



Published in final edited form as:

Magn Reson Med. 2011 April ; 65(4): 1053–1061. doi:10.1002/mrm.22691.

A Novel Method for Quantifying Scanner Instability in fMRI

Douglas N. Greve^{1,*}, Bryon A. Mueller², Thomas Liu³, Jessica A. Turner⁴, James Voyvodic⁵, Elizabeth Yetter⁶, Michele Diaz⁵, Gregory McCarthy⁵, Stuart Wallace¹, Brian J. Roach⁶, Judy M. Ford⁶, Daniel H. Mathalon⁶, Vince D. Calhoun⁷, Cynthia G. Wible⁸, Stephen G. Potkin⁴, Gary Glover⁹, and The FBIRN¹⁰

¹ Martinos Center for Biomedical Imaging, 143 13th Street, Charlestown, MA, USA

² Department of Psychiatry, University of Minnesota, Minneapolis, MN 55455, USA

³ Center for Functional Magnetic Resonance Imaging and Department of Radiology, University of California San Diego, La Jolla CA, USA

⁴ Department of Psychiatry and Human Behavior, University of California Irvine, Irvine, California

⁵ Duke-UNC Brain Imaging and Analysis Center, Duke University Medical Center, Durham, North Carolina

⁶ Department of Psychiatry, Yale University, West Haven, CT

⁷ Mind Research Network & Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM Department of Psychiatry, Yale University

⁸ Department of Psychiatry, Harvard Medical School and Brockton VAMC, Boston, MA

⁹ Department of Radiology, Stanford University, Stanford, CA 94305, USA

¹⁰ FBIRN—Functional Testbed-Biomedical Informatics Research Network, NIH-NCRR, Bethesda, MD 87131, USA

Abstract

A method was developed to quantify the effect of scanner instability on fMRI data by comparing the instability noise to endogenous noise present when scanning a human. The instability noise was computed from agar phantom data collected with two flip angles, allowing for a separation of the instability from the background noise. This method was used on human data collected at four 3T scanners, allowing the physiological noise level to be extracted from the data. In a “well-operating” scanner, the instability noise is generally less than 10% of physiological noise in white matter and only about 2% of physiological noise in cortex. This indicates that instability in a well-operating scanner adds very little noise to fMRI results. This new method allows researchers to make informed decisions about the maximum instability level a scanner can have before it is taken off line for maintenance or rejected from a multisite consortium. This method also provides information about the background noise, which is generally larger in magnitude than the instability noise.

Keywords

quality assurance; fMRI; multi-site; scanner instability

*Corresponding Author: greve@nmr.mgh.harvard.edu.

Introduction

Minimizing image noise is always important for MRI, but is especially critical in functional MRI (fMRI) because the Blood Oxygen Level Dependent (BOLD) contrast is often extremely small, and multiple imaging measurements must be made over time to track cognitive processes. It is therefore crucial to separate and clarify the contributions of the various sources of noise from the scanner and the subject. Researchers are generally interested in characterizing and tracking noise and instability for fMRI because fMRI pulse sequences tend to push a scanner to design specification limits, which may exacerbate instability. This applies to diffusion-weighted imaging (DWI) and arterial spin labeling (ASL) sequences as well. The purpose of this research is to assess the impact of scanner-related noise on fMRI analysis for the purposes of establishing benchmarks of scanner performance.

The noise in fMRI images has components that are either additive or multiplicative. Additive noise is independent of NMR signal and includes thermal contributions from the subject or phantom, electronics, “spike” noise, and spurious RF interference signals injected through failure of the magnet room’s RF shielding. Additive noise is not encoded by the gradients and can therefore be spread across the image, including outside the object being scanned. For this reason it is often referred to as “background noise”, even though it may occur in both the background and the foreground. Even when the additive noise comes solely from the object being scanned, the noise amplitude will be influenced by scanner hardware (e.g., the coil being used), so even this noise can be thought of as having a scanner-dependent component.

Multiplicative noise is a fluctuation in NMR signal intensity and can therefore only appear in the image where there is NMR signal. Multiplicative noise can have several sources. One source is from the scanner in that k-space trajectories, RF pulses, and slice select pulses are not played out in exactly the same way from time point to time point due to fluctuations in, for example, resistive shim currents or gradient, RF, or receiver amplifier gain or phase. These types of instrument-induced fluctuations are known as “scanner instability” (1). Some instability is inevitable even in a well-operating scanner. When scanning a human, additional multiplicative noise results from modulation of signal by physiological processes including bulk head motion, cardiovascular pulsatility in the brain, respiration (2) and from neural processes thought to give evidence for resting state networks(3). Physiological noise has properties similar to scanner instability noise, namely a complicated spatio-temporal correlation structure (2,4). In this work we do not attempt to characterize the structure of instability or physiological noise as we are primarily interested in the magnitude of the noise and not its spatial/temporal correlation.

Thus, both additive and multiplicative noise sources arise from both subject and scanner. Whatever the source, noise will tend to impair the ability to draw conclusions from the data. While the noise from a subject will be inevitable, we can attempt to control the noise from scanner through an appropriate quality assurance (QA) protocol. This paper introduces a new method to quantify scanner noise based on the expected effect on fMRI analysis, with the final metric relating to the extra scan time needed to reverse the effect of having the extra scanner-induced noise (and so concomitant increase in scanner costs, subject fatigue, etc). The method requires a measure of the background noise variance which we compute from two series acquired with different flip angles. We acquire such data on phantom and human subjects to differentiate between instability noise, background noise, and physiological noise. With these data the total variance can be decomposed into three components: background, instability, and physiological, allowing a comparison of the

scanner-induced noise contribution to the total noise. This forms the basis of a new scanner performance metric that can be used for QA or to compare scanners.

Theory

As mentioned above, when scanning a static phantom, there are two main sources of noise: background and instability. Scanning a human introduces a third type of noise due to physiological processes. Background, instability and physiological noise are temporally independent, so their variances will add linearly. The total temporal noise variance σ_α^2 at flip angle α is given by:

$$\sigma_\alpha^2 = \sigma_{p,\alpha}^2 + \sigma_{i,\alpha}^2 + \sigma_{bg}^2, \quad (1)$$

where $\sigma_{p,\alpha}^2$ is the physiological variance at flip angle α , $\sigma_{i,\alpha}^2$ is the scanner instability variance at flip angle α and σ_{bg}^2 is the background noise variance (independent of flip angle). Signal-to-fluctuation-noise (SFNR) is defined as the mean intensity divided by the standard deviation of the total noise (5). We will define several new SFNR measures below (see also Table 1 for a summary).

