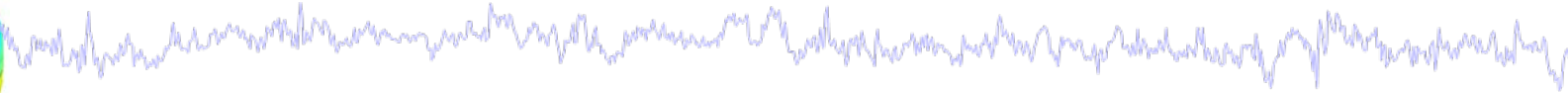# STUDY clustering overview

STEP 1

Build a STUDY

STEP 2

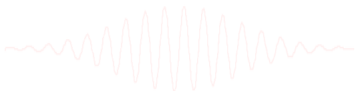Precompute the data

STEP 3

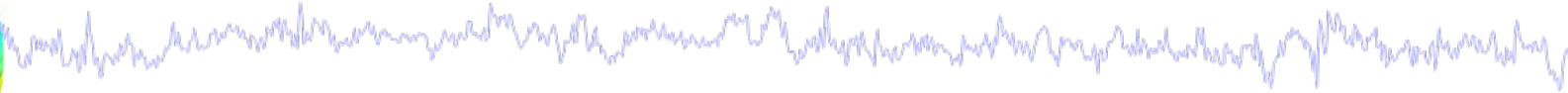Precluster the data

STEP 4

Cluster the data

Exercise...

STUDY clustering overview

STEP 1

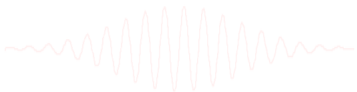   Build a STUDY

STEP 2

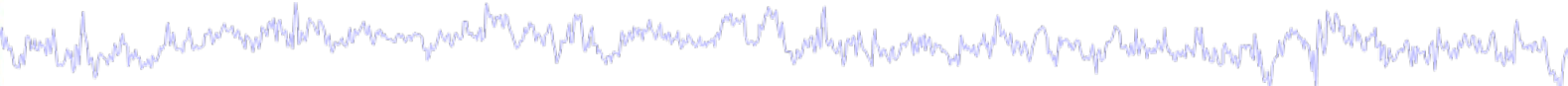   Precompute the data

STEP 3

      Precluster the data

STEP 4

      Cluster the data

Exercise...

Memory options

Memory options should change when using STUDY vs single dataset

Build a STUDY

Build a STUDY, cont'd

Edit dataset info

ICs to cluster

Build a STUDY

Most important option:

- Allows only one dataset to be loaded at once.
- Most STUDYs are too big to have all data loaded at once.

```
% Open eeglab:
[ALLEEG EEG CURRENTSET ALLCOM] = eeglab;

% Set memory options:
pop_editoptions( 'option_storedisk', 1, 'option_savetwofiles', 1,…
 'option_saveica', 1, 'option_single', 0, 'option_memmapdata',…
 0, 'option_computeica', 1, 'option_scaleicarms', 0,…
'option_rememberfolder', 1);
% saves a file 'eeg_options.m' to your current working directory

% Initialize EEGLAB/STUDY variables:
STUDY = []; CURRENTSTUDY = 0; ALLEEG=[]; EEG=[]; CURRENTSET=[];
```

Build a STUDY

```
% Faster alternative to building a STUDY manually
% Example STUDY:  13  subjects, 3 conditions

% Define variables:

basedir = 'C:\...\EEGLAB_WORKSHOP\STUDY\';
setnames = {'Memorize.set','Ignore.set','Probe.set'};
subjs = {'S01','S02','S03','S04','S05','S06','S07',…
'S08','S09','S10','S11','S12','S13'};
studyname = 'Sternberg';
taskname = 'Sternberg';
savename = 'stern.study';
```
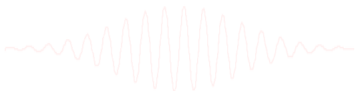
Define variables
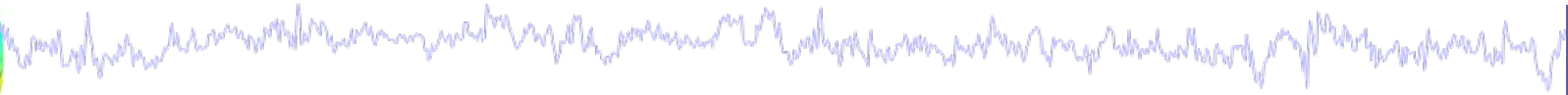
```
% concatenate string variables:

[ ] % strings inside brackets will be concatenated

datset = [basedir,subjs{subj},'\',setnames{cond}];



    C:\EEGLAB_Workshop\STUDY\S01\Memorize.set
```
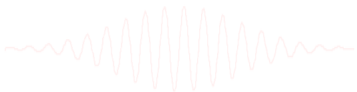
Load dataset info from commandline

```matlab
% Now loop through subjects and add to the STUDY:
index = 1; % initialize STUDY index
for subj = 1:length(subjs) % for each subject
    for cond = 1:length(setnames) % for each condition
        datset = [basedir,subjs{subj},'\',setnames{cond}];
        [STUDY ALLEEG] = std_editset( STUDY, ALLEEG,…
                    'name', studyname, 'task', taskname,...
                    'commands',{{'index',index,'load',datset},…
                    {'dipselect',0.15},{'subject',subjs{subj}},…
                    {'condition',conds{cond}}},…
                    'inbrain','on','updatedat','off',…
                    'savedat', 'off','filename', [basedir, savename]);
        index = index + 1;
        CURRENTSTUDY = 1; EEG = ALLEEG; CURRENTSET = [1:length(EEG)];
        [STUDY, ALLEEG] = std_checkset(STUDY, ALLEEG);
    end;
end;
eeglab redraw
```

STUDY structure

```
STUDY =

              name: 'Sternberg'
              task: 'Sternberg'
       datasetinfo: [1x39 struct]
             notes: ''
          filename: 'stern.study'
          filepath: 'C:\Users\julie\Documents\Workshops\Finland\STUDY'
           history: [1x7332 char]
           subject: {1x13 cell}
             group: {''}
           session: []
         condition: {'ignore'  'memorize'  'probe'}
            setind: [3x13 double]
               etc: [1x1 struct]
          preclust: [1x1 struct]
           cluster: [1x1 struct]
           changrp: [1x71 struct]
             saved: 'yes'

      >>
```
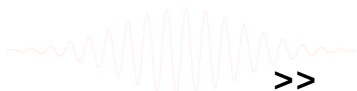
Subject info in STUDY structure

```
>> STUDY.datasetinfo


ans =


1x39 struct array with fields:

      filepath

      filename

      subject

      session

      condition

      group

      index

      comps


>>
```

Gives information for each dataset of each subject
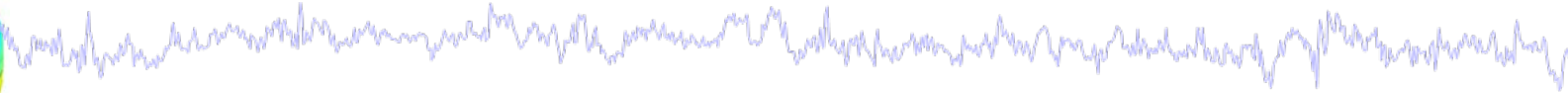
Each DATASET (NOT each subject) has a unique index
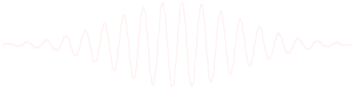
STUDY clustering overview

STEP 1

Build a STUDY

STEP 2

Precompute the data

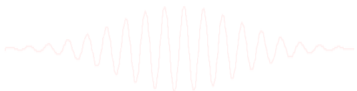**STEP 3**

Precluster the data

STEP 4

Cluster the data

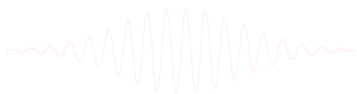**Exercise...**

Precompute data measures

Precompute data measures

Select and compute component measures for later clustering -- pop_precomp()

