Technical Document

FASTER v1.0

Fully Automated Statistical Thresholding for EEG artifact Reduction

Hugh Nolan

Robert Whelan

20th September 2010

Introduction

FASTER is a software suite that works in tandem with EEGLAB. *FASTER* can be used to automatically process batches of files, taking as input raw unprocessed data in .bdf or .set format, and creating as output epoched data with artifacts removed, as well as grand averages of different conditions. Artifact removal is automated, using statistical methods to remove bad channels, epoch and ICA-separated non-EEG signals.

For full details of the rejection algorithms / properties used, see Nolan, Whelan, Reilly: “FASTER: Fully Automated Statistical Thresholding for EEG artifact Reduction”, Journal of Neuroscience Methods, *192*(1), pp. 152-62, 2010.

This manual details the practical working of the user interface designed for this program.

1: Installation

Requirements for FASTER:

- MATLAB v7.4/R2007a or above (an older version *may* work, but version 7.0 has been tested and does not)
- Signal processing toolbox
- EEGLAB (has been tested with several versions from releases 6, 7 and 8).

1) Download the FASTER package from http://www.mee.tcd.ie/~neuraleng/Research/Faster?action=downloadman&upname=FASTER.zip
2) Unzip the package, and add the directory to your MATLAB path.

2: Quick Start

2.1: Checklist

1) Select directory, channel locations, and file type.
2) Set up file filter and prefix, if necessary.
3) Enter number of EEG and external channels, and select reference channel.
4) Select markers for epoching, epoch length and baseline subtraction length.
5) Enter channels for ICA, and then the EOG channels in ICA rejection options.
6) Turn on grand average and enter markers for making separate files.

2.2: Step-by-step

1) Run “FASTER_GUI” from the MATLAB command line.
2) Click the “Job Directory” button, and select the top folder in which the .bdf or .set files are located. The files can be located in subdirectories of the job directory.
3) Click the “Channel locations” button, and select the location file for the channels you wish to use. This should specify locations for the EEG channels and any external (e.g. EOG, EMG, mastoids, etc.) channels you wish to keep in the analysis. The number of channels in this file should be equal to the **number of EEG channels plus the number of External channels** specified in the channel options section.
4) Pick the appropriate file type from the drop down menu.
5) If you wish to process only files containing a certain string in the filename/filepath, enter this string in the “file filter” box.
6) If you wish to put a prefix in front of the output files (for example, to try multiple processing options on the same set of files without overwriting the previous output), enter it in the “Output file prefix” box.
7) Click the “Channels” option button. Double click the “EEG channels” option and enter the channels that are EEG channels (e.g. 1:128, or 1:64). If you are missing one or two channels, due to e.g. bad electrodes, mark them as EEG channels here and list them in the “Known bad channels” option box).
8) Double click the “External channels” options and enter the channels that are EOG, EMG, mastoids, etc (e.g. 129:132, 129:136).
9) Ensure that the reference channel number is a valid EEG channel. Fz is recommended, as this was used for evaluation of the FASTER method, but it is only an **intermediate step** for artifact removal. The final output will be in **average reference**, although the data can be re-referenced to any electrode(s) after processing.
10) If you wish the data to be epoched, click on the “Epoch” options button, and input the markers for epoching (e.g. [1 2 100]).
11) Enter the epoch length in **seconds** (e.g. [-0.5 1]).

12) Enter the baseline subtraction length (e.g. [-0.1 0], or 0 to turn off baseline subtraction).
13) Click the “ICA” options button, and ensure that the ICA channels are valid channels (unless you have a good reason, the best option is to include all EEG and External channels).
14) Open the “Rejection options” in the “ICA” options. Enter the EOG channels, if present. If not, using two frontal channels near the eyes may suffice.
15) If a grand average is desired, click the “Grand average” option button.
16) Turn on the “Make grand average” checkbox.
17) If separate averages for different epoch markers are desired, enter them in the “Grand average markers” option box.
18) If desired, press “Save job” (recommended, allows resuming of cancelled job).
19) Press “Run job”.

3: Option Descriptions

Main panel

- Job Directory: This is the directory in which all the BDF or SET files for processing are stored. Files can be in subdirectories of this. If there is more than one file for processing in a (sub)directory, the files will be moved to their own subdirectories within this, named for each file.

- Channel locations: This specifies the channel location file to allocate to the files. If the files being processed are .set, original channel locations will not be overwritten. FASTER assumes all files have the same number of channels and channel locations – if using .set files with different channel locations, proper operation isn't guaranteed.

