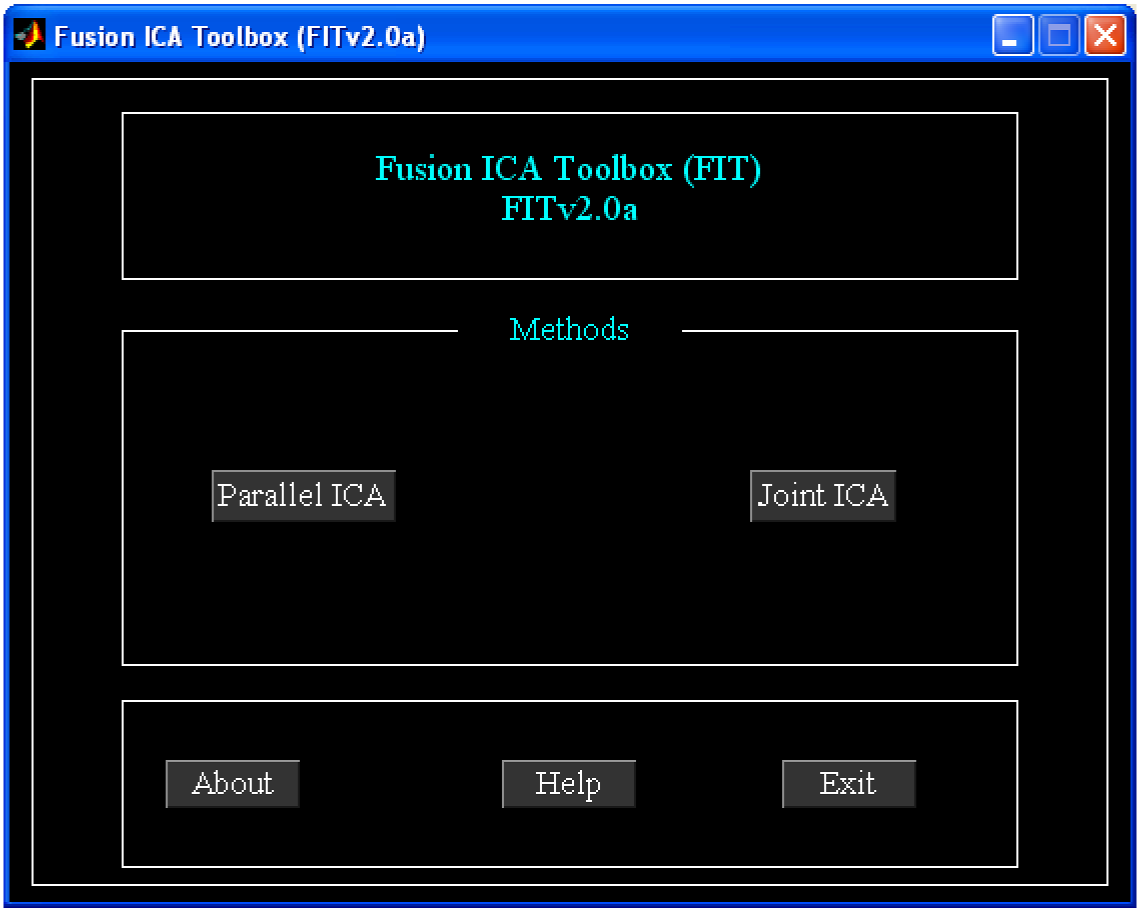
Fusion ICA Toolbox (FIT) Manual

Srinivas Rachakonda[1], Jean Liu[1] and Vince Calhoun[12]
February 8, 2008



---

[1] *The MIND Research Network, Albuquerque, NM*
[2] *Dept. of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM*

Contents

Chapter 1

Introduction

This manual is divided into two chapters. In the first chapter we provide the motivation for using Fusion ICA Toolbox (FIT). In the second chapter, a detailed description of the toolbox is given.

1.1 What is FIT

FIT is an application developed in MATLAB and works on version 6.5 and higher. FIT implements the joint ICA and parallel ICA methods to examine the shared information across modalities.

1.2 Why Joint Analysis?

A single individual often has data collected using multiple modalities including EEG, fMRI, and sMRI. The data from these experiments are usually analyzed separately using statistical parametric mapping (SPM), independent component analysis (ICA) or sometimes directly subtracted from one another on a voxel by voxel basis. The above approach does not examine the shared information between features[1] and voxels. Existing techniques for joint information such as structural equation modeling (SEM) are only applied to certain regions of interest. SEM can be used to look at the correlational structure between regions activated by different tasks. A robust method for examining joint information over full brain is needed. Each imaging technique has a certain advantage, e.g. fMRI has good spatial resolution and EEG has good temporal resolution. A joint ICA model ([3]) for example, was proposed as a technique that combines both spatial and temporal resolution.

1.3 Joint ICA

Joint ICA as applied to two fMRI tasks extracts maximally spatially independent maps for each task that are coupled together by a shared loading parameter. The steps in the joint ICA method are explained as follows:

- Feature Collection - Features are computed for each individual. A feature can be an activation map or EEG signal.
- Feature Normalization - Normalization is done on features using average sum of squares for each task.
- Feature Composition - The data from each feature is stacked across columns and rows represent number of subjects.
- Principal Component Analysis (PCA) - PCA will be used to reduce the data dimension from subjects to components.
- Joint ICA - Spatially independent components will be extracted from the reduced data. Each component shares a common loading or mixing parameter between the tasks.

The advantages of the joint ICA method are given below:

---

[1] Contrast image, ICA spatial map, EEG signal or SNP array.

- Joint ICA can be used to identify an underlying disorder like schizophrenia (SZ). Brain activation patterns from SZ patients may behave similarly for multiple tasks.
- Good spatial and temporal resolution is achieved when fMRI and EEG modalities are fused together.

1.4 Parallel ICA

Parallel ICA is an extension of ICA to accommodate the need for analyzing multiple modalities. Beginning with two modalities, it aims to find hidden factors from both modalities and connections between them. With properly controlled constraints, avoiding over fitting and under fitting caused by multiple reasons, reliable results can be obtained ([1] and [2]).

Comparing with joint ICA, where a shared mixing matrix is used for both modalities, the fundamental difference is that parallel ICA assumes the two data sets are mixed in a similar pattern but not identical. That is, there are two A (mixing) matrices. They are similar but not identical. What is more, certain components from each modality are mixed in a more similar way then the others. Take the analysis of fMRI features and EEG waveforms as an example, it is reasonable to assume inter-subjects variations (the degrees of a component expressed in subjects) are very similar for some components. Contrastingly the same inter-subject variation is assumed in the joint ICA. Parallel ICA pays more attention to individual linked components and their connections, while the joint ICA in reference studies ([5]) inter-effects between EEG and fMRI as a whole.

Chapter 2

Fusion ICA Toolbox

Joint ICA or parallel ICA can be run using the graphical Fusion ICA Toolbox (FIT). Joint ICA can also be run using a batch script (Section 2.3.3). We now provide a step-by-step walk through using example data. First, the example data sets and software must be installed.

2.1 Installing Example Subjects

Unzip `Fusion_Example_Data.zip` file and this contains three folders: `fmri_fmri`, `erp_fmri` and `fmri_gene`. fMRI-fMRI and fMRI-EEG data sets are used to demonstrate joint ICA method whereas fMRI-gene data set is used to explain parallel ICA method. fMRI-fMRI and fMRI-gene data sets consists of two groups like healthy and schizophrenics whereas EEG-fMRI data-set is from a healthy group. fMRI and EEG data were collected while participants performed either an auditory oddball or a sternberg working memory task and are more fully reported in [3] and [5]. The computed "features" which are entered into the ICA analysis, are activation images computed using SPM, an event-related potential (ERP) from a centrally located electrode or SNP array. The data format for each modality is given below:

- Functional MRI (fMRI) or structural MRI (sMRI) data - 3D Analyze or 3D Nifti format.
- ERP or EEG data - ASCII format.
- SNP or Gene data - ASCII format.

2.2 Installing FIT

Unzip `FITv2.0a.zip` file and place it in an appropriate directory. Add folder `ica_fuse` and its sub-folders on MATLAB path. You can also create a `fusion_startup.m` file for setting the path according to your needs. Type fusion at the MATLAB command prompt and this will open FIT (Figure 2.1). FIT contains user interface controls like *Parallel ICA* and *Joint ICA*. We first discuss *Joint ICA* using fMRI-fMRI example data set followed by fMRI-EEG fusion example. *Parallel ICA* is explained using fMRI-gene example data set in Section 2.4.

2.3 Joint ICA Toolbox

When you click *Joint ICA* button in figure 2.1, joint ICA toolbox (JICAT) will open. You can also open JICAT by typing `fusion('jointICA')` at the MATLAB command prompt. Figure 2.2 contains user interface controls like *Setup Analysis*, *Run Analysis*, *Fusion Info*, *Display* and "Utilities". We explain these user interface controls using fMRI-fMRI example data set. fMRI-EEG fusion is discussed in Section 2.3.2.

2.3.1 fMRI-fMRI Fusion

2.3.1.1 *Setup Analysis*

When you click *Setup Analysis* button (Figure 2.2), a figure window will open to select the directory where all the analysis information will be stored. Figure 2.3 shows the initial parameters window. The parameters in the figure are explained below:

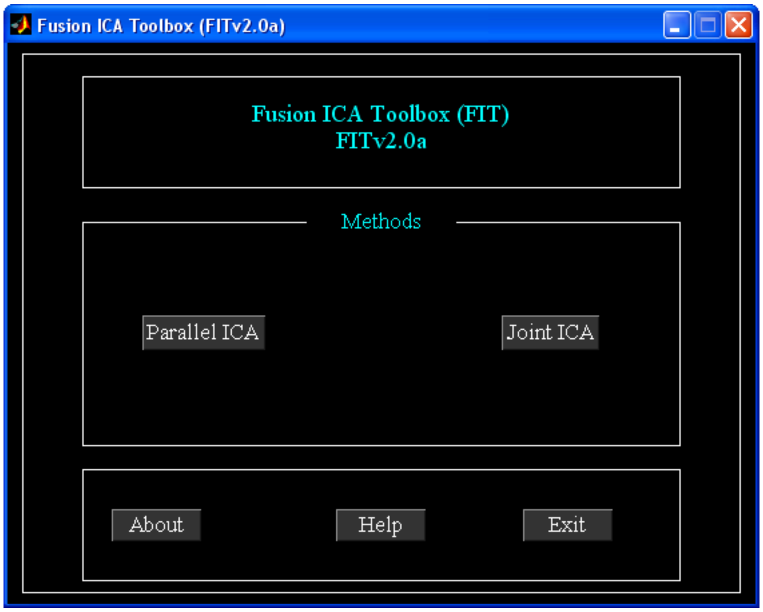


Figure 2.1: GUI for running fusion ICA toolbox.

