Clustering of ICA Components for efficient EEG artifact detection

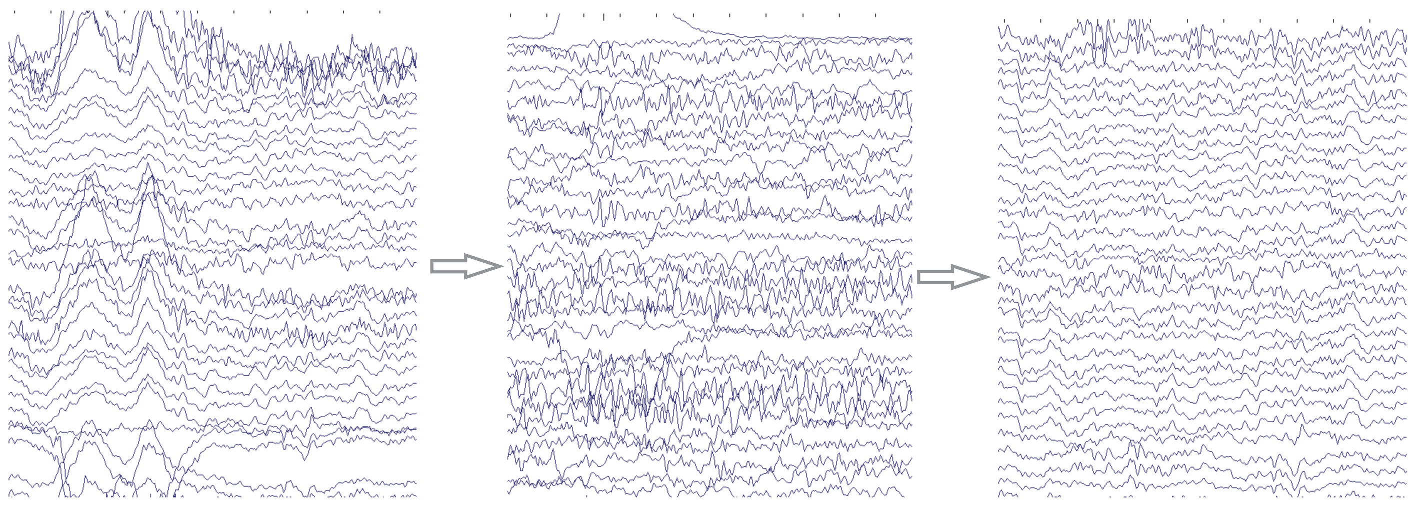
Stefan Scherbaum, Institut für Allgemeine Psychologie, Technische Universität Dresden.
Stefan.Scherbaum@web.de

TECHNISCHE UNIVERSITÄT DRESDEN

Independent Component Analysis and the Rejection of EEG Artifacts

Independent Component Analysis (ICA) is a method to extract statistically independent source signals from a recorded mixture of signals without prior knowledge about the environment (for an introduction, see Stone, 2002)

For EEG data, ICA can separate artifacts like eye blinks or muscle tension from brain sources. The artifacts can then be removed and the electrode data can be reconstructed without the artifactual signals (Jung, Makeig, Humphries, Lee, McKeown, Iragui and Sejnowski, 2000). In comparison to other methods, based on assumptions theoretically not justifiable for EEG data (e.g. principal components analysis), ICA maximally separates artifactual information from the rest of the data enabling a complete reconstruction of the original signals (Iriarte, Urrestarazu, Valencia, Alegre, Malanda, Viteri and Artieda, 2003).

