THEORY AND APPLICATION OF OPTIMAL LINEAR RESOLUTION TO MRI TRUNCATION ARTIFACTS, MULTIEXPONENTIAL DECAYS AND *IN VIVO* MULTIPLE SCLEROSIS PATHOLOGY

By

Keith S Cover

B. Sc. (Physics) University of Waterloo, 1983

M. Sc. (Geophysics) University of British Columbia, 1986

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF PHYSICS AND ASTRONOMY

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

March 2001

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Physics and Astronomy
The University of British Columbia
6224 Agricultural Road
Vancouver, Canada
V6T 1Z1

Date: June 21, 2001

Abstract

It is widely believed that one of the best way to proceed when analysing data is to generate estimates which fit the data. However, when the relationship between the unknown model and data is linear for highly underdetermined systems, is it common practice to find estimates with good linear resolution with no regard for fitting the data. For example, windowed Fourier transforms produces estimates that have good linear resolution but do not fit the data. Surprisingly, many researchers do not seem to be explicitly aware of this fact. This thesis presents a theoretical basis for the linear resolution which demonstrates that, for a wide range of problems, algorithms which produce estimates with good linear resolution can be a more powerful and convenient way of presenting the information in the data, than models that fit the data.

Linear resolution was also applied to two outstanding problems in linear inverse theory. The first was the problem of truncation artifacts in magnetic resonance imaging (MRI). Truncation artifacts were heavily suppressed or eliminated by the choice of one of two novel Fourier transform windows. Complete elimination of truncation artifacts generally led to unexpectedly blurry images. Heavy suppression seemed to be the best compromise between truncation artifacts and blurriness.

The second problem was estimating the relaxation distribution of a multiexponential system from its decay curve. This is an example where hundreds of papers have been written on the subject, yet almost no one has made a substantial effort to apply linear resolution. I found the application to be very successful. As an example, the algorithm was applied to the decay of MRI data from multiple sclerosis patients in an attempt to differentiate between various pathologies.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

> "The Master said, Yu, shall I tell you what knowledge is? When you know a thing, to know that you know it, and when you do not know a thing, to recognize that you do not know it. That is knowledge."
>
> *Analects of Confucius (Waley's translation)* (Jeffreys 1973)

Summary

When trying to estimate a model from data, it is generally believed that finding a model which fits the data is the method of choice. However, for linear forward problems which are highly underdetermined, a very common practice is to find an estimate with good linear resolution, akin to optical resolution, with no regard to fitting the data. A prominent example is the windowed discrete Fourier transform. It is demonstrated that if the linear algorithm, which generates an estimate of the unknown model from the data, is invertible then the estimate, along with it noise statistics and linear resolution, is a complete summary of all the models that fit the data. The confusion between models that fit the data and linear resolution, which stretches back at least 50 years, seems to be due to the success of some algorithms being attributed to models that fit the data but where the good linear resolution is the important property. Wiener deconvolution is a prime example.

1.1 Introduction

Applications of the inversion of linear forward problems include the medical imaging modalities of MRI, computed tomography (CT), single photon spectroscopy (SPECT) and position emission tomography (PET) (Farquhar T H et al 1998). Other applications include (1) the measurement of frequency components of oscillating systems such as digital filter design using the discrete Fourier transform (DFT), and (2) multiexponential relaxation components of a decaying system found in MRI and downhole magnetic resonance used in the oil and gas industry. In each of these applications, the interest is often in estimating the unknown model, $m^U(y)$, over the full range of the variable y of equation 1.1. I will pursue the same goal.

If the inversion problem were purely a mathematical one, it would be simple to solve. The solution is just the subset of the model space which contains all models that fit the data. Listing the data, a measure of the noise and the corresponding data functions completely defines this set. However, the real problem is communicating an understanding of the contents of this subset to an interpreter – especially because the subspace contains an infinite number of models. Dealing with an infinite number of models which fit the data causes conceptual problems as well as computational ones. The conceptual problems may be complicated by the fact that, in some cases, finding even one model which fits the data is a major accomplishment. Generating one, or even many estimates which fit the data, is generally a straight forward, if tedious, task in the inversion of a linear forward problem. Conceptually, linear and nonlinear forward problems are often considered together, again directing a researcher away from the option of linear resolution, since it generally does not exist for nonlinear forward problems.

In the inversion of a linear forward problem, we are trying to learn what we can about an unknown model from measured data linearly related to the unknown model.

The linear transform from the unknown model, $m^U(y)$, to the data is called the linear forward problem. The unknown model is a member of the set of all possible models called model space. For the forward linear problem of interest, the model space is a linear vector space as defined by Parker (1994 p 3). For practical purposes, you can think of model space as containing every possible model.

The unknown model yields the N measured data, d_k, through the linear equation

$$d_k = \int_a^b m^U(y)\, g_k(y)\, dy + e_k \tag{1.1}$$

where the functions $g_k(y)$ are called the data functions and relate the unknown model to the data. The random variables, e_k, are the additive noise and are usually assumed to be independent of each other and stationary (Parker 1994 p 280). The constants a and b are the suitable limits on the integration. While equation 1.1 is only presented in one dimension, it can easily be generalized to higher order dimensions.

Common problems with linear forward problems are the 1D Fourier transform

$$d_k = \int_a^b m^U(f) e^{-2\pi i f t_k} df + e_k, \tag{1.2}$$

convolution

$$d_k = \int_a^b m^U(t) b(t - t_k) dt + e_k, \tag{1.3}$$

and a relative of the Laplace transform, the multiexponential transform

$$d_k = \int_0^\infty m^U(\tau) e^{-t_k/\tau} d\tau + e_k. \tag{1.4}$$

The measured data, d_k, is often thought of as list of numbers. However, for the list of numbers to have meaning it must have a context. The context is provided by the associated data function and noise statistics. Therefore, when I refer to the information provided by the data, I am assuming it is being interpreted in the context of the data functions and noise statistics.

Information other than that provided by the data about the unknown model is usually referred to as *a priori* information. In some problems, the available *a priori* information can be very helpful and should be taken advantage of. For example, if the unknown model is known to within two parameters, fitting to the two parameters is the best way to go. But often *a priori* information is nonexistent, suspect, or of a form which is difficult to integrate into an algorithm estimating the unknown model. In these cases, producing an estimate which does not rely on any *a priori* information can still be valuable and is the subject of this investigation.

At this point it is worth noting the subtle difference in the use of the terms model and estimate. The model always refers to an element of the model space and is a theoretical concept. Estimate refers to the output of an estimation algorithm. There is some overlap in the use of these terms.

There is a major dichotomy in the current application of linear inverse theory to many common problems. The commonly used algorithms for estimating unknown models fall roughly into two mutually exclusive classes, but there is some overlap. In the first type, the algorithms generate models that fit the data to within the noise. The second type contains estimates with good linear resolution. The first type is seen by many to be the "obvious" way to design an algorithm. The second type is widely used in applications including windowed Fourier transforms and most medical imaging modalities. Yet rarely in the literature is the dichotomy discussed in detail.

When most people hear the term resolution they think of how large an object can be seen with a microscope, a telescope or some other optical imaging modality. Figure 1.1 shows three images of the rings of a tree. Figure 1.1(a) has the highest image quality with 1.1(b) being blurrier and figure 1.1(c) being noisier. The relative resolution and noise of the three figures can easily be confirmed by inspection, but quantitative methods are also available. The resolution at various locations in an image can be measured by using

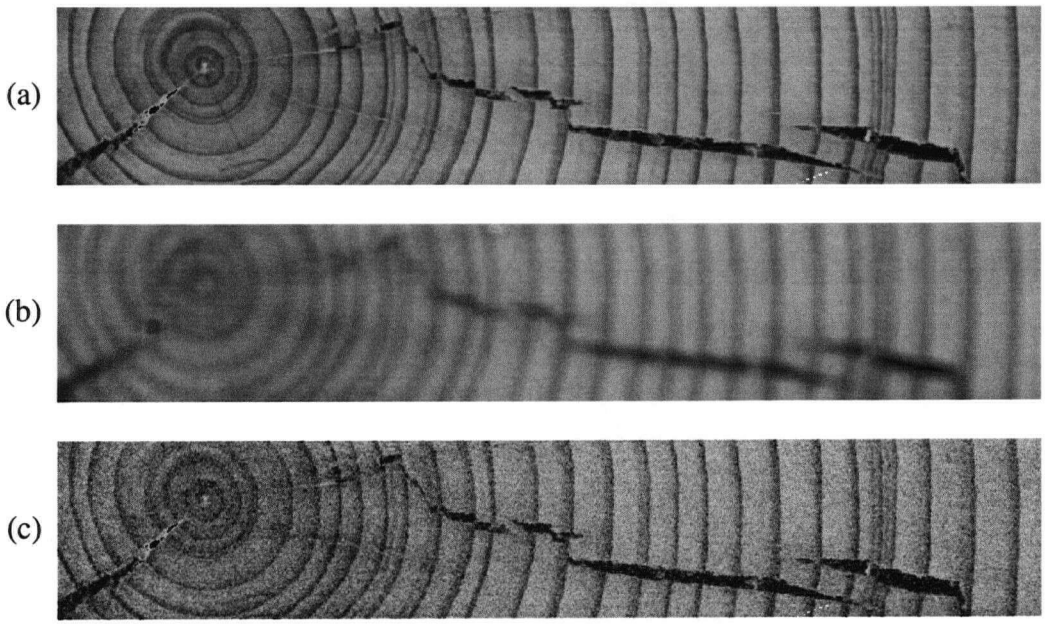

Figure 1.1: Three images of the same tree rings. The first (a) is the best quality image, (b) is blurrier and (c) is noisier.

the same imaging system to observe a variety of point sources. The noise can measured by imaging the tree rings repeatedly and comparing the images.

The term linear resolution, rather than just resolution, is used because, while optical systems are generally linear, some are nonlinear and the mathematical properties of the resolution of linear optics are quite different than those of nonlinear optics. The same is true for linear and nonlinear inversion algorithms as will be described below.

I believe the root cause of the dichotomy is that, at its heart, inversion of a linear forward problem is not just a mathematical problem but also a problem of communicating to an interpreter what is known and not known about a unknown model. An interpreter could be a radiologist reading the images produced by a magnetic resonance imaging (MRI) scanner, a geophysicist reviewing the results of seismic processing or an engineer studying the resonance modes of a bridge, just to name a few.

The goal of this paper is to resolve the dichotomy by providing a theoretical basis for

linear resolution in the context of linear inverse theory. In addition, I will demonstrate that linear resolution provides a much more powerful and efficient way to deal with the ambiguity due to the many models that fit the data in a linear inverse problem. I will also present a brief history of the dichotomy.

1.2 Models that fit the data

The most commonly recommended techniques to find estimates of the unknown model are based on the principle of "models that fit the data". (Menke 1984, Oldenburg 1984, Parker 1994, Sabatier 2000, Tarantola 1987, Twomey 1977). The basic principle of models that fit the data is to find a estimate that reproduces the data to within the noise. The most commonly used criterion to assess how well an model, $m^M(y)$, fits the data is the χ^2 criterion

$$\chi^2 = \sum_k \frac{(d_k - d_k^M)^2}{\sigma_k^2} \tag{1.5}$$

where σ_k is the standard deviation of the noise, e_k, in equation 1.1, and d_k^M are the data values predicted by a particular model, $m^M(y)$,

$$d_k^M = \int_a^b m^M(y)\, g_k(y)\, dy. \tag{1.6}$$

The value of χ^2 must be sufficiently small for the estimate to be considered to fit the data. It common practice to require χ^2 to be approximately equal to the number of data points since it is the most probable value for χ^2 assuming Gaussian noise (Menke, 1984). In equation 1.5 the noise of each data point is assumed to be uncorrelated from all the others. If the noise is correlated with a correlation matrix, C_{kl}^D, the equation for χ^2 is

$$\chi^2 = \sum_k \sum_l (d_k - d_k^M)(C_{kl}^D)^{-1}(d_l - d_l^M). \tag{1.7}$$

Equation 1.7 reduces to equation 1.5 when

$$C_{kl}^D = \sigma_k^2 \delta_{kl}. \tag{1.8}$$

where δ_{kl} is equal to one when $k = l$ and zero otherwise. For more details on χ^2 see Parker (1994 p 127) or any other of the above references.

Techniques which find models that fit the data have to choose among the infinite number of models that fit the data. A great many mathematical algorithms have been introduced in various attempts to deal with this nonuniqueness. However, as Twomey (1977 p vi) points out, while there have been many advances in mathematical techniques for solving the problem of estimating the unknown model, none of these techniques removes the fundamental ambiguity due to the infinitely many models that fit the data. He continues on to say that, in some instances, these procedures can effectively hide the ambiguity. In other words, models that fit the data are good at telling us what we do know – that the generated models could be the unknown model – but they are not very good at telling us what we don't know – all the models that also fit the data which we haven't generated.

There are a variety of views on how to use models that fit the data. I will attempt to summarize those views, but it is best to refer to the original references for a detailed explanation of these techniques.

There are two main schools of thought in inverse theory on how to deal with the nonuniqueness with models that fit the data, although these schools overlap. The first advocates presenting an interpreter with a wide selection of models that fit the data for each data set (Tarantola 1987 p 167, Oldenburg 1984 p 666). This solution is not practical for many problems. For example, in MRI, many sets of scans would have to be generated for each MRI scan of a patient. Each of these sets of scans would have to be examined by a radiologist and stored. Currently, the resources required to examine and store the standard one set of scans per patient is high. It would be totally impractical to considerably increase the resources allocated. The additional time it would take a radiologist to consider many scans for each patient would also be a major burden.

The second school of thought is to devise an algorithm which selects the "best" estimate which fits the data. This is usually done by introducing an ordering, from best to worst, of the set of possible estimates with a norm or other ordering criteria. The best estimate that fits the data to within the noise is then selected. A common method for ordering the estimates is the L_2 norm which is calculated by the equation

$$L_2 = \int_a^b m^U(y)^2 \, dy. \tag{1.9}$$

The estimate which fits the data to within the noise and has the smallest L_2 is considered the best estimate. As we shall see below, the smallest L_2 estimate, also has the property of linear resolution.

Other ordering criteria are maximum entropy (Tarantola 1987 p 151), maximin (Tarantola 1987 p351), smallest L_1 norm (Menke 1984 p 37), and nonnegative least squares (NNLS) (Lawson and Hanson 1974 p 160, Parker 1994 p 359), to name just a few. The question is which ordering to choose. Commonly, an interpreter will try a variety of these orderings on a few data sets and then select one of the criteria to use routinely. One of the risks in this approach is that, as Twomey suggested, the nonuniqueness of the solution may be hidden because only one solution is presented.

On occasion, a hybrid of the two schools of thought has been used where a variety of best estimates are presented to the interpreter.

An example of the prevalence of fitting the data is the attempts made to apply deconvolution to correct the blurry images originally produced by the Hubble Space Telescope (HST). A spherical aberration in the HST's main mirror caused a linear distortion in the images acquired during its first 4 years of operation. Since HST was carefully designed to be a highly linear device with good linear resolution, the forward problem was linear, although the actual linear resolution was less than optimum because of the flawed optics. A wide variety of deconvolution algorithms were applied to the

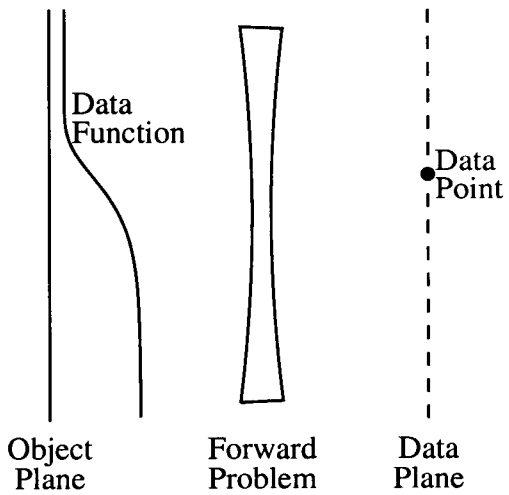

Figure 1.2: The forward problem illustrated as an optics problem. The data function, $g_k(y)$, corresponding to the data point, d_k, illustrates the poor linear resolution.

images to try to correct the distortions. While it turned out the blurriness was too unpredictable to be corrected, it is interesting to note that every algorithm presented at the major conference on the subject (Hanisch and White 1993, Adorf H 1995) assumed the estimate must fit the data. Although a few of the algorithms were linear, most were nonlinear .

Another example is an extensive review paper on estimating an unknown model when the forward problem is a Laplace transform (Istratov and Vyvenko 1999). Again, only algorithms which fit the data were considered.

1.3 Linear resolution

Linear resolution in inversion is mathematically equivalent to linear resolution in optics. This equivalence was appreciated at least as far back as 1952 and probably much earlier. Fellgett and Schmeidler (1952) applied Wiener deconvolution (1949) to sharpening the focus of the limb of the sun during a partial eclipse. Wiener's work involved trading off between resolution and noise in the deconvolution problem.

Figure 1.2 is an optical representation of equation 1.1 for a particular data function. The object plane represents the unknown model, $m^U(y)$, with the continuous variable y varying over the object plane. The data plane represents the measured data, d_k, and the discrete index of the data, k, varies over the data plane. The data function, $g_k(y)$, gives the weighting of the various locations on the image plane that sum to give the intensity at the data point. The data function in figure 1.2 clearly has poor linear resolution because the light arriving at the data point on the data plane is summed from a large part of the object plane. In optics, a second lens could be added to the right of the data plane to focus the image onto an image plane and improve the optical resolution. Likewise, in an inversion problem, an estimate would achieve better resolution than the forward problem by applying a linear operator to the data.

For an estimation algorithm to have linear resolution it must be linear. A detailed definition of linear is given by Parker (1990 p 44), but for the problems considered herein, any inversion algorithm which can be represented as a matrix is linear. This matrix, a_{ij}, will be called the estimation matrix. Therefore, an estimate of the unknown model can be generated by the equation

$$m_j^E = \sum_{k=1}^{N} a_{jk} d_k \tag{1.10}$$

and its covariance matrix, C_{kl}^E, is

$$C_{kl}^E = \sum_p \sum_q a_{kp} C_{pq}^D a_{lq}. \tag{1.11}$$

To precisely define linear resolution, Backus and Gilbert (1967, 1968, 1970) substituted equation 1.1 into 1.10 to get

$$m_j^E = \int_a^b m^U(y) \left[\sum_{k=1}^{N} a_{jk} g_k(y) \right] dy \tag{1.12}$$

and defined a new function, called the averaging or resolution function, to be

$$R_j(y) = \sum_{k=1}^{N} a_{jk} g_k(y). \tag{1.13}$$

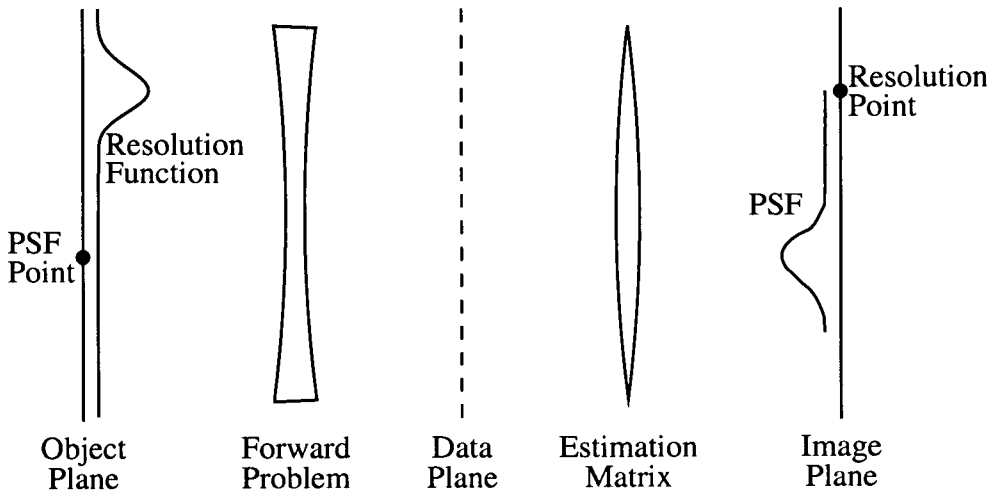

Figure 1.3: The estimation matrix when combined with the forward problem gives good linear resolution.

Proper choice of the estimation matrix, a_{jk}, can often lead to resolution functions which are well localized, or in other words well focused, about a point of interest in the object plane. The resolution function can be thought of as a synthetic data function. Likewise, the values of the estimate m_j^E can be though of as synthetic data. Equation 1.13 is a generalization of Wiener deconvolution.

Figure 1.3 shows the optics representation of the forward problem combined with the estimation matrix. The continuous variable y from equation 1.1 varies over the image plane. Although the image plane is a function of a continuous variable, we only calculate the image at a finite number of points which is usually sufficient to characterize the image plane. The resolution function on the object plane and its corresponding point in the image plane are also displayed. The resolution function replaces the data function of figure 1.2 which highlights the idea of the resolution function as a synthetic data function.

A common way of characterizing the performance of a lens set is to place a point source at a variety of locations in the object plane and measure the corresponding projections in the image plane. The projections are referred to as point spread functions (PSF's)

(Hecht and Zajac 1976). Since the optics are linear, any intensity distribution in the object plane can be decomposed into a series of point sources and each corresponding PSF in the image plane calculated. The PSF's can be added together to get the same image as a projection of the object would produce. Figure 1.3 shows a PSF point source in the object plane and the corresponding PSF in the image plane.

Once you have the complete set of PSF's, it is easy to calculate the resolution functions. The reverse is also true. Therefore, the set of the PSF's contains the same information as the set of the resolution functions. It can be easily shown that if all the resolution functions or all the PSF's are spatially invariant then the PSF's must equal the resolution functions.

As mentioned above, an important requirement for linear resolution is that the estimation algorithm is linear. This simple point seems to be over looked too often. Nonlinear estimation algorithms, which apply the data from linear forward problems of point sources, may give estimates with very narrow peaks. However, since linearity does not apply, the width of the peak has very little meaning when compared to the width of a peak from a linear estimation algorithm. Some estimation algorithms claiming "superresolution" are examples of this oversight.

1.4 Conservation of information

The great power of linear resolution, as will be shown below, is that an estimate, along with its noise statistics and resolution functions, can present all the information about the unknown model in the data – no more, no less. Thus, only one estimate presented to an interpreter gives the interpreter complete information about all of the models that fit the data, provided the interpreter has an understanding of the estimate resolution and noise. This is in contrast to estimates which generate models that fit the data which would have

to generate a great many estimates to supply the interpreter with the same information. In medical imaging for example, radiologists become familiar with the linear resolution and noise of various modalities during their basic training. Thus, examining one image with good linear resolution will provide the radiologists with all the information provided by the data. I believe this is why the common modalities in medical imaging including MRI, CT and SPECT, use linear resolution often with no regard to whether the estimates fit the data or not.

When presenting the information in the data about an unknown model to an interpreter, it is often desirable that the information be presented faithfully and not altered in any way. To assess whether an estimation algorithm accomplishes this goal, we need a mathematical formalism for presenting the information about the unknown model. For this purpose Backus and Gilbert used model space and Tarantola used model space with an additional probability density.

One can use equation 1.7 to assign a χ^2 to every model in the model space. This assignment is in line with Tarantola's (1987 p 1) approach of assigning a probability density to every model since the χ^2 can be converted directly to a probability density function on the model space. If we think about an estimate as synthetic data, the resolution functions as synthetic data functions, and the covariance matrix of the estimate as synthetic noise statistics, an estimate which conserves the information about the unknown model will assign exactly the same χ^2 to the model space as the data, as the following steps will show.

The χ^2 assigned to a model, $m^M(y)$, by a particular estimate, m_k^E, its noise and resolution functions, is given by

$$\chi^2 = \sum_k \sum_l (m_k^E - m_k^M) C_{kl}^{E^{-1}} (m_l^E - m_l^M). \tag{1.14}$$

where

$$m_k^M = \int_a^b m^M(y)\, R_k(y)\, dy. \tag{1.15}$$

Substituting equation 1.13 into 1.15 and 1.10 and then substituting the results, along with equation 1.11, into equation 1.14 yields

$$\chi^2 = \sum_k \sum_l (\sum_r a_{kr}\gamma_r)(\sum_s a_{ls}\gamma_s)\left(\sum_p \sum_q a_{kp} C_{pq}^D a_{lq}\right)^{-1} \tag{1.16}$$

where

$$\gamma_k = d_k - \int_a^b m^M(y)\, g_k(y)\, dy. \tag{1.17}$$

Rearranging the indices in equation 1.16 yields

$$\chi^2 = \sum_r \sum_s \gamma_r\gamma_s \sum_p \sum_q (\sum_k a_{kr} {a_{kp}}^{-1}) C_{pq}^{D^{-1}} (\sum_l a_{ls} {a_{lq}}^{-1}) \tag{1.18}$$

If the estimation matrix is invertible, then this equation reduces to

$$\chi^2 = \sum_r \sum_s \gamma_r\gamma_s C_{rs}^{D^{-1}} \tag{1.19}$$

and it follows that this equation is equal to 1.7. Therefore, if the estimation matrix is invertible, the estimate conveys exactly the same information about the unknown model as the data.

This result is not completely surprising, but I am not aware of it being previously published. Conservation of information can be a very desirable characteristic of an estimation matrix since it means all future calculations can be done with just the estimate, its resolution functions and covariance matrix. The measured data are no longer necessary.

In some cases perfect conservation of information may not be necessary and a small amount of loss may be acceptable. However, to determine how much information has been lost, we need a way to quantify the information. Shannon's measure of entropy offers

one way to accomplish this goal (Shannon 1948, Guiasu 1977). It should be possible to calculate the entropy due to the data, its data functions and noise statistics. The same measure could be applied to the estimate. If the values of entropy are different, then information has not been conserved. However, the values of entropy may be close enough that, for practical purposes, sufficient information can been conserved.

1.5 The Discrete Fourier Transform

The discrete Fourier transform (DFT) is perhaps the most common inversion of a linear forward problem. A vast number of algorithms are available in the literature for generating estimates of an unknown frequency spectrum. Thus the DFT provides an excellent case study of the possible and preferred approaches. The forward problem for the DFT is given in equation 1.2.

The most popular way to estimate the frequency spectrum is the discrete Fourier transform (DFT),

$$m^E(f) = \sum_{k=0}^{N-1} w_k d_k e^{2\pi i k f} \tag{1.20}$$

where

$$f = n/N \quad n = -N/2, ..., N/2 \tag{1.21}$$

and where w_k is a window, N is the number of data points and d_k are the data points to be transformed. The DFT is often implemented as a fast Fourier transform (FFT) because the FFT is much faster and less prone to round off error.

A wide variety of estimation algorithms are available which provide estimates which fit the data (Kay 1988), including the well known maximum entropy method. Many of these algorithms are nonlinear and thus, an interpreter must be very careful not to imbue the PSF's of a nonlinear algorithm with the properties of a linear algorithm. As mentioned earlier, it should also be kept in mind that each one of these estimates, when

interpreted as an estimate that fits the data, supplies information about only one of the infinitely many models that fit the data.

All linear estimation algorithms for the Fourier transform can be implemented by a windowed DFT because of the convolution theorem. Many papers have been written on various windows (Harris 1978). Their purpose is to adjust the linear resolution of the DFT. The variety of windows comes from the different preferences of what is optimal linear resolution.

A special property of the DFT inversion is that the resolution function is spatially invariant and thus is equal to the PSF at all points in the estimate. Another special property of the DFT inversion is that if $w_k = 1$ (no window is applied), then the estimate will fit the data and have linear resolution. However, when a window is applied, the estimate will no longer fit the data, but it will still have linear resolution. It is interesting to note that, to date, I have not located a reference which mentions this point. Also, from personal discussions with physicists and other people who work with data analysis and results, few people seem to be aware of this point and many are surprised when they realize it.

Figure 1.4 gives an example of the application of the Fourier transform. Figure 1.4(a) presents 128 complex data points from the model in figure 1.4(b). The model consists of purely real data with Dirac delta functions at $f = -0.3562$ and $f = 0.0375$ with amplitudes of 8.0 and -4.0 respectively. The box car has a height of 3 and ranges between $f = 0.0136$ and $f = 0.1975$. Gaussian noise with standard deviation of 0.001 has been added to the data. Dimensionless units are used for the frequency and amplitude to keep the example general.

Figure 1.4(c) is an estimate of the unknown model generated using equation 1.20. Figure 1.4(d) was generating using the Hamming window (Harris 1978, Lowe and

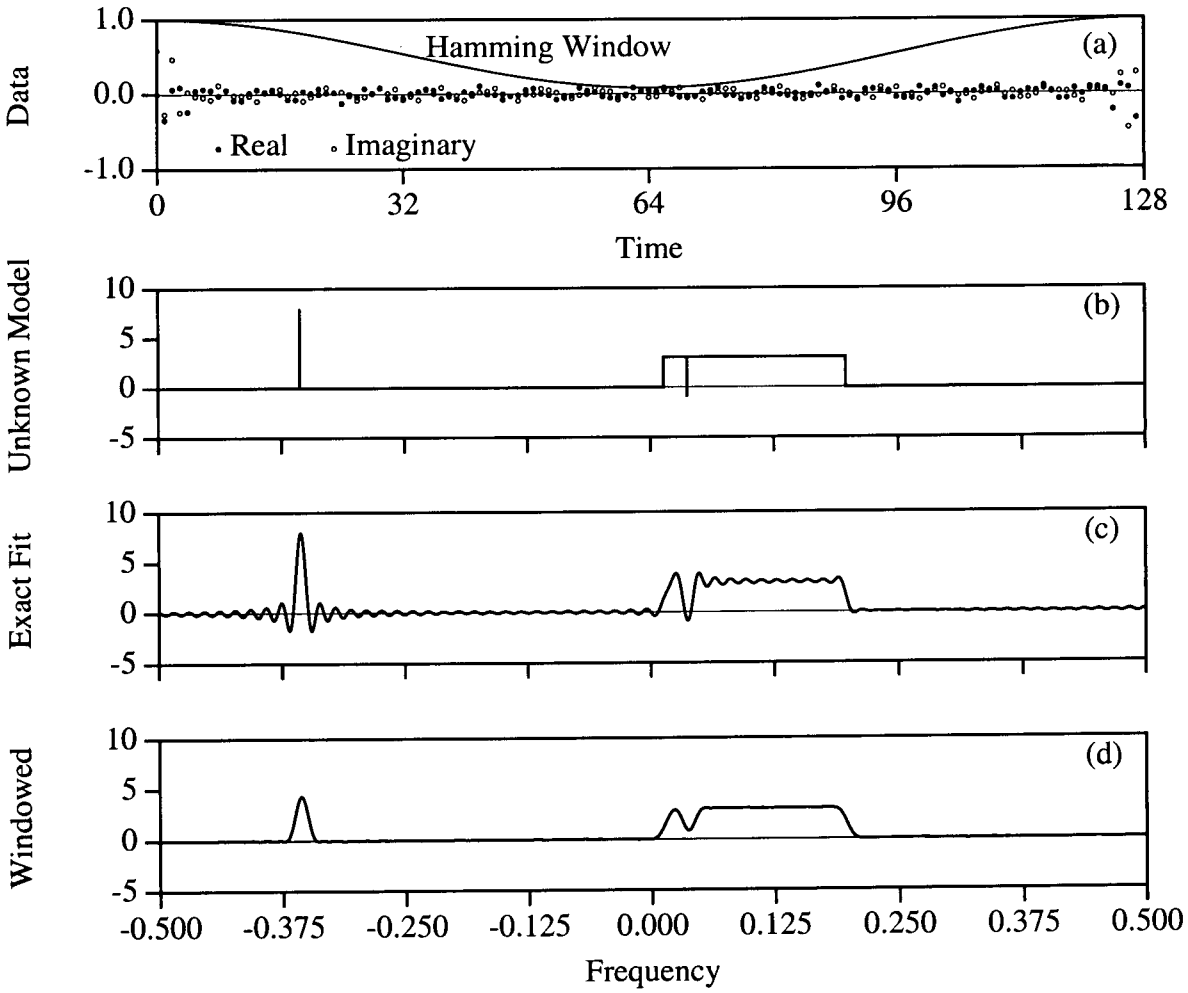

Figure 1.4: Example of discrete Fourier transforms. The exact fit has $\chi^2 = 0.0$ while the windowed data has $\chi^2 = 201118.9$.

Sorenson 1997) shown in figure 1.4(a).

$$H(v) = 0.54 + 0.46cos(2\pi\nu/\nu_{max}) \tag{1.22}$$

From the isolated peak at $f = -0.3562$ you can determine the the PSF. Since, for the Fourier transform, the PSF is spatially invariant, it is equal to the resolution function. The fit, as measured by χ^2, is given in figures 1.4(c) and 1.4(d).

Figure 1.4(c) is an exact fit while figure 1.4(d) has such a large χ^2 that it would not be considered a fit by any standards, but both have linear resolution. Both estimates are interpretable. Therefore, this suggests linear resolution is the important property in interpretation.

When first viewing figures 1.4(c) and (d), I suspect most interpreters use the isolated peak to the left of each spectrum to gauge the PSF and then examine the rest of the spectrum with the PSF in mind. This again suggests linear resolution is being used to interpret the estimate and not models that fit the data.

There is a test which I believe confirms linear resolution as the property of the estimate of which interpreters are taking advantage. Cover either the left or right half of the frequency spectrum in figures 1.4 with a piece of paper or your hand. The visible parts of the spectra are still easily interpretable. If the estimates were being interpreted as models that fit the data, the estimates would no longer be interpretable because we do not have an estimate that fits the data with just one half of the spectrum. But linear resolution only requires a section of a spectrum which is bigger than the PSF to yield useful information about that section of the spectrum.

The windowed Fourier transform is a good example of a linear estimation algorithm which conserves the information about the unknown. Provided the applied window is everywhere non zero then the equivalent estimation matrix in invertible. This is the case for the Hamming window. Since the DFT without a window is invertible, the windowed

form is also invertible. That is why no information is lost when further processing is performed in the frequency spectrum with no reference to the original data. For example, you can fit individual peaks of a spectrum without having to fit the whole data set.

The medical modality of MRI is a good example of an application which uses windowed Fourier transforms, but in 2D rather than one (Lowe and Sorenson 1997). MRI is reconstructed using a windowed 2D DFT. The purpose of windowing the data is to reduce the ringing in the resulting image, also known as truncation artifacts. This is akin to improving the focus of the image with no regard as to whether the model fits the data or not. Normally, the windows used in the MRI 2D DFT conserve information. Two images are produced for each data set, the real and imaginary. It is common practice to combine the two images into a magnitude image with the magnitude image being displayed to the interpreter. The phase information is only provided to the interpreter in special circumstances and for many clinical uses, the additional information it provides in not useful.

Examination of CT, SPECT and PET literature shows a similar situation to MRI. Filtering has been added to original algorithms to improve the linear resolution of the resulting images with no regard for the model fitting the data (King et al 1984, Gilland et at 1988, Farquhar 1998).

1.6 Estimates that fit the data and have linear resolution

How did the dichotomy between models that fit the data and linear resolution come about? It is impossible to say for sure, but careful examination of the literature yields some valuable clues which suggest good linear resolution was a fortunate byproduct of models that fit the data.

The earliest reference with the dichotomy that I have been able to find is Wiener's

deconvolution (1949). This publication is considered by many to be a central publication of deconvolution. In this paper, Wiener used the least squares fit to the data to derive deconvolution digital filters which had linear resolution. In this classic work, he gave no indication that he considered that linear resolution and fitting the data were two independent concepts. The DFT is another widely used early example. As mentioned above, without a window, it produces an estimate which both fits the data and has linear resolution.

There is also a method by Backus and Gilbert (1967) based on the Dirichlet criterion which produces estimation matrices which often have good resolution for the linear forward problem. Menke (1984 p 95) showed that several other methods for fitting the data will produce the same estimation matrices to within the noise handling properties. These fitting methods include the smallest L2 model, the least squares fit to the data and the maximum likelihood fit. Wiener's deconvolution and the DFT are examples of these types of algorithms. I suspect that these algorithms which produced estimates that fit the data and, inadvertently, also had good linear resolution, encouraged researchers to confuse the two independent concepts.

Only rarely in publications describing the application of linear inverse theory are estimates considered which have linear resolution but do not fit the data and the publication acknowledges that fact. Two such examples are the Fourier transform and deconvolution (Oldenburg 1976, 1981) in addition to others (Menke 1984).

Any mention of the dichotomy in the literature, as opposed to discussions of the two parts independently, is an uncommon occurrence. The issues are discussed by Tarantola (1987 p 461) and Menke (1984). Tarantola advocates models that fit the data using the method of generating many models that fit the data for each data set. He talks about Backus and Gilbert giving a "blurry" view of the unknown model. Tarantola states that if the estimate with linear resolution is too blurry, you should use models that fit the

data. This seems to be missing the point about linear resolution. A blurry estimate with good signal to noise means that the data do not tell us as much about the unknown model as we would like. The only way to get around this problem is to get more data with better signal to noise.

1.7 Conclusions

The dichotomy in estimating the unknown model for a linear forward problem has been around for at least the last fifty years. Careful analysis of the dichotomy in the context of linear inverse theory has shown that a single estimate, which was generated from the data by a transform which is representable by an invertible estimation matrix and has good linear resolution, has the very powerful ability to communicate to an interpreter a complete summary of all the models that fit the data. This is provided that the interpreter has an understanding of the noise and linear resolution of the estimate. Such an estimate is much more effective at communicating to an interpreter what is known or not known about an unknown model than one or many models that fit the data.

One important problem which has not yielded to optimal linear resolution is the inverse Laplace transform. Chapter 3 of this thesis discusses this problem in more detail and shows how to modify the forward problem so good optimal linear resolution can be achieved with an estimation matrix.

References

Adorf H 1995 Hubble Space Telescope image restoration in its fourth year *Inverse Problems* **11** 639–653

Backus G and Gilbert F 1967 Numerical application of a formalism for geophysical inverse problems *Geophys. J. R. Astr. Soc.* **13** 247–276

Backus G and Gilbert F 1968 The resolving power of gross earth data *Geophys. J. R. Astr. Soc.* **16** 169–205

Backus G and Gilbert F 1970 Uniqueness in the inversion of inaccurate gross Earth data *Phil. Trans. R. Soc. Lond. Ser. A* **266** 123–192

Farquhar T H, Chatziioannou A, Chinn G, Dahlbom M, Hoffman E J 1998 An investigation of filter choice for filtered back-projection reconstruction of PET *IEEE Trans Nucl Sci* **45** 1133–1137

Fellgett P B and Schmeidler F B 1952 On the sharpening of observational data with special application to the darkening of the solar limb *Roy. Astr. Soc. Notices* **112** 445–451

Gilland D R, Tsui B M W, McCartney W H, Perry J R, Berg J 1988 Determination of the optimum filter function for SPECT imaging *J Nucl Med* **29** 643–650

Guiasu S 1977 *Information theory with applications* (New York: McGraw-Hill)

Hanisch R J and White R L Eds 1993 The Restoration of HST Images and Spectra - II *Space Telescope Science Institute*

Harris F J 1978 On the use of windows for harmonic analysis with the discrete fourier transform *Proc. IEEE* **66** 51–83

Hecht E, Zajac A 1976 **Optics** (Don Mills, Ontario: Addison-Wesley)

Istratov A A, Vyvenko O F 1999 Exponential analysis in physical phenomena *Rev Sci Instr* **70** 1233-1257

Jeffreys H 1973 *Scientific Inference* (Cambridge: Cambridge)

Kay S M 1988 *Modern spectral estimation* (Englewood Cliffs, NJ: Prentice Hall)

King M A, Schwinger R B, Doherty P W, Penney B C 1984 Two-dimensional filtering of SPECT images using the METZ and Wiener Filters *J Nucl Med* **25** 1234–1240

Lawson C L and Hanson R J 1974 *Solving Least Squares Problems* (Englwood Cliffs, NJ: Prentice-Hall)

Lowe M J and Sorenson J A 1997 Spatially filtering functional magnetic resonance imaging data *Mag Res Med* **37** 723-729

Menke W 1984 *Geophysical Data Analysis: Discrete Inverse Theory* (New York: Academic Press)

Oldenburg D W 1976 Calculation of Fourier Transforms by the Backus-Gilbert Method *Geophys J R Astr Soc* **44** 413–431

Oldenburg D W 1981 A comprehensive solution to the linear deconvolution problem *Geophys J R Astr Soc* **65** 331–357

Oldenburg D W 1984 An introduction to linear inverse theory *IEEE Trans Geosci Remote Sens* **GE-22** 665-674

Parker R L 1994 *Geophysical inverse theory* Princeton: Princeton University Press

Parker J A 1990 *Image reconstruction in radiology* Boston: CRC Press

Sabatier P C 2000 Past and future of inverse problems *J Math Phys* **41** 4082–4124

Shannon CE 1948 A mathematical theory of communication *Bell Syst. Techn. J* **27** 379–423

Tarantola A 1987 *Inverse Problem Theory* (New York: Elsevier)

Treitel S and Robinson E A 1966 The design of High-Resolution Digital Filters *IEEE Trans Geosci Elec* **Ge-4** 25–38

Twomey S 1977 *Introduction to the Mathematics of Inversion of Remote Sensing and Indirect Data* (Amsterdam: Elsevier)

Wiener N 1949 *Extrapolation, Interpolation and Smoothing of Stationary Times Series* (New York: John Wiley & Sons)

Chapter 2

MRI Truncation Artifacts as a Linear Resolution Problem

Summary

A Fourier transform window was designed which heavily suppressed truncation artifacts in MR images, improved the signal to noise by 20 to 75% and substantially improved the interpretability of the resulting MR images compared to the commonly used Fermi window. This window should be particularly effective on large pixel images, such as the 32x32 pixel images used in chemical shift imaging (CSI) and functional MRI (fMRI). Much effort has gone into trying to reduce truncation artifacts in MR images as they can interfere with interpretation. By considering truncation artifacts as a problem in linear resolution two Fourier transform windows were designed – one which heavily suppressed them and a second which completely eliminated them. The complete removal of truncation artifacts did not result in the desired improvement in the interpretability of MR images. It appears that a minimal level of truncation artifacts have the benefit of making edges appear sharper analogous to an edge enhancement technique. The sharper edges allowed structures to be picked out more easily.

2.1 Introduction

The Fourier transform of a segment of a sinusoidal wave does not transform to a Dirac delta function but rather to a sinc function which includes sidelobes. These sidelobes are referred to, among other names, as truncation artifacts. Truncation artifacts typically are noticed on an MR image as ringing adjacent to sharp edges although they can have less

obvious manifestations. In most clinical applications of MRI the truncation artifacts are not a serious problem but in specific parts of the body, in particular the spine, truncation artifacts can easily be mistaken for pathology. In addition, small features anywhere in the body, which approach the resolution of a MR image, can be obscured by truncation artifacts. The linear nature of an MRI scan makes the truncation artifact problem an excellent example of a linear resolution problem.

Applications which would benefit from improved truncation artifact suppression include 2D chemical shift imaging (CSI) and functional magnetic resonance imaging (fMRI).

