

DataQuality is a MATLAB tool for checking the quality of fMRI data. The tool is provided by New York University Center for Brain Imaging and implemented by University of Missouri-Columbia Brain Imaging Center to run *automatically* on every fMRI data file transferred to CTP server. A data quality directory will be created and located at the same directory of your fMRI raw data. For a detailed description of the tool, check the website at: <http://cbi.nyu.edu/software/dataQuality.php>

1. Purpose

The tool examines the quality of functional data. It checks the spatial and temporal statistics of the voxel intensities in a series of regions-of-interest (ROIs) automatically defined by the tool, and generates an html report with the quantitative quality indicators (Signal-to-Noise Ratio, Signal-to-Fluctuation-Noise Ratio and Signal-to-Ghost Ratio; defined below). It also reports the results of spike detection and plots of the average image intensity time-course for these ROIs. A function to correct for any spikes detected in the data is also included

2. Technical Description of the analysis

The code loops through the images, reads the data and calculates the spatial mean and standard deviation —for each volume— of the signal in a set of ROIs, which are positioned in the object, the Nyquist ghost and in the background. The location of the ROIs can be calculated automatically by another function (also provided with the package), based on the image dimensions and the phase encoding (PE) direction. These positions can be modified manually by the user.

After obtaining the temporal series of the spatial means and standard deviations for each ROI, a series of statistical indicators are computed that give the user a quantitative index of the data quality:

- 1) **SNR₀**: Signal-to-Noise Ratio. It takes as “signal” the average voxel intensity in all the ROIs defined in the object, averaged across time. It takes as “noise” the standard deviation (across space) of the signal in the ROIs defined in the background, and then averaged across time:

$$SNR_0 = \frac{\langle \langle \text{Subject} \rangle \rangle_t}{1.53 \times \langle [\text{stdx}(\text{Sbackground})] \rangle_t}$$

The 1.53 factor is used to correct for the fact that we took the absolute value of a complex image with 0 mean.

- 2) **SFNR**: Signal-to-Fluctuation-Noise Ratio. It takes as “signal” the same as for the SNR_0 above. It takes as “fluctuation noise” the (temporal) standard deviation of the (spatial) mean in the same ROIs, after the slow drift (quadratic) has been removed from the temporal series.

$$SFNR = \frac{\langle \langle \text{Subject} \rangle x \rangle t}{stdt(\langle \text{Subject} \rangle x)}$$

- 3) **SGR**: Signal-to-Ghost Ratio. It takes as “signal” the average voxel intensity from an ROI inside the object, and then averaged across time. It takes as “ghost” the average voxel intensity from an ROI in the ghost, corresponding to the previous ROI in the object (that is, half a field of view away), and then averaged across time:

$$SGR = \frac{\langle \langle \text{Subject-linked-to-ghost} \rangle x \rangle t}{\langle \langle \text{Sghost} \rangle x \rangle t}$$

Then, it plots the signal time-courses for each ROI group, in which you can check visually for the quality of your data. Next, it searches for spikes in the signal intensity in a set of ROIs in the background. Finally, it writes an html report file which includes some information about the scan, the figures, the statistical indicators and the spike detection results. If any spikes had been detected during the analysis, this tool saves the repetition numbers and the spiky slices in a file, allowing the user to correct the data with an additional function provided in this tool.

3. Interpreting the html report

3.1 Open the html report on CTP server (example to Frey05 data)

ssh -X -Y username@ctp.psyc.missouri.edu and type your university pawprint.

```
cd
```

```
/data/test/storage/dicomserver/dir/Frey/05/subID/1.3.../dataquality/qReport_.../
```

```
firefox --no-remote qReport_2015....a001.html
```

3.2 Always check the ROI location: if an ROI happens to be too close to the edge of the head, even a little bit of motion could artificially introduce fluctuations in the ROI signal, decreasing the SNR numbers, but with no real effect in the actual quality of the data for our functional analysis. Again, check the image with the ROI locations to see if the intensity is zero or not; or, even better if you are not sure, stay away from the slice edges.

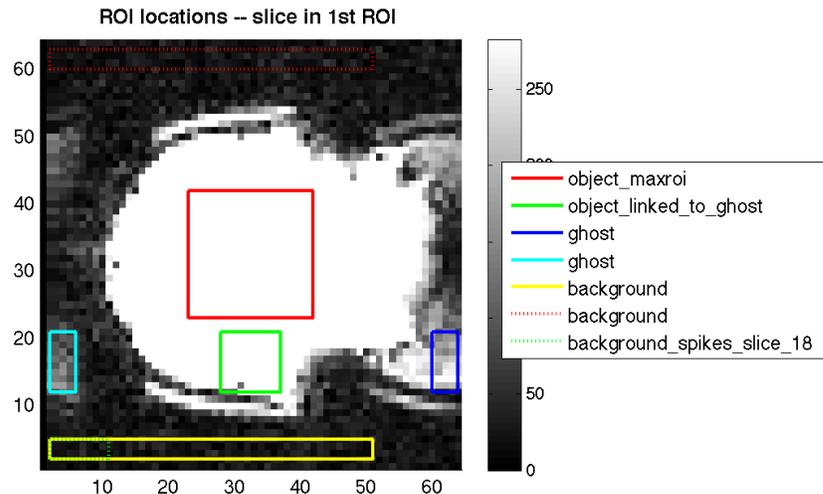


Figure 1: Example of the default locations of the ROIs in the central slice from a representative run (the image intensity has been windowed so that the ghost shows up clearly). This type of plot allows the user to check whether the default placement of the ROIs is correct or if you need to move some of them.