Pre-compute component measures for STUDY 'Sternberg'

☑ Compute ERP/spectrum/ERSP only for components selected by RV (set) or for all components (unset)

List of measures to precompute

☑ ERPs	Baseline ([min max] in ms)	[-200 0]	
☑ Power spectrum	Spectopo parameters		Test
☑ ERSPs	Time/freq. parameters	'cycles', [3 0.5], 'nfreqs', 100	Test
☑ ITCs			
☑ Scalp maps			

☐ Recompute even if present on disk
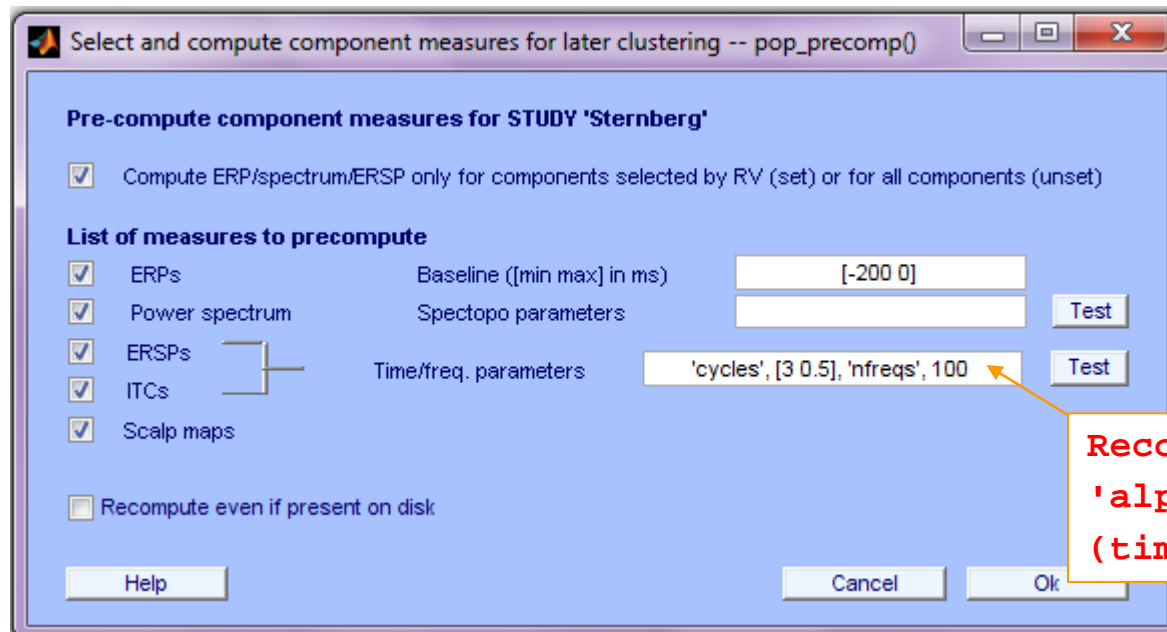
Recommend:
'alpha',.01
(time-consuming)

Help   Cancel   Ok

```
[STUDY ALLEEG] = std_precomp(STUDY, ALLEEG,  'components',…
'erp', 'on', 'rmbase',[-200 0] , 'scalp', 'on', 'spec',…
'on', 'specparams',{}, 'ersp', 'on', 'erspparams',…

{'cycles',[3 0.5] , 'nfreqs',100, 'freqs',[3 70] ,…
'alpha',0.01}, 'itc', 'on');
```

STUDY clustering overview

STEP 1
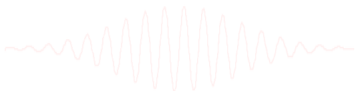
    Build a STUDY

STEP 2

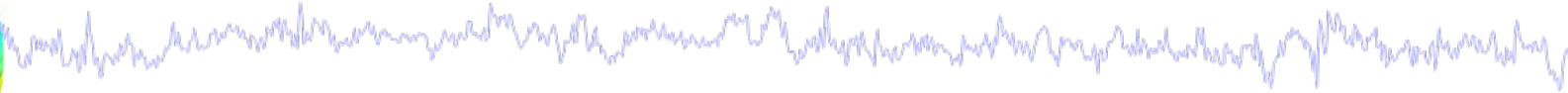   Precompute the data

STEP 3

    Precluster the data
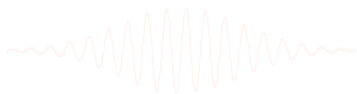
STEP 4

    Cluster the data

Exercise...

Precluster the data

Precluster the data
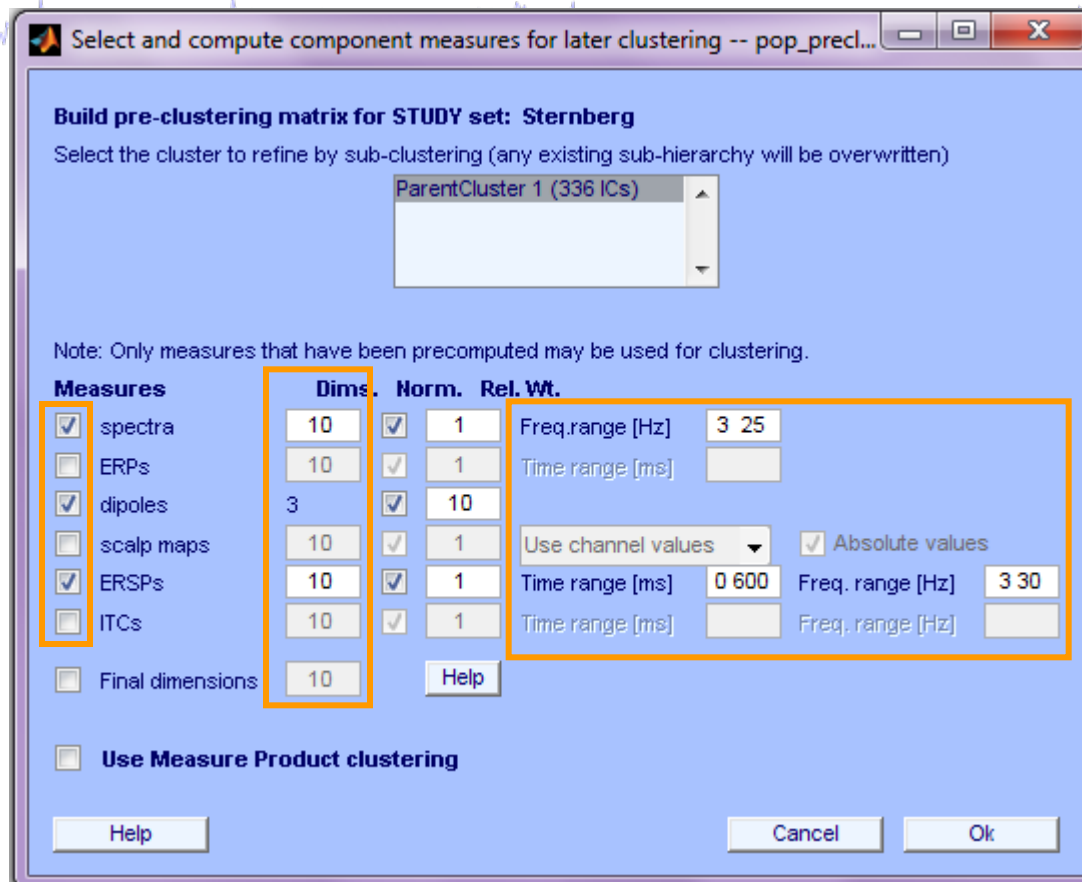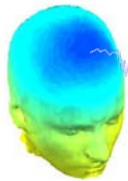

```
parentclust = 1; % cluster 1 is always full parent cluster
[STUDY ALLEEG] = std_preclust(STUDY, ALLEEG, parentclust,{'spec','npca',5,…
    'norm',1,'weight',1,'freqrange',[3 25]},{'erp','npca',6,'norm',1,…
    'weight',1, 'timewindow',[0 400]},{'scalp','npca',10,'norm',1,'weight',1,…
    'abso',1},{'dipoles','norm',1,'weight',10},{'ersp','npca',20,…
    'freqrange',[3 30] ,'timewindow',[0 600],'norm',1,'weight',1},{'itc',…
    'npca',6,'freqrange',[3 30],'timewindow',[0 400] , 'norm',1, 'weight',1});
```
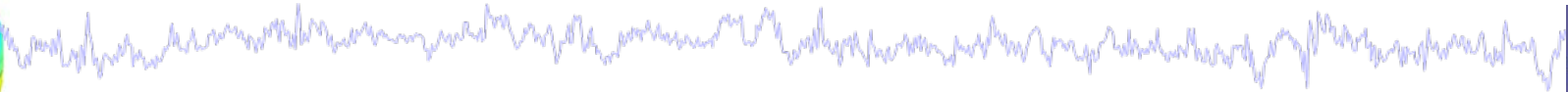
Choosing data measures

