Clustering of ICA components

Arnaud Delorme

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

Steps of clustering

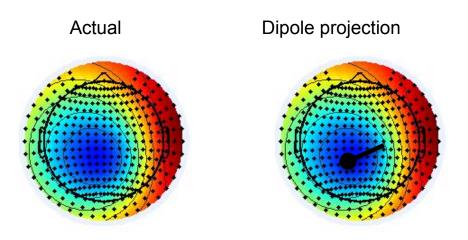
- Select ICA components for clustering
- Precompute measures of interest
- Cluster measures
- Plot clusters and edit them if necessary



Edit dataset info

				•		pop_study	/(): Pre-select compone	ents		
							al (topo map – dipole any existing compon			
							15			
					🖌 Keep on l	y in-brain d	lipoles.			
Cre	ate a new STUDY set pop	_study()			Can	el	Help		Ok	
									1	
Edit	STUDY set information - rer	nember to	save change	es	Chanada ana		T			
	STUDY set name:				Sternberg					
	STUDY set task name:				Sternberg					
	STUDY set notes:									
	dataset filename	browse		session	condition	group	Select by r.v.			
1	C:\\Users\\julie\\Documents\\\		S01		memorize		Comp.: 3 5	Clear		
2	C:\\Users\\julie\\Documents\\\	Nor	S01		ignore		Comp.: 3 5	Clear		
3	C:\\Users\\julie\\Documents\\\	Nor	S01		probe		Comp.: 3 5	Clear		
4	C:\\Users\\julie\\Documents\\\	Nor	S02		memorize		Comp.: 5 6	Clear		
5	C:\\Users\\julie\\Documents\\\	Nor	S02		ignore		Comp.: 5 6	Clear		
6	C:\\Users\\julie\\Documents\\\	Nor	S02		probe		Comp.: 5 6	Clear		
7	C:\\Users\\julie\\Documents\\\	Nor	S03		memorize		Comp.: 6 7	Clear		
8	C:\\Users\\julie\\Documents\\\	Nor	S03		ignore		Comp.: 6 7	Clear		
9	C:\\Users\\julie\\Documents\\\	Nor	S03		probe		Comp.: 6 7	Clear		
10	C:\\Users\\julie\\Documents\\\	Nor	S04		memorize		Comp.: 1 2	Clear		
Impo	rtant note: Removed datasets v	vill not be sa	wed before be	ing delete	d from EEGLAB n	emory				
			< F	age 1	>					
				-						
	Dataset info (condition, group	,) differs	from study info	o. [set] = 0	Overwrite dataset	info.				
	Delete cluster information (to a						stc.)			
			-	,						
	Help					C	ancel C	k		

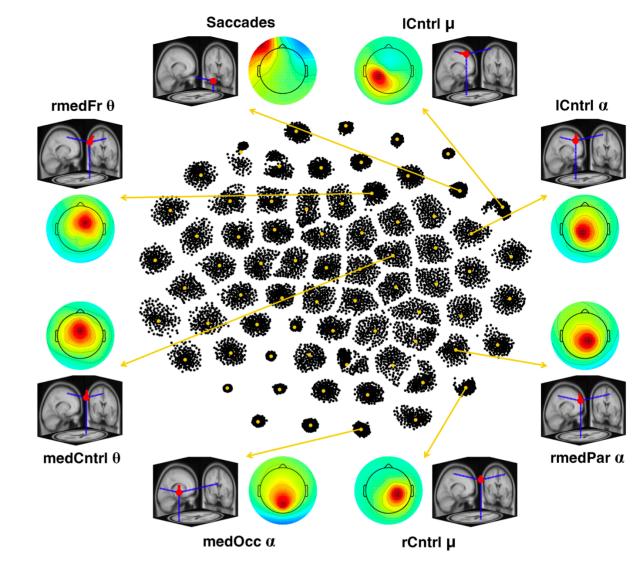
Computing residual variance (%)



$$r = \Sigma (x_i - \tilde{x}_i)^2 / \Sigma x_i^2$$



Clustering results example

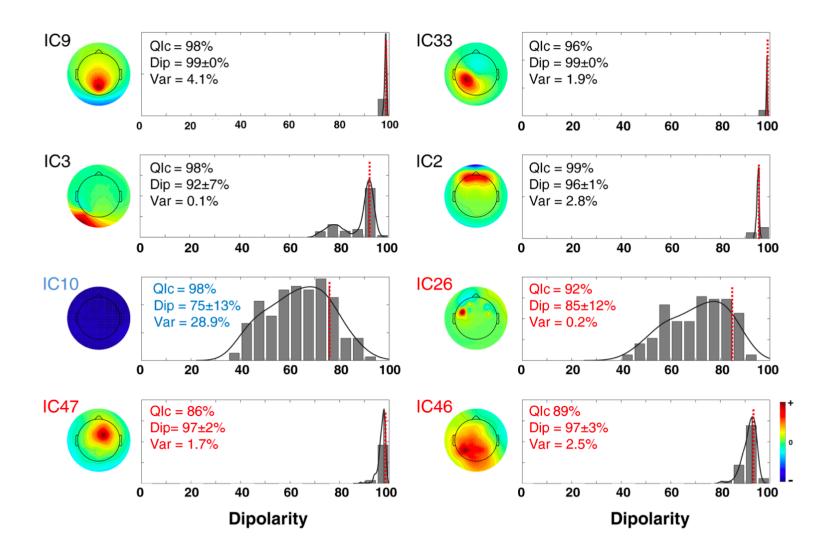


RELICA: A method for estimating the reliability of independent components

Fiorenzo Artoni^{a,*}, Danilo Menicucci^b, Arnaud Delorme^{c,e,f}, Scott Makeig^c, Silvestro Micera^{a,d}

