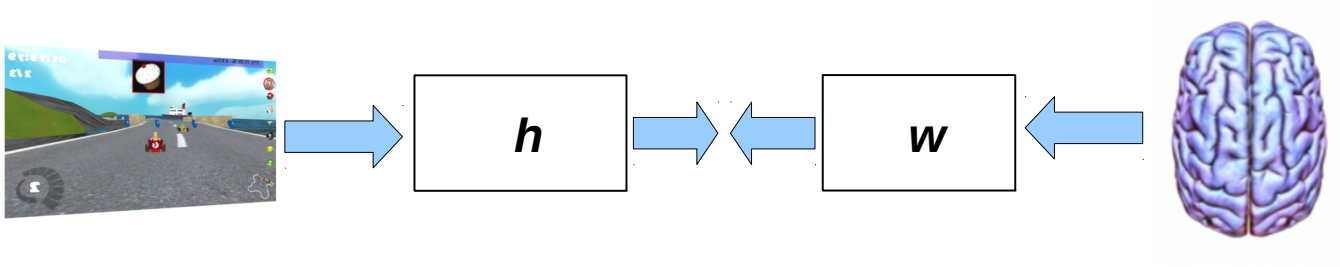
Encoding (LMS)



Decoding (LMS)



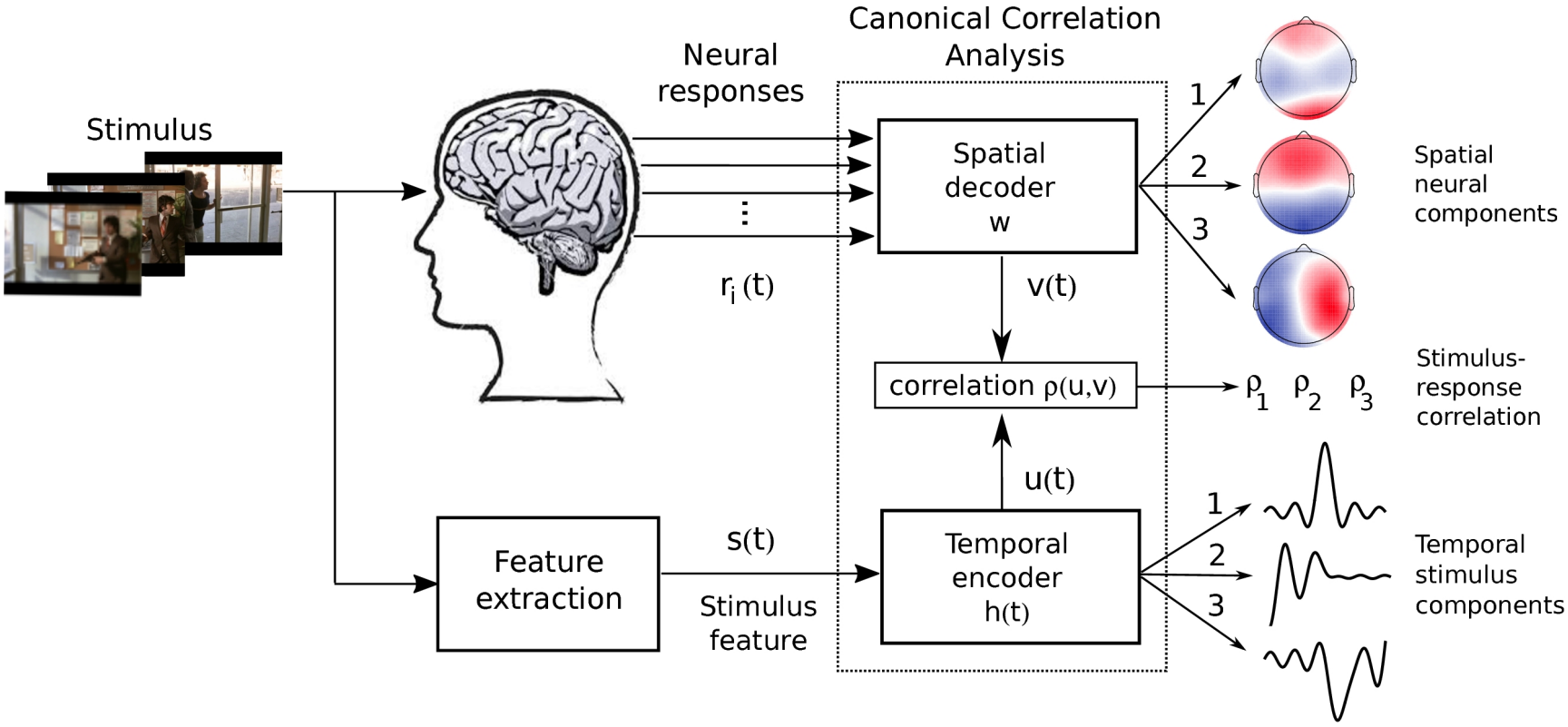
Hybrid (CCA)



