Published in final edited form as:
Magn Reson Med. 2009 August ; 62(2): 365–372. doi:10.1002/mrm.21992.

Automatic quality assessment in structural brain magnetic resonance imaging

Bénédicte Mortamet[1], **Matt A. Bernstein**[2], **Clifford R. Jack Jr**[2], **Jeffrey L. Gunter**[2], **Chadwick Ward**[2], **Paula J. Britson**[2], **Reto Meuli**[4], **Jean-Philippe Thiran**[3], **Gunnar Krueger**[1], and **the Alzheimer's Disease Neuroimaging Initiative***

[1] Advanced Clinical Imaging Technology, Siemens Suisse SA, Healthcare Sector IM&WS – Centre d'Imagerie Biomédicale (CIBM), Lausanne, Switzerland [2] Mayo Clinic, Rochester, MN, USA [3] Ecole Polytechnique Fédérale de Lausanne (EPFL), Signal Processing Laboratory (LTS5), Lausanne, Switzerland [4] Centre Hospitalier Universitaire Vaudois and University of Lausanne, Switzerland

Abstract

MRI has evolved into an important diagnostic technique in medical imaging. However, reliability of the derived diagnosis can be degraded by artifacts, which challenge both radiologists and automatic computer-aided diagnosis. This paper proposes a fully automatic method for measuring image quality of 3D structural MRI. Quality measures are derived by analyzing the air background of magnitude images and are capable of detecting image degradation from several sources, including bulk motion, residual magnetization from incomplete spoiling, blurring, ghosting, etc. The method has been validated on 749 3D T_1-weighted 1.5 T and 3 T head scans acquired at 36 Alzheimer's Disease Neuroimaging Initiative (ADNI) study sites operating with various software and hardware combinations. Results are compared against qualitative grades assigned by the ADNI quality control center (taken as the reference standard). The derived quality indices are independent of the MRI system used and agree with the reference standard quality ratings with high sensitivity and specificity (>85%). The proposed procedures for quality assessment could be of great value for both research and routine clinical imaging. It could greatly improve workflow through its ability to rule-out the need for a repeat scan while the patient is still in the magnet bore.

Keywords

Magnetic resonance imaging; automatic quality assessment; image quality; artifacts detection

Introduction

MR imaging quality can be affected by a wide variety of artifacts. They can be broadly classified into two categories: those that are machine-specific and those that are related to the patient. Some of the machine-specific artifacts are not visually obvious, yet can potentially degrade images. This can cause inaccurate diagnosis or dramatically affect the efficiency of automated quantitative image analysis algorithms that are increasingly used in

Corresponding author: Bénédicte Mortamet, EPFL SB IPMC LIFMET, CH F1 562 (Bâtiment CH), Station 6, CH-1015 Lausanne, +41 21 693 05 99 (office), +41 21 693 79 60 (fax), benedicte.mortamet@epfl.ch.

*Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu\ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators are listed at: www.loni.ucla.edu\ADNI\Collaboration\ADNI_Authorship_list.pdf clinical practice and research. These techniques offer promise for improved clinical workflow including clinical research studies such as longitudinal monitoring of the evolution and the treatment of degenerative and inflammatory diseases (e.g. dementias, multiple sclerosis, Parkinson disease, etc.). In this context, recognizing artifacts becomes fundamental.

Recently, various investigators have proposed standardized quality assurance (QA) protocols and methodologies to test machine-related artifacts (1-3). These protocols are often based on specially designed phantoms to analyze image quality-related system parameters such as gradient linearity (4), geometric accuracy, high-contrast resolution, slice thickness/position accuracy, image intensity uniformity (5-7), percent signal ghosting and low-contrast object detectability (8,9). These QA tests are of high interest to monitor scanner performance and retrospectively correct human images for drifts or discontinuities in gradient calibration (10,11).

Although QA tests are performed as standard procedure during tune-up and service of MR systems and are used in several clinical studies, very little has been reported about detecting and analyzing patient-related artifacts. The importance of such quality control might have been downplayed so far under the assumption that an experienced radiologist is able to “read-through” artifacts. Nevertheless, this issue has been investigated (12,13) and some studies demonstrated that a visual quality assessment of intentionally degraded MR images was not as sensitive as an automated image analysis system to detect low MRI quality (14).

Signal-to-noise ratio (SNR) has traditionally been presented as an important index of image quality in magnitude human MR images (5,15). SNR measures, however, are not necessarily sensitive to patient-related artifacts. These latter ones often appear as signal intensity being mis-mapped to an incorrect spatial location relative to tissues being imaged. Major types of patient-related artifacts are: (a) edge artifacts (chemical shifts, ringing, ghosting from motion), (b) flow artifacts, (c) aliasing artifacts (wraparound from improper patient positioning and protocol planning, e.g., nose wrap) (16-19). Most of these artifactual signal intensities propagate over the image and into the background (i.e. imaged air, whose volume, in this study, typically corresponds to 40 % of the total 3D volume of a structural MRI scan) and corrupt the expected noise distribution in affected regions. In this investigation, we focus on a careful analysis of the background intensity distribution that in most circumstances provides sufficient information to detect the presence of artifacts. It also allows the derivation of two sensitive quality indices through model-based and model-free approaches.

