Task 1
Build a STUDY (via commandline)

Task 2
STUDY structure details

Task 3
Precompute/precluster data measures

Task 4
Choose data measures to cluster

Task 5
Clustering algorithm

Exercise...
Build a STUDY

% Faster alternative to building a STUDY manually
% Example STUDY: 15 subjects, 4 conditions

% Define variables:

basedir = 'C:\EEGLAB_WORKSHOP\STUDY\';
setnames = {'attend1_pos1.set','attend1_pos5.set',...
 'attend5_pos5.set','attend5_pos1.set'};
subjs = {'S01','S02','S03','S04','S05','S06','S07',...
 'S08','S09','S10','S11','S12','S13','S14','S15'};
studynname = 'Attention';
taskname = '5-box';
savename = 'attention.study';

% Note: 'EEG' fields 'condition', 'group', and 'session' need to be
% defined in each dataset of every subject before building the STUDY!
Build a STUDY

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

% Set memory options:
pop_editoptions( 'option_storedisk', 1, 'option_savematlab', 1,...
'option_computeica', 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=[];

Most important option:
- Allows only one dataset to be loaded at once.
- Most STUDYs are too big to have all data loaded at once.
Build a STUDY: GUI (review)

INSTEAD... load from the commandline
Define variables

```matlab
basedir = 'C:\EEGLAB_WORKSHOP\STUDY\';
setnames = {'attend1_pos1.set','attend1_pos5.set',
            'attend5_pos5.set','attend5_pos1.set'};
subjs = {'S01','S02','S03','S04','S05','S06','S07',
         'S08','S09','S10','S11','S12','S13','S14','S15'};

% concatenate string variables:
[] % strings inside brackets will be concatenated
[basedir,subjs{subj},'/',subjs{subj},'_',setnames{cond}];
```

C:\EEGLAB_WORKSHOP\STUDY\S01\S01_attend1_pos1.set
Load dataset info from commandline

% 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
        dataset = [basedir, subjs{subj},'/',subjs{subj},'
',setnames{cond}]; % concatenate strings
        [STUDY ALLEEG] = std_editset( STUDY, ALLEEG,...
            'name', studynname, 'task', taskname,...
            'commands',{['index',index,'load',dataset},...
                {'dipselect',0.15},{'subject',subjs{subj}}},...
            '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

EEGLAB Workshop VII, Apr. 20-22, 2009, Bloomington, IN: Julie Onton –STUDY scripting
STUDY scripting

Task 1
Build a STUDY

Task 2
STUDY structure details

Task 3
Precompute/precluster data measures

Task 4
Choosing data measures to cluster

Task 5
Clustering algorithm

Exercise...
Task 2: STUDY structure details

Question:

How do I know which ICs/subjects are in each cluster?

Where in the STUDY structure can I find this information?
Task 2: Understanding STUDY structure

>> STUDY.cluster

One cluster:

>> STUDY.cluster(6)

ans =

name: 'Cls 6'

parent: {'ParentCluster 1'}

child: []

comps: [4 2 22 1 6 15 2 7 2 6 12 15 17 3]

sets: [4x14 double]

algorithm: {'Kmeans' [19]}

centroid: []

preclust: [1x1 struct]

topo: [67x67 double]

topox: [67x1 double]

topoy: [67x1 double]

topoall: {1x14 cell}

topopol: [-1 -1 -1 1 1 1 -1 -1 1 -1 1 1 1 1 1]

20 = # of clusters

6 = cluster number

1x20 struct array with fields:

name
parent
child
comps
sets
algorithm
centroid
preclust
topo
topox
topoy
topoall
topopol

EEGLAB Workshop VII, Apr. 20-22, 2009, Bloomington, IN: Julie Onton –STUDY analysis
Task 2: Understanding STUDY structure

```matlab
>> STUDY.cluster(6)
ans =
    name: 'Cls 6'
    parent: {'ParentCluster 1'}
    child: []
    comps: [4 2 22 1 6 15 2 7 2 6 12 15 17 3]
    sets: [4x14 double]
    algorithm: {'Kmeans' [19]}
    centroid: []
    preclust: [1x1 struct]
    topo: [67x67 double]
    topox: [67x1 double]
    topoy: [67x1 double]
    topoall: {1x14 cell}
    topopol: [-1 -1 -1 1 1 1 -1 -1 1 1 1 1 1 1]
```

Which subject? Which dataset(s)?

```
>> STUDY.cluster(6).sets
ans =
    2 6 22 ...
    4 8 24 ...
    1 5 21 ...
    3 7 23 ...

Second IC = second column

Condition 4

```

Dataset indices

```matlab
>> STUDY.cluster(6).sets
ans =
    2 6 22 ...
    4 8 24 ...
    1 5 21 ...
    3 7 23 ...
```

Condition 4

Subject 2

Condition 4

EEGLAB Workshop VII, Apr. 20-22, 2009, Bloomington, IN: Julie Onton –STUDY analysis
STUDY scripting