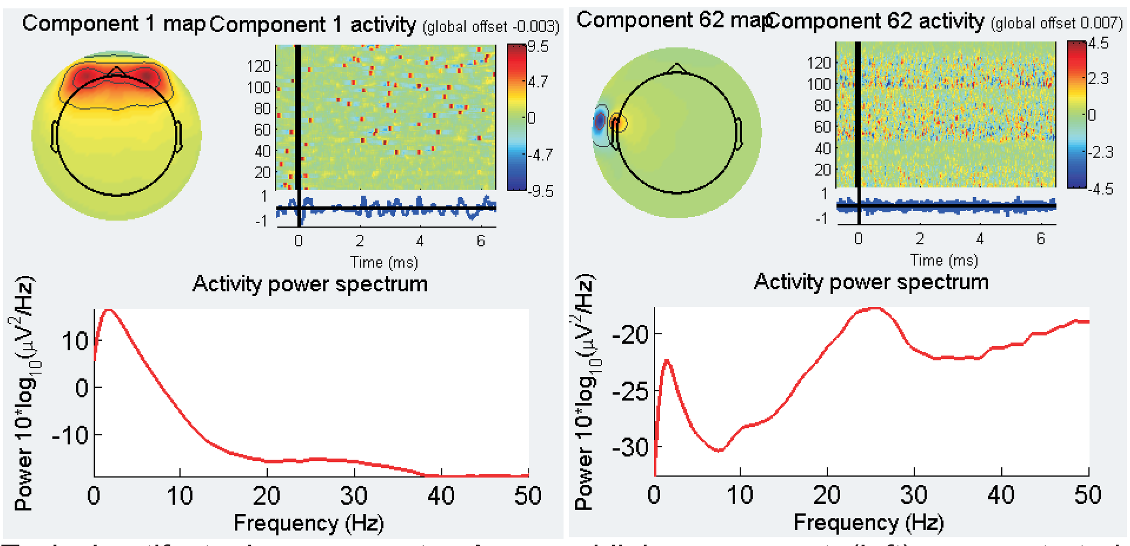


ICA decomposes the original signals (left) into a number of components (middle). Components representing artifactual signals (eye blinks, eye movements, and muscle artifacts) can then be removed without affecting the brain signal part of the recorded data (right).

Looking for Needles in a Haystack - Classification of ICA Components by Clustering

ICA clearly separates EEG data into independent components, but the classification of each component as artifactual or brain based has to be done by the user and cannot be automated easily (Jung, Makeig et al., 2000). Extracted by blind source separation, 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 calls for personal experience and implicates several pitfalls:

- Basing the classification solely on personal experience can lead to inconsistent decisions. Until a user reaches a reasonable level of experience, the task is very erroneous.
- Manual classification of components is time consuming, as one needs to keep track of several features of each component for correct classification. This can also lead to a reasonable loss of attention over time.



Typical artifactual components: An eye blink component (left), concentrated around the frontal electrodes, shows a specific spectrum and a spiky activation pattern. A muscle tension, focused around a specific location, often shows a blocked activation pattern and a high frequency spectrum.

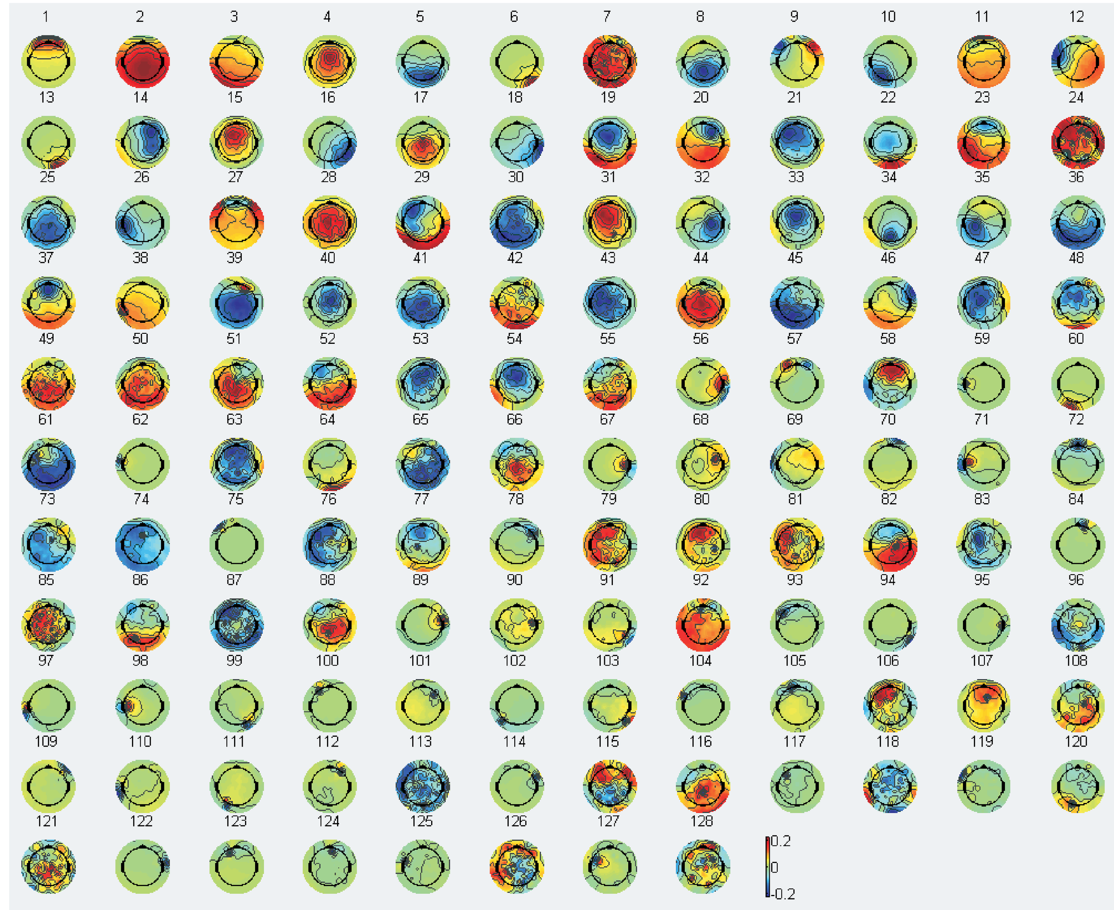
To support the user during this process, I developed a clustering tool that uses statistical measures to pre-classify components. It enables the user to get an overview or gist of each group of components. This idea was inspired by the semi-automatic trial rejection mechanism included in EEGlab (Delorme & Makeig, 2004), which makes trial rejection less arbitrary by proposing trials to reject, based on a classification by several measures as kurtosis, probability of data etc.

