Clustering Independent Components of EEG Data

Scott Makeig
Institute for Neural Computation
University of California San Diego

26th EEGLAB Workshop

Beer Sheva, Israel

October, 2017

Why cluster independent components across subjects or sessions?

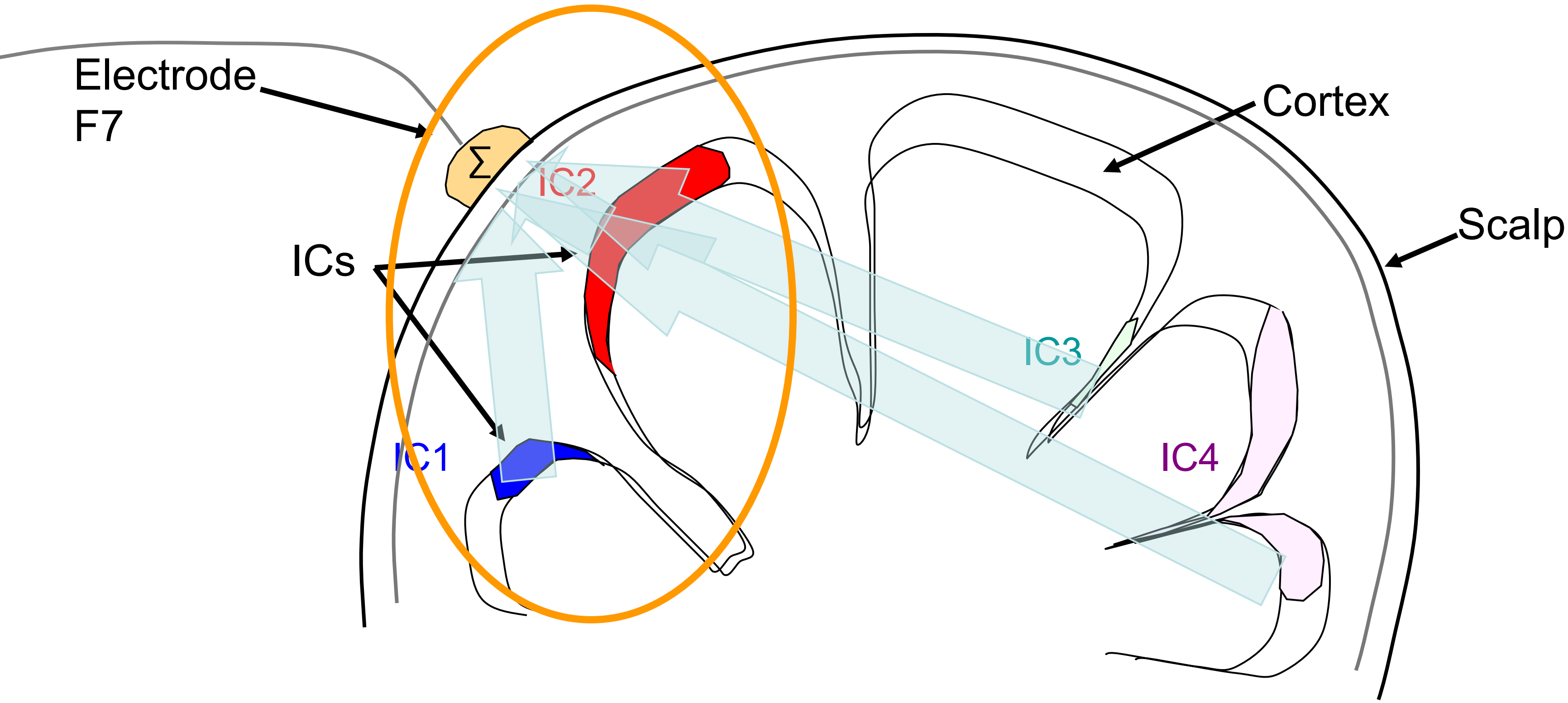
- ICA transforms the data from a channel basis (activity recorded at each channel)
    - to a component basis (activity computed at each IC).
- Normally, EEG researchers assume that, for example, electrode channel F7 == F7 == F7 ... in each subject – and then 'cluster' their data assuming channel equivalence.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

"Your Cz is My Cz!"

- But this is only ***roughly*** correct !

