BRAIN PRODUCTS
Solutions for neurophysiological research

Press Release

Support & Tips

Independent Component Analysis – demystified!

by Dr. Markus Plank, Scientific Consultant at Brain Products' Scientific Support

For some researchers, Independent Component Analysis (ICA) to a certain extent might still be equivalent with a black box, which magically alters the data and produces "cleaner" signals. In this article, I would like to take you by the hand and demystify the theoretical background, requirements and algorithms as well as the implementation of ICA in BrainVision Analyzer 2. Of course, we cannot address all details, but at the end of this article you will certainly have a deeper understanding of the method and hopefully feel more comfortable with integrating ICA into your data processing pipelines.

Before delving into the intricacies of the method, let's consider the following situation: A philharmonic orchestra gives a concert and you would like to record the cleanest signal possible with as little noise as possible – a challenging task since the concert hall is sold out. Wisely, you placed and securely mounted several microphones all over the concert hall, stage and rows of seats. Further, you safely assume that all the players on stage will be stationary throughout the concert (you don't expect them to stand up and walk around while playing), and each instrument group plays their own melody. After an exciting evening, each microphone will have captured a recording of the mixed original signals, in fact you will have as many signal mixtures as you have microphones. Due to the placement of the microphones in the music hall, the mixtures will differ slightly across microphones. Now your ultimate goal is to unmix the mixture and to extract, or reconstruct, the "pure signals". While the different melodies played by the instrument groups should of course be preserved, acoustic noise originating from the audience or – even worse – the nearby airport should be removed.

Importantly, this example can be transferred in a straightforward manner to EEG recordings. The time-domain channel activations (as represented by a time series of microvolt amplitudes) recorded from the EEG electrodes can be considered as a set of mixtures of brain signals which are supposedly generated by synchronization of neural compounds in cortex and subcortical regions, triggering far-field potentials. While the patches of neurons themselves are stationary and do not move, the activation patterns mix and merge based on the principles of volume conduction, propagate through all layers of cortex, skull, and tissue, and are ultimately present at any scalp site. Again, the ideal outcome of the analysis is to unmix this mixture, allowing for the identification of brain-based signals (which should be preserved) and non-brain signals resulting from (oculo-)motor activity, ECG or line noise (which should be removed).

In both examples, the goal is to extract the statistically "pure signals" from the recorded mixture in order to allow for a selection of signals to keep and signals to be discarded. Exactly this can be accomplished with Independent Component Analysis (ICA; Makeig et al. 1996). The technique has been recognized as a powerful tool for attenuating artifacts and analyzing statistically independent cortical processes in scalp and intracranial EEG recordings. Particularly in experimental setups where limited EEG data are available or when stereotypical artifacts such as blinks or muscular activity are contaminating the data (for example in patient groups, children or mobile EEG setups where subjects are moving freely) ICA may be superior to artifact rejection. While artifact rejection removes contaminated data portions, ICA allows for artifact attenuation, thereby preserving the original amount of samples, resulting in a higher signal-to-noise ratio for subsequent analysis steps.

ICA extracts pure signals from signal mixtures given that the following requirements are met (see Jung et al. 1998 for details):

1. Pure signals as extracted by ICA are characterized by their time-course (melodies, neural activation patterns) which is statistically independent of any other signal. These activations in fact are the independent components (ICs).

2. The generators of the pure signals (instrument groups, neural compounds) as well as the recording sites (microphones, electrodes) are stationary throughout the recording. Thereby the topographic projections of the components towards the recording sites are fixed.

3. The mixing is linear and propagation delays are negligible.

4. The probability distributions of the individual IC activations are not precisely Gaussian.

ICA does not put any further requirements on the data. In fact, it is completely agnostic of the nature of the signal, which is why ICA is generally referred to as **Blind Source Separation** algorithm (Hyvärinen and Oja 2000). Please keep in mind that components are purely statistical properties, so they do not map 1:1 onto physiological processes. In the context of EEG/MEG data, the extracted components can be examined based on their time-course (and topographic distribution), and components that represent noise, artifacts or other non-brain processes can be removed. Then, the corrected set of independent components is mixed or back-projected, which will result in a modified signal mixture at the electrodes where artifacts will have been attenuated.