For verification and fine-tuning of the developed method, two quality indices were calculated on 749 structural scans obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu\ADNI). The ability of these two indices to correctly differentiate low-quality from high-quality scans was compared against qualitative grades assigned by an experienced reader from the ADNI MRI quality control center, which were used as a reference (i.e. “gold”) standard. We compared the discriminative abilities between our two quality indices over their ranges of possible cutoff points in order to identify the preferred one. Transforming these indices to a low- vs. high-quality decision requires specific cutoff points. We attempted to determine cutoff points for each quality index but more detailed research could optimize and customize them for a specific application.

The paper is organized as follows. Section 1 presents our automatic quality assessment approach, based on the computation of two indices. The results provided in Section 2 and discussed in Section 3 assess the value of our technique in routine clinical practice.

1. Methods

The proposed automatic quality control method relies on investigation of the background (air) region of the image. For the most relevant artifacts (ghosting, motion, flow, wrap-around), artifactual signal intensity displacement into the background provides the means for a sensitive quality assessment. The proposed approach is achieved in a three-step process comprising: (1) background air region delineation, (2) computation of a model-free quality index (QI_1) and (3) computation of an additional quality index (QI_2) that examines the noise intensity distribution by fitting a noise model.

1.1. Processing steps

1.1.1. Step 1 – Background region segmentation—Figure 1 shows a schematic flowchart of the method proposed to segment the background volume-of-interest (VOI). It consists in the following steps: (a) Segmentation of the whole head and (b) Atlas-based refinement of the VOI. Step a) is based on the establishment of the scalp/air boundary by means of image gradient computation. This transition is quite sharp and is enhanced by a preliminary anisotropic diffusion filtering. To exclude non-scalp/air boundary voxels, we compute a threshold from the magnitude gradient image histogram, which is defined as the intensity corresponding to 1% of the number of non-zero voxels in the image, i.e., an empirically defined frequency threshold. Thresholding produces a set of voxels belonging to the outer scalp boundary refined by a closing operation. Then a hole-filling process creates a single volume containing the entire head. Because we aim to detect artifacts that cause signal fluctuations in brain tissue (i.e. we are not interested in artifacts affecting the neck region), we restrict background noise analysis to a VOI above the plane passing through the nasion-to-posterior-of-the-cerebellum line and perpendicular to sagittal plane. This VOI is codified in a home-built MRI T1-weighted template that is aligned to the subject scan with a 12-parameters affine transformation in step b). Finally, regions below this plane are appended to the head mask resulting from step a) and a background image is obtained after exclusive masking.

1.1.2. Step 2 – Quality index QI_1: detection of artifactual voxels—Let us index N image voxels belonging to the background with $i \in S = \{1,2,\ldots,N\}$ and denote the intensity of voxel i by x_i and the background intensity histogram by H (H plots the normalized number of background voxels (vertical axis) for a given intensity in the magnitude image (horizontal axis)). After background segmentation, the first step of the algorithm consists in extracting artifactual voxels. Artifactual intensities overlap with true noise, so our primary goal is to find an appropriate threshold to remove low-intensity noise in the background. The intensity x_1 at maximum amplitude of H gives an initial estimate of the range of artifactual intensities. Thresholding produces a set of voxels described as $X_{t1} = \{i \in S: x_i > x_1\}$. After thresholding, the volume still contains voxels with intensity due to true noise that are randomly scattered through the volume. Hence, a pure thresholding intensity-based method would not be efficient to capture artifacts. To remove the remaining noise, we apply a modified morphological opening operation (20,21), consisting in an erosion of X_{t1}, using a 3D cross structuring element, and a dilation, performed iteratively with the same kernel and constrained to voxels intensity above x_1. The result of this process is a natural definition of artifact regions $X_{artifacts}$ where statistics can be performed. We consider the proportion of voxels with intensity corrupted by artifacts normalized by the background size as a first quality index QI_1:

$$QI_1 = \frac{\sum_{i \in X_{artifacts}} i}{N} \tag{1}$$

This quality measure ideally reflects the presence ($QI_1 \neq 0$) or absence ($QI_1 = 0$) of artifacts and relies on both artifacts delineation and clustered property of artifactual voxels. This latter assumption is valid for most artifacts (e.g. edge, flow, etc.), but may be violated in case of more subtle artifacts (e.g. blurring, intensity nonuniformities, etc.). Therefore, we propose a second quality index that is complementary to the first one.

