MARA Tutorial

April 23, 2014

Contents

1 Introduction

This is the tutorial for MARA, an open-source EEGLAB plugin which automatizes the process of hand-labeling independent components for artifact rejection [8]. The core of MARA is a supervised machine learning algorithm that learns from expert ratings of 1290 components by extracting six features from the spatial, the spectral and the temporal domain. Features were optimized to solve the binary classification problem "reject vs. accept". Thus, MARA is not limited to a specific type of artifact, and should be able to handle eye artifacts, muscular artifacts and loose electrodes equally well.

MARA is written in MATLAB and makes extensive use of EEGLAB's functionality including its excellent visualization tools [2].

1.1 Requirements

In addition to the requirements of EEGLAB (`http://sccn.ucsd.edu/eeglab/`), MARA needs the Matlab Statistics Toolbox, the Optimization Toolbox and the Signal Processing Toolbox.

MARA requires channel locations with labels in the form FPz, F3, Fz, F4, Cz, Oz, ..., and components need to be at least 15s long for the computation of the temporal feature. Note that MARA might lead to suboptimal results on narrow band-passed filtered data, as the spectrum features are calculated on the power spectrum between 2 Hz and 39 Hz. (You can deactivate the frequency features if you are familiar with MATLAB scripting.)

1.2 Citing MARA

Like EEGLAB, MARA is free software distributed under the GNU General Public License [1]. If you find this toolbox useful, please refer EEGLAB and the following publication:

> I. Winkler, S. Haufe, and M. Tangermann. Automatic classification of artifactual ICA-components for artifact removal in EEG signals. *Behavioral and Brain Functions*, 7:30, 2011.

2 MARA in brief

Artifact rejection with Independent Component Analysis (ICA) is based on the idea that artifactual components and neuronal activiy are generated independently. The EEG signal is decomposed into independent source components (ICs) in the hope that artifactual components are separable from brain signals. MARA's goal is to identify the components that are predominantly driven by artifacts. The EEG can then be reconstructed without the artifactual components.

2.1 Process Chain for ICA Artifact Rejection

The typical process chain for ICA artifact rejection consists of the following steps:

1. A rough pre-cleaning of the data by e.g. channel rejection and trial rejection may be performed. This step is usually helpful for obtaining a good ICA decomposition. EEGLAB uses statistical thresholding to suggest epochs to reject from analysis and provides a rich class of functions to do so (such as Tools>Automatic Channel Rejection or Tools>Automatic Epoch Rejection). For more information, see `http://sccn.ucsd.edu/wiki/Chapter_01:_Rejecting_Artifacts`. EEGLAB authors suggests to run ICA twice, first to reject epochs based on the IC time courses, second to obtain a good ICA decomposition for the cleaner data.

2. As ICA decomposition is known to be sensitive to slow drifts, high pass filtering the data (at 0.5 Hz or even 2 Hz) can sometimes improve the quality of the decomposition. In EEGLAB, this can be achieved using Tools>Filter the data>Basic FIR filter. For more information, see `http://sccn.ucsd.edu/wiki/Chapter_04:_Preprocessing_Tools`.

   Note that MARA might lead to suboptimal results on narrow band-passed filtered data, because its spectrum features are calculated on the power spectrum between 2 Hz and 39 Hz.

3. Tools>Run ICA calculates ICA decompostion (it might take a while), see `http://sccn.ucsd.edu/wiki/Chapter_09:_Decomposing_Data_Using_ICA`.

   The option 'pca' can be set to perform a dimensionaliy reduction prior to IC computation. Such a step may be helpful, in order to reduces the noise level and avoid an unnatural splitting of sources. (It also makes IC computation faster and reduces and the number of components that have to be labeled.)

4. To study and identify artifactual componets by visual inspection, EEGLAB provides the Tools>Reject data using ICA>Reject components by map function. EEGLAB also comes with a powerful class of plotting functions under the Plot menu.

   This is the step where you might want to use MARA, as it automatizes the identification of artifactual components.

5. After artifactual ICs have been identified, the EEG can be reconstructed without them. In EEGLAB, this can be called via Tools>Remove components.

2.2 MARA

MARA realizes automatic IC classification by a linear pre-trained classifier. It is based on the following six features which were determined in a feature selection procedure described in [8]. One feature aims to detect outliers in the time series of an IC, three features are extracted from the spectrum, and two feature extract information form the scalp map of an IC.

