Clustering of ICA components

Arnaud Delorme

(with Julie Onton, Romain Grandchamp, Nima Bigdely Shamlo, Scott Makeig)
Outline

- ICA clusters and reliability within subjects
- ICA clusters and reliability across subjects
- Clustering in EEGLAB theory & Practice
ICA decomposition of multiple data sets from the same individuals

- Experimental protocol
  - Mind wandering experiment
  - 2 subjects
  - 11 x 30 min. sessions
  - 2 sessions per week
  - EEG from Biosemi 64 channels
  - Fs=1024 Hz
Results (Cluster 1)

100 % Sessions contribute
100 % Sessions contribute
Results (Cluster 8)

100 % Sessions contribute
Results (Cluster 13)

63.64% Sessions contribute
Results (Cluster 14)

36.36% Sessions contribute
## Inter iteration Cluster Consistency

<table>
<thead>
<tr>
<th>Clusters</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td><strong>100</strong></td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>90</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td><strong>99</strong></td>
</tr>
<tr>
<td>5</td>
<td>90</td>
<td>40</td>
<td>10</td>
<td>90</td>
<td>90</td>
<td>60</td>
<td>100</td>
<td>10</td>
<td>60</td>
<td>90</td>
<td><strong>64</strong></td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>0</td>
<td>100</td>
<td>60</td>
<td>100</td>
<td>90</td>
<td>60</td>
<td>90</td>
<td>60</td>
<td>60</td>
<td><strong>68</strong></td>
</tr>
<tr>
<td>7</td>
<td>90</td>
<td>100</td>
<td>90</td>
<td>90</td>
<td>60</td>
<td>90</td>
<td>90</td>
<td>100</td>
<td>90</td>
<td>90</td>
<td><strong>89</strong></td>
</tr>
<tr>
<td>8</td>
<td>80</td>
<td>80</td>
<td>60</td>
<td>80</td>
<td>40</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>100</td>
<td><strong>76</strong></td>
</tr>
<tr>
<td>9</td>
<td>60</td>
<td>90</td>
<td>50</td>
<td>60</td>
<td>80</td>
<td>60</td>
<td>0</td>
<td>10</td>
<td>60</td>
<td>50</td>
<td><strong>52</strong></td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>90</td>
<td>10</td>
<td>40</td>
<td>0</td>
<td>50</td>
<td>50</td>
<td>0</td>
<td>50</td>
<td>60</td>
<td><strong>39</strong></td>
</tr>
<tr>
<td>11</td>
<td>60</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>60</td>
<td>10</td>
<td>90</td>
<td>60</td>
<td>60</td>
<td><strong>37</strong></td>
</tr>
<tr>
<td>12</td>
<td>100</td>
<td>50</td>
<td>50</td>
<td>100</td>
<td>50</td>
<td>100</td>
<td>50</td>
<td>100</td>
<td>50</td>
<td>100</td>
<td><strong>75</strong></td>
</tr>
<tr>
<td>13</td>
<td>50</td>
<td>10</td>
<td>20</td>
<td>50</td>
<td>90</td>
<td>50</td>
<td>50</td>
<td>10</td>
<td>50</td>
<td>20</td>
<td><strong>40</strong></td>
</tr>
<tr>
<td>14</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>20</td>
<td>30</td>
<td>20</td>
<td>20</td>
<td>30</td>
<td>30</td>
<td><strong>21</strong></td>
</tr>
</tbody>
</table>
Outline

- ICA clusters and reliability within subjects
- ICA clusters and reliability across subjects
- Clustering in EEGLAB theory & Practice
Localization

ICA component scalp maps

Electrodes

Components

Time (ms)
Computing residual variance (\%) 

\[ r = \frac{\sum (x_i - \bar{x}_i)^2}{\sum x_i^2} \]
Validation of the ICA algorithm for EEG

Data
• 13 subjects performing a memory task
• 71 electrodes including EOGs
• more than 300,000 data points/subject

