PCA/ICA™

User Manual

Compumedics Neuroscan
7850 Paseo Del Norte
El Paso, TX 79912
USA

Voice :915-845-5600
Fax: 915-845-2965
e-mail: sales@neuro.com
techsup@neuro.com
web site: www.neuro.com

Document number 7235, Revision B

For Technical Support.....

If you have any questions or problems, please contact Technical Support through any of the following routes.

If you live outside the USA or Canada, and purchased your system through one of our international distributors, please contact the **distributor** first, especially if your system is under warranty.

In all other cases, please use **techsup@neuro.com**, or see the other Support options on our web site (*http://www.neuro.com/neuroscan/support.htm*).

Or, if you live in the USA or Canada, please call **1-800 474-7875**. International callers should use **915-845-5600, ext 223**.

For Sales related questions, please contact your local distributor, or contact us at **sales@neuro.com**.

Compumedics Neuroscan
7850 Paseo Del Norte
El Paso, TX 79912
Telephone: 915-845-5600
Fax: 915-845-2965
E-mail: techsup@neuro.com
Web: www.neuro.com

Table of Contents

Document number 7235, Revision B

Installing PCA/ICA™

All of the necessary files needed to register and execute the PCA/ICA program are contained on the installation CD. The program is part of Tool Box 2003. To access the Tool Box, you must have a paid maintenance contract, or it may be purchased separately ($1000 for the first license and $250 for additional licenses; contact sales@neuro.com). You will need to reprogram your dongle (software lock) in order to access the program. Instructions for doing this are found on our web site at http://www.neuro.com/licreq.htm.

Introduction to PCA/ICA

Principle Component Analysis (PCA) and Independent Component Analysis (ICA) are statistical techniques akin to factor analysis that are used to (1) to reduce the number of variables and (2) to detect structure in the relationships among variables.

PCA generates patterns and loadings that are orthogonal to each other. After the first factor is extracted (by fitting a regression line to a scatter plot), the second factor is extracted from the remaining variability, and so on until there is essentially no variance left. The resulting components are orthogonal to, or uncorrelated with each other (first order decorrelation). It has been argued that PCA may not be the most appropriate method for use with physiological data.

ICA generates patterns and loadings using a stricter criteria for statistical independence (requires that all second order and higher correlations are zero). The generality of ICA lies in the simple principle that different physical processes tend to generate statistically independent signals. Given that scalp-recorded EEG is the summation of signals from multiple sources, ICA computes individual signals that are statistically independent, and which are therefore likely to have been generated by different physiological processes. ICA has been asserted by some to be the preferred method for use with physiological signals. Some advantages with ICA and a further discussion of the differences between ICA and PCA are found in the *Using Filtering in PCA/ICA* section below.

It is beyond the intent of this manual to present a thorough discussion of PCA and ICA analyses, and the justifications, rationales, and advantages of each. Instead, we urge you to consult the increasing number of original research articles in which PCA and especially ICA are the subject.

Operating PCA/ICA

Online and offline. The PCA/ICA program can be used during online acquisition in ACQUIRE and in offline analysis in EDIT. The functioning is essentially the same, although there is one restriction during the online calculations. The Filter option is not available during acquisition.

Document number 7235, Revision B

Load 3DD electrode positions. Once you have the PCA/ICA program installed, you will see a new icon on the Toolbar  . The PCA/ICA program requires 3D electrode positions. It is always preferable to use the actual electrode position data. If you have the 3DD file from 3DSpaceDx that was measured from this subject, simply retrieve the data file, and right click between the electrode displays. Select the Load 3D Electrode positions option.

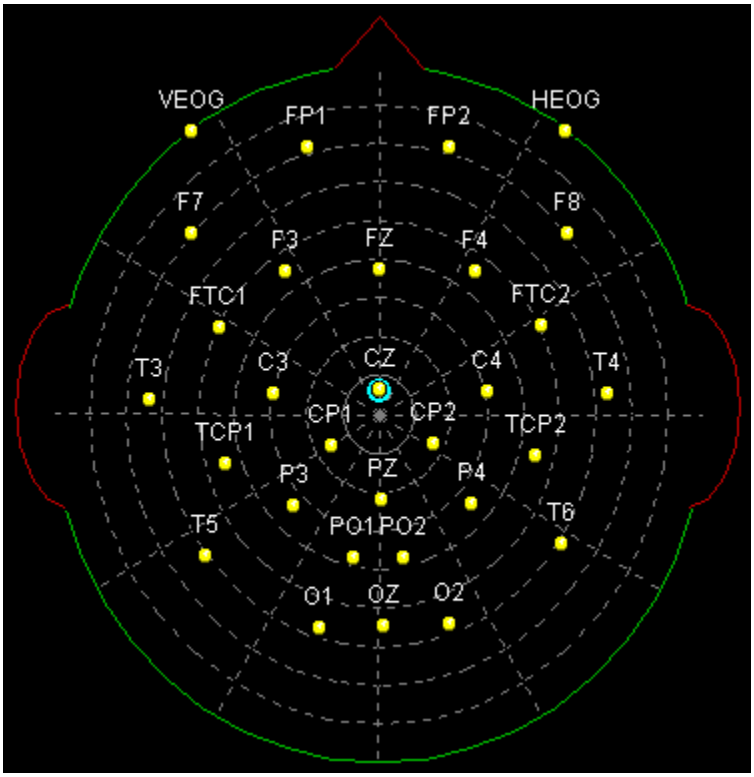
This will display the Open File utility, from which you may select the appropriate 3DD file. Now when you click the PCA/ICA icon, the program will open immediately. If the 3DD electrode positions have not been merged into the data file, you will see the following message after you retrieve your data file and click the  icon.

Click OK, and the "radar" display will appear, with no electrode positions. Click the 

button to place the electrodes automatically according to the 10-20 system. This step presupposes that the electrode labels you have used are standard 10-20 system labels. If not, you will need to relabel them accordingly. *You must have either the 3DD file or standard electrode labels to proceed.*

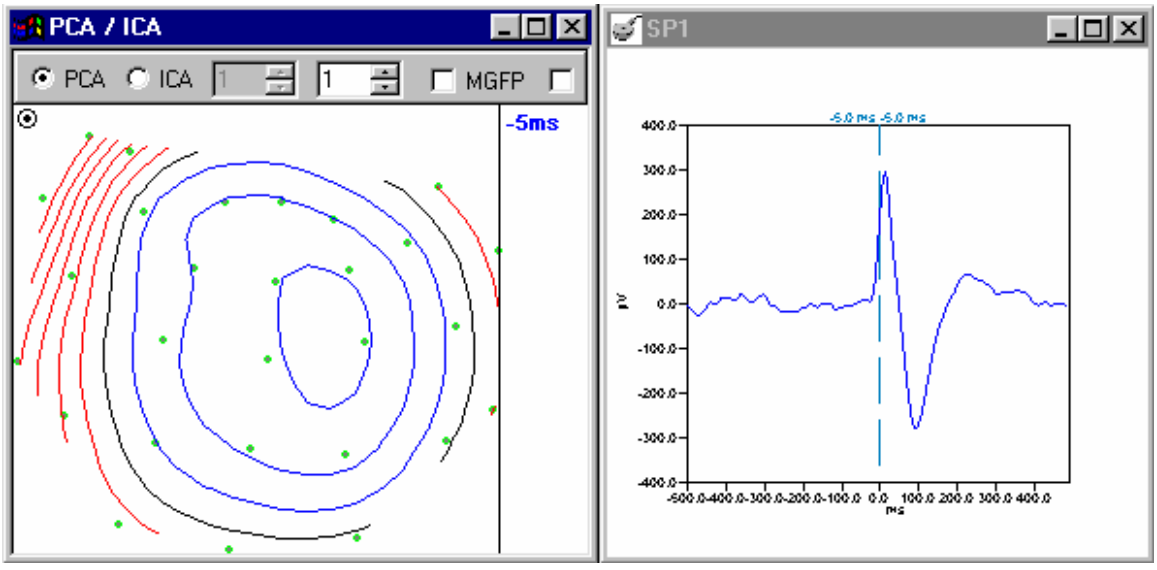
In the steps below, we are using the EpiSpike.avg demo file found in the \ScanData\Demo Files\EEG Spikes folder. This is a single epileptic spike (filtered).

Document number 7235, Revision B



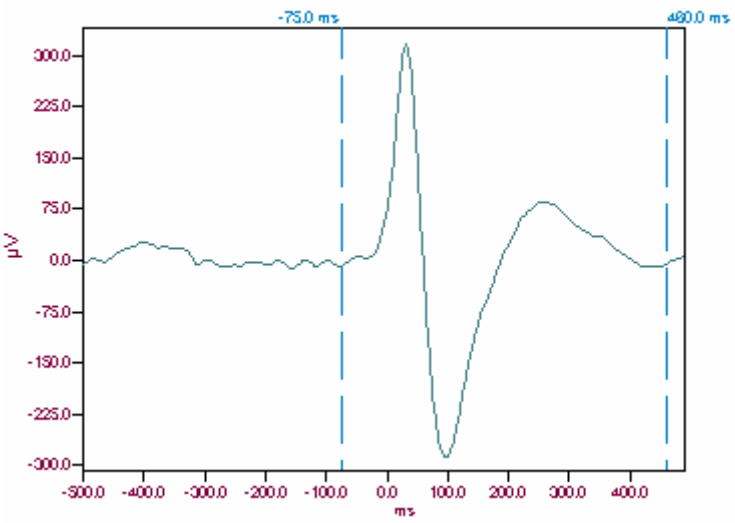
After selecting Match Labels, click OK. If desired, click Adjust Positions to position the displays, then click OK. Now enlarge an electrode display, and click the PCA/ICA icon. You will see the PCA/ICA display appear.

