Automated Quality Assurance for Magnetic Resonance Image with Extensions to Diffusion Tensor Imaging

By

Atiba Fitzpatrick

Thesis submitted to the graduate faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the Degree of

Master of Science
in
Electrical Engineering

Christopher L. Wyatt, Ph.D., Chair
A. A. Beex, Ph.D.
Roger Ehrich, Ph.D.

June, 2005
Blacksburg, Virginia

Keywords: Magnetic Resonance Imaging, Diffusion Tensor Imaging, Automated Quality Assurance, American College of Radiology Accreditation Program, Phantom

Automated Quality Assurance for Magnetic Resonance Image with extensions to Diffusion Tensor Imaging

Atiba Fitzpatrick

(Abstract)

Since its inception, Magnetic Resonance Imaging (MRI) has largely been used for qualitative diagnosis. Radiologists and physicians are increasingly becoming interested in quantitative assessments. The American College of Radiology (ACR) developed an accreditation program that incorporates tests pertaining to quantitative and qualitative analyses. As a result, sites often use the ACR procedure for daily quality assurance (QA) testing.

The ACR accreditation program uses information obtained from clinical and phantom images to assess overall image quality of a scanner. For the phantom assessment, a human observer performs manual tests on T_1 and T_2-*weighted* volumes of the provided phantom. As these tests are tedious and time consuming, the primary goal of this research was to fully automate the procedure for QA purposes. The performance of the automated procedure was assessed by comparing the test results with the decisions made by human observers. The test results of the automated ACR QA procedure were well correlated with that of human observers. The automated ACR QA procedure takes approximately 5 minutes to complete. Upon program completion, the test results are logged in multiple text files.

To this date, no QA procedure has been reported for Diffusion Tensor Imaging (DTI). Therefore, the secondary goal of this thesis was to develop a DTI QA procedure that assess two of the associated features used most in diagnosis, namely, diffusion anisotropy and the direction of primary diffusion. To this end, a physical phantom was constructed to model restricted diffusion, relative to axon size, using water-filled polytetrafluoroethylene (PTFE) microbore capillary tubes. Automated procedures were developed to test fractional anisotropy (FA) map contrast and capillary bundle (axon) orientation accuracy.

Acknowledgements

First, I would like to thank my advisor Dr. Chris Wyatt for his patient and kind instruction over the past two years. I have not met many professors who are as technically competent, yet so down to earth and friendly. Any time that I was stalled and frustrated, I could rely on Dr. Wyatt to give a fresh perspective and inspire further thought.

I must also thank Dr. Karen Sanders, assistant provost and director of the Center for Academic Enrichment and Excellence (CAEE), and Ms. Nini Smiley, assistant director of the Multicultural Academic Opportunities Program (MAOP), for granting me an assistantship that allowed great flexibility for my pursuit of employment and the completion of this research. I must also thank them and all staff of CAEE and MAOP for their kindness and warmth that gave me a sense of family although I was so far from home.

I would like to thank Dr. Roger Ehrich and Dr. A. A. Beex for serving on my committee and taking time out during the summer to critique my defense. Without the efforts of Dr. Robert Kraft, this work would have literally not been possible. Dr. Kraft scanned the DTI phantom and provided MRI data and crucial feedback for the automated ACR QA procedure. I am very appreciative. I would also like to thank my lab mate Jung Wook Suh for implementing the image registration procedure.

I greatly appreciate my family's understanding and forgiveness for the numerous birthdays, holidays, high school football games, and parties that I missed over the last 6 years. Yet, they must know that they truly are my inspiration for pursuing higher education. I am simply following in my mother's footsteps, trying to make my father proud, and trying to be a good role model for my siblings. I would like to thank my uncle for doing much more than is required of an uncle over the years. You have been a second father to me.

On a personal note, I must thank my best friend Danielle Miles for the great advice and encouragement that she constantly gives me. My life has been greatly enhanced since meeting you.

I have to give a "big up" to all members of the Caribbean Student Organization at Virginia Tech. We have definitely had some fun times together, and you have given me a tropical vibe in this often cold town called Blacksburg. Thanks for allowing me to pull away during the last few months. We must all definitely keep in touch. I must also thank my good

friends from my alma mater (RPI) Mandoye, Chitsunge, Max, Abdul, Kayode, Seriak, Momar, Jean-Claude, and everyone else for being unconditional friends.

Much thanks to my lab mates Jung Wook Suh, Lisa Demar, Warren Liu, and Yangming Ou for the great advice and ideas that you helped me foster and for being patient when I got a little talkative. Last but certainly not least, I must give all glory and thanks to the almighty, without whom I would not even exist much less be in this position.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACR – American College of Radiology

ACR-PA – American College of Radiology Phantom Assessment

ADC – Apparent Diffusion Coefficient

BKE/SKE – Background Known Exactly and Signal Known Exactly

CNR – Contrast to Noise Ratio

DSI – Diffusion Spectrum Imaging

DTI – Diffusion Tensor Imaging

DT-MRI – Diffusion Tensor Magnetic Resonance Imaging

EPI – Echo-Planar Imaging

FA – Fractional Anisotropy

FN – False Negative

FOV – Field of View

FP – False Positive

ID – Inner Diameter

MRI – Magnetic Resonance Imaging

MTF – Modulation Transfer Function

NEX – Number of Excitations

OD – Outer Diameter

PIU – Percent Integral Uniformity

PSF – Point Spread Function

PTFE – Polytetrafluoroethylene

QA – Quality Assurance

RA – Relative Anisotropy

RF – Radio Frequency

ROI – Region of Interest

SNR – Signal to Noise Ratio

TE – Echo Time

TN – True Negative

TP – True Positive

TR – Repetition Time

Voxel – Volume Element

VR – Volume Ratio

WFUBMC – Wake Forest University Baptist Medical Center

Chapter 1 Introduction and Background

1.1 Introduction

Since its inception in 1973, magnetic resonance imaging (MRI) has largely been used for qualitative diagnosis of diseases. Physicians or radiologists inspected MR images looking for signs of abnormality. However, more recently, health practitioners are concerned with the quantitative aspects that MRI can contribute to diagnosis. As when any system is quantified, repeatability and comparison of measurements are motivating factors.

For this reason, institutions like the American College of Radiology (ACR) developed accreditation programs for MRI scanners as well as for devices corresponding to other imaging modalities. The ACR MRI accreditation program requires facilities to submit clinical and phantom images for an assessment of scan quality. The ACR also recommends that sites carry out daily quality assurance (QA) testing and they provide a QA procedure that is similar to their accreditation program. However, if the actual accreditation program is used as a daily QA procedure, sites would have very high probability of attaining accreditation on the first trial. The QA procedure could serve to pre-qualify images before submission to the ACR for accreditation. The ACR's program is of particular interest because it is required by many health insurance companies for reimbursement of payment for MRI services by clients [15].

Of primary interest in the ACR accreditation program is the phantom image assessment portion. The ACR provides a physical phantom to participating facilities that provides ground truth measurements for various tests. This phantom is imaged according to predefined protocol and the images are to be sent in for the ACR to perform the various associated tests. With the ACR accreditation program being used for daily QA, the images can also be analyzed on site. There are two issues of concern with this procedure. First, the accreditation program procedure involves individuals performing manual tests. That is, someone views the images and performs the associated tests using various on-screen measurement tools. As one can imagine, this is a very tedious task, which deters site employees from carrying out the QA testing on a daily basis. Secondly, it has been shown that human testing by visual inspection is not objective since factors such as fatigue and visual perception variation come into play [12].

The ACR accreditation procedure only pertains to the two most common scan sequences namely, T_1 and T_2, which pertain to structural analyses. Increasing in popularity and use is a new scan sequence called Diffusion Tensor Magnetic Resonance Imaging (DT-MRI), which is more commonly referred to as Diffusion Tensor Imaging (DTI). It serves as a non-invasive estimate of white matter connectivity by measuring water diffusion. Many diseases are associated with white matter abnormality such as multiple sclerosis, brain ischemia, alcoholism, and schizophrenia. Multiple sclerosis lesions occur in white matter and underlie gray matter. Ischemia is localized tissue anemia that is due to the constriction or obstruction of a blood vessel. Chronic alcoholism has been found to cause neurological abnormalities. Schizophrenia can be characterized by the anisotropic asymmetry pattern differences between healthy and affected patients. Analyzing white matter structure is useful in diagnosing and assessing suitable treatment of these diseases. It is essential that quantitatively accurate and high quality images be used for diagnosis. To achieve some level of certainty that a scanner is functioning as it should, a definitive and repeatable quality assurance procedure should be used.

The central objective of this thesis was to fully automate the phantom portion of the ACR accreditation program for MRI daily QA purposes as well as to develop a novel QA procedure for DTI. In order to fully automate the ACR related QA procedure, a rigid image registration procedure was incorporated, which was sufficient since the ACR phantom is a rigid body. A motion estimation algorithm [37] was used to determine the translational and rotational differences between slices of candidate series and a standard series. The volume element (voxel) locations of all relevant objects were known for the standard series, and after registration, the locations of these objects are also known for the candidate series and can be accessed for the various tests. Upon completion of the QA procedure, the test results are logged to a text file so that longitudinal analysis of machine performance can be done. In developing the DTI QA procedure, the approach was to first design and manufacture a suitable physical phantom to provide ground truth measurements indicative of biological diffusion. The decision was made to produce an inorganic phantom, which mainly consisted of water-filled capillary tubes grouped into bundles of various sizes. Also, suitable scanning parameters were decided upon and software was composed to automate the testing of fractional anisotropy (FA) contrast and bundle orientation accuracy scanner capabilities.

This thesis is organized as follows. In chapter 1, the significance of this research will be stated. The relevant theoretical background and QA literature will be discussed. Chapter 2 will discuss the approach taken in automating the ACR accreditation program. Also, the experimental methods involved in developing the DTI QA procedure will be discussed. In the 3rd and final chapter, the verification of the ACR related QA procedure will be discussed. The results of expected pass testing and third party forced failure testing will be discussed. Also, the conclusions drawn from this experiment and possible future work will be discussed.

1.2 Significance

Up until the present, there has been little standardization in the MRI field. Manufacturers have their own methods for verifying scanner functionality and performance, and sites use their own imaging protocols. With each new year, MR technology improves and as would be expected, mean image quality among different sites increases. However, the image quality variance between sites also increases [36], which is indicative of the need for QA.

A review of literature has uncovered numerous MRI related QA procedures that are either manual, semi automated, or fully automated. Firbank et al. proposed a procedure in which a scanner operator would measure phantom image signal to noise ratio (SNR), image uniformity, and geometric distortion and record these measurements on a weekly basis [17]. Two of the semi automated procedures that were discovered were 1) a method developed by Barker et al. which was tailored to quantitative analyses [6] and 2) a procedure that was developed to assess scanner variability for imaging parameters [19]. Recognizing that manual analysis is an impediment to regular QA, Simmons et al. developed a QA procedure that uses automated Shewart charting to identify shifts in the imaging parameters [34]. Gardener et al. developed an automated procedure to detect the degradation of image quality and compared the procedure's ability to that of trained human observers, and as would be expected, the automated procedure was more sensitive [20]. Two commercially available fully automated QA procedures were also found. Radiological Imaging Technology and Cardinal Health both produce MRI QA software, but these procedures require custom designed phantoms [29 and 13].

It is beneficial to use the ACR accreditation program procedure for MRI QA for the following reasons. Some health insurance agencies, such as Aetna US Health Care and

Pennsylvania Blue Cross Blue Shield, have ACR accreditation as a requirement for compensation for MRI related services [15 and 10]. Also, Rhode Island and Connecticut have made ACR accreditation state requirements [10]. ACR accreditation is also required for research institutions that desire to participate in collaborative cross-site projects. The ACR requires that sites participating in the accreditation program incorporate a QA procedure for daily/weekly testing that is consistent with the tests of the accreditation program, and the ACR provides such a procedure [2]. However, the QA procedure that the ACR provides only contains 3 of the 7 accreditation tests. If the full accreditation program was used for the sites QA procedure, the scanner could be regularly pre-qualified for accreditation. This QA procedure would be cost effective because sites participating in the ACR program could use the provided phantom that they already own and would not have to purchase any additional components. Also, 32 percent of the participating sites fail accreditation the first time, which leads to a $500 resubmission fee and the lost revenue due to additional down time [26]. Pre-qualifying data prior to the initial submission should greatly improve a site's chance of passing.

Without much debate, many would agree that a fully automated QA procedure is more desirable than a manual procedure. The ACR accreditation program is a manual human-observer dependent procedure that is widely regarded as being time consuming and tedious [10]. For the required T1 and T2 series, the procedure will take about a half hour to complete. Even though the staff at various imaging centers might be aware of the benefits of a regular QA program, the time and effort necessary for the ACR procedure can be a great deterrence. With an automated procedure, a technician can simply initiate the sequence, and then return at a later time to view the test results. During the intervening moments, the technician could be performing other unrelated tasks. Another reason for automating the procedure is to avoid the subjectivity of a manual procedure. Especially for qualitative analyses, results can vary greatly between different observers and would depend on an individual's visual perception. One attempt at automating the ACR procedure has been made [24]. However, this procedure had major limitations with the most prominent being that it was not fully automated.

An extensive literature review revealed that no image-quality or imaging parameter variability related QA procedure has been developed for DTI. Existing DTI quality related testing procedures primarily pertain to the correction of eddy current induced image artifacts [9]. Diagnosing and assessing the treatment of diseases such as Schizophrenia largely involve

estimating white matter connectivity by measuring anisotropic diffusion and determining the direction of maximum diffusion [28 and 1]. Therefore, it would be wise to ensure that a particular scanner was functioning well with regard to these features.

1.3 Theoretical Background

1.3.1 Quality of Images

A working definition of image quality for scientific and medical purposes is how well a task can be performed on the image. These tasks are of two categories, namely, detection and parameter estimation [12]. A similar definition is that image quality is measured by how well desired information can be extracted from an image by an observer [7]. The classification category includes signal detection, and discrimination. The estimation category pertains to estimating an unknown quantity of an object known to be contained in the image.

The classification task is simplest in the case known as background known exactly and signal known exactly (BKE/SKE). In this instance, even when the contrast between the signal and background is small, one would expect the task to be trivial. The difficulty is introduced by the random noise that is inherent in imaging systems. Thus, a measure called the SNR is of interest in any image quality consideration. As the name implies, the SNR compares some measurement of the signal of interest and the noise. The measurements used for this purpose vary. Some literature defines SNR as the signal mean divided by the noise variance, which is a sufficient measure when the noise is uncorrelated "white" noise. However, the variance is not a useful measure if the noise is correlated. In this case, determining the SNR is more involved.

Another quantity that is sometimes used in image quality considerations is Pixel SNR, which is defined as pixel signal amplitude divided by the noise variance per pixel. However, unless the signal of interest is comprised of a single pixel, this quantity can be misleading [12]. The SNR in an image can be high, but without sufficient contrast relative to the noise, detectability can still be low. Contrast has many different definitions, but they all involve a difference between signal and background amplitude. Therefore, a measure called the contrast to noise ratio (CNR) is often used in image quality considerations, which is often defined as contrast divided by noise variance. Another image characteristic, which is especially important in

medical imaging, is spatial resolution. This is most commonly defined as the closest proximity of two point sources that allow for the signals to be discernible after imaging. The spread of a point source during imaging can be characterized by the point spread function (PSF) as well as the modulation transfer function (MTF) of the imaging system Transforming the PSF from the spatial domain to the frequency domain results in the MTF. This PSF is convolved with the object during imaging.

The observer, mentioned in the image quality definition, can be a human or an algorithm. When an observer is performing classification tasks, there are 4 possible outcomes to a decision. A true positive (TP) is defined as a positive (e.g. signal present) decision being made when the true state was positive. Alternatively, a false positive (FP) is when the observer decided positive, but the true state was negative. Similarly, the outcome can be a true negative (TN) or a false negative (FN). A way of assessing the effect of a certain parameter on task accuracy is the receiver operating curve (ROC). This curve is produced by plotting the TP rate vs. the FP rate while varying a parameter. The parameter being varied can be any of the image quality measures, or if the observer is an algorithm, the parameter might be the decision threshold. The area under the curve (AUC) is a measure of the performance of the observer relevant to the fixed parameters. Figure 1.1 shows an example of an ROC curve. While the ROC is still the gold standard of performance evaluation, obtaining it is very time consuming and requires many images.

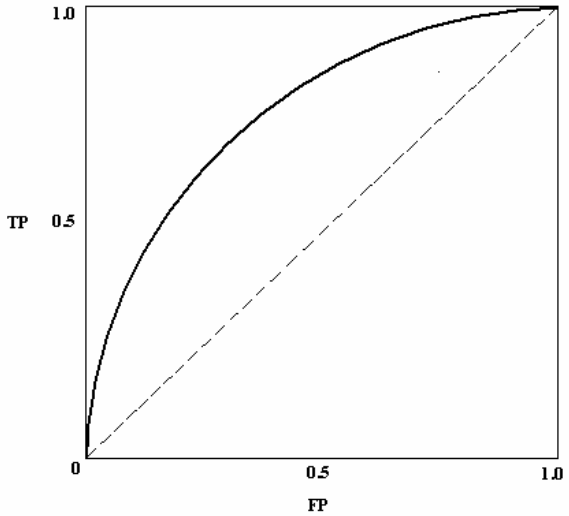


Figure 1.1 – Receiver operator curve. The bold line shows a possible curve that might result from plotting the TP fraction vs. the FP fraction while varying a parameter (such as SNR or decision-threshold). The dashed line corresponds to the case where the diagnosis was random.

When the observer is a human, the task being performed can be affected by two independent factors, which are data quality and display quality, and both of these factors must be optimized. The data quality assessment can be performed with a mathematical model, which is the algorithm observer alluded to earlier. The display quality involves the perceptual capabilities of humans, so a human task performance measurement is used for the assessment [33].

The Rose Model is described in three papers written by Albert Rose between 1946 and 1948 [30-32]. These works were among the first to assess the performance and limitations of film, electronic imaging devices, and the human visual system in signal detection in the presence of noise. The underlying reasoning motivating these considerations was based on the knowledge of the quantum nature of light and its distribution as described using Poisson statistics. Rose considered small circular dots in images corrupted by noise. Of most interest is his work in developing human observer models, which he was the first to do. Both models for static and dynamic images were developed. These models involved showing circular spots in images corrupted by noise with a 50% probability that the spot was present. The observer would then indicate whether a spot was seen or not. Rose then noted all possible outcomes including FPs and FNs. Rose concluded that human decision efficiency was between 20% and 50%. Rose's goal was to determine a SNR threshold that would allow the human observer to detect the signal. This threshold value, which was determined experimentally, could be used to also determine the corresponding contrast threshold. An oversight of Rose was that the threshold determined also depended on individual human perceptual capabilities.

It is understandable that Rose's approach is still used in image quality assessment considering the following conditions of his model. The conditions of the images used were SKE/BKE. Thus, perfect prior signal information was assumed. Most importantly, the signals were flat topped and sharp edged. This meant that photon counting is equivalent to cross correlation with a like template which included a scaled version of the signal. However, one should be mindful of certain limitations and concerns [12]. First, Rose's approach is only valid under the aforementioned conditions of his model. The most limiting of these is the flat top signal. Since imaging systems involve blurring, this limits the validity to synthetic images, which are of little use. Further, as a result of this condition, the model is not valid for filtered images. Secondly, as previously mentioned, the experimentally determined threshold is dependent on

human perceptual ability, and so, lacks objectivity. Lastly, for an image quality measure to be useful, it should be related to the purpose of the image or the task to be performed on the image. Thus, the SNR threshold decision must be made with this consideration.

An alternative to ROC analysis and the Rose Model for image quality assessment is the use of model observers. In general, the model observer should be based on the task involved, so the model observer should be made to predict the behavior of the ultimate observer whether that be a human or an algorithm [12]. These observers have the advantage of not being as time and resource consuming as the ROC procedure while being far more applicable and objective than the Rose Model. However, model observers are not as simple as the aforementioned procedures. Descriptions of two popular model observers are now provided.

The simplest observer model, called the ideal observer, stems from Bayesian math of detection theory. In this case though, the entire image is taken to be a random sample of size M x 1, where the image is of size M. For the sake of explanation, we can assume that the task to be performed is to observe an image **g** and determine the state of an object *f* that produced the image to be of a positive or negative state (where positive can mean signal present; signal normal). Then a likelihood ratio can be determined that is of the form

$$l_{ideal} = \log(\, l(\mathbf{g})) = \log\left[\frac{p(\mathbf{g},1)}{p(\mathbf{g},2)}\right] \tag{1.1}$$

where $p(\mathbf{g},k)$ is the conditional probability distribution of **g** given that the state is *k* (positive/negative) and $l(\mathbf{g})$ is known as the likelihood ratio. If $?_{ideal}$ is greater than a specified threshold, the true state is determined to be positive. In the cases of SKE/BKE when the noise is known to be identically Gaussian distributed for each class, it can be shown that this test statistic simplifies to

$$l_{ideal} = (\bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1)\mathbf{K}^{-1}\mathbf{g} \tag{1.2}$$

where $\bar{\mathbf{g}}_k$ is the mean vector of class k and **K** is the covariance matrix of the noise.

Disadvantages of the ideal observer are that it has little predictive ability of human performance when the noise is not Gaussian distributed or when the image is not SKE/BKE since the likelihood ratio becomes highly nonlinear and mathematically intractable. In these cases, it is best to use another model [7].

Before the description of the second observer can be given, concepts of linear discriminants must be briefly provided. When the ideal observer is intractable due to a highly nonlinear likelihood ratio function, linear discriminants can be used to improve classification. These procedures reduce the dimensions of the feature space such that the classes are most separable. The most famous linear discriminant was created by Fisher [18]. The Fisher discriminant is a measure of how separable classes are relative to a vector $\mathbf{w}$. Each sample is projected onto a line in the direction $\mathbf{w}$ in order to achieve the dimension reduction. When the number of classes is 2, such as for the detection task, the measure is primarily given by the difference between the projection of the class means onto the aforementioned line. However, to negate the effect of the scale of $\mathbf{w}$, the mean difference is divided by the difference of scatter matrices. The scatter matrix for a class is simply the sample covariance matrix scaled by the number of samples in the class. The general Fisher linear discriminant criterion is given by

$$J(\mathbf{w}) = \frac{\mathbf{w}^t \mathbf{S}_B \mathbf{w}}{\mathbf{w}^t \mathbf{S}_W \mathbf{w}} \tag{1.3}$$

where $\mathbf{S}_B$ and $\mathbf{S}_W$ are known as the between-class and within-class scatter matrices respectively and are given by

$$\mathbf{S}_B = \sum_{i=1}^{c} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^t$$

and (1.4)

$$\mathbf{S}_W = \sum_{i=1}^{c} \mathbf{S}_i$$

where c is the number of classes, n_i is the number of samples in class i, $\mathbf{m}_i$ is the sample mean of class i, $\mathbf{m}$ is the total sample mean, and $\mathbf{S}_i$ is the scatter matrix for class i [14].

Based on linear discriminant analysis, an image quality assessment might be taken to be the separability of the classes given a certain imaging system or SNR. However, this method is not practical because $\mathbf{S}_W$ is not invertible if the number of samples does not exceed the number dimensions [7] (underdetermined system). This would mean that the number of images would have to exceed the number of pixels, which is not reasonable.

The Hotelling observer remedies this limitation by using simulated images to calculate or estimate ensemble covariance matrices. Therefore, the Hotelling observer has a measure similar to the linear discriminant criterion that uses these ensemble covariance matrices. The procedure

for calculating the Hotelling criterion is given in [16, 23, and 8] in detail. However, a brief description is now provided. Let $\mathbf{S}_B$ and $\mathbf{S}_W$ for the Hotelling observer be given by

$$\mathbf{S}_B = \sum_{i=1}^{2} P_1 P_2 (\bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1)(\bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1)^t$$

and (1.5)

$$\mathbf{S}_W = \sum_{i=1}^{2} P_i \mathbf{K}_i$$

where P_i and $\mathbf{K}_i$ are the prior probability and ensemble covariance matrix, respectively, of class i. Then the Hotelling criterion is given by

$$J = \text{trace}(\mathbf{S}_W^{-1}\mathbf{S}_B). \quad (1.6)$$

The resulting Hotelling test statistic, which is given by

$$l_{Hot} = (\bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1)\mathbf{S}_W^{-1}\mathbf{g}, \quad (1.7)$$

is quite similar to that of the ideal filter in equation (1.1). The difference is that the sample covariance matrix is replaced by the within-class scatter matrix. In the case of SKE/BKE, both test statistics are identical. However, the Hotelling test statistic is applicable to other cases.

1.3.2 Nuclear Magnetism

Nuclear magnetism is the area of physics concerning the magnetic properties of atomic nuclei. Of primary concern in MRI considerations are the hydrogen atom. This is because hydrogen is the most abundant element in the body, which is understandable since two hydrogen atoms are present in every water molecule [39]. Moving charges produce a magnetic field with the strength of the field being proportional to the charge magnitude and the rate of motion. Although protons possess a very low charge, they spin very fast, and so have a noticeable magnetic field. Therefore, the proton possesses angular momentum resulting in a magnetic moment.

I will now consider the effect of applying an external magnetic field to the proton. To aid in this explanation, I will use the analogy of a spinning top under the influence of gravity. If a top was spinning without being influenced by gravity, it would be perfectly upright as it spun. When gravity is introduced, the top will begin to tilt as it spins and a noticeable wobble would be present. This wobble is known as precession and is represented in Figure 1.2. Similarly, when a

proton is near or in an external magnetic field B_0, it will precess at the Larmor frequency. This precession results because the external magnetic field is causing the proton to align with the field direction. Therefore, it should not be surprising that the Larmor frequency is directly proportional to the magnetic field and be determined with the equation:

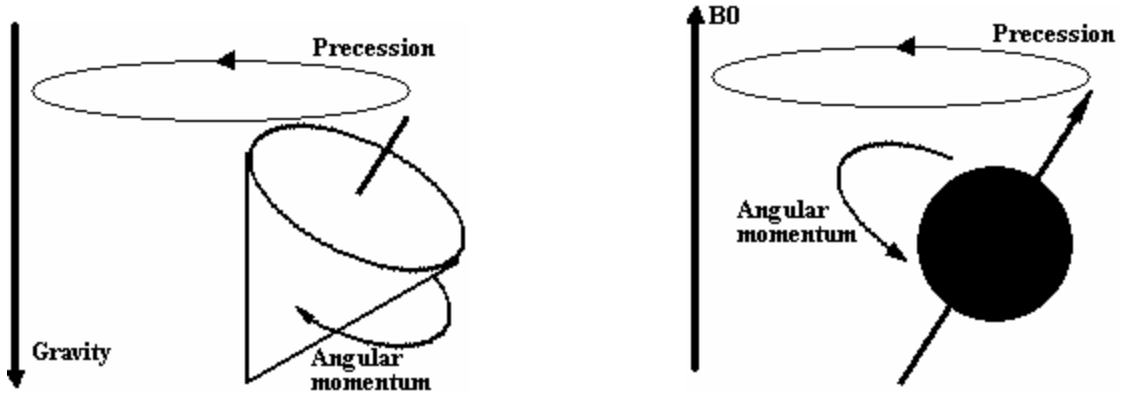
$$w_0 = gB_0 \tag{1.8}$$

where

w_0 : is the Larmor frequency in units of MHz

g : is the gyromagnetic ratio and is characteristic of the element. For protons it is 42.58 MHz/T

B_0 : is the external magnetic field strength in units of T.



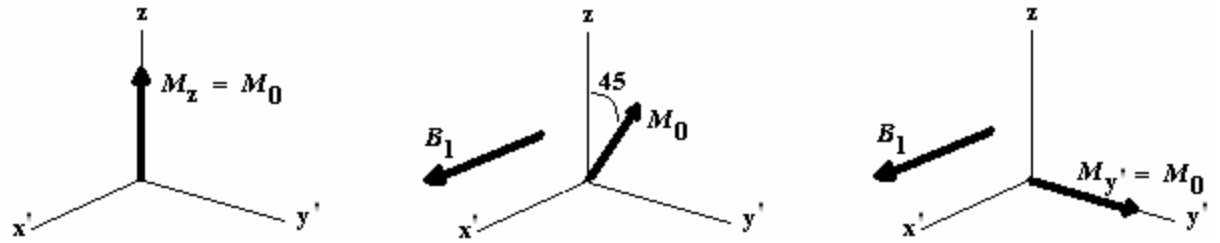
(a) (b)

Figure 1.2 – Illustration of precession. (a) Precessing top. (b) Precessing proton.

Now I will consider what some literature refers to as a spin system [39], which is simply the consideration of many protons or "spins" as would be found in tissue being imaged. Initially, the magnetic moments of the protons are randomly oriented resulting in a zero net magnetization. When the magnetic field is applied, the protons will align themselves either parallel or antiparallel where parallel means that the magnetic moment is pointing in the same direction as the magnetic field. A few more protons are aligned parallel to the magnetic field resulting in a small but measurable net magnetization M_z in the field direction where the field direction is denoted as the z-direction.

After the magnetic field B_0 has been applied causing the spins to be either in a parallel or antiparallel orientation, a radio frequency (RF) pulse B_1 can be applied at the Lamor frequency to

"excite" the spin system. An RF coil is used to transmit this energy into the spin system. Recall that before excitation, the net magnetization pointed in the z-direction. The excitation causes the net magnetization to "flip" such that the vector is pointing between the z-direction and the x-y plane as shown in Figure 1.3. The angle that the vector makes with the z-axis is called the flip angle. The flip angle is proportional to the RF pulse magnitude and duration. The individual protons continue precessing at the same rate causing the net magnetization vector to continually precess about the z-dimension. If the protons are flipped by 90 degrees, the entire net magnetization vector will lie in the x-y (transverse) plane M_{xy}, but will still precess about the z-dimension. Thus, the rotating net magnetization will induce an alternating voltage in the RF coil, which is the MR signal considered in imaging.



(a) (b) (c)

Figure 1.3 – Result of applying RF pulse to spin system. The coordinate frame used is known as the rotating frame since x'-y' rotate around the z axis at the Lamor frequency. This frame is used so that we can temporarily ignore precession. (a) Before the pulse is applied, the net magnetization is in the field direction. (b) A pulse can be applied to cause the net magnetization vector to flip by 45 degrees. (c) A pulse of greater magnitude or longer duration can cause the net magnetization vector to flip by 90 degrees.

After the RF pulse has ended, the spin system immediately begins to transmit energy absorbed from the pulse. The net magnetization in the x-y plane gradually reduces, and so the signal being measured also reduces. There are two independent processes by which this occurs, namely, spin-lattice interaction and spin-spin interaction, which will now be described.

After removing the RF pulse, only the magnetic field is being applied to the spin-system, which results in the spins realigning with the z-direction. As this happens, the net magnetization vector will also realign with the z-direction. Thus, the length of the vector component rotating in the x-y plane decreases causing the measured signal to decrease. Conversely, the vector

component in the z-direction gets longer and so this process is called longitudinal relaxation. During this process, the absorbed energy is released into the surrounding lattice, hence the name spin-lattice interaction. The time taken for longitudinal relaxation of a spin system can be described with the following equation

$$M_z(t) = M_0 \cos(a) + (M_0 - M_0 \cos(a))(1 - e^{-t/T_1}), \tag{1.9}$$

where a is the flip angle and T_1 is the rate of recovery for M_z, which is independent of the magnetic field strength [38].

As the name would imply, in spin-spin interaction the measured signal decreases as a result of the interaction between neighboring spins. As with any other rotating objects, all spins have a certain phase. Immediately after excitation, all spins have the same phase. For simplicity, we can use the rotating frame reference and say that all spins would be pointing in the y'-direction. Thus, the net magnetization would also be in the y'-direction, i.e., $M_{y'} = M_0$. When the spins are far away from each other, they simply precess at the Larmor frequency. However, as spins begin to move close to each other, due to random motion, their own individual magnetic fields will affect the neighbor. Thus, one proton may begin precessing at a faster rate, and the other may precess more slowly. Therefore, all spins gradually dephase causing the net magnetization vector in the x'-y' plane to gradually decrease. When the spins are completely dephased, M_y will equal zero. Since, the absorbed energy is not lost to the lattice but instead exchanged between neighboring spins such that the spins may continue to be oriented in the transverse plane, this process is known as transverse relaxation. The relaxation time can be described with the equation

$$M_{y'}(t) = M_0 \sin(a) e^{-t/T_2} \tag{1.10}$$

where T_2 is the relaxation rate and is independent of the magnetic field strength [38].

Loss of phase coherence also results from another independent mechanism. Spatial variations within the external magnetic field caused by an inhomogeneous applied field and different tissues having different magnetic susceptibilities also cause spins to dephase. The overall dephasing of the spins is given by the equation

$$\frac{1}{T_2^*} = \frac{1}{T_2^+} + \frac{1}{T_2} \tag{1.11}$$

where T_2' is the relaxation time that is caused be the inhomogeneous magnetic field, and the overall relaxation rate T_2^* is a combination of T_2' and T_2 [38].

1.3.3 Magnetic Resonance Imaging

The major advantage that MRI has over most other imaging modalities is that it has excellent soft-tissue contrast [38]. For MRI, three tissue parameters determine the image brightness with respect to tissue type, namely, T_1, T_2, and the proton density [39]. Recall from the previous section that T_1 and T_2 are the longitudinal and transverse relaxation rates respectively. The proton density is simply the number of excitable spins per unit volume, which determines the maximum signal strength that can be emitted for a particular tissue. The values chosen for certain imaging parameters, as will explained below, determines which one of these tissue parameters contributes the most to the image contrast. Therefore, an image can be T_1-*weighted*, T_2-*weighted*, or *density-weigted*. Seeing that the parameter values for tissues are known, one can select the appropriate weighting in order to "best see" a tissue of interest.

