

Clustering Independent Components of EEG Data



Scott Makeig

Institute for Neural Computation
University of California San Diego

28th EEGLAB Workshop

La Jolla, CA

November, 2018



Why cluster independent components across subjects or sessions?



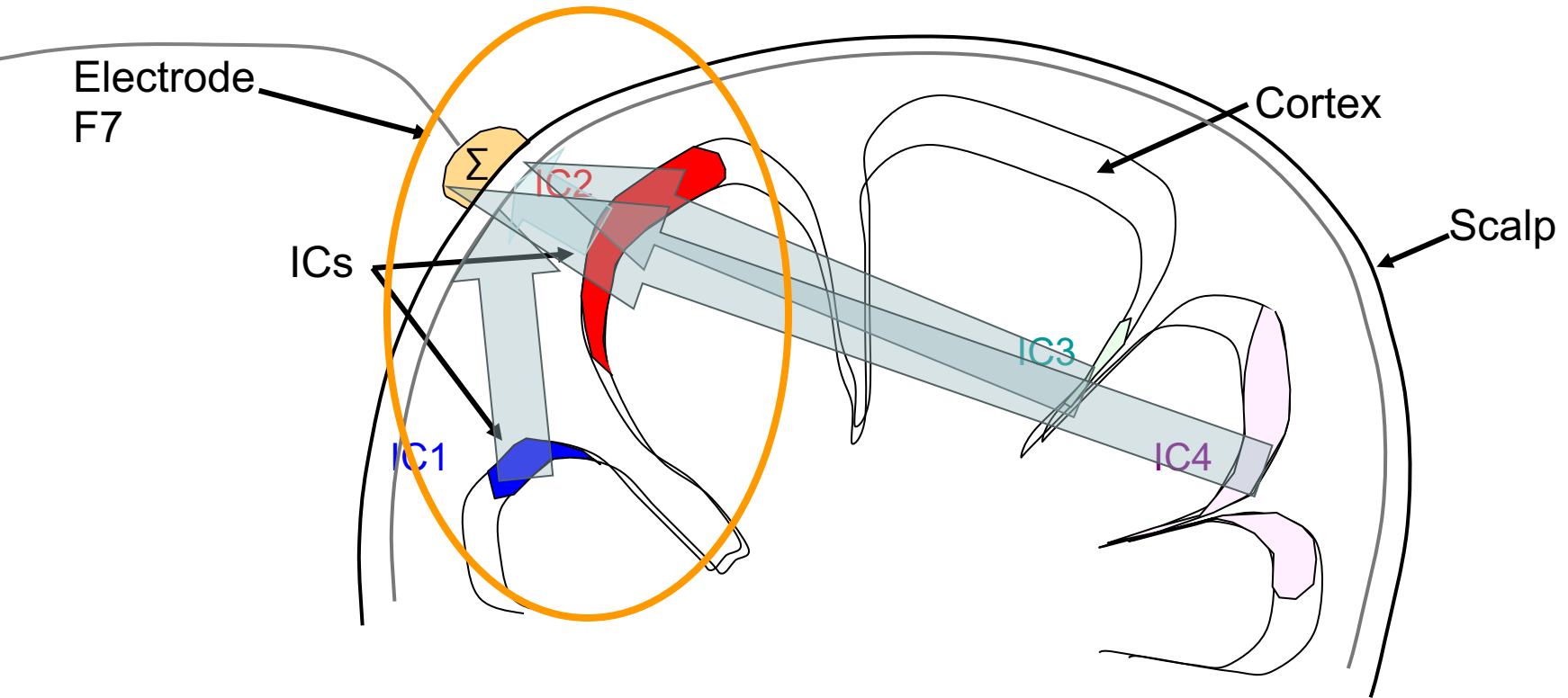
- ICA transforms the data from a channel basis (activity recorded at each channel)
- to a component basis (activity computed at each IC).
- Normally, EEG researchers assume that, for example, electrode channel F7 == F7 == F7 ... in each subject – and then ‘cluster’ their data assuming channel equivalence.
- This amounts to the simple assumption

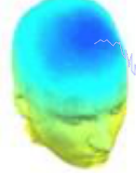
“Your Cz is My Cz!”

- But this is only *roughly* correct !

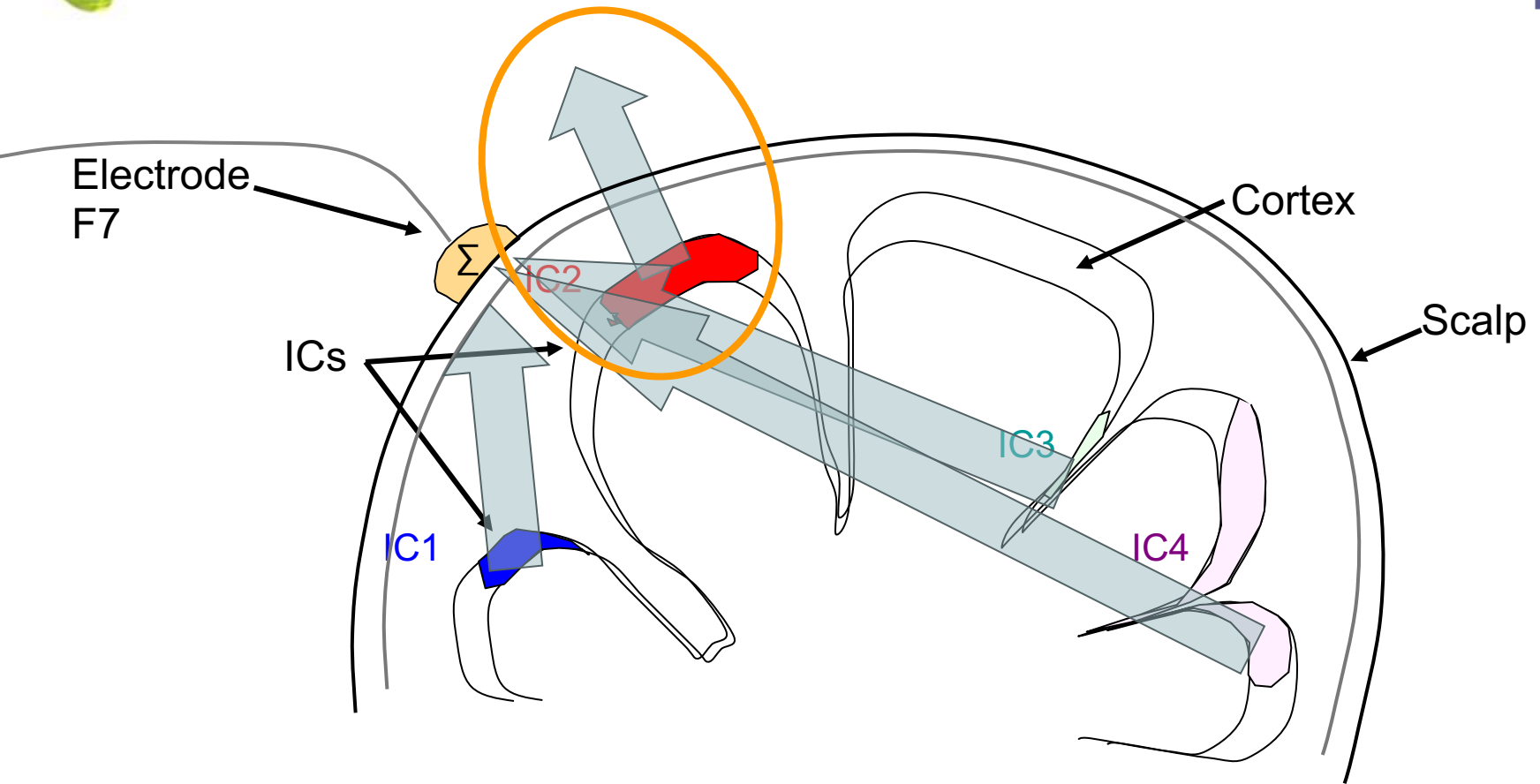


Example: First Subject



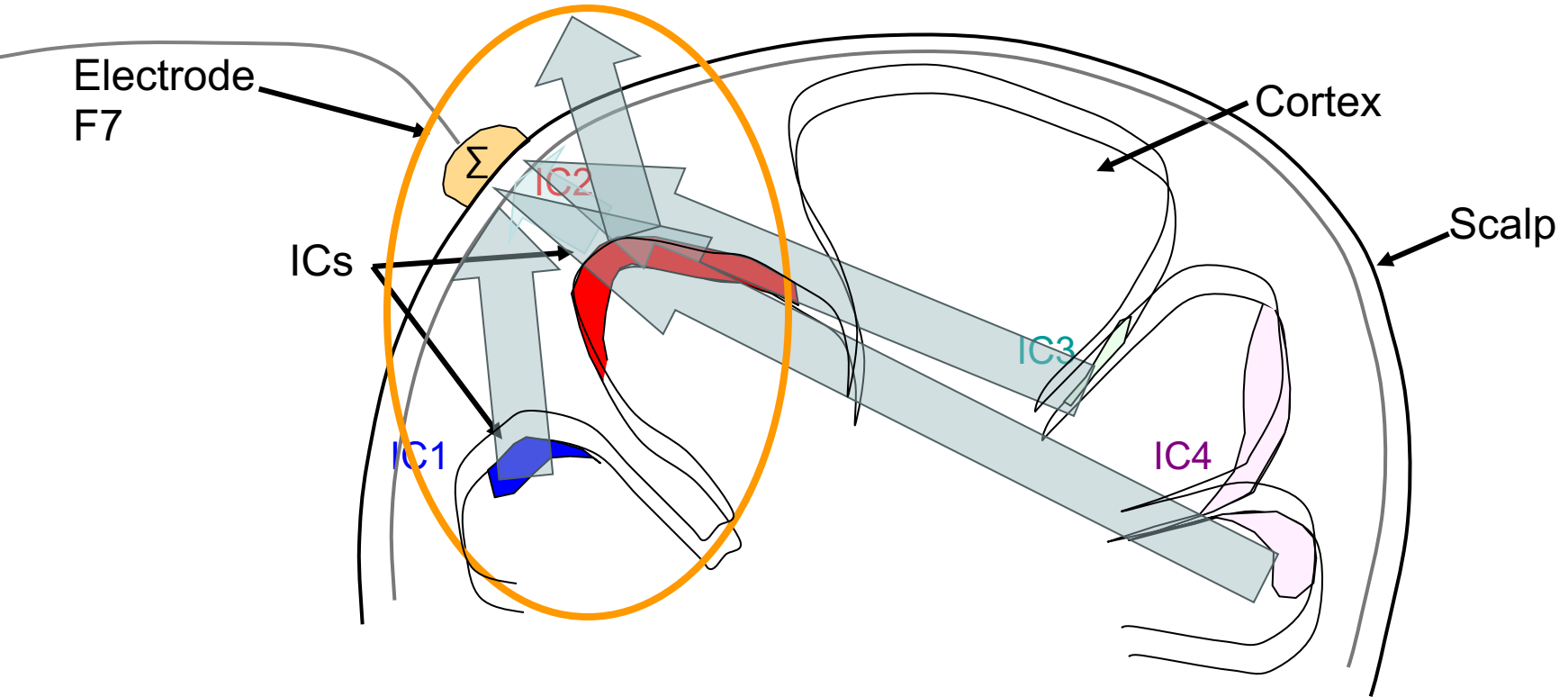


Second Subject



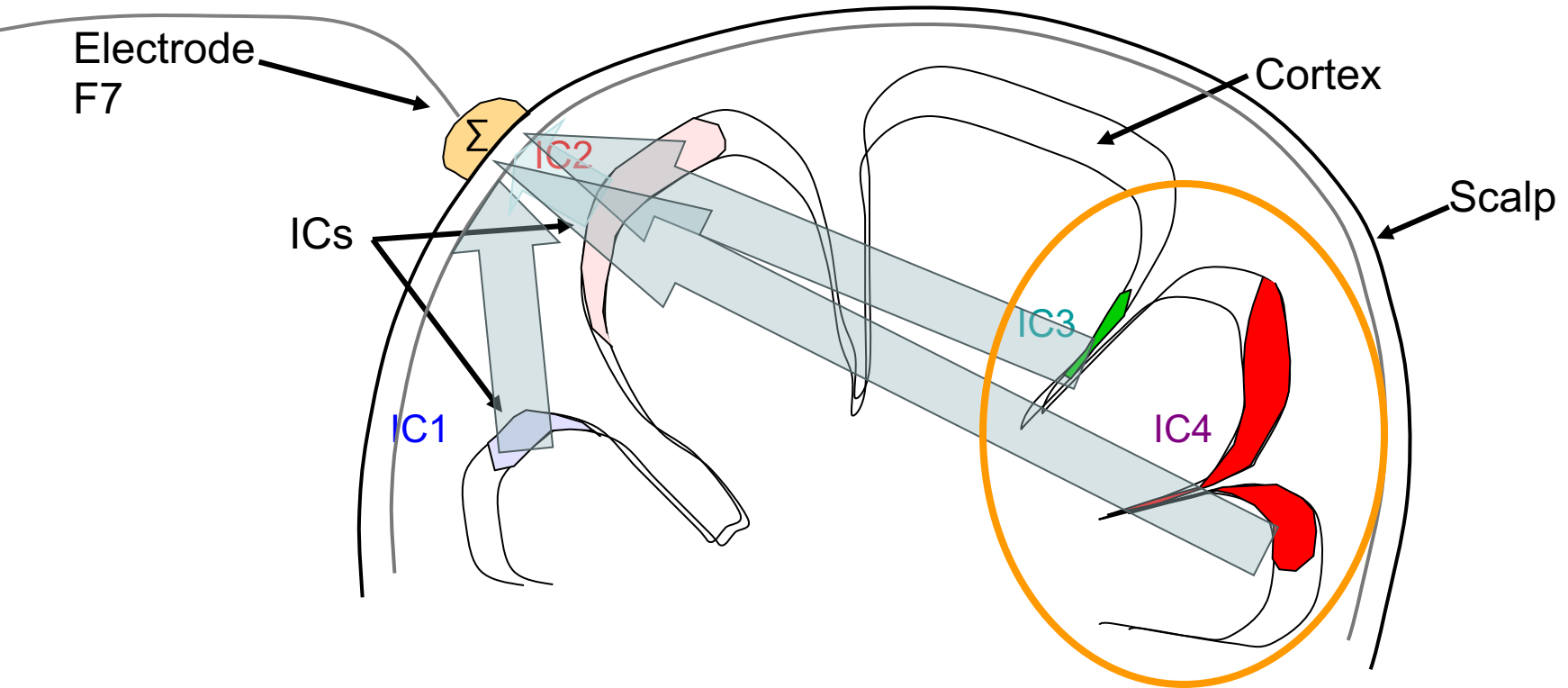


Third Subject





Fourth Subject





Clustering ICA components by eye



Left mu

Right mu





So how to cluster components?

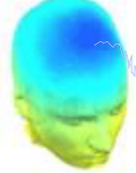


The same problems hold for clustering independent components

Across Ss, components don't even have “the same” scalp maps!

