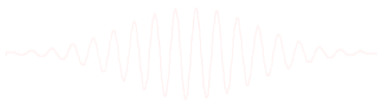


# Clustering of ICA components

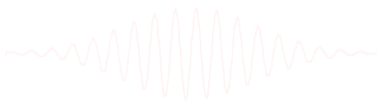
Arnaud Delorme

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



# Steps of clustering

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



# Edit dataset info

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

pop\_study(): Pre-select components

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

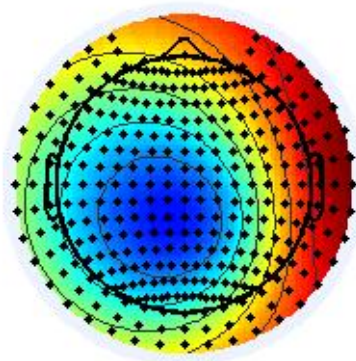
15

Keep only in-brain dipoles.

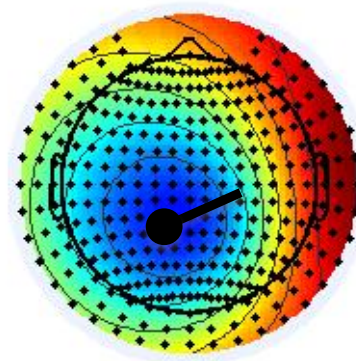
Cancel Help Ok

# Computing residual variance (%)

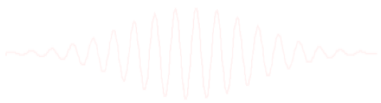
Actual



Dipole projection

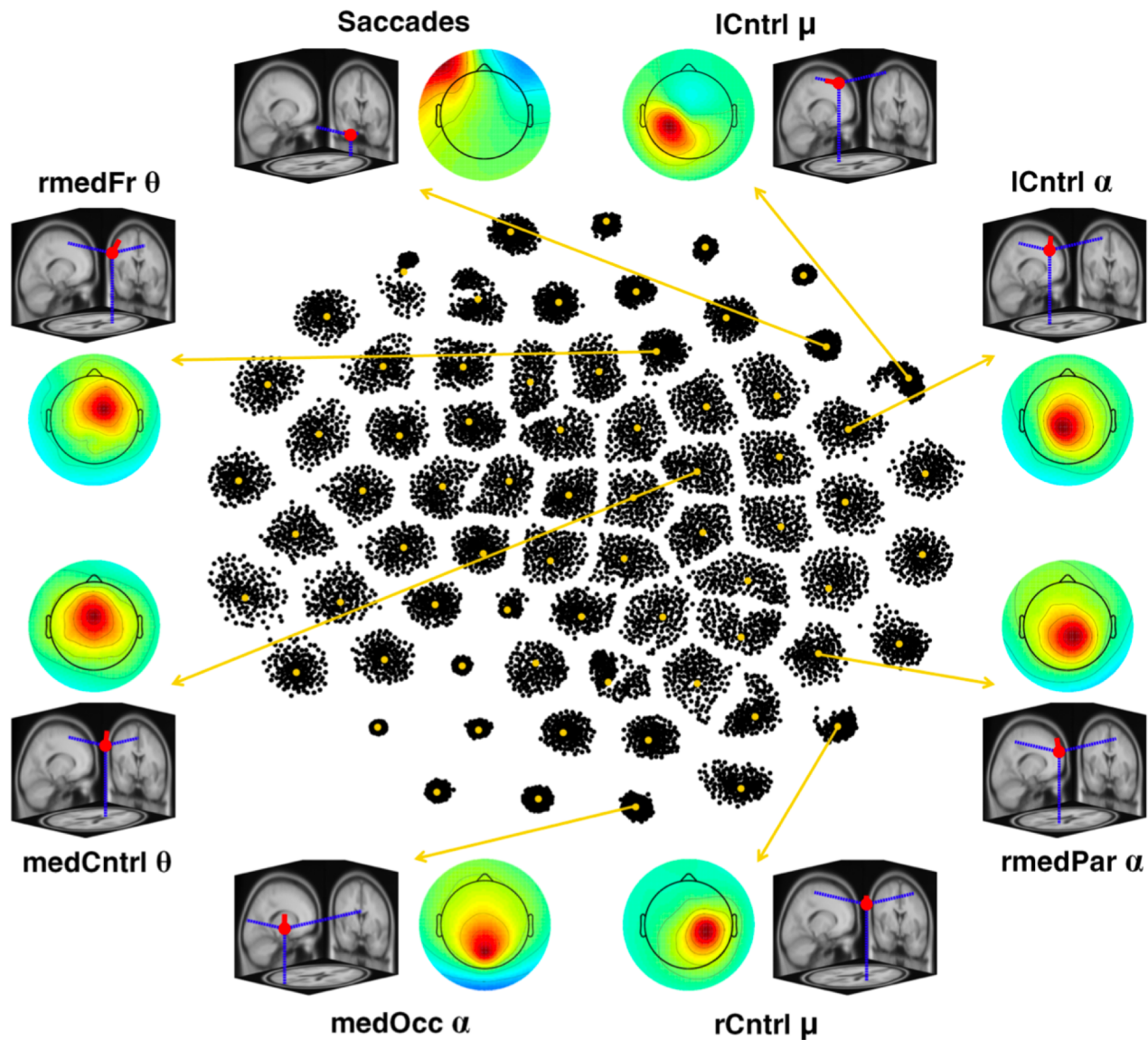


$$r = \frac{\sum(x_i - \tilde{x}_i)^2}{\sum x_i^2}$$





## Clustering results example

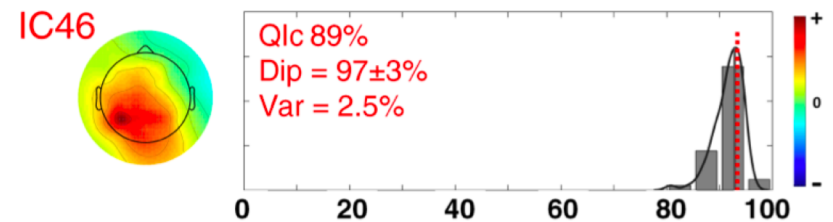
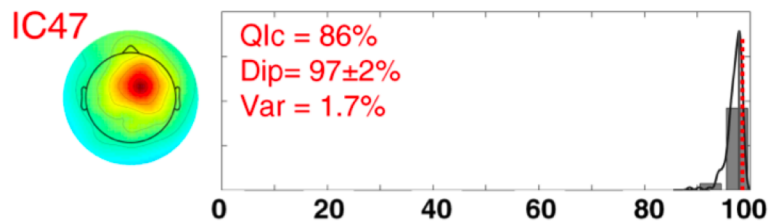
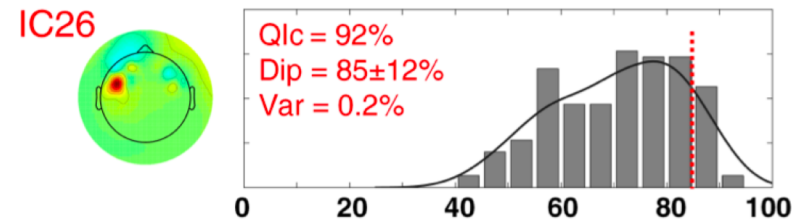
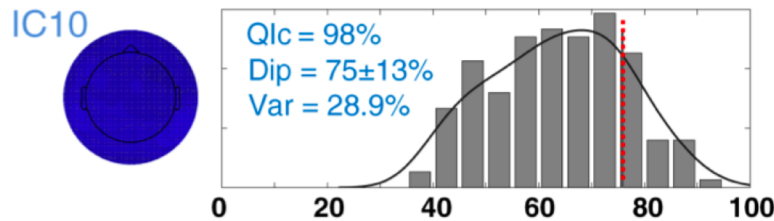
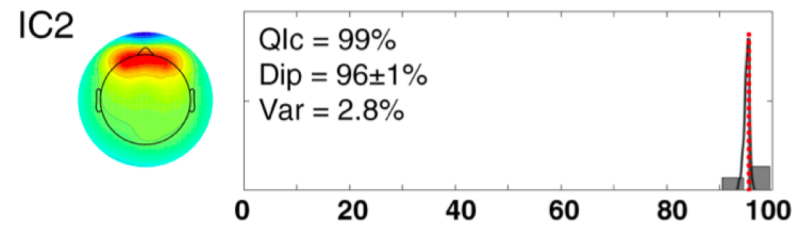
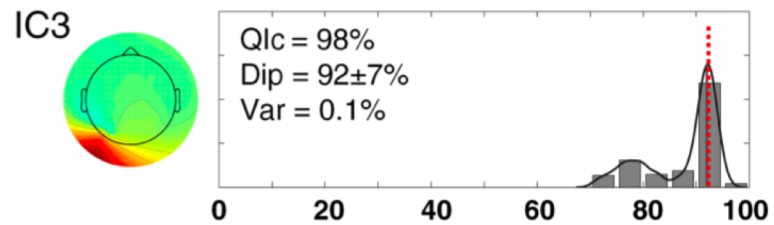
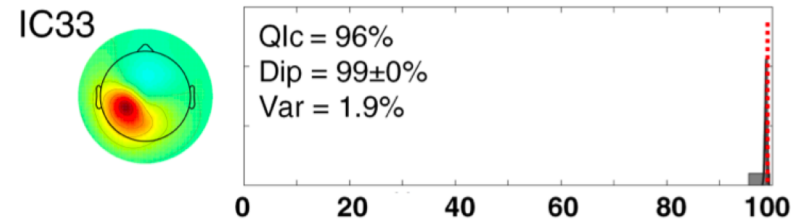
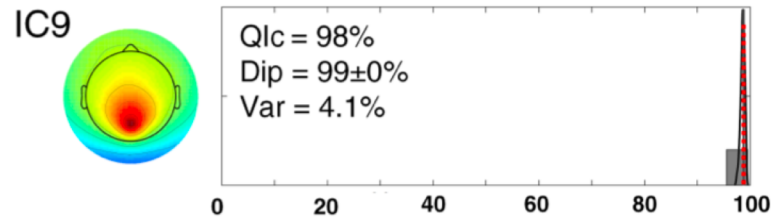


RELICA: A method for estimating the reliability of independent components

Fiorenzo Artoni <sup>a,\*</sup>, Danilo Menicucci <sup>b</sup>, Arnaud Delorme <sup>c,e,f</sup>, Scott Makeig <sup>c</sup>, Silvestro Micera <sup>a,d</sup>

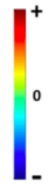
# Within-cluster reliability

The distribution of dipolarity within the cluster helps assessing the **quality** and characteristics of Independent Components



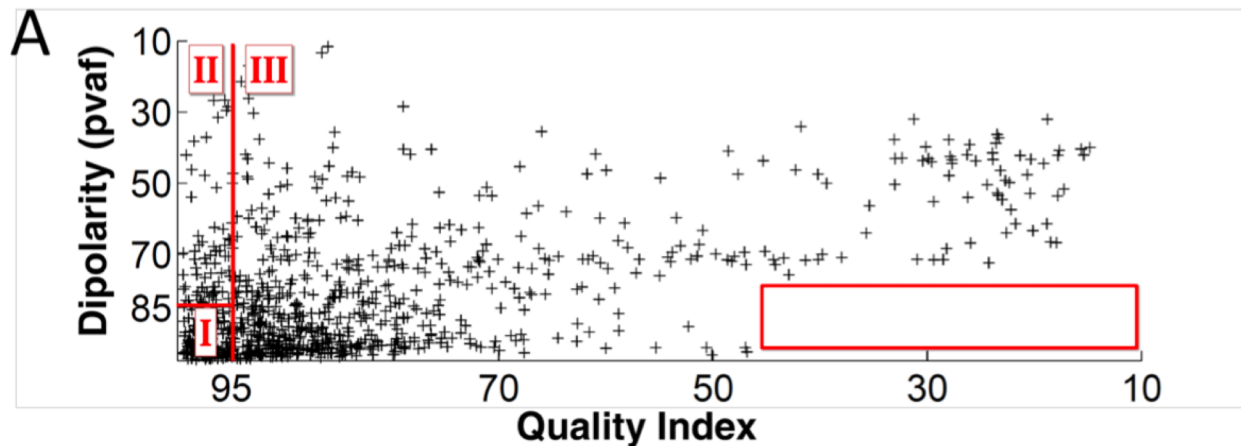
Dipolarity

Dipolarity



# Reliability criteria and the $rv < 15\%$

First justification why we should select an  $rv < 15\%$  for components to include in further analyses: there is a forbidden region underlined in red, that indicates the absence of



**CLASS I**

Quality Index and Dipolarity above Retention threshold: **Good**

**CLASS II**

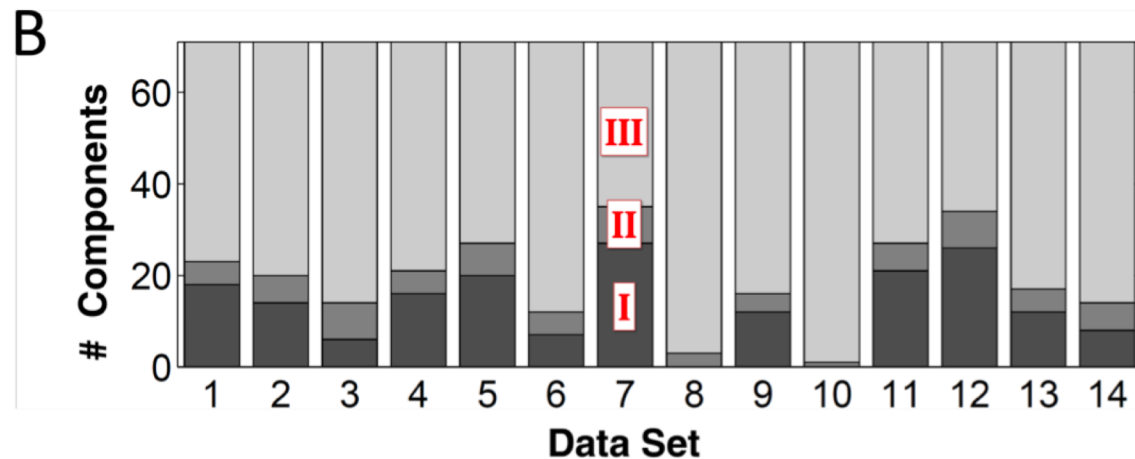
Quality Index above threshold, dipolarity below: **artifact** or mixing of multiple processes

**CLASS III**

Quality Index below retention threshold

$dip \pm std > th$

$dip \pm std < th$



Probable Inseparable noise: variance explained useful or **multiple subject confirmation** **Discard**

# ICs to cluster

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

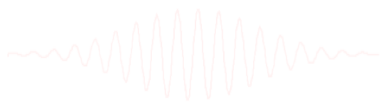
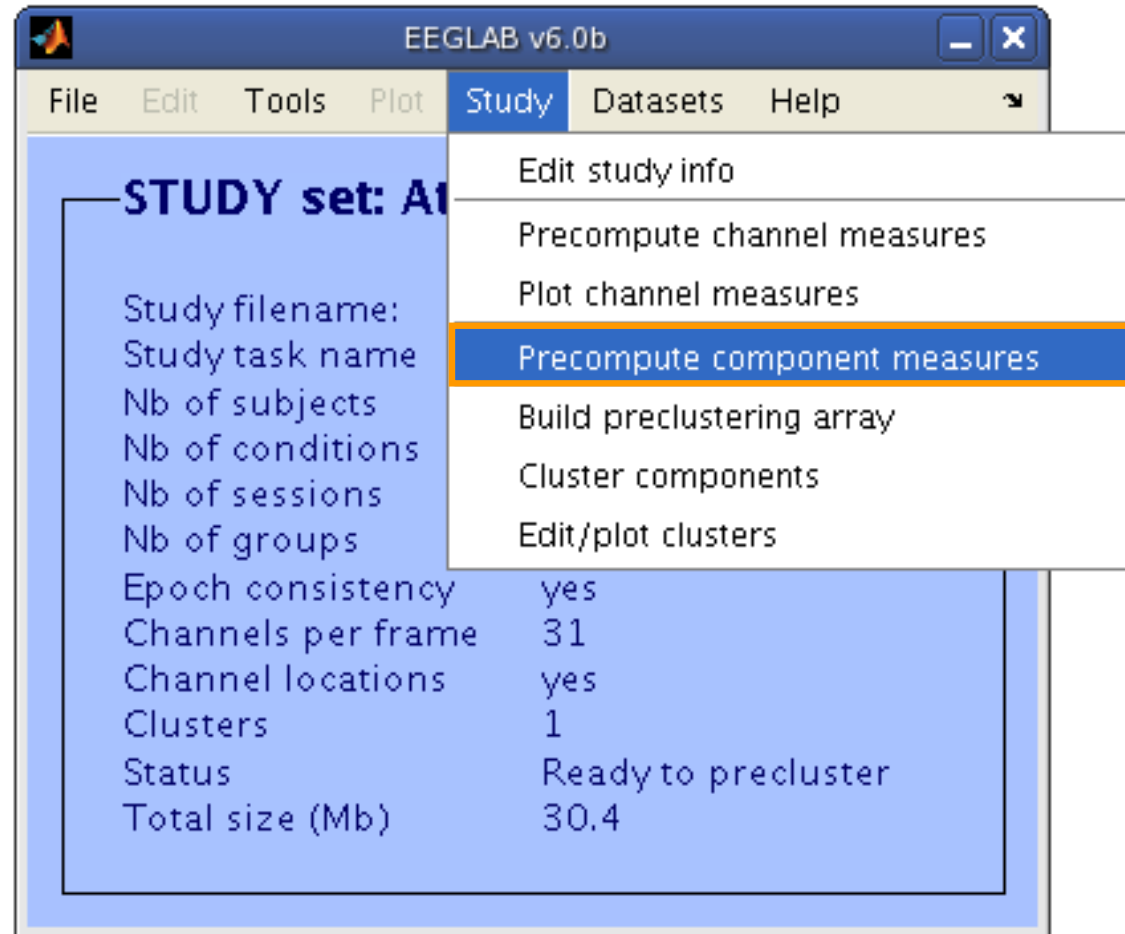
< Page 1 >

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.)