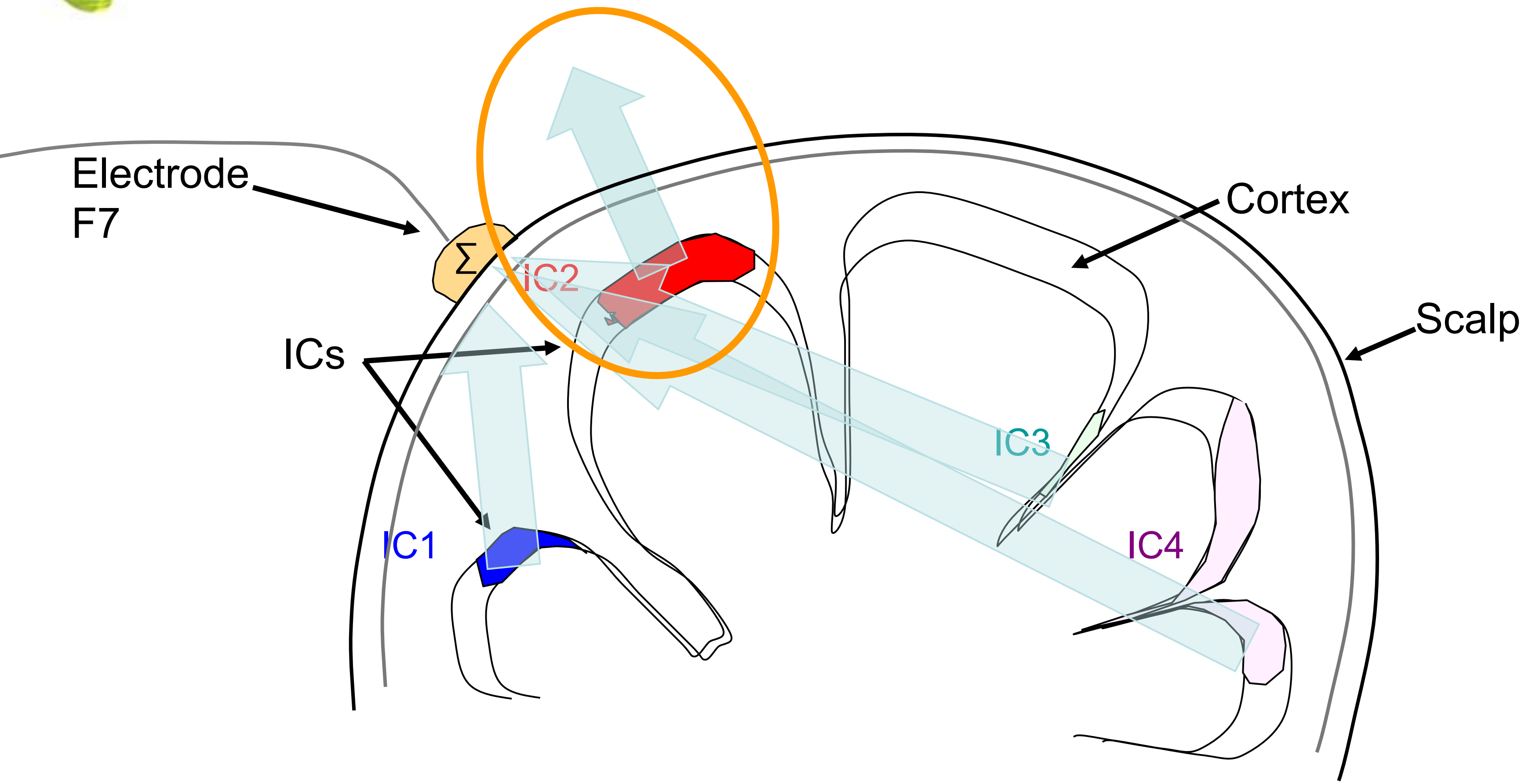
Makeig, 2007

Example: First Subject



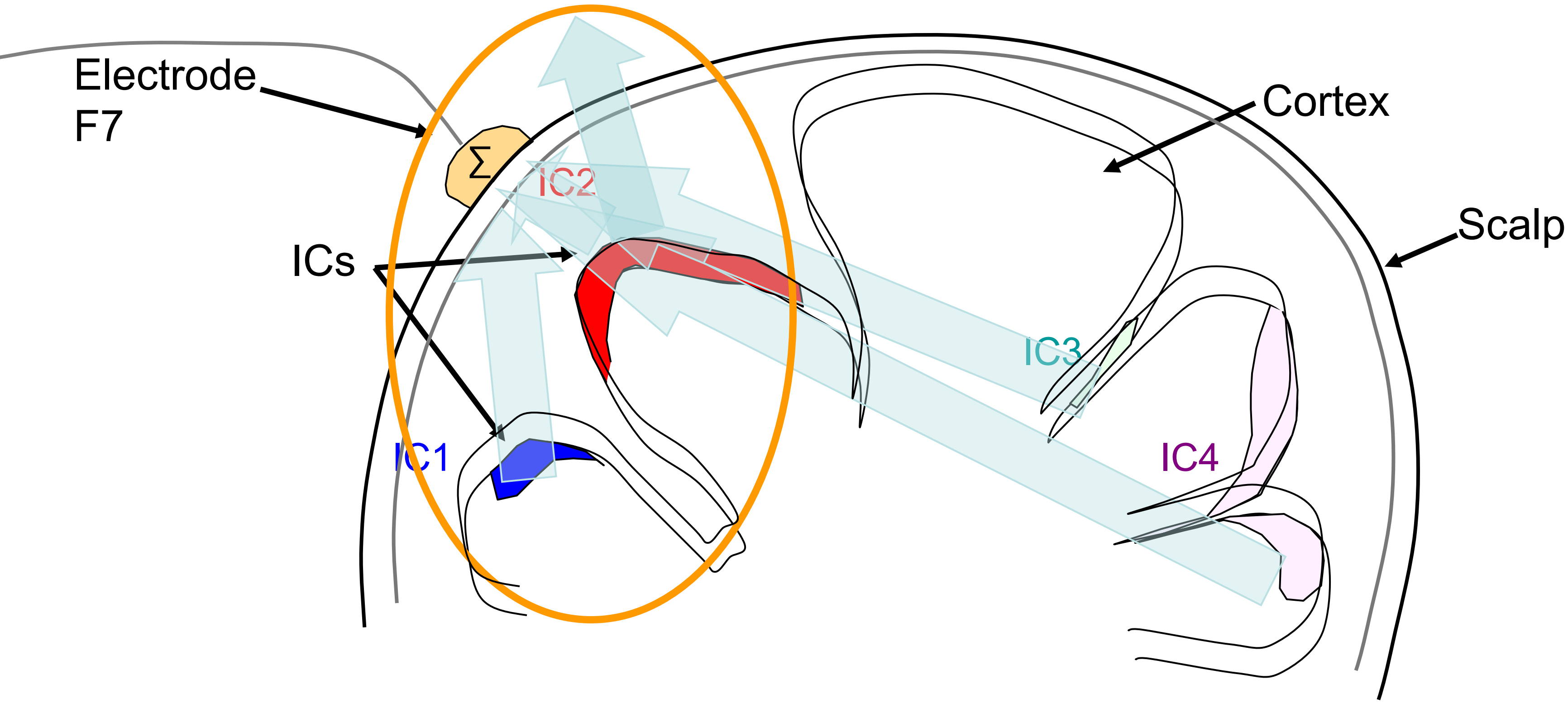
Makeig, 2005

Seond Subject



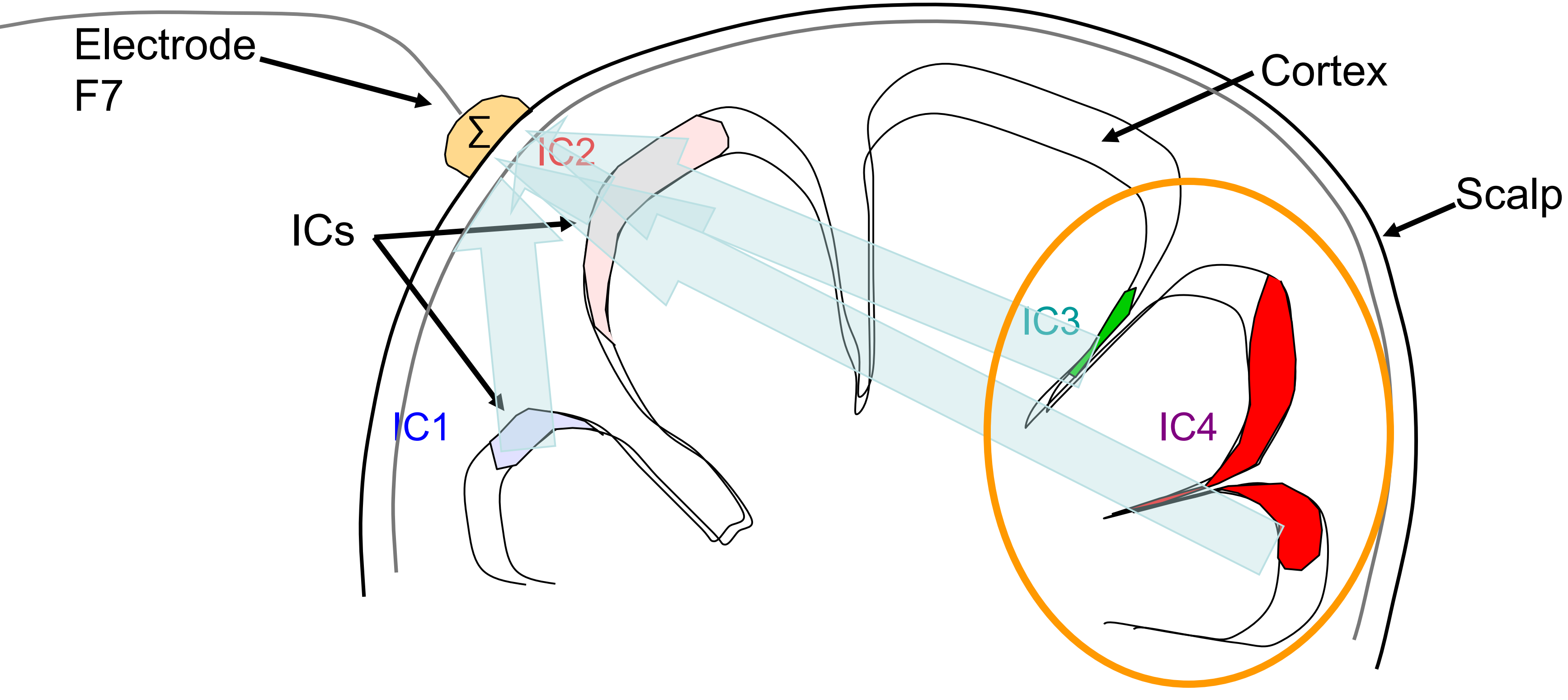
Makeig, 2005

Third Subject



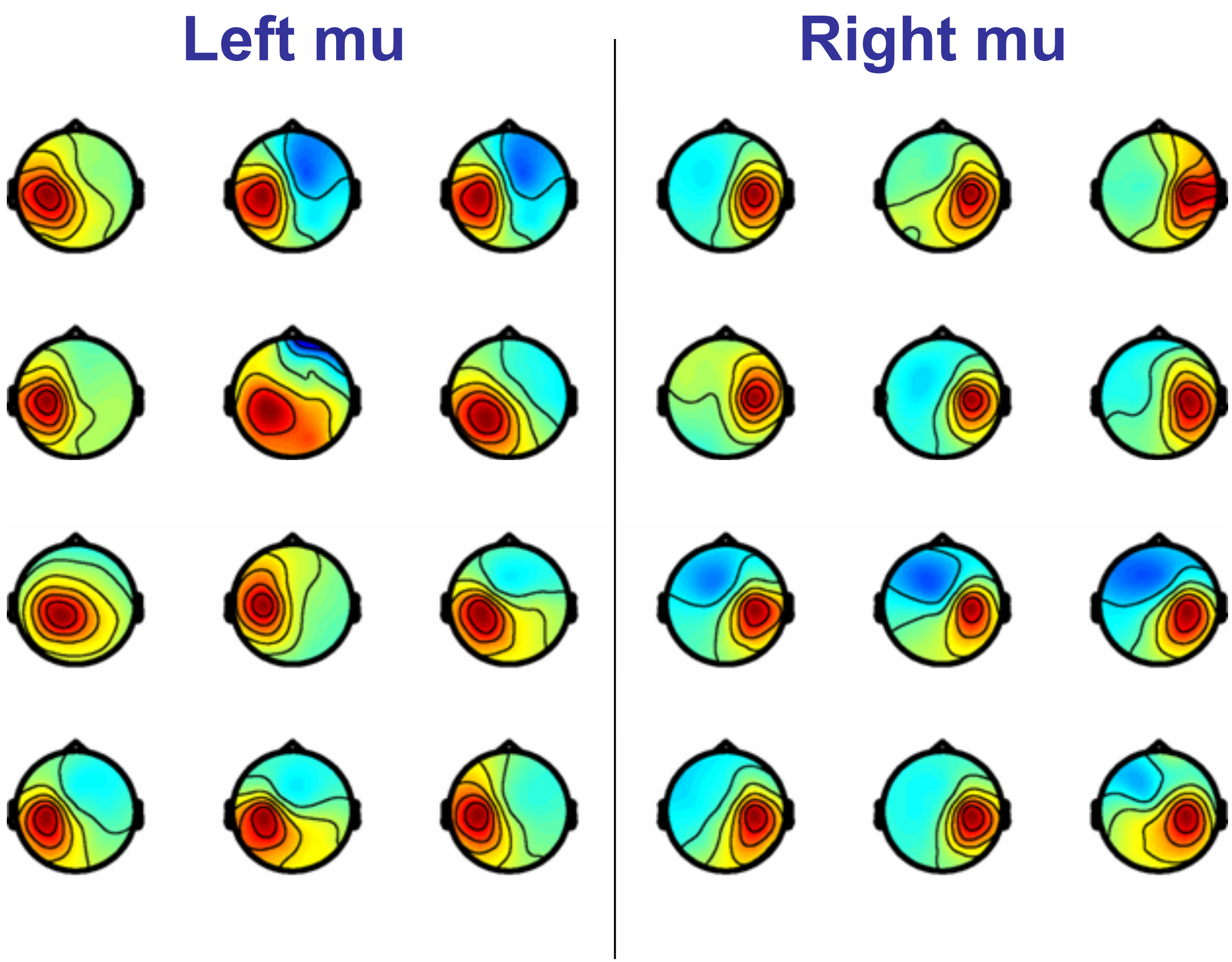
Makeig, 2005

Fourth Subject



Makeig, 2005

Clustering ICA components by eye



Makeig et al., ~2000 *unpublished*

So how to cluster components?

The same problems hold for clustering independent components

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

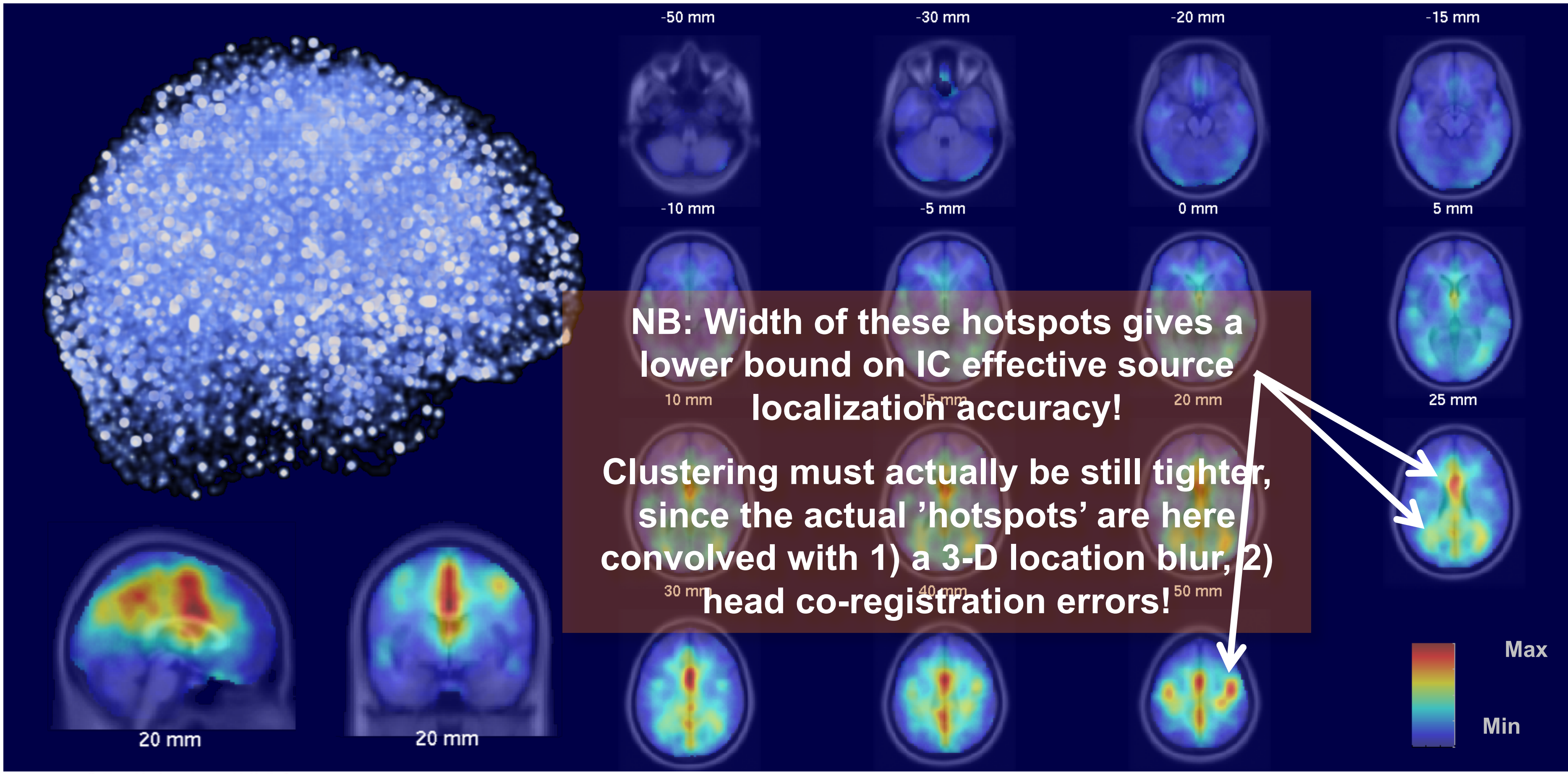
→ Are "the same" components found across subjects?