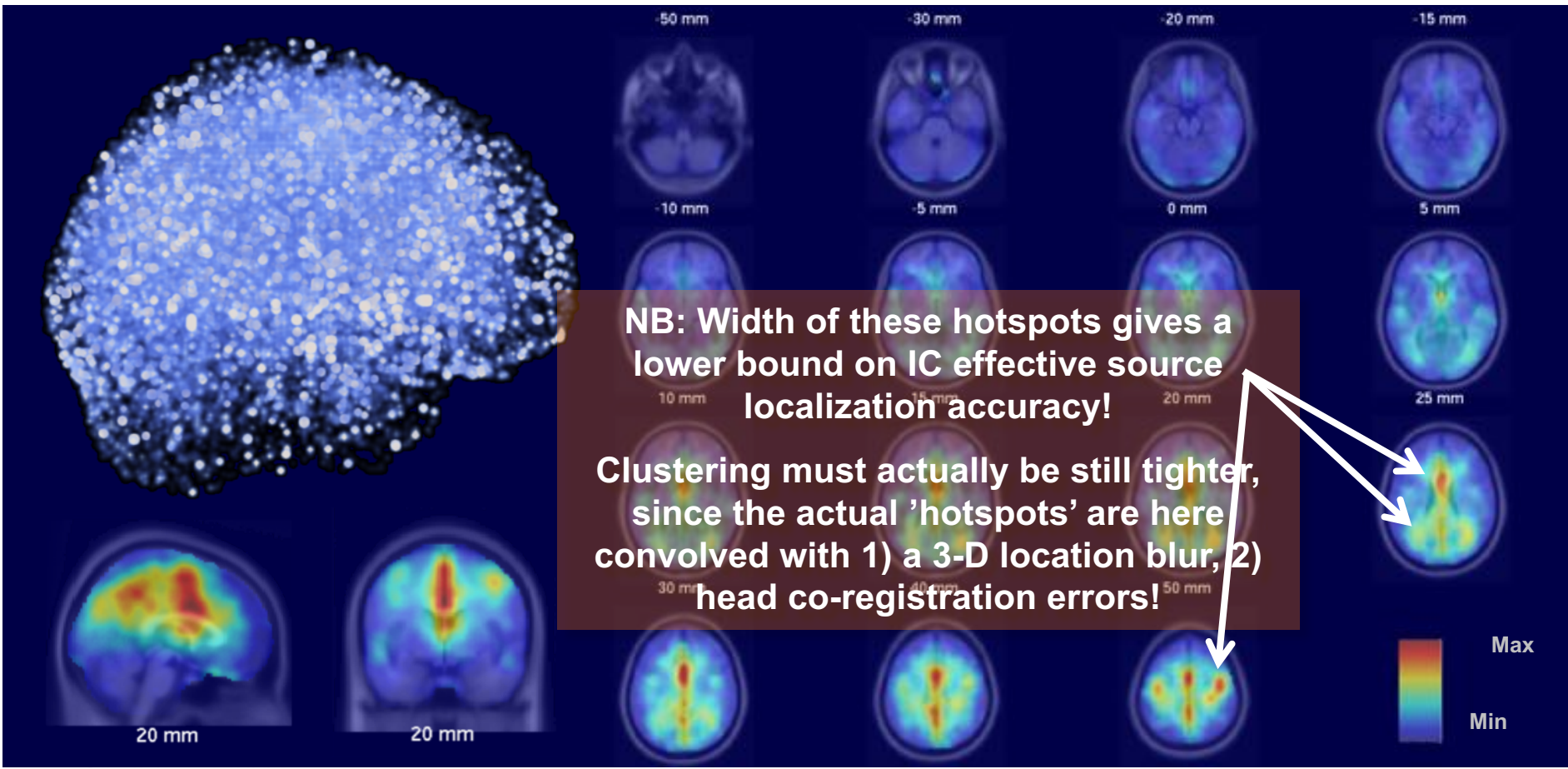
→ Are “the same” components found across subjects?

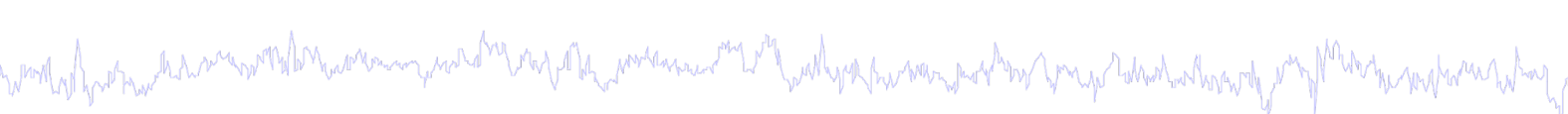
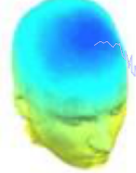
- What should define “*the same*” (i.e., “*component equivalence*”)?
 - Similar scalp maps?
 - Similar cortical or 3-D equivalent dipole locations?
 - Similar activity power spectra?
 - Similar ERPs?
 - Similar ERSPs?
 - Similar ITCs?
 - Or similar *combinations* of the above?? ...



EEG IC Source Locations

(135,794 IC equivalent dipoles!)

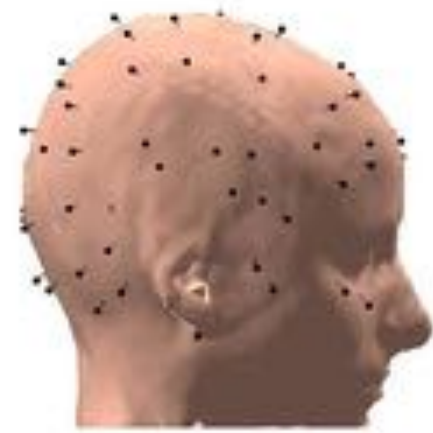
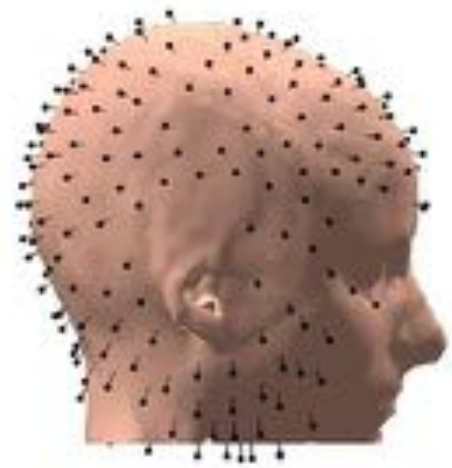




... Some caveats

In this *dipoledensity()* assay ...

- MR head images were not available → brain co-registration crude.
- Single versus dual-dipole model selection was subjective.
- Different electrode montages → mis-localization effects.
- Electrode locations were not all digitized – some ‘gestimated’ !
- Brain geometries differ!





So how to cluster components?

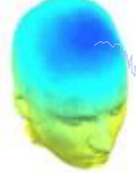


The same problems hold for clustering independent components

Across Ss, components don't even have "the same" scalp maps!

→ Are "the same" components found across subjects?

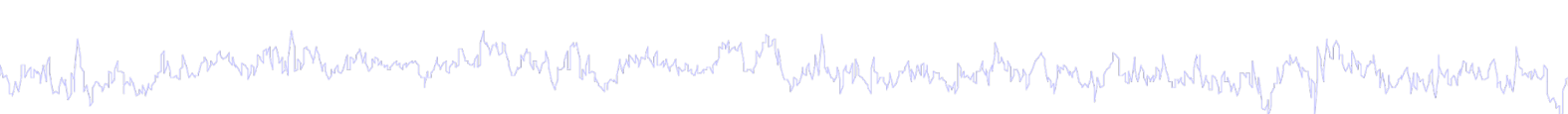
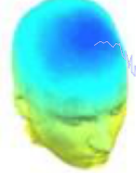
- What should define "the same" (i.e., "component equivalence")?
 - Similar scalp maps?
 - Similar cortical or 3-D equivalent dipole locations?
 - Similar activity power spectra?
 - Similar ERPs?
 - Similar ERSPs?
 - Similar ITCs?
 - Or similar **combinations** of the above?? ...
 - **EEGLAB clustering supports all these possibilities.**



Study IC Clustering: Assumptions

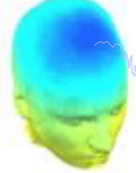
- Assumes there are *functionally equivalent* ICs across most subjects.
- Assumes these ICs have *similar responses* to experimental conditions across **a set** of measures
(ERP, ERSP, ITC...)
- Creates ***non-overlapping IC partitions*** making each IC belong to only one cluster.



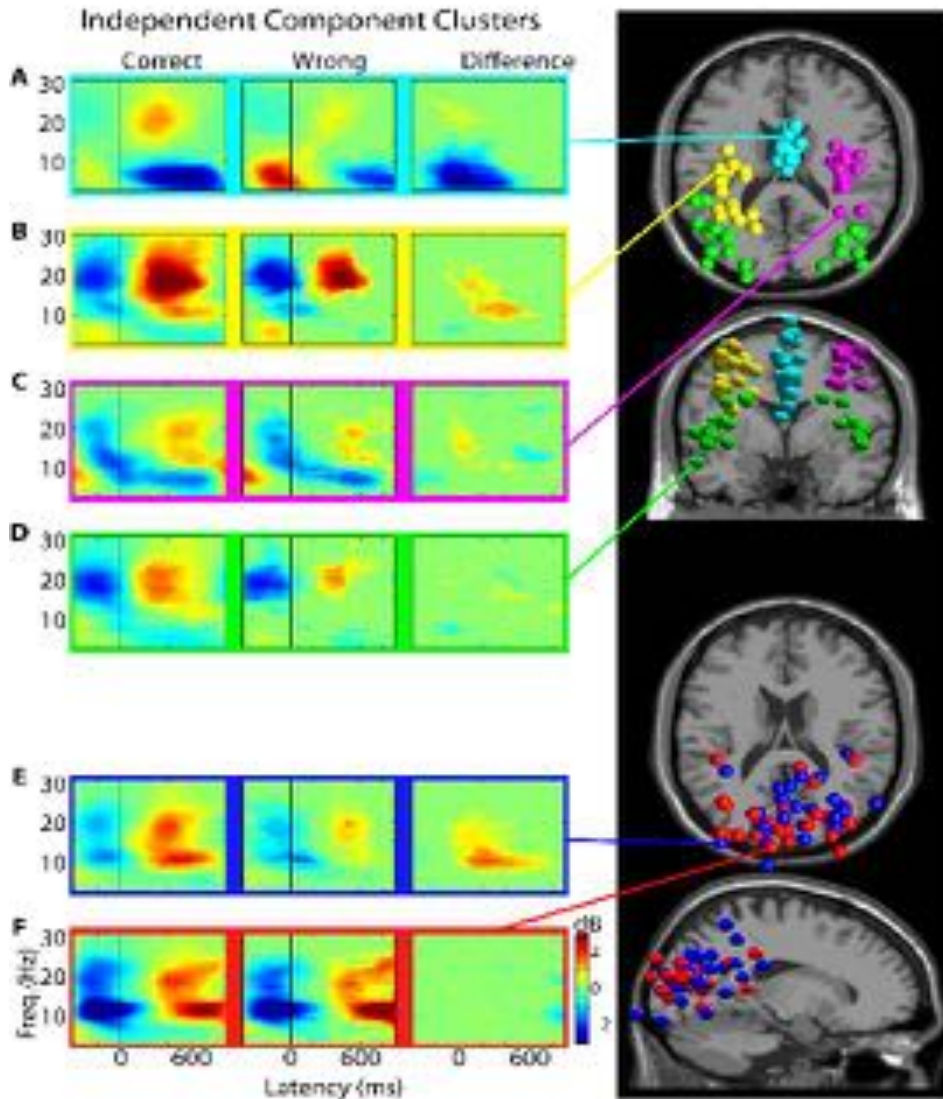


EEGLAB Study Clustering strategy

1. Cluster on **multiple measures (dipole locations, scalp maps, spectra, ERPs, ITCs, ERSPs, ...)** **in one or more conditions.**
2. **Reduce the dimension** of each measure to a principal component subspace.
3. Compose a PCA-reduced **position vector** for each component.
4. **Cluster** the composed component vectors using k-means or other.
5. Use the computed component measures (not PCA-reduced) to **visualize the activities and spatial properties** of the clustered components.
6. Compute and visualize the **cluster-mean measures.**
7. Use **clustered Study set data** as input into 'std_???' functions.



Study IC Clustering



Sometime clusters are spatially separate AND have distinct responses.

In other cases, they may have similar responses or may overlap spatially.

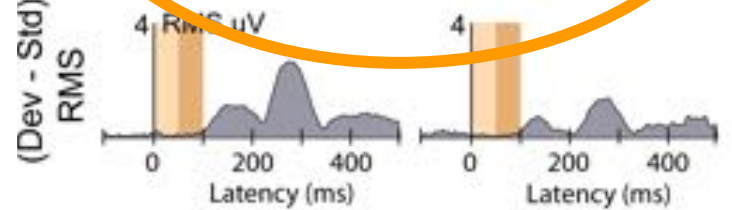
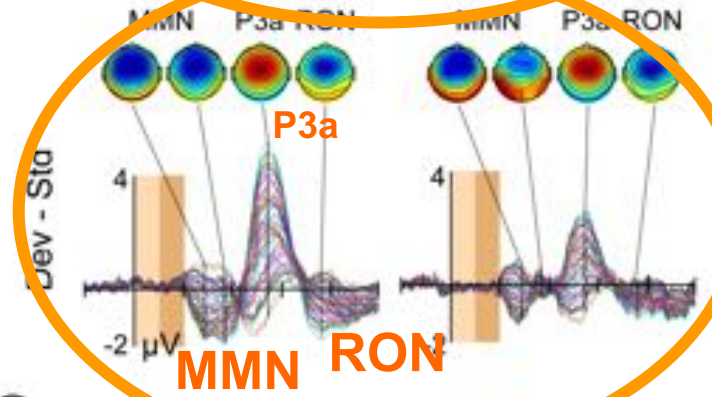
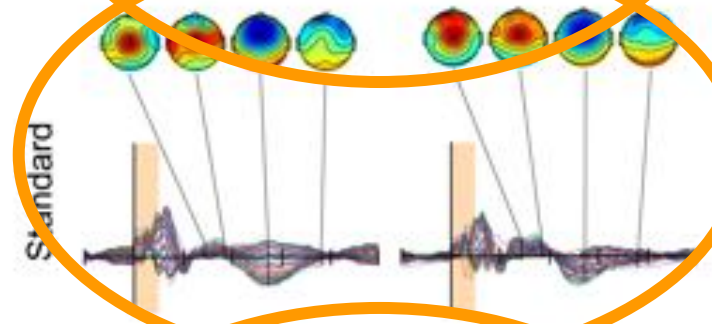
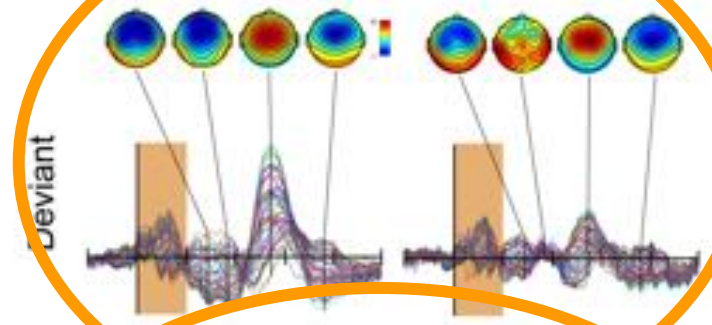