1. ***Current Density Norm***: ICA itself does not provide information about the locations of the independent source components. However, ICA scalp maps can be interpreted as EEG potentials for which the location of the sources can be estimated. MARA considers 2142 locations arranged in a 1 cm spaced 3D-grid and seeks the source distribution with minimal l_2-norm [5, 6] (i.e. the simplest solution of the forward problem formulated according to [3, 4, 7]). Since this source distribution can model cerebral sources only, it is natural that artifactual signals originating outside the brain can only be modeled by rather complicated sources. Those are characterized by a large l_2-norm, which MARA uses as a feature.

2. ***Range Within Pattern***: The logarithm of the difference between the minimal and the maximal activation in a scalp map. Spatially localized scalp maps stemming from e.g muscle artifacts or loose electrodes are typically characterized by a high Range Within Pattern.

3. ***Mean Local Skewness***: The mean absolute local skewness of time intervals of 15 s duration. This feature aims to detect outliers in the time series.

4. $\boldsymbol{\lambda}$ and ***Fit Error***: These two features describe the deviation of a component's spectrum from a prototypical 1/f curve and its shape. The parameters $k_1, \lambda, k_2 > 0$ of the curve

$$f \mapsto \frac{k_1}{f^\lambda} - k_2 \tag{1}$$

are determined by six points of the log spectrum: (1) the log power at 2 Hz, (2) the log power at 3 Hz (3) the point of the local minimum in the band 5–13 Hz, (4) the point one Hertz below the third point of support (5) the point of the local minimum in the band 33–39 Hz, and (6) the point one Hertz below the fifth point of support. MARA uses the logarithm of λ and the mean squared error of the approximation of f to the real spectrum in the 8–15 Hz range as features.

The spectrum of muscle artifacts, characterized by unusual high values in the 20–50 Hz range, is approximated by a comparatively steep curve with high λ.

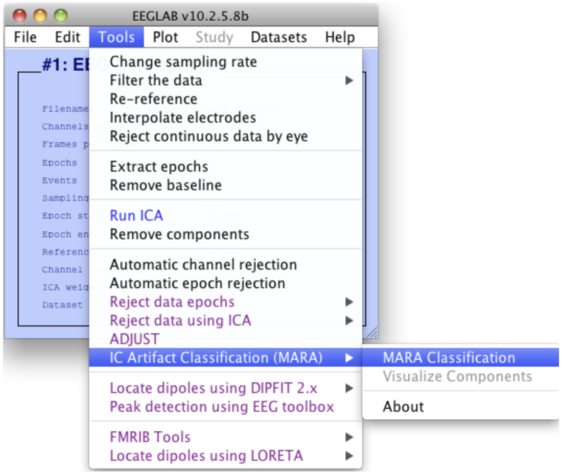
5. ***8-13 Hz***: The average log band power of the α band (8–13 Hz). This feature aims to detect the typical α peak in components of neural origin.

3 Installation

To install MARA, download the compressed plug-in file, uncompress it, and paste it into your EEGLAB `plugins` folder. EEGLAB will now automatically recognise the plugin. The next time you start EEGLAB, the following line should appear in your matlab command window:

```
EEGLAB: adding plugin function "eegplugin_MARA"
```

After loading a data set, MARA can now be found in the tools menu, Tools>IC Artifact Classification (MARA) :



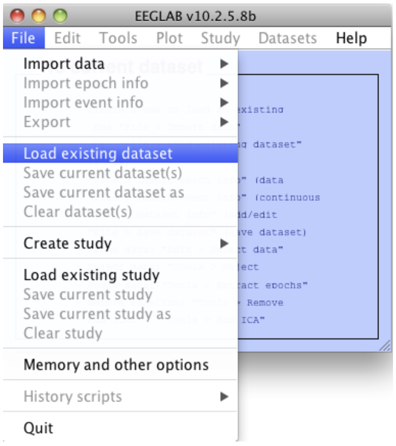
4 Tutorial

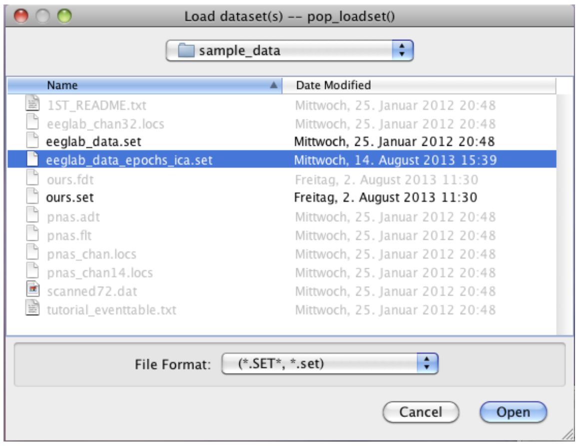
This tutorial takes you step by step through the MARA plugin. For this tutorial, we use the sample data `EEGLAB_data_epochs_ica.set` that you should find in the `sample_data` folder provided with your EEGLAB distribution. Decomposition into IC components has already been performed on this data set.