Initially, you will see the PCA/ICA display with nothing in it. Select an electrode display and enlarge it to mid-size or full-size. Then you will see the following. In the PCA/ICA window, the column on the left with the contours is referred to as the *patterns* column. On the right is the *loadings* column.

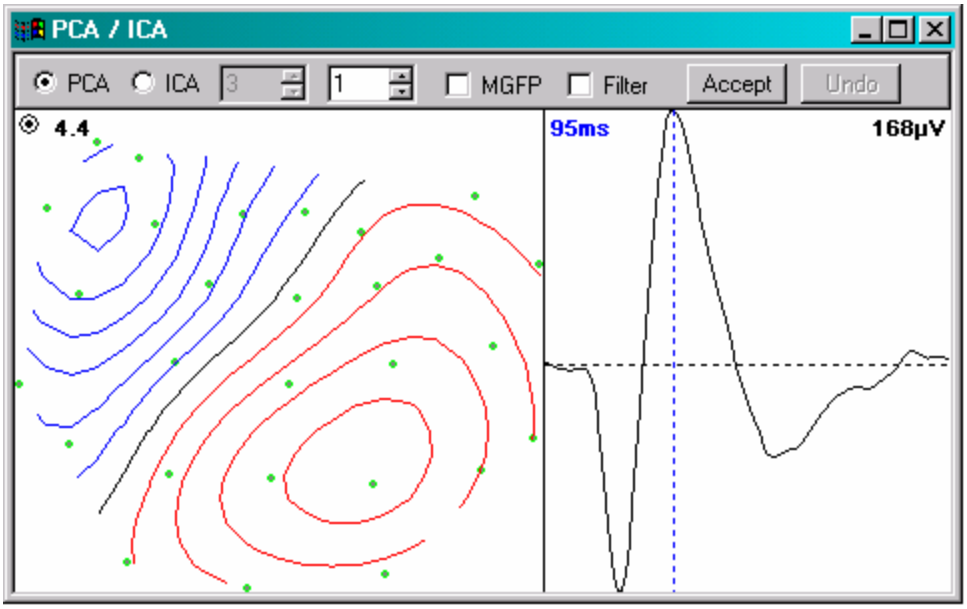


Document number 7235, Revision B

Use the mouse to position the vertical cursors in the waveform display to select an interval of interest. You can also position the cursors using the arrow keys from the keyboard. The left and right arrows alone will move the left cursor one data point at a time. Using the Shift+arrow key combination will move the right cursor. The Ctrl+arrow combination will move both cursors together, maintaining the width between them.

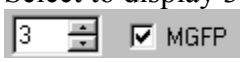


You will then see the first loading in the PCA/ICA display. The display appears inverted in this example. In fact, the valence is arbitrary and not relevant. The PCA/ICA program takes whatever loading is largest, and makes it positive for the display. The pattern contours are correct regardless of whether the loadings are positive or negative.

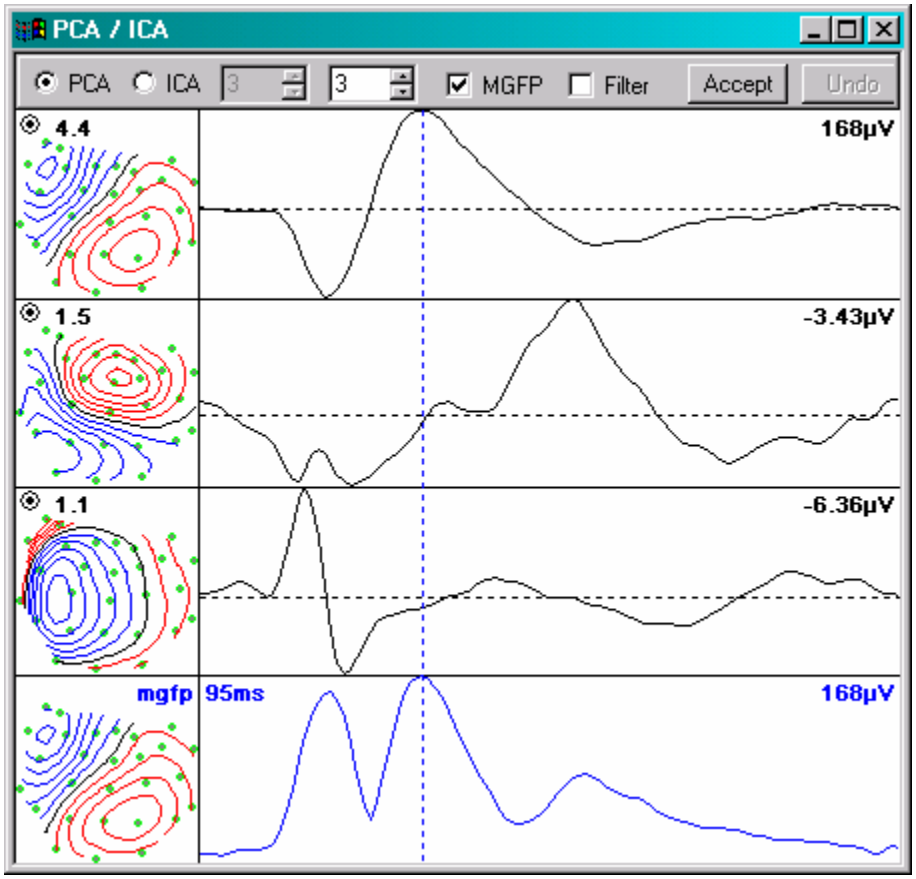


Document number 7235, Revision B

Select to display 3 loadings, and enable MGFP (Mean Global Field Power)



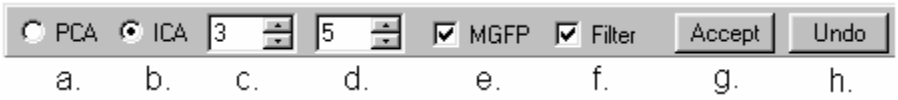
to see a display similar to the following.



The "3" in the field to the left of the MGFP box determines how many loadings are displayed. In fact, up to 10 patterns are available for display (rarely are more than 5-7 patterns significant). *There cannot be more patterns than the number of electrodes or time points.*

Now let's look at the display in more detail.

Option selection fields. The following options are accessed from the command bar at the top of the display.

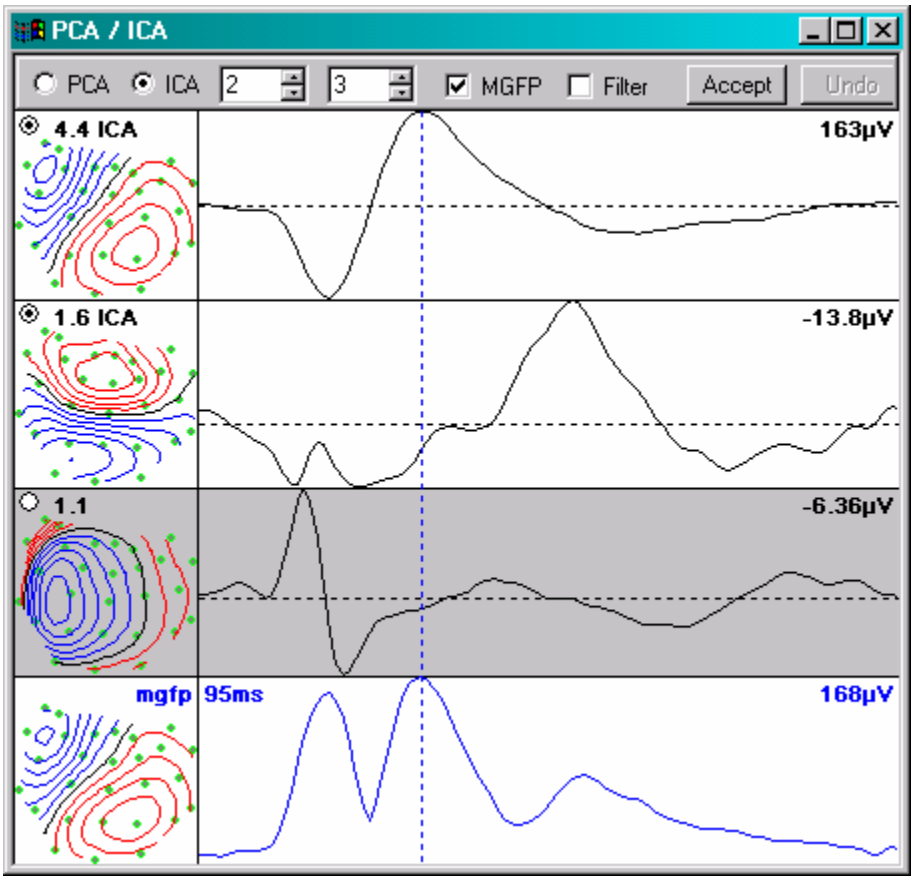


a. Click this field to display the patterns and loadings from the PCA analyses.

b. Click this field to display the patterns and loadings from the ICA analyses.

Document number 7235, Revision B

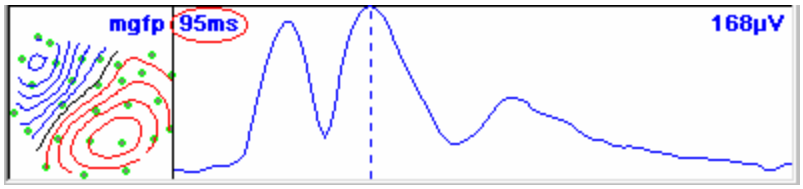
c. This field is active only when you select ICA, and it is used to select the number of ICA loadings to be computed and displayed. In the example below, note that 3 loadings are displayed, and that 2 of them are ICA components. The 3rd (and all additional ones not displayed) is a PCA component. The ICA loadings will always be indicated by "ICA" in the pattern display area.



d. This field is used to select the number of loadings to be displayed (all PCA loadings are calculated, and up to the first 10 can be displayed). Loadings beyond the 10th one are assumed to be statistical noise, and are disregarded.