EEGLAB Study Clustering procedure



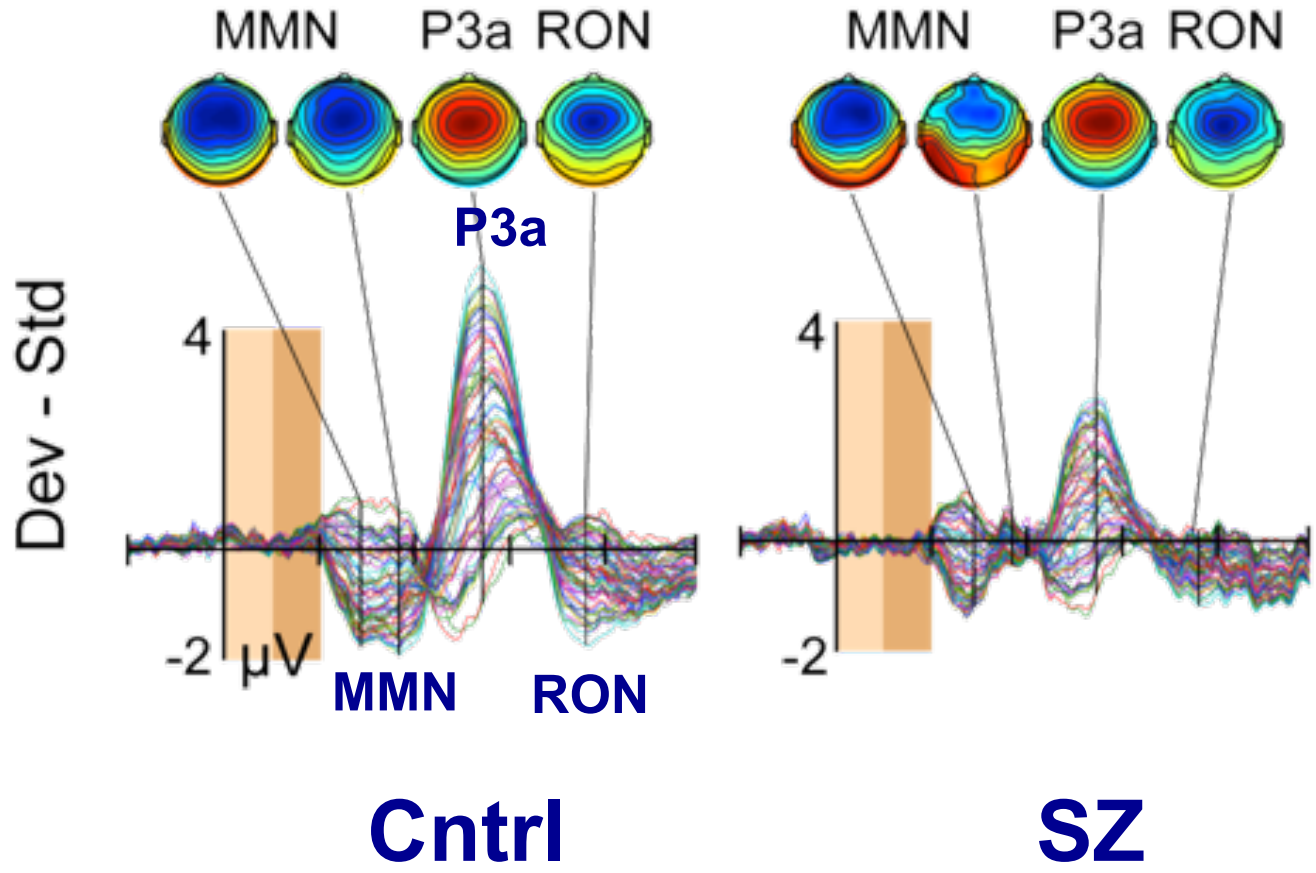
1. Identify a set of datasets as an **EEGLAB Study**.
2. Specify the **subject** code, subject **group**, **condition** and/or **session** for each dataset in the Study.
3. Identify **components to cluster** in each Study dataset.
4. Decide on **component measures** to use in clustering the Study and/or to evaluate the obtained component clusters.
5. **Compute the component measures** for each Study dataset.
6. **Cluster the components on these component measures.**
7. **Review the obtained clusters** (e.g., their scalp maps, dipoles, and activity measures).
8. **Edit the clusters** (manually remove/shift components, make sub-clusters, merge clusters, re-cluster).
9. **Statistically test differences** within or between selected clusters.

Nonpsychiatric Comparison Subjects (NCS) Schizophrenia Patients (SZ)





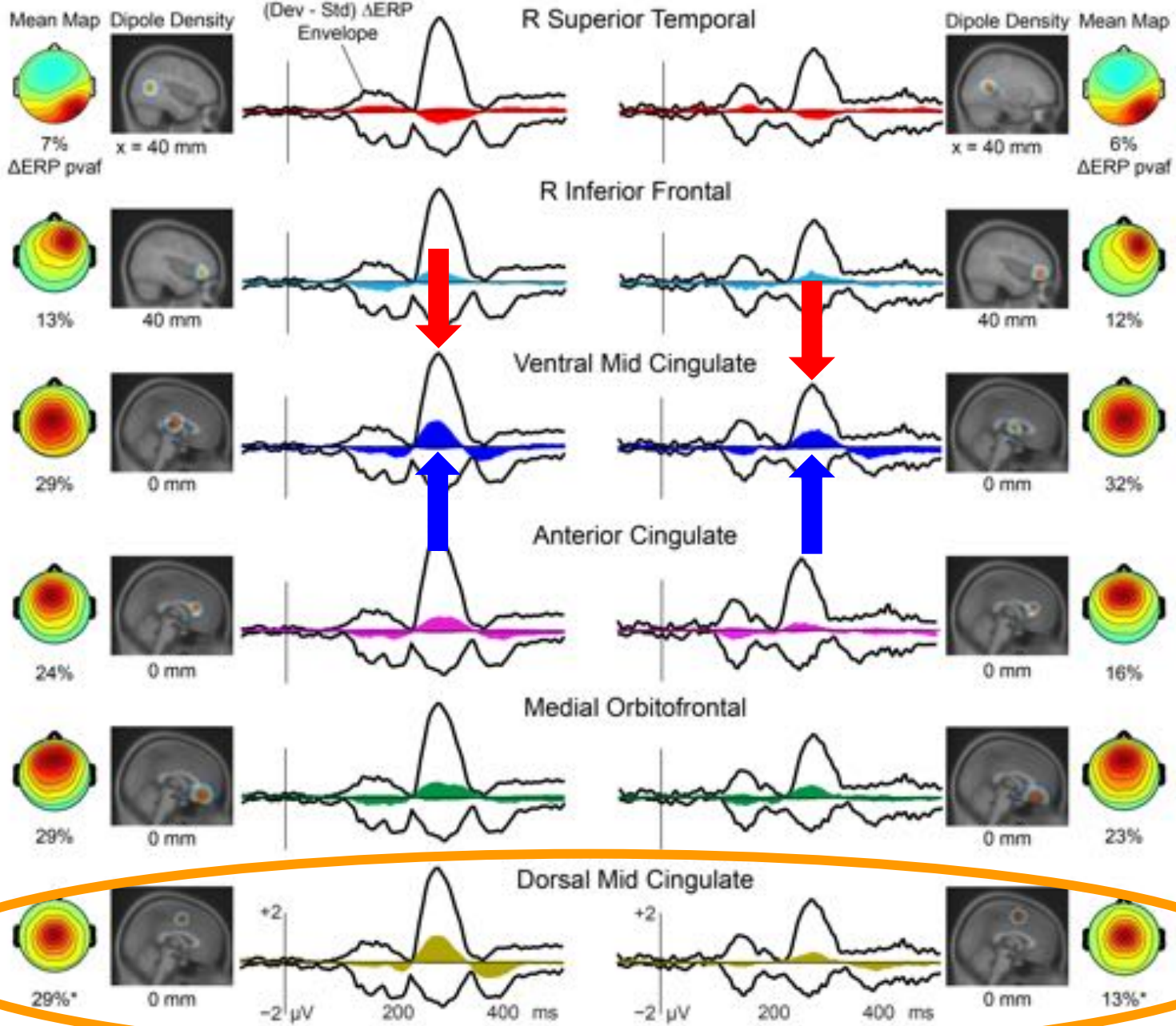
Auditory Deviance Response



The deepest mental trap in electrophysiology lies in the word "THE" !!!

Nonpsychiatric Comparison Subjects (NCS)

Schizophrenia Patients (SZ)



PEAK AMPLITUDES

ERP

r^2



Scalp Electrode (Fz)

Verbal IQ (WRAT)	P3a	0.11
Functional Capacity (UPSA)	RON	0.12

X

R Superior Temporal

Working Memory (LNS Reorder)	RON	0.15
Verbal IQ (WRAT)	RON	0.15
Immediate Verbal Memory (CVLT)	RON	0.28
Delayed Verbal Memory (CVLT)	RON	0.26
Functional Capacity (UPSA)	MMN	0.48
Functional Capacity (UPSA)	RON	0.26

R Inferior Frontal

Negative Symptoms (SANS)	RON	0.36
Psychosocial Functioning (SOF)	RON	0.24
Auditory Attention (LNS Forward)	MMN	0.38
Working Memory (LNS Reorder)	MMN	0.30
Verbal IQ (WRAT)	MMN	0.46

Ventral Mid Cingulate

Positive Symptoms (SAPS)	RON	0.29
Negative Symptoms (SANS)	P3a	0.36
Immediate Verbal Memory (CVLT)	RON	0.41
Delayed Verbal Memory (CVLT)	RON	0.24
Verbal IQ (WRAT)	RON	0.29
Executive Functioning (WCST)	RON	0.24

Anterior Cingulate

Functional Status (GAF)	MMN	0.18
Functional Status (GAF)	RON	0.17
Immediate Verbal Memory (CVLT)	RON	0.25
Delayed Verbal Memory (CVLT)	RON	0.17

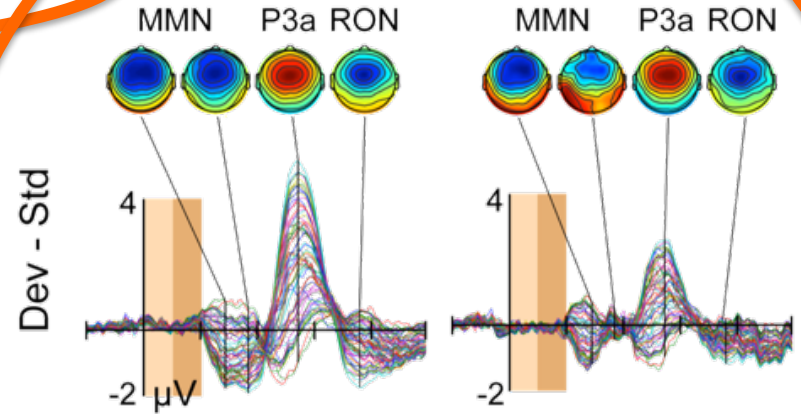
Medial Orbitofrontal

Positive Symptoms (SAPS)	P3a	0.40
Negative Symptoms (SANS)	P3a	0.54
Psychosocial Functioning (SOF)	P3a	0.37
Functional Capacity (UPSA)	P3a	0.32

Dorsal Mid Cingulate

Verbal IQ (WRAT)	P3a	0.15
Executive Functioning (WCST)	MMN	0.18

ADR



Cntrl

SZ

PEAK LATENCIES

ERP

r^2

ADR



Scalp Electrode (Fz)

---n/a---

R Superior Temporal

Functional capacity (UPSA)	MMN	0.25
Delayed Verbal Memory (CVLT)	MMN	0.17

R Inferior Frontal

Negative Symptoms (SANS)	RON	0.51
Psychosocial Functioning (SOF)	RON	0.25
Executive Functioning (WCST)	MMN	0.30
Executive Functioning (WCST)	P3a	0.28

Ventral Mid Cingulate

Negative Symptoms (SANS)	P3a	0.33
Negative Symptoms (SANS)	RON	0.33
Psychosocial Functioning (SOF)	P3a	0.31
Verbal IQ (WRAT)	MMN	0.25
Executive Functioning (WCST)	P3a	0.30

Anterior Cingulate

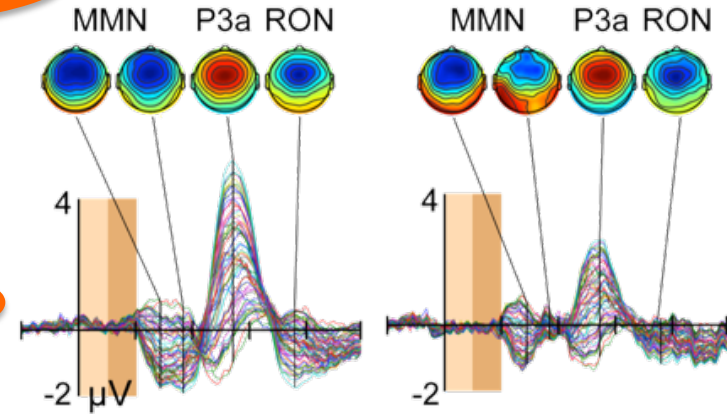
Functional Capacity (UPSA)	RON	0.17
Verbal IQ (WRAT)	MMN	0.24
Auditory Attention (LNS-Forward)	MMN	0.17

Medial Orbitofrontal

Negative Symptoms (SANS)	RON	0.41
Positive Symptoms (SAPS)	RON	0.40
Auditory Attention (LNS-Forward)	MMN	0.29
Executive Functioning (WCST)	P3a	0.32

Dorsal Mid Cingulate

Negative Symptoms (SANS)	MMN	0.20
Negative Symptoms (SANS)	P3a	0.17
Global Functioning (GAF)	RON	0.24
Functional Capacity (UPSA)	P3a	0.13



Cntrl

SZ



Should every subject be included in every cluster?

Not all subjects contribute components to each cluster.