We model the physiological and instability fluctuations as *signal-weighted* (SW), i.e., they are multiplicative fluctuations in signal and therefore variances are proportional to the square of the signal intensity (2). These variances can therefore be grouped together into a single component: $\sigma_{sw,\alpha}^2 = \sigma_{p,\alpha}^2 + \sigma_{i,\alpha}^2$, with the total variance (Equation 1) re-expressed as

$$\sigma_\alpha^2 = \sigma_{sw,\alpha}^2 + \sigma_{bg}^2. \quad (2)$$

We can measure the total variance from the time series, but we also need the background variance in order to compute the signal-weighted variance. Some proposals ((6),(7),(8)) suggest measuring the variance in the actual background, e.g., in the corners of the image or in other areas away from locations where EPI artifacts are suspected. There are several problems with this approach. First, it is difficult to find areas where no artifact signal is deposited due to ghosting and ramp sampling or non-Cartesian k-space trajectories such as spiral. Second, the variance of the noise as it appears in the background is not the same as that for when it appears in the foreground due to the complex-to-magnitude operation. For a single channel coil, the noise is Rayleigh distributed, and it is easy to convert the variance to foreground variance by dividing by $2-\pi/2$ (9). For multiple-channel coils, there is no simple conversion factor. Methods exist (10–12) to compute the background noise variance in multiple coils using a 0° flip angle scan (i.e., no RF power), but these methods require special pulse sequences, access to k-space data, and/or assume that there is no noise correlation across channels. Our method avoids these issues and may be implemented on any scanner.

The definition of signal weighted is that the variance of the SW component will simply scale when going from flip angle α_1 to α_2 , i.e.,

$$\sigma_{sw,\alpha_2}^2 = \frac{\sigma_{sw,\alpha_1}^2}{M^2}, \quad M = \frac{\mu_{\alpha_1}}{\mu_{\alpha_2}}, \quad (3)$$

where μ_{α_i} is the mean intensity at flip angle α_i . Given two acquisitions at different flip angles, the mean intensities μ_{α_i} and total variances $\sigma_{\alpha_i}^2$ can be measured empirically for each data set. This gives us three equations (Equation 2 for each flip angle, plus Equations 3) with three unknowns ($\sigma_{sw\alpha_1}^2, \sigma_{sw\alpha_2}^2, \sigma_{bg}^2$). Substituting and solving for the SW and background components, we get:

$$\sigma_{sw\alpha_1}^2 = \frac{M^2(\sigma_{\alpha_1}^2 - \sigma_{\alpha_2}^2)}{M^2 - 1} \quad (4)$$

$$\sigma_{bg}^2 = \frac{M^2\sigma_{\alpha_2}^2 - \sigma_{\alpha_1}^2}{M^2 - 1} \quad (5)$$

These variances can then be used to compute two new SFNR measures that can be used to assess scanner quality:

$$swSFNR = \frac{\mu_{\alpha_1}}{\sigma_{sw\alpha_1}} \quad (6)$$

$$bgSFNR = \frac{\mu_{\alpha_1}}{\sigma_{bg}}, \quad (7)$$

where $swSFNR$ is the signal weighted component, and $bgSFNR$ is the background component. When scanning a phantom, only instability and background noise are present, and so the $swSFNR$ will only represent instability noise. When scanning a human subject, all three components (physiological, instability, and background) will be present, and the physiological and instability components cannot be separated. However, we can estimate the amount of instability in human scans using the instability in the phantom scan because the instability is proportional to image intensity, so

$$\hat{\sigma}_{I\alpha} = \frac{\mu_{\alpha,H}}{iSFNR} \quad (8)$$

where $iSFNR$ is the instability SFNR measured from the phantom scan, $\mu_{\alpha,H}$ is the voxel-wise mean intensity of the human scan, and $\hat{\sigma}_{I\alpha}$ is the voxel-wise estimate of the instability standard deviation in the human scan. As the total variance, the instability variance, and the background variance are now estimated, Equation 1 can be used to solve for the estimate of the physiological variance $\hat{\sigma}_{P,\alpha}^2$. This yields three new SFNR measures: $bgSFNR$ and $iSFNR$ (both defined above), and $pSFNR = \mu_{\alpha,H} / \hat{\sigma}_{P,\alpha}$, the physiological SFNR.

In an fMRI study, these three sources of noise are independent of any task, so the increase of any source can be reversed by an increase in scanning time proportional to the increase in total variance that it causes. For example, if instability accounted for 10% of the total noise, one would need to scan 10% longer (with concomitant increase in scanner costs, the number stimuli needed, and subject fatigue). Alternatively, one could think of the scanner as running at 90% capacity. While this metric provides no absolute cutoff for scanner acceptance, it provides a clear link between instability and scanner performance, and researchers will be well prepared to set a threshold based on their own comfort level. For this study, we quantify the instability noise with respect to only the physiological noise (i.e., assume the

background noise is, or can be made, negligible). This gives a more conservative measure and also simplifies the quantification because the increase in total noise variance can be expressed as the square of the ratio of iSFNR to pSFNR.

Methods

This method was applied to traveling subject data collected at four MRI sites as part of the Functional Biomedical Informatics Research Network (fBIRN) test bed. We identify these sites as Site03, Site05, Site06, and Site18. All were 3T scanners; two were General Electric (GE) and two were Siemens. See Tables 2 and 3 for the site-independent and -dependent fMRI acquisition parameters.

Human Data Analysis

Over the course of six months (April-October, 2007), the same eighteen subjects were each scanned at all four sites in this study. This study had approval from the institutional review boards (IRB) at each collection site. Each subject was scanned a total of 5 times. The first scan was always at Site18. The remaining 4 scans were randomized across site (including revisiting Site18). Each visit consisted of several acquisitions including: anatomical, task fMRI, B0 map, arterial spin labeling (ASL), and two rest fMRI scans with different flip angles. For this work, we utilized only the two rest fMRI scans and only the anatomical images collected at Site06. The rest scans were at the end of a long protocol and, in some visits, the rest scans were not collected due to time issues. Consequently, only 10 subjects had complete rest data sets for all 5 visits. In September 2007, the Site03 scanner had a large instability due to a faulty transient noise suppressor. The two subjects collected at this time were excluded from our normative data set, leaving only eight subjects.