We assume that you are familiar with EEGLAB and ICA decomposition, which are very well documented in http://sccn.ucsd.edu/eeglab/. More information on ICA decomposition in EEGLAB in particular can be found in http://sccn.ucsd.edu/wiki/Chapter_09:_Decomposing_Data_Using_ICA.

4.1 Loading a sample EEG set

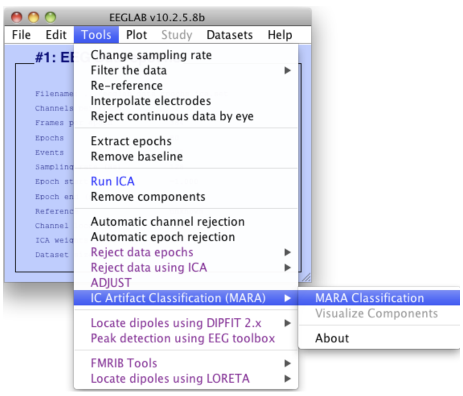
Select File>Load exisiting dataset and browse to your `sample_data` folder. Select `EEGLAB_data_epochs_ica.set`, press 'Open' and the sample data is loaded.



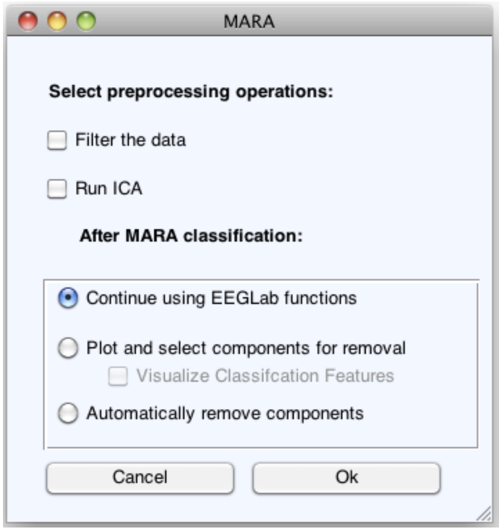


4.2 Starting MARA

To start MARA, go to Tools>IC Artifact Classification (MARA)>MARA classification



The following main menu will open:



4.3 MARA's options

4.3.1 Pre-processing

If you have not pre-processed your data or run ICA yet, you can select to do so by ticking the corresponding checkboxes in the MARA main menu, 'Filter the data' and 'Run ICA'. Note that obtaining a good ICA decomposition is a difficult problem, and you might consider more elaborate pre-processing as outlined in Section 2.1.

1. 'Filter the data' calls the same function as Tools>Filter the data>Basic FIR filter. A pop up appears asking you to provide filtering options. After filtering, you will be asked a name for the new data set.

   Note that MARA might yield suboptimal classification results on narrow-bandpass filtered data as some of its features are based on the power spectrum between 2 Hz and 39 Hz.

2. 'Run ICA' calls the same function as Tools>Run ICA. A pop up appears asking you to provide ICA options. After ICA computation (which might take a while), you will be asked for the name of the new data set.

ICA is already performed on the sample data set we analyse in this tutorial, so we will not go through these steps.

4.3.2 After classification

MARA offers the following three options that concern the course of action after MARA classified the components:

- The default option is to use MARA only to mark components for rejection. The components are not removed and not visualized. You can use standard EEGLAB functions for further investigation and re-labeling of the components.
- If you check 'Plot and select components for removal' MARA will provide its own visualisation of the IC components which allows you to quickly change the classifier's proposed rating. The components will not be removed, you can later inspect components with EEGLAB's plotting functions.

  If you check 'Visualize Classification Features', an additional window will display the six features that MARA classification is based on for each component.
- If you check 'Automatically remove components' you accecpt MARA's decision blindly. The components will be automatically removed.

