AUTOMATIC ARTIFACT REJECTION FOR EEG DATA USING HIGH-ORDER STATISTICS AND INDEPENDENT COMPONENT ANALYSIS

A. Delorme, S. Makeig, T. Sejnowski
CNL, Salk Institute
10010 N. Torrey Pines Road
La Jolla, CA 92107, USA

ABSTRACT

While it is now generally accepted that independent component analysis is a good tool for isolating both artifacts and cognitive related activations in EEG data, there is still little consensus about criteria for automatic rejection of artifactual components and single trials. Here we developed a graphical software to semi-automatically assist experimenter in rejecting independent components and noisy single data trials based on their statistical properties. We used kurtosis to detect peaky activity distributions that are characteristic of some types of artifact and entropy to detect unusual activity patterns. EEG-LAB, a user-friendly graphic interface running under Matlab, allows the user to tune and calibrate the rejection criteria, to accept or override the suggested components and trials labeled for rejection, and to compare the results with other rejection methods.

1. INTRODUCTION

It is critical to detect artifact contamination of evoked-potential EEG data for several reasons. First, artifactual signals usually have high amplitudes. Thus, even if their distribution in the recorded EEG is sparse, they can bias evoked potential averages constructed from the data and, as a consequence, bias the results of an experiment. Rejecting artifacts is also critical when data are used in further processing to interpret brain area activity - e.g. event related spectral perturbations or Independent Component Analysis (ICA) [1].

In most current EEG software packages, single data trials are rejected when they contain out-of-bounds potential values at single electrodes. In event-related experiments, each data epoch normally represents a single experimental trial. Usually, the experimenter first subtracts a baseline – e.g., the average potential before the stimulus occurs – from each trial. Then, the experimenter chooses rejection electrodes at which potential values should not exceed some defined threshold value. The selected electrodes usually include central scalp placements which record a large portion of brain activity, parietal placements including temporal muscle artifacts and/or frontal electrodes containing blinks and eye movement artifacts. The problem with this thresholding process is that it only takes into account low-order signal statistics (min, max). This kind of rejection might not be sufficient to detect muscle activity for instance, which typically involves rapid electromyographic (EMG) signals of moderate size. Higher order statistical properties of the EEG signals might contain more relevant information about this and other types of artifacts.

Independent Component Analysis (ICA) applied to collection of single trials EEG data has proven to be efficient for separating distinct artifactual and neural EEG processes [1]. ICA is able to separate brain activation processes or artifacts whose time waveforms are (maximally) independent of each other. For instance, eye movements and muscle activities produce specific ICA activation patterns and component maps [2][3]. While ICA is now considered an important technique for removing artifacts, there is still little consensus about the characteristics of artifact components. In this paper, we develop a framework for semi-automatic artifact rejection based on data and independent component activity statistics. We first apply these measures to the raw data. Then, we apply ICA to the data and use statistics on the independent component activities to locate and reject both artifactual data trials and artifactual components. EEG-LAB, a user-friendly graphic interface under Matlab, guides the user through this process (www.salk.edu/~arno/eeglab.html). The EEG-LAB software is based in part on the ICA/EEG toolbox of Makeig et al. (www.salk.edu/~scott/ica.html) and implements some of its functions to help the user visualize the data and make decisions about which components to reject. To test and illustrate the rejection process, we used event related EEG data from a go-nogo visual categorization task [4] (32-electrode montage, ears referenced, 1000 Hz sampling rate, 1200 target and non-targets trials per subject).

2. ARTIFACT REJECTION ON RAW DATA

Most artifacts are typically “odd” data in the sense that they are transient and unexpected events. Isolating artifacts thus involves detecting such events. To do so, we chose two measures: probability distribution and kurtosis. The probability

of a data trial activity, given the probability distribution of activity in all data trial, is a measure of the data trial's "oddness". If the probability value is low, it means that the activity values in a data trial at a given electrode are unexpected, given the probability distribution of other data trials' activity at the same electrode. To estimate the relative probability of each trial of our raw data, we first computed the observed probability (D_e) density function of all data value at each electrode *e*. Then, we computed the logarithm of the joint probability $J_e(i)$ of the activity values (A_i) of each single data trial *i* at electrode *e*:

$$J_e(i) = -\log(\prod_{x \in A_i} p_{D_e}(x)) \qquad (1)$$

$p_{D_e}(x)$ being the probability of observing the value x in the probability distribution D_e of activity at electrode *e*. We used the logarithm of the joint probability for better graphic presentation of very low joint probability values. The joint probability was computed for every single data trial at each electrode.