The anatomicals (MP-RAGE: TR=2300ms, TE=2.94ms, TI=1100ms, Flip Angle=9°, .86×.86×1.2mm³) were analyzed in FreeSurfer (13,14) to automatically generate *subject-specific* cortical ((15), (16)) and subcortical (17) regions-of-interest (ROIs), including cortex, cortical white matter (WM), putamen (Put), pallidum (Pal), hippocampus (Hip), and amygdala (Amyg). ROIs for cortex were also generated. The WM ROI was eroded by 3mm in all directions to reduce partial voluming with other tissue types. The cortex was divided into two ROIs, an “inner” cortex (ICtx) and an “outer” cortex (OCtx), where the outer cortex was defined as any cortical voxel within 5mm of the edge of the brain. Cortical regions known to be affected by EPI B0 distortion (inferior and medial temporal gyri, orbital cortex, (18)) were excluded from the cortex ROIs.

The resting state data were acquired at two flip angles, $\alpha_1=77^\circ$ and $\alpha_2=10^\circ$. The full set of fMRI acquisition parameters across all sites is given in Tables 2 and 3. To the largest extent possible, these parameters were matched across site. Note that all sites used phased array receive coils (three sites used 8-channel coils and one site used a 12-channel coil). The data from the separate coils was always combined using the root-sum-of-squares (RSS) method (no acceleration was used). The 77 ° acquisition was chosen because 77 ° is the Ernst angle of gray matter (T1=1350ms; (19)) at 3T for TR=2s and is the flip angle used in our task fMRI acquisitions. The choice of the low flip angle was driven by two considerations. First, the best fit is achieved when the intensity difference between the high and low flip angle acquisitions is maximized, arguing for a scan with as low a flip angle as is possible. However, if the SFNR becomes too low the noise in the foreground will become non-Gaussian (e.g., Rayleigh or Rician), resulting in space-variant background noise and possibly invalidating the Equations 4 and 5. This issue is of particular concern with multi channel coils combined with RSS where each coil has its own SFNR which can change substantially over space. Fortunately, the SFNR requirements to assure a Gaussian noise distribution in the image foreground are actually quite modest. The SFNR can be interpreted

as a z-value, so the probability that the noise will exceed the mean signal can be computed based on the z distribution. For example, for an SFNR of only 2 the noise will exceed the mean signal only once in every 44 samples. At 10°, the raw SFNR for the phantom data always exceeded 40. Even when taking into account the loss of SFNR over a coil profile, it is still very unlikely that enough non-Gaussian noise would have been present to skew the results, at least with the 8 or 12 channel coils and acquisition parameters used in this experiment.

The resting state data were motion corrected to the middle time point (20). No slice-timing correction was applied. The time series at each voxel was analyzed with a general linear model (GLM) (21) in which a 2nd order polynomial plus motion correction parameters modeled the time-series; no temporal whitening was used. The regression coefficient corresponding to the constant in the 2nd order polynomial was used as the estimate of the mean (μ) at each voxel. The residual variance was used as the estimate of the total noise variance (σ^2) at the voxel. The middle time point image volume was registered to the anatomical volume using a 6 degrees-of-freedom (DOF) linear transform (22), and the mean and variance maps were resampled using nearest-neighbor interpolation into the anatomical (1mm³) space where they were averaged within each ROI. This procedure generates a measure of the mean and total variance for each flip angle from which we compute the background and SW variance for each subject at each site for each ROI.

Agar Phantom Data Analysis

Over the course of the study, each site collected phantom data (Site03: N=25; Site05: N=12, Site06: N=18, Site18: N=42) using the same protocol as in the human fMRI. The phantom used at each site was a sphere with diameter 17.5cm filled with agar designed to have comparable T₁ and T₂ to that of gray matter (23). These data were analyzed using a 3D mask of the phantom, which was eroded by five voxels resulting in a spherical ROI approximately 13cm in diameter. The agar phantom was analyzed in the same way as the human data with the exception that the time series was not motion corrected¹, and so no motion correction parameters were used in the regression. This provides a measure of the mean and total variance from which to compute the background and SW noise for the agar phantom for each site. Agar phantom scans from Site03 during the time of scanner instability were excluded.

Other Quality Assurance (QA) Measures—For comparison purposes, we also computed other stability measures for each agar scan, namely R_{dc}, FWHM, and Percent Fluctuation. R_{dc}, the “radius of decorrelation”, was proposed by Friedman, et al, 2006 (23), based on Weisskoff, 1996 (1), who suggested plotting the temporal variance of the waveform averaged over an ROI against the number of voxels in the ROI. For pure white noise, the variance will drop linearly with the number of voxels (N) in the average. For spatially correlated (e.g., instability) noise, the variance reduction will be less than linear, and for large N the variance will reach an asymptotic “floor” level. The R_{dc} is the value of N at which the variance reduction reaches the asymptotic floor. Similarly, Friedman, et al, 2006 (23), suggested using a measure of the spatial full-width/half-maximum (FWHM) computed from the correlation of neighboring voxels as a QA metric, with larger FWHM being indicative of more instability (5,23,24). A third measure is the Percent Fluctuation, defined as the time series standard deviation measured in a large ROI after second order detrending of drift.

¹No motion correction was used because we only look at the center of the phantom away from the edges and our scanners did not exhibit much B0 drift. However others should be aware that drift may require motion correction.

Quality Control and Problem Data Sets

Since the goal is to provide normative data to compare scanner performance, the data used in the analysis needed to be free of artifact. For the agar data, this was achieved by visually looking for outliers in plots of agar SFNRs over time. The instability in the Site03 data mentioned above was discovered by visual inspection of the SFNR time plots. In addition, some agar data sets from Site18 were found to have elevated background noise. These data sets were also excluded from the analysis. When the human data from Site18 were examined more closely, it was found that, while all subjects were similar in physiological noise, there was clear bimodal distribution of background noise variance, with 7/16 data sets (four subjects) having elevated background noise. Although the source of this noise has not been discovered, these data were kept in the analysis to prevent the results from being based on only four subjects. Also, the task data from these subjects have been analyzed, and the results are as expected for this working memory paradigm (results will be shown elsewhere). For these reasons, we believe that the data presented herein represent that from well-operating scanners (except for the elevated background noise in some of the Site18 data).