Unmixing and back-projecting

In this section, the mathematical foundations of ICA are presented, which includes matrix operations. Generally, we start with the time-domain signal mixtures as recorded at the electrodes. The electrode data can be represented as a 2-D matrix, where the rows represent channels and columns represent samples. The values in the matrix are the recorded voltage amplitudes for each channel and sample. Let's call this matrix x. We can now generate an unmixing matrix W (ICA matrix) which, when multiplied with the data matrix x, transforms the mixed data x into IC activations a.

$$Wx = a$$

Each row of matrix a represents a component, and each column represents a sample. The activation of component a can be characterized as linear sum of weighted channel activations: The activation a_{11} of component F00 at time-point 1 is determined based on matrix multiplication of the first time-point $x_{1,1..n}$ across all electrodes with the weights $W_{1..n,1}$ at all electrodes (Figure 1).
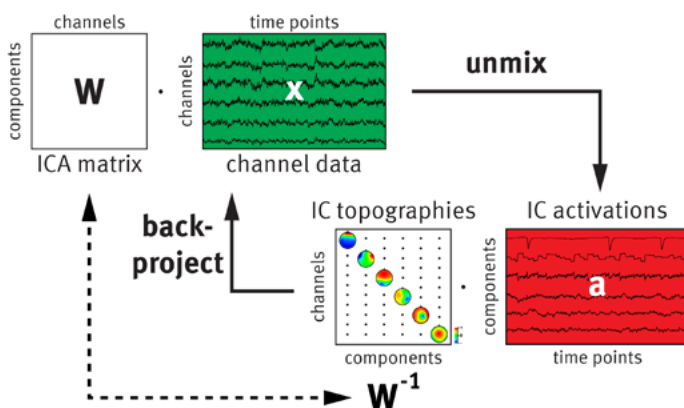

Figure 1: Schematic flowchart of ICA unmixing and back-projecting using six EEG channels, resulting in six independent components, each with a specific IC activation and topography (adapted with permission from Makeig and Onton 2008).

Although the effective number of statistically independent components contributing to the scalp EEG is unknown, from the matrix representation it should become obvious that you can always extract maximally as many components as there are electrodes (Makeig et al. 1997) since the mixed signals from N electrodes are decomposed into a linearly weighted sum of N components. For this reason ICA can be characterized as a complete decomposition technique. This implies that the unmixing process can be reversed by multiplying the IC activation matrix a by the inverse matrix W^{-1}. The columns of the inverse matrix W^{-1} contain the relative weights and polarities of the back-projections from each component to each electrode, which can be plotted as topographic map. For this reason, ICA is sometimes referred to as linear spatial filter since it allows for the evaluation, classification and selection of components based on their **activations** (as stored in the rows in the IC activation matrix a) and **topographies** (as stored in the columns of the inverse matrix W^{-1}).

The computational goal of ICA is to find an unmixing matrix W such that *maximal* temporal independence across all components is achieved – please be aware that complete mathematical independence can never be established for EEG data with finite length (Makeig and Onton 2008). Finding W in fact is the most time-consuming part in the computation, and while it might sound already quite close to magic, it's not. For those of you being interested in the conceptual basics, please refer to the section "How to determine independence and find an optimal unmixing matrix W?". The ICA algorithms available in BrainVision Analyzer can be found in the section "How to apply ICA in BrainVision Analyzer".

How to determine independence and find an optimal unmixing matrix W

Imagine the search for the optimal unmixing matrix W to take place in an abstract landscape of potential matrices W. This theoretical landscape has troughs and (local and global) peaks, corresponding to matrices W that result in low and high temporal independence of the components, respectively. The ICA algorithm now searches for the highest summit (or "global maximum") in this landscape of potential matrices W. While at the beginning W will not be optimal, with each step ICA will ascend and identify a more optimal unmixing matrix W. This stepwise optimization is generally referred to as stochastic gradient ascent. Optimization is limited to a certain number of steps (after which the search is terminated without success), but stops as soon as the summit in the landscape has been reached. In this case the optimal unmixing matrix W which maximizes temporal independence between components has been identified.