The probability measure allows us to detect outlier trials. We might be able to spot other artifact trials based on their unusually peaky distribution of potential values. In some artifact trials, the distribution of activation is very peaky, for example a trial containing strong transient muscle activity. To measure this peakyness, we used the kurtosis of the activity values in each trial. Kurtosis is the 4th cumulant of the data – the mean being the first cumulant and the variance the second. For each trial at each electrode, the kurtosis was calculated using the following equations:

$$K = m_4 - 3m_2^2 \qquad (2)$$

$$m_n = E\{(x - m_1)^n\} \qquad (3)$$

where m_n is the n^{th} central moment of all activity values of the trial, m_1 the mean, and E the expectancy function (in our case the average). Typically, if the kurtosis is highly positive, the distribution of activity is peaked and sparse, and the identified data is likely to be an artifact. If the all activity values are similar, the kurtosis will be highly negative. Once more, this type of activity is not typical of true EEG signals, which reflect non-stationary processes, so low kurtosis values make us suspect the presence of artifactual data. Negative kurtosis values usually reflect AC (alternating current) or DC (direct current) artifacts, for example those induced by line or screen currents or by loose electrode contacts.

Before defining rejection thresholds for both entropy and kurtosis, we first normalized these measures (to 0-mean and standard deviation 1). We were thus able to define thresholds in terms of a number of standard deviations from the mean (figure 1). We usually used thresholds higher than 5, which means than, assuming a gaussian distribution of the true EEG activity values at one electrode, the probability that an identified trial belongs to the distribution is less than $1.6*10^{-12}$. Assuming that all the true EEG activity values belong to a single gaussian distribution is unrealistic. Yet, given the large difference between the true EEG activity values and outliers, this approximation is not unreasonable in this context.

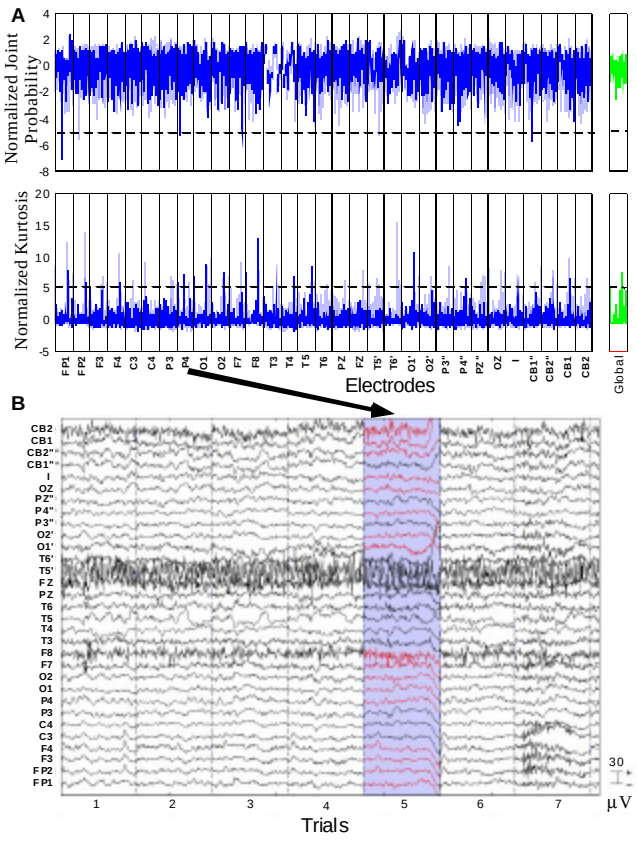


Figure 1. Rejection of artifact-laden trials on the statistics of single electrode activity in the raw data. A. Probability and kurtosis of the observed activity at each electrode for each trial. The horizontal dashed bars indicate rejection thresholds (in numbers of standard deviations from the mean). If the measure for one trial exceed one of these rejection thresholds, the trial is marked for rejection. B. Visualization of data at each electrode with suggested epoch labeled for rejection (the arrow indicates one of the electrodes for which the statistical measure on the labeled trial exceeds the threshold). Before accepting the suggested rejection, the user can review these plots. Rejected trials are depicted with a colored background and rejected components, with a specific color (not shown).

