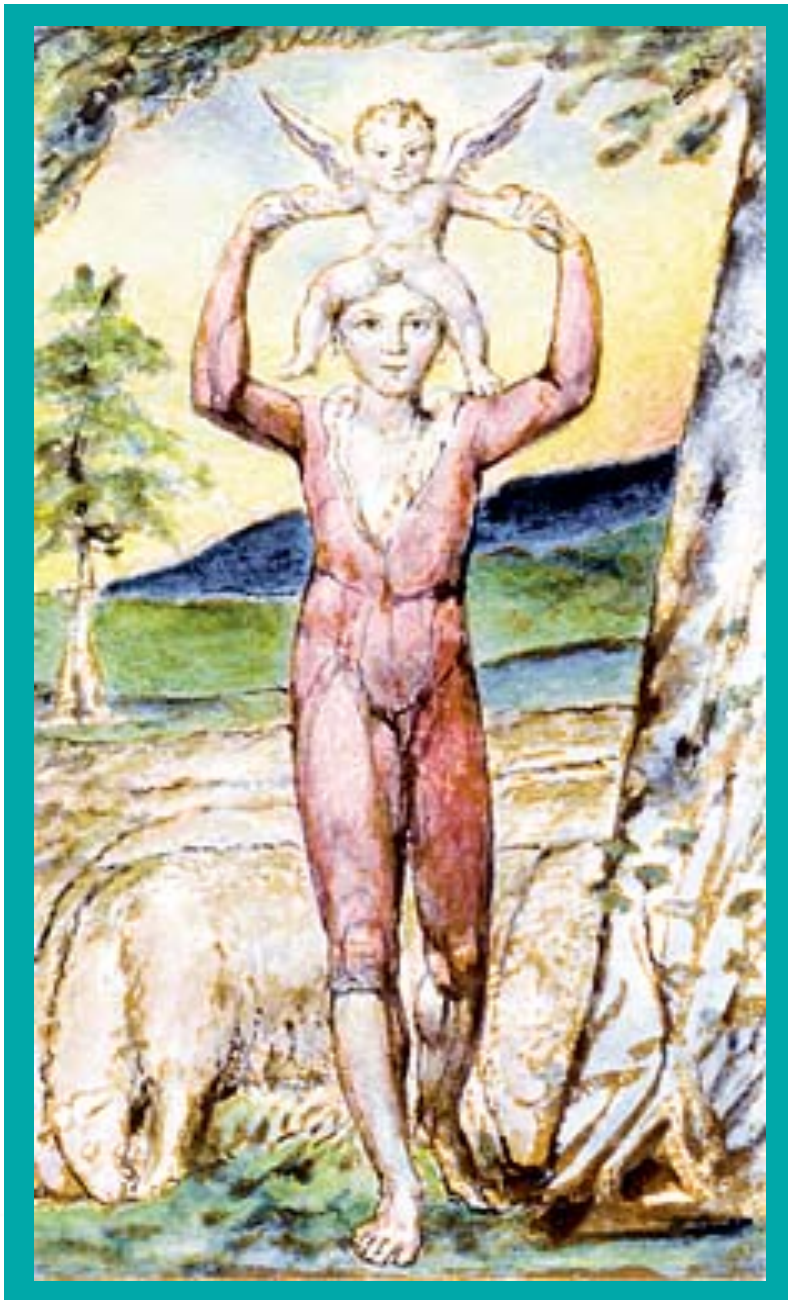
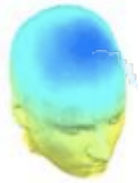


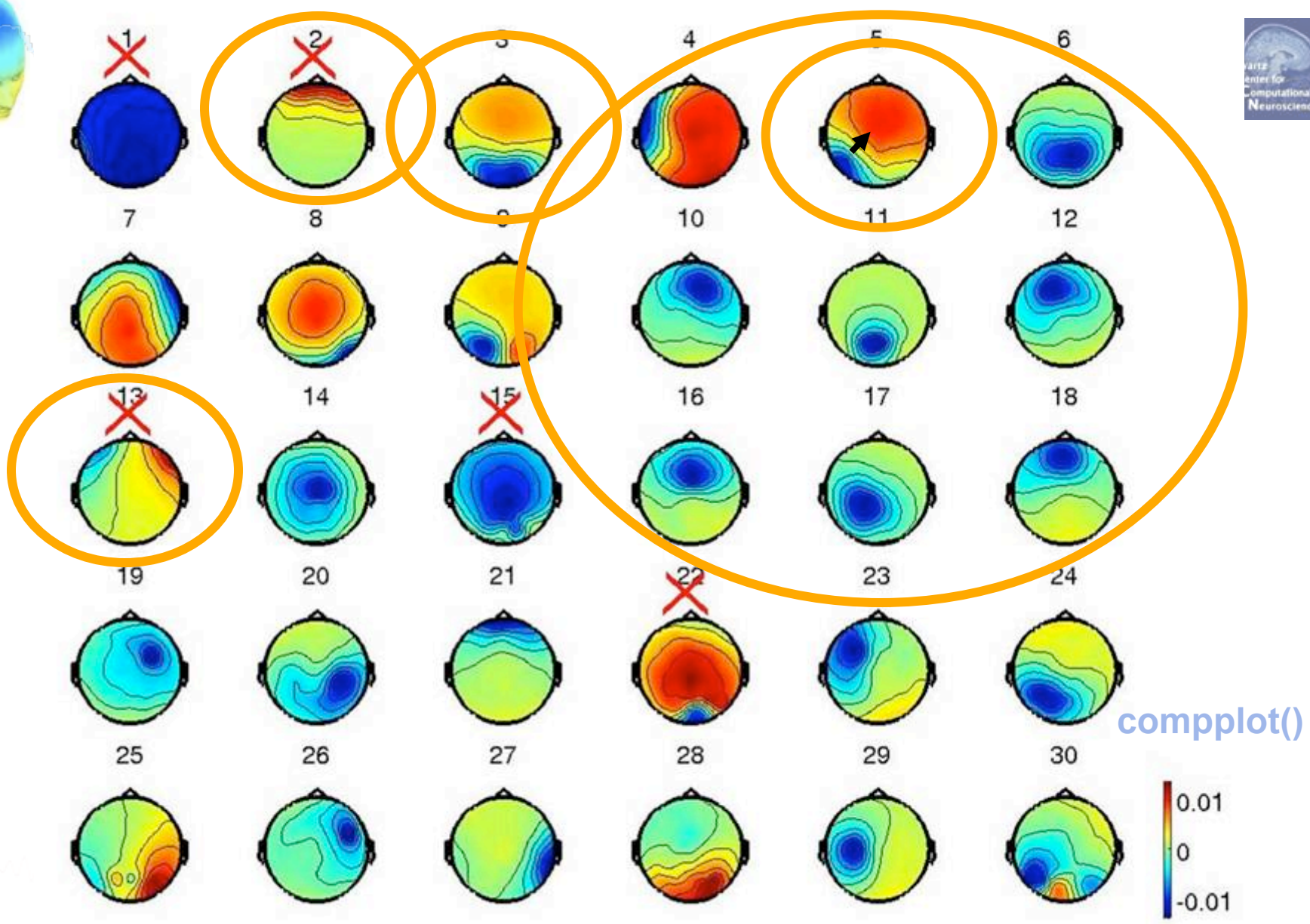
Independent Component Clustering

Scott Makeig
Institute for Neural Computation
University of California San Diego

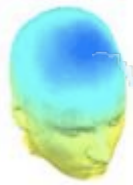
EEGLAB June, 2009
Aspet, France

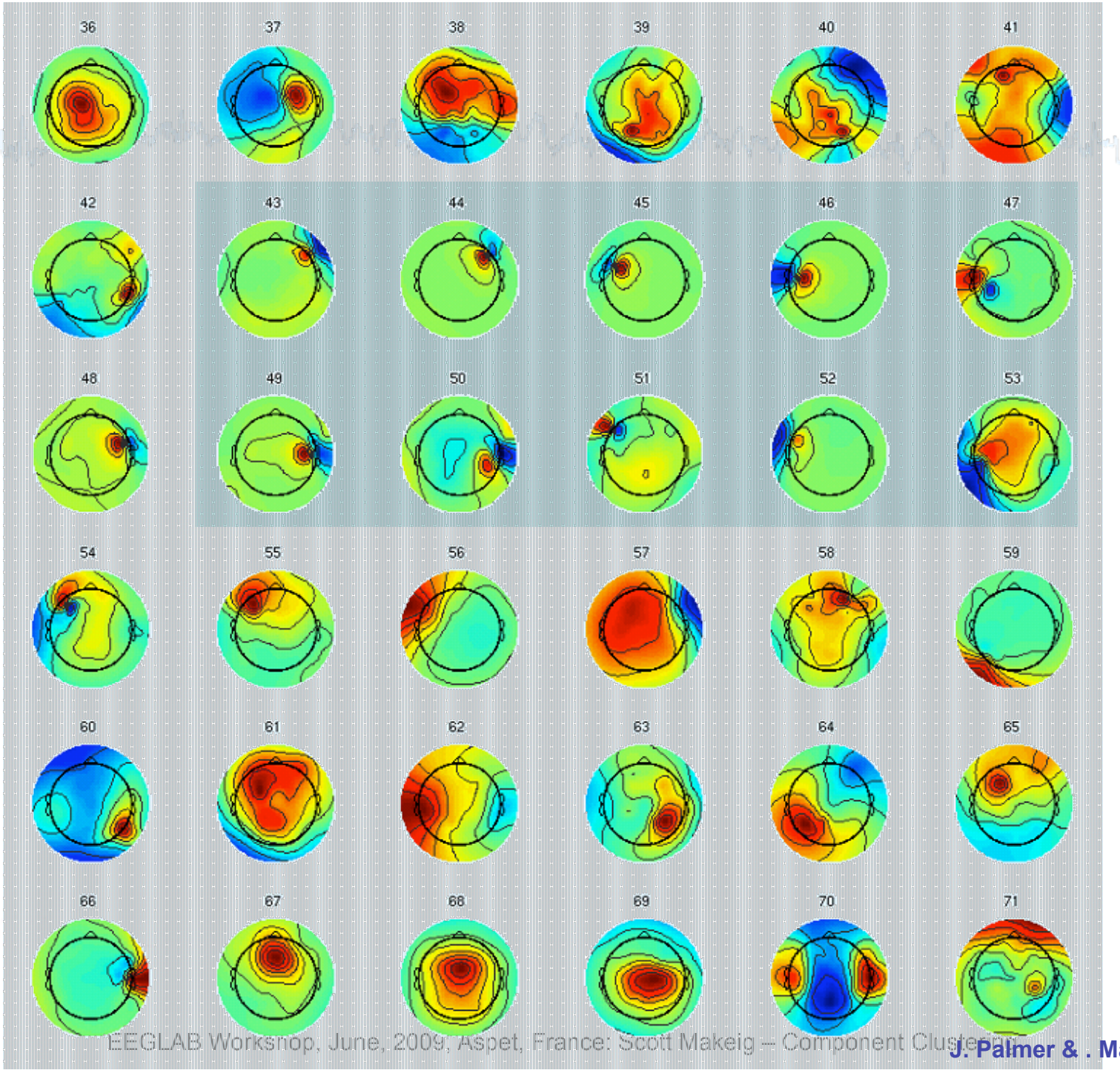
Largest 30 independent components (single subject)





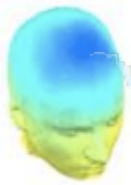
compplot()





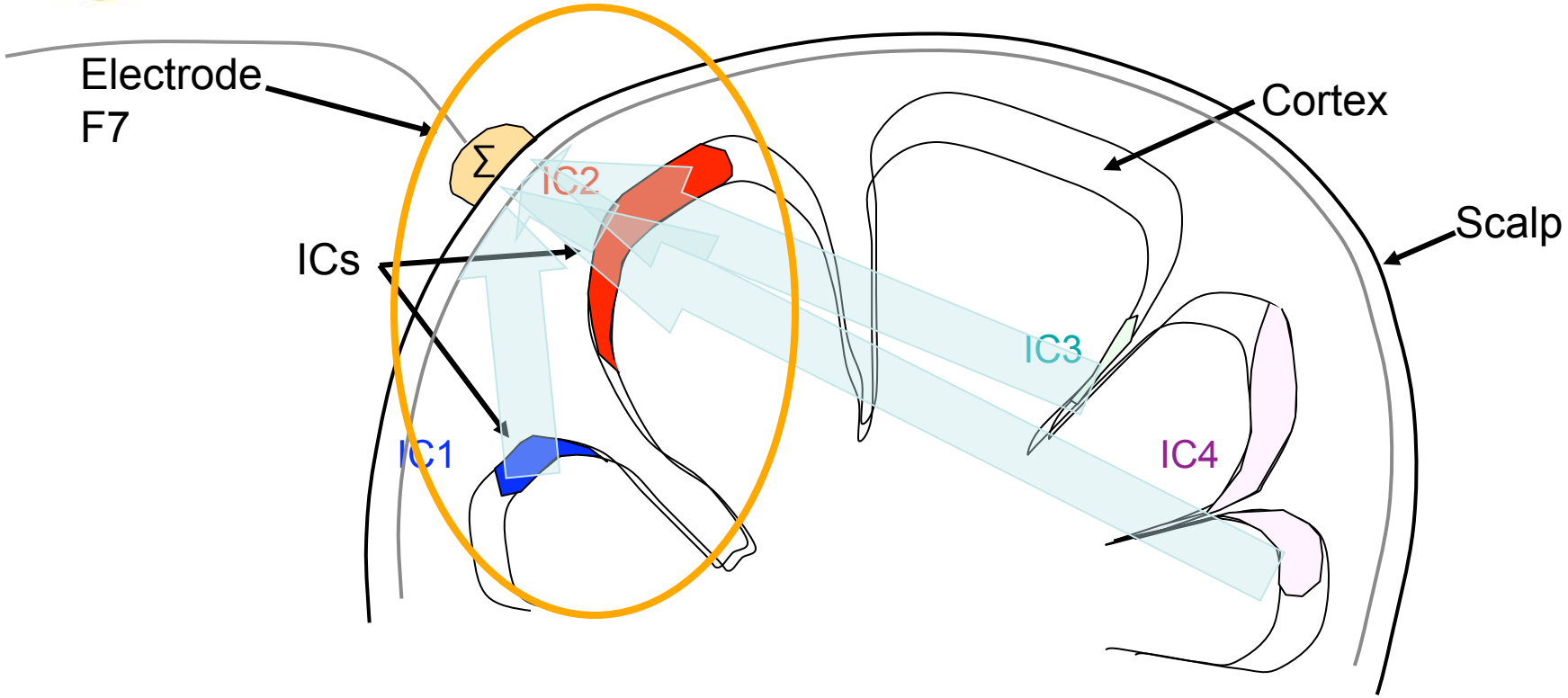
J. Palmer & . Makeig (2007)