The aforementioned imaging parameters that determine the weighting will now be discussed. The repetition time (TR), is the time period between RF excitations of a slice through the volume. The length of TR determines whether an image will have high T_1 weighting or not. To understand this, first note that M_z is considered in determining pixel intensity for a T_1-*weighted* image. When TR is short, the spins of tissues with short T_1 values will fully recover to being oriented in the z-direction after the RF pulse. The corresponding image regions are represented with bright intensities. On the other hand, the spins of tissues with long T_1 values will not fully recover and will be represented with dark intensities. Thus, this image will have high contrast between these two tissues. However, if TR is long, then the spins of both tissues will fully recover resulting in little image contrast between the tissues.

An echo is the time-point when maximum rephasing has occurred after a 180 degree RF pulse. This rephasing is done to undue the dephasing resulting from T_2^* in order to measure the desired quantity T_2. The time period between the RF pulse and this measurement is known as the echo time (TE). The choice of TE determines whether an image will be T_1-*weighted* or not. For this image, the measured signal is proportional to the length of the transverse magnetization vector M_{xy} as described in the previous section. A long TE allows the spins of a tissue with a

short T_2 to become completely dephased. Thus, the corresponding signal measured will be zero and the representative image region will have zero intensity. The spins of a tissue with a long T_2 will not completely dephase, and so the corresponding image region will have a nonzero intensity. Thus, there will be high contrast between these tissues. If TR is chosen to be short, then the spins of neither tissue will completely dephase and the image will have low contrast. For a *density-weigted* image to be created, the T_1 and T_2 weighting have to be low.

So far, we have discussed how the MR signal considered in imaging is created and concepts of image contrast. To create an image, the spins at different spatial locations must be distinguishable. With only the main magnetic field, this is not possible, so mechanisms called gradient coils are used. Three gradient coils, corresponding to the three spatial dimensions, are used to linearly alter the magnetic field. The signal from each gradient is the partial derivative (slope) of the altered magnetic field with respect to position in the corresponding direction. The associated equations are

$$\frac{\partial B_x}{\partial x} = G_x, \qquad \frac{\partial B_y}{\partial y} = G_y, \text{ and} \qquad \frac{\partial B_z}{\partial z} = G_z \tag{1.12}$$

where B_z is the magnetic field experienced only by protons with the same z location. A similar statement can be made for B_x and B_y. Therefore, the magnetic field experienced by all spins at location z is given by

$$B_z = B_0 + zG_z \tag{1.13}$$

This can be seen in Figure 1.4.

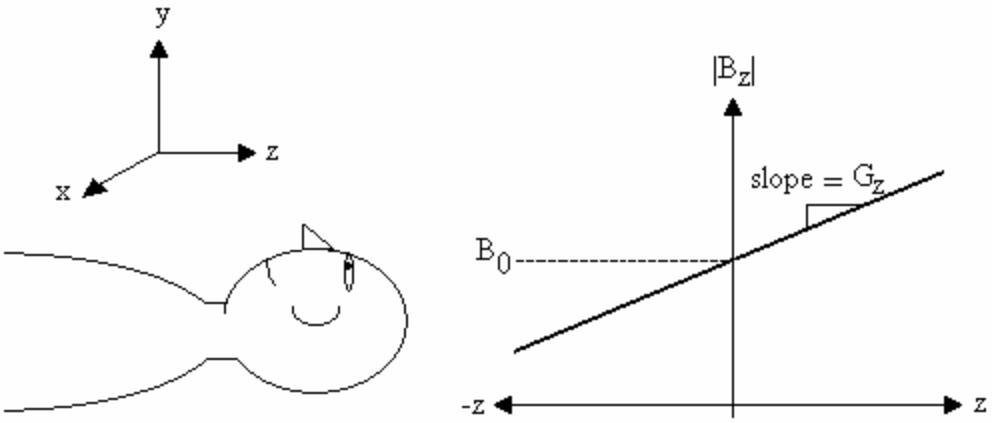


Figure 1.4 – Illustration of common imaging coordinate convention and magnetic field varying linearly in z-direction.

Using these gradient coils, the image is encoded in three steps 1) slice selection, 2) phase encoding, and 3) frequency encoding. Equation 1.1 shows that the precessional frequency of the spins is proportional the magnetic field that the spins are experiencing. Therefore, a certain slice through the volume can be selected by applying an RF pulse with a suitable center frequency. Only spins with frequencies contained in the pulse will be excited, so only a certain slice can be imaged. This slice selection is done by first turning on the gradient of direction orthogonal to the slice-plane of interest. This is explained further in Figure 1.5.

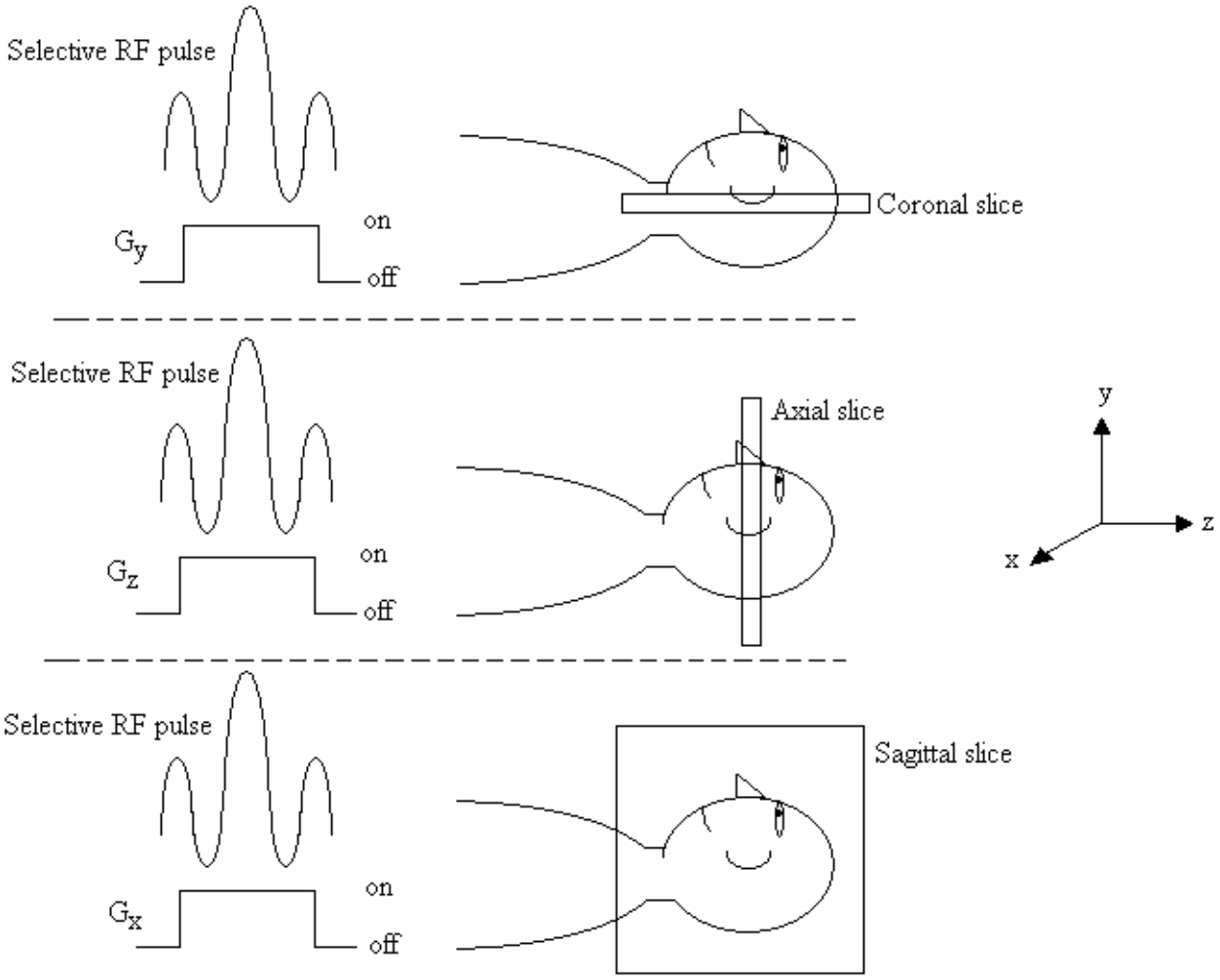


Figure 1.5 – Slice selection via application of 90 degree RF pulse. (top) Excite spin system while y-gradient is turned on to select a coronal slice. (middle) Excite spin system while z-gradient is turned on to select an axial slice. (bottom) Excite spin system while x-gradient is turned on to select a sagittal slice.

The thickness of the selected slice is proportional to the bandwidth of the RF pulse and is given by the equation

$$T = \frac{w_B}{g G_{slice}} \tag{1.14}$$

where T is the slice thickness, $?_B$ is the RF pulse bandwidth, and G_{slice} is the strength of the gradient used for slice selection [38]. The slice thickness can be reduced by decreasing the pulse bandwidth, or increasing the strength of the gradient.

After a slice position has been selected, the other 2 dimensions of the slice must be encoded. This is done by phase and frequency encoding. For phase encoding, the phase encoding pulse is turned on and off prior to data acquisition. For discussion purposes, let us assume that the phase encode direction is the y-direction. The quick pulse of the phase encode gradient causes the phases of the spins to be altered relative to y position, such that spins at a certain y location are distinguishable. For frequency encoding, the corresponding gradient is turned on while data acquisition is occurring. Assuming the x-direction for the frequency encode direction, the precessional frequency of the spins will change relative to their x-position such that the spins are distinguishable along the x-direction.

As a high SNR is very important for qualitative analyses of MRI data, some factors that affect MR SNR will now be discussed. These factors are slice thickness, field of view (FOV) and matrix size, bandwidth, number of excitations (NEX), imaging parameters, magnetic field strength, and the RF coils [39]. There is a tradeoff between SNR and resolution, and this is apparent when considering the slice thickness and the slice matrix size relative to the FOV. The resolution is closely related to voxel-size. Small voxels allow for small objects to be within close proximity of each other and still be discernible. However, the SNR is inversely proportional to the voxel size, since a smaller voxel will contain fewer spins. The slice thickness, expressed in mm, is how thick the voxels of a slice are. The matrix size must be related to the FOV in order to determine the voxel lengths along the other 2 dimensions. The FOV states the height and width of the slice in mm. The bandwidth of a system is the spectrum of frequencies that an MR system can encode. A larger bandwidth allows for faster imaging, but reduces the SNR because more noise at the outskirts of the spectrum is incorporated. The NEX shows the number of times the signal is measured to facilitate signal averaging. Signal averaging increases the SNR, but the

additional measurements increase scan time. We discussed how the imaging parameters TR and TE affect image contrast. However, there is a tradeoff between contrast and SNR. Longer TR increases the SNR but reduces contrast in T_1-*weighted* images. Shorter TE increases the SNR but reduces contrast in T_2-*weighted* images. A high magnetic strength causes the spins to recover to the z-orientation at a faster rate, and so increases the SNR. Thus, a high-strength magnet can be used with a large bandwidth for high-speed imaging. Many of the previously mentioned factors have negative tradeoffs for increased SNR, or they are fixed for a given scanner. The choice and appropriate use of RF coils can greatly improve the SNR without any tradeoff. In making this choice, the goal is to have the coil be as close as possible to and surround the organ(s) of interest.

1.3.4 Diffusion Tensor Imaging

Diffusion Tensor Imaging was introduced in mid 1980s as a means of measuring the microscopic phenomenon of Brownian motion on a macroscopic scale. The thermal energy carried by the molecules results in this random translational motion [11]. The time-distance relationship of a particle undergoing diffusion can be described as

$$D = \frac{1}{6t} r^2 \tag{1.15}$$

where D is known as the diffusion coefficient, t is the time period of the diffusion, and r is the net diffusion.

Nuclear Magnetic Resonance can be used to measure the average diffusion and anisotropy without interfering with the diffusion process [11]. The signal measured by the scanner is attenuated and can be described as

$$A = \exp(-bD), \tag{1.16}$$

where b corresponds to the gradients of the scanner. On the macroscopic scale (on the order of mm), diffusion is likely to be impeded by impermeable obstacles resulting in anisotropic behavior. In the case of anisotropic diffusion, the coefficient becomes a second order tensor describing diffusion along all dimensions, given by

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix}, \tag{1.17}$$

with this tensor being symmetric. The tensor generally depends on particle mass, temperature, and the structure of the medium. In DTI, the temperature and mass of water molecules are assumed to be constant so that the tensor measure can be used to estimate the structure of the medium.

The process of determining the diffusion tensor begins with acquiring a set of diffusion weighted images S_i (where $i = 1,2,\ldots n$) and an image acquired without diffusion weighting S_0. The diffusion weighted images are acquired with a diffusion sensitized pulse sequence such as echo-planar imaging (EPI), which is the fastest MRI pulse sequence. EPI involves using a single RF excitation to obtain full k-space (frequency domain) sampling for a given slice. Thus, an entire volume can be acquired in a few seconds [38]. This high imaging speed is necessary for measuring diffusion. A minimum of $n = 6$ diffusion weighted images are required for all unique tensor elements, and each diffusion weighted image is created using a gradient pulse sampled in the direction of the corresponding tensor element. More than 6 diffusion weighted images can be acquired in order to reduce the amount of noise, present in the diffusion weighted images, that propagates into the tensor image. For arbitrary n, the equation used to solve for the unique tensor components is

$$\begin{bmatrix} x_1^2 & y_1^2 & z_1^2 & 2x_1^2y_1^2 & 2y_1^2z_1^2 & 2x_1^2z_1^2 \\ x_2^2 & y_2^2 & z_2^2 & 2x_2^2y_2^2 & 2y_2^2z_2^2 & 2x_2^2z_2^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^2 & y_n^2 & z_n^2 & 2x_n^2y_n^2 & 2y_n^2z_n^2 & 2x_n^2z_n^2 \end{bmatrix} \begin{bmatrix} Dxx \\ Dyy \\ Dzz \\ Dxy \\ Dyz \\ Dxz \end{bmatrix} = \begin{bmatrix} \frac{-1}{b} \ln \frac{S_1}{S_0} \\ \frac{-1}{b} \ln \frac{S_2}{S_0} \\ \vdots \\ \frac{-1}{b} \ln \frac{S_n}{S_0} \end{bmatrix} \tag{1.18}$$

where the leftmost matrix is the encoding matrix containing the unit normalized gradient measurement directions [36]. When this matrix is not square, its inverse does not exist and the solution must be found using the pseudo-inverse via singular value decomposition.

On a voxel-by-voxel basis, diffusion can be characterized by an ellipsoid, where the dimensions of the ellipsoid correspond to the eigenvectors of the tensor. The eigenvector associated with the principal eigenvalue, points in the direction of maximum diffusion [11]. The ellipsoid is convenient for visualizing the diffusion represented in each voxel, and three primary features can be extracted that are related to the ellipsoid. These features can be used to differentiate between tissues. The apparent diffusion coefficient (ADC), also called the mean diffusitivity, is a measure of ellipsoid size [11]. This measure is rotational invariant, and is given

by the trace of the tensor divided by 3 [1]. Anisotropy is a measure of the eccentricity of the ellipsoid, and hence, how diffusion differs with respect to different dimensions. Three similar rotational invariant indices of this measure are relative anisotropy (RA), volume ratio (VR), and the widely used FA. As FA was used in this research, it will be defined here as

$$FA = \frac{\sqrt{3[(l_1 - \langle l \rangle)^2 + (l_2 - \langle l \rangle)^2 + (l_3 - \langle l \rangle)^2]}}{\sqrt{2(l_1 + l_2 + l_3)^2}}, \tag{1.19}$$

where

$$\langle l \rangle = (l_1 + l_2 + l_3)/3, \tag{1.20}$$

and ?$_i$ is an eigenvalue of the tensor. Notice that $\langle l \rangle$ is the ADC. The third feature is simply the main ellipsoid axis given by the eigenvector of the principal eigenvalue. As previously stated, this indicates the direction of maximum diffusion.

1.4 Literature Review

1.4.1 ACR Accreditation Program

The ACR accreditation program uses information obtained from clinical and phantom images to assess overall image quality of a scanner. For the clinical assessment portion of the program, the site is required to submit example images of the brain, cervical-spine, lumbar-spine and knee. The assessment of these images is qualitative. For the phantom assessment portion of the program, a human observer employed by the ACR performs quantitative manual tests on T_1 and T_2-*weighted* volumes of a provided phantom. The ACR phantom is a cylindrical and acrylic-plastic object designed to be imaged using the scanner's head-coil. It has an end-to-end length of 148 mm and a diameter of 190 mm. On its interior, the phantom has certain objects that allow for the associated tests to be performed. In hollow spaces, it is filled with a solution of nickel chloride and sodium chloride. The phantom assessment portion is being used in this research.

For the phantom assessment portion, the ACR requires that five images be submitted. The first is a single localizer image, which is simply a sagittal slice through the center of the phantom. This localizer is a 20 mm thick spin-echo slice that displays the locations of the various objects in the phantom. The other four images are split between corresponding to T_1 and T_2

scanning sequences. These are all comprised of eleven slices of thickness 5 mm that are spaced 5 mm apart [4]. Two sets are to be taken with the specific scan protocols prescribed by the ACR with T_1 and T_2 scan sequences respectively and the other two are similar with the difference being that they are to be taken with the site's protocols. Figure 1.7 shows the localizer and the axial slices.

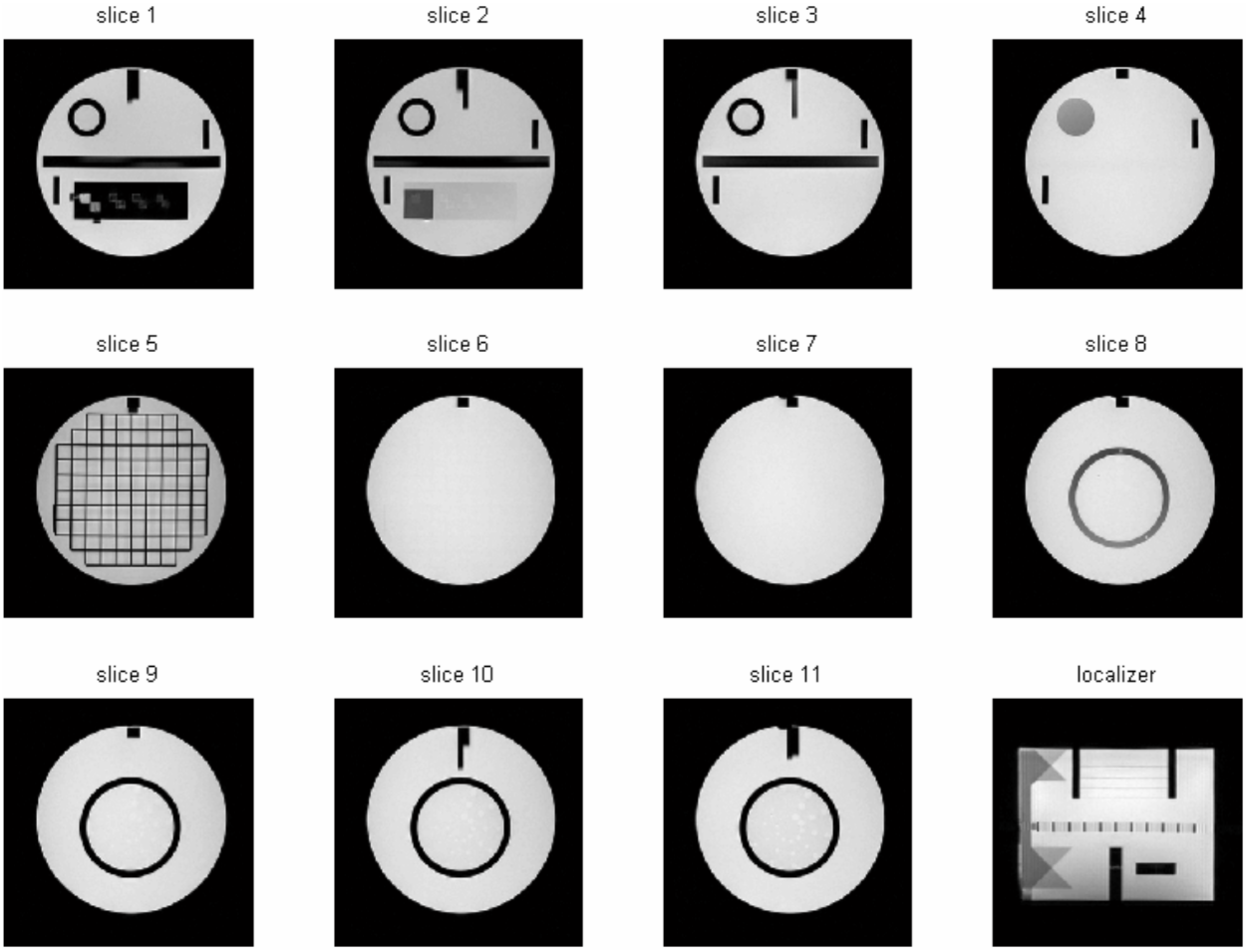


Figure 1.6 – Series used to perform ACR accreditation tests. Slices 1 through 11 are the axial slices of a T_1 series. The sagittal localizer slice is shown in the bottom right image.

The phantom portion of the ACR accreditation program is comprised of seven tests. Three of these pertain to quantitative-imaging quality assurance, namely Geometric Accuracy, Slice Thickness Accuracy, and Slice Position Accuracy [6]. The tests associated with qualitative-imaging are High Contrast Spatial Resolution, Image Intensity Uniformity, Percent Signal

Ghosting, and Low Contrast Object Detectability [3]. These tests will now briefly be described. Refer to Appendix C to view the actual steps for this manual procedure.

Geometric Accuracy

The Geometric Accuracy test is performed to assess whether geometric distortion has occurred during the scan process. The phantom dimensions in the image should be similar to the true dimensions. The ACR procedure, for this test, is to measure phantom dimensions in the localizer as well as slice 1 and 5 of the axial series. The procedure begins by the observer applying the minimum intensity window. The level is then set to a value that is approximately half the mean intensity corresponding to solution only regions of the phantom. This procedure is done adaptively. An onscreen tool is used to perform the measurements by inspection. The failure of this test can be caused by miscalibrated gradients. The gradients shift over time and require recalibration. Also, an acquisition bandwidth that is too low can cause spatial distortions. This bandwidth may be set low to improve SNR.

High Contrast Spatial Resolution

The High Contrast Spatial Resolution test determines the scanners ability to resolve small high-contrast objects within close proximity of each other. A resolution insert is located in slice 1. This insert is comprised of three pairs of hole arrays, which are the items that are actually considered in this test. These pairs vary in the size of the holes that they are comprised of. To perform the assessment, the observer magnifies the section of the slice that contains the insert. He or she then inspects the holes to see if they are distinguishable from their neighbors. The failure of this test may be caused by excessive filtering, which may be done to improve image appearance. Excessive image ghosting can also cause a failure, but this ghosting would be detected by the Percent Image Ghosting test to be described later. Also, large geometric errors, resulting from gradient miscalibration, can cause this test to be failed.

Slice Thickness Accuracy

In the Slice Thickness Accuracy test, the observer views the slice thickness insert located in slice 1 in order to determine whether the prescribed slice thickness was achieved. Two intensity ramps are located in this insert, which are crossed with one having a positive slope and

the other having a negative slope. The observer would adjust the magnification, intensity window, and level so that the ramps are distinct. The on-screen measurement tool is then used to determine the lengths of the ramps. The slope of the intensity ramp is 10:1. Therefore, the intensity ramps will show up in the slice with a length that is 10 times the thickness of the slice. A cause of failure for this test might be RF amplifier non-linearity, which would result in distorted RF pulses. Also, extremely poor gradient calibration and switching can cause a failure.

Slice Position Accuracy

In the Slice Position Accuracy test, slices 1 and 11 are considered to determine if the slices positions are as they were prescribed. The position of slices 1 and 11 are to be prescribed so that they pass through the vertices of the crossed wedges of the phantom. These crossed wedges show up as 2 long vertical bars at the top of the phantom in each slice. When these slices are positioned accurately, the bars will be of equal length. The observer uses the on-screen tool to measure the difference in length between the adjacent bars. The most common reason why this test fails is that the slice positions were prescribed incorrectly prior to scanning. Also, the automatic patient-table positioning mechanism may need adjusting. Lastly, a failure may be caused by very bad gradient or B_0 calibration.

Image Intensity Uniformity

In the Image Intensity Uniformity test, solution only regions are considered to determine whether there is relatively uniform intensity throughout. The observer determines the average intensity value of small regions of interest (ROI) about the minimum and maximum intensity locations in the image. The maximum and minimum pixels are not used because of considerations of noise in the image. These average values are then used to calculate the percent integral uniformity (PIU). One cause of failure may be that the phantom is mispositioned in the head coil. The phantom should be centered in the head coil because the coil's center has the most uniform sensitivity. Other causes of failure might be malfunctioning head coil components or image ghosting.

Percent Signal Ghosting

The Percent Signal Ghosting test determines if a phase encode shadow is present in the image. This phase encode shadow would be comprised of faint copies of the phantom superimposed along the phase encode direction of the image. For this test, 5 ROIs are considered. The central ROI is the largest and is set to include most of the phantom. Then, 4 ROIs are situated at the compass directions relative to the phantom slice i.e. top, bottom, left, and right. Only 2 of these exterior ROIs will contain a possible ghost portion. The other 2 will contain only the background. The evaluator uses the mean intensity values from these ROIs to calculate the ghosting ratio for the image. A failure of this test can be caused by motion or vibration of the phantom during data acquisition. Otherwise, the cause would be instability in the measured signal between pulse cycles, which can be associated with the receiver coil, transmitter coil, or gradient coil subsystems.

Low Contrast Object Detectability

The Low Contrast Object Detectability test assesses the number of low contrast objects that can be detected by an observer. The test is performed on slices 8 through 11. Each of these slices contains 10 spokes of circles spread radially around the phantom. Each spoke contains three circles of the same size. The circles decrease in size moving from spoke to spoke moving in a clockwise direction. Also, the contrast of the circles on each slice is constant, and the overall contrast increases moving from slice 8 to 11. The observer counts the number of complete spokes distinguishable in these slices. The most common cause of failure is incorrect slice positions. Also, a failure can be caused by the phantom being mispositioned or ghosting artifacts. If none of these factors were the cause, then the SNR might be too low.

1.4.2 ACR Phantom Analysis Tool

In fulfillment of a MSE Graduate Capstone project requirement at the Milwaukee School of Engineering, Steve Huff created an automated QA procedure that involved the ACR accreditation program. The system developed was called the American College of Radiology Phantom Analysis (ACR-PA) tool [24], and it used simple and effective image processing techniques to replace the human measurement process. This project had a very broad scope, and the system included data acquisition, image analysis, test failure recommendations, and system-

error descriptions. Verification testing results were included in the paper, and the system showed favorable results in comparison with the alternative of human observers. All aspects of the system could be controlled by the user. In fact, the paper gives much detail about user actions.

The approaches taken for the quantitative-imaging tests implemented in the ACR-PA tool were similar to the approach of this author, which will be discussed in chapter 2. These involved performing the required length measurements using line profiles. Line profiles simply involve extracting the pixel values along a specified line in the image. There was one difference between both approaches for the quantitative imaging tests. In the ACR-PA tool, the line profiles are applied to the original slice. The approach being used here applies the line profile after the slice has been binarized. Therefore, the desired length measurement can be determined by simply summing the corresponding line profile. The approach taken for the High Contrast Spatial Resolution test by Huff involved first creating 3 crops (sub-images) containing the three resolution array pairs. Then an iterative thresholding procedure was used to binarize the images in order to maximize the number of holes detected. The algorithm then scans the relevant row or column counting "0" to "1" transitions. These transitions indicate the number of holes in a given row or column. My approach for this test is similar as will be seen in the next chapter. For the Image Intensity Uniformity test, my approach is also similar to that of the ACR-PA tool. The ACR-PA tool uses a Gaussian filter to remove noise that would corrupt the process of determining max and min intensity locations. The approach taken by Huff for the Percent Signal Ghosting test involved a cross-correlation measurement that produced images that were quantified for peak, mean, and integral ghost measurements. The purpose of developing this new assessment method was to distinguish between "smeared" and correlated ghost artifacts.

However, the ACR-PA tool had two limitations. The first limitation was that the procedure required an associated phantom holder. The holder was constructed such that the phantom could be placed in a consistent position (within +/- 1 mm) prior to imaging. The holder also included a bubble level that would facilitate accurate rotational alignment. The creation of this phantom holder is an achievement. The limitation is that the procedure requires the use of this particular holder. A better solution would not require any holder.

The second limitation is that the procedure required substantial user interaction. Even with the use of the phantom holder, the procedure requires a one-time user-initiated coordinate transformation because of slight mechanical differences in the bridge, cradle, table, and

alignment lights between scanners. Before the system is used for the first time on a certain scanner, the user must display a sagittal and axial slice and specify certain landmarks to facilitate the coordinate transformation. This introduces potential human error into the procedure, and requires possibly undesired human involvement. The approach taken by the ACR-PA for the Low Contrast Object Detectability test involved creating edge enhanced images of the associated axial slices. The enhanced and the original images were displayed adjacent to each other, and a human observer was then required determine what objects were detectable. Therefore, this test was not automated.

1.4.3 Motion Estimator

There are generally two types of motion estimation algorithms. The algorithms that fall into the first category are feature based. These algorithms need to detect some sort of features such as landmarks, surface, or contours before registration can take place. The downfall with these methods is that speed and accuracy are greatly dependent on feature selection. The algorithms of second type of motion estimation procedures use the actual image data to directly determine the transformation. The benefit of this procedure is that high accuracy and speed can be achieved.

The motion estimation algorithm being used in this thesis, developed by Baba C. Vemuri et al. [37], falls into the second category. High speed and robustness were achieved by using a B-spline optical flow network and modified Newton iterative method for computational efficiency. The major contribution of this paper was the development of a novel modified Newton iteration method, in which, the Hessian is precomputed at the optimum solution before the solution is found. Since the Hessian does not have to be computed at each iteration, the algorithm is exceedingly fast and has a larger convergence range. The paper claimed accurate results for intra-modality registration of medical image data. The method has applicability to 2-D rigid, 2-D affine, and 3-D affine cases, but 2-D rigid registration was sufficient for this work.

Considering a reference image I_1 and a displaced image I_2, the paper defines the motion difference between the images using a flow field given by

$$I_2(x_i + u_i, y_i + v_i) = I_1(x_i, y_i) \tag{1.21}$$

where (x_i, y_i) denote a given pixel location, (u_i, v_i) is the flow field at that location, and i is the pixel index. For computational efficiency, this flow field is reduced to a specified number of

control points using bilinear basis functions. The following equations show the relationship between the flow field at each pixel location and the control points $\hat{u}_j$ and $\hat{v}_j$

$$\begin{aligned} u(x_i, y_i) &= \sum_j \hat{u}_j w_{ij} = \sum_j \hat{u}_j B_j(x_i, y_i) \\ v(x_i, y_i) &= \sum_j \hat{v}_j w_{ij} = \sum_j \hat{v}_j B_j(x_i, y_i) \end{aligned} \tag{1.22}$$

where j is the control point index. In the above equation, $w_{ij} = B_j(x_i, y_i)$ are the basis functions, which are given by $B(x, y) = (1-|x|)(1-|y|)$, and both x and y range between -1 and 1. For 2-D rigid flow, the entire flow field can be described with a small set of parameters, which are placed in a motion vector as $\mathbf{T} = [f, d_1, d_2]^T$ where f denotes the rotation angle and d_1 and d_2 denote the x and y translations respectively. Then, the rigid flow model is defined as

$$\begin{bmatrix} u(x, y) \\ v(x, y) \end{bmatrix} = \begin{bmatrix} \cos j & -\sin j \\ \sin j & \cos j \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} - \begin{bmatrix} x \\ y \end{bmatrix}. \tag{1.23}$$

The expectation of squared differences error criterion is used and is defined as

$$J\{E_{sd}(\hat{\mathbf{T}})\} = E\left\{ I_1\left(\mathbf{X}_i + \sum_j w_{ij}(\mathbf{T}_j \mathbf{T}' - \hat{\mathbf{X}}_j) \right) - I_2(\mathbf{X}_i) \right)^2 \right\} \tag{1.24}$$

where $E\{\}$ denotes the expectation operation, $\mathbf{X}_i = (x_i, y_i)$, $\hat{\mathbf{T}}$ pertains to the control points, and $\mathbf{T}' = [\cos f, -\sin f, -d_1, \sin f, -\cos f, -d_2]^T$.

The modified Newton optimization method is used to minimize the error criterion defined above. This paper proposes a method that precomputes the Hessian at the optimum using innovations (incremental transformations) $\tilde{\mathbf{T}}^k$. For the 2-D rigid-motion case, the basic steps of algorithm were given in the paper as follows.