4.4 Running MARA

Pressing 'Ok' on the MARA main menu will start MARA classification and produce the following output on the Matlab command window:

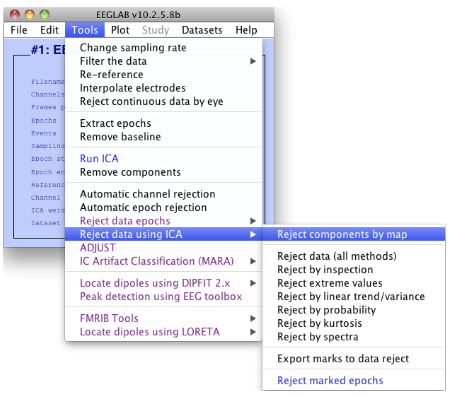
```
MARA is computing features. Please wait
.................................
Features ready
MARA marked the following components for rejection:
     3     11     21     23     24     26     28     31
```

That's it. If you click Tools>Reject data using ICA>Reject components by map you can see the components marked for rejection in red.

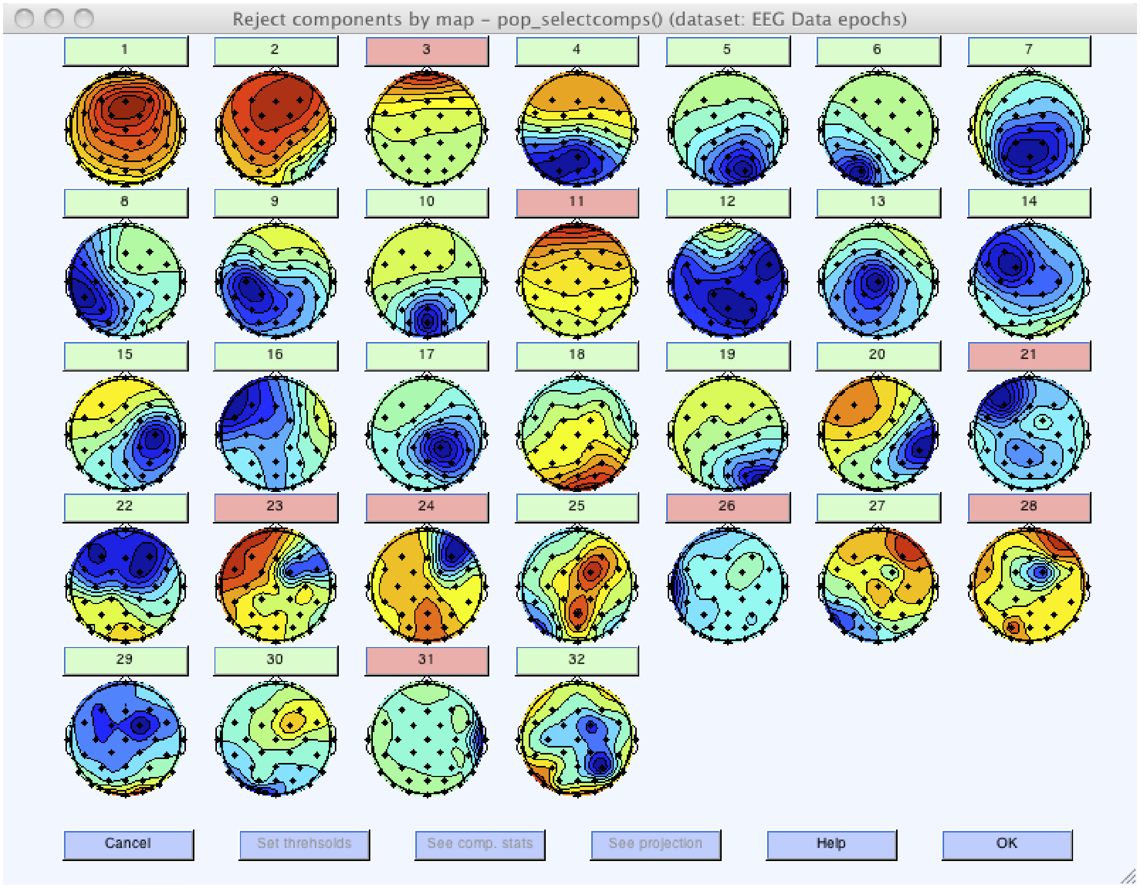
4.5 Visualising components with EEGLAB

After MARA component classification, you can study the selected (and not selected) components using EEGLAB functionaliy that is well described in the EEGLAB Tutorial `http://sccn.ucsd.edu/wiki/Chapter_09:_Decomposing_Data_Using_ICA`. Here, we just give a very brief overview.