Most importantly, ICA completely neglects the time-course of the variables. Instead, statistical independence implies that the distributions, or probability density functions (pdfs) of two random variables y_1 and y_2 are completely unrelated to each other. Knowing the values of variable y_1 does not tell anything about the values of variable y_2, and vice versa:

$$p(y_1,y_2) = p_1(y_1)\, p_2(y_2)$$
i.e. joint pdf of y_1 and y_2 = mutual pdfs of y_1 and y_2

While the joint pdfs represent the combined probability density of both variables, the marginal pdfs represent the probability density of each variable alone. Two variables y_1 and y_2 are statistically independent if and only if the joint pdf can be generated by multiplication of the two marginal pdfs.

This definition can be extended for any number N of random variables. The joint pdf must then simply be a product of N terms. Mathematical proof can be found in Hyvärinen & Oja (2000). We can now extend the definition to the core property of independent random variables. Assuming two functions h_1 and h_2, the expectation E can be expressed as:

$$E[h_1(y_1)\, h_2(y_2)] = E[h_1(y_1)]\, E[h_2(y_2)]$$

This formula is crucial for independence since it states that two random variables are independent not only if the distributions of the actual variables are independent, but further that independence also applies no matter what function is applied to the variables, including all moments of the distribution.

This feature of independent variables to be independent irrespective of the applied function is the central distinction to uncorrelatedness, which is a weaker form of independence. Two random variables y_1 and y_2 are said to be uncorrelated, if their **covariance** (i.e., the **second moment** of the joint pdf) is zero:

$$E[y_1,y_2] - E[y_1] \, E[y_2] = 0$$

If two variables are independent, they are also uncorrelated. However, if two variables are uncorrelated, they do not necessarily have to be independent. While correlation only refers to the **second moment** (covariance) of the joint pdf, independence involves all moments. For this reason, independence places stronger constraints on the data than uncorrelatedness (Hyvärinen and Oja 2000). While computing **all moments** of a distribution is of course not possible in real life, the computation of **kurtosis**, which is based on the computation of the **fourth moment**, has been shown to be sufficient.

$$\text{kurtosis } (y) = E[y^4] - 3$$

Kurtosis is a classical measure of non-Gaussianity of a distribution. The less Gaussian the distribution, the higher the absolute value of kurtosis. While a Gaussian has a kurtosis of zero, *platykurtic* (sub-Gaussian) distributions with broad and wide peaks have negative kurtosis, and *leptokurtic* (super-Gaussian) distributions with acute and narrow peaks have positive kurtosis.

Since signal mixtures usually neither look like any of the pure signals but instead have a more Gaussian distribution (Central Limit Theorem), ICA examines signals with respect to their kurtosis, or non-Gaussianity, where maximum positive or negative kurtosis reflects maximal independence. This implies that ICA can only successfully unmix non-Gaussian signals (see the requirements for ICA). The good thing is that pure signals such as music, EEG voltage amplitudes or speech indeed are highly super-Gaussian. By contrast, highly sub-Gaussian distributions generally reflect AC (alternating current) or DC (direct current) noise, e.g., induced by screen currents, electrical machinery, lighting fixtures, or loose electrode contacts (Delorme et al. 2007).

So coming back to our landscape of potential unmixing matrices W, the criterion which is maximized actually is the kurtosis or another associated measure of non-Gaussianity of all components. Which criterion in detail is evaluated varies across ICA algorithms (the ICA algorithms implemented in BrainVision Analyzer are characterized in more detail in the next paragraph).