- 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?? …

Makeig, 2007

EEG IC Source Locations

(135,794 IC equivalent dipoles!)

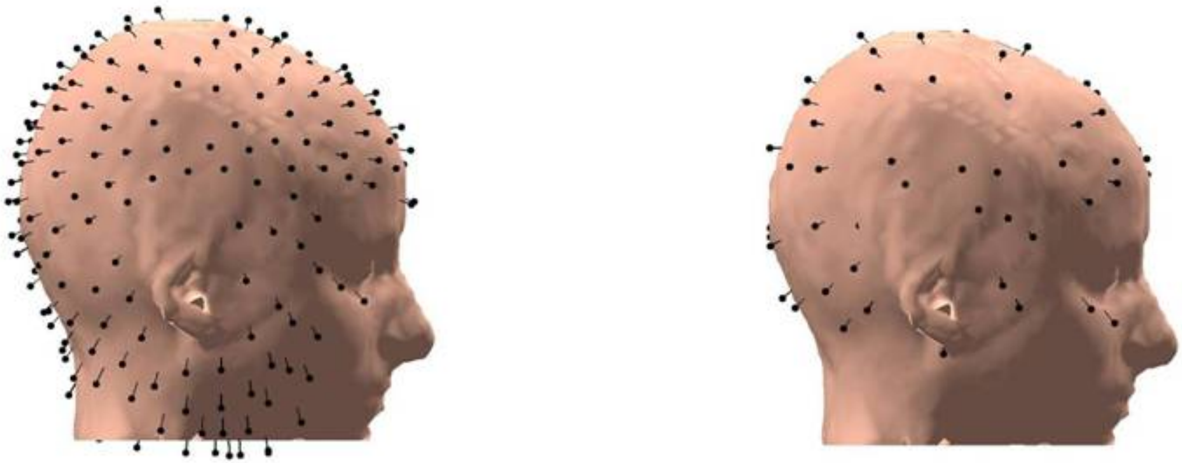


Nima Bigdely-Shamlo, Kay Robbins, Christian Kothe, Jessica His, Scott Makeig, 2013

… Some caveats

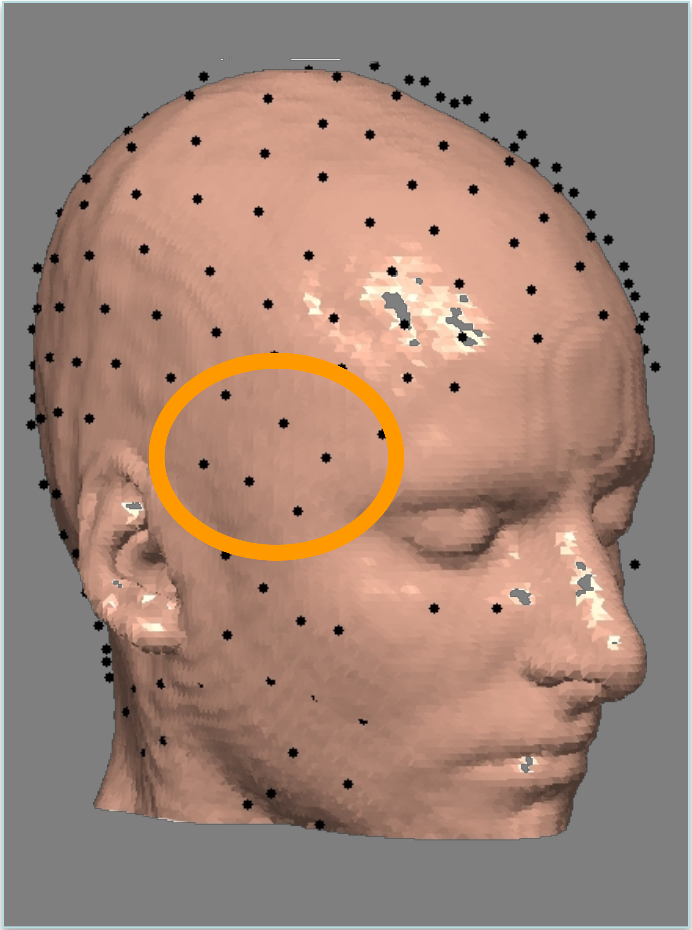
In this *dipoledensity()* assay …

- 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 'guestimated' !
- Brain geometries differ!



Graphics: Julie Onton, 2005

Co-Registration of Electrodes with MR Image



MR + EEG

EEG

Onton & Makeig, 2004

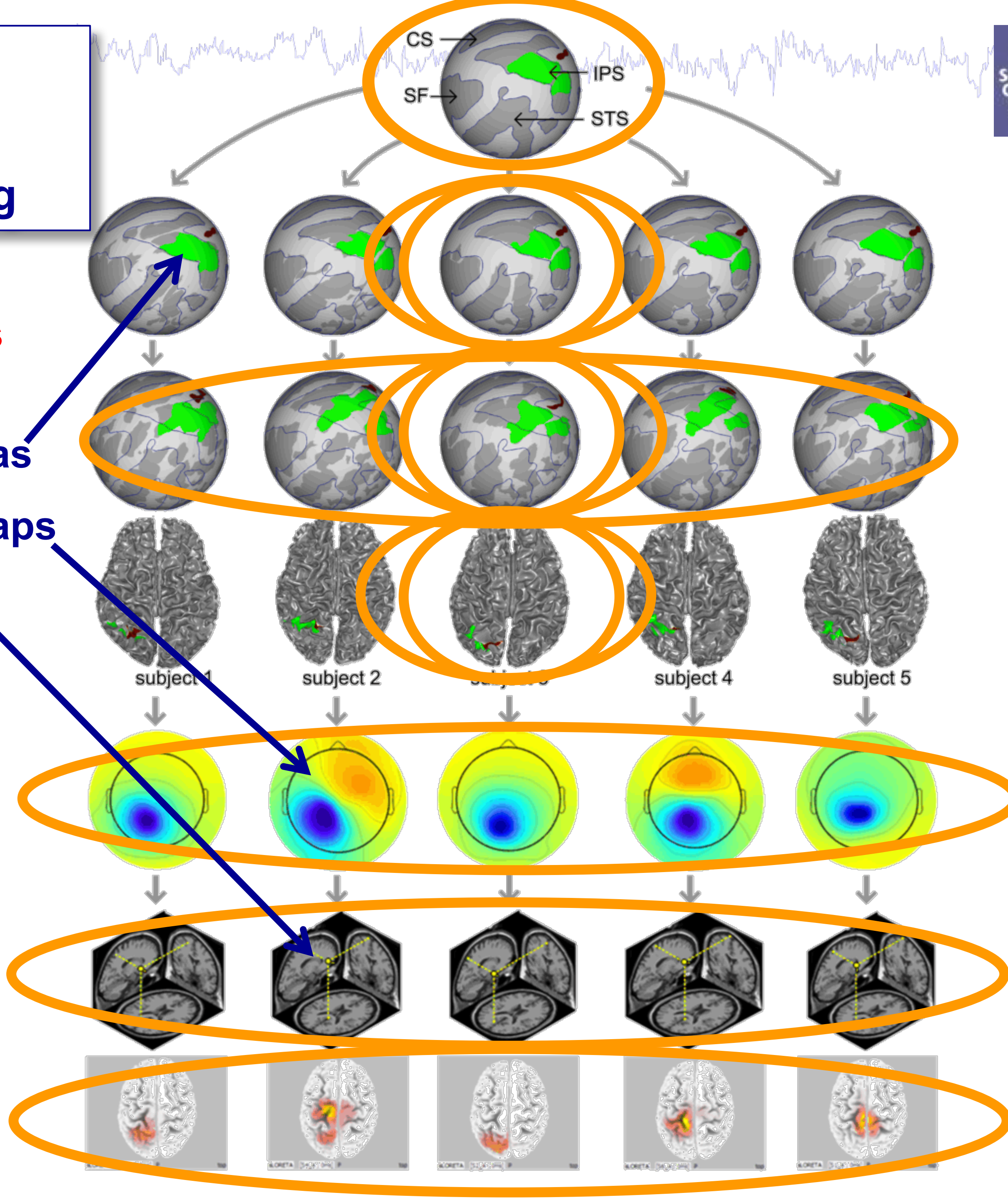
Arthur Tsai – Topological source clustering

Why should IC clusters have breadth?

Equivalent cortical areas

Have different scalp maps

And dipole locations!



Arthur Tsai et al., *NeuroImage*, 2014

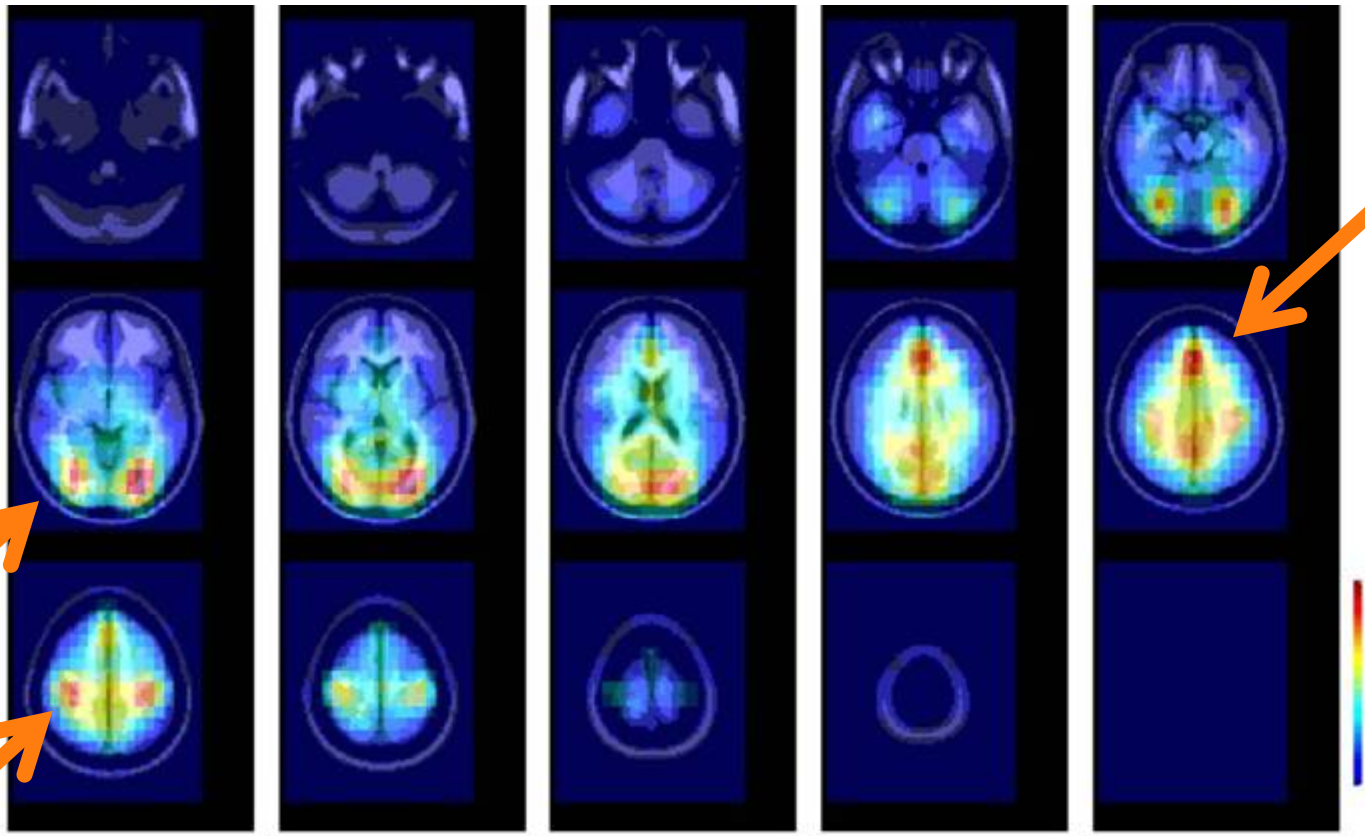
Does the spatial distribution of IC equivalent dipole source locations depend on the task the subject performs?