3. INDEPENDENT COMPONENT ANALYSIS AND EEG

3.1 Why use ICA

Having estimated these higher statistical properties of the signal, one might ask, why should we go further? All the measures we have used so far are based on raw potential values at single electrodes. However, EEG activity at different electrodes is highly correlated and thus contains redundant information. Also, several artifacts might be represented at the same set of electrodes and it would be useful if we could isolate and measure these artifacts based on their projection to overlapping electrode subsets. This is what ICA does [1][5]. Intuitively, one can imagine an *n*-electrode recording array as an *n*-dimensional space. The recorded signals can be projected into a more relevant coordinate frame than the single-electrode space: the independent component space. In this new coordinate frame, the projections of the data on each basis vector – i.e. the independent components – are maximally independent of each other. Assessing the statistical properties of the data reprojected onto these axes, we might be able to isolate and remove the artifacts more easily and efficiently.

ICA has been shown to be more efficient for this purpose than other algorithms such as principal component analysis (PCA) [2][3]. There are several reasons to use ICA. First, several major assumptions of using ICA seem to be fulfilled in the case of EEG recordings (for a detailed justification, see [1][6]). The first assumption is that the ICA component projections sum linearly onto the electrodes. This is actually the case for EEG in which different electrical sources – individual muscle and brain activations – sum linearly at scalp electrodes following Ohm's law. The second hypothesis is that sources are independent. This is not strictly realistic but even if artifact might be related to cognitive activation – a muscle contraction for instance is usually triggered by activity in the motor cortex – the time course of the artifact and the triggering event may be different in some or all trials. Thus, they will be accounted by different independent components [2]. A third assumption concerns the gaussianity of the sources: the source activity – i.e. the projection of the data onto the ICA basis vector – must not have a gaussian distribution. This last condition is quite plausible for artifacts, which are usually sparsely active and thus far from gaussian.

From the original data U, ICA finds a new projection space. Multiplication of U by the unmixing matrix W – found for example by infomax ICA [5] – represents a linear change of coordinate system, from the electrode space to the independent component space:

$$S = W * U \qquad (4)$$

where S is the matrix of activations of the components across time in the new ICA coordinate frame (as shown in figure 3C). In EEG, each component is a linear weighted sum of the activity waveform at all the electrodes (the weights being a row of W). The independent components comprise these activations and the associated scalp maps give the projection weights of the components back into the electrode space. It is possible to derive the component spatial maps, as illustrated in figure 3A, as the columns of W^{-1}. For computing and visualizing ICA components, the software we developed uses the ICA Matlab toolbox for EEG of Makeig et al. (www.salk.edu/~scott/ica.html).

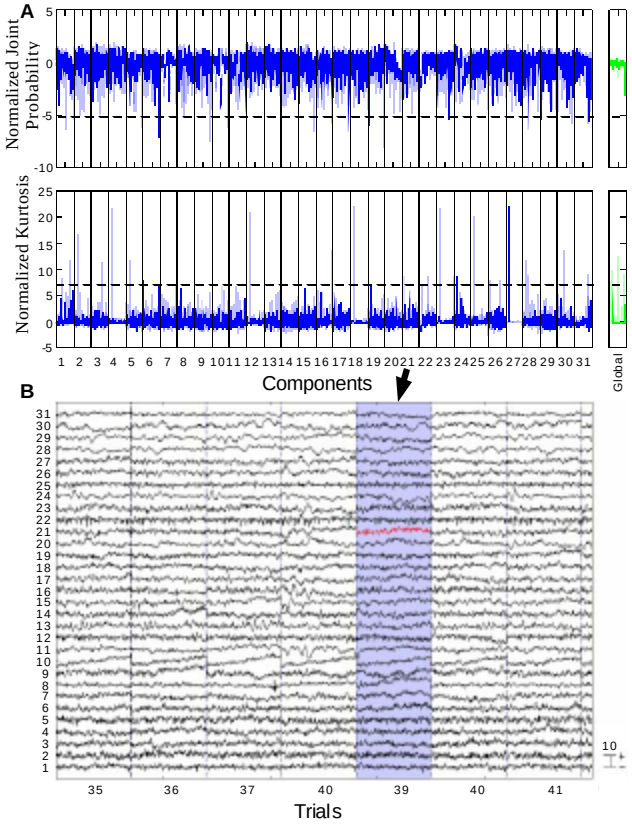


Figure 2. Detection of an artifact-laden trial based on the statistics of spatial independent components of 31-channel EEG. A. Probability and kurtosis of each component's activity in each trial. The horizontal dashed bars indicate rejection thresholds (in numbers of standard deviations from the mean). If the measure at one trial exceeds these rejection thresholds, the trial is marked for rejection. B. Activation of each component with trials marked for rejection (the arrow indicates which component was used to mark the labeled trial for rejection). Before accepting the rejection, the user can review these plots. Rejected trials are depicted with a colored background and rejected components, with a specific color (not shown).