Channel-based clustering

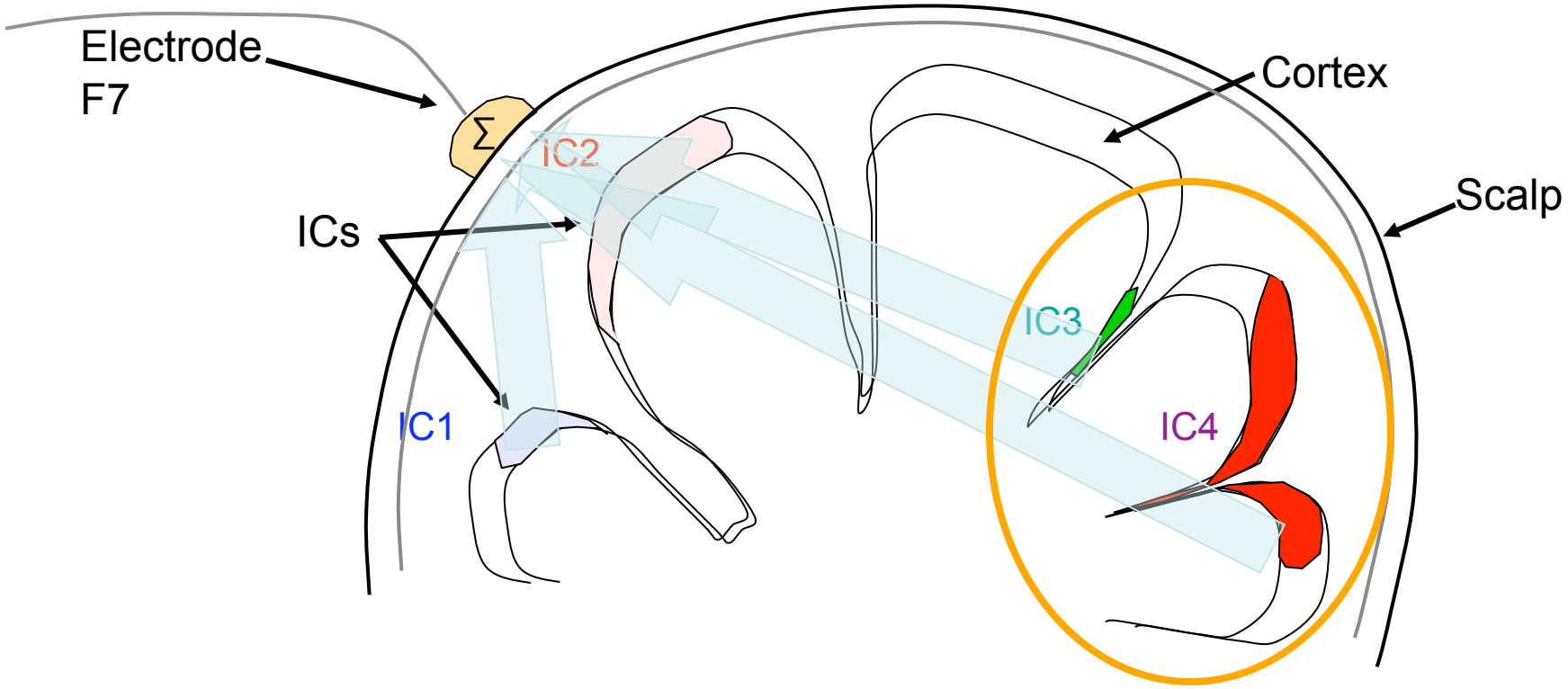


- ICA transforms the data from a ***channel basis*** (activity recorded at each scalp channel)
  - to a ***component basis*** (activity computed at each independent spatially-filtered (cortical or non-cortical) component process).
- Normally, EEG researchers assume that each electrode, say F7 is equivalent to the F7 of each subject – and then ‘cluster‘ their data by identifying channels with equivalent scalp locations ...
- But this is only *roughly* correct!

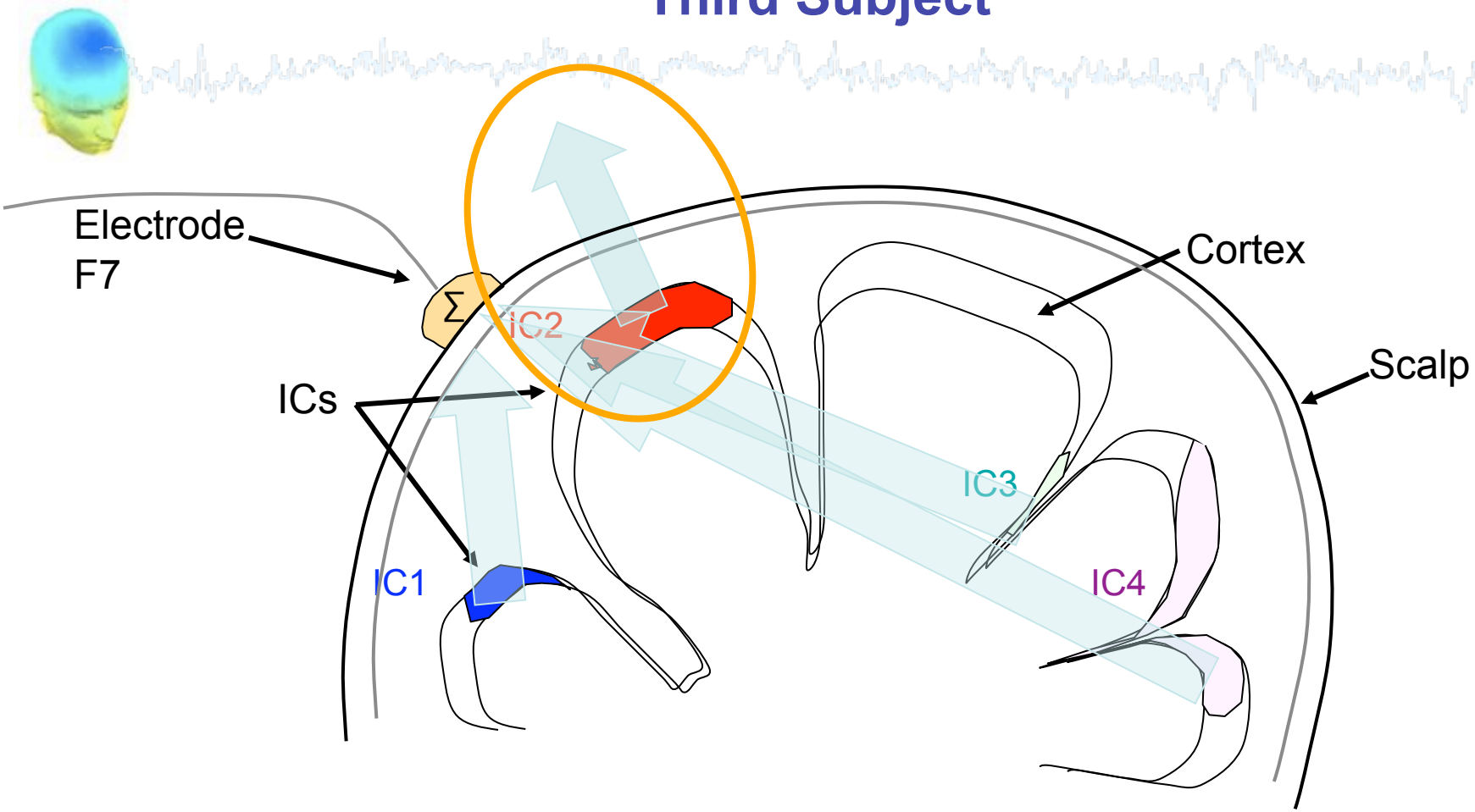
Example: First Subject



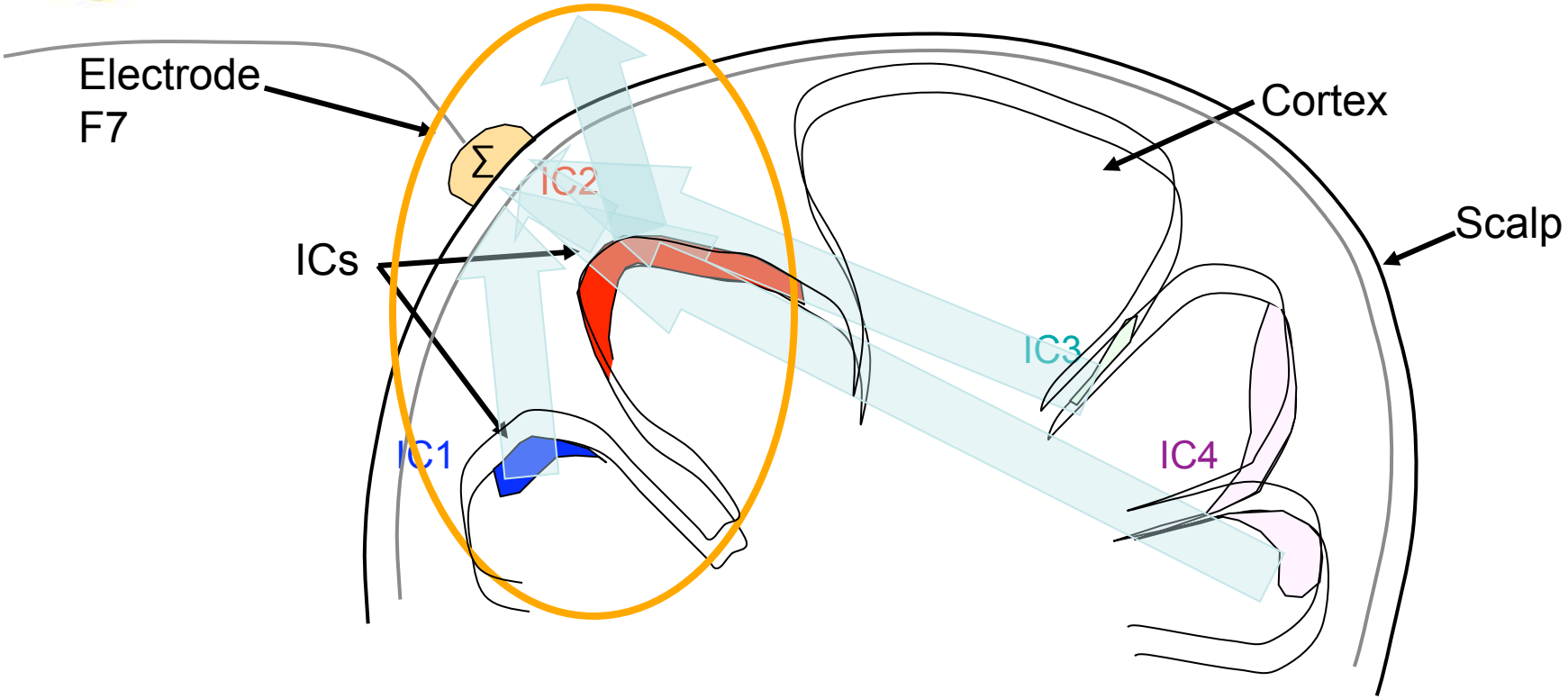
Second Subject



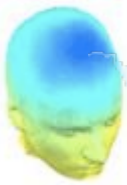
Third Subject



Fourth Subject



How to cluster independent components?