i.e.

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

Makeig, 2007

Equivalent dipole density



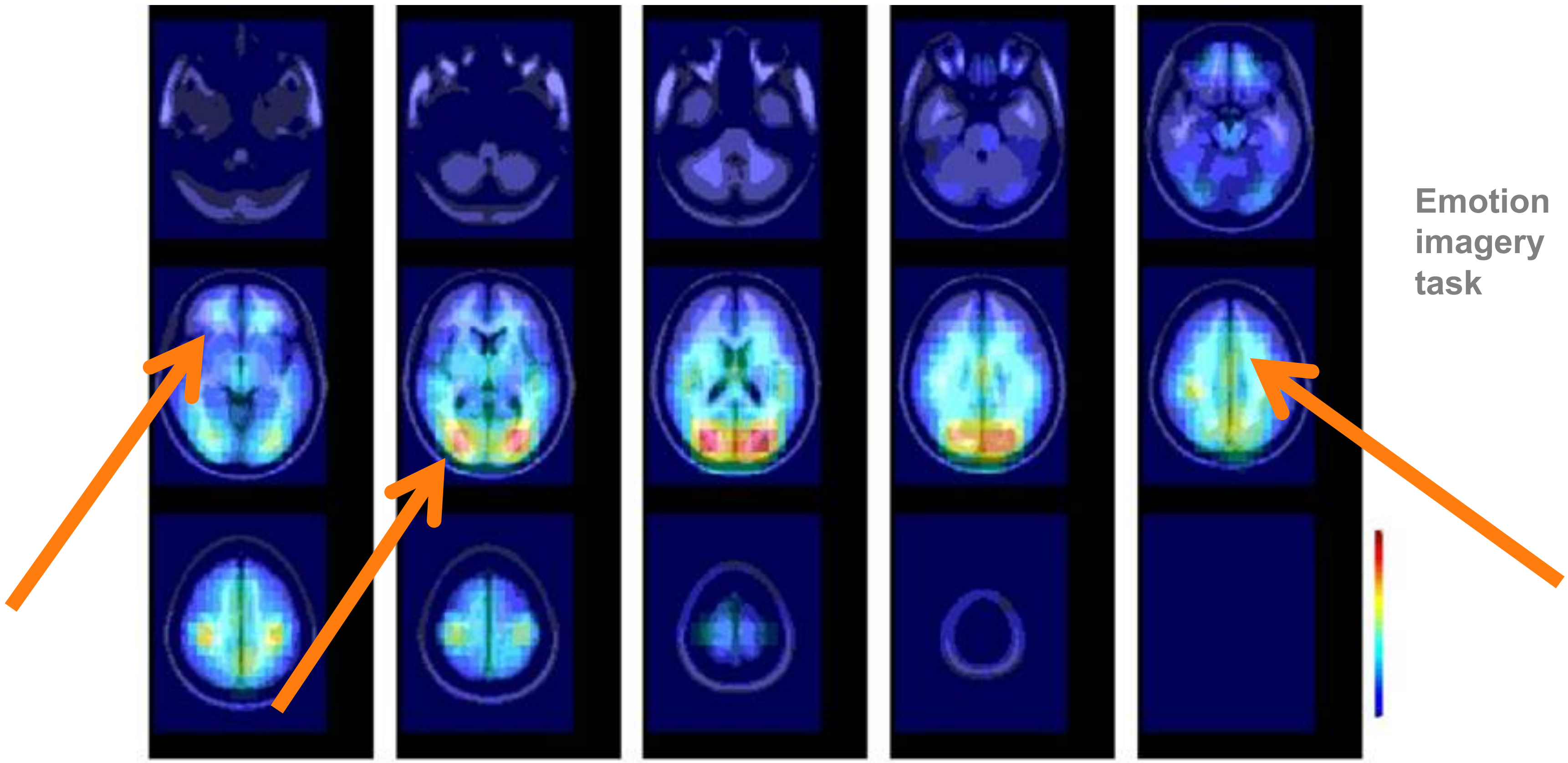
Sternberg letter memory task

>> dipoledensity()

Onton et al., 2

Onton et al., '05

Equivalent dipole density



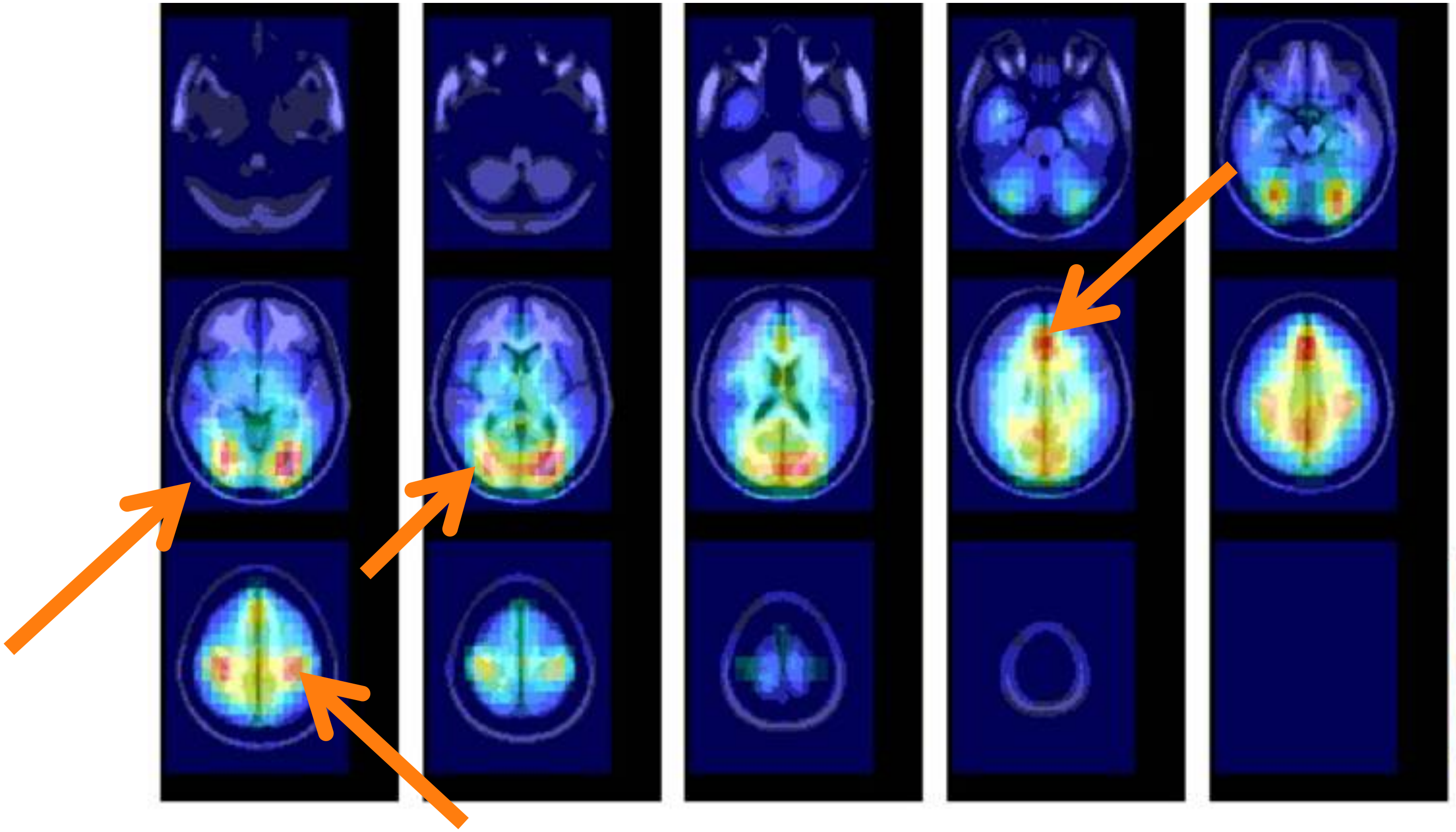
Emotion imagery task

>> dipoledensity()