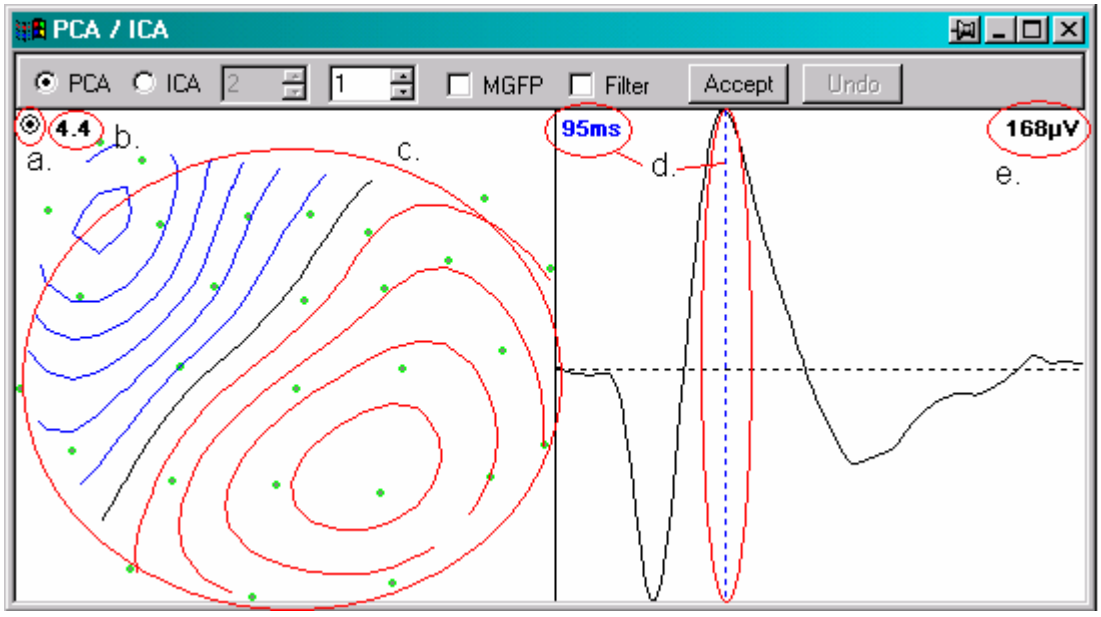
e. This field is used to display the Mean Global Field Power waveform at the bottom of the display. This is a composite waveform based on all EEG channels, and is useful for determining where the power peaks occur. The cursor position (the dashed vertical line) is shown in ms's (to the right of the MGFP label). Use the left and right arrow keys to move the cursor from the keyboard (hold either key down to see the contour pattern change throughout the selected interval).

Document number 7235, Revision B



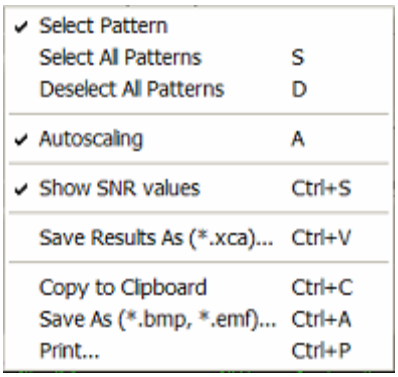
f., **g**., and **h**. The Filter toggle and Accept button are used in conjunction with *selecting* components to remove/retain them in the selected interval in the waveforms. Click the Undo to undo the filtering. Use of these features is described in more detail in the *Using Filtering in PCA/ICA* section below.

Display features. The following information is displayed in the patterns and loadings fields.



a. The radio buttons are used to select or deselect the corresponding pattern/loading. These and other options can also accessed by clicking the right mouse button within the PCA/ICA display.

Document number 7235, Revision B



You may select/deselect each pattern, or you can select or deselect all patterns. (With PCA, if you deselect a component, all trailing components will be deselected automatically). Note that if you Deselect All Patterns, all 10 patterns will be deselected. If you have, for example, 5 patterns displayed, and you manually deselect them, the remaining 5 patterns may still be selected (and you will see activity in the Filtered MGFP display). If all 10 patterns are deselected, you will see nothing in the Filtered MGFP display. It is a good idea to display all 10 components to be sure you are including only the ones you want (especially when using ICA).

The remaining options on the list are as follows. You have the option of displaying the SNR values (default) or the variance percentages in the patterns fileds. The results can be saved to a *.xca file. This is a text file that contains the number of electrodes, the numbers of samples in the selected interval, the XYZ coordinates of the electrode positions, the number of selected patterns, the number of weights, the number of normalized patterns and the number of normalized loadings. The display can be copied to the Windows Clipboard, saved as a BMP or EMF file, or Printed.

b. This number is the SNR value (signal-to-noise ratio). Generally, patterns with SNR's greater than 1.0 are the potentially valid ones. Deselect *Show SNR values* to see the percent of variance explained.

c. The contour pattern shows the distribution of the loading. Blue contour lines show the negative values distribution; red contours show the positive values distribution. Each loading will have it's own distribution. The contour line distance is autoscaled. You can change the number of contour lines using the Up and Down arrows on the keyboard, or with the mouse wheel.

d. The vertical cursor is used in association with the MGFP display, and may be "grabbed" and repositioned with the mouse, it may be repositioned by simply clicking at the point where you wish it to be positioned, or use the left and right arrow keys on the keyboard to move it. (The contour

pattern for the MGFP display will change with each new position). The ms indicator is shown on the lowest component displayed, or with the MGFP display if MGFP is selected.

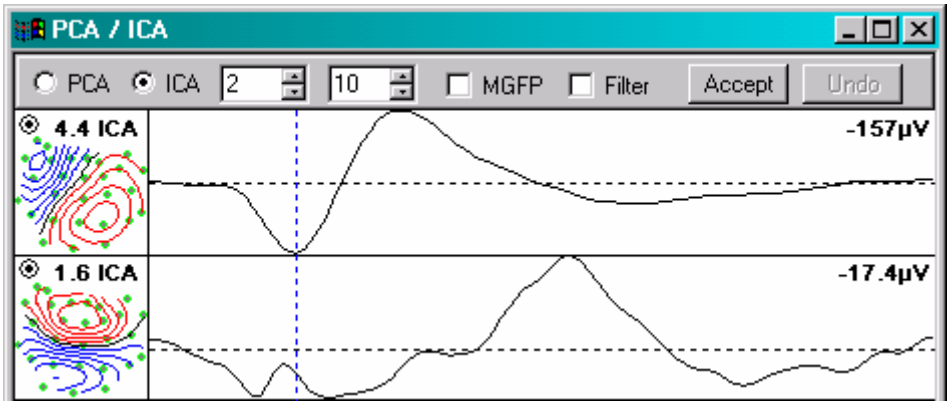
e. The pattern amplitude (in uVs) is displayed. Move the cursor to measure different time points.

PCA versus ICA Loadings

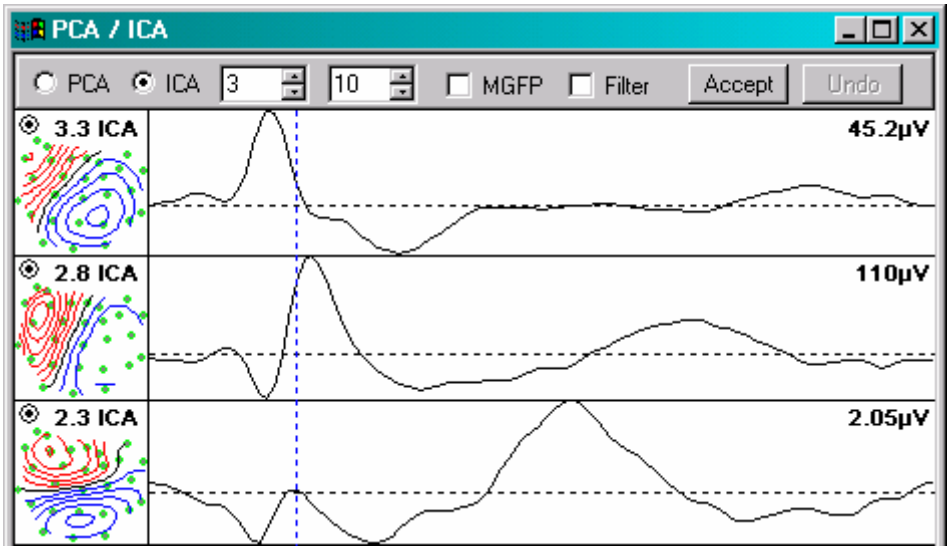
As you use PCA and ICA, you will notice several differences between their operations. When you select PCA and set an interval in the data waveforms, all loadings are computed (there cannot be more loadings than there are electrodes or data points), and your only choice is how many to display (up to 10). Rarely do more than about 5-7 loadings, at most, represent anything other than noise, and these later ones can be disregarded.