What measure(s) should you use?

• It depends on your final cluster criteria…

     - If for example, your priority is dipole location,

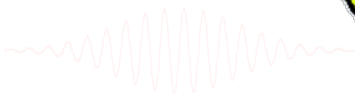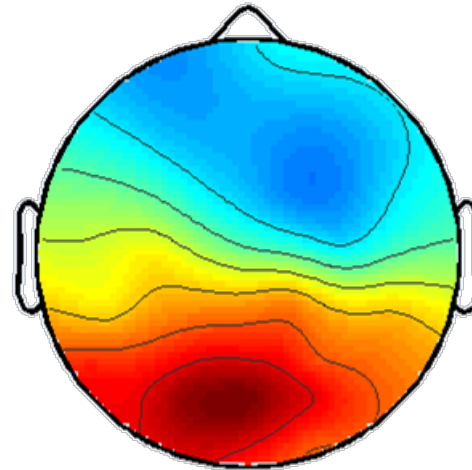   then cluster only based on dipole location…

But consider:

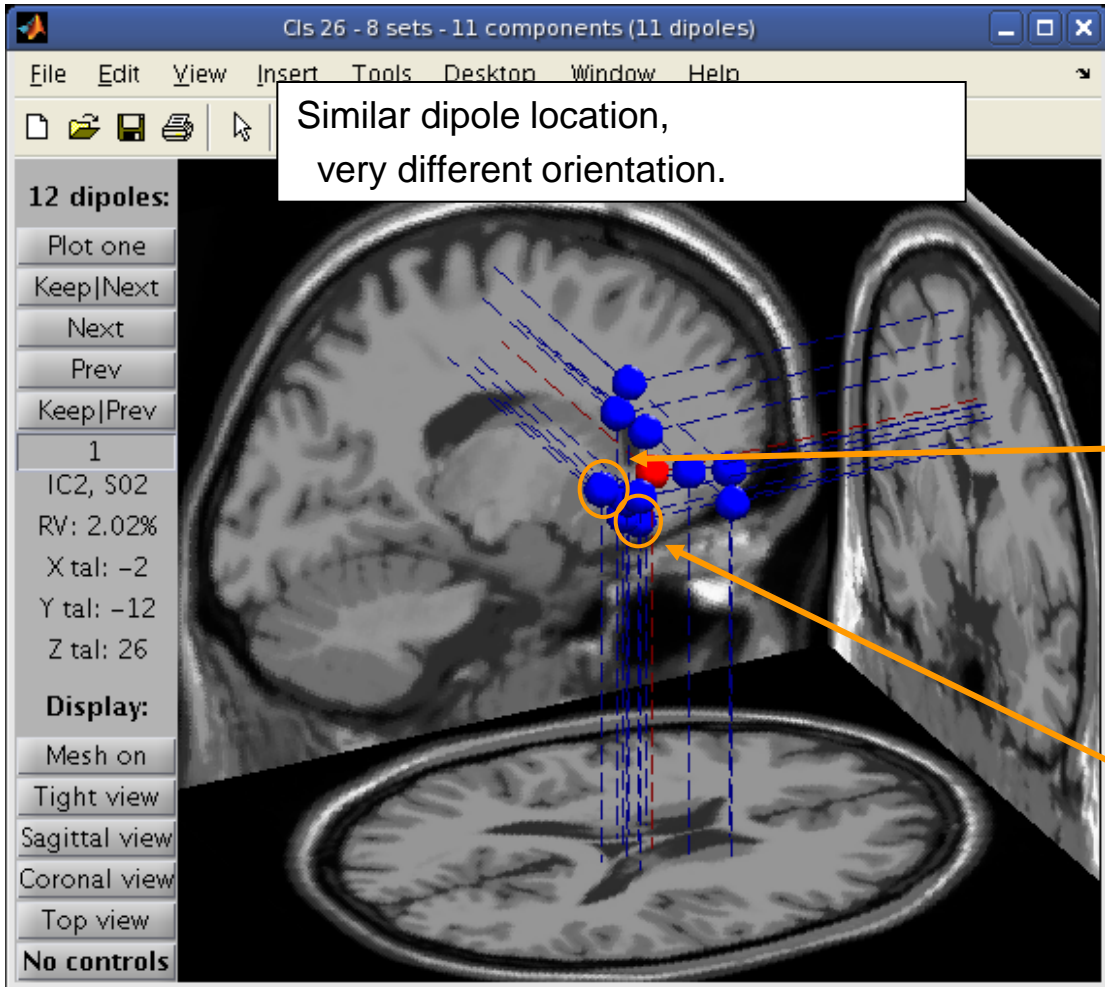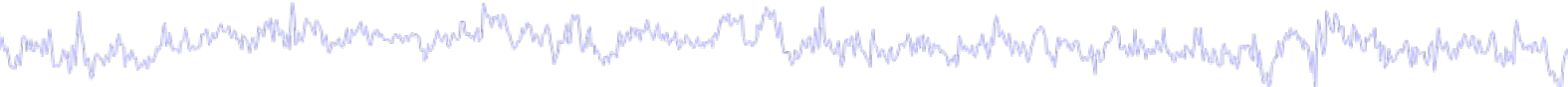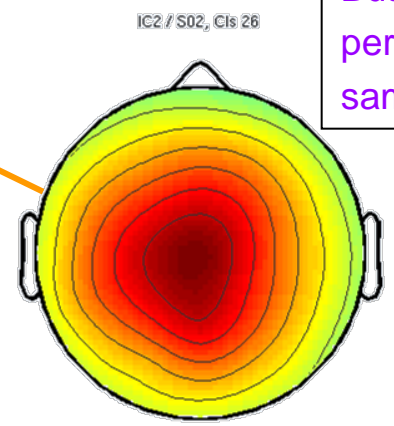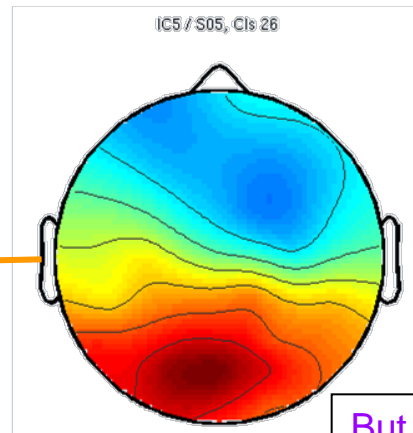   - What is the difference between these two components?

IC2 / S02, Cls 26   IC5 / S05, Cls 26

Choosing data measures

Cls 26 - 8 sets - 11 components (11 dipoles)

File Edit View Insert Tools Desktop Window Help

Similar dipole location,
very different orientation.