1. Precompute the Hessian at the optimum $\tilde{\mathbf{H}}$ using equation (1.25)

$$\tilde{\mathbf{H}} = 2E\left\{ \mathbf{A} \frac{\partial I_1(\mathbf{X})}{\partial(\mathbf{X})} \left(\frac{\partial I_1(\mathbf{X})}{\partial(\mathbf{X})} \right)^T \mathbf{A}^T \right\} \tag{1.25}$$

where

$$\mathbf{A} = \left(\frac{\partial h(\mathbf{X}, \mathbf{T})}{\partial \mathbf{T}} \right) = \begin{bmatrix} -\sum_j w_j \hat{y}_j & \sum_j w_j \hat{x}_j \\ -\sum_j w_j & 0 \\ 0 & -\sum_j w_j \end{bmatrix} \tag{1.26}$$

and h denotes a transformation of image coordinates.

2. Begin iterations and at iteration k, compute the gradient vector using the following equation

$$\mathbf{g}(\hat{\mathbf{T}}^k) = 2E\left\{(I_m - I_2)\mathbf{N}\mathbf{M}^{-T}\frac{\partial I_2(\mathbf{X})}{\partial \mathbf{X}}\right\} \tag{1.27}$$

where

$$\mathbf{M} = \frac{\partial \mathbf{X}^k}{\partial \mathbf{X}} = \begin{bmatrix} 1+\sum_j (w_j)'_x \hat{u}_j^k & \sum_j (w_j)'_x \hat{u}_j^k \\ \sum_j (w_j)'_x \hat{v}_j^k & 1+\sum_j (w_j)'_x \hat{v}_j^k \end{bmatrix} \tag{1.28}$$

and

$$\mathbf{N} = \left(\frac{\partial \mathbf{X}^{k+1})}{\partial \tilde{\mathbf{T}}^{k+1}}\right)^T |_{\tilde{\mathbf{T}}^{k+1}=0} = \begin{bmatrix} -\sum_j w_j \hat{y}_j^k & \sum_j w_j \hat{x}_j^k \\ -\sum_j w_j & 0 \\ 0 & -\sum_j w_j \end{bmatrix}. \tag{1.29}$$

3. Compute the innovation for the next iteration $\tilde{\mathbf{T}}^{k+1}$ using the following equation

$$\tilde{\mathbf{T}}^{k+1} = -\tilde{\mathbf{H}}^{-1}\mathbf{g}(\hat{\mathbf{T}}^k). \tag{1.30}$$

4. Update the motion parameter $\hat{\mathbf{T}}^{k+1}$ for the next iteration using the following equation

$$\hat{\mathbf{T}}^{k+1} = \begin{bmatrix} \hat{j} \\ \hat{d}_1 \\ \hat{d}_2 \end{bmatrix}^{k+1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\hat{j} & -\sin\hat{j} \\ 0 & \sin\hat{j} & \cos\hat{j} \end{bmatrix}^{k+1} \begin{bmatrix} \hat{j} \\ \hat{d}_1 \\ \hat{d}_2 \end{bmatrix}^k + \begin{bmatrix} \tilde{j} \\ \tilde{d}_1 \\ \tilde{d}_2 \end{bmatrix}^{k+1}. \tag{1.31}$$

The algorithm stops whenever there is little change in $\hat{\mathbf{T}}^k$.

1.4.4 DTI phantoms

The primary component of a QA procedure for any medical imaging modality is a suitable physical phantom that models the physical phenomena measured by the modality. In this preliminary attempt, I was limiting the scope to the primary DTI features that are considered for diagnosis and discriminating between tissues, namely, diffusion anisotropy and tissue microstructure based on principal diffusion direction. Therefore, a phantom was desired that modeled restricted diffusion, which occurs on the interior of axons. It was also desired that this phantom be inorganic such that it would not degrade over time.

Three sources were found that described similar phantoms meeting these requirements. As part of the work for his master's thesis [36], Alex Taylor created a physical phantom to evaluate the performance of tractography algorithms, which are used to estimate tissue microstructure. In creating this phantom, Taylor used polytetrafluoroethylene (PTFE) microbore capillary tubing with an inner dimension (ID) of 0.305 mm and an outer diameter of 0.702 mm. Multiple segments of the tubing were cut, filled with water, and assembled into sheets with a 90 degree crossing pattern. The capillary sheets were placed in a small plastic container and surrounded by gelatin to mitigate air related susceptibility artifacts in the images.

Lin et al. [25] created a phantom to be used to compare the effectiveness of DTI and Diffusion Spectrum Imaging (DSI) for correctly determining the orientation of crossed axonal fibers. Lin used PTFE ultramicrobore tubing with ID 0.05 mm and OD 0.35 mm. Segments of the tubing were filled with water and assembled into sheets. Layers of these sheets were stacked at 90 and 45 degrees in reference to each other in an interleaved fashion. The structures were then secured to a firm plastic plate.

Hagen et al. [21] were the first individuals to evaluate the effectiveness of DTI for determining fiber orientation using a physical model. This phantom also consisted of PTFE ultramicrobore tubing. The capillaries were filled with water by a gluing a 22 ½ - gauge needle to each segment. After being filled, the capillaries were sealed by melting both ends and removing the needle. The capillaries were placed in three different orientations, namely, aligned, coiled, and random and placed in placed inside borosilicate glass tubes. It was stated in this paper that the capillary ID is roughly twice that of the larger nerve fibers. It was also stated that the outer diameter of the tubing is much greater than radial myelin thickness, so the density of water in tissues would be much greater than it is in this phantom structure. Thus, a much lower signal would be measured from the phantom.

It is difficult to decide that one of these phantoms is better than any other. The results sections of these sources implied that the phantoms were suitable for the particular purpose. The smaller ultramicrobore tubing is closer to the size of actual axons, but far more difficult to fill with water. Also, the image resolution would have to be increased, which decreases the SNR.

1.5 Chapter Summary

In this chapter, the significance of this thesis was discussed. The automated ACR accreditation program was chosen for an MRI QA procedure because 1) it incorporates tests pertaining to quantitative and qualitative imaging, 2) the procedure can be used to pre-qualify data prior to submission for accreditation, and 3) an automated procedure would remove most of the discomfort that deters regular QA. To this date, no QA procedure has been reported for DTI. Theoretical background including definitions and assessments of image quality, nuclear magnetism, MRI, and DTI were provided. A review of pertinent literature, which included the ACR accreditation program, accomplishments and limitations of prior work in automating this program, motion estimation, and DTI phantoms, was discussed. The automated tests for both the MRI and DTI QA procedures will be described in detail the next chapter along with the experimental methods taken in developing the DTI QA procedure.

Chapter 2 Methods

2.1 ACR Quality Assurance

The automated ACR procedure was implemented such that each of the 7 tests could be executed independently or the total procedure could be executed in full. To better understand the procedural flow, refer to Appendix B. Generally, image registration is a prerequisite for the success of the tests unless great care has been taken in positioning the phantom prior to imaging. Where important, I made an effort to keep the steps of the automated procedure as close as possible to those of the manual ACR procedure. Whenever, an implemented step was a departure from the manual procedure, the decision was made that the difference between results of both methods would be negligible. The major departure taken from the manual procedure is that Otsu's method of automatic thresholding [27] is used to segment the objects of interest from the background as opposed to performing intensity window and level adjustments. In this section, the approaches taken to automate the ACR tests will be described. Refer to section 1.4.1 for details of the ACR accreditation program.

2.1.1 Image Registration

Unless great care has been taken in positioning the phantom prior to scanning, the candidate series must be registered with a standard series prior to executing the automated tests, since the location of the objects of interest for the tests are known for this standard series. The algorithm explained in section 1.4.3 is used for performing rigid motion estimation. The term motion used here is used to refer to the translational and rotational differences between the candidate and the standard series. The sagittal motion is determined by applying the algorithm to the localizer slices of the candidate and standard series. The procedure for determining the axial series motion is a bit more involved. It was discovered that motion estimation using axial slices 1, 2, and 3 provided accurate results, which is because of the high level detail present in these slices. Other slices were more isometric, which caused the motion estimation to be inaccurate. Therefore, the motion vectors determined for slices 1 through 3 are averaged to produce the axial motion vector. All images are shifted and scaled so that their intensities ranged between 0 and 1

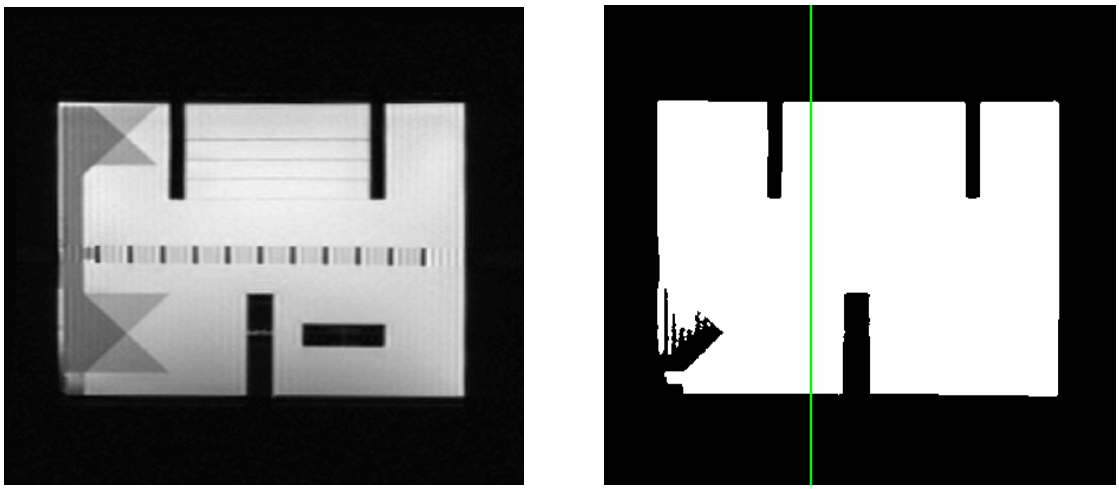
in performing the motion estimation procedure. The axial and sagittal motion vectors are then used to perform rigid registration on the sagittal and axial slices of the candidate series.

2.1.2 Geometric Accuracy Test

The Geometric Accuracy Test involves performing length measurements on the sagittal localizer and axial series slices 1 and 5. Before acquiring the necessary measurements, the images are enlarged by a factor of 3 to allow for sub-voxel accuracy. That is, the number of rows and columns are increased by a factor of 3, and bicubic interpolation is used. Also, the same segmentation procedure is performed on each slice. The steps of the procedure are as follows:

1. The phantom is segmented from the image. First, the voxel intensities are shifted and scaled so that they range from 0 to 1. Then, the phantom regions are segmented from the background using Otsu's method of binarization. Then, objects that consist of 6 or fewer voxels are removed from the image. Finally, holes within foreground objects are filled with an algorithm based on morphological construction [35].

2. The segmentation procedure result for the sagittal localizer is shown in Figure 2.1. After segmentation, the column highlighted in green is summed to provide the length measurement along this direction. The actual measurement considered is this sum minus 1.

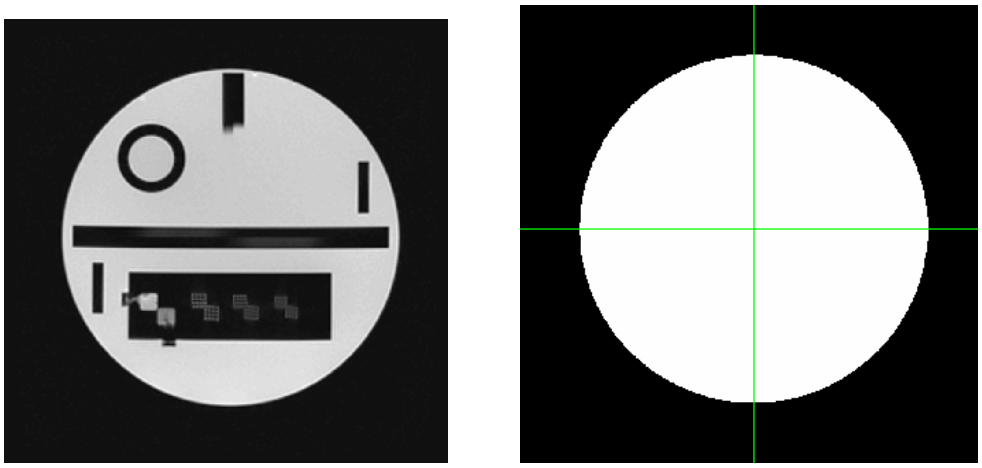


(a) (b)

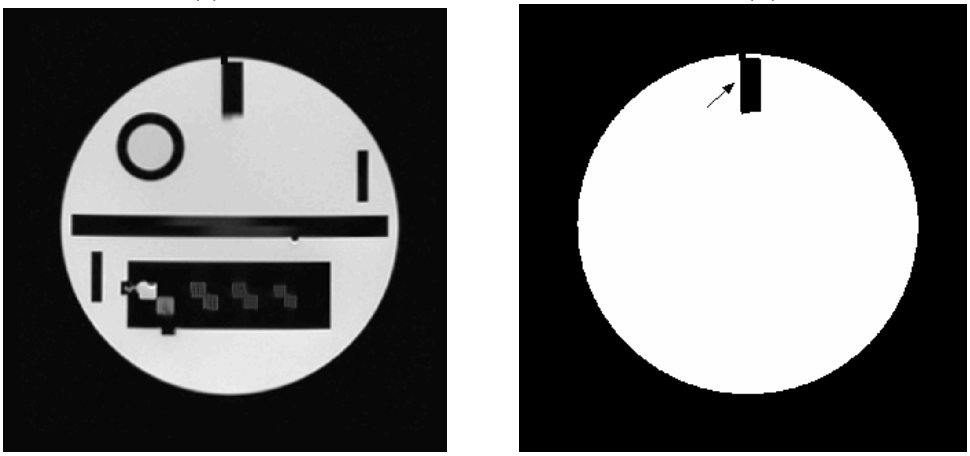
Figure 2.1 – Illustration of segmentation procedure for sagittal localizer. (a) Original image. (b) Binarized sagittal localizer. Green line shows column that is summed for measuring length.

Figure 2.2 (a) and (b) show a likely pair of original slice and segmentation result along with rows and columns that measurements should be performed on for axial slice 1. From this example, one might assume that the measurement procedure used for the sagittal localizer is also applicable here. However, Figure 2.2 (c) and (d) shows images for which this procedure would fail to give an accurate measurement. Since the phantom wall is relatively thin, the edge of the phantom region near the top of the image may be discontinuous. The corresponding binary image would also be discontinuous and so the hole indicated by the arrow in Figure 2.2 (d) will not be filled. Therefore, a more robust measurement procedure is employed for this slice.

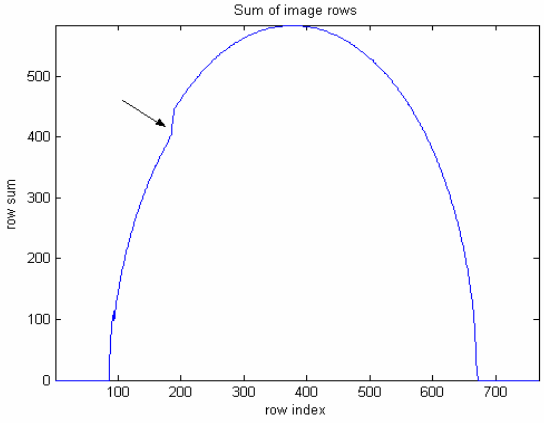
3. To measure the diameter length along the row dimension, the binary image is summed along the column dimension producing an array whose elements are the sum of all intensity values in each row. A plot of this array is shown in Figure 2.2 (e). All nonzero elements in this array are summed to give the desired length measurement with the actual measurement considered being this sum minus 1. A similar procedure is used to measure the diameter along the column dimension with the difference being that the corresponding array is comprised of column sums.



(a) (b)



(c) (d)

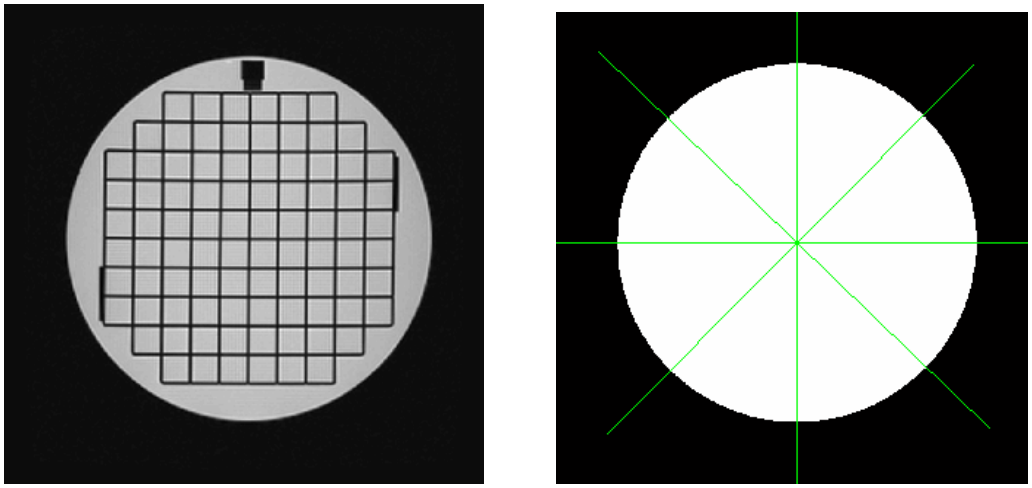


(e)

Figure 2.2 – Illustration of segmentation and measurement procedure for slice 1. (a) Original image. (b) Result of binarizing image in (a). Green lines indicate rows and columns prescribed by ACR for length measurement. (c) Example of slice with discontinuous phantom edge. (d) Result of binarizing image in (c). Notice that hole indicated by arrow has not been filled. (e) Plot

of sum of rows of image in (d). The presence of the hole is evident in this plot and is indicated in the arrow.

4. The diameters of slice 5 are measured at varying degrees as shown in Figure 2.3 (b). For the 0 and 90 degree measurements, the procedure described for slice 1 is also used. To perform diameter measurements at 45 and -45 degrees from the x-axis, line profiles are extracted from the binary image using nearest-neighbor interpolation. The sums of these profiles minus 1 are taken to be the diameter measurements.



(a) (b)

Figure 2.3 – Illustration of segmentation and measurement process for slice 5. (a) Original image. (b) Segmented image showing degrees at which diameter is measured.

All voxel measurements are converted to mm and compared to the true lengths. The true length for the sagittal localizer is 148 mm, and the true diameter of the phantom is 190 mm. The test is determined to have failed if any measurement differs from the standard by an absolute value of more than 2.

2.1.3 High Contrast Spatial Resolution

The phantom contains a resolution insert, which is a small block of plastic that has water filled holes drilled into it. These holes show up as bright spots in the corresponding region of slice 1 as shown in Figure 2.4. Three pairs of not quite square hole-arrays are considered in this test, and these are the left pair, center pair, and right pair. Each hole-array pair is created such that it consists of an upper-left (UL) array and a lower-right (LR) array that both share one hole in common where they meet. The UL array is used to assess resolution along the column (left-right) direction. Each row is offset slightly to the right of the one above it. This is to ensure that at least one row has holes that fall sufficiently in the voxels. The LR array is used to assess resolution along the row direction. The columns of this array are offset such that each is slightly lower than the one to the left. Again, this is to ensure that at least one column has holes that fall sufficiently in the voxels. An illustration of a typical hole-pair array is shown in Figure 2.5 (a). The hole diameters are 1.1, 1.0, and 9 mm for the left pair, center pair, and right pair of arrays respectively. In each case, the center-to-center spacing is twice the diameter. The steps of the procedure are as follows:

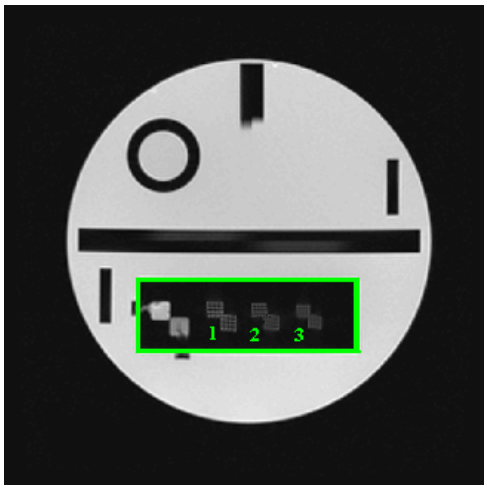
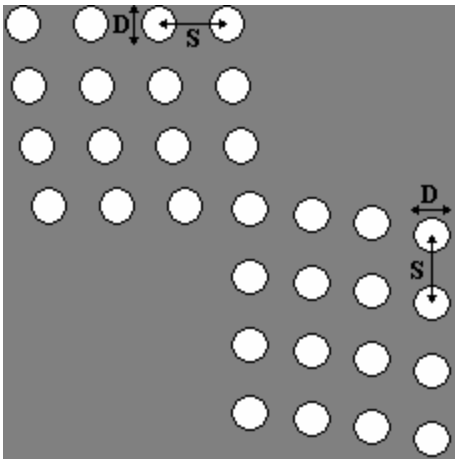


Figure 2.4 – Slice 1 with resolution insert enclosed in green box. Three hole-array pairs are considered in this test, namely, 1) left-pair, 2) center-pair, and 3) right-pair.

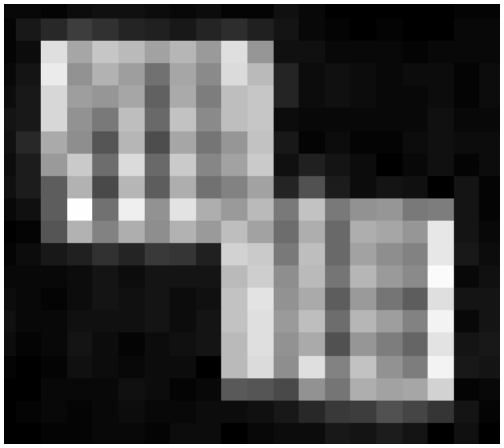
Even though the candidate series would have already been registered, it would not be exactly aligned with the standard set. Instead, there will be an offset of at least a fraction of a

voxel, and a voxel has a width of about 0.965 mm. Since the largest hole-array pair has a width of only about 9.9 mm, this offset is substantial, and further registration is necessary.

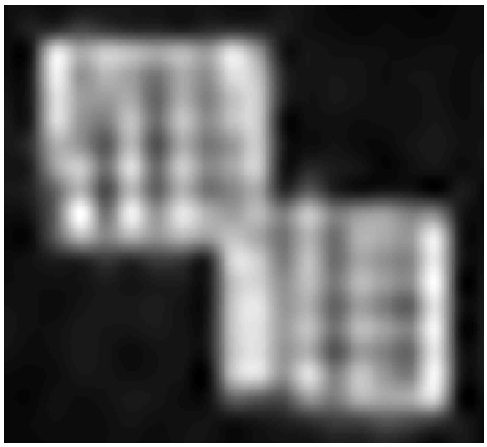
1. The first step in processing a hole-array pair would be to crop it from the original image, and then scale the crop by a factor of 10 using bicubic interpolation. As evident from Figure 2.5 (b), without interpolation, it is quite difficult to discern the holes. The candidate crop must then be registered with the corresponding crop of the standard set with a procedure similar to that described in section 2.1.1. After registration is complete, the locations of the holes are known and further processing can commence.



(a)



(b)



(c)

Figure 2.5 – Hole-array pair. (a) Illustration of hole-array pair. The center-to-center spacing between holes (S) is twice the hole diameter (D). (b) Crop of left hole-array pair. (c) Image in b resized by factor of 10 with bicubic interpolation.

My approach for automating this test was based upon the steps taken by the human observer in the manual procedure. The observer adjusts the display level and window settings until one row/column has 4 distinct holes. The observer would set the window to be narrow, and then would continually adjust the level until satisfied that either 4 distinct holes in a row/column were visible or that there was no possibility of discerning 4 distinct holes. Generally, when the holes are large and merge with adjacent holes, the observer would be inclined to raise the level setting. Alternatively, if less than 4 holes are visible but are distinct and relatively small, the observer would be inclined to lower the level setting.

2. Since, only one row/column must pass for the UL/LR array to pass, I processed each row/column individually by cropping it from the hole-array. Figure 2.6 shows a hole-array pair for which the rows and columns to be processed have been separated.

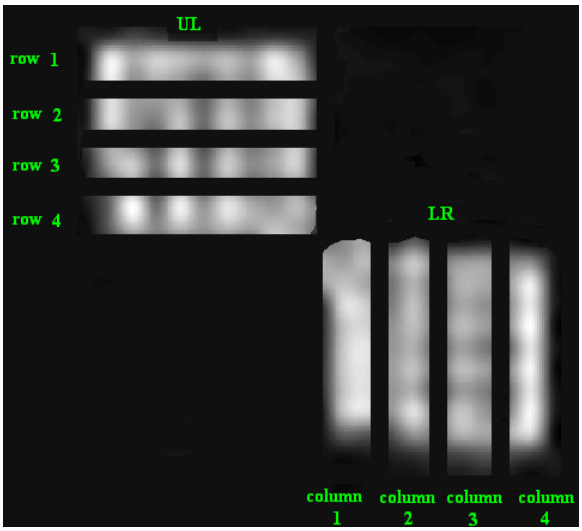


Figure 2.6 – Hole-pair array for which rows and columns to be processed have been separated.

As previously mentioned, segmentation procedure for the automated tests consist of thresholding. However, in this case Otsu's method is not used. Instead, an obvious relationship between the threshold and ratio of on-voxel to ROI size was considered, where the ROI is contains a single row/column to be processed. This relationship is illustrated in Figure 7. It is important to note that the relationship is not necessarily linear, but is non-increasing.

3. Therefore, if I know the optimum ratio for which 4 distinct holes can be found for a row/column, I can perform a binary search to find the corresponding optimum threshold. With each iteration of the search procedure, the number of objects present in the ROI is determined using a connected-component labeling algorithm [22], and the search is ended if 4 objects are found. If the algorithm converges to the optimum ratio without finding 4 holes, the max number of holes is returned.

The optimum ratio value was determined experimentally to be 0.35 for the left and center hole pairs and 0.22 for the right hole pair. However, these ratio choices are not crucial since the objects are counted with each iteration. The ratio choices simply ensure that the search progresses in the right direction.

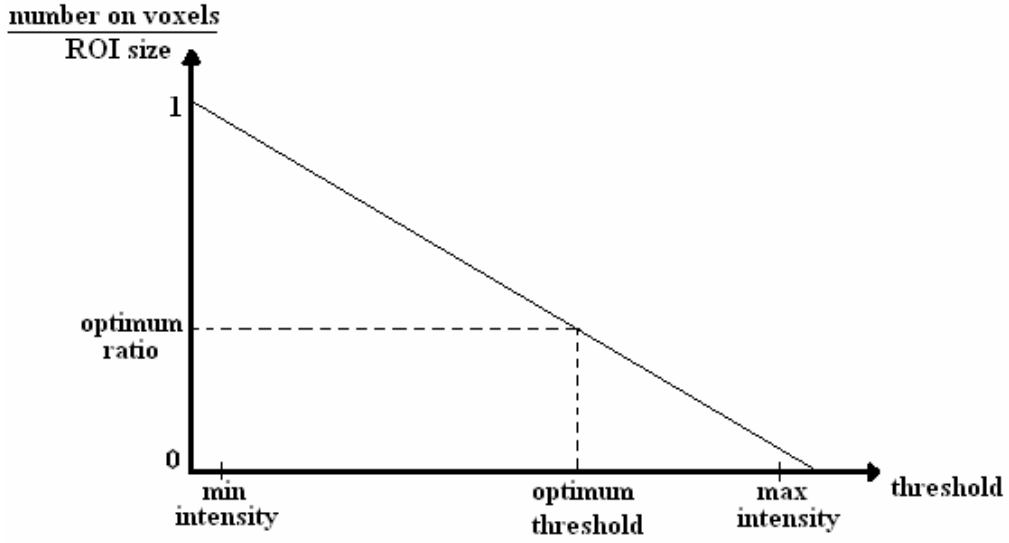


Figure 2.7 – Illustration of relationship between on-voxel/ROI-size ratio and threshold.

Furthermore, the search algorithm will often find the solution in the first iteration. This is because the initial threshold is set to be slightly larger than the max intensity value of the highest relative minima between holes. Figure 2.8 shows max intensity profiles of rows in a UL array. These profiles were created by extracting the max intensity of each column creating a 1 dimensional row array. One can see that the initial threshold choice is optimal because in one case it segments all 4 holes and in the other case no other threshold value can segment all 4 holes.

Still, the aforementioned search algorithm is performed on the ROIs in order to have a more robust procedure.

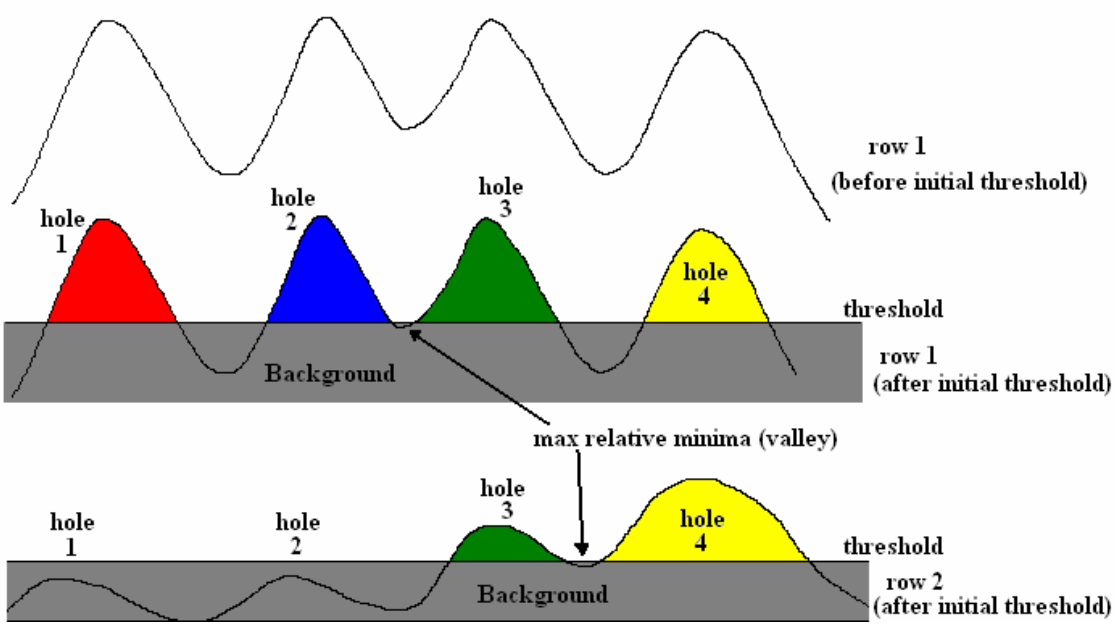


Figure 2.8 – Illustration of initial threshold selection. (top) profile of UL row ROI created by extracting max intensity of each column. (middle) Initial threshold taken to be slightly above maximum relative minima and applied to profile from above. Notice that all holes will be visible. (bottom) A max profile from another row is thresholded with a similar procedure. Notice that only two holes will be visible.

4. Finally, if 4 holes are not found for a row/column, the algorithm tries 60 thresholds varying between 50 percent of the max intensity and the max intensity in order to detect 4 objects. This step is performed, in order to ensure that the binary search did not fail. This procedure is the last resort because it is more computationally expensive than the binary search. The max count of objects detected is returned for each row/column.

At each iteration, morphological closing with a line structuring element is performed on each binary row/column before counting to ensure that the same hole is not counted twice. For the row the line structuring element is oriented vertically, and for the column it is oriented vertically.

In this test, each array of the left, center and right pairs is processed to see if at least one row/column of an array has 4 discernible holes (see Figure 2.9). The measured resolution of a

series is then taken to be the smallest hole-size for which all holes in at least one row/column are discernible. The ACR recommends that a resolution of 1 mm should be achieved along both column and row dimensions, so the left and center UL and LR arrays should pass for the series to pass the test.

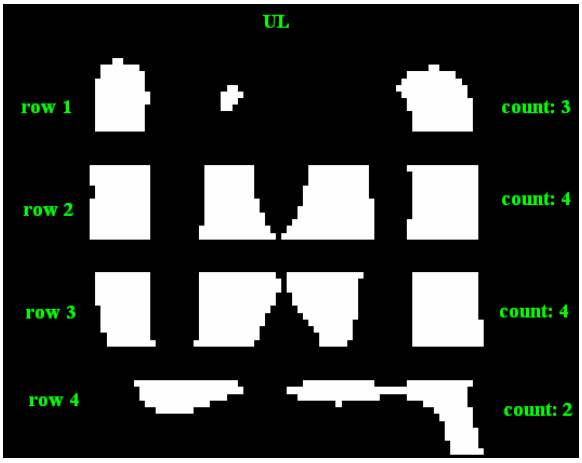


Figure 2.9 – Result of segmenting holes in right UL hole-array. Notice that 4 holes are discernible in rows 2 and 3. Thus, the measured resolution along the column dimension is 0.9 mm for this series.