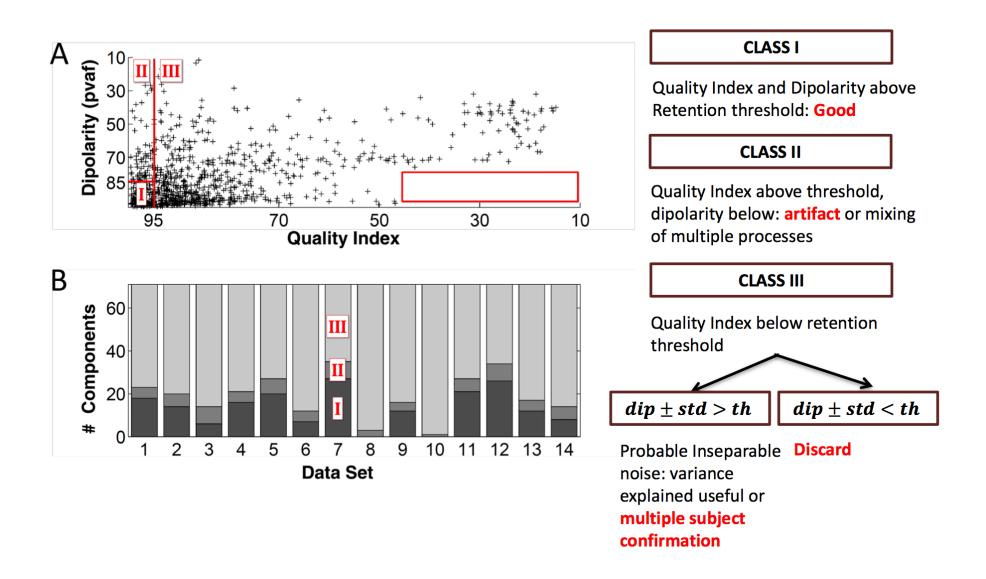
Within-cluster reliability

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



Reliability criteria and the rv<15%

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



ICs to cluster

	STUDY set name	Ņ	(_ _ X	_	Sternberg			
	STUDY set task r		select components		E	Sternberg			
	dataset filename		ic 19 ic 20		sion	condition	group	Select by r.v.	
1	C:\\Users\\julie\\		ic 21 ic 22			memorize		Comp.: 3 5	Clea
2	C:\\Users\\julie\\		ic 23			ignore		Comp.: 3 5	Clea
3	C:\\Users\\julie\\		ic 24			probe		Comp.: 3 5	Clea
4	C:\\Users\\julie\\		ic 25 ic 26			memorize		Comp.: 5 6	Clea
5	C:\\Users\\julie\\		ic 27		П	ignore		Comp.: 5 6	Clea
6	C:\\Users\\julie\\		ic 28 ic 29			probe		Comp.: 5 6	Clea
7	C:\\Users\\julie\\		ic 29		F.	memorize		Comp.: 6 7	Clea
8	C:\\Users\\julie\\		ic 31			ignore		Comp.: 6 7	Clea
9	C:\\Users\\julie\\		ic 32		F.	probe		Comp.: 6 7	Clea
10	C:\\Users\\julie\\	-	Cancel Ok		F.	memorize		Comp.: 1 2	Clea
Impor	Dataset info (conc	lition,	group,) differs from stud	Page	1 st] = O	>	info.		

~~~~

#### **Precompute data measures**

| -    |                    |          | EE    | GLAB v6. | 0b                |             |         |
|------|--------------------|----------|-------|----------|-------------------|-------------|---------|
| File | Edit               | Tools    | Plot  | Study    | Datasets          | Help        | с<br>Ч  |
|      | -STUD              | V se     | 4+ A4 | Edit     | t study info      |             |         |
|      | -3100              | 1 36     |       |          | compute ch        | annel measi | ures    |
|      | Study f            | ilenar   | ne:   | Plot     | channel me        | easures     |         |
|      | Study t            | ask na   | ame   | Pre      | compute co        | mponent me  | easures |
|      | Nb of s<br>Nb of c |          |       | Buil     | d precluste       | ring array  |         |
|      | Nbofs              |          |       | Clu      | ster compor       | nents       |         |
|      | Nb of g            |          |       |          | t/plot cluste     | rs          |         |
|      | Epoch<br>Chann     |          |       |          |                   |             |         |
|      | Chann              |          |       |          | 25                |             |         |
|      | Cluster            | rs       |       | 1        |                   |             |         |
|      | Status<br>Total s  | iza (M   | кY    |          | eady to pr<br>0.4 | ecluster    |         |
|      | TOTALS             | 128 (191 | 0)    | 5        | 0.4               |             |         |
|      |                    |          |       |          |                   |             |         |

## **Pre-compute measures**

| File | Edit     | Tools     | Plot   | Study  | Datasets                  | Help        |       |
|------|----------|-----------|--------|--------|---------------------------|-------------|-------|
|      | STUD     | Y set:    |        |        | tudy info<br>t/Edit study | y design(s  | ;)    |
|      | Study fi | lename: . | s/data | Preco  | mpute cha                 | nnel mea    | sures |
|      | Study ta | sk name   |        |        | hannel me                 |             |       |
|      | Nb of su | bjects    |        |        |                           |             |       |
|      | Nb of co | nditions  |        |        | mpute con                 |             |       |
|      | Nb of se | ssions    |        | Meas   | ure Produc                | t clusterir | ng 🕨  |
|      | Nb of gr | oups      |        | PCA o  | lustering (               | original)   | •     |
|      | Epoch co | nsistency |        | Edit/  | plot cluster              | S           |       |
|      | Channels | per fram  | e      | 61     |                           |             |       |
|      | Channel  | locations |        | ves    |                           |             |       |
|      | Clusters |           |        | 1      |                           |             |       |