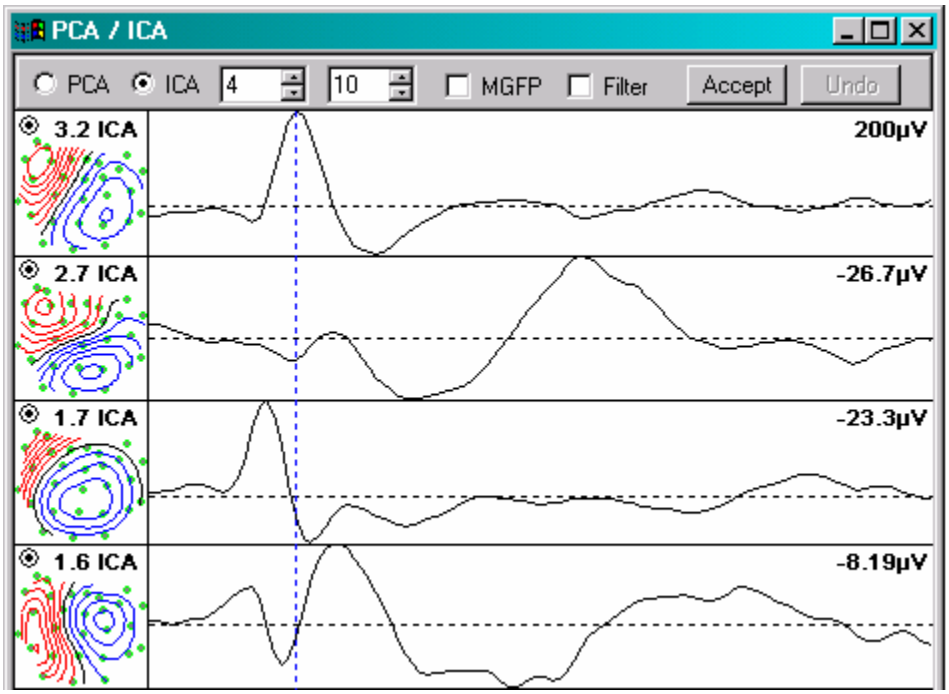
When you select ICA, the program attempts to determine a transformation for the subspace of the PCA results (in a mathematical sense, or, in other words, the number of components you select), such that the components are as independent from each other as possible. If you could select only one ICA component, there is nothing for ICA to do, so the result is the same as the first PCA loading. Therefore, you do not have the option to select only one ICA component - you must select at least two.

As you increase the number of ICA loadings, you may notice that previous loadings may change quite a bit. In the following sequence of figures (using the EpiSpike.avg file), we are increasing the number of ICA loadings from 2 to 5. Note how with each new loading the previous loadings can change.

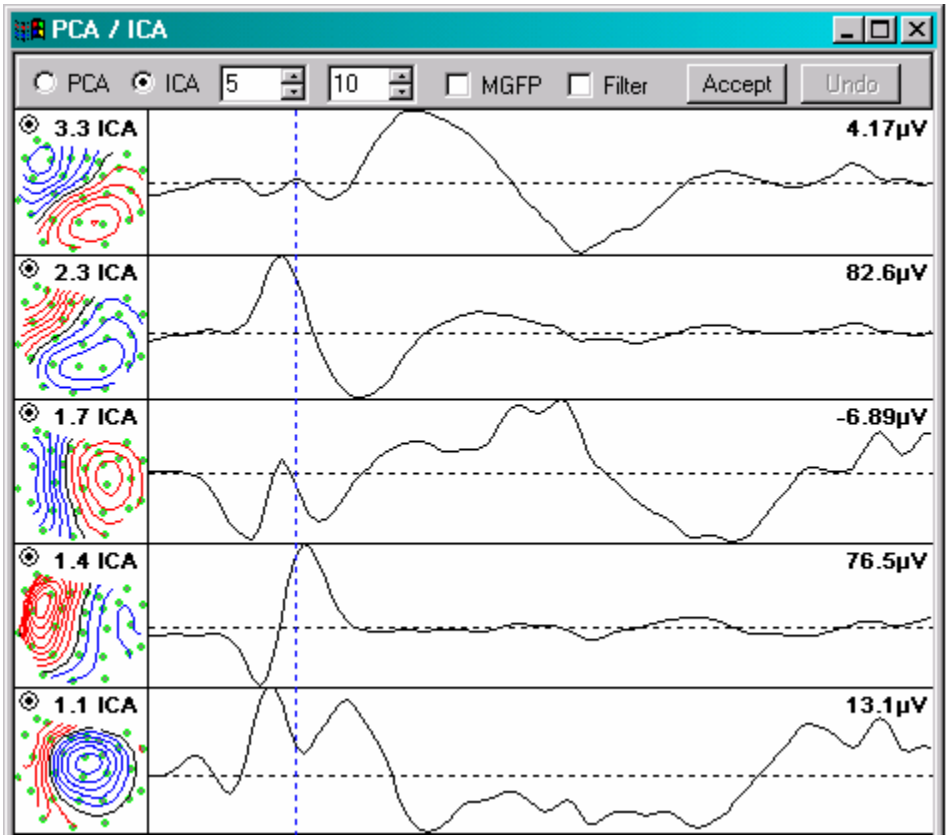


Document number 7235, Revision B





Document number 7235, Revision B

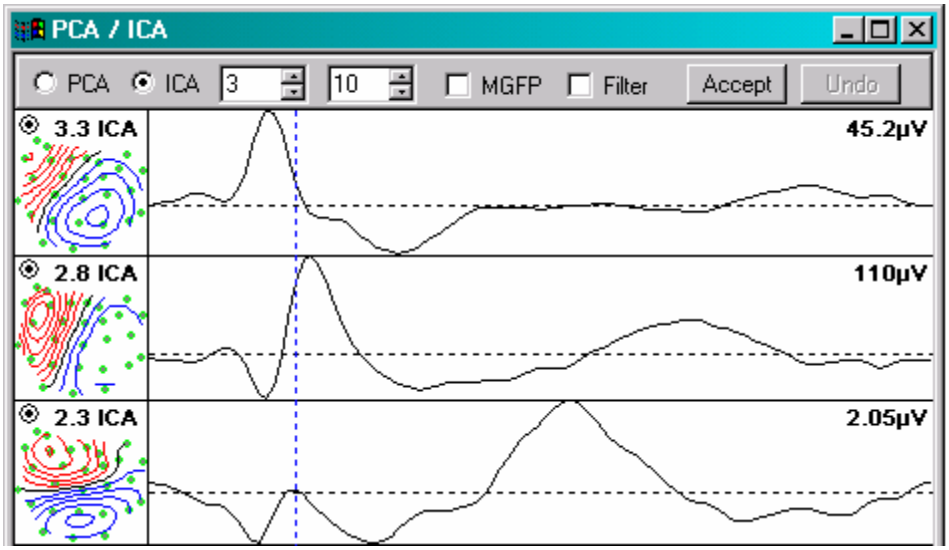


In some instances, the loadings are simply inverted. Again, the PCA/ICA program will take whatever the largest component is and make it positive in each loading display (and the signs are not relevant). In other instances, the order of loadings changes, and, in some cases, features in the loadings may disappear, reappear, or become combined with other loadings. Again, ICA is in each instance redetermining the number of independent components in the signal subspace, making them as statistically independent as possible.

Now, the question arises: if the loadings/patterns vary depending upon how many ICA loadings are computed, how do you know how many loadings to compute? As a general rule, you should use PCA to determine the number of components. If the SNR value is greater than 1.0, the components is potentially valid (1.0 is the noise level). The program will impose this automatically. For example, if there are 3 PCA patterns with SNRs greater than 1.0, ICA will be set automatically to 3 components. (You can manually change the number of ICA components to a larger number, although this is not recommended in general use).

In this example, the PCA results show three patterns with SNRs >1.0, and we therefore perform the ICA analyses with 3 components.

Document number 7235, Revision B



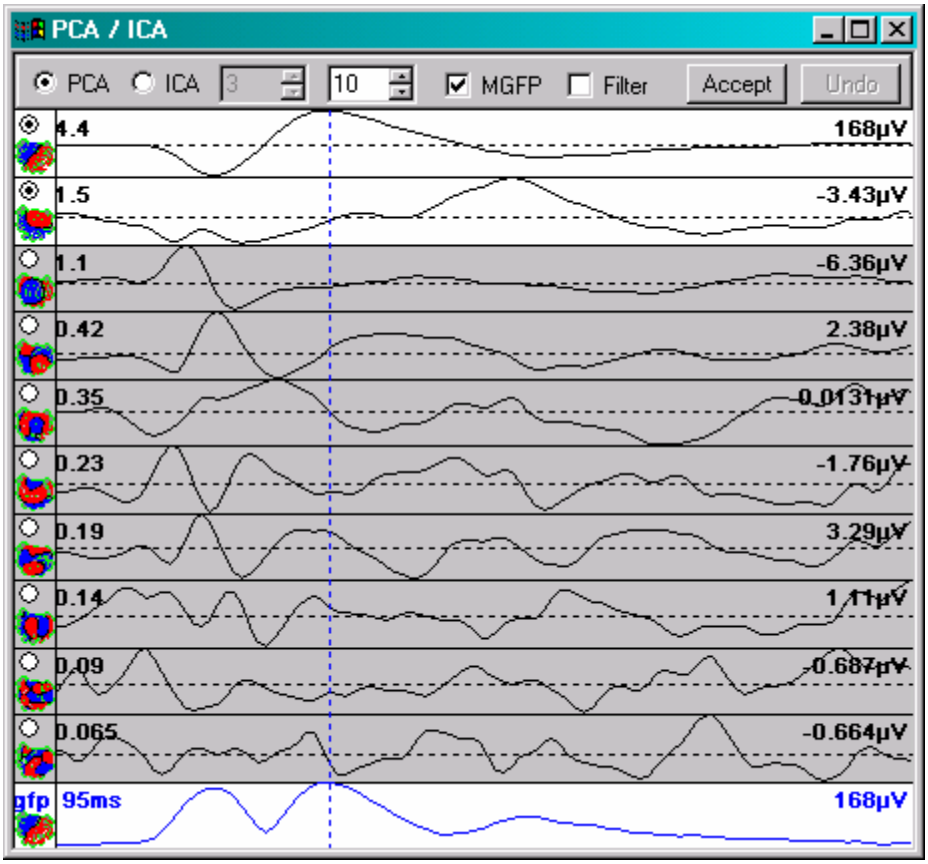
In practice, it may take some trial and error to determine what the best number of loadings should be. Again, there is no absolute right or wrong number. You may need to perform your analyses in several ways, and then determine which way makes the most sense given the effects on the data. The next section describes how to filter the loadings selectively, and how to see the effects of filtering one or more of them out of the waveform data.