1.1.3. Step 3 – Quality index *QI₂*: noise distribution analysis—

Characteristics of noise in a magnitude-reconstructed image obtained from single- or multiple-receiver unit systems has been thoroughly studied (15,22,23). For individual images, the magnitude operation transforms a complex Gaussian (normal) variate to a Rician variate. When component coil images from an array are combined as the sum-of-square magnitudes, the distribution is described by a modified Bessel function of the first kind (or chi function). Let us index N' image voxels belonging to the background from which the region $X_{artifacts}$ delineated in Step 2 has been removed with *j (X00404) S′ = {1,2,…,N′}*, the intensity of voxel j by y_j and its distribution H'. For an array of n coil elements, y_j follows a probability density function (*pdf*) defined by:

$$pdf_n(y_j) = \frac{2}{(\sigma\sqrt{2})^{2n}(n-1)!} y_j^{2n-1} e^{\frac{y_j^2}{2\sigma^2}} \tag{2}$$

where σ is the standard deviation of the true noise (i.e. Gaussian noise in the real and imaginary images). Using single-receiver system, the *pdf* follows a Rayleigh distribution while the use of more receiver units produces a more symmetric *pdf*. Figure 2 shows background noise distributions of phantom scans obtained with various receiver unit configurations. These distributions are in excellent agreement with Constantinides' modeling (22).

We propose to fit the above-outlined model to H' using maximum likelihood estimation (24). As shown on Figure 3, the presence of artifacts enlarges the noise intensity distribution's right tail that extends to higher intensity values and hence increases its skewness. Right-skewness, up to a certain degree, is incorporated in the model, particularly for single-channel receive coil. However, the model becomes less and less accurate with substantially right-skewed distributions. Therefore, we consider the goodness-of-fit (*gof*, i.e. absolute error) to be a sensitive measure of quality, especially if its computation is restricted to the right tail of the distribution, such as:

$$gof = \frac{\sum_{y_j \in [t_2, \max(y_j)]} \left| H'(y_j) - p\widehat{d}f(y_j) \right|}{N} \tag{3}$$

where t_2 is the intensity at half maximum amplitude of H' in the negative slope part.

We propose a second quality index accounting for both clustered artifacts effects (measured by QI_1) and subtle artifacts effects (measured by *gof*), defined as:

$$QI_2=QI_1+gof \quad (4)$$

1.2. Subjects and image acquisition

In this study, we assess overall image quality in 749 T1-weighted structural 3D head MRI data (188 subjects, 72.5 ± 17.5 years old) obtained from the ADNI study. Data have been acquired at 36 different Siemens 1.5T and 3T scanners (6 different scanner types). The scanners were equipped and operating with various models of receive head coils and software versions (see Table 1). Because our purpose is to evaluate image quality from background noise analysis, no screening was performed based on age, clinical dementia rating, or gender. Back-to-back scans are acquired on each subject within each ADNI scanning session. Scan and scan-repeat were considered as two independent scans because different patient-related artifacts can occur differently during these two acquisitions.

T1-weighted sagittal volumes are obtained using the magnetization prepared rapid gradient echo (MP-RAGE) pulse-sequence (25). Used imaging parameters are TR = 2300 ms, TI = 900 ms, flip angle = 9 ° at 3 T (TR = 2400 ms, TI = 1000 ms, flip angle = 8 ° at 1.5 T) minimum full TE, 160 sagittal slices. All 1.5 T subject acquisitions use 1.25×1.25 mm^2 in-plane spatial resolution and 1.2 mm thick sagittal slices. The 3.0 T subject acquisitions also use 1.2 mm thick sagittal slices, but are acquired with 1.0×1.0 mm^2 in-plane spatial resolution. In addition, phantom acquisitions are obtained as part of each imaging session using the same imaging parameters and in-plane resolution but 1.3 mm thick sagittal slices (10,11,26). These phantom scans were used to confirm Constantinides' noise modeling (22) (see 1.1.3). Detailed lists of imaging protocol are available at www.loni.ucla.edu/ADNI/Research/Cores. No pre-processing was applied.

1.3. Information on the Alzheimer's Disease Neuroimaging Initiative

The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI is to improve methods for AD clinical trials. Specific aims include optimizing methods for imaging in multi site studies and comparing the value of serial MRI, positron emission tomography (PET), fluid biomarkers, and clinical and neuropsychological assessment to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). These efforts are designed to improve the effectiveness of clinical trials to discover and validate new treatments for AD, and reduce the duration and cost of clinical trials. For up-to-date information see www.adni-info.org.

1.4. Validation process

As part of the ADNI MRI workflow, each MP-RAGE sequence is graded qualitatively for artifacts by an experienced individual at the ADNI MRI center at the Mayo Clinic. Quality grading is performed on a 4-point scale [None, mild, moderate, severe] on several criteria [blurring/ghosting/ringing; flow artifact; intensity and homogeneity; signal to noise ratio; susceptibility artifacts; gray-white CSF contrast]. General image quality is also graded on a binary basis: passed, failed. The latter criterion is considered in our investigation as a gold standard to assess the predictive performance of quality indices obtained with the proposed quality assessment method.

We evaluate the effectiveness of tests based on our automated quality indices through two attributes: accuracy and consistency.