Onton et al., 2

Onton et al., '05

Equivalent dipole density



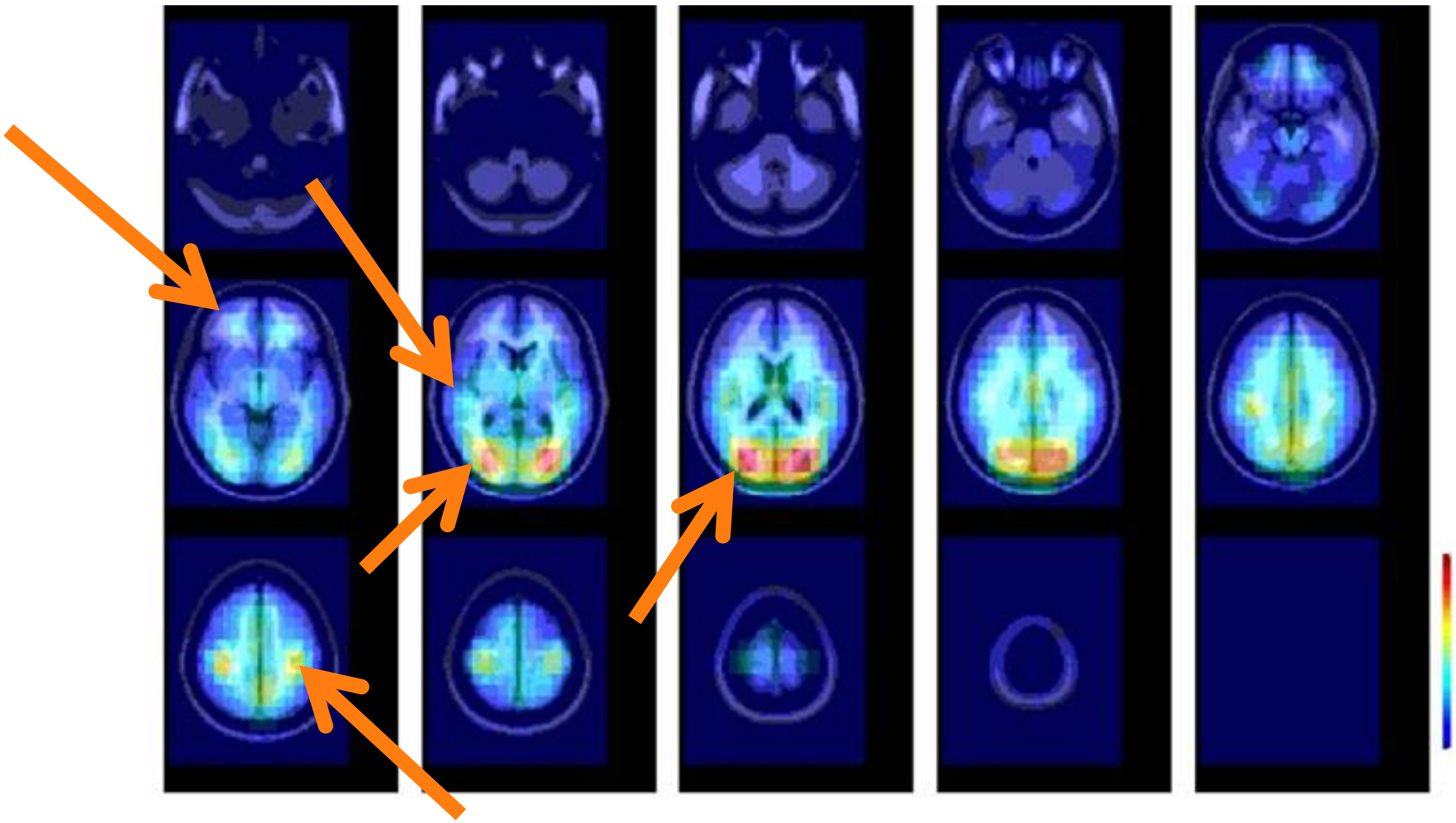
Sternberg letter memory task

>> dipoledensity()

Onton et al., 2

Onton et al., '05

Equivalent dipole density



Emotion imagery task

>> dipoledensity()

Onton et al., 2

Onton et al., '05

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.

Makeig, 2007

Study IC Clustering: Assumptions

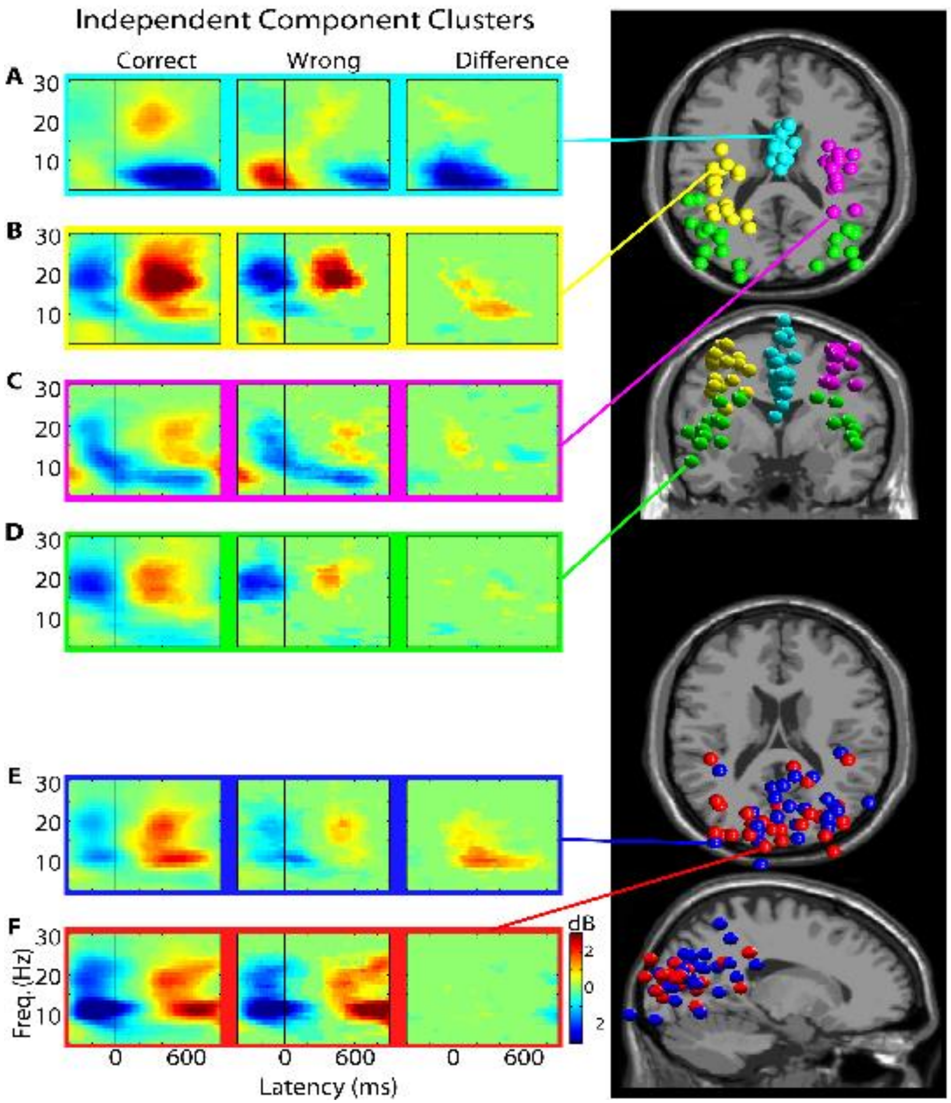
- Assumes there are *functionally equivalent* ICs across most subjects.
- Assumes these ICs have *similar responses* to experimental conditions across **a set** of measures

  (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.

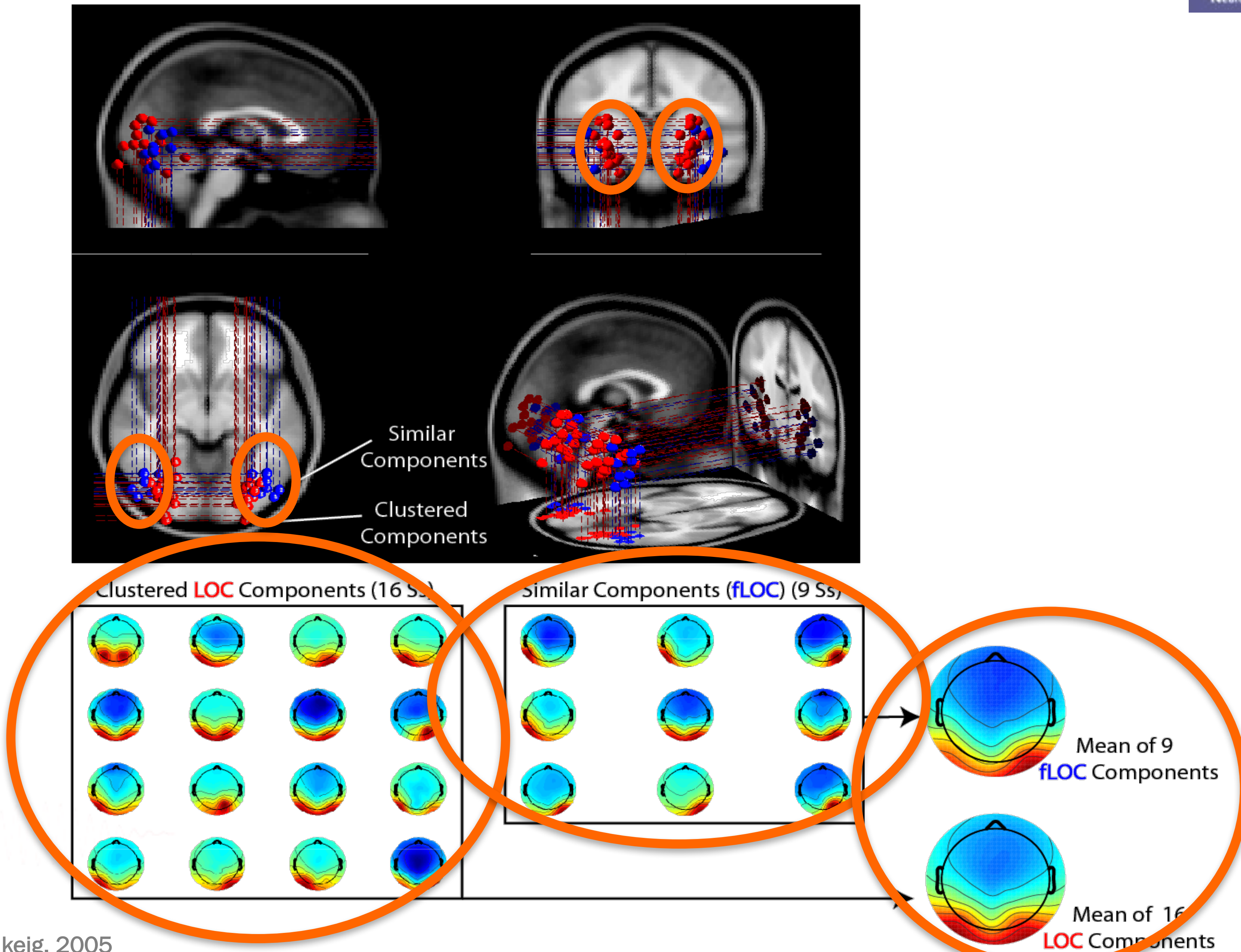
Onton & Makeig, 2007

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.