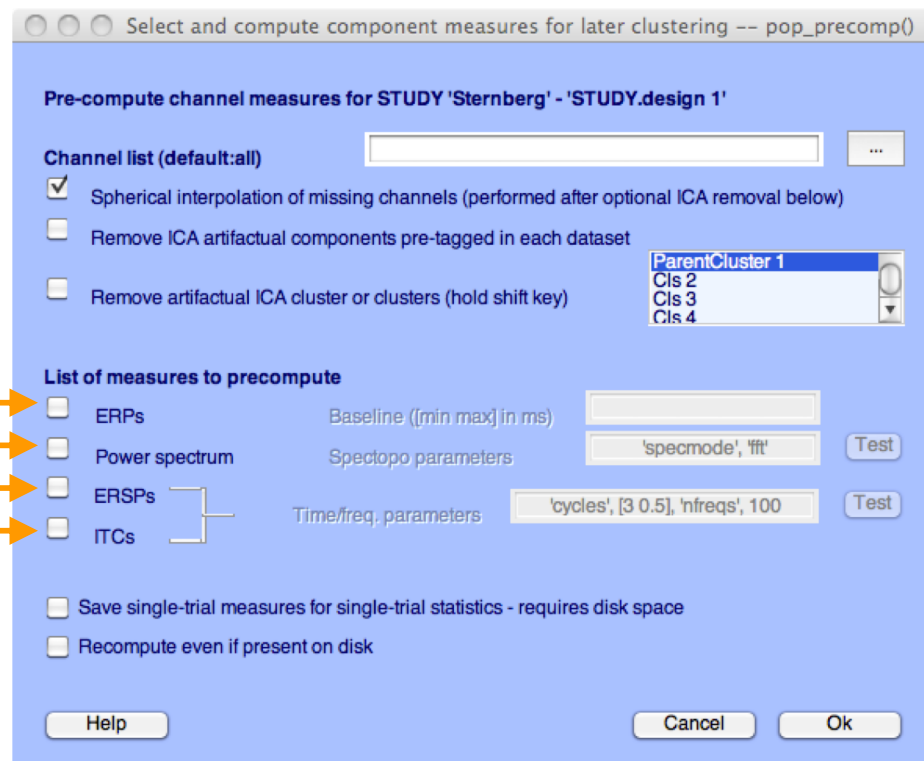
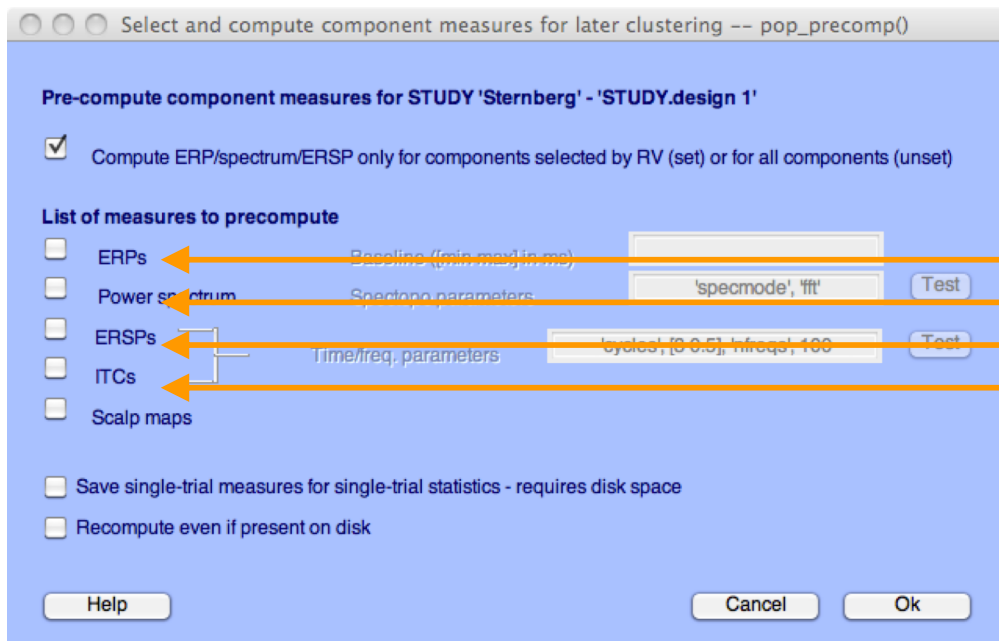
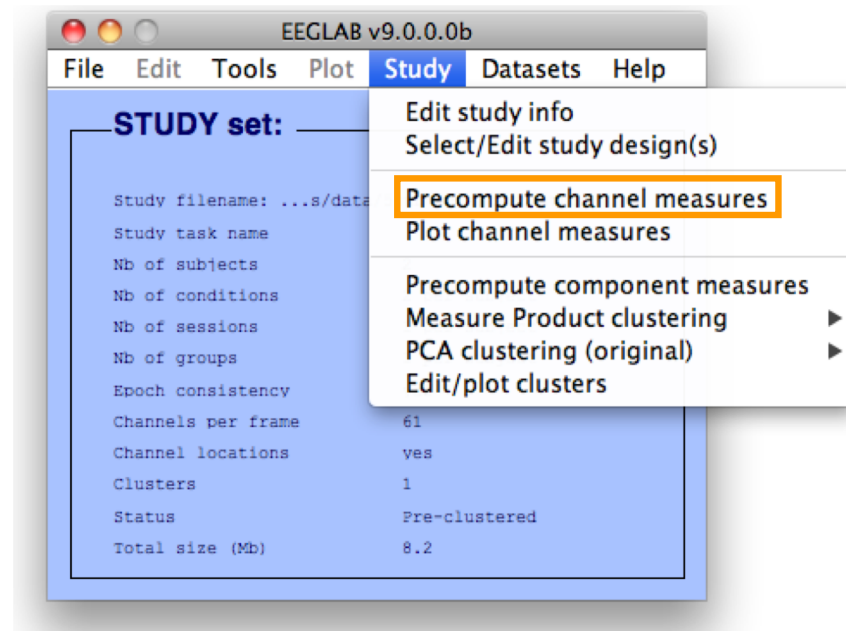
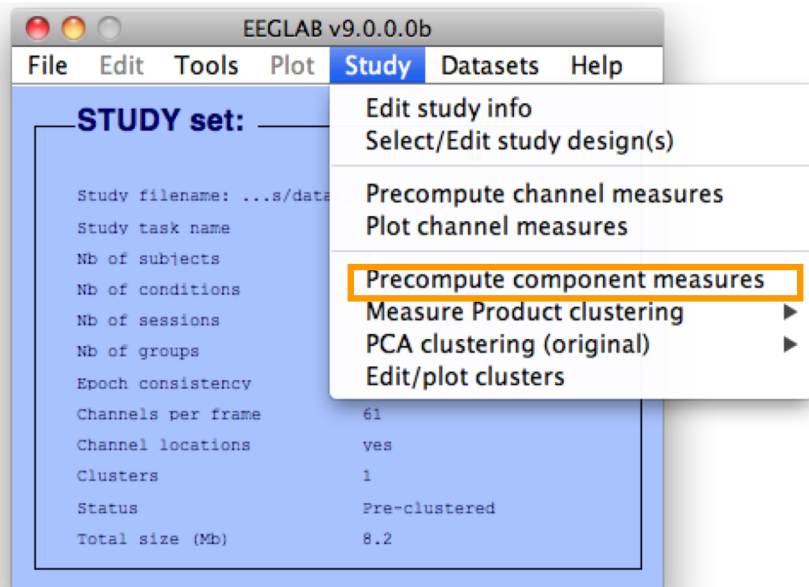
Help Cancel Ok

# Precompute data measures





# Pre-compute 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**

ERPs Baseline ([min max] in ms) [-200 0]

Power spectrum Spectopo parameters Test

ERSPs Time/freq. parameters 'cycles', [3 0.5], 'nfreqs', 100 Test

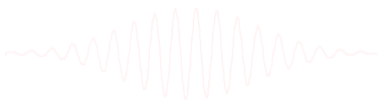
ITCs

Scalp maps

Recompute even if present on disk

Help Cancel Ok

**Time-frequency options**



# Cluster components

EEGLAB v15.x (dev)

File Edit Tools Plot **Study** Datasets Help

**STUDY set: Sternberg**

Study filename:  
Study task name  
Nb of subjects  
Nb of conditions  
Nb of sessions  
Nb of groups  
Epoch consistency: ves  
Channels per frame: 69.70.71  
Channel locations: ves  
Clusters: 1  
Status: Ready to precluster  
Total size (Mb): 229.3

Study menu options:

- Edit study info
- Select/Edit study design(s)
- Precompute channel measures
- Plot channel measures
- Precompute component measures
  - PCA clustering (original)** ▶ Build preclustering array
  - Edit/plot clusters
  - Cluster components

Select and compute component measures for later clustering -- pop\_preclust()

**Build pre-clustering matrix for STUDY set: Sternberg**

Only measures that have been precomputed may be used for clustering

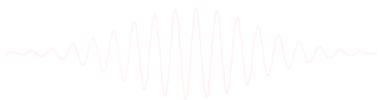
Mixing time-based and location-based measures to cluster might result in Help

Time-based info	PCA	Weight			
<input type="checkbox"/> spectra	10	1	Freq. range [Hz]	3 25	
<input type="checkbox"/> ERPs	10	1	Time range [ms]		
<input type="checkbox"/> ERSPs	10	1	Time range [ms]		Freq. range [Hz]
<input type="checkbox"/> ITCs	10	1	Time range [ms]		Freq. range [Hz]

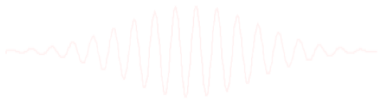
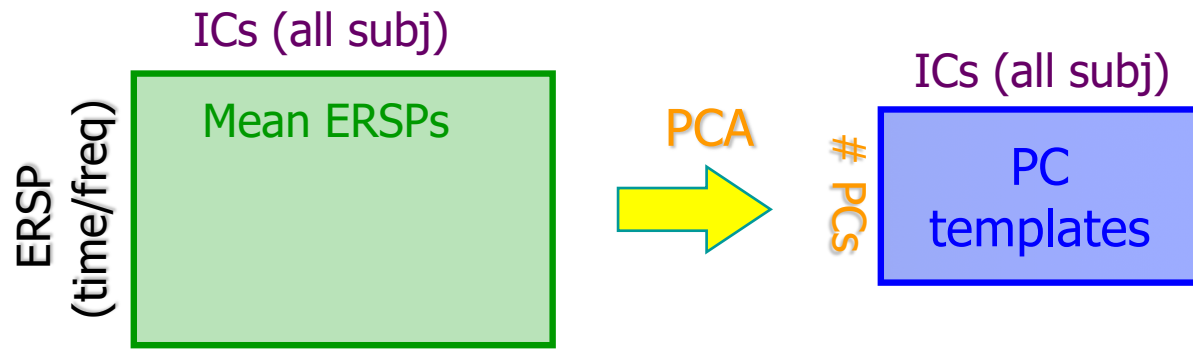
Location-based info	PCA	Weight	
<input checked="" type="checkbox"/> dipole locations	3	1	
<input checked="" type="checkbox"/> dipole orient.	3	1	Amplitude & polarity is ignored
<input type="checkbox"/> scalp maps	10	1	Use channel v... <input checked="" type="checkbox"/> Absolute values

Help Cancel Ok

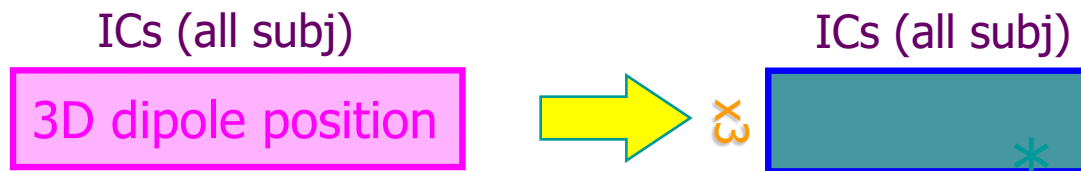
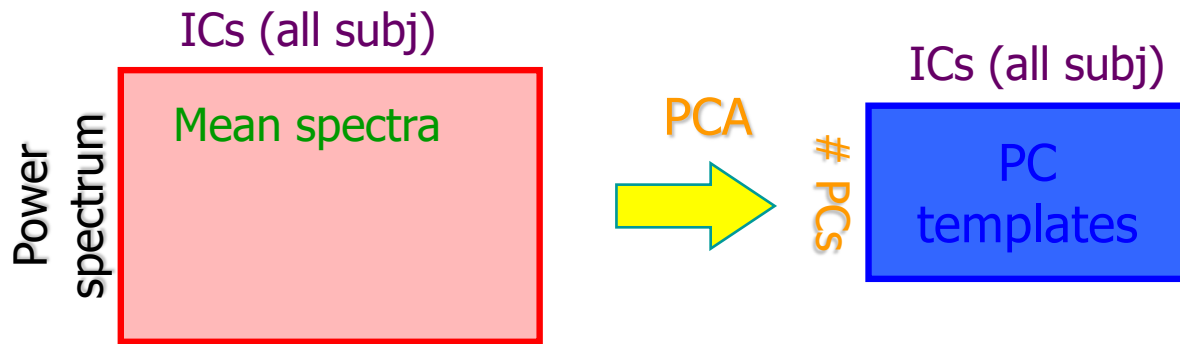
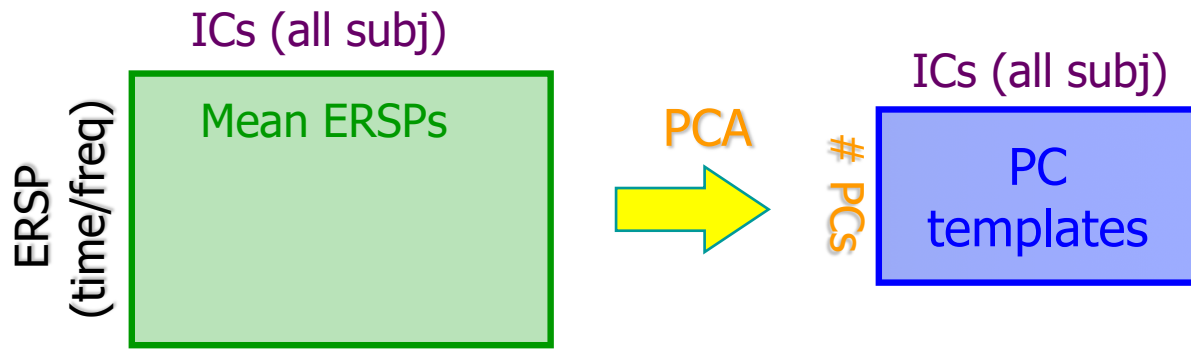