Accuracy refers to the ability of automated quality measures to discriminate between high-quality and low-quality scans and is measured using receiver operating characteristics (ROC) curves (27) (see Figure 4.a). More specifically, the ROC curve for a test represents the range of combinations of sensitivity and specificity achievable as the cutoff value ranges from more stringent (more specific) to less stringent (more sensitive). Sensitivity and specificity need the information from the confusion matrix, which consists of four elements: true positive or appropriate high-quality decision (TP), false positive or outcome of a missed low-quality scan (FP), false negative or outcome of a missed high-quality scan (FN), and true negative or appropriate low-quality decision (TN). Sensitivity is calculated as true positives divided by number of cases “passed” on the Mayo quality control rating. Specificity is calculated as true negatives divided by number of cases “failed” on the Mayo quality control rating. We use ROC curves for three purposes: (1) to measure the performance of each test over its range of possible cutoff points provided by the area under the curve (AUC=1: perfect; 0.9-1: excellent; 0.8-0.9: good; 0.7-0.8: fair; 0.6-0.7: poor; 0.5-0.6: fail), (2) to compare the discriminative abilities between our two quality indices in order to identify the preferred one, and (3) to determine a cutoff point for each quality index (qiT_1 and qiT_2) corresponding to equal sensitivity and specificity. An appropriate modeling technique (e.g. supervised machine learning) could also be used to further determine optimal cutoff points but is not investigated here. Each cutoff point determined in (3) allows the classification of the datasets into two quality groups (high-/low-quality). The statistical difference between these two groups is evaluated by means of a one-way analysis of variance (Fischer's ANOVA) (see Figure 5).

Consistency refers to the degree to which the quality index predicts the same quantity from different field strength/system/receive coil type configurations. For example, a consistent prediction algorithm would predict the same quantity for a particular scan regardless of whether the scan was acquired at 1.5 or 3T, using 8 channels or 12 channels receive coils. To assess consistency, we conduct an ANOVA by considering acquisition configuration as an independent variable.

1.5. Practical implementation

Background noise regions extraction algorithm was written as an integrated package in ITK (Insight Toolkit). Noise distribution analyses are implemented in C++. When run on a Dual-Core Intel Itanium 2 CPU 2.93 GHz, the whole process takes about one minute for images with matrix dimensions 240×256×160. Statistical analyses are performed using Matlab routines.

2. Results

2.1. *QIs* accuracy – discrimination of low- and high-quality images

The quality of a total number of 749 scans consisting of 695 manually rated as high-quality (or pass) and 54 rated as low-quality (or fail) was assessed. On average, 42.5 ± 4.7 % of the total 3D scan volume was considered as background over which quality indices were computed. ROC curves were used to test the accuracy of each automated quality index (see Figure 4.a.). These curves have areas of about 0.93 on average indicating an “excellent” prediction performance, and overall a test based on QI_2 is as accurate as a test based on QI_1. However, QI_2 appears to be more specific in artifact detection (QI_2-based test ROC curve is steeper near the origin than QI_1 one).

Trading-off sensitivity and specificity at approximately equal rates (87.19 % and 85.18 % on average for QI_1 and QI_2, respectively) provides cutoff points of 5.06e-3 and 5.7e-2, respectively. This means that a scan will automatically be rated as low-quality if more than

0.506 % of its background volume is affected by artifacts (according to QI_1 value and 5.7 % according to QI_2 value).

2.2. *QIs* consistency – discriminative abilities independence from acquisition configurations

ANOVA reveals significant differences among the two quality groups (i.e. low- and high-quality) for each *QI* as shown on Figure 5.a. and 5.b. ($p < 0.001$). This significant trend is also valid when considering acquisition configuration as an independent variable ($p < 0.001$ for each configuration). Therefore, the discriminative performances of the proposed indices appear to be independent of differences in field strength or software-hardware combinations used by the MRI systems.

Overall, the model underlying QI_2 fits the data well and our quality indices appear to be both accurate and consistent.

3. Discussion

In response to growing recognition of the value of MRI as an instrument in clinical research, many investigators have launched multicenter studies. Similarly, multi-center trials are widely used to assess pharmaceuticals and other therapies during their various stages of development. To a large extent, the success of these studies will depend on MR data quality, which will influence the post-processing (e.g. volumetric quantification) and diagnostic conclusions.

Our purpose was to identify a simple method to automatically assess image quality from magnitude images, in order to optimize the quality of quantitative measures. This method, if applied just after the scan session and ideally incorporated in image reconstruction, may also inform the MR operator about low-quality directly after the scan and advise the need to rescan while the patient is still in the MR bore. Consequently, the number of call-back examinations and thus overall patient burden could be reduced. Because the proposed method is automated, it offers perfect repeatability (assuming that the cutoff point is held fixed), unlike human graders, where inter- and intra-observer repeatability is imperfect.