2.1.4 Slice Thickness Accuracy

In addition to the resolution insert, the phantom also contains a thickness insert that is a block of plastic with two 1 mm thick slots cut into it. These slots show up as crossing signal ramps in slice 1 of the axial series as shown in Figures 2.10 and 2.11. There is an upper ramp and lower ramp with one having a positive slope and the other having a negative slope. The slope is 10 to 1 with respect to the plane of slice 1. Therefore, the signal ramps will appear in the slice with a length that is about 10 times that of the slice thickness. However, if the phantom is tilted in the left-right direction, the ramp lengths will be altered. This is why 2 ramps are considered. The tilt will cause one ramp to be lengthened and the other to be shortened. A formula that will be defined later takes this into account to determine the true slice thickness. Following are the steps of the procedure:

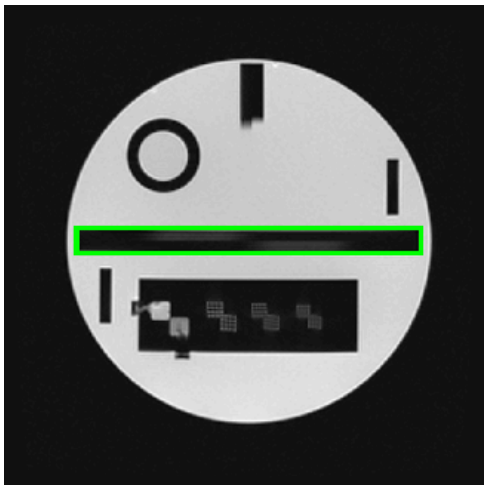


Figure 2.10 – Slice 1 with thickness insert enclosed in green box.

1. The first step of this automated test is to crop the thickness insert. Then, the cropped image is resized by a factor of 3 using bicubic interpolation to allow for sub-voxel accuracy in measurements.

2. The resized crop image is then binarized using automatic thresholding.

3. The connected components are then counted using the labeling algorithm. Since the ramps are so close together, they will often be connected in the binary image. If that is the case, the ramps are separated by simply setting suitable image rows to 0.

4. Finally, the length of each ramp is determined. This is done by summing the image regions corresponding to each ramp. The length for a ramp is taken to be the median of the sum minus 1. The median value is taken because the ends of the ramps will often be ragged due to truncation (Gibbs phenomenon). This procedure is illustrated in Figure 2.11.

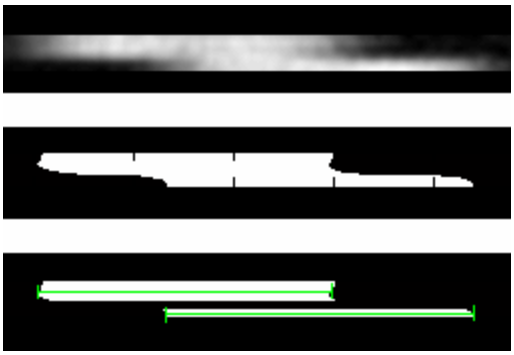


Figure 2.11 – Illustration of processing steps. (top) Cropped intensity ramps. (middle) Crop binarized using automatic thresholding. (bottom) If one representations of ramps are merged in binary image, separate them. Finally, determine median length of ramps.

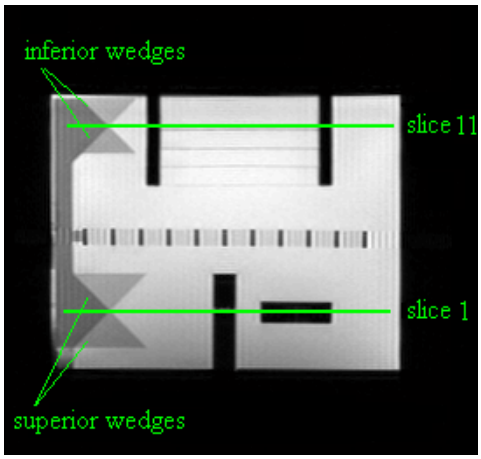
The determined lengths of top and bottom ramps are converted to mm and then used in the following formula for determining the slice thickness.

$$slice\ \ thickness = 0.2\frac{(top * bottom)}{(top + bottom)} \tag{2.1}$$

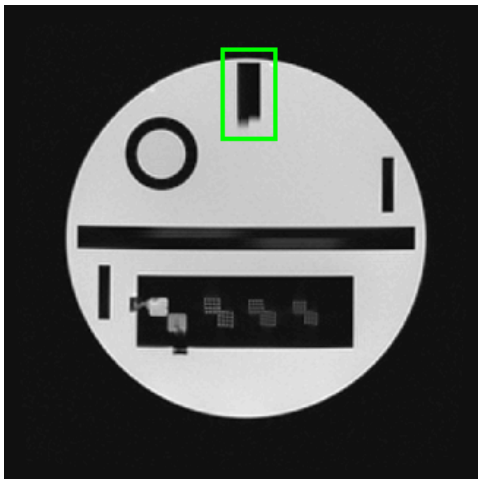
The measured thickness should be 5 mm $\pm$ 0.7 mm.

2.1.5 Slice Position Accuracy

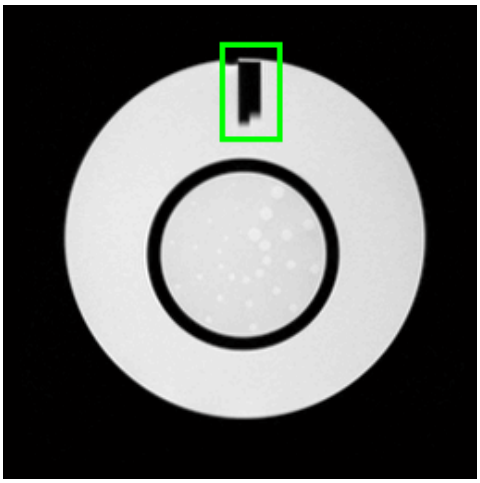
The phantom contains pairs of crossed wedges at its superior and inferior ends with slopes of 45 degrees. The slice positions are prescribed such that slices 1 and 11 intersect the wedges at the vertices as shown in Figure 2.12 (a). The wedges show up as 2 vertical bars in the slices as shown in Figure 2.12 (b) and (c). When the slice is correctly aligned, the bars are of equal length. If the slice is displaced superiorly, the right bar will be longer. If the slice is displaced inferiorly, the left bar will be longer. Therefore, one can simply find the difference between bar lengths to determine the slice is displacement. The steps of the procedure will now be listed as illustrated by Figure 2.13.



(a)



(b)



(c)

Figure 2.12 – Crossed wedges and corresponding bars. (a) Sagittal localizer with crossed wedges. The prescribed positions for slices 1 and 11 are such that they intersect the vertices of the wedges. (b) Slice 1 with the bars enclosed in a green box. (c) Slice 11 with the bars enclosed in a green box.

1. The first processing step is to crop a region containing the bars. The crop is then resized by a factor of 3 to allow for sub-voxel accuracy in measurements.

2. This image is then binarized by performing automatic thresholding.

3. Recall from section 2.1.2 that the phantom edge at the top of the image can be discontinuous causing the bars to be connected to the background. In this case, the bars are separated from the background by simply setting the image region above and 40 rows

below the topmost point of the phantom to zero. This step does not affect the accuracy of the outcome because the difference in bar lengths is of interest.

4. The bars are assumed to be connected since they are adjacent to each other. Therefore, the final binary image is created to only contain the largest object in the present binary image.

5. Finally, the bars are separated and their lengths are determined. The lengths are taken to be the region sum along the row direction minus 1.

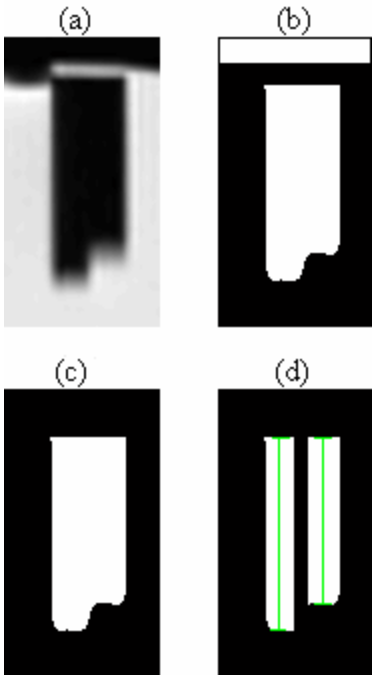


Figure 2.13 – Illustration of processing steps. (a) Crop of bars from slice 11. (b) Complement of binary image produced using automatic thresholding. The bars were separated from the background at top of image. (c) Objects adjacent to image boundary are removed from image. (d) Bars are separated and their lengths are measured.

Because the wedges have slopes of 45 degrees, the actual slice displacement is half of the bar difference. For this test to be passed for slice 1 and 11, the absolute value of the bar difference should be less than or equal to the 5 mm. This means that the both slices must be displaced by less than 2.5 mm.

2.1.6 Image Intensity Uniformity

The Image Uniformity Test assesses a solution only region of the phantom, which is shown in Figure 2.14. This region of the phantom is physically uniform with an equal amount of solution throughout. Therefore, the intensity of the corresponding slice in the series should also be uniform. The intensity uniformity is assessed by determining and comparing the high and low intensity levels within the phantom region of slice 7.

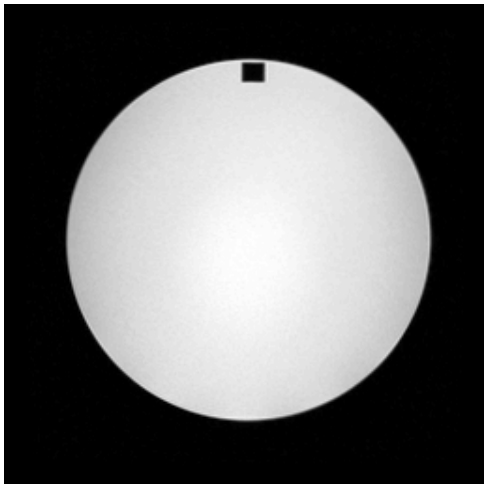


Figure 2.14 – Slice 7 passes through a solution only region of the phantom. Therefore, the phantom region of this slice should have relatively uniform intensity.

The processing steps of this automated test will now be listed. The ACR Program guide advises against simply determining the max and min intensity values within the phantom region of the slice because of the presence of noise in the image. This method would provide max and min measurements with much greater difference than the true values would have. Therefore, the test would be biased towards failure. In light of this, a more elaborate procedure is used.

1. The slice is blurred in order to lessen the effect of noise. A Gaussian kernel with standard deviation 2 and size 6 squared is used. The example of a blurred image used is shown in Figure 2.15.

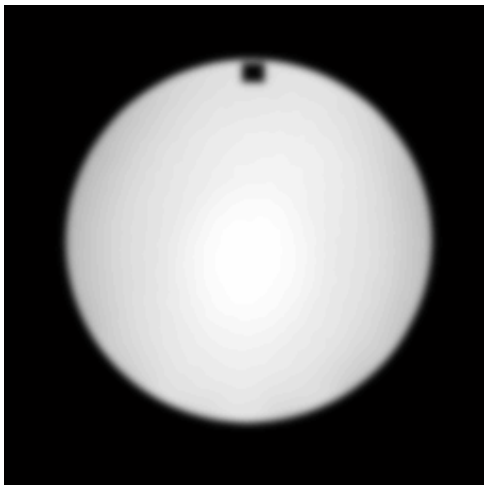
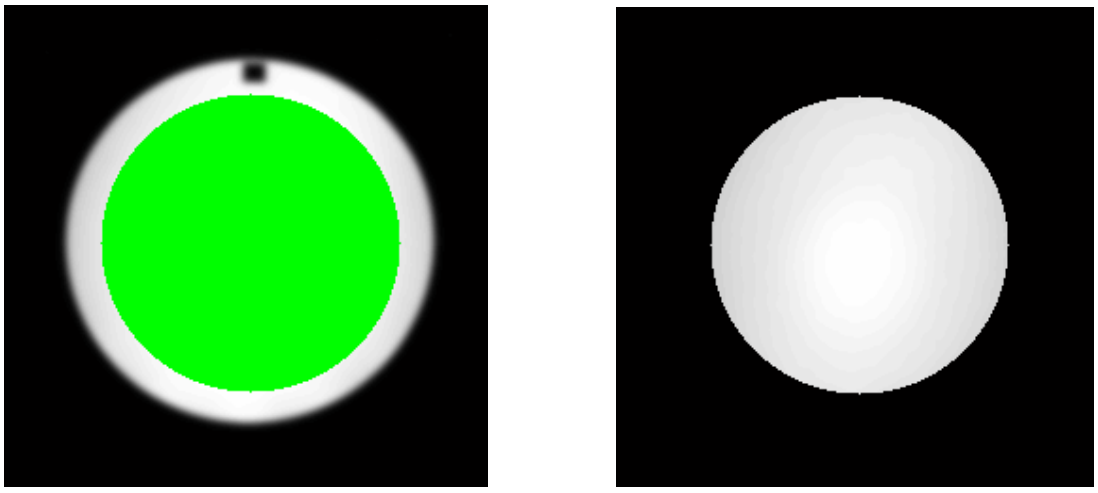


Figure 2.15 – Blurred version of image in Figure 14.

2. The blurred image is used to determine the locations of the average max and min intensities within a large circular ROI shown in Figure 16. These locations will pertain to the true max and min intensities in the image. The ACR Program guide specified that the radius of the ROI should range between 79.8 and 82.7 mm. I used a ROI of radius 80.1 mm in order to allow the procedure to accurately process images in which the phantom diameter varied greatly from the standard.



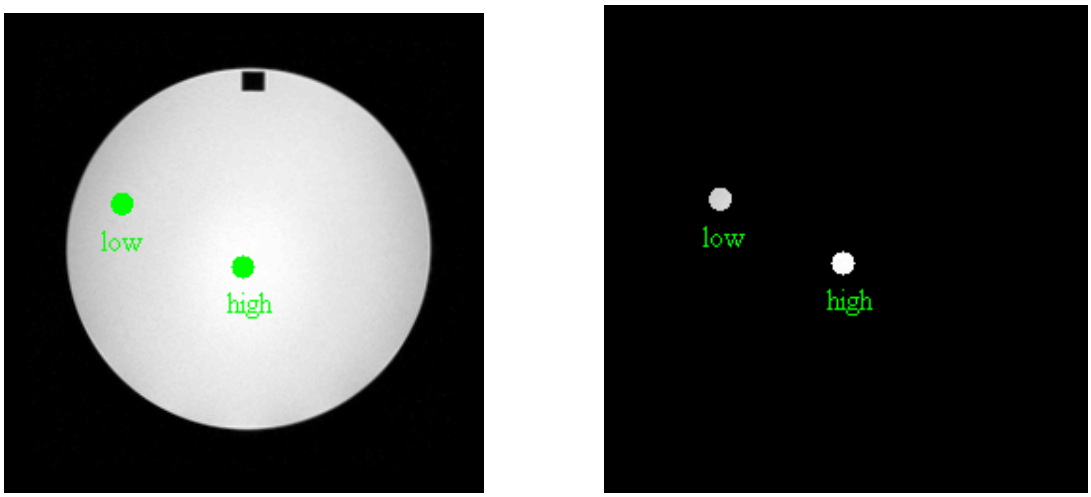
(a) (b)

Figure 2.16 – Large ROI used for determining the location of the max and min intensities in the blurred image. (a) Blurred image with ROI shown in green. (b) Extracted ROI.

3. Two small circular ROIs of area 1 cm^2 are then applied to the original image and are centered about the determined max and min locations as shown in Figure 2.17. The mean intensities in these small ROIs are taken to be the low and high intensities of the image and are used in a formula to calculate the PIU.

$$PIU = 100\left(1-\left(\frac{high-low}{high+low}\right)\right) \tag{2.2}$$

The PIU should be greater or equal to 90 percent for the test to be passed.



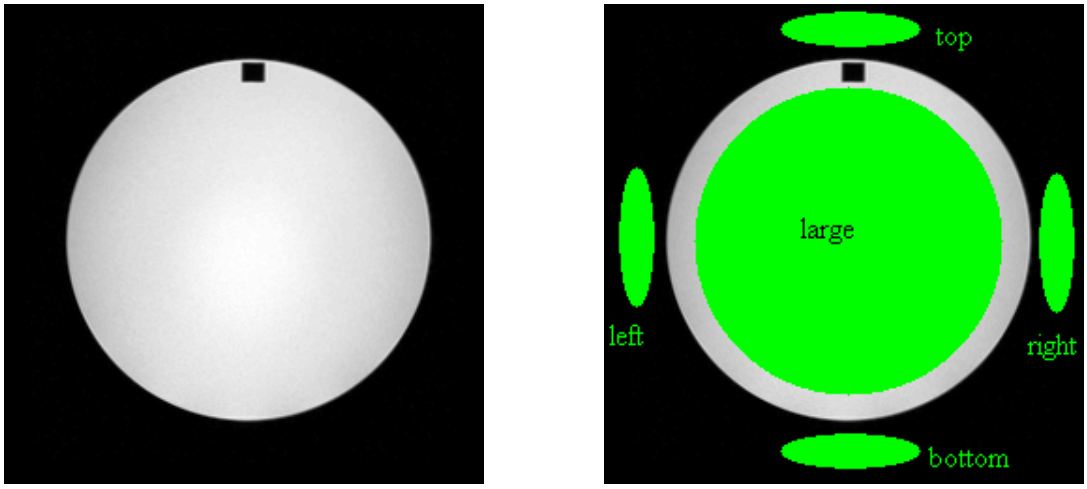
(a) (b)

Figure 2.17 – Small ROIs used to determine the average low and high intensities in the image. (a) Small ROIs applied to original image. (b) Extracted small ROIs.

2.1.7 Percent Signal Ghosting

The Percent Signal Ghosting Test determines whether a substantial copy of the imaged object (phantom) appears in the background of slice 7. Ghosts of an object can only fall within the phase encode shadow, which is the area of the image along the left-right direction that is not above or below the object. In this test, four elliptical ROIs are placed in the background at locations above, below, left, and right of the object. The ROIs that are above and below the phantom region serve as controls of the mean background level since it is certain that these ROIs will contain no ghost signals. A large circular ROI is used to determine the average intensity of the object. The steps of the automated test are now provided.

1. Five ROIs are applied to slice 7 as shown in Figure 18. The large ROI has a radius of 80.1 mm. Each elliptical ROI has a major-axis length of 36.1 mm and a minor-axis length that is ¼th the major-axis length as specified by the ACR. The elliptical ROIs were oriented as shown so that they would not overlap the object.



(a) (b)

Figure 2.18 – ROIs applied in this test. (a) slice 7 which is used for this test. (b) ROIs applied to the slice 7. The mean intensities in these ROIs are used to determine whether any signal ghosting is present.

2. The mean intensity values of the top, bottom, left, right, and large ROIs are used in the following equation to determine the ghosting ratio.

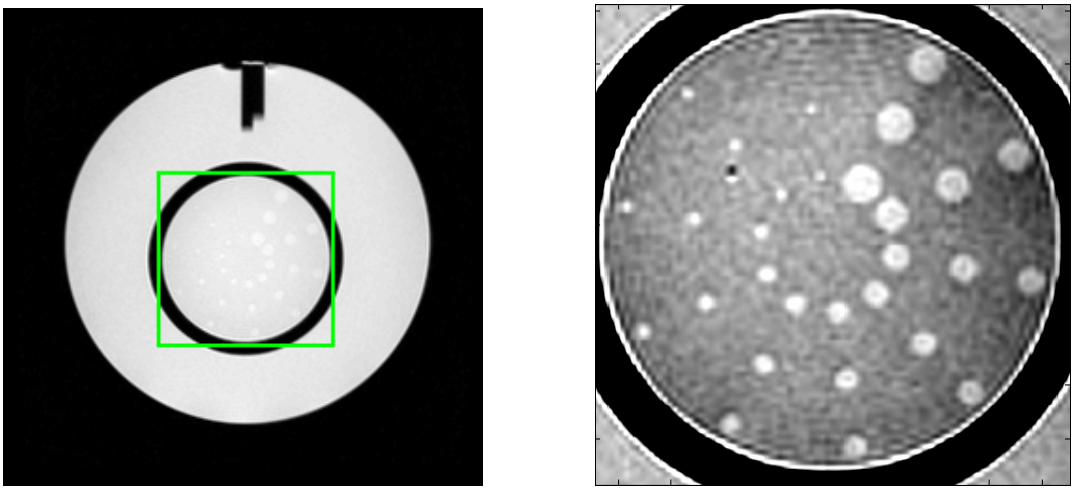
$$ghosting\ \ ratio = \left| \frac{(top + bottom) - (left + right)}{2 * large} \right| \qquad (2.3)$$

The vertical lines in the preceding equation denote that the absolute value is considered. The ghosting ratio must be less than or equal to 0.025 for this test to be passed.

2.1.8 Low Contrast Object Detectability

The phantom contains 4 sheets of plastic of towards its inferior end through which slices 8 through 11 pass. Each sheet has holes that allow the solution to pass through. These holes are grouped into 10 spokes that radiate from the center of the sheet as shown in Figure 2.19, where each spoke contains 3 holes of the same size. The spoke with the largest holes is taken to be

spoke 1, and the hole diameters decrease from 7.0 mm at spoke 1 to 1.5 mm at spoke 10 moving counter clockwise. Moving from slice 8 to 11, the sheets increase in thickness causing the contrast between the holes and background to be 1.4, 2.5, 3.6, and 5.1 percent respectively. Therefore, all holes of a given slice have the same contrast level, but vary in size. Also, all slices have the same spokes with the same size holes, but vary in contrast. The ACR manual procedure involves the observer counting the number of spokes in each slice for which all holes are visible. The steps of the automated test are now provided.



(a) (b)

Figure 2.19 – Contrast sheet for slice 11. (a) Green box encloses contrast sheet in slice 11. (b) Contrast sheet cropped and resized. The display intensity level and window were set to best show the spokes of holes.

1. The contrast sheet is segmented from the rest of the image. This is desired so that the spokes can be more accurately located for processing. First, the image is binarized using the automatic thresholding procedure. Then, holes within the object are filled and relatively small disconnected objects are removed. Next, the complement of the filled image is added to the original binary image to produce an image that only has a small ring around the contrast sheet in the background. Then, connected components adjacent to the image border are removed producing an image with a disk representative of the contrast sheet. Finally, this disk image is multiplied with the original image to extract the contrast sheet. These steps are illustrated in Figure 2.20.

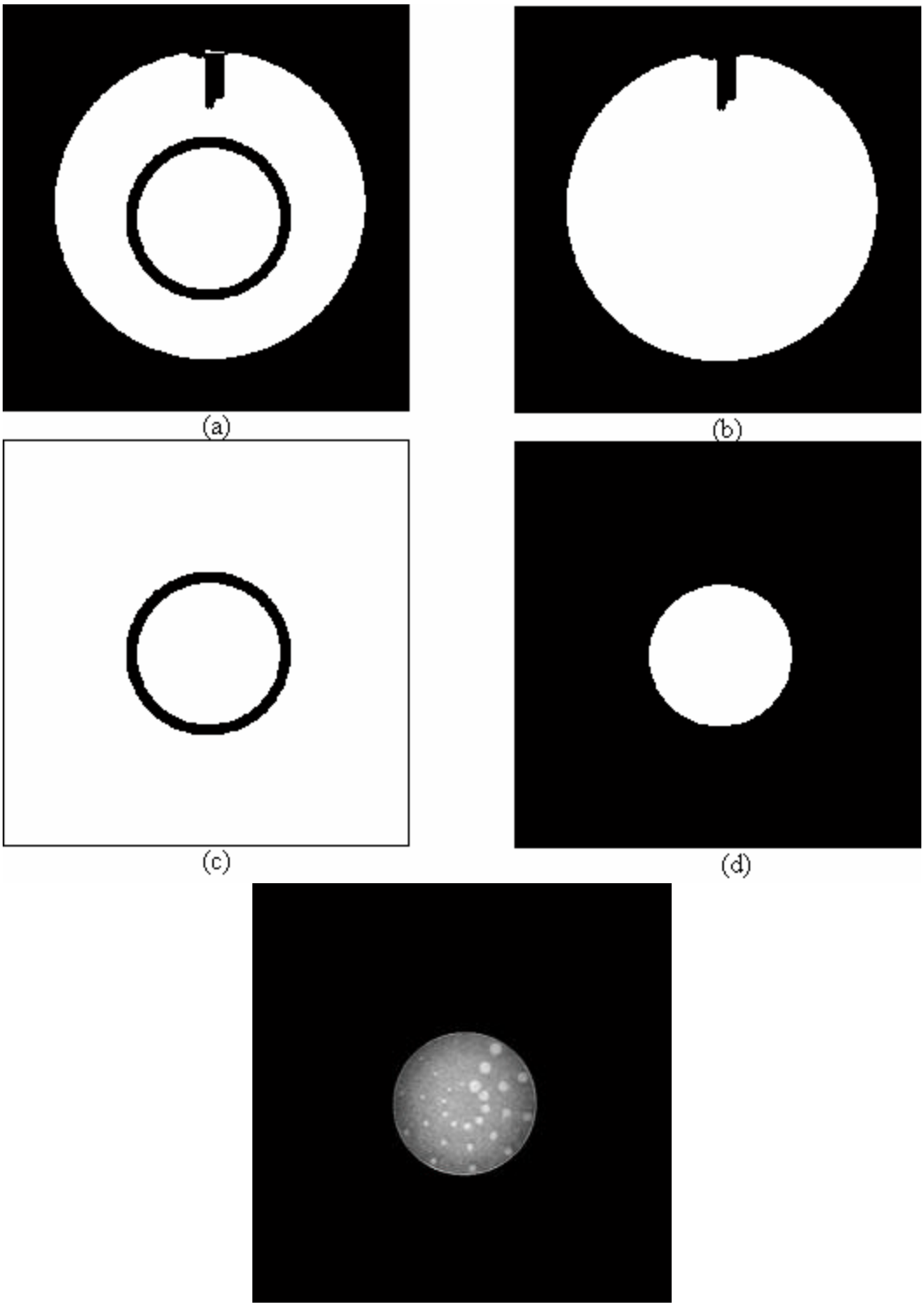


Figure 2.20 – Illustration of contrast sheet segmentation procedure. (a) Binarize slice using automatic thresholding. (b) Fill holes within object. (c) Add complement of image in (a) to image in (b). (d) Remove objects adjacent to image border. (e) Multiply by original slice to extract contrast sheet.

After segmenting the contrast sheet from the image, the holes must be detected. In automating this test, I drew inspiration from the Rose Model described in section 1.3.1. I recognized that the images considered in this test have the conditions necessary for this Rose Model. That is, the signal and background are known exactly, and the signals to be detected are flat topped and sharp edged. The first condition is true because image registration is performed. However, the second condition would only be absolutely true if there was no blurring during imaging, which is not possible. Still, I noticed that holes were relatively flat and sharp edged. With these conditions being satisfied, signal detectability could be assessed by performing cross correlation with a template containing an intensity-scaled version of the signal of interest. Rose also experimentally determined related SNR and CNR thresholds. If the image SNR and CNR were above these thresholds, he decided that the signal would be detectable by an observer. Similarly, I experimentally determined a correlation threshold, and if the correlation result for a given hole exceeds this threshold, the algorithm decides that the hole is detectable by an observer.

2. The image containing the contrast sheet is resized by 2 using bicubic interpolation. For each spoke of a given slice, 11 equal-length paths that will be used for cross correlation are determined as shown in Figure 2.21. The paths radiate from the determined centroid of the contrast sheet and extend just past its radius. Because of likely miniscule rotational discrepancies between the standard series and the registered candidate series, it is probable that none of the eleven paths about a spoke with small holes will pass through the center of all holes in the spoke. However, with 11 spokes being used it is highly probable that each hole will have a path pass through its center.

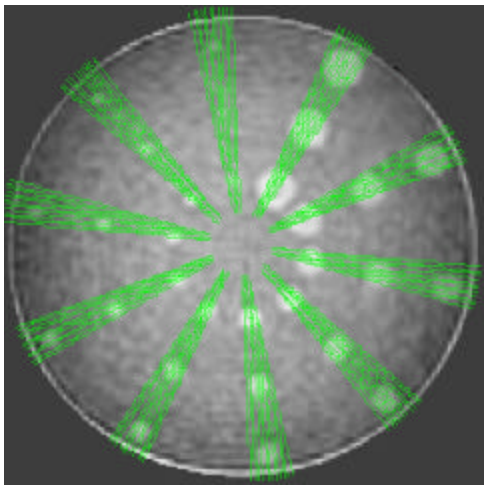


Figure 2.21 – Correlation sheet of slice 11 with correlation paths plotted in green. There are 11 paths about each spoke varying in position by 1 degree.

3. Correlation is performed by sliding one of the templates shown in Figure 2.22 along the 11 paths about each spoke resulting in 11 correlation profiles for each spoke. These 11 profiles are condensed into one by creating an array that consists of the max value of all profile along the path index. Thus, this aggregate correlation profile, shown in Figure 2.23, stores the maximum correlation result for each path index.

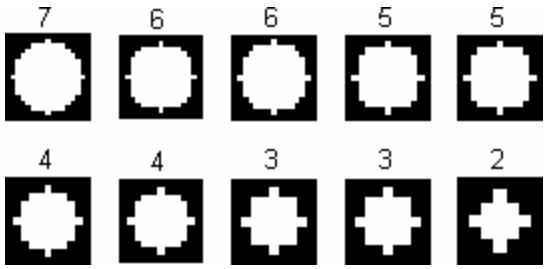
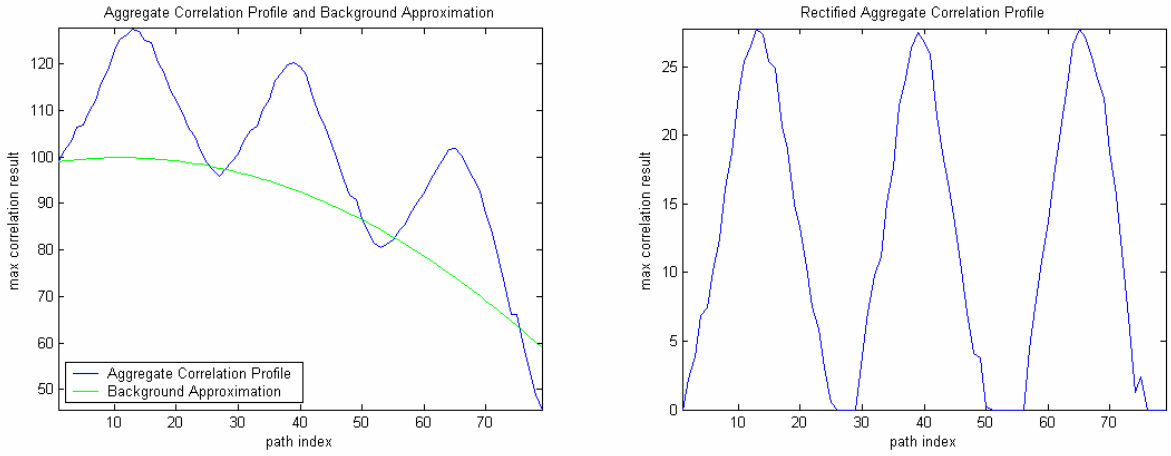


Figure 2.22 – Correlation templates listed in order of corresponding spokes. The radius of the disk contained is indicated above the template. Since the image was resized by a factor of 2, the radius is approximately equal to the hole-diameters listed above. (top) Templates for spokes 10 through 6. (bottom) Templates for spokes 5 through 1.

4. Figure 2.23 (a) shows the aggregate correlation profile for spoke 1 of slice 11. From now on I will refer to the aggregate correlation profile as simply the correlation profile. It is evident that there is a gradually increasing downward shift of the correlation result as one moves along the path index. This shift results because the image intensity gradually

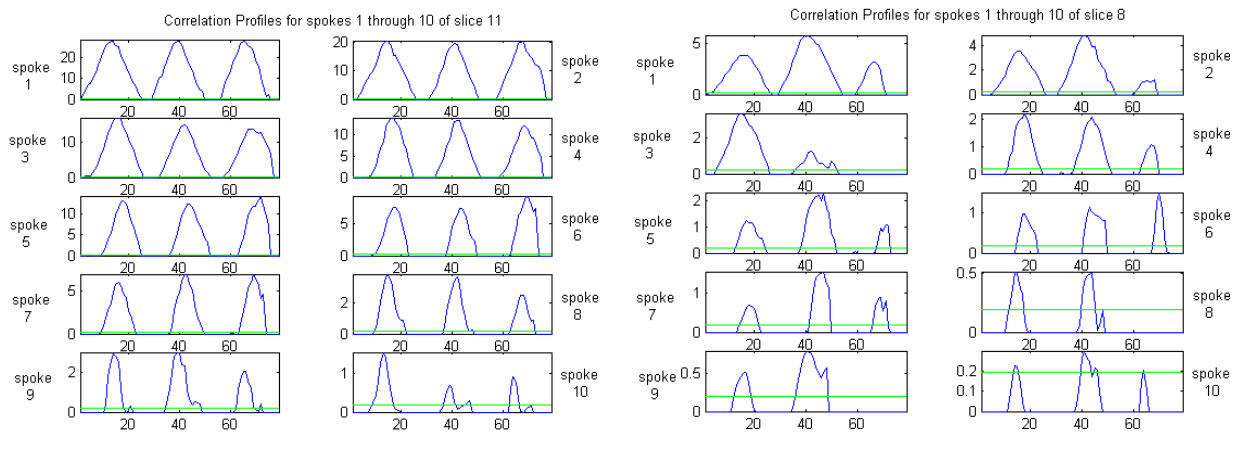
decreases in the contrast sheet region of the image moving away from its center. In order to rectify this, the algorithm performs polynomial regression using values from the valleys of the correlation profiles. The determined polynomial is taken to be an approximation of the background and is subtracted from the correlation profile.



