

Clustering of ICA Components for artifact detection

Stefan Scherbaum, University of Osnabrück
Stefan.Scherbaum@web.de

Independent Component Analysis is a statistical method that does blind source separation. Therefore, there is practically no implied meaning within the results and no easy interpretation. Neither the order of the components nor the seemingly arbitrary weights in the transformation matrix tell anything about the nature of each component per se.

Therefore, classification of components is a task that cannot be automated easily (Jung, Makeig et al., 2000) and that demands personal experience. The nature of this task has several drawbacks:

- For beginners, learning this task is hard and erroneous until they achieve a reasonable level of experience.
- Basing the classification solely on personal experience makes decisions somewhat arbitrary.
- Manual classification of components is time consuming and there might be a reasonable loss of attention during the task.
- It is hard to keep constantly track of all features of each component that have to be cared about for classification.

To address these issues, I decided to write a clustering tool that introduced some statistical measures to classify the components automatically. This way, it enables the user to get an overview or gist of each group of components. I was inspired to do this by the semi-automatic trial rejection mechanism included in EEGLab (Delorme, Makeig et al., 2001). This mechanism makes trial rejection less arbitrary by proposing trials to reject, based on a classification by several measures as kurtosis, probability of data etc.

I used the following properties to classify components as artifacts (compare Delorme and Makeig, 2004):

1. The frequency spectrum: High power in high frequencies is a distinctive feature of muscle artifacts. Eye blinks also have a distinctive spectrum that is different to normal EEG spectra.

Clustering of ICA Components for artifact detection

2. The components topography (the weight matrix): Muscle artifacts often concentrate around few electrodes, while eye blinks concentrate around the frontal electrodes.
3. The distribution of the activation of the component over time: muscle artifacts often occur only for one block of trials. Therefore, the overall activation of such a component is high for one block of trials and then low for the rest of the trials.
4. The strength of the component's ERP: a distinctive ERP at stimulus onset or around events indicates brain signals.

I measured these properties in the following way:

1. To measure the spectrum, I used several ratios of frequency samples:
40/20 Hz, 50/10 Hz, 50/20 Hz, 50/30 Hz, 20/10 Hz, 10/5 Hz.
Additionally, I included the overall power of the spectrum.
2. To measure the component's topography, I counted the weights that were greater than 2 standard deviations of the whole weight matrix.
3. To search for blocked activity of a component over trials, I summed the absolute activity of all trials over time, convoluted the resulting signal with a 30x1 block, and used the maximum value of the resulting signal as an indicator of blocked activity.
4. To measure the ERPness, I calculated the kurtosis of the components ERP.

These measures provided 10 dimensions describing each component. I used the kmeans method to cluster the data with 100 iterations to derive stable clusters. Clustering the components into 7 groups was the best tradeoff between clearness of classification and reliability of classification.

The screenshot in figure 1 shows the result of such a clustering in a 2D view (showing the 2 main dimensions of a PCA on the clustered data). Components already rejected are marked with a red square. For the data shown, there is strong tendency for artifactual components to be on the right side of the plot.