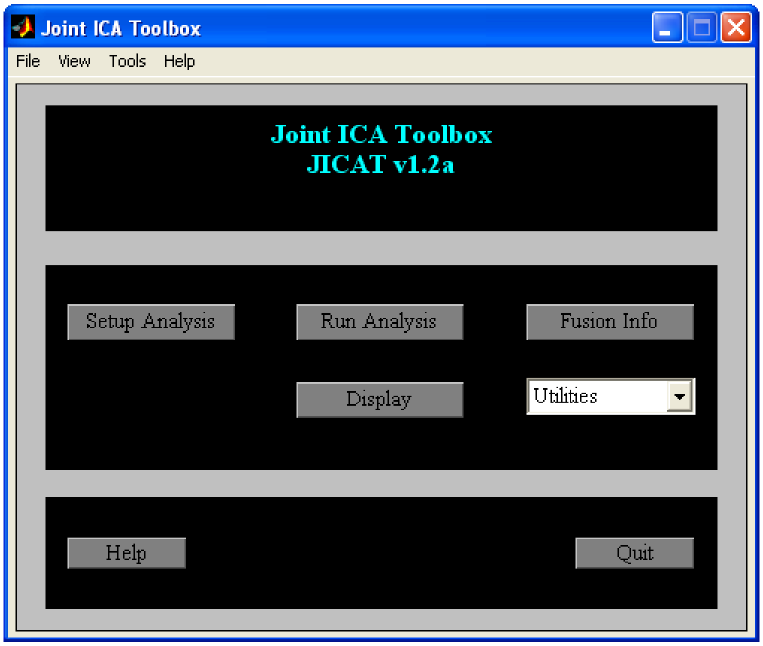
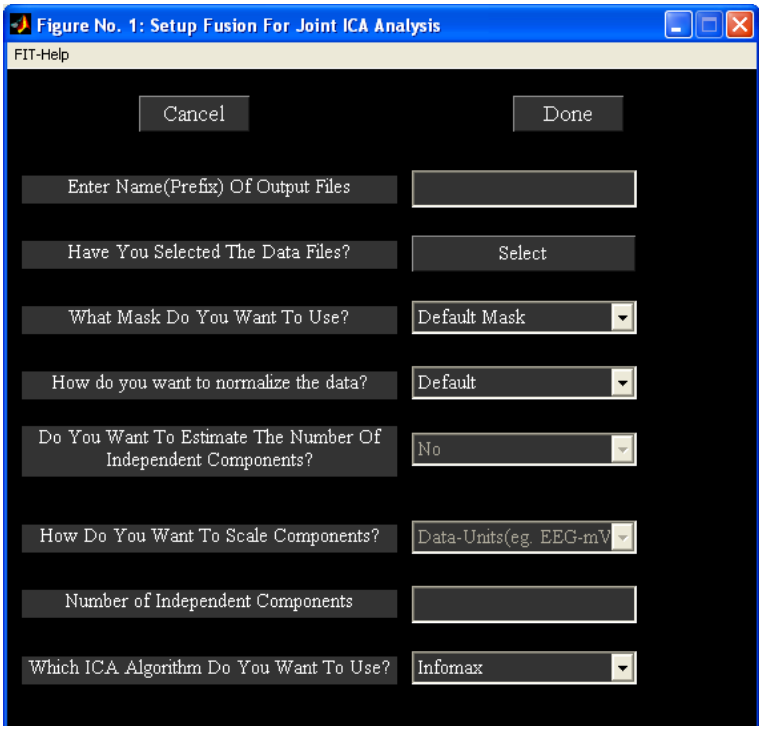


Figure 2.2: GUI for running joint ICA.



Enter Name(Prefix) Of Output Files

Have You Selected The Data Files? Select

What Mask Do You Want To Use? Default Mask

How do you want to normalize the data? Default

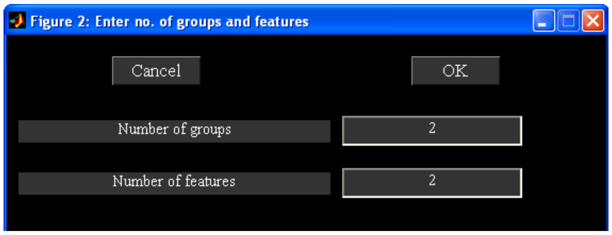
Do You Want To Estimate The Number Of Independent Components? No

How Do You Want To Scale Components? Data-Units(eg. EEG-mV

Number of Independent Components

Which ICA Algorithm Do You Want To Use? Infomax

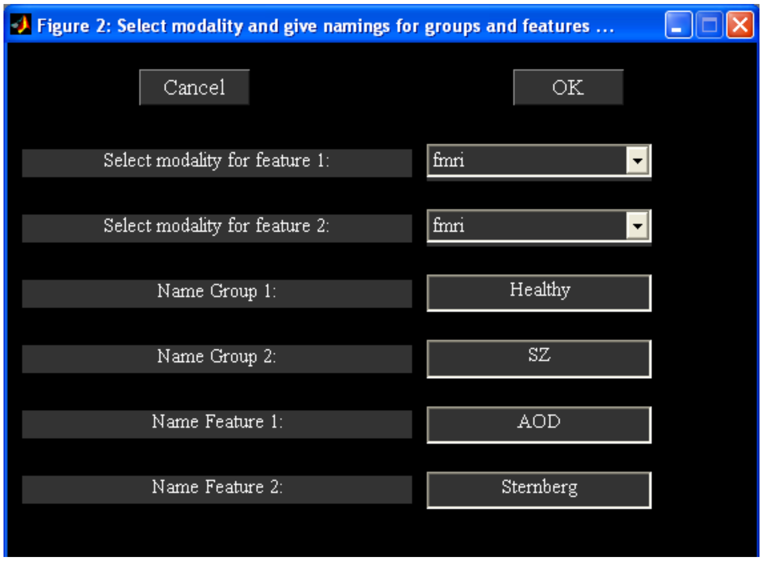
Figure 2.3: Initial parameters select window.



Number of groups 2

Number of features 2

Figure 2.4: Figure window shows number of groups and features selected for the analysis.



Select modality for feature 1: fmri

Select modality for feature 2: fmri

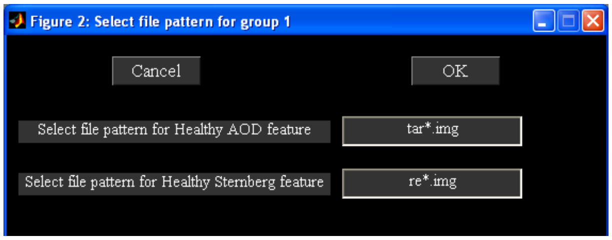
Name Group 1: Healthy

Name Group 2: SZ

Name Feature 1: AOD

Name Feature 2: Sternberg

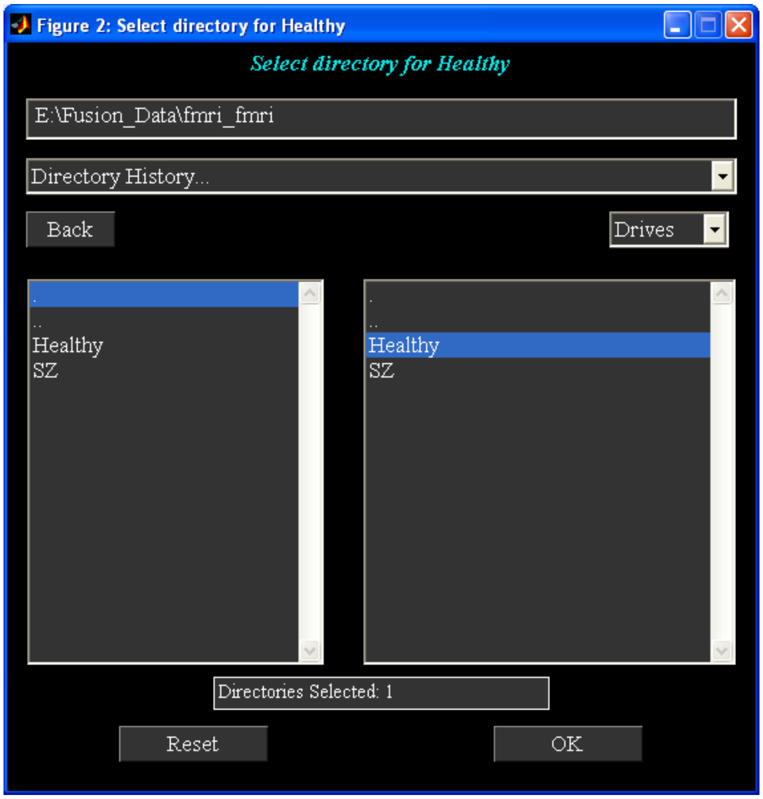
Figure 2.5: Option is provided to enter the names of groups and features.



Select file pattern for Healthy AOD feature: tar*.img

Select file pattern for Healthy Sternberg feature: re*.img

Figure 2.6: File pattern of each feature for first group.



E:\Fusion_Data\fmri_fmri

Directories Selected: 1

Figure 2.7: Selected directory for first group.

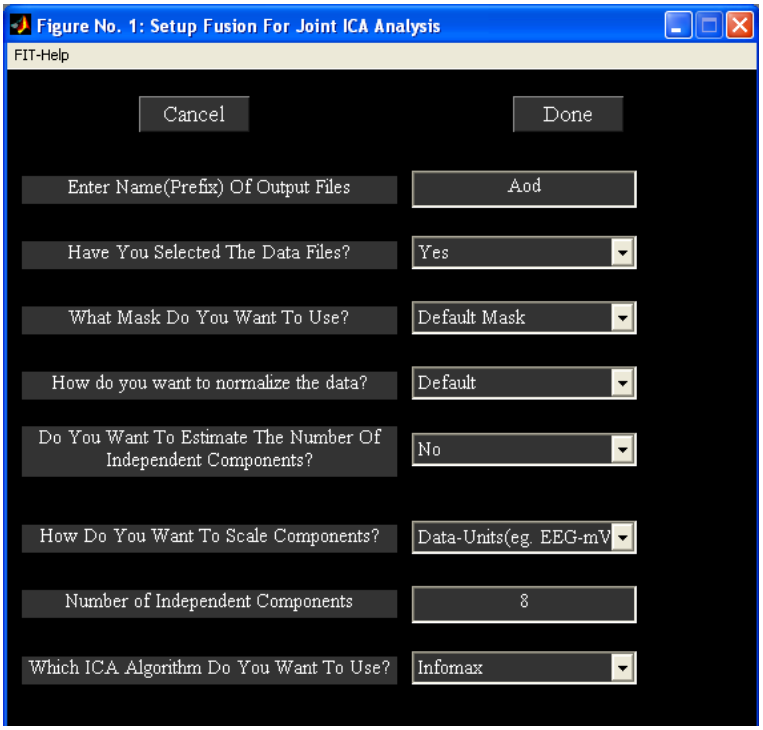
- ”Enter Name (Prefix) Of Output Files” - All the output files will be stored with this prefix.
- ”Have You Selected The Data Files?” - When you click push button *Select*, figure 2.4 will open to enter the number of groups and features for the analysis. After entering the information in figure 2.4, figure 2.5 will open to enter the names of groups and features. For the example data-set, the files filter pattern is the same for both groups. Enter only the file pattern (Figure 2.6) for features of first group. Since the directories where the groups are stored are separate, select ”No” for question ”Is the data organized in one group folder?” This will let you select directory (Figure 2.7) for each group. After the data-sets are selected, a drop down box will show the answer ”Yes” for question ”Have You Selected The Data Files?” The data files selected for the analysis are printed to a text file with suffix `selected_data.txt`.
- ”What Mask Do You Want To Use?” - There are two options like ’Default mask’ and ’Select Mask’. Each option is explained below:
  - ’Default Mask’ - For fMRI and sMRI, the default mask includes only the voxels that are non-zero and not Nan for all the subjects. For EEG default indices used are from global variable EEG_DATA_INDICES in `ica_fuse_defaults.m` file.
  - ’Select Mask’ - When you click ”Select Mask” option, a new figure window will open to select the mask. Mask for each modality can be selected by clicking the feature name in the left listbox of that figure.
- ”How do you want to normalize the data” - Options available are ’Default’, ’Norm2’, ’Std’ and ’None’. Normalization is done separately for each feature over groups. Default normalization uses square root of mean of squared data for all subjects.
- ”Do You Want To Estimate The Number of Independent Components?” - The data used for estimation is stacked across columns. Components are estimated based on the Minimum Description Length (MDL) criteria. The algorithm used for estimation uses a smoothness factor from variable FWHM_VALUE (`ica_fuse_defaults.m`) to account for correlated data samples.
- ”How Do You Want To Scale Components?” - You have the option to scale components to data units or Z-scores. Each option is explained below:
  - ’Data-Units(eg. EEG-mV)’ - Regression fit is calculated by using components of a feature as model and original data of that feature as observation. Components of a feature are scaled by their respective slope or beta weight. This involves flipping of components when slope is negative. You can turn off this option by setting variable FLIP_SIGN_COMPONENTS in `ica_fuse_defaults.m` to 0.
  - ’Z-scores’ - Components are first converted to ’Data-Units’ and then converted to Z-scores.
- ”Number of Independent Components” - Number of independent components that will be extracted from the data.
- ”Which ICA Algorithm Do You Want To Use?” - Presently there are 8 ICA algorithms implemented in the FIT like Infomax, FastICA, ERICA, SIMBEC, EVD, JADE OPAC, AMUSE and SDD ICA.