Proof-of-Concept—to show that our technique actually measures background noise, we scanned an agar phantom at flip angles of 0° , 10° , and 77° using a single-channel coil at Site06. At 0° , the data will be pure background noise but with a Rayleigh distribution. The variance of the original Gaussian noise can be calculated by dividing the Rayleigh variance by $2-\pi/2$ (about .429). This is then used as the gold standard to compare against the background variance computed from our method using the 10° and 77° scans. Single channel coils were used because of the ease with which the background noise can be computed from a single channel 0° flip angle scan. The single channel proof-of-concept should extend to multiple coils combined with RSS except when the noise in an individual channel becomes non-Gaussian. As mentioned earlier, this is a very low probability event for the acquisition parameters used in this study.

Results

Proof-of-Concept

the difference between the background variance computed from the 0° scan and that computed using our method averaged over 12 acquisitions on single channel coils was only 2.3% \pm 2.2%, with 10 out of 12 differing by less than 1%. This shows that the method is reasonably accurate.

Signal-Weighted Noise

the mean swSFNR for the agar phantom and for the ROIs averaged across all sites and all 8 subjects are shown in Figure 1A and B. As expected, WM has the highest swSFNR because it has the least physiological noise (2). Cortex has about 40% of the swSFNR of WM. Throughout most of the brain structures the swSFNR is very similar across site and across the two visits at Site18. This is expected because physiological noise is a property of the subject, not the scanner. Very little site effect is observed in the data, a testament to the repeatability of the proposed noise metric across scanners for this study. We tested the human swSFNR for an effect of site using repeated measures ANOVA. When corrected for 8 comparisons, only one ROI showed a site effect (amygdala, $p=10^{-4}$). White matter, cortex, hippocampus, pallidum, caudate, and putamen showed no significant effect of site. The swSFNR for the agar (Figure 1B) is at least 3 times greater than for the white matter meaning that the instability noise variance is about 10% of the physiological noise in white matter. For cortical gray, the physiological noise variance is about 50 times that of the instability. This indicates that the swSFNR in Figure 1A is very close to the pSFNR. When averaged over all sites and subjects, the pSFNR was 289.3 for white matter and 101.5 for

inner cortex. For all agar scans, the average iSFNR was 1330.2. Researchers looking to use this method to judge their scanner could measure the iSFNR in a phantom and then compare it to the pSFNR values listed above. Note, however, that the estimation of pSFNR is based on limited data from these 8 subjects scanned at these 4 sites, so care must be taken when comparing the pSFNR to the iSFNR, and more studies of this type are needed.

Background Noise

Figure 2 shows the mean background SFNR for the human ROIs (Figure 2A) and agar (Figure 2B), grouped by site instead of structure. Figure 2A shows there is very little variation across brain region. This is expected because most background noise is not spatially encoded and the multiple coils are combined using simple RSS, meaning that it should be uniform across the image. This uniformity property does not hold for “spike” or coherent RF leakage noise or for some k-space reconstruction methods that can impart a spatial pattern on the noise variance (e.g., adaptive combine (25), SENSE (26), or GRAPPA (27)). There appears to be a strong site and manufacturer effect, with the GE sites having larger bgSFNR for both human and agar. However, this is at least partially due to a manufacturer difference between the slice select methods. In this study, each site used the default slice select method for the scanner manufacturer. By default, Siemens scanners use a fat saturation pulse followed by a broadband slice selective excitation pulse while GE scanners use a water-only excitation pulse. Such pulses are two-dimensional in nature (spatial and spectral) and necessitate tradeoffs in spatial selection performance for compactness relative to the fat-saturation/broadband scheme. As a result, water-only excitation pulses generate wider spatial sidebands than the broadband pulse for the same selected slice thickness, which results in a larger effective slice thickness and a larger signal. The background noise magnitude is not affected by signal, so the net result is a larger bgSFNR with the (GE) water excite slice select pulse than with the (Siemens) broadband slice select pulse. We had originally thought that the different levels of background noise were caused by distortion correction being turned on in the GE sites (Table 2). Further investigation of this showed that this had a trivial effect on the agar phantom background noise. Site18 has the worst performance on the human data, partially due to a rogue source of background noise (mentioned above). For the agar results, Site18 was similar to Site06 because agar scans with elevated background noise were removed.

Variance Composition

Table 4 shows the proportion of total variance at 77° for each type of noise for each site for white matter and cortical gray matter averaged over the 8 subjects. In all sites for both tissue types, the instability noise represents a very small proportion of the total noise. The worst case for white matter is only 4%, and only 1% for cortex. This is different than the 10% value give above because that value was based on a more conservative comparison to physiological noise alone (i.e., no background noise). In contrast, the background noise can contribute substantial amounts (40–90% in white matter and 10–50% in cortex). The background noise can be reduced by spatial smoothing (e.g., the variance will reduce by a factor of 2.3 for 5mm of smoothing). Even after smoothing, background noise will likely dominate over instability noise. This proportion may also change with changes in voxel resolution, because, as voxels become smaller, background noise fraction will increase. Note that there is a strong site effect in the background noise due to two factors. First, the two GE sites have a lower proportion of background noise due to the slice select issue mentioned above. Second, the difference between the two Siemens sites (with Site18 having much more background noise) is just a result of the unexplained elevated levels of background noise that contaminated some of the scans from Site18.

Comparison with other QA Measures

Figure 3 shows four other QA measures plotted against the corresponding iSFNR for the agar phantom. The vertical lines indicate the iSFNR that would result in instability noise with 10% of the *variance* of physiological noise in either white matter or cortex (i.e. $\sqrt{10}$ times the iSFNR). Figures 3A and 3B show the FWHM in the readout (RO) and phase encode (PE) directions, respectively. Higher FWHM tends to correspond to lower iSFNR, but the correspondence is not strong. It is difficult to determine where one would set a FWHM threshold to trigger scanner maintenance. Even at the worst FWHM=4.5mm (RO), the instability noise is still more than 10 times lower than that in cortex. As one would expect, the R_{dc} (Figure 3C) tends to increase with iSFNR. The Percent Fluctuation (Figure 3D) is inversely related to the iSFNR as one would expect because Percent Fluctuation is essentially a noise-to-signal measure.