How to apply ICA in BrainVision Analyzer

In the Support Tip on Ocular Correction ICA I already listed optimal data preprocessing steps and recommendations on how many channels and samples should be used for valid ICA decomposition. These of course again apply here. Before you apply the transformation, electrode data should be properly filtered and checked for non-stereotypical artifacts. These data points should be marked as „Bad Intervals", as data portions marked as "Bad Intervals" are disregarded by ICA. Data selected for ICA should include the relevant information content to be decomposed (for example, enough stereotypical EMG, EOG, ECG artifacts). Therefore, I suggest to use longer, representative data intervals or even the whole data (provided it can be afforded with respect to memory constraints).

Also, in order to visualize the **IC topographies** (representing the columns of the inverse matrix W^{-1}), electrode positions should be present. When using channels with unknown positions (Radius, Theta, Phi = [0,0,0]), the IC topographies will remain empty, and evaluation can solely be accomplished based on the **IC activations** (the rows of matrix a).

Once your data fulfills these requirements, you can accomplish the sequence of ICA unmixing, IC examination and selection as well as back-projection by means of the transformations *ICA* and *Inverse ICA* (Figure 2). The steps within each transformation will be described in more detail below.

Figure 2: : Steps of the transformations ICA and Inverse ICA

ICA: Unmixing your data

The transformation *ICA* accomplishes the computational part of unmixing the channel data into temporally maximally independent components. It can be found under T*ransformations › Frequency and Component Analysis › ICA*. In order to apply the transformation to multiple datasets in template mode (e.g., overnight or during a tea break), the automatic mode is recommended. In this case, the transformation is not stalled with an interactive view (requiring user input), but instead accomplishes the computational part and closes the newly created ICA node. This preserves memory capacities and speeds up processing of subsequent datasets.

In the dialog you can save the unmixing matrix W and the inverse matrix W^{-1} to text files, for example in order to apply these matrices to other nodes in your history tree (using the transformation *Linear Derivation*). Based on the remaining options set in the transformation dialog, Analyzer accomplishes the following computational steps:

(1) PCA sphering

As stated above, ICA completely neglects the time-course of the data, but instead only examines the *distributions* of the channel data. In order to optimize the distribution characteristics of the data that is fed into ICA, the data is de-meaned (i.e., for each channel, the mean amplitude is computed and subtracted from each and every sample). By doing that, the data distributions of all channels have an overlapping mean of zero, and joint measures can be computed in a more efficient way.

After de-meaning, **PCA sphering/whitening** is applied (Johnson 2003), where linearly dependent (multicollinear) channels are detected, the data rank is reduced, and the maximal number of ICs to be extracted from the data is determined. Sphering is generally considered a necessary precondition to ICA since it allows for better and faster convergence of the iterative search for the optimal unmixing matrix W.

You might ask what this actually means and why this is important. Let's get back to the statement that ICA extracts maximally as many components as there are real channels, which is referred to as the **data rank**. If you recorded EEG data from only 20 channels and interpolated 100 channels (e.g., to virtually extend coverage and create a smoother mapping view), the 100 interpolated channels do not carry distinct information – they are all linearly dependent on the original 20 channels. The data rank is still 20 and not 120. As a result, only 20 independent components can be validly decomposed out of the data. Of course this is an extreme example, but it occurs when-ever feeding original *plus* linearly interpolated channels into ICA, for example upper and lower VEOG channels *plus* the bipolar VEOG channel (generated by subtracting upper and lower VEOG data). The same happens when using the complete set of average-referenced channels for ICA decomposition, which is why for average-referenced data it is recommended to exclude at least one channel from ICA in order to prevent multicollinearity between channels and, followingly, the overestimation of the number of components to be extracted from the data.

PCA sphering is smart. Based on the covariance of the data matrix x, the transformation detects linear dependencies and raises your attention in case the data rank and, accordingly, the number of components should be reduced. You can safely accept the data rank reduction, since the component time-courses and topographies might in fact be invalid if the rank reduction is not applied. You should, however, consider and examine what might have caused the data rank reduction in the first place and – if possible – adjust the channel selection for ICA accordingly.