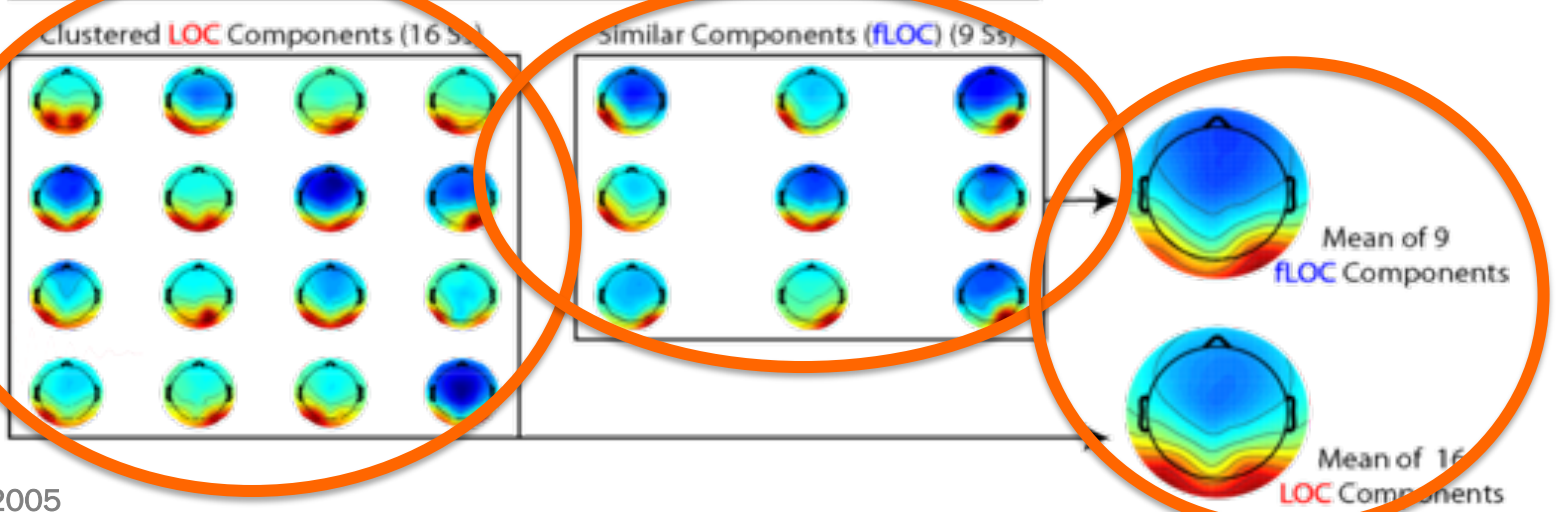
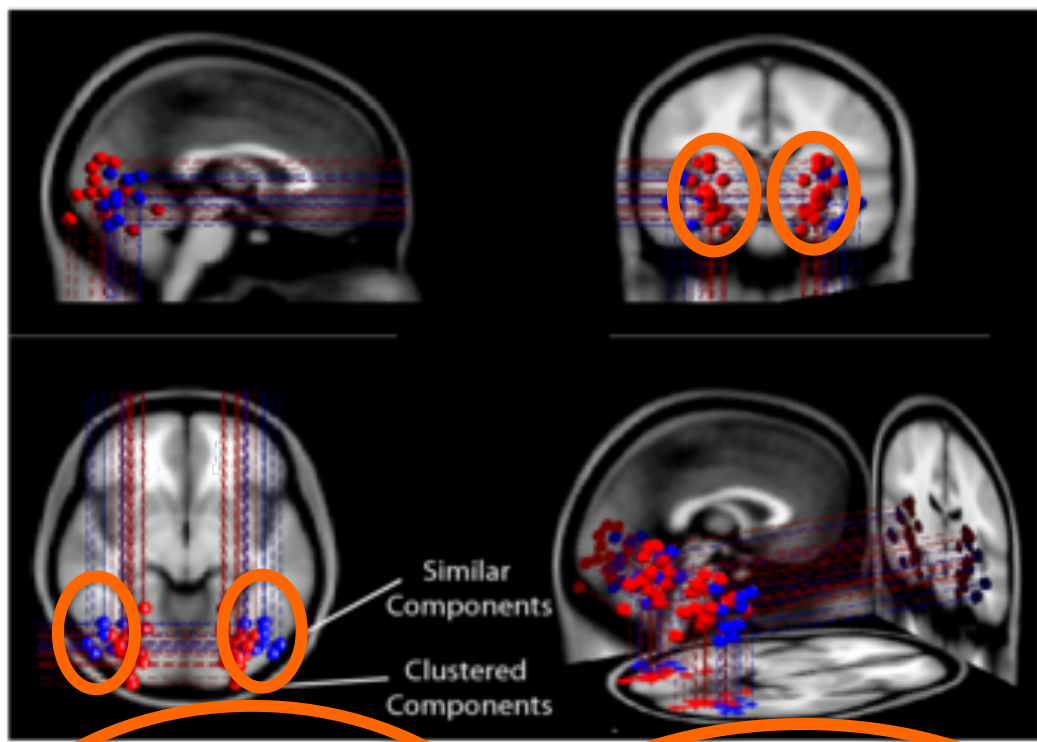
Why not?

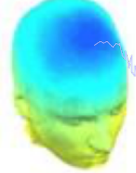
- Different numbers of artifact components
- Subject differences!?
- Does my subject group really exhibit a Gaussian cloud of individual differences around 'a mean subject' in 'subject space' ??





Why aren't all participants in every IC cluster?

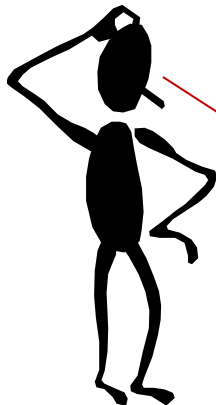




STUDY IC Clustering: Practical Problems

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()

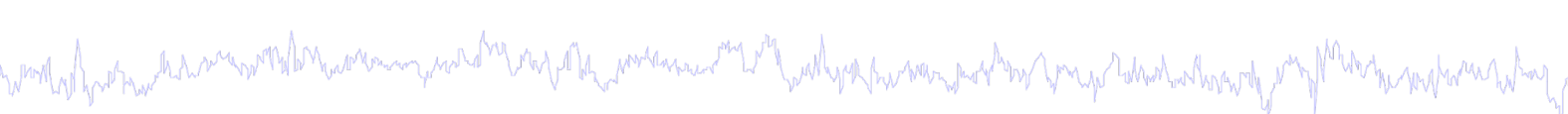
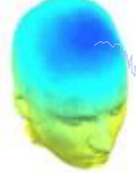
Pre-compute measures on which to cluster components from study 'N400STUDY'
 Select the cluster to refine during sub-clustering (any existing sub-hierarchy will be overwritten)

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	<input type="button" value="Help"/>	

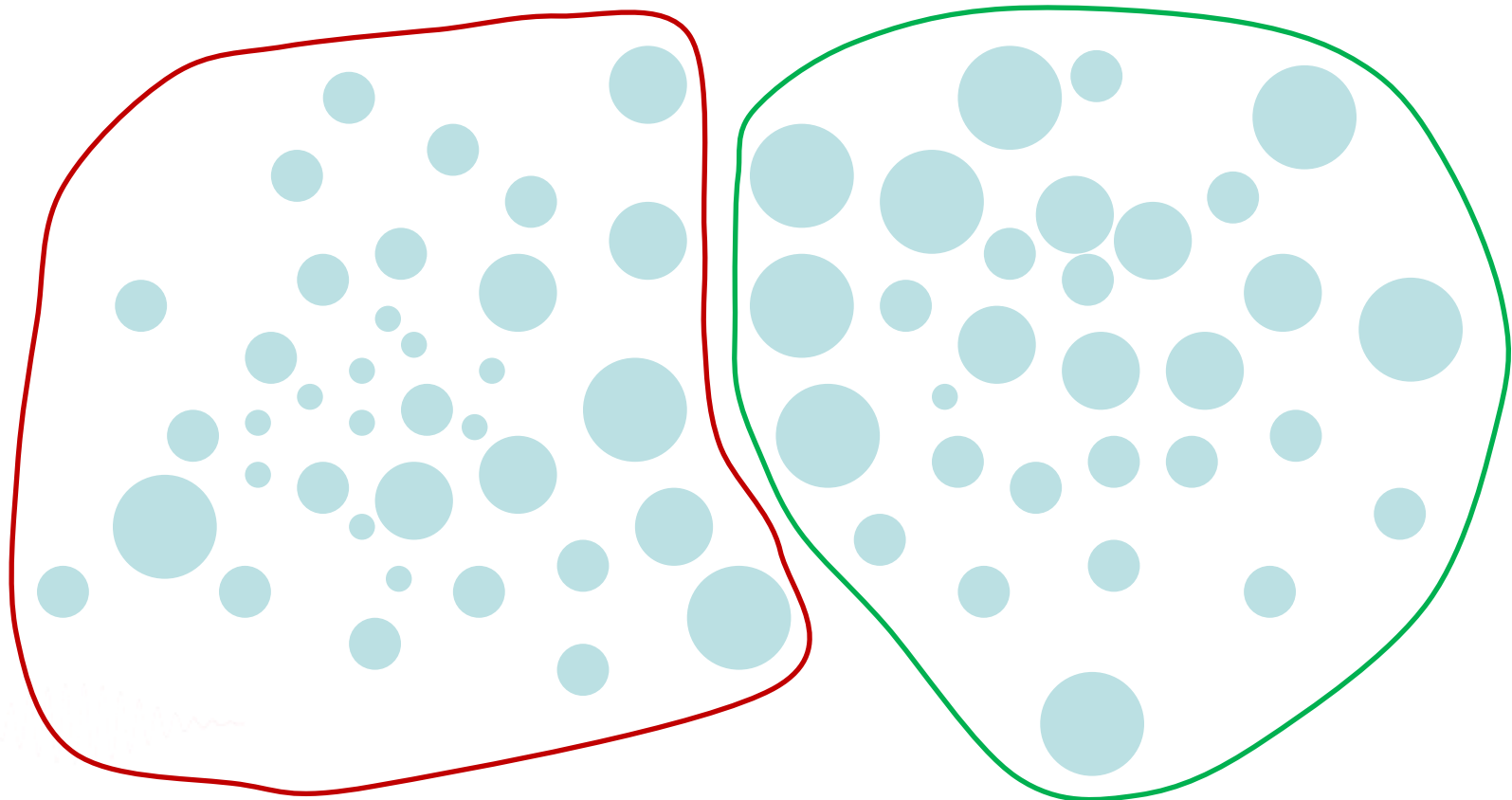
Frequency range [Hz]: 3 25
 Latency range in ms [lo hi]: -2100 1995
 Absolute values:
 Time/freq. parameters: 'p', [3 25], 'cycles', [3 0.5], 'pa'
 Time/freq. parameters: 'p', [3 25], 'cycles', [3 0.5], 'pa'

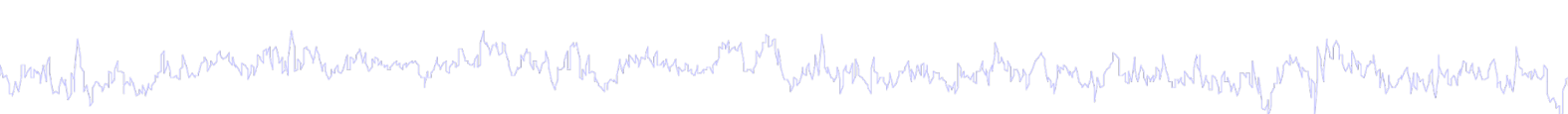
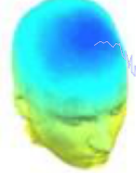
Save STUDY to file: /data/boiman4/ami/5/subjects/N400preclust.study



Problems with multi-measure clustering

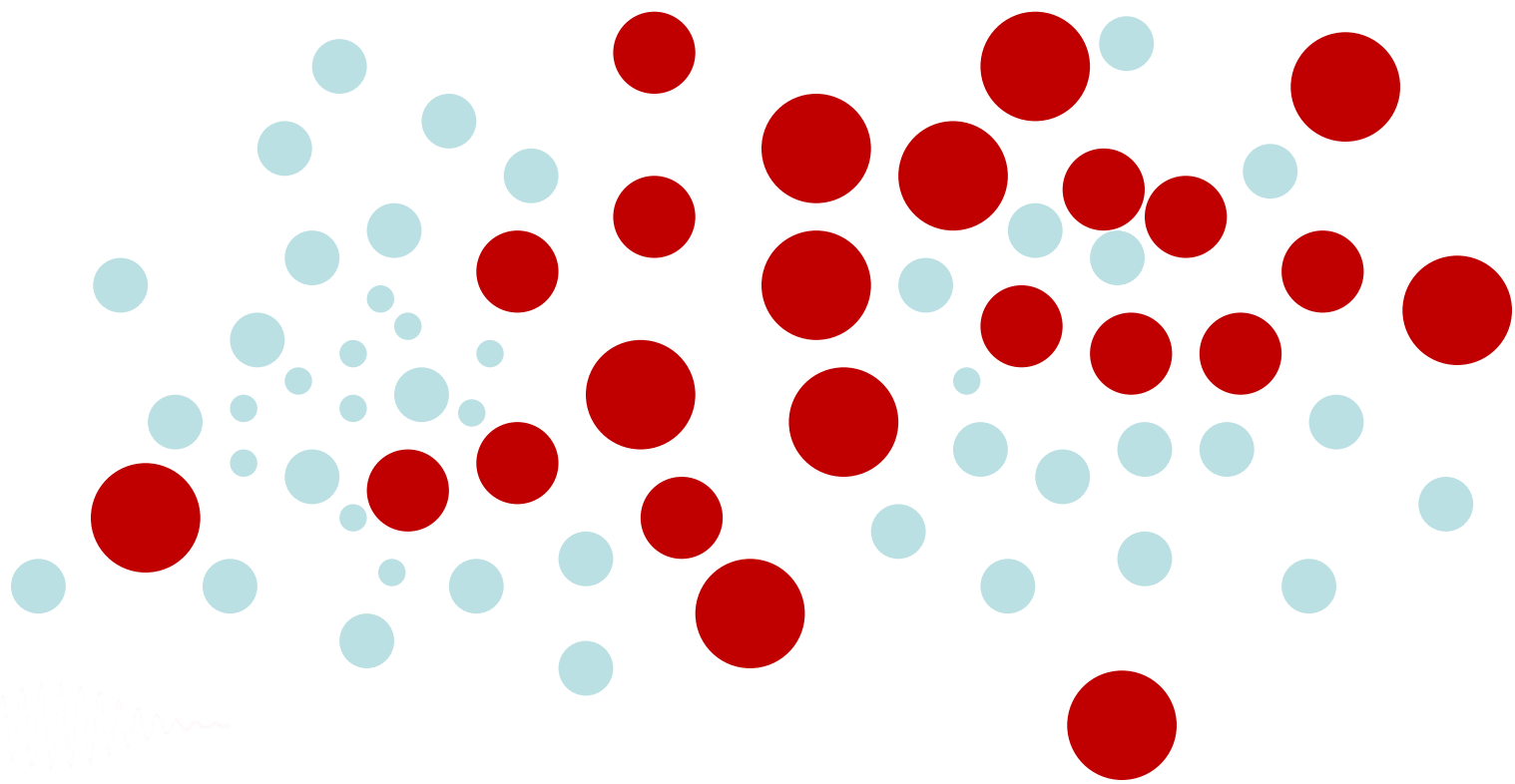
What are the clusters according to location?

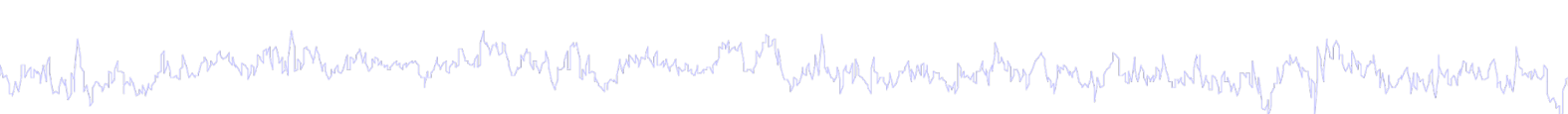
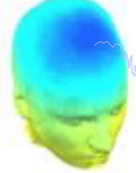




Problems with multi-measure clustering

What are the clusters according to size ?

