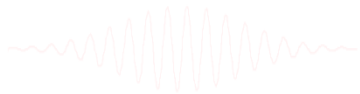


Clustering of ICA components

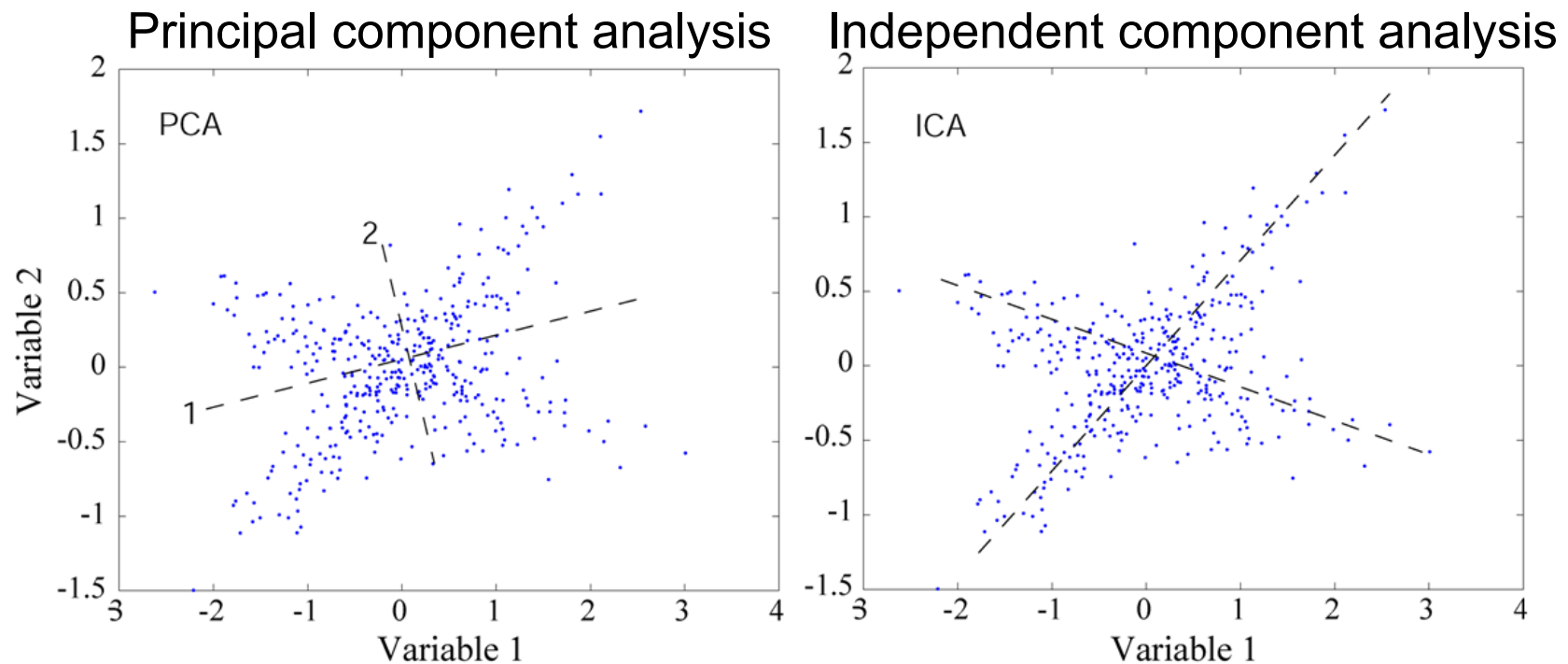
Arnaud Delorme

(with Julie Onton, Romain Grandchamp, Nima Bigdely Shamlo, Scott Makeig)



ICA and PCA

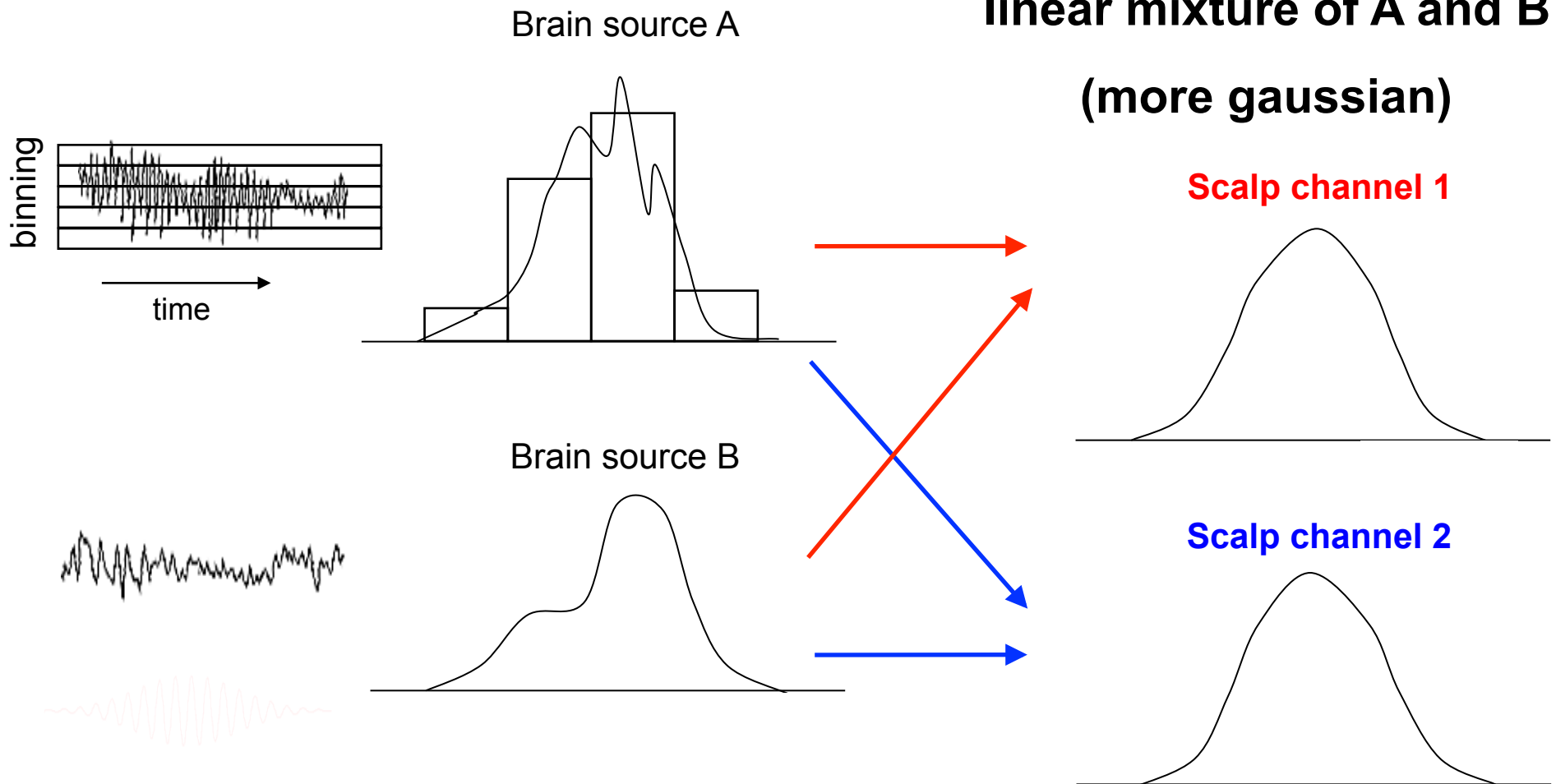
ICA is a method to recover a version, of the original sources by multiplying the data by a unmixing matrix,



While PCA simply decorrelates the outputs (using an orthogonal mixing matrix), ICA attempts to make the outputs **statistically independent**, while placing no constraints on the mixing matrix.

Central limit theorem

**Scalp channels =
linear mixture of A and B
(more gaussian)**

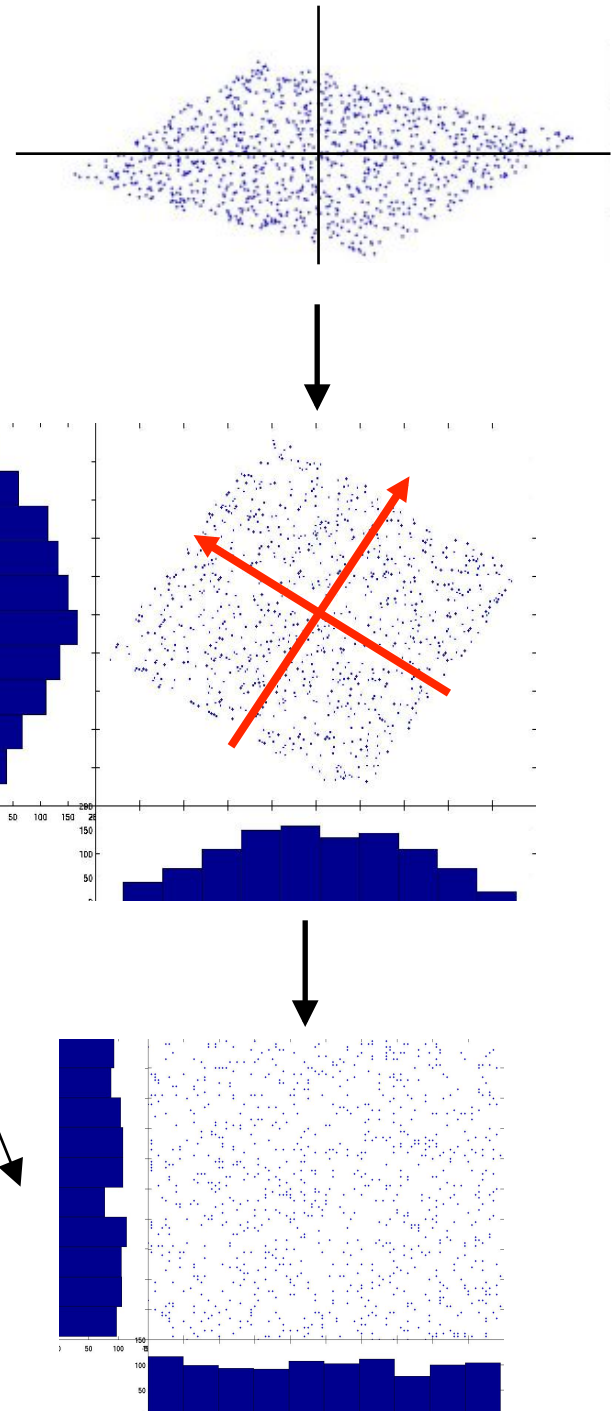


ICA Training Process

Central limit theorem

- Remove the mean
 $\mathbf{x} = \mathbf{x} - \langle \mathbf{x} \rangle$
- 'Sphere' the data by diagonalizing its covariance matrix,
 $\mathbf{x} = \langle \mathbf{x} \mathbf{x}^T \rangle^{-1/2} (\mathbf{x} - \langle \mathbf{x} \rangle)$.
- Update \mathbf{W} according to

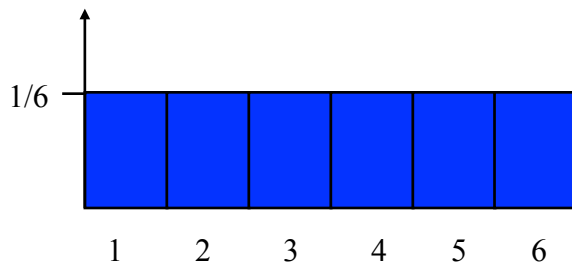
$$\Delta \mathbf{W} \propto \frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} \mathbf{W}^T \mathbf{W}$$



Entropy

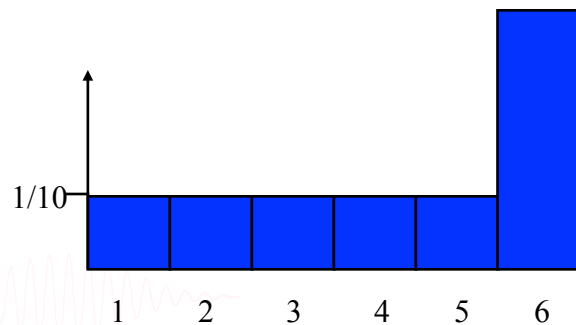
$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log_b p(x).$$

Dice: 1/6



$$H = 6 \left(-\frac{1}{6} \log_2 \left(\frac{1}{6} \right) \right) = 2.58$$

Fake dice (make a 6 half of the time): entropy 2.16 (base 2)



$$H = 5 \left(-\frac{1}{10} \log_2 \left(\frac{1}{10} \right) \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) = 2.16$$

Entropy

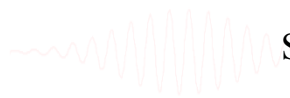
$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log_b p(x).$$

Joint entropy

$$H(X, Y) = - \sum_{(x, y) \in \mathcal{X} \times \mathcal{Y}} p(x, y) \log_b p(x, y).$$

Mutual Information

$$H(y_1, y_2) = H(y_1) + H(y_2) - I(y_1, y_2).$$



Shannon in his landmark 1948 paper "A Mathematical Theory of Communication."