3.2 ICA statistics and trial's rejection

Having computed the independent components of the data, we can try to detect outlier trials. As with raw data, we can compute both the probability and kurtosis of each trial for each component (see equations 1,2 and 3 above). As already point out, these measures can help us detect unexpected activity (using probability) or activity with an unusually peaky distribution (using kurtosis).

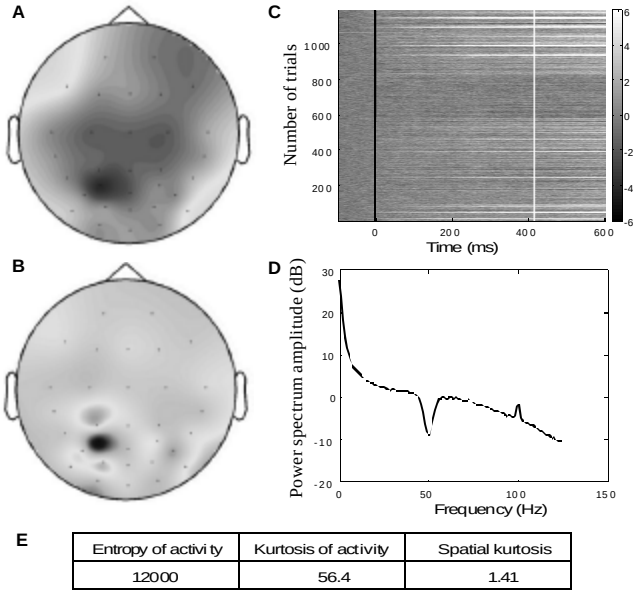
For rejecting single trials, we used either the joint probability or kurtosis of data in all components. As with the electrode data, we normalized the values of probability and kurtosis and then selected rejection thresholds in terms of number of standard deviation from the mean (figure 2). To refine the rejection criteria, we also defined a maximum number of components for which each measure should not exceed the rejection threshold. For instance if the maximum was set to 3 components and the data trial exceeded the rejection threshold at 4 components or more, then the trial was marked for rejection. The basic idea is that if the measure exceeds the rejection threshold at many components – i.e. the trial an outlier many components – then it is very likely to be an outlier trial in the original data. One might imagine that an artifact trial would only be expressed in a single independent component. The reality is more complex for various types of unpredictable noise (i.e. from coughing and other head movements), and several components are affected. Another example might be activity that saturates at some recording sites, thus altering the recording of other electrical sources from true EEG. As a consequence, the spatial and temporal structure of several components, found by preliminary ICA on the raw data, might be biased.

3.3 Component rejection and user calibration

While it is possible to reject artifact trials from component activity, it is also possible to subtract whole artifactual components from the original data. Instead of removing all the trials with muscle artifact activity for example, ICA decomposition allows the experimenter to subtract from the data the activity belonging to the artifactual muscle component. One can manually and visually rejects components based upon their spatiotemporal characteristics (figure 3), but one can also set a rejection threshold for relevant higher order statistics. We used three high-order statistical measures for each component: the entropy of the activity of the component (over all trials), the kurtosis of the activity and the kurtosis of the components' spatial map. Entropy for component *i* is defined as

$$H(i) = -\sum_{x \in D_i} p_{D_i}(x) \log(p_{D_i}(x)) \quad (5)$$

where $p_{D_i}(x)$ is the probability of observing the activity values *x* in the observed probability distribution of activity D_i from component *i*. The use of the term entropy might be abusive because the measure computed in equation 5 is only a rough approximation of the true entropy of a component. The component activity values are not independent across time and trials but, because the distribution of activity of outlier components is very different from the standard distribution, this measure should be able to detect these outlier components. Additional terms could be added to equation 5 for minimizing the dependence of the entropy measure on the time step increment [7]. As a second measure, we used the kurtosis of the component's activity to detect sparse and peaky activity