2D CSI is a method of proton spectroscopy for obtaining relative concentrations of brain metabolites including N-acetylaspartate (NAA), choline, creatine and mobile lipids *in vivo*. Typical measurements collect 32x32 voxels at $1cm^2$ in 20 minutes. The imaging resolution is necessarily poor due to restrictions in time and signal to noise. Improvement of the images' resolution and noise by modifying the image reconstruction algorithm would be valuable.

The area of fMRI is only a few years old but is already responsible for perhaps 30% of MR research. fMRI monitors the function of the brain in real time. A scan of the whole brain is acquired in a few seconds and constantly repeated. Poor signal to noise mandates that a small matrix size, typical 64x64, must be used. However, high resolution and better signal to noise will produce substantial improvements to the location of active brain centers or to the time resolution of brain events.

By far the most commonly used MRI algorithm for estimating the unknown model which produced the data is the 2D Fourier transform (2DFT) (Edelstein *et al* 1980, Chen and Hoult 1989). The 2DFT gives good results when applied to time domain (k-space) data although the resulting image has truncation artifacts. The windowed 2DFT used

for MRI is

$$m^E_{n_1 n_2} = \frac{1}{K_2 K_1} \sum_{k_2=-K_2/2}^{K_2/2-1} \sum_{k_1=-K_1/2}^{K_1/2-1} w_{k_1 k_2} exp(\frac{2\pi i k_2 n_2}{K_2 Z_2})\ exp(\frac{2\pi i k_1 n_1}{K_1 Z_1})\ d_{k_1 k_2} \qquad (2.1)$$

where $d_{k_1 k_2}$ is the k-space data with indices k_1 and k_2 and $m^E_{n_1 n_2}$ is the image of the object being scanned with indices n_1 and n_2. Z_1 and Z_2 are zero filling factors used to interpolate the resulting image and are usually integers between 1 and 8. Equation 2.1 is only correct to within a scaling factor as an estimate of the unknown model. Since the pixel to pixel contrast rather than the absolute intensity of a pixel is the primary means of interpreting the resulting images, accuracy to within a scaling factor is all that is usually necessary.

The coefficients $w_{k_1 k_2}$ are, strictly speaking, not part of the standard 2DFT. The standard 2DFT assumes $w_{k_1 k_2}$ are equal to unity. The coefficients $w_{k_1 k_2}$ are equivalent to a Fourier Transform window (FTW) (Harris 1978). They can be considered to be the Fourier transform coefficients of a deconvolution operator applied to the image resulting from the standard 2DFT. In addition, they can be thought of as a digital contact lens which, if properly shaped, sharpens the focus of the images generated by the 2DFT.

This chapter will be concerned with finding the values of $w_{k_1 k_2}$ which give the best quality image by sharpening the linear resolution and reducing the truncation artifacts

2.2 MR imaging algorithms

Many attempts have been made to reduce truncation artifacts (Haacke et al 1989, Smith and Nichols 1990, Constable and Henkelman 1991A, 1991B, Liang et al 1992) but windowed 2DFT is still the algorithm which is almost universally used. Reducing truncation artifacts is a valuable goal because it may allow more accurate interpretation of the MR images. Furthermore, the improved resolution and signal to noise could be used to shorten scanning time.

2.3 Linear resolution

The object of FTW's is to reduce the leakage of signal from undesirable frequencies into frequencies of interest (Harris 1978). The effectiveness of a FTW is assessed by Fourier transforming it and then determining how localized the resulting point spread function (PSF) is about the main peak. This is analogous to assessing the focus of a glass lens by testing how well it focuses a beam of light.

The property that light focused by a glass lens has in common with the windowed 2DFT is linear resolution. For a system to have linear resolution, it must, of course, be linear. The defining property of a linear system is superposition. Superposition states that if individual functions A_i pass through a system and give corresponding functions B_i then if the function $\sum_i a_i A_i$ is passed through the same linear system the resulting function will be $\sum_i a_i B_i$ where a_i are arbitrary constants.

To probe the performance of a linear system, it is common practice to make A_i a series of Dirac delta functions. In imaging, the corresponding functions B_i are referred to as the PSF. The shape of the PSF defines the linear resolution at the location of the Dirac delta function in A_i. The shape of the PSF gives us a good indication about how much the data can tell us about that particular region of the image, thus dealing with the inherent ambiguity of MRI data. In deconvolution the shape of the PSF's are all the same, only their location changes. Therefore, only one PSF needs to be calculated.

To generate an estimate with linear resolution, the relationship from the unknown model to the image must be linear. The forward problem for MRI is

$$d_{k_1 k_2} = \int_{-1/2}^{+1/2} \int_{-1/2}^{+1/2} m^U(x,y) exp(2\pi i k_2 y)\ exp(2\pi i k_1 x)\ dxdy \tag{2.2}$$

where $m^U(x,y)$ is the unknown model. When the equation 2.2 is substituted into equation 2.1 the result is a linear relationship between the unknown model and the image. It

therefore follows that linear resolution is obtained from MRI scanner data imaged using the 2DFT.

It is a common practice to use Dirac delta functions to try to define some concept of resolution for nonlinear systems. The resulting PSF's may have a spiky shape but since superposition does not apply to nonlinear systems, the shape of the PSF tells us very little about how functions other than Dirac delta functions will be handled by the system. Maximum entropy is a good example of a nonlinear algorithm which is considered to give "high resolution" results because it images Dirac delta functions very accurately. However, it is very difficult to predict how it will handle any other function shapes since it is a nonlinear algorithm. This may be another reason why maximum entropy does such a poor job of estimating the unknown model in MRI (Constable and Henkelman 1990). It is unfortunate that many authors who use nonlinear imaging algorithms do not clearly state that the PSF's are far less informative about the resolution of nonlinear algorithms than for linear algorithms.

2.4 Interpreter requirements

The images produced by an MRI scanner are interpreted by radiologists with the aim of discovering pathology. Providing the radiologists with images which are "interpretable" is the primary goal of MR imaging. But what exactly makes an image interpretable?

It has long been known that MRI k-space data do not give sufficient information to completely determine the unknown model because the data are finite in number and contaminated by noise. It is important that the ambiguity is clearly expressed in the MR images so radiologists can incorporate this information into their interpretation. Everyday, in thousands of hospitals around the world, radiologists interpret MR images produced by the windowed 2DFT. These images must have some property that makes

them interpretable. If this property is discoverable, it may be possible to improve upon it.

One property of the MR images produced by the standard 2DFT is that they fit the data. However, it is unlikely that this property is the important one to the radiologist for two reasons. The first reason is that many models that fit the data need to be produced for each data set to get a good handle on all the possible models that fit a particular data set. Since radiologists make a reliable interpretation from just one image, they must be getting a feel for the ambiguity of the results from the one image. The second, and more important reason, is demonstrated if you black out part of an MR image. You can still interpret the remaining areas. This is important because blacking out part of the image is equivalent to setting the values all to zero. Therefore, the resulting partially blacked out image no longer fits the data, but it is still interpretable.

Equation 2.1 generates a real and imaginary image for each data set. If all the coefficients of the windows are non zero, then equation 2.1 is invertible since the 2DFT is invertible. Thus, from the first chapter, the real and imaginary images, along with their PSF and noise statistics, completely express what is known and not known about the unknown model. This is a very powerful property since radiologists interpreting scans will have all available information in front of them when they are interpreting the scans. In practice the radiologists usually prefer to work with the magnitude image, which is a combination of the real and imaginary images. While this results in a loss of the information, it seems to be outweighed by the removal of the sometimes confusing phase information and halving the number of images.

Estimates of the unknown model which do not have linear resolution cannot express all that is known and not known about the unknown model even if they fit the data. This is probably the main reason why nonlinear algorithms such as maximum entropy have been unsuccessful in supplanting equation 2.1 and the standard inversion algorithm.

Based on the assumption that the characteristics of the MR images which matter to radiologists are linear resolution and noise, I will spend the rest of this chapter trying to improve the linear resolution of the 2DFT by a better choice of the FTW coefficients.

2.5 Calculating the FTW's

The improvement of linear resolution by deconvolution can be traced back to a paper by Wiener (1949). However, the formalism used here will be that of Backus and Gilbert (1970) who generalized Wiener's method to a variety of resolution criteria. The formalism will be presented in the one dimensional form for simplicity, but it can be easily generalized to two or more dimensions. The 2DFT can be factored into 1D DFT's applied along the rows and then the columns. It is common practice to apply the 1D DFT in the same manner to generate the 2DFT's. The same practice will be followed here.

In the Backus and Gilbert formalism, the forward problem is presented in the form

$$d_k = \int_a^b m^U(x)\, g_k(x)\, dx \tag{2.3}$$

where, as before, $m^U(x)$ is the unknown model we are trying to estimate. The function $g_k(x)$ is referred to as the data function and is the mathematical representation of the relationship between the unknown model and the data. Since our algorithm to estimate the unknown model must be linear to have linear resolution, the algorithm must be expressible in the form

$$m_j^E = \sum_k a_{jk} d_k \tag{2.4}$$

where the image, m_j^E, is approximated by a discrete function for calculation purposes. Backus and Gilbert substituted equation 2.3 into equation 2.4 to get

$$m_j^E = \int_a^b R_j(x) m^U(x)\, dx \tag{2.5}$$

where $R_j(x)$, the resolution function, is defined to be

$$R_j(x) = \sum_k a_{jk}\, g_k(x) \tag{2.6}$$

They also added the additional requirement, called the unimodular constraint, that the area of the resolution function must be unity.

It should be noted that Backus and Gilbert, when they devised their formalism and introduced the resolution function, were not trying to come up with a better way to generate an estimate with linear resolution. Rather, they were attempting to formulate a concept of resolution for models that fit the data. The forward problem they considered was nonlinear, so when combined with their linear estimation algorithm, it yielded estimates with nonlinear resolution. Later researchers applied the Backus and Gilbert approach to linear forward problems although it is not clear they all appreciated the importance of linear resolution to the interpretation of the estimate. See Backus and Gilbert (1970) and Parker (1994) for more details.

The resolution function is equal to the PSF for the linear deconvolution problem. This can be shown by substituting the Dirac delta function into equation 2.5 while keeping in mind that the data functions are all equal except they are translated from each other. This property is referred to as spatial invariance. Since we are dealing with a linear deconvolution problem, only one $R_j(x)$ needs to be calculated. Therefore I will drop the subscript and assume the peak of $R(x)$ is located at $x = 0$.

2.5.1 Linear Resolution Criteria

What is the desired shape for a resolution function? The ideal shape would be nonnegative with no side lobes and a narrow center lobe with its peak at the point of interest. Unfortunately, we usually have to compromise between these desired goals. Sometimes, a narrow center lobe is more important than no side lobes or nonnegativeness.

In other cases the reverse may be true.

Several criteria have been presented in the literature for selecting a_k to give a resolution function with the desired shape. The Dirichlet criterion, which I will refer to as the **2DFT** criterion, is given by minimizing the objective function

$$Q^{(}FD) = \int_a^b [R(x) - \delta(x)]^2 \, dx \tag{2.7}$$

where $\delta(x)$ is the Dirac delta function. The criterion is applied by substituting equation 2.6 into equation 2.7 and then using any of a variety of minimization methods. Oldenburg (1976) has shown that the Dirichlet criterion yields the 1D version of the standard 2DFT to within a term.

A commonly used FTW in MR image reconstruction is the **Fermi** FTW with a width, w_f, of 10 and a radius, r_f, of half the number of pixels (Harris 1978, Lowe and Sorenson 1997).

$$H(\nu) = \frac{1}{1 + exp[(\nu - r_f)/w_f]} \tag{2.8}$$

It is a good example of current practices in dealing with truncation artifacts and is the default FTW for the General Electric MRI scanners.

For comparison purposes I also applied the **Hamming** FTW which is recommended by Lowe and Sorenson for MR image reconstruction.

$$H(\nu) = 0.54 + 0.46cos(2\pi\nu/\nu_{max}) \tag{2.9}$$

In an attempt to improve the interpretability of the MR images, I devised two additional resolution criteria. The first, referred to as the least enclosed criterion or **monotonic** criterion, eliminated the truncation artifacts.

The least enclosed criterion accomplished its goal in a round about way. Even though the final area of the resolution function must be unity, it started out by minimizing the area subject to certain constraints. These constraints, described below, ensured the area

was not minimized to zero and the resolution function was localized about the point of interest, x. After the area was minimized, the coefficients were multiplied by the appropriate factor so the area was normalized to unity.

The criterion minimized

$$Q^{Mono} = \int_a^b R(x)dx \tag{2.10}$$

with the constraints that

$$R(x) > 0, \tag{2.11}$$

referred to as the nonnegative constraint,

$$R(0) \geq 1, \tag{2.12}$$

referred to as the peak constraint and

$$\begin{aligned} \tfrac{dR(x)}{dx} &\leq 0 \quad x \geq 0 \\ \tfrac{dR(x)}{dx} &\geq 0 \quad x \leq 0 \end{aligned} \tag{2.13}$$

referred to as the monotonic constraint. The monotonic constraint demands that the resolution function monotonically approaches the lower bound from the peak. Since the lower bound is zero, then the monotonic constraint prevents the formation of side lobes. While nonnegativeness and no side lobes have been often mentioned as desirable characteristics in the literature, I am not aware of them being required in this way previously in a resolution criterion.

An additional constraint which may be applied is the noise gain constraint. It limits the amount the noise may be increased by the FTW. The criterion expressed mathematically is

$$NG \geq \sqrt{\frac{1}{K}\sum_{k=1}^{K} a_k^2} \tag{2.14}$$

where NG is the limit on the noise gain. Decreasing the noise gain cannot improve the linear resolution and almost always makes it worse. The noise of typical MRI scans and

the noise gain the FTW generated by the described resolution criteria are low enough that the noise constraint does not have to be applied. Therefore I will not use it in this chapter.

The second criterion I devised was the **L1** criterion. It uses the peak and noise gain constraints of the monotonic constraint but minimizes the L1 norm of the resolution function

$$Q^{L1} = \int_a^b |R(x)|dx. \tag{2.15}$$

Since there is no nonnegative constraint, some sidelobes will occur. As with the monotonic constraint, after equation 2.15 is minimized, the coefficients must be multiplied by the appropriate coefficient so the area of the resolution function is one.

2.6 Method

The four FTW's as well as the standard 2DFT and a nonlinear algorithm, the Constable Henkelman method, were compared in several ways. The 1D PSF's for 128 points was calculated for each and plotted for the FTW's. The 2D PSF's for 128x256 and 32x32 pixels were also calculated by zero-filling the raw data by a factor of 8 after applying the windows. Then several measures of each 2D PSF were calculated and tabulated or plotted.

For additional comparison, a phantom and two MR images of the human brain were acquired and reconstructed using the various methods. The phantom and one image of the human brain were acquired at 128x256 pixels. A second image of the human brain was acquired at 32x32 pixels.

A comment is necessary on how the coefficients at the edges of the FTW's are handled. As Harris (1978) points out, the periodic nature of the Fourier Transform means the edge points of a FTW must be treated carefully. Fortunately, in the MRI problem, the data

measured near the edge of the FTW are usually near zero. Therefore, there is little need to be concerned about coefficients of the edges of a FTW.

The monotonic window was calculated using the AMPL optimization program (Fourer *et al* 1993) with the LOQO solver. The L1 window was calculated using the AMPL program with the DONLP2 solver.

All the FTW algorithms mentioned herein are based on the 2DFT and are thus linear. A nonlinear algorithm is the reconstruction algorithm of Constable and Henkelman (**CH**) (Constable and Henkelman 1991). In zero filling it is common practice to fill the high frequencies with zeros. Constable and Henkelman replaced these zeroes with information gleamed from the lower frequencies.

Several variations of the Constable and Henkelman algorithm have been published (Amartur and Haacke 1991A, Amartur et al 1991B, Liang ZP 1992). The variation presented by Liang is implemented here and consists of several steps. (1) Fourier transforming the zero filled raw data to spatial domain. Any truncation artifacts are thus clearly evident. (2) A two dimensional modified sigma filter is applied to the magnitude of the image to suppress truncation artifacts. (3) An edge enhancing filter is then applied to the magnitude image. (4) The data is then inverse Fourier transformed back to the frequency domain using the phase information from step (1). (5) Use the new frequency data to extrapolate the original frequency data to high high frequencies. A linear merging filter is applied to the edges of the original data to avoid discontinuities. Steps (1) through (5) are repeated three times.

It should be kept in mind that the Constable Henkelman algorithm does not produce a result with linear resolution nor does it produce a model that fits the data. While part way through the algorithm the original data is reintroduced, it is then filtered before being Fourier transformed to the final image. This filtering prevents the final result from fitting the data.

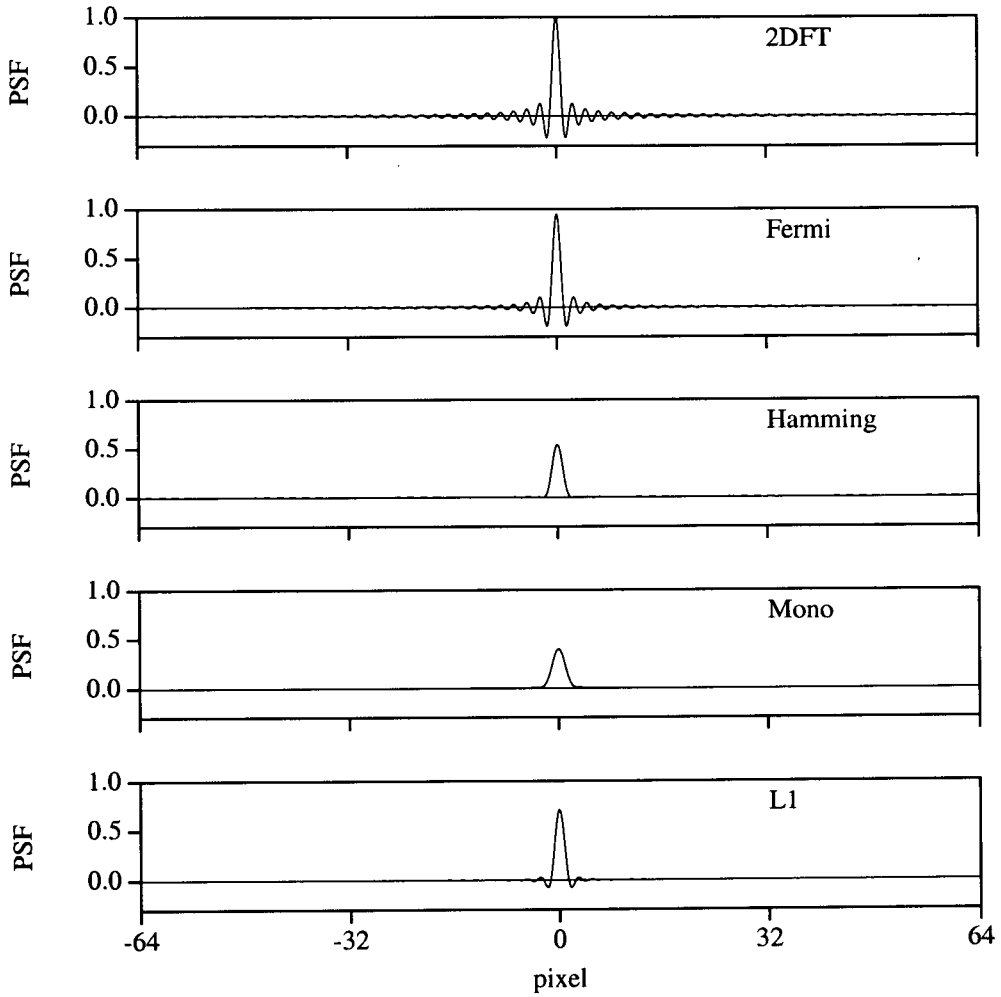

Figure 2.1: 1D PSF

2.7 Results

The 1D form of the 1D PSF's for each of the FTW's is plotted in figure 2.1.

What constitutes the best linear resolution is somewhat dependent on the type of image you are examining. Therefore, I have provided in tables 2.1 and 2.2 a variety of measures by which to compare the PSF's resulting from the various resolution criteria for a 128x256 and 32x32 k-space data set. The measures include gains and signal to noise ratios. The **PSF gain** is the height of the peak of the PSF when the unknown model is a

Code	PSF Gain	Noise Gain	Positive Fraction	FWHM	50% (Pixels)	90% (Pixels)
2DFT	1.000	1.000	0.566	1.203	7.688	50.625
Fermi	0.845	0.851	0.610	1.260	3.250	37.500
Hamming	0.292	0.398	0.923	1.817	0.938	9.250
Monotonic	0.159	0.274	1.000	2.302	1.188	2.500
L1	0.486	0.569	0.811	1.558	0.813	4.750

Table 2.1: Measures of the PSF's for 128x256

Code	PSF Gain	Noise Gain	Positive Fraction	FWHM	50% (Pixels)	90% (Pixels)
2DFT	1.000	1.000	0.615	1.202	2.438	10.187
Fermi	0.671	0.685	0.698	1.298	1.500	8.750
Hamming	0.292	0.397	0.967	1.817	0.813	1.625
Monotonic	0.182	0.293	1.000	2.138	1.063	2.437
L1	0.496	0.570	0.829	1.523	0.750	3.688

Table 2.2: Measures of the PSF's for 32x32

Dirac delta function with unit area. The **noise gain** is defined by the right hand side of equation 2.14. The **positive fraction** is the fraction of the PSF which is positive. The **FWHM** is the full width at half maximum height of the center lobe of the PSF. The units of the FWHM are the size of a pixel if the image data were applied to the k-space data.

Another important measure of the linear resolution, as expressed by the PSF, is the localization of the signal. This can be measured by calculating the fraction of the signal within a circle of a specified radius. The absolute value of the signal is used because both positive and negative parts of the PSF contribute to the signal. The radii of the circles, which includes 50% and 90% of the PSF, are also included in tables 2.1 and 2.2. Figures 2.2 and 2.3 show the complete signal-encircled curves for each of the FTW's considered.

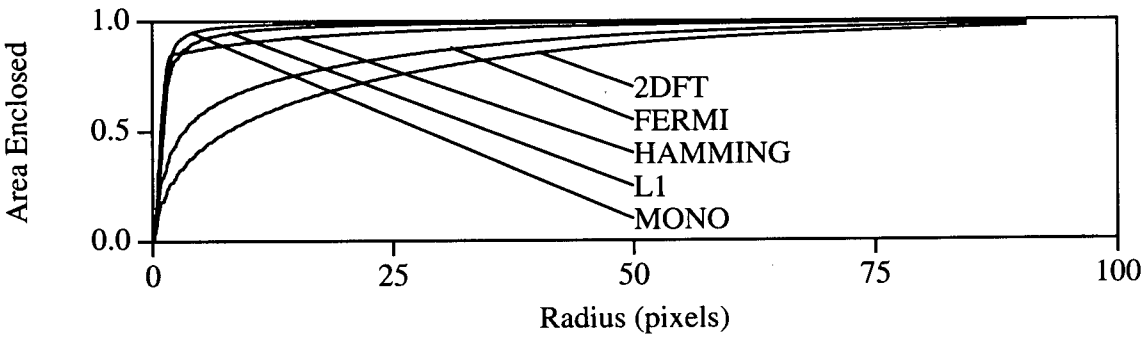

Figure 2.2: Signal-encircled in the PSF's for 128x256 pixel images.

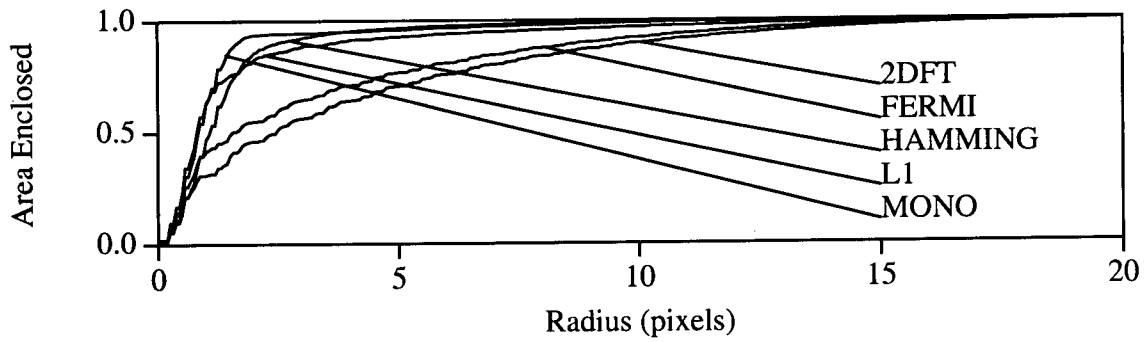

Figure 2.3: Signal-encircled in the PSF's for 32x32 pixel images.

Parameters for the PSF's for the Constable Henkelman algorithm were not represented because they are not defined. While the PSF's are normally calculated by applying the algorithm to a point source, it can be calculated by adding the point source to any other model, imaging the sum of the two and then subtracting off the image of the model. The same PSF will always result no matter what model is chosen provided the algorithm is linear. A nonlinear algorithm will yield many different answers therefore the PSF is undefined.

Each of the six reconstruction methods was applied to three MRI k-space data sets. A subjective assessment of the apparent blurriness of the grey-white matter boundary and well defined structures in the MR images show that the apparent blurriness increases with the FWHM. However, one must be cautious of subjective assessments of resolution. As Pratt mentions (1991 p 303), psychophysical experiments indicate that a photograph or visual signal with accentuated or "crispened" edges is often more subjectively pleasing than an exact photometric reproduction. The truncation artifacts for the 2DFT have the effect of accentuating edges, thus giving the illusion of better linear resolution.

A better and more reliable way to assess the practical linear resolution of the images would be to assess their diagnostic ability. Images acquired from a wide variety of pathologies, with particular emphasis on those which are demanding of resolution, should be processed using the various FTW's and the results compared for diagnostic ability. Unfortunately, such an extensive comparison of diagnostic images is outside the scope of this document.

An examination of the MR images from the FTW algorthims, figures 2.4 to 2.8 and 2.10 to 2.14, by a trained eye also shows that the truncation artifacts in the MR images decreases with the decreasing of the 50% and 90% encircled signals. As expected, the monotonic FTW produces an image with no truncation artifacts. The L1 FTW has surprisingly small truncation artifacts for its FWHM. The image produced by the