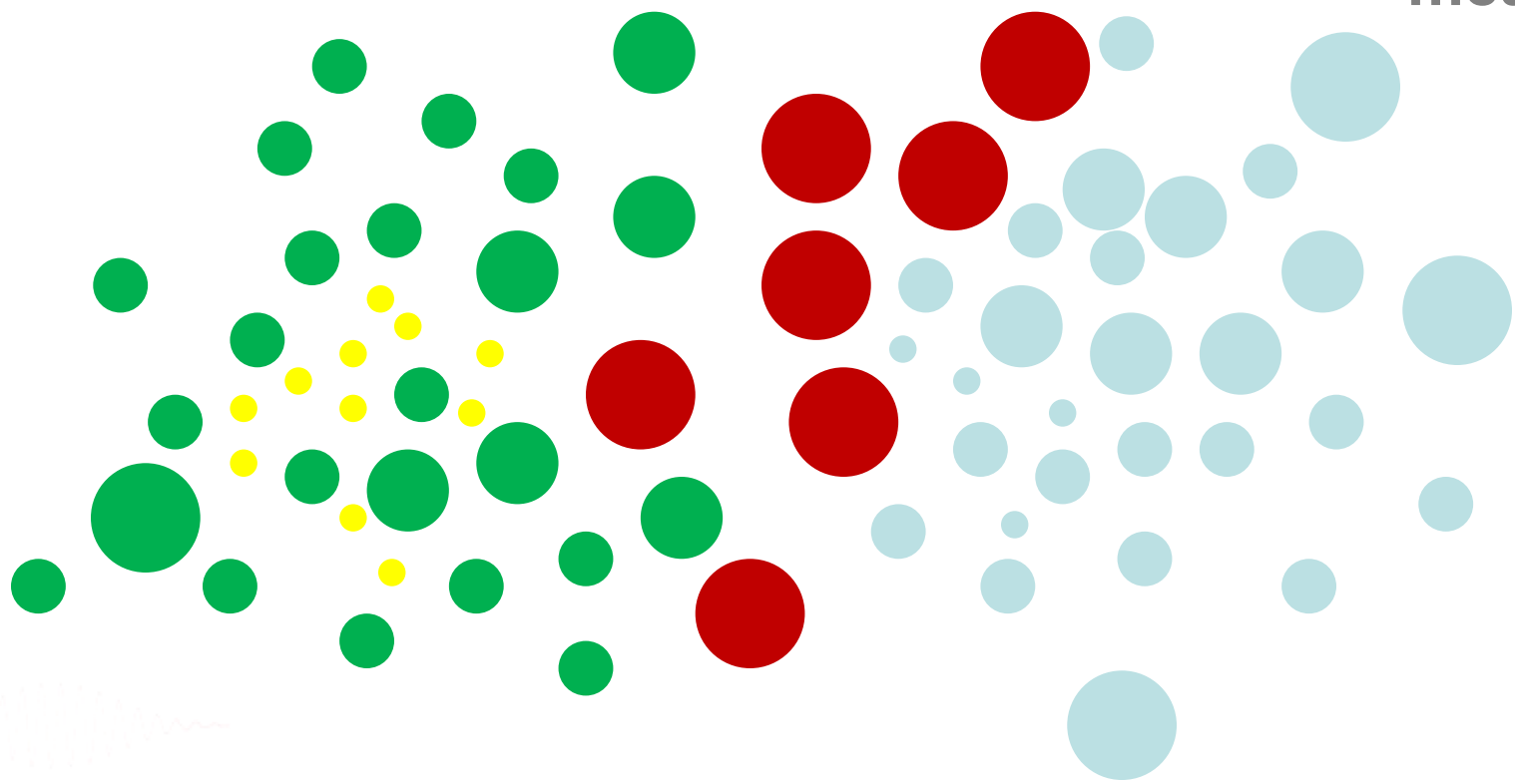


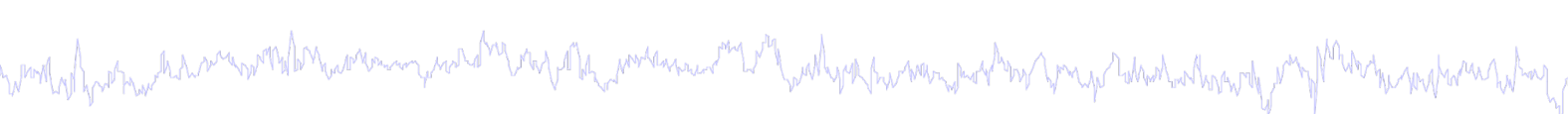
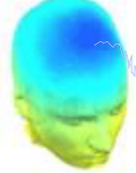


Problems with multi-measure clustering

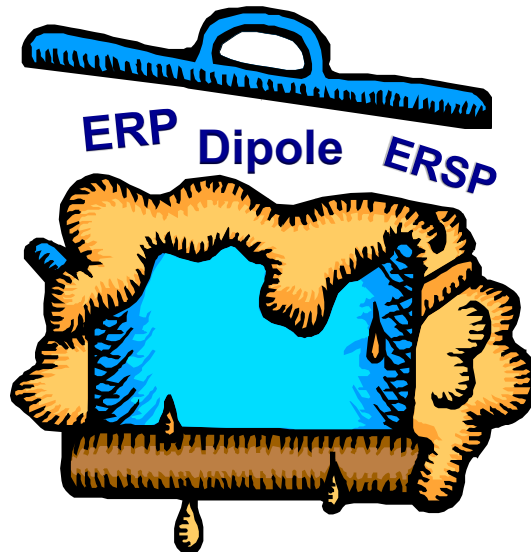
What are the clusters according to location and size?

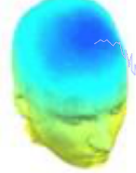
Well, it depends on how much weight we give each measure...



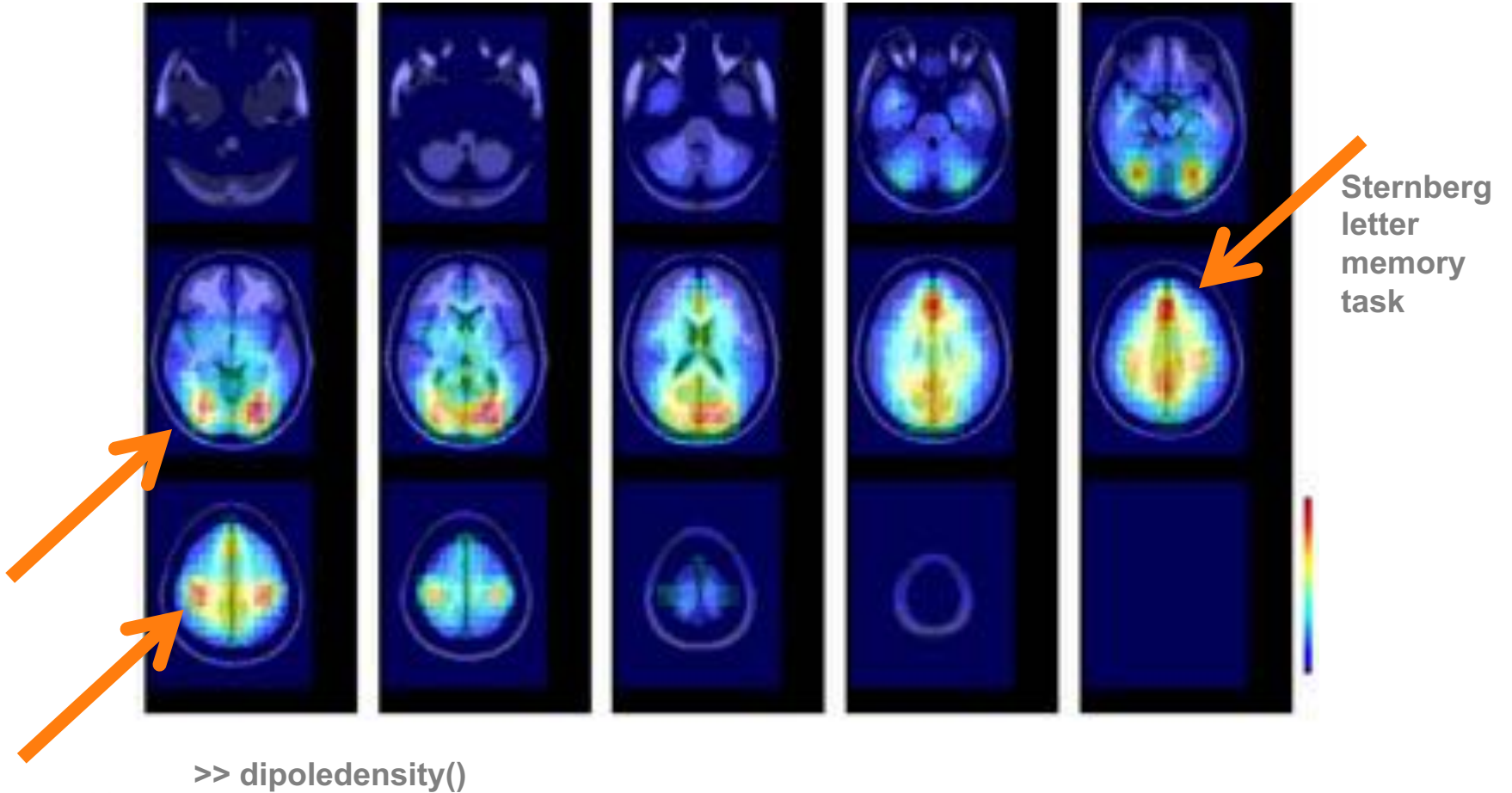


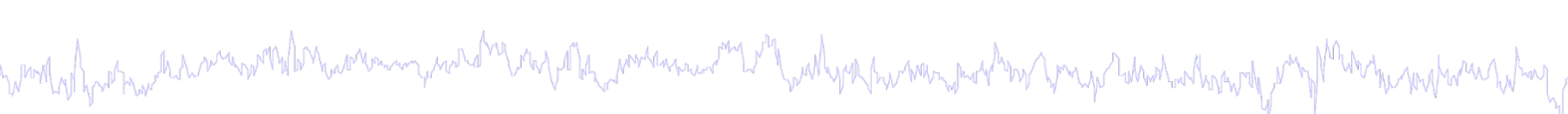
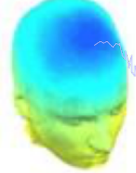
- With either clustering method, we basically mix together distances for a subset of EEG measures (ERP, ERSP, ITC, mean spectrum, dipole location).
- This may make clustering distance less interpretable.





Equivalent dipole density



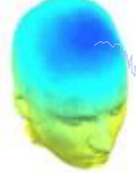


Study IC Clustering by Measure Projection

- **Instead**, we can directly work on pair-wise similarity matrices and prevent ICs with similarities less than certain threshold (e.g., ERSP corr. < 0.5) to be clustered together.
- The most important measure is **equivalent dipole location**.
- Assuming a certain variability estimate for dipole location (due to error in localization and subject variability), one can also estimate an optimum number of clusters.

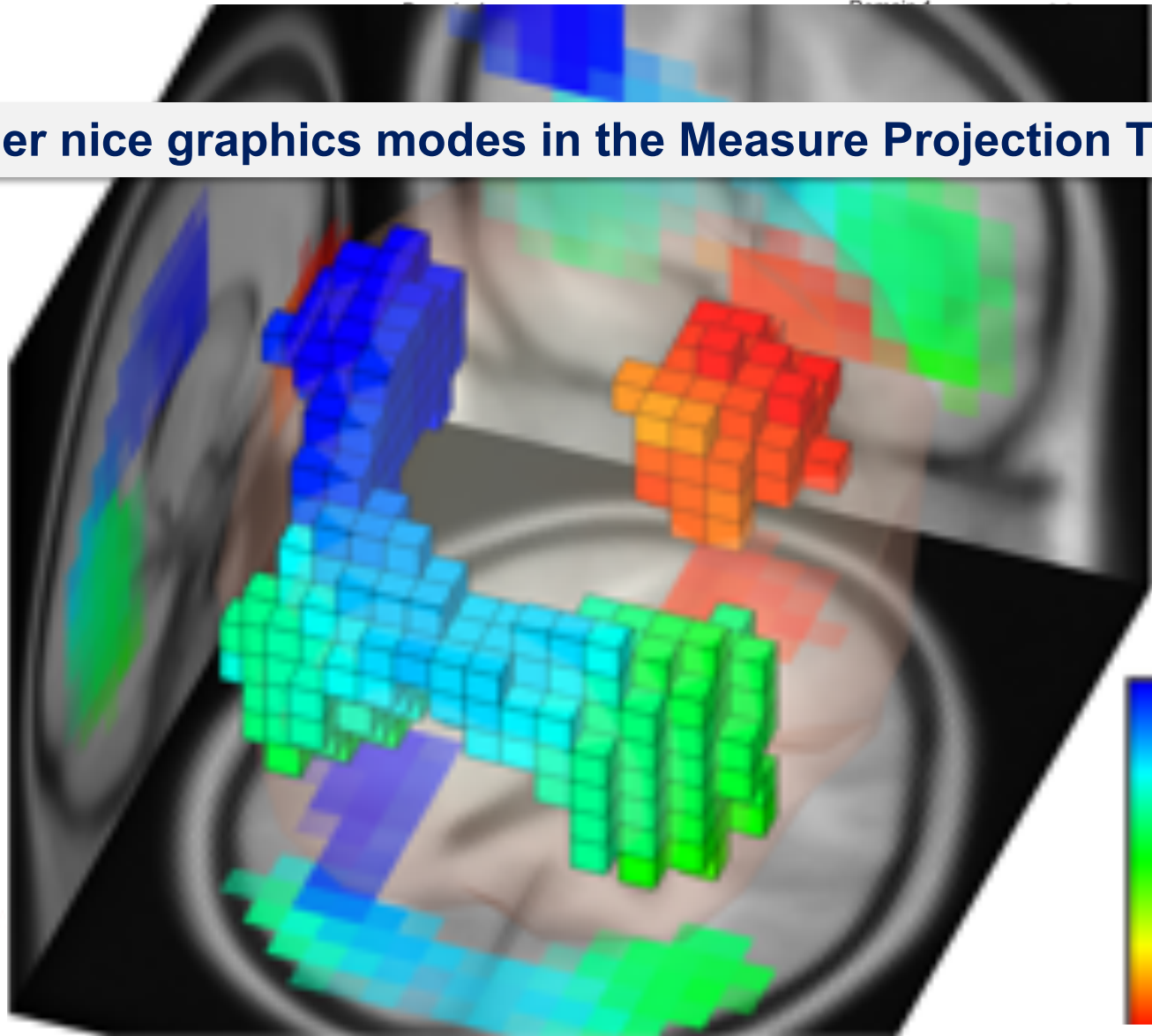
Measure Projection asks:

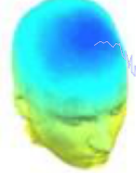
1. Where in 'template brain space' does our data have evidence that **our one measure of interest** is consistent across nearby ICs?
2. Which such brain space voxel *domains* show consistent differences?