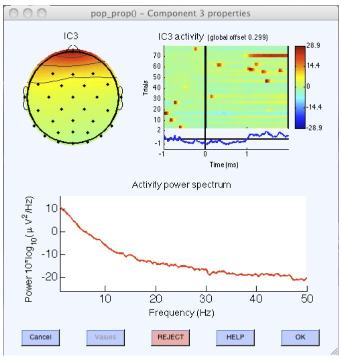
Go to Tools>Reject data using ICA>Reject components by map.

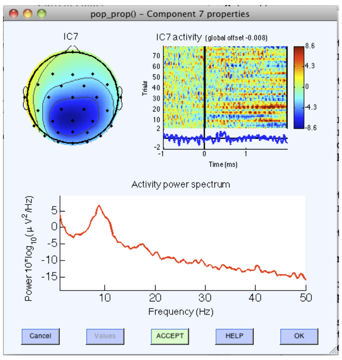


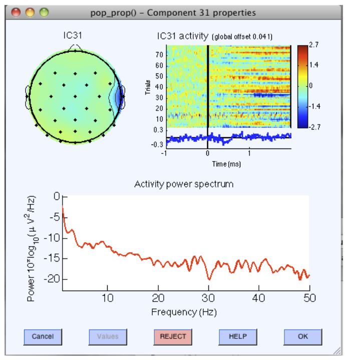
The following window will pop up, where components selected for rejection are marked with red buttons:



You can plot the properties of each component and change its label (Accept / Reject) by clicking on the rectangular button above each component's scalp map:





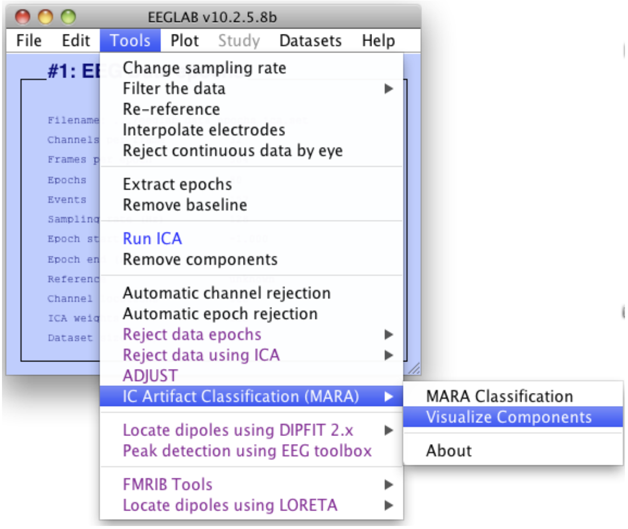


In the examples above, IC3 is a typical eye blink artifact (strong frontal activition, steep power spectrum), IC7 is a neural component (alpha peak around 10 Hz, scalp map indicates an occipital brain source), and IC31 is a typical muscle artifact (spatially localized acitivity, high power above 20 Hz).

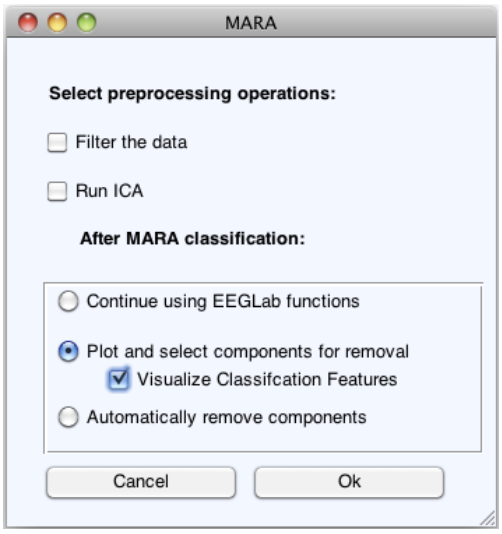
To identify artifacts by visual inspection, it is also helpful to inspect the components time courses, which you can access via Plot>Component activations (scroll).