12 dipoles:
Plot one
Keep|Next
Next
Prev
Keep|Prev
1
IC2, S02
RV: 2.02%
X tal: –2
Y tal: –12
Z tal: 26

Display:
Mesh on
Tight view
Sagittal view
Coronal view
Top view
No controls

Obvious dramatic effect on
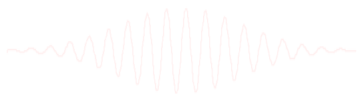scalp map topography:
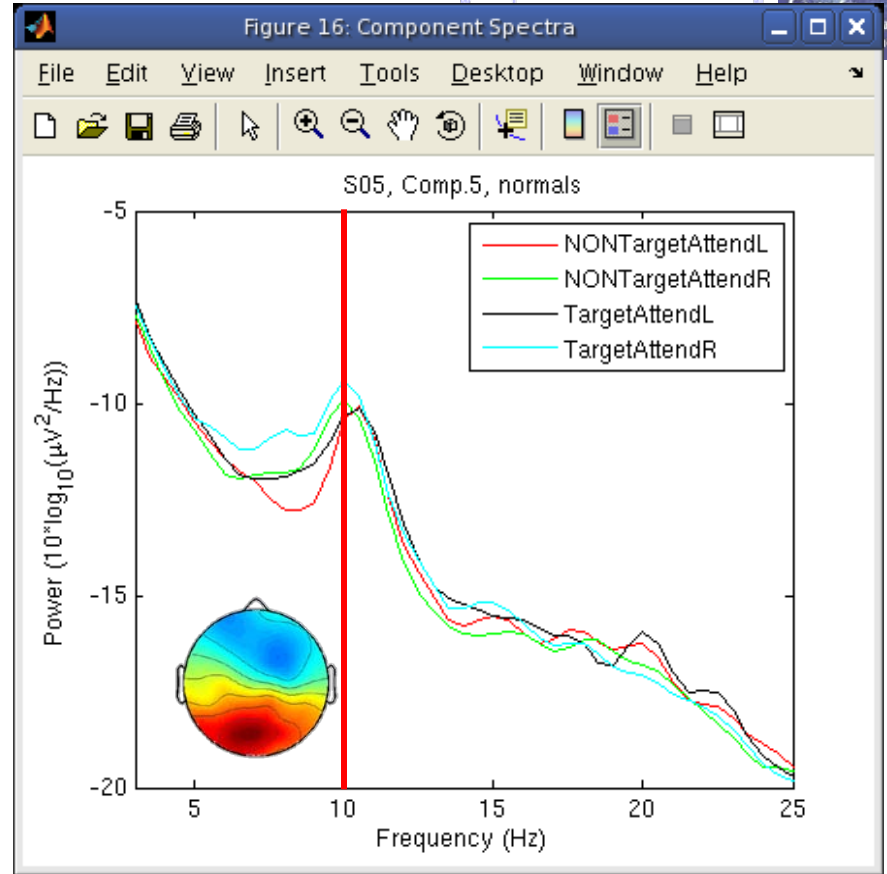
IC5 / S05, Cls 26
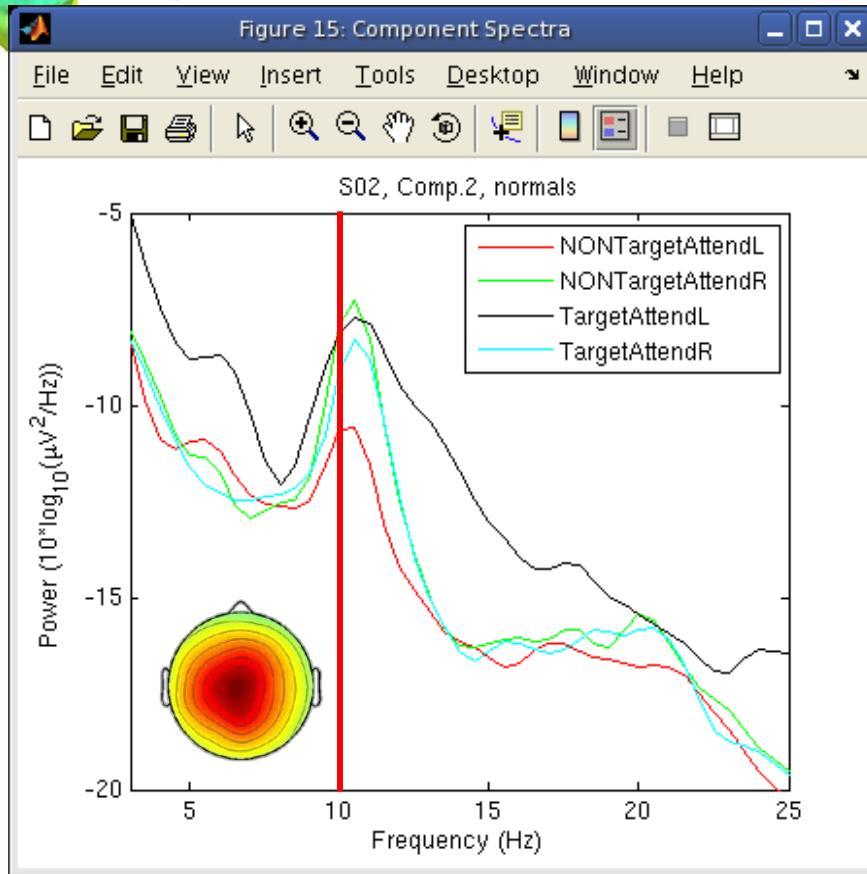
IC2 / S02, Cls 26

But, do they
perform the
same functions?