From <http://planetmath.org/encyclopedia/ShannonsTheoremEntropy.html>

Contingency table for stress and emotionality

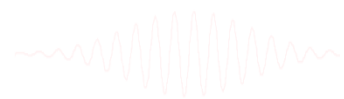
	STRE						
	1	2	3	4	5	6	Total
EMOT= 1	19	4					23
2	11	63	64	3	1		142
3	2	16	18	20	2	2	60
4	1	4	1	9	6	2	23
5			1	2	4	3	10
6				1	1	1	3
Total	33	87	84	35	13	8	

From <http://tecfa.unige.ch/~lemay/thesis/THX-Doctorat/node149.html>

Contingency frequencies for stress and emotionality

	STRE					
	1	2	3	4	5	6
EMOT= 1	0.07	0.02				
2	0.04	0.24	0.25	0.01		
3	0.01	0.06	0.07	0.08	0.01	0.01
4		0.02		0.03	0.02	0.01
5				0.01	0.02	0.01
6						

Joint entropy 3.46; exercise: compute mutual information



$$H(X, Y) = - \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p(x, y) \log_b p(x, y)$$

ICA learning rule

How to make the outputs statistical independent?

Minimize their redundancy or mutual information.

Consider the joint entropy of two components,

$$H(y_1, y_2) = H(y_1) + H(y_2) - I(y_1, y_2).$$

Maximizing $H(y_1, y_2) \implies$ minimizing $I(y_1, y_2)$.

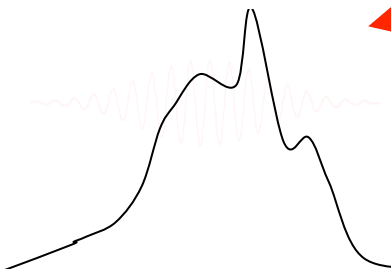
↓
The learning rule:

↓
=0 if the two variables
are independent

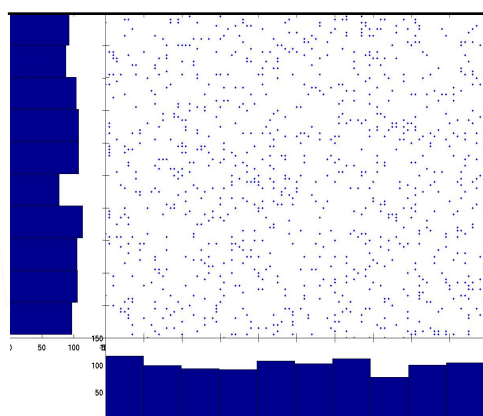
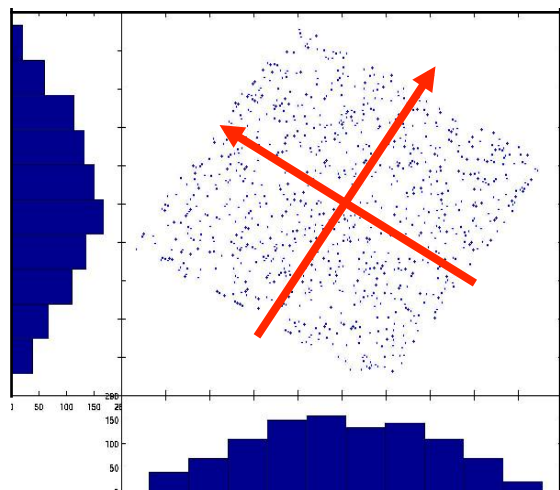
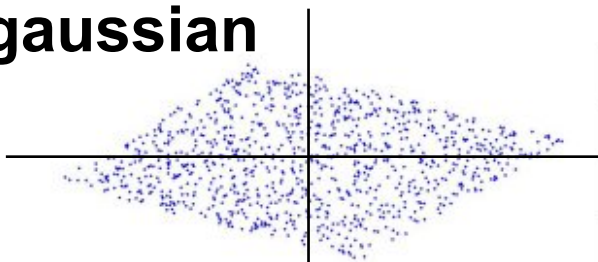
Entropy
extremum

$$\Delta \mathbf{W} \propto \frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} \underbrace{\mathbf{W}^T \mathbf{W}}$$

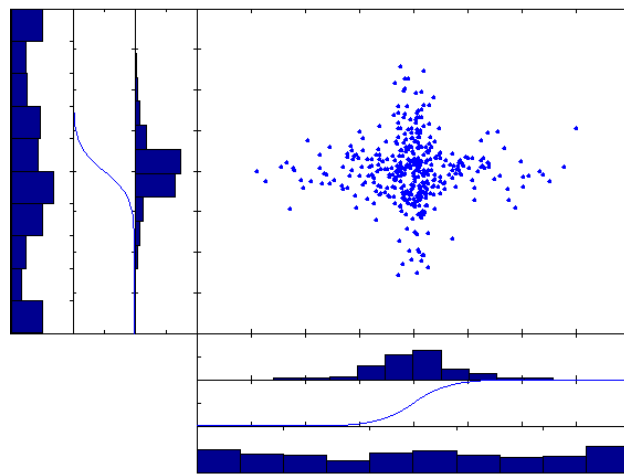
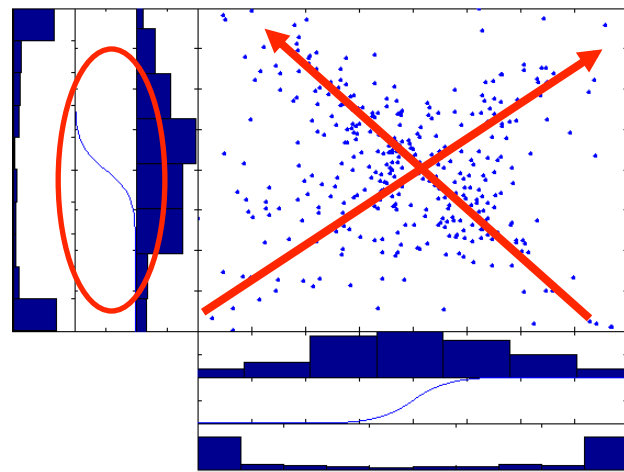
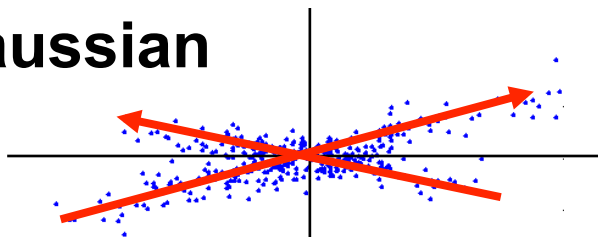
Natural gradient (Amari)



Sub-gaussian



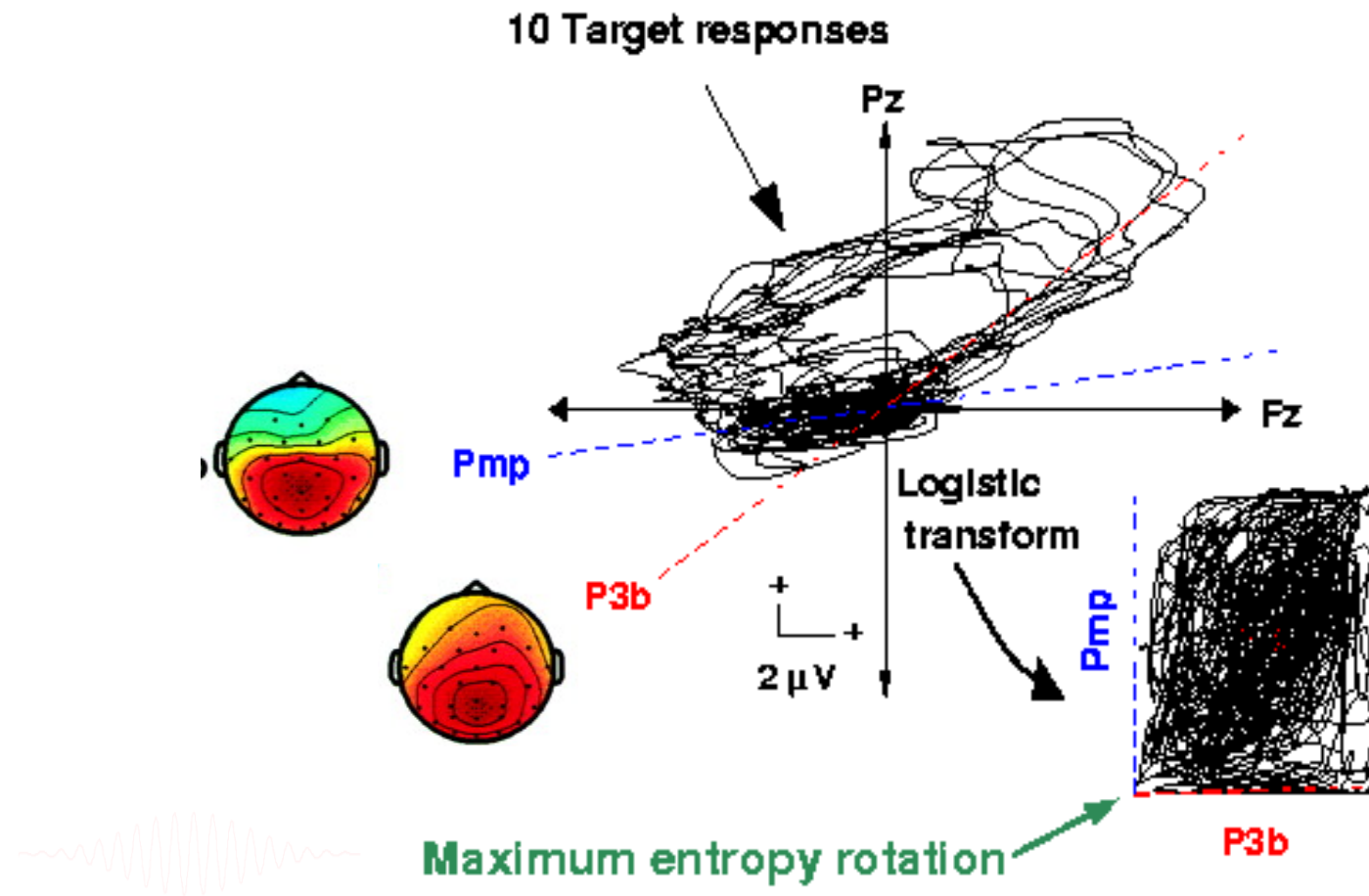
Super-gaussian



Sphering

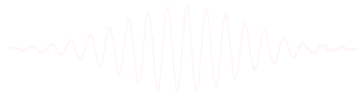
ICA

Independent components of EEG/ERF



Steps of clustering

- Select ICA components for clustering
- Precompute measures of interest
- Cluster measures
- Plot clusters and edit them if necessary



Edit dataset info

pop_study(): Pre-select components

Enter maximum residual (topo map – dipole proj.) var. (in %)
NOTE: This will delete any existing component clusters!

15

☒ Keep only in-brain dipoles.

Cancel Help Ok

Create a new STUDY set -- pop_study()

Edit STUDY set information - remember to save changes

STUDY set name: Sternberg

STUDY set task name: Sternberg

STUDY set notes:

	dataset filename	browse	subject	session	condition	group	Select by r.v.	
1	C:\Users\julie\Documents\Wor	...	S01		memorize		Comp.: 3 5 ...	Clear
2	C:\Users\julie\Documents\Wor	...	S01		ignore		Comp.: 3 5 ...	Clear
3	C:\Users\julie\Documents\Wor	...	S01		probe		Comp.: 3 5 ...	Clear
4	C:\Users\julie\Documents\Wor	...	S02		memorize		Comp.: 5 6 ...	Clear
5	C:\Users\julie\Documents\Wor	...	S02		ignore		Comp.: 5 6 ...	Clear
6	C:\Users\julie\Documents\Wor	...	S02		probe		Comp.: 5 6 ...	Clear
7	C:\Users\julie\Documents\Wor	...	S03		memorize		Comp.: 6 7 ...	Clear
8	C:\Users\julie\Documents\Wor	...	S03		ignore		Comp.: 6 7 ...	Clear
9	C:\Users\julie\Documents\Wor	...	S03		probe		Comp.: 6 7 ...	Clear
10	C:\Users\julie\Documents\Wor	...	S04		memorize		Comp.: 1 2 ...	Clear

Important note: Removed datasets will not be saved before being deleted from EEGLAB memory