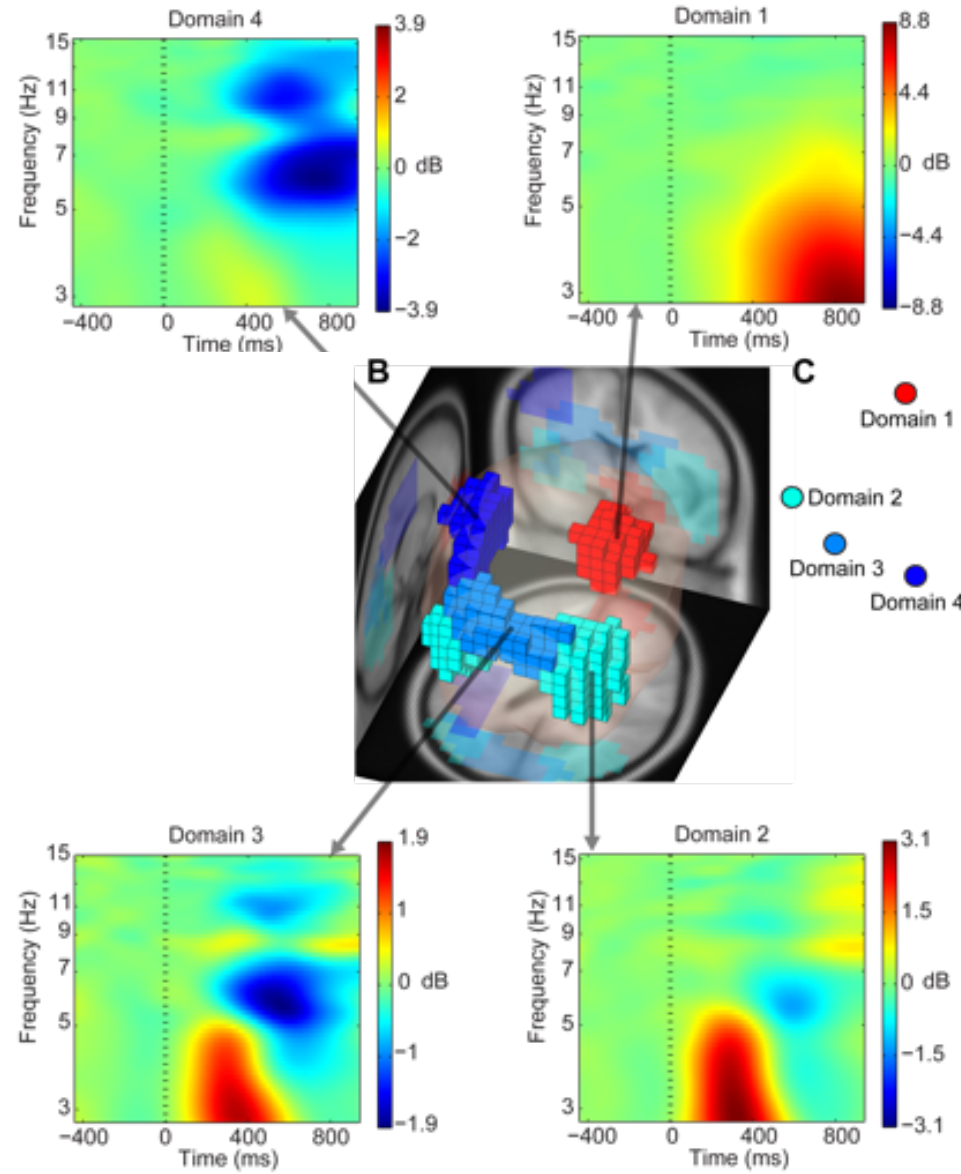
Measure Projection: RSVP Task Example

Other nice graphics modes in the Measure Projection Toolbox



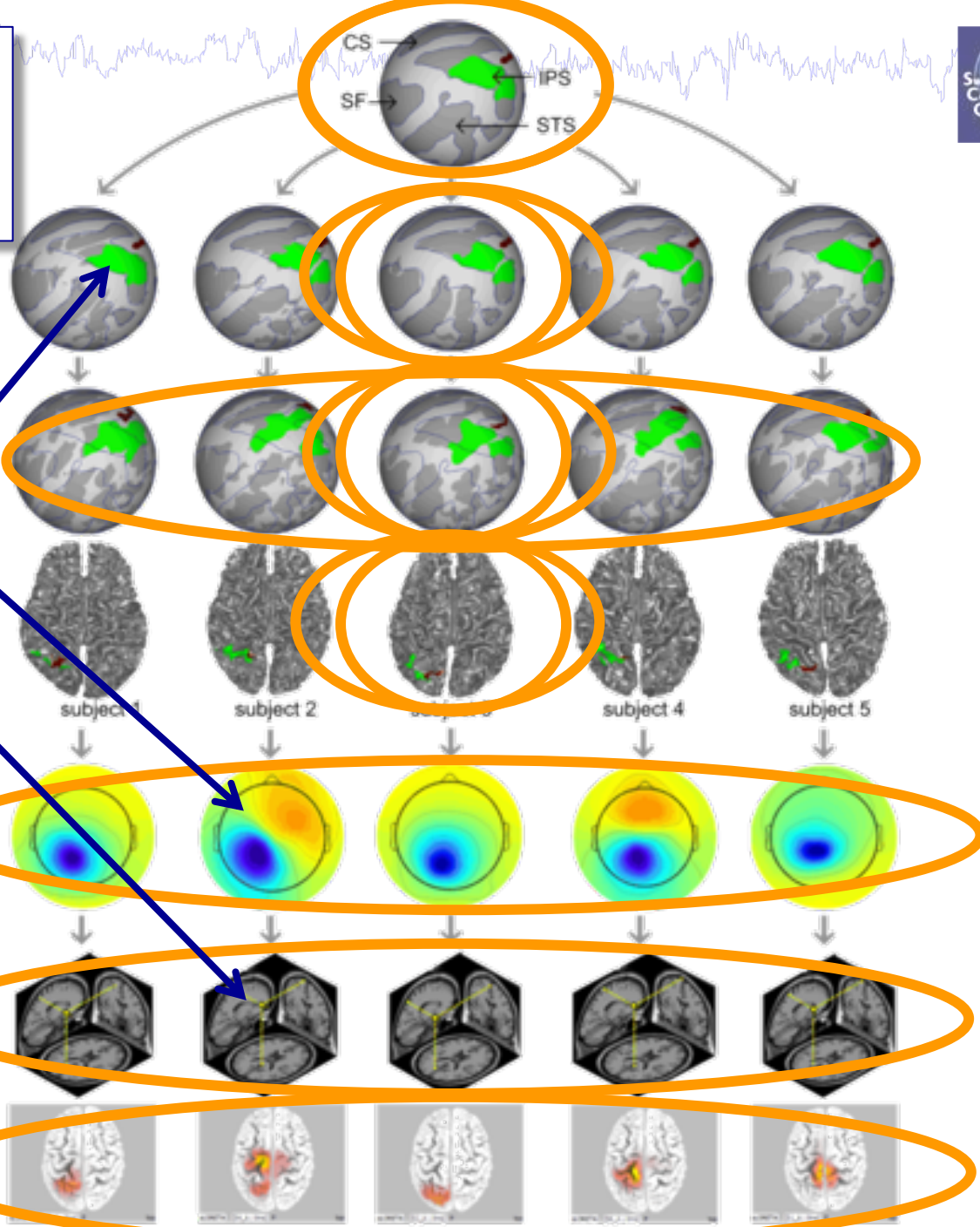


Measure Projection: RSVP Task Example



ERSP Dissimilarity

Arthur Tsai – Topological source clustering



Why should IC clusters
have breadth?

Equivalent cortical areas

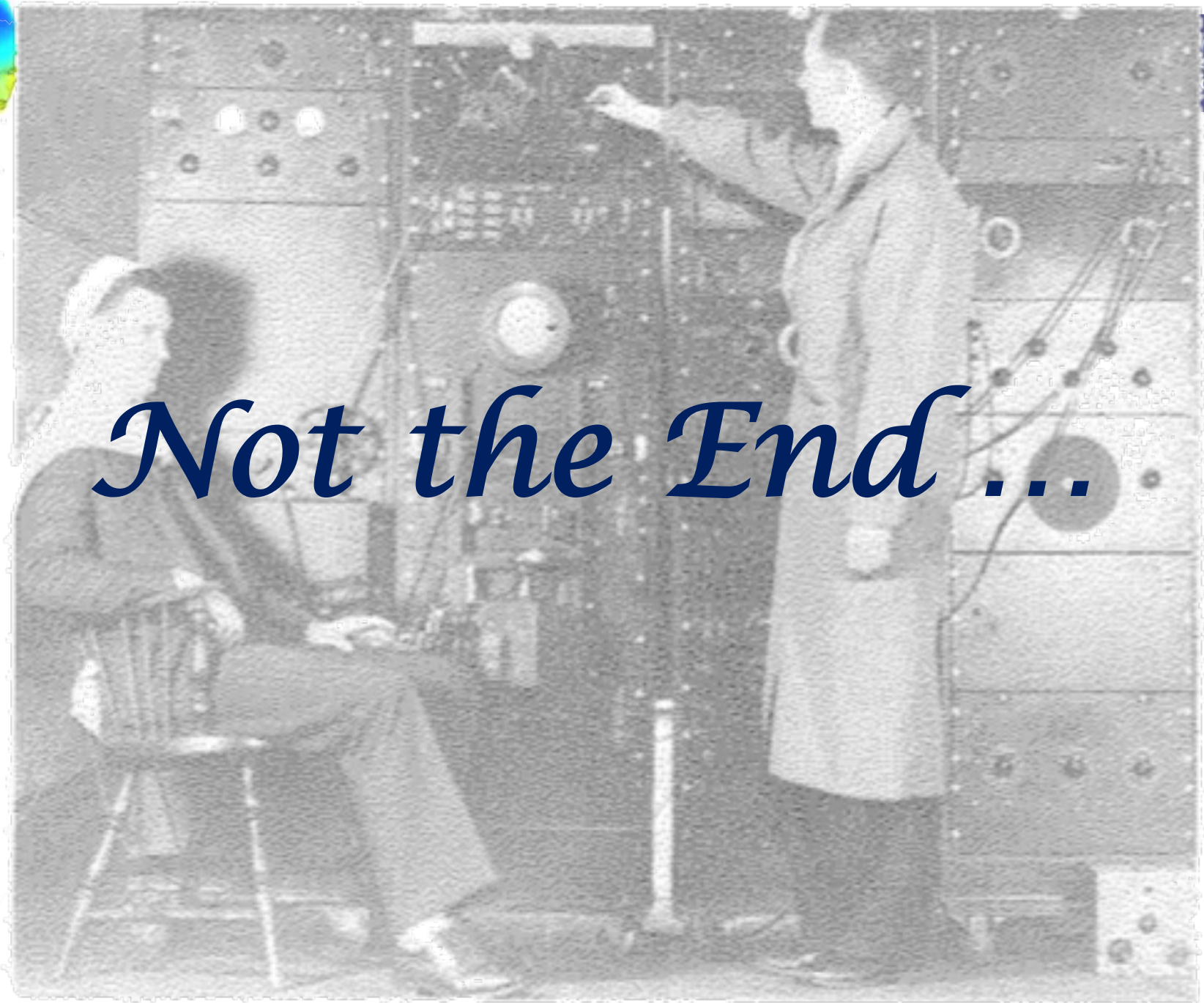
Have different scalp maps

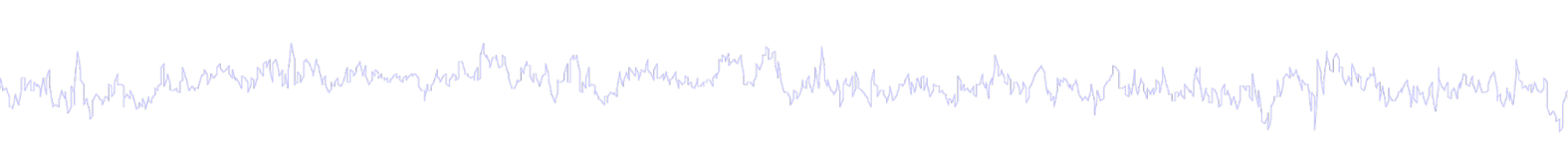
And dipole locations!





Not the End ...



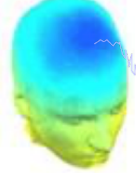


**Does the spatial distribution of IC
equivalent dipole source locations
depend on the task the subject
performs?**

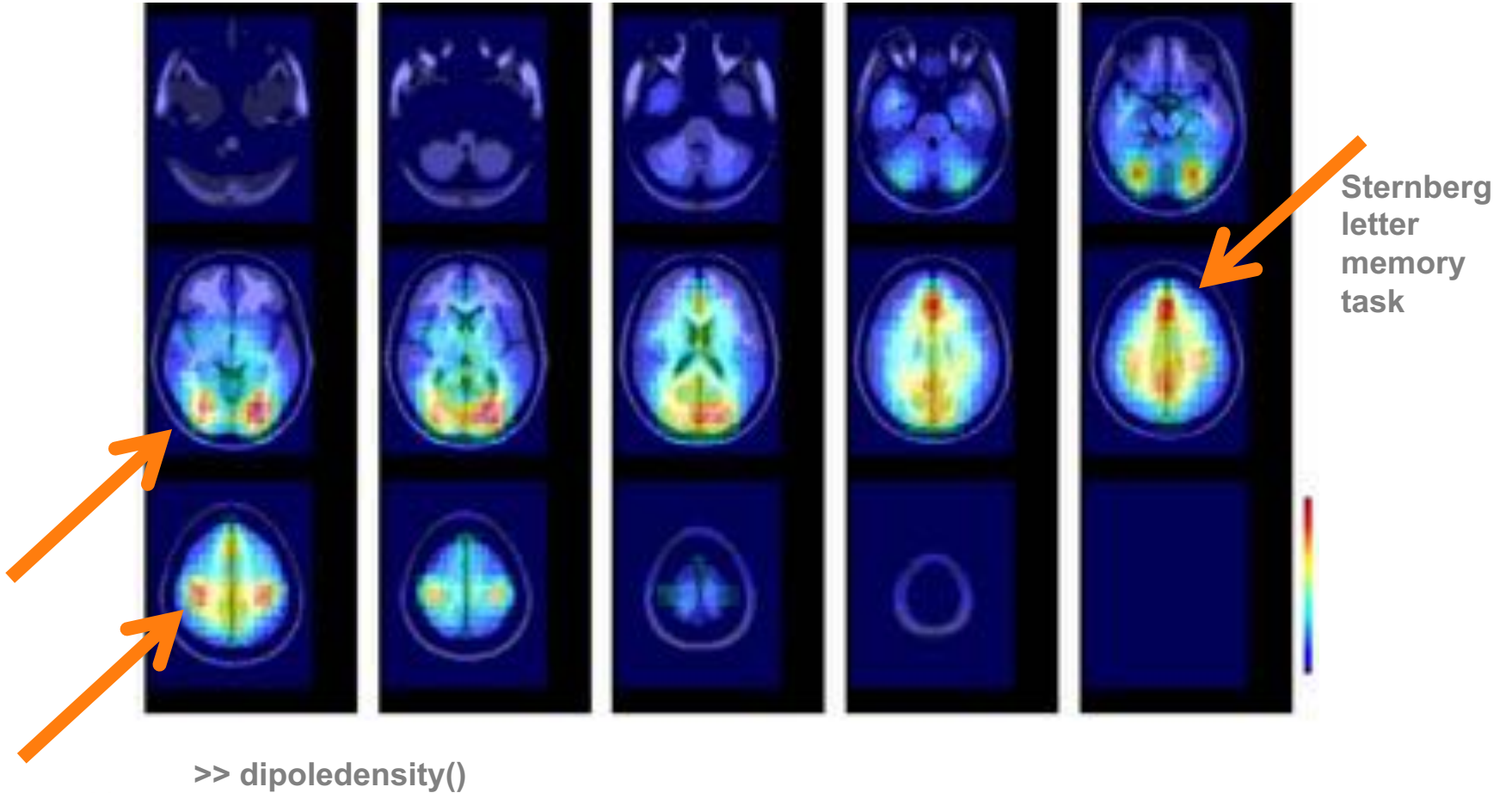
i.e.