(a) (b)

Figure 2.23 – Aggregate Correlation Profile. (a) Original aggregate result of correlation for spoke 1 of slice 11. The background approximation determined using polynomial regression is also shown. (b) Rectified correlation profile determined by subtracting the approximated background.

5. The cross correlation and rectification steps described above are repeated for all spokes producing the profiles shown in Figure 2.24. The next step is to apply the experimentally determined correlation threshold to determine what holes are detectable. This threshold was determined by allowing an objective observer to tell the slice and spoke which was first to be undetectable using the GUI observer tool described in Appendix C. The peak level of correlation result corresponding to the undetectable hole was taken to be the threshold. This procedure was repeated until an optimal threshold value was determined. The threshold is applied to the correlation profiles of all spokes of the series. The correlation profiles with all holes exceeding the threshold are determined to be detectable and are counted.



(a) (b)

Figure 2.24 – Correlation profiles for multiple spokes along with correlation threshold plot in green. (a) Spokes 1 through 10 of slice 11. Since this slice has the greatest contrast, the correlation results for all holes of all spokes exceed the threshold substantially. (b) Spokes 1 through 10 of slice 8. This slice has the lowest contrast, so the correlation results do not exceed the threshold by as much as in slice 11. For spokes 3, 8, and 9 the external hole is not detectable because of the image intensity shift described above.

6. The ACR program description states that for a given slice, the observer should stop counting spokes after the first incomplete spoke is found. Therefore, this rule is incorporated into this automated test. Thus, final count may be less than the true number of detectable spokes.

For this test to be passed, the total spoke count for the series should be greater than or equal to 9.

2.2 Automated DTI Quality Assurance

Two of the most important DTI features considered in diagnosis are diffusion anisotropy and axon orientation. Thus, I designed tests to assess diffusion tensor image quality as it pertains to these considerations. In many instances, the accuracy of DTI images can be compromised by miscalibration of the scan parameters such as pre-correction gradient, gradient strength, and amplifier gain. Generally, physical phantoms provide ground truth measurements that can be

used to assess the performance of a scanner. If the phantom has consistent diffusion behavior, DTI images produced using this phantom should have similar voxel intensities (assuming that noise will have a minimal effect). Substantial intensity deviations would indicate that a parameter was miscalibrated or that there was a problem with the scanner functionality. As this project is a first attempt at developing a DTI QA procedure, I only considered varying the z-gradient strength as a source of quality reduction. The decision was made to vary the z-gradient strength because capillary bundles were oriented such that they were parallel with the z direction. Thus, the maximum diffusion would be in the z direction. In developing this DTI QA procedure, first a suitable phantom was created and imaged. Then, the test procedures were developed to test for low contrast object detectability in FA maps and orientation accuracy.

2.2.1 Construction of Physical Phantom

A physical phantom was desired that could provide ground truth measurements of water diffusivity. In designing this phantom, I consulted [36, 25, and 21] to learn what aspects of the phantoms used were effective. The authors were interested the effect of crossed axonal patterns and changing orientation on diffusion measurements, but I decided to focus on the capabilities of imaging systems relative to axon diameter. The phantom was designed such that FA contrast and orientation accuracy capabilities of the scanner could be assessed. The phantom was primarily comprised of PTFE microbore tubing with dimensions stated in section 1.4.4. (Although ultramicrobore tubing is closer to the size of large axons, it was not used because of the difficulty in filling such small tubing with water.) Lengths of 7 cm were cut, and these capillaries were grouped into bundles increasing in size by powers of 2 with the smallest bundle being a single capillary and the largest bundle consisting of 16 capillaries. The phantom was manufactured such that the bundles were surrounded by gelatin and encased in a plastic container. The surrounding gelatin was necessary to avoid air related susceptibility artifacts being present in the image. The procedure followed for creating the phantom was as follows:

1. Cut capillaries to 7 cm lengths.
2. Use syringe to insert water into capillaries while ensuring that no air bubbles are present.
3. Seal both sides of capillaries with hot-glue.
4. Make bundles with 1, 2, 4, 8 and 16 capillaries using Krazy Glue.

5. Create hardened layer of gelatin in plastic container.
6. Place capillary bundles, on gelatin layer.
7. Create second hardened layer of gelatin atop the bundles such that no air is trapped within the structure.

Figure 2.25 shows the capillary bundles and the completed phantom.

(a) (b)

Figure 2.25 – Photograph of phantom components. (a) Four largest capillary bundles shown with penny to provide a sense of scale. (b) Top view of completed phantom.

2.2.2 Imaging Physical Phantom

The phantom was scanned on a 1.5 T GE TwinSpeed MR scanner (General Electric Healthcare, Milwaukee WI) at the Wake Forest University Baptist Medical Center (WFUBMC). This scanner is equipped with high performance gradients supporting a 40 mT/m magnitude and a 150 T/m/s slew rate. A head coil with a maximum 125 kHz sampling rate was used for both RF transmission and signal reception.

Three z-gradient strength values were used in scanning. The normal gradient strength was 3.9445. The other two magnitude values used were a reduction of 10.12 and 20.12 percent resulting in gradient strengths of 3.5946 and 3.13446 respectively. For each gradient strength, we had 4 repetitions resulting in 16 overall image data sets. The filenames of these datasets are not important, but in order to refer to a particular set, I will list the file number. Table 2.1 shows the z-gradient level and repetition number for all data sets used in developing this procedure.

Table 2.1 – Data set file numbers shown with associated z-gradient level and repetition number.

z-gradient (relative to max)	repetition 1	repetition 2	repetition 3	repetition 4
100 %	005	008	012	016
90%	006	009	013	017
80%	007	010	014	018

We oriented the phantom such that the capillaries were parallel with the z-dimension. That is, the bundles intersected the x-y plane, resulting in cross-sections in each axial slice. The image specifics are shown in Table 2.2.

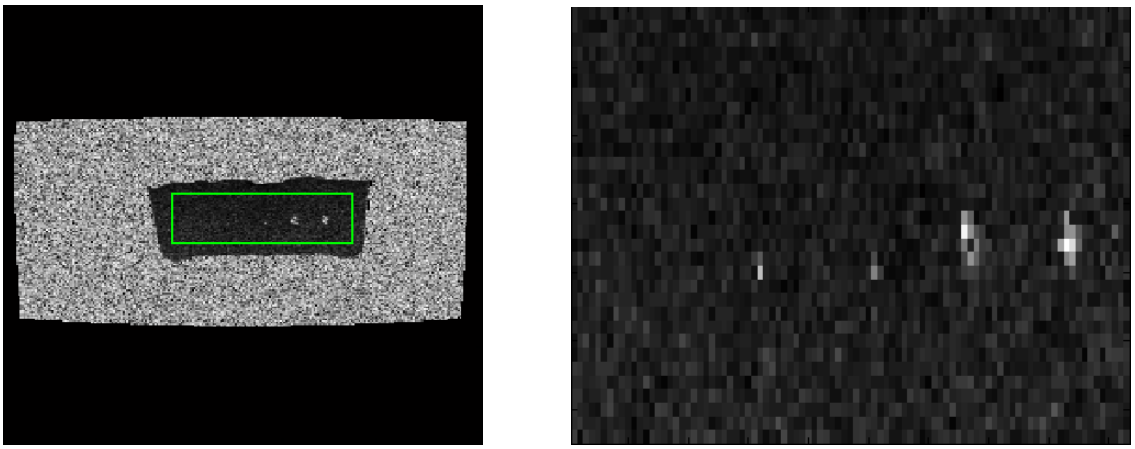
Table 2.2 – Image Specifics

In-plane resolution (x-y plane)	1.016 mm
Between plane resolution (z-dimension)	2 mm
Number of slices	24
Matrix size	256 x 256

Thus, the 3-dimensional image contained 4.8 cm of the phantom along the z-dimension. This was desired because even though the capillaries are 7 cm long, as they were slightly curved due to spool memory, the actual length in the phantom is shorter. Along the z-dimension, the phantom was imaged in its center to avoid artifacts resulting from air susceptibilities. This was because some capillaries may have contained air bubbles at their ends. The small resolution value shown in the table was chosen since the capillaries were so small in diameter. However, lower resolution results in lower SNR, which is compounded by the fact that relatively little signal can be measured from these small capillaries. Therefore, this tradeoff was considered when specifying the in-plane and between plane resolutions. Taylor [36] concluded that insufficient spatial resolution prohibited the correct orientation of the capillaries being determined for his work because of partial volume effects, and he used 2 mm in plane and 4 mm between plane resolutions. Lin et al. [25] used an in-plane resolution of 0.72 mm and noted no SNR issues. However, we decided to use a slightly higher resolution to ensure that sufficient SNR was obtained. The matrix size was chosen to be 256 x 256 even though the width of the phantom is about 13 cm. This matrix size was set twice as large as necessary to avoid aliasing.

2.2.3 Fractional Anisotropy Contrast Test

I decided to create a procedure to assess the capability of an observer to detect low contrast objects in FA images as shown in Figure 2.26. Therefore, the first thing that was done was to create the FA maps from the tensor data using equation 1.19. The objects to be detected are the cross sections of the bundles intersecting the axial slices. Thus, one can expect that the bundles would show up as elliptical objects in the images. For object detection, I considered the 3 methods described in section 1.3.7, namely, the Rose Model, ideal observer model, and Hotelling observer model. The Rose Model and the ideal observer model require the SKE/BKE condition. In fact, [7] states that if there is any randomness in the objects to be detected, it is not possible to calculate the test statistic for the ideal observer. The locations of the bundle cross sections were not known exactly because the bundles had irregular shapes that resulted from a combination of the capillary curve due to spool memory and excessive glue application. Also, the combination of extremely weak signal and high noise amplitude adds to the randomness of the objects. I also considered using the Hotelling observer model. However, enough was not known about the image statistics such as the probabilities of the signal present and not-present states to create the simulated images.



(a) (b)

Figure 2.26 – Axial slice through FA map of series with file number 008. (a) Full axial slice. The green box encloses the region considered for the test (b) Crop of phantom region of axial slice.

I decided that for an object to be absolutely detectable, the signal should be higher than the background with added noise. Therefore, a meaningful decision threshold for detectability

would simply be a suitable intensity threshold that caused the bundles to be placed in the foreground during binarization. The steps of this test procedure will now be provided.

1. The FA map is created from tensor image being considered, and the phantom regions of all axial slices are cropped as shown in Figure 2.27. The max intensity value of regions known to be part of the background is taken to be the threshold used in segmentation.

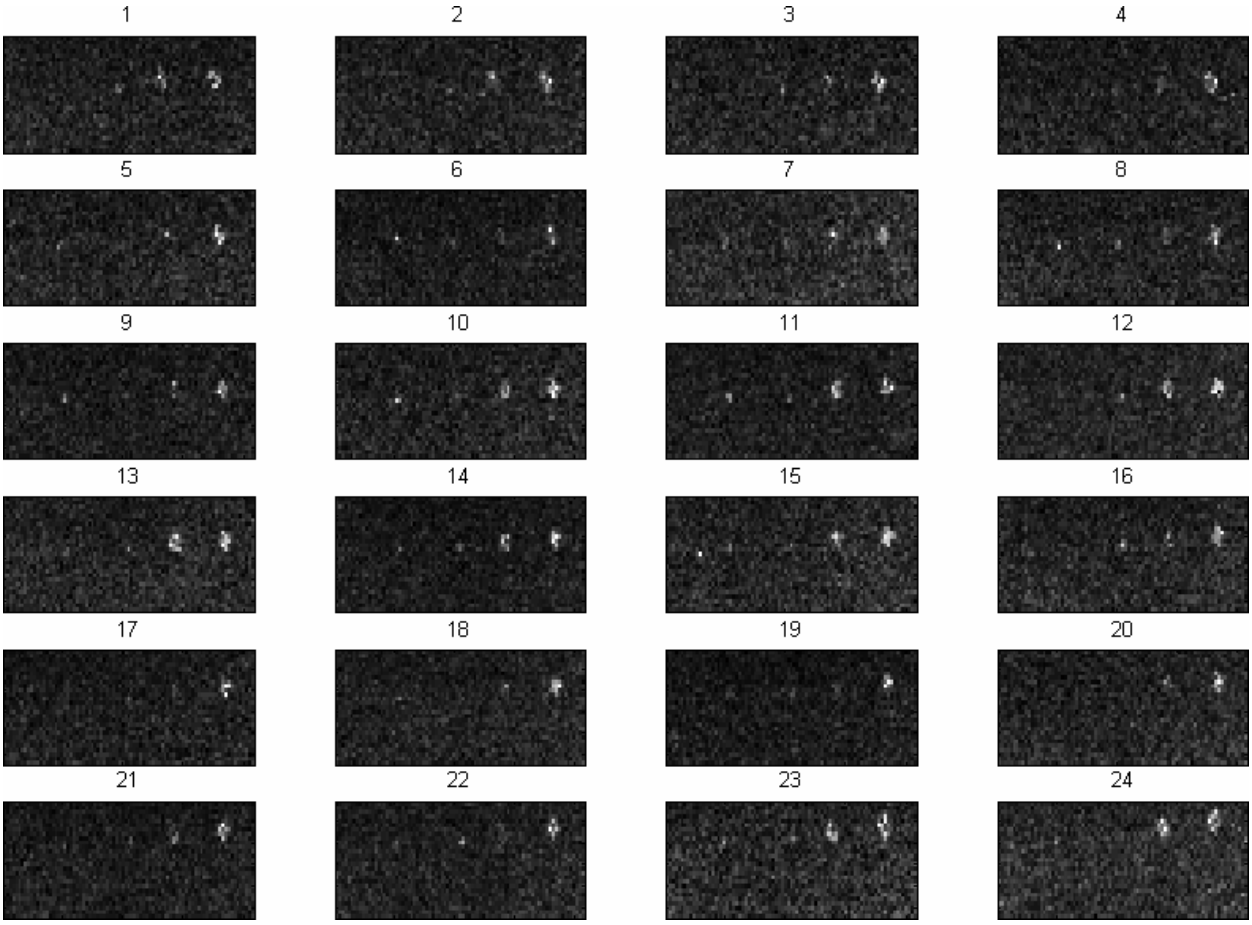


Figure 2.27 – All 24 axial slices of series with file number 008.

2. All axial slices are thresholded to segment the detectable bundles from the background. A morphological closing operation is then performed using a circular structuring element to connect pixels that pertain to a single bundle. A mask is then multiplied to the resulting binary image in order to remove any objects that are outside of the possible bundle locations. This mask is shown in Figure 2.28. The resulting series produced by

performing this segmentation procedure on the original series from Figure 2.27 is shown in Figure 2.29.

Figure 2.28 – Mask applied to images after thresholding step. The vertical location of the foreground regions of the mask is based on the location of the largest object present in the binarized image after thresholding. This is because the vertical position of the bundles changes drastically from slice-to-slice.

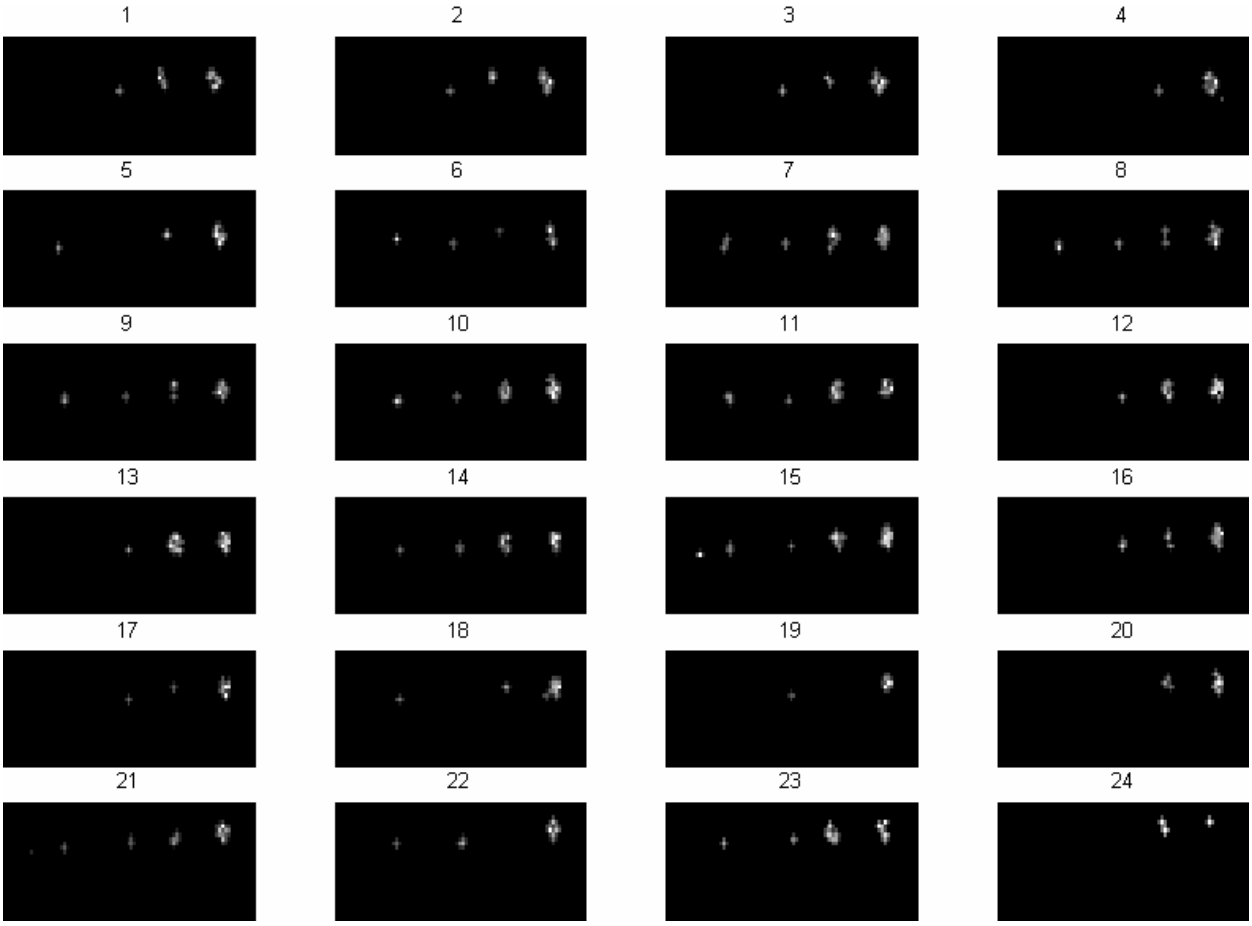


Figure 2.29 – Result of segmentation procedure performed on series with file number 008. These images were produced by multiplying the final binary images with the corresponding original slices. Notice that only in slice 15 can all 5 bundles present.

3. The final step of the procedure would be to count the number of bundles detectable for each slice, and then make a decision as to whether the data set has passed the test.

To decide on a criterion to be used to pass or fail a data set, I considered the object detectability results for the 16 data sets mentioned earlier. Figure 2.30 shows the detectability decisions made for these 16 data sets. The first thing that I noticed was that the detectability decisions varied greatly for series with the same z-gradient strength. This was unexpected but shows that object detectability is highly affected by the noise, so much so that object detection almost becomes a random process. This is a direct result of the low resolution and small signal that can be measured by bundles consisting of very small capillaries. Also, the inner diameter of each

capillary is very small relative to the outer diameter. Thus, not enough water is present to produce a strong signal.

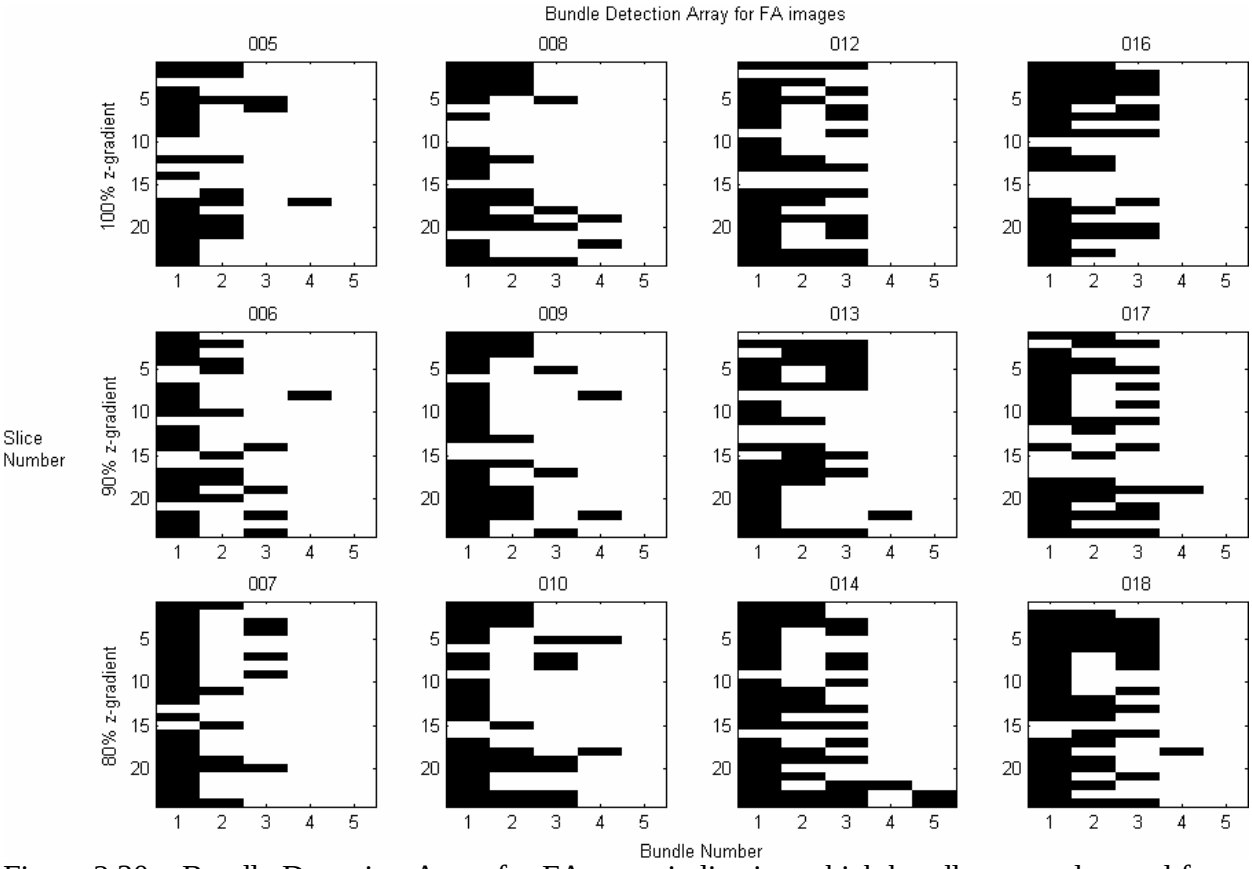


Figure 2.30 – Bundle Detection Array for FA maps indicating which bundles were detected for a given slice. Each image pertains to a data set, and the images are arranged in the same order as in Table 2.1. Images in the same row have the same z-gradient strength. Images in the same column are of the same repetition number. For a particular image, the image-rows pertain to the slice number, and the image-columns pertain to the bundle number. The bundles are listed in the same order as they appear in the axial slices, i.e. from left-to-right. For a given series, if a bundle was detected, the pixel of the corresponding slice and bundle number in Bundle Detection Array is set to white.

After exploring the data shown in Figure 2.30, I discovered the trends shown in the following 3 tables. Table 2.3 shows the total number of bundle cross sections detected for each series. Notice that the low count for a certain z-gradient strength is lower than the high count for the lower z-gradient strengths. This, again, was surprising, and this count could not be used as a criterion.

Table 2.3 – Total number of bundle cross sections detected

z-gradient (relative to max)	repetition 1	repetition 2	repetition 3	repetition 4
100 %	90	86	78	78
90%	89	85	81	79
80%	87	82	71	73

In some slices, small bundles are detected yet larger bundles are not detected, which was also unexpected. However, a similar observation was made for the low contrast objects considered in the final test of the ACR procedure. Recall that the ACR Accreditation Program guidance manual states that the human observer should stop counting spokes as soon as the first incomplete spoke is found even though complete spokes of smaller disks might be detectable. I adopted the counting procedure from the ACR program for this test. Table 2.4 shows the number of slices for a given series that contain at least 4 bundles. However, only slices that are missing the smallest bundle are counted as having 4 bundles. There is not enough difference between the series different z-gradient strengths for this count to be used as a criterion.

Table 2.4 – Number of slices in a series that contained at least 4 bundles.

z-gradient (relative to max)	repetition 1	repetition 2	repetition 3	repetition 4
100 %	14	11	8	9
90%	11	11	10	8
80%	14	12	6	5

Table 2.5 shows the number of slices that contained all 5 bundles in the series. It is evident that if this count is used as the criterion and if a decision threshold of 4 is used, the pass rate for the series with the maximum z-gradient strength would be ¾. The pass rate for the series with z-gradient strength that is 90 percent of the maximum would be ½. All series with the lowest z-gradient strength would fail. Therefore, it appears that this criterion would be sufficient.

Table 2.5 – Number of slices in a series that contained all 5 bundles.

z-gradient (relative to max)	repetition 1	repetition 2	repetition 3	repetition 4
100 %	5	6	3	4
90%	4	3	4	3
80%	1	3	2	2

2.2.4 Orientation Accuracy Test

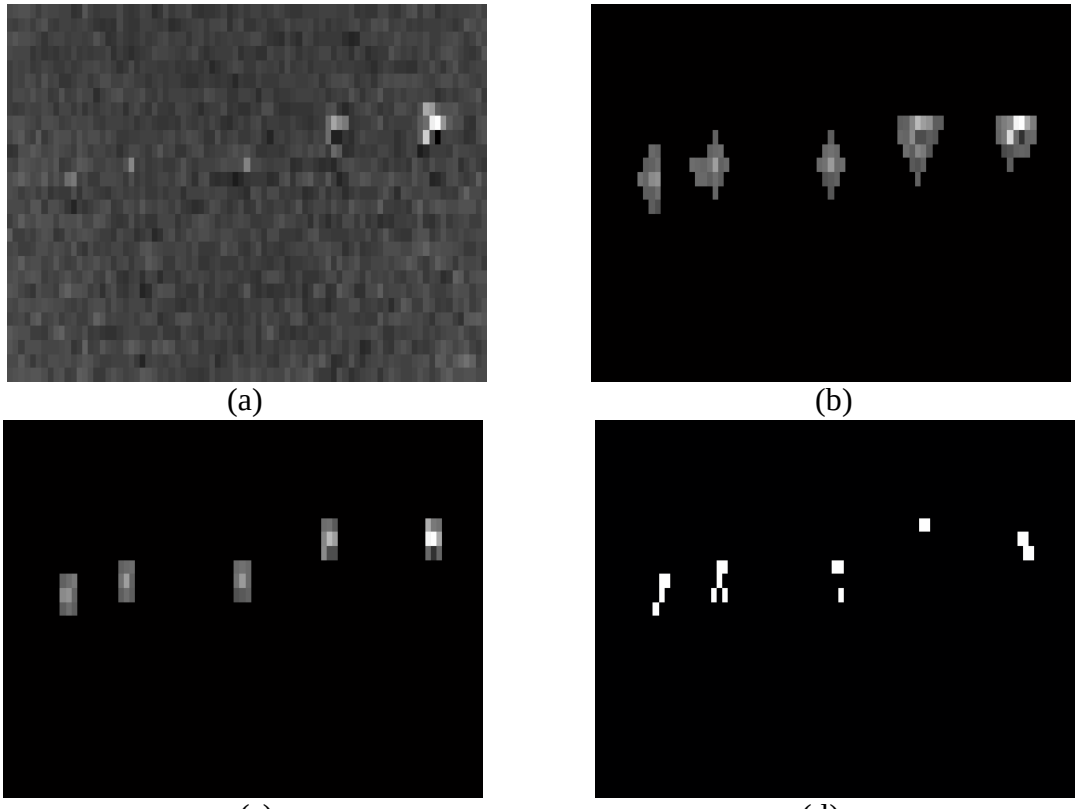
Seeing that phantom was oriented such that the bundles were approximately parallel with the z-direction while imaging, the maximum diffusion should also be in this direction. Therefore, when one tries to determine the orientation of the bundles, one should determine the corresponding orientation. In this test, this orientation is determined from the image by assessing the main direction of diffusivity, which is done by eigenvalue-eigenvector decomposition of the voxel tensors. The steps of the test are now provided as illustrated in Figure 2.31.

1. The max eigenvalue of a voxel tensor is the maximum unidimensional diffusivity coefficient and is a measure of the diffusion in the max-diffusion direction for the voxel. I decided that for a bundle to be considered, the directional diffusivity of its voxels should be larger than that of the background. Therefore, first an image is produced that has the max eigenvalue for each voxel tensor as its voxel values. Then, the bundle segmentation procedure described in the previous section it performed to determine what bundles to consider for orientation accuracy assessment.

2. As is evident in Figure 2.31 (b), the segmented bundles may incorporate portions of the background, which is a result of the morphological closing operation. Therefore, the voxel location of max intensity in each bundle region is determined, and the only a 3x3 neighborhood about each bundle is considered from then on.

3. The eigenvectors of the max eigenvalues in each bundle neighborhood are considered. The number of voxels that have the z-component being the largest in the eigenvector are counted, and this count is stored in an array indexed by the image slice number and bundle number.

4. The final step of the procedure would be to determine whether the series passed by assessing a certain criterion. Therefore, I explored count arrays for the 16 data sets to determine an optimal criterion.



(a) (b) (c) (d)

Figure 2.31 – Illustration of processing steps for this test. (a) Original max-eigenvalue image from data set with file number 008. This is a crop containing the phantom region of an axial slice. (b) Result of segmentation procedure described in section 2.2.3. The image shown here was produced by multiplying the binary image to the corresponding original max-eigenvalue image. (c) The location of the voxel of max intensity is determined for each bundle, and the 3x3 neighborhood about this location is established. (d) The number of voxels for which the z-component of the eigenvector are largest are counted for each 3x3 neighborhood.

Figure 2.32 shows the Bundle Detection Array corresponding to max eigenvalue images of the 16 datasets. If a bundle was not detected, it was decided that the maximum directional diffusion was not large enough to warrant assessing the orientation accuracy. It is evident that the

images shown in this figure are similar to those of the FA contrast test shown in Figure 2.30. However, there is no need to formally consider the trends, as the test criterion will not be drawn from this array.

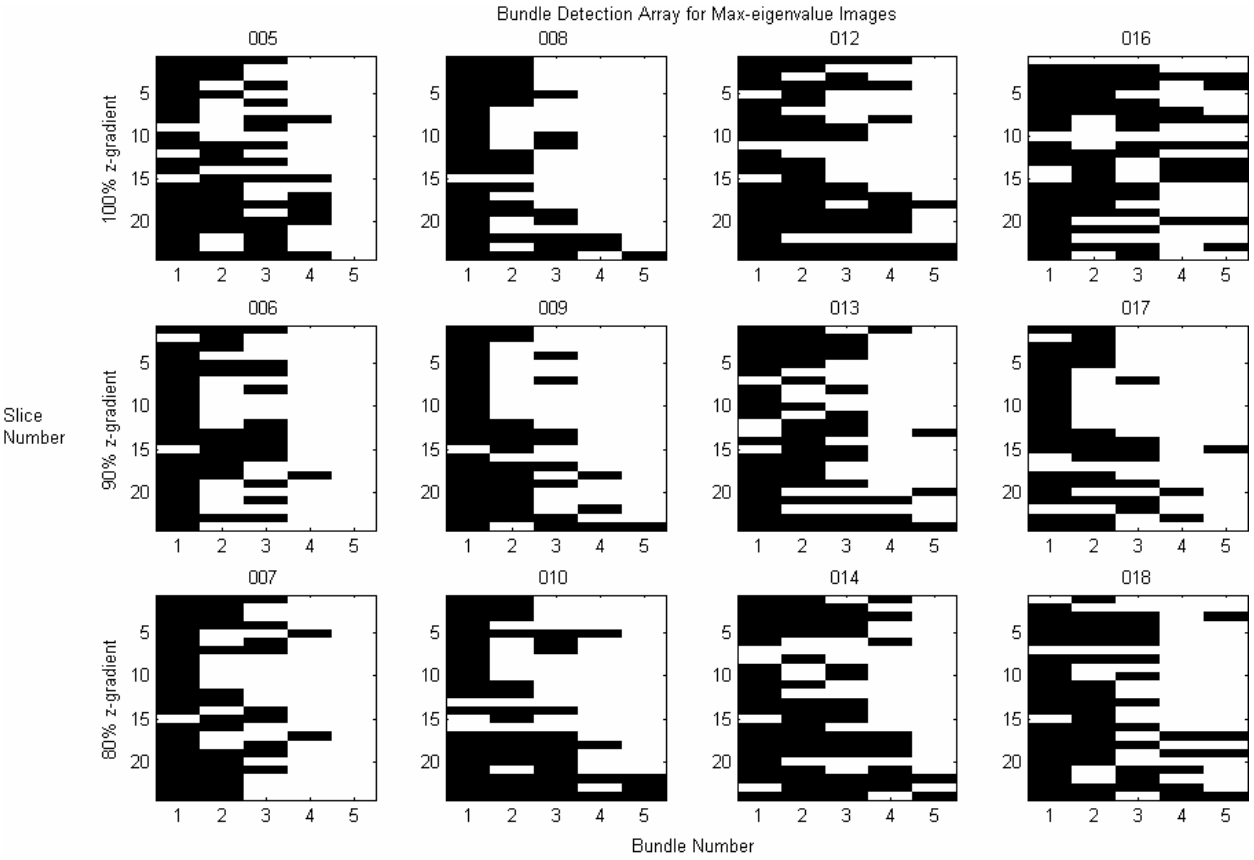


Figure 2.32 – Bundle Detection Array for max-eigenvalue images indicating which bundles were detected for a given slice. Each image pertains to a data set, and the images are arranged in the same order as Table 2.1. Images in the same row have the same z-gradient strength. Images in the same column are of the same repetition number. For a particular image, the image-rows pertain to the slice number, and the image-columns pertain to the bundle number. The bundles are listed in the same order as they appear in the axial slices, i.e. from left-to-right. For a given series, if a bundle was detected, the pixel of the corresponding slice and bundle number in Bundle Detection Array is set to white.