- File type: specifies whether the files to search for are .bdf or .set files. If using the .set option to reprocess a directory, be aware that .set files in the "Intermediate" directories will not be processed, so if you want to include them, they must be moved. Also note that while processing .set files, the original .set file is saved in the "Intermediate" directory, prefixed with "Original_". Therefore if you want to reprocess a directory of .set files, you will need to move the original files.

- File filter: used to filter files within the job directory. Only files with the text specified in this box in the filename or path will be processed. Leave empty to process all files.

- Output file prefix: text in this box is prepended to the .set files produced. This could be used to run separate analyses on the same directory without overwriting the output files.

- Resume: used only for a loaded file. If set, FASTER processing will resume where the operation was cancelled last time. If the operation

finished, no new files will be processed, but a grand average may be created if enabled. Any new files added to the job directory will be ignored. Note that if the file filter is changed and resume is enabled, some files may not be processed, therefore it is better to disable resume if changing the file filter.

- Save job: saves the current job setup to a file. Default file extension is .eegjob. Depending on the operating system, the file extension may not be appended to the title automatically. If so, the file can still be loaded, but may need to be renamed manually to have the .eegjob extension.

- Load job: loads a previously saved job.

Save options

The checkboxes beside each option button allow saving of intermediate files before each step (or more exactly, before any information/noise is removed at each stage). This allows the processing options to be reviewed, and therefore changed if necessary. These files are saved in a new directory “Intermediate” which is created in the directories corresponding to each file.

- Save before filtering: The file has been opened / converted and channel locations have been added if necessary.

- Save before channel interpolation: The data has been filtered according to the filter options.

- Save before epoching: Bad channels have been detected and interpolated, if enabled.

Filter options

For more details on the meaning of the options, see the appendix.

- High pass on: turns on and off the high pass filter.

- High pass frequency: adjusts the high pass filter frequency.

- High pass options:

    - HP ripple: specifies the maximum ripple allowed in the pass band of the high pass filter.
    - HP attenuation: specifies the desired attenuation of the stop band of the high pass filter.
    - HP transition band width: specifies the width of the transition band of the high pass filter.

- Low pass on: turns on and off the low pass filter.

- Low pass frequency: adjusts the low pass filter frequency.

- Low pass options:

    - LP ripple: specifies the maximum ripple allowed in the pass band of the low pass filter.
    - LP attenuation: specifies the desired attenuation of the stop band of the low pass filter.
    - LP transition band width: specifies the width of the transition band of the low pass filter.

- Notch on: turns on and off the notch filter.

- Notch frequency: adjusts the notch filter frequency.

- Notch options:

    - Notch bandwidth: specifies the width of the stop band of the notch filter.
    - Notch ripple: specifies the maximum ripple allowed in the pass band of the notch filter.
    - Notch attenuation: specifies the desired attenuation of the stop band of the notch filter.
    - Notch transition band width: specifies the width of the transition band of the notch filter.

- Resample:

    - Resample on: turns on and off resampling. Resampling has been seen to produce filter artifacts at beginning and end of files, and where any boundary events occur, and is generally not recommended if it can be avoided. If resampling is desired, ensure that there is some non-experimental data at the start / end of the files, or that you can afford to lose some epochs. Note that due to the statistical nature of the epoch rejection algorithm, not all filter artifacts may be removed. For single-trial use, resampling is not recommended.
    - Resample frequency: New sampling frequency.

Channel options

- Reference channel: the reference channel during bad channel detection only. The EEG data in the final output is referenced to the average reference. This value should be a channel which is relatively clean. For evaluation of the FASTER method, Fz was used.

- EEG channels: the number of the channels which contain EEG data. Usually in Biosemi recordings, these will be the first 64 or 128. Normal MATLAB notation is recognised, so 1:128 specifies the first 128 channels

as EEG, 1:64 specifies the first 64, etc. Note that the channel locations file must contain the correct amount of EEG and external channels.

- External channels: the number of the channels which contain EOG, EMG, etc, data. Usually in Biosemi recordings, these will come after the EEG channels, e.g. 129:136 specifies that the 8 channels after 128 EEG channels are external channels. Note that the channel locations file must contain the correct amount of EEG and external channels. Be aware that if a channel is not selected as an EEG or external channel, it is removed from the dataset. This may have consequences for the reference channel. For example, if channels [1:63 66:128] are selected to be EEG channels and channels [129 131 134] are selected to be external channels, and channel 85 was chosen as the reference, channel 85 would correspond to the original channel 87, as two channels (64 and 65) before it have been removed. To avoid confusion in this way, it is best to keep all channels, and indicate missing channels as “known bad channels”. These will be interpolated.