4.6 Visualizing components with MARA

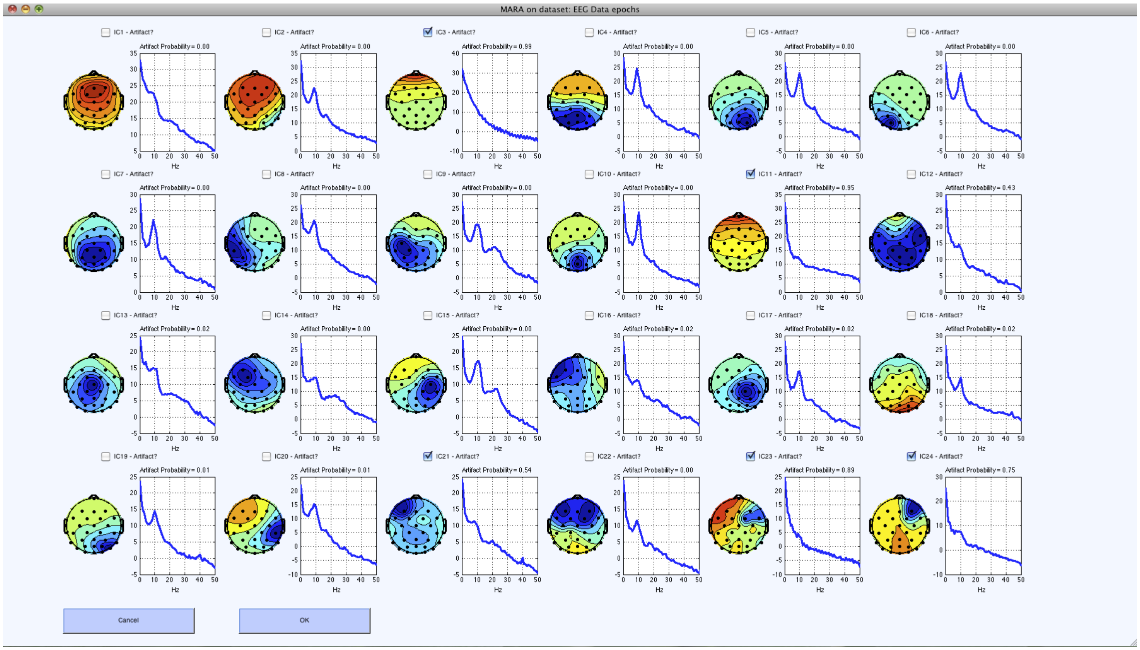
MARA provides a visualization function that allows for a quick change of the classifiers proposed ratings and gives an overview of each component's scalp map and power spectrum. After MARA computation, you can call it from the menu by Tools>IC Artifact Classification (MARA)> Visualize Components.