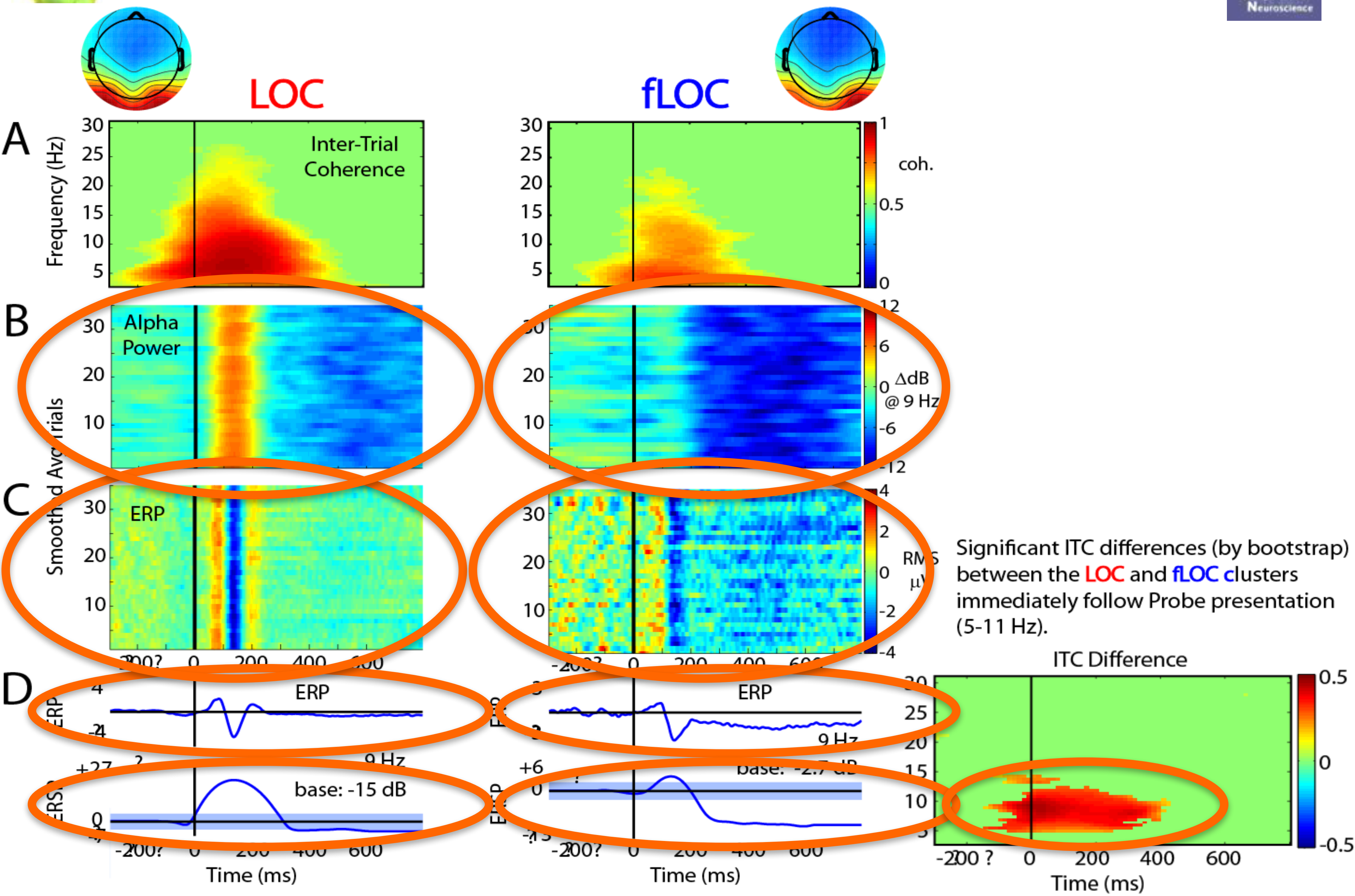
Makeig, 2007

Why aren't all participants in every IC cluster?



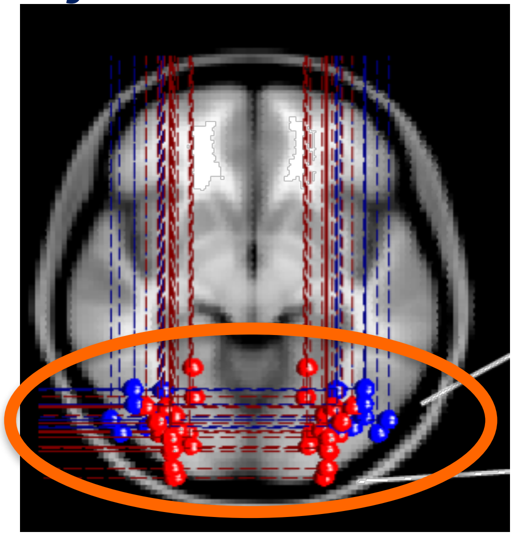
Onton & Makeig, 2005

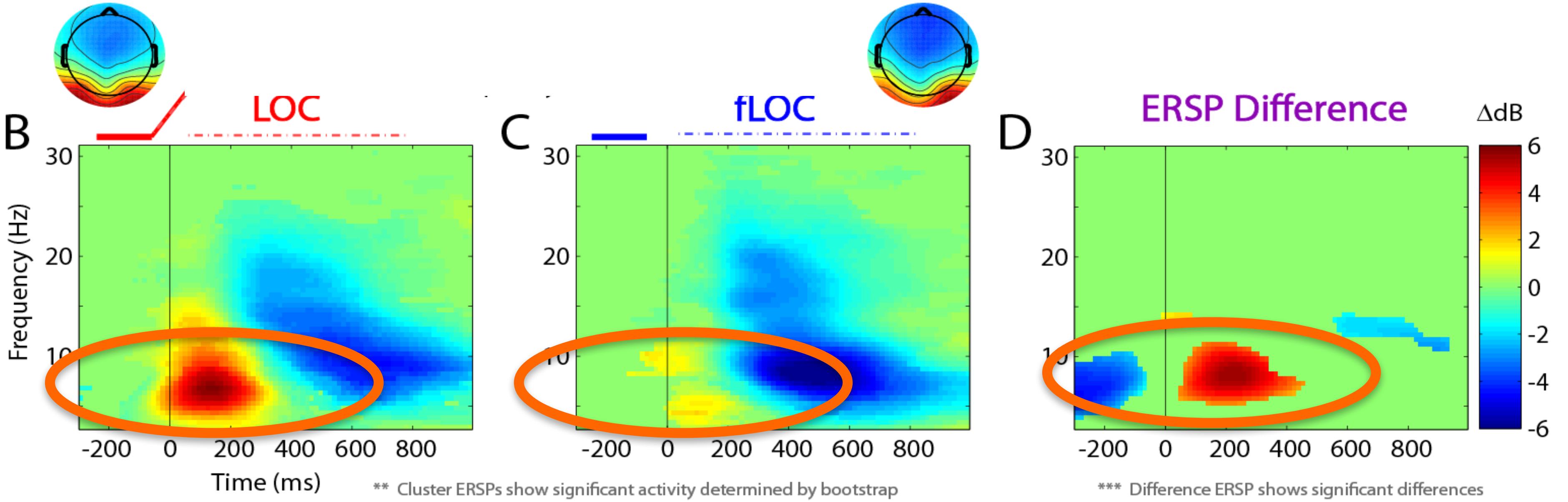
Subject differences?



Significant ITC differences (by bootstrap) between the LOC and fLOC clusters immediately follow Probe presentation (5-11 Hz).

Subject differences?

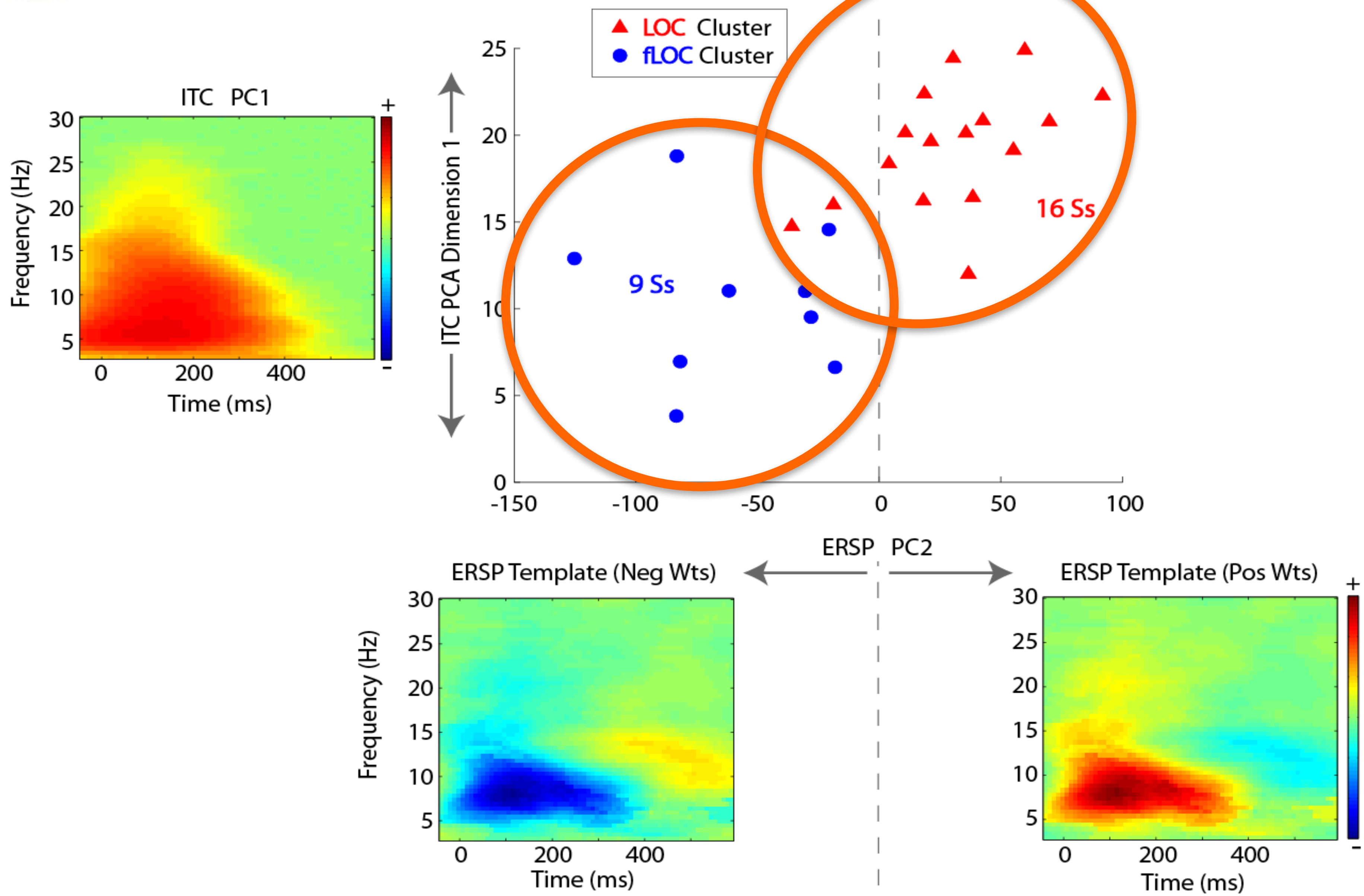




** Cluster ERSPs show significant activity determined by bootstrap statistics within subject and binomial probability between subjects ($p < 0.01$)

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

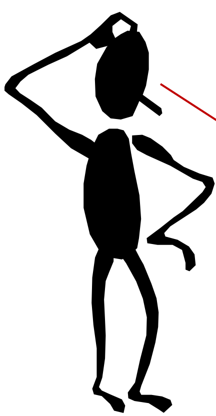
Subject differences?