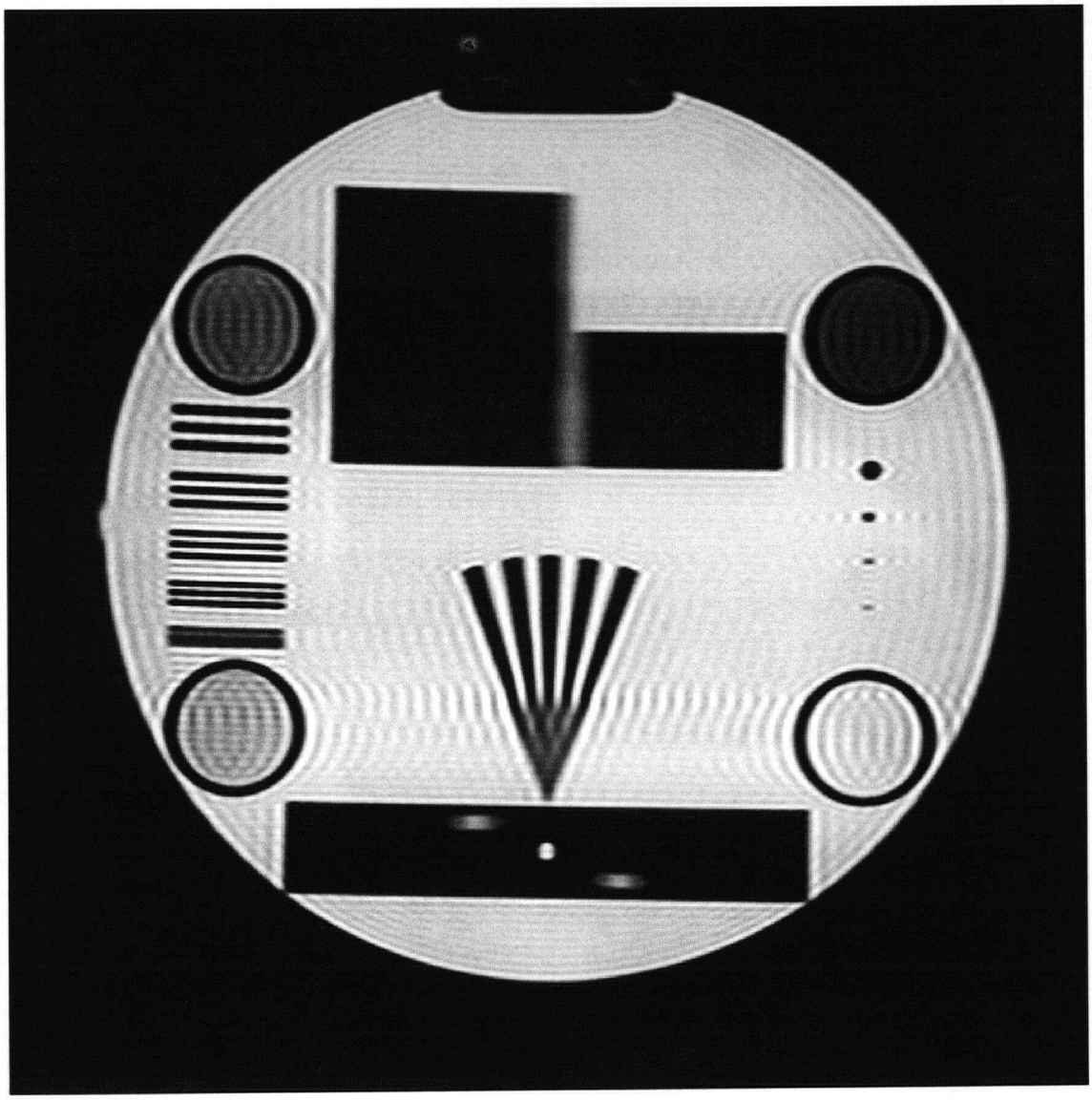

Figure 2.4: 2DFT FTW phantom image

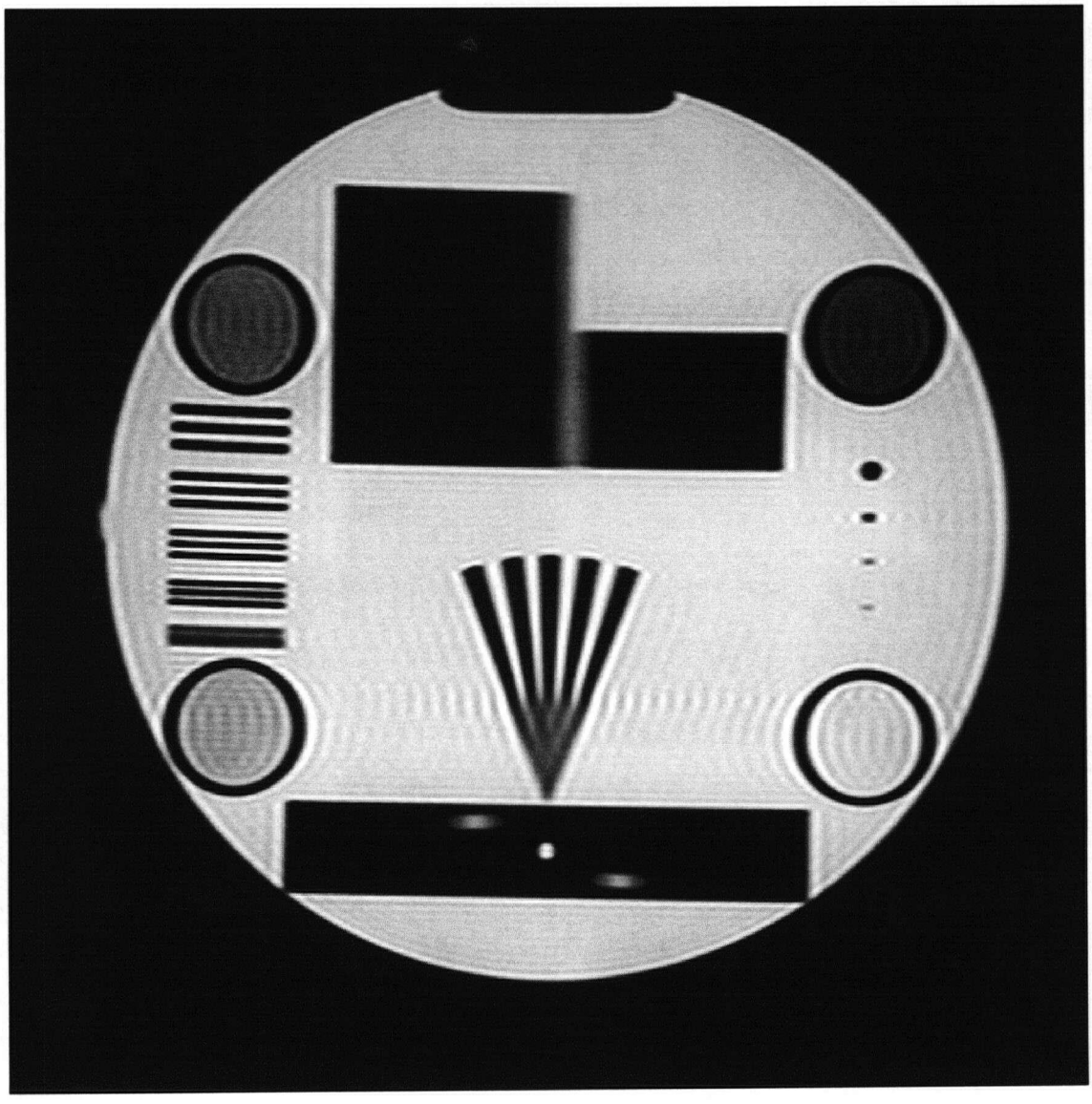

Figure 2.5: Fermi FTW phantom image

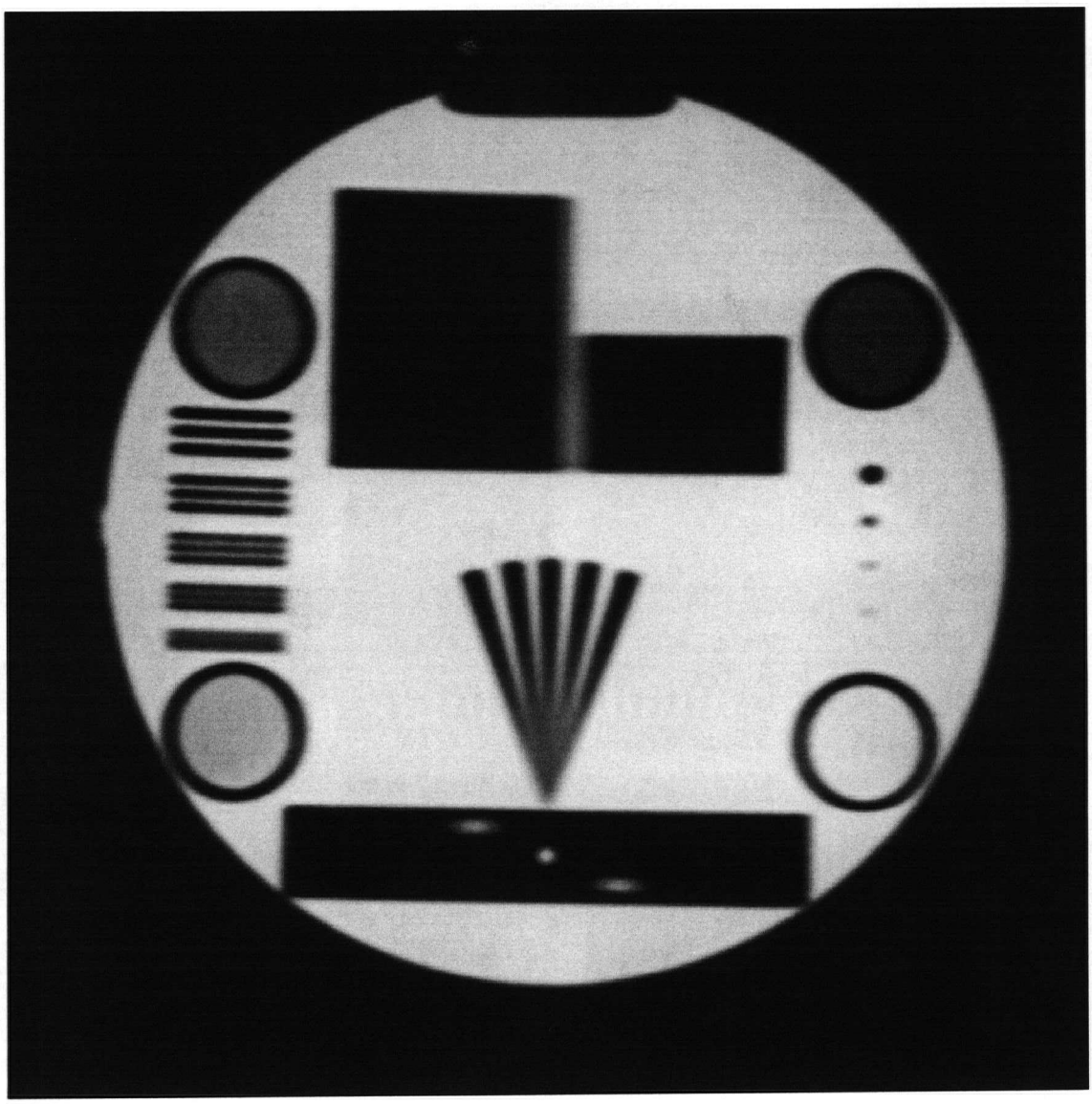

Figure 2.6: Hamming FTW phantom image

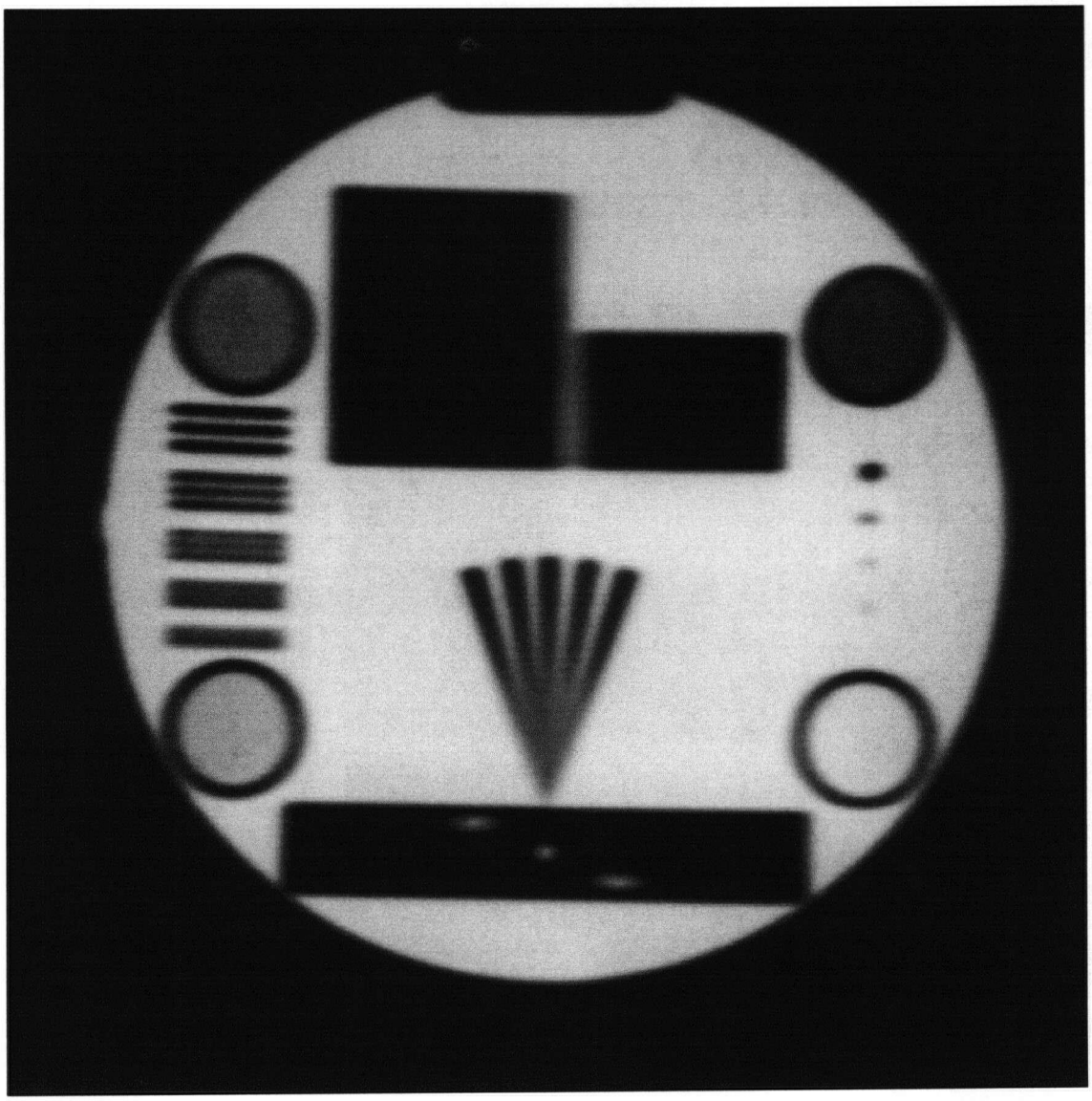

Figure 2.7: Monotonic FTW phantom image

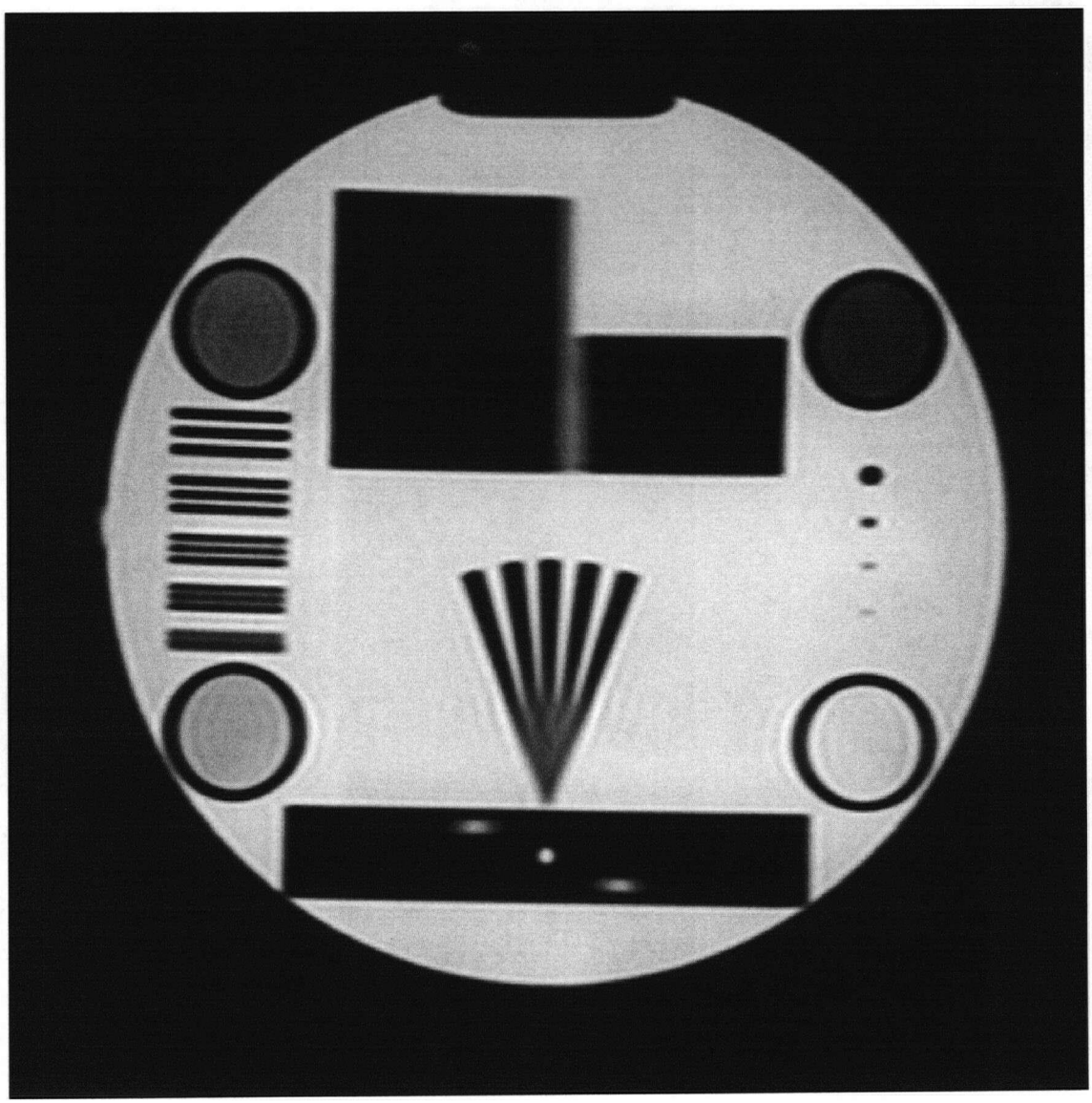

Figure 2.8: L1 FTW phantom image

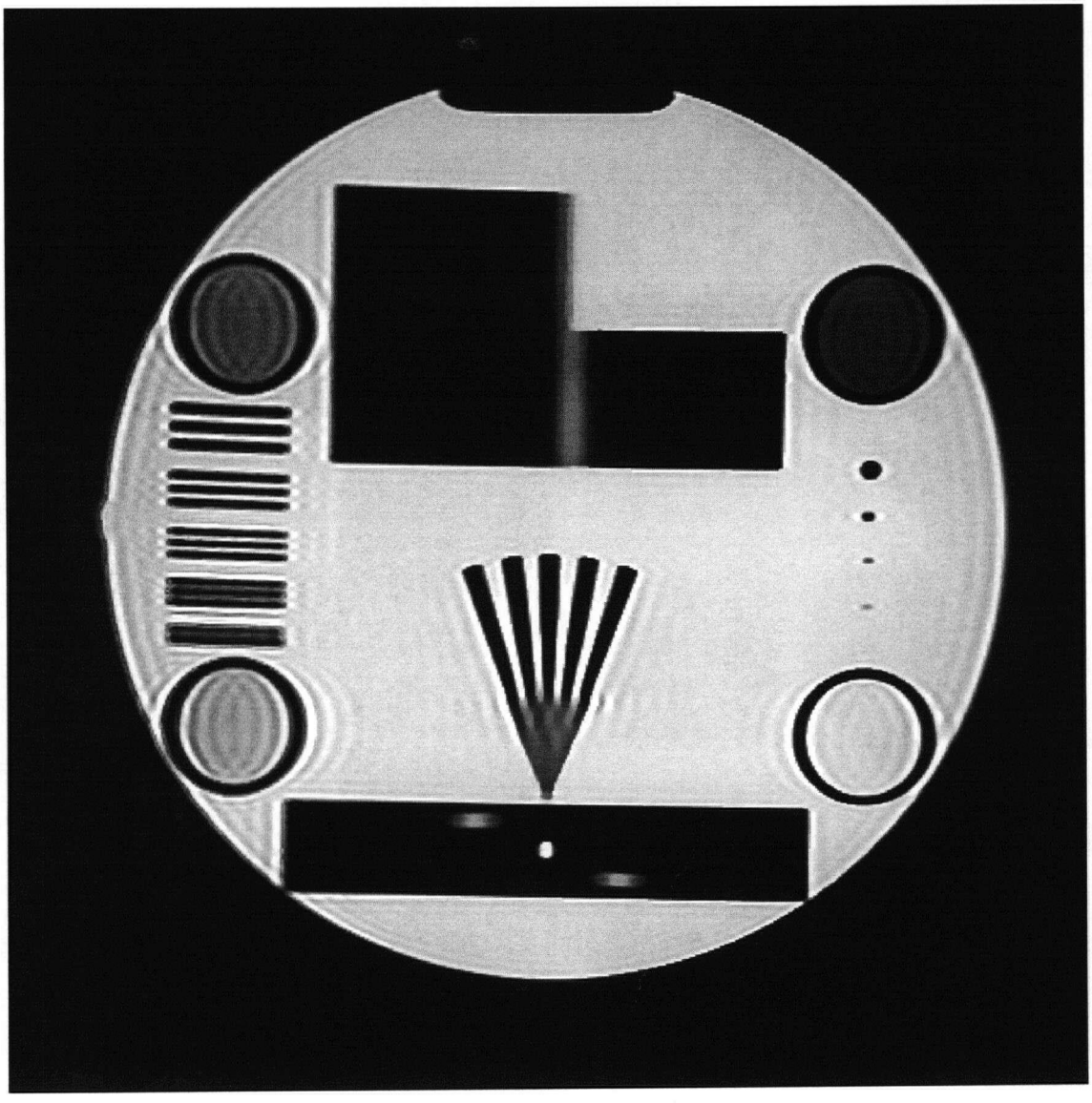

Figure 2.9: Constable Henkelman algorithm phantom image

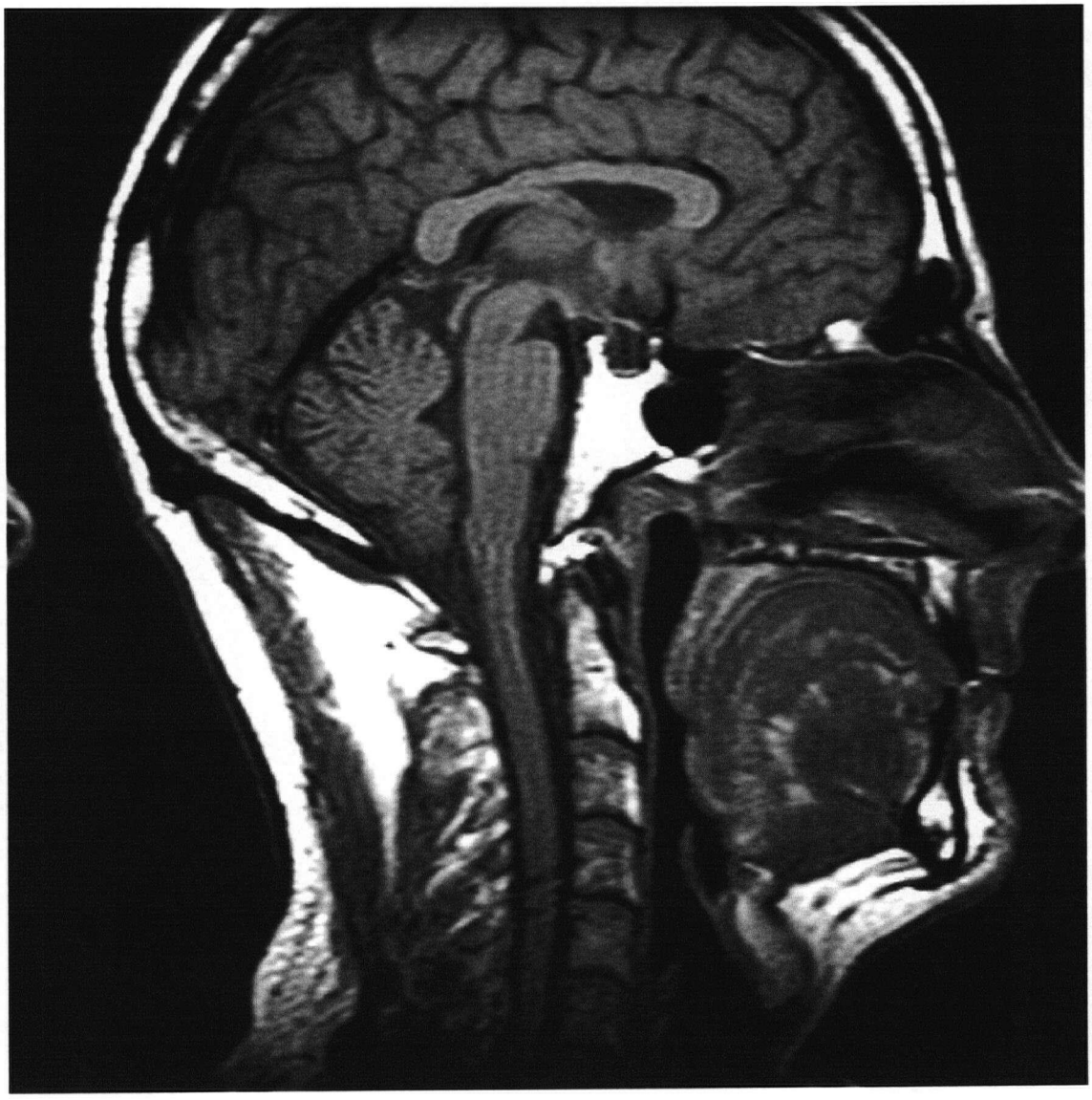

Figure 2.10: 2DFT FTW brain Image

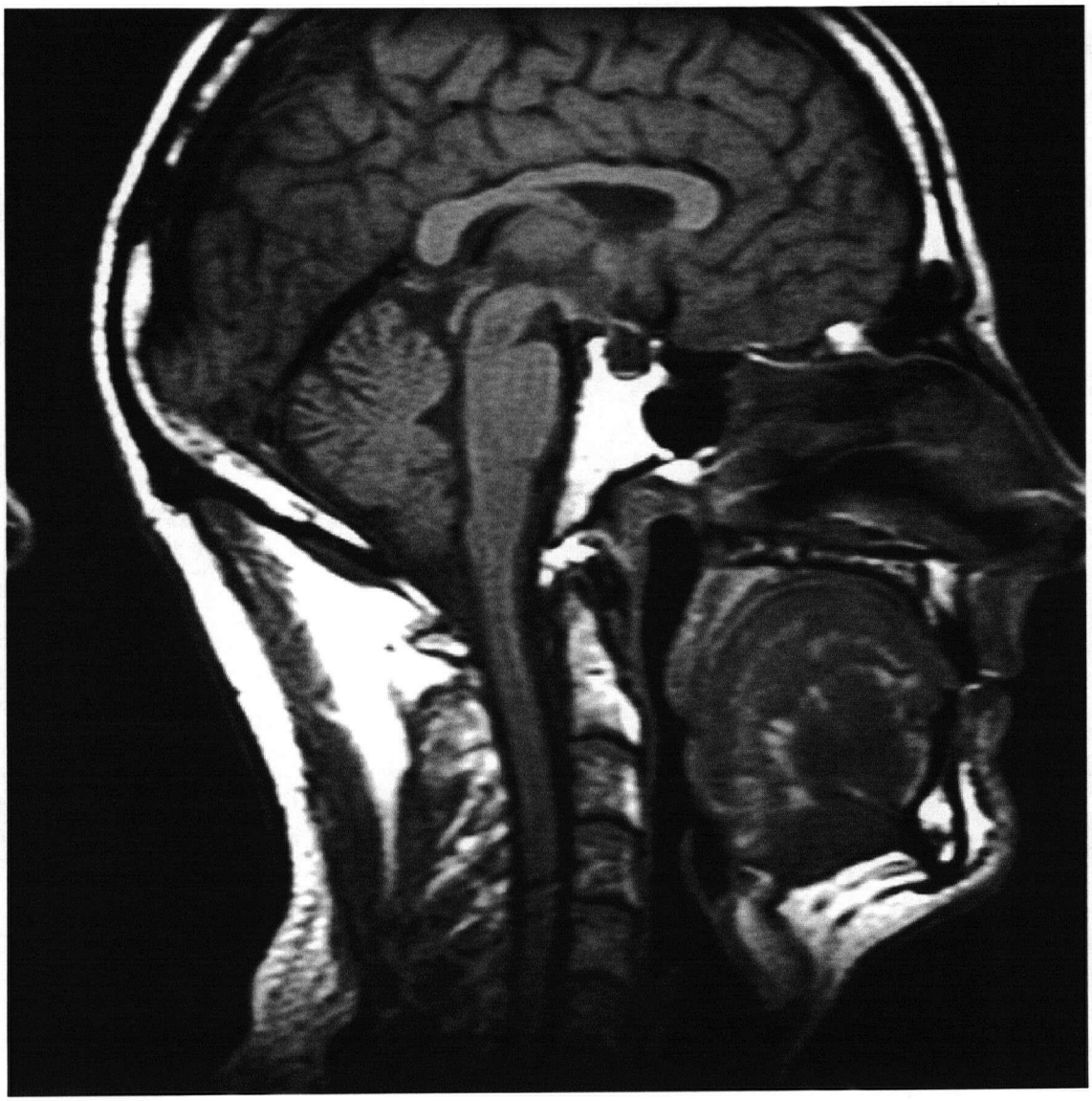

Figure 2.11: Fermi FTW brain image

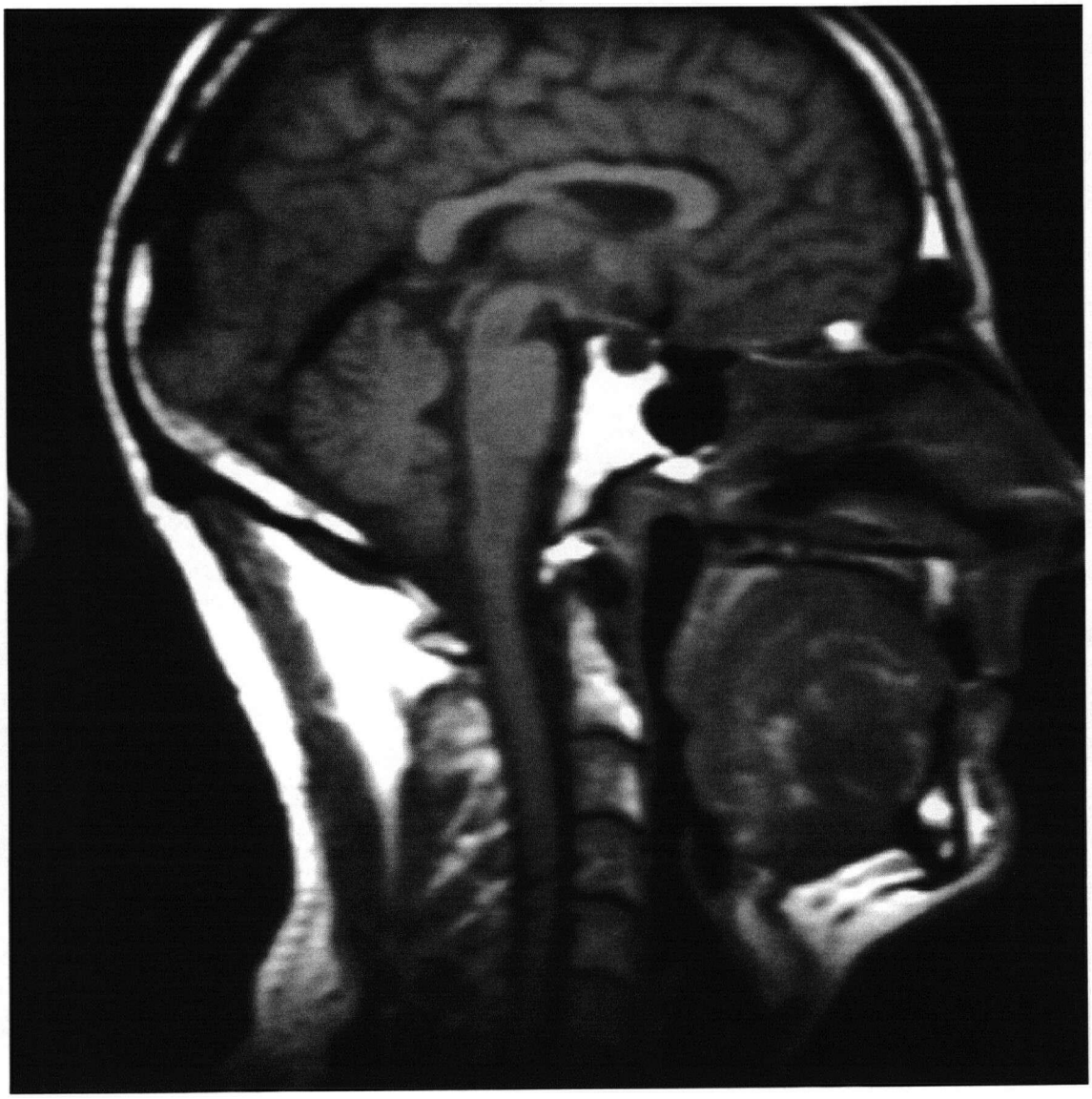

Figure 2.12: Hamming FTW brain image

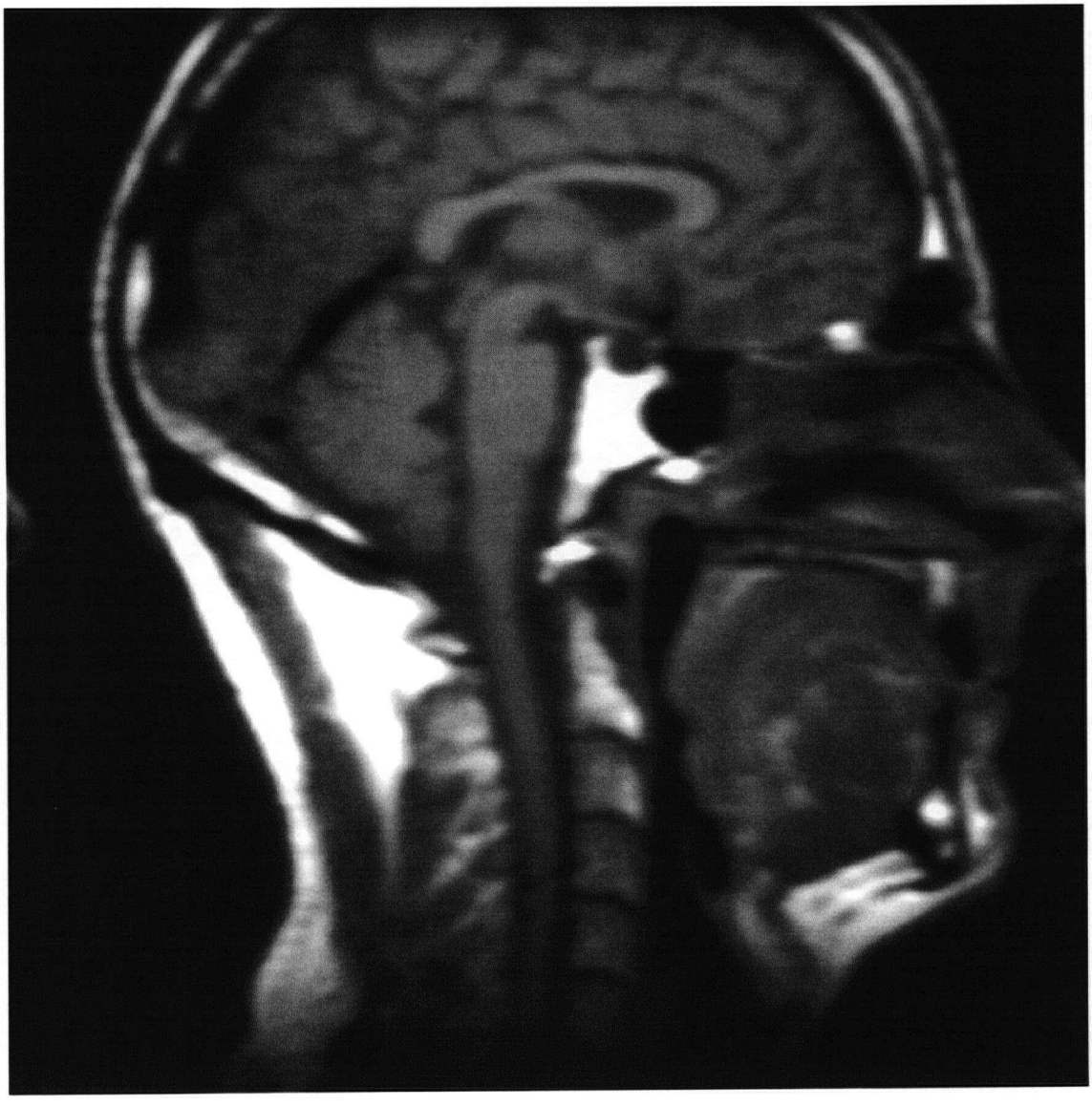

Figure 2.13: Monotonic FTW brain image

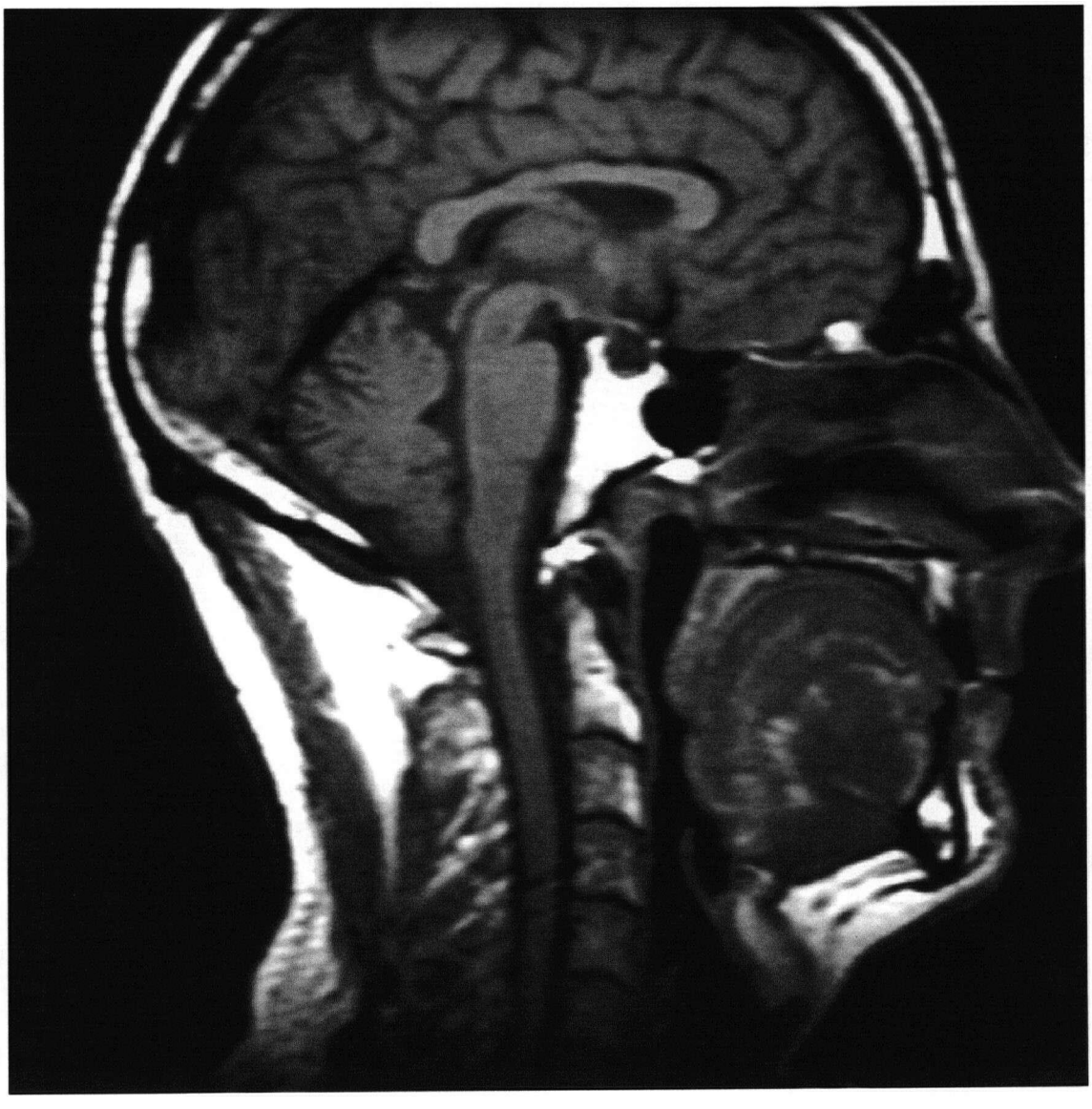

Figure 2.14: L1 FTW brain image

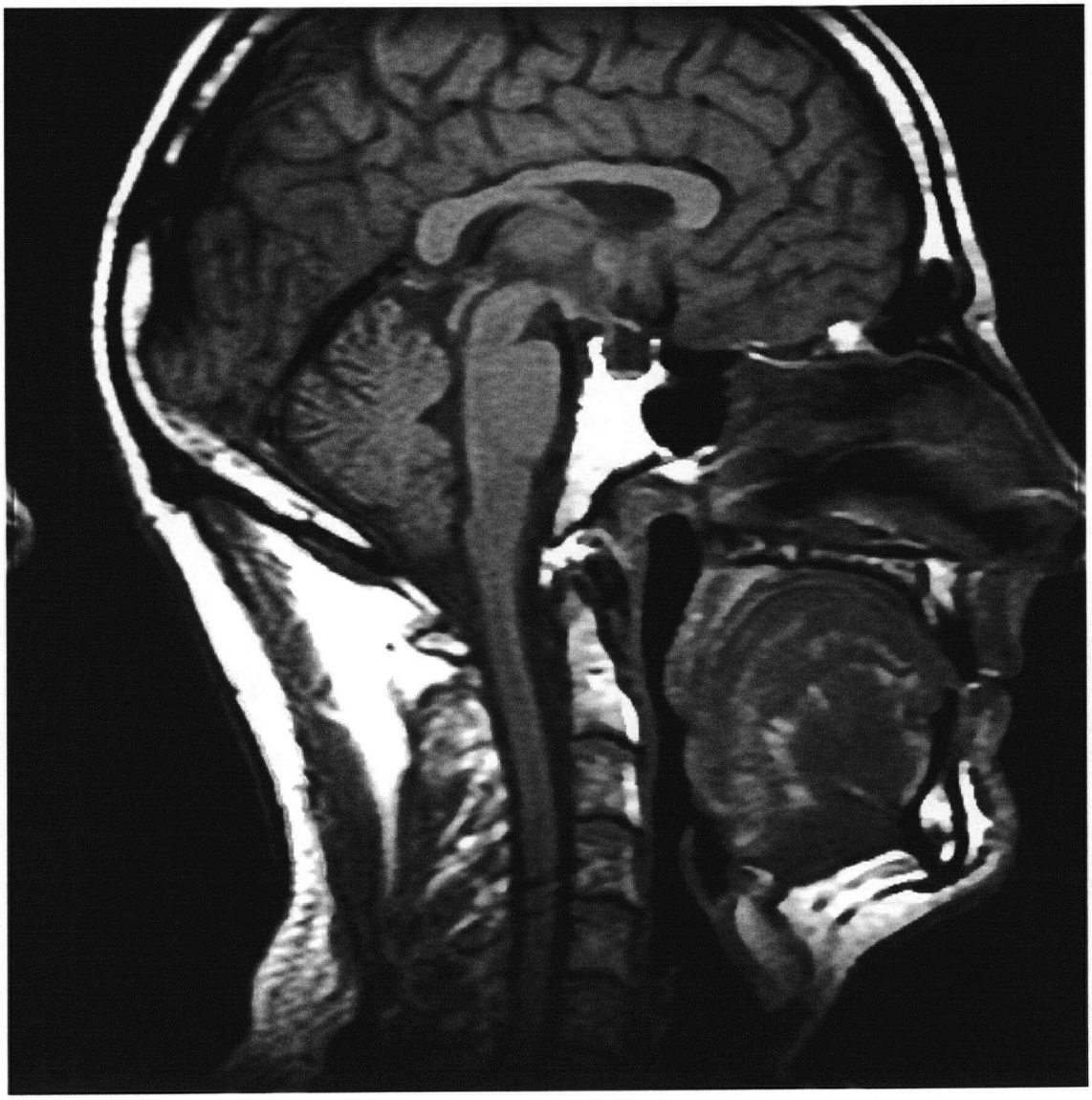

Figure 2.15: Constable Henkelman algorithm brain image

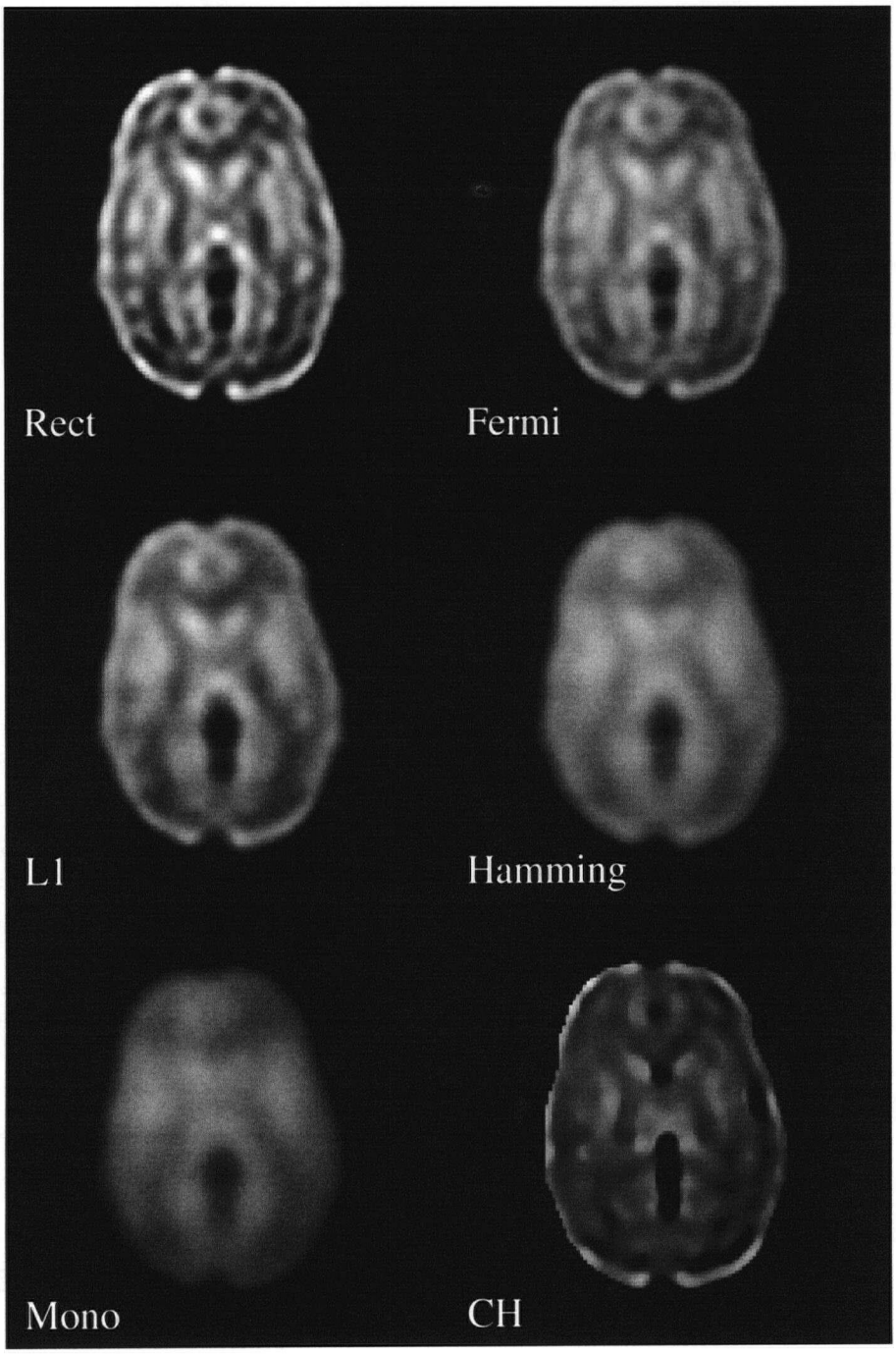

Figure 2.16: Images resulting from applying rectangular, Fermi, L1, Hamming and monotonic Fourier transform windows as well as the Constable Henkelman algorithm to 32x32 matrix raw data.

Hamming FTW also has very small truncation artifacts. For most practical purposes, there is little difference between the performance of the monotonic and Hamming FTW's.

The results for the Constable Henkelman (CH) algorithm are presented in figures 2.9, 2.15 and 2.16. As expected the CH algorithm does much better that the 2DFT or the benchmark Fermi filter for the 128x256 pixel images. Large regions of the phantom and the center part of the 128x256 brain image have almost all the the truncation artifacts suppressed. The CH algorithm performs poorly for the 32x32. This is not surprising considering the algorithms was designed with with a much larger number of pixels in mind. Comparison of CH with the L1 FTW shows the L1 FTW suppresses the truncation artifacts in a more consistent manner.

The L1 FTW seems to be the best compromise between truncation artifacts and blurriness. It has far less truncation artifacts than the Fermi FTW but is only slightly more blurry.

2.8 Conclusions

The results from the monotonic FTW conclusively demonstrate that it is possible to generate MR images with no truncation artifacts. Judging by the area enclosed plot, it should also have the best resolution. However, as evident from the images, the practical resolution is nowhere near as good as the L1 FTW images. Why the discrepancy?

It seems that the side lobes of the PSF's tend to average over areas of uniform intensity or smoothly varying structures. Only very particular types of structures cause the side lobes to reinforce each other and form truncation artifacts. Fortunately, these types of structures are not that common in human anatomy, but they are common enough to be cause for concern. In addition, the sidelobes have the effect of enhancing edges making structures stand out more clearly.

The improvements in signal to noise using the L1 FTW over the standard Fermi are also valuable. Ranging from 20% for the 32x32 image to 50% for 128x256 images, this increase in signal to noise costs nothing in terms of acquisition time or effort.

It should be kept in mind that the images generated by any of the FTW's do not reproduce the data but do provide good linear resolution. The real and imaginary images, along with their PSF and noise statistics, completely specify all that is known and not known about the unknown model. However, usually the real and imaginary images are combined, with some loss of information, into a magnitude image which is then presented to an interpreter.

It should also be kept in mind that removal of the truncation artifacts is just a standard linear resolution problem and the techniques demonstrated in this paper, in particular the L1 resolution criterion, can be applied to any other linear inversion problem in 1D, 2D or more dimensions. This class of problems includes a host of standard imaging modalities which can be found in any number of books and other publications on imaging and linear inverse theory (Menke 1984, Twomey 1977, King 1984, Parker 1994, Parker 1990).

References

Amartur S and Haacke EM 1991A Modified iterative model based on data extrapolation method to reduce Gibbs rings. *J. Magn. Res. Imag.* **1** 307–317

Amartur S, Liang Z P, Boada R, Haacke E M 1991B Phase-constrained data extrapolation method for reduction of truncation artifacts. *J. Magn. Res. Imag.* **1** 721–724

Backus G and Gilbert F 1970 Uniqueness in the inversion of inaccurate gross Earth data *Phil. Trans. R. Soc. Lond. Ser. A* **266** 123–192

Chen C-N and Hoult D I 1989 Biomedical Magnetic Resonance Technology (Adam Hilger:

New York)

Constable R T and Henkelman R M 1991A Data exploration for truncation artifact removal *Mag Res Med* **17** 108–118

Constable R T and Henkelman R M 1991B Removal of truncation artifacts in NMR imaging US Patent 5,001,429

Edelstein W A, Hutchison J M S, Johnson G and Redpath T 1980 Spin warp NMR imaging and applications to human whole-body imaging *Phys. Med. Biol.* **25** 751–756

Fourer R, Gay D M and Kernighan B W 1993 *AMPL: A modeling language for mathematical programming* (Pacific Grove CA: Duxbury Press)

Haacke M E, Liang A P and Izen S H 1989 Constrained reconstruction: a superresolution, optimal signal-to-noise alternative to the Fourier transform in magnetic resonance imaging *Medical Physics* **16** 388–397

Harris F J 1978 On the use of windows for harmonic analysis with the discrete fourier transform *Proc. IEEE* **66** 51–83

Huestis W P 1987 Construction of non-negative resolving kernels in Backus-Gilbert Theory *Geophys. J. R. Astr. Soc.* **90** 495–500

King M A, Schwinger R B, Doherty P W, Penney B C 1984 Two-dimensional filtering of SPECT images using the METZ and Wiener Filters *J Nucl Med* **25** 1234–1240

Liang Z P, Boada F E, Constable R T, Haacke E M, Lauterbur P C, Smith M R 1992 Constrained reconstruction methods in MR Imaging *Rev. Magn. Reson. Med.* **4** 67-185

Lowe M J and Sorenson J A 1997 Spatially filtering functional magnetic resonance imaging data *Mag Res Med* **37** 723-729

Menke W 1984 *Geophysical Data Analysis: Discrete Inverse Theory* (New York: Academic Press)

Oldenburg D W 1976 Calculation of Fourier Transforms by the Backus-Gilbert Method *Geophys J. R. Astr. Soc.* **44** 413–431

Oldenburg D W 1981 A comprehensive solution to the linear deconvolution problem *Geophys. J. R. Astr. Soc.* **65** 331–357

Parker J A 1990 *Image reconstruction in radiology* Boston: CRC Press

Parker R L 1994 *Geophysical inverse theory* Princeton: Princeton University Press

Pratt W K 1991 *Digital Image Processing* (John Wiley & Sons: New York)

Press W H, Teukolsky S A, Vetterling W T, Flannery B P 1992 *Numerical Recipes in C* (Cambridge: Cambridge)

Smith M R and Nichols S T 1990 A comparison of models as alternative magnetic resonance image reconstruction methods *Mag Res Imag* **8** 173–183

Twomey S 1977 *Introduction to the Mathematics of Inversion of Remote Sensing and Indirect Data* (Amsterdam: Elsevier)

Wiener N 1949 *Extrapolation, Interpolation and Smoothing of Stationary Times Series* (New York: John Wiley & Sons)

Chapter 3

Multiexponential decay analysis using optimal linear resolution

Summary

A novel algorithm is presented for estimating the relaxation distribution from multiexponential decay curves. The resulting estimates have the very useful property of optimal linear resolution when plotted against the logarithm of the time constant. Each estimate is shown to be a complete summary of all the models that fit its data. The novel multiexponential forward problem used is closely related to the Laplace transform.

The almost universally practiced method for estimating a relaxation distribution of a multiexponential decay is to calculate one or more of the infinitely many estimates which fit the data to within the noise. Whenever possible, *a priori* information is used to choose which of the estimates to calculate. However, sufficient *a priori* information may be suspect or not available. In these cases, a valuable option is to generate an estimate by linearly resolving each point of the unknown relaxation distribution as well as possible to within the limits of the noise. In general, these estimates do not fit the data and are best interpreted in a similar fashion to images produced by light focused with a glass lens. A novel criterion is introduced for optimizing the linear resolution.

3.1 Introduction

A decaying signal generally has a point, usually defined at $t = 0$ (where t is time), at which the system begins decaying. The signal decays to a constant value, often zero, as t approaches infinity. In magnetic resonance imaging (MRI), the $t = 0$ point is when the

excitation pulse energizes the sample. Particular MRI samples, such as pure water with small amounts of impurities added, have a monoexponential decay. Other MRI systems, such as parts of the human brain, are known not to have pure monoexponential decays. However, for both of these systems, and many others, valuable insight may be obtained by examining an estimate of the relaxation distributions generating the decay curves.

Estimating a relaxation distribution from a decay curve falls into a class of problems referred to as linear inversion. The linear inverse problem is one of communicating to an interpreter what is known and, equally important, what is not known about a unknown model, $m^U(y)$. In the linear inverse problem, we are provided with N data points, d_k, and the corresponding data functionals,

$$d_k = \int_a^b m^U(y)\, g_k(y)\, dy. \tag{3.1}$$

The unknown model, $m^U(y)$, is the function we are interested in estimating as well as possible. The relationship between the unknown model and the data, as expressed in equation 3.1, is referred to as the forward problem.

Each data function, $g_k(y)$, is one of the N functions which compose the data kernel of the linear transform (Tarantola 1987 p 223, Parker 1994). The data function is the mathematical representation of how the unknown model is mapped to a particular data point. It incorporates instrumental performance as well as the theoretical model. It can often be measured directly by measuring a series of $m^U(y)$ which are Dirac delta functions.

A wide variety of inversion algorithms exists for generating estimates which fit the data to within the noise. For the multiexponential transform, and its close relative, the Laplace transform, references for inversion algorithms include Bellman *et al* (1966), Smith and Nichols (1983), Essah and Delves (1988), Whittall and MacKay (1989), Whittall *et al* (1991) and Štěpánek (1993) with Istratov and Vyvenko (1999) giving

a particularly extensive survey. For a single data set, each algorithm will generate its own estimate which fits the data. Reliable *a priori* information may assist in the choice of the inversion algorithm or algorithms. For example, if you happen to know that the unknown model is a single Dirac delta function then simply fitting the data to a two parameter monoexponential is the way to go. Unfortunately, often the *a priori* information is unreliable or insufficient to reduce the ambiguity of the unknown model to a single model.

The concept of estimates with linear resolution is dealt with to a limited extent by some of the above publications, but always for models that fit the data. It is not always appreciated that estimates that have optimum linear resolution but do not fit the data are widely used in data analysis. A common example is the windowed discrete Fourier transform (DFT) which is often implemented as the fast Fourier transform (FFT) (Harris 1978). A DFT without a window produces the smallest L2 model that fits the data. Applying a window before the applying an DFT results in a estimate which fits the data times the window, not the original data. The goal of applying a window is usually to improve the shape of the point spread function (PSF), which is a measure of the linear resolution of the transform. As demonstrated in chapter one, for a Fourier transform, the resulting complex frequency distribution along with the PSF and a measure of the noise, completely quantifies our knowledge of the unknown frequency distribution and is also easily presented to an interpreter.

Haario and Somersalo (1987) applied the Backus and Gilbert (1967, 1968, 1970) spread criterion to estimating the relaxation distribution of the Laplace transform. The linear resolution of the resulting relaxation distribution decreased by a factor of 10 for each factor of 10 increase in the time constant which, for practical purposes, is of little use.

Whittall *et al* start with the data functions given in equation 3.2 and use a novel

resolution criterion to achieve resolution functions with useful characteristics but did not suggest they would be useful for estimating relaxation distributions. Considering the same paper contained several estimates of relaxation distributions which fit the data, this seems to be another example of the predominant belief that only models that fit the data are useful estimates.