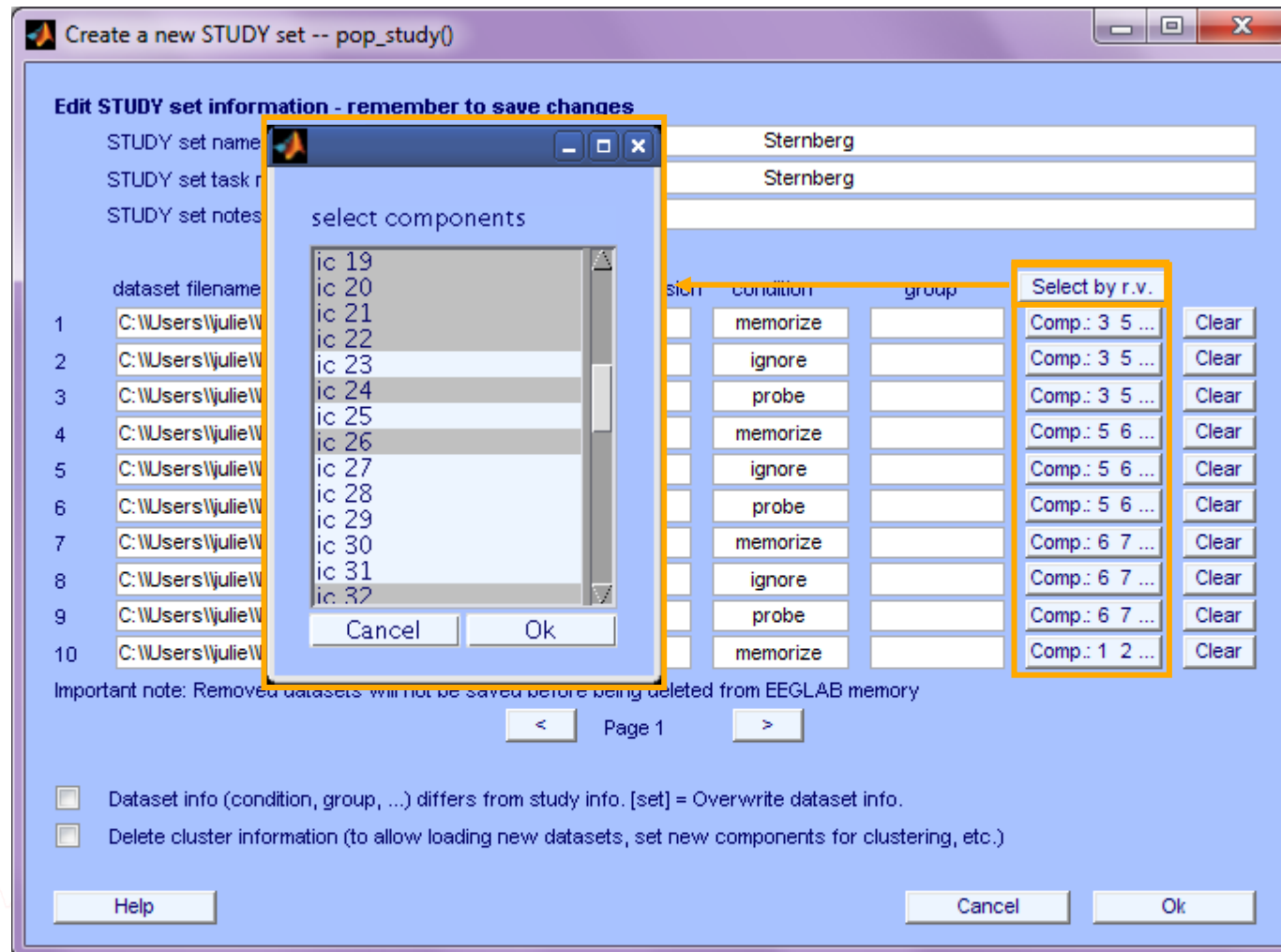
< Page 1 >

☐ Dataset info (condition, group, ...) differs from study info. [set] = Overwrite dataset info.

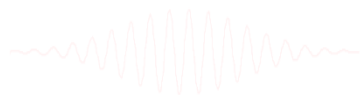
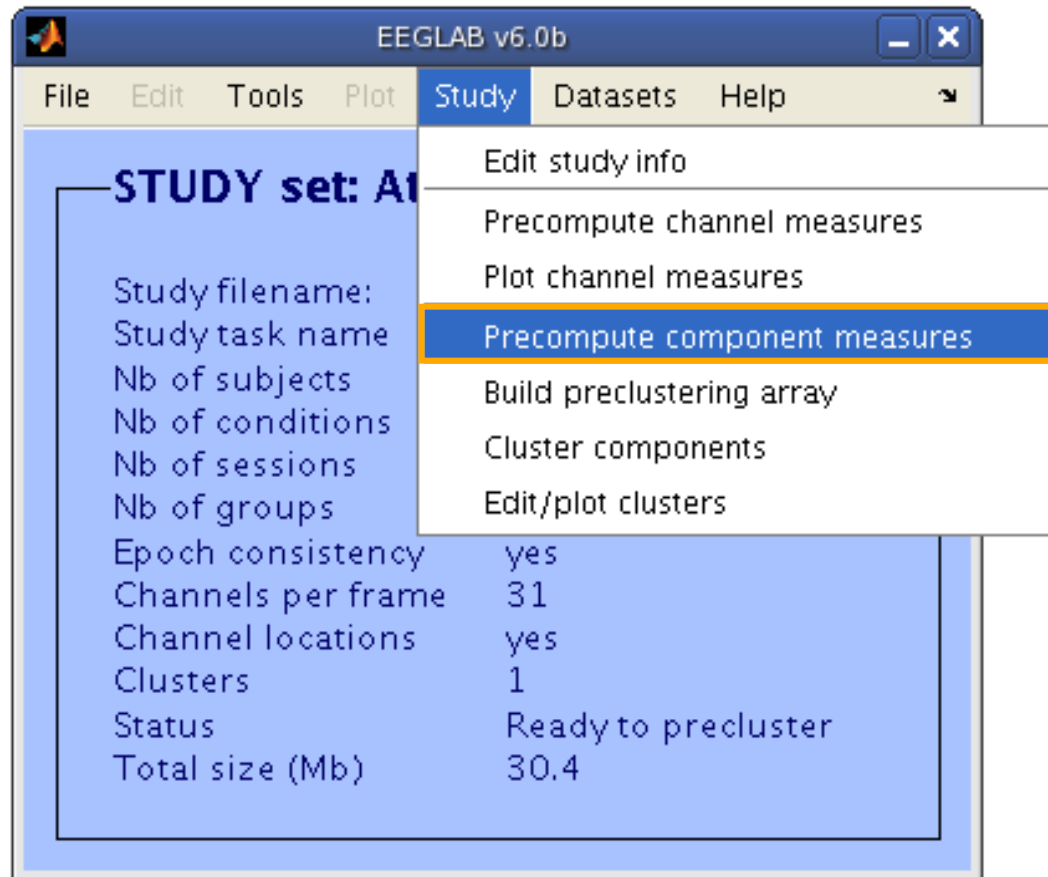
☐ Delete cluster information (to allow loading new datasets, set new components for clustering, etc.)

Help Cancel Ok

ICs to cluster



Precompute data measures



Precompute data measures

TIP: Compute all measures so you can test different combinations for clustering

Select and compute component measures for later clustering -- pop_precomp()

Pre-compute component measures for STUDY 'Sternberg'

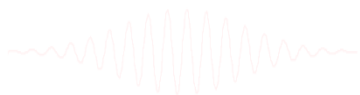
☒ Compute ERP/spectrum/ERSP only for components selected by RV (set) or for all components (unset)

List of measures to precompute

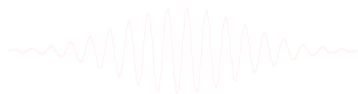
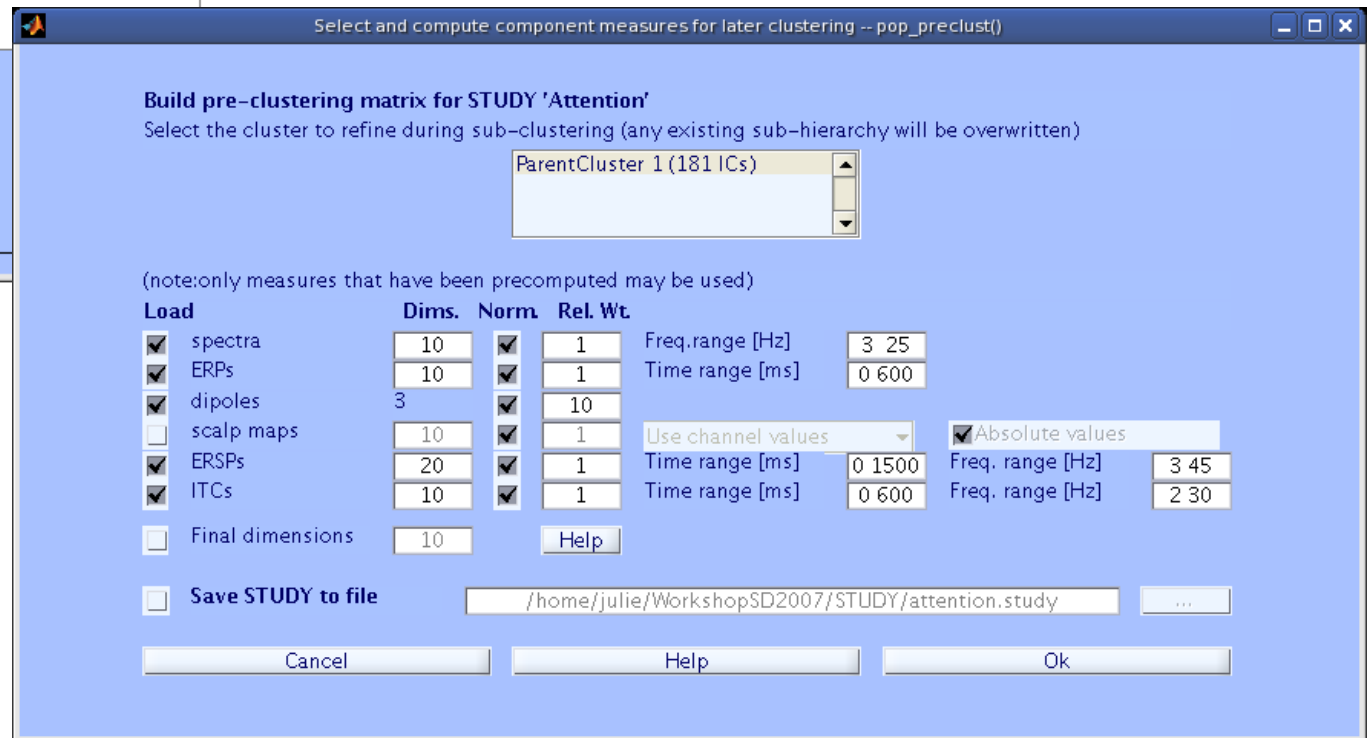
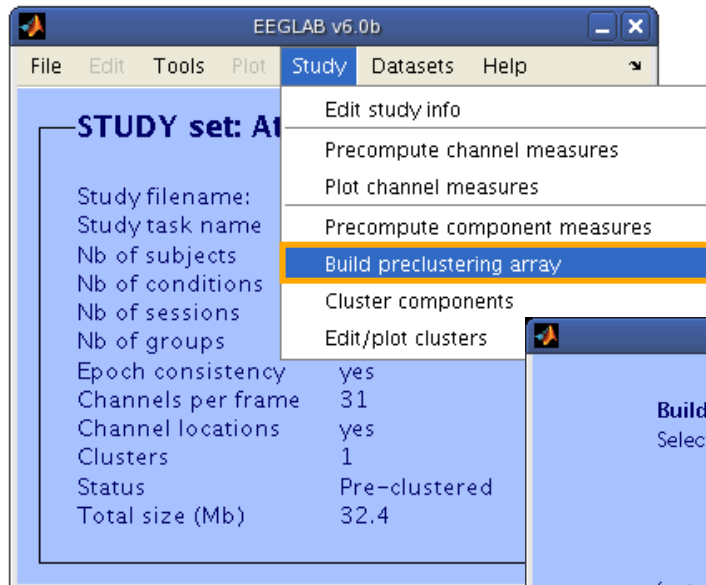
<input checked="" type="checkbox"/> ERPs	Baseline ([min max] in ms)	<input type="text" value="[-200 0]"/>	
<input checked="" type="checkbox"/> Power spectrum	Spectopo parameters	<input type="text" value=""/>	<input type="button" value="Test"/>
<input checked="" type="checkbox"/> ERSPs	Time/freq. parameters	<input type="text" value="'cycles', [3 0.5], 'nfreqs', 100"/>	<input type="button" value="Test"/>
<input checked="" type="checkbox"/> ITCs			
<input checked="" type="checkbox"/> Scalp maps			