Discussion

The impact of scanner instability noise will manifest itself when one attempts to draw conclusions from fMRI data. False negatives will increase because the extra noise will reduce the size of the test statistic. This can be mitigated by increasing the amount of data collected proportional to the square of the total SFNR change, with a proportional increase in stimulus presentations, cost for scanner time, and subject fatigue or other task performance issues. In our results, instability caused a very small increase in total variance of about 4% in white matter and about 1% in cortex (Table 4). We emphasize that the instability is independent of any task and all brain processes, and so it will not bias the amplitude of the hemodynamic response; it will only increase its variance. At a group analysis level, the impact of instability noise will be even less because its proportion of total variance will drop with the addition of inter-subject variance. If an effect size (i.e., z-ratio, t-ratio, F-ratio) is passed to the higher level, then first-level noise will have an unpredictable effect, which is one of the reasons why the hemodynamic response amplitude is preferred. While the first-level variances are used in a mixed-effects group analysis, they are only used as weighting factors or as a means to better estimate the higher level variance (28) and so do not systematically bias the results at the group level. This brings up the important point that different levels of noise from different sites must be taken into account in multi-site group analysis (using mixed effects or weighted least squares mentioned above) as it creates subject-specific noise levels in the data which can skew the p-values.

The impact of background noise on fMRI analysis is similar to the impact of instability noise discussed above. Non-artifact background noise is spatially uncorrelated, so it can be easily decreased with spatial smoothing. In our results, the background noise contribution was much more substantial than the instability noise though usually much smaller than the physiological noise. There was a manufacturer effect in the bgSFNR, but this was apparently due to differences in signal caused by differences in the slice-select methods. Background noise is often seen as inevitable because most of it emerges from the sample being scanned. However, there can be contributions from scanner hardware and environment, and it is possible for these to change over time. This was observed in the Site18 human and agar data where the background variance increased by more than a factor of two over time.

Other QA metrics designed to detect instability (FWHM, R_{dc} , and Percent Fluctuation) are clearly related to Instability SFNR (Figure 3), though they also carry information about background noise as well. These measures have typically been tracked longitudinally at a single site. Over time, the site would accumulate baseline measures against which changes could be compared. Our critique of this is that the longitudinal measures do not relate clearly to the fMRI results, so it is hard to judge what number of standard deviations should be used

as the threshold. Nevertheless, longitudinal QA testing can be quite sensitive in early detection of scanner problems and should be considered an essential ingredient in any QA protocol.

To implement this method, a site would need to collect data on an agar phantom at two flip angles and apply the equations given in the methods section. This method does not require any elaborate configuration of the scanner or access to special data (e.g., raw k-space data). The agar phantom analysis itself is quite straightforward and can easily be implemented with custom software or using a third-party software package (e.g., FSL (www.fmrib.ox.ac.uk/fsl) or AFNI (afni.nimh.nih.gov/afni)). The iSFNR and bgSFNR can then be compared to normative values such as those given in this manuscript. It is not clear how parameter changes to the fMRI protocol will alter the pSFNR measure, so more human data may need to be acquired using the new protocol if the protocol is going to be changed. We also note that it is probably not necessary to use a particular phantom to measure the instability since the composition of the phantom will only affect the mean intensity, and the instability will simply scale with it. This would, however, affect the ability to compare the bgSFNR measured on the phantom.

The QA metrics described herein were computed from acquisitions using a high and low flip angle. In this study, the RSS method was used to combine the multiple receive coils to reconstruct the images. For human data, the ROIs were both large (e.g., white matter) and small (e.g., amygdala). For the agar data, the ROI was a (relatively large) 13 cm sphere. One limitation in this method is that if the noise variance changes over the ROI, then the method may yield inaccurate results. This can happen with the RSS coil combine method if the low flip angle scan noise distribution is significantly non-Gaussian due to low SFNR in an individual coil. We argue that the SFNR has to become extremely low (below 2) for this to have a significant impact on the results and that this is very unlikely to be the case for the acquisition parameters used in this work. However, low SFNR conditions may be met in acquisitions with a substantially smaller voxel size than used here or in scans using a multi-channel coil with many more elements. Other methods to combine data from multiple coils exist. The adaptive combine method (25) performs a weighting of the complex coil images. If there is no spatial variation in the weighting, then the method of extracting noise variances from scans with two flip angles should be valid. If there is spatial weighting, then the noise will also be space-variant. Parallel imaging (e.g., (26) (27)) changes the demands made on the scanner hardware and parallel imaging reconstruction methods will likely impart spatial variation in the noise variance. In these cases, the use of smaller ROIs may be required. It is worth noting that a two flip angle acquisition is not required to compute the SFNR measures developed in this manuscript. Any method that determines the background variance can be used to compute these SFNR measures (e.g. (11,12)), although the two flip angle acquisition can be acquired on virtually any scanner, is simple to analyze, and should be valid for most fMRI acquisitions.

Conclusions

The purpose of this study was to measure the relative contribution of three noise sources (instability, background, and physiological) on the total temporal fMRI variance to establish non-arbitrary scanner performance benchmarks. We have developed a simple method to separate instability and physiological noise from background noise by acquiring fMRI time series at two flip angles. No special pulse sequences or access to k-space data are needed. Such data were acquired on both humans and agar gel phantoms at four 3T sites (two GE, two Siemens) as part of the Functional Biomedical Informatics Research Network (fBIRN) test bed. For the human data, noise summaries were computed for various ROIs customized to each subject. These data show that the physiological SFNR (pSFNR) measurements are

repeatable across visit, site, and manufacturer. There are two primary innovations in this work. First, we quantify the instability based on the amount of extra scanning time needed to mitigate the effect of added noise. Using our fMRI protocol, the iSFNR was easily 3 times that of the pSFNR in white matter, meaning that instability effects could conservatively be mitigated by acquiring about 10% more data. Interestingly, the background noise contributed much more noise than instability in the well-operating scanners of this study. Second, we advocate comparing the instability SFNR (iSFNR) at one scanner to the scanner-independent pSFNR rather than to the iSFNR at another scanner. Both of these innovations can help scanner administrators and researchers to make informed decisions about the maximum instability level a scanner can have before it is taken off line for maintenance or rejected from a multisite consortium.