3.2 The multiexponential forward problem

Before proceeding farther we must first convert the data functions for the forward problem of the multiexponential problem from being a function of τ to a function of $ln(\tau)$. The change of variable must take place in the context of equation 3.1 to ensure the form of the equation is preserved. The multiexponential forward problem commonly used in data analysis and inverse theory has the form

$$d_k = \sum_i m_i e^{-t_k/\tau_i} \tag{3.2}$$

Equation 3.2 can be expressed in the continuous form

$$d_k = \int_0^\infty \sum_i m_i \delta(\tau - \tau_i) d\tau \; e^{-t_k/\tau} \tag{3.3}$$

where $\delta()$ is the Dirac delta function.

Applying the change of variables

$$y = \ln(\tau) \tag{3.4}$$

to equation 3.3, without simplification, yields

$$d_k = \int_{-\infty}^{+\infty} \sum_i m_i \delta(e^y - e^{y_i}) e^{-t_k e^{-y}} (e^y dy) \tag{3.5}$$

Applying the Dirac delta function identity

$$\delta(f(x)) = \sum_i \frac{1}{f'(x_i)} \delta(x - x_i) \; for \; each \; f(x_i) = 0 \tag{3.6}$$

(Cohen-Tannoudji *et al* 1977 p 1470) to equation 3.5 along with standard simplifications yields

$$d_k = \int_{-\infty}^{+\infty} \sum_i m_i \delta(y - y_i) e^{-t_k e^{-y}} dy \tag{3.7}$$

Substituting

$$m^U(y) = \sum_i m_i \delta(y - y_i) \tag{3.8}$$

into equation 3.7 gives

$$d_k = \int_{-\infty}^{+\infty} m^U(y)\, e^{-t_k e^{-y}}\, dy. \tag{3.9}$$

I will use the form of the multiexponential transform in equation 3.9 as the forward problem. The data functions taken from equation 3.9 are

$$g_k(y) = e^{-t_k e^{-y}}. \tag{3.10}$$

It should be noted that all data functions approach 1 as y approaches infinity.

I am not aware of the data function in equation 3.10 being derived previously either rigorously, using change of variable, or otherwise. Many different changes variable have been used in the literature to formulate various multiexponential transforms. A changes of variables very similar to $y = ln(\tau)$ was applied in by Smith and Nichols (1983) but to the analytical inverse of the the Laplace transform. Provencher (1976) applied $z = ln(t_k)$ to the forward Laplace transform. Gardner *et al* (1959) applied the pair of change of variables, one of which was similar to $y = ln(\tau)$, to the Laplace transform but again the result was quite different from equation 3.9. After an extensive review of the literature, I have not found the forward problem reformulated in the form of equation 3.9.

The MRI decay curve of interest had 48 data points with the first 32 separated by 10ms and the last 16 separated by 300ms. Since the last 16 points were spaced at 30 times the first 32, for convenience, all intervals were divided by 10ms. The 10ms factor will be referred to as the dwell time. Data functions are shown in figure 3.1. Note that

while the data functions are plotted as a function of logarithmic variable y, the axis is labeled in units of time constants τ for convenience.

In specific problems it may be desirable to increase the sensitivity selectively to specific time constants and reduce it at other time constants. This can be dealt with by dividing the unknown model by a balancing function, $B(y)$, in equation 3.9. This gives

$$d_k = \int_{-\infty}^{+\infty} \frac{m^U(y)}{B(y)} B(y) e^{-t_k e^{-y}} \, dy. \tag{3.11}$$

The new data function then becomes

$$g_k(y) = B(y) e^{-t_k e^{-y}}. \tag{3.12}$$

While many balancing functions are possible, a useful balancing function is $B(y) = e^{-wy}$ which increases the relaxation at late time constants by, in effect, multiplying the relaxation distribution by τ^w. A value of $w = 1$ is often useful. The applications covered in the rest of this document do not use balancing functions.

In some instruments, data may be acquired by averaging over a window rather than at a particular point in time. To calculate the data function for a window all one needs to do is convolve the point data function with the weighting function of the window.

If the data has 100's, 1000's, or 10,000's of points in the decay curve, it is much more efficient computationally, for both calculating the estimation matrix and applying it to data, to average adjacent data points together to create a new group data point. The corresponding data functions must also be averaged together to get the composite data function corresponding to the new averaged data point. The size of the groups is important. Groups which are too large will cost resolution in the resolution functions. Groups which are too small are inefficient. I have found logarithmic group size to be a good choice for evenly spaced points. The equation

$$G_l = max(1, int(A\ 10^{l/N}) \tag{3.13}$$

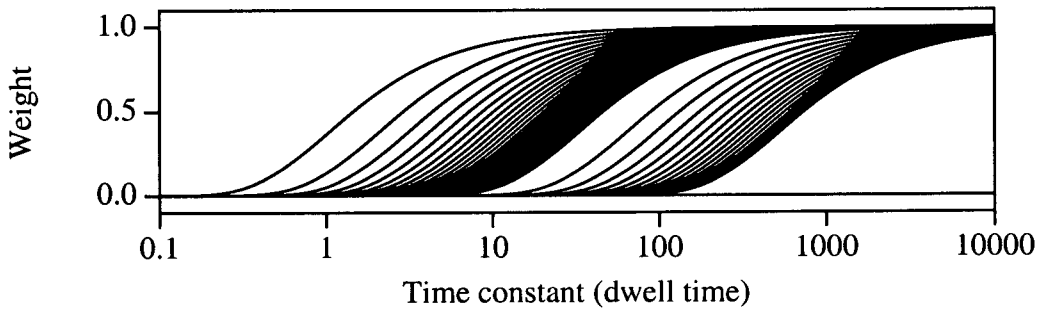

Figure 3.1: Data functions for the modified forward problem.

yields good group sizes for A=0.1 and N=10 where $l = 1, 2, 3, ...$ and G_l is the group size. As an example, for 10,000 points the grouping would be

$$\begin{aligned} &1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 3\ 3\ 5\ 6\ 7\ 10\ 12\ 15\ 19\ 25\ 31\ 39\ 50\ 63\ 79 \\ &100\ 125\ 158.199\ 251\ 316\ 398\ 501\ 630\ 794\ 1000\ 1258\ 1584\ 1995\ 309. \end{aligned} \tag{3.14}$$

According to this list, the first 13 data points would be treated individually. The next two data points would be averaged together followed by the next 3. The total sum of this list of numbers is 10,000, as would be expected.

3.3 Designing an estimation matrix

Backus and Gilbert introduced the resolution function as a way of characterizing the linear resolution of a linear inversion. A linear inversion can always be implemented as a matrix multiplication but is not always implemented that way. Therefore any linear inversion can calculated by

$$m_j^E = \sum_{k=1}^{N} a_{jk} d_k \tag{3.15}$$

where m_j^E is the estimate of the unknown model at points y_j after multiplying the data by estimation matrix a_{jk}.

Backus and Gilbert substituted equation 3.1 into equation 3.15 to get

$$m_j^E = \int_I m^U(y) \,[\sum_{k=1}^{N} a_{jk} g_k(y)] dy \tag{3.16}$$

and defined the resolution function to be

$$R_j(y) = \sum_{k=1}^{N} a_{jk}\, g_k(y). \tag{3.17}$$

The resolution function gives a concise mathematical description of the linear resolution at each point of the estimate. It should be noted that equation 3.17 is completely independent of the unknown model and the data. Thus the linear resolution is independent of the unknown model and data.

One restriction on the value of the coefficients is that the area of the resolution function must be unity,

$$1 = \int_a^b R_j(y)\, dy. \tag{3.18}$$

This requirement ensures each point in the estimate is a local average of the unknown model.

The criterion for selecting the coefficients, which correspond to a row of the estimation matrix, that yield the optimal linear resolution is referred to as the resolution criterion. Together, the resolution function and its noise gain, which is defined below, give a concise formulation of the ambiguity of the estimate of the unknown model from information given by the data and corresponding data functions. Previous linear algorithms for estimating relaxation distributions have had poor resolution. By achieving the maximum resolution for a given noise gain much improved linear resolution is achieved.

Several criteria for finding resolution functions have been published. The most familiar one is the Dirichlet criterion (Tarantola 1987 p 461) and is achieved by finding the coefficients which minimize I_j^{LS} where

$$I_j^{LS} = \int_a^b [\delta(y - y_j) - R_j(y)]^2\, dy. \tag{3.19}$$

It is also referred to as the least squares resolution criterion. This resolution criterion will generate an estimate of the unknown model which is equal to, the least squares fit estimate of Wiener (1949) for the deconvolution problem (Menke 1984 p 67, Twomey 1977). The result is also equal to the smallest L2 model that fits the data, to within the noise.

Backus and Gilbert (1968) introduced the spread criterion

$$I_j^{SP} = \int_a^b (y - y_j)^2 R_j^2(y) dy. \tag{3.20}$$

Other resolution criteria have been published (Oldenburg 1976 and Huestus 1987) but the most popular methods seem to be the Dirichlet and the Backus-Gilbert spread criterion.

3.4 Designing a estimation matrices

When focusing an image with a lens, the goal is to make each point in the image resolve the smallest possible region of the object being imaged. The focusing is accomplished by first grinding a lens to the optimum shape and then positioning it properly between the object and the imaging plane. The goal for generating a estimation matrix is similar. The goal is to calculate linear combinations of the data functions which yield a resolution function that resolves as small a region of the unknown model as possible.

3.4.1 Noise gain and resolution functions

In addition to linear resolution, the other important characteristic at each and every point in an estimate is the standard deviation due to the noise in the data.

If the noise in the data is uncorrelated, has a mean of zero, and has the same standard deviation for all data points, it can be characterized by a single standard deviation. The noise gain for each point in an estimate is defined to be the standard deviation of the

point in the estimate divided by the standard deviation of the noise in the data and can be calculated directly from the coefficients

$$NG_j = \sqrt{\sum_{k=1}^{N} {a_{jk}}^2}. \tag{3.21}$$

Often, the resolution functions resulting from the spread criterion have several undesirable properties. These include peaks, which do not lie exactly at the corresponding point in the estimate, and positive and negative side lobes on the wings of the main lobe. Huestis (1987) proposed a method for calculating resolution functions which do not have negative side lobes. However, these resolution functions still have peak alignment problems and the possibility of positive side lobes. In addition, it is computationally very expensive.

3.4.2 Least enclosed criterion

A novel resolution criterion is described below and is referred to as the least enclosed criterion. It produces resolution functions which are nonnegative, have no side lobes and are relatively narrow in width. In addition, the criterion can incorporate a predefined noise gain.

The least enclosed criterion accomplishes its goal in a round about way. Even though the final area of the resolution function must be unity, it starts out by minimizing the area subject to certain constraints. These constraints, described below, ensured the area was not minimized to zero, and the resolution function is localized about the point of interest, y^{peak}. After the area is minimized, the coefficients are multiplied by the appropriate factor so the area is normalized to unity. Therefore we first minimize a temporary resolution function, $R^{temp}(y)$, and then normalize its area to give the final resolution function $R(y)$.

To satisfy the least enclosed resolution criterion the resolution function, $R_j^{temp}(y)$, must be the minimum of

$$I_j^{LE} = \int_{-\infty}^{+\infty} w_j(y) R_j^{temp}(y) dy \tag{3.22}$$

subject to (1) the nonnegative constraint

$$R_j^{temp}(y) \geq 0, \tag{3.23}$$

(2) the monotonicity constraint

$$\begin{array}{ll} \frac{dR_j^{temp}(y)}{dy} \leq 0 & y \geq y_j^{peak} \\ \frac{dR_j^{temp}(y)}{dy} \geq 0 & y \leq y_j^{peak}, \end{array} \tag{3.24}$$

where y_j^{peak} is the location of the resolution function peak, and (3) the peak constraint

$$R_j^{temp}(y_j^{peak}) \geq 1. \tag{3.25}$$

The peak is located at the point of the relaxation distribution we are interested in estimating with this resolution function.

The constraints demand that $R_j(y)$ is nonnegative, is monotonically decreasing from the peak and that the peak is located at y_j. $R_j(y)$ is localized by requiring the peak to have a value greater than unity and then by minimizing the area subject to the monotonicity constraints. The function $w_j(y)$ is an optional weighting function which can be used to force the resolution function to shift area between the wings and the central region. Values of 1, $|y - y_j^{peak}|$, and $(y - y_j^{peak})^2$ for $w_j(y)$ are good ones to try. I will use $w_j(y) = 1$.

The effectiveness of the minimization can be checked by how close the value of the peak of the resolution function is to unity since the minimum area should yield a peak value of exactly unity. In addition to these constraints $R_j(y)$ should have unit area to better approximate a delta function. This is accomplished by dividing the coefficients a_{jk} resulting from the minimization of the area of $R_j(y)$ by the appropriate factor.

An upper limit on the noise gain, NG_j, of a particular point in an estimate can be applied by the additional constraint

$$\sqrt{\sum_{k=1}^{N} {a_{jk}}^2} \leq NG_j \sum_{k=1}^{N} a_{jk} \int_{-\infty}^{+\infty} g_k(y)dy. \tag{3.26}$$

where the area of the resolution function is included because the area of the resolution function must be normalized before the noise gain is calculated. The constraints on the resolution function will not always be satisfied for every y_j^{peak}. Therefore, no corresponding $R_j(y)$ will exist for such cases. This nonexistence indicates that there is not enough information in the data functions to resolve the unknown model to the desired accuracy at the point of interest within the given noise constraint.

When calculating the coefficients which give the resolution function, the area of an integral in equation 3.22 needs to be calculated. The upper bound of the integral is infinity. This will cause stability problems for the optimization of the integral because many of the non optimum solutions will have infinity area. The optimization can be stabilized by setting the upper bound to a large but a finite value. One hundred times the logarithm of the largest time at which the decay was measured seems to be a good choice for the upper bound. This approximation of infinity by a large value works because all the data functions approach a constant value as τ approaches infinity.

3.4.3 Calculating the estimation matrices

To apply the constraints of the least enclosed resolution criterion, the data functions were discretized into 625 points with a spacing of 20 logarithmically spaced points per decade starting at $t = 0.01$. The simple penalty method given in Fletcher (1987) was used to perform the constrained minimization.

Figure 3.2 shows all the resolution functions for the five estimation matrices. For convenience, the noise gain is constant for each estimation matrix but this is not a

necessary requirement. Initially, the graphs may appear crowded, but, if you keep in mind that the peak of each resolution function is at the point of interest in the relaxation distribution estimate and that every resolution function decreases monotonically from its peak, you should have little difficulty picking out individual resolution functions. Since the area of each resolution function is unity on the logarithmic scale, the height of each peak is inversely proportional to the width of the resolution function. Therefore, the higher the peak, the better the resolution.

As the noise gain of the estimation matrices increases from figure 3.2, the resolution also increases. This is consistent with the trade-off between noise and resolution introduced by Wiener(1949) for the deconvolution problem and generalized for linear forward problems by Backus and Gilbert (1968).

3.5 Estimates of the unknown model

Several different unknown models, $m^U(y)$'s, were chosen to demonstrate the use of the estimation matrices. Figure 3.3 shows the estimates of single exponential functions with no noise for the same estimation matrices as used for figure 3.2. In optics, these images are commonly referred to as point spread functions (PSF's).

While some may argue that the resolution of the estimation matrices is poor because the PSF's are not particularly narrow, this limited resolution is an expression of the ambiguity of the estimation. It tells the interpreter that many different unknown models could have produced the same data to within the noise. To resolve the relaxation distribution better, the interpreter needs to improve the signal to noise of the data or sample more data points.

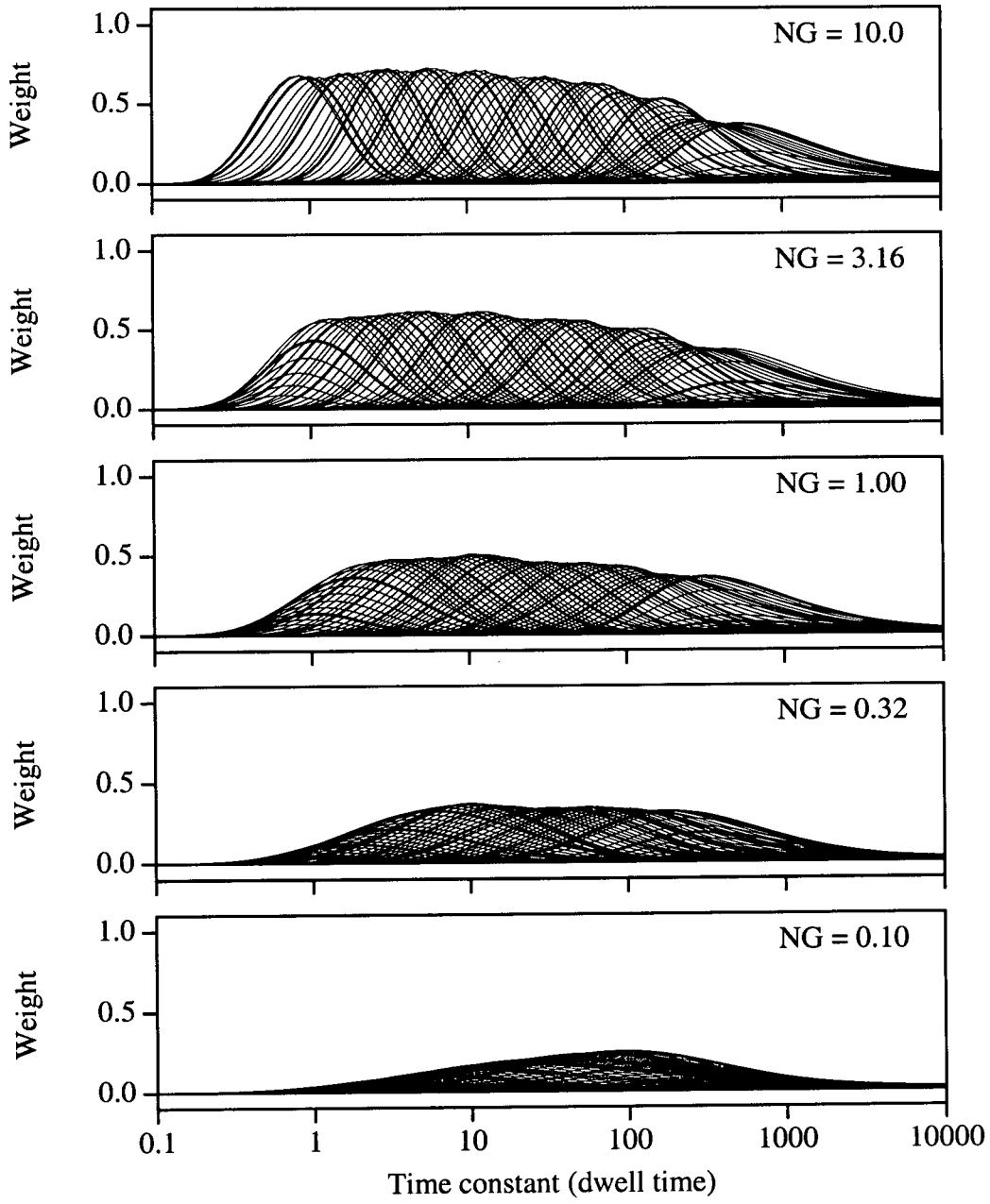

Figure 3.2: Resolution functions for the five estimation matrices. Five resolution functions per decade are highlighted to aid interpretation of the plots.

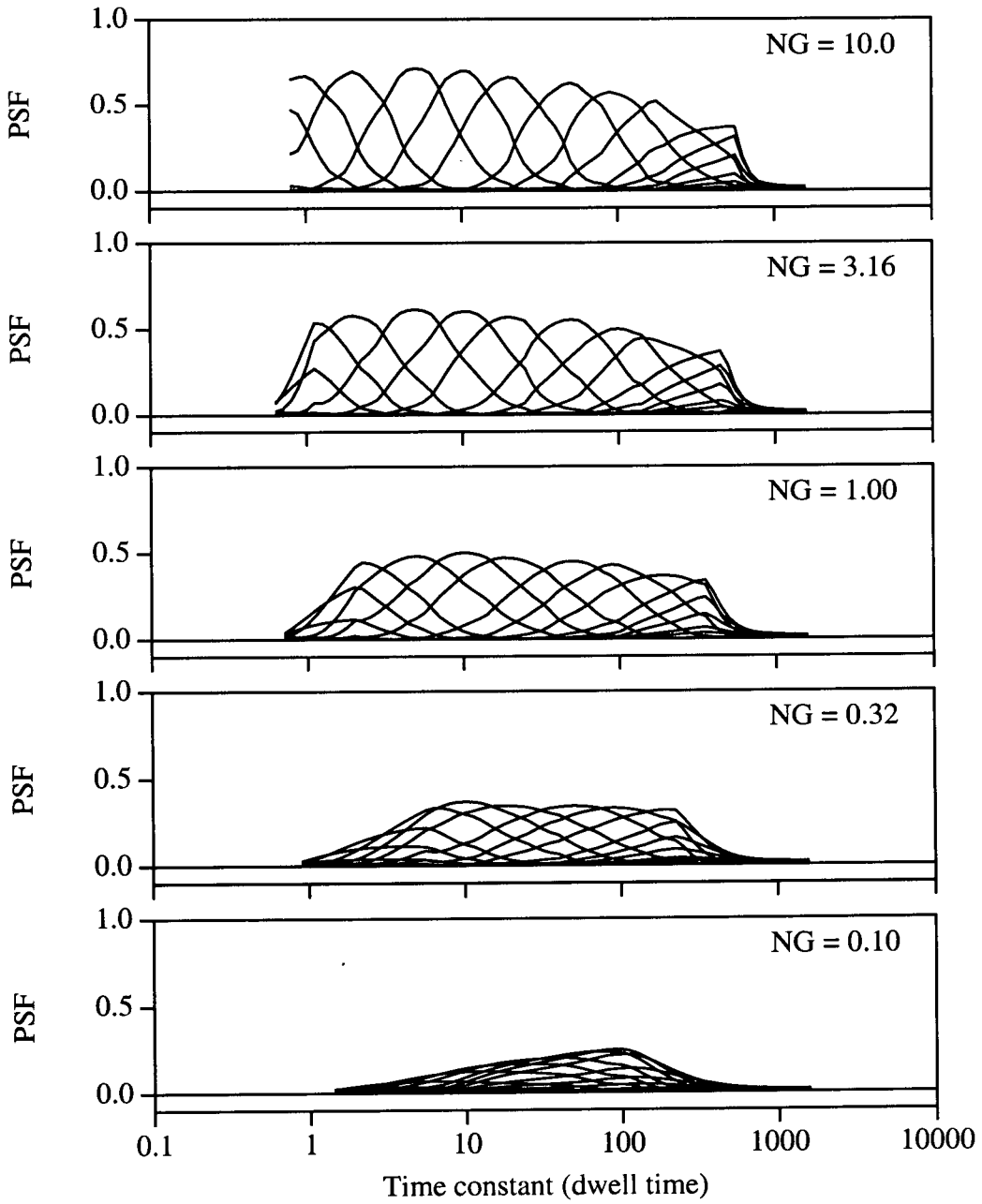

Figure 3.3: PSF's for the five estimation matrices. The point sources are located at 0.5, 1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000 and 10000 dwell times respectively.

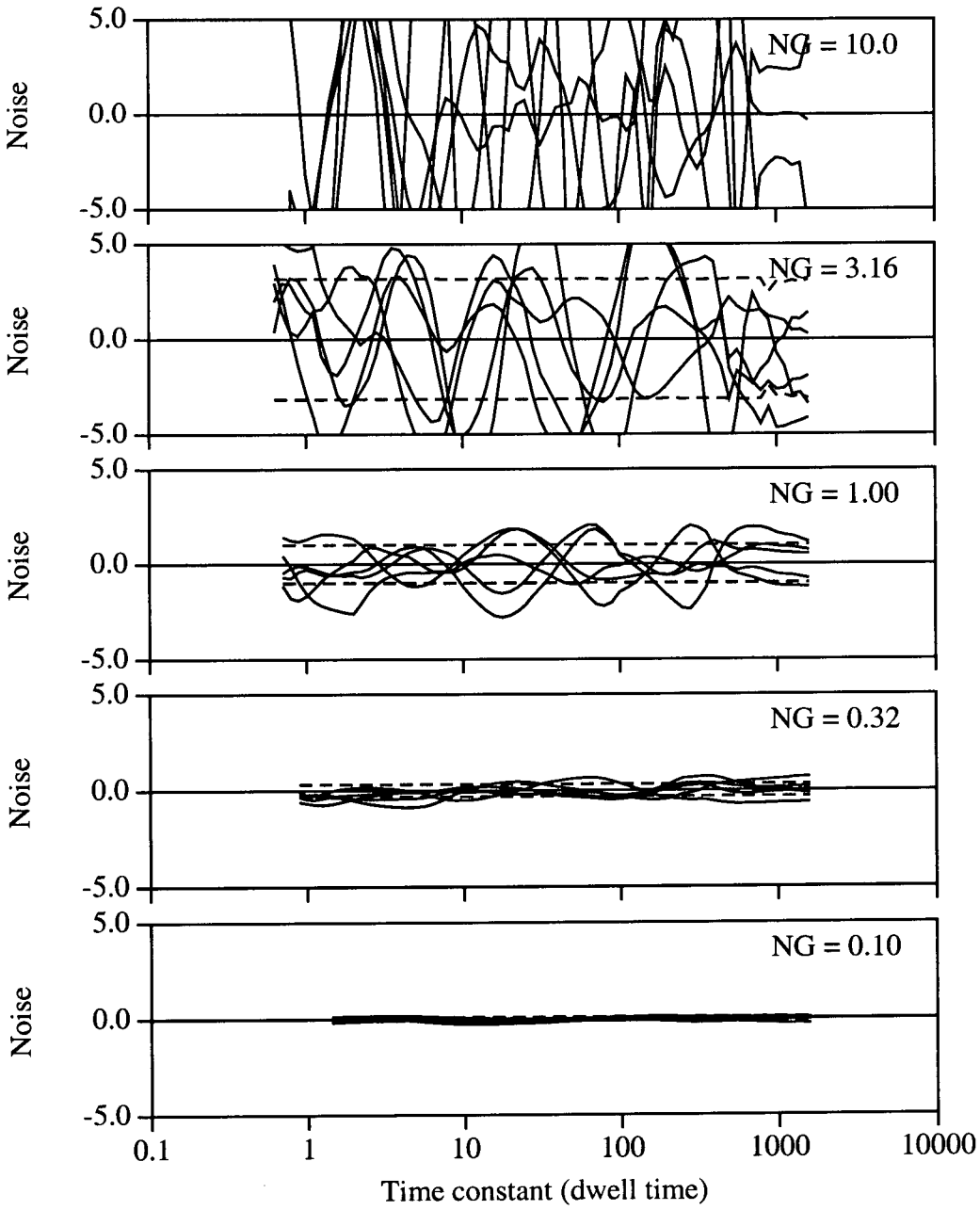

Figure 3.4: Relaxation distributions for pure noise decay curves. The noise was uncorrelated Gaussian with a standard deviation of one. The dashed lines shows calculated standard deviation for each estimation matrix.

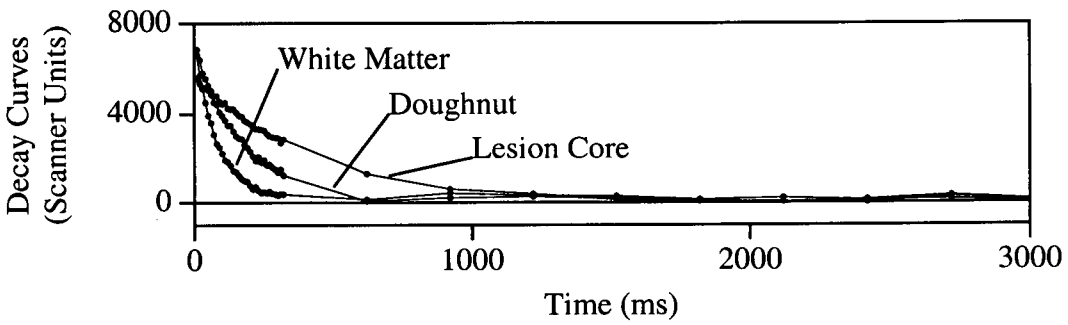

Figure 3.5: Single pixel decay curves from a magnetic resonance measurement. Tissue types examples are white matter, edema and lesion core from a multiple sclerosis patient.

3.5.1 Noise

Figure 3.4 shows the relaxation distributions for five realizations of decay curves consisting only of uncorrelated Gaussian noise with a standard deviation of 1. The first thing to notice is that the noise in the relaxation distribution increases as predicted by the noise gain. It is also important to keep in mind that, while the noise was uncorrelated in the decay curve, it is correlated to adjacent points in the relaxation distribution.

3.5.2 Magnetic Resonance Data

Figure 3.5 shows several decay curves of pixels taken from a series of 48 echoes of a MRI of a multiple sclerosis patient's brain (MacKay *et al* 1994). The T_2 relaxation data were acquired with a 10ms dwell time out to 320ms followed by 16 echoes at 300ms spacing. The strength and decay of the signal contain valuable information about the tissue. The units of the signal are the units used internally by the scanner.

Some of the major tissue types of interest in multiple sclerosis are white matter, white matter with edema and lesion tissue. White matter is present in a healthy brain, while edema and lesions are signs of disease.

Figure 3.6 shows relaxation distributions yielded by applying the estimation matrices

to the decay curves. Each dot of the distribution corresponds to a row of the estimation matrix. The larger the noise gain of the estimate the wider the range of relaxation estimates available. This is a characteristic which shows up in the resolution functions and is due to the fact that the larger the noise gain the more likely there is a feasible resolution function available.

Estimating the noise in the estimates in figure 3.6 can be accomplished in several ways. The first is to estimate the noise in the data and multiply by the noise gain for each estimation matrix. The ideal way to measure the noise in the data is to repeat the measurements a large number of times and calculate mean, standard deviation and covariance. These statistics can then be propagated through the linear matrix of a estimation matrix using standard statistical procedures. Unfortunately, the measurement of the decay curves takes about 30 minutes to complete on a patient, so large numbers of repetitions are impractical.

Another way to measure the noise in the data is to estimate the standard deviation from previous measurements using the same instrument. This is not always reliable because the noise can vary from sample to sample. For example, the noise from a patient is usually at least twice what it is from a test sample. The motion of the patient, including blood flow and breathing, is probably the source of the noise. Unfortunately, these sources of noise may behave more like signal than thermal noise, which can make interpretation of the relaxation distribution more complicated.

Figure 3.5 shows that the signal has decayed to well below the noise before 2s. Therefore, the last few points of the decay curve are pure noise and offer another way to measure the noise. The average standard deviation of the last 6 points of the decay curves is 112 scanner units. However, the measured decay curve is the magnitude of a complex decay. Thus the noise of the last 6 points is Rayleigh distributed. A random variable with a Rayleigh distribution results from squaring a random variable with a

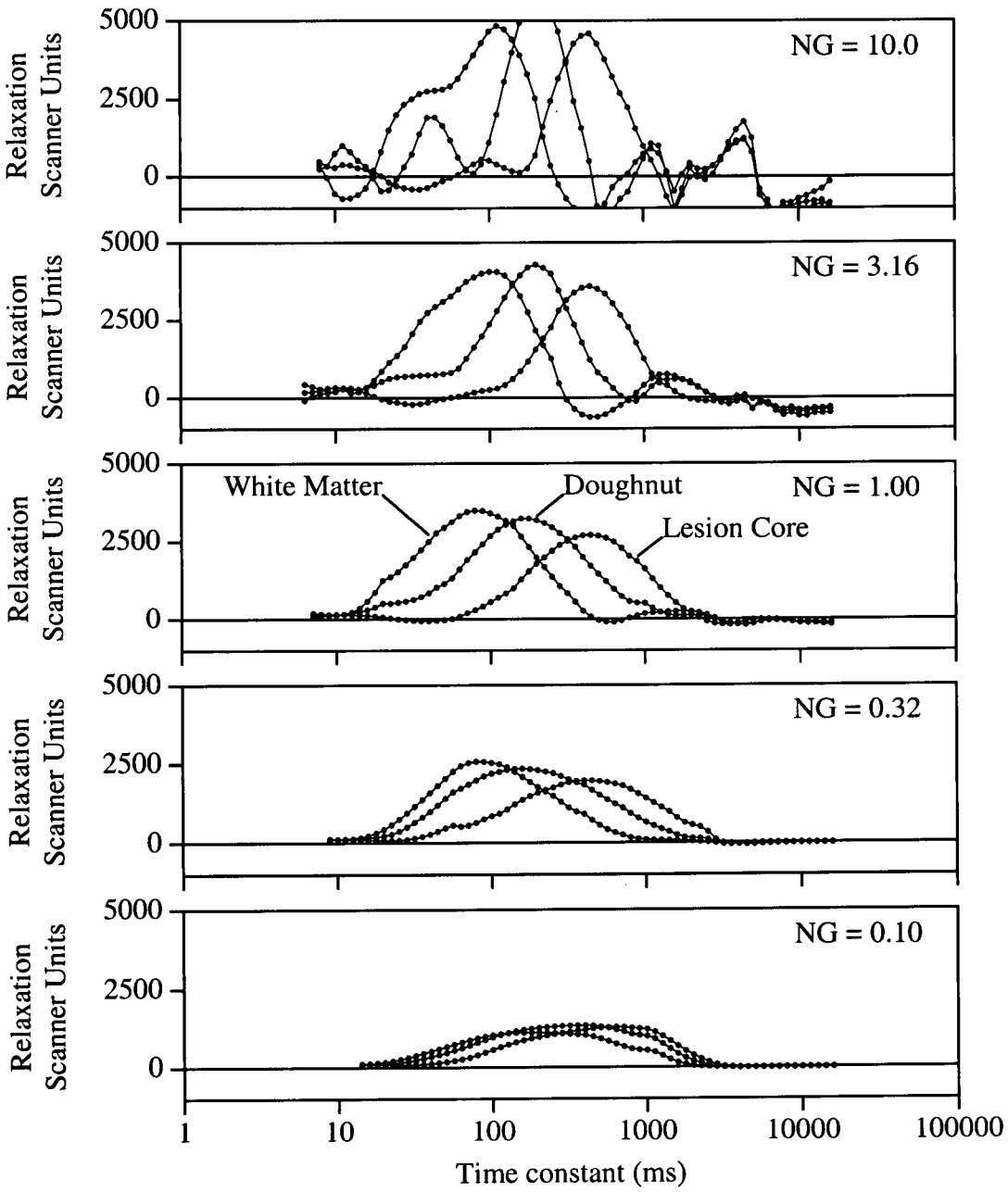

Figure 3.6: Relaxation distributions of the magnetic resonance decays. The estimation matrices used are the same as those used in figures showing the resolutions functions and the PSF's.

Gaussian distribution. At earlier times, when the signal is much larger than the noise, the distribution is, to a very good approximation, Gaussian. The ratio of the standard deviations of the Rayleigh noise to the Gaussian noise is 0.655. Therefore the standard deviation of the noise at early times is 171 (Gudbjartsson and Patz 1995).

3.6 Conservation of Information

As explained in chapter 1, it is a desirable property of an estimation matrix to be invertible. If it is, then the resulting estimate of the relaxation distribution, in combination with the resolution functions and noise of the estimate, conveys all the information about the unknown models that is given by the data and associated data functions and noise statistics. In other words, an estimate is a complete summary of all the models that fit the data. The five estimation matrices used in this paper were decomposed using singular value decomposition (Press *et al* 1992). All five matrices were invertible with a ratio of the largest to smallest eigenvalues ranging from 10^8 to 10^{10}.

3.7 Other Applications

Decaying sinusoids are a common problem in inverse theory. A decaying sinusoid can be handled if it is band pass filtered at the particular bandwidth of interest and then the magnitude of the decay curve taken. The relaxation distribution of the magnitude decay curve can then be calculated. The MR decay curves are generated in this way.

Any row of a estimation matrix, since it corresponds to a resolution function, can be applied to a time series in the same way as digital filters. The "relaxation" digital filter would be useful for applications such as ultrasound and radar. Using quadrature detection, it would be possible to measure the magnitude of a reflection off an interface. If the signal from an interface of interest oscillated for a while after the initial sound wave

had passed, the decay time of the oscillation could tell us more about the interface. Applying various relaxation digital filters to the ultrasound time series would allow characterization of the interfaces.

In addition to estimating the relaxation distribution of a magnetic resonance decay curve, there are other applications where estimating relaxation distributions of decay curves may be useful. These include photoluminescence of biological and other types of samples, electrical signals radiating from ore bodies in geophysical exploration, and acoustic, electrical and electromagnetic decay processes. While the case considered handled data acquired at evenly spaced discrete points in time, it is possible to design an estimation matrix which handles data integrated over a window or which is unevenly spaced in time by modifying the data functions.

3.8 Conclusions

Using the windowed DFT as a guide, I have demonstrated a practical stable algorithm for estimating the relaxation distribution of a decay curve. The algorithm has optimal linear resolution and handles noise in a robust manner, which are both desirable properties of the DFT. When an estimated relaxation distribution is combined with a knowledge of the linear resolution of the relaxation distribution, via the resolution functions or PSF's, and a measure of the noise, it gives a complete description of what is known and not known about the unknown model from the data.

The technique introduced for estimating the relaxation distribution is heavily dependent on the data function of equation 3.10. If equation 3.10 could be calculated so easily, if not rigorously, from equation 3.2, given the strong interest in finding a robust way of estimating relaxation distribution from multiexponential decays, why was it not used previously? Part of the problem may have been due to the conceptual and calculation

difficulties arising from a data function with infinite area. However, I suspect the main reason was the wide spread belief that, for an estimate to be useful, it must fit the data. If you are not interested in linear resolution equation 3.10 offers little benefit over equation 3.2.

References

Backus G and Gilbert F 1967 Numerical application of a formalism for geophysical inverse problems *Geophys. J. R. Astr. Soc.* **13** 247–276

Backus G and Gilbert F 1968 The resolving power of gross earth data *Geophys. J. R. Astr. Soc.* **16** 169–205

Backus G and Gilbert F 1970 Uniqueness in the inversion of inaccurate gross Earth data *Phil. Trans. R. Soc. Lond. Ser. A* **266** 123–192

Bellman R E, Kalaba R E and Lockett J 1966 *Numerical Inversion of the Laplace Transform* (Amsterdam: Elsevier)

Cohen-Tannoudji C, Diu B, Laloë F 1977 *Quantum Mechanics* (New York: John Wiley & Sons)

Essah W A and Delves L M 1988 On the numerical inversion of the Laplace Transform *Inverse Problems* **4** 705–724

Fletcher R 1987 *Practical Methods of Optimization* (Toronto: John Wiley & Sons)

Gardner DG, Gardner JC, Laush G and Meinke WW 1959 Method for the analysis of multicomponent exponentials decay curves *J Chem. Phys.* **31** 978–986

Gudbjartsson J and Patz S 1995 The Rician distribution of noise MRI data *Magn. Res. Med.* **34** 910–914

Haario H and Somersalo E 1987 The Backus-Gilbert method revisted: Background,

implementation and examples *Numerical Functional Analysis and Optimization* **9** 917–943

Harris F J 1978 On the use of windows for harmonic analysis with the discrete fourier transform *Proc. IEEE* **66** 51–83

Huestis S P 1987 Construction of non-negative resolving kernels in Backus-Gilbert Theory *Geophys. J. R. Astr. Soc.* **90** 495–500

Istratov A A, Vyvenko O F 1999 Exponential analysis in physical phenomena *Rev Sci Instr* **70** 1233-1257

MacKay A, Whittall K, Adler J, Li D, Paty D, Graeb D 1994 *In Vivo* Visualization of Myelin Water in Brain by Magnetic Resonance *Magnetic Resonance in Medicine* **31** 673–677

Oldenburg D W 1976 Calculation of the Fourier Transform by the Backus and Gilbert method *Geophys. J. Roy. Astr.* **44** 413–431

Parker R L 1994 *Geophysical inverse theory* Princeton: Princeton University Press

Press W H, Teukolsky S A, Vetterling W T, Flannery B P 1992 *Numerical Recipes in C* (Cambridge: Cambridge)

Provencher S W 1976 A fourier method for the analysis of exponential decay curves *Biophysical Journal* **16** 27–41

Smith M R and Nichols S T 1983 Improved Resolution in the analysis of multicomponent exponential signals *Nuclear Instruments and Methods* **205** 479–483

Štěpánek, P 1993 Data Analysis in Dynamic Light Scattering *Dynamic Light Scattering* 177–241 (Oxford: Clarendon Press)

Tarantola A 1987 *Inverse Problem Theory* (New York: Elsevier)

Twomey S 1977 *Introduction to the Mathematics of Inversion of Remote Sensing and Indirect Data* (Amsterdam: Elsevier)

Whittall K P and MacKay A L 1989 Quantitative interpretation of NMR relaxation data

Journal of Magnetic Resonance **84** 134–152

Whittall K P, Bronskill M J and Henkelman R M 1991 Investigation of analysis techniques for complicated NMR relaxation data *Journal of Magnetic Resonance* **95** 221–234

Wiener N 1949 *Extrapolation, Interpolation and Smoothing of Stationary Times Series* (New York: John Wiley & Sons)

Chapter 4

MRI Peak Time Constant Imaging in Multiple Sclerosis

Summary

The objective of this study was to measure the relaxation of *in vivo* brain tissue for normal volunteers and multiple sclerosis (MS) patients using new techniques. Small voxels, with pixels of 0.86mm by 1.7mm and slice thickness of 3mm, were used to reduce the effects of partial volume. A 48 echo single slice sequence which had echoes from 10ms to 5120ms increased the range of T2 time measurable. And a novel technique was used to analyse the multiexponential decay curves which reliably measured the dominant time constant.

A phantom was scanned to characterize the performance of the extended MRI sequence and novel analysis techniques. Two normal volunteers and four MS patients were scanned once each. One of the MS patients was scanned serially 15 times over a 24 week period after a new lesion was detected.

The phantom's monoexponential T2 relaxation times, when near 100ms, were measurable to 1% accuracy. When estimating the relaxation distribution from the MRI decay curves, the novel technique was very robust in the presence of noise allowing the use of small voxels. However, the technique lost sensitivity below time constants of 50ms and had little sensitivity to time constants below 20ms. With this technique most of the tissues in the normal volunteer and the MS patient showed a monoexponential decay to a very good approximation including the new MS lesion. Exceptions to this rule included "dirty white matter" and the internal capsule. Some lesion voxels had T2 time constants

exceeding 400ms but none exceeded 500ms. The weekly scans of the MS patient showed changes in T1 and T2 parameters over time reflecting changes in lesion pathology.

The technique described proved to be particularly useful in estimating the T2 time constant of nearly monoexponential decay curves for each voxel. Images of T2 time constants in brain had good signal to noise providing what appeared to be pathologically meaningful information. This technique will be useful in providing new insight not only into MS pathology but also in the assessment of other pathologies including brain tumours.

4.1 Introduction

Magnetic resonance has become a major tool in the diagnosis, treatment and monitoring of many diseases including multiple sclerosis (MS). (Filippi M *et al* 1998; Paty and Ebers 1998, MacKay *et al* 1994). Its ability to detect and monitor MS lesions over time with little risk to the patient has become a great asset. While they are able to detect lesions reliably, conventional MR techniques have limited ability to differentiate between various types of pathology within a lesion and various types of lesions. The diagnosis and treatment of a host of other pathologies, such as brain tumors, would also benefit from a better ability to discriminate between pathologies *in vivo* using MRI.