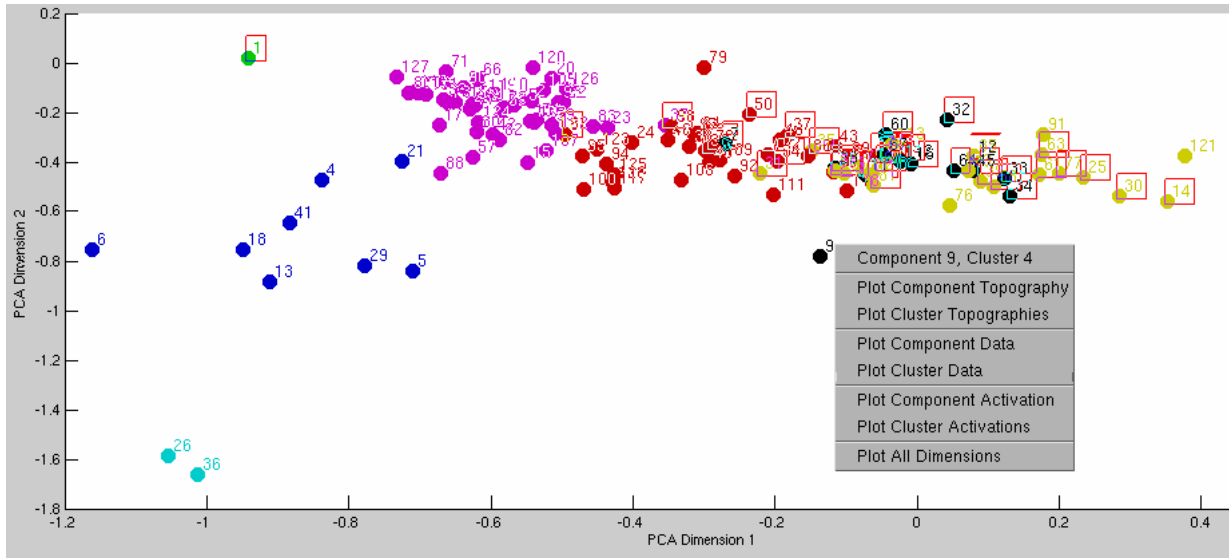


Figure 1 - The results of the clustering process, mapped on the first two PCA dimensions. Artifacts can be found mainly on the right. Components marked for rejection are marked with a red square. A context menu for each component offers further views.

The only one not following this rule is component 1, which represents a cluster on its own and contains eye blinks (see figure 2).

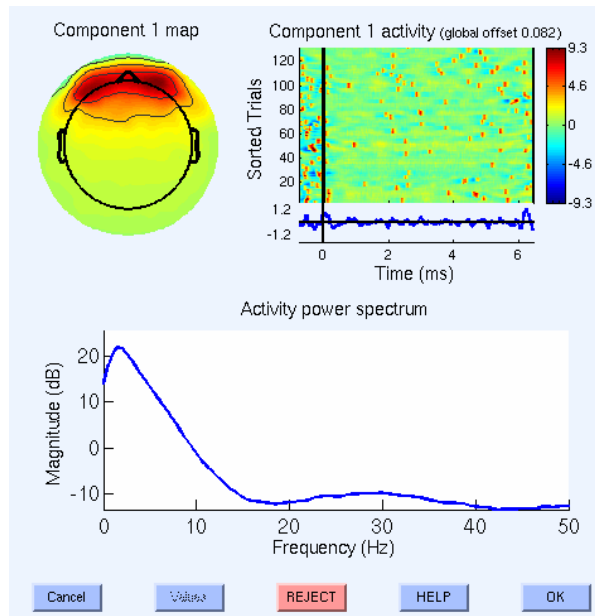


Figure 2 - Plot of component properties, using built in functions of EEGLab. Component 1 shows a typical topography and spectrum for eye blink artifacts.

I integrated the tool into EEGLab as plugin. It offers the user to plot component topographies (see Figure 3 and 4), to scroll component activations (see Figure 5), or to view the component's statistical data (Figure 6).