Acknowledgments

Support for this research was provided in part by the National Center for Research Resources (P41-RR14075, R01 RR16594, P41-009874, the NCRR BIRN Morphometric Project BIRN002, and Functional Imaging Biomedical Informatics Research Network (FBIRN) U24 RR021382), the National Institute for Biomedical Imaging and Bioengineering (R01 EB001550, R01EB006758), as well as by the Department of Energy (DE-F02-99ER62764-A012) to the Mind Research Network (previously known as the MIND Institute). The members of the fBIRN project all deserve acknowledgement for their significant efforts, but unfortunately, they are too numerous to mention. Please visit <http://www.nbirn.net> for more information regarding key personnel. We would also like to thank Thomas Benner for his help in collecting data for this manuscript.

References

1. Weisskoff RM. Simple measurement of scanner stability for functional NMR imaging of activation in the brain. *Magn Reson Med*. 1996; 36(4):643–645. [PubMed: 8892220]
2. Kruger G, Glover GH. Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magn Reson Med*. 2001; 46(4):631–637. [PubMed: 11590638]
3. Biswal, B.; DeYoe, EA.; Cox, RW.; Hyde, JS. fMRI Analysis for Aperiodic Task Activation Using NonParametric Statistics. New York, New York: SMR, Berkeley, CA; 1994. p. 624
4. Triantafyllou C, Hoge RD, Krueger G, Wiggins CJ, Potthast A, Wiggins GC, Wald LL. Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. *Neuroimage*. 2005; 26(1):243–250. [PubMed: 15862224]
5. Friedman L, Glover GH. Reducing interscanner variability of activation in a multicenter fMRI study: controlling for signal-to-fluctuation-noise-ratio (SFNR) differences. *Neuroimage*. 2006b; 33(2):471–481. [PubMed: 16952468]
6. Sijbers J, den Dekker AJ, Van Audekerke J, Verhoye M, Van Dyck D. Estimation of the noise in magnitude MR images. *Magn Reson Imaging*. 1998; 16(1):87–90. [PubMed: 9436952]
7. Sim KS, Lai MA, Tso CP, Teo CC. Single Image Signal-to-Noise Ratio Estimation for Magnetic Resonance Images. *Journal of Medical Systems*. 2009
8. Stocker T, Schneider F, Klein M, Habel U, Kellermann T, Zilles K, Shah NJ. Automated quality assurance routines for fMRI data applied to a multicenter study. *Hum Brain Mapp*. 2005; 25(2): 237–246. [PubMed: 15846770]
9. Papoulis, A. Probability, random variables and stochastic processes. New York: McGraw-Hill; 1984.
10. Constantinides CD, Atalar E, McVeigh ER. Signal-to-noise measurements in magnitude images from NMR phased arrays. *Magn Reson Med*. 1997; 38(5):852–857. [PubMed: 9358462]
11. Kellman P, McVeigh ER. Image reconstruction in SNR units: a general method for SNR measurement. *Magn Reson Med*. 2005; 54(6):1439–1447. [PubMed: 16261576]
12. Robson PM, Grant AK, Madhuranthakam AJ, Lattanzi R, Sodickson DK, McKenzie CA. Comprehensive quantification of signal-to-noise ratio and g-factor for image-based and k-space-based parallel imaging reconstructions. *Magn Reson Med*. 2008; 60(4):895–907. [PubMed: 18816810]

13. Dale AM, Fischl B, Sereno MI. Cortical Surface-Based Analysis I: Segmentation and Surface Reconstruction. *NeuroImage*. 1999; 9:179–194. [PubMed: 9931268]
14. Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*. 1999; 9(2):195–207. [PubMed: 9931269]
15. Fischl B, van der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat D, Busa E, Seidman L, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale A. Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*. 2003 in press.
16. Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 2006; 31(3):968–980. [PubMed: 16530430]
17. Fischl B, Salat DH, Albert M, Dieterich M, Haselgrove C, Kouwe Avd, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*. 2002; 33(3):341–355. [PubMed: 11832223]
18. Jezzard P, Balaban RS. Correction for geometric distortion in echo planar images from Bo field variations. *Magn Reson Med*. 1995; 34:65–73. [PubMed: 7674900]
19. Wansapura JP, Holland SK, Dunn RS, Ball WS Jr. NMR relaxation times in the human brain at 3.0 tesla. *J Magn Reson Imaging*. 1999; 9(4):531–538. [PubMed: 10232510]
20. Cox R, Jesmanowicz A. Real-time 3D image registration for functional MRI. *Magn Reson Imaging*. 1999; 42:1014–1018.
21. Friston KJ, Jezzard P, Turner R. Analysis of functional MRI time-series. *Human Brain Mapping*. 1994; 1:153–171.
22. Greve DN, Fischl B. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*. 2009; 48:63–72. [PubMed: 19573611]
23. Friedman L, Glover GH. Report on a multicenter fMRI quality assurance protocol. *J Magn Reson Imaging*. 2006a; 23(6):827–839. [PubMed: 16649196]
24. Friedman L, Glover GH, Krenz D, Magnotta V. Reducing inter-scanner variability of activation in a multicenter fMRI study: role of smoothness equalization. *Neuroimage*. 2006; 32(4):1656–1668. [PubMed: 16875843]
25. Walsh DO, Gmitro AF, Marcellin MW. Adaptive reconstruction of phased array MR imagery. *Magn Reson Med*. 2000; 43(5):682–690. [PubMed: 10800033]
26. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med*. 1999; 42(5):952–962. [PubMed: 10542355]
27. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med*. 2002; 47(6):1202–1210. [PubMed: 12111967]
28. Beckmann C, Jenkinson M, Smith S. General multi-level linear modelling for group analysis in fMRI. *Neuroimage*. 2003; 20:1052–1063. [PubMed: 14568475]