Figure 2.8 shows the completed parameters for data fusion analysis. When you click *Done* button, ICA options window (Figure 2.9) will open. Presently, ICA options are available for Infomax, FastICA and SDD ICA. You can use the defaults or enter values for the parameters within permissible limits that are shown in the prompt string. All the user input is stored in a MAT file having suffix `ica_fusion.mat`.

2.3.1.2 *Run Analysis*

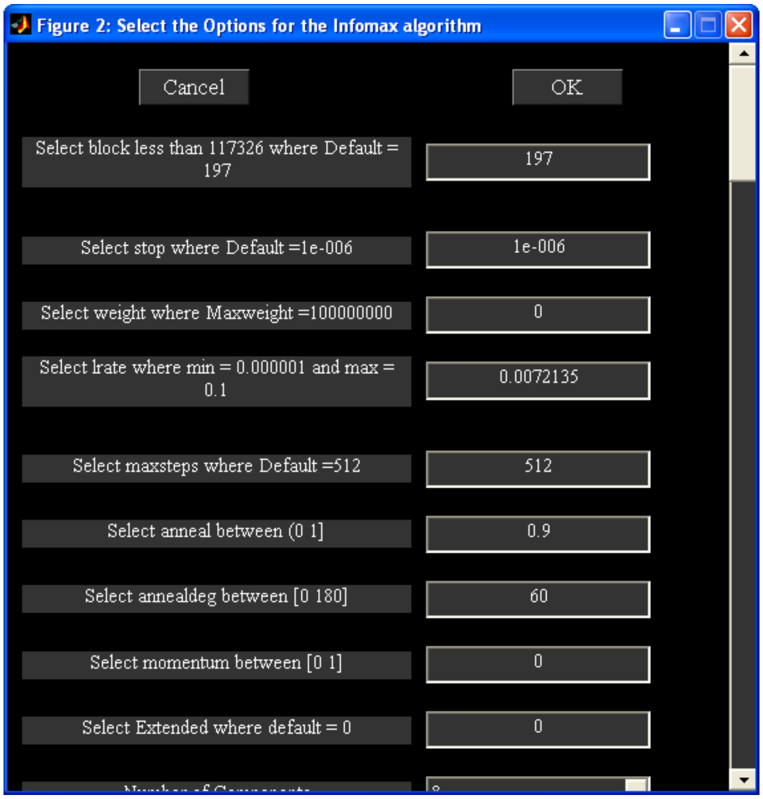
Fusion analysis can be done through *Run Analysis* button or by selecting ”Run” under ”Tools” menu (Figure 2.2). Select the fusion parameter file (`*ica_fusion.mat`) which was created after setting up the analysis and wait for the analysis to complete. The steps involved in the analysis are as follows:

- Principal Component Analysis (PCA) - Data will be reduced using PCA. The information about data reduction is stored in a MAT file with the suffix `pca_comb`. First combination is all the features stacked together in columns.
- Independent Component Analysis (ICA) - ICA will be run on the reduced data obtained from the PCA step. The information about ICA is stored in a MAT file with suffix `ica_comb`.



Enter Name(Prefix) Of Output Files	Aod
Have You Selected The Data Files?	Yes
What Mask Do You Want To Use?	Default Mask
How do you want to normalize the data?	Default
Do You Want To Estimate The Number Of Independent Components?	No
How Do You Want To Scale Components?	Data-Units(eg. EEG-mV
Number of Independent Components	8
Which ICA Algorithm Do You Want To Use?	Infomax

Figure 2.8: Figure window shows the completed parameters.



Select block less than 117326 where Default = 197	197
Select stop where Default =1e-006	1e-006
Select weight where Maxweight =100000000	0
Select lrate where min = 0.000001 and max = 0.1	0.0072135
Select maxsteps where Default =512	512
Select anneal between (0 1]	0.9
Select annealdeg between [0 180]	60
Select momentum between [0 1]	0
Select Extended where default = 0	0

Figure 2.9: ICA options for Infomax algorithm.

- Back Reconstruction - Components mixing matrix is multiplied with de-whitening matrix obtained from the PCA step. This will be stored as a MAT file with suffix `br_comb`.
- Scaling Joint ICA Components: The components obtained after ICA step are in arbitrary units. These will be scaled by doing a multiple regression of components (model) with the original data (observation) of the features. The sign of the component will be flipped depending on the beta weight information. This information is saved as a MAT file with suffix `sc_comb`.
- Output Files: Joint ICA components will be saved in output files for the respective features with the suffix `joint_comp_ica_feature`.

Note:

- All the analysis information is stored in a log file with suffix `results.log`. After the analysis is done successfully, display GUI (Figure 2.11) will open. You can turn off this option by setting variable OPEN_DISPLAY_WINDOW to 0 in `ica_fuse_defaults.m` file.
- PCA and ICA will be run on combination of features, if you set variable OPTIMIZE_FEATURES to 'Yes' in `ica_fuse_defaults.m` file.
- You can run ICA several times by setting variable NUM_RUNS_ICA in defaults.

2.3.1.3 *Fusion Info*

Fusion Info button is used to view analysis information. Figure 2.10 will open after you had selected the fusion information file. The function of each button in the figure is given below:

- *Parameter Info* - User input information is shown.
- *Analysis Info* - Analysis information is shown like how many combinations of features are run.
- *Output Files* - Output files information is shown.

2.3.1.4 *Display*

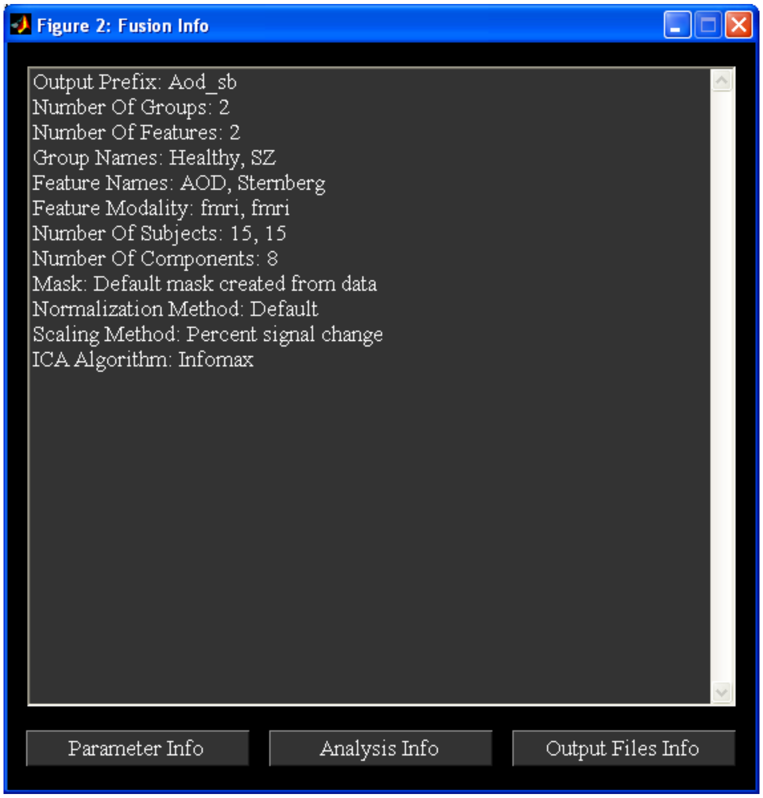
When you click *Display* button (Figure 2.2) and have selected the fusion parameter file, figure 2.11 will be displayed. An alternative way to display *Display* GUI figure is to select "Display" under "Tools" menu. *Display* GUI is used to display joint ICA components. Display GUI contains main user interface controls, hidden user interface controls (display defaults) and "Utilities" menu. Hidden user interface controls (Figure 2.12) will be displayed when you click "Display Defaults" menu. There is an option to create ERP-fMRI (Section 2.3.2) movie under "Utilities" menu. We next explain main user interface controls followed by hidden user interface controls.

Main User Interface Controls

- "Component No" - Component numbers to display. By default all components will be selected.
- "Feature" - Features to display. By default all features will be displayed.
- "Do You Want To Sort Components" - Sorting joint ICA components is explained in section 2.3.1.5.

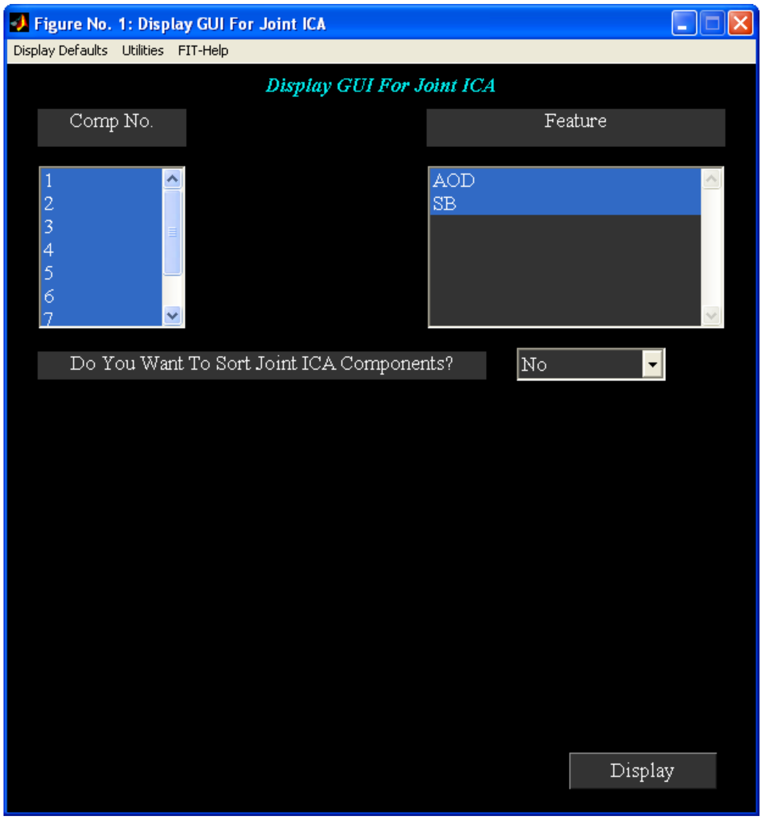
Hidden User Interface Controls

- "Convert To Z-scores" - Component images will be converted to z-scores.
- "Threshold" - Z-threshold used for displaying images.
- "Image Values" - Options available are "Positive and Negative", "Positive", "Absolute" and "Negative".
- "Components per figure" - Options available are "1", "4" and "9".
- "Anatomical Plane" - Options available are "Axial", 'Sagittal" and "Coronal".
- "Slices (in mm)" - Slices in mm to be plotted.