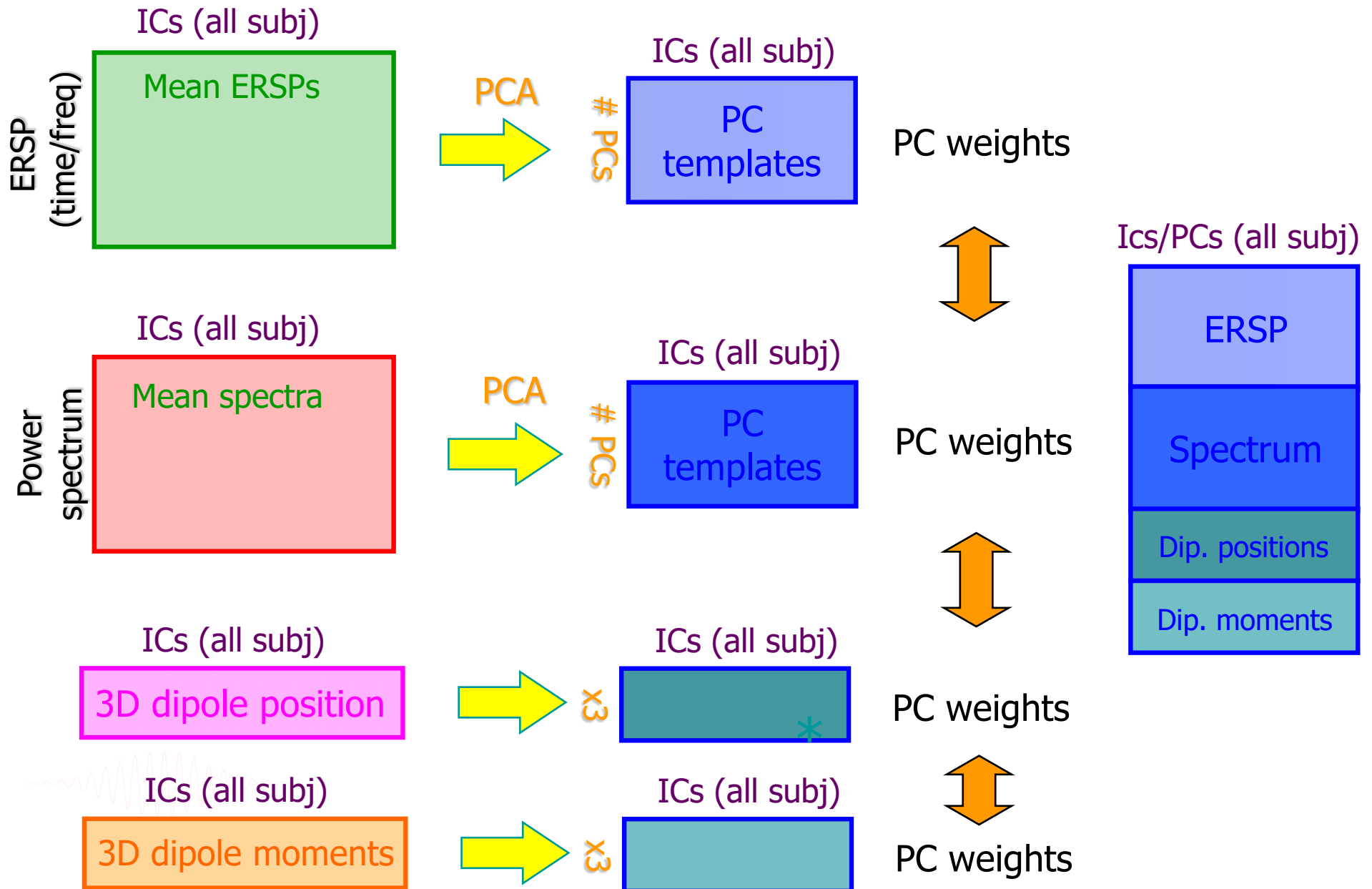
# Precluster schematic



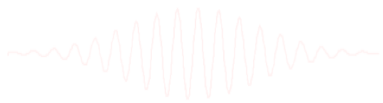
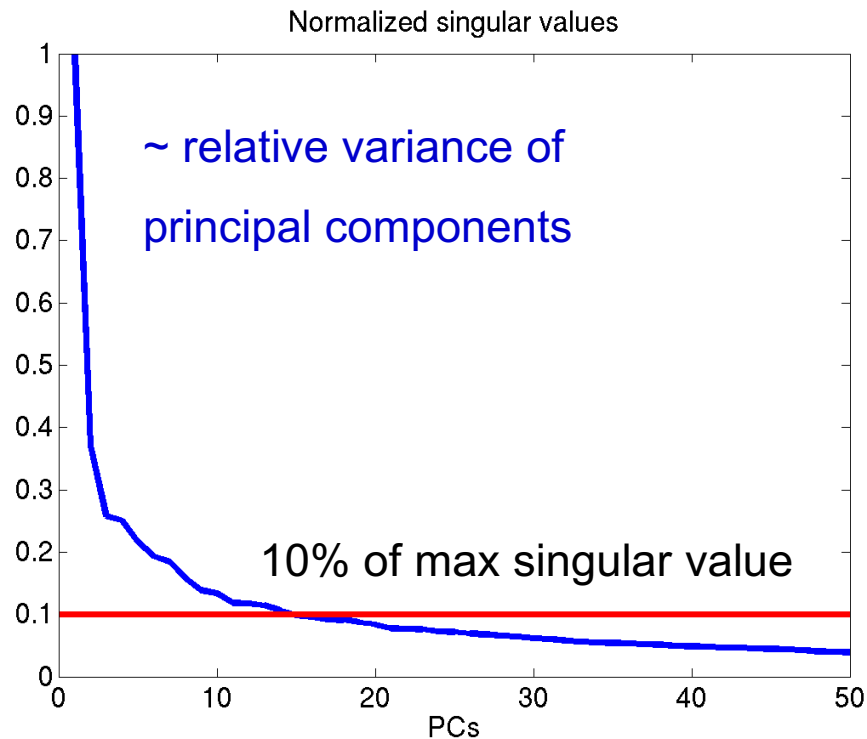
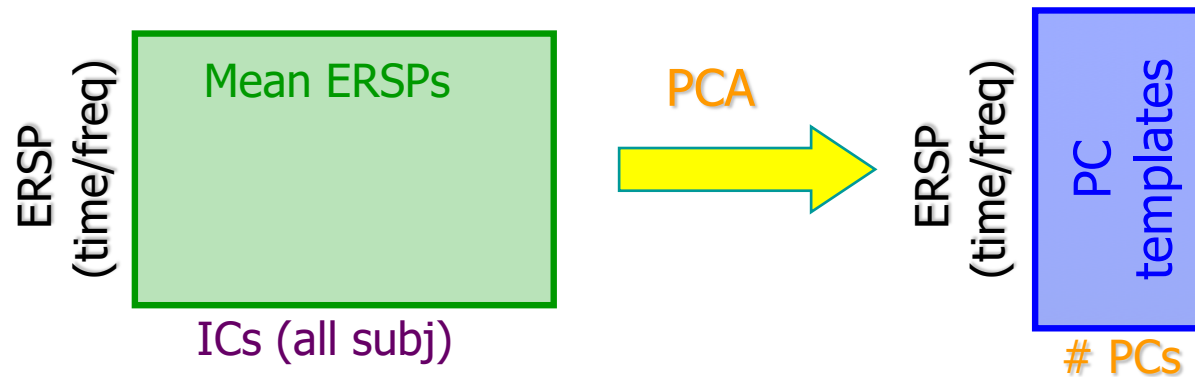
# Precluster schematic



# Precluster schematic

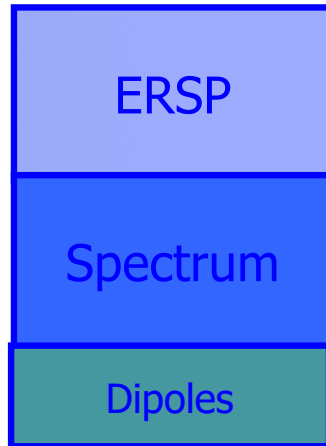


# Precluster: Use singular values from PCA

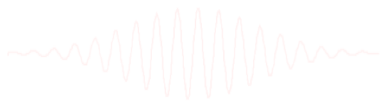
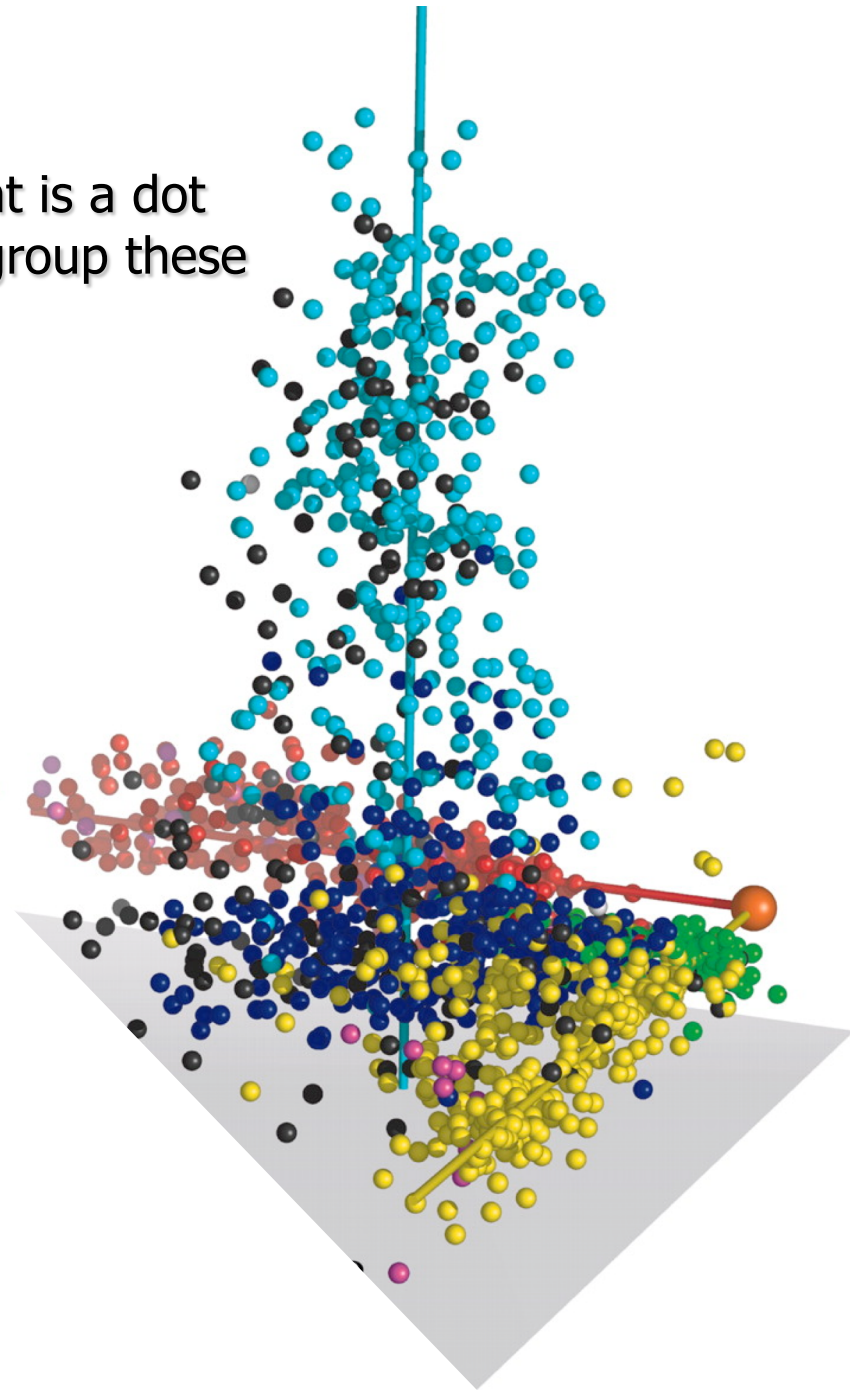


# Precluster schematic

ICs (all subj)



Each component is a dot  
Clustering will group these  
dots



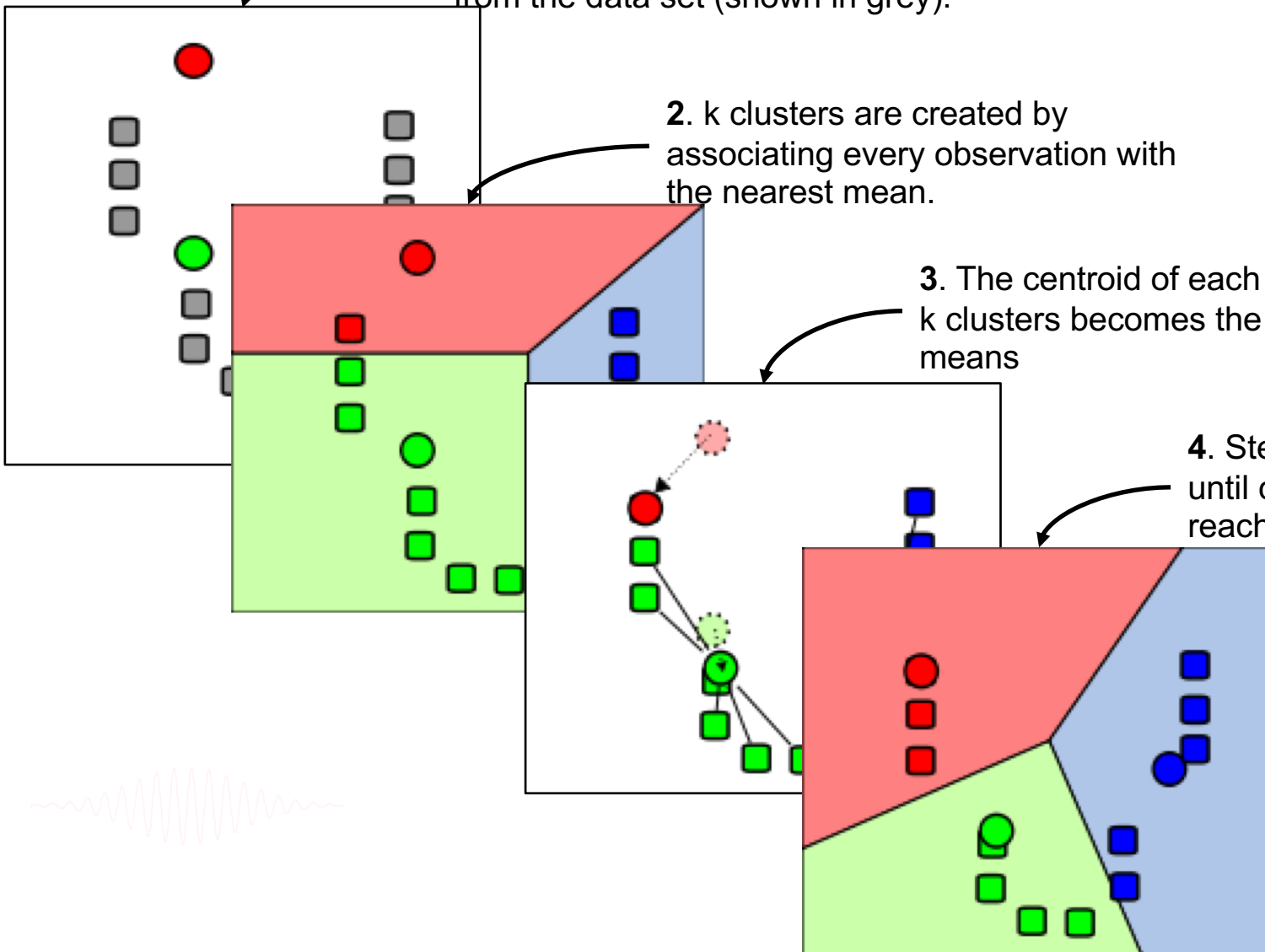
# Classical KMean

1.  $k$  initial "means" (in this case  $k=3$ , (shown in color)) are randomly selected from the data set (shown in grey).

2.  $k$  clusters are created by associating every observation with the nearest mean.