Two PCA sphering methods are available in the ICA dialog: **Classic sphering** assumes the covariance of the data matrix x to be solely based on the linearly mixed channel data (Bell and Sejnowski 1995; Oja 1985). By contrast, **probabilistic sphering**

assumes the covariance matrix to consist of the linearly mixed channel data plus some added random Gaussian noise (Tipping and Bishop 1999). In case you decide to stay at the component level and do not intend a back-projection onto the original channels, then probabilistic sphering might be recommendable since the resulting component activations will be statistically de-noised. However, if you intend a back-projection onto the channels, classical sphering is recommended since probabilistic sphering would also back-project the added Gaussian noise, therefore resulting in noisier data compared to the pre-ICA data – which is exactly the opposite of the desired outcome.

(2) ICA decomposition

In this step the sphered data is fed into ICA. The maximum number of components to be unmixed is always equal to the number of channels (this is the default option). However, you can specify the number of components manually, or limit it to components whose Eigenvalues during PCA sphering exceeded a certain criterion (e.g., Eigenvalue › 0.001). However, in case of rank reductions due to linear dependencies between channels, Analyzer suggests to further reduce the number of components to the actual data rank, which should be accepted.

The ICA algorithms available to you are **Infomax ICA** (Bell and Sejnowski 1995; Lee et al. 1999) and **FastICA** (Hyvärinen and Oja 2000), which are the most common ones being applied to EEG data (a detailed comparison of almost all available ICA algorithms can be found in Delorme et al. 2012). While both of them employ iterative fitting procedures using systems of nonlinear functions in order to maximize independence, they differ with respect to the parameters used for evaluating the unmixing matrix W.

- **Infomax ICA** identifies an optimal unmixing matrix W by naturalistic gradient ascent. Since the naturalistic gradient ascent is dynamically searching for an optimal unmixing matrix W in an unsupervised fashion where all extracted components are evaluated concurrently, each ICA run will produce slightly different sets of components where the component sequence generally differs across different runs of ICA. The search for the optimal unmixing matrix W is based on maximizing negentropy, i.e., the uncertainty in a random variable. The concept of negentropy is closely related to non-Gaussianity and kurtosis, since all of them are maximized as the joint distribution of two or more variables can be fully predicted from the mutual distributions of each variable (see section „How to determine independence and find an optimal unmixing matrix W?" for details). **Restricted Infomax ICA** is able to unmix the data into components with positive kurtosis, which is typical for pure signals such as music, speech and EEG data. **Extended Infomax ICA** can unmix components with positive and negative kurtosis – the latter is typical for AC (alternating current) or DC (direct current) noise as

induced by screen currents, electrical machinery, lighting fixtures, or loose electrode contacts (Delorme et al. 2007). Extended Infomax ICA is therefore recommended when intending to unmix both components reflecting biological signals as well as components reflecting line noise. The option **Biased** reflects a logistic function that was introduced by Makeig and colleagues (1997), which unmixes the data in favor of components with positive kurtosis.

- **FastICA** can be considered as a computationally optimized version of Infomax ICA, which is deterministic and always produces the same IC sequence across different runs (Hyvärinen and Oja 2000). **Restricted FastICA** finds an optimal unmixing matrix **W** by means of a fixed-point algorithm maximizing negentropy, **extended FastICA** is based on maximum likelihood estimation (Hyvärinen et al. 2001; Koldovsky et al. 2006). In contrast to the naturalistic gradient method of Infomax ICA, the FastICA algorithms converge much faster. However, the increased speed of FastICA procedures comes with a price: FastICA has issues with "weak" components, i.e., components whose distributions are close to Gaussianity or close to each other (Chevalier et al. 2004) – and in fact EEG data often contains such weak components! In this case, FastICA takes much longer to converge or might even produce incorrect decompositions. In this case, the unmixed components might not be temporally maximally independent. This is the major drawback of the FastICA procedures.

When you apply the transformation as recommended in automatic mode, the newly created ICA node contains the decomposed **IC activations** – labeled by "F" followed by a number (channels that have not been selected for ICA will be displayed below the IC activations) – as well as the **IC topographies** which are stored in the node properties (together with the un-mixing and inverse weights).