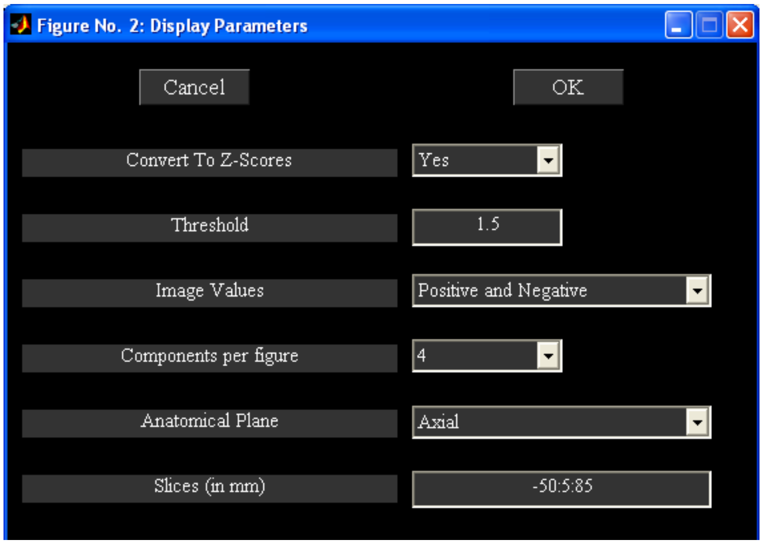
Output Prefix: Aod_sb
Number Of Groups: 2
Number Of Features: 2
Group Names: Healthy, SZ
Feature Names: AOD, Sternberg
Feature Modality: fmri, fmri
Number Of Subjects: 15, 15
Number Of Components: 8
Mask: Default mask created from data
Normalization Method: Default
Scaling Method: Percent signal change
ICA Algorithm: Infomax

Figure 2.10: Figure shows the analysis information.



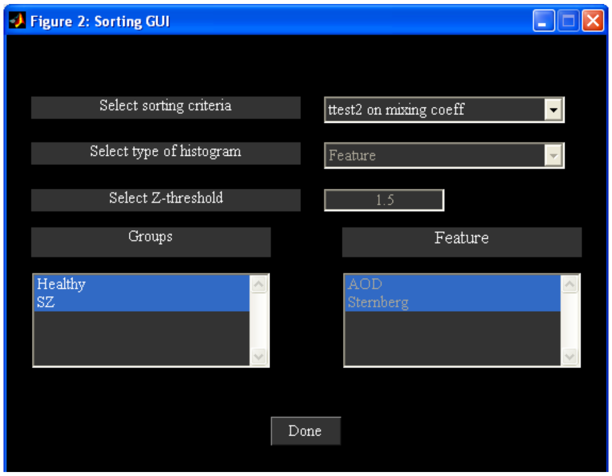
Do You Want To Sort Joint ICA Components? No

Figure 2.11: Display GUI.



Convert To Z-Scores	Yes
Threshold	1.5
Image Values	Positive and Negative
Components per figure	4
Anatomical Plane	Axial
Slices (in mm)	-50:5:85

Figure 2.12: Display GUI defaults.



Select sorting criteria	ttest2 on mixing coeff
Select type of histogram	Feature
Select Z-threshold	1.5

Groups	Feature
Healthy	AOD
SZ	Sternberg

Figure 2.13: Figure shows the utility to sort joint ICA components.

Note: The defaults for these display parameters are available in `ica_fuse_defaults.m` file. Anatomical image used for displaying images is from variable ANATOMICAL_FILE in `ica_fuse_defaults.m` file. After entering the display parameters, click on *Display* (Figure 2.11) button to display the features. Figure 2.14 shows the components in groupings of four.

2.3.1.5 Sorting Components

In order to sort joint ICA components, statistics toolbox must be installed. Select "Yes" for "Do You Want To Sort Joint ICA Components?" in figure 2.11. Figure 2.13 shows the utility to sort components. The parameters in the figure 2.13 are explained below:

- "Select sorting criteria" - Presently there are two options like "ttest2 on mixing coeff" and "Spatial div". Explanation of each option is given below:
  - "ttest2 on mixing coeff" - Two sample t-test is done on mixing coefficients between the selected groups. Figure 2.15 shows components sorted based on p-value. The results from two sample t-test are stored in a text file with suffix `ttest2_mixing_coeff.txt`.

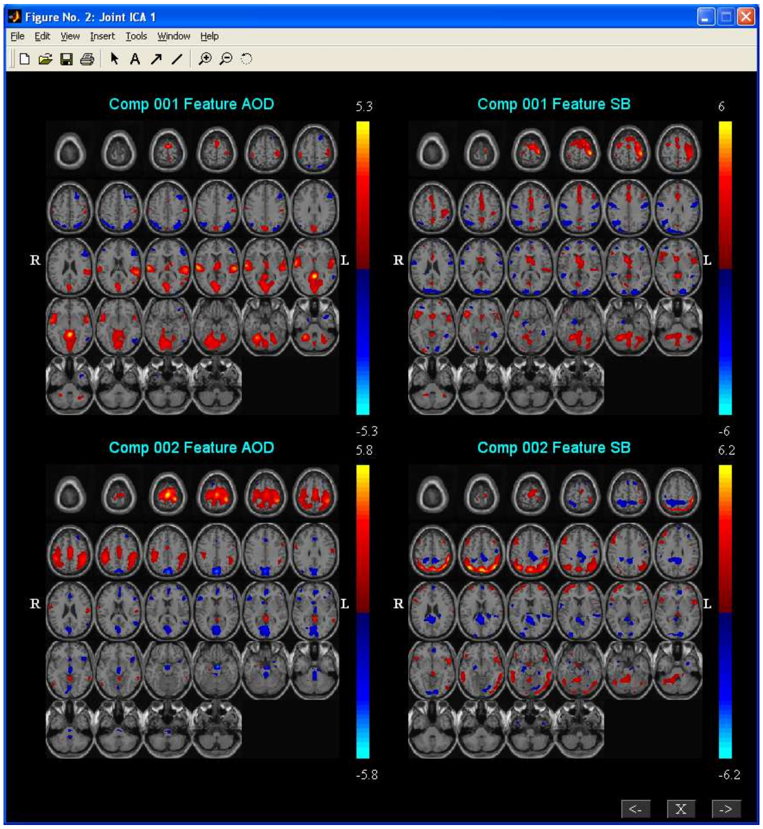


Figure 2.14: Figure shows components in groupings of four.

  - ”Spatial div” - To calculate spatial divergence between the groups, a distribution must be created for each component of a group. We use cross-task histogram as distribution for each selected group from spatial maps. You need to select the features from features listbox and groups from groups listbox (Figure 2.16). Divergence is calculated between the selected groups. Figure 2.17 shows components sorted based on spatial divergence. The results from divergence are stored in a text file with suffix `divergence_groups.txt`.

- ”Select Type of Histogram” - There are two types of histograms like ”Feature” and ”Component”. We explain the type of histograms below:
  - Feature - Z-threshold is applied on each component map of a feature and voxels are sorted in descending order. The resulting set of voxels is used as a mask to original data. Cross-task histogram is computed based on the selected features from features listbox.
  - Component - Component voxels for each feature are sorted in descending order. Cross-task histogram for each component is generated based on selected features from features listbox.
- ”Select Z-threshold” - This option is available only when you select ”Spatial div” as the sorting criteria and select ”Feature” as type of histogram.
- ”Groups” - Select at most two groups for sorting components.
- ”Features” - You can select features only when ”Spatial div” is used as the sorting criteria.

When you click *Done* button, components are sorted based on the criteria selected.

Note:

- Options for divergence criteria are ”kl” (Kullback Leibler), ”J” (J), ”alpha” (Alpha) and ”renyi” (Renyi). For setting the appropriate divergence criteria see defaults (`ica_fuse_defaults.m`).

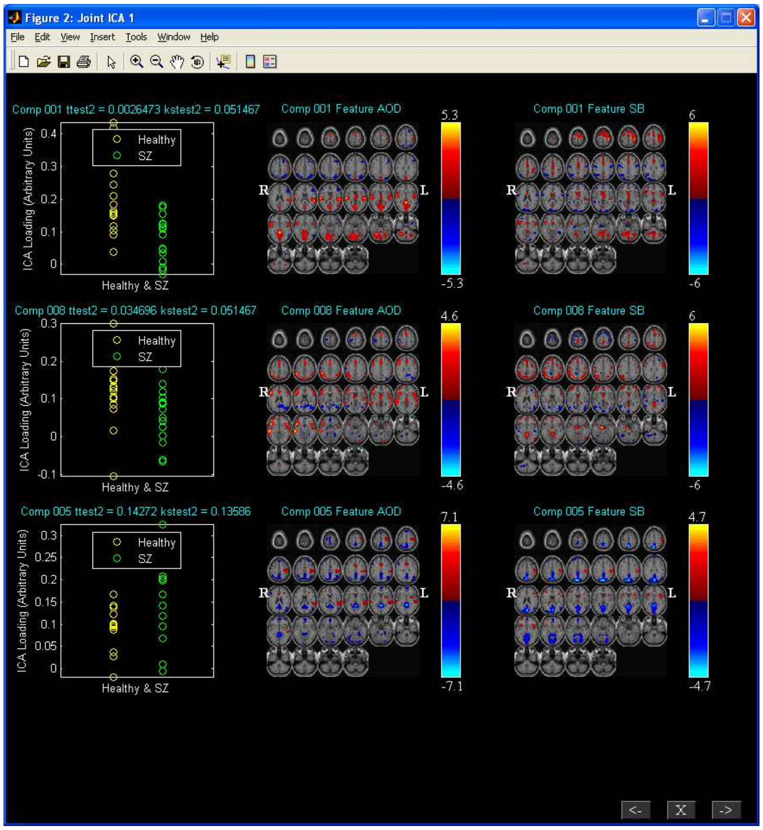
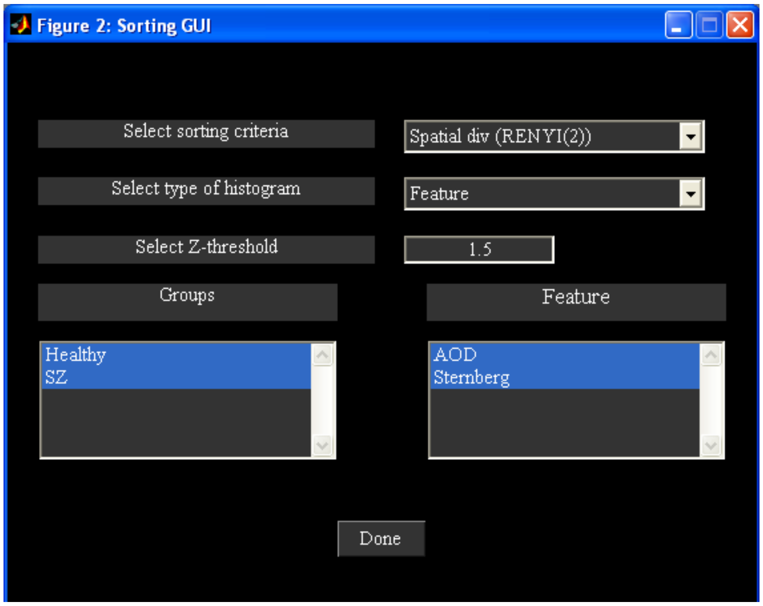


Figure 2.15: Figure shows components sorted using two sample t-test on mixing coefficients between the groups.



Select sorting criteria: Spatial div (RENYI(2))

Select type of histogram: Feature

Select Z-threshold: 1.5

Groups: Healthy, SZ

Feature: AOD, Sternberg

Figure 2.16: Sorting GUI shows the selected parameters for spatial divergence.