Task 1
Build a STUDY

Task 2
STUDY structure details

Task 3
Precompute/precluster data measures

Task 4
Choosing data measures to cluster

Task 5
Clustering algorithm

Exercise...
Precompute data measures
Precompute data measures

[STUDY ALLEEG] = std_precomp(STUDY, ALLEEG,'components',...
'erp','on', 'spec','on', 'scalp','on', 'allcomps','off',...
'itc','on', 'ersp','on', 'recompute','off',...
'erspparams',{ 'cycles',[3 0.5], 'padratio', 1, 'alpha', 0.01});
Precluster data measures

parentclust = 1; % cluster 1 is always full parent cluster

[STUDY ALLEEG] = std_preclust( STUDY, ALLEEG, parentclust,...
{ 'spec', 'npca', 10, 'norm', 1, 'weight', 1, 'freqrange', [3 25] },... 
{ 'dipoles', 'norm', 1, 'weight', 10 },...
{ 'ersp', 'npca', 20, 'norm', 1, 'weight', 1, 'freqrange', [3 40],...
 'timewindow', [0 1500] },...
{ 'itc', 'npca', 10, 'norm', 1, 'weight', 1, 'freqrange', [3 30],...
 'timewindow', [0 600] });
STUDY scripting

Task 1
Build a STUDY

Task 2
STUDY structure details

Task 3
Precompute/precluster data measures

Task 4
Choosing data measures to cluster

Task 5
Clustering algorithm

Exercise...
Choosing data measures

What measure(s) should you use?

• It depends on your final cluster criteria...
  - If for example, your priority is dipoles, then cluster only based on dipole location...

But consider:
  - What is the difference between these two components?
Choosing data measures

Similar dipole location, very different orientation.

Obvious dramatic effect on scalp map topography:

But, does this indicate a difference in function?
Choosing data measures

ERPs seem different...
Choosing data measures

Spectra are similar, though different condition effects...
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.
STUDY scripting

Task 1
Build a STUDY

Task 2
STUDY structure details

Task 3
Precompute/precluster data measures

Task 4
Choosing data measures to cluster

Task 5
Clustering algorithm

Exercise...
Precluster schematic

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

PC weights

Mean spectra
Power spectrum
ICs (all subj)

PC weights

ERS
(timetime/freq)

PC weights

Rel. weight

Dipole location
ICs (all subj)

ICs (all subj)

EEGLAB Workshop VII, Apr. 20-22, 2009, Bloomington, IN: Julie Onton –STUDY scripting
Precluster: Use singular values from PCA

%% Do it yourself:
%% Load all ERSP data
%% decompose with PCA
%% plot singular values
(See code in ‘practicum_10.m’)

~ relative variance of principal components

10% of max singular value
The **k-means algorithm** is an algorithm to cluster objects into $k$ partitions.

- It attempts to find the centers of **natural clusters** in the data by minimizing the total **intra-cluster variance**, or the squared error function:

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} |x_j - u_i|^2$$

where there are $k$ clusters $S_i$, $i = 1,2,...,k$ and $u_i$ is the **centroid** or mean of all the points.

- A drawback of the algorithm is that it has to be told the number of clusters (i.e. $k$) to find.

- If the data is not naturally clustered, you get some strange results.
Clustering: K-means

Note: If too many clusters were requested, you may get this error.

\[ \text{nclusts} = 25; \quad \% \text{ choose } \# \text{ of clusters to create} \]

\[ [\text{STUDY}] = \text{pop_clust} \left( \text{STUDY, ALLEEG,'algorithm','kmeans','clus_num',nclusts}; \right) \]
Successful clustering

Study 'Attention': 181 of 181 components clustered

- Select cluster to plot
  - All cluster centroids
  - ParentCluster 1 (181 ICs)
  - outlier 2 (1 ICs)
  - ... (more options)
- Plot scalp maps
- Plot dipoles
- Plot ERPs
- Plot spectra
- Plot ERSPs
- Plot ITCs
- Plot cluster properties
- Create new cluster
- Rename selected cluster
- Merge clusters
- Save STUDY set to disk
- Reassign selected component(s)
- Remove selected outlier comps.
- Auto-reject outlier components

- Select component(s) to plot
  - outlier 2' comp. 1 (S12 IC12)
  - 'Cls 3' comp. 1 (S01 IC11)
  - 'Cls 3' comp. 2 (S05 IC11)
  - ... (more options)
- Plot scalp map(s)
- Plot dipoles
- Plot ERPs
- Plot spectra
- Plot ERSPs
- Plot ITCs
- Plot component properties
Cluster properties

Mean scalp map

Cluster dipoles with centroid

Mean ERSP (un-masked)

Mean inter-trial coherence (un-masked)
Cluster properties

Some clusters may be artifact components!

Outside the brain volume
Exercise

• **Novice**
  - Use the GUI to build a STUDY (for practice, try just a few subjects).
  - Choose 'precompute' options (but do not recalculate for the sake of time).
  - Choose 'precluster' options and cluster.

• **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 Intermediate/Advanced exercises can be found in …/EEGLAB_WORKSHOP/Scripts/practicum_10.m**