We evaluated the performance of the method against the best gold standard available corresponding to experts' quality rankings on a relatively large number of datasets. Our results indicate that the automated measures are highly efficient in predicting image quality (AUC > 0.9). As shown on Figure 4.a, QI_1- and QI_2-based test ROC curves cross. Therefore, no definite conclusion can be drawn concerning which test is the best to perform. This holds true as long as we consider investigation of quality measures based on untreated raw data (no filter, no parallel imaging) as used within ADNI. This ROC curves crossing suggests that QI_2-based test may still be preferable over a certain range of sensitivities and specificities, even though its AUC is the lowest. For example, an investigator who wishes to exclude low-quality scans might prefer to operate in the left portion of the curve where QI_2-based test appear to work better. In the opposite case of an investigator who is more concerned about sensitivity, it might be preferable to opt for the QI_1-based test, which appears more effective in the upper right end of the curve. In other words, the selection of cutoff points could be application-specific. Further investigation would therefore consist of customizing cutoff levels according to the required performance of a target application (e.g. brain structures segmentation). In the meantime, for a given sensitivity/specificity combination, corresponding cutoff values for QI_1 and QI_2 can be easily obtained using Figures 4.b. and 4.c. When considering hardware configuration as an independent variable, ANOVA revealed highly significant differences between low- and high-quality groups for each index ($p < 0.001$). Therefore, these cutoff values can be considered as “universal” and thus can be applied to every system. Moreover, introduction of a new system would not require the acquisition of phantom scan or large-scale data collection for adjustment of the respective cutoff points and there is every reason to think that our quality indices might perform equally well for data acquired with scanners from different vendors.

One potential concern was that our artifacts delineation algorithm would isolate clustered artifacts voxels but could omit subtle ones. Accordingly, model-based QI_2 test would overall perform better in detecting the latter. This hypothesis appears to be valid since we observed a trend towards an increased specificity when using the QI_2-based test. This finding suggests that incorporating additional models may improve the discriminative abilities of our quality tests. One immediate possibility could be to differentiate artifacts into subtypes with regard to their spatial location. Such differentiation could be easily obtained by incorporating subregions into the VOI template, thus having an *a priori* knowledge where specific artifacts are most likely supposed to take place. In addition, artifacts such as noise spikes would be difficult to encode in a template as they create wave-like patterns over the whole volume. In this particular case, the background intensity distribution $\boldsymbol{H}$ is no longer smooth, which can be simply detected. Consequently, measuring the degree of background histogram smoothness could provide an efficient quality criterion for noise spike type artifacts.

Although many of our quality measures were consistent with the independent visual ADNI qualitative ratings of the datasets, a few were not. At a positivity criterion of < 5.06E-03, the sensitivity of QI_1-based test is 87.34 % (607 scans appropriately rated as high-quality over 695) and the specificity is 87.04 % (47 scans appropriately rated as low-quality over 54). As seen on Figure 4.a., this sensitivity/specificity combination dominates the one achievable with QI_2. Therefore, the number of false positives and false negatives cannot be further decreased by varying the cutoff point. However, on close examination of the false positives, these discrepancies are mostly attributable to the fact that artifacts noticed by the expert do not propagate into the background (e.g. failed according to ADNI because of nose wrap or low SNR due to use of wrong coil). An interesting strategy to automatically detect quality degradation in those cases would be to derive additional quality measures from, check of imaging parameters and brain tissue intensity analysis.

This latter approach, however, raises several questions. First, given a certain contrast, is it realistic to predict brain compartment intensity characteristics such as gray/white matter or cerebrospinal fluid mean intensities? Second, there is a real question of age-related or pathology-related changes in the MR imaging characteristics of gray and white matter that could affect the significance of our quality indices. This points to a potentially important new area of research. In an initial approach, we implemented a check of sequence parameters as saved in the DICOM header of each image. Thus, we were able to detect a scan that turned out to be a false negative case according to our *QIs*. This particular scan was performed using the body coil instead of the head matrix, which resulted in low quality due to poor SNR. In another example a structural scan with the nose entering brain tissue was manually rated as low-quality. Our quality assessment, however, did not exhibit any signs of quality degradation. Nevertheless, the atlas-based background segmentation procedure provided us with the information of the erroneous positioning or choice of FoV (not incorporated in the discussed sensitivity and specificity analysis) since the nose can be spatially encoded in the VOI template. Other head features such as the chin or ears can also be incorporated into the template to detect poor slice positioning. Initial results show that our atlas-based technique performs remarkably well in detecting the degree of nose wrapping but are not presented here in the interest of conciseness.

Summarizing, additional analysis steps such as automated protocol checks, more detailed tissue analysis, smoothness analysis of the background intensity distribution, k-space analysis or introducing artifacts spatial *a priori* may provide valuable information to even further increase sensitivity and specificity in the automated quality assessment.

The proposed quality assessment is restricted to imaged background air areas and anticipates that a reasonable number of background voxels is provided. This assumption is valid in our study as about 40 % of the area in 3D volumes accounts to background (here ranging from 30 % to 55 %). Further investigation could aim at determining a minimum number of background voxels required. Nevertheless, for typical protocols such as used in this particular study, we presume that both *QIs* considered here can be directly extended to other contrasts. However, parallel-imaging techniques or corrections for B1-intensity variation of the phased array receive coils performed inline (e.g. product correction known as prescan-normalize on Siemens scanners) are known to alter background noise distribution that can no longer be modeled by central chi statistics (28). As a result, model-based QI_2 test becomes questionable in such cases. The model-free QI_1 test, however, is predicted to perform reliably and be independent of this effect, since the 3D connected structure of artifacts is typically not affected. Initial results testing the influence of prescan-normalize confirm this hypothesis as we did not observe differences in the quality rating.