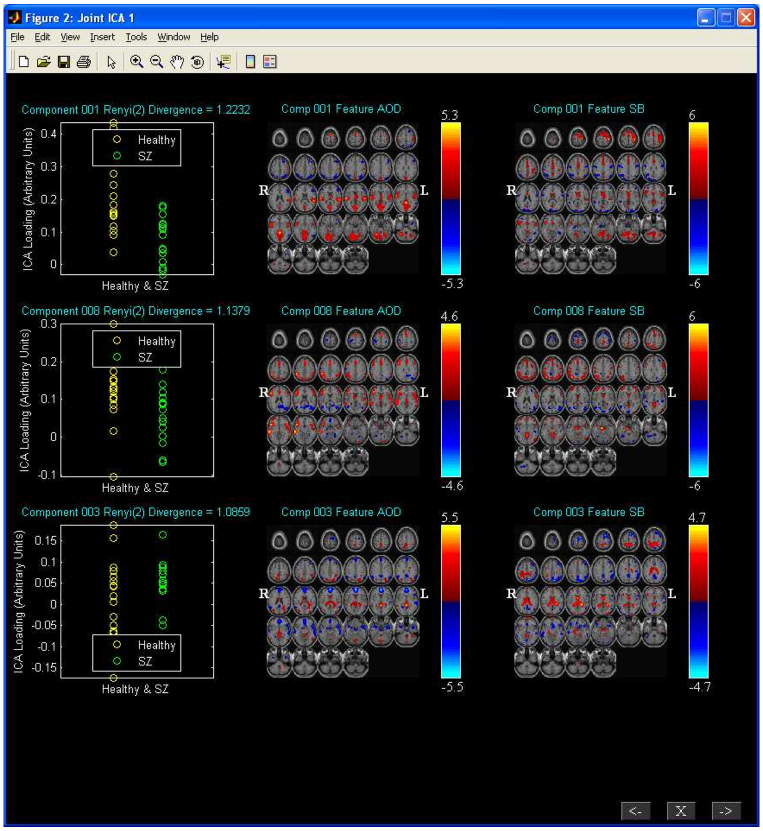


Figure 2.17: Figure shows components sorted using spatial divergence between the groups.

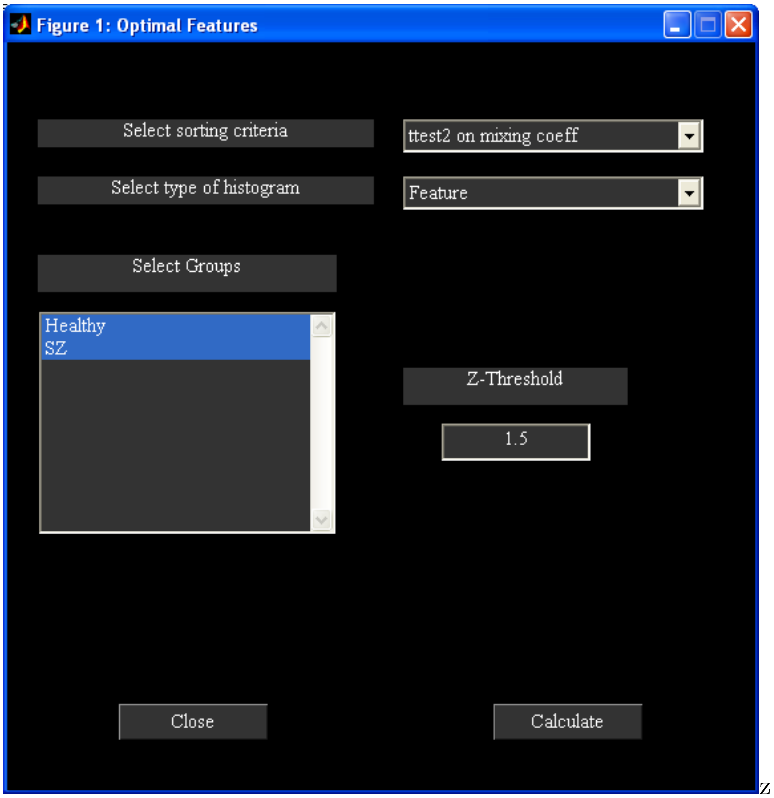


Figure 2.18: GUI for running optimal features.

- You can change the number of bins used in calculating histogram by changing variable NUM_BINS (See `ica_fuse_defaults.m`).
- You can open a spatial map or EEG signal when you double click or click on the corresponding axis.
- Option is provided to plot cross-task histograms when you use right click on the loading coefficients axis.

2.3.1.6 Utilities

"Utilities" drop down box in figure 2.1 contains options like "Optimal Features" and "Histogram Plot".

- "Optimal Features" - Before running this utility, set variable OPTIMIZE_FEATURES in `ica_fuse_defaults.m` to 'Yes' and run analysis. This will run ICA on different combinations of features in order to identify which combination of features most differentiates the groups. After you click "Optimal Features" in "Utilities" (Figure 2.1) drop down box, figure 2.18 will open. The parameters in the figure 2.18 are as follows:
  - "Select sorting criteria" - "Best" component is determined based on the sorting criteria selected. For spatial divergence sorting criteria information from "Select type of histogram" is used.
  - "Select type of histogram" - There are two options like 'Feature' and 'Component'.
    - 'Feature' - "Best" component voxels are sorted in descending order and Z-threshold is applied. This set of voxels is used as a mask to the original data. Cross-task histogram is constructed for each selected group.
    - 'Component' - "Best" component voxels are sorted in descending order and cross-task histogram is constructed for each selected group.
  - "Select Groups" - Select at most two groups. When you click *Calculate* button spatial divergence between the selected groups is calculated using the cross-task histograms. The default divergence criteria is "renyi" with $\alpha = 2$. Figure 2.19 shows the features ranked in descending order of their divergence.
  - "Select Z threshold" - Z-threshold applied on "Best" component when 'Feature' histogram is used.

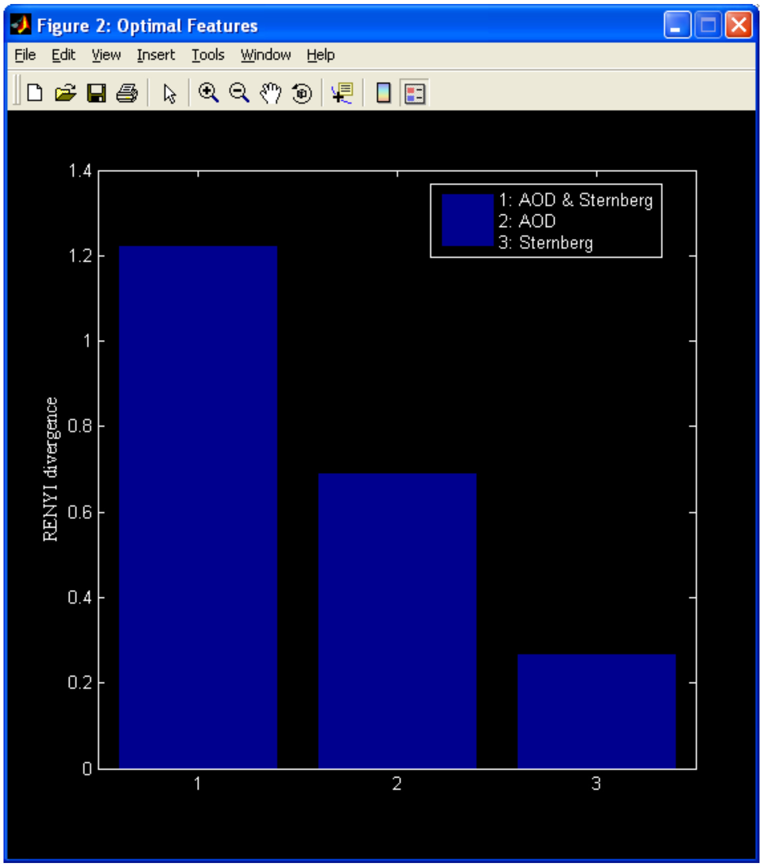


Figure 2.19: Figure shows features ranked based on spatial divergence.

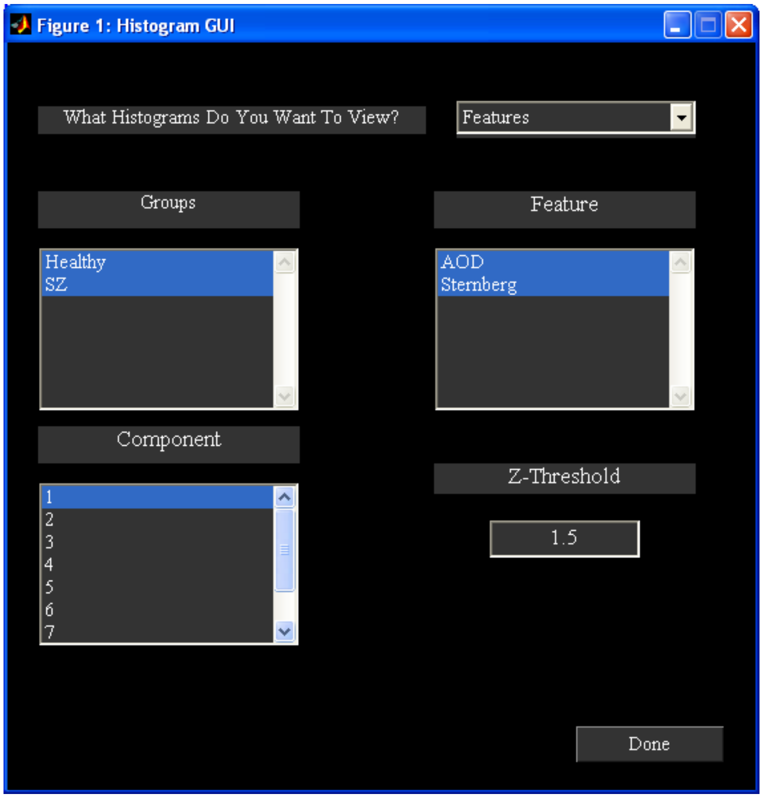


Figure 2.20: GUI for plotting cross-task histograms.

- "Histogram Plot" - Figure 2.20 will open after you have selected the fusion information file. The parameters in the figure are as follows:
  - "Groups" - Select at most two groups.
  - "Features" - Select at most two features.
  - "Component" - Select a component of interest.
  - "Z-Threshold" - Z-threshold to be applied on spatial maps.

  When you click *Done* button, Z-threshold is applied on the selected component map and voxels are sorted in descending order. This is used as a mask to the original data and cross-task histograms are generated. We calculate cross-task histogram for each subject of a group. Mean histogram is calculated for each group and a group difference histogram (Figure 2.21) is also calculated.

2.3.2 fMRI-EEG Fusion

We used fMRI-EEG example data set to do fMRI-EEG fusion. The data set contained one group and 23 subjects. We selected 12 components to be extracted from the data. Figure 2.22 shows the joint ICA components ordered based on the peak of EEG signal. Option is provided in figure 2.11 to create fMRI-EEG movie. Figure 2.23 will open when you select "Create ERP-fMRI movie" from "Utilities" menu in figure 2.11. The parameters in the figure 2.23 are as follows:

- "Enter output file name to save movie" - Movie in AVI format will be saved with this name.
- "fMRI" - Select fMRI modality.
- "EEG" - Select EEG modality.
- "Select a group" - Select a group.