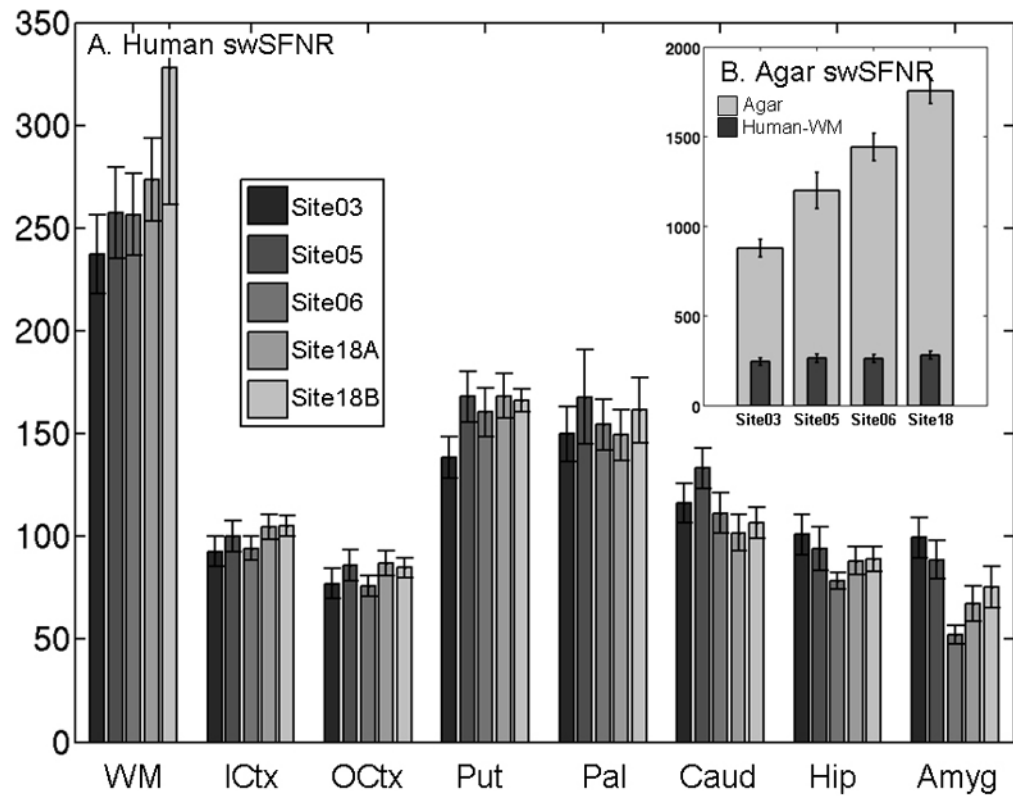


Figure 1. (A) Human Signal-Weighted SFNR for each site sorted by ROI. Site18A and Site18B refer to the two visits at Site18. (B) Agar Instability SFNR. The vertical scale is different than Panel A, so the Human WM swSFNR is replicated from Panel A for reference. Confidence intervals are 1 standard error measured across subjects.

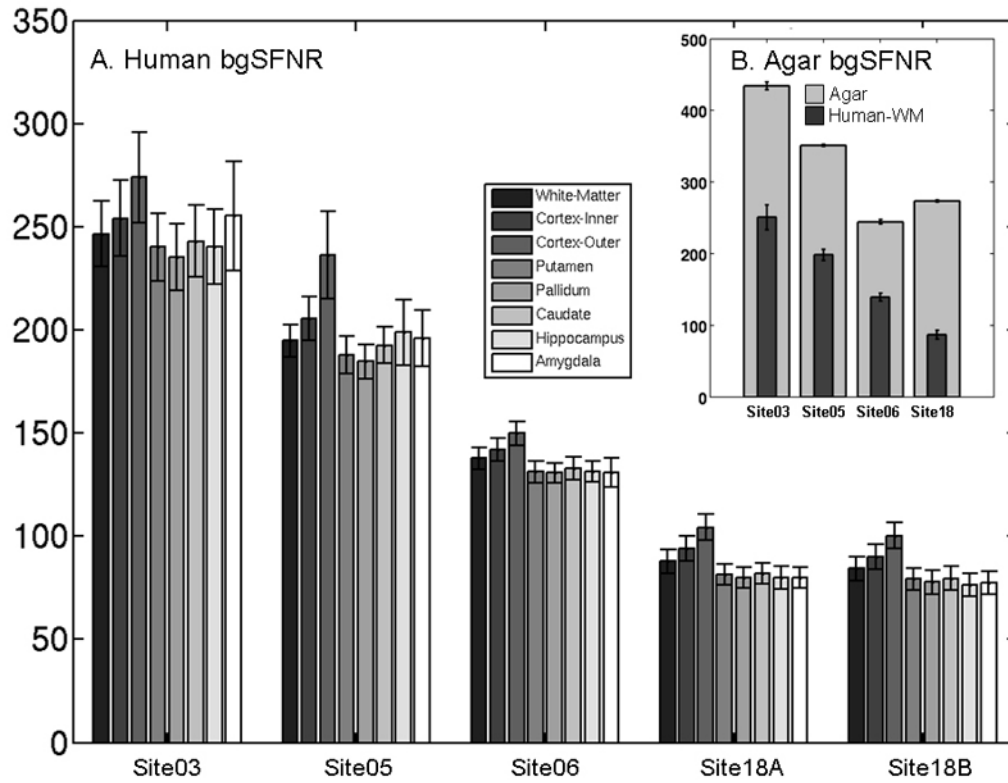


Figure 2.

Background SFNR for (A) the human ROIs sorted by site, and (B) the agar phantom. The vertical scale is different than Panel A, so the Human WM background SFNR is replicated from Panel A for reference. Confidence intervals are 1 standard error measured across subject. Note that some of the difference in the background SFNR performance between sites is caused by a difference in the default slice select mechanisms between Siemens and GE scanners. The agar and human background SFNRs are hard to compare because the background noise generally depends on the size of the object being scanned and its ionic conductivity. Note that rank of Site18 changes between the human and agar data. This is because Site18 was affected by a sporadic source of background noise, and affected agar scans were removed but affected human scans were kept in order to maintain a reasonable number of subjects in the study.