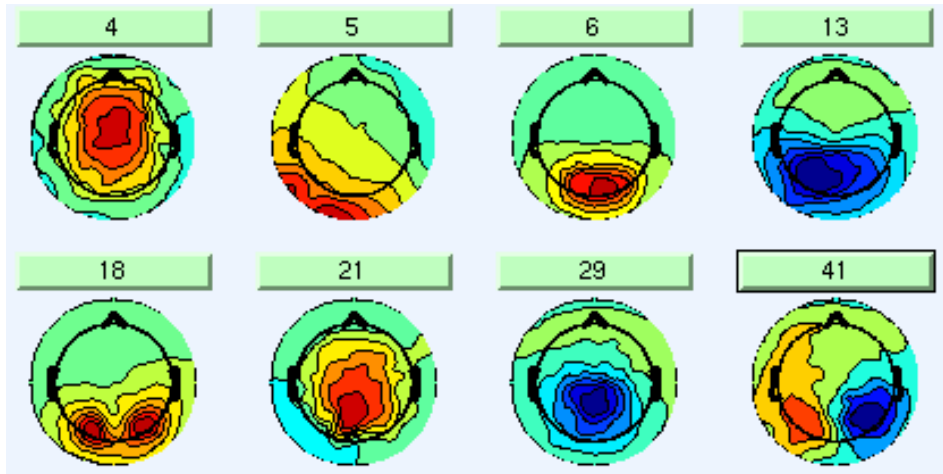


Figure 3 - Topographies of all components in one cluster.
Topographies can be plotted using the built in functions of EEGLab.

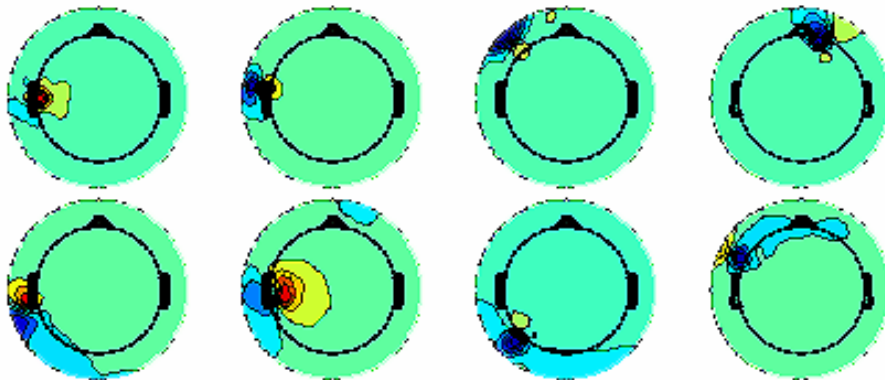


Figure 4 - A cluster of possibly artifactual components found by the algorithm.
The topographies, centered on single electrodes, are typical for muscle tension artifacts.

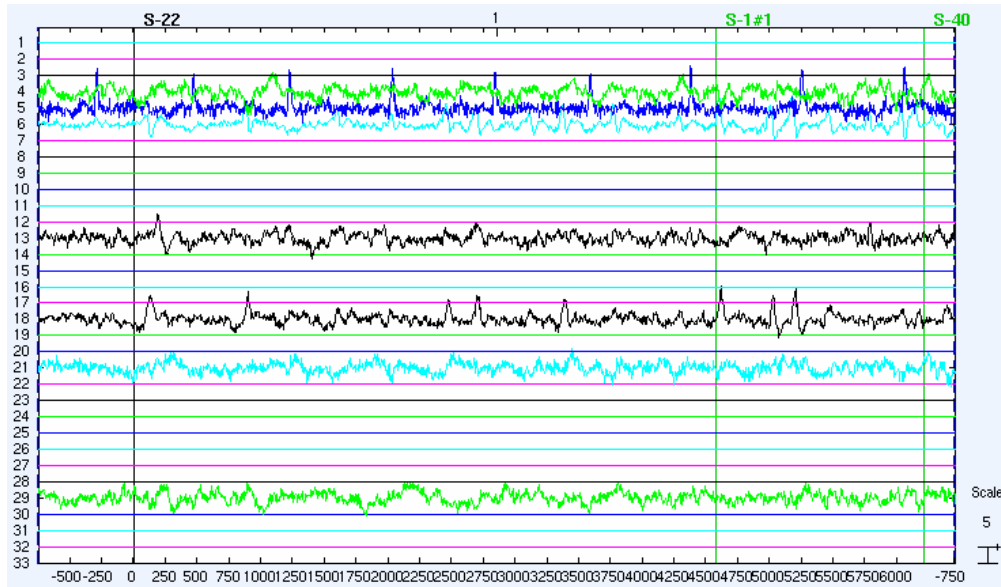


Figure 5 – Waveforms of a cluster of components.

