Clustering of ICA components

Arnaud Delorme

(with Julie Onton, Romain Grandchamp, Nima Bigdely Shamlo, Scott Makeig)
Steps of clustering

- Select ICA components for clustering
- Precompute measures of interest
- Cluster measures
- Plot clusters and edit them if necessary
Edit dataset info
ICs to cluster

![Image of software interface](image_url)

**Edit STUDY set information - remember to save changes:**

- **STUDY set name**: [Input field]
- **STUDY set task**: [Input field]
- **STUDY set notes**: [Input field]

**Dataset filename**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Condition</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>ic 19</td>
<td>memorize</td>
<td></td>
</tr>
<tr>
<td>ic 20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ic 21</td>
<td>ignore</td>
<td></td>
</tr>
<tr>
<td>ic 22</td>
<td>probe</td>
<td></td>
</tr>
<tr>
<td>ic 23</td>
<td>memorize</td>
<td></td>
</tr>
<tr>
<td>ic 24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ic 25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ic 26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ic 27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ic 28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ic 29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ic 30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ic 31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ic 32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Select components**

- ic 19
- ic 20
- ic 21
- ic 22
- ic 23
- ic 24
- ic 25
- ic 26
- ic 27
- ic 28
- ic 29
- ic 30
- ic 31
- ic 32

**Select by r.v.**

- Comp. 3
- Comp. 5
- Comp. 6
- Comp. 7
- Comp. 1

**Important note:** Removed datasets will not be saved before being deleted from EEGLAB memory.

- Dataset info (condition, group) differs from study info. [set] = Overwrite dataset info.
- Delete cluster information (to allow loading new datasets, set new components for clustering, etc.)

**Options:**

- Help
- Cancel
- Ok
Precompute data measures
Precompute data measures

**TIP:** Compute all measures so you can test different combinations for clustering

Time-frequency options
Cluster components
Precluster schematic

- Mean ERSPs (PCs)
- Power spectrum (PCs)
- 3D dipole position (PCs)

PCs connected to:
- Concatenation
- PCA
- Mean ERSPs
- Spectrum
- Dipoles
Precluster: Use singular values from PCA

Mean ERSPs

ICs (all subj)

PC templates

# PCs

ERSP (time/freq)

ERSP (time/freq)

Normalized singular values

~ relative variance of principal components

10% of max singular value

Precluster: Use singular values from PCA

Mean ERSPs

ICs (all subj)

PC templates

# PCs

ERSP (time/freq)

ERSP (time/freq)

Normalized singular values

~ relative variance of principal components

10% of max singular value
Precluster schematic

Each component is a dot. Clustering will group these dots.
1. k initial "means" (in this case k=3, shown in color) are randomly selected from the data set (shown in grey).

2. k clusters are created by associating every observation with the nearest mean.

3. The centroid of each of the k clusters becomes the new means.

4. Steps 2 and 3 are repeated until convergence has been reached.

Classical KMean
Cluster components
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?
Choosing data measures

Similar dipole location, very different orientation.

Obvious dramatic effect on scalp map topography:

But, do they perform the same functions?
Choosing data measures

ERPs seem different…
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?
Plot/edit clusters

[Image of a software interface for plotting and editing clusters]
Plot cluster data

Plot mean scalp maps for easy reference
Plot cluster data

Choose which cluster

Choose which components

Study 'Attention': 181 of 181 components clustered

Select cluster to plot
- Cls 5 (3 ICs)
- Cls 7 (10 ICs)
- Cls 8 (5 ICs)
- Cls 9 (12 ICs)

Select component(s) to plot
- All components
- S01 IC6
- S05 IC9
- S06 IC12

Plot scalp maps
- Plot dipoles
- Plot ERPs
- Plot spectra
- Plot ITCs
- Plot scalp map(s)

Create new cluster
- Rename selected cluster
- Merge clusters

Save STUDY set to disk
/home/julie/WorkshopSD2

Figure 4

Cls 21 (7 ICs, 7 Ss)

ic1/S01
ic1/S08
ic11/S02
ic2/S10
ic3/S12
ic3/S15
Plot cluster data
Plot cluster ERP
Reassigning components
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.
Measure projection

(EEGLAB extension by Nima Bigdely Shamlo) only has one pre-clustering parameter.

(Affinity clustering by Pernet, Martinez, Delorme)
Exercise

• Load the STUDY
• Precompute ERP, power spectrum and scalp topographies
• Precluster and cluster components using spectrum and dipoles location
• Look at your cluster. Identify frontal midline theta cluster and occipital alpha cluster
• Plot significant difference for one component cluster spectrum between the two conditions in the default design