|      | Status   |           |        | Pre-cl | ustered                   |             |       |
|      | Total si | ze (Mb)   |        | 8.2    |                           |             |       |

#### ○ ○ ○ Select and compute component measures for later clustering -- pop\_precomp()

| Pre- | compute component measures for STUDY 'Sternberg' - 'STUDY.design 1'                              |  |
|------|--------------------------------------------------------------------------------------------------|--|
| ☑    | Compute ERP/spectrum/ERSP only for components selected by RV (set) or for all components (unset) |  |
| List | of measures to precompute                                                                        |  |
|      | ERPs deceline ((min mavilin me)                                                                  |  |
|      | Power se citrum Spectropo parameters 'specmode', 'fit' (Test)                                    |  |
|      | ERSPs                                                                                            |  |
|      | Time/freq, parameters                                                                            |  |
|      | Scalp maps                                                                                       |  |
|      |                                                                                                  |  |
|      | Save single-trial measures for single-trial statistics - requires disk space                     |  |
|      | Recompute even if present on disk                                                                |  |
|      |                                                                                                  |  |
|      | Help Cancel Ok                                                                                   |  |

| File | Edit                 | Tools                | Plot   | Study                                                                                                          | Datasets                  | Help       |       |  |  |  |
|------|----------------------|----------------------|--------|----------------------------------------------------------------------------------------------------------------|---------------------------|------------|-------|--|--|--|
|      | STUD                 | Y set:               | _      |                                                                                                                | tudy info<br>t/Edit study | y design(s | 5)    |  |  |  |
|      | Study fi<br>Study ta | lename: .<br>sk name | s/data |                                                                                                                | mpute cha<br>hannel me    |            | sures |  |  |  |
|      | Nb of se<br>Nb of gr | nditions<br>ssions   |        | Precompute component measures<br>Measure Product clustering<br>PCA clustering (original)<br>Edit/plot clusters |                           |            |       |  |  |  |
|      | Channels             | per fram             | e      | 61                                                                                                             |                           |            |       |  |  |  |
|      | Channel              | locations            |        | yes                                                                                                            |                           |            |       |  |  |  |
|      | Clusters             |                      |        | 1                                                                                                              |                           |            |       |  |  |  |
|      | Status               |                      |        | Pre-cl                                                                                                         | ustered                   |            |       |  |  |  |
|      | Total si             | ze (Mb)              |        | 8.2                                                                                                            |                           |            |       |  |  |  |