☐ Recompute even if present on disk

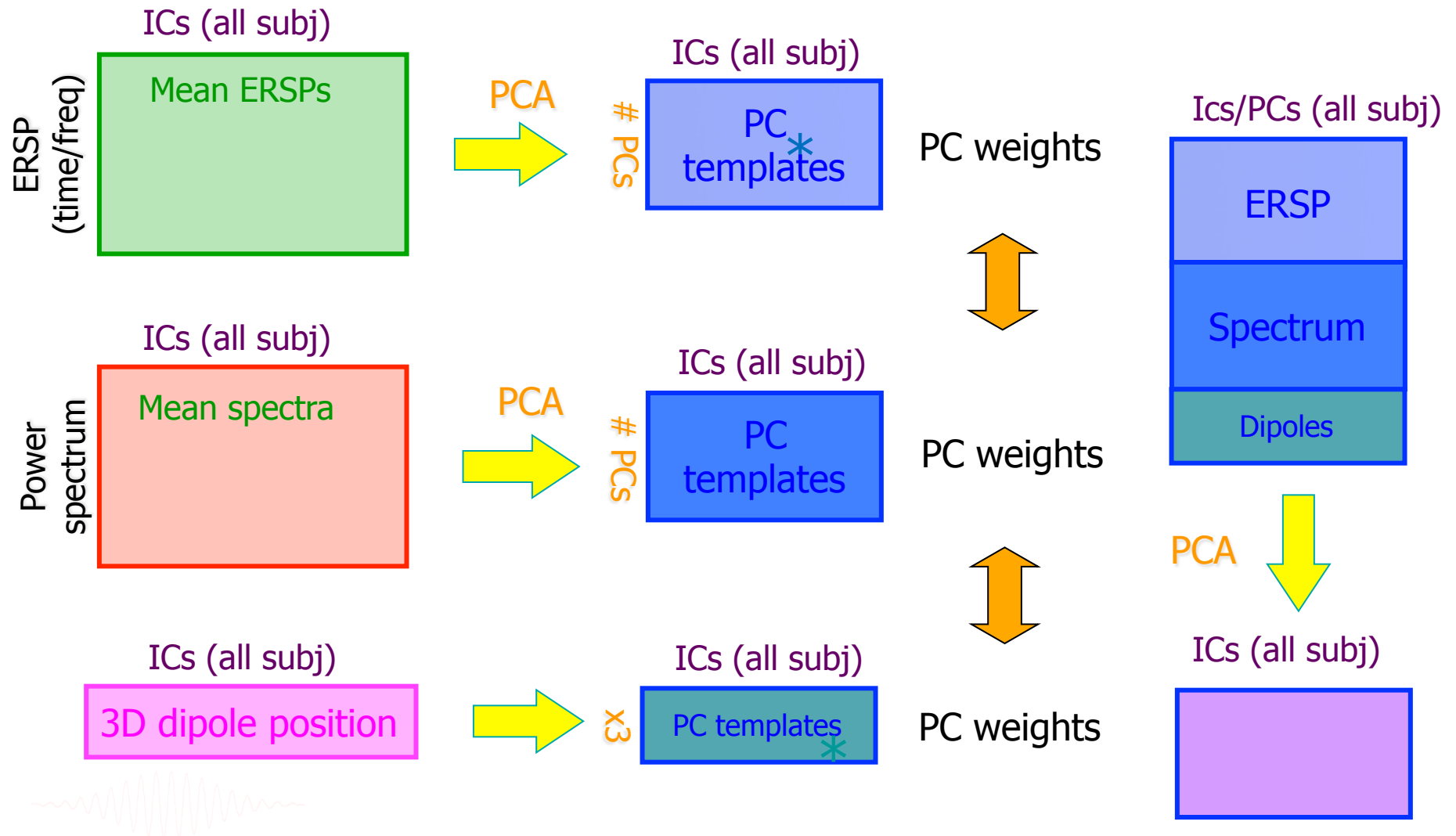
Time-frequency options



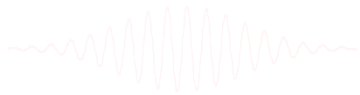
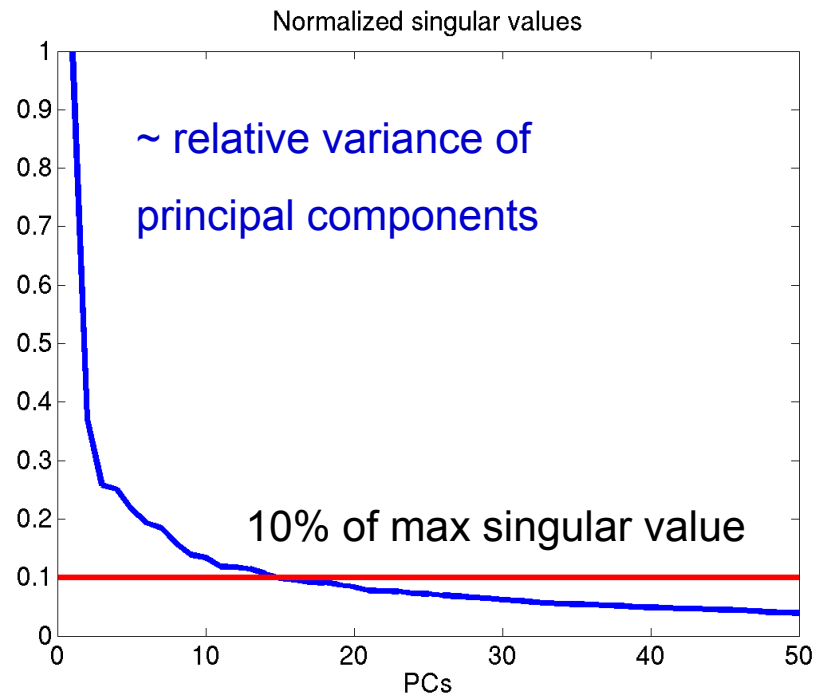
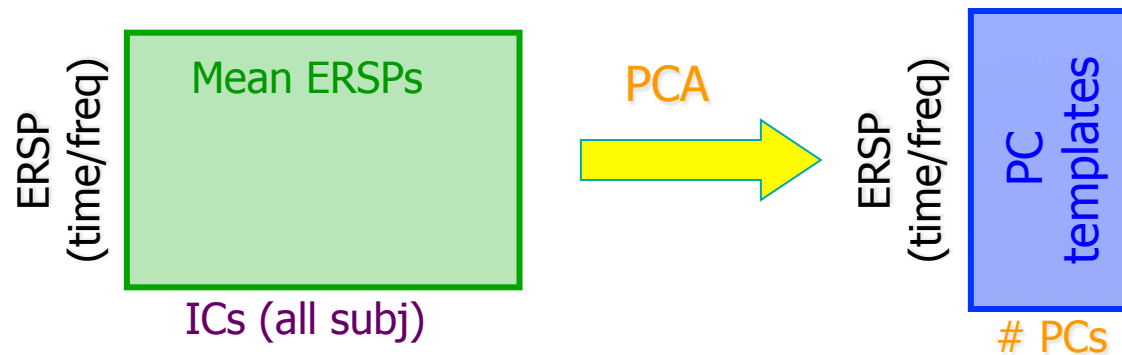
Cluster components



Precluster schematic



Precluster: Use singular values from PCA



Precluster schematic

ICs (all subj)

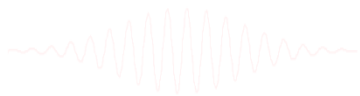
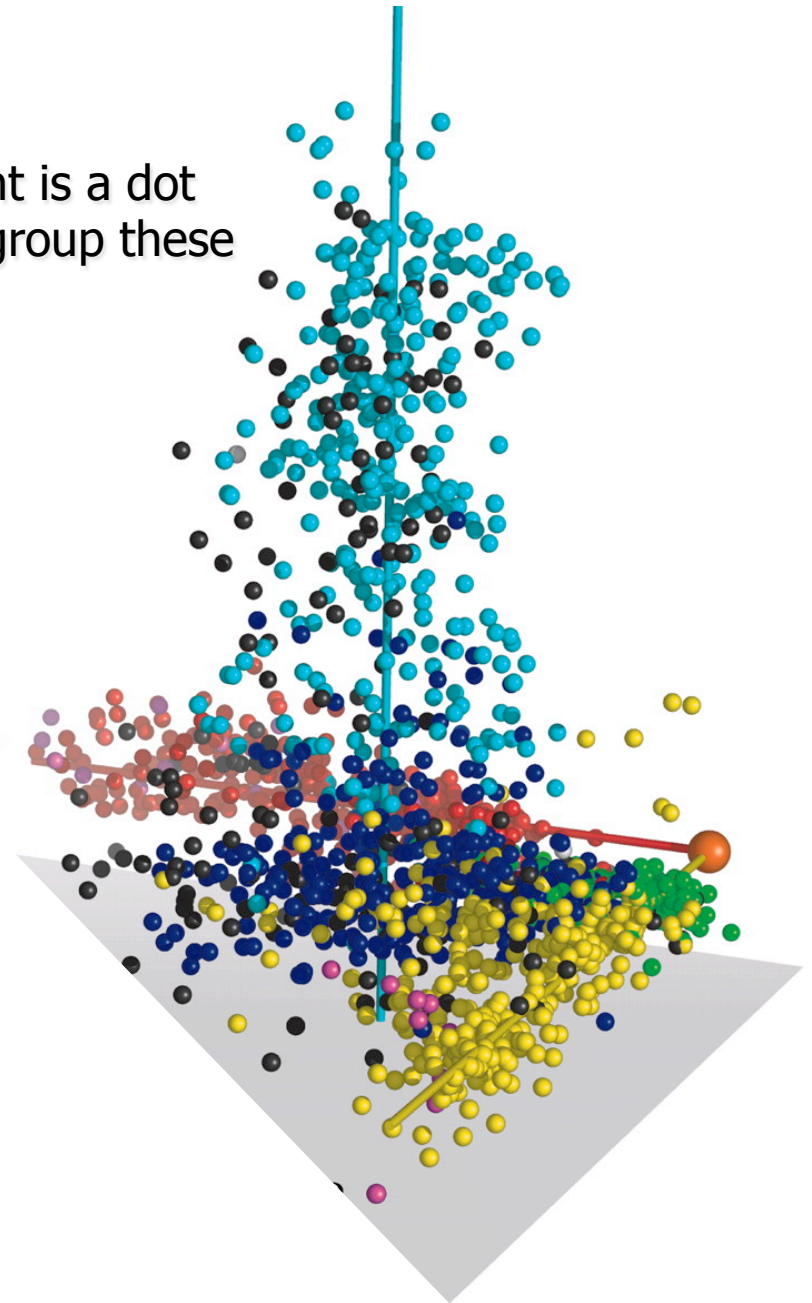


OR

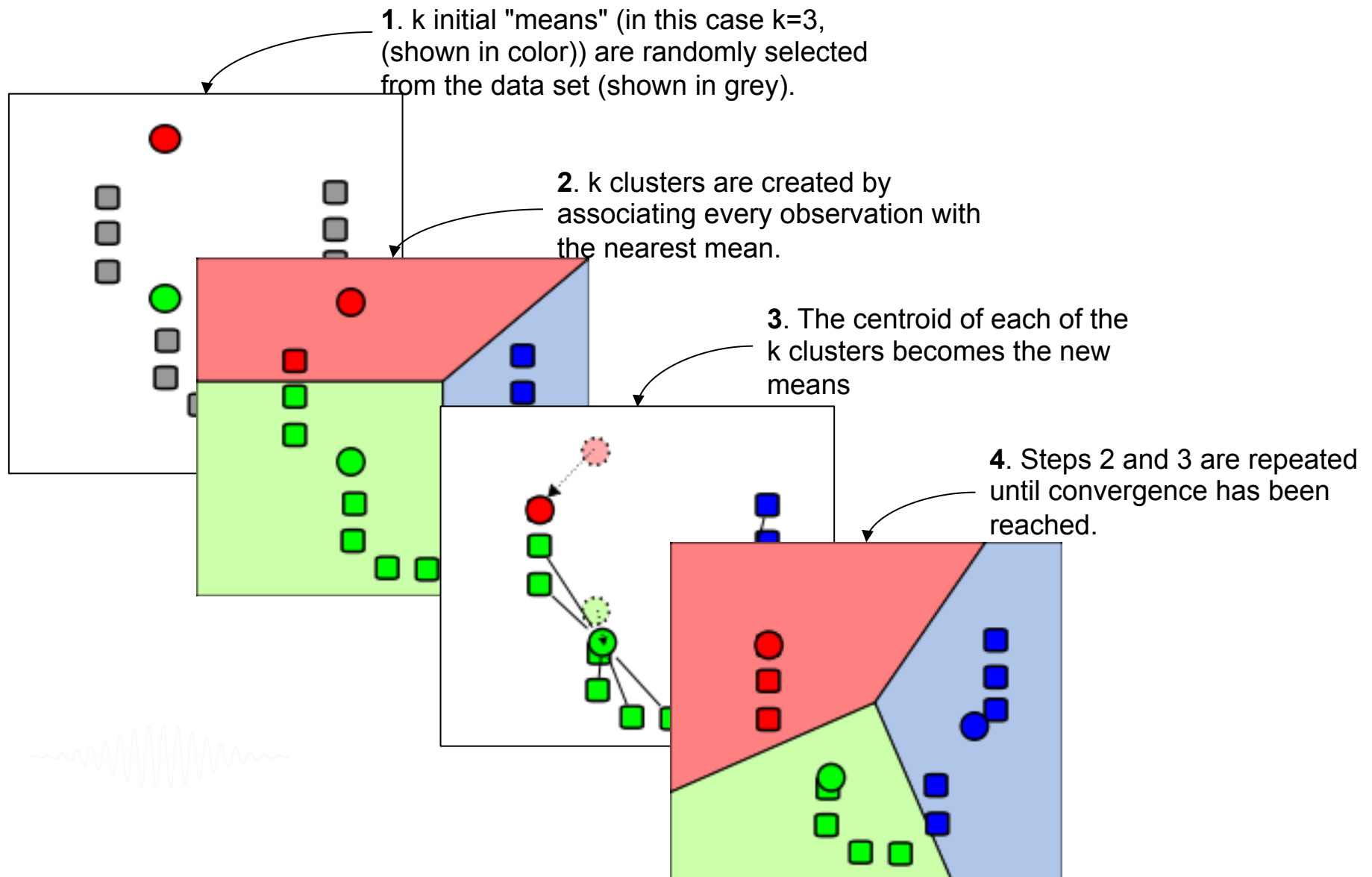
ICs (all subj)