Dmochowski, *NeuroImage*, 2017

Stimulus-Response Correlation (SRC)

Jacek Dmochowski



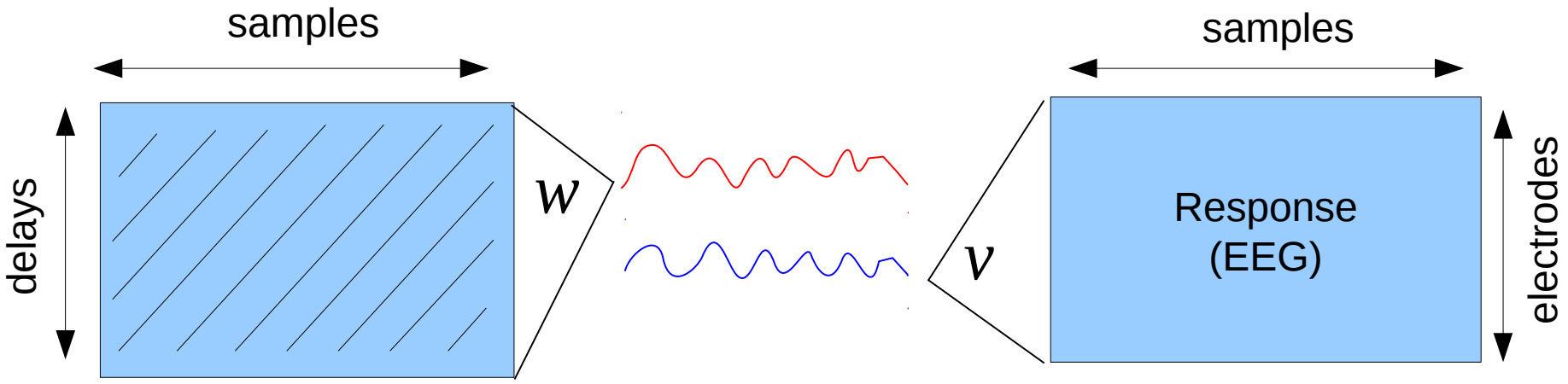
Code: www.parralab.org/resources.html

Dmochowski, *NeuroImage*, 2017

Stimulus-Response Correlation (SRC)

Jacek Dmochowski

Apply Canonical Correlation analysis to the two matrices:



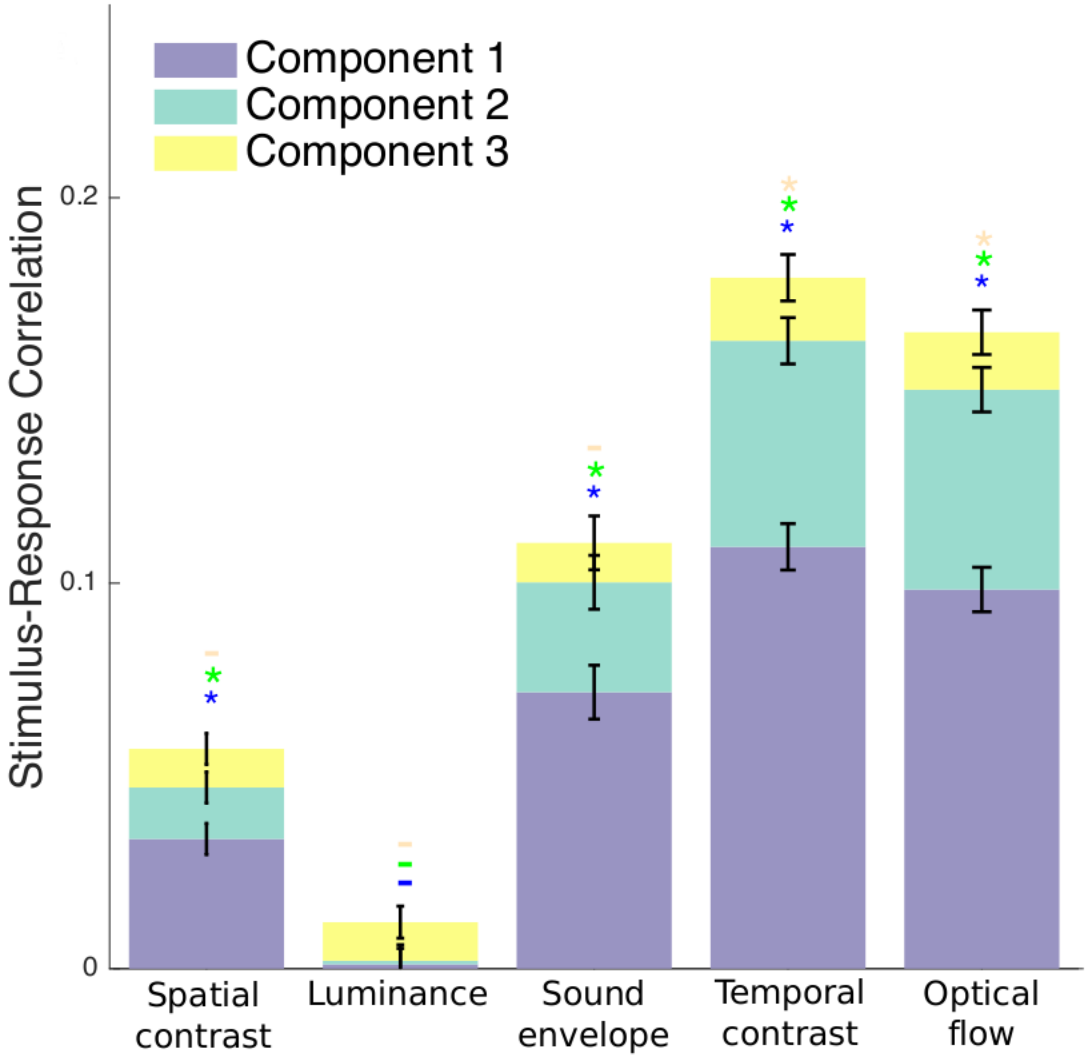

```
>> [W,V] = cca(toeplitz(s), EEG)
```