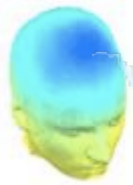


The same problems hold for clustering independent components

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? …



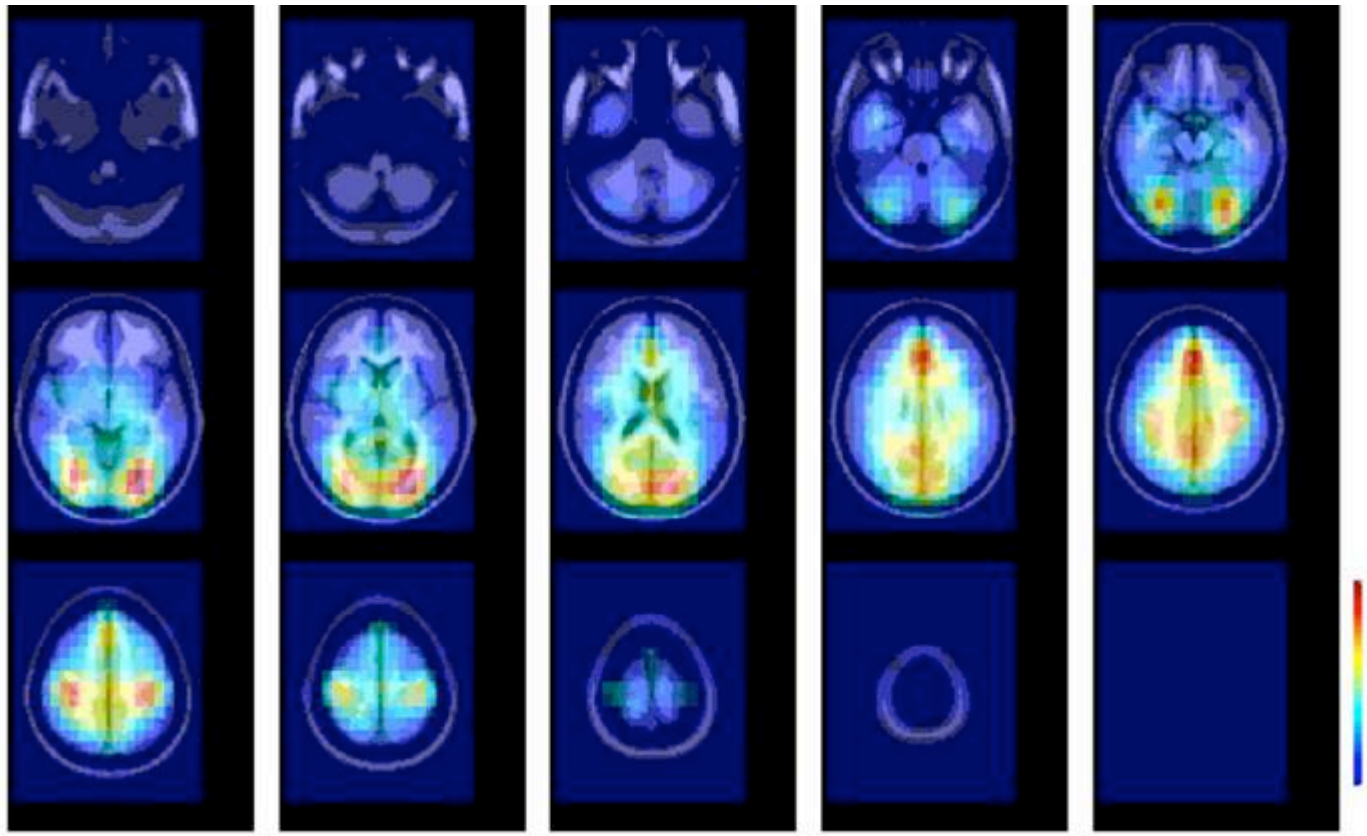
First, we should ask…

Does the spatial distribution of independent components depend on the task the subject performs?

i.e.

Do “the same” components (and component clusters) appear for every subject task?

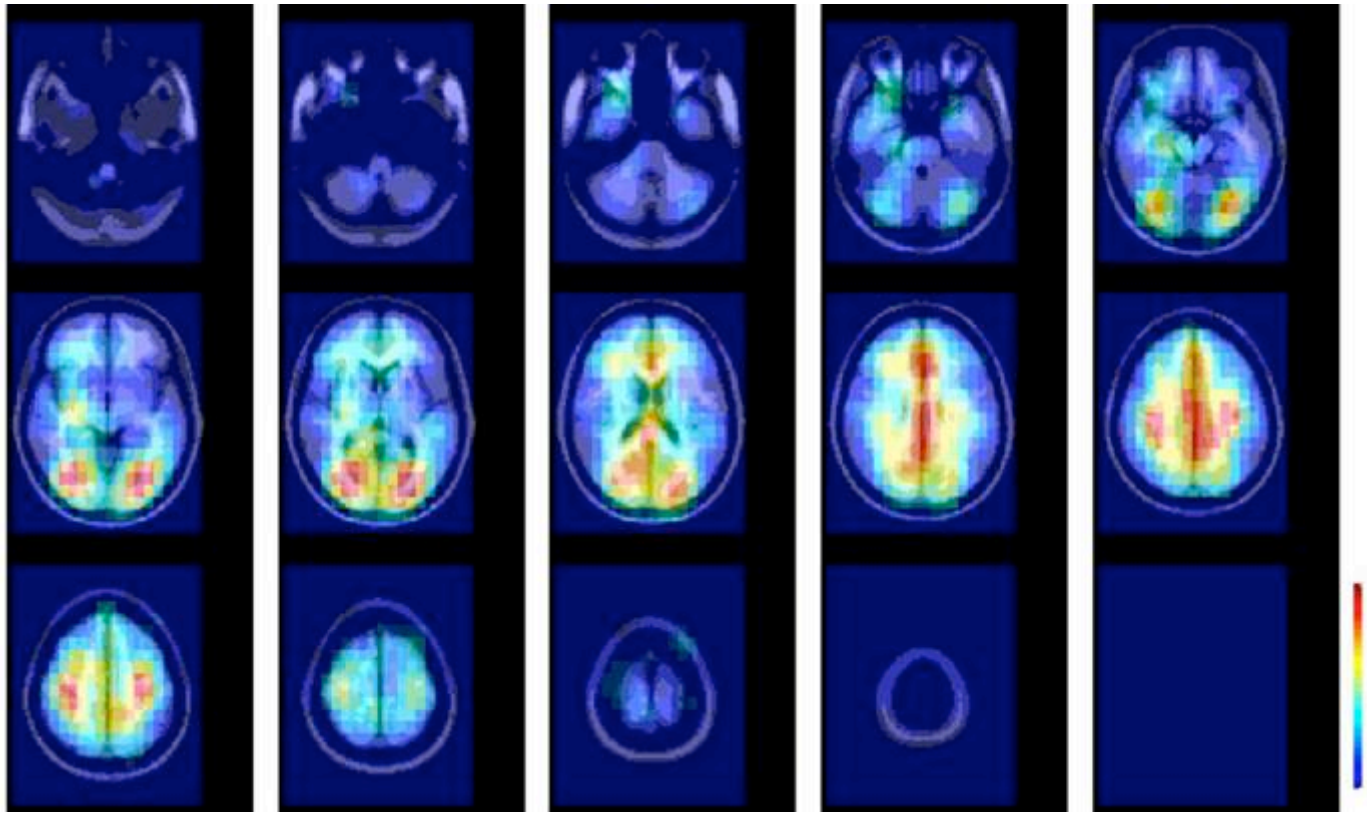
Equivalent dipole density



Sternberg letter memory task

>> dipoledensity()

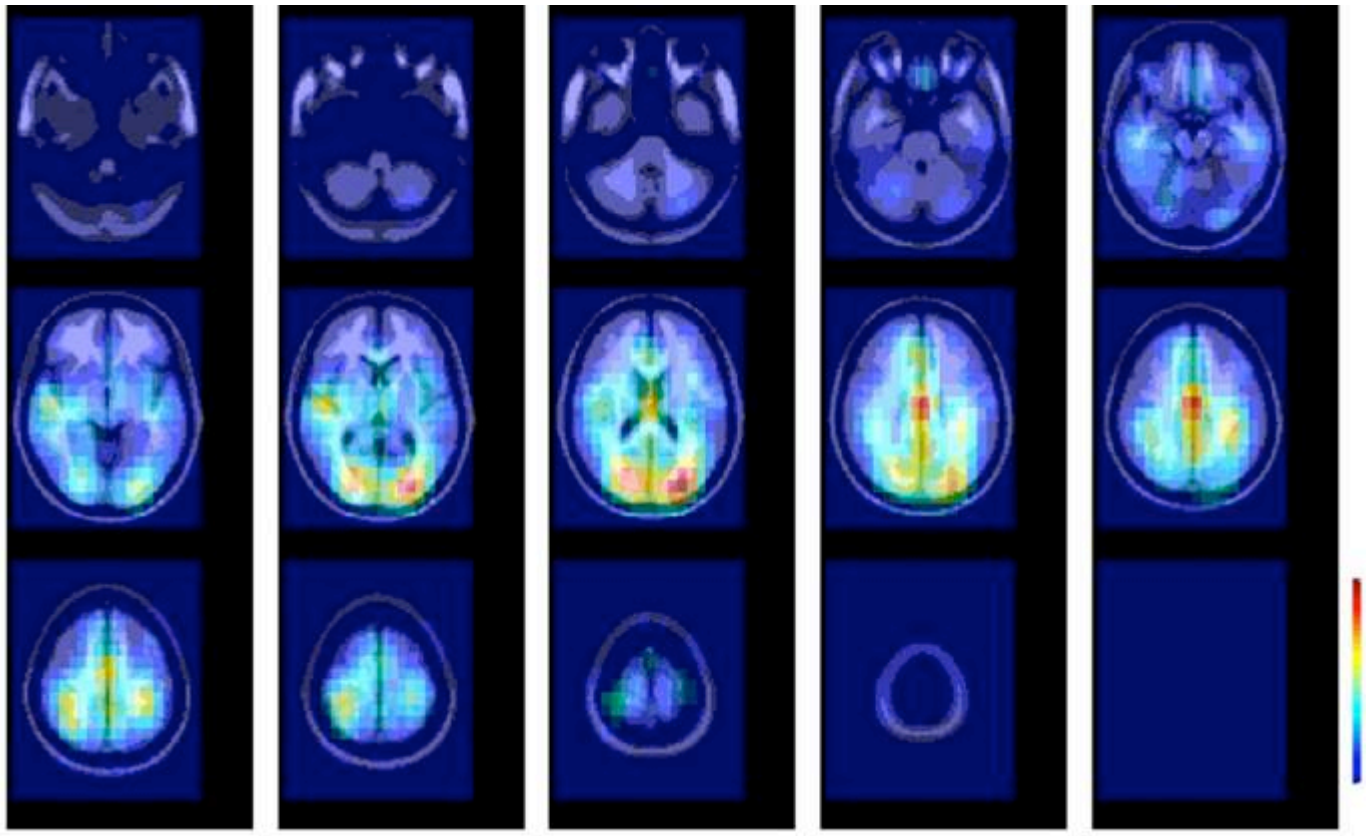
Equivalent dipole density



Letter twoback with feedback

>> dipoledensity()

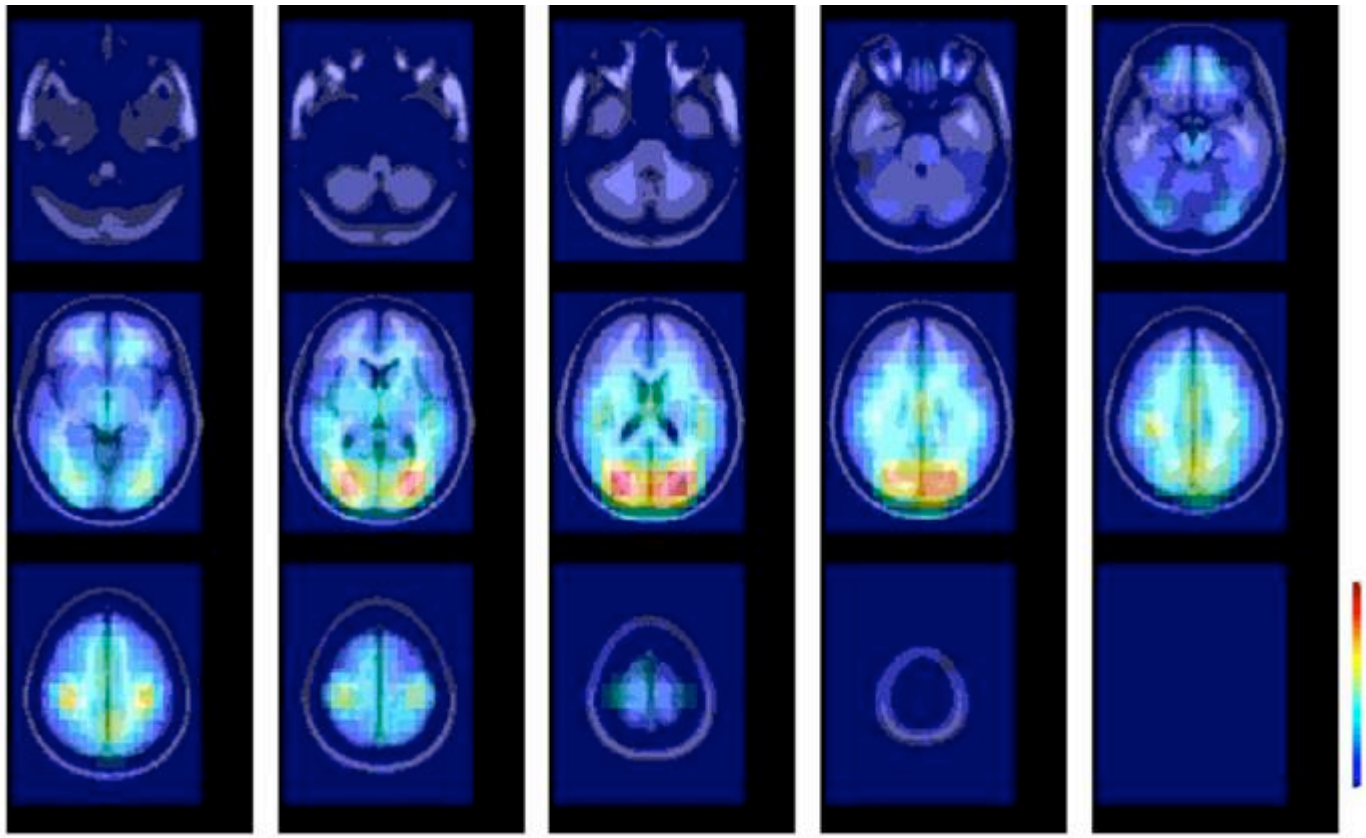
Equivalent dipole density



Auditory oddball plus novel sounds

>> dipoledensity()