Decomposition
• 23 ICA algorithms plus PCA and Promax

Analysis
• Localization of all components with a single dipole
  (4-shell spherical model)
<table>
<thead>
<tr>
<th>Algorithm (Matlab func.)</th>
<th>D%</th>
<th>LL</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extended Infomax (runica)</td>
<td>29.9</td>
<td>178</td>
<td>EEGLAB 4.515</td>
</tr>
<tr>
<td>Pearson</td>
<td>29.1</td>
<td>169</td>
<td>ICAcentral (6)</td>
</tr>
<tr>
<td>Infomax (runica)</td>
<td>28.2</td>
<td>160</td>
<td>EEGLAB 4.515</td>
</tr>
<tr>
<td>ERICA</td>
<td>26.9</td>
<td>184</td>
<td>ICAcentral (5)</td>
</tr>
<tr>
<td>SONS</td>
<td>25.4</td>
<td>183</td>
<td>ICAcentral (5)</td>
</tr>
<tr>
<td>SHIBBS</td>
<td>23.7</td>
<td>169</td>
<td>ICAcentral (2)</td>
</tr>
<tr>
<td>FastICA*</td>
<td>23.5</td>
<td>169</td>
<td>ICAcentral (2)</td>
</tr>
<tr>
<td>JADE (jader)</td>
<td>23.4</td>
<td>169</td>
<td>EEGLAB 4.515</td>
</tr>
<tr>
<td>TICA</td>
<td>23.4</td>
<td>169</td>
<td>ICAcentral (5)</td>
</tr>
<tr>
<td>JADE optimized (jade_op)</td>
<td>21.4</td>
<td>169</td>
<td>ICAcentral (5)</td>
</tr>
<tr>
<td>JADE w/ time delay (jade_td)</td>
<td>20.2</td>
<td>169</td>
<td>ICAcentral (5)</td>
</tr>
<tr>
<td>eeA</td>
<td>19.0</td>
<td>305</td>
<td>ICAcentral (8)</td>
</tr>
<tr>
<td>Infomax (icaML) †</td>
<td>18.8</td>
<td>212</td>
<td>ICA DTU Tbox</td>
</tr>
<tr>
<td>FOBI</td>
<td>18.6</td>
<td>169</td>
<td>ICAcentral (5)</td>
</tr>
<tr>
<td>SOBIRO (acsohiro)</td>
<td>17.9</td>
<td>167</td>
<td>EEGLAB 4.515</td>
</tr>
<tr>
<td>EVD 24</td>
<td>17.7</td>
<td>169</td>
<td>ICAcentral (5)</td>
</tr>
<tr>
<td>EVD</td>
<td>17.0</td>
<td>169</td>
<td>ICAcentral (5)</td>
</tr>
<tr>
<td>SOBI</td>
<td>16.1</td>
<td>583</td>
<td>EEGLAB 4.515</td>
</tr>
<tr>
<td>icaMS†</td>
<td>10.6</td>
<td>169</td>
<td>ICA DTU Tbox</td>
</tr>
<tr>
<td>AMUSE</td>
<td>8.5</td>
<td>169</td>
<td>ICAcentral (5)</td>
</tr>
<tr>
<td>PCA</td>
<td>3.1</td>
<td>583</td>
<td>EEGLAB 4.515</td>
</tr>
<tr>
<td>Promax</td>
<td>33.7</td>
<td>467</td>
<td>EEGLAB 4.515</td>
</tr>
<tr>
<td>Whitening/Sphering</td>
<td>57.6</td>
<td>164</td>
<td>EEGLAB 4.515</td>
</tr>
</tbody>
</table>
Component examples
Cumulative number of component below residual variance threshold
More independence -> more biological components
C

Percent near-dipolar (r.v. < 5%) components

- Ext. Infomax
- Pearson Infomax
- FastICA
- JADE
- SHiBBS
- TICA
- JADE opt.

$R^2 = 0.71$
$p = 10^{-5}$

Remaining Pairwise Mutual Information (10^{-3} %)

$R^2$ value

Probability ($10^{-3}$)

Near dipolar cutoff (% r.v.)
Correlations between decompositions
Right $\mu$ cluster

S1 IC51

S3 IC41

S5 IC51

S7 IC48

S11 IC49

S2 IC41

S4 IC50

S6 IC60

S9 IC39

S14 IC49

Frequency (Hz)

Power (dB)
Occipital $\alpha$ cluster
Frontal Midline $\theta$ cluster
D

Mean cluster tightness (mm)

R^2 = 0.74
p < 10^{-5}

Mutual Information Reduction (kbits/sec)
Outline

- ICA clusters and reliability within subjects
- ICA clusters and reliability across subjects
- Clustering in EEGLAB theory & Practice
**Edit dataset info**

### Create a new STUDY set -- pop_study()

#### Edit STUDY set information - remember to save changes

<table>
<thead>
<tr>
<th>STUDY set name:</th>
<th>Sternberg</th>
</tr>
</thead>
<tbody>
<tr>
<td>STUDY set task name:</td>
<td>Sternberg</td>
</tr>
<tr>
<td>STUDY set notes:</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dataset filename</th>
<th>browse</th>
<th>subject</th>
<th>session</th>
<th>condition</th>
<th>group</th>
</tr>
</thead>
<tbody>
<tr>
<td>C:Users</td>
<td>Julie</td>
<td>Documents</td>
<td>Work...</td>
<td>SC1</td>
<td>memory</td>
</tr>
<tr>
<td>C:Users</td>
<td>Julie</td>
<td>Documents</td>
<td>Work...</td>
<td>SC1</td>
<td>ignore</td>
</tr>
<tr>
<td>C:Users</td>
<td>Julie</td>
<td>Documents</td>
<td>Work...</td>
<td>SC2</td>
<td>probe</td>
</tr>
<tr>
<td>C:Users</td>
<td>Julie</td>
<td>Documents</td>
<td>Work...</td>
<td>SC2</td>
<td>memory</td>
</tr>
<tr>
<td>C:Users</td>
<td>Julie</td>
<td>Documents</td>
<td>Work...</td>
<td>SC3</td>
<td>ignore</td>
</tr>
<tr>
<td>C:Users</td>
<td>Julie</td>
<td>Documents</td>
<td>Work...</td>
<td>SC3</td>
<td>probe</td>
</tr>
<tr>
<td>C:Users</td>
<td>Julie</td>
<td>Documents</td>
<td>Work...</td>
<td>SC3</td>
<td>memory</td>
</tr>
<tr>
<td>C:Users</td>
<td>Julie</td>
<td>Documents</td>
<td>Work...</td>
<td>SC4</td>
<td>memory</td>
</tr>
</tbody>
</table>