Entropy of activity	Kurtosis of activity	Spatial kurtosis
12000	56.4	1.41

Figure 3. Characteristics of a non-muscle artifact component. A. Plot of the spatial map of the independent EEG component. B. plot of the laplacian of the spatial map of the component. The laplacian helps to determine which electrodes are the most active in the component. If only one electrode is active (as it is approximately the case here), the component is likely to be an artifact. C. Plot of the component activity in 1200 single trials. Each single line is a different trial as in the ERP-image plotting format [6]. The color scale limits is arbitrary as ICA cannot retrieve the absolute amplitude of source activity. If the component is only active in a few trials, if it is not related the timing of the stimulus, or if it contains very high frequencies (>100Hz), it is very likely to be an artifact. D. The power spectrum of the component activity showing which frequencies dominate the component activity. If the component contain a homogenous distribution of high frequencies, it is likely to be an artifact component (note the notch filter at 50 Hz to remove noise from electrical lines). E. the three high-order statistical characteristics of this component (see text for details).

distribution, which are characteristic of some artifactual components (see equation 2 and 3). The last measure we used was the kurtosis of the spatial projection of the component. If the spatial projection of an artifactual component is peaky – for instance only one electrode is activated – then its kurtosis will have a high value. We will then be able to automatically detect and reject such components by setting an adequate rejection threshold.

Besides these three measures, the EEG-LAB user is presented with 4 graphical representations of the component: its spatial map, the laplacian of its spatial map, its individual trial activity and its frequency content. These representations provide the user with several types of information on which he can base his judgment for rejecting components (figure 3). Some spatial distributions for instance are typical for eye or muscle movements. They usually involve a single electrical polarity (positive or negative) and extend towards the edge of the skull (not shown).

To assess the efficiency of the three component statistics, we considered four classes of artifact components – eye blinks, eye movements, muscle activity and other types of artifacts. We visually rejected the components for four subjects performing a visual go-nogo categorization task of natural images (see the introduction for details on the experiment). This manual rejection was based upon the visual component characteristics as shown in figure 3. We then computed the probability that a component was rejected by this method as a function of the three statistical measure values. Figure 4 shows that all three measure values are correlated with selection as artifact and so can be used to detect artifacts. We observed that highest values of these measures seem to be characteristic of muscle components whereas intermediate values may be characteristic of non-muscle artifacts, and low values of non-artifact components. For eye blinks and eye movements, we only had two components per subject and we were not able to compute any statistics. However, the measures we computed might also be used to reject these components (for eye blink components: activity entropy 13400±3200, activity kurtosis 86±68 and spatial kurtosis 4.8±1.5; for eye movements: activity entropy 8580±4900, activity kurtosis 82±55 and spatial kurtosis 9.8±4.9). Though these results are still preliminary, they show that there is a relation between the statistics of the components and their rejection by at least one human experimenter. In the software we developed, the user can define thresholds for these three measures. Another screen shows the scalp maps of all the components, those marked for rejection by the program colored red. Buttons allow the user to review each component and to add and subtract from the rejection list. The final list chosen might be saved and used to calibrate subsequent processing.