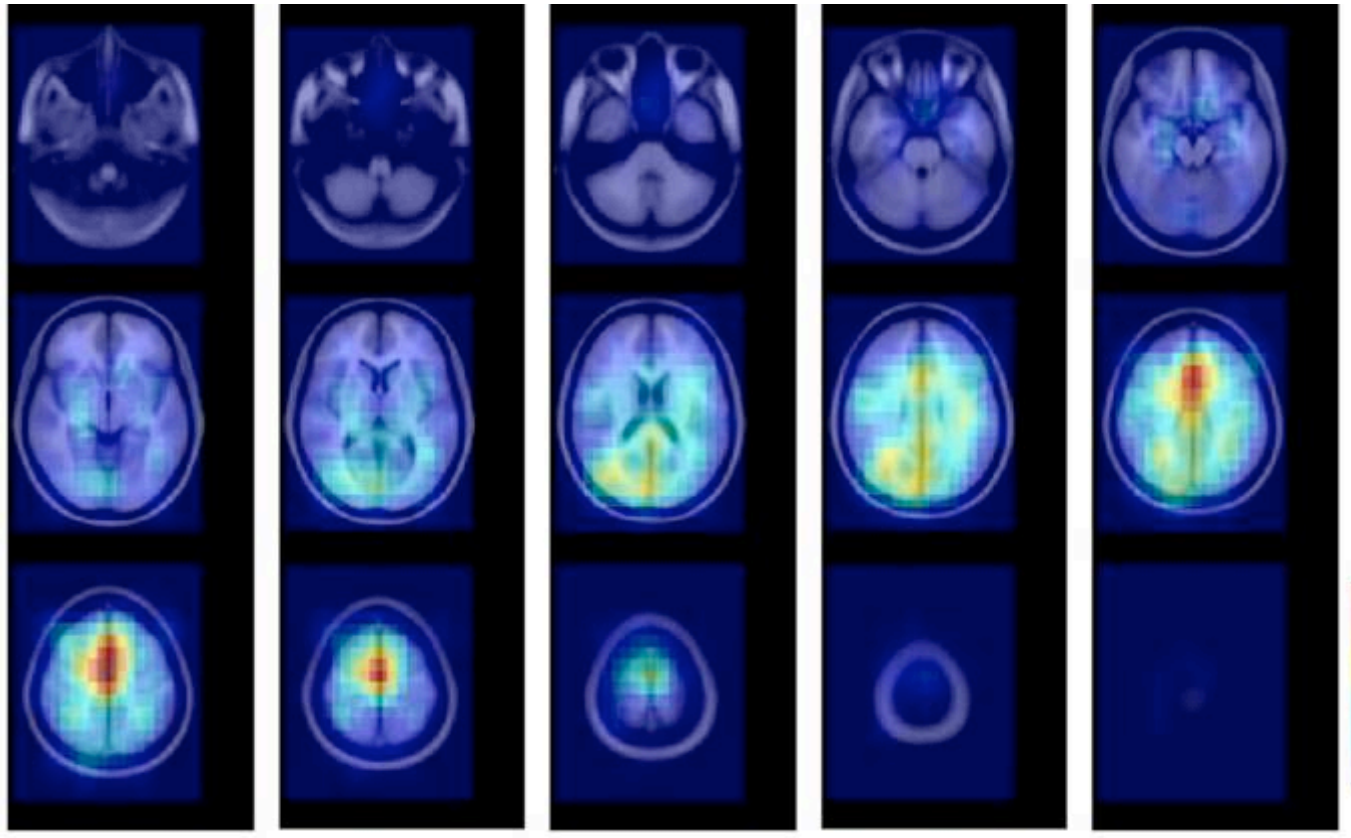
Equivalent dipole density



Emotion imagery task

>> dipoledensity()

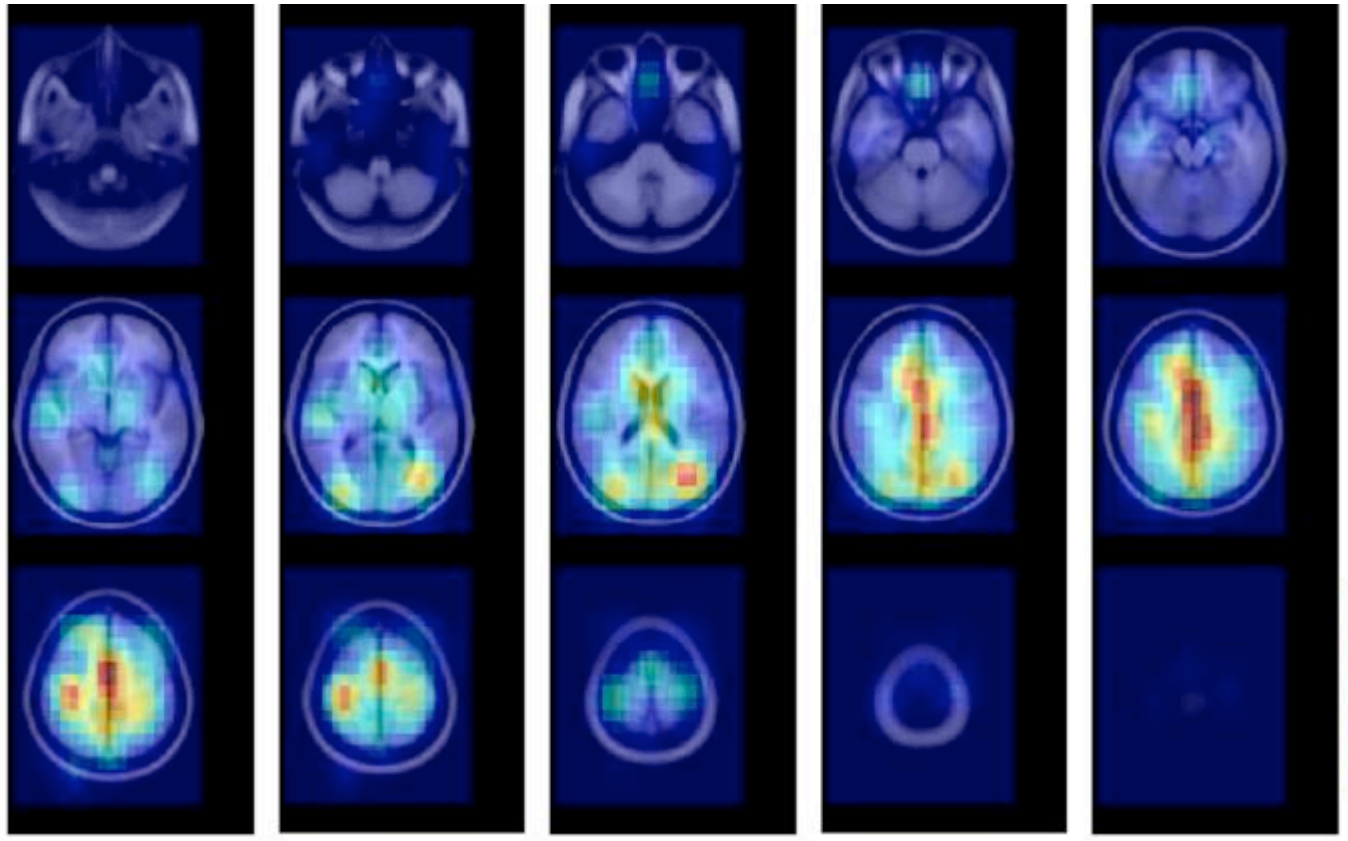
Equivalent dipole density Exp I



Word memory (old/new) task

>> dipoledensity()

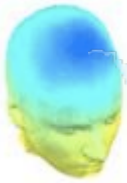
Equivalent dipole density Exp II



Visually cued button press task

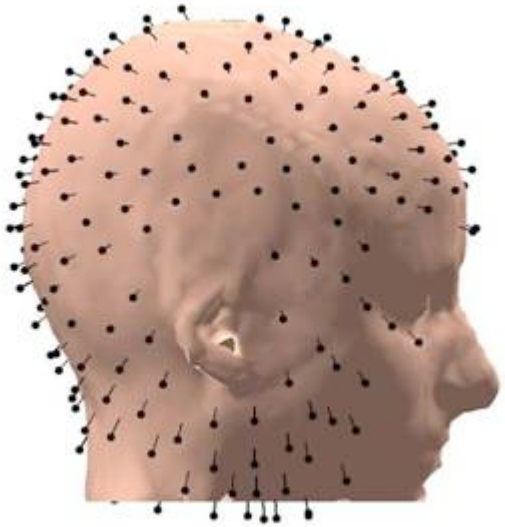
>> dipoledensity()

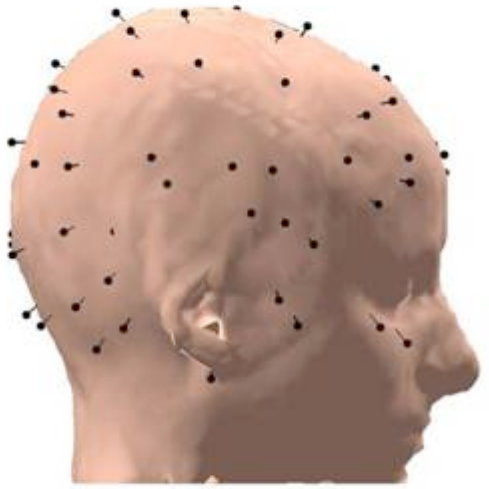
… Some caveats



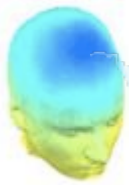
In this preliminary study …

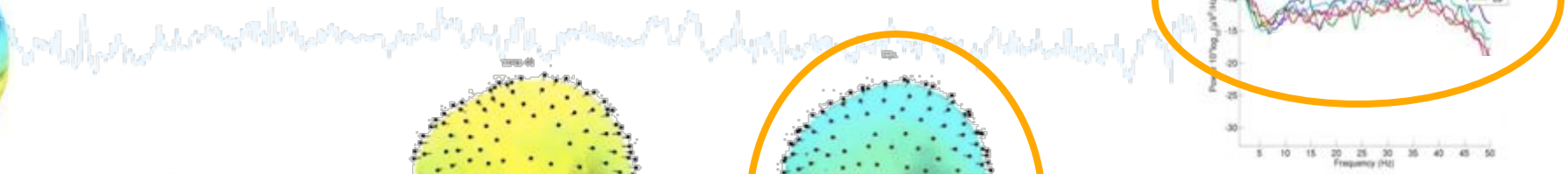
- The electrode locations were not individualized.
- MR images were not available → co-registration crude.
- Single versus dual-dipole model selection was subjective.
- Different electrode montages → possible location effects

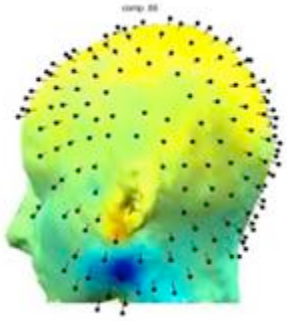


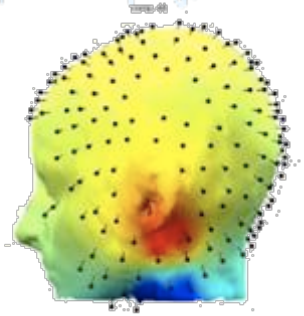


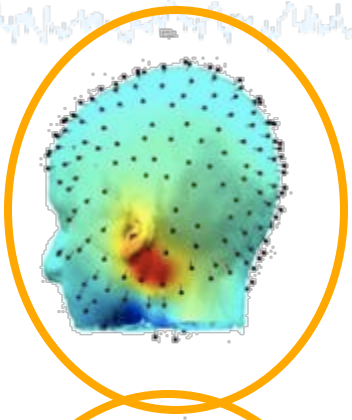
Clustering by spectra (1 subject)