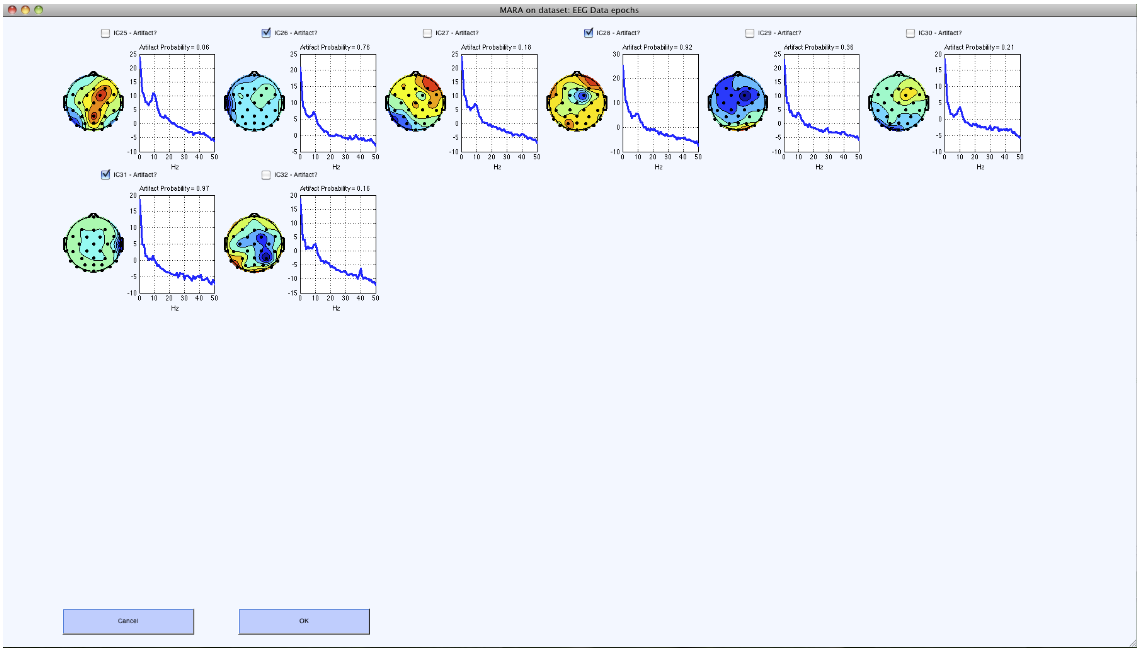


Alternatively, you can perform both MARA and its visualization by checking 'Plot and select components for removal' in the MARA main menu.



The visualization MARA provides shows each component scalp map, its spectrum, the current label of the component (artifact or not), as well as the probability of each component of being an artifact as computed by MARA. You can change the label of each component by checking or un-checking the checkbox above each component.

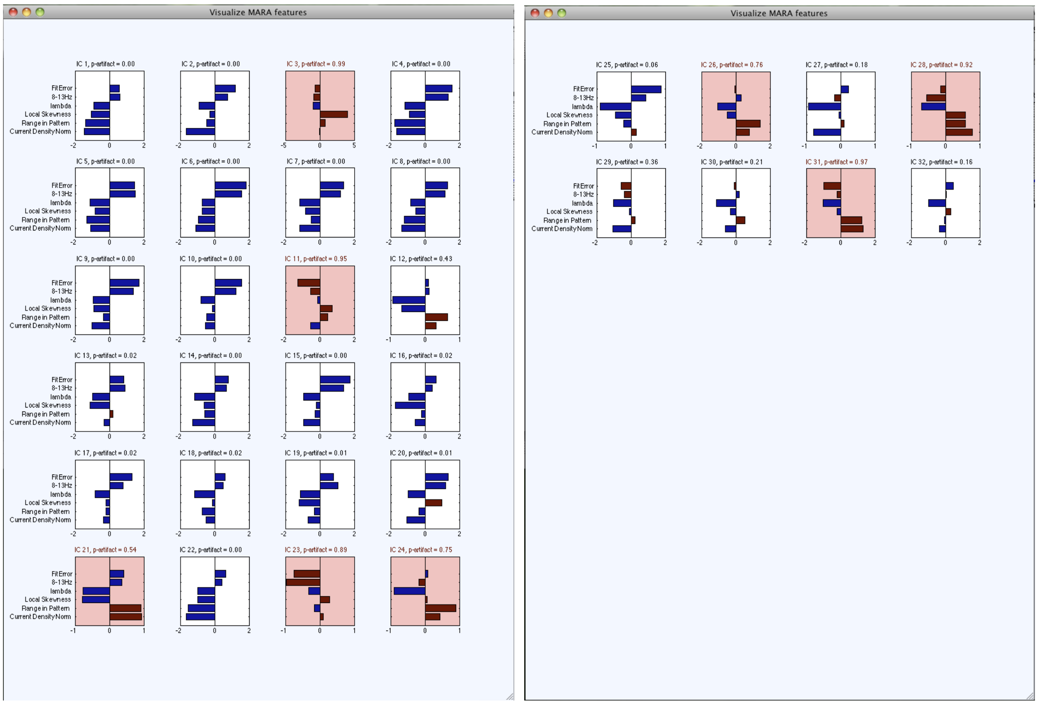




Mark all the ICs you wish to reject from the data and press 'OK'. You can now continue inspecting the ICs with EEGLAB functions as described in Section 4.5. To substract their activiy from your data, go to Tools>Remove components as outlined in the follwoing section.

MARA features

For information purposes, windows that display for each component the six features that MARA classification is based on will pop up:



Features that contribute towards a component to be marked as an artifact are plotted as red bars, features that indicate the component contains neuronal activity are marked in blue.

These features may be helpful in understanding MARA's decision. Recall the following from Section 2.2: A high *FitError* and a high *8-13Hz* feature indicate that the component contains an α peak and thus neuronal activity. A high λ value indicates a spectrum with high values in the 20-50 Hz range, which is a sign that the component contains muscle activity. A high *Local Skewness* value indicates outliers in the component's time series. A high *Range in Pattern* and *Current Density Norm* indicates a scalp map that is unusual for a neuronal component.

Lets consider the example components we visualized in Section 4.5.