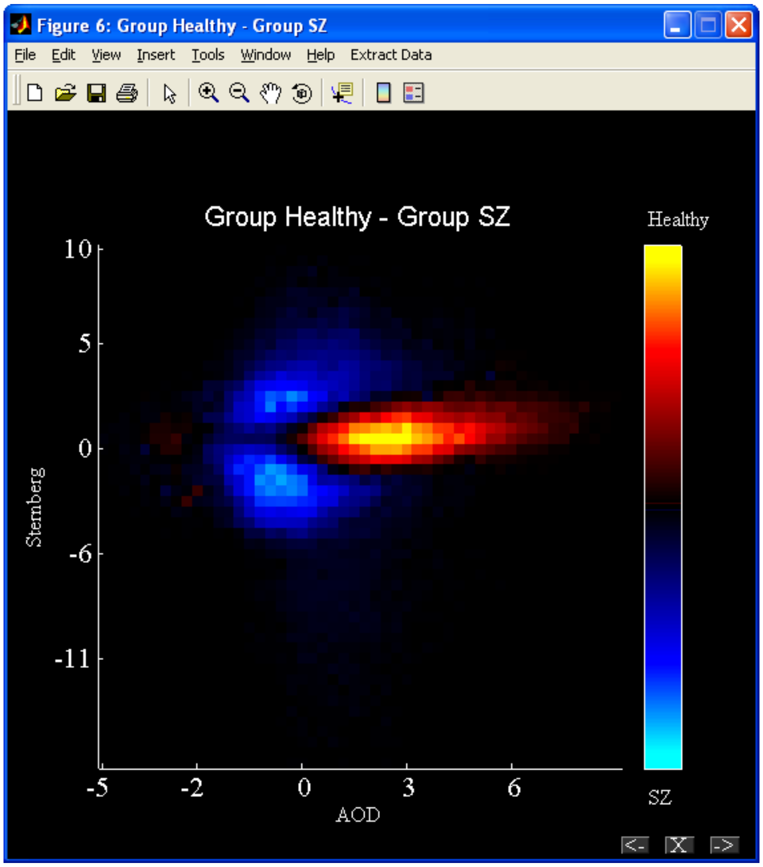


Figure 2.21: Group difference histogram.

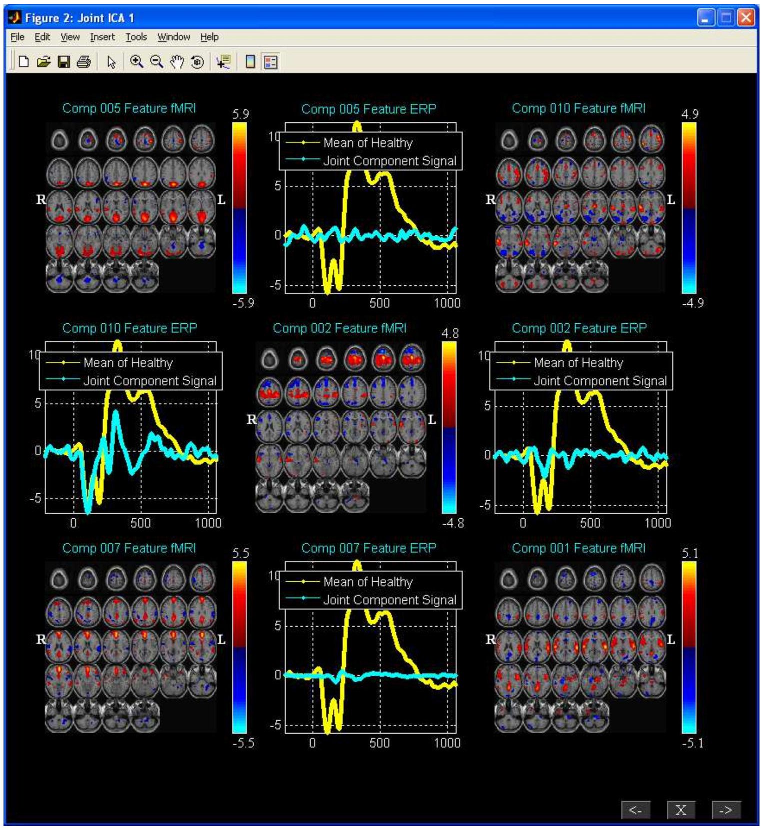


Figure 2.22: Figure shows components of fMRI-EEG fusion that are ordered based on the peak of EEG signal.

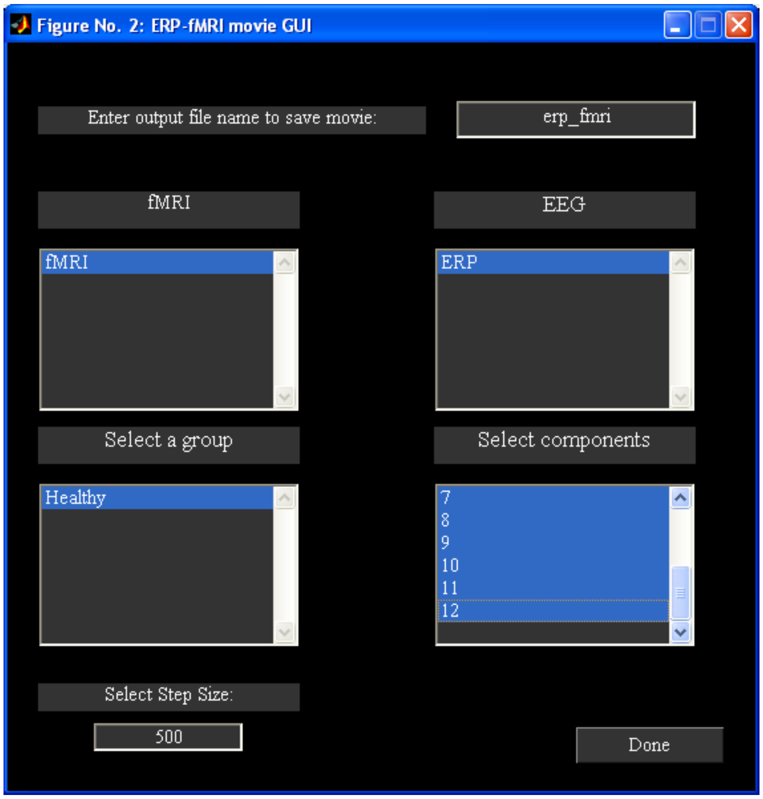
- ”Select Components” - Select the components of interest.
- ”Select Step Size” - Enter step size for time.

When you click *Done* button, weighted component data is calculated by multiplying fMRI components with the EEG components. The resulting component data will be voxels by time. Figure 2.24 shows a screen shot of ERP-fMRI movie. Spatial map is plotted on left and mean EEG signal (solid yellow line) and components EEG signals are plotted on right.

2.3.3 Batch Script

Joint ICA fusion analysis can also be done using a batch file. The syntax for the function is `ica_fuse_batch_file(inputFile)` where `inputFile` variable refers to full file path of the input file. Example input files like `input_data_fusion_1.m` and `input_data_fusion_2.m` are provided in folder `ica_fuse/ica_fuse_batch_files`. The parameters in the input file are explained below:

- `outputDir` - All the results will be stored in this directory.
- `prefix` - Output files will have this prefix.
- `maskFile` - There are two options like ’Default Mask’ and ’Select Mask’.
  - ’Default Mask’ - Default mask includes non-zero and not Nan voxels for fMRI and sMRI modalities whereas for EEG modality indices in global variable EEG_DATA_INDICES are used.
  - ’Select Mask’ - Mask must be entered in a cell array. You need to specify full file path of the image for fMRI and sMRI modalities. For EEG modality specify indices.
- `normalize` - Normalization is done for each feature. There are four options like ’Default’, ’Norm2’, ’Std’ and ’None’. Default normalization uses square root of mean of squared data for all subjects.



Enter output file name to save movie: erp_fmri

fMRI: fMRI

EEG: ERP

Select a group: Healthy

Select components:
7
8
9
10
11
12

Select Step Size: 500

Figure 2.23: GUI for creating ERP-fMRI movie

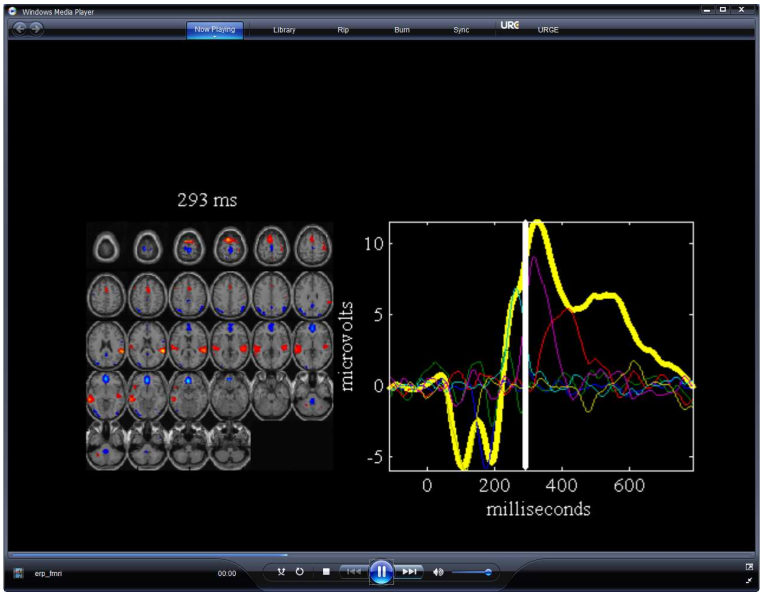


Figure 2.24: ERP-fMRI movie

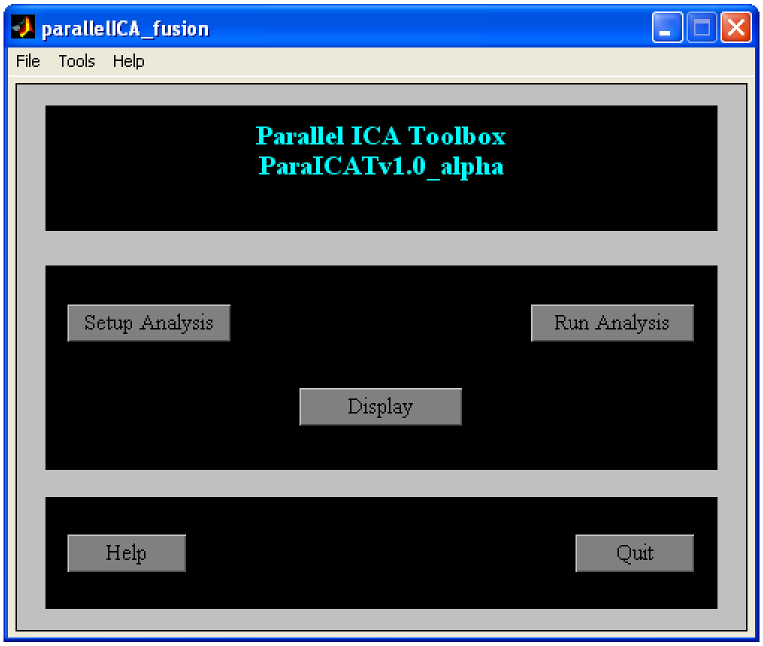
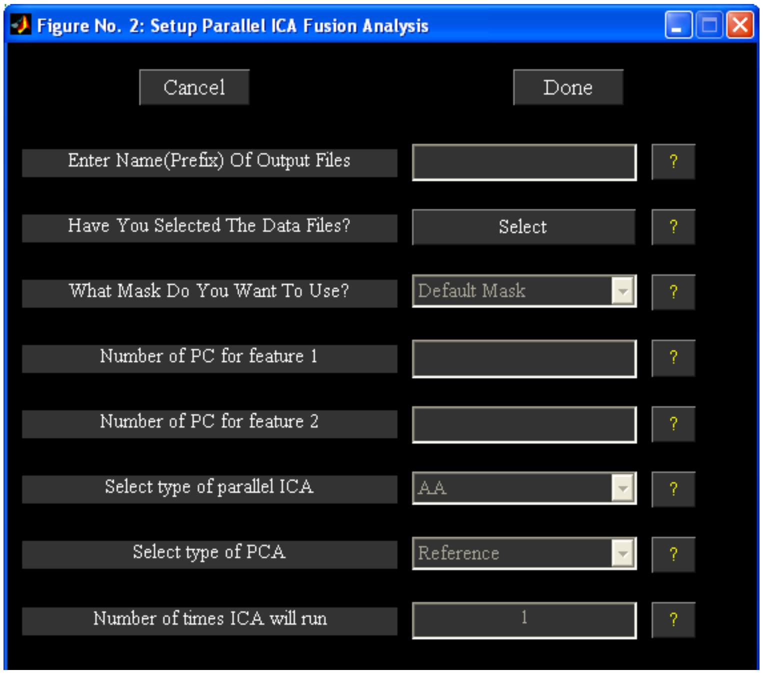


Figure 2.25: Parallel ICA Toolbox.