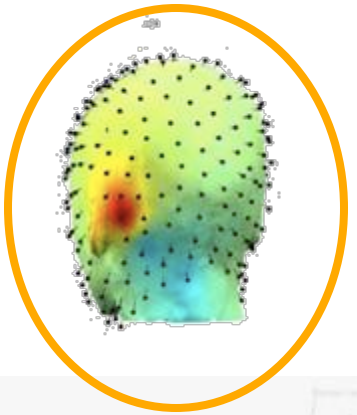


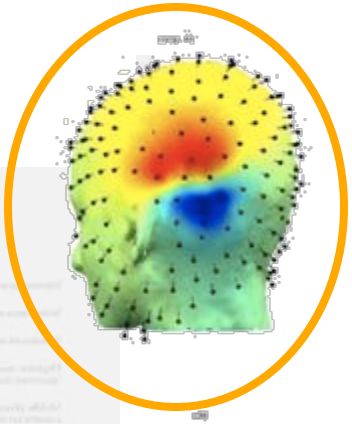


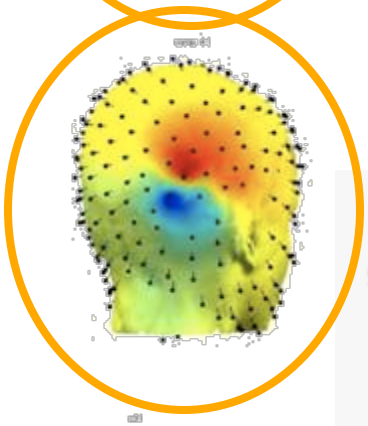


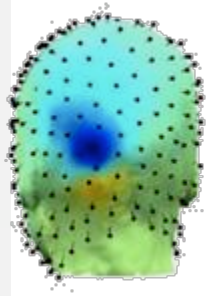


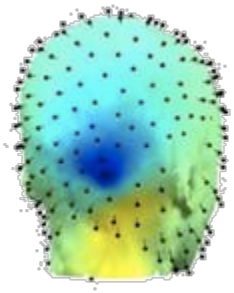


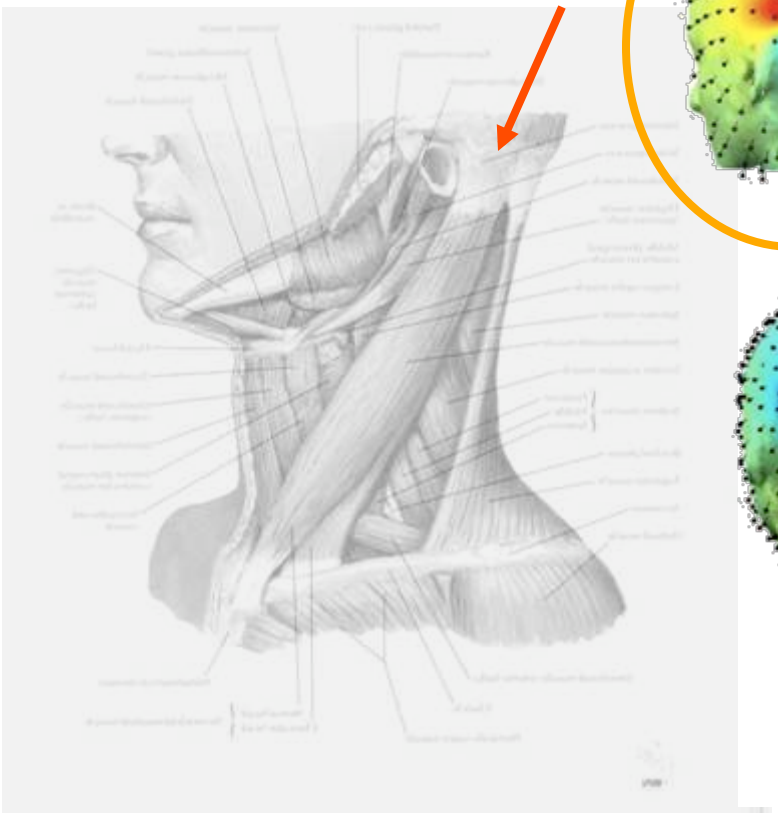


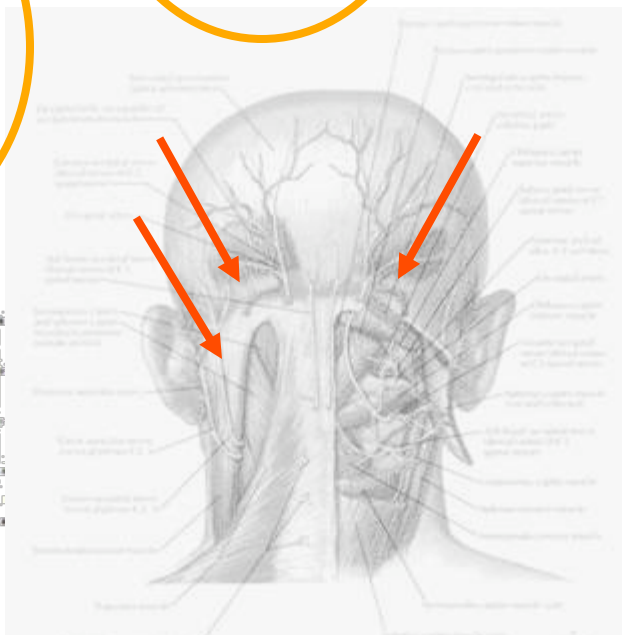






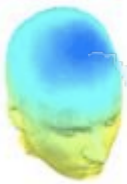






e, 2009, Aspet, France: Scott Makeig – Compone

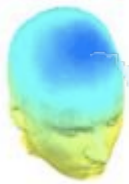
Makeig et al., *unpublished*

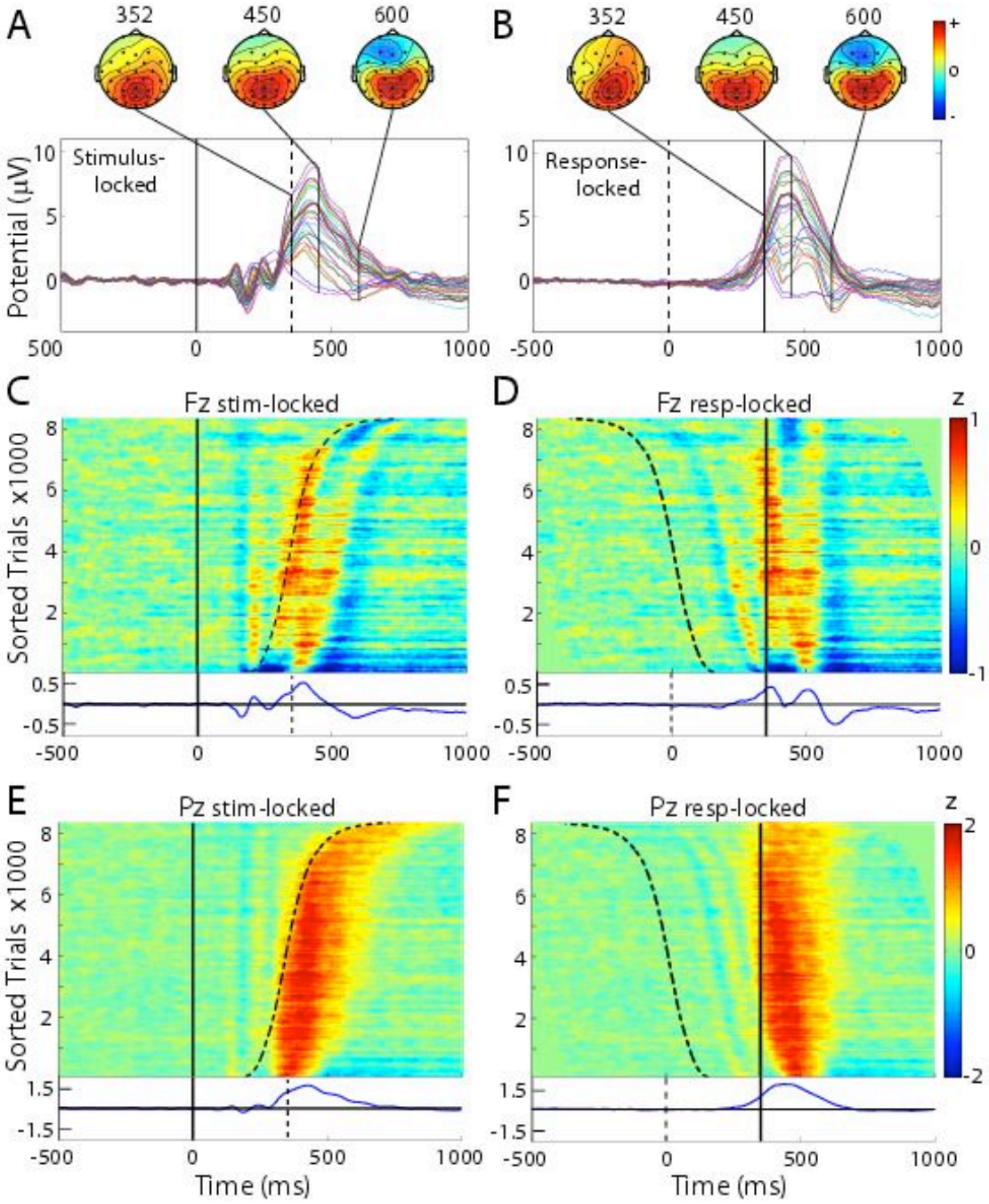


Visual Selective Attention Task

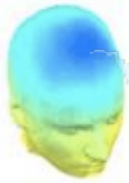
+

15 subjects



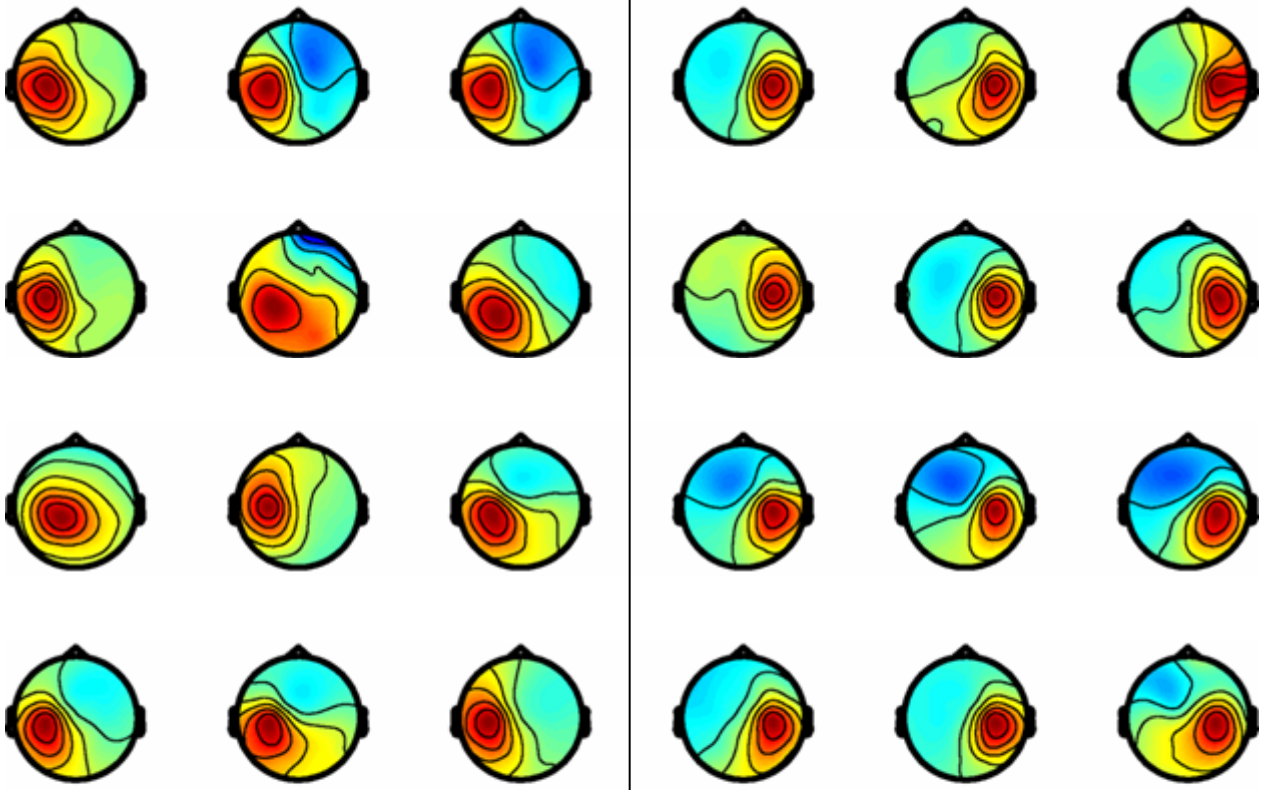

A
352
450
600
Stimulus-locked
Potential (μV)
B
Response-locked
C
Fz stim-locked
Sorted Trials x1000
D
Fz resp-locked
z
E
Pz stim-locked
F
Pz resp-locked
Time (ms)