Figure 2.33 shows an array that stores the count of voxels with the eigenvector having a max z-component for the bundles. This array is similar to the Bundle Detection Array with the difference being that the max z-component voxel count is indicated instead of just the bundle

presence. By comparing Figures 2.32 and 2.33, one can see that for some bundles that were detected, no voxel had an eigenvector with the z component being the largest. This occurs more often as the z-gradient strength decreases, which is expected. This means that, as the z-gradient strength decreases, the orientation accuracy also decreases. Table 2.6 shows the total number of bundle cross sections that were determined to be correct with regard to orientation. The orientation was determined to be correct if at least one voxel of a bundle cross section had the z-component being the largest component of the eigenvector. If a decision threshold is set between 50 and 46, only the data sets that pertain to the largest z-gradient would pass using this criterion.

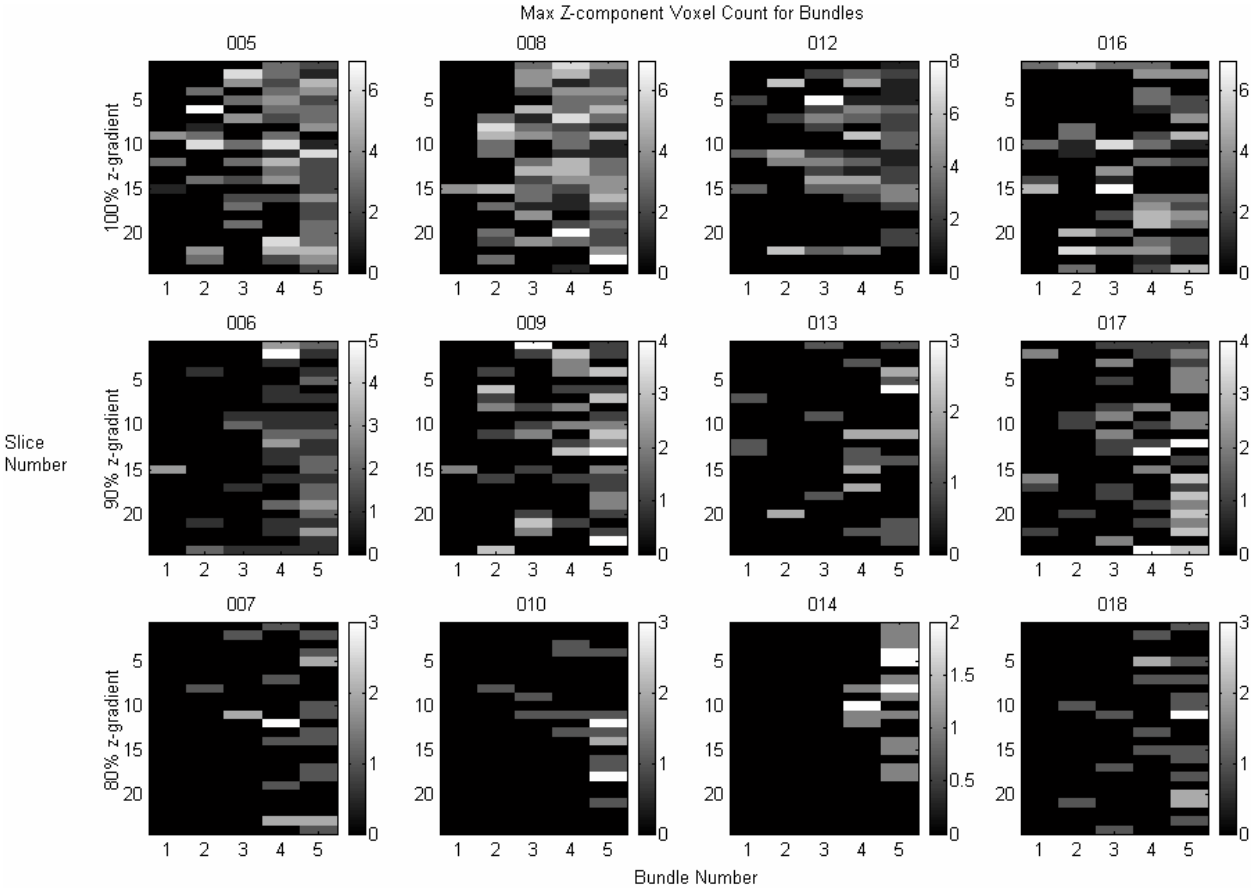


Figure 2.33 – Max z-component voxel count for bundles. Each image pertains to a data set, and the images are arranged in the same order as Table 2.1. Images in the same row have the same z-gradient strength. Images in the same column are of the same repetition number. For a particular image, the image-rows pertain to the slice number, and the image-columns pertain to the bundle number. The bundles are listed in the same order as they appear in the axial slices, i.e. from left-to-right. For a given series, if a bundle was detected, the pixel of the corresponding slice and

bundle number in the Bundle Detection Array is assigned the number of voxels in the corresponding 3x3 neighborhood that for which the z-component of the eigenvector was largest.

Table 2.6 – Number of bundle cross sections that consist of at least 1 voxel with the z component being largest in the corresponding eigenvector for entire series.

z-gradient (relative to max)	repetition 1	repetition 2	repetition 3	repetition 4
100 %	61	71	51	50
90%	45	45	23	46
80%	20	16	17	22

The previously mentioned criterion does not incorporate the varying bundle sizes. Tables 2.7 and 2.8 show the number of slices that contain at least 4 and 5 bundles of correct orientation respectively. Similar to the FA Contrast test, a slice is decided to have 4 correct bundles only if these are the 4 largest bundles. If I use a criterion that related to the presence of 4 correct bundles in a slice, the decision threshold can be set to 2, and only the series pertaining to the largest z-gradient strength would pass. However, if the criterion of all 5 bundles being correct was used, there is no threshold that can be chosen to separate the series of the 100 and 90 percent z-gradient strengths.

Table 2.7 – Number of slices in a series that contained at least 4 bundles of correct orientation.

z-gradient (relative to max)	repetition 1	repetition 2	repetition 3	repetition 4
100 %	2	6	4	2
90%	1	0	0	0
80%	0	0	0	0

Table 2.8 – Number of slices in a series for which all 5 bundles were of correct orientation.

z-gradient (relative to max)	repetition 1	repetition 2	repetition 3	repetition 4
100 %	0	1	1	1
90%	0	0	0	0
80%	0	0	0	0

2.3 Chapter Summary

In this chapter, algorithmic steps of the MRI and DTI QA procedures were described in detail. For MRI QA, the components described included the image registration procedure and the automated ACR accreditation program tests. For DTI QA, FA Contrast and axon Orientation Accuracy tests were introduced and described. Also, the steps taken in constructing and imaging the physical phantom were described. In the next chapter, verification of the MRI QA procedure is discussed followed by the longitudinal QA analysis of the scanner at WFUBMC. Finally, the conclusions and a discussion of future work will be provided.

Chapter 3 Results and Conclusions

3.1 Verification of ACR Quality Assurance

The verification of the ACR QA procedure was performed in a hierarchical nature. The system validity was assessed on three levels, which were the 1) procedural level, 2) test level, and 3) measurement level. Sections 3.1.1 through 3.1.3 describe this hierarchical verification using T_1-*weighted* axial data while similar testing using T_2-*weighted* data is described in section 3.1.4.

3.1.1 Procedural Level Verification

For the procedural level, one sagittal localizer and 16 axial series were used. Two of these axial series were base scans that had the normal imaging parameters prescribed by the ACR, but the other 14 axial series were purposefully altered prior to scanning to facilitate forced failure testing. The decision was made to use one sagittal localizer in this test in order to limit the variability to only the axial series, which is reasonable since only one measurement is performed on the localizer. This localizer was created with the ACR prescribed imaging parameters. The automated ACR QA procedure was executed using these series, and assuming that the scanner had no deficiencies, it was expected that the base scans would be the only axial series that would pass all tests, and thus, the entire procedure. As expected, the automated procedure correctly detected the base scans, which can be seen in Table 3.1 (along with the altered parameters).

Table 3.1 – Forced failure specifications and full procedure result for axial series 5 through 20.

Axial Series Number	Full Procedure Result	Alteration(s)	Gradient Calibration Scale Factor		Slice Thickness (mm)	Slice Spacing (mm)	NEX	TG
			cfxfull	cfyfull				
5	Pass	None: Base Scan	25639	25486	5	5	1	111
6	Fail	Gradient: cfxfull 10% delta	23075	25486	5	5	1	111
7	Fail	Gradient: cfxfull 5% delta	24357	25486	5	5	1	111
8	Fail	Gradient: cfyfull 10% delta	25639	22937	5	5	1	111
9	Fail	Gradient: cfxfull 5% delta, cfyfull 10% delta	24357	22937	5	5	1	111
10	Fail	Gradient: cfyfull 5% delta	25639	24212	5	5	1	111
11	Fail	Gradient: cfyfull 2% delta	25639	24976	5	5	1	111
12	Fail	Gradient: cfxfull 2% delta, cfyfull 2% delta	25126	24976	5	5	1	111
13	Fail	Slice Thickness. Slice Spacing	25639	25486	1	9	1	111
14	Fail	Slice Thickness. Slice Spacing. NEX	25639	25486	1	9	4	111
15	Fail	Slice Thickness	25639	25486	8	2	1	111
16	Fail	Slice Thickness	25639	25486	6	4	1	111
17	Pass	None: Base Scan	25639	25486	5	5	1	111
18	Fail	Uniformity: TG. Gradient: cfxfull 5% delta	24212	25486	5	5	1	133
19	Fail	Uniformity: TG. Gradient: cfxfull 5% delta	24212	25486	5	5	1	88
20	Fail	No Shim	25639	25486	5	5	1	111

3.1.2 Test Level Verification

For test level verification, the objective was to consider the axial series that were forced to fail and determine whether the expected tests were failed or passed based on the parameters that were altered for a given series. Four parameters were altered for this experiment. The gradient calibration scale factors, for the x (cfxfull) and y (cfyfull) gradients, were reduced by

varying percentages, and the expectation was that the diameters measured should also reduce by the corresponding percentages. This was done for series 6 through 12, 18 and 19. A reduction in the x and y gradient calibration scale factors should reduce the horizontal and vertical diameters respectively by the same percentage. For series 13 through 16, the slice thickness and slice spacing were changed such that they would still sum to 10 mm. (Note that the ACR prescribed slice thickness and spacing are both 5 mm.) For series 14, a NEX value of 4 was used, which would increase the series SNR. The transmit gain (TG) was used to miscalibrate the flip angle in order to reduce the SNR for series 18 and 19. Series 20 was created without the use of any shims. Metal shims or correction coils are normally used to correct the inhomogeneity of the magnetic field. Without these shims, the quantitative aspects of imaging, such as geometric dimensions, slice thickness, and slice position can be affected. In some instances, the validity of the pass/fail decisions for certain tests was assessed by performing the associated steps of the manual procedure.

The series for which the gradient calibration scale factor(s) were altered were the only ones to fail the Geometric Accuracy test, which was expected. Also, the reduction in the measured diameter was about the same percentage as the reduction of the calibration scale factors. Refer to Appendix A to view the automated procedure measurements.

Series 13 failed the High contrast Spatial Resolution test because the slice thickness was only 1 mm, which resulted in a low SNR. Series 14 had the same imaging parameters as series 13 except for an increased NEX, which resulted in an increased SNR. Thus, the holes in the resolution insert had higher intensity, and so series 14 passed this second test. However, with careful window and level adjustments in the manual procedure, series 13 can also pass the spatial resolution test. That is, an observer would be able to recognize all four holes in the first row of the UL array as points of brighter signal intensity than the spaces between them [3]. These intensity peaks are not high enough for any threshold (used by automated procedure) to segment all 4, so the automated test result was a FN. Series 16 fails the spatial resolution test because the slices are displaced greatly causing slice 1 to not pass through the center of the resolution insert. Therefore, there is very little contrast between the holes and the background. It was also discovered that, with a 5 percent or greater change in the diameter, there was an increased likelihood that the spatial resolution test would be incorrectly failed. Series 6, 7, and 10 were such cases. This failure may result from misregistration of the hole-array pairs prior to

processing. However, series 9 correctly passed the test and had diameter reductions in the vertical and horizontal directions of 10 and 5 percent respectively. All other series passed this test.

The series for which the slice thickness was altered failed the Slice Thickness Accuracy test. The slice thicknesses measured for these series were within +/- 0.41 mm of the thickness specified except for series 16 with the reason being that the substantial slice displacement caused the signal ramps to be truncated at the ends of the thickness insert. Series 18 also failed this test although the correct thickness was prescribed. This series most likely failed because of the 5 percent reduction in the x gradient calibration scale factor causing the phantom to be scaled down in size along the horizontal direction. Thus, the signal ramps were also shortened enough to cause this failure.

As mentioned earlier, the slice thickness and slice spacing were changed for series 13 through 16 such that both still summed to 10 mm. Therefore, the center-to-center distance between slices would be 10 mm for these series. Thus, assuming that the same positions were specified, the slices should be centered about the same locations. However, of these four series, only 13 and 14 passed the Slice Position Accuracy test. This may be because the thin slices (1 mm) can pass directly through the vertex of the crossed wedges while the thicker slices of series 15 and 16 (8 and 6 mm respectively) do not. Series 20 also failed this test, which is understandable since no shim was used and an excessively inhomogeneous magnetic field would not facilitate correct slice locations.

For the Image Intensity Uniformity test, the series that had the TG parameter altered failed as expected. Since no parameters were altered that would affect the intensity uniformity for series 8 and 9, the failure of these series was unexpected. These failures were due to the 10 percent reduction in the y gradient calibration scale factor. Thus, the phantom was shortened along the vertical direction causing the low intensity box at the top of the phantom to be partially present in the large ROI. Part of this low intensity box contributes to the low intensity average determined in this test. The large ROI was on the low end of the size range specified in the ACR guidance manual, and so can not be made smaller. However, in practice, it is unlikely that the phantom diameter will vary from the truth by 10 percent. Since no steps were performed to induce any signal ghosts, all series passed the Percent Signal Ghosting test. Also, with the ACR prescribed passing criteria being low, all series passed the Low Contrast Object Detectability test.

3.1.3 Measurement Level Verification

In the measurement level verification, the validity of the measurements for each test of the automated procedure was assessed. The truth for the ACR QA tests can be provided solely by manual testing. Therefore, the automated procedure measurements must be compared to manual procedure measurements. Of primary concern were the measurements pertaining to the forced failure testing. The associated series would allow for an understanding to be gained about the FP rate versus the FN rate of the automated procedural tests. Because of time limitations, only 5 altered series and participants were used for this assessment. Each participant performed the full manual procedure on 2 series, and the series overlapped between participants such that each series was assessed twice. This was so the inter-observer variability could be assessed. The decision was made to use the altered series 15, 16, 18, 19 and 20 so that the effect of all alterations could be assessed. The results from the manual procedure can be viewed in Appendix A.

To provide a frame of reference, I performed the manual procedure on series 17, which was one of the base scans. As expected, all tests were passed, but I decided to compare my measurements with that of the automated procedure. Table 3.2 shows the absolute differences between the final measurements for the manual and automated procedures. Tests 1 and 4 have multiple final measurements (measurements compared to criteria for pass/fail decision), so the maximum difference is shown. To provide a sense of the relative difference, the percent difference is also shown for each test. This was calculated using the ratio of the absolute difference to the manual measurement. Tests 3, 4, and 6 have high percent differences because the associated measurements have relatively low values. This table shows that my measurements were very close to those of the automated procedure. However, one might argue that this result is biased since I am the creator of the automated procedure. For this reason, independent third party investigators were asked to perform the testing on the altered series, and their measurements will now be compared to those of the automated procedure.

Table 3.2 – Differences between human (author) and automated observer measurements for base scan: axial series 17

Test	Max absolute difference	Percent difference
1. Geometric Accuracy	0.7073	0.3719
2. High Contrast Spatial Resolution	0	0
3. Slice Thickness Accuracy	0.1582	3.2608
4. Slice Position Accuracy	0.2330	10.5500
5. Image Intensity Uniformity	0.2755	0.2893
6. Percent Signal Ghosting	0.0008	24.6546
7. Low Contrast Object Detectability	0	0

The first measurement performed in the Geometric Accuracy test is on the sagittal localizer. While assessing two axial series, the participants unknowingly assessed the same localizer twice. The intra-observer variability between these measurements can be observed in Figure 3.1. Though, the same procedure was repeated minutes after the first measurement, human observers 1 and 2 had different measurements. One benefit of the automated procedure is that its measurements will be consistent for the same data, so it is more predictable. From this figure, one can also see the inter-observer variability. However, half of the measurements are about 148.3 +/- 0.005 mm, which is relatively close to the automated procedure measurement of 148.112 mm. The red line and arrows of this figure show that all measurements pass the criteria of being within +/- 2 mm of the true end-to-end length, as explained in chapter 2.

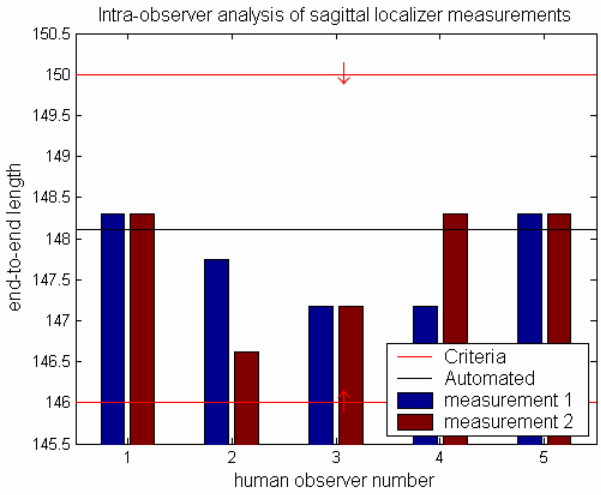
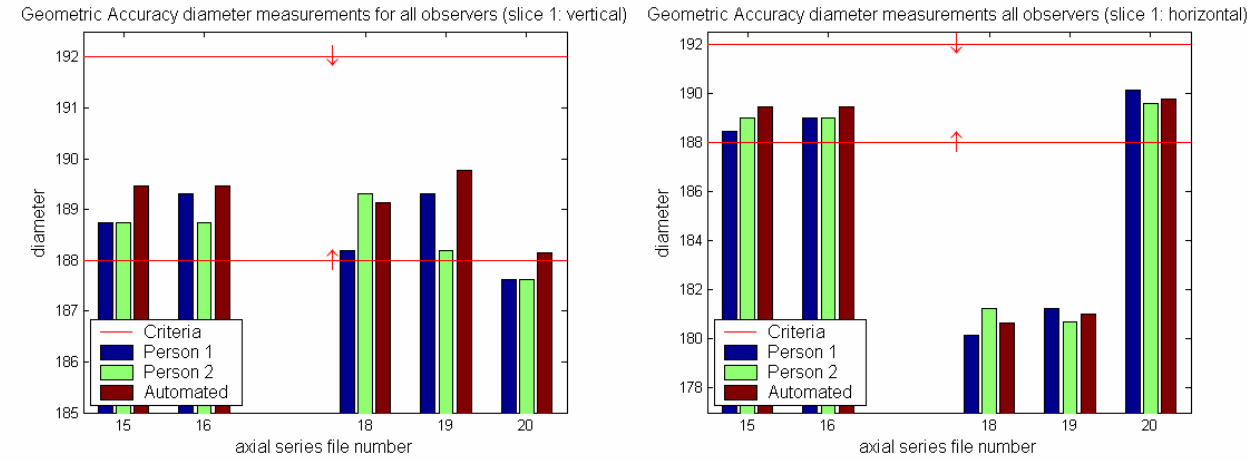


Figure 3.1 End-to-end length measurements of sagittal localizer for human and automated observers. The intra and inter-variability of the measurements can be seen.

Subsequently, the diameters of slices 1 and 5 of the axial series were measured. Figure 3.2 shows the vertical and horizontal measurements of slice 1. The observer measurements are plotted for each of the altered series being considered. For a given series, the human observer listed as "Person 1" is the first person listed in the tables of Appendix A. The pass/fail results corresponding to the manual measurements of this slice were the same as those for the automated procedure except for series 20. The low vertical measurements of series 20 (for all observers) were a result of an inhomogeneous field caused by the lack of shims. The measurement of the automated observer is just above the low criterion margin causing a FP. Not shown in this figure are the measurements for slice 5. One point of interest for these measurements is that the second human observer of series 16 has diagonal measurements that are below the low criteria margin. However, this is a FN because the first human observer of the series, the automated observer, and I agree that series 16 should pass this test. Also, no parameters were altered to cause this series to fail this test.



(a) (b)

Figure 3.2 –Diameter measurements for slice 1 of considered altered series performed by human and manual observers. (a) Vertical measurements. (b) Horizontal measurements.

For the High Contrast Resolution test, the pass/fail decisions based on the results from all human observers were consistent with those of the automated procedure. As expected, series 16 was the only one to fail as made evident by Figure 3.3. However, there are a few instances where the hole-counts are inconsistent. To resolve these inconsistencies, I performed the manual tests.

For the left pair of arrays of series 16, I agreed with the first human observer disagreeing with the other observers. For the center-pair of series 16, I agreed with the first human and automated observers. I disagreed with the automated procedure for the right-pair of series, 15, 16, and 18. For 15 and 18, the threshold method of hole-segmentation fails for the automated procedure, but one of the human observers was able to suitable set the display level and widow in order to correctly count the number of holes in the LR array. For the right-pair of series 16, the large slice displacement results in very little contrast between the holes and the background and I could not detect any holes. I agreed with the automated observer for the right-pair of series 20 disagreeing with the other observers. In summary, the counts for the automated procedure were correct 11 out of 15 times. The incorrect cases pertained to either large slice displacement or the limitations of the threshold method.

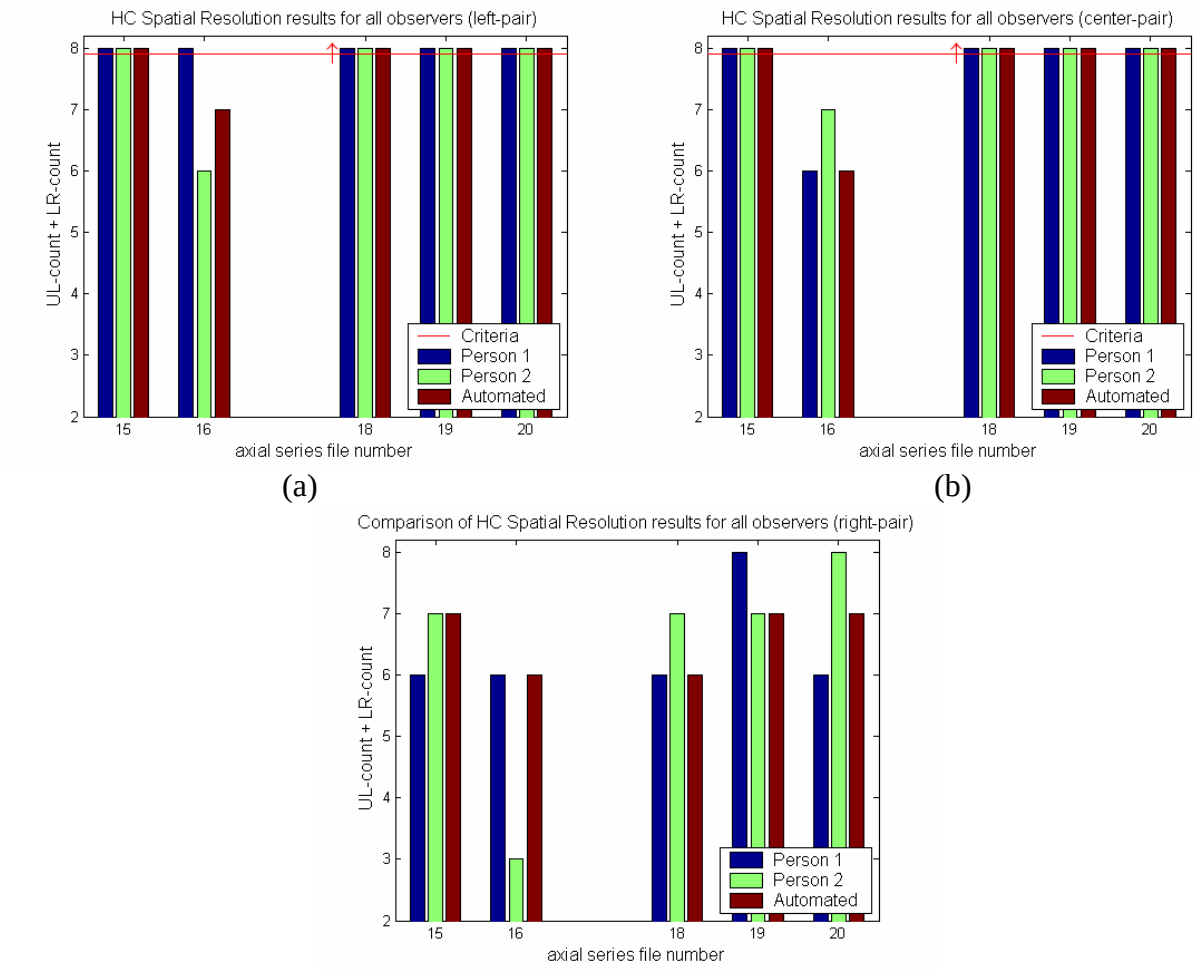


Figure 3.3 –High Contrast Spatial Resolution results for considered altered series performed by human and automated observers. For brevity, the UL and LR array counts are summed. (a) Left-pair. (b) Center-pair. (c) Right-pair.

The first person to perform the Slice thickness accuracy for both series 16 and 18 had measurements that varied substantially from those of the other observers as can be seen in Figure 3.4. I measured a slice thickness of 3.399 for series 16, which is along the lines of the measurements for Person 2 and the automated observer. As stated earlier, the large slice displacement caused the signal ramps to be located at the ends of the thickness insert, which is why the thickness measured is not what was prescribed. Another consequence of this is that it was possible for a human observer to accidentally incorporate phantom regions to the left and right of the thickness insert into the measurement, which is probably what Person 1 did. The first

person to measure the slice thickness for series 19 had a low measurement that would cause a FN. My measurement was 4.885, which is similar to the measurements of the other observers.

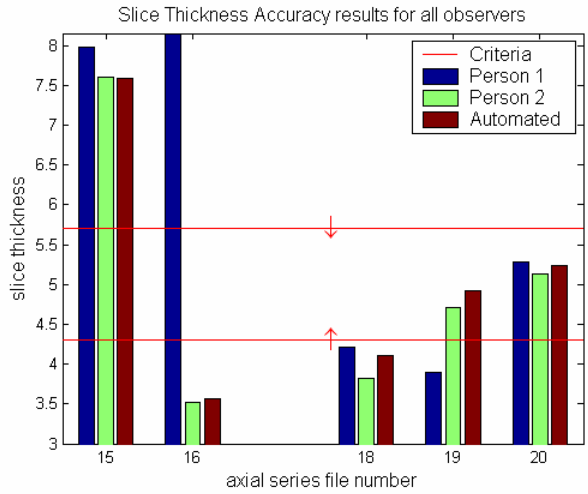
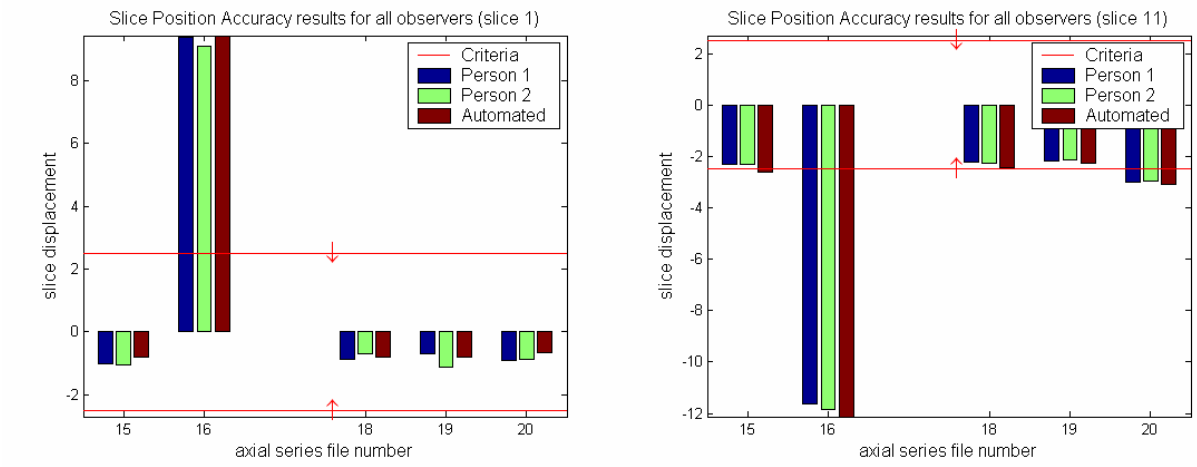


Figure 3.4 – Slice Thickness Accuracy Results for considered altered series performed by human and automated observers.

The human observer results for the Slice Position Accuracy test were as expected for all series except for 15. The automated procedure reported a displacement that was too long resulting in a FN for slice 11. This is made evident by Figure 3.5. I measured a displacement of 2.934 mm, which is along the lines of the other observer's measurements. Notice that slice 1 of series 16 is displaced superiorly while slice 11 is displaced inferiorly. This is a consequence of the altered slice thickness and spacing and incorrect prescribed slice locations.



(a) (b)

Figure 3.5 – Slice Position Accuracy results for considered altered series performed by human and automated observers. (a) Slice 1. (b) Slice 11.

The size and placement of the Large ROI can greatly affect the value of the average low signal determined in the Image Intensity Uniformity test. Thus, there can be substantial variability between PIU values for a series. This can be seen when viewing the measurements for series 19 and 20 in Figure 3.6. Because of this, it is difficult to say that an error was made in any of the measurements, but Person 1 had a low measurement that caused a FN for series 20. I determined a PIU value of 90.164, which is just above the criterion threshold.

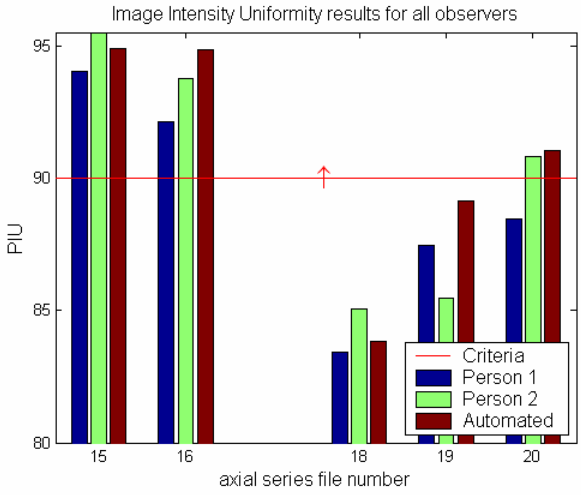


Figure 3.6 – Image Intensity Uniformity results for considered altered series performed by human and automated observers.

Since no ghost signals were induced, it was not surprising that the human observer's measurements were well below the criteria. Figure 3.7 also shows that these measurements were relatively close to the automated procedure measurements.

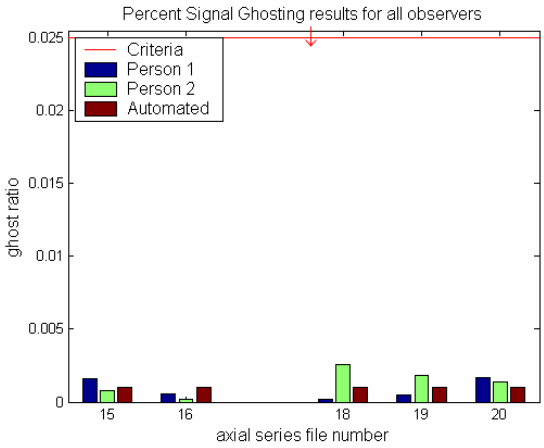


Figure 3.7 – Percent Signal Ghosting results for considered altered series performed by human and automated observers.