Choosing data measures
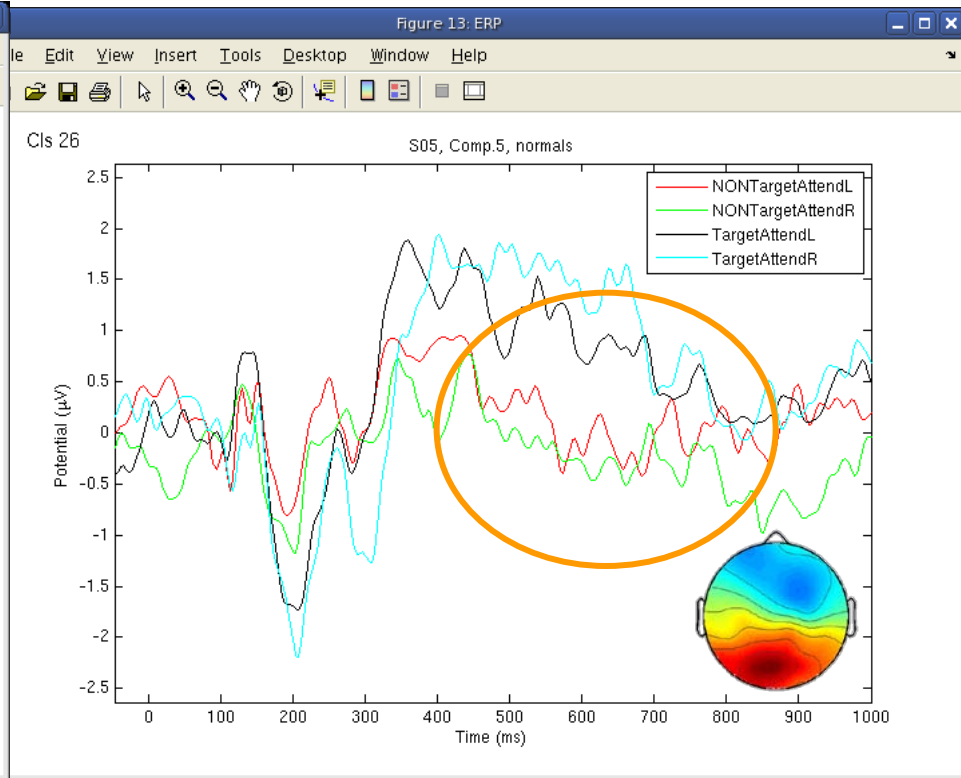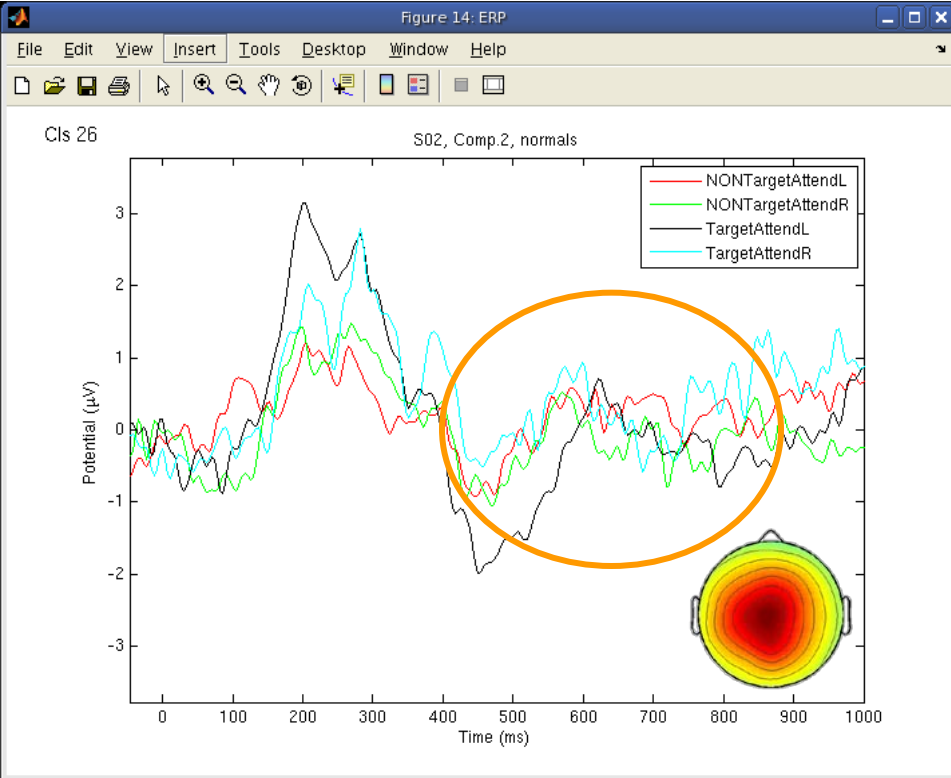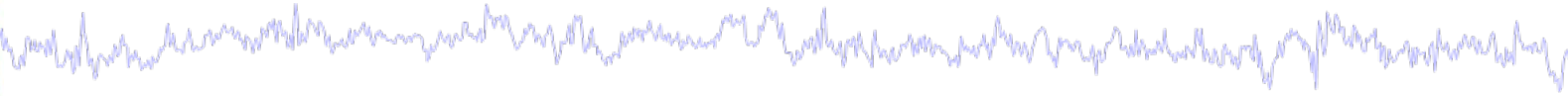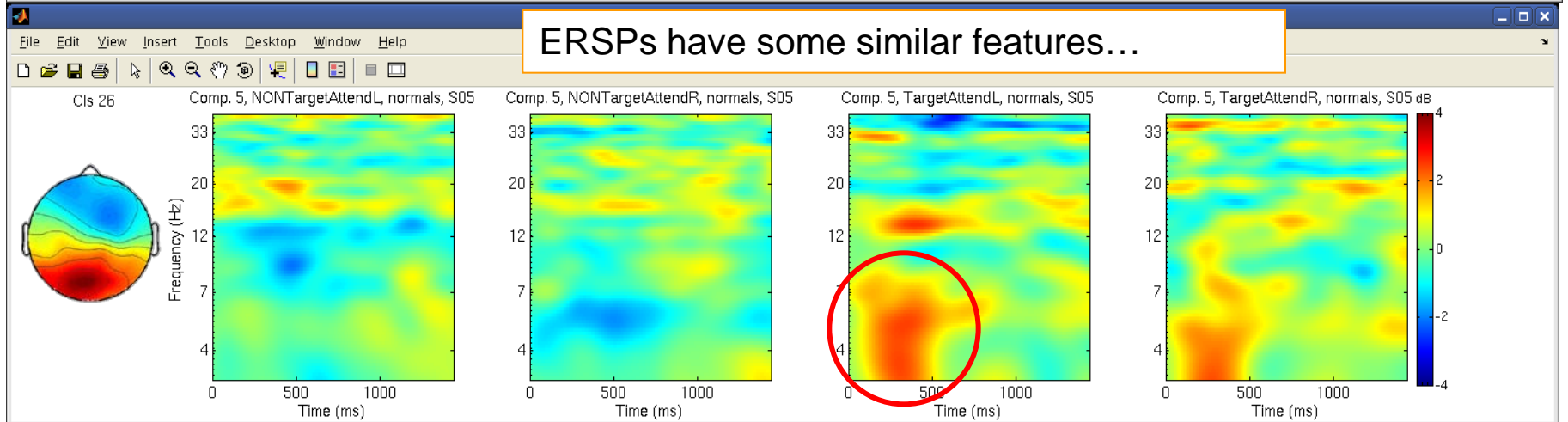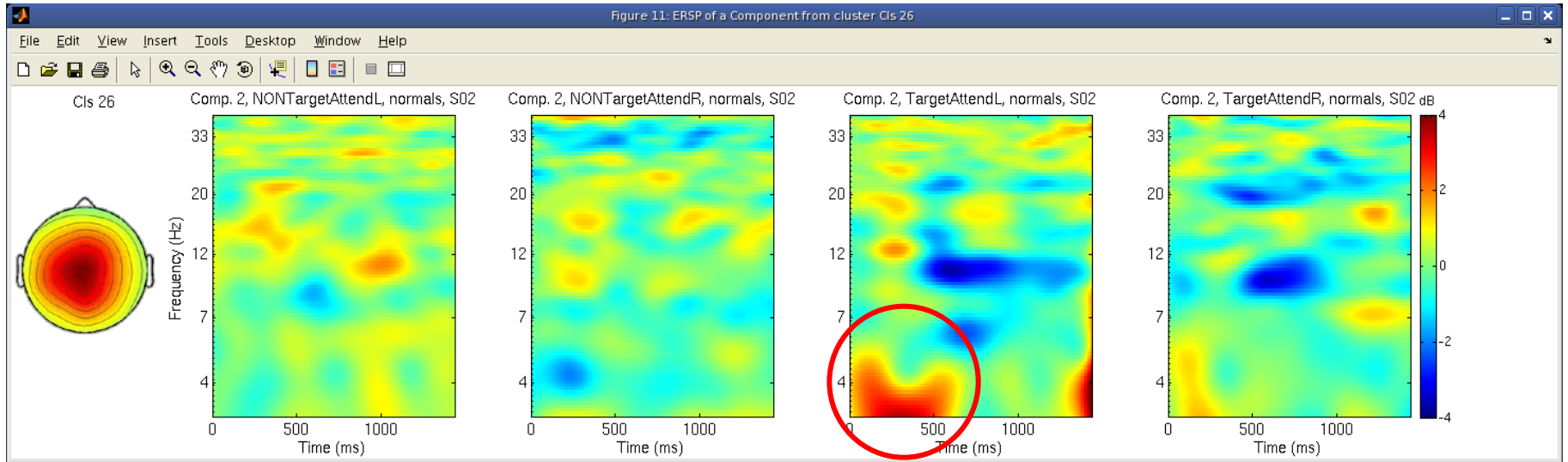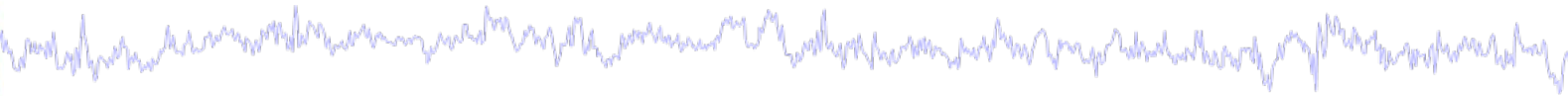