Using Filtering in PCA/ICA

One of the fundamental applications of the PCA and especially ICA techniques involves the removal and retention of the various components that are computed from your original EEG waveform data. The Filter option, along with Selecting/Deselecting Patterns, and the Accept and Undo buttons are used in this process. Filtering is possible in the EDIT program only, not during online acquisition in ACQUIRE.

There are some initial considerations to bear in mind when deciding whether to apply filtering based on PCA versus ICA analyses. With PCA, since the extracted field patterns are orthogonal to each other, it is generally not appropriate to remove the leading or distinct non-noise components because this will have no effect on the trailing components. It is appropriate, however, to remove the trailing noise components. In fact, you will find that the program does not let you selectively filter leading PCA components. Whenever you deselect a PCA component, all trailing components are also deselected.

Document number 7235, Revision B



With ICA, the extracted field patterns are not orthogonal and may overlap. You may therefore extract artifact or distinct noise patterns from their temporal independence. In this case, it is appropriate to filter the data by omitting selected patterns.

To elaborate further on this point, PCA is a unique decomposition of the data without any additional parameters other than the time range and the number of sensors. It gives you MIN (NumSamples, NumSensors) components (= field patterns, sorted by their mean amplitude / field strength), and their corresponding loadings (time courses). All patterns are orthogonal to each other (scalar product = 0: no overlap), and all loadings are orthogonal to each other (scalar product = 0: no overlap). Thus, in a mathematical sense, the PCA gives you a unique basis of the data space.

If one has an estimate of the noise in the data, one can omit all trailing patterns that have an amplitude smaller than this noise level. The remaining leading components make up new filtered data. The number of the leading components gives a hint for the number of spatially fixed dipole configurations that are responsible for the signals in the analyzed data range (since spatially fixed dipole configurations lead to fixed field patterns = components). The simplest approach of a fixed dipole configuration is a single fixed dipole; however, the dipoles would not necessarily produce orthogonal field patterns (only in very rare cases like SEPs: tangential N20, radial P23). The PCA patterns, however, span the

Document number 7235, Revision B

whole data space since they form a (orthogonal) basis, which means that the dipole patterns are linear combinations the PCA patterns having overlap to these PCA patterns (scalar product = 0).

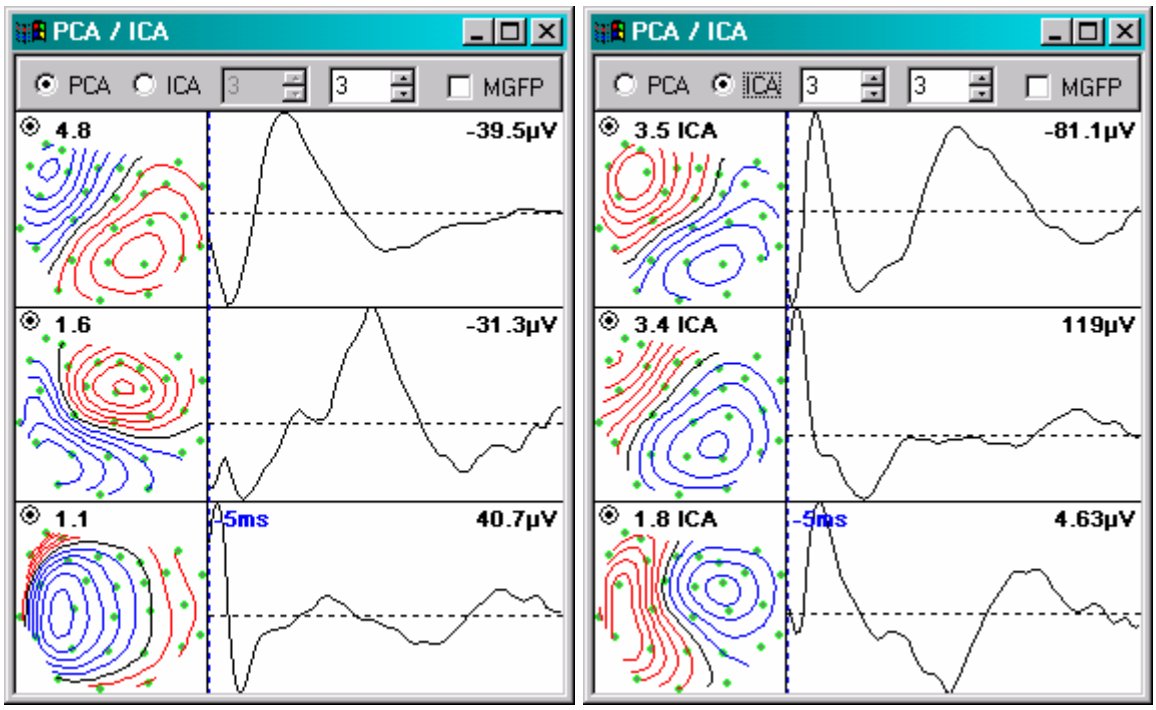
From this you can see that, with PCA, *one must not filter any non-trailing signal components*, since otherwise one would damage (skip) the projections of the dipole patterns to these components. If one would filter the data in this way, one would have to remove the PCA pattern from the computed leadfield of the dipoles as well. This is called projection of the patterns from the data and the leadfield. Only the trailing noise components can be filtered without leadfield projection, since they are assumed to be orthogonal to the remaining signal space. The main information from the PCA is thus the number of relevant components in the analyzed time range!

The ICA implementation uses this information (number of relevant components in the data time range) for speeding up the otherwise slow computation of the ICA decomposition. The number of ICA components is adjusted manually in the PCA/ICA program. If you could select "1" ICA pattern, nothing would happen as compared to the PCA, since one ICA component cannot be independent (independent to what?). An ICA with one component is the same as the PCA first component. Therefore, the program will not allow you to compute a single ICA component; you must select a minimum of two ICA components. With more than one ICA component, one decomposes the data into a signal space (selected ICA components) and a noise space (remaining PCA components). With ICA, *you can extract artifact or distinct noise patterns from their temporal independence*, and it is appropriate to filter the data by omitting selected patterns.

The number of ICA components is NOT a "free" parameter. It has to be determined from the noise properties of the data in the analyzed time range. If the data are composed of independent enough components, the ICA can separate these. These sources in general have overlapping field patterns, and ICA thus gives a non-orthogonal basis of the data space. In this situation, filtering (rejection) of artifact components is possible by synthesizing the data without projection of these components from the leadfield, since the source patterns survive in the remaining components (there is no projection of the artifact patterns from the data).

It is also true that filtering the first *x* PCA loadings from the data file will create a file that is *identical* to the file created by filtering the first *x* ICA loadings (using the same original data file and interval). Even though the selected PCA SNRs, contour patterns, loadings, and loading amplitudes are different from the ICA ones, you will obtain the identical effects when you filter all but the selected loadings. Consider the following.