Code: www.parralab.org/resources.html

Dmochowski, *NeuroImage*, 2017

SRC for audio/visual features in video

Jacek Dmochowski

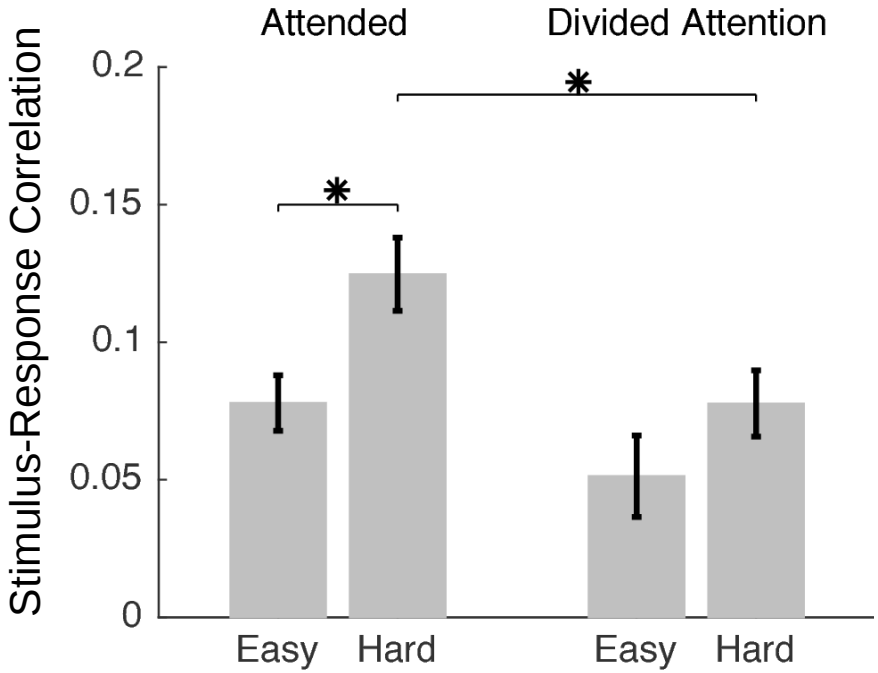


Dmochowski, *NeuroImage*, 2017

SRC modulated by task

Jacek Dmochowski

A



Dmochowski, *NeuroImage*, 2017

Robustness and regularization

- Many of the methods presented require an estimate of the covariance $\boldsymbol{R}_{xx}$ that is robust to ***noise*** and can be estimated from a ***small sample***.
- Inversion in particular is sensitive to uncertainty resulting from small sample size, i.e. matrix is singular or ill-conditioned, so that inverting magnifies even small estimation errors.

Standard techniques to address this problems:

- Subspace reduction
- Automated outlier rejection (robust PCA)
- Shrinkage (J. Schafer, K. Strimmer, Stat Appl Genet Mol Biol, vol. 4(32), 2005.)

Robust PCA

$$M = L + S$$

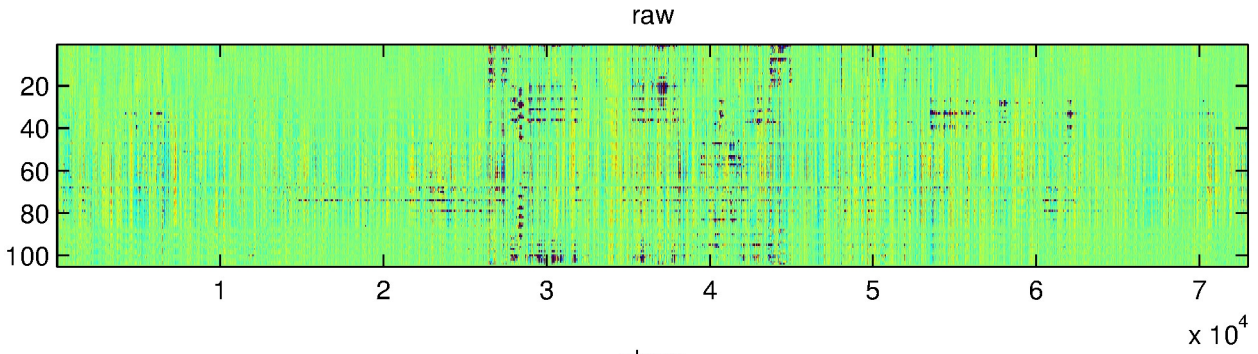
$$argmin \left\| L \right\|_2 + \lambda \left\| S \right\|_1$$

- Candès, JACM, 2011: convex, proofs that it “does the right thing” provided rank L is “small”.
- Fast algorithm: Inexact augmented Lagrange multiplier (Z Lin, M Chen, Y Ma – arXiv 2010)
- Does not scale well with dimensions, but scales fine with length of signal
- Works great for EEG.