The following properties were used to classify components as artifacts (compare Delorme, Makeig, et al., 2001):

1. The frequency spectrum: Specific artifacts show specific distributions of frequency power that are different to normal EEG spectra. These differences are quantified by several ratios of spectral power (40/20 Hz, 50/10 Hz, 50/20 Hz, 50/30 Hz, 20/10 Hz, 10/5 Hz) and the overall spectral power.

2. The topography (the weight matrix): Muscle artifacts concentrate around few electrodes, while eye blinks concentrate around frontal electrodes. This property can be quantified by the median of each component's weights.

3. The activation over trials: Muscle artifacts often occur only for one block of trials, showing low activation for the rest of the trials. I measured this property by convoluting a gaussian function with the added absolute values of each trials activation. The size of the gaussian is determined by the number of trials in each block of the experiment.



Detecting artifacts within a huge number of components is a difficult and time-consuming task.

4. The ERP: A distinctive ERP indicates brain signals. The ERPness was detected by the kurtosis of the average over trials.

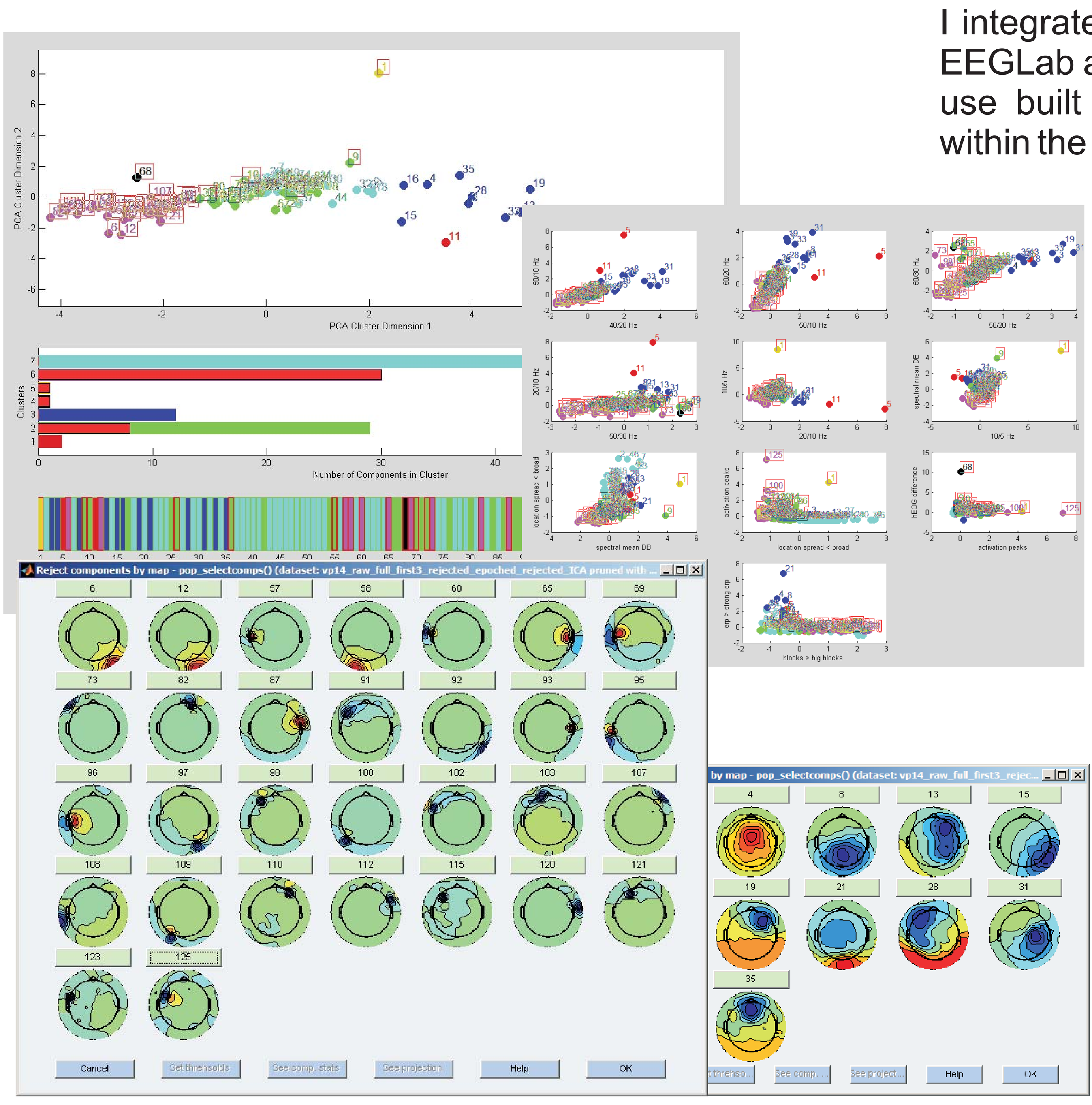
5. The peakiness of the activation over time: Artifactual signals often show a very peaky signature. This property was measured by the kurtosis of the component's activation.