- Known bad channels: these channels are known to be bad in all data sets. This could be due to missing or faulty electrodes throughout data collection. They will be interpolated, but will remain present in the output file, therefore their channel locations will not change. This is the recommended way of dealing with known missing channels – leave their locations in the channel location file, and enter the number in this box.

- Channel rejection: enables or disables the automated bad channel detection feature. Bad channels are interpolated when detected.

- Rejection options: allows you to turn on and off separate parameters of the channel rejection algorithm, and modify their statistical thresholds.

    - Channel correlation: rejects based on the channel’s mean correlation with all other channels.
    - Channel variance: rejects based on the channel’s variance.
    - Hurst exponent: rejects based on the Hurst exponent.

Epoching options

- Markers for epoching: the marker numbers (currently only numerical markers supported, but this will be augmented in future releases) that your data will be epoched on. If this is left empty, data are assumed to be single trial.

- Epoch limits: the length of the epoch, in **seconds**. It takes two values, one the minimum (usually a number of seconds before each marker, therefore a negative number) and the other a maximum (usually a number of seconds before each marker, therefore a positive number).

- Baseline subtraction: the times over which to take a mean value to subtract from the ERP. This occurs on a per-channel, per-epoch basis. This is used to line up ERPs to have a similar baseline period. This is in **seconds**, and must fit within the epoch limits.

- Epoch rejection: turns on or off the bad epoch detection feature. Bad epochs are removed from the dataset.

- Rejection options: allows you to turn on and off separate parameters of the epoch rejection algorithm, or modify their statistical thresholds.

    - Deviation from mean: rejects based on the epoch deviation from the mean value of each channel.
    - Variance: rejects based on the epoch's variance.
    - Amplitude range: rejects based on the epoch's amplitude range.

ICA options

- Run ICA: enables or disables running ICA on the dataset. The ICA algorithm used is the Infomax ICA algorithm.

- Channels for ICA: selects what channels to use for ICA. Unless you have a good reason, this should be all of the EEG channels and all of the external channels.

- ICA k value: this is a quality value, which determines how many ICA components can be output, based on the length of the dataset and the number of channels. Unless you have a good reason, the default value of 25 should work well.

- ICA component rejection: enables or disables the automatic ICA component rejection feature.

- Rejection options: allows you to turn on and off separate parameters of the ICA component rejection algorithm, or modify their statistical thresholds.

    - Median gradient: rejects based on the median gradient value. This tends to detect spurious high frequency activity.
    - Spectral slope: rejects based on the slope of the frequency spectrum in the low-pass transition band. This detects white noise components.
    - Spatial kurtosis: rejects based on the peakedness of the scalp maps which map ICA components back to scalp space. This detects artifactual activity which occurs on only one or two channels.
    - Hurst exponent: rejects based on the Hurst exponent of the ICA component time-series. This detects general non-biological signals.
    - EOG correlation: rejects based on the ICA component time-series correlation with the EOG channels. This detects EOG activity.

- EOG channels: specify any EOG channels recorded here. These will be a subset of the external channels. Alternatively, if no EOG channels have been recorded, try using one or two of the most frontal electrodes recorded – the EOG signal should still be suitably high to allow the correlation measure to detect an EOG component.

Epoch interpolation options

- Epoch interpolation: enables or disables interpolation of single channels within single epochs.

- Rejection options: allows you to turn on and off separate parameters of the epoch interpolation algorithm, or modify their statistical thresholds.

    - Median gradient: rejects based on the median gradient (higher gradient implies quick change, so this detects unusual high frequency noise) of channels within epochs.
    - Variance: rejects based on the variance (unusually high or low activity) of channels within epochs.
    - Amplitude range: rejects based on the amplitude range of channels within epochs.
    - Deviation from mean: rejects based on the deviation of channels within epochs from the whole dataset channel mean. Detects abnormal slow wave activity or uncorrected drift.

Grand average options

- Make grand average: enables or disables the grand average feature. Grand averages are saved in the job directory.

- Grand average markers: if markers are entered, separate grand averages will be made for each marker. These will be prefixed with the marker number.

- Subject removal: enables or disables automation detection and removal of bad subjects based on their ERP.

- Rejection options: allows you to turn on and off separate parameters of the subject removal algorithm, or modify their statistical thresholds.