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

Pre-compute channel measures for STUDY 'Sternberg' - 'STUDY.design 1'

| Cha<br>✓ |                                                  | n of missing channels (p<br>al components pre-tagg |                     | ParentCluster 1             | <br>w)   |
|----------|--------------------------------------------------|----------------------------------------------------|---------------------|-----------------------------|----------|
|          | Remove artifactual IC                            | A cluster or clusters (ho                          | ld shift key)       | Cls 2<br>Cls 3<br>Cls 4     | <u> </u> |
| List     | of measures to prece                             | ompute                                             |                     |                             |          |
|          | ERPs                                             | Baseline ([min m                                   | ax] in ms)          |                             |          |
|          | Power spectrum                                   | Spectopo parame                                    | eters               | 'specmode', 'fft'           | Test     |
|          | ERSPs                                            | Time/freq, parameter                               | s <sup>i</sup> cycl | es', [3 0.5], 'nfreqs', 100 | Test     |
| _        | Save single-trial measu<br>Recompute even if pre | ures for single-trial statis<br>sent on disk       | tics - requires     | disk space                  |          |
|          | Help                                             |                                                    |                     | Cancel                      | Ok       |

#### **Precompute data measures**

#### TIP: Compute all measures so you can

#### test different combinations for clustering

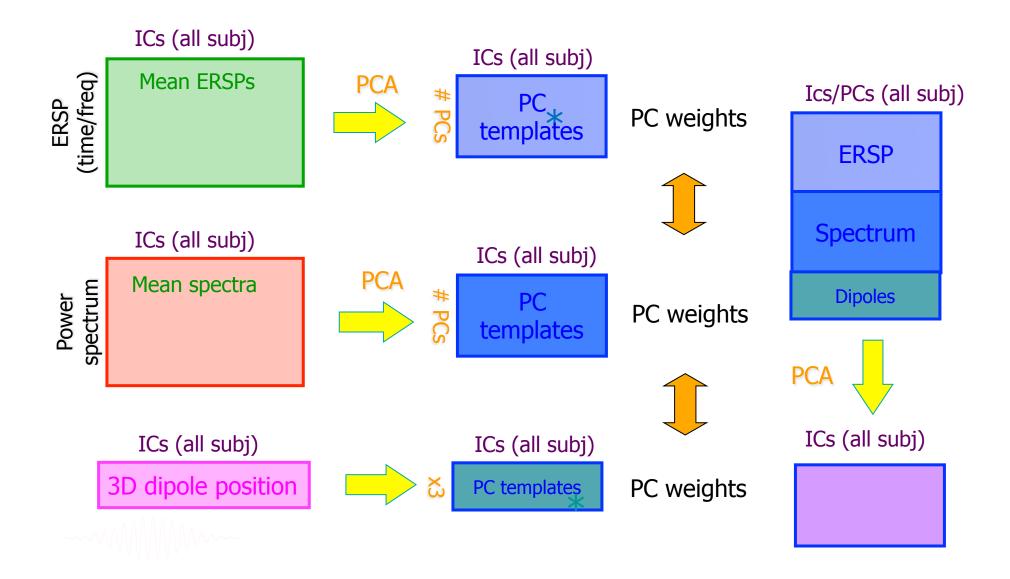
|   | Select and compute comp                                           | oonent measures for later clu | stering - | pop_precomp()                  | • ×            |                   |
|---|-------------------------------------------------------------------|-------------------------------|-----------|--------------------------------|----------------|-------------------|
|   | Pre-compute component                                             |                               |           |                                |                |                   |
| l | Compute ERP/spectrum                                              |                               |           |                                |                |                   |
| L | List of measures to prec                                          | ompute                        |           |                                |                |                   |
| L | 🔽 ERPs                                                            | Baseline ([min max] in m      | s)        | [-200 0]                       |                |                   |
| L | Power spectrum                                                    | Spectopo parameters           |           |                                | Test           |                   |
|   | ERSPs     ITCs                                                    | Time/freq.parameters          | 'cyc      | les', [3 0.5], 'nfreqs', 100 🔻 | Test           |                   |
|   | <ul> <li>Scalp maps</li> <li>Recompute even if present</li> </ul> | nt on disk                    |           |                                | Time-<br>optic | -frequency<br>ons |
|   | Help                                                              |                               |           | Cancel                         | Ok             |                   |

-----

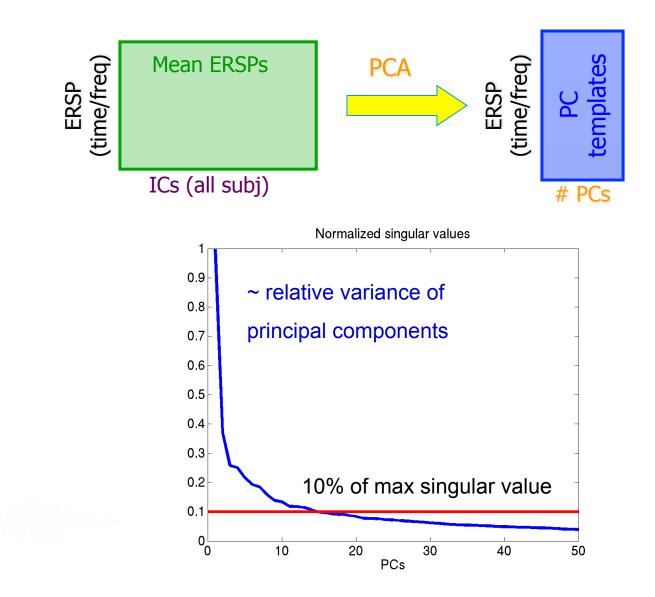
## **Cluster components**

|      | EE                                                               | GLAB v6.0b                    |                                                       |                |           |                           |                               |                  |      |  |