Clustering ICA components by eye



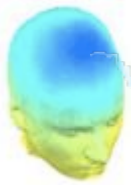
Left mu

Right mu



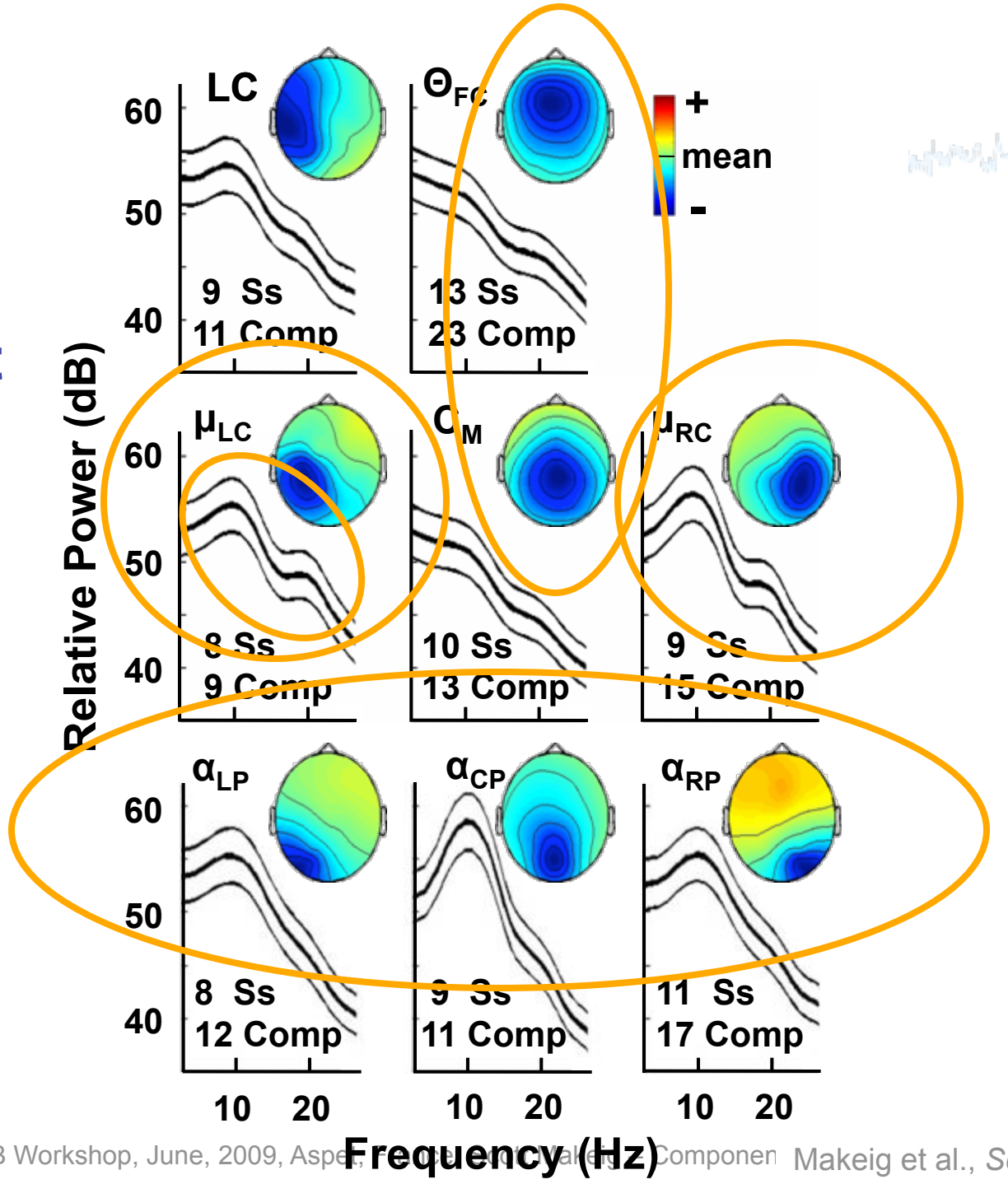
Makeig et al., *submitted*

Semi-automated clustering

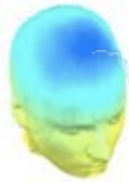


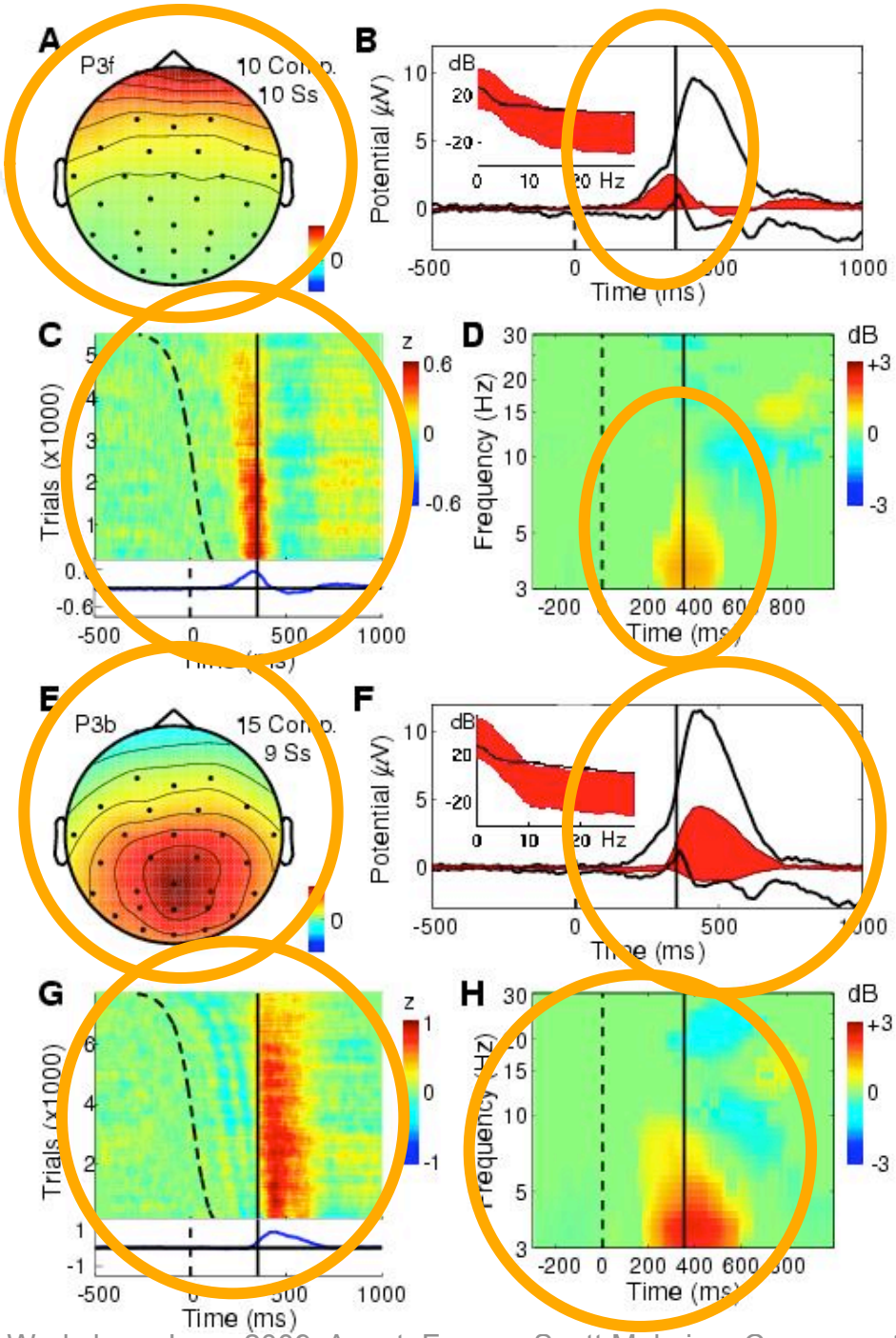
- Clustered components from 15 Ss using a ‘component distance metric’ incorporating differences between their (weighted) scalp maps, dipole locations, spectra, ERP, ERSP, and ITC patterns.
- Hand-adjusted clusters to remove outliers.
- Determined time/frequency regions of significant ERSP and ITC for each component using permutation-based statistics.
- Used binomial statistics to highlight time/frequency regions significantly active within clusters.