- `groupNames` - All the group names must be entered in a cell array. The number of groups is determined by the length of the `groupNames` vector.
- `featureNames` - Feature names must be entered in a cell array and the number of features is determined by the length of this vector.
- `modality` - The modality of each feature must be entered in a cell array. The number of modalities must equal the number of features. Presently available modalities are 'fmri', 'smri' and 'eeg'.
- `answerFilePattern` - The data is selected using the file pattern. Explanation of this variable is given below:
  - 1 - File pattern is the same between the groups.
  - 0 - File pattern is different between the groups.
- `group1_file_pattern` - File pattern for group 1. If the file pattern is same between the groups this variable will be read. Length of this vector must equal the number of features.
- `answerDir` - Value of 1 means all the data is in one folder whereas 0 means the data for each group is in a separate folder.
- `group1_dir` - Enter input directory for group 1.
- `numComp` - Number of components that can be extracted from the data.
- `algorithm` - Selected ICA algorithm. Presently, eight ICA algorithms are available. Please type `ica_fuse_icaAlgorithm` at the MATLAB command prompt for the available list of ICA algorithms.

2.4 Parallel ICA Toolbox

Parallel ICA toolbox (ParaICAT) is implemented based on parallel ICA algorithm ([1]). We explain ParaICAT using fMRI-gene data set. This data set consists of two folders like `Healthy` and `SZ`. There are 43 subjects in healthy group and 20 subjects in schizophrenics group. When you click on *Parallel ICA* button (Figure 2.1), ParaICAT (Figure 2.25) will open. You can also open ParaICAT by typing `fusion('paraICA')` at the MATLAB command prompt. The user interface controls in the figure 2.25 are *Setup Analysis*, *Run Analysis* and *Display*.



Enter Name(Prefix) Of Output Files

Have You Selected The Data Files?

What Mask Do You Want To Use? Default Mask

Number of PC for feature 1

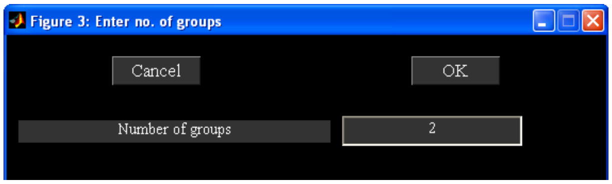
Number of PC for feature 2

Select type of parallel ICA AA

Select type of PCA Reference

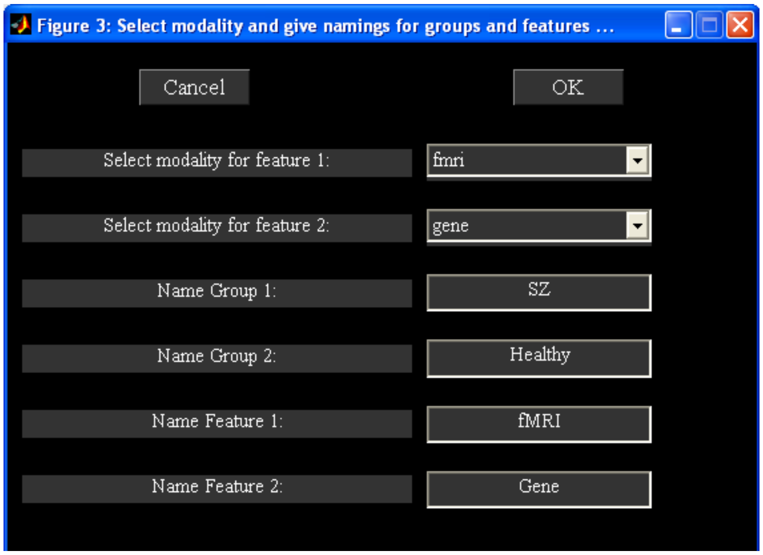
Number of times ICA will run 1

Figure 2.26: Setup ICA GUI for parallel ICA fusion.



Number of groups 2

Figure 2.27: Enter number of groups for the analysis.



Select modality for feature 1: fmri

Select modality for feature 2: gene

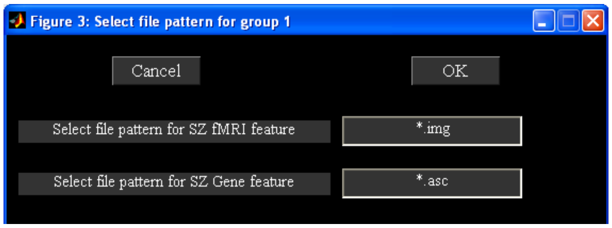
Name Group 1: SZ

Name Group 2: Healthy

Name Feature 1: fMRI

Name Feature 2: Gene

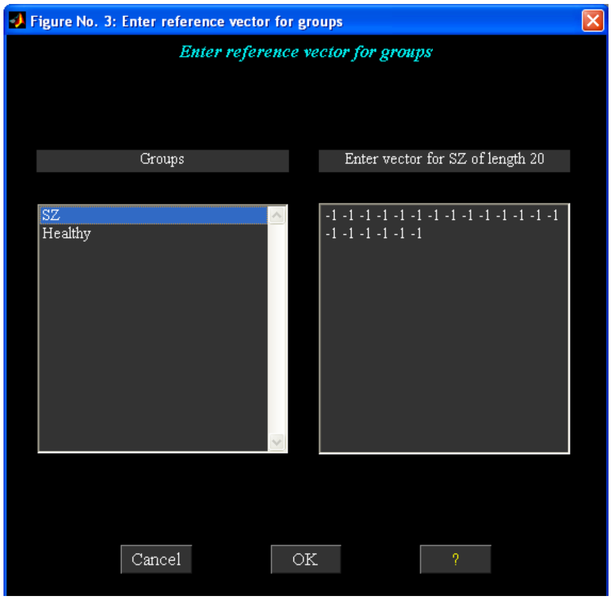
Figure 2.28: Name groups and features.



Select file pattern for SZ fMRI feature: *.img

Select file pattern for SZ Gene feature: *.asc

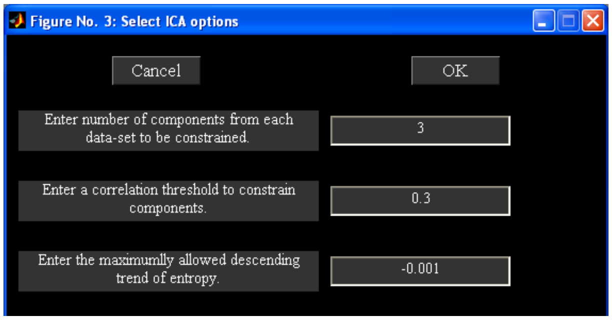
Figure 2.29: Enter filter pattern for features.



Groups: SZ, Healthy

Enter vector for SZ of length 20: -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1

Figure 2.30: Enter group information in the right edit box. Option is provided to load group information from ascii file when you use right click on edit box.

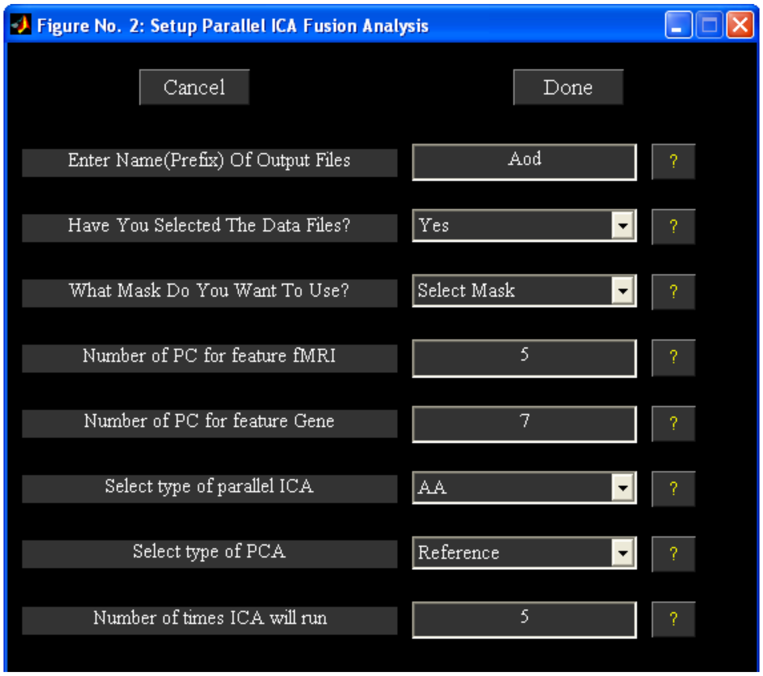


Enter number of components from each data-set to be constrained. 3

Enter a correlation threshold to constrain components. 0.3

Enter the maximumlly allowed descending trend of entropy. -0.001

Figure 2.31: Options for parallel ICA algorithm.



Enter Name(Prefix) Of Output Files	Aod
Have You Selected The Data Files?	Yes
What Mask Do You Want To Use?	Select Mask
Number of PC for feature fMRI	5
Number of PC for feature Gene	7
Select type of parallel ICA	AA
Select type of PCA	Reference
Number of times ICA will run	5

Figure 2.32: Figure window shows the completed parameters for the analysis.

2.4.1 *Setup Analysis*

When you click *Setup Analysis* button in figure 2.25 and had selected the output directory for analysis, figure 2.26 will open. The parameters in the figure are as follows:

- "Enter Name(Prefix) Of Output Files" - All the output files will be saved with this prefix.
- "Have You Selected The Data Files?" - When you click *Select* button, a figure window (Figure 2.27) will open to enter the number of groups for the analysis. Figure 2.28 will open to enter names for groups and features. For the example data set filter pattern is same between the groups, we selected "yes" for question "Is the file pattern same between the groups?" Figure 2.29 will open to enter the file patterns for features. After entering the above parameters, a figure window will open to select the directory where the data is located for groups.
- "What Mask Do You Want To Use?" - There are two options like "Default mask" and "Select Mask". The explanation for the mask is given in Section 2.3.1.1. We used "Select Mask" option and selected file `myMask_t3.img` as mask for fMRI data and all SNP indices for gene modality.
- "Number Of PC for modality fMRI" - Enter number of principal components to be extracted from the data for feature "fMRI".
- "Number Of PC for modality Gene" - Enter number of principal components to be extracted from the data for feature "Gene".
- "Select Type Of Parallel ICA" - There are two options "AA" and "AS". The "AA" option uses correlation measure between mixing coefficient of modality 1 with mixing coefficient of modality 2 whereas "AS" option uses correlation measure between mixing coefficient of modality 1 with source of modality 2.
- "Select type of PCA" - There are two options like "Reference" and "Standard". "Reference" option uses information from groups to project eigen vectors to that dimension. Please see [1] for more information. Figure 2.30 will open when you click "Reference" option. We entered a value of -1 for SZ group and a value of 1 for healthy group.
- "Number of times ICA will run" - Number of times you want ICA to be run on the data.

Figure 2.32 shows the completed parameters for the analysis. When you click *Done* button, parallel ICA options will be shown in an input dialog box (Figure 2.31). After the parallel ICA options are selected, a parameter file with suffix `para_ica_fusion.mat` will be saved in the analysis output directory.