Several studies have used magnetic resonance relaxation (MRR) to study MS lesions *in vivo* (Larsson, Barker et al 1998) with some success. They have reported numerous cases of both monoexponential and biexponential decay curves to lesions. However, as the studies mention, the biexponential nature could be due to partial volume effects. The volume over which the decay was averaged in this study was at least 50 times smaller than previously referred studies, minimizing partial volume effects.

There were two purposes to this study. The first was to measure the MRR

characteristics of MS lesions for much smaller voxel sizes than previously used. The second was to provide an initial assessment of two new tools for MRR of MS lesions.

The first tool was a modified MRR sequence. The sequence acquired 48 echoes for a single slice using a Carr-Purcell-Meiboomm-Gill (CPMG) pulse sequence (Mackay et al 1994). The echo times ranged from 10ms to 5.12s so a wide range of T2 time constants could be measured.

The second tool was a new way to estimate the relaxation distribution of a voxel from its decay curve. The method was very robust under noise and thus allowed for the use of small voxel volumes. It was implemented with an estimation matrix where the coefficients of the matrix have been specially chosen to maximize the linear resolution for a given level of noise. The new method used for estimating the relaxation distribution is particularly good at estimating the time constant of the monoexponential associated with the peak of the distribution. Images composed of the value of the peak T2 time constant for each voxel appear to be particularly useful.

Much effort has gone into MR spectroscopy (MRS) as a means to study the natural history of MS lesions *in vivo* (Arnold *et al* 1998). Although MRR and MRS measure very different parameters of tissue, one major advantage of MRR over MRS is that the smallest voxel sizes that provide reasonable signal to noise in MRS are about 800mm^3. This is about 180 times larger than a 4.3mm^3 voxel, the smallest used in the MRR study described herein. In MS applications, this small sample size means that the variation of the relaxation within many lesions can be measured. The reason relaxation can handle much smaller voxels than spectroscopy is that relaxation measures the main water signal while spectroscopy measures the much weaker signal from metabolites.

4.2 Method

For the MRR sequence the first 32 echoes were spaced 10ms apart with the next 16 spaced 300ms apart. The last echo was at 5.12s and the TR was 7.0s. The combination of many early echoes with several late ones allowed a large range of relaxation times to be examined. As a phantom study showed, the change in echo time from 10ms to 300ms during the sequence did not effect the decay curves in most circumstances.

Images were acquired at a range of voxel sizes and shapes. In this study, the smallest voxel size which gave good signal to noise was $4.3mm^3$. The ideal shape for the voxel requires a trade off between small pixel dimensions on an image and a homogeneous sampling of tissue in a voxel. In standard imaging sequences, to achieve small pixel dimensions and reasonable voxel volumes, slice thicknesses many times larger than the pixel dimensions are commonly used. But these pencil shape voxels were more likely to have heterogeneous tissues than a cubic voxel of the same volume unless the long dimension of the pencil shaped voxel happened to lie along elongated anatomical structures or pathology. For this study we acquired images at a variety of voxel shapes and sizes. Pixels of the dimensions $0.86 \times 1.7mm$ with a 3mm thick slice seemed to be a good compromise for this study but future applications may benefit from more cubic voxels.

A wide variety of methods are used to estimate the relaxation distribution from a decay curve (Istratov and Vyveno 1999, Whittall et al 1991). However, all these methods differ in two crucial but related ways from the standard method of estimating the frequency distribution in MRS. In MRS, the discrete Fourier transform (DFT), usually windowed, is the most common way of estimating a frequency spectrum. The DFT is usually implemented as the fast Fourier transform (FFT) because of the FFT's speed. The purpose of adding the window to the DFT transform is to improve the line shape

in the resulting estimates and no regard is given to whether the resulting estimate fits the data or not. Since the DFT is a linear transform, the method of using a window to improve the line shape is equivalent to improving the linear resolution of the transform.

Non windowed DFT's also have linear resolution in addition to fitting the data. Since the process of interpreting windowed and non windowed DFT's seem to be the same, I strongly suspect that interpreters are relying solely on the linear resolution for interpretation.

For this study, the method selected to estimate the relaxation distribution was the estimation matrix described in chapter 3. As with the windowed DFT, its goal is good linear resolution with no requirement that the estimate reproduce the data. The estimation matrix was applied with a matrix multiplication of the form

$$m_i^e = \sum_j a_{ij} d_j \tag{4.1}$$

where d_j are the 48 data points from the decay curve and m_i^e is the estimate of the relaxation distribution. Because the estimation matrix was linear, the sum of several decay curves was also the sum of the relaxation distributions of each curve transformed individually.

Figure 4.1a shows the estimates of the relaxation distributions of a range of monoexponential decay curves. Since all the monoexponential decays had a value of one at $t = 0$ the resulting estimates can be referred to as point spread functions (PSF's). As can be seen from the PSF's, for the data acquired, the linear resolution is best for monoexponentials with time constants between 50 and 1000ms. Above and below these time constants, the resolution is poorer and the location of the peaks is shifted in toward the intermediate time constants. The wide PSF's are an unavoidable consequence of how little information the data provides about the relaxation distribution.

Another characteristic of the chosen estimation matrix that is in common with the

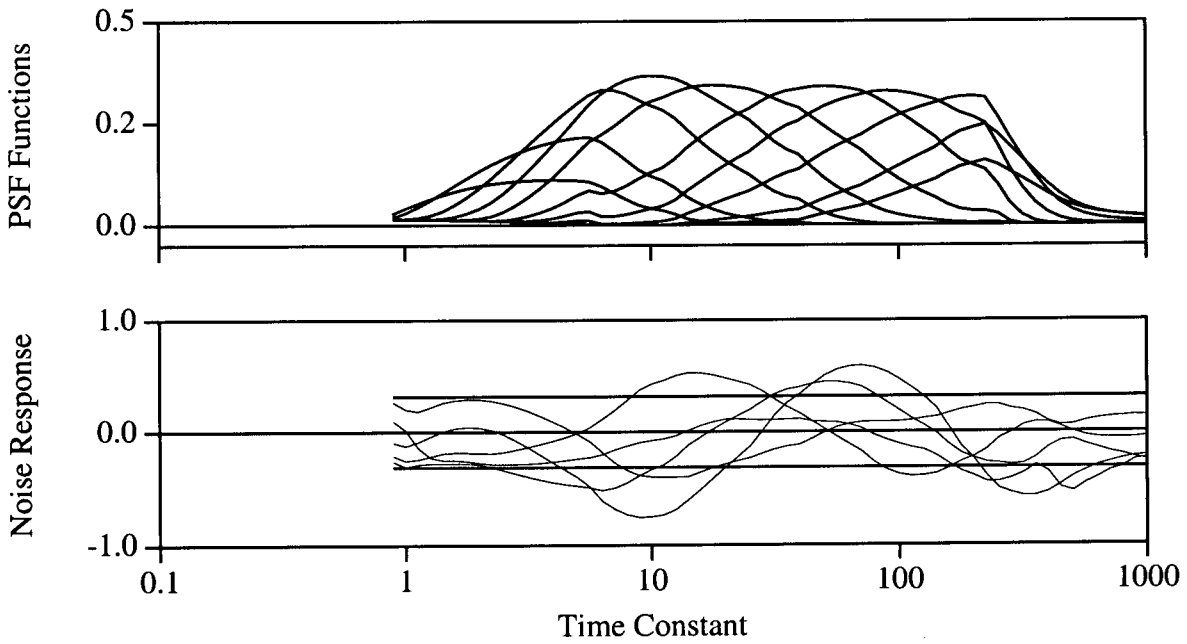

Figure 4.1: (a) Point Spread Functions: Estimates of the relaxation distributions for monoexponential decays with time constants equal to 10, 20, 50, 100, 200, 500 and 1000, 2000, 5000, 10000ms for relaxation distribution transformation (b) Estimates, generated by the transform matrix, of the relaxation distribution of five decay curves consisting purely of uncorrelated Gaussian noise with a standard deviation of one. The solid lines indicate the predicted standard deviation of the noise.

windowed DFT, is that, provided that all data points have noise with the same standard deviation, the points in the estimate all have noise of the same standard deviation. Figure 4.1b shows five realizations of decays with no signal but only uncorrelated Gaussian noise with a standard deviation of one.

The ratio of the standard deviation of the data to the standard deviation of the noise at a point in the relaxation distribution is defined as the noise gain. It can be predicted by calculating the square root of the sum of the squares of the coefficients in the row of the estimation matrix corresponding to the desired time constant. The noise gain for each row of the estimation matrix is plotted as two solid lines in Figure 4.1b. It has the desirable property of being the same for every time constant and has a value of 0.316. As with the windowed DFT estimation, having the same standard deviation of noise across the relaxation distribution makes the distribution easier to interpret.

While an estimation matrix which produced noise with a lower standard deviation could have been chosen, it would have reduced the resolution of the PSF's. Conversely, an estimation matrix which produced noise with a higher standard deviation would have produced PSF's with higher resolution. The estimation matrix chosen had what was felt to be the best tradeoff between the linear resolution and noise for the data acquired.

The estimates of the frequency distributions generated by the windowed DFT are used to determine the frequency of various metabolites as well as identifying the metabolites. Similarly, the location of the peaks in an estimate of a relaxation distribution can be instructive about the tissue which generated the corresponding decay curves. While the location of a peak can be read off a plot of an estimate, a more accurate way to determine it is to fit the relaxation distribution of a monoexponential to points near the peak of the relaxation distribution. As an example, Figure 4.4 shows the calculated relaxation distributions as a dotted line with one dot per row of the estimation matrix. Fitted monoexponentials are shown as a smooth line.

By testing various fitting algorithms I found only three points near the peak of the relaxation distribution, the peak point and the two points immediately adjacent, need to be fitted to get a good estimate of the monoexponential's amplitude and time constant. By only fitting to the points at the peak, effects of other exponentials are minimized. The estimation matrix used had a time constant spacing of 16 points per decade so the three points fitted would only span a factor of 0.125 in time constant as can be seen from Figure 4.4. Since only peak information is used in calculating the monoexponential, it is referred to as the peak monoexponential. The peak monoexponentials parameters are referred to as the peak amplitude and peak T2 time constant.

How well a relaxation distribution is approximated by a monoexponential fit to its peak can be measured by calculating the area between original distribution and the fitted monoexponential distribution. It should be kept in mind when calculating the areas of relaxation distributions that the metric of the horizontal is the natural logarithm of the time constant.

4.3 Results

A phantom, a normal volunteer and an MS patient were scanned with the 48 echo sequence to assess the information provided by the decay curves and relaxation distributions. The peak of each distribution was fitted with a monoexponential and the logarithm of the corresponding time constants displayed as an image. In additional, the relaxation distribution for selected pixels were plotted, along with their monoexponential peak fits.

4.3.1 Phantom Study

The phantom used was a 0.5l water bottle filled with tap water. It contained a cylindrical glass tube with 1% agarose doped with 2mM NCl giving a predicted T2 of 100ms and a T1 of 600ms (Christofferson *et al* 1991). Figure 4.2 shows a peak amplitude image of a cross section through the phantom. Figure 4.3 shows a peak time constant image of a cross section through the same phantom. The intensity of the image corresponds to the logarithm of the time constant. The outer ring contains the tap water and the inner the agarose. The uniform intensity in the agarose and the mostly uniform intensity in the tap water are an indication of the precision of the methods used to determine the peak time constant.

There is an artifact at the top of the tap water in the image. This artifact shows up in the decay curves of the corresponding voxels. The artifact is characterized by low time constant during the echoes spaced at 10ms and abruptly increasing to a new value when the echo spacing increasing to 300ms. The artifact seems to be due to 180 degree pulses in the MRR sequence which are not exactly 180 degrees at all regions of the slice. It is likely due to a B_0 or B_1 inhomogeneity at the water/air interface found at the edge of the bottle. Similar artifacts may occur in regions of the human brain where there are air/tissue interfaces and need to be watched for. The decay curves of the human subjects described herein were examined for indications of similar problems but none were found.

Figure 4.4(a) shows the decays curves for the 4.3mm^3 voxel indicated in figure 4.3. The dots show the measured value and the continuous lines show the monoexponentials fitted to the peaks of the relaxation distributions. The relaxation distribution in figure 4.4 (b) shows the estimates of the relaxation distributions as dots. The continuous line through the dots shows the relaxation distribution of the peak monoexponentials. As is clearly evident, the monoexponential parameters, as determined by the peak time constant,

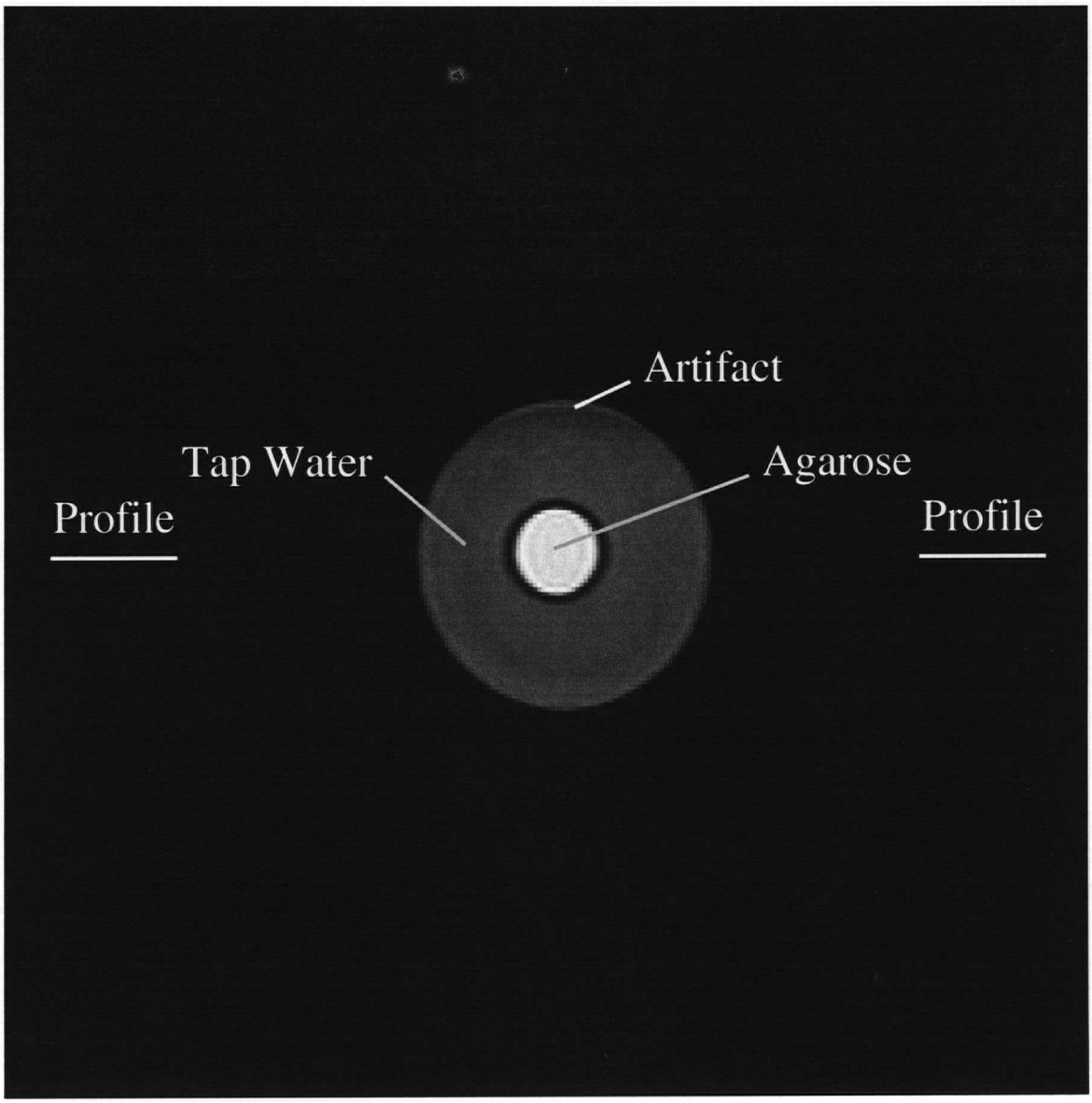

Figure 4.2: Peak amplitude constant image of the tap water-agarose phantom. The water has less signal because of T1 effects.

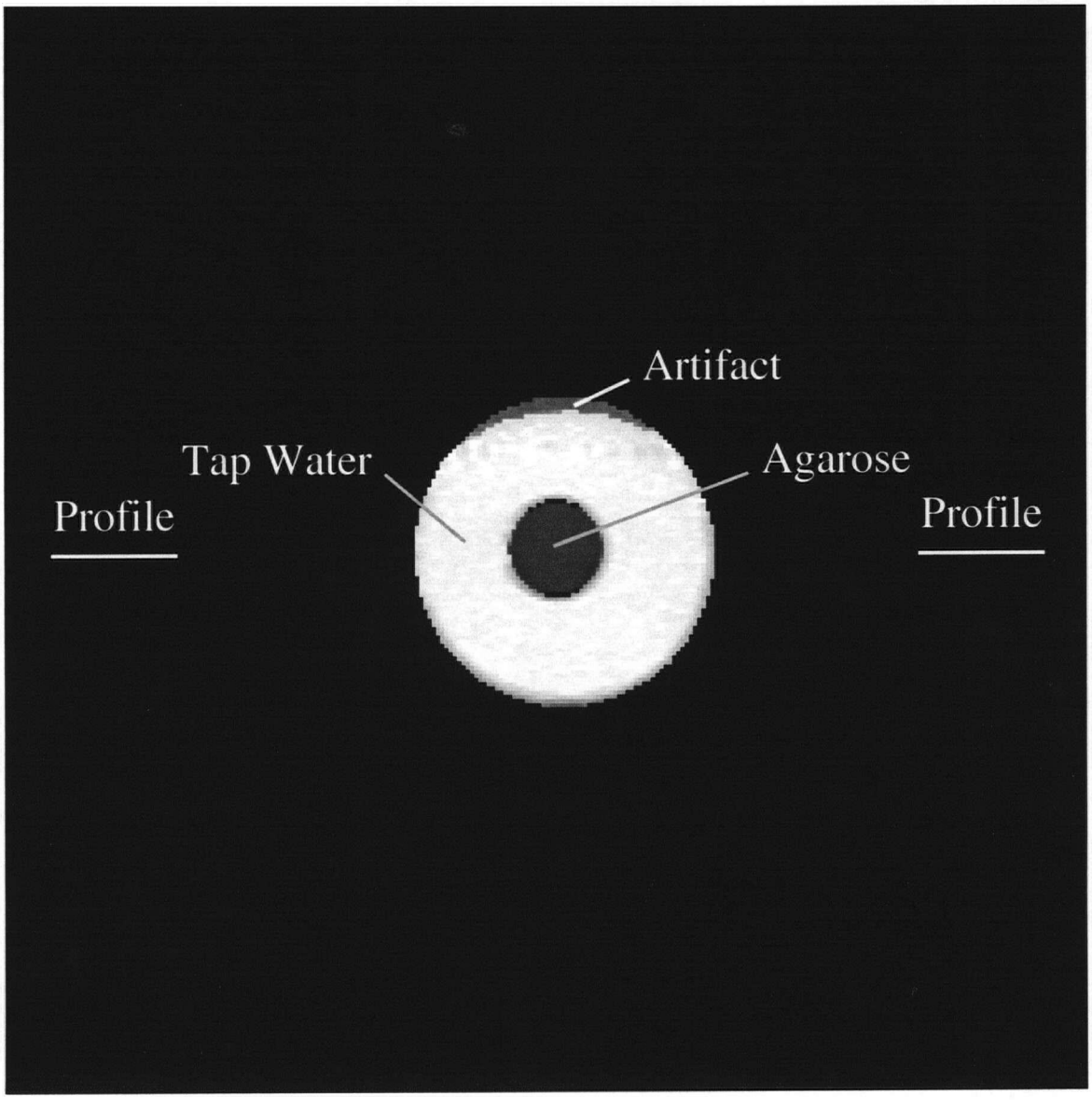

Figure 4.3: Peak time constant image of the tap water-agarose phantom. The outer diameter of the phantom is 55mm. The time constant for the tap water and agarose, as measured from the peak time constant, are 1875ms and 99.5ms respectively. The contrast of the image is between 10ms and 2500ms with the intensity, as with all peak time images, proportional to the logarithm of the time constant.

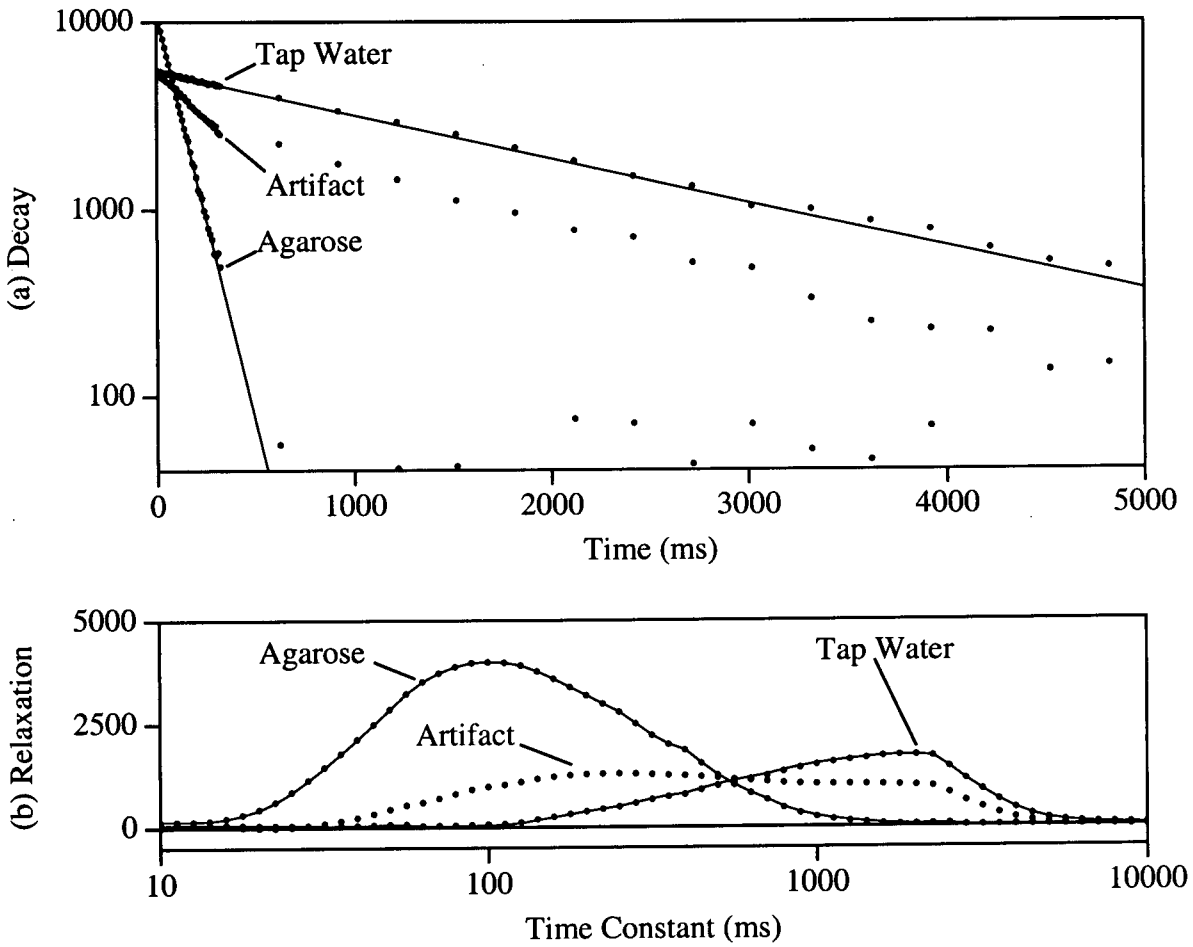

Figure 4.4: Decay curves and relaxation distributions for agarose and water phantom

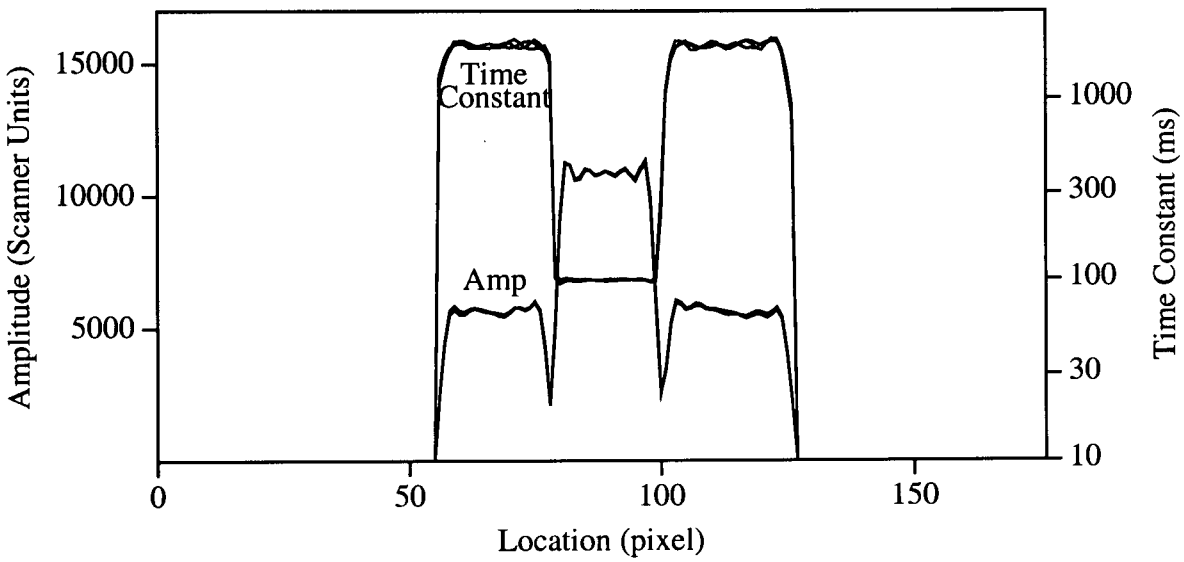

Figure 4.5: Repeated profiles (3 times) through the center of the tap water-agarose phantom along with the peak amplitude and time constant images.

model the decay curves and relaxation distributions extremely well.

This example also clearly demonstrates the robustness of the peak monoexponential in approximating a relaxation distribution. The noise after 500ms in the time domain is all positive because the decay curve is a magnitude measurement. This can be confusing to methods which fit the data with a monoexponential decay because the positive noise may be mistaken for a second monoexponential decay curve with a very large time constant.

The MRR sequence also performed very well. This is indicated by measured points lying on the solid line. If the sequence had poor 180 degree pulses the points might lie alternately above and below the line. This problem is common in CPMG sequences.

Figure 4.5 shows three repeats of a profile between the profile lines indicated on figures 4.2 and 4.3. The scan was repeated three times to get a measure of the repeatability of the peak time constants. As clearly evident from the profile, the measured peak time constant of the agarose was very repeatable while the measurement of the tap water peak time constant was not quite as good. The agarose had a peak time constant of

99.5ms with a standard deviation of 0.4%. The tap water had a peak time constant of 1875ms with a standard deviation of 1.9%. The agarose had a peak time constant within the range of brain tissue while the tap water has a peak time constant which was much longer but not far from cerebrospinal fluid (CSF). The peak amplitude profile is also highly reproducible.

The time constant of the tap water is only a factor of 0.37 of the last echo. This put it above the ideal range of peak time constants to be measured with the MRR echo sequence used. From examination of figure 4.5 it can be seen that the variation in tap water peak time constant is correlated between adjacent pixels. This could be due to some combination of truncation artifact and small motions of the agarose tube in the tap water due to gradient vibrations in the scanner.

4.3.2 Normal Volunteers

The 48 echo sequence was performed on two normal volunteers. Both volunteers gave similar results.

Figure 4.6 show a 30ms conventional spin echo image from the 48 echo sequence from one of the normal volunteers. This image is similar of those acquired in a standard clinical scan. Figure 4.7 shows the amplitude of the monoexponential fit to the peak of the relaxation distribution for each voxel of the normal volunteer. Figure 4.8 is a peak time constant image calculated from the same data set as figure 4.6. Figure 4.9 is a profile through figures 4.7 and 4.8. From the profile, peak time constants for white matter range between 70 and 100ms.

The profile more clearly displays the variation over the brain. The peak time constant image and the profile clearly show the noise to be quite low.

The area between the relaxation distribution for each voxel and the peak monoexponential is presented as an image in figure 4.10. The dark area represents a very

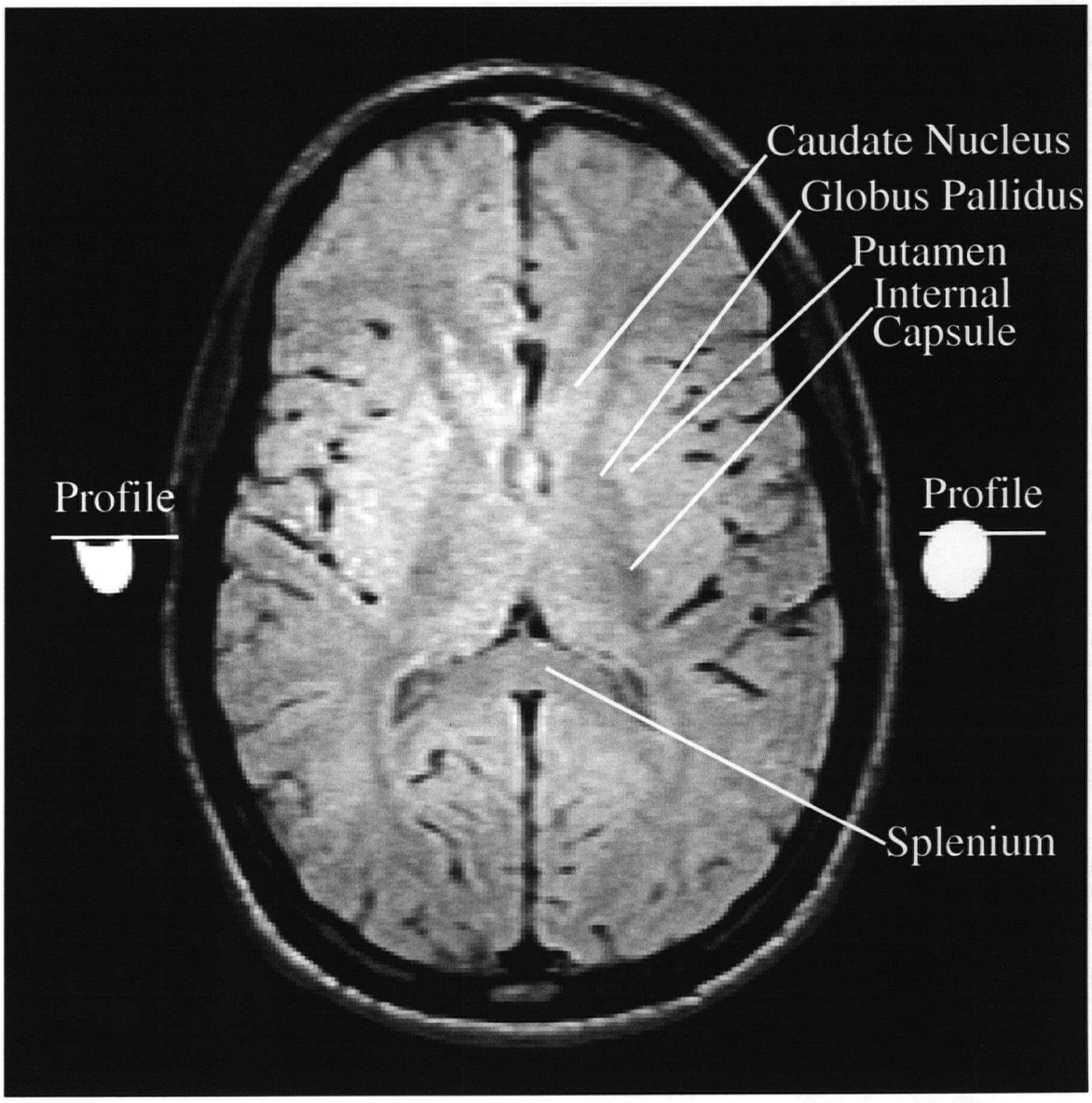

Figure 4.6: Conventional spin echo image at 30ms from the 48 echo sequence for the normal volunteer.

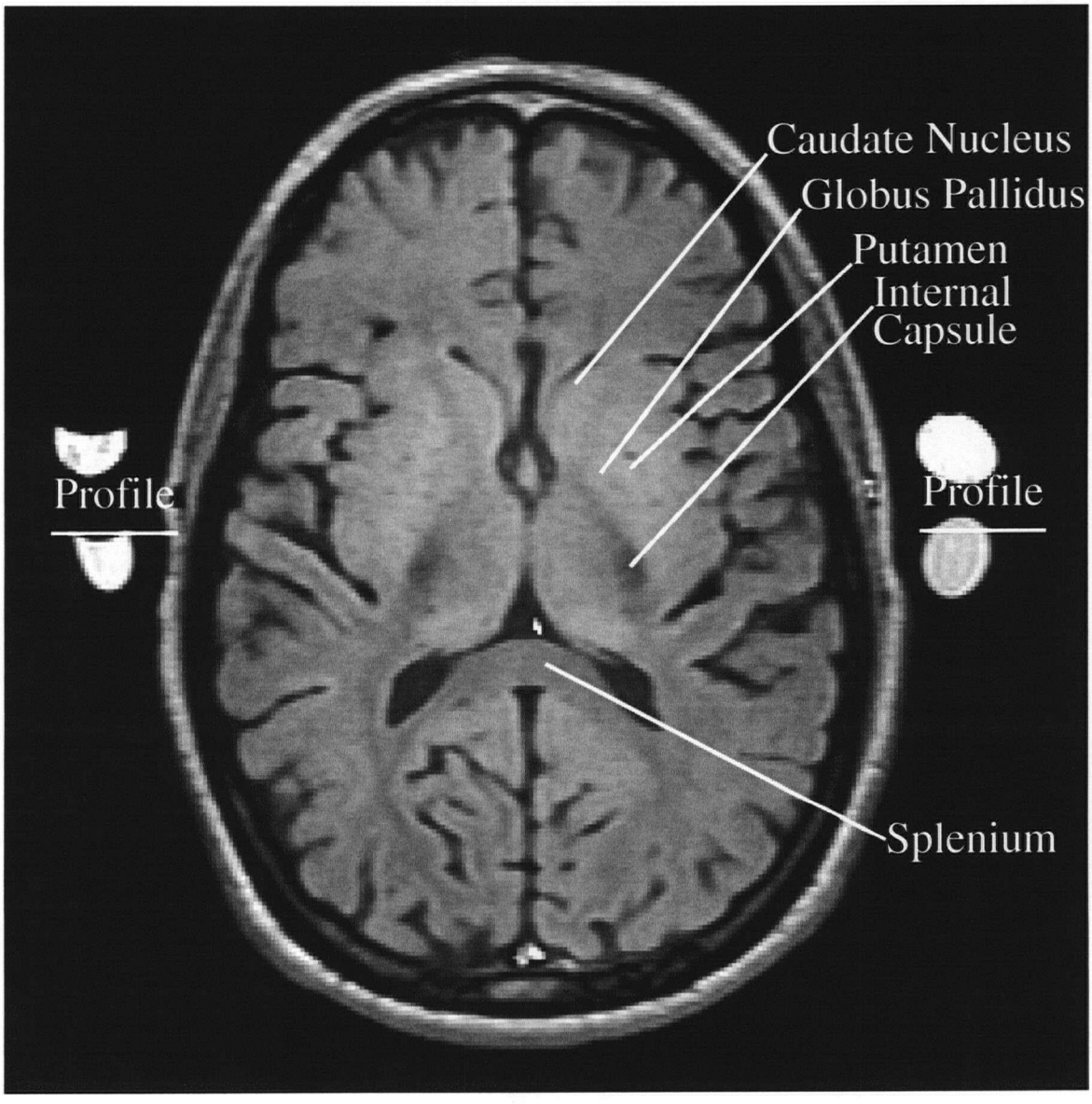

Figure 4.7: Peak amplitude image of the normal volunteer.

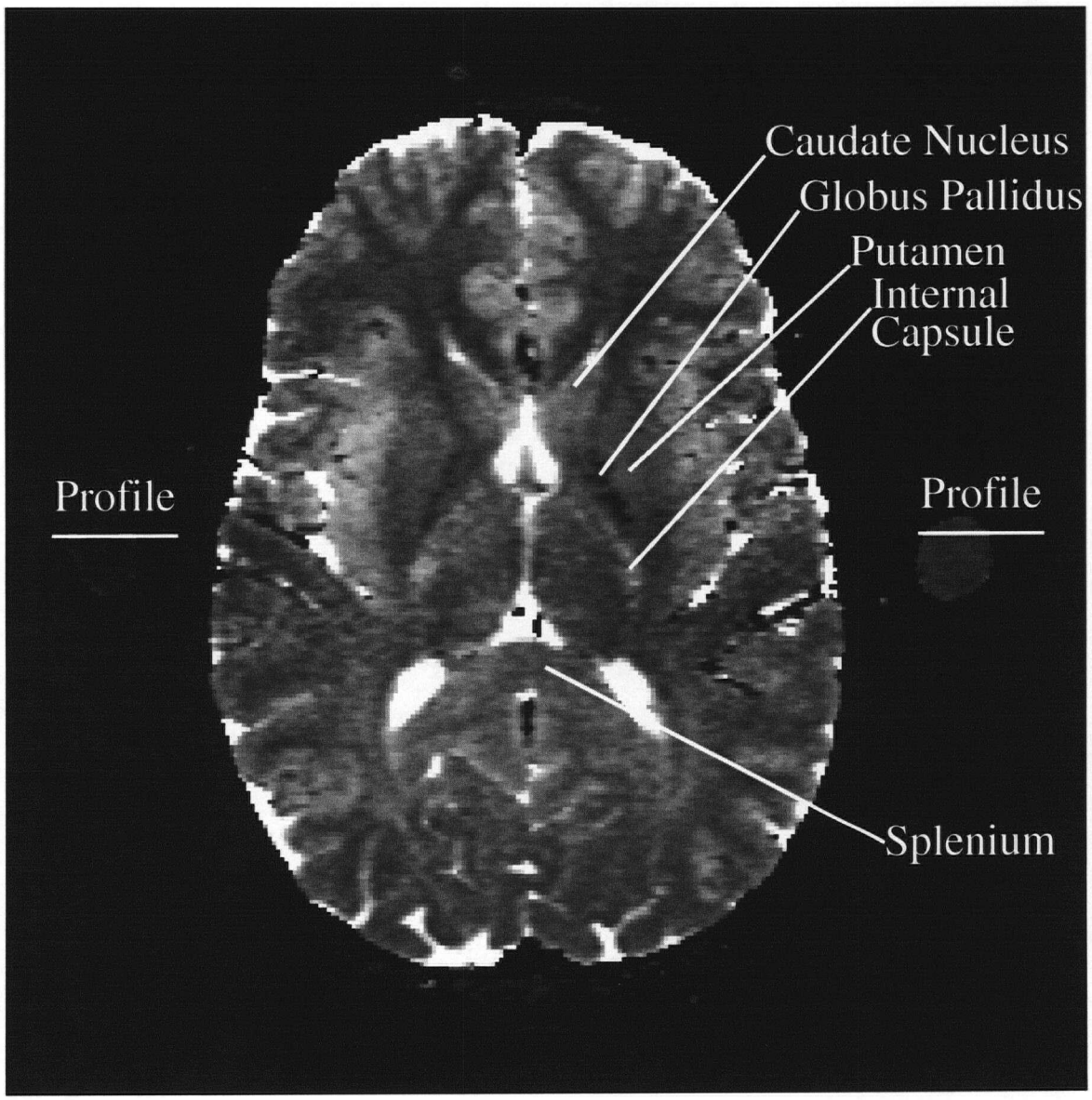

Figure 4.8: Peak time constant image of the normal volunteer with contrast between 50ms and 125ms.

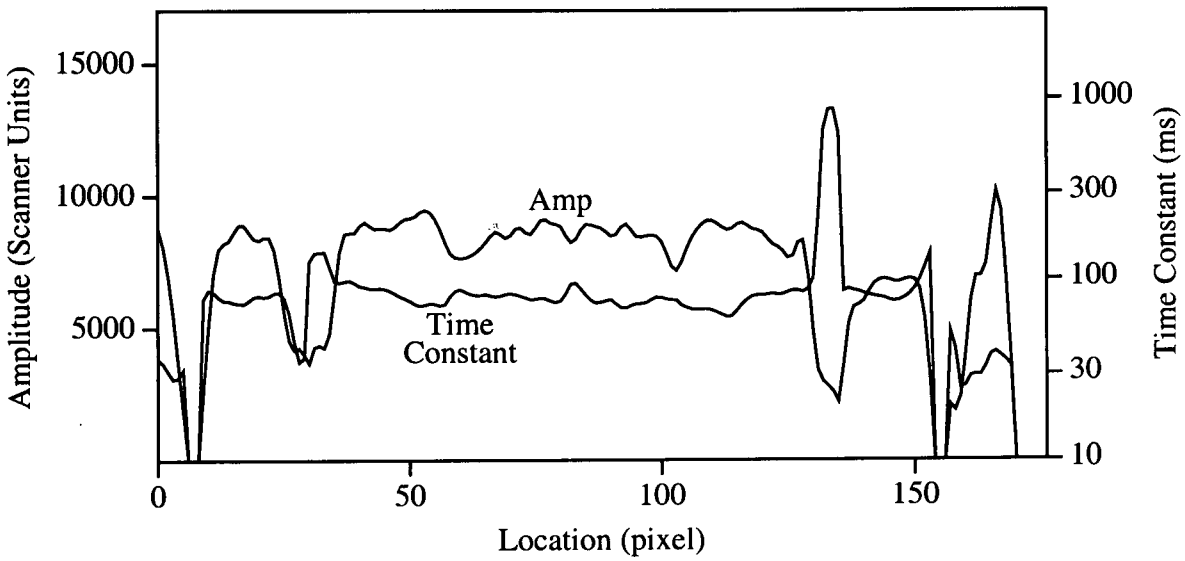

Figure 4.9: Profile of peak amplitude and time constant for the normal volunteer.

good approximation of measured distribution by the peak monoexponential. Surprisingly, most of the voxels in the brain have distributions which are very well approximated by peak monoexponentials. The main exceptions being the CSF, internal capsule and, to a lesser extent, the splenium. The CSF's non-monoexponential nature is easily explained by flow artifacts. The characteristic is also evident in the MS patient considered below.

Figure 4.11 shows the relaxation distribution and peak monoexponential fit for the caudate nucleus which figure 4.10 indicated was well approximated by a peak monoexponential. Examination of the relaxation distribution clearly shows how well the peak monoexponential approximates the measured relaxation distribution.