Before proceeding, please always examine the Operation Infos of the ICA node by right-clicking the node in the history explorer and selecting Operation Infos… from the context menu. I advise checking whether ICA converged in the given number of steps. Only in this case the unmixing has been successful, and the displayed activations truly reflect temporally maximally

independent components, allowing you to proceed with the component evaluation.

Inverse ICA:
Component examination, selection and back-projection

After examining the Operation Infos of the ICA node, the components can be examined, selected and back-projected onto the channels using the transformation *Inverse ICA*, which can be found in *Transformations › Frequency and Component Analysis*. If you run this transformation automatically, all components are back-projected onto the channels and the original data is simply restored, which is equivalent to a complete back-projection. However, this is not desired when applying ICA for the purpose of artifact attenuation. In this case, it is mandatory to apply the *Inverse ICA* in semiautomatic mode since only then you get access to the flexible and powerful visualization tools of the interactive view, allowing you to carefully evaluate and ultimately select or reject components based on their activations and topographies.

After all the automatic computation steps this is where your input is desired! As can be seen in Figure 3, the interactive view displays the IC activations (labeled by "F" followed by a number) in the main view. The table in the upper right corner lists the inverse weights of all components towards the data channels (each row is one component and each column is a channel). The color of the first column in each row indicates whether or not this component has been selected for back-projection. Initially, all components are colored green. By double-clicking a row in the table you can change the color to red, indicating that
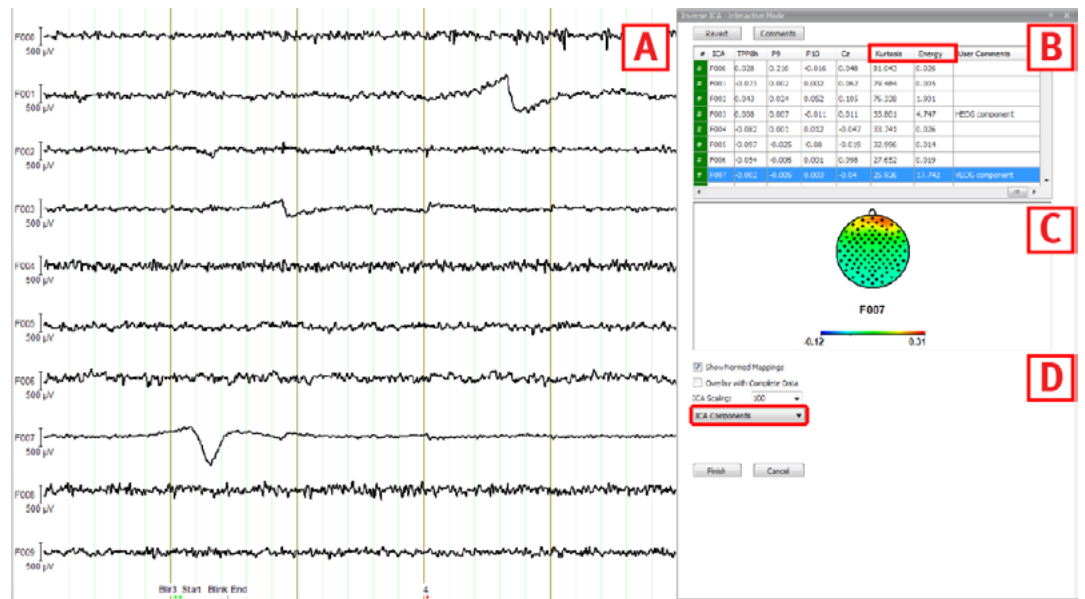

Figure 3: [A] IC activations (scaled by a factor of 100) of components F000 – F009. [B] Table containing the inverse weights, kurtosis and energy values of all components as well as a color code on whether the component will be selected for back-projection (green) or not (red). In the current example, components have been sorted based on their kurtosis. [C] Topography of F007 highlighted in the criterion table. The IC topography represents the normed inverse weights of the currently selected component towards all channels (unit-less weights, units can be neglected). [D] Display settings. In this part of the interactive view, you can adjust the information displayed in the main window, for example the component scaling with respect to the channel amplitudes.

the component reflects artifactual or non-brain processes and should not be back-projected onto the channels.