**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]
ICs to cluster
Precompute data measures
Precompute data measures

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

Time-frequency options
3. Cluster components
Precluster schematic

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

Power spectrum
Mean spectra
ICs (all subj)

3D dipole position

PCA

PC templates
ICs (all subj)

Mean spectra
PC templates
ICs (all subj)

3D dipole position
PC templates
ICs (all subj)

PC weights

concatenate

PC weights

concatenate

PC weights

PCA

ICs (all subj)
Precluster: Use singular values from PCA

Mean ERSPs

ICs (all subj)

PCA

ERSP (time/freq)

IC templates

# PCs

Normalized singular values

~ relative variance of principal components

10% of max singular value

Credit: Julie Onton
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.
Customized KMean
(no more than 1 session per cluster)

1. A first KMean solution is computed for N clusters
2. Select the cluster with minimum residual distance to centroid
3. Keeps at most one component per session (min dist. to centroid)
4. Store the resulting cluster
5. Remove the cluster’s ICs from the pool of all ICs
6. Compute a new KMean solution for N-1 clusters on the new pool
7. Loop until the desired number of selected clusters is reached
Cluster components

- **Study set:** At
- **Study filename:**
- **Study task name:**
- **Nb of subjects:**
- **Nb of conditions:**
- **Nb of sessions:**
- **Nb of groups:**
- **Epoch consistency:** yes
- **Channels per frame:** 31
- **Channel locations:** yes
- **Clusters:** 1
- **Status:** Pre-clustered
- **Total size (Mb):** 32.4

**Cluster components**
- Edit study info
- Precompute channel measures
- Plot channel measures
- Precompute component measures
- Build preclustering array
- Cluster components
- Edit/plot clusters

**Set clustering algorithm:** pop_clust0
- **Clustering algorithm:** Kmeans (stat. toolbox)
- **Number of clusters to compute:**
  - Separate outliers (enter std.): 24
  - 3

**Performing clustering on cluster "ParentCluster 0"**

**Help** | **Cancel** | **Ok**
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…
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.
Plot/edit clusters
Plot cluster data

Study 'Attention': 181 of 181 components clustered

Select cluster to plot
- All cluster centroids
- ParentCluster 1 (181 ICs)
- outlier 2 (1 ICs)
- Cls 3 (5 ICs)

Select component(s) to plot
- outlier 2' comp. 1 (S1 IC12)
- Cls 3' comp. 1 (S01 IC11)
- Cls 3' comp. 2 (S05 IC11)
- Cls 3' comp. 3 (S06 IC15)

Plot scalp maps
- Plot dipoles
- Plot ERPs
- Plot spectra
- Plot ERPs
- Plot cluster properties

Create new cluster
- Rename selected cluster
- Merge clusters

Save STUDY set to disk: /home/julie

Plot mean scalp maps for easy reference
Plot cluster data

Choose which cluster

Choose which components
Plot cluster data

Select component(s) to plot

Plot scalp map(s)

Plot dipole(s)

Plot ERPs

Plot spectra

Plot ITCs

Plot cluster properties

Create new cluster

Rename selected cluster

Merge clusters

Save STUDY set to disk

/home/julie/workshop06/5subjects/W1study/study

...
Plot cluster data

Study #: 151 of 151 components clustered

Select cluster to plot:
- Cls 15 (4 ICs)
- Cls 16 (6 ICs)
- Cls 17 (4 ICs)
- Cls 18 (4 ICs)
- Cls 19 (4 ICs)

Select component(s) to plot:
- All components
- S02 IC13
- S02 IC12
- S02 IC17

Set ERP plotting parameters -- pop_erpparams()

- Time range in ms [low high]
- Plot scalp map at latency [ms] NaN
- Plot conditions on the same panel
- Plot groups on the same panel
- Statistical method to use: Parametric
- Statistical threshold (p<)
- Compute condition statistics
- Compute group statistics
- Use single trials (when available)
- Use False Discovery Rate to correct for multiple comparisons

[Options: Help, Cancel, Ok]
Plot cluster ERP
Other plotting options…

Figure 6 (2)

- High dose
- Low dose
- Placebo

- KAN
- NONKAN
Reassigning components
Reassigning components
Outlier cluster reassignment
Measure Projection Toolbox
Measure Projection Toolbox
Exercise

Precluster (pre-computation already done) and cluster components using measures of your choice. Experiment with different measures.