4. Conclusion

Considered together, these results establish the feasibility of providing automatic image quality assessment in structural brain MRI data. Quality tests performed on 749 datasets indicate that our quality indices have excellent predictive value (area under receiver operating characteristics curve > 0.9) and correlate remarkably well with quality ratings from an independent gold standard source (both sensitivity and specificity > 85 %). The proposed automated procedures for quality assessment could be of great value for both clinical routine and research imaging. For example, this approach can provide unbiased exclusion criteria for research studies, and can greatly improve clinical workflow through its ability to rule-out the need for a repeat scan while the patient is still in the magnet bore.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI; Principal Investigator: Michael Weiner; NIH grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering (NIBIB), and through generous contributions from the following: Pfizer Inc., Wyeth Research, Bristol-Myers Squibb, Eli Lilly and Company, GlaxoSmithKline, Merck & Co. Inc., AstraZeneca AB, Novartis Pharmaceuticals Corporation, Alzheimer's Association, Eisai Global Clinical Development, Elan Corporation plc, Forest Laboratories, and the Institute for the Study of Aging, with participation from the U.S. Food and Drug Administration. Industry partnerships are coordinated through the Foundation for the National Institutes of Health. The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory of Neuro Imaging at the University of California, Los Angeles.

The authors would like to thank Meritxell Cuadra Bach for helpful discussions on data fitting, which benefited the performance of our second quality index. We also gratefully acknowledge reviewers for their constructive criticisms on drafts of this article.

References

1. Ihalainen T, Sipila O, Savolainen S. MRI quality control: six imagers studied using eleven unified image quality parameters. Eur Radiol. 2004; 14(10):1859–1865. [PubMed: 14997335]
2. Mascaro L, Strocchi S, Colombo P, Del Corona M, Baldassarri AM. Definition criteria for a magnetic resonance quality assurance program: multicenter study. Radiol Med (Torino). 1999; 97(5):389–397. [PubMed: 10432972]

3. Phantom test guidance for the ACR MRI accreditation program. American College of Radiology; Reston: 1998.
4. Mallozzi, RP.; Blezek, DJ.; Gunter, JL.; Jack, CR., Jr; Levy, JR. Phantom-based evaluation of gradient non-linearity for quantitative neurological MRI studies. In proceedings of the 14th Annual Meeting of ISMRM; Seattle, WA, USA. 2006. p. 281
5. Kaufman L, Kramer DM, Crooks LE, Ortendahl DA. Measuring signal-to-noise ratios in MR imaging. Radiology. 1989; 173(1):265–267. [PubMed: 2781018]
6. Bourel P, Gibon D, Coste E, Daanen V, Rousseau J. Automatic quality assessment protocol for MRI equipment. Med Phys. 1999; 26(12):2693–2700. [PubMed: 10619255]
7. Podo F. Quality control in magnetic resonance for clinical use. Ann Ist Super Sanita. 1994; 30(1): 123–137. [PubMed: 7832394]
8. Bernstein MA, Thomasson DM, Perman WH. Improved detectability in low signal-to-noise ratio magnetic resonance images by means of a phase-corrected real reconstruction. Med Phys. 1989; 16(5):813–817. [PubMed: 2811764]
9. Chen CC, Wan YL, Wai YY, Liu HL. Quality assurance of clinical MRI scanners using ACR MRI phantom: preliminary results. J Digit Imaging. 2004; 17(4):279–284. [PubMed: 15692871]
10. Gunter, JL.; Bernstein, MA.; Borowski, B.; Felmlee, JP.; Blezek, DJ.; Mallozzi, RP. Validation testing of the MRI calibration phantom for the Alzheimer's Disease Neuroimaging Initiative Study. In proceedings of the 14th Annual Meeting of ISMRM; Seattle, WA, USA. 2006. p. 511
11. Mallozzi, RP.; Blezek, DJ.; Ward, CP.; Gunter, JL.; Jack, CR, Jr. Phantom-based geometric distortion correction for volumetric imaging of Alzheimer's disease. In proceedings of the 12th Annual Meeting of ISMRM; Kyoto, Japan. 2004. p. 259
12. Woodard JP, Carley-Spencer MP. No-reference image quality metrics for structural MRI. Neuroinformatics. 2006; 4(3):243–262. [PubMed: 16943630]
13. Magnotta VA, Friedman L. Measurement of Signal-to-Noise and Contrast-to-Noise in the fBIRN Multicenter Imaging Study. J Digit Imaging. 2006; 19(2):140–147. [PubMed: 16598643]
14. Gardner EA, Ellis JH, Hyde RJ, Aisen AM, Quint DJ, Carson PL. Detection of degradation of magnetic resonance (MR) images: comparison of an automated MR image-quality analysis system with trained human observers. Acad Radiol. 1995; 2(4):277–281. [PubMed: 9419562]
15. Henkelman RM. Measurement of signal intensities in the presence of noise in MR images. Med Phys. 1985; 12(2):232–233. [PubMed: 4000083]
16. Mirowitz SA. MR imaging artifacts. Challenges and solutions. Magn Reson Imaging Clin N Am. 1999; 7(4):717–732. [PubMed: 10631675]
17. Clark JA 2nd, Kelly WM. Common artifacts encountered in magnetic resonance imaging. Radiol Clin North Am. 1988; 26(5):893–920. [PubMed: 3420238]
18. Hedley M, Yan H. Motion artifact suppression: a review of post-processing techniques. Magn Reson Imaging. 1992; 10(4):627–635. [PubMed: 1501533]
19. Saloner D. Flow and motion. Magn Reson Imaging Clin N Am. 1999; 7(4):699–715. [PubMed: 10631674]
20. Serra, J. Image Analysis and Mathematical Morphology. Academic Press Inc.; Orlando, FL, USA: 1983.
21. Haralick RM, Sternberg SS, Zhuang X. Image analysis using mathematical morphology. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1987; 9(4):532–550.
22. 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]
23. Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. Magn Reson Med. 1995; 34(6): 910–914. [PubMed: 8598820]
24. Duda, RO.; Hart, PE. Pattern Classification and Scene Analysis. Institute SR. , editor. New York: Wiley; 1973. p. 44-84.
25. Mugler JP 3rd, Brookeman JR. Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). Magn Reson Med. 1990; 15(1):152–157. [PubMed: 2374495]
26. Jack CR Jr, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, LW J, Ward C, Dale AM, Felmlee JP, Gunter JL, Hill DL, Killiany R, Schuff N, Fox-Bosetti