I generally suggest dedicating some time to the careful examination of all components one by one. You can, for example, sort the table according to the inverse weights. Also, quantitative measures that you might want to use for selecting components can be found at the very right end of the table: While **kurtosis** reflects the non-Gaussianity of the component activations (with smaller values indicating more Gaussian distributions), **energy** represents the conjunctive variance of the component activations and topographies. Components with 'checkerboard'-like topographies and/or shallow activations will have a lower energy compared to components with well-defined, consistent topographies and/or activations.

The mapping view below the table visualizes the IC topography of the component currently selected in the table. The map represents an interpolated representation of the inverse weights towards the channels (from matrix W^{-1}). By using the options available below the mapping view, you can scale the component activations with respect to the channel amplitudes (*ICA Scaling*), or use the drop-down list to switch between the component activations (*ICA Components*), the cleaned/corrected channel data after excluding certain components from the back-projection (*Correction*), or the simultaneous display of all IC topographies (*Topographies*). The elements are displayed in Figure 3.

If you select *Correction* from the drop-down list, you can display the channel data after excluding artifactual components from back-projection (Figure 4). When combining this selection with the option *Overlay with Original Data*, you can compare the data

before (red graphs) and after (black graphs) removing the currently selected artifactual components in the main window. This option is quite useful since you can observe the immediate effects of your component selection.

As you can see, artifactual processes in your data can be properly attenuated based on careful, conscious, and informed decisions about the "artifactual nature" of a component using its kurtosis and energy and evaluating its activation and topography. "Artifactual" components ideally reflect processes purely based on non-brain processes while at the same time containing absolutely no brain-based EEG signals. As soon as you are satisfied with your component selection, you can click *Finish* in order for the changes to take effect. Now only the components marked green are back-projected onto the channels, while the inverse weights of the components marked red are set to zero. In other words, the IC activation matrix a is multiplied by the modified inverse matrix W^{-1}', and a new data matrix x' is generated which does not contain any of the components that you excluded from back-projection.

Taken together, the transformations *ICA* and *Inverse ICA* as implemented in BrainVision Analyzer 2 are easy to use tools to diminish the effects of artifactual processes in your EEG recordings based on data-driven methods. In contrast to the previously presented transformation *Ocular Correction ICA*, you have the full freedom to further process and examine your data at the component level, for example by creating subnodes containing frequency- or time-frequency-domain representations of component dynamics, which may help you to determine which components to keep and which ones to remove from your data.
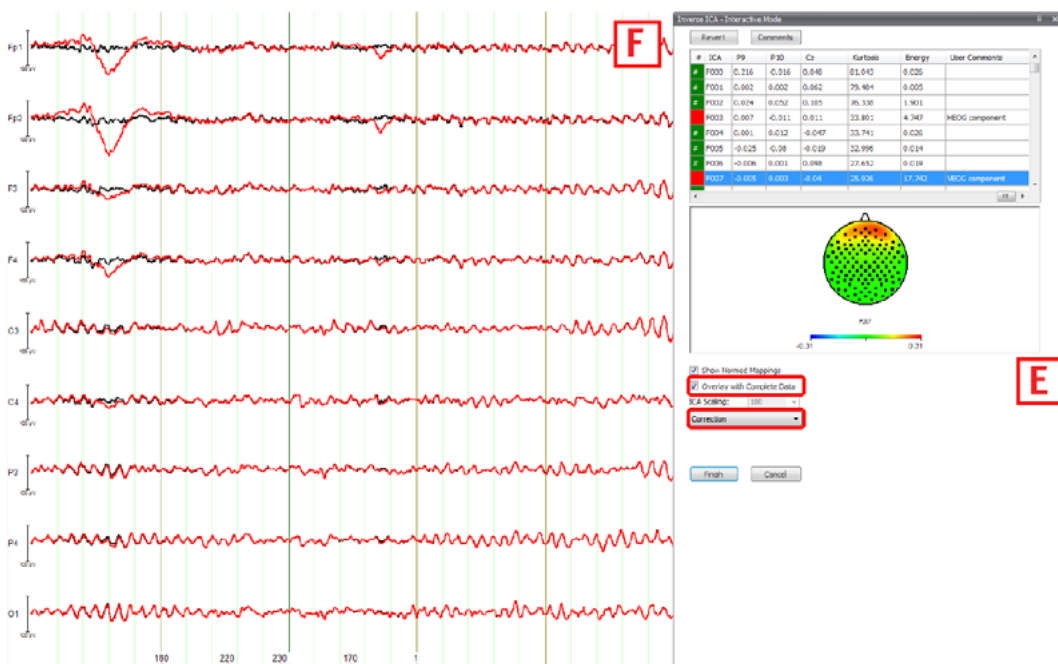