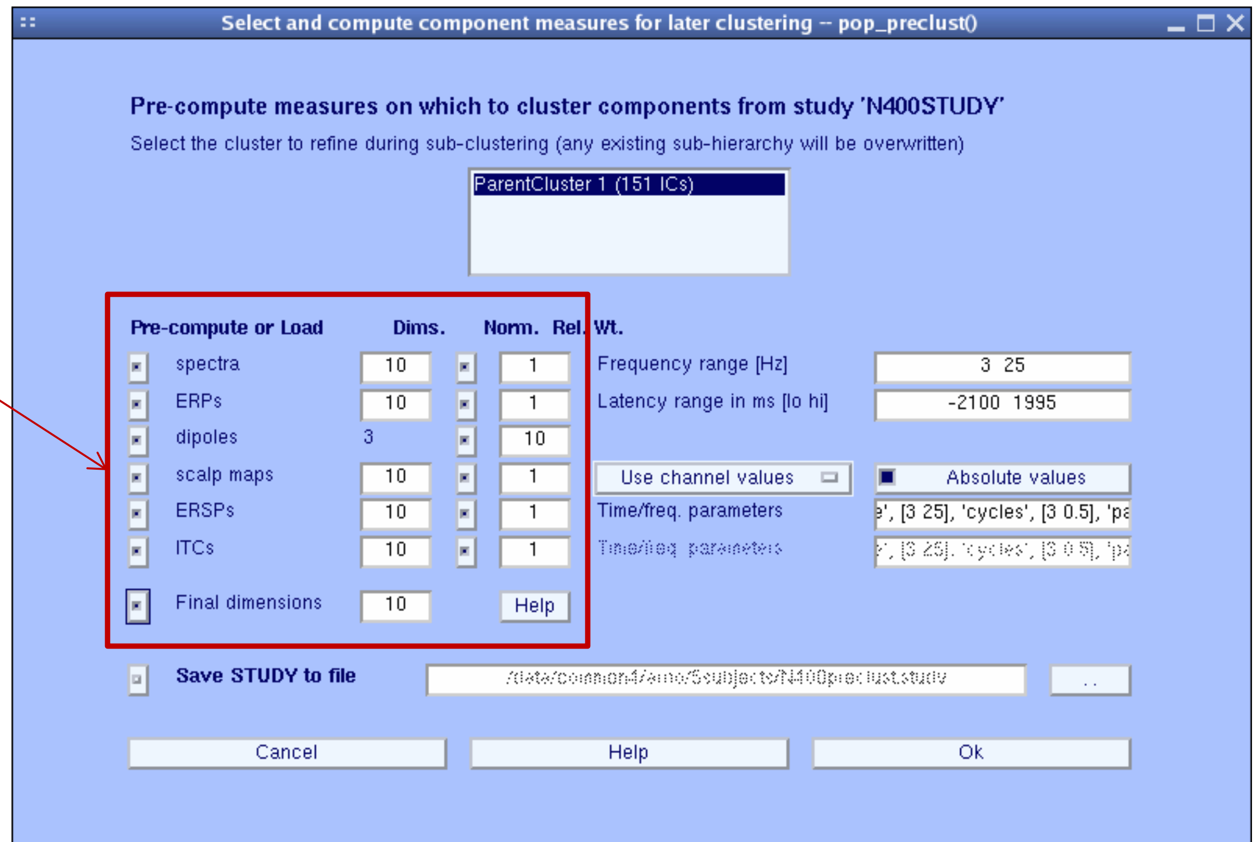


STUDY IC Clustering: Practical Problems

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



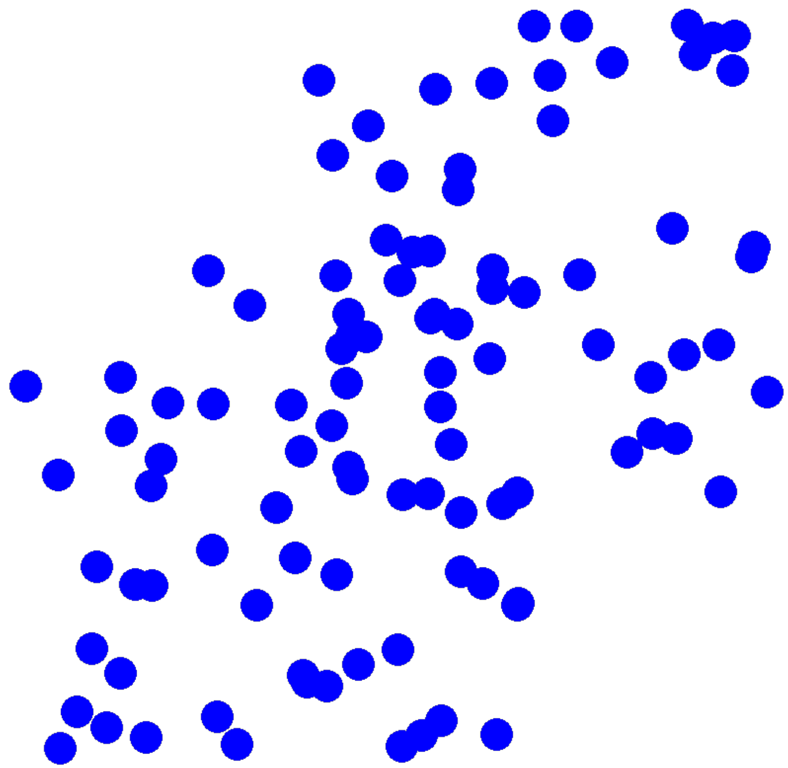


N. Bigdely-Shamlo, 2010

Problems with multi-measure clustering

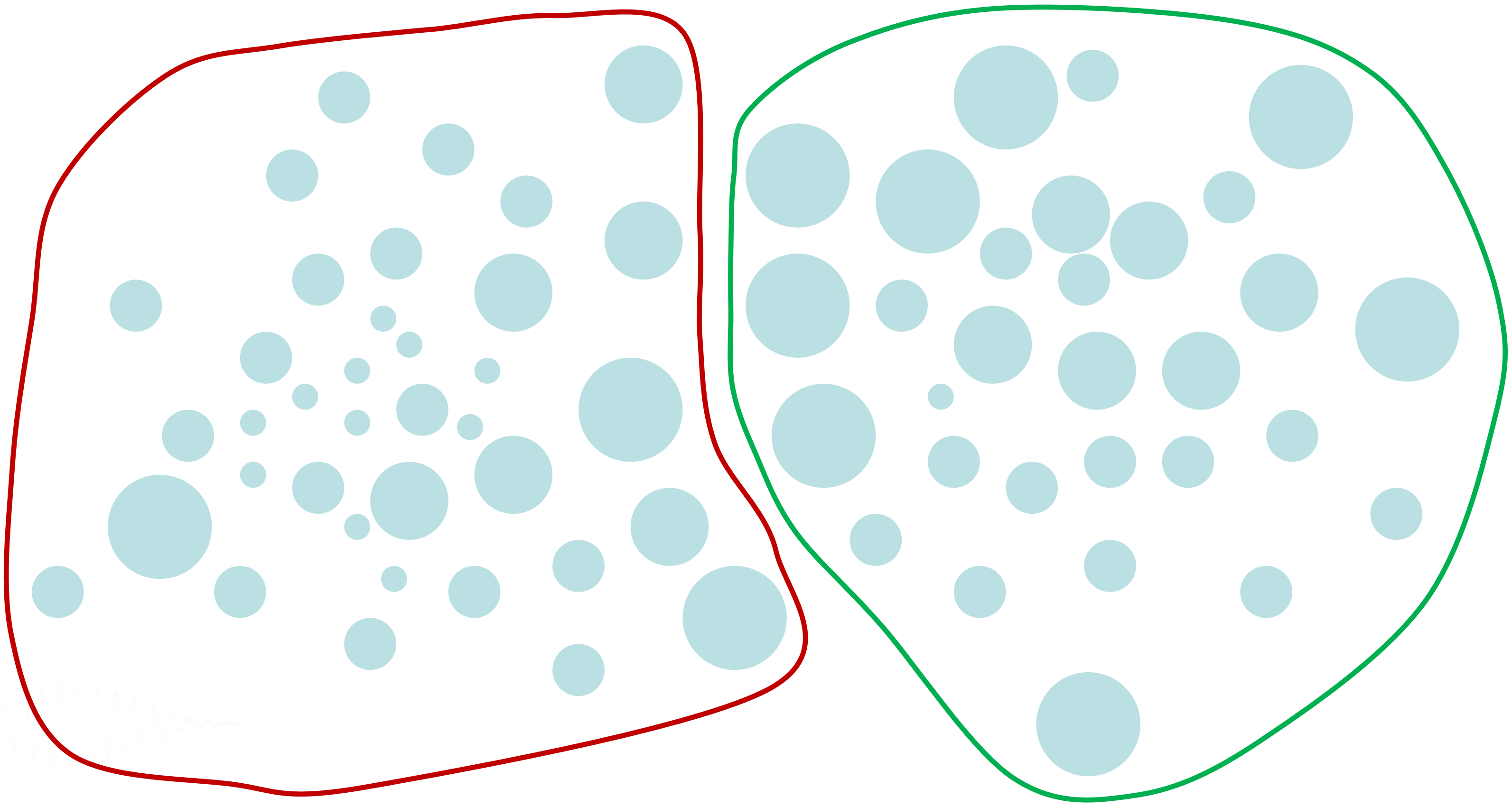
In a uniform density distribution,

where are the clusters by location?