Figure 4.10 shows that the internal capsule is the structure which is least well approximated by a peak monoexponential. Figure 4.13 shows the measured decay curves and relaxation distribution along with the peak monoexponential. Since the peak monoexponential did not do a good job of approximating the measured relaxation distribution, a biexponential was manually fitted to the relaxation distribution. This is

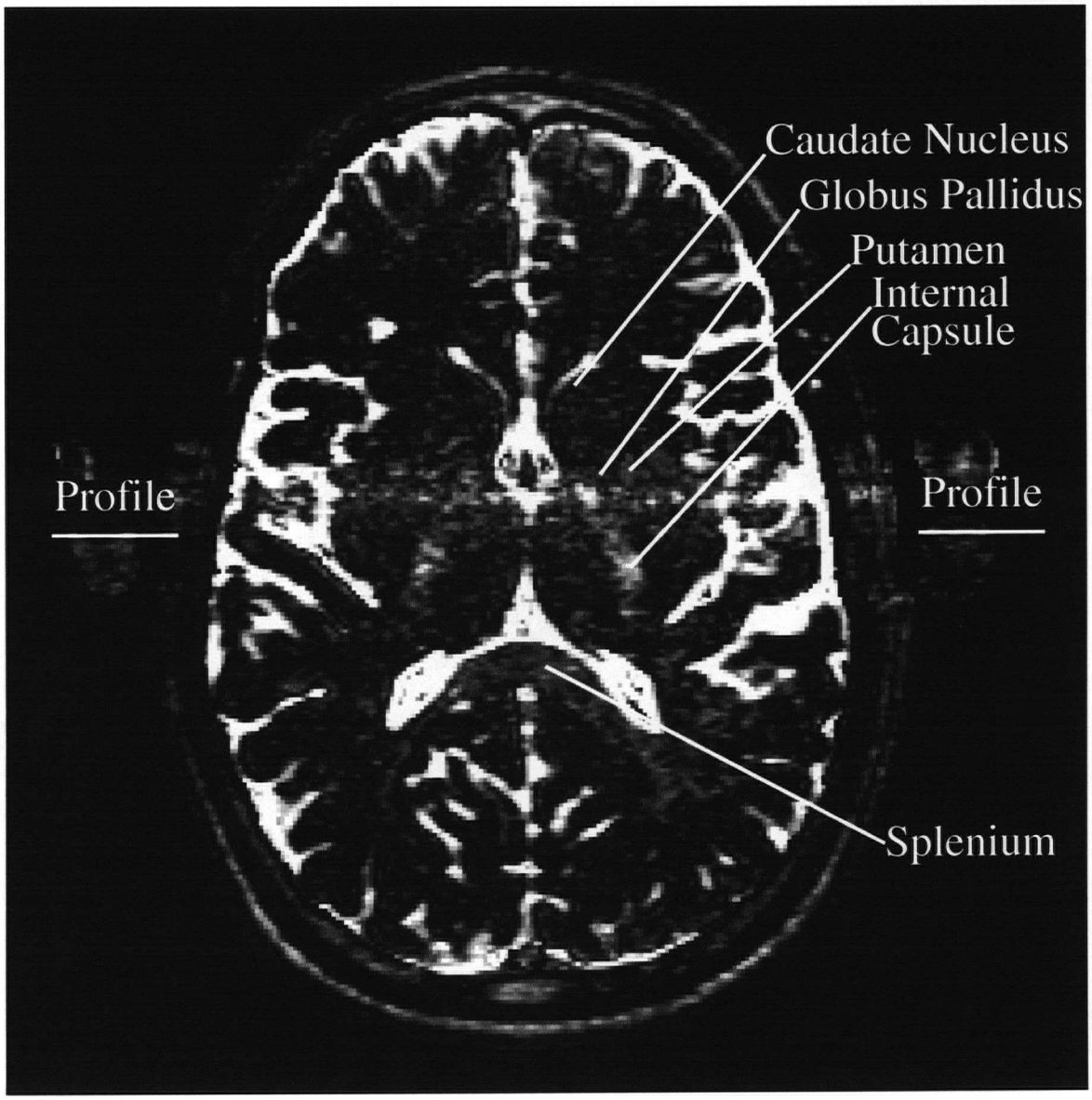

Figure 4.10: Image of how well the relaxation distribution each pixel is approximated by a monoexponential for the normal volunteer. The lower the pixel value, the better the approximation. The parameter imaged is the area between the relaxation distributions for the pixel and its monoexponential peak fit.

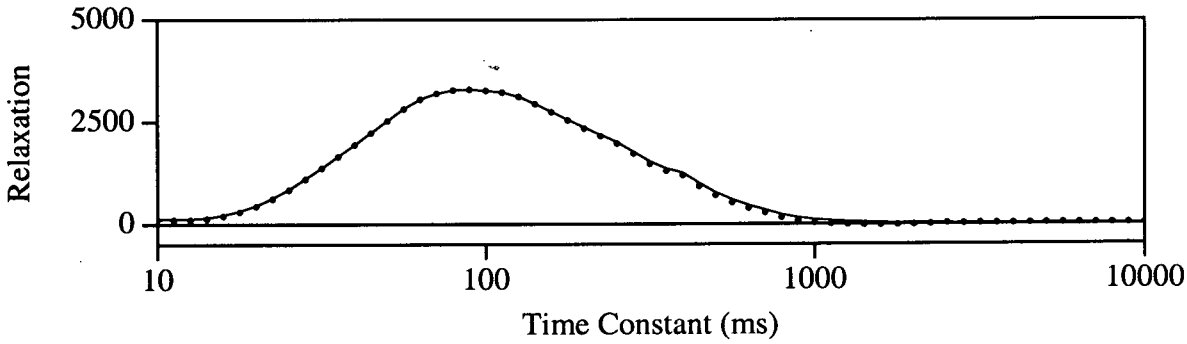

Figure 4.11: Relaxation distributions of the caudate nucleus for the normal volunteer including peak monoexponential.

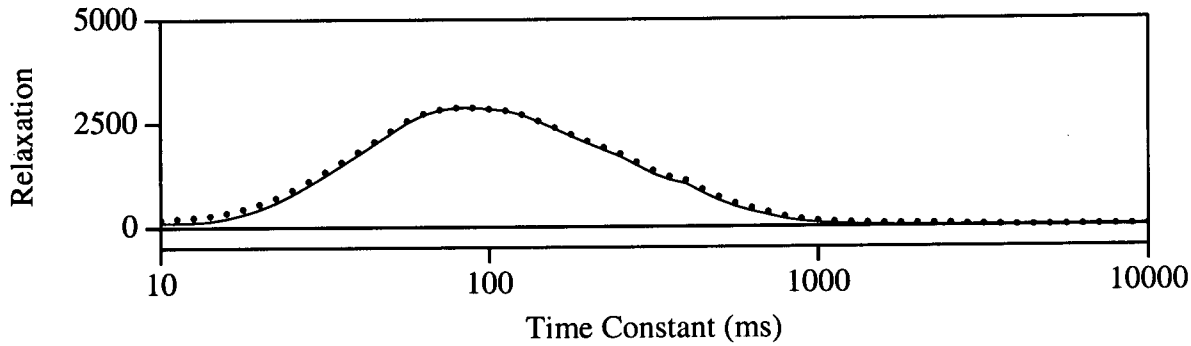

Figure 4.12: Relaxation distribution of the splenium for the normal volunteer including peak monoexponential.

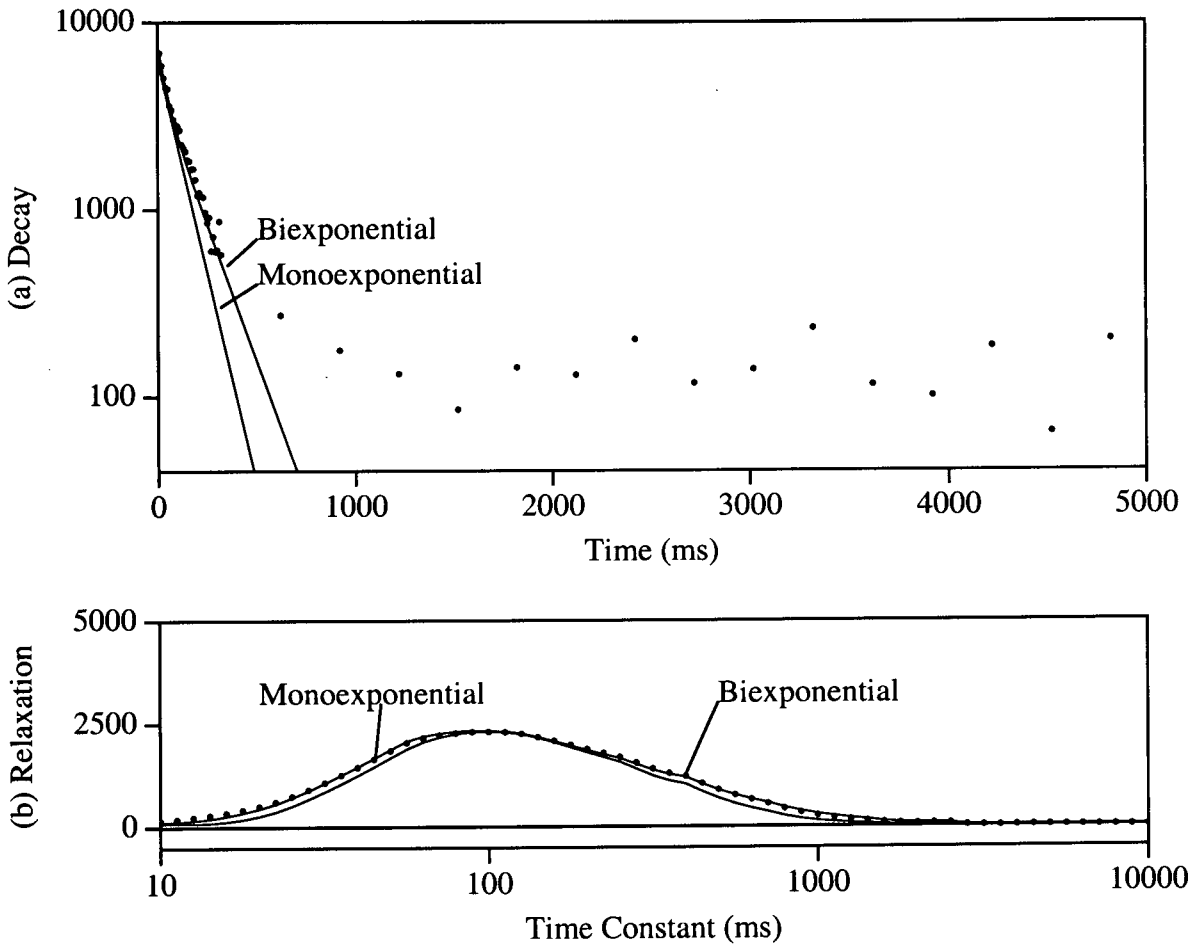

Figure 4.13: Decay curves and relaxation distributions of posterior internal capsule for the normal volunteer including peak time constant and biexponential fit. The biexponential fit has amplitudes of 4439 and 3217 and time constants of 148.9 and 39.3ms respectively while the monoexponential fit has amplitude of 6295 and a time constant of 95.5ms.

only one example of several biexponentials which could do a reasonable job of reproducing the internal capsule's relaxation distribution.

It should be noted that comparing the relaxation distribution to a peak monoexponential fit is not very sensitive to large or small time constants. The PSF's in figure 4.1a shows the sensitivity drops off below 50ms and above 2000ms. The only way to improve this sensitivity while preserving linear resolution is to sample earlier that 10ms for smaller time constants and later than 5.12s for longer time constants. For the MRR scans considered herein, the smaller time constants seem to be the main problem.

4.3.3 MS patients

The 48 echo sequence was also performed on four MS patients. One MS patient had a new lesion and was followed serially. The results presented here are for the MS patient that was followed serially but the range of T1 and T2 time constants was similar in the single scans of each of the other three patients. Figure 4.14 show a 30ms conventional spin echo image from the 48 echo sequence for one of the MS patients.

The prominent feature in Figure 4.14 is the lesion to the right side of the brain. The lesion consists of a core, a thin ring and a thick "doughnut" like structure which is possibly edema. But it impossible to determine if the "doughnut" is indeed edema from the conventional spin echo MRI scans.

Figure 4.15 shows the peak amplitude image for the MS patient. The lesion core clearly has a lower signal than the adjacent white matter, most likely due to T1 saturation since the time between the last 180 degree pulse and the next excitation pulse is 2.03s. Also visible in this image is so called "dirty white matter". It is white matter which appears altered from normal but does not appear as intense as a focal lesion. It is visible on both sides of the ventricle although only one side is labeled.

Figures 4.16 and 4.17 show the peak time constant image at two different contrasts

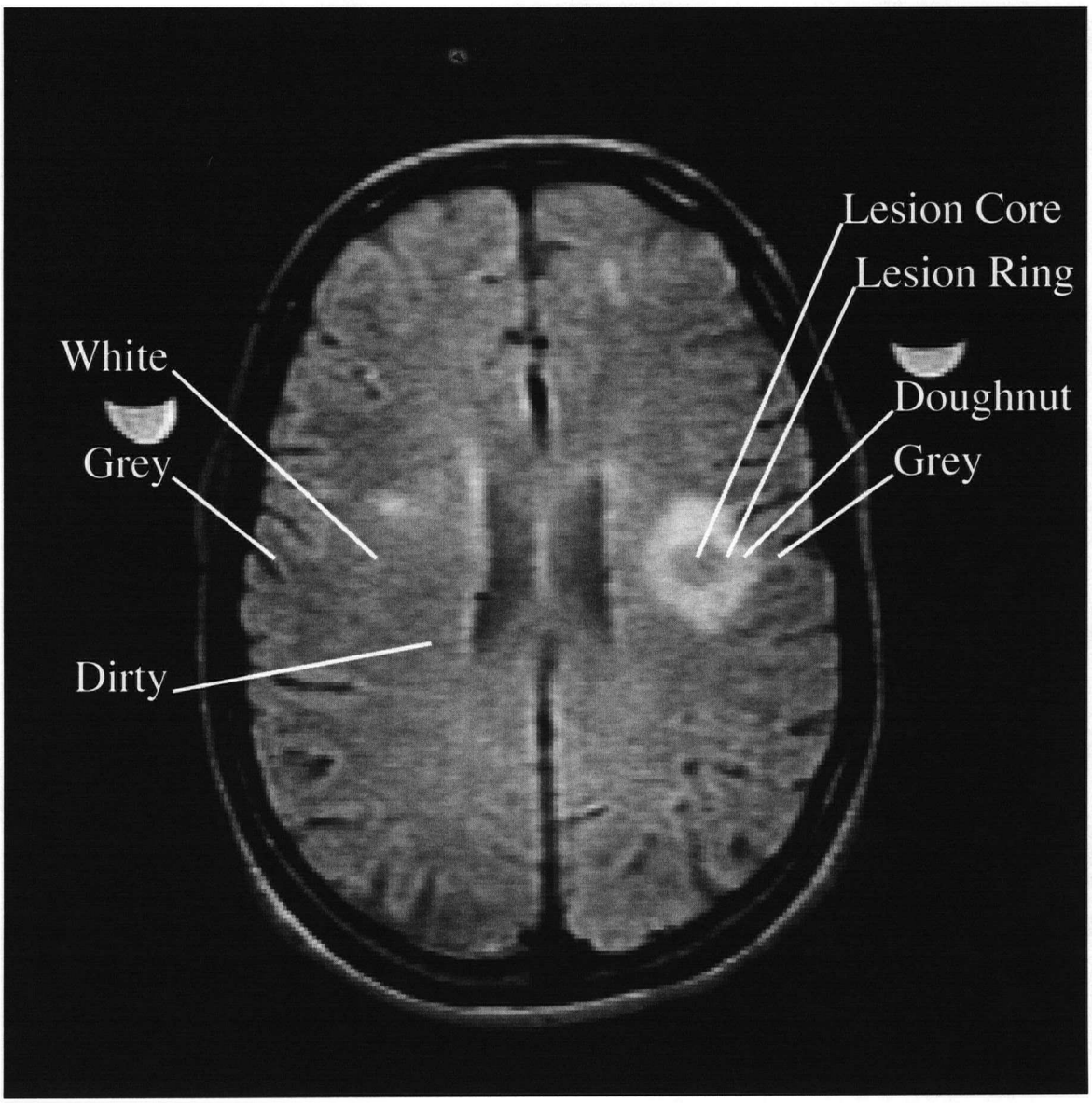

Figure 4.14: Spin echo image at 30ms from the 48 echo sequence for the MS patient.

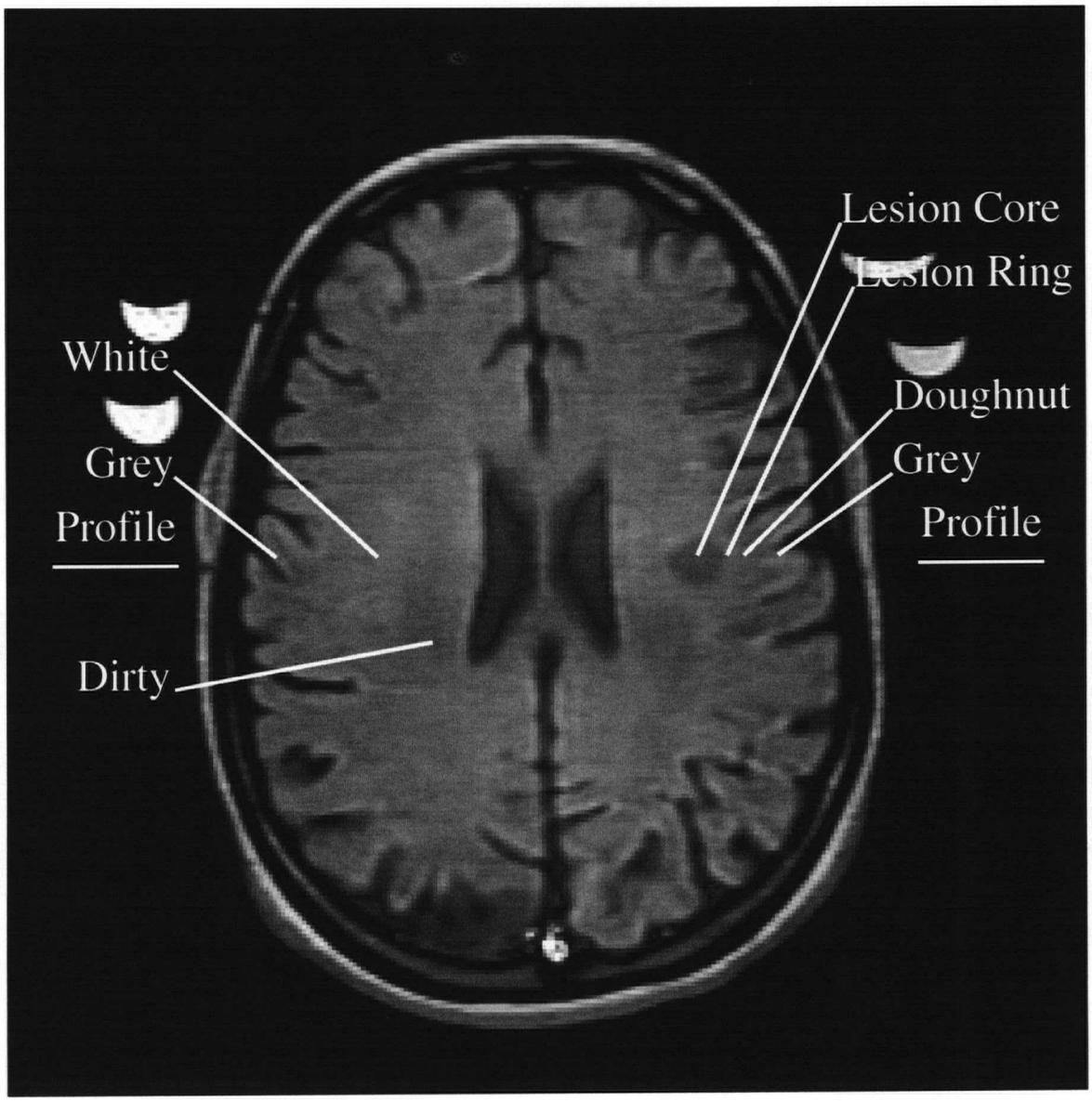

Figure 4.15: Peak amplitude image of the MS patient.

in order to show the structure of the lesion as well as the rest of the brain. The diameter of the lesion shown on this scan is about 20mm. It is interesting to note the "doughnut" surrounding the lesion core is barely visible on the peak amplitude image.

Figure 4.18 shows a profile from figures 4.15 and 4.17. The core of the lesion has peak time constants ranging from 400ms to 430ms. The "doughnut" has a time constant of about 170ms.

Figure 4.19 shows an image of the area between the measured relaxation distribution and the peak monoexponential for each pixel. As for the normal volunteer, most of the voxels in the brain have a near monoexponential relaxation distribution. Surprisingly this is also true within the lesion. An exception is the non-monoexponential nature of the dirty white matter.

Figure 4.20 shows the relaxation distributions for four voxels from a variety of pathologies. All have relaxation distributions which are well approximated by a monoexponential.

Figure 4.21 shows the relaxation distributions for grey, white and dirty white matter on the opposite side of the head from the lesion. The relaxation distributions of grey and white are almost identical. The dirty white has a clear non-monoexponential distribution with excess signal at long time constants.

The MRI scans presented for the MS patient were completed as part of a serial study. In addition to the MRR sequence, the weekly scans consisted of a dual echo conventional spin echo with TE of 30ms and 90ms and TR of 2500ms and a Fast spoiled phase gradient recall (Fast SPGR) for acquiring a proton density image. The fast SPGR sequence used a gradient echo and had a flip angle of 20 degrees with a TE of 2ms and a TR of 20,000ms. The dual echo was acquired with 1 average and the Fast SPGR with 3 averages. The slice thickness of the dual echo and Fast SPGR was 5mm as compared to the 3mm for the MRR sequence but all slices had the same center.

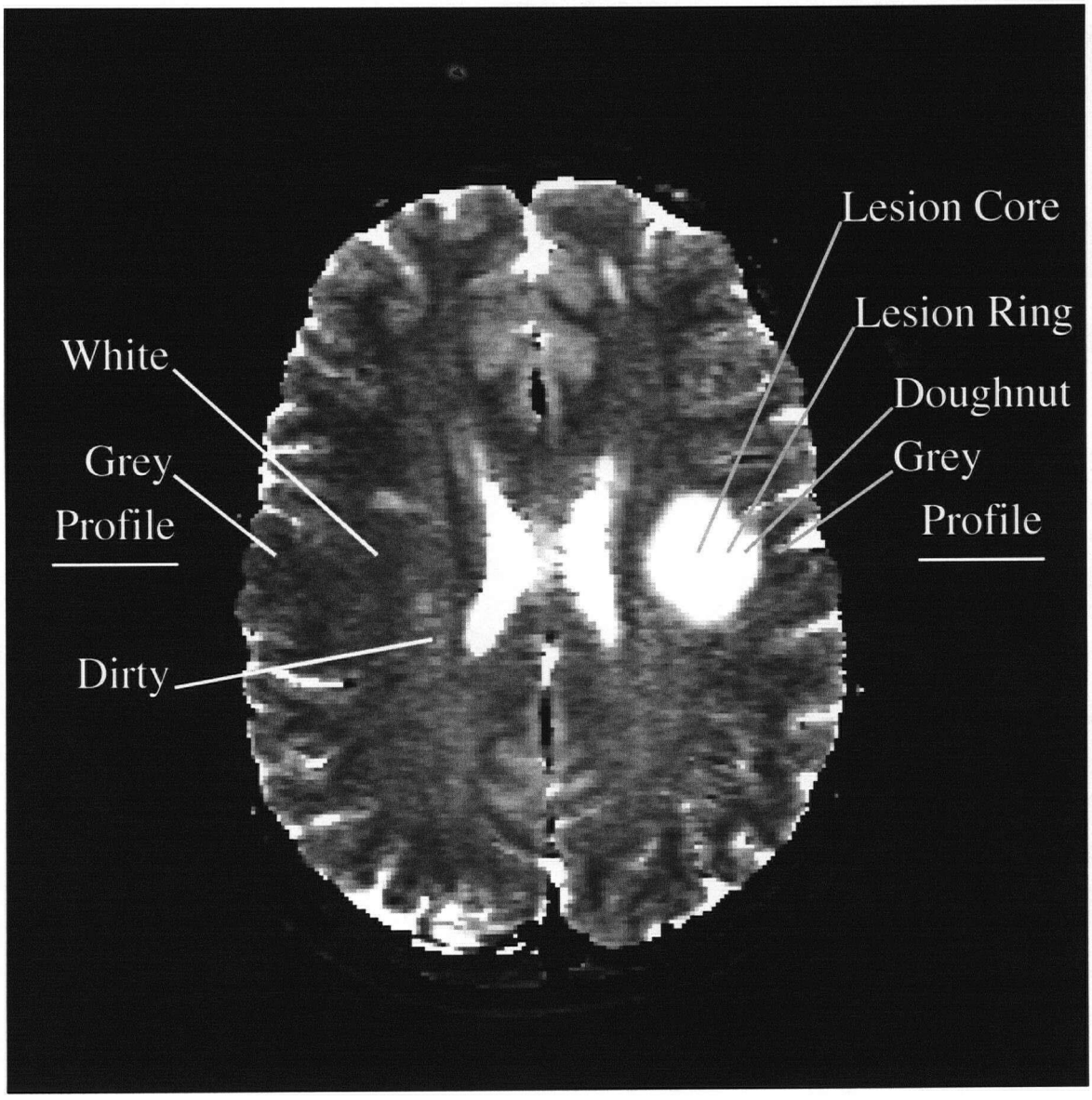

Figure 4.16: Peak time constant images of the MS patient with contrast between 50ms and 125ms.

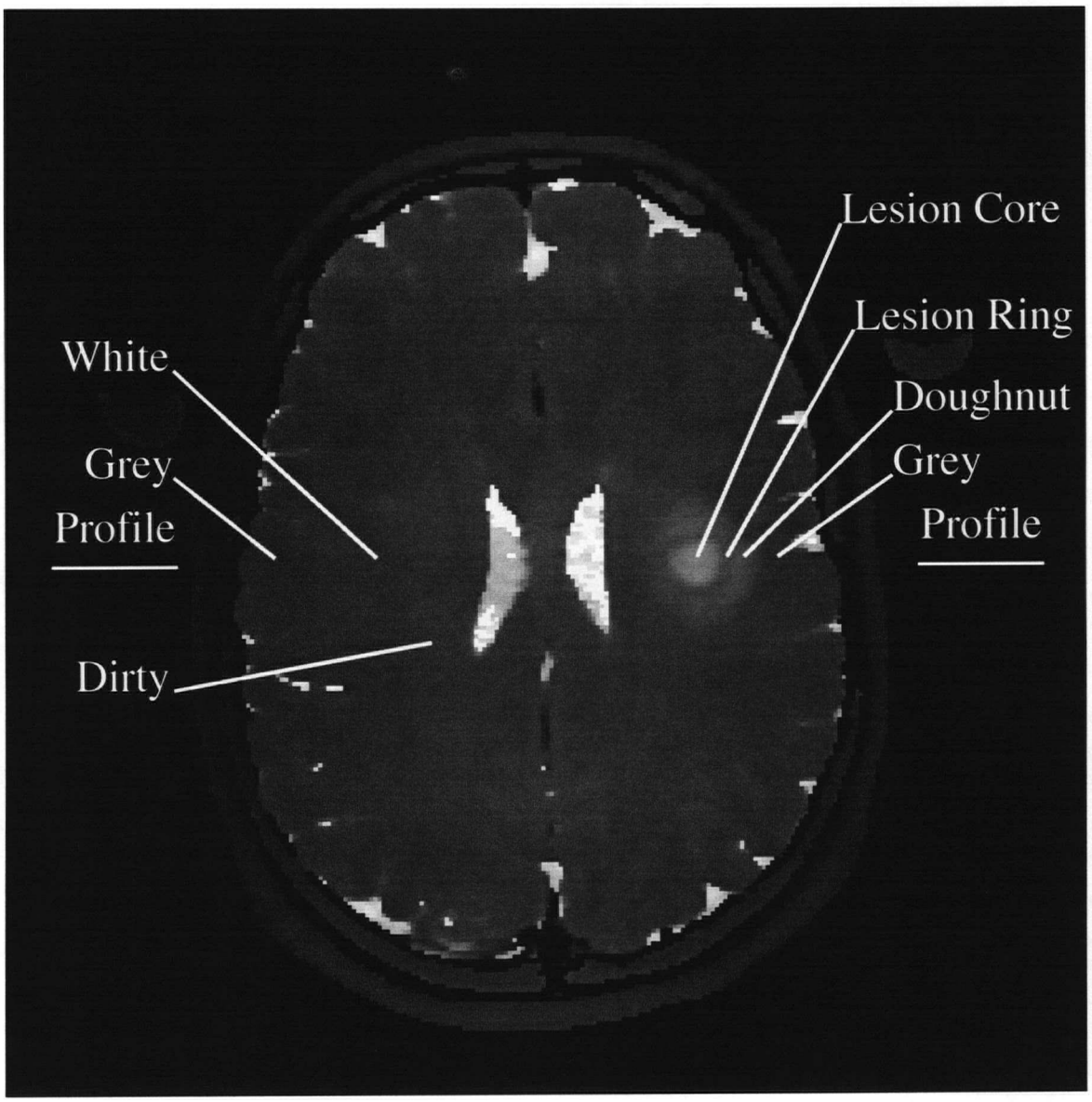

Figure 4.17: Peak time constant image of the MS patient with contrast between 10ms and 2500ms.

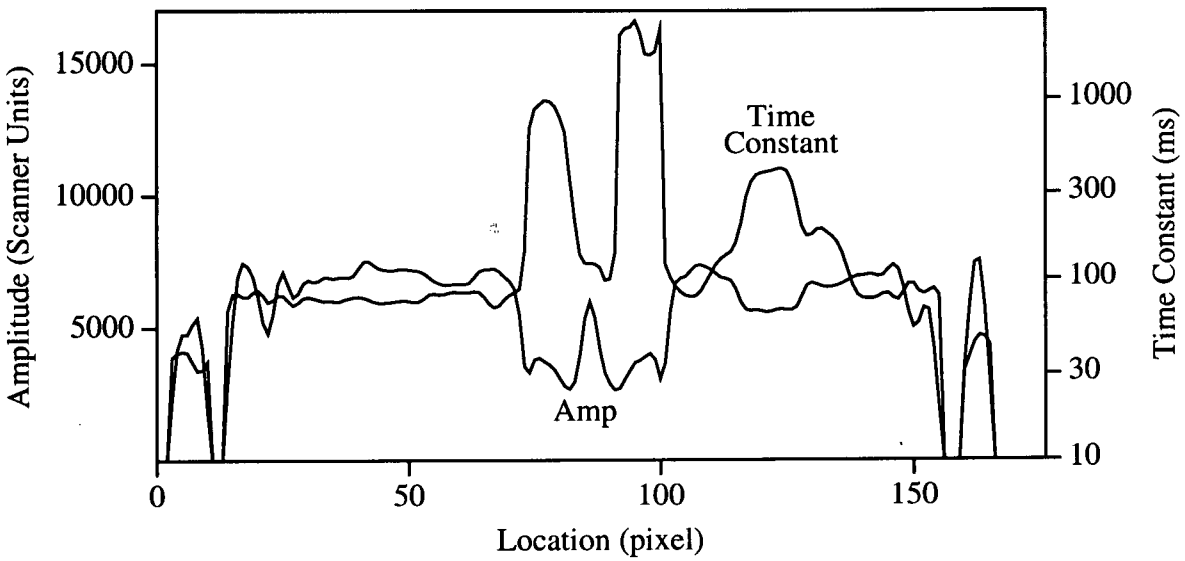

Figure 4.18: Profile of peak time constant for the MS patient.
A MS lesion is centered at the pixel location of 120 pixels. CSF is located at 80 and 95 pixels.

A selection of the scans from the study is shown in figure 4.22. The new lesion was first observed in the week 0 scan. The week -5 scan was the most most recent prior scan of the site of the new lesion and only a dual echo was performed.

The main trends over time are the disappearing of the "doughnut" surrounding the lesion core, the brightening of the lesion core on the peak amplitude images, and the darkening of the lesion core on the peak time constant images.

The darkening of the peak time constant image is clear indication of the decrease of the heightened T2 value of the lesion with time. The proton density image of the lesion core stayed constant in intensity over time, although the bright area shrunk over time. The peak amplitude images show the lesion core getting brighter over time. This trend, combined with the constant intensity of the proton density, indicates the heightened T1 of the lesion core is decreasing over time.

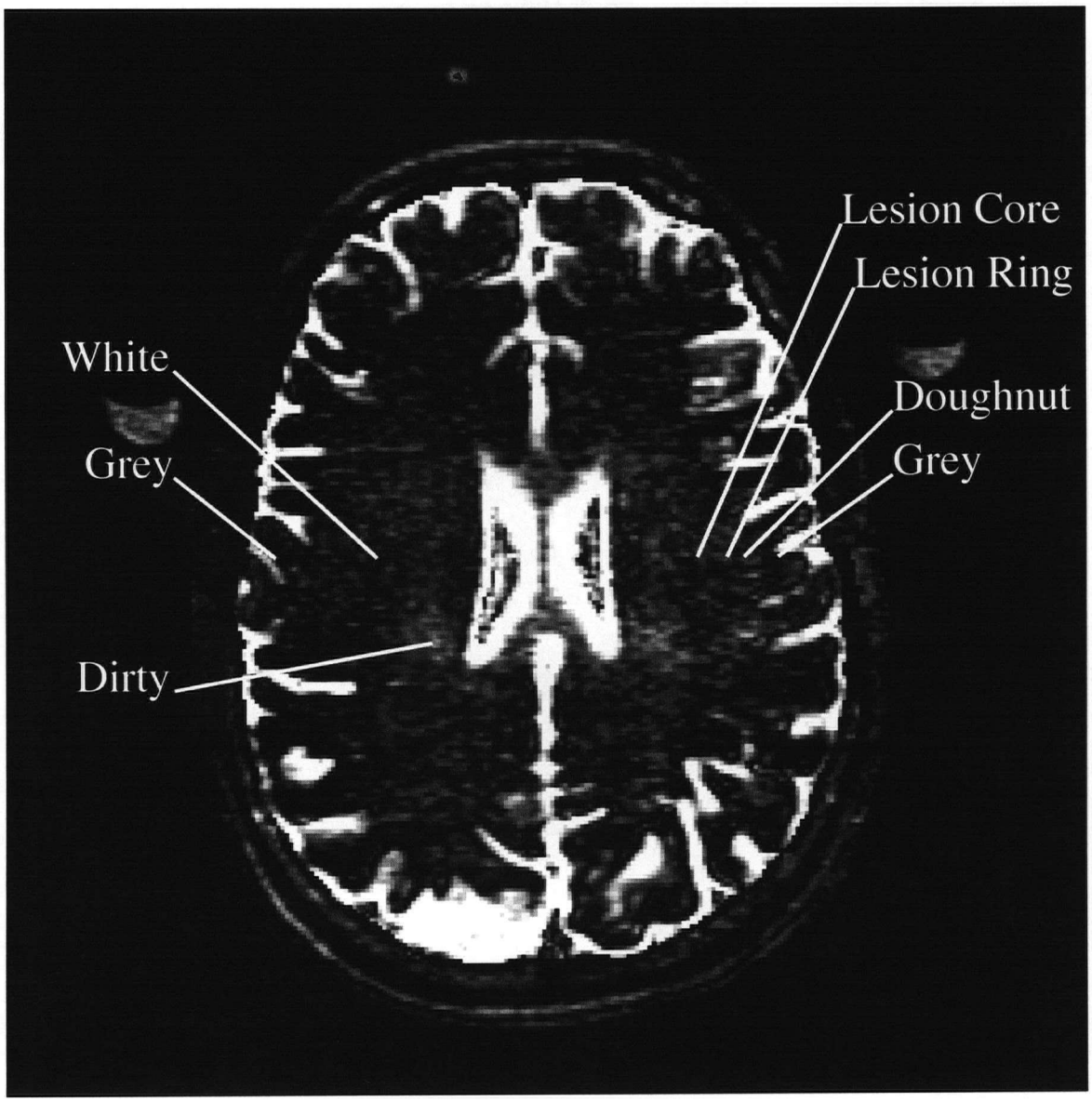

Figure 4.19: Image of how well the relaxation distribution of each pixel is approximated by a monoexponential for the MS patient.

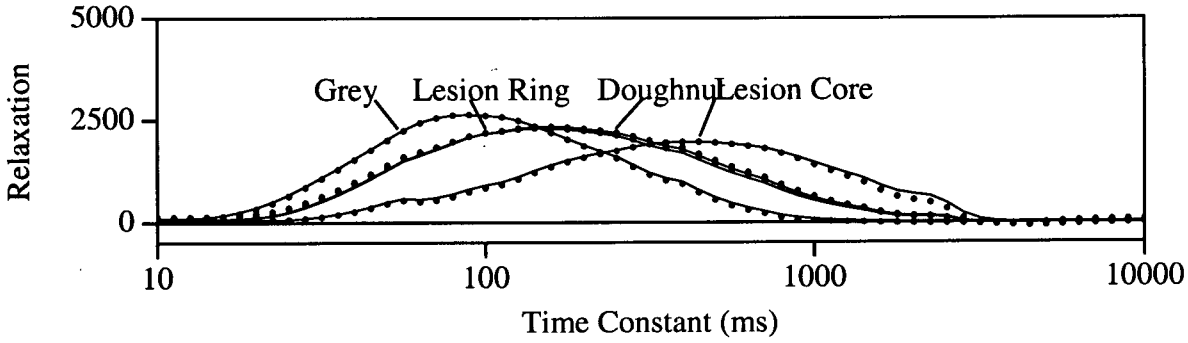

Figure 4.20: Relaxation distributions for voxels adjacent to the lesion in the MS patient. The "doughnut" relaxation distribution is slightly above and to the right of the relaxation distribution of the lesion ring.

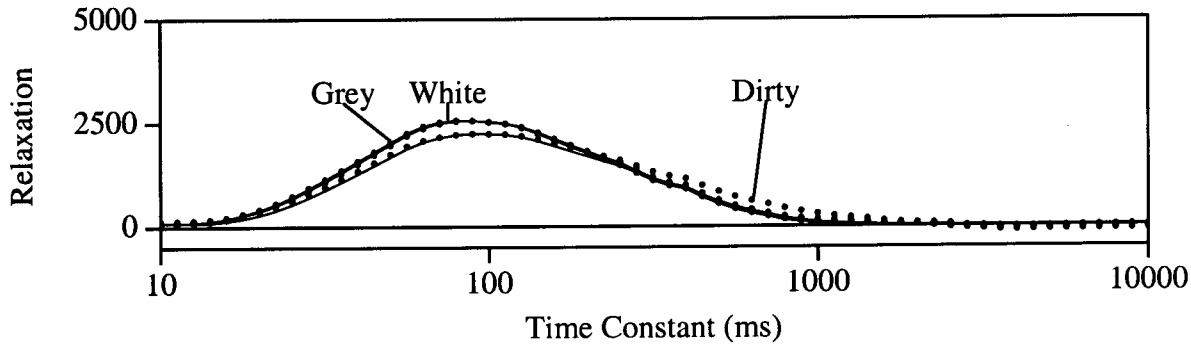

Figure 4.21: Relaxation distributions for voxels on the opposite side of the brain from the MS patient.

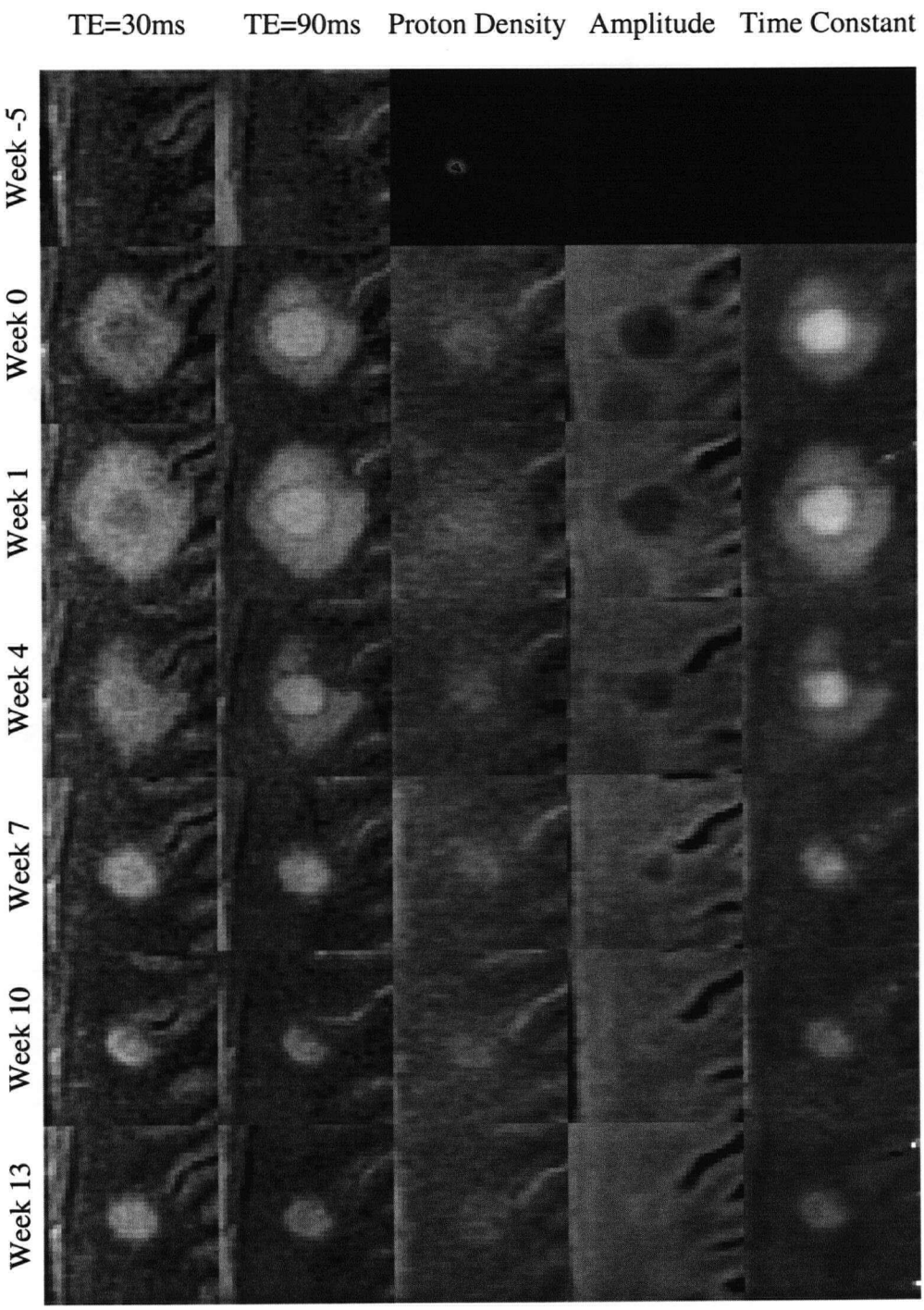

Figure 4.22: Sample of weekly scans from the MS patient.