To show specifically the activations of the components in one cluster, the tool removes all components, not included in this cluster, and then uses the built in EEGLab function.

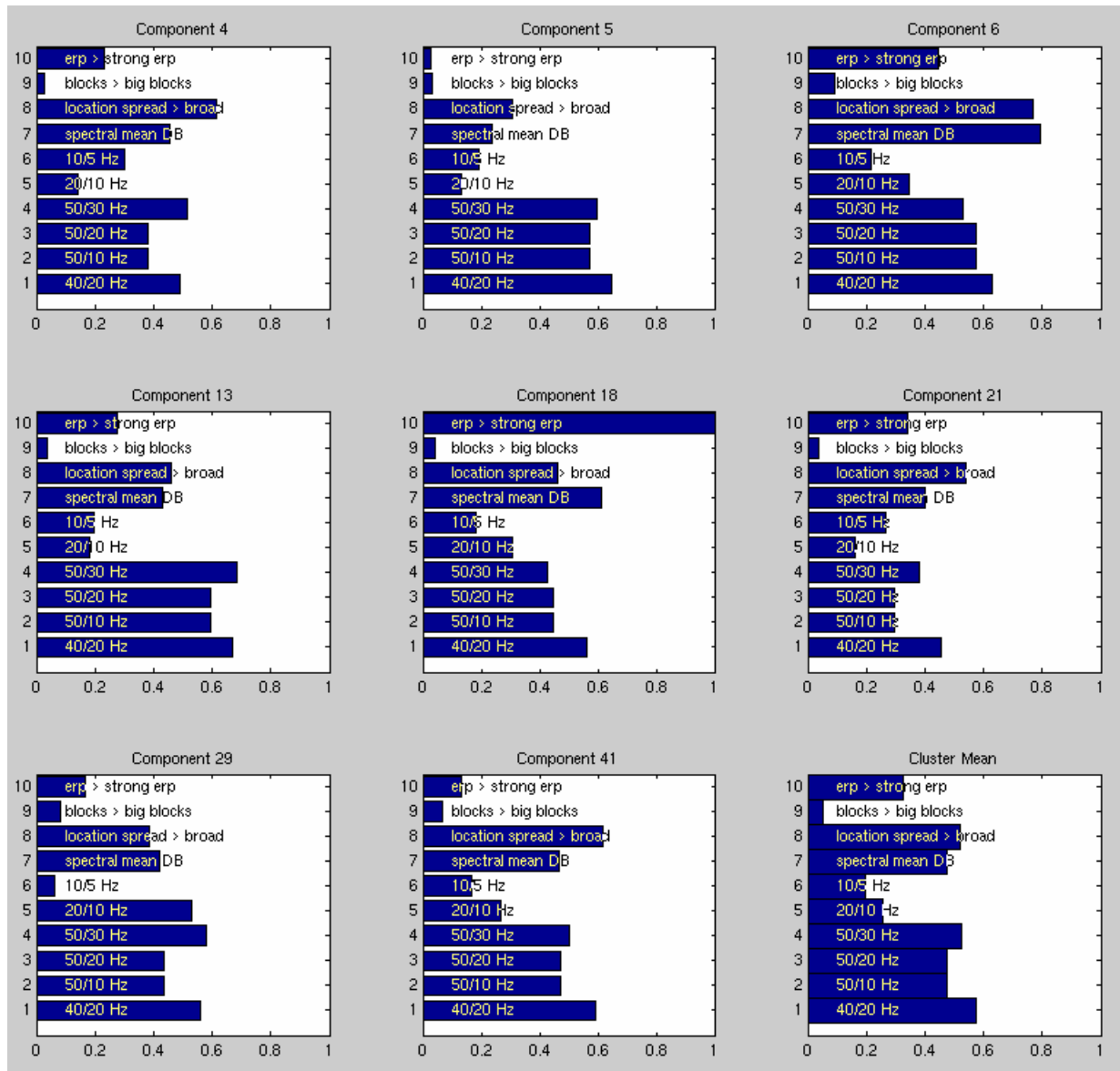


Figure 6 – Overview of cluster data.