|------|------------------------------------------------------------------|-------------------------------|-------------------------------------------------------|----------------|-----------|---------------------------|-------------------------------|------------------|------|--|
| File | e Edit <b>Tools</b> Plot                                         | Study Datasets Help 🛥         |                                                       |                |           |                           |                               |                  |      |  |
|      | -STUDY set: AI                                                   | Edit study info               |                                                       |                |           |                           |                               |                  |      |  |
|      | -STODT SEL AI                                                    | Precompute channel measures   |                                                       |                |           |                           |                               |                  |      |  |
|      | Study filename:                                                  | Plot channel measures         |                                                       |                |           |                           |                               |                  |      |  |
|      | Study task name                                                  | Precompute component measures |                                                       |                |           |                           |                               |                  |      |  |
|      | Nb of subjects<br>Nb of conditions                               | Build preclustering array     |                                                       |                |           |                           |                               |                  |      |  |
|      | Nb of sessions                                                   | Cluster components            |                                                       |                |           |                           |                               |                  |      |  |
|      | Nb of groups                                                     | Edit/plot clusters            | Select ar                                             | nd compute com | ponent me | asures for later clusteri | ng pop_p                      | reclust()        |      |  |
|      | <ul> <li>Epoch consistency</li> <li>Channels per fram</li> </ul> |                               |                                                       |                |           |                           |                               |                  |      |  |
|      | Channel locations                                                | DU                            | uild pre-clustering ma<br>elect the cluster to refine |                |           |                           | archy will                    | he overwritten)  |      |  |
|      | Clusters                                                         | 1                             | lect the cluster to renne                             | -              |           |                           | <ul> <li>arcny win</li> </ul> | be overwritten)  |      |  |
|      | Status<br>Total size (Mb)                                        | Pre-clustered<br>32.4         |                                                       |                |           | (,                        |                               |                  |      |  |
|      | rotar size (mby                                                  | 52.1                          |                                                       |                |           |                           | -                             |                  |      |  |