Overall, each component is described by 11 dimensions. Using kmeans to cluster the data with 100 iterations, 7 clusters emerged as the best tradeoff between clearness of classification and reliability of classification.

An EEGLab Plugin for the Clustering and Rejection of EEG Artifacts

EEGLab (Delorme & Makeig, 2004) is a free open source EEG software, based on Matlab. It uses ICA for artifact rejection and for the analysis of EEG data. The latest version supports the clustering of ICA components over subjects for statistical analysis based on similar components.

I integrated the component clustering tool into EEGLab as a plugin, allowing the researcher to use built in displays and new displays from within the main clustering display.

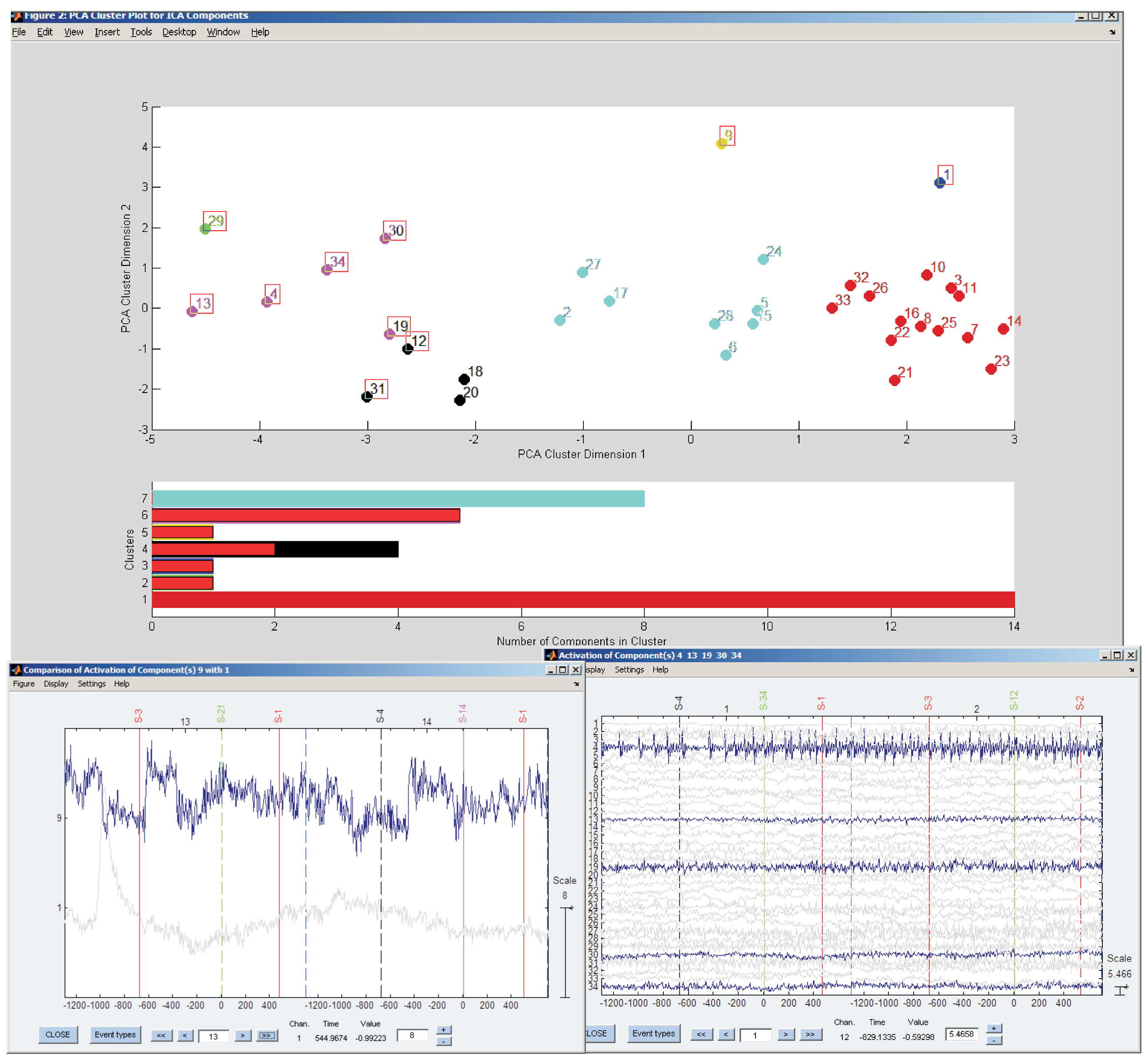


The Main View (upper left) shows the 126 clustered components. Artifacts concentrate in the upper left area while "good" components concentrate in the lower right area. Red rectangles show components already marked for rejection by the user. From this view, the user can choose several other informative views, e.g. a view of the clustering data (upper right) or topographical plots (using built in EEGLab functions) of the components of the left-most cluster (containing only artifacts (lower left)) or of the components of the right-most cluster (containing only brain signal components (lower right)).

After finishing the clustering process, the tool shows a plot of the clustered dimensions (PCA transformed) with all components. From here, the user can choose to view the topographies, clustering data, or activation curves of each component or of all components of a specific cluster. It is also possible to compare the activity of selected components and to watch plots of all clustering dimensions.

In general, the PCA overview offers a gist of the structure of the ICA data. Artifactual components are grouped on one side of the plot while brain signals are found on the opposite side. Only clusters in the middle of the display contain components that can't be securely classified, thus needing more intensive investigation.