3. The centroid of each of the  $k$  clusters becomes the new means

4. Steps 2 and 3 are repeated until convergence has been reached.



# Cluster components

EEGLAB v15.x (dev)

File Edit Tools Plot **Study** Datasets Help

**STUDY set: Sternbe**

- Edit study info
- Select/Edit study design(s)
- Precompute channel measures
- Plot channel measures
- Precompute component measures
- PCA clustering (original)
- Edit/plot clusters**

Study filename:	
Study task name	
Nb of subjects	
Nb of conditions	
Nb of sessions	
Nb of groups	
Epoch consistency	yes
Channels per frame	69,70,71
Channel locations	yes
Clusters	7
Status	Pre-clustered
Total size (Mb)	229.3

Set clustering algorithm -- pop\_clust()

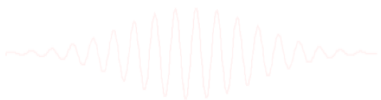
Performing clustering on cluster 'ParentCluster 1'

Clustering algorithm: Kmeans (stat. toolbox)

Number of clusters to compute: 24

Separate outliers (enter std.) 3

Help Cancel Ok



# Choosing data measures

What measure(s) should you use?

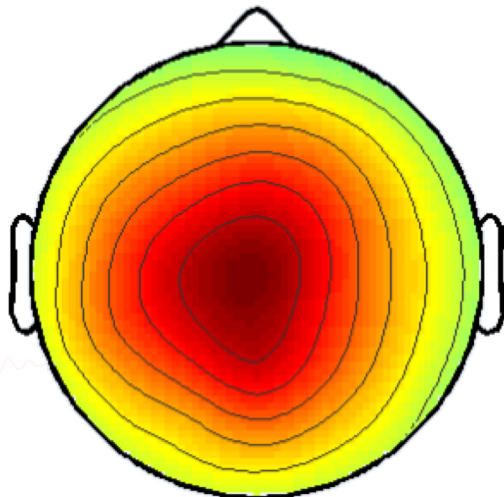
It depends on your final cluster criteria...

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

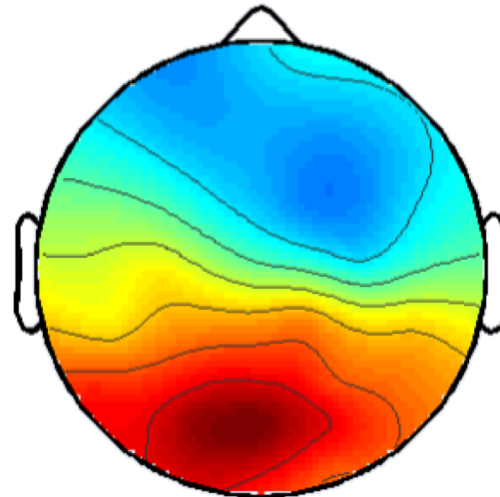
But consider:

- What is the difference between these two components?

IC2 / S02, Cls 26

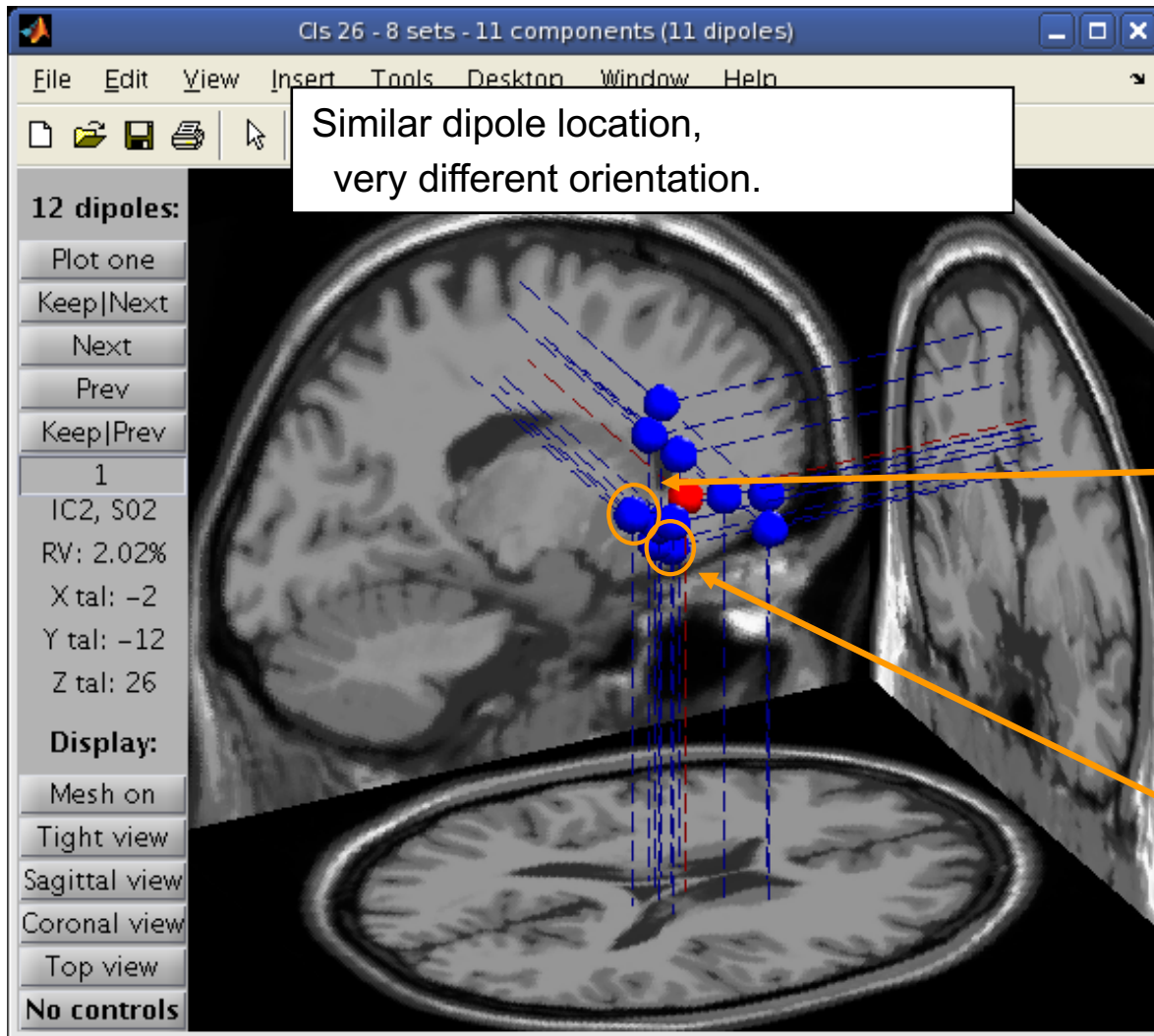


IC5 / S05, Cls 26



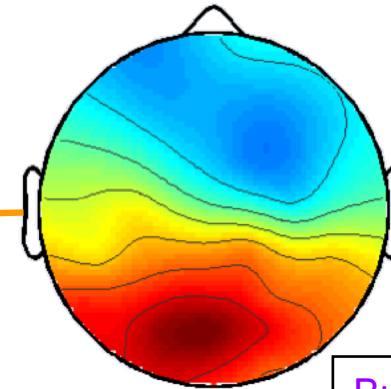


# Choosing data measures

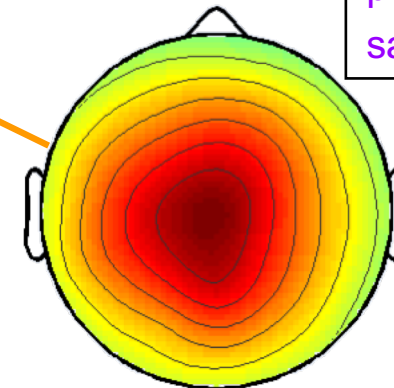


Obvious dramatic effect on scalp map topography:

IC5 / S05, Cls 26

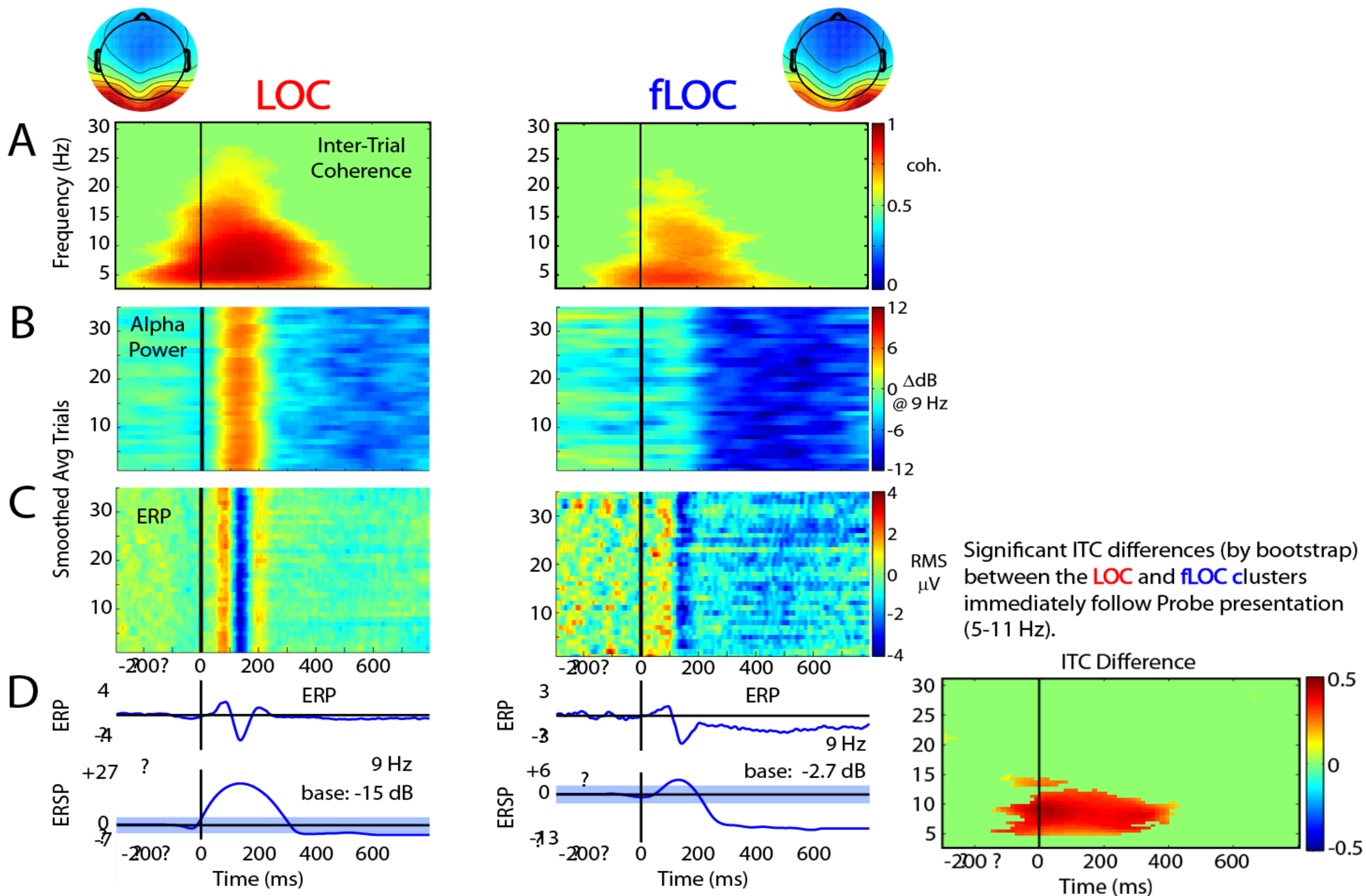


IC2 / S02, Cls 26



But, do they perform the same functions?

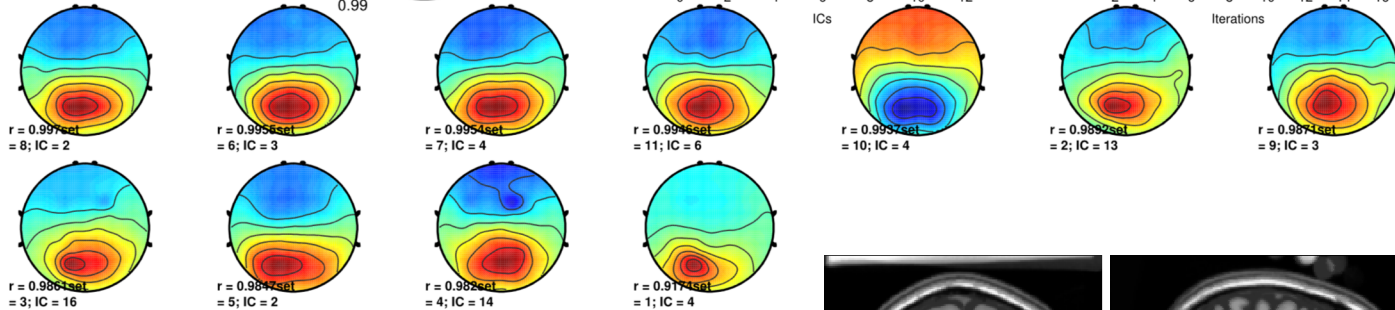
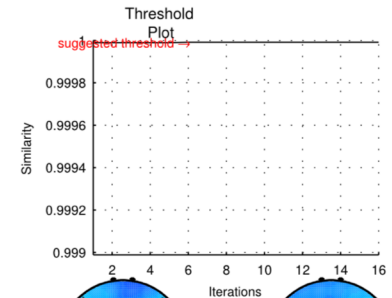
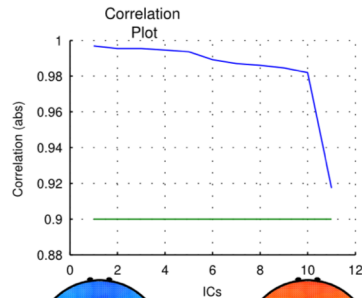
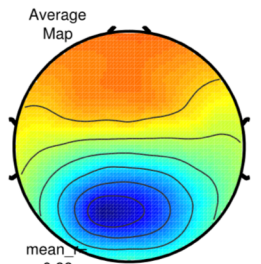
# Subject differences?