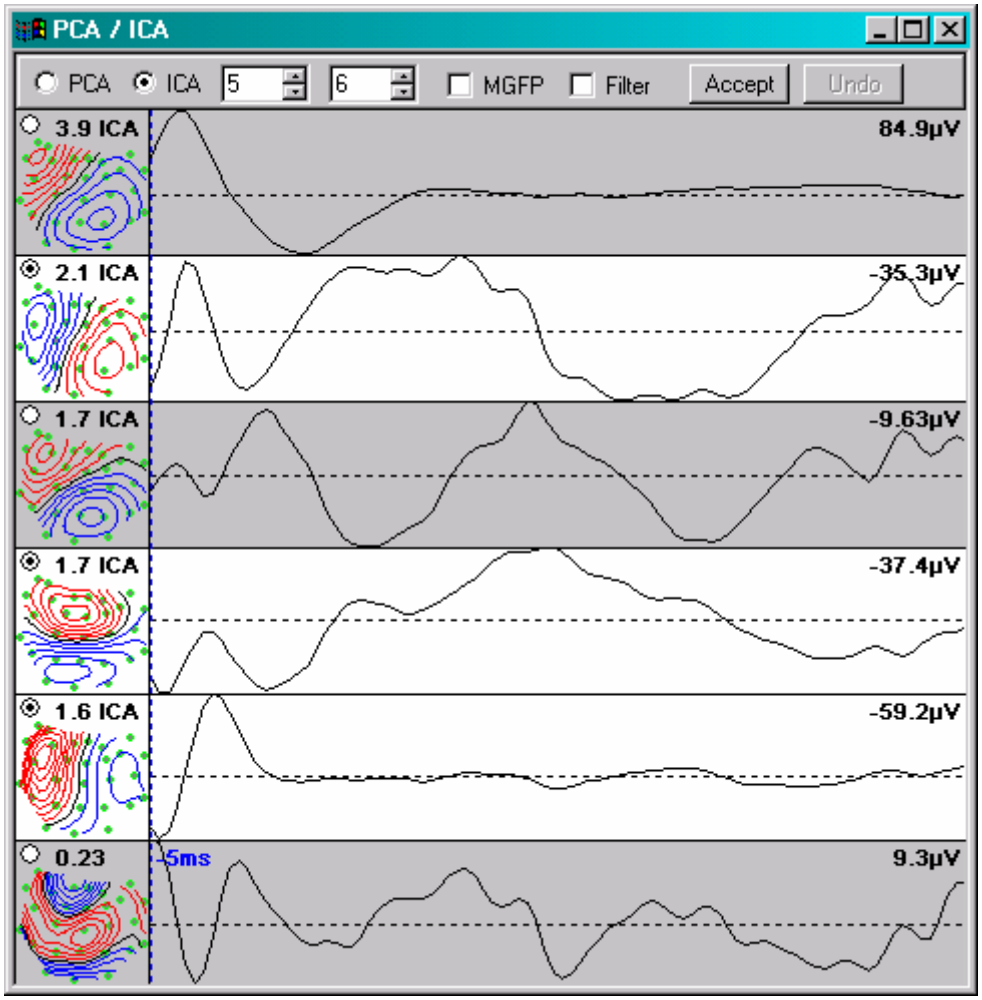
Document number 7235, Revision B



The contours, SNRs, loadings, etc. are fairly different for PCA and ICA. Yet if you filter all but these three loadings from the data files, the results waveforms will be identical between files. Why is that? Think of signal subspace and noise subspace. The deselected components (all of them are PCA components in this example) comprise the noise subspace, Therefore, the noise subspace is the same in both cases. The noise subspace plus the signal subspace equals the total subspace, which is the same for both files. Therefore, the signal subspace must be equal for both PCA and ICA loadings. The "size" of the signal subspace is the same; it is merely subdivided into components differently using PCA versus ICA. When the noise subspace is removed from the waveforms, you must be left with the same signal subspace, and therefore the output files will be identical.

This leads to the advantage inherent with ICA. That is, with ICA you may select leading and trailing components independently. With ICA, you may, for example, deselect components as follows.

Document number 7235, Revision B



You cannot do that with PCA - all trailing components are always deselected. In this case, the filtered data file with selected ICA components will be different from the PCA filtered data file. That is the advantage with ICA. This will be demonstrated below in an example where the first loading is an artifact that we wish to remove in order to obtain more accurate source localization of the remaining P300 component.

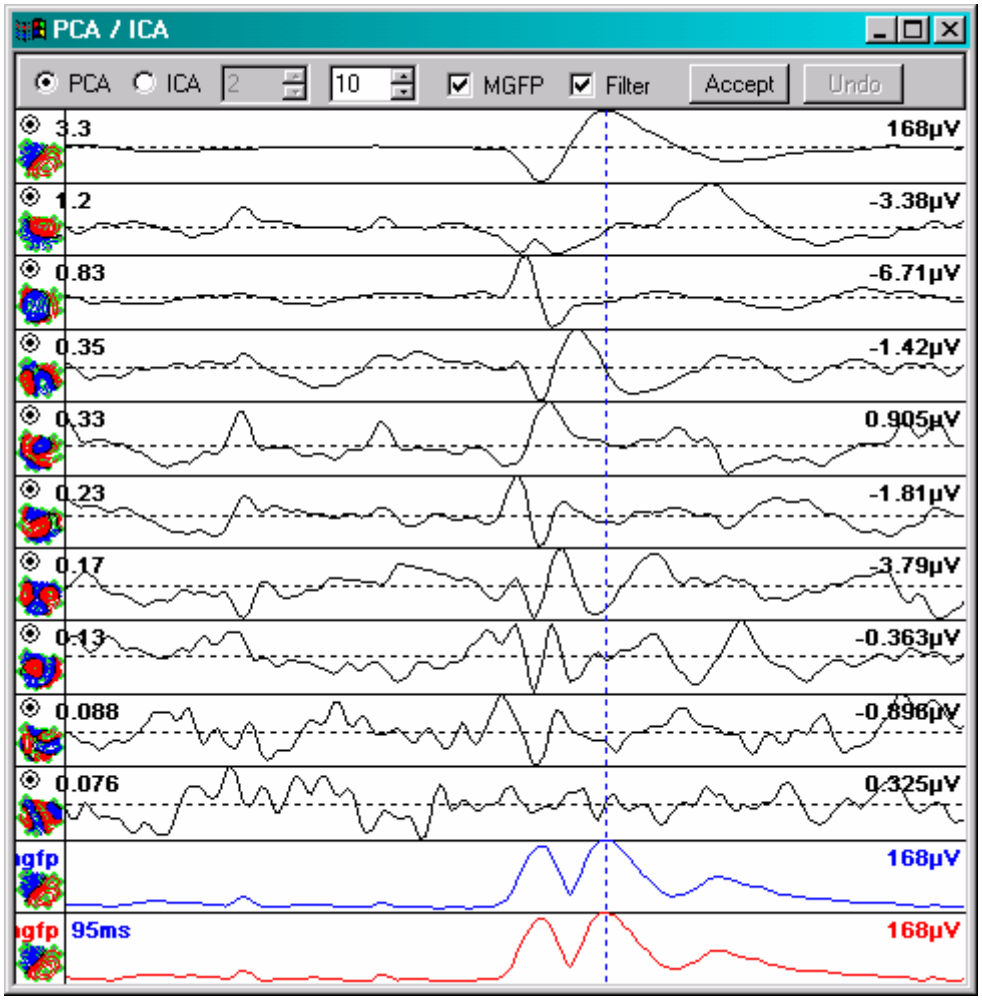
As mentioned above, all PCA loadings are computed, even though only the first 10 can be displayed. The default selection of all components after 10 is "deselected", and they are never displayed nor retained in the filtered data (they are assumed to be noise, and are disregarded). The important point to understand is that, *when selecting and deselecting loadings, you should display all 10 of them to be sure you are including/excluding the ones you want*.

The key to understanding Filtering is: *loadings in gray (deselected) will be removed; loadings in white (selected) will be retained. In other words, you are filtering the data to RETAIN the SELECTED loadings, and REMOVE the DESELECTED loadings*

Document number 7235, Revision B

Using PCA/ICA in source analysis. The components used in dipole source analyses, like all scalp recorded EEG, are potentially influenced by various background or artifactual activities. If these distorting influences can be removed prior to the source analyses, then the dipole localizations should be more accurate. ICA will allow you to select components to retain and to remove any other ones. PCA is used primarily to help determine how many components are valid ones. There are no absolute rules regarding the process to follow; it is part art and part science, with occasional trial and error.

Example 1. In this example we will demonstrate the basic analysis process we recommend. We wish to determine the most accurate dipole source localization for an epileptic spike. Retrieve the EpiSpike.avg file, set the interval for the entire sweep range, and display all 10 PCA loadings. Display the MGFP and filtered MGFP results.



Looking at the relative field strengths, we see that there are 2 valid components (where the SNR > 1.0).

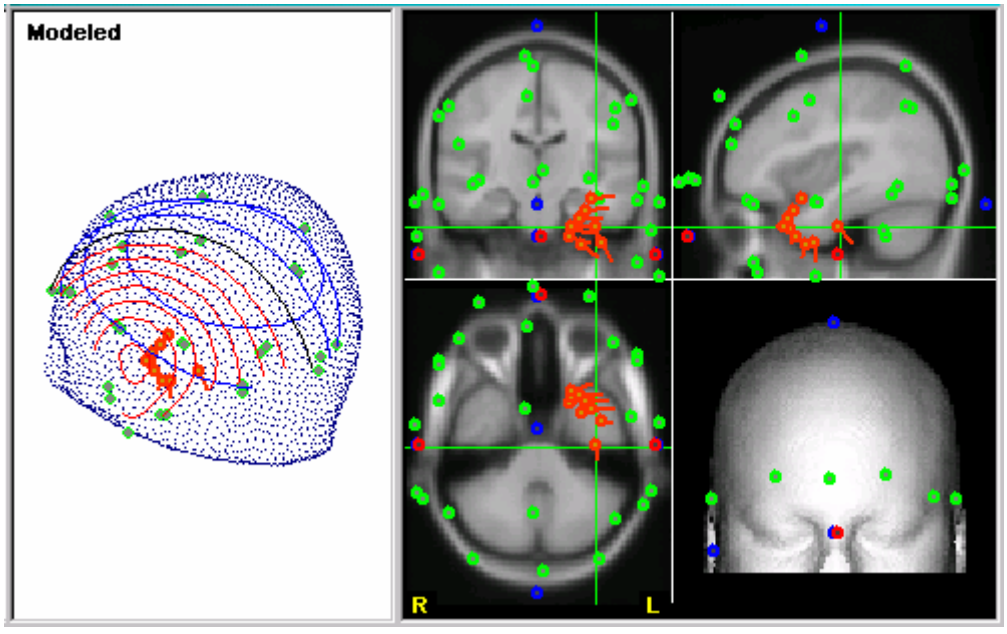
Document number 7235, Revision B

Switch to ICA, compute as many components as possible (10). Select ALL components, and note that the MGFP and filtered MGFP results are identical - nothing has been filtered out. Deselect ALL loadings, and note that the filtered MGFP power is a flat line at zero - everything has been filtered out.

Start selecting any of the ICA components, one at a time, and see the effect each has on the filtered MGFP. The MGFP and filtered MGFP results are autoscaled, so move the vertical cursor to measure the amplitude. Recall, with ICA, the components are statistically independent, and are assumed to have different sources for each.