|      |                                                                  | (n                            | ote:only measures that                                | have been prec | omputed r | mav be used)              |                               |                  |      |  |
|      |                                                                  |                               | ad                                                    | Dims. Norm     |           |                           |                               |                  |      |  |
|      |                                                                  | ×                             |                                                       | 10             | 1         | Freq.range [Hz]           | 3 25                          |                  |      |  |
|      |                                                                  | ×                             |                                                       | 10 🗹           | 1 10      | Time range [ms]           | 0 600                         |                  |      |  |
|      |                                                                  |                               | scalp maps                                            | 10             | 10        | Use channel values        | -                             | Absolute values  |      |  |
|      |                                                                  | ×                             |                                                       | 20             | 1         | Time range [ms]           | 0 1500                        | Freq. range [Hz] | 3 45 |  |
|      |                                                                  | ~                             | ITCs                                                  | 10 🖌           | 1         | Time range [ms]           | 0 600                         | Freq. range [Hz] | 2 30 |  |
|      |                                                                  |                               | Final dimensions                                      | 10             | Help      |                           |                               |                  |      |  |
|      |                                                                  |                               | Save STUDY to file                                    | /              | home/juli | e/WorkshopSD2007/         | STUDY/ati                     | tention.study    |      |  |
|      |                                                                  |                               |                                                       |                |           |                           |                               | 01               |      |  |
|      |                                                                  |                               | Cancel                                                |                |           | Help                      |                               | Ok               |      |  |
|      |                                                                  |                               |                                                       |                |           |                           |                               |                  |      |  |

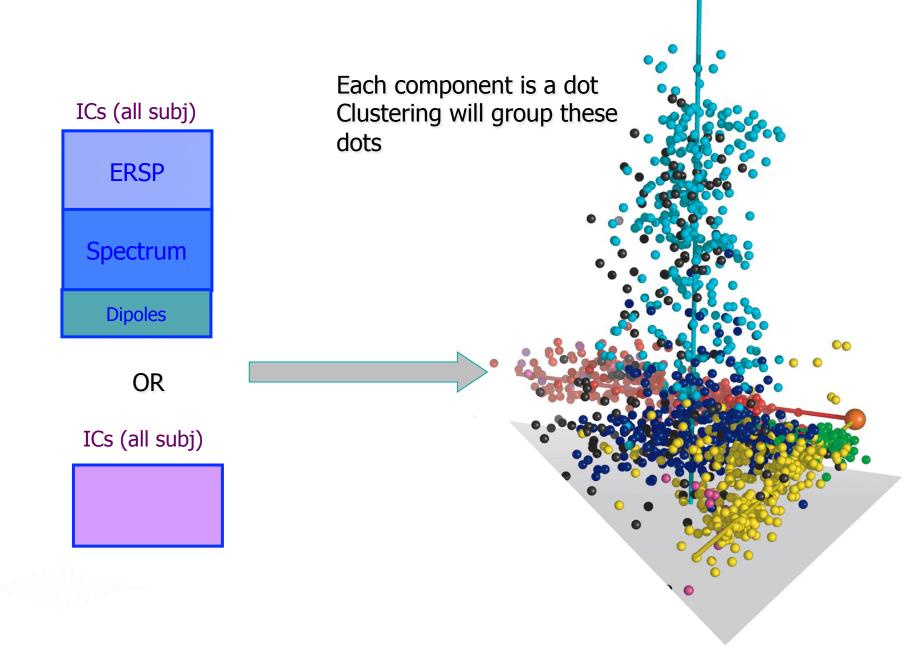
#### **Precluster schematic**



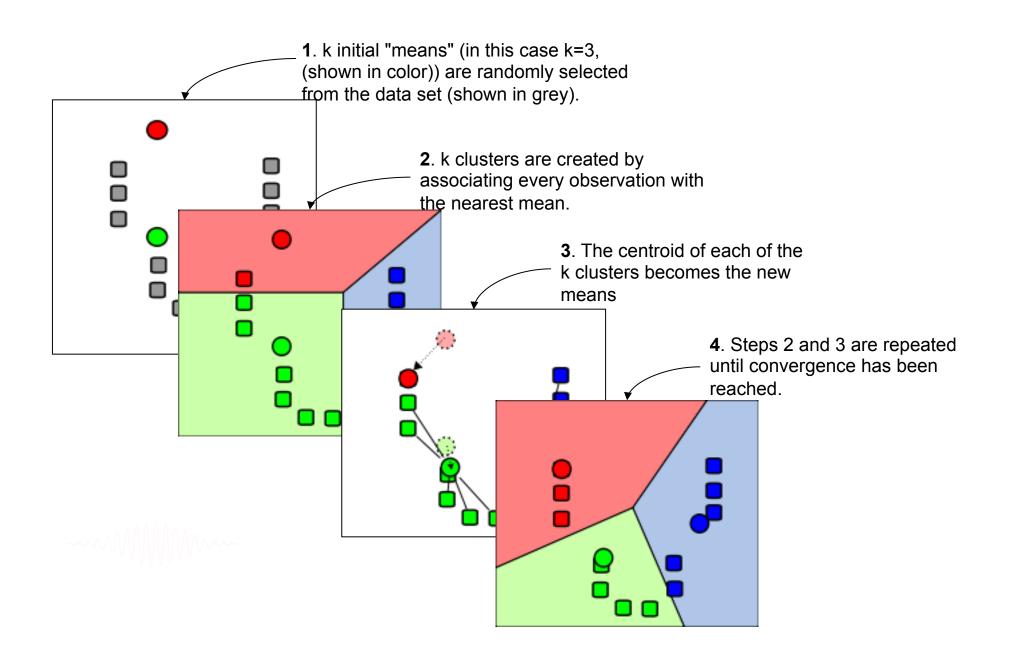
#### **Precluster: Use singular values from PCA**



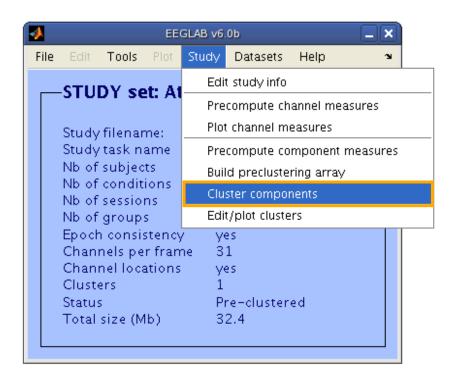
#### **Precluster schematic**

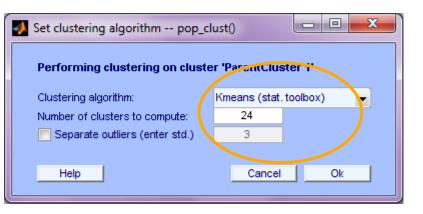


# **Classical KMean**



#### **Cluster components**





#### Choosing data measures

What measure(s) should you use?

It depends on your final cluster criteria...

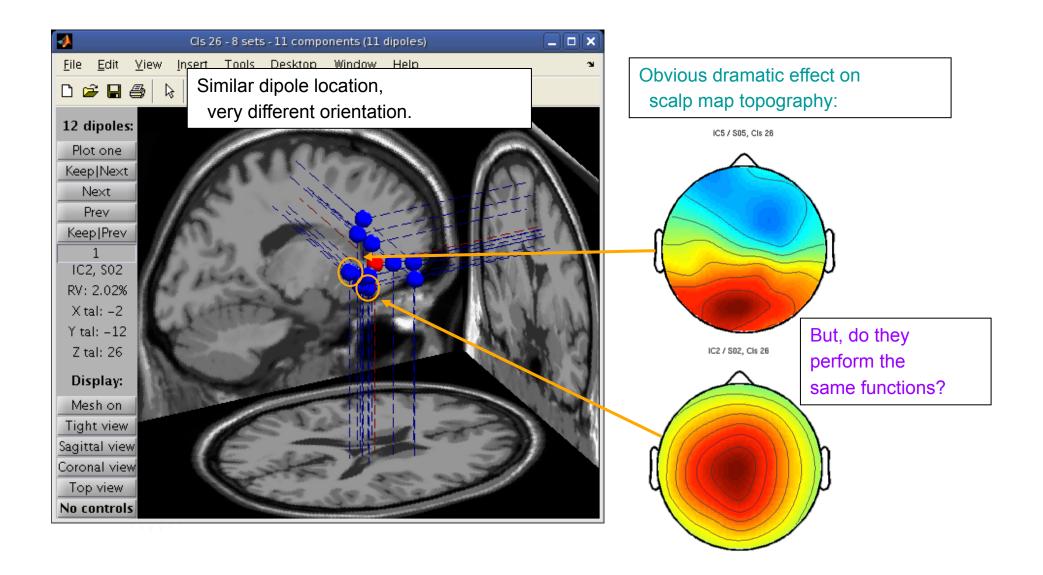
- If for example, your priority is dipole location, then cluster only based on dipole location...

But consider:

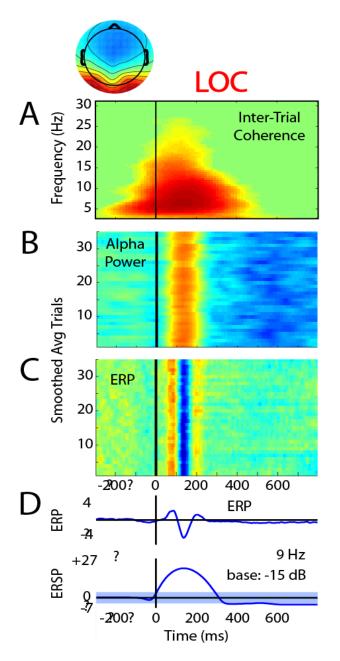
- What is the difference between these two components?

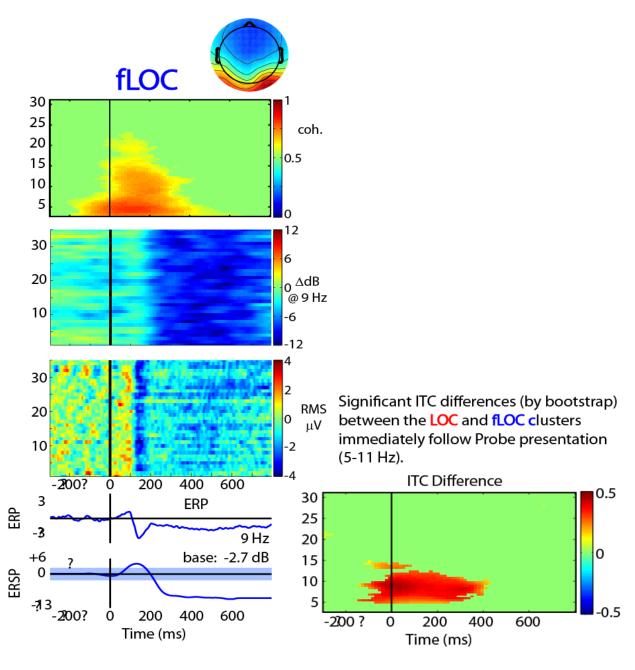


#### **Choosing data measures**

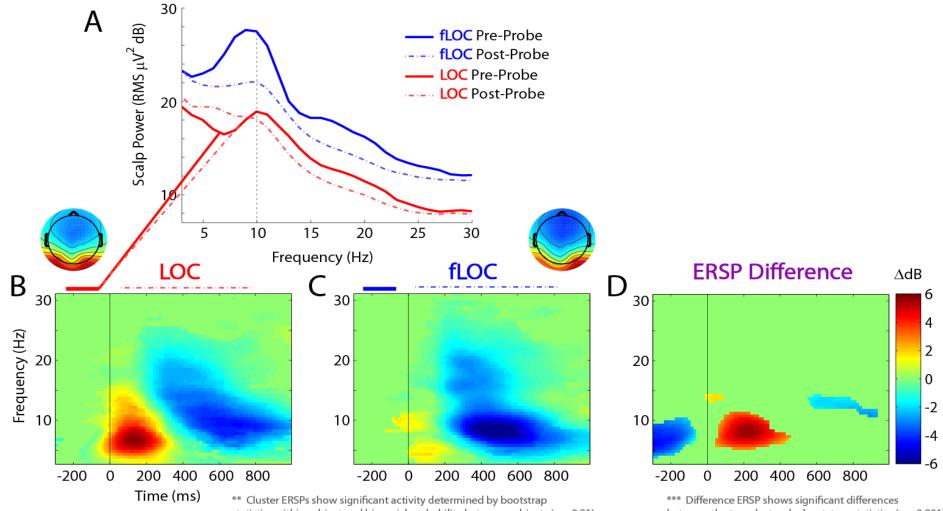