To see, why the tool clustered components together, the user may have a look at the statistical data fed into the clustering process.

All views are available for only one component or for the components of a whole cluster. It is also possible to view 2D plots of all clustered dimensions (see Figure 59).

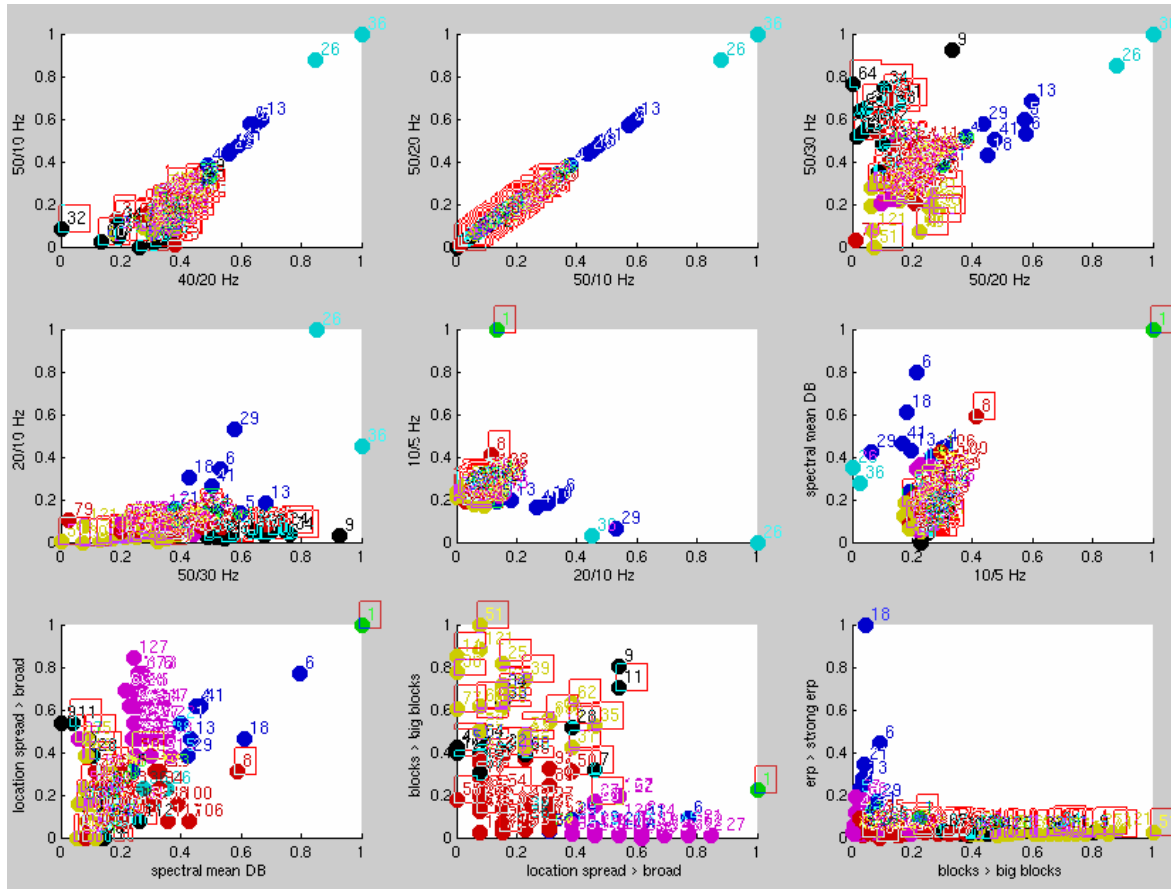


Figure 7 – View of all clustering dimensions.

As the PCA view of the clusters is a very reduced view on the data, the user can view all dimensions underlying the clustering process.

Summarizing, the tool offers a quite reliable method to suggest components for rejection and to match specific rejection criteria persistently. It does not automate the process of artifact rejection, but it helps to solve or reduce the problems of the process described above.