S, Lin C, Studholme C, DeCarli CS, Krueger G, Ward HA, Metzger GJ, Scott KT, Mallozzi R, Blezek D, Levy J, Debbins JP, Fleisher AS, Albert M, Green R, Bartzokis G, Glover G, Mugler J, Weiner MW. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging. 2008; 27(4):685–691. [PubMed: 18302232]

27. Cantor SB, Sun CC, Tortolero-Luna G, Richards-Kortum R, Follen M. A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. J Clin Epidemiol. 1999; 52(9): 885–892. [PubMed: 10529029]

28. Yeh, EN.; McKenzie, CA.; Grant, AK.; Ohliger, MA.; Willig-Onwuachi, JD.; S, DK. Generalized Noise Analysis for Magnitude Image Combinations in Parallel MRI. In proceedings of the 12th Annual Meeting of ISMRM; Seattle, WA, USA. 2003.

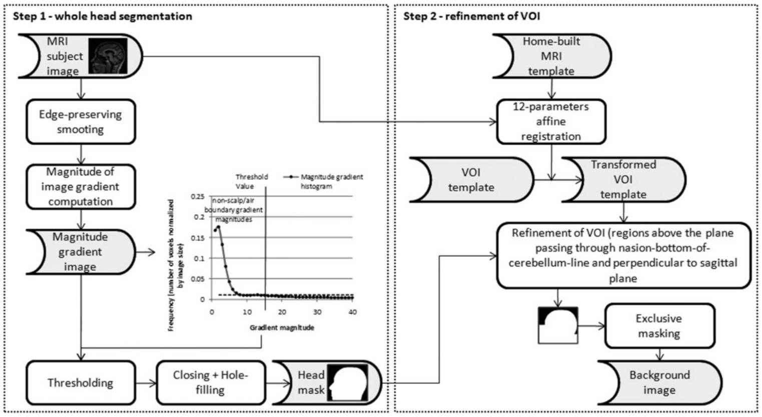
Step 1 - whole head segmentation
MRI subject image
Edge-preserving smooting
Magnitude of image gradient computation
Magnitude gradient image
Threshold Value
Magnitude gradient histogram
non-scalp/air boundary gradient magnitudes
Frequency (number of voxels normalized by image size)
0.25
0.2
0.15
0.1
0.05
0
0
10
20
30
40
Gradient magnitude
Thresholding
Closing + Hole-filling
Head mask
Step 2 - refinement of VOI
Home-built MRI template
12-parameters affine registration
VOI template
Transformed VOI template
Refinement of VOI (regions above the plane passing through nasion-bottom-of-cerebellum-line and perpendicular to sagittal plane
Exclusive masking
Background image

Figure 1.

Flow chart showing the 2 steps involved in background region segmentation

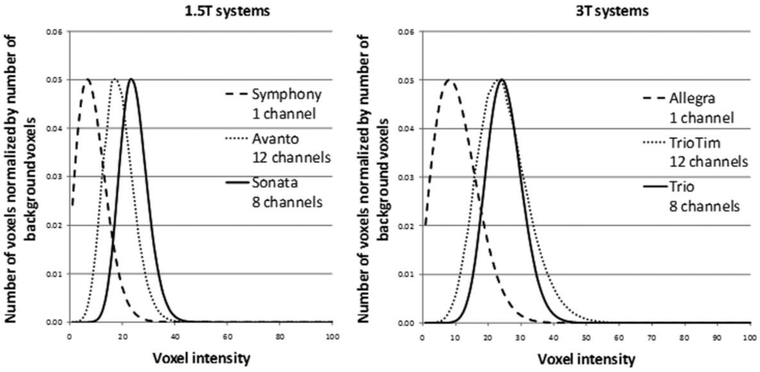

Figure 2.