4.4 Conclusion

The new tools of the 48 echo sequence and the new way to calculate the relaxation distribution provided new information on the pathology in MRI scans. The peak amplitude and time constant are particularly promising but further work is needed to assess their usefulness.

A surprising result of this study was that most voxels in an image had a relaxation distribution much closer to monoexponential than expected. This is in contrast to the published results that indicate a non-monoexponential of decay of some lesions measured as a whole. (Larsson, Barker et al 1998). However, as suggested in the literature, this non-monoexponential behaviour was probably due to the variation of the pathology of the lesion. The small voxel volume, $4.3mm^3$, used in the 48 sequence probably made the decay curves more monoexponential because, the larger the voxel, the more likely it is to contain a variety of pathologies. The small voxel size was in part possible because of the robust ability of the new way to calculate the relaxation distribution when noise is present.

One of the major problems in assessing the effectiveness of any method for monitoring MS pathology *in vivo* is the general lack of direct knowledge of the state of the pathology at the time of the MR scan. If in even a few cases we knew the exact *in vivo* pathology at the time of the MR scan, it would provide a valuable way to calibrate both MRR and MRS measurements.

The 48 echo sequence, because of its large number of echoes and TR of 7 seconds, takes about 30 minutes to run and only acquires a single slice. With more knowledge of the time constants of common pathologies it should be possible to reduce the time of the last echo and repeat time, thus speeding up the sequence. In addition, multi slice MRR sequences may also be possible.

In summary, MRR with a large number of echoes and the new method of calculating the relaxation distribution looks very promising as a way to obtain additional information about *in vivo* pathology non-invasively.

References

Arnold D J, Wolinsky J S, Matthews J S, Falini A 1998 The use of magnetic resonance spectroscopy in the evaluation of the natural history of multiple sclerosis *J Neurol Neurosurg Psychiatry* **94** (S95–S101)

Christofferson J O, Olsson L E, and Sjöberg S 1991 Nickel-doped agarose gel phantoms in NMR imaging *Acta Rad* **32** 426–431

Filippi M, Horsfield MA, Ader HJ, Barkhof F, Bruzzi P, Evans A, Frank JA, Grossman RI, McFarland HF, Molyneux P, Paty DW, Simon J, Tofts PS, Wolinsky JS, Miller DH 1998 Guidelines for using quantitative measures of brain magnetic resonance imaging abnormalities in monitoring the treatment of multiple sclerosis *Ann Neurol* **43** (499–506)

Istratov A A and Vyenko O F 1999 Exponential analysis in physical phenomena *Rev Scien Instr* **70** 1233–1257

Larsson HBW, Barker GJ, MacKay A 1998 Nuclear Magnetic resonance relaxation in multiple sclerosis *J Neur, Neurosurgery and Psychiatry* **64** S70-S76

MacKay A, Whittall K, Adler J, Li D, Paty D, Graeb D 1994 *In Vivo* Visualization of Myelin Water in Brain by Magnetic Resonance *Magnetic Resonance in Medicine* **31** 673–677

Paty DW and Ebers GC 1998 *Multiple Sclerosis* (Philadelphia: F A Davis Company)

Whittall K P, Bronskill M J and Henkelman R M 1991 Investigation of analysis techniques for complicated NMR relaxation data *Journal of Magnetic Resonance* **95** 221–234

$Id: MP.tex, v1.402001/06/1921:49:38 kscover Exp kscover$

Chapter 5

The Myelin Peak

5.1 Introduction

The detection of a MR relaxation signal, believed to be due to the water compartmentalized in the myelin surrounding the axons in the human brain, has been reported in the literature (MacKay et al 1991, MacKay et al 1994, Whittall 1997). The algorithm used to calculate the relaxation distribution from the magnetic resonance decay curves was a regularized nonnegative least squares technique (NNLS) which produced a model which fit the data. The NNLS algorithm takes advantage of *a priori* information, information beyond that supplied by the data, to select one of the many models that fit the data. But there are many other models which fit the data for a given data set. The question is whether all the models that fit the data are consistent with the myelin peak or whether the algorithm used happened to pick a model which was consistent with the myelin peak hypothesis when other models were not.

The function of the myelin in the brain is to increase nerve signal propagation velocities by ten to a hundred times. Demyelination in multiple sclerosis is responsible for conduction blocks or delays leading to disabilities. A robust technique for measuring myelin by MR would be a valuable tool in the understanding of multiple sclerosis.

From anatomical studies the literature predicts a time constant for the myelin signal with in the range of 10 to 50ms but most likely around 15ms. Additional signal is predicted to come from the tissue in the 80 to 100ms and above range, based on *in vitro*

NMR relaxation measurements. *In vivo* MR investigations of the myelin peak have been unable to localize its T2 time beyond limiting it to the 10 to 50ms window. Important questions to answer are (1) whether MR relaxation measurements can independently confirm the exisitance of the myelin peak, (2) what additional information we can learn about the peak and (3) how much *a priori* information can and should be invoked as part of the analysis.

5.2 Method

MR decay curves from a slice of human brain were measured *in vivo*. The CPMG sequence used to measure the decays had 48 echoes acquired on a 32x32 acquisition matrix with a 220mm field of view and a 5mm slice thickness. The large voxel size was used to achieve high signal to noise ratio with the smallest possible volume. The first 32 echoes were spaced 10ms apart followed by 16 echoes at 100ms spacing. The sequence had a TR of 4s. The L1 Fourier transform window described previously was applied to suppress truncation artifacts before zero filling to 128x128 and Fourier transforming.

To calculate a relaxation distribution with sufficient linear resolution, an estimation matrix with a noise gain of at least 10 would be needed. The PSF, noise response and resolution function for the selected estimation matrix are shown in figure 5.1. The estimation matrix is invertible; thus, the relaxation distribution conserves all the information in the data. Therefore, once the estimate, its noise statistics and resolution functions are calculated the original data, data functions and noise statistics can be safely ignored and fits performed to regions of the estimate.

The resolution functions for the estimation matrix clearly demonstrate that any signal in the 15ms region of the relaxation distribution will have little sensitivity to signals in the 80 to 100ms region. The resolution functions also have the desirable property of

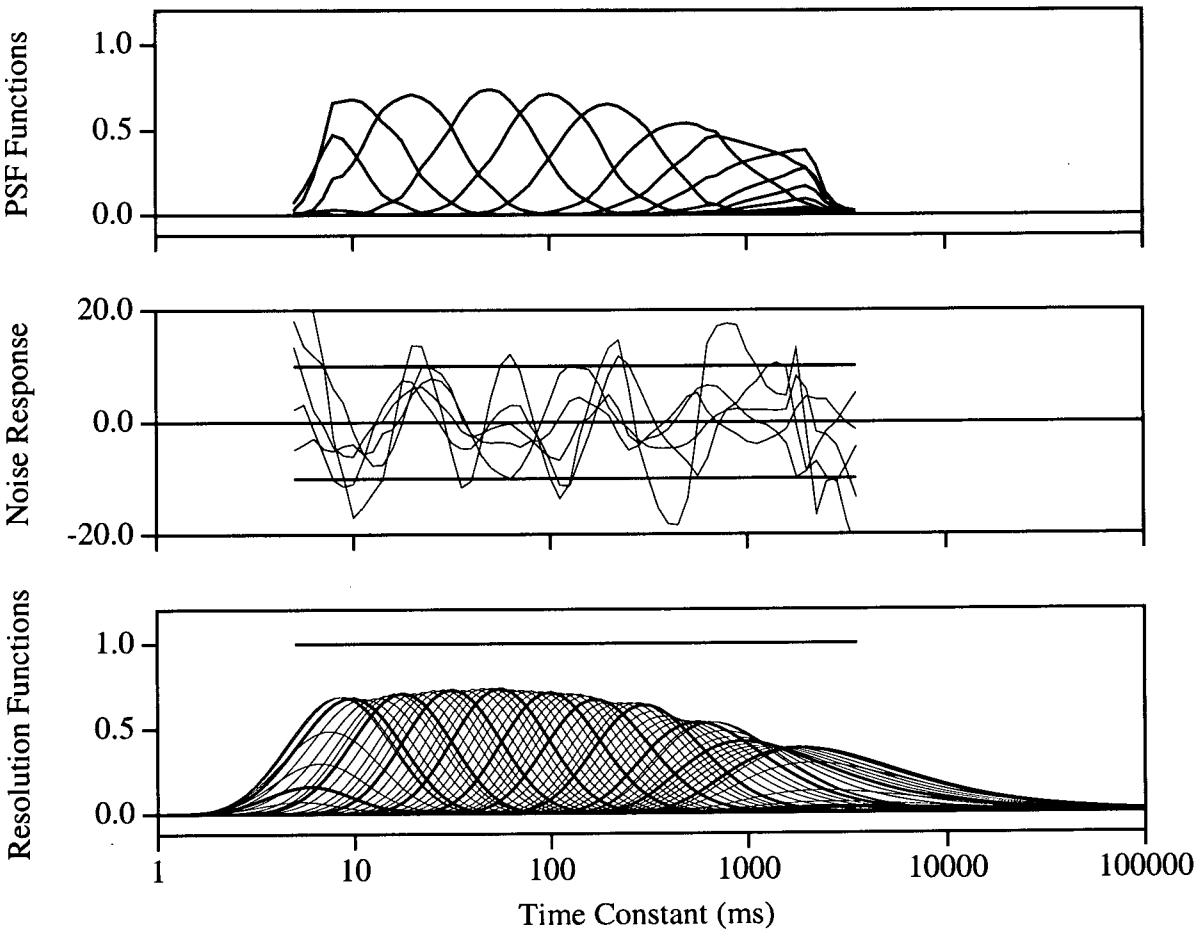

Figure 5.1: PSF, noise response and resolution functions for an estimation matrix with 32 echoes at 10ms spacing and 16 echoes at 100ms spacing. The noise gain had a value of 10.

having nearly uniform logarithmic resolution in the 10 to 200ms range. The superior resolution of a higher noise gain estimation matrix would have been preferred but the signal to noise ratio of the decay curves was too low to implement such a matrix.

To compare several methods that generate models that the fit the data in the literature to the estimation matrix results, the relaxation distribution of the decay curve for the posterior internal capsule (PIC) was calculated using several fitting methods.

MacKay *et al* (1994) used a regularized form of the nonnegative least squares

algorithm (NNLS) to calculate a model that fit the data which, they believed based on prior experience, was similar in nature to the true model. The full details are described in Whittall and MacKay (1989) and Vavasour *et al* (1998). NNLS finds the nonnegative model which fits the data which has the smallest χ^2. The equation for χ^2 is given in equation 1.5. For calculation purposes the model as composed a set of monoexponentials spaced at approximately 50 points per decade from 15ms to 10,000 ms and only the amplitudes of the monoexponentials were allowed to vary.

The regularized from of NNLS (RNNLS) caps the χ^2 at a few percentage larger than the values given by NNLS for the same data set (Vavasour *et al*, 1998). Within this χ^2 constraint the smallest L2 model is found by minimizing

$$I = \int_a^b m^2(y)dy \tag{5.1}$$

where the bounds a and b are the upper and lower time constants. The RNNLS model useful because it less sensitive to noise than the NNLS. Again the model was composed of a set of monoexponentials spaced at approximately 50 points per decade from 15ms to 10,000 ms and only the amplitudes of the monoexponentials were allowed to vary.

Another common modelling method used in the literature is to sum a few monoexponentials. The χ^2 is minimized while the amplitude and time constant of each monoexponential is allowed to vary. For the PIC decay curve 3 monoexponentials were found to yield a good fit.

The fourth method, NNL4, found the smallest model in the L4 norm sense that fits the data within a reasonable χ^2. The goal of this model was to find a solution that was fairly flat. From the literature, a reasonable χ^2 is defined to be the number of data points (Menke 1984). The PIC decay curve has 48 data points, therefore χ^2 was capped at 48. Then the L4 norm, defined as

$$L4 = \int_a^b m^4(y)dy, \tag{5.2}$$

was minimized. Again, for calculation purposes, the model was composed of a set of monoexponentials spaced at approximately 50 points per decade from 15ms to 10,000 ms and only the amplitudes of the monoexponentials were allowed to vary.

In related work, Vavasour *et al* (2000) added to the NNLS method the additional *a priori* information that the time constants of the decay rates may only be 20, 80, 120 or 2000ms. While Vavasour *et al* made it clear that, although this additional four peak *a priori* information was imperfect, they felt the fitted amplitudes of the time constants were still informative. The additional *a priori* information was required because the decay curves they were analysing were too noisy for RNNLS.

As an additional comparison of the performance of the linear resolution algorithm with RNNLS, six simulations were performed. The simulations were divided into two groups, one group per model, and three different signal to noise ratios, 50:1, 200:1 and 1000:1 within each group. The first model was an idealized form of the signal from white matter. It was composed of two peaks at 20 and 80ms. The peaks had amplitudes of 2000 and 18000 respectively in arbitrary units. The 20ms peak is intended to represent the signal from water in myelin while the main peak represented the signal from the rest of the tissue.

The second model effectively had uniform intensity from 1ms to 10s on the logarithmic scale and is referred to as the flat model. It was generated by summing together 201 monoexponentials uniformly spaced at 50 points per decade. The sum of the amplitudes of the monoexponentials was 20000, the same as for the white matter model.

The decay curves were initially sampled at the same 48 time points as for the PIC. But results of RNNLS prompted the simulations to be repeated for only the first 32 time points. Estimation matrices with noise gains of 1.0, 3.2 and 10.0 were used to estimated the relaxation distributions from the decay curves signal to noise ratios of 50:1, 200:1 and 1000:1 respectively. The resolution functions of the estimation matrices used are shown

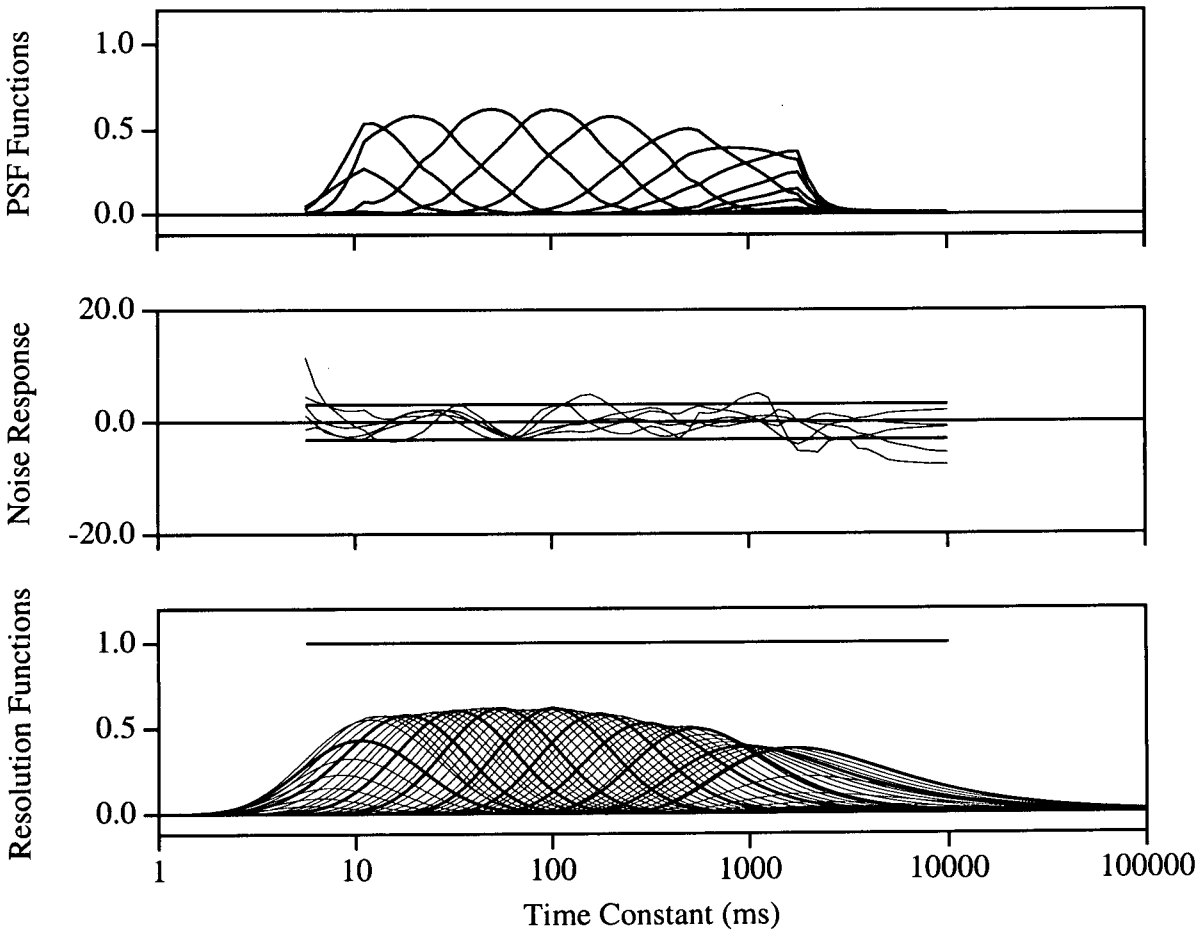

Figure 5.2: PSF, noise response and resolution functions for an estimation matrix with 32 echoes at 10ms spacing and 16 echoes at 100ms spacing. The noise gain had a value of 3.2.

in figures 5.1, 5.3 and 5.2.

5.3 Results

Figure 5.4 shows the slice from which the decay curves were acquired but in higher resolution. Figure 5.5 shows the same slice acquired by the CPMG sequence with a 32×32 acquisition matrix at the 30ms echo. The raw data was multiplied by the L1 windows and zero filled to 128×128 before image reconstruction for display purposes.

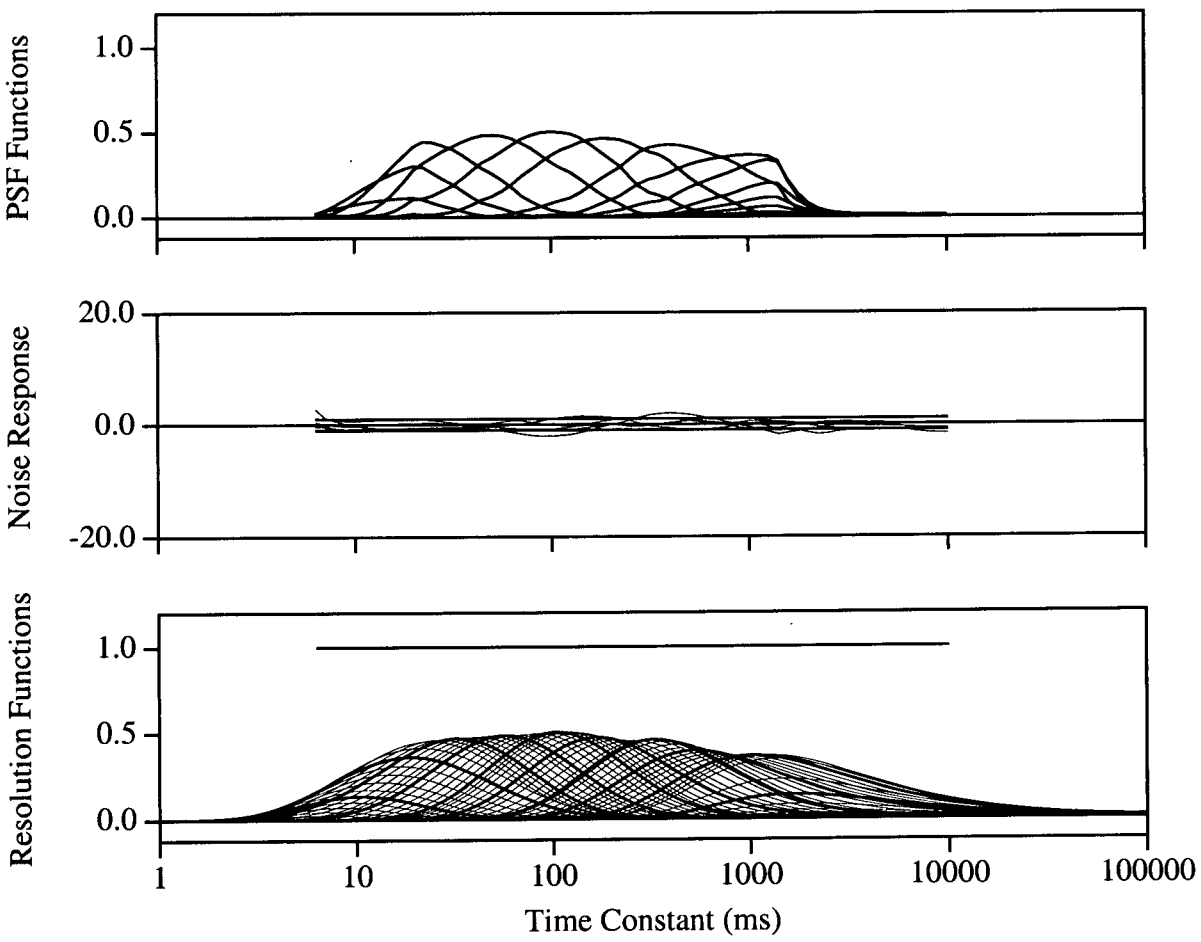

Figure 5.3: PSF, noise response and resolution functions for an estimation matrix with 32 echoes at 10ms spacing and 16 echoes at 100ms spacing. The noise gain had a value of 1.0.

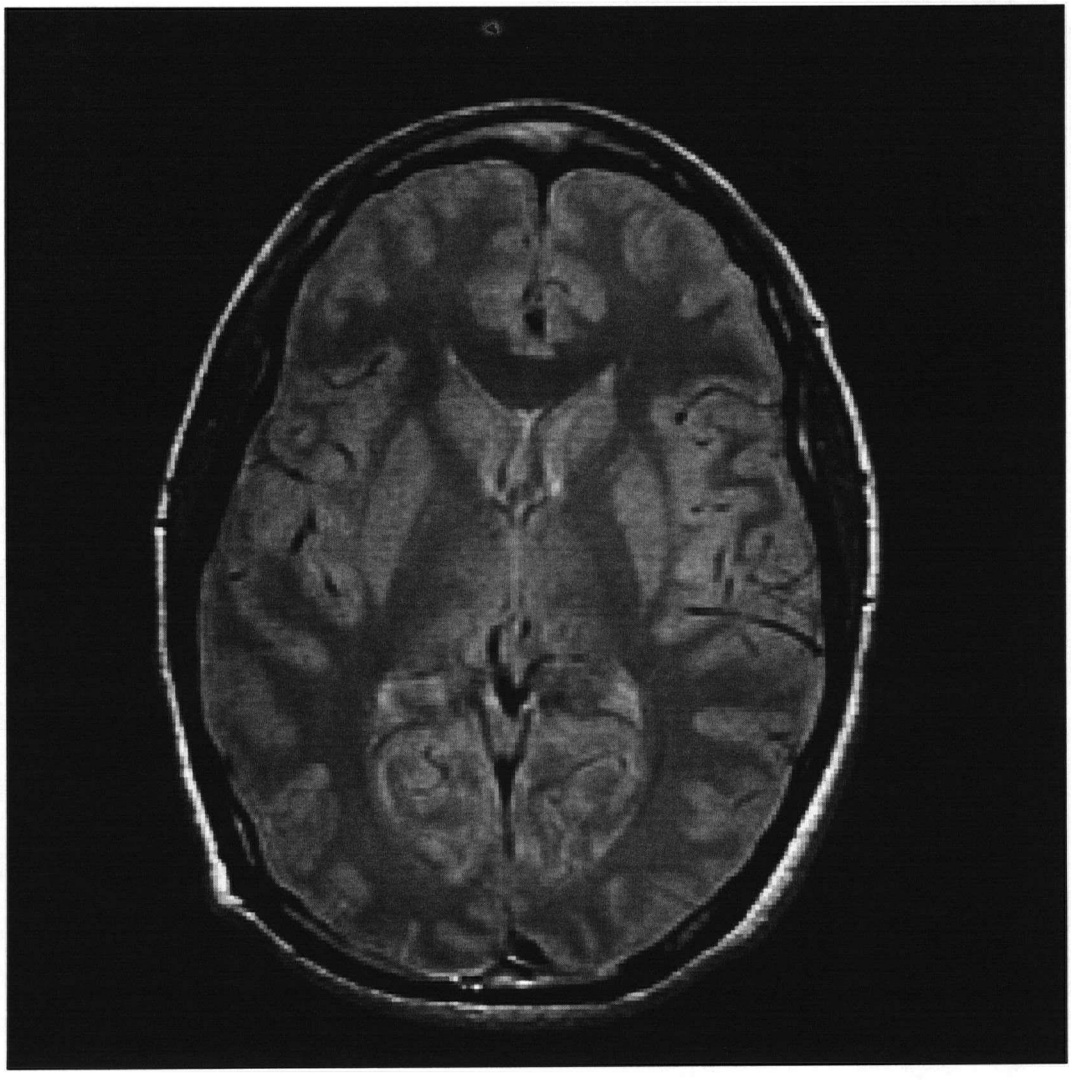

Figure 5.4: Fast Spin Echo Image of selected slice with TE=34ms, TR=2200ms and 256x256 acquisition matrix.

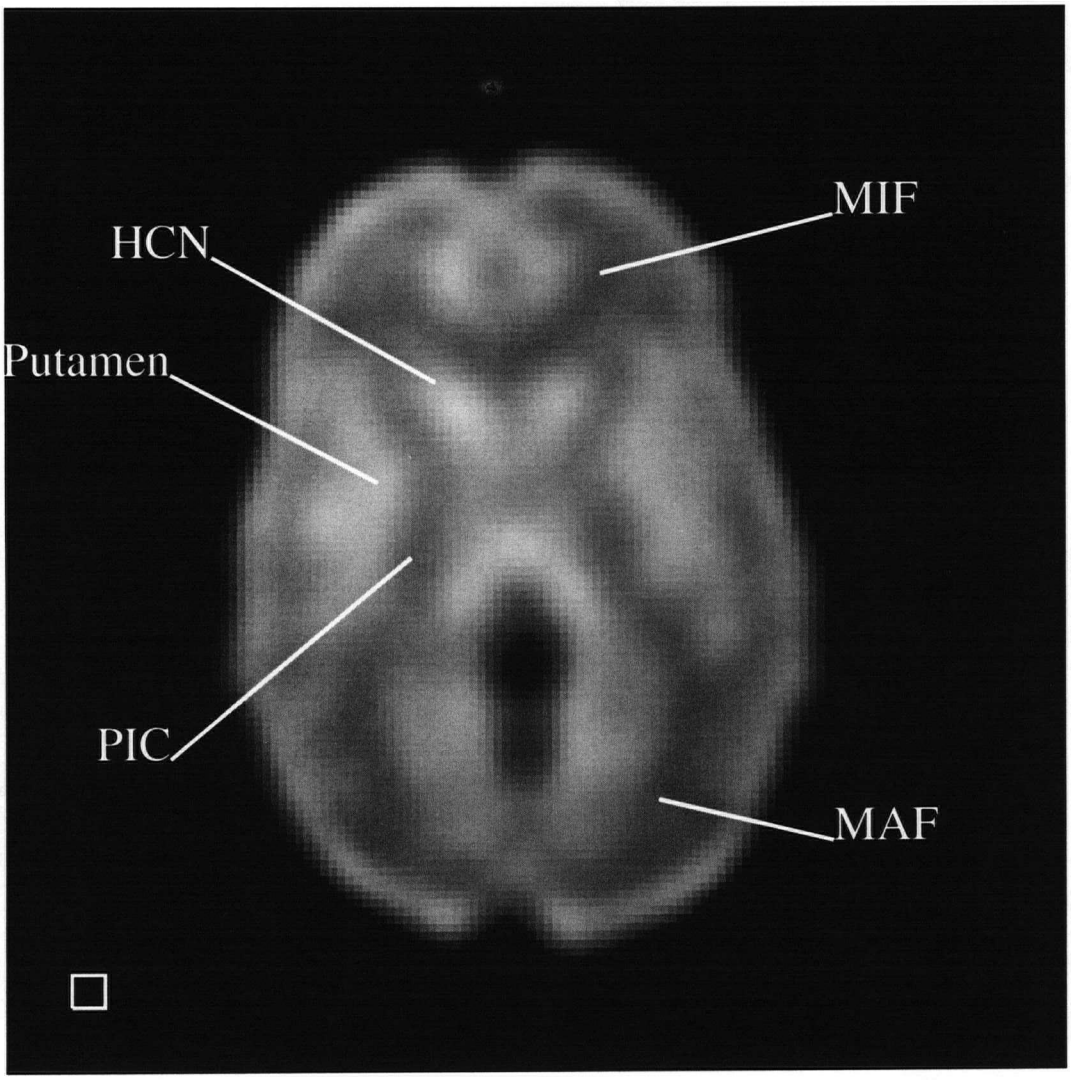

Figure 5.5: TE=30ms image from the 32x32 acquisition matrix CPMG sequence. Before reconstruction the raw data was multiplied by the L1 window and zero filled to 128x128. The sampled pixels are located in the Minor Forceps (MIF), Head of the Caudate Nucleus (HCN), Putamen, Posterior Internal Capsule (PIC), and Major Forceps (MAF). The square in the lower left corner indicates the size of a pixel in the acquisition matrix.

Five decay curves were sampled from the 32x32 image shown in figure 5.5. Figure 5.6 shows the decay curves and relaxation distributions for the putamen and the PIC. The relaxation distributions have a monoexponential fitted to their peaks as described previously. The PIC clearly has excess signal in the 15ms region as compared to the putamen.

To accomplish proper analysis of the PIC decay curve in figure 5.6, a measure of the noise was required. This was accomplished by taking the standard deviation of the last 10 points of the PIC decay curve yielding a value of 14.49. Since the decay curve contains the magnitude data, the noise in the last part of the decay curve was Raleigh in nature. Therefore the standard deviation must be divided by 0.655 to get the standard deviation of the corresponding Gaussian noise. This process yielded a standard deviation for the noise of 22.12. It should be kept in mind that the character of the noise is unlikely to be simply Gaussian since motion artifacts due to blood flow, among other sources, will produce partially time correlated noise.

Application of the χ^2 statistical test to the different between the putamen and the PIC at the peak point (p=0.02) and the 5 points near and at the peak, allowing for correlated noise, (p=0.0002) yielded a significant difference. From pathology studies it is known that PIC has a relatively high concentration of myelin whereas the putamen has very little. Figure 5.7 shows three more relaxation distributions. The major forceps and minor forceps have excess signal in the 15ms region while the head of the caudate nucleus has little signal. Again, the signals are consistent with their respective myelin concentrations.

A similar statistical test applied to the negative signal region around 1s in the relaxation distribution of the PIC in figure 5.8 showed the signal to be indistinguishable for zero.

Since all the information in the original data is preserved in the estimate, we can fit

monoexponentials to part of the relaxation distributions to predict the signals that may compose them. The PIC relaxation distribution for figure 5.6 has a monoexponential fit to the peak of the distribution. It has an amplitude of 12952 and a time constant of 73.8ms. This monoexponential was subtracted from the PIC measured data leaving two peaks. The left peak, corresponding to the myelin, was fitted with a monoexponential with an amplitude of 1580 and a time constant of 12.1ms. The sum of the two monoexponentials are plotted on top of the original PIC distribution in figure 5.8. There is clearly a good fit to the data up to 100ms giving a good indication that the myelin peak is well modeled by a single exponential.

Figure 5.9 shows an image of the signal at 14ms. This image was calculated by calculating the signal intensity at 14ms in the relaxation distributions for each of the pixels in figure 5.4. Again, the high intensity signal correlates with the myelin location in the brain, with a few exceptions. The bright regions on the outside of the brain correspond to fat. Three bright regions along the mid-line of the brain include regions with cerebral spinal fluid (CSF). The CSF may have been contaminated with flow artifacts and thus were not analysed. Note that the CSF is much easier to recognize on the high resolution images.

Figure 5.10 shows five models that fit the data for the PIC decay curve. The χ^2 for each of the five models indicate a good fit with the exception of the 4 peak NNLS model. The values of 157.12 indicates a poor fit. This is probably due to some inconsistency between the four peak *a priori* information and the data.

All five model show signal in the 15ms peak region as predicted by the estimate with linear resolution. The sum of the in these regions are 1167 (NNLS), 1649 (RNNLS), 1220 (L4) and 1692 (3 Monoexponential) and 2788 (4 peaks). The values for the the 4 peaks should be treated with caution because of the high χ^2.

Figures 5.11 through 5.16 show the results for the six simulations. The first three plots

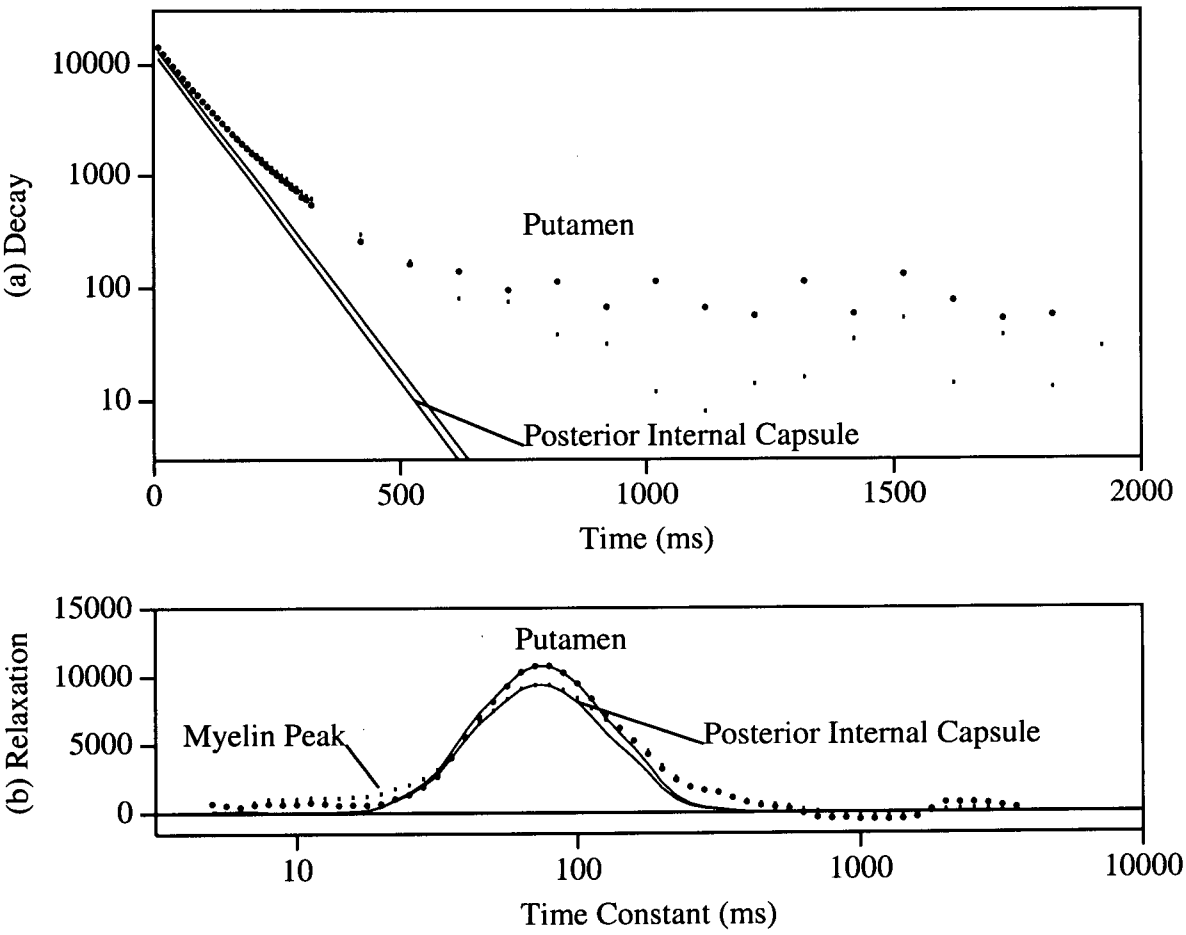

Figure 5.6: Decay curves and relaxation distributions of two sample pixels from the 32x32 acquisition matrix image.

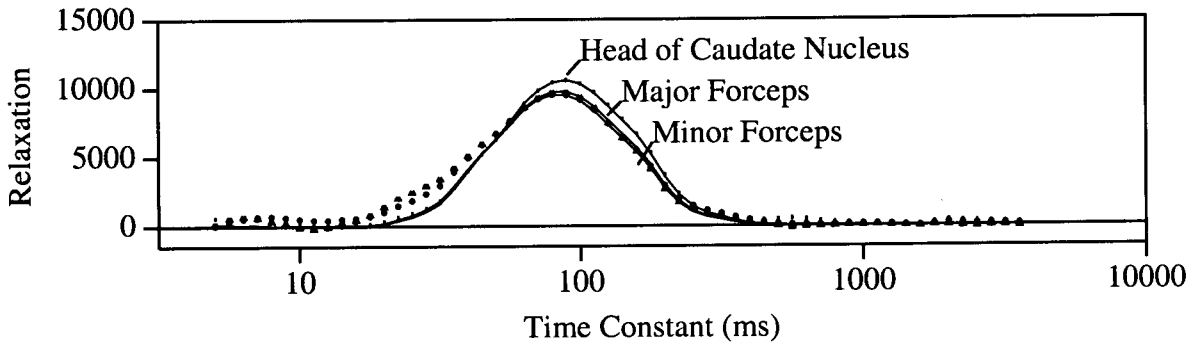

Figure 5.7: Three additional relaxation distributions for three sample pixels for the 32x32 acquisition matrix image.

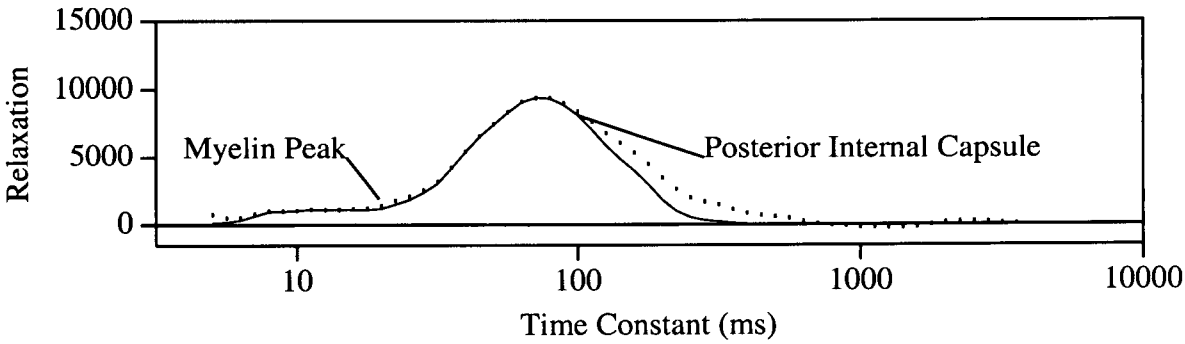

Figure 5.8: Relaxation distribution for PIC (dotted line) with biexponential fit with amplitudes of 12952 and 1580 and time constants 73.8ms and 12.1ms respectively (solid line). Note that the distribution above 100ms is deliberately left unfit.

Model or Estimate	Myelin Peak Amplitude	Main Peak Amplitude	χ^2
NNLS	1,667	16,568	36.90
RNNLS	1,649	16,564	37.12
3 Monoexp	1,220	16,729	45.06
NNL4	1,692	16,434	48.00
4 Peak NNLS	2,788	16,628	157.12
Linear Res	1,580	14,532	undefined

Table 5.1: Synopsis of estimate and models of the unknown data. The myelin peak and main peak for the linear estimate are from the monoexponential fits.

of each figure show the first three realizations of the RNNLS analysis. The fourth plot shows the average of RNNLS results for the 1024 realization for each simulation. The fifth plot shows the linear resolution results for first realization of each simulation along with the 95% confidence interval due to the noise. The confidence interval was calculated from multiplying twice the standard deviation of the noise in the decay curve by the noise gain of the estimation matrix. Only one realization of the relaxation distribution is presented for linear resolution as opposed to three for RNNLS because linear resolution also has the associated resolution functions to convey information on the nonuniqueness of the

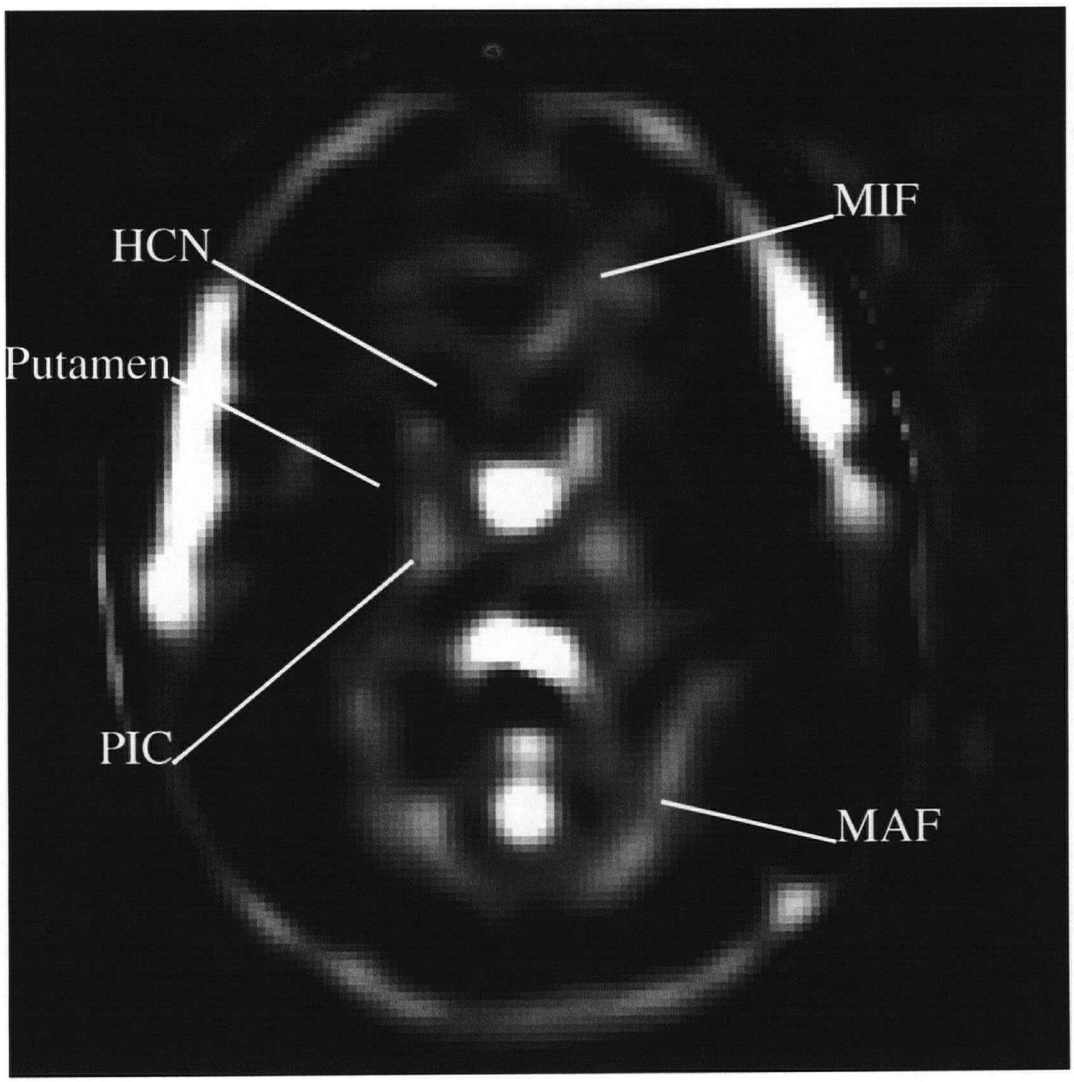

Figure 5.9: Image with time constant at 14ms from the same CPMG data set presented in figure 5.3. The image was calculated by using the 14ms signal for the relaxation distribution of each pixel.