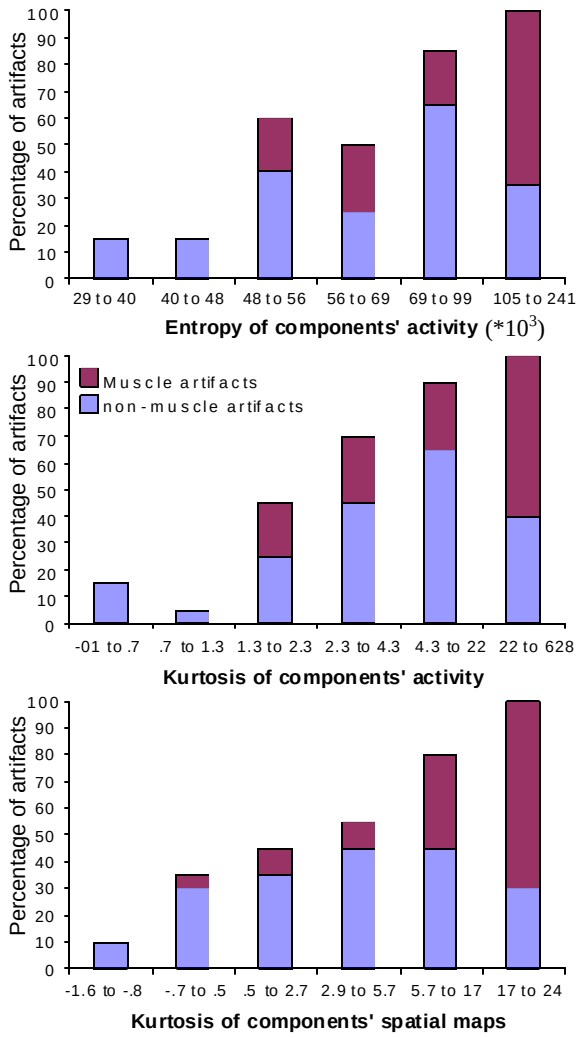


Figure 4. Component entropy and kurtosis as a function of experimenter rejections based on visual inspection of plots like figure 3 (4 subjects, 120 components out of which 65 were labeled as artifacts on visual inspection by the experimenter, A.D.). For each measure, we assessed the probability of the experimenter rejections. To compute the probabilities, we sorted the components by their entropy or kurtosis values and then partitioned the components into 6 consecutive groups of 20 (if 10 components in one of the groups are artifacts, then the artifact probability is 0.5). We considered 2 subconditions: muscle components (26), non-muscle components (39). We can see that the measures were quite accurate for detecting artifact. Extreme positive values of any of the measures corresponded to 100% of artifact components.

4. SUMMARY

As shown in figure 5, the EEG-LAB software we developed allows the user to combine and compare all types of rejection. We showed that using high-order statistics of the raw data and of the independent components, we may be able to semi-automatically reject trial artifacts. We also showed that high-order statistics of independent component activity are strongly correlated with the artifact/non-artifact distinction at least as defined by common practice in our laboratory. After artifact rejection, the cleaned data can be decomposed by ICA and/or other analysis method to study true EEG activity.

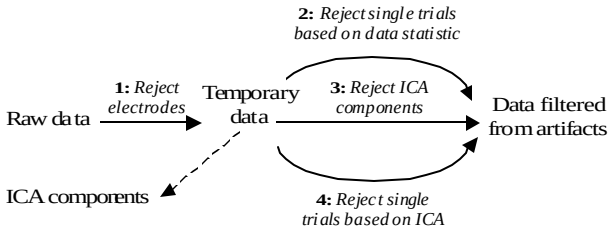


Figure 5. Schema for combining different types of rejection. After rejecting bad electrodes (1) and computing ICA, the algorithm combines three types of rejection. Trials can be rejected depending on the data statistics or the independent component statistics (respectively 2 and 4). Subtraction of artifactual independent components can also be performed (3).

We believe that one may detect artifacts more accurately using high-order statistical measures of the signals, regardless of the exact implementation of these measures. This approach allows experimenters to use information in the data that was taken into account by standard rejection methods. The graphical software we developed is open-source (software available at www.salk.edu/~arno/eeglab.html), with full documented source's code, and is organized hierarchically, meaning that complex graphical functions call simpler functions. It is thus easy to add new features. It also allows the user to directly call artifact rejection commands from a Matlab batch file and to combine these functions with other Matlab commands.

5. REFERENCES

[1] Makeig S., Bell A.J., Jung T.P. and Sejnowski T. J. "Independent component analysis of electroencephalographic data". *Advances in Neural Information Processing Systems*, Touretzky D., Mozer M., and Hasselmo M., editors, 1996, pages 145-151.

[2] Jung T. P., Makeig S., Humphries C., Lee T. W., McKeown M. J., Iragui V. and Sejnowski T. J. "Removing electroencephalographic artifacts by blind source separation". *Psychophysiology*, 37(2):163-178, 2000.

[3] Jung T. P., Makeig S., Westerfield M., Townsend J., Courchesne E. and Sejnowski T. J. "Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects". *Clinical Neurophysiology*, 111(10):1745-1758, 2000.

[4] Fabre-Thorpe M., Delorme A. and Richard G. "A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes". *Journal of Cognitive Neuroscience*, 13(2):171-180, 2001.

[5] Bell A.J. and Sejnowski T.J. "An information-maximization approach to blind separation and blind deconvolution". *Neural Computation*, 7(6):1129-1159, 1995.

[6] Makeig S., Westerfield M., Jung T. P., Covington J., Townsend J., Sejnowski T. J. and Courchesne E. "Functionally independent components of the late positive event-related potential during visual spatial attention". *Journal of Neuroscience*, 19(7):2665-2680, 1999.

[7] Cover T.M. and Thomas J.A. *Elements of Information Theory*. John Wiley & Son, Inc., 1991.