It was interesting to see how the hole detection procedure, based on the Rose Model, faired compared with human observers for the Low Contrast Object Detectability test. The spoke-counts for the automated procedure were generally well correlated with those of the human observers except for series 18 and 19. This manual test is fairly subjective, and results would differ depending on a person's choice of display window and level settings and his or her visual perception. Therefore, I decided to perform the manual procedure for comparison. For series 15, 16, and 20, the automated observer's spoke-count was within 3 of the count from one of the observers. I counted 32 spokes for series 15, which happened to be the same as the automated observer count. For series 16, my count of 20 was higher than the counts for the other observers. I counted 28 spokes for series 20, which was 1 less than the counts of the other humans and 4 less than that of the automated observer. However, for series 18 and 19 my spoke-counts of 32 and 29 were similar to that of the human observers, but the automated observer's count was substantially higher. Generally, I think that the results of automated procedure are suitable. It would probably be best to slightly raise the correlation-threshold used to decide whether a hole is detectable. Further experiments would be needed in this regard though. Notice that the large slice displacement in series 16 caused the spoke-counts to be relatively low, but the

series still passed easily. For a series to have fewer than 9 complete spokes counted, the CNR would have to be very low since each of the 4 slices considered has 10 spokes.

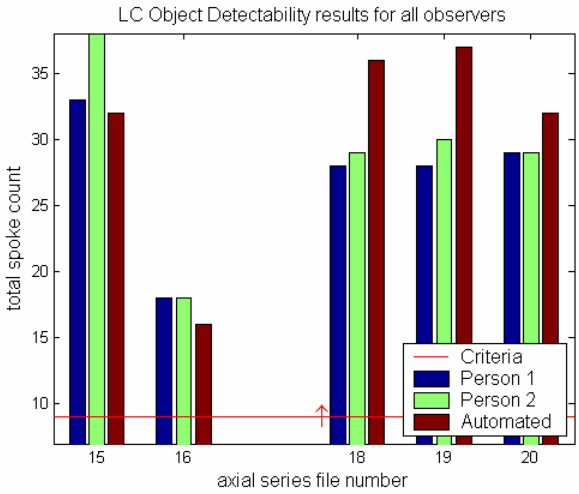


Figure 3.8 – Low Contrast Object Detectability results for considered altered series performed by human and automated observers.

3.1.4 Verification Using T_2-*weigted* Axial Data

The axial data used for the aforementioned verification testing was T_1-*weigted*. Similar verification testing was performed with 12 T_2-*weigted* axial series and 5 additional sagittal localizer slices. Imaging parameters that pertain to quantitative imaging were altered such that certain measurements were expected for the Geometric Accuracy, Slice Thickness Accuracy, and Slice Position Accuracy tests. Four base scans (axial series) were included to provide a frame of reference, and these series passed the full procedure as expected. For two of the axial series, the x gradient calibration scale factor was reduced slightly to cause the horizontal diameter measurements in slices 1 and 5 to also be reduced resulting in a failure of the first test. For two other axial series, the slice thickness was prescribed to be 3 mm. The measured thickness was within +/- 0.109 mm of the specified value, and so the Slice Thickness Accuracy test was failed. Incorrect, slice positions were prescribed for the other two series causing the Slice Position Accuracy test to be failed. For verification testing of the sagittal localizer measurement process, two base scans were assessed, and the other three slices were created with less than 2.6 percent reduction in the x and y gradient calibration scales. However, these reductions were not enough

to cause failures in the Geometric Accuracy test. Since the measured length should vary by less than about 1.4 percent from the truth for this test to be passed, it was possible that the results might have been FPs. However, I measured the end-to-end lengths manually, and determined that all localizer slices correctly passed.

3.2 Longitudinal QA Analysis of scanner at WFUBMC

This automated MRI QA procedure is intended for daily or weekly image quality assessment of scanners. Therefore, a longitudinal QA analysis of a 1.5 T GE TwinSpeed MR scanner (with specifications listed in section 2.2.2) at the WFUBMC was performed in order to determine the usefulness of the procedure for this purpose. For this study, the ACR phantom was imaged 7 times between March 29th and May 31st 2005 with the ACR prescribed scanning protocols. The other 5 dates were April 12th and 19th and May 9th, 10th, and 17th. The QA procedure was passed for each date, and the statistics of the measurements across the dates are shown in Table 3.3.

Table 3.3 – Statistics of measurements across dates for each test.

Test / measurement	Mean of measurements	Variance of measurements
1. Geometric Accuracy (localizer – length)	147.50714	0.22702
1. Geometric Accuracy (axial – diameter)	189.63157	0.03729
2. High Contrast Spatial Resolution (right-pair)	3.35714	0.24725
3. Slice Thickness Accuracy	5.00929	0.00269
4. Slice Position Accuracy	1.04650	0.70388
5. Image Intensity Uniformity	94.76343	0.41640
6. Percent Signal Ghosting	0.00029	0.00000
7. Low Contrast Object Detectability	33.00000	1.00000
SNR	39.07614	8.32278

Recall that 6 diameter measurements are assessed in the first test, so all measurements were used to calculate these statistics. Similarly, the absolute values of the displacements corresponding to both slices 1 and 11 were used to calculate the statistics for the Slice Position Accuracy Test. For the High Contrast Spatial Resolution test to be passed, the left and center pairs must have 4 detectable holes for the UL and LR arrays, so only the right pair of arrays was used to calculate these statistics.

No interesting trends were noticed for the measurements, such as them gradually decreasing or increasing, but the variance of the measurements shows that there was little variation relative to the ACR criteria. Notice that for the quantitative imaging related tests (1, 3, and 4) that the means are relatively close to the truth when compared with the criteria margins. With the exception of the fourth test, the means of the qualitative imaging related tests well exceed the pass criteria. These means and variances indicate that based on the ACR pass criteria, the likelihood of future failures is low. This is understandable because the ACR criteria are indicative of minimum level of performance that can be expected from a well functioning scanner [3]. Further statistical longitudinal analysis can allow for stricter yet suitable criteria to be decided upon in order to attain a desired level of scanner performance. The statistics for the High Contrast Resolution test indicate that it is more likely for 3 holes to be detected than 4 for the right pair of arrays. This is understandable because these holes are smaller than the voxels. Thus, the current ACR pass criterion is optimal for this test. The ACR requires that a SNR test be incorporated into a site's QA procedure. The SNR statistics can be used help in defining an appropriate passing criterion.

3.3 Conclusions and Discussions

3.3.1 Conclusions

The primary focus and accomplishment of this thesis was the full automation of the ACR accreditation program for use in daily or weekly QA. It was determined that this procedure returns reliable and repeatable measurements that are generally well correlated with those of human observers. The measurements obtained from the manual procedure are the only truth, but subjectivity inherent in pre-measurement steps cause intra-observer variability. The automated procedure is more predictable. The full automated procedure takes approximately 5 minutes to complete (on a Dell Precision Workstation with a 800 MHz Pentium 4 processor) compared to 15 minutes for the manual procedure. This is not a significant reduction in time, but a site employee can be free to perform other tasks while the daily QA is being done. The results of the procedure are logged to two types of files. The first file type was designed to allow users to easily view and understand results for all tests in one central location. Results for each test are

also logged to individual files, and the measurements are ordered in respective columns. The intent here was to make longitudinal analysis easier.

Forced failure testing allowed for some limitations to be discovered. Substantial geometric distortion compromises the accuracy of the High Contrast Spatial Resolution test. However, an incorrect result should only be a FN, which is more acceptable than a FP. Another limitation of this test is that the automated threshold method is less sensitive than the manual window and level adjustment method at segmenting all holes, but this only contributes to the FNs. Great geometric distortion can also affect the efficacy of the other tests largely because the image registration procedure may fail. The measurements returned by the Geometric Accuracy test can be considered in deciding whether the other test results are valid. Still, in practice, the correct imaging protocols will be used and it will be highly unlikely that the measured phantom dimensions will vary by 5 percent or more from the truth. Longitudinal analysis has shown that the measurements have low variance relative to the ACR prescribed pass criteria. Thus stricter criteria can be set for all tests except for the High Contrast Spatial Resolution tests in order to reduce the number of FPs relative to the ACR criteria. Note that for most tests, FPs and FNs occurred more often for the third-party investigators than for the automated procedure. The motion parameter range for which the registration procedure can be expected to converge correctly is +/- 30 mm for translation and +/- 20 degrees for rotation, which should be more than sufficient.

A DTI QA procedure was developed that tested whether a scanner was producing images with suitable FA contrast and axonal orientation accuracy relative to axon size. A physical phantom was created to model the restrictive diffusion that occurs on the interior of the axons. It was discovered that the choice of imaging parameters (especially resolution) were appropriate such that the bundle orientation could be determined yet the SNR was high enough to allow for sufficient contrast. Criteria were selected that would allow only images created with the correct z-gradient magnitude to pass the FA contrast and bundle Orientation Accuracy tests.

3.3.2 Discussion of Further Work

As stated earlier, the ACR requires that an SNR test be incorporated into site's QA procedures. Currently, the automated procedure performs the SNR measurement as described in

section 2.1.7, but this measurement is not compared to any criteria in any formal test. Further longitudinal analysis can allow sufficiently strict criteria to be determined for this test. Typically, one cannot have a SNR that is too high so only a lower criteria threshold would be needed, but an upper threshold can be used to ensure that the SNR is not uncharacteristically high. It would have to be decided if a failure of this test should cause a full-procedure failure.

The ACR provides very specific instruction for positioning and aligning the phantom when creating data for submission for accreditation. As a pre-qualifying step, the correct phantom alignment can be assessed by assessing the rotation and translation parameters returned by the motion estimation algorithm. Currently, these values are logged in a report with each scan, but a formal test can be formulated that compares the parameters to appropriate criteria margins. As for SNR, it would have to be decided if a failure of this test should cause a full-procedure failure. Great phantom misalignment can cause incorrect motion parameters to be returned, so another consideration would be to save the sagittal localizer and slice 1 of the axial series as external image files for later viewing.

The work done in developing the DTI QA procedure was a preliminary attempt. A simple phantom was developed that only pertained to restricted diffusion. A more realistic phantom is desired that also incorporates hindered diffusion, which occurs outside the axon [5]. The effect of varying axon orientation would also have to be incorporated. Ultramicrobore tubing should be used since it is a better model of axons in terms of size. In using ultramicrobore tubing, the resolution of the images must increase largely because of the comparatively low water density in the phantom. In order to maintain an acceptable SNR, signal averaging can be performed at a high rate. Pass criteria can either be determined experimentally or with the consideration of a theoretical model.

Bibliography

1. I. Agartz, J.L. Andersson and S. Skare. "Abnormal brain white matter in schizophrenia: a diffusion tensor imaging study," *Neuroreport*, Vol.12, No.10, July, 2001.

2. American College of Radiology, "MRI Quality Control Manual," 2001.

3. American College of Radiology, "Phantom Test Guidance for the ACR MRI Accreditation Program," 1998.

4. American College of Radiology, "Site Scanning Instructions for Use of the MR Phantom for the ACR MRI Accreditation Program," 1998.

5. Yaniv Assaf, Raisa Z. Freidlin, Gustavo K. Rhode, and Peter J. Basser, "New Modeling and Experimental Framework to Characterize Hindered and Restricted Water Diffusion in Brain White Matter", *Magnetic Resonance in Medicine*, 52:965–978 (2004)

6. G. J. Barker, P.S. Tofts, "Semiautomated Quality Assurance for Quantitative Magnetic Resonance Imaging," Magnetic Resonance Imaging, Vol. 10, pp. 585-595, 1992.

7. Harrison H. Barrett, Jie Yao, and Jannick P. Rolland, "Model observers for the assessment of image quality," *Proceedings on National Academy of Sciences*, Vol. 90, pp 9758-9765, November 1993, Colloquium Paper.

8. H. H. Barrett, T. A. Gooley, K. A. Girodias, (1991) XIIth International Conference on Information Processing in Medical Imaging, eds. Colchester, A. C. F & Hawkes, D. J. (Springer, New York), pp. 458-473.

9. Mark E. Bastin, "On the use of water phantom images to calibrate and correct eddy current induced artifacts in MR diffusion tensor imaging," Magnetic Resonance Imaging 18 (2000) 681-687.

10. Robert Bell, "ACR MRI Accreditation: Myth and Reality," [Internet WWW], *Address:* http://www.imagingeconomics.com

11. Denis Le Bihan, Jean-Francois Mangin, Cyril Poupon, "Diffusion Tensor Imaging: Concepts and Applications," Journal of Magnetic Resonance Imaging, 2001, 13: 534-546

12. Arthur E. Burgess, "The Rose Model, revisited," *Journal of Optical Society of America*, Vol. 16, No. 3. March 1999.

13. Cardinal Health. "MRI PC/AutoQA Software," [Internet WWW]. *Address:* http://www.cardinal.com/rms

14. Richard O. Duda, Peter E. Hart, David G. Stork, Patter Recognition, Second Edition, John Wiley and Sons, Inc., 2001.

15. Leslie Farnsworth, "ACR Prepares for MRI Accreditation Demands of Next Millennium," [Internet WWW], *Address:* http://www.radinfonet.com/standards/accred_update.html

16. R. D. Fiete, H. H. Barrett, W. E. Smith, (1987) *J. Opt. Soc. Am. A. Opt. Image Sci.* 4, 945-953.

17. M. J. Firbank, R. M. Harrison, and E. D. Williams, "Quality Assurance for MRI: practical experience," The British Journal of Radiology, vol. 73, pp. 376-383, 2000.

18. R. A. Fisher, (1936) *Ann Eugenics* 7, 179-188.

19. E. A. Gardener, J. H. Ellis, and R. J. Hyde, " MRI Scanner Variability Studies using a Semi-Automated Analysis System," Magnetic Resonance Imaging, pp. 1087-97, 12(7), 1994.

20. E. A. Gardener, J. H. Ellis, R. J. Hyde, "Detection of Degradation of Magnetic Resonance Images: Comparison of an Automated MR Image-quality Analysis System with trained human observers," Academic Radiology, pp. 277-81, Apr: 2(4) 1995.

21. Elisabeth A. H. von dem Hagen and R. Mark Henkelman, "Orientational Diffusion Reflects Fiber Structure Within a Voxel," *Magnetic Resonance in Medicine*, 48: 454–459 (2002)

22. Robert M. Haralick and Linda G. Shapiro. Computer and Robot Vision, Volume I. Addison-Wesley, 1992. pp. 28-48.

23. H. Hotelling, (1931) Ann. Math. Stat. 2, 360-378.

24. Steve J. Huff, "Software Development and Automation of the American College of Radiology MR Phantom Testing Accreditation Criteria For the General Electric Magnetic Resonance Imaging Whole Body Scanner," MSOE. May, 2003.

25. Ching-Po Lin, Van Jay Wedeen, and Jyh-Horng Chen, "Validation of diffusion spectrum magnetic resonance imaging with manganese-enhanced rat optic tracts and ex vivo phantoms," *NeoroImage*, vol19 (2003) 482-495

26. Moriel NessAiver, "Simply Physics," [Internet WWW], *Address:* http://www.simplyphysics.com/ACR_QC.HTM

27. N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62 – 66, 1979.

28. H. C. Park, Westin, M. Kubicki, et al. "White matter hemisphere asymmetries in healthy subjects and in schizophrenia: a diffusion tensor MRI study," *NeuroImage*, 2004

29. Radiological Imaging Technology. “Diagnostic Image QC Measurement Software,” [Internet WWW]. *Address:* http://www.radimage.com

30. A. Rose, “A unified approach to the performance of photographic film, television pickup tubes and the human eye,” *Journal of the Optical Society of America*, 47, 273-294 (1946).

31. A. Rose, “The sensitivity performance of the human eye on an absolute scale,” *Journal of the Optical Society of America*, 38, 196-208, (1948).

32. A. Rose, “Television pickup tubes and the problem of vision,” *Advances in Electronics and Electron Physics*, L. Marton, ed. (Academic, New York, 1948), Vol. 1, p. 131-166.

33. P. F. Sharp, C. E. Metz, R. F. Wagner, “Medical Imaging: the assessment of image quality,” *International Commission on Radiology Units and Measurements*, Bethesda, Md., 1996.

34. Andrew Simmons, Elizabeth Moore, and Steven C. R. Williams, “Quality Control for Functional Magnetic Resonance Imaging Using Automated Data Analysis and Shewhart Charting,” Magnetic Resonance in Medicine, 41, pp. 1274-1278, 1999.

35. Pierre Soille, Morphological Image Analysis: Principles and Applications, Springer-Verlag, 1999, pp. 173-174.

36. Alexander J. Taylor, “Diffusion Tensor Imaging: Evaluation of tractography algorithm using ground truth phantoms,” Virginia Tech, May 2004

37. Baba C. Vemuri and Shuangying Huang. “An Efficient Motion Estimator with Application to Medical Image Registration”. Medical Image Analysis (1998) volume 2, number 1, pp 79-98

38. Andrew Webb, Introduction to Biomedical Imaging, John Wiley and Sons, Inc., Hoboken, New Jersey, 2003.

39. Dominik Weishaupt, Victor D. Kochli, Borut Marincek, How does MRI work?, Springer-Verlag Berlin Heidelberg New York, 2003.

Appendix A: Automated ACR QA Verification Results for Human and Automated observers

Table A.1 – Automated procedure results for Geometric Accuracy test: only axial series.

Axial Series Number	slice 1 - diameter (mm)		slice 5 - diameter (mm)				PASS/FAIL
	vertical	horizontal	vertical	horizontal	45 diag	- 45 diag	
5	189.453	189.779	189.453	189.779	189.453	189.453	pass
6	189.128	171.875	189.453	171.875	179.688	179.688	fail
7	189.128	180.990	189.453	180.664	185.221	184.570	fail
8	171.549	189.779	171.549	189.779	180.013	180.013	fail
9	171.549	180.664	171.549	180.990	175.456	176.107	fail
10	180.664	189.779	180.664	189.779	185.221	184.896	fail
11	187.500	189.779	187.500	189.779	188.802	188.477	fail
12	184.896	186.849	184.896	187.174	185.872	185.872	fail
13	189.453	189.779	189.453	189.779	189.779	189.453	pass
14	189.453	189.779	189.453	189.779	189.453	189.453	pass
15	189.453	189.453	189.453	189.779	189.453	189.453	pass
16	189.453	189.453	189.453	189.779	189.453	189.453	pass
17	189.453	189.779	189.453	189.779	189.453	189.453	pass
18	189.128	180.664	189.453	180.664	185.221	184.570	fail
19	189.779	180.990	189.453	180.664	184.570	184.245	fail
20	188.151	189.779	188.477	189.453	189.453	188.151	pass

Table A.2 – Human observer (author) results for Geometric Accuracy test: localizer and base axial series.

Axial Series Number	localizer - length (mm)	slice 1 - diameter (mm)		slice 5 - diameter (mm)				PASS/FAIL
	vertical	vertical	horizontal	vertical	horizontal	45 diag	- 45 diag	
17	147.739	189.870	189.562	190.432	190.116	190.518	190.160	pass

Table A.3 – Human observer results for Geometric Accuracy test: localizer and altered axial series.

Participant Name	Axial Series Number	localizer - length (mm)	slice 1 - diameter (mm)		slice 5 - diameter (mm)				PASS /FAIL
		vert	vert	hor	vert	hor	45 diag	-45 diag	
Danielle Miles	15	148.305	188.746	188.466	189.308	189.564	189.382	189.052	pass
	16	148.302	189.309	189.007	189.309	189.010	189.399	189.481	pass
Lisa Demar	16	147.740	188.747	189.007	189.308	189.562	187.804	187.441	fail
	18	146.615	188.188	180.139	188.746	180.138	185.577	183.517	fail
Curtis Cox	18	147.173	189.309	181.247	189.871	180.693	185.167	184.759	fail
	19	147.177	189.309	181.247	189.308	181.247	185.140	185.198	fail
Ruth Cox	19	147.177	188.185	180.693	188.185	181.247	184.651	184.651	fail
	20	148.301	187.624	190.115	188.746	189.561	190.160	188.094	fail
Warren Liu	20	148.302	187.624	189.561	188.746	189.010	189.399	188.988	fail
	15	148.302	188.746	189.007	188.746	189.562	189.010	188.999	pass

Table A.4 – Automated procedure results for High Contrast Spatial Resolution test

Axial Series Number	left-pair		center-pair		right-pair		PASS/FAIL
	UL	LR	UL	LR	UL	LR	
5	4	4	4	4	4	3	pass
6	3	4	3	4	3	3	fail
7	4	4	4	4	3	3	pass
8	4	4	4	3	4	3	fail
9	4	4	4	4	3	3	pass
10	4	4	4	3	4	3	fail
11	4	4	4	4	4	3	pass
12	4	4	4	4	3	3	pass
13	4	4	3	4	3	3	fail
14	4	4	4	4	4	4	pass
15	4	4	4	4	4	3	pass
16	4	3	3	3	3	3	fail
17	4	4	4	4	4	3	pass
18	4	4	4	4	3	3	pass
19	4	4	4	4	3	4	pass
20	4	4	4	4	4	3	pass

Table A.5 – Human observer (author) results for High Contrast Spatial Resolution test: base axial series.

Axial Series Number	left-pair		center-pair		right-pair		PASS/FAIL
	UL	LR	UL	LR	UL	LR	
17	4	4	4	4	4	3	pass

Table A.6 – Human observer results for High Contrast Spatial Resolution test: altered axial series.

Participant Name	Axial Series Number	left-pair		center-pair		right-pair		PASS/FAIL
		UL	LR	UL	LR	UL	LR	
Danielle Miles	15	4	4	4	4	3	3	pass
	16	4	4	3	3	3	3	fail
Lisa Demar	16	3	3	3	4	1	2	fail
	18	4	4	4	4	3	3	pass
Curtis Cox	18	4	4	4	4	3	4	pass
	19	4	4	4	4	4	4	pass
Ruth Cox	19	4	4	4	4	3	4	pass
	20	4	4	4	4	3	3	pass
Warren Liu	20	4	4	4	4	4	4	pass
	15	4	4	4	4	4	3	pass

Table A.7 – Automated procedure results for Slice Thickness Accuracy test

Axial Series Number	Slice Thickness (mm)	PASS/FAIL
5	4.938	pass
6	4.394	pass
7	4.661	pass
8	5.034	pass
9	4.734	pass
10	5.053	pass
11	4.996	pass
12	4.817	pass
13	1.342	fail
14	1.397	fail
15	7.59	fail
16	3.564	fail
17	5.011	pass
18	4.109	fail
19	4.913	pass
20	5.241	pass

Table A.8 – Human observer (author) results for Slice Thickness Accuracy test: base axial series.

Axial Series Number	signal ramp lengths (mm)		Slice Thickness (mm)	PASS/FAIL
	top	bottom		
17	47.652	49.436	4.853	pass

Table A.9 – Human observer results for Slice Thickness Accuracy test: altered axial series.

Participant Name	Axial Series Number	signal ramp lengths (mm)		Slice Thickness (mm)	PASS/FAIL
		top	bottom		
Danielle Miles	15	77.859	81.849	7.980	fail
	16	78.092	85.240	8.151	fail
Lisa Demar	16	36.909	33.700	3.523	fail
	18	40.823	43.493	4.212	pass
Curtis Cox	18	37.910	38.646	3.827	fail
	19	37.948	39.894	3.890	fail
Ruth Cox	19	46.573	47.778	4.717	pass
	20	54.948	50.937	5.287	pass
Warren Liu	20	52.138	50.483	5.130	pass
	15	73.857	78.306	7.602	fail

Table A.10 – Automated procedure results for Slice Position Accuracy test

Axial Series Number	Slice displacement (mm)		PASS/FAIL
	slice 1	slice 11	
5	-0.814	-2.441	pass
6	-0.814	-2.360	pass
7	-0.814	-2.441	pass
8	-0.488	-2.116	pass
9	-0.651	-2.116	pass
10	-0.651	-2.279	pass
11	-0.651	-2.441	pass
12	-0.651	-2.279	pass
13	-0.488	-2.116	pass
14	-0.488	-2.116	pass
15	-0.814	-2.604	fail
16	9.440	-12.126	fail
17	-0.651	-2.441	pass
18	-0.814	-2.441	pass
19	-0.814	-2.279	pass
20	-0.651	-3.092	fail

Table A.11 – Human observer (author) results for Slice Position Accuracy test: base axial series.

Axial Series Number	Bar-length difference (mm)		Slice displacement (mm)		PASS/FAIL
	slice 1	slice 11	slice 1	slice 11	
17	-1.608	-4.416	-0.804	-2.208	pass

Table A.12 – Human observer results for Slice Position Accuracy test: altered axial series.

Participant Name	Axial Series Number	Bar-length difference (mm)		Slice displacement (mm)		PASS/FAIL
		slice 1	slice 11	slice 1	slice 11	
Danielle Miles	15	-2.050	-4.652	-1.025	-2.326	pass
	16	18.798	-23.255	9.399	-11.628	fail
Lisa Demar	16	18.174	-23.659	9.087	-11.829	fail
	18	-1.713	-4.479	-0.857	-2.240	pass
Curtis Cox	18	-1.393	-4.502	-0.696	-2.251	pass
	19	-1.371	-4.385	-0.686	-2.193	pass
Ruth Cox	19	-2.243	-4.289	-1.122	-2.144	pass
	20	-1.815	-6.035	-0.907	-3.017	fail
Warren Liu	20	-1.751	-5.949	-0.875	-2.974	fail
	15	-2.092	-4.639	-1.046	-2.319	pass

Table A.13 – Automated procedure results for Image Intensity Uniformity test

Axial Series Number	PIU	PASS/FAIL
5	95.945	pass
6	92.830	pass
7	93.866	pass
8	87.544	fail
9	86.762	fail
10	97.233	pass
11	94.776	pass
12	94.669	pass
13	95.399	pass
14	94.648	pass
15	94.902	pass
16	94.860	pass
17	95.496	pass
18	83.854	fail
19	89.146	fail
20	91.070	pass

Table A.14 – Human observer (author) results for Image Intensity Uniformity test: base axial series.

Axial Series Number	Mean Intensity		PIU	PASS/FAIL
	low	high		
17	512.465	563.911	95.220	pass

Table A.15 – Human observer results for Image Intensity Uniformity test: altered axial series.

Participant Name	Axial Series Number	Mean Intensity		PIU	PASS/FAIL
		low	high		
Danielle Miles	15	495.396	558.050	94.053	pass
	16	482.901	564.960	92.169	pass
Lisa Demar	16	500.356	566.832	93.771	pass
	18	331.109	462.584	83.435	fail
Curtis Cox	18	331.347	447.911	85.042	fail
	19	421.228	541.792	87.481	fail
Ruth Cox	19	404.267	541.515	85.488	fail
	20	407.208	513.158	88.488	fail
Warren Liu	20	427.129	513.267	90.840	pass
	15	510.029	557.723	95.533	pass

Table A.16 – Automated procedure results for Percent Signal Ghost test

Axial Series Number	Ghosting Ratio	PASS/FAIL
5	0.000	pass
6	0.001	pass
7	0.000	pass
8	0.000	pass
9	0.001	pass
10	0.001	pass
11	0.000	pass
12	0.000	pass
13	0.004	pass
14	0.002	pass
15	0.001	pass
16	0.001	pass
17	0.000	pass
18	0.001	pass
19	0.001	pass
20	0.001	pass

Table A.17 – Human observer (author) results for Percent Signal Ghost test: base axial series.

Axial Series Number	Mean Intensity					Ghosting Ratio	PASS/FAIL
	large	top	bottom	left	right		
17	543.943	3.586	3.725	3.198	2.962	0.001	pass

Table A.18 – Human observer results for Percent Signal Ghost test: altered axial series.

Participant Name	Axial Series Number	Mean Intensity					Ghosting Ratio	PASS/FAIL
		large	top	bottom	left	right		
Danielle Miles	15	537.244	2.318	2.088	1.419	1.266	0.002	pass
	16	546.541	2.308	2.946	2.083	2.544	0.001	pass
Lisa Demar	16	546.374	2.865	3.181	2.671	3.199	0.000	pass
	18	378.247	3.784	3.835	3.588	3.912	0.000	pass
Curtis Cox	18	378.220	3.509	3.617	5.062	4.003	0.003	pass
	19	488.548	3.546	3.995	3.806	3.282	0.000	pass
Ruth Cox	19	488.579	4.807	5.762	4.403	4.346	0.002	pass
	20	466.297	3.777	3.925	3.516	2.614	0.002	pass
Warren Liu	20	466.239	3.630	3.852	3.323	2.899	0.001	pass
	15	537.793	2.456	2.518	2.241	1.853	0.001	pass

Table A.19 – Automated procedure results for Low Contrast Object Detectability test

Axial Series Number	Sum Spokes	PASS/FAIL
5	33	pass
6	30	pass
7	38	pass
8	32	pass
9	38	pass
10	35	pass
11	35	pass
12	37	pass
13	37	pass
14	35	pass
15	32	pass
16	16	pass
17	34	pass
18	36	pass
19	37	pass
20	32	pass

Table A.20 – Human observer (author) results for Low Contrast Object Detectability test: base axial series.

Axial Series Number	Spoke Count				Sum Spokes	PASS/FAIL
	slice 11	slice 10	slice 9	slice 8		
17	10	10	9	5	34	pass

Table A.21 – Human observer results for Low Contrast Object Detectability test: altered axial series.

Participant Name	Axial Series Number	Spoke Count				Sum Spokes	PASS/FAIL
		slice 11	slice 10	slice 9	slice 8		
Danielle Miles	15	10	10	9	4	33	pass
	16	10	8	0	0	18	pass
Lisa Demar	16	9	9	0	0	18	pass
	18	10	9	8	1	28	pass
Curtis Cox	18	10	9	8	2	29	pass
	19	9	9	8	2	28	pass
Ruth Cox	19	10	9	9	2	30	pass
	20	10	10	9	0	29	pass
Warren Liu	20	10	10	7	2	29	pass
	15	10	10	10	8	38	pass

Appendix B: Automated MRI Quality Assurance Procedure User-Manual

Written: May 18, 2005
Last revised: May 18, 2005

Introduction

This document is a user-guide for the automated magnetic resonance imaging (MRI) quality assurance (QA) procedure developed by Atiba Fitzpatrick and Dr. Chris L. Wyatt. This document will briefly describe the program flow as well as the most important particulars of all components of the code. First, an overview of the procedure will be provided. Then, details about individual components will be provided. Then, possible user actions will be described. Finally, limitations and other areas of concern will be discussed.

Overview

This QA procedure involves the 7 tests of the Phantom Assessment portion of the American College of Radiology (ACR) accreditation program. The program requires a human observer to carry out the various tests using an assortment of tools for display intensity window and level adjustment and on-screen length and mean intensity measurements. However, the procedure being described in this user-manual is fully automated. The main steps of the automated procedure are as follows:

1. Read axial and sagittal series
2. Perform image registration
3. Perform Geometric Accuracy Test
4. Perform High Contrast Spatial Resolution Test
5. Perform Slice Thickness Accuracy Test
6. Perform Slice Position Accuracy Test
7. Perform Image Intensity Uniformity Test
8. Perform Percent Signal Ghosting Test
9. Perform Low Contrast Object Detectability Test
10. Log results

Steps 3 through 9 are the automated ACR tests.

The steps would be executed in this order if the full procedure was being executed. However, the tests are independent of each other and can be executed individually. Still, image registration is generally necessary for the success of the tests unless great care was taken in positioning the phantom prior to scanning.

Component Descriptions

The program flow and particulars will now be described in more detail. The motion estimation algorithm of the image registration procedure is written in C++. All other code is written in MATLAB. The following descriptions will be will be in the order of the general procedure flow stated above. The related files of each step will be listed, but sub-functions will not be.

Main Function

Primary related files: RunFullProcedure.m, RunFullProcedure.mat

When executing the full automated procedure, a main function is used that controls the program flow. This function, named RunFullProcedure.m, calls the various functions associated with the procedure steps listed above. The test functions return their associated measurements. The measurements are analyzed in the main function to determine if each test is passed. The results are then logged in various text files (to be discussed later) as well as saved in a file named RunFullProcedure.mat. The criteria for the tests are defined in this main function, and can be altered at the user's discretion. This function requires the '.img' filenames for the axial and sagittal series as the only inputs.

Image Registration

Primary related files: findMotionVector.m, TransformImage.m, CalculateDiff2.m FixCalNewImage.cpp, FixComputeLoop.cpp, FixCal_gT.cpp, FixComputeLoop.mexglx, acrslices.mat, acr_transformed.mat

The main file of this section named findMotionVector.m first uses various files to read the .img files that contain the candidate series data. The files used to read the data begin with the letters 'wfu'. A description of these files will not be provided here as it is not pertinent. The findMotionVector.m file then loads the standard series data stored in the file acrslices.mat. This acrslices.mat file contains an array named acrslices that contains 11 axial slices in ascending order with a sagittal localizer slice appended at the end. This data set is the ground truth used for registering the candidate series. Files named CalculateDiff2.m FixCalNewImage.cpp,

FixComputeLoop.cpp, FixCal_gT.cpp, and FixComputeLoop.mexglx are used for motion estimation. The motion referred to here is simply the translational and rotational differences between the candidate and standard data sets. The file named TransformImage.m uses the estimated motion vector to perform the rigid transformation of the candidate series. The registered candidate data set is then saved in a file named acr_transformed.mat, which is subsequently loaded by the test functions. The acr_transformed.mat file contains an array named acr_transformed that contains the 11 axial slices in ascending order with the localizer appended at the end. This file also stores the name of the related '.img' files. The findMotionVector.m file accepts as its only inputs the '.img' file names for the sagittal and axial series to be registered.

Geometric Accuracy Test

Primary related files: geometric.m, acr_transformed.mat

The measurement process for the Geometric Accuracy Test is executed with the file named geometric.m. This file loads the file named acr_transformed.mat that contains the registered candidate series. The phantom length and diameter measurements are returned to the calling function for analysis.