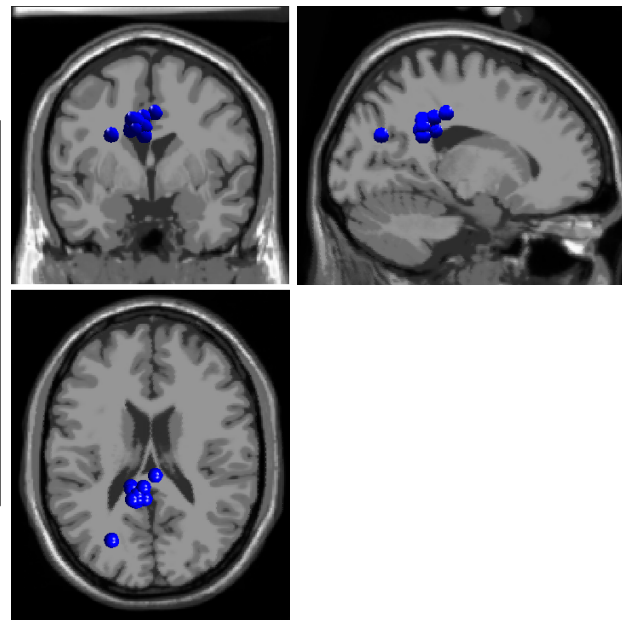
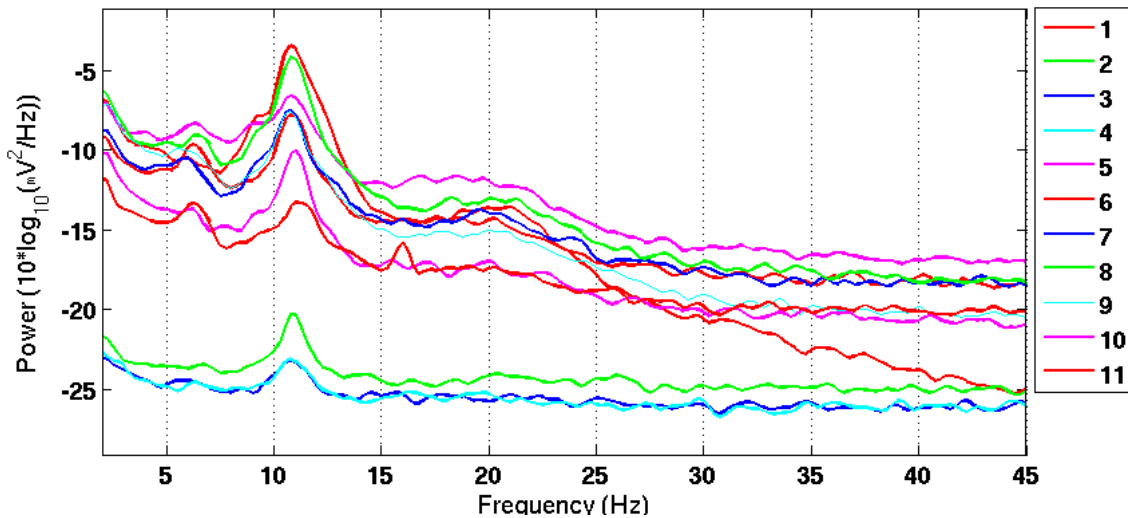
# Results (Cluster 1 within subject)

100 % Sessions contribute

INFO:  
 Template: CB Session 7 PREPROC:STEP 2; Set 7; IC 3;  
 Number of datasets: 11  
 Correlation threshold: 0.9 (green line)  
 Max ICs from each dataset: 1  
 Cluster: 11 ICs from 11 sets  
 All datasets contribute.  
 Similarity = 1.0000



Cls 3 Spectrum



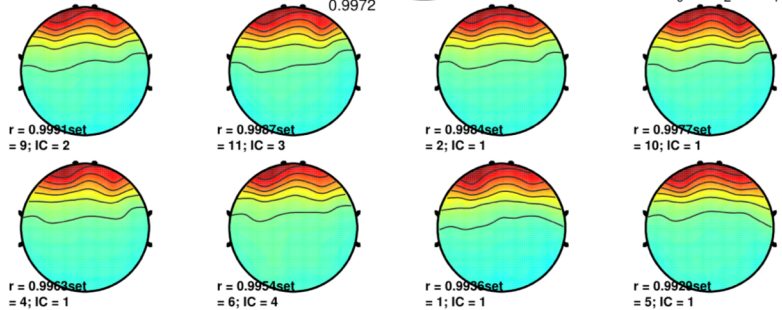
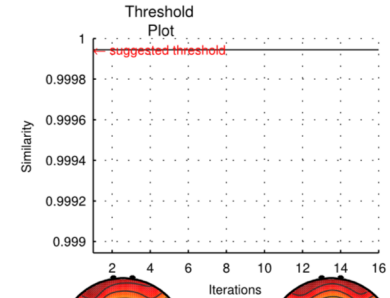
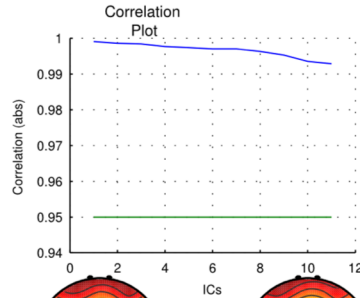
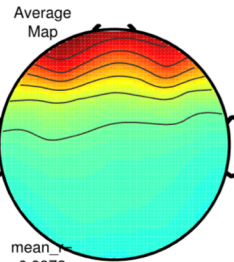
# Results (Cluster 2 within subject)

100 % Sessions contribute

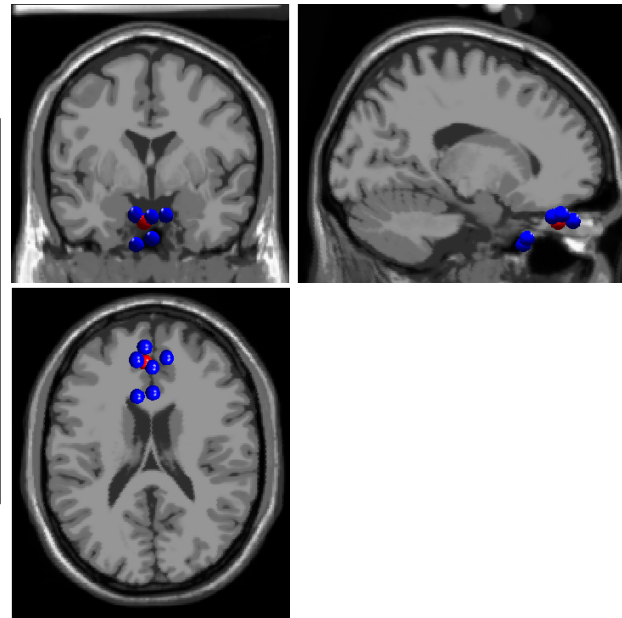
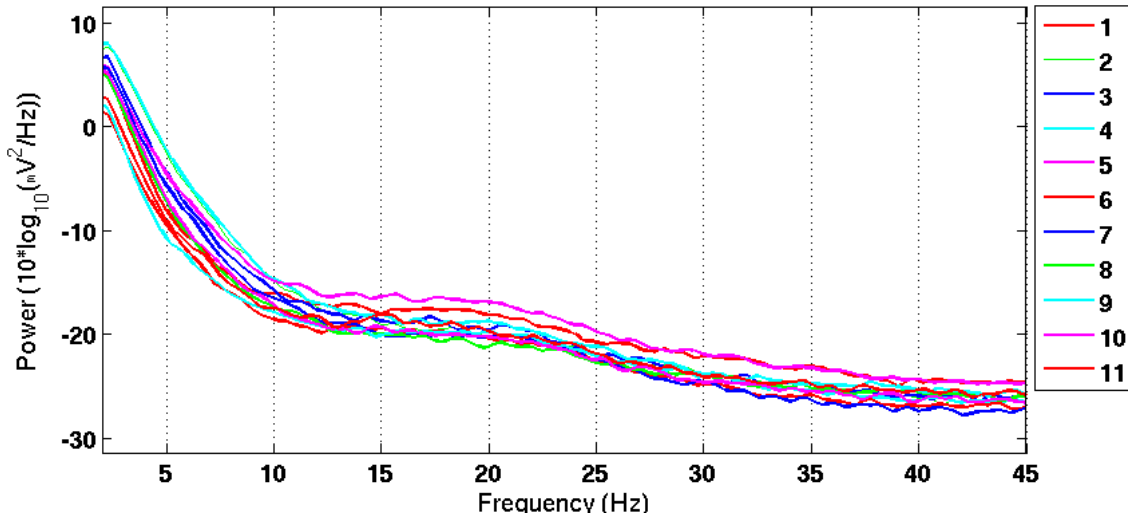
INFO:

Template: CB Session 5 PREPROC:STEP 2; Set 5; IC 1;  
 Number of datasets: 11  
 Correlation threshold: 0.95 (green line)  
 Max ICs from each dataset: 1  
 Cluster: 11 ICs from 11 sets  
 All datasets contribute.

Similarity = 0.9999



Cls 4 Spectrum





# Results (Cluster 8 within subject)

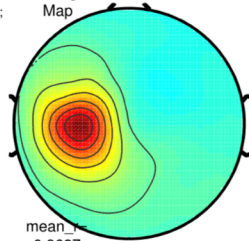
100 % Sessions contribute

INFO:

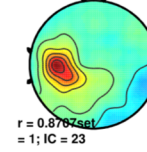
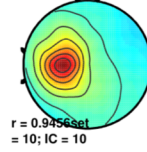
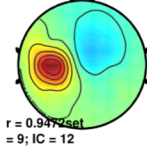
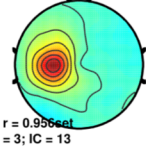
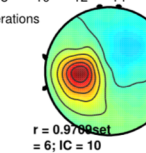
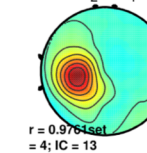
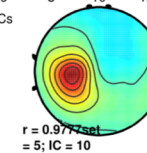
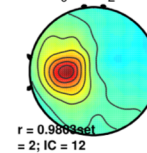
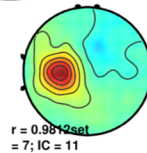
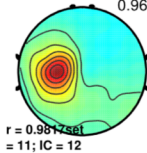
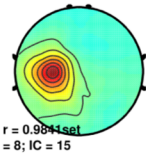
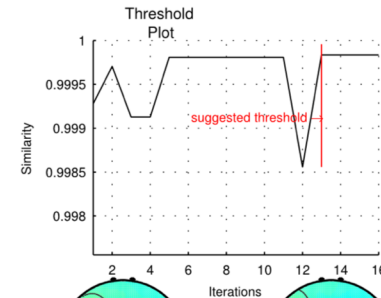
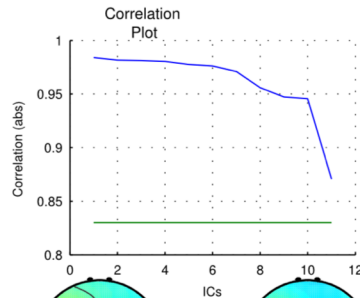
Template: CB Session 7 PREPROC:STEP 2; Set 7; IC 11;  
 Number of datasets: 11  
 Correlation threshold: 0.83 (green line)  
 Max ICs from each dataset: 1  
 Cluster: 11 ICs from 11 sets  
 All datasets contribute.

Similarity = 0.9998

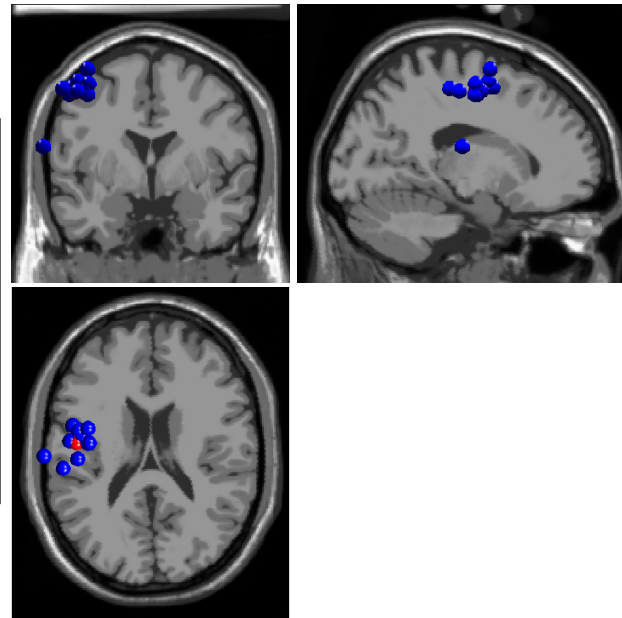
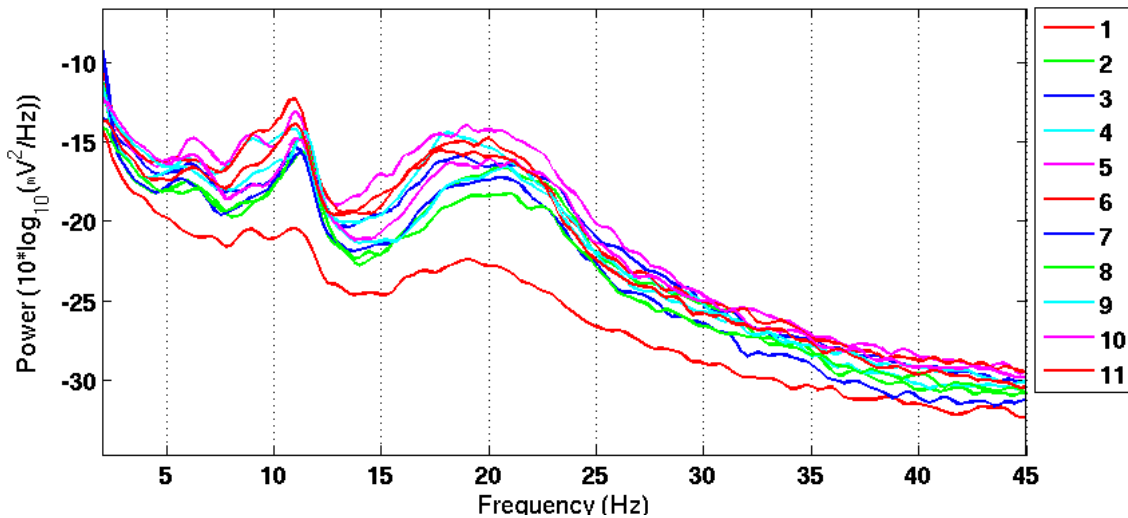
Average  
Map



mean  $r = 0.9687$



Cls 8 Spectrum



# Results (Cluster 13 within subject)

63.64% Sessions contribute

INFO:

Template: CB Session 2 PREPROC:STEP 2; Set 2; IC 24;

Number of datasets: 11

Correlation threshold: 0.8 (green line)

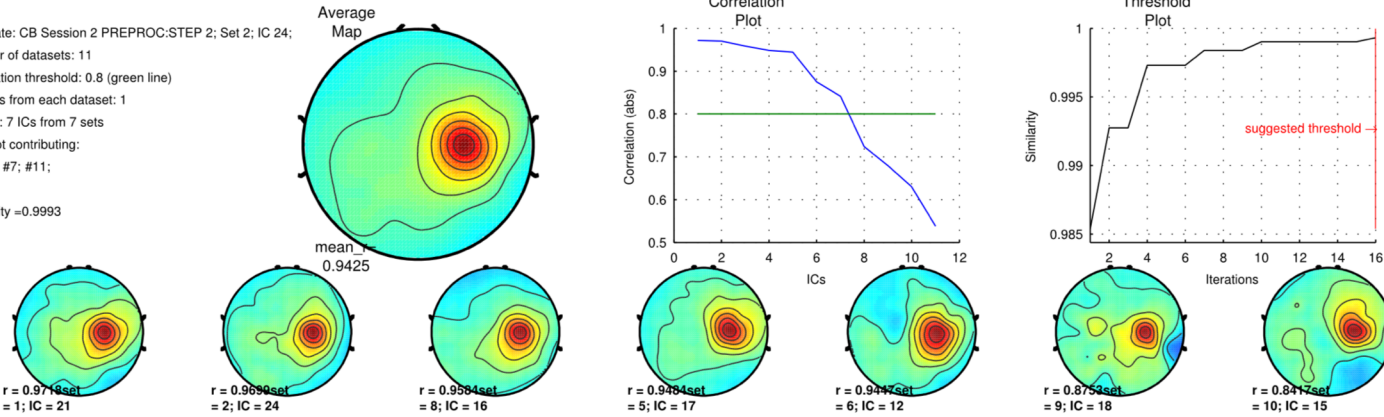
Max ICs from each dataset: 1

Cluster: 7 ICs from 7 sets

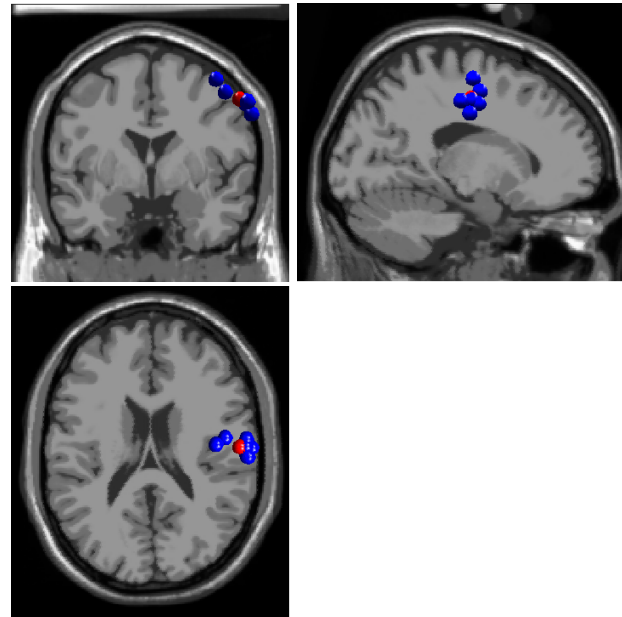
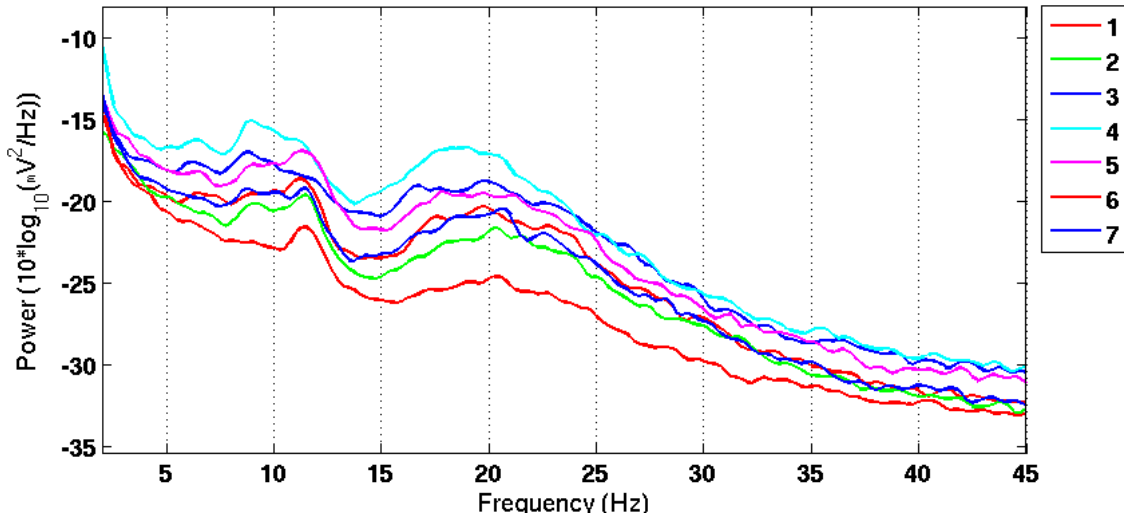
Sets not contributing:

#3; #4; #7; #11;

Similarity = 0.9993



Cls 13 Spectrum



# Results (Cluster 14 within subject)

36.36% Sessions contribute

INFO:

Template: CB Session 1 PREPROC:STEP 2; Set 1; IC 5;

Number of datasets: 11

Correlation threshold: 0.93 (green line)

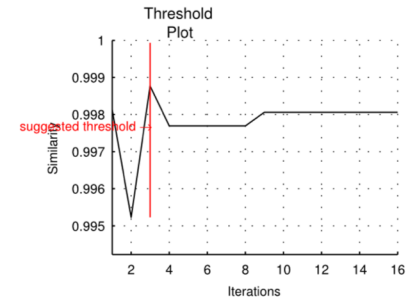
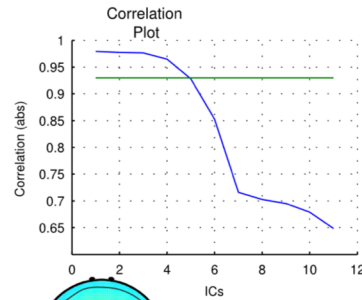
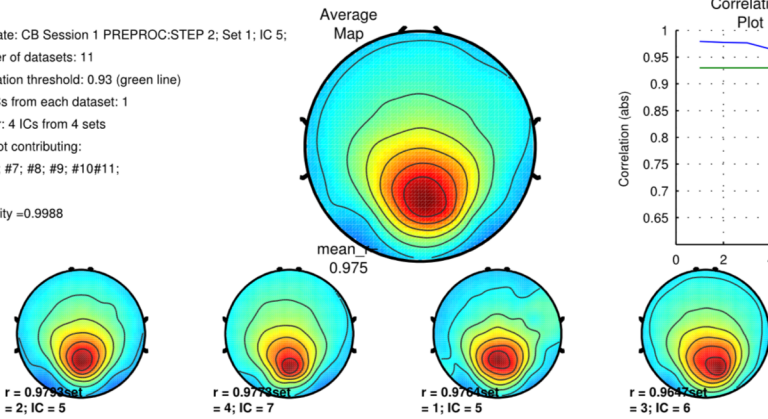
Max ICs from each dataset: 1

Cluster: 4 ICs from 4 sets

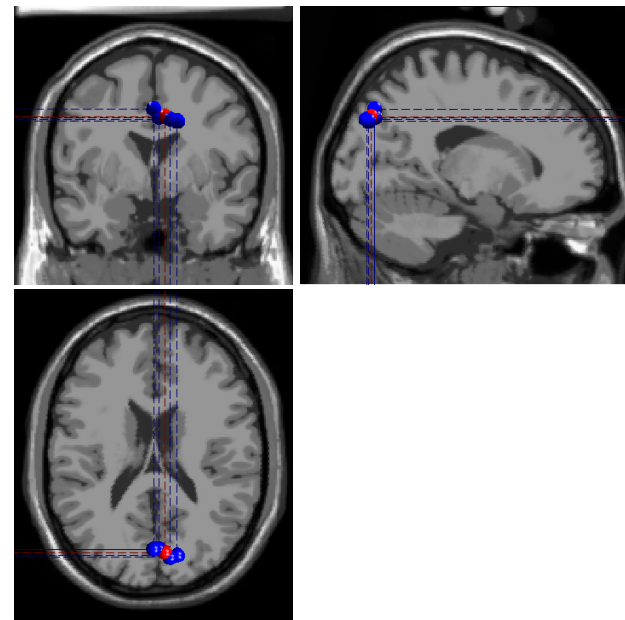
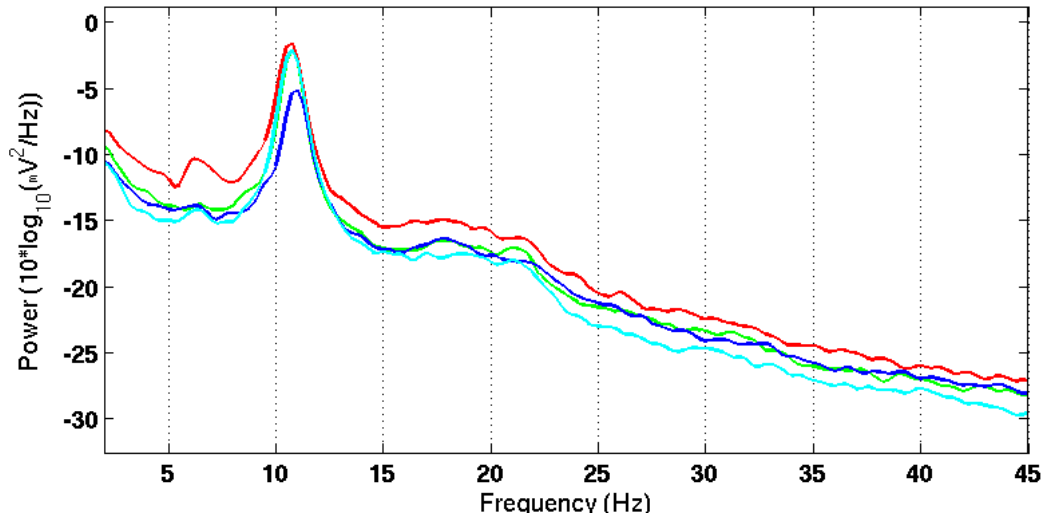
Sets not contributing:

#5; #6; #7; #8; #9; #10#11;

Similarity = 0.9988

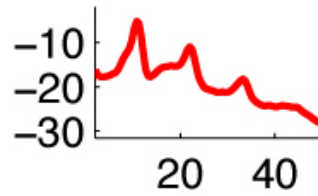
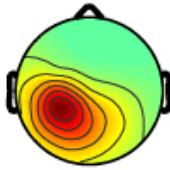


Cls 14 Spectrum

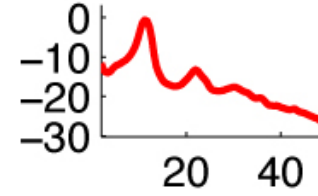
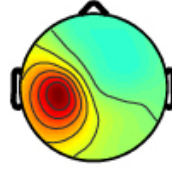


# Left $\mu$ cluster (across subjects)

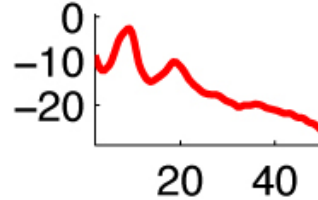
S2 IC47



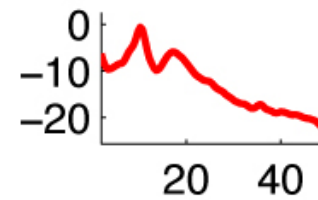
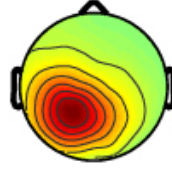
S3 IC47



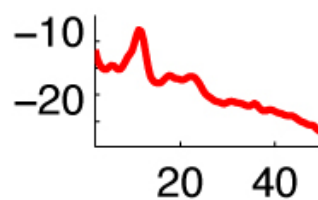
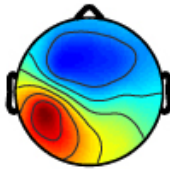
S4 IC37



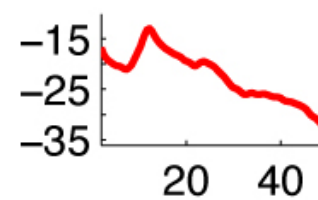
S5 IC48



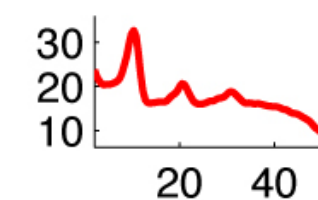
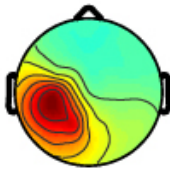
S6 IC46



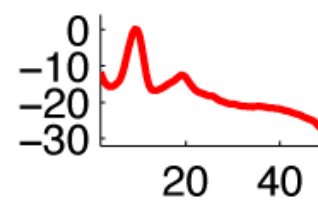
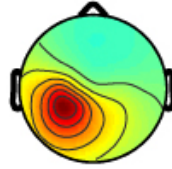
S7 IC35



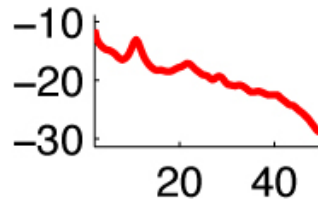
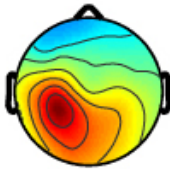
S9 IC7



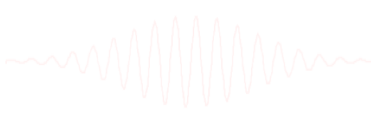
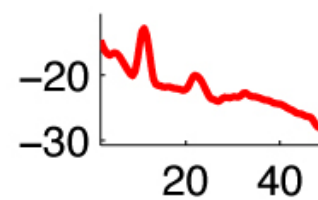
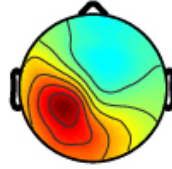
S11 IC45



S12 IC45



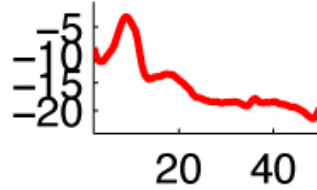
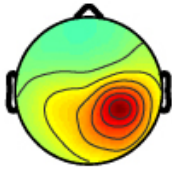
S14 IC45



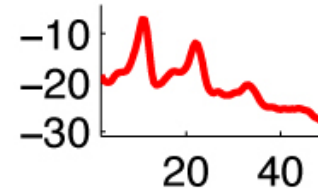
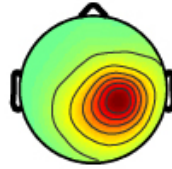


# Right $\mu$ cluster

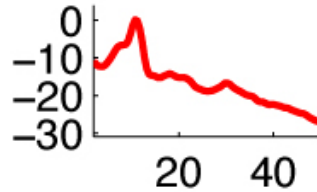
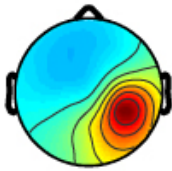
S1 IC51



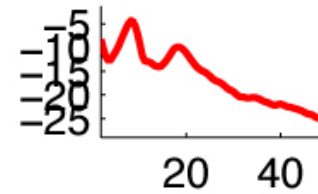
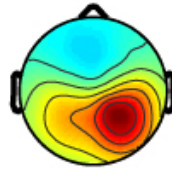
S2 IC41



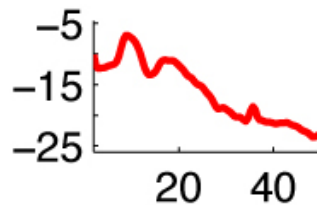
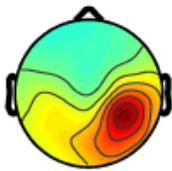
S3 IC41



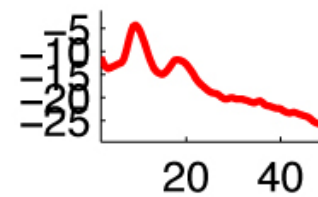
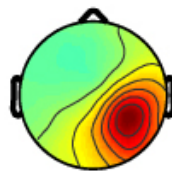
S4 IC50



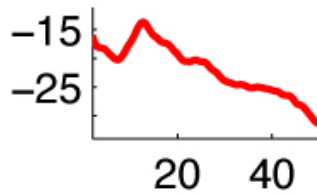
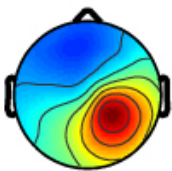
S5 IC51



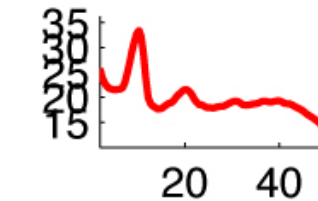
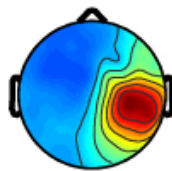
S6 IC6<sup>0</sup>



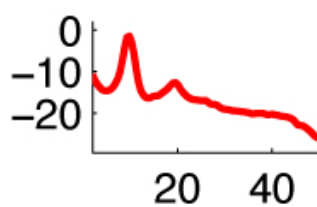
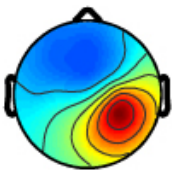
S7 IC48



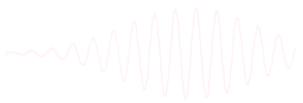
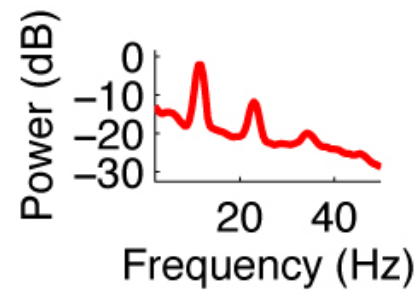
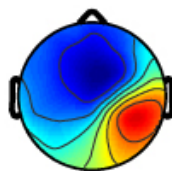
S9 IC39



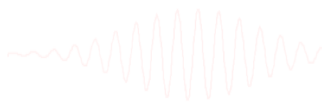
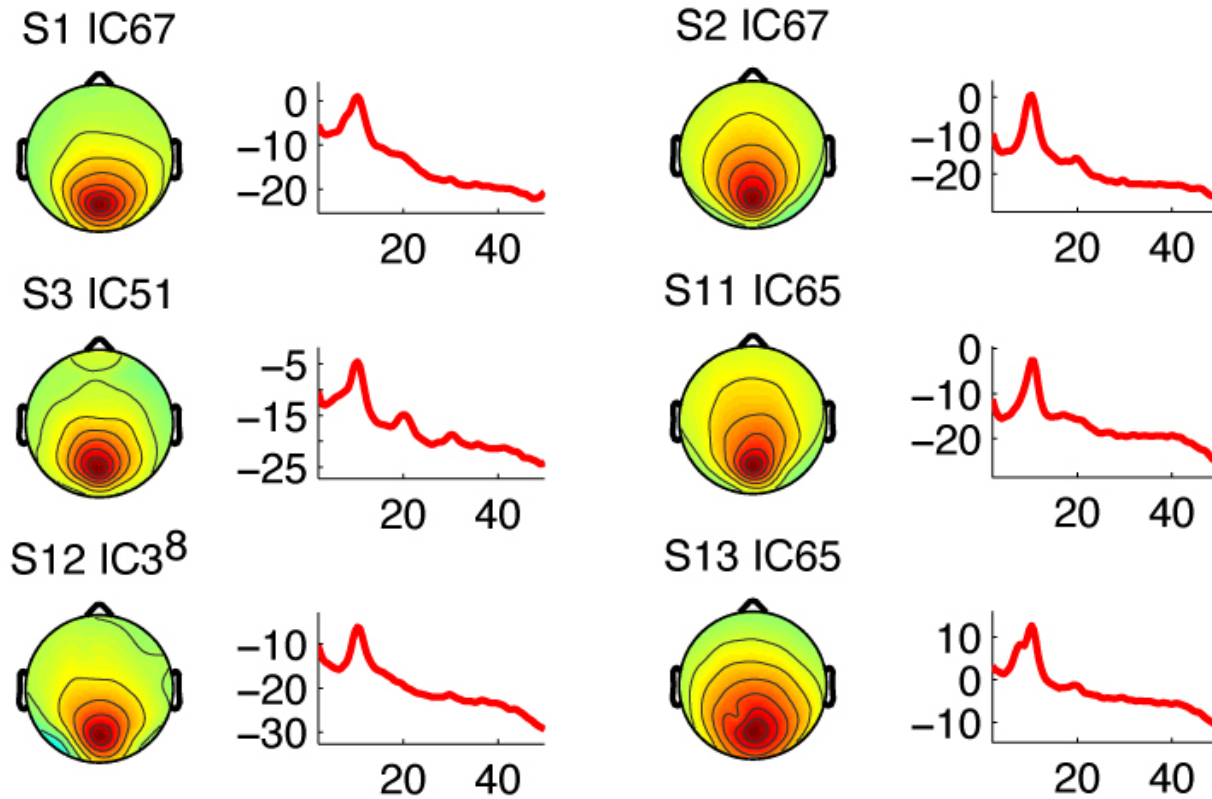
S11 IC49