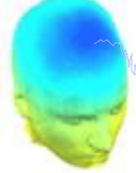
**Do “the same” ICs (and IC clusters)
appear for every task?**



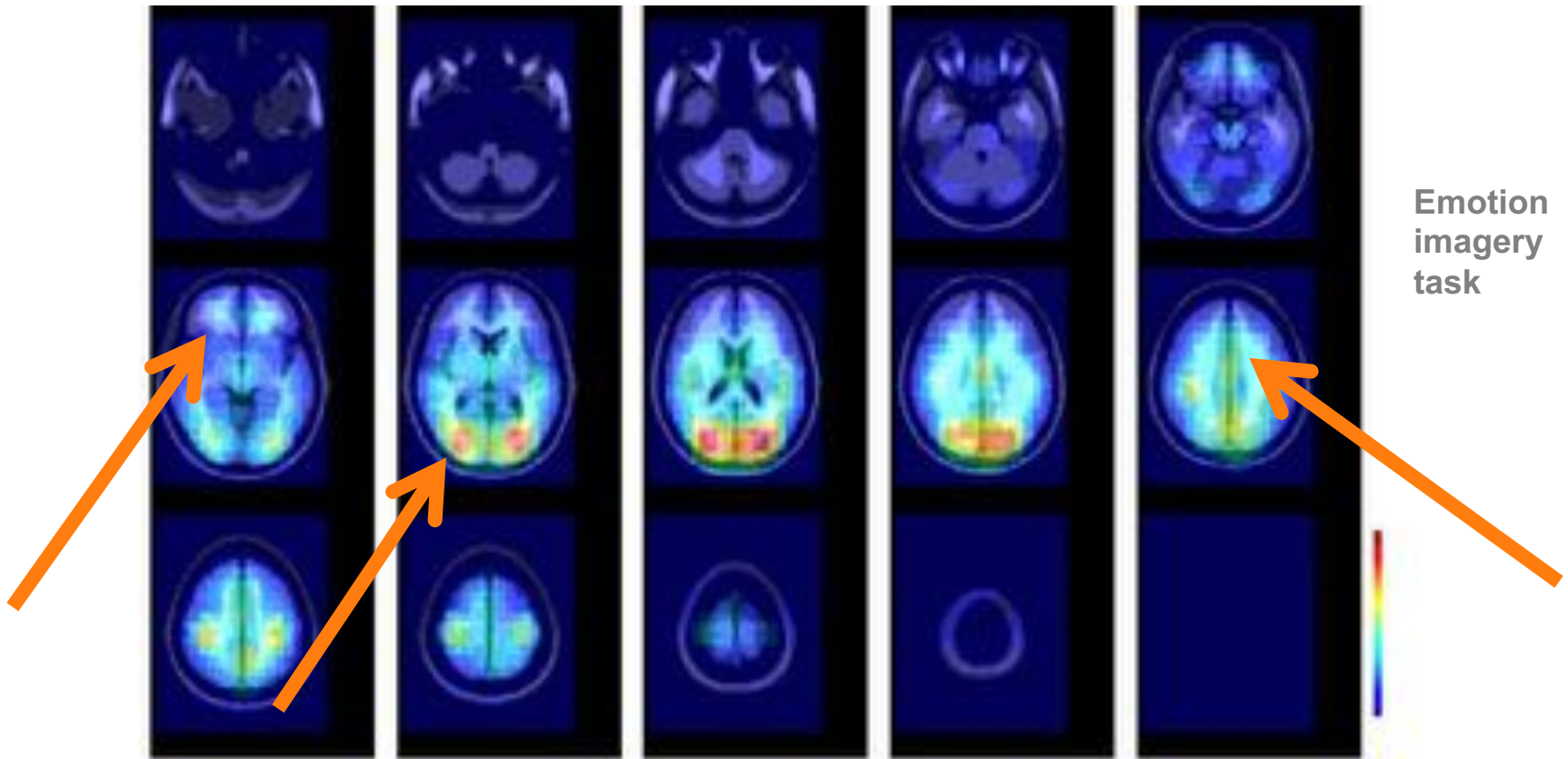


Equivalent dipole density



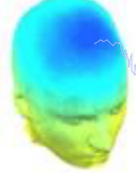


Equivalent dipole density

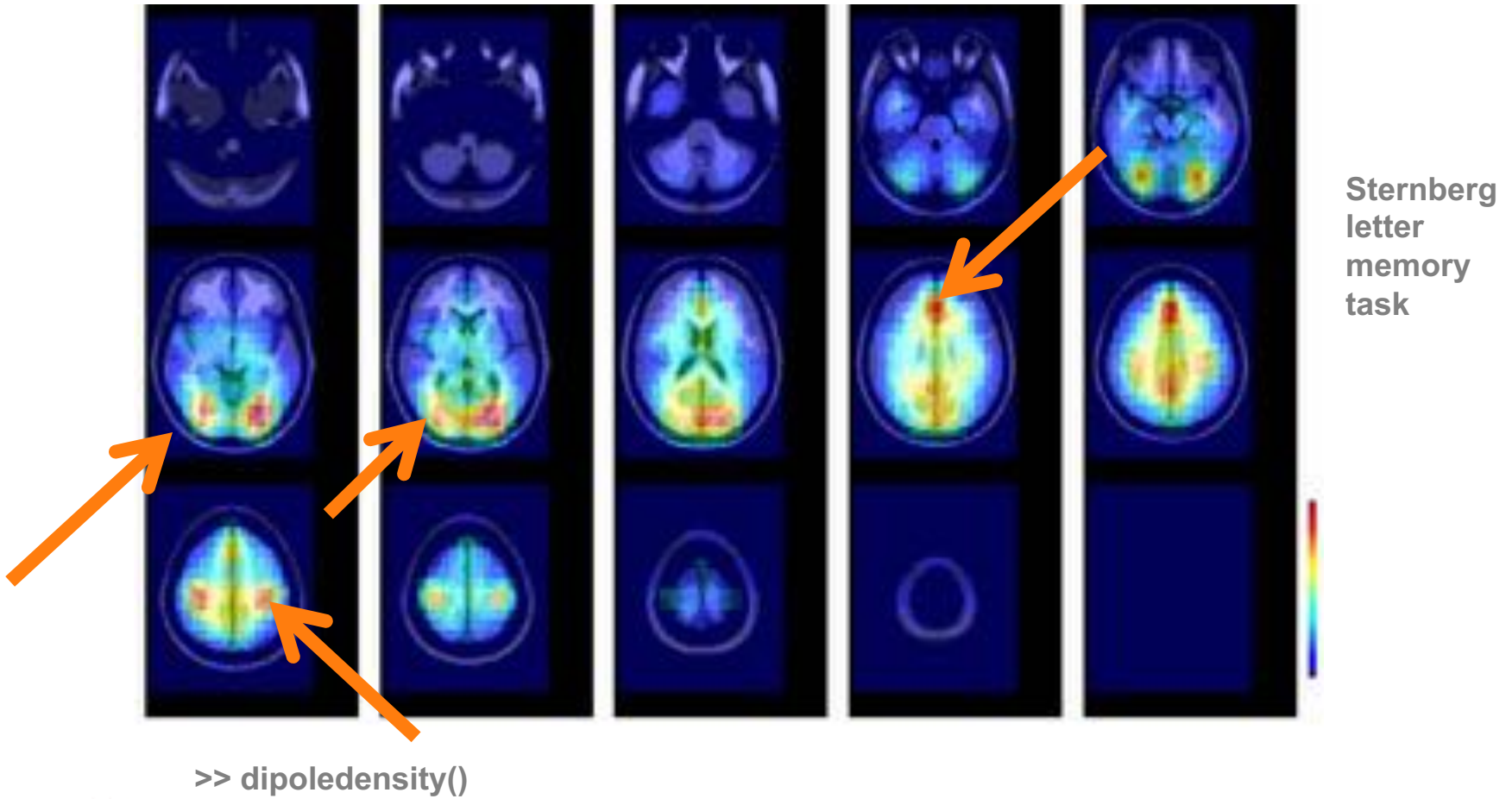


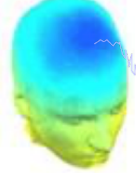
>> dipoledensity()



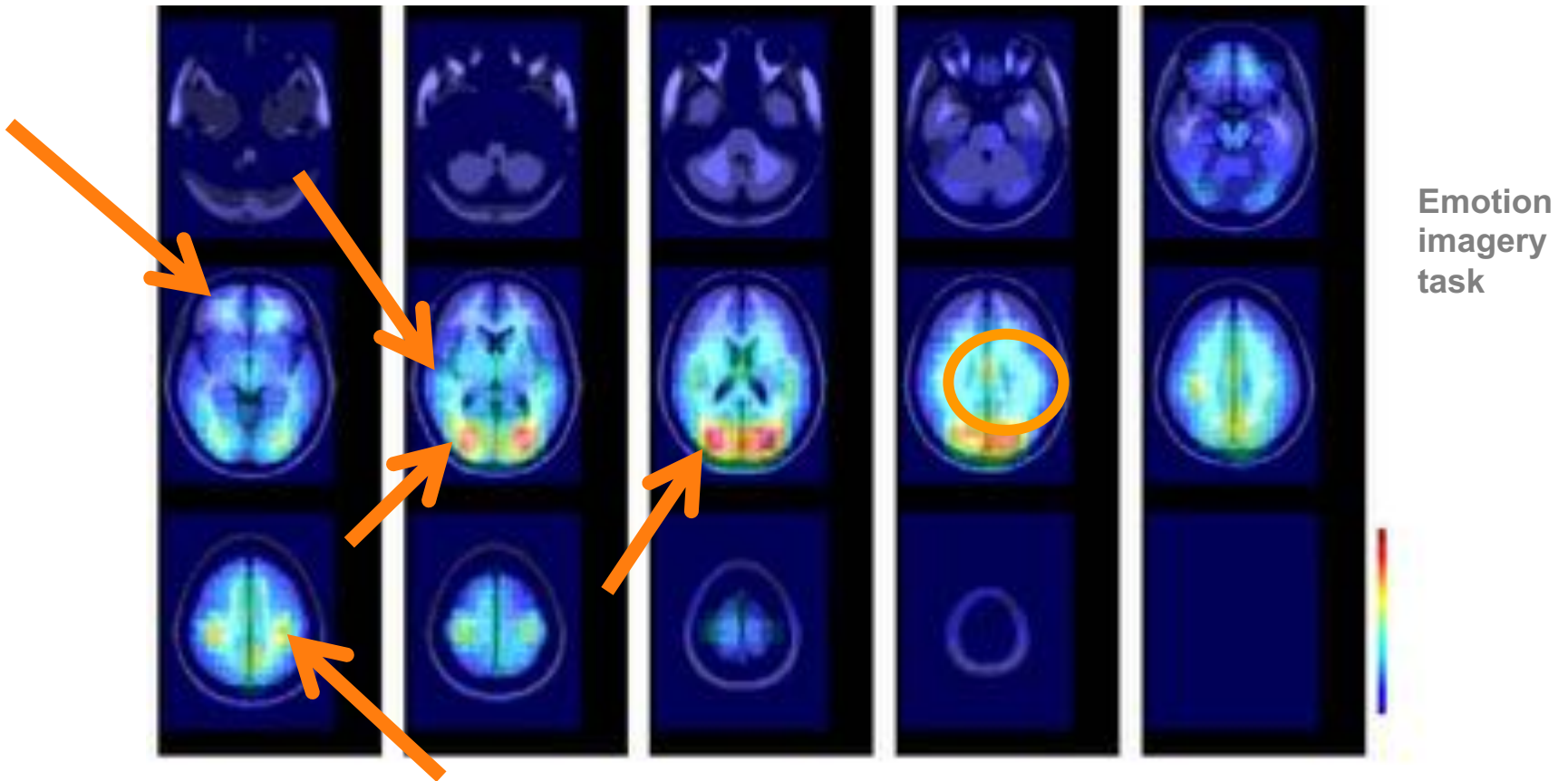


Equivalent dipole density





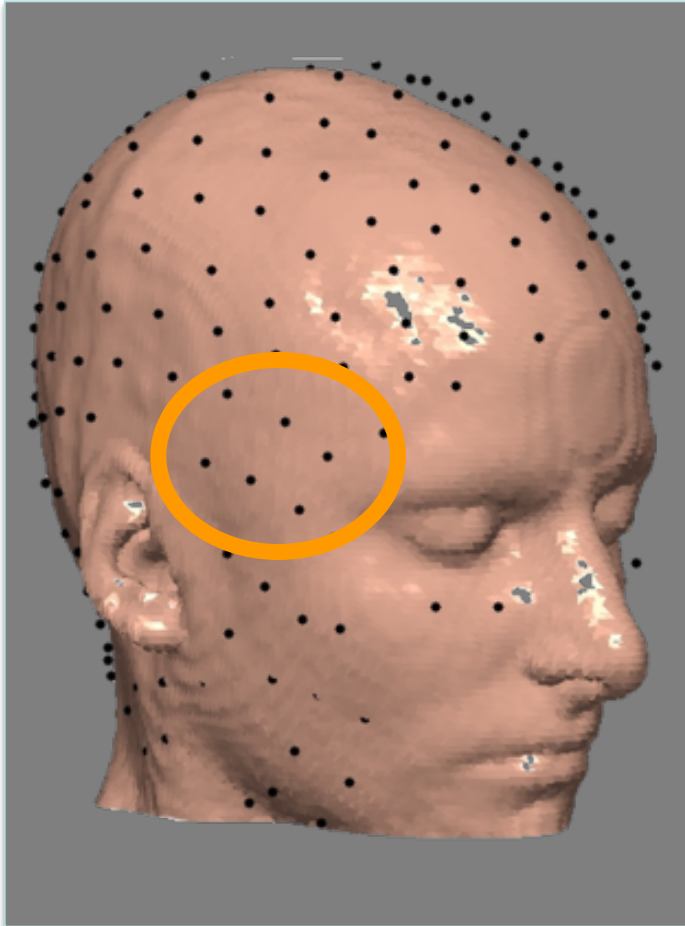
Equivalent dipole density



>> dipoledensity()



Co-Registration of Electrodes with MR Image



MR + EEG

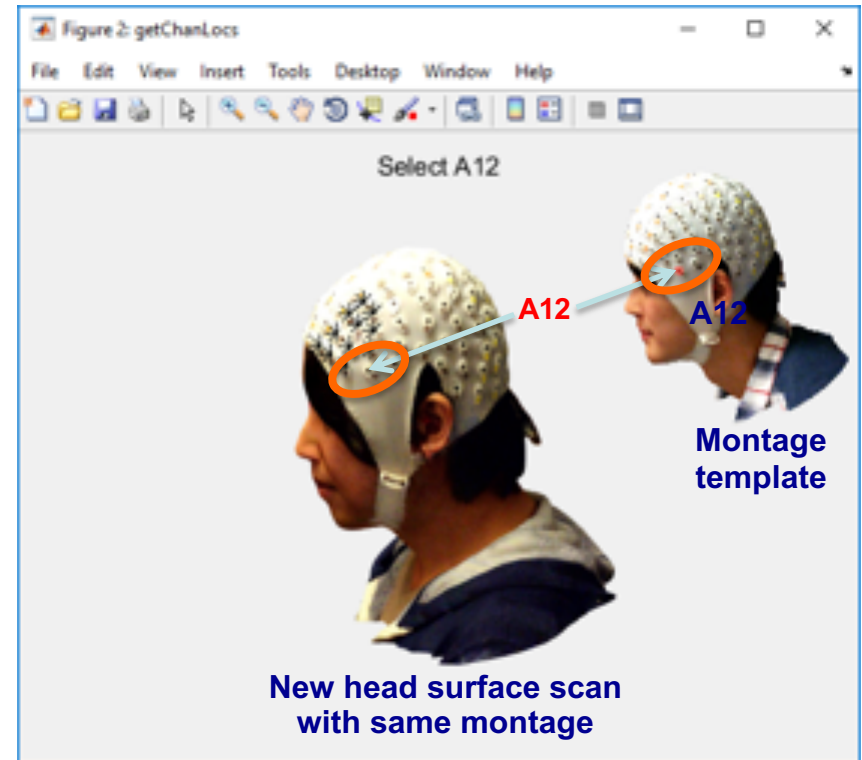


EEG



Handheld 3-D electrode position recording

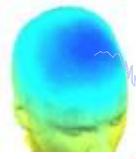
(‘electrode digitizing’)