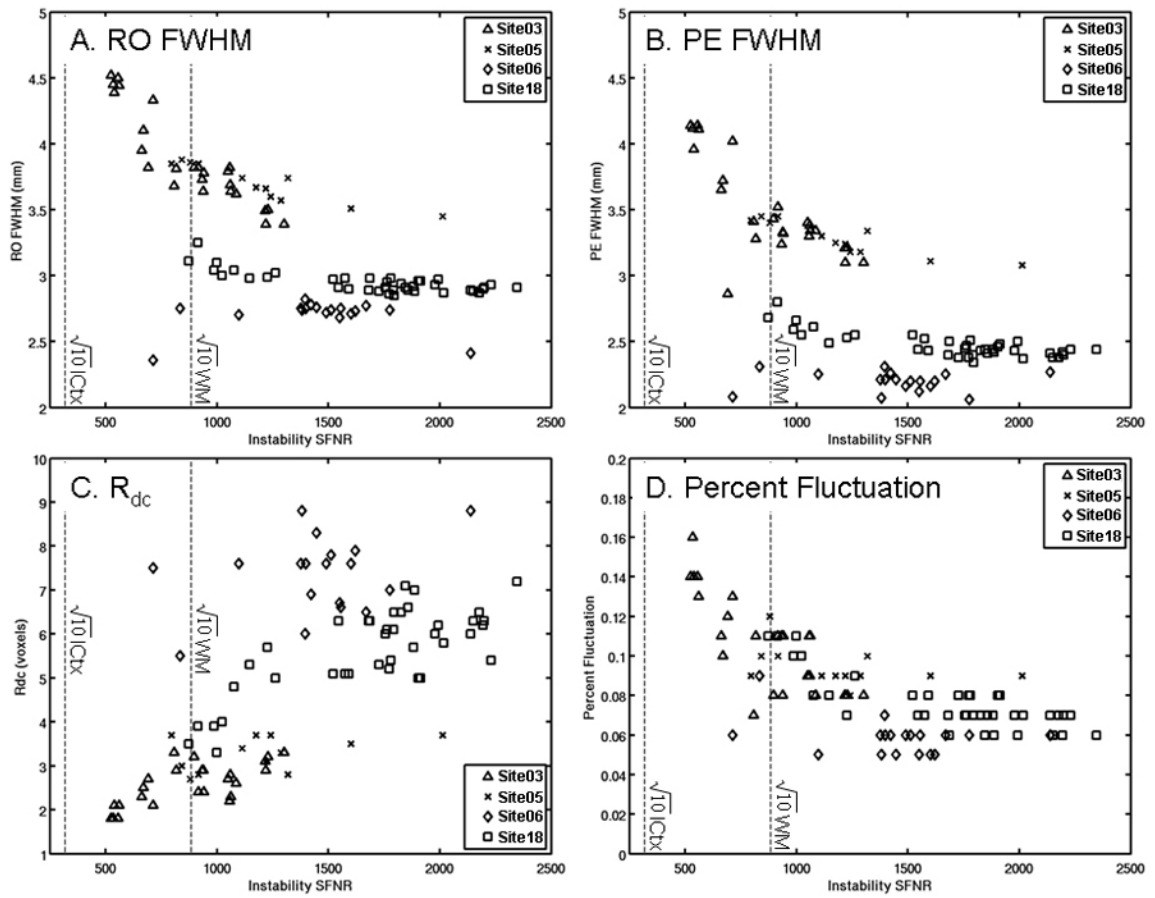


Figure 3. Other agar QA measures compared against Instability SFNR in well-operating scanners. (A) FWHM in the readout (RO) direction. (B) FWHM in the phase encode (PE) direction. (C) Weisskoff plot R_{dc} . (D) Percent Fluctuation. The vertical dashed lines indicate $3.16 (\sqrt{10})$ times the SFNR for inner cortex (ICtx) and white matter (WM) indicating that the instability would account for about 10% of the noise variance in the given structure.

Table 1

SFNR abbreviations and definitions.

Name	Definition
SFNR	Total SFNR (5)
swSFNR	Signal-Weighted SFNR (Equation 6)
bgSFNR	Background SFNR (Equation 7)
iSFNR	Instability SFNR
pSFNR	Physiological SFNR

Table 2

Site-Independent Acquisition Parameters

Field Strength	3T
TR	2000 ms
TE	30 ms
Flip Angle	$\alpha_1=77^\circ$ and $\alpha_2=10^\circ$
Time Points	100 (77°), 50 (10°)
In-Plane Voxel Size	3.44 mm
In-Plane Field-of-View	220 mm
In-Plane Matrix Size	64×64
Slice Spacing	5 mm (4mm + 1mm gap)
Slice Orientation	Axial, AC-PC
Number of Slices	30
Readout Trajectory	EPI with Ramp Sampling
Phase Encode Direction	Posterior to Anterior
Spatial Filtering ¹ (Fermi, Elliptical)	Off

¹ GE sites had gradient distortion correction turned on.

Table 3

Site-Dependent Acquisition Parameters.

Parameter\Site	Site03	Site05	Site06	Site18
Manufacturer	GE	GE	Siemens	Siemens
Scanner Model	Signa Excite	Signa HDx	TimTrio	Trio
Scanner Software	12.0	14.0	VB13	VA25
Echo Spacing	492 ms	492 ms	500 ms	490 ms
Bandwidth ^I	500 kHz	500 kHz	294.4 kHz (2298Hz/pixel)	302.7 kHz (2368Hz/pixel)
Slice Order	Interleaved	Sequential	Sequential	Sequential
Coil	8 Chan	8 Chan	12 Chan	8 Chan
Gradient Distortion Correction	On	On	Off	Off
Discarded Acquisitions	3	3	3	2
Stimulus Display	Goggles	Goggles	Projection	Projection
Excitation	Water	Water	Broadband, Fat Saturation	Broadband, Fat Saturation
Coil Combine	Weighted Sum- of-Squares	Weighted Sum- of-Squares	Sum-of-Squares	Sum-of-Squares

^I“Bandwidth” here refers to raw bandwidth and does not take into account filtering or averaging the vendor may apply during k-space reconstruction.

Table 4

Contribution of each noise source (% of total variance).

Site	White Matter			Cortex (Inner)		
	Physiological	Instability	Background	Physiological	Instability	Background
Site03	47.6	4.0	48.4	89.8	1.0	9.1
Site05	36.3	1.7	62.1	83.0	0.6	16.4
Site06	22.9	0.7	76.4	74.9	0.3	24.8
Site18A	9.6	0.2	90.2	52.0	0.2	47.8
Site18B	8.0	0.2	91.8	48.1	0.2	51.7