High Contrast Spatial Resolution Test

Primary related files: spatialResolution.m, TransformImage.m, CalculateDiff2.m FixCalNewImage.cpp, FixComputeLoop.cpp, FixCal_gT.cpp, FixComputeLoop.mexglx, acr_transformed.mat

The file named spatialResolution.m performs the High Contrast Spatial Resolution Test. Even though the candidate series would have already been registered, remaining small translational differences between the registered candidate (from acr_transformed.mat) and standard series require subsequent registration of resized small regions of interest (ROIs). These ROIs contain the hole-pair arrays assessed in the test, which are scaled sufficiently for optimal registration. The files named TransformImage.m, CalculateDiff2.m FixCalNewImage.cpp, FixComputeLoop.cpp, FixCal_gT.cpp, FixComputeLoop.mexglx are used for image registration as described above. The file named spatialResolution returns the hole-counts for each

row/column of the hole-pair arrays as well as the resolution decision for each hole-size to the calling function.

Slice Thickness Accuracy Test

Primary related files: thickness_accuracy.m, acr_transformed.mat

The file named thickness_accuracy.m first loads the file named acr_transformed.mat that contains the registered candidate series. The measured lengths of the signal ramps are returned to the calling function for analysis.

Slice Position Accuracy Test

Primary related files: position_accuracy.m, acr_transformed.mat

The file named position_accuracy.m performs the necessary measurements for the Slice Position Accuracy Test. The file named acr_transformed.mat, which contains the registered candidate series must first be loaded. The measured bar lengths are returned to the calling function for further analysis.

Perform Image Intensity Uniformity Test

Primary related files: imageUniformity.m, acr_transformed.mat

The file named imageUniformity.m first loads the file named acr_transformed.mat that contains the registered candidate series. The average maximum and minimum intensities in the phantom region of slice 7 are determined and returned to the calling function for analysis.

Perform Percent Signal Ghosting Test

Primary related files: percentGhost.m, acr_transformed.mat

The file named percentGhost.m first loads the file named acr_transformed.mat that contains the registered candidate series. The intensity-means of the associated ROIs are determined and

returned to the calling function for analysis. The series signal-to-noise ratio (SNR) is calculated from the large ROI and also returned to the calling function.

Low Contrast Object Detectability Test

Primary related files: objectDetectability.m, acr_transformed.mat

The file named objectDetectability.m first loads the file named acr_transformed.mat that contains the registered candidate series. This objectDetectability.m file counts the number of complete spokes in the series. The function returns the total spoke count and an array depicting which spokes were determined to be detectable.

Logging Results

Primary related files: LogAllData.m, LogIndividualTests.m, ACR_Results.txt, Geometric_Sagittal_Results.txt, Geometric_Axial_Results.txt, Resolution_Results.txt, Thickness_Results.txt, Position_Results.txt, Uniformity_Results.txt, Percent_Ghost_Results.txt, Object_Detect_Results.txt, SNR_Results.txt, RunFullProcedure.mat

The main function (RunFullProcedure.m) described above calls two functions to report procedure results to various text files. Data is not passed from the main function to the reporting functions. Instead, the file named RunFullProcedure.mat, which the data from the main function is stored in, is loaded by the reporting functions. The function named LogAllData.m is called to report data from all tests to a text file called ACR_Results.txt. If this text file already exists, the data will simply be appended at the end of the file. Otherwise, the text file will be created. This report shows the date and time that the procedure ended, the filenames of the series considered, as well as other relevant information. Of primary interest to the user, the report shows the results, measurements, criteria used, and the difference of measurements from criteria for all tests. This report also includes brief descriptions of each test for user not very familiar with the ACR program. The SNR measurement is also reported as can be seen in Figure B.1. This report should be used when the user wants to be able to view data from all tests at once.

The function named LogIndividualTests.m reports the data for each test as well as the SNR measurement to separate text files, namely, Geometric_Sagittal_Results.txt, Geometric_Axial_Results.txt, Resolution_Results.txt, Thickness_Results.txt, Position_Results.txt, Uniformity_Results.txt, Percent_Ghost_Results.txt, Object_Detect_Results.txt, and SNR_Results.txt. Notice that each test has a corresponding text file. The exception being that the Geometric Accuracy Test results are reported to 2 separate files with one corresponding to the sagittal measurement, and the other corresponding to the axial measurements. This distinction was made because the sagittal and axial series come from different '.img' files. These reports should be used when the user wants to perform longitudinal analysis of one test or the SNR measurement. The results for each execution will be stored in a row along with the date, time and series filename information. Thus, the user can simply scan down a results column to see how the measurement varies for different series. Each time the procedure is executed, the new data is appended to the end of the existing text files. An example of one of these files is shown in Figure B.2.

Date: 19-May-2005 Time: 15:06
axial filename: WFUQA-0009-001-0011-memp sagittal filename: WFUQA-0004-001-0001-memp
Localizer offset angle(degrees): 0.330 translation(mm): x:0.160 y:-2.738
Axial slices offset angle(degrees): 0.912 translation(mm): x:0.452 y:2.793

TEST	RESULT	MEASUREMENT(mm)	CRITERIA(mm)	DISCREPANCY
1 Geometric	fail	loc: 148.112	148.000	0.112
		s1: vert:171.549 hor:180.664	190.000	vert:18.451 hor:9.336
		s5: vert:171.549 hor:180.990	190.000	vert:18.451 hor:9.010
		s5: 45d:175.456 -45d:176.107	190.000	45_d:14.544 - 45_d:13.893
2 Resolution	fail	left-pair: UL:4 LR:3	4	UL:0 LR:1
		center-pair: UL:4 LR:3	4	UL:0 LR:1
		right-pair: UL:3 LR:4	4	UL:1 LR:0
3 Thickness	fail	thickness: 2.701	5.000	2.299
4 Position	pass	displace: s1:-0.326 s11:0.732	2.500	s1:2.826 s11:1.768
5 Uniformity	fail	PIU: 81.576	90.000	8.424
6 Percent Ghost	pass	ratio: 0.001	0.025	0.024
7 Object Detect	pass	num spokes: 38	9	29
TOTAL	fail			
SNR		18.497		

Notes:
1. Geometric Accuracy Test measures phantom length with localizer and diameter with axial slices 1 and 5. Measurements should vary from criteria by 2 mm or less.
2. Spatial Resolution Test determines lowest resolution for axial series by detecting holes in slice 1. Num holes detected should be 4 for both UL and LR of left and center pairs.
3. Thickness Accuracy Test calculates slice thickness by measuring lengths of signal ramps in slice 1. Sice Thickness should vary from criteria by 0.7 mm or less.
4. Position Accuracy Test determines slice displacement by measuring lengths of vertical bars in slices 1 and 11. Displacement should be less than or equal to criteria.
5. Percent Image Uniformity Test determines image uniformity in slice 7. PIU should be greater than or equal to criteria.
6. Percent Signal Ghosting Test determines amount of signal ghosting present in slice 7. Ghost ratio should be less than or equal to criteria.
7. Object Detectability counts the number of complete spokes detectable in slices 8 - 11. Total number of spokes should be greater than or equal to criterial.
SNR is determined from large ROI of slice 7. SNR = mean(ROI) / std(ROI).

Figure B.1 – An example of the file named ACR_Results.txt used to log the results for all tests. The result, measurement, criteria used, and difference between criteria and measurement are reported for each test. The SNR measurement is also reported.

Axial Filename	Date	Time	Result	Thickness (mm)
WFUQA-0008-001-0011-memp	20-May-2005	15:01	pass	5.115
WFUQA-0009-001-0011-memp	20-May-2005	15:05	pass	4.775
WFUQA-0010-001-0011-memp	20-May-2005	15:09	pass	5.126
WFUQA-0011-001-0011-memp	20-May-2005	15:13	pass	5.094
WFUQA-0012-001-0011-memp	20-May-2005	15:17	pass	4.915
WFUQA-0013-001-0011-memp	20-May-2005	15:21	fail	0.000
WFUQA-0014-001-0011-memp	20-May-2005	15:25	fail	1.422
WFUQA-0015-001-0011-memp	20-May-2005	15:29	fail	7.705
WFUQA-0016-001-0011-memp	20-May-2005	15:33	fail	3.638
WFUQA-0017-001-0011-memp	20-May-2005	15:37	pass	5.076

Figure B.2 – An example of the file named Thickness_Results.txt that is used to log the results of the Slice Thickness Accuracy Test. The associated result, and measurement for each axial '.img' file is recorded along with the date and time the procedure ended.

User Actions

To execute the full procedure, the user must simply call the main function named RunFullProcedure.m while specifying the '.img' filenames of the series to be considered. At any time after image registration has been performed, the user can execute individual tests since the registered candidate data is stored in the file named acr_transformed.mat. However, with the exception of the High Contrast Spatial Resolution Test, the necessary analysis of measurements will have to be performed in the calling function. These analysis formulas and related criteria can be simply copied from the main function. If the user decides to forgo the image registration procedure, the candidate data must still be saved to the file named acr_transformed.mat in the format described above.

Procedure Limitations

Since rigid transformation is performed by this procedure, with the exception of the Geometric Accuracy Test, the test results are affected by substantial geometric distortion. If the phantom region of the slices is sheared or scaled, the objects measure by certain tests may be distorted or not be located where expected. This may cause an automated test to incorrectly fail. The tests that are most affected by substantial geometric distortion are the High Contrast Spatial Resolution Test and Slice Thickness Accuracy Test. It is important to note, however, that the geometric distortion must be substantial to affect other tests. In most instances, the Geometric Accuracy Test can be failed without the accuracy of the other tests being compromised.

Practical Issues

The full procedure takes approximately 4 minutes to complete per sagittal/axial series pair. However, almost 3/4 of that time is spent on image registration (2.8 minutes). Thus, there is very little benefit gained from skipping any tests. This QA procedure works for T1 and T2 series, without the user indicating the series type. The author makes this statement with some reservation because further testing is needed. However, procedure generality was the intention when the tests were being implemented. Lastly, a graphical user-interface (GUI) manual procedure tool was created to allow a user to perform the manual ACR accreditation program procedure. The tool allows the user to select a slice from a loaded series, adjust display level and window settings, resize a crop of the slice, measure lengths, and determine mean intensities in specified ROIs. This GUI tool can be used for verifying the results from the automated procedure. If more detail is desired, request a copy of the associated Master's Thesis from Dr. Chris Wyatt. Also, the previously listed functions include descriptive comments to aid in understanding the program flow.

Appendix C: ACR Accreditation Program Phantom Assessment Procedure

(abridged)
Written: May 25, 2005
Last revised: May 28, 2005

Introduction

This document is a summarized version of the Phantom Assessment portion of the American College of Radiology (ACR) Accreditation Program guidance manual to be used for verification of the Automated MRI Quality Assurance Procedure by third party investigators. Only the descriptions of the measurement steps have been adapted for this document. For information on how test results should be analyzed and possible causes of failure, the reader is advised to refer to the ACR Accreditation Program guidance manual. For brevity, many statements have been paraphrased from the guidance manual. All other statements have been taken directly from the manual. First, a brief overview of the accreditation program procedure will be provided. Then, the steps for each test will be listed along with descriptive figures.

Overview

The Phantom Assessment portion of the ACR accreditation program requires a human observer to carry out 7 tests on images produced by scanning a provided physical phantom. The tests involved are for:

11. Geometric Accuracy,
12. High Contrast Spatial Resolution,
13. Slice Thickness Accuracy,
14. Slice Position Accuracy,
15. Image Intensity Uniformity,
16. Percent Signal Ghosting, and
17. Low Contrast Object Detectability.

A graphical user interface (GUI) tool, which is shown in Figure C.1, was created to allow a user to perform all measurements necessary for the accreditation program. This tool allows the user to select one of the 11 axial slices or the localizer slice of the series being considered. The user can adjust the display level and window for optimal visualization. The display intensity range is shown to the right of the displayed slice. The user can magnify a specified region using bicubic interpolation. Circular and elliptical regions of interest (ROIs) can be specified for the GUI to return the associated mean intensity. Finally, lengths can be measured which will be relative to the image field of view (FOV). The default FOV is 250x250 mm.

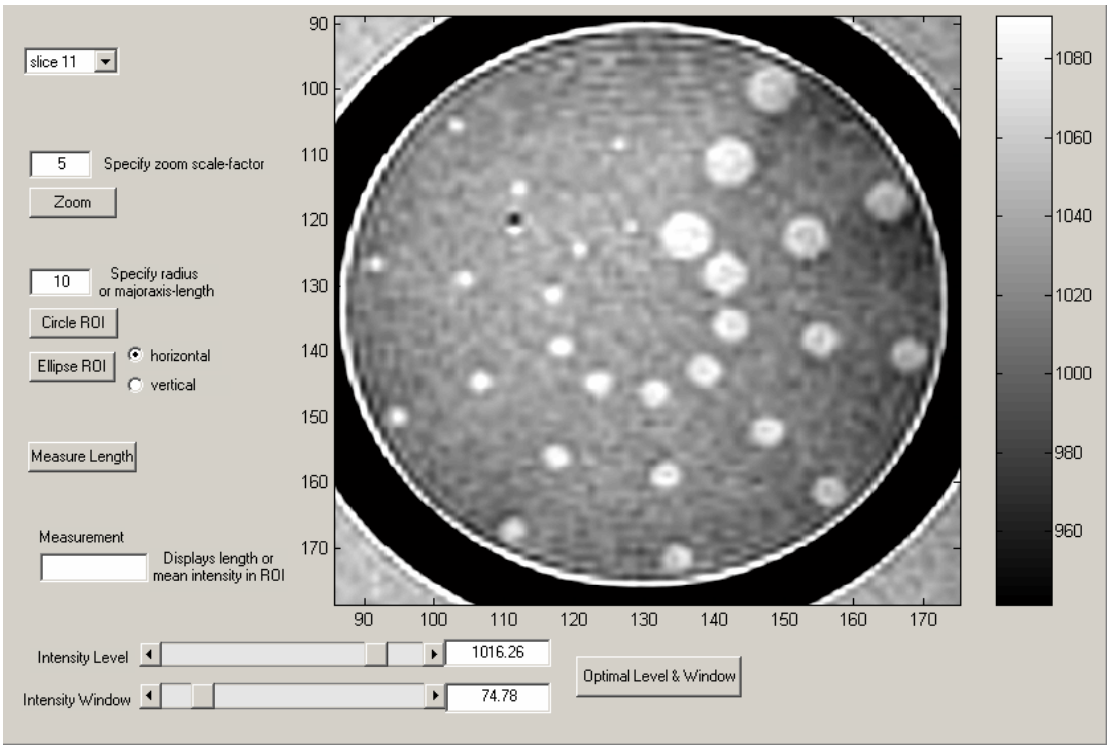


Figure C.1 – GUI observer tool. This GUI allows user to select a slice from a loaded series, resize a specified region using bicubic interpolation, adjust the display intensity window and level, measure lengths, and specify circular or elliptical ROIs for which the intensity means are determined.

1. Geometric Accuracy

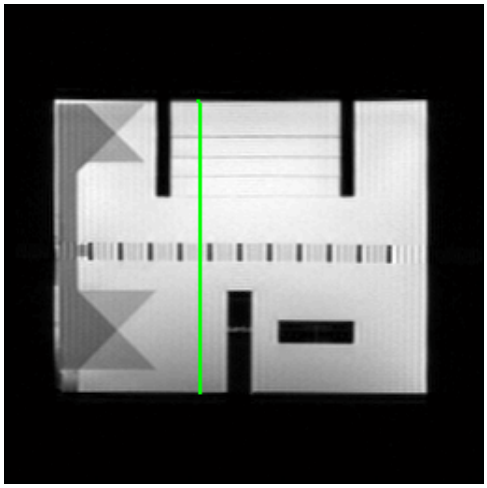
1. Select localizer slice, and adjust display level and window as shown below. Measure end-to-end length of the phantom (in top-to-bottom) direction near center (left-to-right) of phantom (see Figure C.2 (a)).
2. Select slice 1, and adjust display level and window as shown below. Measure the phantom diameter in the top-to-bottom and left-to-right-direction (see Figure C.2 (b)).
3. Select slice 5, and adjust display level and window as shown below. Measure the phantom diameter in the top-to-bottom, left-to-right-direction, and along the 45 and -45 degree diagonals (see Figure C.2 (c)).

Display level and window adjustment instructions:

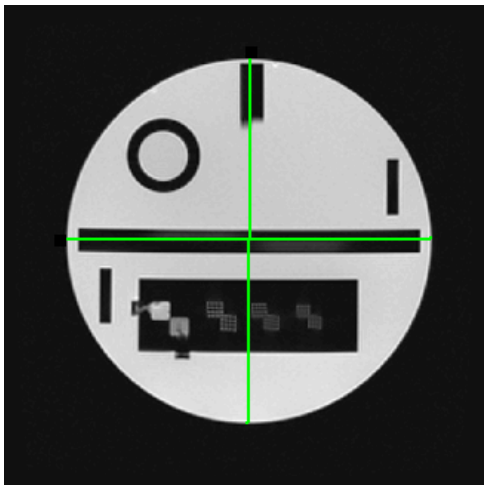
i. Set the display window to its minimum.
ii. Lower display level until all solution only regions are white.

iii. Raise the display level until about half of the water only regions turn black. Note this value.
iv. Set display level to half of the value from (iii).

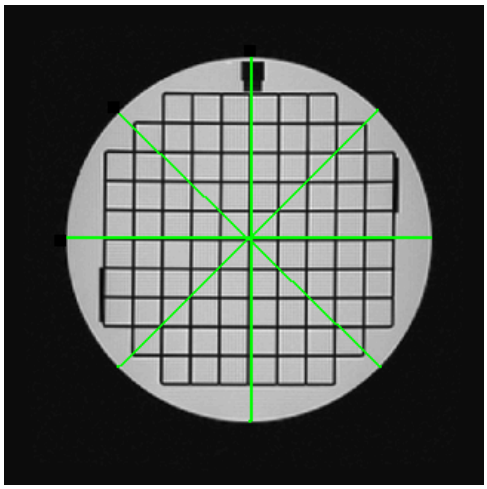
The true end-to-end length of the phantom is 148 mm. The true diameter of the phantom is 190 mm. The measurements should be within +/- 2 mm of the true values for this test to be passed.



(a)



(b)



(c)

Figure C.2 – (a) Illustration for localizer assessment. (b) Illustration for slice 1 assessment. (c) Illustration for slice 5 assessment.

2. High Contrast Spatial Resolution

1. Select slice 1.

2. Crop image region containing resolution insert and magnify (using zoom tool of GUI) by a factor of between 2 and 4 (see Figure C.3).
3. Begin with left pair. Look at the upper left (UL) array and adjust the display window and level to best show the holes in each row as being distinct from one another (see Figure C.4).
4. Count the max number of holes discernible in any row.
5. Now look at the lower right (LR) array and adjust the display window and level to best show the holes in each column as being distinct from one another.
6. Count the max number of holes discernible in any column.
7. Repeat steps 3 through 6 for the center and right pairs.

For this test to be passed, 4 holes should be detected for both the UL and LR arrays of the left and center pairs.

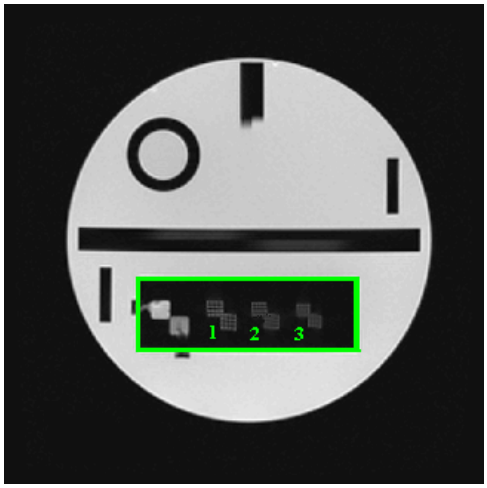
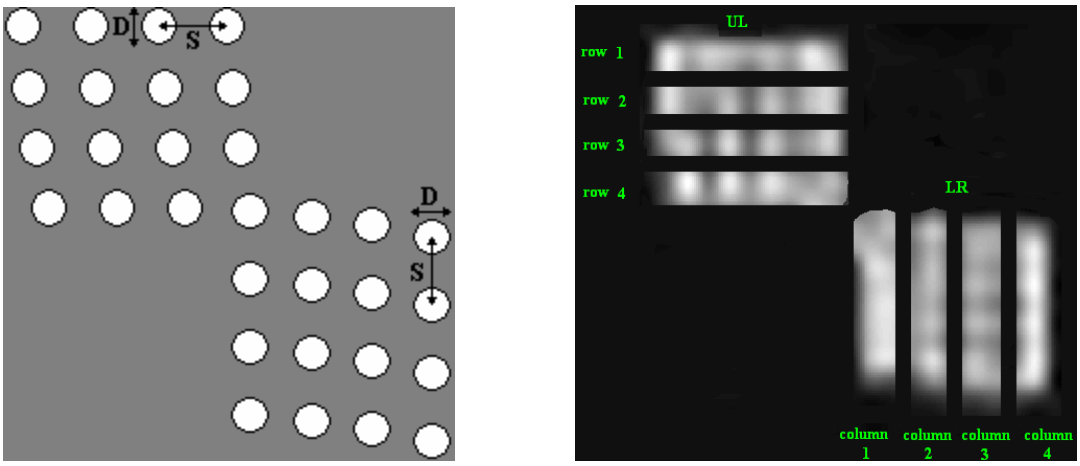


Figure C.3 – Slice 1 with resolution insert enclosed in green box. Three hole pair arrays are considered in this test, namely, 1) left-pair, 2) center-pair, and 3) right-pair.



(a) (b)

Figure C.4 – (a) Illustration of hole-array pair. The center-to-center spacing between holes (S) is twice the hole diameter (D). (b) Hole array pair for which rows and columns to be processed have been separated.

3. Slice Thickness Accuracy

1. Select slice 1, and magnify a region containing the slice thickness insert by a factor of 2 to 4 (see Figure C.5).
2. Adjust the display level and window such that the signal ramps are well visualized. Both the level and window values should be relatively low.
3. Use the elliptical ROI tool to determine the average intensity of the ramps.
4. Lower the display level to half the average intensity.
5. Measure the lengths of the ramps (in left-to-right direction).

There are 2 ramps denoted top and bottom. The measured lengths are used in the following formula to determine the slice thickness.

$$slice\ \ thickness = 0.2\frac{(top * bottom)}{(top + bottom)}$$

The slice thickness specified should have been 5.0 mm. The measured slice thickness should be within +/- 0.7 mm of this value for this test to be passed.

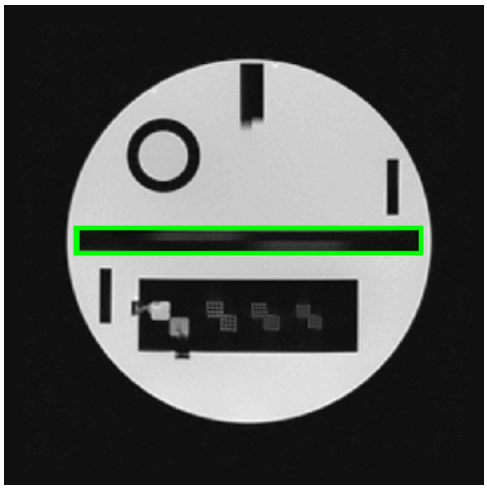
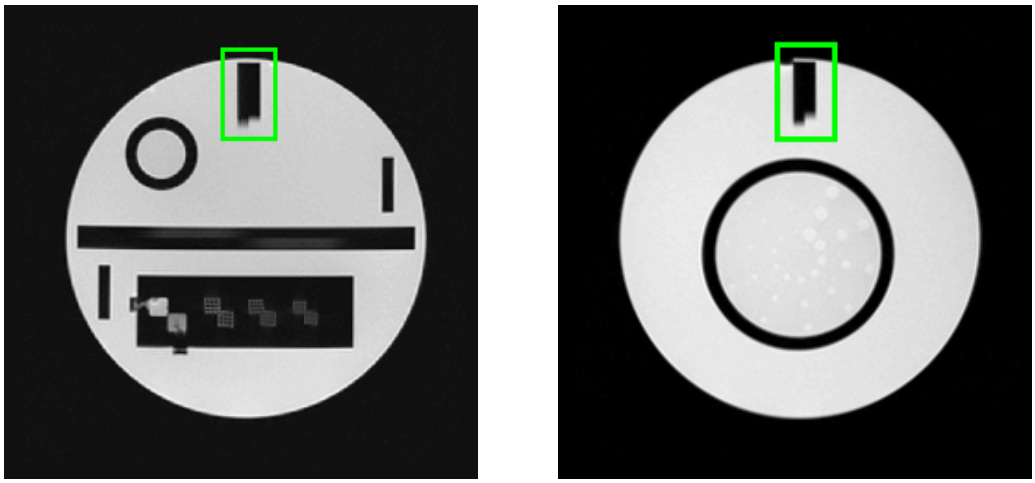


Figure C.5 – Slice 1 with thickness insert enclosed in green box

4. Slice Position Accuracy

1. Select slice 1. Magnify a region containing the vertical bars by a factor of between 2 to 4 (see Figure C.6).
2. Adjust the display level and window such that the ends of the bars are well defined.
3. Measure the difference between the top-to-bottom lengths of the bars. If the left bar is longer, assign a negative value to the length difference. Otherwise, the value should be positive.
4. Repeat the steps 1 though 3 for slice 11.

The absolute value of the difference should be less than or equal to 5 mm for the test to be passed. However, note that the true slice displacement is half the value of the length difference. Thus, the true slice displacement should be less than or equal to 2.5 mm.



(a) (b)

Figure C.6 – Vertical bars. (a) Slice 1 with the bars enclosed in a green box. (b) Slice 11 with the bars enclosed in a green box.

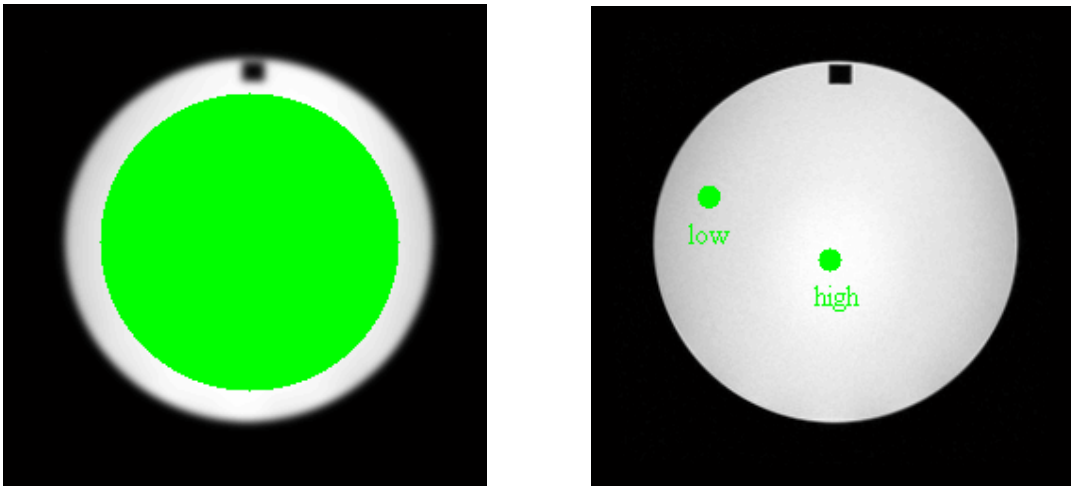
5. Image Intensity Uniformity

1. Select slice 7.
2. Center a large circular ROI with a radius between 79.8 and 82.7 mm about the phantom region of the slice (see Figure C.7 (a)).
3. Set the display window to its minimum, and lower the level until the entire area inside the large ROI is white. Now raise the level slowly until a small region of dark pixels of area approx. 1 cm^2 develops inside the ROI. This is the area if lowest signal in the large ROI.
4. Center a circular ROI of area 1 cm^2 (radius 5.6 mm) on this low signal region and record the mean intensity as the low signal value (see Figure C.7 (b)).
5. Now raise the level until the only a small 1 cm^2 area of white pixels remain in the image. This is the high signal region of the large ROI.
6. Center a circular ROI of area 1 cm^2 (radius 5.6 mm) on this high signal region and record the mean intensity as the high signal value.

The high and low signal values are used in the following formula to compute the percent integral uniformity (PIU).

$$PIU = 100\left(1-\left(\frac{high-low}{high+low}\right)\right)$$

For this test to be passed, the PIU should be greater than or equal to 90 percent.



(a) (b)

Figure C.7 – ROIs used for determining the max and min intensities in slice 7. (a) Slice 7 with ROI shown in green. (b) Small ROIs applied to slice 7.

6. Percent Signal Ghosting

1. Select slice 7.
2. Center a large circular ROI with a radius between 79.8 and 82.7 mm about the phantom region of the slice. Make sure that the ROI does not include any of the dark square at the top of the phantom. Record the mean intensity of this ROI (see Figure C.8).
3. Place elliptical ROIs of major axis length 65 mm along the top, bottom, left, and right sides of the phantom. The elliptical ROIs should be oriented such that only the background is inside. Record the mean intensity corresponding to each elliptical ROI.

Use the mean intensity values from the ROIs in the following formula to compute the ghosting ratio.

$$ghosting\ \ ratio = \left| \frac{(top + bottom) - (left + right)}{2 * large} \right|$$

For this test to be passed, the ghosing ratio should be less than or equal to 0.025.

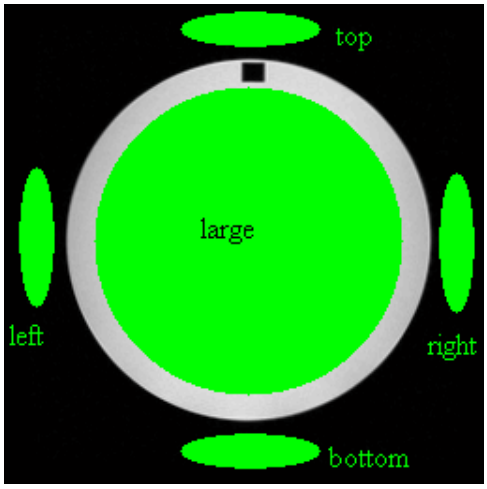


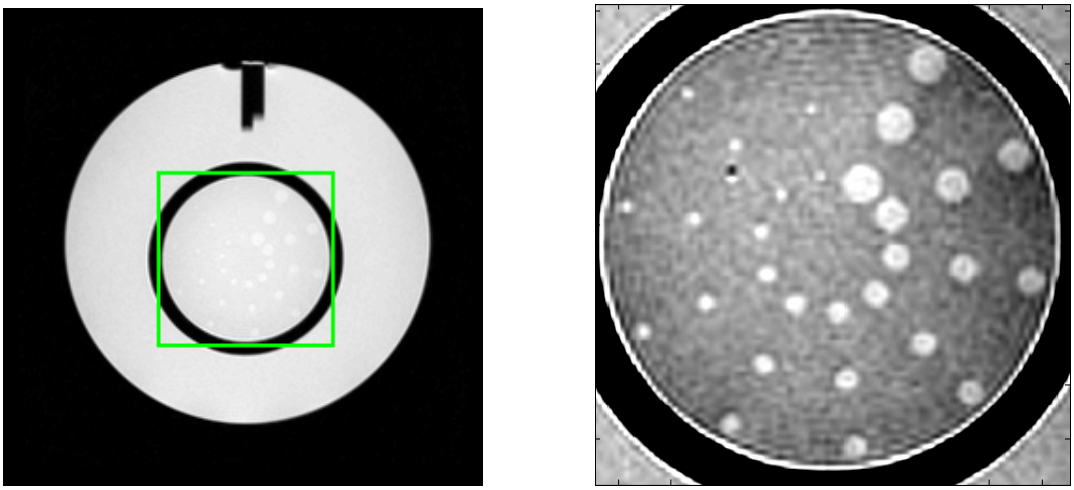
Figure C.8 – ROIs applied to the slice 7. The intensity-means in these ROIs are used to determine whether any signal ghosting is present.

7. Low Contrast Object Detectability

1. Select slice 11.
2. Adjust the display window and level for the best visibility of the disks. This will usually require a narrow window (see Figure C.9).

3. Now count the number of complete spokes beginning with the spoke containing the largest disks and moving clockwise. A spoke is complete only if all three of its disks are discernible for a single level setting. Stop counting spokes of a slice when the first incomplete spoke is found. Record the final count.
4. Repeat steps 2 and 3 for slices 10 through 8.
5. Sum the spoke count for slices 11 through 8.

The total spoke count for the series should be greater than or equal to 9 for this test to be passed.



(a) (b)

Figure C.9 – Contrast sheet for slice 11. (a) Green box encloses contrast sheet in slice 11. (b) Contrast sheet cropped and resize. The display intensity level and window were set to best show the spokes of holes.

Vita

Atiba Fitzpatrick

Atiba Fitzpatrick was born in Kingstown, St. Vincent and the Grenadines (SVG), but also spent much of his childhood living in the British Virgin Islands, Philadelphia, and Brooklyn. During his formative years, Fitzpatrick attended the Biabou Methodist School in St. Vincent, and his time in high school was divided between the SVG Grammar School, and the School of the Future in Manhattan. In December 2002, he received his BS in Electrical Engineering from the Rensselaer Polytechnic Institute, which happens to be the oldest technical school in the English speaking world (founded in 1824). This thesis is the partial fulfillment of requirements for an MS in Electrical Engineering from Virginia Tech. Fitzpatrick is also finishing with a Biomedical Engineering option.