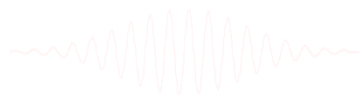
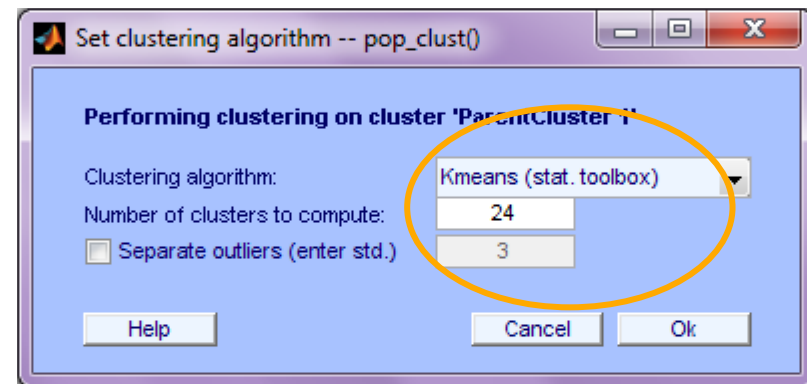
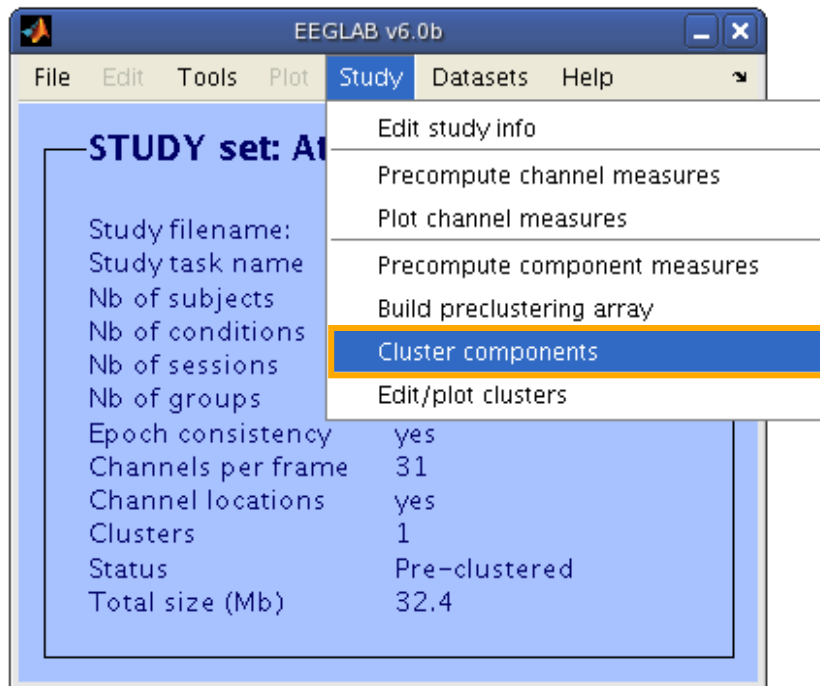
Each component is a dot
Clustering will group these
dots



Classical KMean



Cluster components



Choosing data measures

What measure(s) should you use?

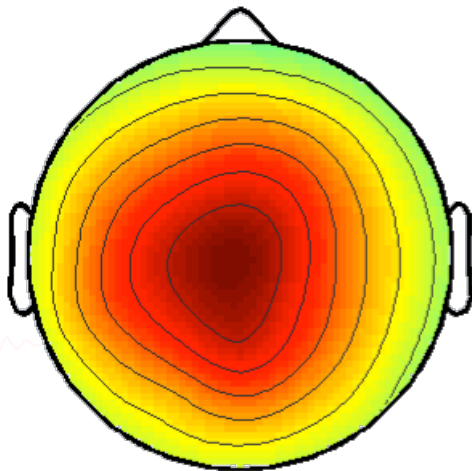
It depends on your final cluster criteria...

- If for example, your priority is dipole location, then cluster only based on dipole location...

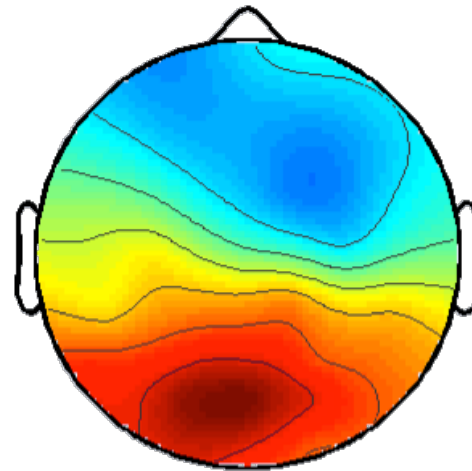
But consider:

- What is the difference between these two components?

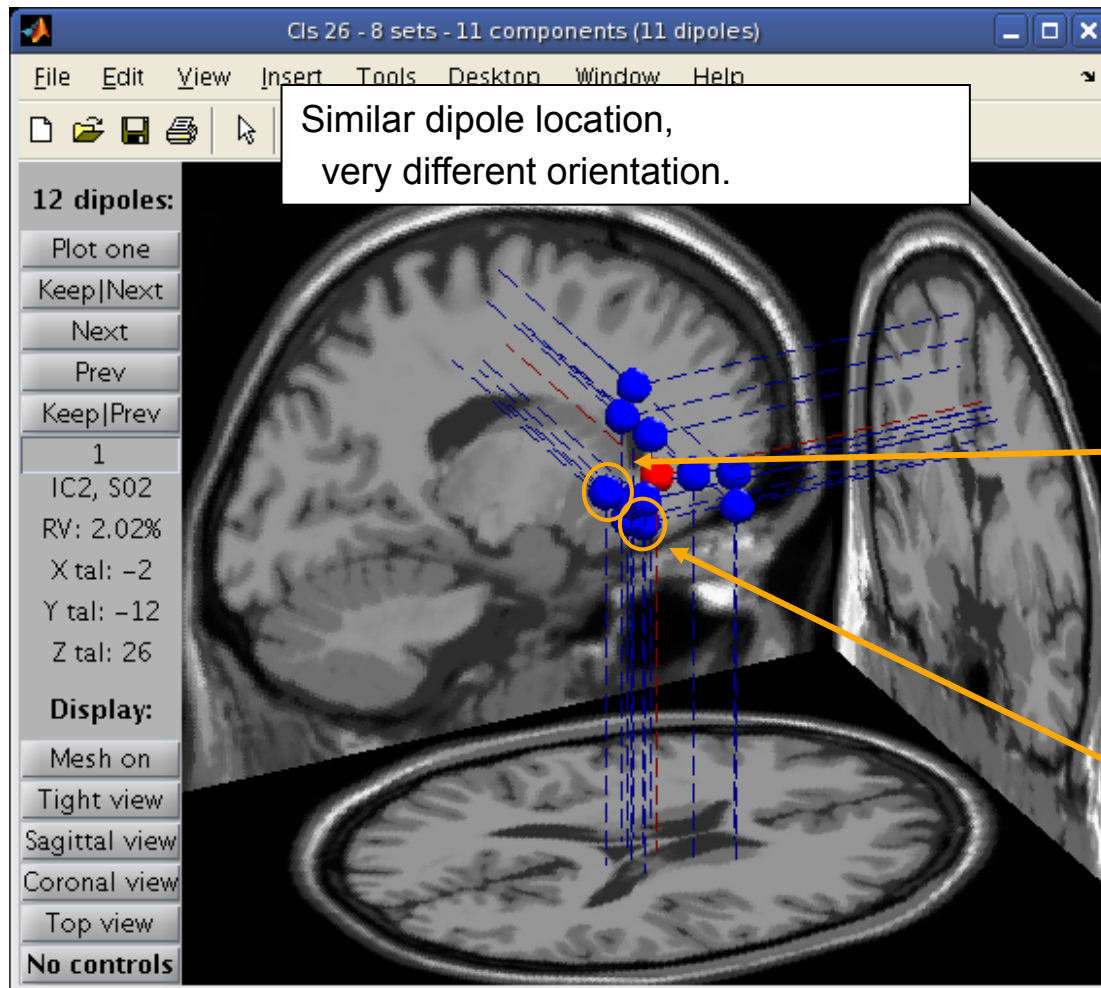
IC2 / S02, Cls 26



IC5 / S05, Cls 26

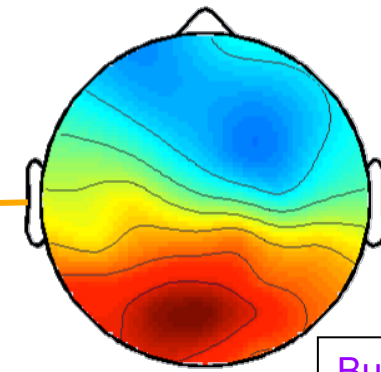


Choosing data measures

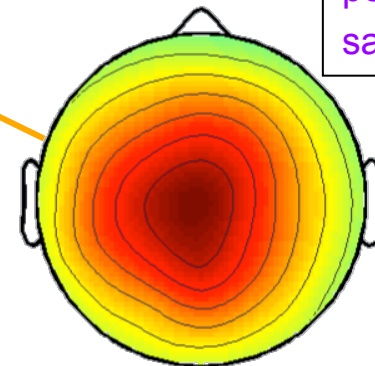


Obvious dramatic effect on scalp map topography:

IC5 / S05, Cls 26

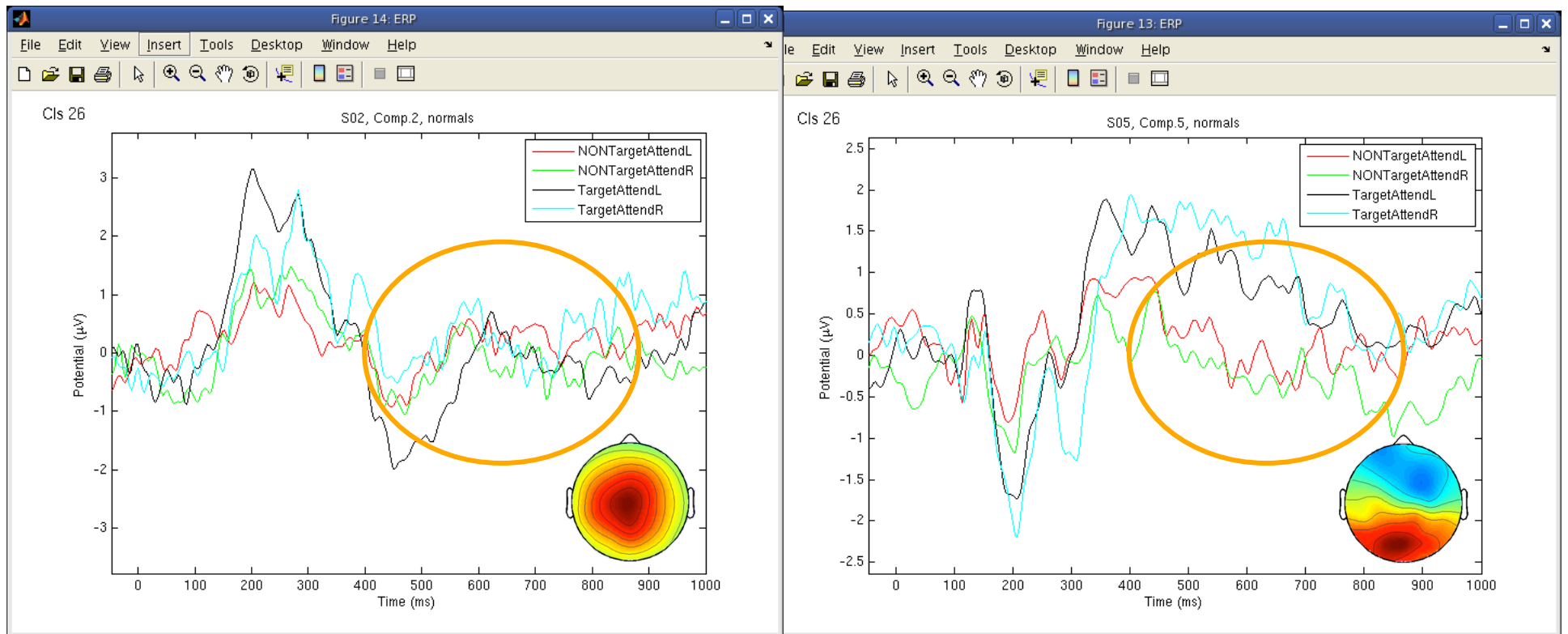


IC2 / S02, Cls 26

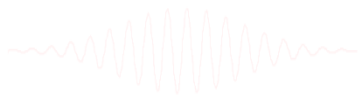


But, do they perform the same functions?