Now change the number of ICA loadings from 10 to two. The two remaining components change somewhat. You are telling the program to find either two, or more, independent components, and it matters whether you select two or some other number. This is where the PCA part of the program is most useful. From it (from the distribution of relative field strengths), we found there were two valid loadings, so we would use two ICA loadings.

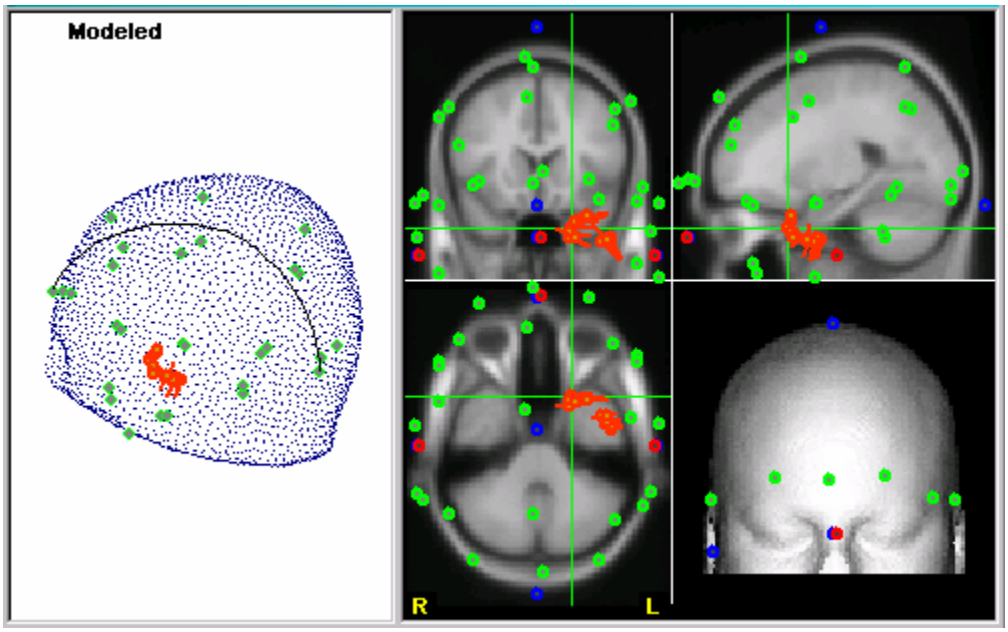
We decide to use a Moving dipole model (in SOURCE), where the two ICA components are computed. In the first figure below, we are using the unfiltered EpiSpike.avg file, an interval of -25 to 65 ms, with a Moving dipole model, 80% threshold, and the standard BEM. The results trace a path around the tip of the left anterior temporal lobe.



After removing all but the first two components, the source appears to be more in the vicinity of the medial surface of the left anterior temporal lobe. This could be a significant difference if surgery were involved. Which is correct? Evidence

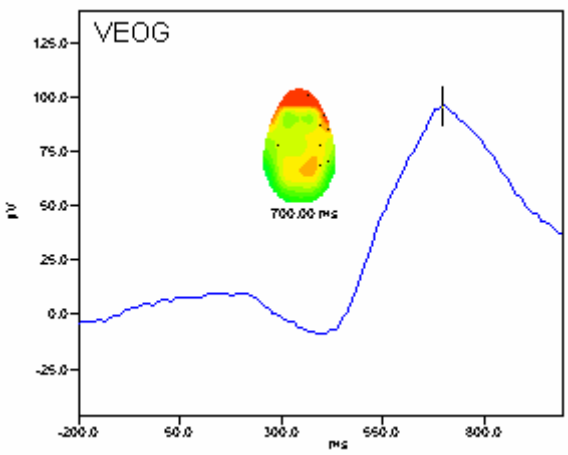
Document number 7235, Revision B

from other sources was consistent with the two loading results above, as suggested by the PCA results.



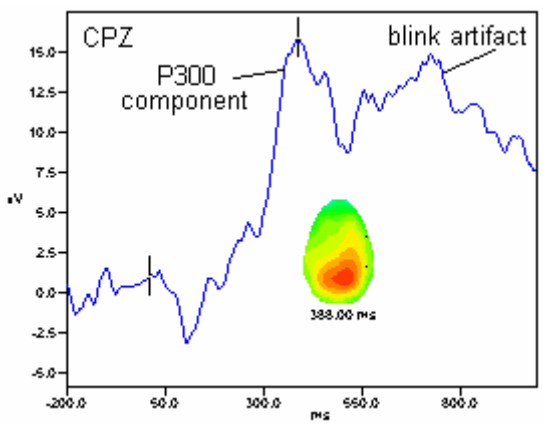
You might want to experiment with different dipole and volume conductor models, as well as number of dipoles computed.

Example 2. In this example, we will use the p300.eeg demo file to demonstrate how to remove blink artifact. Responses to the "oddball" stimuli (type 2) were averaged, without any attempt to remove the blink artifact that is in the file. The result is an AVG file that contains anterior blink artifact and the posterior P300 component. We want to calculate the dipole source localization of the P300 component, without any VEOG artifact. Looking at a 2D map and the VEOG channel, we find the peak blink artifact around 700ms, primarily from anterior frontal sites.



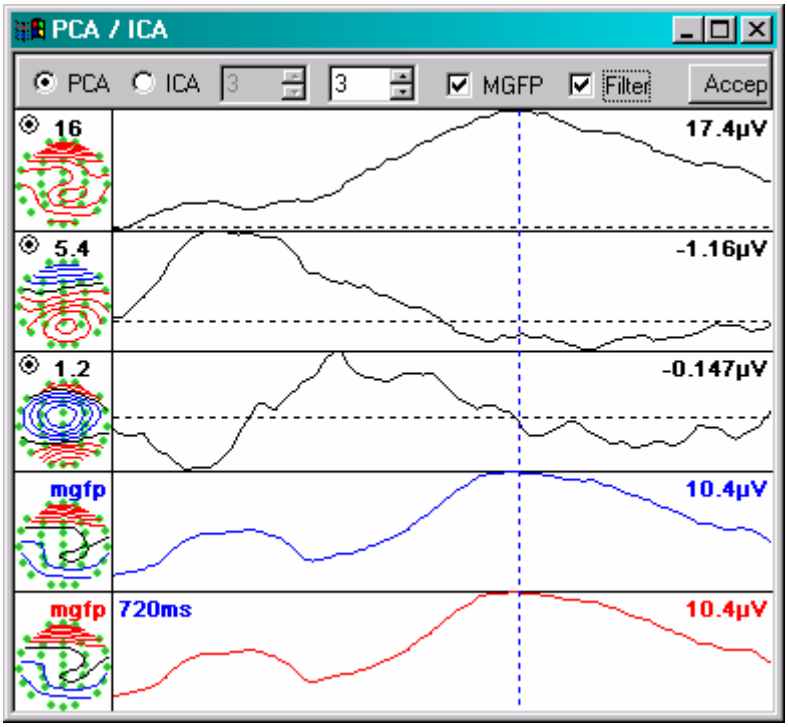
Document number 7235, Revision B

The P300 component peaks at midline posterior sites at about 388ms, with what appears to be blink artifact in the latter part of the waveforms.



We will use PCA/ICA to remove the blink artifact while preserving the P300 component. Start the PCA/ICA program. (If you are following along with the example, you will need to add the electrode position information by clicking the Match Labels button on the "radar" screen when it appears, and then OK and OK).

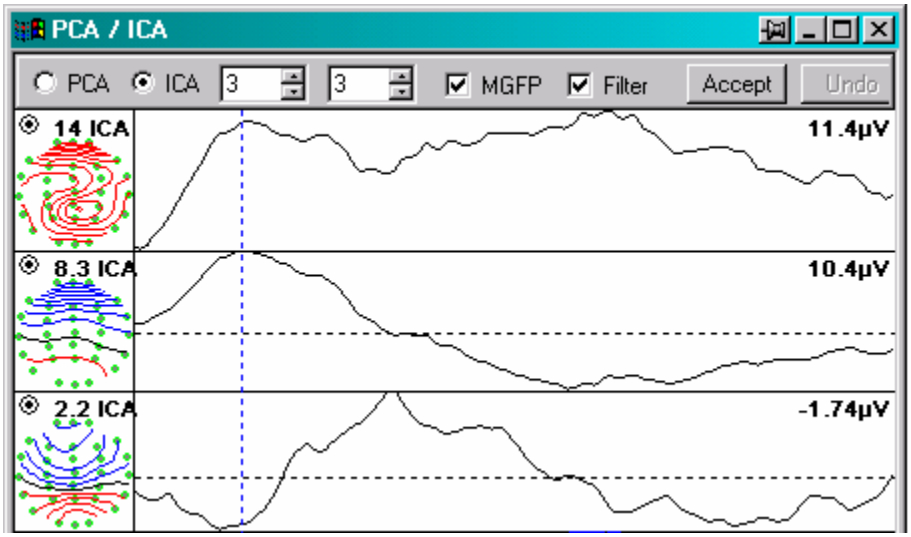
Select the interval between 280ms to 996ms. Display all 10 PCA components, as well as the MGFP and Filtered MGFP results. Looking at the relative field strengths, there are three valid loadings.