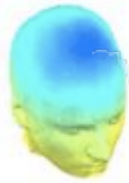
N1 Study Component Clusters

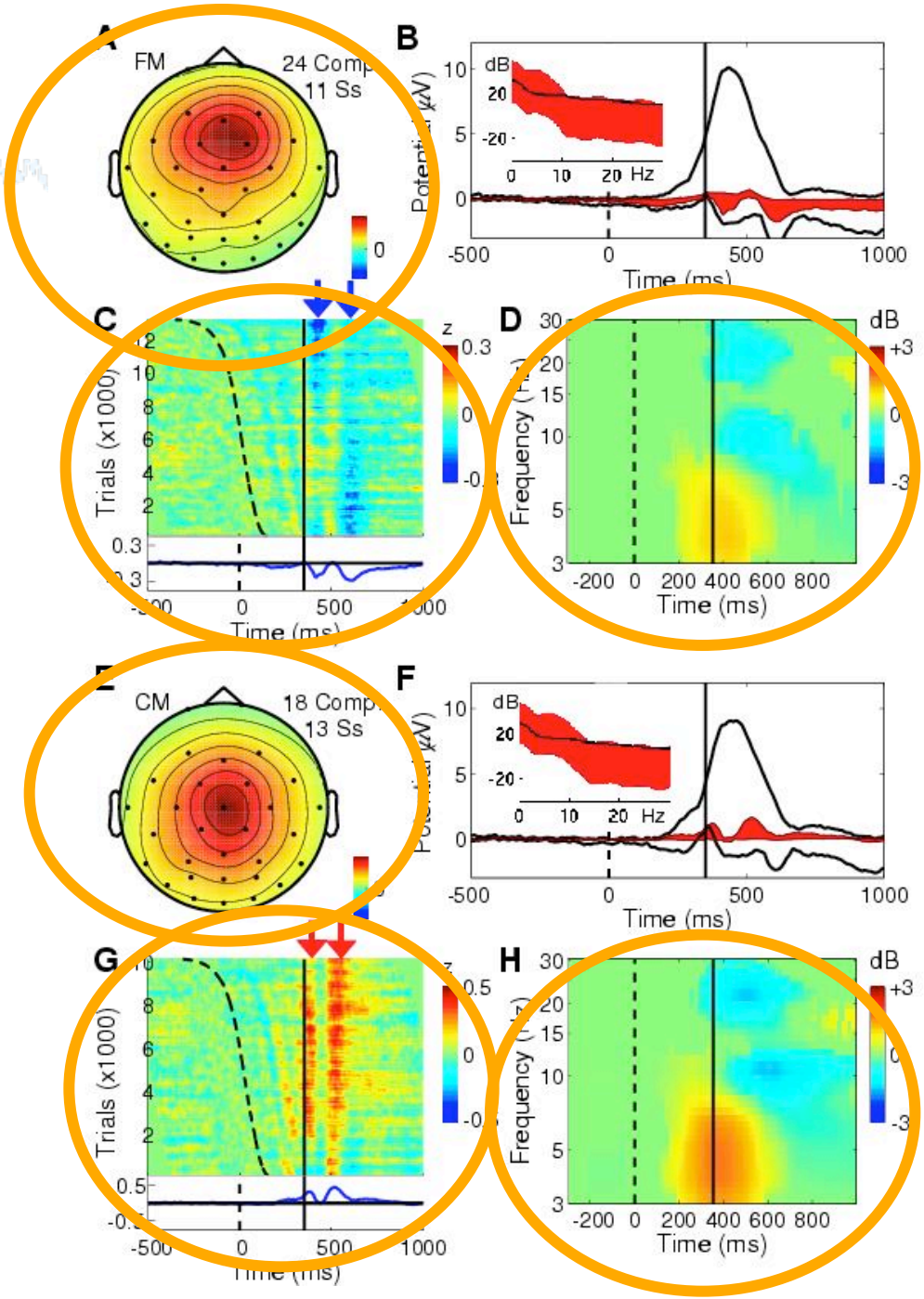


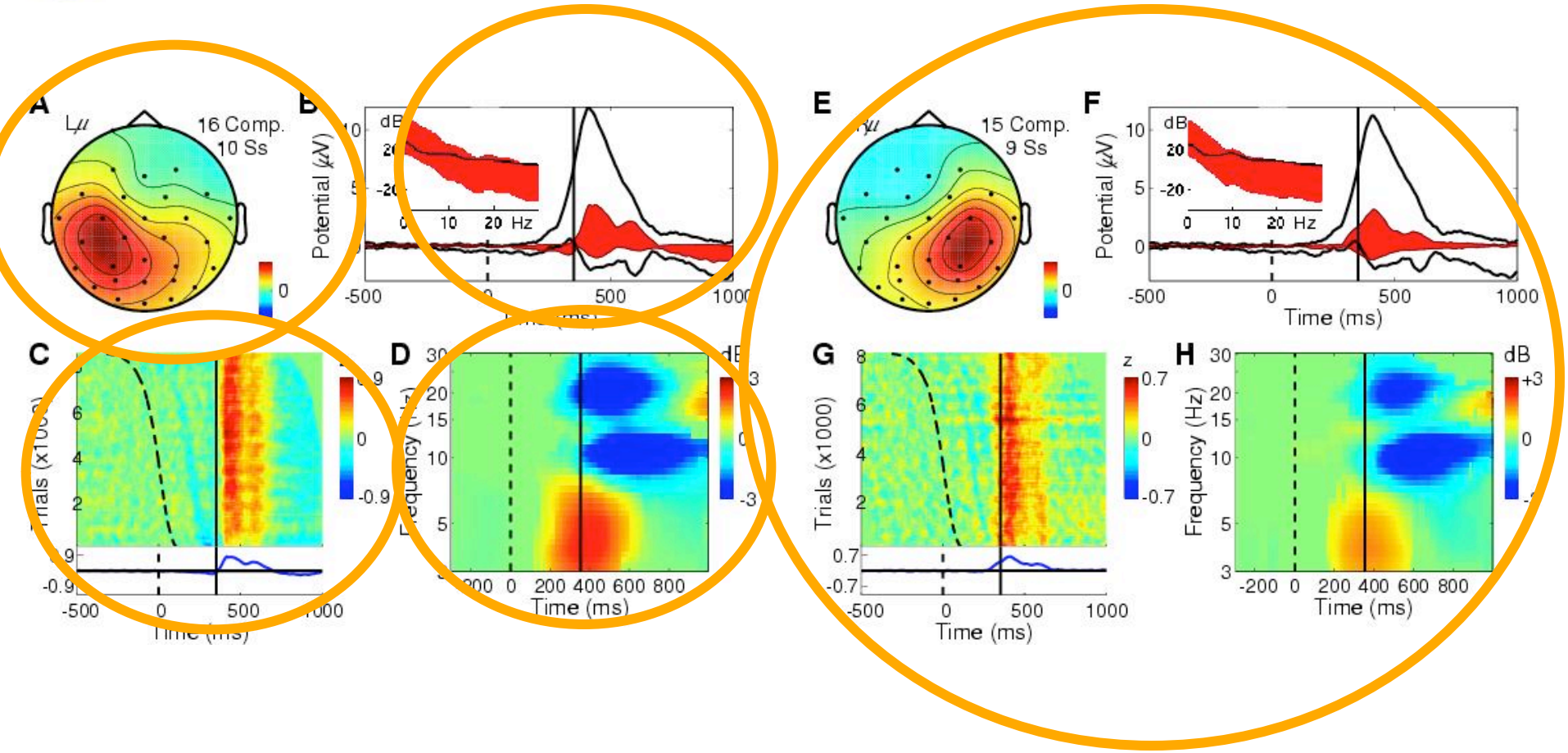
Makeig et al., *Science* 2002

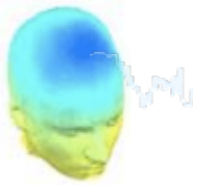


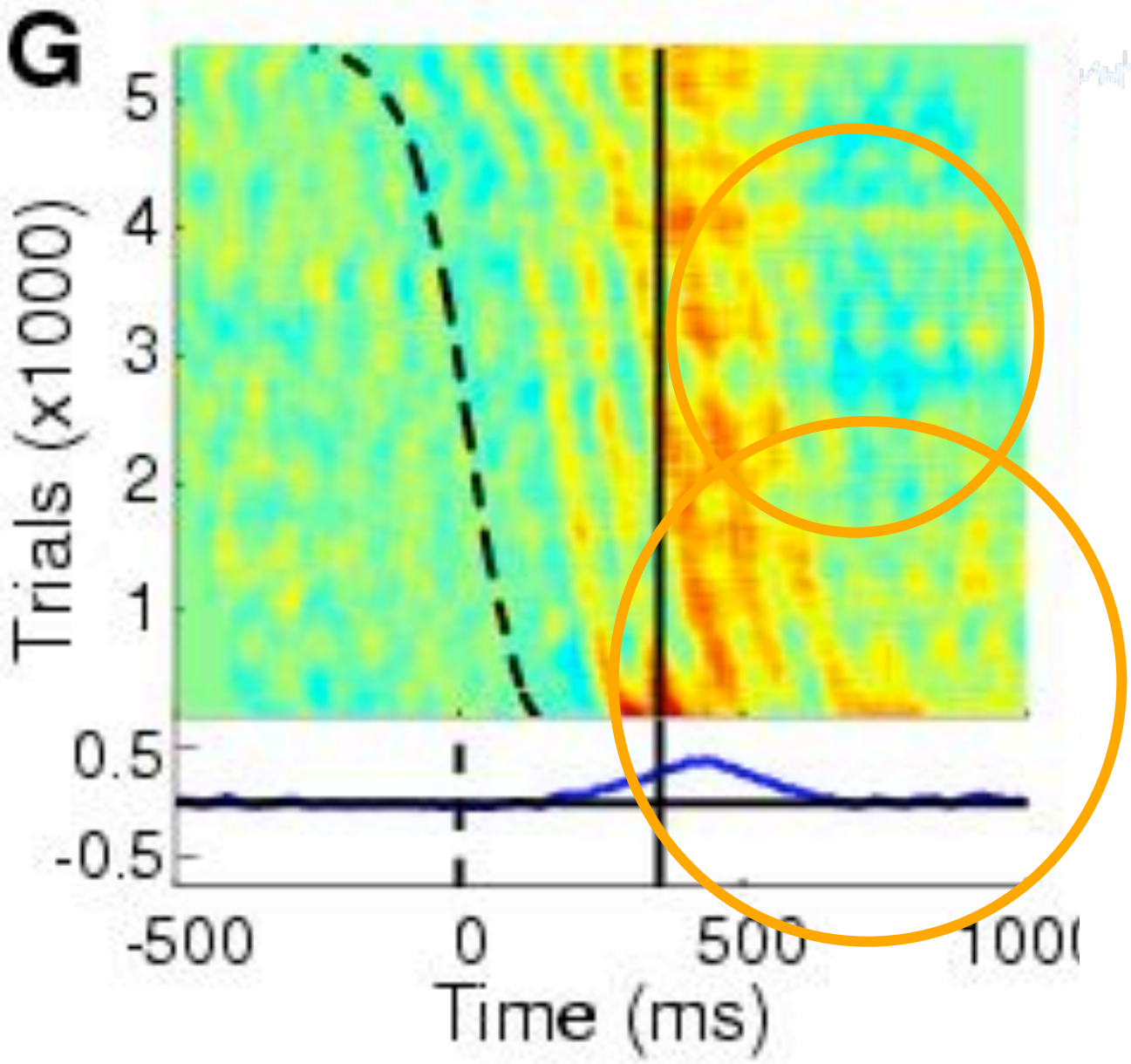

A
P3f
10 Comp.
10 Ss
B
dB
Potential (μV)
Time (ms)
C
Trials (x1000)
z
D
Frequency (Hz)
dB
Time (ms)
E
P3b
15 Comp.
9 Ss
F
dB
Potential (μV)
Time (ms)
G
Trials (x1000)
z
Time (ms)
H
Frequency (Hz)
dB
Time (ms)

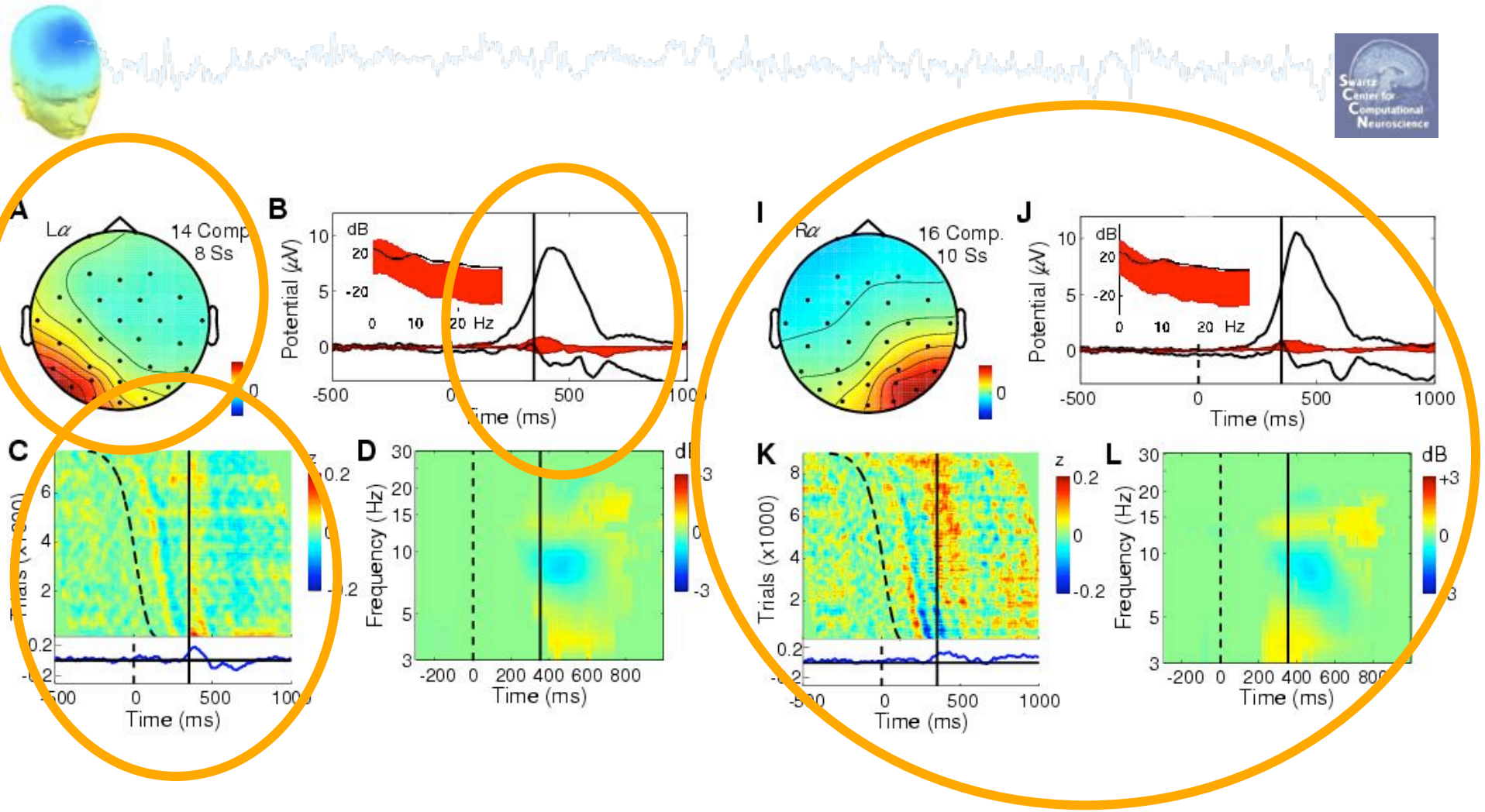



A
FM
24 Comp
11 Ss
0
B
dB
20
-20
0 10 20 Hz
10
5
0
-500 0 500 1000
Time (ms)
C
z
0.3
0
Trials (x1000)
12
10
8
6
4
2
0.3
-500 0 500 1000
Time (ms)
D
dB
+3
0
-3
Frequency (Hz)
30
20
15
10
5
3
-200 0 200 400 600 800
Time (ms)
E
CM
18 Comp
13 Ss
F
dB
20
-20
0 10 20 Hz
10
5
0
-500 0 500 1000
Time (ms)
G
z
0.5
0
Trials (x1000)
10
8
6
4
2
0.5
-500 0 500 1000
Time (ms)
H
dB
+3
0
-3
Frequency (Hz)
30
20
15
10
5
3
-200 0 200 400 600 800
Time (ms)

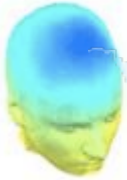

A
16 Comp.
10 Ss
B
Potential (μV)
dB
10 20 Hz
-500
500
C
Trials (x1000)
-500
0
500
Time (ms)
D
Frequency (Hz)
-200 0 200 400 600 800
Time (ms)
E
15 Comp.
9 Ss
F
Potential (μV)
dB
10 20 Hz
-500
0
500
1000
Time (ms)
G
Trials (x1000)
z
0.7
0
-0.7
-500
0
500
1000
Time (ms)
H
Frequency (Hz)
dB
+3
0
-200 0 200 400 600 800
Time (ms)

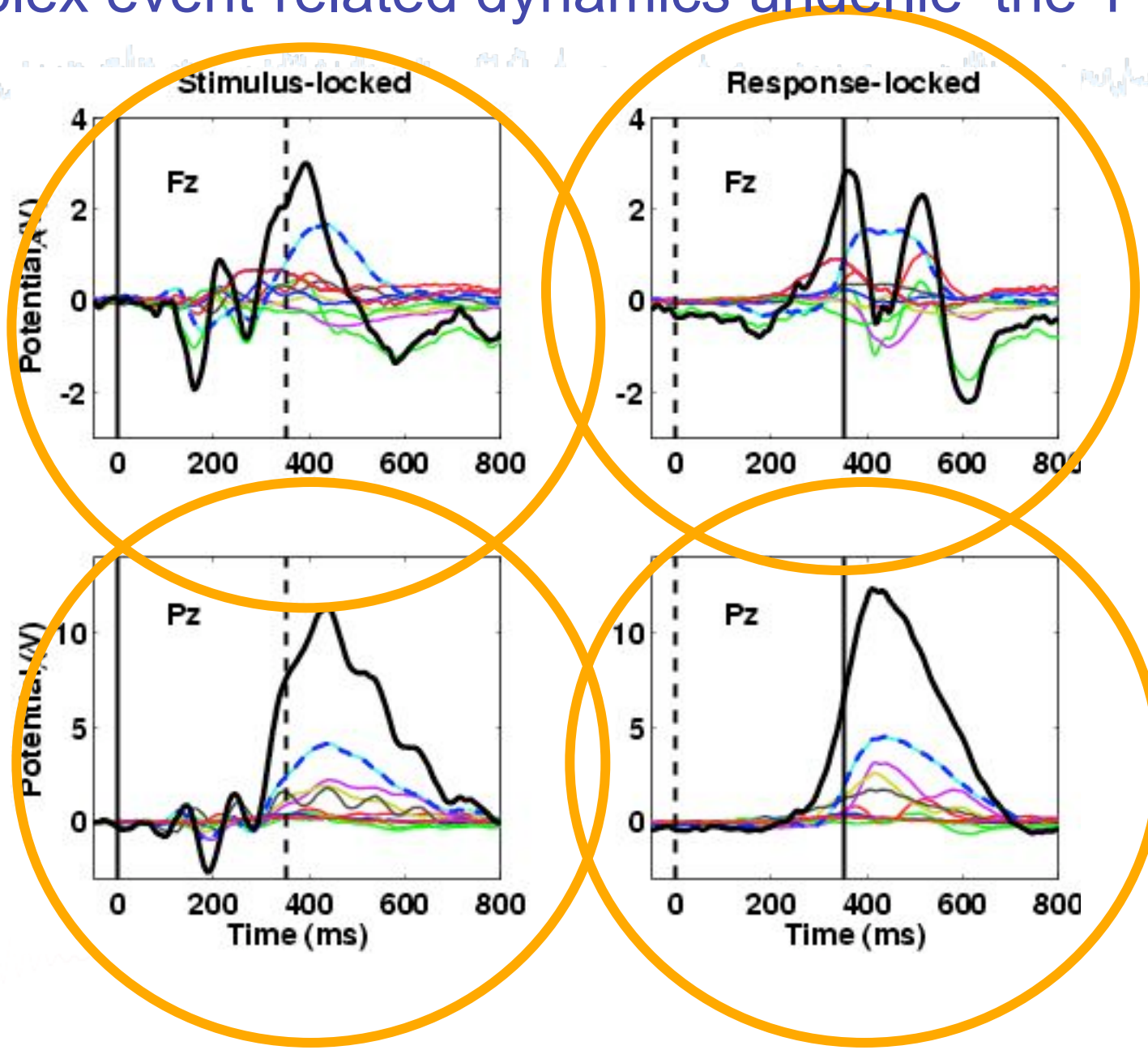



G
Trials (x1000)
5
4
3
2
1
0.5
-0.5
-500
0
500
Time (ms)

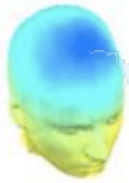

A
Lα
14 Comp.
8 Ss
B
dB
20
-20
0 10 20 Hz
Potential (μV)
10
5
0
-500
0
500
Time (ms)
C
Trials (x1000)
6
4
2
0.2
-500
0
500
Time (ms)
z
0.2
0
D
Frequency (Hz)
30
20
15
10
5
3
-200 0 200 400 600 800
Time (ms)
I
Rα
16 Comp.
10 Ss
0
J
dB
20
-20
0 10 20 Hz
Potential (μV)
10
5
0
-500
0
500
1000
Time (ms)
K
Trials (x1000)
8
6
4
2
0.2
0
500
1000
Time (ms)
z
0.2
0
-0.2
L
Frequency (Hz)
30
20
15
10
5
3
-200 0 200 400 600
Time (ms)
dB
+3
0