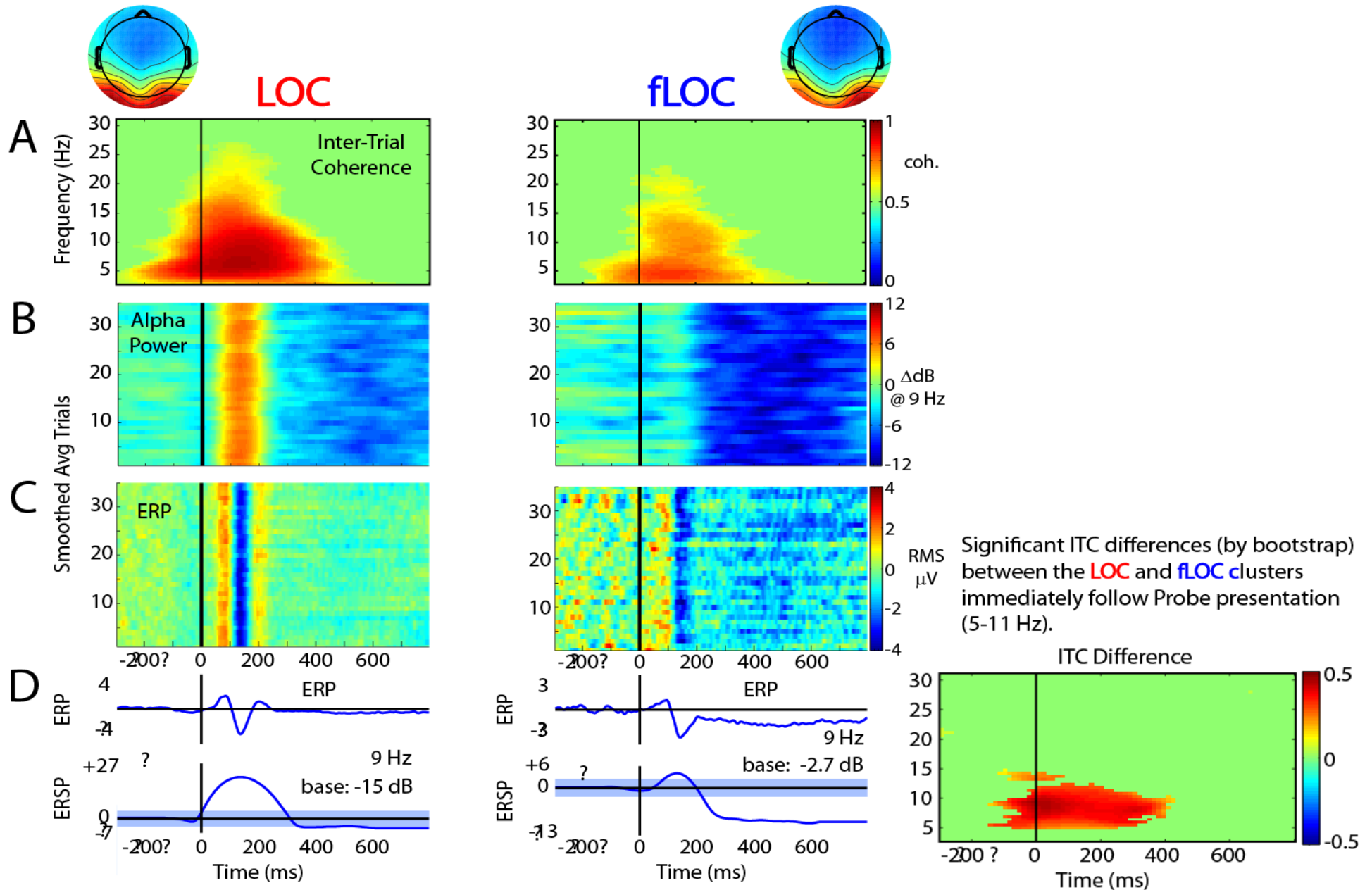
Choosing data measures



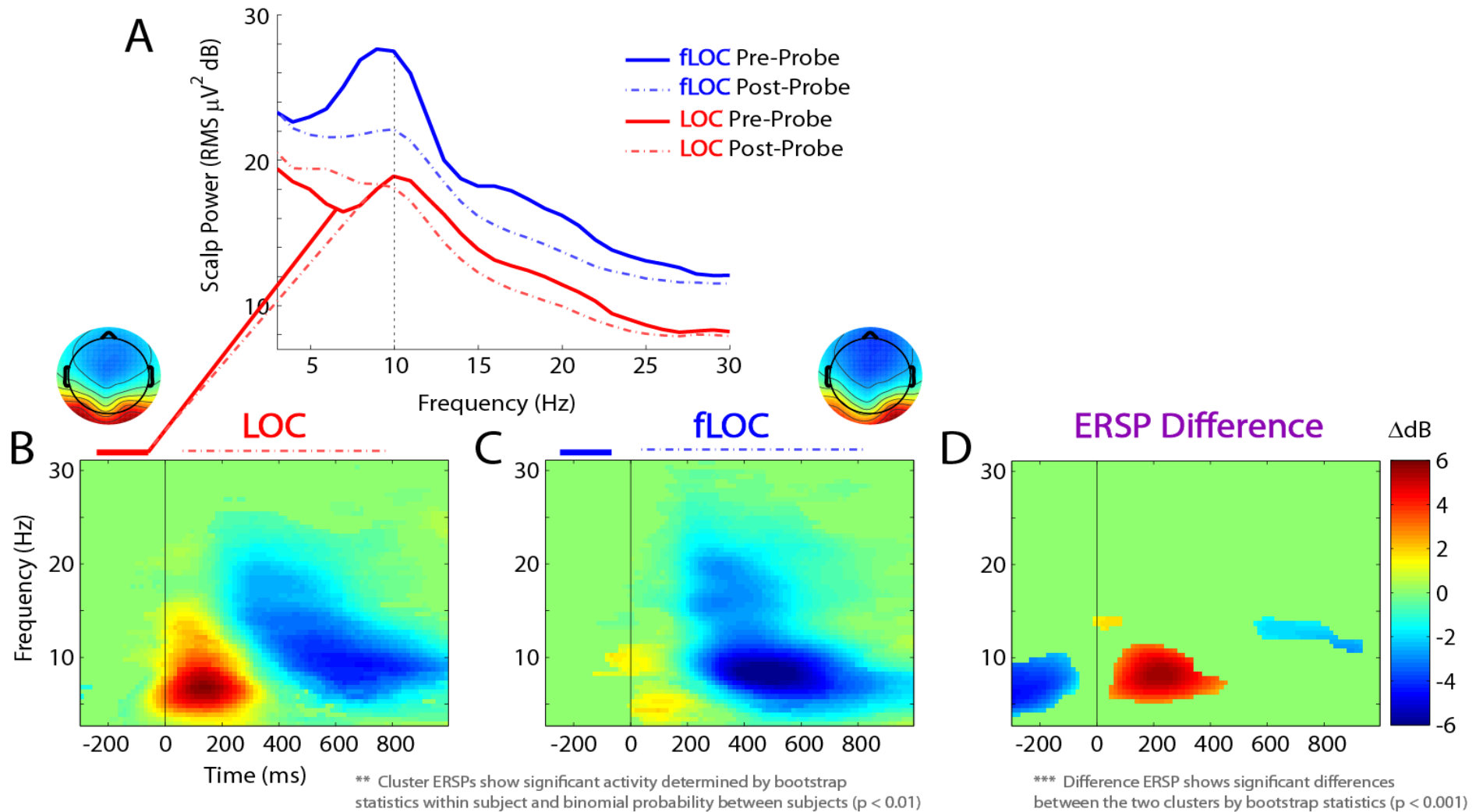
ERPs seem different...



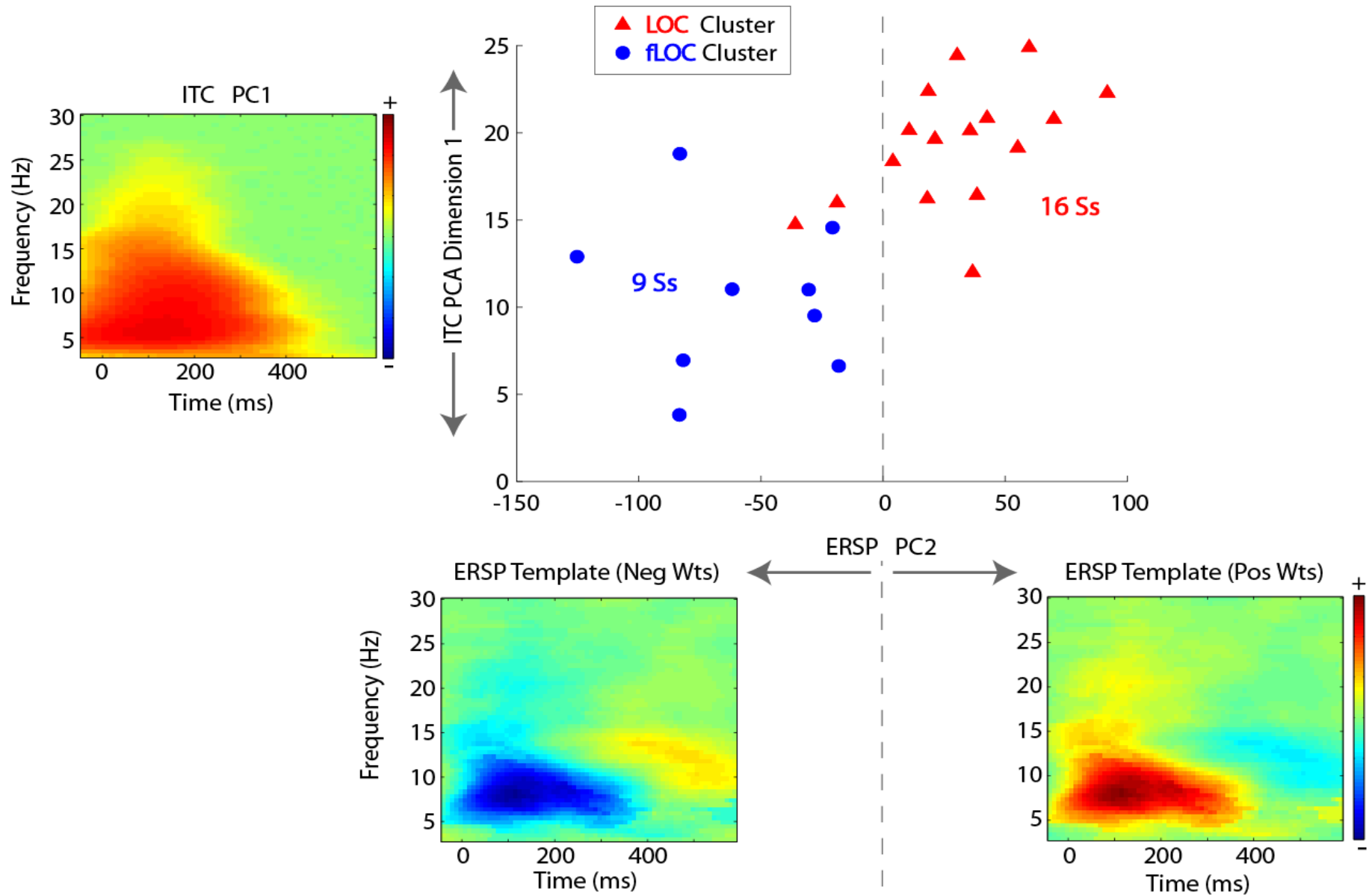
Subject differences?