This preclassification supports the user by providing a suggestion if a component should be accepted or rejected. Offering convenient access to all information about one component speeds up the classification process.



For the 32 clustered components, the Main View (upper image) shows a clear separation of components into artifactual and signal clusters. The statistical overview shows that only cluster 4 (black) contains uncertain components. The activation comparison view (lower right) extends built in EEGLab plots and helps to clearify if a component contains artifactual information by selectively comparing it with other components. The cluster activation view (lower left) helps to get a better impression of the components of one cluster.

In summary, the tool offers gradual suggestions for the rejection of artifactual ICA components. It does not automate the process of artifact rejection, but it helps to reduce the pitfalls of the manual component rejection and thereby can speed up this process remarkably.

References

- Stone, J.V., 2002. Independent component analysis: an introduction. Trends Cogn. Sci. 6, 59–64.
- Jung T-P, Makeig S, Humphries C, Lee T-W, McKeown MJ, Iragui V,Sejnowski TJ, 2000. Removing electroencephalographic artifacts by blind source separation. Psychophysiology 37, pp 163–178.
- Iriarte J, Urrestarazu E, Valencia M, Alegre M, Malanda A, Viteri C,Artieda J., 2003. Independent component analysis as a tool to eliminate artifacts in EEG: a quantitative study. J. Clin. Neurophysiol. Jul-Aug 20:4, pp. 249-57.
- Delorme, A. and Makeig,S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics.J. Neuroscience Methods, 134:9-21, 2004, 134, pp 9–21.
- Delorme A., Makeig S. and Sejnowski T. J., 2001. Automatic artifact rejection for EEG data using high-order statistics and Independent Component Analysis. In Proc. Of 3rd International Independent Component Analysis and Blind Source Decomposition Conference, December 9-12, San Diego, USA.