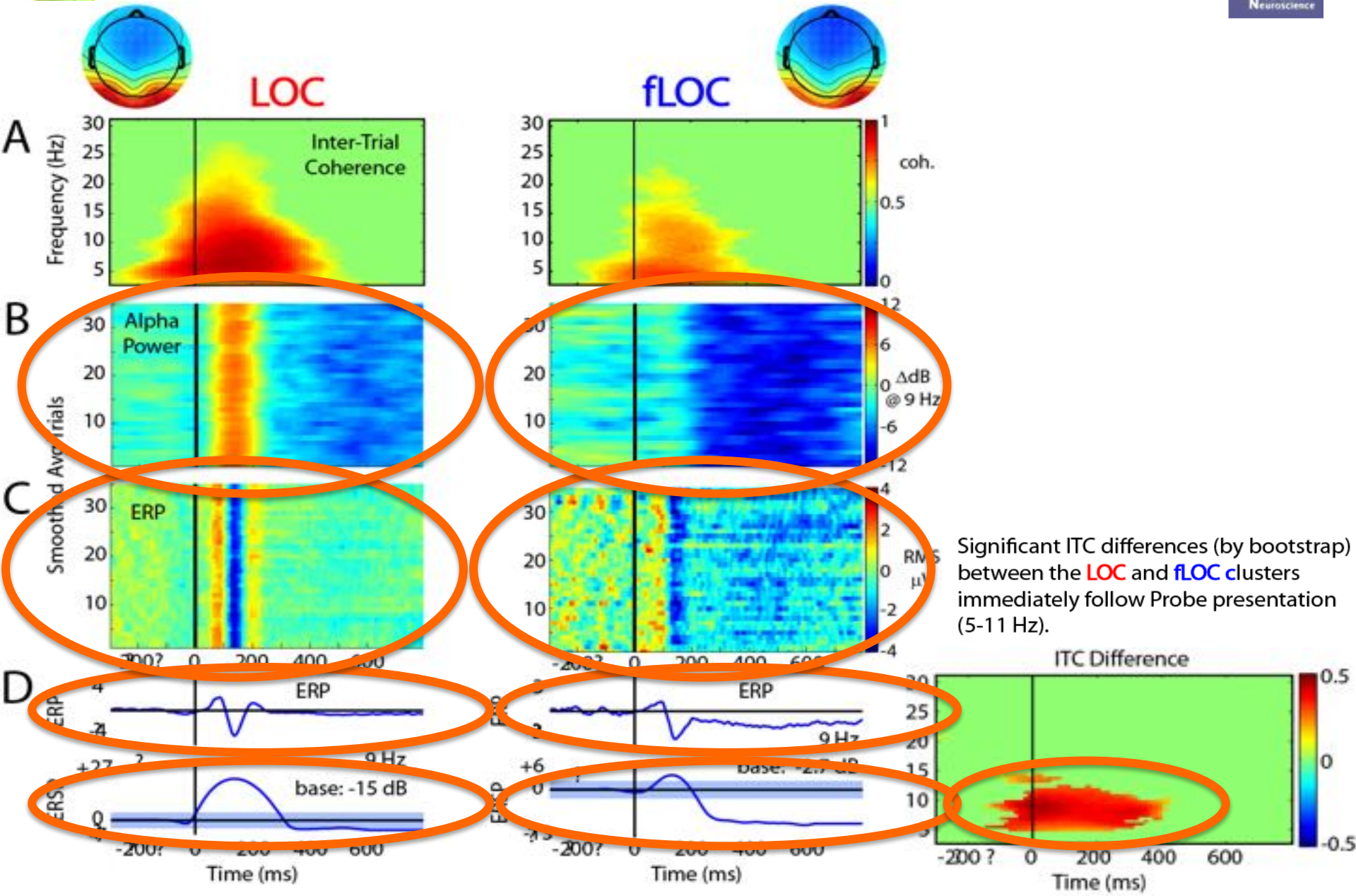
Clement Lee & S. Makeig, 2018

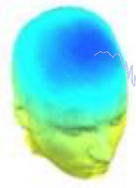
The `get_chanlocs()` GUI –
post hoc 3-D electrode
location determination from
a 3-D head image.

Cl. Lee & S. Makeig, 2014

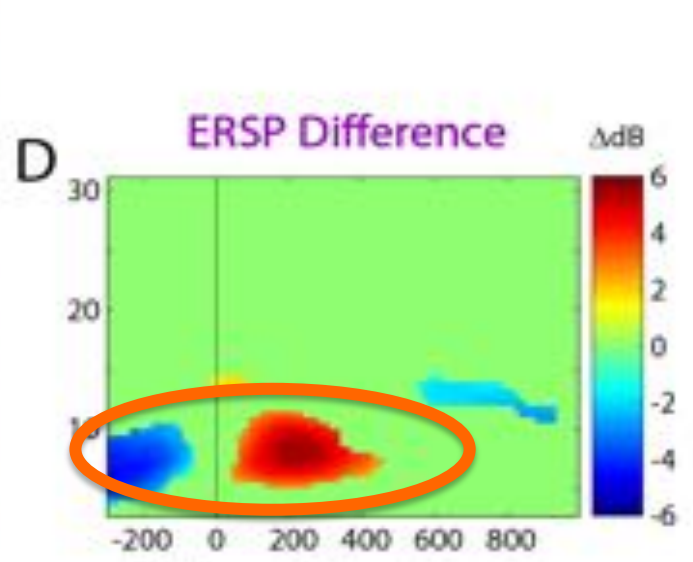
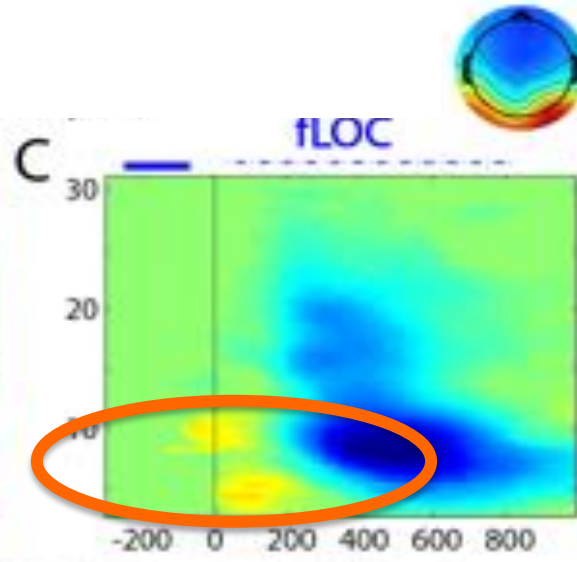
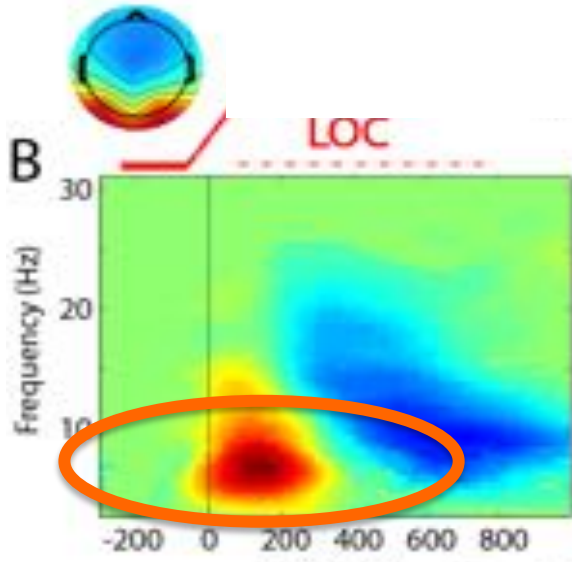
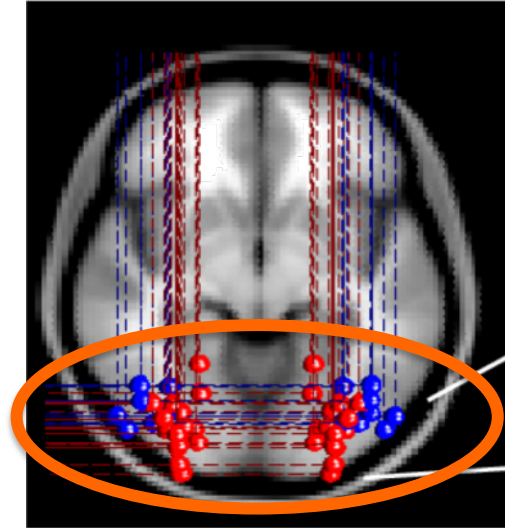


Subject differences?



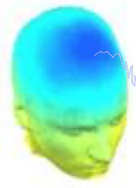


Subject differences?

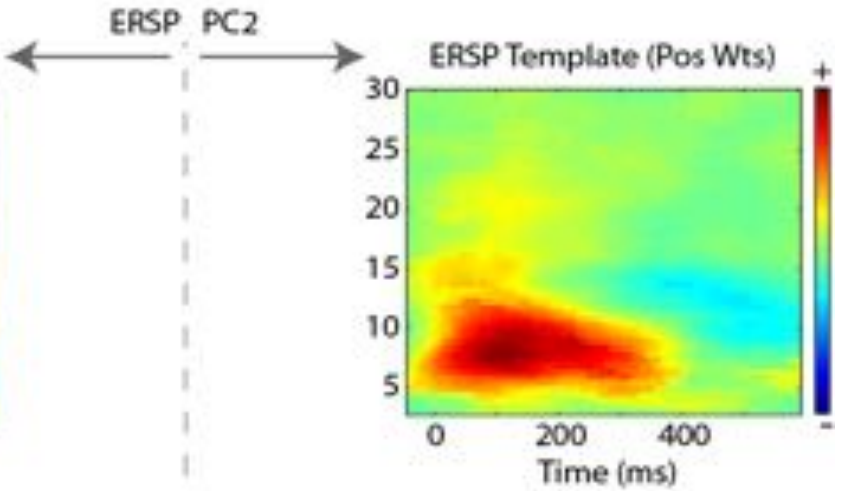
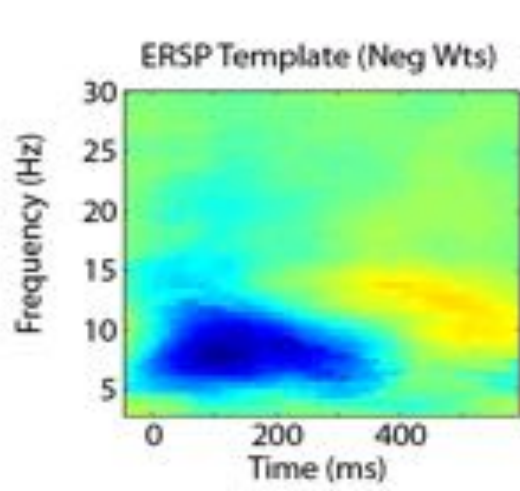
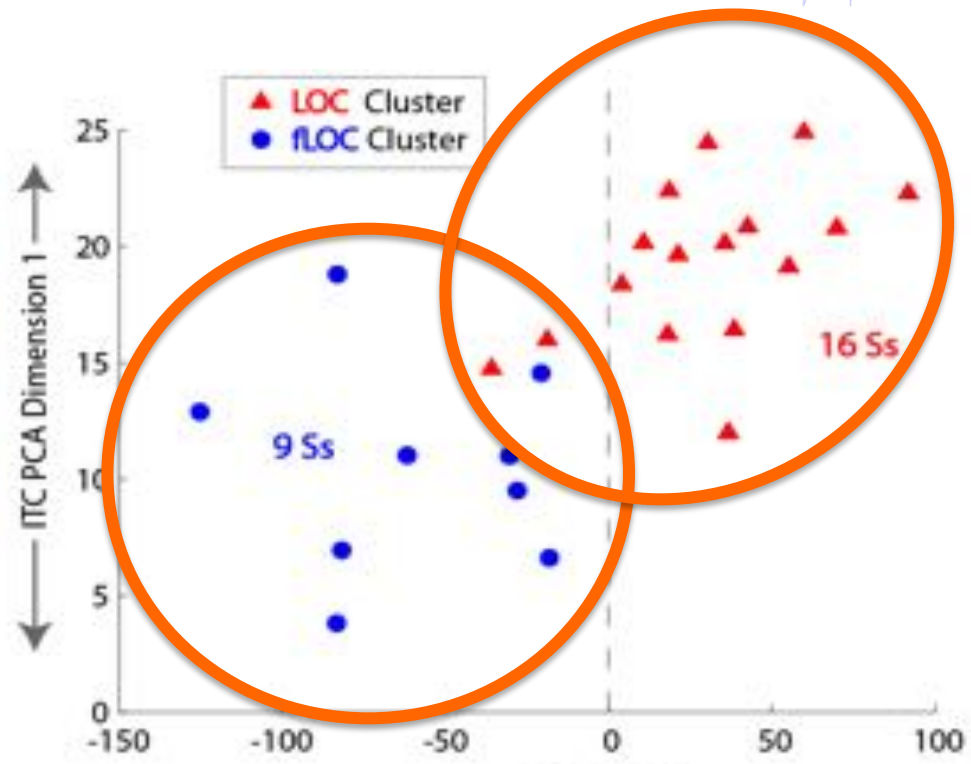
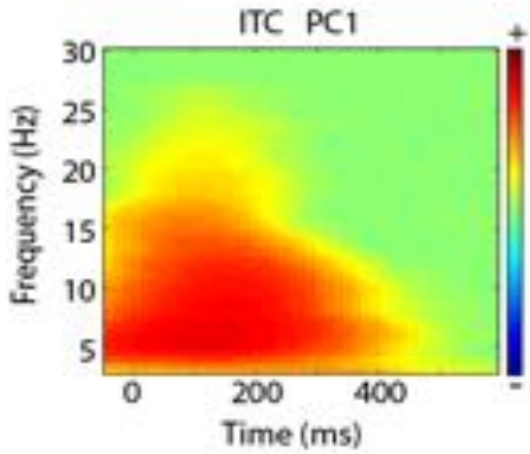


** Cluster ERSPs show significant activity determined by bootstrap statistics within subject and binomial probability between subjects ($p < 0.01$)

*** Difference ERSP shows significant differences between the two clusters by bootstrap statistics ($p < 0.001$)



Subject differences?



ERSP PC2