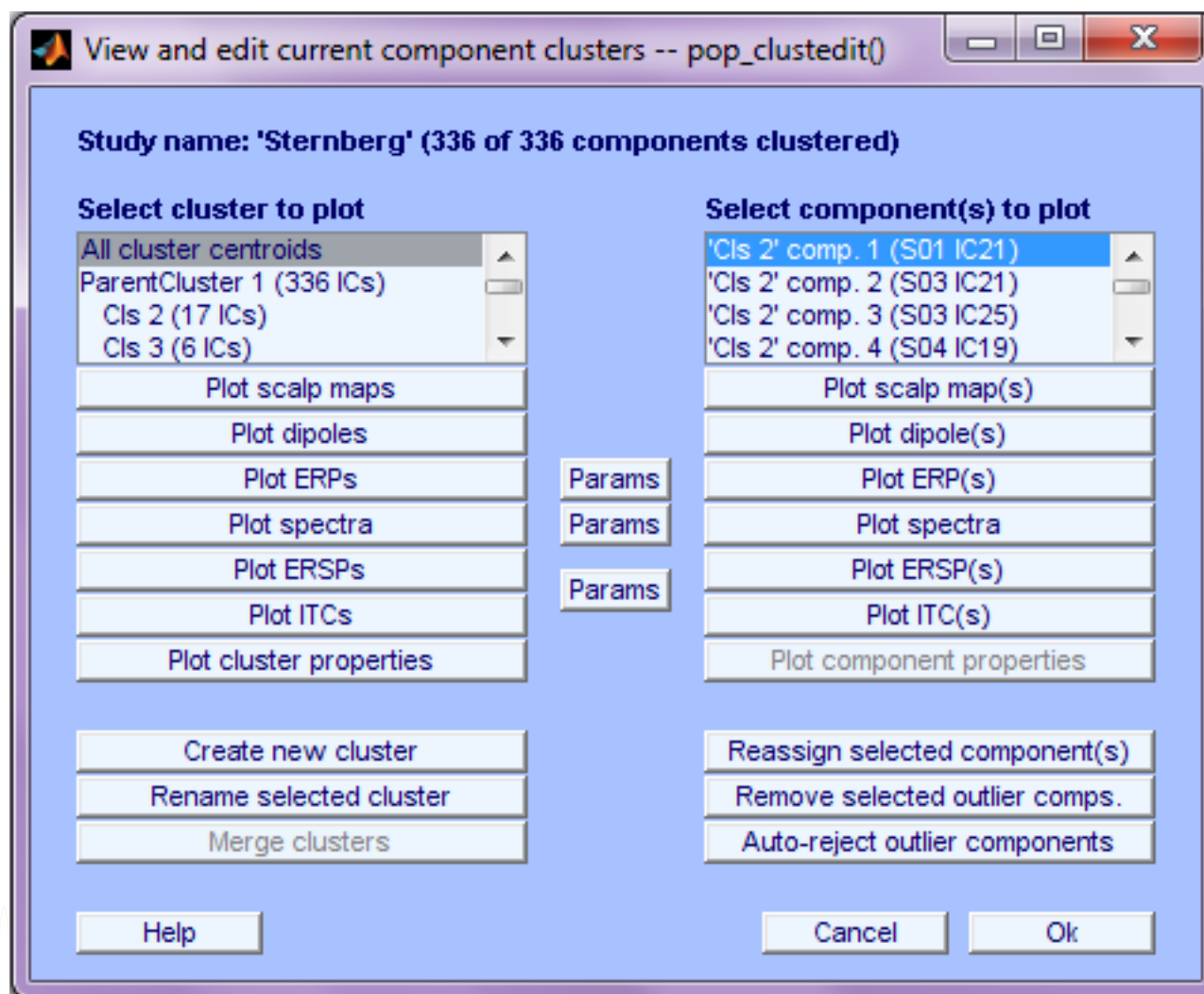
Subject differences?



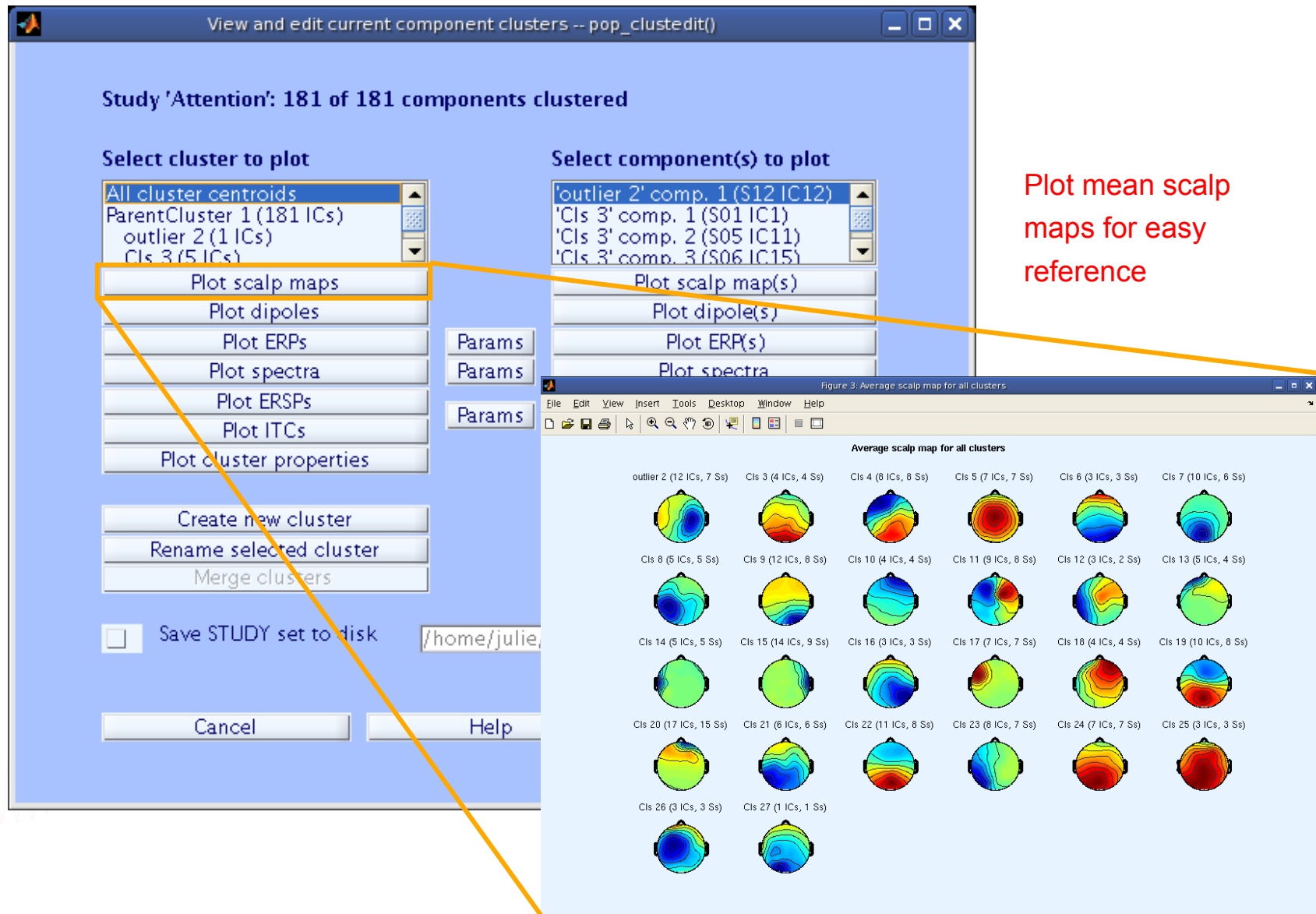
Subject differences?



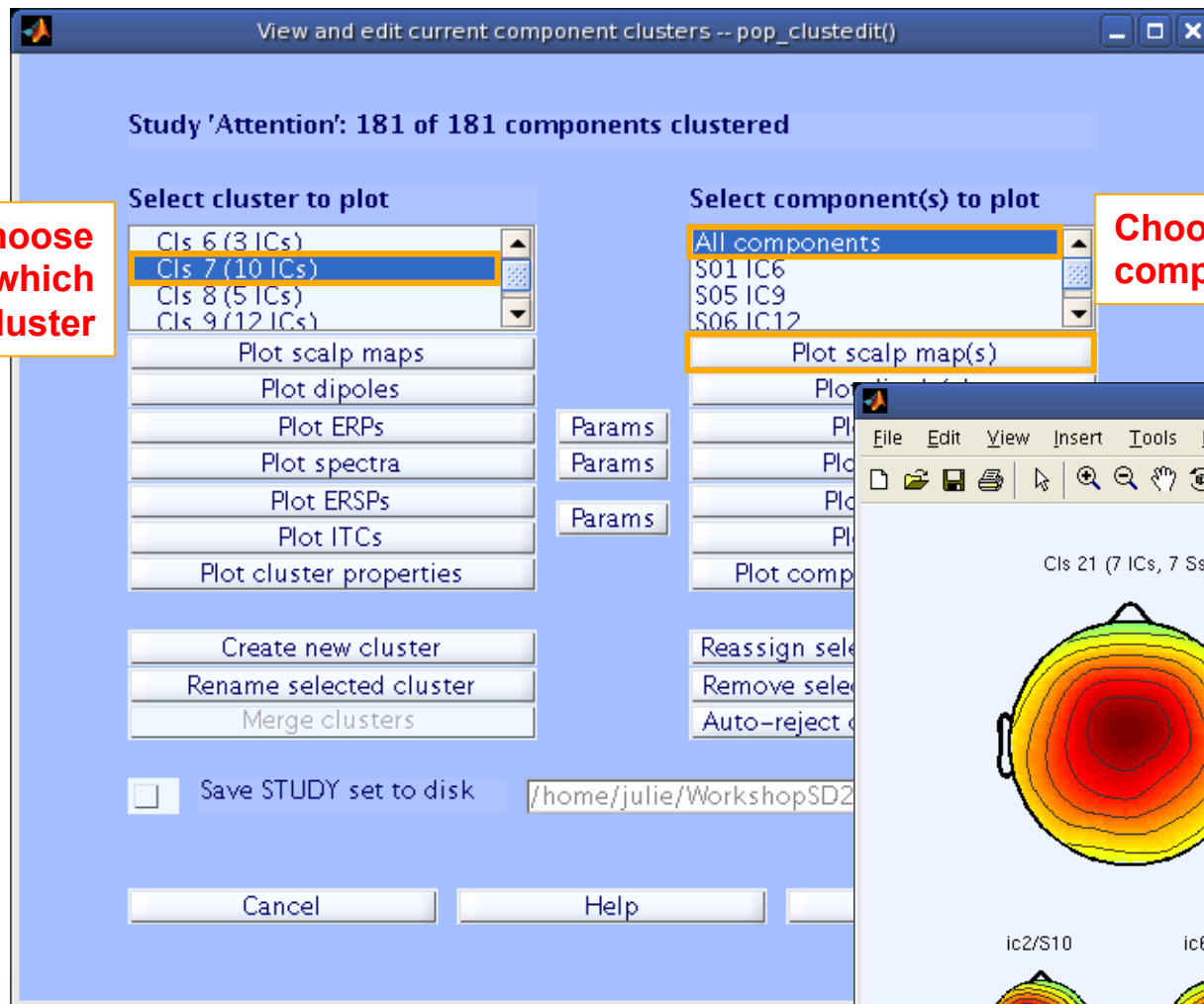
Plot/edit clusters



Plot cluster data

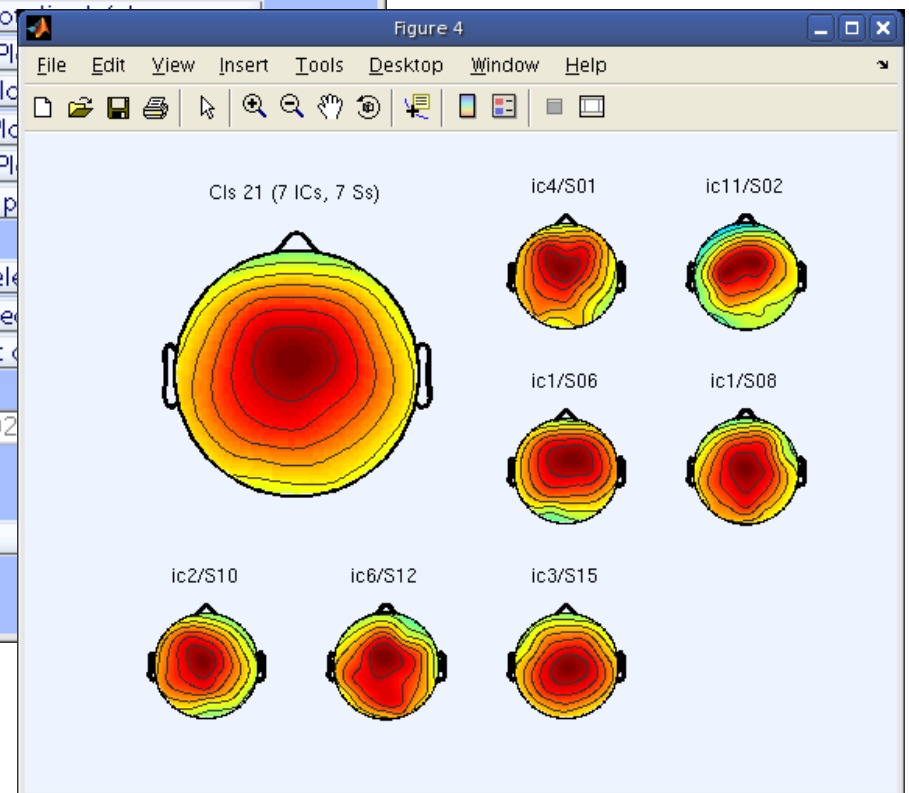


Plot cluster data



Choose
which
cluster

Choose which
components



Plot cluster data

The screenshot displays the MNE software interface for plotting cluster data. The main window, titled "component clusters -- pop_clustedit()", contains several panels and buttons.