ERPs seem different…

Choosing data measures

Spectra are similar, but they have

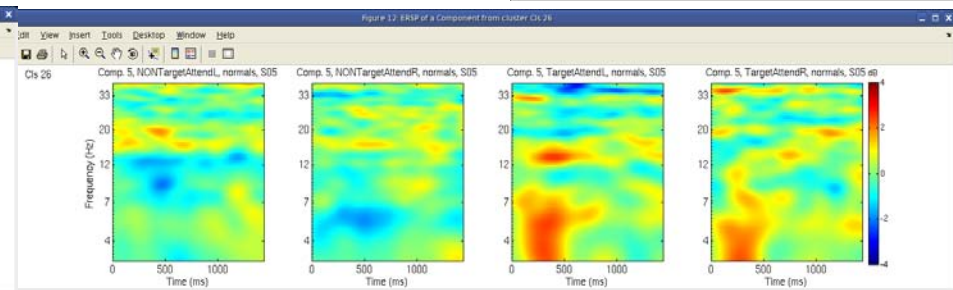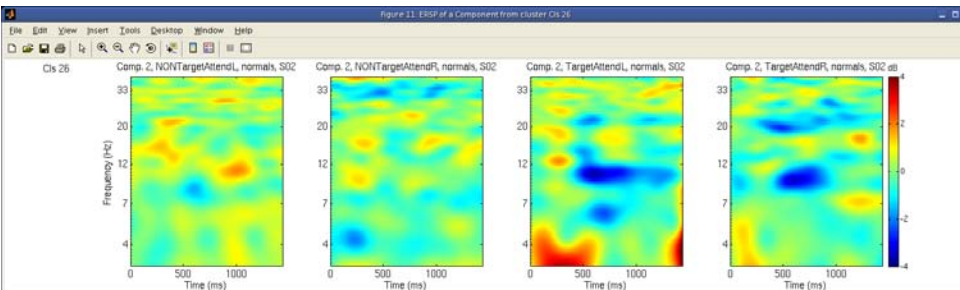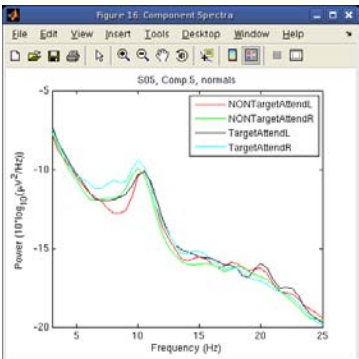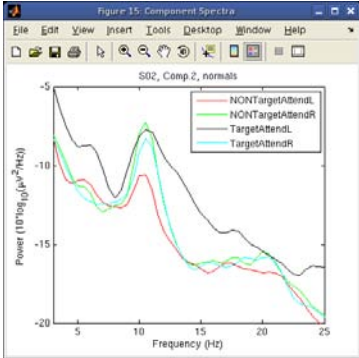variable responses to different conditions…

Choosing data measures

ERSPs have some similar features…

Choosing data measures

What data measures
should you use?

It depends…

• broadly-matched ICs: use
  many/all of the measures.
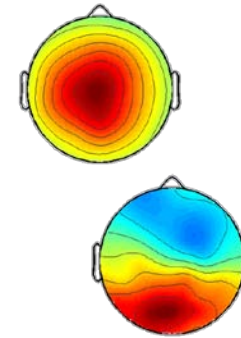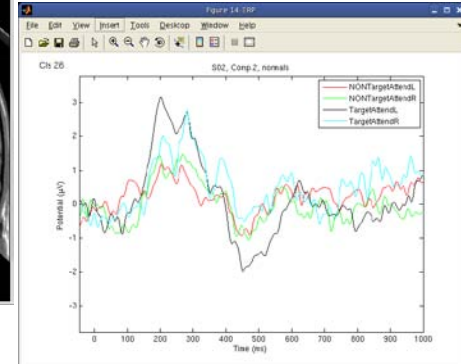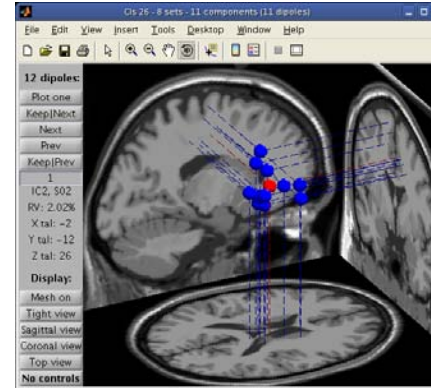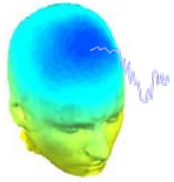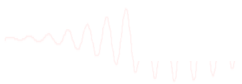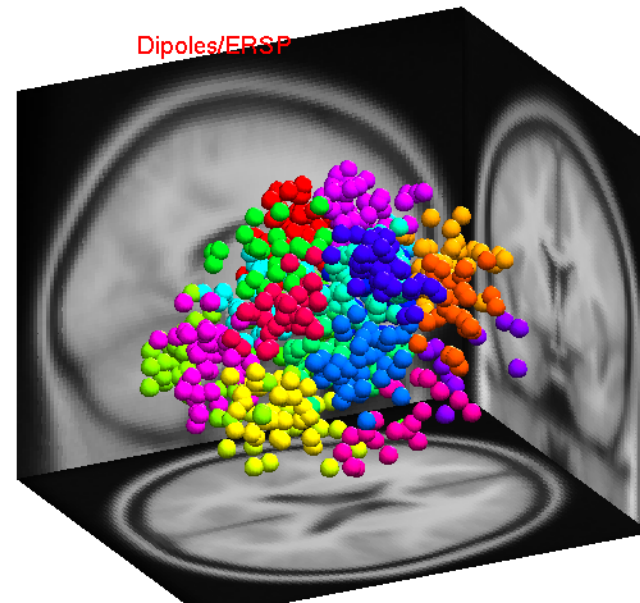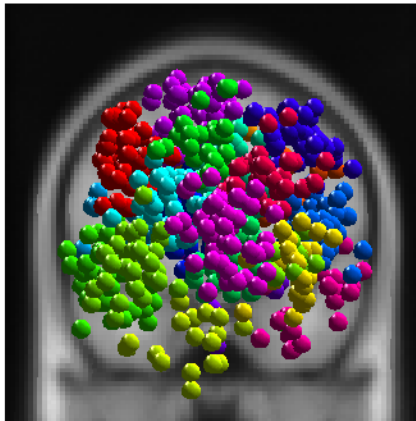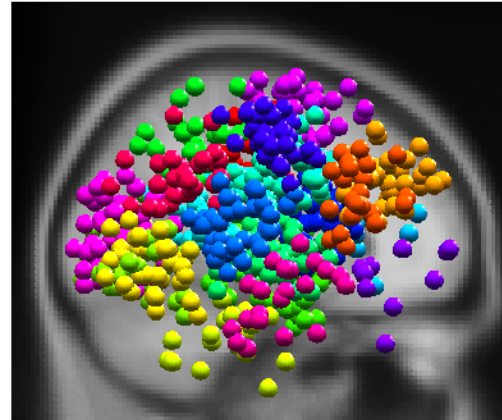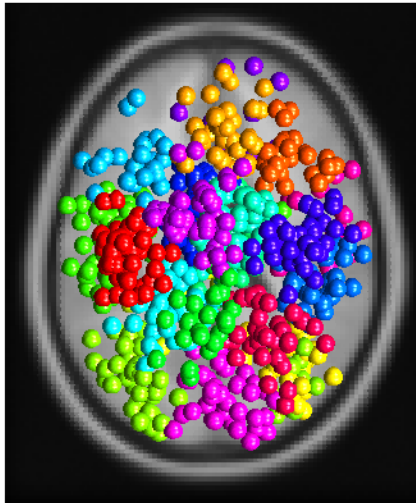
• specifically-matched ICs: use
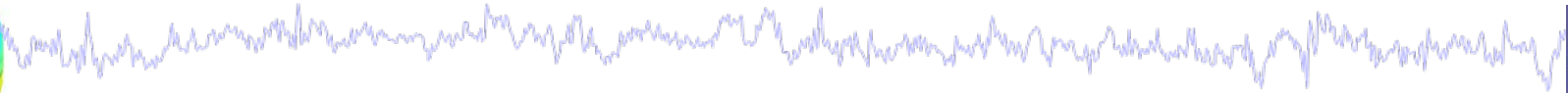 one/few of the measures.
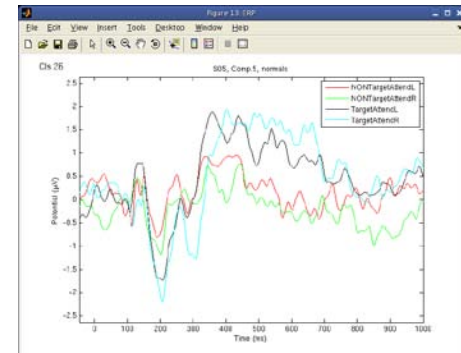
What should clusters look like?