3.3 If the script doesn't print any spike-candidates, one can be sure there are no spikes in the data. If it does print some, the user has to look at the plots of the specific slice/time-point in the data (which will be printed in the report) to see what the problem is. Because the script picked that point, there will be some background irregularities. You have to see how big those irregularities are compared to your background signal and compared to the normal fluctuations in the time-course of the object ROIs: it is possible that they are insignificantly small.

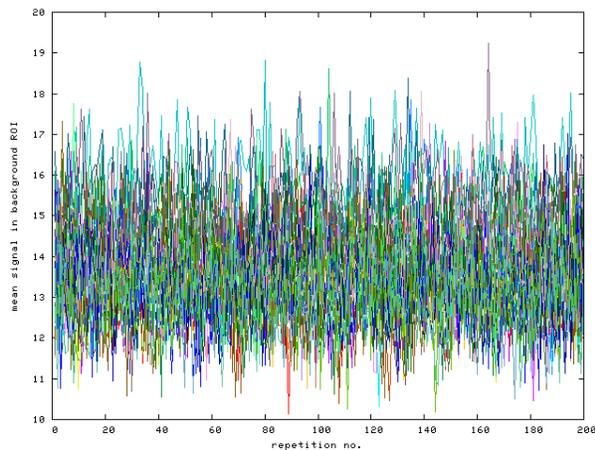


Figure 2: Example of the time-course in the spikes ROIs for a normal run. Note how the noise fluctuation size is very constant across time and the baseline is similar for all ROIs.

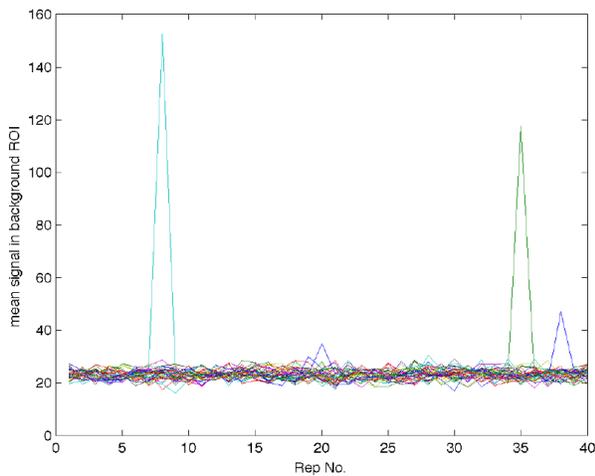


Figure 3: Example of the time-course in the spikes ROIs for a "spiky" run. There are three peaks standing up above the baseline level.

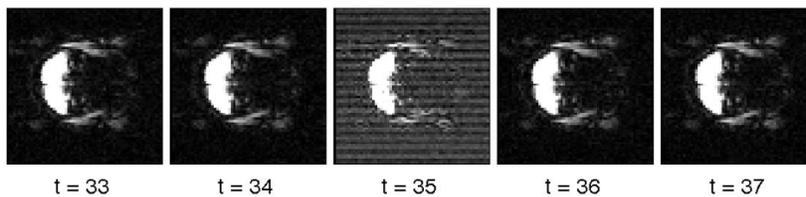


Figure 4: Images of the slice showing a peak in the intensity at time-point 35 in Fig. 3, and the same slice for the two time points before and after. Note the corduroy pattern across the image, typical of spikes.

4. Discussion

This tool calculates several statistical indicators that address different aspects of data quality. Each has its own relevance, and it is very important to monitor their values across different runs within the same session and across different sessions and subjects. Note that the specific values will depend on many things, like where you place your ROIs (especially when using surface coils), reconstruction type (online vs. offline), the length of the run, the bandwidth, the specific task, TR, TE, FA, etc. After analyzing several runs, one can get an idea of what values to expect for a given imaging protocol, and get a feeling of what constitutes a “bad run.”

In general, one should pay special attention to the values reported for **SFNR**. There are several contributions to the temporal intensity fluctuations at any given voxel: the functional activity itself, the electronics noise (mostly white noise from the receiver), physiological changes (respiration and heart beat), scanner noise (B0 drift due to heating ...), etc. Because the goal of checking the data quality is to know how good our data are going to be for the functional analysis, one wants to make sure that the fluctuations not due to functional activity are as small as possible. Therefore, one wants to estimate the size of the temporal noise (temporal fluctuations in the signal) with respect to the average temporal signal using a large number of voxels. For each time point, using the average intensity in a large number of voxels will average the BOLD contributions from different functional parts of the brain, so the variations due to functional activity over time will be significantly washed out, giving only an average functional contribution. Most of the electronics white noise will be averaged out as well, since its contribution is spatially independent. This way, the main source of temporal variation of the average intensity will come from the scanner noise (which is what we want to estimate) and from the physiological activity (which can be recorded and subtracted from each voxel time-course, improving the statistical power of the functional analysis of the data). Therefore, SFNR —which is the indicator tuned to measure this aspect of the noise— is probably the most relevant indicator of the quality of our functional data.

The 1.53 factor in Eq. 1 is based on the assumption that the noise distribution in the background is Rayleigh. This is only valid for data collected using a single coil. If

the data were collected using a different number of coils, the factor would be different. But, for purposes of SNR_0 comparisons across runs and across subjects, it doesn't make any difference.

For questions regarding your report, please consult Dr. Huiling Peng, BIC MRI Physicist.