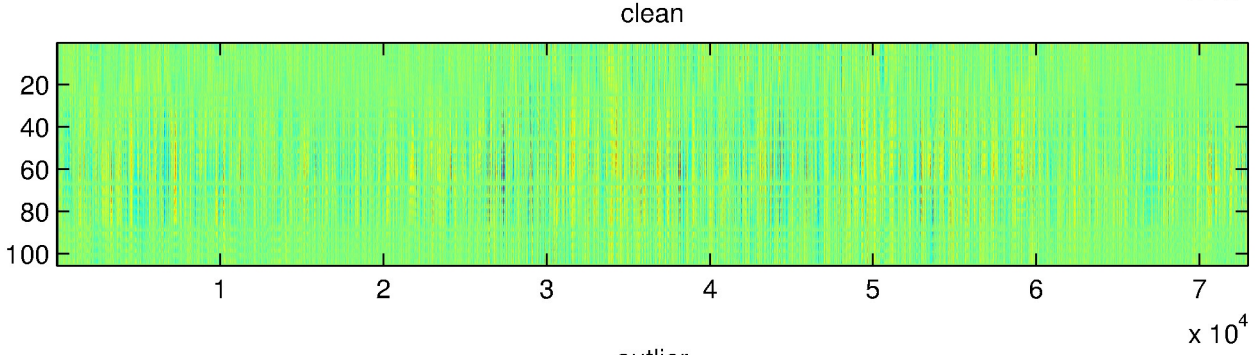
Robust PCA on EEG

$$M = L + S$$

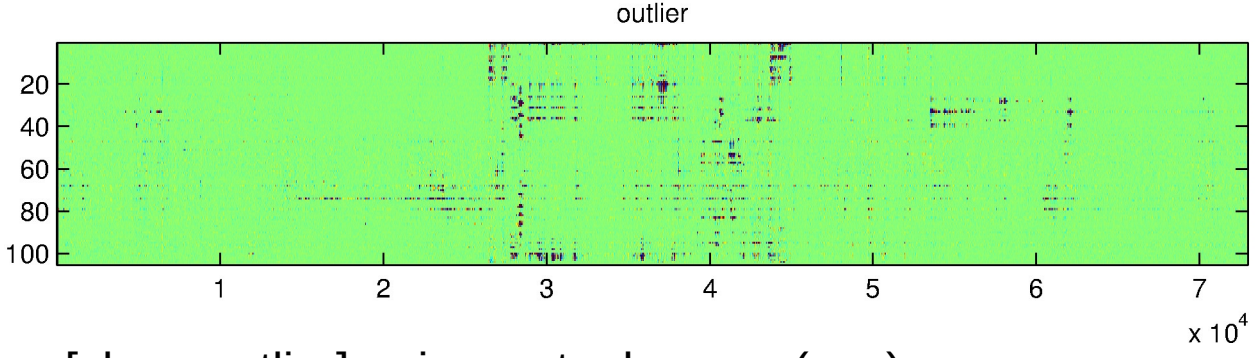
M



L



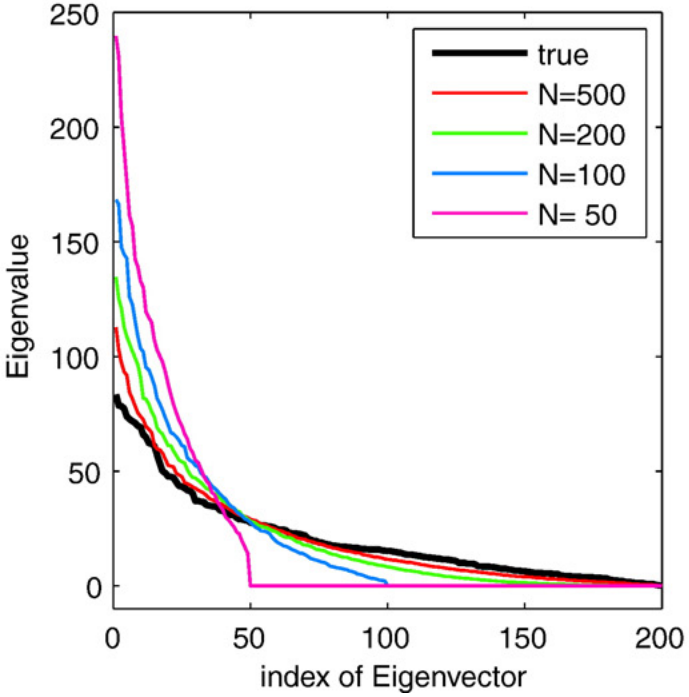