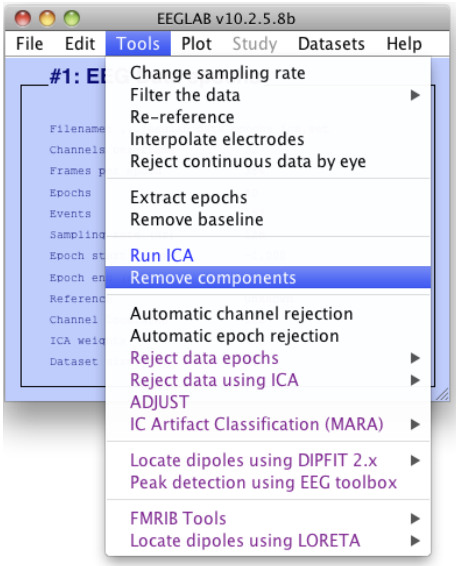
1. The eye blink artefact component IC3 is characterized by strong outliers in the time series (rare blinks) and contains no α peak. MARA labels IC3 as an artifact due to its low *FitError*, low *8-13Hz*, and high *Local Skewness*.

2. The neuronal component IC7 contains an α peak, no strong outliers in the time series and a smooth, dipole-like pattern. All features of MARA indicate it is a neuronal component.

3. The muscle artifact IC31 is recognized by MARA due to its lack of α peak (low *FitError*, low *8-13Hz*) and its very localized scalp map (high *Range in Pattern* and high *Current Density Norm*)

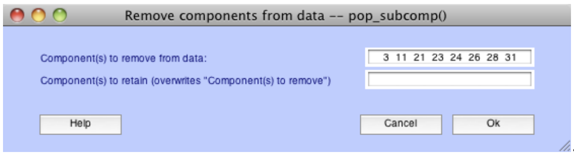
4.7 Removing rejected components

To remove the rejected components, go to Tools>Remove components



.

Press 'Ok' to remove the ICs from data.

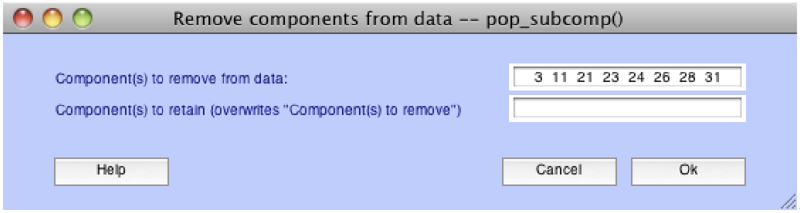


.

EEGLAB will ask for confirmation and a name to the new data set.

4.8 Automatic Mode of MARA

You can automatically remove components without inspection by clicking 'Automatically remove components' in the MARA main menu. In this case, there will be no visualization of the components, and you will be directly forwarded to EEGLAB's remove components menu.



.

EEGLAB will ask for a confirmation and a name of the new data set.

4.9 Command Line Usage

You can start MARA from the command line as follows:

```
[ALLEEG,EEG,CURRENTSET] = processMARA(ALLEEG,EEG,CURRENTSET)
```

For more information, please see its help function.

5 Contact and Support

If you have questions, suggestions or bug reports, please contact ~~irene.winkler@tu-berlin.de~~ irenedowding@web.de

References

[1] `http://www.gnu.org/licenses/gpl.txt`.

[2] A. Delorme and S. Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9 – 21, 2004.

[3] V. Fonov, A. Evans, R. McKinstry, C. Almli, and D. Collins. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47, Supplement 1(0):S102 –, 2009.

[4] V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, and D. L. Collins. Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54(1):313 – 327, 2011.

[5] M. S. Hämäläinen and R. Ilmoniemi. Interpreting magnetic fields of the brain: minimum-norm estimates. *Med. Biol. Eng. Comput.*, 32:35–42, 1994.

[6] S. Haufe, V. V. Nikulin, A. Ziehe, K.-R. Müller, and G. Nolte. Combining sparsity and rotational invariance in EEG/MEG source reconstruction. *NeuroImage*, 42:726–738, 2008.

[7] G. Nolte and G. Dassios. Analytic expansion of the EEG lead field for realistic volume conductors. *Physics in Medicine and Biology*, 50(16):3807, 2005.

[8] I. Winkler, S. Haufe, and M. Tangermann. Automatic classification of artifactual ICA-components for artifact removal in EEG signals. *Behavioral and Brain Functions*, 7:30, 2011.