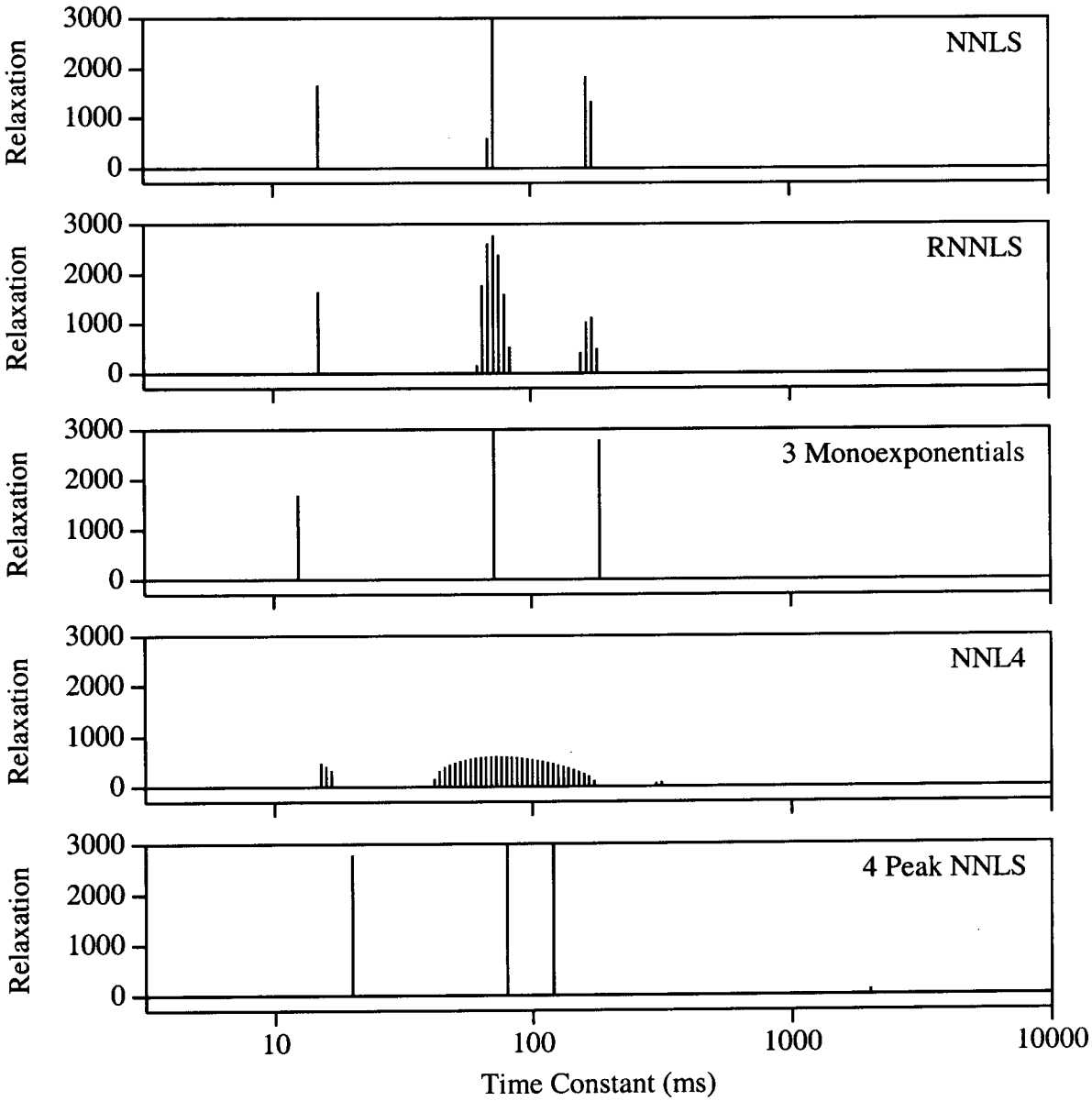

Figure 5.10: Five fits to the PIC decay curve. The χ^2 are 36.90, 37.12, 45.06, 48.00 and 157.12 respectively. Several peaks extend beyond the top of the plots. The largest peak for the NNLS solution has an amplitude of 11,116. The largest peak of the 3 monoexponentials is 12258. The peaks at 80 and 120ms for the 4 peak NNLS solution have heights of 9,342 and 4,386 respectively.

results.

Table 5.2 shows the results from the 1024 simulations for the myelin peak from the RNNLS and linear resolution for the three simulations with the white matter model. The standard method for calculating the amplitude of the myelin peak from the RNNLS is to integrate between 10ms and 40ms. This was done for all the realizations for each simulation. The 2.5, 50 and 97.5 percentiles from myelin amplitudes for each simulation were extracted and presented, with some simple arithmetic, as the average myelin peak and 95% confidence intervals. The time constant of the myelin peak in the RNNLS is not normally calculated because it is primarily determined by the lower bound of the RNNLS algorithm used. In the simulations the lower bound was set at the standard value of 15ms.

Table 5.3 shows the results from the 1024 simulations of RNNLS for just 32 echoes. These additional measurements allowed the comparison of the effects of Rayleigh noise on RNNLS. Two myelin windows were applied to examine the effect of the different sizes on the myelin peak amplitude.

The linear resolution myelin peak amplitude and time constant were calculated for each realization by first fitting a monoexponential to the main peak as explained in chapter 4. The fitted monoexponential was then subtracted off to leave the myelin peak. The myelin peak was then fitted with a monoexponential to yield its amplitude and time constant. The same method as for the RNNLS myelin peak was used to calculate the average and confidence intervals for the amplitude and time constant of the myelin peak.

5.4 Discussion

A major advantage of estimates with linear resolution over models that fit the data is demonstrated for the PIC decay curve in figures 5.6 and 5.10. Many estimates of models

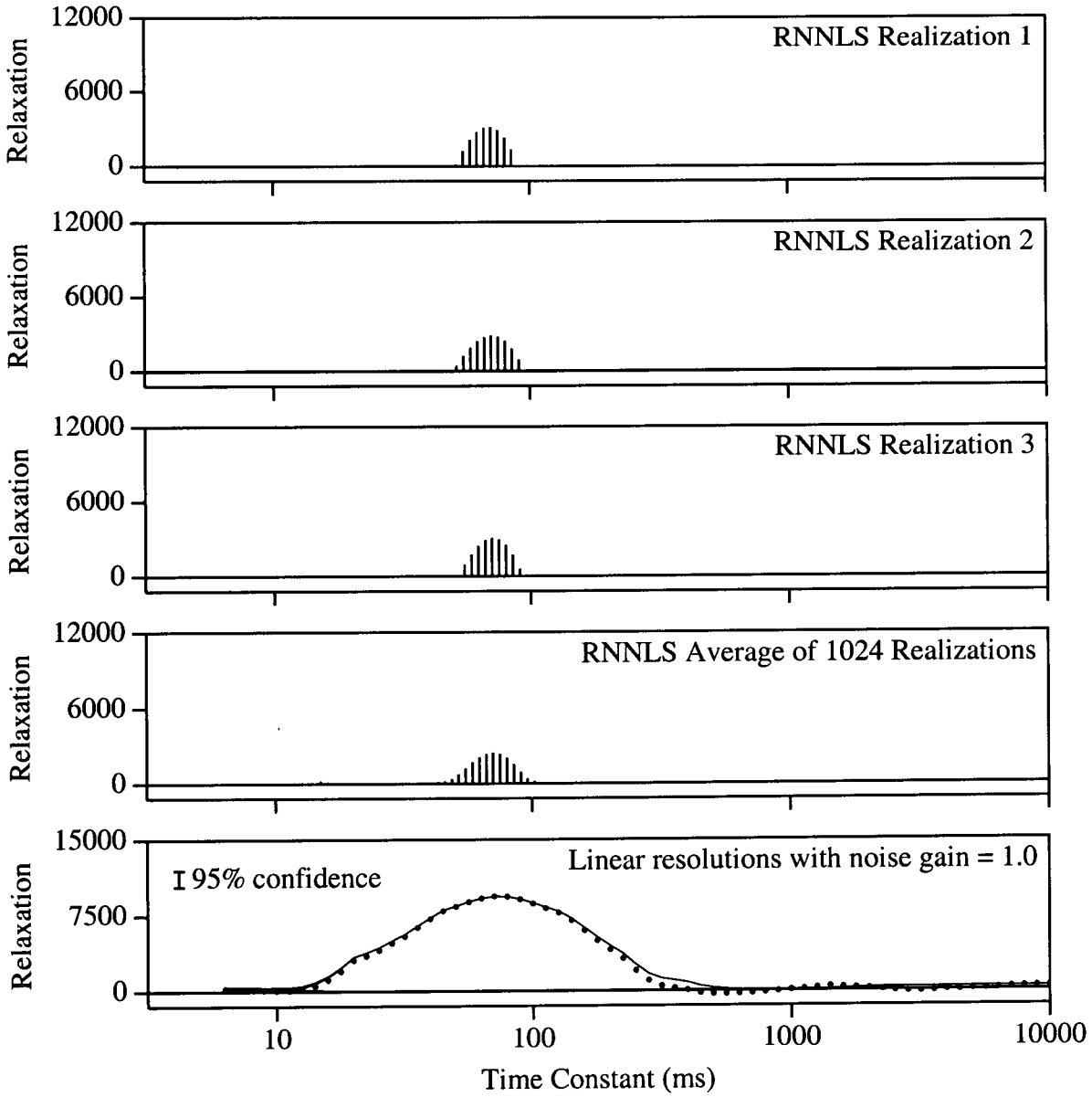

Figure 5.11: White matter model simulation with 50:1 signal to noise ratio. An estimation matrix with noise gain of 1 was used for the linear resolution estimate.

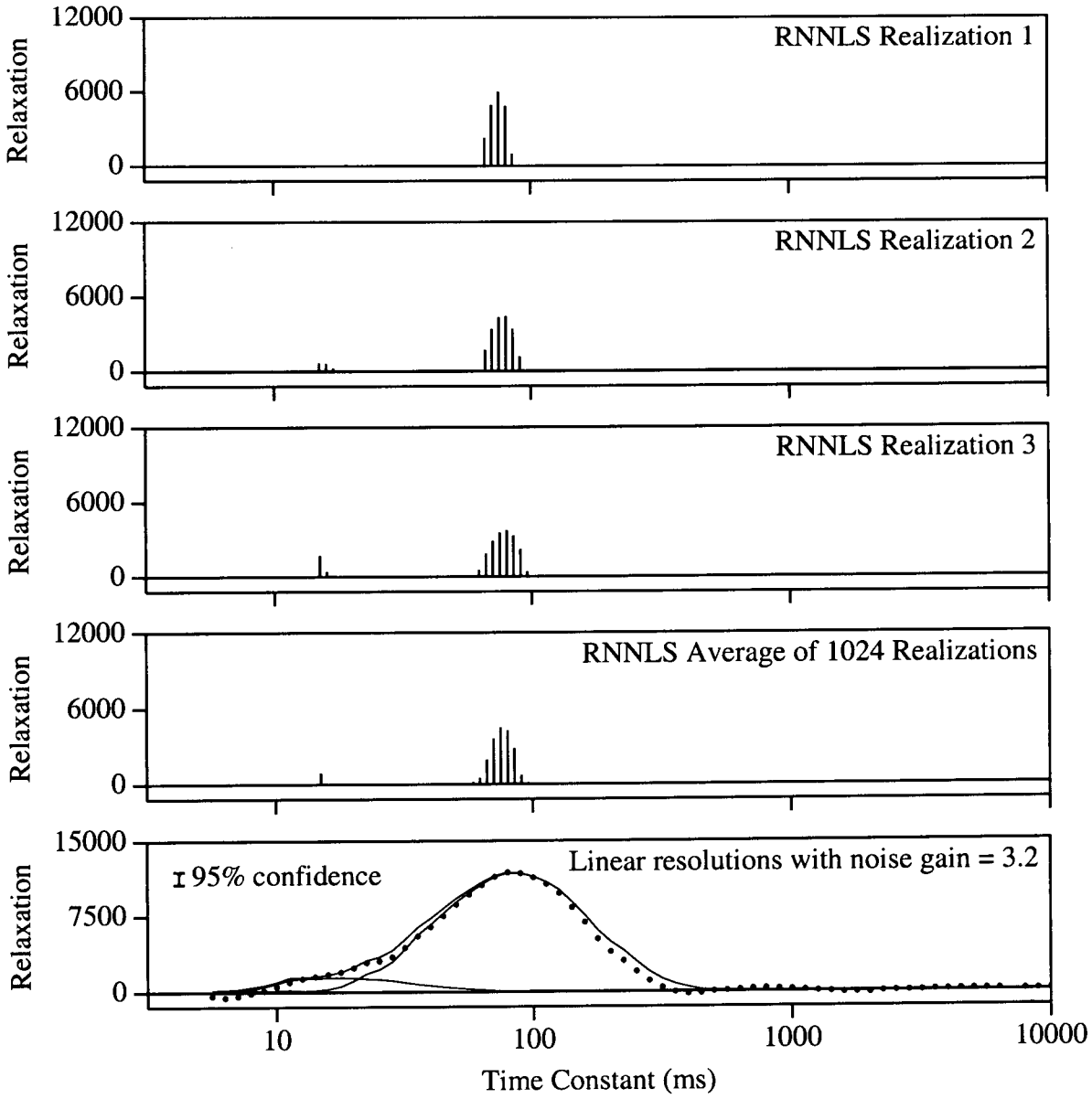

Figure 5.12: White matter model simulation with 200:1 signal to noise ratio. An estimation matrix with noise gain of 3.2 was used for the linear resolution estimate.

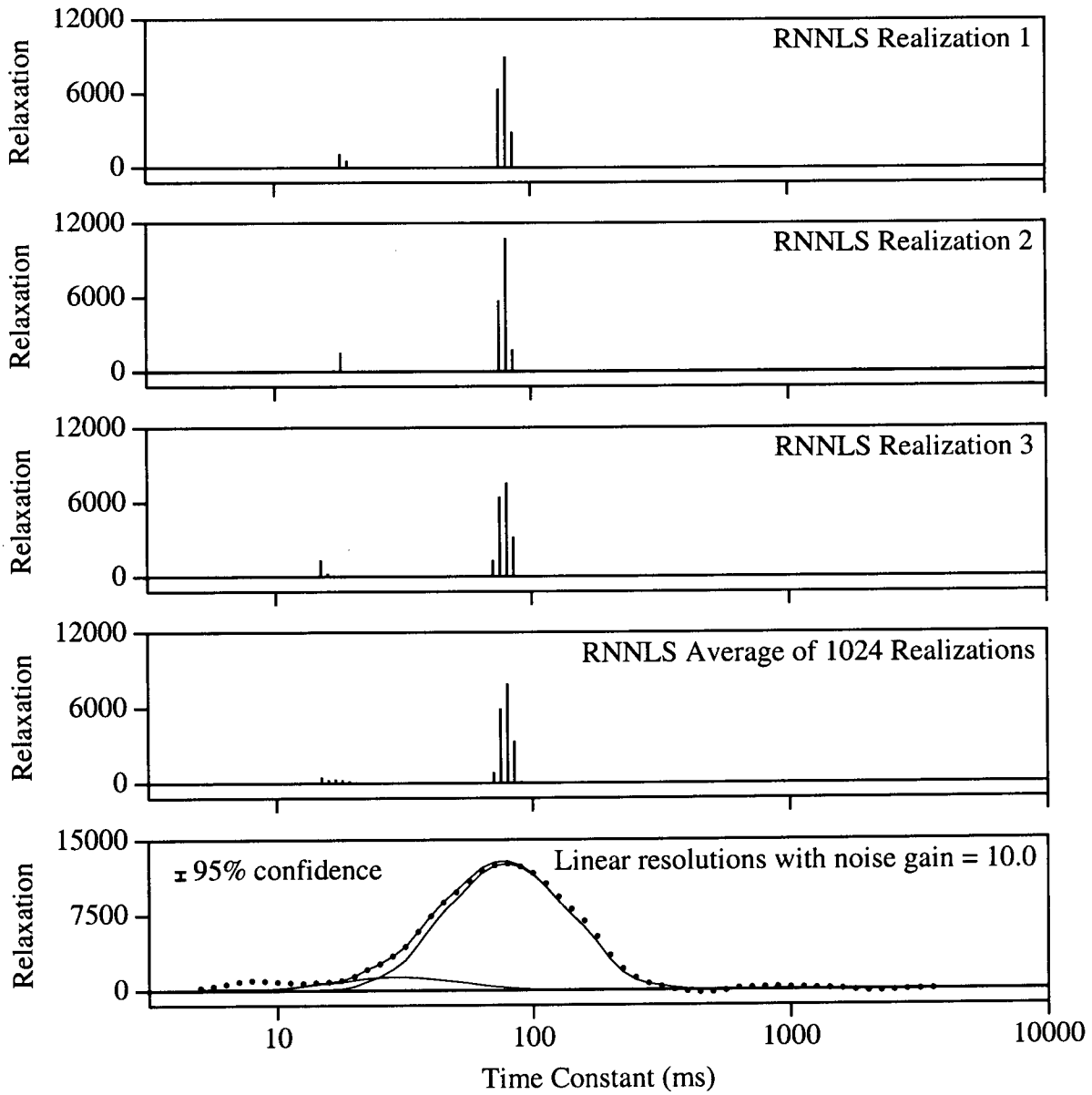

Figure 5.13: White matter model simulation with 1000:1 signal to noise ratio. An estimation matrix with noise gain of 10.0 was used for the linear resolution estimate.

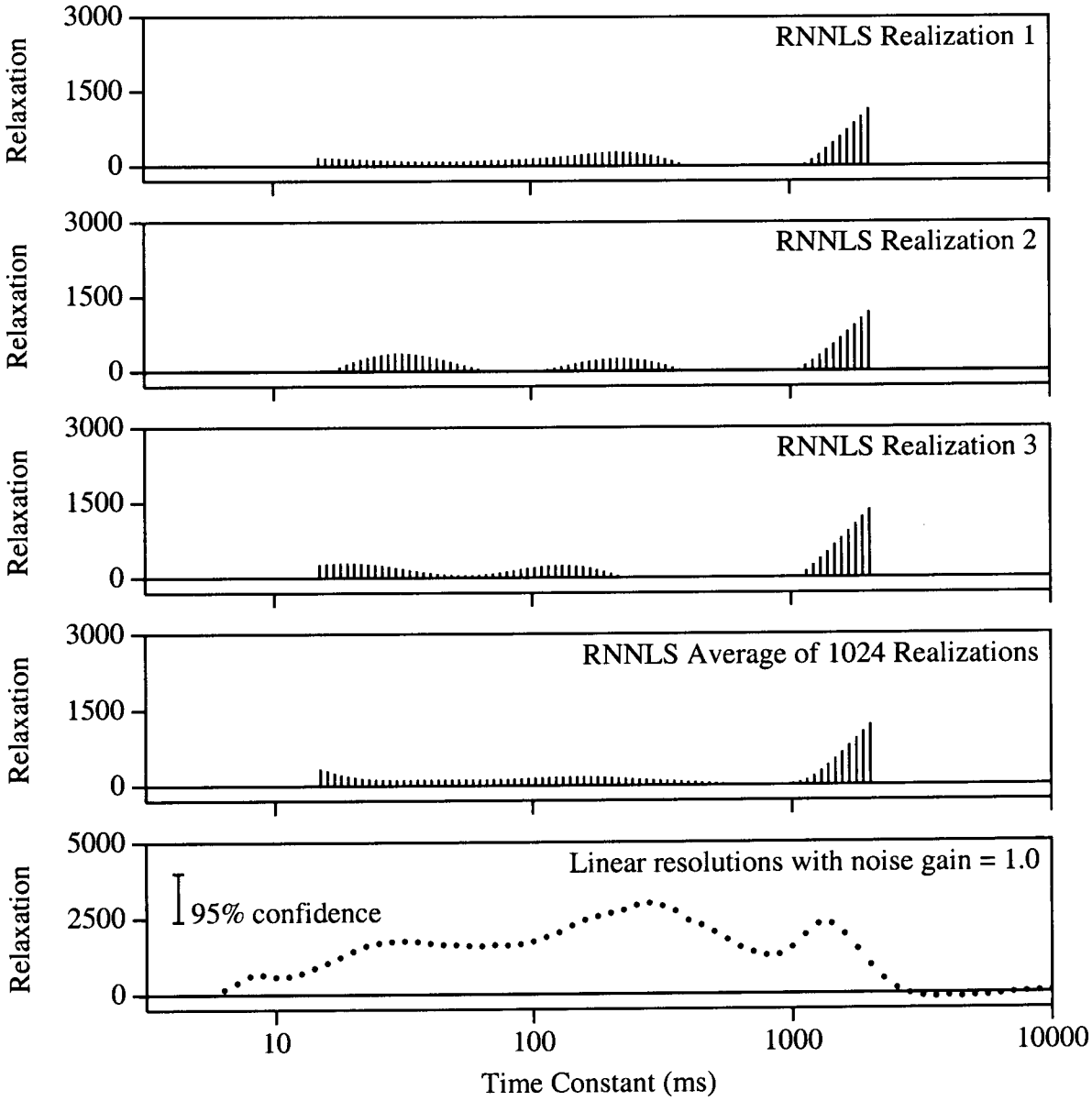

Figure 5.14: Flat model simulation with 50:1 signal to noise ratio. An estimation matrix with noise gain of 1.0 was used for the linear resolution estimate.

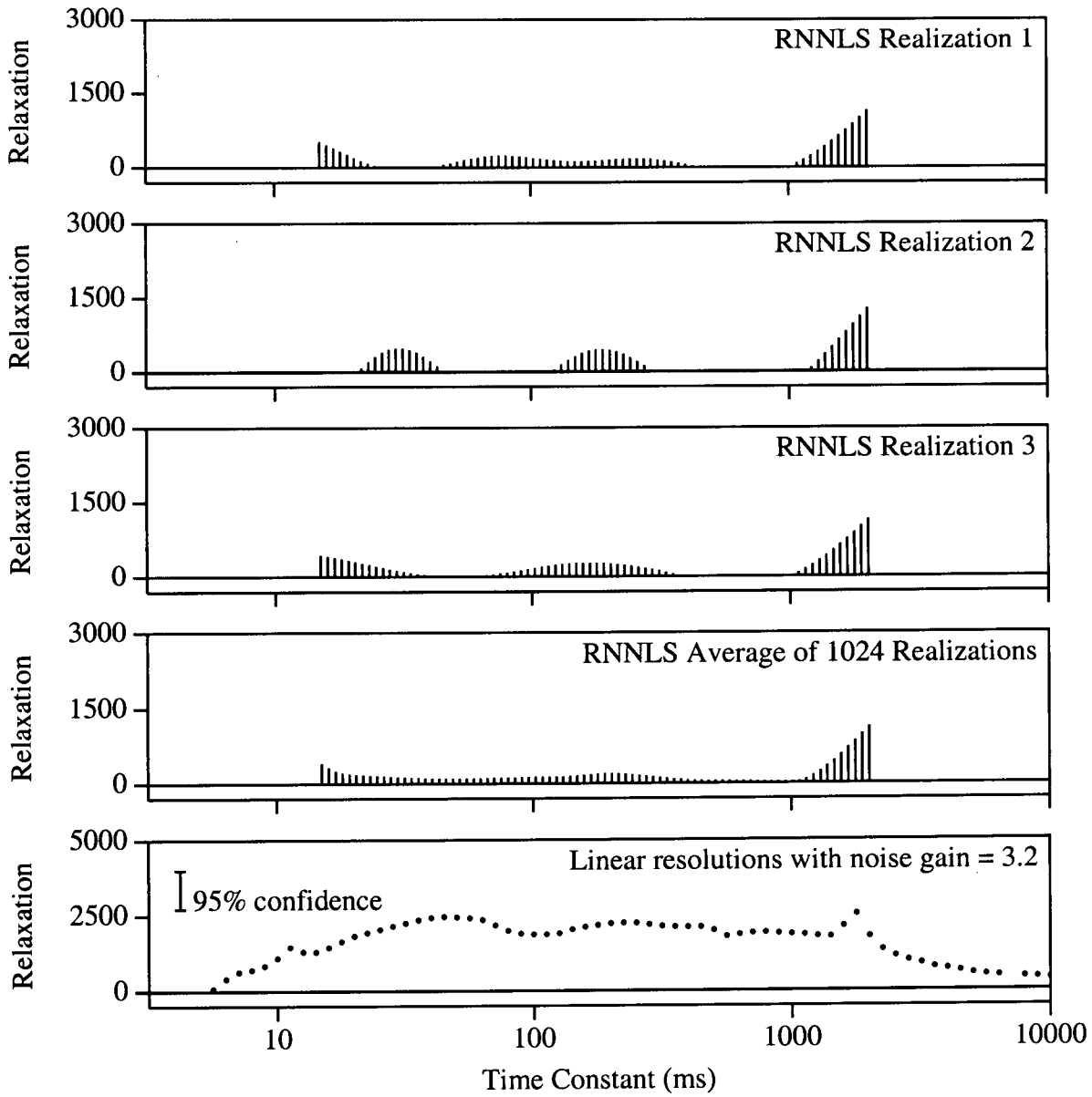

Figure 5.15: Flat model simulation with 200:1 signal to noise ratio. An estimation matrix with noise gain of 3.2 was used for the linear resolution estimate.

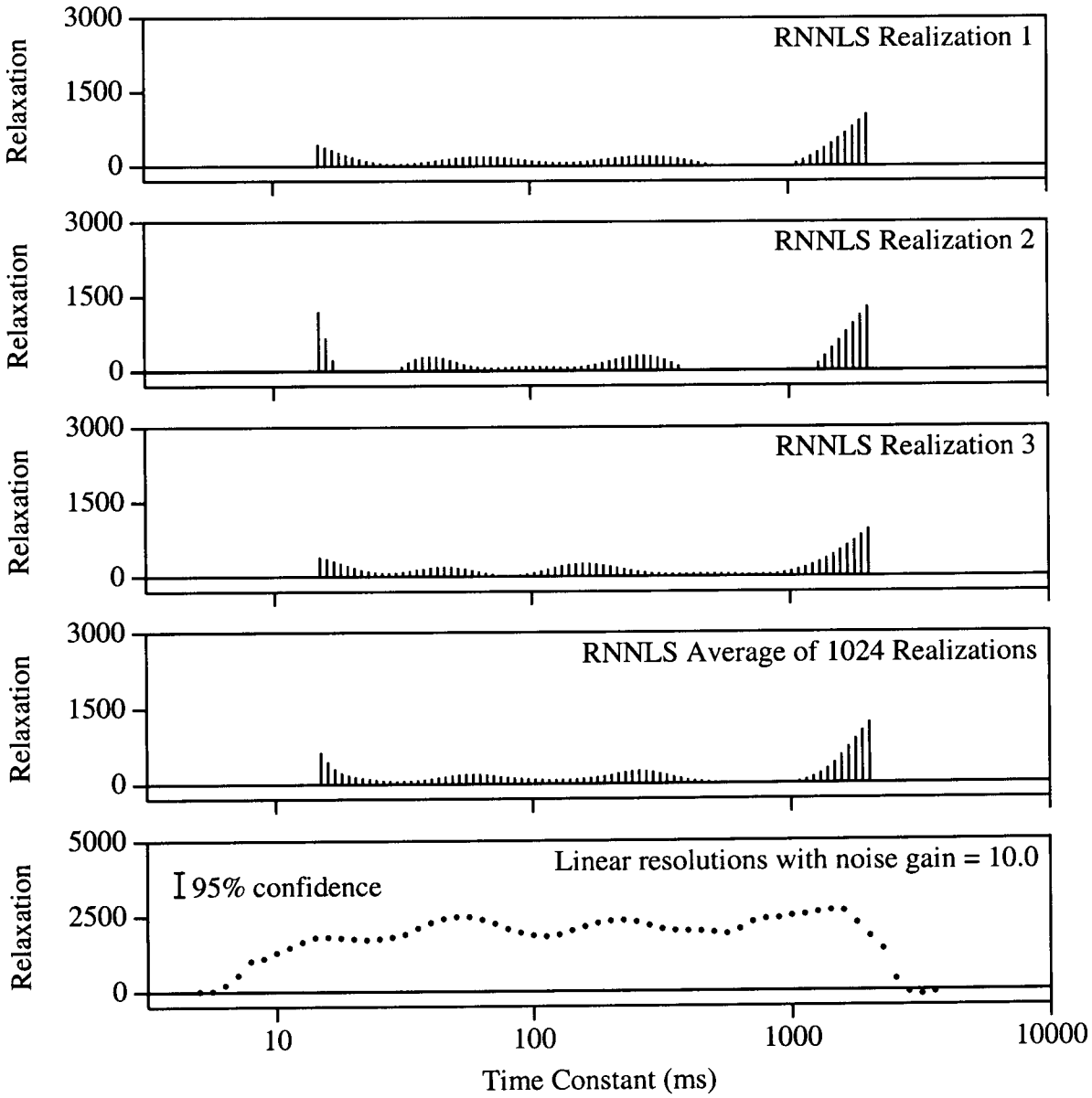

Figure 5.16: Flat model simulation with 100:1 signal to noise ratio. An estimation matrix with noise gain of 10.0 was used for the linear resolution estimate.

Signal to Noise	RNNLS Myelin Amplitude	LR Myelin Amplitude	LR Myelin Time Constant
50:1	undefined	undefined	undefined
200:1	1124±391±1144	2518±600±2222	18.5±5.5±90ms
1000:1	1766±84±237	2233±241±747	20.6±4.2±9.9ms

Table 5.2: Myelin peak amplitudes for RNNLS and linear resolution as well as the time constant for linear resolution (LR) based on 1024 realizations of the noise for 48 echo sequence. The myelin window on NNLS was from 15ms to 40.4ms. The confidence intervals are for the 50% and 95% percentiles. The undefined entries indicate the results was not reproducible enough to determine a meaningful median or confidence interval.

Signal to Noise	15 - 40.4ms window	15 - 51.8ms window
200:1	1448±519±1390	1732±480±1410
1000:1	1834±135±377	1834±135±377

Table 5.3: Comparison of various myelin peak window sizes on RNLLS. Myelin peak amplitudes for RNNLS based on 1024 realizations of the noise. The confidence intervals are for the 50% and 95% percentiles.

that fit the data for each decay curve are required to communicate to an interpreter what the data tells us and does not tell us about the unknown model. The five models presented are only a small sample of the range that could be generated and it should be kept in mind that one of the five has a large χ^2. While evidence of a signal in the 15ms region is present in all five models, there could be other estimates which fit the data and do not have a signal in the 15ms region. We would need to calculate many more estimates that fit the data before we could be sure the data demanded a signal in the 15ms region.

In contrast, the estimate generated by the invertible estimation matrix, combined

with the resolution functions and noise statistics, confirms the existence of the signal in the 15ms region with a single estimate and some statistical analysis.

Why are there no errors bars on the models that fit the data plots? Error bars are an excellent way to present the nonuniqueness of a single random variable with a Gaussian distribution. They also works well for other single variable distributions that can be summarizes by a mean and a standard deviation. But if the distribution is multi peaked error bars will be much less effective, and perhaps misleading, in communicating in the information.

The χ^2 fit to the four peak model is several times larger than ideal. This indicates there is some inconsistency between the four peak assumption and the data. Vavasour *et al* report good success with this *a priori* information. That is probably because they were dealing with much noisier data. This an example which demonstrates *a priori* information is not always helpful.

For the relaxation distribution the non negative *a priori* information has strong experimental evidence to back it up. Both in NMR relaxation and spectroscopy, a wide range of systems produce positive only distributions. But I think it is telling that in spectroscopy, the standard procedure for estimating spectra is the windowed Fourier transform which preserves linear resolution and does not invoke any non negative assumptions. Any *a priori* information is usually then invoked to analyze the spectrum. An example of this is Provencher's LCModel method (1993). Using an invertible estimation matrix to estimate a relaxation distribution provides a similar option to multiexponential analysis – an option that was previously unavailable.

The simulation of the white matter model and the flat model provide insight into the comparison of the RNNLS and linear resolutions algorithms. While the white matter model is highly simplified it is somewhat indicative of the respective sensitivity of the two algorithms to the ideal myelin peak. The RNNLS algorithm performed slightly better.

But both algorithms took advantage of the knowledge that the model being composed of two peaks to come up with a better estimate of the myelin peak parameters.

One interesting characteristic of the performance of RNNLS is the effect of the change in the number of echoes and the myelin window. From comparing tables 5.2 and 5.3 fewer echoes gave a myelin peak amplitude closer to the correct value of 2000. This is counter intuitive since the less information should not improve the quality of the results. But the later echoes have virtually zero signal and are thus dominated by Rayleigh noise. This Rayleigh noise is apparently confusing the RNNLS algorithms degrading the estimate of the myelin peak amplitude.

The effect of increasing the myelin window upper limit from a value of 40ms to 52ms was examined. Not surprisingly, as table 5.3 shows, the value of the RNNLS myelin peak amplitude increased for a signal to noise ratio of 200:1. But it did not increase for the 1000:1 signal to noise ratio. This is probably due to the 20ms and 80ms peaks being well separated at high signal to noise.

While the white matter model gave an indication of sensitivity of the respective algorithms, the flat model gave an indication of their specificity. The RNNLS algorithm consistently produced a false "myelin peak" at approximately 15ms even through the true model was completely flat. In contrast, the linear resolution estimate was flat with the signal dropping off at each end as would be expected from the resolution functions associated with the respective estimation matrices. This result is consistent with the hypothesis that linear resolution is less likely to give misleading results.

5.5 Conclusions

The relaxation distributions in figures 5.6, 5.7 and 5.9 show a signal in the region of 15ms which corresponds to the regions in the brain with high myelin signal. This is

strong evidence that the signal in the 15ms region is associated with myelin because the relaxation distributions were generated by an invertible estimation matrix with sufficient linear resolution to resolve a signal in the 15ms region from the other signals present. A monoexponential fitted to the myelin signal had an amplitude 12% of the main peak and was located at 12.1ms. This is similar in value to the myelin peaks in the four models that fit the data with good χ^2 fit.

The myelin peak also demonstrated the advantages of estimates which have linear resolution and conserve the information in the data over models the fit the data, in particular. The linear resolution algorithm only required one relaxation distribution to summarize all the models that fit the data. Monoexponentials could then be fitted to the regions of the relaxation distribution of interest and the other areas could be ignored because independence of the information at difference time constants due to the linear resolution. In addition, the simulations demonstrated the the linear resolution method is less likely to see peaks, such as the "myelin peak" detected by RNNLS in the flat model, that are not there.

References

MacKay AL, Whittall KP, Cover KS, Li DKB, Paty DW 1991 in "Abstracts, Society of Magnetic Resonance in Medicine, 1991" pg 917.

MacKay A, Whittall K, Adler J, Li D, Paty D, Graeb D 1994 *In Vivo* visualization of myelin water in brain by magnetic resonance *Magn. Reson. Med.* **31** 673–677

Menke W 1984 *Geophysical Data Analysis: Discrete Inverse Theory* (New York: Academic Press)

Provencher SW 1993 Estimation of metabolite concentrations from localized *in vivo* NMR spectra *Magn. Reson. Med.* **30** 672–679

Vavasour IM, Whittall KP, MacKay KL, Li DKB, Vorobeychik G and Paty DW 1998 A comparison between magnetization transfer ratios and myelin water percentages in normals and multiple sclerosis patients. *Magn. Reson. Med.* **40** 763–768.

Vavasour MV, Whittall KP, Li DKB, MacKay AL 2000 Different magnetization transfer effect exhibited by the short and long T2 components in human brain. *Magn. Reson. Med.* **44** 860–866.

Whittall K P, Bronskill M J and Henkelman R M 1991 Investigation of analysis techniques for complicated NMR relaxation data *J. Magn. Reson.* **95** 221–234

Whittall K P, MacKay A L, Graeb D A, *et al.* 1997 *In vivo* measurement of T2 distributions and water contents in normal human brain. *Magn. Reson. Med.* **38** 35–44

Whittall K P and MacKay A L 1989 Quantitative interpretation of NMR relaxation data *J. Magn. Reson.* **84** 134–152

Appendix A

Lens Grinding Software

During my thesis a great deal of time and effort went into writing software to implement several resolution criteria for various applications. This appendix details several of the procedures I used and developed to aid in the implementation.

All the resolution criteria in this thesis was expressed as a constrained objective function to be optimized. I chose the simple penalty method as the optimization algorithm because it was simple, straight forward and could handle linear and nonlinear constraints. When combined with the conjugate gradient method (Press *et al*, 1992), the simple penalty method was computationally fast and was also efficient it in its use of memory. Memory requirements go as $O(N)+O(M)$ where N is the number of coefficients in a digital lens, which corresponds to the number of data points, and M is the number of points the model space has been discretized into.

A.1 Termination conditions

Termination conditions for optimization code is a difficult problem. Terminating too early will results in the resolution criterion not being properly satisfied. Terminating too late wastes computer time. I leaned toward terminating later rather than earlier because the quality of the digital lenses generated was more important than computer time because the main purpose of this thesis was to assess the usefulness of digital lenses.

There were 3 nested converging sequences that required termination criteria; (1) the simple penalty method, (2) the conjugate gradient method (Press *et al* 1992), and (3)

the 1D optimization (Press *et al* 1992). The normal criterion for terminating the 1D optimization was if the functional value did not decrease by a certain fraction each step. The fraction was usually set to 10^{-8}. There was also a step limit of 1000 on each 1D optimization to prevent infinite or unproductive near infinite loops.

Both the conjugate gradient and simple penalty methods used a slightly more complicated termination criterion than the 1D optimization. They terminated when the function value failed to improve by a certain factor for a specified number of consecutive steps. Requiring a consecutive number of failures-to-converge should make the convergence more robust under roundoff error. The conjugate gradient method used a factor of 10^{-7} and a limit on consecutive failures of 25. The total number of iterations was also limited but this limit was application dependent. The simple penalty method had a factor of 10^{-6} and a limit on consecutive failures of 4.

A.2 Objective Functions

The most general form of the resolution criterion objective function, as given in Chapter 2, is

$$Q = \int_a^b W(x)|R(x)|^r dx. \tag{A.1}$$

where $W(x)$ is the weighting function and $R(x)$ is the resolution criteria. Some resolution criteria may drop the absolute value sign around the resolution function. When the resolution function is discretized it can be expressed as

$$Q = \sum_k W(x_k)|R(x_k)|^r. \tag{A.2}$$

The above equation was used to calculate the objective function for most of the applications in this thesis.

If $r = 1$ and the resolution function is constrained to be nonnegative or does not require the absolute value of the resolution function, then the objective function can be

calculated directly from the coefficients of the digital lens, a_j, via the equation

$$Q = \sum_j W_j O_j a_j. \tag{A.3}$$

where W_j is calculated by

$$W_j = \sum_k g_{jk} W(x_k) \tag{A.4}$$

and O_j is calculated by

$$O_j = \sum_k g_{jk} \delta_{kk}. \tag{A.5}$$

The matrix g_{jk} is the discretized form of the data functions.

A.3 Smoothing the Objective Functions

Finding the minimum of equation A.1 can be problematic if $0 < r < 2$. This problem is because the the first derivative of

$$z = |y|^r \tag{A.6}$$

is discontinuous at y=0 in such a way as to form a notch for $0 < r < 2$. The conjugate gradient method, when applied to the sum of a series of functions with this notch, such as equation A.2, easily gets caught in one or more of the notches. The discontinuity can be confirmed by differentiating equation A.6 for $y < 0$ and $y > 0$.

This problem was remedied for $0 < r < 2$ by replacing the notch for small value of y with the function

$$z = P * y^2 \tag{A.7}$$

where P is the penalty amount from the simple penalty method. This replacement function rounds out the bottom of the notch so the conjugate gradient method does not get caught in it. The transition point between equation A.6 and A.7 can easily be calculated by substituting one equation into the other. As the simple penalty method's

penalty, P increases, the range of value of y for which the equation A.7 replaces equation A.6 automatically reduces.

At the transition point between the two functions, the replacement function is steeper than the original function. This insures the conjugate gradient method does not get caught at the transition. The replacement function should always be steeper than the original function at the transition point. If $r > 2$ then the replacement function should use a power which is positive integer multiple of 2 which is large enough to insure the replacement function is always steeper than the original function at the transition point.

This procedure dramatically improved the speed and stability of the convergence of the objective function. For example, the truncation artifacts problem considered in chapter 2 has only linear constraints and a linear objective for some of the resolution criteria. Therefore, the digital lens could have been solved using the simplex method, the standard algorithm for solving linear programming problems (Press *et al* 1992). I considered that possibility and found it to be grossly inefficient. For example, the memory requirements for the simplex method would be O(NM). In contrast, the simple penalty implemented in this thesis required $O(N) + O(M)$ as mentioned earlier in this chapter.

A.4 Algorithmic Efficiencies

Several special properties common to all the resolution criteria were exploited to speed up the code.

The linear transform for the trial digital lens to the trial resolution function must be computed a large number of times during an optimization which is an computer intensive process. The relationship between the digital lens and the resolution function is always linear. This linearity can be used to greatly reduce the number of times the resolution function has to be directly calculated from the digital lens.

The conjugate gradient method reduced the multidimensional optimization problem down finding the lowest value of the objective function along a series of lines in parameter space. The scalar parameter c was varied along a line with direction d_n^a. Therefore, for the trial digital lens a_n, the line was

$$a_n = a_n^0 + cd_n^a. \tag{A.8}$$

Because of the linearity, a similar equation could be written for the corresponding trial resolution function R_m,

$$R_m = R_m^0 + cd_m^R \tag{A.9}$$

where

$$R_m^0 = a_n^0 g_{nm} \tag{A.10}$$

and

$$d_m^R = d_n^a g_{nm}. \tag{A.11}$$

Therefore, along each line of optimization, the linear transform from the trial digital lens to the trial resolution function only needs to be calculated twice, instead of once for every value of c considered. This can result in a many fold reduction in the conjugate gradient calculation time.

References

Fletcher R 1987 *Practical Methods of Optimization* (Toronto: John Wiley & Sons)

Press W H, Teukolsky S A, Vetterling W T, Flannery B P 1992 *Numerical Recipes in C* (Cambridge: Cambridge)