S

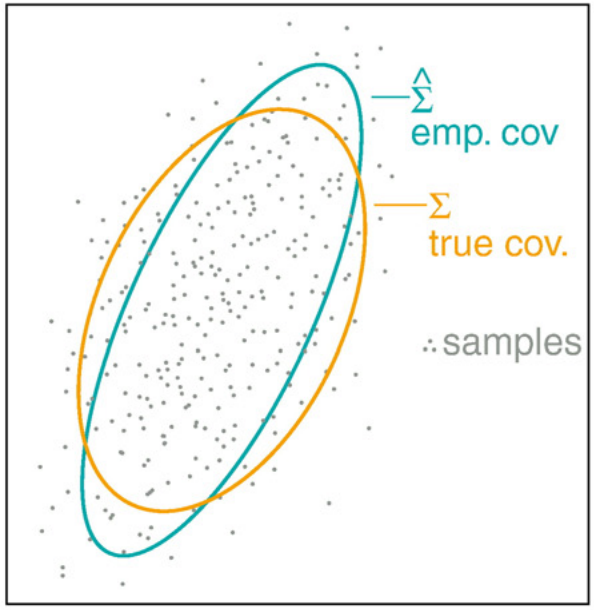


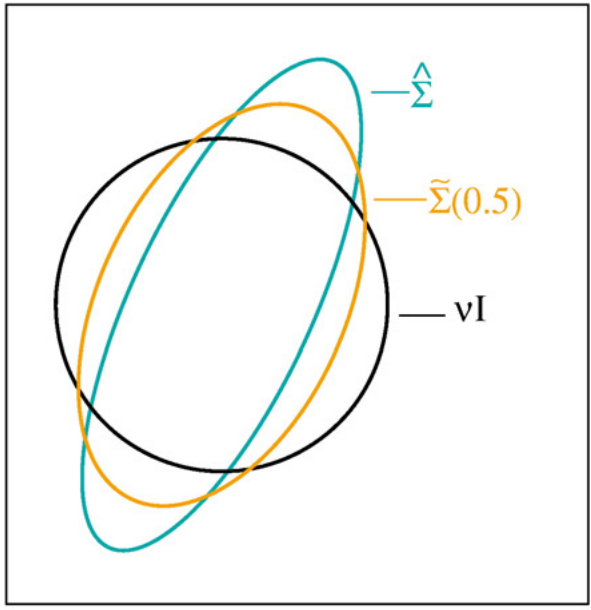

```
[clean,outlier]  = inexact_alm_rpca(raw);
```