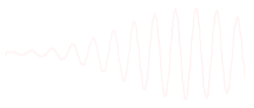
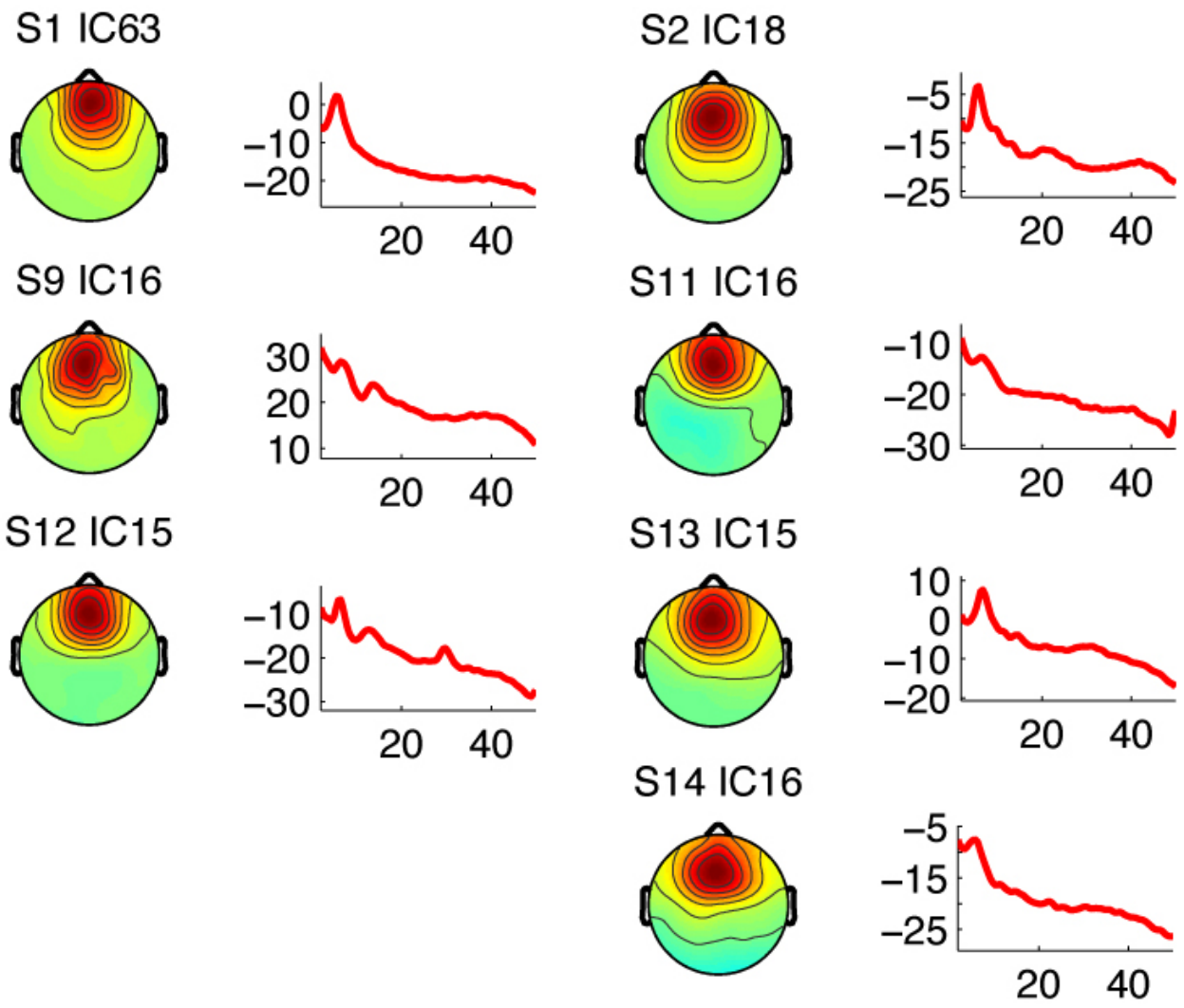
S14 IC49



# Occipital $\alpha$ cluster



# Frontal Midline $\theta$ cluster

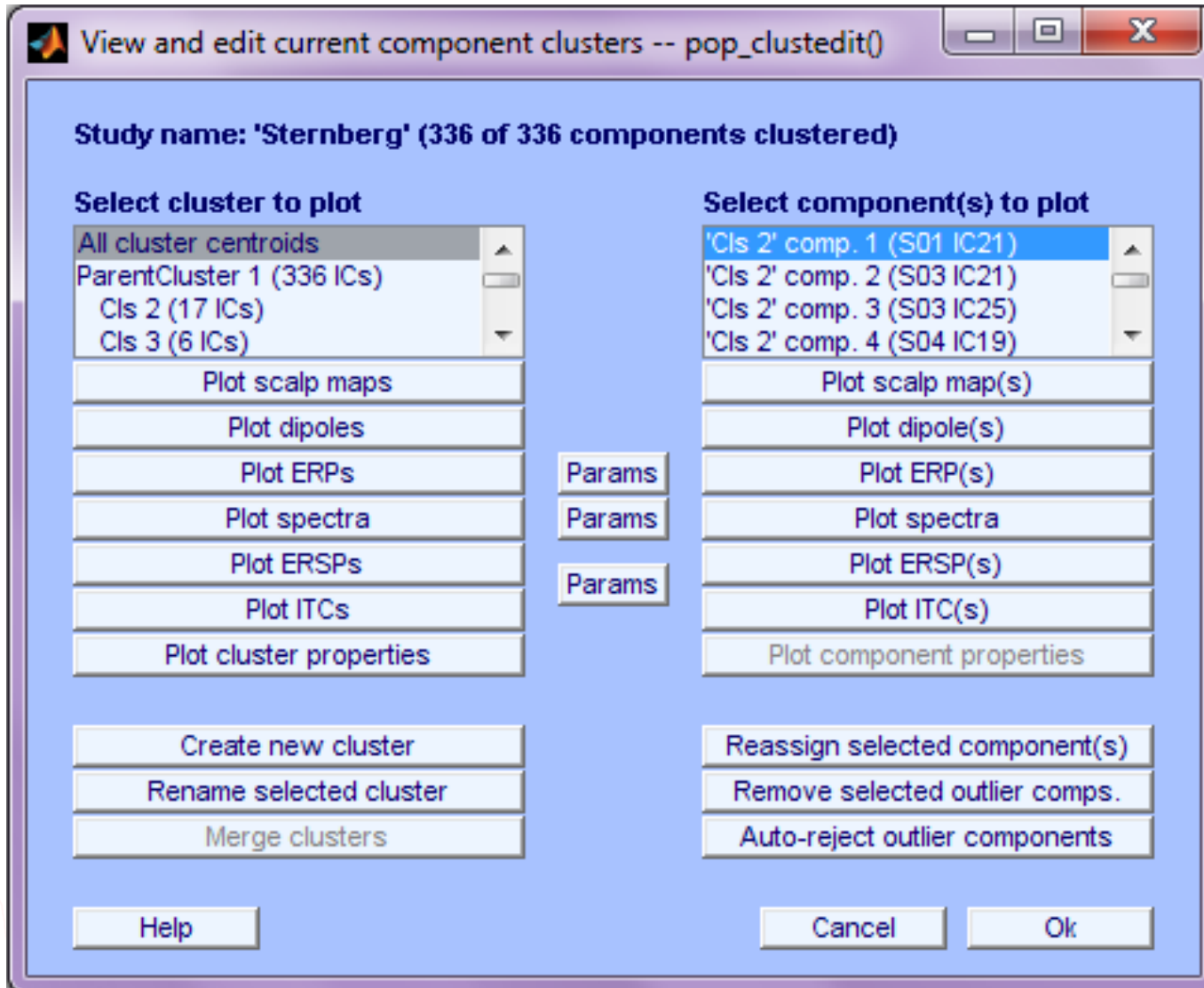


# View and edit clusters

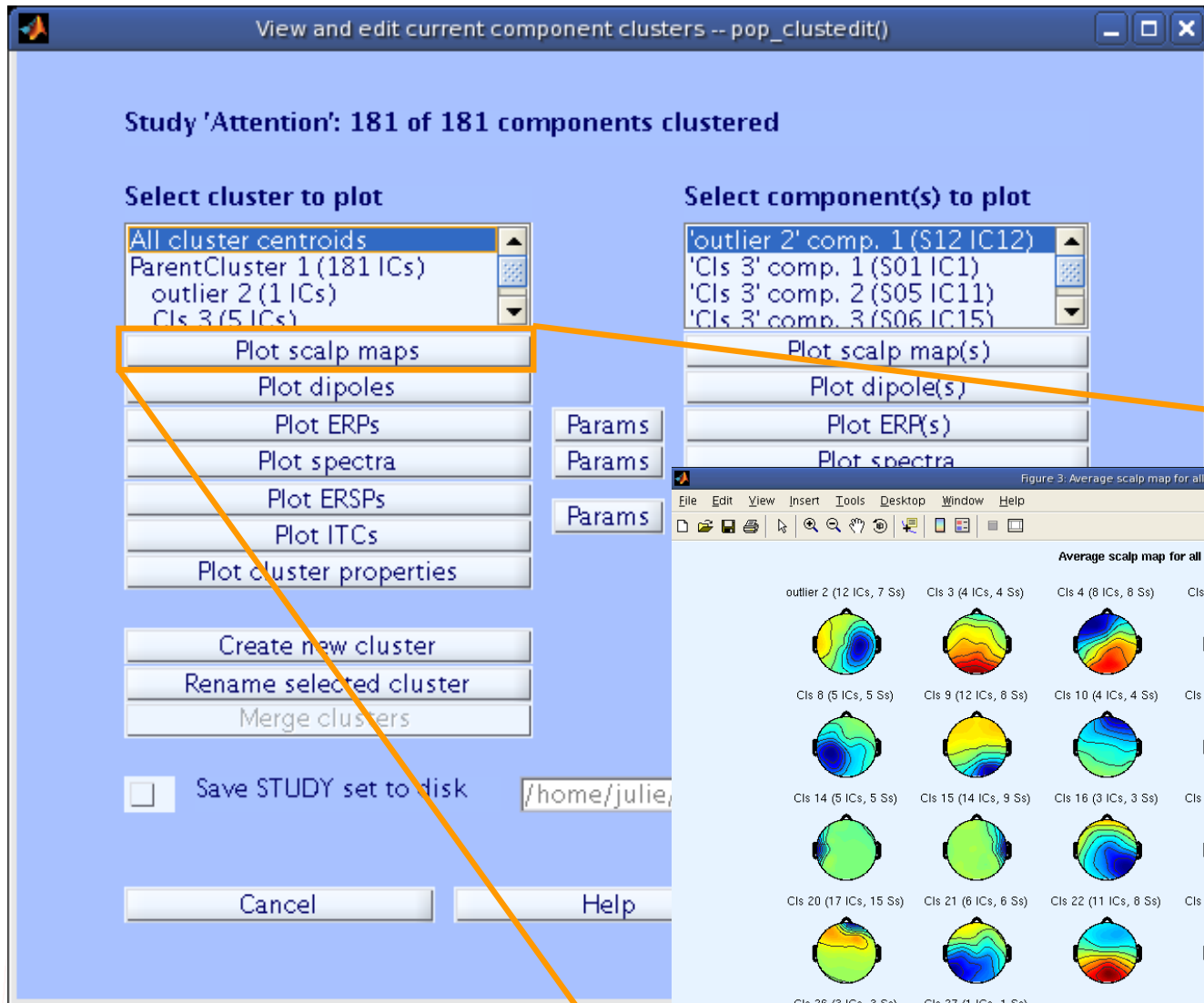
The screenshot shows the EEGLAB v15.x (dev) interface. The 'Study' menu is open, displaying options for editing study information, precomputing channel and component measures, and editing/plotting clusters. The 'Edit/plot clusters' option is highlighted. In the background, a window titled 'STUDY set: Sternberg' displays the following parameters:

Study filename:	
Study task name	
Nb of subjects	
Nb of conditions	
Nb of sessions	
Nb of groups	
Epoch consistency	yes
Channels per frame	69.70.71
Channel locations	yes
Clusters	7
Status	Pre-clustered
Total size (Mb)	229.3