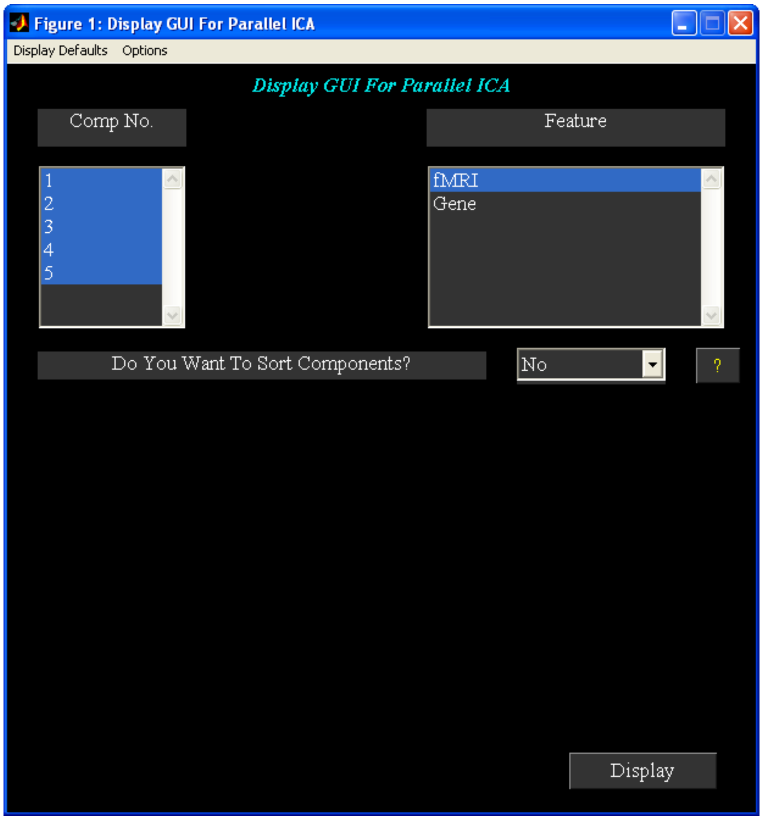


Figure 2.33: Display GUI.

2.4.2 *Run Analysis*

When you click *Run Analysis* button and had selected the parameter file (`para_ica_fusion.mat`), parallel ICA will be run. The analysis steps are as follows:

- Principal Component Analysis (PCA) - PCA is run on each modality separately. The number of components extracted for each modality depends on the number you have entered during *Setup Analysis* (Section 2.4.1). PCA information is stored in a file with suffix `para_ica_pca.mat`.

- Parallel ICA - ICA is run on each modality and a correlation measure is enforced between the modalities during the analysis. Correlation measure is dependent on the type of parallel ICA. There are two types of parallel ICA like "AA" and "AS". "AA" enforces the correlation measure between mixing coefficient of modality 1 with mixing coefficient of modality 2 whereas "AS" uses correlation measure between the mixing coefficient of modality 1 with the source of modality 2. ICA information is stored in a file with suffix `para_ica_ica.mat`.

Note: All the analysis information is stored in a log file with suffix `para_ica_results.log`.

2.4.3 *Display*

Figure 2.33 will open after you click *Display* button (Figure 2.1) and had selected the parameter file. The parameters in the display GUI are as follows:

- "Comp No:" - Component numbers to display. This is dependent on the feature you have selected in the feature listbox. This option will be disabled when you sort components.

- "Feature:" - You can select at most one feature to display. This option will be disabled when you sort components.

- "Do You Want To Sort Components?" - You have the option to sort components based on correlation. When you select "Yes" for sorting components and click *Display* button, components will be sorted based on the

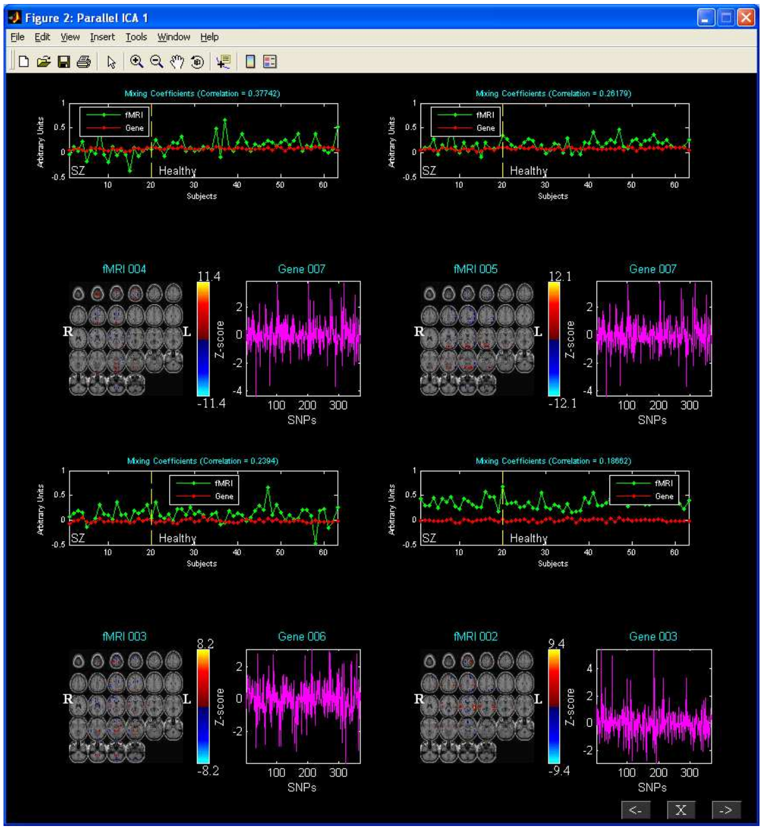


Figure 2.34: Components are sorted based on the correlation.

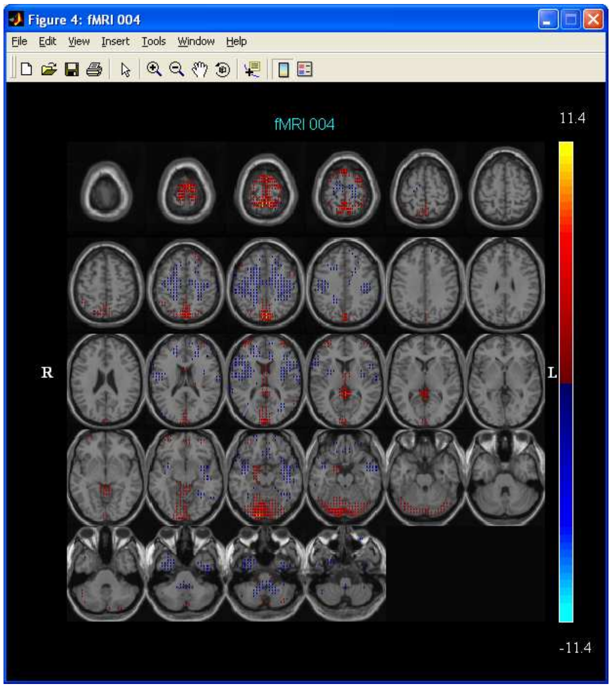


Figure 2.35: Enlarged view of a spatial map.

correlation. Figure 2.34 shows the results of sorting. When you click on any axes, enlarged view of that plot is shown in a different figure (Figure 2.35).

- ”Display Defaults” - Spatial map defaults will be shown in a figure window (Figure 2.12) when you click on ”Display Defaults” menu. The display defaults shown are the same as those explained under hidden user interface controls in Section 2.3.1.4.
- ”Options” - There is an option to select the SNP’s locus file. We have included locus file `LocusNames.txt` for the example data set. Dominant SNP’s will be printed to a text file having suffix `dominant_snps.txt` when components are sorted based on correlation. Dominant SNP’s are determined based on Z-threshold for SNP’s. You can change Z-threshold by changing variable SNP_Z_THRESHOLD in `ica_fuse_defaults.m` file.

2.5 Appendix

2.5.1 Defaults

We explain some of the defaults used in FIT. The defaults file (`ica_fuse_defaults.m`) is in folder `ica_fuse`.

- Colors: Colors are RGB values. The colors used for figure and user interface controls are as follows:
  - FIG_BG_COLOR: Figure background color.
  - FIG_FG_COLOR: Figure foreground color.
  - UI_BG_COLOR: All user interface controls except push button have the same background color.
  - UI_FG_COLOR: All user interface controls except push button have the same font color.
  - BUTTON_BG_COLOR: Push button background color.
  - BUTTON_FG_COLOR: Push button font color.
  - AX_COLOR: Axes background color.
  - LEGEND_COLOR: Legend color.
  - Font defaults are as follows:
    - UI_FONT_NAME: Font name.
    - UI_FONT_UNITS: Font units.
    - UI_FONT_SIZE: Font size.
- Display defaults: Display defaults for spatial maps are as follows:
  - CONVERT_TO_Z: Convert images to Z-scores. Options are ’Yes’ and ’No’.
  - Z_THRESHOLD: Z threshold for spatial maps. Default value is 1.5.
  - IMAGE_VALUES: Option is provided to display activations (’Positive’), activations and de-activations (’Positive’ and ’Negative’), absolute value (’Absolute’) and de-activations (’Negative’).
  - IMAGES_PER_FIGURE: Number of images plotted in one figure. Options are ’1’, ’4’ and ’9’.
  - ANATOMICAL_FILE: Anatomical file used for overlaying components.
  - ANATOMICAL_PLANE: Options are ’Axial’, ’Sagittal’ and ’Coronal’.
- OPTIMIZE_FEATURES: Option is provided to run ICA on different combinations of features. You can turn off this option by setting a value of ’No’.
- FWHM_VALUE: Smoothness factor used for dimensionality estimation.
- DIVERGENCE_PARAMETERS: Divergence criteria and number used when calculating spatial divergence. Options are ’kl’ (Kullback Liebler), ’j’ (J), ’alpha’ (Apha), ’renyi’ (Renyi). Kullback Liebler and J divergence doesn’t need a number.
- EEG_DATA_INDICES: Default mask used for EEG modality. Default value is [50:500].
- NUM_RUNS_ICA: Number of times you want ICA to be run. Default value is 1.

- STANDARDIZE_SUBJECTS: This option lets you convert feature data of each subject to Z-scores.
- Z_THRESHOLD_HISTOGRAM: Z threshold applied on the selected component when calculating histograms of features.
- NUM_BINS: Number of bins used for calculating histogram.
- FLIP_SIGN_COMPONENTS: Components will be flipped when sign of beta weight is negative with respect to a feature. You can turn off this by setting a value of 0.
- FLIP_ANALYZE_IM: Flip parameter for analyze images. Default value is 0.

Bibliography

[1] J. Liu, G. D. Pearlson, A. Windemuth, G. Ruano, N. I. Perrone-Bizzozero, and V. D. Calhoun, ”Combining fMRI and SNP Data to Investigate Connections between Brain Function and Genetics Using Parallel ICA”, HBM, December 2007.

[2] J. Liu, O. Demirci, and V. D. Calhoun, ”A Parallel Independent Component Analysis Approach to Investigate Genomic Influence on Brain Function”, In Press.

[3] V. D. Calhoun, T. Adali, K. A. Kiehl, R. Astur, J. J. Pekar and G. D. Pearlson, ”A Method for Multitask fMRI Data Fusion Applied to Schizophrenia”, HBM 2005.

[4] V. D. Calhoun and T. Adali, ”A Feature-Based Approach To Combine Functional MRI, Structural MRI and EEG Brain Imaging Data”, Proc. EMBS 2006.

[5] V. D. Calhoun, T. Adali, G. D. Pearlson and K. A. Kiehl, ”Neuronal chronometry of target detection: Fusion of hemodynamic and event-related potential data”, NeuroImage 2006.