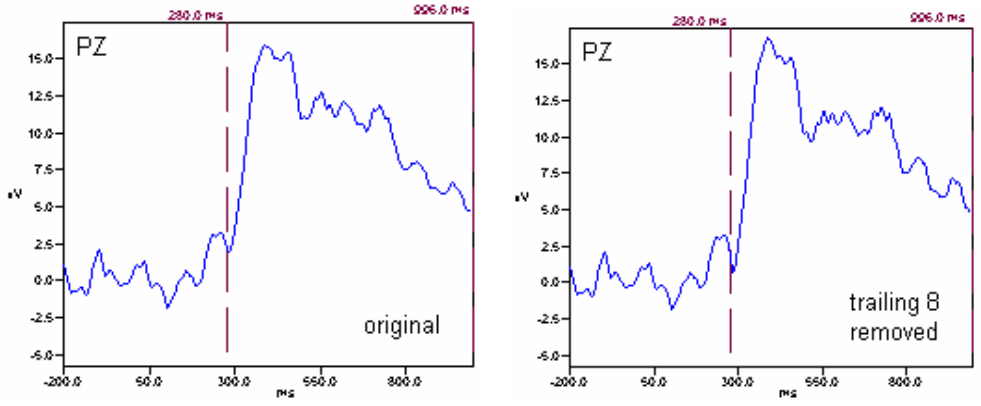
Document number 7235, Revision B

Note that the first one is clearly the blink component. The distribution is frontal, and it peaks around 700ms. The second loading is the P300 component - the distribution is mid-posterior, and it peaks at about 388ms. The third loading is less clear.

Now switch to ICA, and it will be set for three loadings automatically. Here it is less clear what each loading represents, although the first is probably the blink, and the second is probably the P300 component.

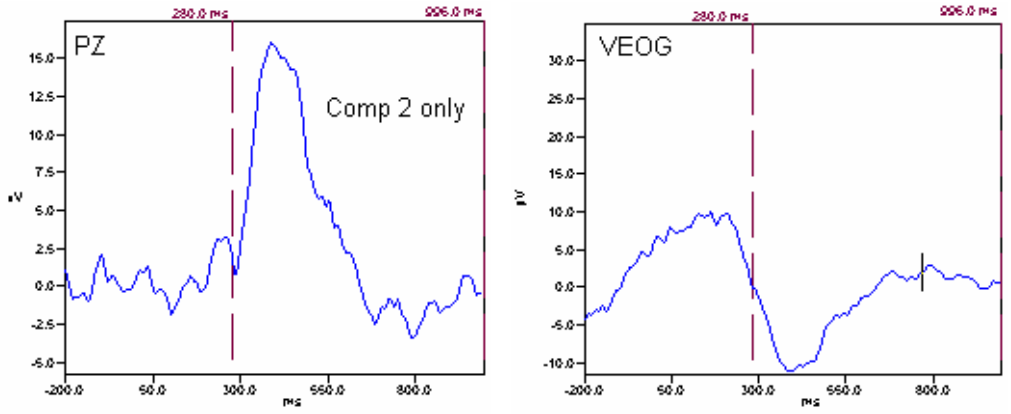


The next step is to try retaining different combinations of loadings to see what effects there are on the waveforms. For example, *deselect* all but the first two loadings. Click Accept and look at the waveforms. Click Undo to restore the originals. Filtering all but the first two components has a small effect.



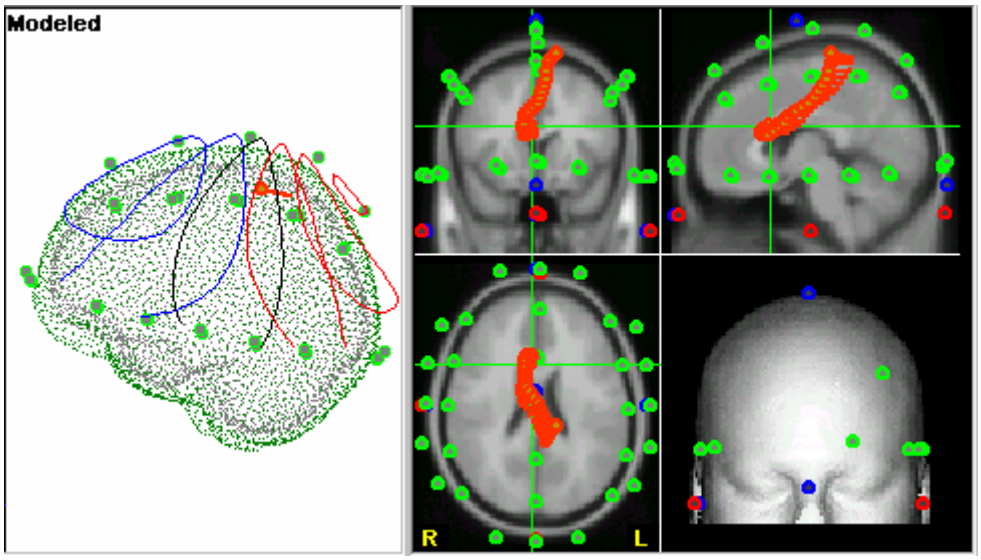
Now deselect all but the second loading. The blink component has been removed, leaving the P300 component intact.

Document number 7235, Revision B

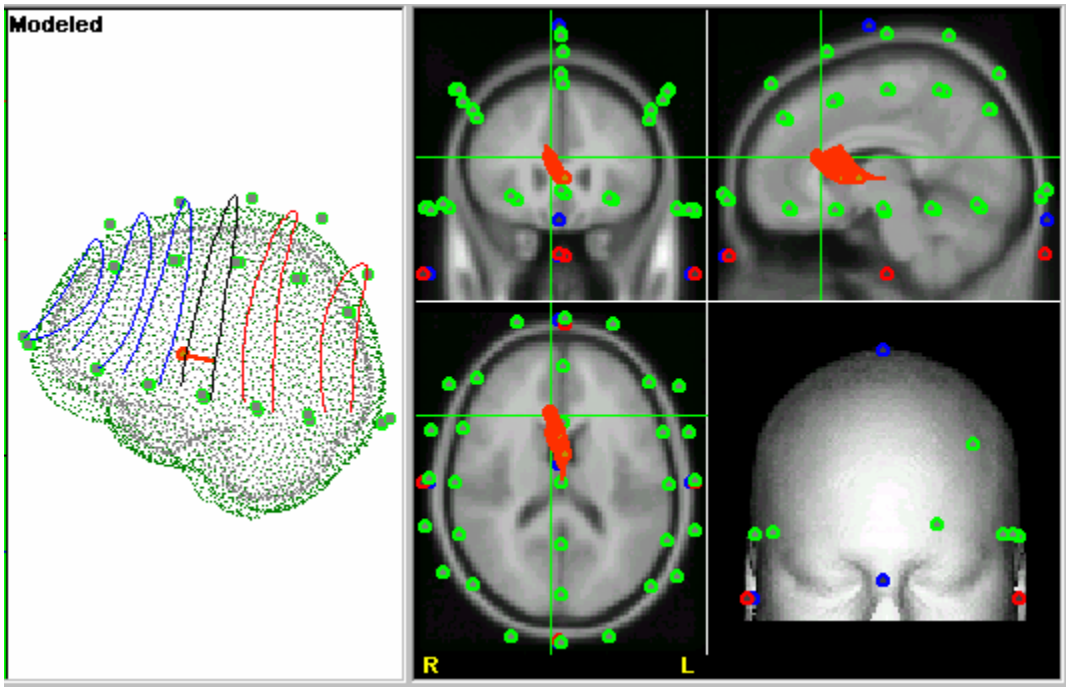


At this point you might try Undoing the filtering, and selecting loadings 2-3 only. There is minimal effect on the P300 component. Save the file with *all but the second and third* loadings filtered (for use in SOURCE). The reason for including the third component will be explained shortly.

Now, we will see what effect - if any - removing the blink component made on the dipole source solution of the P300 component. First, we retrieve the original data file, and compute the P300 source solution using the 300-500ms interval, a Moving dipole model, 70% threshold, and the standard BEM. An interesting albeit somewhat dubious progression is computed. The progression goes from lower left (vicinity of the anterior part of the corpus callosum) to upper right (vertex), in the sagittal view, across the interval.



Next, we take the filtered file, leaving only the second and third components (removing the blink and trailing loadings), and use the same parameters to plot the moving dipole.



Now, we find that the dipole solutions cluster in the vicinity of the anterior part of the corpus callosum. The later drifting toward the vertex is likely due to VEOG artifact. This demonstrates how important it is to remove loadings that may adversely affect the dipole source solutions.

Why did we include the third component as well as the second one? Any time you use only a single component from PCA/ICA, you will obtain dipole solutions in SOURCE that are the same. Moving dipoles will have all dipoles in the same position, and appear not to move. The various dipole models will all give the same results. With only a single component retained, the created data file will have a stable pattern scaled by a multiplicative factor from sample to sample (the ICA loading). Therefore all fits account for the same pattern, giving the same result, and the source strength changes to account for the multiplicative factor. For that reason, we included the third ICA component in the example above (since we were removing the first one).

The examples above illustrate some of the possible applications of PCA and ICA. We encourage you to try a variety of applications with different data sets. We welcome your feedback regarding successful and less successful applications. Please send any comments to techsup@neuro.com.

Document number 7235, Revision B

Summary of Keystrokes

Single and combination keystrokes can be used in place of the mouse for some operations.

When the data display has the focus:

The **left** and **right arrows** move the left vertical interval marker (the start of the interval).
Shift+left or **right arrow** moves the right interval marker (the end of the interval).
Ctrl+left or **right arrow** moves both markers, keeping the interval between them constant.

When the PCA/ICA display has the focus:

The **left** and **right arrow** keys move the cursor in the loadings displays (after clicking the focus there).
The **arrows** and **Tab** keys position the mouse on the Command line (after clicking the focus there).
Shift+arrow and **Ctrl+arrow** have the same function as when the data display has the focus.
The mouse **wheel**, or the **up** and **down** keyboard **arrows**, are used to change the number of lines in the contour map (after clicking inside the patterns/loadings area).

Document number 7235, Revision B