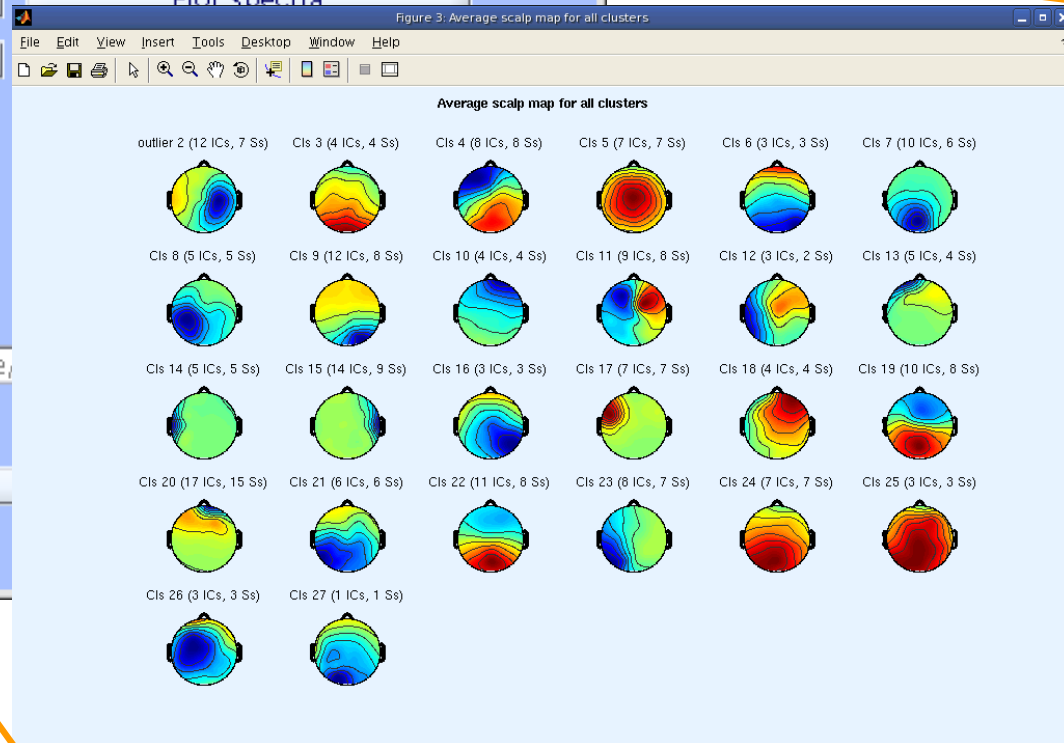
# Plot/edit clusters



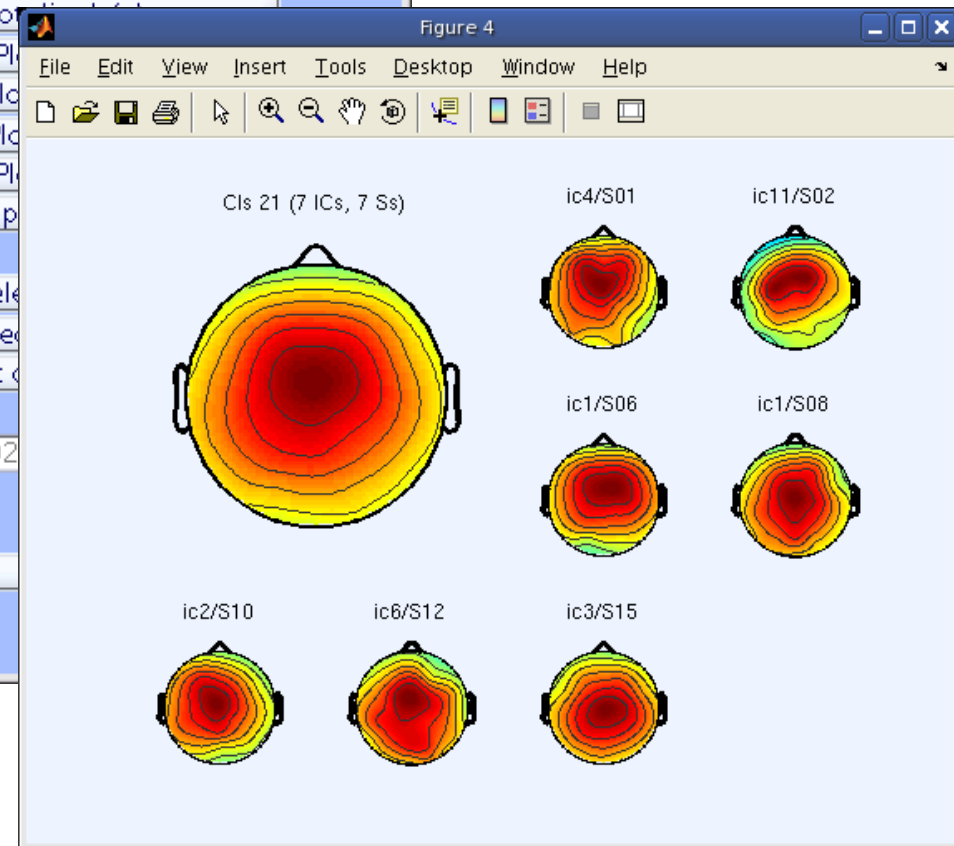
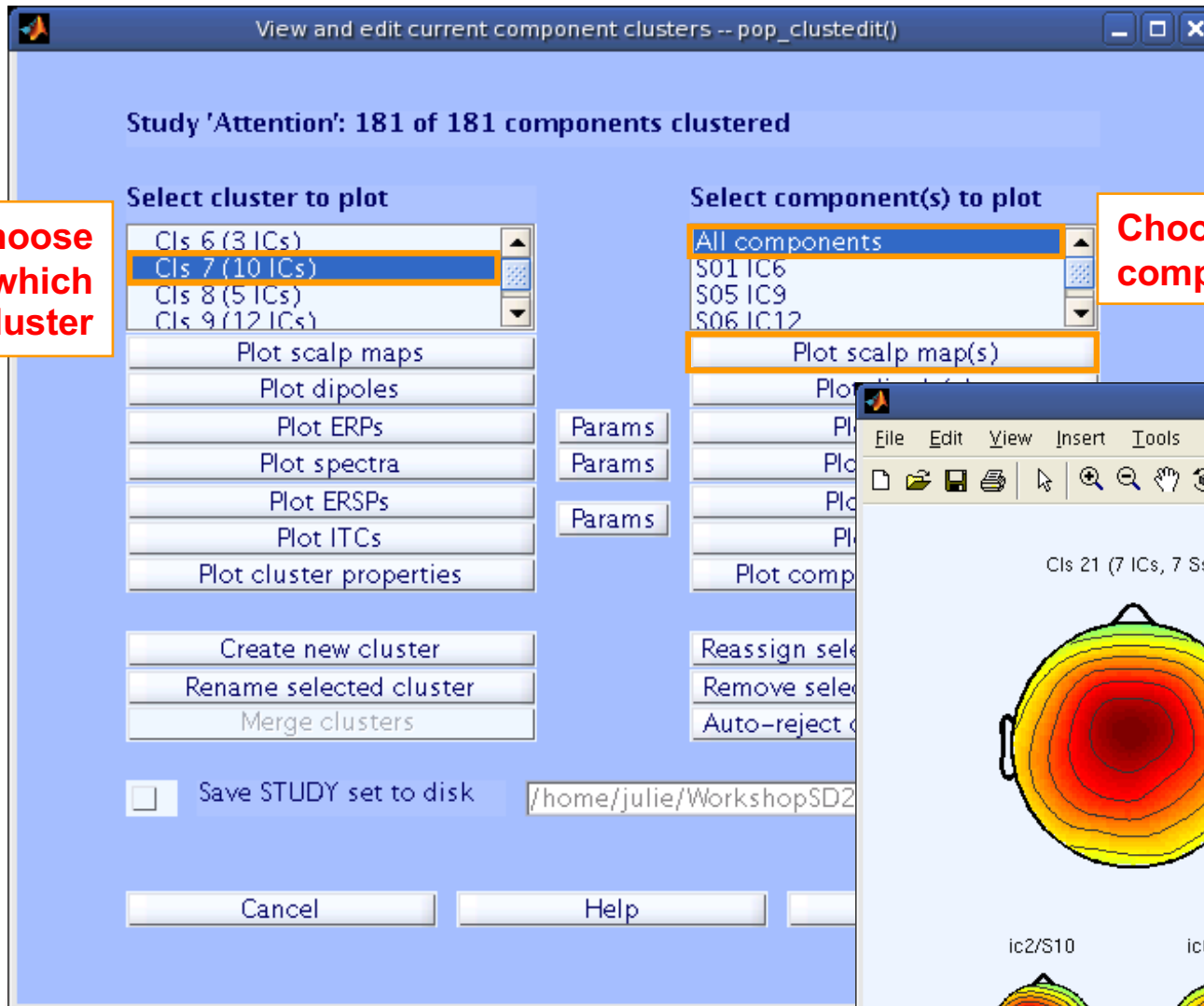
# Plot cluster data



Plot mean scalp maps for easy reference



# Plot cluster data



# Plot cluster data

Cls 3 - 9 sets - 15 components (15 dipoles)

File Edit View Insert Tools Desktop Window Help

16 dipoles:  
Plot one  
Keep|Next  
Next  
Prev  
Keep|Prev  
1  
S01, IC6  
RV: 2.92%  
X tal: -39  
Y tal: -10  
Z tal: -1  
Display:  
Mesh on  
Tight view  
Sagittal view  
Coronal view  
Top view  
No controls

edit current component clusters -- pop\_clustedit()

memorize vs ignore

Select component to plot

- All components
- S01 IC6
- S01 IC8
- S02 IC8
- S02 IC23
- S02 IC40

Plot scalp map(s)

Plot dipole(s)

Plot ERP(s)

Plot spectra

Plot ERPimage(s)

Plot ERPimage(s)

Plot ERSP(s)

Plot ITC(s)

Plot component properties

Create new cluster

Rename selected cluster

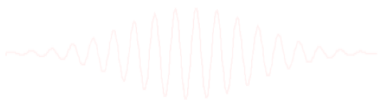
Merge clusters

Reassign selected component(s)

Remove selected outlier comps.

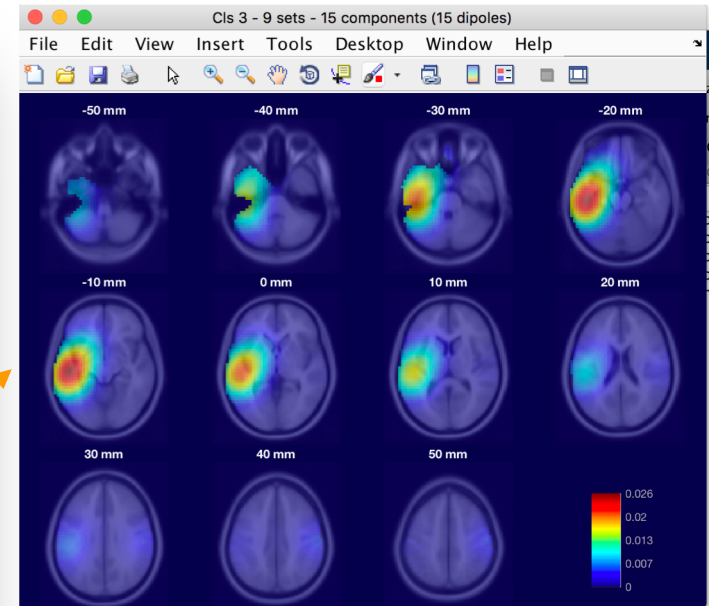
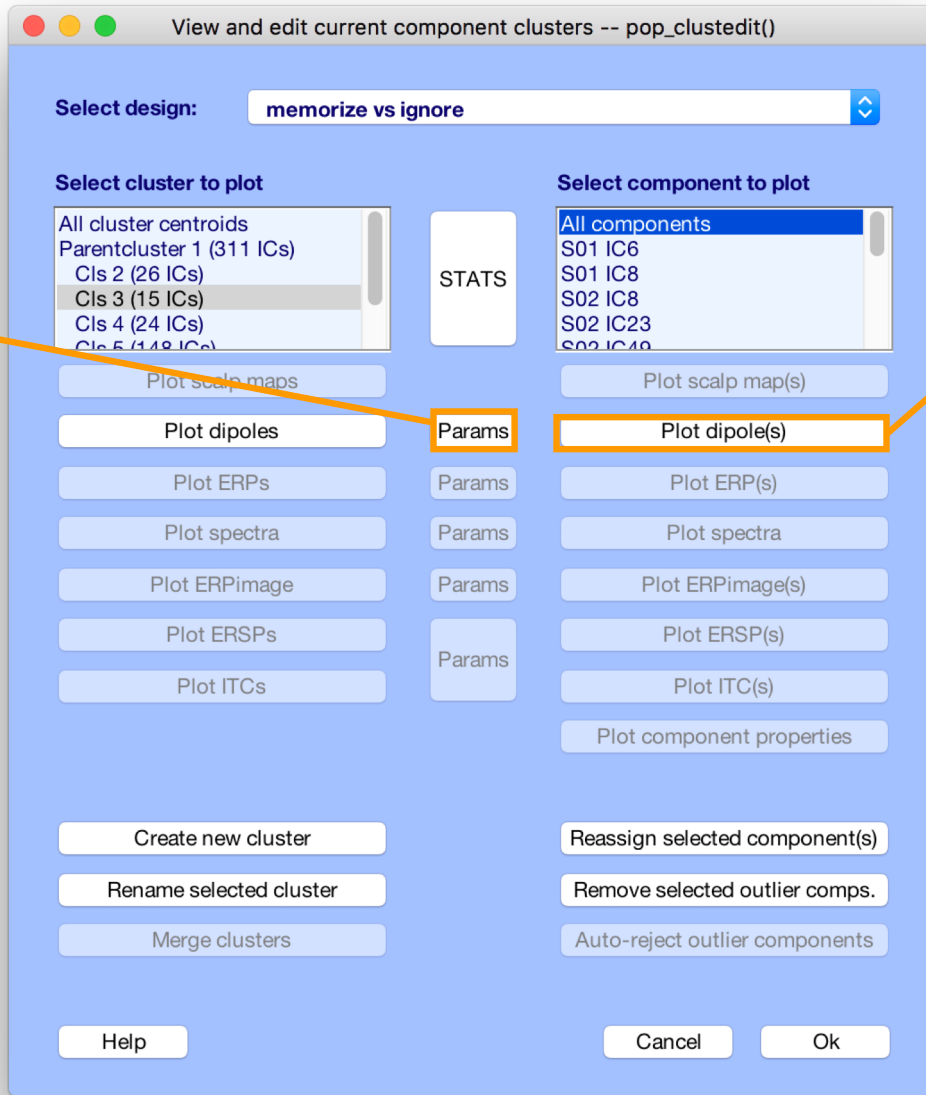
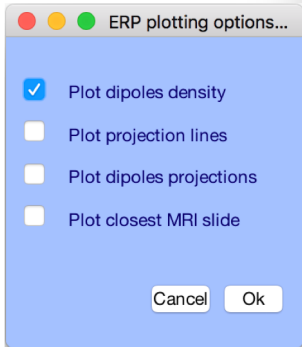
Auto-reject outlier components

Help Cancel Ok

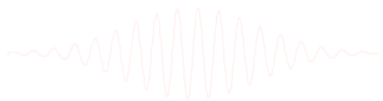
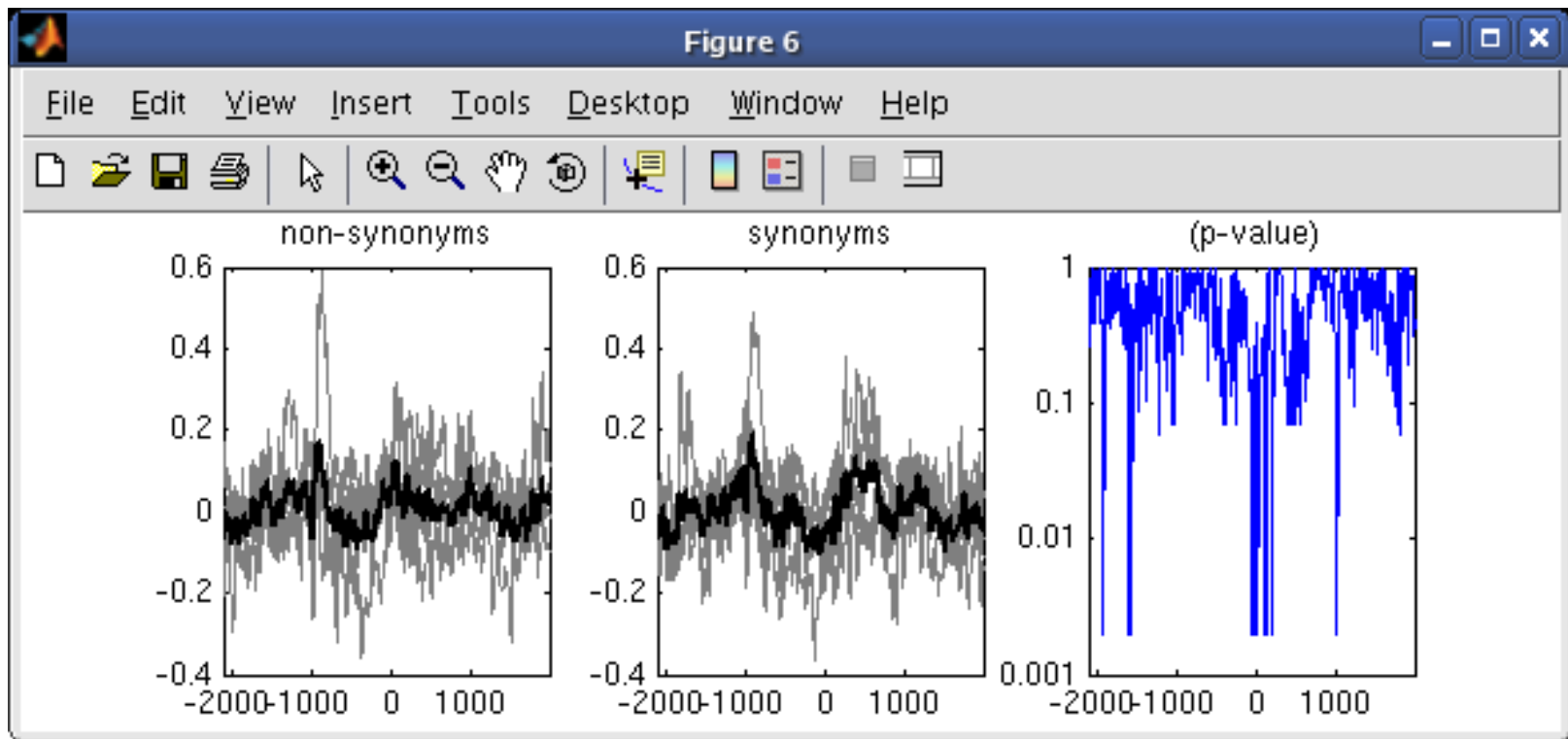




# Plot cluster data



# Plot cluster ERP



# Exercise

- Load the STUDY stern.study
- Precompute **spectrum**, **ERP** and **scalp maps** for components
- Precluster and cluster components using **dipole locations** and **dipole moments** (affinity clustering)
- Look at your cluster. Identify frontal midline theta cluster and occipital alpha cluster
- Remove outliers if any
- Plot significant difference (parametric statistics) for one component cluster spectrum between the two conditions ignore vs memorize