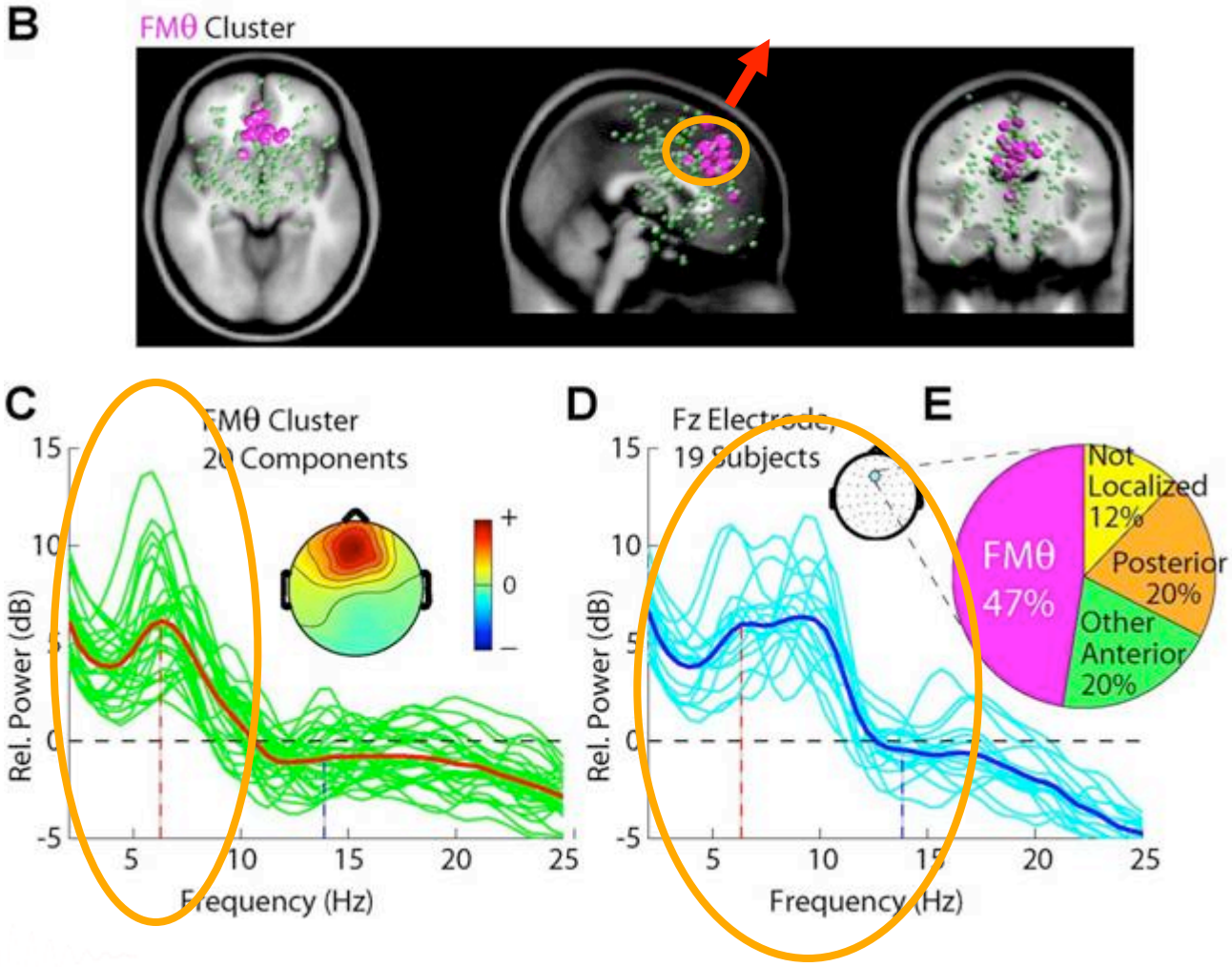
Complex event-related dynamics underlie ‘the’ P300





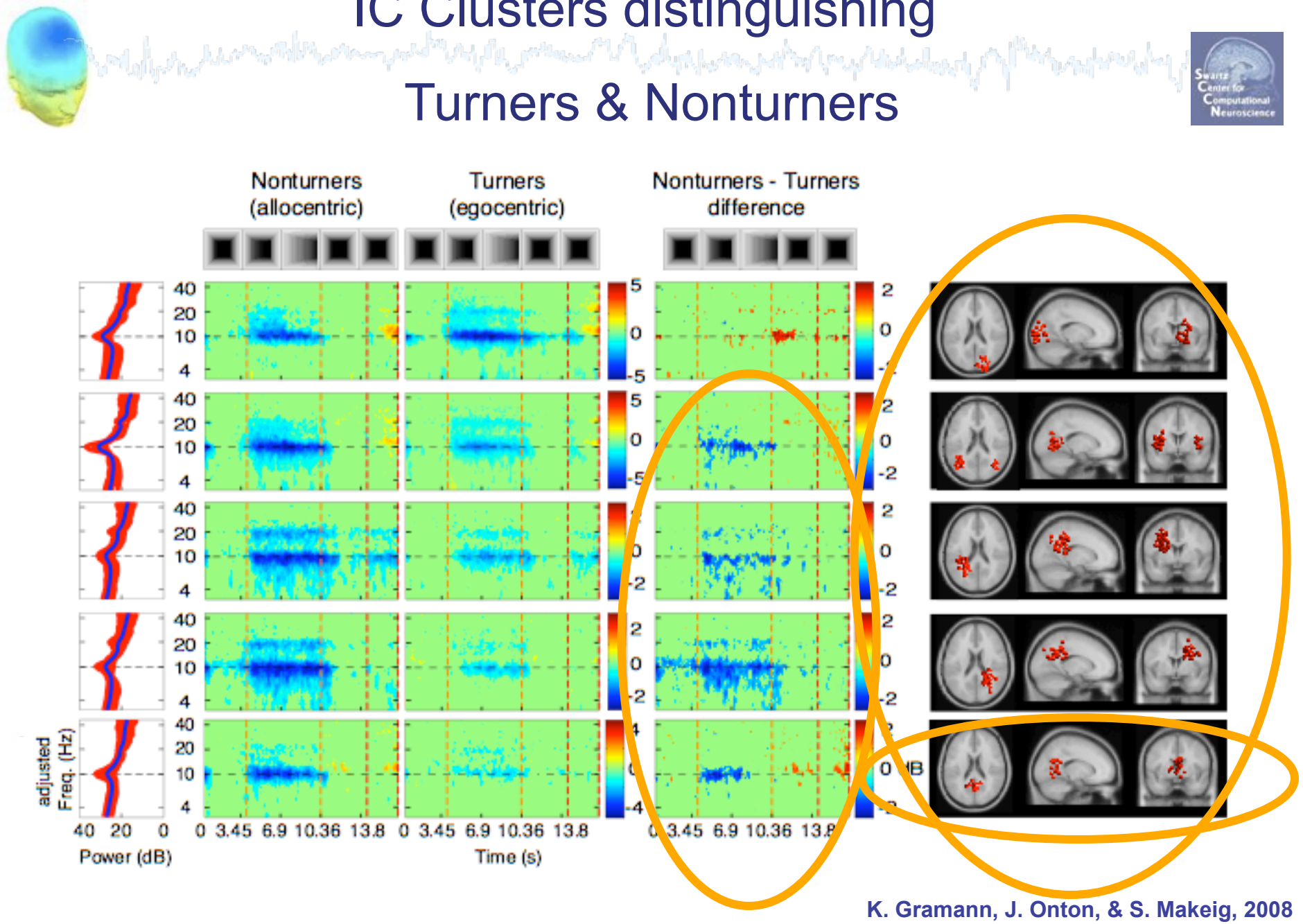
A FMΘ cluster during working memory





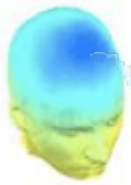
Onton et al., *NeuroImage* 2005

IC Clusters distinguishing Turners & Nonturners



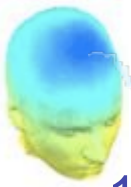
K. Gramann, J. Onton, & S. Makeig, 2008

Are obtained component clusters “real“?



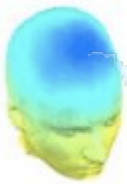
- **Naïve realism** (a.k.a. “expertise”)
  - “Yes! … because I know one when I see one!”
    - “If it appears where Mu components appear,
      and acts like Mu components act,
      then it IS a Mu component!”
- **Convergent evidence** (a.k.a., “double-checking”)
  - Two possible approaches:
    - Cluster on PLACE → Check ACTIVITY consistency (re task)
    - Cluster on ACTIVITY → Check PLACE consistency
- **Absolute truth:**
    - More ideal forward and inverse models
    - Invasive multiscale recordings + modeling

EEGLAB clustering procedure

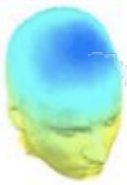


1. Identify a set of datasets as an EEGLAB **study** or '**studyset**'.
2. Specify the subject **group**, **subject** code, **condition** and **session** of each dataset in the study.
3. Identify **components to cluster** in each study dataset.
4. Decide on **component measures** to use in clustering 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. Perform **signal processing** within or between selected clusters.

EEGLAB 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 the **clustered study set data** as input into **std_** functions.



Need ICs from all subjects be included in each cluster?

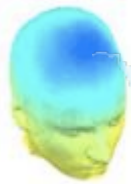
Not all subjects contribute components to each cluster.

Why not?

- Different numbers of artifact components (~INR)
- Subject differences!?
- Is my subject group a Gaussian cloud??

  → 'subject space'

Beyond Clustering



Physiologic Data

Multiple Modes

Distributed Dynamics

Perturbations Modulations

Cognitive Events

Anticipation

Re-evaluation

Imagination

Value

Model

Individual Subjects

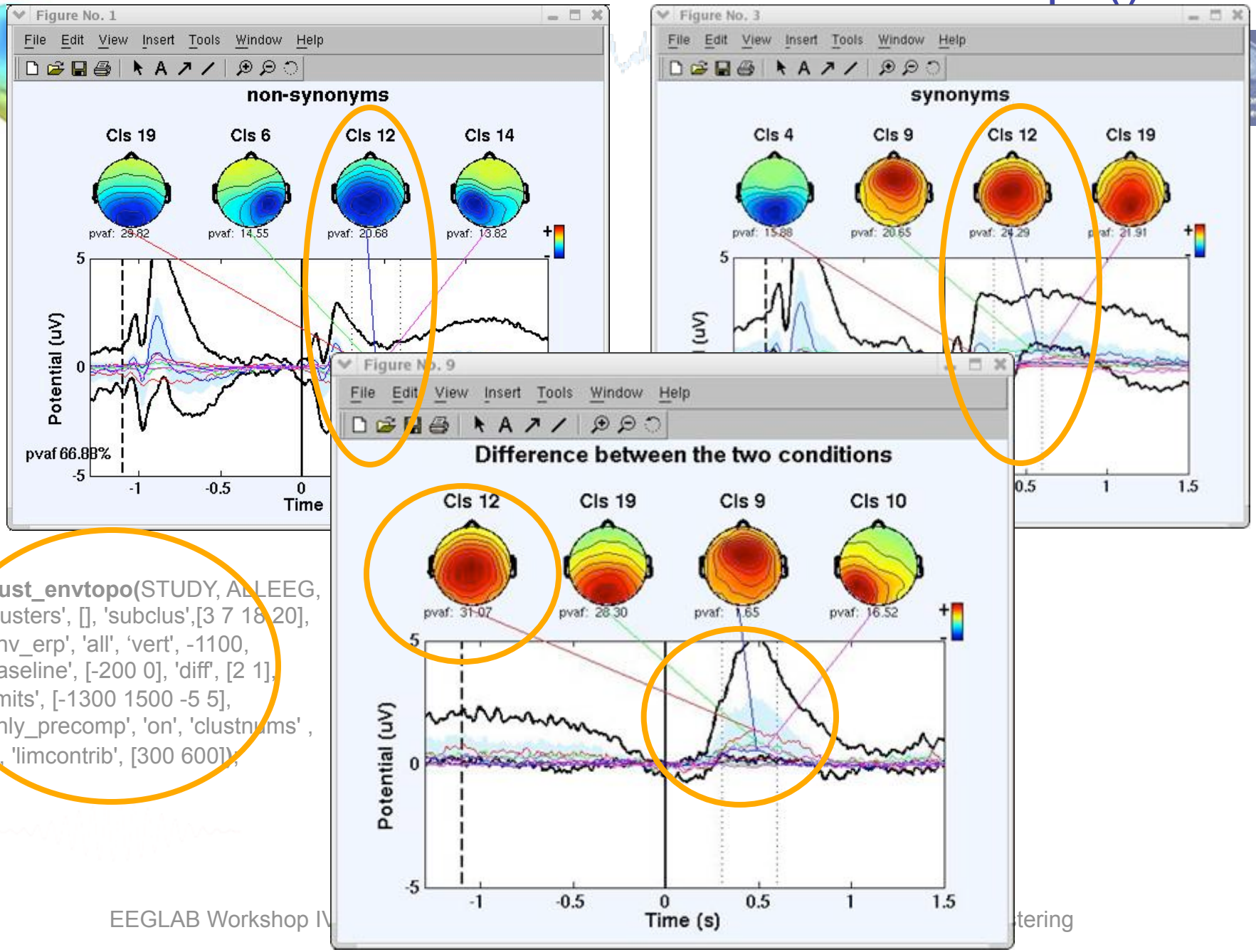
Culture

Personality

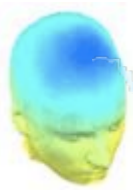
Genetics

Age / Gender

Cluster ERP contributions - clust_envtopo()



clust_envtopo(STUDY, ALLEEG,
'clusters', [], 'subclus',[3 7 18 20],
'env_erp', 'all', 'vert', -1100,
'baseline', [-200 0], 'diff', [2 1],
'limits', [-1300 1500 -5 5],
'only_precomp', 'on', 'clustnums' ,
4, 'limcontrib', [300 600]);



Component clustering research issues

Research Issues:

- Alternative clustering methods (soon)
- Clustering goodness-of-fit
- Cluster plotting details
- Add new pre-clustering measures
- Study subject differences!?

Development Plan:

- New EEGLAB 'STUDY.DESIGN' definition:
  - Vector of conditions → Uncoupled experimental STUDY.DESIGN structures, each with a vector of included conditions.
  - More flexible and complete STUDY.DESIGN statistics