Figure 4: [E] Selecting "Correction" from the drop-down list displays the channel data in the main window where the artifactual components (which you marked red in the table) have been excluded from back-projection. Additionally activating the check box "Overlay with Original Data" allows for a comparison of the data before (red) and after (black) excluding the artifactual components [F].

In the end I hope that this comprehensive introduction into ICA elicited your interest – maybe even some of my personal enthusiasm for ICA has passed onto you – and demystified the technique. While artifact correction of signals (such as speech, music or EEG) using ICA might appear to be close to magic, it is a purely statistical method which is based on certain assumptions and methodological principles, delivering a representation of temporally maximally independent component activations with temporally stable topographies.

For any further questions regarding the transformation, please contact us at support@brainproducts.com.

References

- Bell AJ, and Sejnowski TJ. *An information-maximization approach to blind separation and blind deconvolution. Neural Comput 7: 1129-1159, 1995.*

- Chevalier P, Albera L, Comon P, and Ferreol A. *Comparative performance analysis of eight blind source separation methods on radiocommunications signals. In: International Joint Conference on Neural Networks. Budapest, Hungary: 2004.*

- Delorme A, Palmer J, Onton J, Oostenveld R, and Makeig S. *Independent EEG sources are dipolar. PLoS One 7: e30135, 2012.*

- Delorme A, Sejnowski T, and Makeig S. *Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. Neuroimage 34: 1443-1449, 2007.*

- Hyvärinen A, Karhunen J, and Oja E. *Indepenent Component Analysis. New York: Wiley & Sons, 2001.*

- Hyvärinen A, and Oja E. *Independent component analysis: Algorithms and applications. Neural Networks 13: 411-430, 2000.*

- Johnson D. *Colored Gaussian Noise* http://cnx.org/content/m11260/latest/

- Jung TP, Humphries C, Lee TW, Makeig S, McKeown MJ, Iragui V, and Sejnowski TJ. *Extended ICA removes artifacts from electroencephalographic recordings. Adv Neur In 10: 894-900, 1998.*

- Koldovsky Z, Tichavsky P, and Oja E. *Efficient variant of algorithm FastICA for independent component analysis attaining the Cramer-Rao lower bound. IEEE Trans Neural Netw 17: 1265-1277, 2006.*

- Lee TW, Girolami M, and Sejnowski TJ. *Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural Comput 11: 417-441, 1999.*

- Makeig S, Bell AJ, Jung TP, and Sejnowski TJ. *Independent component analysis of electroencephalographic data. Advances in Neural Information Processing Systems 8: 145-151, 1996.*

- Makeig S, Jung TP, Bell AJ, Ghahremani D, and Sejnowski TJ. *Blind separation of auditory event-related brain responses into independent components. Proc Natl Acad Sci U S A 94: 10979-10984, 1997.*

- Makeig S, and Onton J. *ERP features and EEG dynamics: An ICA perspective. In: Oxford Handbook of Event Related Potential Components, edited by Luck S, and Kappenmann E. New York: Oxford University Press, 2008.*

- Oja E. *On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. Journal of Mathematical Analysis and Applications 106: 69-84, 1985.*

- Tipping ME, and Bishop CM. *Mixtures of probabilistic principal component analyzers. Neural Comput 11: 443-482, 1999.*