#### **Subject differences?**





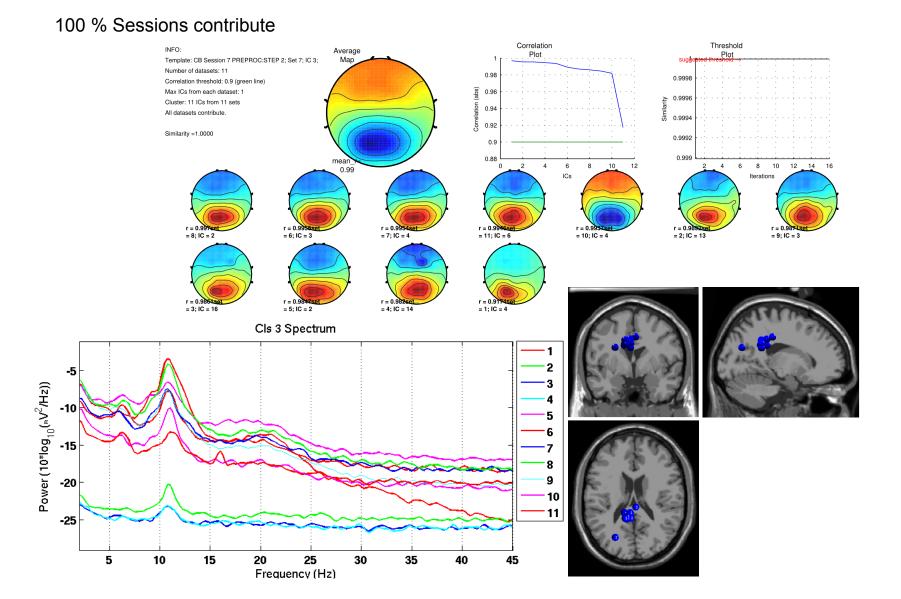
#### **Subject differences?**



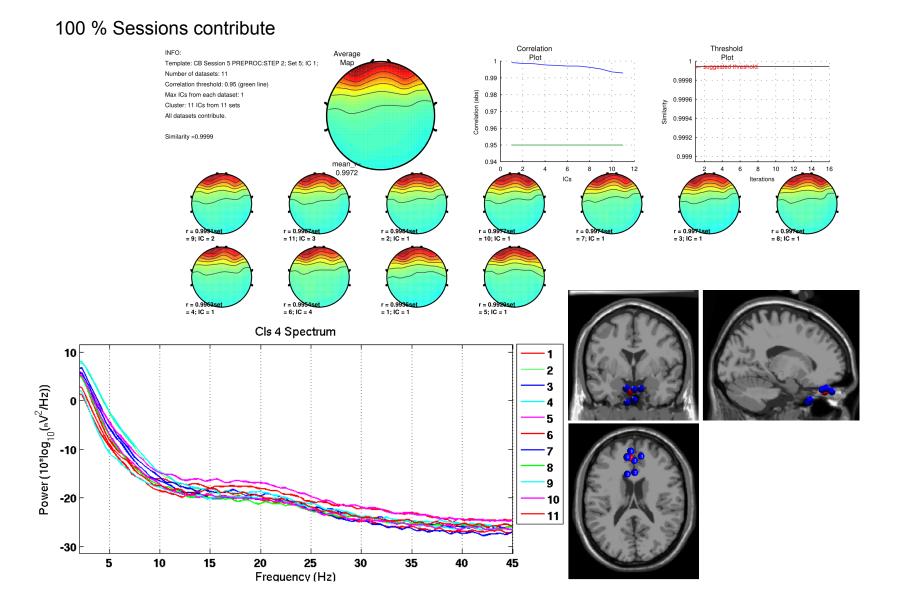
statistics within subject and binomial probability between subjects (p < 0.01)

between the two clusters by bootstrap statistics (p < 0.001)

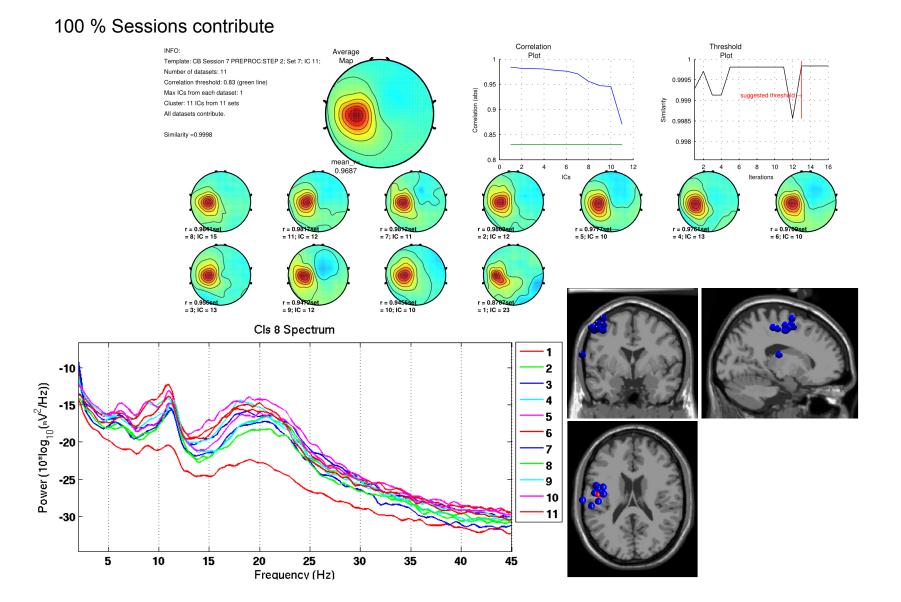
## **Results (Cluster 1)**



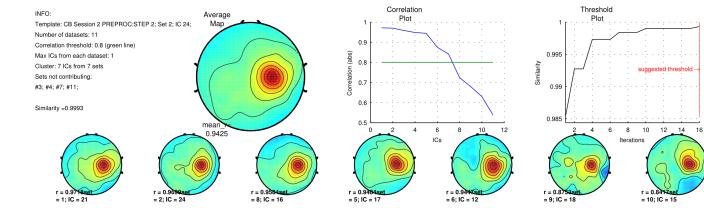
## **Results (Cluster 2)**



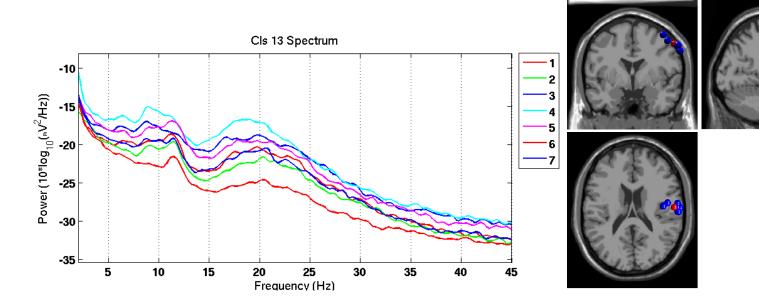
## **Results (Cluster 8)**



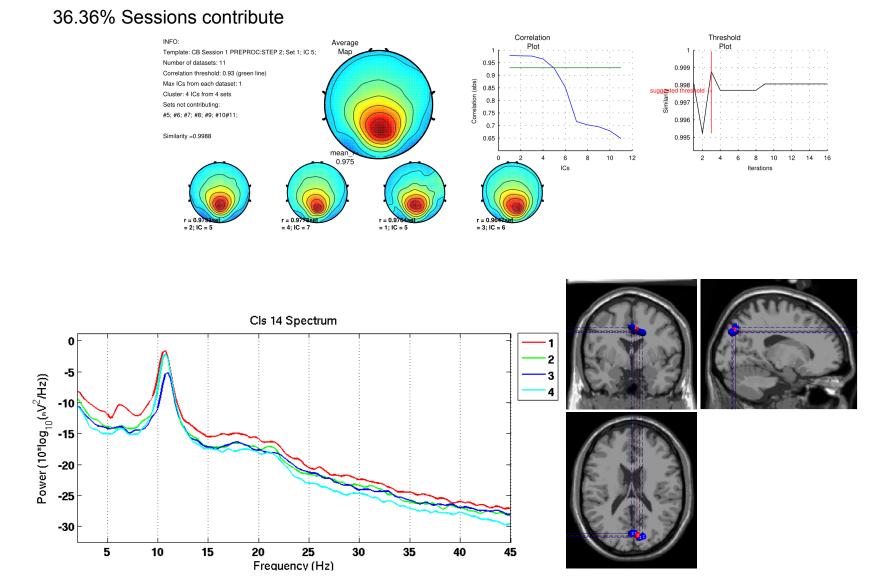
#### **Results (Cluster 13)**



#### 63.64% Sessions contribute



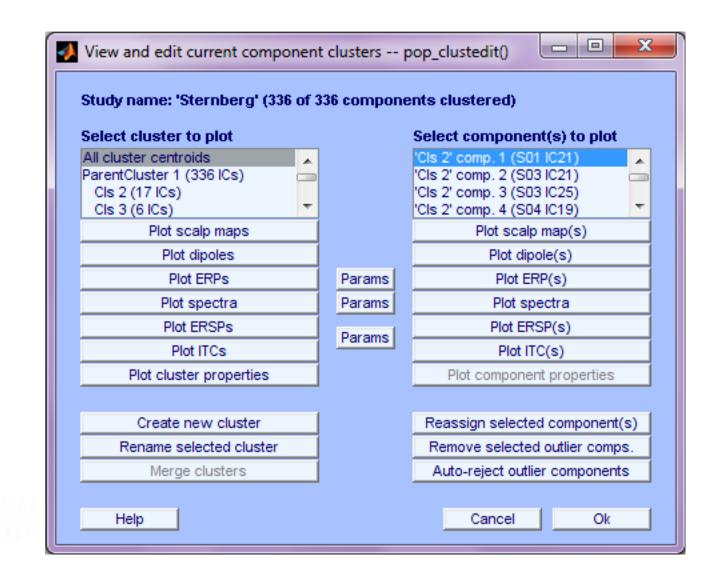
#### **Results (Cluster 14)**



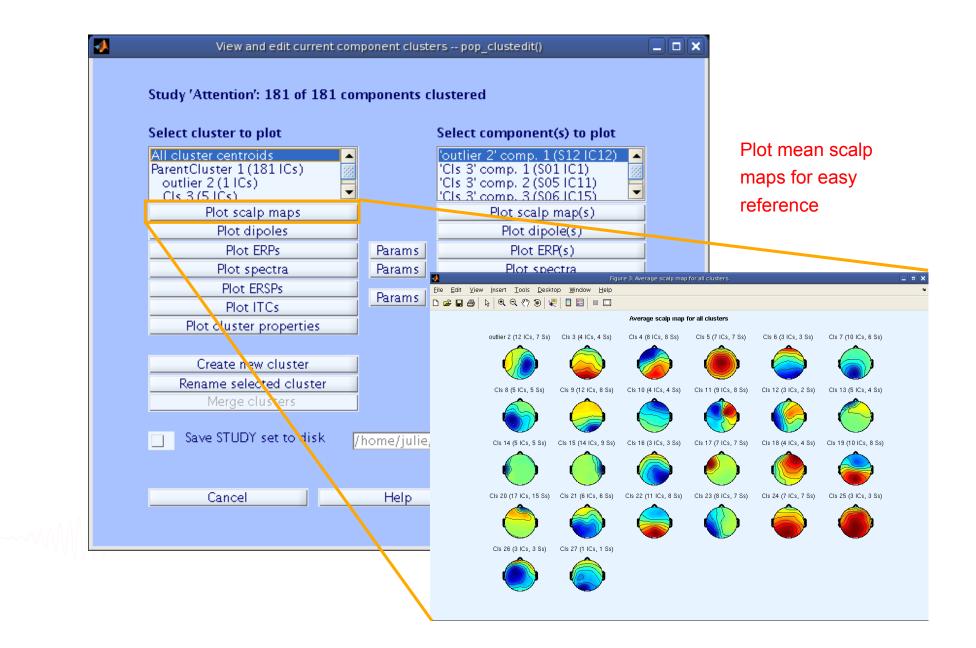
#### View and edit clusters

| -    |                                          |                                             | EE                         | GLAB v6.                      | .0b                              |            |         |  |
|------|------------------------------------------|---------------------------------------------|----------------------------|-------------------------------|----------------------------------|------------|---------|--|
| File | Edit                                     | Tools                                       | Plot                       | Study                         | Datasets                         | Help       | Ľ       |  |
|      | сти                                      | DY se                                       |                            | Edi                           | t study info                     |            |         |  |
|      | 510                                      | 51.36                                       | AI                         | 1                             | compute ch                       | annel meas | ures    |  |
|      | Study                                    | filenar                                     | ne:                        | Plot                          | t channel me                     | easures    |         |  |
|      | · ·                                      | task n                                      |                            | Pre                           | compute co                       | mponent m  | easures |  |
|      |                                          | isubjec<br>icondit                          |                            | Buil                          | ld precluste                     | ring array |         |  |
|      |                                          | sessio                                      |                            | Clu                           | ster compor                      | nents      |         |  |
|      |                                          | group                                       |                            |                               | t/plot cluste                    | rs         |         |  |
|      |                                          |                                             |                            |                               | es<br>1                          |            |         |  |
|      |                                          |                                             |                            |                               | -                                |            |         |  |
|      | Clust                                    |                                             | 1010115                    | 2                             |                                  |            |         |  |
|      | Status                                   | 5                                           |                            | P                             | re-cluster                       | ed         |         |  |
|      | Total                                    | size (M                                     | lb)                        | 3                             | 9.1                              |            |         |  |