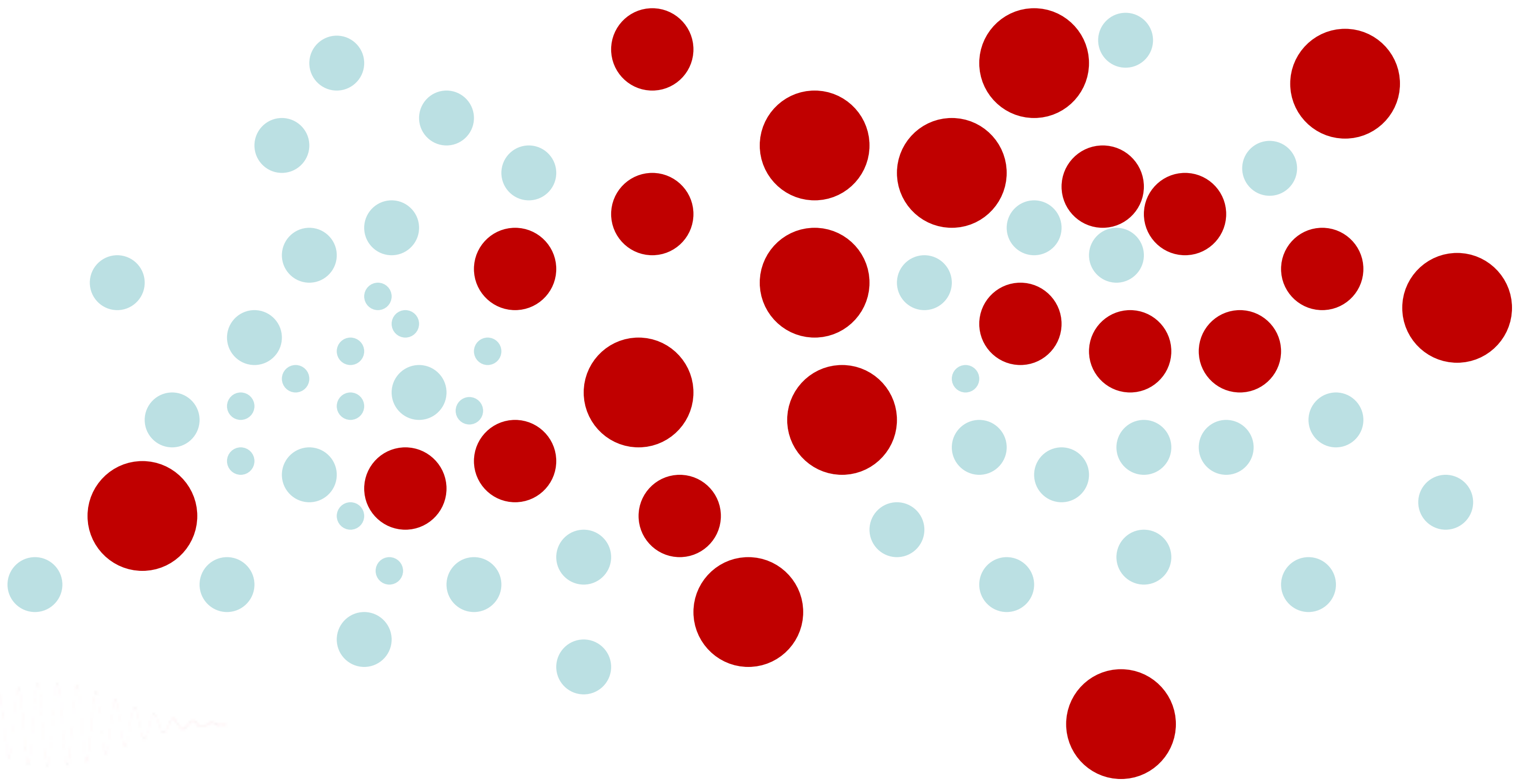
Problems with multi-measure clustering

What are the clusters according to location?



Problems with multi-measure clustering

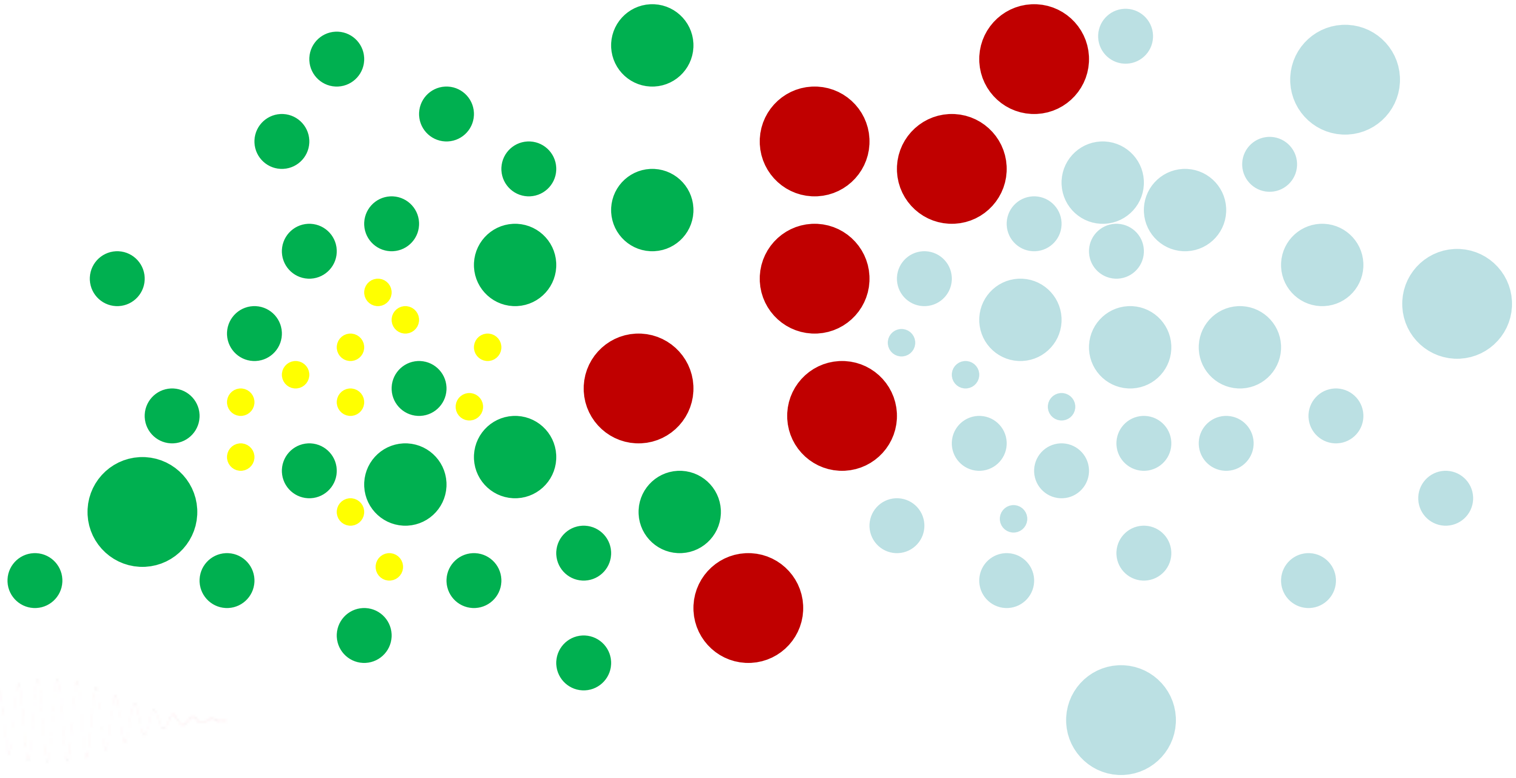
What are the clusters according to size ?



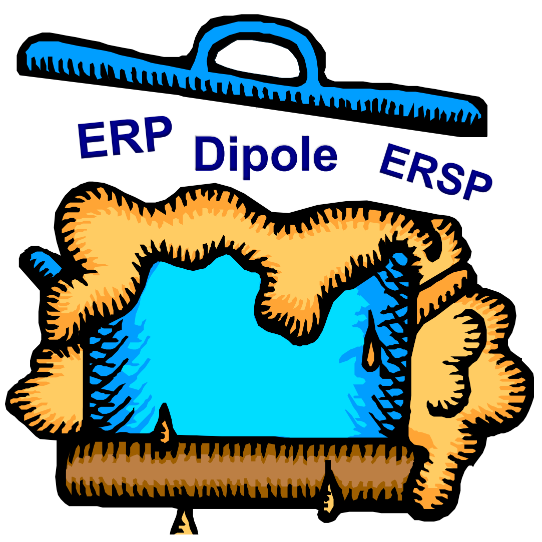
Problems with multi-measure clustering

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.



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:

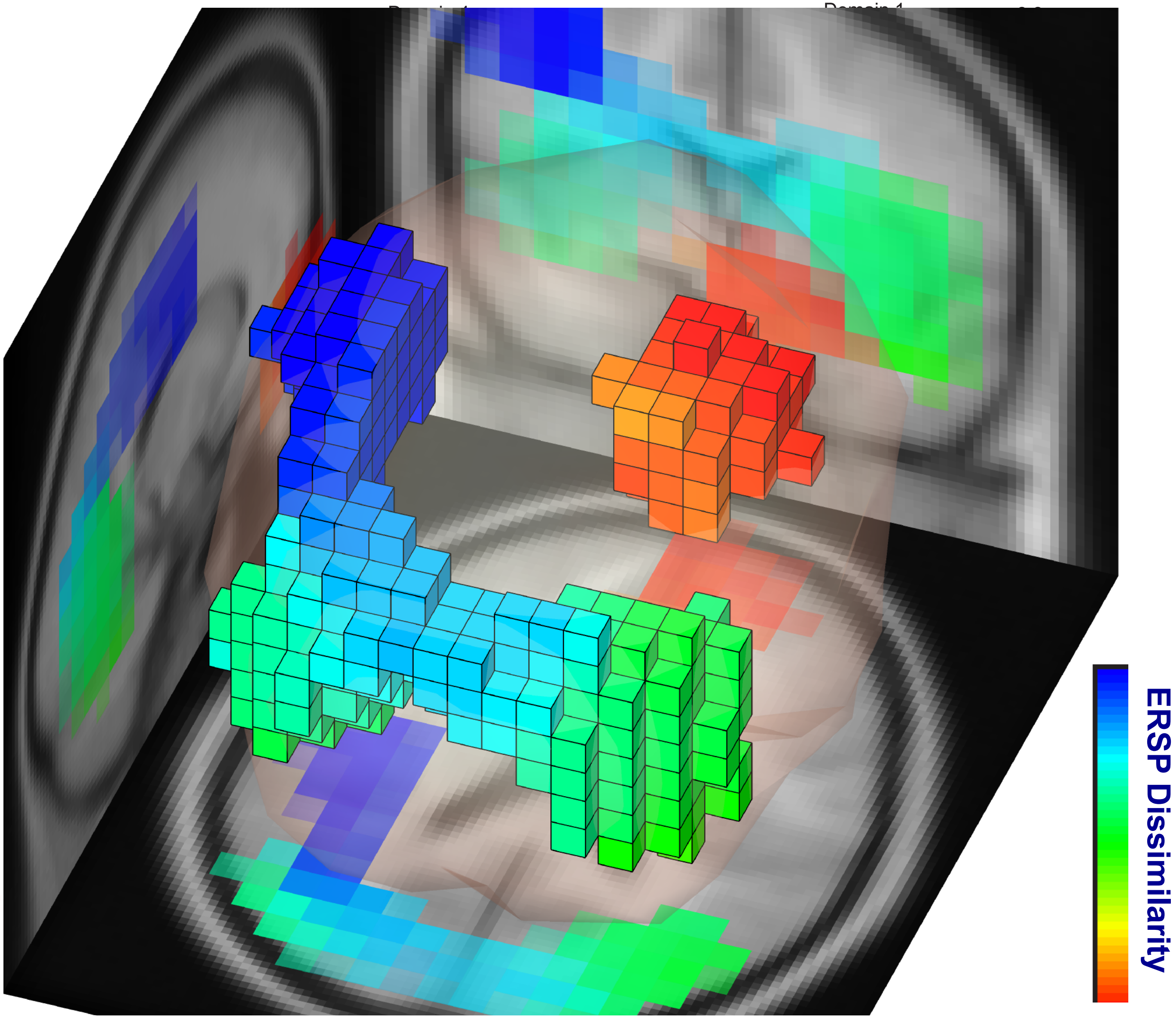
1. Where in ‘template brain space’ does our data have evidence that our measure of interest is consistent across nearby ICs?

2. Which such brain space voxel *domains* show consistent differences?

N. Bigdely-Shamlo, 2010

Project Target ERSPs on Equivalent Dipole Locations

Measure Projection: RSVP Task Example



N. Bigdely-Shamlo, 2011

Questions?