Left Panel (15 dipoles):

- Buttons: Plot one, Keep|Next, Next, Prev, Keep|Prev.
- Display: Mesh on, Tight view, Sagittal view, Coronal view, Top view.
- No controls

Brain Plot Window (Top Left):

Cl: 19 - 5 sets - 14 components (14 dipoles)

15 dipoles:

- 1
- IC3, S02
- RV: 2.62%
- X tal: -6
- Y tal: -13
- Z tal: 21

Main Panel (Buttons):

- Plot scalp maps
- Plot dipoles
- Plot ERPs
- Plot spectra
- Plot ERSPs
- Plot ITCs
- Plot cluster properties
- Create new cluster
- Rename selected cluster
- Merge clusters

Right Panel (Select component(s) to plot):

Select component(s) to plot

All components

- S02 IC3
- S02 IC11
- S02 IC12
- S02 IC17

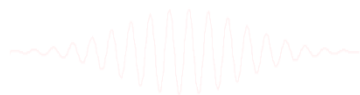
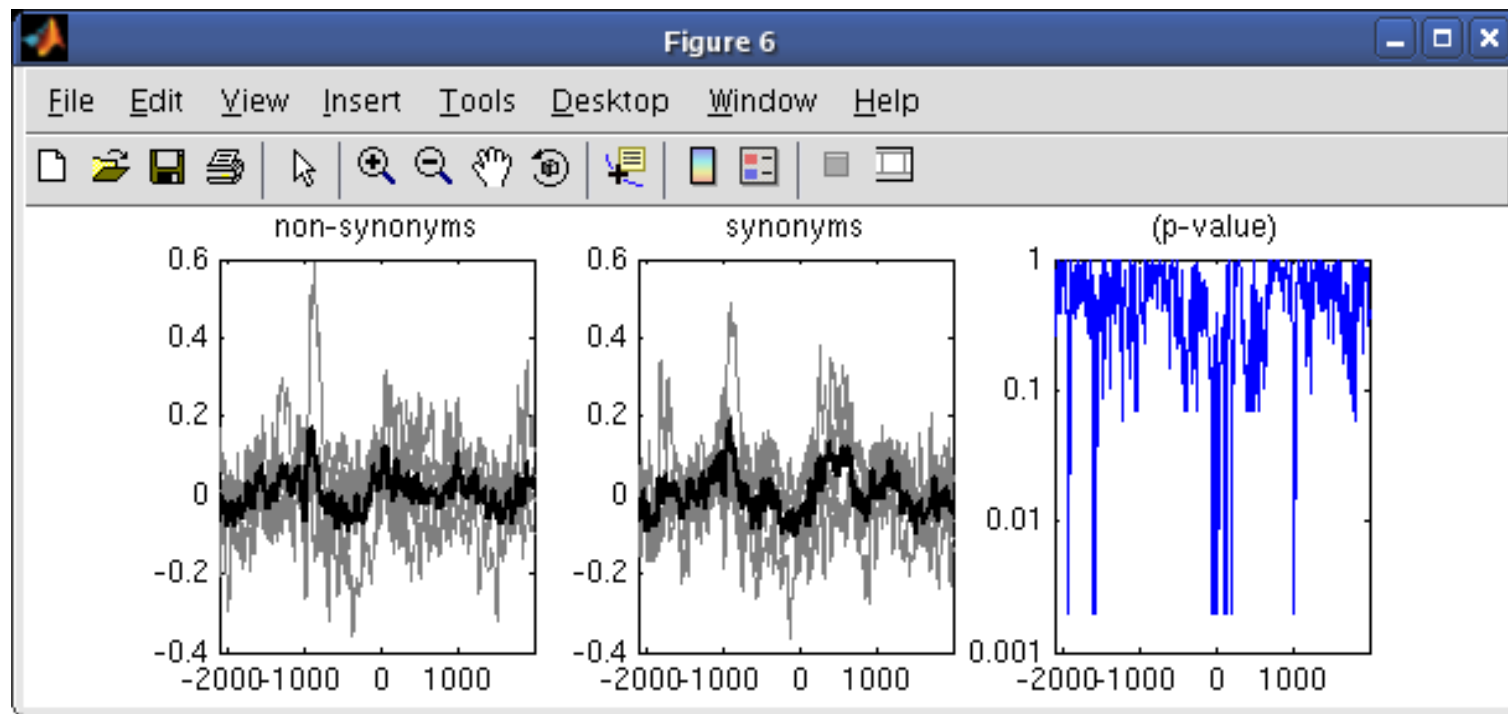
Plot Options (Right Panel):

- Plot scalp map(s)
- Plot dipole(s)** (highlighted with an orange box)
- Plot ERP(s)
- Plot spectra
- Plot ERSP(s)
- Plot ITC(s)
- Plot component properties

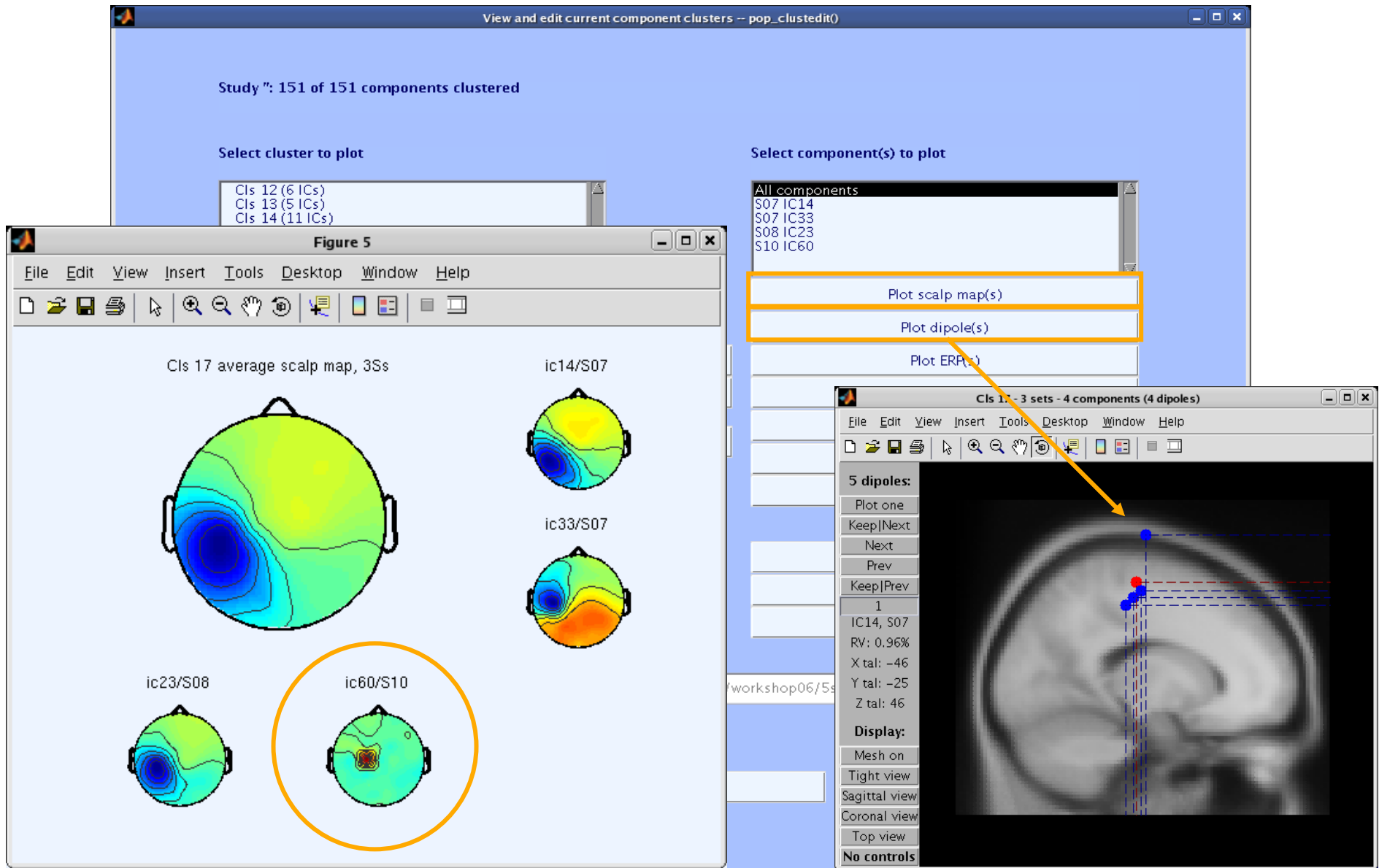
Bottom Panel:

- Params (three buttons)
- Reassign selected component(s)
- Remove selected outlier comps.
- Auto-reject outlier components
- Save STUDY set to disk (checkbox)
- File path: /home/julie/workshop06/5subjects/WSstudy.study
- Buttons: Cancel, Help, Ok

Plot cluster ERP



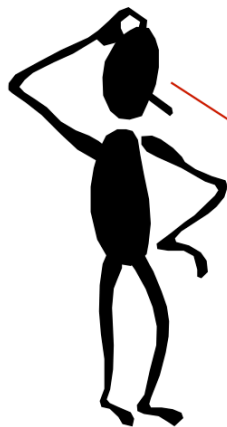
Reassigning components



Issue with standard clustering

Large parameter space problem: many different clustering solutions can be produced by changing parameters and measure subsets. Which one should we choose?

EEGLAB clustering has ~12 parameters



Select and compute component measures for later clustering -- pop_preclust()

Pre-compute measures on which to cluster components from study 'N400STUDY'
Select the cluster to refine during sub-clustering (any existing sub-hierarchy will be overwritten)

ParentCluster 1 (151 ICs)

Pre-compute or Load	Dims.	Norm.	Rel. Wt.
<input checked="" type="checkbox"/> spectra	10	<input checked="" type="checkbox"/> 1	Frequency range [Hz]
<input checked="" type="checkbox"/> ERPs	10	<input checked="" type="checkbox"/> 1	Latency range in ms [lo hi]
<input checked="" type="checkbox"/> dipoles	3	<input checked="" type="checkbox"/> 10	
<input checked="" type="checkbox"/> scalp maps	10	<input checked="" type="checkbox"/> 1	Use channel values <input type="checkbox"/>
<input checked="" type="checkbox"/> ERSPs	10	<input checked="" type="checkbox"/> 1	Time/freq. parameters
<input checked="" type="checkbox"/> ITCs	10	<input checked="" type="checkbox"/> 1	Time/freq. parameters
<input checked="" type="checkbox"/> Final dimensions	10	Help	

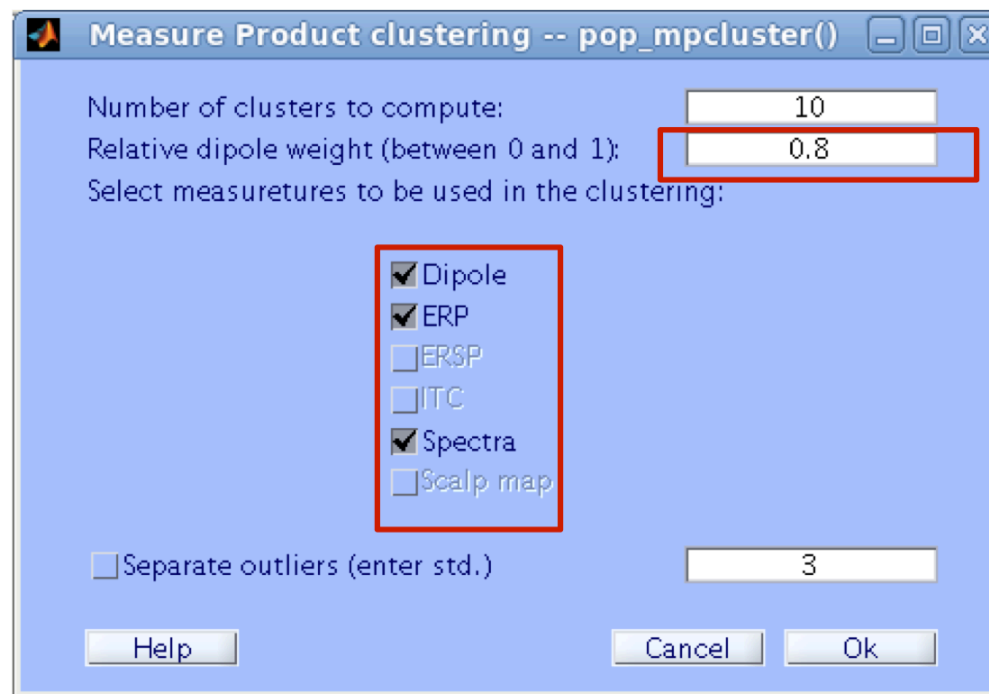
☒ Save STUDY to file /data/common4/xxw/5subjects/N400precluststudy

Cancel Help Ok



Measure projection

(EEGLAB extension by Nima Bigdely Shamlo)
only has one pre-clustering parameter.



(Affinity clustering by Pernet, Martinez, Delorme)

Exercise

- Load the STUDY
- Precluster and cluster components using spectrum and dipoles location
- Look at your cluster. Identify frontal midline theta cluster and occipital alpha cluster
- Plot significant difference for one component cluster spectrum between the two conditions in the default design