    - Deviation from mean: rejects based on the deviation of the ERP from the mean channel values of all ERPs.
    - Variance: rejects based on the variance (unusually high or low activity) of the ERP.
    - Amplitude range: rejects based on the amplitude range of each ERP.
    - Max EOG value: rejects based on the highest value in the EOG channels per ERP.

More options

- Save as default options: saves the current option set as that which is loaded on startup of the FASTER GUI.

- Load default options: loads the default options.

4. Appendix

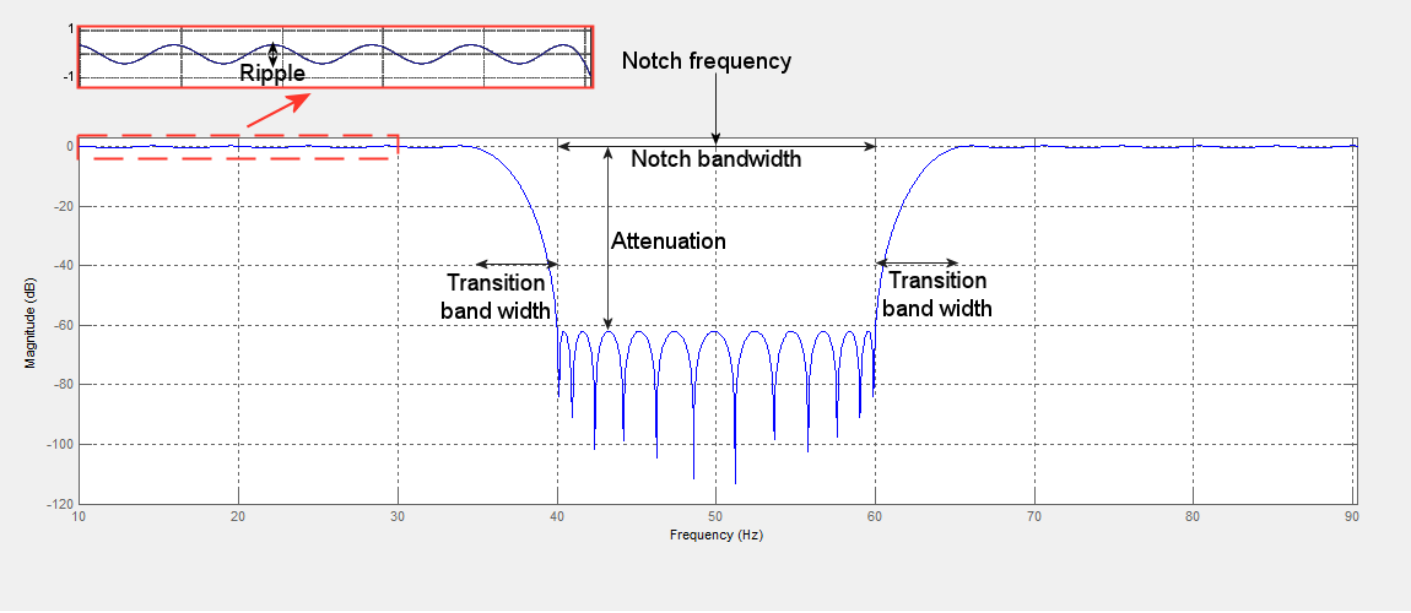

Figure 1. Notch filter design describing the different parameters available in the filter options section. High- and low-pass filters work similarly.

- Regarding filtering, Figure 1 depicts a notch filter design, exaggerated for clarity. The design features are noted on the figure. High- and low-pass filters share the same design options excepting the notch bandwidth, so their operation can be extrapolated from this diagram. If you are unsure about a parameter, the default options give sensible values. Adjust with care, as trying to use extreme values (very small transition band, very low frequencies, very high attenuation, etc.) may not be possible to design, and so the output from the filtering process may not be stable. **This will not be detected as an error.**
    - Notch frequency: the centre frequency of the notch filter.
    - Notch bandwidth: how wide the notch band is (corner frequencies are therefore centre frequency – bandwidth/2 and centre frequency + bandwidth/2).
    - Ripple: how much change there is in the passband. This should be small.
    - Attenuation: the minimum that the frequencies in the notch band are attenuated (then may be attenuated more, as can be seen in the ripple).

    - Transition band width: how wide the transition between the notch band and the passband is.

- Regarding the statistical thresholding, the threshold which can be changed for each measure is a Z threshold, i.e. a measure of standard deviation from the mean. Increasing the threshold means that less data will be rejected and decreasing means that more data will be rejected. The standard value of 3Z was used for all testing, however other values may give better results for different recording paradigms / EEG setups.