ICs clustered by dipole location and ERSP activity

Plot STUDY dipoles

```matlab
% std_diopleclusters() variables:


clusters = [3:length(STUDY.cluster)]; % clusters to plot

title = 'Cluster Dipoles'; % figure title

plot_params = [2,2,1]; % [nrows,ncols,subplot]

views = [1,2,3,4]; % 1=top,2=side,3=rear,4=oblique

cols = hsv(length(clusters));


% std_dipoleclusters function call:


std_dipoleclusters(STUDY,ALLEEG,'clusters',clusters,…

'title',title,'viewnum',views,'rowcolplace',plot_param,…

'centroid','off','colors',cols);
```
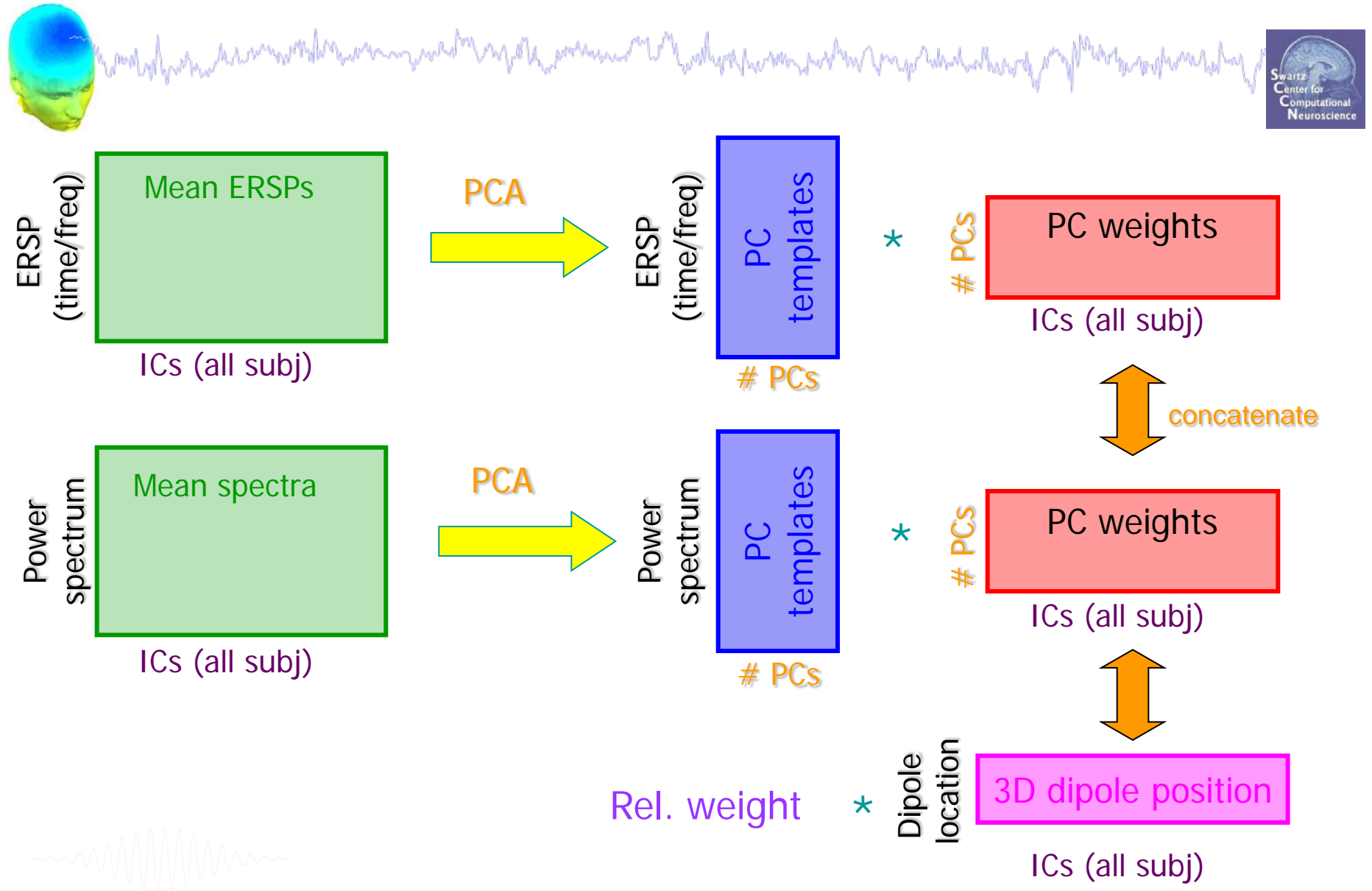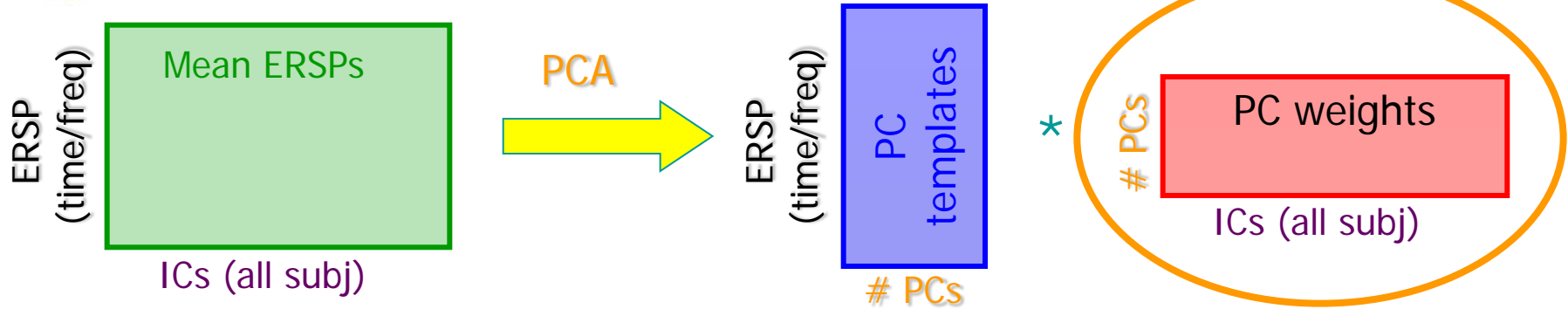
Precluster schematic
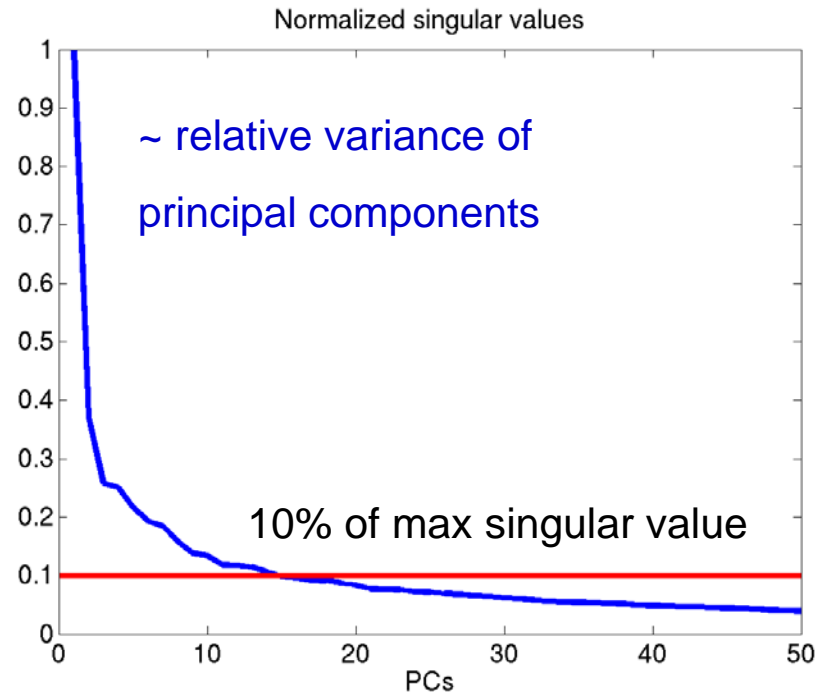