Shrinkage

$$\hat{R}_{xx} = R_{xx} + \lambda I$$





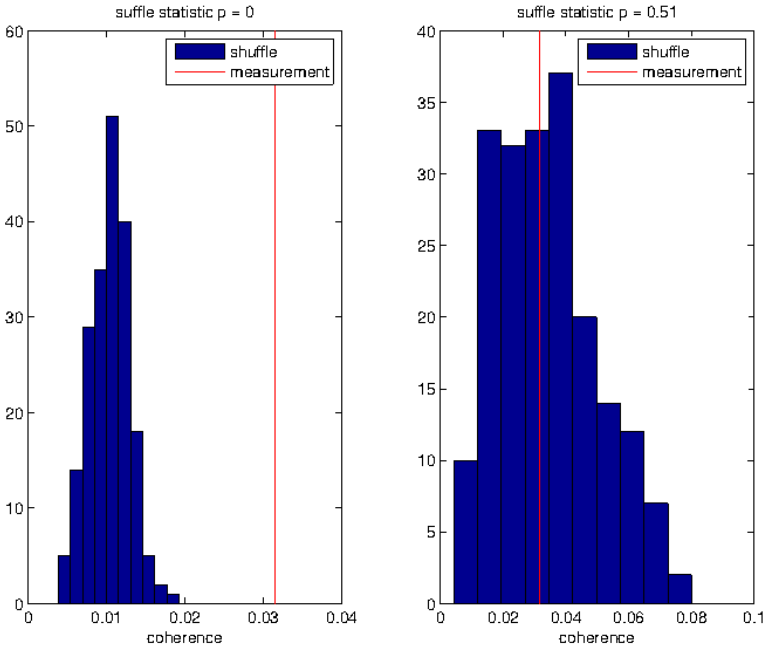


Blankertz, Lemm, Treder, Haufe, Müller, Neuroimage 2011

Randomization statistic

- In many of these examples we have specifically picked a spatial filter which maximizes a desired statistic, e.g. t-statistic. We can no longer use it with standard tables to compute p-values as it is now biased.
- Recall that p-value represents the probability of something happening by chance. Thus, we generate 'random' data and find the optimal filter ***w*** for this random data to see what values we obtain by chance.
- We can then ask what fraction of these random values is larger than the actual value we observed – this is the p-value.

Most important in generate this random data is to **preserve the correlation structure!**



Broken correlation

Preserved correlation

Not significant!

Conclusion …

- Combining electrodes into a single component gives a **large boost in statistical significance**.

- However, there is no magic bullet on how to do this.

  Because
  Finding genuine current sources in the brain from EEG is an **ill-posed problem**: there are more unknowns than observations.

  Thus
  We are **forced to make assumptions** about sources.

… Conclusion

Assumptions:

- **Anatomy**: This leads to inverse modeling, e.g. dipole fit, LORETA, etc.

- **Sources**: This leads to various “blind” sources separation algorithms, e.g. independence, non-stationarity, differing spectral properties, etc.

Alternatively, forget claims about sources and anatomy.

- Instead, extract “components” with favorable properties: largest effect size, most discriminant, maximum power, maximum change in power, most reproducible, etc.
- Properties are to be selected based on need, e.g.
    - demonstrate small effects
    - detect earliest onset
    - single-trial detection, etc.

References

- Parra, Spence, Gerson, Sajda, "Recipes for the Linear Analysis of EEG", Neuroimage, 2005.
- Parra, Sajda, "Blind Source Separation via Generalized Eigenvalue Decomposition", Journal of Machine Learning Research, 2003.
- Lucas Parra, Paul Sajda, "Blind Source Separation via Generalized Eigenvalue Decomposition", Journal of Machine Learning Research, vol. 4, pp. 1261-1269, Dec 2003.
- Dias, Dmochowski, Sajda, Parra, "EEG precursors of detected and missed targets during free-viewing search", Journal of Vision, 2013.
- de Cheveigne, Parra, Joint decorrelation, a versatile tool for multichannel data analysis, NeuroImage, 2014
- Yu Huang, Lucas C. Parra, Stefan Haufe, "The New York Head - A precise standardized volume conductor model for EEG source localization and tES targeting", NeuroImage, 140: 150-162, October 2016
- Dmochowski, Ki, DeGuzman, Sajda, Parra, Multidimensional stimulus-response correlation reveals supramodal neural responses to naturalistic stimuli, accepted *NeuroImage*, May 2017
- Cohen, Parra, Memorable audiovisual narratives synchronize sensory and supramodal neural responses, *eNeuro*, 3(6), November 2016. (best explanation of ISC code)
- Dmochowski, Sajda, Dias, Parra, "Correlated components of ongoing EEG point to emotionally laden attention - a possible marker of engagement?" Frontiers of Human Neuroscience, 2012.

Preprints, and latest publications at http://parralab.org

Code

- ISC http://parralab.org/isc
- SRC http://parralab.org/resources.html
- This tutorial: http://parralab.org/teaching/eeg
  (with code including penalized logistic regression)