Background noise distribution for magnitude sum-of-squares phantom images (MP-RAGE) using single-, eight-, and twelve-channel receiving system at 1.5 T and 3 T.

12-channel systems were driven in ADNI in CP-matrix mode, which equals a 4-channel mode.

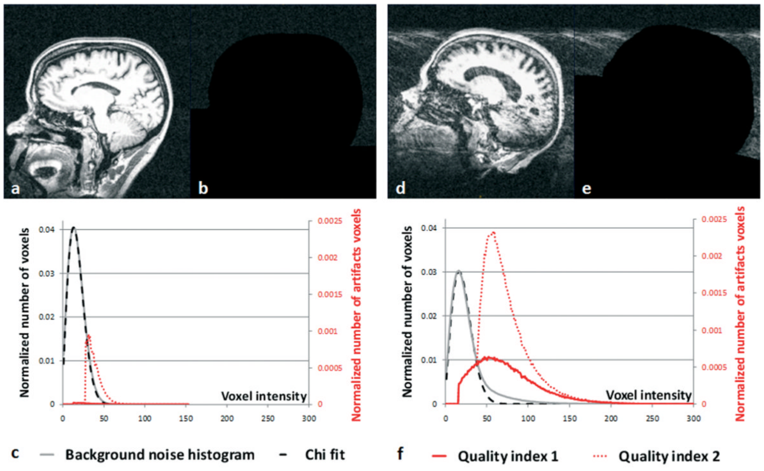

Figure 3.

Original sagittal slices of high- (a) and low- (d) quality datasets (MPRAGE, Allegra, single channel RX coil) along with their background air mask used for analysis (b-e)

On the graphs (c-f), solid gray line corresponds to histogram of background voxels intensities, dash dark line to chi fit, solid red line to QI_1 intensity distribution and dot red line to QI_2 intensity distribution, the number of voxels on both left and right axes is normalized by the total number of voxels belonging to the background.

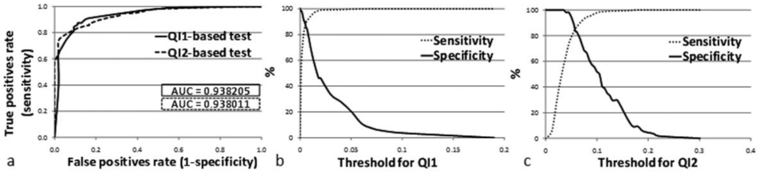

Figure 4.

Performance of quality tests demonstrated by area under ROC curves (a) and corresponding cutoff value ranges used to generate sensitivity-specificity pairs for QI_1 (b) and QI_2 (c).

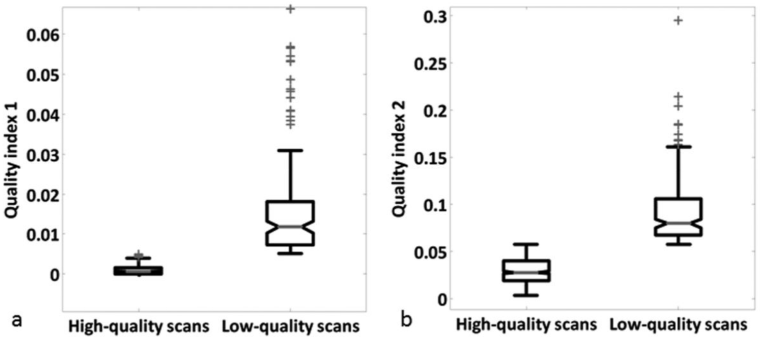

Figure 5.

Comparison of QI_1 (a) and QI_2 (b) variability ($p < 0.001$ for each QI).

Top and bottoms of each box are the 25th and 75th percentiles of the samples, respectively. The line in the middle of each box is the sample median. The whiskers are lines showing adjacent values. Observations beyond the whisker are outlier.

Table 1

1.5T and 3T Siemens scanner types and software/hardware combinations involved in this study.

Field strength	System	Receive coil type	Number of sites involved	Number of subjects scanned	Software versions	Number of datasets
1.5 T	Symphony	1 channel CP Mode	9	44	VA21A;VA25A;VA30A	185
	Sonata	1 channel CP Mode	4	25	VA25A;VA30A	113
	Sonata	8 channels PA	2	17	VA25A	71
	Avanto	12 channels PA	4	18	VB11D;VB13A;VB15	62
3 T	Allegra	1 channel BC	3	19	VA25A	67
	Trio	8 channels PA	8	46	VA25A	178
	TimTrio	12 channels PA	6	19	VB12T;VB13A;VB15	72
Total	6	4	36	188	7	749

The number of sites, the receive coil type (birdcage BC, Phased-Array PA), with operating mode (Circularly Polarized CP if applicable) and software versions, the number of subjects (who had baseline or/and follow-up scans back-to-back) and the total number of datasets for each system are indicated.