|      |                                          |                                             |                            |                               |                                  |            |         |  |
|      | Epoch<br>Chan<br>Chan<br>Clust<br>Status | n consi:<br>nels pe<br>nel loca<br>ers<br>s | stency<br>r fran<br>ations | / ye<br>ne 3<br>ye<br>2<br>Pi | es<br>1<br>es<br>6<br>re-cluster |            |         |  |

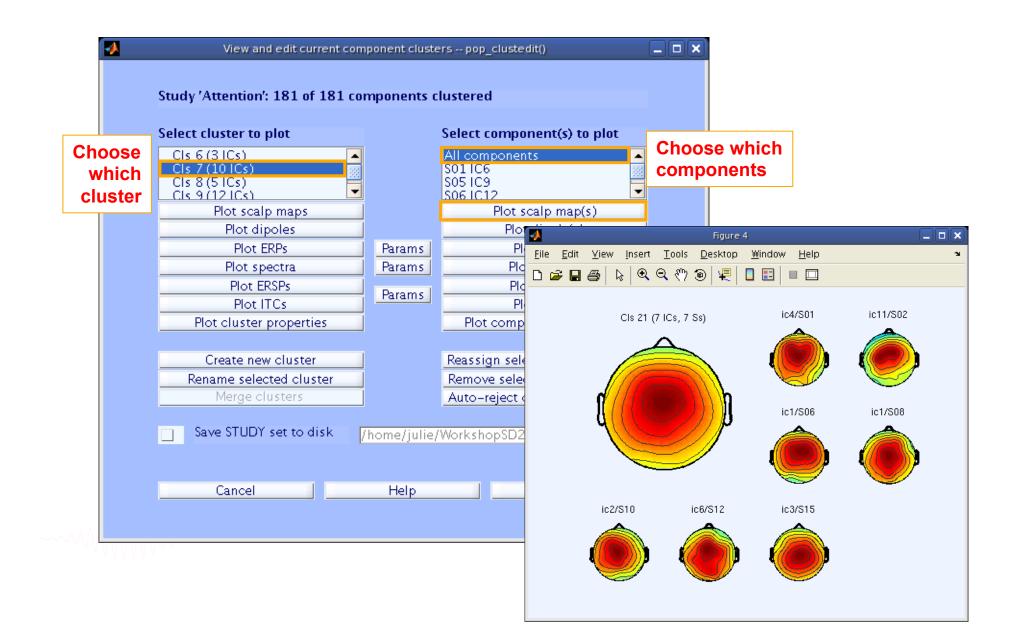
#### **Plot/edit clusters**



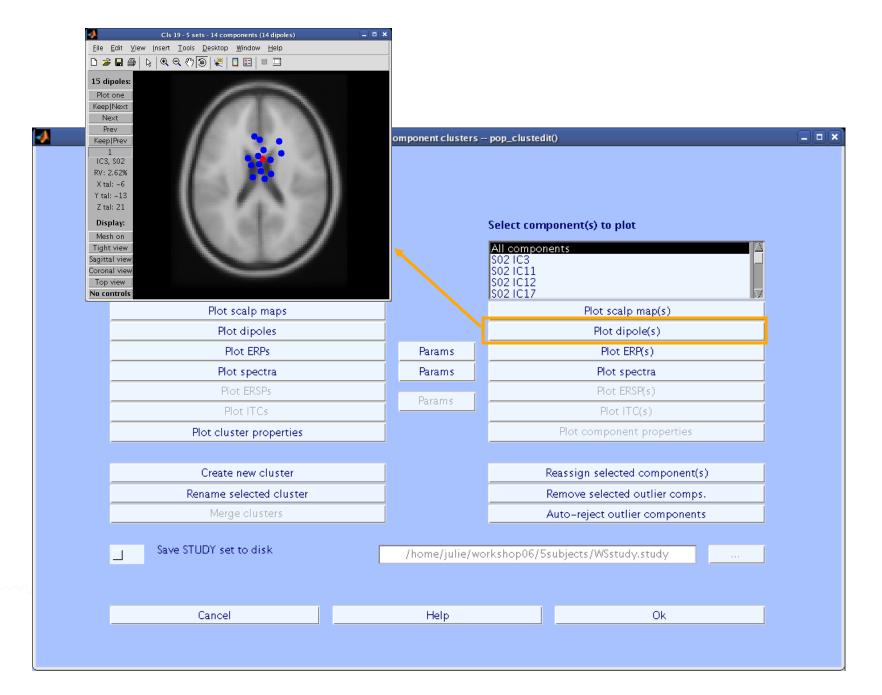
#### **Plot cluster data**



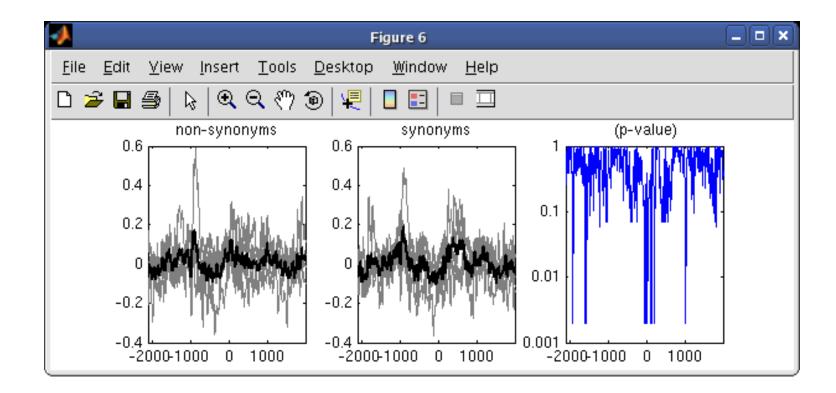
#### **Plot cluster data**



#### **Plot cluster data**



#### **Plot cluster ERP**



#### **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



|     |                    |      | ub-cl | ustering (an | r components from study "<br>y existing sub-hierarchy will be o<br>r1 (151 ICs) |                                     |
|-----|--------------------|------|-------|--------------|---------------------------------------------------------------------------------|-------------------------------------|
| Pre | -compute or Load   | Dims | i     | Norm. Rel    | wt.                                                                             |                                     |
|     | spectra.           | 10   |       | 1            | Frequency range [Hz]                                                            | 3 25                                |
| F   | ERPs               | 10   |       | 1            | Latency range in ms [lo hi]                                                     | -2100 1995                          |
| -   | dipoles            | 3    | •     | 10           |                                                                                 |                                     |
| Ā   | scalp maps         | 10   | Ē     | 1            | Use channel values 😐                                                            | <ul> <li>Absolute values</li> </ul> |
|     | ERSPs              | 10   |       | 1            | Time/freq. parameters                                                           | s', [3 25], 'cycles', [3 0.5], 'pa  |
|     | ITCs               | 10   |       | 1            | Timm/ling pareipeters                                                           | 91, [3-25], "cycles", [3-0-5], "pe  |
| •   | Final dimensions   | 10   |       | Help         |                                                                                 |                                     |
| ۰   | Save STUDY to file | • [  |       | /6444/001    | nnan4/acro/Seubjacto/N400pcac                                                   |                                     |
|     | Cancel             |      | 11    |              | Help                                                                            | Ok                                  |

#### **Measure projection**

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

| • | Measure Product        | clustering p       | oop_mpcluster() |    |
|---|------------------------|--------------------|-----------------|----|
|   | Number of clusters to  | compute:           | 10              |    |
|   | Relative dipole weight |                    |                 |    |
|   | Select measuretures to | ) be used in the ( | clustering:     |    |
|   |                        | <b>⊠</b> Dipole    |                 |    |
|   |                        | RP ERP             |                 |    |
|   |                        |                    |                 |    |
|   |                        | Spectra            |                 |    |
|   |                        | ∐Scalp map         |                 |    |
|   | Separate outliers (e   | nter std.)         | 3               |    |
|   | Help                   |                    | Cancel          | Ok |

(Affinity clustering by Pernet, Martinez, Delorme)

# Exercise

- Load the STUDY stern.study
- Precompute spectrum, ERP and scalp maps for components
- Precluster and cluster components using spectrum and dipoles
- 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