Mean ERSPs

ERSP (time/freq) × ICs (all subj)

PCA →

PC templates

ERSP (time/freq) × # PCs

*

PC weights

PCs × ICs (all subj)

Mean spectra

Power spectrum × ICs (all subj)

PCA →

PC templates

Power spectrum × # PCs

*

PC weights

PCs × ICs (all subj)

concatenate

Rel. weight *

3D dipole position

Dipole location × ICs (all subj)

Precluster: Use singular values from PCA

Mean ERSPs
ERSP (time/freq)
ICs (all subj)

PCA →

PC templates
ERSP (time/freq)
PCs

*

PC weights
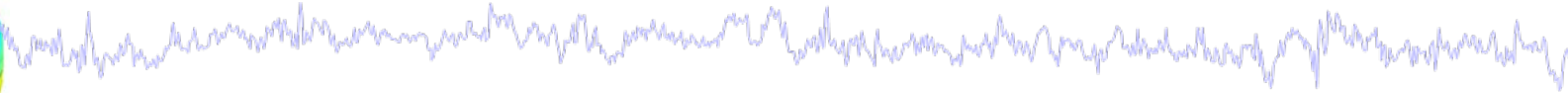# PCs
ICs (all subj)

%% Do it yourself:

%% Load all ERSP data

%% decompose with PCA

%% plot singular values

(See code in 'Tutorial_7_BuildSTUDY.m')
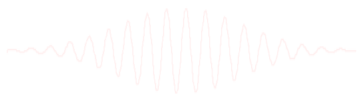
Normalized singular values

~ relative variance of

principal components

10% of max singular value

PCs

STUDY clustering overview

STEP 1

Build a STUDY

STEP 2

Precompute the data

STEP 3

Precluster the data

STEP 4

Cluster the data

Exercise...

Cluster components


```
nclusts = 25; % choose # of clusters to create

[STUDY] = pop_clust(STUDY, ALLEEG,'algorithm','kmeans','clus_num',nclusts);
```
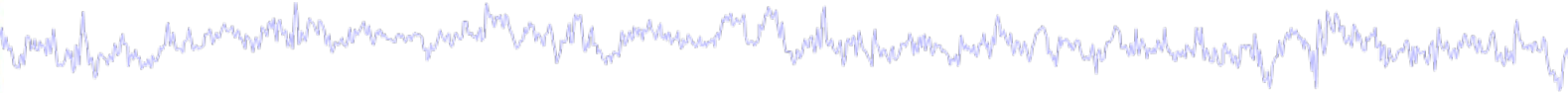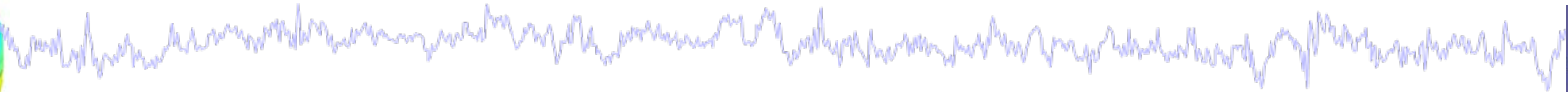
Plot/edit clusters

Plot cluster data

Plot mean scalp maps for easy reference
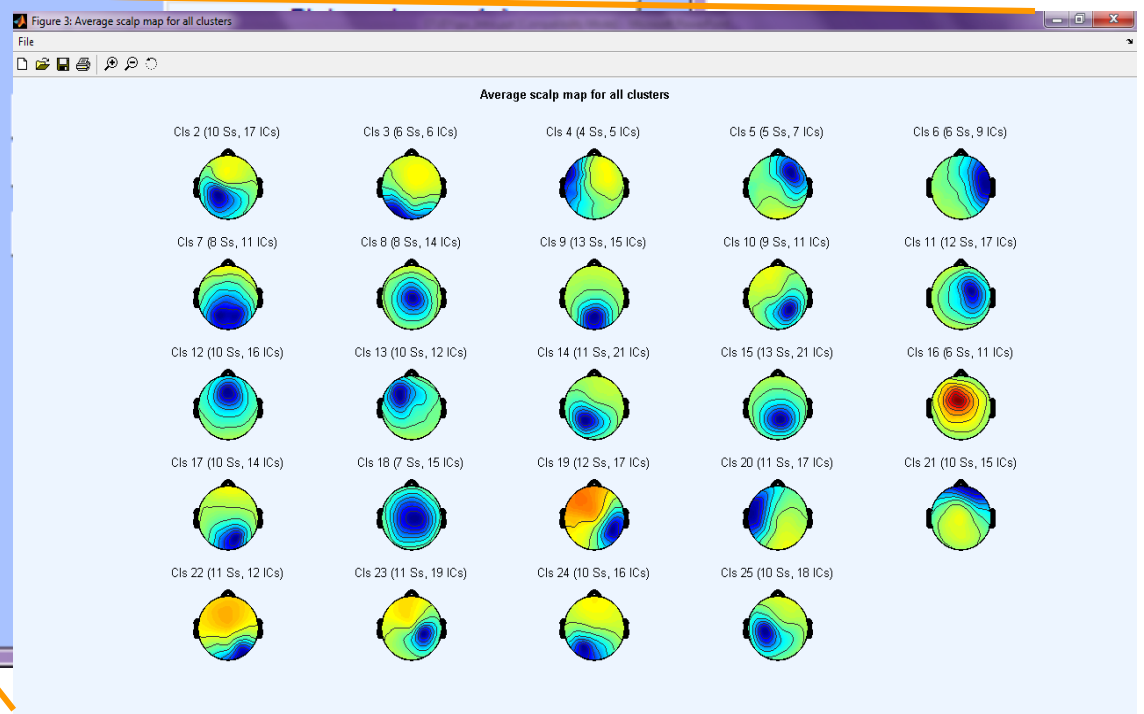
Exercise

- **Novice**
  -Open stern.study and practice plotting the existing clusters

- **Intermediate**
  - Script a loop to build a STUDY from the commandline
  - Precluster (pre-computation already done) and cluster components using measures of your choice.

- **Advanced**
  - Load raw data measures and run PCA to determine the relative size of PCA dimensions for each data measure.
  - Try preclustering/clustering based on your observations

** All scripts for exercises can be found in:
…/Scripts/Tutorial_7_BuildSTUDY.m